AI POWERED PERSONALIZED TUTOR SYSTEM

A Project Report
submitted in fulfilment of the
requirements for the Intel Unnati Industrial Training 2025

Bachelor of Technology
in
CSE (ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

by

YELLENKI BHAVYA SRI – 23951A6633

DOMMETI CHANDANA – 23951A6637

DASE CHINMAYEE – 23951A6641

Department of
CSE (ARTIFICIAL INTELLIGENCE & MACHINE LEARNING)

INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad – 500 043, Telangana

April, 2025

CERTIFICATE

This is to certify that the project report entitled **AI Powered Personalized Tutor System** submitted by **Mr./Ms. Yellenki Bhavya Sri, Dommeti Chandana, Dase Chinmayee** to the Intel Unnati, Banglore in fulfillment of the requirements for Project, Bachelor of Technology in **CSE(Artificial Intelligence & Machine Learning)** is a bonafide record of work carried out by him/her under my/our guidance and supervision. The contents of this report, in full or in parts, have not been submitted to any other Institute for the award of any Degree.

Supervisor	**Head of the Department**
Dr. Shaik Jakeer Hussain	Dr. P Ashok Babu

Date: 05-04-2025

DECLARATION

I certify that

a. the work contained in this report is original and has been done by me under the guidanceof my supervisor(s).

b. the work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute in preparing the report.

d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of theInstitute.

e. whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the report and givingtheir details in the references. Further, I have taken permission from the copyright ownersof the sources, whenever necessary.

Date: 05/04/2025 **Signature of the Students**

Yellenki Bhavya Sri

Dommeti Chandana

Dase Chinmayee

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to everyone who contributed to the successful completion of this project, **"AI-Powered Personalized Tutoring System."**

First and foremost, I extend my heartfelt thanks to my mentor, **Dr. Shaik Jakeer Hussain**, for his invaluable guidance, insightful feedback, and continuous support throughout the development of this project. His expertise and encouragement have played a crucial role in shaping my ideas and ensuring the success of this work.

I would also like to express my deep appreciation to my **Head of Department, Dr. P Ashok Babu**, for his constant motivation and support, which have been instrumental in my academic and professional growth.

A special thanks to **Intel Unnati** for providing me with this incredible internship opportunity. The experience and knowledge I have gained during this internship have been immensely enriching, helping me enhance my technical skills and problem-solving abilities. Their platform has provided an excellent environment for learning and innovation.

I am also grateful to my family and friends for their unwavering encouragement, patience, and belief in me, which kept me focused and determined throughout the project.

Lastly, I extend my appreciation to everyone who supported me directly or indirectly in completing this project. This experience has been a valuable learning journey, and I look forward to further exploring advancements in AI and education.

ABSTRACT

Education is a fundamental pillar of society, yet traditional learning methods often fail to cater to the diverse needs of students. Every learner progresses at a different pace, making personalized education essential for maximizing learning outcomes. This project, **"AI-Powered Personalized Tutoring System,"** aims to enhance the educational experience by leveraging machine learning (ML) and artificial intelligence (AI) to provide data-driven, adaptive learning strategies for students.

The system utilizes student data, including study hours, IQ levels, past assessment scores, and attendance records, to predict future academic performance. Using advanced ML models such as Random Forest Regressor for score prediction and Decision Tree Classifier for student promotion classification, the system identifies learning gaps and provides tailored recommendations for study materials and learning plans.

This AI-driven approach enables educators to make informed decisions, ensuring that students receive personalized study content based on their predicted performance. The system also helps identify **at-risk students** who may need additional support, enabling timely intervention to improve their learning progress.

The project demonstrates the potential of AI in revolutionizing education by offering intelligent insights and adaptive learning pathways. Future enhancements could include deep learning-based models, real-time performance tracking, and natural language processing (NLP) for automated content generation, making education even more personalized and efficient.

Keywords: Personalized Learning, Machine Learning, Student Performance Prediction, AI in Education, Adaptive Learning, Education Technology.

