International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: https://ijwos.com
Volume 3 Issue 7 , July 2026
E-ISSN : 3049-2424
DOI: https://doi.org/10.71366/ijwos

An Intelligent Framework for Student Performance Analysis and Prediction Using Machine Learning and Educational Data Mining

Abhishek R[1], Dr. H Girisha[2]
[1] MTech, CSE, Rao Bahadur Y. Mahabaleshwarappa Engineering College (RYMEC), Ballari, Karnataka, India
[2] Professor & Head of Department, CSE, Rao Bahadur Y. Mahabaleshwarappa Engineering College (RYMEC), Ballari, Karnataka, India.

Article Info

Article History:

Published: 18 July 2026

Publication Issue:
Volume 3, Issue 7
July-2026

Page Number:
367-375

Corresponding Author:
Abhishek R

Abstract:

Educational institutions generate a vast amount of academic data through examinations, assignments, attendance records, internal assessments, and semester evaluations. Analyzing this data manually to identify students who may require academic support is a complex and time-consuming process. Traditional methods primarily rely on statistical analysis and human judgment, which often fail to identify hidden patterns influencing student performance. The rapid advancement of Machine Learning (ML) and Educational Data Mining (EDM) has created opportunities to develop intelligent systems capable of predicting academic outcomes and assisting educators in making informed decisions. Accurate prediction of student performance enables institutions to provide timely interventions, improve learning outcomes, and enhance overall educational quality.

This paper presents An Intelligent Framework for Student Performance Analysis and Prediction Using Machine Learning and Educational Data Mining, a web-based system designed to analyze historical student academic records and predict future performance. The proposed framework utilizes multiple academic attributes, including student marks, attendance, grades, SGPA, credits earned, and pass/fail status, to evaluate academic progress and identify performance trends. The collected data undergoes preprocessing techniques such as data cleaning, missing value handling, feature selection, and normalization to improve the quality of the dataset before model development. Exploratory Data Analysis (EDA) is performed using statistical methods and graphical visualizations to understand relationships among different academic factors and identify key indicators affecting student success.

The prediction framework employs two supervised machine learning algorithms to perform different prediction tasks. A Random Forest Classifier is implemented to predict whether a student is likely to pass or fail based on academic performance indicators, while a Linear Regression model is used to estimate future SGPA values. The system is developed using Python, Streamlit, MongoDB, Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn, providing an interactive platform for academic data analysis, visualization, and prediction. The generated results are presented through intuitive dashboards, charts, and performance reports that enable educators to monitor student progress efficiently and identify academically weak students at an early stage.

Keywords: Intelligent Framework

1. Introduction

Education plays a vital role in the development of individuals and society, making continuous assessment of student learning an essential part of the academic process. Educational institutions collect large amounts of data through examinations, assignments, attendance records, laboratory performance, internal assessments, and semester

evaluations. This academic data contains valuable information that can be utilized to understand student learning behavior, evaluate academic progress, and improve the overall quality of education. However, extracting meaningful insights from such large and complex datasets using conventional methods remains a significant challenge.

