The Economic Research Guardian – Vol. 16(2)2026
Semi-annual Online Journal, www.ecrg.ro
ISSN: 2247-8531, ISSN-L: 2247-8531
Econ Res Guard 16(2): 296-325

Forecasting the Industry Employment Considering the Democracy and Digitalization Indicators by Using Machine Learning Algorithms

Semin Topaloglu-Paksoy
Department of Econometrics, Faculty of Economics and Administrative Sciences, Çukurova University, Adana, Turkiye
Email: spaksoy@cu.edu.tr

Orhan Cengiz*
Department of Accounting and Taxation, Pozanti Vocational School, Çukurova University, Adana, Turkiye
Email: ocengiz@cu.edu.tr

Engin Paksoy
Faculty of Engineering, Department of Computer Engineering, Çukurova University, Adana, Turkiye
Email: engin1236544@gmail.com

Abstract
The current study aims to forecast industry employment considering the economic growth, digitalization, participatory democracy, liberal democracy, human capital, and foreign direct investment (FDI) inflow for a panel group of G7 (Canada, France, Germany, Italy, Japan, the United Kingdom, the United States of America (USA)) and E7 (Brazil, China, India, Indonesia, Mexico, Russia, and Turkiye) separately by using annual data covering the period 2000-2022. Moreover, we perform the information and communication technologies (ICTs) index as a proxy for digitalization following the existing literature. Selected indicators are the fundamental indicators used widely to measure countries' levels of economic development, technological advancement, quality of governance, and social progress, and therefore preferred for forecasting industry employment that indicates the well-being and welfare in the countries. To forecast industry employment, this paper employs five machine learning (ML) algorithms, including Artificial Neural Networks (ANNs), Random Forest (RF), Gradient Boosting (GBoost), Support Vector Machines (SVM), and AdaBoost. The forecasting model is evaluated for the G7 and E7 groups, considering the Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2) values. According to the results, for the case of E7 countries, the RF outperform the others in terms of MSE, RMSE, MAE, and R^2 values. Moreover, the results for the G7 countries case indicate that RF is the best predictor compared to other methods, considering MSE, RMSE, MAE, and R^2 values as evaluation criteria. The result of analysis has indicated that ML algorithms provide reliable results, enabling their use in industry employment forecasting.

Keywords: Employment, Industry employment, Machine learning, Digitalization, Democracy

* Corresponding author.

JEL classification: C44, C61, E27, O11, O14

Received: 10 September 2026; Received in revised form: 17 October 2026; Accepted: 19 October 2026

1. Introduction

Industrialization plays a key role in sustaining long-run economic growth, productivity, and structural transformation, and historical development shows that it is an engine of growth. In particular, the manufacturing sector produces tradable goods, and demand for industrial commodities comes not only from domestic markets but also from global markets, which may offer small countries opportunities to export industrial goods (Amodio et al., 2026). In particular, the manufacturing sector, as a sub-component of industry, plays a key role in economic growth, and its potential for economic growth comes from the capacity to stimulate accumulation, economies of scale, and technological improvement (Haraguchi et al., 2017). However, in the age of globalization and digitalization, the world is becoming more service-led, and the share of industry employment is fluctuating in many countries. Based on World Bank (2026) data, the share of industry employment was around 22% at the beginning of the 1990s and increased to 24% in 2025. In contrast, the share of service sector employment jumped from 36% to 51% during this period.

Today's world rapidly becomes more digital, covering artificial intelligence (AI), robots, and the Internet of Things, and it is an integral part of our lives. Undoubtedly, the potential impacts of digitalization are discussed in terms of its association with the labor force. The main arguments focus on the nexus between digitalization and employment. Despite the large consensus that innovation has a crucial role in promoting productivity, resulting in economic growth and employment, it is not clear whether innovation drives employment in the age of digitalization (Balsmeier & Woerter, 2019). Within the rising digitalization, there is a large set of definitions and concepts of digitalization (Yunxia et al., 2023). For example, Walkme (2024) defines "the process of leveraging digital technologies to transform a business model, creating new revenue streams and value-producing opportunities." Information Technology (IT) Glossary of Gartner (2024) describes digitalization as "the use of digital technologies to change a business model and provide new revenue and value-producing opportunities; it is the process of moving to a digital business." Moreover, the IMF (2018) broadly defines the digital economy as representing the integration of the digital economy with ICTs and traditional sectors. Moreover, digitalization is understood as the comprehensive integration of digital tools across different spheres of life, covering economic, social, and environmental dimensions (Rezgui & Bakari, 2025).

Arising big data, cloud, and AI push companies to adapt to innovation trends and technological transformations (García-Romanos & Martínez-Ros, 2024). According to O'Grady (2024), the size of the digital economy will reach $16.5 trillion in 2028, constituting 17% of global GDP. The biggest digital economies consist of the USA, China, the United Kingdom (UK), Japan, Germany, and South Korea. According to Huawei Technologies (2024) analysis, the global digital supply has significantly increased. For example, global investment in ICTs, telecom-service expenditure, data centers, and renewable energy reached US$4.89, US$2.94, US$0.17, and US$3.12 trillion, respectively.

Computing power has emerged since 1945, and routine tasks have been automated. Hence, indispensably, routine tasks have been replaced with computers, reducing the demand for labor who accomplish routine tasks (Arntz et al., 2019). The relationship between digital transformations and employment has received significant concern from different perspectives in recent years.

Therefore, a vast literature focuses on exploring the association between these two factors. A pioneering study by Frey and Osborne (2017) documented that approximately 47% of USA employment faces the high-risk category. Likewise, Aaron Smith's analysis from the Pew Research Center (2016) shows that 65% of Americans think that robots and computers will displace humans within 50 years. Further, the study conducted by Arntz et al. (2016) for 21 OECD countries reported that an average of 9% of jobs are automatable, indicating that the risk of automation seems lower effect compared to the occupation-based views.

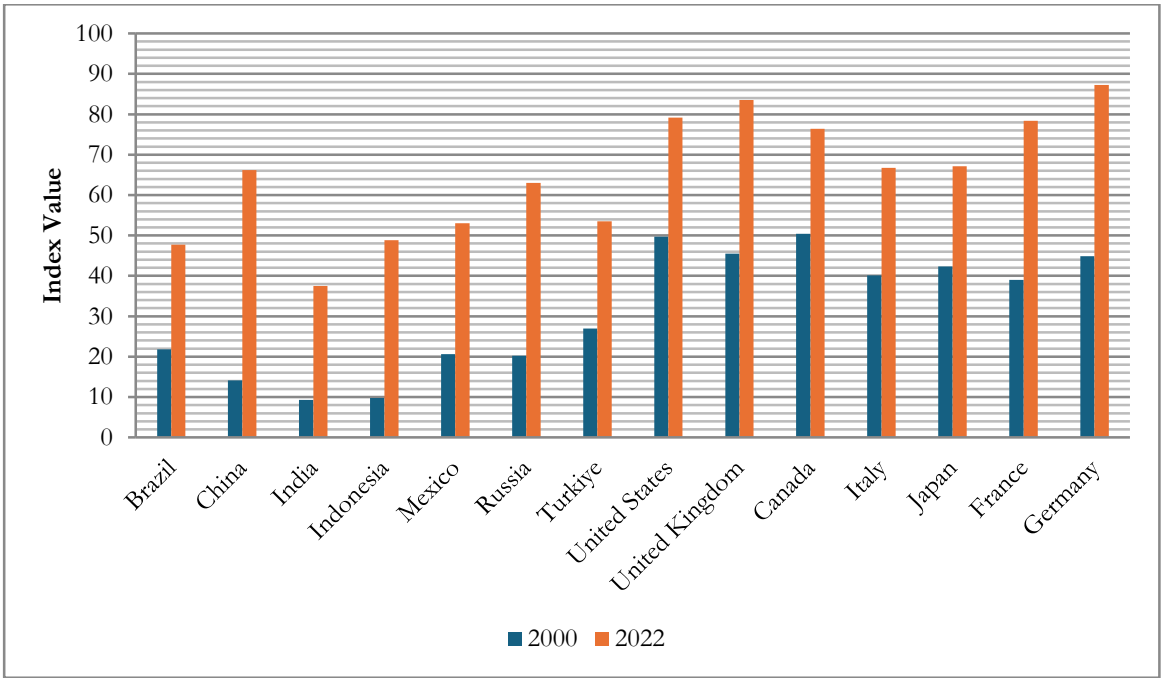
There are different mechanisms explaining the relationship between digitalization and employment. On the first side, in order to keep the competition, firms adopt the use of more digital tools in their operations, resulting in decreasing demand for low-skilled labor forces. In contrast, since they prefer to use digital tools, there is an increasing demand for a high-skilled labor force specializing in AI, robots, and other types of digital instruments (Pandey, 2024). Thus, it needs to be clarified how digitalization shapes our world, including the labor market. Moreover, different pathways are causing opposite effects on employment. From the perspective of displacement effects, it is argued that with an increase in the transformation of the economy toward digitalization, the new technological tools perform the tasks overtaken by people (Richiardi et al., 2025). The digitalization approach certainly causes unemployment, but it may not be correct. Some approaches based on productivity effects ascertain the positive influence of digitalization on the labor market. For instance, digitalization promotes a rise in labor demand for non-automated tasks in automated and non-automated sectors. As discussed by Acemoglu and Restrepo (2018) in detail, the effects of productivity may emerge in different ways. Firstly, as the cost of producing automated tasks declines, the scale of the economy enlarges in non-automated sectors, resulting increase in labor demand for non-automated tasks. Simultaneously, it can mitigate the demand for labor in automated sectors. Second, capital accumulation driven by digitalization may raise demand for capital and labor force. Thirdly, automatization speedly overtakes tasks performed by labor, and at the same time, it triggers the productivity of given machines as a deepening of automation does not cause labor displacement; it increases labor demand (Acemoglu & Restrepo, 2018). Digitalization development and spread also provide new job opportunities and tasks requiring intensive knowledge. Significantly, new forms of employment, such as platform employment, develop along with digitalization, and this promotes labor demand. Further, AI, robots, big data, and cloud computing encourage the automatization of production systems, reducing production costs and increasing productivity. Since productivity rises with automation, the scale of the economy enlarges as well by increasing employment (Yunxia et al., 2023).

On regional and global scales, the G7 (the USA, the UK, Canada, Italy, Japan, France, Germany) and E7 (Brazil, China, India, Indonesia, Mexico, Russia, and Turkiye) are among the striking groups in the world economy. They are leading in GDP, trade, investment, energy, and most other parts of the indicator. The report of PwC (2017) assessed that the world economy is expected to double by 2042, with the E7 countries driving global economic growth. Noticeably, it is forecasted that the E7 countries' share of world GDP will rise from 35% to 50%. Among E7 countries, China will be the largest economy in the world, based on the purchasing power parity (PPP) measurement, followed by India and Indonesia, which will be the fourth-largest economies. More importantly, the E7 countries are expected to grow at an annual rate of approximately 3.5% over the next 34 years. However, during that period, it is expected that G7 countries will grow by around 1.6%.

However, they are also a pioneer in establishing a global digital economy. Based on Huawei Technologies (2024) data, the global digitalization index (GDI) calculated for 77 countries from all over the world, the GDI is 78.8, 61.3, 50.2, 58.8, 62.2, 63.4, and 66.8 for the USA, Canada, Italy, Japan, France, Germany, and the UK, respectively in 2023. Among the G7 countries, all countries are labeled as frontrunners except Italy, which is labeled as an adopter. In addition, the GDI is 44.8, 69.2, 40.3, 33.1, 39.6, and 41.4 for Brazil, China, India, Indonesia, Mexico, and Turkiye,

respectively. Among the E7 countries, China is labeled as a frontrunner, Indonesia is a starter, and the rest of the members are adopters (Huawei Technologies, 2024).

