Trade Analysis and Prediction in Dark web

[1]M. Eliazer, [2]Shruti Gupta, [3]Bharti Gupta

Abstract

The non-referenced web is estimated five hundred times the width of the web's surface. Dark web accounts for about a few percentage of the anonymous web and contains all types of criminal activities: drug trafficking, counterfeiting, hacking, etc. Dark Net Markets (DNM) are e-markets usually hosted as Tor hidden networks offering escrow services between buyers and sellers trading in Bit coin or other crypt currencies, usually for marketing of drugs or related illegal/regulated goods; Agora was one of the most common DNMs. Agora was a dark net market that hosts on the Tor network itself, launched in 2013 and due to some reasons it was shut down in August 2015. In this paper we propose analyzing and predicting the trade in a particular Dark Web Market (Agora) using neural networks.

Keywords: dark net, dark net markets, illegal trading, cryptocurrency, Agora

I. Introduction

Darkweb Marketplaces generally host as online markets on Tor Hidden Services, normally providing illegal trading in goods. Identifying and evaluating goods that are bought and distributed on such sites can have significant inferences, especially for applications related to threat intelligence and security.The votality of the dark net makes it so difficult to study. It is huge and really vast as compared to the normal internet. Adequate assessment of such systems can help us to identify systems that are at-risk and can help us in identification of possible future cyber threats. Websites specializing in the selling of malicious hacking goods on the deep-web and darkweb have become the place of choice for cyber criminals to trade these items online. In this project we aim to analyze and predict the trades in Dark Net markets. The Analysis and prediction will be done using neural networks. We are using the data-set of Agora Dark Web available on Kaggle and present an analysis of trades present in it. In the Agora data-set, we will analyze and predict the data according to the columns named - Item Name and Item description to identify the category

[1] *Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India.*
[2] *Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India.*

they belong to and then compare the results to the expected results and predict the accuracy of the system.

1.1. Data Mining

The method of finding out the patterns in large amount of data available to us and finding out useful insights from data mining process. It is a field of Computer Science which helps us in modifying Machine Learning, Statistics and Database Systems available to us. This process has various steps to be followed in order to obtain proper and useful insights from the data being analyzed and mined. The data is needed to be pre-processed in order to remove useless and useless data which may lead to production of false and incorrect results.The process of Data Mining is usually followed in order to cut costs, increase the profit of the companies, upgrade relationships with customers. It is basically used to convert the raw, useless data into a data which can be used in order to produce a specific outcome. In order to produce outcomes, several steps are to be followed such as, first, the data is pre-processed, then analyze and predicted and visualized and hence results can be collected.

1.2. Dark Web

Dark Web is a term usually used to reference the encrypted part of the internet which is not accessible by normal users. 'Dark Net' is another term which can be used in order to reference to the Dark Web. The dark web can not be accessed using our conventional browsers but a special browser or a network called the 'tor'. Almost all such sites hide their identity by using the encryption protocol. Not all websites on the network use the same protocol, some also use the I2P. The Dark Web is almost like the internet available to us containing similar services except for the fact that we can not find out who are behind these websites and services i.e they are anonymous. The are some websites like Silkroad, Agora, etc which host as e-commerce websites on the dark web similar to websites like flipkart, Amazon, etc on the internet. Most of the illegal trade and hacking is done using the dark web because the data is encrypted and un-traceable and hence more secure in order to protect the identity. On the dark net a person can get all the services which he might not get on a local general store like credit card numbers, guns, hackers, assasins, etc

II. Literature Survey

Othmane Cherqi et all [1] proposed a system for Exploratory analysis of the illegal trade which develops in this Darknet marketplace using algorithms such as Logistic Regression, , Linear Support Vector Machine, Naive Bayes, Random Forest.

3 Student, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India.

Eric Nunes et all [2] proposed a system for Analysing Forums using DeLP structured argumentation system to identifying the systems that maybe at-risk based on discussions availble on-site using the algorithms such as Decision tree, Logistic Regression, Naive Bayes Classifier, Random Forest, Support Vector Machine.

Ericsson Marin et all [3] proposed a system for examination of malware products from 17 malicious hacker markets using the techniques such as Manual Labeling and unsupervised clustering, N-grams, K-means algorithm.

