**'Unemployment in EU': Unlocking forecasting techniques Using advanced machine learning and statistical models.**

**Sba22250**

**A Partial Research Thesis**

**of the requirement of**

**Degree of**

**Master of Science in Data Analytics**

A picture containing icon

Description automatically generated

**May 2023**

**1. Research Title..........................................................................................................................3**

**2. Abstract...................................................................................................................................3**

**3. Introduction……………………………………………………………………………………………………………………………3**

**4. Research Objective ..................................................................................................................3**

**5. Validity …………………………………………………………………………………………………………………….……………5**

**6. Relevance………………………………………………………………………………………………………………………………5**

**7. Literature Review.....................................................................................................................6**

**7.1 Machine Learning Models…………………………………………………………………………………………….........8**

**7.2 Autoregressive Integrated Moving Average (ARIMA) Model...................................................8**

**7.3 Vector Autoregression (VAR) Model………………………………………………………………………………….…9**

**7.4 Artificial Neural Network (ANN) Model……………………………………………………………………………..…9**

**7.5 Comparative Analysis…………………………………………………………………………………………………………11**

**8. Sampling………………………………………………………………………………………………………………………………11**

**9. Primary Research Methodology……………………………………………………………………………………………12**

**10. Ethics………………………………………………………………………………………………………………………………….14**

**11.Conclusions...........................................................................................................................15**

**12. References...........................................................................................................................17**

1. **Research Title(**Unlocking - be careful here as it's not clear as to where you wish to progress to. The topic content has a social science element to it - be careful to keep it focussed on the DA domain area. 5)

'Unemployment in EU': Unlocking forecasting techniques Using advanced machine learning and statistical models.

1. **Abstract**

This study looks into the problem of unemployment in the European Union (EU) and makes predictions about how it will affect employment in the future using cutting-edge machine learning algorithms, forecasting techniques, and statistical models. The official statistics agency of the European Commission, Eurostat, confirmed the validity and accuracy of the data, which was the analysis. We are certain that this study will fulfill the necessity for a careful comprehension of business improvements in the EU and convey significant data to partners and policymakers. We work to produce accurate projections and forecasts by utilizing cutting-edge analytical techniques, allowing pre-emptive actions to be made to successfully solve unemployment concerns.

The projected study results will provide insight into the dynamics of employment in the EU going forward, aiding in the development of evidence-based policies and initiatives to encourage inclusive and sustainable growth.

1. **Introduction**

Unemployment is an important economic and social problem impacting individuals, communities, and entire countries. Long-term work possibilities and combating unemployment have been priorities for the European Union (EU). It is necessary for policymakers, economists, and other stakeholders to comprehensively understand the EU's unemployment rate to make accurate predictions about the future. This review examines possible effects on work in the EU utilizing state-of-the-art AI calculations, factual models, and anticipating procedures. We try to give top-to-bottom evaluations of the elements of joblessness in the EU and produce exact projections for informed navigation by using an exhaustive dataset on joblessness from Eurostat, the authority measurements office of the European Commission.

1. **Research Objectives(**These are still general - more focus needed. Some over emphasis on TO content, again be careful to ensure ROs are relevant and focussed. 8)

Calculating and evaluating how future unemployment will impact employment in the European Union (EU) is the goal of the data analysis study. The project's objective is to thoroughly examine the available data, mostly from Eurostat, to currently available, mostly from Eurostat, which is to comprehend trends in EU unemployment and produce projections. “These goals can be met by carefully examining a range of time series models, including machine learning methods such as ARIMA, VAR, SVM, Decision Tree, Random Forest, and exponential smoothing. The study compares how well different models perform when it comes to correctly predicting unemployment rates in the EU while considering seasonality, trends, and other economic indicators. The purpose of the study is to shed light on the usefulness and applicability of various time series models for forecasting unemployment rates.”

**1. Analysing Historical Unemployment Data**: This study's primary objective is to analyse the historical unemployment data that it got from Eurostat. We must examine historical trends, patterns, and fluctuations in unemployment rates to fully understand the dynamics and characteristics of unemployment in the EU. Future trends in unemployment will be predicted using the analysis as a basis.

2. **Forecasting Future Patterns in Unemployment:** The goal of the project is to create forecasting models that can foresee future patterns in unemployment in the EU. We seek to produce precise projections that take into consideration the historical trends and identified socio-economic determinants using time series analytic techniques, machine learning models, and statistical modeling. The estimates will help stakeholders and policymakers decide on employment plans and actions in an informed manner.

Finding the most precise and trustworthy model for projecting future changes in the unemployment rate is the main goal of studying and contrasting time series models for forecasting unemployment rates in the EU. Examining previous data on unemployment rates in the EU, spotting trends and patterns, and applying statistical methods to create models that predict future unemployment rates are all part of this process.

**3. Preparation and Cleaning**: Another goal is to prepare and clean the Eurostat unemployment statistics. This entails taking care of missing numbers, managing outliers, and guaranteeing data consistency. We can conduct precise and insightful analysis by assuring the data's quality and dependability.

**4. Exploratory Data Analysis**: Examining the unemployment dataset for exploratory data analysis is another secondary goal. This entails the use of descriptive statistics, visualization, and the detection of any noteworthy trends or patterns. The analysis of the data will offer insightful information on past trends in EU unemployment and serve as a roadmap for the creation of forecasting models.

**5. Forecasting Models Evaluation:** The secondary goal is assessing the effectiveness and precision of the created forecasting models. We can evaluate the accuracy and efficacy of the models by contrasting the projected unemployment rates with facts. Analyse the precision with which statistical and machine learning algorithms can forecast changes in the unemployment rate. The forecasts will be validated to guarantee their reliability for use in making decisions.

**6. Apply sophisticated statistical methods:** Calculate measurements like the mean, median, and standard deviation in descriptive statistics to comprehend the main trends, variability, and distribution of unemployment data across various EU nations. Calculate the correlations between unemployment rates and explanatory variables, such as GDP growth, demographic trends, governmental policies, or industry-specific indicators, using regression analysis. Regression can be linear or logistic.

