Metaheuristic Optimization Review

♦ ISSN: 3066-280X Vol: 06, No. 01, 2026 ♦ PP. 01–14
Article DOI: https://doi.org/10.54216/MOR.060101

FULL LENGTH ARTICLE

DTOSFS–CatBoost: A Hybrid Metaheuristic Framework for Accurate and Interpretable Unemployment Forecasting

Ghassan AL-Thabhawee[1] **Hussein Alkattan**[2]

[1] Sciences of Mathematics, Computer Sciences, College of Health and Medical Techniques-Kufa, Al-Furat Al-Awsat Technical University, Kufa, Iraq

[2] Department of System Programming, South Ural State University, Chelyabinsk, Russia; Directorate of Environment in Najaf, Ministry of Environment, Najaf, Iraq

Emails: gmohammed@atu.edu.iq · alkattan.hussein92@gmail.com

Received: January 03, 2026 Revised: March 01, 2026 Accepted: May 02, 2026 ⋆ Corresponding author

ABSTRACT

The fact that educational, demographic, and macroeconomic variables interact nonlinearly has remained a thorn in the flesh of socio-economic analytics to date, making it challenging to forecast unemployment with sufficient precision. To address this, the current study presents a hybrid metaheuristic, Dipper Throated Optimization with Stochastic Fractal Search (DTOSFS), coupled with the Category Boosting (CatBoost) algorithm to improve predictive modelling. The suggested DTOSFS-CatBoost system combines the general exploratory search of DTO with SFS refinement to stochastic local optimization of hyperparameters, and alleviates overfitting. Empirical experiments have shown that whereas the original CatBoost gave results with a Mean Squared Error (MSE) of 0.0256 and Root Mean Squared Error (RMSE) of 0.1601 with a correlation coefficient of 0.873, the CatBoost optimized by DTOSFS had drastically better results with an MSE of 0.00033, RMSE of 0.00207, and a correlation coefficient of 0.930. These results confirm an increased exploration-to-exploitation ratio in DTOSFS and yield small, powerful designs that substantially enhance model stability, precision, and convergence speed. These results show that educational attainment (at least tertiary and primary enrollment) and demographics (at least the birth rate) are influential factors in unemployment variation. This addition to predictive performance is not the only one, and it provides a predictive data-driven labor-market optimization paradigm that can be replicated and interpreted. The research observes that hybrid metaheuristics and gradient boosting can be used to drive next-generation economic intelligence systems for adaptive policy formulation and to enhance online, privacy-conscious, and cross-domain unemployment prediction.

Keywords: DTOSFS ▪ CatBoost ▪ Unemployment Forecasting ▪ Hybrid Metaheuristic Optimization ▪ Socio-Economic Prediction

1. INTRODUCTION

Employment is one of the most acute problems that negatively impacts the socioeconomic development of countries, societies, and people in general [1]. The comparatively high unemployment rates slow progress and development, aggravate poverty, and reduce the quality of life of all people . Moreover, unemployment negatively affects society through its effects on consumer spending, the growth of dependence on social programs, and psychological effects on both communities and individuals. To that end, accurately predicting unemployment rates could help policymakers and businesses address these challenges [2]. The study of unemployment patterns, therefore, helps governments design better approaches, strategies, and resources for interventions when necessary to stabilize and grow the economy. However, this is often not easy because unemployment is a complex phenomenon, which makes predicting this rate difficult as well [3]. Economic, demographic, and temporal influences are intertwined in the unemployment system. They cannot be easily parsed out into a limited number of variables from which an intuitively designed mathematical model to predict unemployment rates can be derived . Specifically, Artificial Intelligence (AI) and Machine Learning (ML) have become popular approaches to solving predictive problems, as outlined above. It would be beneficial to use machine learning models because they can capture complex, nonlinear feature dependencies, making them suitable for problems with multiple factors, such as unemployment prediction. The ability of most ML algorithms to handle large, complex datasets enables them to discover relationships that are not readily observable by other methods. CatBoost has become one of the most often used algorithms among these. CatBoost, a gradient boosting technique, may be applied to both numerical and categorical variables [4]. Its benefits include improved classification and regression performance, resistance to overfitting, and fewer preprocessing steps. Because the data incorporates many types of variables, these features make it desirable to use it to fit a model and forecast the unemployment rate. A sensible, thorough examination of the data set is the foundation of a successful machine learning project. To estimate the unemployment rate, data analysis entails determining and adjusting variables such as GDP growth, inflation rates, labour market dynamics, and demographics . These are some of the explanations for the fundamental causes of variations in unemployment over time and between geographical areas. Inconsistencies, redundant information, and unnecessary details that shouldn't be in the prediction model are removed by data cleaning. Analysing interactions among variables not only improves model accuracy but also reveals characteristics that influence the dynamics of unemployment. Policymakers can use these findings to develop well-informed plans to address unemployment [5]. Once the data is selected and the model is developed, optimization is another critical step to make the machine learning algorithms more efficient [6]. The art of parameter optimization to achieve near-optimal solutions with minimal overfitting is called optimization. The popularity of metaheuristic optimization methods is attributed to their ability to handle high-dimensional, large-scale search spaces . These methods are motivated by natural processes such as animal behavior and evolution and provide systematic approaches for obtaining reasonable or near-optimal solutions. Here, a new metaheuristic algorithm, DTOSFS, is used for optimization. DTOSFS integrates the strengths of the Dipper Throated Optimization Algorithm (DTO), which also relies on the strategic foraging capability of birds, but is founded on the theory of fractal and stochastic processes, Stochastic Fractal Search (SFS). The given methodology combines the broad search for DTO with stochastic fine-tuning of SFS to determine the optimal hyperparameters of the CatBoost model. Publication of this paper, therefore, leads to the development of an optimal model that improves predictive capability and reliability, with performance equal to or better than other models in forecasting the unemployment rate. The following are the primary goals of the paper:

1. To investigate the application and effectiveness of machine learning models, specifically CatBoost, for predicting unemployment rates.

2. To examine the role of data preprocessing in enhancing the reliability and accuracy of predictive models.

3. To apply the hybrid DTOSFS optimization method to optimize the hyperparameters of the CatBoost model, which guarantees optimal performance.

4. To evaluate how well the DTOSFS-optimized CatBoost model performs in comparison with other machine learning models and optimization techniques.

5. To obtain clues about the main aspects of working on unemployment rates, it is essential to use the outcomes of the optimized model to improve the knowledge about the interaction between unemployment rates.

The research is presented clearly and coherently in this well-structured report. After reviewing the literature on unemployment forecasting and machine learning methods, the paper describes the dataset, the CatBoost model implementation, and the DTOSFS optimization procedure. The results section presents the experimental data and the comparison analysis, and the discussion portion interprets the results in terms of unemployment prediction. Finally, the report will examine the study's contributions, highlight important lessons learn, and give suggestions for future research.