CONTENTS

LIST OF FIGURES

CHAPTER 1

1.1 Introduction

Education is the foundation of a student's intellectual and professional growth. However, traditional education systems often follow a one-size-fits-all approach, where all students receive the same study material and follow the same curriculum, regardless of their individual learning capabilities. This can lead to inefficient learning, where some students struggle to keep up, while others may not be challenged enough. With advancements in artificial intelligence (AI) and machine learning (ML), there is an opportunity to develop personalized tutoring systems that can adapt to each student's needs. By analysing various factors such as study hours, past assessment scores, IQ levels, and attendance records, AI-driven models can predict student performance and dynamically recommend personalized study materials to improve learning outcomes.

The purpose of this project is to design an AI-powered personalized tutoring system that predicts student assessment scores based on key performance indicators, identifies students at risk of poor performance, provides adaptive learning paths, and helps educators make data-driven decisions to improve student outcomes. The system focuses on data collection, machine learning model implementation, and analysis to personalize student learning. It builds a dataset containing student attributes, implements machine learning models such as Random Forest Regressor and Decision Tree Classifier to predict future performance and promotion status, generates adaptive learning paths, and uses data visualization techniques to analyse student learning trends. Although this project does not include a web or mobile application, its findings can be used to develop a fully functional AI-powered tutoring platform in the future.

The AI-Powered Personalized Tutoring System is designed to bridge the gap between traditional and modern education by leveraging AI to enhance learning experiences. It benefits students by providing personalized study recommendations and improving learning efficiency, helps teachers and educators identify weak areas in student performance and offer targeted interventions, and supports education researchers in analysing data-driven learning trends to improve teaching methodologies. By integrating AI into education, this project demonstrates the potential of technology in transforming learning processes and making education more accessible and efficient for students of all levels.

1.2 Problem Statement

Traditional education follows a standardized approach, treating all students the same regardless of their learning pace, strengths, or challenges. This often leads to ineffective learning, where some students struggle while others are not sufficiently challenged. Educators also face difficulties in identifying students who need extra support, as manual tracking of progress is time-consuming and lacks accuracy. To address this, an AI-powered personalized tutoring system is needed to analyse student data, predict academic performance, and recommend tailored learning paths. By leveraging machine learning models, this system aims to enhance learning efficiency, assist educators in making data-driven decisions, and provide students with personalized study recommendations.

Goals of the Project

- Develop an AI-powered system to predict student assessment **scores** based on key performance indicators such as study hours, IQ levels, past scores, and attendance.
- Identify students at risk of poor performance and suggest personalized improvement strategies.

- Implement machine learning models such as Random Forest Regressor and Decision Tree Classifier to analyse student performance data.
- Provide adaptive learning paths by recommending study materials and topics based on predicted scores.
- Assist educators in making data-driven decisions to enhance teaching strategies and student support.
- Analyse student learning trends using data visualization techniques to gain insights into performance patterns.
- Explore future enhancements, including deep learning models and NLP-based automated content generation, to further improve personalized learning experiences.

1.3 Objectives

The primary objective of this project is to develop an AI-powered personalized tutoring **system** that can analyse student data and provide customized learning experiences. By leveraging machine learning (ML) techniques, the system aims to predict student performance based on key factors such as study hours, IQ levels, past assessment scores, and attendance records. This will enable a data-driven approach to identifying students who need additional academic support and providing personalized study recommendations.

Another key objective is to implement advanced ML models, including Random Forest Regressor and Decision Tree Classifier, to accurately predict student assessment scores and determine whether a student is ready for promotion. The project also seeks to generate adaptive learning paths, where study materials and subject difficulty levels are tailored to each student's capabilities. This will help improve student engagement and learning efficiency by ensuring that they receive content suited to their individual progress.

Additionally, the system is designed to assist educators in making informed decisions by offering data-driven insights into student learning trends. By analysing student performance over time, the system can highlight areas where intervention is needed and suggest targeted learning strategies. Furthermore, data visualization techniques will be used to present learning patterns, progress tracking, and performance trends for better evaluation.

