Sustainable Development Goals Attainment Prediction: A Hierarchical Framework using Time Series Modelling

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Yassir Mosaed Alharbi

June 2023

Dedication

To my Father. **Mosaed Mobarak Alharbi**. May your soul rest in the heavens.

Acknowledgements

At the final stage of my PhD journey, I feel proud of myself and extremely grateful to others. I would like to acknowledge the support of many individuals who have helped me on my journey and without whose support I could not have completed this endeavour. In particular I would not have been able to get to this stage without the support of my PhD supervisor, Professor Frans Coenen, who supported and encouraged me throughout my PhD journey, offering fundamental guidance, making valuable suggestions for my research, and providing critical feedback on my thesis. My wife, Feryal Alharbi, who deserves my deepest gratitude. Without her support and encouragement, finishing this journey would have been impossible. My sons Ammar and Raef, they are the infinite source of joy and laughter in my life, and for that, I am very proud and grateful to have them in my life. To my mother, Fatemeh, and my eldest brother, Sami, who made me into who I am today. And for that, I will be in forever in their debt.

Abstract

This thesis presents work focused on finding an effective mechanism for Sustainable Development Goal (SDG) attainment predictions for geographical entities. Each SDG has several targets, and each target has several indicators to predict. The motivation is the desire to utilise the published SDG data in a bottom-up, hierarchical classification based on time series modelling, to predict which geographical entities will meet their SDGs on time, if ever. The main research question that this thesis seeks to address is *"How can the tools and techniques of machine learning be harnessed to effectively and efficiently conduct attainment prediction in the context of the UN Sustainable Development Goals?"*. Three frameworks are proposed and evaluated across 38 geographical entities, spanning four geographical regions. The 38 geographical regions included in this thesis have a total of 200, 742 individual observations, each representing a single time series data point. These observations were collected and compiled from the data sets of all 38 regions. To facilitate data management and analysis, the data set was transformed, with each column corresponding to a unique time series and the index denoting specific points in time (Year). This transformation effectively converted the 200,742 individual observations into 36,421 unique time series. The first approach, SDG-AP, assumed that time series were unrelated and independent, so a univariate time series forecasting approach could be adopted. A number of univariate forecast models were considered. The second approach, SDG-CAP, tested the hypothesis that intra-entity relationships existed between SDG indicators within the context of a single geographic region, and hence a multivariate time series forecasting approach could be adopted, which would produce better SDG attainment predictions than those produced using SDG-AP. A number of approaches for identifying such relationships were considered. Finally, the last method, SDG-TTF, tested the hypothesis that both intra- and inter-entity relations between the different time series could be found and that utilisation of these relationships for multivariate time series forecasting would produce a more effective SDG attainment prediction than in the case of the previous two frameworks considered. Root Mean Square Error (RMSE) was used to evaluate SDG-AP as well as Critical Difference Diagram . Whereas the SDG-CAP and SDG-TTF frameworks were evaluated using RMSE, Borda Count and Critical Difference Diagrams.

Contents

List of Figures

List of Tables

List of Abbreviations

The following acronyms and abbreviations are found throughout this thesis:

ACF	Autocorrelation Function
AI	Artificial Intelegence
ALI	Attainment Likelihood Index
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
CSV	Comma-Separated Values
DTW	Dynamic Time Warping
DVI	Distributed Value Imputation
EMA	Estimated Moving Average
FBprophet	Facebook Profit
GC	Granger Causality
GDP	Gross Domestic Product
GTI	Goal Target Indicator
HTS	Hierarchchal Time Series
IS	Individual Series
LASSO	Least Absolute Shrinkage and Selection Operator
LOO	Leave One Out

LSTM Long short-term memory

MA Moving Average

MAPE Mean Absolute Percentage Error

MAR Missing At Random

MASE Mean Absolute Scaled Eerror

MDG Millennium Development Goals

MNAR Missing Not At Random

MW Mann–Whitney U test

NLP Natural Language Processing

NN Neural Networks

ONS Office for National Statistics

PACF Partial Autocorrelation Function

PC Pearson Correlations

PVI Predictive Value Imputation

RMSE Root Mean Squared Error

RNN Recurrent Neural Network

SDG Sustainable Development Goals

SDG-AP The Sustainable Development Goal Attainment Prediction Framework

SDG-CAP The Sustainable Development Goal Correlated Attainment Prediction Framework

SDG-ENS Sustainable Development Goal Ensemble

SDG-TTF The Sustainable Development Goal Track Trace and Forecast Framework

SMA Simple Moving Average

UID Unique ID

UN United Nation

VARIMA Vector Autoregressive Moving Average

WMA Weighted Moving Average

Chapter 1

Introduction

1.1 Overview

In September 2000, at the end of The United Nations (UN) Millennium Summit, world leaders adopted eight Millennium Development Goals (MDGs) to be achieved before 2015. The MDGs are listed in Table 1.1. Most of the specified goals were achieved by most countries [3], and the MDG initiative was declared to be a success. For example, extreme poverty (MDG 1), defined as living on less than $1.25 a day was significantly reduced, dropping from 47% in 1990 to 14% in 2015. Globally, people living in extreme poverty declined from 1.9 billion in 2000, to 836 million by 2015. With respect to universal primary education (MDG 2), primary school enrolment increased from 83% in 2000 to 91% in 2015. The global under-five mortality rate (MDG 4) dropped from 90 deaths per 1000 in 1990, to 43 deaths per 1000 in 2015.

1.	To eradicate extreme poverty and hunger.
2.	To achieve universal primary education.
3.	To promote gender equality and empower women;
4.	To reduce child mortality.
5.	To improve maternal health.
6.	To combat HIV/AIDS, malaria, and other diseases.
7.	To ensure environmental sustainability.
8.	To develop a global partnership for development.

Table 1.1: The eight 2000 Millennium Development Goals (MDGs)

1.	No Poverty.
2.	Zero Hunger.
3.	Good Health and Well-being.
4.	Quality Education.
5.	Gender Equality.
6.	Clean Water and Sanitation.
7.	Affordable and Clean Energy.
8.	Decent Work and Economic Growth.
9.	Industry, Innovation and Infrastructure.
10.	Reduced Inequality.
11.	Sustainable Cities and Communities.
12.	Responsible Consumption and Production.
13.	Climate Action.
14.	Life Below Water.
15.	Life on Land.
16.	Peace and Justice Strong Institutions.
17.	Partnerships to Achieve the Goals.

Table 1.2: The seventeen 2005 Sustainable Development Goals (SDG)

The success of the MDG initiative paved the way for another set of goals. In September 2015, the UN introduced the Sustainable Development Goals (SDG), listed in Table 1.2, to be achieved by 2030 [4, 5]. However, this time the goals covered a broader range of domains. The vision was that the achievement of these goals would provide for a world free from hunger and poverty, and would ensure the sustainability of natural resources and the protection of the environment. The philosophical underpinning for the SDG initiative, and the MDG initiative, was the idea that the world is a connected place, and that all UN members should therefore work together to ensure the attainment of these goals for all member states [6, 1]. The SDG can be categorised into 4 different levels:

Biosphere : Environmental-related goals.

Society : Goals related to empowering society, such as by: (i) eradicating poverty, (ii) promoting health and equality and (iii) promoting sustainable cities.

Economy : Goals related to economic growth using responsible consumption while reducing workforce inequality.

Figure 1.1: Interconnectedness of the UN SDGs as presented by the Azote for Stockholm Resilience Centre, Stockholm University [1]

Partnership : Goal 17, which exists to promote a global effort to ensure that the SDGs are attained in all countries.

By considering the SDG in terms of the above categorisation, it can be seen how the SDG are related and connected as illustrated in Figure 1.1. For example, without a sustained water source and sanitation (SDG 6), growing food will be harder; thus, it will affect the attainability of SDG 2 (Zero Hunger); and therefore SDG 8 (Decent work and economic growth) will not be fulfilled. Given the interconnected nature of the SDG one can see that there is a causal relation between individual SDGs.

Part of the UN strategy with respect to the SDG initiative was the timely release of SDG data with public access. SDG Data is downloadable from the SDG website [1]. The downloadable file is a relatively large (400MB) CSV file that contains data related to over 200 different geographical entities and regional groupings. The data can be used to monitor the attainment likelihood of individual SDGs over time. In other words, the data can be used to forecast whether a particular geographic entity, e_i, in the set of geographic entities, E, will meet its SDGs or not. This is the central theme of the work presented in this thesis.

[1]https://unstats.un.org/sdgs/indicators/database/

Whether e_i meets its SDGs or not can be expressed as a simple Boolean (yes/no) or as an index. The latter would then indicate how close a geographic entity that had not met its SDGs was to meeting them.

Without going into too much detail in this introductory chapter, the published Sustainable Development Goals (SDG) data can generally be conceived as time-series data, as it consists of measurements taken over time to monitor progress towards the 2030 Agenda for Sustainable Development [7]. Therefore, forecasting SDG attainment can be seen as a time series analysis problem. Each SDG has several targets, each consisting of numerous indicators (and in some cases, sub-indicators and even sub-sub-indicators). SDG attainment can thus be perceived as a large hierarchical topography. A fragment of such a topography for Egypt is illustrated in Figure 1.2. Each leaf node is associated with a set attainment threshold to be achieved by a specific date. The challenge of forecasting whether a geographic region will meet its SDGs can be expressed as a hierarchical time series prediction problem [8]. However, it is crucial to distinguish between forecasting a time series and determining if a target will be achieved within a given timeframe; these are two distinct tasks.

Leaf nodes contain time series data; in this instance, indicator 1.1.1 is the focal point containing two different sets of sub-sub-indicators: SI_POV_DAY1 and SI_POV_EMP1. Each of these indicators is further divided into different time series. For example, the indicator SI_POV_EMP1 will contain time series about the 'Employed population below the international poverty line, by sex and age (%)'. Achieving this specific goal necessitates two distinct operations: firstly, forecasting the time series as accurately as possible, and secondly, determining based on the forecasted values and the target thresholds whether the goal will be met or not. For this particular indicator, the target date is set for 2030, with the aim to end poverty. Therefore, a target is considered achieved if, by 2030, the forecasted value is less than 0.05%.

There are other application domains where data can be represented and utilised in this manner. For instance, store level forecasting can be used by a retail company that owns a chain of stores with the aim of boosting sales. This involves augmenting sales per store, subsequently per region, and finally throughout the higher levels. Another example is stock market index prediction, such as the Dow Jones Industrial Average (DJIA), a widely used benchmark index in the USA. Figure 1.3 shows a fragment of the Dow Junes hierarchical topography

Referencing the hierarchical topographies displayed in Figures 1.2 and 1.3, it is evident

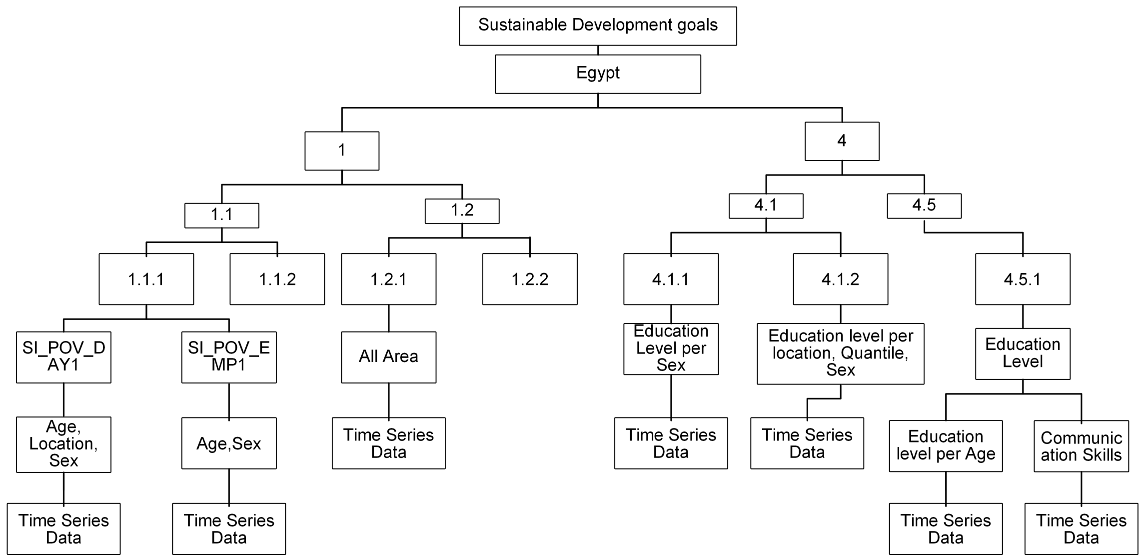


Figure 1.2: A fragment of the SDG hierarchical topography for Egypt

that data from each level contributes to the values of the upper tiers. In the context of the Dow Jones hierarchical topography shown in Figure 1.3, the value of Boeing Commercial Air planes (BCA) contributes to Boeing's value, which influences the Dow Jones Aerospace sector's value, thereby impacting the overall Dow Jones Index. In the case of the SDG hierarchical topography shown in Figure 1.2, whether a sub-sub-indicator is realised or not influences whether a sub-indicator is achieved, and so forth.

Regardless of whether we wish to predict the Dow Jones Index or the likelihood of a geographic entity meeting its SDGs, in both cases the leaf nodes of the hierarchical topography will hold data which, when assigned, can be used to "populate" the rest of the hierarchy. In the case of the SDG hierarchical topography the data at the leaf nodes will be in the form of time series. A mechanism is therefore required for predicting values at the leaf nodes which can then be passed up the hierarchy so as to populate the rest of the hierarchy, level-by-level, until the root node is reached.

So, the challenge here boils down to predicting a hierarchical time series. While this thesis mainly concentrates on achieving SDGs, it's important to note that this approach of prediction has a wide range of uses.

Forecasting SDG attainment is challenging because each time series in the data set consists of only 22 observations from 2000 to 2021. In other words, the time series are

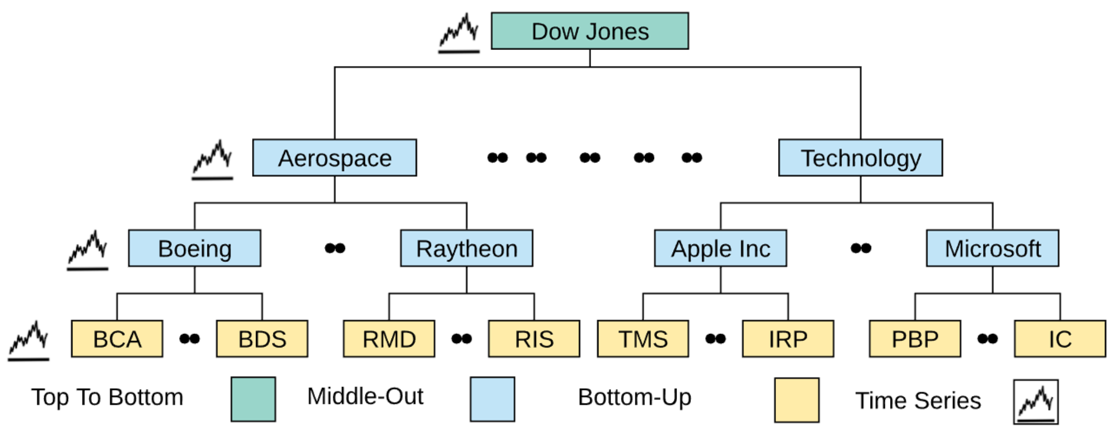


Figure 1.3: A fragment of the Dow Jones Hierarchical topography

very short. This challenge is aggravated by the observation that few SDG time series are complete with all 22 observations, usually because of delays in recording the data. In addition, many SDG time series feature gaps (sequences of missing values). These missing values can be categorised as either: (i) Missing-At-Random (MAR) or (ii) Missing-Not-At-Random (MNAR) [9]. In the case of MAR values, there is no obvious explanation as to why the data is missing. One example of this can be found with respect to Target 1.1.1, "Proportion of population below international poverty line (%)" for the geographic entity Egypt where data is missing with respect to every year except for the years 2003, 2007 and 2009. Another example where we might have MAR data is as a consequence of natural disasters, events that do not usually happen and are therefore unexpected. In the cases of MNAR values, the gaps are regular and caused by the sequence in which data is collected. For example, forestation data is frequently gathered over five year cycles. This is typically conducted using satellite imagery [10] whereby "snapshots" separated by five year intervals are compared and the level of forestation (or deforestation) measured. Thus a five years data collection cycle.

Given the above discussion, SDG attainment prediction or forecasting can be conceptualised as a bottom-up hierarchical time series problem. In other words, what is required is a hierarchical classification model. However, unlike established hierarchical classification systems, which work in a top down manner [11], the classification model needs to work in a bottom up manner. In both cases, the objective is to establish the class label, or some value, of an entity with respect to some predefined hierarchical taxonomy. In both cases, the classification will operate in a level-by-level manner. However, the branches in the taxonomy in the top down case represent disjunctions, while the branches in the bottom up case represent conjunctions. In the top down case, the path in the hierarchy from the root node, which is in this case Goal 1, to the target levels then to the indicator levels and then to the leaf nodes which hold a time series. This is illustrated in Figure 1.4(a) where a classification path is highlighted in bold lines.

In the bottom up case, labels associated with the leaf nodes need to be established before labels associated with parent nodes can be established, all the way up to the root node as shown in Figure 1.4(b). In other words, time series need to be forecasted first, then converted into attainment labels. The hierarchical taxonomy used in the case of bottom up hierarchical classification can thus be thought of as a "dependency tree" [12]. An alternative way of differentiating the two approaches is to describe top down hierarchical classification as adopting a "coarse-to-fine" classification approach, whilst bottom up hierarchical classification adopts a "fine-to-coarse" classification approach. It should also be noted that top down hierarchical classification was originally proposed as a mechanism for addressing classification problems that featured a large number of classes. Techniques for top down hierarchical classification are well established, techniques for bottom up hierarchical classification have been less well studied. In SDG time series data, yearly time-stamped observations only exist at the leaf nodes. While the upper nodes only hold a hierarchical identifier. Consequently, a bottom-up approach is the only viable way to predict indicators, target, and goal attainment. The generation of such bottom up hierarchical classification models, focused on SDG attainment prediction, is thus the central challenge that this thesis seeks to address.

1.2 Motivations

Given the foregoing, the motivation for the work presented in this thesis is the desire to be able to predict which geographic entities will meet their SDG using the tools and techniques of machine learning. More specifically to use some kind of bottom up, time series-based, hierarchical classification model to achieve this. This entails three principal challenges. The first is the size of the data with 260 different geographical entities and some six hundred time series for each entity ($260 \times 600 = 156{,}000$ time series). The second is that the time series available are very short, a maximum of 22 observations for each time

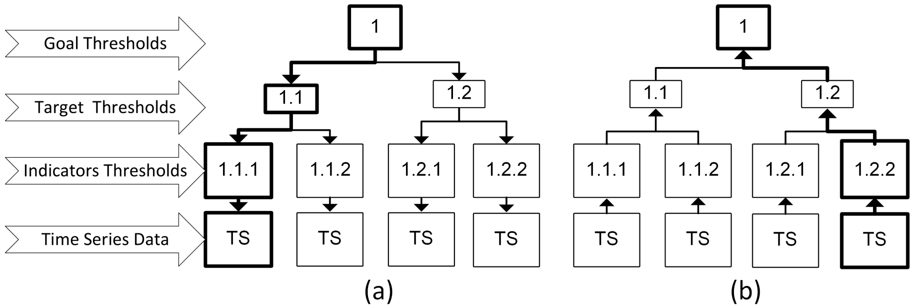


Figure 1.4: Hierarchical Classification; (a) Top Down, (b) Bottom Up.

series; although in many cases the time series are shorter and/or feature missing values. The third is the nature of the prediction; there are four options:

1. Treat each item set in isolation, the assumption is that there is no dependency (correlation) between individual time series.
2. Assume that there are correlations between time series pertaining to a particular geographic entity and that these correlations can be used to enhance SDG attainment prediction.
3. Assume that there are correlations between time series pertaining to a particular SDG (across geographic entities), and that these correlations can be used to enhance SDG attainment prediction.
4. A combination of 2 and 3.

1.3 Research Questions

The primary research question that the PhD thesis intends to provide an answer to is:

How can the tools and techniques of machine learning be harnessed to effectively and efficiently conduct attainment prediction in the context of the UN Sustainable Development Goals?

The provision of an answer to this primary research question will entail the resolution of several subsidiary questions:

S1. What is the most efficient and effective way to derive a taxonomy for SDG data?

S2. How best can machine learning be used to forecast whether individual SDGs will be met?

S3. Assuming a hierarchical taxonomy, how can a prediction label be derived for the root node of the hierarchy?

S4. Is it possible to generate more sophisticated machine learning methods to be held at the leaf nodes in the hierarchy by combining data using different instances?

S5. Can some form of feature combination be applied to improve overall forecasting accuracy?

S6. Can we identify an effective mechanism than a simple Boolean yes/no for establishing when a goal will be reached, if ever?

1.4 Research Methodology and Evaluation

The preceding discussion suggests that using a single approach will not provide an answer to all the research questions. A four-stage work plan was therefore adopted featuring four Work Packages (WPs) for the research:

WP1. Taxonomy generation.

WP2. Forecasting individual SDG attainment.

WP3. Forecasting SDG attainment in the context of intra-entity correlations.

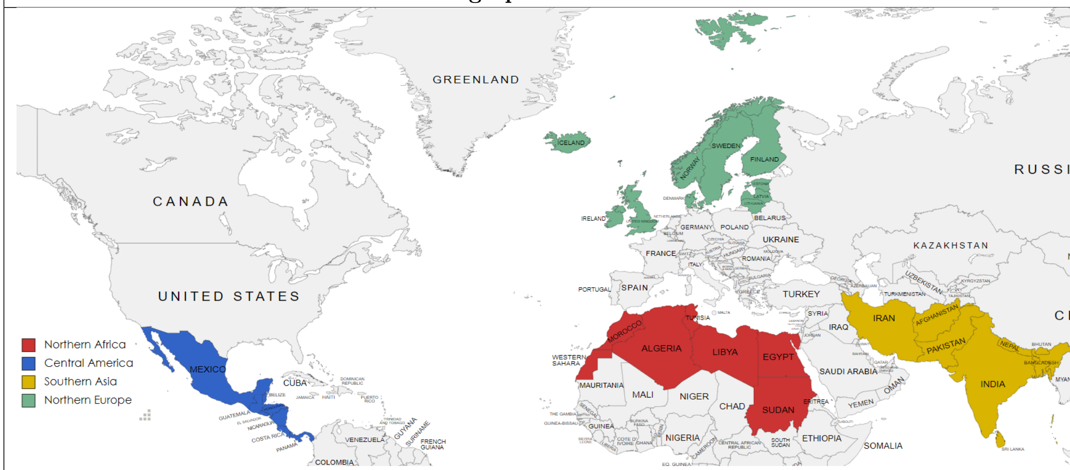
WP4. Forecasting SDG attainment in the context of intra- and inter-entity correlations.

Each work package was designed to provide an answer, or at least a partial answer, to one or more of the subsidiary questions listed above (S1 to S6); and consequently to provide an answer to the primary research question. Each work package is discussed in some further detail later in this section in Sub-sections 1.4.1 to 1.4.4.

The SDG data contains a total of 260 geographical regions, comprised of 156,000 time series. To act as a focus, and for detailed evaluation purposes, four geographic regions were considered, each comprised of a number of geographic entities: (i) Northern

Europe, (ii) Southern Asia, (iii) North Africa and (iv) Central America. These regions were chosen because of their diverse geographical locations also as to include wide range of economic state, geographical locations, and development levels. The total number of time series within those four geographical regions was 36,421. Further detail concerning these geographic regions is given in Table 1.3 the main evaluation metric used throughout the thesis was Root Mean Squared Error (RMSE).

Geographical Entities



#	Northern Europe	Southern Asia	North Africa	Central America
1	Aland Islands	Afghanistan	Algeria	Belize
2	Denmark	Bangladesh	Egypt	Honduras
3	Estonia	Bhutan	Libya	Guatemala
4	Faroe Islands	India	Morocco	Mexico
5	Finland	Iran	Sudan	Nicaragua
6	Isle of Man	Maldives	Tunisia	Panama
7	United Kingdom and Northern Ireland	Nepal	Western Sahara	Costa Rica
8	Iceland	Pakistan		El Salvador
9	Ireland	Sri Lanka		
10	Latvia			
11	Lithuania			
12	Norway			
13	Sweden			
14	Svalbard and Jan Mayen Islands			

Table 1.3: List of the Geographic Regions used as a focus in this thesis

1.4.1 Taxonomy Generation (WP1)

The first Work Package, WP1, was taxonomy generation. The concept of an SDG taxonomy was introduced earlier in this chapter; a fragment of this taxonomy was given in Figure

1.2. The taxonomy generation work package was directed specifically at subsidiary research questions S1, S3 and S6. Prior to investigating taxonomy generation two initial challenges first needed to be considered:

1. Although the SDG raw data contained information about all goals for all countries, there were no Unique IDs (UIDs) for each time series. Without a unique ID, tracking the progress of an exact leaf node time series would be challenging, especially if the goal was to compare the progress of two identical series from two geographical regions.

2. In many cases the data features an absence of a specific threshold values for when a goal can be considered to have been attained.

The final intended deliverable of WP1 was a fully operational SDG taxonomy structure generation mechanism which could be populated using the techniques investigated in later work packages and ultimately be used to make SDG attainment predictions for given geographic entities.

1.4.2 Forecasting Individual SDG Attainment (WP2)

Work Package WP2 was designed to address Subsidiary Questions S2 and S3. The most straightforward approach to SDG attainment prediction given an entity $e_i \in E$ (where E is the global set of geographic entities) is to assume each time series is independent. Time series analysis can be conducted in either a univariate and multivariate, of which the first is the simplest to implement. The independence assumption allowed for univariate time series prediction; in other words, for WP2, a univariate approach was adopted. The fundamental idea was to investigate the usage of some form of regression applied to each time series associated with a leaf node in the taxonomy and then use the results from the generated regression models to populate the taxonomy to consequently to produce a final prediction. The intention was to consider a number of different regression mechanisms.

1.4.3 Intra-Entity Approach (WP3)

The third work package, WP3, was the first of two work packages designed to provide answers to Subsidiary Questions S4 and S5. The intuition here was that, as noted above, the univariate approach assumed no relationship between the time series associated with an entity $e_i \in E$ (where E is the global set of geographic entities). If we consider a target such as: “Number of Durable Structures” and “By 2030, significantly reduce the number of deaths and the number of people affected and substantially decrease the direct economic losses relative to global gross domestic product caused by disasters, including water-related disasters, with a focus on protecting the poor and people in vulnerable situations”, we can see that the associated time series clearly have an impact on each other. The identification mechanism to identify these relationships for all time series $T_i = \{t_1, t_2, \dots\}$ associated with e_i was thus the objective of WP3. The intention was to compare each time series $t_1 \in T$ with all other time series in T, identify all possible relationships, select the top k time series that were most closely related to the target time series and then use this collection of time series to build a prediction for the leaf node in the taxonomy represented by t_i. In this manner, for example, relationships were found between SDG 1 (“No poverty”), SDG 2(“End hunger”) and SDG 3 (“Good Health and Well-being”). The intuition was that better predictors might be produced by grouping time series into multivariate forecasting scenarios. The intention was to consider the following techniques for identifying correlations between SDGs:

1. Pearson Correlation [13]
2. Least Absolute Shrinkage Selector Operator (LASSO) regression [14].
3. Granger Causality [15].
4. Mann-Whitney U-test [16]
5. Dynamic time warping [17].
6. An ensemble technique combing the above five.

1.4.4 Inter-Entity Approach and Ali Index (WP4)

Work Package WP4 consists of two parts. The first, as in the case of WP3, was directed at Subsidiary goals S4 and S5. The second was directed at Subsidiary goal S6. With respect to the first the intuition was that time series associated with a geographic entity e_i might behave similarly, or be related to time series associated with neighbouring entity e_j. A better SDG attainment prediction mechanism might, therefore, be built if using time series data from correlated geographic regions. WP4 was therefore designed to investigate this potential. The second part of WP4 was motivated by the observation that where a geographic entity $e_i \in E$ is predicted to not attain its SDGs it would be good to know by how far. The idea was that instead of a simple Boolean (True/False) attainment outcome, it would be better if some likelihood index was produced, an SDG Attainment Likelihood Index (ALI). This would then, given an entity e_i which was predicted to not attain its SDGs, an indication of how far off the entity was from SDG attainment. The aim of the final part of WP4 was to investigate this idea.

1.5 Contributions

This thesis makes a number of significant contributions. These are itemised in this section as follows:

The Sustainable Development Goals Taxonomy and Thresholds. A data set, and mathematical conversion of the SDG textual targets, that anyone conducting study of SDG attainment can readily use or modify for their own research purposes.

The Sustainable Development Goals Attainment Prediction (SDG-AP) Framework. A framework that employs a univariate method to forecast the SDG attainment designed as a benchmark with which alternatives can be compared.

The Sustainable Development Goals SDG-ENS A mechanism whereby time series within geographical entities can be compared.

The Sustainable Development Goals Correlated Attainment Predictions (SDG-CAP) Framework. A framework that uses dependent time series within the current geographic entity, intra-entity dependencies, to create multivariate forecast models for SDG prediction.

The Sustainable Development Goals Track, Trace and Forecast (SDG-TTF) Framework. A framework that uses dependent time series both within the current geographic entity and across neighbouring entities, intra- and inter-entity dependencies, to create multivariate forecast models for SDG prediction.

The Attainment Likelihood Index. A mechanism whereby the likelihood of a geographic entity attaining its SDGs can be measured.

1.6 Published Work

Many of the main contributions of this thesis have been published in relevant peer-reviewed conferences, journals and workshops. A list is provided in Table 1.4.

#	Title	Type	Venue	Related Chapter
1	Sustainable Development Goal Attainment Prediction: A Hierarchical Framework Using Time Series Modelling [18]	Workshop	IJCAI	3,4,5
2	Sustainable Development Goal Attainment Prediction: A Hierarchical Framework using Time Series Modelling	Conference	IC3K	3,4,5
3	Sustainable Development Goal Relational Modelling: Introducing the SDG-CAP Methodology [19]	Conference	DAWAK	6
4	Sustainable Development Goal Relational Modelling: Introducing the SDG-RMF Methodology	Workshop	Harvard University	6,7,8,9
5	Sustainable Development Goals Monitoring and Forecasting using Time Series Analysis[20]	Conference	DELTA21	7
6	Sustainable Development Goal Relational Modelling And Prediction: Introducing THE SDG-CAP-EXT Methodology [21]	Journal (invitation from DAWAK)	Journal of Data Intelligence	8,9
7	Forecasting The UN Sustainable Development Goals (Submitted)	Journal (invitation from DELTA)	Communications in Computer and Information Science	8,9

Table 1.4: List of publications

1.7 Thesis Outline

In this section of this introductory chapter the structure of the rest of this thesis is presented as follows.

Chapter 2: Literature Review. The second chapter presents the necessary background and related work concerning the subject matter of this thesis. The chapter commences with an overview of research published on SDG attainment prediction and machine learning. This is followed by a review of the relevant previous work with respect to Time Series analysis, causal and correlation inference, time series missing value resolution and hierarchical time series analysis.

Chapter 3: The SDG Data Set. The third chapter of this thesis which presents the SDG data used throughout this thesis to act as a focus for the work and for evaluation purposes. The chapter details the required pre-processing of the data and introduces the proposed SDG taxonomy (the central structure used throughout this thesis for SDG attainment prediction). The chapter covers, in part, Work Package WP1 and is directed at Subsidiary Research Question S3.

Chapter 4: SDG Taxonomy Threshold Derivation and Utilisation. Chapter 4 examines how individual SDG attainment thresholds can be determined and incorporated into the SDG taxonomy from Chapter 3. The chapter also considers the operationalisation of the proposed SDG taxonomy and how results can be visualised. The chapter covers the rest of Work Package WP1, and is directed at providing a final answer to Subsidiary Research Question S3.

Chapter 5: The sustainable Development Goal Attainment Prediction Framework. Chapter 5 presents the first of the prediction frameworks presented in this thesis, the benchmark Sustainable Development Goal Attainment Prediction (SDG-AP) framework, which utilised a univariate approach to SDG attainment prediction (Work Package WP2) where it was assumed that all SDG times series are independent of one another. The chapter reports on several methods for achieving this. The work presented in this chapter is directed at Subsidiary Research Questions S2 and S3.

Chapter 6: The Sustainable Development Goal Correlated Attainment Prediction Framework. Chapter 6 presents the second of the prediction frameworks proposed in this thesis, the Sustainable Development Goal Correlated Attainment Prediction (SDG-CAP) Framework which incorporated intra-geographic region dependencies. The framework incorporated the idea of causal inference within a single geographic entity, intra-entity correlations, with a view of improving the effectiveness of the SDG-AP univariate approach from the previous chapter. The chapter reports on work conducted on Work Package WP3 and is directed at Subsidiary Research Questions S4 and S5.

Chapter 7: Sustainable Development Goals Track Trace and Forecast Framework. Chapter 7 presents the third and final prediction frameworks proposed in this thesis, the Sustainable Development Goal Track Trace and Forecast (SDG-TTF) Framework which incorporated intra-geographic region dependencies. The framework incorporated the idea of causal inference both within a single geographic entity, intra-entity correlations, and across a number of geographic entities, inter-entity correlations. The chapter reports on the investigations undertaken with respect to the idea of a Attainment Likelihood Index (ALI) that ranks entities based on their goals attainment potential. The work presented in the chapter is directed at WP4 and subsidiary research questions S4, S5 and S6.

Chapter 8: Conclusion. Chapter 8, the final chapter of the thesis, provides a closing summary, the main findings regarding the primary research question and the associated subsidiary research questions, and some concluding remarks. The chapter also provides some discussions about possible directions for future work.

1.8 Conclusion

This opening chapter has provided the background and motivation for the work presented in this thesis. The chapter has also provided the primary research question that the thesis seeks to address and a number of subsidiary research questions, the research methodology adopted to provide answers to the primary and subsidiary research questions, and the main contributions of the work presented in the thesis. The chapter was concluded with an itemisation of the papers resulting from the work described and an overview of the remainder of the thesis. The following chapter is the literature review chapter, where a review is presented of the previous work that underpins the research presented later in this thesis.

Chapter 2

Literature Review

2.1 Introduction

As discussed in Chapter 1, the work conducted in this thesis seeks to explore Sustainable Development Goal attainment prediction using time series forecasting and machine learning. This chapter provides a review of the relevant previous published work. The chapter commences, Section 2.2, by providing some background to the idea of SDG attainment prediction using forecasting and tracking. The subsequent section, Section 2.3, provides a comprehensive review of the existing work on time series analysis within the context of the SDG subject domain of this thesis. Next, in Section 2.4, a discussion on the identification of relationships between time series is presented. The section reviews several methods for finding relationships (causation, correlation, similarity) in SDG data: (i) Pearson Correlation (PC), (ii) Granger Causality (GC), (iii) Least Absolute Selection and Shrinkage Operator (LASSO), (iv)the Mann-Whitney U T-test (MW) and (v) Dynamic Time warping (DTW). This is followed by a review of the various techniques which can be used to address the missing values problem, which is a feature of the SDG data, in Section 2.5. As noted earlier in Chapter 1, the solutions presented in this thesis to address the SDG attainment prediction problem are founded on the idea of a hierarchical topology describing the relationship between individual SDGs and their associated targets, indicators, sub-indicators and in some cases sub-sub-indicators. Thus, in Section 2.6, previous work directed at Hierarchical Time Series analysis is presented. The chapter is then completed, in Section 2.7, with a summary of the material presented and a “look ahead”.

2.2 SDG Forecasting

The concept of the Sustainable Development Goals (SDG) started as a political initiative [5] directed at ensuring a "prosperous future for all". The SDG concept was incorporated into the development plans for many countries. These development plans were typically accompanied by a range of national initiatives designed to support SDG attainment. A common feature of these initiatives was the idea that Artificial Intelligence (AI) could be used to support SDG attainment. As a consequence a significant number of high-profile organisations with an interest in AI started to get involved. Organisations such as Google and Amazon started using the term "AI for social good" (AI4SG) [22]. In 2021 Harvard University organised an AI4SG workshop [1]. AI has significant potential to advance and enhance prediction of attainment of SDG goals for a given country, or at least obtaining information concerning progress towards attainment. Given the foregoing, there has been significant research effort directed at utilising AI for SDG forecasting. Generally, the previous work directed at the forecasting of SDG attainment can be divided into three categories: (i) single target forecasting, (ii) multiple target forecasting and (iii) SDG Indexing. Each of these is discussed in detail in the following three sub-sections, Sub-sections 2.2.1 to 2.2.3. With reference to this previous work, it is noteworthy that there is no existing work, at least to the best knowledge of the author of this thesis, directed at a holistic approach to forecasting. Existing work has been limited to forecasting with respect to a single SDG target or, at best, a small number of SDG targets. The work presented in this thesis takes a holistic approach to SDG forecasting taking into account not only the individual SDGs with respect to an individual region, but also the impact of SDG attainment in the context of neighbouring regions. This dimension makes the work presented in this thesis distinct from the previous existing work considered in the remainder of this section.

2.2.1 Single Target Forecasting

Single target forecasting, as the name suggests, is where only a single SDG goal, target or indicator, is considered with respect to a single geographic region. Much of the existing work on SDG forecasting falls under this category [23, 24, 25, 26, 27, 10]. In [23] and [24], forecasting was undertaken in the context of electricity supply in Ukraine with a

[1] https://crcs.seas.harvard.edu/event/2021-ai-social-good-workshop

view to SDG-7 attainment, “Ensure access to affordable, reliable, sustainable and modern energy for all”. In [25] the focus was on “air pollution particulate matter forecasting” using a neural-based ensemble approach in the context of SDG-3 “Good health and well being”. In [26], the focus was on Thailand’s carbon dioxide emissions using a Vector Autoregessive Moving Average (VARIMA) model, relevant with respect to SDG-9 “Build resilient infrastructure, promote inclusive and sustainable industrialisation and foster innovation”. In [27], Gross Domestic Profit (GDP) forecasting was considered, using a regression-based estimation model, in the context of SDG-1, “End poverty in all its forms everywhere” was considered. The author used a regression-based estimation model to forecast the attainment of SDG-1. In [10], a forecasting mechanism was proposed that utilised satellite imagery and a neural network to classify forest areas and consequently contribute to forecasting SDG-15, ”Life on land”, attainment. There are many more examples.

2.2.2 Multiple Target Forecasting

Multiple target forecasting, as opposed to single target forecasting, is directed at forecasting either: (i) a number of SDG targets simultaneously, or (ii) forecasting with respect to a number of factors in the context of a single SDG or (iii) forecasting with respect to a number of geographic regions in the context of a single SDG. As noted in this thesis, many SDG targets are interconnected in some way [1]. An example of forecasting with respect to a number of SDG can be found in [28], where a methodology was proposed (named DNNSDGs1-6) to monitor and forecast SDGs 1 to 6 using deep neural networks. An example of forecasting a single SDG but using a number of factors can be found in [29] where multiple regression was used, with respect to the number of years that early pre-primary education children attended dance classes and math/science test scores, to measure the attainability of SDG Target 4.2, “by 2030 ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education”. A secondary objective of the work presented in [29] was to establish an association between the number of years that early pre-primary education children had attended dance classes and their maths and science test scores. In [30] a study was presented focused on forecasting the mortality rate of traffic accidents at a global level (across geographic regions), using Bayesian forecasting methods, in connection with SDG Indicator 3.6.1 “Death rate due to road traffic injuries”. In [31] satellite imagery was used with respect to 15 SDG indicators spread over seven SDG goals. For SDG-1, “No poverty", satellite imagery was used to track changes and produce an "Index of change in asset wealth". For SDG-2 "Zero Hunger", satellite images of cropland were used to generate sequence data describing progress to which time series forecasting was applied to predict future progress.

2.2.3 SDG Indexing

Another approach to predict the attainability of SDGs is by ranking countries using an aggregated index. Of course the index on its own will not forecast whether a geographic region will attain its SDGs; to identify which countries will or will not attain their SDGs, a "cut-off" threshold index is required. An example of such indexing can be found in [32]. The indexing idea is also incorporated into the International Futures Modelling System (IFMS) [33] hosted by the Pardee Centre for International Futures[2]. The Pardee IFMS uses a range of equations to generate indexes which are then put together to produce a global index. The equations are generated in a variety of manners. In the case of health goals, regression models were used to generate the desired equations. An example equation, for determining the mortality rate per 100K of population index, is given in Equation (2.1). In the equation: c is the age category, p is gender, d is cause of death, r is region, β is the regression coefficient, Y is the GDP, HC is the number of years of adult education over 25, r is a region identifier, T is time, SI is smoking impact and k is a correction factor. However, the main disadvantage of the IFMS approach is that the equations are fixed, there is no continuous learning involved as in the case of the approaches presented in this thesis.

$$M_{c,p,d,r} = (C_{c,p,d} + \beta_1) \times (Y_r + \beta_2) \times (HC_r + \beta_3) \times (Y_r^2 + \beta_4) \times (T + \beta_5) \times SI_{c,k,p,r} \quad (2.1)$$

.

The approach presented in this thesis also, in part, adopts an indexing approach. However, the index in this case is intended to demonstrate how close a particular geographic region is to attaining its SDGs where 100% attainment is not predicted. The intuition here is that a simple binary, yes-no (will attain - will not attain), classification is not very helpful if the classification is no (will not attain).

[2]https://korbel.du.edu/pardee

2.3 Time Series Analysis

The approaches to SDG forecasting presented in this thesis all incorporate time series analysis. Time stamped data associated with SDG indicators, sub-indicators and/or sub-sub-indicators is used to build time series forecasting models which in turn are used to predict SDG attainment. This section reviews previous work directed at time series analysis, specifically time series forecasting. The time series forecasting mechanism that the work presented in this thesis builds upon includes:(i) autoregressive modelling, (ii) Facebook Prophet (FBprophet) and (iii) Neural Networks. Each of these is discussed in further detail in the following three sub-sections, Sub-sections 2.3.1, 2.3.2 and 2.3.3.

2.3.1 Autoregressive Modelling

Autoregressive (AR) modelling is concerned with building a time series data model using known historical time series values that can be used to predict future values for a given time series; in other words predicting future values according to past values. In [34], AR is defined as a scheme that assumes that a given value of a time series is a weighted linear sum of past values and/or residual deviations. Since its inception in 1926 [35], AR modelling has been widely used, with respect to many application domains, for predictive analysis and forecasting. Example application use cases include: commodity price forecasting in global markets [36], analysis of patients rehabilitation progress over time [37], and Natural Language Processing (NLP) tasks such as next sentence and next-word prediction given past context [38]. Two forms of autoregressive modelling are considered in this thesis, Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA). The distinction is that ARMA is applicable to stationary time series (time series that do not feature changing trends or seasonal variations) while ARIMA is applicable to non-stationary time series (time series that do feature changing trends or seasonal variations). Both use the concept of a Moving Average (MA). A Moving Average can be calculated in multiple ways, as discussed in the following sub-sections, before further describing ARMA and its extensions, such as VARMA and STARMA, as well as ARIMA.

Moving Average

A Moving Average (MA) is an average of some subset n of numbers in a time series. The average can be calculated for any time series sub-sequence [39]. Several variations of MA have evolved over time including:

1. Simple moving Average (SMA): This is an average of n equally weighted previous data points calculated as follows:

$$SMA_n = \frac{1}{n}\sum_{t=1}^{t} = nx_t \tag{2.2}$$

Where x_t is the actual value at time t in the time series. Thus SMA assumes that each data point is equally accountable in the computation of future estimations.

2. Weighted moving average (WMA): This is a weighted version of SMA where the average of n differently weighted previous data points is used. The weights are intended to reflect progress of the time series. Typically, the most recent data points are assigned greater weights compared to the older ones. Thus, the weights are linearly related to the change in time. In [40], WMA was computed as follows:

$$WMA_n = \frac{\Sigma_{t=k-n+1}^{k} w_t A_t}{\Sigma_{t=k-n+1}^{k} w_t} \tag{2.3}$$

Where n is the total number of observations, k is the window size, At is the actual value at time t and Wt is the weight of the data.

3. Exponential Moving Average (EMA): EMA is similar to WMA except that the weights increase exponentially from one data point to the next. In [41], EMA was computed as follows:

$$S_1 = Y_1, \tag{3}$$

$$for\ t > 1, S_t = \alpha \cdot Y_t + (1 - \alpha) \cdot S_{t-1} \tag{4}$$

Where Y_t is the value at time t, S_t is the value of EMA at time t and α is the weighting factor that can be computed as $\alpha = 2/(n + 1)$.

MA schemes can be use to obtain an overall indication of the trends in a data set which in turn can be used for forecasting. In [40], a comparison of the above techniques for forecasting Forex (the foreign exchange currency market) was presented. Three major currency pairs were considered: EUR/USD, AUD/USD and GBP/USD. Using Mean

Squared Error (MSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE) as the evaluation metrics, the order of performance from best to worst was found to be EMA, WMA, SMA in all the experiments across all currency pairs considered.

Autoregressive Moving Average (ARMA)

As the acronym suggests, ARMA is a combination of AR and MA. The most frequently referenced methodology for calculating ARMA is the Box-Jenkins methodology [42]. The Box-Jenkins methodology has been shown to work well with respect to long time series [42], provided that the number of AR terms (p) and the number of MA terms (q) are appropriately specified. This combination therefore suggested that a time series model y_t can be modelled as,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-1} + \ldots + \phi_p x_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-1} - \ldots - \theta_q x_{t-q} \tag{2.4}$$

The Box-Jenkins methodology in its basic form, comprises the following three steps:

1. **Identification:** Determining the order of the model to make sure that the time series is stationary (ARMA is only applicable to stationary time series). This is done by plotting the Auto Correlation and Partial Auto Correlation coefficients, calculated using the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) respectively, so as to determine appropriate values of p and q. An ACF plot can be used to summarise the correlation of observations with lag values. A PACF plot, in turn, can be used to summarises the correlation of observations with lag values not accounted for by prior lagged observations.

2. **Estimation:** Estimation of the parameters ϕ and θ (Equation 2.4). This involves using non-linear optimisation methods that minimise the sum of square errors such as Ordinary Least Square [43], or Maximum Likelihood [44].

3. **Diagnostics and Forecasting:** The evaluation of the model to ensure that it does not overfit (is unable to generalise with respect to test data), and that the residuals are not auto-correlating. Once the model orders are sufficiently estimated, optimal values for p and q are arrived at and the model can be used for forecasting.

The ARMA model has several specialized extensions, including Space-Time Autoregressive Moving Average (STARMA) and Vector Autoregressive Moving Average (VARMA). In areas like geostatistics, environmental studies, and transportation where data exhibit spatial and temporal structures, STARMA is especially helpful because it is built to handle space-time data by accounting for both spatial and temporal dependencies. On the other hand, VARMA is an extension for multivariate time series in which multiple time series are modelled concurrently. the idea is to capture the interdependencies between various time series and makes it appropriate for situations in which multiple time series are connected or influence one another [42]. When numerous time series need to be analysed simultaneously, VARMA models have been used in a variety of fields, including economics, finance, and environmental research [45].

Autoregressive Integrated Moving Average (ARIMA)

ARIMA is a variation of ARMA. While ARMA assumes stationary time series, ARIMA assumes non-stationary time series. The "I" within ARIMA stands for Integrated and it represents the difference or combining the previous differences (previous lags along with the residual errors) in order to make the time series stationary [46].

2.3.2 Facebook Prophet

The well documented disadvantages of modelling (such as ARMA and ARIMA modelling) are that they do not perform well with long term forecasts or using short time series. ARIMA is also prone to large trend errors whenever there is a change in trend. This was shown in [47] where the naivety of exponential smoothing in capturing weekly seasonality but missing long-term seasonality was noted. Given that the available SDG time series were short and that, given the shortness of the time series, the forecasts to be made were relatively long term, it was considered appropriate, with respect to the work presented in this thesis, to also consider an alternative forecasting mechanism that would address these disadvantages. To this end, Facebook Prophet (FBprophet), as proposed in [47], was adopted. FBprophet comprises a three component time series model:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{2.5}$$

Where; $g(t)$ is a trend function modelling non-periodic changes in the value of the time series, $s(t)$ models periodic changes (such as weekly and yearly seasonality), and $h(t)$ represents the effects of "holidays" which occur on potentially irregular schedules over one or more days, and finally ϵ_t is an error term to account for situations not accommodated by the model. The FBprophet model has been extensively utilised, as reported in the literature to forecast real-world data, including its use in predicting the number of future COVID-19 cases in Bangladesh, as demonstrated in [48]. Additionally, the model has been applied to other areas such as predicting stock prices, electricity demand, and website traffic due to its capability to capture diverse seasonal patterns, trends, and anomalies present in time series data. Its popularity stems from its effectiveness in analysing and predicting time series data in numerous domains.

