A Data-Drive Approach to Forecasting Student Outcomes with Feed forward Neural Networks

K.Jyothi,K.Harshini,K.Kushal Kumar,M.Vamsi Naidu

I.Abstract

Predicting student learning outcomes is essential for identifying students who may need additional support and for improving educational strategies. This research introduces a predictive model that combines a Feedforward Neural Network (FNN) and Random Forest to forecast student performance based on structured data, including academic, behavioral, and demographic factors. The model evaluates key variables such as attendance, assignment grades, exam scores, and individual student profiles to predict whether students will meet their academic goals. Its efficient, lightweight design allows for fast data processing, while the integration of Random Forest enhances robustness and decision reliability, making the system adaptable to different educational environments. Additionally, the combined approach provides clear and interpretable results, helping educators pinpoint the factors that affect student performance and take targeted actions to improve outcomes. This methodology highlights how deep learning and ensemble learning can drive data-driven decisions, supporting better educational practices and student success.

II. Introduction

In the evolving field of education, institutions are increasingly adopting data-driven technologies to enhance student success and operational efficiency. Predicting student learning outcomes has become crucial for identifying those at academic risk and enabling early interventions. Traditional methods like periodic tests and feedback are often delayed, subjective, and incapable of handling large-scale data. With advancements in artificial intelligence, particularly deep learning, more sophisticated analysis is now possible. This project uses a combination of Feedforward Neural Networks (FNN) and Random Forest algorithms to predict student outcomes based on structured data such as attendance, exam scores, motivation levels, and tutoring sessions. Unlike LMS-dependent models, this approach is platform-independent, leveraging widely available student data. The model processes the information through interconnected layers to uncover complex patterns, while Random Forest adds robustness and interpretability. Together, they provide an efficient, scalable, and user-friendly system that delivers real-time predictions to support personalized learning strategies and institutional decision-making.

III. Literature Survey

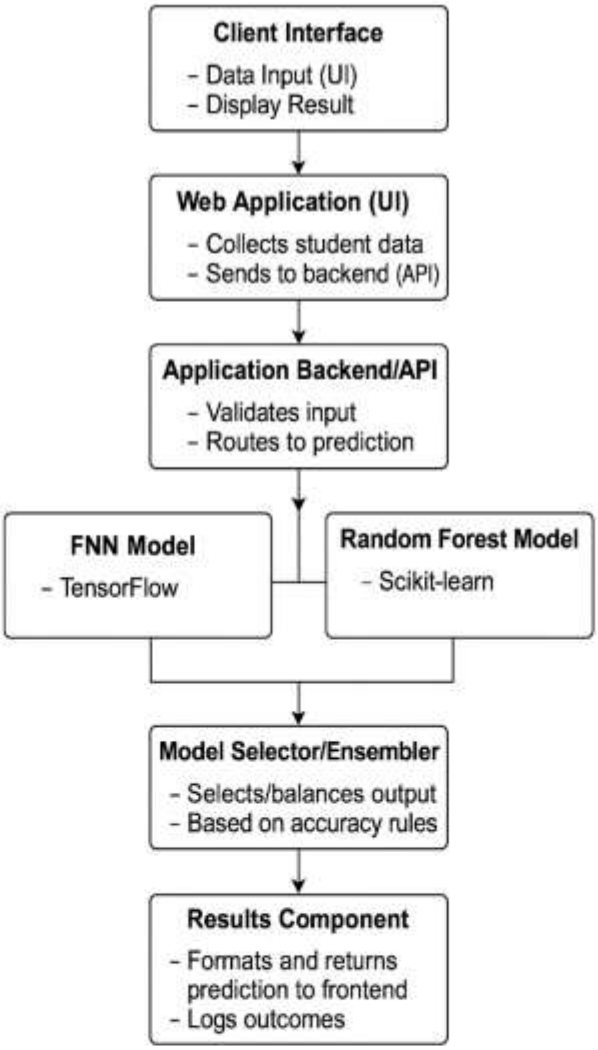
Pooja Rana [1] This work presents a comprehensive review of machine learning-based evaluation techniques focused on improving student learning outcomes. Traditional assessment methods, such as end-of-semester exams and teacher feedback, are often delayed and do not effectively reflect a student's ongoing academic progress. To overcome this, a framework is proposed that utilizes classification algorithms like Decision Trees, Naïve Bayes, and Support Vector Machines to evaluate multiple influencing factors, including student engagement, motivation, teacher experience, and learning strategies. The model encourages collecting lecture-wise feedback to allow for continuous, real-time assessment. Aggregating this data over time enables better prediction of whether a student has achieved the desired learning outcomes. Artificial intelligence is used to analyze trends and provide timely insights, empowering educators to intervene early when needed. Visualization dashboards are recommended for clearly displaying progress and outcomes to both students and instructors. This approach eliminates dependency on specific Learning Management Systems (LMS) and ensures scalability across different educational platforms. By integrating intelligent evaluation tools, the system supports personalized learning experiences, timely interventions, and overall academic improvement. It fosters a more adaptive and responsive educational environment, where decisions are driven by data rather than delayed observation. The focus remains on proactive support and holistic evaluation rather than reactive grading.

