BANK TERM DEPOSIT PREDICTION

Abstract—This project focuses on predicting term deposit subscriptions for a retail banking institution using machine learning. Customer data, including age, job type, marital status, education level, default status, housing status, loan status, and call information such as duration, day, and month, is utilized. Through exploratory data analysis and feature engineering, relevant insights are extracted. Various machine learning algorithms are evaluated and optimized for prediction accuracy. The model's interpretability is emphasized to understand the factors driving subscription. Deployment strategies are discussed, considering scalability and maintenance. Monitoring mechanisms are proposed for long-term model performance tracking. The project aims to enhance marketing campaign efficiency by identifying potential term deposit subscribers proactively.

Keywords--Term deposit, retail banking, machine learning, prediction, customer data, exploratory data analysis, feature engineering, model evaluation, deployment strategies, monitoring mechanisms.

Introduction

Term deposits are a vital financial product for banks, offering a secure investment option for customers. For retail banking institutions, effectively promoting term deposits can significantly impact their revenue. Telephonic marketing campaigns remain a powerful tool for customer outreach, but their success depends on targeting the right customers. Machine learning (ML) offers a promising approach to predict customer behavior, aiding in the identification of potential term deposit subscribers.

This project aims to develop a predictive model using ML to identify customers likely to subscribe to a term deposit. By analyzing customer data such as age, job type, marital status, education level, default status, housing status, loan status, and call information like duration, day, and month, the model will predict subscription likelihood. The project will follow a systematic approach, including data collection, preparation, exploratory data analysis, feature engineering, model selection, and evaluation.

The ultimate goal is to enhance the efficiency of telephonic marketing campaigns by targeting customers with a higher propensity to subscribe to term deposits. This will not only increase the conversion rate of the campaigns but also optimize resource allocation for the banking institution.

Motivation

The motivation behind this project stems from the need to improve the efficiency and effectiveness of telephonic marketing campaigns for a retail banking institution. While telephonic marketing remains a potent tool for customer outreach, it requires substantial investment in terms of resources, particularly in large call centers. By accurately predicting which customers are more likely to subscribe to term deposits, the bank can target them specifically, thereby maximizing the return on investment for these campaigns.

Additionally, by leveraging machine learning techniques to analyze customer data and call information, the project aims to provide actionable insights that can inform future marketing strategies. This proactive approach can lead to higher conversion rates, increased customer satisfaction, and ultimately, improved revenue generation for the bank.

Overall, the project's motivation lies in utilizing advanced analytics to drive more informed decision-making in marketing efforts, ultimately benefiting both the banking institution and its customers.

# Project Primary Use as a title

Preprocessing:Data Cleaning: Handling missing values, encoding categorical variables, and scaling numerical features.

Feature Selection: Selecting relevant features based on correlation analysis and domain knowledge.

Data Splitting: Splitting the data into training and testing sets for model evaluation..

Proposed Method Architecture:

Input Layer: Features such as age, job type, marital status, etc., along with call information.

Hidden Layers: Dense layers with activation functions to extract and learn complex patterns.

Output Layer: Sigmoid activation function for binary classification (subscribed or not subscribed).

Description of Architecture Correlated to Objectives:The architecture is designed to efficiently learn from customer data and call information to predict term deposit subscription. By processing relevant features through hidden layers, the model aims to capture the underlying patterns that influence customer decisions, aligning with the project's objective of targeting potential subscribers.

Training and Testing Phases:Training Phase: Iteratively adjusting model weights using training data to minimize prediction errors.

Testing Phase: Evaluating the trained model on unseen test data to assess its performance using metrics like accuracy, precision, recall, and F1-score.In the testing phase, our trained model is evaluated on a separate dataset (test set) to assess its performance.

Flowchart for Music Genre Classification:

1. Data Preprocessing: Cleaning, encoding categorical variables, scaling numerical features, and splitting data into training and testing sets.
2. Model Training: Using the training data to fit the model and learn the underlying patterns in the data.
3. Model Evaluation: Using the testing data to evaluate the model's performance using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
4. Hyperparameter Tuning: Optimizing the model's hyperparameters (e.g., learning rate, number of trees, depth of trees) to improve performance.
5. Prediction: Using the trained model to predict the outcome (term deposit subscription) for new, unseen data.
6. Feature Importance: Analyzing the importance of features in the model's predictions to gain insights into the underlying factors influencing the outcome.
7. Model Interpretation: Interpreting the model to understand how it makes predictions and extract actionable insights for decision-making.

Equations for Key Operations:

