

# MA429 Summative Project

May 2020

## 1 Introduction

This report will focus on analysing data of Portuguese direct marketing campaigns of banking institutions. The data related to the success of selling long term bank deposits through phone campaigns [2]. The data provided by the UC Irvine Machine Learning repository includes input variables relating to personal information, banking information, and contact information of potential clients. It includes 41,188 observations and 21 variables.

The analysis of banking data is extremely important for many reasons. This data can be used to promote equity, reduce fraud, and build a more resilient economy. The data that we are analysing is Portuguese bank marketing including communication, banking, and demographic data. The analysis of such data though important comes with many pitfalls, demographic data has been purposefully used in conjunction with banking data to deny access to capital to certain disadvantaged groups such as in North American redlining of the mid 20th century [1]. Even if there is no malicious intent these models can have a significant impact on who gets access and who does not and therefore it is important to understand biases and shortcomings of any model that is built as well as imperfections in the data itself.

## 2 Methodology

### 2.1 Overview

### 2.2 Data Set

### 2.3 Pre-processing

## 3 Conclusions

## References

- [1] Richard Harris and Doris Forrester. The Suburban Origins of Redlining: A Canadian Case Study, 1935-54. *Urban Studies*, 40(13):2661–2686, December 2003.
- [2] S. Moro, R. Laureano, and P. Cortez. *A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems*. Elsevier, 2014.