Mhd Wesam Al Nabki et all[4] have proposed a system called ' Darknet Usage Text Addresses ' (DUTA) which has been made publicly available. The pipeline elements were fixed and the aspects were identified which had a critical impact on the classification. This was done using the algorithms such as TF-IDF, Support Vector Machine.

Andres Baravalle et all [5] proposed a method and work was performed to gain information about the products and services offered on a larger market using techniques such as web scraping, data tainting and turing checking.

Massimo La Morgia et all [6] proposed a system to uncover the geolocation of the crowds accessing the dark web and group them time-zone wise using the techniques such as Gaussian Distriution, naive bayes, web scraping.

Robert Koch [7] proposed an analytical framework that could be used to determine guidelines on how to mitigate the risk. This was done using the program known as the Memex framework of DARPA.

Faizan et all [8] proposed a system to analyse and categorize the data into 31 different categories. This was done using a focused Web Crawler.

Eric Nunes et all [9] proposed a system that was developed for collecting information from the sites which can be used for identification of emerging cyber threats. The techniques used were Web crawler, Parser, Naive Bayes, Logistic Regression, Random Forest.

Ericsson Marin et all [10] proposed a system that discovered user posting rules through user posts sequences, to use those rules later in the near future for predictions. The algorithms used were Sequential Mining, T-Rule Growth.

III.   Proposed Work

This system will help us account the types of trade that are prevailing in the dark web such as Selling of Drugs, Forgeries, Arms Supply, Services, etc.using the item description available int the data set. The data set available will be pre cleaned and then analyzed later. The new clean data will analyzed using a Data Mining algorithm called - The neural Network.

IV.   Implementation

The system consists of an application which will analyse the data set that is available to us I.e Agora Marketplace. It will then depict and help in further understanding according to the data available to us. So far only a limited amount or a specific category of trade has been analyzed. The whole or the types of trade that exist in the dark web are not analyzed. In our system we analyze all the types of trades that exist in agora market place and then use the knowledge for prediction purposes. The system is proposed to have two modules.

4.1 Data Pre-processing

It helps in handling raw data and converting it into useful and efficient format.

4.1.1 Data Cleaning

The data available may be irrelevant or contain missing parts. To handle the inaccuracy, data cleaning is done.

4.1.2 Data Transformation

Some steps are followed in order to transform the data into useful values. The methods being used are normalization, Stemming, Lemmatization, etc.

4.1.3 Data Reduction

Non required values of data are removed from the data set making it more reliable and useful. After the data is cleaned, the cleaned data is saved in a new file and is ready to use for further prediction.

4.2 Analysis and Prediction

The data Analysis and Prediction is done using a Data Mining algorithm known as The Neural Networks.

4.2.1 Neural Network

A Neural Network is a mathematical model inspired by the neural-biological network. Simple artificial nodes, called "neurons," "processing components" or "modules" are linked together in an artificial neural network to form a network that behaves like a neural-biological network. This includes a network of simple elements that pose complex global action dictated by the relations between the processing elements and parameters of the elements. Models of neural networks are part of theoretical neuroscience and computational neuroscience. For example, we use neural networks for predictions of the time series, data detection of anomalies and comprehension of natural languages. Neural networks in Data Science are useful in providing information on associations, classifications, clusters, and forecasting.

