**FIT 3162/ FIT COMPUTER SCIENCE PROJECT**

**User Guides**

**Project title: Improving Software Testing using Software Fault Prediction Methods when data is highly imbalanced**

**Team:**

FIT3162\_MA\_4

**Team Members:**

* Tah Wen Zhong
* Jason Toh Zhern Wee
* Ethan Hor Sheng Jian

**Project Supervisor:**

Golnoush Abaei

# Table of contents

[**Table of contents**](#_3vwv6sxrunw8) **2**

[**End User Guide**](#_az8lllhhti9h) **3**

[1. Introduction](#_lxn2reoprvhi) 3

[2. Launching the program](#_d3rot8y2le0) 3

[3. Login screen](#_ezwtdxbriqv5) 3

[4. Home screen](#_am0xoqcb0cr5) 4

[5. Main algorithm](#_k19zdg6fp14x) 5

[5.1 Program settings](#_kt4w80cgi3b7) 5

[5.1.1 Prediction settings](#_b1aszhhn81s6) 6

[5.1.2 Training settings](#_9iuhs2arxugc) 7

[5.2 Feature selection menu](#_9ki8636f4e66) 8

[5.3 Result](#_kow8lt4xnzg1) 9

[5.3.1 Table view](#_aaauj7xtx2jm) 9

[5.3.2 Chart view](#_hta0m0yuy2ty) 10

[5.3.3 CSV results](#_doo8irqymewv) 11

[6. Reports](#_wnuy9s3jwjoh) 12

[7. About us](#_k1ppo39nuy30) 14

[8. Instructions](#_lpnl17keigi4) 15

[9. Step-by-step example of how the program is run](#_qp61nlvsw4re) 16

[9.1 Upload a dataset](#_cgoaiy9ukv2w) 16

[9.2 Configuring settings](#_6aabc9m2x96b) 17

[9.2.1 Selecting the prediction models](#_s7uslbyx5riw) 17

[9.2.2 Training Settings](#_p8w2a8903amc) 18

[9.2.3 Selecting the feature selection method](#_kxb275z96rg3) 19

[9.3 Viewing the Results](#_tqz4woux2rb) 20

[9.3.1 Table View](#_s9ko2m21mgbr) 20

[9.3.2 Chart View](#_e1m8swvo47zt) 21

[9.3.3 CSV Results](#_tnw5fcz0pljx) 23

[9.4 Going back to the homepage](#_yij7dxukyt58) 24

[10 Exiting the application](#_inxprevnf4cs) 25

[**Technical Guide**](#_funi5ev3h4ov) **26**

[1. Software downloads](#_pto5hnsy2q5y) 26

[1.1 Python](#_owht1jhc2az4) 26

[1.2 Python packages](#_2v0tkby8nbs6) 26

[2. Installation](#_7ruc9n8le7d9) 27

[2.1 Zip](#_52z2yqu3apr3) 27

[2.2 NSIS](#_naqmpw69p9ml) 27

# End User Guide

## 1. Introduction

When it comes to developing software these days, testing proves to be a challenge for most parties. This is because of the amount of budget that needs to be allocated towards testing. To tackle the expensive budget, we have developed a program that is capable of predicting sections in a software which may be more fault-prone, and therefore allocate the testing budget accordingly.

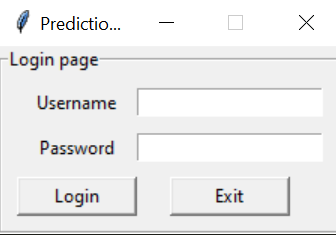
This document is a user manual which aims to guide the user on how to properly use this program.

## 2. Launching the program

In the program folder, select and run the file name “main.psw”. This will launch our prediction

software and you will be presented with the login page.

## 3. Login screen



**Fig. 1 Screenshot of login screen**

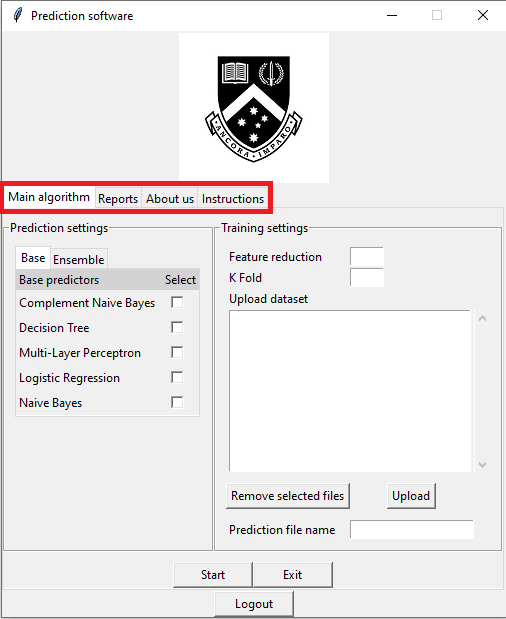
The user will have to enter the correct credentials in order to proceed to the homepage of the program. An error message will be displayed if either the username and password provided is incorrect or invalid. An exit button is also shown to allow the user to safely close the application.

