STUDENT PERFORMANCE PREDICTION SYSTEM

PROJECT SYNOPSIS REPORT

GULZAR GROUP OF INSTITUTIONS

(Affiliated to IKG Punjab Technical University, Kapurthala)

Submitted by:
Garv Sehgal (2339273)
Vasu Mahajan (2232187)
Tanuja Khumanthem (2232181)
Passang Tshering Sherpa (2339280)
Badal Rai (2339267)

Under the Guidance of:
(Er. Mandeep Kaur)

Department: Computer Science Engineering
Year: 4th (2025)

Academic Session: 2025 – 2026

Introduction & Background

In recent years, schools and colleges have started using data and technology to improve student learning. A major challenge is finding students who may underperform early so that teachers can help them on time. Traditional methods like attendance, grades, and assignments are useful but often fail to predict results accurately. Machine Learning (ML) can analyze multiple factors together, find hidden patterns, and predict student performance more effectively, enabling early support.

Relevance & Importance

An academic success prediction system can:

- **Give Early Alerts –** Identify weak students before exams so they can get extra help.
- **Support Data-Based Decisions –** Teachers can plan remedial classes or counseling based on actual data.
- **Save Resources –** Focus help only where needed instead of for all students equally.
- **Improve Results –** Over time, this can raise pass rates and reduce dropouts.

Literature Review / Related Work

Here are several recent studies relevant to your project that use similar datasets, machine learning models, and techniques. These give insight into what has been done, what works well, and possible considerations for your project.

Study (Year)	What They Did / Data Used	Key Insights for Our Project
Althaqafi et al. (2025) – Springer	Used academic, LMS, psychological data; handled class imbalance; tested Logistic Regression, Random Forest.	Class imbalance matters; Random Forest + ensemble models improve accuracy.
Ahmed et al. (2025) – Nature	Compared multiple ML models incl. ensemble methods on wide feature sets.	Ensemble models give better accuracy; preprocessing is very important.
Research (2024) – ACM	Used Random Forest to predict grades and analyze feature importance.	Shows which factors (attendance, assignments, participation) matter most.
Orji et al. (2022) – arXiv	Included motivation, study habits with academic data; compared RF & Logistic Regression.	Non-academic factors help; RF is accurate, LR is more interpretable.
Jimenez Martinez et al. (2024) – arXiv	Tested multiple ML models to detect at-risk students at different times in the semester.	Timing matters; early prediction gives teachers time to act.

Gaps / Issues Noted in Literature:

- Class Imbalance: Many studies report skewed distribution between high-performing and low-performing students; oversampling or methods like SMOTE help avoid bias.
- Feature Diversity: Adding non-academic features like motivation or participation improves prediction accuracy.
- Interpretability vs Accuracy: Random Forests give better accuracy; Logistic Regression offers better interpretability, useful for teachers.

Problem Statement

Academic underperformance among students is often identified only after examinations or end-of-semester evaluations, leaving minimal opportunities for corrective measures. Traditional methods of monitoring—such as manual observation of attendance or reviewing past grades—are reactive and limited in scope. Existing academic monitoring systems often fail to provide timely predictions or actionable insights to educators. This gap highlights the need for a proactive, data-driven solution that can predict students' academic outcomes early in the semester and enable targeted interventions for at-risk learners.

Objectives

1. To design and develop a machine learning-based system that predicts student academic success using key factors such as attendance, past grades, assignment scores, and participation.

2. To evaluate and compare predictive models (e.g., Random Forest, Logistic Regression) using performance metrics like accuracy, precision, recall, and F1-score.

3. To provide early alerts for at-risk students so that teachers can implement timely interventions and improve academic outcomes.

4. To identify the most influential academic factors affecting performance through feature selection and analysis.

5. To visualize prediction outcomes and trends using graphs for better interpretability by educators.

Scope of the Project

Focus: Predict student success using academic factors (attendance, grades, assignments, participation).

Included:

- Data collection & preprocessing.
- Feature selection & ML model development (Random Forest, Logistic Regression).
- Performance evaluation (accuracy, precision, recall, F1-score).
- Early alerts for at-risk students.
- Visualization of trends.

