Configuration Manual

MSc Internship
Cyber Security

David Damilola Ibinaiye
Student ID: x18170455

School of Computing
National College of Ireland

Supervisor: Imran Khan

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	David Damilola Ibinaiye
Student ID:	X18170455
Programme:	M.Sc. Cyber Security **Year:** 2019
Module:	Academic Internship
Lecturer:	Imran Khan
Submission Due Date:	Thursday, 12 December 2019
Project Title:	**Identification and Classification of Phishing Websites Using Machine Learning – Random Forest**
Word Count:	450 **Page Count:** 11

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: ……

Date: 12 Dec. 19

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

David Damilola Ibinaiye
Student ID: x18170455

1 WEKA Download

1.1 Feature confirmation of WEKA application.

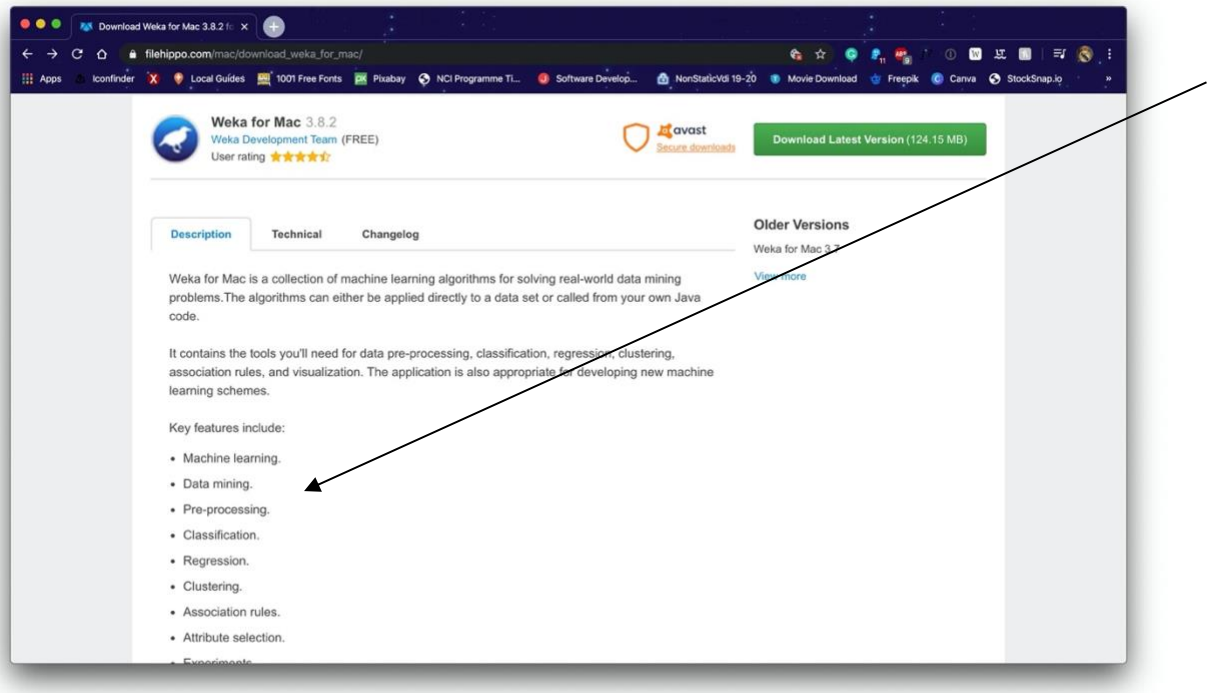

Fig. 1. The image above shows the Key features of the WEKA application which includes machine learning, data mining, pre-processing, classification, regression, etc. The link to the mac application would be at reference [1].