As we do not have a live database to manage accounts, here is a set of valid credentials you can use to login to the program:

**Username:** FIT3162

**Password:** FYProx

## 4. Home screen



**Fig. 2 Home page screenshot with highlights**

Once the user successfully logs in to the program, the user will be authorized to enter the home page which contains the main items within our program. Here there are several sections in our program which the user can view using the tabs highlighted in Figure 2. Each section will be further elaborated in the remainder of the guide. If the user wants to log out from the program, there is a “Logout” button located at the bottom of every page.

# 

**Fig. 3 Screenshot of Logout warning message box**

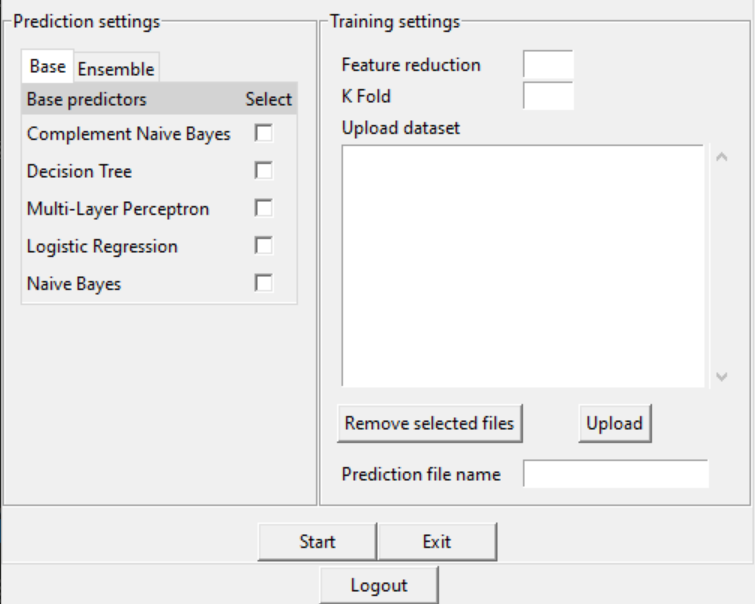
A warning message will be displayed when the user clicks the logout button. This is to prevent an accidental logout by the user that may halt the main algorithm. So the purpose of this warning box is for confirmation on the user’s action.

# 

## 5. Main algorithm

This section contains information regarding the interaction between the end users and the main program. The following sections will serve as a detailed guide on how to effectively utilize the prediction software to the absolute max potential.

### 5.1 Program settings



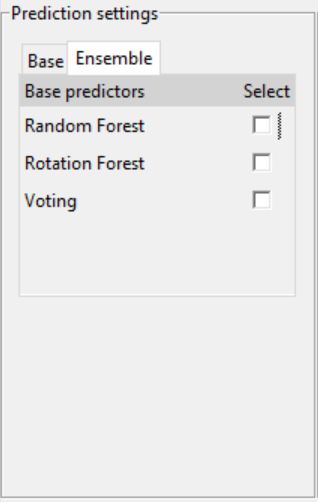
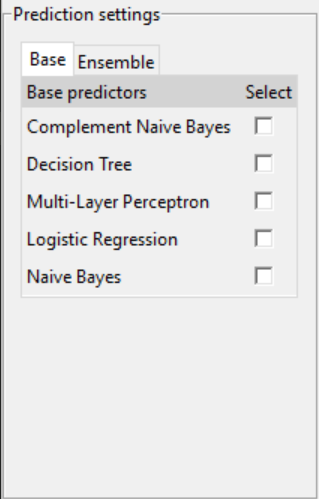
**Fig. 4 Screenshot of Program Settings**

The program settings menu contains both the prediction settings and training settings. Further information regarding these two settings can be found in the following section. Three buttons will be situated at the bottom.

When a user clicks on the Start button, it will transition to the feature selection menu.

Clicking on the Exit button will safely close the application whereas clicking on the Logout button will transition to the login screen. A confirmation message will be displayed if a user clicks on the Logout button to reaffirm the user’s choice.

#### 5.1.1 Prediction settings



**Fig. 5 Base Predictors Settings Fig. 6 Ensemble Predictors Settings**

The prediction settings menu consists of two sections. One for Base predictors and one for ensemble predictors. The base predictors have 5 options as seen in Figure 5 and the ensemble predictor has 3 options as seen in Figure 6. The user will be able to select multiple from either base, ensemble or both. If a user has not selected at least one predictor and proceeds to click on the ‘Start’ button, an error message will be displayed which will prompt the user to select at least one predictor.

#### 5.1.2 Training settings

## 

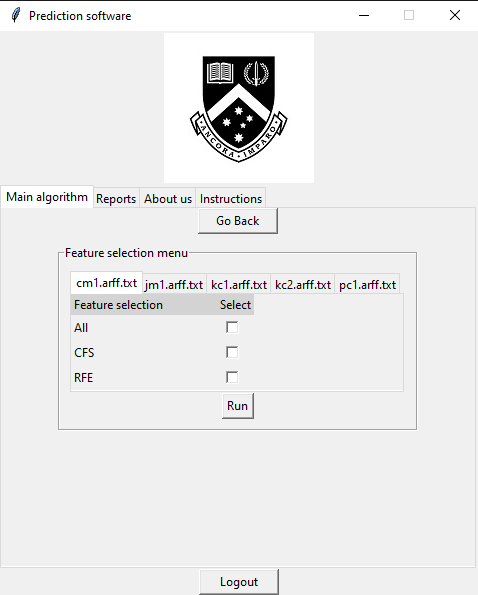
**Fig. 7 Screenshot of Training settings**

The training settings menu allows users to manipulate various parameters for training the prediction models. Below is a description of each section.

