# **Prediction of prospect customers response “YES” or “NO” to open a term deposit account of Bank Marketing Campaign.**

# **Using Classification and Regression Machine Learning tools**

# **In Python.**

**By: Marina Golberg 501072689**

[**marina.golberg@ryerson.ca**](mailto:marina.golberg@ryerson.ca)

**Ryerson University: The Chang School of Continuing Education**

**CIND820: Big Data Analytics Project**

**Date: 09/27/2021**

**Instructor: Dr. Tamer Abdou**

**Abstract:**

Businesses these days are using every available platform to do marketing of their company, products, and services. They use telemarketing, email marketing, advertisements, and many other different tools to gain sales and expand their business. Telemarketing is still the most used method of increasing sales of small to big companies. It is a cost-effective and flexible marketing strategy that offers a high level of personal contact.

Predictive analytics uses data models, statistics, and machine learning to predict future events.

It's a discipline that helps you to analyze your marketing campaigns, assess their efficiency, and see possible improvements to lead an increase in sales in future.

In this capstone project, the theme Classification and Regression. The goal of these classification models is to help the company more reliably predict future customer subscription before it occurs to secure deposits more effectively and increase customer satisfaction by reducing undesirable advertisements for certain customers.

This project would be able to answer following questions:

Predict prospect customer response “yes” or “no” to term deposit subscription?

What type of customers is more likely to subscribe to term deposit and which feature has higher influence?

What is the best time in a year for marketing campaign and ideal amount of calls to potential clients?

**Data Set Information:**

The data is related with 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, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed. The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y). Data can be found here: <https://archive.ics.uci.edu/ml/datasets/bank+marketing>

## **What Is a Term Deposit?**

A term deposit is a fixed-term investment that includes the [deposit of money](https://www.investopedia.com/terms/d/deposit.asp) into an account at a financial institution. Term deposit investments usually carry short-term maturities ranging from one month to a few years and will have varying levels of required minimum deposits.

The investor must understand when buying a term deposit that they can withdraw their funds only after the term ends. In some cases, the account holder may allow the investor early termination—or withdrawal—if they give several days notification. Also, there will be a penalty assessed for early termination.

<https://www.investopedia.com/terms/t/termdeposit.asp>

**The tools that will be used for this project are:**

Jupyter notebook for Python environment, Pandas to perform data manipulation and analysis, NumPy will be used to perform a wide variety of mathematical operations on arrays, Seaborn, Plotly and Matplotlib will be used for data visualization statistical graphing and plotting, Sklearn will be used for classification and regression these library contains a lot of efficient tools for machine learning and statistical modeling.

**The techniques that will be used to answer questions in this project are:**

Predict prospect customer response “yes” or “no” to term deposit subscription?

Data preprocessing with sklearn to encode object variables to numerical labels. SMOTE oversampling as data is imbalance, I will Normalize Data so the algorithms perform better, Classifiers are Logistic Regression, Decision Tree Classifier, Random Forest, k-nearest neighbors, Neural Network. Then I will compere the scores of the algorithms to choose the best one.

What type of customers is more likely to subscribe to term deposit and which feature has higher influence? What is the best time in a year for marketing campaign and ideal amount of calls to potential clients?

Feature selection technique to check the contribution of each attribute in explaining of client's subscription, correlation to see the relationship between the attributes, exploratory data analysis to visually see the relationship between variables and the dependent variable y.

,

**References:**

1. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. StefanLessmanna BartBaesensbc Hsin-VonnSeowd Lyn C.Thomasc
2. [Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014
3. Brownlee, J (2020) SMOTE for Imbalanced Classification with Python (Last accessed at: https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/ on 12th April, 2021)
4. 6Chawla, N et al (2002) SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16 (2002) 321–357
5. <https://www.investopedia.com/terms/t/termdeposit.asp>
6. <https://archive.ics.uci.edu/ml/datasets/bank+marketing>