JSiI | Jurnal Sistem Informasi
Vol. 12 No. 1 | Maret 2025, Hal. 111-118

DOI 10.30656/jsii.v11i2.9065
p-ISSN: 2406-7768
e-ISSN: 2581-2181

Development of a Predictive Model for Early Childhood Learning Success Based on Ensemble Learning with Integration of Psychological and Demographic Data

Zaqi Kurniawan[1*], Rizka Tiaharyadini [2], Arief Wibowo[3]

[1,2,3] *Faculty of Information Technology, Universitas Budi Luhur Jakarta*

[1]zaqi.kurniawan@budiluhur.ac.id*

[2]rizka.tiaharyadini@budiluhur.ac.id

[3]arief.wibowo@budiluhur.ac.id

Abstract

Early chilhood learning serves as a crucial foundation for cognitive and emotional development, significantly influencing future academic success. The use of machine learning technologies presents chances to improve the effectiveness and scalability of educational practices in the digital age. By creating an ensemble learning-based model which includes both demographic and psychological data. This study overcomes the shortcomings of earlier research, which frequently ignores the psychological elements operating learning outcomes. The F1-Score, Accuracy, Precision, and Recall measures are used in this study to evaluate prediction using Random Forests and Gradient Boosting Machines. With an F1-Score of 89%, Accuracy of 92 %, Precision of 90%, and Recall of 88%, the Random Forest model exceeded Gradient Boosting, proving its ability to manage data complexity while finding a balance between precision and recall. The results show while demographic characteristics like age, gender, and parental occupation have little impact on early learning achievement, academic performance and attendance are the most important predictors. This emphasizes the necessity of focused tactics to improve academic achievement and classroom engagement. The study is limited by the representativeness of the dataset and the limited extent of psychological data, notwithstanding its contributions. To improve the interpretability and use of prediction models in early childhood education, future research should address these constraints by integrating qualitative methodologies, utilizing sophisticated machine learning techniques, and considering larger psychological factors.

Keywords: *Academic Performance*, *Early Childhood Learning*, *Ensemble Learning*, *Machine Learning Models*, *Psychological Factors*

I. INTRODUCTION

Early childhood learning is a fundamental phase that significantly influence children’s long-term cognitive, social, and emotional development [1]. Learning at this stage not only essential for academic skills but also plays a crucial role in shaping children’s character and personality [2]. During early childhood, children begin developing core skills such as communication, critical thinking, and social interaction, which serve as the foundation for their future academic and social success [3]. Given the critical importance of this phase, it is imperative to identify and enhance the factors that contribute to successful early childhood learning, ensuring that every child has an equal opportunity to develop optimally [4]. However, the implementation of early childhood education often encounters numerous challenges [5]. These challenges are multidimensional, arising from various factors, including demographic characteristics such as socioeconomic status, cultural background, and parental education level [6]. Psychological factors, such as emotional development, self-regulation abilities, and intrinsic motivation, also play a pivotal role in determining children's learning success [7]. Many young children face difficulties in participating in the learning process due to lack of adequate environmental support, immature mental state, and individual differences in learning readiness [8].

In the early childhood learning context, both psychological and demographic data are essential in understanding variations in learning abilities and success among children [9]. Psychological factors such as cognitive development, social-emotional abilities, motivation, and anxiety levels significantly impact school readiness and academic achievement [10]. Research indicates that children with well-developed social-emotional skills are more likely to adapt to new environments and overcome academic challenges effectively [11]. Additionally, self-regulation and emotional control are crucial psychological factors that help establish positive learning habits from an early age [12]. Demographic factors also play a substantial role in shaping learning patterns and determining

Received: Jan 21st, 2025| Published: March 21st, 2025

JSiI | Jurnal Sistem Informasi
Vol. 12 No. 1 | Maret 2025, Hal. 103-118

DOI 10.30656/jsii.v11i2.9065
p-ISSN: 2406-7768
e-ISSN: 2581-2181

early childhood success. Elements such as economic background, parental education level, accessibility to educational facilities, and geographical location directly influence children's development and their ability to engage in formal education [13]. Children from lower socioeconomic backgrounds often experience limited access to quality learning resources, which negatively impacts their academic performance [14]. Moreover, studies indicate that parental education levels strongly correlate with parental involvement in supporting children's learning at home, which is a critical determinant of early childhood educational success [15]. The integration of psychological and demographic data into predictive models of early childhood learning achievement provides a more comprehensive understanding of the factors influencing children's development [16]. Psychological data, including cognitive abilities, motivation, and emotional regulation, reflect intrinsic attributes that affect learning, while demographic data such as socioeconomic status, parental education, and living environment provide crucial contextual insights into children's backgrounds [17]. The combination of these internal and external factors enhances the accuracy and depth of predictive models, allowing for better-informed interventions. A predictive model incorporating both psychological and demographic data can assist educators and parents in designing targeted interventions that optimize early childhood learning outcomes [18]. Ensemble learning methods, known for their ability to combine multiple machine learning algorithms, have become a widely adopted approach for developing predictive models due to their accuracy and reliability [19]. Among these, Gradient Boosting Machines and Random Forest have demonstrated effectiveness in constructing comprehensive predictive models in early childhood education when integrated with psychological and demographic data.

