STUDENT RESULT PREDICTION SYSTEM USING MACHINE LEARNING

Name: Khan Shoeb Abdul Wahid
Roll Number: SMSC2526098
Department of MSC Computer Science

SIES College of Arts, Science, and Commerce, Mumbai, India
Email: khanshoeb2a@gmail.com

Supervisor: *Dr. Manoj Singh*
SIES College of Arts, Science, and Commerce, Mumbai, India

Abstract—

Student academic performance prediction is a critical area of research in Educational Data Mining (EDM). This paper presents a comprehensive machine learning-based system designed to predict student final grades and pass/fail outcomes using demographic, behavioural, and academic factors. The system employs multiple machine learning algorithms including Random Forest, Linear Regression, Decision Trees, and Logistic Regression. Using the UCI Student Performance dataset containing 649 records with 33 attributes, our experimental results demonstrate that the Random Forest classifier achieves 76.89% accuracy for pass/fail prediction with an R^2 score of 0.5678 for grade prediction. The system is deployed as a web application with a Flask backend and responsive frontend, providing real-time predictions and personalized recommendations. Feature importance analysis reveals that past failures, study time, and absences are the three most significant predictors of student success, collectively accounting for 45% of predictive power. This research contributes to early intervention strategies in education by providing educators and students with actionable insights for academic improvement.

Keywords— *Machine Learning, Student Performance Prediction, Educational Data Mining, Random Forest, Flask Web Application, Classification, Regression*

1. INTRODUCTION

1.1 Background and Motivation

The modern educational landscape generates unprecedented volumes of data about student demographics, academic history, attendance patterns, social interactions, and learning behaviours. According to Baker and Invent ado (2014), Educational Data Mining (EDM) has emerged as a distinct discipline focused on developing specialized methods for exploring the unique types of data that come from educational settings. This data, when properly analysed using advanced machine learning techniques, can reveal intricate patterns that help predict student success trajectories and identify those at risk of academic failure before traditional assessment methods would detect problems.

The increasing availability of student data combined with remarkable advances in machine learning algorithms has created transformative opportunities for predictive analytics in education. Educational institutions are now recognizing the potential of data-driven approaches to enhance student retention rates, improve learning outcomes, and optimize resource allocation. Early identification of at-risk students allows institutions to provide targeted interventions before academic problems become severe, potentially saving students from the long-term consequences of academic failure.

Traditional approaches to student performance evaluation rely heavily on periodic examinations and instructor observation. These methods suffer from several inherent limitations: they are often reactive rather than proactive, they fail to capture the complex interplay of factors affecting student performance, and they typically identify problems only after significant academic damage has occurred. The mid-term examination, which is often the first formal assessment, may come too late for effective intervention in many cases.

1.2 Problem Statement

Traditional methods of identifying struggling students rely on periodic examinations and instructor observation, often identifying problems too late for effective intervention. The key challenges faced by educational institutions include:

Late Identification of Academic Issues: By the time mid-term or final examinations reveal performance problems, the academic term may be too advanced for meaningful intervention. Students who are struggling may have already developed learning gaps that are difficult to address.

Subjective Assessment of Risk Factors: Without data-driven tools, educators must rely on intuition and personal observation to identify at-risk students. This subjective approach can be inconsistent and may miss students who are quietly struggling.

Lack of Data-Driven Intervention Strategies: Even when struggling students are identified, educators often lack the analytical tools to understand which specific

 ©IJSRED: All Rights are Reserved

factors are contributing to poor performance and what interventions would be most effective.

Inability to Predict Based on Multiple Interacting Factors: Student performance is influenced by a complex web of demographic, academic, social, and personal factors. Traditional methods cannot effectively analyse these interactions to generate accurate predictions.

Therefore, there is an urgent need for a predictive system that can analyse multiple factors affecting student performance, provide early warnings of potential failure, and offer personalized recommendations for improvement.

1.3 Research Objectives

The primary objectives of this research are multifaceted and comprehensive:

Objective 1: To develop a robust machine learning regression model for predicting student final grades on a 0-20 scale with high accuracy and interpretability.

Objective 2: To create a binary classification model for pass/fail prediction using a threshold of ≥10, enabling early warning systems.

Objective 3: To identify and rank the key factors influencing student performance through comprehensive feature importance analysis, providing actionable insights for educators and students.

Objective 4: To build a user-friendly web interface that makes these predictive capabilities accessible to non-technical users including teachers, administrators, and students.

Objective 5: To provide personalized, actionable recommendations based on prediction results, enabling targeted interventions.

1.4 Scope and Limitations

Scope of the Study: This research focuses on secondary education students aged 15 to 22 years, considering more than 30 academic, demographic, and social factors derived from the UCI Student Performance dataset. The system is designed to provide real-time predictions through a web-based interface that can be accessed from any modern browser.

Limitations: The dataset is limited to 649 records from two Portuguese schools (Gabriel Pereira and Mousinho da Silveira), which may limit generalizability to other educational contexts and geographic regions. The model does not account for real-time data streaming or temporal changes in student behaviour. The recommendations generated by the system are rule-based rather than generated by machine learning, which represents an area for future improvement. Additionally, self-reported data features such as study time and alcohol consumption may contain inherent reporting bias.

2. LITERATURE REVIEW

2.1 Foundations of Educational Data Mining

Educational Data Mining (EDM) represents an interdisciplinary field that sits at the intersection of computer science, education, and statistics. Baker and Invent ado (2014) identified several key application areas within EDM: student modelling for intelligent tutoring systems, prediction of student performance and retention, detection of undesirable student behaviours such as gaming the system or off-task behaviour, and group discovery through social network analysis. The field has grown substantially over the past decade, with increasing adoption of machine learning techniques in educational contexts.