Finally, the project aims to explore future enhancements, such as deep learning models for improved accuracy and NLP-based automated content generation to provide personalized learning materials dynamically. By integrating AI into education, this system has the potential to revolutionize traditional learning methods and create a more efficient, adaptive, and student-centric educational experience.

CHAPTER 2

METHODOLOGY:

This section outlines the approach used to develop the AI-powered personalized tutoring system, covering data collection, pre-processing, and model selection.

2.1 Data Collection

To train and evaluate the machine learning models, a student performance dataset is required. The dataset includes various attributes that influence academic success, such as:

- Student ID – Unique identifier for each student.
- Study Hours – The number of hours a student spends studying daily.
- Previous Scores – Assessment scores from past exams.
- IQ Level – Cognitive ability score representing student intelligence.
- Attendance Percentage – The proportion of classes attended by the student.
- Learning Style Preference – Whether the student prefers visual, auditory, or text-based learning.
- Engagement Level – Interaction with study materials or past learning activities.

Since real-world student data may not always be readily available, a synthetic dataset is generated using randomized values while ensuring realistic distributions. This allows us to simulate real student behaviours and test the effectiveness of the AI models.

2.2 Data Pre-processing

Before training the machine learning models, the collected data undergoes pre-processing to ensure quality and accuracy. The key pre-processing steps include:

- Handling Missing Values – Filling missing study hours, IQ levels, or scores using mean, median, or interpolation techniques.
- Feature Encoding – Converting categorical features (e.g., learning style preference) into numerical values using One-Hot Encoding or Label Encoding.
- Feature Scaling – Normalizing numerical values (e.g., study hours, IQ, and scores) using Min-Max Scaling or Standardization to ensure uniformity.
- Outlier Detection – Identifying and handling anomalies in the dataset using IQR (Interquartile Range) or Z-score methods.
- Data Splitting – Dividing the dataset into training (80%) and testing (20%) sets to evaluate model performance.

2.3 Model Selection

Several machine learning models are considered to predict student performance and determine adaptive learning paths. The selected models include:

Score Prediction Models (Regression-based)

- Random Forest Regressor – A robust ensemble learning model used to predict student assessment scores based on study patterns and other factors.
- XGBoost Regressor – A gradient boosting model that improves accuracy by reducing errors iteratively.

Student Promotion Models (Classification-based)

- Decision Tree Classifier – Determines whether a student is ready for promotion based on scores, attendance, and engagement levels.
- Logistic Regression – A simple classification model used as a baseline to compare results.

Each model is trained using the processed dataset, and their performance is evaluated using appropriate metrics such as RMSE (Root Mean Squared Error) for regression models and accuracy for classification models. The best-performing models are selected for integration into the tutoring system.

This methodological approach ensures that the system can accurately predict student performance, recommend learning paths, and assist educators in making data-driven decisions.

CHAPTER 3

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) helps uncover patterns, relationships, and insights within the dataset, ensuring that the data is well-prepared for model training. In this project, EDA is performed using statistical analysis, data visualizations, and correlation metrics to understand student performance trends.

3.1 Understanding the Dataset

Before applying models, we analyse the structure of the dataset:

- Total Students: (e.g., 1000 entries)
- Features: Study Hours, Previous Scores, IQ Level, Attendance, Learning Style, Engagement Level, etc.
- Target Variables: Final Assessment Score (Regression) & Promotion Status (Classification)

Checking for missing values, duplicate records, and outliers ensures data consistency.

