Data-driven soft sensing towards quality monitoring of industrial pasteurization processes

Gabriel Filios[1,3], Andreas Kyriakopoulos[2], Stavros Livanios[1], Fotis Manolopoulos[2], Sotiris Nikoletseas[1,3], Stefanos H. Panagiotou[1], Paul Spirakis[1,4]

[1]Computer Engineering and Informatics Department, University of Patras, Greece
Email: {filiosg, livanios, spanagiotou}@ceid.upatras.gr

[2]Heineken, Patras Plant, Greece
Email: {andreas_kyriakopoulos, fotis_manolopoulos}@heineken.com

[3]Computer Technology Institute and Press "Diophantus" (CTI), Greece
Email: {nikole}@cti.gr

[4]Department of Computer Science, University of Liverpool, UK
Email: {P.Spirakis}@liverpool.ac.uk

***Abstract*—In the food and beverage industry many foods, beers and soft drinks usually need to get pasteurized, a process that holds a significant role in the quality and taste of the final product but is difficult to monitor due to the process nature. Soft sensing techniques, also called virtual sensing or surrogate sensing, can be leveraged to monitor the product quality, by using information available from other measurements and process parameters to calculate an estimation of the quantity of interest. In this paper, we develop a soft sensing methodology that is based on machine learning algorithms for continuous, end-to-end estimation of the temperature of products during the pasteurization process, with the vision to serve as an intermediate step towards monitoring live the final quality of the pasteurized products. This work studies a real beer pasteurization process in collaboration with Heineken's plant in Patras, Greece and the results demonstrate notable performance in temperature prediction accuracy, with average root mean square error (RMSE) of 1.85°C in the test sets. Thus, we claim that it is possible to obtain measurements quite similar to the ones by the respective physical sensors with sufficient accuracy, and our methodology can be considered as a virtual low-cost solution for monitoring product quality in legacy pasteurizer operation.**

***Index Terms*—Machine Learning, industrial monitoring, virtual sensing, pasteurization process, quality assurance**

I. Introduction

A. Beer pasteurization process

In the food and beverage industry many foods, beers and soft drinks need to be pasteurized to minimize the effect of microorganisms on the physical stability and flavour of products. Pasteurization is an important procedure in which the filled cans or bottles travel through a tunnel with several thermal zones where water at different temperatures is sprayed onto the packages to control their temperature.

The tunnel has a low ceiling with spray heads at regular intervals and the bottles or cans move through the pasteurizer slowly on either a walking beam or conveyor belt. The tunnel is divided into different temperature zones to slowly bring the products up to temperature, keep them at a specified holding temperature and then bring them back down to room temperature. The slow transition in the product temperatures is deliberate in order to avoid thermal shock and damage to the products' containers. Modern tunnel pasteurizers contain sophisticated control systems to manage the temperatures, deal with line hold-ups and slowdowns in a way to prevent over or under pasteurization of the product.

The temperature monitoring inside any packed product is not feasible and in practice the operators pass a monitoring kit through the tunnel (denoted hereafter as thermograph recorder), which consists of a product with a temperature sensor installed at its cold spot (close to the bottom of the center of the can or bottle) and a device for logging the measurements. The operators must follow this procedure periodically several times each shift, to collect enough samples and to assure that the process operates under the relevant quality standards. However, due to production and time constraints it is not possible to perform live monitoring this way.

B. Soft sensing in the process industries

In the past decades researchers started to make use of the large amounts of data being measured and stored in the process industry by building predictive models based on this data. In the context of process industry, these predictive models are called Soft Sensors. This term is a combination of the words "software", because the models are usually computer programs, and "sensors", because the models are delivering similar information as their hardware counterparts [1], and nowadays it is practically a combination of data processing, data-driven modeling, and software building techniques [2]. In this way, Soft Sensors are established as a valuable alternative to the traditional means for the acquisition of hard-to-measure quality variables based on other easy-to-measure process variables, in the context of process monitoring and control in modern industrial processes [1].

The range of tasks fulfilled by Soft Sensors is broad. The original and still most dominant application area of Soft Sensors is the prediction of process variables that are normally determined at low sampling rates or through infrequent, off-line analysis only, as they are intractable to directly measure online through physical devices [1], [2], [3].

At a very high level one can distinguish two different classes of Soft Sensors, namely model-driven and data-driven. The former are traditional first principle methods that synthetize prior knowledge or experiences but the latter are more flexible and have attracted the major interest of researchers in the past years [1], [2]. In this work, we develop data-driven soft sensors based on machine learning (ML) algorithms.

C. Motivation and contribution of this work

Today, with the recent advances in new digitalization technologies, the latest pasteurization machines (as well as automation upgrading companies) integrate software and automation techniques for precise quality control and monitoring. However, there are still many industrial plants that operate with legacy pasteurizers due to the high costs of upgrading or technical expertise barriers. Shop-floor workers operating these legacy machines do not have a consistent, real time view on the product quality (product temperatures, accumulated pasteurization units), except when they manually use thermograph recorders, which is a time-consuming operation and therefore occurs only a few times per working shift. Consequently, it is not possible to monitor the quality and taste of the final product without spending manual worker hours.