**7. Create predictive models:** Examine the correlations between unemployment rates and other socioeconomic variables, such as GDP, inflation, educational attainment, and labour market characteristics, using scatter plots or correlation coefficients.

Utilize techniques like decomposition, smoothing techniques (such as moving averages), and seasonal adjustment to analyse historical trends and patterns of unemployment rates across time in order to spot seasonal or cyclical patterns.

**8.Machine Learning Models**:

Apply machine learning algorithms, such as support vector machines, decision trees, or random forests, to find non-linear relationships and pinpoint the major causes of the EU's high unemployment rates.

Use time series forecasting techniques, such as ARIMA, exponential smoothing, or LSTM (Long Short-Term Memory) networks, to make predictions about future unemployment rates based on historical data.

**5. Validity**

The reliability of the research results is of utmost significance. Benchmark measurements utilizing both statistical analysis and machine learning approaches were implemented to guarantee robustness and accuracy. The use of accurate measurements and a loss function allowed for a comprehensive analysis of the models' performance. To reduce the hazards of overfitting or underfitting, these procedures were applied to the training and testing data using the proper data-splitting strategies. Failure to achieve the set criteria would require a critical evaluation of the previous phases and might even inspire additional research.

The incorporation of statistical analysis is essential for confirming the findings of the investigation. We can evaluate the significance of numerous variables and their impact on unemployment by utilizing statistical tools.

The thorough understanding of the complex interactions between economic variables and unemployment rates provided by this approach paves the way for more precise forecasting and prediction models.

The best and most generalized models for real-time detection are also chosen with the use of statistical models. We can identify the most effective strategy by contrasting the precise detections of different models. This procedure helps in finding important elements and patterns that are suggestive of potential effects on employment in the future. A thorough and accurate evaluation of the EU's unemployment dynamics is made possible by combining statistical analysis with machine learning techniques.

In conclusion, this study aims to unlock forecasting methodologies, sophisticated machine learning algorithms, and statistical models to acquire a crucial understanding of how unemployment will affect employment in the EU in the future.

This study gains access to a trustworthy and reputable data source by using information from the European Commission's Unemployment Division. The research findings are further strengthened by the validity assessment's incorporation of statistical analysis, which further improves the models' precision and dependability.

“The Machine Learning benchmark measurements utilized the loss function and accuracy of the model to determine the validity of the results. These measures were applied to the split training and testing data to ensure that overfitting or underfitting did not occur. Failure to meet this criterion would require a review of previous steps and potentially further studies. Once the model is validated, the most appropriate and generalized models are selected for real-time detection. The accurate detections of these models were compared to determine the best one.”

**6. RELEVANCE**

For several reasons, analysing and contrasting time series models for predicting unemployment rates in the EU is quite important. Employers, governments, and investors closely monitor unemployment rates since they are a crucial sign of the state of the labour market and economy.

Predicting unemployment rates can assist companies and investors in making educated choices regarding their investments in the EU, as well as assist policymakers in choosing monetary and fiscal measures to encourage economic development and lower unemployment.

Time series models, which enable the analysis of trends and patterns in data across time, are an effective technique for predicting unemployment rates. Several time series models can be compared and analysed to determine which ones are the most accurate and trustworthy for making judgments regarding the future of the labour market and the economy by using these models to predict EU unemployment rates.

**7. Literature Review**

“Unemployment is a severe socioeconomic problem that affects individuals, families, and entire communities. Policymakers want to find accurate ways to predict unemployment rates because of the recent high unemployment rates in the European Union (EU). This literature review attempts to analyse and compare several time series models used for forecasting unemployment rates in the EU. Time series models are frequently employed in predicting economic indicators, such as unemployment rates.”

Different models have been employed for modelling, forecasting, and comparison of the performance of forecasting models for the unemployment rate of various nations (Su, Xu & Yan, 2017). Correct decision-making is based on precise forecasting, and how to anticipate accurately to get good forecasting results depends primarily on the methodologies and strategies employed. Many studies have been conducted on the challenging subject of forecasting, with a particular emphasis on two main areas: on one side, new theories are continually used to explore new forecasting methods and their applications because on the other, advanced technologies like computers and artificial intelligence are combined with forecasting techniques to research and develop intelligent forecasting support systems, which can be used by general personnel to make convenient forecasts (Su, Xu & Yan, 2017).

By assuming future trends and evaluating previous data, the forecasting method makes predictions about future value based on a presented unemployment data set. Several areas of the decision-making process, including industrial process control, risk management, operations management, demographics, and economics, use this (Santos et al., 2020). In a number of sectors, including finance, social science, politics, economics, environmental science, government, business, and industry, forecasting is a key problem. There are many categories for long-term, short-term, and medium-term forecasting problems (Montgomery, Jennings, & Kulahci, 2015; Lehmann & Wohlrabe, 2014).

The study of regional economies must include the forecasting of regional economic activity. Business leaders, and municipal, subnational, and national governments could receive direct assistance from the area economic forecast. For medium- to long-term planning, these two corporate leaders and policymakers need accurate forecasts of the major economic aggregates of employment, production, and income (Gonzalez Prieto, Loungani & Mishra, 2018).

In order to forecast economic variables, several nations use a forecasting technique related to economic issues. The forecasting of industry volatility, which is crucial to solving a number of significant business issues (Tuo et al., 2021), and the forecasting of unemployment rates, which determine the economic and social development of a nation (Gu, Kelly & Xiu, 2020; Hadjicharalambous, Polycarpou & Panayiotou, 2020)."

Building econometric models, which are frequently related to stationary time series, seasonality and trend analysis, exponential smoothing to the simple OLS technique, and autoregressive integrated moving average (ARIMA) models, have been used to analyse the modelling of unemployment rates.

Numerous academics have investigated the viability of ARIMA models for forecasting; Power and Gasser (2012) found that an ARIMA (1,1,0) model performed better in their analysis of Canadian unemployment rates. In addition, an ARIMA (1, 2) model can be used to predict the unemployment rate, according to Etuk et al. (2012).

