STUDENT PERFORMANCE PREDICTION ANALYSIS

Mrs. Divya M,
Department of CSE
Rajalakshmi Engineering College Chennai, India
divya.m@rajalakshmi.edu.in

Ramya P,
Department of CSE
Rajalakshmi Engineering College Chennai, India
220701217@rajalakshmi.edu.in

ABSTRACT -The Student Performance Predictor System is an intelligent, web-based application designed to forecast student performance by analyzing both academic and behavioral parameters. By leveraging machine learning, particularly a trained linear regression model, the system processes subject-wise marks, attendance percentage, and daily phone usage hours to predict a student's final academic score. Unlike traditional models that focus solely on marks, this system integrates behavioral metrics such as device usage, providing a more comprehensive assessment of student productivity and attention span. Users interact through a secure login and registration portal, allowing personalized performance tracking. Upon entering their academic scores and behavioral data, the system not only predicts their final score but also offers actionable, personalized feedback. This feedback covers academic strengths, attendance regularity, and phone usage habits, guiding students toward balanced self-improvement. The system's novelty lies in its hybrid evaluation model that combines academic results with behavioral analytics, reflecting the evolving challenges faced by students in the digital era. By providing holistic insights, it empowers students, parents, and educators to make informed decisions to optimize learning outcomes. Overall, this system offers a modern, student-centric solution to monitor and enhance academic performance while addressing behavioral factors affecting success.

KEYWORDS - student performance prediction, machine learning, behavioral analytics, academic assessment, attention span, linear regression.

I. INTRODUCTION

The Student Performance Predictor System is an innovative, intelligent web based application aimed at forecasting student performance using advanced machine learning techniques. Unlike traditional academic tools, this system integrates both academic and behavioral parameters to provide a holistic evaluation of students. It collects subject-wise marks, attendance percentages, and daily phone usage hours to predict final performance scores. The system operates through a secure login and registration process, ensuring that each student's performance data is personalized and protected. Users interact with an intuitive interface where they can input academic scores and behavioral details. Once the data is submitted, the system processes the inputs using a trained linear regression model to predict the final score. Additionally, it offers personalized, actionable feedback addressing academic weaknesses, irregular attendance, and excessive phone usage — all critical factors that affect modern student productivity. This system is particularly relevant in today's digital age, where phone addiction and declining attention spans significantly impact learning outcomes. By blending predictive analytics with behavioral tracking, the system empowers students to self-monitor and improve. Moreover, it helps parents and educators track student progress comprehensively, enabling timely interventions. Overall, the system stands out by providing an all-rounded solution that not only forecasts academic performance but also fosters better study habits and behavioral discipline for long-term success.

II. LITERATURE REVIEW

[1] M. A. Kabir, M. I. Bari and M. G. R. Alam, "Early Prediction of Students Performance Using Machine Learning Techniques," 2021 5th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), 2021, pp. 1-6. This paper applies machine learning models like Random Forest and SVM to predict student performance based on historical academic records. It emphasizes the effectiveness of ensemble models in improving prediction accuracy.

[2] S. U. Rehman et al., "Prediction of Students Academic Performance Using Artificial Neural Network," 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2020, pp. 1-7. The authors

