Forecasting Unemployment in the Euro-Area with Machine Learning

Article *in* Journal of Forecasting · September 2021

DOI: 10.1002/for.2824

Periklis Gogas
Democritus University of Thrace

Theophilos Papadimitriou
Democritus University of Thrace

Emmanouil Sofianos
Democritus University of Thrace

Forecasting Unemployment in the Euro-Area with Machine Learning

Journal:	*Journal of Forecasting*
Manuscript ID	FOR-21-0042
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	27-Jan-2021

Complete List of Authors:	Sofianos, Emmanouil; Demokriteio Panepistemio Thrakes Gogas, Periklis; Demokriteio Panepistemio Thrakes Papadimitriou, Theophilos; Demokriteio Panepistemio Thrakes

Keywords:	
Keywords:	Unemployment Rate, Euro Area, Machine Learning, Forecasting

Abstract:	Unemployment has a direct impact on public finances and yields serious sociopolitical implications. This study aims to directionally forecast the euro-area unemployment rate. To the best of our knowledge, no other studies forecast the euro-area unemployment rate as a whole. The data set includes the unemployment rate and 36 explanatory variables, as suggested by theory and the relevant literature, spanning the period from 1998:4 to 2019:9 in monthly frequency. These variables are fed to three machine learning methodologies: Decision Trees (DT), Random Forests (RF), and Support Vector Machines (SVM), while an Elastic-Net Logistic Regression (Logit) model is used from the area of Econometrics. The results show that the optimal random forest model outperforms the other models by reaching a full-dataset forecasting accuracy of 88.5% and 85.4% on the out-of-sample.

Forecasting Unemployment in the Euro-Area with Machine Learning

Periklis Gogas, Theophilos Papadimitriou, Emmanouil Sofianos*

Department of Economics, Democritus University of Thrace, 69100 Komotini, Greece;

pgkogkas@econ.duth.gr (P.G.); papadimi@econ.duth.gr (T.P.)

* Correspondence: esofiano@econ.duth.gr

Abstract

Unemployment has a direct impact on public finances and yields serious sociopolitical implications. This study aims to directionally forecast the euro-area unemployment rate. To the best of our knowledge, no other studies forecast the euro-area unemployment rate as a whole. The data set includes the unemployment rate and 36 explanatory variables, as suggested by theory and the relevant literature, spanning the period from 1998:4 to 2019:9 in monthly frequency. These variables are fed to three machine learning methodologies: Decision Trees (DT), Random Forests (RF), and Support Vector Machines (SVM), while an Elastic-Net Logistic Regression (Logit) model is used from the area of Econometrics. The results show that the optimal random forest model outperforms the other models by reaching a full-dataset forecasting accuracy of 88.5% and 85.4% on the out-of-sample.

JEL classification: C53; E24; E27

Keywords: Unemployment Rate; Euro Area; Decision Trees; Random Forest; SVM; Machine Learning; Forecasting;

1. Introduction

The unemployment rate is an important macroeconomic variable. It significantly impacts the individuals that remain out of the work force both financially and psychologically as much as the households that they support. At the macroeconomic level, a high unemployment rate has detrimental effects both on the society and the national economy. An increasing unemployment rate reduces tax revenue (due to the loss of income by the unemployed) and increases government spending in terms of unemployment benefits and the associated re-training and placement programs. In countries where the central bank's mandate includes not only the preservation of monetary stability but also the goal of full employment, an increased unemployment rate may increase inflationary pressure. Therefore, being able to timely forecast changes in the unemployment rate, is important for fiscal and monetary policy makers in order to implement relevant policies.

The significance of unemployment rate forecasting was first acknowledged during the 1970s, when for the first time the problem of stagflation (i.e. the coexistence of high inflation with high unemployment and slow economic growth) emerged (Brunner et al., 1980, Grubb et al., 1982). Before the 1973 stagflation, economists relied in the traditional Phillips Curve that assumed an inverse relationship between the unemployment and the inflation rate. This relation, eventually, proved to be a mere statistical stylized fact, valid only in the short-run when monetary policy is unexpected and/or not credible (Muth, 1961).

Many statistical and econometric techniques have been employed in the voluminous and rich literature of unemployment forecasting, most of them burdened with shortcomings, ranging from high sensitivity to model specification, to extreme requirements with respect to the data (Cook and Smalter Hall, 2017). Recently, economists turned their attention to models free from these shortcomings from the area of Machine Learning.

In the last 20 years, the ARIMA and GARCH models were used extensively to forecast the unemployment rate. Dobre and Alexandru (2008) used ARIMA in the case of Romania and Floros (2005) used GARCH in the case of the UK. Mladenovic et al. (2017)

used ARIMA models to forecast unemployment in the EU28. Kurita (2010) used ARFIMA models to forecast the unemployment of Japan.

Moreover, models using Google Index (job-search index) and data mining methods are applied to forecast the unemployment rate for the US (D'Amuri and Marcucci 2010, Xu, et al. 2013).

With respect to Machine Learning models, Sermpinis et al. (2014), use a hybrid genetic algorithm-support vector regression (GA-SVR) to forecast the US inflation and the unemployment. The forecasting performance of the GA-SVR model is tested against a random walk model, an autoregressive moving average model, a moving average convergence/divergence model, a multi-layer perceptron, a recurrent neural network, and a genetic programming algorithm. The GA-SVR outperforms the competition. Cook and Smalter Hall, (2017) created four different neural network architectures to forecast the US unemployment rate. Their results outperformed the SPF (Survey of Professional Forecasters), used as a benchmark. Similarly, Kreiner and Duca (2019) use Artificial Neural Networks (ANN) to forecast the US unemployment outperforming the SPF benchmark results. Stasinakis et al. (2014) forecast the US unemployment rate using radial basis function neural networks (RBFNN), Kalman filters and support vector regression. The models are tested against an ARMA, a STAR and three different ANN. Their results show that the RBFNN statistically outperforms all other models.

Despite the rich literature on forecasting the US unemployment rate, there is a limited number of papers forecasting unemployment in the eurozone. Most of the relevant studies focus their attention in the national level and to the case of specific countries. Dumisic et al. (2015) use double exponential smoothing and the Holt-Winters' methods to forecast the unemployment rate in Greece, Spain, Croatia, Italy and Portugal. The authors conclude that in all the analyzed countries, the double exponential smoothing method, is the most accurate with the exception of Italy where the Holt-Winters' method dominated. Barot (2004), assesses the accuracy of the Swedish domestic forecasters for GDP growth, CPI and unemployment. The best directional forecasting model for unemployment reached an accuracy of 75%.