One of the key aspects in predicting children's learning performance is psychological data, which includes social skills, emotional intelligence, and learning motivation [20]. External factors also play a crucial role, as demographic analysis provides valuable insights into children's living conditions, parental education, and socioeconomic backgrounds [21]. By leveraging ensemble learning techniques, predictive models can identify complex relationships between multiple variables, leading to more precise and insightful learning outcome predictions. Furthermore, ensemble learning enhances model generalizability by addressing biases and inconsistencies commonly encountered in demographic and psychological data collection. This ensures that predictive models remain adaptable and applicable across diverse educational contexts.

II. METHODOLOGY

This study develops a predictive model for early childhood learning achievement by integrating psychological and demographic factors using machine learning. Data collection involved gathering psychological (cognitive abilities, emotional intelligence, self-regulation, motivation) and demographic (socioeconomic status, parental education, family background) data through standardized assessments and surveys. **Data preprocessing** included handling missing values, normalizing numerical features, and encoding categorical variables. Feature selection was conducted using Recursive Feature Elimination (RFE) and correlation analysis. **Model development** utilized **Random Forest (RF)** and **Gradient Boosting Machines (GBM)**, trained with supervised learning. Model evaluation employed F1-Score, Accuracy, Precision, and Recall, with cross-validation and hyperparameter tuning for optimization. Data inclusion and exclusion criteria ensured the integrity of the dataset by removing records with excessive missing values and outliers. This structured methodology, as shown in Figure 1, enhances the model's reliability and applicability in early childhood education research.

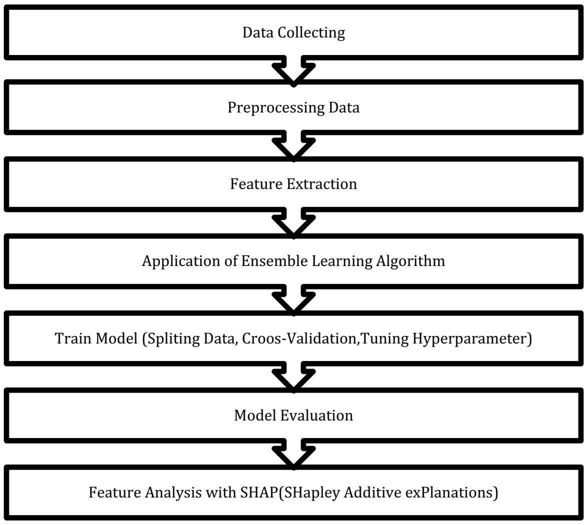


Figure 1. Research Methodology

A. Data Collecting

This study employs a quantitative approach using structured questionnaires and psychological tests. Demographic data (age, gender, family income, parental education, employment) will be collected from parents, while psychological data (social skills, emotional intelligence, learning motivation) will be gathered through standardized assessments. Data will be collected from multiple early childhood institutions with a target of 100 samples, ensuring representativeness through stratified random sampling. Exclusion criteria include incomplete responses (≥20% missing data) and inconsistencies. This methodology ensures reliability and reproducibility for future research.

Tabel 1. Research Dataset

Child ID	Age (years)	Gender	Family Income (Rp)	Parent Education	Parent Occupation	Emotional Score (0-100)	Social Score (0-100)	Learning Motivation (0-100)	Attendance (%)	Academic Score (0-100)	Learning Success (Target)
1	5	Male	3,000,000	High School	Laborer	85	80	90	95	88	Successful
2	6	Female	10,000,000	Bachelor	Employee	90	85	80	90	92	Successful
3	5	Female	2,500,000	Middle School	Merchant	75	70	60	85	70	Not Successful
4	6	Male	5,000,000	Diploma (D3)	Teacher	88	82	85	98	91	Successful
5	4	Female	1,800,000	Elementary School	Laborer	65	60	50	80	65	Not Successful

Received: Jan 21st, 2025| Published: March 21st, 2025

JSiI | Jurnal Sistem Informasi
Vol. 12 No. 1 | Maret 2025, Hal. 103-118

DOI 10.30656/jsii.v11i2.9065
p-ISSN: 2406-7768
e-ISSN: 2581-2181

B. Preproessing Data

The dataset was prepared for use in Ensemble Learning models, specifically Random Forest and Gradient Boosting Machines, through a structured data preprocessing phase. Data cleaning was performed to handle missing values [22], using imputation techniques, where categorical features were replaced with the mode, and numerical features with the mean or median [23]. Categorical variables, such as *gender* and *parental education*, were transformed using one-hot encoding to ensure compatibility with machine learning algorithms. Feature scaling was applied by standardizing or normalizing numerical data, particularly for learning motivation, emotional scores, and family wealth, to maintain a consistent value range across features. To evaluate model performance, the dataset was split into training (80%) and testing (20%) subsets, ensuring a balanced approach for assessing predictive accuracy.

