

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





School of Computing

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

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.

n filehippo.com/mac/download_weka_for_mac/ X 🗣 Local Guides 🗮 1001 Free Forts 🐹 Pluabay 📀 NCI Programme TL. 🔮 Software Develop 🙆 NonStaticVid 18-20	 A O B LE III =7 O LE LE III =7 O E Movie Download
Weka for Mac 3.8.2 Weka Development Team (FREE) User rating *****	Download Latest Version (124.15 MB)
Description Technical Changelog	Dider Versions Reka for Mac 3
Weka for Mac is a collection of machine learning algorithms for solving real-world data mining V problems. The algorithms can either be applied directly to a data set or called from your own Java code.	service -
It contains the tools you'll need for data pre-processing, classification, regression, clustering, association rules, and visualization. The application is also appropriate for developing new machine learning schemes.	
Key features include:	
Machine learning.	
Data mining.	
Pre-processing.	
Classification.	
Regression.	
Clustering.	
Association rules.	
Attribute selection.	

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

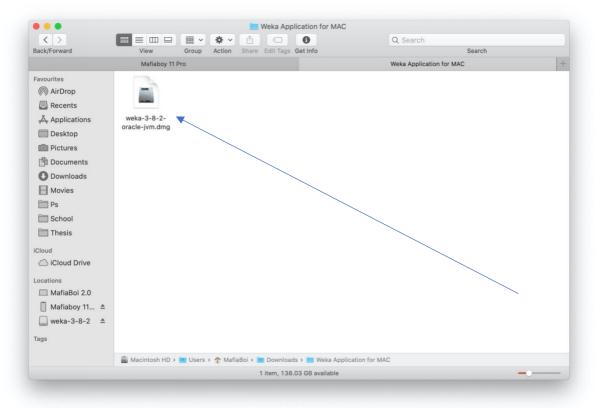


Fig. 2 Double click on the *weka-3-8-2-oracle-jvm.dmg* file to extract the WEKA application for installation, which is shown in the figure below.

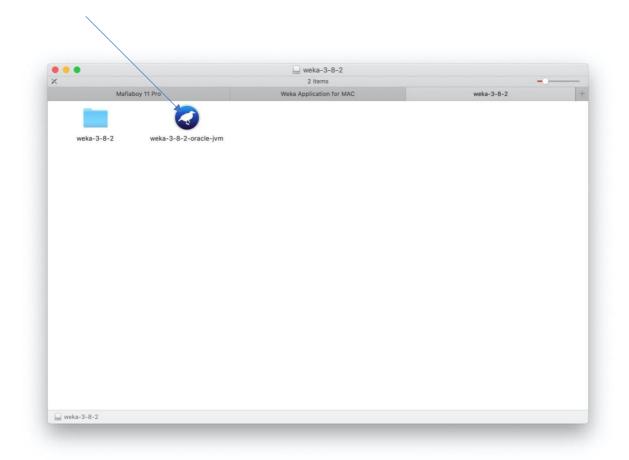


Fig. 3 Launching the WEKA application, double click on the *weka-3-8-2-oracle-jvm* file to install WEKA and JRE into your computer, which would bring up a message box shown in the figure below.



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.

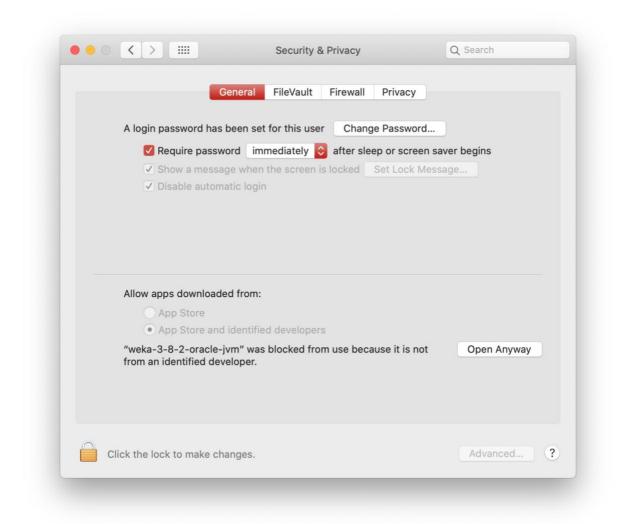


Fig. 5 When in System preferences, select Security & Privacy and click on the General tab. Locate at the bottom the Open Anyway button and click it which would bring up a prompt message shown below.