Claveria (2019), investigates the unemployment rate for eight European countries with ARIMA models including as predictors both an indicator of unemployment, based on the degree of agreement in consumer unemployment expectations, and a measure of disagreement based on the dispersion of expectations. The forecasting models are tested on out-of-sample data. According to the tests the degree of agreement in consumers' expectations contains useful information in predicting unemployment rates, especially for the detection of turning points. Katris (2019), forecasts the unemployment rate in 22 European countries using the FARIMA model with GARCH errors. The author considers the non-linearity of the data and uses artificial neural networks, support vector regression and multivariate adaptive regression splines on his tests. FARIMA models are the optimal choice for the 1-step ahead forecasts, while for the longer forecasting period (h=12) the neural network approaches achieves comparable results with the FARIMA-based models.

In this paper we employ several Machine Learning models to directionally forecast the euro-area unemployment rate as a whole. The models used are the Support Vector Machines (SVM), the Decision Trees (DT), the Random Forests (RF) and the Elastic-Net Logistic Regression (Logit) model. The innovations of our approach are: a) we focus our study in the supranational level of the euro area as a whole and not to specific countries within the E.U. and b) we directionally forecast the unemployment rate using several classification models from the area of Machine Learning. To the best of our knowledge, no other study attempts to forecast the direction of euro-area unemployment rate. Moreover, we use the permutation Variable Importance Measure (VIM) in the RF methodology to identify the significance of the predictors used in the model.

The paper is organized as follows: in Section 2 we will briefly discuss the methodologies and the data while in Section 3 we describe our empirical results. Finally, Section 4 concludes the paper.

2. Methodology and dataset

2.1. SVM

Support vector machines (SVM) is a set of methods for data classification and regression based on the maximization of the interclass distance: the basic concept of the SVM is to define the optimal[1] linear separator that separates the data points into two classes.

Let's consider a dataset of N m-sized vectors $\mathbf{x}_i \in \mathrm{R}^m$ (i=1, 2…, N). These vectors are labelled according to the class they belong $y_i \in \{-1, +1\}$.

If the two classes are linearly separable, we define a boundary as:

$$f(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i - b = 0. \tag{1}$$

Subject to:

$$\mathbf{w}^T\mathbf{x}_i - b > 0 \; \forall i: y_i = +1,$$
$$\mathbf{w}^T\mathbf{x}_i - b < 0 \; \forall i: y_i = -1, \tag{2}$$

$\mathbf{w}$ is the parameter vector, and b is the bias (Figure 1). So $y_i f(\mathbf{x}_i) > 0, \forall i$.

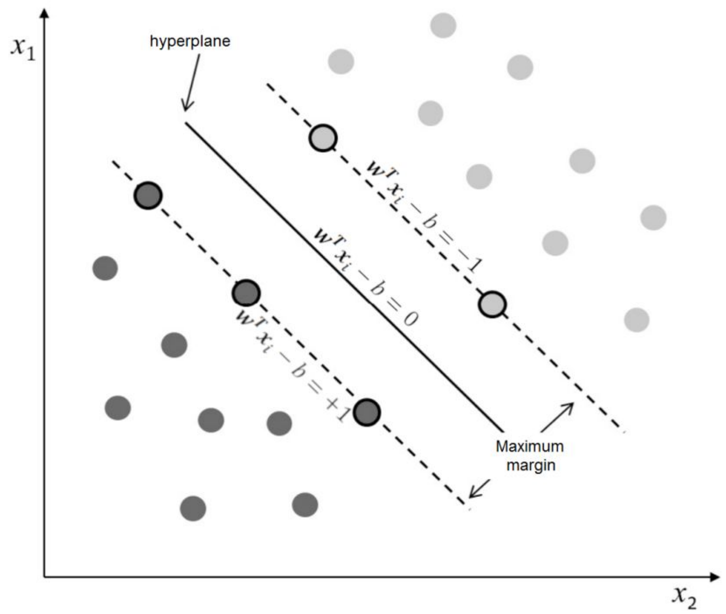


Figure 1. Hyperplane selection and support vectors. The SVs (represented with the pronounced black contour) define the margins that are represented with the dashed lines and outline the separating hyperplane represented by the continuous line.

[1] optimal in the sense of the model generalization to unknown data

In the linearly separable case, the separator (line in two dimensions, plane in three dimensions, and hyperplane for higher dimensions spaces) is defined as the decision boundary that classifies each data point to the correct class. A small subset of datapoints called support vectors (SV) define the position of the separator. In Figure 1, the SV are represented with the pronounced black contour, the margin lines (parallel lines to the optimal separator passing from the SV) are represented by dashed lines, and the separator is represented by a continuous line. The distance between the two margin lines is the distance between the two classes. The goal of SVM is to identify the linear separator that maximizes the distance between the two classes.

In real-life phenomena, datasets are often contaminated with noise and may contain outliers. These cases cannot be approached using the presented methodology. Cortes and Vapnik (1995), introduced non-negative slack variables and a parameter C, describing the desired tolerance to classification errors to treat these cases.

When the dataset is not linearly separable, the SVM model is paired with kernels: The initial data space is projected through a kernel function into a space of higher dimensionality (called feature space) where the dataset may be linearly separable. In Figure 2, we show a dataset that is not linearly separable in the two-dimensional data space (left graph), denoted by the red and blue circles. By projecting the dataset in a three-dimensional feature space (right graph) using a kernel function, the linear separation is feasible.

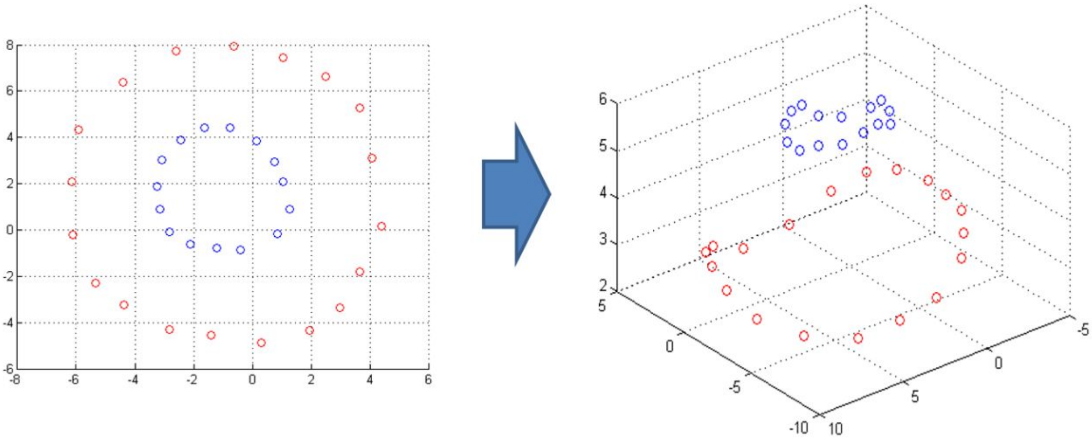


Figure 2. The non-separable two-class scenario in the data space (left) and the separable case in the feature space after the projection (right).

In this paper, we examine three kernels, the linear, the radial basis function (RBF) and the polynomial kernel. The linear kernel defines the separating hyperplane in the original dimensions of the data space, while the RBF and the polynomial project the initial dataset to a higher dimensional space. You may find a detailed analysis of the SVM methodology in Gogas et al. (2019). (Our implementation of SVM models is based on LIBSVM[2], Chang and Lin., 2011).