Excluded:

- Non-academic/sensitive factors (socio-economic, psychological).
- Real-time ERP/LMS integration.
- Personalized recommendations (future scope).

Beneficiaries:

- Students → Early alerts.
- Teachers → Insights for intervention.
- Institutions → Data-driven planning.

Proposed Methodology

1. **Requirement Analysis –** Define problem & academic parameters.

2. **Data Collection & Preprocessing –** Gather records, clean & normalize (Python, Pandas, NumPy).

3. **Feature Selection –** Identify key predictors & transform data.

4. **Model Development –** Train Random Forest & Logistic Regression (Scikit-learn).

5. **Testing & Evaluation –** Measure accuracy, precision, recall, F1-score, ROC.

6. **Prediction & Alerts –** Categorize students (high, moderate, low) + early alerts.

7. **Visualization & Reporting –** Graphs, insights, and report.

8. **Final Output –** Trained ML model + performance analysis + alert mechanism.

Expected Outcomes

- Accurate prediction of students' academic performance (grades or pass/fail) based on historical data such as past grades, attendance, behavior, and other relevant features.
- Identification of key factors or indicators influencing student performance through feature selection techniques.
- An ML model that can classify or predict student performance early, helping educators to take timely intervention measures.
- Improved decision-making for teachers and administrators by providing insights into student learning trends.
- Enhancement of educational strategies such as tiered instruction or personalized learning based on prediction results.
- Visualization of student performance data through charts and reports for easy interpretation by stakeholders.