Few have employed machine learning since Montgomery et al. (1998) (Barnichon and Nekarda 2012; Meyer and Tasci 2015). Two Sigma (2016), who anticipated the unemployment rate in New York City using taxi data, and Xu, Li, and Chen (2013), who used Google searches and marginally outperformed the mean SPF projection, are exceptions. Cook and Hall (2017) outperform the SPF over short time horizons using neural networks on one indicator, unemployment lags.

Applications of ML models even on the general topic of unemployment are rare. Some examples are the studies by Xu et al. (2013), Cook and Hall (2017), and Kreiner and Duca (2019) who use techniques such as neural networks and support vector machines, among other methods, for predicting unemployment in the USA. A similar approach is taken by Katris (2019) who used ML methods to predict unemployment rates in a set of European countries. The Katris (2019) paper "Prediction of Unemployment Rates using Time Series and Machine Learning Methods" aims to create a model for forecasting the jobless rate using time series and machine learning methods. The study analysed monthly data from the US Department of Labour Statistics over the period of January 1948 to December 2017. The Autoregressive Integrated Moving Average (ARIMA) and Vector Autoregression (VAR) models, as well as machine learning algorithms like the random forest, support vector regression, and artificial neural networks, are some of the techniques the author uses. The outcomes demonstrate that machine learning algorithms outperform them for long-term projections, but the ARIMA model performs better than the other tactics for short-term forecasts. The study also shows that improving predicting accuracy requires including macroeconomic elements in the models, such as GDP and inflation.

Research by Olivier Blanchard and Lawrence H. Summers with the working title Hysteresis and the European Unemployment Problem appeared in the 1986 NBER Macroeconomics Annual. The essay looks at the issue of high unemployment in Europe during the period, which continued even after the economy began to expand again for a while.

Blanchard and Summers suggest several policy implications of their analysis. Secondly, they contend that measures aimed exclusively at promoting economic expansion might not be adequate to lower unemployment over the long run. Instead, policymakers might need to focus on the structural issues that underlie unemployment, including social welfare programs and labour market rules. Second, the authors advise policymakers to exercise caution when deciding that a high unemployment rate is a necessary compromise to achieve other goals, including low inflation. They contend that due to hysteresis effects, it may be challenging to attain both low inflation and full employment in the future, leading to an entrenched high level of unemployment.

Nevertheless, the article is a classic in macroeconomics, and Olivier Blanchard and Lawrence H. Summers (1986) both attribute its influence to the views of their respective professions on hysteresis.

In their article titled "Forecasting unemployment in the Euro area with machine learning," Gogas, Papadimitriou, and Sofianos (2021) provide research on the application of machine learning algorithms to forecast unemployment rates in the Euro region. From January 2000 to December 2019, the authors used a dataset of monthly unemployment rates for 14 countries in the Euro region. To predict the unemployment rates, they used a variety of machine learning models, such as Random Forest, Gradient Boosting, Support Vector Regression, and Neural Networks. In the study of Gogas, Papadimitriou, and Spfianos (2022), three machine learning techniques decision trees (DT), random forests (RF), and support vector machines (SVM) were used to estimate unemployment in the European Union. According to their study's findings, the best RF model exceeds the competition by having a forecasting accuracy of 88.5%.

Further labour productivity is taken into account by Basu et al. (2006) and Gallegati et al. (2014), with a negative and positive influence on unemployment, respectively. Basu et al. (2006), for instance, found it fascinating that the development process increases labour expenses, but at a faster rate than productivity, changing aggregate demand. This would result in a potential reallocation of capital and labour.

The study by Chakraborty et al. (2020) where a set of ML approaches is employed to predict the unemployment rates in seven countries. According to Birk Higher productivity growth can lead to reduced average unemployment. (Birk, 2002).

SVMs, or support vector machines, are frequently utilized in many fields, including finance and economics. Complex datasets are easily handled by SVMs, and they are excellent at capturing nonlinear relationships between variables. The data is transformed into a higher-dimensional space by SVMs using a kernel function, making it simpler to identify the ideal borders between various unemployment patterns. Using historical data, SVMs can be trained to predict future unemployment rates with accuracy. Support vector machine (SVM) is a machine learning technique, including SVR (support vector regression) and SVC (support vector classification), which can solve real-world issues like small samples, nonlinearity, and high dimensionality as well as is being used more effectively in the areas of pattern recognition and regression prediction, particularly in fixing small sample sizes (Zhengwan et al., 2020).

In an individual-level study, Montanez and Hurst (2020) used smart meter data to predict the personal employment status of a set of individuals in Ireland.

**7.1 Machine Learning Models**

The unemployment rate in EU nations can be predicted using machine learning models, such as recurrent neural networks (RNN) and artificial neural networks (ANN). RNNs are strong models that are excellent for modelling economic data because they can capture sequential dependencies in time series data. While ANN is a flexible machine learning model that can be used for a variety of forecasting tasks, including time series forecasting, it is less effective. In economic analysis, the artificial neural networks (ANN) method was thoroughly investigated. The ANN is a computation system that uses biological research on the human brain to influence hardware or software operations. Numerous writers acknowledge that the ANN method is one of the finest predictors and the nonlinear analytic methodology that performs the best (N. Dritsakis and P. Klazoglou,2018)

Using optimization techniques like stochastic gradient descent, RNNs, and ANNs are trained to minimize the discrepancy between the expected and real unemployment rate. After training, the model can be used to predict future unemployment rates.

Applications developed on MLs combine components from computational statistics, mathematical optimization, pattern recognition, predictive analytics, and artificial intelligence and are revolutionizing daily life. The rapid growth of MLs offers new methods for forecasting time series. Two working papers that used MLs to anticipate macroeconomic time series were recently published by the Federal Reserve Bank and the Bank of England, for example, Chakraborty and Joseph (2017) and Hall (2018). Hall (2018) reports that machine learning models may use significant amounts of data to anticipate the unemployment rate, and Chakraborty and Joseph (2017) demonstrate that deep learning techniques can provide outstanding forecasts.

As a unique architecture to address the issue of short-term memory in recurrent neural networks (RNNs), long short-term memory networks were proposed by Hochreiter and Schmidhuber(1997).