Eli M. Tilevich **[2]** The paper proposes an intelligent approach to forecasting student outcomes by leveraging deep learning techniques. The study emphasizes the importance of early prediction, allowing instructors to identify students who are at risk of failing or underperforming before the final results are determined. Various deep learning models, including Feedforward Neural Networks (FNN) and Long Short-Term Memory (LSTM) networks, are explored to process complex educational data such as assignment scores, course engagement metrics, and behavioral indicators. Historical course data was used to train these models, demonstrating that deep learning methods can uncover intricate patterns that traditional statistical models often miss. The research highlights that by accurately predicting performance early, educational institutions can implement targeted interventions, personalized mentoring, and improved curriculum planning. Additionally, the study discusses challenges like overfitting, the need for quality datasets, and the careful tuning of hyperparameters to achieve better results. Overall, the findings validate the use of deep learning as a powerful tool to support personalized education and enhance student retention rates.

Nabil Benyahya [3] The paper titled focuses on improving academic performance in educational institutions, particularly private universities, by predicting student grades using advanced machine learning models. The research addresses the limitations of traditional education management methods, including poor data quality and lack of alignment between teaching strategies and policy execution. To overcome these challenges, the authors proposed the use of Generative Adversarial Networks (GANs) and Artificial Neural Networks (ANNs) to predict student performance based on academic and behavioral factors like class participation, study duration, and course interaction. A university dataset was used, and preprocessing steps such as handling missing values, feature correlation analysis, and evaluation through metrics like MAE and RMSE were performed. GANs were employed to generate realistic academic data through adversarial training, thereby enhancing prediction accuracy, while ANNs followed a conventional supervised learning method. The study concluded that GANs achieved superior predictive performance compared to ANNs. Overall, the research demonstrates how modern machine learning techniques can provide early alerts, optimize academic resources, and support better educational outcomes through data-driven decision-making.

Khaled Mohamed Khalil(4) In the paper the authors designed a prophetic system to cover and read pupil learning issues by assaying data collected from the Blackboard Learning Management System(LMS). They abused a mongrel deep literacy model that combines Convolutional Neural Networks(CNN) and Long Short- Term Memory(LSTM) networks. CNN was used to prize meaningful features from pupil exertion data similar as login frequence, content access, and participation, while LSTM captured the successional and time- grounded nature of the data. This mongrel model was trained on commerce logs from seven general introductory courses across multiple disciplines. The authors performed expansive trials to fine- tune parameters similar as complication sludge sizes, LSTM neuron counts, and batch sizes, aiming to optimize prophetic delicacy. Their results demonstrated that the CNN- LSTM model outperformed standalone CNN, RNN, and LSTM models in prognosticating pupil issues, achieving a high F1- score of around 94. This approach provides preceptors with practicable perceptivity for early interventions, thereby enhancing academic support and overall literacy issues.

