***Malware detection using Machine Learning/Deep Learning and create recommendation engine***

Malware is a worldwide pandemic. It is designed to damage computer systems without the knowledge of the owner using the system. Software’s from reputable vendors also contain malicious code that affects the system or leaks information’s to remote servers. Malware includes computer viruses, spyware, dishonest ad-ware, rootkits, Trojans, etc.

Malicious software disrupts IT and computer processes and in extreme cases can delete,

steal or complete breakdown of the corporate network or the loss of business critical data. Following can be the impacts of malware on business:

 Attack sites and disable services

 Identity Theft/Identity Spoofing

 Affect network performance or complete breakdown of the corporate network

 Steal sensitive information

 Control over the applications running in your systems

 Hardware Failure

According to security firm Kaspersky Lab reports, cybercriminals do not steal just data. They also stole up to $1 billion from 100 different financial institutions across the U.S., Germany, Russia, Ukraine, and China over the past two years.

[According to Coveware Q2 Ransomware Marketplace Report “](https://www.coveware.com/blog/2019/7/15/ransomware-amounts-rise-3x-in-q2-as-ryuk-amp-sodinokibi-spread)Ransomware Attacks Costs

Nearly Triple in 2019 to over $36K per Attack”.

According to the FBI's Internet Crime Complaint Center, ransomware--malicious programs that infect a computer or network and hold data hostage until a ransom is paid--has cost companies $18 million in the past 15 months.

The rapid rise of the Internet and the ensuing growth in malware meant that manually created detection rules were no longer practical and new advanced protection technologies are needed. Today, due to constantly evolving malware, **signature-based antivirus** software is no longer an effective solution that’s why Anti-malware companies turned to machine learning, machine learning train computers to recognize and differentiate between benign and malicious files. They teach the machine what’s good and what’s bad, so that eventually, the machine can sort the files on its own. Today, machine learning boosts malware

detection using various kinds of data on host, network and cloud-based anti-malware components.

Design a solution to detect malware using Machine learning / deep learning on given dataset and create recommendation engine.

1. Link for the dataset - <https://github.com/fabriciojoc/brazilian-malware-dataset>

(download and extract zip file)

 Dataset is with label, 0 for goodware and 1 for malware, remove the particular row if the ‘label’ column has value other than 0 and 1, and given

dataset contains features extracted from PE (Portable executable) files. E.g.

.exe, .dll

 Drop this columns “FirstSeenDate, Identify, PE\_TYPE, SHA1 “

o **Note:** Imported symbol also known as **imported function/apis** from dll

2. Create a classification model to detect malware using any ML/DL algorithm.

3. Create a command line interface based recommendation engine, provide Scoring:

 Low (0-0.3)

 Medium (0.4 to 0.6)

 High (0.7 to 1)

**Note:** Don’t get too much into to understand PE file format, focus on Data pre-processing and normalizing.

Useful links:

[https://docs.microsoft.com/en-us/windows/win32/debug/pe-format (To understand features)](https://docs.microsoft.com/en-us/windows/win32/debug/pe-format)

<https://arxiv.org/pdf/1804.04637.pdf>

<https://towardsdatascience.com/nlp-for-beginners-cleaning-preprocessing-text-data-ae8e306bef0f>