📄 Predictive Modeling for Course Demand and Revenue Forecasting on EduPro

---

1️⃣ Abstract

- Online learning platforms require predictive intelligence for strategic planning.
- EduPro currently relies on historical reporting instead of forecasting models.
- The project develops machine learning models to predict:
    - Enrollment count per course
    - Course-level revenue
    - Category-level revenue
- Multi-dimensional datasets are integrated and engineered.
- Random Forest regression models are implemented.
- A Streamlit web dashboard is deployed for stakeholder use.
- The system transforms reactive planning into proactive decision-making.

2️⃣ Introduction

- Online education industry is highly competitive and dynamic.
- Course performance depends on:
    - Pricing strategy
    - Course level
    - Instructor quality
    - Market demand
- EduPro faces uncertainty in:
    - Launching new courses
    - Setting optimal pricing
    - Allocating instructors
- Lack of forecasting increases business risk.

- Predictive modeling provides data-driven planning capability.
- This project introduces demand forecasting for operational optimization.

3️⃣ Problem Statement

EduPro currently lacks:

- Predictive models for enrollment forecasting
- Revenue estimation at course and category level
- Quantitative evaluation of pricing sensitivity
- Automated instructor-course performance mapping

As a result:

- Course launches depend on intuition
- Pricing decisions lack data support
- Revenue estimation is reactive
- Resource allocation is inefficient

4️⃣ Project Objectives

- Develop predictive models for enrollment count.
- Develop predictive models for course revenue.
- Analyze category-level revenue trends.
- Identify key demand drivers.
- Build an interactive forecasting dashboard.
- Enable simulation of pricing and duration changes.
- Provide business insights for strategic planning.

5️⃣ Dataset Description

5.1 Courses Dataset

- CourseID
- CourseCategory

- CourseType (Technical / Non-Technical)
- CourseLevel (Beginner / Intermediate / Advanced)
- CoursePrice
- CourseDuration
- CourseRating

5.2 Teachers Dataset

- TeacherID
- Expertise
- YearsOfExperience
- TeacherRating

5.3 Transactions Dataset

- TransactionID
- CourseID
- TransactionDate
- Amount

6️⃣ Data Preprocessing

- Merged Courses and Transactions using CourseID.
- Aggregated transaction data to compute:
    - Enrollment_Count
    - Total_Revenue
- Handled missing values using:
    - Zero-fill strategy for sparse enrollments.
- Encoded categorical variables using:
    - Label Encoding for CourseCategory
    - Label Encoding for CourseLevel
    - Label Encoding for CourseType
- Ensured numerical features were standardized where necessary.

7️⃣ Feature Engineering

Course-Level Features

- Price bands (Low / Medium / High)
- Duration buckets
- Rating tiers
- Course level encoding

Instructor-Level Features

- Experience buckets
- Instructor rating score
- Expertise-category alignment

Historical Performance Features

- Past enrollment count
- Revenue per enrollment
- Average course revenue

8️⃣ Target Variables

The project predicts:

- Enrollment_Count (Regression Target)
- Total_Revenue (Regression Target)
- Category-Level Revenue (Aggregated Analysis)

9️⃣ Machine Learning Methodology

9.1 Model Selection

- Linear Regression (Baseline)
- Ridge Regression
- Random Forest Regressor
- Gradient Boosting Regressor

9.2 Final Model Used

- Random Forest Regressor selected for:
    - Handling nonlinear relationships
    - Robustness to overfitting
    - Better performance on structured tabular data

🔟 Model Training Process

- Split dataset using:
    - 80% training data
    - 20% testing data
- Trained separate models for:
    - Enrollment prediction
    - Revenue prediction
- Saved models using Joblib serialization.
- Stored encoders for production use.

1️⃣1️⃣ Model Evaluation Metrics

Mean Absolute Error (MAE)

- Measures average prediction error.
- Lower MAE indicates better accuracy.

Root Mean Square Error (RMSE)

- Penalizes larger prediction errors.
- Sensitive to outliers.

R^2 Score

- Indicates proportion of variance explained.
- Values closer to 1 indicate better model fit.

1️⃣2️⃣ Feature Importance Analysis

Major demand drivers identified:

- Course Price sensitivity
- Course Rating influence
- Course Level impact
- Course Duration effect
- Category-based demand variation

Insights:

- Lower-priced beginner courses show higher enrollment.
- Higher-rated courses generate greater revenue.
- Technical categories outperform non-technical categories in revenue.

1️⃣3️⃣ Streamlit Web Application

Core Functionalities

- Course performance overview table.
- Category-level revenue visualization.
- Enrollment prediction simulation.
- Revenue forecasting simulation.
- Interactive input controls for:
    - Price
    - Duration
    - Level
    - Course category

Technical Integration

- Loaded trained models (.pkl files).
- Loaded encoders for consistent prediction.
- Implemented fallback demo mode for stability.
- Deployed using Streamlit Community Cloud.

1️⃣4️⃣ Business Impact

The system enables:

- Data-driven course launch decisions.
- Revenue forecasting before investment.
- Optimized pricing strategies.
- Instructor allocation based on predicted demand.
- Reduced financial uncertainty.

Stakeholder benefits:

- Improved budget planning.
- Increased profitability.
- Strategic content roadmap development.

1️⃣5️⃣ Limitations

- Dataset size is limited (synthetic dataset).
- No external market demand indicators included.
- No time-series forecasting implemented.
- Revenue influenced primarily by price and enrollment only.

1️⃣6️⃣ Future Improvements

- Integrate real-time enrollment data.
- Add time-series forecasting (ARIMA / LSTM).
- Include student feedback sentiment analysis.
- Implement advanced boosting models (XGBoost).
- Deploy automated retraining pipeline.
- Add dashboard analytics for trend detection.

1️⃣7️⃣ Conclusion

- Predictive modeling enhances strategic planning for EduPro.

- Random Forest regression provides accurate enrollment and revenue forecasting.
- Feature importance analysis identifies key business drivers.
- Deployment via Streamlit enables practical stakeholder use.
- The project successfully transitions EduPro from reactive reporting to proactive forecasting.

1️⃣8️⃣ Keywords

- Predictive Modeling
- Enrollment Forecasting
- Revenue Prediction
- Machine Learning
- Random Forest
- Educational Analytics
- Business Intelligence
- Streamlit Deployment