ICTs index is a key component of digitalization. Figure 1 depicts the ICTs trend in both E7 and G7 countries. Generally, the ICTs index increased across both E7 and G7 members.



Source: Authors' preparation based on UNCTAD (2024) data.

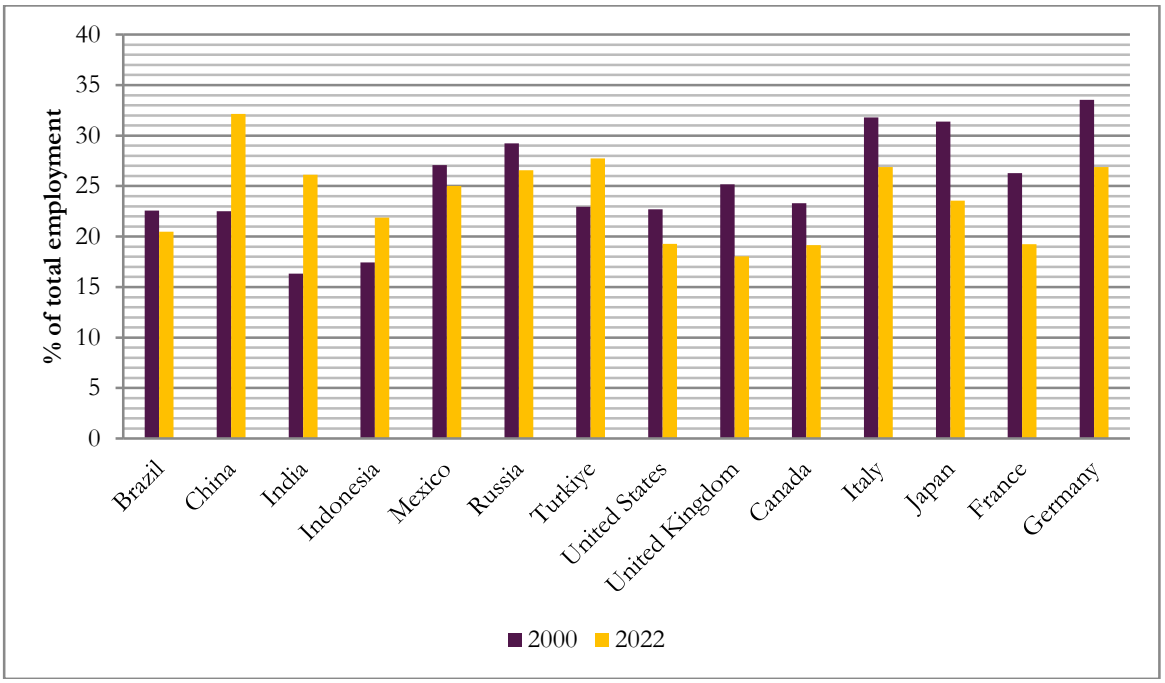
Figure 1 - ICTs Index in E7 and G7 countries

However, the level and magnitude vary across these two groups. For example, in E7, the index value was under 30 in 2000, and it increased, with values ranging from around 38 to 66 in 2022. Hence, ICTs improved remarkably in E7 countries. During this period, the greatest improvement has been in China. Besides, the G7 countries had higher values ranging from 39 to 50 in 2000, and they increased to values ranging from 67 to 82 in 2022. Germany has seen enormous improvement across the G7 countries. Since the two groups are compared, it is clear that although both groups improved significantly, G7 countries continued to have higher ICTs values over the period. More importantly, although there is a difference between the two groups, the gap between them has narrowed progressively. In addition, the comparative advantages of the G7 countries can be attributed to their developed technological infrastructures and earlier investment. Also, the E7 countries struggle to converge with the G7 countries by adopting comprehensive policies. As Aydin et al. (2026) argue, digitalization is not solely related to the technological development process. It also affects the labour markets, investment flows, efficiency, productivity, and institutional arrangements. In the case of the E7 countries, digitalization can promote economic complexity by facilitating access to knowledge networks and innovation, and by transforming the economic structure by shifting from a resource-based economy to a knowledge- and innovation-based one.

One of the fundamental arguments go on the relationship between digitalization and employment. Particularly, the association between digitalization and employment in industry has great attention. One of the crucial factors debating on the nexus between digitalization and industrial employment is the automation of industry expands rapidly. Takki (2024) pointed out that the global automation market transforms manufacturing using developed technology. It has experienced significant growth, and the global industrial market is forecasted to reach approximately 415 billion by 2030.

Therefore, it is a logical attitude and well-founded fear of how digitalization will shape employment in industry sectors. Digitalization transforms the industry in different ways. Specifically, Industry 4.0 has caused a significant transformation that has made the manufacturing process more flexible and endowed with automation, AI, and the Internet of Things (IoT). Because of rapid industry transformation, digitalization has created challenges for the labour force, and employment relationships have become more complex. In particular, new job models require more digital skills, and skilled labour is in short supply. As workplace digitalization spreads across economies increasingly reliant on digital labour capabilities, it affects people in different ways. For example, economic and social challenges may increase. Besides, although digitalization provides employment opportunities for different groups, a lack of digital infrastructure, inadequate education, and regulatory controls can make it difficult for workers to adopt new technologies, making it hard to achieve targeted outcomes for this group (Charles et al., 2022).

In terms of E7 and G7, the employment in industry trend varies in both inter-groups and intra-groups. For example, as depicted in Figure 2, industry employment declined from 22.59%, 27.07%, and 29.24% to 20.49%, 25.03%, and 26.57% in Brazil, Mexico, and Russia, respectively, during 2000-2022.



Source: Authors' preparation based on World Bank (2024) data.

Figure 2 - Employment in industry in E7 and G7 countries

In contrast, it has increased from 22.50%, 16.32%, 17.44%, and 22.95% to 32.15%, 26.12%, 21.87%, and 27.73% in China, India, Indonesia, and Turkiye, respectively, during the same period. It is clear that as of 2022, China has the highest industry employment among E7 members. Moreover, in the G7, employment in the industry declined across all members. For instance, it reduced from 22.70%, 25.17%, 23.29%, 26.27%, 33.53%, 31.79%, and 31.40% to 19.28%, 18.06%, 19.16%, 19.26%, 26.87%, 26.87%, and 23.55% in the USA, the UK, Canada, France, Germany, Italy, and Japan, respectively spanning the period 2000-2022. The comparison of these two groups indicates that the reduction of employment in the industry has been conspicuous in the last two decades in G7 countries. However, in E7 countries, it is changeable, and industry has a key role in economic development in some members of the E7 countries.

In the era of digitalization, industry employment may be influenced by various factors. Among these factors, democracy plays a crucial role in many countries. As digitalization changes the structures of work and working conditions of the labor force, the labor force's rights become under pressure. Specifically, the new forms of work that have emerged from digitalization cause significant labor and social security law challenges. It is widely recognized that the adoption of digital tools reduces the importance of the workplace and increases the platform economies. Since platform economies become more widespread, self-employed people may be challenged with two disadvantageous statuses. Firstly, they may lack the possibility of a minimum wage through a collective agreement, and secondly, they may have less social protection than standard employees (Pärli, 2022).

In the current study, five ML algorithms- Artificial Neural Networks (ANNs), Random Forest (RF), Gradient Boosting (GBoost), Support Vector Machines (SVM), and AdaBoost- are utilized for forecasting industry employment in the E7 and G7 panel groups using annual data from 2000-2022, considering economic growth, digitalization, participatory democracy, liberal democracy, human capital, and FDI inflow. The relevant literature includes studies that focus on employment using traditional statistical or econometric approaches. However, studies that use ML algorithms to forecast employment are relatively new, and the literature remains limited.

The main motivation for choosing the G7 and E7 countries for the sample is due to several reasons. First, Huang et al. (2022) assess the G7 and E7 countries as two leading country groups. The G7 countries are responsible for 58% of global GDP, while the E7 countries are advancing technologically and economically, particularly in human capital growth and patterns. Undoubtedly, these two groups play a substantial role in the global economy. According to UNCTAD (2026) data, as of 2025, the GDP of the E7 countries are more than $30.5 trillion and accounts for 29.47% of world GDP. The GDP of the G7 countries is more than $45.01 trillion and accounts for 43.49% of world GDP. In 2025, merchandise exports were $6.205 trillion for E7 and $7.207 trillion for G7. In addition, the total population of the E7 is more than 3.7 billion, and the G7 is 790 million in 2025. Thus, these the groups have remarkable and different importance in the global economy in terms of the different aspects. The E7 countries as representing one of the most striking developing countries groups, have experienced substantial development in the global economy, while G7 countries represent the advanced countries group and with having developed technological and digital infrastructure. The E7 countries increase their economic power in the global economy, along with digital transformation. However, as a institutional factor the democracy is clearly different in these two groups that G7 countries have a relatively developed democracy compared to E7 countries. In other words, the differences between two groups are not only related to the economic development but also democracy and digitalization infrastructure also are different. Therefore, the forecasting the industry employment by using digitalization and democracy level for heterogenous groups may provide meaningful perspective in the existing literature. Second, as depicted in Figure 1, the E7 countries have improved digitalization over the digitalization for last two decades, and G7 countries have mature digital infrastructures. These two groups offer significant digitalization development at different stages of economic development. More importantly, the E7 countries are heterogeneous in terms of institutional and democratic governance. In contrast, the G7 countries display stable democratic governance. Thirdly, in case of the E7 and G7 countries, the trend for industry employment is multidirectional. For instance, as highlighted in Figure 2, the share of industry employment rose in China, India, Indonesia, and Turkiye, while it decreased in Brazil, Mexico, and Russia during 2000-2022. In contrast, the share of industry employment declined in all G7 members except the United Kingdom and Italy over the same period. Therefore, it is important to examine and compare industry employment by considering digitalization and democracy across these groups at different levels of development.

The expected contributions of the current study to the existing literature are threefold. Firstly, to the best of our knowledge, this is the first study to forecast industry employment for the E7 and G7 groups.

Secondly, in this paper, we attempt to forecast industry employment. However, most studies that employ ML algorithms for forecasting the labour market primarily consider employment or unemployment. Specifically, digitalization threats mainly affect the unskilled labour force who are employed in the industry sector. Hence, it is essential to measure the impact of digitalization on industry employment using ML algorithms for policymakers rather than the entire employment sector. Thirdly, we examine the impact of digitalization, democracy, economic growth, mean years of schooling, and FDI inflow on industry employment. As a proxy for digitalization, the ICTs index is utilized, following the literature. Moreover, as indicators of democracy, we used two different measures: the participatory democracy index and the liberal democracy index.

The rest of the study is categorized as follows: Section 2 summarizes the empirical literature review of studies that utilized statistical methods and ML algorithms. Section 3 explains the features of the ML algorithms used in the study. Section 4 provides the data, material and method. Section 5 documents the ML implication results. Section 6 discusses the findings, and Section 7 concludes with policy recommendations.

2. Empirical literature summary

We categorize the empirical literature into two sub-sections. In the first subsection, the studies related to the determinants of employment are summarized, along with the traditional econometric methods employed. In the second subsection, studies that utilized machine learning tools for forecasting employment were investigated.