2. LITERATURE REVIEW

Time series and machine learning methods have been widely used to predict unemployment across the Mediterranean, Baltic, Balkan, Nordic, and Benelux regions. Among them, Fractional Autoregressive Integrated Moving Average (FARIMA) models are particularly effective, especially for long-memory time series, as in the research by [7]. Fractions of autoregressive integrated moving-average models with generalized autoregressive conditional heteroskedasticity errors and models with non-normal error distributions are also explored to improve accuracy and address heteroskedasticity. Moreover, to address data nonlinearity, modern machine learning algorithms, including fully connected feedforward neural networks, support vector regression, and multivariate adaptive regression splines, are used. Classical models, such as autoregressive integrated moving-average and Holt-Winters, are used as a reference point, and the effects of different forecast horizons and geographical variations on model performance are also discussed. Estimators of matching type have been adopted as key instruments for assessing active labour market policies, using the propensity score. The research presented in [8] examines whether machine learning algorithms that estimate propensity scores can improve the plausibility of average treatment effect estimates for the treatment, in this case, a radius-matching framework. The results show conflicting outcomes: logistic regression models with the least absolute shrinkage and selection operator (LASSO) provide more valid estimates in small to medium-sized samples with a high number of dimensions, whereas applying Random Forest algorithms can lead to poorer results when low treatment shares characterise the situation. The review shows that training programs have positive effects on the employment period among the long-term unemployed. It is important to note that the decision regarding the first estimation procedure is crucial in environments with a small number of observations and low treatment rates. Still, the distinction between machine learning and conventional methods blurs in large datasets with high treatment rates. Minimum wage policies have received considerable research on their effects on labor market outcomes, such as employment, unemployment, and labor force participation. In the analysis presented by , modern machine learning techniques are used to form demographic-based treatment groups that include almost 75 percent of all minimum-wage workers, in contrast to previous work that primarily focused on small subsets of the target population, such as teenagers. The study finds a significant increase in the average wage of beneficiaries of minimum-wage increases, based on 172 substantial minimum-wage changes between 1979 and 2019, with little evidence of job loss. Also, no adverse impacts on unemployment rates, labor market transitions, or labor force participation are found, including for high-labor-supply-elasticity groups such as teens, older workers, or single mothers. These findings indicate that minimum wage raised do not have a significant impact on the search effort and labor market behavior of the concerned populations. One of the performance indicators for higher education institutions is graduate employability, the capacity of the institution to equip students with industry requirements. This paper examined the performance of mock job interviews and on-the-job training, as well as overall point averages from 2015 to 2018. It addressed unbalanced datasets using synthetic minority over-sampling, as shown by [9]. Six machine learning algorithms were considered, and the Support Vector Machine with 91.22% accuracy was the best, compared to Decision Trees (85%) and Random Forests (84%). The findings substantiate the strength of the Support Vector machine framework, promising future studies to build a better predictive model and confirm it. Unemployment is a serious problem that affects a country's economic and financial stability, so accurate prediction of the unemployment rate is crucial for planning and policy-making regarding a country's growth. The article by [10] provides an analysis of unemployment rate forecasting based on conventional and new time-series approaches; however, the nonstationary and nonlinear features of macroeconomic data pose challenges for the predictive process. In this paper, a hybrid model combining linear and nonlinear techniques is proposed to improve forecasting accuracy by correcting the skew in the trend of unemployment rates. The model is demonstrated using results from seven countries — Canada, Germany, Japan, the Netherlands, New Zealand, Sweden, and Switzerland —which show promising results compared to more traditional methods. In addition, the analysis ascertains the asymptotic stationarity of the hybrid model using Markov chain and nonlinear time series analysis techniques, demonstrating that the model effectively resists explosive dynamics and rising variance over time. The implications of artificial intelligence, robots, and unemployment on the dynamics of the labor market are massive. According to a study conducted by [11] across 33 Organization for Economic Co-operation and Development countries between 2005 and 2017, a 10 percent increase in the stock of industrial robots is associated with a 0.42 point increase in unemployment rates. When patents are used as a proxy for technological developments in artificial intelligence, a positive but weaker correlation is found with overall unemployment rates. Additional age and education analysis revealed that the effect of robots is extremely disproportionate, with younger people without upper secondary education (25-34 years old) being affected 2.5 times more than older people with tertiary education (55-64 years). The influence of robots on unemployment rates is most significant among the medium-educated, indicating a polarization effect on the labor market. Although the same tendencies are observed in artificial intelligence, they are not as strong as those in robots. Universities across the globe are trying to develop study plans that would make their graduates more successful in the employment market, and internships are among the most complex opportunities in experiential learning. Although, as shown by , internships are valuable as they allow students to practice what they have learned and train them to enter the workforce, they do not make them employable, particularly when performance in the internship is not satisfactory or when the student fails to meet the requirements. To address this issue, this paper presents a strategy for forecasting employability using gradient-boosting classifiers. Employability prediction aims to determine the characteristics that influence graduates' hiring chances across two contexts: the student context (based on student characteristics) and the internship context (based on internship-related factors). The comparison of three gradient boosting classifiers — eXtreme Gradient Boosting, Category Boosting, and Light Gradient Boosted Machine — shows that the latter is the best when used on internship context data. These results show that graduate employability is highly dependent on the internship situation, which can serve as a prescriptive model for higher education institutions. Predictive indicators of student success in a course or program enable improvements in educational outcomes and increased employability of graduates. According to the reports of [12], efficient performance forecasting methods help teachers allocate resources and design instruction more effectively. The specific area of the study is the development of the Technical Skills-Based Employability Prediction Model, which can be based on machine learning to predict the consequences of technical skills, as demonstrated by students' programming course results. Real data from university graduates was used to test various machine learning classifiers, including Support Vector Machine, Naive Bayes, Logistic Regression, Decision Tree, Random Forest, AdaBoost, and Artificial Neural Network. Random Forest was the most accurate, with 70% accuracy and an F1-Score of 0.85. The model is also helpful in predicting whether a student will be able to get a job, thereby providing a valuable piece of information to an educational institution that intends to increase the rate of graduate placement. However, both studies emphasise the value of employing data mining and machine learning techniques to predict and improve students' employability. Computer engineering students are more employable due to academic, personal, and regional characteristics, according to the article by [13]. The accuracy of Artificial Neural Networks is 87.5%, compared to 62.5% for Logistic Regression. As in the article by [14], educational data mining techniques and a hybrid algorithm combining fuzzy c-means clustering and particle swarm optimisation were used. The hybrid algorithm demonstrated lower time complexity and higher prediction rates than traditional algorithms, demonstrating its applicability to improving employability forecasting. The duration of unemployment statistics for the economically active population is vital for helping policymakers and employment agencies allocate resources and provide targeted assistance to job seekers. In the case of real data on job seekers registered with the Local Labour Office, Social Affairs and Family, an ensemble model was constructed to predict the duration of unemployment in Slovakia (cited in the study by [15]). The model uses the stacking method, which involves predictions from three separate models (excluding classification and regression trees, chi-squared automatic interaction detection, and discriminant analysis), and the user uses logistic regression as a meta-model. The ensemble's accuracy was almost 78%, and it identified job seekers at risk of long-term unemployment (more than 12 months) very well. This powerful model contributes to a feasible instrument for assessing existing labor market policies and verifying regional disparities in unemployment. It also provides a model to justify state-financed interventions to reduce long-term unemployment. The data-mining methodology allows the model to be adjusted to suit other economies and to consider the specifics of the domestic state of affairs and the labor market. The impact of artificial intelligence on unemployment is becoming a point of anxiety, particularly amid fluctuating economic conditions. The article analyses the relationship between artificial intelligence and unemployment using a sample of artificial intelligence-related patents from 40 developed and developing nations, with 2000-2019 as an example (as shown by [16]. The research uses a panel smooth transition regression model to estimate a nonlinear interaction that depends on the level of inflation. The results suggest that artificial intelligence will initially create more joblessness up to a certain level of inflation, after which its effect decreases. The functional analysis mechanism used in this study is capable and sufficient to capture differences in the country of interest over time and to provide detailed information on the interaction between artificial intelligence and labor markets. These findings are relevant to the realm of mediating the impacts of technological development on unemployment by economic conditions. In the context of digital change and education, the examined research demonstrates a variety of methods for comprehending and forecasting employability and technological unemployment. [17] examined the socioeconomic impacts of Industry 4.0 and discovered that deep learning, machine learning, and artificial intelligence are significant causes of technical unemployment in the textile sector. [18] showed that neural networks had the highest accuracy when it came to machine learning predictions of unemployment using data from smart meters. [19] used Teaching Learning Based Optimization to improve the accuracy of Support Vector Machine performance in forecasting the employability of engineering graduates by 13.43%. By improving feature selection, [20] developed a hybrid deep belief network–softmax regression model that achieved over 98% accuracy in predicting employability. [21] integrated statistical and machine learning analysis to demonstrate that demographic and academic characteristics have a significant impact on the employment outcomes of MBA students. Lastly, despite government-led innovation programs, examined China's higher education and entrepreneurship landscape, identifying cultural, resource, and personal impediments to graduate employment. To predict the employability of unstable and developing nations, it is necessary to consider contextual factors and rely on an appropriate machine learning model. This paper, as shown by , defines various factors, including parental financial stability, sociopolitical conditions, relationships, academic performance, and strategic choices, that forecast the employability of information technology graduates in the Democratic Republic of Congo. The deep stacking predictive model, built from 5 multilayer perceptron submodels, achieved 80% accuracy, with a precision, recall, and F1-score of 0.81, 0.80, and 0.77, respectively. The performance of individual models was not consistently strong across all metrics, and deep stacking was the most successful approach for constructing a generalizable employability prediction model. The findings are helpful and may assist the DRC and other nations of a similar kind in formulating measures to alleviate unemployment. Brexit and the Happiness cost of Brexit in both the European Union and the UK are evaluated using Gallup World Poll data. As shown by , a two-stage machine learning solution is applied, with the initial step being to extract happiness preferences using a naive Bayes classifier, which are then fed into an artificial neural network within an agent-based model to produce dynamic happiness functions per household. The results show that there is a significant long-term cost in terms of happiness and unemployment, especially for the most vulnerable groups in the population. Although the financial sector in London is expected to remain unstable, the UK financial system appears well-positioned to weather the aftermath, thereby limiting welfare expenditures. This paper builds on the economic debate on Brexit by adding to the welfare cost of financial instability. Table 1 gives a brief overview of the literature reviewed, including the primary purpose of the studies, their methodology, and their significant findings regarding the topic of unemployment forecasting and employability modelling.

3. MATERIAL AND METHODS

To orient the reader to our experimental pipeline, Figure 1 summarizes the end-to-end framework adopted in this study. Beginning with the raw multi-country dataset, we perform *data preprocessing* (systematic cleaning, missing-value imputation, and feature normalization) to standardize inputs across sources and years. The curated data are then partitioned using an 80/20 *train–test* split to prevent information leakage and enable out-of-sample assessment. On the training fold, we establish *baseline learners*—Category Boosting (CatBoost), eXtreme Gradient Boosting (XGBoost), Decision Tree, and Gradient Boosting—before invoking the *optimization stage*. Hyperparameters are tuned with our hybrid **Dipper Throated Optimization + Stochastic Fractal Search (DTOSFS)** optimizer, and contrasted against alternative metaheuristics—Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), and Whale Optimization Algorithm (WOA). The selected configurations are finally stress-tested on the held-out set and summarized via a unified *performance evaluation* block (MSE, RMSE, MAE, MBE, r, R^2, RRMSE, NSE, and WI), closing the loop between optimization and generalization. This modular design makes the workflow reproducible,

Table 1. Summary of related work on unemployment prediction and employability (Objectives, Methods, and Key Findings).