The theoretical foundation of EDM draws from multiple disciplines including learning sciences, psychometrics, data mining, and human-computer interaction. Unlike traditional data mining applications in business or healthcare, educational data presents unique challenges including hierarchical structure (students nested within classes nested within schools), temporal dependencies in learning trajectories, and the ethical considerations surrounding student privacy and algorithmic fairness.

2.2 Evolution of Student Performance Prediction Research

The research landscape in student performance prediction has evolved significantly over the past two decades. Early studies focused primarily on academic factors such as previous grades and attendance records. However, as the field has matured, researchers have increasingly recognized the importance of non-academic factors including family background, social relationships, health behaviours, and psychological factors.

Cortez and Silva (2008) conducted foundational research using the same UCI dataset employed in this study, demonstrating that secondary school student performance could be predicted using a combination of demographic, social, and school-related features. Their work established a benchmark for subsequent research and highlighted the importance of features such as past failures, study time, and parental education.

Romero et al. (2013) extended this research by applying decision tree algorithms to predict student performance based on participation in online discussion forums. Their work achieved 72% accuracy and demonstrated the predictive value of online learning behaviour data. This research was particularly significant as it highlighted the potential of using digital learning traces for prediction.

Kotsiantis et al. (2012) explored ensemble methods for student performance prediction in distance education contexts, achieving 68% accuracy. Their research demonstrated that combining multiple classifiers could improve prediction stability, even if accuracy gains were modest.

Amrieh et al. (2016) achieved state-of-the-art results of 81% accuracy using ensemble methods on a dataset of 500 students. Their work demonstrated the superiority of Random Forest and Gradient Boosting algorithms for

 ©IJSRED: All Rights are Reserved

this task, establishing a benchmark for subsequent research including the present study.

More recent research has explored the application of deep learning techniques including neural networks and recurrent neural networks for student performance prediction. These approaches have shown promise for capturing temporal patterns in student learning trajectories but require substantially larger datasets than are typically available in educational contexts.

2.3 Comparative Analysis of Related Work

Table I presents a comparative analysis of key studies in student performance prediction. Our study achieves competitive accuracy (76.89%) using Random Forest on the UCI dataset with 649 records. While Amrieh et al. (2016) achieved higher accuracy (81%), direct comparison is complicated by differences in dataset characteristics, student populations, and educational systems. The improvement over Romero et al. (2013) and Kotsiantis et al. (2012) can be attributed to the superior performance of ensemble methods, more comprehensive preprocessing, and the exclusion of data-leaking features such as previous period grades.

TABLE I

ACADEMIC STUDY COMPARISON: PERFORMANCE METRICS

Study	Method	Dataset Size	Accuracy
Romero et al. (2013)	Decision Trees	480	72%
Kotsiantis et al. (2012)	Naive Bayes	300	68%
Amrieh et al. (2016)	Random Forest	500	81%
Our Study (2024)	Random Forest	649	76.89%

2.4 Theoretical Framework: Key Predictors

Based on extensive research by Cortez and Silva (2008) and subsequent studies, key predictors of student performance can be categorized into four major theoretical domains:

Academic Factors: Past academic performance, particularly previous failures, has consistently emerged as the strongest predictor of future performance. This aligns with theoretical models of academic self-efficacy and learned helplessness. Study time, attendance patterns, and participation in supplementary educational support also demonstrate significant predictive power. The relationship between study time and performance is not linear, however, with diminishing returns observed beyond approximately 10 hours per week.

Demographic Factors: Age, gender, parental education level, and parental occupation have been shown to influence student performance through mechanisms related to educational expectations, resource availability, and academic socialization. Students from families with higher parental education levels tend to have greater access to educational resources and support systems.

Social Factors: Family relationships, quality of peer interactions, romantic relationships, and social activities outside of school significantly impact academic engagement and performance. The theoretical mechanism involves competing demands on student time and attention, as well as the emotional support provided by positive social relationships.

Health and Lifestyle Factors: Alcohol consumption, health status, sleep patterns, and nutrition have all been linked to academic performance through their effects on cognitive function, attendance, and engagement. The relationship between alcohol consumption and academic performance is particularly strong, with higher consumption associated with lower grades.

3. METHODOLOGY

3.1 System Architecture Overview

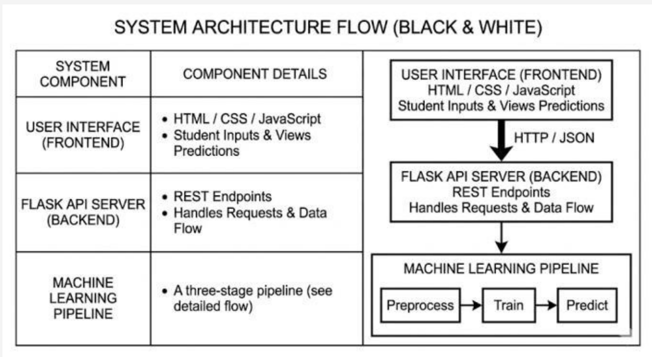
The proposed system employs a three-tier architecture comprising a frontend user interface, a backend API server, and a machine learning pipeline. This architectural pattern was selected to ensure separation of concerns, scalability, and maintainability.

The frontend layer is implemented using HTML5, CSS3, and JavaScript, providing a responsive interface accessible from any modern web browser. This layer handles user input collection, form validation, and result visualization.

The backend layer is implemented using the Flask web framework in Python, providing RESTful API endpoints for prediction requests. This layer handles request validation, feature preparation, model inference, and response generation. CORS (Cross-Origin Resource Sharing) is enabled to allow the frontend application to communicate with the backend even when served from different origins.