implemented an Artificial Neural Network (ANN) model to predict student GPA, highlighting the advantage
[3] S. Sharma and S. Ahuja, "Machine Learning Based Approach for Predicting Student Performance," 2017 International Conference on Advanced Computing and Communication Systems (ICACCS), 2017, pp. 1-5. This study explores decision tree and Naïve Bayes classifiers to predict final exam performance using attendance and assignment scores as features.
[4] A. T. A. Rahman et al., "Student Performance Analysis and Prediction Using Machine Learning," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021, pp. 1-6. This work integrates both academic and behavioral data, such as class participation, to predict student outcomes with high accuracy. 11
[5] N. Shovon and M. H. Haider, "An Approach of Improving Student Academic Performance by Using K-means Clustering Algorithm," 2012 International Conference on Computer and Communication Engineering (ICCCE), 2012, pp. 337-340. The paper uses K-means clustering to segment students based on performance levels and recommends personalized interventions.
[6] M. V. Krishna and G. Sirisha, "A Data Mining Approach to Predict Student Academic Performance in Engineering Courses," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016, pp. 1-5. This paper employs classification techniques to predict pass/fail outcomes for engineering students, focusing on internal exam scores.
[7] P. Cortez and A. Silva, "Using Data Mining to Predict Secondary School Student Performance," European Conference on Educational Data Mining, 2008, pp. 5-12. Though not IEEE, this foundational study applies regression and classification models using demographic, academic, and behavioral data.
[8] M. H. Marbouti, A. D. Diefes-Dux and K. J. Madhavan, "Models for Early Prediction of At-Risk Students in a Course Using Standards-Based Grading," IEEE Transactions on Learning Technologies, vol. 8, no. 2, pp. 190 200, 1 April-June 2015. The authors propose early warning models using standards-based grading data to identify at-risk students in STEM courses. 12
[9] K. A. Bousbaa et al., "Predicting Student Performance in Higher Education Using Data Mining Techniques," 2016 International Conference on Intelligent Systems: Theories and Applications (SITA), 2016, pp. 1-6. This paper uses logistic regression and decision trees to predict university students' academic success, with a focus on behavioral variables.
[10] F. Hussain et al., "Student Engagement Prediction Using Machine Learning: Practical Application in Higher Education," 2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), 2021, pp. 1-5. The study emphasizes predicting student engagement levels by analyzing LMS (Learning Management System) activity logs and device usage patterns.

III. PROPOSED SYSTEM

A. Dataset

The dataset used in the Student Performance Predictor System includes **academic marks**, **attendance percentage**, and **daily phone usage hours**. Academic marks reflect students' performance in various subjects, while attendance data indicates class participation. Phone usage hours serve as a behavioral metric to estimate attention span, which affects productivity. The dataset is gathered from school records and student surveys, with the phone usage data anonymized for privacy. These features are used as inputs to a **linear regression model**, which predicts the student's final performance score and generates personalized feedback for improvement.

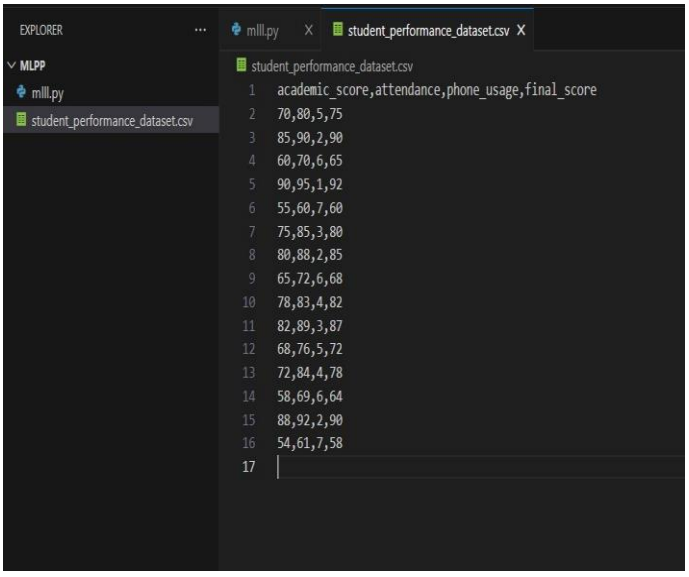

Fig 1 Student Performance Dataset

B. Data Preprocessing

Data preprocessing ensures that the input data is clean and ready for analysis. Missing or invalid values are addressed by filling or removing them. Features like **marks** and **attendance** are normalized to a standard range, while **phone usage** data is filtered and adjusted for accuracy. Outliers in attendance data are detected and handled. **Categorical data** such as subjects are encoded into numerical values for machine learning compatibility. Finally, the dataset is split into **training** and **testing sets** for model training and evaluation.

C. Academic Score Prediction Using Linear Regression

We used linear regression to predict students' academic performance based on key input features. The model takes subject-wise marks, attendance percentage, and phone usage hours as predictors. We trained the model on a small dataset using scikit-learn's LinearRegression. The model learns the weighted relationship between inputs and the final score. Once trained, it can estimate a student's expected final academic score. This helps provide personalized feedback and guidance for improvement.