Therefore the **motivation** behind this work is to tackle the lack of continuous quality monitoring in legacy pasteurizers through AI/ML and IoT technologies that informs continuously the operators about the quality of every batch of pasteurized products. In particular, our focus is on the accurate prediction of the temperatures for a) the sprayed water on the product and b) the product cold spot, inside the machine during the pasteurization process. This is a fundamental step towards quality monitoring of the pasteurization process and in particular for predicting the accumulated pasteurization units for each batch of products that enters the pasteurizer machine, but is out of the current paper's scope.

We also declare that the focus of the current paper lies on the virtual sensing flow and prototype feasibility, thus so far in our methodology we have not explored advanced modeling, such as time series forecasting techniques [4] or advanced machine learning techniques such as deep learning methods [2].

The **contributions** of this work can be summarized as follows:

- We develop a data-driven, end-to-end, soft sensing approach that utilizes machine learning models for water and product temperature estimation during tunnel pasteurization processes.
- In collaboration with Athenian Brewery S.A plant in Patras, Greece (member of the Heineken international group), we gathered requirements and validate the soft sensing approach in a real production line with beer pasteurization process.
- The generic methodology we follow makes the soft sensing approach applicable to any tunnel pasteurizer used in packaging production lines and independent to the industry domain, the legacy of systems and the produced products.
- We provide a virtual, low-cost solution for product quality monitoring with no need for equipment replacement or upgrade of the legacy pasteurization machines or manual thermograph recordings.

Roadmap of the paper. The rest of this paper is organized as follows. Section II elaborates on the related work in virtual sensing in process industries and afterwards in Section III we present the methodology of our work, by describing the dataset used in the experiments, our data processing actions and the models used to realize the proposed soft sensing approach. Section IV contains the outcomes of the experiments and finally, in V we summarize the subject of the work and report our next steps for further exploitation.

II. Related work

In this section we list relevant papers in terms of soft sensing and machine learning for temperature prediction in industrial processes. In [5], Riverol et al. demonstrate the application of soft control strategies in the thermal treatment in the dairy industry. In particular, they apply fuzzy logic and neural networks to predict pasteurization temperature in a plate heat exchanger (PHE) that is used to pasteurize milk and they conduct a limited number of real time experiments to confirm the feasibility of their approach.

Dai et al., the authors of [6], propose a hybrid modeling framework for roller kiln temperature prediction. The framework is composed from a first-principle model for each single temperature zone of the roller kiln and then a data-driven model is developed using moving window-double locally weighted kernel principal component regression. The data-driven model is used for error compensation to improve the estimation accuracy and the modeling results demonstrate that the developed hybrid prediction model can correctly estimate the roller kiln temperature.

In [7], Zhang et al. provide a comparative analysis of deep and shallow predictive techniques to predict hot metal temperature (HMT) in blast furnace ironmaking. The results demonstrate that a) shallow neural network is preferred for current time HMT prediction, b) Gaussian process regression and support vector regression are preferred for multi-step-ahead HMT predictions.

A novel nonlinear feature representation method named nonlocal and local structure preserving stacked autoencoder (NLSP-SAE) is proposed for soft sensor modeling from Liu et al. in [8]. The soft sensors are developed to predict the 90% distillate temperature of heavy naphth in industrial hydrocracking process and the experimental results indicate that the NLSP-SAE-based soft sensor outperforms other methods in terms of smaller RMSE and larger R^2.

In [9], Kabugo et al. employ data-driven soft sensors to predict syngas heating value and hot flue gas temperature. They studied methods such as multivariable linear regression, principal component regression and partial least squares regression and a nonlinear dynamic method, namely a neural network-based NARX model. The latter was able to describe the dynamic behavior of the combustion process and demonstrated better performance in the prediction of both problems.

Futhermore, Shang et al. apply a deep neural network based soft sensor for online quality prediction of the 95% cut point temperature of heavy diesel, which is the primary controlled variable in crude distillation units (CDU) [10]. The predictions made by the deep neural network match real values much better in comparison with other traditional data-driven methods. This demonstrates that the deep learning technique can extract nonlinear latent variables effectively, making the neural network a desired latent variable model.

In [11], Leon-Medina et al. present an applied deep learning temperature prediction model for a 75 MW electric arc furnace, which is used for ferronickel production. Their methodology considers two steps: a data cleaning process to increase the quality of the data, and second, a multivariate time series deep learning model to predict the temperatures in the furnace lining. Their deep learning model is a sequential one based on GRU (gated recurrent unit) which achieved an average root mean square error (RMSE) of 1.19 °C in the test set.

Sala et al. [12] propose a data-driven approach to predict the endpoint temperature at the basic oxygen furnace (BOF). Three regression models are evaluated for predicting the targets; one classical linear approach with Ridge Regression and two non-linear multivariate models based on decision trees, Random Forest and Gradient Boosted Regression Trees. The obtained results on the first set of features showed improvements over the analytical models currently used in the steel production pipeline.