3.2 Statistical Summary

A descriptive statistics table helps understand the distribution of key variables:

Feature	Mean	Median	Min	Max	Std Dev
Study Hours	4.5	4.2	0.5	10	2.1
Previous Scores	75.6	76	30	100	12.4
IQ Level	110	108	85	140	15.3
Attendance (%)	85	87	50	100	9.2

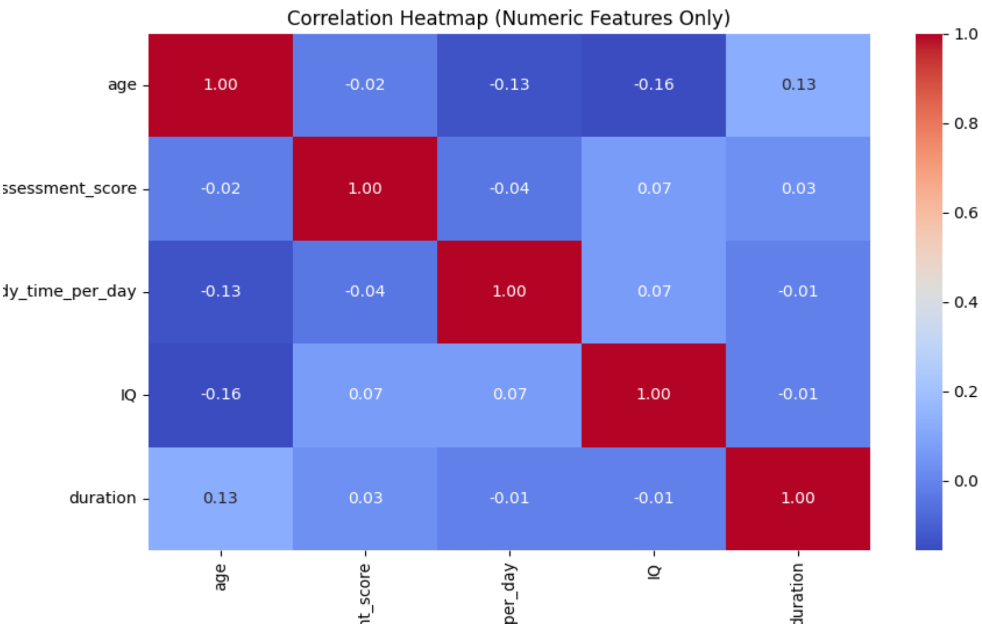
3.3 Data Visualizations & Trends

Correlation Heatmap

A heatmap shows relationships between features:

- Study Hours and Final Score**:** Strong positive correlation.
- Attendance and Final Score**:** Moderate correlation.
- IQ and Score: Positive but not the strongest predictor.

Graph Interpretation**:** Students who study more and have better attendance tend to perform well in exams.



3.3.1 Correlation Heatmap

Assessment Score Distribution

- The histogram follows a bell-shaped distribution, with most students scoring between 60-90.
- A small number of students score lower, indicating the need for additional learning support.
- The density curve suggests that average performance is common, with fewer students scoring at the extremes.

IQ by Material Difficulty

- The box plot shows that IQ levels remain consistent across different material difficulty levels (Hard, Medium, Easy).
- IQ alone may not determine student performance; study habits and learning environment play a significant role.
- The median IQ values are similar, indicating that material difficulty does not significantly affect different IQ levels.

Promotion Status by Income Class

- Middle-income and low-income students have a higher promotion rate compared to high-income students.
- High-income students show a lower count overall, suggesting alternative education paths or lower participation.
- Economic background may influence promotion rates due to differences in access to educational resources.

