**Project Title: Customer Subscription Prediction Using Logistic Regression and Decision Trees**

**1. Introduction:**

* Description: This project aims to predict whether a customer will subscribe to a service based on various demographic and behavioral features.
* Language and Libraries: Python 3, Jupyter Notebook, Pandas, NumPy, scikit-learn.

**2. Data Collection and Preprocessing:**

* Data Source: Dataset obtained from [source].
* Features: Features include age, gender, income, past subscription history, etc.
* Preprocessing: Data cleaning, handling missing values, encoding categorical variables using one-hot encoding.

**3. Exploratory Data Analysis (EDA):**

* Statistical Analysis: Descriptive statistics, correlation analysis.
* Visualization: Histograms, box plots, scatter plots to understand data distribution and relationships.
* Libraries Used: Pandas, Matplotlib, Seaborn.



**4. Modeling:**

* Logistic Regression:
  + Method: Binary classification model to predict subscription likelihood.
  + Libraries: scikit-learn.
  + Model Training: Fitting the logistic regression model to the training data.
  + Hyperparameter Tuning: Default hyperparameters used.
* Decision Trees:
  + Method: Classification algorithm to capture non-linear relationships.
  + Libraries: scikit-learn.
  + Model Training: Training a decision tree classifier with a max depth of 4.
  + Hyperparameter Tuning: Limited to adjusting max depth for simplicity.

**5. Results and Evaluation:**

* Evaluation Metric: Accuracy score on validation set.
* Interpretation: Comparison of accuracy scores between logistic regression and decision trees.
* Libraries Used: scikit-learn.

**6. Prediction on Test Dataset:**

* Preprocessing: Similar preprocessing steps applied to the test dataset.
* Prediction: Using the trained decision tree model to predict subscription status.
* Libraries Used: Pandas, scikit-learn.

**7. Submission File Creation:**

* Creating Submission File: Storing test predictions in a CSV file.
* Conversion: Converting predicted values (1/0) to 'yes'/'no' for better interpretation.
* Libraries Used: Pandas.

**8. Final Steps:**

* Instructions: Detailed steps for using the submission file to generate accuracy score.
* Additional Notes: Suggestions for model improvement and potential pitfalls.
* Language Used: Clear and concise instructions in English.

**9. Conclusion:**

* Summary: Recap of project objectives, methodologies, and outcomes.
* Future Work: Suggestions for further analysis or enhancements.
* Language Used: Reflective and forward-looking language.

**10. References:** - Credits: Acknowledgment of dataset sources, libraries, and code snippets. - Language Used: Proper citation format in English.

This detailed documentation provides a step-by-step explanation of the project, including the choice of language, libraries, and methodologies employed. It ensures clarity and reproducibility of the analysis while facilitating understanding for stakeholders.

Top of Form