

# MSiA 401: Mini Project - Group 6

Alejandra Lelo de Larrea Ibarra

Kiran Jyothi Sheena

Lixuan Chen

Wencheng Zhang

Abstract

HERE GOES THE ABSTRACT

## Contents

1	Introduction	1
2	Data	2
2.1	Original Data . . . . .	2
2.2	EDA . . . . .	2
2.3	Data Cleaning & Feature Engineering . . . . .	2
2.4	Final Data Set Description . . . . .	2
3	Modelling	2
3.1	Train, Test & Cross Validation . . . . .	2
3.2	Naive Bayes (benchmark) . . . . .	2
3.3	Logistic Regression . . . . .	2
3.4	Neural Networks . . . . .	2
3.5	Tree . . . . .	2
3.6	Nearest Neighbors . . . . .	2
3.7	Local Logistic Regression . . . . .	2
3.8	Weighted Kernel . . . . .	2
3.9	Boosting . . . . .	2
3.10	Random Forest . . . . .	2
3.11	Support Vector Machines . . . . .	2
4	Model Comparison	2
5	Final Considerations	2
	References	2

## 1 Introduction

This is a way to cite papers (Wirth and Hipp 2000).

## 2 Data

### 2.1 Original Data

### 2.2 EDA

### 2.3 Data Cleaning & Feature Engineering

### 2.4 Final Data Set Description

## 3 Modelling

### 3.1 Train, Test & Cross Validation

### 3.2 Naive Bayes (benchmark)

### 3.3 Logistic Regression

### 3.4 Neural Networks

### 3.5 Tree

### 3.6 Nearest Neighbors

### 3.7 Local Logistic Regression

### 3.8 Weighted Kernel

### 3.9 Boosting

### 3.10 Random Forest

### 3.11 Support Vector Machines

## 4 Model Comparison

## 5 Final Considerations

## References

Wirth, Rüdiger, and Jochen Hipp. 2000. "CRISP-DM: Towards a Standard Process Model for Data Mining." In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 1:29–39. Manchester.