* **Feature Reduction**
  + Number of features to be reduced from the dataset
  + Minimum: 1
  + Maximum: 20
* **K Fold**
  + Number of folds to split the dataset into training and testing dataset
  + Minimum: 2
  + Maximum: 20
* **Upload Dataset**
  + Area to upload datasets, able to upload more than one dataset
  + Dataset must be written in arff format
  + Minimum: 1
* **Prediction File Name**
  + File name of the output csv file
  + Cannot be empty

For each section, there are criterias that need to be fulfilled which would throw an error message otherwise.

### 5.2 Feature selection menu



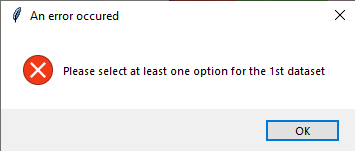
**Fig. 8 Image of the feature selection menu**

Once the user sets the program settings, the feature selection menu is presented as shown in Figure 8. This interface allows users to select their preferred feature selection methods for each dataset. The feature selection methods are used to reduce the metrics used to fit the models. The number of metrics to be reduced to is set by the user on the training settings screen which is shown in section 4.1.2. For flexibility, the program provides a number of tabs to allow different selections for each dataset. Hence, the number of tabs is based on the number of datasets uploaded by the user.

Each tab contains a checkbox which has a fixed set of options for the user to select. The checkbox allows multiple selection so the user can select more than one choice. Here are brief descriptions of each option available:

* **All:** All metrics within the dataset will be used
* **CFS:** Metrics will be chosen using correlation based feature selection
* **RFE:** Metric will be chosen using recursive feature elimination

Once the user is done with the selection, they can click the “Run” button to execute the program’s main algorithm. If the user wants to return to the previous step then the “Go back” button can be used instead.



**Fig. 9 Image of the error message for missing feature selection**

The user is required to select at least one option for each dataset, so an error message will display if any dataset has no options selected. As shown in Figure 9, the message will inform the user which dataset needs selection.

### 5.3 Result

Once the main algorithm finishes its computation, the results will be obtained and displayed to the user. From the interface, the result data can be viewed in tabular and chart form. Furthermore, a csv file is produced which also contains the results. If the user wants to reuse the algorithm, the “Go Back” can be found at the top of the program.

#### 5.3.1 Table view

**Fig. 10 Screenshot of Table view tab**

The table view is one of two viewing methods which display the results in a tabular form. The results displayed are the evaluation score of each model for every dataset. The columns represent the model used whereas the rows represent the uploaded dataset. Since the user can select multiple feature selections, the method which each data underwent is appended to the name of the dataset. Figure 10 shows a sample of how the table view will look like.

