Unemployment Insights: A Machine Learning Approach

Cristina Boboc
The Bucharest University of Economic Studies; Institute of National Economy – Romanian Academy

Alexandra Roberta Rosca
The Bucharest University of Economic Studies

Ana-Maria Ciuhu (anamaria.ciuhu@insse.ro)
Institute of National Economy – Romanian Academy; Romanian National Institute of Statistics

Valentina Vasile
Institute of National Economy – Romanian Academy

ABSTRACT

This study offers a comprehensive exploration of unemployment, spanning short-term, medium-term and long-term periods, highlighting the necessity for in-depth analysis and timely intervention. The primary goal is to develop machine learning models proficient in identifying the characteristics of individuals undergoing unemployment. Utilizing data from the European Social Survey, the research employs data processing techniques and diverse machine learning algorithms, including Logistic Regression, Random Forest, Ada Boost, LightGBM, and Gradient Boosting. The comparative analysis of unemployment across different durations reveals significant variations in predicting factors. Short-term unemployment is notably influenced by age. For those unemployed for twelve months or more, factors such as perceived equal opportunities and attachment to Europe become influential. In the context of unemployment lasting five years, variables related to happiness and marital status prove crucial, impacting motivation and potentially contributing to extended unemployment. This study marks the initiation of a comprehensive exploration into understanding the traits of unemployed individuals, extending beyond the individual level to incorporate macroeconomic considerations. The application of machine learning algorithms provides valuable insights into addressing this the societal challenge of unemployment.

Keywords: *unemployment, labour force, economic impact, social impact, machine learning*

JEL classification: *E24, J64*

1. INTRODUCTION

Unemployment is a state of imbalance between the demand of the business environment for work and the offer expressed by individuals. It encompasses natural unemployment, driven by changes in labor demand due to developments in the business environment or technological shifts that redefine the quantity and nature of work. Additionally, unemployment can result from crises or institutional reforms. From an individual perspective, unemployment signifies a situation in which the aspirations and expectations of an individual regarding a specific job do not align with the requirements of the business environment. Consequently, an unemployed person is someone who is capable of employment, actively.

The effects of unemployment are multiple. At the national level, it serves as a whistleblower for the education system, signaling potential changes in competences, abilities, specializations, and even shifts in professions. Additionally, it functions as a barometer for assessing the state of the economy, indicating imbalances in development or the escalation of crises. From an individual perspective, voluntary unemployment (as discussed by Jung and Winkelmann in 1993 and Boadway and Cuff in 2016) represents an individual's proactive choice for change, reflecting a pursuit of a job aligned with professional aspirations, career development, material well-being, and overall job satisfaction. Conversely, involuntary unemployment leads to significant adverse effects, both economic (such as the inability to generate income for personal consumption, financial difficulties, loss of stability and autonomy, professional disqualification, and poverty) and social (including discrimination, marginalization, and social tensions).From a cultural perspective, being employed is a shared value among individuals, and when the workplace aligns with individual expectations, the joy derived from work enhances performance and fosters engagement in continuous personal development, including ongoing training and career progression.

In contemporary society, securing employment is crucial for maintaining the expected standard of living. The availability of job opportunities hinges on factors such as possessed skills, educational level, specializations, and the demand within the business environment, varying by field of activity.

From a societal standpoint, unemployment is a constant; there will always be individuals seeking employment, and employers, driven by new entrepreneurial initiatives or business development, will require new employees. Nevertheless, an excessively high level of unemployment can lead to various negative economic consequences. For instance, a decline in income and purchasing power may result in homelessness, causing a reduction in the

demand for goods and services, ultimately leading to economic stagnation. Additionally, heightened unemployment can create social and political pressures, giving rise to tensions and instability within a community.

2. A BRIEF OVERVIEW OF THE UNEMPLOYMENT STATE

According to the Corporate Finance Institute (2023), unemployment can be defined as the phenomenon where individuals who are in perfect working condition and actively seeking a job, unfortunately, fail to find one. Additionally, this category also includes individuals who have a job but not the suitable one.

The unemployment rate measures this phenomenon as share of the unemployed people in the total number of people constituting the labour force. Also, the economic state of a country can be reflected by its unemployment rate. There is a negative correlation between the unemployment rate and the economic situation of a country, so that, an increase in the unemployment rate will highlight an economic decline.

From an economic and behavioral perspective, unemployment among young people pushes them to explore how they can optimize their skills through participation in various training stages (Mroz and Savage 2020). The same experts point out in their paper that unemployment has a negative impact on an individual's income for a minimum of ten years, even if the individual was unemployed for six months or less. Furthermore, they argue that young people are predisposed to accept job offers below the salary they deserve because of the fear of not being employed in the future. Arulamaplam (2001) considers that in the UK, unemployment permanently affects a person's financial component. Moreover, he mention that people who have been unemployed in the past earn 6 percent less when they are rehired and 14 percent less after three years. Kessler et al. (1988), after a community study conducted in agreement with the previous ones, concluded that unemployment has a clinically visible impact on the health of people who experienced this status. They also stated that the main pawn in the degradation of health is financial pressure because, without it, the negative effects on health are halved, and that unemployment also negatively impacts other usually insignificant events. Among the negative effects on health are both physical and mental tensions, including anxiety and depression. Mastekaasa (1996) approaches another perspective, based on a study conducted in Norway, and supports the idea that individuals who are mentally healthy have a lower chance of being dismissed from a job and a higher chance of quickly rehiring, opposite to those who suffer from mental disorders, who present a higher risk

of dismissal. Dooley et al (1996) focused to another approach and demonstrate that, at the aggregate level, there is a strong link between the suicide rate and long-term employment, and, at the individual level, substance dependence and depression are the consequences of long-term unemployment. Moreover, Pohlan (2019) asserts that unemployment also has harmful effects on the quality of life, including life satisfaction, integration into society, access to economic resources, and also on mental health, as this leads to social exclusion, and isolation from society. The author also demonstrated that those who have a partner and also have higher education reduce the harmful effects of job loss. Costescu (2013) underline that, although education is indispensable in the modern world, individuals with higher education face major problems in finding a job suitable for their level and end up working in a lower job compared to their qualifications, which also leads to the degradation of mental health. Other authors (Engel-Yeger B. et al. 2016) focused on analyzing the connection between sensory processing patterns and their relationship with encompassed medical conditions and concluded that there is a strong link between unemployment and depression.