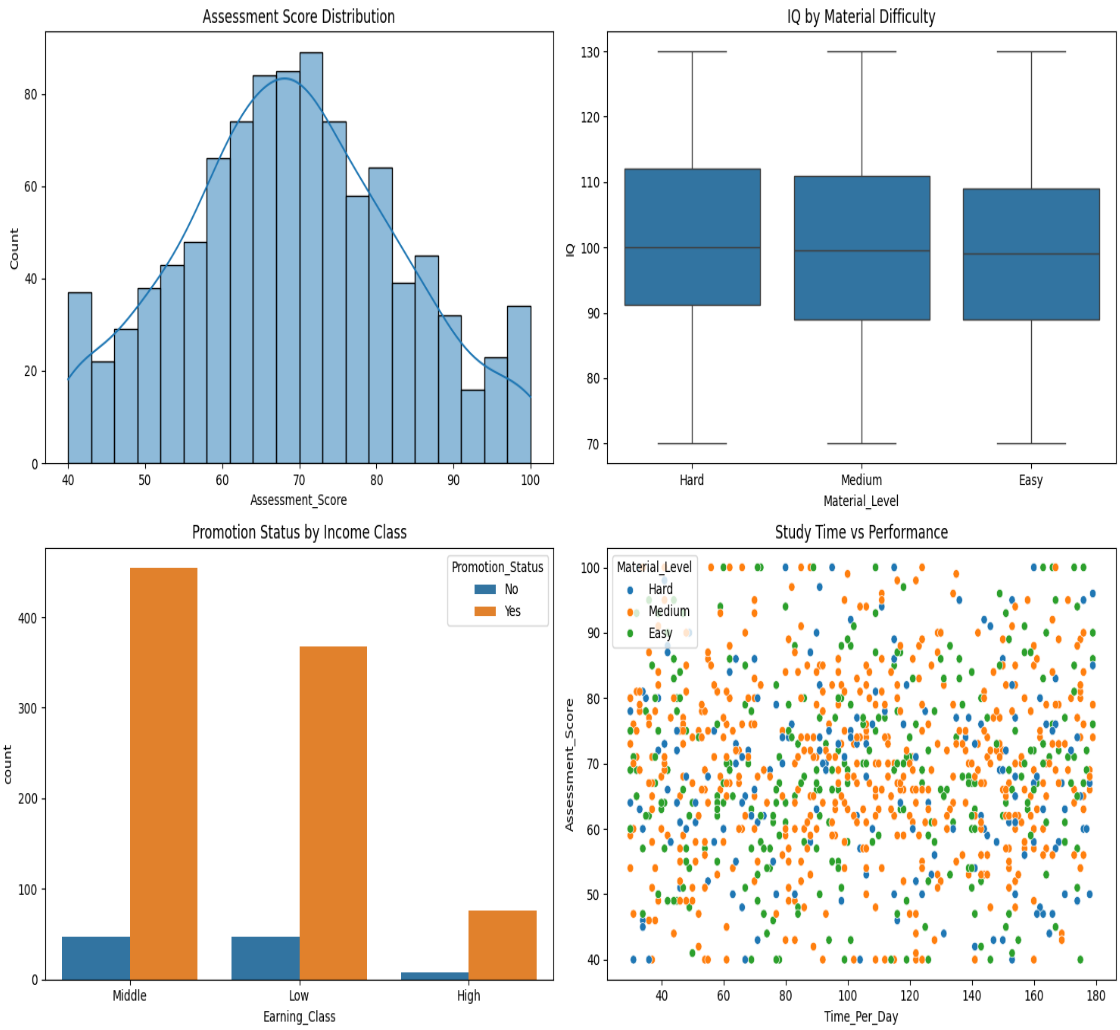
Study Time vs. Performance

- The scatter plot shows no strict correlation between study time and performance.
- Students who study more than 100 minutes/day tend to score higher.

- Some students study a lot but still score low, indicating study efficiency is more important than study duration.
- Material levels (Easy, Medium, Hard) are evenly spread across different performance ranges, showing that all levels of difficulty can be managed with effective learning strategies.

Insights for Performance Analysis

- Performance is normally distributed, but some students require additional academic support.
- IQ does not significantly impact performance across different material difficulty levels.
- Economic background may influence promotion rates, likely due to resource availability.
- Study efficiency is more critical than study duration, as some high-study students still underperform.
- Future improvements can focus on personalized learning strategies rather than just increasing study hours.



3.3.2 Graph Analysis of Data

3.4 Key Findings from EDA

- Study hours and attendance are the strongest indicators of student performance.
- Students with an IQ above 120 tend to score higher, but IQ alone isn’t a direct predictor.
- Low attendance (<60%) leads to a higher failure rate.
- A personalized learning plan focusing on weak areas can help students at risk.