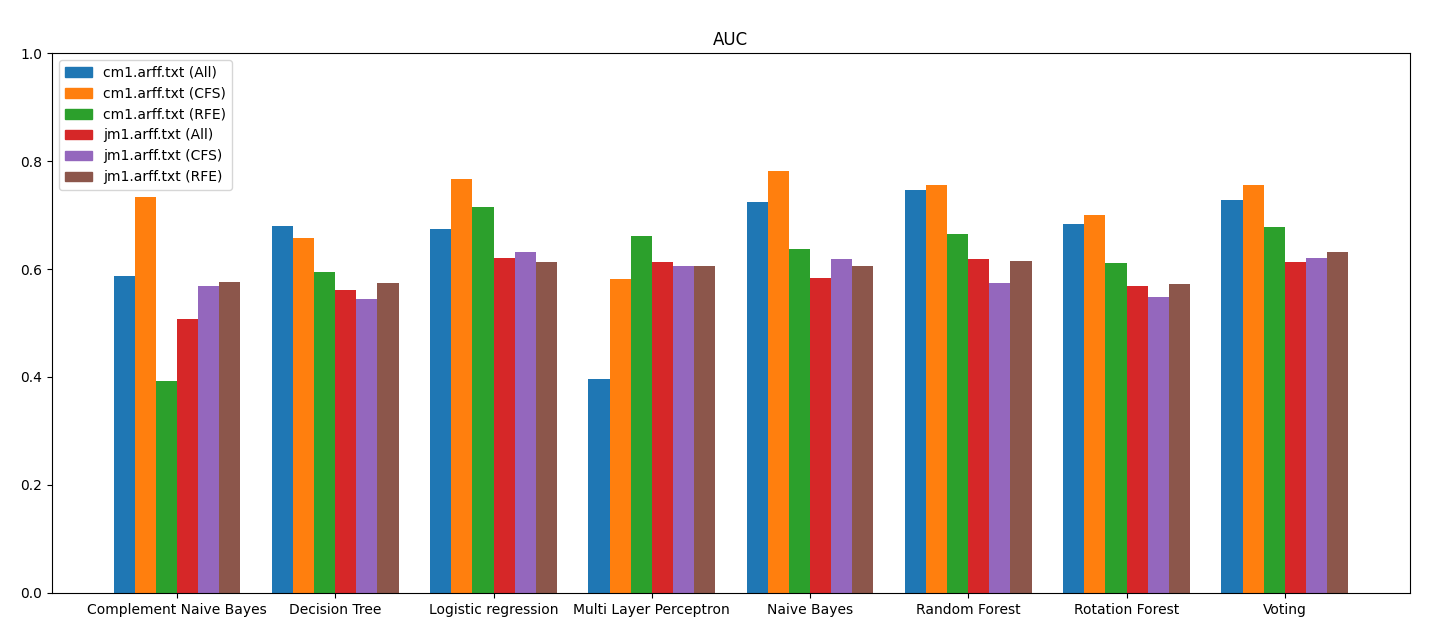
### 

#### 5.3.2 Chart view

### 

**Fig. 11 Screenshot of Chart view tab**

The chart view is the second viewing method which allows the user to view the evaluation results in chart forms. There are 4 charts available for each evaluation score, the user can click on the “Show chart” button next to any of the evaluation metrics to show a bar graph formed based on the results.

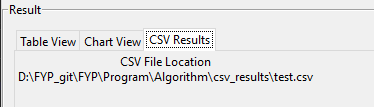


**Fig. 12 Sample chart for AUC results**

Figure 12 shows an example of how the charts would look like, unlike the table view the results for each evaluation metrics are separated to ease viewer’s interpretation.

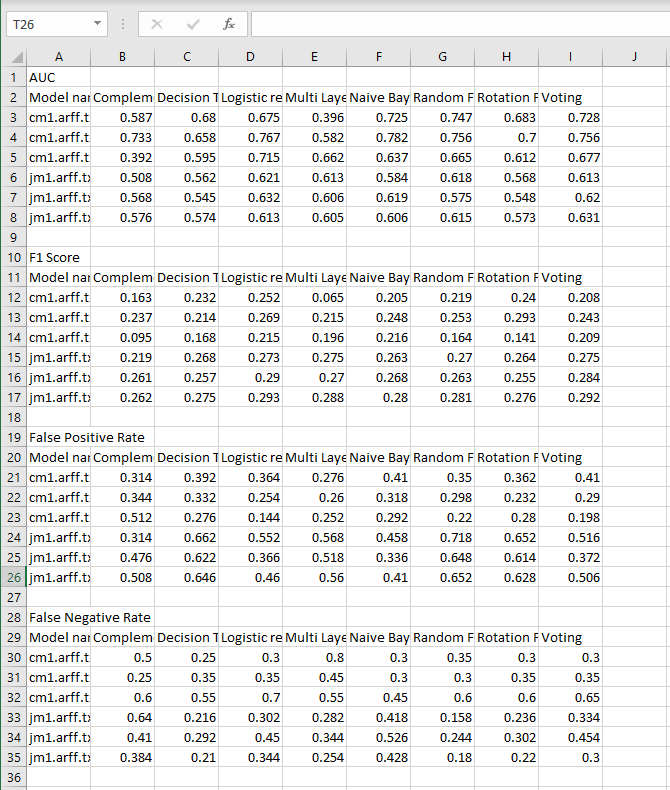
### 

#### 5.3.3 CSV results



**Fig. 13 CSV Results tab**

As mentioned previously, the algorithm will produce a csv file which contains the result of the program. The name of this csv file is set by the user at the training settings 4.1.2. The result will always be added in a folder within the program called csv\_result. As shown in Figure 13 the CSV Results tab will inform the location of this file from the root directory.

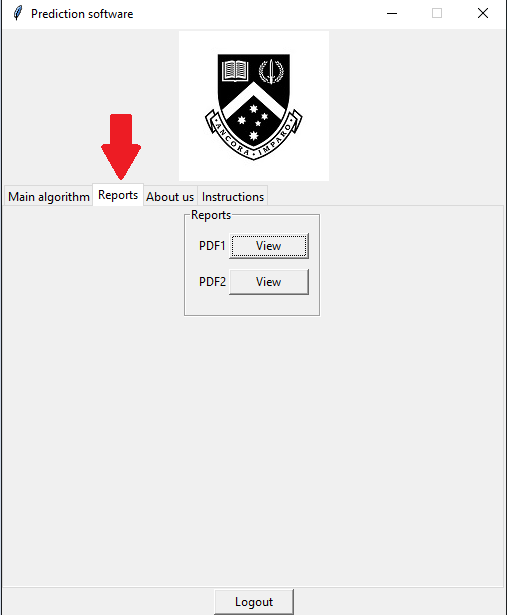


**Fig. 14 Sample of the csv file**

Figure 14 shows a sample of how the data is stored in the csv file. The csv will store the data in a similar format as the table view where each evaluation score is separated with rows and columns representing the models and datasets.