3. MACHINE LEARNING IN UNEMPLOYMENT FORECASTING

Unemployment is an extremely complex and highly significant phenomenon, making its understanding and forecasting it indispensable. To better comprehend how unemployment operates at the individual level, it is necessary to consider the person's entire history, including periods of unemployment or employment, to identify when an individual might face unemployment. Moreover, if a well-constructed model exists, forecasts can be made throughout an individual's life regarding unemployment.

In the literature there is an interest for time series forecasting of unemployment (Katris 2020; Tsung 2022) and for the rate of exiting unemployment, computed by using logistic regression (Kutuk and Guloglu 2019), but it was found that other machine learning algorithms, such as gradient boosting or random forest, perform better.

Other approach (Viljanen and Pahikkala 2020) developed a Markov Chain model to forecast the probability associated with a person exiting unemployment, the likelihood of a person becoming unemployed, and the probability of being unemployed at a specific time. Unemployment status forecasting is revealed by the state probabilities of the Markov Chain. Equilibrium state probabilities incorporate the fact that it can be predicted whether an individual will be unemployed or not. The authors investigated both a statistical model and a machine learning model for transition rates

from different situations, where predictions can be helpful in a future moment or for an entirely new individual. In this work, Markov Chain was used in a statistical context for modeling unemployment, with the authors focusing on unemployment dynamics and the effect of variables, not on the model's capacity. In the model created by the authors, each prediction is based on two sources: the individual's gender, work experience, education level, age, field of study, and labor market history. Thus, after applying the model, they concluded that people aged between 55 and 60 have double the chances of being unemployed compared to their reference value. Moreover, they assert that gender does not influence unemployment and that the fewer years of work experience an individual has, the higher the probability of entering unemployment. The model's result also emphasizes that lack of work experience predicts almost four times more than the reference value, and those with sufficient experience only a quarter. People approaching retirement age and young people have visibly lower chances of exiting unemployment compared to adults, so unemployment increases roughly simultaneously with age. Education and field of study are also representative factors, so people with minimal education are more likely to enter unemployment. The authors (Viljanen and Pahikkala 2020) maintain that the exact prediction of entering unemployment can be extremely difficult, but they express satisfaction with the performance of the machine learning model.

This study provides a comprehensive view of unemployment, considering both short-term and long-term periods, underscores the importance of in-depth examination and timely intervention. In contrast, other studies focus on specific aspects of unemployment, such as forecasting trends or exit rates, employing diverse methodologies. The Markov Chain model by Viljanen and Pahikkala (2020) shares some similarities with the this research in its attention on individual factors and machine learning but differs in its emphasis and conclusions. Together, these studies provide a multifaceted view of unemployment, each contributing unique insights and methodologies to the field.

4. IDENTIFYING THE PROFILE OF UNEMPLOYED INDIVIDUALS USING MACHINE LEARNING ALGORITHMS

The main objective of this study is to develop machine learning models capable of identifying the characteristics of individuals undergoing a period of unemployment. Leveraging data obtained from the European Social Survey, the research will employ data processing techniques and machine learning algorithms to pinpoint the significant variables influencing the likelihood of a person being in either a short-term or long-term unemployment period.

1.1. Data overview

The data utilized in this research originate from the European Social Survey (ESS), an academically driven cross-national survey conducted across Europe since its inception in 2001. The survey employs face-to-face interviews with newly selected cross-sectional samples every two years, capturing the attitudes, beliefs, and behavior patterns of diverse populations in over thirty nations. The initial dataset comprises 51,365 records and 321 variables. A comprehensive analysis of the current dataset will be undertaken, identifying the relevant variables aligned with the paper's objectives. Various data preprocessing steps will be implemented, addressing missing values, encoding categorical variables, and performing other necessary operations. The analysis will be carried out using Python software (Van Rossum & Drake 2009).

Throughout the data processing phase, it was observed that the dataset contained missing values. To address this issue, the interpolate method with the nearest parameter was employed, replacing each missing value with the nearest existing value in the dataset. This replacement was determined by calculating the distance between the missing value and the existing values, selecting the nearest value based on this distance. Furthermore, all duplicates were removed from the database by using the identification number of each observation.

In the process of recoding, categorical variables in the dataset were transformed into dummy variables, where each category was represented by a dummy variable, incorporating also a strategy for selecting relevant categories. The initial step involved identifying the top five categories with the highest frequency of occurrence, which were then transformed into separate dummy variables. Categories not included in the top five were grouped into a new category called „other".

To identify outliers in the dataset, we computed the z-score for each variable. The z-score provides a measure of how far a value diverges from the dataset mean in terms of standard deviations (Pearson 1895). Consequently, values with a z-score exceeding 3 were pinpointed, signifying substantial deviation from the variable's mean. In the case of numeric variables, outlier values were substituted with the mean of the respective variable, while for categorical variables, outlier values were substituted with the mode.

1.2. Main results

In order to categorize unemployed individuals, machine learning models have been devised to predict distinct durations of unemployment:

- Short-term unemployed: Individuals without employment for more than 3 months (variable name: uemp3m)

- Medium-term unemployed: Individuals without employment for more than 12 months (variable name: uemp12m)

- Long-term unemployed: Individuals without employment for more than 5 years (variable name: uemp5y)

Algorithms such as Logistic Regression (Cox 1958), Random Forest (Breiman 2001), Ada Boost (Freund & Schapire 1997), LightGBM (Ke et al. 2017), or Gradient Boosting (Friedman 2001) will be employed to create accurate and robust models. Each of these methods possesses distinct characteristics, differentiating them in terms of functionality and performance (Oladipupo 2010). The database is divided into a training dataset, representing 70% of the initial dataset, and a testing dataset, representing 30% of the initial dataset.

