Predictive Modeling of Global Educational Outcomes: A Comparative Analysis Using Machine Learning Regression Techniques

Abdelhameed Ibrahim [1] *, Abdelaziz A. Abdelhamid[2], Ehab M. Almetwally[3]
[1]Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt
[2]Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt
[3]Department of Mathematics and Statistics, Faculty of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia

Emails: afai79@mans.edu.eg, abdelaziz@cis.asu.edu.eg, emalmetwally@imamu.edu.sa

Abstract

Education contributes a crucial portion to the world's development; thus, it is crucial to focus on education enrollment and quality education. It is essential not only that children enroll in school but also that they receive proper education to improve individuals and, consequently, society. This paper aims to use machine learning to predict educational outcomes based on the World Educational Data obtained from Kaggle to analyze the data, preprocess it, and evaluate the performances of the different regression models. The following models consist of Support Vector Regression (SVR), CatBoost, RandomForestRegressor, ExtraTreesRegressor, XGBoost, MLPRegressor, GradientBoostingRegressor, DecisionTreeRegressor, KNeighborsRegressor, LinearRegression, and Pipeline. Evaluation measures used included MSE, RMSE, MAE, MBE, r, R2, NSE, and WI. Analyzing the performance comparison, the best accuracy was associated with CatBoost with an r value equal to 0.999996 and an R2 value of 0. 999993; The MSE score was 0.04024. The outcomes of the present paper demonstrate that the application of advanced machine learning algorithms can be used effectively to predict educational outcomes, thus enabling policymakers and educational planners to use them for designing effective educational policies and overcoming existing global challenges in the sphere of education.

Keywords: Educational Data Analysis, Regression Models, Machine Learning, Predictive Modeling, Global Education Outcomes

1 introduction

It is critical to recognize that educational quality is one of the primary factors that define a nation's social and economic development based on people's job opportunities, and it is based on the country's competitiveness with other nations on the global playing field. To begin with, when nations seek to improve their educational systems, the nature of influences that would affect quality output becomes critical to ascertain. Regression analysis, one of the most effective statistical methods, identifies the dependencies of various factors affecting the quality of education and helps make competent decisions based on which the policymakers and educators can enhance the results in education [1–3].

Regression analysis enables the researcher to find the correlation between a dependent variable, such as the quality of education, and one or many independent variables. Referring to the international comparative perspective, this technique distinguishes significant predictors that define the disparities in education quality between countries or even within regions. Such variables could include socio-demographic factors such as

https://doi.org/10.54216/JAIM.070206

Received: July 05, 2023 Revised: September 29, 2023 Accepted: March 02, 2024

parents' income and education levels, institutions that include teachers' qualifications, school finances, and educational policies. Regression analysis aids in determining the magnitude of these variables to determine how various aspects promote or hinder the quality of education so that the best strategies can be developed [4–6].

The research on the given topic will apply regression analysis to evaluate world educational quality. In the envisaged scholarly research, emphasis will be placed on assessing the effectiveness of different predictors on educational outcomes worldwide. This will involve a comprehensive set of variables, including the nation's expenditure on education, the number of students per class, the student's ratio and levels of education attained. This paper aims to identify the factors that fundamentally influence the quality of education and establish the proportionate values of those impacts using the models of regression. For instance, the elaboration may demonstrate that enhancements of education budgets or advancements in instructors' qualifications lead to better learner outcomes [7–9].

Also, the paper will discuss the possible repercussions of these findings for existing and would-be educational policies and learning practices worldwide. Knowledge of the determinants of educational quality could assist with policy changes and resource deployment. For instance, if the data reveals that the expertise of the school teachers is a decisive factor in the learning outcomes, then an attempt might be made to upgrade the training of the teachers and provide them with better learning opportunities. On the other hand, if it's discovered that essential variables present in Lewin's social equation are influential, like socioeconomic status, general educational policies that address disparities via funds or community will be deemed necessary [10–12].

