National College of Ireland

Configuration Manual

MSc Research Project
Data Analytics

Laksana Vongchalerm
X19101538

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……. Laksana Vongchalerm……

Student ID: ……X19101538……

Programme: ……MSc. In Data Analytics…… **Year:** …2021-2022

Module: ……MSc. Research Project……

Supervisor: ……Vladimir Milosavljevic……

Submission Due Date: ……31st January 2022……

Project Title: ……Analysis of predicting the success of the banking telemarketing campaigns by using machine learning techniques…

Word Count: ……769……..…….…….. **Page Count**……8…….

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……Laksana Vongchalerm……

Date: ……31st January 2022……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Laksana Vongchalerm
X19101538

1 Introduction

The configuration manual provides information about software were applied in the research project. Furthermore, screenshots and descriptions from the codes and outputs.

2 Data Set

The dataset is collected from UCI Machine Learning Repository as a CSV file. After downloading the file, it was imported into R Studio to solve the missing data.

3 Hardware Requirements

The project was used to perform artefacts as below.

Operating System	Windows 11
Processor	Intel Core i5-10210U CPU @ 1.60GHz 2.11 GHz
RAM	8 GB
Memory	500 GB

4 R Studio

The dataset is collected from UCI Machine Learning Repository as a CSV file. After downloading the CSV file, it was imported into R Studio to adjust the empty data. Figure 1 shows that the screenshot in R Studio.

```
setwd("D:/Work/R")
File[is.na(File)] = "unknown"
write.csv(File,"File.csv", row.names = FALSE)
View(File)
```

Figure 1: Missing Data

5 IBM SPSS Statistics

IBM SPSS Statistics is software that assists initial data analysis. After solving the empty data, we used the new file to import into the IBM SPSS Statistics to consider the missing data as an unknown.

6 WEKA Application

We used WEKA to do the imbalance data and create models like Logistic Regression and Decision Tree.