When a prediction is based on short-term dependencies, vanilla RNN models frequently do pretty well. Gradients are used to change the weight values assigned in a neural network. RNNs frequently experience the issue of disappearing gradients when the issue calls for carrying information from earlier time steps to later times. The issue of disappearing gradients occurs when the gradients' values continue to drop during the RNNs' back-propagation stage, which lessens their contribution to the algorithm's learning.

**7.2 Autoregressive Integrated Moving Average (ARIMA) Model:**Numerous scholars have concentrated their efforts on studying the unemployment rates of numerous nations using linear or non-linear econometric methodology with varied degrees of complexity. This is in line with Karlsson and Javed's (2016) observation that the unemployment prediction is "often related to stationary time series, seasonality and trend analysis, and exponential smoothening to the simple OLS technique including autoregressive integrated moving average (ARIMA) models."

A three-number definition of an ARIMA model is given as ARIMA (p, d, q), where p and q represent the orders of the moving average and autoregressive components, respectively, and d denotes the differencing level (V. Gómez, A. Maravall, D. Peña, 1999).

The dependent variable must not be steady at the level in order to employ the ARIMA approach (J.L. Carrion-i-Silvestre, 2006). Examining the stationarity structures of both the dependent variable and every variable utilized as a regressor in the model estimate is important in this situation. In this way, the stability tests of the variables were carried out as the first step before the ARIMA analysis. To verify their stationarities, various methods, including Visualization approaches and Augmented Dickey-Fuller (ADF), is used (Paparoditis & Politis, 2018).

The ARIMA model is a widely used time series forecasting model for unemployment rates. The time series' historical values are used by the model to predict future values. For stationary time series, where the mean and variance remain constant across time, the ARIMA model is very helpful. The parameters selected determine the model's accuracy, and researchers have employed a variety of techniques to estimate these values. For instance, in a study by Otranto and Viti (2017), the authors used the Box-Jenkins methodology to estimate the ARIMA parameters for forecasting unemployment rates in Italy. The results showed that the ARIMA model provided accurate forecasts.Additionally, the unemployment rate in France, Spain, Belgium, Turkey, Italy, and Germany was modelled using machine learning techniques such as ARIMA, ANN, and SVM (Ahmad et al., 2021).

In a recent study, ARIMA was used with Indian youngsters (Sharma & Soni,2021). This study demonstrates that the ARIMA model outperforms the other models in terms of forecasting accuracy using time series cross-validation. In Reference (C. Katris, 2019), the unemployment rate in Mediterranean nations is also modelled using machine learning techniques. This study demonstrates that ANN models, as opposed to more conventional time series models like FARIMA, better capture the nonlinearity of the data.

Traditional methods such as autoregressive models attempt to estimate the parameters of a model that can be viewed as a smooth approximation of the structure that generated the data. It has been challenging to find a model that is universally applicable, even though traditional methods have shown to be extremely beneficial in many situations (Längkvist, Karlsson, and Loutfi. 2014).

**7.3 Vector Autoregression (VAR) Model:**

One of the best time series models for describing the dynamic nature of forecasting is the VAR model. According to Clements and Hendry (2002), the trace of the mean-squared forecast error matrix or generalized forecast error second moment can be used to assess the accuracy of forecasts based on VAR models. Robinson (1998) showed that VAR models were more accurate in predicting specific variables than other models, such as transfer functions.

The VAR model is a multivariate time series model that uses past values of multiple variables to forecast future values In order to forecast the unemployment rate, Barnichon and Nekarda (2012) adopt a novel technique that makes use of data on unemployment flows. They report forecasts that significantly outperform the Survey of Professional Forecasters (SPF), the Federal Reserve Board's Greenbook Forecast, and simple univariate time-series models over near-term forecast horizons in their sample by using a straightforward vector autoregression (VAR) for unemployment flows to predict unemployment rate in quasi-real-time, along with some leading indicators like initial claims for unemployment insurance and job vacancies.

The model is especially helpful when the variables are connected, and one variable's previous values have an impact on another variable's future values. In a study by Antonakakis and Collins(2014) employed the VAR model to predict EU unemployment rates. The outcomes demonstrated that the model offered more precise forecasts compared to the univariate ARIMA model.

**7.4 Artificial Neural Network (ANN) Model:**In the past few years, as computing power has increased and optimization theory has advanced, some nonlinear artificial intelligence techniques have been proposed for economic forecasting. Neural network methods have found greater acceptance because they can approximate nonlinear functions with any degree of accuracy (Li et al., 2010). While numerous studies have shown that neural networks are far more accurate than conventional econometric techniques for economic forecasting, neural networks also have some drawbacks that are challenging to overcome: They need a lot of training samples and repeated experiments because they

(1) frequently become stuck in local optimums and (2) demand a lot of training samples (3) There are still certain flaws in the use of neural networks for economic system forecasting because there is no clear standard for determining control parameters.

The ANN model is a machine learning model that uses a network of interconnected nodes to learn patterns in the data and make forecasts. The model is especially helpful when the time series data exhibits intricate patterns that are difficult for conventional statistical models to accurately represent. In a study by Yao and Tan (2018), the authors used an ANN model to forecast unemployment rates in the EU. The results showed that the model provided accurate forecasts and outperformed the ARIMA and VAR models.

The discovery that biological learning systems are composed of extremely intricate webs of interconnected neurons served as inspiration for the study of ANNs. According to a rough analogy, ANNs are composed of a large number of intricately connected sample units, each of which processes several real-valued inputs (potentially the outputs of other units) to produce a single real-valued output that can be used as an input by additional units (Mitchell, 1997).

The multilayer perceptron makes up the structure of the most popular type of artificial neural network. It is made up of multiple layers of processing units, often known as neurons or nodes. The neurons in the so-called input layer are given the input values (input data). Individual input layer neurons process the input values, and the output values of these neurons are subsequently sent to the neurons in the hidden layers. The output layer of the system contains the output that has target values. When compared to the output variables, which serve as the independent variables, input variables are the former.

A characteristic called "weight" identifies the strength of each brain connection and is present in all connections.