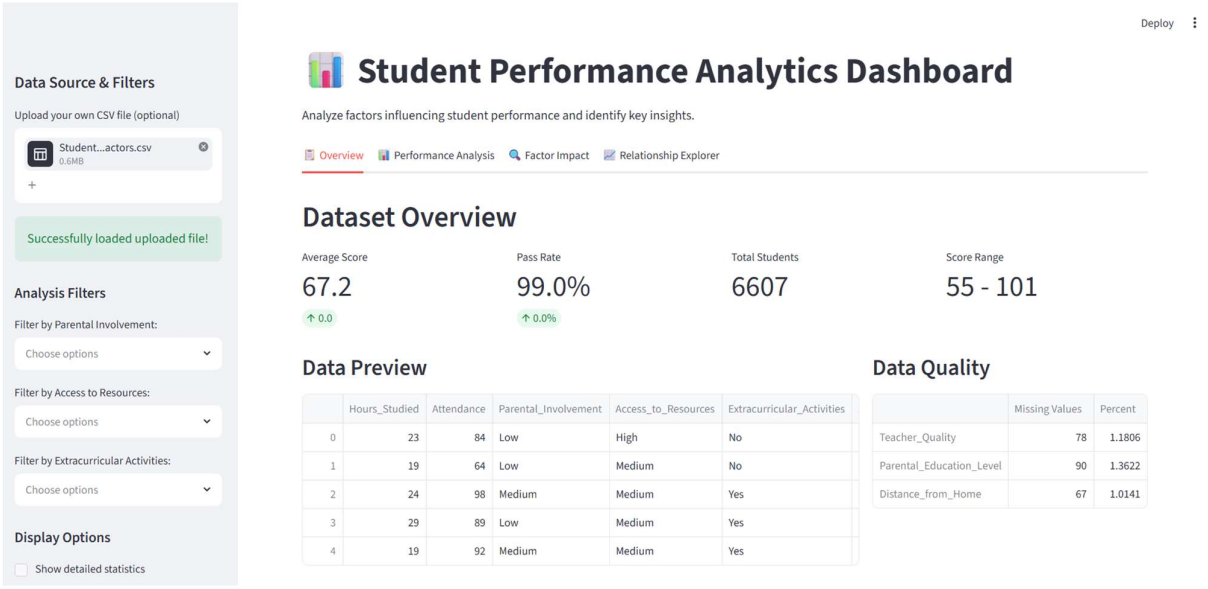
Traditional approaches to evaluating student performance mainly rely on manual analysis, statistical reports, and educators' observations. Although these methods provide basic information about students' academic achievements, they often fail to identify hidden patterns and relationships among multiple academic factors. As a result, students who are academically at risk may not be identified at an early stage, reducing the opportunity for timely intervention and personalized academic support. The increasing volume of educational data has therefore created a demand for intelligent analytical systems capable of transforming raw academic information into meaningful knowledge.

Recent advances in **Educational Data Mining (EDM)**, **Artificial Intelligence (AI)**, and **Machine Learning (ML)** have introduced new possibilities for analyzing educational datasets and predicting student performance with greater accuracy. Machine learning algorithms can automatically learn from historical academic records, recognize complex relationships among educational attributes, and generate reliable predictions regarding future academic outcomes. These predictive models assist educators in identifying low-performing students, understanding the factors affecting academic success, and implementing effective intervention strategies to improve learning outcomes.

The proposed research introduces **An Intelligent Framework for Student Performance Analysis and Prediction Using Machine Learning and Educational Data Mining**, which integrates educational data analysis, visualization, and predictive modeling into a unified web-based platform. The framework analyzes student information such as attendance, marks, grades, SGPA, credits earned, and pass/fail status to evaluate academic performance and generate predictive insights. By applying data preprocessing, exploratory data analysis, and supervised machine learning algorithms, the system provides accurate predictions and meaningful visual reports that support educators in making informed academic decisions.

The proposed framework employs a **Random Forest Classifier** to predict student pass/fail outcomes and a **Linear Regression** model to estimate future SGPA values. These models are trained using historical student records after performing preprocessing techniques such as data cleaning, normalization, feature selection, and missing value handling. The predicted results, together with graphical visualizations and statistical analysis, enable institutions to monitor academic progress, identify students requiring additional support, and enhance educational planning through data-driven decision-making.

Furthermore, the system is developed using **Python, Streamlit, MongoDB, Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn**, providing an interactive and user-friendly environment for managing student records, analyzing educational data, and generating prediction results. The proposed framework aims to reduce manual effort, improve prediction accuracy, and provide educational institutions with an intelligent solution for monitoring and improving student academic performance.



