The Effect of Economic and Individual Variables on Retirement Decisions

Jose de Jesus Rocha Salazar

May 2019

University of Liverpool

https://www.liverpool.ac.uk/

Content

Abstract

Introduction

1 How Macroeconomic and Financial Fluctuations Affect Retirement: The Case of an Oil Producing Country

1.1 Introduction………………………………………….. 9
1.2 Pension System in Mexico…………………………... 11
1.3 Methodology and Model ……………………………. 14
1.4 Data…………………………………………………...16
1.5 Results……………………………………………….. 21
1.6 Conclusions…………………………………………...29

2 Machine learning techniques: Application to the pension industry

2.1 Introduction…………………………………………... 30
2.2 Machine Learning Methodologies…………………….34
2.3 Data…………………………………………………....37
2.4 Model………………………………………………….44
2.5 Results……………………………………………….... 49
2.6 Conclusions…………………………………………....54

3 Conclusions…………………………………………………....55

References

Appendix A: Pension Schemes in Mexico by Law

Abstract

Pension systems in several countries have begun to show a lack of efficiency and sustainability, which motivates governments to investigate which factors cause this loss of balance. It is commonly known that traditional pension systems where active workers finance the pensions of retired workers as pay-as-you-go (PAYGO) have a strong dependence on the demographic structure of countries. Here is where factors such as the increase in life expectancy and a decrease in birth rate, have an impact on this balance. Several measures aimed at keeping the sustainability of pension schemes are focused on the modification of variables, such as retirement age and contribution of workers, to maintain the solvency of the system. Some articles have gone beyond and studied how macroeconomic and individual variables can influence the decision of early retirement.

This study contributes to written literature about the effect of macroeconomic variables on retirement decisions. It expands in the inclusion of oil prices as an important variable that can affect retirement, specifically in countries that export oil and whose economy depends on this income. Additionally, this study presents an innovative model of artificial intelligence developed for the industry to predict the early retirement of individuals. This model takes into account economic factors and individuals characteristics to create an alert system that constantly is fed as new data is generated.

This research has a strong impact in recent years. Oil prices have had several fluctuations, mainly downward, affecting the economy of countries in different ways. Early retirement is one example of the consequences in the economy caused by the fluctuating oil prices. Currently, there is no machine-learning model implemented internally in insurance companies or government institutions to track early retirement. The alert system allows controlling early retirement at a company level. The governments and administrators of pensions should be aware of this.

Introduction

The concept of retirement began to be institutionalized and become a reality in the life of workers in the 19th century, Costa (1998). Later it began to expand and gain strength especially in developed and industrialized countries in the 20th century, Atchley (1996). The concept of retirement did not originate spontaneously or suddenly; two reasons explain its emergence. The first one is that workers agreed with their employers the retirement age because as finite beings at some point in their lives they were going to face weakening of their physical abilities to perform the job, Atchley (1993). Second, old workers faced difficulties with finding work because employers preferred younger and stronger workers. This last reason limited the obtaining of income by old individuals and it made it necessary to develop a system that provided financial sustenance in this new stage of life.

The definition of retirement has changed over the years, Zickar (2012). Retirement is colloquially defined as a change of state in the working life of individuals, a definitive abandonment of the labour force. Definitive abandonment is key in this definition because many individuals change jobs several times throughout their lives passing through a short period of unemployment. This short period of time away from the labour force does not necessarily mean a permanent retirement, Denton and Spencer (2009).

Retirement is also defined as the event where workers leave the labour force due to a decrease in psychological commitment to work and an increase in behavioural withdrawal, Wang and Shi (2014). The individual's new financial support would be the pension, which is different to the salary from an economic activity, Szinovac (2003). The individual's new status will be composed of leisure activities, volunteering or secondary careers.

Nowadays it has become important for public and private institutions around the world, especially for western countries, to study the retirement behaviours of individuals, Hakola and Uusitalo (2004). The ageing of the baby boomer generation has enhanced this importance, Beehr and Bennett (2007). Pension systems in several countries have lost their balance due to the ageing of this generation. This has awakened the interest of Governments, institutions and researchers to investigate which factors, besides natural ageing, affect the financial solvency of the pension schemes. Another alarming factor that has strengthened the study in this issue is the drastic fall in the labour participation rate of old individuals in all OECD countries, Dorn and Sousa (2010). The latter derives the economic concern that ageing is exacerbated as more and more people become solely consumers rather than both consumers and producers.

It is clear that the increase in life expectancy, demographic structure and decrease in birth rate are factors that weaken the financial health of pension schemes, as well as the point in time in which the worker's retirement plans are consumed. The moment in which the decision of retirement is taken for a worker determines whether the retirement is early or late. Early retirement is usually defined as definitively leaving a long career before the age of 65 years old, Feldman (1994) and Wang and Alterman (2017). Early retirement begins to be built during the work stage of individuals and evolves to consummate in retirement at some point

Early retirement can be voluntary or involuntary. Voluntary early retirement is decided by the individual with all the willingness to do leisure activities rather than to be working. Involuntary early retirement refers to that which arises unexpectedly from situations that the individual cannot control. Both have impact on individual's consumption after retirement. Involuntary early retirement has a greater impact than the voluntary one due to the factor of uncertainty present, Smith (2006). In some cases, both types of early retirement might be present at the same time.

Early retirement is caused by several events, whether voluntary or involuntary. Studies, such as that of Breinegaard et al. (2017), associate early retirement with organizational and management factors. Stress has also been a factor associated with early retirement, Want et al. (2008). The family is a variable that can also influence early retirement, Figueira et al. (2017), for example, if the spouse receives a high income that allows the couple to have a good quality of life, the other one might retire early. The social environment at work is often the second family and home of the employee. This is why conditions at work and how comfortable employee feels influence retirement decisions. If the worker does not have a good relationship with colleagues or the environment is simply unfriendly and stressful he will retire early, Elovainio et al. (2005) and Carr et al. (2016). Studies such as Burkhause (1979), Gustman and Steinmeier (1986), Blundell et al. (2002), Queiroz and Souza (2017) explain that the more generous the pension amount the earlier the individuals will retire. Other international studies such as Blöndal and Scarpetta (1998) and Dorn and Souza (2005) confirm that early retirement is more frequent in countries with more flexible retirement regulations.

Another important cause of early retirement is a deteriorated health condition, Leinonen et al. (2016). Studies such as Burtless and Quinn (2000) explain that workers with a poor health status and who perform very demanding jobs tend to be the first to retire.

In addition to affecting the financial balance of pension systems, the early retirement has bad consequences on the economy and society. Early retirement causes loss of skilled labour replacing the positions with unskilled workers affecting the productivity of companies. This replacement is basically to lay off workers when they are old. In addition to this, companies are reluctant to hire people in advanced age, focusing on hiring only young workers who often lack of knowledge and experience. The latter may be due to young workers are preferred over old workers because they demand a smaller salary, the expertise of older workers may be non-transferrable and older workers seeking employment may represent a subset that is less productive. Second, early retirement reduces the individual's range of social contacts and hastens social support. Finally, early retirement implies less tax revenue and the provision of benefits and pensions[1].
Identifying and controlling the factors that influence early retirement is important, not only for a balanced pension system but also for good economic performance. Some nordic countries, particularly Finland, has promoted a longer career. These countries value old workers for their years of experience and motivate them to work as much as possible. Also, countries in Europe joined this practice and began to improve the working conditions of older workers so that they feel comfortable and stay working for longer, Reday and Mulvey (2005).

In addition to these measures, studies have focused on analysing how macroeconomic conditions can influence retirement decisions. An example of this is the article of Coile and Levine (2011) where they study how an economic crisis impacts retirement decisions. They focus on fluctuations of variables such as stock market, housing prices and unemployment rate.

Following the above interest, this thesis is focused on studying how certain macroeconomic, financial and individual factors affect retirement decisions. The project is divided into two chapters.

Chapter one uses time series analysis and econometric techniques to determine the impact of variables such as unemployment rate, the stock market and petroleum prices on retirement in an oil producing country. Emphasis is placed on oil prices since this variable has fluctuated in recent years due to the increase in its production and geopolitical issues. It is interesting to analyse how the macroeconomic variables mentioned above together with oil prices influence retirement decisions. This study is interesting for oil importing and exporting countries. China and India, as importing countries, have an accelerated economic growth that fluctuates around 15% and 10% respectively. This economic growth is linked to a greater consumption of oil. If oil prices decrease, the costs of the manufacturing, transport and other industries decrease as well, causing an increase in the consumption of individuals, improving their well-being. It is also proven that a decrease in oil prices has an effect on inflation and economic growth in the country. In the case of exporting countries, such as Kuwait, Venezuela and Saudi Arabia, the effect is totally the opposite in the same issues mentioned above. Drops in oil prices can cause an economic slowdown in these countries since they are highly dependent on the income from exports of this resource. This decrease in income directly affects the government budget that invests part of the earnings in social programs and internal investment[2]. Based on this information, it is natural to think that retirement decisions of workers are influenced by fluctuations in petroleum prices.

Chapter two applies machine learning algorithms in an insurance company to predict the retirement of workers in function of macroeconomic factors and individual characteristics. The objective is to develop a computer system that alerts private insurance companies when an individual has a high probability of retiring at an early age. This allows for tracking early retirement and take action on time to avoid it or be financially prepared. This is a first attempt to implement artificial intelligence in a private insurance company. Previously, the models only identified causality between explanatory variables and retirement decisions without predicting exactly when an early retirement could occur. The model proposed here determines an early retirement probability, which allows us to predict this event with a high level of confidence and make it visible in the computer systems of private insurance companies. The objective of this chapter is purely to classify and generate alerts identifying certain patterns. It does not focus deeply on finding causal relationships. This is why the machine learning methodologies were used in this chapter.

Basically, chapter one analyses retirement decisions at country level and chapter two at a company level.

[1] See “The Social Consequences of Early Retirement”, Alan Walker.
[2] Essays, UK. (November 2013). Effects of an oil price shock on importing and exporting countries.

1

How Macroeconomic and Financial Fluctuations Affect Retirement: The Case of an Oil Producing Country

José de Jesus Rocha Salazar[a] and Maria Del Carmen Boado-Penas[a]

This paper analyzes the impact of macroeconomic and financial variables on retirement. Special attention is given to petroleum prices since this variable has not been taken into account in previous studies and is essential for those countries immersed in the trade of this natural resource. Not only the unemployment rate but also stock market (Price Index of the Mexican Stock Exchange) and petroleum prices are considered as explanatory variables in our model. The study considers the reactions of retirement by gender, age and level of education. We conclude that in the long term, there is an increase in the number of new pensioners when oil prices decrease.

Keywords: Financial effect; macroeconomic effect; petroleum prices; retirement decisions

1. Introduction

The retirement of individuals has become of great importance for governments because of its social implications. Over the years in western countries the proportion of men over 65 years old in the workforce decreased from 31% in 1951 to 12% in 1981, Walker (1982). This was caused by early retirement and because unemployment is common among individuals close to retirement, Munnell (2006). The major interest in understanding retirement is the provision, level of retirement pension, and consequences on the economy of a country, Walker (1982).

Many people decide to delay their retirement age because economic conditions allow it. In other cases, as explained by Bosworth and Burtless (2010), when the unemployment rate grows, people decide to retire early. Some individuals want to meet the retirement requirements as quickly as possible because they prefer enjoying leisure activities than working as explained by Muldoon and Kopcke (2008).

The economic crisis in general also has an impact on retirement decisions as explained by Hurd and Rohwedde (2010). Using HRS data between 2008 and 2009, the authors showed that many were planning to work longer and delay their retirement as a result of the crisis experienced by the USA in those years.

Coile and Levine (2011) find that workers between 62 and 69 years old are affected by fluctuations in the unemployment rate and the long-run stock market. In particular, workers with lower education are affected by the unemployment rate and those with higher education are affected by the stock market. In the same line, Bosworth and Burtless (2010) conclude that a decrease in the value of stock market and housing prices cause a delay in retirement because people decide to stay in the labour market longer to rebuild their lost wealth.

For oil producers, the economy reacts to petroleum prices and its fluctuations, according to Bach, Sunila and Kumar (2015). In the last few years, petroleum prices have experienced fluctuations that range from 60 USD per barrel to a peak of 146 USD in 2009 and subsequently descended again to below 50 USD in 2015, Organization of the Petroleum Exporting Countries (2015). These changes in prices affect in different ways the economy of the countries involved in the international trade of petroleum. Brazil, as producer oil country, the decrease in petroleum prices caused a slowdown in the growth of GDP, investment strategies, construction sector and industries linked to the oil production, Cavalcanti and Jalles (2013) and Florêncio (2016). China, as an importing country, the falling prices led to better economic growth especially in the industrial sector due to a decrease in expenditures. Industries, such as airline and agricultural, might experience a reduction in their sale prices provoking less inflation and more consumption when oil prices remain low, Qianqian (2011) and Tian (2016).

Mexico is the eighth largest oil producer in the world. The production of petroleum and related products represent between 7 and 10.5 per cent of GDP, and the income from sales represents about 33 per cent of tax revenue. The fall in oil prices between 2014 and 2016 caused a cut reduction in federal and sub-federal public spending which affected the investments in several sectors such as education, health and various social and internal investment programs[3]. This paper, using data from Mexico, examines the effect of macroeconomic and financial variables on retirement and extends previous studies observing the impact of oil prices fluctuations. Following this introduction, in Section II, we briefly explain the Mexican pension system. Section III presents the methodology and model used. Section IV shows the data used. The paper ends with the main results and conclusions.