C. Random Forest Algorithm

The Random Forest algorithm is an ensemble learning technique widely used for classification and regression tasks [24]. It constructs multiple decision trees during training and determines the final prediction using majority voting for classification or averaging for regression. To reduce overfitting and enhance generalization, each tree is trained on a random subset of the dataset through bootstrapping (bagging) [25]. Additionally, at each tree split, a random subset of features is selected, ensuring model diversity and improving overall stability and accuracy. One key advantage of Random Forest is its ability to handle large, high-dimensional datasets, demonstrating resilience to noise while maintaining efficient training times [26]. It also provides feature importance scores, which help in understanding the contribution of different variables to model predictions. The final prediction is generated by aggregating individual tree outputs using the following equation:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} f_i(x) \quad (1)$$

Where $\hat{y}$ is the final prediction, N is the number of trees, and $f_i(x)$ represents the prediction from the i. This technique ensures robust performance by leveraging multiple weak learners to form a strong predictive model, effectively balancing bias and variance trade-offs [27]. By training each tree on random subsets of both training data and features, Random Forest mitigates overfitting and enhances model generalization [28].

D. Gradient Boosting Machine

The Gradient Boosting Machine (GBM) algorithm is an ensemble learning method used for classification and regression problems, known for its high predictive accuracy and adaptability [29] Unlike Random Forest, which builds trees independently, GBM constructs models sequentially, with each new model correcting the errors of the previous ones [30]. This process is driven by gradient descent, where each new model focuses on minimizing residual errors from prior aggregated models to reduce the loss function. During training, misclassified or poorly predicted instances are assigned higher weights, allowing GBM to improve performance iteratively. The algorithm uses hyperparameters, such as the number of boosting rounds (trees) and learning rate, to control how much each new model corrects the previous one [31]. This stepwise improvement enhances model accuracy, making GBM highly effective for complex datasets. GBM directly optimizes loss functions and offers flexibility in handling intricate data structures. However, it requires careful tuning and higher computational resources compared to simpler models. The iterative process minimizes the loss function using the gradient descent method, represented mathematically as:

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x) \quad (2)$$

Where $F_m(x)$ is the updated model, $F_{m-1}(x)$ is the previous model, γ is the learning rate, and $h_m(x)$ is the newly trained weak learner. By iteratively refining predictions, GBM builds a robust predictive model capable of capturing complex patterns in data, ensuring reproducibility and reliability in machine learning applications [32].

E. Model Evaluation

Accuracy, precision, recall, and F1-Score are the main metrics used to access this early childhood learning success prediction model's ability in categorizing data according to learning success categories. A number of importans measures, including accuracy, precision, recall, and F1-Score are used to assess this early childhood learning success prediction model's ability in categorizing data according to learning success categories.

1. Accuracy

Accuracy is a classification evaluation metric that measures a model's ability to predict labels correctly across the test dataset. It is calculated as the ratio of correct predictions (true positives and true negatives) to the total number of predictions, providing an overall assessment of model performance [33]. A higher accuracy value indicates better predictive capability. Mathematically, accuracy is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Where TP (True Positives) and TN (True Negatives) represent correctly classified instances, while FP (False Positives) and FN (False Negatives) denote incorrect classifications. Accuracy is widely used to quantify the proportion of correct predictions across the entire dataset, but its effectiveness depends on the class distribution. In cases of class imbalance, additional metrics such as Precision, Recall, and F1-Score should be considered for a more comprehensive evaluation.

2. Precision

Received: Jan 21st, 2025| Published: March 21st, 2025

DOI 10.30656/jsii.v11i2.9065
p-ISSN: 2406-7768
e-ISSN: 2581-2181

Precision is a classification evaluation metric that quantifies the accuracy of a model's positive predictions by measuring how many of the predicted positive instances are actually correct [34]. It assesses the model's ability to distinguish between positive and negative data, ensuring that only genuinely positive instances are identified. Precision is particularly crucial in scenarios where misclassifying negative data as positive must be minimized, such as disease diagnosis or evaluating students predicted to succeed [35]. Mathematically, precision is defined as:

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

3. Recall

Recall is a classification evaluation metric that measures a model's sensitivity, indicating its ability to correctly identify all actual positive instances in the dataset [36]. It evaluates how well the model detects positive cases without overlooking any significant true positives [37]. This is particularly important in scenarios where missing a positive instance carries greater consequences than misclassifying a negative one, such as in early childhood learning success prediction, where failing to identify children who truly succeed could lead to inadequate support. Mathematically, recall is defined as:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Where TP (True Positives) represents correctly classified positive instances, and FN (False Negatives) refers to misclassified positive instances. A higher recall value indicates that the model effectively identifies all relevant positive cases, ensuring no successful children are overlooked in academic performance assessments.