1.2 Installation of WEKA

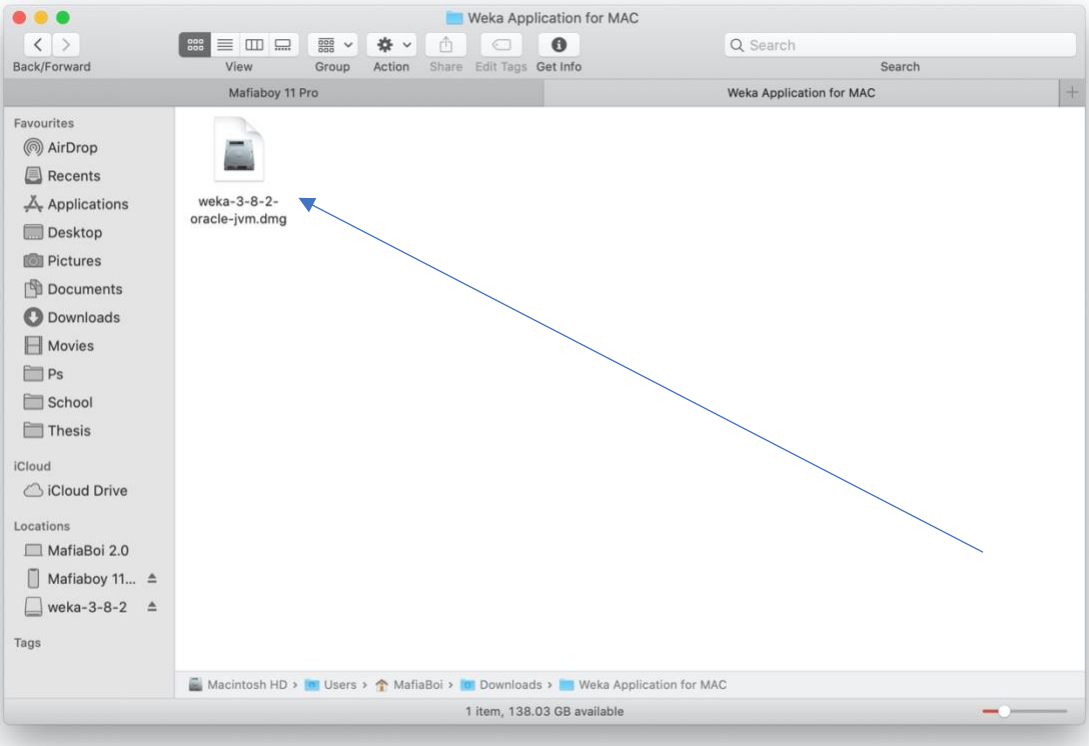

Fig. 2 Double click on the ***weka-3-8-2-oracle-jvm.dmg*** file to extract the WEKA application for installation, which is shown in the figure below.

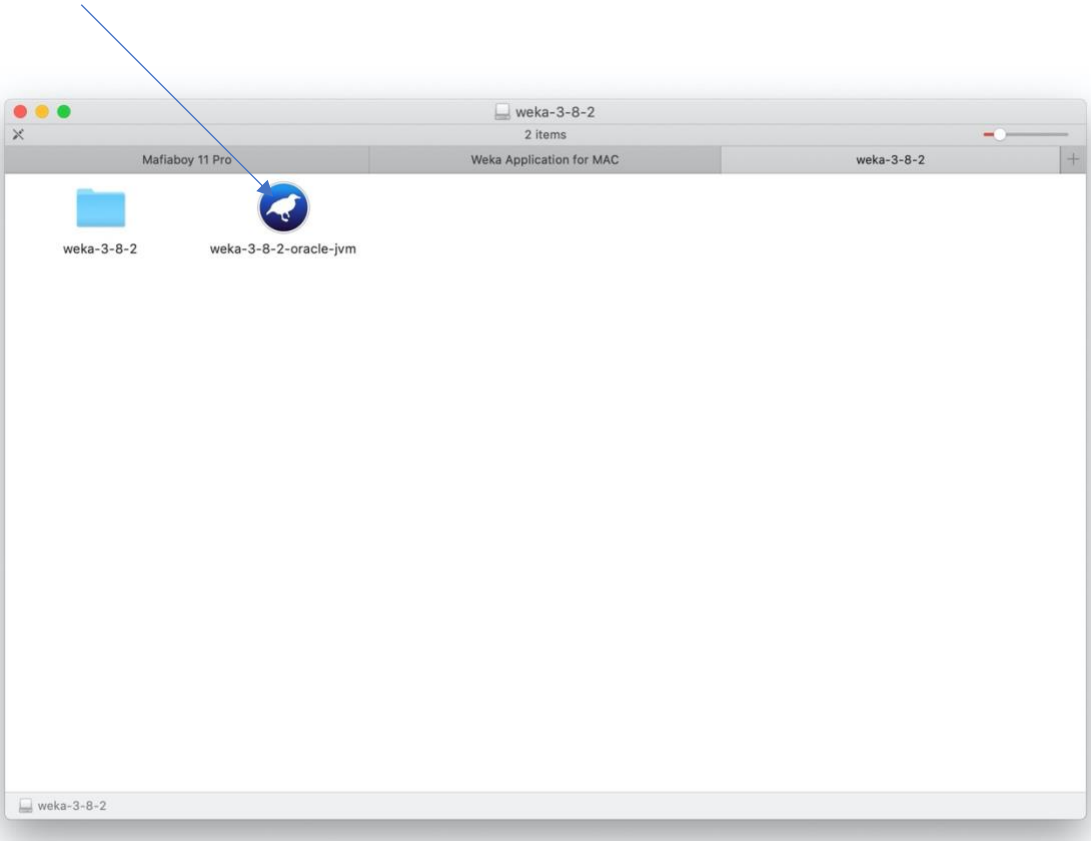

Fig. 3 Launching the WEKA application, double click on the ***weka-3-8-2-oracle-jvm*** file to install WEKA and JRE into your computer, which would bring up a message box shown in the figure below.