	macOS cannot verify the developer of "weka-3-8-2-oracle-jvm". Are you sure you want to open it?
	By opening this app, you will be overriding system security which can expose your computer and personal information to malware that may harm your Mac or compromise your privacy.
	This item is on the disk image "weka-3-8-2-oracle-jvm.dmg" Chrome downloaded this disk image today at 09:27 from dl5.filehippo.com
?	Move to Bin Open Cancel

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

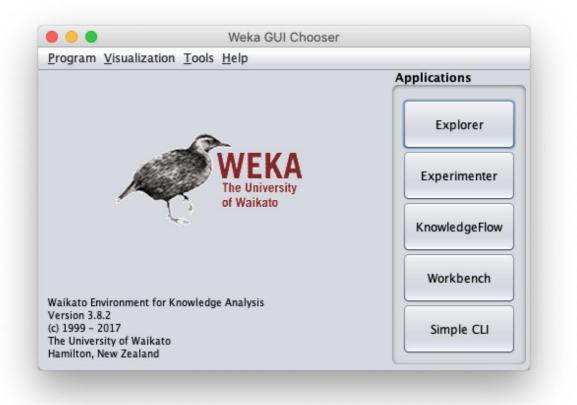


Fig. 7 WEKA launch window. Once the WEKA application loads up, click on Explorer on the top right corner to launch the window where the classification experiment would be carried out.

2.2 Load phishing data file into the WEKA Application

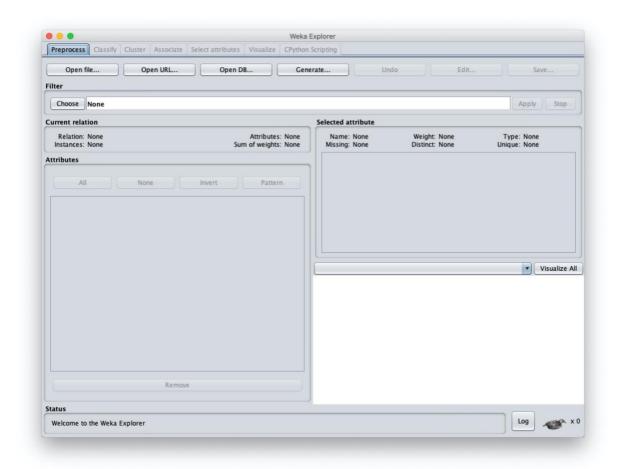


Fig. 8 Select Open file to select the raw dataset file that would be coverted for the classificattion experiment.

ſ	Preprocess Classify Open file	Weka Explorer Y Cluster Associate Select attributes Visualize CPython Scripting Open URL Open DB Cenerate	Undo Edit Save	Walat-6-6-2
5.7.1	Filter			1410 A
MA	Choose None		Apply Stop	
1.11	Current relation	Selected a	attribute	lingers
L	Relation: None Instances: None	• • • Open		
	Attributes	Look in: DEMO		
	Al	Exp Mp4 Exp Videos P Gence.ppt Phishingdata.arff Phishingdata.csv	Invoke options dialog Note: Some file formats offer additional options which can be customized when invoking the options dialog. Visualize All	Errenshots
Service Contraction		File Name: phishingdata.csv		Windows
		Files of Type: All Files		
			Open Cancel	
		Remove		
	Status			
a state of the	Welcome to the Weka	a Explorer	Log 🛷 x 0	

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.

Ope	n file Open URL Open DB Gene	erate Undo	Ec	dit Save
er				
Choose	None			Apply Stop
rent rela	tion	Selected attribute		
Relation: nstances:	phishingdata Attributes: 31 11055 Sum of weights: 11055	Name: having_IP_Address Missing: 0 (0%)	Distinct: 2	Type: Nominal Unique: 0 (0%)
ributes		No. Label	Count	Weight
induces		1 phishy	3793	3793.0
		2 legitimate	7262	7262.0
14	All None Invert Pattern			
lo.	Name			
1	having IP_Address			
2	URL_Length			
3 [Shortining_Service			
4	having_At_Symbol			
	double_slash_redirecting			
	Prefix_Suffix			
	having_Sub_Domain			
	SSLfinal_State			
	Domain_registeration_length			
	Favicon	Class: Result (Nom)		Visualize
	port			
	HTTPS_token			262
	Request_URL		4	262
14 [URL_of_Anchor Links_in_tags			
	SFH			
	Submitting_to_email			
	Abnormal_URL			
	Redirect			
	on_mouseover	3793		
21	RightClick			
	popUpWidnow			
	Iframe			
24 [age_of_domain			
	Remove			
tus				

