ANAMOLY DETECTION USING CNN ALGORITHM

Muralidhara S,

Assistant Professor,
Computer Science and Engineering,
NHCE, Bengaluru, India

Abstract: Anomaly Detection on Facebook images plays a key role on identification of anomalous or behavioral analysis of Facebook. It targets detection of quality, violence, perfection of the images of the Facebook. Using specific Artificial Neural network algorithm of the machine learning detection of the anomalous images is performed by the project. Convolution neural network algorithm with 5 layers has been selected in this project to detect some of the anomalous images from a large dataset of the images downloaded from the Facebook. The total dataset of the images is divided into 3 parts such as train dataset, validation dataset, and Test dataset. The accuracy of the algorithm depends on the volume of the training dataset. The larger training dataset gives more accurate results compare to the smaller dataset. Kernel matrix has been formed from the RGB pixel of the images downloaded from Facebook. Each image have kernel matrix. From the kernel matrix algorithm is being trained. The objective of the project is to detect quality; violence and perfection of downloaded images from the Facebook which intern detect the activities and behavior of the user. Facebook has its own immune system to safeguard its users from unwanted malicious content. Despite the immune system deployed by Facebook, unwanted spam, phishing, and other malicious content continues to exist on Facebook. Objective is to detection of malicious content on Facebook .The goal is to decide for each user whether the account is compromised or not.

IndexTerms **– Artificial neural Network, Machine Learning, Social Networks**

I. INTRODUCTION

A system that monitors network movement for suspicious activity and generates alerts such activity is called intrusion detection system (IDS). Anomaly detection and reporting is the primary function of intrusion detection systems and capable of taking actions when malicious activity or anomalous traffic is detected. Security mechanism of a system is designed to prevent unauthorized access to system resources and data. To detect the intrusion attempts that action taken to repair the damage .This field is called as intrusion detection. Intrusion detection systems derived in different essences and detect doubtful activities using different methods, including the following:

A network intrusion detection system (NIDS) is run within network to monitor inbound and outbound traffic to and from all the devices on the network.

A host intrusion detection system (HIDS) is run on all computers or devices in the network with direct access to both the internet and the enterprise internal network. HIDS is also able to identify malicious traffic that originates from the host itself. When the host has been infested with malware then it attempt to spread to other systems.

Signature-based intrusion detection systems monitor the packets traversing the network and compare them against a database of signatures or attributes of known malicious threats.

Anomaly-based intrusion detection systems monitor network traffic and compare it against a recognized baseline, to determine which is considered normal for the network with respect to bandwidth, protocols, ports and other devices. This type of IDS alerts administrators to potentially malicious activity.

II. ANOMALY DETECTION

Anomaly detection is the identification of data points, items, observations or events that do not conform to the expected pattern of a given group. These anomalies occur very infrequently but may signify a large and significant threat such as cyber intrusions or fraud. Anomaly detection is mainly a data-mining process and is used to determine the types of anomalies occurring in a given data set and to determine details about their occurrences.

Anomalies or outliers come in three types.

1. Point Anomalies: If an individual data instance can be considered as anomalous with respect to the rest of the data (e.g. purchase with large transaction value)

2. Contextual Anomalies: If a data instance is anomalous in a specific context but not otherwise if occur at a certain time or a certain region. e.g. large spike at the middle of the night). These are calculated by focusing on the segments of data and applying collective anomaly techniques within each segment independently.

3. Collective Anomalies. If a collection of related data instances is anomalous with respect to the entire dataset, but not individual values. They have two variations.
1. Events in unexpected order (ordered. e.g. breaking rhythm in ECG)

2. Unexpected value combinations (unordered. e.g. buying a large number of expensive items).

SUPERVISED MACHINE LEARNING FOR ANOMALY DETECTION

This method requires a labeled training set that contains both normal and anomalous samples for constructing the predictive model. Theoretically, supervised methods are believed to provide better detection rate than unsupervised methods. The most common supervised algorithms are supervised neural networks, parameterization of training model, support vector machine learning, k-nearest neighbors, Bayesian networks and decision trees. K-nearest neighbor (k-NN) is one of the most conventional nonparametric techniques that are used in supervised learning for anomaly detection. It calculates the approximate distances between different points on the input vectors and then assigns the unlabeled point to the class of its K-nearest neighbors. The Bayesian network is another popular model that can encode probabilistic relationships among variables interest. This technique is generally used for anomaly detection in combination with statistical schemes. These supervised techniques have several advantages, including the capability of encoding interdependencies between variables and of predicting events, along with the ability to incorporate both prior knowledge and data.