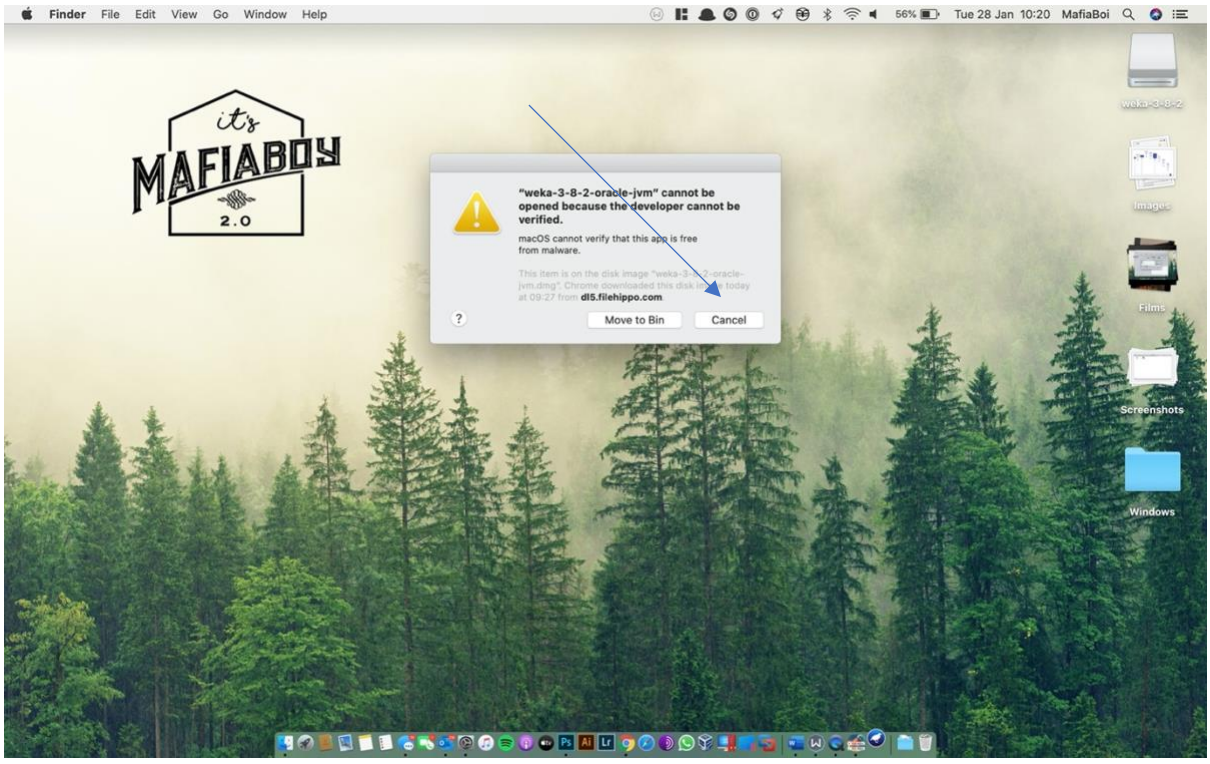

Fig. 4 This prompt message is alerting you that the application you are about to install cannot be verified by the Mac OS. Select Cancel and open System preferences.

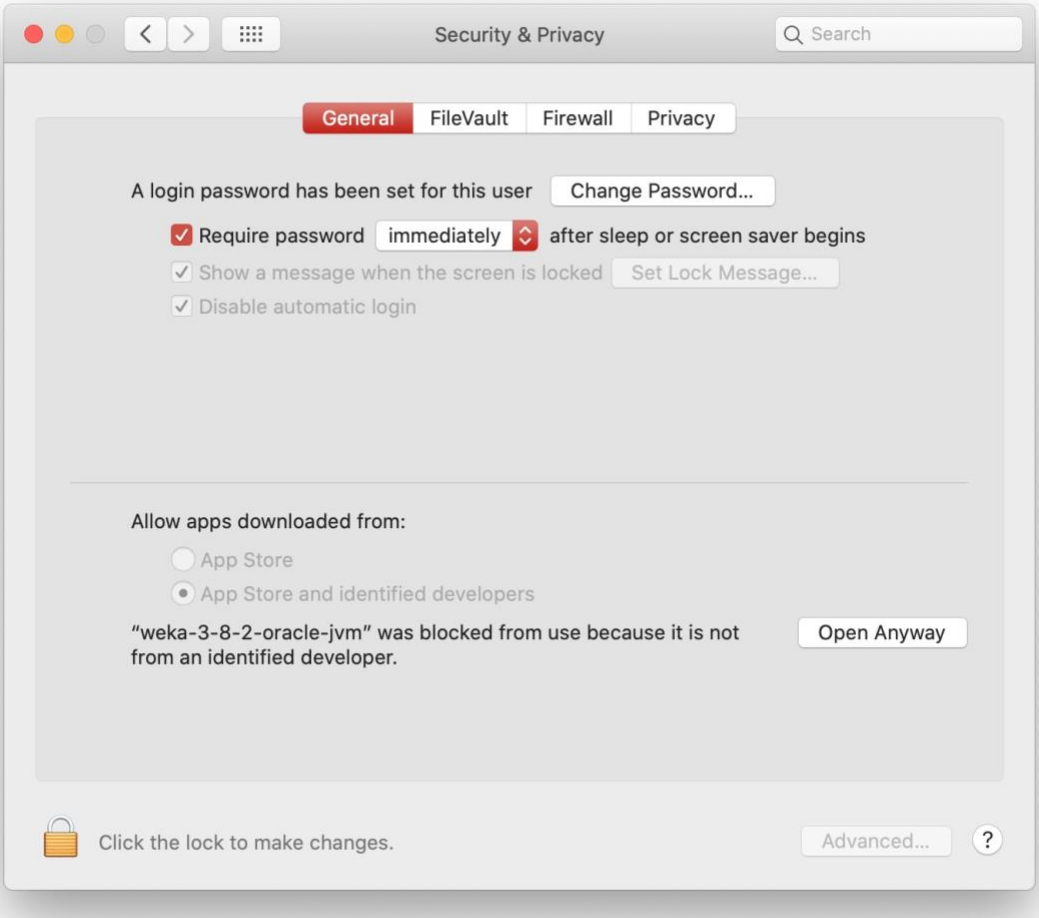

Fig. 5 When in System preferences, select Security & Privacy and click on the General tab. Locate at the bottom the Open Anyway button and click it which would bring up a prompt message shown below.