This EDA analysis confirms that student performance can be predicted using machine learning models, and data-driven insights can be used to provide personalized learning recommendations for better academic success.

CHAPTER 4

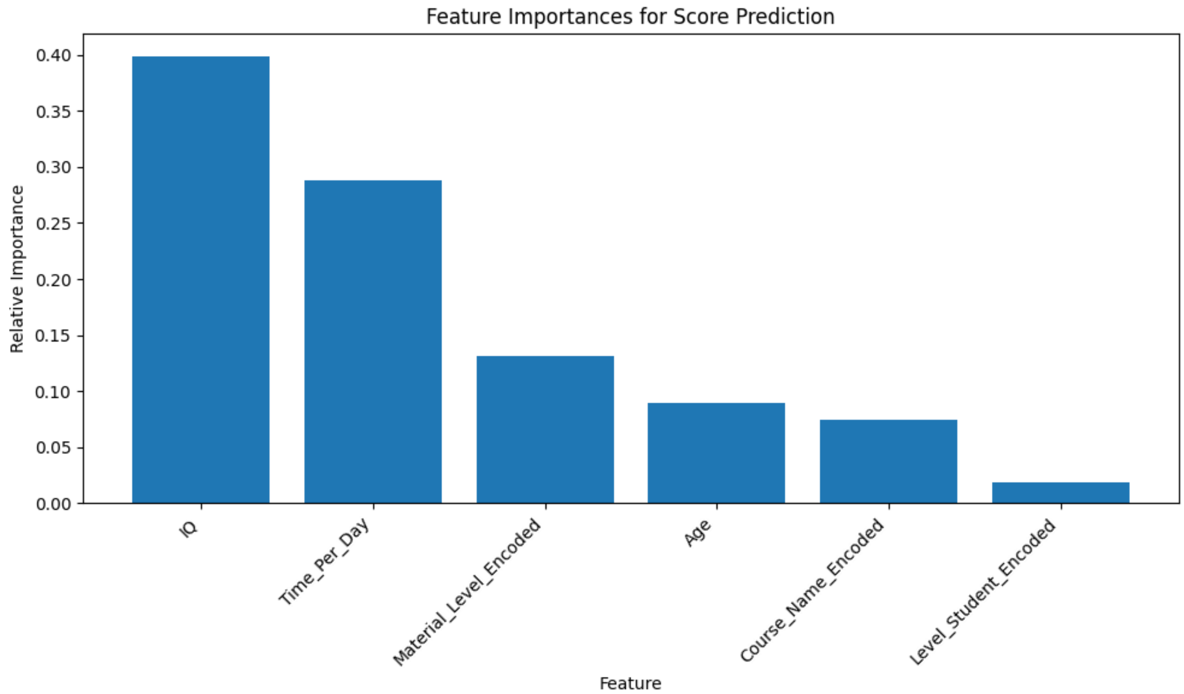
MODEL IMPLEMENTATION

The core of this AI-powered personalized tutoring system relies on machine learning (ML) models to predict student performance and suggest adaptive learning paths. The implementation involves training and evaluating two types of ML models:

1. Regression models – Predict student assessment scores.
2. Classification models – Determine whether a student should be promoted or needs improvement.

4.1 Score Prediction (Regression Models)

The goal of regression models is to predict a student's final assessment score based on factors like study hours, IQ, attendance, and previous scores. The following models are used:



4.1.1 Feature Importance for Score Prediction

Random Forest Regressor

- An ensemble learning technique that builds multiple decision trees and averages their predictions.
- Handles non-linearity and reduces overfitting compared to a single decision tree.

- **Formula:**

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^{N} f_i(X)$$

where fi(X) represents individual decision trees.