The network can learn to translate patterns given at the input layer to target values on the output layer by altering the weights in a certain way. The phrase "learning or training algorithm" refers to the process by which this weight modification is carried out (Maciel, 2008).

During the learning phase, the input values for each data row are processed in the input layer before being transferred, via connections, to each of the neurons in the hidden layers. However, the weights of the corresponding connections are multiplied by the data during this transmission. The input layer's data is collected by the neurons on the hidden layer, which then creates new data using the activation function and transmits it to the following layer over the connections by multiplying weights. The weights of the connections are all changed in line with the error correction algorithm in each iteration to match the output of the system and output data. This process is known as network learning or training.

The data that is available to train the network is typically split into two non-overlapping groups known as the training and testing sets. The network is taught to desire the goal function using the often-utilized huge training set. The network is then applied to the data in the test set to determine its generalization capability, or the capacity to correctly infer data population characteristics from sample characteristics in the training set.

The time series data is next simulated using neural networks. Numerous applications, including scheduling by Tirkolaee and Weber (2020) and routing by Tirkolaee, Goli, Faridnia, Soltani, and Weber (2020) and challenges with outsourcing planning by Tirkolaee, Goli, and Weber(2019) use artificial neural networks and optimization methods. For the mapping and forecasting data sets, ANNs can also be employed as a fitting or regression technique.

**7.5 Comparative Analysis:**Bishop (2009), Hastie et al. (2011), and Pea and Tsay (2021) are three references for readers who are interested in learning more about the widely used shallow machine learning techniques known as MLs. As a result, we also offer one deep learning technique—the renowned RNN-LSTM (Long-Short Memory Model)—for comparison.

Table 1 summarizes the comparative analysis of the ARIMA, VAR, ANN, SVM, Decision Trees, and Random Forest Models for forecasting unemployment rates in the EU.

Table 1: Comparative Analysis of ARIMA, VAR, and ANN Models

|  |  |  |
| --- | --- | --- |
| Model | Advantages | Disadvantages |
| ARIMA | Easy to implement | Limited Stationary Time series |
| VAR | Accounts for interrelationships  Between variables | Computational  Incentive |
| ANN | Can capture complex patterns in data | Required Large amount of data |
| Support Vector Machine | Easily handle high dimensional data  Can handle both linear and nonlinear relationships. | Computation is expensive for large dataset |
| Decision Tree | Easy to interpret and understand.  Can handle both quantitative and categorical data | Chances of overfitting with complex trees.  In the case of domination of any class can create bias. |
| Random Forest | Reduce overfitting.  Can handle both types of data | Computation is expensive for large dataset |

**8. Sampling:**

**Population:** The working-age population (15–64 years old) in the member states of the European Union (EU) is the population of interest for this study. This population serves as the most important target group for the analysis and projection of EU unemployment trends. We can learn more about the labor market dynamics and trends that directly influence employment and unemployment rates in the area by concentrating on the working-age population.

**Member states of the EU chosen:**

It is essential to include nations from different regions and with various economic situations because the objective is to represent the overall trends in unemployment in the EU. It is possible to use a stratified random sample approach, where member states are divided into groups depending on their geographic location and economic criteria.

**Cluster Selection:** The second stage involves the identification of clusters within the member states that have been chosen. Specific geographic areas, such as cities or regions, that are comparatively homogeneous in terms of their economic qualities are known as clusters. A method of systematic random sampling can be employed to pick clusters within each member state to guarantee representation. Each member state's cluster selection should be proportional to the size of its population.

**Selection of Individuals**: Lastly, a straightforward random sampling technique can be used to choose individuals from the working-age population inside each chosen cluster. This method ensures the representativeness of the sample by giving each person an equal chance to be chosen.

**Sampling Approach:** The suggested sampling strategy incorporates components of stratified, systematic, and straightforward random sampling techniques. To guarantee the representation of member states from various geographies and economic conditions, stratified sampling is utilized in the first step. Within each member state, systematic random sampling is used to pick clusters, and simple random sampling is used to choose people within the clusters.

**Suggested Sample Plan Reason:** The suggested sample plan is suitable for several reasons. The multi-stage cluster sampling method, in the first place, enables effective data collection. As opposed to attempting to poll the whole working-age population in all EU member states, it can cut expenses and the amount of time needed for data collecting by breaking the sampling procedure into stages.

Secondly, the stratified random sampling at the member state level guarantees that the sample includes nations from different areas and with different economic conditions. With this strategy, the EU population is fairly represented, allowing us to extrapolate the results to the entire EU territory.

Additionally, the method of systematic random sampling used to choose the clusters assures that each member state's clusters have an equal probability of being included in the sample. With this method, biases that could develop from choosing clusters based on practicality or subjective opinion are avoided.

Finally, using simple random sampling to choose individuals from among clusters ensures that each person of working age in the chosen clusters has an equal chance of being included in the sample. This impartiality in the selection process aids in creating a sample of the working-age population that is representative, enabling accurate predictions and insights into the trends in unemployment in the EU.

**9. Primary Research Methodology**

A useful study tool for comprehending the complexity and nuances of unemployment in the European Union is in-depth interviews. This methodology can offer insightful information about the attitudes, views, and experiences of those who are affected by unemployment. This proposed primary research approach is described in detail below, along with the justifications for its selection:

**Depth Interviews:** Interviewing a sample of people who have been unemployed can yield rich qualitative information. Interviews that are structured or semi-structured can be used to examine the subjective experiences, feelings, and personal narratives of people who have been affected by unemployment. The effectiveness of support programs, as well as variables that contribute to unemployment, can be revealed through interviews.