Fig. 6 This message is restating what was mentioned in the previous message about the Mac OS not recognising the developer. Locate the Open button at the bottom and click it.

2 WEKA Application

2.1 Select Explorer

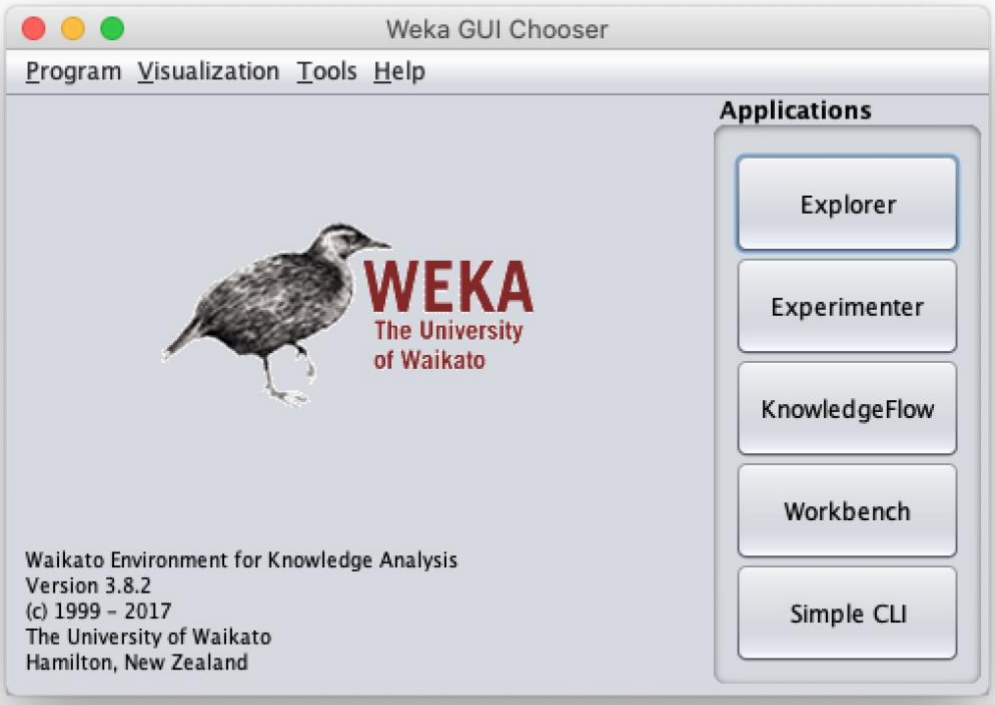

Fig. 7 WEKA launch window. Once the WEKA application loads up, click on Explorer on the top right corner to launch the window where the classification experiment would be carried out.