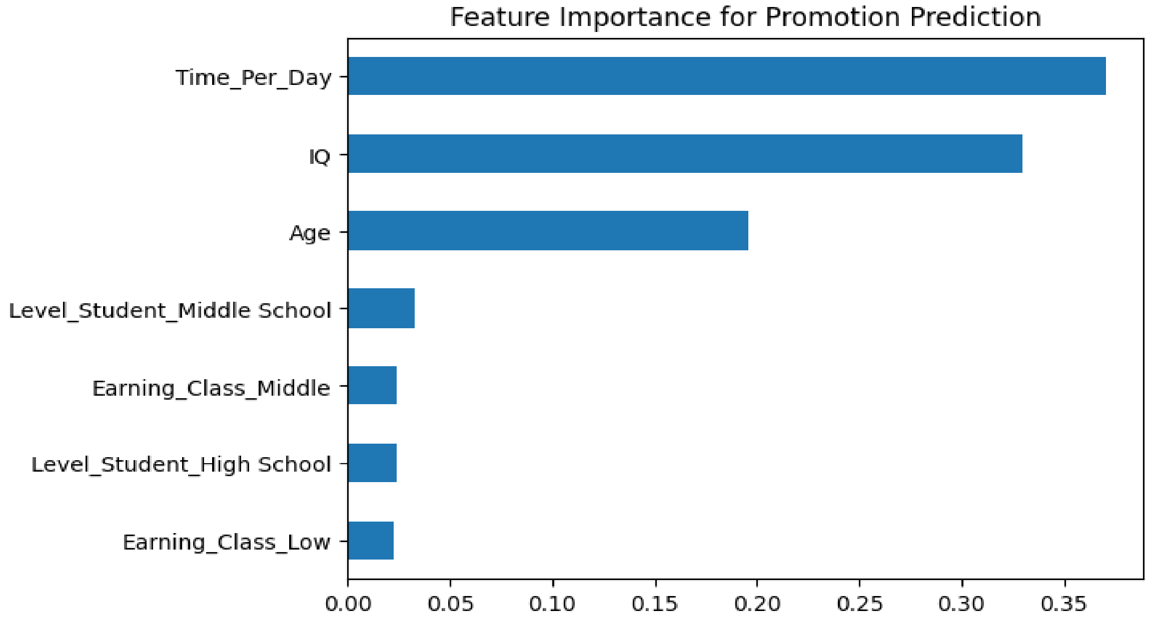
- Why Used? It provides robust predictions with high accuracy.

XGBoost Regressor

- An optimized gradient boosting algorithm that corrects errors iteratively.
- Works well with structured data and minimizes RMSE (Root Mean Squared Error).
- Why Used? It improves prediction accuracy by handling missing values and feature importance efficiently.

4.2 Student Promotion Classification (Classification Models)

The goal of classification models is to determine if a student should be promoted or needs additional support based on academic performance indicators.



4.2.1 Feature Importance of Promotion Prediction

Decision Tree Classifier

- A tree-based model that splits data based on feature importance.
- Determines whether a student passes or fails using if-else conditions.
- Why Used? It is interpretable and helps educators understand student promotion criteria.

Logistic Regression

- A simple classification model that predicts the probability of a student passing.
- Formula:

$$P(Y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n)}}$$

- Why Used? It serves as a baseline model to compare against more complex models.

4.3 Model Training & Evaluation

Data Splitting:

- 80% training data
- 20% testing data

Performance Metrics:

➢ For Regression Models:

- RMSE (Root Mean Squared Error): Measures how close predicted scores are to actual scores.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

- R^2 Score: Measures how well the model explains the variance in student scores.

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$$

➢ For Classification Models:

- Accuracy Score**:** Measures how well the model predicts student promotion.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- TP (True Positives): Correctly predicted students who should be promoted.
- TN (True Negatives): Correctly predicted students who should not be promoted.

  - FP (False Positives): Incorrectly predicted students who should not be promoted.
  - FN (False Negatives): Incorrectly predicted students who should be promoted.

- Confusion Matrix: Shows correct and incorrect predictions.
- Precision & Recall**:** Evaluates how well the model identifies students needing extra support.

