

# EduSmart: AI-Powered Academic Performance Analysis System for Smarter Educational Decisions

Miss. Sanjana Kamble

Department of Computer Science  
ATSS College of Business Studies and Computer  
Application, Pune, India

Miss Swati Chavan

Department of Computer Science  
ATSS College of Business Studies and Computer  
Application, Pune, India

**Abstract:** - Student performance evaluation is a critical component of academic quality assurance and continuous improvement in higher education. With the increasing availability of educational data, analytical techniques can be effectively applied to understand learning outcomes in a structured manner. This research paper presents a systematic analysis of student academic performance using examination marks as the primary data source. The study focuses on identifying performance trends, subject-wise strengths and weaknesses, and overall academic consistency among undergraduate Computer Science students.

The dataset used in this research consists of semester-wise examination results collected from institutional academic records. Python-based data analysis techniques are employed, utilizing libraries such as Pandas and NumPy for data preprocessing, computation, and statistical evaluation. Visualization techniques, including graphs and charts, are applied to enhance interpretability and facilitate comparative analysis across subjects and semesters.

The findings of the study provide meaningful insights into student learning patterns, enabling early identification of weak academic areas and recognition of high-performing subjects. Such insights can support students in improving their academic strategies and assist educators in designing targeted interventions. The research demonstrates how data-driven approaches can contribute to effective academic monitoring, informed decision-making, and improved educational outcomes

**Keywords:** *Student Performance Analysis, Data Mining, Python-Based Analysis, Subject-Wise Performance, Learning Outcome Assessment, Data Visualization in Education*

## I. INTRODUCTION:

In the contemporary educational environment, academic performance assessment has evolved beyond traditional result reporting to include systematic analysis and interpretation of student data. Evaluating student marks is not only essential for measuring academic achievement but also plays a significant role in understanding learning behaviors, subject comprehension, and instructional effectiveness. Analyzing

performance data enables educational institutions to enhance teaching methodologies and support student development more effectively (Al-Din & Al Abdulqader, 2024).

With the rapid advancement of data analytics and programming tools, educational datasets can now be processed efficiently to uncover meaningful patterns. Student examination results, when analyzed systematically, provide valuable insights into subject-wise performance variations, consistency across semesters, and overall academic progress. Such analysis helps identify students who require additional academic support and subjects that may need curriculum or instructional improvements (Al-Din & Al Abdulqader, 2024).

Educational Data Mining (EDM) and Machine Learning (ML) techniques have increasingly been applied to academic datasets to extract hidden patterns and support data-driven decision making in education. The EDM cycle emphasizes stages such as problem definition, data collection, feature selection, model training, evaluation, and deployment, which collectively transform raw educational data into actionable insights (Al-Din & Al Abdulqader, 2024).

This research focuses on the analysis of examination marks of undergraduate Computer Science students using Python-based data analysis techniques. By applying computational methods to real academic records, the study aims to bridge the gap between raw academic data and actionable insights. The use of data visualization further enhances understanding by presenting complex results in a clear and interpretable format, a practice strongly supported in EDM research for improving interpretability of academic analysis (Yağcı, 2022). The significance of this study lies in its practical application of data analytics in the education domain. It highlights how academic performance data can be transformed into useful knowledge that benefits both students and educators. Through systematic analysis, the research contributes to the development of data-driven academic

monitoring systems that promote continuous improvement in teaching and learning processes (Yağcı, 2022; Al-Din & Al Abdulqader, 2024)

## II. LITERATURE REVIEW

Educational Data Mining (EDM) and Machine Learning (ML) have emerged as powerful approaches for analyzing student data and predicting academic performance. Researchers have increasingly focused on transforming raw educational records into meaningful insights that can support instructors, institutions, and students in improving learning outcomes.

Yağcı (2022) conducted an experimental study to predict students' final exam performance using machine learning algorithms such as Random Forest, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes. The study used a dataset of 1854 undergraduate students and demonstrated that even limited academic features such as midterm marks, department, and faculty information could achieve an accuracy between 70% and 75%. The research highlighted the importance of early prediction systems to identify students at risk before final examinations and emphasized the practical applicability of ML models in academic monitoring systems.

In contrast to single experimental studies, Al-Din and Al Abdulqader (2024) presented a comprehensive systematic review of 91 research papers published between 2015 and 2023 related to student performance prediction using EDM and ML techniques. Their review identified the most commonly used data sources such as Learning Management Systems (LMS), Student Information Systems (SIS), attendance records, and academic results. The study also explained the complete EDM cycle, which includes problem definition, data collection, data preprocessing, feature selection, model building, evaluation, and deployment. This structured approach provides a methodological foundation for developing student performance prediction systems.

