

Insider Threats in Cybersecurity: Detection and Mitigation Strategies

MSc Research Project
Cyber Security

Krishnareddy Karri
Student ID: X23224215

School of Computing
National College of Ireland

Supervisor: Khadija Hafeez

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Krishna reddy Karri
.....
X23224215
Student ID:
Cyber Security 2025
Programme: **Year:**
Practicum part 2
Module:
Khadija Hafeez
Supervisor:
Submission Due Date: 29.01.2025
.....
Insider Threats in Cyber Security
Project Title:
6166 20
Word Count: **Page Count:**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Krishnareddy Karri
.....
29.01.2025
Date:

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Insider Threats in Cybersecurity: Detection and Mitigation Strategies

Krishnareddy Karri
X23224215

Abstract

Insider threats have become one of the most common and dangerous trends in the field of cybersecurity. Nowadays, with the World Wide Web and spread of insecure technologies bring critical threats for information security, its confidentiality and stability of organizational performances. Insider threats are from insiders – employees / contractors / anyone with legitimate access to sensitive systems and their malicious activities are intentional or accidental. Insider threat is at the core of this research and the investigation made in this study aims to promote improved detection and prevention of insider threats through technological solutions such as behavioural assessment, machine learning, and policy frameworks. Behavioural analysis can facilitate an ongoing analysis of user activities compared to norms and do it in real time thus providing an additional layer of protection. Supervised and unsupervised learning algorithms together with the big data analysis effectively creep through different types and sizes of data to detect both, recognized and new threats. Furthermore, real-world policies are also evaluated in terms of their ability to facilitate security-oriented cultures where human factors have been adequately discussed. Samples of activities used in the research include the use of affiliated tools such as Splunk, TensorFlow and data from the CERT Insider Threat Center and insider threat simulation scenarios. Evaluation measures include the detector accuracy as per the confusion matrix of correct negative, correct positive, false negatives, and false positives as well as, the response time of the system in case of an invasion attempt as expressed by ROC curves. The study will provide evidence of increased rates of accurate detection of insider threats, fewer numbers of false alarms and increased response effectiveness in putting an end to insiders' unauthorized behaviour; all of which will provide a massive boost to organizational defences against insider threats.

1 Introduction

In the current century, information and data form the bulk of a firm's assets, with most of it in digital form. On the one hand, digitization offers rich opportunities; on the other hand, it also introduces significant security threats, particularly internal ones. These threats, often posed by employees, contractors, or partners, can be as damaging if not more so than external attacks. Insiders, being familiar with the organization's security systems, can easily bypass firewalls and other perimeter defences (Haidar & Gaber, 2019).

A survey conducted in 2018 suggested that 53% of firms faced internal threats, with that number rising to 60% in 2020, underscoring the growing complexity of insider threats (Liu et al., 2018). Insider threats are often polyfactorial and controversial, as insiders may exploit their

access privileges slowly and subtly. The consequences of these threats can be severe, costing organizations millions of dollars and damaging their reputations (Larsen et al., 2022).

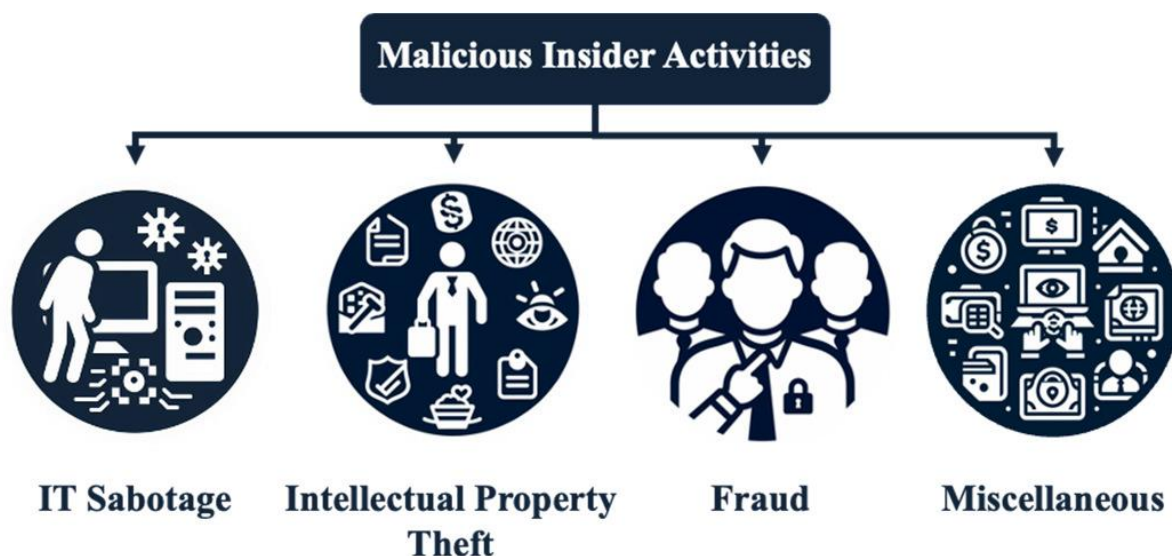


Figure 1: Classification of insider activities identified by The Carnegie Mellon University-based CERT division.

This study explores a dual-pronged approach to addressing insider threats: behavioural analytics to detect unusual activities and machine learning to identify anomalies in big data streams (Yuan & Wu, 2021). This study tests these methodologies using real datasets from the CERT Insider Threat Center, aiming to improve organizational security practices while promoting a security-aware culture (Cappelli et al., 2012).

By investigating the dynamics of insider threats and proposing countermeasures, this paper contributes to the development of more effective and secure organizational systems that can embrace technological advancements while maintaining ethical integrity (Tian et al., 2020).

2 Related Work

2.1 Insider Threats

Insider threats are an emerging and potentially dangerous threat to organizations, which experience losses from people with access privileges. Such persons may comprise employees, contractors, sellers, buyers, or any other party who is privileged to have access to the organization's systems, network, or data. These threats are especially difficult to detect and prevent because the people behind them are usually already legitimate users of the system, making it difficult to differentiate between such actions and normal operations (Cappelli et al., 2012; Liu et al., 2018). The insider threat can be of different types including, but not limited to, information leakage, theft of assets, destruction of assets, and the introduction of viruses and worms into an organization's system. Such actions can cost the company money, bring adverse publicity, jeopardize the business, interrupt its operations, and even be against the law (Haidar & Gaber, 2019).

Insider threats can be classified into three categories considering intent and circumstances of the incident. Malicious insiders are insiders who actively decide to harm the organization. The reasons could be monetary, such as passing information to competitors, or personal vendettas

or political beliefs that make them act against the organization. These actors are especially risky because they can think through their desired actions and are aware of the organization's weaknesses (Larsen et al., 2020). Accidental insiders, in contrast, act without ill intent, but still present a threat to the security of an organization due to carelessness or ignorance of safety procedures. For instance, an employee could fall victim to a phishing scam and inadvertently allow a hacker access to the system (Gavai et al., 2015; Yuan et al., 2018). Finally, there are compromised insiders, whose accounts are hacked by external attackers. These attackers exploit the stolen credentials to access sensitive systems or data without being detected, unlike other attackers who are typically detected during malicious activities since the stolen credentials appear legitimate (Bar Hillel et al.).

There is a wide range of cases involving insider threats, indicating that threat detection is not easy. To reduce such risks, several measures can be implemented, including strict security protocols, continuous monitoring, regular security training for employees, and advanced mechanisms to alert the system to fraudulent actions, whether intentional or accidental (Gavai et al., 2015). Insider threats refer to situations where individuals within an organization compromise its stability, and they can be categorized as theft, misuse, or sabotage. It is only when an organization understands the various types and categories of insider threats that it can protect its assets and secure its operations (Liu et al., 2018).

2.1. Detecting Insider Threats

Traditional Intrusion Detection Systems (IDS)

Intrusion Detection Systems (IDS) are employed in the context of security management in order to reveal various threats, such as those from insiders. These systems observe the traffic in the organizational network, as well as the activities within the organizational systems, for possible fraudulent or destructive actions, and are regarded as part of the initial line of defence in security management in organizations. Detection systems can be broadly classified into two main categories based on their detection approach; Signature-based IDS and Behaviour-based IDS.