1.2.1. Short-term unemployment

In this analysis, various algorithms were analysed and their performance was evaluated to identify the most efficient algorithm for the classification model. Consequently, LightGBM emerged as the top-performing model, followed by Histogram based Gradient Boosting, achieving an F1-Score of 50% and an accuracy of 77%. Random Forest and Logistic Regression exhibited comparatively weaker performance. Given its superior performance, LightGBM was selected as the reference model.

Machine learning algorithms for Short-Term Unemployment

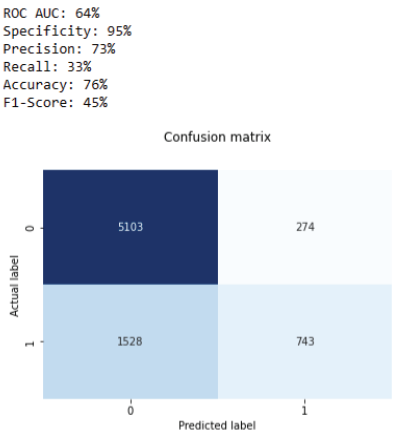
Table 1

Model	ROC-AUC	Specificity	Precision	Recall	Accuracy	F1-Score
LightGBM	68%	91%	66%	45%	78%	54%
Random Forest	63%	92%	64%	35%	75%	45%
Ada Boost	64%	86%	56%	43%	73%	48%
Gradient Boost	65%	89%	60%	41%	74%	48%
Hist Gradient Boosting	66%	92%	69%	40%	77%	50%
Logistic Regression	63%	93%	66%	34%	75%	45%
Support Vector Machines	64%	90%	61%	39%	74%	47%

Source: Created by the authors in Jupyter Notebook.

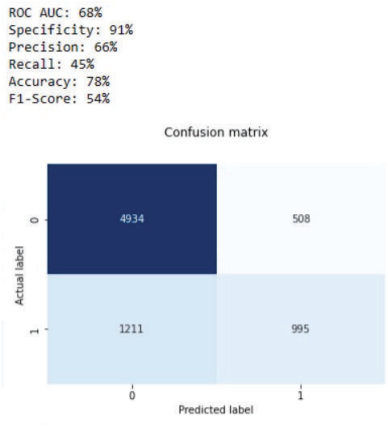
Confusion Matrix using the LightGBM algorithm before applying SMOTE method, for short-term unemployment

Figure 1



. Confusion Matrix using the LightGBM algorithm after applying SMOTE method, for short-term unemployment

Figure 2

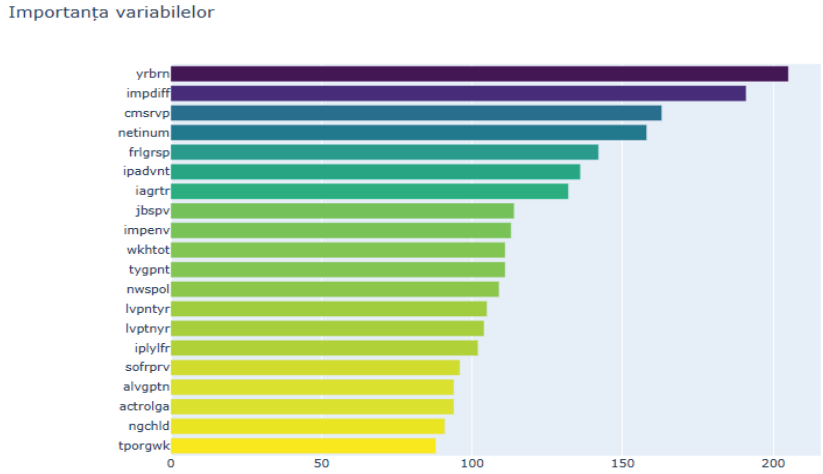


Source: Created by the authors in Jupyter Notebook.

Before implementing hyperparameter optimization techniques, the model successfully classified 76% of the respondents accurately. Considering the imbalanced nature of the variable, the Synthetic Minority Oversampling Technique (SMOTE) method was employed for adjustment. SMOTE creates synthetic records to balance the dataset, preventing the minority class from adversely impacting the model's performance and avoiding suboptimal results. Consequently, following the application of the SMOTE method and hyperparameter optimization techniques, the model achieved an accuracy of 78%, correctly classifying 995 respondents as unemployed (Recall = 45% from the total number of unemployed individuals).

Classification of Variables by Importance for short-term unemployment

Figure 3



Source: Created by the authors in Jupyter Notebook.

The chart above illustrates the top twenty most crucial variables for this model. Consequently, the paramount variable in this model is „yrbrn," signifying the respondent's age and highlighting age as the foremost characteristic in profiling short-term unemployment. Following closely, the variable „impdiff" claims the second position in importance, encapsulating the notion of whether it is essential for an individual to explore new or different aspects of life. This encompasses desires for personal experiences such as increased travel, dissatisfaction with current employment restrictions, or the exploration of new passions. Similarly, the variable „ipadvnt," indicating whether a respondent seeks an interesting life, is also among the most pivotal variables.

1.2.2. Medium-term unemployment

In this case, the dependent variable used is „uemp12m," where the value is 1 if an individual experienced unemployment for twelve months or more and 0 otherwise. To enhance the model, the Grid Search technique was employed, systematically evaluating all possible combinations of the model's parameters to enhance efficiency. Observable improvements are noted across all performance indicators assessing the model's effectiveness.