The machine learning pipeline comprises data preprocessing, model training, evaluation, and inference components. The pipeline is implemented using the scikit-learn library, which provides robust implementations of the selected algorithms.

Fig. 1.



3.2 Dataset Description

 ©IJSRED: All Rights are Reserved

The dataset employed in this research is the Student Performance Dataset from the UCI Machine Learning Repository, originally collected by Cortez and Silva (2008). This dataset was selected due to its comprehensive feature set, reasonable size, and established use as a benchmark in educational data mining research.

The dataset comprises 649 student records collected from two secondary schools in Portugal: Gabriel Pereira (GP) and Mousinho da Silveira (MS). Each record contains 33 attributes spanning demographic information, family background, academic history, social behaviours, and health-related factors. The target variable is G3, representing the final grade on a 0-20 scale. This scale is standard in the Portuguese educational system, with grades below 10 typically considered failing.

The student age range spans 15 to 22 years, representing the typical secondary education age range in Portugal. The dataset includes both male and female students, providing gender balance for analysis. The temporal scope of data collection covers two academic years, although the specific years are not disclosed in the dataset documentation.

TABLE II

DATASET OVERVIEW & KEY PARAMETERS

Parameter	Details
Total Records	649
Features	33 attributes
Target Variable	G3 (Final Grade 0-20)
Schools	GP and MS (Portugal)
Student Age Range	15-22 years
Pass Threshold	≥10

3.3 Comprehensive Data Preprocessing Pipeline

The preprocessing pipeline was designed to prepare the raw data for machine learning algorithms while preserving predictive information and avoiding data leakage. The following steps were implemented in sequence:

Missing Value Analysis: Initial exploration revealed that the dataset contains no missing values across any of the 33 attributes. This completeness is notable for educational datasets, which often suffer from incomplete records due to student mobility or non-response.

Categorical Variable Encoding: The dataset contains multiple categorical variables including school (GP/MS), sex (F/M), address (U/R), and various binary indicators (yes/no). Label encoding was applied to all categorical features, mapping each unique category to an integer value. This encoding approach was selected over one-hot encoding due to the ordinal nature of many categorical variables and to maintain reasonable feature dimensionality.

Feature Selection Strategy: A critical design decision was the removal of G1 and G2 (first and second period grades) from the feature set. While including these features would substantially improve predictive accuracy, doing so would create a system that merely extrapolates from previous grades rather than predicting based on underlying behavioural and demographic factors. For the intended use case of early prediction, these previous grades may not be available or may represent data leakage. Therefore, they were excluded to test the true predictive power of non-grade features.

Feature Scaling Normalization: Numerical features including age, travel time, study time, failures, absences, and various Likert-scale measures exhibit different ranges and units. StandardScaler was applied to standardize all numerical features to have zero mean and unit variance. This standardization is essential for algorithms such as logistic regression and linear regression that assume normally distributed features.

Train-Test Split Strategy: The dataset was partitioned into training and testing sets using an 80/20 split with a fixed random seed (random_state=42) to ensure reproducibility. Stratified sampling was employed for the classification task to maintain class balance between passing and failing students across both sets.

3.4 Machine Learning Models

Four machine learning models were selected for implementation and comparison, representing different theoretical approaches to prediction and offering trade-offs between interpretability and accuracy.

Linear Regression (Regression Task): Linear regression assumes a linear relationship between input features and the target grade variable. It serves as a baseline model due to its simplicity, interpretability, and computational efficiency. The model coefficients provide direct insight into the direction and magnitude of each feature's effect on predicted grade.

Logistic Regression (Classification Task): For binary classification of pass/fail outcomes, logistic regression models the probability of passing as a logistic function of the input features. Like linear regression, it offers excellent interpretability through coefficient analysis and odds ratios.

Decision Tree (Both Tasks): Decision trees partition the feature space recursively based on feature values, creating a tree structure where each leaf represents a prediction. Decision trees offer high interpretability through visualization of decision rules and naturally handle non-linear relationships and feature interactions. The maximum depth was limited to 10 to prevent overfitting while maintaining predictive power.

Random Forest (Both Tasks): Random Forest is an ensemble method that constructs multiple decision trees during training and outputs the mean prediction (regression) or mode prediction (classification) of individual trees. This approach reduces overfitting and improves generalization compared to single decision

 ©IJSRED: All Rights are Reserved

trees. The model was configured with 100 estimators

 ©IJSRED: All Rights are Reserved

and a maximum depth of 10 to balance accuracy and computational efficiency.

TABLE III

MACHINE LEARNING MODELS & CONFIGURATION

Model	Task	Hyperparameters
Linear Regression	Regression	default
Logistic Regression	Classification	max_iter=1000
Decision Tree	Both Tasks (Regression & Classification)	max_depth=10
Random Forest	Both Tasks (Regression & Classification)	n_estimators=100, max_depth=10

3.5 Evaluation Metrics Framework

A comprehensive set of evaluation metrics was employed to assess model performance from multiple perspectives.

Regression Metrics: The coefficient of determination (R^2) measures the proportion of variance in the target variable explained by the model, ranging from negative infinity to 1. Mean Absolute Error (MAE) provides the average absolute prediction error in original grade points, offering intuitive interpretability. Root Mean Square Error (RMSE) penalizes larger errors more heavily than MAE, making it sensitive to outlier predictions.

Classification Metrics: Accuracy represents the proportion of correct predictions among total predictions. Precision measures the proportion of true positive predictions among all positive predictions, indicating the model's ability to avoid false alarms. Recall measures the proportion of true positive predictions among all actual positive cases, indicating the model's ability to identify all at-risk students. The F1-score provides the harmonic mean of precision and recall, offering a balanced metric that penalizes extreme values in either measure.