UNSUPERVISED MACHINE LEARNING FOR ANOMALY DETECTION

These techniques do not require training data. They are based on two basic assumptions. First, they presume that most of the network connections are normal traffic and only a small amount of percentage is abnormal. Second, they anticipate that malicious traffic is statistically different from normal traffic. Based on these two assumptions, data groups of similar instances that appear frequently are assumed to be normal traffic and those data groups that are infrequent are considered to be malicious. The most common unsupervised algorithms are self-organizing maps (SOM), K-means, C-means, expectation-maximization meta-algorithm (EM), adaptive resonance theory (ART), and one-class support vector machine. One popular technique is the self-organizing map (SOM). The main objective of the SOM is to reduce the dimension of data visualization.

III. LITERATURE SURVEY

Data Mining Approach for Anomaly Detection in Social Network Analysis [1] has inputs such as huge data from single database.

It has the methodology such as 1. A dataset is a collection of data corresponds to the content of single database table.

2. From the dataset remove the null values.

3. Risk assessment approach is based on the idea of estimating the user's risk on the basis of how much his\her behavior deviates from the normal user. Risk assessment is composed of two phase clustering: The first phase consists of organizing the users according to group identification features. b. In the second phase, users are categorized using behavioral features using K-means and Expectation Maximization algorithm.

4. The risk score is estimated using Membership probability based on a value of group identification features. a. High membership probability-normal user. b. Low membership probability-abnormal user. It helps to detect risk to the all OSN user the principle of outlier detection.

Security Threats and Defensive Techniques of Machine Learning: A Data Driven View [2] presented a systematic survey on security concerns with a variety of machine learning techniques, Reactive defense and proactive defense. It uses large volume of data for security threat taxonomy to detect anomalous activities such as evasion, Impersonate, Inversion attack. Result of these systems such as revisited existing security threats towards machine learning from two aspects, the training phase and the testing/inferring phase. The training phase, those in the testing or inferring phase, data security and privacy.

Detection of Malicious Applications on Facebook using Machine Learning Algorithm [3] is possible .It has taken input from user login fields Facebook as developer account Project Application ID Token of per user Web services of User information. SVM Algorithm as a classifier is used to detect the anomalous activities such as common behavioral and structural features of the Facebook data. The result of the anomaly detection of malicious applications successfully and adding it to our app, post warning on users wall. Hide and prevent the sharing of bad posts and comments, to determine whether the image is appropriate or not.

Early Detection and Classification of Bearing Faults using Support Vector Machine Algorithm [4]. Fault and health present in the bearing is determined by the using machine learning algorithms. A huge amount of dataset is used to detect the health of the bearing .A classifier algorithm such as SVM is applied to a large data set to detect fault and health of bearing.

Facebook Inspector (FbI): Towards Automatic Real Time Detection of Malicious Content on Facebook [5] will detect anomalous activities of the Facebook on real time. It contains input sources of malicious content, legitimate content with URLs, and all legitimate content. Mobile platforms were preferred over web for posting legitimate content. It contains a large volume of dataset such as 4.4 million public Facebook posts generated by over 3.3 million unique users and pages. Taking large volume of dataset into account real-time anomalous activities of the content can be determined. Methodologies such as phishing, advertising campaigns, content originating from compromised profiles, artificial reputation gained through fake likes are the sources where the anomalous activities are detected. The outcome of the analysis is the anomalous posts in Facebook.

IV. METHODOLOGY

ANNs have the ability to learn and model non-linear and complex relationships, which is really important because in real-life, many of the relationships between inputs and outputs are non-linear as well as complex. After learning from the initial inputs and their relationships, it can infer unseen relationships on unseen data as well, thus making the model generalize and predict on unseen data.