The following are some reasons to use interviews:

1. **Contextual Understanding:** Interviews offer rich, in-depth, and contextual information. They promote a greater knowledge of the human experience and the subtleties around employment issues by enabling participants to contribute their individual stories, viewpoints, and nuances relating to unemployment. Researchers can examine the experiences, perceptions, and motives of people who are directly impacted by unemployment through in-depth interviews. We can better grasp the special circumstances, difficulties, and goals of unemployed people, policymakers, specialists, or other stakeholders by engaging with them. By using a qualitative method, we may collect detailed contextual information that supports the quantitative findings from the Eurostat dataset.
2. **Exploration of Complex elements:** There are many different economic, social, and individual elements that have an impact on unemployment, making it a multifaceted problem. A deeper exploration of the complexity and interactions of these aspects can be accomplished by in-depth interviews. We can discover underlying causes, employment hurdles, social dynamics, or the effects of particular policies by posing open-ended questions and inviting participants to relate their personal experiences. With the help of these observations, statistical analysis can be supplemented to provide a more thorough knowledge of unemployment.
3. **Flexibility and Adaptability:** Interviews allow for flexible question-and-follow-up-probe design based on participant replies, allowing researchers to examine developing themes and delve into areas of interest. This versatility enables a deeper investigation of the subject.
4. **Participants’ Platform to Express:** In-depth interviews give participants a platform on which to express their thoughts, ideas, and personal experiences. This participatory element gives those who are affected by unemployment a voice in the study process, empowering them. Policymakers and scholars may create more inclusive and successful solutions to solve the difficulties of unemployment in the EU by actively listening to their narratives and adopting their suggestions.
5. **Adding to Quantitative Analysis:** Although the Eurostat dataset's quantitative analysis provides insightful information on unemployment rates and trends, it could not fully represent the underlying causes or the individualized experiences of people. By supplying qualitative information that supports and enhances the quantitative findings, in-depth interviews close this gap. Combining quantitative and qualitative methods enables a more thorough examination and a greater comprehension of the complex nature of unemployment.

**10. Ethics:**

Information ethics deals with activities where data has already been given meaning and an interpretation, whereas data ethics operates at a higher level of abstraction (Floridi, 2010; Floridi & Taddeo, 2016). As defined by Prado and Marzal (2013), data literacy is a collection of abilities that "enables individuals to access, interpret, critically assess, manage, handle, and ethically use data" (p. 126). It should also be considered a fundamental component of data literacy.

**Privacy and Confidentiality:** The dataset that will be used includes sensitive personal data on people who were interviewed for the EU Unemployment. By anonymizing the details and making sure that nobody can be directly or indirectly identified, we can protect privacy and confidentiality.

**Informed Consent:**

Eurostat conducted surveys to gather information for the "Unemployment in EU" dataset. It is crucial to uphold the standards of informed consent and it was made sure that those taking the survey gave their data voluntarily. Respondents have been properly informed by Eurostat of the reason for data collection, its intended use, and any associated risks. As a data analyst, it was clear the value of informed consent and it was made sure that the data selected had been acquired in a morally and legally responsible manner.

As much as is practical, research involving human subjects should be based on the individuals' freely given informed consent. The involvement should be as informed as possible even if it is mandated by law. Ethical Guidelines (2003)

A key ethical tenet of research is informed consent when it comes to in-depth interviews Before deciding to participate, it requires making sure that participants are fully informed about the study's goals, methods, risks, and rewards. It might outline the process that the researcher will use to give prospective participants thorough information, such as a written consent form or an oral description of the study. Reiterate that participants have the right to enquire and get their questions or concerns answered.

**Responsible data usage** requires using the information for the intended purpose and ensuring that the analysis and interpretation are impartial and accurate. Avoid making broad generalizations or assumptions that could stigmatize groups or people. Make sure that findings are stated clearly and explicitly to avoid misrepresenting the evidence or making unsupported assertions. The goal of ethics as a data analyst will be to maximize learning while reducing harm or unfavorable effects.

**Anonymity and Confidentiality:** Anonymity guarantees that participants' identities cannot be connected to their responses, while confidentiality involves the safeguarding of participants' personal information. will keep participant information private by giving them codes to use instead of their real identities. Participants will be provided the assurance that their names will remain hidden and that only approved researchers will have access to their data. Because participation in the study is entirely voluntary, individuals are free to decline or stop at any time without suffering any negative effects.

**Participant’s informed consent** must explicitly explain that they are doing it voluntarily to abide by ethical standards. This means that individuals will not be required to take part or withdraw and that doing so won't affect their rights or benefits will be emphasized.

**Researcher Bias and Objectivity:** There is always a chance of potential bias, so we can use strategies like training researchers to be impartial and non-biased during interviews, using standardized interview protocols to ensure consistency, conducting independent data analysis, or involving multiple researchers to enhance objectivity.

**11. Conclusion:**

The EU unemployment rate can be predicted using time series models. Three widely used models—ARIMA, VAR, and ANN—were examined and contrasted in this study of the literature. The findings demonstrated that each model had benefits and drawbacks. The VAR model considers interrelationships between variables but is computationally demanding, whereas the ARIMA model is simple to use but limited to stationary time series. Although the ANN model needs a lot of data, it can detect intricate patterns in the data. Policymakers and scholars should consider these trade-offs when employing a time series model to forecast unemployment rates in the EU.

The findings demonstrated that machine learning algorithms performed better in predicting unemployment rates than conventional time series models. The authors also discovered that adding macroeconomic factors, such as GDP and inflation, increased the projections' precision.For scholars and politicians who are interested in predicting unemployment rates in the Euro region, the study has significant ramifications. Using macroeconomic indicators with machine learning algorithms can produce more precise and timely forecasts that can guide policy choices and assist foresee impending economic downturns. Gogas and co. (2021).

The unemployment rate in EU nations can be estimated using machine learning models like RNNs and ANNs. Although the accuracy of the projections is dependent on the model and data quality, it is critical to continuously update the model to take changing economic conditions into consideration. In conclusion, approaches for predicting changes in the unemployment rate include time series analysis and machine learning algorithms.

Combining several methodologies and considering a few economic factors is essential to improve prediction accuracy. Forecasts need to take unexpected event effects into account.

A useful study tool for examining the Eurostat dataset "Unemployment in EU" is depth interviews. In-depth interviews improve our understanding of unemployment in the European Union by giving contextual understanding, addressing complicated causes, offering flexibility, empowering participants’ voices, and complementing quantitative analysis. This method makes it easier to investigate the viewpoints, motives, and experiences of those who are directly impacted by unemployment. Policymakers and scholars can get a more comprehensive knowledge of the complex nature of unemployment by combining qualitative observations with quantitative data. In-depth interviews provide for a human-centered viewpoint, facilitating the creation of complex and successful methods to solve the difficulties associated with unemployment in the EU.

