Faculty of Engineering and Technology

Degree/ program	B.Tech	**Specialisation**	Computer Science	**Project Team ID** CSE02310184
Academic Year	2020	**Semester**	8	

Name of student	**Register Number**	**Department**	**Mobile Number**	**Email ID**
Shaik Riyaz	RA1611003010231	CSE	9952919265	sr9087@srmist.edu.in
R.B.Srinivas Raju	RA1611003010184	CSE	8331055577	rr4725@srmist.edu.in

Course Code	15CS496L	**Course Title**	Major Project

Mission Statement

Working Title of the Project:	AN EFFECTIVE MODEL OF TERMINATING PHISHING WEBSITES AND DETECTION BASED ON LOGISTIC REGRESSION
Project Site / Location	Chennai
Name and address of the company / organisation (Applicable for projects with industry or industry support)	SRM Institute of Science and Technology, Kattankulathur, Kanchipuram District-603203

Supervision Team

	Supervisor	**Co-Supervisor**	**External Supervisor (If applicable)**
Name	Mrs.K.R.Jansi		
Designation	Associate Professor		
Department	Computer Science and Engineering		
Campus	Kattankulathur		
Telephone	9600082712		
E-mail	jansi.k@ktr.srmuniv.ac.in		

Faculty of Engineering and Technology

Problem (or) Product Description

Mission Statement

Problem (or) Product Description: TERMINATING PHISHING WEBSITES AND DETECTION
These days phishing is one of the greatest and the quickest developing risk in light of the fact that phishing assailants will get data which was entered by the client and phishing aggressors will utilize that data without the client information. At present personal information is the most important thing compared to money, So phishing website attackers are focusing on the personal information of the user and they taking advantage when the user enters there personal information in the phishing websites. So the main theme of these projects is to avoid misuse of personal information by phishing attackers. To overcome this problem machine learning algorithm (Logistic regression) is used and provided with the massive dataset. From that dataset, it trains the algorithm and helps in detecting the new web links which are fraudulent links. Attackers disguise their website as legitimate and try to get data from the user for which they make users visit a website and get the personal information that is needed. Before trying to enter into those websites it is important to check whether the given link is good or a phishing link. By checking the link we can save ourselves from the attackers and can keep our data safe.
Assumptions and Constraints
The system is Should run on anaconda navigator

Division of work and contributors

Time period		Activities or components of the project	Name/Register Number of the Individual Contributor	Names/Register Number of the Joint Contributors
From Date	To Date			
23/12/19	24/12/19	Read and analyzed various papers on phishing detection. Read about the existing systems and discussed the pros and cons of each.	RA1611003010231	RA1611003010184
26/12/19	26/12/19	Prepared a presentation for the first review and noted down any required changes.	RA1611003010231	RA1611003010184
05/01/20	06/01/20	Analysed papers and read about classifier model used in the project.	RA1611003010231	RA1611003010184
09/01/20	11/01/20	Worked on the classifier model to create the first part of the project.	RA1611003010184	RA1611003010231
20/01/20	24/01/20	Dealt with the Logistic Regression and SVM Working Methodology	RA1611003010231	RA1611003010184
01/02/20	03/02/20	Learned the basics Logistic Regression and its advantages over SVM classifier and maximum entropy methods	RA1611003010184	RA1611003010231
04/02/20	05/02/20	Finished the basic Interface in the project.	RA1611003010184	RA1611003010231
08/02/20	08/02/20	Prepared 1st draft of research paper and showed to guide for feedback.	RA1611003010231	RA1611003010184
11/02/20	13/02/20	Studied about phishing websites and methods to extract the features	RA1611003010184	RA1611003010231
18/02/20	21/02/20	Paper presentation at the ICIOT conference.	RA1611003010231	RA1611003010184