2.2. Decision Trees

The decision trees are supervised machine learning methodologies for classification and regression. They are flowchart-like top-down structures of nodes and branches. Each node represents a splitting criterion (calculated with an impurity measure) and each branch the corresponding outcome. The top node is the root node representing the complete dataset; the other ones are called decision nodes. The nodes that don't split any further are called leaves (or terminal nodes) and depict the final outcomes of the decision-making process (Figure 3).

---

[2] The software is available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/

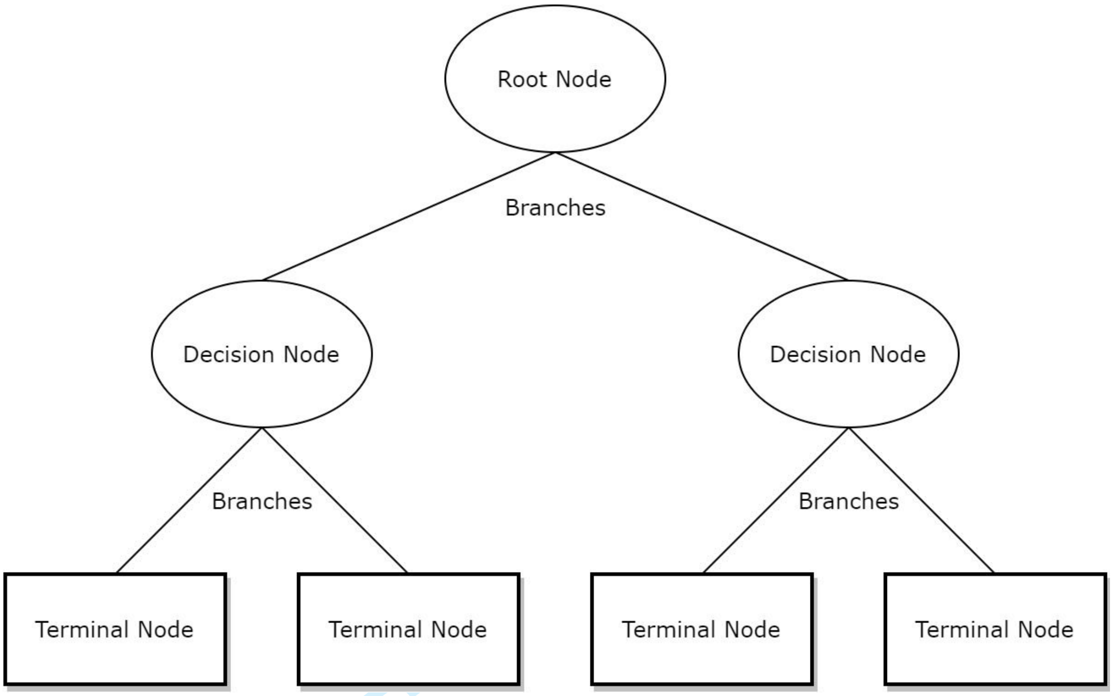


Figure 3. Example of a decision tree.

In this paper we use the CART (Classification And Regression Trees) algorithm. The CART algorithm creates binary decision trees using the Gini impurity split criterion (Breiman et al. 1984). Given the true labels distribution, Gini impurity measures the frequency that a randomly selected element in a set, may be randomly labeled incorrectly. Trivially, in the case that every leaf contains elements of one class, the Gini impurity is zero.

In a set D of m-sized vectors distributed in L classes, $i \in \{1,\ldots,L\}$, the subset of D with the elements that belong to the i-th class is denoted $\mathrm{D}^{(i)}$. The Gini impurity of D is calculated as

$$Gini(\mathrm{D}) = 1 - \sum_{i=1}^{L} p_i^2, \tag{3}$$

where p_i is the relative frequency of appearance of class i in D calculated as $\frac{|\mathrm{D}^{(i)}|}{|\mathrm{D}|}$, (where |D| denotes the cardinality of the set D).

The CART growing procedure stops when one of the two standard stopping criteria is met: a. Every leaf has elements of only one class or b. Some user specified limit on either the maximum number of splits or the maximum number of leaf nodes is reached.

2.3. Random Forests

Decision trees are easy to interpret but their main drawback is that their generalization ability is low and they perform poorly in out-of-sample data. In other words, they tend to have a low bias but a high variance. Breiman (2001a), introduced the random forests algorithm. It is a learning method that combines the concept of decision trees with the bootstrap aggregating algorithm that is usually called bagging[3] (Breiman, 1996). The method avoids the issue of overfitting that may occur in decision trees by combining the result of many decision trees.

When training random forest models, we create many alternative classification trees. For each individual tree, we draw from the initial data set of size m, a randomly selected with replacement (bootstrapped) sample of the same size. From this procedure, each of the bootstrapped samples contains approximately 2/3 of the initial dataset observations (some of them multiple times) while the rest that is not selected, called the out-of-bag (OOB), is used to test the generalization ability of the trained model. From the initial set of features of cardinality n, a randomly selected subset of $\sqrt{n}$ features is used in every tree and at every split. This is done to reduce the dependence of the model to the training set, i.e. to reduce the bias of the model. No observation is a member of every OOB's in a forest. The response of the RF for any observation of the dataset, is the dominant outcome from all the cases (trees) in which the observation is member of the OOB. The accuracy of the RF model is calculated as the rate of the correctly classified observations.

An important aspect of the random forests methodology is the ability to calculate Variable Importance Measures (VIMs) which can be used to rank the predictors according to

[3] from Bootstrapping AGGregatING.

their relevant importance in forecasting the response variable, the target. There are two main strains of VIMs: a) the mean decrease in impurity (Gini) VIM and b) the permutation VIM. The former is mainly used for continuous variables (Breiman et al. 1984) and thus in our empirical work we use the latter. The permutation VIM is defined to be the decrease in a model's score (OOB score in our case) when a single feature is randomly shuffled (Breiman, 2001b). This process breaks the connection between the feature and the target variable; hence the decrease in the OOB score is indicative of how much the model relies upon the feature. This helps to shed some light in the workings of the so-called black-box of the methodology. The ranking of the features is of special interest to policy makers as they can indirectly infer and assess the relative importance of the independent variables.

2.4. Elastic-Net Logistic Regression (Logit)

The logit model is a well-established econometric methodology that is used widely in the empirical forecasting literature for binary classification. It estimates the probability:

$$\mathrm{P_r}\,(y = 1) = \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}}} \tag{4}$$

Where $\mathrm{P_r}(y = 1)$ is the probability that an event y will occur and is assumed to be determined by a vector of independent variables x. In this setting, y, the dependent binary variable (dummy variable), takes the value of 1 or 0. In our case, a value of 0 indicates a decrease in the unemployment rate and a value of 1 indicates an increase while β is the vector of the estimated coefficients.