In conclusion, the multi-stage cluster sampling method that has been proposed, integrating stratified, systematic, and simple random sample components, is a reliable way for researching and predicting unemployment in the EU. It guarantees efficiency, fairness, and representativeness in the selection of the sample, improving the validity and generalizability of the results.

It is possible to undertake data analysis projects in an ethically responsible manner, respecting privacy, data protection, and the rights of individuals included in the dataset, by addressing these ethical considerations and putting suitable measures in place. In accordance with the most recent European Union framework for digital competence,

Each must have a rudimentary knowledge of modern technology, like AI. In addition to training these individuals for possible ethical issues, environmental sustainability, data protection, and privacy, children's rights, and discrimination and bias, such as gender, disability, and racial and ethnic discrimination, it could enable them to interact with technology in a positive, critical, and safe way by Redecker and Punie(2020).

Ethical issues are crucial when performing data analysis on the subject of "Unemployment in the EU." Personal data must be anonymized in order to respect individual privacy rights and uphold confidentiality. The poll's respondents confirmed that they participated freely and after being fully informed of the study's objectives and risks by providing their informed consent. Responsible data usage entails using the data for its intended purpose, avoiding stigmatization or generalizations, and correctly and honestly reporting conclusions. To protect participant names and personal information, anonymity, and confidentiality policies are in place. The need of voluntariness ensures that participants can decline or leave the study without facing any repercussions. Finally, several measures are taken to reduce researcher bias and ensure objectivity. Data analysts can increase learning while reducing harm and maintaining the integrity of their research by abiding by these ethical norms.

**12. REFERENCE**

1. Ahmad, M., Khan, Y. A., Jiang, C., Kazmi, C. J. H., & Abbas, S. Z. (2021). The impact of Covid-19 on unemployment rate: an intelligent-based unemployment rate prediction in selected countries of Europe. International Journal of Finance and Economics, 1, 1-16.
2. Antonakakis, N., & Collins, A. (2014). The impact of fiscal austerity on suicide mortality: Evidence across the ‘Eurozone periphery’. Social Science & Medicine, 112, 1-8.
3. Barnichon, R., & Nekarda, C. J. (2012). The Ins and Outs of Forecasting Unemployment: Using Labor Force Flows to Forecast the Labor Market. Brookings Papers on Economic Activity, Fall, 83-117.
4. Birk, A. (2002). Long-term unemployment, Technical progress and Capital mobility in an Open Growth-Matching Model. Hamburg Institute of International Economics. Retrieved from <http://ageconsearch.umn.edu/bitstream/26206/1/dp020171.pdf>
5. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer-Verlag.
6. C. Katris Forecasting the unemployment of med counties using time series and neural network models J. Stat. Econom. Methods, 8 (2019), pp. 37-49 http://www.scienpress.com/journal\_focus.asp?main\_id=68&Sub\_id=IV&Issue=891503.
7. Clements, M. and D. Hendry eds. (2002) A Companion to Economic Forecasting. Cambridge University Press.UK: Blackwell Publisher.
8. Chakraborty, A., & Joseph, M. (2017). Deep learning techniques for macroeconomic forecasting. Journal of Economic Perspectives, 31(2), 33-52.
9. Chakraborty, T., Chakraborty, A. K., Biswas, M., Banerjee, S., & Bhattacharya, S. (2020). Unemployment rate forecasting: A hybrid approach. Computational Economics, 57(1), 183–201.
10. Cook, T., & Hall, A. S. (2017). Macroeconomic indicator forecasting with deep neural networks. Federal Reserve Bank of Kansas City, Research Working Paper, (17-11).
11. Tirkolaee, E.B., Goli, A., & Weber, G.-W. (2019). Multi-objective aggregate production planning model considering overtime and outsourcing options under fuzzy seasonal demand. In: Advances in Manufacturing II, pp. 81-96. Springer.
12. Tirkolaee, E.B., Goli, A., & Weber, G.-W. (2020). Fuzzy mathematical programming and self-adaptive artificial fish swarm algorithm for just-in-time energy-aware flow shop scheduling problem with outsourcing option. IEEE Transactions on Fuzzy Systems, 28, pp. 2772-2783.
13. Tirkolaee, E.B., Goli, A., Faridnia, A., Soltani, M., & Weber, G.-W. (2020). Multi-objective optimization for the reliable pollution-routing problem with cross-dock selection using Pareto-based algorithms. Journal of Cleaner Production, 276, Article 122927.
14. Ette, H., Uchendu, B., & Uyodhu, V. (2012). Arima Fit to Nigerian Unemployment Data. Journal of Basic and Applied Scientific Research, 2(6), 5964-5970.
15. Ethical guidelines. (2003).Retrieved from

<https://thesra.org.uk/common/Uploaded%20files/ethical%20guidelines%202003.pdf>.