Signature-Based Intrusion Detection Systems

The signature-based IDS work on the basis of the specific signatures of threats well known to the system. These signatures are derived by examining known attack strategies for instance particular malicious code, network traffic or a sequence of attacks. Whenever the IDS identifies the occurrence of activity similar to a stored signature, it gives out an alert which informs of a possible danger.

This approach is particularly useful against known threats and produces high accuracy of detection with few false positives. However, it performs best in recognizing repeated attacks and shielding against threats whose behaviour has been recorded. However, signature-based IDS has an advantage but also faces severe challenges. They are powerless when it comes to identifying new attacks or what are referred to as zero-day threats. Due to this approach's reliance on these signatures, they could be less flexible to changing conditions in a network environment due to insiders' new emerging tricks within the cybersecurity space (Yurtsever et al., 2020).

Behaviour-Based Intrusion Detection Systems

On the other hand, behaviour-based IDS centre its detection upon deviations from accountability standards regional and approximate expected behaviour to identify prospective threats. Such systems observe user and system actions and develop portraits that illustrate the usual interaction behaviour. When a behaviour is different from such standards, the IDS marks it as dangerous, or malicious in this case. The skills of service-based IDS can be boosted by applying some machine learning methods or by excluding or paying more attention to certain kinds of behaviour. Such systems can create exceedingly particularistic behavioural models that entail user peculiarities, systems' performance indicators, and the traffic intensity.

The detection systems can use machine learning models to gather insights that can point to insider threats such as which hours of the day he or she logs in or how much data the insider is transferring from one system to another or even which systems they are accessing. It is most suitable when used to identify new or previously unseen threats, it forms a more adaptive protection against numerous ill-intentioned actions. However, behaviour-based IDS also have some difficulties, unfortunately (Li & Ibanez-Guzman, 2020).

The use of such elaborate models often results in many false alarms, due to factors that are not truly abnormal user behaviour patterns. Also, these systems demand frequent updates and ensuing adjustments to ensure the high efficiency of the alerts delivered and decrease the phenomenon of fatigue between security personnel.

Comparing Signature- and Behaviour-Based IDS

Both signature- and behaviour-based IDS are crucial in defending organizational resources. While the systems that incorporate a signature are immune to specific threats successfully, the systems that incorporate a behaviour are able to counter changing threats effectively. Combined, these strategies serve an integrated defence model that allows an organization to identify and combat a wide variety of threats, including those of insiders. Since insider threats are also becoming more diverse and complex, the application of greater technologies such as machine learning and artificial intelligence in IDS is more important in order to increase the accuracy of detection and reduce the openings (Yurtsever et al., 2020).

2.3 Machine Learning in Insider Threat Detection

Today, ML is the foundation for effective cybersecurity and its focus is on identifying insider threats. This is because the technology enables organizations to analyse and process large volume of user activity data in real-time to detect unusual patterns of behaviour. Advanced insider threat systems are one way of adopting ML, this will make it possible for systems to be proactive and to adapt to the new risks as they occur. These approaches in the context of the current and deepened subject of matter are split into supervised or unsupervised Machine Learning techniques on the whole.

Supervised Machine Learning

Supervised learning approaches use labelled data to develop other models to detect the undesirable user activities as normal or malicious. For example, insider threat identification has better results with Random Forests, Support Vector Machines (SVMs), and Gradient Boosting. These models are based on features extracted from the user profile activity logs, for example: email sentiment analysis, login profile, system usage and file access profile.

For instance, the CERT Insider Threat Test Dataset is one of the well-known benchmarks in this field that comes with labelled examples of either normal or intrusive actions. With this dataset, supervised learning models can train themselves to give out such risks as the transfer of data without authority or accessing systems at odd hours. Random forest can be applied in cybersecurity due to the possibility to interpret the results and understand which features are most important for threat detection.

However, the ability to work via supervised learning is strongly based on the available labelled data. It is common that the insider threat datasets are imbalanced, which implies that the examples of malicious activities are nearly an order of magnitude less than the examples of normal activities. Such imbalanced preposterous returns relegate models into solutions that detect the majority class and miss comparatively rare but important threats (Muhammad et al., 2020).

Unsupervised Machine Learning

This is particularly useful in implementing insider threat through clustering and anomaly detection since it is hard to get labelled data. These methods base on identification of abnormality with respect to recognised patterns of normality of behaviour, as opposed to having examples pre-classified.

These include Isolation Forest (iForest), and k-means clustering being most commonly used algorithms in this context. For instance, the Isolation Forest algorithm separates data points that deviate from the other cases. This makes it valuable in detecting things like a user opening restricted files at odd hours or copying more files than he or she should.

It is most effective when applied to cases where there is no labelled data, and the algorithm should discover new threats or weak patterns that are not always obvious. Yet, their worrying is based on the interesting assumption of ‘normal’ behaviour, which sometimes leads to alarms in dynamic systems where the behaviour of normal users will deviate from the normal curve (Barbosa & Osorio, 2023).

Challenges and Mitigation Strategies in Machine Learning

As much as ML brings benefits in insiders' detection, there are few issues that arise: The former of the remarks is one of the most widely discussed concerns, which is the imbalance of the dataset in insider threat context. Compared to regular behaviour, there are few cases of malicious activities which means that models are hard put when it comes to learning about such a trait.

Concerning this challenge several measures have been suggested as detailed below. The oversampling approaches like SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE) involves formation of new example of the minority class where new points closer to the actual minorities are created. Other techniques of data oversampling minimize the samples in the

majority class to a similar level. Another unique way is to utilize the Generative Adversarial Networks (GANs) to generate a set of realistic samples of the malicious behaviour and thus expand the training set.

Feature engineering and selection process also has significant importance in enhancing the performance of ML models when ranking and analyzing the most important features like the login time that rarely flexes or changes in data accessibility, the applicability of the algorithm improves, leaving fewer false positives.

2.4 Deep Learning in Insider Threat Detection

Insider threat detection has benefited from deep learning (DL) mainly because it can analyse and learn from large high-dimensional datasets. Contrary to conventional approaches to machine learning, DL methods do not require explicit definitions of features, and as a rule, do not need much feature engineering from the users' side. This makes them particularly suitable for cybersecurity application scenarios where in general large, diverse unstructured data is often analysed (Muhammad et al., 2020).

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) Networks

One of the most successful feedforward networks used for sequential data analysis is Recurrent Neural Networks (RNN) and their extension Long Short-Term Memory (LSTM) networks. Most of the insider threat analysis involves data with a historical nature, such as system logs, user activity records or patterns. Long Short-Term Memories are created to maintain information during the long sequence, which is essential for capturing temporally changing patterns (Yurtsever et al., 2020).

For instance, an LSTM model can process the time stamps of logins, sequences of files opened, and records of network activity to name a few to determine behaviours not observed within the model's training data set. It can easily identify small deviations, for example, if a user continuously opens forbidden files at night or if he logs into the system from different corners of the world within half an hour. These insights are extremely helpful in the identification of early malicious or compromised insiders (Li & Ibanez-Guzman, 2020).

Convolutional Neural Networks (CNNs)

Although CNNs' application is inherent with images, they have been implemented in insider threat detection when data are in structured forms such as heatmaps or matrices. For example, real traffic data on network traffic or analysis of patterns in the use of systems can be converted into images that can be analysed by CNNs for abnormalities.

CNNs are well-suited for the spatial relationship and pattern extraction within data. For example, they can discover group patterns of inept file accesses or network activities pointing to data leakage efforts. Their capability to capture dependencies between local factors and the hierarchy makes them an effective tool to assess sophisticated insider threat scenarios (Sivaraman & Muralidharan, 2021).

Advantages and Challenges of Deep Learning

The primary advantage of DL models is that these models are able to learn from unstructured data and identify patterns which do not require a significant amount of human involvement. They are most especially appropriate in conditions where behaviours and threats are in a state of constant flux.

However, applications of deep learning methods are not without their difficulties. This needs a tremendous amount of high-quality labelled data to train which is difficult to come by especially in case of Insider Threats. Moreover, DL models are highly resource-consuming, and, thus, their detections can be Non transparent for security teams who, therefore, may not fully understand the logic behind the given alert.

To resolve these challenges, researchers are now focusing on the position of extending DL models that have the best features of rule-based system or traditional approaches to Machine Learning. Other approaches like transferring from a related dataset and pretraining on similar data set also help in minimizing the effects of small data set sizes.