2.3.3 Neural Network Based Schemes

The third category of forecasting schemes considered in this thesis is Neural Networks (NNs) based schemes. NN schemes provide another alternative to AR-based forecasting methods [49, 50, 51]. NNs are mathematical models that perform complex operations in a manner inspired by the functioning of biological neurons in order to learn past or observed information [52]. Broadly speaking, NNs can be categorised as either Feed-Forward Networks (FFN) or Recurrent Neural Networks (RNNs). FFNs perform non-linear transformations on a given input to produce an output. Information flows from the input layer to the output layer in a "forward" direction, hence the term feed-forward. A FFN that has multiple layers performing non-linear transformations is referred to as a multi-layer network. RNNs are distinct from FFNs in that, they include paths from an NN layer that loops back to the starting layer called a feedback connection [53]. This allows RNNs to model sequential data. Consequently, RNNs have been widely used in time series forecasting and various other sequential processing tasks such as: speech recognition [54], dialogue systems [55], and machine language translation [56]. In addition, NNs have a pattern-recognition ability that is hinged upon an algorithm that invokes the backward flow of information through the NN layers called backpropagation. The intuition behind backpropagation is that it computes gradients of the errors made by the NN, by computing small changes to the error with respect to the weight and bias terms of the connections in the NN [57]. The challenge is that these gradients can "explode", or "vanish" (decay), in ways that ends up minimally contributing to the NNs ability to learn or recognise patterns [58]. A remedy to this problem was proposed in [58], namely Long Short Term Memory (LSTM) networks. For the work descrined in chapter 6 encoder-decoder LSTMs were adopted because of their reported recent success in sequence-to-sequence prediction (or sequence labelling) problems [59]. Further details concerning LSTMs and Encoder-decoder models is presented in the remainder of this sub-section.

Long Short Term Memory (LSTM)

The idea of the Long Short Term Memory (LSTM) RNNs was proposed in [58]. LSTMs were designed to overcome the vanishing gradient problem of RNNs. The LSTM architecture has internal mechanisms, called "gates", that can regulate the flow of information. The gates learn which information to keep and which one to disregard, thus enabling the network to capture good and informative patterns in long sequences.

Encoder-Decoder Models

The RNN Encoder-decoder model was proposed in [59]. The idea was to enhance the machine translation task, which essentially aimed to align and translate sequences from one language to another [56]. The model comprised a pair of RNN networks aligned next to each other, one operating as an encoder and the other as a decoder. The encoder projects an input variable sequence of vectors to a fixed-length vector c:

$$\begin{aligned} h_t &= f(x_t, h_{t-1}) \\ c &= q(h_1, \ldots, h_N) \end{aligned} \tag{2.6}$$

where h_t is a hidden state vector obtained at time t when an input x_t is received, and f and q are nonlinear functions. The decoder maps this vector back to a variable length sequence of vectors. Given the context vector c in Equation 2.6, and all previous decoded features/variables, the decoder predicts the next variable/feature in the sequence as shown in Equation 2.7.

$$p(y_t) = \prod_{t=1}^{T} p(y_t | \{y_1, \ldots, y_{t-1}\}, c) \tag{2.7}$$

Encoder-Decoder Models in Time Series Forecasting

In the context of time series forecasting, encoder-decoder models have become increasingly popular due to their ability to model complex temporal dependencies. Specifically, they are able to learn from the historical data of a time series (input sequence to the encoder) to predict future values (output from the decoder) [60]. One of the critical elements when employing these models in time series forecasting is the choice of window size, i.e., the length of the input sequence. The window size essentially determines the amount of past information the model has access to for making the prediction. Too short a window may not capture essential past information, while too long a window could lead to overfitting or unnecessary computational complexity [61]. After training, the model can be used for out-of-sample forecasting, where the aim is to predict the time series values beyond the observed data. The length of this forecasting horizon can also impact the model's performance. Shorter horizons may not provide comprehensive insights into future trends, while longer horizons can increase the uncertainty of the predictions [62]. Importantly, the encoder-decoder model architecture allows for multi-step forecasting, which can provide a forecasted sequence rather than a single point prediction. This is particularly useful in time series forecasting tasks where knowing the trend or pattern of future values is just as important as knowing individual future points [63]. In the context of Multi-Input Multi-Output (MIMO) forecasting, encoder-decoder models can employ a methodology where the output from the decoder (the forecasted values) is fed back as input to the model for future predictions. This technique, often referred to as autoregressive forecasting, allows the model to leverage its own predictions to generate forecasts for further time steps [60]. This becomes particularly useful when forecasting multiple steps ahead, as the model uses its initial predictions to inform subsequent ones, effectively creating a loop of self-informed forecasting. This approach allows the model to capture the inherent temporal dependencies in the time series beyond the observed data [64].

However, while this approach can increase the model's ability to capture complex temporal patterns, it may also amplify prediction errors over time due to the propagation of errors from one forecasting step to the next. Hence, while employing this method, it is crucial to balance the benefits of self-informed forecasting against the potential increase in forecast uncertainty [65]. The encoder-decoder model is a powerful tool in time series forecasting due to its flexibility and capacity to manage sequences of different lengths. But, the performance of the model hinges a lot on the careful picking of parameters such as the window size and forecast horizon. Additionally, it's crucial to take into account the effect of other time series. After the identifying of potential causer relationships, these factors can greatly influence how well the model performs.

2.4 Time Series Causality and Correlation Inference

A central theme of the work presented in this thesis is the view that individual SDGs are not independent of one another. This idea is central to the Sustainable Development Goal Correlated Attainment Prediction (SDG-CAP) and Sustainable Development Goal Track Trace and Forecast (SDG-TTF) frameworks presented in Chapters 6 and 7. The motivation for these two frameworks, as already noted, was the observation that the attainment of one SDG was often related to, or at least influenced by, the attainment of another SDG, either with respect to the same geographic region, or with respect to a neighbouring region. To establish whether two or more SDG's are connected requires some form of causality/correlation inference applied to the time series associated with the SDGs. Causal inference is concerned with the process of establishing a connection (or the lack of a connection) between events or instances. For two time series to be causally related, a change in one should cause a change in the other. Given two time series, $A = \{a_1, a_s, \dots, a_n\}$ and $B = \{b_1, b_2, \dots, b_m\}$, where we wish to establish that B is causality-related to A, this is typically established using a prediction mechanism that uses the "lag" $\{b_1, \dots, b_{m-1}\}$ in B to predict a_n in A. We then compare the predicted value for a_n with the known value, for example using Root Mean Square Error (RMSE). If the RMSE is small this means that the two values are similar and thus we can conclude that "time series A is causality-related to time series B". Correlation describes an association between two time series. When one changes so does the other. Correlation does not necessarily imply causality. Causal and correlation inference have been applied in the context of SDG analysis. In [66], an SDG forecasting mechanism was described, focusing on socioeconomic indicators in the Lipetsk region of Russia, that used correlation analysis to find statistically significant relationships between SDG indicators. The study concluded that employment rates are a determining factor in a region's ability to meet its SDGs.

There are a number of mechanisms that can be adopted for causal and correlation inference. As it pertains to the work presented in this thesis five such mechanisms were considered: (i) Pearson Correlation (PC), (ii) Granger Causality (GC), (iii) Least Absolute Shrinkage and Selection Operator (LASSO), (iv)) the Mann-Whitney U Test (MW) and

(vi) Dynamic Time Warping (DTW). Each is discussed in further detail later on in this thesis (in Chapter 6), however, for completeness in the context of this literature review chapter, each is summarised in the following five sub-sections, Sub-sections 2.4.1 to 2.4.5.

2.4.1 Pearson Correlation

Pearson Correlation [13] (PC) has been used to measure the correlations between any given pair of time series. The mechanism assumes the linearity of the data. This assumption holds with respect to many SDG time series that are typically linearly spaced, and therefore Pearson seems like a good choice for determining whether a relationship exists between a pair of SDG time series or not.

2.4.2 Granger Causality

Granger Causality (GC) is one of the most widely used causal inference mechanisms found in the literature [6, 67]. It was introduced in the 1960s and is calculated as shown in Equation 2.8 where: (i) X and Y are time series, (ii) a and b are the lags of X and Y, (iii) t is the current time step and (iv) e is a residual error. The idea is that if time series X "granger causes" time series Y, then the past values of X should contain helpful information to forecast Y in a manner that would be better than when forecasting y using only Y's historical data. GC has been used previously in the context of SDG prediction, for example in [6] where 20,000 pairs of time series that featured causal relationship were found.

$$Xt = a_1 X_{t-1} + b_1 Y_{t-1} + e \tag{2.8}$$

The variation of GC that was used with respect to the research presented in this thesis was the Stats-model variation [68].

2.4.3 LASSO

Least Absolute Shrinkage and Selection Operator [14] (LASSO) is an L1 regularisation technique frequently used to reduce high dimensionality data, which can also be employed to establish the existence of a causality between variables [69, 14]. LASSO reduces the dimensionality of the input data set by penalising variances to zero, thus allowing irrelevant variables to be removed. Equation 2.9 shows the LASSO cost function. Inspection of the equation indicates that the first part is the *squared error* function, while the second part is a penalty applied to the regression slope. If λ is equal to 0, then the function becomes a normal regression. However, if λ is not 0 coefficients are penalised accordingly, leaving only coefficients that can explain the variance in the data.

$$L\ C\ F = \sum_{i=1}^{n} \left(y_i - \sum_{j} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{2.9}$$

2.4.4 Mann-Whitney U Test

The Mann-Whitney U Test (MW) [16] is the fourth mechanism considered in this thesis. The test is used to determine if any two pairs of time series are statistically different. It is a non-parametric test (unlike, for example, LASSO).

2.4.5 Dynamic Time Warping

Dynamic Time Warping (DTW) is a metric that indicates the similarity of two time series[17]; in other words a special form of correlation. However, unlike Euclidean distance measurement [70], DTW is not constrained by an equal length requirement for time series. DTW can be used to measure the distance between time series of varying lengths and takes into account misalignment between time series. The "warping distance" w is the similarity value produced by DTW. A value of $w = 0$ indicates that the two time series under consideration are identical.

2.5 Time Series Missing Values

SDG data as will become noticeable in the following chapter has a significant amount of missing data. In [71] it was observed that the challenge of missing data is inescapable when analysing real-world data [71] such as SDG data. Many machine learning techniques do not work well or do not work at all in the presence of missing data. In [72] it was argued that missing data results in increased model complexity and consequently learning time; which, in turn, tends to degrade model performance. Moreover, several authors have indicate that, if inappropriately handled, missing data problems can lead to fewer extracted patterns, analysis complications, overfitting and/or underfitting, which all collectively or individually lead to misinformative conclusions [73].

Dealing with missing data in both multivariate and univariate time series analysis is an ongoing research problem. There is no single "best" strategy for handling missing data. A given strategy may be appropriate for one task but not another. The most obvious approach to handle missing data is simply to ignore it, provided our machine learning algorithms will still run with missing data (often not the case). The obvious alternative is to delete records that feature missing data [74]. However, this means that a lot of potentially useful data may be thrown away. This will be a disadvantage particularly in the case of SDG data, because of the large number of missing values in the data. Another alternative to handling the missing data problem is to impute (estimate) missing values from the existing data. The intuition behind imputation is that, if an important feature is missing for a particular instance, it can be estimated from other instances [73]. We can identify two broad categories of imputation: (i) Predictive Value Imputation (PVI) where the mean or mode of a discrete- or real-valued sequence is used as the replacement value [75], and (ii) Distributed Value Imputation (DVI) where the expected distribution of the target variable is obtained and used to replace missing values. With respect to the work presented in this thesis, PVI was used, more specifically polynomial interpolation imputation. There are a variety of ways in which PVI can be implemented. The most common are listed below. A comparative examinations between all of these methods is presented later in this thesis in Chapter 3.

Linear Interpolation: Involves fitting a straight line between the known "end-points" before and after a sequence of missing values and using a linear function to compute each missing value y as indicated by the equation below, where $k = \frac{y_2 - y_1}{x_2 - x_1}$; and $x_1 \leq x \leq x_2$ and $y_1 \leq y \leq y\_2$.

$$y = y_1 + k(x + x_1) \tag{2.10}$$

Spline Interpolation: The joining of values surrounding missing values using cubic splines. The determination of the spline is based on fitting cubic polynomials to a series of observed data (x_i, y_i). The fitting (imputation) is done so that the knots (where piece wise portions joins the function) and its first two derivatives are continuous. A cubic spline with knots $x_{i=1}^{n}$, is defined as follows (provided If $x_i \leq x \leq x_{i+1}$):

$$f(x) = a_i + b_i x + c_i x^2 + d_i x^3 \tag{2.11}$$

Time Interpolation: Imputation followed by normalisation of the imputed data to project it onto a distribution with a standard deviation of 1, or into a scale between 0 and 1.

Finally, in this section, it should be noted that there has been recent work where machine learning models have been used to learn missing values [76].

2.6 Hierarchical Time Series Analysis

In Chapter 1 it was noted that the solutions presented in this thesis stem from the idea of Hierarchical Time Series (HTS) analysis. The idea is that individual time series forecasting models would be held at the lowest level of the hierarchy which will then feed up to the top level of the hierarchy, hence we have a bottom-up hierarchy. The alternative is a top-down hierarchy. There has been previous work directed at the use of HTS in the context of machine learning. In [77], an example was discussed focused on tourism in Australia. The hierarchy here comprised four levels of aggregated data. Level 0 comprised a single time series representing the total number of visitor's nights in Australia. Level 1 contained seven time series representing the seven states and territories within Australia; and Levels 2 and 3 individual regions and sub-regions (76 with respect to the latter). The HTS concept has also featured with respect to Stock indexes, budgeting and medical applications [78, 79, 80]. Whatever the case HTS can be used to forecast from the bottom or top level while using the remaining levels as "Regressors", or forecasting can take place at a middle level and "percolated" to the remaining levels.

2.7 Summary

This chapter has provided a review of the previous work considered relevant to the work presented in this thesis. The chapter commenced by considering previous work directed at SDG forecasting, and then went on to consider time series analysis, time series causality and correlation inference, missing data within the context of time series and hierarchical time series analysis. The following chapter presents the nature of the SDG data used with respect to the evaluations reported on throughout the work presented in this thesis.

Chapter 3

The SDG Data Set

3.1 Introduction

Throughout the work described in this thesis, the SDG data set was used both as a driver for the work and for evaluating the work. This chapter provides an overview of the SDG data set and the adopted mechanisms for pre-processing this data to obtain a taxonomy describing the structure of the data and the time series required to build the prediction models to be held at the leaf nodes of the taxonomy. The preprocessing described in this chapter is accessible in the project GitHub repository found at: [1].Recall that the central idea promoted in this thesis is that of bottom-up hierarchical classification facilitated by a taxonomy of the domain of discourse. The motivation for this, as noted in Chapter 1, was the observation that the SDG data could be arranged in a hierarchical format that lent itself to hierarchical classification. For example, a geographic region will satisfy the UN SDG expectations if the individual SDGs are satisfied, which requires that the individual targets are satisfied, which in turn requires individual indicators, sub-indicators and sub-sub-indicators to be satisfied. However, it's important to recognise that forecasting the future of an SDGs and predicting whether a target will meet its goals are two distinct operations. The first requires a time series forecasting operation, focusing on the progression and trends of the SDGs over time. The latter, on the other hand, requires a classification operation based on the actual SDGs targets. These two operations, while interconnected, address different aspects of SDGs attainment prediction.A fragment of the SDG taxonomy was presented previously in Figure 1.2. With reference to this taxonomy, each leaf node

[1]https://github.com/Yassir-Alharbi/Sustainable-Development-goals

holds time series data to be used to generate a forecasting (prediction) model. Once the models have been generated, they can be used to predict SDG attainment with respect to a particular indicator and particular country or geographic region. Results are passed up the taxonomy. The question is then how best to generate such a taxonomy. This is one of the subsidiary research questions that this thesis seeks to provide an answer to, Subsidiary Research Question S1 from Chapter 1:

S1. What is the most efficient and effective way to derive a taxonomy for SDG data?

The rest of this chapter is structured as follows. Section 3.2 provides an overview of the raw SDG data. Section 3.3 presents the adopted preprocessing of the SDG data to firstly generate the required SDG taxonomy and secondly to produce *country data files* to which the approaches for SDG attainment prediction described later in this thesis can be applied. Section 3.4 presents the interconnectedness and statistical nature of the SDGs. The chapter is completed in Section 3.5 with a summary of the work presented.

3.2 Overview of The Sustainable Development Goal Raw Data

This section provides an overview of the SDG raw data. The aim is to facilitate the reader's understanding with respect to the remaining sections of this chapter. To maintain the visibility of the SDG agenda, the UN periodically releases SDG related data on the www platform of the United Nations Department of Economic and Social Affairs' Statistics Division[2]. For example, in 2020, there was a total of 19 releases [81]. Once on the SDG data website, the data can be downloaded partially or wholly in a CSV format. The SDG platform contains data related to 346 countries and geographical regions. However, it is important to note that some countries no longer exist, such as the Former Sudan. In addition, the platform features collated data for regional groupings, such as Sub-Saharan Africa, Northern Africa, Western Asia, Central and Southern Asia, and so on.

The SDG data set D comprises a set of records $\{R_1, R_2, \dots\}$. Each record R_i comprises a set of values $\{v_1, v_2, \dots\}$ where each value corresponds to a set of attributes $A = \{a_1, a_2, \dots\}$ where each individual attribute a_i is an individual categorical or numeric attribute, which takes a corresponding categorical or numeric value. Thus D comprises

[2]https://unstats.un.org/sdgs/indicators/database

a single, very large table with columns representing a range of numerical and categorical attributes and rows representing single observations related to individual SDG indicators (sub-indicators or sub-sub-indicators). Each record R_i is date stamped. The set A (the columns in the table) represents the complete set of attributes for all 561 indicators. However, only a small sub-set of the available set of attributes will be relevant for any one indicator.

Table 3.1 gives an example of an SDG record for the country Afghanistan for the year 2015. The example refers to Goal-16 “Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels”, Target 16.1 is then “Significantly reduce all forms of violence and related death rates everywhere”, which has associated with it indicator 16.1.1, “Number of victims of intentional homicide per 100,000 population, by sex (victims per 100,000 population)”, which in this case has a value of 0.55597 per 100,000 head of population. For ease of reading, the table has been arranged in a multi-column format. The record is actually a single row in the SDG database D. In the table, each entry comprises an attribute name and value pair. Recall that the set of attributes is a global set of attributes that encompasses all indicators. From the example, it can be seen that many of the attributes are not relevant with respect to certain SDG indicators (such as “Mountain elevation” or “type Of Speed”). This is why the value for many attributes in Table 3.1 has been set to “NA” (Not Applicable). To act as a focus for the work reported in this thesis 38 different countries or geographical regions, represented by 200,742 time series observations, were selected. The list of countries was given in Section 1.4 of Chapter 1.

The SDG data, as described above, first needed to be preprocessed into a structured format that would facilitate the generation of the desired taxonomy. The structured format also had to facilitate the population of the taxonomy, once generated, with respect to individual countries (geographic regions). A vital element of populating the taxonomy was collating the time series values to be held at the taxonomy leaf nodes, which would later be used to build the desired forecasting (prediction) models. Recall that each time series represents the data associated with an individual indicator (sub-indicator or sub-sub-indicator) represented by a leaf node in the taxonomy. Note that the indicator reference, 16.1.1 in the example, incorporates the goal and target references; hence for further processing, we only need the indicator reference; in the remainder of this thesis, we will refer to this as the Goal-Target-Indicator (GTI). The GTI is sufficient to identify individual rows

Attribute	Goal	TimeCoverage	Cities
Value	16	NA	NA
Attribute	Target	UpperBound	Counterpart
Value	16.1	NA	NA
Attribute	Indicator	LowerBound	Disability status
Value	16.1.1	NA	NA
Attribute	SeriesCode	BasePeriod	Education level
Value	VC_IHR_PSRC	NA	NA
Attribute	SeriesDescription	Source	Fiscal intervention stage
Value	Number of victims of intentional homicide per 100,000 population, by sex	National Statistical Organization	NA
Attribute	GeoAreaCode	GeoInfoUrl	Food Waste Sector
Value	4	NA	NA
Attribute	GeoAreaName	FootNote	Freq
Value	Afghanistan	NA	NA
Attribute	TimePeriod	Activity	Frequency of Chlorophyll-A concentration
Value	2015	NA	NA
Attribute	Value	Age	Grounds of discrimination
Value	0.55597	NA	NA
Attribute	Time_Detail	Cause of death	Hazard type
Value	2015	NA	NA
Attribute	IHR Capacity	Mode of transportation	Observation Status
Value	NA	NA	NA
Attribute	Level of requirement	Mountain Elevation	Parliamentary committees
Value	NA	NA	NA
Attribute	Level/Status	Name of international institution	Policy Domains
Value	NA	NA	NA
Attribute	Location	Name of non-communicable disease	Policy instruments
Value	NA	NA	NA
Attribute	Nature	Quantile	Migratory status
Value	C	NA	NA
Attribute	Report Ordinal	Substance use disorders	Type of speed
Value	NA	NA	NA
Attribute	Reporting Type	Type of occupation	Type of support
Value	G	NA	NA
Attribute	Sampling Stations	Type of product	Type of waste treatment
Value	NA	NA	NA
Attribute	Sex	Type of skill	Units
Value	FEMALE	NA	PER_100000_POP

Table 3.1: SDGs Example Record, dated 2015, for the region and GTI pair Afghanistan and 16.1.1

in D if there are no sub-indicators or sub-sub-indicators. Hence, to differentiate between individual time series, a unique “Individual Series” (IS) identifier, made up of the Series Code (the fourth attribute in Table 3.1) and further identifying characters, was devised. Note, from the table, that the series code is made up of three *text segments* separated by underscore characters. IS values were constructed by adding a fourth text segment (but more on this later).

3.3 Data Pre-Prossessing

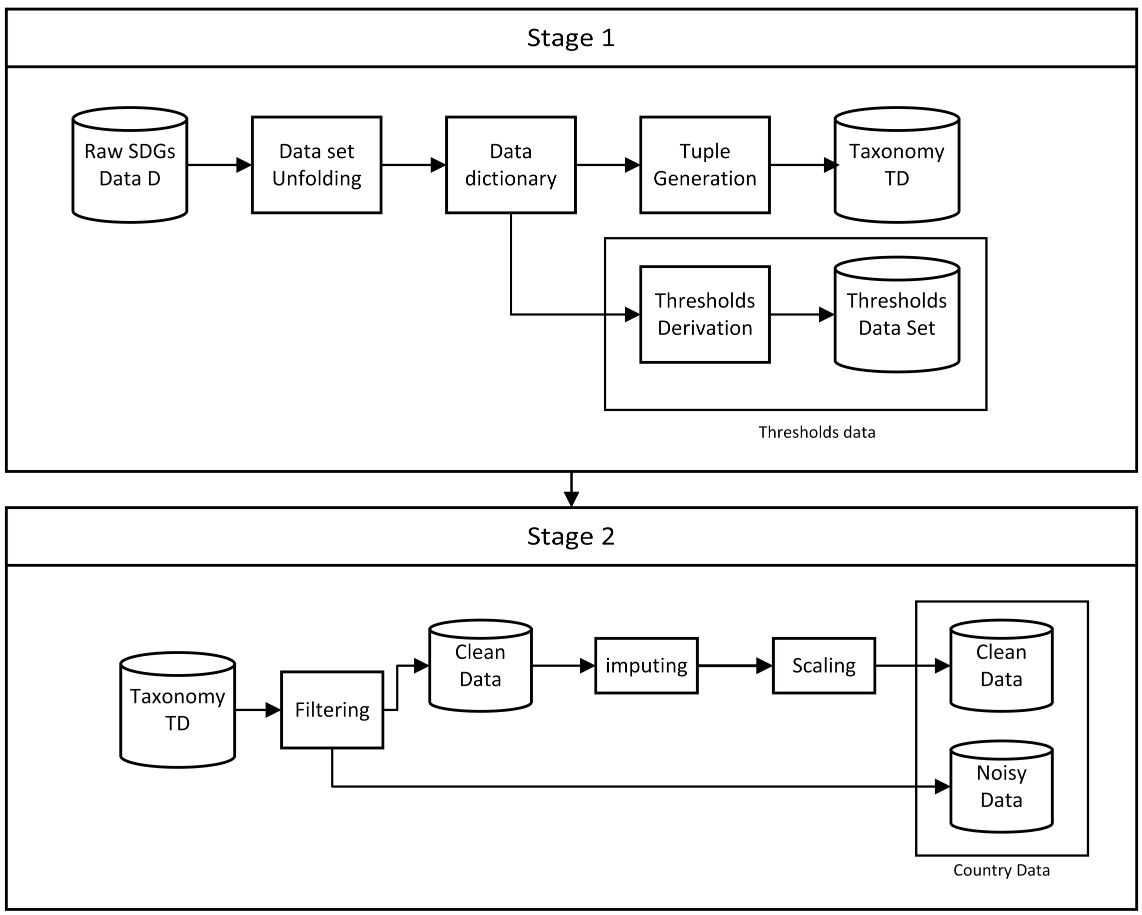


Figure 3.1: Preprocessing Schematic

This section presents the pre-processing of the SDG data to establish and populate the desired taxonomy. A schematic of the adopted pre-processing mechanism is given in Figure 3.1. From the figure, it can be seen that the mechanism comprised two stages.

Stage 1 Taxonomy Generation (Stage 1a) and Threshold Derivation (Stage 1b)

Stage 2 Missing Value Imputation and Scaling (Stage 2a), and the Generation of Country Data Files (Stage 2b).

Threshold derivation (Stage 1b) is a complex process and therefore discussion of this process is left to the following Chapter, Chapter 4 (Sub-section 4.2). Detail concerning Stage 1a is presented in Sub-section 3.3.1; whilst detail concerning Stages 2a and 2b is provided in Sub-sections 3.3.2 and 3.3.3 respectively.

3.3.1 Taxonomy Generation (Pre-processing Stage 1)

Stage 1, as noted above, comprises taxonomy generation. The taxonomy, although visualised as a tree structure, is actually stored as a set of tuples in a data set $TD = \{TR_1, TR_2, \dots\}$, where each $TR_i \in T$ is a tuple of the form:

$$TR_i = \langle Geographical\_Region, GTI, ID, T \rangle \tag{3.1}$$

where: (i) "Geographical_Region" is the name of the country or region of interest, (ii) GTI is the relevant Goal-Target-Indicator, (iii) ID is a unique identifier for a leaf node in the taxonomy for a given region, and (iv) T is the time series associated with the leaf node, $T = [t_1, t_2, \dots]$. Note that the format of TD was specifically designed so as to allow comparison between records within a given geographic region and comparison across geographic regions. The significance of this is that two of the frameworks considered in this thesis, the Sustainable Development Goal Correlated Attainment Prediction (SDG-CAP) framework and Sustainable Development Goal Track Trace and Forecast (SDG-TTF) framework, require intra- and inter-region comparison of time series.

Data Pre-processing Stage 1 is thus concerned with the transforming of the raw SDG input data D into a data set TD. The process, as shown in Figure 3.1, commences by applying a depth-first "data set unfolding" operation to D to collect the *taxonomy paths* that will comprise the SDG taxonomy hierarchy for particular region and indicator pairs. The path information is stored in a data dictionary; an intermediate data repository, that can be conceptualised as "sitting between" D and TD, designed to facilitate the

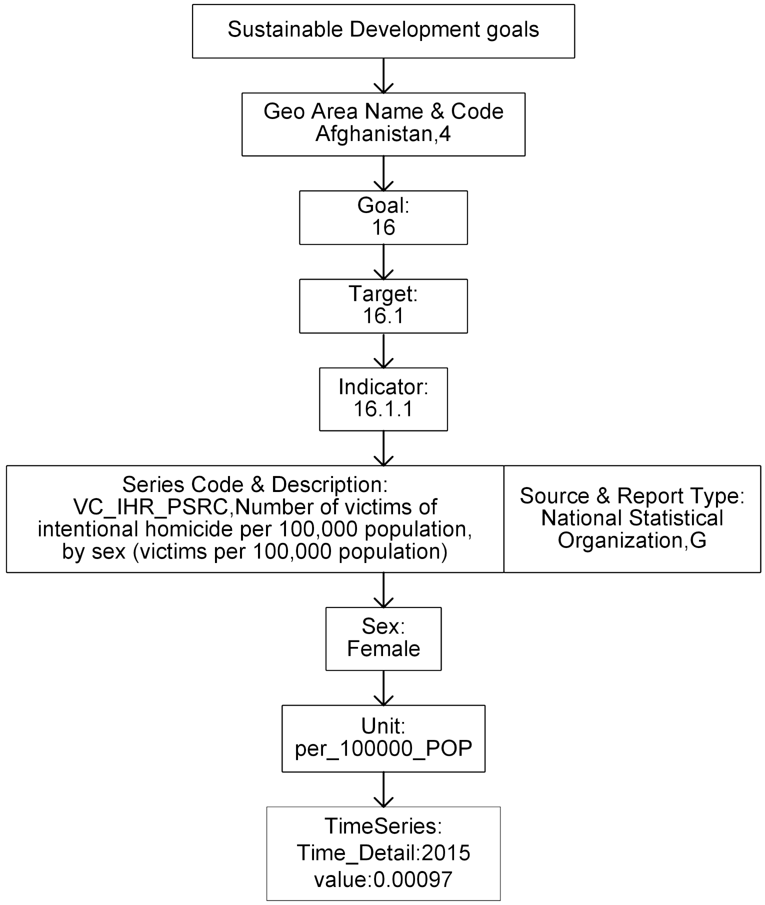


Figure 3.2: Illustration of SDG raw data depth-first "unfolding" for the region-GTI pair Afghanistan and 16.1.1

transformation from D to TD. An example of this unfolding, using the region and indicator pair Afghanistan and GTI 16.1.1, shown previously in Table 3.1, is given in Figure 3.2. A number of example dictionary entries are given below:

1. {*(Goal : 16), (Target : 16.1), (Indicator : 16.1.1), (Series Code : VC_IHR_PSRC), (Series Description : Number of victims of intentional homicide per 100,000 population , by sex (victims per 100,000 population))) (Geo Area Name : Afghanistan), (Units : PER_100000_POP), (Sex : FEMALE)*}

2. {*(Goal : 11), (Target : 13.1), (Indicator : 11.b.2), (Series Code : SG_GOV_LOGV), (Series Description : Number of local governments (number)), (Geo Area Name :*

Afghanistan), (Units : NUMBER)}

3. {*(Goal : 15), (Target : 15.4), (Indicator : 15.4.2), (Series Code : ER_MTN_GRNCVI), (Series Description : Mountain Green Cover Index), (Geo Area Name : Afghanistan), (Units : PERCENT), (Observation Status : A), (Mountain Elevation : 4)*}

The first of the above examples is for the region-indicator pair Afghanistan and 16.1.1 used for illustrative purposes in Table 3.1 and Figure 3.2. The second two are two additional examples. Note that only the most salient information is stored in the dictionary, the information needed for the taxonomy generation.

The dictionary content is then used to generate the SDG taxonomy. Table 3.2 lists details of 55 example leaf nodes. In the table, Column 1 gives the Goal-Target-Indicator (GTI), Column 2 gives the "Series Description", Column 3 the Individual Series (IS) indicator and Column 4 the unique ID number for the leaf node. The series description, a textual description of the indicator, is taken straight from the raw data (see Table 3.1). The IS indicator is the hash code for the record. As noted earlier, the IS indicator comprises four text segments separated by underscore characters. The first three are the series code taken from the raw data. For example, considering the fifth example in Table 3.2, "VC_IHR_PSRC". The fourth text segment in this case that has been added is "FEMALE" to give an IS indicator of "VC_IHR_PSRC_FEMALE". This ID provides a simple unique identifier which is simpler to use than the IS indicator; and which, with practice, can be easily interpreted. The ID number (Column 4) is used for convenience of processing to link records to nodes in the taxonomy as it is constructed.

The final step in Stage 1a (see Figure 3.1) is to transpose the data in the dictionary into a set of tuples ("Tuple Generation") to be held in a set of the form $TD = \{TR_1, TR_2, ...\}$ where each $TR_i \in TR$ is a taxonomy tuple of the form given in Equation 3.1. Note that to reference the time series associated with a taxonomy tuple TR_i the notation T_i will be used. Table 3.3 gives a series of example time series for several indicators for the country Afghanistan. Each column represents a time series. For reasons of conciseness only values for the years 2000, 2005, 2009, 2015 and 2018 have been given. The column headings give the relevant GTI and the unique time series identifier. What can be clearly seen from the table is that there are many missing values.

G.T.I	Series Description	Individual Series	ID
1.1.1	Proportion of population below international poverty line (%)	SI_POV_EMP_15-24_MALE	1
		SI_POV_EMP_BOTHSEX_15+	2
		SI_POV_EMP_BOTHSEX_15-24	3
		SI_POV_EMP_BOTHSEX_25+	4
		SI_POV_EMP_FEMALE_15+	5
		SI_POV_EMP_FEMALE_15-24	6
		SI_POV_EMP_FEMALE_25+	7
		SI_POV_EMP_MALE_15+	8
		SI_POV_EMP_MALE_25+	9
		SI_POV_DAY	10
1.3.1	[ILO] Proportion of children/households receiving child/family cash benefit (%)	SI_COV_CHLD_BOTHSEX	12
		SI_COV_CHLD_FEMALE	13
		SI_COV_CHLD_MALE	14
	[ILO] Proportion of employed population covered in the event of work injury (%)	SI_COV_WKINJRY_BOTHSEX	15
		SI_COV_WKINJRY_FEMALE	16
		SI_COV_WKINJRY_MALE	17
	[ILO] Proportion of poor population receiving social assistance cash benefit (%)	SI_COV_POOR_BOTHSEX	20
	[ILO] Proportion of population above statutory pensionable age receiving a pension, by sex (%)	SI_COV_PENSN_BOTHSEX	21
		SI_COV_PENSN_FEMALE	22
		SI_COV_PENSN_MALE	23
	[ILO] Proportion of population covered by at least one social protection benefit (%)	SI_COV_BENFTS_BOTHSEX	24
		SI_COV_BENFTS_FEMALE	25
		SI_COV_BENFTS_MALE	26
	[ILO] Proportion of population with severe disabilities receiving disability cash benefit (%)	SI_COV_DISAB_BOTHSEX	27
		SI_COV_DISAB_FEMALE	28
		SI_COV_DISAB_MALE	29
	[ILO] Proportion of unemployed persons receiving unemployment cash benefit, by sex (%)	SI_COV_UEMP_BOTHSEX	30
		SI_COV_UEMP_FEMALE	31
		SI_COV_UEMP_MALE	32
	[ILO] Proportion of vulnerable population receiving social assistance cash benefit (%)	SI_COV_VULN_BOTHSEX	33
		SI_COV_VULN_FEMALE	34
		SI_COV_VULN_MALE	35
	[World Bank] Poorest quintile covered by labour market programs (%)	SI_COV_LMKTPQ	36
	[World Bank] Poorest quintile covered by social assistance programs (%)	SI_COV_SOCASTPQ	37
		SI_COV_SOCINSPQ	38
	[World Bank] Proportion of population covered by labour market programs (%)	SI_COV_LMKT	39
	[World Bank] Proportion of population covered by social assistance programs (%)	SI_COV_SOCAST	40
		SI_COV_SOCINS	41
1.5.1	Number damaged dwellings attributed to disasters (number)	VC_DSR_DADN	42
	Number damaged dwellings attributed to disasters, by hazard type (number)	VC_DSR_DDHN_ACIDR	43
		VC_DSR_DDHN_ACIDT	44
		VC_DSR_DDHN_ALLUV	45
		VC_DSR_DDHN_ANIAK	46
		VC_DSR_DDHN_AVALE	47
		VC_DSR_DDHN_BIOGL	48
		VC_DSR_DDHN_CHESP	49
		VC_DSR_DDHN_COLDW	50
		VC_DSR_DDHN_CONTM	51
		VC_DSR_DDHN_COSER	52
		VC_DSR_DDHN_COSFL	53
		VC_DSR_DDHN_CSOLD	54
		VC_DSR_DDHN_CYCLN	55

Table 3.2: Taxonomy leaf node examples

Years	1.2.1_11	1.4.1_2269	1.4.1_2271	1.4.1_2270	1.5.1_2957	8.4.2_1071	8.4.2_1073	8.4.2_1087	8.4.2_1076
2000		22.74099	31.29478	24.64515		21000	6945939	1925978	200178
...									
2005	33.7	27.47609	40.17754	30.41177		243004	9645728	2188895	113076
...									
2009	38.3	31.29914	48.66158	35.4571		725012	9408891	2350763	113076
...									
2015		37.0523	62.26144	43.41761	17	1820623	12948523	2395294	107546
...									
2018									

Table 3.3: Example time series for Afghanistan and a number of indicators arranged in columns

3.3.2 Missing Value Imputation and Scaling (Pre-processing Stage 2a)

As noted earlier with reference to Table 3.3, the SDG data set features a significant number of missing attribute values. A further issue is the different units of measurement used with respect to the different indicators (and the consequent time series). Stage 2 of the data pre-processing applied to the SDG data was therefore directed at addressing these twin issues.

At the time of writing, the theoretical maximum length of any SDG time series is 22 points, covering 22 years of observations from 2000 to 2021. Some indicators, for a small number of countries, have few data points going back to 1967. Figure 3.3 shows the number of observations per year with respect to the 38 different countries considered for evaluation purposes in this thesis. Inspection of the figure indicates that the majority of the data falls between 2000 and 2018. Recall that Millennium Development Goals (MDGs) initiative, the precursor to the SDG initiative, was introduced in 2000. Beyond 2018 the data is frequently not yet available (although, in some cases, it may never become available). Thus, in the context of the research presented in this thesis, only data from 2000 to 2018 was considered; 34,526 time series in total. The total number of time series was calculated as a count of all time series related to the countries considered in this thesis

One of the significant challenges of SDG attainment analysis is data availability. Each country is responsible for collecting their regional data and reporting back to the UN. The United Kingdom, for example, uses The Office for National Statistics (ONS) to collect and report UK data [82]. Countries that do not have an established national agency responsible for collecting national statistics, such as some developing countries, are not as able to provide SDG data, on a regular basis, as countries which do have such an agency. The consequence is missing data.

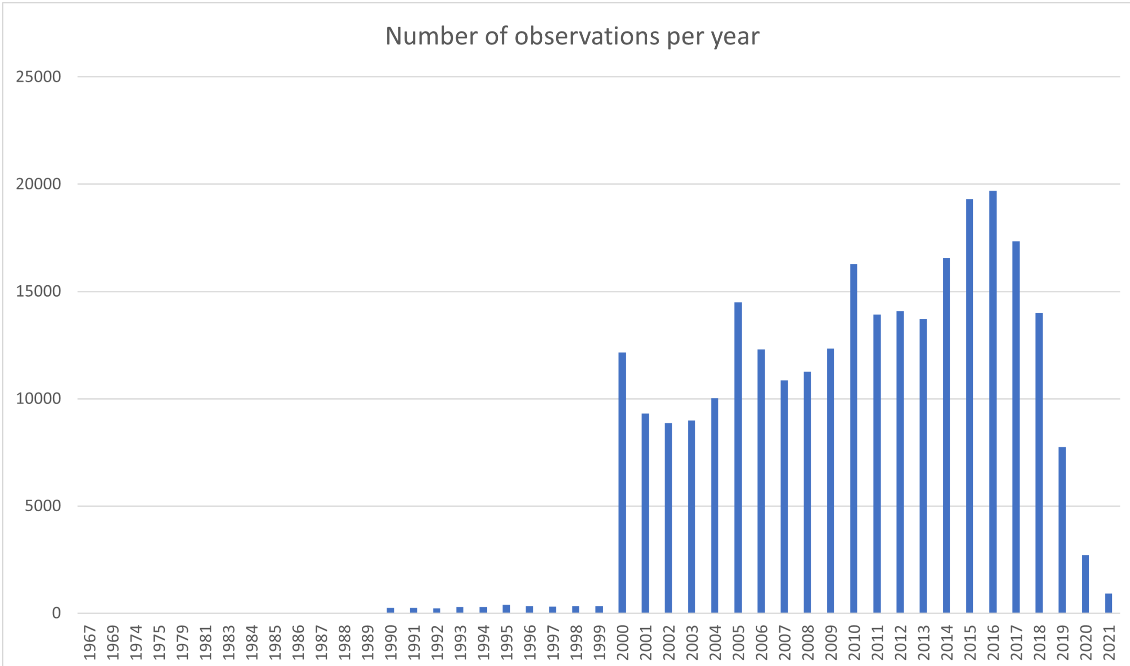


Figure 3.3: Number of Observation per year with respect to the 38 different countries considered for evaluation purposes in this thesis

The reasons for missing data in the collated SDG time series are varied but can be categorised as either: (i) Missing At Random (MAR) or (ii) Not Missing At Random (NMAR) [9]. We can illustrate the distinction by considering the example time series given in Table 3.4 for the country Egypt. Column 2 of the table, 11.5.2_3189, presents the time series is for the indicator "Direct economic loss attributed to disasters (current United States dollars)". Inspection of the associated time series reveals a time series with only one recorded value (the value 311). However, the data describes "loss attributed to disasters", which by definition are not regular occurrences; hence financial losses resulting from disasters are not recorded every year. This type of missing data is thus considered to be NMAR data. The example in Column 3, 9.c.1_2712, describes a time series for the indicator "Proportion of population covered by a mobile network". In this case, the missing data is indicated by a 0 character. The reason for the missing data is unclear because Egypt did have mobile services prior to 2014. The use of 0 characters to indicate missing data is also an irregular use of the value 0. Whatever the case, this type of missing data is considered MAR data. It was assumed that any time series featuring a sequence

Year	11.5.2_3189	9.c.1_2712	3.5.2_642
2000		0	
2001		0	
2002		0	
2003		0	0.1527
2004		0	
2005		0	
2006		0	
2007		0	0.1959
2008		0	
2009		0	
2010		0	
2011		0	0.1727
2012		0	
2013		0	
2014		61	
2015		89	0.1361
2016	311	0	
2017		95.1	
2018			

Table 3.4: Example time series that feature MAR and NMAR data for the geographic region of Egypt;
11.5.2_3189 = Direct economic loss attributed to disasters (current United States dollars)
9.c.1_2712 = Proportion of population covered by a mobile network, by technology
3.5.2_642=Alcohol consumption per capita (aged 15 years and older) within a calendar year (litres of pure alcohol)

of four or more consecutive 0 characters will be considered to feature missing data. The third example in Table 3.4, Column 4, 3.5.2_642, describes a time series for the indicator "Alcohol consumption per capita (aged 15 years and older) within a calendar year (litres of pure alcohol)". In this case, a gap of three years between observations is evident; therefore, it is safe to assume that this particular series is reported every three years and thus, missing values can be categorised as NMAR. Whatever the case, some strategy had to be adopted to address the missing or irregular values contained in TD'.

To address the missing data problem, the idea was to adopt some kind of data imputation, the process of assigning values to missing attribute values according to neighbouring values. This will only work if sufficient neighbouring values are available. Some preliminary experiments, not reported here, indicated that for the imputation to have a chance of success, a minimum of 25% of the values were required. In other words, given that the usable SDG time series of interest were of length 18, we needed values for four or more of the points.

Thus, with reference to Figure 3.1, the first step in Stage 2a was to filter TD into two sub-sets: (i) clean data $TD_{\geq 5}$, and (ii) noisy data $TD_{<5}$. The noisy data sub-set was comprised of time series that either: (i) had less than 5 values, or (ii) feature sequences of four or more 0 characters. The clean data sub-set was then comprised of time series that: (i) comprise more that 4 values and (ii) feature no sequences of four or more 0 characters. Imputation could then be applied to $TD_{\geq 5}$ clean data sub-set.

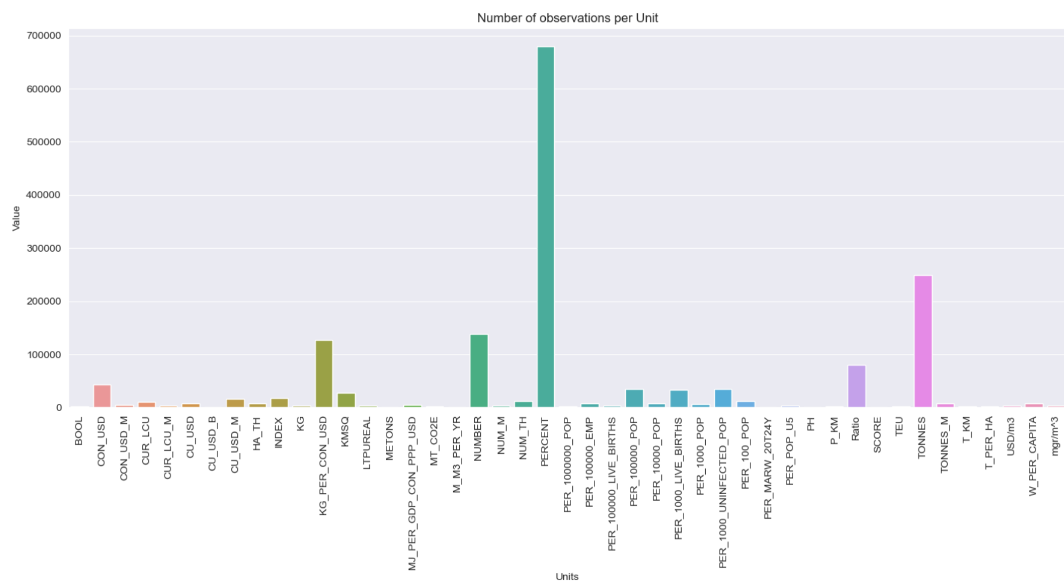


Figure 3.4: Number of observations per unit

A further issue with the collated time series was the different measures used with respect to the different indicators. For each country, as of February 2021, there were up to 3,408 different time series covering a wide range of domains[3]. For each of these time series, one of 45 different units of measurement was used. Figure 3.4 lists each unit and the number of times it appeared in the data (up to February 2021). The dominant measuring

[3]Note that all indicators are not necessarily relevant to all countries, for example, indicators concerned with forestation will not be relevant to a desert country; hence all countries do not feature exactly the same number of time series.

unit is the percentage, followed by Tonnes and number. The percentage unit is widely used in the SDG data, as it is applicable to many different scenarios. The Tonnes unit of measurement was used most frequently with respect to Goal 8, “Promote sustained, inclusive and sustainable economic growth, full and productive employment decent work for all” where many indicators measure material consumed in a country. The number unit measurement was often used to describe a monetary figure or a population. With this in mind, any consideration on building multivariate time series models will require the data to be on the same scale. Without scaling of the time series, a series of population counts or monetary values will always dominate over (say) a series comprised of percentages values. There was thus a requirement for some form of scaling to be applied to the data.

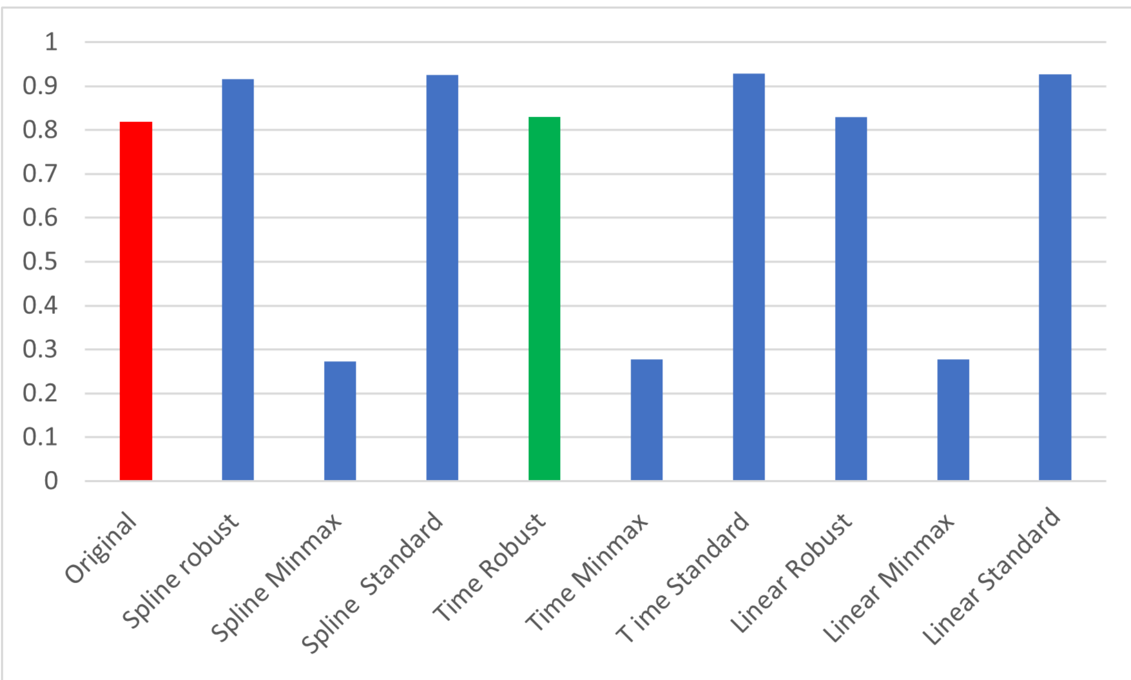


Figure 3.5: comparison between each interpolation and scaling method

A set of experiments was conducted to find the best mechanism for imputing missing values and scaling the data. Three different imputation methods and three different scaling algorithms were considered for the experimentation. The imputation methods were: (i) Spline, (ii) Time and (iii) Linear [83, 84, 85]. The scaling algorithms considered were: (i) Robust, (ii) Minmax and (iii) Standard [85]. These methods and algorithms were chosen because of their popularity in the literature. The three imputation methods and three scaling algorithms could be combined in nine different ways. Only complete time series were used for the experimentation; time series where all 18 values were available. There were 218 of these out of the 36,421 time series extracted for the 38 countries used as a focus for the work presented in this thesis.