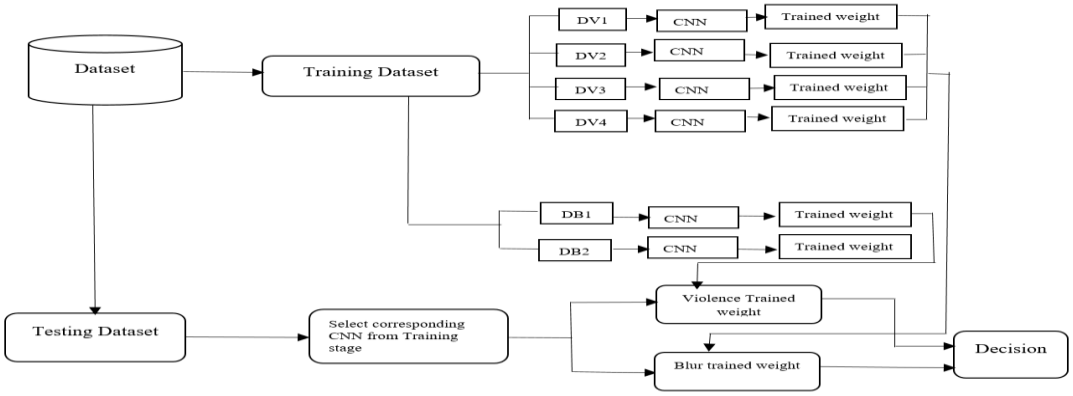

Fig 1: Methodology of Proposed System

In first step data being collected from facebook to form dataset to be analyzed by the proposed system. The data is being collected from Facebook using Facebook graph API after creating a developer account in Facebook. The graph API is the primary way to get data out off and to put data into the Facebook platform .To query data programmatically, upload photos, post new stories in Facebook low level http-based API is used in graph API. Information in Facebook composed of three parts such as node, edges, and fields. The Facebook graph API allowed developers to access pretty much same data as regular user.

In second step of the processing consists of cleaning of data. In this system to pre-process data tools are not used. Manually data is being cleaned and put into the respective dataset. The formation of the dataset is the important step in case of machine learning development. Dataset lead the proposed system performing accurate result according to the requirement. So cleaning of data to get perfect dataset is necessary process in machine learning development.

In the third step of the process the implementation of the Convolutional neural network algorithm is applied to the dataset to find out the violence image and blurred image. In this process the datasets is stored in list of image. Input layer is taking the dependent variable, after undergone by four layer processing such as ReLu, pooling, and convolution ReLu, and pooling layer .The output of the pooling layer is connected to the Full convolution layer. Input variable x is input through the input layer of the convolution neural network. The output layer contains the dependent variable y as an output. ReLu stands for Rectified Linear Unit for a non-linear operation f(x) =max (0, x).performance of the ReLu is better than tanh and sigmoid. Pooling layer: Pooling layer reduces the number of parameter when the image is too large. There are variety type of pooling such as max pooling, average pooling and sum pooling. Fully Connected layer: Fully connected layer is the layer where matrix flattened into vector and feed into neural network. The output of the fully control model is transferred to the network. The five layered model of convolution neural network is applied to the list of data .According to the model the algorithm trained .Accuracy, f1 score, Recall and precision for the development find out.

V. CONCLUSION

Anomaly detection of the Facebook images using convolutional neural network (CNN) is proposed. CNN algorithm is an efficient algorithm for anomaly detection on images as compare to other machine learning algorithms. Anomalous activities of the Facebook can be determined by the anomalous images. CNN algorithm has a kernel which significantly smaller than the input and simplifies the number of computation required to train the model. This algorithm focuses to the relevant features of the input image for which it require fewer parameters. Weight sharing is another important feature of the CNN algorithm which helps in reducing memory as compared to other algorithms. The datasets are formed to avoid outfitting on CNN algorithm. There are two data sets such as violence dataset to detect violence activities and blur dataset to detect quality of images. The violence dataset detect for images with axe, gun, hammer and knife which intern detect the violence activities. The blur image dataset compares original image quality with blur image which intern detect blur image.

REFERENCES

[1] Dr. Valarmathi K, "Data Mining Approach for Anomaly Detection in Social Network Analysis", ICICCT 2018.

[2] QIANG LIU1, PAN LI1, WENTAO ZHAO1, WEI CAI2, SHUI YU3, VICTOR C. M. LEUNG2, (Fellow, IEEE), "Can A Survey on Security Threats and Defensive Techniques of Machine Learning: A Data Driven View" Bitcoin Magazine, 2016.

[3] Kibin Immadisetti Naga Venkata Durga Naveen, Manamohana K, Rohit Verma, "Detection of Malicious URLs using Machine Learning Techniques ", March, 2019.

[4] Jagath Sri Lal Senanayaka, Surya Teja Kandukuri, Huynh Van Khang, Kjell G. Robbersmyr, "Early Detection and Classification of Bearing Faults using Support Vector Machine Algorithm". 2017.

[5] Prateek Dewan, Ponnurangam Kumaraguru, "Facebook Inspector (FbI): Towards automatic real-time detection of malicious content on Facebook",2017.