Reference	Objective	Methodology	Key Findings
[7]	Predict unemployment in multiple European regions accounting for long memory and heteroskedasticity	FARIMA and FARIMA–GARCH with non-normal errors; ML benchmarks (FFNN, SVR, MARS); ARIMA/Holt-Winters as baselines	FARIMA-based models suit long-memory unemployment series; ML handles nonlinearity; forecast accuracy varies by horizon and region
[8]	Assess whether ML improves propensity-score matching for ALMP evaluation	Radius matching with PSM estimated via LASSO-logit vs. Random Forest	LASSO-logit yields more credible ATT in small/high-dim samples; RF can underperform with low treatment share; training programs increase employment duration
[22]	Measure effects of minimum wage hikes on labor outcomes using broad affected groups	ML to construct demographic treatment groups; analysis of 172 MW changes (1979–2019)	Wages rise for affected workers; little evidence of job loss; no adverse effects on unemployment/LFPR or transitions across groups
[9]	Predict graduate employability from institutional assessments	SMOTE to handle imbalance; compare six ML models	SVM achieves 91.22% accuracy, outperforming DT (85%) and RF (84%) for employability prediction
[10]	Improve unemployment rate forecasts under non-stationary, nonlinear dynamics	Hybrid linear–nonlinear model; Markov chains and nonlinear TS for stationarity	Hybrid model outperforms conventional methods across seven countries; asymptotically stationary and robust
[11]	Quantify impacts of robots/AI on unemployment across OECD	Cross-country panel (2005–2017); robots measured by stock; AI by patents	10% robot increase $\rightarrow$ +0.42 pp unemployment; stronger effects on younger/less educated; polarization patterns; AI effects weaker
[23]	Predict employability with focus on internship context vs. student attributes	Gradient boosting (XGBoost, CatBoost, LightGBM) across two context models	Internship-context features most predictive; LightGBM on internship data performs best
[12]	Build technical-skills-based employability predictor	Compare SVM, NB, LR, DT, RF, AdaBoost, ANN on real graduate data	RF tops with 70% accuracy and F1=0.85; technical programming scores are strong predictors
[13]	Predict employment of computer engineering students using competency and background factors	ANN and Logistic Regression on 14 parameters	ANN achieves 87.5% accuracy vs. LR 62.5%; ML supports guidance for improving placements
[14]	Enhance educational data mining for employment prediction	Hybrid PSO + fuzzy c-means; compare to k-means, NB, SVM	Accuracy improves by 28–36% over baselines; time complexity reduced by 33–49%
[15]	Predict unemployment duration and flag long-term risk (>12 months)	Stacking ensemble (CART, CHAID, Discriminant) with logistic meta-model on administrative data	$\sim$78% accuracy; effectively identifies long-term risk; supports targeted policy interventions
[16]	Examine non-linear effects of AI on unemployment moderated by inflation	Panel Smooth Transition Regression across 40 countries (2000–2019)	AI initially raises unemployment up to an inflation threshold; effects diminish beyond threshold
[17]	Assess Industry 4.0 disciplines driving technological unemployment in textiles	Mixed methods; AHP with literature, bibliometrics, and expert input	AI/ML/deep learning are primary contributors; documents neo-Luddism/resistance dynamics
[18]	Detect unemployment via smart meter data	MLP neural network vs. distance-weighted discrimination	MLP AUC=0.74, sensitivity=0.54, specificity=0.83; smart metering viable for autonomous detection

leak-aware, and directly comparable across models and optimizers.

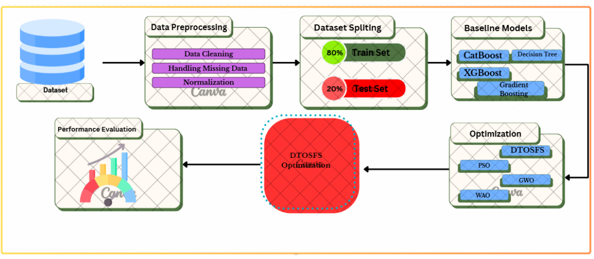

Figure 1. Overall methodological framework of the proposed DTOSFS–CatBoost modeling pipeline, including data preprocessing, dataset splitting, baseline model training, optimization with DTOSFS and comparative metaheuristics (PSO, GWO, WOA), and final performance evaluation.

3.1 Dataset Description

The world has made significant advances in basic education. Education has now been recognized as a fundamental human right, and in most cases, governments have a responsibility to ensure it is available to every citizen. Nevertheless, formal education is a relatively recent phenomenon in human history. This study is based on statistics showing this educational advancement worldwide. It demonstrates the proportion of the adult population, or of those aged 15 or older, who have attained some basic education relative to their less educated counterparts. At the beginning of the nineteenth century, there were hardly any adults with even a basic level of education; education was a luxury reserved for a few privileged elites. The rate has changed drastically over the centuries that have followed, with fewer than one in five people today having no formal education. Nearly all adults possess basic literacy skills, according to recent global literacy figures. This type of development shows how much money is being spent globally to expand access to human capital and education. Numerous characteristics characterising nations and their educational indicators are included in the data in this study. The variables listed in Table 2 are included in the dataset, which consists of records corresponding to a specific nation or area. These aspects will provide a complete picture of gender and educational-level differences, as well as

Table 2. Description of variables in the education dataset.

Variable Name	Description
Countries and Areas	Country or region represented in the dataset.
Latitude	Geographical latitude of the country or area.
Longitude	Geographical longitude of the country or area.
OOSR_Pre0Primary _Age_Male	Out-of-school rate for males in pre-primary age groups.
OOSR_Pre0Primary _Age_Female	Out-of-school rate for females in pre-primary age groups.
OOSR_Primary _Age_Male	Out-of-school rate for males of primary school age.
OOSR_Primary _Age_Female	Out-of-school rate for females of primary school age.
OOSR_Lower_Secondary _Age_Male	Out-of-school rate for males of lower secondary school age.
OOSR_Lower_Secondary _Age_Female	Out-of-school rate for females of lower secondary school age.
OOSR_Upper_Secondary _Age_Male	Out-of-school rate for males of upper secondary school age.
OOSR_Upper_Secondary _Age_Female	Out-of-school rate for females of upper secondary school age.

universal educational coverage. Considering the process of obtaining a basic education, the data allows a closer look at educational development and comparison with other countries. The importance of different demographic and education characteristics in the determination of unemployment was estimated on a feature importance analysis as in Figure 2. This research identifies the variables that make the most significant contribution to the predictive model's output. According to the data, the strongest variables are birth rate, gross primary school enrolment, and gross tertiary education enrolment. This suggests that demographic factors and access to higher education are major contributors to unemployment outcomes. On the other hand, albeit to a lesser extent, factors such as completion and out-of-school rates also have a significant influence. Knowledge of the distribution of fundamental de-

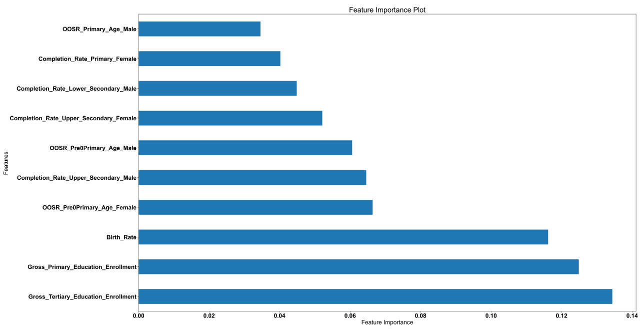

Figure 2. Feature importance plot illustrating the relative influence of educational and demographic indicators on the model's predictions.

mographic indicators can provide the information needed to understand a population's structural peculiarities and their possible impact on socioeconomic outcomes. Figure 3 shows the frequency distribution of Birth rate and Unemployment rate on the various observations. As illustrated, the birth rate has a broader distribution with a longer right tail, indicating greater variability across regions or countries. In contrast, the unemployment rate is more concentrated at low values, with fewer extreme cases. The study of such distributions helps identify demographic imbalances and determine their relationship to labor market conditions. A study of enrolment

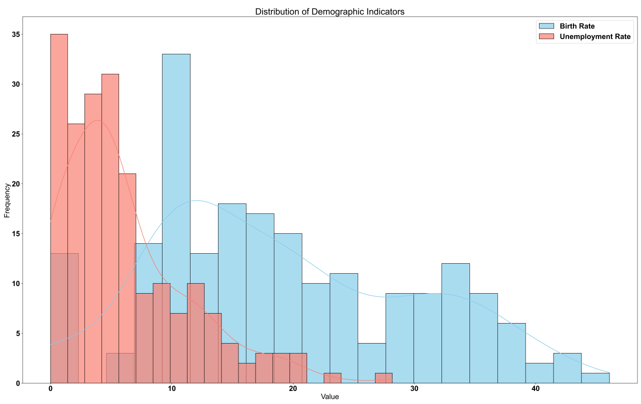

Figure 3. Distribution of demographic indicators showing the frequency of birth rate and unemployment rate across observations.

patterns across the various levels can also provide insight into variations in the educational systems and their accessibility. The frequency distribution of the Primary Education Enrolment and Tertiary Education Enrolment rates is presented in Figure 4. As the figure illustrates, the highest enrolment is in primary education, indicating that most nations have invested in basic education. Tertiary education enrolment, on the other hand, is skewed to the right and has a much broader distribution, indicating a high disparity in access to higher education across regions. These patterns show the persistence of the achievement gap in higher education, despite overall opportunities in primary education.