Each time series was split into a training part and a testing part, $T_{i_{train}}$ and $T_{i_{test}}$, a prediction using FBprohpet [86] was used to obtain a baseline value. Four values were then removed from each $T_{i_{train}}$ and then the selected imputation method and scaling algorithm were applied to $T_{i_{train}}$ which was then used to predict the values in the test part. The closest prediction match was then selected for use in Stage 2 of the SDG data preprocessing. The adopted evaluation metric was the averaged Root Mean Square Error (RMSE). Although normally a lower RMSE value means better results, in this experiment, the goal was to stay as close as possible to the baseline RMSE value. The results of the experiments are given in Figure 3.5 which shows the average RMSE per combination. From the figure it can be seen that the Time Robust imputation and scaling method combination (the green bar in the figure) produced an average RMSE of 0.8296 which is the closest average RMSE values to the original value of 0.8185. This was then the combination adopted for Stage 2 of pre-processing applied to the SDG data to address the missing value problem and many measurement units used problem.

Returning to Figure 3.1, imputing and scaling was applied to $TD_{\geq 5}$ to give $TD'_{\geq 5}$. However, note that $TD'_{<5}$ was not thrown away. The reason for the later will become clear later in this thesis.

3.3.3 Generation of Country Data Files (Pre-processing Stage 2b)

The final step in Stage 2, Stage 2b, of the SDG data pre-processing was country file generation. From the foregoing, the previous steps in Stage 2 resulted in two data sets $TD_{\geq 5}$ and $TD_{<5}$. The distinction is that the first set includes time series of 18 points (after imputation), while the second set includes time series of 5 points or fewer or noisy data featuring sequences of four or more consecutive zeroes. The two data sets represent all the data contained in the SDG data repository. Typically, we are interested in individual countries or geographic regions; in other words, the SDG data for a specific country or geographic region. A total of 38 countries were considered with respect to the work presented in this thesis. The final step in Stage 2 was therefore the generation of "country files"; data files that hold all the processed SDG data for a particular country or geographic region. For illustrative purposes Table 3.5 gives a fragment of the Afghanistan country file. The table features five records from the relevant $TD_{\geq 5}$ (clean) data subset and five records from the $TD_{<5}$ (noisy) data subset. The clean data was used to find dependencies (causalities/correlations) between the time series with respect to two of the frameworks considered in this thesis. Even though the Noisy Data cannot be used for identifying dependencies (there is not enough of it), it can still be used for *univariate forecasting*, the significance of this will become clear later in the thesis.

Years	Clean Data					Noisy Data				
	afghanistan_1.4.1_2269	afghanistan_1.4.1_2271	afghanistan_1.4.1_2270	afghanistan_1.4.1_2272	afghanistan_1.4.1_2274	afghanistan_15.1.2_2810	afghanistan_15.4.2_3213	afghanistan_15.4.2_3209	afghanistan_15.4.2_3211	afghanistan_15.1.2_2811
2000	-0.27428	-0.10011	-1.263	-1.19738	-0.9901	0				0
2001	-0.03908	-0.14823	-0.57359	-1.03868	-0.8754	0				0
2002	0.150842	-0.56668	0.021679	-0.88485	-0.76028	0				0
2003	0.29518	0.024638	0.522797	-0.73804	-0.64527	0				0
2004	0.395056	-0.64881	0.311655	-0.59146	-0.52845	0				0
2005	0.449014	0.533828	0.978321	-0.45229	-0.41206	0				0
2006	0.459213	-0.15604	1.644988	-0.318	-0.29524	0				0
2007	0.421092	0.698083	2.311655	-0.18859	-0.178	5.74028				0
2008	0.336188	-0.10677	2.644988	-0.05592	-0.05807	5.74028	71.02053	64.65522	75.03884	0
2009	0.213163	0.320297	2.311655	0.055915	0.058069	5.74028				0
2010	0.039078	-0.15604	0.978321	0.170674	0.17658	5.74028				0
2011	-0.17876	0.895189	-0.02168	0.280552	0.295514	5.74028				0
2012	-0.44267	-0.17247	-0.02168	0.385832	0.415199	5.74028				0
2013	-0.7512	-0.02464	-0.02168	0.462499	0.52913	5.74028	70.98188	64.53427	75.02409	0
2014	-1.105	1.387953	-0.02168	0.598924	0.659619	5.74028				0
2015	-1.50409	1.765739	-0.02168	0.671511	0.775814	5.74028				0
2016	-1.94973	1.471019	-1.53835	0.757205	0.897199	5.74028	70.58569	64.13984	74.39861	0
2017	-2.4395	1.734879	-2.3563	0.837778	1.018675	5.74028				0
2018	-1.53901	1.40055	-0.28354	1.068226	1.127194					

Table 3.5: Fragment of Afghanistan Country Data Files

3.4 SDGs Interconnectedness and Statistical Insights

This discussion section comprises two subsections: the first explores the interconnectedness among the Sustainable Development Goals (SDGs), and the latter focuses more into the statistical nature of the data.

Interconnectedness of Sustainable Development Goals

This subsection presents an overview of an experiment designed to understand the relational dynamics among the SDGs. The causal data used, referenced in Chapter 7 Table, 7.1, was compiled into a single dataset, stripped of all geographic identifiers and the ID definition as described in Equation 3.1. The data was then aggregated to the goal and target level

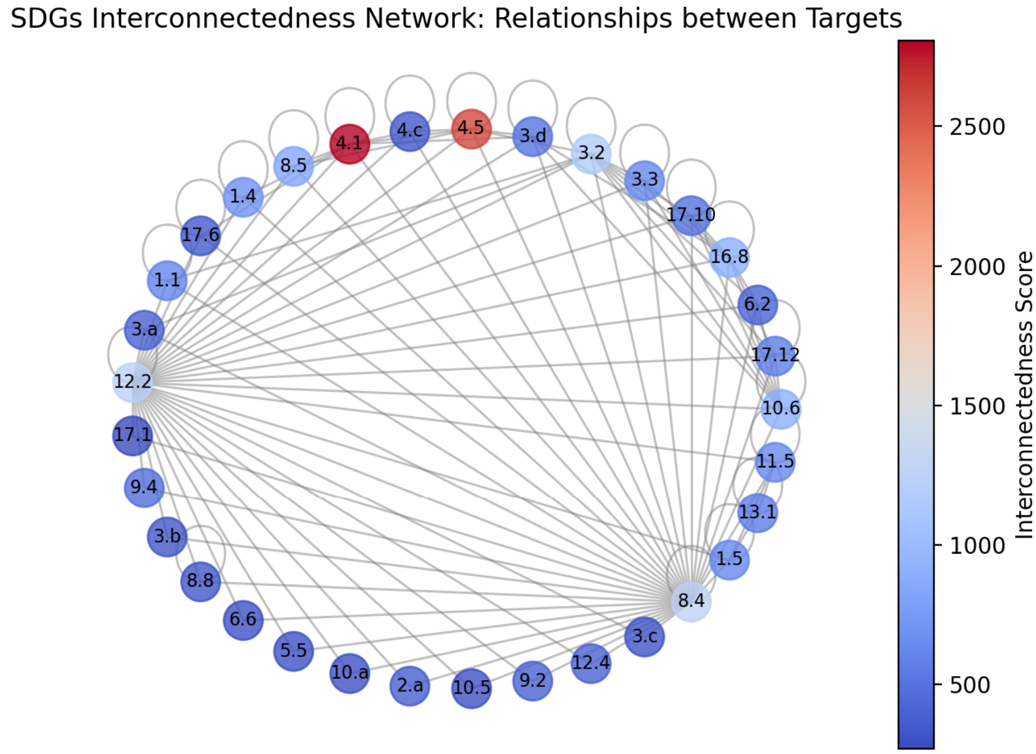


Figure 3.6: Visualising the Relationships and Interdependencies between Targets (Nodes with more than 350 connections

G.T, encapsulating numerous indicators. Table 7.1 presents the top 20 relationships for the country Algeria These were obtained using the SDG-TTF Fame work described later in this Thesis However the same framework can be used, but after stripping regional and low-level data, it can now be used to determine which target is most linked to other targets across all data from all countries in the thesis.

Figure 3.6 illustrates the interconnectedness generated from the experiment. The strength of these connections was determined by frequency . The total number of linked target pairs exceeded 7000. For demonstration purposes, a threshold was set to only display in Figure 3.6, those targets connected over 350 times. Notably, leaf nodes like G.T._4.1 and G.T._8.4 correspond to the SDG targets “By 2030, ensure that all girls and boys complete free, equitable, and quality primary and secondary education leading to relevant and effective learning outcomes” and “Improve resource efficiency in consumption and production” respectively.

Node “4.1” has the highest importance score, implying substantial interplay with many other SDG targets or indicators. Similarly, Node ”8.4” also has a high importance score, denoting its significant role in the SDG network.

The Statistical Nature of the SDGs Data

Goal	Unique Targets	Unique Indicators	Unique Sub-Indicators	Min	Max	Mean
1	4	12	20	0.00	22558608924.00	12094063.86
2	8	15	27	-4969.40	846289.00	5107.44
3	15	43	56	-1.74	846009271.00	289952.44
4	11	19	19	0.00	114.05	35.97
5	10	20	20	-50.00	4796.00	41.65
6	9	18	18	-9.99	800389.00	3020.63
7	6	8	8	0.00	4192.78	69.62
8	13	25	29	-62.60	10240430775.00	18288001.39
9	9	18	18	0.00	6890000000000.00	42280500121.53
10	9	16	16	-5859.67	63461.21	117.25
11	12	21	23	0.00	22558608924.00	34310422.06
12	10	18	18	0.00	10240430775.00	26857164.02
13	5	7	7	0.00	12384370.00	105424.07
14	9	11	11	-9.99	146784466.00	359861.82
15	17	33	36	-3.71	785374.22	4679.48
16	8	14	14	0.00	32915.74	106.13
17	19	37	41	-66346.39	1117190000.00	641109.08

Table 3.6: Detailed metrics for each SDG, including counts of unique targets, indicators, and sub-indicators, along with the minimum, maximum, and mean values.

The time series in this thesis are complex, captured in various measuring units as shown in Figure 3.4. The Table referenced in subsection 3.6 provides a granular summary of each Sustainable Development Goal (SDG), showcasing the unique targets, indicators, and sub-indicators that compose each goal. However it becomes evident that the SDGs are not equally represented in terms of their associated targets, indicators, and sub-indicators. For instance, the most complex goal, Goal 17 “Partnership for the Goals”, consists of 19 unique targets, 37 unique indicators, and 41 unique sub-indicators. In contrast, Goal 13 “Climate Action”, is relatively less complex with only 5 unique targets, 7 unique indicators, and 7 sub-indicators. It is also noteworthy that although Goal 17 comprises 19 targets, 37 indicators, and 41 sub-indicators, the total count of the time series related to this goal is not $19 * 37 * 41 = 28897$ individual components, but rather only 282. This is because not all of the 19 targets have 37 indicators. Furthermore, not all targets are relevant to all geographical regions or countries as mentioned earlier in the chapter . For example, landlocked countries would have fewer data points relevant to Goal 14“Life Below

Water". The complexity of the data, both in terms of composition and breadth, accentuates the need for a taxonomical representation, which will be discussed in more detail in the following chapter.The complexity of the SDGs data varies significantly from goal to goal, and even within individual indicators and sub-indicators of each goal. For instance, the maximum value for Goal 9 "Industry, Innovation, and Infrastructur" reaches up to 6.89 trillion, attributed to the "Passenger volume (passenger kilometres), by mode of transport" sub-indicator in India. This highlights how each country's specific circumstances and level of development can contribute to the diversity of the data. On the other hand, the minimum value for Goal 2, "Zero Hunger" is -4969.40, associated with an unusual food price anomaly reported by the Aland Islands in 2014. These instances underscore the complexities inherent in the data, including occasional outliers and extreme values.

3.5 Summary

This chapter has presented detail concerning the publicly available Sustainable Development Goal (SDG) data and the associated pre-processing of this data. The pre-processing was directed at generating the desired SDG taxonomy, for a particular country or region, together with the associated time series data to be held at the leaf nodes of the taxonomy. A major issue to be addressed was the prevalence of missing data and the many different units of measurement used. It was found that using a combination of Time interpolation and Robust scaling produced the best results. An important element not included in the discussion presented in this chapter was the derivation of the various thresholds required for the taxonomy to be used. This is discussed in the following chapter. The following chapter also discusses the proposed Bottom-up Hierarchical Classification process, the foundation for the work presented later in the thesis with respect to the SDG attainment prediction frameworks presented in Chapters 5 to 7, and the proposed mechanism for the visualisation of results.

Chapter 4

SDG Taxonomy Threshold Derivation and Utilisation

4.1 Introduction

The previous chapter presented the proposed mechanism for generating the SDG taxonomy to be used, in a bottom-up manner, to determine whether individual countries or regions will attain their SDGs or not. This chapter considers the operationalisation of this taxonomy. Recall that the central idea was that the leaf nodes of the taxonomy would hold prediction models which would be used to predict values. The predicted value would be compared to a threshold σ and a classification generated using the following identity, where c is the class value, v_{pred} is a predicted value and a σ a threshold of some kind:

$$c = \begin{cases} \text{Met} & \text{if } v_{pred} \geq \sigma \\ \text{Met} & \text{if } \sigma = \text{unknown} \\ \text{Not Met} & \text{otherwise} \end{cases} \tag{4.1}$$

An explanation of the "unknown" threshold will be presented later in this chapter. Whatever the case, the class is then passed up the hierarchy and combined with classifications from other leaf nodes to obtain an overall classification. For this to happen a set of thresholds $S = \{\sigma_1, \sigma_2, \dots\}$ is required. The challenge is that the in the SDG guidelines [4], thresholds are expressed in a textual manner; the thresholds are not explicitly specified in terms of a mathematical formula. For instance, if we consider SDG 1, "End poverty", there is no explicit high-level definition of what constitutes poverty. This is defined in terms of a set of targets and indicators, but even then, the definition is not defined as a mathematical identity but in textual terms. For example, Target 1.1 states that "*By 2030, eradicate extreme poverty for all people everywhere, measured as people living on less than $1.25 a day*". To derive a mathematical equation from this sentence requires manual analysis. Inspection of the Target 1.1 example sentence indicates two critical phrases: (i) the date by which the target should be achieved, in this case 2030; and (ii) and the word "eradicate". However, the word "eradicate" is not defined. An absolute value of zero could be used, however, this would mean that a country where 99.999% of the population lived on $1.25 or more a day would not be classified as not attaining the target; this is clearly not what was intended. An alternative strategy was therefore required. More specifically, before any attainment prediction could be made, an analysis of the SDG targets and indicators had to be undertaken so that mathematical attainment identities could be derived, which could then be used to measure the progress towards attainment of the SDGs.

This chapter therefore commences, Section 4.2, by presenting a proposed manual process whereby the thresholds to be associated with leaf nodes can be defined. The idea is to identify a categorisation of thresholds within the SDG guidelines and use this categorisation to identify the desired thresholds. The work addresses subsidiary research questions S3 from Chapter 1:

S3. Assuming a hierarchical taxonomy, how can a prediction label be derived for the root node of the hierarchy?

Once the set of thresholds S has been identified the remainder of the SDG taxonomy can be populated. Once populated the taxonomy is ready to be operationalised. Central to this operationalisation is the generation of predicted values to be compared to the thresholds. The process whereby such predictions can be generated is the subject for Chapters 5, 6 and 7 of this thesis. The general process whereby the taxonomy can be operationalised is common regardless of the prediction mechanism adopted. Results are stored in a *Country Table*. It was also considered important to provide a mechanism whereby the results could be visualised. The general bottom-up classification process, Country Table generation and outcome visualisation are thus presented in Section 4.3 of this chapter. The chapter is concluded in Section 4.4 with a summary of the work presented and the main findings.

4.2 Threshold Derivation

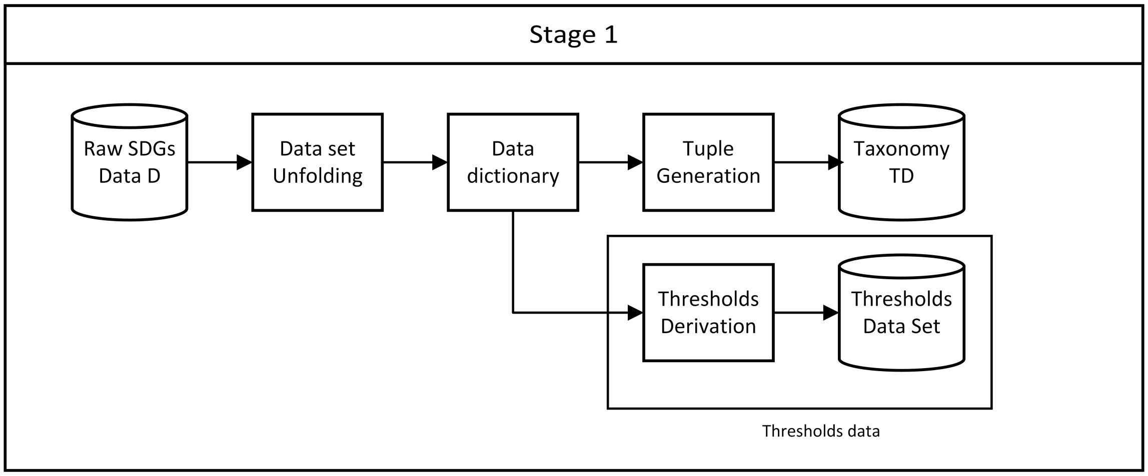


Figure 4.1: SDG Taxonomy Generation

Detail concerning the proposed approach for converting the textual SDG threshold information associated with targets and indicators is presented in this section. A schematic of the required SDG data pre-processing was presented in Chapter 3, Figure 3.1. This pre-processing comprised two stages: Taxonomy Generation and Missing Value Imputation and Scaling. The taxonomy generation schematic is presented again in Figure 4.1. The top row steps "Raw SDGs Data D", "Data set Unfolding", "Data dictionary", "Tuple Generation" and "Taxonomy TD" were discussed in Chapter 3. The bottom row steps, "Threshold Derivation" and the final "Threshold Data Set Σ" will be discussed here. Threshold derivation was conducted manually. However, it only needed to be done once (unless the SDGs change). The guidelines were manually applied to translate the textual descriptions of SDGs targets into mathematical ones for any keywords that match the guidelines.

The challenge of taxonomy threshold generation was noted in the introduction to this chapter. Some targets do include sufficient specific detail to allow the straightforward derivation of a threshold. There are many targets with ambiguous objectives. For example, Target 3.3 states "*By 2030 end the epidemics of AIDS, tuberculosis, malaria, neglected tropical diseases, combat hepatitis, water borne diseases and other communicable diseases*", and Target 3.5 states "*Strengthen the prevention and treatment of substance abuse, including* *narcotic drug abuse and harmful use of alcohol*". The keyword "end" could, arguably, be any number close to zero. Some targets are considered opposites, such as Target 2.5 "*By 2020, maintain the genetic diversity of seeds, cultivated plants and farmed and domesticated animals and their related wild species, including through soundly managed and diversified seed and plant banks at the national, regional and international levels, and promote access to and fair and equitable sharing of benefits arising from the utilisation of genetic resources and associated traditional knowledge, as internationally agreed*". Indicator 2.5.1 refers to "*Number of local breeds kept in the country*" and Indicator 2.5.2 to "*Local breeds classified as being at level of risk of extinction (number)*"; the goal is to increase the first while decreasing the second one.

Inspection of the SDG documentation [5] indicates that the way that thresholds are expressed can be categorised according to a small number of "templates" as follows:

1. Explicit Threshold Included
2. Floor and Ceilings
3. Rankings
4. Ratios
5. Fixed Percentiles
6. Unknown Thresholds or Date

Each of these templates will be considered in further detail later in this section. However, the point to note here is that there are a small number of templates, and hence mechanisms can be devised whereby each can be processed. The start point for the work was to refer to the existing literature on SDG attainment and analysis. As noted in Chapter 2 there has been previous work directed at analysing SDG data [23, 24, 27, 10]. Although, as also noted in Chapter 2, much of this previous work was directed at individual goals, targets or indicators rather than a holistic view of SDG attainment as considered in this thesis. From the literature authors have published guidelines on how to interpret the health-related targets, including mathematical definitions. An example is given in Figure 4.2, taken from [2], for Target 2.2 "*By 2030, end all forms of malnutrition, including achieving, by 2025, the internationally agreed targets on stunting and wasting in children younger than 5 years, and address the nutritional needs of adolescent girls, pregnant and lactating women, and*

	Health-related SDG indicator	Indicator definition	Currently measured by GBD	Further details	SDG target	SDG target used in this analysis
Goal 1: End poverty in all its forms everywhere						
Target 1.5: By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social, and environmental shocks and disasters	Disaster mortality (1.5.1; same as indicators 11.5.1 and 13.1.1)	Death rate due to exposure to forces of nature, per 100 000 population	Yes	Existing datasets do not comprehensively measure missing persons and people affected by natural disasters; we thus report deaths due to exposure to forces of nature	Undefined	..
Goal 2: End hunger, achieve food security and improved nutrition, and promote sustainable agriculture						
Target 2.2: By 2030, end all forms of malnutrition, including achieving, by 2025, the internationally agreed targets on stunting and wasting in children younger than 5 years, and address the nutritional needs of adolescent girls, pregnant and lactating women, and older people	Child stunting (2.2.1)	Prevalence of stunting in children younger than 5 years, %	Yes	Stunting is defined as below −2 SDs from the median height-for-age of the WHO reference population. No indicator modifications are required	Eliminate by 2030	≤0·5%
Target 2.2 (as above)	Child wasting (2.2.2a)	Prevalence of wasting in children younger than 5 years, %	Yes	We have separated reporting for indicator 2.2.2 into wasting (2.2.2a) and overweight (2.2.2b). Wasting is defined as below −2 SDs from the median weight-for-height of the WHO reference population	Eliminate by 2030	≤0·5%
Target 2.2 (as above)	Child overweight (2.2.2b)	Prevalence of overweight in children aged 2–4 years, %	Yes	We have separated reporting for indicator 2.2.2 into wasting (2.2.2a) and overweight (2.2.2b). We used the IOTF thresholds because the WHO cutoff at age 5 years can lead to an artificial shift in prevalence estimates when the analysis covers more age groups. Furthermore, considerably more studies use IOTF cutoffs, which allowed us to build a larger database for estimating child overweight	Eliminate by 2030	≤0·5%

Figure 4.2: Health goals conversion courtesy of "Measuring progress from 1990 to 2017 and projecting attainment to 2030 of the health-related Sustainable Development Goals for 195 countries and territories: a systematic analysis for the Global Burden of Disease Study 2017" [2]

older people". In this example, the keyword 'end' was converted into a 0.5%, on the opposite side, 99.5% was used.

As stated in the introduction to this chapter, a significant challenge facing attainment prediction for measuring SDG progress is the lack of clear guidelines published by the UN to assess the progress towards SDG attainment. Measuring, in general, requires a starting point, an endpoint, and a threshold of some sort to be achieved. The starting point for SDGs is the year 2015 [5], and the endpoint in most cases is 2030, with some targets set to be achieved before 2030. On the other hand, numerically defined thresholds are much more difficult to obtain. At the start of this section, a number of templates were listed, and the point was made that each can be processed in a bespoke manner. The processing associated with each template is discussed in the remainder of this section. Although the total number of thresholds in the entire data set exceeds 3000, the unique ones are attached to Appendix A A.1.

Explicit Threshold Included

Recall that the SDGs comprise a total of 169 different targets spread over 17 goals. A small number of the targets can be immediately converted to a mathematical representation. For example, Target 3.1 states "*By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births*". In other words, the threshold equates to:

$$\sigma = \frac{\text{live births} \times 70}{100,000} \tag{4.2}$$

For this example the geographical entity in question needs to meet this target by 2030. Target 3.4 states "*By 2030, reduce by one third premature mortality from non-communicable diseases through prevention and treatment and promote mental health and well-being*". The associated Indicator 3.4.1 states "*Mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease*" and 3.4.2 "Suicide mortality rate" for this target to be considered to be met the forecasted value for 2030 must be <33.33% of the 2015 value.

Floors and Ceilings

There are many targets where the goal is to "Eliminate", "End", "attain universal access" or "Ensure All". For example, to bring some value x down to 0% or 100%. As already noted, achieving absolute 0% or 100% is not realistic. Thus for these targets, the concept of floors and ceilings was used. A ceiling of 99.5% for maximum coverage and a floor of 0.05% for minimum coverage.

Rankings

It is not unusual for an SDG target to be expressed in the form of a set of rankings. For example "1 = Requires further progress; 2 = Partially meets; 3 = Meets; 4 = Fully meets", or "legislative and/or regulatory provision has been made for managing disaster risk (1 = YES; 0 = NO)". In such a case the conversion to a mathematical equivalence could be easily undertaken, because of the way in which the target was expressed.

Ratios

Several targets, mainly health related targets, reference a ratio. For example "By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births" or "By

2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births". Again each was directly converted into the sentinel number.

Fixed Percentiles

A small number of SDG targets are directed at a specific increase or decrease expressed in terms of percentages. For example, "By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions", or "By 2030, double the global rate of improvement in energy efficiency". The thresholds associated with such targets were expressed accordingly.

Unknown thresholds or date

In some cases, either the date is not specifically mentioned or the textual target has no specific achievement. For example indicator 4.6.1 references the "Proportion of population achieving at least a fixed level of proficiency in functional skills, by sex, age and type of skill (%))", however, the "fixed level" is not documented. Where a threshold cannot be determined it is specified as "Unknown". As shown in Equation 4.1. If the target date is not explicitly given, it is presumed to be 2030.

4.3 Bottom up Classification and Visualisation

Once all thresholds have been calculated and incorporated into the taxonomy the next stage is to calculate the individual predictors to be held at the leaf nodes. How these predictors can best be generated is the central focus of this thesis. However, a high-level view of the classification process is presented here, in Sub-section 4.3.1, in preparation for the work presented in later chapters. A schematic of the prediction process is presented in Figure 4.3. On the left of the figure we have forecasting results (derived using one of the prediction methods considered later in this thesis) and the derived set of thresholds $S = \{\sigma_1, \sigma_2, \dots\}$. These are combined to do the Bottom-up Classification. The result of the classification, as can be seen from the figure, is a *Country Table* and a visualisation. Each of these will be discussed in further detail later in this section in Sub-sections 4.3.2 and 4.3.3.

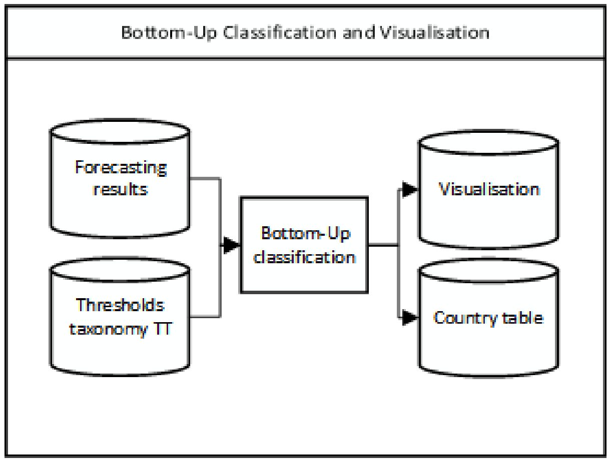


Figure 4.3: The Bottom-Up classification

4.3.1 Bottom-up Classification

Recall that predicting whether a geographic region will reach its SDGs is based on whether the lower levels classes meet their targets (thresholds) on time. This is demonstrated in Figures 4.4,4.5,4.6, presents Goal 3 targets and thresholds. For those thresholds related to the targets the forecasted values must be better than the threshold. The total number of thresholds for Goal 3 is 13, each uniquely different. For instance, the threshold ≤ 12 implies that every time series related to GTI 3.1.2 or GTI 3.2.2 must be less than or equal to the target, which is 12 or fewer per 1000, as can be observed in 4.4. Once a forecasted value is converted to a label then the classifications made at the leaf nodes are "passed-up" the hierarchy as shown in Figure 4.7. The general idea is shown in Figure 4.7(a); and a specific example, using the country Afghanistan, in Figure 4.7(b). The forecasting models considered in later chapters in this thesis are used to make predictions for individual indicators (leaf nodes in the topology), the predictions are compared to the threshold values held in the Thresholds Taxonomy generated as described in the previous section. The results are then passed up the SDG topology hierarchy to the root node. At each intermediate node, a Boolean "and" operation, Equation 4.3, will be applied, and the result passed further up the tree till the root node is eventually reached. In Equation 4.3 the function $condition(f_{i_j}, \omega_j)$ returns true if the condition expressed by ω_j holds true with respect to the forecast value f_{i_j}. The condition expressed by ω, at is simplest, might be a threshold which the value f_{i_j} should be greater than or less than for the condition to be

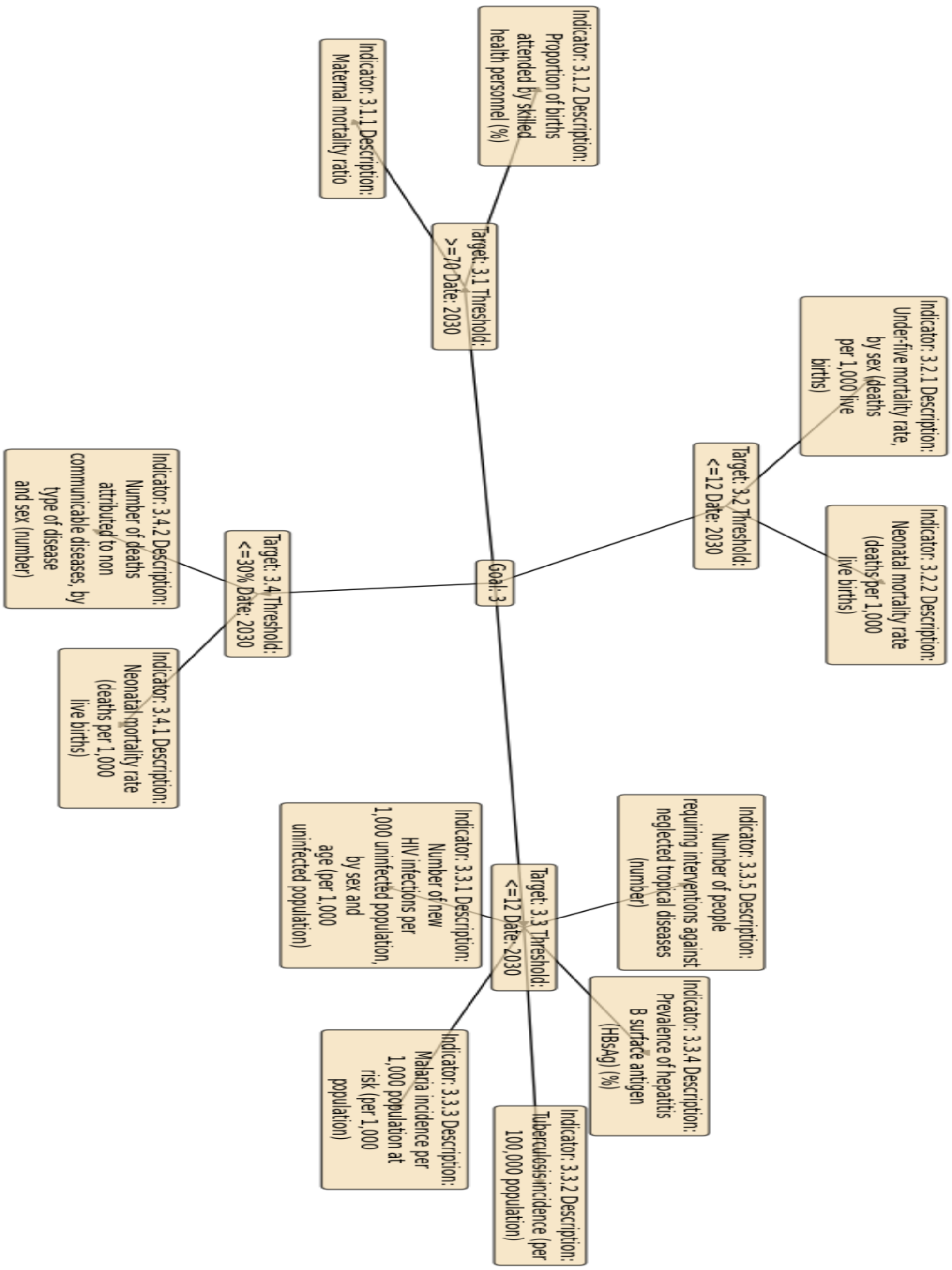


Figure 4.4: Goal 3 thresholds and targets part1

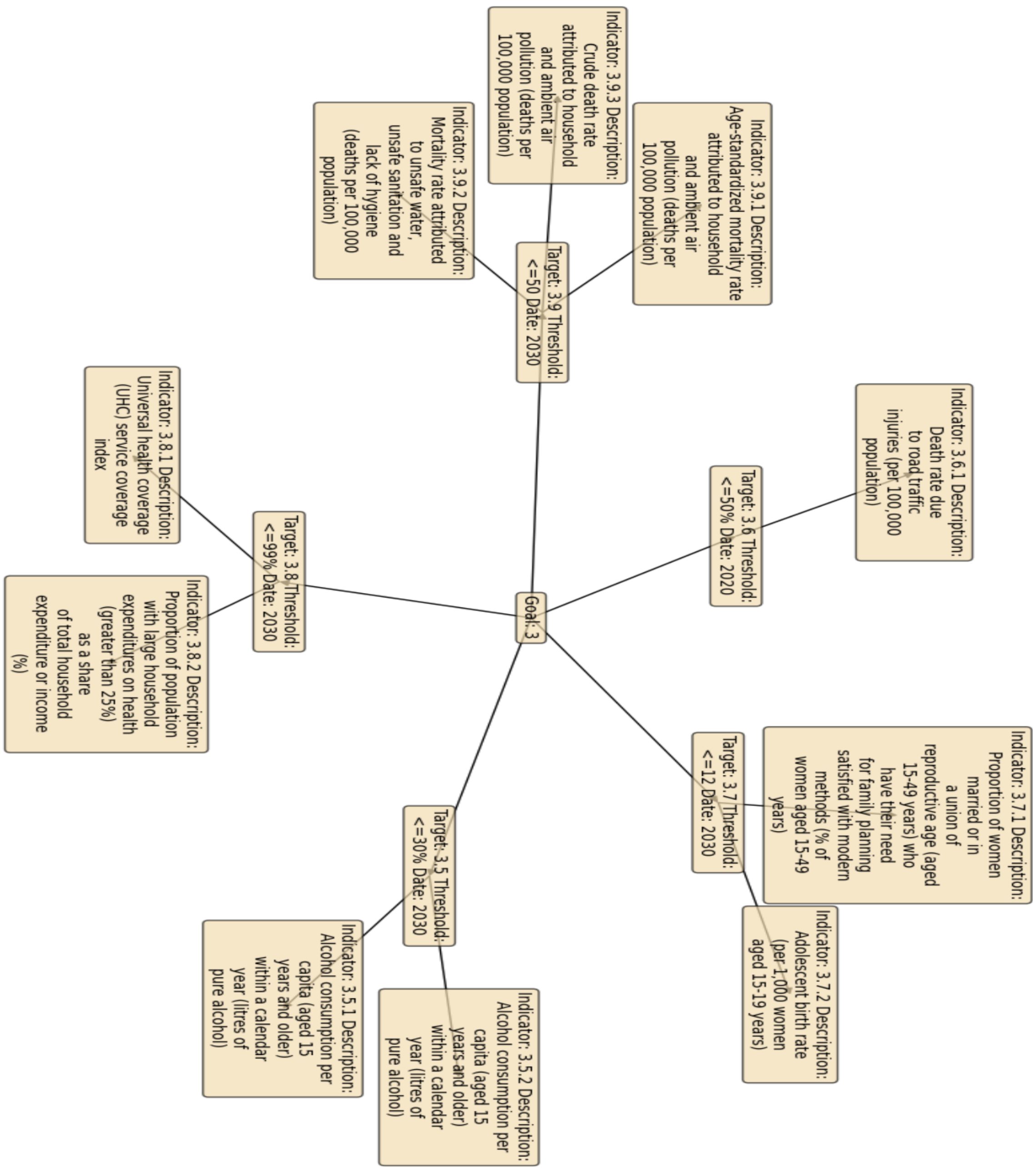


Figure 4.5: Goal 3 thresholds and targets part2

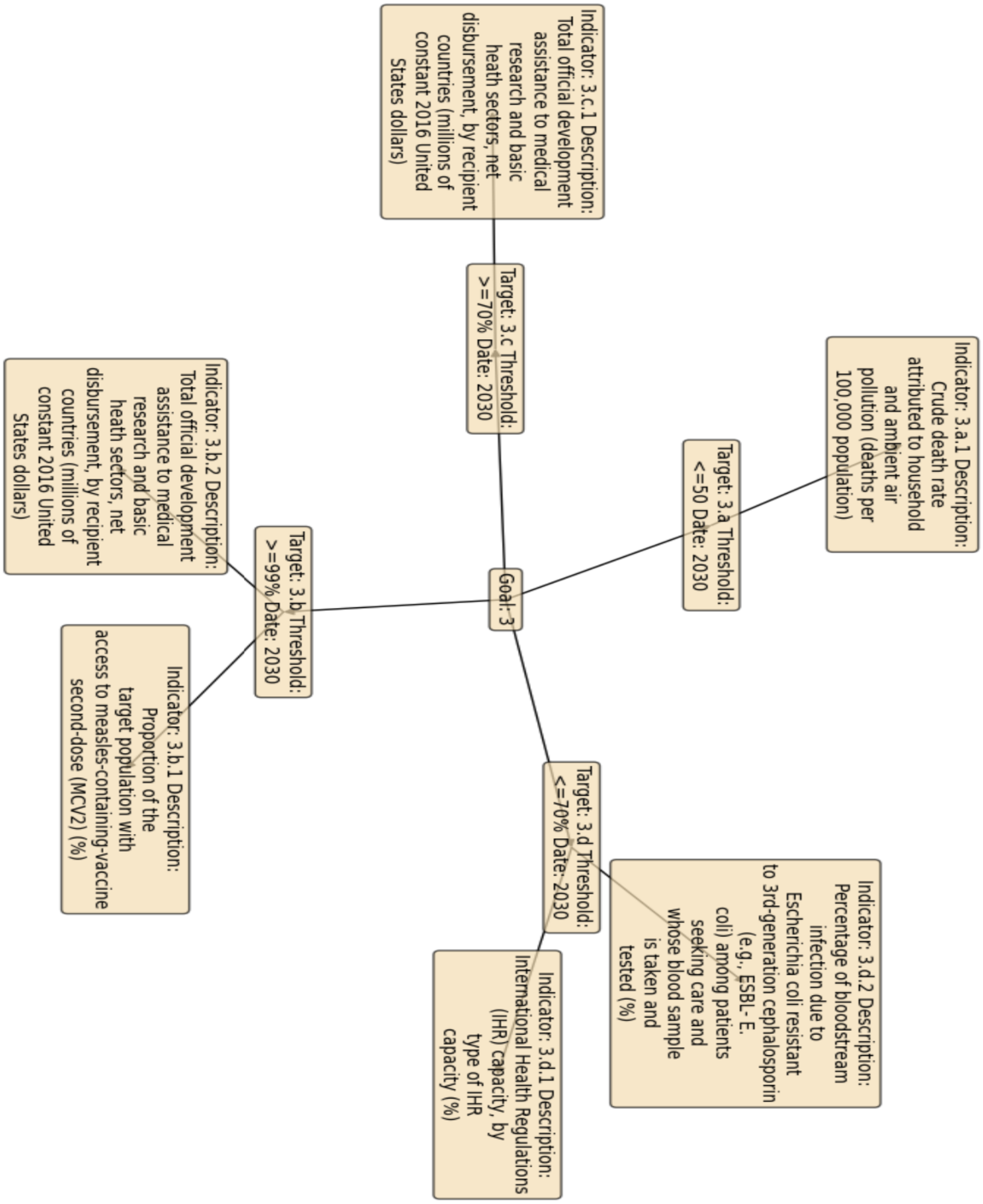


Figure 4.6: Goal 3 thresholds and targets part3

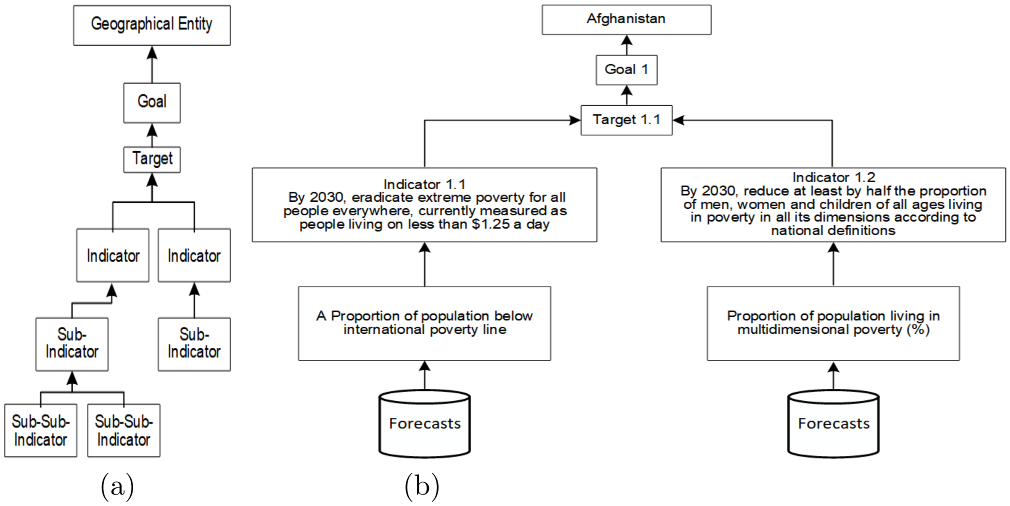


Figure 4.7: Hierarchical Classification; (a) The general nature of SDGs hierarchy, (b) SDGs target 1.1 hierarchy

true. Thus the set of attributes A has a corresponding set of conditions $\Omega = \{\omega_1, \omega_2, \dots\}$.

$$class(e_i, F_i) = \begin{cases} true & if\ \forall f_{i_j} \in F_i\ condition(f_{i_j}, \omega_j) = true \\ false & otherwise \end{cases} \tag{4.3}$$

The final result as to whether a given country will or will not attain its SDGs will culminate at the root node of the taxonomy. The results are stored in a *Country Table* to which a visualisation can be applied, as discussed in the following two sub-sections.

4.3.2 Country Table

The results of the binary classification were stored in a tabular format, referred to as a *Country Table*. A fragment of a Country Table is given in Table 4.1. A full Country Table will, on average, comprise some 600 rows. In the table the first column gives the GTI (Goal-Target-Indicator) reference. For example, GTI 3.2.1 is concerned with "children mortality rate". The"Meta data" column gives the leaf node sub-indicator, age and gender in this case. The"Initial" column gives the initial value for the year 2015, the reference value. The "Target" column then gives the threshold (calculated as described above). The "Forecast" gives the predictions. The result column indicates whether the threshold has

Target	**Meta Data**	**Threshold**	**2000**	**2001**	**2002**	∼	**2013**	**2014**	**2015**	**2016**	**2017**
3.1.1	Ubber Bound	<=70	60	57	54	∼	39	38	**37**	**37**	**37**
3.2.1	1Y	<=25	31.6	30.1	28.8	∼	18.5	17.8	**17.2**	**16.7**	**16.1**
	Mid point	<=25	33.5	32	30.6	∼	19.9	19.2	**18.5**	**17.9**	**17.3**
	Females <1Y	<=25	35.3	33.8	32.3	∼	21.2	20.4	**19.8**	**19.1**	**18.5**
	Males <1Y	<=25	41.5	39.4	37.5	∼	23.4	22.6	**21.7**	**21**	**20.3**
	Females <5Y	<=25	43.2	41	39.1	∼	24.7	23.9	**23**	**22.2**	**21.5**
	Males <5Y	<=12	20.8	20.1	19.5	∼	12.8	12.4	**12**	**11.5**	**11.1**

Table 4.1: A fragment of the Egypt country table. It includes two targets: 3.1.1, 'By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births', and 3.2.1, 'By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births, and under-5 mortality to at least as low as 25 per 1,000 live births'. Values in bold text indicate that the goal was reached.

been met by the required date or not, either" "Met", "Not Me" or "Uknown". The result "unknown" is used where it has not been possible to determine a threshold (as discussed in the foregoing section). To demonstrate that the bottom-up classification works, thresholds were applied to the data from the year 2015. It was found that Egypt met its goals for those GTIs starting from that year. For the GTI presented on the last row of the table, the target was met only starting from 2015. However, if the threshold had been applied to the data from 2014, the target would not have been met.

4.3.3 Visualisation

It was considered that a simple yes/no SDG attainment prediction was not sufficient where the result was "no". Consultation with potential end users[1] indicated that in the event of a "no" prediction, the reason for the prediction would be of interest. Further investigation and discussion indicated that some means of visualising the result would serve to explain a SDG attainment prediction. End users should be able to view the taxonomy. This was implemented using the D3.js JavaScript library for producing dynamic and interactive data visualisations [87]. An example of the result is given in Figure 4.8 using predictions for the country Afghanistan. Nodes coloured in green highlight goals, targets, indicators and sub-indicators that will be attained on time. Nodes coloured in red highlight goals,

[1]Staff at the University of Liverpool's School of Environmental Sciences, https://www.liverpool.ac.uk/environmental-sciences/

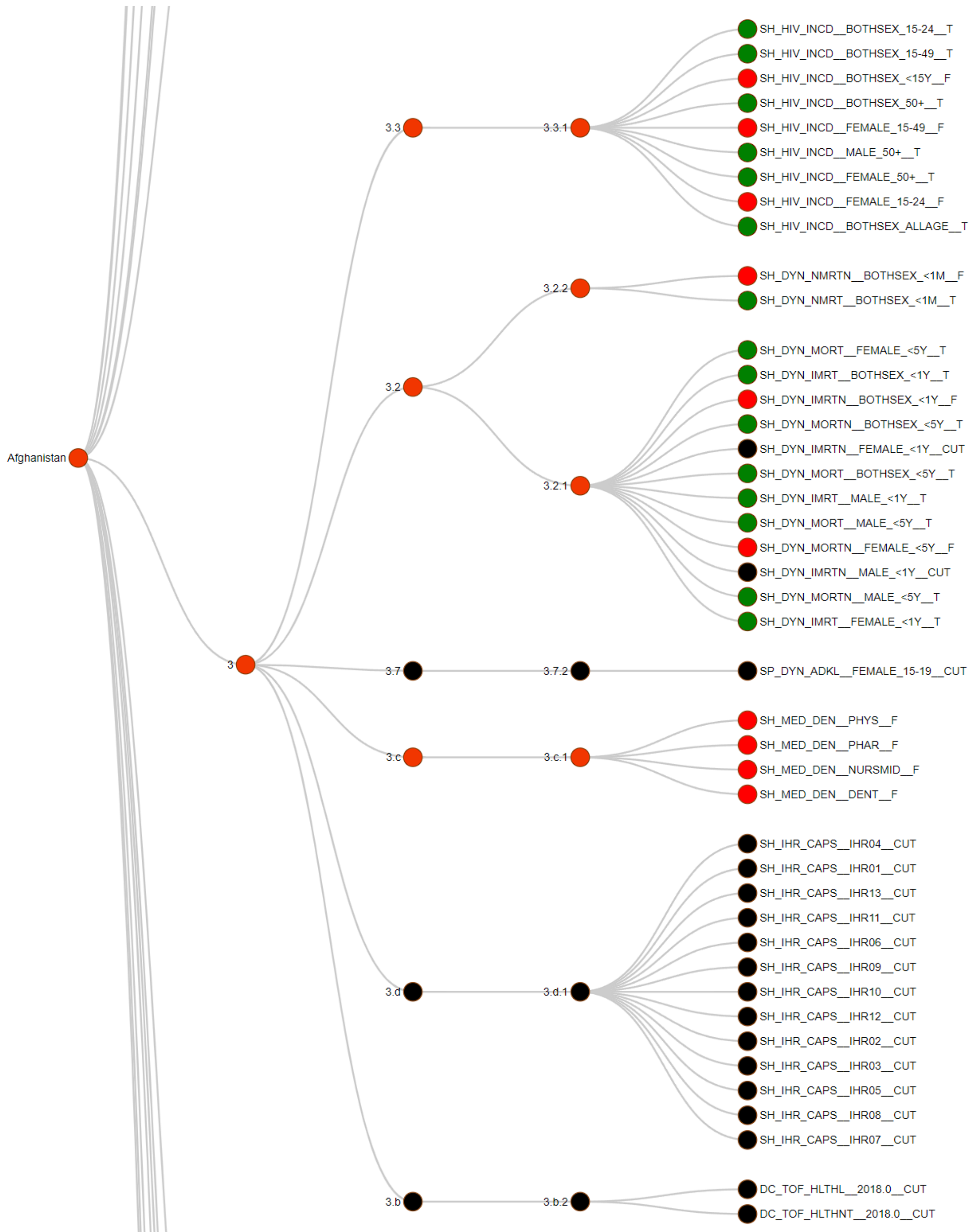


Figure 4.8: SDG taxonomy visualisation generated using the D3.js JavaScript library

targets, indicators and sub-indicators that will not be attained on time. Nodes coloured in black represent goals, targets, indicators, and sub-indicators where the thresholds are unknown.From the figure, it can be seen that using the visualisation it is easy to identify goal attainment (or non-attainment, as in the case of the example). The visualisation allowed: (i) inspection of the taxonomy with respect to individual SDGs of interest, and (ii) tracing of individual leaf node predictions as they were "passed up" the taxonomy.

