

PROJECT PROPOSAL

Title: Data-Driven Course Demand Prediction and Resource Planning System

University: Vijaybhoomi University

Name: Laxmi Mahesh Parab

Registration Number: 2024SEPVUGP0028

1. Abstract

Universities often face challenges in academic planning due to changing student enrollment patterns. Inaccurate demand forecasting can lead to overcrowded classrooms, underutilized faculty, and inefficient use of institutional resources. This project proposes a machine learning-based Course Demand Prediction System that analyzes historical enrollment data along with course characteristics and faculty attributes to predict expected student enrollment.

The system will use three traditional machine learning models—Linear Regression, Decision Tree, and Random Forest—to generate accurate predictions. Based on these predictions, the system will provide data-driven recommendations for faculty allocation, classroom planning, and section scheduling. An interactive dashboard will be developed to visualize trends, compare model performance, and generate real-time predictions along with resource planning insights.

2. Problem Statement

University administrators must make critical academic planning decisions weeks in advance, including the number of course sections to offer, faculty assignments, classroom allocation, and resource budgeting. Currently, these decisions are often based on historical trends and manual estimation.

Such approaches fail to account for dynamic factors such as changing student preferences, faculty reputation, course difficulty, elective demand, and industry relevance. As a result, institutions frequently experience overcrowded classes, course cancellations due to low enrollment, uneven faculty workload distribution, and inefficient classroom utilization.

A data-driven system is therefore required to accurately forecast course demand and support effective academic resource planning.

3. Proposed Solution

The proposed system will consist of the following integrated components:

Data Pipeline

Synthetic data will be generated to simulate realistic university enrollment scenarios. The dataset will include features such as course category, semester, faculty rating (3.0–5.0),

previous enrollment, course difficulty (1–5), elective status, lab component, and industry demand score (1–10). The data will be stored in an SQLite database for structured and efficient management.

Model Development

Three regression models will be developed and compared:

- Linear Regression for baseline analysis
- Decision Tree Regressor for capturing non-linear relationships
- Random Forest Regressor for improved accuracy through ensemble learning

Model performance will be evaluated using R^2 Score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

Interactive Dashboard

A Streamlit-based web dashboard will provide:

- Data exploration and visualization
- Model performance comparison
- Interactive prediction interface
- Resource planning recommendations such as number of sections, faculty requirement, and classroom capacity

Prediction System

The system will support both single-course and batch predictions. Users can input course parameters and receive enrollment forecasts along with recommended resource allocation.

Model Lifecycle Management

The project will include model versioning, performance tracking, periodic retraining simulation, and model serialization using joblib.

Repository Structure

A well-organized GitHub repository will include data generation scripts, database integration, model training and prediction modules, dashboard code, and documentation. Large files will be excluded.

4. Expected Outputs

Technical Deliverables

- Trained machine learning models with target performance ($R^2 > 0.90$ and $MAE < 15$ students)
- SQLite database containing versioned datasets
- Interactive Streamlit dashboard
- Python scripts for data generation, training, and prediction

Analytical Deliverables

- Model comparison report
- Feature importance analysis
- Residual and prediction accuracy visualizations

Decision Support Outputs

- Course enrollment predictions
 - Resource planning recommendations (sections, faculty, classrooms)
 - Risk indicators for courses with unusually high or low demand
-

5. Project Timeline

Week 1: Repository setup, proposal preparation, and synthetic data generation (1000+ records)

Week 2: SQL integration, data preprocessing, and exploratory data analysis

Week 3: Model training, evaluation, and serialization

Week 4: Streamlit dashboard development and prediction interface

Week 5: Model retraining simulation, version tracking, and documentation

Week 6: System integration testing, repository cleanup, and demo preparation

Week 7: Final submission

Maintenance Plan

- Quarterly model retraining using updated data
 - Monthly dashboard performance monitoring
 - Semester-end accuracy evaluation against actual enrollments
 - Annual feature engineering review
-

6. Technology Stack

- **Programming:** Python 3.8+, Jupyter Notebook / Google Colab
 - **Machine Learning:** Scikit-learn, Pandas, NumPy, Joblib
 - **Database:** SQLite, SQLAlchemy
 - **Visualization & Dashboard:** Matplotlib, Seaborn, Streamlit
 - **Version Control:** Git, GitHub
-

7. Conclusion

This project demonstrates the practical application of traditional machine learning techniques to address real-world challenges in university operations. By developing a complete pipeline—from data generation and preprocessing to model deployment and visualization—the system provides actionable insights for academic planning. It also reflects best practices in model evaluation, version control, and lifecycle management, making it both academically relevant and practically useful.