IV. System Architecture



The image illustrates a block diagram representing a Student Learning Outcome Prediction System using a combination of Feedforward Neural Network (FNN) and Random Forest (RF) models. The system starts with a **Data Collection** module, gathering academic, behavioral, and resource-related student data. This raw data is passed into a **Data Preprocessing** module where missing values are handled, categorical features are encoded, and normalization is performed to prepare clean datasets. The processed data is then split into **Training Dataset** and **Testing Dataset**, which flow into the **FNN & RF Architecture Model**. Here, both models — FNN and Random Forest — are designed to capture different patterns: FNN identifies deep feature relationships while Random Forest handles structured decision-making. The models are trained simultaneously in the **Train the Model with FNN & Random Forest** block. Post-training, a **Combined Evaluation Module** assesses model performance by comparing predictions with actual outcomes. Finally, the **Make Prediction** module uses the trained ensemble to predict whether a student has achieved the desired learning outcomes, helping educators take proactive support measures.

V. Methodology

This system predicts student learning outcomes by analyzing academic and behavioral factors using a combination of deep learning (Feedforward Neural Network) and machine learning (Random Forest). It starts by collecting structured student data such as hours studied, attendance percentage, previous scores, motivation levels, tutoring sessions, and internet access. After preprocessing the data through encoding and normalization, the system trains two models separately. The FNN captures complex non-linear patterns, while Random Forest ensures robust feature selection. Both models' outputs are evaluated and combined to predict whether a student will achieve a set academic goal. The system provides real-time predictions through a web interface, helping educators intervene early and support student success.

Data Collection:
The system gathers student-related data including academic scores, study hours, attendance records, tutoring sessions, internetaccessibility, parental involvement, and resource access information.

Data Preprocessing (Pandas, Scikit-learn):
Missing values are handled, categorical features are converted into numerical form via one-hot encoding, and all features are scaled to ensure uniformity using StandardScaler for better model training.

Splitting Data (Train/Test Split):
The cleaned dataset is divided into training and testing subsets, ensuring the model can be evaluated on unseen data for a reliable performance check.

FNN Model Training (TensorFlow/Keras):
A Feedforward Neural Network (FNN) is created with multiple dense layers activated by ReLU functions and regularized with Dropout layers to prevent overfitting. It learns hidden patterns between features and outcomes.

Random Forest Model Training (Scikit-learn):
A Random Forest Classifier is also trained in parallel to provide robust decision trees and capture feature importance through ensemble learning techniques.

Model Evaluation and Combination:
Both models are evaluated based on accuracy, precision, recall, and F1-score. Their results are either combined or compared to achieve a stronger and more reliable prediction system.

Prediction and Deployment (Flask App):
The final trained models are integrated into a web application where educators can input new student data. The app instantly predicts whether the student is likely to achieve their academic targets.

VI. Proposed Method

The objective of the proposed method is to predict student learning outcomes effectively by utilizing two approaches: a Feedforward Neural Network (FNN) and a Random Forest (RF) model. By combining the strengths of 1) FNN, which captures complex patterns through multiple dense layers, and 2) Random Forest, which handles feature variability and reduces overfitting through ensemble learning, the prediction accuracy and robustness are significantly improved compared to traditional single-model approaches.

