# Project Overview

This project focuses on the analysis of data from direct marketing campaigns (phone calls) of a Portuguese banking institution, with the aim to predict whether clients will subscribe to a term deposit. The challenge is a classification problem with the target attribute being "Subscribed", which can have two values: "yes" or "no". The project involves data exploration, preprocessing, training various classification models, and evaluating these models on a test set.

# Data Analysis and Visualization

**Data Exploration**

* Initial exploration of the data involved examining distributions and counts of various features- Visualization techniques such as histograms and count plots were extensively used to understand the data.

**Correlation Analysis**

* Correlation matrices were generated to understand the relationships between different numerical features.

# Data Preprocessing

* Categorical features were identified and transformed using one-hot encoding.
* Rows containing 'unknown' values were removed to ensure data quality.
* High correlation pairs were identified for potential feature selection or dimensionality reduction.

# Model Training

Several machine learning models were trained to predict the target variable. These include:

* Decision Tree Classifier
* Logistic Regression
* Linear Regression
* Random Forest Classifier
* Gradient Boosting Classifier
* Gaussian Naive Bayes
* AdaBoost Classifier
* Support Vector Machine (SVM)

Each model was trained using the preprocessed dataset, and predictions were made on a test set.

# Model Performance Evaluation

* \*\*Decision Tree Classifier\*\*
* Accuracy: 14.22%
* \*\*Logistic Regression\*\*
* Accuracy: 87.96%
* \*\*Linear Regression\*\*
* Mean Absolute Error: 0.7917
* Mean Squared Error: 0.7376
* Root Mean Squared Error: 0.8589
* \*\*Random Forest Classifier\*\*
* Accuracy: 14.33%
* \*\*Gradient Boosting Classifier\*\*
* Accuracy: 14.31%
* \*\*Naive Bayes\*\*
* Accuracy: 20.12%
* \*\*AdaBoost Classifier\*\*
* Accuracy: 14.22%
* \*\*Support Vector Machine (SVM)\*\*
* Accuracy: 65.04%

# Conclusion

In conclusion, this project demonstrated the application of various machine learning techniques to a real-world problem in the banking sector. The diverse set of models used highlights the importance of experimenting with different approaches to find the most effective solution for a given

classification problem. While some models showed higher accuracy than others, each model provided unique insights into the dataset's characteristics.

# Team Member

Vladimir Gvaramadze

Uriel Vortia

Xinfu Guo