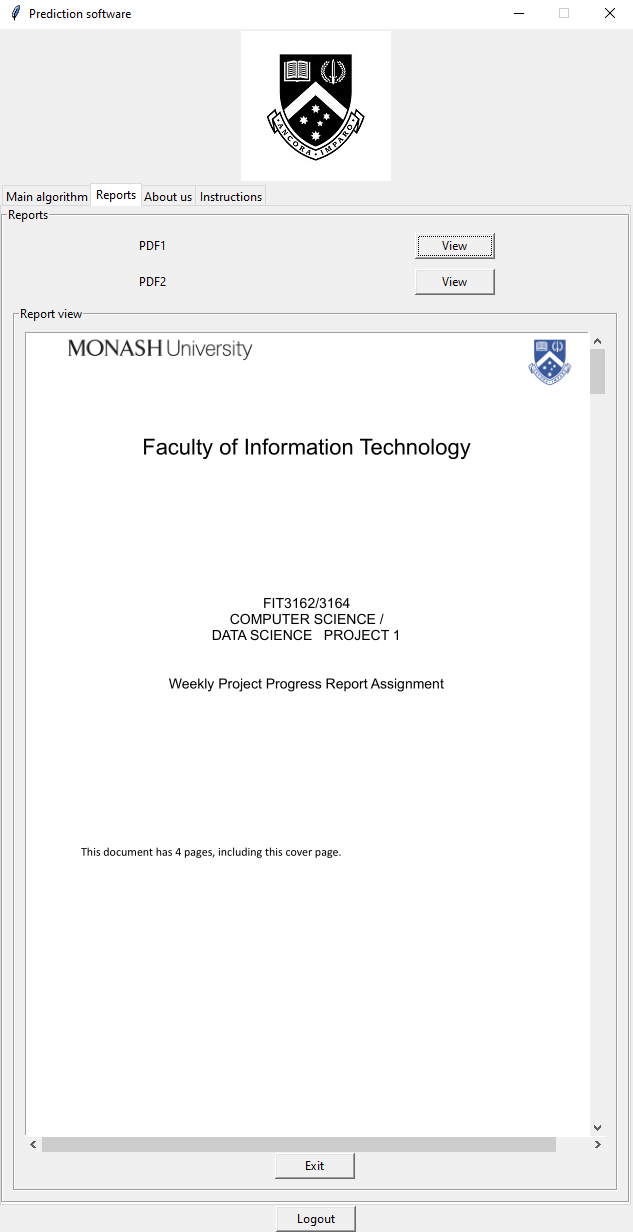
# 

## 6. Reports



**Fig. 15 Switching to the “Reports” section in the homepage**

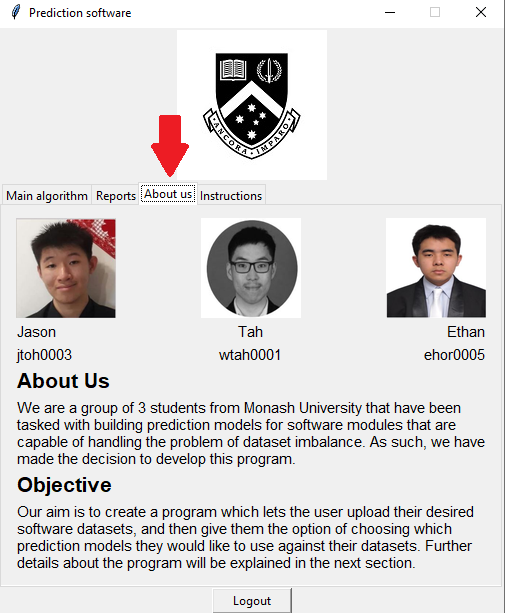
To access the Reports section, simply click the “Reports” tab after reaching the homepage. Then, you will be taken to a screen which looks similar to the one in Figure 15.



**Fig. 16 Viewing a report in the Reports section**

In this section, you can click on the View button for any of the reports you wish to view, and it will pop up just as shown in Figure 16. When you are done looking at the report, click the exit button to close the report.

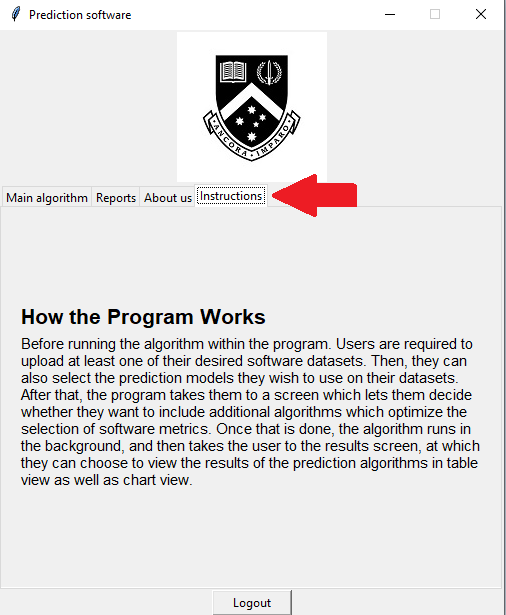
## 7. About us



**Fig. 17 Switching to the “About Us” section**

The “About Us” section of the program is the section which contains the profile of our team. To access it, simply click the “About us” tab after reaching the homepage. Then, you will be taken to a screen which looks similar to the one in Figure 17.