Finally, Zhao et al. propose a machine learning based multidimensional soft sensor and a calibration scheme and validate them in the case of steam reforming solid oxide fuel cell (SR-SOFC) system [13]. They estimate the temporal-spatial temperature distribution (TSTD) with Multivariable Linear Regression, the central node temperature with Least Square Support Vector Machine and they calibrate the temperature with Stochastic Gradient Descent algorithm. The simulation results show that the proposed method can obtain the SR-SOFC stack temperature distribution in time and effectively, with an average error less than 1 K.

III. Methodology

In this section we describe in detail our proposed soft sensing approach. We first define the problem and prediction targets and then elaborate on the data collection process and preprocessing steps to prepare the data for the machine learning models. The latter are presented in the last part of this section, together with the implementation details. To make our methodology clear, we summarize the various steps of our approach into a high level representation in Figure 1 and more details for each step are provided in the following subsections.

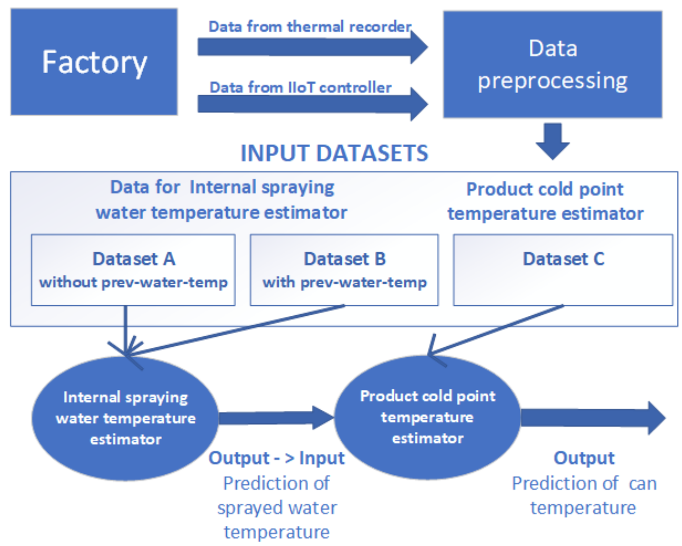


Fig. 1: Outline of our methodology

A. Problem description

In this work we develop two data-driven soft sensors for the estimation of process and product temperatures, as they are useful for monitoring the quality of pasteurized beer products in the future. We formulate the soft sensing as a machine learning regression problem, which corresponds to two ML models that solve the following continuous variables/targets (see Figure 2b), accordingly:

- Target 1: Internal spraying water temperature
- Target 2: Product cold spot temperature

The first target concerns the internal spraying water that impacts the products as they move on inside the pasteurizer during the pasteurization process. The second target concerns the cold spot of the packed products, that is located in the centre of the bottom of the been can or bottles, as seen in Figure 2.

At this point, we highlight the reasons for using a soft sensor to predict the temperature of water sprayed on the products (Target 1). The temperature of the water in the water tanks of the pasteurizer, in which PT sensors and IIoT controllers are installed to collect data, is not equal to the temperature of the water sprayed from the sprinklers that impacts the product. The water sprayed on the products is also not consistent across the whole zone because the temperature is affected as a result of the different temperatures of the previous and next zones.

B. Our proposed soft sensing approach

In this part, we detail our soft sensing approach with respect to the prediction flow of the two ML-driven soft sensors (estimators) and the pasteurization process.

At first, the output of the ML model for spraying water estimation (Target 1) is used as one of the input features for the ML model that predicts the cold spot temperature (Target 2). This input is denoted hereafter in section IV and Tables II and III, as “previous spraying water”. We use this as an input because the cold spot temperature is affected by the spraying water, and the physical thermal recorder calculates the temperature in the same way. It is thus clear that the first model needs high accuracy in the predictions in order to have a successful model for cold spot estimation. For this reason, we also feed the cold spot model with the real water temperature values to predict the cold spot temperature and use it as a baseline to assess the models predictive capacity in section IV. The values of the sprayed water are recorded using the thermograph recorder.

Furthermore, we note that to evaluate the models, we develop a simulation of an end-to-end pasteurization run for a given batch of products. Thus our approach simulates a real time, step by step prediction flow (see Figure 2a), i.e., we predict each temperature target every 10 seconds as the thermograph recorder would do so. To achieve this, we use a variable that tracks the position of the can in the pasteurizer and uses that value to collect the corresponding values from the IIoT controllers and relevant data (pasteurization program etc., see section III-C) to create the input data for the estimators. This input will be used to predict the value of the sprayed water on the can and the product cold post temperature on the next instance of the pasteurization. This prediction process takes place every 10 seconds until the pasteurization process is finished. Using such kind of simulation allows us to evaluate the result of the whole process of pasteurization and not individually predicted temperatures. When the pasteurization is completed the predicted temperatures of the sprayed water and cold spot are compared with the values recorded from the thermal recorder when passed through the pasteurizer. We repeat this procedure for all the unique pasteurization runs we use as test data. An example of the cold spot predictions can be seen on Figure 2a.