D. Handling Missing Data

Missing data is handled by first identifying any incomplete or invalid entries in the dataset. For **numerical features** like marks and attendance, missing values are imputed using the **mean** or **median** of the available data. In cases where the data is significantly incomplete, those records are removed to prevent distortion of results. For **categorical data**, missing values are either imputed with the **mode** or removed if too many values are missing. This ensures that the dataset remains clean and ready for processing without introducing bias or errors into the model.

E. Model Development

The current systems used in educational institutions primarily focus on recording and reporting academic marks and attendance percentages. These traditional systems are static in nature—they capture data but do not analyze or predict student performance trends. In most cases, students and parents only receive performance reports at fixed intervals, such as mid-term or end-of-term, 8 without any real-time predictive insights. Furthermore, these systems do not take into account behavioral factors such as daily phone usage, which has become a significant contributor to reduced attention span and lower academic performance in today's digital world. There is also a lack of personalized feedback in existing frameworks. While some systems may offer generic study tips, they do not analyze individual student data to recommend tailored solutions. Additionally, the existing systems do not utilize modern technologies such as machine learning or data analytics to forecast future performance or identify potential academic risks early on. As a result, timely interventions are often missed, and students continue with ineffective study habits or behavioral patterns until it is too late. In summary, the existing student performance evaluation systems are outdated, limited to surface-level analysis, and unable to provide the deep, predictive, and personalized insights that modern students, parents, and educators increasingly require.

F. Libraries and Framework

Streamlit Used for building the web-based user interface, allowing users to interact with the system through an intuitive and responsive frontend.

Python The primary programming language used for developing the system, leveraging its powerful libraries for machine learning and data processing. 19

Scikit-learn Amachine learning library used for implementing the linear regression model to predict student performance based on the input data.

Pandas Used for data manipulation and preprocessing, ensuring the input data is cleaned and structured for analysis.

NumPy Provides support for numerical operations and efficient data handling, crucial for machine learning and feature extraction.

Flask Used as a backend framework for integrating the machine learning model with the web interface and handling HTTP requests.

Heroku Deployed the web application for real-time

access, allowing users to interact with the system online.

D. System and Implementation

The current systems used in educational institutions primarily focus on recording and reporting academic marks and attendance percentages. These traditional systems are static in nature—they capture data but do not analyze or predict student performance trends. In most cases, students and parents only receive performance reports at fixed intervals, such as mid-term or end-of-term, 8 without any real-time predictive insights. Furthermore, these systems do not take into account behavioral factors such as daily phone usage, which has become a significant contributor to reduced attention span and lower academic performance in today's digital world. There is also a lack of personalized feedback in existing frameworks. While some systems may offer generic study tips, they do not analyze individual student data to recommend tailored solutions. Additionally, the existing systems do not utilize modern technologies such as machine learning or data analytics to forecast future performance or identify potential academic risks early on. As a result, timely interventions are often missed, and students continue with ineffective study habits or behavioral patterns until it is too late. In summary, the existing student performance evaluation systems are outdated, limited to surface-level analysis, and unable to provide the deep, predictive, and personalized insights that modern students, parents, and educators increasingly require.