2.2 Load phishing data file into the WEKA Application

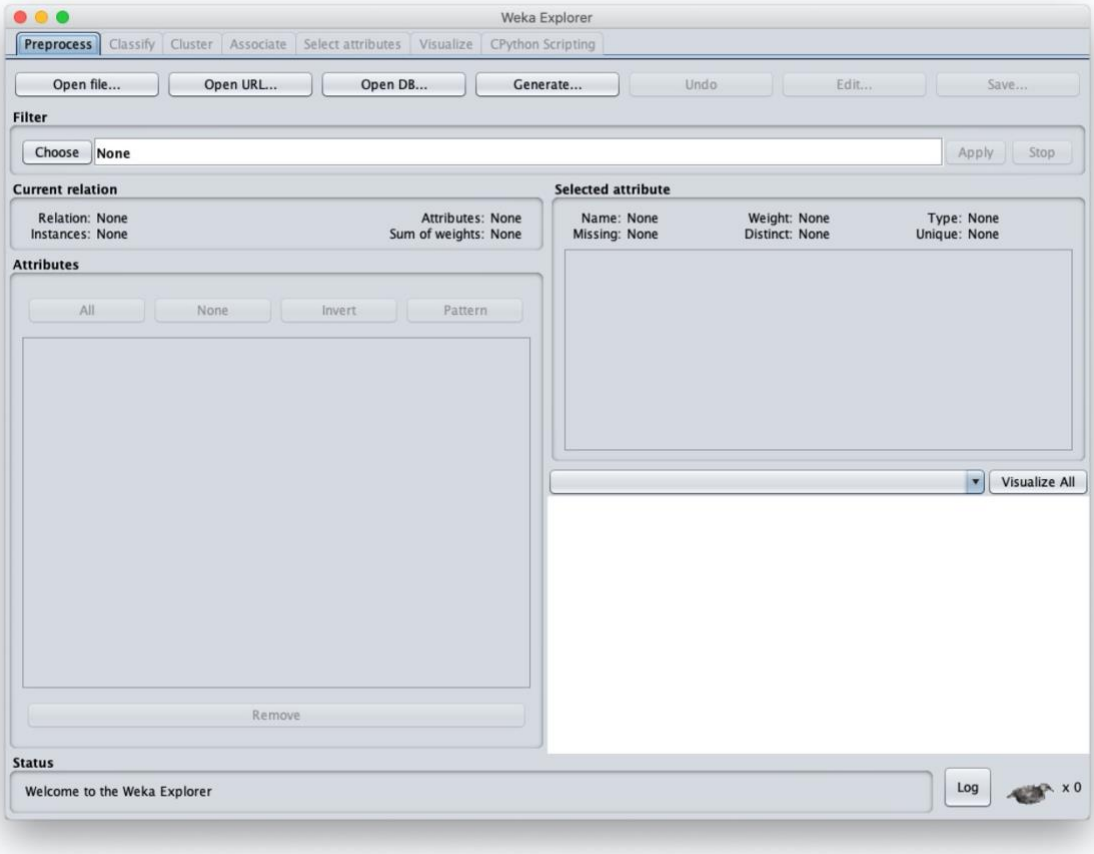

Fig. 8 Select Open file to select the raw dataset file that would be coverted for the classificattion experiment.

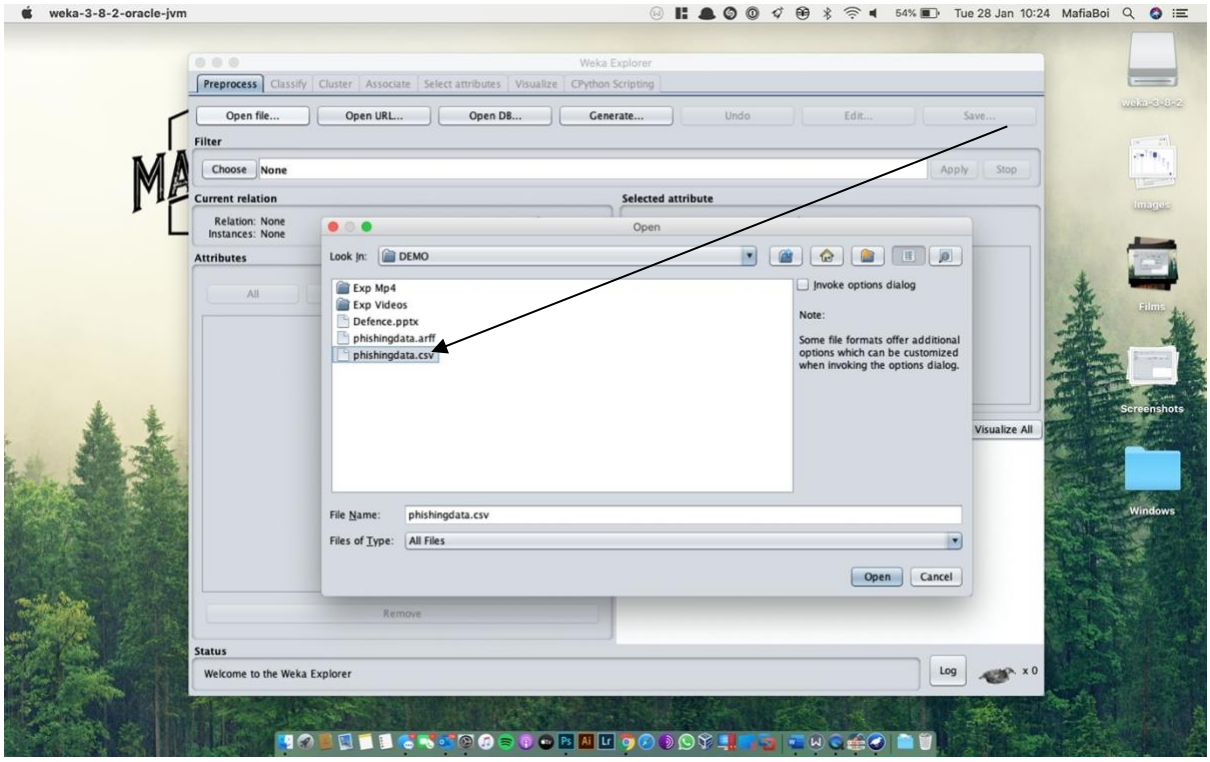

Fig. 9. Locate the dataset file, set Files of Type to All Files and select the dataset file, click the open button at the bottom.