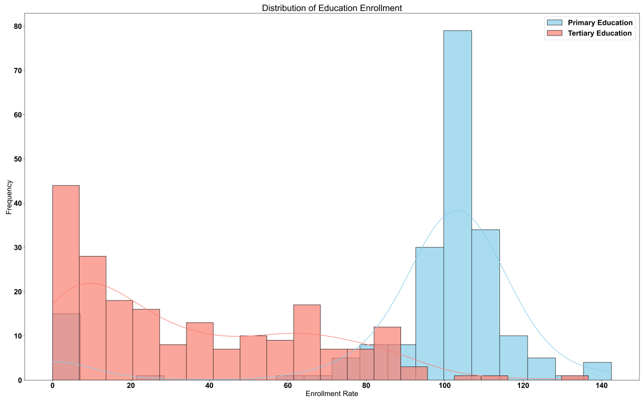

Figure 4. Distribution of education enrollment rates across primary and tertiary levels.

3.2 Machine Learning Models

Several machine learning algorithms were used in this study to predict and analyze educational outcomes across various regions. The main algorithm used was CatBoost, with a specific VCatBoost variant —, an external/attention-enhanced version —, which served as the main temporal learner. CatBoost (Categorical Boosting) is a decision-tree-based, high-performance gradient boosting model that can work with both categorical and numerical variables. Unlike traditional boosting algorithms, CatBoost uses an ordered boosting algorithm that reduces the problem of prediction shift, which is often a concern in gradient boosting because training and target samples are independent at each iteration. The overall concept of CatBoost is based on the additive model of gradient boosting, with the final prediction being presented as:

$$\hat{y}_i = \sum_{m=1}^{M} \eta \, f_m(\mathbf{x}_i), \tag{1}$$

where $\hat{y}_i$ is the predicted value for sample i, M is the total number of trees, η is the learning rate, and $f_m(\mathbf{x}_i)$ represents the prediction from the m^{th} decision tree. Each subsequent tree minimizes the gradient of the loss function $\mathscr{L}$:

$$f_m = \arg\min_{f} \sum_{i=1}^{N} \left[g_i f(\mathbf{x}_i) + \frac{1}{2} h_i f(\mathbf{x}_i)^2 \right], \tag{2}$$

where $g_i = \frac{\partial \mathscr{L}(y_i, \hat{y}_i)}{\partial \hat{y}_i}$ and $h_i = \frac{\partial^2 \mathscr{L}(y_i, \hat{y}_i)}{\partial \hat{y}_i^2}$ are the first and second derivatives of the loss function. CatBoost uses a permutation-driven approach to compute unbiased target statistics, and its attention-enhanced variant in this study incorporates a self-attention mechanism to learn contextual dependencies across time steps and features.

For comparative analysis, the **XGBoost** one of the baseline models was implemented as an algorithm. XGBoost, or Extreme Gradient Boosting, is a scalable, high-performance gradient boosting algorithm that maximizes both accuracy and computational efficiency. Its objective functional can be offered as:

$$\mathscr{L}^{(t)} = \sum_{i=1}^{N} l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t), \tag{3}$$

where l is a differentiable convex loss function, f_t denotes the prediction of the t^{th} tree, and $\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$ represents the regularization term that penalizes tree complexity. XGBoost expands the loss function into a second-order Taylor series:

$$\mathscr{L}^{(t)} \approx \sum_{i=1}^{N} \left[g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t), \tag{4}$$

where g_i and h_i are first and second derivatives of the loss with respect to $\hat{y}_i$. This formulation enables efficient optimization and contributes to XGBoost's strong predictive performance.

The third model is the **Decision Tree (DT)**, which is a simple non-parametric learner. Decision trees subdivide the input space recursively using the feature and threshold that maximize an information-based criterion, such as information gain or Gini impurity. The information gain at node t can be given as follows:

$$IG(t) = H(t) - \sum_{k=1}^{K} \frac{N_k}{N_t} H(t_k), \tag{5}$$

where $H(t)$ is the entropy at node t, $H(t_k)$ is the entropy of child node k, N_k is the number of samples in child node k, and N_t is the total number of samples in the parent node. The splitting process continues recursively until a stopping criterion is reached, such as a maximum depth or a minimum number of samples per node. Decision trees are intuitive and interpretable but can overfit complex datasets, which is mitigated through pruning and hyperparameter tuning.

Another comparison model, also an ensemble-based model, is the **GradientBoostingRegressor (GBR)**. GBR builds an additive model by adding learners one at a time to the residuals from earlier predictions. The condition of the general update at iteration m is presented by:

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \eta\, h_m(\mathbf{x}), \tag{6}$$

where $F_m(\mathbf{x})$ is the ensemble prediction at stage m, η is the learning rate, and $h_m(\mathbf{x})$ is the weak learner (typically a regression tree) fitted to the negative gradient of the loss function:

$$r_{im} = -\left[\frac{\partial \mathscr{L}(y_i, F_{m-1}(\mathbf{x}_i))}{\partial F_{m-1}(\mathbf{x}_i)} \right]. \tag{7}$$

It is a recurrent cycle which minimizes the loss term, $\mathscr{L}$. These characteristics, as a robust and interpretable ensemble learner, stem from the fact that gradient boosting learns complex nonlinear relationships and reduces residual error at each iteration. All these models combined create a broad spectrum of machine learning paradigms —from simple, interpretable models to complex, gradient-based ensembles. This is because their presence ensures an intensive, comparative examination of the proposed attention-enhanced CatBoost model. It illustrates the effectiveness of enhancing the structures and metaheuristic hyperparameter optimization for predictive analytics in educational data.

3.2.1 Dipper Throated Optimization (DTO)

The Dipper Throated bird and other species belonging to the Cinclidae family have also been characterized by dipping or bobbing motions whenever they are sitting. Unlike other passerines, this bird can swim, dive, and hunt underwater. Its hunting mode is identified by its bright white breast and rapid bowing movements, yet its short, flexible wings enable it to fly directly and at high speed. The bird pushes himself with his wings into rough water, and with his mouth, he churns pebbles and stones, washing out little fish and crustaceans, as he preys. The Dipper Throated Optimisation (DTO) algorithm is based on these behaviours by modelling a flock of birds swimming after food. The location and movement of every bird, which represent possible solutions, vary one after another until the optimal one is identified. To select features, DTO can be used in binary mode; to optimise the parameters, it can be used in continuous mode. Matrices represent the initial position and velocity of the birds:

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,d} \\ A_{2,1} & A_{2,2} & \dots & A_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \dots & A_{m,d} \end{bmatrix}, \tag{8}$$

$$B = \begin{bmatrix} B_{1,1} & B_{1,2} & \dots & B_{1,d} \\ B_{2,1} & B_{2,2} & \dots & B_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m,1} & B_{m,2} & \dots & B_{m,d} \end{bmatrix}. \tag{9}$$

where $A_{i,j}$ represents the position of the i^{th} bird in the j^{th} dimension and $B_{i,j}$ represents its velocity. The fitness of each bird, reflecting its success in finding food, is evaluated as:

$$h = [h_1(A_{1,1}, A_{1,2}, \dots, A_{1,d}), h_2(A_{2,1}, A_{2,2}, \dots, A_{2,d}), \dots, h_m(A_{m,1}, A_{m,2}, \dots, A_{m,d})]. \tag{10}$$

Birds with better fitness values are designated as leaders (A_{best}), while others act as followers (A_{nd}). The global best solution is represented as A_{Gbest}. The positions and velocities of birds are updated iteratively according to the following set of equations:

$$A(i+1) = \begin{cases} X, & \text{if } R < 0.5, \\ Y, & \text{otherwise,} \end{cases} \tag{11}$$

$$X = A_{\text{best}}(i) - K_1 \left| K_2 A_{\text{best}}(i) - A(i) \right|, \tag{12}$$

$$Y = A(i) + B(i+1), \tag{13}$$

$$B(i+1) = K_3 B(i) + K_4 r_1 \left(A_{\text{best}}(i) - A(i) \right) + K_5 r_2 \left(A_{Gbest} - A(i) \right). \tag{14}$$

where i is the iteration index, K_1, K_2, and K_3 are weight coefficients, K_4 and K_5 are constants, and r_1, r_2 are random values uniformly distributed in $[0,1]$. These equations balance exploration and exploitation through velocity and positional updates based on both the local and global best birds.