2. Pension System in Mexico

This section explains the background and structure of the pension system in Mexico since data from this country are used to do the study.

Social security issues began in Mexico in 1943 when the social security law was enacted. The law started to be effective in 1944. It covered affairs such as pensions and medical services for workers and their families. As a developing country and a growing economy, the coverage was only provided in the capital city of the country, leaving the provincial states unprotected. This action was not a pilot test to study how a social security system worked in the country; rather it was a limited start due to a lack of resources.

Later between the 1950s and 1960s and thanks to the investment of foreign companies, mainly from the United States, Mexico began to experience accelerated economic growth. Industries such as the oil industry began to generate income for the country. Thanks to these resources, the social security service expanded throughout the country. More hospitals, clinics were made and a pension system in the pay-as-you-go scheme was consolidated and administered by the Mexican Institute of Social Security (IMSS).

Since its implementation, the IMSS has been financed by contributions of workers, companies and the government. The IMSS covers workers in the private sector. Public workers are affiliated to the Institute of Security and Social Services of State Workers (ISSSTE) and the armed forces are covered by Mexican Petroleum (PEMEX).

Initially, the Mexican pension system was governed by a pay-as-you-go scheme where the contributions of the active workers paid the pension of the retirees in a particular period of time. Over time the system began to lose sustainability due to the increase in life expectancy and a decrease in the birth rate. All the negative issues that affected the distribution scheme were debated and finally, in December 1995, the Social Security Law was reformed. The scheme went from a distribution scheme to one of the individual defined contribution accounts managed by AFORE (National Commission of the Savings System for Retirement). The law began to be effective on July 1, 1997.

---

[3]Ministry of Finance and Public Credit (SHCP) 2017, Opportunistic Public Finance Statistics. Gentleman, J. 1984. “Mexican Oil and Dependent Development”. New York: P. Lang.

In recent years about 50% of insureds are covered by the IMSS, as shown in Table 1.

Table 1. The population of insured individuals by an institution.

Institution	Number of insured individuals	Percentage
IMSS	74,032,437	49.35%
ISSSTE	12,973,731	8.65%
Seguro Popular	57,105,622	38.07%
PEMEX/SEDENA/SEMAR	1,893,946	1.26%
Private Institutions	2,182,514	1.45%
Other Public Institutions	1,824,595	1.22%

Own elaboration. Report to the President and Congress about the financial and risk situation of IMSS 2015-2016.

Currently, the National Commission of the Savings System for Retirement (CONSAR) regulates the AFORES that administer the individual accounts of the workers. They invest resources in investment funds called SIEFORES (Specialized Investment Companies in Retirement Funds). Recently, the 11 existing AFORES administer about 57,432,774 accounts as shown in Table 2.

Table 2. Accounts managed by AFORES.

Afore	Number of Registered Workers	Resources deposited in SIEFORES (Millions of Mexican Pesos)	Resources deposited in Bank of Mexico(Millions of Mexican Pesos)	Total of accounts managed by AFORES(Millions of Mexican Pesos)
Azteca	1,818,313	241	0	1,818,554
Banamex	7,118,698	3,036,765	0	10,155,463
Coppel	8,135,890	9,198	0	8,145,088
Inbursa	1,078,835	451	0	1,079,286
Invercap	1,859,578	610,807	0	2,470,385
Metlife	428,796	293,587	0	722,383
Pension ISSSTE	1,450,467	618,652	0	2,069,119
Principal	2,185,106	565,745	0	2,750,851
GNP	2,707,348	1,204,151	0	3,911,499
SURA	4,078,125	3,330,254	0	7,408,379
XXI Banorte	8,156,052	1,485,094	7,260,621	16,901,767
Total	**39,017,208**	**11,154,945**	**7,260,621**	**57,432,774**

Own elaboration. CONSAR statistics, March 2017.

It can be said that since 1997 a mixed pension system operates in Mexico. All the workers in Mexico have an individual account in the AFORES independently of the time they started to contribute. People who started contributing before the first of July 1997 have the right to decide to retire between the previous scheme and the new scheme. People who started contributing after the enactment of the law of 1997 only have the right to a pension under the new scheme of individual accounts. It is likely that most of the individuals with the option to decide between the schemes choose to retire by the law of 1973 since the retirement requirements are less and the benefits are better[4].

The pension scheme in Mexico is going through a period of transition and the responsibility in the last 15 years has fallen on the federal government that finances pensions through the income from exports and contributions from workers. Oil exports nourish fiscal income and fluctuations in their prices have a direct impact. In this sense, the drop in oil prices directly affects the sustainability of the pension system by motivating reforms that increase contributions and retirement age. For this reason, is natural to think that fluctuations in oil prices can affect the retirement decisions of workers. The unemployment rate affects in the sense that older workers will want to leave the labour market when the rate is high as explained by Bosworth and Burtless (2010). The effect of the stock price index is associated with the impact of unemployment rate on this index, Gonzalo and Taamouti (2017).

3. Methodology and Model

Monthly information of times series is available from 2005 to 2016.

In this paper, we model the frequency of new pensioners using the unemployment rate, the stock market and petroleum prices as explanatory variables.

For the purpose of this study the frequency of the retirement in a period of time will be observed through the following proportion [5]:

$$\frac{N_t^{i,j}}{T_t^{i,j}} \quad (1)$$

Where:

$N_t^{i,j}$: is the number of new pensioners at a time *"t"* with individual characteristics *"i"*, "j".

$T_t^{i,j}$: is the total population at time *"t"* with individual characteristics *"i"*, "j".

Where i and j can correspond to gender, age or level of education of the individuals. In the case of this study we consider two combinations, $(i,j) = (gender, age)$ and $(i,j) = (gender, level\ of\ education)$.

The proportion $\frac{N_t^{i,j}}{T_t^{i,j}}$ can also be interpreted as the probability of new retirees with characteristics *"i"*, "j" in the period "t".

The steps are the following:

1.-Correlations between the explanatory variables are determined. If there is a high correlation between some of them, the analysis of principal components analysis is used to reduce the number of variables and avoid possible multicollinearity problems.

---

[4]For more details seehttps://www.gob.mx/consar/articulos/pension-por-regimen-73.

[5]This proportion was constructed with the available data, assuming a constantly decreasing rate of mortality. At the beginning we considered three characteristics but the variables were non-significant in many cases, maybe because the number of observation in the sample considering three variables was not enough.

2.- Tests of unitary roots for the explanatory variables, dependent variable and their first differences are done to know the nature of their stationarity.

3.- Cointegration tests are carried out following the methodology of Johansen (1991).

4.- Long-term effects are estimated following the methodology of Engle and Granger (1987).

5.- An error correction model is used to estimate the short-term effects.

For the tests of unit roots, the Dickey-Fuller test is carried out, which proposes the following model:

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t \quad (2)$$

$$\Delta y_t = \mu - (1-\rho)\, y_{t-1} + \varepsilon_t \quad (3)$$

Where μ and ρ are the parameters to be estimated and $\varepsilon_t$is the error term that is supposed to be "white noise", under the null hypothesis that $(1-\rho)$ is biased downwards.

If the proportions of retirement, unemployment rate, stock market and petroleum prices are stationary of the same order, cointegration tests are carried out following the methodology of Johansen (1991), otherwise, the relationship cannot be determined.

This approach considers the form:

$$\Delta y_t = \pi y_{t-1} + \sum_{i=1}^{p-1} \psi \Delta y_{t-i} + B x_t + \varepsilon_t \quad (4)$$

Where y_t is a vector of $k$nown stationary variables integrated of order 1, $I(1)$, x_t is a vector of d deterministic variables and ε_t it is a vector of innovations.

Johansen's method estimates the matrix π in a restricted form so that it analyses whether the implicit restrictions can be rejected by the reduced order of π. If the variables are effectively cointegrated, the estimation by OLS (Ordinary Least Squares) is consistent to estimate the long-term effects, Engle and Granger (1987). To estimate the short-term effects of the macroeconomic variables on the probability of retiring, an Error Correction Model (ECM) is used:

$$\Delta y_t = \sum_{i=1}^{n} \theta_i \, \Delta x_{it} + \delta \left(y_{t-1} - \alpha - \sum_{i=1}^{n} \beta_i x_{it-1} \right) + \eta_t \quad (5)$$

Where $\sum_{i=1}^{n} \theta_i \Delta x_{it}$ represents the adjustments for short-term movements and $\delta(y_{t-1} - \alpha - \sum_{i=1}^{n} \beta_i x_{it-1})$is the linear combination of long-term variables.

The coefficient of the second term δ, is the error correction coefficient that represents the speed of adjustment of the variables[6].

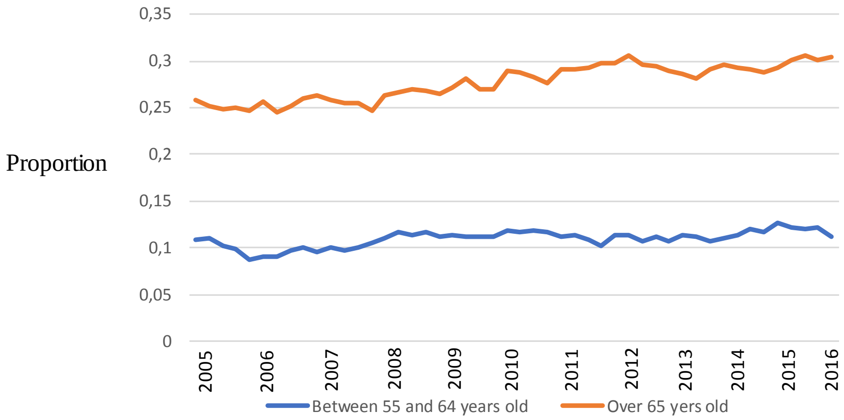
4. Data

The quarterly variables are converted into monthly following the methodology of Boot, Feibes and Lisman (1967). Based on this methodology we minimize the sum of the squares of the differences between the successive monthly values (estimated values), subject to the constraints that during each quarter the average of monthly values should equal the quarterly total (observed values). The daily variables were averaged to obtain a monthly value. Two sources of data are used for this analysis.

The National Employment and Occupation Survey (ENOE)

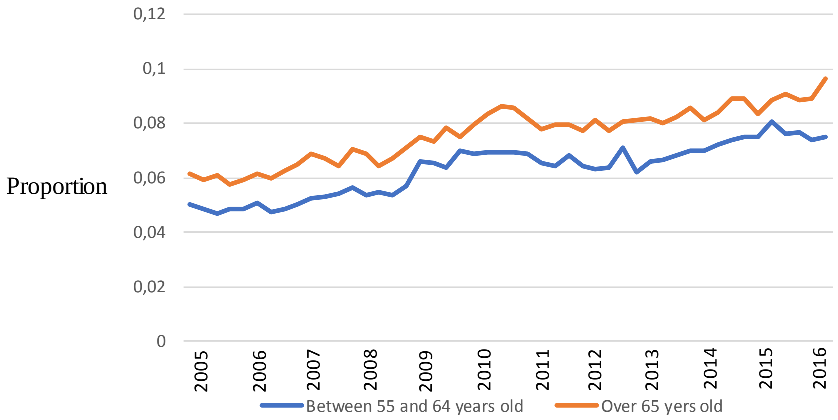
Since 2005, this survey is carried out by INEGI on a monthly basis. From this source, the number of retired people and population by age, gender and level of education from 2005 to 2016 are obtained quarterly. The ENOE divides the information of pensioners into 4 schooling ranges which are incomplete primary education, complete primary education, complete secondary education and high school or higher. In our analysis, incomplete primary education will be considered as incomplete elementary education and complete primary and secondary education will be considered as complete elementary education[7]. The ENOE presents information in different age ranges, but for the purposes of this study, attention is only given to pensioners over 55 years old. The pensioners will be divided into two groups, aged between 55 and 64 years old and over 65 years old. We divide into two groups to analyse if workers who decide to retire before 65 years old are affected in a different way than those who retire after 65 years old. Figures 1 and 2 show the evolution in the proportion of retired men and women by age range. As expected, the proportion of retired men and women over 65 years old is larger than those between 55 and 64 years old. This difference is more marked in men. It is also observed that the proportions of men are larger than the proportion of retired women. This is common for a culture such as Mexico where men will normally provide the main source of income to cover the expenses for the entire family. The proportion of retired women has an increasing trend no matter the age. The proportion of men, on the other hand, shows a growing trend but a bit more stable than women.

Figure 1. Proportion of retired men by age



Own elaboration. Quartely series from 2005 to 2016.

Figure 2. Proportion of retired women by age



Own elaboration. Quartely series from 2005 to 2016.

Figures 3 and 4 show the evolution in the proportion of retired men and women by educational level. The proportion of retirement is larger in those individuals with high school or higher. The smallest proportion is found among those individuals with incomplete primary education. This is because the population of individuals with incomplete primary education is larger than that one with high school or higher. The graph seems to suggest that the more level of schooling the more likely the person is to retire. This can be linked to the level of knowledge individual has about the benefits in their pension scheme.