6.1 Open WEKA Application and click at the Explorer button.

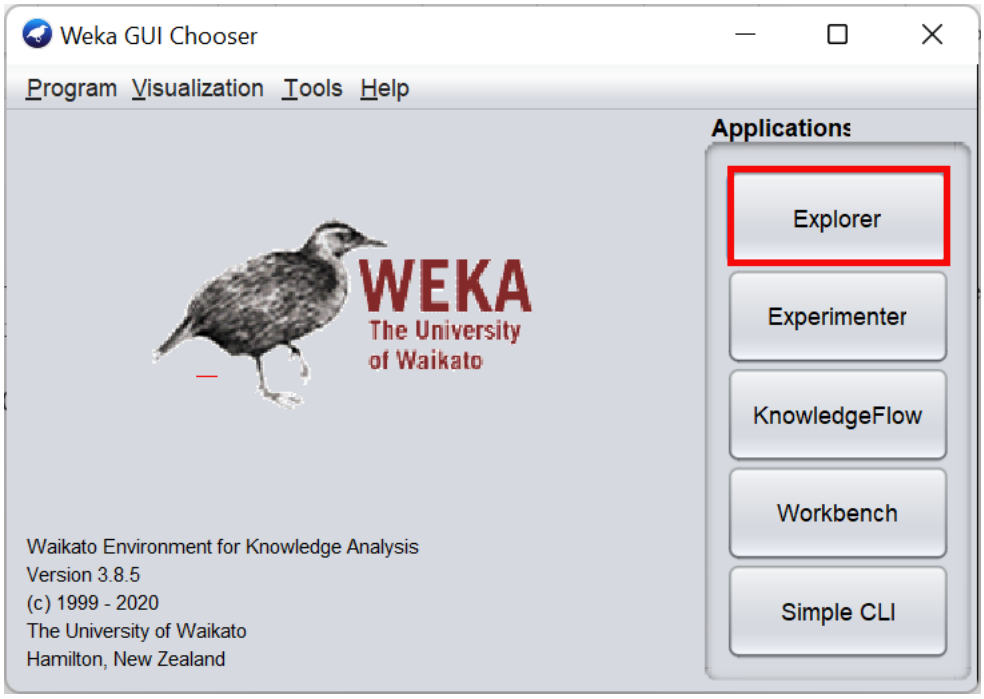

Figure 2: WEKA Application

6.2 Click Open files > select the dataset > Open. The application shows WEKA interface, and the data is imbalance.

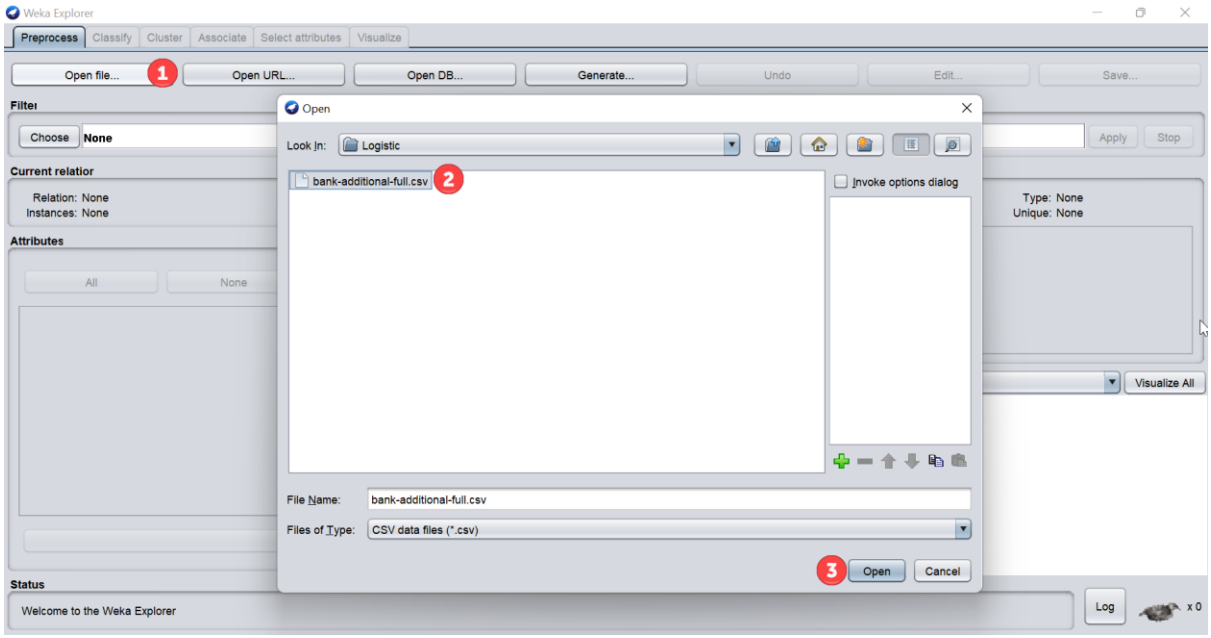