The Ridge regression and LASSO (Least Absolute Shrinkage and Selection Operator) methodologies are simple techniques that reduce the complexity and possible multicollinearity in models with many explanatory variables. Moreover, they prevent overfitting that may result from simple linear models, such as the logit, introducing a tolerable amount of bias in the model.

Lasso adds a penalty for non-zero coefficients, but unlike ridge regression which penalizes the sum of squared coefficients (L_2 penalty), lasso penalizes the sum of their absolute values (L_1 penalty). As a result, employing lasso, many coefficients are exactly zeroed. This is never the case in ridge regression. Thus, lasso can be also used as a feature selection method leading to parsimonious models. The penalty terms are controlled by the parameter lambda (λ).

Elastic net, proposed by Zou and Hastie (2005), is a penalized linear regression model that includes both the lasso (L_1) and ridge (L_2) penalties during training. A parameter $\alpha \in (0, 1]$, is provided to determine the weight imposed to each of the penalties. As α (alpha) shrinks toward 0, elastic net approaches to ridge regression and when $\alpha=1$ elastic reduces to lasso. In our case, the parameters α and λ are optimized within a 5-fold cross validation scheme.

2.5. The Dataset

The dataset includes the euro-area monthly unemployment rate and 36 additional explanatory variables. The frequency is monthly, and the data span the period from 1998:4 to 2019:9 for a total of 243 observations for each variable. The 36 explanatory variables were selected based on economic theory and previous relevant empirical studies (Chen and Ranciere, 2019, Groen and Kapetanios, 2016, Kim and Swanson, 2014, Shen, 1996). The data for the unemployment rate were obtained from the ECB Statistical Data Warehouse and the explanatory variables from the ECB Statistical Data Warehouse, the Federal Reserve Bank of St. Louis, the Center for Economic Policy Research (CEPR) and Yahoo Finance. Table 1 summarizes these variables.

Table 1. List of explanatory variables used with sources and type of variable.

Variable	Source	Type
Unemployment rate female	ECB Statistical Data Warehouse	percent
Unemployment rate male	ECB Statistical Data Warehouse	percent
EuroCoin Index	Center for Economic Policy Research	index
WTI Oil Prices (US dollars)	Federal Reserve Bank of St. Louis	average
WTI Oil Prices (US dollars)	Federal Reserve	end of period

	Bank of St. Louis	
10 Year Euro Bond Rate	Federal Reserve Bank of St. Louis	percent
Henry Hub Natural Gas Spot Price	Federal Reserve Bank of St. Louis	Dollars per million Btu, average
Henry Hub Natural Gas Spot Price	Federal Reserve Bank of St. Louis	Dollars per million Btu, end of period
USD/EUR	ECB Statistical Data Warehouse	end of period
AUD/EUR	ECB Statistical Data Warehouse	end of period
CAD/EUR	ECB Statistical Data Warehouse	end of period
JPY/EUR	ECB Statistical Data Warehouse	end of period
GBP/EUR	ECB Statistical Data Warehouse	end of period
Trade Weighted U.S. Dollar Index: Broad, Goods	Federal Reserve Bank of St. Louis	average, Index Jan 1997=100
Trade Weighted U.S. Dollar Index: Broad, Goods	Federal Reserve Bank of St. Louis	end of period, Index Jan 1997=100
Trade Weighted U.S. Dollar Index: Major Currencies, Goods	Federal Reserve Bank of St. Louis	average, Index Mar 1973=100
Trade Weighted U.S. Dollar Index: Major Currencies, Goods	Federal Reserve Bank of St. Louis	end of period, Index Mar 1973=100
M1 euro area	Federal Reserve Bank of St. Louis	Billions of dollars
M3 euro area	Federal Reserve Bank of St. Louis	Billions of dollars
M1 US	Federal Reserve Bank of St. Louis	Billions of dollars
M2 US	Federal Reserve Bank of St. Louis	Billions of dollars
M3 US	Federal Reserve Bank of St. Louis	Billions of dollars
MZM US	Federal Reserve Bank of St. Louis	Billions of dollars
Dow Jones	Yahoo Finance	adjusted close, end of period
NASDAQ	Yahoo Finance	adjusted close, end of period
S&P 500	Yahoo Finance	adjusted close, end of period
DAX (German Stock Exchange)	Yahoo Finance	adjusted close, end of period
CAC 40 (French Stock Exchange)	Yahoo Finance	adjusted close, end of period
International Trade: Exports: Value (goods): Total for the Euro Area	Federal Reserve Bank of St. Louis	levels
International Trade: Imports: Value (goods): Total for the Euro Area	Federal Reserve Bank of St. Louis	levels
Consumer Price Index: Total All Items for the	Federal Reserve	Growth rate

United States	Bank of St. Louis	
Consumer Price Index: Total All Items for the United States	Federal Reserve Bank of St. Louis	index
Consumer Price Index: All items: Total: Total for the Euro Area	Federal Reserve Bank of St. Louis	Growth rate
Consumer Price Index: All items: Total: Total for the Euro Area	Federal Reserve Bank of St. Louis	index
Gold Fixing Price 3:00 P.M. (London time) in London Bullion Market, based in U.S. Dollars	Federal Reserve Bank of St. Louis	average
Gold Fixing Price 3:00 P.M. (London time) in London Bullion Market, based in U.S. Dollars	Federal Reserve Bank of St. Louis	end of period

The target variable (label) is defined by the first difference of the unemployment rate: 0 when the difference is negative implying that the unemployment rate decreases and 1 when the difference is positive and the unemployment rate rises. In the cases where the unemployment rate does not change in two subsequent months and the difference is 0, we assigned the label of the previous difference. All variables except for the ones expressing percentages are transformed into natural logarithms and the unemployment rate is transformed in first differences. The dataset was randomly permutated to establish that the training set and the out-of-sample set have a comparable ratio of both classes. The permutation procedure does not use future information as the whole rows are shuffled.

We tested the data for stationarity using the ADF test. All data are stationary due to the permutation. For the SVM, the data were normalized to the $[-1, 1]$ range.

We created various data subsets for training and evaluating our models (Table 2):

- **Training Set (TS)**: The optimal parameter values for all models are identified in the training step. For the SVM (in all three kernels) and the DT models we employ a 5-fold cross-validation procedure. The optimal parameters are the ones that provide the maximum average accuracy in all five folds. For all models the training set included the same 80% of the full dataset, i.e. 195 observations.
- **Out-Of-Bag (OOB) set**: In the case of the RF, there is no need for cross-validation to avoid over-fitting, as the OOB observations are selected randomly for each tree in each forest. Out-Of-Bag is composed by the observations that are not used when creating the bootstrapped sets for training the trees. For the RF models, the

optimal parameter values are the ones that maximize the accuracy in the OOB set. For every tree, the OOB set is not used anywhere in the training step.