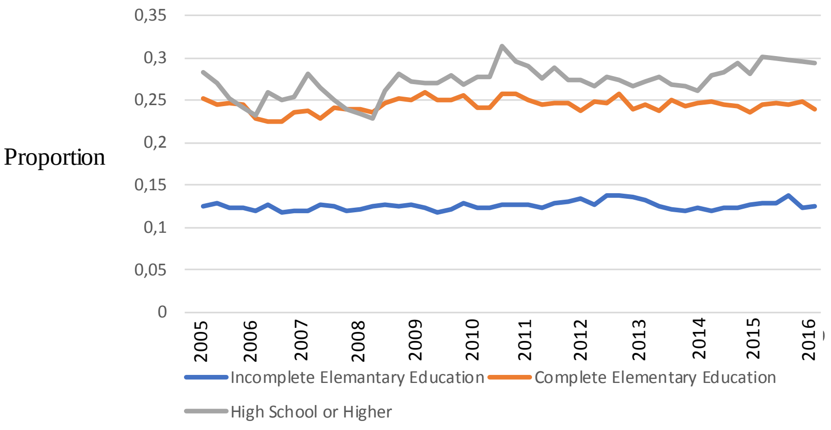
[6]When this coefficient is significant, the existence of cointegration is confirmed.

[7]When complete primary and secondary education were separately studied, no difference was observed.

People with more education have more knowledge about the requirements for retirement and are aware of their benefits and regulations and know the exact time they could retire. Also with higher levels of education come higher salaries and retirement needs will be met earlier in life.

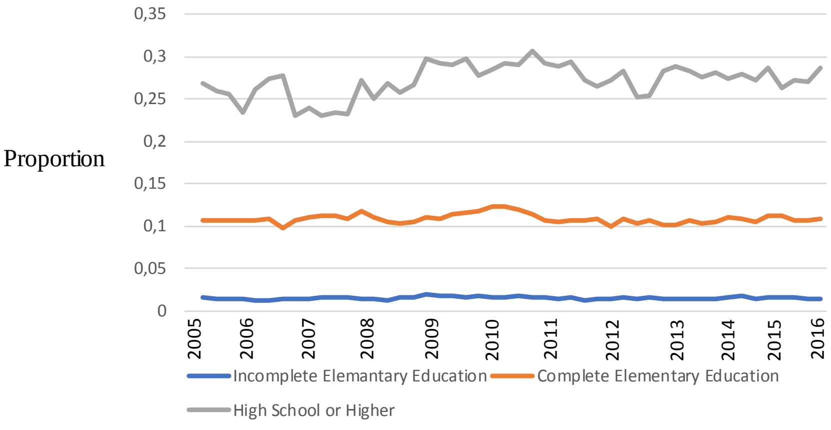
On the other hand, people with a lower level of education will lack knowledge about their benefits and will be more detached from the regulations of pension schemes. Also with lower levels of education come low salaries and retirement needs will be met later in life.

Figure 3. Proportion of retired men by level of education.



Own elaboration. Quartely series from 2005 to 2016.

Figure 4. Proportion of retired women by level of education.



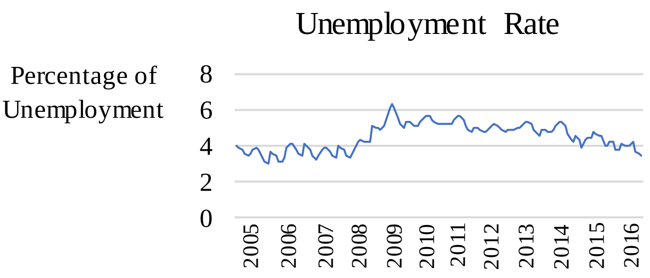
Own elaboration. Quartely series from 2005 to 2016.

It is observed that between 2010 and 2012 there was a slight increase in the proportions of retirees with high school or higher, women over 55 years old and men over 65 years old. This might be related to the crisis that began in the United States between 2008 and 2009. The proportion of retirees with incomplete and complete elementary education and men between 55 and 64 years old remained almost without effect.

Bank of Mexico (BM)

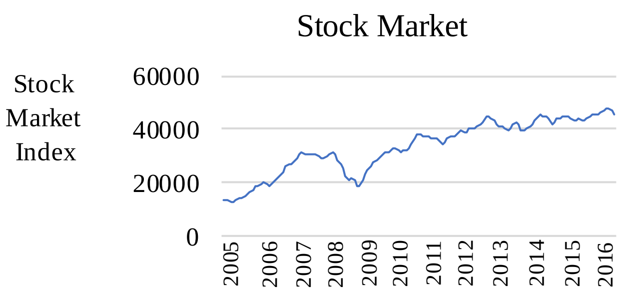
The Bank of Mexico generates and disseminates the statistics of the main macroeconomic and financial variables in Mexico on a daily, monthly or quarterly basis. From this database, we obtained[8] series of the unemployment rate, price index of the Mexican Stock Exchange (as stock market) and petroleum prices with a quarterly, monthly and daily frequency respectively. Petroleum prices are in US dollars and will be deflated using the US consumer price index. Figures 5, 6 and 7 show the trend of the explanatory variables over the period 2005-2016. It is observed that the stock market and petroleum prices seem to be non-stationary and unemployment rate stationary. These hypotheses will be proved using Dickey Fuller test for the unit root.

Figure 5



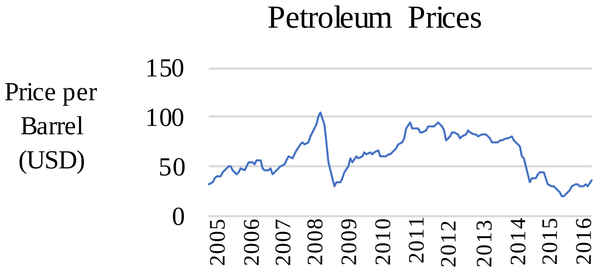
Own elaboration.

Figure 6



Own elaboration.

Figure 7



Own elaboration.

The unemployment rate remained below 4% and stable between 2005 and 2007. Between 2008 and 2009 there was a sharp increase in the unemployment rate to exceed 6%. Afterwards, it drops again until it reaches below 4% in 2016.

The stock price index has had an upward trend in the period analysed except between 2008 and 2009 where a fall in stock prices is observed.

Between 2005 and 2007 oil prices maintained a growing trend. Between 2008 and 2009 there was a marked fall in oil prices to below USD 50 per barrel. After this period, prices seem to increase until falling again after 2011.

It is observed that between 2008 and 2009 the three series reached their maximum and minimum peak in the period analysed. This effect coincides with the crisis experienced in the USA in those years that affected the related economies.

[8]At the beginning we considered to include interest rate and housing prices in our analysis. The strong correlation of these variables with the stock market makes us to keep just stock market in the analysis based on the literature review.

5. Results

Table 3 shows that all the series considered are not stationary but in their first difference, they are stationary. This means the series are integrated of order 1. Here it is verified that although in graphs some of the variables seemed stationary, the statistical tests showed that this is not the case. The fact that all series are integrated in the same order is vital to review if they are effectively cointegrated.

Table 3. Dickey-Fuller test for stationarity.

Variable	p-value for the original series, Z(t)	p-value for the first difference, Z(t)
Men between 55 and 64 years old	0.5204	0.0000
Men over 65 years old	0.8684	0.0000
Women between 55 and 64 years old	0.8026	0.0000
Women over 65 years old	0.9628	0.0001
Men with incomplete elementary education	0.1403	0.0000
Men with complete elementary education	0.0843	0.0000
Men with high school or higher	0.4188	0.0000
Women with incomplete elementary education	0.0525	0.0000
Women with complete elementary education	0.2068	0.0000
Women with high school or higher	0.2568	0.0000
Unemployment Rate	0.3949	0.0000.
Stock Market	0.3972	0.0000
Petroleum Prices	0.5532	0.0000

Own elaboration. Test made in Stata software. The independent variables are the proportion of retired individuals with the characteristics mentioned.

The results of the cointegration test are shown in table 4. It is observed that all the combination of variables are cointegrated since they have rank 1, one vector of cointegration. This shows that there is indeed a causal relationship between the variables in the long term. With this, it is possible to use an error correction model to obtain the short-term effects.

Table 4. Johansen tests for cointegration.

Combination of variables	Maximum rank	Trace statistic	CriticalValue 5%
Men between 55 and 64 years old-Unemployment Rate-Stock Market-Petroleum Prices	1	19.9231*	29.68
Men over 65 years old-Unemployment Rate-Stock Market-Petroleum Prices	1	24.5358*	29.68
Women between 55 and 64 years old-Unemployment Rate-Stock Market-Petroleum Prices	1	19.6030*	29.68
Women over 65 years old-Unemployment Rate-Stock Market-Petroleum Prices	1	21.6787*	29.68
Men with incomplete elementary education-Unemployment Rate-Stock Market-Petroleum Prices	1	24.3349*	29.68
Men with complete elementary education-Unemployment Rate-Stock Market-Petroleum Prices	1	25.5193*	29.68
Men with high school or higher-Unemployment Rate-Stock Market-Petroleum Prices	1	22.4731*	29.68
Women with incomplete elementary education-Unemployment Rate-Stock Market-Petroleum Prices	1	21.0928*	29.68
Women with complete elementary education-Unemployment Rate-Stock Market-Petroleum Prices	1	22.2396*	29.68
Women with high school or higher-Unemployment Rate-Stock Market-Petroleum Prices	1	23.8656*	29.68

Own elaboration. When the trace value is less than the critical value the hypothesis of maximum rank is accepted.

The results of the estimates are shown below in the following tables. It is important to mention that in several cases the coefficients of the variables seem to be very small, due to the proportions we take as dependent variables are very small. Also it is important to say that to interpret the results we focus just on the sign of coefficients and not the magnitudes (we want to see just the direction of the effect and not its magnitude).

Table 5. Long-term effects on the proportion of retired people by gender and age.

	Unemployment Rate	Stock Market	Petroleum Prices	Unemployment Rate	Stock Market	Petroleum Prices
	Men			**Women**		
From 55 to 64 years old	.0060384***	3.80e-07***	-.000167***	.0048007***	7.55e-07***	-.0001518***
	(7.54)	(6.87)	(-6.15)	(10.39)	(23.63)	(-9.67)
Over 65 years old	.0050913***	1.59e-06***	-.0001384***	.0039252***	8.57e-07***	-.0001307***
	(4.89)	(22.11)	(-3.92)	(8.31)	(26.22)	(-8.15)

Own elaboration. The symbol "" expresses that the coefficient is significant at 10%, "**" significant at 5% and "***" significant at 1%. The value in parentheses is the t statistic.*

Table 6. Short-term effects on the proportion of retired people by gender and age.

	ΔUnemployment Rate	ΔStock Market	ΔPetroleum Prices	ECT
	Men			
From 55 to 64 years old	-.0009579	-2.12e-07	-.0000233	-.0633231**
	(-1.23)	(-0.99)	(-0.56)	(-2.01)
Over 65 years old	-.0020413**	3.78e-07	-.000041	-.0838594***
	(-2.05)	(1.37)	(-0.75)	(-2.73)

Own elaboration. The symbol "" expresses that the coefficient is significant at 10%, "**" significant at 5% and "***" significant at 1%. The value in parentheses is the t statistic. ECT is the Error Correction Term.*

Table 7. Short-term effects on the proportion of retired people by gender and age.

	ΔUnemployment Rate	ΔStock Market	ΔPetroleum Prices	ECT
		Women		
From 55 to 64 years old	-.0000519	9.47e-08	.0000374	-0.0872553***
	(-0.11)	(0.73)	(1.47)	(-2.65)
Over 65 years old	-.0008502*	-6.95e-08	6.89e-06	-0.0895238**
	(-1.80)	(-0.54)	(0.26)	(-2.58)

Own elaboration. The symbol "" expresses that the coefficient is significant at 10%, "**" significant at 5% and "***" significant at 1%. The value in parentheses is the t statistic. ECT is the Error Correction Term.*

When we analyse the proportion of retired individuals by gender and age, as shown in table 5, it is observed that in the long term, the unemployment rate has a positive effect on the proportion of retired men and women, no matter the age of retirement. This means that when the unemployment rate increases, the proportion of new pensioners increases and vice versa. As previously explained, this can be linked to the fact that when there is an increase in unemployment, affecting mainly elderly people, individuals decide to retire because there are no jobs. The stock market has a similar effect on the proportion of retired individuals. When the stock market increases, the proportion of retired individuals increases and vice versa. This is explained by the effect of the crisis that began in the United States between 2008 and 2009. When the stock market fell, people considered working longer to recover their lost wealth. Thus, when there is an increase in the stock market, there is no lost wealth and individuals have no interest in working for a longer period of time. Oil prices have a negative impact on the proportion of retired individuals. This means that when oil prices increase, the proportion of new retirees decreases and vice versa. When oil prices are high, the country receives more income and has the opportunity to invest in social programs that distribute wealth among individuals. This way, individuals have more wealth, the economy is more stable and there is no interest in early retirement. In all cases, the coefficient was significant at 1%.

In the short term, as shown in tables 6 and 7, only the unemployment rate has a negative effect on the proportion of retired men and women over 65 years old with 5% and 10%of significance respectively. Individuals between 55 and 64 years old do not react to economic fluctuations in the short term.

Table 8. Long-term effects on the proportion of retired people by gender and level of education.

	Unemployment Rate	Stock Market	Petroleum Prices	Unemployment Rate	Stock Market	Petroleum Prices
	Men			**Women**		
Incomplete Elementary Education	-.0006012	1.76e-07***	.0000371*	.0013118***	-5.50e-09	-.0000342***
	(-0.97)	(4.10)	(1.76)	(7.24)	(-0.44)	(-5.56)
Complete Elementary Education	.006617***	-2.37e-08	-.000061**	.0016769**	-3.62e-08	-.000024
	(7.50)	(-0.39)	(-2.04)	(2.37)	(-0.74)	(-1.00)
High School or Higher	.0108124***	8.43e-07***	-.0004633***	.0161058***	2.49e-07*	-.0000572
	(6.33)	(7.13)	(-7.99)	(8.15)	(1.82)	(-0.85)

Own elaboration. The symbol "" expresses that the coefficient is significant at 10%, "**" significant at 5% and "***" significant at 1%. The value in parentheses is the t statistic.*

Table 9. Short-term effects on the proportion of retired people by gender and level of education.

