Group 13 - Gogulkumar Mathi, Infant Angeline Rose Stephen, Raj Aswat Mahesh

IE: 7374 MACHINE LEARNING FINAL PROJECT ABSTRACT

PREDICTING BANK CAMPAIGN SUBSCRIPTIONS

INTRODUCTION

In order to enhance sales and maintain track of clients, business organizations have been researching online and direct database marketing as part of their marketing strategy in recent years. The older marketing options have contributed minimally to increasing the business of banks. They offered long-term deposits with good interest rates to the people using a direct marketing strategy but contacting many people takes a lot of time and the success rate is also less. Nowadays, banks invest a lot in marketing, and related research has emerged to become mainstream in academic literature. Corporate organizations hold data on customers and develop the capacity to analyze and make good use of such transactional data.

PROBLEM STATEMENT

Understanding topics such as the psychology, thinking, behavior, and motivation of consumers, as well as the approaches utilized in successful campaigns, may help businesses and organizations enhance their marketing strategies. The purpose of these campaigns is to persuade customers to join up for one of the bank's financial products (term deposit is taken into consideration here). Since customers hardly respond positively, data prediction models can help in selecting the most likely prospective customers also.

GOAL

The goal of the project is to boost confidence in the ability to accurately predict which customers and market segments would respond positively to a bank's marketing initiatives and to find out the characteristics that aid banks to make a customer subscribe to their campaigns successfully. It would also help in assisting the banks in selecting high-value customers with lower risks.

METHODOLOGIES

For the given problem statement, the client was supposed to notify the institution whether they intended to subscribe to the product (marking it as a successful campaign) or not after each call (marking it as an unsuccessful campaign) which helps with the evaluation of the campaigns.

In order to predict if the campaigns have been successful or not 3 classification machine learning algorithms are to be used: Logistic regression, Naive Bayes, and Support Vector Machine. The result shows which is the better classifier for the prediction in selling the deposit (success of campaign) in the case of three algorithms with a binary outcome. (yes/no)

DATASET

The data is related to direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, to assess if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

The dataset consists of 11640 records which represent the instances of calls to the clients and 21 features(including the target feature indicating if the marketing campaign was a success or not). The data collected is almost for 2 years(between the years 2008 and 2010).

The 21 features used in the project have been categorized into 5 types for better understanding and further analysis.

Client based features

age (numeric) ; job(categorical) ; marital(categorical) ; education (categorical) ; credit_in_default(categorical) ; housing(categorical) ; loan(categorical)

Previous contract based features

contact(categorical) ; month(categorical) ; day_of_week(categorical) ; duration(numeric)

Socio and Economic based features

employment_variation_rate(numeric) ; consumer_price_index(numeric) ; consumer_confidence_index(numeric) ; euribor_3_month_rate:(numeric) ; number_of_employees(numeric)

Other Features

campaign(numeric) ; pdays (numeric) ; previous(numeric); poutcome(categorical)

Output variable (desired target):

has_subscribed(binary)