## 8. Instructions

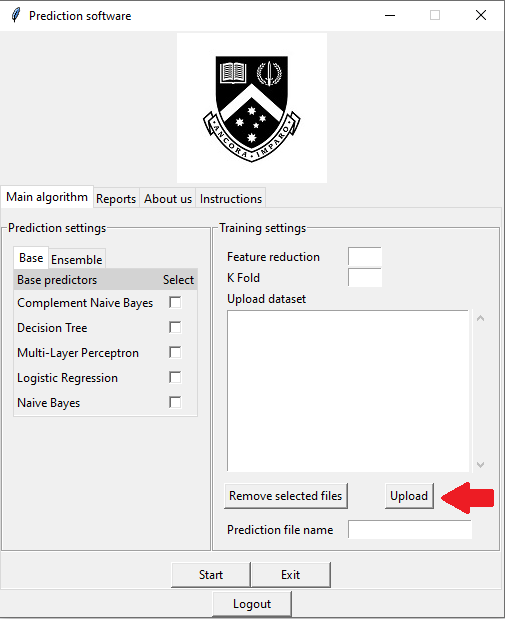


**Fig. 18 Switching to the “Instructions” section**

The “Instructions” section contains a briefer guide on how to use the program. Despite it being there, it is recommended to go through the user manual before using the program. To access it, simply click the “Instructions” tab after reaching the homepage. Then, you will be taken to a screen which looks similar to the one in Figure 18.

## 9. Step-by-step example of how the program is run

### 9.1 Upload a dataset

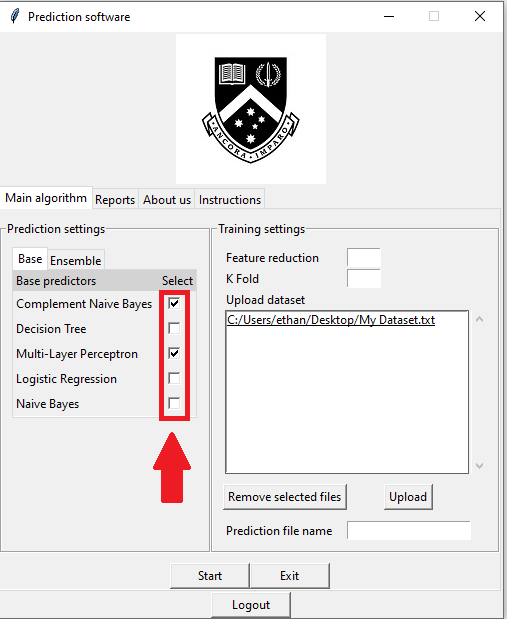


**Fig. 20 Uploading a dataset**

After logging in, click on the Upload button. It should open up a file directory for you. From there, select a dataset which is in the file format of .arff or .txt.

### 9.2 Configuring settings

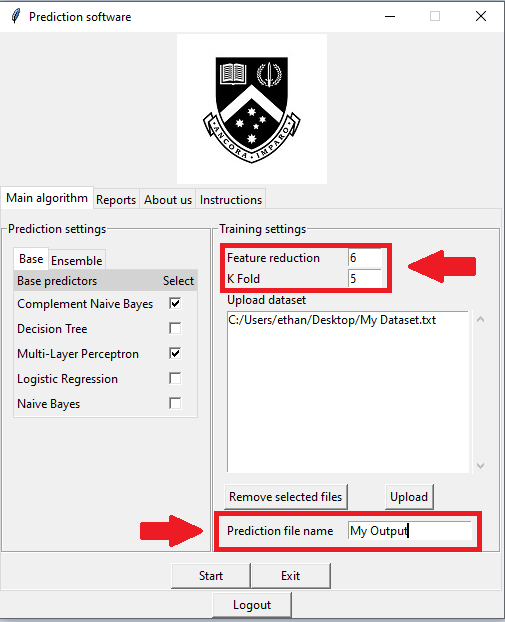
#### 9.2.1 Selecting the prediction models



**Fig. 21 Selecting the models**

Once the dataset is selected, it will be shown in the field on the right side. After this, we should now select the prediction models we want to use. Selected prediction models are indicated with a tick in the checkbox. For this example, we will be selecting Complement Naive Bayes and Multi-Layer Perceptron.

#### 9.2.2 Training Settings



**Fig. 22 Configuring the Training settings**

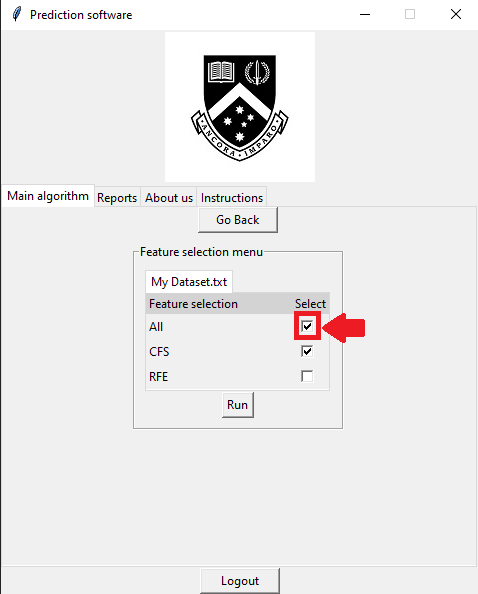
Once the models have been selected, we now want to configure the Training Settings. We will be inputting the values 6 for feature reduction, and 5 for K-fold. After doing that, we will be naming the output CSV file “My Output”. We are now ready to run the program, so the start button will be clicked on.