2. Literature Review

Student performance prediction has become an important research area within **Educational Data Mining (EDM)** and **Machine Learning (ML)** because of its ability to support academic planning and improve educational outcomes. Earlier studies primarily relied on traditional statistical methods to evaluate student performance based on examination scores and attendance records. Although these methods provided useful insights, they were limited in identifying complex relationships among multiple academic attributes and often lacked predictive capabilities. As educational institutions began generating large volumes of digital data, researchers increasingly explored data mining and machine learning techniques to develop more accurate and efficient prediction models.

Several researchers have applied machine learning algorithms such as **Decision Trees, Random Forest, Support Vector Machines (SVM), Naïve Bayes, K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Linear Regression** for predicting student academic performance. These models analyze historical student records, including marks, attendance, assignments, internal assessments, and Semester Grade Point Average (SGPA), to estimate future academic outcomes. Among these techniques, Random Forest has demonstrated high prediction accuracy due to its ensemble learning capability and robustness against overfitting, while Linear Regression has been widely adopted for predicting continuous academic indicators such as SGPA and overall performance scores.

Recent research has also emphasized the importance of **Educational Data Mining** in discovering hidden patterns within academic datasets. Data preprocessing techniques, including missing value handling, normalization, feature selection, and categorical data transformation, have significantly improved the quality of educational datasets before model training. Furthermore, Exploratory Data Analysis (EDA) and visualization methods such as histograms, scatter plots, box plots, correlation matrices, and heatmaps have enabled educators to better understand student learning behavior, identify influential academic factors, and support evidence-based educational decision-making.

Although existing studies have reported encouraging results, many proposed systems concentrate only on performance prediction without providing comprehensive academic analysis or interactive visualization. Some systems utilize a single machine learning algorithm, limiting their flexibility and prediction capability across different educational scenarios. In addition, several existing solutions lack user-friendly interfaces that allow educators to analyze academic data, visualize performance trends, and obtain prediction results through an integrated platform. These limitations reduce the practical applicability of such systems in real educational environments.

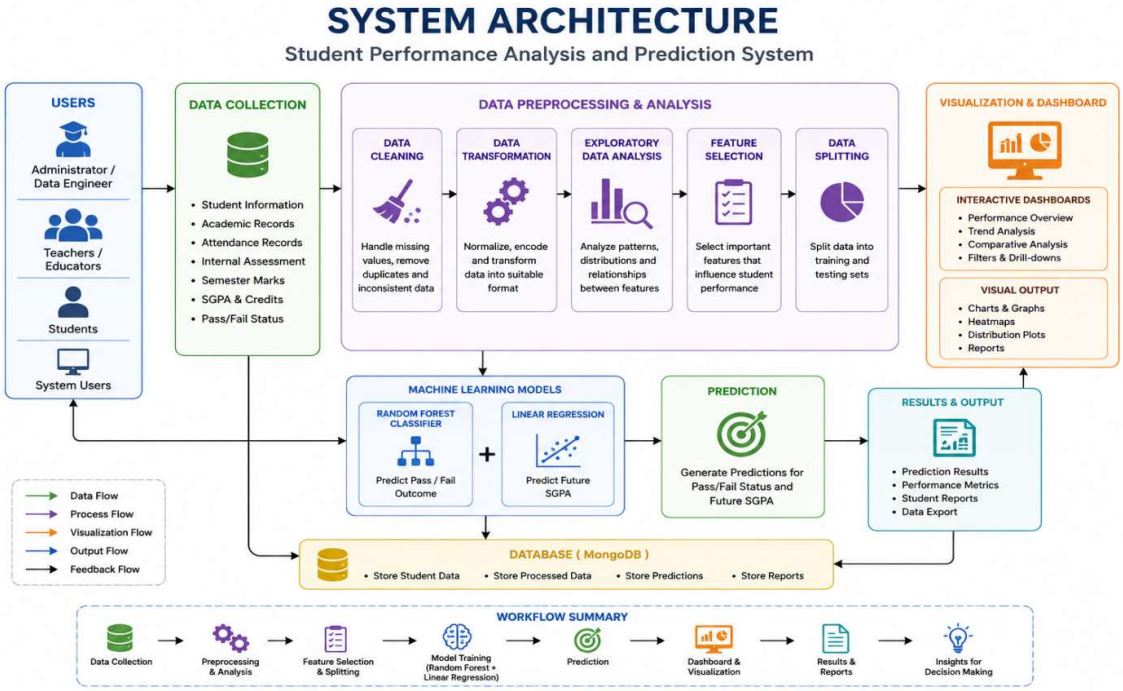
The proposed **Student Performance Analysis and Prediction System** addresses these challenges by integrating data preprocessing, exploratory data analysis, machine learning-based prediction, and interactive visualization within a single web-based application. The system employs a **Random Forest Classifier** to predict student pass/fail outcomes and a **Linear Regression** model to estimate future SGPA values.