Machine Learning Algorithms for Long-Term Unemployment of Twelve Months

Table 2

Model	ROC-AUC	Specificity	Precision	Recall	Accuracy	F1-Score
LightGBM	73%	80%	74%	67%	74%	70%
Random Forest	65%	81%	67%	49%	66%	57%
Ada Boost	63%	72%	62%	53%	63%	57%
Gradient Boost	63%	74%	62%	53%	64%	57%
Hist Gradient Boosting	66%	78%	67%	55%	68%	60%
Logistic Regression	51%	91%	48%	10%	55%	17%
Support Vector Machines	60%	86%	67%	35%	63%	46%

Source: Created by the authors in Jupyter Notebook.

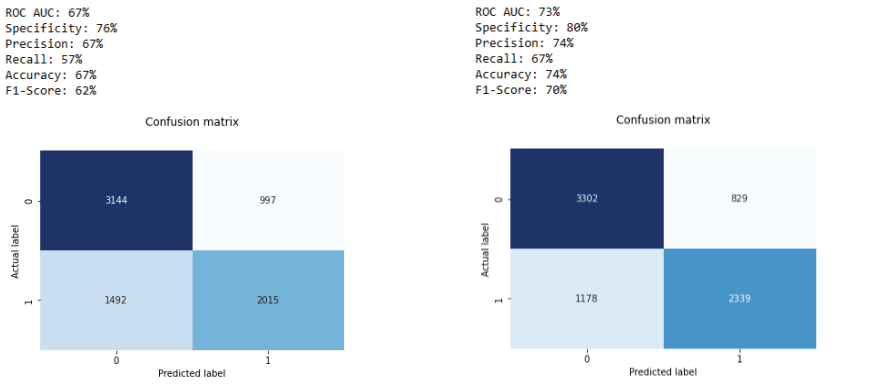
Table 2 presents the outcomes for all the algorithms employed in forecasting the dependent variable. LightGBM demonstrated the most outstanding performance, but there were also other models that proved effective in the context of this analysis. The second-best model for predicting the variable indicating whether an individual will experience unemployment for twelve months or more is, once again, Histogram based Gradient Boosting, achieving an F1-Score of 60% and an accuracy of 68%. In contrast, the Logistic Regression model performed the least effectively, yielding only a 17% F1-Score and 55% accuracy.

Confusion Matrix using the LightGBM algorithm before applying SMOTE method, for long-term unemployment of twelve months

Figure 4

Confusion Matrix using the LightGBM algorithm after applying SMOTE method, for long-term unemployment of twelve months

Figure 5

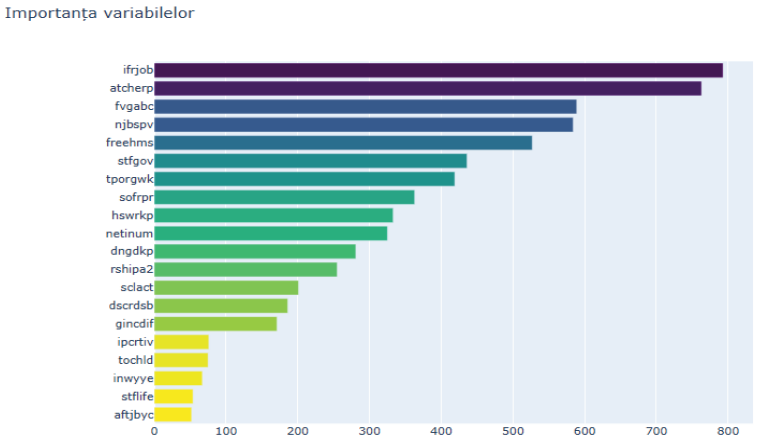


Source: Created by the authors in Jupyter Notebook

Figure 4 depicts the initial model for the variable „uemp12m" without improvements through hyperparameter tuning, and the second figure illustrates the model adjusted through hyperparameter tuning. In the first scenario, the model correctly classified 67% of respondents, while in the latter case, the model significantly improved its performance, showing an 7-percentage-point increase and accurately classifying 74% of respondents. Taking the second model as the benchmark, 2339 out of 3517 individuals experiencing unemployment for twelve months or more, were accurately classified as, placing them in the true positive category (Recall = 67%).

Classification of Variables by Importance for Long-term Unemployment of Twelve Months

Figure 6



Source: Created by the authors in Jupyter Notebook

Figure 6 is a bar chart illustrating the top twenty variables that exert the most influence on the dependent variable in this analysis. Some of these variables describe the situation of individuals before becoming unemployed, while others pertain to their state during medium-term unemployment. Among all the variables integrated into the model, the most pivotal one is the variable reflecting individuals' opinions on equal job opportunities compared to others in their country ("ifrjob"). It suggests that when faced with numerous candidates for a single job, individuals may feel discouraged and insecure if they perceive unequal chances of securing employment compared to others. This could stem from various reasons, such as low self-esteem, particularly in cases of prolonged unemployment, where feelings of insecurity, low self-

esteem, and inferiority may emerge, potentially leading to a reluctance to apply for jobs due to a fear of rejection. The second influential variable in the model is the variable "atcherp," gauging one's attachment to Europe in general. The intensity of attachment to Europe can influence individuals' choices and opportunities, shaping their preferences, mobility, and adaptability in the job market.

Other noteworthy features in this model are the type of organization the respondents have worked in ("tporgwk,"), and the overall life satisfaction ("stflife"). Factors such as organizational stability, growth opportunities, and work culture may influence an individual's decision to stay or leave. A negative perception or dissatisfaction with the organization might motivate individuals to seek alternative employment, contributing to the risk of prolonged unemployment. Individuals that are in medium term unemployment situation are dissatisfied with life in general, whether due to personal or professional reasons. If individuals are dissatisfied with various aspects of their lives, including their work, they may be less inclined to actively seek new employment opportunities. This dissatisfaction could contribute to the risk of unemployment for an extended period of time.