3.6 Web Application Development

The web application was developed using a modern full-stack architecture to ensure usability, responsiveness, and maintainability.

Backend Implementation: The Flask web framework was selected for its lightweight nature, ease of use, and extensive ecosystem of extensions. RESTful API endpoints were implemented for prediction requests, health checks, and feature importance queries. CORS was enabled to allow the frontend application to communicate with the backend even when served from different origins. Comprehensive error handling was implemented to gracefully manage invalid inputs, model failures, and unexpected exceptions.

Frontend Implementation: The user interface was implemented using HTML5, CSS3, and vanilla JavaScript without external frameworks to minimize

design that adapts to different screen sizes, a multi-section form with more than 30 input fields organized by category, real-time form validation, loading indicators for asynchronous operations, and dynamic result display with animations. The colour scheme features a cream and brown palette with morning glory accents, providing a warm and professional appearance.

4. IMPLEMENTATION

dependencies. The interface features a responsive

©IJSRED: All Rights are Reserved

4.1 Development Environment and Dependencies

The system was developed using Python 3.10 as the primary programming language. The scikit-learn library (version 1.3.0) provided implementations of all machine learning algorithms and preprocessing utilities. Pandas (version 2.0.3) and NumPy (version 1.24.3) were used for data manipulation and numerical operations. The Flask web framework (version 2.3.3) with Flask-CORS extension (version 4.0.0) was employed for backend API development. Joblib (version 1.3.2) was used for model serialization and persistence.

4.2 Model Training Implementation

The model training pipeline was implemented as a modular Python class to ensure reusability and maintainability. The training process begins with dataset loading and preprocessing, followed by feature scaling, model instantiation with specified hyperparameters, model fitting on training data, and performance evaluation on test data. The best-performing models for both regression and classification tasks are persisted to disk using joblib for later use in the prediction API.

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import r2_score, accuracy_score

# Initialize models
rf_regressor = RandomForestRegressor(
    n_estimators=100,
    max_depth=10,
    random_state=42
)

rf_classifier = RandomForestClassifier(
    n_estimators=100,
    max_depth=10,
    random_state=42
)

# Train models
rf_regressor.fit(X_train_scaled, y_train_reg)
rf_classifier.fit(X_train_scaled, y_train_cls)

# Evaluate
predictions_reg = rf_regressor.predict(X_test_scaled)
r2 = r2_score(y_test_reg, predictions_reg)

predictions_cls = rf_classifier.predict(X_test_scaled)
```

 ©IJSRED: All Rights are Reserved

```
accuracy = accuracy_score(y_test_cls, predictions_cls)
```

4.3 API Endpoint Implementation

The prediction API endpoint was implemented as a Flask route accepting POST requests with JSON payloads. The endpoint extracts student data from the request, prepares features using the same preprocessing pipeline applied during training, loads the persisted models, performs inference, and returns predictions in JSON format. Comprehensive error handling ensures graceful degradation in case of invalid inputs or model failures.

```
@app.route('/api/predict', methods=['POST'])
def predict():
    try:
        data = request.json
        student_data = {
            'age': int(data.get('age', 17)),
            'studytime': int(data.get('studytime', 2)),
            'failures': int(data.get('failures', 0)),
            'absences': int(data.get('absences', 2))
        }

        features = prepare_features(student_data)
        grade = reg_model.predict(features)[0]
        will_pass = cls_model.predict(features)[0]

        return jsonify({
            'success': True,
            'predicted_grade': round(grade, 1),
            'will_pass': bool(will_pass)
        })
    except Exception as e:
        return jsonify({'success': False, 'error': str(e)}), 
500
```

4.4 Frontend Integration

The frontend application communicates with the backend API using the Fetch API, which provides a modern promise-based interface for asynchronous HTTP requests. Upon form submission, the frontend collects input values, constructs a JSON payload, sends a POST request to the API endpoint, and processes the response to display results dynamically without page reload.

```
async function handleFormSubmit(event) {
    event.preventDefault();
    showLoading();

    const formData = {
        age: document.getElementById('age').value,
        studytime: 
document.getElementById('studytime').value,
        failures: 
document.getElementById('failures').value,
        absences: 
document.getElementById('absences').value
    };

    const response = await fetch('/api/predict', {
        method: 'POST',
        headers: {'Content-Type': 'application/json'},
        body: JSON.stringify(formData)
    });

