COMPUTER SCIENCE AND DATA ANALYTICS

Project Proposal

Paper title: ***Malware Detection using Machine Learning***

Student: Mammadzada Asiman

Instructor: **Dr. Jamaladdin Hasanov**

Dr. Stephen Kaisler

Baku 2023

1. What are we going to do?

The spread of malware poses serious risks to computer networks and systems, resulting in data loss, privacy violations, and monetary losses. Systems for malware detection must be reliable and effective in order to address the constantly changing threat landscape. The objective of this research proposal is to create a machine learning-based automated system for malware analysis and detection. We want to improve the precision and efficiency of malware detection methods by utilizing machine learning.

Heuristics, static analysis methods, and conventional signature-based approaches are still the mainstays of malware detection today. Although these techniques have been somewhat successful, they have certain drawbacks. These methods struggle to keep up with the dynamic nature of malware due to the fast growth of malware and the extensive usage of obfuscation techniques. Due to the high rate of false positives and negatives, detection is ineffective, and security concerns are raised.

The suggested strategy makes use of machine learning algorithms in an effort to overcome the shortcomings of the most recent malware detection methods. We can create a dynamic and adaptable system that can identify patterns and behaviors suggestive of malicious software by training models on vast datasets of both known malware samples and lawful software. The machine learning models will be taught to recognize typical traits associated with malware, including code structures, system calls, network traffic patterns, and behavioral abnormalities. This strategy may improve detection precision, cut down on false positives, and adapt to new malware threats.

2. How is it done today? Current Limitations

Currently, malware detection depends on static analysis methods, heuristics, and conventional signature-based approaches. These techniques frequently experience a large number of false positives and negatives as a result of their inability to keep up with the continually changing malware field. Advanced malware strains also use obfuscation tactics to avoid detection.

3. What is your idea to do something better?

The proposal involves leveraging machine learning algorithms to create a dynamic and adaptive system for malware detection. By training models on large datasets of both known malware samples and legitimate software, the system can learn to recognize patterns and behaviors indicative of malicious software. This approach has the potential to enhance detection accuracy, reduce false positives, and adapt to emerging malware threats.

4. Who will benefit from your work? Why?

My study has potential benefits for many stakeholders. Organizations and security experts will get access to more powerful tools for detecting and reducing malware risks. End users will benefit

from enhanced malware defenses that protect their private information and sensitive data. The improvements in the field of machine learning-based malware detection will also assist the research community.

5. What risks do you anticipate?

False positives, when benign software is mistakenly labeled as malware, and false negatives, where malware is found but not stopped, are two possible concerns. There could also be difficulties because of the complexity and variety of malware samples, the necessity for reliable training data, and the possibility of adversarial assaults directed at the machine learning models themselves.

6. Out of pocket costs? Complete within 11 weeks?

Obtaining or gaining access to pertinent datasets, setting up computational resources for training and testing machine learning models, and any fees associated with presenting or publishing research results might all be considered out-of-pocket expenses for this study. The extent and complexity of the project, the accessibility of resources, and the research style used will all affect whether this study can be finished in 11 weeks.