Developed using **Python, Streamlit, MongoDB, Pandas, Matplotlib, Seaborn, and Scikit-learn**, the framework provides educators with a comprehensive platform for analyzing student performance, identifying academically at-risk students, and supporting data-driven academic planning. By combining predictive analytics with intuitive visual dashboards, the proposed system enhances both the accuracy and usability of student performance analysis compared to many existing approaches.

3. Research Methodology

The proposed Student Performance Analysis and Prediction System follows a systematic machine learning workflow to analyze student academic records and predict future academic performance. The methodology consists of several stages, including data collection, preprocessing, exploratory data analysis, feature engineering, model training,

prediction, and result visualization. Each stage contributes to improving the accuracy and reliability of the prediction models.



Data Collection

The first stage of the proposed system involves collecting student academic records from educational datasets. The dataset contains important attributes such as student marks, attendance percentage, internal assessment scores, semester examination marks, grades, Semester Grade Point Average (SGPA), credits earned, and pass/fail status. These academic parameters provide valuable information for understanding student learning behavior and form the foundation for performance analysis and prediction. The collected data is stored in a structured format and used as the primary input for the machine learning models.

Data Preprocessing

After data collection, the dataset undergoes preprocessing to improve its quality and prepare it for machine learning analysis. This stage involves handling missing values, removing duplicate records, correcting inconsistent entries, converting categorical data into numerical form, and normalizing numerical attributes. Data preprocessing ensures that the dataset is clean, consistent, and free from errors, thereby improving the efficiency and prediction accuracy of the machine learning models.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis is performed to understand the characteristics of the dataset and identify relationships among different academic attributes. Various statistical methods and visualization techniques, including histograms, scatter plots, box plots, pie charts, count plots, and correlation matrices, are used to analyze student performance trends. This analysis helps educators identify significant factors affecting academic performance and provides meaningful insights before the prediction models are developed.

Feature Selection

Feature selection is carried out to identify the most influential academic attributes that contribute to student performance prediction. Important features such as attendance, internal assessment marks, semester examination scores, grades, SGPA, and credits earned are selected for model training, while irrelevant or redundant attributes are removed. This process reduces computational complexity, improves model efficiency, and enhances the overall prediction accuracy of the proposed system.

Model Training

The processed dataset is divided into training and testing datasets to develop reliable prediction models. The proposed framework utilizes two supervised machine learning algorithms: a Random Forest Classifier for predicting student pass/fail outcomes and a Linear Regression model for estimating future SGPA values. During the training phase, the algorithms learn patterns from historical academic records and establish relationships among different educational attributes, enabling them to generate accurate predictions for unseen student data.

Prediction Process

Once the machine learning models are successfully trained, they are used to predict student academic performance using new academic records. The system analyzes the input data based on the learned patterns and generates predictions such as pass/fail status and expected SGPA. These prediction results help educators identify academically weak students at an early stage, allowing timely intervention and appropriate academic support to improve learning outcomes.

