**CUSTOMER CONVERSION PREDICTION**

**PROBLEM STATEMENT**

You are working for a new-age insurance company and employ multiple outreach plans to sell term insurance to your customers. Telephonic marketing campaigns still remain one of the most effective way to reach out to people however they incur a lot of cost. Hence, it is important to identify the customers that are most likely to convert beforehand so that they can be specifically targeted via call. We are given the historical marketing data of the insurance company and are required to build a ML model that will predict if a client will subscribe to the insurance.

**INTRODUCTION**

In today's competitive insurance sales environment, effective marketing resource allocation is critical. Telephonic marketing, while successful, is expensive. As a result, recognizing potential subscribers ahead of time can considerably improve the cost-effectiveness of marketing campaigns. This study describes the process of creating and assessing a predictive model to achieve this goal.

**DATA OVERVIEW**

The dataset consists of historical marketing data with the following features:

* **age**: Age of the client (numeric)
* **job**: Type of job (e.g., admin., blue-collar, entrepreneur)
* **marital**: Marital status (e.g., single, married, divorced)
* **educational\_qual**: Education level (e.g., primary, secondary, tertiary)
* **call\_type**: Type of contact communication (e.g., cellular, telephone)
* **day**: Day of the month when last contacted (numeric)
* **mon**: Month when last contacted (e.g., jan, feb, mar)
* **dur**: Duration of the last contact in seconds (numeric)
* **num\_calls**: Number of contacts performed during this campaign for this client (numeric)
* **prev\_outcome**: Outcome of the previous marketing campaign (e.g., unknown, other, failure, success)
* **y**: Target variable indicating whether the client subscribed to the insurance (binary)

**DATA EXPLORATION AND PREPROCESSING**

**1. Data Loading and Initial Inspection**

The dataset was loaded and examined for structure and completeness. I Identified the data types, checked for missing values, and ensured there were no anomalies.

**2. Data Cleaning**

* **Missing Values**: Handled through imputation or deletion based on the context and impact on model performance.
* **Outliers**: Detected and managed to prevent them from clipping.

**3. Exploratory Data Analysis (EDA)**

EDA revealed the following key insights:

* **Age Distribution**: Most clients fall within the 30-50 age range.
* **Job Types**: Distribution across different job types with a focus on certain high-concentration areas like blue-collar and administrative roles.
* **Marital Status**: Predominantly married individuals.
* **Education Level**: A mix of secondary and tertiary education levels.
* **Call Type and Timing**: Patterns in call duration and timing that correlate with subscription rates.
* **Previous Campaign Outcome**: Significant impact of previous campaign outcomes on the likelihood of subscription.

**4. Encoding and Scaling**

Categorical Features:

Transformed categorical features into numerical format using Label encoding.

Numerical Features:

Scaling is not performed since it is not mandatory for the models chosen.

**MODEL BUILDING**

**1. Model Selection**

Evaluated various models including:

* Logistic Regression
* Decision Trees
* Random Forests
* XGBoost

**2. Model Training**

Models were trained using a split of 80% training data and 20% testing data. Cross-validation was employed to ensure robust performance estimates.

**MODEL EVALUATION**

**Performance Metrics**

* **F1-Score**: Primary metric used to balance precision and recall.
* **Accuracy**: Provided a general sense of overall model performance.

The best model achieved an F1-Score of **0.89** is **XGBoost.**

#### Feature Importance

Analysed the importance of each feature in the model’s decision-making process:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Feature** | **Importance** | **Importance\_abs** |
| 7 | dur | 0.452242 | 0.452242 |
| 6 | mon | 0.175747 | 0.175747 |
| 9 | prev\_outcome | 0.119672 | 0.119672 |
| 5 | day | 0.083127 | 0.083127 |
| 0 | age | 0.072050 | 0.072050 |
| 1 | job | 0.033387 | 0.033387 |
| 8 | num\_calls | 0.020204 | 0.020204 |
| 3 | education\_qual | 0.016009 | 0.016009 |
| 4 | qual\_type | 0.015884 | 0.015884 |
| 2 | marital | 0.011679 | 0.011679 |

**RESULTS AND BUSINESS IMPACTS**

**Summary of Findings**

* Identified key factors influencing client subscriptions, with call duration and previous campaign outcomes being the most significant.
* Developed a model that effectively predicts potential subscribers, achieving an F1-Score of 0.89.

**Business Impact**

* The model’s insights can guide strategic adjustments in marketing efforts, focusing on high-impact areas.

### **CONCLUSION**

This project successfully developed a predictive model that can significantly enhance the efficiency of telephonic marketing campaigns for the insurance company. The model’s deployment is expected to lead to substantial cost savings and higher conversion rates. Continuous monitoring and refinement will ensure the model remains effective as market dynamics evolve.