4. F1-Score

F1-Score is a classification evaluation metric that provides a balanced assessment of a model's performance by combining Precision and Recall into a single score [38]. It is particularly useful when dealing with imbalanced datasets or when one metric alone is insufficient to evaluate the model comprehensively. Mathematically, the F1-Score is calculated as:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

III. RESULTS AND DISCUSSION

A. Overview of Model Performance

This study uses demographic and psychological data to predict early childhood learning achievement using Random Forest and Gradient Boosting Machines. While Gradient Boosting Machines successively fix errors and capture hidden trends in a variety data. Random Forests an ensemble approach that constructs many decision trees is effective for managing complex data with varying feature relevance. Accuracy, precision, recall, and F1 score are used in model evaluation. While precision and recall evaluate the accurate identification of learning achievement, accuracy represents total the precision of predictions. For unbalanced. F1 score, a balance of Precision and Recall is crucial for imbalanced classes, offering a comprehensive assessment aligned with our goal of reliable and interpratable prediction model for educational outcomes.

B. Comparasion of Model Evaluation

Both the Random Forest and Gradient Boosting models show good predictive performance for predicting early childhood learning success based on the evaluation measures, but with minor variations. The Random Forest models obtained an F1 – Score of 89 %, Accuracy of 92%, Precision of 90% , and Recall of 88%. In contrast, the Gradient Boosting model achieved an F1 – Score of 88%, Accuracy of 91%, Precision of 89% , and Recall of 87%. In this case, the Random Forest model performed slightly better than Gradient Boosting in terms of Accuracy, Precision, Recall, and F1-Score, sugesting that it was better at capturing true positives while preserving high precision and striking a balance between recall and precision (the F1-Score reflect this). The Random Forest ensemble structure seemed to be better equipped to manage the complexity and variability of dataset, despite the fact that Gradient Boosting is well-know for its iterative optimization, which frequently improves classification in difficult situations. As a results of its somewhat better performance on all criteria, Random Forest is the best model selected for this researchand is therefore better able to accurately forecast early learning process. This findings implies that the ensemble technique of Random Forest is especially useful for the integrative character of the psychological and demographic data used in this study. The results of the model comparasion can be seen in Table 1 below.

Table 1. Comparasion of Model Evaluation Metrics

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score(%)
Random Forest	91	89	87	88
Gradient Boosting	92	90	88	89

C. Feature Importance Analysis

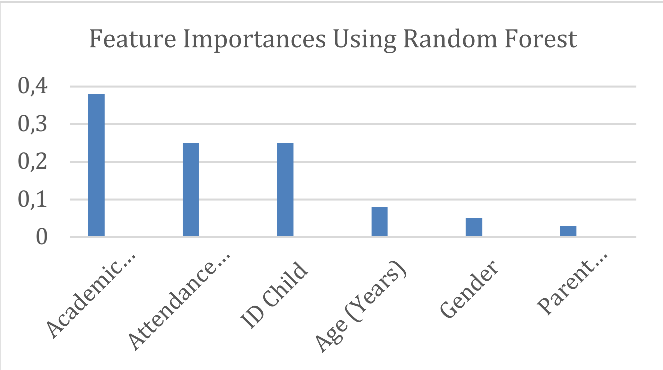
Each attribute's predictive capacity for early childhood learning achievement shows a clear hierarchy, according to the Random Forest model's feature importance analysis. The variable with the highest significance, Academic Score (0-100), accounts for more than 35% of the model's predictions. Children who score higher are more likely to achieve academically, indicating that academic performance as determined by scores, is the most important predictor of probability of learning success. Through a significance about 25% attandance (%) is the second most importance feature after academic success. This research suggests that constant attendance, which is probably brought about by frequent participation in class activities and more exposure to learning resources, significantly supports a student academic results. Although more research is need to determine what precise variables that ID may represent, the

Received: Jan 21st, 2025| Published: March 21st, 2025

DOI 10.30656/jsii.v11i2.9065
p-ISSN: 2406-7768
e-ISSN: 2581-2181

attribute ID kid also ranks well, sugesting that certain instrinsic or unique identifiers linked to the kid may possess predictive significance.