	ΔUnemployment Rate	ΔStock Market	ΔPetroleum Prices	ECT
		Men		
Incomplete Elementary Education	-.0019783***	-8.04e-08	.000038	-.0885679**
	(-2.95)	(-0.43)	(1.03)	(-2.56)
Complete Elementary Education	.0017255	1.50e-07	-7.54e-06	-.141031***
	(1.50)	(0.47)	(-0.12)	(-3.37)
High School or Higher	-.0046261**	1.10e-07	-.0002318**	-.1118527***
	(-2.48)	(0.21)	(-2.27)	(-3.20)

Own elaboration. The symbol "" expresses that the coefficient is significant at 10%, "**" significant at 5% and "***" significant at 1%. The value in parentheses is the t statistic. ECT is the Error Correction Term.*

Table 10. Short-term effects on the proportion of retired people by gender and level of education.

	ΔUnemployment Rate	ΔStock Market	ΔPetroleum Prices	ECT
		Women		
Incomplete Elementary Education	-.000559**	6.13e-09	.0000122	-.176771***
	(-2.34)	(0.10)	(0.92)	(-4.03)
Complete Elementary Education	-.0007802	7.61e-08	.0000251	-.0875397***
	(-1.15)	(0.40)	(0.63)	(-2.69)
High School or Higher	.0060549**	-6.22e-07	.0003586***	-.0909926**
	(2.59)	(-0.96)	(2.81)	(-2.38)

Own elaboration. The symbol "" expresses that the coefficient is significant at 10%, "**" significant at 5% and "***" significant at 1%. The value in parentheses is the t statistic. ECT is the Error Correction Term.*

When the analysis is done by gender and level of education, the results are more varied. In the long term, as shown in table 8, the unemployment rate affects positively the proportion of retired men and women, except men with incomplete elementary education. The stock market has a positive effect on the proportion of men and women retired with high school or higher education, and men with incomplete elementary education. Oil prices have a negative effect on the proportion of retired men with complete elementary education and high school or higher. For men with incomplete elementary education, the effect of oil prices is positive. As for the proportion of retired women, oil prices have a negative effect on those with incomplete elementary education. The effects were similar to those obtained when the proportion is analysed by gender and age.

In the short term, as shown in tables 9 and 10, the unemployment rate has a negative impact on the proportion of retired men with incomplete elementary education and high school or higher and the proportion of retired women with incomplete elementary education.

The unemployment rate has a positive effect on the proportion of retired women with high school or higher. Oil prices have a negative impact on the proportion of retired men with high school or higher and positive on women with the same level of education. The stock market has no impact on retirement. Individuals with complete elementary education do not react to economic fluctuations in the short term.

Some short-term effects were different from those obtained in the long term. This might be due to the process of adjustment that individuals make in their decisions to face a change in macroeconomic variables.

6. Conclusions

It was analysed how the proportion of retired men and women with certain age ranges and levels of education reacts to fluctuations in the unemployment rate, stock market and oil prices. In this study, the cointegration theory was used to determine these effects.

The theory of cointegration requires the series to be integrated of the same order and the existence of cointegration vectors. It was proved that all the series are integrated of order 1 and the combinations of the proportions and the explanatory variables are cointegrated. It was shown that there is at least one vector of cointegration among the variables.

Regarding the results, it was obtained that the unemployment rate has an impact on the retirement of men and women in the short and long term. The stock market and oil prices have an impact mainly in the long term. Particularly, oil prices have a negative effect on the proportion of retirement and affect mainly to men.

In the long term, the effects of the unemployment rate and stock market in the proportion of retired individuals coincide with the results obtained by previous studies. In the short term, the effects of these variables change.

In recent years, it has been observed that oil prices have had several fluctuations, mainly downward. This is due to geopolitical problems and an increase in the production of petroleum. As shown in this article, this can influence the retirement decisions of workers in the short and long term.

The government should consider this effect to take measures to ensure the sustainability of the system when petroleum prices fluctuate outside the normal pattern.

Future research aims to analyse the pattern of new retirees' decisions for other oil-producing countries with different economic and political conditions such as Norway.

2

Machine learning techniques: Application to the pension industry

Jose de Jesus Rocha Salazar[1] · Maria del Carmen Boado Penas[2]

Abstract: Artificial intelligence techniques have become very popular in different public and private organizations because they allow the development of software capable of being adapted to rapid change. Private insurance companies are not the exception and have ventured into this new field by implementing algorithms that allow a better understanding of available data to make forecasts. The knowledge of retirement decisions allows the insurance companies to detect retirement at a given time so that they have the adequate budgetary provision in place. In this paper, using data from a Mexican Insurance Company, we apply machine learning algorithms to predict whether a person retires before or after 65 years old in function of both individual characteristics and macroeconomic factors.

Keywords: insurance company; machine learning; Mexico; retirement; supervised learning

1. Introduction

Machine learning algorithms allow the analysis of Big Data and at the same time automate its processing and pattern finding. One of the main uses of machine learning techniques is to create predictive models. Automating the processing of large amounts of data allows algorithms to create more reliable and accurate predictive models, Talwar and Kumar (2013).

It is true that organizations do not need to have a large amount of data to work with machine learning models, however, a large amount of data improves their accuracy. This is why a data culture has awakened in companies and they invest part of their capital in reliable data collection, McAfee and Brynjolfsson(2012).

Many machine learning algorithms are online and update as new data are entered. For example, the platforms of Amazon and Netflix predict the preferences of the customers based on their previous choices, Bell et al. (2008). Models for prediction of weather changes also update continuously online as the environmental conditions change, Alavi, Gandomy and Larry (2016). The accuracy of these models is the result of the training process and automation that is part of machine learning.

Within the financial sector, machine learning algorithms such as decision tree and neural networks have been used to detect credit card fraud. These algorithms help companies to minimize the loss from this financial crime, Bolton and Hand (2001), Delamaire et al. (2009), Juszczak et al. (2008) and Pozzolo et al. (2014).

Studies in the insurance sector have used artificial neural networks to evaluate the financial capability of insurance companies and predict insolvency Olaniyi et al. (2012). Other studies have used feedforward neural networks with the back-propagation algorithm to build decision models for five insurances including life, annuity, health, accident, and investment-oriented insurances Lin et al. (2008). Machine learning algorithms also have been used to analyze the quality of the mortality models Deprez et al. (2017).

In the pension field, various studies have been carried out to determine the explanatory factors on retirement decisions. Some studies explain that macroeconomic factors such as unemployment rate and stock market affect the decision of people to advance their retirement or delay it, Bosworth and Burtless (2010) and Coile and Levine (2011). Other articles show that personality traits can predict the timing and routes of people's retirement, Blekesaune and Skirbekk (2012) and Feldman and Beehr (2011).

In the context of a private company such as an insurance company which offers retirement plans, the retirement decisions are of high interest because an early retirement involves fewer contributions to the system, less investment, and therefore less net earnings. Besides insurance, companies need to predict when and how much the budgetary provision will be needed to make payments almost immediately when the pensions are claimed.

Ideally for an insurance company, is that their clients contribute to the system as much as possible because it implies more investments and profits. For public pensions, the late retirement is also beneficial because it implies contributions for longer and fewer liabilities. For many reasons, this is not always possible and many people decide to retire early from the system and claim their accumulated pension.

In the 1960s the labour rate of participation for people over 60 years old was above 70% for OECD countries. In recent decades the rate of participation has decreased even below 20%. For example, in the United Kingdom, the average age of retirement for men and women passed from 67.2 to 62.7 years old from 1950 to 1995, Blöndal and Scarpetta (1999).

There are several causes of early retirement, for example, workers aged 55-64 due to their low level of schooling and old skills are more likely to be removed from the labour market and replaced by younger ones Fallick (1996) and Munnell et al. (2006).

The policies and conditions for the claim of pensions can encourage early retirement since people tend to claim their pension as soon as they meet the conditions of age and time worked as established by Muldoon and Kopcke (2008). Also, the absence of recessions and the presence of a stable economic environment, stimulate an early retirement since workers have no personal financial reasons to stay longer, Hurd and Rohwedder(2010).

This phenomenon not only affects the private insurance companies which are fed from returns of contributions but also the society in terms of lower productive capacity, Herbertsson (2001).

Despite these studies to understand the early retirement, a methodology and computing system has not been developed to predict the retirement of people in real time and act on it by implementing it, in the internal systems of a company using machine learning theory.

The purpose of this paper is developing an alert system using machine learning techniques to predict whether an individual with certain characteristics and macroeconomic conditions is more likely to retire before or after 65 years old[9]. Even machine learning algorithms are often used in actuarial science; this is the first time these methodologies are used on retirement.

The paper uses data offered by a private insurance company in Mexico but the model could also be applied to the public sector.

After this introduction, the paper is structured as follows. Section 2 gives summarised explanations about the concept and use of machine learning. In section 3 a brief explanation is given about the risk of early retirement faced by insurance companies. In section 4 the data used in the analysis is presented. Section 5 presents the models of machine learning used on the prediction. In section 6 the results are shown. Section 7 provides conclusions.

[9] 65 years old is the normal statutory age of retirement in the country where the data are taking from.

2. Machine Learning Methodologies

With the presence of a globalized world and the advancement of technology in various disciplines, artificial intelligence has become a vital element in this contemporary era. Within artificial intelligence, machine learning has become a powerful tool in the manufacturing industry, financial institutions and services that work with large amounts of data. Previously these companies processed information with data available at that time and calibrated models in the systems for that period of time. Subsequently, there was no updating of the model despite the fact that new data was constantly being generated. The innovative idea of machine learning is that methodologies are designed for computers to learn constantly as data is generated over time. This makes predictions of the future more accurately.

The use of machine learning tools is a complex process that requires the intervention of engineers, business analysts, programmers, statisticians, etc. Machine learning uses various mathematical models, heuristic learning and knowledge acquisitions for decision making and prediction. This makes it a more versatile methodology and adaptable for different disciplines.

The concept of machine learning has its origin in the 1950s. Arthur Lee developed an algorithm in machine learning programs for playing checkers in 1952. In 1957 Frank Rosenblatt developed the first model of neural networks that simulated the processing of thoughts in the human brain. In 1967 the "nearest behaviour" algorithm was developed and allowed computers to recognize patterns. In the 90s researchers began using machine learning to analyze large amounts of data. In the year 2006, Geoffrey Hinton coined the term "deep learning" to explain new algorithms that let computers "see" and distinguish objects and text in images and videos, Hinton et al. (2006). In 2014 Facebook developed DeepFace, a software algorithm that is able to recognize or verify individuals on photos at the same level as humans can, Taigman et al. (2014).

In recent years and thanks to the greater storage capacity of processors, the applications of machine learning algorithms have focused on the development of predictive models that are constantly fed with Big Data. Big Data is a concept linked to Machine Learning.

Big Data is any data source that has the following characteristics: it is extremely large in volume, it can be processed at very high speed by means of machine learning algorithms, it is constantly expanding in variety and size and truthfully represents reality. It is not strictly necessary to have a large database to work with machine learning algorithms, but it is true that the larger the database the more accurate the prediction will be.

The vital part of machine learning methodologies is learning. It is assumed that there exists a function $"f"$ applied to a set of data and the objective of the apprentice system is to guess what type of function it is. A hypothesis is created about the function to be learned, let is say $"h"$. Both $"f"$ and $"h"$ are evaluated in the input vector $x = (x_1, x_2, \ldots, x_n)$ with $"n"$ components. It is understood that *h(x)* is the function to be estimated and implemented by the system that has as input the vector *x*. The creation of $"h"$ is based on a set of data called training set with $"m"$ input vectors. In several cases, there is also a test set to evaluate the fit and accuracy of the $"h"$ function with respect to the $"f"$ function.

The learning process in machine learning can be done in two ways: supervised learning and unsupervised learning. In supervised learning, we have some real values of the $"f"$ function for the $"m"$ input vectors of the training set. The goal here is to fit a function $"h"$ to the function $"f"$ that allows us to label the training set in a precise way. Common algorithms in this type of learning are support vector machine, random forest and logit regression. They can be used for example in the data on banking transactions previously identified as fraudulent or non-fraudulent to create a function that classifies future transactions and minimize the risk of fraud, Perols (2011).
In unsupervised learning, values of the function $"f"$ are not available for the training set. The objective in this type of learning is reduced to splitting the training set into subsets where the member data share similar taxonomic characteristics. Common algorithms in this type of learning are k-means for clustering problems and apriori algorithm for association rule learning problems. They can be used for example in the information generated in social networks through comments, blogs, reactions, etc. This information can be classified as positive or negative identifying their similar characteristics, Nivedha and Sairam (2015).

The methodologies of machine learning have taken force in different fields of the industry mainly in those dedicated to information technologies. In data processing, machine learning has some differences and advantages over traditional econometric models from a computational point of view. Econometrics focuses more on finding causality relationships between the variables by estimating partial correlations (ceteris paribus) and under certain statistical assumptions. Machine learning focuses on the prediction and classification using data, very computational not necessarily in a statistical way. Machine learning is not very focused on finding causality between the variables, Zheng et al. (2017).

A new trend is arising in which it is recommended to use both econometric and machine learning disciplines in an integral way to strengthen the result on prediction both in short and long term. There is evidence that machine learning may be better at making short-term predictions and econometric techniques turn out to be better in the long term. This is because machine learning methodologies deal with the heterogeneity of data and therefore are better at capturing short-term predictions. Econometrics, on the other hand, is better with long-run trends, i.e. linear or regular patterns, Liu and Xie (2018).