## 

## 

## 

#### 9.2.3 Selecting the feature selection method

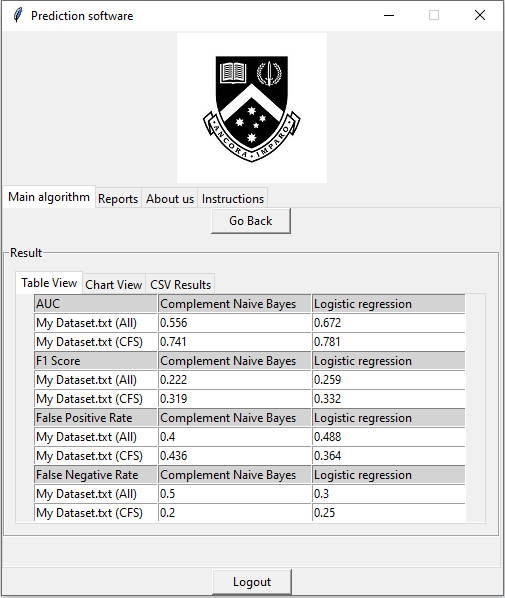


**Fig. 23 Configuring the feature selection settings**

After clicking the start button, the user will be brought to the feature selection menu. For this example, we will only be using All and CFS so click on the checkboxes. Selected methods are indicated with a tick in their checkbox.

### 9.3 Viewing the Results

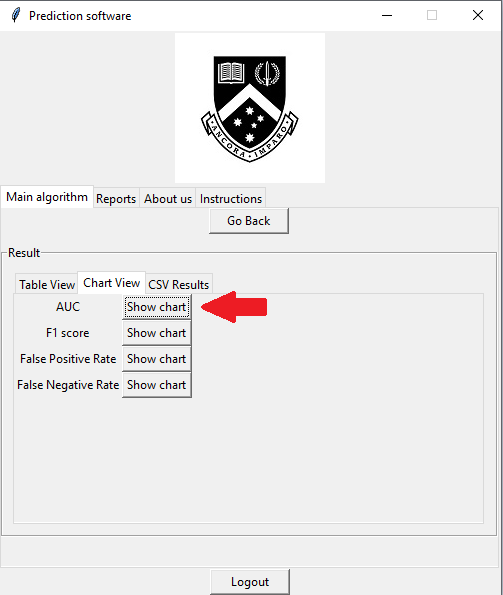
#### 9.3.1 Table View



**Fig. 24 Table View of the results**

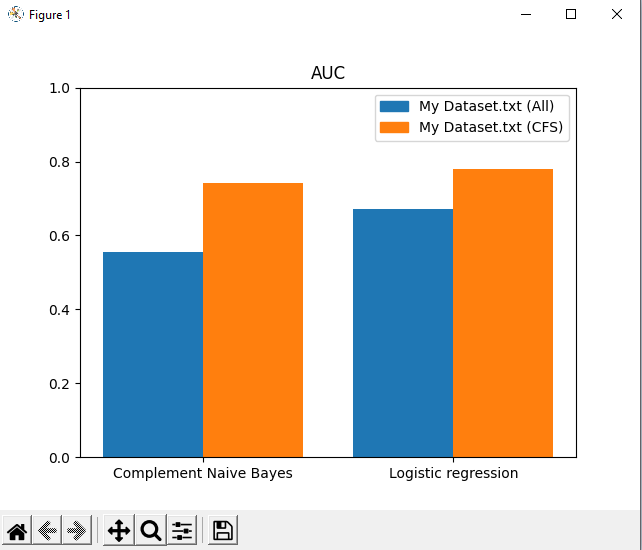
After the program finishes running, the user is taken to the results screen. This is where we can view the results through multiple methods. The main method of how the results are displayed is in a table view, as shown by Figure 24.

#### 9.3.2 Chart View



**Fig. 25 Chart View tab**

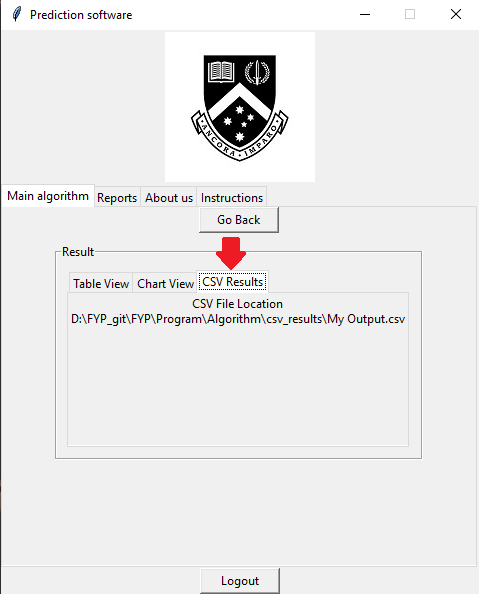
To switch to chart view, just click on the Chart View button, then the program will switch to a separate tab, where each of the performance metrics can be viewed in the form of a chart. To view a chart, click on the “Show Chart” button for any of the performance metrics. For this example, we will be viewing the results of AUC in the form of a chart.



**Fig. 26 Chart View results for AUC**

Once the “Show Chart” button is clicked on, a chart similar to the one in Figure 25 will appear on a separate window.