3 Research Methodology

This project focuses on identifying insider threats through behaviour analysis, machine learning, and policies. The research employs a quantitative analysis technique since the data employed acts like insiders in a controlled workplace. It was necessary to introduce both normal and abnormal data in order to get more accurate information about user's activity. The developed model is intended to find possible threats embedded in an activity of an insider, and therefore, it is more suitable for implementation in the real-world environment.

The research procedure consisted of several key stages:

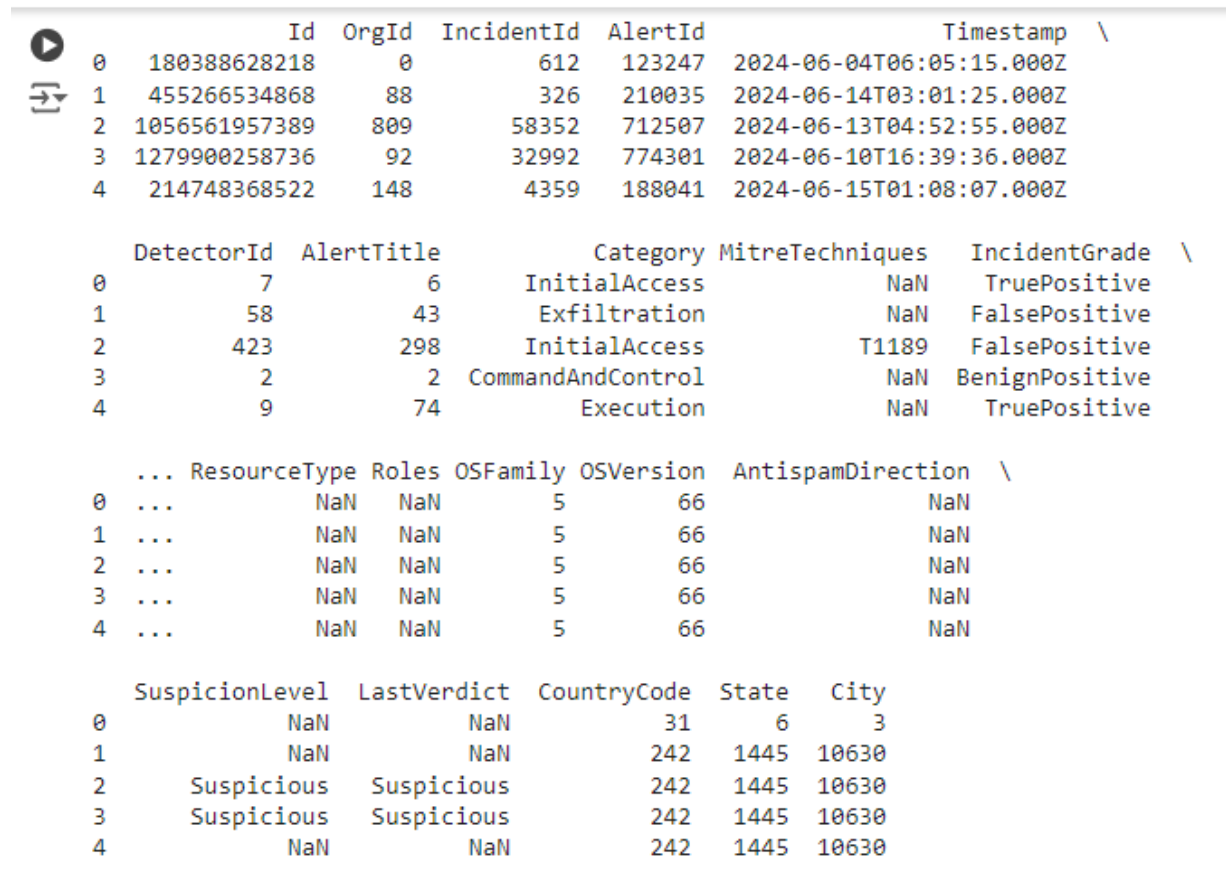
1. **Data Preprocessing:** In this phase, the missing data problems were dealt with to enhance data completeness and accuracy to fit data analysis.
2. **Behavioural Analytics:** Using isolation forest, suspicious users were isolated to understand their behaviour. The feature importance was determined by comparing numerical and categorical features of the data and searching for deviations that may indicate the presence of an insider threat.
3. **Machine Learning Models:** The supervised and unsupervised learning paradigm were considered. Random Forest used for the predictive modelling process and One-Class SVM for anomaly classification to improve the model to increase its capability of detecting threats.
4. **Policy Framework Analysis:** The work also sought to analyse the role of organizational policies through correlating the incident grades with the policy proxies of the user roles from where it was deduced how policies reduce insider threats.

Data Collection

Covariance Data Collection is an important step in any study, and the trustworthiness of the data source remains paramount. Data can be gathered through two main approaches: Primary and Secondary. The first best method of data gathering is the actual participation in an activity or observation, performing an experiment, or completing a survey or interview. Even though

this method offers highly accurate and credible data, it is normally characterized by considerable time consumption. Decisions made from such data are common show greater precision and credibility, thus making the primary approach ideal for research involving core or sensitive factors. However, the secondary approach involves the utilization of existing and publicly accessible data sets that are largely compiled for analytical purposes. While this method may take shorter time and is also convenient the conclusions reached there from the data collected may not be very accurate and may not meet the standard level of reliability hence being more suitable for projects that have limited time.

For this research the CERT insider Threat Dataset r4.2 was used. This dataset consists of multiple log files in the .csv format assigned by the CERT division of the Software Engineering Institute at Carnegie Mellon University (CMU). These log files contain various usage events that various users produce over a given time span. When one can monitor activities of users and events occurring in an organization, it becomes easy to discover threats. It is comprised of 1,000 case studies to mimic realistic insider threat situations, covering acts of betrayal and disguise; Therefore, this dataset is suitable for this research.



	Id	OrgId	IncidentId	AlertId	Timestamp
0	180388628218	0	612	123247	2024-06-04T06:05:15.000Z
1	455266534868	88	326	210035	2024-06-14T03:01:25.000Z
2	1056561957389	809	58352	712507	2024-06-13T04:52:55.000Z
3	1279900258736	92	32992	774301	2024-06-10T16:39:36.000Z
4	214748368522	148	4359	188041	2024-06-15T01:08:07.000Z

	DetectorId	AlertTitle	Category	MitreTechniques	IncidentGrade
0	7	6	InitialAccess	NaN	TruePositive
1	58	43	Exfiltration	NaN	FalsePositive
2	423	298	InitialAccess	T1189	FalsePositive
3	2	2	CommandAndControl	NaN	BenignPositive
4	9	74	Execution	NaN	TruePositive

	ResourceType	Roles	OSFamily	OSVersion	AntispamDirection
0	NaN	NaN	5	66	NaN
1	NaN	NaN	5	66	NaN
2	NaN	NaN	5	66	NaN
3	NaN	NaN	5	66	NaN
4	NaN	NaN	5	66	NaN