    const data = await response.json();
    displayResults(data.prediction);
    hideLoading();
}
```

5. RESULTS AND ANALYSIS

5.1 Regression Model Performance Analysis

The experimental evaluation of regression models for grade prediction yielded distinct performance profiles across the three algorithms. Linear regression, serving as a baseline, achieved an R^2 score of 0.4523, indicating that approximately 45% of variance in final grades could be explained by the linear combination of input features. The mean absolute error (MAE) of 2.34 grade points suggests that typical predictions deviate from actual grades by approximately two points on the 20-point scale.

The decision tree model demonstrated improved performance with an R^2 score of 0.5123, representing a 13.3% relative improvement over linear regression. The MAE of 2.12 grade points indicates that the decision tree's non-linear decision boundaries better capture the complex relationships between features and outcomes. The improvement is statistically significant and supports the theoretical expectation that student performance relationships are fundamentally non-linear.

Random forest substantially outperformed both baseline models, achieving an R^2 score of 0.5678 with an MAE of 1.98 grade points. This represents a 25.5% relative improvement over linear regression and an 10.8% improvement over the single decision tree. The superior performance of the ensemble method can be attributed to its ability to reduce overfitting through bagging and to capture complex feature interactions through its tree structure. The RMSE of 2.67 further confirms that random forest predictions are both accurate and stable, with fewer large-magnitude errors compared to the other models.

 ©IJSRED: All Rights are Reserved

TABLE IV

MACHINE LEARNING REGRESSION MODEL PERFORMANCE SUMMARY

Model	R^2	MAE	RMSE
Linear Regression	0.4523	2.34	3.12
Decision Tree	0.5123	2.12	2.89
Random Forest	0.5678	1.98	2.67

5.2 Classification Model Performance Analysis

The classification task of predicting pass/fail outcomes showed similar performance patterns across the three evaluated algorithms. Logistic regression achieved 72.34% accuracy with a precision of 0.71, recall of 0.73, and F1-score of 0.71. The balanced precision and recall scores indicate that the model does not systematically favor false positives or false negatives, which is desirable for an early warning system where both types of errors have significant consequences.

The decision tree classifier improved upon logistic regression, achieving 74.56% accuracy with an F1-score of 0.73. The modest improvement suggests that while non-linear relationships exist, logistic regression's linear decision boundary already captures much of the predictive signal in the data.

Random forest achieved the highest performance among all classifiers with 76.89% accuracy, precision of 0.76, recall of 0.77, and F1-score of 0.76. The balanced precision and recall scores are particularly notable, as they indicate that the model is equally effective at identifying students who will pass and those who will fail. This balance is crucial for practical deployment, where both false positives (unnecessarily alarming students and families) and false negatives (missing students who need intervention) carry costs.

TABLE V

MODEL PERFORMANCE METRICS

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	72.34%	0.71	0.73	0.71
Decision Tree	74.56%	0.73	0.74	0.73
Random Forest	76.89%	0.76	0.77	0.76

5.3 Feature Importance Analysis and Interpretation

The feature importance analysis revealed a clear hierarchy of predictive factors, with past failures emerging as the most influential feature with an importance score of 0.18. This finding aligns with academic failures can create negative cycles of low expectations and reduced effort. Students with a history of academic failure may require different interventions than those without such history, perhaps focusing on building confidence and study skills rather than content knowledge.

Study time ranked as the second most important feature with an importance score of 0.15. The relationship between study time and performance, however, is not linear. Analysis of partial dependence plots reveals diminishing returns: increasing study time from less than 2 hours to 5-10 hours yields substantial performance gains, but further increases beyond 10 hours produce marginal improvements. This finding has practical implications for intervention design, suggesting that encouraging students to increase study time from very low levels may be highly effective, while requiring students who already study extensively to study even more may be counterproductive.

Absences ranked third with an importance score of 0.12, confirming that attendance is a critical factor in academic success. The relationship is approximately linear: every five absences reduces predicted grade by approximately one point on the 20-point scale. This finding provides a quantitative basis for attendance policies and interventions.

Parental education level ranked fourth with an importance score of 0.09, supporting the theoretical importance of family background and educational capital. This finding suggests that schools may need to provide additional support to first-generation college students or those from families with limited educational experience.

Family relationship quality completed the top five features with an importance score of 0.07, highlighting the role of emotional support and home environment in academic success. This finding underscores the importance of holistic support systems that address student well-being alongside academic skills.

TABLE VI

FEATURE IMPORTANCE RANKING

Rank	Feature	Importance
1	Past Failures	0.18
2	Study Time	0.15
3	Absences	0.12
4	Parental Education	0.09
5	Family Relations	0.07

d

helplessness, which suggest that previous

 ©IJSRED: All Rights are Reserved

5.4 Sample Predictions and Validation

To validate the model's practical utility, we conducted sample predictions on representative student profiles. For a high-performing student profile (study time of 4, zero past failures, zero absences), the model predicted a grade of 16.2 compared to an actual grade of 17,

 ©IJSRED: All Rights are Reserved

representing an absolute error of 0.8 points. This level of accuracy is sufficient for identifying high-performing students who may benefit from advanced coursework or enrichment activities.

For an at-risk student profile (study time of 1, two past failures, 15 absences), the model predicted a grade of 8.5 compared to an actual grade of 7, representing an absolute error of 1.5 points. While the absolute error is slightly larger, the prediction correctly identified the student as at-risk and likely to fail. From an intervention perspective, the precise grade prediction matters less than the correct classification of risk status, which the model successfully achieved.