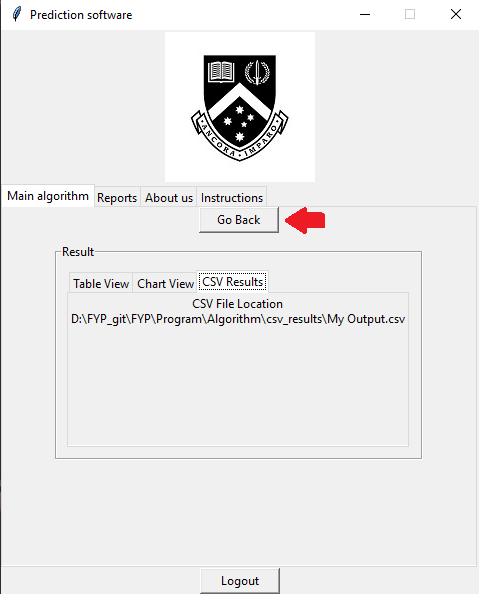
#### 9.3.3 CSV Results



**Fig. 27 CSV Results tab**

Finally, we can also view the results in its raw CSV file. Click on the “CSV Results” tab, this will show you the directory at which the CSV file of the results is located. To view the CSV file, navigate to that directory and open the CSV file with a software such as Notepad or Microsoft Excel.

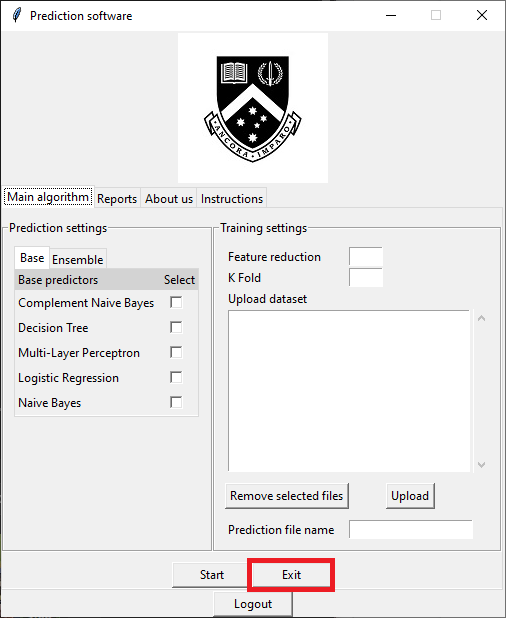
### 9.4 Going back to the homepage



**Fig. 28 Table View of the results**

After we are done with the program, click on the “Go Back” button to return to the homepage. This can be done at any step of the program if you feel that some settings need to be changed.

## 10 Exiting the application



**Fig. 29 Exit button**

To exit the program, there is an “Exit” button located in the Main algorithms section as shown in Figure 29. When the user clicks the exit button, a confirmation message will appear to prevent any accidental exits from the user.

# Technical Guide

## 1. Software downloads

### 1.1 Python

****

**Fig. 30 Python icon**

The program was built mainly using Python, so Python is required to run our application. For installation, you can visit <https://www.python.org/downloads/> and download the Python installer. For releases, we recommend installing any Python builds version above 3.8.

### 1.2 Python packages

For this section, you will need the package-management system Pip installed. Usually, Pip will be automatically installed along with Python. If pip is not installed in your machine then you can visit <https://pip.pypa.io/en/stable/installation/> to learn how to manually install it.

Our application requires several python packages to run its internal computation. These packages can be installed using Pip. Below is a table listing the packages used, along with the Pip install command lines for installation through an operating system’s shell.

**Table 1: Descriptions of Python packages**

| **Package name** | **Details** | **Pip install command line** |
| --- | --- | --- |
| Scikit Learn | A machine learning library that contains various classification, regression and clustering techniques and is designed for Python Programming Language. | pip install sklearn |
| NumPy | A library that supports mathematical functions that operates large multidimensional arrays and matrices. | pip install numpy |
| SciPy | A library that is used for scientific computing and technical computing and contains modules for optimization, linear algebra as well as image processing | pip install scipy |
| Matplotlib | A plotting library that creates static, animated and interactive visualizations in python. | pip matplotlib |

**Table 1: Descriptions of Python packages (continued)**

| **Package name** | **Details** | **Pip install command line** |
| --- | --- | --- |
| Tkinter | The standard Python GUI library, used to create interfaces for the application. | pip install tk |
| rotation-forest | Library containing the implementation of the Rotation Forest algorithm in Sklearn style. | pip install rotation-forest |
| Imblearn | A library containing tools and techniques for handling datasets with high degree of imbalance. | pip install imblearn |
| Pandas | A data handling library, containing operations for analyzing and manipulating data structures. | pip install pandas |
| tkPDFViewer | tkiner library extension that allows PDF files to be displayed within the GUI | pip install tkPDFViewer |

## 2. Installation

Our program will be packaged using Zip or NSIS. The following sections will explain how to install the program based on the package type you received

### 2.1 Zip

For the zip version of our program package, the files can be extracted using softwares such as WinZip or WinRAR. After extraction, a folder should appear which contains all the program files required.

### 

### 2.2 NSIS

### 

**Fig. 31 Installer screen**

For the NSIS version of our program package, no external software is required. The exe installer can be run normally, and you will be presented with a screen as shown in Figure 31. Here you can set the destination folder and click on the “install” button to begin installation.