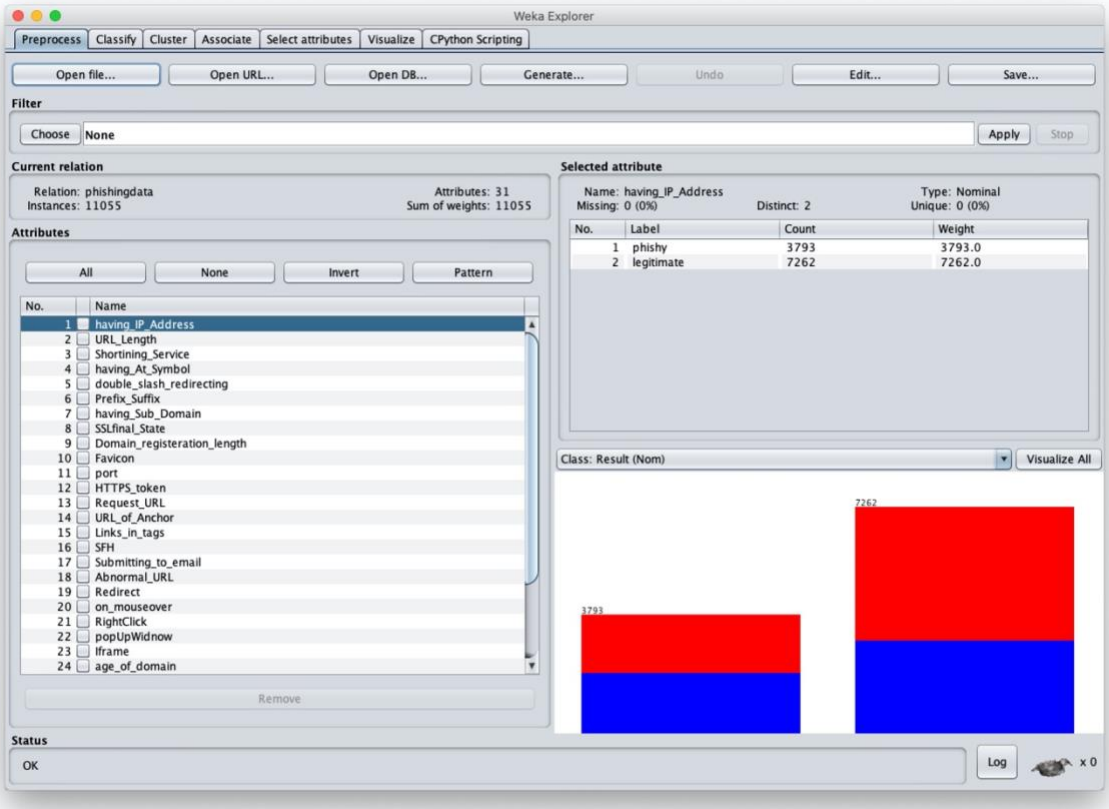

Fig. 10. The image above shows the WEKA application interface displaying the phishing data loaded with a list of file attributes. The image shows the selected file has 31 attributes and 11055 instances.

2.3 Convert dataset file to .arff

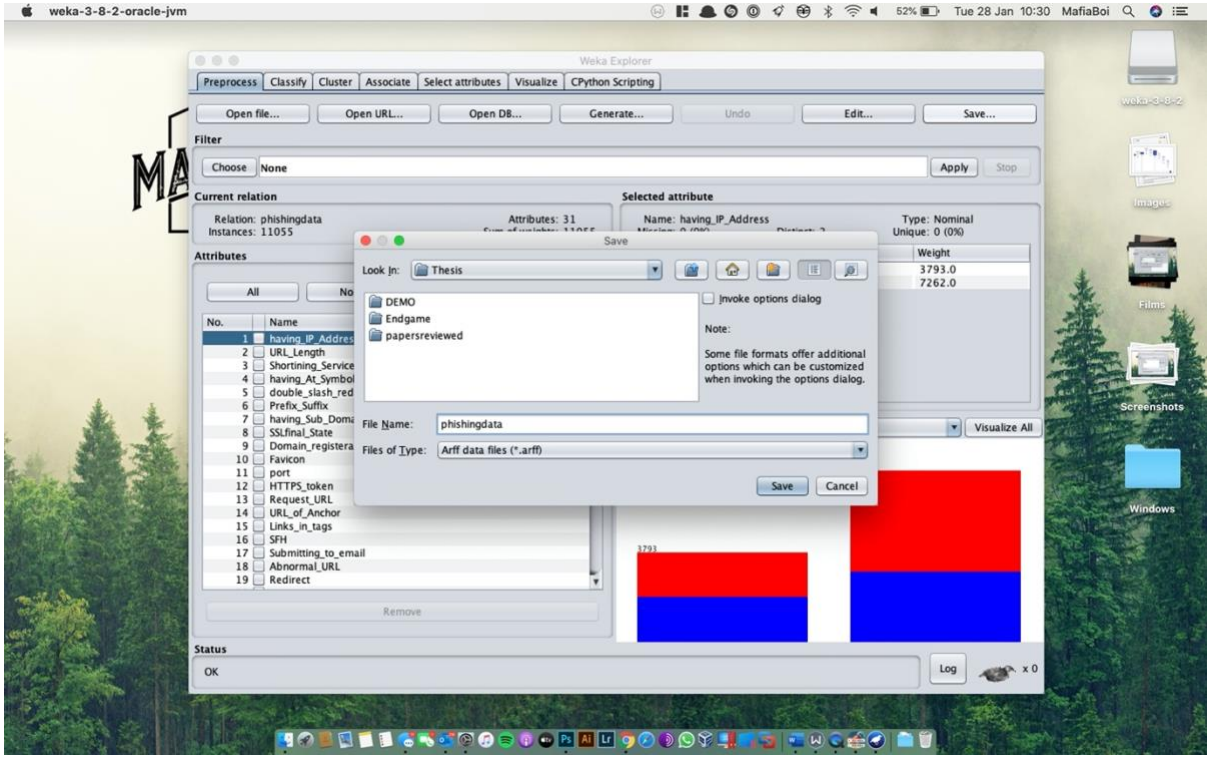

Fig. 11 Click the Save button on the top and on the pop up box set Files of Type to Arff data files. Click the Save button below to save the loaded phishing data file in .arff format.

