**Project Title**:

**Bank Marketing Analysis and Deposit Prediction Using XGBoost**

**Description**:  
This project involves analyzing a bank marketing dataset to understand customer behavior and predict whether a customer will subscribe to a term deposit. The analysis includes exploratory data analysis (EDA) and visualization, followed by the application of the XGBoost machine learning algorithm for prediction.

**Executive Summary**

This report presents a comprehensive analysis of a bank marketing dataset obtained from Kaggle. The primary goal is to predict whether a customer will subscribe to a term deposit based on various attributes such as age, job, marital status, and more. The analysis is crucial for banks to tailor their marketing strategies and optimize their promotional efforts. The XGBoost algorithm, known for its performance in classification tasks, is employed for this prediction task. The model achieved a notable accuracy, demonstrating its effectiveness in solving this problem. This report details the steps taken from data preprocessing, visualization, and model training to evaluation, offering insights into customer behaviour and the predictive power of the model.

**Table of Contents**

1. **Title Page**
2. **Executive Summary**
3. **Introduction**
   * 3.1 Project Overview
   * 3.2 Objective
4. **Reason for Choosing the Subject**
5. **Data Source**
6. **Why XGBoost?**
7. **Methodology**
   * 7.1 Data Reading and Cleaning
   * 7.2 Exploratory Data Analysis (EDA) and Visualization
   * 7.3 Data Preprocessing for Model
   * 7.4 Model Training and Evaluation
8. **In-Depth Code and Insights**
   * 8.1 Data Loading and Initial Inspection
   * 8.2 Handling Missing Values and Duplicates
   * 8.3 Visualizations: Pie Charts, Count Plots, Histograms, and More
   * 8.4 Feature Engineering and Transformation
   * 8.5 Model Building with XGBoost
   * 8.6 Model Performance and Insights
9. **Description of Dataset**
10. **Conclusion**
11. **References**

**Introduction**

**3.1 Project Overview**

The project focuses on analyzing a bank's marketing dataset to predict customer behavior, specifically whether a customer will subscribe to a term deposit. Understanding customer behavior is essential for banks to improve their marketing strategies and drive growth.

**3.2 Objective**

The primary objective is to explore the dataset, understand key patterns through visualizations, and build a machine learning model using XGBoost to predict term deposit subscriptions.

**Reason for Choosing the Subject**

The banking sector heavily relies on targeted marketing strategies to reach potential customers. By analyzing customer data, banks can identify key segments likely to subscribe to specific products like term deposits. This analysis not only helps in improving marketing efficiency but also aids in customer retention and satisfaction. The XGBoost algorithm is chosen due to its high accuracy and performance in classification tasks, making it a suitable choice for this problem.

**Data Source**

The dataset used in this project is sourced from Kaggle, a popular platform for datasets and machine learning challenges. The "Bank Marketing" dataset contains information about customer demographics, bank attributes, and the outcome of previous marketing campaigns.

**Why XGBoost?**

XGBoost is a powerful machine learning algorithm known for its speed and accuracy in handling structured data. It is particularly effective for classification tasks, making it an ideal choice for predicting whether a customer will subscribe to a term deposit. Additionally, XGBoost handles missing data well, provides feature importance, and is highly customizable, which adds to its appeal.

**Methodology**

**7.1 Data Reading and Cleaning**

The dataset is first read into a Pandas DataFrame, followed by an initial inspection to understand its structure. Cleaning involves handling missing values, removing duplicates, and correcting data types.

**7.2 Exploratory Data Analysis (EDA) and Visualization**

EDA is conducted to explore the distribution of variables and relationships between them. Various plots such as pie charts, count plots, histograms, and heatmaps are used to gain insights into the data.

**7.3 Data Preprocessing for Model**

Categorical variables are encoded using Label Encoder to prepare the data for model training. The data is then split into training and testing sets.

**7.4 Model Training and Evaluation**

The XGBoost model is trained on the training set and evaluated on the test set using metrics like accuracy, confusion matrix, and classification report.

**Description of dataset**[**¶**](https://www.kaggle.com/code/kareemellithy/bank-market-deposit-prediction-xgboost/notebook#Description-of-dataset)

1-age: Represents the age of the individual.

2-job: Describes the occupation or job of the person.

3-marital: Indicates the marital status of the person (e.g., married, single, divorced).

4-education: Represents the educational level of the person (e.g., primary, secondary, tertiary).

5-default: Indicates whether the person has credit in default ('yes', 'no', or 'unknown').

6-housing: Shows whether the person has a housing loan ('yes', 'no', or 'unknown').

7-loan: Indicates whether the person has a personal loan ('yes', 'no', or 'unknown').

8-contact: Describes the method of communication used to contact the person (e.g., 'cellular', 'telephone').

9-day: Indicates the day of the week of the last contact.

10-month: Represents the month of the last contact.

11-Duration: Represents the duration of the last contact in seconds

12-campaign: Indicates the number of contacts made during this campaign.

13-pdays: Describes the number of days since the person was last contacted or -1 if they were not previously contacted.

14-previous: Represents the number of contacts made before this campaign.

15-poutcome: Indicates the outcome of the previous marketing campaign.

16-deposit: The target variable, indicating whether the person subscribed to a term deposit ('yes' or 'no').

**In-Depth Code and Insights:**

**Conclusion**

The project successfully demonstrates how exploratory data analysis and machine learning can be used to predict customer behavior in the banking sector. The use of XGBoost proved effective, achieving a commendable accuracy in predicting term deposit subscriptions. This analysis provides valuable insights for banks to target their marketing efforts more effectively.

**References**

* Kaggle: Bank Marketing Dataset (https://www.kaggle.com/datasets)
* XGBoost Documentation (<https://xgboost.readthedocs.io/>)