Figure 3: Select dataset

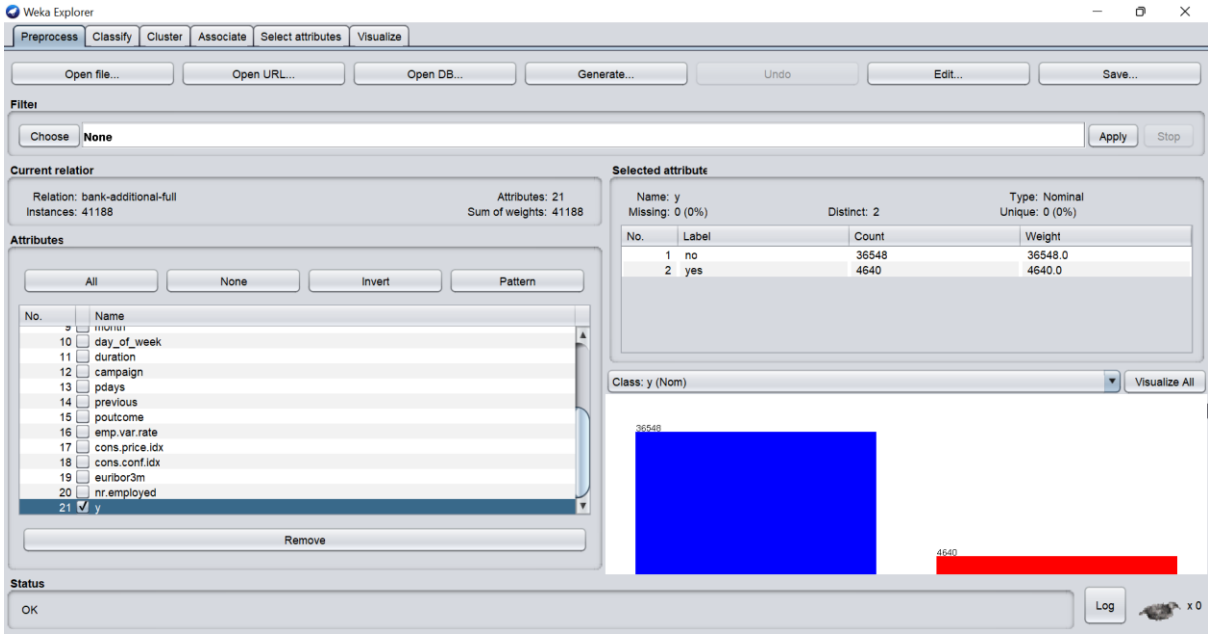

Figure 4: WEKA application interface

6.3 To solve the imbalance data, click choose > filters > supervised > Instance > SpreadSubsample. The application displays the balance data.