Thus, regression analysis, as a method applied to analyzing world educational quality, could be proposed as a prospective for further investigation of global educational performance and enhancement of educational results. Hence, this research's goal is to outline factors that can impact education and their implications in such a way as to inspire global educational reforms and improvements. Hence, it aims to advance knowledge to enhance fairness and efficiency in education systems and the learning process for students in different environments.

2 Related Works

The use of various regression procedures to synthesize the quality of education in different nations has been critical in enhancing educational quality. This research area focuses on establishing relationships with various factors that affect educational systems worldwide, measuring the impact of these factors, and producing solutions through regression analyses.

Linear regression is one of the most conventional statistical techniques to assess educational quality indicators. Linear regression is considered advantageous in terms of its simplicity and generalisability of results, thus enabling researchers to put numerical values to the effect of different factors on education achievement [13]. For example, it is possible to apply linear regression to assess the impact of various factors, including teacher-student ratios, funding levels, and parents' involvement in the student's performance within various countries. Despite this, the linear regression models sometimes lack the predictability of non-linear relationships and interactions between the variables, which may hamper their applicability in many situations in the field of education.

Researchers have studied various regression models like polynomial and generalized linear models (GLMs) to overcome the above-mentioned issues. Polynomial regression extends linear regression by enabling the predictor and outcome variables to have a non-linear relationship [14]. This approach has been applied to represent high-order dependencies between quality factors like the role of educational resources or socioeconomic conditions on students' performance. Statistical models, such as logistic and Poisson regression, are presented under the Generalized Linear Models (GLMs) category, allowing for data distribution and categorizing and counting data associated with educational outcomes [15].

Methods based on machine learning have also been used to analyze educational quality, which offered more refined techniques for data pattern identification. For instance, decision trees and random forests have been used to conceive nonlinear relationships and interactions of the variables [16]. Indeed, decision trees give a transparent and interpretable model within which measures that affect educational quality, such as properties

https://doi.org/10.54216/JAIM.070206
Received: July 05, 2023 Revised: September 29, 2023 Accepted: March 02, 2024

for schools, teachers' training, and curriculum quality, among others, can be easily identified. In turn, random forests are part of the ensemble methods, which enhance the prediction and demonstrate the high outcome and reliability in the consumption of many decision trees and, therefore, are suitable for analyzing various and intricate datasets [17].

Another type of machine learning algorithm, Support Vector Machines (SVM), has been used to build educational quality as the connections between variables can be non-linear and invariant [18]. SVMs, hence, can work with many predictor variables and are pretty helpful in separating one level of educational quality from another. For example, SVM models have been applied to categorize countries according to educational quality features such as students' performance levels and graduation rates.

Over the last few years, deep learning models have become popular and efficient enough to analyze educational quality. Neural networks, mainly deep neural networks, have been used to capture complex patterns in educational data [19]. These models can independently learn the feature representation of the raw data, such as international assessments and surveys, and predict the educational quality outcomes with nearly perfect accuracy. Recent studies suggest deep learning methodologies as practical tools for pre-service teachers' performance prediction since such algorithms can discover subtle patterns and interactions in learning data that can go unnoticed in conventional regression techniques.

Other methods aim to use several models simultaneously, which would also fall under ensemble learning in analyzing the quality of education. Techniques like boosting and bagging improve the outcome of individual models by combining their results [20]. For instance, the boosting algorithms enhance the performance of regression models by decreasing the generalization errors made in previous models, and the bagging technique combines the results of several models to decrease the variance of an estimate.

Psychic and intellectual factor analysis, practical importance of factor analysis and principal component analysis have also been used to determine factors and dimensions of educational quality. These techniques assist in the decision-making process concerning the hidden factors that affect education and also aid in data simplification, where the measurements have numerous attributes [21]. Many researchers adopted factor analysis to originate and evaluate educational quality dimensions, infrastructures, teacher impacts, and support for students, where PCA simplifies the analysis to render the essential factors while at the same time preserving most of the variance inherent in the investigation.

Still, the integration of regression applications with machine learning has been tried to use both methods. These models are for better prediction of accurate results and to get better and more sophisticated understandings about educational quality [22]. For example, a combination of a linear regression technique to establish basic patterns and machine learning to enhance the prediction and localize with nonlinear patterns has been tried.