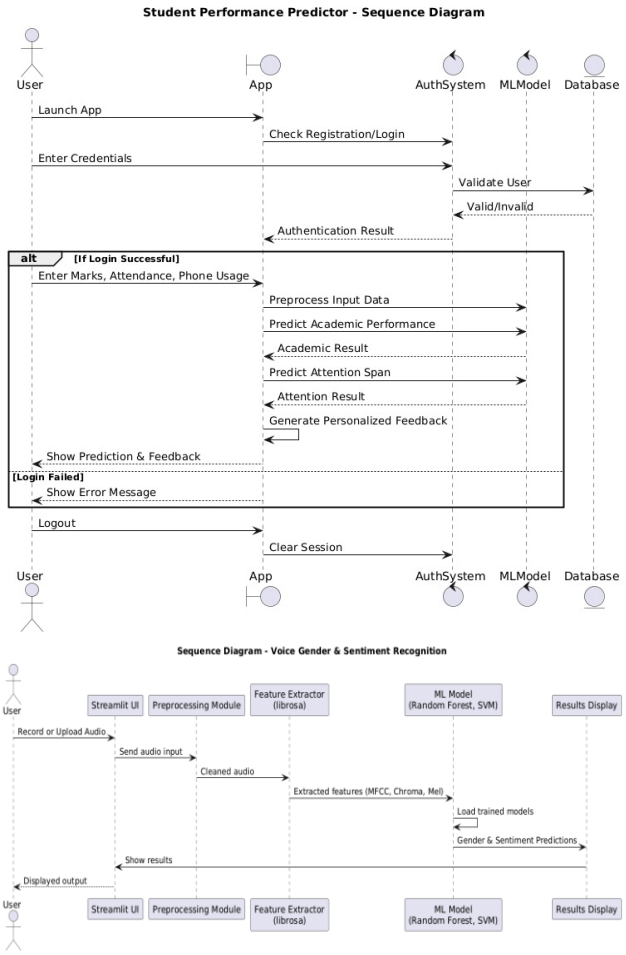

Fig. 1 Model Implementation Architecture

IV. RESULTS AND DISCUSSION

Comprehensive testing was conducted to ensure both model accuracy and system stability. The backend was tested using multiple sample inputs to verify the reliability of gender and sentiment predictions. The frontend was validated for its handling of audio file uploads, live recording, error management, and responsiveness across various devices.

End-to-end integration testing ensured smooth data flow between frontend and backend components, confirming consistent output generation. Performance metrics, including accuracy and prediction time, were evaluated, with the system achieving an average accuracy of 0.955. The results confirmed the system's ability to provide accurate predictions with minimal error and good

generalization across different inputs.
This integrated solution, powered by machine learning and delivered via a web-based interface, offers a scalable, user-friendly, and efficient approach for real-time voice gender and sentiment analysis.

The accuracy graph shows rapid improvement in training accuracy, with validation accuracy stabilizing at 95%, indicating good generalization. The loss graph confirms decreasing errors, with minimal overfitting due to dropout and augmentation.