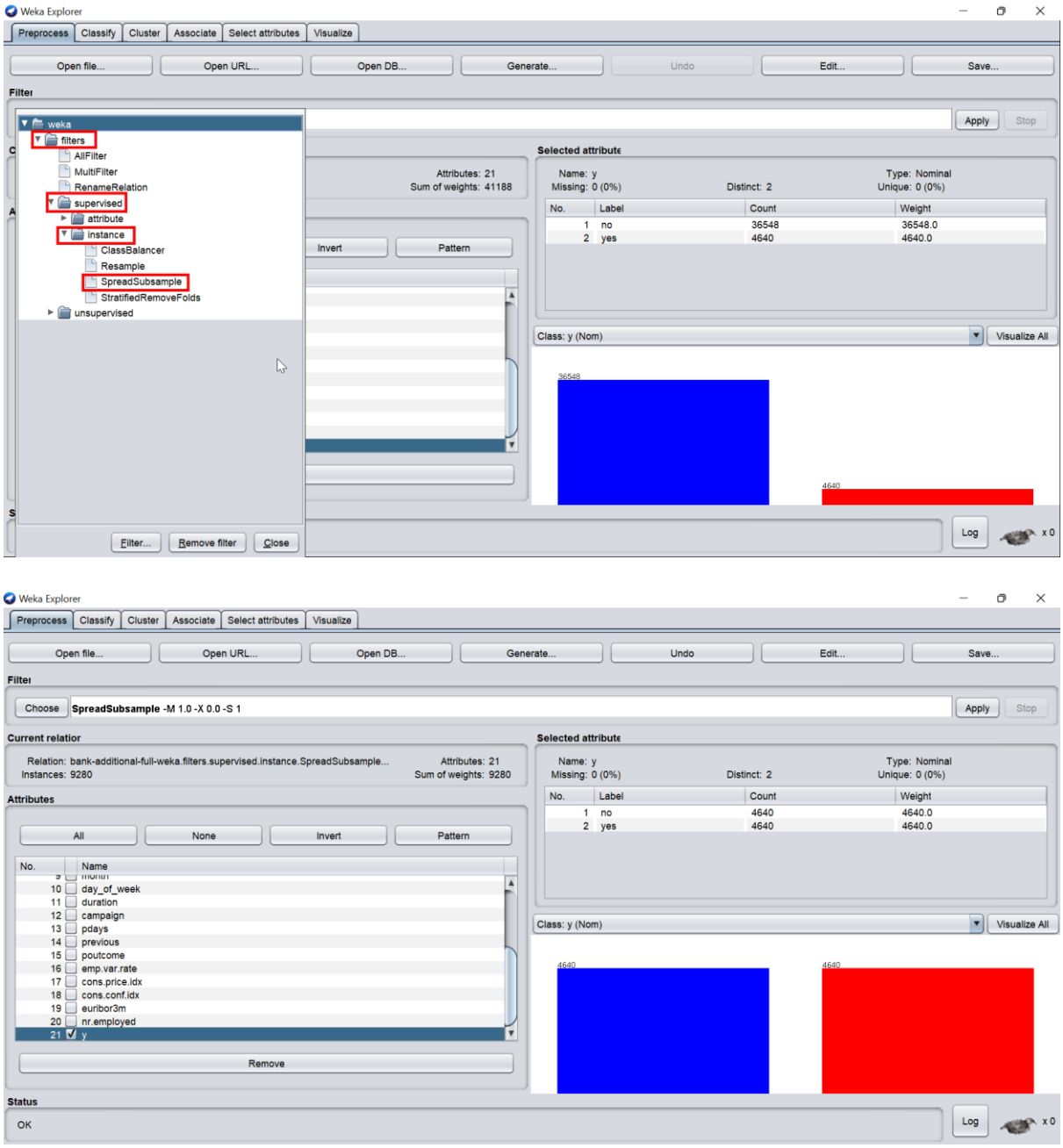

Figure 5: Balance Data

6.4 After the data is balanced, select the Classify tab and then choose the machine learning algorithm. In this paper selected Logistic.

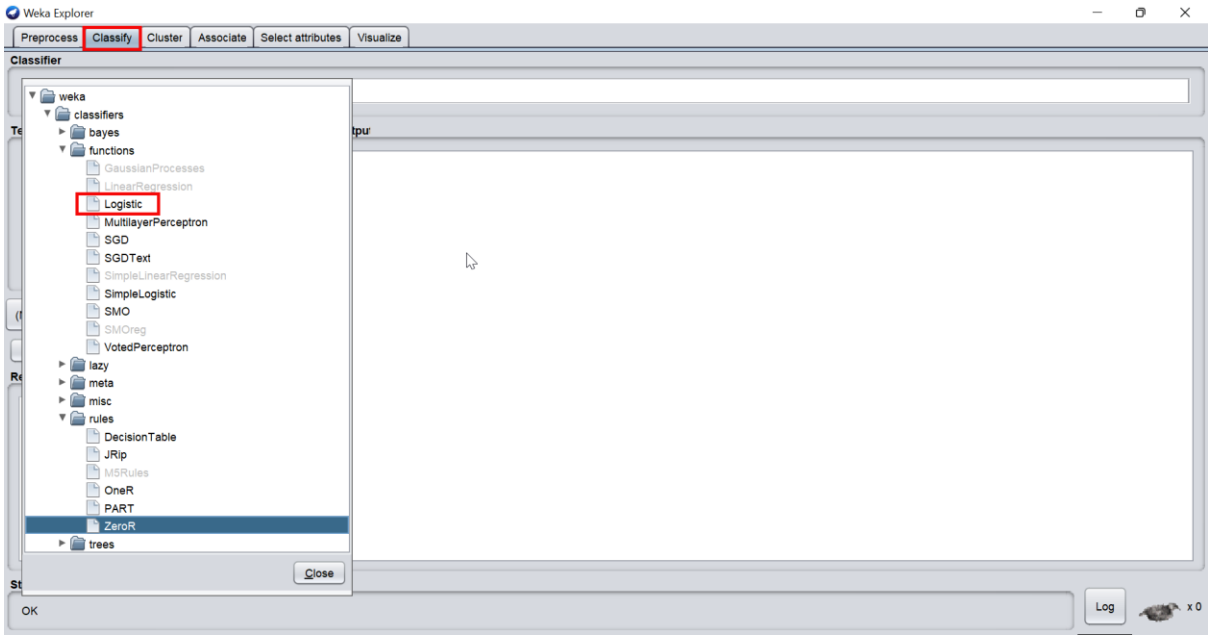

Figure 6: Selecting the machine learning algorithm: Logistic Regression

6.5 Confirmation to select the machine learning algorithm by clicking the Start button then the application shows the results. Figure 8 shows the results of Logistic Regression with the AUC (Area under ROC Curve) was 93.4%, with precision and recall values of 87% and 86.9%, respectively.