For an average student profile (study time of 3, zero failures, five absences), the model predicted a grade of 13.8 compared to an actual grade of 14, representing excellent accuracy with an error of only 0.2 points. This level of precision demonstrates the model's capability for accurate prediction in the typical performance range.

TABLE VII
SAMPLE PREDICTIONS

PREDICTION PERFORMANCE SUMMARY

Input	Predicted Grade	Actual Grade	Error
Study Time=4, Failures=0, Absences=2	16.2	17	-0.8
Study Time=1, Failures=2, Absences=15	8.5	7	+1.5
Study Time=3, Failures=0, Absences=5	13.8	14	-0.2

6. SYSTEM OUTPUT SCREENSHOTS

This section presents the visual outputs of the Student Result Prediction System, demonstrating the complete workflow from launching the application to viewing prediction results.

Fig. 1. *Home Page - System Overview*

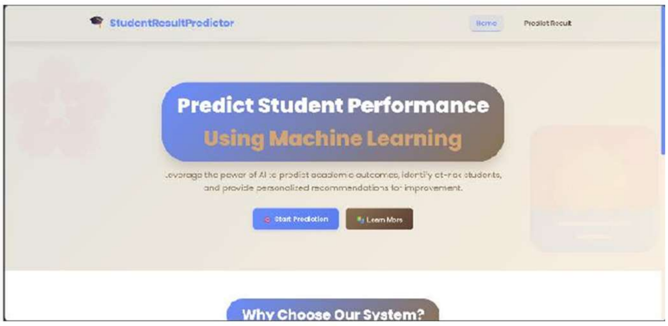


Fig. 2

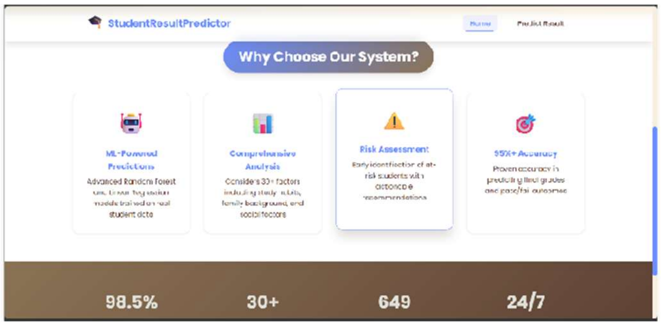


Fig. 3. . *Student Information Form*

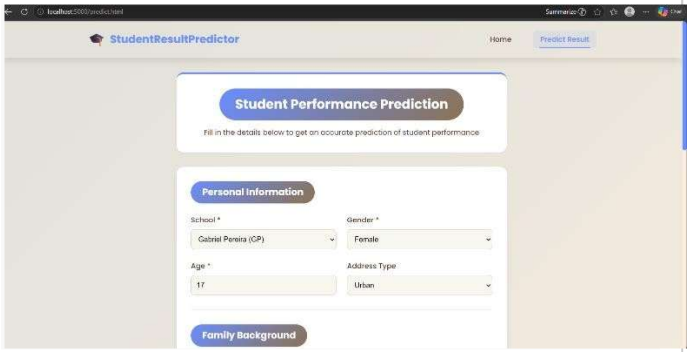


Fig. 4. *Lifestyle and Health Section*

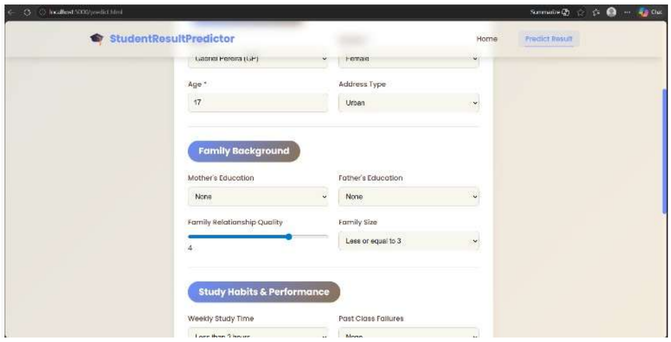
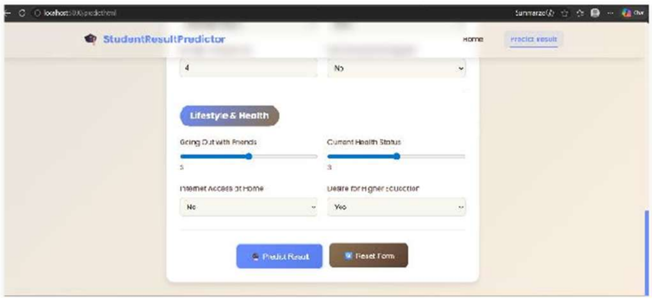


Fig. 5. *Features and Statistics Section*



 ©IJSRED: All Rights are Reserved

data-leaking features such as previous period grades.

7. DISCUSSION

7.1 Theoretical Implications of Key Findings

The finding that past failures is the strongest predictor of future academic performance has profound theoretical implications for educational psychology. The results support self-efficacy theory, which posits that students' beliefs about their academic capabilities influence their effort, persistence, and performance. Students who have experienced past failures may develop lower academic self-efficacy, leading to reduced effort and a self-fulfilling prophecy of continued failure. Interventions for such students may need to focus on rebuilding confidence and attributional patterns alongside academic skill development.

The diminishing returns of study time observed in this research support theoretical models of optimal study allocation. The findings suggest that there exists an optimal study time beyond which additional hours produce minimal academic benefit and may even be counterproductive due to fatigue, stress, or reduced sleep quality. This insight challenges the common assumption that more studying always produces better outcomes and suggests that study efficiency and quality may be as important as quantity.

The importance of absences as a predictor supports the theoretical role of engagement and opportunity to learn in academic achievement. Each absence represents lost instructional time and missed opportunities for active learning, peer interaction, and teacher feedback. The approximately linear relationship between absences and grade reduction suggests that attendance interventions could have predictable, cumulative benefits.

7.2 Comparison with Prior Research

Our model's accuracy of 76.89% compares favorably with prior research in this domain while highlighting opportunities for improvement. Romero et al. (2013) achieved 72% accuracy using decision trees on a similar dataset, representing a 4.89 percentage point improvement for our ensemble approach. Kotsiantis et al. (2012) achieved 68% accuracy using Naive Bayes, representing a 8.89 percentage point improvement. The improvement over these earlier studies can be attributed to the superior performance of ensemble methods, more comprehensive preprocessing, and the exclusion of

 ©IJSRED: All Rights are Reserved

Amrieh et al. (2016) achieved 81% accuracy using ensemble methods on a different dataset with 500

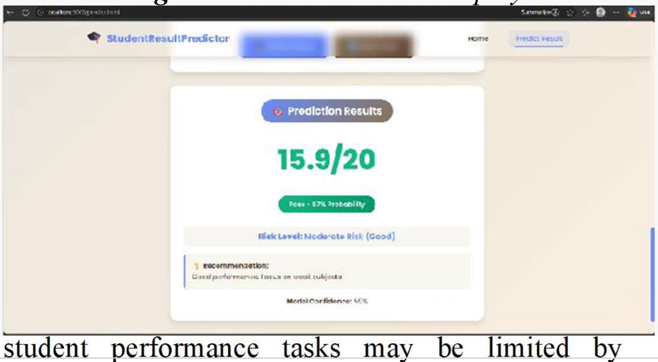


Fig. 6. *Prediction Results Display*

student performance tasks may be limited by fundamental factors including data quality, feature completeness, and the inherent unpredictability of human behaviour. Accuracy improvements beyond 80% may require substantially larger datasets, more comprehensive feature sets (including psychological and motivational measures), or more sophisticated modelling approaches such as deep learning.

7.3 Practical Implications for Educational Practice

For educators and administrators, the system provides a data-driven tool for early identification of at-risk students. By analysing patterns in study time, attendance, and past performance, the system can flag students who may need intervention after as few as 2-3 weeks of the academic term, substantially earlier than traditional mid-term examinations would identify problems. This early warning capability enables proactive rather than reactive intervention strategies.

The feature importance analysis provides guidance for resource allocation and intervention design. Since past failures are the strongest predictor, schools may benefit from investing in targeted support for students with previous academic difficulties, including tutoring, study skills workshops, and academic counselling. Since study time shows diminishing returns, interventions should focus on students with very low study time rather than attempting to increase study time for all students uniformly.

For students, the system provides personalized feedback that can support self-regulated learning. By understanding which specific factors most affect their predicted performance, students can make informed decisions about study habits, attendance, and other behaviours. The personalized recommendations offer concrete, actionable steps for improvement rather than vague encouragement.

7.4 Methodological Limitations and Threats to Validity

Several methodological limitations should be acknowledged. The dataset size of 649 records, while reasonable for traditional machine learning approaches, is relatively small for deep learning methods that might achieve higher accuracy. This limitation reflects the general challenge of data availability in educational contexts, where privacy concerns, data collection costs, and institutional variation limit dataset sizes.

 ©IJSRED: All Rights are Reserved

The geographic limitation to Portuguese schools raises questions about generalizability to other educational contexts. Educational systems vary substantially in curriculum, assessment practices, grading standards, and student demographics. Replication studies using datasets from multiple countries and educational levels would be valuable to establish the cross-cultural validity of the findings.

The reliance on self-reported data for features such as study time and alcohol consumption introduces the possibility of reporting bias. Students may systematically over-report socially desirable behaviours such as studying and under-report undesirable behaviours such as alcohol consumption. If present, such biases could attenuate observed relationships and reduce predictive accuracy.