- **Out-Of-Sample (OOS) set**: This is 20% of the full dataset, i.e. 48 observations that were left aside from the training or the selection of the best model from each methodology. The OOS set is used to gauge the performance of the optimal trained models from each methodology to new data that were never seen by the models and eventually detect overfitting.

- **Full Dataset (FD):** The overall optimal model was identified by feeding the top model from each methodology with the initial full dataset (FD=TS∪OOS).

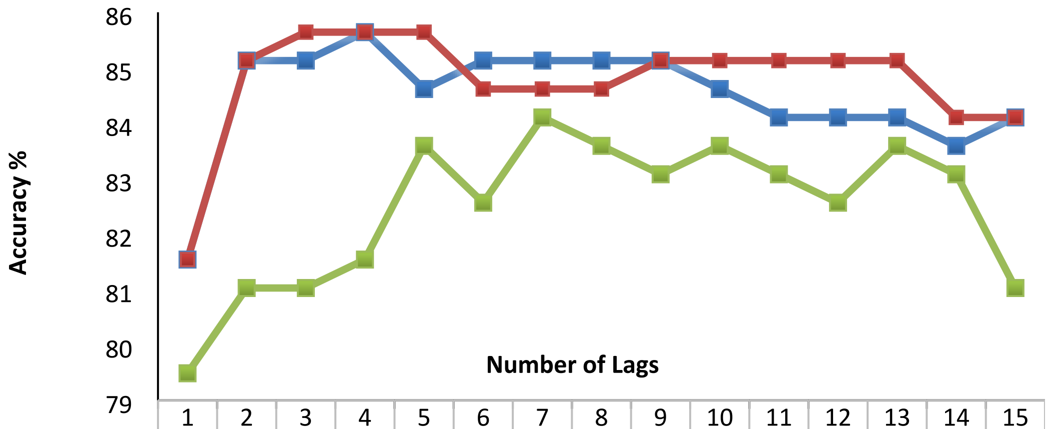
Table 2. Model selection procedure

Methodology	Model selection via	Optimal parameters	Test overfitting	Optimal overall model
SVM-Linear	Training	Max average CV-test	OOS	Full Dataset
SVM-RBF	Training	Max average CV-test	OOS	
SVM-Polynomial	Training	Max average CV-test	OOS	
Decision Trees	Training	Max average CV-test	OOS	
Random Forests	Training	Max average OOB	OOS	
Elastic-Net Logit	Estimation	Min deviance CV-test	OOS	

3. Empirical results

3.1. SVM models

With the SVM models we proceeded in two steps: first we produced the best autoregressive model AR(q) for each kernel (linear, RBF and polynomial) and then we augmented it to a structural model by adding the explanatory variables one by one. The autoregressive models were created using lagged values of the unemployment rate first differences. For the optimization of the hyper-parameters we used a cross-validation procedure with a coarse-to-fine grid search scheme. We considered a maximum of 15 lags. The model with the highest training set accuracy is chosen as the best AR(q) model for each kernel (Figure 4). According to this, the optimal AR(q) models are an AR(4) for the linear kernel, an AR(3) for the RBF kernel and an AR(7) for the polynomial kernel. Both the linear and the RBF kernel based AR models achieved a training set accuracy of 85.64% and the polynomial kernel based models a training set accuracy of 84.1%.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
linear	81.54	85.13	85.13	85.64	84.62	85.13	85.13	85.13	85.13	84.62	84.10	84.10	84.10	83.59	84.10
rbf	81.54	85.13	85.64	85.64	85.64	84.62	84.62	84.62	85.13	85.13	85.13	85.13	85.13	84.10	84.10
polynomial	79.49	81.03	81.03	81.54	83.59	82.56	84.10	83.59	83.08	83.59	83.08	82.56	83.59	83.08	81.03

Figure 4. Training set accuracy of SVM-AR models for the three kernels.

In the second step, the explanatory variables were sequentially added, one by one, in the best autoregressive models identified in the previous step for each kernel. The training set accuracy of these augmented models is reported in Table 3. The best models, for all three kernels, are the ones that are augmented with the DAX. The linear and the RBF[4] kernel models reached an 86.67% accuracy and the polynomial 85.13%. We proceeded by adding sequentially a second explanatory variable, but this did not increase the model accuracy in any case.

[4] For the RBF kernel the augmented models with WTI prices (end of period) and M3 for eurozone also achieve an accuracy of 86.67% but the train accuracy (average train cross-validation) is higher for the model with the DAX variable and is selected as the best augmented RBF model.

Table 3. Training set directional forecasting accuracy of the explanatory variables: linear, RBF and polynomial kernels.

Variable	Linear	RBF	Polynomial
Best AR	85,64%	85,64%	84,10%
Unemployment rate eurozone female	84,62%	85,13%	83,08%
Unemployment rate eurozone male	84,10%	85,13%	84,10%
10y Euro Bond rate	85,64%	85,13%	83,59%
CPI EU growth rate	85,64%	86,15%	82,56%
CPI EU index	85,13%	85,13%	84,10%
CPI US growth rate	85,13%	84,62%	83,08%
CPI US index	85,13%	85,13%	83,08%
Trade Weighted US goods, average	85,64%	85,64%	83,59%
Trade Weighted US goods, end of period	85,64%	85,13%	83,08%
Trade Weighted US currencies, average	84,62%	84,10%	83,59%
Trade Weighted US currencies, end of period	85,13%	84,10%	82,05%
WTI average LN	85,64%	86,15%	83,59%
WTI end of period LN	85,64%	86,67% *	84,10%
Henry Hub natural gas spot prices, average LN	85,64%	83,08%	83,08%
Henry Hub natural gas spot prices, end of period LN	86,15%	83,08%	82,56%
M1 EU LN	85,13%	85,64%	83,08%
M3 EU LN	85,13%	86,67% *	84,62%
M1 US LN	85,64%	85,13%	84,10%
M2 US LN	85,13%	85,64%	83,59%
M3 US LN	85,13%	85,64%	83,59%
MZM US LN	85,13%	85,64%	84,10%
Exports EU levels LN	85,13%	84,62%	83,59%
Imports EU levels LN	85,13%	85,64%	84,10%
Gold prices, average LN	85,13%	86,15%	83,59%
Gold prices, end of period LN	85,64%	86,15%	83,59%
Eurocoin index	84,62%	83,59%	84,62%
USD/EUR	85,64%	84,62%	81,54%
AUD/EUR	84,10%	86,15%	84,10%
CAD/EUR	83,08%	85,13%	83,59%
JPY/EUR	84,62%	82,05%	81,54%
GBP/EUR	85,13%	85,13%	83,59%
Dow Jones LN	86,15%	85,64%	84,10%
Nasdaq LN	85,13%	85,64%	83,59%
S&P 500 LN	85,13%	86,15%	83,59%
DAX LN	86,67% *	86,67% *	85,13% *
CAC40 LN	86,15%	86,15%	83,59%

An asterisk (*) denotes the best augmented model for each kernel.