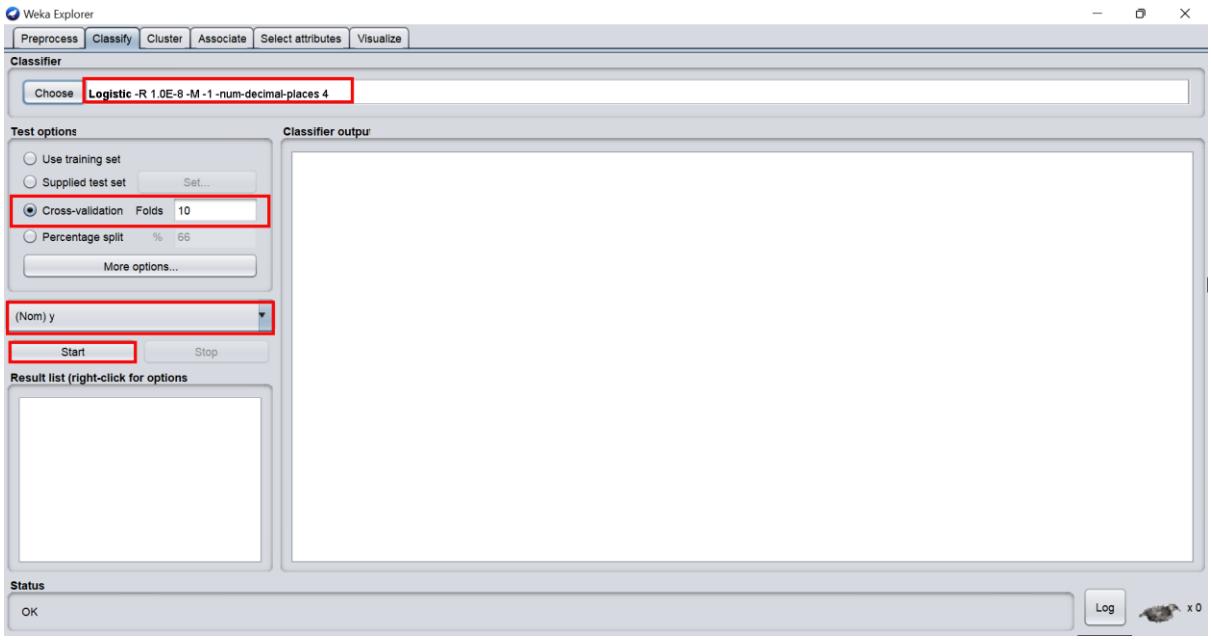

Figure 7: Confirmation the machine learning algorithm: Logistic Regression

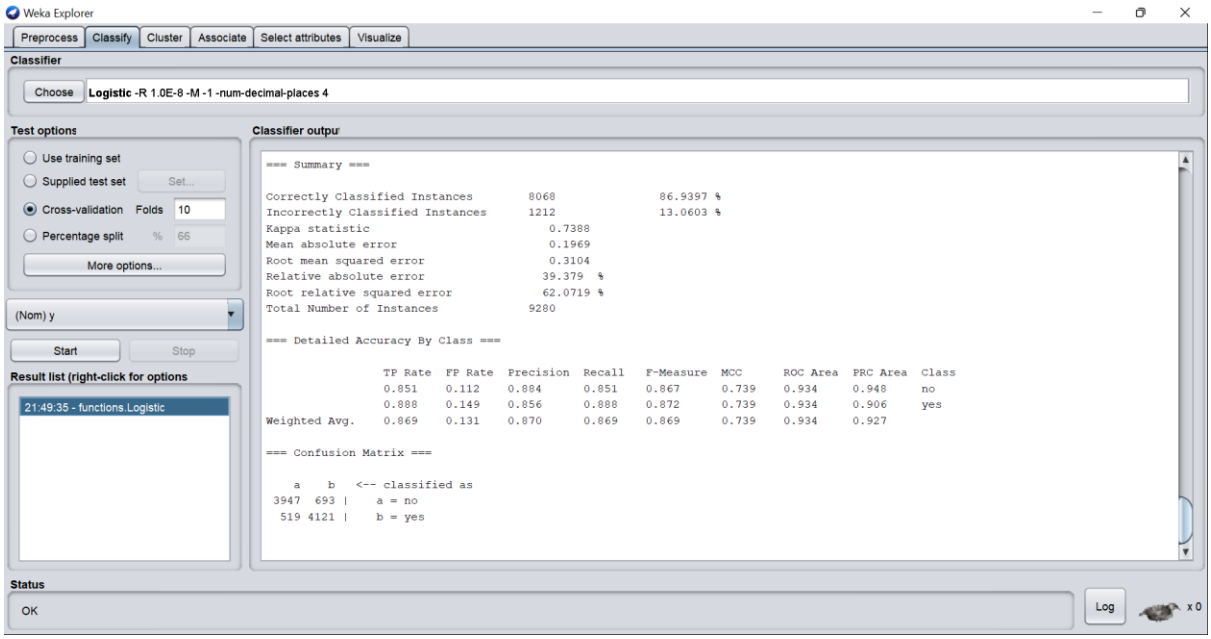

Figure 8: Logistic Regression result

6.6 For decision tree, select the Classify tab and then choose the machine learning algorithm. In this case selected trees > J48

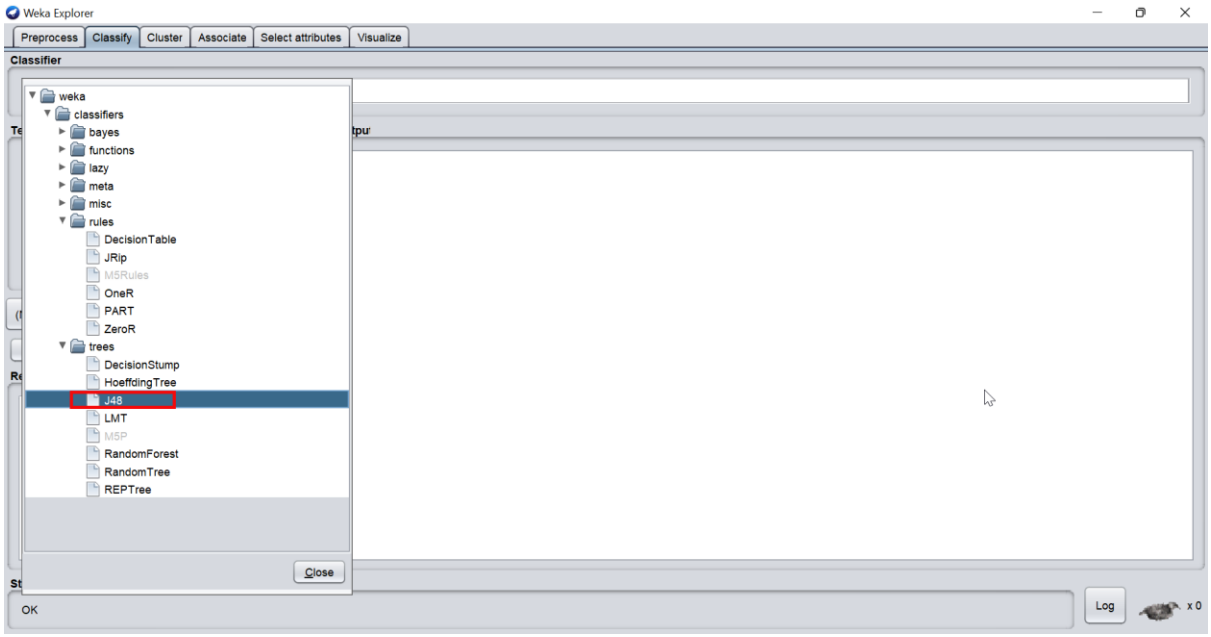

Figure 9: Selecting the machine learning algorithm: Decision Tree

6.7 Confirmation to select the machine learning algorithm by clicking the Start button then the application shows the results. Figure 11 shows the results of Decision Tree with the AUC (Area under ROC Curve) was 86.6%, with precision and recall values of 84.4%.