2.1. Studies utilized statistical methods to analyze the determinants of employment

Adeem et al. (2019) performed the ARDL bounds approach to investigate the effect of socio-economic factors on employment in Pakistan during the period 1972-2016. The empirical outcomes indicate that economic growth, FDI, human capital, government expenditure, political stability, and capital accumulation have statistically significant and positive influences on employment. Pattanaik and Nayak (2014) researched the macroeconomic determinants of employment intensity in India from 1993–94 to 2009–10. The pooled OLS, fixed effects (FE), and random effects (RE) results indicate that labour supply, economic structure, inflation, and human capital are key factors influencing employment intensity. Maroof et al. (2019) investigated the determinants of industrial employment in South Asian economies over the 1996–2015 period by utilizing panel ARDL and Granger causality estimation techniques. The author's empirical findings suggest that institutional quality, FDI, financial openness, and inflation are driving factors of industrial employment. Anjola et al. (2018) examined the factors that affect industrial output in Nigeria, using data covering the period from 1986 to 2016. The ARDL results indicate that employment and political stability have a mitigating effect on industrial output. However, high inflation, ineffective exchange rates, and power outages are negatively associated with industrial output.

The study conducted by Wajebo (2024) using industry-level panel data for 100 industries in Ethiopia, spanning the period 2002–2020, and employing the two-step GMM estimator, results indicate that real wages and competition negatively affect labour demand. However, local energy supply, FDI, import of raw materials, and capital accumulation positively affect labour demand.

Oueghlissi and Derbali (2024) attempted to examine the association between democracy, corruption, and unemployment in 80 developing countries from 1990 to 2018. Two-step GMM estimator results report that democracy reduces corruption. However, as unemployment rises, the benefit of democracy on corruption is lessened. Ilyas and Khan (2018) investigated the role of political regimes on unemployment in Pakistan covering the period 1977-2012. The authors divided political regimes into democratic and dictatorial regimes and utilized the Johansen co-integration and error correction model to reveal the relationship between the variables. Empirical findings show that democratic regimes are associated with higher unemployment rates than dictatorships. Van Noort (2024) studied whether industrialization is related to democracy in 145 countries between 1845 and 2015 and found that industrialization is significantly associated with democracy.

Based on other main dimensions of the current paper, it is possible to find a wide variety of studies discussing the effect of digitalization on employment. For example, Balsmeier and Woerter (2019) discussed, based on data from Swiss firms, that although an increase in digitalization fosters skilled labour employment, it reduces the employment of low-skilled workers. Gómez-Plana and Latorre (2019) investigated the role of digitalization on employment by considering two types of firms, national and foreign multinationals (MNEs), using data from 1964 to 2014. According to the results, digitalization, represented by ICTs, promotes employment in both national firms and MNEs. Shuttleworth et al. (2022) utilized survey data from German industrial workers for two periods, 2014 and 2020, to examine the relationship between digitalization and employment in the industry. The results concluded that digitalization creates more employment opportunities for skilled workers. Moreover, it is found that large companies tend to utilize more digital tools than small companies do. Zhang et al. (2022) empirically investigated the impact of digitalization on employment polarization in 30 provinces of China by collecting data for 2014 and 2019 using OLS and FE estimators. The empirical outcomes highlighted that digitalization increases employment polarization between skilled and unskilled workers. However, effective public policy may mitigate the polarization between workers. Likewise, Zhao and Said (2023) examined the relationship between digitalization and employment structure in 31 provinces of China from 2013 to 2020. The FE estimator outcomes indicate that, alongside digitalization, employment in the secondary sector decreases while employment in the tertiary sector increases. From a regional dynamics perspective, in the eastern and middle parts of the country, digitalization has a negative impact on primary sector employment but a positive impact on tertiary sector employment. Additionally, as part of the secondary sector, manufacturing and construction are negatively impacted by digitalization. García-Romanos and Martínez-Ros (2024) investigated the nexus between digitalization and employment in Europe, considering the role of firm size and the complementarity of research and development (R&D) investments with digitalization. The author collected data from Innobarometer surveys in 2015 and 2016. They explored an inverted U-shaped association between digitalization and employment. This is more obvious in small enterprises. Furthermore, it is approved that there are positive links between employment and investment in information technology. Reljic et al. (2021) investigated how diffusion of technologies affects employment and labour skills in Germany, France, Spain, Italy, the Netherlands, and the UK, covering the period 2009–2014 via performing a weighted least squares estimator. Also, the authors used industries' consumption of intermediate inputs from digitally intensive sectors and investment in ICTs as indicators of digitalization. The empirical results suggest that employment is positively related to the share of digital tools in total inputs, whereas it is negatively associated with the deepening of ICTs capital.

2.2. Studies utilized ML methods to analyze the determinants of employment

Borgeous works are investigating the determinants of employment in the existing literature by performing traditional statistical methods. However, in recent years, with the rise of ML methods, scholars have attempted to apply these methods to forecasting employment. For example, Güler et al. (2024) employed the ANNs, SVM, and XGBoost as ML methods to forecast Turkiye's unemployment rate using monthly data from 2005 to 2023, incorporating variables such as inflation, exchange rate, and labour force. The author built two models for forecasting. According to the results of the ML algorithms, ANNs are the most suitable algorithm for Model 1, and XGBoost is the most suitable forecasting method for Model II.

Gogas et al. (2022) adopted the decision trees (DT), RF, and SVM algorithms as ML methods and the logit model as a statistical method to forecast unemployment in Euro area with monthly data during the 1998:4-2019:9. The results indicate that RF is the best forecasting algorithm with a 88.5% accuracy rate in full-Turkiye set and 85.4% rate in out-of-sample.

Gajewski et al. (2023) employed the Support Vector Regressor (SVR), RF, and ARIMA models to forecast employment in the Netherlands between January 1, 2014, and December 31, 2021. The findings suggest that RF is suitable for forecasting in large sample sizes, whereas SVR is convenient for small sample sizes. Additionally, SVR and RF are generally superior to ARIMA in forecasting complex time series. Jafor et al. (2023) employed SVM, Logistic Regression (LR), ANNs, RF, XGBoost, and AdaBoost to forecast employee promotions. The results show that ANNs and AdaBoost offer more reliable results rather than other ML algorithms. Specifically, the modified ANNs algorithm outperforms with a 95.30% accuracy rate. Karahan and Çetintaş (2022) applied the ANNs method to forecast the unemployment rate in Turkiye for the period 2005-2018, using the inflation rate, employment rate, exchange rate, current deficit, and industrial production. The authors concluded that the results are reliable and closely aligned with actual data. Davidescu et al. (2021) applied the seasonal model autoregressive integrated moving average (SARIMA), self-exciting threshold autoregressive (SETAR), Holt-Winters, ETS (error, trend, seasonal), and neural network autoregression (NNAR) forecasting model to predicting unemployment in the Romanian economy. The study's data for identification cover the period from January 2000 to December 2017, while the estimation and out-of-sample data cover the period from 2018 to 2020, and the forecasting unemployment rate range is 2021-2022. The results indicate that the Holt-Winters method is superior to others in the sample forecast, considering the RMSE, MAE and mean absolute per cent error (MAPE) measurements. In the out-of-sample evaluation considering RMSE and MAE, NNAR is the most suitable. Finally, considering the MAPE, the SARIMA has the highest accuracy. Mulaudzi and Ajoodha (2020) used the linear regression, elastic net, ridge regression, least absolute shrinkage and selection operator (LASSO), SVR, multi-layer perceptron (MLP), and long short-term memory (LSTM) models to forecast the unemployment in South Africa by using data spanning the January 1970-December 2019. They reported that MLP has the lowest error rate and ridge regression has the highest R-squared value.

Yurtsever (2023) performed the LTSM-Gated recurrent unit (GRU) hybrid model to forecast unemployment in the USA, United Kingdom, France and Italy, using monthly data from January 1983-May 2022. Considering the RMSE, MAPE, and MAE as performance criteria, the hybrid model outperforms the results of LSTM and GRU alone in the USA, United Kingdom, and France. However, in Italy, the GRU model has the best results. Ibrahim et al. (2023) adopted a hybrid model comprising ARIMA, ANNs, SVM, and LSTM to forecast unemployment in Nigeria. The authors consider two different datasets, NBS and Macrotrend. They concluded that the combination of ARIMA-ANNs as a hybrid model outperforms others, considering the MAE,

MAPE, and RMSE evaluation criteria in the NBS dataset. In contrast, the combination of ARIMA-SVM as a hybrid model outperforms others in the Macrotrend dataset. Olalekan et al. (2021) proposed LR, Naïve Bayes, SVM, RF, and DT as ML algorithms to forecast job employment status and found that LR is superior to the others, achieving a 93% accuracy level. Swetha et al. (2023) studied job appeasement and employee corrosion with SVM, RF, DT, LR, and stacking models and concluded that RF and stacking classifiers outperform with a 98% accuracy level. Orozco-Castañeda et al. (2024) employed SVM, neural networks (NN), and autoregressive models to predict monthly occupation and unemployment rates in Colombia, using data from January 10, 2005, to November 30, 2020. They revealed that NN is superior to autoregressive models. More importantly, across autoregressive models and SVR models, NN outperforms others, considering the RMSE and MAE as evaluation criteria. Senthurvelautham and Senanayake (2023) implement the LSTM model to forecast the job market using a dataset of 522,180 job listings in the software industry. They document that this algorithm is successful in predicting job trends at least for the following 12 months with a significant accuracy rate of 95.71%.

3. Data and method

3.1. Data

In the current study, we utilized the five ML algorithms to forecast the industry employment in the E7 and G7 country groups by using annual data over the period 2000-2022. To forecast industry employment, we used the following variables: economic growth, digitalization, participatory democracy, liberal democracy, human capital, and FDI inflow. The measurement, unit, and sources of the variables are reported in Table 1.

Table 1 - Description, measurement and sources of variables

Variable	Symbol	Measurement	Source
Industry employment	EMP	Employment in industry (% of total employment) (modeled ILO estimate)	World Bank (2024)
Economic growth	GROWTH	GDP growth (annual %)	World Bank (2024)
Digitalization	DIG	Information and Communication Technologies (ICTs) index	UNCTAD (2024)
Participatory democracy	PAR_DEM	Participatory democracy index	Our World in Data (2024)
Liberal democracy	LIB_DEM	Liberal democracy index	Our World in Data (2024)
Human capital	SCHOOL	Mean Years of Schooling (years)	World Bank (2024)
FDI inflow	FDI	Foreign direct investment, net inflows (% of GDP)	World Bank (2024)

Source: Authors' compilation.

3.2. Machine Learning Methods for Employment Forecasting

ML methods are an essential part of AI, where their algorithms learn from datasets to forecast better certain outcomes based on patterns that humans struggle to identify. As a new method, ML can model non-linear relationships and identify complex dependencies in time series (Farooq et al., 2025). In the ML field, datasets are typically divided into two main parts: training and testing datasets. These parts serve different purposes in the model development and evaluation process. In the current paper, we use an expanding-window approach. In this method, the training window is progressively extended by including the following year, and each model is tested on the next year. This evaluates model fit as more data becomes available and offers realistic predictions without data leakage (Jyothi et al., 2026).

On the other hand, the testing dataset is used to evaluate the overall performance of the model produced at the training stage. The testing dataset helps predict the model's performance on new data rather than the training data used previously (Uçar et al., 2020). Another noticeable feature of ML methods is that there is no consensus on whether a ML method can always provide better results compared to others. Several hyperparameters used to manage ML during the training stage, such as the size and structure of the dataset employed in the study and the selected algorithms for prediction, play a crucial role.

Table 2 presents the procedure used to select hyperparameters for E7 countries.