The review by Al-Din and Al Abdulqader (2024) further revealed that Decision Tree, Random Forest, Logistic Regression, and SVM are the most frequently used algorithms in educational prediction tasks due to their interpretability and accuracy. The authors emphasized that nearly 62% of researchers preferred Decision Tree models because they are easy for educators to understand and interpret. Additionally, the study highlighted the importance of feature selection techniques such as Chi-square, Information Gain, and Gain Ratio in improving model performance.

Both studies emphasize the growing need for data-driven academic monitoring systems. While Yağcı (2022) demonstrated practical implementation and effectiveness of ML models on real student data, Al-Din and Al Abdulqader (2024) provided a broad overview of existing methodologies, tools, challenges, and research gaps in the field. One key issue identified is that many existing systems lack interpretability for teachers and fail to connect predictions with actionable educational strategies.

These research contributions form a strong foundation for the development of the Edusmart system. They justify the use of EDM and ML techniques, guide the selection of appropriate algorithms and features, and highlight the importance of creating interpretable and practical student performance analysis systems.

### A. Research Motivation

The motivation behind Edusmart stems from the growing recognition that educational institutions require advanced analytics to address student dropout, performance decline, and personalized learning needs. Traditional evaluation methods focusing on final exam results or end-of-term assessments often fail to provide proactive insights. However, predictive models that leverage historical and real-time student data enable early detection of trends that may lead to poor performance. Machine learning and EDM techniques have demonstrated strong potential in extracting meaningful patterns from large academic datasets, motivating the development of frameworks like Edusmart (Al-Din & Al Abdulqader, 2024; Junejo et al., 2025).

### B. Research Gap

Although substantial progress has been made in predictive analytics for education, several gaps remain:

Lack of interpretability in many high-accuracy models.

Limited integration of diverse educational data sources.

Insufficient focus on early warning capabilities that support real-time decision-making.

Challenges related to model fairness and ethical data use.

These gaps indicate the need for a comprehensive system that combines performance prediction accuracy with transparency and actionable insights for educators.

### C. Problem Statement

Existing academic performance prediction systems often focus on either accuracy or model complexity, but do not fully address the need for explainable predictions that can be effectively used by educators to improve learning. There is a need for a predictive framework that not only forecasts student performance reliably but also provides clear insights into the contributing factors, enabling personalized academic support.

#### D. Objectives of the Study

The objectives of this research are:

To analyze recent developments in student performance prediction using machine learning.

To propose an architecture for Edusmart that integrates explainable predictive models.

To identify the key data sources and preprocessing steps required for accurate forecasting.

To discuss the role of transparent analytics in supporting educator interventions

### III. RESEARCH METHODOLOGY:

This study adopts a quantitative and descriptive research design to analyze student academic performance using examination marks. The methodology is structured to systematically collect, process, analyze, and interpret academic data in order to derive meaningful insights related to student performance patterns, subject-wise strengths, and areas requiring improvement

#### A. Introduction to the Proposed System :

The proposed system is designed to analyze student academic performance using a structured, data-driven approach. Its primary objective is to convert raw examination marks into meaningful insights that reflect subject-wise strengths, weaknesses, and overall academic trends. Unlike traditional manual result analysis, the proposed system automates data handling and analysis, thereby improving accuracy, efficiency, and interpretability. The system supports both students and educators by providing clear analytical outcomes that can assist in academic planning, performance monitoring, and decision-making.

#### B. System Architecture :

The given figure 1 indicates architecture of edusmart, the system architecture follows a modular and layered design to ensure systematic processing of academic data. The architecture consists of the following interconnected components:

#### 1. Data Input Module

This module is responsible for acquiring student examination data from institutional academic records. The data includes semester-wise and subject-wise marks.

#### 2. Data Preprocessing Module

In this stage, the collected data is cleaned, validated, and formatted to ensure consistency and accuracy. Errors, missing values, and inconsistencies are addressed before analysis.

#### 3. Analysis Module

The analysis module applies computational techniques to evaluate student performance. It calculates performance indicators such as averages, subject-wise scores, highest and lowest marks, and semester-wise trends.

#### 4. Visualization Module

This module generates graphical representations such as bar charts and comparative graphs to visually present performance patterns. Visualization improves clarity and enhances understanding for non-technical users.