1.2.3. **Long-term unemployment**

In this case, the dependent variable used is unemployment for a period of five years („ uemp5yr „). As in previous case, the Grid Search technique was employed, systematically evaluating all possible combinations of the model's parameters to enhance efficiency. Table 3 presents the outcomes for all the algorithms employed in forecasting the dependent variable „uemp5yr". Similar to the two previously developed models, LightGBM demonstrated the best performance, attaining an accuracy of 77%. Before applying hyperparameter optimization techniques, the model correctly classified 59% of individuals who had experienced a period of unemployment lasting more than 5 years out of a total of 4136 (Recall = 59%). After endeavors to improve its performance, the model achieved an accuracy of 77% and a Recall of 62%.

Machine Learning Algorithms for Long-Term Unemployment of Five Years

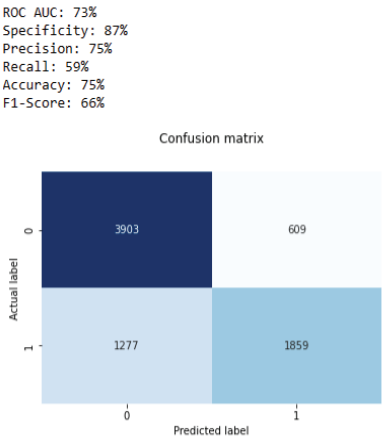
Table 3

Model	ROC-AUC	Specificity	Precision	Recall	Accuracy	F1-Score
LightGBM	75%	87%	77%	62%	77%	69%
Random Forest	62%	82%	62%	43%	66%	51%
Ada Boost	62%	82%	62%	43%	66%	51%
Gradient Boost	62%	81%	62%	44%	66%	51%
Hist Gradient Boosting	68%	86%	71%	50%	71%	58%
Logistic Regression	62%	82%	61%	42%	65%	50%
Support Vector Machines	58%	92%	67%	25%	65%	37%

Source: Created by the authors in Jupyter Notebook

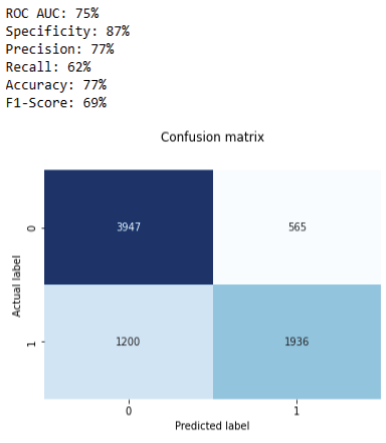
Confusion Matrix using the LightGBM algorithm before applying SMOTE method, for long-term unemployment of five years

Figure 7



Confusion Matrix using the LightGBM algorithm after applying SMOTE method, for long-term unemployment of five years

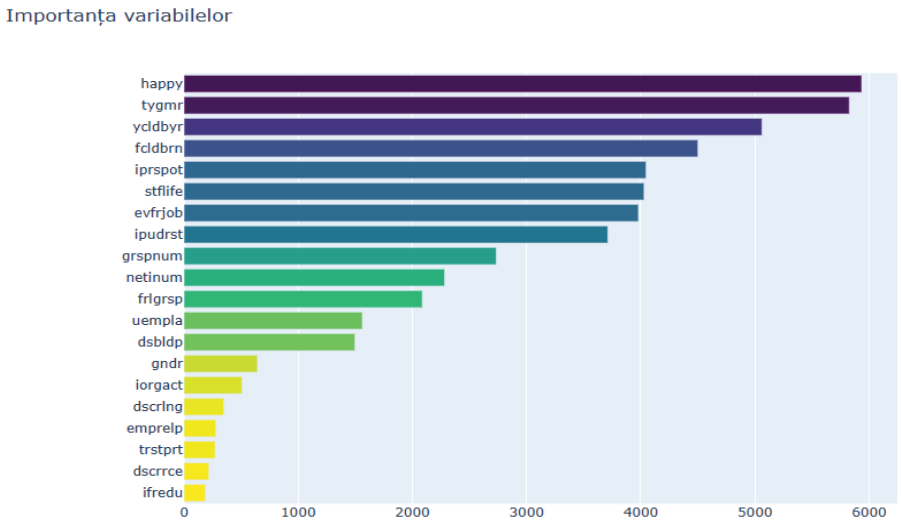
Figure 8



Source: Created by the authors in Jupyter Notebook

Classification of Variables by Importance for long-term unemployment of five years

Figure 9



Source: Created by the authors in Jupyter Notebook

Figure 9 underscores the significance of the top twenty variables on the dependent variable in this analysis. Some of these variables depict the circumstances of individuals before experiencing unemployment, while others relate to their state during medium-term unemployment. The variable „happy," reflecting the respondent's overall level of happiness, holds the top position. Consequently, it appears that the level of happiness is substantially influenced by long-term unemployment status. However, if an individual is unhappy, this discontent may stem from their current status as an unemployed person or could be linked to other aspects of their personal life. Consequently, this dissatisfaction might relegate the job search to a secondary priority, leading to insufficient attention to securing employment.

The second-ranking variable, „tygmr," is particularly intriguing, as it signifies whether the individual got married at a young age. This variable holds significance as it can be viewed as an influential factor impacting an individual's career trajectory. Getting married at a young age might introduce both short and long-term career breaks, contingent on the specific circumstances. In some instances, individuals may voluntarily take breaks in their professional careers to focus on family responsibilities or adjust to married life. On the other hand, the influence could extend to the partner's

desires or expectations, potentially leading to the respondent refraining from active employment. In both scenarios, the decision to marry at a young age appears to have a dual impact, affecting the individual's career trajectory both directly through personal choices and indirectly through external factors such as a partner's preferences. The third influential variable in the model encompasses the year in which the family's youngest child was born („ycldbyr") being closely linked to the previous variable. This factor appears to exert a substantial impact on professional life. The introduction of a child can significantly affect one's career, particularly in the child's early years, where the individual may need to take a break from employment, leading to a temporary halt in their professional trajectory.