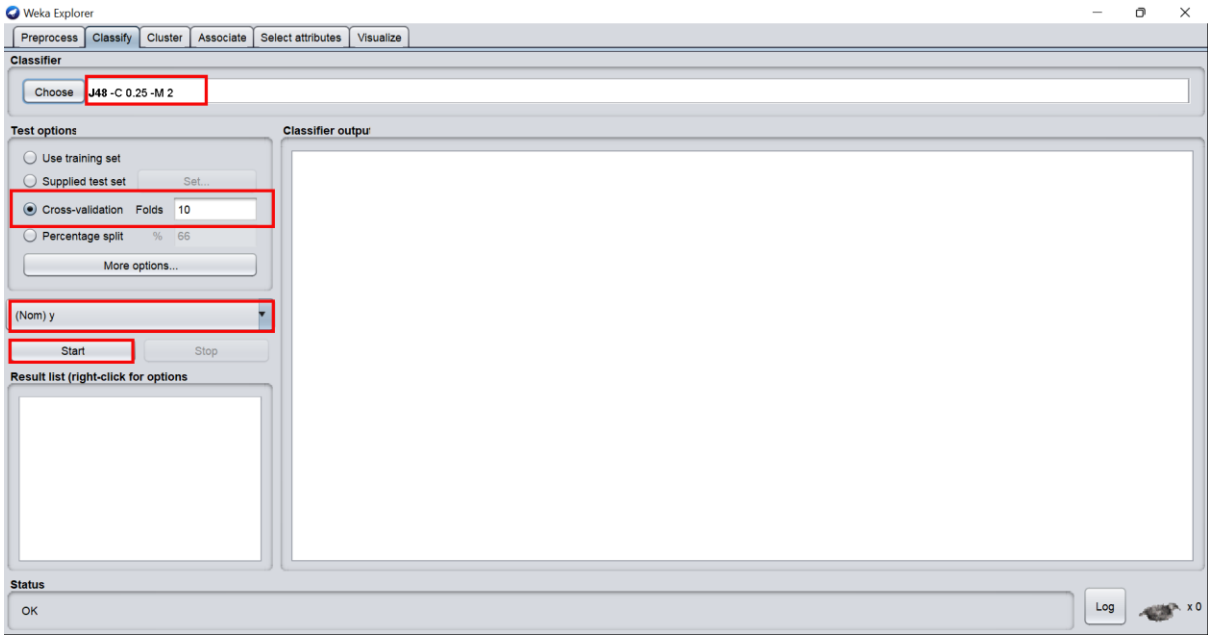

Figure 10: Confirmation the machine learning algorithm: Decision Tree

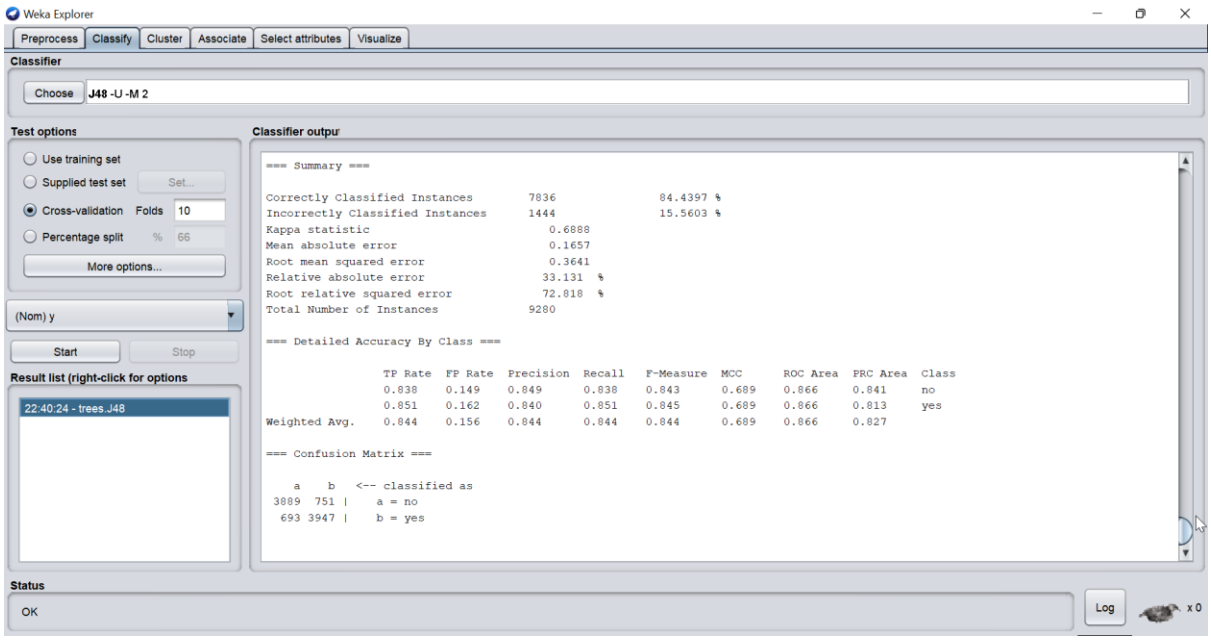

Figure 11: Decision Tree Result

6.8 The most performance was Logistic Regression and then used the model