3.2. Decision Trees models

For the decision trees we trained the model using all the explanatory variables. We trained various models with the number of decision nodes ranging from 1 to 156 and the leaf size ranging from 1 to 50 using cross validation and grid search for the hyper-parameter optimization. The optimal forecasting model has 2 decision nodes and 3 leaf nodes. The first node (parent node) is defined by the criterion "the first difference of the unemployment rate at lag 2 is less than 0.015". If the criterion is not satisfied, the model forecasts that the unemployment rate will increase (label=1). Otherwise, the tree continues in the second node, where the criterion is "the Nasdaq is less than 7.64". If this criterion is satisfied the model forecasts that the unemployment rate will rise (label=1); in the opposite case the model forecasts that the unemployment rate will decrease (label=0) (Figure 5). The training set forecasting accuracy of the best decision tree model is 83.08%.

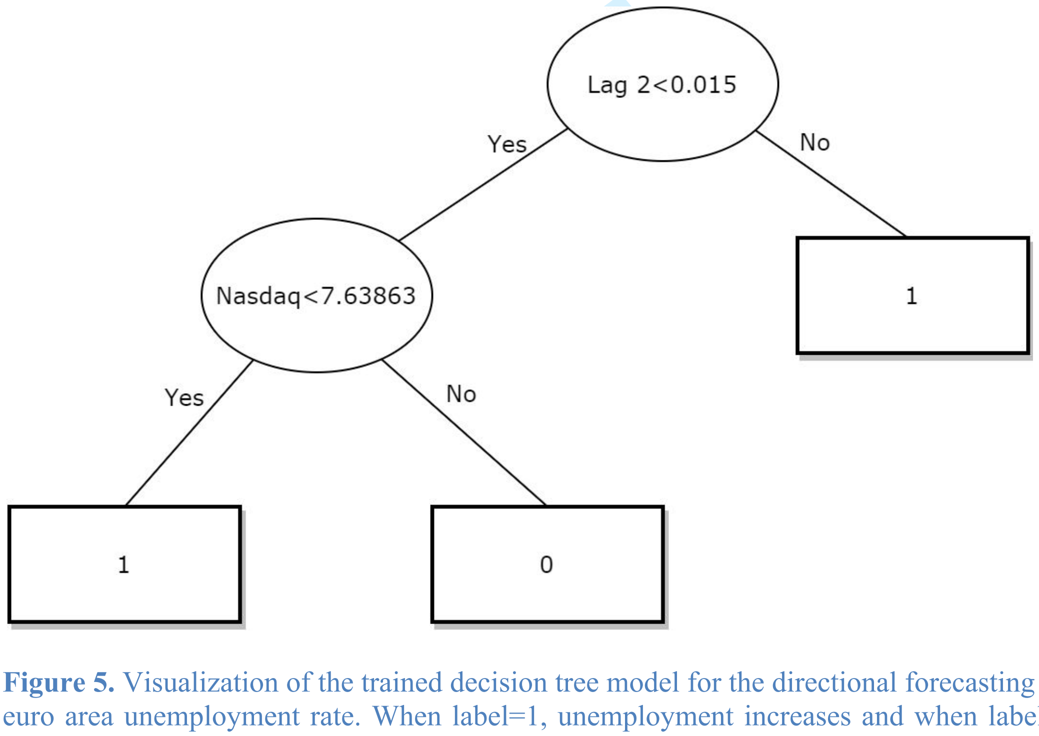


Figure 5. Visualization of the trained decision tree model for the directional forecasting of the euro area unemployment rate. When label=1, unemployment increases and when label=0, it decreases.

3.3. Random Forest models

As we mention earlier, our training sample includes 195 observations. With these we created 950 alternative forests that include trees ranging from 50 to 1000. Each tree in these models uses a bootstrapped sample (sampling with replacement) of the initial size of 195 observations. At every node a different, randomly selected, subset of explanatory variables is used. The cardinality of this subset is $\sqrt{n}$ as proposed by Geurts et al. (2006), where n is the number of explanatory variables (features); in our sample, N=51 (36 explanatory variables and 15 lags of the unemployment rate first differences). This is done to reduce the dependence of the model to the training data, which is the basic cause of overfitting (low bias and high variance).

We trained our models and used grid search for the hyper-parameter optimization. Namely, we trained models with maximum number of splits ranging from 1 to 5 and number of trees ranging from 50 to 1000. The OOB accuracy was used to select the best model. The optimal forest includes 88 trees and the maximum number of splits is equal to 5. The corresponding training set accuracy (OOB accuracy) is 85.64%.

3.4. Elastic-Net Logistic Regression (Logit) model

We optimized the parameters α and λ in a 5-fold cross-validation scheme. The model with minimum binomial deviance is selected as optimal (Zou and Hastie, 2005, Hastie et. al, 2009). Deviance is defined as the difference of likelihoods between the fitted model and the saturated model (perfect model). The optimal model is an elastic net regularization logit model with α=0.7. The selected variables (non-zero fitted coefficients) are shown in Table 4. The training set accuracy of this model is 86.67%.

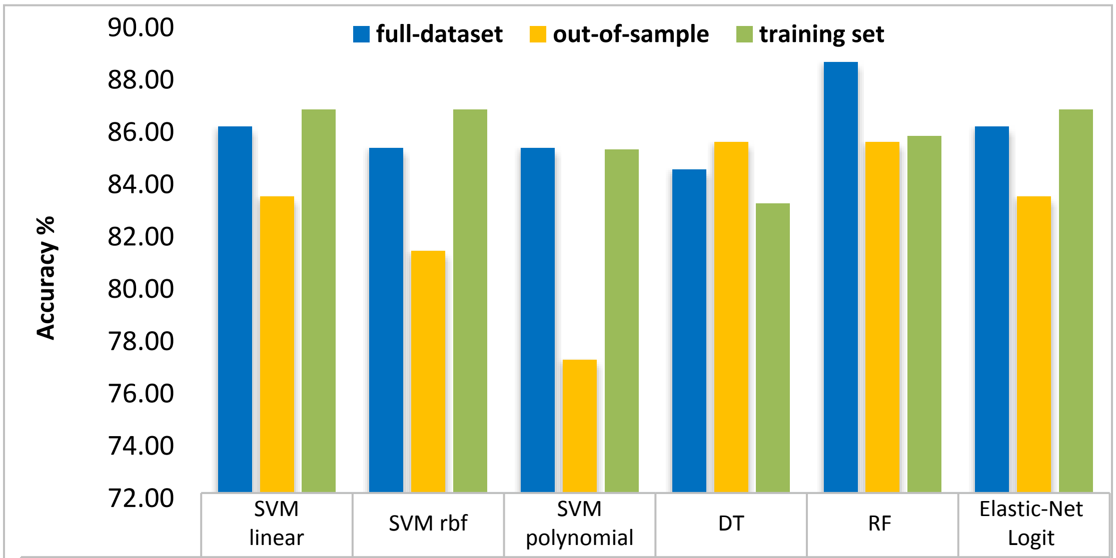
Table 4. Selected variables with non-zero fitted coefficients