C. Data collection

The data to be collected fall on two categories, the data collected from IIoT controllers on the pasteurizer and the data obtained from passing a thermal recorder from the pasteurizer while in use, along the cans filled with beer (see Figure 2b).

For the first category of data, PT100 type of sensors were installed on the pasteurization machine and by using an IoT controller setup (based on a custom industrial version of Raspberry Pi, namely Revolution Pi), those data can be accessed and stored on the cloud. The pasteurizer is divided in six bath zones where water is spayed on the cans with different temperatures on each zone. The exact data collected by the sensors are the following:

1) A binary value that indicates if the belt of the pasteurizer is moving.
2) The room temperature of the building where the pasteurizer is housed.
3) A binary value that indicates if the previous machine in the factory line is in operation (as it affects the performance of the pasteurizer).
4) The temperature of the water stored in the water tank in the first to fourth zones of the pasteurizer.

The second category of data are recorded by using the thermograph recorder that passes through the pasteurization tunnel during normal operation alongside with the cans.

The data collected and used for building the machine learning models come from 265 passing’s of the thermal recorder from the pasteurizer. These passings contributed to 57.179 rows for the internal spraying water temperature dataset and 58.940 rows for the product cold spot temperature dataset.

D. Data preprocessing

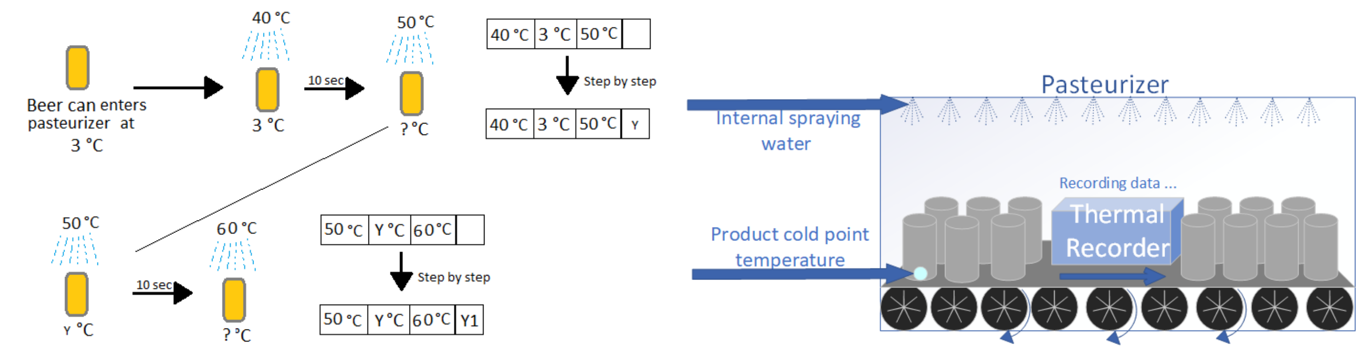
Three datasets are created for training the machine learning models (discussed later in section III-E). Datasets A and B are used for spraying water estimator and dataset C for the cold spot estimator. The only difference between dataset A and B is the additional usage of the 9nth feature (prev-water-temp) in dataset B. We summarize the input features of each dataset along with the source of collection in Table I.

		Dataset Names			
		Spraying water		Cold Spot	
No.	Feature Name	A	B	C	Source
1	Factory-temp	Yes		No	IIoT controller
2	Paster-run	Yes		No	IIoT controller
3	Paster-time	Yes		Yes	preprocessing
4	Paster-program	Yes		No	IIoT controller
5	Bath-number	Yes		Yes	preprocessing
6	Bath-temp	Yes		Yes	IIoT controller
7	Prev-bath-temp	Yes		No	IIoT controller
8	Next-bath-temp	Yes		No	IIoT controller
9	Prev-water-temp	**No**	**Yes**	Yes	Thermograph
10	Prev-can-temp	No	No	Yes	Thermograph
11	Water-temp	No	No	Yes	Thermograph

TABLE I: *Datasets and their corresponding features*

In the following list we provide a brief description concerning each feature in the table:

1) *Factory-temp.*: Temperature of the building that houses the pasteurizer.
2) *Paster-run*: Binary value that shows if the belt of the pasteurizer is moving or not (cans change position).
3) *Paster-time*: The time needed for the pasteurization process to be completed is divided in instances of 10 seconds and the value of this feature indicates in what instance of the pasteurization the can has reached so far.
4) *Paster-program*: The pasteurizer has different pasteurization programs for different products, and each program may require different length of time to complete



(a) Step by step prediction flow for cold spot temperature estimation

(b) Abstract representation of the prediction targets

Fig. 2: Outline of prediction flow and pasteurization targets

the pasteurization process. So, Paster-program contains the value of the program used on the pasteurizer.