Faculty of Engineering and Technology

Project Work – Student Log Book

Division of work and contributors

Time period		Activities or components of the project	Name/Register Number of the Individual Contributor	Names/Register Number of the Joint Contributors
From Date	To Date			
03/3/20	05/03/20	Worked on the accuracy of the logistic regression	RA1611003010231	RA1611003010184
10/03/20	11/03/20	Prepared the presentation for the third review.	RA1611003010231	RA1611003010184
14/03/20	14/03/20	Presented a semi working demo and showed to supervisor for inputs.	RA1611003010231	RA1611003010184
12/04/20	12/04/20	Discussed about the final presentation through online mode..	RA1611003010184	RA1611003010231
03/04/20	13/05/20	Worked on making the final report using latex..	RA1611003010184	RA1611003010231
15/05/20	15/05/20	Submitted the paper to the journal for publication	RA1611003010184	RA1611003010231
23/05/20	23/05/20	Received the mail for the paper publishing is under process	RA1611003010184	RA1611003010231
25/05/20	28/05/20	Started working on making a voice over presentation to be presented through online mode.	RA1611003010231	RA1611003010184
02/06/20	02/06/20	Submitted the final report, voice over presentation and the video recording of the demo to the guide.	RA1611003010184	RA1611003010231
05/06/20	05/06/20	Project demo and documentation completed and submitted	RA1611003010231	RA1611003010184

Faculty of Engineering and Technology

Project Work – Student Log Book

Summary record of major progress meetings with supervisors			**Working title of dissertation/research project:**	
Meeting date & supervisors present	**Progress since last meeting**	**Agreed programme of work and target dates**	**Other issues, e.g. facilities, supervision, training needs, etc.**	**Date of next meeting**
13/10/19 Mrs.K.R.Jansi	Selected the base paper and prepared the project abstract.	Preparation of Power point presentation with details about the project such as its prototypes, architecture, software description.	Phyton programming and machine learning methodologies for phishing detection.	26/12/19
26/12/19 Mrs.K.R.Jansi	We prepared for the 1st review. Power point Presentation and presented.	70% code to be prepared and begin writing the research paper.	Choose reference paper and work accordingly on the paper	06/02/20
06/02/20 Mrs.K.R.Jansi	Prepared 70% of the code and added all basic functionalities	Finish the coding part. Finish the research paper	Shift from video input to a live demo presentation	10/03/20