#### 5. Output Module

The final module presents analyzed results in an interpretable format, highlighting strong and weak subjects and overall academic performance for students and educators

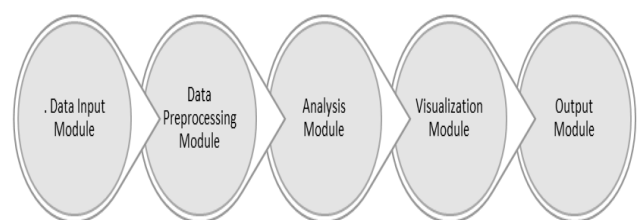


Figure 1 : Architecture Diagram of Edusmart

#### C. Data sources

- The data used in this research is obtained from atss college of bussiness studies & computer application , academic records of undergraduate computer science students. . Data includes 78 records of students (results of sem iii & sem iv). Includes internal marks and exam data.
- The dataset includes semester-wise examination marks across multiple subjects. Since the data originates from

real institutional sources, it reflects actual academic conditions and enhances the practical relevance of the study. Using authentic student data ensures that the analytical outcomes are realistic and applicable in real educational environments

#### IV. EXPERIMENTAL SETUP :

The research employs modern computational tools to ensure effective data analysis and visualization:

Python Programming Language serves as the primary platform for implementing the analysis due to its simplicity and analytical capabilities. Pandas Library used for data manipulation, filtering, aggregation, and tabular analysis. NumPy Library applied for numerical computations and statistical calculations. Data Visualization Tools utilized to generate charts and graphs that represent academic performance trends. Spreadsheet Software used for storing and organizing the raw dataset before analysis.

##### Data Preprocessing Steps

Data preprocessing is a critical phase in the research methodology, as the accuracy of results depends on data quality. The following steps are performed:

##### 1. Data Cleaning

Removal of duplicate records and correction of inconsistent or incorrect entries.

##### 2. Handling Missing Values

Identification of missing or null values and appropriate handling to avoid analytical bias.

##### 3. Data Formatting

Conversion of raw data into a structured and standardized tabular format suitable for computational analysis.

##### 4. Data Validation

Verification of data correctness to ensure marks are accurately recorded and within valid ranges.

##### 5. Data Organization

Arrangement of data subject-wise and semester-wise to enable efficient performance analysis.

#### V. DATA ANALYSIS & VISUALIZATION:

##### (a) Bar Graph — Semester III vs Semester IV Comparison

The subject-wise marks distribution comparison is shown in Figure 2. The first bar chart compares average marks of students in Semester III and Semester IV for each subject. Each subject (like EVS, English, C1, C2, Practical's, etc.) has two bars — one blue (Sem III) and one orange (Sem IV). The chart's title: "Subject-wise Marks Distribution Comparison (Sem III vs Sem IV)". This helps visualize improvement or decline in marks from one semester to another

Observations:

- Most subjects show a slight increase in Semester IV.
- Some subjects like "Practical's" and "M1" show notable improvement.

- The difference between bars shows the performance trend over time.

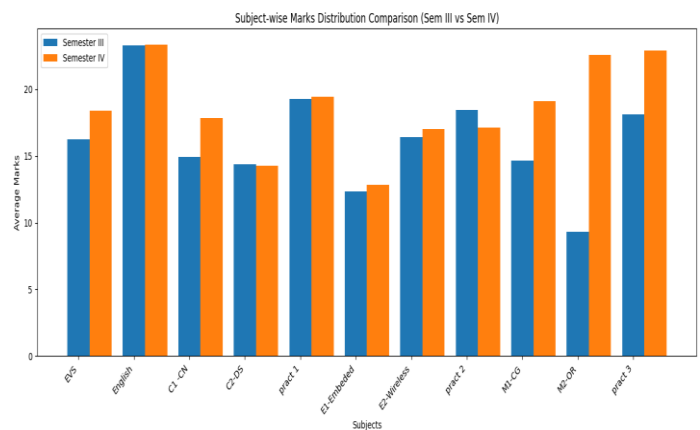


Figure 2 : subject-wise marks distribution comparison

##### b) Line Graph — Semester III Subject-wise Marks

The figure 3 shows second graph which is a line plot showing average marks per subject in Semester III only.

X-axis: Subjects (EVS, English, DS, SE, etc.)

Y-axis: Average marks

The line connects points showing how marks fluctuate across subjects.

Observations:

English and SE subjects have higher averages (~22–23 marks).

M3 (Graph Theory) has the lowest average, indicating a weak area for students.

The zigzag line pattern reflects uneven performance across subjects.