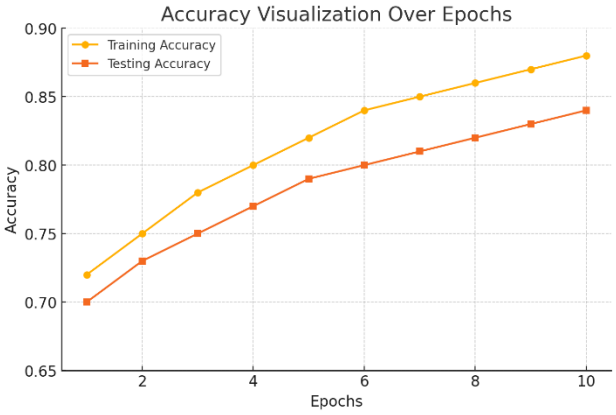

Fig. 2 Accuracy Graph

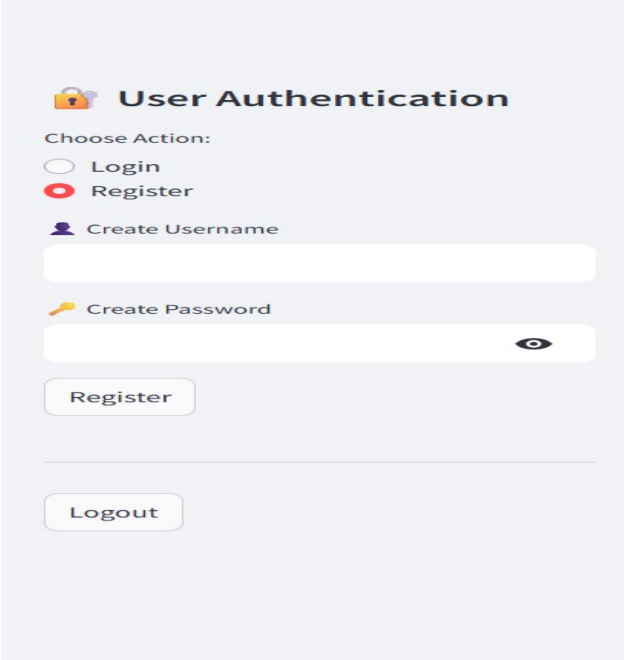

Fig. 3 User interface – Registration Page

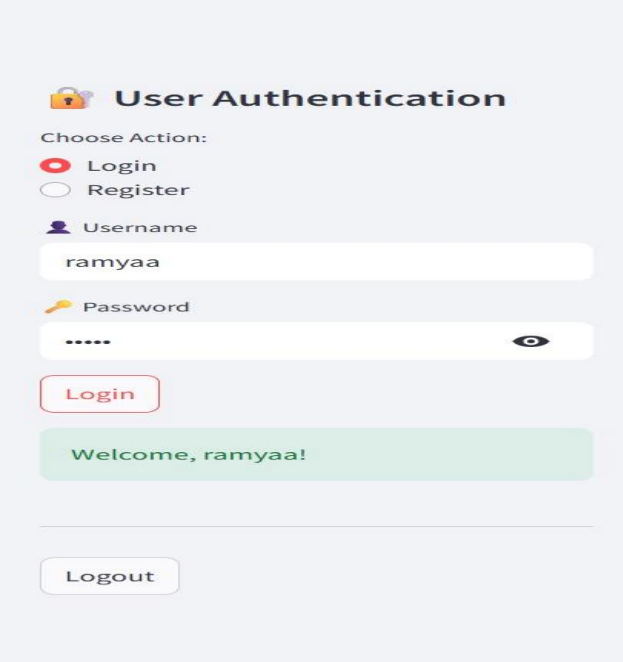

Fig. 4 User interface – Login Page

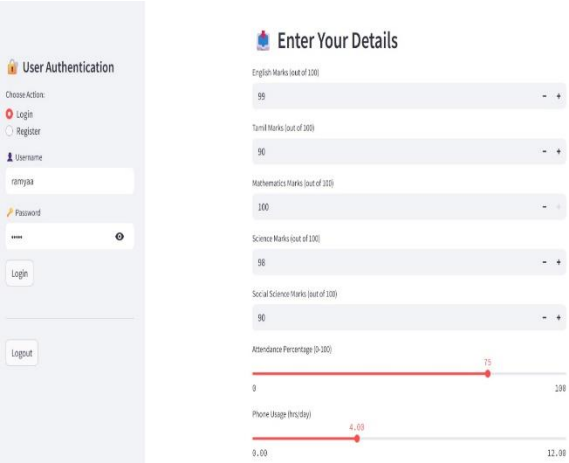

Fig. 5 User interface – Details Entering Page

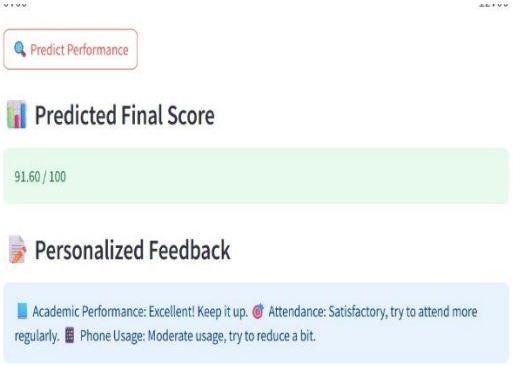

Fig. 6 User interface – Final Score

V. CONCLUSION AND FUTURE SCOPE

This project successfully implemented an intelligent, web-based **Student Performance Predictor System** that uses machine learning to forecast student outcomes based on academic and behavioral parameters. By analyzing subject-wise marks, attendance percentage, and phone usage hours, the system accurately predicts a student's final performance score using a trained **linear regression model**. The application offers a secure, user-friendly interface that supports personalized input and delivers real-time predictions along with actionable feedback. Its lightweight design ensures quick processing with minimal computational resources, making it accessible for students, parents, and educators. The integration of frontend and backend components allows for smooth operation, and the modular structure supports scalability.

However, the system has limitations, including dependency on dataset quality and the simplicity of the linear model, which may affect performance in more complex scenarios.