Faculty of Engineering and Technology

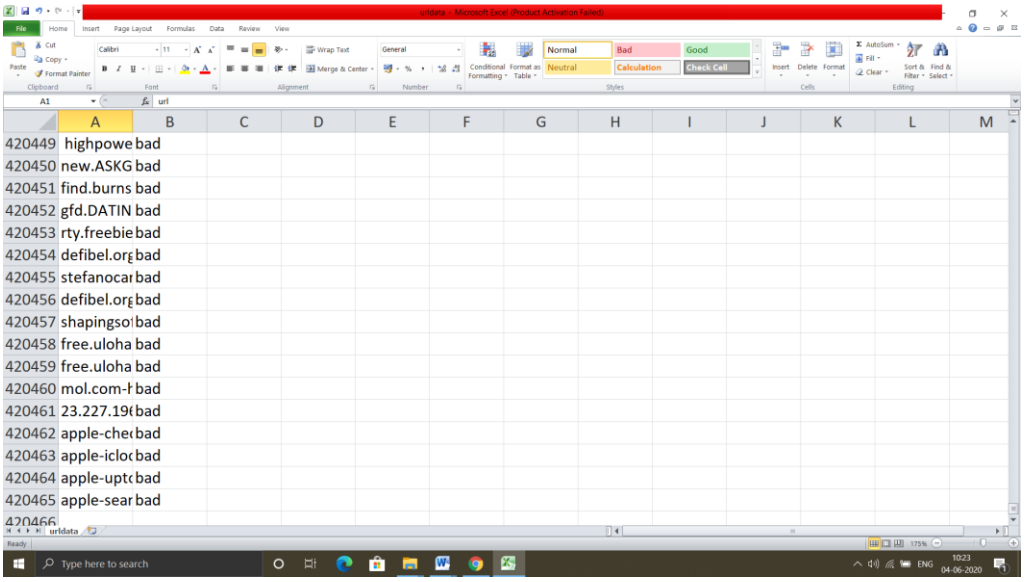
Project Work – Student Log Book

Summary record of major progress meetings with supervisors			**Working title of dissertation/research project**	
Meeting date & supervisors present	**Progress since last meeting**	**Agreed programme of work and target dates**	**Other issues, e.g. facilities, supervision, training needs, etc.**	**Date of next meeting**
10/03/20	Presented the final demo for review 3	Check for plagiarism in the research paper before final submission and make required changes.	Paraphrasing changes required in the paper.	14/04/20
03/05/20	Submitted the paper to the journal for publication.	Status of research paper publication will also be clear by the end of next month.	Conduct meetings through video conferencing.	20/05/20
05/06/20	Completed all the documentation work along with voice over presentation and demo video.	Documents submitted.	Nil.	09/06/20
09/06/20	Project successfully completed and presented in the final review through online mode.	Nil	Nil	-

Faculty of Engineering and Technology

Worksheet / Data collection / Observation etc

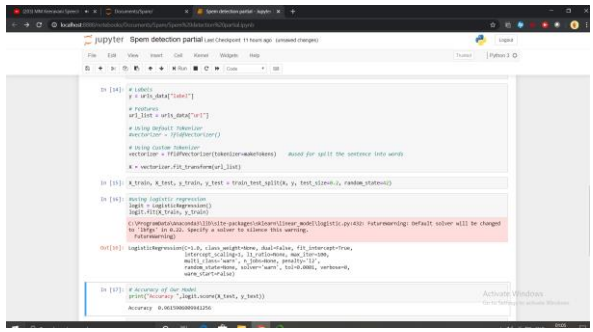
- In order to train we need a data set from which logistic regression can extract features and apply those on the newly given link.
- Those phishing links are taken from WHOIS data base a data base where the phishing links are stored for the future use.
- The data set we are using to train has 4 lakh plus weblinks combinatation of good and bad links.
- All these data links are stored in spread sheet.

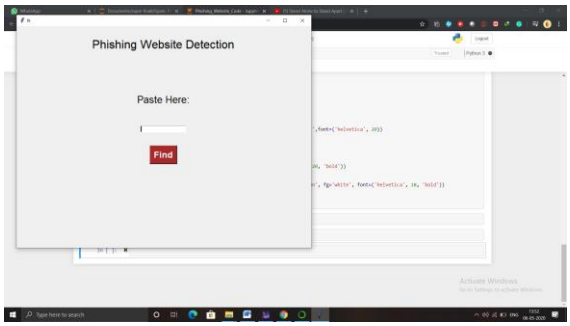


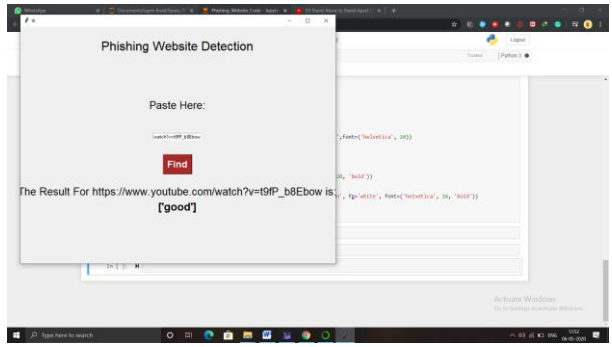
Worksheet / Data collection / Observation etc

LOGISTIC REGRESSION:

- We choose logistic regression over other algorithm because of its speed and the accuracy
- Accuracy is found by dividing the data set into two 80% and 20% .The 80% data is used for the training of the machine learning algorithm.20% is the testing on the data set. In that 20% how accurately the algorithm gives the result gives the accuracy.
- In this system the accuracy is 96%
- This system takes in the input as a link and checks it from the features extracted and gives the output as good or bad.







Journal Publication

Paper - AN EFFECTIVE MODEL OF TERMINATING PHISHING WEBSITES AND DETECTION BASED ON LOGISTIC REGRESSION.

Publication – International Journal of Advanced Science and Technology [IJAST]

Status – Paper publishing is under process.