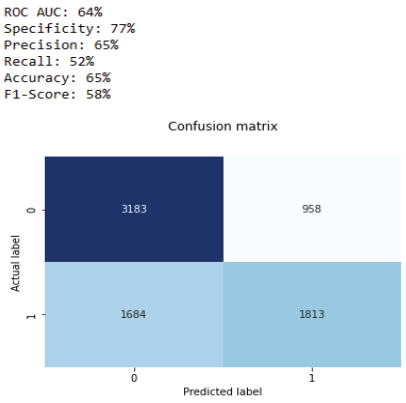
Another crucial variable for the model is the one indicating whether a person believes they have an equal chance of securing a job compared to others in their country (evfrjob). This variable is instrumental in elucidating the influence of personal beliefs and the cyclical effect on unemployment status. The state of being unemployed leads individuals to believe that they do not have equal opportunities as everyone else. Additionally, this belief can discourage individuals from actively seeking specific jobs due to the fear that they might not have an equal chance of success, thereby perpetuating their unemployment status.

1.2.4. **Building a new model for each case with a limited number of variables**

Methods for feature selection were employed to identify the most impactful variables affecting the likelihood of unemployment. Variable selection relies on techniques that pinpoint a subset of pertinent variables for a given dataset. A reduced set of variables can enhance the efficiency and speed of machine learning algorithms. Some algorithms may yield suboptimal predictions when confronted with an extensive array of irrelevant features. In this context, Recursive Feature Elimination (RFE) (Brownlee J. 2020) was utilized as a variable selection algorithm. This involves incorporating a different machine learning algorithm into the core of the method, enveloped by RFE to aid in feature selection. Therefore, this technique serves as a wrapper-type feature selection algorithm that internally incorporates filter-based feature selection.

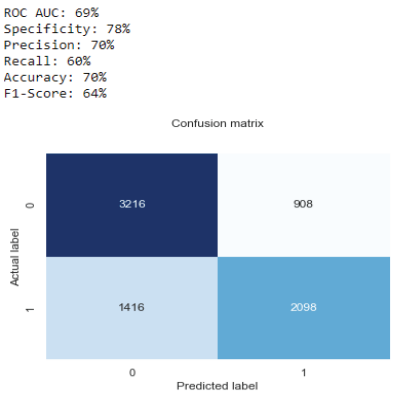
Confusion Matrix using the LightGBM algorithm before applying SMOTE method, for limited number of variables, for long-term unemployment of twelve months

Figure 10



Confusion Matrix using the LightGBM algorithm after applying SMOTE method, for limited number of variables, for long-term unemployment of twelve months

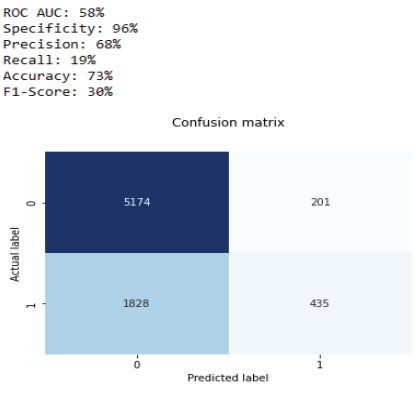
Figure 11



Source: Created by the authors in Jupyter Notebook

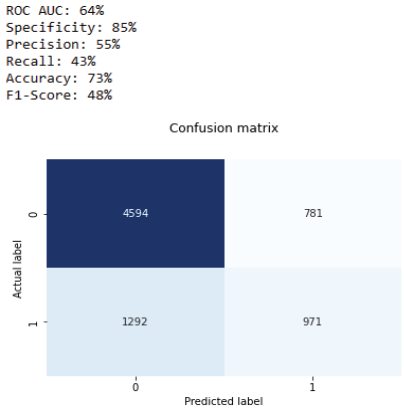
Confusion Matrix using the LightGBM algorithm before applying SMOTE method, for limited number of variables, for short- term unemployment of three months

Figure 12



Confusion Matrix using the LightGBM algorithm after applying SMOTE method, for limited number of variables, for short-term unemployment of three months

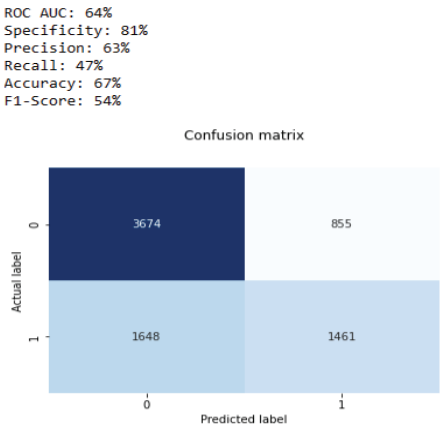
Figure 13



Source: Created by the authors in Jupyter Notebook

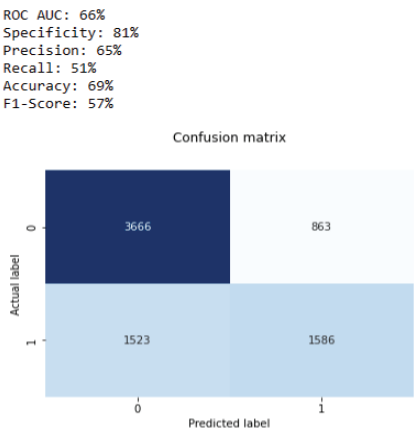
Confusion Matrix using the LightGBM algorithm before applying SMOTE method, for limited number of variables, for long term unemployment of five years

Figure 14



Confusion Matrix using the LightGBM algorithm after applying SMOTE method, for limited number of variables, for long term unemployment of five years

Figure 15



Source: Created by the authors in Jupyter Notebook

The underlying principle of this technique is as follows: initially, the algorithm selects the initial training set, and the dataset's variables gradually decrease until the desired number is achieved. Beginning with fitting the designated machine learning algorithm at the model's core, the features are classified by importance, the least significant features are eliminated, and the model is reconstructed. This iterative process persists until the specified number of features is reached. Through the application of this technique, we identified the first 100 most crucial variables from the three models designed to predict unemployment-related variables. Figures 10-15 depict the confusion matrices for all six variables, three before applying the SMOTE method and three after applying the SMOTE method. The results indicate that by employing RFE, the complexity of the models is effectively reduced, concentrating on the most significant variables. The findings demonstrate that comparable predictive performance can be achieved with a considerably diminished number of variables, as evident in the cases of short-term unemployment (uemp3m) and long-term unemployment (uemp5yr). However, for medium-term unemployment (unemp12m), this reduction may occasionally result in significantly weaker performances. Overall, the approach underscores the importance of meticulous feature selection in constructing robust and efficient

predictive models, emphasizing the potential trade-offs between complexity and performance.

