Methodology

Studies suggest that two main detection methodologies are used. The first is anomaly detection, where the system makes a baseline profile of the normal system, network, or program activities. Any abnormality from the learnt baseline is labelled a malicious insider. The second is signature-based detection, which identifies a previously known malicious insider when such activities match the stored signature or rule-based protocol to model the used behaviours on the system. Figure 1 shows that the majority of existing solutions are based on anomaly-based detection or address insider threats as a classification issue (Al-Mhiqani, M.N. et al., 2020).

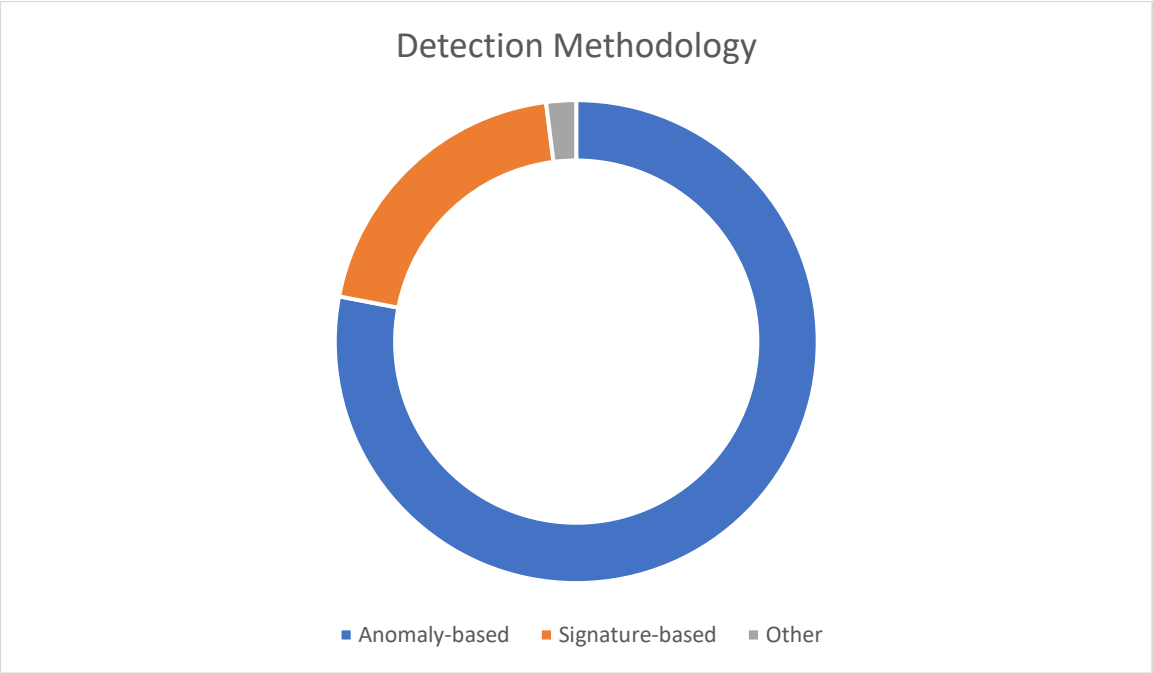

Figure 1: Detection Methodology

Building on this foundation, this study will apply sentiment analysis along with Anomaly-based techniques to improve on existing methods. We also draw inspiration from the anomaly detection methodology applied by Bin Sarhan, B. and Altwaijry, N., 2023. The adopted methodology is broken down into five crucial steps which are graphically displayed in Figure 2 below.

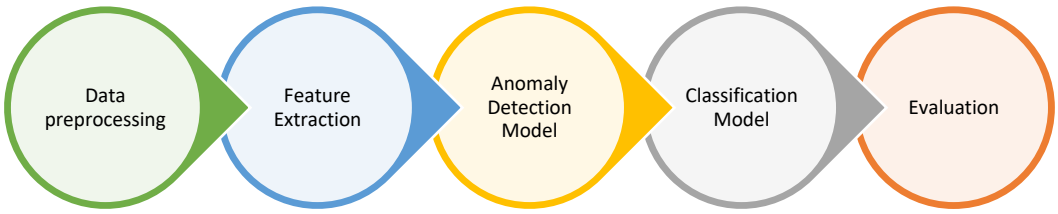

Figure 2: Methodological Process (Bin Sarhan, B. and Altwaijry, N., 2023)

Data Preprocessing

The process of data collection and preprocessing is essential in any machine learning paradigm. The collection and preprocessing of this data is not only essential for insider threat detection but also for various cyber security activities (Bin Sarhan, B. and Altwaijry, N., 2023). Furthermore, in order to apply sentiment analysis the datasets collected will need to undergo further data preprocessing such as the removal of noise through text cleaning, tokenization and handling of emotions and slang. These steps are explained in more detail during the implementation phase of the study.

