**Classifying Direct Marketing Campaigns using Naive Bayes**

Name: Amos Lipsys

Student Number: 13017795

Lecturer: Ickjai Lee

Date: 3rd June 2017

Subject: CP3403 Data Mining

**ABSTRACT**

The main objective was to create a classification model of a Portuguese banking institution direct marketing scheme. This model could then be used to increase effectiveness of future marketing campaigns. Data mining techniques were used to achieve this goal. First the data had to be pre-processed into a more useable format. Numerical data was summarised into categorical data, variables with a majority of missing data were excluded. Naïve Bayes classification method was then used to classify the data to create the model. This model was then used to analyse test data. The resulting model had an accuracy of around 60%. This is thought to be due to the classifier not including Laplace Smoothing. This smoothing would have increased the models accuracy to over 80%. This has a large impact on the model created. This new model would allow for direct marking to be far more effective by contacting only those with a high success chance.

**Introduction**

Data mining also known as Knowledge Discovery in Data (KDD) is the process of finding relevant unknown patterns within data. [1] Data mining is separate however from standard statistical analysis being more than just quantifying data. The algorithms in data mining, allow for patterns and relationships to be formed. Models are created which allow for accurate predictions to be made based on the information available to the model. [2]

There are a few main categories of data mining all with their own advantages and disadvantages. Data mining is exploration of data and because of this there is never one solution that will fit all situations perfectly. The data miner’s job is to use the tools available to create a model. If one solution proves to be inadequate another tool should be used instead. [3] Classification, clustering and association rules mining are three main data mining processes.

Classification is the processes of creating models to predict future inputs. This may be analysing the characteristics of a person to determine how risky it would be to offer them a loan or determining what things are contained in an image to give the illusion of computer vision. [4] .

1 <https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm#CHDFGCIJ>

2 <https://www.simplilearn.com/data-mining-vs-statistics-article>

3 <https://www.kdnuggets.com/2015/06/nine-laws-data-mining-part-1.html/2>

4 <https://www.tutorialspoint.com/data_mining/dm_classification_prediction.htm>

* + Introduction (1 page)
    - Brief background
  + Description on Dataset (1-2 pages)
    - Details about the dataset including how many instances, attributes etc
  + Preprocessing
    - Details about the preprocessing done for the dataset including cleaning, transformation etc
  + DM Area and DM Algorithm
    - Brief introduction into DM area & algorithm of your choice and provide justifications for the choice
  + Scenarios (Exploration of the Effect of k in k-means clustering)
    - Explanation on what you are going to do in your DM project
  + Results and Analysis
  + Conclusion
  + Reference