Domain: Model for AI Security
Content Topic: Anti-Threat Framework

Abstract

The heart of Artificial Intelligence is data. Everything in AI is data. Suppose if the data itself corrupted, it could lead to serious damage to the system and to the environment. An attacker could make the data invalid, that could mistake the AI to change the behaviour of it. In order to avoid this kind of attack, the AI must be trained to detect these kinds of inputs along with its required dataset for proper functioning. Security of an Intelligent Systems is that the most vital factor for single users and businesses depending on it. The increasing range of the cyber-attacks, anti-virus scanners cannot fulfil the necessity for cover. Hence, the skill level has to be increased to identify and prevent in the event of cyber threats. Therefore, the availability of the attacking tools on the web, may attack the Artificial Intelligence system, where the existing tools such as firewalls, Honeypot may not detect such kind of threats and vulnerability present in the system. The Recurrent Neural Networks (RNN) detects the vulnerabilities and the artificial intelligence-based generative models do the prevention process and improves reliability. We will implement a Time Series function along with its supporting models called ARCH/GARCH which can be used to predict the outcome of AI model ahead of one period which could make the fatal error and misbehaviour of AI. The invalid data will be corrected by the trained RNN model and if it fails to convert, an alert will be triggered.

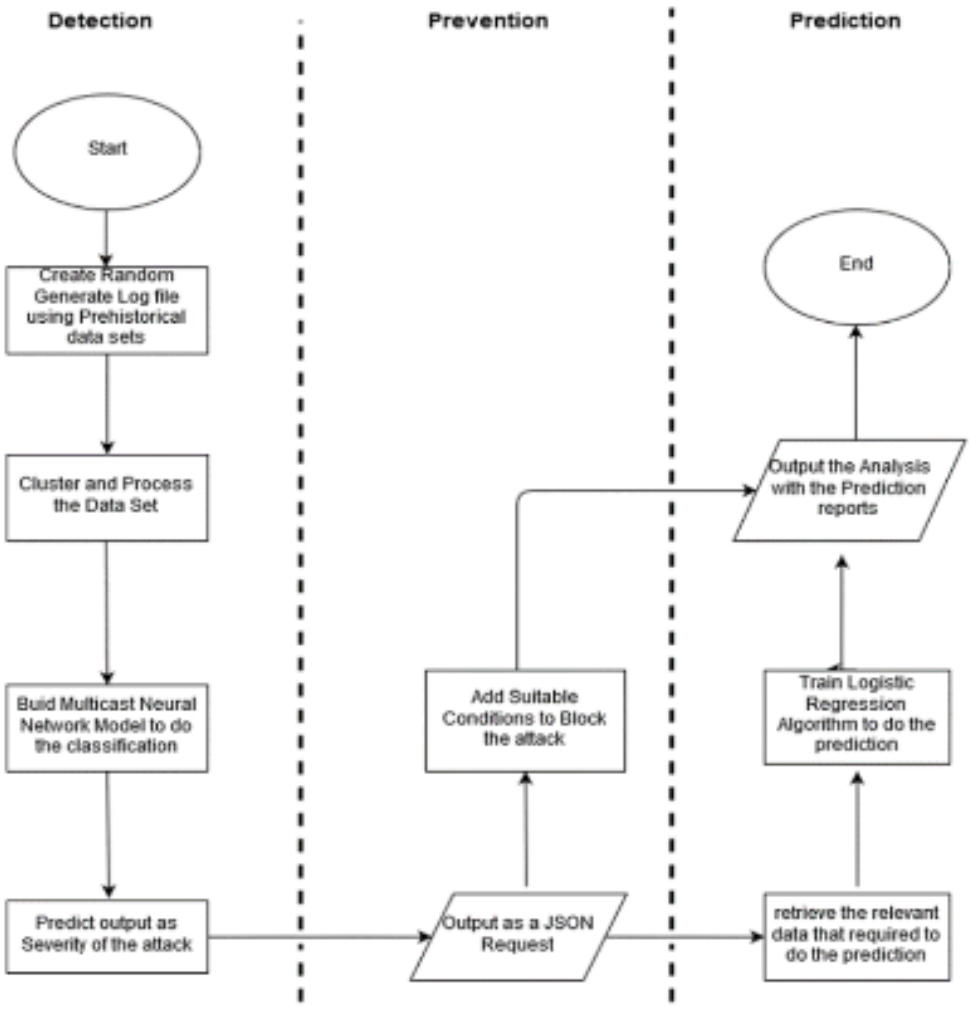
Keywords— cyber-attacks; cyber-attack prediction; data mining; AI attacks; incident response; attack mitigation;

Concept

The heart of Artificial Intelligence is data. Everything in AI is data. Suppose if the data itself corrupted, it could lead to serious damage to the system and to the environment. An attacker could make the data invalid, that could mistake the AI to change the behaviour of it. In order to avoid this kind attack, the AI must be trained to detect these kinds of inputs along with its required dataset for proper functioning. With this pre-defined training we also need to make separate process that runs along with AI to detect the various invalid inputs in such case the AI is failed to detect it. For that purpose, we will implement a Time Series function along with its supporting models called ARCH/GARCH which can be used to predict the outcome of AI model ahead of one period which could make the fatal error and misbehaviour of AI.

Existing Methodology

In Existing Methodology, the researchers divided the process into four phases such that in each phases *data collection*, *detection*, *prevention and prediction are* made by the model during execution.



Proposed Methodology

- In our proposed Methodology the *data collection* takes place as like existing one, but the data is pre-processed for both the AI system and the framework which working along the AI system.
- While the AI system working on the data, the framework model will read the logs and compare it with current input to check whether it is invalid or not.
- Then the framework runs Recurrent Neural Network on the new input it received for the input validation and sends it to the Time Series Function.
- The Time Series Function will predict the output of the new input in which the AI system should produce and if it is predicted that it is incorrect for the system it will directly terminate the process.

- The Recurrent Neural Network can send it values for AI system if it has invalid input for normal behavior of the AI system.
- In order to avoid further unresponsiveness or bottlenecks, if the input is too large or many inputs come at same time the entire system can adapt by running the framework and later the AI system.
- By running the AI system and the framework model parallel, it could save the time as well as if any one fails to produce the output, the other one will take care.

Proposed Methodology Flowchart