In the case of this study, only machine learning algorithms will be used to make predictions since the objective of the model is purely of computational interest to create an alert system.

3. Data

Factors such as economic policies, financial situation, state of health, working environment, etc. are all taken into account in studying the retirement decisions of workers. These factors act as forces which push individuals to their retirement decisions, Shultz et al. (1998).

The early retirement is a personal choice influenced mainly by high replacement rates with more leisure time, high preferences for retirement, health state and lay-off risk, Heyma (2004).

In the case of this study, the data of individual characteristics were given for an insurance company in Mexico[10] and the macroeconomic environment of the country was also considered. We analyse a sample of 1500 claims of retirement over a period of 12 years, from 2005 to 2017. These claims coincide with the claims of the Mexican Institute of Social Security (IMSS) pensions[11]. This indicates that the sample is composed of definitive retirement. The information was given by the department of systems and selected randomly. The data will be divided into two parts: 70% for the training set and 30% for the test set. The sample for the test set is useful to evaluate the adjustments of the models. The split is made randomly thereby each model considers different training and test sets.

The dependent variable is a binary variable that takes 1 if the person retired before 65 years old and 0 otherwise.

[10]This insurance company offers occupational and individual plans. Our sample takes into account both plans.

[11]IMSS is the institution in Mexico which manages the public pensions governed by the law of 1973.

In table 1, the variables used on the training and test sets are shown.

Table 1. Variables used in the model:

Variable	Representation
Gender	This variable takes value 1 if the person is male and 0 female.
Disease	This variable is 1 if the person has a disease related to the 10 most common mortal diseases in Mexico and 0 otherwise.
Level of Education	1 means the person has elementary or secondary education, 2 high school, 3 graduate studies and 4 postgraduate studies.
Marital Status	1 if the person is married, 0 otherwise.
Salary	The last salary obtained at the moment of retirement.
Employment Status	This variable takes value 1 if the person is an employer and 0 if the person is an employee.
Dependants	The number of people who depend economically from the pensioner.
Unemployment Rate	Unemployment rate at the moment of retirement.
Credit Score	The score fluctuates between 400 and 850, 850 is the best rating.
Return of Government Bonds	Return of the government bonds at the moment of retirement.
Stock Market	Stock Market at the moment of retirement.

Own elaboration. A sample of claims in private retirement plans. The gender, Disease, Level of Education, Marital Status, Salary, Employment Status, Credit Score and Dependents are obtained from the Insurance Company. The Unemployment Rate, Return of Government Bonds and Stock Market are obtained from the database of Bank of Mexico.

The total sample has 909 men and 591 women, 864 individuals are married and 602 individuals are employers.

Disease

The most common mortal diseases in Mexico are the following according to INEGI statistics[12]:

1. Heart diseases
2. Diabetes Mellitus
3. Cancer
4. Liver disease
5. Cerebral-vascular illnesses
6. Lung diseases
7. Pneumonia
8. kidney failure
9. Congenital malformations

10. Bronchitis

640 individuals in the sample have a disease related to the 10 most common mortal diseases in Mexico. It is expected that individuals with an illness related to the above list will be more likely to retire and leave the labour market.

Salary

Table 2 shows some statistics on the salary of individuals considered in the sample. The average salary is 25,053.33 Mexican Pesos (MXN) a month, which is around 1,300.00 American Dollars a month. The salary in the sample varies from 15,000.00 to 40,000.00 MXN, which is considered the salary of the middle class and upper middle class. It is important to mention that around 50% of the population in Mexico has a salary between 2,400.00 and 7,200.00 MXN[13] a month.

Table 2

Monthly Salary	Obs	Mean	Std. Dev.	Min	Max
	1,500.00	25,053.33	7,816.17	15,000.00	40,000.00

Own elaboration.

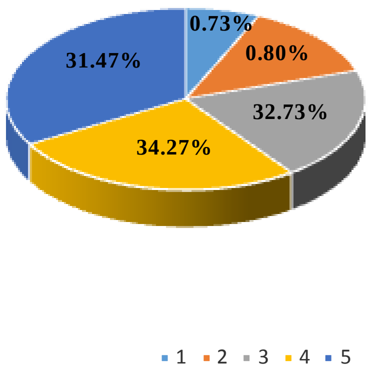
It is expected that workers with high salary remain in the labour market for longer because they will prefer the current lifestyle to one with a pension lower than the salary.

Dependants

Graph 1shows the number of dependants of individuals in the sample. 34.71% of individuals have 4 dependants while only 0.73% has 1 dependant. This might be related to the level of salary. Usually, people with higher salary tend to have more children. At first glance it seems to be a large number of dependents compared to European countries or northern America. The sample considers people who have retired in the last 12 years, people who were born in the fifties and who began to procreate in the sixties and seventies. In Mexico, according to statistics from the National Population Council (CONAPO), the average number of children per woman in the seventies was 6.1.

Taking this into account, we see that the number of dependents in the sample is below the average for that generation.

Graph 1. Number of dependants in the sample.



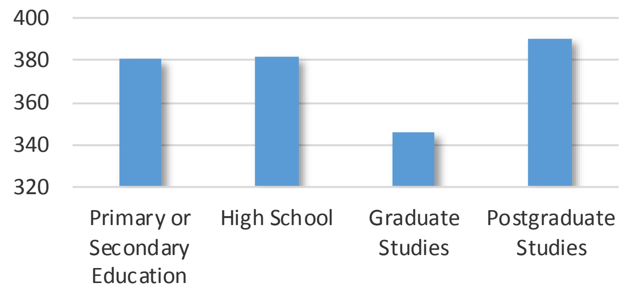
Own elaboration.

It is expected that the more the number of dependants the longer the workers will remain in the labour market because of the responsibility of providing economic support.

Level of Education

Graph 2 shows how the individuals are distributed by education level. It is observed that most of the workers in our sample have postgraduate studies. Also, workers with elementary education or high school have a high frequency. According to statistics from the National Employment and Occupation Survey (ENOE), the generations of the sixties and seventies used to finish only elementary education and high school. This explains the high frequency of individuals with only basic education. Over the years and as Mexico grew economically and became involved in a globalized and competitive world, individuals chose to acquire higher educational degrees. On the other hand, the high frequency of individuals with postgraduate studies in the sample may be due to the profile that is commonly associated with a private insurance company. Individuals with higher educational levels tend to have better jobs, higher income and therefore have the possibility of choosing a private pension system.

Graph 2. Number of individuals by Level of Education.



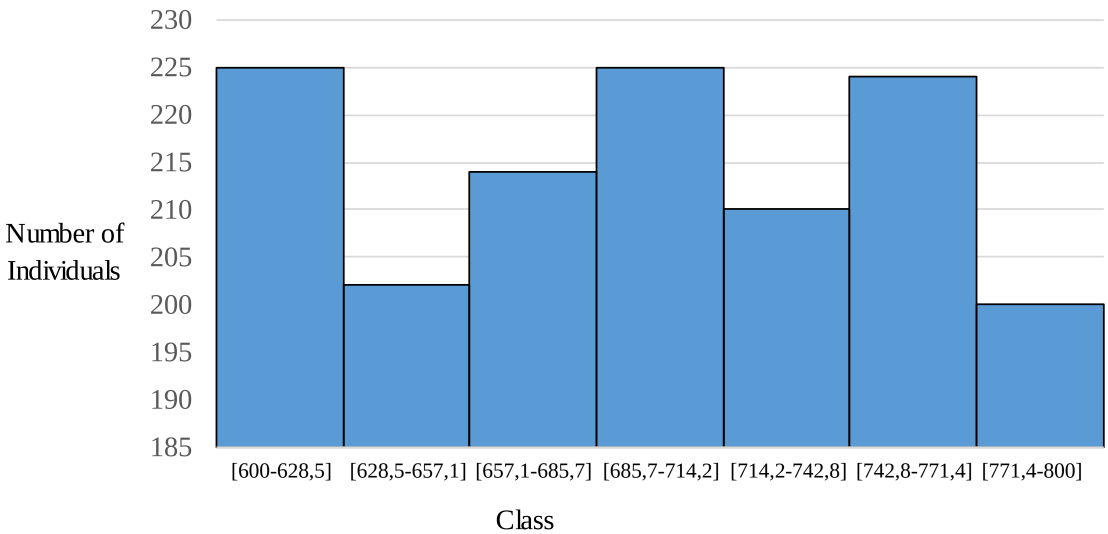
Own elaboration.

Credit Score

This variable was included as a measure of an individual's financial situation. The higher the score the better the financial health of workers. It can be observed in Graph 3 that all the workers have a score between 600 and 800. This reflects good credit behaviour. Only 25% of the population in Mexico used to be in this range[14]. This variable is commonly used in the banking sector to predict the credit default of consumers and companies. Individuals and companies with a high probability of default are rejected or given smaller credits. A consumer with a high credit risk at some point in his life has not had sufficient solvency or liquidity to face his financial obligations. Following this logic, an individual with a high risk of default is more likely to stop contributing to the private pension system and unsubscribe, Madeira (2003).

[12]For more details see the statistics of INEGI, http://www.inegi.org.mx/est/contenidos/proyectos/registros/vitales/mortalidad/tabulados/ConsultaMortalidad.asp.
[13]For more details see the statistics of income and expenditure of the families, http://www.beta.inegi.org.mx/proyectos/enchogares/regulares/enigh/nc/2016/default.html.

Graph 3. Score of Individuals.

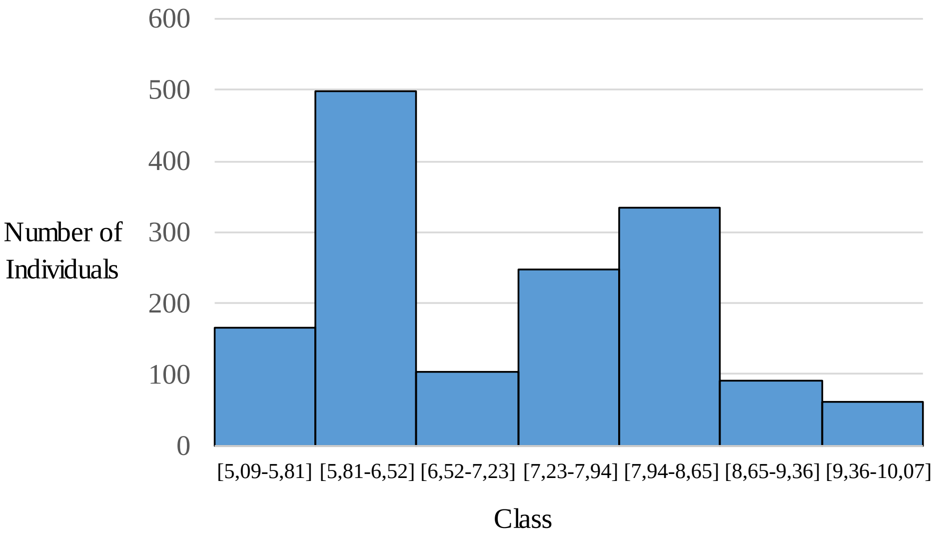


Own elaboration. The horizontal axis shows the classes of Credit Score.

Government Bonds

Graph 4 shows the return of government bonds of individuals in the sample. It has various fluctuations that vary in a range of between 6% and 8%. This variable was introduced as an indicator of a fixed rate.

Graph 4. Return of Government Bonds.



Own elaboration. The horizontal axis shows the classes of Government Bond Returns.

[14]For more details see the interpretation of the score, https://www.burodecredito.com.mx/.

Unemployment Rate and Stock Market

These variables were included as macroeconomic indicators that can influence retirement decisions. Bosworth and Burtless (2010) explain that when the unemployment rate increases people tend to retire early because there are no jobs in the market. Coile and Levine (2011) show that fluctuations in the stock market have an impact on the retirement decisions of workers with a high level of education.

In summary, we can say that our sample is composed of people with high income, good credit behaviour and a high level of education. This is expected because being insured by a private institution requires the capacity to save in a larger proportion. People with lower wages could hardly keep contributing to the private system. The only viable option for them is the public scheme.

4. Model

This section shows the supervised machine learning algorithms used to predict early retirement. In the supervised learning two kinds of data are used, the training set and the test set. The idea for the learner or computer is to learn from the training set to develop a rule that classifies new examples such as the test set.

Formally a supervised learning consists of a set of n ordered pairs $(x_{1,}\, y_1), \dots, (x_{n,}\, y_n)$, where x_i is a vector of characteristics and y_i is the label of that vector. For example, in our case the vector x_i represents the characteristics or attributes of retired people, x_i=(Gender, Disease, Level of Education, Salary, Employment Status, Dependants, Unemployment Rate, Credit Score, Return of Government Bonds, Stock Market) and y_i is the classification of the pensioner, "retired before 65 years old", "retired at 65 or after". The test data is another set of m vectors without classification: $(x_{n+1}, \dots, x_{n+m})$. As described above, the goal is to make a model to label the test set as “retired before 65 years old”, “retired at 65 or after”.

This paper is focused on three models of supervised learning, random forest, logistic regression and support vector machine. These three models are the most used for binary classification problems.

Random Forest

Random forest is a classifier that consists of a collection of tree-structured classifiers *{h(x, k), k = 1,...}* where the *{k}* are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input *x*, Breiman (2001). It is a substantial modification of bagging that builds a large collection of de-correlated trees and then averages them. In many ways, the performance of random forests is very similar to boosting, but they are simpler to train and tune.