Result Visualization

The final stage of the proposed system presents the analytical and predictive results through an interactive web-based dashboard developed using Streamlit. The dashboard displays various charts, graphs, and performance reports that allow users to analyze student performance, compare academic trends, and interpret prediction results effectively. These visualizations provide a clear understanding of student progress and support educational institutions in making informed, data-driven academic decisions.

4. Results & Analysis

The proposed Student Performance Analysis and Prediction System was successfully implemented and evaluated using student academic records containing attributes such as attendance, internal assessment marks, semester examination scores, grades, SGPA, credits earned, and pass/fail status. The system effectively performed data preprocessing, exploratory data analysis, visualization, and machine learning-based prediction through an interactive web application. Various analytical dashboards and graphical representations, including histograms, pie charts, scatter plots, box plots, and correlation matrices, were generated to examine student performance patterns and identify the factors influencing academic achievement. These visualizations enabled educators to understand the distribution of academic performance and evaluate relationships among different educational attributes.

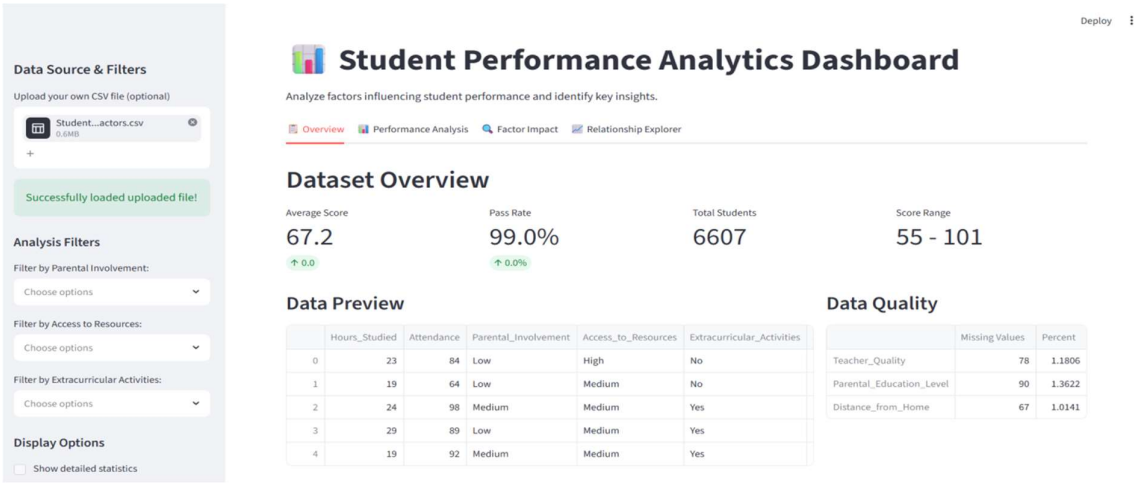


Figure 1. Student Performance Analytics Dashboard

The Student Performance Analytics Dashboard provides an interactive interface for analyzing student academic data. It allows users to upload datasets, apply filters, and view key performance indicators such as average score, pass rate, total students, and score range. The dashboard also displays a preview of the uploaded data and identifies missing values to ensure data quality. These features enable educators to analyze student performance efficiently and support data-driven academic decision-making.

