Classification of Engineering College Fees in India using Random Forest

A Machine Learning Approach to Categorize Fees as High, Medium, or Low

By Anushka Rathour

Problem Statement

1. **Rising Education Costs**: The increasing cost of higher education makes it essential for students and parents to choose colleges wisely.

2. **Lack of Fee Categorization**: There is no systematic approach to categorize colleges based on their fee structures.

3. **Decision-Making Challenge**: Students and parents struggle to identify colleges that fit their financial situation.

4. **Objective**: Develop a machine learning model to classify college fees into high, medium, and low categories based on various features.

z

Unique Idea Brief

A machine learning-based approach for accurate fee category prediction.

Data Acquisition

Gathering comprehensive data on various engineering colleges

- Fee structures
- College rankings
- Facilities
- Total Student Enrollment

Model Training

Training a machine learning model using the collected data

- Supervised learning algorithms
- Feature engineering and selection
- Model evaluation and optimization

Prediction

Predicting the fee category based on user input

- State
- College Type
- Average Fees
- Established Year
- Gender Accepted

Help students and parents make informed decisions by providing categorized college lists.

Features Offered

The solution offers features to enhance transparency and accessibility.

Accurate Fee Classification:

Classify fees into high, medium, and low categories with high accuracy.

Detailed College Information:

Analyse college demographics, facilities, and academic offerings.

Categorized Lists:

Generate and present lists of colleges based on predicted fee categories.

Insightful Analysis:

Provide insights into the factors influencing college fees.

Processflow

The processflow outlines the steps involved in predicting fee categories.

1. **Data Collection**

   Gather data from the “engineering colleges in India.csv” by Kaggle (https://www.kaggle.com/datasets/shrirangmhalgi/engineering-colleges-in-india)

2. **Data Preprocessing**

   Clean and prepare the data, handling missing values and encoding categorical variables.

3. **Model Training & Evaluation**

   Training a machine learning model to predict fee categories(here we have trained model in Random Forest). Evaluate model performance using metrics like accuracy.

4. **Prediction**

   Using the trained model to predict fee categories based on user input into high , medium and low.

5. **Visualization**

   Displaying predictions in an easy-to-understand format

Architecture Diagram

The system architecture showcases the various components involved in the solution.

Data Storage

Storage for all relevant data

Prediction Engine

The core component responsible for predicting fee categories

User Interface

Provides a user-friendly interface for accessing the prediction system

Technologies Used

The project leverages various technologies to build a robust and efficient solution.

Language	Python
Machine Learning Library	Scikit-learn
Data Visualization Library	Matplotlib, Seaborn
Web Framework	Flask
Database	Pandas & NumPy

Conclusion and Recommendations

The fee category prediction system provides valuable insights to students.

1. **Effective Classification**

   The random forest classifier successfully categorizes college fee into high, medium and low categories.

2. **Useful Insights**

   The model provides valuable insights for prospective students and educational planners.

3. **Accessibility**

   Makes information readily available to students.

4. **Future Development**

   Integration with other relevant data sources for more accurate predictions