**Title: Enhancing Bank Marketing Strategies through Advanced Predictive Modelling**

<https://www.kaggle.com/datasets/aaditshukla/bank-marketing-dataset>

**Problem Description:** Traditional in-person product selling methods have proven to be time-consuming and costly. In contrast, modern telemarketing allows banks to reach a broader audience at a reduced expense. However, telemarketing's efficacy is contingent on the bank's ability to identify the most responsive customers. The bank marketing problem revolves around optimizing marketing campaigns for term deposits. By analyzing the Dataset sourced from Kaggle, containing extensive customer data and campaign outcomes, the objective is to develop predictive models to ascertain customers' likelihood of subscribing to term deposits. Leveraging Spark's distributed computing and advanced analytics, the project aims to unveil pivotal factors influencing customer decisions. The ultimate goal is to enhance the bank's marketing efficiency, fostering heightened customer engagement and increased subscription rates.

**Description of Dataset:** The Bank Marketing Dataset is a comprehensive dataset of 45,211 individual records, each representing distinct customer interactions with the bank's marketing endeavours. The dataset comprises 18 columns, including features such as age, job type, marital status, education, previous campaign outcomes, and term deposit subscriptions (the target variable). This dataset enables analysis and predictive model development through Spark's distributed computing capabilities.

**Goal:** The project's primary aim is to predict the factors behind customers' decisions to subscribe to term deposit. To achieve this aim, the following tasks have been established:

1. To Conduct data analysis to understand feature distributions patterns using spark
2. To Develop predictive models such as random forest, logistic regression, and gradient boosting, to forecast customers' likelihood of term deposit subscriptions.
3. To Create a data visualization dashboard using Tableau to facilitate monitoring.

**Project Plan:**

* **Data Preprocessing:** By Employ Spark's data manipulation techniques and processing capabilities
* **Predictive Modelling & Evaluation:** Implement machine learning algorithms and Evaluate model performance using metrics and optimize the models to enhance predictive accuracy.
* **Report and Presentation:** Summarize the project's findings in a detailed report and deliver a comprehensive presentation using Tableau.