Feature Extraction

Insider threat detection has a challenge in the feature engineering phase, where there's no standard for how many features should be extracted from each log source. As this varies based on the study. Traditionally, researchers manually extract features from the relationships between entities. They aim to get as many high-quality features as they can. Deep Feature Synthesis (DFS) is a notable automated tool for this, popular among scholars in recent years (Bin Sarhan, B. and Altwaijry, N., 2023). It mimics human intuition to engineer features for datasets often seen in databases or logs years (Bin Sarhan, B. and Altwaijry, N., 2023). We seek to apply the same feature extraction techniques in this study as research has shown how effective this method can be.

Anomaly Detection

The next step in the methodological approach we have chosen to implement is the Anomaly Detection phase, which will be conducted through the development of an anomaly detection model. The insider threat detection problem is a problem solved based on the availability of data sources related to the issue (Bin Sarhan, B. and Altwaijry, N., 2023). Several methods have been used to detect anomalies, with majority of these methods stemming from either supervised or unsupervised machine learning techniques (Bin Sarhan, B. and Altwaijry, N., 2023). With the availability or lack thereof of datasets, in this study, both supervised and unsupervised machine learning techniques will be applied in order to identify whether supervised or unsupervised machine learning techniques perform better and to identify which of these machine learning models perform best in detecting anomalies. Furthermore, the development of an ensemble model to further improve the accuracy of will be conducted. Some of the models that have been considered for this study include:

- Long Short-Term Memory – which was evaluated by Bin Sarhan, B. and Altwaijry, N., 2023 over a series of data log datasets and their studies suggested that the LSTM algorithm produced great results with an average accuracy of 93.85%. The same algorithm was recommended by Al-Mhiqani, M.N. et al., 2020 who suggested that the LSTM algorithm was well suited for classifying time series and that it could potentially be used to record temporal behavioural patterns.
- iForest – which was also evaluated by Bin Sarhan, B. and Altwaijry, N., 2023. Through their study, they evaluated that the iForest algorithm, tested by several log data datasets had an average accuracy of 82% in detecting anomalies.

By making use of these two unsupervised machine learning algorithms, we seek to detect anomalies before then classifying them before evaluation.

Classification

Once we have managed to detect these anomalies, we then need to classify them. During this phase of our methodological approach two supervised learning classification algorithms will be applied:

Support Vector Machine (SVM) and Random Forest. The two algorithms where also applied in a study by Bin Sarhan, B. and Altwaijry, N., 2023. Oladimeji et al 2019 also reviewed several classification techniques and the same algorithms were said to be effective in the study. Al-Mhiqani, M.N. et al., 2020 also applied the same models in their study with their SVM model producing an average accuracy of approximately 98% while their Random Forest model produced a Best AUC of 0.979.

We will further our study by implementing sentiment analysis techniques during this classification phase to improve on the performance of our predictions. By looking at how sentiment analysis has been implemented in similar studies, the techniques applied could be used in assessing the sentiments of insiders and the data collected from various data sources. The techniques implemented for this study are further explained in the implementation section.

Evaluation

Once we have trained and tested our model, we then need to evaluate its effectiveness. Currently, no standard has been set that addresses the evaluation of insider threat detection systems (Al-Mhiqani, M.N. et al., 2020), which is why the selection of the best detection methods is still a challenging task. For this study, we implement machine learning evaluation techniques in order to evaluate the effectiveness of our models. This will be done by looking at four factors: Accuracy, Precision, Recall and the F1 – Score. This evaluation method is a common model evaluation method that has been applied in several studies including studies by Bin Sarhan, B. and Altwaijry, N., 2023, as well as Al-Mhiqani, M.N. et al., 2020. Below is the list of equations that will be used to calculate the various evaluation metrics:

1. Accuracy = (TP + TN)/(TP + TN + FP + FN)
2. Precision = TP/(TP + FP)
3. Recall = TP/(TP + FN)
4. F1 – Score = (2 x precision x recall)/(precision + recall)

Conclusion

The methodology applied in several depends mostly on the type of study, intention and data available. Our methodology stems from techniques applied in several studies in the field of machine learning. Majority of the steps included are inherited from studies cantered around anomaly detection. The implementation of Sentiment analysis is then added to further improve on the already considered techniques as this is a gap that has not been looked at in the study. By applying sentiment analysis techniques in insider threat detection, we plan on improving the accuracy at which these threats can be detected. The above steps are further explained in the implementation phase that follows.

Reference