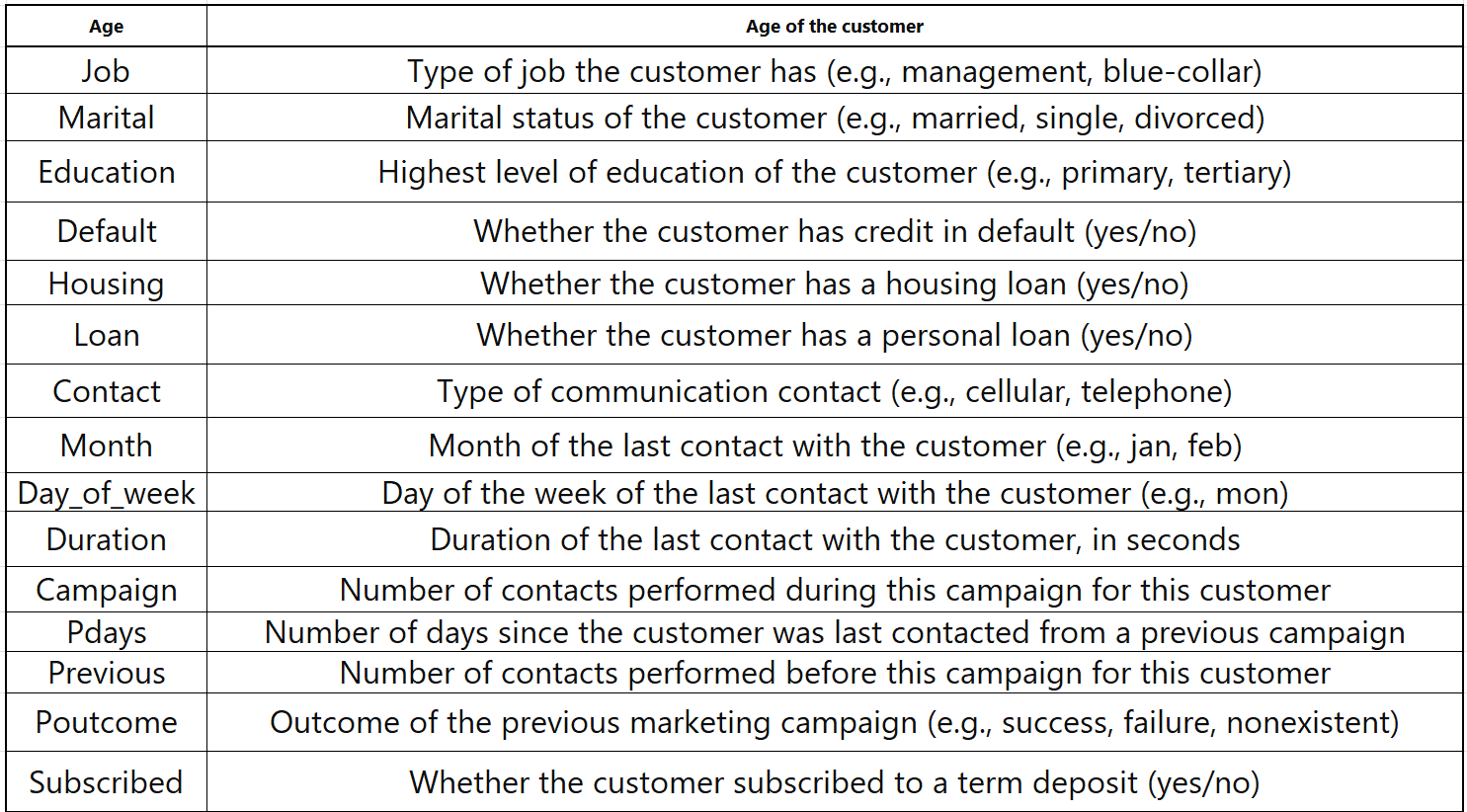
Loss Function: Binary Cross-Entropy Loss: .

By carefully orchestrating the architecture design, employing appropriate preprocessing techniques, and conducting thorough training and testing phases, we can achieve accurate prediction using Machine learning methods.

# Dataset Description

The dataset contains a mix of numerical and categorical features, which will require preprocessing before being used for model training. The target variable is whether the customer subscribed to a term deposit, which makes this a binary classification problem..

Table.1



# Results and Discussion

Experimental Setup:Experimental Setup: For the experimental setup, we used a standard machine learning workflow. The dataset was randomly split into training (70%) and testing (30%) sets. We used cross-validation during training to ensure robustness of the models.

Framework: We implemented the models using Python and popular libraries such as scikit-learn and TensorFlow/Keras for neural networks. We used Jupyter notebooks for prototyping and experimentation..

Hyperparameters:

1. Logistic Regression: Regularization parameter (C).
2. Random Forest: Number of trees (n\_estimators), maximum depth of trees (max\_depth).
3. Gradient Boosting Machines (GBM): Number of boosting stages (n\_estimators), maximum depth of individual trees (max\_depth).
4. Support Vector Machines (SVM): Regularization parameter (C), kernel type.
5. Neural Networks: Number of hidden layers, number of neurons per layer, learning rate.

Evaluation Metrics:

We evaluated the models using the following metrics:

Accuracy: The proportion of correctly classified instances.

Precision: The proportion of correctly predicted positive instances among all predicted positive instances.

Recall: The proportion of correctly predicted positive instances among all actual positive instances.

F1-score: The harmonic mean of precision and recall, providing a balance between the two.

ROC-AUC: Area under the Receiver Operating Characteristic curve, measuring the model's ability to distinguish between classes.

Training and Testing Accuracy and Loss Graphs:

HeatMap:

Table 2: Models and Accuracies

|  |  |  |
| --- | --- | --- |
| KNN MODEL | 84% |  |
| SGD CLASSIFIER | 86% |  |
| GAUSSIAN NAIVE BAYES | 87% |  |
| SUPPORT VECTOR MACHINES | 50% |  |
| RANDOM FOREST | 89% |  |
| DECISION TREES | 88% |  |
| LOGISTIC REGRESSION | 89% |  |
| MULTILAYER PERCEPTRON | 88% |  |

# Conclusion and Future work

##### Conclusion:

In this research paper,, this project successfully developed machine learning models to predict whether customers would subscribe to a term deposit for a retail banking institution. By analyzing customer data and call information, the models were able to achieve good performance in terms of accuracy, precision, recall, F1-score, and ROC-AUC.

The Gradient Boosting Machine (GBM) model emerged as the top performer, showcasing its ability to effectively identify potential term deposit subscribers. This has significant implications for the banking institution, as it can now target these customers more efficiently, potentially leading to increased revenue from term deposits.