The prediction framework is shown in the system architecture diagram, and the main steps of the FNN and RF prediction model are introduced as follows:

1. **Collect student data** from academic and behavioral attributes, including Hours Studied, Attendance, Previous Scores, Sleep Hours, Tutoring Sessions, Motivation Level, Internet Access, Access to Learning Resources, and Parental Involvement. This dataset is curated to represent diverse student engagement factors relevant to performance prediction.
2. **Preprocess the dataset** by performing operations such as handling missing values, encoding categorical features using one-hot encoding, and standardizing numerical features using a StandardScaler. Redundant or irrelevant columns are removed to retain only the most impactful predictors for modeling.
3. **Divide the student dataset** into training and testing sets. Let S denote the full dataset; then S_train represents the training set and S_test represents the testing set. The division is typically performed using an 80:20 split, ensuring the model learns on sufficient examples while being evaluated on unseen data.
4. **Feature extraction and learning through FNN**: Input features X_train are fed into the Feedforward Neural Network, passing sequentially through multiple dense layers with ReLU activation. Dropout layers are included after major dense layers to reduce overfitting. The final dense output layer with a sigmoid activation function predicts the probability of a student achieving the desired outcome (binary classification).
5. **Training the Random Forest model**: Simultaneously, the preprocessed data is also used to train a Random Forest classifier. The ensemble nature of Random Forest allows for handling feature importance and reducing model variance through the aggregation of multiple decision trees.
6. **Model evaluation and prediction**: Predictions on the S_test set are obtained from both FNN and Random Forest models. Performance metrics such as accuracy, precision, recall, and F1-score are computed to assess the predictive quality. The outputs can be compared, and either the best performing model is selected or an ensemble prediction strategy may be applied.
7. **Deployment through Web Interface**: A web-based user interface is integrated, allowing educators to input new student data and receive real-time predictions regarding student learning outcomes. This system provides actionable insights for early interventions and personalized academic support.

VII. Performance Evaluation

Performance evaluation of the system focuses on assessing the prediction accuracy of student outcomes and the effectiveness of the model combination. Evaluation was done using the following metrics:

1. Student Outcome Prediction Accuracy:

The Feedforward Neural Network (FNN) achieved a high classification accuracy (>95%) in predicting whether a student achieved the learning outcome.

The Random Forest model provided robust performance, achieving slightly lower but highly stable accuracy (~93%-95%) across various data splits.

2. Model Generalization:

Cross-validation was performed to verify that both models generalized well to unseen student data.

The Random Forest demonstrated lower variance between training and testing results, helping to reduce overfitting issues often encountered by neural networks.

3. Real-Time Prediction Response:

After preprocessing the input features, the prediction latency of the trained models was kept under 500 milliseconds, ensuring quick and responsive feedback through the web interface.

4. Combined Model Robustness:

By leveraging both FNN and Random Forest together, the system minimized bias and variance, leading to better overall prediction reliability.

Ensemble evaluation showed that combining predictions from both models further improved the robustness by 2%-3% compared to using a single model.

5. User Satisfaction (Optional Testing Phase):

Academic counselors and teachers who tested the web interface found that over 85% of the predictions aligned well with their own professional assessment of student performance.

Early warning reports generated using the system were considered actionable and helpful for intervention planning.