The random forest technique can be used in two ways, to estimate a regression or classification. The regression is used when the output variable is continuous; the classification function is used when the output variable is categorical. For example, if we want to determine or predict the systolic pressure of a person based on height, weight and age, a regression would be used. If we want to determine if a person will retire before or after 65 years old (as is our case, yes/no) depending on his gender, health status, monthly salary, etc., the classification function would be used.

As the name states, the objective of this technique is to create many trees randomly much like a forest. The more trees are created the more accurate the classification will be. The trees have the same root. Random forest is composed of two stages. The first stage consists in the creation of the forest. The second stage consists in making predictions based on the forest.

This technique has some advantages over other classification models. First, if the variable is continuous the same algorithm and trees can be used. Second, the random forest methodology identifies the most important and relevant variables for the prediction.

The random forest algorithm has been extensively used in various disciplines for its accuracy in classification and processing speed, Du et al. (2015).

In the banking sector, a random forest has been used to identify the loyal consumer and the defaulting consumer. The loyal consumer is one who obtains large amounts of loans and pays appropriately. On the other hand, the defaulting client is the one who does not pay a loan appropriately in the established times and amounts, Mu J et al. (2013). Banks can use the individual characteristics of consumers and behaviour patterns in past credits to discriminate between good and bad customers before granting a loan, Akbari S. and Mardukhi F. (2014). In medicine, a random forest is used to predict Alzheimer disease using several Magnetic Resonance Imaging (MRI) measures, Lebedev A.V. et al. (2014). In e-commerce, the random forest has been useful to give recommendations to clients about products that they might like depending on other customers who have bought similar products, Joshi R. et al. (2018).

Logistic Regression

The logistic regression is a generalized linear model of binary response, widely used at the actuarial level, McCullagh and Nelder (1989). The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression generates the coefficients (and its standard errors and significance levels) of a formula to predict a *logit transformation* of the probability of the presence of the characteristic of interest.
This model starts with the following.
Let y_i , u_i and x_i be the binary variable, error term and vector of explanatory variables respectively. We define:

$$P_i = Prob(y_i = 1) = Prob\left[u_i > -\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{i,j}\right)\right]$$

$$= 1 - F\left[-\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{i,j}\right)\right] = F\left(\beta_0 + \sum_{j=1}^{k} \beta_j x_{i,j}\right) \quad (1)$$

Where F is the cumulative distribution function of u_i and P_i is the probability of retiring before 65 years old. $y_i = 1$ if individuals retire before 65 years old and 0 otherwise.
The model can be written as:

$$Log\left(\frac{P_i}{1-P_i}\right) = Logit(P_i) = \beta_0 + \sum\nolimits_{j=1}^{k} \beta_j x_{i,j} \quad (2)$$

We can write the model in terms of odds:

$$\left(\frac{P_i}{1-P_i}\right) = exp\left(\beta_0 + \sum\nolimits_{j=1}^{k} \beta_j x_{i,j}\right) \quad (3)$$

Or in terms of probability of early retirement occurring as:

$$P_i = \frac{exp\{x_i'\beta\}}{1+exp\{x_i'\beta\}} \quad (4)$$

Where P_i is the probability of early retirement, x_i the vector of explanatory variables and β the parameters to be estimated.

The idea of logistic regression arose knowing that proportions and probabilities take values between 0 and 1. Therefore the assumptions of normality over proportion become unusable and the binomial distribution is the one that best fits this type of variable.

One of the advantages of the logistic regression is that it does not require all the assumptions of the linear regressions to generate good estimators. It is also possible to work with non-linear relationships between the dependent variable and the explanatory variables since a logarithmic transformation of the linear regression is used.

In the education sector, the logistic regression has been used to predict whether a student is a slow learner or not, Peng, Lee and Ingersoll (2002). Logistic regression also has been used in the financial sector to predict bankruptcy, Ohlson (1980). In the insurance sector, the logistic regression has been used to determine the probability of claiming an insurance policy based on the characteristics of the individual, Viaene et al. (2007). The last application is useful to give preferential prices to individuals with low risk and offer a high price to those individuals with high claim risk.

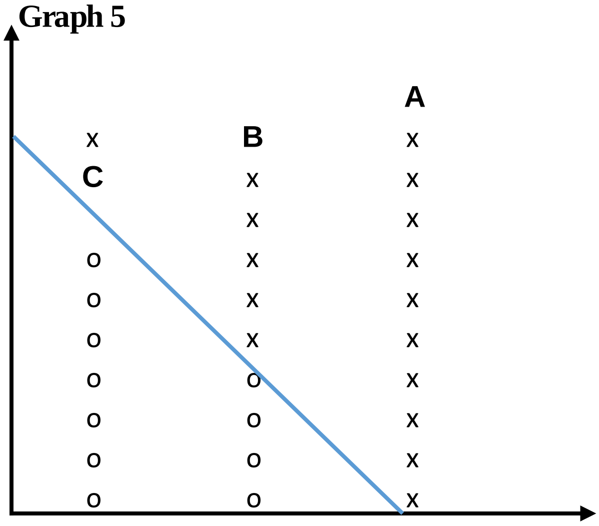
Support Vector Machine

Support vector machine is a method the foundation of which was developed by Cortes and Vapnik (1995) and is gaining popularity due to many attractive features and promising empirical performance. The formulation embodies the Structural Risk Minimisation (SRM) principle, which has been shown to be superior, Gunn et al. (1997), to the traditional Empirical Risk Minimisation (ERM) principle, employed by conventional neural networks. SRM minimizes an upper bound on the expected risk, as opposed to ERM that minimizes the error on the training data. It is this difference that equips SVM with a greater ability to generalize, which is the goal in statistical learning.

SVMs were developed to solve the classification problem, but recently they have been extended to the domain of regression problems, Vapnik et al. (1997).

The intuition behind Support Vector Machine is the following: consider the following graph, in which x's represent individuals who retired early, o's denoting individuals who retired late and the decision boundary which is a line given by the equation $\theta x^T = 0$ (this line is also called the separating hyperplane). Three points have also been labelled A, B and C.

Graph 5



Own elaboration.

SVM technique represents the sample points in space, separating the classes into 2 spaces as wide as possible by a hyper-plane defined as the vector between the 2 points, of the 2 classes, closest to which is called support vector. When the new samples are put in correspondence with said model, depending on the spaces to which they belong, they can be classified in one or other class. For example, point A is very far from the hyper-plane, so it can be classified as early retirement with a high level of confidence. On the other hand, point C is closer to the borderline, being labelled as early retirement but not with the same level of confidence as point A.

Mathematically the problem is the following.

$$min_{\gamma,\omega,b} \frac{1}{2}|\omega^2| \quad (5)$$

$$s.t.\ \ y^i(\omega^T x^i + b) \geq 1, i = 1,2,\dots,m$$

Where γ is the distance to the decision boundary, (ω, b) is the orthogonal vector of the hyperplane and y^i the label of the training data. The objective is to find the optimal hyperplane that allows a correct classification.

5. Results

The data set was split in the training and test set. The performance of the model on the prediction can be measured by the test set and the confusion matrix. This matrix shows the proportion of true positives (proportion of positive cases that were correctly classified TP), false positives (proportion of positive cases that were incorrectly classified FP), true negatives (proportion of negatives cases that were correctly classified TN) and false negatives (proportion of negative cases that were incorrectly classified FN). The three models were run separately selecting randomly the sample of the test set. In table 3, the confusion matrix of the three models is shown.

Table 3. Confusion Matrix

Random Forest

		Predicted	
		Early	Late
Real	Early	173	41
	Late	44	192

Logistic Regression

		Predicted	
		Early	Late
Real	Early	175	33
	Late	47	195

Support Vector Machine

		Predicted	
		Early	Late
Real	Early	179	46
	Late	43	182

Own elaboration.

Note that the sums of the rows of the matrixes are different. This is because the test sets used the models are selected randomly. It can also be observed that models predict very well. Logistic regression predicts over 84% of the cases as early retirement when actually individuals retired early, random forest over 80% and support vector machine over 79%. The three models correctly predict more cases of late retirement than early retirement.

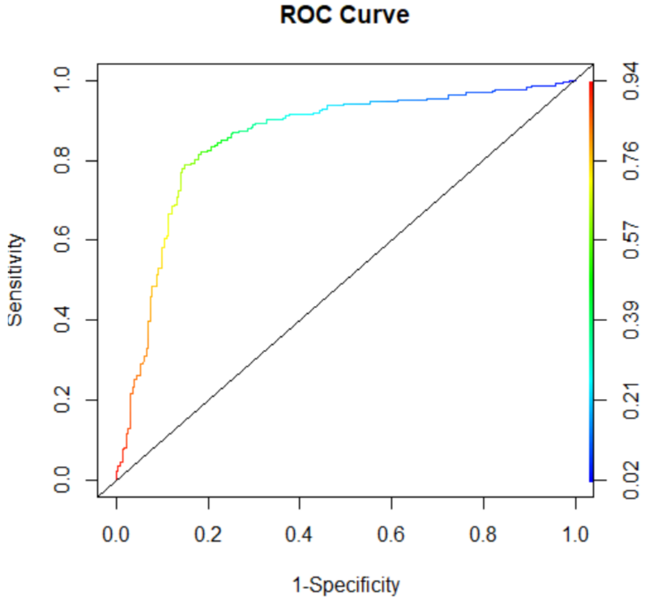
The proportion of true positives is also known as *sensitivity,* which is the probability of predicting an early retirement on the test set when the person actually retired early. The proportion of true negatives is also known as *specificity* which means the probability of predicting a late retirement on the test set when actually the person retired late.

The measures of sensitivity and specificity are used to construct the ROC Curve which is the graph of 1-specificity versus the sensitivity for each possible threshold value or cut-off point on the scale of results of the test under study, Fawcett (2006). This is, $y = f(x)$,

$$ROC(c) = \begin{cases} y = S(c) \\ x = 1 - E(c) \end{cases} \quad (6)$$

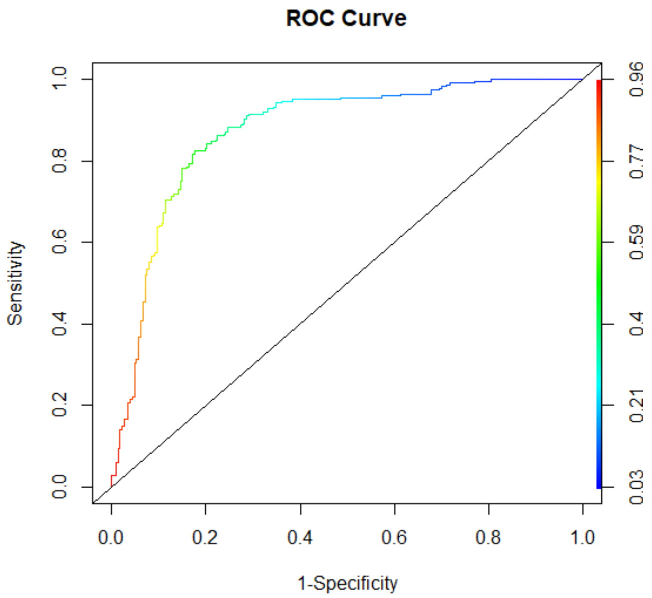
The different thresholds from 0 to 1 are applied to the outputs of the models to predict class membership (confusion matrixes). Since in both axes we have probabilities, the ROC curve will be contained in the square $[0,1]X[0,1]$. In the following graphs, the ROC Curves are shown.

Graph 6. Random Forest



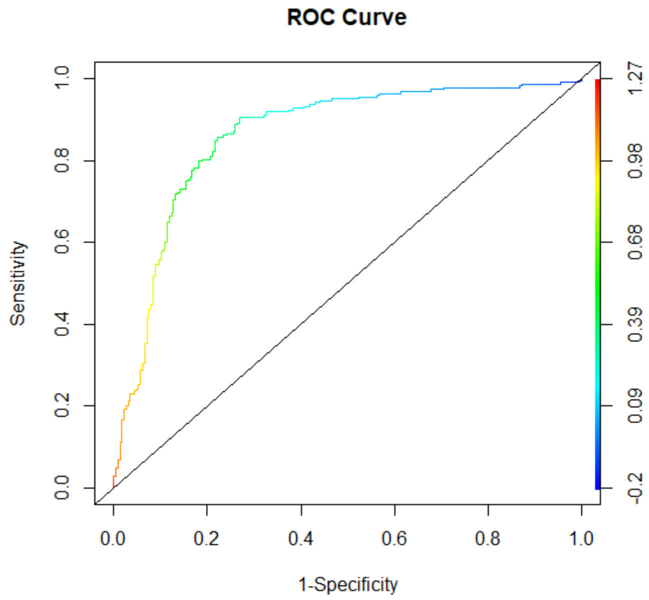
Own elaboration.

Graph 7. Logistic Regression



Own elaboration.

Graph 8. Support Vector Machine



Own elaboration.

The area under the ROC Curve is known as accuracy and represents the probability of discriminating correctly. The accuracies of the models are shown in table 4.

Table 4: Accuracy of the models (AUC)

Random Forest	Logistic Regression	Support Vector Machine
0.8426	0.8880	0.8657

Own elaboration.

It is observed that the models have good accuracy at discriminating early retirement mainly logistic regression.

The three ROC curves have almost the same shape and curvature. This means that any of the three can be used to classify the workers. This can also be seen in the accuracy where it is observed that the three models differ in accuracy only by decimals. For the creation of a score, the most suitable model for its design is the logistic regression. It is possible to create scores with support vector machine and random forest but it requires additional steps in the internal system of companies. The logit model generates the scores directly.