Work Plan & Timeline

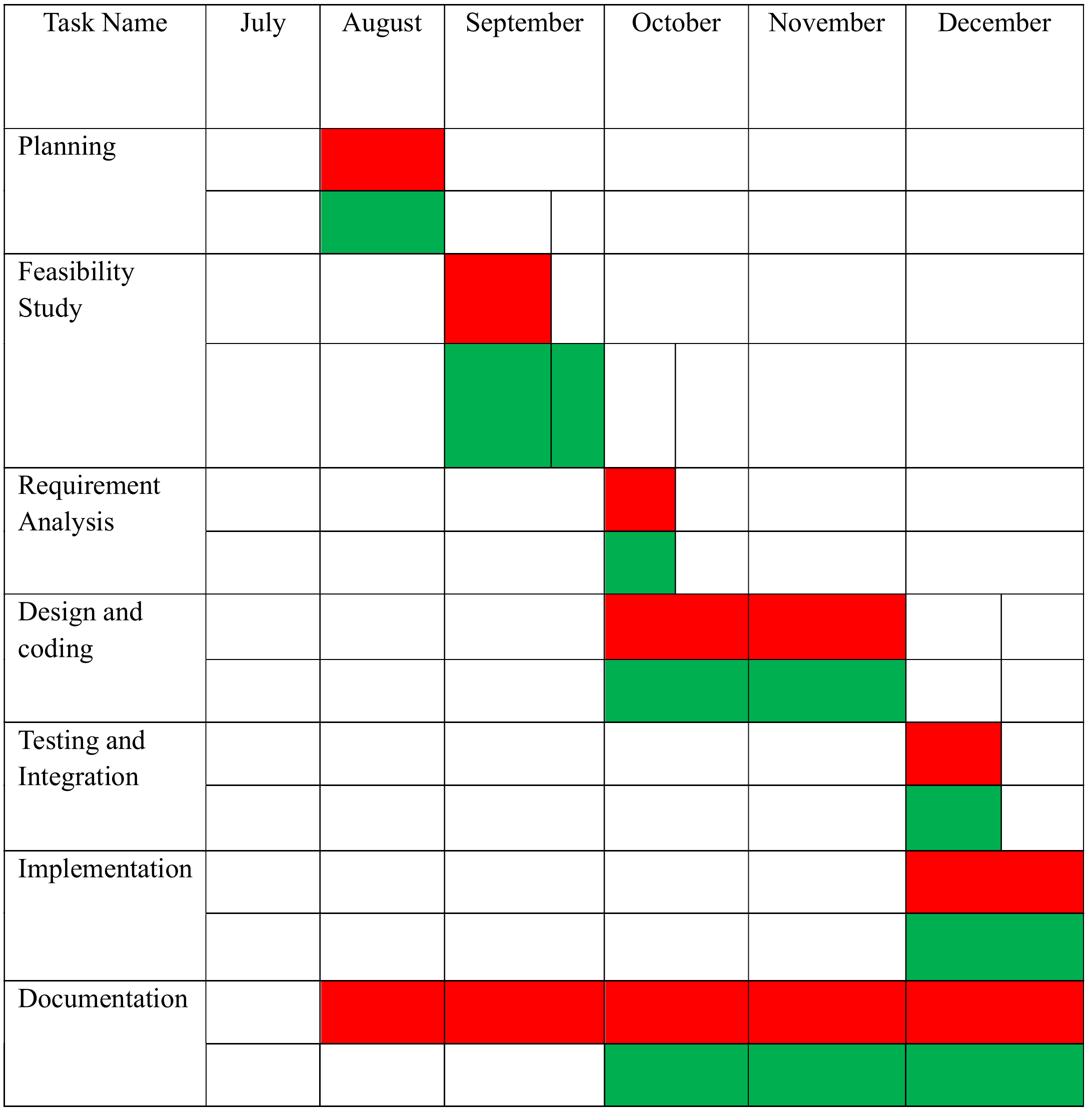

Fig : Gantt Chart

Resources Required

1. Software Resources:

- Programming Language: Python
- Libraries:
    - Scikit-learn (ML algorithms)
    - Pandas, NumPy (data preprocessing)
    - Matplotlib, Seaborn (visualization)
    - TensorFlow/Keras (if using deep learning)
- IDE/Tools: Jupyter Notebook / Google Colab

2. Hardware Resources:

- A computer with at least:
    - i3/i5 processor or higher
    - 8GB RAM (16GB recommended for deep learning)
    - 256GB+ storage

3. Dataset:

- **Public datasets**:
    - UCI Student Performance Dataset
    - Kaggle datasets (Student grades, attendance, socio-economic factors, etc.)

References

1. Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences*, *11*(9), 552.
2. Chen, Y., & Zhai, L. (2023). A comparative study on student performance prediction using machine learning. *Education and Information Technologies*, *28*(9), 12039-12057.
3. Balcioğlu, Y. S., & Artar, M. (2025). Predicting academic performance of students with machine learning. *Information Development*, *41*(3), 896-915.
4. Rizwan, S., Nee, C. K., & Garfan, S. (2025). Identifying the factors affecting student academic performance and engagement prediction in mooc using deep learning: A systematic literature review. *IEEE Access*.
5. Bisri, A., Heryatun, Y., & Navira, A. (2025). Educational Data Mining Model Using Support Vector Machine for Student Academic Performance Evaluation. *Journal of Education and Learning (EduLearn)*, *19*(1), 478-486.
6. Bisri, A., Heryatun, Y., & Navira, A. (2025). Educational Data Mining Model Using Support Vector Machine for Student Academic Performance Evaluation. *Journal of Education and Learning (EduLearn)*, *19*(1), 478-486.
7. Kordbagheri, A., Kordbagheri, M., Tayim, N., Fakhrou, A., & Davoudi, M. (2025). Using advanced machine learning algorithms to predict academic major completion: A cross-sectional study. *Computers in Biology and Medicine*, *184*, 109372.
8. Ha, D. T., Loan, P. T. T., Giap, C. N., & Huong, N. T. L. (2020). An empirical study for student academic performance prediction using machine learning techniques. *International Journal of Computer Science and Information Security (IJCSIS)*, *18*(3), 75-82.
9. Hasan, H. R., Rabby, A. S. A., Islam, M. T., & Hossain, S. A. (2019, July). Machine learning algorithm for student's performance prediction. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.
10. Kishor, K., Sharma, R., & Chhabra, M. (2021, September). Student performance prediction using technology of machine learning. In *International Conference on Micro-Electronics and Telecommunication Engineering* (pp. 541-551). Singapore: Springer Nature Singapore.