3 Random Forest Classification using WEKA

3.1 Selecting the Random Forest Application

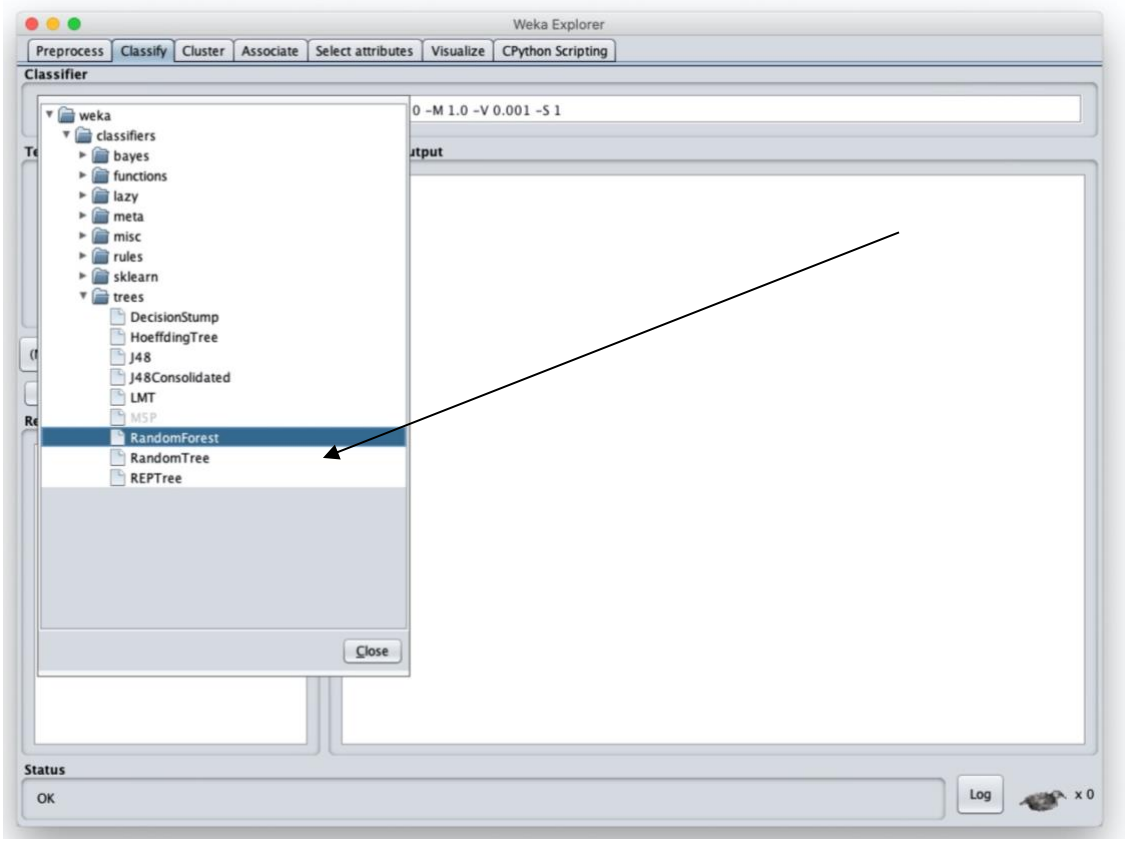

Fig. 12 After saving the file in .arff format, select the Classify tab at the top of the application, then select the Machine learning algorithm of choice; in this case Random Forest.