5) *Bath-number*: This value indicates in which zone of the pasteurizer the can is passing through, when the prediction is made (1, 2, 3, 4, 5, 6).
6) *Bath-temp*: The temperature of the water in the tank of the zone where the can is currently passing through.
7) *Prev-bath-temp*: The temperature of the water in the tank of the previous zone where the can passed through.
8) *Next-bath-temp*: The temperature of the water in the tank of the next zone from where the can is currently passing through.
9) *Prev-water-temp*: Temperature of the water falling on can 10 seconds before current estimation.
10) *Prev-can-temp*: Temperature of the cold spot of can 10 seconds before current estimation.
11) *Water-temp*: This is the temperature of the water falling on the cans (Target of Internal spraying water temperature estimator and used as input for prediction for Product cold spot temperature estimator)
12) *Can-temp*: The temperature of the cold spot of the can (Target of Product cold spot temperature estimator)

Considering the source of the last four features (Prev-water-temp, Prev-can-temp, Water-temp, Can-temp), their values are taken from two sources according to the ML model building stage. In particular, when the datasets are created these values are taken from the recorded values of the thermograph to train the models. However, when making the simulation predictions for a pasteurization run (i.e., when these values are used as input for the estimators to run the ML models as a real time simulation) the values are taken from the previous predictions of the estimators. Eventually, in each time step the predictions of the estimators are evaluated against the actual thermograph measurements.

Finally, after the aforementioned dataset creation procedure, each dataset is splitted into the following datasets: train (for ML model building) and test (for ML model evaluation on real, unseen data) with a 80/20 % split.

E. Machine learning algorithms

After experimentation with different machine learning algorithms, we present only our top findings to provide a comprehensible and detailed analysis of the results in the next section.

The algorithms we therefore present include the decision tree (DT), ridge regression and stacked ensembles (SE). Decision trees are used for the internal spraying water estimator, while ridge for the cold spot temperature estimator and the stacked ensembles were used for both estimators. The first two models are implemented using scikit-learn [14] library, while the stacked ensemble models using the TPOT AutoML library [15]). The experiments were held on a local workstation with the following specs; OS: Windows 10, RAM: 16.0 GB, CPU: AMD Ryzen 7 3700X (3.60 GHz).

Stacked ensembles consist of multiple single models that are individually trained. The predictions of the single models are then used as features to train a meta-model with the aim to achieve better results, by combining the predictions of the single models and improving on the flaws that these single models may have. We explore stacked ensemble models to find more complex data representations, in comparison with the other aforementioned algorithms, and for this reason we leverage TPOT as an AutoML tool to extract the optimal stacking models for this problem.

The final models and their main hyperparameters for **Internal spraying water temperature estimator**: Stacked ensemble:

- SGDRegressor: alpha=0.0, learning_rate="constant", loss="epsilon_insensitive", eta0=0.01, fit_intercept=True, l1_ratio=0.5, penalty= "elasticnet", power_t=50.0
- ExtraTreesRegressor: min_samples_leaf=6, bootstrap=True, max_features=1.0, min_samples_split=15, n_estimators=100
- ExtraTreesRegressor: min_samples_leaf=7, bootstrap=False, max_features=0.8, min_samples_split=17, n_estimators=100

Decision tree: criterion = "gini", splitter = "best", min_samples_splitint = 2, min_samples_leafint = 1.

Similarly, we list the models and main hyperparameters for **Product cold spot temperature estimator**. Stacked ensemble:

- LassoLarsCV: normalize=True
- ExtraTreesRegressor: threshold=0.0, max_features=0.15, n_estimators=100
- RandomForestRegressor: min_samples_leaf=3, bootstrap=True, max_features=0.75, min_samples_split=9, n_estimators=100)

Ridge: alpha = 1.0 fit_interceptbool = True, normalizebool = False, copy_Xbool = True, max_iterint = None, tolfloat = 1e-3, positivebool = False, random_stateint = None

IV. Experimental Results and Evaluation

In this section we present and discuss the results from the experiments that are worth showcasing for predicting the internal spraying water temperature as well as the product cold spot temperature. Details on our soft sensing approach and ML algorithms are provided in sections III-B and III-E.

A. Internal spraying water temperature estimation

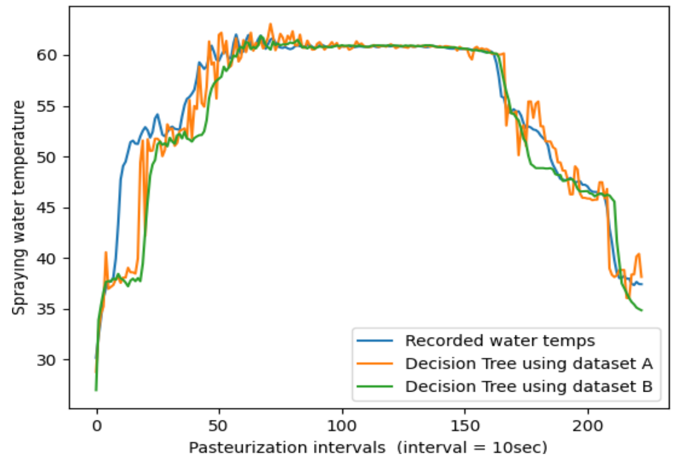
Model	Prev. spraying water	MAE	MAPE (%)	RMSE	R^2
DT	✓	3.61	7.39	5.24	0.40
DT	×	1.99	3.99	3.20	0.77
SE	✓	1.41	2.80	2.16	0.85
SE	×	**1.29**	**2.56**	**1.91**	**0.89**