In the system, Neural Networks are used to analyse and create connection between the data provided the to the network during the training period. When useful and proper connections are created by the system it is then further used to make predictions about the test data.
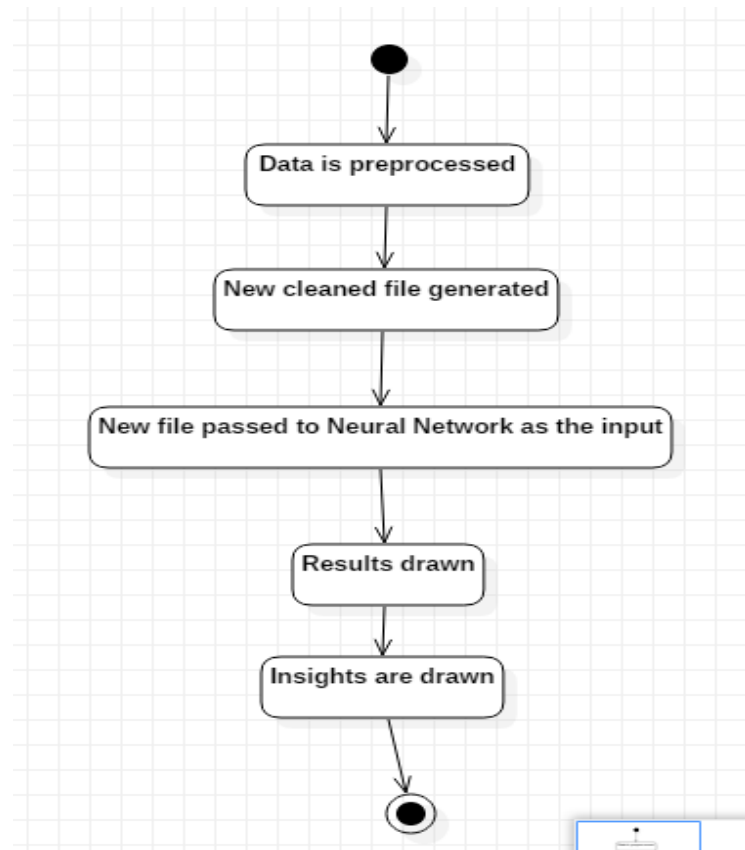
V. Performance

The System Provides the output for the given data set with 94% accuracy.

VI. Future Scope

The non-referenced nature of dark net market has made it very difficult to keep a track of the illegal activities that take place on the Tor Network. The application developed can be further used to analyze data sets of other darknet markets that re surface online. It can also be used to identify potential cyber security threats.

VII.   Conclusion

A large number of illegal trade takes place on the dark web which is very difficult to track and analyze. This system will help to analyze the extracted data from the dark net markets and help us understand the depth and kind of illegal activities that take place on the network efficiently. It efficiently provides predictions for the items that we may not be able to categorize and identify individually.

References

[1] Othmane Cherqi, Ghita Mezzour, Mounir Ghogho , Mohammed El Koutbi" Analysis of Hacking Related Trade in the Darkweb" Published in 2018 IEEE International Conference on Intelligence and Security Informatics (ISI).

[2] Eric Nunes, Paulo Shakarian,Gerardo I. Simari "At-Risk System Identification via Analysis of Discussions on the Darkweb" Published in 2018 APWG Symposium on Electronic Crime Research (eCrime).

[3] Ericsson Marin, Ahmad Diab and Paulo Shakarian "Product Offerings in Malicious Hacker Markets" published in 2016 IEEE Conference on Intelligence and Security Informatics (ISI).

[4] Mhd Wesam Al Nabki, Eduardo Fidalgo, Enrique Alegre, and Ivan de Paz "Classifying Illegal Activities on Tor Network Based on Web Textual Contents" published in 2017 Association for Computational Linguistics

[5] Andres Baravalle, Mauro Sanchez Lopez, Sin Wee Lee "Mining the Dark Web: Drugs and Fake Ids" published in 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)

[6]Massimo La Morgia, Alessandro Mei, Simone Raponi, Julinda Stefa "Time-Zone Geolocation of Crowds in the Dark Web" published in 2018 IEEE 38th International Conference on Distributed Computing Systems

[7] Robert Koch "Hidden in the Shadow: The Dark Web – A Growing Risk for Military Operations?" publishded in 2019 11th International Conference on Cyber Conflict

[8] Faizan and Raees Ahmad Khan "Exploring and analyzing the dark Web: A new alchemy" published in 2018

[9] Eric Nunes, Ahmad Diab, Andrew Gunn, Ericsson Marin , Vineet Mishra, Vivin Paliath, John Robertson, Jana Shakarian, Amanda Thart, Paulo Shakarian "Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence" Published in 2016 IEEE Conference on Intelligence and Security Informatics (ISI).

[10] Ericsson Marin, Mohammed Almukaynizi, Eric Nunes, Jana Shakarian, Paulo Shakarian "Predicting Hacker Adoption on Darkweb Forums" published in 2018 IEEE Intl Conf on Parallel & Distributed Processing.