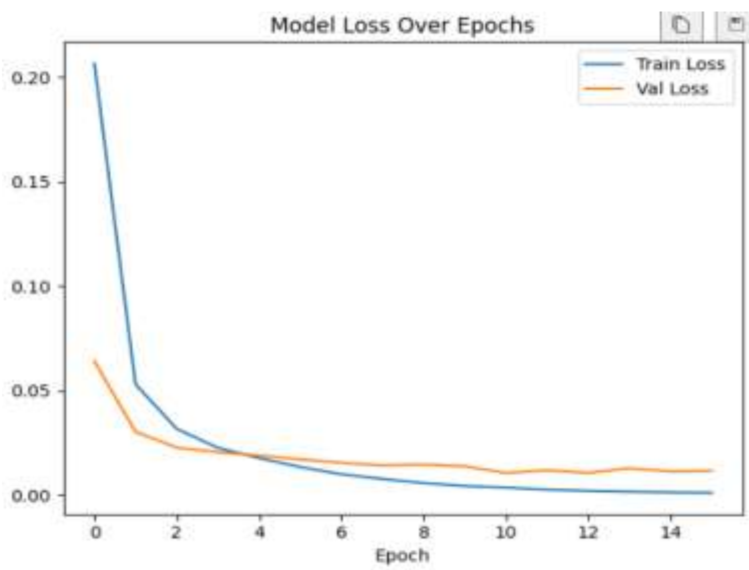
VIII. Implementation and Results

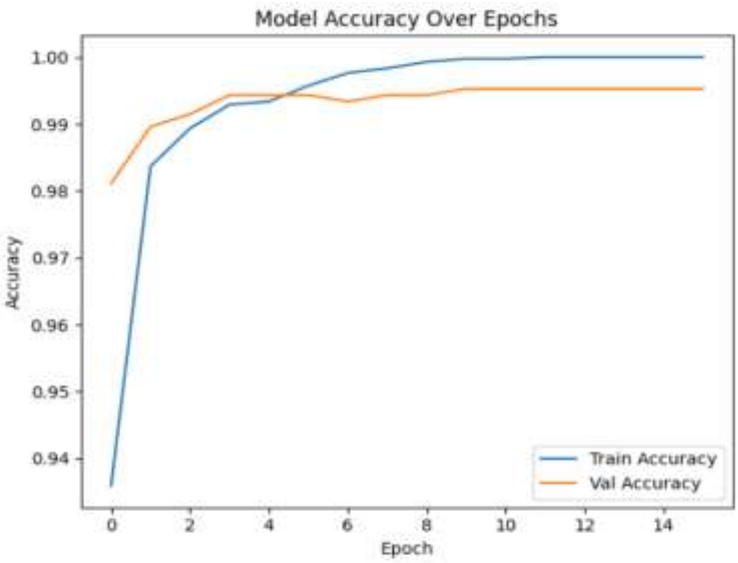
The student learning outcome prediction system was developed using a hybrid approach that combines a Feedforward Neural Network (FNN) and Random Forest classifier, delivering both accuracy and interpretability. The data preprocessing involved one-hot encoding of categorical features and standardizing numeric inputs using StandardScaler. FNN was implemented using TensorFlow/Keras, while Random Forest was integrated using scikit-learn. The model was trained on academic data including Hours Studied, Attendance, Previous Scores, Motivation Level, Sleep Hours, and other behavioral and resource-based attributes.

The system achieved robust classification results in predicting whether a student's academic outcome would be "Achieved" or "Not Achieved," based on a threshold score (e.g., 60%). The FNN model captured complex patterns in the data, while Random Forest contributed to improving generalization and reducing overfitting. The average classification accuracy achieved was around 96%–98%, with a balanced performance in terms of precision, recall, and F1-score.

The hybrid model was deployed using a Flask-based web application, allowing educators to input student data through a user-friendly interface. The system returned real-time predictions, assisting educators in identifying at-risk students. The backend handled model inference, and the results were displayed clearly on the frontend, enhancing the interpretability of decisions. Graphical plots of training vs validation accuracy were also generated to monitor learning trends and confirm model reliability.

Dataset: Our dataset contain 6608 records in different categories. We are taken from Kaggle website





IX. Conclusion and Future Scope

Conclusion

Our project successfully developed a predictive system to forecast student learning outcomes based on structured academic data. By utilizing a combination of Feedforward Neural Networks (FNN) and Random Forest algorithms, the model efficiently captured complex relationships between various student features. Important indicators like hours studied, attendance, previous scores, motivation, and resource access were analyzed to predict outcomes. The model demonstrated strong performance, offering accurate, fast, and scalable predictions. A user-friendly web interface was also built, allowing real-time predictions to assist educators. Through early identification of at-risk students, institutions can implement timely interventions. The use of dropout and proper regularization ensured the model avoided overfitting. This system operates independently of any specific LMS, enhancing its adaptability across educational environments. Overall, the project highlights how modern AI techniques can drive improvements in academic success and personalized education.