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

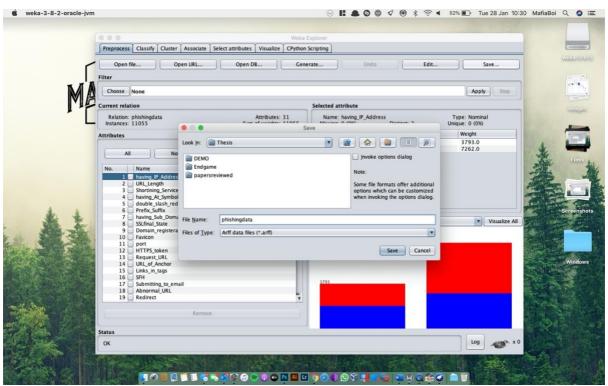


Fig. 11 Click the Save button on the top and on the pop up box set Files of Type to Arff data files. Click the Save button below to save the loaded phishing data file in .arff format.

3 Random Forest Classification using WEKA

3.1 Selecting the Random Forest Application

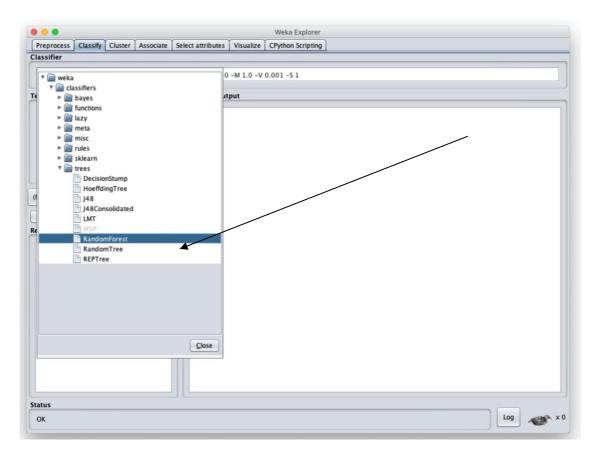


Fig. 12 After saving the file in .arff format, select the Classify tab at the top of the application, then select the Machine learning algorithm of choice; in this case Random Forest.

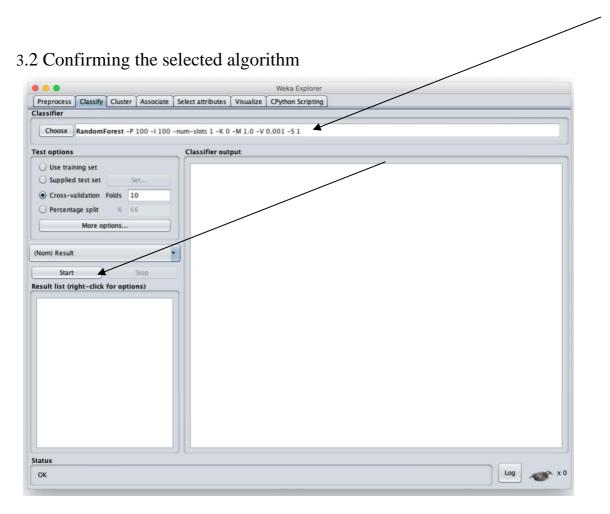


Fig. 13. Make the sure the selected algorithm has been loaded. Select Cross-validation and set to 10. Select the Start button on the left panel to begin the classification process using the selected algorithm.

3.3 Result Evaluation

Preprocess Classify Cluster Asso	ciate Select attributes Visualize CPython Scripting
lassifier	
Choose RandomForest -P 100 -	I 100 - num-slots 1 - K 0 - M 1.0 - V 0.001 - S 1
est options	Classifier output
Use training set Supplied test set Set Cross-validation Folds 10 Percentage split % 66 More options Nom) Result Start Stop esult list (right-click for options) 15:18:45 - trees.RandomForest	<pre></pre>
	a b < classified as 4785 193 a = phishy 108 6049 b = legitimate
tatus	
OK	

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].