Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Detecting Insider Threat through Psychometric Scores and Work Environment

Manas Kumar Yogi[1], Yamuna Mundru[2]

Assistant Professor , Computer Science and Engineering Department, Pragati Engineering College (Autonomous), Surampalem, East Godavari District, Andhra Pradesh, India

E-mail: [1]manas.yogi@gmail.com, [2]yamuna.lakkamsani@gmail.com

Abstract

Nowadays the cyber-threat is looming large from disgruntled employees rather than external attackers. With the advent of modern threat models which help in development of adversary attack scenarios, the security designers can get an idea of attack mitigation strategies. The analysis of threat model helps in knowing the unforeseen circumstances which can lead to a cyber-security risk. If the weakness in the network or in any other element of a cyber-ecosystem is identified during the design of security models, then eventually it will lead to stronger security applications. This paper formulates a framework of identification of insider threat which may be a result of various factors in an organization. The proposed technique considers psychometric condition of an employee along with other factors which may give rise as a threat to the organization. The dataset has been used to train the model and the experimental results have found an effective way to detect the insider threat in specific cases.

Keywords: CERT, SVM, decision trees, threat, cyber security

1. Introduction

An insider threat is defined as an attack perpetrated by a user or malicious code that is already present in defending the perimeter of a system or an organization. An insider could be a current employee, former employee, a contractor or a business partner, anyone with legitimate access to the company's network, databases and applications. Insider threats are dangerous because often these attackers know, how the system is configured and its weaknesses. An external attacker has to breach the firewall and then get into the system, whereas for an internal attacker all that information is known which makes insider attacks hard to defend against.

DOI: https://doi.org/10.36548/jscp.2022.3.008
Received: 21.08.2022, received in revised form: 17.09.2022, accepted: 02.10.2022, published: 12.10.2022

In a popular survey related to cyber-attack incidents, it was found that 64% of the threats were due to negligence; 23% were pertained to insider attacks, and 13% emerged from theft of credentials.

1.1 Motivation behind an insider attack

There are many reasons for an insider attack, and a few reasons to show how these attacks might happen has been briefed in this work. Employees sometimes fall for cyber-attacks like phishing that compromises their safety along with the company's. Another motivation could be a financial gain; a company might be paying the employee to leak the information. Moreover, it can also be for the thrill or curiosity, or it can be some disgruntled employee trying to take revenge on the company. There can also be causes that are bigger like terrorism or political or ideological views. Another motivation would be for a personal gain or to satisfy their ego. There can be various reasons for which an employee would try to attack their own company.

1.2 Insider threat damage and impact

The biggest threat and impact an insider threat has on a company is that, the company loses trust among its customers and negative reports on the media will tarnish the image of the company making the company lose business and clients. Insider threats are often coupled with external attacks which would be used to steal intellectual property or for a ransom ware attack which would also cost the company a lot of money and reputation. Insider threats are very hard to detect and before identifying the insider a lot of damage is done. There have been instances of the attacker being active in the company, years before they were detected. The scope of an insider attack is enormous as the company is interconnected and a significant amount of damage to the company can be done through any department. The workplace would also be toxic if the employees know about the insider attacks, because they would suspect their own co-workers in the workplace.

1.3 Concept of insider threat detection

Insider threat works on detecting a potential attacker based on a few characteristics. There can be many features that can be used such as network activity, activity logs, biometric data such as pulse, temperature etc. Based on these features, different machine learning techniques either supervised or unsupervised are used to predict if a person is going to be a potential attacker.

1.4 Insider threat classification

Insider threat can be classified as intentional as well as unintentional. Unintentional insider threats arise due to negligence or accidental aspects. Intentional threats are taken in consideration with collusive threats which arise when an insider collaborates with an external malicious entity who has a strong intention of a cyber attack. Also, chances of third party threats cannot be ruled out. These members are not part of formal organisation but they have full potential to harm the organisation due to all the privileges given to them during their work tenure.

Insider attacks originate from within the organization whereas regular cyberattacks originate mainly from external entities who have motive to harm the business value of the organisation and its services. There are majority of defense techniques to prevent the regular cyberattack but insider attacks are a challenge due to identification of such attacks. There are no fixed threat models for such insider attacks and due to the sensitive and dynamic nature of insider attacks, this area of research is actively pursued in current times.