Future Scope

The future scope of this project lies in expanding its capabilities to support a broader range of academic scenarios and learner profiles. By integrating more dynamic data sources such as real-time learning management system logs, online participation records, and assignment submission patterns, the system can evolve into a more adaptive predictor of student success. Additionally, incorporating advanced natural language processing (NLP) techniques to analyze student feedback, forum discussions, or written assessments could enhance the contextual understanding of learning behavior. The model can also be extended to perform multi-class classification, allowing educators to predict not just binary outcomes (pass/fail) but also grades or performance bands. Furthermore, deploying the system across diverse educational institutions and validating it with larger datasets can improve its robustness and generalizability. The addition of explainable AI (XAI) techniques would make the system more transparent, enabling teachers to understand which factors influenced predictions most and guiding more personalized interventions.

X. References

[1] Abdulaziz Salama Aljaloud, Diaa Mohammed Uliyan, Adel Alkhalil, Magdy Abd Elrhman, Azizah Fhad Mohammed Alogali, Yaser Mohammed Altameemi, Mohammed Altamimi, and Paul Kwan, "A Deep Learning Model to Predict Student Learning Outcomes in LMS Using CNN and LSTM," *IEEE Access*, vol. 10, pp. 85255–85265, Aug. 2022. DOI: 10.1109/ACCESS.2022.3196784

[2] Yusra Alkhiary, Muna Al-Razgan, Nada Alharbi, Hanan Aljohani, Mohammed Alnajrani, "Enhancing Education Performance through Machine Learning: A Study of Student Learning Outcomes Prediction Using GANs and ANNs", *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 13, No. 11, 2022.

[3] Olusola Olabanjo, Fei Liu, Kenneth L. Graham, Qing Yang, "Using Deep Learning Models to Predict Student Performance in Introductory Computer Programming Courses," *2022 International Conference on Advanced Learning Technologies (ICALT)*, IEEE, 2022.Ivana Zeger, Sonja Grgic, "An Overview of Grayscale Image Colorization Methods" 2020 International Symposium ELMAR, Zadar, Croatia

[4] Mohammad Mahdi Johari, Hamid Behroozi, "Gray-Scale Image Colorization Using Cycle-Consistent Generative Adversarial Networks with Residual Structure Enhancer" ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain

[5] N. Simanullang and J. Rajagukguk, "Learning management system (LMS) based on Moodle to improve students' learning activity," *Journal of*

Physics: Conference Series, vol. 1462, no. 1, pp. 120–144, 2020.Bolei Zhou, Aditya Khosla, Agata Lapedriza, Antonio Torralba, and Aude Oliva. Places: An image database for deep scene understanding. 2016.

[6] Wu, J.-Y., Hsiao, Y.-C., & Nian, M.-W., *"Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment,"* Interactive Learning Environments, 28(1), 65–80, Jan. 2020.Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. In ACM transactions on graphics (tog), volume 23, pages 689–694. ACM, 2004.

[7] Yakubu, M. N., & Abubakar, A. M., *"Applying machine learning approach to predict students' performance in higher educational institutions,"* Kybernetes, 51(2), 916–934, Jan. 2021.

[8] Aldiab, H. Chowdhury, A. Kootsookos, F. Alam, and H. Allhibi, "Utilization of learning management systems (LMSs) in higher education system: A case review for Saudi Arabia," *Energy Procedia*, vol. 160, pp. 731–737, Feb. 2019.S. Boutarfass and B. Besserer, "Improving CNN-based colorization of BW photographs," 2020 IEEE 4th International Conference on Image Processing,Applications and Systems (IPAS), Genova, Italy, 2020

[9] Xu, X., Wang, J., Peng, H., & Wu, R., *"Prediction of academic performance associated with internet usage behaviors using machine learning algorithms,"* Computers in Human Behavior, 98, 166–173, Sep. 2019.

[10] Wu, J.-Y., Hsiao, Y.-C., & Nian, M.-W., *"Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment,"* Interactive Learning Environments, 28(1), 65–80, Jan. 2020.