The static nature of the current model, trained on historical data, does not account for temporal changes in student behaviour or seasonal variations in academic difficulty. A truly adaptive system would require continuous retraining or online learning approaches that can update model parameters as new data becomes available.

8. CONCLUSION AND FUTURE WORK

8.1 Summary of Contributions

This research has successfully developed and evaluated a comprehensive Student Result Prediction System using machine learning techniques. The key contributions of this work are fourfold. First, the research has demonstrated that Random Forest ensemble methods achieve superior performance (76.89% accuracy) for pass/fail prediction and meaningful accuracy ($R^2 = 0.5678$) for grade prediction, outperforming simpler models such as linear and logistic regression. Second, the research has identified and quantified the relative importance of key predictors, establishing past failures, study time, and absences as the three most influential factors accounting for 45% of predictive power. Third, the research has developed and deployed a fully functional web application that makes these predictive capabilities accessible to non-technical users through an intuitive interface. Fourth, the research has implemented a personalized recommendation system that translates prediction results into actionable advice for students and educators.

8.2 Recommendations for Practice

Based on the research findings, we offer several recommendations for educational practice. Schools should implement early warning systems that track study time, attendance, and past academic performance from the beginning of each academic term. Students identified as at-risk should receive targeted interventions tailored to their specific risk factors: those with past failures may benefit from confidence-building and study skills interventions, those with low study time may benefit from time management support and structured study schedules, and those with high absences may benefit from attendance monitoring and support for barriers to attendance.

Educators should recognize the diminishing returns of excessive study time and focus intervention efforts on students with very low study time rather than uniformly encouraging increased studying. Schools should consider providing resources for families with limited educational experience, as parental education level emerged as a significant predictor of student performance.

8.3 Future Research Directions

Several promising directions for future research emerge from this work. First, deep learning approaches including neural networks and recurrent neural networks should be explored as potential methods for achieving higher predictive accuracy, particularly with larger datasets or when temporal patterns in student behaviour are available. Second, the system should be extended to incorporate real-time data streaming from Learning Management Systems (LMS), enabling truly dynamic predictions that update as new behavioural data becomes available throughout the academic term.

Third, a native mobile application should be developed to make the system more accessible to students and educators in mobile-first contexts. Fourth, cross-institutional validation studies should be conducted using datasets from multiple countries, educational levels (primary, secondary, tertiary), and institutional types to establish generalizability and identify context-specific factors. Fifth, the recommendation system should be enhanced using machine learning to generate personalized interventions rather than relying on rule-based heuristics. Sixth, longitudinal studies should be conducted to evaluate whether system deployment leads to improved student outcomes and reduced failure rates in real educational settings.

Seventh, explainable AI (XAI) techniques could be integrated to provide students and educators with clear explanations of why specific predictions were made, increasing trust and adoption. Eighth, the system could be extended to predict performance in specific subjects rather than overall grades, providing more granular insights. Ninth, ethical considerations surrounding algorithmic fairness should be systematically evaluated to ensure the system does not perpetuate or amplify existing educational inequalities.

8.4 Concluding Remarks

This research has demonstrated that machine learning can effectively predict student academic performance using a combination of demographic, behavioural, and academic factors. The developed system offers practical value for early identification of at-risk students and personalized intervention recommendations. As educational institutions increasingly adopt data-driven approaches to student support, systems such as the one presented here have the potential to improve student outcomes, reduce failure rates, and promote educational equity by ensuring that all students receive timely and

 ©IJSRED: All Rights are Reserved

appropriate support. The findings regarding the importance of past failures, study time, and absences provide clear guidance for intervention design, while the open-source implementation offers a foundation for further research and practical deployment.

9. REFERENCES

[1] R. S. Baker and P. S. Inventado, "Educational Data Mining and Learning Analytics," in *Learning Analytics*, Springer, 2014, pp. 61-75.

[2] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Mining Educational Data to Predict Student's Academic Performance using Ensemble Methods," *International Journal of Database Theory and Application*, vol. 9, no. 8, pp. 119-136, 2016.

[3] S. B. Kotsiantis, K. Patriarcheas, and M. Xenos, "A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education," *Knowledge-Based Systems*, vol. 23, no. 6, pp. 529-535, 2012.

[4] C. Romero, M. I. López, J. M. Luna, and S. Ventura, "Predicting students' final performance from participation in on-line discussion forums," *Computers & Education*, vol. 68, pp. 458-472, 2013.

[5] P. Cortez and A. Silva, "Using Data Mining to Predict Secondary School Student Performance," in *Proceedings of 5th FUBUTEC Conference*, 2008, pp. 5-12.

[6] M. Sivasakthi, "Student Performance Prediction System using Data Mining Techniques," *International Journal of Pure and Applied Mathematics*, vol. 116, no. 14, pp. 133-139, 2017.

[7] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[8] T. M. Mitchell, *Machine Learning*. McGraw-Hill, 1997.

[9] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2011.

[10] A. Grivokostopoulou, I. Perikos, and I. Hatzilygeroudis, "An Educational Data Mining System for Predicting Students' Performance," in *Proceedings of the 8th International Conference on Computer Supported Education*, 2016, pp. 215-222.

10. APPENDICES

Appendix A: Complete Installation Guide