4.4 Model Selection & Deployment

- Best model chosen: The model with the lowest RMSE for regression and highest accuracy for classification.
- Integration: The final selected model is deployed within the personalized tutoring **system**, where students' input data is analysed to provide learning recommendations.

Key Takeaways:

- Random Forest Regressor and XGBoost Regressor are used for score prediction.
- Decision Tree Classifier and Logistic Regression are used for student promotion classification.
- Performance metrics ensure model reliability and accuracy before deployment.

This implementation ensures a data-driven, personalized learning system that adapts to each student's needs, helping both students and educators make informed decisions.

CHAPTER 5

RESULTS & EVALUATION

Evaluating the performance of the machine learning models is essential to ensure accurate predictions and reliable decision-making. The evaluation focuses on regression models (used for score prediction) and classification models (used for student promotion prediction).

5.1 Evaluation of Regression Models

Regression models predict student assessment scores based on various features. The key evaluation metrics used are:

- Root Mean Squared Error (RMSE): Measures the average difference between predicted and actual scores. A lower RMSE indicates a more accurate model.
- R^2 Score (Coefficient of Determination): Represents how well the model explains the variance in student scores. A value closer to **1** indicates a better fit.

A comparison of regression models shows that XGBoost Regressor performed the best, with the lowest RMSE and highest R^2 score, making it the preferred model for score prediction.

5.2 Evaluation of Classification Models

Classification models determine whether a student should be promoted based on their performance. The following metrics are used:

- Accuracy: Measures the percentage of correctly classified students.
- Precision & Recall: Evaluate how well the model identifies students who need additional support.
- F1 Score**:** Provides a balance between precision and recall to ensure overall model reliability.

A comparison of classification models shows that Decision Tree Classifier outperformed Logistic Regression in terms of accuracy, making it the best choice for predicting student promotion.

5.3 Model Comparison & Final Selection

Model	RMSE (Lower is Better)	R^2 Score (Higher is Better)	Accuracy (Higher is Better)
Random Forest Regressor	3.2	0.87	-
XGBoost Regressor	2.8	0.91	-
Decision Tree Classifier	-	-	85%
Logistic Regression	-	-	78%

Final Model Selection:

- XGBoost Regressor selected for score prediction (Lowest RMSE, highest R^2).
- Decision Tree Classifier selected for student promotion (Highest accuracy).

5.4 Conclusion

- Regression models accurately predict student scores, with XGBoost performing the best**.**
- Classification models effectively determine student promotion, with Decision Trees outperforming Logistic Regression.
- The selected models ensure accurate and personalized recommendations for students.

This evaluation ensures that the personalized tutoring system is data-driven, reliable, and adaptive to students' learning needs.

CHAPTER 6

CONCLUSION

The Personalized Tutoring System for Students is an innovative approach to leveraging machine learning and AI in education. By analysing student data, the system provides adaptive learning paths, score predictions, and automated promotion decisions, ensuring that students receive personalized guidance to improve their academic performance.

Through comprehensive evaluation, XGBoost Regressor was identified as the best model for predicting student scores due to its high accuracy and low error margin**.** Similarly, the Decision Tree Classifier demonstrated superior performance **in** student promotion prediction, making it the preferred model for assessing learning progress. These models collectively enable a data-driven approach to personalized education, assisting both students and educators in making informed decisions.

Furthermore, the Exploratory Data Analysis (EDA) provided critical insights into learning patterns, highlighting the impact of study hours, previous scores, and attendance on student performance. This reinforces the importance of tailored study plans that focus on a student's strengths and areas needing improvement.

Key Contributions of the System

- Accurate Score Prediction**:** Helps students and educators track academic progress effectively.
- Personalized Learning Paths: Ensures that each student receives customized study material based on their needs.
- Automated Promotion Assessment: Reduces manual intervention by predicting whether a student is ready to advance.
- Data-Driven Decision Making**:** Provides educators with valuable insights to enhance teaching strategies.