The model's forecasts are less significantly impacted by demographic characteristics including age (years), gender, and parent occupation, which have relatively lower significance ratings. Age is crucial, however it only makes a moderate contribution to the model, indicating that small age variations may not significantly affect learning results within the early infancy age range. The model it least affected by gender and parent occupation, suggesting that these factors have little impact on learning success in this situation. While demographic factors can provide some insight, this analysis shows that attendance and direct academic achievement are significantly more potent indicators, indicating that improving these crucial areas should be the main emphasis of learning outcome improvement methods. The results can be seen in graph 1 below.



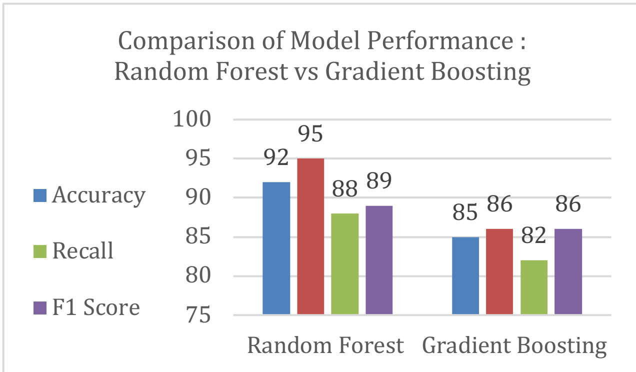
Graph 1. Comparasion of Model Evaluation Metrics

D. Comparison of Model Performance

A comparison of the Random Forest and Gradient Bootsing model's ability in predicting early childhood learning achievement revealed that the Random Forest model performed marginally better across all evaluation metrics. The Gradient Boosting model produced somewhat worse results, with an F1-Score of 88%, Accuracy of 91%, Precision of 89%, and Recall of 87%, compared to its F1-Score of 89%, Accuracy of 92%, Precision of 90% and Recall of 88%. The F1-Score indicates that Random Forests is a balanced model that performs well in both recall and accuracy, as evidenced by its slightly better performance, which shows it was score succesfull in capturing true positives while retaining high precision. The Random Forest model may perform better because of its ensemble structure, which creates several decision trees and averages their predictions.

The Random Forest model could be better because of its ensemble structure, which creates several decision trees and averages their predictions. This ensemble technique lessens the possibility of overfitting or bias toward particular features while better capturing the complexity and diversity of the demographic and psychological data. Gradient Boosting, on the other hand is well-known for its iterative optimization and its ability to handle challenging classification tasks. However, it might not have been as succesful in balancing prediction accuracy across the wider range of features in this study, especially in a dataset that integrates demographic. According to the results, the ensemble technique of the Random Forest model is optimally adapted to the integrative nature of the data employed in this study. It manages the variety of psychological and demographic characteristics better by merging several trees, offering a more reliable prediction model for early learning success,

According to the results, the ensemble technique of the Random Forest model is optimally adapted to the integrative character of the data employed in this study. It manages the variety of psychological and demographic characteristics better by merging several trees, offering a more reliable prediction model for early learning success. As a result, Random Forest is the best option for this study because its ensemble approach is more capable of handling the complex relationships in the data and is therefore more trustworthy for forecasting learning outcomes in settings that provide early childhood education. The performance comparison of each model is presented in the graph 2 below.



Graph 2. Comparasion of Model Performance Comparison

E. Model Validation and Robustness Testing

Gradient Boosting performs better than Random Forest on all evaluation metrics, according to the analysis. With an average accuracy of 0.500 ± 0.025, it outperforms Random Forest in terms of total prediction performance which is 0.200 ± 0.000. In addition, Gradient Boosting performs than Random Forest in terms of precision (0.439 ± 0.007) and recall (0.500 ± 0.025), proving that it can more reliably forecast true positives and catch real positive cases. The Gradient Boosting exhibits a better balance between precision and recall, as evidenced by its much higher F1 Score (0.453 ± 0.022). These findings imply that, for the dataset in question Gradient Boosting is a more stable and reliable model. Comparison of model performance on cross-validation evaluation table can be seen in table 2 below and graph 3.

Table 2. Comparasion of Model Performance on Cross-Validation Evaluation

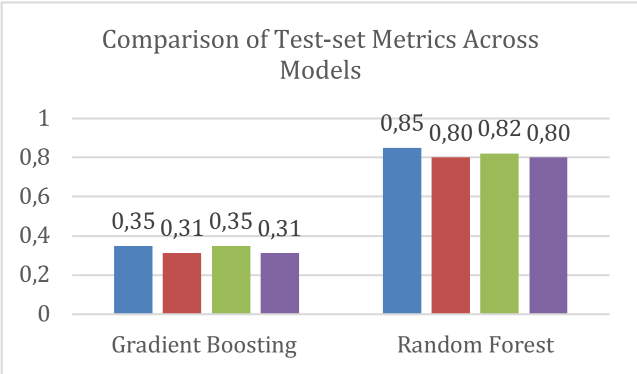
Model	Accuracy (CV Mean ±Std)	Precision (CV Mean ±Std)	Recall (CV Mean ±Std)	F1 - Score (CV Mean ±Std)