Table 2 - Procedure used for hyperparameter selection for E7 countries

Model	Parameter	Value
ANNs	activation	relu
	alpha	0.0001
	learning_rate_init	0.01
	Hidden_Layer_Count	4
	Hidden_Layer_1_Nodes	64
	Hidden_Layer_2_Nodes	64
	Hidden_Layer_3_Nodes	32
	Hidden_Layer_4_Nodes	16
AdaBoost	learning_rate	1.0
	loss	square
	n_estimators	100
GBoost	learning_rate	0.1
	max_depth	2
	n_estimators	200
RF	max_depth	10
	min_samples_leaf	1
	n_estimators	100
SVM	C	10
	epsilon	0.1
	gamma	scale
	kernel	rbf

Source: Authors' compilation.

Table 3 presents the procedure used to select hyperparameters for G7 countries. In the current study, the hyperparameters were determined with random search for ANNs, and grid search for RF, GBoost, AdaBoost, and SVM.

Table 3 - Procedure used for hyperparameter selection for G7 countries

Model	Parameter	Value
ANNs	activation	tanh
	alpha	0.01
	learning_rate_init	0.01
	Hidden_Layer_Count	4
	Hidden_Layer_1_Nodes	32
	Hidden_Layer_2_Nodes	16
	Hidden_Layer_3_Nodes	8
	Hidden_Layer_4_Nodes	4
AdaBoost	learning_rate	1.0
	loss	Square
	n_estimators	200
GBoost	learning_rate	0.1
	max_depth	3
	n_estimators	200
RF	max_depth	10
	min_samples_leaf	1
	n_estimators	100
SVM	C	10
	epsilon	0.01
	gamma	scale
	kernel	rbf

Source: Authors' compilation.

The grid search is one of the conventional methods used in ML for optimizing and tuning the parameters for a given model. In this approach, firstly the grid of all possible hyperparameter combination is created and latter the test and training the model on validation is performed. The optimal set of hyperparameter denotes the set of hyperparameters that provides the best performance on validation (Shams et al., 2024). As assessed by Belete and Huchaiah (2022) the implementation of the grid search is simple. In contrast, in the case of high dimensionality hyperparameters, the number of required assessment increase exponentially and it may loss its effectiveness. Regarding the structure of the current paper, it has a relatively small sample size (2000-2022 and 7 cross-sections) and the grid search is applicable. Besides, in terms of the ANNs, it is known that ANNs has a greater number of hidden layers, number of neurons, activiation function, the random search becomes more suitable to determine the optimal hyperparameters. As Bergstra and Bengio (2012) discussed, random search is more efficient when all hyperparameters are equally important.

In the research area, multiple ML methods are therefore utilized and compared to obtain the most valuable or optimal forecasting (Sterkenburg & Grünwald 2021; Rainio et al., 2024). For this reason, the current paper employs various forecasting algorithms, including ANNs, GBoost, SVM, RF, and AdaBoost. Using different ML algorithms, it is attempted to obtain more powerful and reliable models for forecasting. These artificial forecasting algorithms are explained in the following subsection.

3.2.1. Artificial Neural Networks (ANNs)

The ANNs have been one of the popular computational systems in AI since the 1980s. The main feature of the ANNs is that it emancipates the human brain neural network from information processing and creates a simple model and various networks (Wu & Feng 2018). Using interconnected processing elements, it attempts to imitate the functioning of human brains by determining patterns from input and output data (Güler et al., 2024). Therefore, the ANNs behave like a structure of the human brain. For example, billions of neurons may create a biological neural network in the human brain that is fulfilled by humans, speaking, reading, recognizing, resolving problems (El_Jerjawi & Abu-Naser 2018). According to Güler et al., (2024), the ANNs proved its efficiency in forecasting non-parametric data. More importantly, ANNs are superior to traditional estimation techniques that can be performed under the conditions where these assumptions are fulfilled. In other words, the ANNs reduce the dependence on assumptions. Moreover, it has the advantage of effectively describing non-linear interactions depending on the selection of functions (Güler et al., 2024).

3.2.2. Gradient Boosting (GBoost)

GBoost as a ML model was improved by Friedman (2001). It combines the output of the various individual predictors to create a powerful committee (Biau et al., 2019). The principal operation of the GBoost is to combine the many simple models, known as weak learners, into a strong learner with improved prediction accuracy (Touzani et al., 2018). It solves classification problems by minimizing errors through the introduction of supplementary models (Carmona et al., 2019). This model begins by creating a single leaf and then building regression trees. In that model, the regression tree represents a type of decision tree that is composed of an iterative process, separating the data into small groups (Yoon, 2021). GBoost is a stagewise additive model. In the first iteration, it tries to determine the weak learner (first tree) in order to reduce the training error. In the second iteration, algorithm learns the second tree to reduce the error caused by the first tree. This procedure is repeated till capture the quality model (Sarker, 2021; Nguyen et al., 2021).

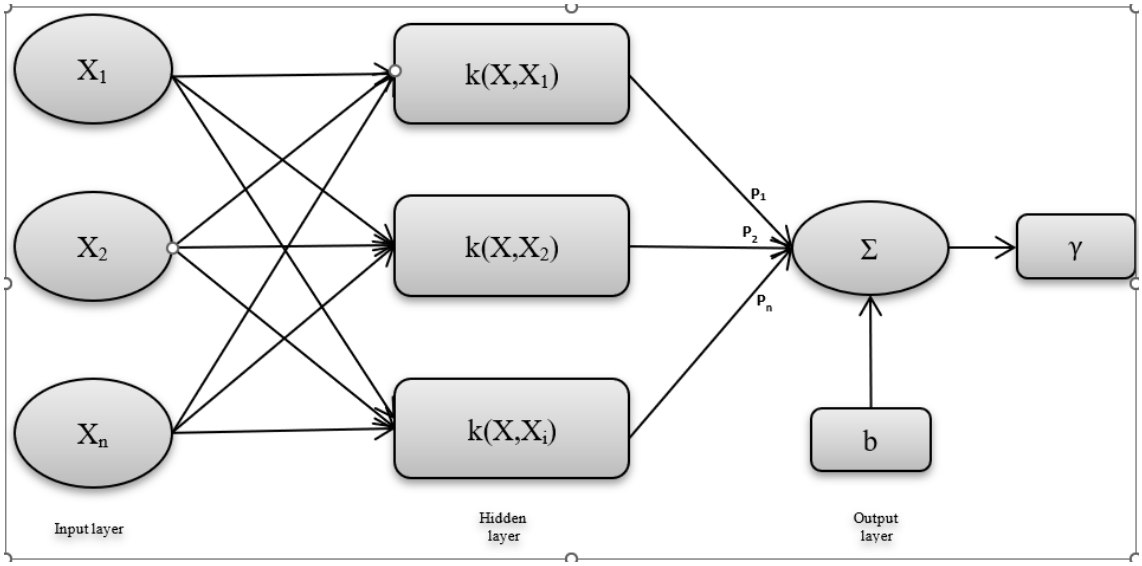
3.2.3. Support Vector Machine (SVM)

Another ML tool is the support vector machine (SVM). This model can be used in regression and classification tasks to classify and regress both linear and non-linear data (Güler et al., 2024). This method was first introduced by Vapnik (1995) and contains two basic categories, support vector classification (SVC) and support vector regression (SVR) (Katris, 2020). Particularly, this model is based on statistical learning theory and structural risk minimization. The primary task of this model is to transform the nonlinear input area into an area with high-dimensional properties, thereby creating a hyperplane (Zendehboudi et al., 2018). Although researchers widely prefer the SVM due to its ease of implementation and allowing for use for both linear and non-linear data, it also has some disadvantages. For example, it can be challenging to interpret model parameters and long model training (Katipoğlu & Sarıgöl, 2023).

As illustrated in Figure 3, the regression model of SVM can be constructed, as shown in Eq. (1), to predict the γ value depending on the input values from X1 to Xn.

$$\gamma = w^T \theta(x) + b \quad (1)$$

where, w is a weight vector, θ is a mapping function, for instance, a kernel function to produce higher dimensional space, X is a one-dimensional input vector, and b is the bias term.



Source: Authors' own drawing based on Zendehboudi et al. (2018).

Figure 3 - Network operation of the SVM

In each iteration, the goal is to obtain a more efficient weight vector and a higher-dimensional hyperplane with stronger separation power (Zendehboudi et al., 2018).

3.2.4. Random Forest (RF)

RF is another ML algorithm that addresses both regression and classification problems. More importantly, it consists of defining different independent trees, and many trees are constructed in parallel on different kernels. This makes the RF easy to obtain as a parallel and application to forecasting (Genuer et al., 2017; Katipoğlu & Sarıgöl, 2023). RF is firstly introduced by Breiman (2001) and this algorithm develops various classification trees. The primary concept behind RF is to combine the outputs of multiple DT to achieve a single result. In order to classify the new factor, it describes a vector into each tree in the forest (Fahlapi et al., 2020). RF performs bootstrap aggregation, which includes creating n trees and capturing the results by averaging the predictions from all estimators. Moreover, it contains several parameters related to the size of the model, the number of estimators, the maximum number of nodes, and the maximum depth of each tree (Gajewski et al., 2023).

3.2.5. AdaBoost

AdaBoost (or Adaptive Boosting) is a boosting algorithm commonly used to refine imbalanced data. It creates a model and distributes weight equally to all data points (Jafor et al., 2023). As an iterative ensemble method, it sets the weights of classifiers and trains the data sample in each stage of iteration (Navlani, 2018). The main idea behind AdaBoost is to adjust the weights in the learners and training process in every iteration, aiming to achieve precise predictions with significant accuracy (Subasi et al., 2022). It combines with weak classifiers to construct a learning algorithm with strong classifiers (Heo & Yang, 2014).

3.2.6. Performance metrics

All ML models, whether their task is linear regression or classification, need a metric to judge performance (Rainio et al., 2024). A metric is based on calculating some sort of distance between predicted and actual value. For robust predictions, the predictions made by data-driven models should not be too far from the actual output. In the current study, popular and recently used metrics are used to measure the performance of the ML models.

The performance of the predicted models was determined according to MAE, MSE, RMSE, and R^2 values. MAE is an indicator that defines only mean errors, ignoring the directions of deviation between actual values and predicted values. MAE values of 0 indicate a possible perfect fit, while a value of ∞ indicates the worst fit. MSE measures the average of the squares of the errors. This indicator, which takes a value from 0 to ∞, shows the deviation between the actual and predicted data. The RMSE values closest to 0 represent the optimum model among the predicted models. Finally, R^2 is a statistical measure that determines the proportion of the variance in the dependent variable that can be explained by the independent variable (Rainio et al., 2024). It provides a measure of how the model replicates well-observed outcomes.

The formulas of the statistical metrics used are shown in Table 4. The notations are as follows: O_i represents the observed (or actual, true) values or the Employment values, P_i represents the Employment values predicted by the models, n represents the number of the sample (or data), $\overline{O}$ represents the mean of the employment values in the data set used.

Table 4 - The statistical metrics employed to assess the performance of the models

Statistical indices	Symbol	Equation
Mean Absolute Error	MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert O_i - P_i\rvert$
Mean Squared Error	MSE	$MSE = \frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2$
Root Mean Square Error	RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2}$
Coefficient of Determination	R^2	$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2}$

Source: Patel et al. (2025).

The particularly noticeable fact that these indicators point to the same model implies that the pattern representing the data set has been well learned by the ML methods used, and therefore, the models produced fit the data at a satisfactory level. Otherwise, if the model has not trained well in the learning stage and its predictions obtained in the testing stage are unreliable.

4. Results

The results of the ML algorithms applied to the panel groups of the E7 and G7 countries are compared, and the results are evaluated based on MSE, RMSE, MAE, and R^2 values.

4.1. Results of ML algorithms for the E7 countries

Our results begin with the provision of descriptive statistics and a correlation matrix for both country groups.