TABLE II: *Results for spraying water temperature estimation per model and spraying water input values*

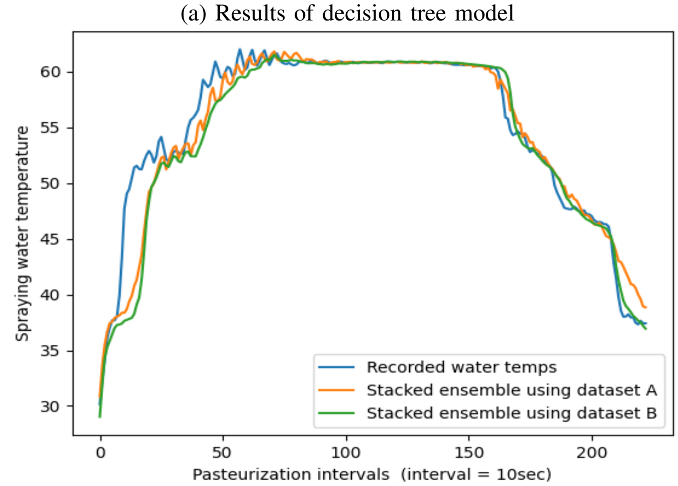
In Table II we present the results for the models that estimate the internal spraying water temperature. The first column concerns the ML algorithms to be compared, in this case, decision tree (DT) and stacking ensemble (SE) estimator, which are both described in section III-E. The second column of the table concerns the dataset used to train the models. In particular, if the previous value of the spraying water temperature is included as an input feature to predict the temperature in the next 10 seconds (dataset B) or not (dataset A), it is indicated with a check-mark or cross-mark respectively. Both datasets have been explained in section III-D concerning data preprocessing. In the next four columns the metrics of each case are displayed.

From the table we easily notice that the model which produces the best results is the one using the stacked ensemble discovered with TPOT and trained without using the previous spraying water temperature value (dataset A). However, the usage of this feature (dataset B) in the same model has similar performance with small deviations, comparing to decision tree that achieves significantly lower performance in each metric.

In Figure 3 we demonstrate the results for a single simulation of predictions, i.e, a single end-to-end pasteurization process run for a given batch of products. In particular, Figures 3a and 3b represent the results based on the model used, i.e., decision tree and stacked ensemble respectively, and also each of the Figures compares the model performance for each dataset: A or B, i.e., if previous spraying water value is used.



(a) Results of decision tree model



(b) Results of stacked ensemble model

Fig. 3: Spraying water temperature estimation per model and dataset for a single pasteurization run

At first, we observe that the stacked ensemble produces more consistent predictions than the decision tree. The latter has higher fluctuations, which leads to the overall inferior performance. We also notice that the models that take into account the value of the previous 10 seconds of the water temperature (dataset B), tend to follow a smoother prediction line. This extra feature seems to prevent the models of producing high deviations between consecutive predictions, as the previous values have a high feature importance and they influence the predicted next temperature. On the other hand, the models that do not leverage this information (dataset A), produce more spikes and predict better the sudden temperature deviations.

B. Product cold spot temperature estimation

In Table III the results of the models for product cold spot temperature estimation are displayed. The results are measured the same way as for the internal spraying water temperature using the same process of simulation. All models for this case are trained with dataset C.

The table is organized in a similar manner as Table II in the previous case. The first column concerns the ML algorithms

Model	Spraying water data: Input	Spraying water data: Prev. water	MAE	MAPE (%)	RMSE	R^2
Ridge	DT	✓	3.24	6.63	4.54	0.62
		×	1.66	3.53	2.30	0.91
	SE	✓	1.41	2.97	1.99	0.93
		×	1.46	3.09	2	0.92
	Recorded values		*0.77*	*1.77*	*1.05*	*0.98*
SE	DT	✓	3.79	7.48	5.09	0.37
		×	1.53	3.18	2.16	0.92
	SE	✓	**1.24**	**2.58**	**1.80**	**0.94**
		×	1.32	2.74	1.85	0.93
	Recorded values		*0.60*	*1.31*	*0.85*	*0.99*

TABLE III: *Results for cold spot temperature estimation per model and spraying water input values*

to be compared, in this case, ridge regression and stacking ensemble (SE), which are both described in section III-E. The models for product cold spot temperature estimation are using as an input feature the results of the internal spraying water temperature estimator and so it is important to show for each experiment which spraying water model and training dataset was used to predict those results (referred as "Spraying water data" in the table). In particular, the second column of the table is referring to the model used to create the internal spraying water temperature estimator (DT = Decision Tree, SE = Staking Ensemble). The third column of the table concerns the dataset used to train the models. If the previous value of the spraying water temperature is included as an input feature to predict the temperature in the next 10 seconds (dataset B) or not (dataset A), it is indicated with a check-mark or cross-mark respectively. Both datasets have been explained in section III-D concerning data preprocessing. The next four columns present the resulting metrics of the experiments.