1	Lag 1 of unemployment rate (first differences)
2	Lag 2 of unemployment rate (first differences)
3	Lag 6 of unemployment rate (first differences)
4	Unemployment rate female

5	10 year Euro Bond rate
6	Eurocoin index
7	CAD/EUR
8	Nasdaq
9	CAC40

3.5. Overall Best Model

Thus far, we have identified the best model from each methodology (SVM, DT, RF, Elastic-net logit) using the same training set. In this section, we feed these models with the full dataset and we select the best overall model. This is the model that achieves the highest accuracy in directionally forecasting the monthly unemployment rate in euro-area. We use the full-dataset accuracy since we used different training schemes (OOB for the random forest, cross-validation for the SVM, the DT and the Elastic-net logit), and it would be wrong to compare the performance of the created models using the training set accuracy. We also check the performance of the models in the out-of-sample part only to evaluate the forecasting performance on data that was not used during the training process i.e. the possibility of having low bias (high accuracy in training set) and high variance (low generalization ability). The results are summarized in Figure 6.



	SVM linear	SVM rbf	SVM polynomial	DT	RF	Elastic-Net Logit
full-dataset	86.01	85.19	85.19	84.36	88.48	86.01
out-of-sample	83.33	81.25	77.08	85.42	85.42	83.33
training set	86.67	86.67	85.13	83.08	85.64	86.67

Figure 6. Full-dataset, out-of-sample and training set directional forecasting accuracy of euro area unemployment rate: decision trees, random forest SVM (linear, RBF and polynomial) and Elastic-net logit models.

According to the results, the SVM coupled with the linear kernel reached a full-dataset accuracy of 86.01% and an out-of-sample accuracy of 83.33%. The SVM coupled with the RBF kernel reached a full-dataset accuracy of 85.18% and an out-of-sample accuracy of 81.25%. The SVM model with the polynomial kernel reached a full-dataset accuracy of 85.19% and an out-of-sample accuracy of 77.08%. The best SVM model is the one coupled with the linear kernel. The decision tree model reached a full-dataset accuracy of 84.36% and an out of sample accuracy of 85.42%, while the random forest model has a full-dataset accuracy of 88.48% and out-of-sample 85.42%. Finally, the elastic-net logit model achieved a full-dataset accuracy of 86.01% and an out-of-sample accuracy of 83.33%. According to these results, the best overall directional forecasting model for the euro area unemployment rate is the random forest model.

3.6. Variable Importance Measure results

As mentioned in the empirical part, we used the permutation VIM to identify and interpret the forecasting ability of the explanatory variables, for the random forest model. The permutation VIM calculates the loss in the OOB score after the permutation of each variable, showing the dependence of the model on each variable. In Table 5 we present the ranking of the variables according to permutation VIM. According to this ranking, the 10 most important variables are: the first 4 lags of the euro area unemployment rate, the Eurocoin index, 4 financial indices (S&P500, Nasdaq, Dow Jones and CAC40) and the WTI prices.

Table 5. Permutation feature importance with random forest ensemble method.

Variable	Permutation VIM
lag1	0,5310
lag2	0,4889
S&P 500	0,4322
Eurocoin index	0,3601
lag3	0,3352
Nasdaq	0,3347
Dow Jones	0,3324
lag4	0,3162
WTI average	0,3129
CAC40	0,2846
Trade Weighted US currencies average	0,2758
Trade Weighted US goods end of period	0,2648
lag5	0,2626
DAX	0,2612
USD/EUR	0,2486
lag6	0,2299
CAD/EUR	0,2211
AUD/EUR	0,2200
M3 US	0,2137
Gold prices average	0,2122
10y Euro Bond rate	0,2080
lag7	0,2004
CPI EU growth rate	0,1991
MZM US	0,1957
Trade Weighted US currencies end of period	0,1851
lag8	0,1796
CPI US index	0,1714
M1 EU	0,1703
Trade Weighted US goods average	0,1638

lag9	0,1578
CPI EU index	0,1438
M1 US	0,1434
WTI end of period	0,1397
Gold prices end of period	0,1379
Unemployment rate eurozone female	0,1349
M2 US	0,1060
lag12	0,1060
Imports EU levels	0,1037
CPI US growth rate	0,1019
GBP/EUR	0,0912
lag15	0,0676
lag14	0,0582
lag11	0,0534
Unemployment rate eurozone male	0,0255
JPY/EUR	0,0171
Exports EU levels	0,0000
M3 EU	0,0000
lag10	-0,0038
Henry Hub natural gas spot prices average	-0,0455
Henry Hub natural gas spot prices end of period	-0,1494
lag13	-0,1716

Lags seem to play an important role in forecasting the unemployment rate. It seems that unemployment follows short-run trends. In the decision tree model the lag2 is selected in one out of the 2 decision nodes. In the case of the random forests, that as we have seen achieves the best accuracy over all the tested methodologies, the permutation VIM assigns the highest scores to lag1 and lag2 of the unemployment rate.

Stock market indices seem to play an important role in euro-area unemployment rate forecasting. Stock prices reflect the present value of the market's expectations of future corporate profits. Thus, it is expected that either directly or indirectly there may be a positive link between stock indices, output and employment. Economic theory suggests that the expectation of an increased demand in the foreseeable future will be followed by a rise in supply. This increase in supply can be achieved through new investment and/or increased usage of the existing production capacity. In both cases, it is expected to observe an increase in the demand for the two factors of production: capital and labor. As a result, the unemployment rate in the near future is expected to fall.

In the SVM models, the DAX provides the highest forecasting accuracy for all SVM kernels and the Nasdaq is selected in one of the two decision nodes in the decision trees model. The permutation VIM for the random forests reports 4 financial indices among the 10 most important variables: the S&P500, the Nasdaq, the Dow Jones (related to the U.S. stock market), and the CAC40 (related to the French stock market). Similar findings are reported in several papers in the literature. Farmer in 2010 and 2015 shows that the stock market contains significant information about future unemployment. Pan (2018) found a particularly strong and one-way causal direction from stock prices to the unemployment rate in G7 countries and a strong bilateral causal relationship between stock prices and unemployment for other advanced countries. Sibande et al., 2019 use DCC-MGARCH tests, to analyze time-varying causality between stock market returns and unemployment in the UK.

4. Conclusion

The unemployment, obviously, has many serious social, political and personal negative implications as the right to work is one of the basic rights in a modern democracy. Additionally, from the economics standpoint, unemployment creates significant inefficiencies. The GDP is sub-optimal, as part of the human capital is not used in production. Moreover, unemployment imposes significant costs to governments in terms of the associated benefits, re-training programs and the mechanisms that are put in place for tracking, monitoring and minimizing the people that are out of jobs.