Table 5 highlights the descriptive statistics for the E7 countries. According to the results, the average value is 24.3, 4.53, 38.4, 0.33, 0.38, 7.769, and 2.18 for EMP, GROWTH, DIG, PAR_DEM, LIB_DEM, SCHOOL, and FDI, respectively. Besides, DIG has the highest standard deviation, and PAR_DEM has the lowest one.

Table 5 - Descriptive Statistics of E7

	EMP	GROWTH	DIG	PAR_DEM	LIB_DEM	SCHOOL	FDI
Mean	24.325	4.533	38.437	0.336	0.380	7.769	2.188
Median	24.129	5.019	39.500	0.393	0.434	7.412	2.092
Maximum	32.152	14.230	69.600	0.656	0.790	12.412	5.033
Minimum	16.319	-8.624	9.300	0.038	0.043	4.144	-2.757
Std. Dev.	3.561	3.917	15.344	0.175	0.226	2.181	1.217
Skewness	-0.065	-0.755	-0.131	-0.216	-0.076	0.928	-0.568
Kurtosis	2.475	4.113	2.055	2.158	1.864	3.211	4.674
Jarque-Bera	1.958	23.633	6.444	6.002	8.807	23.445	27.475
Prob.	0.375	0.000	0.039	0.049	0.012	0.000	0.000
Sum	3916.359	729.940	6188.400	54.246	61.189	1250.872	352.371
Sum Sq. Dev.	2029.873	2455.059	37672.36	4.909	8.174	761.332	237.359
Observations	161	161	161	161	161	161	161

Source: Authors' compilation.

Following the provision of descriptive statistics for the E7 countries, the correlation matrix is reported in Table 6, which indicates a significant negative correlation between participatory democracy (PAR_DEM) and industry employment (EMP), as well as between liberal democracy (LIB_DEM) and industry employment (EMP).

Table 6 - Correlation matrix of E7

	EMP	GROWTH	DIG	PAR_DEM	LIB_DEM	SCHOOL	FDI
EMP	1.000						
GROWTH	0.101	1.000					
DIG	0.554	-0.322	1.000				
PAR_DEM	-0.658	-0.278	-0.298	1.000			
LIB_DEM	-0.686	-0.232	-0.357	0.985	1.000		
SCHOOL	0.511	-0.259	0.606	-0.397	-0.499	1.000	
FDI	0.126	0.048	0.127	0.019	0.040	-0.001	1.000

Source: Authors' compilation.

Moreover, there is a positive association between economic growth (GROWTH) and industry employment (EMP), digitalization (DIG) and industry employment (EMP), human capital (SCHOOL) and industry employment (EMP), and foreign direct investment (FDI) inflow and

industry employment (EMP). Furthermore, the highest correlation exists between liberal democracy (LIB_DEM) and industry employment (EMP), with a value of -0.68, while the lowest correlation exists between economic growth (GROWTH) and industry employment (EMP), with a value of 0.10.

Table 7 documents the descriptive statistics for the G7 countries. According to the results, the average values are 24.03, 1.37, 65.13, 0.62, 0.79, 10701.02, and 2.18 for EMP, GROWTH, DIG, PAR_DEM, LIB_DEM, SCHOOL, and FDI, respectively.

Table 7 - Descriptive Statistics of G7

	EMP	GROWTH	DIG	PAR_DEM	LIB_DEM	SCHOOL	FDI
Mean	24.033	1.377	65.136	0.620	0.794	12.311	2.180
Median	23.294	1.792	66.200	0.632	0.792	12.765	1.726
Maximum	33.531	8.674	87.300	0.690	0.872	14.256	12.731
Minimum	18.056	-10.359	39.000	0.521	0.728	8.755	-0.898
Std. Dev.	4.124	2.614	9.859	0.043	0.037	1.443	2.204
Skewness	0.361	-1.502	-0.502	-0.513	0.432	-0.769	2.328
Kurtosis	1.967	7.732	2.810	2.154	2.286	2.450	9.862
Jarque-Bera	10.656	210.779	7.005	11.882	8.425	17.780	461.390
Prob.	0.004	0.000	0.030	0.002	0.014	0.000	0.000
Sum	3869.358	221.742	10486.90	99.914	127.864	1982.15	351.048
Sum Sq. Dev.	2721.233	1094.074	15552.87	0.302	0.222	333.4663	777.388
Observations	161	161	161	161	161	161	161

Source: Authors' compilation.

Likewise, DIG has the highest standard deviation in the results of the G7 countries. However, the lowest standard deviation belongs to the LIB_DEM.

Following the provision of descriptive statistics for the G7 countries, the correlation matrix is reported in Table 8.

Table 8 - Correlation matrix of G7

	EMP	GROWTH	DIG	PAR_DEM	LIB_DEM	SCHOOL	FDI
EMP	1.000						
GROWTH	-0.102	1.000					
DIG	-0.509	-0.014	1.000				
PAR_DEM	0.195	-0.057	-0.095	1.000			
LIB_DEM	0.140	-0.012	0.092	0.620	1.000		
SCHOOL	-0.356	0.076	0.555	-0.404	0.200	1.000	
FDI	-0.126	0.176	-0.104	0.068	0.140	0.070	1.000

Source: Authors' compilation.

The results suggest a different correlation against the results of the E7 countries. For example, there is a negative correlation between economic growth (GROWTH) and industry employment (EMP), digitalization (DIG) and industry employment (EMP), human capital (SCHOOL) and industry employment (EMP), and foreign direct investment (FDI) inflow and industry employment (EMP). In contrast, there is a positive correlation between participatory democracy (PAR_DEM) and industry employment (EMP), as well as between liberal democracy (LIB_DEM) and industry employment (EMP). Furthermore, the highest correlation exists between digitalization (DIG) and industry employment (EMP), with a value of -0.51, while the lowest correlation exists between economic growth (GROWTH) and industry employment (EMP), with a value of -0.10.

Moreover, the current research employed box plots and violin plots to visually represent the distribution of numerical data obtained by ML methods. In this part, introduced tables and graphs that help us gain a better understanding of data and that provide visual support for improved decision making. In describing numerical data, a graph of the five-number summary called a box and whiskers plot is used. The five-number summary refers to the five descriptive measures: minimum, first quartile, median, third quartile and maximum value of the data series with the following order;

$$\text{minimum} < Q1 < \text{median} < Q3 < \text{maximum} \quad (2)$$

The first quartile Q1 separetes the smallest 25% of data from remaining data while the third quartile Q3, (or 75th percentile) separates approximately the smallest 75% of the data from the remaining largest 25% of the data. If median of the data is equal to the second quartile (Q2 = 0.50), distribution of the data is normal. If a distribution is skewed, then the median will not be in the middle of the box, and instead off to the side. That type of the graph is very compact in the summarization of data and also in comparing the groups through the box and whiskers markings' positions. The small circles in the graph represent the outliers in the data.

Table 9 shows the training and test periods for E7 and G7 countries using an expanding window. The first training set covers 2000-2012 (13 years, 91 observations), and 2013 is held out as the test period. At each step, one year is included in the training set while keeping the starting period fixed.

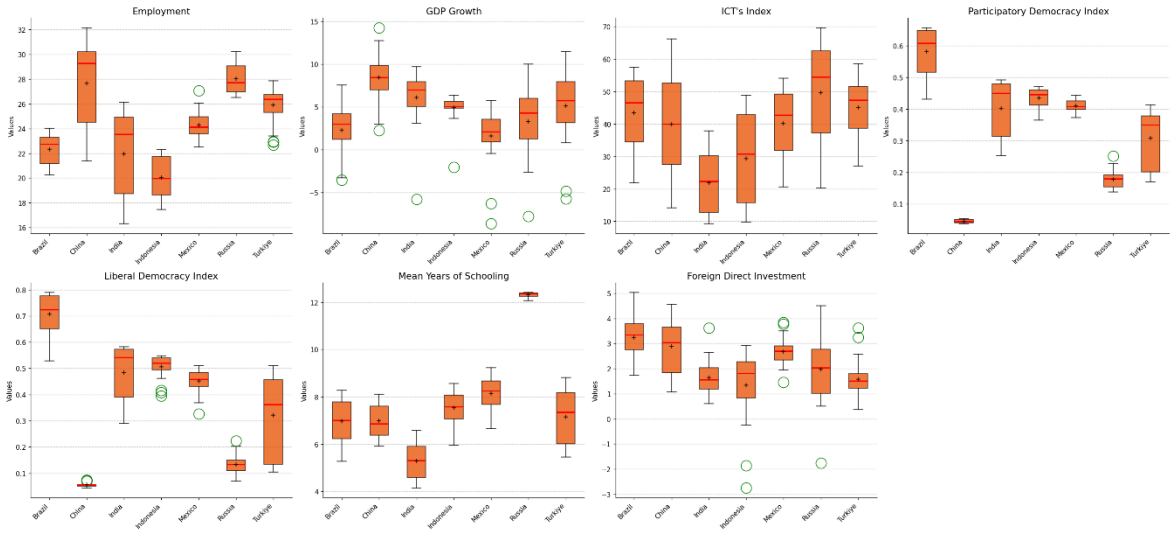
Table 9 - Training and test periods of the model for E7 and G7 countries

Window ID	Train Start Year	Train End Year	Test Year	Train Year Count	Train Size	Test Size
1	2000	2012	2013	13	91	7
2	2000	2013	2014	14	98	7
3	2000	2014	2015	15	105	7
4	2000	2015	2016	16	112	7
5	2000	2016	2017	17	119	7
6	2000	2017	2018	18	126	7
7	2000	2018	2019	19	133	7
8	2000	2019	2020	20	140	7
9	2000	2020	2021	21	147	7
10	2000	2021	2022	22	154	7

Source: Authors' compilation.

Hence, the training set progressively increased from 91 to 154. Thus, we used 10 windows, each with a test set of 7 observations. More importantly, to prevent data leakage, the training data preceded the test period.

Figure 4 shows the box plots of indicators for E7 countries, where China has the highest mean, quartiles, and maximum values for the EMP and GDP.

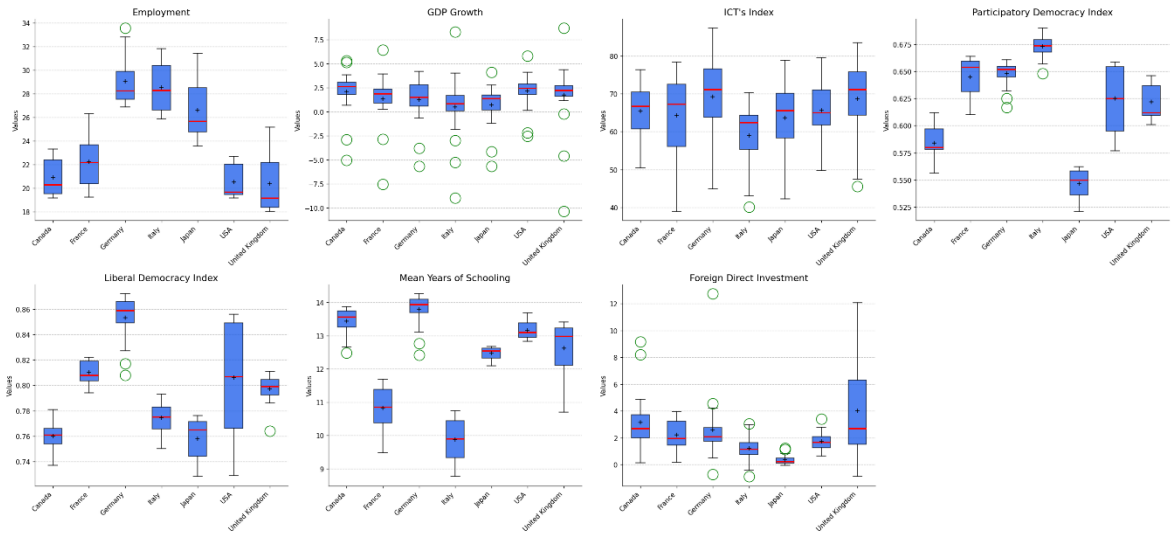


Source: Authors' compilation.