As noted in section III-B, we also feed the ML models that predict the cold spot temperature, with the actual, real (recorded from the thermograph) water temperature values. Therefore, "Recorded values" rows in the table correspond to the results of the Ridge and stacked ensemble (SE) models using the recorded water values and we use these results as a baseline for assessing the models predictive capacity and showcase the magnitude of the effect the water input has on the cold spot models.

Table III indicates clearly that the best results come from the stacked ensemble (SE) and in particular using as input the stacked ensemble-based spraying water estimator that is trained with the previous spraying water temperature values (dataset B). However, the results of SE are still very close even without using that feature, with only unit differences in the metrics (e.g., 1.80 vs 1.85 RMSE respectively).

Considering the performance on recorded water values using either Ridge or the stacked ensemble, the performance is exceptional (e.g., for the SE, RMSE of 0.85, 0.99 R^2), meaning that with a more accurate spraying water input from the spraying water estimator, the cold spot estimator can perform very well. At the same time, there is a potential for slight improvements on the cold spot model itself, but still this mainly depends on the quality of water input from the spraying water estimator.

We also note the overall comparable performance of the Ridge model when using as an input the stacked ensemble-based spraying water estimator. However, since there are small but clear deviations from the stacked ensemble-based cold spot estimator using either the spraying water estimator or the recorded water values, the stacked ensemble is selected as the best model for predicting the product cold spot temperature.

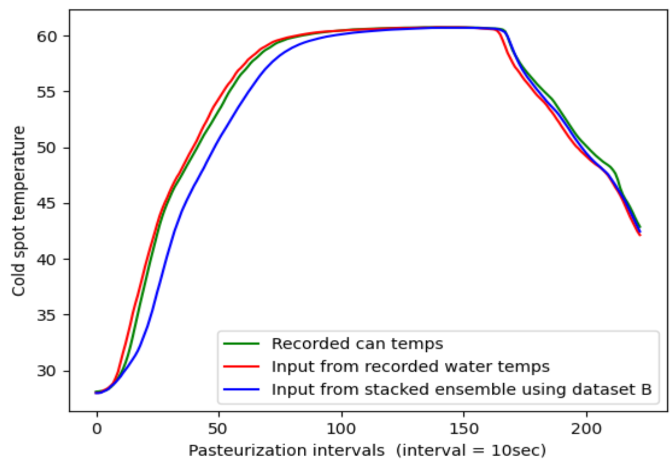


Fig. 4: Model performance when using real (recorded) water values and predicted output of spraying water estimator

Figure 4 refers to model performance when using real (recorded) water values and predicted output of spraying water estimator. The red line displays the predictions of the SE-based cold spot estimator using as part of its input the real values of water temperatures sprayed on the cans. The blue line depicts the predictions of the cold spot estimator using as part of its input the predicted values of water temperatures sprayed on the cans and the green line represents the real recorded values of the product cold spot temperatures.

By using the recorded sprayed water temperatures as an input, the results of the cold spot estimator are almost identical with the real values recorded from the thermal recorder. The blue line also shows how impactful is the water input data to the estimator, as the same pattern of errors (e.g underestimation in the ascending trend of temperatures) can be observed in Figure 3 that displays the results of the internal spraying water temperature estimator.

To sum up, the results indicate clearly that the quality of the spraying water estimator constitutes a significant factor for the quality of the cold spot estimator, since the spraying water temperature is given as an input to the cold spot models. Therefore, future works will focus on improving the spraying water estimator and reduce as much as possible the prediction errors. Moreover, we point out that the stacked ensembles (SE) perform better than less complex, single models and thus we argue that using AutoML tools, at least in the prototype stage of a virtual sensing approach, is beneficial in finding the optimal models and results and thus proving the advantages of developing machine learning-based soft sensors.

Finally, we highlight that the usage of the previous value of the spraying water temperature results to ambiguous results since it is beneficial for the cold spot estimator but not for the spraying water estimator. The differences in the overall model performance are very close using either dataset (A or B), so the highest impact comes from the rest of the features that represent the state of the pasteurization (e.g. bath number, pasteurization program and time, bath temperature, etc.). However, since using this input leads to better predictions in certain values we still consider experimenting with it, especially for future works that it might have a greater impact, such as the prediction of total accumulated pasteurization units (PUs) during the pasteurization process.

V. Conclusion and future work

In this paper we developed two soft (virtual) sensors based on machine learning models, as a means to realize automated quality monitoring of products in a real world tunnel pasteurization process with data-driven techniques. The first model predicts the spraying water temperature and the second one the temperature of the packed product at its cold spot. We compare different machine learning algorithms, namely decision tree, ridge regression and stacked ensembles, with the latter achieving the highest performance for both cases/targets, having a RMSE of 1.91°C and 1.8°C respectively.