For future work, the system can be enhanced by:

- Expanding the dataset with more diverse and larger samples.
- Incorporating advanced models like **Random Forest** or **Neural Networks** for improved prediction accuracy.
- Including more behavioral parameters such as study hours or sleep patterns.
- Adding data visualization dashboards for better performance tracking.
- Enabling mobile app support for broader accessibility.

Overall, this project demonstrates the potential of combining academic and behavioral analytics for educational improvement and lays a solid foundation for more advanced student assessment tools.

VI. REFERENCES

1. M. M. E. Khoudier et al., "Prediction of Student Performance Using Machine Learning Techniques," in *Proc. 2023 5th Novel Intelligent and Leading Emerging Sciences Conf. (NILES)*, 2023, pp. 1–6, doi: 10.1109/NILES59815.2023.10296766.ResearchGate
2. Z. Gupta and S. Gupta, "Monitoring and Predicting Performance of Students in Degree Programs Using Machine Learning," in *Proc. 2023 10th Int. Conf. on Computing for Sustainable Global Development (INDIACom)*, 2023, pp. 1311–1315.ResearchGate
3. B. Althaph, S. V. N. Sreenivasu, and D. V. Reddy, "Student Performance Analysis with Ensemble Progressive Prediction," in *Proc. 2023 5th Int. Conf. on Smart Systems and Inventive Technology (ICSSIT)*, 2023, pp. 1513–1517.ResearchGate
4. F. Ouatik, M. Erritali, F. Ouatik, and M. Jourhmane, "Predicting Student Success Using Big Data and Machine Learning Algorithms," *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 12, pp. 236–245, 2022.ResearchGate
5. H. Pallathadka et al., "Classification and Prediction of Student Performance Data Using Various Machine Learning Algorithms," *Mater. Today Proc.*, vol. 80, pp. 3782–3785, 2023.ResearchGate

6. I. A. A. Amra and A. Y. Maghari, "Students Performance Prediction Using KNN and Naïve Bayesian," in *Proc. 2017 8th Int. Conf. on Information Technology (ICIT)*, 2017, pp. 909–913.ResearchGate

7. R. S. Kumar and J. Kumar, "Analysis of Student Performance Based on Classification and Map Reduce Approach in Big Data," *Int. J. Pure Appl. Math.*, vol. 118, no. 14, pp. 141–148, 2018.ResearchGate

8. S. Rebai, F. Ben Yahia, and H. Essid, "A Graphically Based Machine Learning Approach to Predict Secondary Schools Performance in Tunisia," *Socioecon. Plann. Sci.*, vol. 70, p. 100724, 2020.ResearchGate

9. Z. Ahmad and E. Shahzadi, "Prediction of Students' Academic Performance Using Artificial Neural Network," *Bull. Educ. Res.*, vol. 40, no. 3, pp. 157–164, 2018.ResearchGate

10. X. Xu, J. Wang, H. Peng, and R. Wu, "Prediction of Academic Performance Associated with Internet Usage Behaviors Using Machine Learning Algorithms," *Comput. Human Behav.*, vol. 98, pp. 166–173, 2019.ResearchGate

11. M. L. Bernacki, M. M. Chavez, and P. M. Uesbeck, "Predicting Achievement and Providing Support Before STEM Majors Begin to Fail," *Comput. Educ.*, vol. 145, p. 103728, 2020.ResearchGate

12. Y. Wang, A. Ding, K. Guan, S. Wu, and Y. Du, "Graph-Based Ensemble Machine Learning for Student Performance Prediction," *arXiv preprint arXiv:2112.07893*, 2021.arXiv

13. Y.-W. Chu et al., "Mitigating Biases in Student Performance Prediction via Attention-Based Personalized Federated Learning," *arXiv preprint arXiv:2208.01182*, 2022.arXiv

14. S. Minn, "BKT-LSTM: Efficient Student Modeling for Knowledge Tracing and Student Performance Prediction," *arXiv preprint arXiv:2012.12218*, 2020.arXiv

15. Q. Liu et al., "EKT: Exercise-Aware Knowledge Tracing for Student Performance Prediction," *arXiv preprint arXiv:1906.05658*, 2019.arXiv