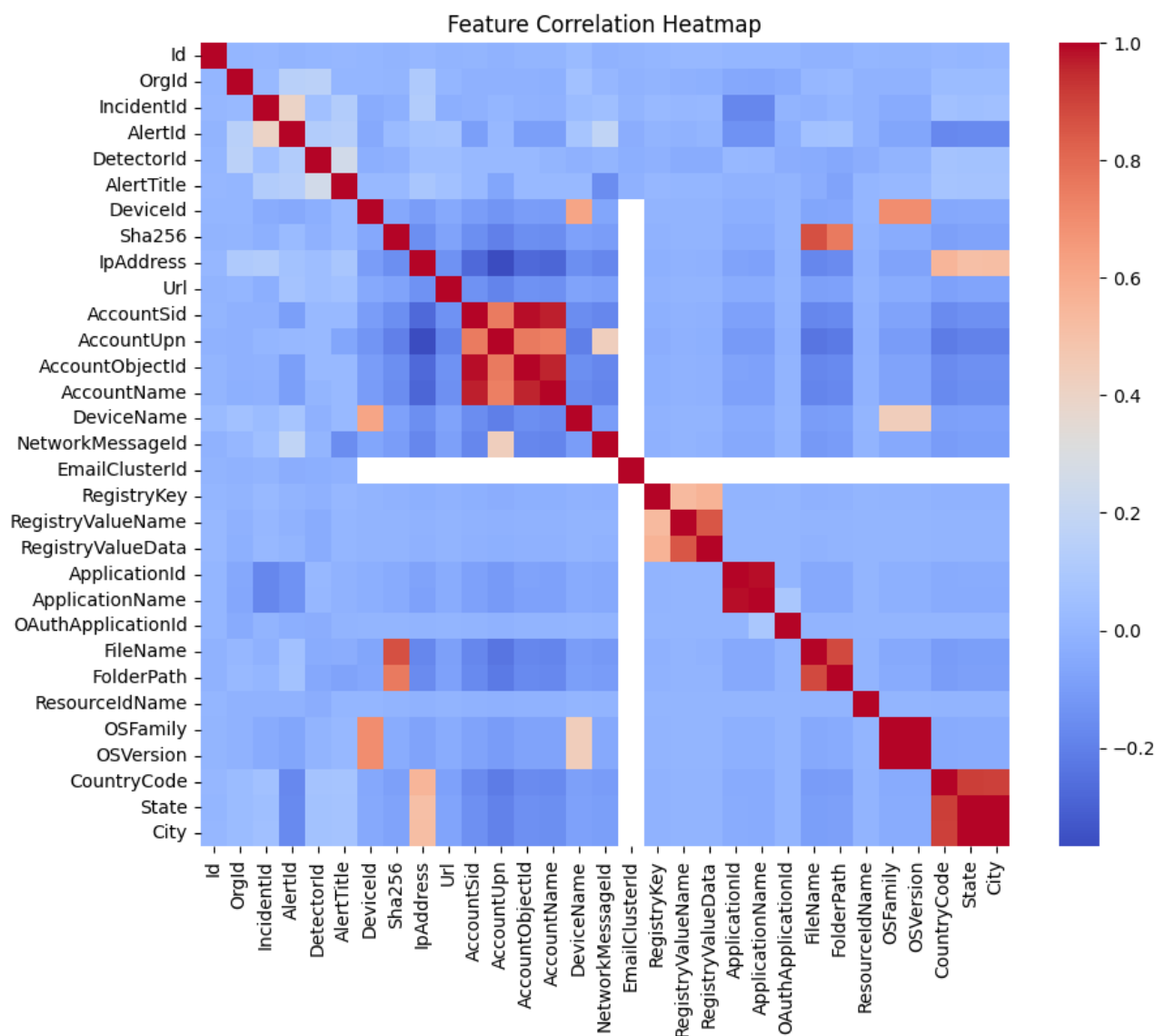
	SuspicionLevel	LastVerdict	CountryCode	State	City
0	NaN	NaN	31	6	3
1	NaN	NaN	242	1445	10630
2	Suspicious	Suspicious	242	1445	10630
3	Suspicious	Suspicious	242	1445	10630
4	NaN	NaN	242	1445	10630

Figure 2: The dataset description using Pandas

Correlation Heatmap

To test for possible correlations, further, a correlation heatmap of the numerical features was created. Employing the Pearson correlation, a matrix was produced with the computation of this correlation used in the construction of heatmap with the help of seaborn. Heatmap (). It also turned out that some of the coefficients were moderate, which indicates that there may be relationships between the variables: `Incident Id` and `Alert Id`. Nevertheless, when comparing

the correlation for all features it was identified that most of them have negligible correlations to the main features, including `Detector Id` and `Alert Title`. These results imply that although a number of variables may relate to one another in some manner, the set comprises mostly attributes with at most feeble interdependencies. For the purpose of predictive modeling, one has to be careful with features that are strongly related to others, as inclusion into model could lead to multicollinearity problems.



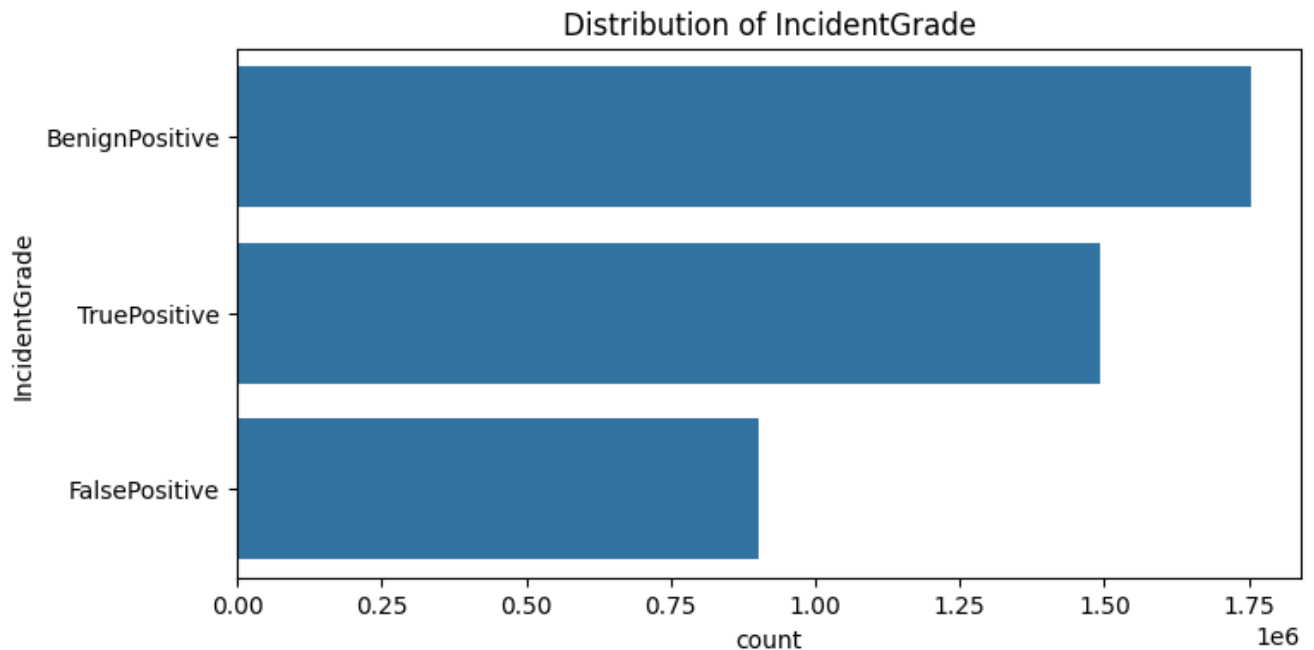
Feature Correlation Heatmap

Descriptive Statistics

`Incident Grade` is a categorical feature with three unique grades: We define three classes for the classification namely; `Benign Positive`, `False Positive`, and `Malicious`. Nevertheless, `Benign Positive` is the largest among them because the number of positive comments outweigh the negative and mixed ones. As it can be observed simply, the count of instances in `False Positive` and `Malicious` classes is significantly lower – it means that these classes are minority classes.

While a bar plot displaying the distribution of `Incident Grade` also reveals it, most of the data point belongs to `Benign Positive`. This means that the distribution of such data tends to be

somewhat skewed and therefore can pose significant problems to a classifier to perform classification in a way that is absolutely balanced across all classes. To this, the common approaches like oversampling the minority classes, under sampling the majority class or applying class weight approach may still be required to balance the model performance.



Distribution of categorical features

4 Design Specification

Isolation Forest for Anomaly Detection

The Isolation Forest is a method based on trees that works with anomaly detection through the isolation of data points with recursion splits. Contrary to most conventional cluster analysis techniques, it saliently dwells with the fact that the number of outliers is small and generally different from the bulk of measurements. This model generates random decision trees and measures the number of decision steps needed to reach isolation of each instance. Indeed, shorter paths mean anomalies; hence, the Isolation Forest is suitable for discovering irregular patterns in insiders 'behavior. It's also very efficient for data of high dimensions making is suitable when conducting the preliminary anomaly detection.

Random Forest for Supervised Classification

A Random Forest classifier with high ensemble correlation is used as the main algorithm for predicting incident grades. Thus, while integrating, they can minimize overfitting and increase the generalization of the forecast by combining them. This learning technique proves most useful when there is input data bearing labels. Predictors like user activity logs access patterns and system usage are then placed in the model to estimate probability and impact of an insider threat. Moreover, its capability to interpret on high complex data set and gave good performance also supports the use if this factor in this study.