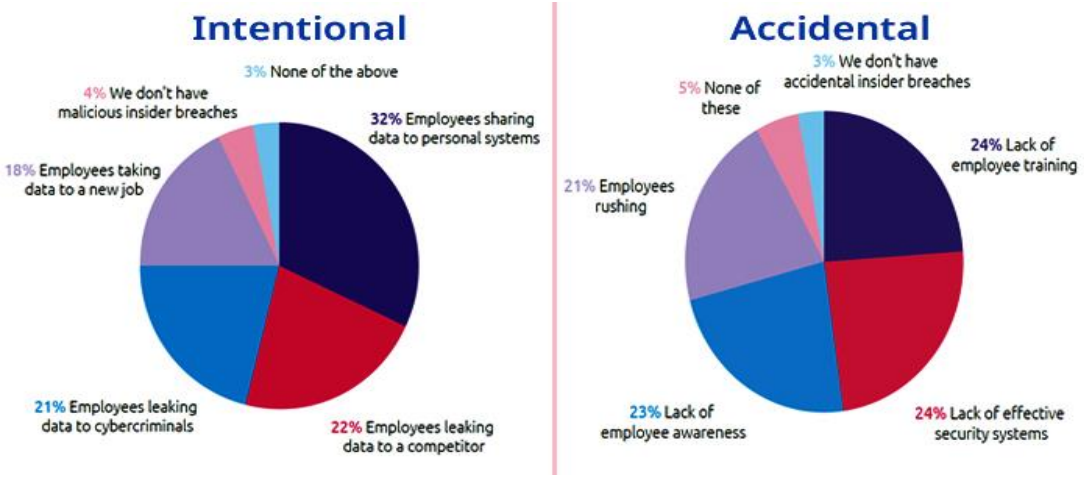

Figure 1. Causes of data breaches resulting in insider threats [6]

2. Existing models

There are only a few papers in this field and these helped to get an understanding of what is happening in the file of insider threat detection. The paper [1] is based on sentiment analysis. They used two different modes of learning: supervised and unsupervised. Algorithms used are Naive Bayes, SVM, Decision trees and linear regression in unsupervised and KNN, EM and dbscan in supervised. They found that the decision tree had a highest accuracy of 99.7%.

Table 1. Study of existing methodologies

Reference	Concept / Theoretical model / Framework	Methodology used / Implementation	Dataset details / Analysis	Relevant Findings	Limitations / Future Research / Gaps identified
[1]	NLP model consisting of ML algorithms like Adaboost, SVM, KNN, Naive Bayes, Linear and Logistic Regression	Pre-processing techniques - stop word removal, stemming, tokenization	2 datasets including behaviour of 24 users that were collected over 5 days span	AdaBoost outperformed other models with a high accuracy of 98%	Testing deep learning classifiers
[2]	Generating description form models to identify the pattern used for enhancing trust.	SHAP, LIME, Contrastive Explanations Method (CEM), ProtoDash and Boolean Decision Rules via Column Generation (BRCG)	NSL KDD dataset with 42 features in total	Identifying trustful patterns	
[3]	NLTK tool used to do a sentimental analysis with techniques such as Decision Trees, Linear Regression, Naive Bayes and SVM, K means, EM and DBSCAN	Pre-processed the data by removing unessential words like pronouns, prepositions, and articles	Sentiment 140 dataset was used	Decision Tree algorithm classified malicious insiders with a high accuracy of 99.7%	Analysis of individual user's sentiment

[4]	Mental state of the insider before attack was initiated	It sees the personalities of the insider and identifies those related to the dark triad	Includes theoretical foundation for insider threat attacks and their causes	Found a few causes that can be contributing to the development of an insider threat	Empirical relationship between factors which lead to an insider attacks are to be established
[5]	Development of an insider threat detection system using machine learning algorithms such as Isolation Forest and one class SVM	Due to huge data, set feature extraction was done from attributes from each log file such as weekday login, after working hours login, weekend login, logoff, email in, email out, email out with attachment, popular http and psychometric scores	CERT dataset which is over 20GB of various system log files recording all activities of 1,000 users in a duration of 500 days including weekends	SVM performed well compared to Isolation Forest with an AUC score of 0.97	Explore the idea of calculating the trust score based on other factors, such as access control policy of the system

