Assignment Project: Exploratory Data Analytics and Predictive Modeling

Course: Predictive Analytics
Total Marks: 10
Deadline: 29 November 2024

Objective:

The aim of this project is to showcase your understanding of Exploratory Data Analysis (EDA) and predictive modeling. You will work with a dataset of your choice, analyze it thoroughly, and apply predictive analytics to uncover patterns and insights.

Instructions:

1. Dataset Selection:
   - Choose a dataset relevant to any area of interest (e.g., finance, healthcare, retail, social media).
   - The dataset should have sufficient data points (at least 200 rows) and a mix of numerical and categorical features.

2. Exploratory Data Analysis (EDA):

   Conduct a thorough analysis of the dataset, including:
   - Summary statistics and data distributions.
   - Data cleaning (handling missing values, duplicates).
   - Feature engineering and transformation, if necessary.
   - Visualizations to reveal insights (e.g., histograms, scatter plots, correlations).

3. Predictive Modeling:
   - Choose an appropriate predictive modeling technique (e.g., linear regression, decision trees, or any classification method if applicable).
   - Split the dataset into training and testing sets (e.g., 80/20).
   - Train and evaluate your model, providing metrics such as accuracy, precision, recall, or RMSE (for regression models).
   - Interpret and document your model's performance.

4. Report and Submission:
   - Compile your findings, EDA insights, and model evaluation in a report.
   - Submit the code used in a Jupyter Notebook or a similar format alongside your report.

Report Template for Predictive Analytics Project

Cover Page

- Assignment Project: Exploratory Data Analytics and Predictive Modeling
- Course: Predictive Analytics
- Submitted By: [Your Name, Roll Number]
- Submitted To: [Instructor's Name]
- Submission Date: [Date]

[Table of Contents]

1. Introduction

Briefly introduce the objective of the assignment. State the significance of Exploratory Data Analysis (EDA) and predictive modeling.

2. Dataset Overview

- Dataset Source: Describe the source of the dataset (e.g., open-source, created, etc.).
- Dataset Description: Outline key details like the number of rows, columns, and general types of features.
- Objective: State what you aim to achieve by analyzing and modeling this data.

3. Data Preparation

- Data Cleaning: Describe any preprocessing steps taken (e.g., handling missing values, removing duplicates).
- Feature Engineering: Mention any new features created or existing features transformed.
- Summary Statistics: Provide a table with summary statistics (mean, median, mode, etc.) for each feature.

4. Exploratory Data Analysis

- Data Distributions: Include visualizations (histograms, bar charts) to show distributions.
- Correlation Analysis: Present correlation tables or heatmaps for feature

relationships.
- Key Insights: Summarize observations and findings from EDA, such as trends, outliers, or patterns.

5. Predictive Modeling

- Model Selection: Explain the model chosen and why it fits the data.
- Training and Testing: Describe the train-test split used (e.g., 80/20).
- Model Training: Document model parameters and settings during training.
- Model Evaluation: Provide evaluation metrics such as accuracy, RMSE, precision, recall, or others relevant to the model type.

6. Results and Model Evaluation

- Results Summary: Include key results, comparing model performance.
- Interpretation: Explain the relevance of model outcomes and how they answer the project objective.
- Visualization: Where applicable, provide charts or graphs showing model performance (e.g., confusion matrix, ROC curve).

7. Conclusion and Future Work

- Summary of Findings: Briefly summarize what the project accomplished.
- Limitations: Note any limitations or challenges faced in modeling.
- Future Improvements: Suggest improvements or additional analyses that could enhance the model.

8. References

List any sources used, including dataset sources, libraries, or tools referenced in the project.