

BankBytes

Problem Description

Data Glacier – Data Science Internship

LISUM33

Week – 10

Team Members:

Name: Mark Kimutai Kitur
Email: kiturmark@gmail.com
Country: Kenya
College: Dedan Kimathi University of Technology
Specialization: Data Science

Name: Yue Hu
Email: hy1550278246@gmail.com
Country: The United Kingdom
College: Durham University
Specialization: Data Science

Name: Rakshith Nagaraju
Email: rakshith.nagaraj6@gmail.com
Country: The United Kingdom
College: University of Liverpool
Specialization: Data Science

Business Understanding.

The aim of this project using the direct marketing campaigns of ABC Bank to sell their term deposit is to create a predictive model(s) in order to drive more sales for the term deposit by using targeted marketing channels. This process will save company resources and help indicate to the marketing team which demographics to target with the specific channels.

The goal is to derive insights from the marketing campaign data from previous programs and build predictive models to see what factors played the most important roles in whether customers bought the term deposit or not, once these factors are predictable, the marketing channels (telemarketing, sms, etc.) can be tailored to target those clients who are more likely to buy the term deposit, increasing the rate of return on the resources spent on marketing.

There is an output variable present in the list of features in the dataset which indicates whether the client bought the term deposit or not, this comes along with several features that are related to the client, the response of the client during previous marketing events, social and economic context variables and other campaign factors.

If the predictive model accurately predicts whether the client would be willing to buy the term deposit based on factors that we provide, then the project can be considered a success.

The primary stakeholders will be the executives of ABC Bank who want to increase revenue by way of selling more term deposits to a more selective audience who will respond to their marketing channels.

Predicting that a client is likely to buy the product but in actuality would not is acceptable, but predicting that a client is not going to buy the product but in actuality would buy it, is not acceptable, in more technical terms, false positives (Type I errors) are acceptable but false negatives (Type II errors) are not.

Framing the Problem Statement.

From initial analysis there seems to be very few clients contacted, if contacted at all. This is due to there being no structure in who to contact in the client list, this problem is exactly what the project is aiming to solve.

Increasing conversion rate by 50% from unsubscribed customers in 2 months to reduce the redundancy of spending resources on clients who are not willing to purchase the product and focusing more on the clients who are more likely to purchase or are already existing customers is the key focus of the project.

Breakdown.

There are two types of data present in the dataset, numerical and categorical data. There are no null values in the dataset and hence there is no need for filling NA values first.

The numerical data should be analysed for outliers and skewness, these outliers and skewness of each parameter should be taken care of by using methods such as log transformations or square root transformations, this is because all the data is right skewed. There are a few outliers in the which can also be taken care of with the use of these methods. The data can then be normalized and standardized in order to make it more efficient for use in model preparation and classification.

The categorical data in the dataset has no outlier values, i.e. there is no instance of wrong casing or extra categories. All the categories fall under the specified categories for that specific feature. But for model training, we cannot use raw categorical data, therefore we should prepare the data by means of methods such as one-hot encoding or label encoding, the data though categorical none of it is "Ordinal data", which means that the order in which the categories occur is irrelevant and can be encoded in any order.

Upon analysis of the economic features, 'pddays', 'empvrrate', 'nremployed' and, 'euroborn3m', they do not heavily influence 'y'. These features can be factored out during model training to improve the predictions of the model.

Once all the features have been transformed, data preparation will be complete, this data can now be processed through linear regression models and classification models in order to arrive at predictions on the marketing data in order to arrive at a business proposal for the marketing team to focus on, instead of having a vague approach or a broad approach to target every client at random.

Conclusion.

In conclusion, the data from previous marketing programs should be clean and then analysed in order to see how the data will help to provide insights.

This data is then put through the process of data transformations and can now be used to train models for regression and classification.

Once these models are trained on the data, it can be tested on how accurately it predicts whether a client would buy the product or not, and can also be used to classify which demographics are more likely to buy the product so that clients are targeted instead of a broad advertisement which does not allocate resources in a proper way.