The dataset used in this proposed work is the same used in the study [2]; however, this work uses the activity log as the main basis along with the psychometric score. Work [2] used unsupervised learning algorithms of Isolation Forest and one class SVM. It worked on a test score that was generated after a certain period of time and this score was used in the next cycle to find the intruder; the results vary depending on the length of each cycle. It is useful in raising flags in the future without auditing any logs. Paper [3] focused on the root causes of why insider threat happens and discussed a few traits and motives why anyone would turn to become an insider threat. Information on the attributes of the employees were gathered and found that with the right trigger the motive people turn out to be an insider threat. They classified the insider threats as 4 categories: espionage, IT Sabotage, fraud, theft of IP. Paper [4] explains why ML models have to be transparent for the end user and there are answers to why a certain model has taken a particular decision. This wasn't done easily and for this they

used various methods like SHAP, LIME, Contrastive Explanations Method (CEM), ProtoDash and Boolean Decision Rules via Column Generation (BRCG). Paper [5] used the TWOS dataset to detect the insider threat. They chose supervised machine learning models as they provided better recall. Adaboost, Naive Bayes (NB), Logistic Regression (LR), KNN, Linear Regression (LR) and Support Vector Machine (SVM) are the algorithms used. Adaboost gave an accuracy of 98.3% to detect malicious emails. The dataset used was small and detecting malicious emails isn't detecting insider threats.

3. Proposed Work

If an employee is happy in the company and has a good psychometric score, there are less chances of him taking part in an insider attack. Using this idea, people who are willing to take part in an insider attack and harm the company can also be found. The environment in which the employee works is also important in keeping the employees loyal and avoiding these insider attacks. For people to like the company they work in, they have to be happy with the working conditions of that company. Each person is different, and so the personality of each employee has to be considered. A survey which asks questions that will help to determine if the employee is happy in the company now or unhappy with something in the company, has been designed and its results are used in making the prediction along with their psychometric scores. To determine the questions to be asked in the survey, a few senior employees in various companies are contacted after which 9 questions are chosen to be in the survey. Using the results from the survey and combining it with the psychometric test score data from the CERT dataset, supervised learning models are generated to determine the insider based on these survey responses. There are many supervised techniques, but this paper focuses on Support Vector Machines, AdaBoost, XGBoost, Decision tree, K- Nearest Neighbours and Random forest algorithms.

3.1 Methodology adapted

Step 1: Identification of factors which may give rise to the intent of inside attack.

Step 2: Data acquisition through questionnaire/survey among employees of an organization.

Step 3: Manipulation of data on the proposed model.

Step 4: Training the data with CERT data on the model.

Step 5: Testing the data on the model.

Step 6: Generation of results and inferencing from the results in the form of visualizations.

3.2 Algorithm used in the proposed approach

An ensemble algorithm has been used for classification of a input value as belonging to a class. There are 2 classes. One class is identified as insider threat and other class is not an insider threat.

3.2.1 Novelty of Ensemble algorithms

It denotes to a group of algorithms that integrates the predictions of two or more algorithms for the following benefits: A group of algorithm can derive high prediction rate when compared to a single high-performance model, thereby causing performance enhancement. The weaknesses inheriting in some learning algorithms are compensated by the strengths of others. An optimal approach always gives overall beneficial outputs. Ensemble learning helps to improve the stability of individual algorithms by minimizing the error factors such as noise, variance, and bias.