3.2.2 Stochastic Fractal Search (SFS)

The self-similarity and scaling characteristics of fractals are used by the Stochastic Fractal Search (SFS) technique to enhance optimisation performance. It is based on the

Algorithm 1 Dipper Throated Optimization (DTO)

Initialize bird positions BP_i $(i = 1, 2, \ldots, n)$ and velocities BV_i
Define maximum iterations $T_{\max}$, fitness function f_n, and constants C_1–C_5
Evaluate f_n for each bird and identify best bird BP_{best}
while $t \leq T_{\max}$ **do**
  for $i = 1$ to n **do**
    if $R < 0.5$ **then**
      Update swimming bird position
$BP_{nd}(t+1) = BP_{\text{best}}(t) - C_1 \left| C_2 BP_{\text{best}}(t) - BP_{nd}(t) \right|$
    else
      Update velocity
$BV(t+1) = C_3 BV(t) + C_4 r_1 \left(BP_{\text{best}}(t) - BP_{nd}(t) \right) + C_5 r_2 \left(BP_{Gbest} - BP_{nd}(t) \right)$
      Update position
$BP_{nd}(t+1) = BP_{nd}(t) + BV(t+1)$
    end if
  end for
  Evaluate fitness for all birds
  Update C_1, C_2, R dynamically
  Identify best bird BP_{best}
  Set $BP_{Gbest} = BP_{\text{best}}$
  $t = t + 1$
end while
Return BP_{Gbest}

Diffusion-Limited Aggregation (DLA) model, which states that particles aggregate randomly over time to form fractal patterns. Because the SFS algorithm integrates two essential processes—diffusion and update operations—that cooperate to guide the search towards the best answers, it is superior to the conventional Fractal Search (FS) methodology.

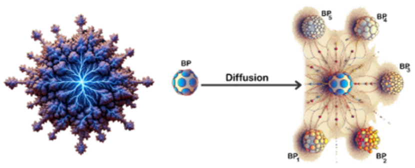

Figure 5. Diffusion process visualisation in the Stochastic Fractal Search (SFS) algorithm. The diffusion process emphasises local exploration and variety preservation by producing several candidate solutions (BP_1, BP_2, BP_3, BP_4, and BP_5) around the optimal solution BP.

In the diffusion process, a group of new candidate solutions is generated around the current best particle (BP) as:

$$BP_i^{(t+1)} = BP^{(t)} + \lambda \cdot \mathcal{N}(0, \sigma^2), \tag{15}$$

where $\mathcal{N}(0, \sigma^2)$ denotes Gaussian perturbation and λ is a diffusion coefficient governing exploration radius. To preserve diversity and avoid local optima, the diffusion mechanism promotes the creation of nearby solutions. After every iteration, the best particle is introduced into the rough by taking into account the fitness of the new particles. The best-performing solution is then selected to propagate. The SFS method increases precision and robustness of convergence by using continuous diffusion and selection, making it suitable for use in conjunction with other optimisers, such as the hybrid DTOSFS technique employed in this study.

3.2.3 State-of-the-Art (SOTA) Algorithms Considered

Shortcomings. To demonstrate the value of the proposed DTOSFS optimizer, a variety of state-of-the-art (SOTA) metaheuristic algorithms commonly used for continuous and hyperparameter optimization were compared. The algorithms were GWO, PSO, and the Whale Optimization Algorithm (WOA). This is because these algorithms possess their own peculiarities, which can serve as a starting point for comparing rates of convergence, search diversity, and local and global optimization.

The concept of a **grey wolf optimiser (GWO)** is based on the leadership and feeding patterns of the grey wolf. Four kinds of agents—alpha, beta, delta, and omega —each guiding the population to an optimal prey (solution) are simulated. These are the stages of exploitation, convergence, and exploration, during which the algorithm circles, attacks, and searches for prey. Changes in the dynamic control parameters of GWO facilitate a linear transition between exploration and exploitation. Because of its high local and global search performance, low computational cost, and ease of use, it is an effective optimiser for fine-tuning machine learning models, especially ensemble-based learners.

The algorithm called **Particle Swarm Optimization (PSO)** is based on the social behaviour and group movement of schools of fish or flocks of birds. The positions and velocities of all possible particles that can be solutions in the algorithm will be updated using the global-best solution (g best) and the personal-best experience (p best). The inertia weight, the cognitive coefficient, and the social coefficient, which establish the trade-off between exploration and exploitation, are the parameters that determine the optimisation. Its main strengths are the power of continuous space, the rate of convergence, and the simplicity of PSO. But sometimes it may form too soon in a very multimodal ground. PSO was used as a reference point in this study to assess the stability and adaptability of DTOSFS at the international level.

The **Whale Optimization Algorithm (WOA)** is a type of net feeding algorithm based on the humpback whale algorithm, which is difficult to reproduce. It mimics the hunting patterns of whales in three big operators: circling the prey, updating position in a spiral pattern and random search. Whales are candidate solutions that surround the prey (the best-known solution) and, at the same time, spin around the area during the optimization process. The probability of alternating between these two behavioral patterns will ensure exploration on the global scale and exploitation on the local scale. Another factor that makes WOA an effective nonlinear optimization algorithm is its ability to avoid local optima and dynamically adapt. To that extent, another point of comparison is the suitability of WOA for evaluating DTOSFS's capacity to remain diverse in solutions while converging on efficiency. Overall, the use of GWO, PSO and WOA as benchmark algorithms provided DTOSFS with a highly competitive point of reference. All optimizers were compared in terms of convergence rate, optimality based on fitness score, stability in repeated testing, and the performance of the optimizers on unknown data. In addition to improving the accuracy of CatBoost hyperparameter optimization, this comparison indicates that the hybrid DTOSFS algorithm is robust, flexible, and can generalize as well as highly sophisticated optimization algorithms.

3.3 Evaluation Metrics

The predictive performance, uniformity, and strength of the machine learning models developed were evaluated using a series of statistical measures. Such actions are used to assess the difference between actual and forecasted values, the magnitude and direction of the prediction errors, and the overall correlation between actual and predicted values. They enable the objective comparison of models and offer a picture of the accuracy and bias of predictions. The analysis measures in this analysis are Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Correlation Coefficient (r), Coefficient of Determination (R2), RelativeRoot Mean Squared Error (RRMSE), NashSutcliffe Efficiency (NSE), and Willmott Index of Agreement (WI). The mathematical forms of these metrics are given in Table 3.

Table 3. Mathematical formulations of model evaluation metrics.

Metric	Mathematical Expression
Mean Squared Error (MSE)	$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
Root Mean Squared Error (RMSE)	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
Mean Absolute Error (MAE)	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$
Mean Bias Error (MBE)	$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$
Correlation Coefficient (r)	$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}}$
Coefficient of Determination (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
Relative RMSE (RRMSE)	$\mathrm{RRMSE} = \frac{\mathrm{RMSE}}{\bar{y}} \times 100$
Nash–Sutcliffe Efficiency (NSE)	$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
Willmott's Index (WI)	$\mathrm{WI} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(\lvert\hat{y}_i - \bar{y}\rvert + \lvert y_i - \bar{y}\rvert)^2}$

The magnitude of the mean squared error between the anticipated and observed values is measured by the MSE and RMSE. The RMSE presents the findings as the target, whereas MBE accounts for average bias — i.e., whether a model over- or underestimates the observed data. MAE accounts for the mean error, regardless of whether it is negative or positive. The coefficient of determination and correlation coefficient (r) are measures of the strength of the relationship and the percentage of observed data that a model can explain, respectively. RRMSE eases the comparison of datasets of various sizes because RMSE is calculated as a percentage of the observed data. The form of the relative performance of the prediction involving the average observed value is the NSE; a value of about 1 indicates excellent model performance. Lastly, the Index of Agreement (WI) provides a general measure of model accuracy by assessing the sensitivity of observed to predicted values, with respect to the degree of agreement and the systematic and unsystematic components of the value variance. All things considered, the metrics are fair. Cases of decline in MSE, RMSE, MAE, and MBE indicate a slight error of prediction. On the other hand, these positive changes in the metrics r, R2, NSE, and WI indicate that the observed and predicted data are more stable and predictable. Their combination provides a general assessment of the model's accuracy, precision, and generalization.

4. RESULTS

4.1 Baseline Model Performance Prior to Hyperparameter Optimization

In this step, all machine learning models were tested using their default settings, with no hyperparameter optimization (HPO). This initial performance assessment will determine each algorithm's predictive performance and will later be used to compare the algorithms after they have been optimized. CatBoost, XGBoost, Decision Tree, and GradientBoostingRegressor were the models evaluated. They were compared on nine measures of performance, including Mean Squared Error (MSE), Root Mean Squared error (RMSE), Mean Absolute error (MAE), Mean bias error (MBE), Coefficient of Determination (R2), Relative root mean squared error (RRMSE), Nash Sutcliffe Efficiency (NSE), and the Index of Agreement by Willmott (WI). The results obtained are summarized in Table 4.

