Title: Customer Bank Deposit Prediction using Classification Analysis

1. Introduction: The project focuses on utilizing classification analysis and machine learning techniques to create a model that predicts whether a customer will make a deposit or not, based on their demographic and financial information. The dataset used for this project was obtained from Kaggle, a well-known platform for data science and machine learning competitions.

2. Data Collection and Preprocessing: Data collection involved obtaining the dataset from the Kaggle website, which contains information about customers, including demographic details, financial information, and whether they made a deposit or not. The dataset was preprocessed to handle missing values. Any null values present were removed during the Exploratory Data Analysis (EDA) phase.

3. Exploratory Data Analysis (EDA): EDA was conducted to gain insights into the dataset's structure, understand the distribution of features, and identify potential relationships between variables. Null values, if any, were addressed to ensure the dataset's integrity and completeness.

4. Model Selection: The project aimed to predict a binary outcome (whether a customer will make a deposit or not). Hence, classification models were considered for this task. The following algorithms were used:

Logistic Regression: A popular linear classification algorithm that models the relationship between the input features and the probability of the target class.

K Neighbors Classifier: A non-parametric classification algorithm that assigns a data point to the class most common among its k-nearest neighbors.

Decision Tree Classifier: A tree-based algorithm that recursively splits the data into subsets based on feature conditions to classify the target.

Support Vector Machine (SVM) Classifier: A powerful algorithm that finds the hyperplane that best separates the data points of different classes.

5. Model Training and Evaluation: The dataset was split into training and testing sets using the train-test split technique. The selected classification algorithms were trained on the training set and evaluated on the testing set. Evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix were used to assess the models' performance.

6. Model Selection and Hyperparameter Tuning: After evaluating the models, it was found that the Decision Tree Classifier provided the highest accuracy of 78.38%. As a result, the Decision Tree algorithm was chosen as the best-performing model for this particular problem.

To further optimize the Decision Tree model's performance, hyperparameter tuning was performed using a loop function. Hyperparameters are configuration settings that are not

learned during the training process, and they significantly impact the model's performance. Techniques like Grid Search or Random Search were used to find the best combination of hyperparameters that yielded the highest accuracy score.

7. Conclusion: In conclusion, the project successfully developed a predictive model using classification analysis to identify potential customers who are likely to make a deposit. The Decision Tree Classifier was chosen as the best-performing model, offering an accuracy of 78.38%. This model can be valuable for banks to target potential customers who are more likely to make deposits based on their demographic and financial information. By using this predictive model, banks can optimize their marketing strategies and improve their customer engagement and conversion rates.