**README**

This repository contains a dataset and Jupyter notebook related to a bank marketing project. The goal of the project is to predict whether a client will subscribe to a bank term deposit. The dataset used in this project is an enriched version of the "Bank Marketing" dataset from the UCI Machine Learning repository, including additional social and economic features/attributes.

**Dataset Information:**

The dataset consists of two files: "bank-additional-full.csv" and "bank-additional.csv". The "bank-additional-full.csv" file contains 41,188 instances, while the "bank-additional.csv" file is a smaller subset containing 10% of the instances. The dataset has 20 input attributes (features) and 1 output attribute (target variable).

**Attribute Information:**

Age (numeric)

Job: type of job (categorical)

Marital: marital status (categorical)

Education: level of education (categorical)

Default: has credit in default? (categorical)

Housing: has a housing loan? (categorical)

Loan: has a personal loan? (categorical)

Contact: contact communication type (categorical)

Month: last contact month of the year (categorical)

Day\_of\_week: last contact day of the week (categorical)

Duration: last contact duration in seconds (numeric)

Campaign: number of contacts performed during this campaign and for this client (numeric)

Pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means the client was not previously contacted)

Previous: number of contacts performed before this campaign and for this client (numeric)

Poutcome: outcome of the previous marketing campaign (categorical)

Emp.var.rate: employment variation rate - quarterly indicator (numeric)

Cons.price.idx: consumer price index - monthly indicator (numeric)

Cons.conf.idx: consumer confidence index - monthly indicator (numeric)

Euribor3m: euribor 3-month rate - daily indicator (numeric)

Nr.employed: number of employees - quarterly indicator (numeric)

Project Organization:

This repository is well-organized with a README file and a Jupyter notebook. The README file provides important information about the project, while the Jupyter notebook contains the code and analysis.

**Summary of Findings:**

After analyzing the bank marketing dataset and evaluating different classifiers, the following conclusions were drawn:

**Model Performance:**

K-Nearest Neighbors achieved an accuracy of 0.8674.

Logistic Regression achieved an accuracy of 0.8865.

Decision Trees achieved an accuracy of 0.8870.

Support Vector Machines achieved an accuracy of 0.8865.

Based on the accuracy scores, the best classifier for this project is Decision Trees.

Best Parameters:

For Decision Trees, the best parameters were 'max\_depth': 5 and 'min\_samples\_split': 2.

For K-Nearest Neighbors, the best parameters were 'n\_neighbors': 7 and 'weights': 'distance'.

For Logistic Regression, the best parameters were 'C': 0.1 and 'penalty': 'l2'.

**Conclusion**:

After evaluating the performance of different classifiers, Decision Trees outperformed the other models with an accuracy of 0.8870.

The best parameters for Decision Trees were 'max\_depth': 5 and 'min\_samples\_split': 2.

This indicates that the decision tree model with limited depth and minimum samples required to split is the most effective in predicting whether a client will subscribe to a bank term deposit based on the given attributes.