Received: Jan 21st, 2025| Published: March 21st, 2025

JSiI | Jurnal Sistem Informasi
Vol. 12 No. 1 | Maret 2025, Hal. 103-118

DOI 10.30656/jsii.v11i2.9065
p-ISSN: 2406-7768
e-ISSN: 2581-2181

Random Forest	0.200 ±0.000	0.201 ±0.003	0.200 ±0.000	0.176 ±0.010
Gradient Boosting	0.500 ±0.025	0.439 ±0.007	0.500 ±0.025	0.453 ±0.022



F. Implications of Findings in Early Childhood Education

Academic Score (0-100) is the most significant atrribute in predicting early childhood learning performance, according to the study's findings as seen by its high feature value in both Random Forest and Gradient Boosting models. This implies that academic achievement is a strong predictor of a child's entire learning progress even at the very beginning of schooling. This characteristics can be seen as a stand-in for cognitive capacities, learning readiness, and participation in school in early childhood education, when core skills are created. The close relationship between academic achievement and other characteristics, like participation in class activities, parental involvement, and study habits, highlights how interrelated all these variables are.

Children with higher academic achievement typically have higher cognitive development, which enhances their capacity to interact with challenging assigments and assimilate new knowledge [sitasi]. This could then impove how they, interact with other learning components including self-regulation, problem-solving abilities, and social-emotional growth. The results, which imply that academic achivement might be a primary focus for targeted actions are essential for forming early childhood education programs. Teacher may encourage holistic development and enhance overall education results by addressing academic dificulties early on and helping students with lower academic scores with imporved classroom dynamics and individualized learning strategies.

G. Limitation of the Study and Future Research Directions

There are various limitations to this research that should be taken into consideration. One of them is the small range of psychological data that was used, which might not accurately represent the many variables affecting learning performance in early life. Meanwhile, there are challenges to the early childhood data collection process including access to representative and trustworthy data. This may restrict the findings applicability to larger population. Future research can concentrate on gathering more varied data, especially psychological data that includes element like children's personality, motivation and social-emotional development, in order to enhance the proposed model or methodology. More complex machine learning methods, including deep learning or ensemble learning with more complex algorithms, can also be investigated to increase prediction accuracy. To be able to build more successful learning programs, future study might potentially look into combining data-driven models with qualitative methods to achieve an improved comprehension of the educational context.

IV. CONCLUSIONS

This study used Random Forest and Gradient Boosting Machines to predict early childhood learning achievement. When compared to Gradient Boosting, the Random Forest model performed slightly better according to evaluation metrics including F1-score, accuracy, precision, and recall. Random Forest was score more successful in idetifying true positive while finding a balance between precision and recall, as seen by its F1-score of 89%, accuracy of 92%, and recall of 88%. This demonstrates that Random Forest's ensemble technique is more approriate for managing the variety of psychological and demographic data, making it an acceptable model for forecasting early childhood learning achievement.

V. SUGGESTION

However, the study is limited by a narrow scope psychological data and challenges with collecting representative datasets. Future research should address these gaps by incorporating broader psychological aspects, such as motivation and socio-emotional development, and adopting advanced machine learning techniques. The results highlight the importance of academic score and attandance as primary predictors of early childhood learning achievement, while demographic factors such as age, gender, and parental occupation have less influence. In this way, the study highlights the need for targeted interventions to improve classroom participation and academic performance.

REFERENSI

[1] L. Bakken, N. Brown, and B. Downing, "Early Childhood Education: The Long-Term Benefits," *Journal of Research in Childhood Education*, vol. 31, no. 2, pp. 255–269, Apr. 2017, doi: 10.1080/02568543.2016.1273285.

[2] K. A. A. Gamage, D. M. S. C. P. K. Dehideniya, and S. Y. Ekanayake, "The Role of Personal Values in Learning Approaches and Student Achievements," *Behavioral Sciences*, vol. 11, no. 7, p. 102, Jul. 2021, doi: 10.3390/bs11070102.

[3] M.-A. Sørlie, K. A. Hagen, and K. B. Nordahl, "Development of social skills during middle childhood: Growth trajectories and school-related predictors," *Int J Sch Educ Psychol*, vol. 9, no. sup1, pp. S69–S87, Dec. 2021, doi: 10.1080/21683603.2020.1744492.

Received: Jan 21st, 2025| Published: March 21st, 2025

DOI 10.30656/jsii.v11i2.9065
p-ISSN: 2406-7768
e-ISSN: 2581-2181

[4] M. Hazegh, "Characteristics of Effective Early Childhood Leaders," in *Research Anthology on Early Childhood Development and School Transition in the Digital Era*, IGI Global, 2022, pp. 862–888. doi: 10.4018/978-1-6684-7468-6.ch043.