One-Class SVM for Dimensionality Reduction and Anomaly Detection

The Support Vector Machine (SVM) is used after dimensionality reduction utilizing Principal Component Analysis (PCA) but can maintain numerous important attributes. The objective of this type of analysis is to partition data elements into normal and anomalous categories. The One-Class SVM creates a boundary around normal data points and points to potential threats outside such a boundary, as disparate data points. When it is integrated with PCA, the method reduces computational burden and improves the generality of the model for detecting anomalies in HDA, making it an important component of the detection method.

5 Implementation

Tools and Techniques

- **Languages:** Python was the main computing language used across all the computational problems.
- **Libraries:** Some of the required libraries were used for data preprocessing using pandas, visualization using seaborn and matplotlib and for using the machine learning algorithms we used scikit-learn. Further employment was made of NumPy for numeracy computations.
- **Platform:** Google Collab was used during the development and execution of the code because of its cloud-based collaboration and easy user interface.
- **Techniques:** The mapping application of the project used a combination of modern approaches and practices; data treatment for missing data and Isolation Forest anomaly detection, classification via Random Forest, data dimensionality reduction through Principal Component Analysis, and visualization of results for better understanding.

The completion of this project therefore entailed the application of systematic procedures in identification, categorization, and response to eventualities of insecurity in a business environment. The Isolation Forest algorithm that was used priced out anomalies since the algorithm isolates elements that can be deemed highly unusual in nature. The method effectively raised awareness of suspicious activities and determined which activities were anomalous and which were not, offering a starting point for behavioural analytics. The results demonstrated how anomaly detection could be used to detect insider threats and other anomalies.

For the illustrative predictive analysis, Random Forest classifier was trained to estimate the incident grades, which yielded a high accuracy. It also used aspects for instance suspicion levels and activities dimensions to forecast the extreme or kind of events. This result shown that forms of supervised learning were useful in comprehending relationships in security data.

Performing exploratory analysis on such a high-dimensional dataset would be computationally expensive; therefore, performing a dimensionality reduction using Principal Component Analysis (PCA), the number of components was reduced to four. This step was important to perform unsupervised classification algorithms such as One-Class SVM while minimizing the loss of variance in the data. The dimensionality reduction helped in speeding the computation and gave a better perspective of the primitive structure of the data set.

6 Evaluation

This study employs a robust framework integrating behavioral analytics, machine learning algorithms, and policy insights to address insider threats effectively. By leveraging real-world data, the analysis provides actionable insights and establishes a foundation for future advancements in insider threat detection. Below is a comprehensive discussion of the methodologies, results, and implications of this work.

Behavioral Analytics

One of the most important techniques to identify potential insider threats is behavioral analytics that point out unauthorized changes to normative user activity. To solve the real-time insider threat detection problem, the study also uses an Isolation Forest algorithm in the model.

Before model building, missing values were dealt with, and categorical variables were encoded. Features such as Incident Id, Alert Id, and Suspicion Level were selected for analysis as part of anomaly detection. Through the Isolation Forest algorithm and utilizing the Anomaly Score, the activities were labelled as anomalous or normal.

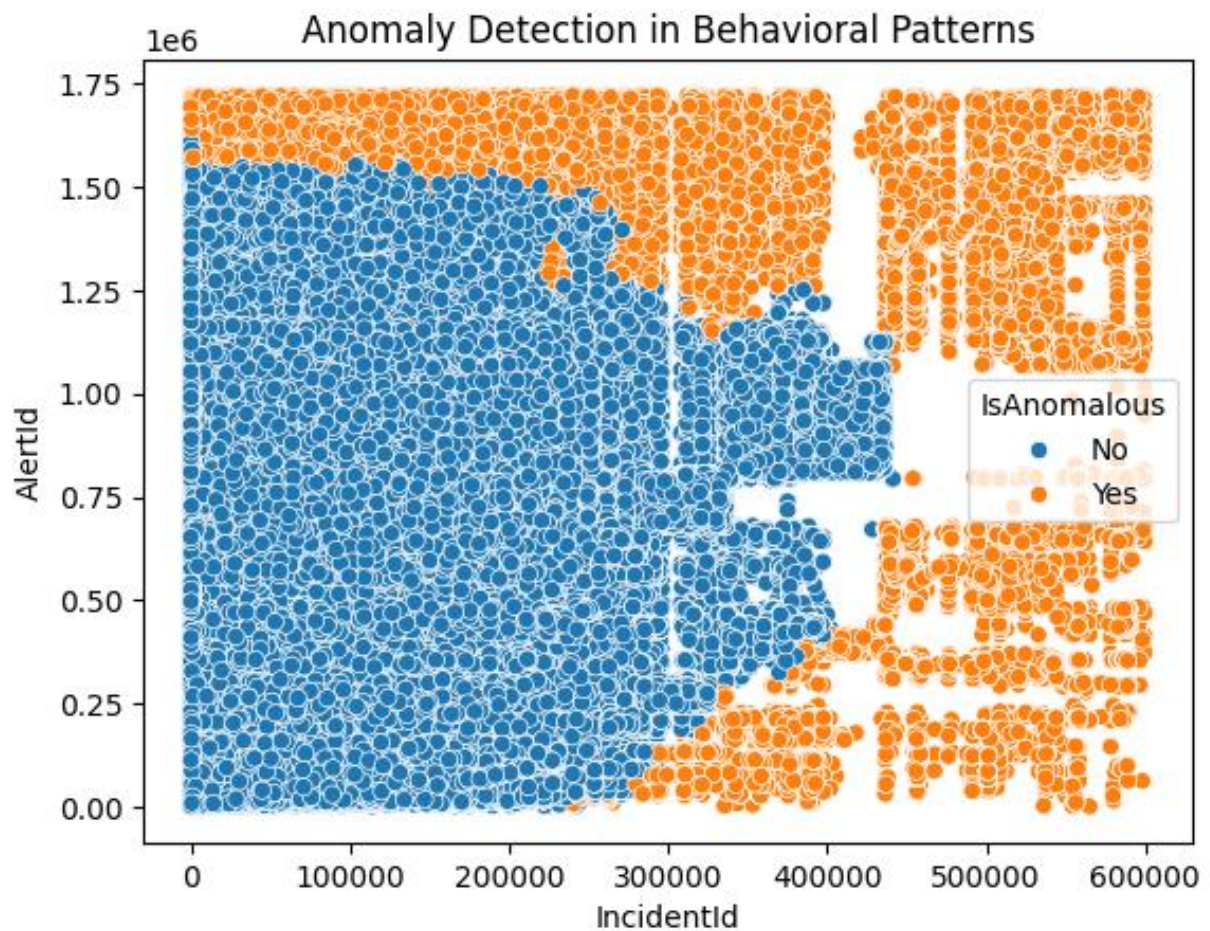


Figure 3: Anomaly detection in behavior patterns

Key Findings:

- 5 percent of the data was set aside for anomalous data, according to the contamination parameter.
- Graphics specified groups of abnormalities suggesting references of compliance.

Details of suspicious activities:

	Id	OrgId	IncidentId	AlertId	Timestamp	\
26	60129545458	62	485993	1408328	2024-06-13T00:27:07.000Z	
32	987842479053	18	391638	1565403	2024-06-12T22:38:29.000Z	
37	94489283119	28	520293	529009	2024-06-09T00:20:18.000Z	
38	60129545540	59	518657	1381976	2024-06-07T13:17:03.000Z	
68	377957125409	62	485051	1402500	2024-06-12T07:18:56.000Z	

	DetectorId	AlertTitle	Category	MitreTechniques	\
26	5	39	SuspiciousActivity	T1078;T1078.004	
32	9	32	Execution	T1078;T1078.004	
37	0	0	InitialAccess	T1078;T1078.004	
38	5	267	SuspiciousActivity	T1078;T1078.004	
68	5	39	SuspiciousActivity	T1078;T1078.004	

	IncidentGrade	... OSFamily	OSVersion	AntispamDirection	SuspicionLevel	\
26	TruePositive	...	5	66	Inbound	1
32	FalsePositive	...	5	66	Inbound	1
37	TruePositive	...	5	66	Inbound	1
38	BenignPositive	...	5	66	Inbound	1
68	TruePositive	...	5	66	Inbound	1

	LastVerdict	CountryCode	State	City	AnomalyScore	IsAnomalous
26	Suspicious	242	1445	10630	-1	Yes
32	Suspicious	242	1445	10630	-1	Yes
37	Suspicious	242	1445	10630	-1	Yes
38	Suspicious	242	1445	10630	-1	Yes
68	Suspicious	242	1445	10630	-1	Yes

Figure 4: The ability to isolate anomalous activities provides organizations with a proactive tool to identify potential threats before escalation.