This means that our soft sensing approach can be considered as a promising, virtual low-cost "upgrade" of legacy tunnel pasteurizers, by exploiting their data with ML models and IIoT controllers and providing real time transparency to their operation, without the need for manual thermograph samplings. We also argue that our approach is designed to be compliant with pasteurization processes of different properties compared to the studied process in this work. If a pasteurizer machine has differences, e.g., different number of temperature zones or different pasteurization time in each zone, etc., we expect similar results in terms of accuracy as in this paper, given that adequate quality data are collected and some light preprocessing modifications are applied in the datasets to fit for the different properties.

Considering future extensions to this work, we aim to gather additional data that is required from thermograph samplings to extend the soft sensing and predict an essential quality KPI which is the accumulated Pasteurization Units (PUs) in each batch of pasteurized products. Finally, another research venue to be investigated is the experimentation and utilization of more advanced models such as deep neural networks (e.g. LSTMs to process the temporal pattern included in the recorded time series data) or time series forecasting techniques (e.g. ARIMA) to model the temperature based on the historic sequence of the time series.

Acknowledgment

The present work was financially supported by the "Andreas Mentzelopoulos Foundation".

References

[1] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Computers & Chemical Engineering*, vol. 33, no. 4, pp. 795–814, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0098135409000076

[2] L. Yao and Z. Ge, "Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1490–1498, 2018.

[3] M. Drakoulelis, G. Filios, V. G. Ninos, I. Katsidimas, and S. Nikoletseas, "Virtual sensors: an industrial application for illumination attributes based on machine learning techniques," *Annals of Telecommunications*, vol. 76, no. 7, pp. 529–535, 2021.

[4] G. Filios, I. Katsidimas, S. Nikoletseas, S. Panagiotou, and T. P. Raptis, "An agnostic data-driven approach to predict stoppages of industrial packing machine in near future," in *2020 16th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2020, pp. 236–243.

[5] C. Riverol, G. Ricart, C. Carosi, and C. Di Santis, "Application of advanced soft control strategies into the dairy industry," *Innovative Food Science & Emerging Technologies*, vol. 9, no. 3, pp. 298–305, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1466856407001063

[6] J. Dai, N. Chen, X. Yuan, W. Gui, and L. Luo, "Temperature prediction for roller kiln based on hybrid first-principle model and data-driven MW-DLWKPCR model," *ISA Transactions*, vol. 98, pp. 403–417, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019057819303593

[7] X. Zhang, M. Kano, and S. Matsuzaki, "A comparative study of deep and shallow predictive techniques for hot metal temperature prediction in blast furnace ironmaking," *Computers & Chemical Engineering*, vol. 130, p. 106575, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S009813541930674X

[8] C. Liu, Y. Wang, K. Wang, and X. Yuan, "Deep learning with nonlocal and local structure preserving stacked autoencoder for soft sensor in industrial processes," *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104341, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197621001895

[9] J. C. Kabugo, S.-L. Jämsä-Jounela, R. Schiemann, and C. Binder, "Industry 4.0 based process data analytics platform: A waste-to-energy plant case study," *International Journal of Electrical Power & Energy Systems*, vol. 115, p. 105508, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142061518336731

[10] C. Shang, F. Yang, D. Huang, and W. Lyu, "Data-driven soft sensor development based on deep learning technique," *Journal of Process Control*, vol. 24, no. 3, pp. 223–233, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959152414000365

[11] J. X. Leon-Medina, J. Camacho, C. Gutierrez-Osorio, J. E. Salomón, B. Rueda, W. Vargas, J. Sofrony, F. Restrepo-Calle, C. Pedraza, and D. Tibaduiza, "Temperature prediction using multivariate time series deep learning in the lining of an electric arc furnace for ferronickel production," *Sensors*, vol. 21, no. 20, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/20/6894

[12] D. A. Sala, A. Jalalvand, A. Van Yperen-De Deyne, and E. Mannens, "Multivariate time series for data-driven endpoint prediction in the basic oxygen furnace," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018, pp. 1419–1426.

[13] W. Zhao, J. Jiang, H. Qin, X. Li, and J. Li, "Machine learning based soft sensor and long-term calibration scheme: A solid oxide fuel cell system case," *International Journal of Hydrogen Energy*, vol. 46, no. 33, pp. 17 322–17 342, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360319921006637

[14] 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*, vol. 12, pp. 2825–2830, 2011.

[15] R. S. Olson, N. Bartley, R. J. Urbanowicz, and J. H. Moore, "Evaluation of a tree-based pipeline optimization tool for automating data science," in *Proceedings of the Genetic and Evolutionary Computation Conference 2016*, ser. GECCO '16. New York, NY, USA: ACM, 2016, pp. 485–492. [Online]. Available: http://doi.acm.org/10.1145/2908812.2908918