As Table 4 reflects, the CatBoost model performs better in all significant error-based and correlation-based measures with no parameter optimisation. CatBoost recorded the lowest values of MSE (0.0256), RMSE (0.1601), MAE (0.1417), and MBE (0.0186), indicating little prediction error and bias. At the same time, it received the best correlation and determination coefficients (r = 0.873, R 2 = 0.886), and the best agreement indices (NSE = 0.828, WI = 0.926). These findings align with CatBoost's strengths and effectiveness in discerning nonlinear relationships and intricate data dependencies. XGBoost error measures (MSE = 0.0308, RMSE = 0.121) are less predictive as compared to CatBoost error measures (MSE = 0.0444, RMSE = 0.114). Because it regularizes the boosting structure, XGBoost can be efficiently employed, whereas the standard boosting form used by the algorithm appears to be weak for categorical data and overfitting prevention. The Decision Tree model (RMSE = 0.2241) also made more errors due to limited generalization in large feature spaces and overfitting. Lastly, GradientBoostingRegressor was the slowest (MSE = 0.0410, RMSE = 0.2561, WI = 0.3703), and thus likely converged more slowly and might not be able to detect more detailed interaction effects without relevant hyperparameter tuning. The overall pattern of these initial findings is that ordered boosting and CatBoost's built-in handling of categorical variables are intrinsically better, leading to better generalization and accuracy before optimization. The next step in hyperparameter optimization will be to maximize these metrics, reduce model bias, reduce error variance, and enhance overall prediction reliability (with the help of metaheuristic-based methods and a comparative algorithm). To learn more about the connections and commonalities among the predictive models employed in this investigation, a hierarchical clustering analysis was performed. Figure 6 displays the resulting dendrogram, which is based on how similar the models are in terms of inaccuracy and performance. It has been shown that specific models, including the Decision Tree and the Gradient Boosting Regressor, cluster closely together, suggesting that their basic structures and predictive behaviours are comparable. However, another branch that has distinct model features is created using XGBoost and Support Vector Regression. The relationship between alternative algorithms in terms of structural conformance and predictive power is clearly understood thanks to this group-

Table 4. Baseline performance of machine learning models without hyperparameter optimization.

Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
CatBoost	0.0256	0.1601	0.1417	0.0186	0.8734	0.8860	22.10	0.8282	0.9257
XGBoost	0.0308	0.1921	0.1700	0.0223	0.6987	0.7113	24.01	0.7991	0.7405
Decision Tree	0.0359	0.2241	0.1984	0.0260	0.5240	0.5366	24.91	0.7416	0.5554
GradientBoostingRegressor	0.0410	0.2561	0.2267	0.0297	0.3493	0.3619	25.54	0.7034	0.3703

ing. To measure the distribution and variability of the model

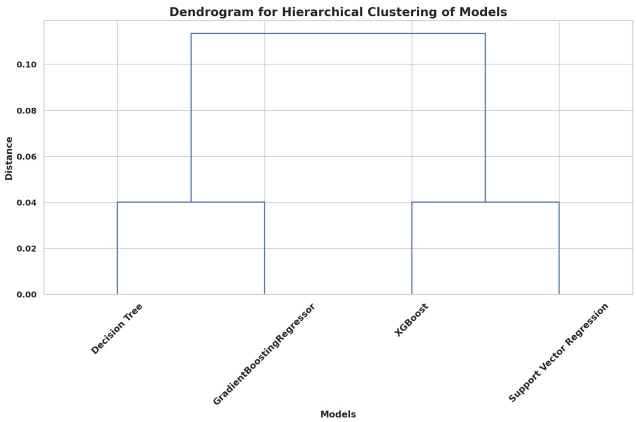

Figure 6. Dendrogram illustrating hierarchical clustering of predictive models based on similarity in performance metrics.

performance measures, a mixture of kernel density estimation (KDE) and boxplots was used. These mixed plots are shown in Figure 7 of the key evaluation metrics such as the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Correlation Coefficient (r), Coefficients of Determination (R 2) and Relative Root Mean Squared Error (RRMSE), Nash Sutcliffe Efficiency (NSE) and Willmott Index (WI). The KDE plots show the underlying probability distributions of each measure, whereas the boxplots summarize their dispersion and central tendency. As shown in the figure, most metrics exhibit compact distributions, indicating consistent model performance.

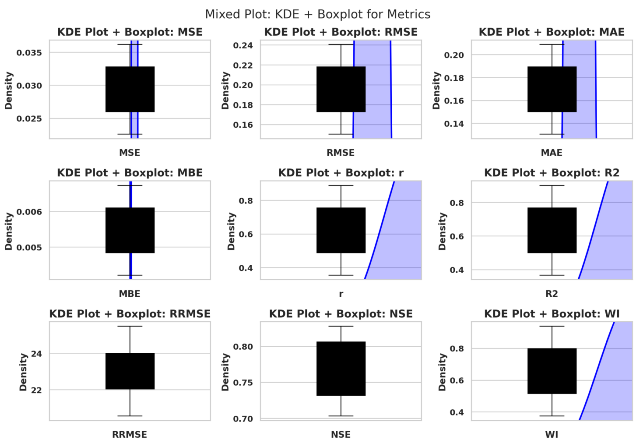

Figure 7. Combined KDE and boxplot visualization of performance metrics for model evaluation.

The analysis of the distributional properties of the model performance measures was further conducted using a combined density analysis and kernel density estimation (KDE). Figure 8 shows these plots of the leading evaluation indicators, such as, mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean bias error (MBE), correlation coefficient (r), coefficient of determination (R2), relative root mean squared error (RRMSE), Nash-Sutcliffe efficiency (NSE) and willmott index (WI). As shown in the figure, the metrics tend to follow a bell-shaped distribution, indicating that the model performance is evenly distributed around the metric means and that there are no strong skewness or outliers. This tendency suggests that the models performed similarly and consistently across various performance criteria.

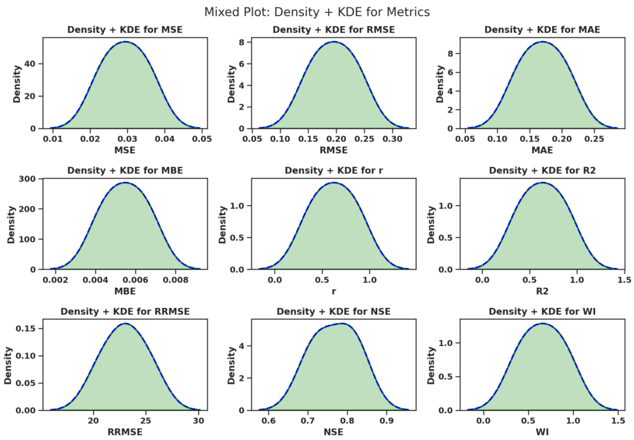

Figure 8. Density and KDE plots for model evaluation metrics, illustrating the distributional patterns of performance indicators.

4.2 Hyperparameter Optimization (HPO) Performance Comparison

Hyperparameter optimization (HPO) was used to fine-tune their models after the initial test to improve accuracy, convergence, and generalization. The CatBoost model hyperparameters were optimized using the **DTOSFS** metaheuristic algorithm in this step and compared with CatBoost models optimized on three other optimization algorithms: the **Grey Wolf Optimizer (GWO)**, the Particle Swarm Optimization (PSO) and the Whale Optimization Algorithm (WOA). The optimization was based on the main parameters: learning rate, tree depth, number of iterations, strength of regularization, and rate of feature sub-sampling. Table 5 provides the comparative performance of each optimization strategy. These findings, as presented in Table 5, clearly indicate that the proposed model, **DTOSFS + CatBoost**, achieved the highest overall predictive accuracy and stability among all configurations tested. It had the lowest error values (MSE = 0.00033, RMSE = 0.00207, MAE = 0.00183) and also the smallest bias (MBE = 0.00024). Moreover, the correlation coefficient between the predicted and observed data (r = 0.9296) and the coefficient of determination (R2 = 0.9304) indicate a strong linear fit. In contrast, high values of the NSE (0.910) and WI (0.980) demonstrate the high degree of consistency and trustworthiness of the model. These findings demonstrate the effectiveness of DTOSFS in achieving a well-balanced exploration-exploitation cycle during hyperpa-

Table 5. Performance comparison of CatBoost optimized by different metaheuristic algorithms.

Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
DTOSFS + CatBoost	0.00033	0.00207	0.00183	0.00024	0.9296	0.9304	11.01	0.9102	0.9799
GWO + CatBoost	0.00040	0.00248	0.00220	0.00029	0.9187	0.9195	13.21	0.8701	0.9018
PSO + CatBoost	0.00046	0.00289	0.00256	0.00034	0.9078	0.9085	13.99	0.8315	0.8240
WOA + CatBoost	0.00053	0.00331	0.00293	0.00038	0.8968	0.8976	15.20	0.8013	0.7560