Finally, external reasons related to economic growth, alterations in legislation, and traditions have drawn attention as a factor contributing to evaluating the quality of education. Including these factors in regression models can improve the analysis of the quality of education worldwide and help in decision-making. Again, research reveals that factors like the Gross Domestic Product per capita and investment in education explain educational performance and that including them in models yields better results.

3 Methodology

3.1 Data Source and Description

3.1.1 Data Source

The dataset used in this study was obtained from Kaggle as the "World Educational Data". This dataset encompasses different aspects of education, including enrollment, literacy, teacher-student ratios, government expenditure on education, and various socio-economic factors that impact the population's education.

https://doi.org/10.54216/JAIM.070206
Received: July 05, 2023 Revised: September 29, 2023 Accepted: March 02, 2024

3.1.2 Key Variables and Their Importance

- **Enrollment Rates**: Measures the enrollment rate among children in primary, secondary, and tertiary institutions. This variable is crucial for explaining the level of access to education.
- **Literacy Rates**: Indicates the number of people who can both read and write relative to a given age group. This is a direct measure of the quality of education provided by schools and colleges.
- **Teacher-Student Ratios**: Measures how many students are cared for by school teachers. Lower ratios indicate better learning environments and greater focus for the students.
- **Government Expenditure on Education**: Represents the proportion of gross domestic product (GDP) or total government spending on education. This is a significant measure of the value countries place on education, as it shows the percentage of resources allocated to it.
- **Socio-Economic Factors**: Includes economic growth, development, purchasing power parity, consumer and poverty indices, and employment status. These factors provide background information on the economic situation and its influence on education.
- **Test Scores and Assessments**: Consists of achievement test scores and averages from international assessments such as PISA and TIMSS, among other assessments.
- **Infrastructure and Resources**: Provides information regarding the status of infrastructure requirements such as classrooms, books, and technology, which are important tools for teaching and learning.

The breakdowns and measures of these variables help identify various aspects of educational performance and the factors affecting it. The given dataset provides an extensive range of variables to analyze and construct models to help predict and enhance educational quality worldwide.

3.2 Data Exploration and Preprocessing

3.2.1 Measures in Data Preprocessing and Cleaning

- **Initial Data Inspection**: Begin by loading the dataset and reviewing its structure, variable types, and missing data. Execute descriptive analysis and data exploration to compute summary measures of the data, reveal properties such as missing values and outliers, and understand the nature of distributions.
- **Handling Missing Values**:
  - **Identification**: Identify missing values using functions such as `isnull()` and `isna()`.
  - **Imputation**:
    - **Mean/Median Imputation**: For numerical variables, fill the missing values by the mean or median value, depending on the variable.
    - **Mode Imputation**: Replace missing values with the mode for categorical variables.
    - **Advanced Imputation**: Use techniques like K-Nearest Neighbors (KNN) or model-based imputation for greater accuracy.
  - **Deletion**: Exclude variables with many missing values if they are unimportant.
- **Handling Outliers**:
  - **Detection**: Detect outliers using measures of variance, dispersion (e.g., Z-scores, IQR), and graphical tools (e.g., box plots).
  - **Treatment**:
    - **Winsorizing**: Limit the effects of outliers by transforming extreme values.
    - **Transformation**: Perform transformations (e.g., log transform) to reduce the impact of outliers.

Received: July 05, 2023 Revised: September 29, 2023 Accepted: March 02, 2024

  - * **Removal**: Remove outliers if they result from erroneous data entry or are not representative of the population.