Table 5 shows the score for some workers based on the logit model.

Table 5. Score based on logit model.

Variables	**Individual 1**	**Individual 2**	**Individual 3**	**individual 4**
Gender	Female	Male	Male	Male
Disease	Disease	No Disease	No Disease	Disease
Level of Education	Elementary and Secondary Education	High School	High School	High School
Marital Status	Married	Single	Married	Married
Salary	35 000	24 000	26 000	17 000
Employment Status	Employer	Employee	Employer	Employer
Dependants	4	4	3	3
unemployment rate	3.8436	4.9853	5.0549	4.6000
Credit Score	649	643	605	659
Return of Government Bonds	7.7158	5.3995	6.1988	6.0575
Stock Market	24066.1667	40216.8333	37846.1667	44100.1667
Probability of Early Retirement	**0.0407**	**0.3285**	**0.5194**	**0.9518**

Own elaboration.

If we observe individual 1 and 4 with the lowest and highest probability of early retirement respectively, we can observe certain patterns. Individual 4 has a higher level of education than individual 1, which could indicate that workers with a higher level of education tend to retire early. Individual 1 earns twice as much as individual 4, which means that workers with higher incomes tend to work for longer. Individual 4 is male and 1 female, which implies that men remain in the labour force less time than women. Individual 1 has 4 dependents and individual 4 has 3, which means that a smaller number of dependents implies less responsibility and fewer reasons to work for a longer time. Individual 4 is subject to a higher unemployment rate than individual 1, which confirms the idea that a high unemployment rate encourages early retirement. Individual 4 is subject to a higher stock market corroborating the idea that a high stock market implies more wealth and therefore there is no reason to work for longer. Regarding the returns on bonds, we see that individual 4 who faced a lower return obtained a high probability of retirement before 65 years. There is no significant difference in the credit score of both individuals.

As for health condition it can be observed certain pattern among individuals 2, 3 and 4. The presence of a risky disease is related to a higher probability of early retirement.

6. Conclusions

In this article, it was shown that it is possible to predict early retirement in an insurance company using machine learning algorithms and individual characteristics.

The three algorithms used were: support vector machine, logistic regression and random forest. There was not much difference between the results obtained for the three models in the confusion matrix and precision measures.
The alert system consists of two functions, the classification and creation of a score. For early or late classification, any of the models can be used. For the generation of an early retirement probability score, the most suitable model is a logistic regression.

This alert system can be used in any insurance company which has the necessary data to make predictions. The advantage of machine learning techniques is that new variables can always be incorporated into the model in order to improve the accuracy of the results.

This work also gives an example of how big data and machine learning methodologies might be used, if the availability of data was bigger. For example, in the case of a movement from a public pension system to a private one of individual accounts, the availability of data would be BIG.

The classification and creation of a score in a retirement plan system are innovative and offers insurance and government companies an alternative to track and control early retirement. Predictive data modelling is gaining strength in different fields of the industry. The pension sector must implement artificial intelligence in its processes to improve its results.

Future research aims to analyse how the predictions would change if we apply the models in the public sector or in countries with a different macroeconomic environment and different behaviour patterns.

3

Conclusion

Countries in the world have to make crucial decisions regarding the pension system due to the loss of financial efficiency. This problem is mainly due to the demographic transition that is being experienced in the world: an increase in the number of older people and a decrease in the birth rate. Researchers and policymakers study how other variables besides the demographic structure affect the retirement decisions in order to restore the balance in pension systems. One key variable here is early retirement.

Early retirement clearly benefits and contributes positively to the well-being of individuals. The problem is that current pension systems are not designed to maximize worker welfare. Its structure is very dependent on the population pyramid, mortality rate, life expectancy, contributions to the system and retirement decisions. In this dependency, the objective of maximizing the welfare of workers, private companies and government finances is missed.

This study showed evidence that macroeconomic and individual factors can influence retirement decisions of individuals at country and company level. Social security reforms in various countries have been directed on increasing the labour participation of older individuals, restricting access to early retirement by either increasing the age of retirement or increasing the period of contribution. These measures are usually very general since they do not contemplate the particular case of each individual. Early retirement can be controlled by identifying the specific economic factors and individual characteristics that push individuals to make the decision of retirement. Controlling these variables requires a joint effort of governments, public and private institutions.

Chapter 1 showed that for an oil producing country, downward fluctuations on petroleum prices cause an increase in a proportion of retirement. Exporting countries whose GDP depends in large part on the production of this resource must diversify their economies to minimize the effects of a fall in prices over a long period of time. Increase in the unemployment rate and stock market causes an increase in the proportion of retired workers. In this sense, the countries can boost the generation of jobs through internal and external investment plans to maintain a low unemployment rate. The stock market can be kept stable by ensuring low-interest rates, not overvalue the price of shares and minimizing the pressure of speculative agents.

Information was provided about the effects of oil prices as a new variable that may affect retirement decisions. Measures such as increasing the contribution by workers or increasing the retirement age are usually very unilateral, improving the balance of pension systems but affecting the welfare of workers. The measures in trying to maintain the actuarial balance of the pension systems should go beyond than just modifying inert parameters such as retirement age, contributions, etc., and pay attention to the economic fluctuations in which the pension systems are immersed.

By common sense, it could be inferred that an oil-importing country would behave in the opposite manner to the exporting country analysed in this study, but it would be important to make an analysis of the same type to corroborate our hypotheses.

Chapter 2 designed an alert system to predict early retirement at the company level. This system generates early retirement probabilities based on individual characteristics and macroeconomic variables. Private insurance companies can use the score generated by this system to know at what time individuals with certain characteristics have high probability of retirement. Once this is known, insurance companies can use their insurance executives to negotiate with policy holders and avoid retirement. If the act of convincing does not work, companies will have no choice, but at least they will know when it is highly likely to disburse the financial resource of a pension.

Governments and insurance companies can also use this system to minimize the pension claim rate by identifying which variables are influencing this behaviour. For example, it was observed in this study that people with a deteriorated health status are more likely to retire at an early age, Bloemen et al. (2017). In this sense, the government can provide an efficient health system to minimize the frequency of retirement caused by this risk. Insurance companies by its part can avoid ensuring workers with risky diseases. It is also shown that individuals with higher incomes might work for longer, Quinn (1977), Leonosio (1996) and Milbank (2011). In this case, governments can increase the salary and make a better distribution of wealth in the country. Insurance companies by its part can focus on ensuring workers with high purchasing power.

The model of artificial intelligence presented here allows us to calculate probability of early retirement and automate the tracking of this behaviour. It has the advantage that more variables can be included if they are considered relevant to improve the accuracy of the alert system. This model can be also applied in public institutions since they generally have large amounts of data at the country level.

This study lets us know more variables through which early retirement can be controlled. Variables such as credit score, number of dependents, marital status, worker's gender and oil prices must be taken into account to minimize the frequency of retirement.

References

Alavi A.H., A.H. Gandomi and D.J. Larry. 2016.“The progress of Machine Learning in Geosciences: Preface”. Geoscience Frontiers 7(1): 21-31. https://doi.org/10.1016/j.gsf.2015.10.006.

Akbari S. and F. Mardukhi. 2014. “Classification of Bank Customers Using the Random Forest Algorithm”. *International Journal of Mathematics and Computer Sciences (IJMCS)* 29. http://www.scholarism.net.

Atchley R. C. 1993. “Critical Perspectives on Retirement”. In T. R. Cole, W. A. Achenbaum, P. L. Jakobi, & R. Kastenbaum (Eds.), Voices and visions of ageing: Toward a critical gerontology, 3–19. New York: Springer.

Atchley R. C. 1996. “Retirement”. Encyclopedia of Gerontology, 2:437–449. San Diego, CA: Academic Press.

Bach D.H., S. Sunila and P. Kumar. 2015. “Oil Price and Stock Returns of Consumers and Producers of Crude Oil”. *Journal of International Financial Markets, Institutions and Money* 34: 245-262. doi.org/10.1016/j.intfin.2014.11.010.

Beehr T. A. and M.M. Bennett. 2007. “Examining Retirement From a Multi-level Perspective”. *Applied psychology series*. Ageing and work in the 21st century, 277-302. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.http://psycnet.apa.org/record/2007-15684-013.

Bell R., Y. Koren and C. Volinsky. 2008. The BellKor 2008 Solution to the Netflix Prize.

Blekesaune M. and V. Skirbekk. 2012.“Can Personality Predict Retirement Behaviour? A Longitudinal Analysis Combining Survey and Register Data from Norway”. *European J*ournal of Ageing 9 (3): 199 – 206. https://doi.org/10.1007/s10433-011-0212-3.

Blöndal S. and S. Scarpetta. 1999.The Retirement Decision in OECD Countries. OECD Economics Department Working Papers 202. http://dx.doi.org/10.1787/565174210530.

Blommestein H., P. Janssen, N. Kortleve and J. Yermo. 2009. "Evaluating Risk Sharing in Private Pensions Plans". *OECD Journal: Financial Market Trends* 2009(1), https://doi.org/10.1787/fmt-v2009-art7-en.

Blundell R., C. Meghir and S. Smith. 2002. Pension Incentives and the Pattern of Early Retirement. *The Economic Journal* 112(478): C153-C170. https://doi.org/10.1111/1468-0297.00031.

Bolton R.J. and D.J. Hand. 2001. Unsupervised Profiling Methods for Fraud Detection. Credit Scoring and Credit Control 7: 235–255. http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=EA8DC3FC4AD4C449E74F96A7631DC2B1?doi=10.1.1.24.5743.

Boot J.C.G., W. Feibes and J.H.C. Lisman. 1967. “Further Methods of Derivation of Quarterly Figures from Annual Data”. *Applied Statistics* 16(1): 66-75. doi: 10.2307/2985238.

Bosworth B. P.and G. Burtless. 2010. "Recessions, Wealth Destruction, and the Timing of Retirement”. Working Paper 2010-22, Chestnut Hill, Center for Retirement Research at Boston College.

Breiman L. 2001.“Random Forests”. Machine Learning 45(1): 5
32. https://doi.org/10.1023/A:1010933404324.

Breinegaard N., J.H. Jensen and J.P. Bonde. 2017. Organizational Change, Psychosocial Work Environment, and Non-Disability Early Retirement: A prospective study among senior public employees. *Scandinavian Journal of Work, Environment and Health* 43(3):234-240. doi:10.5271/sjweh.3624.

Burtless G. and J.F. Quinn. 2002. Is Working Longer the Answer for an Aging Workforce?. Issue Brief No. 11. Chestnut Hill, MA: The Center for Retirement Research at Boston College.

Carr E., G.H. Johnson, J. Head, N. Shelton, M. Stafford, S. Stansfeld and P. Zaninotto. 2016. Working Conditions as Predictors of Retirement Intentions and Exit from Paid Employment: A 10-year follow-up of the English Longitudinal Study of Ageing. *European Journal of Ageing* 13:39-48. doi: 10.1007/s10433-015-0357-9.

Cavalcanti T. and J. T. Jalles. 2013. “Macroeconomic Effects of Oil Price Shocks in Brazil and in the United States”. *Applied Energy* 104: 475-486. doi:10.1016/j.apenergy.2012.10.039.

Coile C. and P. B. Levine. 2011. “The Market Crash and Mass Layoffs: How the Current Economic Crisis May Affect Retirement”. *The B.E. Journal of Economic Analysis and Policy* 11(1): 22. doi: 10.2202/1935-1682.2568.

Cortes C. and V. Vapnik. 1995.“Support-Vector Networks”. Machine Learning 20(3): 273-297. https://doi.org/10.1023/A:1022627411411.

Costa D.L. 1998.“The Evolution of Retirement: An American Economic History, 1880-1990”. *The National Bureau of Economic Research* 6-31. http://www.nber.org/chapters/c6108.

Delamaire L., H. Abdou and J. Pointon. 2009.“Credit Card Fraud and Detection Techniques: A Review”. Banks and Bank Systems 4(2): 57–68. http://eprints.hud.ac.uk/id/eprint/19069.

Denton F. T. and B.G. Spencer. 2009. “What is Retirement? A Review and Assessment of Alternative Concepts and Measures”. *Canadian Journal on Aging/La revue canadienne du vieillissement 28*:63–76. https://doi.org/10.1017/S0714980809090047.

Deprez P., P.V. Shevchenko and M.V. Wuthrich. 2017.“Machine Learning Techniques for Mortality Modeling”. *European Actuarial Journal* 7:337–352. https://doi.org/10.1007/s13385-017-0152-4.

Dorn D. and A. Sousa. 2010. “Voluntary’ and ‘Involuntary’ Early Retirement: an International Analysis”. *Applied Economics* 42(4):427-438, doi: 10.1080/00036840701663277.

Du P., A. Samat, B. Waske, S. Liu and Z. Li. 2015. “Random Forest and Rotation Forest for Fully Polarized SAR image Classification Using Polarimetric and Spatial Features”. ISPRS J. Photogramm. Remote Sens. 105, 38–53. https://doi.org/10.1016/j.isprsjprs.2015.03.002.

Elovainio M., K. Van Den Bos, A. Linna, M. Kivimäki, L. Ala-Mursula, J. Pentti and J. Vahtera. 2005. “Combined Effects of Uncertainty and Organizational Justice on Employee Health: Testing the Uncertainty Management Model of Fairness Judgments Among Finnish Public Sector Employees”. Social Science and Medicine 61: 2501-2512.doi: 10.1016/j.socscimed.2005.04.046.