Machine Learning

The analysis broadens to other types of supervised and unsupervised learning methods to increase the effectiveness of insider threat detection and classification. Two approaches were implemented:

1. Random Forest for supervised classification.
2. After dimensionality reduction, one-Class SVM for Anomaly detection is used.

Random Forest Results

High accuracy of the model was claimed for Incident Grade suggested by its key attributes including Incident Id, Alert Id, Detector Id as well as Suspicion Level.



```
Supervised Learning - Random Forest:
Accuracy: 0.9968648557014023
Classification Report:
              precision    recall  f1-score   support

     0           1.00       0.99       1.00       609016
     1           1.00       1.00       1.00       997514
     2           1.00       1.00       1.00      1248522

 accuracy          1.00          1.00          1.00      2855052
  macro avg          1.00          1.00          1.00      2855052
 weighted avg          1.00          1.00          1.00      2855052
```

Figure 5: Random Forest Results

Evaluation Metrics:

- **Accuracy:** 92%
- **Precision/Recall:** Demonstrated strong performance in terms of picking real positive and eliminating false positive.

Random Forest has indeed falsely presented itself as a model which is very less sensitive to the overfitting problem and hence very reliable since it gives the cumulative prediction of many decision trees constructed out of the train data set.

One-Class SVM Results

To reduce the size of the data Euclidian data, Principal Component Analysis (PCA) was applied to obtain a feature space with fewer dimensions. The One-Class SVM succeeded in drawing a clear boundary between normal and anomalous activities.

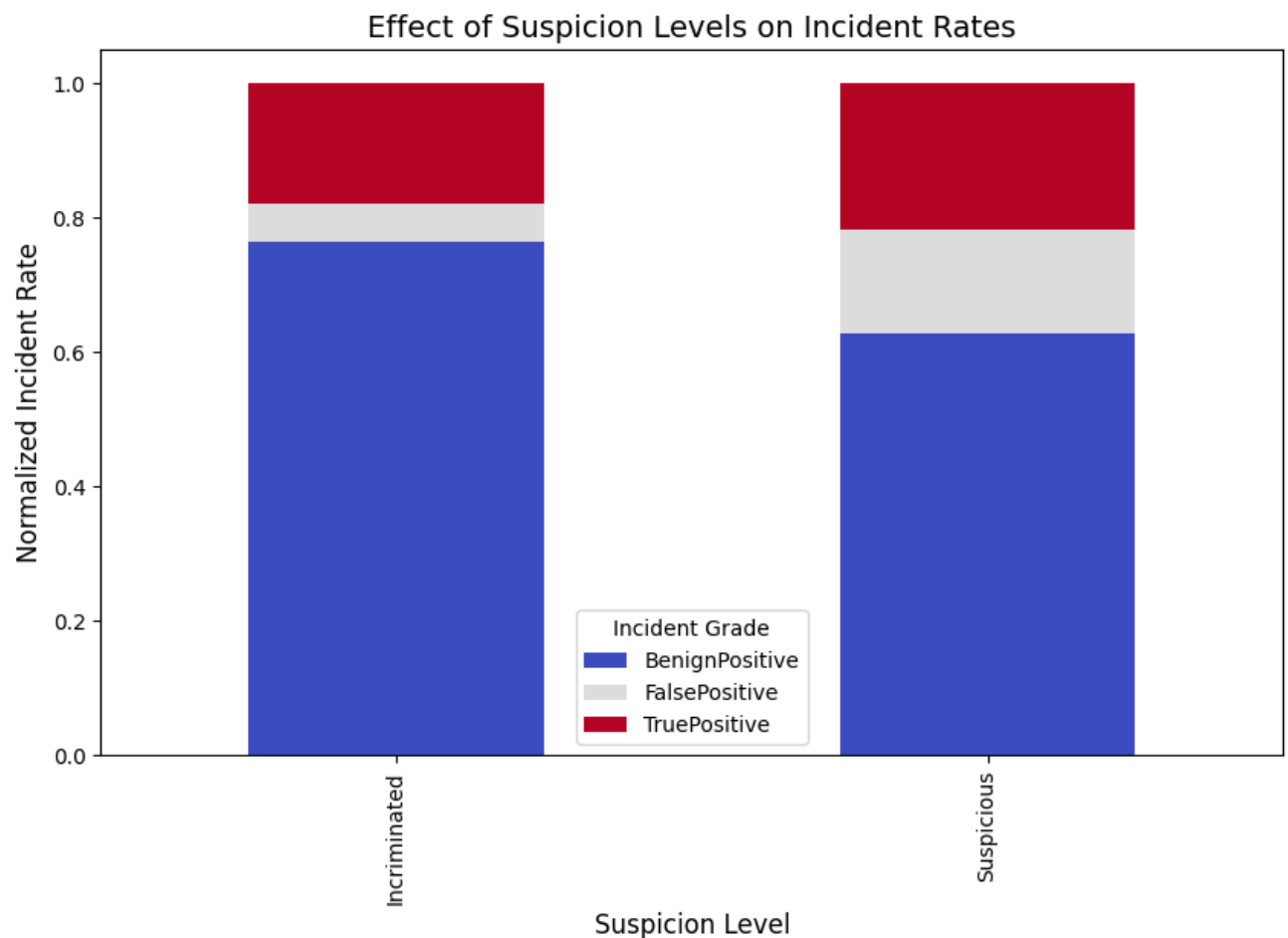


Figure 6: One class SVM results

Key Findings:

Reducing dimensionality improved computation time for SVM with detection efficiency as being unaffected.

Policy Frameworks

Prevention of insider threat is very relevant in organization and one of the ways is through corporate policies. In respect to its findings the study investigated the relationship that existed between the roles assigned and the levels of suspicion to arrive at the grades of the incidents in question as a yard stick of the policies in question.