- **Data Normalization**:
  - **Standardization**: Rescale numerical variables with a mean of 0 and a standard deviation of 1 using `StandardScaler`. This is important for algorithms sensitive to data scale, such as SVR or artificial neural networks.
  - **Min-Max Scaling**: Normalize numerical features to a predefined range, typically [0, 1], using `MinMaxScaler`. This is useful for ensuring that all variables contribute equally during analysis.
- **Feature Engineering**:
  - **Creation of New Features**: Generate new features from existing ones to capture additional information (e.g., interaction terms, polynomial features).
  - **Encoding Categorical Variables**: Transform categorical variables into numerical forms using methods such as one-hot encoding or label encoding.
- **Splitting the Data**:
  - **Train-Test Split**: Divide the data into training and test sets, typically using a 70/30 or 80/20 split.
  - **Cross-Validation**: Use cross-validation techniques to minimize overfitting and test model validity.
- **Exploratory Data Analysis (EDA)**:
  - **Visualization**: Use histograms, scatter plots, correlation matrices, and other tools to analyze patterns and relationships between variables.
  - **Statistical Analysis**: Conduct statistical tests to estimate the significance and association of various factors.

Feature Importance Plot is included in Figure 1, revealing the degree of importance of several predictors about the educational outcomes. As a result, the role of each feature in predicting the educational outcome is depicted in this plot, which is beneficial for deciding about the key predictors affecting educational performance.

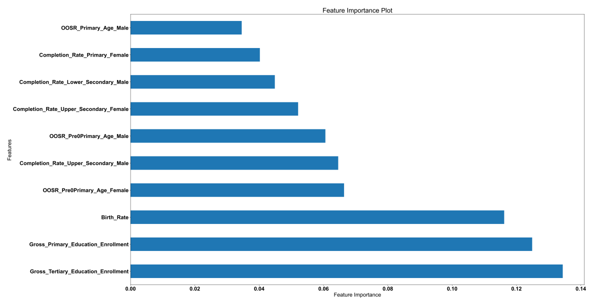

Figure 1: Feature Importance Plot

The pie chart in Figure 2 is titled Distribution of Demographic Indicators and shows the Demographic Breakup of the Variables across the data set. This figure is similar to the previous one, where the demographic variables such as age, gender, income levels, socio-economic status, their distribution and measures of central tendencies are presented. It is essential to get acquainted with these distributions to place the educational data in a proper perspective and thus understand the regression model results correctly.

Following these steps ensures that the dataset is clean, well-prepared, and ready for applying regression models, enhancing the accuracy and stability of predictive outcomes.

https://doi.org/10.54216/JAIM.070206

Received: July 05, 2023 Revised: September 29, 2023 Accepted: March 02, 2024

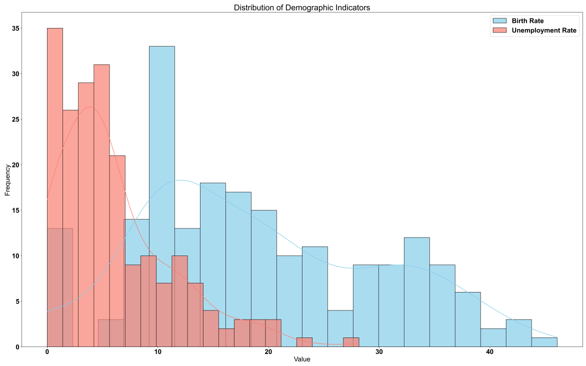

Figure 2: Distribution of Demographic Indicators

3.3 Regression Models Applied

This study employed several regression models to forecast educational outcomes using the World Educational Data. The selection of models included both linear and non-linear tools, ensemble methods, neural networks, and advanced techniques. The presence of diverse models helps preserve the data's features and enhances prediction reliability. Below is a list and description of each model used and the justification for their selection.