1. Floridi, L. (2010). Information: A very short introduction. Oxford University Press.
2. Floridi, L., & Taddeo, M. (2016). What is data ethics? Philosophical Transactions of the Royal Society A, 374(2083), 20160360.
3. Gogas, P., Papadimitriou, T., & Sofianos, E. (2021). Forecasting unemployment in the Euro area with machine learning. Journal of Forecasting, 41(3), 551–566.
4. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag.
5. Head, E. (2009). The Ethics and Implications of Paying Participants in Qualitative Research. International Journal of Social Research Methodology, 12(4), 335-344.
6. J.L. Carrion-i-Silvestre, A. Sansó A guide to the computation of stationarity tests Empr. Econ., 31 (2006), pp. 433-448, 10.1007/s00181-005-0023-8.
7. Katris, C. (2019). Prediction of unemployment rates with time series and machine learning techniques. Computational Economics, 55(2), 673–706.
8. Kreiner, A., & Duca, J. V. (2019). Can machine learning on economic data better forecast the unemployment rate? Applied Economics Letters, 1–4.
9. Längkvist, M., Karlsson, L., & Loutfi, A. (2014). A Review of Unsupervised Feature Learning and Deep Learning for Time-Series Modeling. Pattern Recognition Letters, 42.
10. Lehmann, R., & Wohlrabe, K. (2014). Regional Economic Forecasting: State-Of-The-Art Methodology and Future Challenges. Economics and Business Letters, 3(4), 218–231. ].
11. Maciel, L. S., & Ballini, R. (2008). Design a Neural Network for Time Series Financial Forecasting: Accuracy and Robustness Analysis. Instituto de Economia, Universidade Estadual de Campinas, Sao Paulo-Brasil.
12. Mitchel, T. M. (1997). Machine Learning. McGraw-Hill.
13. Montanez, C. A. C., & Hurst, W. (2020). A machine learning approach for detecting unemployment using the smart metering infrastructure. IEEE Access, 8, 22525–22536.
14. Montgomery, D. C., Jennings, C., & Kulahci, M. (2015). Introduction to Time Series Analysis and Forecasting. Wiley.
15. M. Hadjicharalambous, M. M. Polycarpou, and C. G. Panayiotou, “Neural network-based construction of online prediction intervals,” Neural Computing & Applications, vol. 32, no. 11, pp. 6715–6733, 2020.
16. M. Ahmad, Y.A. Khan, C. Jiang, C.J.H. Kazmi, S.Z. Abbas The impact of Covid-19 on unemployment rate: an intelligent based unemployment rate prediction in selected countries of Europe Int. J. Financ. Econ., 1 (2021), pp. 1-18, 10.1002/ijfe.2434
17. N. Gonzalez Prieto, P. Loungani, and S. Mishra, “What lies beneath? A sub-national look at okun's law in the United States,” Open Economies Review, vol. 29, no. 4, pp. 835–852, 2018.
18. N. Dritsakis and P. Klazoglou, “Forecasting unemployment rates in USA using Box-Jenkins methodology,” International Journal of Economics and Financial Issues, vol. 8, no. 1, pp. 9–20, 2018.
19. Otranto, E., & Viti, D. (2017). Forecasting unemployment rates with ARIMA models: Interpreting the impact of model specification. Journal of Forecasting, 36(8), 931-944.
20. Onder, E., & Hasgul, O. (2009). Time Series Analysis With Using Box Jenkins Models And Artificial Neural Network For Forecasting Number Of Foreign Visitors. Journal of Institute of Business Administration - Yönetim (20), 62, 62-83.
21. Peña, D., & Tsay, R. S. (2021). Statistical Learning for Big Dependent Data. Wiley.
22. P. H. Li, H. H. Kwon, L. Sun, U. Lall, and J. J. Kao, “A modified support vector machine based prediction model on streamflow at the Shihmen reservoir, Taiwan,” International Journal of Climatology, vol. 30, no. 8, pp. 1256–1268, 2010.
23. P. Paparoditis, D.N. Politis The asymptotic size and power of the augmented Dickey–Fuller test for a unit root Econ. Rev., 37 (2018), pp. 955-973, 10.1080/00927872.2016.1178887.
24. Power, B., & Gasser, K. (2012). Forecasting Future Unemployment Rates. ECON 452 First Report.
25. Prado, J., & Marzal, M. A. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. Libri, 63(2), 123–134. <https://doi.org/10.1515/libri-2013-0010>
26. Santos, C. H., Lima, R. D. C., Leal, F., Queiroz, J. A. de, Balestrassi, P. P., & Montevechi, J. A. B. (2020). A decision support tool for operational planning: a Digital Twin using simulation and forecasting methods. Production, 30, Article ID e20200018.
27. Smalter, H. A. (2018). Machine Learning Approaches to Macroeconomic Forecasting. Economic Review, Federal Reserve Bank of Kansas. Retrieved from <https://www.kansascityfed.org/~/media/files/publicat/econrev/econrevarchive/2018/4q18smalterhall.pdf>
28. S. Gu, B. Kelly, and D. Xiu, “Empirical asset pricing via machine learning,” Review of Financial Studies, vol. 33, no. 5, pp. 2223–2273, 2020.
29. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., 1997.
30. S. Sharma, H.K. Soni An unemployment prediction rate for Indian youth through time series forecasting Mach. Intel. Smart Syst. (2021), pp. 315-335, [10.1007/978-981-33-4893-6\_29](https://doi.org/10.1007/978-981-33-4893-6_29).
31. S. Tuo, T. Chen, H. He et al., “A regional industrial economic forecasting model based on a deep convolutional neural network and big data,” Sustainability, vol. 13, no. 22, Article ID 12789, 2021.
32. Two Sigma. 2016. Predicting New York City Unemployment Rate Using Taxi Cab Data.][ Xu, W., Z. Li, and Q. Chen. 2013. “Forecasting the Unemployment Rate by Neural Networks Using Search Engine Query Data.” Service Oriented Computing Applications 7 (1): 33–42. doi:10.1007/s11761-012-0122-2.
33. V. Gómez, A. Maravall, D. Peña Missing observations in ARIMA models: skipping approach versus additive outlier approach J. Econ., 88 (1999), pp. 341-363, 10.1016/S0304-4076(98)00036-0.
34. Yao, Y., & Tan, C. (2018). Application of artificial neural network model in time series forecasting. Journal of Physics: Conference Series, 1069(1), 012027. doi: 10.1088/1742-6596/1069/1/012027.
35. Y. X. Su, H. Xu, and L. J. Yan, “Support vector machine-based open crop model (SBOCM): case of rice production in China,” Saudi Journal of Biological Sciences, vol. 24, no. 3, pp. 537–547, 2017.
36. Xu, W., Li, Z., Cheng, C., & Zheng, T. (2013). Data mining for unemployment rate prediction using search engine query data (Vol. 7, pp. 33–42). Berlin: Springer.
37. Z. Zhengwan, Z. Chunjiong, L. I. Hongbing, and X. I. E. Tao, “Multipath transmission selection algorithm based on immune connectivity model,” Journal of Computer Applications, vol. 40, no. 12, article 3571, 2020.
38. Meeting Points
39. Research questions are very general. Problems should be interesting