A
Progress report for Project

on

Student Marks Performance Prediction

for

Submitted in partial fulfillment for the Award of

MASTER OF COMPUTER APPLICATION

Submitted to:

Mr. Zahid Ahmed

Computer Science & Application

Submitted by:

ShivamGuljani(23CSA3BC084)

Department of Computer Science & Application

Vivekananda Global University, Jaipur

Year- 2025

Table of Contents

1 Project Title

Student Marks Performance Prediction

2 Project Overview

The "Student Marks Performance Prediction" system is designed to analyze various academic and non-academic features of students to estimate their future performance in terms of grades or marks. By analyzing data trends and using predictive algorithms, the system can identify students who may be at risk of underperforming, enabling educators and institutions to provide timely support.

The dataset used for this project was sourced from the Kaggle platform. It includes features such as attendance, study hours, previous exam scores, and more. A MongoDB database is used to store the data and results.

3 Objectives

- To collect and preprocess student academic data.
- To analyze key factors that influence academic performance.
- To develop and train predictive machine learning models.
- To integrate the model with MongoDB for data storage.
- To deploy the model through a web-based interface.
- To visualize results and provide insights to end users.

4 Work Completed

Task	Status	Description
Problem Definition	Completed	Defined project scope, goals, and measurable objectives.
Dataset Acquisition	Completed	Downloaded and prepared dataset from Kaggle.
Data Cleaning & Preprocessing	Completed	Handled missing values, normalization, and formatting.
Exploratory Data Analysis (EDA)	Completed	Identified trends, distributions, and patterns in data.
Feature Selection & Engineering	Completed	Selected impactful attributes for model input.
Model Development	Completed	Implemented algorithms like Linear Regression, Random Forest.
MongoDB Integration	Completed	Configured MongoDB to store and retrieve model input/output data.
GitHub Maintenance	Ongoing	Codebase is continuously updated and pushed to GitHub.
Documentation	In Progress	README, setup files, and reports being maintained.

5 Project Structure

```
Student-Marks-Prediction/
│
├── .git/                    # Git version control
├── logs/                    # Logging info from training & predictions
├── mlproject.egg-info/      # Python packaging info
├── Notebook/                # Jupyter notebooks (EDA, modeling, etc.)
├── src/                     # Core source code (data loaders, models, utils)
├── venv/                    # Python virtual environment
├── .gitignore               # Files/folders excluded from Git tracking
├── README                   # Project summary and usage instructions
├── requirements.txt         # List of dependencies
└── setup.py                 # Setup script for packaging the module
```

6 Technologies to be used

6.1 Software Platform

- **Programming Language:** Python, due to its extensive support for data science and machine learning.
- **Development Environment:** Jupyter Notebook, Google Colab, or VS Code for coding and testing.
- **Frameworks and Libraries:** Scikit-Learn for machine learning, Pandas and NumPy for data manipulation, Matplotlib and Seaborn for data visualization.
- **Database Management:** MongoDB for storing and retrieving student performance data.
- **Visualization Tools:** Streamlit for building an interactive user interface to display predictions and insights.

6.2 Hardware Platform

- **Local Machine Requirements:** Minimum 8GB RAM, Intel i5 Processor (or equivalent) for efficient computation.
- **Storage:** At least 50GB free disk space to accommodate the dataset and model files.
- **Network Requirements:** A stable internet connection to access the MongoDB database remotely.

6.3 Tools

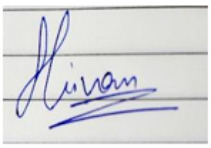
- Libraries: Pandas, NumPy, Scikit-Learn, Matplotlib, Seaborn
- Machine Learning Algorithms: Regression, Random Forest, Neural Networks
- Deployment Tools: Flask/Django for web application, Streamlit for visualization
- **Database:** MongoDB for storing student performance data and related records

7 Challenges Faced

- **Data Quality and Completeness:** The dataset had numerous missing values and inconsistencies. Various imputation and normalization techniques had to be applied to make the data model-ready.
- **Feature Selection Complexity:** Identifying the most influential features required extensive experimentation with statistical techniques and model-based selection methods.

- **Model Accuracy vs. Interpretability:** Balancing highly accurate models (like ensemble methods) with more interpretable ones (like linear regression) posed a trade-off challenge. Interpretability is key for educators to trust the system.
- **Database Integration:** Ensuring smooth communication between Python-based applications and MongoDB involved resolving issues with data serialization, schema management, and secure remote access.
- **Scalability Concerns:** Early system designs were not optimized for scalability, especially when handling larger datasets and real-time data ingestion. Refactoring the pipeline was necessary.
- **Collaboration & Version Control:** Managing multiple updates, dependency conflicts, and consistent documentation on GitHub required the use of strict branching, pull request reviews, and proper commit strategies.

8 Personal Details

Project Name	Course Name	Student ID	Student Name	Role	Signature
Student marks performance prediction	MCA (IVth SEM)	23CSA3BC084	Shivam Guljani	Developer	

9 Conclusion

The Student Marks Performance Prediction project has made significant progress, with 70% of the work completed. The system has successfully integrated machine learning with a backend MongoDB database. The predictive models are functional, and the results are promising. The next steps involve final optimization, UI development, deployment, and full documentation. Upon completion, the project will offer valuable insights to educational institutions, helping to improve learning outcomes for students.

10 References

- https://www.researchgate.net/publication/332893829_Predicting_Students'_Performance_Using_Machine_Learning_Techniques
- Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, Third Edition

- Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems
- https://www.w3schools.com/datascience/
- https://www.mongodb.com/docs/atlas/.
- https://www.kaggle.com/datasets/spscientist/students-performance-in-exams?datasetId=74977