4.4 Conclusion

This chapter commenced by presenting a manual process whereby SDG thresholds could be derived given that specific mathematical threshold identities were not included in the SDG guidelines. The threshold generation process was founded on the idea of a categorisation of the SDG targets and indicators into a set of threshold templates. Once generated the SDG taxonomy could be populated with the thresholds at which stage the taxonomy was ready for use. A criticism of the proposed threshold derivation mechanism was that it was a manual approach. Given the presented categorisation it might be possible to automate this process using the tools and techniques of Natural Language Processing (NLP) and the derived templates, however, this was considered beyond the scope of this thesis and therefore left as a topic for future work. The chapter went on to consider the utilisation of the SDG taxonomy, in general terms, and the visualisation of the results. In combination, the material presented in this chapter was designed to provide an answer to Subsidiary Research Question S3. The following chapter presents the first of the SDG prediction mechanisms presented in this thesis. Namely, the Sustainable Development Goal Attainment Prediction (SDG-AP) framework, the first of the SDG prediction mechanisms presented in this thesis, which adopts an "independence assumption" and therefore uses a univariate time series approach to prediction model generation.

Chapter 5

The Sustainable Development Goal Attainment Prediction Framework

5.1 Introduction

This chapter presents the first Sustainable Development Goal (SDG) attainment predictions framework considered in this thesis, the Sustainable Development Goal Attainment Prediction (SDG-AP) framework. This is a univariate approach, where each time series T in a given country's SDGs data is considered to be entirely independent of any other time series pertaining to the same country or any other country. The SDG-AP framework presented in this chapter was also proposed as a base model, with which the more sophisticated models presented in the following chapters could be compared. The work presented in this chapter is directed at subsidiary research questions S2 and S3 from Chapter 1:

S2. How best can machine learning be used to forecast whether individual SDGs will be met?

S3. Assuming a hierarchical taxonomy, how can a prediction label be derived for the root node of the hierarchy?

As noted earlier, the hierarchical taxonomy, on which all the frameworks presented in this thesis are based, required that predictors (forecasters) were held at the leaf nodes, and that the outcomes from the predictors were "passed up" the taxonomy. Recall that the taxonomy was designed to answer questions such as "will a geographic area x meet its goal y by time t". The forecasting approach incorporated into all the frameworks presented in this thesis can thus be categorised as a "bottom-up hierarchical forecasting" approach. In each case, the predictors associated with the taxonomy's leaf nodes were built using the time series available from the UN SDG data set preprocessed as described in the preceding two chapters, to form a data set $TD = \{TR_1, TR_2, \dots\}$, where each $TR_i \in T$ is a tuple of the form:

$$TR_i = \langle Geographical\_Region, GTI, ID, T \rangle \tag{5.1}$$

as presented previously in Equation 3.1. The remaining nodes in the hierarchy held simple Boolean (yes-no) operators that returned an output in a deterministic manner according to the input from their "child" nodes. It's noteworthy to mention that the data used in this process was normalised, which led to a certain level of leakage. This is an issue that requires further attention and will be considered for future work.

For reference, Figure 5.1 shows a number of example time series for the geographical regions Sweden, United Kingdom, Ireland and Isle of Man. For each country four time series are presented, each associated with a different leaf node (indicator) in the taxonomy. Each indicator is labelled with its identifier as was described in Chapter 3, for example "sweden_1.4.1_2270". Although the nature of the particular indicator in each case is not of relevance here, from the figure it can be seen that a trend line can be observed in most cases, thus suggesting that predictions can be made.

The forecast models at the leaf nodes were required to predict what the value associated with the indicator in question would be at time t and then to determine whether that value met its specified threshold value σ or not. For the work presented in this chapter, four different predictions (forecasting) models were investigated: (i) Auto-Regressive Moving Average (ARMA) [88], (ii) Auto-Regressive Integrated Moving Average (ARIMA) [89], (iii) Facebook Prophet (FBprophet) [86] and (iv) Univariate Long Short Term Memory (LSTM) forecasting. These forecasting models were selected because of their popularity in the literature.

The rest of this chapter is structured as follows. Section 5.2 provides an overview of the proposed SDG-AP framework. Section 5.3 presents the evaluation of the proposed framework. Section 5.4 then concludes the chapter with a summary of the work presented,

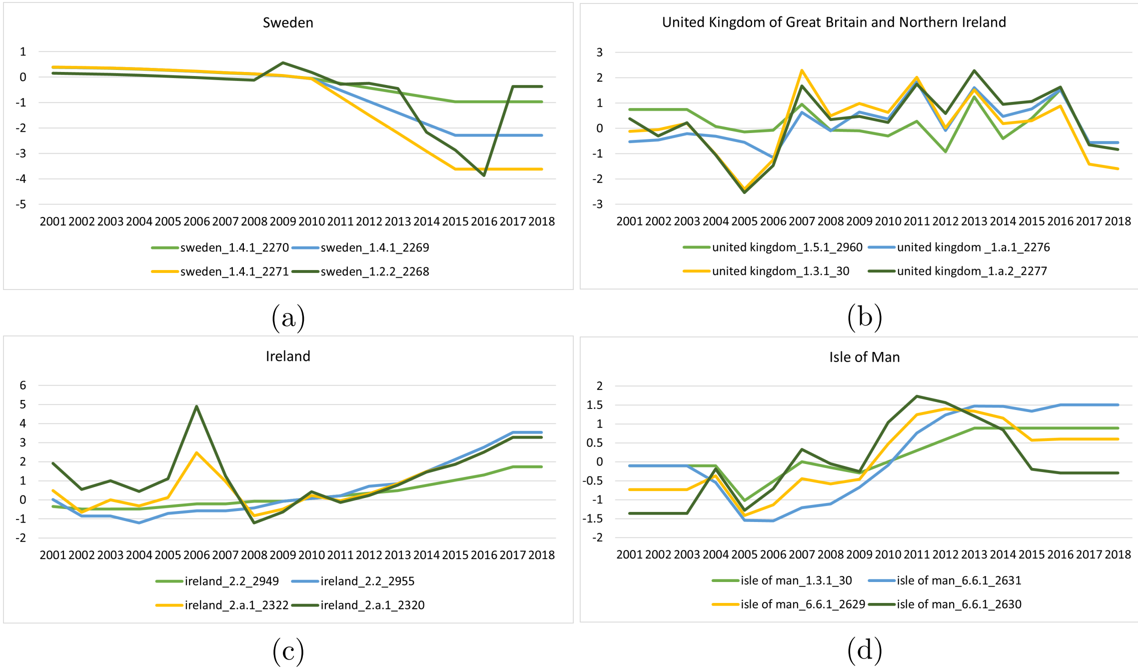


Figure 5.1: A number of example SDG time series

and the main findings.

5.2 The SDG-AP Framework

Detail concerning the proposed SDG-AP framework is presented in this section. A schematic of the framework is presented in Figure 5.2. The input to the framework is the SDG data for the country of interest that has been pre-processed as discussed in Chapter 3. Recall that, during the pre-processing, the input data was separated into two categories: the set of time series with less than five values, $TD'_{<5}$, and the set of times series with five values or more, $TD'_{\geq 5}$. However, for the SDG-AP framework $TD'_{\geq 5}$ was split into two: $TD'_{<5\&\geq 3}$ and $TD'_{<3}$. The sets $TD'_{<5}$ and $TD'_{<5\&\geq 3}$ were combined to $TD'_{\geq 3}$. The set $TD'_{<3}$ was discarded. Time series with less than three observations were discarded because it would not be possible to construct a meaningful trend line with just one or two observations. The output was a Country Table of the form described previously in Sub-section 4.3.2 and a D3.Js visualisation. In between the input and output, as indicated

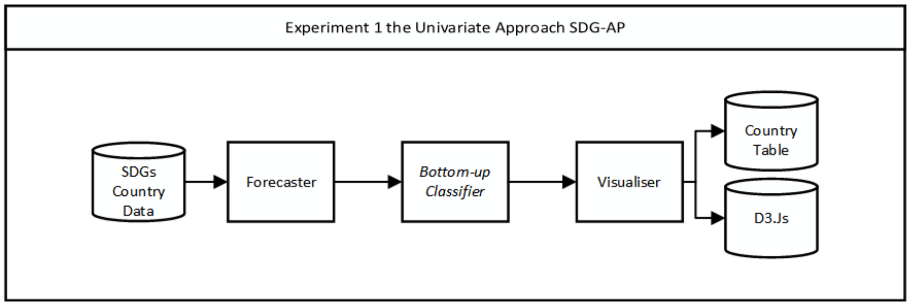


Figure 5.2: The SDG-AP Framework

by the schematic given in Figure 5.2 are three processing steps: (i) forecasting, (ii) bottom up classification and (iii) visualisation. Detail concerning each of these processing steps will be discussed in the following three sub-sections.

5.2.1 Forecasting

Recall, from Chapter 4, that the adopted approach was to conduct a bottom-up classification process starting with forecast models held at the leaf nodes of the taxonomy described in Chapter 3. The first step within the SDSG-AP framework was thus to generate the forecast models from the given data. Whatever forecasting (prediction) model is incorporated into the framework it must be able to operate using short time series. Recall that the current maximum length for any SDG time series will not exceed 22 observations. As noted above, four different prediction models were considered: (i) Autoregressive moving average (Arma) [88], (ii) AutoRegressive Integrated Moving Average (Arima) [89], (iii) FaceBook Prophet (FBProphet) [86] and (iv) Univariate Long short-term memory (LSTM).

5.2.2 Bottom-Up Classification

The leaf node forecast models, once generated, were used to make predictions for individual indicators (leaf nodes in the topology) which were then compared to the threshold values held in the Data Dictionary generated as described in the previous chapter. The results were then passed up the SDG topology hierarchy to the root node. At each intermediate node a Boolean "and" operation was applied and the result passed further up the tree. The final result was a Boolean SDG attainment prediction and a Country Table, for the input geographic region (country).

5.2.3 Visualisation

The third processing step with the proposed SDG-AP framework was the visualisation of the results held in the Country Table. The visualisation, as already noted, was in the form of a D3.js visualisation generated using the D3.js JavaScript library. An example of a D3.js visualisation was given in Figure 4.8 in the previous chapter. The same example is considered later in this chapter.

5.3 Evaluation

The evaluation of the proposed SDG-AP framework is presented in this section. The objectives of the evaluation were:

1. To identify the most appropriate forecast model generation technique to be used to create the forecast models to be held at the leaf nodes of the taxonomy.

2. To analyse the operation of the proposed framework as a whole.

The experiments were conducted using the following countries from the SDG country list:

North Africa: Algeria, Egypt, Libya, Morocco, Sudan, Tunisia and Western Sahara.

South Asia: Afghanistan, Bangladesh, Bhutan, India, Iran, Maldives, Nepal, Pakistan and Sri Lanka.

Northern Europe: Aland Islands, Denmark, Estonia, Faroe Islands, Finland, Iceland, Ireland, Isle of Man, Latvia, Lithuania, Norway, Svalbard and Jan Mayen Islands, Sweden and United Kingdom.

Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, and Panama

Recall from Section 1.4 that the selected countries were chosen to encompass a broad geographical and economic range. Collectively these countries referenced 23,068 time series covering the 17 SDGs. After pre-processing (see Section 3.3) the sets $TD'_{\geq 5}$, when combined, comprised 8,629 time series; whilst the sets $TD_{<5\&\geq 3}$ 6,430, when combined, comprised 15,059. imputation and scaling where applied as discussed in section 3.3 .

Experiments were conducted using a windows machine running Python 3 on a Ryzen 9 processor with 40 Gig of ram and Nvidia RTX2060 GPU. The total run time for the experiments was 35 hours. The remainder of this section comprises two sub-sections, Sub-sections 5.3.1 and 5.3.2, directed at the two evaluation objectives listed above respectively.

5.3.1 Forecast Model Selection

Four forecasting models were considered: (i) ARMA, (ii) ARIMA, (iii) FBProphet, and (iv) Univariate LSTM, in order to find the most suitable forecast model generation technique to be used to create the forecast models to be held at the leaf nodes of the taxonomy.

ARMA, ARIMA and FBprophet were evaluated by simply dividing each time series T_i into a training part, a validation part and a test part by defining the last n values to be the test part; $n = 4$ was used in this case. The right-hand panel in Table 5.1 shows an example evaluation record for ARMA, ARIMA and FBprophet forecasting. From the panel, it can be seen that the training part comprises 12 observations, and the test and validation parts are 4 observations each. The next 11 points, up to 2030, in the time series were then forecast (recall that we have reliable data for 2000 to 2019).

The input required to train an LSTM model [90] is different to the straightforward training input for ARMA, ARIMA and FBprophet, as described above. For the evaluation presented here, each time series T_i was divided into a training and test set in the same manner as before, reserving the last $n = 4$ values for the test set. However, training an LSTM model requires the training set to be divided into a set of overlapping input / output *samples*. Each sample comprised of three input values, followed by an output value. Each sample was used as a *prediction step* in the overall LSTM model generation process. The left-hand panel in Table 5.1 shows an example LSTM training record. The first column lists the sample IDs in terms of the prediction step number, the second column the year for the associated value, and the third and fourth columns the values for the sample input and output as appropriate.

The evaluation metrics used with respect to all four forecasting models considered were: Root Means Square Error (RMSE) and Means Absolute Percentage Error (MAPE) [91]. RMSE values were calculated as shown in Equation 5.2, where f is the forecast value and o is the observed value. RMSE provides results with the same unit of measurement as the forecast values, it is therefore easy to compare RMSE values generated by alternative forecasting methods. However, it is not an intuitive measure. MAPE is calculated as

Univariate LSTM data shape			
Steps	**Year**	**Input**	**Output**
Train 1	2000	-0.93993	
	2001	-0.93993	
	2002	-0.8371	
	2003	Validate	-0.73351
Train 2	2001	-0.93993	
	2002	-0.8371	
	2003	-0.73351	
	2004	Validate	-0.62962
Train3	2002	-0.8371	
	2003	-0.73351	
	2004	-0.62962	
	2006	Validate	-0.40462
***	***	***	***
Test1	2014	0.665215	
	2015	0.665215	
	2016	0.665215	
	2017		0.665215
Test2	2015	0.665215	
	2016	0.665215	
	2017	0.665215	
	2018		0.972637

ARMA, ARIMA, FBprophet		
Year	**Ts**	**Steps**
2000	-0.93993	Train
2001	-0.93993	
2002	-0.8371	
2003	-0.73351	
2004	-0.62962	
2005	-0.40462	
2006	-0.29048	
2007	-0.17521	
2008	-0.05879	
2009	0.058793	
2010	0.177561	
2011	0.297533	Validate
2012	0.418749	
2013	0.541211	
2014	0.665215	
2015	0.665215	Test
2016	0.665215	
2017	0.665215	
2018	0.972637	

Table 5.1: Example training, validation and test data split for Univariate LSTM, ARMA, ARIMA and FBpropohet forecast model generation

shown in Equation 5.3 where f is the forecasted value, n the sample size, which is 4 and o is the observed value for each $T_i \in TD'_{\geq 3}$ then an average was taken for all the results within each country. MAPE offers an easy to understand forecasting error expressed in terms of a percentage.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5.2}$$

$$\text{MAPE} = \left(\frac{1}{n}\sum \frac{\text{o-f)}}{\text{o}}\right) * 100 \tag{5.3}$$

The results are presented in Table 5.2. In the table, the countries are listed in column one. The overall RMSE and MAPE values and number (count) of correctly forecast time series are given in the following three column blocks, one block per forecast model generation method. The count value represents a simple counter of how many time series were successfully forecasted for any given country. In many cases a model can fail to successfully forecast a time series due to the time series nature self such as the data contains too many zero values, extremely large or small values. The overall average RMSE and MAPE, and the count average, are given in the penultimate rows. The final row gives the "Total Number of Time Series" that were correctly forecasted. A critical difference diagram is presented in Figure 5.3 encompassing the four techniques considered. From the diagram it can be seen that here is no clear statistical difference between the operation of the four techniques

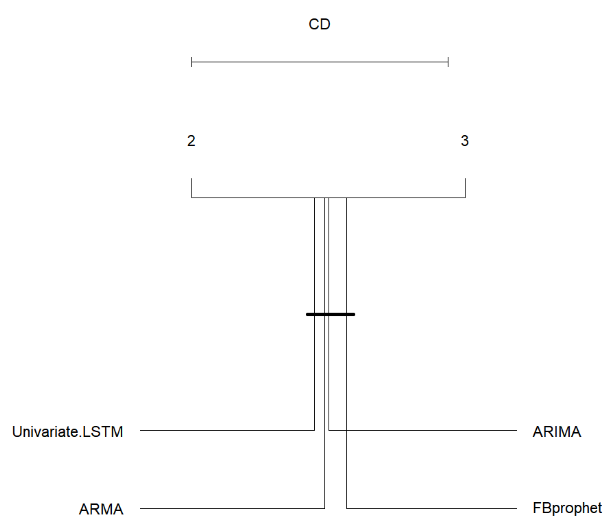
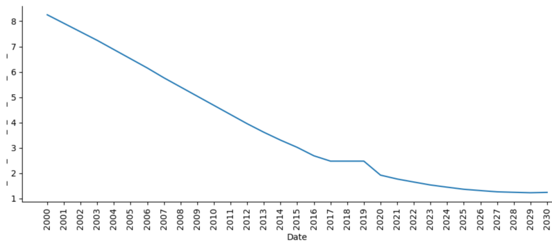


Figure 5.3: A critical difference diagram for Arma, Arima, Fbprophet and Univariate LSTM

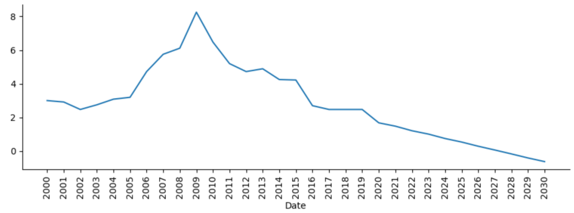
Although from Figure 5.3 no statistical difference between the four techniques considered was observed and from Table 5.2, it can be observed that despite the univariate LSTM producing the best average RMSE result argument can be made for the adoption off FBProphet This is because, FBProphet emerges as a superior choice considering its versatility and performance consistency. With a total of 15018 time series handled, exceeding the numbers of ARIMA (11554), ARMA (14056), and Univariate LSTM (14436), FBProphet demonstrates its adaptability to a wide range of data, including short or complex time series, proving its robustness and reliability. A number of examples where FBprophet has been used to forecast time series values, using the country Afghanistan, are given in Figure 5.4. In the figure, the 20 values for the years from 2000 to 2018 are the known historical values, and the 11 values for the years 2020 to 2030 are the predicted values. As a result of this analysis, FBprophet was selected as the preferred univariate forecasting method and was used where appropriate in comparison to the other forecasting frameworks discussed later in the thesis.

Table 5.2: Evaluation results using four different forecast model generators

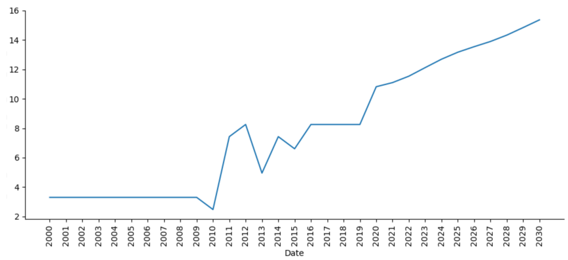
Method	**ARIMA**			**ARMA**			**FBprophet**			**Univariate LSTM**		
Country	e_{RMSE}	e_{MAPE}	COUNT	e_{RMSE}	e_{MAPE}	COUNT	e_{RMSE}	e_{MAPE}	COUNT	e_{RMSE}	e_{MAPE}	COUNT
Alnd islands	15.230	-1414%	2	2.324	-232%	2	1.509	-151%	9	0.504	34%	3
Afghanistan	0.704	-11%	246	0.612	-20%	289	0.781	-34%	334	0.791	-33%	250
Algeria	0.644	11%	267	0.869	-3%	342	1.180	19%	368	0.888	-12%	361
Bangladesh	0.669	13%	307	0.627	-2%	390	0.740	10%	412	0.962	-9%	405
Belize	0.635	-26%	282	0.643	-1%	333	0.666	-23%	348	0.661	-14%	348
Bhutan	0.843	-7%	265	0.868	-14%	328	0.845	14%	342	0.729	-29%	340
Costa Rica	0.723	8%	667	0.795	-20%	790	0.710	-11%	815	0.754	-31%	813
Denmark	0.802	2%	304	0.679	15%	370	0.727	8%	415	0.800	16%	408
Egypt	0.794	4%	313	1.074	-13%	407	1.297	-9%	432	1.286	-33%	429
El Salvador	0.752	-9%	646	0.865	-28%	731	0.707	-9%	752	0.794	-36%	650
Estonia	2.099	155%	302	1.675	114%	352	1.651	124%	394	1.551	110%	390
Faroe Islands	0.872	16%	16	0.625	25%	22	0.381	13%	40	0.550	21%	32
Finland	0.703	-13%	350	0.679	3%	440	0.812	25%	489	0.662	10%	484
Guatemala	0.751	-14%	368	0.745	5%	435	0.863	0%	453	0.550	-9%	453
Honduras	1.292	-6%	566	1.033	-37%	719	1.116	-19%	736	1.042	-49%	733
Iceland	1.165	-17%	289	0.763	20%	360	0.929	-23%	386	0.868	-13%	380
India	0.666	23%	301	0.674	20%	362	0.684	29%	384	0.904	19%	381
Iran	0.830	10%	319	0.494	3%	371	0.634	0%	399	0.491	-13%	394
Ireland	0.832	41%	287	0.823	29%	343	0.807	40%	368	0.824	34%	361
Isle of Man	0.363	-21%	8	0.308	7%	8	0.156	2%	27	0.175	-3%	21
Latvia	1.244	58%	328	0.862	34%	377	0.969	40%	396	0.937	16%	391
Libya	0.748	-21%	170	0.606	19%	242	0.729	20%	264	0.823	9%	256
Lithuania	0.612	-6%	314	0.856	32%	374	0.641	7%	396	0.631	-4%	390
Maldives	1.104	8%	250	0.687	8%	290	0.991	15%	315	0.656	-14%	306
Mexico	0.900	31%	548	0.611	-9%	739	0.702	7%	763	0.729	-23%	661
Morocco	0.638	-2%	351	0.921	21%	418	1.066	2%	446	0.784	-8%	445
Nepal	2.775	174%	317	0.832	44%	387	1.518	105%	409	0.699	8%	404
Nicaragua	1.760	-89%	334	1.936	72%	370	0.905	15%	387	0.622	-11%	383
Norway	0.819	-23%	311	0.772	-3%	383	0.728	5%	447	0.822	1%	437
Pakistan	0.826	0%	367	0.601	8%	418	0.650	-1%	442	0.820	5%	437
Panama	0.711	4%	619	0.733	-1%	730	0.716	7%	748	0.694	-21%	648
Sri Lanka	1.213	-38%	323	0.978	-19%	425	1.058	-33%	442	1.061	-43%	433
Sudan	1.022	-44%	247	1.050	4%	323	1.221	-32%	336	1.190	8%	338
Svalbard and Jan Mayen Islands	0.073	-4%	1	0.010	0%	1	0.008	1%	11	0.418	42%	2
Sweden	1.863	-144%	308	1.595	-87%	397	1.946	-135%	445	1.600	-76%	420
Tunisia	0.640	-8%	337	0.707	19%	392	0.827	3%	413	0.730	7%	410
United Kingdom Northern Ireland	1.318	-80%	315	1.130	-47%	371	1.849	-1%	425	1.237	-40%	413
Western Sahara	10.722	-625%	9	4.148	-251%	25	1.056	6%	30	3.722	-186%	26
Averages	1.588	-54%	304.1	0.953	-7%	369.9	**0.915**	**1%**	**395.2**	0.894	-10%	379.9
Total Number of Time Series	11554			14056			**15018**			14436		



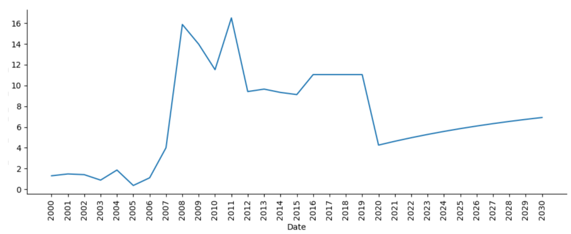
(Under five mortality rate)



(Neonatal mortality rate)



(Domestic coal consumption per capita)



(Health worker density and distribution)

Figure 5.4: A number of examples SDGs Time series with forecast values For Afghanistan generated using FBprophet

5.3.2 Overall Performance

The second objective of the evaluation was to analyse the overall performance of the proposed SDG-AP framework. This was done by inspecting the various Country Tables that were generated. A fraction of the Country Table for Afghanistan, produced using the proposed framework, is given in Table 5.3. Each row represents a time series. The first column gives the GTI for the time series. The “Meta data” column gives the sub-indicator description for the leaf node target time series, the variable D with respect to the adopted time series format given in Chapter 3. The “Initial Value” column gives the current time series value (the year 2015 in the case of the evaluation presented here). The “Predicted” column, as the name suggests, gives the predicted value for the target time series. The “Result” column gives the classification produced using the SDG-AP methodology. Recall that this can be: (i) T indicating that the sub-indicator (represented by the target time series) has been met, or is predicted to be met, on or before the deadline; (ii) F, indicating that the sub-indicator is predicted not to be met by the deadline; or (iii) Unknown, either because the threshold is not clearly defined by the UN or the length of the associated target time series is less than three.

GTI	Meta data	Initial Value	Prediction	Result
1.1.1	BOTHSEX_25+	65.38757	53.44	T
11.5.1	PLGUE	0.92	0.79	F
11.5.1	SUBSD	0.0903	-3.93	T
11.5.1	POLUT	54.76	54.78	F
11.5.1		1.55	1.6	F
11.5.1	VOLER	98	72.99	F
11.5.1		18.4	26.65	F
11.5.2	COLDW	46	32.2	F
11.5.2	CONTM	90	87.74	F
11.5.2	NUCIN	1.91006	1.95	F
4.5.1	GRAD23	0.05427	0.04	T
4.5.1	GRAD23	0.06742	0.07	F
4.5.1	PRIMAR	0.00312	0.04	F
4.5.1	NUME	0.05124	0.04	T
4.5.1	LOWSEC	0.00303	0	T
4.5.1	GRAD23	0.35824	0.86	F
4.5.1	LOWSEC	0.07282	0.07	T
4.5.1	LITE	1.77896	1.8	F
4.5.1	LOWSEC	0.36444	0.35	T
4.5.1	PRIMAR	0.98884	1.04	F
4.5.1	GRAD23	0.06655	0.07	F
4.5.1	GRAD23	0.00401	0.01	F
17.10.1	AGR	2587403	2866154	Unknown

Table 5.3: Fragment of Afghanistan Country Table

Figure 5.5 (a repeat of Figure 4.8) gives an example of the taxonomy visualisation provided using the D3.js Javascript data visualisation package. The example gives part of the populated SDG taxonomy for Afghanistan and SDG 3 (good health and well-being). Green nodes indicate that the indicator will be met (class T), Red nodes that the indicator will be not met (class F) and black nodes that we do not know whether the indicator will be met because of a lack of data (class Unknown). The nodes are labelled with the relevant Goal, Target or Indicator ID, or the relevant descriptor, as appropriate.

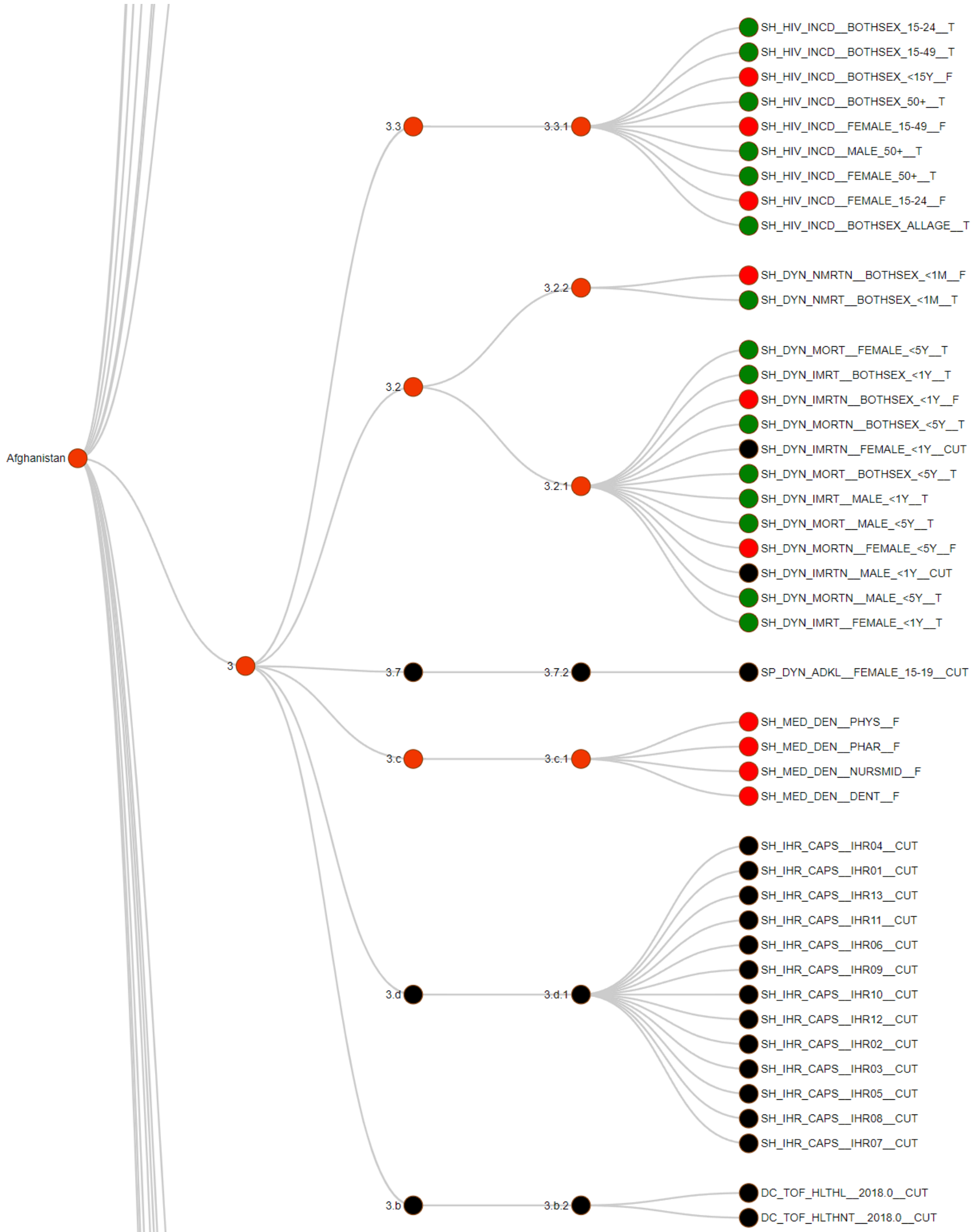


Figure 5.5: Example fragment of the D3.js taxonomy visualisation for the country Afghanistan and SDG 3 (repeat of Figure 4.8)

5.4 Conclusion

This chapter has presented the Sustainable Development Goal Attainment Prediction (SDG-AP) framework for SDG attainment prediction. A number of alternative forecast model generation mechanisms were considered whereby the prediction models to be included at the leaf nodes in the taxonomy could be produced. Out of the four generation models considered, FBprophet was argued to be the most effective. FBprophet was therefore adopted with respect to the univariate elements of the more sophisticated frameworks presented later in this thesis. A criticism of the techniques used is that in cases of percentage forecastings, the forecast should not exceed the range (0-100) which is left as a topic for future work The chapter concluded with an examination of the proposed SDG-AP framework's overall performance. Good results were obtained, however, the proposed SDG-AP framework assumed all indicators (time series) were independent. In the following chapter, the SDG-CAP framework is presented, which considers the potential for intra-causal relationships between indicators (time series) for a given country.

Chapter 6

The Sustainable Development Goal Correlated Attainment Prediction Framework

6.1 Introduction

This chapter presents the second SDG attainment predictions framework considered in this thesis, the Sustainable Development Goal Correlated Attainment Prediction (SDG-CAP) framework. Unlike the SDG-AP framework, discussed in the previous chapter, Chapter 5, which assumed all indicators (time series) were independent, this chapter considers the intra-relationships that might exist between indicators (time series) for a given geographic region (country). Each time series x is considered in relation to all other time series in the set of SDG time series T associated with the same geographic region. The SDG-CAP framework presented in this chapter was therefore proposed as an enhancement to the SDG-AP framework from the previous chapter. The work presented in this chapter also addressed subsidiary research questions S4 and S5 from Chapter 1:

S4. Is it possible to generate more sophisticated machine learning methods to be held at the leaf nodes in the hierarchy by combining data using different instances?

S5. Can some form of feature combination be applied to improve overall forecasting accuracy?

As noted earlier, the SDG-AP framework worked well within the context of the independence assumption of the time series for a given geographic region. However, going back to the discussion in Chapter 1, and Figure 1.1, it can be observed that many SDGs are related. For example, SDG 1 “End poverty in all its forms everywhere”, and SDG 2 “End hunger, achieve food security and improved nutrition and promote sustainable agriculture”, are related as ending hunger will entail eliminating poverty. SDG 3, “Ensure healthy lives and promote well-being for all at all ages” is also clearly related to SDGs 1 and 2. Figure 6.1 shows a number of pairs of time series, identified by their GTIs, for the geographic region Egypt. Examination of the figure indicates clear relationships between the time series. This chapter, therefore, hypothesises that it might be better if the time series data sets used to build the classifiers held at the leaf nodes of the taxonomy, defined in Chapter 3.1, were more holistic. In other words, founded on a set of co-related time series (thus multivariate time series forecasting).

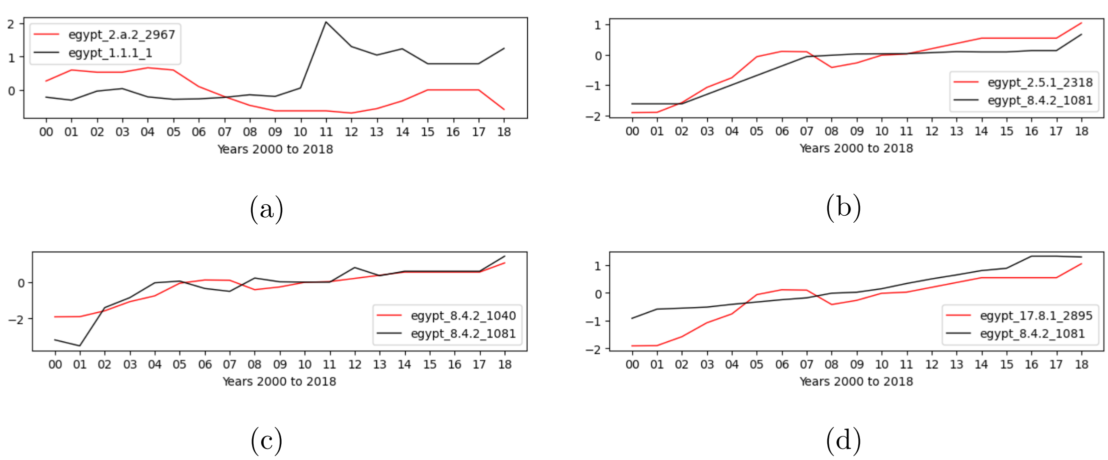


Figure 6.1: Examples of possible relations within SDG time series for the geographic region Egypt

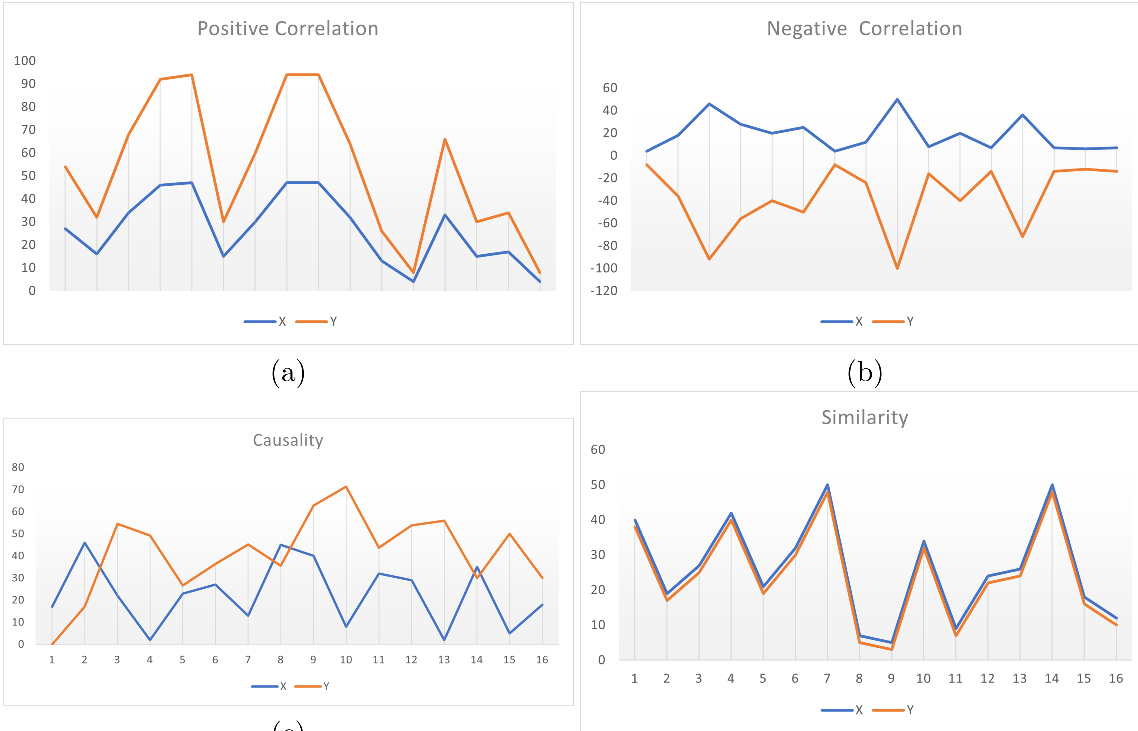
For the work presented in this chapter, a relationship between equally length SDG time series was defined as some identifiable correlation, causation or similarity between the two time series in terms of amplitude and/or magnitude. Six techniques were considered whereby such relationships could be identified (inferred):

1. Pearson Correlation (PC) [13]

2. Least Absolute Shrinkage Selector Operator (LASSO) regression [14].
3. Granger Causality (GC) [15].
4. Mann-Whitney U-test (MW) [16]
5. Dynamic Time Warping (DTW) [17].
6. A bespoke ensemble technique that combined the above five techniques (SDG-ENS).

Numerous alternative methods are proposed in the literature for detecting relationships in time series data, such as cross-correlation and mutual information. However, these techniques present limitations that render them less fitting for this research. Cross-correlation, although valuable under certain circumstances, becomes difficult to interpret when dealing with trending data. specially for short time series that can't be detrended, such as in the case of SDG time series. On the other hand, mutual information, while efficient in identifying a wide array of relationships, is computationally demanding for larger datasets. Consequently, the six techniques listed above were chosen. The selected techniques vary in their operational approaches but all can be used to discern some form of relationship. It's important here to differentiate between correlation, causation, and similarity. In general terms, correlation does not imply causation, although both can coexist. Correlation suggests an observable relationship between two time series x and y, not that a change in x would necessarily cause a corresponding change in y. We can also consider positive and negative correlations. In the latter scenario, the correlation magnitude alters in an inverse manner between time series x and y. Figure 6.2(a) displays a positive correlation and Figure 6.2(b) a negative correlation.

Conversely, causation is where a trend change in a time series x affects the nature of a time series y "causing" y to change in a manner that reflects the change in x. An example of this relation can be seen in Figure 6.2 (c), where the time series y is $n - 1 * 0.5$ where n is a time step in x. In other words, causation requires that the values in x can "explain" the values in y. As in the case of correlation, causation can be positive and negative. Another form of a relation between time series is "Similarity", an example of which is given in Figure 6.2(d). Similarity can be regarded as a special form of correlation (neither positive or negative) With respect to the SDG-CAP framework all three were considered significant. In statistics, the phrase *multicollinearity* is sometimes used to describe the situation where two or more variables are in some way related [92]. The



(a) (b) (c) (d)

Figure 6.2: Examples of Correlation, Causality and Similarity

phrase can equally well be applied to time series. In this chapter, we will use the term relationship as an umbrella term describing causality, correlation and similarity. As noted above, six techniques were experimented with to identify relationships: Pearson, LASSO, Granger, Mann Whitney, DTW and the SDG-ENS ensemble technique. Of these Pearson and LASSO identified correlations, Granger identified causality, and Mann Whitney and DTW similarity. The SDG-ENS ensemble technique was a combination of all five.

Determining the relationship, regardless of the nature of this relationship, between SDG time series is a time-consuming task if done manually and will require a domain expert. For example, considering the geographic region of Egypt, the set of time series T comprises 391 SDG time series. Manual examination of each time series against every other time series would mean $391 \times 390 = 152{,}490$ comparisons. Also, considering pairs of time series only identifies direct, pair wise, relationships where x is related to y; and not indirect relationships where, for example, x is related to y and y is related to z, meaning that

x is indirectly related to z. To consider all possible direct and indirect relationships would feature exponential complexity as the number of time series to be considered increases. For the work presented in this chapter, only pair-wise direct relationships were considered (because of the time complexity of considering non-direct relationships).

In the context of SDG attainment prediction, the challenge of identifying relationships is also hindered by the limited number of observations (points) associated with many of the SDG time series to be considered, making it difficult to establish a conclusive relationship. Recall that the length of the time series considered to establish relationships was 20 points (in many cases, augmentation had been applied to address the missing value problem).

The foregoing was the motivation for the SDG Correlated Attainment Prediction (SDG-CAP) framework presented in this chapter. The challenge, as already alluded to, was how best to identify co-related time series. The rest of this chapter is structured as follows. Section 6.2 provides an overview of the proposed SDG-CAP framework, whilst Section 6.3 presents the evaluation of the proposed framework. The aim of the evaluation was firstly to identify the most appropriate relationship identification technique. The second was to compare the operation of the proposed SDG-CAP framework with that of the SDG-AP framework presented in the previous chapter. The chapter is concluded in Section 6.4 with a summary of the work presented.

6.2 The SDG-CAP Framework

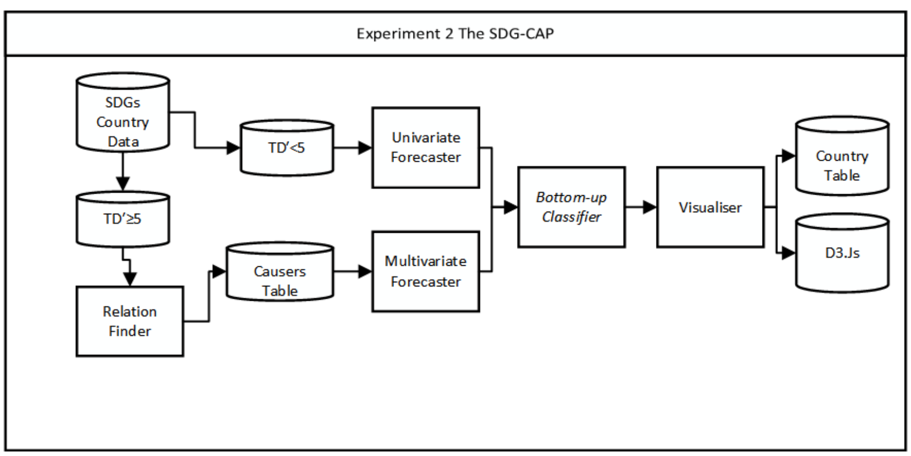


Figure 6.3: The SDG-CAP Framework

Details of the proposed SDG-CAP framework are presented in this section. Figure 6.3 presents a simple schematic of the framework. As in the case of the SDG-AP framework described in the previous chapter, the input is SDG country data generated as described earlier in this thesis in Chapter 3. The country data was divided into two sets, $TD'_{<5}$ and $TD'_{\geq 5}$. The time series in $TD'_{<5}$ are feature less than 5 points and/or feature four or more missing values, whilst the time series in $TD'_{\geq 5}$ feature complete time series, although in some cases missing values may have been imputed. The set $TD'_{<5}$ was handled in the same manner as described for the SDG-AP framework in Chapter 5, where the SDG-AP framework was described, using univariate forecasting. The relationship finder, bottom left of the figure, processes the set $TD'_{\geq 5}$ by comparing each time series $x \in TD'_{\geq 5}$ with every other time series in $TD'_{\geq 5}$. The output was a "Causers Table", a table with the Goal-Target-Indicator (GTI) for the individual time series along the x and y axes, and causality/correlations values at the row and column x and y intersections. As indicated in the introduction to this chapter, six techniques were considered whereby relationships between SDGs time series could be identified. Each of these techniques is discussed in further detail in the following six sub-sections, Sub-sections 6.2.1 to 6.2.6. The discussion regarding each technique is supported by a fragment of a Causer Table generated using the Sri Lanka geographic region.

Once all possible correlations have been discovered the next stage is to build the forecast models for the root nodes in the SDG taxonomy. However, unlike in the case of SDG-AP framework, where uni-variate forecasting was considered, in the case of the SDG-CAP framework, multivariate forecasting was considered for $TD'_{\geq 5}$. Each time series $x \in TD'_{\geq 5}$ will have one or more time series associated with it (identified using the correlation relationship models considered). Where the number of time series was more than one, a multivariate model was used to generate the predictor. A maximum of the five most highly correlated time series was adopted so as to limit the computational complexity. Note that the forecasting complexity increases with the amount of data used. It should also be noted here that the situation will not arise where five time series are not available as it was found that in practice, many time series were associated with each GTI An Encoder-Decoder LSTM was used for the multivariate time series forecasting and FBprophet for the set $TD'_{\leq 5}$ Further discussion concerning the adopted multivariate forecasting is given in Sub-section 6.2.7.

6.2.1 Pearson Correlation

Pearson Correlation [13] was the first of the relationship identification techniques considered with respect to the work presented in this chapter. Pearson Correlation is the most widely used measure to determine whether two variables are correlated in some way, or not, taking into consideration both negative and positive correlation. Pearson measures the extent to which two separate variables co-vary in terms of a linear relationship. A correlation coefficient, r, is determined in the range of -1 to $+1$, where -1 indicates a perfect negative correlation and $+1$ a perfect positive correlation. A value of 0 signifies no correlation. The Pearson correlation coefficient r can be used to measure the association between each pair of SDG time series within the set $TD'_{\geq 5}$ of SDG time series. From an implementational perspective the *scipy.stats* python library [93] was used.

Pearson Correlation values were calculated for each time series pair in the set of time series T. The results were then saved in a "Causer Table". Table 6.1 shows a fragment of the resulting Causer Table for the Sri Lanka geographic region. Inspection of the table indicates that, as to be expected, each time series features a perfect positive correlation with itself ($r = 1.000$). Some strong positive and negative relationships can be identified; r values in excess of 0.900 and -0.900. For example, the time series pair 1.4.1_2273 and 1.4.1_2270 has an r value of 1.000. Time series pairs that feature only very weak relationship can also be identified; r values of less than 0.200. For example, the time series pair 1.5.1_2961 and 1.5.1_3307, and 1.5.1_3307 and 1.5.1_2961, have a r value of -0.236 indicating a weak negative correlation. Note also that the r values are symmetrical about the leading diagonal, this is to be expected, and correlation is a two=way relationship. Using absolute values, the top five most strongly correlated time series were used to formulate multivariate prediction models to be held at the leaf nodes of the topology.