4. CONCLUDING REMARKS

In conclusion, unemployment status varies significantly from individual to individual. A variety of factors influence the individual's state of being unemployed or not, demonstrated both in specialized studies mentioned above and in present work. Also was demonstrate that unemployment has significant effects both economically and socially. Its impact extends into various domains, bringing significant changes to society.

In order to summarize the results of the estimates in this paper, a clarification is required, namely that unemployment is a warning of the health of the economy, but also a permanent risk for individuals present on the labor market, both as people seeking for a job but and as already employed individuals.

The dynamics of society's development and the acceleration of the incorporation of technological progress, to which is added the perception of young people regarding the desired workplace, changes the paradigm of the unemployment risk, generating the need for estimates regarding the dynamics of the unemployment rate.

The influencing factors are diverse, as are the effects of unemployment, with differences depending on the particular characteristics of individuals, on the economic strength of each country.

On the one hand, at the national level, unemployment alters macro balances, adjusts economic growth, effectiveness and efficiency in all fields of activity. Reducing the number of vacant jobs and/or changing the employment structure generates unemployment. The unemployed, through the lack of financial resources available during this period of unemployment, reduces the demand for goods and services and, from here, the business environment is affected, consumption and production are reduced and hence the need for investments for development decreases. We are witnessing a severe economic contraction, up to recession, and, from here, multiple social effects - from reducing the standard of living to increasing the share of people at risk of poverty, from increasing social inequalities to exclusion and fueling social tensions, etc. The two groups of effects of the unemployed status, produce complementary effects on the individual, related to the state of physical health and mental stability, generate stress, anxiety and depression, loss of self-esteem and a decrease in confidence in their abilities, with a significant impact on their mental health. They become socially isolated and their sense

of belonging to a professional group or community is affected, they become a potential risk for other members of the community.

On the other hand, unemployment can be caused by the lack of skills and the asymmetry between the level of education required by the workplace and the offer of the educational system (through graduates and individuals who have upgraded their knowledge and skills). In this case, the employment crisis deepens at the national level, the newly created jobs are at risk of remaining unfilled, high rates of unemployment coexist (including, or especially long-term, among young people) with high rates of vacancies, asymmetry which can be corrected through investments in education and continuous training, therefore, additional costs for individuals in voluntary unemployment, but also for society (responsible for offering training for the involuntary unemployed, in the vast majority of them, being deprived of the financial resources necessary to upgrade the skills or re-qualification).

Therefore, unemployment represents a complex and serious problem, with significant implications both economically and socially. Its effects are felt in all aspects of life and require holistic approaches and appropriate policies to reduce the negative impact and promote inclusion and prosperity in society.

The application of machine learning methods in this study enabled classifications and a more profound understanding of the variables involved in the process. Through these advanced techniques, valuable information was extracted from the dataset. One of the primary advantages of utilizing machine learning in the classification process was the capacity to identify and assess the importance of variables. Employing algorithms such as Random Forest, LightGBM, or Gradient Boosting, measures of feature importance were conducted, aiding in recognizing the influence of variables on the final outcome.

This approach permitted the selection and incorporation into the model of only those variables with a significant impact on classification, eliminating irrelevant or negligible variables. Consequently, a reduction in the dimensionality of the dataset was achieved, leading to the development of more efficient and robust models. Additionally, the use of machine learning methods allowed the discovery of complex relationships and interactions among the variables. The algorithms identified hidden patterns and correlations that were not immediately apparent, providing a deeper and more comprehensive understanding of our data. This, in turn, led to the identification of key factors contributing to unemployment. The importance of each variable within the model was evaluated, and a set of relevant and significant features was obtained.

When comparing unemployment across three different durations—three months, twelve months, and five years—significant variations were evident in the most crucial predicting factors. For individuals unemployed

for a short period, age emerges as a key factor. Young people or those in the earlier stages of their careers may be more susceptible to this situation due to their limited workforce experience or challenges in securing a suitable job. Additionally, the inclination to explore new opportunities or try different things may play a pivotal role in this category.

In contrast, for those facing unemployment for twelve months or more, the most influential factors differ. A noteworthy aspect examined was the respondents' perception of having equal opportunities to secure a job compared to the rest of the population. This sense of inequality can detrimentally impact the job search process and contribute to prolonged unemployment. Moreover, the level of attachment to Europe or involvement in various European projects may hold significance in this category.

In the context of unemployment lasting for five years, the significance of variables related to respondents' happiness level and marital status becomes evident. The level of happiness might be influenced by employment status, and personal satisfaction might impact the motivation to actively seek employment. Furthermore, marrying at an early age may have consequences on occupational status, potentially contributing to an extended period of unemployment.

Indeed, this study marks the commencement of a comprehensive exploration into understanding the traits of unemployed individuals, examining the manifestations, impacts on diverse demographic groups, and broader economic repercussions. This inquiry goes beyond the individual level, extending its scope to incorporate macroeconomic considerations. Addressing the multifaceted issue of unemployment requires a detailed study, prompting timely measures for fostering a balanced economy and enabling individuals, irrespective of age or gender, to lead lives that ensure a minimum standard of living. Furthermore, this paper demonstrates how the application of machine learning algorithms in analyzing the complexities of unemployment, could offer valuable insights to addressing this societal challenge.

