**INSA Cyber talent center**

**Project progress report**

**Title: Malware Detection System Using machine learning**

**Prepared by;**

**Getachew Getu Enyew**

**Submitted to Mr. Dejene**

**Date: July 19, 2022**

**Introduction**

In today’s world, malware has been growing exponentially, causing substantial different asset losses to various organizations. The term malware is malicious software which are designed to meet the harmful intent of a malicious attacker. The program consists viruses, worms, trojans, spyware, bots, rootkits, ransomware. it compromises computers, smart devices, steal confidential information, penetrate networks, and cripple critical. when the computer infected by malicious software, then the file or any asset inside the computer, server or smartphone would be;

* Delete all data on the disc or Change in the file size,
* Damage in the file allocation table, which makes it impossible to read the information on the disk,
* Slowing down the working speed of the computer until it crashes worms. Computer worms are programs with damaging effects which use communication between computers to spread. Worms have similar features with viruses, worms are able spread like viruses in the system, but not locally, but on other computers. It uses computer networks to spread to other systems.

Different anti-malware companies have been proposing solutions to defend attacks from these malwares. Antivirus software such as Norton, McAfee, Avast, Kaspersky, AVG, Bitdefender are a major line of defense for malware attacks. The velocity, volume, and the complexity of malware are posing new challenges to the anti-malware community.

Worlds become inter into more digitalize and intelligent, but in some cases malicious software have a critical negative impact on the development of new eras. To control or avoid assets from attack in 100% is difficult, but some techniques are used to minimize those impact. The concept of machine learning is utilized in the detection of the malwares, various sub-approaches like the ANN, CNN are used for detection purpose.

To successively achieve this project, I will use various machine learning models (like SVM, Naïve Bayes, Neural Networks, ensemble learning) with python programming languages. The algorithm is chosen based on best accuracy score yields the feeding dataset, and based the required problem to be solved.

**Methodology**

Machine learning model is used to predict the class for a given file based on a previously trained model. To analyze data patterns, algorithms must be taught. Multiple phase and tasks are performed to accomplish those tasks;

* collection of the dataset,
* disassembling of executable files,
* feature extraction,
* dimension reduction,
* building classification models, and
* empirical analysis of the results based on different metrics.

**Conclusion**

Machine learning is a good solution to malware problems. Due to the increasing rise of malware, I require automatic solutions to detect infected files. The we teach the machine by feeding the dataset that are constructed in excitable file and run in python scripts. Best algorithms are chosen based on the highest accuracy score for detection malware problems.