Pearson's Correlation	1.5.1_3307	1.5.1_2961	1.5.1_2957	1.4.1_2273	1.4.1_2271	1.4.1_2270	1.4.1_2269
1.5.1_3307	1.000	-0.236	-0.193	-0.132	-0.229	-0.138	-0.110
1.5.1_2961	-0.236	1.000	0.967	0.564	0.565	0.565	0.560
1.5.1_2957	-0.193	0.967	1.000	0.558	0.539	0.557	0.557
1.4.1_2273	-0.132	0.564	0.558	1.000	0.972	1.000	0.999
1.4.1_2271	-0.229	0.565	0.539	0.972	1.000	0.975	0.959
1.4.1_2270	-0.138	0.565	0.557	1.000	0.975	1.000	0.998
1.4.1_2269	-0.110	0.560	0.557	0.999	0.959	0.998	1.000

Table 6.1: A fragment of the Causer Table holding r values for the Sri Lanka geographic region generated using a Pearson Correlation model

6.2.2 Least Absolute Shrinkage Selector Operator (LASSO) Regression

Least Absolute Shrinkage Selector Operator (LASSO) regression [14] is often used for variable selection to reduce model complexity, but can also be used to explain the relationship between two variables, for example a time series x and a time series y. If x can be used to predict y then a correlated relationship exists. A "coefficient of determination", R^2, is derived to indicate the extent to which the data points in a time series y correlate to data points in a time series x. LASSO regression also uses a penalty constraint to shrink the regression coefficients so that they fall below a predefined threshold. Usually, time series whose coefficients are close to zero are subsequently excluded from the model. Effectively LASSO reduces the coefficient estimates' instability so as to reduce overfitting. L1 regularisation is applied as a shrinkage penalty. A Leave-One-Out (LOO) approach was adopted where a target time series y was removed from the set of time series T, and the rest of the time series used to estimate y using Equation 6.1, where β_0 is a constant, x_i is a predictor time series and y is the target time series (obtained through LOO) and ϵ is an error term. The model is optimised by minimising the mean squared error between the estimated/actual value and the expected value of y using the objective (loss) function given in Equation 6.2, where $(y_i - \hat{y}_i)^2$ is a mean squared error with y_i being the expected value and $\hat{y}_i$ the actual or model estimated value. The end part of the equation is the penalty term, which is the sum of the L1-norms of the coefficient parameters multiplied by a regularising parameter or constraint λ. The parameter β_j is one of the predictor time series. Table6.2 present the generated R^2 values using LASSO Regression experiments, for the geographic region Sri Lanka. Inspection of the table indicates that the leading diagonal features R^2 values of 1.00 indicating that each time series is correlated with itself. From the table we can also see, for example, that the time series pairs 1.5.1_2961 and 1.5.1_3307, and 1.5.1_3307 and 1.5.1_2961, have a R^2 value of 0.056 indicating that a very weak correlated relationship can be identified. Again the R^2 values are symmetric about the leading diagonal, as to be expected. To select the best causers the time series with the highest R^2 values were used.

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \epsilon \tag{6.1}$$

$$L\ C\ F = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{k} |\beta_j| \tag{6.2}$$

LASSO R^2	1.5.1_3307	1.5.1_2961	1.5.1_2957	1.4.1_2273	1.4.1_2271	1.4.1_2270	1.4.1_2269
1.5.1_3307	1.000	0.056	0.037	0.017	0.052	0.019	0.012
1.5.1_2961	0.056	1.000	0.936	0.319	0.319	0.319	0.314
1.5.1_2957	0.037	0.936	1.000	0.311	0.290	0.310	0.311
1.4.1_2273	0.017	0.319	0.311	1.000	0.944	1.000	0.998
1.4.1_2271	0.052	0.319	0.290	0.944	1.000	0.951	0.919
1.4.1_2270	0.019	0.319	0.310	1.000	0.951	1.000	0.996
1.4.1_2269	0.012	0.314	0.311	0.998	0.919	0.996	1.000

Table 6.2: A fragment of the Causer Table showing R^2 values for the Sri Lanka geographic region generated using a LASSO Regression model.

6.2.3 Granger Causality

Granger Causality (GC) is used to determine whether a causal relationship exists between two time series x and y in $TD'_{\geq 5}$ using the concept of prediction. The fundamental idea is that if a lag in time series x can be successfully employed to predict the follow on values in time series y a causal relationship can be said to exist; x is said to "Granger cause" y. The presence, or otherwise, of a causal relationship is determined by calculating a probability value p which is than used to accept or reject the "null hypothesis" that y does not Granger cause x. If $p < 0.05$ the null hypothesis can be rejected and a causal relationship said to exist. The Python *Statsmodels library* [68] was used to implement the Granger model. Part of the Causer Table generated using the geographic region Sri Lanka is presented in Table 6.3. The table shows p values for pairs of SDG indicator time series. Inspection of the table indicates that the leading diagonal features p values of 1.00 indicating that they maximally do not Granger cause themselves . Looking at the pairs of time series indicated by GTIs 1.5.1_2961 and 1.5.1_3307, and those indicated by GTIs 1.5.1_3307 and 1.5.1_2961 (also considered in the previous two Sub-sections), these have p values of $p \geq 0.05$ indicating the acceptance of the null hypothesis that there is no causal relationship in both cases. It is worth mentioning here that a weak negative correlation between these time series was identified using Pearson (Table 6.1), whilst no LASSO relation was identified (Table 6.2); indicating that the various methods considered do not necessarily identify the same pairs of time series. With reference to Table 6.3, it should also be noted that the p values are not symmetric about the leading diagonal. This is because a causality relationship between x and y does not mean that the same causality relationship exists between y and x. In this case, to generate the forecast models to be included within the leaf nodes of the SDG taxonomy, the five time series with the lowest p value were selected, provided that $p \leq 0.05$

Granger Causality	1.5.1_3307	1.5.1_2961	1.5.1_2957	1.4.1_2273	1.4.1_2271	1.4.1_2270	1.4.1_2269
1.5.1_3307	1.000	0.126	0.148	0.400	0.145	0.342	0.528
1.5.1_2961	0.635	1.000	0.005	0.423	0.640	0.409	0.456
1.5.1_2957	0.453	0.005	1.000	0.252	0.546	0.253	0.272
1.4.1_2273	0.391	0.063	0.030	1.000	0.000	0.099	0.037
1.4.1_2271	0.183	0.066	0.038	0.034	1.000	0.016	0.058
1.4.1_2270	0.375	0.062	0.030	0.143	0.000	1.000	0.060
1.4.1_2269	0.448	0.065	0.031	0.017	0.000	0.017	1.000

Table 6.3: Fragment of the Causer Table showing p values for the Sri Lanka geographic region generated using a Granger Causality model

6.2.4 Mann-Whitney U-test

U-tests are statistical tests used to assess whether the mean of two population samples differ from each other or not. U-tests produce a p value, as in the case of Granger Causality (discussed in Sub-section 6.2.3 above). If the calculated p value is less than 0.05 the "null hypothesis", that there is no similarities between the population samples, can be rejected. A normal distribution is usually assumed. U-tests can equally well be applied to time series data. There are a number of variations of U-tests. Parametric U-tests have been shown to be susceptible to Type 1 errors where the null hypothesis is rejected when it should not have been; and Type 2 errors where the null hypothesis is not rejected when it should have been[16]. For the work presented in this chapter, the Mann-Whitney U-tests was adopted; a non-parametric version of the paired U-tests. Using the proposed approach (Figure 6.3) each time series in $TD'_{\geq 5}$ was paired with every other time series in $TD'_{\geq 5}$, and the Mann-Whitney U-tests applied to each pair to determine whether there was a causality relationship in each case. The Python implementation of the Mann-Whitney U-tests, available within the *scipy stats* module [93], was used. If $p \leq 0.05$ the null hypothesis (that time series $x \in TD'_{\geq 5}$ have no similarity with time series $y \in TD'_{\geq 5}$) could be rejected otherwise the hypothesis was accepted. A fragment of the Causer Table generated using the Mann-Whitney U-tests, for the Sri Lanka geographic region, is given in in Table 6.4. From the table it can be seen that the leading diagonal features p values of 1.00 indicating a maximum similarity, this was because the time series in question was compared with itself. Looking, for example, at the 1.5.1_2961 and 1.5.1_3307, and 1.5.1_3307 and 1.5.1_2961 (also considered in the previous Sub-sections), a p value of 0.382, indicating no similarity between the pairs, was recorded. Inspection of Table 6.4 also indicates that the p values are, as to be expected because we are considering similarity, symmetric about the leading diagonal. Again the five time series that featured the highest p value were selected for prediction model generation (described in Sub-section 6.2.7).

U-Test	1.5.1_3307	1.5.1_2961	1.5.1_2957	1.4.1_2273	1.4.1_2271	1.4.1_2270	1.4.1_2269
1.5.1_3307	1.000	0.382	0.258	0.406	0.406	0.406	0.323
1.5.1_2961	0.382	1.000	0.312	0.394	0.312	0.394	0.443
1.5.1_2957	0.258	0.312	1.000	0.228	0.167	0.228	0.312
1.4.1_2273	0.406	0.394	0.228	1.000	0.258	0.443	0.370
1.4.1_2271	0.406	0.312	0.167	0.258	1.000	0.258	0.258
1.4.1_2270	0.406	0.394	0.228	0.443	0.258	1.000	0.358
1.4.1_2269	0.323	0.443	0.312	0.370	0.258	0.358	1.000

Table 6.4: Fragment of the Causer Table showing p values for the Sri Lanka geographic region generated using a Mann-Whitney U-tests model

6.2.5 Dynamic Time Warping

Dynamic Time Warping (DTW) is a measure used to indicate the similarity between two-time series. An alternative might be Euclidean distance measurement. However, unlike Euclidean distance measurement, DTW it is not bound by an equal time series length requirement. DTW can be used to measure the distance between time series of differing lengths, and takes account of miss-alignment between time series. DTW produces a similarity value w called the "warping distance". A w value of zero means that the two-time series under consideration are identical. The value for w has an upper bound that is dependent on the nature of the time series considered, however, the further away from zero the greater the degree of non-similarity. For the purpose of the work presented in this chapter, similarity was assumed to imply a relationship; a special form of correlation. As in the case of the previous measures considered, the w values for all pairs in the set of time series $TD'_{\geq 5}$ were calculated. A fragment of the Causer Table, generated using DTW for the Sri Lanka geographic region, is given in Table 6.5. From the table, it can be observed that the leading diagonal features all zero values because each time series will be identical to itself. For example, looking at the pairs 1.5.1_2961 and 1.5.1_3307, and 1.5.1_3307 and 1.5.1_2961, these have w values of 13.301. Further inspection of the table indicates a symmetry about the leading diagonal (because we are measuring similarity). The five time series with the lowest w values were selected.

DTW	1.5.1_3307	1.5.1_2961	1.5.1_2957	1.4.1_2273	1.4.1_2271	1.4.1_2270	1.4.1_2269
1.5.1_3307	0.000	13.301	12.705	14.267	12.974	14.173	14.598
1.5.1_2961	13.301	0.000	4.864	11.718	12.871	11.752	11.617
1.5.1_2957	12.705	4.864	0.000	10.549	11.547	10.623	10.358
1.4.1_2273	14.267	11.718	10.549	0.000	1.890	0.125	0.464
1.4.1_2271	12.974	12.871	11.547	1.890	0.000	1.767	2.393
1.4.1_2270	14.173	11.752	10.623	0.125	1.767	0.000	0.485
1.4.1_2269	14.598	11.617	10.358	0.464	2.393	0.485	0.000

Table 6.5: Fragment of the Causer Table showing w values for the Sri Lanka geographic region generated using a DTW model

6.2.6 The SDG-ENS Ensemble Approach

Applications of ensemble methods to improve performance have been successful in multiple machine learning and statistical modelling tasks [94]. The intuition behind ensembles is that different predictive models are likely to produce various errors for different input instances. Therefore combining their predictions can lead to a reduction in task-specific errors [95]. In [96] the philosophical underpinning for the ensemble idea was well described by stating that "the ensemble approach leverages the performance of a set of models to achieve better prediction accuracy than that of the individual model or technique".

Ensemble methods are often adopted for supervised classification tasks. Moreover, they are used to mitigate against the class imbalance problem where there is an unequal distribution of instances across the classes in a given data set. There are many variations of ensemble approaches; a frequently adopted approach is bagging [97]. For example in [98] bagging was used to significantly improve the predictive performance on an imbalanced dataset. Ensemble learning has additionally been used to lessen the impact of high dimensionality in the input data; a high number of input features. For example in [99] Attribute Bagging (AB) was used, an approach in which a model is trained using a randomly selected subset of features.

Motivated by the recent success of ensemble approaches, such as those mentioned in [95, 96, 98], an ensemble approach to combine the five causal predictions mechanisms described in the foregoing sub-sections, was also considered. The hypothesis was that this would improve the performance of the overall relationship identification process. A heterogeneous ensemble model was constructed, the SDG-ENS model, similar to that presented in [100], to identify relationships. Heterogeneous because the relationship prediction models utilised were all based on different approaches. The SDG-ENS time series comparison mechanism is one of the contributions of this thesis.

Model	Pearson Correlation	LASSO Regression	Granger Causality	Mann-Whitney U-test	DTW
Metric	Correlation coefficient r	R^2	p-value	p-value	distance w

Table 6.6: The relationship measures used with with respect to SDG-ENS

Generically, the heterogeneous ensemble for the identification of relationships on time series data comprises a set M of j relationship identification models, $M = M_1, \ldots, M_j$. To aggregate information from the various models, a tuple of the form $\langle y, s, c\rangle$ was generated for each time series pair, where: (i) y is the actual relation score computed by a model $M_j \in M = M_1$, (ii) s is an indication of the "strength" of the relationship, and (iii) c is a Boolean value indicating whether a relationship exists or not. The value of the variable s was either bound by the range $[0.00, 1.00]$ or $[0.0500, 0.0001]$ for models which were evaluated by monitoring non p-value metrics and p-value metrics respectively. The value for the variable c was set to 1 indicating that a relationship does exist, or 0 indicating that a relationship does not exist. The mechanism whereby the values for s and c were calculated was dependent on the nature of the relationship calculation as described below. For reference, Table 6.6 summarises the measures computed to establish a relationships between pairs of time series using the different models that make up the ensemble. The calculation was as follows.

Pearson Correlation and LASSO Regression: The Pearson Correlation and LASSO Regression models both produced a value within the range $[0, 1]$; an absolute correlation coefficient value r in the case of Pearson Correlation, and an R^2 value in the case of LASSO Regression. In this case the value for s was calculated using a set of threshold values of the form: $\{0, 0.25, 0.5, 0.75\}$. The threshold values were used to indicate the strength (in increasing order) of the relationship in question (the y value in the $\langle y, s, c\rangle$ tuple). Using these thresholds the values for s and c were assigned using the criterion shown below (Equation 6.3).

$$\langle y, s, c\rangle = \begin{cases} \langle y, 1.00, 1\rangle & \text{if } y \ \textit{or } r/r^2 \geq 0.75 \\ \langle y, 0.75, 1\rangle & \text{if } y \ \textit{or } r/r^2 \geq 0.50 \\ \langle y, 0.50, 1\rangle & \text{if } y \ \textit{or } r/r^2 \geq 0.25 \\ \langle y, 0.25, 0\rangle & \text{otherwise} \end{cases} \tag{6.3}$$

Granger Causality and Mann-Whitney U-test: Both Granger Causality and the Mann-Whitney U-test produced p values, which were used to test the null hypotheses that a relationship did not exist between the pair of time series considered. If the p value was less than 0.05 the null hypotheses that there was no influence between x and y was rejected. A threshold score scale of $\{0.050, 0.01, 0.001, 0.0001\}$ was defined to calculate s and c. Influenced by the ideas presented in [101], the smaller the p value, the more contradictory the data is to the null hypothesis, the greater that the value of s should be. Consequently s and c were calculated as follows (Equation 6.4):

$$\langle y, s, c \rangle = \begin{cases} \langle y, 1.00, 1 \rangle & \text{if } y \;\; or \; p \leq 0.0001 \\ \langle y, 0.75, 1 \rangle & \text{if } y \;\; or \; p \leq 0.001 \\ \langle y, 0.50, 1 \rangle & \text{if } y \;\; or \; p \leq 0.01 \\ \langle y, 0.25, 1 \rangle & \text{if } y \;\; or \; p \leq 0.05 \\ \langle y, 0.00, 0 \rangle & \text{otherwise} \end{cases} \tag{6.4}$$

DTW: In the case of the DTW measure, a "warping distance" was produced of between 0.00 and some maxima dependent on the nature of the time series data set under consideration. A high warping distance equated to low similarity, whereas a low warping distance equated to high similarity. In other words, the closer the warping distance was to 0 the greater the similarity. To determine y, s and c a range of percentiles $\{10, 20, 30, 40, 50, 60, 70, 80, 90\}$ were used as thresholds. The values y s and c were subsequently calculated as shown below (Equation 6.5):

$$\langle y, s, c\rangle = \begin{cases} \langle 1, 1.00, 1\rangle & \text{if } w \geq 0.0 \text{ percentile } \textit{or } w \leq 10 \text{ percentile} \\ \langle 0.9, 0.9, 1\rangle & \text{if } w > 10 \text{ percentile } \textit{or } w \leq 20 \text{ percentile} \\ \langle 0.8, 0.8, 1\rangle & \text{if } w > 20 \text{ percentile } \textit{or } w \leq 30 \text{ percentile} \\ \langle 0.7, 0.7, 1\rangle & \text{if } w > 30 \text{ percentile } \textit{or } w \leq 40 \text{ percentile} \\ \langle 0.6, 0.6, 0\rangle & \text{if } w > 40 \text{ percentile } \textit{or } w \leq 50 \text{ percentile} \\ \langle 0.5, 0.5, 0\rangle & \text{if } w > 50 \text{ percentile } \textit{or } w \leq 60 \text{ percentile} \\ \langle 0.4, 0.4, 0\rangle & \text{if } w > 60 \text{ percentile } \textit{or } w \leq 70 \text{ percentile} \\ \langle 0.3, 0.3, 0\rangle & \text{if } w > 70 \text{ percentile } \textit{or } w \leq 80 \text{ percentile} \\ \langle 0.2, 0.2, 0\rangle & \text{if } w > 80 \text{ percentile } \textit{or } w \leq 90 \text{ percentile} \\ \langle 0.1, 0.0, 0\rangle & \text{otherwise} \end{cases} \tag{6.5}$$

GTI	LASSO			PC			GC			MW			DTW		
	y	s	c	y	s	c	y	s	c	y	s	c	y	s	c
1.1.1_1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	1.000	1.000
1.1.1_2	0.999	1.000	1.000	1.000	1.000	1.000	0.443	0.000	0.000	0.802	0.000	0.000	1.000	1.000	1.000
1.1.1_3	0.999	1.000	1.000	1.000	1.000	1.000	0.481	0.000	0.000	0.418	0.000	0.000	1.000	1.000	1.000
1.1.1_4	0.999	1.000	1.000	1.000	1.000	1.000	0.406	0.000	0.000	0.451	0.000	0.000	1.000	1.000	1.000
1.1.1_5	0.999	1.000	1.000	1.000	1.000	1.000	0.494	0.000	0.000	0.323	0.000	0.000	1.000	1.000	1.000
1.1.1_6	0.998	1.000	1.000	0.999	1.000	1.000	0.431	0.000	0.000	0.683	0.000	0.000	1.000	1.000	1.000
1.1.1_7	1.000	1.000	1.000	1.000	1.000	1.000	0.494	0.000	0.000	0.923	0.000	0.000	1.000	1.000	1.000
1.1.1_8	0.999	1.000	1.000	1.000	1.000	1.000	0.431	0.000	0.000	0.266	0.000	0.000	1.000	1.000	1.000

Table 6.7: Example $\langle y, s, c\rangle$ values for LASSO , Pearson's Correlation (PC), Granger Causality (GC), Man Whitney U-test (MW) and DTW, for the Sri Lanka geographical region

Some example $\langle y, s, c\rangle$ results for a number of SDG time series for the geographic region Sri Lanka are given in Table 6.7. Following the computation of the $\langle y, s, c\rangle$ tuples across all the compared time series pairs, a weight factor λ was computed for each model M_j as an indication of the model's contribution to the relationship identification. For each compared pair of time series, we will have a set of tuples $\{\langle y_1, s_1, c_1\rangle, \ldots, \langle y_1, s_j, c_j\rangle\}$ where j is the number of models considered. The set of tuples for a single model j across all compared pairs of time series was used to compute the weighting λ_j for the model. This value was then used as a scaling factor indicating the overall contribution of model j in establishing the relationship between pairs of time series. The computation of λ_j was formulated as a linear regression problem in the range $[0, 1]$ that attempts to fit a line over the values y, s

and c held in the tuples. In other words, λ_j is computed as an R^2 value. Equation 6.6 gives the linear regression model, where: (i) b_j is a bias term for model j, (ii) y_j is the value produced by the model in question for a given pair, (iii) s_j is the calculated relationship strength for the pair in question, and (iv) w_j^y and w_j^s are the regression coefficients. The regression model is used to predict a value for c_j, the relationship existence indicator. Note that the predicted value will be a real value while the calculated value will be a Boolean value. The λ_j weighting (R^2) value for model $M_j \in M$ will then be calculated using Equation 6.7 where: (i) $|T|$ is the number of time series considered, (ii) c_j^i is the actual value, (iii) $\hat{c}_j^i$ is the predicted value and (iv) $\bar{c}_j^i$ is the mean of the available c_j^i values. In other words, in Equation 6.7 the numerator is the Residual sum of the squared errors and the denominator is the total sum of the squared errors.

$$c_j = b_j + w_j^y y_j + w_j^s s_j \tag{6.6}$$

$$\lambda_j = 1 - \frac{\sum_i^{|T|}(c_j^i - \hat{c}_j^i)^2}{\sum_i^{|T|}(c_j^i - \bar{c}_j^i)^2} \tag{6.7}$$

The λ_j weighting was then used as part of the computation of the relationship score, the SDG-ENS score, between two given time series, a score that combines the prediction scores from the models included in the ensemble. The value for SDG-ENS was bound between 0 and 5, where 0 implies there is no correlated relationship between a pair of time series, and 5 implies a maximal correlation between a pair of time series.

Given a pair of compared time series x and y, and a set of scores $\{y_1, y_2, \dots y_{|M|}\}$ obtained for each model M_j, e is computed as the weighted sum of these scores where each score y_j is multiplied by the corresponding weighting factor λ_j for model j. Equation 6.8 shows the computation of the SDG-ENS value.

$$SDG - ENS = \sum_{j=1}^{j=|M|} (\lambda_j . y_j) \tag{6.8}$$

A fragment of the Causer Table generated using the SDG-ENS model and the Sri Lanka geographic region is given in Table 6.8. The GTIs listed along the y and x axes are the same as those presented with respect to the Causer Tables presented earlier in this chapter for the other relationship identification models considered (see Tables 6.1, 6.2, 6.3, 6.4 and 6.5). From Table 6.8 it can be observed that the leading diagonal features the largest values. This is because in most methods used each time series will related to itself. Looking at the pairs 1.5.1_2961 and 1.5.1_3307, and 1.5.1_3307 and 1.5.1_2961, these have SDG-ENS values of 1.12 and 0.94. Further inspection of the table indicates a non-symmetry about the leading diagonal. This is because the SDG-ENS values include causality, and causality is not a reciprocal relationship. The five time series with the highest SDG-ENS values were selected.

SDG-ENS	1.5.1_3307	1.5.1_2961	1.5.1_2957	1.4.1_2273	1.4.1_2271	1.4.1_2270	1.4.1_2269
1.5.1_3307	1.99	0.94	0.31	1.98	1.30	1.84	0.63
1.5.1_2961	1.12	2.30	0.53	2.13	1.00	1.59	0.96
1.5.1_2957	1.00	0.91	2.00	1.83	0.79	1.41	1.43
1.4.1_2273	1.23	1.19	0.42	2.59	1.11	1.69	0.85
1.4.1_2271	0.70	0.47	0.48	1.05	2.54	2.52	1.65
1.4.1_2270	0.69	0.48	0.49	1.06	2.33	4.07	1.47
1.4.1_2269	0.82	1.01	0.75	0.90	0.51	1.19	5.00

Table 6.8: Fragment of the Causer Table showing SDG-ENS values for the Sri Lanka geographic region generated using the ensemble model

6.2.7 Encoder-Decoder LSTMs

Returning to Figure 6.2, once the Causer Table has been created the next stage is to generate the multivariate forecast models to be held at the leaf nodes of the taxonomy. Any number of forecast model generation mechanisms could have been used. However, a multivariate Encoder-Decoder LSTMs was adopted because of its popularity in the literature in the context of time series prediction [102, 103, 104]. An Encoder-Decoder LSTMs is a type of Recurrent Neural Network (RNN) specifically designed to address sequence-to-sequence prediction problems in various Artificial Intelligence (AI) domains, particularly Natural Language Processing (NLP) [56]. Encoder-decoder LSTM models have also been used in the context of the time series forecasting domain. From a modelling perspective, LSTMs are intuitive because they can encode multiple inputs in the form of a sequence (from multiple variables/features) into a fixed vector representation and also decode this fixed vector representation into another sequence which can be of any length [56]. Furthermore, different from the most commonly attempted forecasting problems, which aim to predict one-time step value at a time, a multi-step forecast problem was framed, in which the forecasting model is tasked to predict multiple steps at once; in other words long-term duration forecasting. The auto regressive nature of the encoder-decoder model allows future predictions of a sequence to be conditioned on previously seen sequence values. Thus the model computes $P(x_i|x_{<i})$, the probability of the current value of a time series x at position i given all previous seen values of x in positions before i [105]. During encoding, at each time step i, a hidden state vector h_i is computed:

$$h_i = LSTM(h_{i-1}, t_i^{(j)}) \tag{6.9}$$

Where: (i) h_{i-1} is a hidden state in a given time series up to step $i-1$, (ii) t_i is the input for the current step and (iii) j is the batch or sample being encoded. For each training iteration, the LSTM encodes an entire batch of dimension $m \times f$. It is important to note that, with respect to the work presented here, the LSTM was not initialised with any pre-trained data.

The decoder receives a context vector as an initial state. It then generates a representation that can be used to predict an output sequence using a linear layer placed on top of the LSTM. The conditional distribution at each step is as follows:

$$P(y_i|y_{<i}, c) \tag{6.10}$$

where future values are decoded auto-regressively. At each future time step i, the probability of output y_i is conditioned on previously decoded values at positions before i, $y_{\leq i}$. In the above equation, c is the contextualised representation of the encoder's entire input sequence.

The training loss objective of the encoder-decoder model is to minimise a mean squared error loss calculated as follows:

$$\|y_i - \hat{y}_i\|^2 \tag{6.11}$$

Where y_i is the expected output at time step i and $\hat{y}_i$ is the actual or predicted output.

6.3 Evaluation

The evaluation of the proposed SDG-CAP framework is presented in this section. It should be noted that the evaluation focused on the multivariate elements of the proposed SDG-CAP framework. The elements directed at the data set $TD'_{\geq 5}$. The univariate forecasting applied to $TD'_{<5}$ was the same as that as used with respect to the SDG-AP framework presented and evaluated in the previous chapter, and is therefore not considered further here. The objectives of the evaluation were:

1. **Relation Inference Model Selection**: To identify the most appropriate relationship inference model out of the six considered.

2. **SDG-CAP versus SDG-AP**: To evaluate the performance of SDG-CAP against the performance of SDG-AP to determine if the conjecture that better SDG prediction attainment can be obtained if correlations between individual SDG time series in a given region are taken into consideration, and to analyse the overall performance of the proposed SDG-CAP framework.

The experiments were conducted twice: (i) using the University of Liverpool Barkla high-performance cluster, and (ii) an 8 RTX3060 GPU machine. Usage of the Barkla system entailed a significant "learning curve", hence the author of this thesis took the opportunity to write "The Barkla 5 Minutes Guide to Setting up a Python 3 Conda Environment", this guide has been included in Appendix B of this thesis. The usage of Barkla proved disappointing, the total run time for an experiment directed at one country was between 15 to 20 hours. The experiments with respect to the best performing relationship models were therefore repeated using the 8 RTX3060 GPU machine, this reduced the run time to less than 3 hours per country. For the SDG-CAP evaluation the same geographic regions (countries) as used in the case of the evaluation of the SDG-AP framework presented in the previous chapter, Chapter 5, were used so that comparisons could be made. For convenience these are listed below:

North Africa: Algeria, Egypt, Libya, Morocco, Sudan, Tunisia and Western Sahara.

South Asia: Afghanistan, Bangladesh, Bhutan, India, Iran, Maldives, Nepal, Pakistan and Sri Lanka.

Northern Europe: Aland Islands, Denmark, Estonia, Faroe Islands, Finland, Iceland, Ireland, Isle of Man, Latvia, Lithuania, Norway, Svalbard and Jan Mayen Islands, Sweden and United Kingdom.

Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, and Panama

As noted above, the adopted forecast model was an Encoder/Decoder LSTM [90]. The input required to train the LSTM model was similar to the LSTM univariate model used with respect to the SDG-AP framework presented in the previous chapter, Chapter 5, but with the setup described in Sub-section 6.2.7. For the evaluation presented here the time series in $TD'_{\geq 5}$ were divided into a training part and a test part. Recall that the time series were short, each comprising of 18 points. So each time series was divided into a 14 point training part and a 4 point test part. Thus the encoder-decoder LSTM models (one per leaf node in the taxonomy) was constructed using 14 point time series and tested by determining how well the model predicted the last 14. Accuracy was determined by calculating the Root Mean Square Error (RMSE). The training set was divided into a set of overlapping input/output *samples*. Each sample was comprised of a sequence of 4 input values, followed by a sequence of three time steps validation output values, expressed as a 3, output matrix where the last column of the matrix represented the targeted time series. Each sample was used as a *prediction step* in the overall LSTM model generation process. Table 6.9 illustrates part of the process, and a complete input-output illustration is given in Appendix C.

Multivariate Enc/Dec data shape												
Steps	Year	Input					Output					
		8.4.2_2664	1.4.1_2271	12.2.2_2754	8.4.2_1071	8.4.2_1086	8.4.2_2664	1.4.1_2271	12.2.2_2754	8.4.2_1071	8.4.2_1086	1.4.1_2270 TARGET
Train 1	2000	-0.828	-0.899	2.457	-0.493	-0.967						
	2001	-0.819	-0.807	2.753	-0.494	-0.967						
	2002	-0.541	-0.715	2.113	-0.494	-1.135						
	2003	-0.871	-0.624	1.041	-0.493	-1.135						
	2004						-0.743	-0.514	0.034	-0.347	-1.135	-0.518
	2005	VALIDATE 1					-0.755	-0.402	0.717	-0.274	0.713	-0.405
	2006						-0.228	-0.289	0.281	-0.167	0.940	-0.290
Train 2	2001	-0.819	-0.807	2.753	-0.494	-0.967						
	2002	-0.541	-0.715	2.113	-0.494	-1.135						
	2003	-0.871	-0.624	1.041	-0.493	-1.135						
	2004	-0.743	-0.514	0.034	-0.347	-1.135						
	2005						-0.755	-0.402	0.717	-0.274	0.713	-0.405
	2006	VALIDATE 2					-0.228	-0.289	0.281	-0.167	0.940	-0.290
	2007						-0.073	-0.175	0.454	-0.009	-0.018	-0.175
	**											
Test	2013	0.327	0.545	-0.350	0.689	0.183						
	2014	0.406	0.670	-0.399	0.760	0.270						
	2015	0.406	0.670	-0.399	0.760	0.270						
	2016	0.406	0.670	-0.399	0.760	0.270						
	2017						0.406	0.670	-0.399	0.760	0.270	0.665
	2018	Validate					0.702	0.969	-1.172	1.030	0.616	0.973
	2019						0.790	1.072	-1.338	1.124	0.694	1.078

Table 6.9: Multivariate forecasting Input-Output sequential

The remainder of this section is organised as follows. The evaluation results are presented in Sub-section 6.3.1. The above two evaluation objectives are then considered, in terms of the results obtained, in Sub-sections 6.3.2 and 6.3.3 respectively.

Geographical Region	**SDG-CAP**						**SDG-AP**
	DTW	**GC**	**PC**	**LASSO**	**MW**	**SDG-ENS**	
Afghanistan	0.298	1.294	1.482	1.484	0.999	**0.285**	0.781
Aland Islands	0.091	0.298	0.197	0.169	**0.058**	0.091	1.509
Algeria	0.983	0.891	0.886	**0.475**	0.877	0.912	1.180
Bangladesh	0.541	1.135	1.089	1.024	1.160	**0.494**	0.740
Belize	1.310	0.761	0.771	0.782	0.724	1.349	**0.666**
Bhutan	0.912	1.582	1.492	1.312	1.413	0.972	**0.845**
Costa Rica	0.934	0.735	0.697	0.709	**0.686**	0.931	0.710
Denmark	2.069	5.357	0.745	5.394	5.396	2.788	**0.727**
Egypt	0.444	1.035	0.938	1.064	0.839	**0.429**	1.297
El Salvador	**0.562**	0.717	0.722	0.722	0.723	0.572	0.707
Estonia	**0.380**	1.971	2.004	0.387	1.992	0.430	1.651
Faroe Islands	0.549	0.729	0.703	0.983	1.133	0.527	**0.381**
Finland	0.864	1.861	1.763	1.990	1.931	0.868	**0.812**
Guatemala	0.938	0.848	0.986	0.909	**0.825**	0.921	0.863
Honduras	0.737	1.295	1.347	0.881	1.484	**0.737**	1.116
Iceland	0.652	0.378	0.364	0.333	**0.324**	0.689	0.929
India	**0.379**	0.732	0.728	0.739	0.768	0.582	0.684
Iran	**0.394**	0.774	0.749	0.785	0.726	0.397	0.634
Ireland	0.548	0.660	0.679	0.613	0.661	**0.547**	0.807
Isle of Man	0.232	0.590	0.475	0.581	0.666	0.256	**0.156**
Latvia	**0.427**	0.530	0.574	0.576	0.573	0.443	0.969
Libia	0.329	1.228	0.600	1.177	1.263	**0.318**	0.729
Lithuania	6.332	**0.490**	0.606	0.611	0.582	3.044	0.641
Maldives	2.930	4.609	3.550	3.983	3.433	2.872	**0.991**
Mexico	**0.383**	1.282	1.229	1.218	1.271	0.392	0.702
Morocco	0.658	1.206	1.216	0.665	1.226	**0.655**	1.066
Nepal	**0.668**	0.985	1.369	1.455	1.445	0.673	1.518
Nicaragua	**0.510**	0.870	0.893	0.872	0.910	0.522	0.905
Norway	0.344	0.201	0.226	0.223	**0.186**	0.339	0.728
Pakistan	**0.298**	0.972	0.938	0.932	0.966	0.504	0.650
Panama	0.891	0.971	0.846	0.984	0.986	0.915	**0.716**
Sri Lanka	0.843	1.226	1.103	1.268	1.141	**0.841**	1.058
Sudan	0.547	0.512	**0.507**	0.548	0.547	0.548	1.221
Svalbard	0.365	*	0.708	0.618	0.578	0.322	**0.008**
Sweden	0.578	0.713	0.605	0.762	**0.521**	0.618	1.946
Tunisia	**0.344**	0.802	0.846	0.874	0.882	0.650	0.827
United Kingdom	0.485	0.802	1.385	1.249	1.301	**0.482**	1.849
Western Sahara	0.416	0.859	0.890	0.944	0.685	**0.407**	1.056
Average	0.820	1.132	0.971	1.060	1.102	**0.772**	0.915
Std	1.038	1.013	0.585	0.955	0.922	0.679	**0.403**

Table 6.10: RMSE Results Obtained per Geographic Region

6.3.1 Evaluation Results

The RMSE results obtained are presented in Table 6.10. The first column gives the country's name. The following six columns give the RMSE results obtained using the six different relationship inference techniques considered coupled with the SDG-CAP framework. The final column gives the RMSE value obtained using the SDG-AP framework as reported in Chapter 5. Best results are highlighted in bod fontIn each case, the average RMSE value obtained, and the Standard Deviation (SD), are given at the bottom of the table for each relation inference technique considered. The table features one missing values: The GC value for the geographic region Svalbard. This was because GC found no relationships (causality/correlation/similarity) due to the limited numbers of time serious available for this particular region.

Time Sereis	DTW	LASSO	PC	MW	GC	SDG-ENS
1.4.1_2270	8.4.2_2664	1.4.1_2269	1.4.1_2269	1.4.1_2269	8.1.1_2639	1.4.1_2269
	1.4.1_2271	12.2.2_1891	17.10.1_3275	12.2.2_2756	3.3.1_433	12.2.2_2756
	12.2.2_2754	17.10.1_3275	17.10.1_3271	17.10.1_3269	12.2.2_1856	17.10.1_3271
	8.4.2_1071	12.2.2_1893	17.10.1_3270	1.4.1_2272	10.b.1_3185	17.10.1_3270
	8.4.2_1086	17.10.1_3271	17.9.1_3254	3.2.1_416	7.1.1_2635	17.9.1_3254

Table 6.11: The five best time series, using the different relationship identification techniques considered, for G.T.I 1.4.1_2270 and the Afghanistan geographical region

DTW	**GC**	**PC**	**LASSO**	**MW**	**SDG-ENS**
13	1	3	1	7	13

Table 6.12: Number of best results obtained using both the SDG-CAP Framework (with each of the six relationship inference techniques considered) for the 38 geographic regions considered

6.3.2 Relationship Inference Model Selection

As noted above, the SDG-CAP framework used multivariate time series forecasting. Each forecast model was trained using the time series in question and the five most closely related time series selected. Table 6.11 shows some example "top five" time series with respect to GTI 1.4.1_2270 and the Afghanistan geographic region. From the table it can be seen that the set of selected time series was not always the same for each model. The purpose of the evaluation discussed in this sub-section was thus to identify the best performing model. For the evaluation the following evaluation metrics and strategies were adopted:

1. The overall average RMSE value obtained with respect to each model.
2. The number of best results produced with respect to each model.
3. The standard deviation of the average RMSE values obtained.
4. A Borda count analysis of the results.
5. A critical difference analysis of the results.

Each is considered in turn as follows.

Best Overall Average RMSE Value. The overall average RMSE values are listed in the penultimate row of Table 6.10. The average RMSE results recorded using DTW, GC, PC, LASSO, MW and SDG-ENS were: 0.820, 1.132, 0.971, 1.060, 1.102 and 0.772 respectively. Inspection of these values indicates that SDG-ENS was the best performing intra-geographic region relationship identification technique out of the six considered, with the DTW technique second. The GC technique produced the worst average RMSE result.

Number of Best Results. The number of best results using each model are given in Table 6.12. Inspection of the table indicates, allowing for ties, that the number of best results produced using DTW, GC, PC, LASSO, MW and SDG-ENS were: 13, 1, 3, 1, 7 and 13, respectively. It could therefore be concluded, considering the number of best results only, that, out of the techniques considered, DTW and SDG-ENS were the best performing techniques. It was interesting to note that the third best performing technique was the Man-Whitney U-test technique. In other words techniques that identified similarities outperformed the techniques founded on correlations (Pearson and LASSO) and those founded on causality (Granger).

Standard Deviation of Overall Average RMSE Values. The standard deviation values associated with the overall average RMSE values obtained are given in the final row of Table 6.10. Standard deviation is a measure of the spread of results, a low standard deviation indicates a low spread and, therefore, a consistency of operation of the technique in question. Inspection of these Standard Deviation values indicates that PC technique was the most consistent (a standard deviation of 0.585), with SDG-ENS second (a standard deviation of 0.765). The worst performing technique was DTW with a standard deviation of 1.038. It is this argued here that the SDG-ENS technique produced a more consistent performance than DTW.

Borda Count. A Borda count is a ranked system where the items of interest are ranked in order of preference. Typically a value of 1 is allocated to the most preferred item, a value of 2 to the second most preferred item, and so on. The final Borda count is calculated as shown in Equation 6.12 where n is the number of rankings. Table 6.13 gives the Borda count rankings for the models considered. The total Borda count is given in the last row of the table. The recorded total Borda counts for DTW, GC, PC, LASSO, MW and SDG-ENS were: 170, 171, 180, 175 and 106, respectively. From the table, it can be seen that the DTW technique achieved the best total Borda count of 103, with SDG-ENS in second place with a Borda count of 106. The other four techniques considered were significantly behind DTW and SDG-ENS.

Critical difference diagram. The critical difference diagram, shown in Figure 6.4, compares the performance of the techniques based on their critical differences. The diagram reveals that there was no statistical difference the in the results obtained

$$\text{Borda Count} = \sum_{i=1}^{i=n} Rank_i \times i \tag{6.12}$$

From Figure 6.4 it can be seen that there was no statistical significance in the operation of the six techniques considered However the case. However it can be made DTW and SDG-ENS both were equal in terms of the number of best results produced. SDG-ENS produced a better overall average RMSE value than DTW and SDG-ENS presented a more consistent mode of operation than DTW (a lower standard deviation) However, DTW present a better BORDA count. It is thus difficult to choose between them. It can be argued, from the results obtained, that many of the relationships discovered using SDG-ENS were the same as the similarity relationships discovered using the DTW technique, because of the similar outcomes obtained. Recall that the DTW technique was founded on DTW similarity measurement that took into consideration offsets between time series. The MW technique also considered similarity, but operated in a different manner; and did not, according to the results presented here, work well. What can be concluded is that techniques that identified

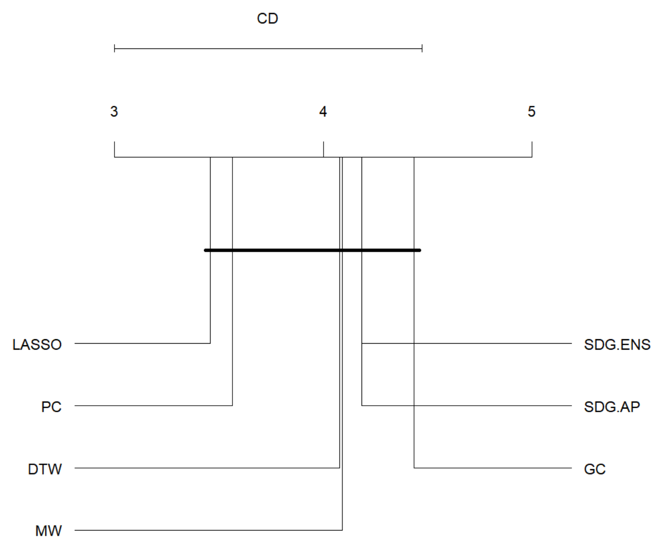


Figure 6.4: The SDG-CAP Critical Difference Diagram

correlations or causality relationships, PC, LASSO and GC, did not work well. It might even be argued that the inclusion of GC, PC and LASSO in the SDG-ENS ensemble might have had a detrimental effect.

6.3.3 Comparison of SDG-CAP and SDG-AP

The second objective of the evaluation was to evaluate the performance of proposed SDG-CAP framework against the performance of SDG-AP "benchmark" framework described in Chapter 5. For the evaluation the FBProphet variation of SDG-AP was considered, because this had been shown to provide best results (see Table 5.2 presented in Chapter 5). For the evaluation, the same criterion were considered as used in the foregoing subsection to identify the most appropriate relationship identification techniques: (i) best overall average RMSE value, (ii) number of best results, (iii) standard deviation of the overall average RMSE, (iv) Borda Count and (v) Critical difference diagram. In terms of average RMSE value, SDG-AP produced a value of 0.902 compared to best average RMSE values of 0.772 and 0.802 using the SDG-ENS and DTW variations of SDG-CAP respectively (see Table 6.10). In terms of the number of best results obtained (Table 6.14) SDG-AP obtained a score of 9 while the SDG-ENS and DTW variations of SDG-

Geographical Region	**SDG-CAP**					
	DTW	**GC**	**PC**	**LASSO**	**MW**	**SDG-ENS**
Afghanistan	2	4	5	6	3	**1**
Aland Islands	3	6	5	4	**1**	2
Algeria	6	4	3	**1**	2	5
Bangladesh	2	5	4	3	6	**1**
Belize	5	2	3	4	1	6
Bhutan	1	6	5	3	4	2
Costa Rica	6	4	2	3	**1**	5
Denmark	2	4	1	5	6	3
Egypt	2	5	4	6	3	**1**
El Salvador	**1**	3	5	4	6	2
Estonia	**1**	4	6	2	5	3
Faroe Islands	2	4	3	5	6	1
Finland	1	4	3	6	5	2
Guatemala	5	2	6	3	**1**	4
Honduras	2	4	5	3	6	**1**
Iceland	5	4	3	2	**1**	6
India	**1**	4	3	5	6	2
Iran	**1**	5	4	6	3	2
Ireland	2	4	6	3	5	**1**
Isle of Man	1	5	3	4	6	2
Latvia	**1**	3	5	6	4	2
Libia	2	5	3	4	6	**1**
Lithuania	6	**1**	3	4	2	5
Maldives	2	6	4	5	3	1
Mexico	**1**	6	4	3	5	2
Morocco	2	4	5	3	6	**1**
Nepal	1	3	4	6	5	2
Nicaragua	1	3	5	4	6	2
Norway	6	2	4	3	**1**	5
Pakistan	**1**	6	4	3	5	2
Panama	2	4	1	5	6	3
Sri Lanka	2	5	3	6	4	**1**
Sudan	3	2	**1**	5	3	5
Svalbard	2		5	4	3	1
Sweden	2	5	3	6	**1**	4
Tunisia	**1**	3	4	5	6	2
United Kingdom	2	3	6	4	5	**1**
Western Sahara	2	4	5	6	3	**1**
Totals	**90**	148	148	160	151	93

Table 6.13: Borda count ranking of the result. Rank rang 1-7 where 1 is the best rank

Ranks	DTW	GC	PC	LASSO	MW	SDG-ENS	SDG-AP
1	**10**	1	1	1	6	**10**	9
2	14	3	3	2	2	13	1
3	7	6	4	5	4	4	9
4	0	4	10	9	5	3	6
5	1	13	10	6	5	3	1
6	3	6	8	11	4	4	1
7	3	4	2	4	12	1	11
Total Votes	**103**	170	171	180	175	106	150

Table 6.14: Summary of the Borda count ranking of the result. Rank range 1-7, where 1 is the best rank

CAP obtained scores of 10. In terms of consistency of performance (standard deviation) SDG-AP produced the best result (see Table 6.10). In terms of Borda count (Table 6.14) and Critical Difference, SDG-AP lagged significantly behind the SDG-ENS and DTW variations of SDG-CAP. Thus it can be concluded that the hypothesis that incorporating intra-country relationships into SDG attainment prediction produces a better prediction accuracy than when such relationships are not included.

6.4 Conclusion

This chapter has presented the Sustainable Development Goal Correlated Attainment Prediction (SDG-CAP) framework for SDG attainment predictions. The central hypothesis considered in this chapter was that better SDG attainment prediction could be realised if intra-geographic region relationships were taken into consideration and used to build multivariate time series forecast models. A range of mechanisms were considered to identify such relationships: Pearson Correlation (PC), LASSO, Granger Causality (GC), the Mann-Whitney U-test (MW) and Dynamic Time Warping (DTW). A bespoke ensemble technique that combined the these five techniques, the SDG-ENS technique, was also considered. The reported evaluation indicated that:

1. Using intra-geographic region relationships did indeed serve to produce better SDG attainment predictions than when such relationships were not considered (as in the case of the SDG-AP framework considered in the previous chapter)
2. The DTW and SDG-ENS variations of SDG-CAP could argued to produced the best results.

3. The SDG-CAP techniques founded on correlation and causality relationships, PC, LASSO and GC, did not work well.
4. It is likely that SDG-ENS utilised the same time series as identified by DTW (although no in depth analysis was conducted).

It was thus concluded that the proposed SDG-CAP framework worked well compared to the SDG-AP framework presented in Chapter 5. However, the proposed SDG-CAP framework assumed all indicators (time series) were not affected by the targeted country's neighbouring countries. Therefore, in the following chapter, the SDG-TTF framework is presented, which considers both inter- and intra- geographic region relationships.

Chapter 7

The Sustainable Development Goal Track Trace and Forecast Framework

7.1 introduction

This chapter presents the third Sustainable Development Goal (SDG) attainment prediction framework considered in this thesis, the Sustainable Development Goal Track Trace and Forecast (SDG-TTF) framework. The hypothesis considered in this chapter is that it is possible to obtain a more accurate SDG attainment prediction by examining the possible relationship between, not only the SDG time series within a given country, but between a given country and its neighbouring countries. In other words both intra- and inter country relationships. The intuition here is that SDG attainment within the countries making up a geographic region will be correlated in some way. Consequently SDG attainment prediction for a given country will be more accurate if the SDG attainment progress in neighbouring countries is taken into consideration. This idea is illustrated in Figure 7.1 which shows four time series for the countries Ethiopia and Egypt with respected to two SDG indicators: (i) Renewable electricity output (% of total electricity output) and (ii) Annual freshwater withdrawals (billions of cubic meters). Initial inspection of the figure does not highlight any particular correlation. However, closer inspection indicates that there is a negative correlation with respect to the renewable electricity output indicator.

As the amount of renewable electricity output increases in Ethiopia, it decreases in Egypt. Similarly, a negative correlation with respect to the annual freshwater withdrawals indicator can be observed. As withdrawals increase in Ethiopia, they decrease in Egypt. The cause for both is a long geopolitical dispute between the two nations over the Nile river water [106]. Note that the renewable energy, in both cases, is renewable hydroelectric energy. Of course, inter-geographic SDG time series relationships do not have to be caused by geopolitical disputes. For example, relationships have been traced to the semiconductor shortages during the COVID-19 pandemic, where many manufacturing domains suffered supply chain issues [107]. Other examples can be found with respect to financial issues, such as the dot-com bubble or the 2008 housing mortgage crisis, both of which have been argued to send SDG attainment into a "chaotic spiral" [108]. Thus, from the foregoing, the SDG-TTF framework uses predictors constructed using both intra- and inter-country relationships. Recall that the SDG-AP framework considered in Chapter 5 assumed that each time series was independent, while the SDG-CAP framework considered in Chapter 6 considered intra-country relationships only.