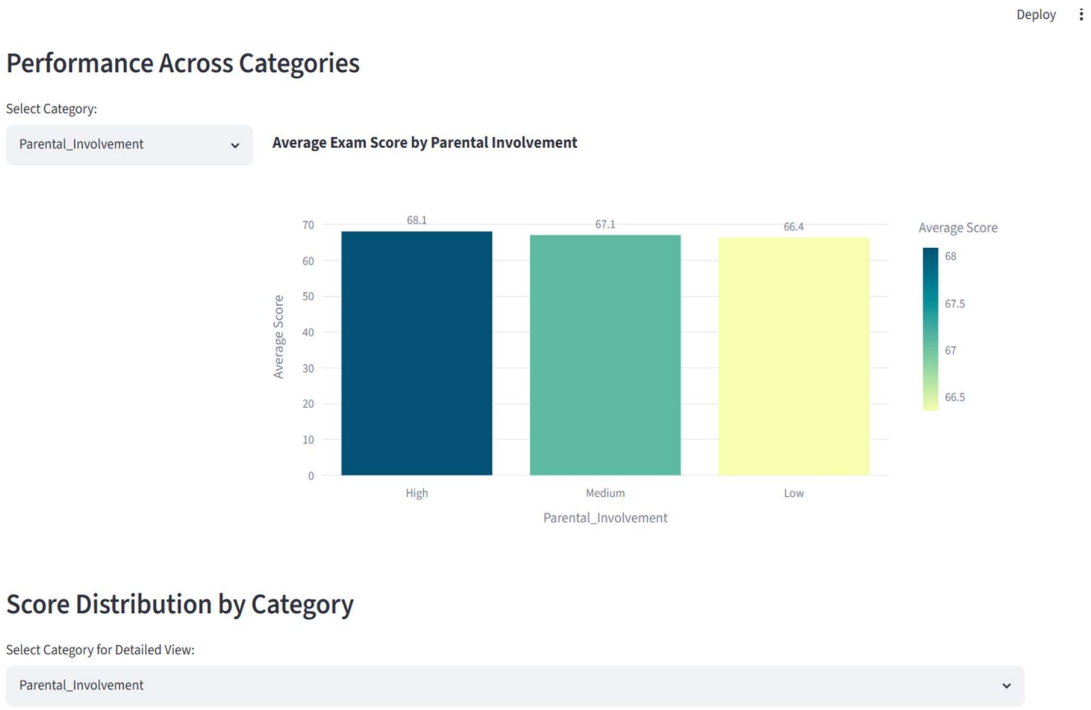


Figure 2. Performance Analysis Dashboard

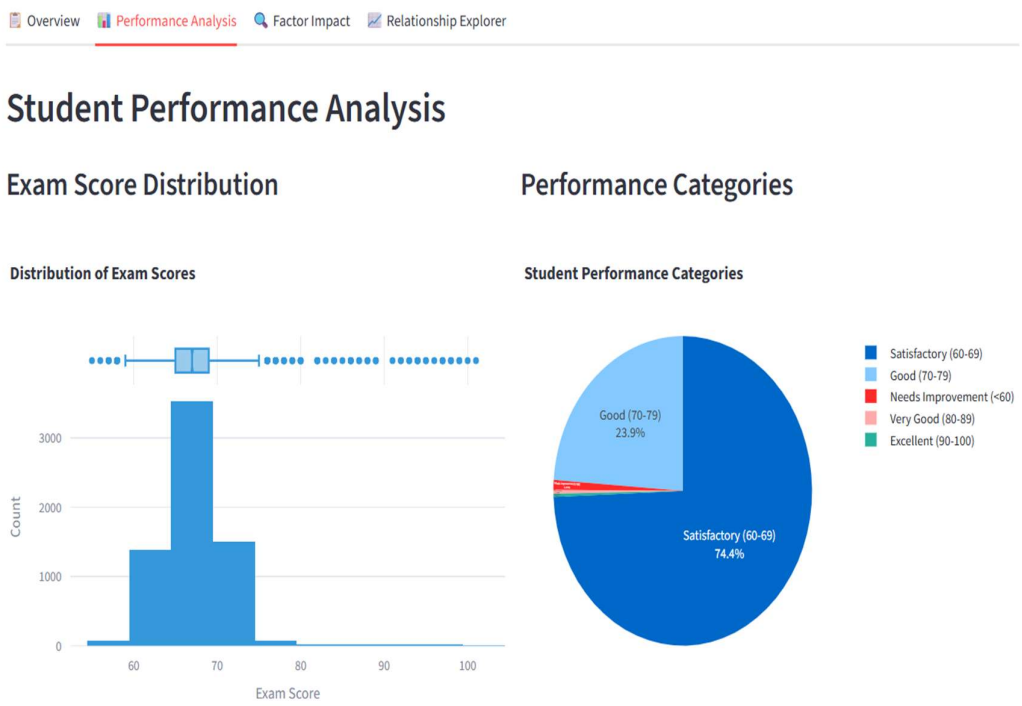


Figure 3. Detailed Factor Analysis

Figure 4. Performance Across Categories

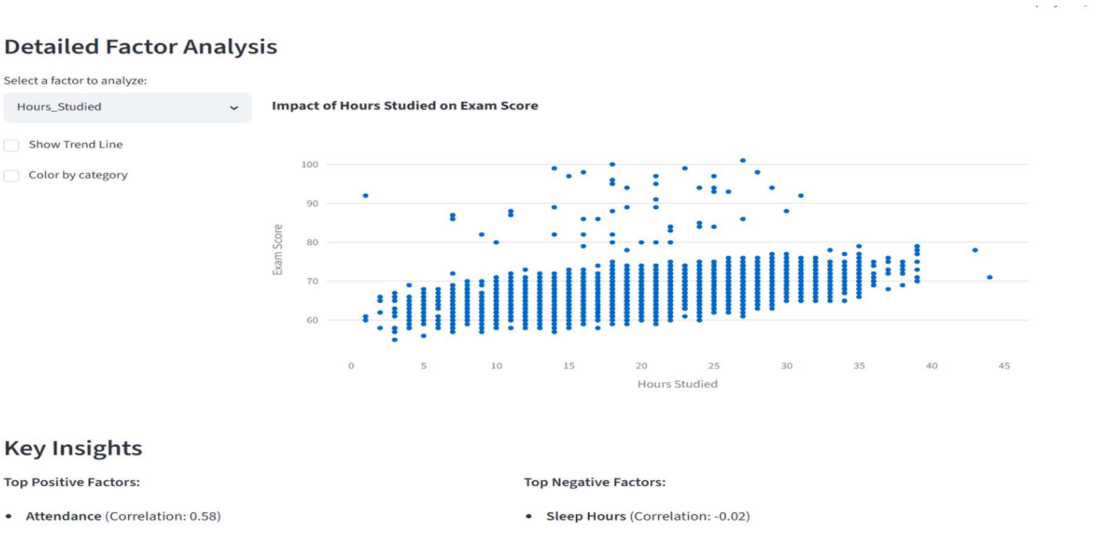


Figure 5. Factor Analysis

The Detailed Factor Analysis module visualizes the relationship between study hours and exam scores using a scatter plot. It helps identify performance trends and highlights the key factors that positively or negatively influence student academic performance.