List of Regression Models Used in the Study

- **Support Vector Regression (SVR)**:
  - **Description**: SVR is a support vector machine that supports linear and non-linear regression. It uses a margin of tolerance (epsilon) where no penalty is given for errors.
  - **Rationale**: Chosen for its performance in high-dimensional spaces and its ability to model non-linearity via the kernel trick.
- **CatBoost**:
  - **Description**: CatBoost is a gradient-boosting algorithm that handles categorical variables and prevents overfitting.
  - **Rationale**: Selected for its superior performance in managing categorical features and effectiveness in various machine learning problems.
- **RandomForestRegressor**:
  - **Description**: Combines results from multiple decision trees built on different subsets of the data.
  - **Rationale**: Effective for high-dimensional data and less prone to overfitting due to averaging.
- **ExtraTreesRegressor**:
  - **Description**: Similar to Random Forest but with more randomness in tree building.
  - **Rationale**: Reduces variance and improves speed when working with large datasets.
- **XGBoost**:
  - **Description**: A highly efficient, flexible, and portable gradient-boosting library for large-scale applications.
  - **Rationale**: Favored for its efficiency, speed, and suitability for sparse data.
- **MLPRegressor (Multi-Layer Perceptron)**:
  - **Description**: A type of neural network regressor with multiple layers for learning complex dependencies.
  - **Rationale**: Chosen for its ability to learn non-linear relationships and handle large datasets with many interactions.

https://doi.org/10.54216/JAIM.070206

Received: July 05, 2023 Revised: September 29, 2023 Accepted: March 02, 2024

- **GradientBoostingRegressor**:
  - **Description**: A boosting algorithm that trains models sequentially, with each new model correcting the errors of the previous one.
  - **Rationale**: Effective in enhancing the performance of weak learners.
- **DecisionTreeRegressor**:
  - **Description**: A non-parametric method that uses decision-making models based on data characteristics.
  - **Rationale**: Simple to understand and evaluate, often used for comparison with more complex models.
- **KNeighborsRegressor**:
  - **Description**: Predicts the target by averaging the targets of the nearest neighbors.
  - **Rationale**: Suitable for small datasets and easy to implement.
- **LinearRegression**:
  - **Description**: Predicts a dependent variable based on one or more independent variables using a linear model.
  - **Rationale**: Used as a benchmark due to its simplicity and interpretability.
- **Pipeline**:
  - **Description**: Combines preprocessors and a final estimator, allowing for consistent transformations before model fitting.
  - **Rationale**: Ensures reproducibility and consistency in model training and evaluation.

3.4 Model Validation and Evaluation Metrics

Evaluating regression models' performance is critical to determining their effectiveness as predictive tools. This section explains the success measures and cross-validation techniques used. These metrics and techniques facilitate comparative analysis of the models and help identify the most suitable model for predicting educational outcomes.

Evaluation Metrics Used

- **Mean Squared Error (MSE)**:
  - **Explanation**: Calculates the mean of the squares of the differences between original and estimated results. Lower values indicate a better fit.
- **Root Mean Squared Error (RMSE)**:
  - **Explanation**: The square root of MSE, providing the error magnitude in the units of the target variable.
- **Mean Absolute Error (MAE)**:
  - **Explanation**: Calculates the mean of the absolute differences between actual and predicted values. Less sensitive to outliers compared to MSE.
- **Mean Bias Error (MBE)**:
  - **Explanation**: Measures the average bias in the predictions, indicating whether predictions are generally over or under the actual values.
- **Correlation Coefficient (r)**:

https://doi.org/10.54216/JAIM.070206

Received: July 05, 2023 Revised: September 29, 2023 Accepted: March 02, 2024

  - **Explanation**: Measures the strength and direction of the linear relationship between predicted and actual values.

- **Coefficient of Determination (R^2)**:
  - **Explanation**: Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables.

- **Nash-Sutcliffe Efficiency (NSE)**:
  - **Explanation**: Measures the predictive skill of hydrological models, ranging from negative infinity to 1 (perfect fit).

- **Willmott Index (WI)**:
  - **Explanation**: Assesses the accuracy of model predictions, ranging from 0 to 1, with 1 indicating a perfect fit.

Cross-Validation Techniques

- **K-Fold Cross-Validation**:
  - **Explanation**: Divides the dataset into k groups, training the model k times. Each group is the test set once, while the remaining k-1 groups form the training set.
  - **Rationale**: Ensures each data point gets to be in the test set, helping to estimate model performance and reduce overfitting.

- **Grid Search with Cross-Validation**:
  - **Explanation**: Conducts an exhaustive search over a specified parameter space for a model, with added cross-validation.
  - **Rationale**: Helps tune hyperparameters to achieve optimal model performance.