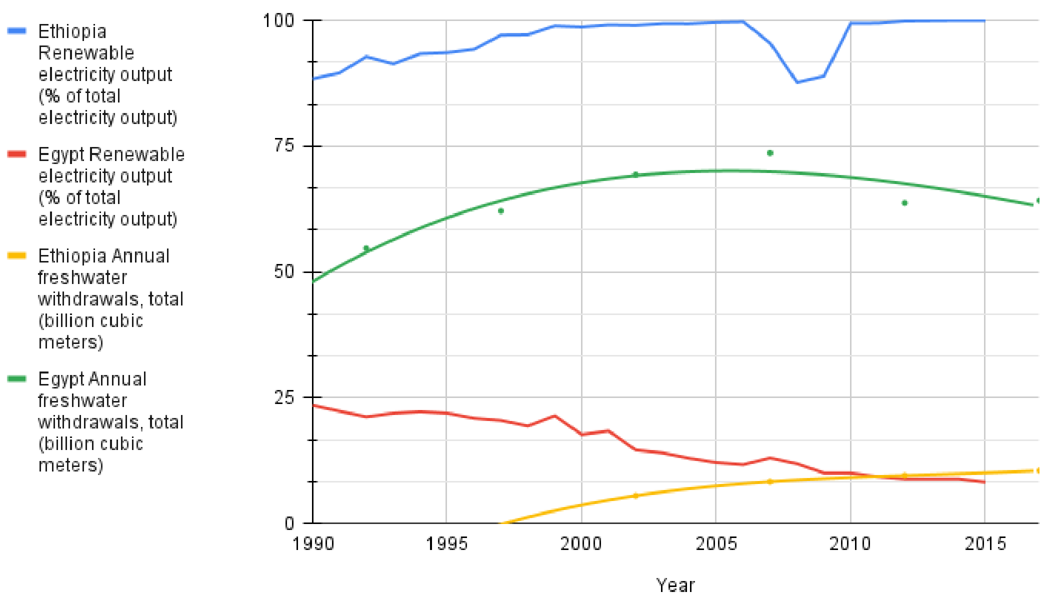


Figure 7.1: Comparison of SDG time series for Ethiopia and Egypt with respect to the "Renewable electricity output" and "Annual freshwater withdrawals, total".

As noted in the previous chapters, the hierarchical taxonomy, on which all the frameworks presented in this thesis are based, requires the generation of predictor (forecaster) models, held at the leaf nodes of the taxonomy, whose outcomes are "passed up" the taxonomy. Recall that the taxonomy was designed to answer questions such as "will a geographical area x meet its goal y by time t". The forecasting approach incorporated into all the frameworks presented in this thesis can thus be categorised as a "bottom-up hierarchical forecasting" approaches. In each case, the predictors associated with the taxonomy's leaf nodes were built using the time series available from the SDG data set maintained by the UN. The remaining nodes in the hierarchy held simple Boolean True/False (T/F) functions which took input from their "child" nodes. However, a single F for any leaf node would be passed up the hierarchy and result in an overall classification of F. It is argued here that the T/F outcome is not very helpful and that a more granular output might be more beneficial. Therefore, with respect to the SDG-TTF framework, the binary classification associated with the previous two frameworks has been replaced with an Attainment Likelihood Index (ALI).

In summary, the work presented in this chapter is directed at a number of the subsidiary research questions presented in Chapter 1, subsidiary research questions S3, S4, S5 and S6:

S3. Assuming a hierarchical taxonomy, how can a prediction label be derived for the root node of the hierarchy?

S4. Is it possible to generate more sophisticated machine learning methods to be held at the leaf nodes in the hierarchy by combining data using different instances?

S5. Can some form of feature combination be applied to improve overall forecasting accuracy?

S6. Can we identify an effective mechanism than a simple Boolean yes/no for establishing when a goal will be reached, if ever?

The chapter is structured as follows. Section 7.2 provides an overview of the proposed SDG-TTF framework, and Section 7.3 presents the evaluation of the proposed framework. The chapter is completed with Section 7.4, which provides a summary of the work presented and the main findings.

7.2 The SDG-TTF Framework

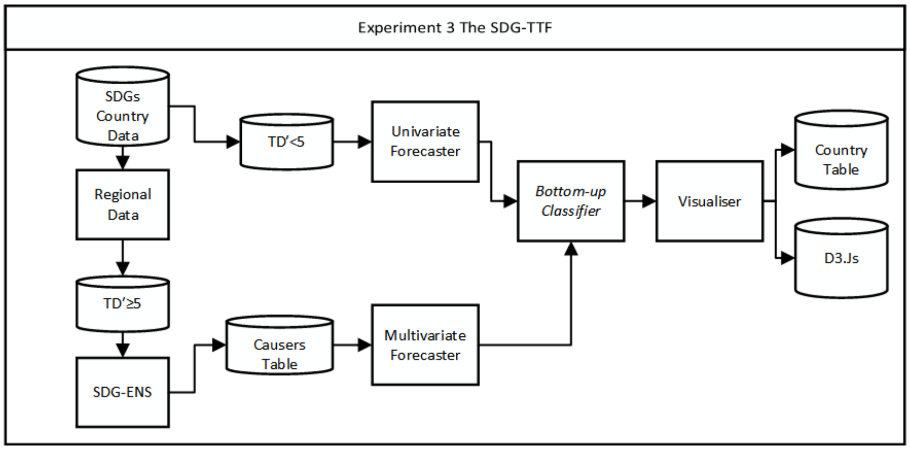


Figure 7.2: The SDG-TTF Framework

Detail concerning the proposed SDG-TTF framework is presented in this section. A schematic of the framework is given in Figure 7.2. The input to the process (top left of the figure), as in the case of the SDG-AP and SDG-CAP frameworks presented earlier, is the SDG country data set $TD = \{TD'_{<5}, TD'_{\geq 5}\}$ generated as described in Chapter 3. Recall that $TD'_{<5} = \{TR'_{<5_1}, TR'_{<5_2}, \dots\}$ where each $TR'_{<5_i}$ is a tuple of the form $\langle Geographical\_Region, GTI, ID, T\rangle$ where T is a time series of less than 5 points and/or featuring four or more missing values. Conversely, $TD'_{\geq 5} = \{TR'_{\geq 5_1}, TR'_{\geq 5_2}, \dots\}$ comprises tuples with complete time series (in some cases time series where missing values have been imputed).

As in the case of the previous two frameworks, $TD'_{<5}$ and $TD'_{\geq 5}$ are processed differently. For $TD'_{<5}$ univariate forecasting was applied at the leaf nodes of the SDG taxonomy, using FBprophet, in the same manner as in the case of the SDG-AP framework described in Chapter 5. With respect to the SDG-TTF framework the set $TD'_{\geq 5}$ was decomposed into a set of regional groupings $G = \{G_1, G_2, \dots\}$. The groupings used were those defined by the UN in [109]: North Africa, South Asia, Northern Europe and Central America. For the evaluation presented later in this chapter $G = \{\text{North Africa}, \text{South Asia}, \text{Northern Europe}, \text{Central America}\}$. For each grouping G, the presence of relationships (causality, correlation and/or similarity) were then identified by comparing each time series within G_i with every other time series in G_i (a computationally expensive task). Mechanisms for identifying relationships were discussed in the previous chapter, Chapter 6, in the context of the SDG-CAP framework. Experiments, also reported in Chapter 6, indicated that the DTW and SDG-ENS variations of SDG-CAP produced the best performance. The SDG-ENS variation produced the best total average RMSE value. Thus, on balance, the SDG-ENS ensemble approach to relationship identification was selected for use with respect to the SDG-TTF framework presented here. Thus, the regional data were examined for relationships occurring across the region using SDG-ENS. The result, as in the case of the SDG-CAP framework presented in the previous chapter, was a "causers table" from which multi-variate forecast models could be generated in the same manner as in the case of the SDG-CAP framework presented in Chapter 6.

Once generated, the models held at the taxonomy leaf nodes can be used to make forecasts for the geographic region in question in a similar manner as described for both the SDG-AP and SDG-CAP frameworks. The leaf node forecasts were compared to the threshold values held in the Data Dictionary generated as described in Chapter 4.1, and the result passed up the taxonomy to achieve a final True/False (T/F) classification stored in a country table. As with respect to the previous frameworks considered in this thesis, the outcome could be inspected using a D3.js visualisation. In the binary classification scenario, as depicted in Figure 7.3, if a single leaf node was predicted as not attaining its goals in time, it would lead to a cascading effect, deeming all the upper levels as unattained.Thus in the example shown in Figure 7.3, the leaf nodes are all determined to be not attained, Even if all were attaining but just a single one was not attained the result of the 7.1.1 node would have been not attained as a consequence of using an & operation It was felt that such a negative outcome would be more in favour if an indication was provided of how close the country in question was to achieving its SDGs. Therefore, the SDG-TTF process is completed, as shown in Figure 7.2. by calculating the ALI value using the data from a country table. This was calculated using Equation 7.1, where $|class = T|$ is the number of indicator time series classified as True and $|class = F|$ is the number of indicator time series classified as False. Thus Equation 7.1 describes a linear relationship. An ALI value of 1 will indicate that all the leaf node indicators in the taxonomy have been met. It is suggested here that the ALI value is particularly useful when a country has failed to meet its SDGs because it indicates how far it was away from meeting its SDGs. Note that the ALI concept can also indicate how far a country is from meeting a particular SDG, Target or Indicator a visualisation of the ALI is shown in 7.4 unlike the binary classification scenario an unattained leaf node will not lead to a domino effect cascading to the upper levels. From the Figure 7.4, unattained leaf node would have a negative impact to the branch, but not a cascading effect to the upper nodes. Thus it will make a better way to visualise and determine which time series is responsible for effecting the attainment status of the goals .

$$ALI = \frac{|class = T|}{|class = T| + |class = F|} \tag{7.1}$$

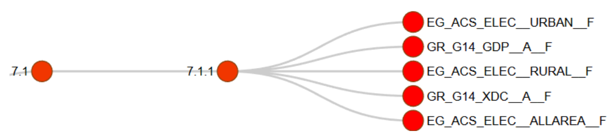


Figure 7.3: simple True/False (T/F) binary classification

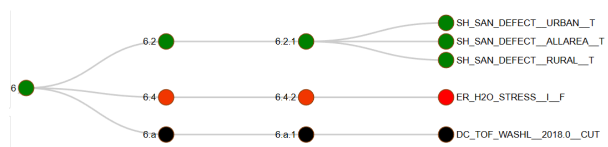


Figure 7.4: ALI classification.

7.3 Evaluation

The evaluation of the proposed SDG-TTF framework is presented in this section. The objectives of the evaluation were:

1. To analysis the nature of the inter-geographic region relationships identified using the proposed SDG-TTF framework.

2. To examine the utility of the proposed ALI concept.

3. To compare the predictive accuracy of the proposed SDG-TTF framework with the SDG-AP and SDG-CAP frameworks presented previously.

As in the case of earlier experiments reported on in this thesis, experiments were conducted using the following countries from the SDG country list for the North Africa, South Asia, Northern Europe and Central America geographic regions:

North Africa: Algeria, Egypt, Libya, Morocco, Sudan, Tunisia and Western Sahara.

South Asia: Afghanistan, Bangladesh, Bhutan, India, Iran, Maldives, Nepal, Pakistan and Sri Lanka.

Northern Europe: Aland Islands, Denmark, Estonia, Faroe Islands, Finland, Iceland, Ireland, Isle of Man, Latvia, Lithuania, Norway, Svalbard and Jan Mayen Islands, Sweden and United Kingdom.

Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, and Panama

Table 7.2 includes a map of the countries and regions considered. The North African region, for example, consisted of 7 countries with a total number of 2,093 different time series. Recall each one of them was compared to every other, in other words, $2{,}093 \times 2{,}092 = 4{,}378{,}556$ comparisons.

Experiments were conducted using a windows machine running Python 3 on a Ryzen 9 processor with 8 cores, 40 Gig of ram and Nvidia RTX2060 GPU was used. Although in the previous chapter, high-performance clusters were used to evaluate the SDG-CAP framework, the environment was changed, and multiprocessing was used for evaluating the SDG-TTF framework. Consequently, instead of considering only the five most related time series in each case, as in the case of the SDG-CAP evaluation considered in the previous chapter, the twenty most related time series were used to build multivariate time series models. A number of twenty was used because this number could still be processed in a reasonable time; the total run time for the experiments was around one week for each region.

The remainder of this section comprises three sub-sections, Sub-sections 7.3.2, 7.3.1 and 7.3.3, each directed respectively at one of the three evaluation objectives listed above.

7.3.1 Inter-region Relationship Identification

The nature of the identified inter-geographic region (country) relationships was analysed by visual inspection. A fraction of the causer table for the country Algeria, produced using the proposed SDG-TTF framework, is given in Table 7.1. Each column lists the top 20 GTIs that are most closely related to the "target" GTI given in the first row; these are the time series to be used for the multivariate forecast model generation for the

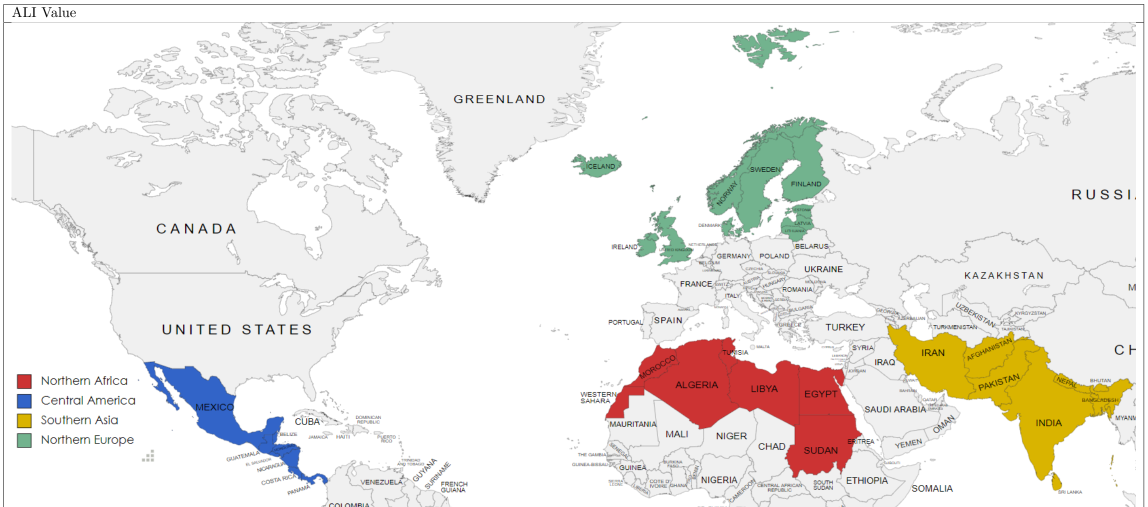
Target Time Series	algeria_17.10.1_3268	algeria_17.10.1_3269	algeria_17.10.1_3270
Top 20 relationships obtained using SDG-TTF	algeria_17.6.1_2894	algeria_17.6.1_2894	algeria_17.6.1_2894
	sudan_11.5.2_3391	algeria_17.10.1_3268	sudan_17.12.1_3252
	sudan_1.5.2_3325	egypt_15.6.1_2822	morocco_17.12.1_3252
	sudan_17.12.1_3252	morocco_7.b.1_3173	sudan_15.1.2_2810
	sudan_11.5.2_3387	morocco_12.a.1_3194	egypt_10.a.1_1306
	sudan_1.5.2_3318	sudan_4.b.1_3085	sudan_9.c.1_2712
	morocco_17.12.1_3252	libya_10.7.3_2739	sudan_5.5.1_935
	libya_10.7.3_2739	morocco_17.6.1_2889	tunisia_9.c.1_2712
	morocco_1.5.1_3315	morocco_17.6.1_2893	morocco_4.4.1_548
	morocco_13.1.1_3401	egypt_8.a.1_3182	algeria_17.6.1_2889
	morocco_11.5.1_3386	egypt_12.a.1_3194	egypt_17.4.1_3246
	algeria_10.7.3_2739	egypt_7.b.1_3173	algeria_17.6.1_2893
	sudan_17.12.1_3259	morocco_9.1.2_2699	libya_3.1.1_2324
	algeria_17.10.1_3269	egypt_10.a.1_1310	egypt_4.3.1_537
	egypt_9.c.1_2712	egypt_3.3.1_432	algeria_3.d.1_512
	egypt_17.6.1_2893	algeria_17.6.1_2888	sudan_5.5.1_937
	egypt_17.6.1_2889	algeria_17.6.1_2892	egypt_4.3.1_539
	egypt_15.6.1_2822	egypt_6.a.1_3165	egypt_4.3.1_538
	sudan_4.b.1_3085	egypt_3.3.1_430	tunisia_2.2_2949
	egypt_7.b.1_3173	morocco_10.7.3_2739	algeria_12.a.1_3194

Table 7.1: Some of the top 20 identified relationships obtained using the SDG-TTF framework for the country Algeria

target GTI (the forecast models to be held at the leaf nodes of the relevant taxonomy tree). For example, the first GTI, Algeria_17.10.1_3268 is "Promote a universal, rules-based, open, non-discriminatory and equitable multilateral trading system under the World Trade Organisation, including through the conclusion of negotiations under its Doha Development Agenda" is strongly related to: GTI Algeria_17.6.1 "Enhance North-South, South-South and triangular regional and international cooperation on and access to science, technology and innovation and enhance knowledge-sharing on mutually agreed terms, including through improved coordination among existing mechanisms, in particular at the United Nations level, and through a global technology facilitation mechanism"; and GTI Sudan_11.5.2 for , "By 2030, significantly reduce the number of deaths and the number of people affected and substantially decrease the direct economic losses relative to global gross domestic product caused by disasters, including water-related disasters, with a focus on protecting the poor and people in vulnerable situations". In other words by enhancing the cooperation between those nations a more beneficial outcome might be achieved. For completeness, Appendix D gives the full Sri Lankan attainment results, and Appendix E the Sri Lankan D3.js visualisation, nodes coloured in green highlight goals, targets, indicators and sub-indicators that will be attained on time. Nodes coloured in red highlight goals, targets, indicators and sub-indicators that will not be attained on time. Nodes coloured in lime highlight goals, targets, and indicators that have a mix of attained and not attained nodes.

7.3.2 Attainment Likelihood Index

ALI Value



Northern Europe		Southern Asia		North Africa		Central America	
Country	ALI Value	Country	ALI Value	Country	ALI Value	Country	ALI Value
Aland Islands	0.50	Afghanistan	0.19	Algeria	0.31	Belize	0.23
Denmark	.050	Bangladesh	0.07	Egypt	0.34	Honduras	0.19
Estonia	0.50	Bhutan	0.43	Libya	0.25	Guatemala	0.20
Faroe Islands	0.0	India	0.75	Morocco	0.36	Nicaragua	0.23
Finland	0.20	Iran	0.07	Sudan	0.31	Mexico	0.22
Isle of Man	0.0	Maldives	0.13	Tunisia	0.30	Panama	0.22
UK & Northern Ireland	0.50	Nepal	0.63	Western Sahara	0.40	Costa Rica	0.23
Iceland	0.50	Pakistan	0.13			El Salvador	0.98
Ireland	0.50	Sri Lanka	0.11				
Latvia	0.50						
Lithuania	0.50						
Norway	0.50						
Svalbard and Jan Mayen Islands	0.50						
Sweden	0.50						

Table 7.2: Visualisation of the ALI values

The second objective of the evaluation presented here was to analyse the ALI concept. Recall that the ALI concept's motivation was to provide a more comprehensive output than that obtained using a simple True/False (T/F) binary classification. The ALI values for the countries considered with respect to the evaluation presented in this section are given in Table 7.2. From the table, it can be seen that ALI values indicate overall progression towards SDG attainment. From the table, El Salvador was ranked the highest, with an ALI value of 0.98; whilst many countries in the North Europe Region had values of 0.50. North European countries have a lower ALI because they started from a much higher level, while El Salvador started from a much lower level. With respect to many SDG targets, it is more difficult for a developed country to reach the target than an underdeveloped country. This is illustrated by considering GTI "17.6.1" and population per country in Figure 7.5. The figure compares the number of internet subscribers as a proportion of the population. A country like the United Kingdom begins with a far higher percentage of individuals already subscribed, far more than a country like (say) Afghanistan. The threshold for this specific example, as specified in Chapter 4, is if the number of internet subscribers (in 2030) is $\geq 60\%$ of the value recorded in 2015. Thus it is easier for a country like Afghanistan to attain this target than it is for (say) the United Kingdom (in this case, the United Kingdom may not be able to attain this particular SDG target). There are numerous other reasons why it may be difficult for a developed country to reach its SDG goals, while an underdeveloped country may be able to do so.

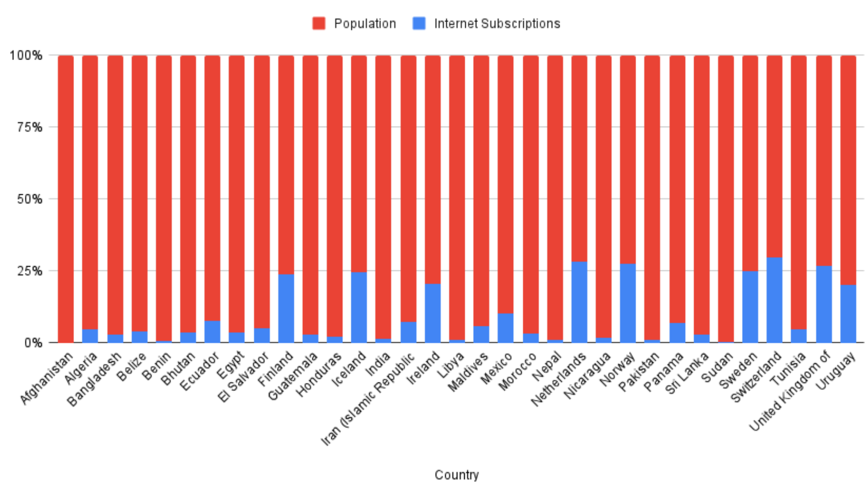


Figure 7.5: Internet subscribers in comparison to population

7.3.3 Comparative Effectiveness of the SDG-TTF Framework

As in the case of the previous frameworks considered, the operation of the SDG-TTF framework was evaluated by using it to predict the last n=4 values in SDG time series. Table 7.3 gives the RMSE values obtained for the 38 countries considered. The countries are listed in column one. The RMSE values obtained using the proposed SDG-TTF framework are listed in column 4. The table also includes the best RMSE results obtained using SDG-AP framework (the FBProfit variation) and the SDG-CAP framework (the SDG-ENS variation) in columns 2 and 3 (as reported in the previous two chapters). The average $RMSE$ value for each framework, and the associated standard deviation, are given in the last two rows. From the table, it can be seen that SDG-TTF produced the best overall performance, thus indicating that the hypothesis that taking both intra- and inter-country relationships into account would provide for a better SDG attainment prediction accuracy than when only intra-country relationships were considered (SDG-CAP) or when all time series were considered to be independent (SDG-AP). To ensure the accuracy of the claim a Borda count evaluation and a critical difference diagram (as also adopted with respect to the evaluation of the SDG-CAP framework presented in Chapter 6) was produced to validate the results. The Borda count rankings are presented in Table 7.4; a ranking of 1 represents the best result, a ranking of 2 is the second best result, and so on. The Borda count, in this case, was calculated as shown in Equation 7.2. The Borda count scores are provided in the final column of Table 7.4. The table indicates that SDG-TTF achieved a higher Borda count score of 60, outperforming both SDG-AP and SDG-CAP, which obtained counts of 93 and 75, respectively. The critical difference diagram is presented in Figure 7.6. From the diagram it can be seen that there was no statistical significance in operation between the three systems considered. Although the Borda count results indicates the SDG-TTF produced a better performance

$$\text{Borda Count} = \sum_{i=1}^{i=3} Rank_i \times i \tag{7.2}$$

7.4 Conclusion

This chapter has presented the Sustainable Development Goal Track Trace and Forecast (SDG-TTF) framework for SDG attainment prediction. The third and final framework

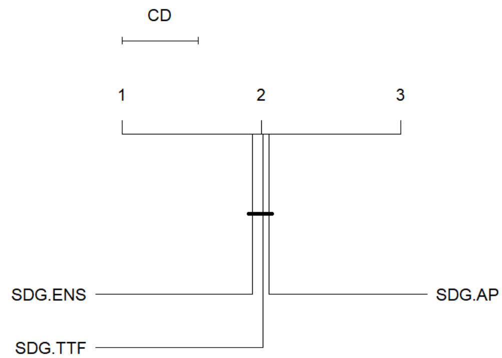


Figure 7.6: Critical Difference Diagram for SDG-TTF.

considered in this thesis. Unlike the previous framework, the SDG-CAP framework, where only intra-country relationships were considered, relationships limited to a single geographical area, the SDG-TTF also considered inter-country relationships, relationships between countries in the same geographic region (as defined by the UN). The presented evaluation indicated that by considering both intra- and inter country relationships better SDG attainment prediction could be obtained. Furthermore, the concept of an Attainment Likelihood Index (ALI) was considered. Analysis of the ALI values obtained indicated that the concept could be used to rank countries based on their SDG attainment likelihood. It was suggested that this was a more informative value than the simple True/False (T/F) binary classification produced using the SDG-AP and SDG-CAP frameworks. The next chapter presents the conclusion of this thesis where the main findings are presented together with some suggestions for future work.

Geographical Region	**SDG-AP**	**SDG-ENS**	**SDG-TTF**
Afghanistan	0.781	**0.285**	1.305
Aland Islands	1.509	**0.091**	0.187
Algeria	1.180	0.912	**0.300**
Belize	**0.666**	1.349	0.748
Bhutan	**0.845**	0.972	1.481
Costa Rica	**0.710**	0.931	**0.710**
Denmark	0.727	2.788	**0.637**
Egypt	1.297	0.429	**0.400**
El Salvador	0.707	**0.572**	0.701
Estonia	1.651	**0.430**	0.644
Faroe Islands	**0.381**	0.527	0.647
Finland	0.812	0.868	**0.644**
Guatemala	0.863	0.921	**0.835**
Honduras	1.116	**0.737**	0.743
Iceland	0.929	0.689	**0.644**
India	0.684	**0.582**	1.480
Iran	0.634	**0.397**	1.321
Ireland	0.807	**0.547**	0.608
Isle of Man	**0.156**	0.256	0.607
Latvia	0.969	0.443	**0.348**
Libya	0.729	**0.318**	0.800
Lithuania	0.641	3.044	**0.348**
Maldives	**0.991**	2.872	1.298
Mexico	0.702	**0.392**	0.821
Morocco	1.066	0.655	**0.600**
Nepal	1.518	0.673	**0.658**
Nicaragua	0.905	**0.522**	0.740
Norway	0.728	**0.339**	1.675
Pakistan	0.650	0.504	**0.490**
Panama	**0.716**	0.915	0.894
Sri Lanka	1.058	0.841	**0.436**
Sudan	1.221	0.548	**0.200**
Svalbard	**0.008**	0.322	0.401
Sweden	1.946	0.618	**0.313**
Tunisia	0.827	0.650	**0.400**
United Kingdom	1.849	0.482	**0.336**
Western Sahara	1.056	**0.407**	0.500
Average	0.915	0.779	**0.700**
Standard Deviation	0.403	0.686	**0.374**

Table 7.3: Comparison of RMSE results obtained using the SDG-AP, SDG-CAP and SDG-TTF frameworks and 38 countries

Framework	Rank 1	Rank 2	Rank 3	Borda Count
SDG-AP	8	10	17	93
SDG-CAP	13	15	10	75
SDG-TTF	17	9	9	60

Table 7.4: Number of best results obtained using SDG-AP , SDG-CAP and SDG-TTF Framework for the 38 geographic regions considered, and the associated Borda count scores

Chapter 8

Conclusion

8.1 introduction

This chapter presents the conclusion to the research presented in this thesis. The chapter is structured as follows. In section 8.2 a thesis summary is provided. The main findings and contributions with respect to the main research question and the subsidiary research questions are then considered in Section 8.3. Finally, some suggestions for potential future work related to the work presented in this thesis are given in Section 8.4.

8.2 Summary

This thesis commenced with an introductory chapter, Chapter 1. This highlighted the motivation for the work presented in the thesis, namely using machine learning tools and techniques to forecast SDG attainment using a Bottom-Up hierarchical approach to predict if a geographical region would reach its goals on time, if at all. Then in Chapter 2, previous work-related to the research reported in the thesis was presented.

The following two chapters, Chapter 3 and 4, considered the SDG data available and the idea of using an active taxonomy to facilitate SDG attainment prediction. Chapter 3 reviewed the SDG data that has been made publicly available by the UN, and the adopted pre-processing of the data so that it can be used for machine learning and especially forecasting. Chapter 4 focused on the construction of a bottom-up hierarchical model, a taxonomy, where the leaf nodes would hold forecasting models, the result of which would be passed up the hierarchy until a final SDG attainment prediction classification could be obtained at the hierarchy root node.

The following three chapters, Chapters 5, 6 and 7, considered the three SDG attainment prediction frameworks proposed in this thesis. Chapter 5 presented the SDG-AP framework. This was a baseline framework that assumed that every SDG time series was independent of any other time series. Several methods were explored to create the desired baseline forecasting models, of which FBprophet proved to be the most accurate. Chapter 6 presented the second framework, the SDG-CAP framework, which considered the possibility of intra-geographic region relationships. A number of different approaches were experimented with whereby relationships could be identified, of which two approaches, the DTW approach and the SDG-ENS ensemble approach, proved to be effective. Of these the SDG-ENS approach produced a better overall average RMSE and was thus argued to be the most effective. Chapter 7 presented the third framework considered in this thesis, the SDG-TTF framework, which considers the possibility of both intra- and inter-geographic regions relationships. The SDG-TTF framework was found to be the most effective of the three proposed frameworks.

8.3 Main Findings

This section provides the main findings of the work presented in this thesis in terms of the overriding research question and the subsidiary research questions. The section commences by considering the subsidiary research questions and then goes on to consider the main research question.

1. [S1.] **What is the most efficient and effective way to derive a taxonomy for SDG data?** Chapter 3 focused on answering this question. It was proposed that the most straightforward manner in which an SDG taxonomy could be created, and consequently the most efficient and effective manner, was to organically grow the hierarchy. An "unpacking" process was proposed whereby leaf nodes in the hierarchical taxonomy represented either indicators, sub-indicators or sub-sub-indicators, which were then "unpacked" to reveal upper levels targets and entire goals. The taxonomy concept was successfully employed with respect to all three frameworks considered in this thesis.

2. [S2.] **How best can machine learning be used to forecast whether individual SDGs will be met?** The fundamental idea presented in this thesis was that machine

learning could best be used to build time series forecast models to be held at the leaf nodes of the taxonomy. A variety of techniques were considered for building the desired forecasting models. Some are directed at univariate forecasting, and some at multivariate forecasting. For univariate forecasting, FBprophet was found to be the most effective when incorporated into the SDG-AP framework an average RMSE value of 0.907 was recorded (see Sub-section 5.3.2 in Chapter 5). An Encoder-Decoder Long Short Term Memory (LSTM) Recurrent Neural Network (RNN) was used for multivariate forecasting. This was used with respect to both the SDG-CAP and SDG-TTF frameworks. Best average RMSE values of 0.779 and 0.700 were recorded using the SDG-CAP and SDG-TTF frameworks, respectively (see Sub-section 6.3.3 in Chapter 6, and Sub-section 7.3.3 in Chapter 7).

3. [S3.] **Assuming a hierarchical taxonomy, how can a prediction label be derived for the root node of the hierarchy?** The proposed mechanism for establishing a prediction label was by using the bottom-up approach facilitated by the taxonomy (this was partly the reason for choosing a taxonomy-based framework to address the SDG attainment prediction problem in the first place). Two approaches were considered for establishing a prediction label at the root node of a country hierarchy: (i) a True/False (T/F) binary classification approach and (ii) an Attainment Likelihood Index (ALI) approach. Values were forecasted at the leaf nodes, which were then converted into a binary True/False (T/F) by comparing with threshold values, and the results passed up the hierarchy. At each parent node a "logical and" operator was applied until the root of the tree was arrived at. The number of T and F values were then used to calculate an ALI value.

4. [S4.] **Is it possible to generate more sophisticated machine learning methods to be held at the leaf nodes in the hierarchy by combining data using different instances?** In Chapter 5, different Univariate forecasting models to be used in the leaf nodes were considered. In Chapters, 6, and 7, a more sophisticated model, an LSTM, was adopted for multivariate forecasting, illustrating that more sophisticated machine learning methods can be used for generating forecast models to be held at the leaf nodes in the taxonomy.

5. [S5.] **Can some form of feature combination be applied to improve overall forecasting accuracy?** Feature combination was considered in terms of intra- and

inter-country relationships (correlations, causalities and/or similarities). The first was presented in Chapter 6, where six mechanisms for identifying relationships were considered with respect to the SDG-CAP framework. Of these, two approaches, the DTW approach and the SDG-ENS ensemble approach, were found to work well. The SDG-ENS approach produced the least average total RMSE, and was thus argued to be the best performing approach. The second was presented in Chapter 6 where the SDG-ENS approach was used again (because it had been found to be the most effective approach for intra-relationship discovery) with respect to the SDG-TTF framework. The evaluation indicated that the SDG-CAP framework outperformed the SDG-AP framework, which was attributed to intra-country relationships. Similarly, the SDG-TTF framework outperformed the SDG-CAP framework, which was attributed to the addition of inter-country relationships.

6. **[S6.] Can we identify an effective mechanism than a simple Boolean yes/no for establishing when a goal will be reached, if ever?** The United Nations, in their SDG documentation, has published broad descriptions of how SDGs should be met that can be interpreted in many different ways. Words such as "End", "Eliminate" and "Insure" was used, as was discussed in Chapter 4. A process was proposed, founded on the idea of templates, whereby these vaguely described thresholds could be converted into numerical representations. Thus when an indicator value was forecast, this could be compared with the numerical representation of the thresholds and used to determine a Boolean classification. The concept of an ALI value was also proposed to encapsulate the likelihood of attainment. This proved to be an effective mechanism for establishing SDG attainment with respect to a given geographic region (typically a country).

Returning to the overriding research question that this thesis sought to answer:

How can the tools and techniques of machine learning be harnessed to effectively and efficiently conduct attainment prediction in the context of the UN Sustainable Development Goals?

The central idea presented in this thesis was the idea of an SDG topology whose leaf nodes could be populated with forecast models, which could then be used to make forecasts that were passed up the topology. The phrase "bottom-up classification" was used to describe this process. The idea was incorporated into three frameworks of increasing sophistication: SDG-AP, SDG-CAP and SDG-TTF. The SDG-AP framework assumed all indicators were independent, the SDG-CAP framework took into account the potential for intra-country relationships between indicators, and the SDG-TTF the potential for both intra- and inter-country relationships. A best RMSE value of 0.700 was obtained when using the SDG-TTF framework, compared to best values of 0.915 and 0.779 when using the SDG-AP and SDG-CAP frameworks. It was thus concluded that the tools and techniques of machine learning could best be harnessed to effectively and efficiently conduct SDG attainment by using the proposed taxonomy to enable bottom-up classification using forecasting models that took into account both intra- and inter-country relationships.

8.4 Future Works

The work presented in this thesis has proposed three frameworks for SDG attainment prediction founded on the idea of an SDG hierarchical taxonomy and a bottom-up classification process. In this final section, a number of future research directions are suggested as follows:

1. **Data Shortage** As noted in Chapter 3, SDG time series are short (at the time of writing, the maximum was 20 points). Some SDGs, for particular countries, have longer time series associated with them, but this was because of some country-specific initiatives that predated the SDG initiative. In addition, many geographical areas lack proper data collection agencies. Thus, there is a shortage of data to train SDG attainment prediction models. One suggested area of future work is this to increase the amount of data available by using external data sources and then expanding the current data. Possible external data sources include aerial or satellite imagery.

2. **Taxonomy Thresholds** From the work presented in Chapter 4, the total number of different thresholds for the different indicators is 3819. As also noted in Chapter 4 the nature of the individual thresholds were all manually identified using a set of templates; a more automated approach would be beneficial. One solution might be to use Natural Language Processing (NLP) techniques to process the UN SDG documentation, with respect to the suggested templates, so that thresholds can be identified in an automated manner. This is then the second area for possible future work suggested here.

3. **Relationship Identification and in-Depth analysis** In Chapters 7 and 7, DTW and SDG-ENS were found to produce the best results a further in-depth analysis between the two techniques should be considered for future work. Another suggested research area could be extending the SDG-TTF framework to accommodate larger geographical groupings, it was assumed that relationships between time series would only exist within a given region. Some evidence was provided to support this assumption. However, it may be the case that relationships exist beyond a country's geographic region. Investigating this potential may provide a fruitful direction for future research, although this would incur increased computational complexity.

4. **Forecasting** It was noted in Chapters 5 and 7that where percentages were considered, some of the forecast models predicted percentage values that exceeded 100% or went below 0%. This is clearly incorrect, and a method needs to be derived to prevent this. It would also be worthwhile to investigate data leakage from the preprocessing discussed in Chapter 3. Lastly forecasting models than the LSTM model used with respect to the work presented in this thesis. For example, more enhanced multivariate neural network techniques such as those presented in [110].

5. **Visualisation** The D3.Js is a Javascript package that was used to visualise the output from the proposed frameworks. However, it was found that D3.Js was difficult to use. Consequently, the visualisations from work presented in this thesis lacked any interactivity. Investigating a more user-friendly tool that supports interactive data visualisation would, therefore, also be beneficial, such as HEAVY.AI or Tableau.

6. **Significance Testing** A number of univariate time series forecasting methods (ARMA, ARIMA, FBprophet and LSTM) were considered in this thesis, and FBprophet was identified as the most appropriate. Similarly, a number of relationship identification approaches were considered (Pearson Correlation, LASSO, Granger Causality, Mann-Whitney, DTW and SDG-ENS), and SDG-ENS was argued to be the most appropriate, although the proposed DTW approach also worked well. However, no statistical significance testing was undertaken (due to time constraints), this is thus suggested as a final item of future work in the list of items provided here.

Bibliography

[1] Johan Rockström and Sukhdev Pavan. A new way of viewing the Sustainable Development Goals and how they are all linked to food, 2016.

[2] Rafael Lozano and Christopher J.L. Fullman. Measuring progress from 1990 to 2017 and projecting attainment to 2030 of the health-related Sustainable Development Goals for 195 countries and territories: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 392(10159):2091–2138, 2018.

[3] United Nations. The Millennium Development Goals Report. *United Nations*, page 72, 2015.

[4] Gerald G Singh, Andrés M Cisneros-Montemayor, Wilf Swartz, William Cheung, J Adam Guy, Tiff-Annie Kenny, Chris J McOwen, Rebecca Asch, Jan Laurens Geffert, Colette CC Wabnitz, et al. A rapid assessment of co-benefits and trade-offs among sustainable development goals. *Marine Policy*, 93:223–231, 2018.

[5] Paul Pacheco (UNSD) and Ze Yar Min. E-Handbook on Sustainable Development Goals Indicators, 2021.

[6] Gyula Dörgo, Viktor Sebestyén, and János Abonyi. Evaluating the interconnectedness of the sustainable development goals based on the causality analysis of sustainability indicators. *Sustainability (Switzerland)*, 10(10):3766, 2018.

[7] P. J. Brockwell. Time series analysis. In *International Encyclopedia of Education*, pages 474–481. 2010.

[8] Rob J. Hyndman, Roman A. Ahmed, George Athanasopoulos, and Han Lin Shang. Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, 55(9):2579–2589, 2011.

[9] Daniel F Heitjan and Srabashi Basu. Distinguishing "missing at random" and "missing completely at random". *The American Statistician*, 50(3):207–213, 1996.

[10] Leonid Shumilo, Andrii Kolotii, Mykola Lavreniuk, and Bohdan Yailymov. Use of land cover maps as indicators for achieving sustainable development goals. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 2018-July, pages 830–833, jul 2018.

[11] Carlos N. Silla and Alex A. Freitas. A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, 22(1-2):31–72, 2011.

[12] Chao Zhang, Fangbo Tao, Xiusi Chen, Jiaming Shen, Meng Jiang, Brian Sadler, Michelle Vanni, and Jiawei Han. TaxoGen: Unsupervised topic taxonomy construction by adaptive term embedding and clustering. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '18, pages 2701–2709, New York, NY, USA, 2018. ACM.

[13] Bruce B. Frey. Pearson Correlation Coefficient. In *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*, pages 1–4. Springer, 2018.

[14] Robert Tibshirani. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[15] C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424, 1969.

[16] Noor Alam and Cynthia Rudin. Robust Nonparametric Testing for Causal Inference in Observational Studies. *Optimization Online, Dec*, pages 1–39, 2015.

[17] Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

[18] Yassir Alharbi, Daniel Arribas-Be, and Frans Coenen. Sustainable development goal attainment prediction: A Hierarchical Framework using Time Series Modelling. *IC3K 2019 - Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 1:297–304, 2019.

[19] Yassir Alharbi, Frans Coenen, and Daniel Arribas-Bel. Sustainable development goal relational modelling: Introducing the sdg-cap methodology. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 12393 LNCS, pages 183–196, 2020.

[20] Yassir Alharbi, Daniel Arribas-Bel, and Frans Coenen. Sustainable development goals monitoring and forecasting using time series analysis. 2021.

[21] Yassir Alharbi, Daniel Aribas-Bel, and Frans Coenen. Sustainable development goal relational modelling and prediction: Introducing the sdg-cap-ext methodology. *Journal of Data Intelligence*, 2(3):348–367, 2021.

[22] Ruben De, Tom Schaul, and Claudia Clopath. AI for social good: unlocking the opportunity for positive impact. *Nature Communications*, 11(1):1–6, 2020.

[23] Luis Rivera-González, David Bolonio, Luis F. Mazadiego, and Robert Valencia-Chapi. Long-term electricity supply and demand forecast (2018-2040): A LEAP model application towards a sustainable power generation system in Ecuador. *Sustainability (Switzerland)*, 11(19):5316, 2019.

[24] Alla Tkachenko, Nataliia Levchenko, Ganna Shyshkanova, Dmytro Plynokos, and Marharyta Kovalenko. Efficiency forecasting for municipal solid waste recycling in the context on sustainable development of economy. In *E3S Web of Conferences*, volume 166, page 13021. EDP Sciences, 2020.

[25] Heikki Junninen, Harri Niska, Kari Tuppurainen, Juhani Ruuskanen, and Mikko Kolehmainen. Methods for imputation of missing values in air quality data sets. *Atmospheric Environment*, 38(18):2895–2907, 2004.

[26] Pruethsan Sutthichaimethee. A forecasting model in managing future scenarios to achieve the sustainable development goals of thailands environmental law: enriching the path analysis varima-ovi model. *670216917*, 2021.

[27] Jesús Crespo Cuaresma, Wolfgang Fengler, Homi Kharas, Karim Bekhtiar, Michael Brottrager, and Martin Hofer. Will the Sustainable Development Goals be fulfilled? Assessing present and future global poverty. *Palgrave Communications*, 4(1), 2018.

[28] Okewu Emmanuel, M. Ananya, Sanjay Misra, and Murat Koyuncu. A deep neural network-based advisory framework for attainment of sustainable development goals 1-6. *Sustainability (Switzerland)*, 12(24):1–16, 2020.

[29] Linda M. Richter, Jere R. Behrman, Pia Britto, Claudia Cappa, Caroline Cohrssen, Jorge Cuartas, Bernadette Daelmans, Amanda E. Devercelli, Günther Fink, Sandra Fredman, Jody Heymann, Florencia Lopez Boo, Chunling Lu, Elizabeth Lule, Dana Charles McCoy, Sara N. Naicker, Nirmalo Rao, Abbie Raikes, Alan Stein, Claudia Vazquez, and Hirokazu Yoshikawa. Measuring and forecasting progress in education: what about early childhood? *npj Science of Learning*, 6(1):1–7, 2021.

[30] Haruhiko Inada, Qingfeng Li, Abdulgafoor Bachani, and Adnan A. Hyder. Forecasting global road traffic injury mortality for 2030. *Injury Prevention*, 26(4):339–343, 2020.

[31] Christopher Yeh, Chenlin Meng, Sherrie Wang, Anne Driscoll, Erik Rozi, Patrick Liu, Jihyeon Lee, Marshall Burke, David B. Lobell, and Stefano Ermon. SustainBench: Benchmarks for Monitoring the Sustainable Development Goals with Machine Learning. *arXiv preprint arXiv:2111.04724*, 2021.

[32] Jeffrey Sachs, Christian Kroll, Guillame Lafortune, Grayson Fuller, and Finn Woelm. *Sustainable Development Report 2021*. Cambridge University Press, 2021.

[33] Barry B Hughes, Cecilia Peterson, Dale s. Rothman, and Jose R. Solorzano. IFs Health Model Documentation. *Denver: Pardee Center for International Futures*, (June):1–85, 2014.

[34] Spyros Makridakis. A Survey of Time Series. *International Statistical Review / Revue Internationale de Statistique*, 44(1):29, 1976.

[35] Spyros Makridakis and Michèle Hibon. ARMA models and the Box-Jenkins methodology. *Journal of Forecasting*, 16(3):147–163, 1997.

[36] Jin Liang Zhang, Yue Jun Zhang, and Lu Zhang. A novel hybrid method for crude oil price forecasting. *Energy Economics*, 49:649–659, 2014.

[37] Paolo Tormene, Toni Giorgino, Silvana Quaglini, and Mario Stefanelli. Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation. *Artificial Intelligence in Medicine*, 45(1):11–34, 2009.

[38] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Recompile 25 Dario Amodei, Ilya Sutskever, and Others. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[39] Rob J Hyndman. Moving Averages., 2011.

[40] Seng Hansun and Marcel Bonar Kristanda. Performance analysis of conventional moving average methods in forex forecasting. In *Proceeding of 2017 International Conference on Smart Cities, Automation and Intelligent Computing Systems, ICON-SONICS 2017*, volume 2018-January, pages 11–17. IEEE, 2017.

[41] Seng Hansun. A new approach of moving average method in time series analysis. In *2013 International Conference on New Media Studies, CoNMedia 2013*, pages 1–4. IEEE, 2013.

[42] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.

[43] Graeme D Hutcheson. Ordinary least-squares regression. *L. Moutinho and GD Hutcheson, The SAGE dictionary of quantitative management research*, pages 224–228, 2011.

[44] Lucien Le Cam. Maximum likelihood: an introduction. *International Statistical Review/Revue Internationale de Statistique*, pages 153–171, 1990.

[45] James D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.

[46] Peter G. Zhang. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50:159–175, 2003.

[47] Sean J. Taylor and Benjamin Letham. Forecasting at Scale. *American Statistician*, 72(1):37–45, 2018.

[48] Sakib Mahmud. Bangladesh covid-19 daily cases time series analysis using facebook prophet model. *Available at SSRN 3660368*, 2020.

[49] R. A. Aliev, B. Fazlollahi, Rashad R. Aliev, and B. Guirimov. Linguistic time series forecasting using fuzzy recurrent neural network. *Soft Computing*, 12(2):183–190, 2008.

[50] An Sing Chen and Mark T. Leung. Regression neural network for error correction in foreign exchange forecasting and trading. *Computers and Operations Research*, 31(7):1049–1068, 2004.

[51] G. Peter Zhang and Min Qi. Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2):501–514, 2005.

[52] Peter A Jansson. Neural networks: An overview. *Analytical chemistry*, 63(6):357A—-362A, 1991.

[53] G Dreyfus. Neural networks: an overview. *Neural networks*, pages 1–83, 2005.

[54] William Chan, Navdeep Jaitly, Quoc Le, and Oriol Vinyals. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 2016-May, pages 4960–4964. IEEE, 2016.

[55] Tsung-Hsien Wen, David Vandyke, Nikola Mrksic, Milica Gasic, Lina M Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young. A network-based end-to-end trainable task-oriented dialogue system. *arXiv preprint arXiv:1604.04562*, 2016.

[56] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[57] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural networks for perception*, pages 593–605. Elsevier, 1989.

[58] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[59] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pages 1724–1734, 2014.

[60] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.

[61] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1):388–427, 2021.

[62] Souhaib Ben Taieb, Rob J Hyndman, et al. *Recursive and direct multi-step forecasting: the best of both worlds*, volume 19. Department of Econometrics and Business Statistics, Monash Univ., 2012.

[63] Anastasia Borovykh, Sander M. Bohté, and Cornelis W. Oosterlee. Conditional time series forecasting with convolutional neural networks. *arXiv: Machine Learning*, 2017.

[64] David Salinas, Michael Bohlke-Schneider, Laurent Callot, Roberto Medico, and Jan Gasthaus. High-dimensional multivariate forecasting with low-rank gaussian copula processes. *Advances in neural information processing systems*, 32, 2019.

[65] Ruofeng Wen, Kari Torkkola, Balakrishnan Narayanaswamy, and Dhruv Madeka. A multi-horizon quantile recurrent forecaster. *arXiv preprint arXiv:1711.11053*, 2017.

[66] Elena Kuznetsova, Margarita Karlova, and Tatiana Fomina. Modeling and Forecasting of the Lipetsk Region Socioeconomic Indicators in the Context of Sustainable Development. In *Proceedings - 2020 2nd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency, SUMMA 2020*, pages 225–227, 2020.

[67] Paresh Kumar Narayan and Russell Smyth. Multivariate granger causality between electricity consumption, exports and GDP: Evidence from a panel of Middle Eastern countries. *Energy Policy*, 37(1):229–236, 2009.

[68] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and Statistical Modeling with Python. In *Proceedings of the 9th Python in Science Conference*, pages 92–96, 2010.

