**Oral**

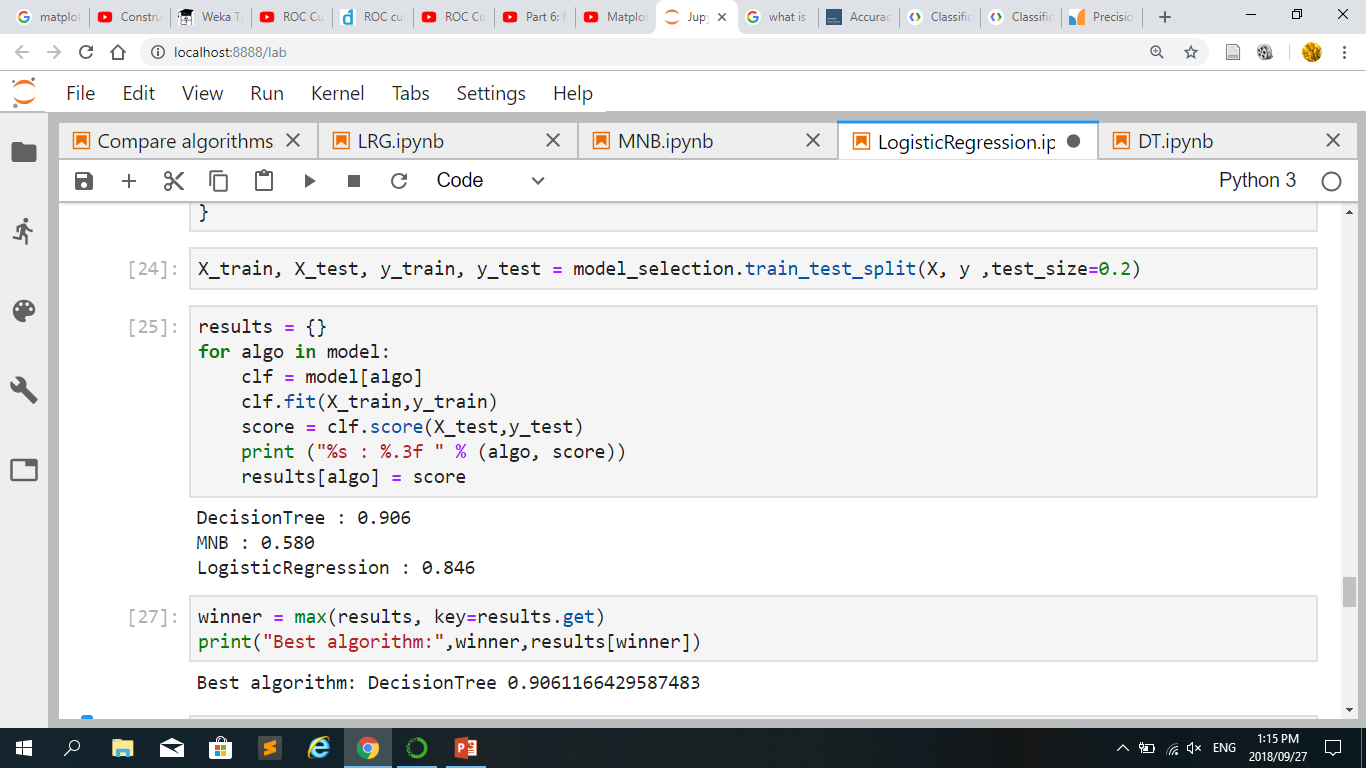
**A prototype of Comparison of the performance of the learning algorithms for verification phishing uniform resource locator (URLs) using machine learning.**

**L.M. Kephas1\*, G.T. Nhinda2, and N. Suresh3**

*1, 2, 3 Department of Information Technology, School of computing, University of Namibia*

Email: [Kephas.loide.m@gmail.com](mailto:Kephas.loide.m@gmail.com); Tel: +264 81 433 9296

Phishing is an online criminal act that occurs when a malicious webpage mimics a legitimate webpage to acquire sensitive information from the user [1]. Detecting phishing websites is one of the crucial problems facing the internet community specifically via emails because of its high impact on the day-to-day online transactions performed. Moreover, there is no doubt that phishing, as a phenomenon, is both highly successful and generally difficult to detect and prevent in a reasonable amount of time [2]. Namibia has experienced its own share in cyber-attacks in the ream of electronic banking transactions which prompt the Namibia government to come up with a draft bill on electronic transactions and cybercrime [3]. Despite a number of solutions to mitigate phishing by previous researchers, there is still no conclusive solution to phishing attacks particularly in the universities environment, and university of Namibia (UNAM) is not an exception. Therefore, this study aims to evaluate the performance of learning algorithms (Naïve Bayesian, Decision tree, and Logistic regression) for verification of phishing URLs using machine learning techniques. Furthermore, the study provide a better understanding on two or more machine learning algorithms that could be used to verify and confirm compromised and phishing URLs in the cyberspace. The study focused mainly on experimental research approach and principle of Personal extreme programming (PXP) development methodology is used for this prototype. PXP is designed to be applied by individual software engineers and is iterative. Applying its practices allows developer to be more flexible and responsive to changes [4]. The experiment is performed using a 7030 URLs dataset, which were divided into two samples: training and testing, 80% for training and 20% for testing and the observed result showed that decision tree provided the best accuracy of 91% as compared to Naïve Bayesian 58% and Logistic regression 85% respectively. Figure1 show the models results.



*Figure 1: The models results*

**Keywords**- Phishing URL, Naïve Bayesian, Decision Tree, Logistic regression, Machine learning, lexical features.

**References**

[1] S. Marchal, J. François, R. State, and T. Engel, “PhishStorm: Detecting Phishing With Streaming Analytics,” *IEEE Trans. Netw. Serv. Manag.*, vol. 11, no. 4, pp. 458–471, Dec.

[2] P. D. Dudhe and P. L. Ramteke, “Detection of Websites Based on Phishing Websites Characteristics,” vol. 3, no. 4, p. 7, 2007.

2014.

[3] N. E. S. Reporter, “Cybercrime a threat to national security – Shanghala,” *New Era Newspaper Namibia*, 15-May-2018.

[4] Y. Dzhurov, I. Krasteva, and S. Ilieva, “Personal Extreme Programming – An Agile Process for Autonomous Developers,” p. 8.