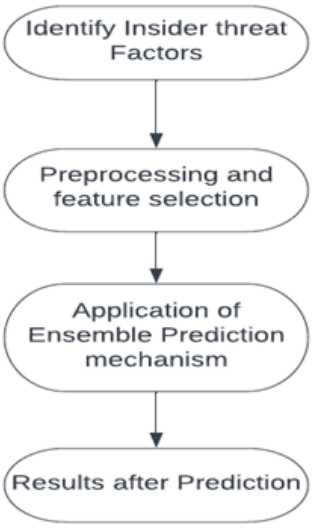

Figure 2. Block diagram of the proposed method [6]

A) Decision Tree

A decision tree originates as a node, which is compared with other properties in the dataset, for a split after a decision is made. A split denotes the number of outputs of one class that comes on one side of the tree and another class on the other side of the tree. For example,

if a node splits the data in such a way that all the outputs which are “No” come on one side (left/right) and outputs which are “Yes” come on the other side (left/right). It can be observed that every node gets divided as a decision until it arrives at a perfect split; the decision of a perfect split becomes the terminal node of a tree. The utmost challenge in developing a decision tree is the selection of attributes that is which attribute to be used as root node or an internal node; since there are a lot of attributes, it is quite difficult to select. There are two techniques for selection: information gain or gini index.

B) Random Forest

As the name Random Forest suggests, it works in the principle that the forest consists of a lot of trees. To make these models, first a bootstrapped dataset should be made, and in each step a subset of variables is considered as a candidate for the root node, and these steps are called bagging or bootstrap aggregating. These steps keep repeating till there is a good amount of trees. When there's a new input, the prediction is taken from all these trees, and the prediction that is most prominent among them will be the chosen prediction.

C) Support Vector Machine

The primary objective of SVM is to find a plane that separates an n-dimensional space into classes. It selects two data points, which are nearest to each other, from two classes as support vectors. The distance between the vectors and the plane which separates the classes is called margin. The main goal here is to find a plane which has maximum margin. So, the plane which separates with a maximum margin is an ideal hyperplane. When a new data point is given and asked to classify, the separating plane is noted and depending on the side it lies on, the point is classified as that. The kernel functions are used to make the SVM classify. When there is no plane that divides the points properly, these kernel functions transform the points to another form which could be divisible by the plane.

D) KNN

k Nearest Neighbor model is used in classification and regression. It works in a way that when an input is given, it finds each type of neighbors that are close to it and the number of neighbors taken into account can be varied. After finding the closest neighbors, the largest type of neighbors that are close to it will be the ones taking into account. If a classification is done, the input will also be of the same class, and if its regression, the average of these neighbors will be the answer. To find the closest neighbors, the method used to measure the

distance must be considered, and few popular methods to estimate the distances are Euclidean, Manhattan, Hamming and Murkowski.

E) XGBOOST

Extreme Gradient Boost (XGBoost), an ensemble tree based machine learning technique which is implemented on top of gradient boost, is used for both classification and regression problems. What distinguishes it from other models is its speed and performance; it is extremely speed and well performed than other models. The inbuilt features of the XGBoost namely parallelization, cache optimization, and out of core computation makes it work faster than the other techniques. As far as the performance is concerned, XGBoost has an inbuilt regularizer which helps to prevent overfitting. Moreover, it has another feature called auto pruning which allows to grow the tree up to a limit maintaining the bias variance, which makes the model more effective and robust.

F) Adaboost

Adaboost is also an ensemble technique. Like Random Forest algorithm, this also makes a number of trees but in this case it just has two leaf nodes. When the first tree is made, the incorrectly classified are sent to the second model, and depending on the error of the first one the next tree is made in a way to reduce the error. This is continued till the particular number specified earlier, is reached. As this is a boosting algorithm, the rows in the dataset can be repeated. Like Random Forest, the votes from each model are taken in, and the majority class is what will be the output.

4. Experimental Results

4.1 CERT dataset

CERT is a dataset on the insider threat generated by CERT along with a few other partners. The dataset covers various cases of traitor instances and masquerading activities. It contains various features like login data logs, HTTP or browsing history, emails, file access logs and device usage, but the psychometric test data has been used for this model. The responses from the survey, for which the questions made by a few experienced senior employees which help to determine how an employee feels about the company, have also been used. The information from the psychometric test and the survey responses combined together is the dataset used for this work. From the CERT dataset of year 2020, nearly 2000

records have been used, out of which 1600 records are used for training the ensemble model and 400 records are used for testing the model prediction. For pre-processing and feature selection, the EFS (Ensemble Feature Selection) tool downloaded as an R-Package from CRAN is used. It stabilises the problem of overfitting. Thereby a quantification of the importance scores of features can be obtained, and the method-specific biases can be compensated.