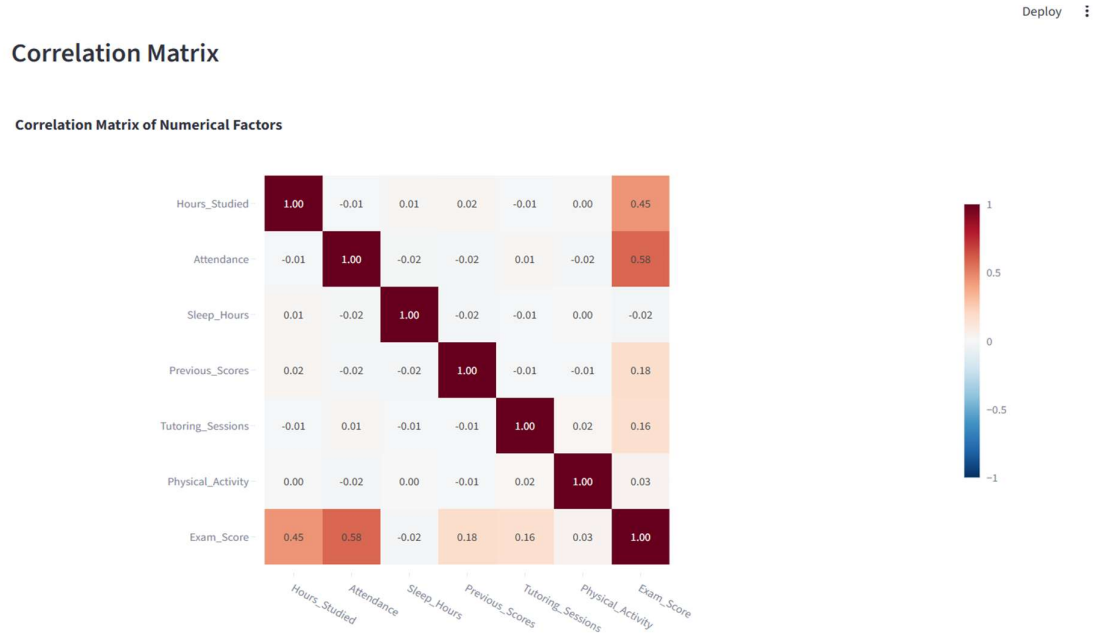


Figure 6. Correlation Matrix

5. Conclusion

This paper presented An Intelligent Framework for Student Performance Analysis and Prediction Using Machine Learning and Educational Data Mining to assist educational institutions in analyzing academic records and predicting student performance. The proposed system integrates data preprocessing, exploratory data analysis, visualization, and supervised machine learning algorithms within a single web-based platform. By utilizing a Random Forest Classifier

for pass/fail prediction and Linear Regression for SGPA estimation, the framework provides reliable predictions and meaningful analytical insights that support educators in identifying academically at-risk students, monitoring student progress, and making data-driven academic decisions. The interactive dashboards and visualization modules further simplify the interpretation of educational data, enabling institutions to improve academic planning and student performance evaluation.

The experimental results demonstrate that the proposed framework effectively enhances the accuracy and efficiency of student performance prediction compared to traditional manual analysis methods. By combining machine learning techniques with educational data analytics, the system reduces manual effort while providing timely and actionable insights for educators. In future work, the proposed framework can be extended by incorporating deep learning models, real-time academic monitoring, personalized learning recommendations, and additional educational attributes to further improve prediction accuracy and support intelligent educational decision-making in modern learning environments.

References

[1] Han, J., Kamber, M., & Pei, J., *Data Mining: Concepts and Techniques*, 4th Edition, Morgan Kaufmann Publishers, 2023.

[2] Géron, A., *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd Edition, O'Reilly Media, 2023.

[3] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J., *Data Mining: Practical Machine Learning Tools and Techniques*, 5th Edition, Morgan Kaufmann, 2023.

[4] Pedregosa, F., Varoquaux, G., Gramfort, A., et al., "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830.

[5] Chollet, F., *Deep Learning with Python*, 2nd Edition, Manning Publications, 2022.

[6] McKinney, W., *Python for Data Analysis*, 3rd Edition, O'Reilly Media, 2022.

[7] Goodfellow, I., Bengio, Y., & Courville, A., *Deep Learning*, MIT Press, Updated Edition, 2023.

[8] Educational Data Mining Society, "Educational Data Mining Research and Applications," Available: https://educationaldatamining.org

[9] Kaggle, "Student Performance Dataset," Available: https://www.kaggle.com/datasets

[10] Streamlit Inc., "Streamlit Documentation," Available: https://docs.streamlit.io

[11] MongoDB Inc., "MongoDB Documentation," Available: https://www.mongodb.com/docs

[12] Python Software Foundation, "Python Documentation," Available: https://docs.python.org/3/

[13] Scikit-learn Developers, "Random Forest Classifier Documentation," Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[14] Scikit-learn Developers, "Linear Regression Documentation," Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

[15] Aggarwal, C. C., *Data Mining: The Textbook*, Springer International Publishing, 2023.