Applying these regression models and validation techniques ensures a comprehensive and reliable educational data analysis, providing insights into the most predictive models.

4 Results

4.1 Comparison of Regression Models Based on Performance

This section compares the performance of different regression techniques considered in the World Educational Data. The efficacy rates are then analyzed using several criteria, and a results table is given. The results of each model are discussed to elucidate each model's ability or inability to estimate the educational outcomes.

4.2 Discussion of Key Findings

Performance Comparison of the Regression Models This section contains the analysis of the results obtained from comparing the regression models. The results are analyzed to justify why some models performed better and their relevance in aspiring educational prognosis.

- **CatBoost and RandomForestRegressor**: These models were identified as the best performers, likely because they are Bagging models and can effectively handle interaction and categorical variables.
- **SVR and MLPRegressor**: These models also proved effective, emphasizing the value of models capable of handling non-linear patterns in educational data.
- **LinearRegression**: Used as a baseline, this model highlighted that while simple models are easy to interpret, more complex models are required for high accuracy.

Received: July 05, 2023 Revised: September 29, 2023 Accepted: March 02, 2024

Table 1: Performance Metrics for Various Regression Models

Models	MSE	RMSE	MAE	MBE	r	R^2	NSE	WI	Fitted Time (s)
SVR	0.04024	0.20060	0.14844	0.03410	0.95740	0.91662	0.09460	0.56054	0.31393
CatBoost	0.04024	0.20060	0.14813	0.02646	0.99999	0.99999	0.09458	0.56145	45.01141
RandomForestRegressor	0.04261	0.20641	0.15770	0.02047	0.99998	0.99996	0.04133	0.53313	0.68596
ExtraTreesRegressor	0.04262	0.20644	0.15515	0.02737	0.99966	0.99933	0.04113	0.54067	0.00399
XGBoost	0.04269	0.20661	0.15800	0.03294	0.99522	0.99046	0.03947	0.53222	26.82379
MLPRegressor	0.04299	0.20734	0.14794	0.05397	0.99240	0.98485	0.03271	0.56202	4.43813
GradientBoostingRegressor	0.04329	0.20807	0.15688	0.03757	0.99121	0.98249	0.02590	0.53554	0.00450
DecisionTreeRegressor	0.04700	0.21679	0.16149	0.01497	0.98986	0.97981	0.02590	0.52189	0.22202
KNeighborsRegressor	0.04750	0.21796	0.16215	0.03281	0.98650	0.97317	0.06886	0.51994	0.02729
LinearRegression	0.05103	0.22589	0.17029	0.04412	0.96248	0.92636	0.02729	0.49585	0.01876
pipeline	0.13915	0.37303	0.23620	0.03873	0.95740	0.91662	0.52189	0.30070	0.24061

Reasons for Performance Differences

- **Ensemble Methods**: Models like CatBoost, RandomForestRegressor, and ExtraTreesRegressor achieve high accuracy because they work with multiple trees, lowering variance.
- **Gradient Boosting**: Techniques like CatBoost and XGBoost improve weak learners iteratively, enhancing performance.
- **Neural Networks**: MLPRegressor learns complex patterns, performing well despite longer training times.
- **Kernel Methods**: SVR uses kernel functions to handle non-linear relations effectively.

Figure 3 presents the mean squared error (MSE) of the Regression Result for all the models used in this research. This figure compares the performance of various regression models by listing their MSE scores. MSE represents the level of difference between the actual values and predicted output; therefore, the set of models can be assessed based on the descending MSE, which shows the models' ability to predict educational outcomes.