Engle R. F. and C. W. J. Granger. 1987. “Co-Integration and Error Correction: Representation, Estimation and Testing”. *Econometrica* 55(2): 251-276. doi: 0012-9682(198703)55:2<251:CAECRE>2.0.CO;2-T.

Fallick B. 1996.“A Review of the Recent Empirical Literature on Displaced Workers”. ILR Review 50(1): 5-16. https://doi.org/10.1177/001979399605000101.

Fawcett T. 2006.“An introduction to ROC Analysis”. *Pattern Recognition Letters* 27(8): 861-874. https://doi.org/10.1016/j.patrec.2005.10.010.

Feldman D. C. 1994. “The Decision to Retire Early: A review and conceptualization”. *The Academy of Management Review* 19(2):285–311. doi: 10.2307/258706.

Feldman D.C. and T.A. Beehr. 2011. “A Three-Phase Model of Retirement Decision Making”. *American Psychologist* 66(3):193-203. doi: 10.1037/a0022153.

Figueira D.A.M., M.C.L. Haddad, R. Gvozd and P.S.C. Pissinati (2017). Retirement Decision-making Influenced by Family and Work Relationships. *Revista Brasileira de Geriatria e Gerontologia* 20(2): 206-2013. http://dx.doi.org/10.1590/1981-22562017020.160127.

Florêncio P. 2016. “The Brazilian 2010 Oil Regulatory Framework and its Crowding-Out Investment Effects”. *Energy Policy* 98: 378-389.doi:10.1016/j.enpol.2016.08.038.

Gentleman J. 1984. “Mexican Oil and Dependent Development”. New York: P. Lang. ISBN:978-0-8204-0063-1.

Gunn S.R., M. Brown, K.M. Bossley. 1997.“Network Performance Assessment for Neurofuzzy Data Modelling”. Lecture Notes in Computer Science, Springer 1280: 313–323.https://doi.org/10.1007/BFb0052850.

Gustman A.L. and T.L. Steinmeier. 1986. “A Structural Retirement Model”. *Econometrica* 54(3): 555-584. https://www.jstor.org/stable/1911308.

Hakola T. and R. Uusitalo. 2004. “Not so Voluntary Retirement Decisions? Evidence from a Pension Reform”. *Journal of Public Economics* 89(11-12):2121-2136. https://doi.org/10.1016/j.jpubeco.2004.12.001.

Herbertsson T.T.2001. “Shrinking Labour Forces and Early Retirement”. Ifo Institute for Economic Research 2(4): 31-37. www.ifo.de/w/BJAoJWyj.

Hernández R., C. Fernández and P. Baptista. 2006. Metodología de la investigación. (4 ª Ed.). México: McGraw Hill.La cultura del ahorro y la previsión en el México actual. (2010, 5 de octubre).

Heyman A. 2004.“A Structural Dynamic Analysis of Retirement Behaviour in the Netherlands”. *Journal of Applied Econometrics* 19(6): 739-759. https://doi.org/10.1002/jae.781.

Hinton G.E., S. Osindero and Y.W. Teh. 2006. A Fast Learning Algorithm for Deep Belief Nets. Neural Computation, 18:1527–1554.

Hurd M.D. and S.Rohwedder. 2010.“Effects of the Financial Crisis and Great Recession on American Households”. *The National Bureau of Economic Research*. Working Paper 16407. doi: 10.3386/w16407.

Johansen S. 1991. “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models”. *Econometrica* 59(6): 1551-1580. doi: 0012-9682(199111)59:6<1551:EAHTOC>2.0.CO;2-P.

Joshi R., R. Gupte and P. Saravanan. 2018. “A Random Forest Approach for Predicting Online Buying Behavior of Indian Customers”. Theoretical Economics Letter 8: 448-475. http://www.scirp.org/journal/tel.

Juszczak P., N.M. Adams, D.J. Hand, C. Whitrow and D.J. Weston. 2008.“The Peg and Bespoke Classifiers for Fraud Detection”. Computational Statistics and Data Analysis 52(9):4521–4532. https://doi.org/10.1016/j.csda.2008.03.014.

Lebedev A.V., E. Westman, G.J.P. Van Westen, M.G. Kramberger, A. Lundervold, D. Aarsland, H. Soininen, I. Kłoszewska, P. Mecocci, M. Tsolaki, B. Vellas, S. Lovestone and A. Simmons. 2014. “Random Forest Ensembles for Detection and Prediction of Alzheimer's Disease with a Good Between-cohort Robustness”. NeuroImage Clinic 28(6):115-125. https://doi.org/10.1016/j.nicl.2014.08.023.

Leinonen T., M. Laaksonen, T. Chandola and P. Martikainen. 2016. “Health as a Predictor of Early Retirement before and after Introduction of a Flexible Statutory Pension Age in Finland”. Social Science and Medicine 158: 149-157. https://doi.org/10.1016/j.socscimed.2016.04.029.

Lin T.C. 2001. “Letter to the Editor: Impact of Job Stress on Early Retirement Intention”. International Journal of Stress Management 8(3): 243-247. https://doi.org/10.1023/A:1011395227349.

Lin Y.J., C.S. Huang, C.C.C Lin. 2008. Determination of Insurance Policy Using Neural Networks and Simplified Models with Factor Analysis Technique.

Liu Y. and T. Xie. 2018. “Machine Learning Versus Econometrics: prediction of box office”. Applied Economics Letters. https://doi.org/10.1080/13504851.2018.1441499.

Madeira S.C., A.L. Oliveira, and C.S. Conceição. 2003. A Data Mining Approach to Credit Risk Evaluation and Behaviour Scoring. In: Pires F.M., Abreu S. (eds) Progress in Artificial Intelligence. EPIA 2003. Lecture Notes in Computer Science, vol 2902. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24580-3_25.

McAfee A. and E. Brynjolfsson. 2012. Big Data: The Management Revolution. Harvard Business Review 90(10):60-6, 68, 128.

McCullagh P. and J.A. Nelder. 1989. Generalized Linear Models 37. London Chapman and Hall.

Ministry of Finance and Public Credit (SHCP). 2017. Opportunistic Public Finance Statistics. www.shcp.gob.mx.

Mu J., L. Xu, X. Duan and H. Pu. 2013. “Study on Customer Loyalty Prediction Based on RF Algorithm”. Journal of Computers 8(8):2134-2138. doi:10.4304/jcp.8.8.2134-2138.

Muldoon D. and R.W. Kopcke. 2008. “Are People Claiming Social Security Benefits Later?”. CRR Issue Brief Number 8-7. Chestnut Hill, Center for Retirement Research at Boston College.

Munnell A.H., S.A. Sass, M. Soto and N. Zhivan. 2006. “Has the Displacement of Older Workers Increased?”. Working Paper 2006-17, Center for Retirement Research at Boston College.doi:10.2139/ssrn.1304736.

Nivedha R. and N. Sairam. 2015. “A Machine Learning Based Classification for Social Media Messages”. *Indian Journal of Science and Technology* 8(16). doi**:** 10.17485/ijst/2015/v8i16/63640.

OECD Pensions Outlook 2016, OECD Publishing, Paris, https://doi.org/10.1787/pens_outlook-2016-en.

OECD Reviews of Pension Systems: Mexico, 2016.

Ohlson J.A. 1980. “Financial Ratios and the Probabilistic Prediction of Bankruptcy”. *Journal of Accounting Research* 18(1):109-131. https://www.jstor.org/stable/2490395.

Organization of the Petroleum Exporting Countries. 2015. Yearly Basket Prices. Retrieved from http://www.opec.org/opec_web/en/data_graphs/40.htm (October 19).

Qianqian Z. 2011. “The Impact of International Oil Price Fluctuation on China's Economy”. *EnergyProcedia* 5: 1360-1364. doi:10.1016/j.egypro.2011.03.235.

Olaniyi O., E. Ajibola, A. Ibiwoye and A.B. Sogunro. 2012.“Artificial Neural Network Model for Predicting Insolvency in Insurance Industry”. *International Journal* of *Management and Business Research* 2(1): 59-68. http://ijmbr.srbiau.ac.ir/article_63.html.

Peng C.Y., K.L. Lee and G.M. Ingersoll. 2002. “An Introduction to Logistic Regression Analysis and Reporting”. *The Journal of Educational Research* 96: 3-14. http://dx.doi.org/10.1080/00220670209598786

Perols J. 2011. Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms. *AUDITING: A Journal of Practice and Theory* 30(2):19-50. https://doi.org/10.2308/ajpt-50009.

Pozzolo A.D., O. Caelen, Y.A. Le Borgne and S. Waterschoot. 2014. Learned Lessons in Credit Card Fraud Detection from a Practitioner Perspective. Expert Systems with Applications 41(10): 4915-4928. https://doi.org/10.1016/j.eswa.2014.02.026.

Queiroz B.L. and L.R. Souza. 2017. “Retirement Incentives and Couple’s Retirement Decisions in Brazil”. *The Journal of the Economics of Ageing* 9(C): 1-13. http://www.sciencedirect.com/science/article/pii/S2212828X15300062.

Reday-Mulvey G. 2005. “Working Beyond 60 – Introduction: Key Policy Issues and Recommendations”. The Geneva Papers on Risk and Insurance - Issues and Practice 30(4): 604-619. https://doi.org/10.1057/palgrave.gpp.2510053.

Rivera F.J. and A.A. Nava. 2012. CIENCIA-UANL, ISSN 1405-9177, ISSN-e 2007-1175 15(57): 33-38. https://dialnet.unirioja.es/servlet/articulo?codigo=3834983.

Shultz K.S., K.R. Morton and J.R. Weckerle. 1998. The Influence of Push and Pull Factors on Voluntary and Involuntary EearlyRetirees´Retirement Decision and Adjustment. Journal of Vocational Behaviour 53(1): 45-57.

Smith S. 2006. “The Retirement Consumption Puzzle and Involuntary Early Retirement: Evidence from the British Household Panel Survey”. *The Economic Journal* 116(510): C130-C148. https://doi.org/10.1111/j.1468-0297.2006.01080.x.

Szinovacz M. 2003. Contexts and Pathways: retirement as an institution, process and experience. En G. Adams y T. Beehr (Eds.), Retirement: reasons, processes and results, 6-52. New York: Springer.

Taigman Y., M. Yang, M.A. Ranzato and L. Wolf. 2014. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. doi:10.1109/CVPR.2014.220.

Talwar A. and Y. Kumar. 2013. "Machine Learning: An Artificial Intelligence Methodology". *International Journal of Engineering and Computer Science* 2(12): 3400-3404. http://ijecs.in/index.php/ijecs/article/view/2261.

Tian H. 2016. "The Role of China's Real Economic Activity in Oil Price Fluctuations". *Procedia-Social and Behavioural Science* 220: 522-530. doi:10.1016/j.sbspro.2016.05.528.

Ulloa B. 2006. "Evaluación de la transición del sistema provisional chileno: comparaciones y simulaciones". Repositorio académico de la Universidad de Chile. http://www.repositorio.uchile.cl/handle/2250/108392.

Vapnik V., S. Golowich, and A. Smola. 1997. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. Advances in Neural Information Processing Systems 9: 281– 287. https://papers.nips.cc/paper/1187-support-vector-method-for-function-approximation-regression-estimation-and-signal-processing.

Viaene S., M. Ayuso, M. Guillen, D. Van Gheel and G. Dedene. 2007. "Strategies for Detecting Fraudulent Claims in The Automobile Insurance Industry". *European Journal of Operational Research* 176(1): 565–583. https://doi.org/10.1016/j.ejor.2005.08.005.

Walker A. 1982."The Social Consequences of Early Retirement". *The Political Quarterly* 53(1):61 – 72. doi.org/10.1111/j.1467-923X.1982.tb01198.x.

Wang M. and J. Shi. 2014. "Psychological Research on Retirement". *Annual Review of Psychology* 65: 209–233.https://doi.org/10.1146/annurev-psych-010213-115131.

Wang M. and V. Alterman. 2017. "Retirement". Oxford Research Encyclopedia of Psychology. doi: 10.1093/acrefore/9780190236557.013.11.

Wang M., Y. Zhan, S. Liu and K.S. Shultz. 2008. "Antecedents of Bridge Employment: a Longitudinal Investigation". *Journal of Applied Psychology* 93(4):818–830. 10.1037/0021-9010.93.4.818.

Zheng E., Y. Tan, P. Goes, R. Chellappa, D.J. Wu, M. Shaw, O. Sheng and A. Gupta. 2017. "When Econometrics Meets Machine Learning". Data and Information Management 1(2):75-83. https://doi.org/10.1515/dim-2017-0012.

Zickar M.J. 2012. "The Evolving History of Retirement Within the United States". In M. Wang (Ed.), *The Oxford handbook of retirement* 10–21. New York: Oxford University Press.

A

Pension Schemes in Mexico by Law

Mexico has two pension schemes. They are commonly known as the scheme of 73 and the scheme of 97. If a worker started working and contributing before July 1, 1997, he has the right to choose between the two schemes. When an individual started working and contributing after July 1, 1997, he will only be entitled to the scheme of 97. The scheme of 73 is financed by the federal government and the scheme of 97 is financed by the contributions of workers saved in AFORES.

The main characteristics of the two schemes are shown in the following table.

	Scheme of 1973	**Scheme of 1997**
Weeks of contribution	500	1250
Calculation of the pension	Average salary last 5 years	Amount saved in the individual account (Afore) plus the yields obtained
Family allowance	NO	YES
Additional bonus payment equivalent to one month of pension	YES	NO
Duration of the pension	Life	Depending on the initial calculation of the life expectancy of the insured
Widowhood pension	YES	NO
pension update	Annual according to inflation	It does not actualize