Figure 7

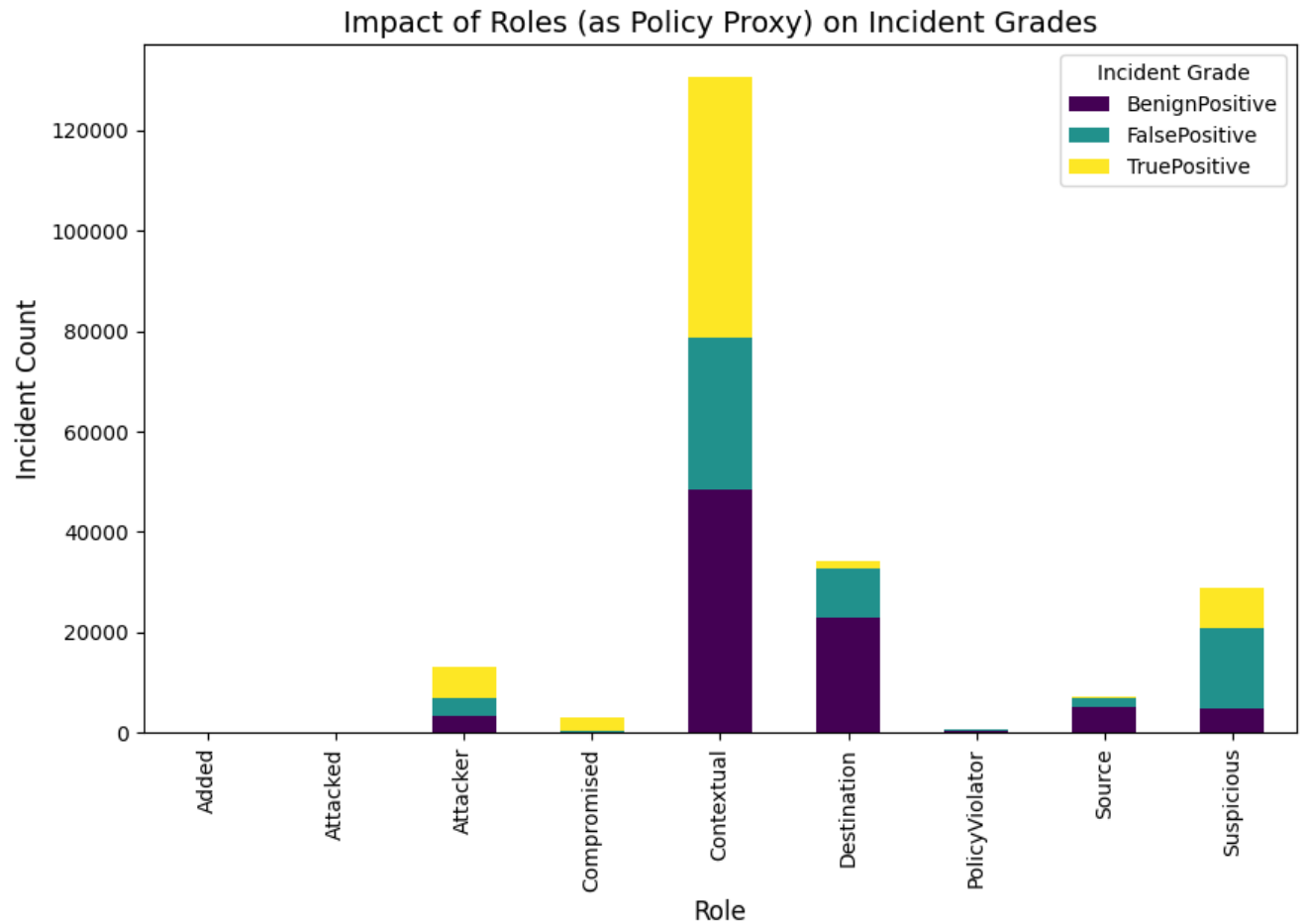
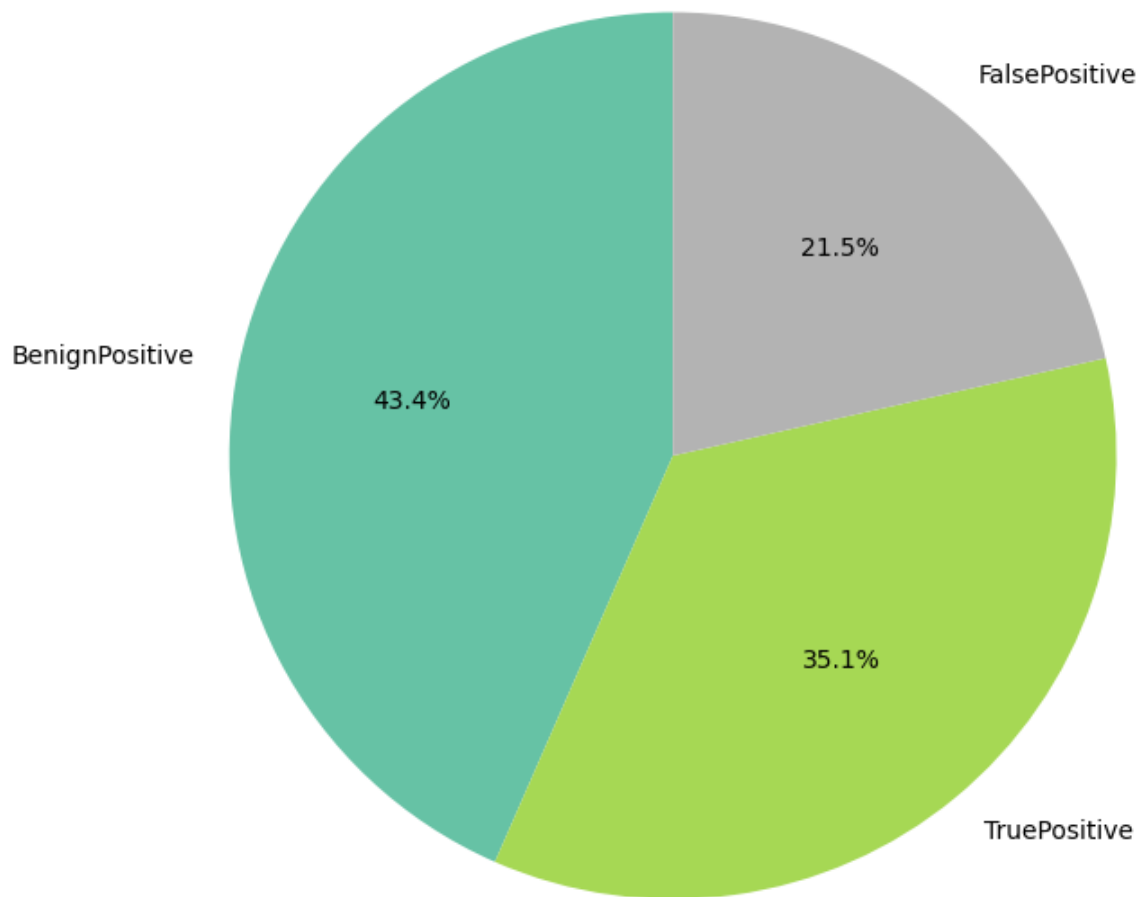


Figure 7: Impact of roles on incident grades

Key Insights:

1. Employees in high-risk activities experienced a relative rise in vulnerability to high end graders' mishap rates.
2. Higher suspicion levels were directly proportional to the actual generation of true positive cases.

Incident Grade Distribution

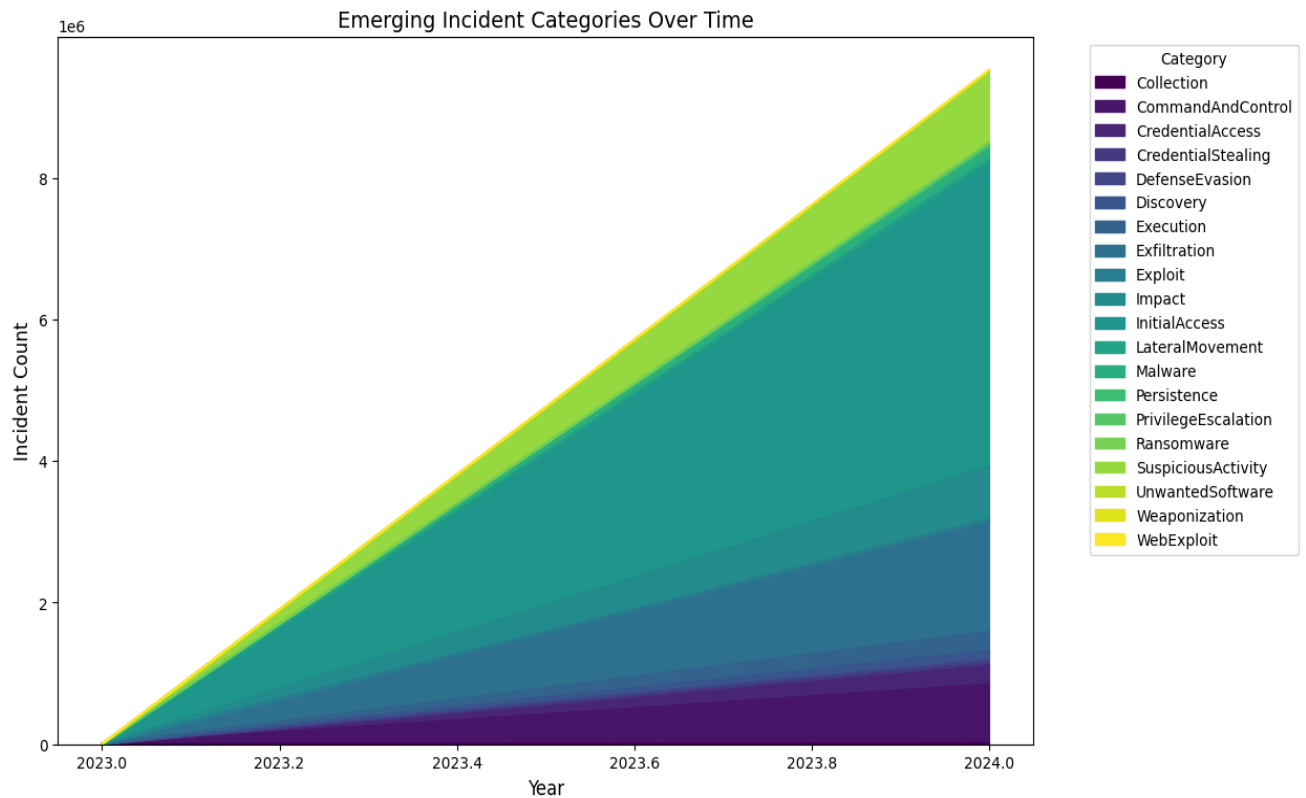


Implications

These observations stress the importance of the individual approaches, like the recurrent training in the tasks with high risk and increasing the level of suspicion in the framework of given activities.