[5] Z. Zulkarnaen and Z. Zulfakar, "The Effectiveness of Early Childhood Teachers," *International Journal of Multicultural and Multireligious Understanding*, vol. 8, no. 8, p. 138, Aug. 2021, doi: 10.18415/ijmmu.v8i8.2859.

[6] J. Munir, M. Faiza, B. Jamal, S. Daud, and K. Iqbal, "The Impact of Socio-economic Status on Academic Achievement," *Journal of Social Sciences Review*, vol. 3, no. 2, pp. 695–705, Jun. 2023, doi: 10.54183/jssr.v3i2.308.

[7] M. N. Wangid, "The Role of Self-Motivation in Self-Regulated Learning," *PSIKOPEDAGOGIA Jurnal Bimbingan dan Konseling*, vol. 11, no. 1, p. 14, Jul. 2022, doi: 10.12928/psikopedagogia.v11i1.14175.

[8] T. E. Benjamin, R. G. Lucas-Thompson, L. M. Little, P. L. Davies, and M. A. Khetani, "Participation in Early Childhood Educational Environments for Young Children with and Without Developmental Disabilities and Delays: A Mixed Methods Study," *Phys Occup Ther Pediatr*, vol. 37, no. 1, pp. 87–107, Jan. 2017, doi: 10.3109/01942638.2015.1130007.

[9] J. J. Montroy, R. P. Bowles, L. E. Skibbe, M. M. McClelland, and F. J. Morrison, "The development of self-regulation across early childhood.," *Dev Psychol*, vol. 52, no. 11, pp. 1744–1762, Nov. 2016, doi: 10.1037/dev0000159.

[10] W. H. Saputri and E. Risnawati, "Preparing for the School Readiness of Early Childhood by Enhancing the Well-Being and Family Support," *JPUD - Jurnal Pendidikan Usia Dini*, vol. 18, no. 1, pp. 270–286, Apr. 2024, doi: 10.21009/JPUD.181.19.

[11] M. Alzahrani, M. Alharbi, and A. Alodwani, "The Effect of Social-Emotional Competence on Children Academic Achievement and Behavioral Development," *International Education Studies*, vol. 12, no. 12, p. 141, Nov. 2019, doi: 10.5539/ies.v12n12p141.

[12] P. A. Graziano, R. D. Reavis, S. P. Keane, and S. D. Calkins, "The role of emotion regulation in children's early academic success," *J Sch Psychol*, vol. 45, no. 1, pp. 3–19, Feb. 2007, doi: 10.1016/j.jsp.2006.09.002.

[13] Andy Prasetyo Wati, Jefry Aulia Martha, Buyung Adi Dharma, Annisya' Annisya', Nur Anita Yunikawati, and Lifa Farida Panduwinata, "The Secret to Successful Investment in Children's Education: Active Parental Involvement in Education," *Ekuitas: Jurnal Pendidikan Ekonomi*, vol. 12, no. 1, pp. 97–104, 2024.

[14] B. Vadivel, S. Alam, I. Nikpoo, and B. Ajanil, "The Impact of Low Socioeconomic Background on a Child's Educational Achievements," *Educ Res Int*, vol. 2023, pp. 1–11, Jan. 2023, doi: 10.1155/2023/6565088.

[15] K. Kantova, "Parental involvement and education outcomes of their children," *Appl Econ*, vol. 56, no. 48, pp. 5683–5698, Oct. 2024, doi: 10.1080/00036846.2024.2314569.

[16] M. T. Greenberg *et al.*, "Predicting developmental outcomes at school entry using a multiple-risk model: Four American communities.," *Dev Psychol*, vol. 35, no. 2, pp. 403–417, 1999, doi: 10.1037/0012-1649.35.2.403.

[17] E. S. Dalmaijer *et al.*, "Direct and indirect links between children's socio-economic status and education: pathways via mental health, attitude, and cognition," *Current Psychology*, vol. 42, no. 12, pp. 9637–9651, Apr. 2023, doi: 10.1007/s12144-021-02232-2.

[18] J. Jeong, E. E. Franchett, C. V. Ramos de Oliveira, K. Rehmani, and A. K. Yousafzai, "Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis," *PLoS Med*, vol. 18, no. 5, p. e1003602, May 2021, doi: 10.1371/journal.pmed.1003602.

[19] M. Kayacı Çodur, "Ensemble Machine Learning Approaches for Prediction of Türkiye's Energy Demand," *Energies (Basel)*, vol. 17, no. 1, p. 74, Dec. 2023, doi: 10.3390/en17010074.