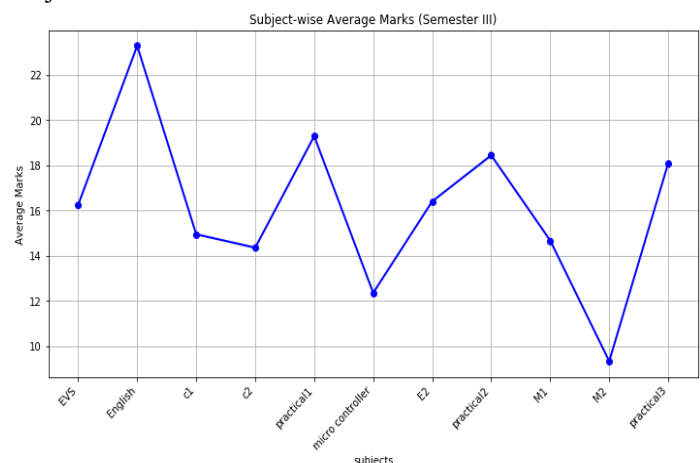


Figure 3 : Subject-wise Average Marks (semester III)

##### (c) Bar + Histogram — Overall Performance

Figure 4 shows one plots: Bar chart: Average marks per subject.

Histogram: Distribution of total marks across all students.

Shows that average marks per subject range roughly from 10 to 14.

EVS, DS, and Practical's have better averages compared to other subjects.

Shows how total marks are distributed.

Most students fall in the 120–150 marks range, meaning performance is clustered around average.

Few students are below 100 or above 150.

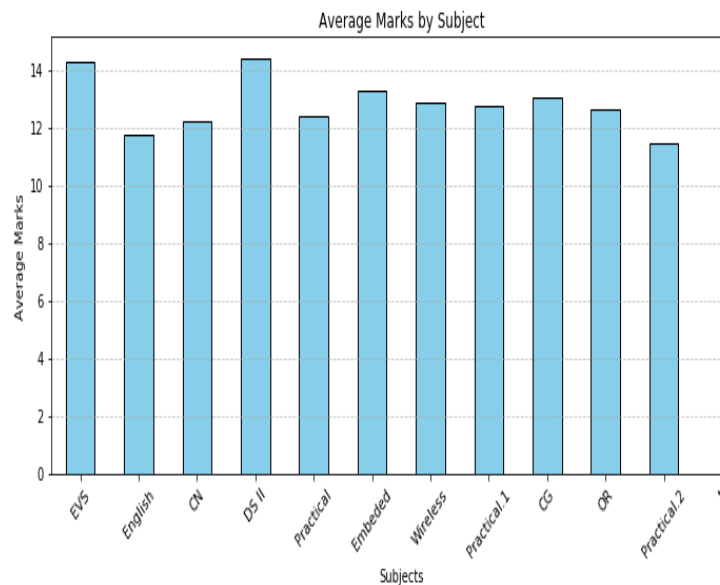


Figure 4 : Average marks by subjects

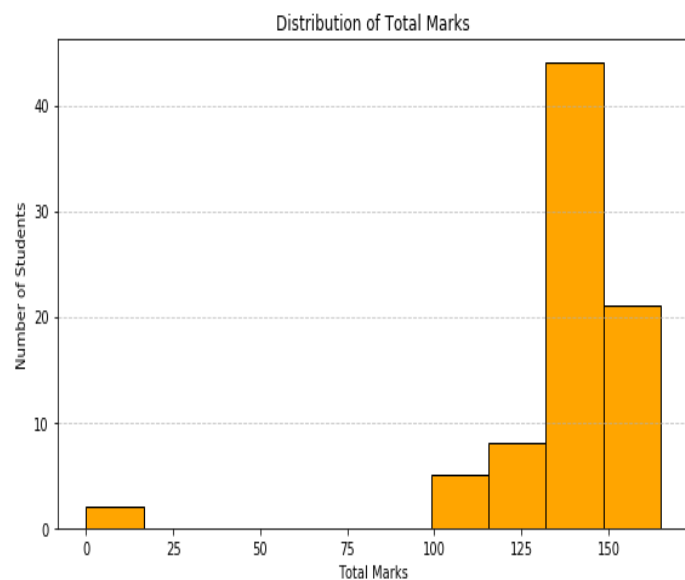


Figure 5 : Distribution of total marks

## VI. INTERPRETATION OF RESULTS:

### 1. Semester Comparison:

There's an overall improvement from Semester III to Semester IV.

Indicates that students are performing better with time, possibly due to familiarity with exam patterns or effective teaching strategies.

### 2. Subject Performance:

Subjects like English, SE, and Practical's have higher average marks.

Mathematical and theoretical subjects (e.g., M1, CO) have lower averages, showing they are challenging areas for students.

