Capstone Project Submission

Team Member's Name, Email and Contribution:

Prabhat Patel (prabhatpatel51@gmail.com)

Contribution- Analyzed some of the attributes of the dataset provided a better visualization in a graphical manner which was very significant in model prediction. Performed oversampling techniques due to imbalance dataset and implemented Logistic Regression, Random Forest classifier, KNN and XGBoost classifier.

Annayan Bose (annayanbose@gmail.com)

Contribution- Observed some of the key factors and found out some important insights in the dataset. Did label encoding on the features which was quite significant and a task of feature selection among various features. Implemented various models such as Logistic Regression, DecisionTree, KNN, Random Forest, Naive Bayes classifier.

Please paste the GitHub Repo link.

Github Link:- https://github.com/prabhat7777/Bank_marketing_effectiveness.git

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

PROBLEM STATEMENT

The data was related to direct marketing campaigns (phone calls) of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to assess if the product (bank term deposit) would be subscribed(yes) or not. The classification goal was to predict if the client will subscribe to a term deposit (variable y).

The dataset consists of direct marketing campaigns data of a banking institution which consisted of 45211 data points with 17 independent variables out of which 7 were numeric features and 10 were categorical features.

Exploratory Data Analysis:-

We performed exploratory data analysis with to get insights from the data to observe following things:-

- The dataset was imbalanced, where the number of negative classes is close to 8 times the number of positive classes.
- The customers who had a job of admin had the highest rate of subscribing a term deposit, but they were also the highest when it comes to not subscribing. This is simply because we have more customers working as admin than any other profession.
- Majority of the customers were married. Followed by Single, divorced and unknown.
- Majority of the customers had a housing loan.

Models Implementation:-

- **Logistic Regression**
- **KNN**
- **Naive Bayes classifier**
- **Random Forest**
- **XGBoost**

Model Performance:

Model	Test AUC	Test Accuracy	F1_score	Precision
Logistic Regression	0.83	0.76	0.77	0.74
KNN	0.91	0.84	0.83	0.87
Naive bayes classifier	0.82	0.72	0.76	0.66
Random Forest	0.96	0.91	0.91	0.90
XGBoost	0.95	0.89	0.88	0.90

Among all models, **Random Forest** and **XGBoost** work the best and provide a reliable prediction.