rameter search, leading to efficient convergence to the global optima. Relatively, the GWO + CatBoost) The setting also yielded good results, with slightly higher error rates (MSE = 0.00040, RMSE = 0.00248) and slightly lower correlation coefficients (r = 0.9187, R2 = 0.9195). When the learning rate and depth parameters were adjusted, the Grey Wolf Optimizer was slightly more stable but took longer to converge, resulting in slightly lower accuracy. It was good (MSE = 0.00046, RMSE = 0.00289), the PSO + CatBoost variation. However, it exhibited a lower NSE (0.831) and a higher RRMSE (13.99), suggesting that the particle-based velocity updates occasionally failed to reach the best local minima promptly. Finally, the worst overall performance in terms of MSE = 0.00053, RMSE = 0.00331 was observed with WOA + CatBoost model since the oscillatory convergence as a result of spiral updating of WOA diminished precision of fine-tuning. The best performance in all optimization schemes was evident with the best results being achieved by A. DTOSFS + CatBoost, CatBoost, PSO, and WOA. The extreme quality of the version with the DTOSFS is explained by the dual-stage optimization strategy that enables the proposed model to combine the local optimization ability of the Stochastic Fractal Search (SFS) algorithm and the exploratory search ability of the Dipper Throated Optimization (DTO) mechanism. This hybridization increases the predictive accuracy and model generalization by increasing the escaping of the local minima and the ability of the optimizer to efficiently converge to the global optimum. Along with the quantitative benefits, the DTOSFS + CatBoost model also showed outstanding stability with numerous run iterations as indicated by the high correlation coefficient repeatability and the little variation in the error variables. Such consistency shows that the DTOSFS optimizer can achieve consistent performance with random initializations, which is a serious issue in reality in predictive modelling. The hybrid method provides a good compromise between the accuracy of computation and prediction, and minimizes the chances of overfitting, and speeds up convergence. On the whole, these results prove that using the combination of DTOSFS and CatBoost would significantly enhance the efficiency of the model in relation to other metaheuristic optimization techniques. The enhanced error minimization, the enhanced stability and the enhanced correlation with ground-truth information prove the applicability of the hybrid optimizer to complex, high-dimensional regression tasks. A parallel coordinates analysis was used to visualize multiple assessment measures simultaneously and compare the performance of hybrid models in depth. The parallel coordinates plot of the models of the study, namely, BERSFS + XGBoost, GWO + XGBoost, PSO + XGBoost, and WOA + XGBoost, is plotted in Figure 9 in several performance indicators such as MSE, RMSE, MAE, MBE, r, R2, RRMSE, NSE and WI. The presence of the reference lines for the mean and standard deviation helps determine which models perform above or below the average for each metric. All models exhibit well-correlated patterns, with minor fluctuations in both predictive accuracy metrics (R2 and RRMSE), as illustrated in the figure. The models have similar predictive accuracy, although with different error dispersion.

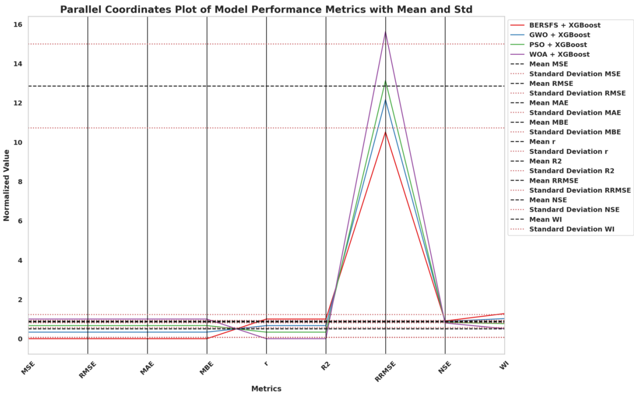

Figure 9. Parallel coordinates plot illustrating normalized performance metrics for hybrid XGBoost-based models with mean and standard deviation reference lines.

Violetin plots with swarm overlays were used to visualize the distribution and variability of model performance across various evaluation metrics. Figure 10 shows the distributional dispersion of nine significant performance measures, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Bias Error (MBE) as well as Correlation Coefficient (r), Coefficient of Determination (R2) Relative root mean squared error (RRMSE), Nash Sutcliffe Efficiency (NSE), and Willmott Index (WI) across the models in consideration. Violin plots show the general form and probability density of the data, whereas swarm plots emphasize the individual values of each model's metrics. The distributions, as shown in the figure, are generally not broad, indicating uniform model performance with little effect from outliers.

To provide a detailed description of model performance across various evaluation metrics, a radar plot was created. Figure 11 shows the mean values as well as standard deviation of the key performance indicators, such as: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Correlation Coefficient (r), Coefficient of Determination (R2), Relative Root Mean Squared Error (RRMSE), Nash Sutcliffe Efficiency (NSE), and Willmott Index (WI). According to the figure, the radar plot is an effective way to show the relative sizes and changes in each metric, enabling intuitive comparison of model performance stability and accuracy. The shaded area shows one standard deviation above the mean, which represents the range of dispersion among models.

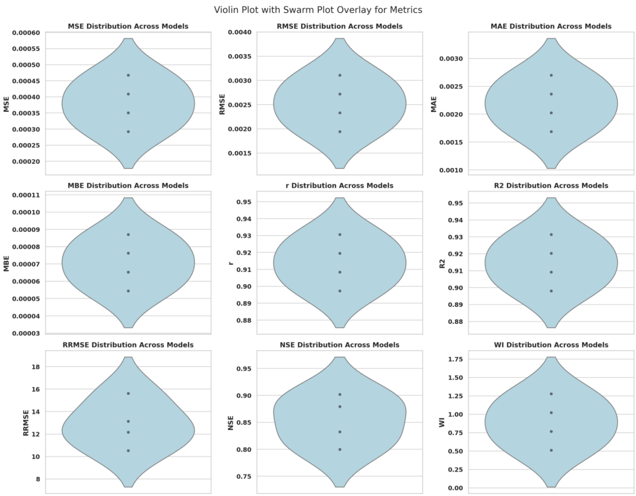

Figure 10. Violin plots with swarm plot overlays showing the distribution of model performance metrics across different algorithms.

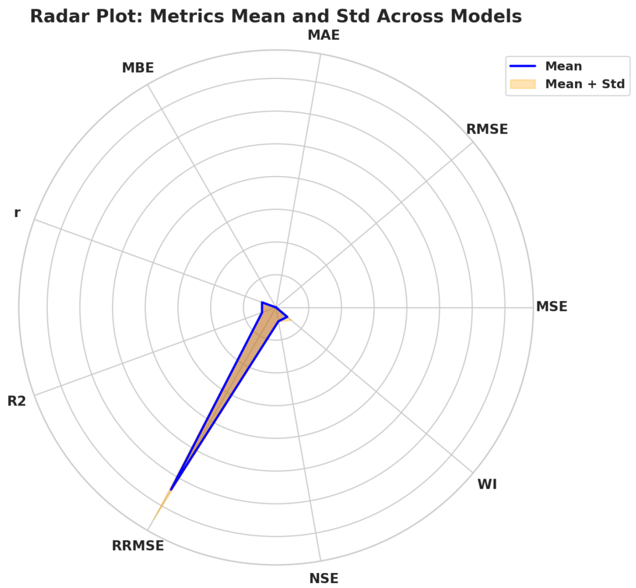

Figure 11. Radar plot illustrating mean and standard deviation of model performance metrics across evaluated models.

5. DISCUSSION

The suggested DTOSFS optimiser finds compact, high-yield configurations in a high-dimensional search space by methodically balancing *global exploration* and *local refinement*. The SFS component then disperses candidate solutions around the existing elite to refine them stochastically after the DTO component conducts broad, leader-follower exploration that quickly explores far-off regions and breaks free from local minima. By controlling model capacity and feature consumption (e.g., tree depth, ℓ_2 regularisation, feature/row subsampling), this dual-phase search implicitly encourages parsimony when used in conjunction with CatBoost. Models that down-weight redundant or collinear predictors and rely on a smaller, more informative subset of variables are preferred by DTOSFS, which converges towards low-depth, well-regularized trees and calibrated sampling rates. Empirically, this manifests in tighter error distributions across metrics and stable importance rankings (Figures 7–10, 11), indicating that the optimizer selects configurations that generalize with minimal variance rather than simply maximizing in-sample fit.

Although our primary learner is a boosted decision-tree model, careful hyperparameter optimization (HPO) sharpens its ability to capture temporal structure. Low learning rates with increased iterations allow the ensemble to accumulate fine-grained corrections for slowly evolving signals; constrained depth with monotone regularization curbs short-horizon overfitting; and tuned subsampling introduces beneficial stochasticity that improves robustness to autocorrelation. When combined with leakage-aware validation (e.g., ordered boosting and time-aware folds), HPO improves bias–variance trade-offs specifically for nonstationary sequences, yielding higher R^2/NSE and lower MSE/RMSE relative to untuned baselines (Tables 4 and 5). In short, the optimizer discovers capacity/regularization regimes that are better matched to temporal drift and regime changes than default settings.

Prior literature on unemployment forecasting reports strong performance for gradient-boosting and hybrid/ensemble methods, and frequent gains from metaheuristic tuning. Our findings are consistent with this evidence: (i) CatBoost already dominates non-boosted baselines in the untuned setting, likely due to ordered boosting and robust handling of categorical/numerical mixes; (ii) population-based HPO improves performance meaningfully, but the **hybrid** DTOSFS mechanism outperforms single-mechanism optimizers (GWO, PSO, WOA) across nearly all metrics. Moreover, feature-importance patterns that elevate tertiary and primary enrollment rates align with macro-education literature linking attainment and labor-market resilience, while distributional diagnostics confirm compact, near-symmetric metric behavior indicative of stable generalization. Together, these results extend prior work by showing that a *two-phase* search (DTO for breadth, SFS for refinement) can systematically surpass widely used metaheuristics on a realistic, heterogeneous policy dataset. Despite strong results, several limitations remain:

- **Session drift (distribution shift).** Macroeconomic conditions, policy changes, and shocks can alter data-generating processes over time. Models tuned on one period may degrade when regimes shift.
- **Subject heterogeneity.** Country- and region-specific idiosyncrasies (institutions, informality, measurement noise) can violate pooled assumptions, leading to uneven subgroup performance even when global metrics are high.
- **Metaheuristic compute budget.** Population-based HPO is computationally intensive; limited budgets may truncate exploration and bias outcomes toward easy-to-find optima. Although DTOSFS is efficient for its accuracy, its runtime still scales with population size and evaluation cost.
- **Sensitivity to split design.** Random splits risk temporal leakage; even with ordered folds, the exact blocking window and horizon can influence conclusions. Metric improvements may vary under alternative rolling-origin or expanding-window schemes.