As a result, accurately measuring, reducing and controlling the unemployment rate is one of the main goals of modern governments. Policy makers from both the government and the central bank, usually take into account as one the variables of main concern the unemployment rate when they design and implement fiscal and monetary policy.

In this study, we attempted to directionally forecast, the euro-area unemployment rate. To the best of our knowledge, this is the first time that the EU-area over-all unemployment rate is forecasted. In the empirical section, we employed decision trees, random forests and SVM models with the linear, the RBF and the polynomial kernels. These methodologies were further compared to a logit forecasting model under an elastic-net framework from the area of econometrics. The direction of the euro-area unemployment rate is the dependent variable and 36 explanatory variables spanning the period from 1998:4 to 2019:9 comprised our dataset. We train all models using 80% of the dataset and the remaining 20% was used as the out-of-sample data to detect possible overfitting and evaluate the generalization ability of our models to new and unknown data.

Overall, the best accuracy was achieved by the random forest model: 88.48% on the full-dataset and 85.42% on the out-of-sample. According to the permutation VIM the most important of the explanatory variables are: the first 4 lags of the unemployment rate, 4 out of 5 tested financial indices (S&P500, Nasdaq, Dow Jones and the CAC40), the Eurocoin index and the WTI prices. Stock markets seem to play an important role in forecasting unemployment in all employed models. In all SVM kernels, the use of the DAX provides a higher forecasting accuracy, the Nasdaq is selected in the decision trees models and 4 financial indices are included in the best random forest model. Thus, we find strong evidence that the use of financial indices seems to play an important role in the euro-area unemployment forecasting.

Acknowledgements: «This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project “Strengthening Human Resources Research Potential via Doctorate Research” (MIS-5000432), implemented by the State Scholarships Foundation (IKY)»

References

- Barot, B. (2004), How accurate are the Swedish forecasters on GDB-Growth, CPI-inflation and unemployment? (1993 - 2001). *Brussels Economic Review.* 47. 249-278.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), pp.123-140.
- Breiman, L. (2001)a. Random forests. *Machine Learning*, 45 pp.5-32
- Breiman, L. (2001)b. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3), pp.199-231.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth. *Inc. Monterey, California, USA*.
- Brunner, K., Cukierman, A. and Meltzer, A. (1980). Stagflation, persistent unemployment and the permanence of economic shocks. *Journal of Monetary Economics*, 6(4), pp.467-492.
- Chang, C. and Lin, C. (2011). LIBSVM. *ACM Transactions on Intelligent Systems and Technology*, 2(3), pp.1-27.
- Chen, S. and Ranciere, R. (2019). Financial information and macroeconomic forecasts. *International Journal of Forecasting*, 35(3), pp.1160-1174.
- Claveria, O. (2019). Forecasting the unemployment rate using the degree of agreement in consumer unemployment expectations. *Journal for Labour Market Research*, 53(1).
- Cook, T. and Smalter Hall, A. (2017). Macroeconomic Indicator Forecasting with Deep Neural Networks. *The Federal Reserve Bank of Kansas City Research Working Papers*,.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), pp.273-297.
- D'Amuri, F. and Marcucci, J. (2010). 'Google It!' Forecasting the US Unemployment Rate with A Google Job Search Index. *SSRN Electronic Journal*,.

- Dobre, I. and Alexandru, A. A. (2008). Modelling unemployment rate using Box–Jenkins procedure. *Journal of Applied Quantitative Methods, 3*(2), 156–166.
- Dumičić, K., Časni, A., Žmuk, B. (2015). Forecasting Unemployment Rate in Selected European Countries Using Smoothing Methods. *World Academy of Science, Engineering and Technology: International Journal of Social Education, Economics and Management Engineering*, 9(4), 867 - 872.
- Farmer, R. (2010). How to reduce unemployment: A new policy proposal. *Journal of Monetary Economics*, 57(5), pp.557-572.
- Farmer, R. (2015). The Stock Market Crash Really Did Cause the Great Recession. *Oxford Bulletin of Economics and Statistics*, 77(5), pp.617-633.
- Floros, C. (2005). Forecasting the UK unemployment rate: Model comparisons. *International Journal of Applied Econometrics and Quantitative Studies, 2*(4), 57–72.
- Geurts, P., Ernst, D. and Wehenkel, L. (2006). Extremely randomized trees. *Mach Learn,* 63, pp. 3–42
- Gogas, P., Papadimitriou, T. and Sofianos, E. (2019). Money Neutrality, Monetary Aggregates and Machine Learning. *Algorithms*, 12(7), p.137.
- Groen, J. and Kapetanios, G. (2016). Revisiting useful approaches to data-rich macroeconomic forecasting. *Computational Statistics & Data Analysis*, 100, pp.221-239.
- Grubb, D., Jackman, R. and Layard, R. (1982). Causes of the Current Stagflation. *The Review of Economic Studies*, 49(5), p.707.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. *Springer Science & Business Media*.
- Katris, C. (2019). Prediction of Unemployment Rates with Time Series and Machine Learning Techniques. *Computational Economics*, 55(2), pp.673-706.

- Kim, H. and Swanson, N. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, 178, pp.352-367.
- Kreiner, A. and Duca, J. (2019). Can machine learning on economic data better forecast the unemployment rate? *Applied Economics Letters*, pp.1-4.
- Kurita T. (2010). A Forecasting Model for Japan's Unemployment Rate. *Eurasian Journal of Business and Economics*, 3 (5), 127-134
- Mladenovic, J., Ilic, I., & Kostic, Z. (2017). Modeling the unemployment rate at the eu level by using Box–Jenkins methodology. *KnE Social Sciences, 1*(2), 1–13.
- Muth, J. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica*, 29(3), p.315.
- Pan, W. (2018). Does the stock market really cause unemployment? A cross-country analysis. *The North American Journal of Economics and Finance*, 44, pp.34-43.
- Sermpinis, G., Stasinakis, C., Theofilatos, K. and Karathanasopoulos, A. (2014). Inflation and Unemployment Forecasting with Genetic Support Vector Regression. *Journal of Forecasting*, 33(6), pp.471-487.
- Shen, C. (1996). Forecasting macroeconomic variables using data of different periodicities. *International Journal of Forecasting*, 12(2), pp.269-282.
- Sibande, X., Gupta, R. and Wohar, M. (2019). Time-varying causal relationship between stock market and unemployment in the United Kingdom: Historical evidence from 1855 to 2017. *Journal of Multinational Financial Management*, 49, pp.81-88.
- Stasinakis, C., Sermpinis, G., Theofilatos, K. and Karathanasopoulos, A. (2014). Forecasting US Unemployment with Radial Basis Neural Networks, Kalman Filters and Support Vector Regressions. *Computational Economics*, 47(4), pp.569-587.
- Xu, W., Li, Z., Cheng, C. and Zheng, T. (2013). Data mining for unemployment rate prediction using search engine query data. *Service Oriented Computing and Applications*, 7(1), pp.33-42.

- Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), pp.301-320.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.