Figure 4 - Box plots of indicators for E7 Countries

In contrast, it has the lowest values in the liberal democracy and participatory democracy indices among the E7 countries. Moreover, Brazil is the leading country in terms of liberal democracy, participatory democracy, and FDI values. In addition, Russia has the highest value in terms of the ICT's index and mean years of schooling.

In Figure 5, it is depicted that in terms of industry employment, although Germany has the highest maximum value, Italy has the highest median and quartile three value among the G7 countries.



Source: Authors' compilation.

Figure 5 - Box plots of indicators for G7 Countries

Moreover, the G7 members are heterogeneous in terms of the participatory democracy index. Italy has the highest index value, while Japan has the lowest. Germany is the leading country in terms of having the highest growth rate and a liberal democracy index. Additionally, Germany has the highest human capital value, as represented by the school variable, whereas Italy has the lowest among the G7 countries.

4.2. Results for the Comparison of ML methods

The current investigation assessed the performance of the models using various metrics, including MSE, RMSE, MAE, and R^2. As mentioned above, in the evaluation criteria, the values of the MSE, RMSE, and MAE should be lower, and R^2 should be high.

Table 10 documents the ML algorithm results for E7 countries, where RF achieves the best performance in terms of the lowest values for the MSE, RMSE, MAE, MAPE (%), RE (%), RRMSE (%), and RRSE (%) with values of 1.317, 1.147, 0.892, 3.632, 3.529, 4.538, and 37.439, respectively, and the highest value for the R^2, at 0.859.

Table 10 - Comparison of employment prediction results for E7 countries

Models	MSE	RMSE	MAE	MAPE (%)	RE (%)	RRMSE (%)	RRSE (%)	R^2
ANNs	1.7496	1.3227	1.0924	4.3529	4.3200	5.2306	43.1453	0.8138
AdaBoost	1.8420	1.3572	1.0953	4.4427	4.3314	5.3669	44.2697	0.8040
GBoost	1.5743	1.2547	0.9626	3.9256	3.8068	4.9616	40.9268	0.8324
RF	1.3175	1.1478	0.8925	3.6321	3.5295	4.5389	37.4399	0.8598
SVM	1.7488	1.3224	1.0049	4.0942	3.9735	5.2293	43.1345	0.8139

Note: MAPE (%), RE (%), RRMSE (%), and RRSE (%) denote Mean Absolute Percentage Error, Relative Error, Relative Root Mean Squared Error, and Root Relative Squared Error, respectively.
Source: Authors' compilation.

It is followed by GBoost, which yields values of 1.574, 1.254, 0.962, 3.925, 3.806, 4.961, and 40.926 for the MSE, RMSE, MAE, MAPE (%), RE (%), RRMSE (%), and RRSE (%), respectively, and 0.832 for R^2.

Table 11 provides the ML algorithm results for G7 countries. Likewise, the RF achieves the best performance in terms of the lowest values for the MSE, RMSE, MAE, MAPE (%), RE (%), RRMSE (%), and RRSE (%), with values of 2.150, 1.466, 1.024, 4.725, 4.593, 6.574, and 42.093, respectively, and the highest value for the R^2, at 0.822.

Table 11 - Comparison of employment prediction results for G7 countries

Models	MSE	RMSE	MAE	MAPE (%)	RE (%)	RRMSE (%)	RRSE (%)	R^2
ANNs	16.2875	4.0357	3.6055	17.5353	16.1636	18.0924	115.8377	-0.3418
AdaBoost	2.8456	1.6869	1.3324	6.4190	5.9735	7.5624	48.4190	0.7655
GBoost	2.4726	1.5724	1.0165	4.6129	4.5572	7.0493	45.1338	0.7962
RF	2.1507	1.4665	1.0246	4.7256	4.5934	6.5745	42.0939	0.8228
SVM	6.2606	2.5021	1.4952	7.3692	6.7031	11.2171	71.8178	0.4842

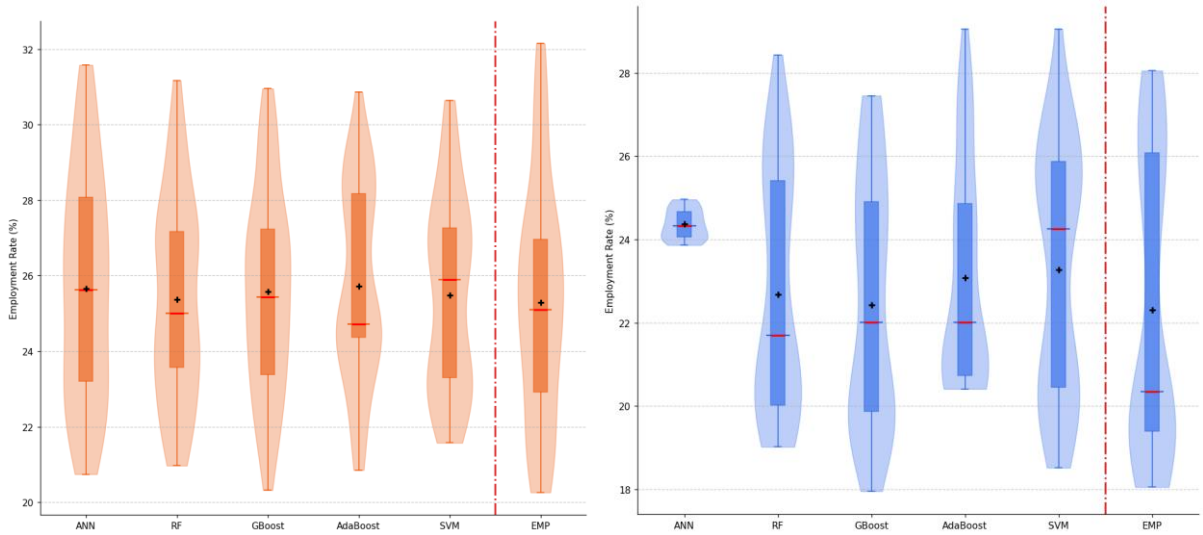
Source: Authors' compilation.

It is followed by GBoost, which yields values of 2.472, 1.572, 1.016, 4.612, 4.557, 7.049, and 45.133 for MSE, RMSE, MAE, MAPE (%), RE (%), RRMSE (%), and RRSE (%), respectively, and 0.796 for R^2.

A violin plot is another statistical graphic and similar to a box and whisker plot with the addition of a rotated kernel density plot on each side. In Figure 6, the violin plot shows the full distribution

of the data in addition to the summary statistics as mean/median and quartile ranges that the box and whisker plot shows. Therefore, the violin plot is also preferred in comparing the outcomes of ML models.

In Figure 6, the violin plots for E7 and G7 countries indicates that the violin plot of the AdaBoost results are very similar to the violin plot of actual Employment data, separated by a dash line within same graph.



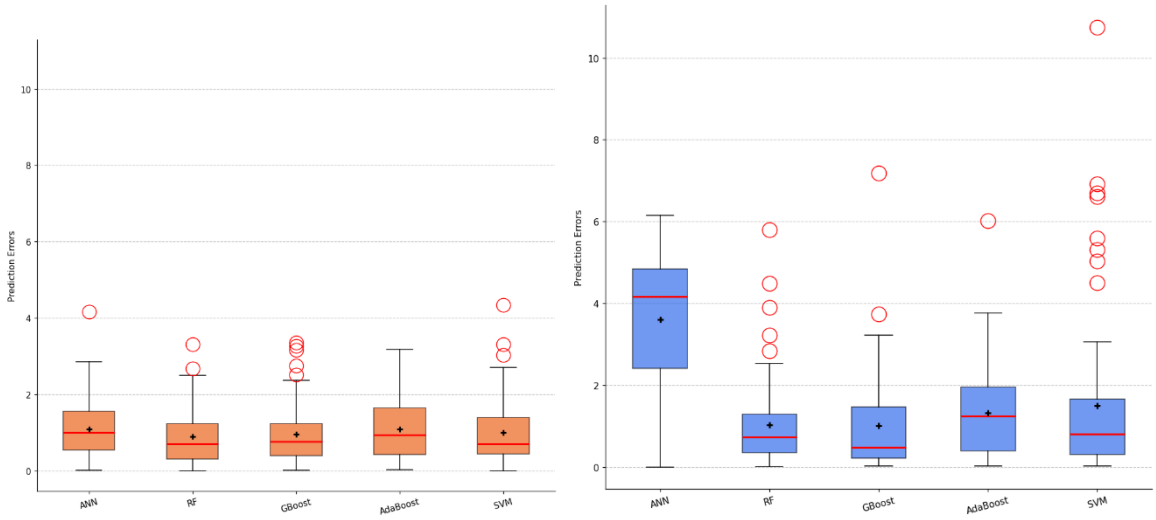
Panel A: E7 Countries Panel B: G7 Countries

Source: Authors' compilation.

Figure 6 - Violin plots of outcomes for E7 and G7 countries

This high-level similarity implies that AdaBoost produced the best forecasts for the Employment data. It is seen that the violin drawings of GBoost and ANNs immediately follow the AdaBoost model and are quite successful.

When ML methods are compared in terms of prediction errors in Figure 7, RF stands out as the model with the smallest prediction error values in both E7 and G7 country groups. It is also observed that the average prediction error of RF is lower than the averages of the other models in the country groups.



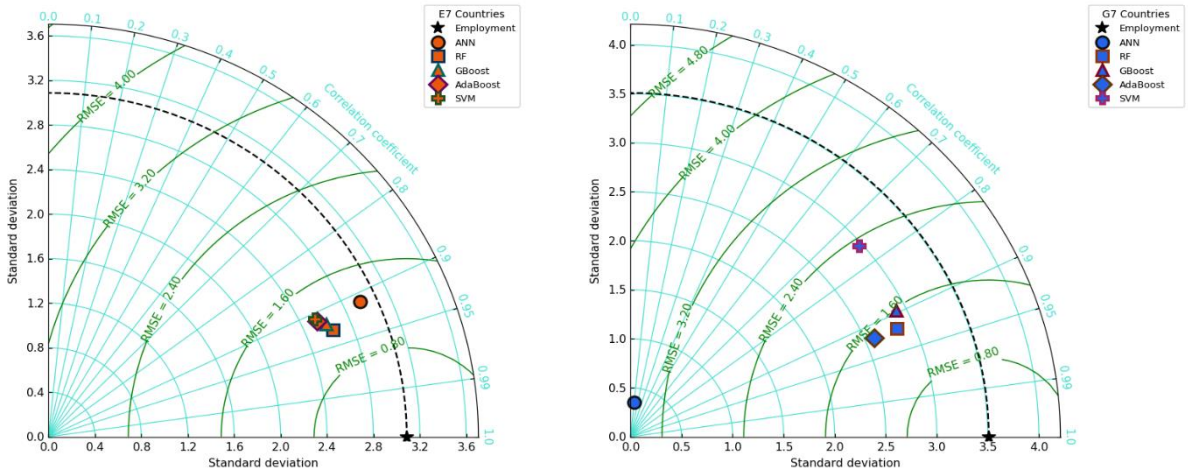
Panel A: E7 Countries Panel B: G7 Countries

Source: Authors' compilation.