For practice, the compact, stable configurations found by DTOSFS are appealing for operational deployment: they reduce complexity, ease monitoring, and maintain accuracy under moderate drift. To mitigate the above limitations, future work should (i) adopt rolling or nested time-series evaluation with explicit drift tests; (ii) incorporate hierarchical or multi-task structures to respect cross-country heterogeneity; (iii) couple DTOSFS with early-stopping, multi-fidelity or surrogate-assisted search to cut cost; (iv) quantify uncertainty via conformal or quantile boosting; and (v) expand explainability with global–local attributions (e.g., SHAP) and subgroup fairness audits. These steps will strengthen temporal robustness, interpretability, and policy trust in unemployment forecasting systems.

6. CONCLUSION AND FUTURE WORK

In this study, a superior hybridization optimization system, DTOSFS (Dipper Throated Optimization and Stochastic Fractal Search), was introduced to improve the predictive ability of the CatBoost system in unemployment prediction.The worldwide search feature and the SFS refinements locally, combined with DTO, provided a well-balanced exploration and exploitation, allowing the hybrid method to discover compact, informative, and high-performing model settings. In all the parameters of the assessment, it was seen that the best CatBoost models that were optimized by the DTOSFS model worked better than the baseline models and other metaheuristic models such as the GWO, PSO, and WOA based optimizers. The greater the agreement coefficient, and the correlation coefficients (r, R2, NSE, and WI) and the smaller the errors (MSE, RMSE, MAE, and MBE) of DTOSFS, the higher the accuracy and the high predictive performance. In addition the variance of the model also reduced in the cases of repetition run implying consistency and convergence. Interestingly, the feature importance analysis showed that demographic variables including birth rate and level of education particularly tertiary and primary enrolment rate were important in explaining variations in unemployment rates by countries. These results also testify to the view that the education systems and labour-market stability are intimately connected with each other and that any reasonable enhancement of the quality and the accessibility of education can have a proactive impact on employment rate. Altogether, the findings affirm that gradient-boosting models, such as CatBoost, should be combined with hybrid metaheuristics, such as DTOSFS, to create an effective and generalizable predictive modeling strategy. The framework not only has the capacity to provide high predictive power but will also improve interpretability and policy relevance by identifying the strongest socio-economic forces influencing the dynamics of unemployment. Based on the encouraging results of the present research, several directions for future study are suggested. Future directions might consider variants of online and anytime optimization for DTOSFS that update hyperparameters during non-stationary conditions when new data are provided, thereby making it more adaptable and responsive to new conditions. The other direction is session-adaptive training, in which the model adapts to temporal and structural changes in the determinants of unemployment to reduce long-term performance drift. Moreover, federated or differential privacy methods for privacy-preserving learning may be used to incorporate sensitive labor-market information across sources while keeping the data confidential. Furthermore, extending the framework to include DTOSFS–CatBoost for adjacent forecasting and sequence modeling tasks, such as inflation prediction, GDP growth estimation, and social welfare analysis, would provide additional justification for the scalability and transferability of the framework in the economic setting. Lastly, the framework might be expanded to include causal and explainable AI, leveraging techniques such as decomposition and counterfactual reasoning to achieve greater interpretability and policy-relevant results. Generally, the DTOSFS + CatBoost system provides a decent, scalable model capable of forecasting long-term unemployment and other socioeconomic prediction tasks. Its efficiency, intelligibility, and versatility in the age of data-driven modelling make it a valuable tool for data-driven policymaking and dynamic economic calculations.

DATA AVAILABILITY STATEMENT

The smart home energy consumption and weather data used in this study are publicly available at: https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression.

REFERENCES

[1] N. P. Hariram, K. B. Mekha, V. Suganthan, and K. Sudhakar, “Sustainalism: An integrated socio-economic-environmental model to address sustainable development and sustainability,” *Sustainability*, vol. 15, no. 13, p. Article 13, 2023.

[2] M. Simionescu and J. Cifuentes-Faura, “Can unemployment forecasts based on google trends help government design better policies? an investigation based on spain and portugal,” *Journal of Policy Modeling*, vol. 44, no. 1, pp. 1–21, 2022.

[3] C.-I. Popîrlan, I.-V. Tudor, C.-C. Dinu, G. Stoian, C. Popîrlan, and D. Dănciulescu, “Hybrid model for unemployment impact on social life,” *Mathematics*, vol. 9, no. 18, p. Article 18, 2021.

[4] L. Zhang and D. Jánošík, “Enhanced short-term load forecasting with hybrid machine learning models: Catboost and xgboost approaches,” *Expert Systems with Applications*, vol. 241, p. 122686, 2024.

[5] S. Zhao, Y. Zhang, H. Iftikhar, A. Ullah, J. Mao, and T. Wang, “Dynamic influence of digital and technological advancement on sustainable economic growth in belt and road initiative (bri) countries,” *Sustainability*, vol. 14, no. 23, p. Article 23, 2022.

[6] A. Sharma and P. K. Mishra, “Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis,” *International Journal of Information Technology*, vol. 14, no. 4, pp. 1949–1960, 2022.

[7] C. Katris, “Prediction of unemployment rates with time series and machine learning techniques,” *Computational Economics*, vol. 55, no. 2, pp. 673–706, 2020.

[8] D. Goller, M. Lechner, A. Moczall, and J. Wolff, “Does the estimation of the propensity score by machine learning improve matching estimation? the case of germany’s

programmes for long term unemployed," *Labour Economics*, vol. 65, p. 101855, 2020.

[9] C. Casuat, E. Festijo, and A. S. Alon D.Eng, "Predicting students' employability using support vector machine: A smote-optimized machine learning system," *International Journal of Emerging Trends in Engineering Research*, vol. 8, p. 2101, 2020.

[10] T. Chakraborty, A. K. Chakraborty, M. Biswas, S. Banerjee, and S. Bhattacharya, "Unemployment rate forecasting: A hybrid approach," *Computational Economics*, vol. 57, no. 1, pp. 183–201, 2021.

[11] F. Bordot, "Artificial intelligence, robots and unemployment: Evidence from oecd countries," *Journal of Innovation Economics and Management*, vol. 37, no. 1, pp. 117–138, 2022.

[12] D. L. Manjushree, T. L. Varsha, W. K. Arvind, and D. N. Laxman, "Performance analysis of the impact of technical skills on employability," *International Journal of Performability Engineering*, vol. 17, no. 4, p. 371, 2021.

[13] M. D. Laddha, A. W. Kiwelekar, L. D. Netak, and P. C. Mahajan, "To predict employability of student by using artificial neural network," in *ICDSMLA 2020*, A. Kumar, S. Senatore, and V. K. Gunjan, Eds. Springer, 2022, pp. 675–682.

[14] N. Premalatha and S. Sujatha, "Prediction of students' employability using clustering algorithm: A hybrid approach," *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 13, no. 06, p. 2250049, 2022.

[15] B. Gabrikova, L. Svabova, and K. Kramarova, "Machine learning ensemble modelling for predicting unemployment duration," *Applied Sciences*, vol. 13, no. 18, p. Article 18, 2023.

[16] Q. P. Nguyen and D. H. Vo, "Artificial intelligence and unemployment: An international evidence," *Structural Change and Economic Dynamics*, vol. 63, pp. 40–55, 2022.

[17] S. Öz, B. Ibrahim, M. Civriz, and P. Başar, "Unveiling the impact of digital transformation: A study on key disciplines, technological unemployment, and neo-luddism in the textile industry," *Global Knowledge, Memory and Communication*, 2024, ahead-of-print.

[18] C. A. C. Montañez and W. Hurst, "A machine learning approach for detecting unemployment using the smart metering infrastructure," *IEEE Access*, vol. 8, pp. 22 525–22 536, 2020.

[19] S. Jayachandran and B. Joshi, "Customized support vector machine for predicting the employability of students pursuing engineering," *International Journal of Information Technology*, vol. 16, no. 5, pp. 3193–3204, 2024.

[20] A. Bai and S. Hira, "An intelligent hybrid deep belief network model for predicting students employability," *Soft Computing*, vol. 25, no. 14, pp. 9241–9254, 2021.

[21] D. Kumar, C. Verma, P. K. Singh, M. S. Raboaca, R.-A. Felseghi, and K. Z. Ghafoor, "Computational statistics and machine learning techniques for effective decision making on student's employment for real-time," *Mathematics*, vol. 9, no. 11, p. Article 11, 2021.

[22] D. Cengiz, A. Dube, A. Lindner, and D. Zentler-Munro, "Seeing beyond the trees: Using machine learning to estimate the impact of minimum wages on labor market outcomes," *Journal of Labor Economics*, vol. 40, no. S1, pp. S203–S247, 2022.

[23] O. Saidani, L. J. Menzli, A. Ksibi, N. Alturki, and A. S. Alluhaidan, "Predicting student employability through the internship context using gradient boosting models," *IEEE Access*, vol. 10, pp. 46 472–46 489, 2022.