[69] Camila Epprecht, D Guegan, and Á Veiga. *Comparing variable selection techniques for linear regression: LASSO and Autometrics.* Centre d'économie de la Sorbonne, 2013.

[70] Per-Erik Danielsson. Euclidean distance mapping. *Computer Graphics and image processing*, 14(3):227–248, 1980.

[71] Shivani Gupta and Atul Gupta. Dealing with noise problem in machine learning data-sets: A systematic review. *Procedia Computer Science*, 161:466–474, 2019.

[72] Maytal Saar-Tsechansky and Foster Provost. Handling missing values when applying classification models. *Journal of Machine Learning Research*, 8:1625–1657, 2007.

[73] Salvador García, Julián Luengo, and Francisco Herrera. Dealing with missing values. In *Intelligent Systems Reference Library*, volume 72, pages 59–105. Springer, 2015.

[74] W. Z. Liu, A. P. White, S. G. Thompson, and M. A. Bramer. Techniques for dealing with missing values in classification. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 1280, pages 527–536. Springer, 1997.

[75] Deepti Vibha and Kameshwar Prasad. How to Deal with Missing Data? *Neurology India*, 68(4):886–888, 2020.

[76] Zhiwu Lu, Zhenyong Fu, Tao Xiang, Peng Han, Liwei Wang, and Xin Gao. Learning from weak and noisy labels for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(3):486–500, 2017.

[77] George Athanasopoulos, Roman A. Ahmed, and Rob J. Hyndman. Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, 25(1):146–166, 2009.

[78] Ju Jie Wang, Jian Zhou Wang, Zhe George Zhang, and Shu Po Guo. Stock index forecasting based on a hybrid model. *Omega*, 40(6):758–766, 2012.

[79] Elaine Deschamps. The impact of institutional change on forecast accuracy: A case study of budget forecasting in Washington State. *International Journal of Forecasting*, 20(4):647–657, 2004.

[80] J.C. Sackellares, L.D. Iasemidis, D.S. Shiau, and L. Dance. Multi-dimensional multi-parameter time series processing for seizure warning and prediction, 2004.

[81] Zin Lin. Update History of the Global SDG Indicators Database and Metadata Repository. Technical report.

[82] J. Evans. Sustainable Development Goals data update, UK - Office for National Statistics, 2020.

[83] Charles A. Hall and W. Weston Meyer. Optimal error bounds for cubic spline interpolation. *Journal of Approximation Theory*, 16(2):105–122, 1976.

[84] Heikki Junninen, Harri Niska, Kari Tuppurainen, Juhani Ruuskanen, and Mikko Kolehmainen. Methods for imputation of missing values in air quality data sets. *Atmospheric Environment*, 38(18):2895–2907, 2004.

[85] F Pedregosa, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, P Prettenhofer, R Weiss, V Dubourg, J Vanderplas, A Passos, D Cournapeau, M Brucher, M Perrot, and E Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[86] Sean J. Taylor and Benjamin Letham. Forecasting at Scale. *American Statistician*, 72(1):37–45, 2018.

[87] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, dec 2011.

[88] A. J. Lawrance and P. A. W. Lewis. The Exponential Autoregressive-Moving Average Earma(P,Q) Process . *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):150–161, 1980.

[89] William R Kinney Jr. ARIMA and Regression in Analytical Review: An Empirical Test. *The Accounting Review*, 53(1):48–60, 1978.

[90] Heidi L. Roth, Andrea N. Lora, and Kenneth M. Heilman. *Effects of monocular viewing and eye dominance on spatial attention*, volume 125. Machine Learning Mastery, 2002.

[91] Rob J Hyndman and George Athanasopoulos. Forecasting: Principles and Practice, may 2018.

[92] Aylin Alin. Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3):370–374, 2010.

[93] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R.J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T. Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, and Yoshiki Vázquez-Baeza. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020.

[94] Omer Sagi and Lior Rokach. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1249, 2018.

[95] Bartosz Krawczyk, Leandro L. Minku, João Gama, Jerzy Stefanowski, and Michał Woźniak. Ensemble learning for data stream analysis: A survey. *Information Fusion*, 37:132–156, 2017.

[96] Abdelhamid Bouchachia and Saliha Bouchachia. Ensemble Learning for Time Series Prediction. *First International workshop on Nonlinear Dynamics and Synchronization*, 2008.

[97] Peter Bühlmann. Bagging, boosting and ensemble methods. In *Handbook of Computational Statistics: Concepts and Methods: Second Edition*, pages 985–1022. Springer, 2012.

[98] Bee Wah Yap, Khatijahhusna Abd Rani, Hezlin Aryani Abd Rahman, Simon Fong, Zuraida Khairudin, and Nik Nairan Abdullah. An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. In *Lecture Notes in Electrical Engineering*, volume 285 LNEE, pages 13–22. Springer, 2014.

[99] Robert Bryll, Ricardo Gutierrez-Osuna, and Francis Quek. Attribute bagging: Improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition*, 36(6):1291–1302, 2003.

[100] Muhammad Atif Tahir, Josef Kittler, and Ahmed Bouridane. Multilabel classification using heterogeneous ensemble of multi-label classifiers. *Pattern Recognition Letters*, 33(5):513–523, 2012.

[101] Huanfei Ma, Kazuyuki Aihara, and Luonan Chen. Detecting causality from nonlinear dynamics with short-term time series. *Scientific Reports*, 4(1):1–10, 2014.

[102] Kai Chen, Yi Zhou, and Fangyan Dai. A LSTM-based method for stock returns prediction: A case study of China stock market. In *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, pages 2823–2824. IEEE, 2015.

[103] Yao Qin, Dongjin Song, Haifeng Cheng, Wei Cheng, Guofei Jiang, and Garrison W. Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. *IJCAI International Joint Conference on Artificial Intelligence*, 0:2627–2633, 2017.

[104] M Battula. Time series forecasting with deep stacked unidirectional and bidirectional LSTMs, 2019.

[105] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 4(January):3104–3112, 2014.

[106] Ashok Swain. Ethiopia, the Sudan, and Egypt: The Nile River dispute. *Journal of Modern African Studies*, 35(4):675–694, 1997.

[107] Jeffrey Voas, Nir Kshetri, and Joanna F. Defranco. Scarcity and Global Insecurity: The Semiconductor Shortage. *IT Professional*, 23(5):78–82, 2021.

[108] Nathan Burks, Adetokunbo Fadahunsi, and Ann Marie Hibbert. Financial Contagion: A Tale of Three Bubbles. *Journal of Risk and Financial Management*, 14(5):229, 2021.

[109] United Division. SDG Indicators, 2021.

[110] Christos Koutlis, Symeon Papadopoulos, Manos Schinas, and Ioannis Kompatsiaris. Lavarnet: Neural network modeling of causal variable relationships for multivariate time series forecasting. *Applied Soft Computing*, 96:106685, 2020.

Appendix A

Sustainable Development Goals Taxonomy

The following Table gives the unique thresholds found in the SDG data presented in Chapter 4.

#	GTI	Thresholds	Date	Target Description	Series Description
1	1.1.1	<=0.5%	2030	By 2030,eradicate extreme poverty for all people everywhere, measured as people living on less than $1.25 a day	A Proportion of population below international poverty line
2	1.2.1	<=50%	2030	By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions	Proportion of population living below the national poverty line (%)

3	1.3.1	>=80%	2030	Implement nationally appropriate social protection systems and measures for all, including floors, and by 2030 achieve substantial coverage of the poor and the vulnerable	[ILO] Proportion of population covered by at least one social protection benefit (%)
4	1.4.1	>=99.5%	2030	By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including micro finance	Proportion of population using basic drinking water services, by location (%)
5	1.5.3	>=1	2020	By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental disasters	Countries with legislative and/or regulatory provisions been made for managing disaster risk (1 = YES; 0 = NO)

6	1.5.3	>=60%	2030	By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental disasters	Score of adoption and implementation of national DRR strategies in line with the Sendai Framework
7	1.a.1	>=70%	2030	Ensure significant mobilisation of resources from a variety of sources, including through enhanced development cooperation, in order to provide adequate and predictable means for developing countries, in particular least developed countries, to implement programmes and policies to end poverty in all its dimensions	Official development assistance grants for poverty reduction, by donor countries (percentage of GNI)
8	2.3.1	>=100%	2030	By 2030, double the agricultural productivity and incomes of small-scale food producers, in particular women, indigenous peoples, family farmers, pastoralists and fishers, including through secure and equal access to land, other productive resources and inputs, knowledge, financial services, markets and opportunities for value addition and non-farm employment	Productivity of large-scale food producers (agricultural output per labour day, PPP) (constant 2011 international $)

9	2.5.1	>=60%	2030	By 2020, maintain the genetic diversity of seeds, cultivated plants and farmed and domesticated animals and their related wild species, including through soundly managed and diversified seed and plant banks at the national, regional and international levels, and promote access to and fair and equitable sharing of benefits arising from the utilisation of genetic resources and associated traditional knowledge, as internationally agreed	number of local breeds kept in the country
10	2.5.2	<=60%	2030	By 2020, maintain the genetic diversity of seeds, cultivated plants and farmed and domesticated animals and their related wild species, including through soundly managed and diversified seed and plant banks at the national, regional and international levels, and promote access to and fair and equitable sharing of benefits arising from the utilisation of genetic resources and associated traditional knowledge, as internationally agreed	Local breeds classified as being at level of risk of extinction (number)

11	3.1.1	<=70	2030	By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births	Maternal mortality ratio
12	3.1.2	>=70	2030	By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births	Proportion of births attended by skilled health personnel (%)
13	3.2.1	<=25	2030	By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births	Infant mortality rate (deaths per 1,000 live births)
14	3.2.2	<=12	2030	By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births	Neonatal mortality rate (deaths per 1,000 live births)
15	3.4.1	<=30%	2030	By 2030, reduce by one third premature mortality from non-communicable diseases through prevention and treatment and promote mental health and well-being	number of deaths attributed to non-communicable diseases, by type of disease and sex (number)

16	3.7.1	>=99%	2030	By 2030, ensure universal access to sexual and reproductive health-care services, including for family planning, information and education, and the integration of reproductive health into national strategies and programmes	Proportion of women married or in a union of reproductive age (aged 15-49 years) who have their need for family planning satisfied with modern methods (% of women aged 15-49 years)
17	3.8.2	<=99%	2030	Achieve universal health coverage, including financial risk protection, access to quality essential health-care services and access to safe, effective, quality and affordable essential medicines and vaccines for all	Proportion of population with large household expenditures on health (greater than 10%) as a share of total household expenditure or income (%)
18	3.9.1	<=50	2030	By 2030, substantially reduce the number of deaths and illnesses from hazardous chemicals and air, water and soil pollution and contamination	Age-standardized mortality rate attributed to ambient air pollution (deaths per 100,000 population)

19	3.d.2	<=70%	2030	Strengthen the capacity of all countries, in particular developing countries, for early warning, risk reduction and management of national and global health risks	Percentage of bloodstream infection due to Escherichia coli resistant to 3rd-generation cephalosporin (e.g., ESBL-E. coli) among patients seeking care and whose blood sample is taken and tested (%)
20	4.5.1	>=99.5%	2030	4.5 By 2030, eliminate gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples and children in vulnerable situations	Proportion of youth and adults with information and communications technology (ICT) skills, by sex and type of skill (%)
21	4.6.1	>=99.5%	2030	By 2030, ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy and numeracy	Proportion of population achieving at least a fixed level of proficiency in functional skills, by sex, age and type of skill (%)

22	5.5.1	Unknown	2030	Ensure women's full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic and public life	Proportion of seats held by women in national parliaments (% of total number of seats)
23	6.b.1	>=10	2030	Support and strengthen the participation of local communities in improving water and sanitation management	Countries with procedures in law or policy for participation by service users/communities in planning program in water resources planning and management, by level of definition in procedures (10 = Clearly defined; 5 = Not clearly defined ; 0 = NA)
24	7.3.1	>=100%	2030	By 2030, double the global rate of improvement in energy efficiency	Energy intensity level of primary energy (megajoules per constant 2011 purchasing power parity GDP)

25	8.1.1	>=7%	2030	Sustain per capita economic growth in accordance with national circumstances and, in particular, at least 7 per cent gross domestic product growth per annum in the least developed countries	Annual growth rate of real GDP per capita (%)
26	8.4.2	<=40%	2030	Improve progressively, through 2030, global resource efficiency in consumption and production and endeavour to decouple economic growth from environmental degradation, in accordance with the 10-Year Framework of Programmes on Sustainable Consumption and Production, with developed countries taking the lead	Domestic material consumption per capita, by type of raw material (tonnes)
27	8.8.1	<=100	2030	Protect labour rights and promote safe and secure working environments for all workers, including migrant workers, in particular women migrants, and those in precarious employment	Non-fatal occupational injuries among employees, by sex and migrant status (per 100,000 employees)

28	10.7.2	>=4	2030	Facilitate orderly, safe, regular and responsible migration and mobility of people, including through the implementation of planned and well-managed migration policies	Countries with migration policies to facilitate orderly, safe, regular and responsible migration and mobility of people, by policy domain (1 = Requires further progress; 2 = Partially meets; 3 = Meets; 4 = Fully meets)
29	10.c.1	<=3%	2030	By 2030, reduce to less than 3 per cent the transaction costs of migrant remittances and eliminate remittance corridors with costs higher than 5 per cent	Average remittance costs of sending $200 to a receiving country as a proportion of the amount remitted (%)
30	14.5.1	>=10%	2020	By 2020, conserve at least 10 per cent of coastal and marine areas, consistent with national and international law and based on the best available scientific information	Coverage of protected areas in relation to marine areas (Exclusive Economic Zones) (%)

31	14.6.1	>=5	2020	By 2020, prohibit certain forms of fisheries subsidies which contribute to overcapacity and overfishing, eliminate subsidies that contribute to illegal, unreported and unregulated fishing and refrain from introducing new such subsidies, recognising that appropriate and effective special and differential treatment for developing and least developed countries should be an integral part of the World Trade Organization fisheries subsidies negotiation4	progress by countries in the degree of implementation of international instruments aiming to combat illegal, unreported and unregulated fishing (level of implementation: 1 lowest to 5 highest)
32	15.2.1	>=40%	2020	By 2020, promote the implementation of sustainable management of all types of forests, halt deforestation, restore degraded forests and substantially increase afforestation and reforestation globally	Above-ground biomass in forest per hectare (tonnes per hectare)

33	16.2.1	<=99.5%	2030	End abuse, exploitation, trafficking and all forms of violence against and torture of children	Proportion of children aged 1-14 years who experienced physical punishment and/or psychological aggression by caregivers in last month (% of children aged 1-14 years)
34	16.9.1	>=99.5%	2030	By 2030, provide legal identity for all, including birth registration	Proportion of children under 5 years of age whose births have been registered with a civil authority (% of children under 5 years of age)

35	17.2.1	>=20%	2030	Developed countries to implement fully their official development assistance commitments, including the commitment by many developed countries to achieve the target of 0.7 per cent of gross national income for official development assistance (ODA/GNI) to developing countries and 0.15 to 0.20 per cent of ODA/GNI to least developed countries; ODA providers are encouraged to consider setting a target to provide at least 0.20 per cent of ODA/GNI to least developed countries	Net official development assistance (ODA) to LDCs as a percentage of OECD-DAC donors' GNI, by donor countries (%)

Table A.1:

Appendix B

Barkla Guide

The following document was written as an introductory guideline to High-performance clusters for first-time users, which was presented in Chapter 4.

The Barkla 5 Minutes Guide to Setting up a Python 3 Conda Environment

Yassir Alharbi

April 30, 2021

1 Introduction

Barkla is a high-performance computing cluster that is available for research at the University of Liverpool. This guide will walk you through a "short cut" to setting up a python 3 Conda environment and then running python code in interactive mode. If you want more than just simply running code and obtaining the results on the screen, like for example specifying the hardware needed for your experiment, or running a job that might take more than 12 hours to finish, you should read the "proper" documentation prepared by the computer services department, "Brief_Barkla_User_Guide-v2.0.pdf". The rest of this guide covers: (i) registration in Section 2, (ii) connecting to Barkla in Section 3, (iii) file uploading and downloading in Section 4, (iv) setting up a Conda Python environment and running your python code in Section 5, and (v) quota allowance in Section 6. Finally, at the end of this guide, a brief summary is provided in Section 7.

2 Registrations

Before you can use Barkla, you need to register to use the cluster, also you will need to register to use the university VPN if you plan to connect from outside the university network. The Barkla registration form can be accessed by navigating to the Computer Services pages on the UoL www site, and then Advance Research Computing[1]. To register for VPN, navigate to the Computer Services pages on the UoL www site and then to VPN where you will find the registration form[2].

3 Connection

Barkla requires a command-line interface. Familiarising your self with basic Linux commands can shorten the learning curve tremendously. Once your registration is approved for using the system, activate the VPN, then launch a terminal if you are using a Mac or

[1] https://www.liverpool.ac.uk/csd/advanced-research-computing/access/
[2] https://www.liverpool.ac.uk/csd/vpn/

Figure 1: Barkla login screen

Linux, or a MobaXterm if you are using Windows. The login credentials are the same as your email username (your MWS credentials). Type the following to login to Barkla. `ssh ⟨USER_NAME⟩@barkla4.liv.ac.uk`, then enter your password. If unable to login the problem is likely to be that your VPN is not configured correctly, or your MWS credentials are not valid. After successfully logging in, you should see a similar screen to that shown in Figure 1.

4 Uploading-Downloading files

Now you are ready to move your files from your local machine to Barkla, and this can be done through the terminal by using the Linux `scp` (Secure Copy) command

- In Barkla, use the Linux command, `mkdir ⟨FOLDER_NAME⟩`, to create a folder (make a directory) for your project
- On your local machine use the terminal to navigate to your local project files using the unix command `cd` (change directory).
- Once in the folder initiate the SCP tool by typing:
  `scp -r * ⟨USER_NAME⟩@barkla4.liv.ac.uk: /⟨FOLDER_NAME⟩`
  and then enter your password. This will cause all the files in your folder to be uploaded to Barkla. If you want to upload a specific file replace the * with the name of the file to be uploaded.

- In Barkla navigate to your project file and type the unix command `ls` (list), you should see all your files.
- In your local terminal run `scp -r ⟨USER_NAME⟩@barkla4.liv.ac.uk: /⟨FOLDER_NAME⟩ ../⟨LOCAL_FOLDER⟩`.

5 Conda environment

Note that Barkla includes 2 login nodes. Barkla4 which we were using above, and Barkla5. Booth nodes are available for "light" jobs only, like moving your files. For more substantial jobs it is better to use the SSH protocol, the commands `ssh viz01` or `ssh viz02`, to log into the visualisation nodes which can handle more intensive operations. After doing that, you are ready to create your own bespoke Conda environment. To do this you need to first load the Conda module, then create your Conda environment and finally install your required packages. The following commands can be used:

- `module add apps/anaconda3/5.2.0`
- `conda create --name ⟨ENVIRONMENT_NAME⟩`
- `conda activate ⟨ENVIRONMENT_NAME⟩`
- `pip install ⟨YOUR_PACKAGES⟩`

now run your code by typing `python ⟨YOUR_PYTHON_FILE⟩`. If you discover packages that are missing use the pip tool to install them.

6 Quota

Barkla is a shared service. As all shared services, it should be used sensibly. Thus there is a quota system. Your quota can be consumed very quickly if your code generates a lot of files. The limitation at the time of writing this document was 100,000 files and 50GB storage. Note that installing torch, for example, require 700MG, running out of space can happen and then you cannot get anything done. You can check your quota allowance by typing `quota -s`.

7 Summary

This 5 minutes guide to setting up a Python 3 Conda environment on the Barkla system has walked you through: registration, logging in, moving your files, creating your environment, running your code and checking your quota allowance. This guide is not meant to replace the "proper" instructional guide that can be obtained from Computer Services. Familiarising your self with Linux commands can speed up significantly your ability to use Barkla.

Appendix C

Encoder-Decoder Input/Output

The following two-part Table shows the complete input-output sequence, which was presented in Chapter 6.

Steps	**Year**	**Input**					**Output**					
		8.4.2_2664	1.4.1_2271	12.2.2_2754	8.4.2_1071	8.4.2_1086	8.4.2_2664	1.4.1_2271	12.2.2_2754	8.4.2_1071	8.4.2_1086	1.4.1_2270 TARGET
1 Test	2000	-0.828	-0.899	2.457	-0.493	-0.967						
	2001	-0.819	-0.807	2.753	-0.494	-0.967						
	2002	-0.541	-0.715	2.113	-0.494	-1.135						
	2003	-0.871	-0.624	1.041	-0.493	-1.135						
	2004	Validate 1					-0.743	-0.514	0.034	-0.347	-1.135	**-0.518**
	2005						-0.755	-0.402	0.717	-0.274	0.713	**-0.405**
	2006						-0.228	-0.289	0.281	-0.167	0.940	**-0.290**
2 Test	2001	-0.819	-0.807	2.753	-0.494	-0.967						
	2002	-0.541	-0.715	2.113	-0.494	-1.135						
	2003	-0.871	-0.624	1.041	-0.493	-1.135						
	2004	-0.743	-0.514	0.034	-0.347	-1.135						
	2005	Validate 2					-0.755	-0.402	0.717	-0.274	0.713	**-0.405**
	2006						-0.228	-0.289	0.281	-0.167	0.940	**-0.290**
	2007						-0.073	-0.175	0.454	-0.009	-0.018	**-0.175**
3 Test	2002	-0.541	-0.715	2.113	-0.494	-1.135						
	2003	-0.871	-0.624	1.041	-0.493	-1.135						
	2004	-0.743	-0.514	0.034	-0.347	-1.135						
	2005	-0.755	-0.402	0.717	-0.274	0.713						
	2006	Validate 3					-0.228	-0.289	0.281	-0.167	0.940	**-0.290**
	2007						-0.073	-0.175	0.454	-0.009	-0.018	**-0.175**
	2008						-0.170	-0.059	0.278	-0.009	-0.011	**-0.059**
4 Test	2003	-0.871	-0.624	1.041	-0.493	-1.135						
	2004	-0.743	-0.514	0.034	-0.347	-1.135						
	2005	-0.755	-0.402	0.717	-0.274	0.713						
	2006	-0.228	-0.289	0.281	-0.167	0.940						
	2007	Validate 4					-0.073	-0.175	0.454	-0.009	-0.018	**-0.175**
	2008						-0.170	-0.059	0.278	-0.009	-0.011	**-0.059**
	2009						0.073	0.059	-0.034	0.009	-0.010	**0.059**
5 Test	2004	-0.743	-0.514	0.034	-0.347	-1.135						
	2005	-0.755	-0.402	0.717	-0.274	0.713						
	2006	-0.228	-0.289	0.281	-0.167	0.940						
	2007	-0.073	-0.175	0.454	-0.009	-0.018						
	2008	Validate 5					-0.170	-0.059	0.278	-0.009	-0.011	**-0.059**
	2009						0.073	0.059	-0.034	0.009	-0.010	**0.059**
	2010						0.163	0.178	-0.315	0.428	-0.007	**0.178**
6 Test	2005	-0.755	-0.402	0.717	-0.274	0.713						
	2006	-0.228	-0.289	0.281	-0.167	0.940						
	2007	-0.073	-0.175	0.454	-0.009	-0.018						
	2008	-0.170	-0.059	0.278	-0.009	-0.011						
	2009	Validate 6					0.073	0.059	-0.034	0.009	-0.010	**0.059**
	2010						0.163	0.178	-0.315	0.428	-0.007	**0.178**
	2011						0.172	0.299	-0.344	0.547	0.007	**0.298**
7 Test	2006	-0.228	-0.289	0.281	-0.167	0.940						
	2007	-0.073	-0.175	0.454	-0.009	-0.018						
	2008	-0.170	-0.059	0.278	-0.009	-0.011						
	2009	0.073	0.059	-0.034	0.009	-0.010						
	2010	Validate 7					0.163	0.178	-0.315	0.428	-0.007	**0.178**
	2011						0.172	0.299	-0.344	0.547	0.007	**0.298**
	2012						0.249	0.421	-0.271	0.618	0.095	**0.419**

Table C.1: Complete Multivariate forecasting Input-Output sequential to forecast Time series Afghanistan_1.4.1_2270 Part1

Steps	Year	Input					Output					
		8.4.2_2664	1.4.1_2271	12.2.2_2754	8.4.2_1071	8.4.2_1086	8.4.2_2664	1.4.1_2271	12.2.2_2754	8.4.2_1071	8.4.2_1086	1.4.1_2270 TARGET
	2007	-0.073	-0.175	0.454	-0.009	-0.018						
	2008	-0.170	-0.059	0.278	-0.009	-0.011						
	2009	0.073	0.059	-0.034	0.009	-0.010						
8 Test	2010	0.163	0.178	-0.315	0.428	-0.007						
	2011						0.172	0.299	-0.344	0.547	0.007	**0.298**
	2012	Validate 8					0.249	0.421	-0.271	0.618	0.095	**0.419**
	2013						0.327	0.545	-0.350	0.689	0.183	**0.541**
	2008	-0.170	-0.059	0.278	-0.009	-0.011						
	2009	0.073	0.059	-0.034	0.009	-0.010						
	2010	0.163	0.178	-0.315	0.428	-0.007						
9 Test	2011	0.172	0.299	-0.344	0.547	0.007						
	2012						0.249	0.421	-0.271	0.618	0.095	**0.419**
	2013	Validate 9					0.327	0.545	-0.350	0.689	0.183	**0.541**
	2014						0.406	0.670	-0.399	0.760	0.270	**0.665**
	2009	0.073	0.059	-0.034	0.009	-0.010						
	2010	0.163	0.178	-0.315	0.428	-0.007						
	2011	0.172	0.299	-0.344	0.547	0.007						
10 Test	2012	0.249	0.421	-0.271	0.618	0.095						
	2013						0.327	0.545	-0.350	0.689	0.183	**0.541**
	2014	Validate 10					0.406	0.670	-0.399	0.760	0.270	**0.665**
	2015						0.406	0.670	-0.399	0.760	0.270	**0.665**
	2010	0.163	0.178	-0.315	0.428	-0.007						
	2011	0.172	0.299	-0.344	0.547	0.007						
	2012	0.249	0.421	-0.271	0.618	0.095						
11 Test	2013	0.327	0.545	-0.350	0.689	0.183						
	2014						0.406	0.670	-0.399	0.760	0.270	**0.665**
	2015	Validate 11					0.406	0.670	-0.399	0.760	0.270	**0.665**
	2016						0.406	0.670	-0.399	0.760	0.270	**0.665**
	2011	0.172	0.299	-0.344	0.547	0.007						
	2012	0.249	0.421	-0.271	0.618	0.095						
	2013	0.327	0.545	-0.350	0.689	0.183						
12 Test	2014	0.406	0.670	-0.399	0.760	0.270						
	2015						0.406	0.670	-0.399	0.760	0.270	**0.665**
	2016	Validate 12					0.406	0.670	-0.399	0.760	0.270	**0.665**
	2017						0.406	0.670	-0.399	0.760	0.270	**0.665**
	2012	0.249	0.421	-0.271	0.618	0.095						
	2013	0.327	0.545	-0.350	0.689	0.183						
	2014	0.406	0.670	-0.399	0.760	0.270						
13 test	2015	0.406	0.670	-0.399	0.760	0.270						
	2016						0.406	0.670	-0.399	0.760	0.270	**0.665**
	2017	Validate 13					0.406	0.670	-0.399	0.760	0.270	**0.665**
	2018						0.702	0.969	-1.172	1.030	0.616	**0.973**
	2013	0.327	0.545	-0.350	0.689	0.183						
	2014	0.406	0.670	-0.399	0.760	0.270						
	2015	0.406	0.670	-0.399	0.760	0.270						
Test	2016	0.406	0.670	-0.399	0.760	0.270						
	2017						0.406	0.670	-0.399	0.760	0.270	**0.665**
	2018	Validate					0.702	0.969	-1.172	1.030	0.616	**0.973**
	2019						0.790	1.072	-1.338	1.124	0.694	**1.078**

Table C.2: Complete Multivariate forecasting Input-Output sequential to forecast Time series Afghanistan_1.4.1_2270 Part2

Appendix D

Sri Lanka SDG Attainment Prediction Results Obtained Using The SDG-TTF Framework

The following table gives the full Sri Lankan attainment predictions results obtained using the SDG-TTF framework presented in Chapter 7.

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
1.4.1	SP_ACS_BSRVSAN	ALLAREA	95.782	93.150	F
1.4.1	SP_ACS_BSRVSAN	RURAL	95.999	93.490	F
1.4.1	SP_ACS_BSRVSAN	URBAN	94.819	90.650	F
1.4.1	SP_ACS_BSRVH2O	ALLAREA	89.416	86.990	F
1.4.1	SP_ACS_BSRVH2O	RURAL	87.720	85.290	F
1.4.1	SP_ACS_BSRVH2O	URBAN	96.949	97.320	F
1.1.1	SI_POV_EMP1	15-24_MALE	0.900	1.970	F
1.1.1	SI_POV_EMP1	MALE_15+	0.700	1.620	F
1.1.1	SI_POV_EMP1	MALE_25+	0.600	1.440	F
1.1.1	SI_POV_EMP1	FEMALE_15-24	0.500	1.140	F
1.1.1	SI_POV_EMP1	FEMALE_15+	0.600	1.360	F
1.1.1	SI_POV_EMP1	FEMALE_25+	0.600	1.370	F
1.1.1	SI_POV_EMP1	BOTHSEX_15-24	0.700	1.460	F
1.1.1	SI_POV_EMP1	BOTHSEX_15+	0.600	1.380	F
1.1.1	SI_POV_EMP1	BOTHSEX_25+	0.600	1.300	F
2.2.2	SH_STA_ANEM	15-49_FEMALE	34.100	45.140	F
2.a.1	AG_PRD_AGVAS	A	7.782	7.850	F
2.c.1	AG_FPA_COMM	RIC	0.400	-1.060	F
2.c.1	AG_FPA_COMM	WHE	-0.400	-3.730	F
2.2.2	SH_STA_ANEM	2.2.3_15-49_FEMALE	34.100	34.320	F
2.2.2	SH_STA_ANEM	2.2.3_15-49_FEMALE	34.300	34.140	F
2.a.1	AG_PRD_ORTIND	A	0.687	0.630	F
2.a.1	AG_XPD_AGSGB	A	5.343	4.790	F
2.a.2	DC_TOF_AGRL	2018	55.649	55.680	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
3.2.1	SH_DYN_IMRTN	MALE_<1Y	1273.000	101.020	T
3.2.1	SH_DYN_IMRTN	FEMALE_<1Y	1006.000	1301.390	F
3.2.1	SH_DYN_IMRTN	BOTHSEX_<1Y	2279.000	3097.280	F
3.a.1	SH_PRV_SMOK	MALE_15+	43.400	44.130	F
3.a.1	SH_PRV_SMOK	FEMALE_15+	2.900	3.530	F
3.a.1	SH_PRV_SMOK	BOTHSEX_15+	23.100	23.710	F
3.2.1	SH_DYN_IMRT	MALE_<1Y	7.300	8.530	F
3.2.1	SH_DYN_IMRT	FEMALE_<1Y	6.000	7.490	F
3.2.1	SH_DYN_IMRT	BOTHSEX_<1Y	6.700	7.960	F
3.3.1	SH_HIV_INCD	15-24_MALE	0.020	0.020	T
3.3.1	SH_HIV_INCD	MALE_50+	0.010	0.010	T
3.3.1	SH_HIV_INCD	15-49_MALE	0.020	0.030	F
3.3.1	SH_HIV_INCD	ALLAGE_MALE	0.010	0.020	F
3.3.1	SH_HIV_INCD	FEMALE_15-24	0.010	0.000	T
3.3.1	SH_HIV_INCD	FEMALE_15-49	0.010	0.010	T
3.3.1	SH_HIV_INCD	ALLAGE_FEMALE	0.000	0.000	T
3.3.1	SH_HIV_INCD	BOTHSEX_15-24	0.010	0.020	F
3.3.1	SH_HIV_INCD	BOTHSEX_50+	0.010	-0.010	T
3.3.1	SH_HIV_INCD	BOTHSEX_15-49	0.010	0.020	F
3.3.1	SH_HIV_INCD	BOTHSEX_ALLAGE	0.010	0.010	T
3.b.2	DC_TOF_HLTHL	2018	51.133	17.680	F
3.2.1	SH_DYN_MORT	MALE_<5Y	8.500	1.520	T
3.2.1	SH_DYN_MORT	FEMALE_<5Y	7.100	8.820	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
3.2.1	SH_DYN_MORT	BOTHSEX_<5Y	7.800	8.120	F
3.c.1	SH_MED_DEN	DENT	0.697	0.660	F
3.c.1	SH_MED_DEN	PHAR	0.770	0.690	F
3.c.1	SH_MED_DEN	PHYS	9.277	8.050	F
3.c.1	SH_MED_DEN	NURSMID	19.966	19.310	F
3.2.1	SH_DYN_MORTN	MALE_<5Y	1485.000	867.240	T
3.2.1	SH_DYN_MORTN	FEMALE_<5Y	1188.000	1530.320	F
3.2.1	SH_DYN_MORTN	BOTHSEX_<5Y	2673.000	-31.210	T
3.b.2	DC_TOF_HLTHNT	2018	48.175	20.660	F
3.2.2	SH_DYN_NMRTN	BOTHSEX_<1M	1601.000	1941.400	F
3.d.1	SH_IHR_CAPS	IHR01	100.000	32.000	F
3.d.1	SH_IHR_CAPS	IHR10	56.000	6.070	F
3.d.1	SH_IHR_CAPS	IHR06	57.000	21.910	F
3.d.1	SH_IHR_CAPS	IHR03	95.000	30.120	F
3.d.1	SH_IHR_CAPS	IHR07	80.000	172.580	T
3.d.1	SH_IHR_CAPS	IHR11	80.000	1386.570	T
3.d.1	SH_IHR_CAPS	IHR09	97.000	94.840	F
3.d.1	SH_IHR_CAPS	IHR12	62.000	30.500	F
3.d.1	SH_IHR_CAPS	IHR04	94.000	-397.150	F
3.d.1	SH_IHR_CAPS	IHR02	100.000	110.450	F
3.d.1	SH_IHR_CAPS	IHR05	45.000	25.550	F
3.d.1	SH_IHR_CAPS	IHR13	62.000	-105.780	F
3.d.1	SH_IHR_CAPS	IHR08	58.000	54.250	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
3.2.2	SH_DYN_NMRT	BOTHSEX_<1M	4.700	5.780	F
3.7.2	SP_DYN_ADKL	FEMALE_15-19	21.000	22.300	F
4.c.1	SE_TRA_GRDL	BOTHSEX_PRIMAR	85.479	83.240	U
4.b.1	DC_TOF_SCHIPSL	2018	6.477	6.770	U
4.3.1	SE_ADT_EDUCTRN	BOTHSEX	0.832	0.500	F
5.5.2	IC_GEN_MGTL	FEMALE	27.580	24.300	F
5.5.1	SG_GEN_PARL	FEMALE	5.780	5.720	F
5.5.2	IC_GEN_MGTN	FEMALE	23.170	25.510	F
5.5.1	SG_GEN_PARLN	FEMALE	13.000	12.870	F
6.a.1	DC_TOF_WASHL	2018	155.774	154.620	F
6.2.1	SH_SAN_DEFECT	ALLAREA	0.559	1.490	T
6.2.1	SH_SAN_DEFECT	RURAL	0.685	1.050	F
6.2.1	SH_SAN_DEFECT	URBAN	0.000	0.580	T
6.4.2	ER_H2O_STRESS	I	90.790	90.800	F
6.1.1	SH_H2O_SAFE	URBAN	90.848	89.580	F
7.3.1	EG_EGY_PRIM	2011	1.810	1.950	F
7.1.1	EG_ACS_ELEC	ALLAREA	98.000	90.620	F
7.1.1	EG_ACS_ELEC	RURAL	97.000	87.910	F
7.1.1	EG_ACS_ELEC	URBAN	100.000	130.710	F
7.a.1	EG_IFF_RANDN	2018	215.460	58.210	F
8.8.1	SL_EMP_INJUR	MALE__T	23.000	18.140	T
8.8.1	SL_EMP_INJUR	FEMALE__T	12.100	16.140	F
8.8.1	SL_EMP_INJUR	BOTHSEX__T	19.000	18.250	T

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
8.a.1	DC_TOF_TRDDBML	2018	269.003	371.350	F
8.a.1	DC_TOF_TRDCML	2018	502.345	469.820	F
8.3.1	SL_ISV_IFEM	MALE_NONAGR	63.900	65.660	F
8.3.1	SL_ISV_IFEM	MALE_TOTAL	70.500	69.310	F
8.3.1	SL_ISV_IFEM	MALE_ISIC4_A	91.200	180.390	F
8.3.1	SL_ISV_IFEM	FEMALE_NONAGR	54.900	55.670	F
8.3.1	SL_ISV_IFEM	FEMALE_TOTAL	64.000	64.590	F
8.3.1	SL_ISV_IFEM	FEMALE_ISIC4_A	86.000	87.850	F
8.3.1	SL_ISV_IFEM	BOTHSEX_NONAGR	60.800	62.400	F
8.3.1	SL_ISV_IFEM	BOTHSEX_TOTAL	68.200	69.660	F
8.3.1	SL_ISV_IFEM	BOTHSEX_ISIC4_A	89.100	89.700	F
8.10.1	FB_ATM_TOTL	A	17.204	15.980	F
8.5.2	SL_TLF_UEM	15-24_MALE	14.600	-8.330	T
8.5.2	SL_TLF_UEM	MALE_15+	2.800	3.250	F
8.5.2	SL_TLF_UEM	MALE_25+	1.300	1.330	F
8.5.2	SL_TLF_UEM	FEMALE_15-24	23.700	29.190	F
8.5.2	SL_TLF_UEM	FEMALE_15+	6.200	7.230	F
8.5.2	SL_TLF_UEM	FEMALE_25+	3.900	4.370	F
8.5.2	SL_TLF_UEM	BOTHSEX_15-24	18.000	22.460	F
8.5.2	SL_TLF_UEM	BOTHSEX_15+	4.100	4.730	F
8.5.2	SL_TLF_UEM	BOTHSEX_25+	2.300	2.390	F
8.4.2	EN_MAT_DOMCMPC	NFO	0.008	0.000	T
8.4.2	EN_MAT_DOMCMPC	FOF	0.339	0.310	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
8.4.2	EN_MAT_DOMCMPC	FEO	0.062	0.050	F
8.4.2	EN_MAT_DOMCMPC	WOD	0.238	0.240	F
8.4.2	EN_MAT_DOMCMPC	COL	0.106	0.060	T
8.4.2	EN_MAT_DOMCMPC	CRO	0.585	0.570	F
8.4.2	EN_MAT_DOMCMPC	CPR	0.571	0.550	F
8.4.2	EN_MAT_DOMCMPC	NMM	3.653	3.130	F
8.4.2	EN_MAT_DOMCMPC	BIM	1.531	1.500	F
8.4.2	EN_MAT_DOMCMPC	GBO	0.105	0.110	F
8.4.2	EN_MAT_DOMCMPC	PET	0.233	0.220	F
8.4.2	EN_MAT_DOMCMPC	ALP	5.592	5.090	F
8.4.2	EN_MAT_DOMCMPC	NMC	3.618	5.310	F
8.4.2	EN_MAT_DOMCMPC	MEO	0.070	0.060	F
8.4.2	EN_MAT_DOMCMPC	NMA	0.035	0.030	F
8.4.2	EN_MAT_DOMCMPC	WCH	0.032	0.030	F
8.10.1	FB_CBK_BRCH	15+_A	18.631	17.670	F
8.8.1	SL_EMP_FTLINJUR	MALE__T	1.200	1.640	F
8.8.1	SL_EMP_FTLINJUR	FEMALE__T	0.100	0.150	F
8.8.1	SL_EMP_FTLINJUR	BOTHSEX__T	0.800	1.360	F
8.4.2	EN_MAT_DOMCMPT	NFO	158953.000	99704.930	F
8.4.2	EN_MAT_DOMCMPT	FOF	7080750.000	5896192.000	F
8.4.2	EN_MAT_DOMCMPT	FEO	1299972.000	1036825.630	F
8.4.2	EN_MAT_DOMCMPT	WOD	4974538.000	4889493.320	F
8.4.2	EN_MAT_DOMCMPT	COL	2215964.000	1592646.600	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
8.4.2	EN_MAT_DOMCMPT	CRO	12220050.000	11971871.030	F
8.4.2	EN_MAT_DOMCMPT	CPR	11916566.000	11590355.790	F
8.4.2	EN_MAT_DOMCMPT	NMM	76256864.000	65233780.580	F
8.4.2	EN_MAT_DOMCMPT	BIM	31957230.000	31716814.260	F
8.4.2	EN_MAT_DOMCMPT	GBO	2187862.000	2322538.540	F
8.4.2	EN_MAT_DOMCMPT	PET	4864786.000	4639052.880	F
8.4.2	EN_MAT_DOMCMPT	ALP	116753769.700	104835724.670	F
8.4.2	EN_MAT_DOMCMPT	NMC	75525674.000	65309779.030	F
8.4.2	EN_MAT_DOMCMPT	MEO	1458925.000	1130305.440	F
8.4.2	EN_MAT_DOMCMPT	NMA	731191.000	699094.690	F
8.4.2	EN_MAT_DOMCMPT	WCH	658214.000	612513.530	F
8.4.2	EN_MAT_DOMCMPG	2010.0_NFO	0.002	0.000	T
8.4.2	EN_MAT_DOMCMPG	2010.0_FOF	0.086	0.090	F
8.4.2	EN_MAT_DOMCMPG	2010.0_FEO	0.016	0.020	F
8.4.2	EN_MAT_DOMCMPG	2010.0_WOD	0.060	0.040	F
8.4.2	EN_MAT_DOMCMPG	2010.0_COL	0.027	0.090	F
8.4.2	EN_MAT_DOMCMPG	2010.0_CRO	0.148	0.180	F
8.4.2	EN_MAT_DOMCMPG	2010.0_BIM	0.387	0.480	F
8.4.2	EN_MAT_DOMCMPG	2010.0_NMM	0.924	0.930	F
8.4.2	EN_MAT_DOMCMPG	2010.0_CPR	0.144	0.170	F
8.4.2	EN_MAT_DOMCMPG	2010.0_GBO	0.027	0.040	F
8.4.2	EN_MAT_DOMCMPG	2010.0_PET	0.059	0.060	F
8.4.2	EN_MAT_DOMCMPG	2010.0_MEO	0.018	0.020	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
8.4.2	EN_MAT_DOMCMPG	2010.0_NMC	0.915	0.920	F
8.4.2	EN_MAT_DOMCMPG	2010.0_ALP	1.414	1.440	F
8.4.2	EN_MAT_DOMCMPG	2010.0_NMA	0.009	0.010	F
8.4.2	EN_MAT_DOMCMPG	2010.0_WCH	0.008	0.010	F
8.5.1	SL_EMP_EARN	isco08-2_MALE	308.500	323.510	F
8.5.1	SL_EMP_EARN	isco08-6_MALE	88.100	60.460	F
8.5.1	SL_EMP_EARN	isco08-3_MALE	201.100	228.210	F
8.5.1	SL_EMP_EARN	isco08-7_MALE	136.200	140.950	F
8.5.1	SL_EMP_EARN	isco08-0_MALE	199.300	216.190	F
8.5.1	SL_EMP_EARN	isco08-8_MALE	132.500	149.070	F
8.5.1	SL_EMP_EARN	isco08-1_MALE	321.000	226.760	F
8.5.1	SL_EMP_EARN	isco08-9_MALE	99.700	93.660	F
8.5.1	SL_EMP_EARN	isco08_MALE	151.500	127.060	F
8.5.1	SL_EMP_EARN	isco08-4_MALE	174.700	210.070	F
8.5.1	SL_EMP_EARN	isco08-5_MALE	125.400	84.560	F
8.5.1	SL_EMP_EARN	isco08-2_FEMALE	238.100	257.020	F
8.5.1	SL_EMP_EARN	isco08-6_FEMALE	78.300	54.670	F
8.5.1	SL_EMP_EARN	isco08-3_FEMALE	183.700	175.560	F
8.5.1	SL_EMP_EARN	isco08-7_FEMALE	79.400	57.250	F
8.5.1	SL_EMP_EARN	isco08-8_FEMALE	89.800	138.400	T
8.5.1	SL_EMP_EARN	isco08-1_FEMALE	310.000	-89.460	F
8.5.1	SL_EMP_EARN	isco08-9_FEMALE	71.700	60.530	F
8.5.1	SL_EMP_EARN	isco08_FEMALE	138.800	131.870	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
8.5.1	SL_EMP_EARN	isco08-4_FEMALE	157.000	116.390	F
8.5.1	SL_EMP_EARN	isco08-5_FEMALE	111.900	100.080	F
8.5.1	SL_EMP_EARN	isco08-2_BOTHSEX	262.900	290.380	F
8.5.1	SL_EMP_EARN	isco08-6_BOTHSEX	86.300	74.430	F
8.5.1	SL_EMP_EARN	isco08-3_BOTHSEX	194.500	221.300	F
8.5.1	SL_EMP_EARN	isco08-7_BOTHSEX	119.200	111.530	F
8.5.1	SL_EMP_EARN	isco08-0_BOTHSEX	197.600	198.710	F
8.5.1	SL_EMP_EARN	isco08-8_BOTHSEX	123.600	141.660	F
8.5.1	SL_EMP_EARN	isco08-1_BOTHSEX	318.200	595.900	T
8.5.1	SL_EMP_EARN	isco08-9_BOTHSEX	89.800	65.030	F
8.5.1	SL_EMP_EARN	isco08_BOTHSEX	146.900	120.570	F
8.5.1	SL_EMP_EARN	isco08-4_BOTHSEX	165.300	176.370	F
8.5.1	SL_EMP_EARN	isco08-5_BOTHSEX	122.300	95.320	F
8.6.1	SL_TLF_NEET	15-24_MALE	16.800	16.430	F
8.6.1	SL_TLF_NEET	FEMALE_15-24	32.200	31.750	F
8.6.1	SL_TLF_NEET	BOTHSEX_15-24	24.700	24.380	F
9.2.1	NV_IND_MANF	2015	16.130	16.700	F
9.4.1	EN_ATM_CO2	ISIC4_C10T32X19	0.704	1.160	F
9.4.1	EN_ATM_CO2	TOTAL	22.285	15.960	F
9.2.1	NV_IND_MANFPC	2015	665.900	598.960	F
9.a.1	DC_TOF_INFRAL	2018	481.370	514.880	U
9.1.2	IS_RDP_PORFVOL	SEA	6200000.000	4624244.480	F
9.4.1	EN_ATM_CO2MVA	2015	0.050	0.090	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
9.4.1	EN_ATM_CO2GDP	2017	0.083	0.080	F
10.6.1	SG_INT_VRTDEV	IBRD	0.248	0.240	F
10.6.1	SG_INT_VRTDEV	WTO	0.610	0.630	F
10.6.1	SG_INT_VRTDEV	IMF	0.144	0.180	F
10.6.1	SG_INT_VRTDEV	ADB	0.762	0.780	F
10.6.1	SG_INT_VRTDEV	UNGA	0.518	0.520	F
10.6.1	SG_INT_VRTDEV	IFC	0.306	0.310	F
10.5.1	FI_FSI_FSLS	A	28.559	67.560	T
10.a.1	TM_TRF_ZERO	AGR	49.911	40.410	F
10.a.1	TM_TRF_ZERO	TEX	29.685	34.540	F
10.a.1	TM_TRF_ZERO	CLO	14.814	19.450	F
10.a.1	TM_TRF_ZERO	ALP	38.945	34.740	F
10.a.1	TM_TRF_ZERO	IND	49.268	40.270	F
10.6.1	SG_INT_MBRDEV	IBRD	0.529	0.530	F
10.6.1	SG_INT_MBRDEV	WTO	0.610	0.640	F
10.6.1	SG_INT_MBRDEV	IMF	0.529	0.540	F
10.6.1	SG_INT_MBRDEV	ADB	1.493	1.500	F
10.6.1	SG_INT_MBRDEV	UNGA	0.518	0.520	F
10.6.1	SG_INT_MBRDEV	IFC	0.543	0.550	F
10.5.1	FI_FSI_FSKRTC	A	12.356	13.510	F
10.5.1	FI_FSI_FSKA	A	8.423	9.040	F
10.5.1	FI_FSI_FSANL	A	2.498	4.360	T
10.5.1	FI_FSI_FSKNL	A	5.587	26.710	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
10.5.1	FI_FSI_FSERA	A	1.992	2.080	F
11.6.2	EN_ATM_PM25	ALLAREA	16.536	13.550	F
11.6.2	EN_ATM_PM25	RURAL	16.334	12.880	F
11.6.2	EN_ATM_PM25	URBAN	16.769	13.690	F
12.2.2	EN_MAT_DOMCMPC	NFO	0.008	0.000	T
12.2.2	EN_MAT_DOMCMPC	FOF	0.339	0.280	F
12.2.2	EN_MAT_DOMCMPC	FEO	0.062	0.050	F
12.2.2	EN_MAT_DOMCMPC	WOD	0.238	0.240	F
12.2.2	EN_MAT_DOMCMPC	COL	0.106	0.160	F
12.2.2	EN_MAT_DOMCMPC	CRO	0.585	0.570	F
12.2.2	EN_MAT_DOMCMPC	CPR	0.571	0.550	F
12.2.2	EN_MAT_DOMCMPC	NMM	3.653	3.160	F
12.2.2	EN_MAT_DOMCMPC	BIM	1.531	1.500	F
12.2.2	EN_MAT_DOMCMPC	GBO	0.105	0.110	F
12.2.2	EN_MAT_DOMCMPC	PET	0.233	0.230	F
12.2.2	EN_MAT_DOMCMPC	ALP	5.592	5.140	F
12.2.2	EN_MAT_DOMCMPC	NMC	3.618	3.170	F
12.2.2	EN_MAT_DOMCMPC	MEO	0.070	0.060	F
12.2.2	EN_MAT_DOMCMPC	NMA	0.035	0.030	F
12.2.2	EN_MAT_DOMCMPC	WCH	0.032	0.030	F
12.2.2	EN_MAT_DOMCMPT	NFO	158953.000	91124.840	F
12.2.2	EN_MAT_DOMCMPT	FOF	7080750.000	5822251.460	F
12.2.2	EN_MAT_DOMCMPT	FEO	1299972.000	1053212.960	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
12.2.2	EN_MAT_DOMCMPT	WOD	4974538.000	4906839.070	F
12.2.2	EN_MAT_DOMCMPT	COL	2215964.000	1452115.930	F
12.2.2	EN_MAT_DOMCMPT	CRO	12220050.000	12118418.710	F
12.2.2	EN_MAT_DOMCMPT	CPR	11916566.000	11649042.230	F
12.2.2	EN_MAT_DOMCMPT	NMM	76256864.000	65757833.540	F
12.2.2	EN_MAT_DOMCMPT	BIM	31957230.000	31715009.460	F
12.2.2	EN_MAT_DOMCMPT	GBO	2187862.000	2306826.490	F
12.2.2	EN_MAT_DOMCMPT	PET	4864786.000	4650471.940	F
12.2.2	EN_MAT_DOMCMPT	ALP	116753769.700	103174832.750	F
12.2.2	EN_MAT_DOMCMPT	NMC	75525674.000	71300676.600	F
12.2.2	EN_MAT_DOMCMPT	MEO	1458925.000	1141217.920	F
12.2.2	EN_MAT_DOMCMPT	NMA	731191.000	710906.140	F
12.2.2	EN_MAT_DOMCMPT	WCH	658214.000	622366.500	F
12.2.2	EN_MAT_DOMCMPG	2010.0_NFO	0.002	0.000	T
12.2.2	EN_MAT_DOMCMPG	2010.0_FOF	0.086	0.090	F
12.2.2	EN_MAT_DOMCMPG	2010.0_FEO	0.016	0.020	F
12.2.2	EN_MAT_DOMCMPG	2010.0_WOD	0.060	0.070	F
12.2.2	EN_MAT_DOMCMPG	2010.0_COL	0.027	0.020	F
12.2.2	EN_MAT_DOMCMPG	2010.0_CRO	0.148	0.180	F
12.2.2	EN_MAT_DOMCMPG	2010.0_BIM	0.387	0.470	F
12.2.2	EN_MAT_DOMCMPG	2010.0_NMM	0.924	0.930	F
12.2.2	EN_MAT_DOMCMPG	2010.0_CPR	0.144	0.090	F
12.2.2	EN_MAT_DOMCMPG	2010.0_GBO	0.027	0.040	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
12.2.2	EN_MAT_DOMCMPG	2010.0_PET	0.059	-0.080	T
12.2.2	EN_MAT_DOMCMPG	2010.0_MEO	0.018	0.020	F
12.2.2	EN_MAT_DOMCMPG	2010.0_NMC	0.915	0.950	F
12.2.2	EN_MAT_DOMCMPG	2010.0_ALP	1.414	1.450	F
12.2.2	EN_MAT_DOMCMPG	2010.0_NMA	0.009	0.010	F
12.2.2	EN_MAT_DOMCMPG	2010.0_WCH	0.008	0.010	F
14.1.1	EN_MAR_CHLDEV	ANNUAL_ALLAREA_isco08_ TOTAL_ALLAGE_BOTHSEX ALP	4.380	5.560	F
15.a.1	DC_ODA_BDVL	2018	7.250	67.610	T
15.b.1	DC_ODA_BDVL	2018	7.250	205.340	T
16.8.1	SG_INT_VRTDEV	IBRD	0.248	0.240	F
16.8.1	SG_INT_VRTDEV	WTO	0.610	0.630	F
16.8.1	SG_INT_VRTDEV	IMF	0.144	0.180	F
16.8.1	SG_INT_VRTDEV	ADB	0.762	0.780	F
16.8.1	SG_INT_VRTDEV	UNGA	0.518	0.520	F
16.8.1	SG_INT_VRTDEV	IFC	0.306	0.310	F
16.1.1	VC_IHR_PSRC	MALE	3.928	5.020	F
16.1.1	VC_IHR_PSRC	FEMALE	0.793	1.350	F
16.8.1	SG_INT_MBRDEV	IBRD	0.529	0.530	F
16.8.1	SG_INT_MBRDEV	WTO	0.610	0.640	F
16.8.1	SG_INT_MBRDEV	IMF	0.529	0.540	F
16.8.1	SG_INT_MBRDEV	ADB	1.493	1.510	F