3.2 Confirming the selected algorithm

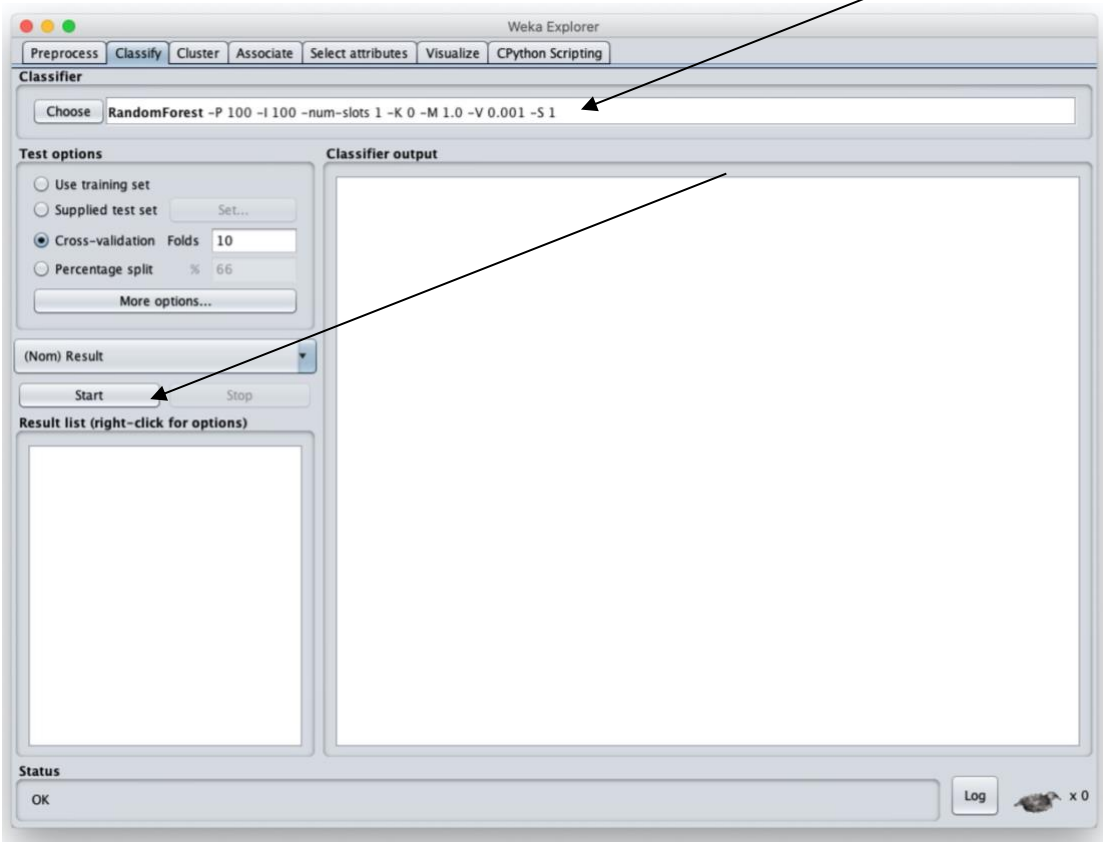

Fig. 13. Make the sure the selected algorithm has been loaded. Select Cross-validation and set to 10. Select the Start button on the left panel to begin the classification process using the selected algorithm.

3.3 Result Evaluation

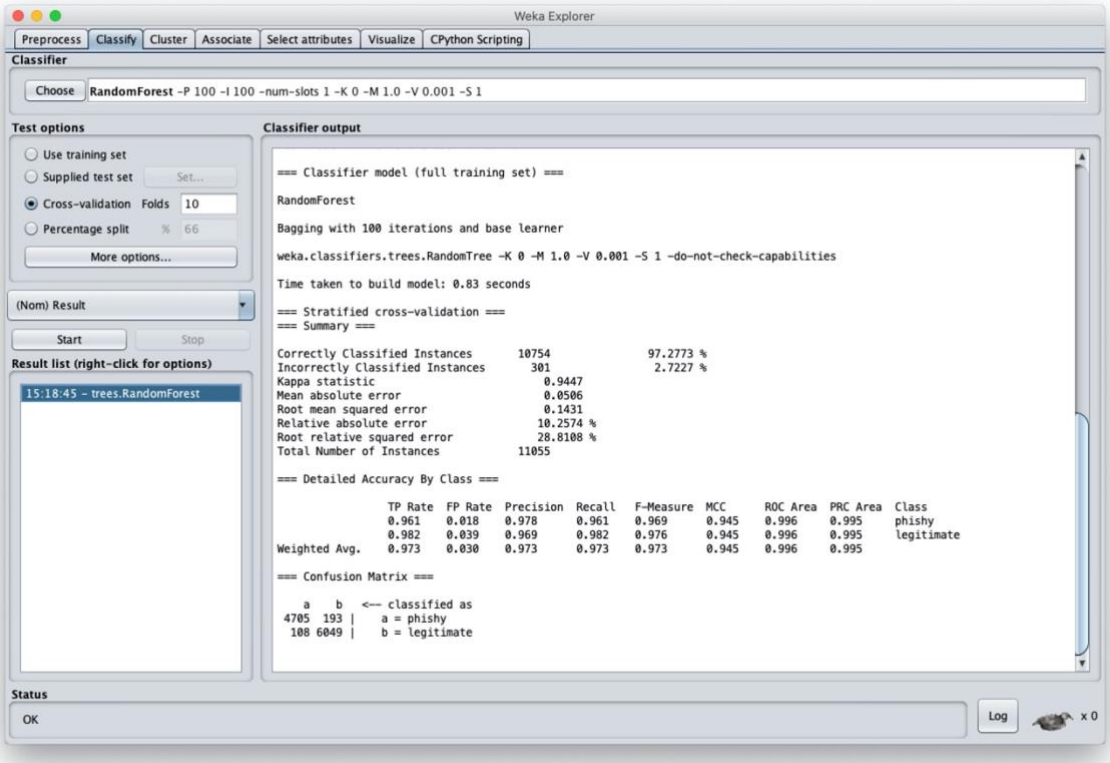

Fig. 14 The image below show the classification result of the experiment. The random forest algorithm yielded an accuracy result of 97.2%, with 10754 out of 11055 correctly classified. The confusion matrix table shows a True Negative (TN) result of 4705 and a True Positive (TP) result of 6049.

4 References

[1] Weka, "Filehippo," Weka, 2018. [Online]. Available: https://filehippo.com/mac/download_weka_for_mac/. [Accessed 11 December 2019].