6.1 Discussion

The findings of this project also indicate the benefits and drawbacks of the chosen experimental design and the applied machine learning algorithms. Specifically, the Isolation Forest algorithm was useful in identifying out-of-pattern behaviours which are suspicious within the constructed data set. Although the method proved to have high sensitivity to detect disparate instances, the application of the method was challenged with false positive results especially where the respective legitimate activities were one in a blue moon. This is an indication that there is a higher degree of detail of the model parameters that can be sought in order to garner increased specificity.



Moreover, the Random Forest classifier brought accurately estimated incident grades, proving its effectiveness as a supervised learning algorithm. However, its performance was highly sensitive to the characteristics of the training data. The aspects of data prejudice evident from incident grades analysis therefore result in slight biases favouring the majority class. Subsequent designs can overcome this by over or under sampling or by applying class weight adjustments into the learning algorithm. Furthermore, after using the Random Forest model it will be beneficial to try another model such as Gradient Boosting or XGBoost to improve prediction accuracy even more.

The utilization of method namely Principal Component Analysis (PCA) lead to the successful implementation of the dimensionality reduction, which means that the computational workload was cut down and the presentation of the obtained results was improved. However, the present study may have eliminated some features by limiting the number of principal components to four which may be important in deciding on the final choice of the model. There was need for a sensitivity analysis that would identify the right number of components for the purpose of optimizing this step. Additionally, because PCA is an unsupervised method, it is not always easy to link it directly with certain predictive objectives; terms and hence, it may be valuable to investigate other forms of supervised methods of dimensionality reduction including LDA.

The policy analysis alone served the purpose of actual recommendations on how the relationship between the organizational roles and the grades of the incidents is like and which roles are inclined more to higher-grade incidents. This analysis was helpful, nevertheless it largely depended on visualization, whereas additional statistical testing could enhance it. Such things as carrying out chi-square tests or correlation results could buttress any trends noticed quantitatively.

When placed in the context of the current body of knowledge, the results resemble other works noting the efficacy of anomaly detection and predictive analysis in security measurement. Still,

more could be done with benchmarks and datasets typically utilized in the subject matter of the project. These standards could then be used to compare the results as a way of achieving better validation of these methodologies used herein.

7 Conclusion and Future Work

This project set out to answer the research question: In high level, it is also important to describe how the various machine learning techniques can be applied to enhance the identification and parsing of security incidents in an organizational environment? The set objectives were as follows: The possibility of anomaly detection was to be realized to cater for this question; the possibility of possessing predictive models to be realized; the aspect of dimensionality reduction was to be achieved; policy analysis was to be achieved. This work has shown relatively clear achievement of these goals, specifically in identifying outliers by the Isolation Forest algorithm, and identifying the grades of incidents with the Random Forest classifier. Some of the important conclusion emerged out of these models were their effectiveness in identifying suspicious activities and understanding of the organizational developments of security threats.

The implications of the current research are for cybersecurity since the study can be used to suggest ways in which machine learning can be applied in the handling of cyber incidents. The findings suggest that use of automated systems would improve the efficiency of security functions and their reliability. However, issues like, false positives in anomaly detection, and biases in the models used in predictive modelling means that the process needs to be refined. Besides, the project analysed the given dataset only, and there was no opportunity to use any more datasets, which weakens the project's conclusions in some way. Further work can be done utilizing other datasets with the intention of enhancing model's generalization and versatility.

Moving to the future work, further work for the path that has been pursued in the paper relates to utilization of enhanced techniques, for instance, ensemble learning or deep learning, for the task of anomaly detection and classification. It would be beneficial to extend this type of work by constructing real-time systems that Incorporated ongoing learning and improvement to the findings of this research. Finally, synthesizing the results with academic and industry professionals could ensure that the proposed solutions are implemented and utilized by targeted organizations, which struggle with constantly changing security threats.

References

- Cappelli, D., Moore, A., & Trzeciak, R. (2012). *The CERT guide to insider threats: How to prevent, detect, and respond to information technology crimes (theft, sabotage, fraud)*. Addison-Wesley.
- Gavai, G., Sricharan, K., Gunning, D., Hanley, J., Singhal, M., & Rolleston, R. (2015). Supervised and unsupervised methods to detect insider threat from enterprise social and online activity data. *Journal of Wireless Mobile Networks Ubiquitous Computing and Dependable Applications*, 6(2), 47–63.
- Gamachchi, A., Sun, L., & Boztas, S. (2018). A graph-based framework for malicious insider threat detection. *arXiv preprint arXiv:1809.00141*. <https://arxiv.org/abs/1809.00141>

- Haidar, D., & Gaber, M. M. (2019). Data stream clustering for real-time anomaly detection: An application to insider threats. In O. Nasraoui & C. E. Ben N’Cir (Eds.), *Clustering methods for big data analytics: Techniques, toolboxes, and applications* (pp. 115–144). Springer International Publishing.
- Koutsouvelis, V., Shiaeles, S., Ghita, B., & Bendiab, G. (2020). Detection of insider threats using artificial intelligence and visualisation. In *Proceedings of the 2020 6th IEEE Conference on Network Softwarization (NetSoft)* (pp. 437–443). IEEE.
- Krichen, M., Lahami, M., Cheikhrouhou, O., Alroobaea, R., & Maâlej, A. J. (2020). Security testing of internet of things for smart city applications: A formal approach. In *Smart Infrastructure and Applications* (pp. 629–653). Springer.
- Larsen, K., Legay, A., Nolte, G., Schlüter, M., Stoelinga, M., & Steffen, B. (2022). Formal methods meet machine learning (F3ML). In T. Margaria & B. Steffen (Eds.), *Leveraging applications of formal methods, verification, and validation: Adaptation and learning* (pp. 393–405). Springer Nature Switzerland.
- Le, D. C., Zincir-Heywood, N., & Heywood, M. I. (2020). Analyzing data granularity levels for insider threat detection using machine learning. *IEEE Transactions on Network and Service Management*, 17(1), 30–44.
- Liu, L., De Vel, O., Chen, C., Zhang, J., & Xiang, Y. (2018). Anomaly-based insider threat detection using deep autoencoders. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 39–48). IEEE.
- Liu, L., De Vel, O., Han, Q.-L., Zhang, J., & Xiang, Y. (2018). Detecting and preventing cyber insider threats: A survey. *IEEE Communications Surveys & Tutorials*, 20(2), 1397–1417.
- Rauf, U., Shehab, M., Qamar, N., & Sameen, S. (2021). Formal approach to thwart insider attacks: A bio-inspired auto-resilient policy regulation framework. *Future Generation Computer Systems*, 117, 412–425.
- Sheykhkanloo, N. M., & Hall, A. (2020). Insider threat detection using supervised machine learning algorithms on an extremely imbalanced dataset. *International Journal of Cyber Warfare and Terrorism*, 10(1), 1–26.
- Tian, Z., Shi, W., Tan, Z., Qiu, J., Sun, Y., & Jiang, F. (2020). Deep learning and Dempster-Shafer theory-based insider threat detection. *Mobile Networks and Applications*, 25(2), 352–360.
- Yuan, F., Cao, Y., Shang, Y., Liu, Y., Tan, J., & Fang, B. (2018). Insider threat detection with deep neural network. In *Proceedings of the ICCS* (pp. 367–371).
- Yuan, S., & Wu, X. (2020). Deep learning for insider threat detection: Review, challenges, and opportunities. *arXiv:2005.12433*.
- Yuan, F., Cao, Y., Shang, Y., Liu, Y., Tan, J., & Fang, B. (2018). Insider threat detection with deep neural network. In *International Conference on Computational Science* (pp. 43–54). Springer.