to find the ranking by clicking the Select attributes tab and set the parameter as Figure 12. After set the parameter, click the Start button. The application shows the ranking as Figure 12.

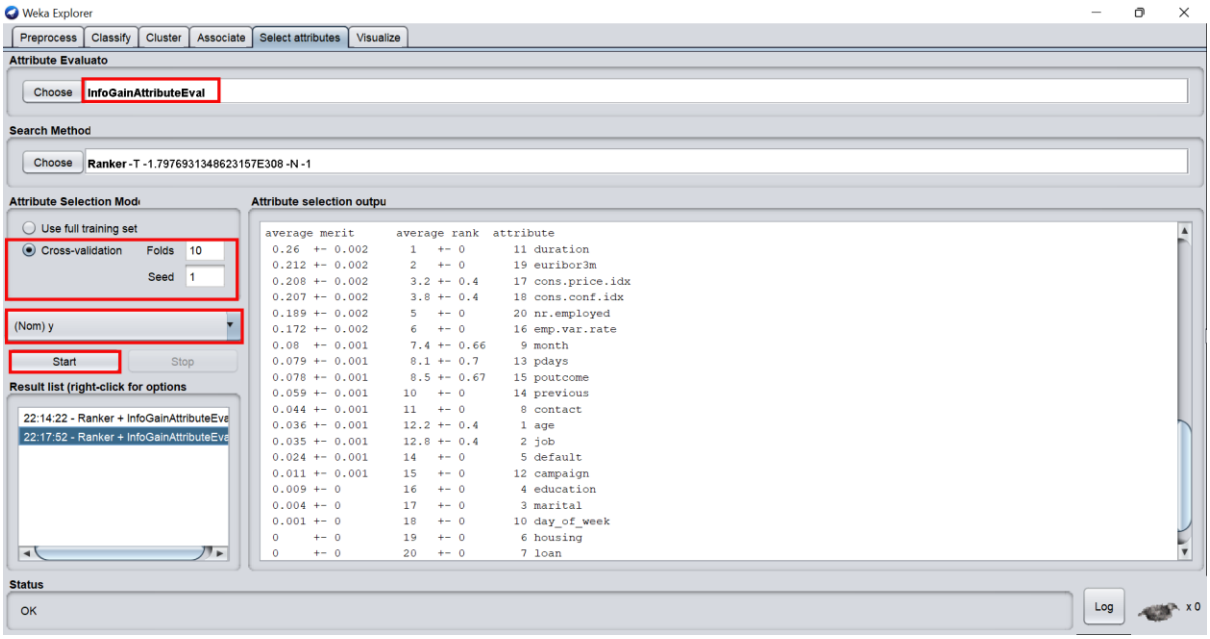

Figure 12: Ranking Attributes