**GROUP REPORT**

Table

Description automatically generated

# INTRODUCTION

## PROJECT OBJECTIVE (0.5 MARKS)

## DATA DESCRIPTION (0.5 MARKS)

## SUMMARY OF APPROACH AND METHODS (1 MARKS)

# METHODS

CHALLENGES (1 MARK)

Imbalanced Data Set

Small Dataset

High Dimensionality

Multi-Collinearity

## SELECTION OF APPROACHES (3 MARKS)

WHAT IS THE GENERAL IDEA?

* To solve the four(or more) challenges present with this dataset
* Select appropriate methods for the problem
* Idea is to conduct sampling to solve imbalance issue, then feature selection to reduce complexity, then apply appropriate machine learning model which is relatively explainable

### 1- RESAMPLING METHODS

HOW IT SOLVES CHALLENGE 1?

#### NEAR MISS V2

#### SMOTE

#### SMOTEEN

#### SMOTETOMEK

### 2 - FEATURE SELECTION

HOW IT SOLVES CHALLENGE 2/3/4

#### FILTER METHODS

#### EMBEDDED METHODS

#### WRAPPER METHODS

### 3- CLASSIFIERS

WHAT CLASSIFIERS WERE CHOSEN INITIALLY AND WHY?

#### DECISION TREES

#### LOGISTIC REGRESSION

#### SUPPORT VECTOR MACHINES

## MACHINE LEARNING APPROACHES (4 MARKS)

## DECISION TREES

JESSE APPROACH

### EXPERIMENTAL RESULTS

DAVID APPROACH

### EXPERIMENTAL RESULTS

## LOGISTIC REGRESSION

ANDREW APPROACH

### EXPERIMENTAL RESULTS

DENNIS APPROACH

### EXPERIMENTAL RESULTS

## SUPPORT VECTOR MACHINES

ANDY APPROACH

### EXPERIMENTAL RESULTS

EDWARD APPROACH

### EXPERIMENTAL RESULTS

# DISCUSSION (8 MARKS)

## COMPARISON OF MODELS (3 MARKS)

BIG SUMMARY TABLE OF RESULTS

## LIMITATIONS OF MODELS (2 MARKS)

BIG SUMMARY TABLE OF PROS AND CONS

## MODEL EXPLAINABILITY (3 MARKS)

DISCUSSION ON HOW EXPLAINABLE EACH MODEL IS

# CONCLUSION (2 MARKS)

## SUMMARY OF RESEARCH

## FUTURE IMPROVEMENTS

# REFERENCES

# APPENDIX