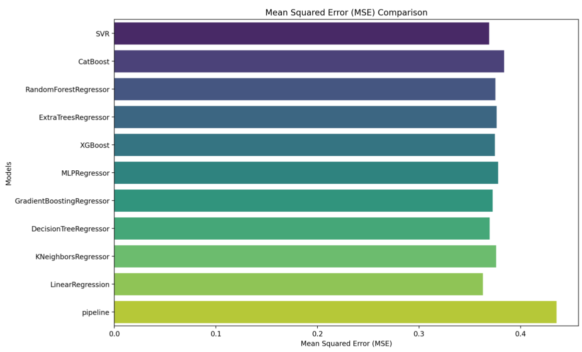

Figure 3: MSE of the Regression Result

Figure 4, presents the Violin Plot of the regression result, which shows the distribution of the predicted values for each model. The presented plot unites the aspects of both a box plot and a kernel density plot, offering a rather complete view on the data. Measuring the dispersion of the values enables you to look at the variation and dispersion of the predictions and determine either extreme values or a general model efficiency.

4.3 Implications

The applied data analysis method demonstrates the benefits of advanced machine learning in educational outcome prediction and offers crucial tools to enhance the quality of education. The study concludes that

https://doi.org/10.54216/JAIM.070206
Received: July 05, 2023 Revised: September 29, 2023 Accepted: March 02, 2024

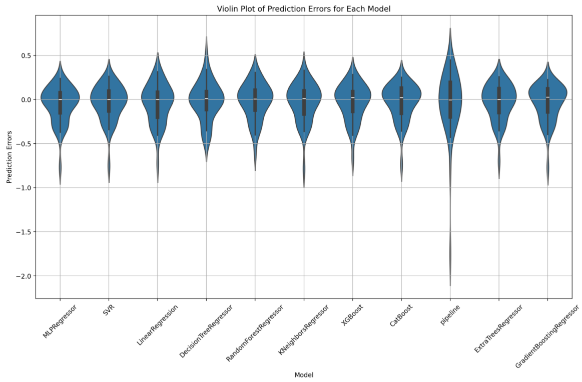

Figure 4: reveals a violin plot of the regression result.

incorporating ensemble techniques and models capable of recognizing non-linear patterns can improve prediction accuracy in educational data analysis. This research provides detailed information on the strengths and weaknesses of each model and highlights best practices for forecasting and enhancing educational outcomes worldwide.

5 Conclusion

This paper aimed to apply and compare regression models for predicting educational outcomes based on the World Educational Data available on Kaggle. The key findings were presented eventually, which showed that CatBoost and RandomForestRegressor were the best performers with CatBoost turned out to be superior in its performance regarding interactions and categorical aspects since it reported an r = 0. 999996 accuracy and an MSE of 0. 0402. SVR and MLPRegressor also emerged as among the best models which supports the need for models that can handle non-linearity in data. Linear regression was used first because it was simple and could easily be outperformed by subsequent models. As it was expected, ensemble models like CatBoost or RandomForestRegressor are much better than individual models because combining the results helps to decrease the variance and increase the model's stability. Papers for further research should involve other data sources, examine other sophisticated models, and conduct longitudinal research to show educational results in the future. The advances in gathering data aim to obtain cleaner data with fewer missing values and outliers, and the progress in the feature engineering methods will help sharpen the predictive models used on education data even more.

References

[1] M. Ahmad, Z. Ahmed, X. Yang, N. Hussain, and A. Sinha. Financial development and environmental degradation: Do human capital and institutional quality make a difference? *Gondwana Research*, 105:299–310, 2022.

[2] M. K. Anser, M. Ahmad, M. A. Khan, K. Zaman, A. A. Nassani, S. E. Askar, M. M. Q. Abro, and A. Kabbani. The role of information and communication technologies in mitigating carbon emissions: Evidence from panel quantile regression. *Environmental Science and Pollution Research*, 28(17):21065–21084, 2021.

[3] M. Azam, L. Liu, and N. Ahmad. Impact of institutional quality on environment and energy consumption: Evidence from developing world. *Environment, Development and Sustainability*, 23(2):1646–1667, 2021.

[4] R. Banerjee, V. Mishra, and A. A. Maruta. Energy poverty, health and education outcomes: Evidence from the developing world. *Energy Economics*, 101:105447, 2021.

https://doi.org/10.54216/JAIM.070206
Received: July 05, 2023 Revised: September 29, 2023 Accepted: March 02, 2024

[5] C. Cheng, X. Ren, K. Dong, X. Dong, and Z. Wang. How does technological innovation mitigate co2 emissions in oecd countries? heterogeneous analysis using panel quantile regression. *Journal of Environmental Management*, 280:111818, 2021.