Figure 7 - Prediction Errors for E7 and G7 countries

The conformance between the actual values of employment in industry and the modelled values has been assessed using a Taylor diagram (in Figure 8), employing the Pearson correlation coefficient, RMSE, and standard deviation statistics. This method facilitated the identification of the most realistic model. The Taylor diagram, introduced by Taylor (2001), is generally used to quantify the degree of correspondence between modelled and observed behaviour in terms of three statistics: the Pearson correlation coefficient, the RMSE, and the standard deviation (Taylor, 2001).

Figure 8 displays a statistical comparison of 5 ML model predictions of the employment in the industry for E7 and G7 countries.



Panel A: E7 Countries Panel B: G7 Countries

Source: Authors' compilation.

Figure 8 - Taylor diagrams for E7 and G7 countries

These graphs indicate that RF is the model that makes the closest prediction to actual employment with the lowest standard deviation and RMSE values, and with the highest R^2 value for both

country groups, E7 and G7. GBoost is following the RF in the E7 group, it has a very close performance to RF. It is notable that the performance of RF, GBoost, and AdaBoost is similar in both countries. A particularly notable fact is that all the ML models used in the current study point to the same model, achieving satisfactory accuracy values in both country groups. It implies that the ML methods have well learned the pattern representing the data set, and therefore, the predictions of the produced models fit the data.

5. Discussion

This paper aims to forecast industry employment by applying five ML algorithms. For both groups, we obtained the remarkable R^2 at the highest 0.859 and lowest 0.804 for E7 countries. Besides, it is found to be the highest at 0.822 and the lowest at -0.341 for the G7 countries. In the E7 countries, the RF has the highest R^2 with 0.859. Moreover, it is also superior to others with the lowest MSE (1.317), RMSE (1.147), MAE (0.892), MAPE (%) (3.632), RE (%) (3.529), RRMSE (%) (4.538), and RRSE (%) (37.439). The RF is followed by GBoost, SVM, ANNs, and AdaBoost with R^2 values of 0.832, 0.8139, 0.8138, and 0.804, respectively. Additionally, RF displays the best performance for the G7 countries in terms of R^2 with 0.822. Furthermore, it outperforms the others with the lowest MSE (2.150), RMSE (1.466), MAE (1.024), MAPE (%) (4.725), RE (%) (4.593), RRMSE (%) (6.574), and RRSE (%) (42.093). The RF is followed by GBoost, AdaBoost, SVM, and ANNs, with R^2 values of 0.796, 0.765, 0.484, and -0.341, respectively, in the G7 countries. Hence, it is reported that RF predicts the model with a high accuracy level in both country groups.

Considering the box plots of the data used for the E7 countries, it is observed that the countries exhibit heterogeneous patterns. However, box plots of the data for the G7 countries display that these countries have homogeneous patterns. Despite the differences between the two countries, the RF outperforms forecasting others in both groups.

It is possible to compare our results with studies that forecast employment or unemployment using ML algorithms. For instance, in the study by Güler et al. (2024), the R^2 values are determined to range from 0.149 to 0.697 for the ANNs, SVM, and XGBoost models. Moreover, MAE values change from 0.395 to 0.954, RMSE from 0.466 to 1.07, and MAPE from 0.104 to 0.606. According to the analysis by Fahlapi et al. (2020), the accuracy rates are 76.26%, 79.37%, and 78.58% for Random Tree, RF, and C.45, respectively, in the study of Jafor et al. (2023) who concluded that the AdaBoost algorithm outperforms with a 95.30% accuracy rate. El_Jerjawi and Abu-Naser (2018) reached an 87.3% accuracy rate for the ANNs model. Gogas et al. (2022) results indicate that RF is the best prediction algorithm, with an 88.5% accuracy rate in the whole dataset and an 85.4% rate in the out-of-sample dataset. Therefore, the current paper's results align with the studies in the relevant literature, capturing significant and reliable performance criteria.

6. Conclusion

The current paper aims to forecast industry employment in the E7 and G7 countries, and the study's results demonstrate the consistency of ML algorithms in economic data. As Yoon (2021) highlighted, traditional econometric methods aim to reveal the potential causal relationship among variables. However, the ML techniques allow us to test the forecasting power. The current study contributes to the existing literature by comparing the results of the ML algorithms, with a particular focus on considering democracy and digitalization. In terms of democracy, we utilized the liberal democracy index and the participatory democracy index. Moreover, following the literature, the ICTs index is widely recognized as an included model, serving as a proxy for

digitalization. The results consistently show that RF has the best performance in forecasting for both country groups.

Overall, the current study proposes the use of ML algorithms as a significant method for forecasting economic data for policymakers with reliable and consistent policy implications. Although the RF model offered promising findings, the study still has some limitations. Firstly, due to the availability of data for the ICTs index, our analysis period spans from 2000 to 2022. When it is possible to access the updated data, it may be possible to use a wide period for forecasting. Secondly, the prediction performance may vary based on the sample, data coverage, and indicators. Hence, future studies may consider the related variables that potentially may affect industry employment in forecasting.

Acknowledgements

No fund received.

Conflict of interest

No conflict of interest

References

Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. *National Bureau of Economic Research Working Paper 24196*, 1-41. https://www.nber.org/papers/w24196

Adeem, M. A., Sibt-e-Ali, M., & Akhtar, R. (2019). Socio-economic determinants of employment: a case study of Pakistan. *Journal of Accounting and Finance in Emerging Economies, 5*(1), 71-82. https://doi.org/10.26710/jafee.v5i1.724

Amodio, F., Poschke, M., Caprettini, B., Choi, J., Huang, H., Lei, Y. H., Reed, T., Rodrigo, R., Sáenz, L. F., Sanfilippo, M., Schwartzman, M., de Souza, G., Sposi, M., & Wiedemann, V. (2026). Industrial Development. *VoxDevLit, 22*(1), 1-60.

Anjola, A., Kehinde, A., Oluwatobi, F., Timilehin, I. D., & Osabohien, R. (2018). Macro-economic determinants of industrial output in Nigeria: exchange rate and employment dimension. *International Journal of Development Research, 8*(6), 21310-21316.

Arntz, M., Gregory. T., & Zierahn, U. (2016). The risk of automation for jobs in OECD countries: a comparative analysis. *OECD Social, Employment and Migration Working Papers, No. 189*. OECD Publishing, Paris. https://doi.org/10.1787/5jlz9h56dvq7-en

Arntz, M., Terry, G., & Zierahn, U. (2019). Digitalization and the future of work: macroeconomic consequences. *ZEW Discussion Papers, No. 19-024*, ZEW-Leibniz-Zentrum für Europäische Wirtschaftsforschung, Mannheim. https://hdl.handle.net/10419/200063

Aydin, M., Bagci, A., Demirtas, N., Sogut, Y., Altundemir, M. E., & Gürdal, T. (2026). The relationship of digitalization, industrialization, financial freedom and institutional quality with economic complexity: A panel data analysis on E− 7 countries. *Central Bank Review, 26*(1), 100241. https://doi.org/10.1016/j.cbrev.2026.100241

Balsmeier, B., & Woerter, M. (2019). Is this time different? How digitalization influences job creation and destruction. *Research Policy, 48*, 103765. https://doi.org/10.1016/j.respol.2019.03.010

Belete, D. M., & Huchaiah, M. D. (2022) Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *International Journal of Computers and Applications, 44*(9), 875-886. https://doi.org/10.1080/1206212X.2021.1974663

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*(2), 281-305.

Biau, G., Cadre, B., & Rouvière, L. (2019). Accelerated gradient boosting. *Machine Learning,* 108, 971-992. https://doi.org/10.1007/s10994-019-05787-1

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32. https://doi.org/10.1023/A:1010933404324

Carmona, P., Climent. F., & Momparler, A. (2019). Predicting failure in the US banking sector: an extreme gradient boosting approach. *International Review of Economics & Finance, 61*, 304-323. http://dx.doi.org/10.1016/j.iref.2018.03.008

Charles, L., Xia, S., & Coutts, A. P. (2022). Digitalization and Employment: A Review. *International Labour Organization*. www.ilo.org/publns

Davidescu, A. A., Apostu, S. A., & Paul, A. (2021). Comparative analysis of different univariate forecasting methods in modelling and predicting the Romanian unemployment rate for the period 2021–2022. *Entropy, 23*, 325. https://doi.org/10.3390/e23030325

El_Jerjawi, N. S., & Abu-Naser, S. S. (2018). Diabetes prediction using artificial neural network. *International Journal of Advanced Science and Technology, 121*, 55-64. http://dx.doi.org/10.14257/ijast.2018.121.05

Fahlapi, R., Hermanto, H., Kuntoro, A. Y., Effendi, L., & Nitra, R. O. (2020). Prediction of employee attendance factors using C4. 5 algorithm, random tree, random forest. *Semesta Teknika, 23*(1), 39-53. DOI: 10.18196/st.231254

Farooq, M., Ayesha, A., & Gul, A. (2025). Forecasting the future: Neural networks vs. time series in Pakistan's CPI. *Economic Research Guardian*, *15*(2), 219-231.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation?. *Technological Forecasting and Social Change, 114*, 254-280. http://dx.doi.org/10.1016/j.techfore.2016.08.019

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*, 1189–1232. DOI: 10.1214/aos/1013203451

Gajewski, P., Čule, B., & Rankovic, N. (2023). Unveiling the power of ARIMA, support vector and random forest regressors for the future of the Dutch employment market. *Journal of Theoretical and Applied Electronic Commerce Research, 18*, 1365-1403. https://doi.org/10.3390/jtaer18030069

García-Romanos, J., & Martínez-Ros, E. (2024). Digitalization and employment in Europe: the role of firm's size and the complementarity of R&D. *R&D Management, 54*(4), 695-712. https://doi.org/10.1111/radm.12656

Gartner (2024). Digitalization. https://www.gartner.com/en/information-technology/glossary/digitalization (accessed in December 25, 2024).

Genuer, R., Poggi, J. M., Tuleau-Malot, C., & Villa-Vialaneix, N. (2017). Random forests for big data. *Big Data Res, 9*, 28–46. https://doi.org/10.1016/j.bdr.2017.07.003

Gogas, P., Papadimitriou, T., & Sofianos, E. (2022). Forecasting unemployment in the euro area with machine learning. *Journal of Forecasting, 41*, 551-566. DOI: 10.1002/for.2824

Gómez-Plana, A. G., & Latorre, M. C. (2019). Digitalization, multinationals and employment: an empirical analysis of their causal relationships. *Journal of Economics and Statistics, 239*(3), 399-439. https://doi.org/10.1515/jbnst-2017-0153

Güler, M., Kabakçı, A., Koç, Ö., Eraslan, E., Derin, K. H., Güler, M., Ünlü, R., Türkan, Y. S., & Namlı, E. (2024). Forecasting of the unemployment rate in Turkey: comparison of the machine learning models. *Sustainability, 16*, 6509. https://doi.org/10.3390/su16156509

Haraguchi, N., Cheng, C. F. C., & Smeets, E. (2017). The importance of manufacturing in economic development: has this changed?. *World Development*, *93*, 293-315. http://dx.doi.org/10.1016/j.worlddev.2016.12.013

Heo, J., & Yang, J. Y. (2014). AdaBoost based bankruptcy forecasting of Korean construction companies. *Applied Soft Computing, 24*, 494-499. http://dx.doi.org/10.1016/j.asoc.2014.08.009

Huang, Y., Haseeb, M., Usman, M., & Ozturk, I. (2022). Dynamic association between ICT, renewable energy, economic complexity and ecological footprint: is there any difference between E-7 (developing) and G-7 (developed) countries?. *Technology in Society, 68*, 101853. https://doi.org/10.1016/j.techsoc.2021.101853

Huawei Technologies (2024). Global digitalization index 2024.