Table D.1 continued from previous page

G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
16.8.1	SG_INT_MBRDEV	UNGA	0.518	0.520	F
16.8.1	SG_INT_MBRDEV	IFC	0.543	0.550	F
16.1.1	VC_IHR_PSRCN	MALE	399.000	482.620	F
16.1.1	VC_IHR_PSRCN	FEMALE	87.000	203.160	F
17.12.1	TM_TAX_DMFN	AGR	4.285	4.340	U
17.12.1	TM_TAX_DMFN	TEX	7.405	7.410	U
17.12.1	TM_TAX_DMFN	CLO	11.263	11.510	U
17.12.1	TM_TAX_DMFN	ALP	5.105	5.130	U
17.12.1	TM_TAX_DMFN	IND	3.026	3.020	U
17.1.2	GC_GOB_TAXD	A	65.088	77.240	F
17.10.1	TM_TAX_WMFN	AGR	32.250	23.980	U
17.10.1	TM_TAX_WMFN	TEX	0.830	2.000	U
17.10.1	TM_TAX_WMFN	CLO	0.000	11.050	U
17.10.1	TM_TAX_WMFN	IND	6.820	8.630	U
17.10.1	TM_TAX_WMFN	ALP	9.180	9.710	U
17.12.1	TM_TAX_DPRF	AGR	3.484	3.350	U
17.12.1	TM_TAX_DPRF	TEX	5.796	6.760	U
17.12.1	TM_TAX_DPRF	CLO	10.478	8.990	U
17.12.1	TM_TAX_DPRF	IND	1.615	1.360	U
17.12.1	TM_TAX_DPRF	ALP	3.853	9.990	U
17.1.1	GR_G14_GDP	A	13.802	13.270	F
17.6.1	IT_NET_BBND	10MBPS	2.040	7.150	T
17.6.1	IT_NET_BBND	ANYS	5.777	3.160	F

Table D.1 continued from previous page

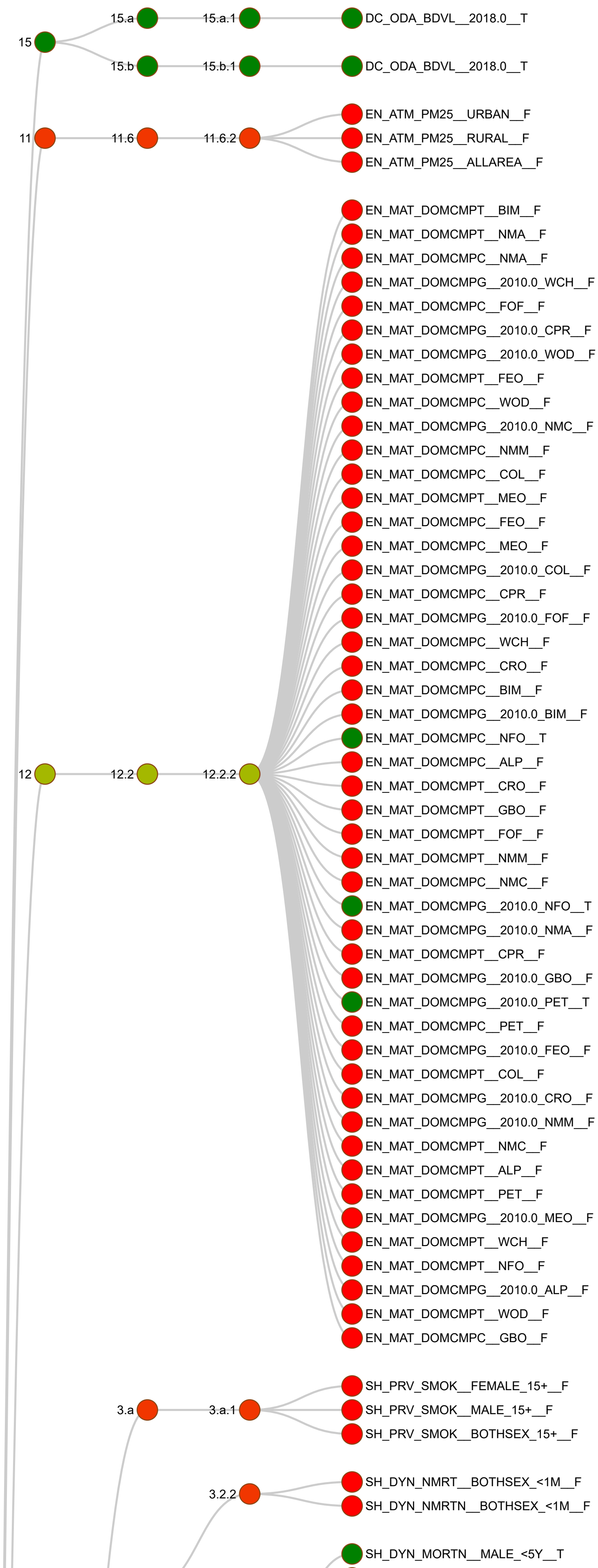
G.T.I	Series Code	Meta data	Initial Value	Prediction	Result
17.6.1	IT_NET_BBND	2MT10MBPS	2.743	41.120	T
17.17.1	GF_COM_PPPI	17.17.1_2019.0	234.673	72.940	F
17.6.1	IT_NET_BBNDN	10MBPS	431044.000	1049732.270	T
17.6.1	IT_NET_BBNDN	ANYS	1220504.000	9069905.960	T
17.6.1	IT_NET_BBNDN	2MT10MBPS	579531.000	352225.360	F
17.1.1	GR_G14_XDC	A	1839561.852	1257338.450	F
17.9.1	DC_FTA_TOTAL	2018	131.324	65.040	F
17.10.1	TM_TAX_WMPS	AGR	31.770	22.420	U
17.10.1	TM_TAX_WMPS	TEX	0.650	2.940	U
17.10.1	TM_TAX_WMPS	CLO	0.000	-4.390	U
17.10.1	TM_TAX_WMPS	ALP	8.730	8.970	U
17.10.1	TM_TAX_WMPS	IND	6.310	8.090	U

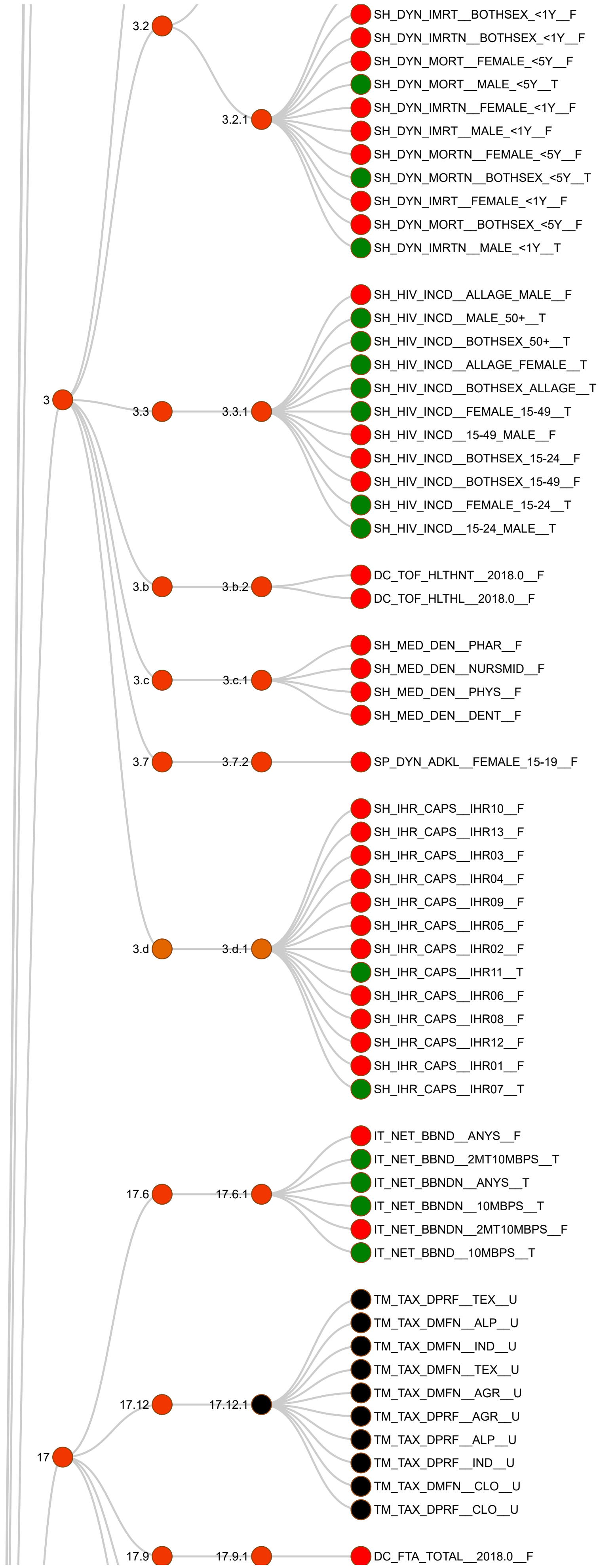
Table D.1: SDG_TTF Seri Lanks

Appendix E

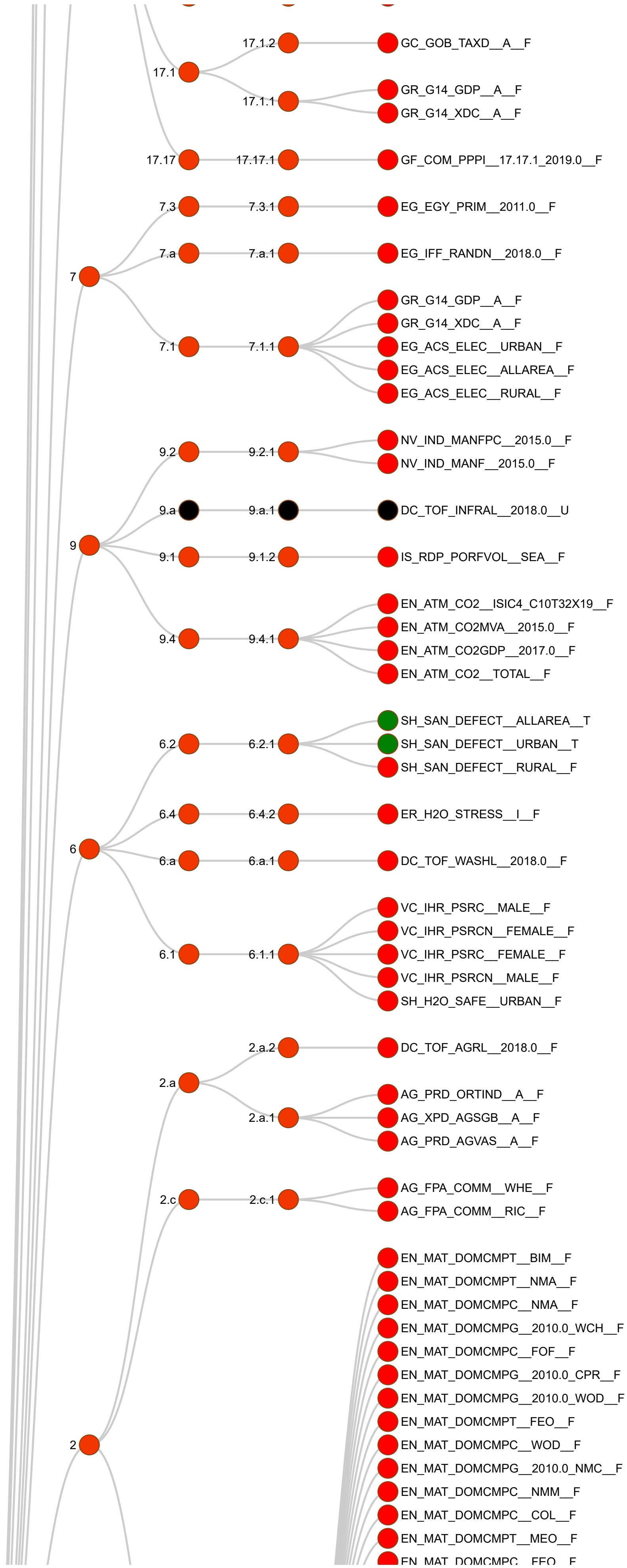
Seri Lanks D3.js Visualisation

The following figure gives the full Sri Lankan D3.js visualisation of the SDG attainment prediction results given in Appendix app:appendixE, obtained using the SDG-TTF framework presented in Chapter 7.

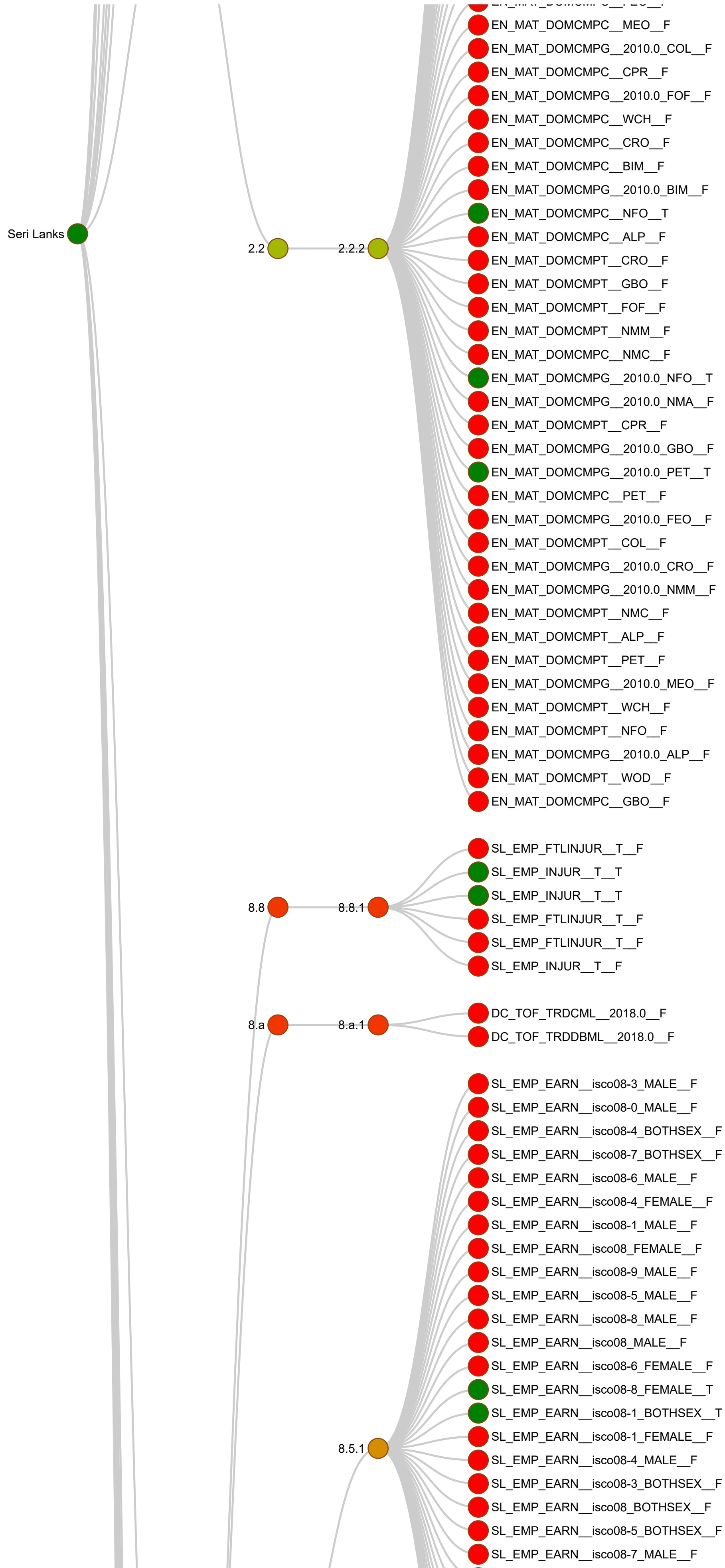

15
15.a
15.a.1
DC_ODA_BDVL__2018.0__T
15.b
15.b.1
DC_ODA_BDVL__2018.0__T
11
11.6
11.6.2
EN_ATM_PM25__URBAN__F
EN_ATM_PM25__RURAL__F
EN_ATM_PM25__ALLAREA__F
12
12.2
12.2.2
EN_MAT_DOMCMPT__BIM__F
EN_MAT_DOMCMPT__NMA__F
EN_MAT_DOMCMPC__NMA__F
EN_MAT_DOMCMPG__2010.0_WCH__F
EN_MAT_DOMCMPC__FOF__F
EN_MAT_DOMCMPG__2010.0_CPR__F
EN_MAT_DOMCMPG__2010.0_WOD__F
EN_MAT_DOMCMPT__FEO__F
EN_MAT_DOMCMPC__WOD__F
EN_MAT_DOMCMPG__2010.0_NMC__F
EN_MAT_DOMCMPC__NMM__F
EN_MAT_DOMCMPC__COL__F
EN_MAT_DOMCMPT__MEO__F
EN_MAT_DOMCMPC__FEO__F
EN_MAT_DOMCMPC__MEO__F
EN_MAT_DOMCMPG__2010.0_COL__F
EN_MAT_DOMCMPC__CPR__F
EN_MAT_DOMCMPG__2010.0_FOF__F
EN_MAT_DOMCMPC__WCH__F
EN_MAT_DOMCMPC__CRO__F
EN_MAT_DOMCMPC__BIM__F
EN_MAT_DOMCMPG__2010.0_BIM__F
EN_MAT_DOMCMPC__NFO__T
EN_MAT_DOMCMPC__ALP__F
EN_MAT_DOMCMPT__CRO__F
EN_MAT_DOMCMPT__GBO__F
EN_MAT_DOMCMPT__FOF__F
EN_MAT_DOMCMPT__NMM__F
EN_MAT_DOMCMPC__NMC__F
EN_MAT_DOMCMPG__2010.0_NFO__T
EN_MAT_DOMCMPG__2010.0_NMA__F
EN_MAT_DOMCMPT__CPR__F
EN_MAT_DOMCMPG__2010.0_GBO__F
EN_MAT_DOMCMPG__2010.0_PET__T
EN_MAT_DOMCMPC__PET__F
EN_MAT_DOMCMPG__2010.0_FEO__F
EN_MAT_DOMCMPT__COL__F
EN_MAT_DOMCMPG__2010.0_CRO__F
EN_MAT_DOMCMPG__2010.0_NMM__F
EN_MAT_DOMCMPT__NMC__F
EN_MAT_DOMCMPT__ALP__F
EN_MAT_DOMCMPT__PET__F
EN_MAT_DOMCMPG__2010.0_MEO__F
EN_MAT_DOMCMPT__WCH__F
EN_MAT_DOMCMPT__NFO__F
EN_MAT_DOMCMPG__2010.0_ALP__F
EN_MAT_DOMCMPT__WOD__F
EN_MAT_DOMCMPC__GBO__F
3.a
3.a.1
SH_PRV_SMOK__FEMALE_15+__F
SH_PRV_SMOK__MALE_15+__F
SH_PRV_SMOK__BOTHSEX_15+__F
3.2.2
SH_DYN_NMRT__BOTHSEX_<1M__F
SH_DYN_NMRTN__BOTHSEX_<1M__F
SH_DYN_MORTN__MALE_<5Y__T

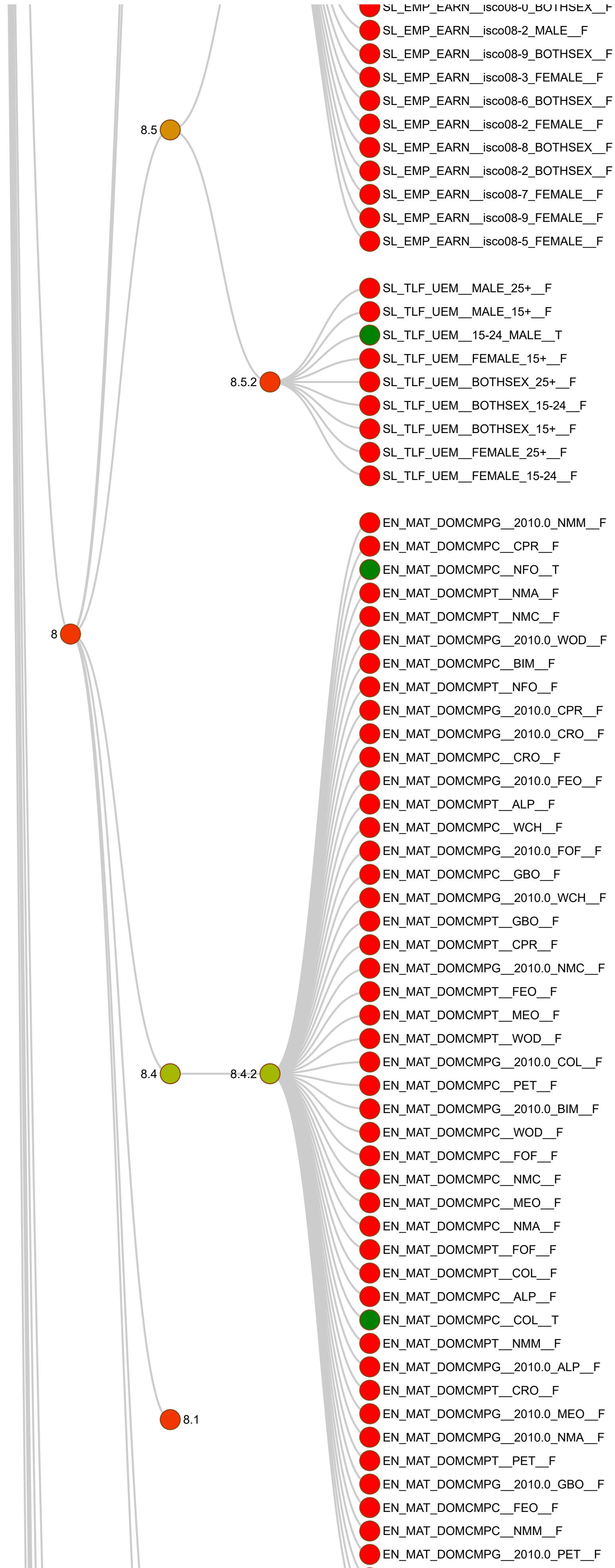


3
3.2
3.2.1
SH_DYN_IMRT__BOTHSEX_<1Y__F
SH_DYN_IMRTN__BOTHSEX_<1Y__F
SH_DYN_MORT__FEMALE_<5Y__F
SH_DYN_MORT__MALE_<5Y__T
SH_DYN_IMRTN__FEMALE_<1Y__F
SH_DYN_IMRT__MALE_<1Y__F
SH_DYN_MORTN__FEMALE_<5Y__F
SH_DYN_MORTN__BOTHSEX_<5Y__T
SH_DYN_IMRT__FEMALE_<1Y__F
SH_DYN_MORT__BOTHSEX_<5Y__F
SH_DYN_IMRTN__MALE_<1Y__T
3.3
3.3.1
SH_HIV_INCD__ALLAGE_MALE__F
SH_HIV_INCD__MALE_50+__T
SH_HIV_INCD__BOTHSEX_50+__T
SH_HIV_INCD__ALLAGE_FEMALE__T
SH_HIV_INCD__BOTHSEX_ALLAGE__T
SH_HIV_INCD__FEMALE_15-49__T
SH_HIV_INCD__15-49_MALE__F
SH_HIV_INCD__BOTHSEX_15-24__F
SH_HIV_INCD__BOTHSEX_15-49__F
SH_HIV_INCD__FEMALE_15-24__T
SH_HIV_INCD__15-24_MALE__T
3.b
3.b.2
DC_TOF_HLTHNT__2018.0__F
DC_TOF_HLTHL__2018.0__F
3.c
3.c.1
SH_MED_DEN__PHAR__F
SH_MED_DEN__NURSMID__F
SH_MED_DEN__PHYS__F
SH_MED_DEN__DENT__F
3.7
3.7.2
SP_DYN_ADKL__FEMALE_15-19__F
3.d
3.d.1
SH_IHR_CAPS__IHR10__F
SH_IHR_CAPS__IHR13__F
SH_IHR_CAPS__IHR03__F
SH_IHR_CAPS__IHR04__F
SH_IHR_CAPS__IHR09__F
SH_IHR_CAPS__IHR05__F
SH_IHR_CAPS__IHR02__F
SH_IHR_CAPS__IHR11__T
SH_IHR_CAPS__IHR06__F
SH_IHR_CAPS__IHR08__F
SH_IHR_CAPS__IHR12__F
SH_IHR_CAPS__IHR01__F
SH_IHR_CAPS__IHR07__T
17
17.6
17.6.1
IT_NET_BBND__ANYS__F
IT_NET_BBND__2MT10MBPS__T
IT_NET_BBNDN__ANYS__T
IT_NET_BBNDN__10MBPS__T
IT_NET_BBNDN__2MT10MBPS__F
IT_NET_BBND__10MBPS__T
17.12
17.12.1
TM_TAX_DPRF__TEX__U
TM_TAX_DMFN__ALP__U
TM_TAX_DMFN__IND__U
TM_TAX_DMFN__TEX__U
TM_TAX_DMFN__AGR__U
TM_TAX_DPRF__AGR__U
TM_TAX_DPRF__ALP__U
TM_TAX_DPRF__IND__U
TM_TAX_DMFN__CLO__U
TM_TAX_DPRF__CLO__U
17.9
17.9.1
DC_FTA_TOTAL__2018.0__F

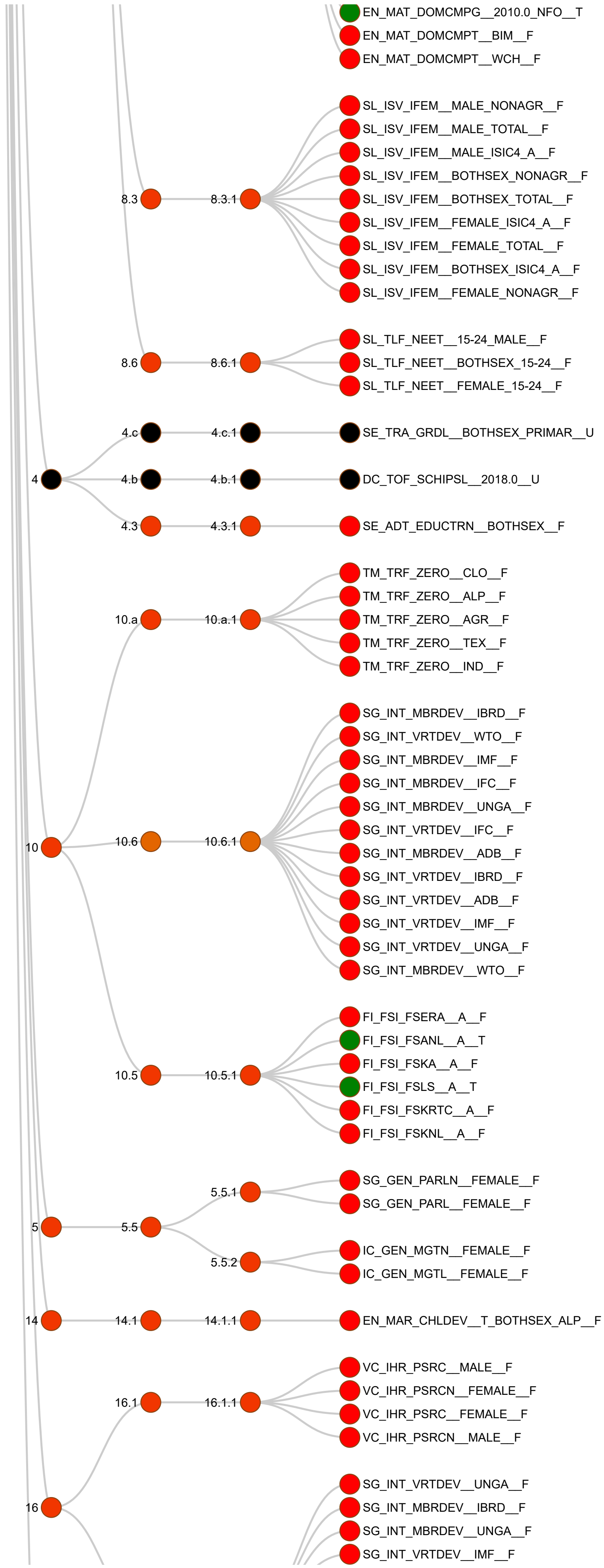


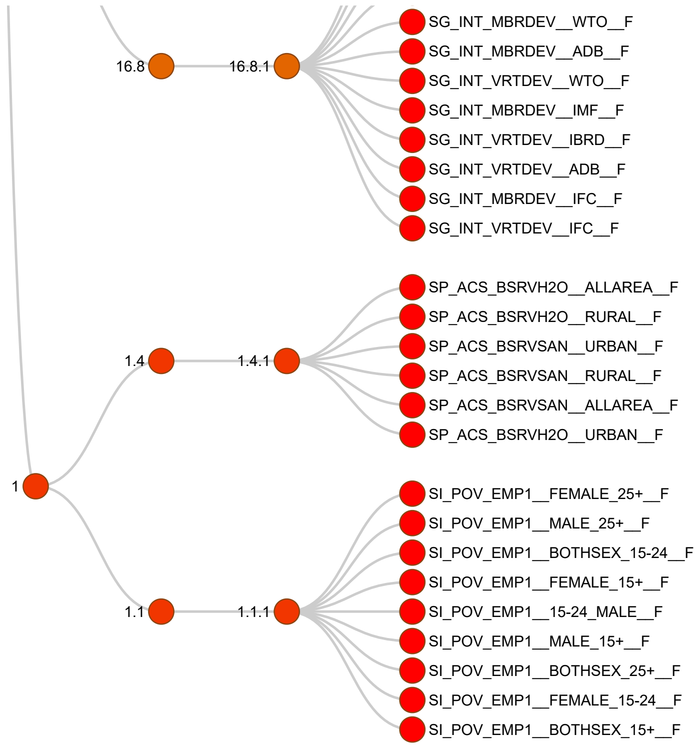
17.1.2
GC_GOB_TAXD__A__F
17.1
17.1.1
GR_G14_GDP__A__F
GR_G14_XDC__A__F
17.17
17.17.1
GF_COM_PPPI__17.17.1_2019.0__F
7.3
7.3.1
EG_EGY_PRIM__2011.0__F
7.a
7.a.1
EG_IFF_RANDN__2018.0__F
7
GR_G14_GDP__A__F
GR_G14_XDC__A__F
7.1
7.1.1
EG_ACS_ELEC__URBAN__F
EG_ACS_ELEC__ALLAREA__F
EG_ACS_ELEC__RURAL__F
9.2
9.2.1
NV_IND_MANFPC__2015.0__F
NV_IND_MANF__2015.0__F
9.a
9.a.1
DC_TOF_INFRAL__2018.0__U
9
9.1
9.1.2
IS_RDP_PORFVOL__SEA__F
EN_ATM_CO2__ISIC4_C10T32X19__F
9.4
9.4.1
EN_ATM_CO2MVA__2015.0__F
EN_ATM_CO2GDP__2017.0__F
EN_ATM_CO2__TOTAL__F
SH_SAN_DEFECT__ALLAREA__T
6.2
6.2.1
SH_SAN_DEFECT__URBAN__T
SH_SAN_DEFECT__RURAL__F
6.4
6.4.2
ER_H2O_STRESS__I__F
6
6.a
6.a.1
DC_TOF_WASHL__2018.0__F
VC_IHR_PSRC__MALE__F
VC_IHR_PSRCN__FEMALE__F
6.1
6.1.1
VC_IHR_PSRC__FEMALE__F
VC_IHR_PSRCN__MALE__F
SH_H2O_SAFE__URBAN__F
2.a.2
DC_TOF_AGRL__2018.0__F
2.a
AG_PRD_ORTIND__A__F
2.a.1
AG_XPD_AGSGB__A__F
AG_PRD_AGVAS__A__F
2.c
2.c.1
AG_FPA_COMM__WHE__F
AG_FPA_COMM__RIC__F
EN_MAT_DOMCMPT__BIM__F
EN_MAT_DOMCMPT__NMA__F
EN_MAT_DOMCMPC__NMA__F
EN_MAT_DOMCMPG__2010.0_WCH__F
EN_MAT_DOMCMPC__FOF__F
EN_MAT_DOMCMPG__2010.0_CPR__F
EN_MAT_DOMCMPG__2010.0_WOD__F
EN_MAT_DOMCMPT__FEO__F
2
EN_MAT_DOMCMPC__WOD__F
EN_MAT_DOMCMPG__2010.0_NMC__F
EN_MAT_DOMCMPC__NMM__F
EN_MAT_DOMCMPC__COL__F
EN_MAT_DOMCMPT__MEO__F


Seri Lanks
2.2
2.2.2
EN_MAT_DOMCMPC__MEO__F
EN_MAT_DOMCMPG__2010.0_COL__F
EN_MAT_DOMCMPC__CPR__F
EN_MAT_DOMCMPG__2010.0_FOF__F
EN_MAT_DOMCMPC__WCH__F
EN_MAT_DOMCMPC__CRO__F
EN_MAT_DOMCMPC__BIM__F
EN_MAT_DOMCMPG__2010.0_BIM__F
EN_MAT_DOMCMPC__NFO__T
EN_MAT_DOMCMPC__ALP__F
EN_MAT_DOMCMPT__CRO__F
EN_MAT_DOMCMPT__GBO__F
EN_MAT_DOMCMPT__FOF__F
EN_MAT_DOMCMPT__NMM__F
EN_MAT_DOMCMPC__NMC__F
EN_MAT_DOMCMPG__2010.0_NFO__T
EN_MAT_DOMCMPG__2010.0_NMA__F
EN_MAT_DOMCMPT__CPR__F
EN_MAT_DOMCMPG__2010.0_GBO__F
EN_MAT_DOMCMPG__2010.0_PET__T
EN_MAT_DOMCMPC__PET__F
EN_MAT_DOMCMPG__2010.0_FEO__F
EN_MAT_DOMCMPT__COL__F
EN_MAT_DOMCMPG__2010.0_CRO__F
EN_MAT_DOMCMPG__2010.0_NMM__F
EN_MAT_DOMCMPT__NMC__F
EN_MAT_DOMCMPT__ALP__F
EN_MAT_DOMCMPT__PET__F
EN_MAT_DOMCMPG__2010.0_MEO__F
EN_MAT_DOMCMPT__WCH__F
EN_MAT_DOMCMPT__NFO__F
EN_MAT_DOMCMPG__2010.0_ALP__F
EN_MAT_DOMCMPT__WOD__F
EN_MAT_DOMCMPC__GBO__F
8.8
8.8.1
SL_EMP_FTLINJUR__T__F
SL_EMP_INJUR__T__T
SL_EMP_INJUR__T__T
SL_EMP_FTLINJUR__T__F
SL_EMP_FTLINJUR__T__F
SL_EMP_INJUR__T__F
8.a
8.a.1
DC_TOF_TRDCML__2018.0__F
DC_TOF_TRDDBML__2018.0__F
8.5.1
SL_EMP_EARN__isco08-3_MALE__F
SL_EMP_EARN__isco08-0_MALE__F
SL_EMP_EARN__isco08-4_BOTHSEX__F
SL_EMP_EARN__isco08-7_BOTHSEX__F
SL_EMP_EARN__isco08-6_MALE__F
SL_EMP_EARN__isco08-4_FEMALE__F
SL_EMP_EARN__isco08-1_MALE__F
SL_EMP_EARN__isco08_FEMALE__F
SL_EMP_EARN__isco08-9_MALE__F
SL_EMP_EARN__isco08-5_MALE__F
SL_EMP_EARN__isco08-8_MALE__F
SL_EMP_EARN__isco08_MALE__F
SL_EMP_EARN__isco08-6_FEMALE__F
SL_EMP_EARN__isco08-8_FEMALE__T
SL_EMP_EARN__isco08-1_BOTHSEX__T
SL_EMP_EARN__isco08-1_FEMALE__F
SL_EMP_EARN__isco08-4_MALE__F
SL_EMP_EARN__isco08-3_BOTHSEX__F
SL_EMP_EARN__isco08_BOTHSEX__F
SL_EMP_EARN__isco08-5_BOTHSEX__F
SL_EMP_EARN__isco08-7_MALE__F



SL_EMP_EARN__isco08-0_BOTHSEX__F
SL_EMP_EARN__isco08-2_MALE__F
SL_EMP_EARN__isco08-9_BOTHSEX__F
SL_EMP_EARN__isco08-3_FEMALE__F
SL_EMP_EARN__isco08-6_BOTHSEX__F
SL_EMP_EARN__isco08-2_FEMALE__F
SL_EMP_EARN__isco08-8_BOTHSEX__F
SL_EMP_EARN__isco08-2_BOTHSEX__F
SL_EMP_EARN__isco08-7_FEMALE__F
SL_EMP_EARN__isco08-9_FEMALE__F
SL_EMP_EARN__isco08-5_FEMALE__F
8.5
SL_TLF_UEM__MALE_25+__F
SL_TLF_UEM__MALE_15+__F
SL_TLF_UEM__15-24_MALE__T
SL_TLF_UEM__FEMALE_15+__F
SL_TLF_UEM__BOTHSEX_25+__F
SL_TLF_UEM__BOTHSEX_15-24__F
SL_TLF_UEM__BOTHSEX_15+__F
SL_TLF_UEM__FEMALE_25+__F
SL_TLF_UEM__FEMALE_15-24__F
8.5.2
EN_MAT_DOMCMPG__2010.0_NMM__F
EN_MAT_DOMCMPC__CPR__F
EN_MAT_DOMCMPC__NFO__T
EN_MAT_DOMCMPT__NMA__F
EN_MAT_DOMCMPT__NMC__F
EN_MAT_DOMCMPG__2010.0_WOD__F
EN_MAT_DOMCMPC__BIM__F
EN_MAT_DOMCMPT__NFO__F
EN_MAT_DOMCMPG__2010.0_CPR__F
EN_MAT_DOMCMPG__2010.0_CRO__F
EN_MAT_DOMCMPC__CRO__F
EN_MAT_DOMCMPG__2010.0_FEO__F
EN_MAT_DOMCMPT__ALP__F
EN_MAT_DOMCMPC__WCH__F
EN_MAT_DOMCMPG__2010.0_FOF__F
EN_MAT_DOMCMPC__GBO__F
EN_MAT_DOMCMPG__2010.0_WCH__F
EN_MAT_DOMCMPT__GBO__F
EN_MAT_DOMCMPT__CPR__F
EN_MAT_DOMCMPG__2010.0_NMC__F
EN_MAT_DOMCMPT__FEO__F
EN_MAT_DOMCMPT__MEO__F
EN_MAT_DOMCMPT__WOD__F
EN_MAT_DOMCMPG__2010.0_COL__F
EN_MAT_DOMCMPT__PET__F
EN_MAT_DOMCMPG__2010.0_BIM__F
EN_MAT_DOMCMPC__WOD__F
EN_MAT_DOMCMPC__FOF__F
EN_MAT_DOMCMPC__NMC__F
EN_MAT_DOMCMPC__MEO__F
EN_MAT_DOMCMPC__NMA__F
EN_MAT_DOMCMPT__FOF__F
EN_MAT_DOMCMPT__COL__F
EN_MAT_DOMCMPC__ALP__F
EN_MAT_DOMCMPC__COL__T
EN_MAT_DOMCMPT__NMM__F
EN_MAT_DOMCMPG__2010.0_ALP__F
EN_MAT_DOMCMPT__CRO__F
EN_MAT_DOMCMPG__2010.0_MEO__F
EN_MAT_DOMCMPG__2010.0_NMA__F
EN_MAT_DOMCMPT__PET__F
EN_MAT_DOMCMPG__2010.0_GBO__F
EN_MAT_DOMCMPC__FEO__F
EN_MAT_DOMCMPC__NMM__F
EN_MAT_DOMCMPG__2010.0_PET__F
8
8.4
8.4.2
8.1


EN_MAT_DOMCMPG__2010.0_NFO__T
EN_MAT_DOMCMPT__BIM__F
EN_MAT_DOMCMPT__WCH__F
SL_ISV_IFEM__MALE_NONAGR__F
SL_ISV_IFEM__MALE_TOTAL__F
SL_ISV_IFEM__MALE_ISIC4_A__F
SL_ISV_IFEM__BOTHSEX_NONAGR__F
8.3
8.3.1
SL_ISV_IFEM__BOTHSEX_TOTAL__F
SL_ISV_IFEM__FEMALE_ISIC4_A__F
SL_ISV_IFEM__FEMALE_TOTAL__F
SL_ISV_IFEM__BOTHSEX_ISIC4_A__F
SL_ISV_IFEM__FEMALE_NONAGR__F
SL_TLF_NEET__15-24_MALE__F
8.6
8.6.1
SL_TLF_NEET__BOTHSEX_15-24__F
SL_TLF_NEET__FEMALE_15-24__F
4.c
4.c.1
SE_TRA_GRDL__BOTHSEX_PRIMAR__U
4
4.b
4.b.1
DC_TOF_SCHIPSL__2018.0__U
4.3
4.3.1
SE_ADT_EDUCTRN__BOTHSEX__F
TM_TRF_ZERO__CLO__F
TM_TRF_ZERO__ALP__F
10.a
10.a.1
TM_TRF_ZERO__AGR__F
TM_TRF_ZERO__TEX__F
TM_TRF_ZERO__IND__F
SG_INT_MBRDEV__IBRD__F
SG_INT_VRTDEV__WTO__F
SG_INT_MBRDEV__IMF__F
SG_INT_MBRDEV__IFC__F
SG_INT_MBRDEV__UNGA__F
10
10.6
10.6.1
SG_INT_VRTDEV__IFC__F
SG_INT_MBRDEV__ADB__F
SG_INT_VRTDEV__IBRD__F
SG_INT_VRTDEV__ADB__F
SG_INT_VRTDEV__IMF__F
SG_INT_VRTDEV__UNGA__F
SG_INT_MBRDEV__WTO__F
FI_FSI_FSERA__A__F
FI_FSI_FSANL__A__T
10.5
10.5.1
FI_FSI_FSKA__A__F
FI_FSI_FSLS__A__T
FI_FSI_FSKRTC__A__F
FI_FSI_FSKNL__A__F
SG_GEN_PARLN__FEMALE__F
5.5.1
SG_GEN_PARL__FEMALE__F
5
5.5
IC_GEN_MGTN__FEMALE__F
5.5.2
IC_GEN_MGTL__FEMALE__F
14
14.1
14.1.1
EN_MAR_CHLDEV__T_BOTHSEX_ALP__F
VC_IHR_PSRC__MALE__F
VC_IHR_PSRCN__FEMALE__F
16.1
16.1.1
VC_IHR_PSRC__FEMALE__F
VC_IHR_PSRCN__MALE__F
SG_INT_VRTDEV__UNGA__F
16
SG_INT_MBRDEV__IBRD__F
SG_INT_MBRDEV__UNGA__F
SG_INT_VRTDEV__IMF__F


16.8
16.8.1
SG_INT_MBRDEV__WTO__F
SG_INT_MBRDEV__ADB__F
SG_INT_VRTDEV__WTO__F
SG_INT_MBRDEV__IMF__F
SG_INT_VRTDEV__IBRD__F
SG_INT_VRTDEV__ADB__F
SG_INT_MBRDEV__IFC__F
SG_INT_VRTDEV__IFC__F
1.4
1.4.1
SP_ACS_BSRVH2O__ALLAREA__F
SP_ACS_BSRVH2O__RURAL__F
SP_ACS_BSRVSAN__URBAN__F
SP_ACS_BSRVSAN__RURAL__F
SP_ACS_BSRVSAN__ALLAREA__F
SP_ACS_BSRVH2O__URBAN__F
1
1.1
1.1.1
SI_POV_EMP1__FEMALE_25+__F
SI_POV_EMP1__MALE_25+__F
SI_POV_EMP1__BOTHSEX_15-24__F
SI_POV_EMP1__FEMALE_15+__F
SI_POV_EMP1__15-24_MALE__F
SI_POV_EMP1__MALE_15+__F
SI_POV_EMP1__BOTHSEX_25+__F
SI_POV_EMP1__FEMALE_15-24__F
SI_POV_EMP1__BOTHSEX_15+__F