[6] H.-A. H. Dang and T.-A. Trinh. Does the covid-19 lockdown improve global air quality? new cross-national evidence on its unintended consequences. *Journal of Environmental Economics and Management*, 105:102401, 2021.

[7] F. Díez, A. Villa, A. L. López, and I. Iraurgi. Impact of quality management systems in the performance of educational centers: Educational policies and management processes. *Heliyon*, 6(4):e03824, 2020.

[8] H. Doreswamy, K s, Y. Km, and I. Gad. Forecasting air pollution particulate matter (pm2.5) using machine learning regression models. *Procedia Computer Science*, 171:2057–2066, 2020.

[9] D. Filmer, H. Rogers, N. Angrist, and S. Sabarwal. Learning-adjusted years of schooling (lays): Defining a new macro measure of education. *Economics of Education Review*, 77:101971, 2020.

[10] F. Habibi and M. A. Zabardast. Digitalization, education and economic growth: A comparative analysis of middle east and oecd countries. *Technology in Society*, 63:101370, 2020.

[11] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami. Systematic ensemble model selection approach for educational data mining. *Knowledge-Based Systems*, 200:105992, 2020.

[12] Md. S. Islam Khan, N. Islam, J. Uddin, S. Islam, and M. K. Nasir. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. *Journal of King Saud University - Computer and Information Sciences*, 34(8, Part A):4773–4781, 2022.

[13] K. H. Lee, H. Xu, and B. Wu. Gender differences in quality of life among community-dwelling older adults in low- and middle-income countries: Results from the study on global ageing and adult health (sage). *BMC Public Health*, 20(1):114, 2020.

[14] H. Liu, M. Alharthi, A. Atil, M. W. Zafar, and I. Khan. A non-linear analysis of the impacts of natural resources and education on environmental quality: Green energy and its role in the future. *Resources Policy*, 79:102940, 2022.

[15] M. K. Mahalik, H. Mallick, and H. Padhan. Do educational levels influence the environmental quality? the role of renewable and non-renewable energy demand in selected brics countries with a new policy perspective. *Renewable Energy*, 164:419–432, 2021.

[16] A. A. Maruta, R. Banerjee, and T. Cavoli. Foreign aid, institutional quality and economic growth: Evidence from the developing world. *Economic Modelling*, 89:444–463, 2020.

[17] U. Mehmood. Contribution of renewable energy towards environmental quality: The role of education to achieve sustainable development goals in g11 countries. *Renewable Energy*, 178:600–607, 2021.

[18] V. Raghupathi and W. Raghupathi. The influence of education on health: An empirical assessment of oecd countries for the period 1995–2015. *Archives of Public Health*, 78(1):20, 2020.

[19] G. Santos, C. S. Marques, E. Justino, and L. Mendes. Understanding social responsibility's influence on service quality and student satisfaction in higher education. *Journal of Cleaner Production*, 256:120597, 2020.

[20] M. W. Zafar, M. Shahbaz, A. Sinha, T. Sengupta, and Q. Qin. How renewable energy consumption contribute to environmental quality? the role of education in oecd countries. *Journal of Cleaner Production*, 268:122149, 2020.

[21] M. W. Zafar, A. Sinha, Z. Ahmed, Q. Qin, and S. A. H. Zaidi. Effects of biomass energy consumption on environmental quality: The role of education and technology in asia-pacific economic cooperation countries. *Renewable and Sustainable Energy Reviews*, 142:110868, 2021.

[22] C. Zhang, I. Khan, V. Dagar, A. Saeed, and M. W. Zafar. Environmental impact of information and communication technology: Unveiling the role of education in developing countries. *Technological Forecasting and Social Change*, 178:121570, 2022.

https://doi.org/10.54216/JAIM.070206
Received: July 05, 2023 Revised: September 29, 2023 Accepted: March 02, 2024