Ibrahim, F. I. J., Umar, H. A., Bichi, A. S., Ahmad, I. S., Rabiu, N. B., & Ahmad, A. M. (2023). Prediction of unemployment rates with time series and machine learning techniques. In *Proceedings Book of International Conference on Computing and Advances in Information Technology (ICCAIT)*, 21-23 November, 2023, Ahmadu Bello University, Zaria, Nigeria, 183-189.

Ilyas, M., & Khan, M. A. (2018). Democracy versus dictatorship: an empirical investigation of determinants of unemployment. *Pakistan Business Review, 20*(3), 525-534.

IMF. (2018). The digital revolution in Asia: disruptor or new growth engine (or both). *Chapter 5 in regional economic outlook: Asia and Pacific*. World Economic and Financial Surveys (Washington).

Jafor, M. A., Wadud, M. A. H., Nur, K., & Rahman, M. M. (2023). Employee promotion prediction using improved AdaBoost machine learning approach. *AIUB Journal of Science and Engineering, 22*(3), 258-266.

Jyothi, K., Srinu, N., Babu, D. B., & Manda, R. (2026). An expanding window temporal validation framework for global AQI and PM2. 5 forecasting using machine learning and deep learning models. *International Journal of Artificial Intelligence and Machine Learning, 6*(8s), 1090-1100. https://www.svedbergopen.com/

Karahan, M., & Çetintaş, F. (2022). Forecasting of Turkey's unemployment rate for future periods with artificial neural networks. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 62*, 163-184. Doi: 10.18070/erciyesiibd.1056618

Katipoğlu, O. M., & Sarıgöl, M. (2023). Prediction of flood routing results in the Central Anatolian region of Türkiye with various machine learning models. *Stochastic Environmental Research and Risk Assessment, 37*, 2205-2224. https://doi.org/10.1007/s00477-023-02389-1

Katris, C. (2020). Prediction of unemployment rates with time series and machine learning techniques. *Computational Economics, 55*, 673-706. https://doi.org/10.1007/s10614-019-09908-9

Maroof, Z., Hussain, S., Jawad, M., & Naz, M. (2019). Determinants of industrial development: a panel analysis of South Asian economies. *Quality & Quantity, 53*, 1391-1419. https://doi.org/10.1007/s11135-018-0820-8

Mulaudzi, R., & Ajoodha, R. (2020). An exploration of machine learning models to forecast the unemployment rate of South Africa: a univariate approach. In *Proceedings Book of 2nd international Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, 25-27 November, 2020, Kimberley, South Africa,1-7.

Navlani, A. (2018). AdaBoost Classifier in Python. DataCamp, available at, https://www.datacamp.com/tutorial/adaboost-classifier-python (accessed in May 15, 2025).

Nguyen, H., Vu, T., Vo, T. P., & Thai, H. T. (2021). Efficient machine learning models for prediction of concrete strengths. *Construction and Building Materials*, *266(Part B)*, 120950. https://doi.org/10.1016/j.conbuildmat.2020.120950

O'Grady, M. (2024). The global digital economy will reach $16.5 trillion and capture 17% of global GDP by 2028. *Forrester.* https://www.forrester.com/blogs/the-global-digital-economy-will-reach-16-5-trillion-and-capture-17-of-global-gdp-by-2028/ (accessed in January 01, 2025).

Olalekan, A. J., Odion, P. O., Irhebhude, M. E., & Aminu, H. (2021). Performance evaluation of machine learning predictive analytical model for determining the job applicants employment status. *Malaysian Journal of Applied Sciences, 6*(1), 67-79. https://doi.org/10.37231/myjas.2021.6.1.276

Orozco-Castañeda, J. M., Sierra-Suárez, L. P., & Vidal, P. (2024). Labor market forecasting in unprecedented times: a machine learning approach. *Bulletin of Economic Research, 76*, 893-915. DOI: 10.1111/boer.12451

Oueghlissi, R., & Derbali, A. (2024). Democracy, corruption and unemployment: empirical evidence from developing countries. *Journal of the Knowledge Economy, 15*(2), 7475-7496. https://doi.org/10.1007/s13132-023-01204-0

Our World in Data. (2024). Democracy. https://ourworldindata.org/democracy (accessed in December 15, 2024).

Pandey, D. (2024). Impact of digitalization on employment pattern in India. *International Journal for Multidisciplinary Research (IJFMR), 6*(1), 1-12.

Pärli, K. (2022). Impacts of digitalisation on employment relationships and the need for more democracy at work. *Industrial Law Journal, 51*(1), 84-108. https://doi.org/10.1093/indlaw/dwaa029

Patel, J. N., Kızılay, A., Şahin, Z., Sercan, B., Toprak, S., & Çakar, T. (2025). Ensemble-Based Stock Prediction for Retail-XGBoost and LightGBM with Rolling Window Training. In *33rd Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE. DOI: 10.1109/SIU66497.2025.11112029

Pattanaik, F., & Nayak, N. C. (2014). Macroeconomic determinants of employment intensity of growth in India. *Margin: The Journal of Applied Economic Research, 8*(2), 137-154. DOI: 10.1177/0973801013519997

Pew Research Center (2016). Public predictions for the future of workforce automation. www.pewresearch.org (accessed in December 15, 2024).

PwC (2017). Shift of global economic power to emerging economies set to continue in long run, with India, Indonesia and Vietnam among star performers. https://www.pwc.com/id/en/ (accessed in May 05, 2024).

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports, 14*, 6086. https://doi.org/10.1038/s41598-024-56706-x

Rezgui, S., & Bakari, S. (2025). The Impact of digitalization on the three pillars of sustainable development: An empirical analysis in the context of African countries. *Economic Research Guardian*, *15*(1), 62-87.

Reljic, J., Evangelista, R., & Pianta, M. (2021). Digital technologies, employment, and skills", *Industrial and Corporate Change*, 1-22. DOI: https://doi.org/10.1093/icc/dtab059

Richiardi, M. G., Westhoff, L., Astarita, C., Ernst, E., Fenwick, C., Khabirpour, N., & Pelizzari, L. (2025). The impact of a decade of digital transformation on employment, wages, and inequality in the EU: a "conveyor belt" hypothesis. *Socio-Economic Review*, 1-27. https://doi.org/10.1093/ser/mwaf011

Sarker, I. H. (2021). Machine learning: algorithms, real-world applications and research directions. *SN Computer Science, 2*, 160. https://doi.org/10.1007/s42979-021-00592-x

Senthurvelautham, S., & Senanayake, N. (2023). A machine learning-based job forecasting and trend analysis system to predict future job markets using historical data. In *Proceedings of the 8th IEEE (Institute of Electrical and Electronics Engineers) International Conference for Convergence in Technology*, 7-9 April 2023, Pune, India, 1-7. https://doi.org/10.1109/I2CT57861.2023.10126233

Shams, M. Y., Elshewey, A. M., El-Kenawy, E. S. M., Ibrahim, A., Talaat, F. M., & Tarek, Z. (2024). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications*, *83*(12), 35307-35334. https://doi.org/10.1007/s11042-023-16737-4

Shuttleworth, L., Schmitz, S., & Beier, G. (2022). Impacts of industry 4.0 on industrial employment in Germany: a comparison of industrial workers' expectations and experiences from two surveys in 2014 and 2020. *Production & Manufacturing Research, 10*(1), 583-605. https://doi.org/10.1080/21693277.2022.2104400

Sterkenburg, T. F., & Grünwald, P. D. (2021). The no-free-lunch theorems of supervised learning. *Synthese, 199*(3-4), 9979-10015. https://doi.org/10.1007/s11229-021-03233-1

Subasi, A., Panigrahi, S. S., Patil, B. S., Canbaz, M. A., & Klén, R. (2022). Advanced pattern recognition tools for disease diagnosis. In *5G Iot and edge computing for smart healthcare* (pp. 195-229). Academic Press. https://doi.org/10.1016/B978-0-323-90548-0.00011-5

Swetha, M. S., Mahalakshmi, S., Pushpa, S. K., Madihalli, A. T., & Ananya, B. A. (2023). Machine learning approach for forecasting job appeasement and employee corrosion. In *Proceedings Book of the Eighth International Conference on Research in Intelligent Computing in Engineering*, 1-2 December, 2023, Hyderabad, India, 85–91. DOI: 10.15439/2023R17

Takki, K. (2024). Future of ındustrial automation: market trends, key players, and growth projections. *Statzon.* https://statzon.com/insights/industrial-automation-market (accessed in December 27, 2024).

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research, 106*(D7), pp. 7183–7192. https://doi.org/10.1029/2000JD900719

Touzani, S., Granderson, J., & Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, *158*, 1533-1543. https://doi.org/10.1016/j.enbuild.2017.11.039

Uçar, M. K., Nour, M., Sindi, H., & Polat, K. (2020). The effect of training and testing process on machine learning in biomedical datasets. *Mathematical Problems in Engineering, 2020*, 2836236. https://doi.org/10.1155/2020/2836236

UNCTAD. (2024). Statistics. https://unctad.org/statistics referred on 25/12/2024 (accessed in December 25, 2024).

UNCTAD. (2026). Gross domestic product: Total and per capita – Annual. https://unctadstat.unctad.org/datacentre/dataviewer/US.GDPTotal (accessed in October 01, 2026).

Van Noort, S. (2024). Industrialization and democracy. *World Politics, 76*(3), 457-498. DOI: 10.1353/wp.2024.a933069

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. New York: Springer.

Wajebo, T. W. (2024). Determinants of labour demand in manufacturing sector in Ethiopia. *The Indian Journal of Labour Economics, 67*, 751-782. https://doi.org/10.1007/s41027-024-00515-3

Walkme (2024). What is digitalization?. https://www.walkme.com/glossary/digitalization/ (accessed in 25 December, 2024).

World Bank (2024). World development indicator. https://data.worldbank.org/ (accessed in December 25, 2024).

World Bank (2026). Indicators. https://data.worldbank.org/ (accessed in October 05, 2026).

Wu, Y. C., & Feng, J. W. (2018). Development and application of artificial neural network. *Wireless Personal Communications, 102*, 1645-1656. https://doi.org/10.1007/s11277-017-5224-x

Yoon, J. (2021). Forecasting of real GDP growth using machine learning models: gradient boosting and random forest approach. *Computational Economics, 57*, 247-265. https://doi.org/10.1007/s10614-020-10054-w

Yunxia, W., Neng, H., & Yechi, M. (2023). The effect of digital economy development on labor employment: empirical evidence from listed companies in China. *Journal of Global Information Management (JGIM)*, *31*(6), 1-27. DOI: 10.4018/JGIM.321180

Yurtsever, M. (2023). Unemployment rate forecasting: LSTM-GRU hybrid approach. *Journal for Labour Market Research, 57*, 18. https://doi.org/10.1186/s12651-023-00345-8

Zendehboudi, A., Baseer, M. A., & Saidur, R. (2018). Application of support vector machine models for forecasting solar and wind energy resources: a review. *Journal of Cleaner Production, 199*, 272-285. https://doi.org/10.1016/j.jclepro.2018.07.164

Zhang, X., Lin, F., Wang, Y., & Wang, M. (2022). The impact of digital economy on employment polarization: an analysis based on Chinese provincial panel data. *Labor History, 63*(5), 636-651. https://doi.org/10.1080/0023656X.2022.2133101

Zhao, Y., & Said, R. (2023). The effect of the digital economy on the employment structure in China. *Economies, 11*, 227. https://doi.org/10.3390/economies11090227