3. Performance Distribution: The histogram shows a normal-like distribution, with most students performing around the mean range.

Only a few students perform exceptionally well or poorly — this indicates consistent average performance across the class.

### 4. Overall, Class Standing:

As per. Analysis the mean performance does not reach high distinction levels.

Hence, class performance is stable but needs improvement.

### Insights and Findings:

#### 1. Academic Strengths:

Subjects like English, EVS, and Software Engineering have higher averages — indicating students understand conceptual and descriptive subjects well.

Practical subjects show better performance, meaning students are good at applied learning.

#### 2. Areas of Improvement:

Subjects involving Mathematical reasoning (e.g., M1, CO, Numerical methods) need more focus.

Additional tutorials, remedial classes, or lab sessions could improve marks here.

#### 3. Performance Trend Over Time:

From Semester III to Semester IV, most subjects show improvement, which suggests:

Better preparation and study patterns.

Curriculum familiarity.

Possibly improved teaching or assessment methods.

#### 4. Overall, Class Performance:

The class's total marks distribution shows a majority near the average mark range — not too high or low

Indicates a moderate academic standing

Need to raise overall performance for better results next term.

## VII. CONCLUSION FROM THE GRAPHS

The analysis shows that students have improved their marks from Semester III to Semester IV, meaning their performance is getting better over time. Subjects like English, EVS, and Practical's have higher scores, showing students understand these topics well. However, subjects like M1 and CO are weaker and need more attention. Most students scored around the average range, with only a few performing very high or low. This means the class is doing okay but can do better. With extra help in difficult subjects and regular performance checks, overall results can improve further in the next semester.

## VIII. DISCUSSION

Edusmart underscores the transformative potential of predictive analytics in education. By integrating accuracy with interpretability, predictive systems can offer educators actionable insights. Explainable models help identify which factors, such as attendance or prior scores, most influence student outcomes, allowing for targeted learning interventions. Ethical considerations, such as data privacy and fairness, are important when deploying such systems in real educational settings.

## IX. CONCLUSION

This paper presented a comprehensive framework for Edusmart, an educational performance prediction system that leverages machine learning and explainable analytics. By reviewing recent research from 2023 to 2025, the study highlighted key methodologies, data sources, and the importance of explainability in educational predictive systems. The proposed framework aims to support educators and institutions in the early identification of at-risk students and the implementation of data-driven interventions.

This study can be extended to automated academic alerts along with the predictive analysis techniques to forecast student academic performance in future semesters. Recommendations can be implemented to notify students about performance improvement areas. Also, the automated system can be integrated with institutional learning management systems for seamless data flow.

## REFERENCES

- [1] Abukader, M., Alzubi, J., & Adegboye, O. (2025). An intelligent student performance prediction system using metaheuristic-optimized LightGBM with SHAP explainability. *IEEE Access*.
- [2] Al-Din, M. S. N., & Al Abdulqader, H. A. (2024). Students' academic performance prediction using educational data mining and machine learning: A systematic review. *International Journal of Research and Innovation in Social Science*, 8(8), 1264–1291. <https://doi.org/10.47772/IJRIS.2024.808095>
- [3] Alhassan, M., & Lawal, A. (2024). Students' academic performance prediction using educational data mining and machine learning. *International Journal of Research and Innovation in Social Science*.
- [4] Alshabandar, R., Hussain, A. J., Keight, R., Khan, W., & Al-Jumeily, D. (2023). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Electronics*.
- [5] Chavan, S. (2025). *EduSmart: Academic performance analysis using Python data analytics* (Unpublished project report).
- [6] Junejo, A., et al. (2025). Neural network-based multi-category student performance forecasting for early online education intervention. *Scientific Reports*.
- [7] Khan, M., et al. (2024). Student performance prediction with regression approach and data generation. *Applied Sciences*, 14.
- [8] Kumar, A., & Singh, P. (2023). Student performance prediction approach based on educational data mining. *IEEE Journals & Magazines*.
- [9] Parkavi, R., & Karthikeyan, T. (2025). Enhancing student performance prediction in higher education using a data-driven ensemble approach. *International Journal of Emerging Technologies in Learning (IJET)*.
- [10] Ramesh, V., Parkavi, R., & Karthikeyan, T. (2023). Student academic performance prediction using supervised learning techniques. *International Journal of Emerging Technologies in Learning (IJET)*.
- [11] Sharma, S., & Gupta, P. (2023). Student academic performance prediction using educational data mining. *IEEE Conference Publication*.
- [12] Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1). <https://doi.org/10.1186/s40561-022-00199-4>