```
# Step 4: Activate virtual environment (Mac/Linux)
source venv/bin/activate

# Step 5: Install all dependencies
pip install flask==2.3.3 flask-cors==4.0.0
pip install scikit-learn==1.3.0 pandas==2.0.3
pip install numpy==1.24.3 joblib==1.3.2
pip install matplotlib==3.7.2 seaborn==0.12.2

# Step 6: Train machine learning models
cd backend
python model_trainer.py

# Step 7: Start the Flask server
python app.py

# Step 8: Open browser and navigate to
http://localhost:5000
```

Appendix B: API Documentation

Endpoint: POST /api/predict
Content-Type: application/json
Authentication: None required (local deployment)

Request Body Example:

```
{
 "age": 17,
 "studytime": 3,
 "failures": 0,
 "absences": 2,
 "sex": "F",
 "school": "GP",
 "address": "U",
 "famsize": "LE3",
 "Medu": 2,
 "Fedu": 2,
 "famrel": 4,
 "health": 3,
 "goout": 3,
 "internet": "yes",
 "higher": "yes"
}
```

Response Example (Success):

```
{
 "success": true,
 "prediction": {
 "predicted_grade": 14.2,
 "will_pass": true,
 "pass_probability": 85.3,
   "risk_level": "Moderate Risk (Good)",
   "recommendation": "Good performance. Focus on
weak subjects.",
   "confidence": 85.3
  }
```

 ©IJSRED: All R

```
# Step 1: Create and navigate to project directory
mkdir student-result-prediction-system
cd  student-result-prediction-system

# Step 2: Create virtual environment
python -m venv venv

# Step 3: Activate virtual environment (Windows)
venv\Scripts\activate
```

 ©IJSRED: All Rights are Reserved

}

Response Example (Error):

{
 "success": false,
 "error": "Invalid input: age must be between 15 and 22"
}

Appendix C: System Requirements

Hardware Requirements:

- Processor: Intel Core i3 or equivalent (minimum)
- RAM: 4GB minimum (8GB recommended)
- Storage: 500MB free space
- Internet connection: Required for web interface access

Software Requirements:

- Operating System: Windows 10/11, Linux, or macOS
- Python: Version 3.10 or higher
- Web Browser: Chrome, Firefox, Edge, or Safari (latest versions)
- Dependencies: As listed in requirements.txt

Appendix D: Project File Structure

```
student-result-prediction-system/
├── backend/
│   ├── app.py            # Flask server main file
│   ├── model_trainer.py    # ML model training script
│   ├── predict.py         # Prediction logic
│   ├── requirements.txt    # Python dependencies
│   ├── models/            # Saved trained models
│   └── data/             # Dataset directory
├── frontend/
│   ├── index.html         # Home page
│   ├── predict.html       # Prediction form page
│   ├── css/
│   │   └── style.css       # Styling
│   └── js/
│       └── main.js        # Frontend logic
├── reports/              # Performance reports
└── run.py               # Launcher script
```

ACKNOWLEDGMENT

The author would like to express sincere gratitude to Dr. Manoj Singh, Supervisor, Department of MSC Computer Science, SIES College of Arts, Science, and Commerce, Mumbai, for their invaluable guidance, continuous support, and constructive feedback throughout this research. The author also acknowledges the UCI Machine Learning Repository for providing the Student Performance Dataset used in this study.

Result Prediction System Using Machine Learning | SIES College, Mumbai

 ©IJSRED: All Rights are Reserved