References

1. Arulampalam W. 2001. "Is Unemployment Really Scarring? Effects of Unemployment Experiences on Wages." *The Economic Journal*, 111(475): F585–606. JSTOR, http://www.jstor.org/stable/798307. Accessed 31 July 2023.
2. Boadway R., Cuff K. 2016. "Optimal Unemployment Insurance and Redistribution". Queen's Economics Department Working Paper No. 1375, https://ageconsearch.umn.edu/record/274701/files/qed_wp_1375.pdf. Accessed 31 July 2023.
3. Breiman L. 2001. "Random Forests". *Machine Learning*, 45(1): 5-32. https://link.springer.com/article/10.1023/A:1010933404324 Accessed 31 July 2023
4. Brownlee J. 2020. *Recursive Feature Elimination (RFE) for Feature Selection in Python*. https://machinelearningmastery.com/rfe-feature-selection-in-python/. Accessed 31 July 2023.

5. Corporate Finance Institute. 2023. "Unemployment", Published November 26, 2019. Updated May 2, 2023. https://corporatefinanceinstitute.com/resources/economics/unemployment/, Accessed 31 July 2023.
6. Costescu E. M. 2013. „The Education – An Important Factor On Unemployment And Profession," *Annals - Economy Series*, Constantin Brancusi University, Faculty of Economics, vol. 6: 219-226, December.
7. Cox D.R. 1958. „The Regression Analysis of Binary Sequences". *Journal of the Royal Statistical Society*. Series B (Methodological), 20(2): 215-242.
8. Dooley D., Fielding J., & Levi L. (1996). "Health and Unemployment". *Annual Review of Public Health*. 17:449-465. doi: 10.1146/annurev.pu.17.050196.002313.
9. Engel-Yeger B., Muzio C., Rinosi G., Solano P., Geoffroy P. A., Pompili M., Amore M., Serafini G. 2016. "Extreme sensory processing patterns and their relation with clinical conditions among individuals with major affective disorders" *Psychiatry Research.* 236:112-118. doi: 10.1016/j.psychres.2015.12.022
10. European Social Survey 2021. https://www.europeansocialsurvey.org/ Accessed 31 July 2023.
11. Freund, Y., Schapire, R. E. 1997. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". *Journal of Computer and System Sciences*, 55(1):119-139. https://www.sciencedirect.com/science/article/pii/S002200009791504X Accessed 31 July 2023.
12. Friedman, J. H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine". *Annals of Statistics*, 29(5): 1189-1232. https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full Accessed 31 July 2023.
13. Ho T.-W. 2022. „Forecasting Unemployment via Machine Learning: The use of Average Windows Forecasts". Available at SSRN: https://ssrn.com/abstract=3496138 or http://dx.doi.org/10.2139/ssrn.3496138. Accessed 31 July 2023.
14. Kutuk Y., Guloglu B. 2019. „Prediction of Transition Probabilities From Unemployment To Employment For Turkey Via Machine Learning And Econometrics: A Comparative Study". *Journal Of Research In Economics*, 58-75. ISSN: 2636-8307. DOI: 10.24954/JORE.2019.29
15. Jung R. C; Winkelmann R. 1993. "Two aspects of labor mobility: a bivariate Poisson regression approach". Empirical Economics,18 (2):543-556. DOI: https://doi.org/10.1007/BF01176203. Accessed 31 July 2023.
16. Katris C. 2020. "Prediction of Unemployment Rates with Time Series and Machine Learning Techniques". *Computational Economics*. 55: 673–706 (2020). https://doi.org/10.1007/s10614-019-09908-9. Accessed 31 July 2023.
17. Ke G., Meng Qi, Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.Y. 2017. "Light GBM: A Highly Efficient Gradient Boosting Decision Tree". *Advances in Neural Information Processing Systems* 30. (NIP 2017) 3149–3157
18. Kessler R.C., Turner J.B. and House J.S. 1988. "Effects of Unemployment on Health in a Community Survey: Main, Modifying, and Mediating Effects". *Journal of Social Issues*, 44: 69-85. https://doi.org/10.1111/j.1540-4560.1988.tb02092.x, Accessed 31 July 2023.
18. Mastekaasa A. 1996. "Unemployment and Health: Selection Effects", *Journal of Community & Applied Social Psychology*. 6(3): 189-205, August 1996. https://doi.org/10.1002/(SICI)1099-1298(199608)6:3<189::AID-CASP366>3.0.CO;2-O.
20. Mroz T., Savage T. 2006. "The Long-Term Effects of Youth Unemployment". *Journal of Human Resources* 41 (2). https://EconPapers.repec.org/RePEc:uwp:jhriss:v:41:y:2006:i:2:p259-293. Accessed 31 July 2023.
21. Oladipupo T. 2010. "Types of Machine Learning Algorithms". New Advances in Machine Learning, InTech, 1 Feb. 2010. Crossref, doi:10.5772/9385.

22. Pearson K. 1895. “Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. Philosophical Transactions of the Royal Society” *Philosophical Transactions of the Royal Society of London. (A.)* **186:** 343–414 http://doi.org/10.1098/rsta.1895.0010. https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.1895.0010 Accessed 31 July 2023.
23. Pohlan L. 2019. “Unemployment and social exclusion”. *Journal of Economic Behavior & Organization*, 164(C): 273-299. https://www.sciencedirect.com/science/article/pii/S0167268119301969. Accessed 31 July 2023.
24. Van Rossum G. & Drake F.L., 2009. *Python 3 Reference Manual*, Scotts Valley, CA: CreateSpace.
25. Viljanen M., Pahikkala T. 2020. “Predicting Unemployment with Machine Learning Based on Registry Data” In Dalpiaz, F., Zdravkovic, J., Loucopoulos, P. (eds) *Research Challenges in Information Science*. RCIS 2020. Lecture Notes in Business Information Processing, vol 385. Springer, Cham. https://doi.org/10.1007/978-3-030-50316-1_21 Accessed 31 July 2023.