4.2 R-factor

The human behavioural elements over time are required, and for this a new feature called Resentment factor (R-factor) is utilized. To make the R-factor, various senior employees of a few companies were requested, and the most common parameters were, the time from the last salary appraisal, if they had an internal inquiry due to poor performance, job satisfaction, does the role assigned match the role promised, thoughts on competence of their manager, feelings towards the boss, job security and their thoughts of the work environment. Some factors in these are more important than the others, hence weights are assigned to these factors by taking the average of the weights that the employees assigned.

$$R\_factor = \sum(weight\_of\_factor \times factor\_score) \div (\sum(weight\_of\_factor)) \quad (1)$$

The dataset used is the mixture of these two i.e., taking the psychometric score from the CERT dataset and the R-factor from own survey and assigning it to the employees in the CERT dataset. The models are trained to detect an insider threat. The CERT dataset has various types of insiders. The dataset is changed to, whether a person is an insider or not an insider, a binary classification. For evaluating the models, an 80-20 split for training and testing the proposed model is done. Few metrics used to compare the models are accuracy, precision, recall and f-measure.

Accuracy: The percentage of correct predictions over the dataset is called Accuracy.

Mathematically,

$$Accuracy = Number\ of\ Correct\ Predictions / Total\ Predictions \quad (2)$$

Precision: Precision is the fraction of positive predictions that are actually correct.

Mathematically,

$$Precision = TP/TP+FP \quad (3)$$

Recall: Recall is the fraction of actual positives that are predicted correctly.

Mathematically,

$$Recall=TP/TP+FN \quad (4)$$

F-measure: F-score is the harmonic mean of Precision and recall.

Mathematically,

$$F\text{-score}= (2*Precision*Recall) / (Precision+Recall) \quad (5)$$

Here, TP is True Positive, FP is False Positive and FN is False Negative.

Python 3.8 on Jupyter Notebook interface is used for developing simulations for the proposed mechanism.