[20] C. MacCann, Y. Jiang, L. E. R. Brown, K. S. Double, M. Bucich, and A. Minbashian, "Emotional intelligence predicts academic performance: A meta-analysis.," *Psychol Bull*, vol. 146, no. 2, pp. 150–186, Feb. 2020, doi: 10.1037/bul0000219.

[21] B. Vadivel, S. Alam, I. Nikpoo, and B. Ajanil, "The Impact of Low Socioeconomic Background on a Child's Educational Achievements," *Educ Res Int*, vol. 2023, pp. 1–11, Jan. 2023, doi: 10.1155/2023/6565088.

[22] H. Kang, "The prevention and handling of the missing data," *Korean J Anesthesiol*, vol. 64, no. 5, p. 402, 2013, doi: 10.4097/kjae.2013.64.5.402.

[23] S. Alam, M. S. Ayub, S. Arora, and M. A. Khan, "An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity," *Decision Analytics Journal*, vol. 9, p. 100341, Dec. 2023, doi: 10.1016/j.dajour.2023.100341.

[24] Y. Liu, Y. Wang, and J. Zhang, "New Machine Learning Algorithm: Random Forest," 2012, pp. 246–252. doi: 10.1007/978-3-642-34062-8_32.

[25] T.-H. Lee, A. Ullah, and R. Wang, "Bootstrap Aggregating and Random Forest," 2020, pp. 389–429. doi: 10.1007/978-3-030-31150-6_13.

Received: Jan 21st, 2025| Published: March 21st, 2025

DOI 10.30656/jsii.v11i2.9065
p-ISSN: 2406-7768
e-ISSN: 2581-2181

[26] A. A. Khan, O. Chaudhari, and R. Chandra, “A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation,” *Expert Syst Appl*, vol. 244, p. 122778, Jun. 2024, doi: 10.1016/j.eswa.2023.122778.

[27] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, “Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling,” *J Chem Inf Comput Sci*, vol. 43, no. 6, pp. 1947–1958, Nov. 2003, doi: 10.1021/ci034160g.

[28] H. A. Salman, A. Kalakech, and A. Steiti, “Random Forest Algorithm Overview,” *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, Jun. 2024, doi: 10.58496/BJML/2024/007.

[29] L. K. Shrivastav and R. Kumar, “An Ensemble of Random Forest Gradient Boosting Machine and Deep Learning Methods for Stock Price Prediction,” *Journal of Information Technology Research*, vol. 15, no. 1, pp. 1–19, Nov. 2021, doi: 10.4018/JITR.2022010102.

[30] G.-W. Cha, H.-J. Moon, and Y.-C. Kim, “Comparison of Random Forest and Gradient Boosting Machine Models for Predicting Demolition Waste Based on Small Datasets and Categorical Variables,” *Int J Environ Res Public Health*, vol. 18, no. 16, p. 8530, Aug. 2021, doi: 10.3390/ijerph18168530.

[31] D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, “Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine (LightGBM).,” *Diagnostics (Basel)*, vol. 11, no. 9, Sep. 2021, doi: 10.3390/diagnostics11091714.

[32] G. Seni and J. F. Elder, “Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions,” *Synthesis Lectures on Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 1–126, Jan. 2010, doi: 10.2200/S00240ED1V01Y200912DMK002.

[33] Hafiz Aryan Siregar, Muhammad Zacky Raditya, Aditya Nugraha Yesa, and Inggih Permana, “Comparison of Classification Algorithm Performance for Diabetes Prediction Using Orange Data Mining,” *Indonesian Journal of Data and Science*, vol. 4, no. 3, Jan. 2024, doi: 10.56705/ijodas.v4i3.103.

[34] L. Janowski, J. Nawała, T. Hoßfeld, and M. Seufert, “Experiment Precision Measures and Methods for Experiment Comparisons,” in *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, IEEE, Jun. 2023, pp. 49–54. doi: 10.1109/QoMEX58391.2023.10178599.

[35] D. Gray, D. Bowes, N. Davey, Yi Sun, and B. Christianson, “Further thoughts on precision,” in *15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2011)*, IET, 2011, pp. 129–133. doi: 10.1049/ic.2011.0016.

[36] Y. SUN, A. K. C. WONG, and M. S. KAMEL, “CLASSIFICATION OF IMBALANCED DATA: A REVIEW,” *Intern J Pattern Recognit Artif Intell*, vol. 23, no. 04, pp. 687–719, Jun. 2009, doi: 10.1142/S0218001409007326.

[37] R. Yacouby and D. Axman, “Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models,” in *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 79–91. doi: 10.18653/v1/2020.eval4nlp-1.9.

[38] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” *J Big Data*, vol. 6, no. 1, p. 27, Dec. 2019, doi: 10.1186/s40537-019-0192-5.

Received: Jan 21st, 2025| Published: March 21st, 2025