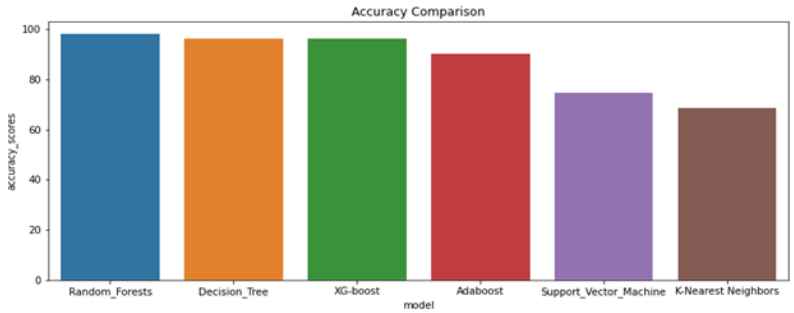

Figure 3. Accuracy comparison graph with different models

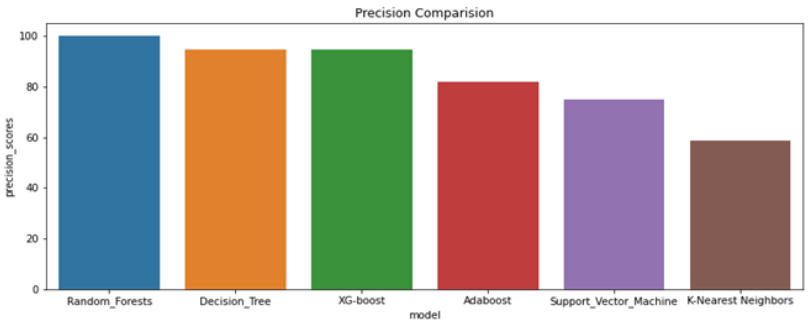

Figure 4. Precision comparison graph with different models

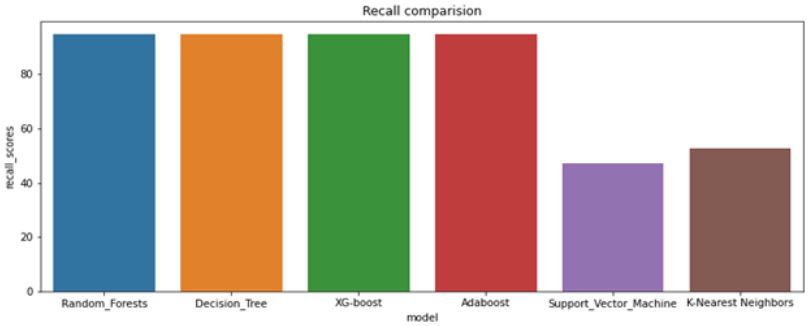

Figure 5. Recall comparison graph with different models

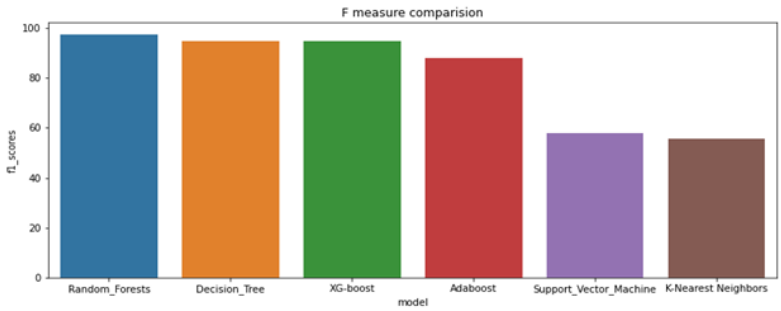

Figure 6. F measure comparison graph with different models

The highest performance is from the Random Forest model with an accuracy of 98% and the lowest accuracy is K-nearest neighbors model which is 68.6%. The XGBoost model and the Decision Tree model show the same performance measure for all 4 metrics. Random Forest performed very well in all the chosen metrics when compared to the other supervised algorithms used. AdaBoost algorithm is unsuitable for applying R-factor due to its sensitiveness towards noisy data and outliers. It needs a quality dataset but the R-factor values contain outlier properties in this study. Hence, the R-factor could not be used on AdaBoost algorithm for this proposed approach.

5. Conclusion

This paper proposes an Intrusion Detection System to identify the insiders who are imposters to the organization based on their Psychometric scores and Resentment factor (R-factor), with the help of Machine Learning techniques. Based on the experimental results, the Random Forest model with an accuracy of 98%, outperforms other supervised machine learning algorithms chosen. This study uses the most widely used and publicly available CERT dataset. A new function (R-factor) is added to the dataset. Several employees in multiple companies are surveyed and based on the survey results, R-factor is determined. Weights are added to each question and the weighted average of the data is calculated. The R-factor is a valuable addition to the dataset as it has been considered as the most important feature in all the models in the paper other than the AdaBoost. In future, few new features would be added to the data and applying advanced deep learning techniques could improve the metrics of the models.

References

[1] MediaWonPark ,Youngin You , and Kyungho Lee:Detecting Potential Insider Threat: Analyzing Insiders' Sentiment Exposed in Social Networks.

[2] Aldairi, Maryam, Leila Karimi, and James Joshi. "A trust aware unsupervised learning approach for insider threat detection." 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI). IEEE, 2019.

[3] Michele Maasberg; John Warren; Nicole L. Beebe:The Dark Side of the Insider: Detecting the Insider Threat through Examination of Dark Triad Personality Traits.2015 48th Hawaii International Conference on System Sciences, ISBN:978-1-4799-7367-5, ISSN: 1530-1605, DOI: 10.1109/HICSS.2015.423

[4] Mane, Shraddha, and Dattaraj Rao. "Explaining network intrusion detection system using explainable AI framework." arXiv preprint arXiv:2103.07110 (2021).

[5] FaisalJanjuaaAsifMasoodaHaiderAbbasaImranRashida Handling Insider Threat through Supervised Machine Learning Techniques, Procedia Computer Science Volume 177, 2020.

[6] www.stealthlabs.com/blog/infographic-20-alarming-insider-threats-statistics/