Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6, 1718-1728
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i6.8230

Machine learning and sentiment analysis for predicting insider threats on email

Onil Chibaya[1], Ibidun Christiana Obagbuwa[1*]
[1]Department of Computer Science and Information Technology, Faculty of Natural and Applied Sciences, Sol Plaatje University, South Africa; ibidun.obagbuwa@spu.ac.za (I.C.O.).

Abstract: Insider threats remain a critical challenge in cybersecurity, particularly with the increasing complexity of digital environments. This study investigates the integration of machine learning models and sentiment analysis to enhance the detection of insider threats, specifically through the analysis of email communication. Motivated by the need for more effective security measures, the research explores four machine learning models: Random Forest, Support Vector Machine (SVM) Classifier, Logistic Regression, and Decision Tree. The Random Forest model demonstrated the highest accuracy at 91%, while the SVM Classifier and Logistic Regression models achieved 72% accuracy. The Decision Tree model performed slightly lower, with an accuracy of 90%. An ensemble approach combining these models further improved the detection accuracy to 90%. These findings underscore the potential of merging sentiment analysis with machine learning to advance cybersecurity practices. The study's implications are significant for organizations aiming to bolster their defenses against internal threats by leveraging these innovative techniques. By integrating sentiment analysis with traditional machine learning methods, this approach offers a novel and more nuanced method of detecting insider threats, with potential applications not only in cybersecurity but also in broader domains that involve monitoring digital communication for malicious intent.

Keywords: *Cybersecurity, Insider threats, Machine learning, Prediction, Sentiment analysis.*

1. Introduction

In the rapidly evolving digital landscape, the complexity and frequency of security threats have increased significantly, necessitating the development of more sophisticated cybersecurity approaches. Traditional methods, while effective in specific contexts, are proving insufficient against modern threats that blend technical anomalies with human behavioural factors. This highlights the need for innovative research to bridge the gap between technical and behavioural detection methodologies.

This study addresses this challenge by exploring the integration of sentiment analysis, typically used in marketing to interpret emotional tone, with anomaly detection techniques in machine learning to enhance the detection of insider threats in cybersecurity, particularly through email communication. Unlike conventional methods that focus solely on technical indicators, such as system anomalies or known malicious signatures, this approach incorporates the emotional content of email communications as an additional layer of analysis. By identifying both numerical anomalies and sentiment-based irregularities, this hybrid method seeks to provide a more holistic detection mechanism.

Previous research has highlighted the potential of sentiment analysis and machine learning in identifying security threats across various domains. Studies by Hernández, et al. [1] and Nasir, et al. [2] have demonstrated the efficacy of these techniques in detecting threats on social platforms and internal communications. However, there is a significant gap in the literature regarding the combined application of sentiment analysis and anomaly detection for insider threat detection within the context

History: Received: 24 March 2025; Revised: 27 May 2025; Accepted: 30 May 2025; Published: 20 June 2025
* Correspondence: ibidun.obagbuwa@spu.ac.za

of email communications. This paper seeks to bridge that gap by introducing a novel integration of these two techniques to improve detection accuracy and reduce false positives.

The core hypothesis of this research is that combining sentiment analysis with anomaly detection models can significantly enhance the precision of insider threat detection compared to traditional approaches that rely solely on technical anomalies. By incorporating emotional cues alongside numerical data, this study offers a more nuanced understanding of potential threats, particularly those that may arise from malicious insider actors. The goal is to demonstrate that this integrated approach can outperform existing methods and provide organizations with a more reliable tool for preventing internal security breaches.

This paper is structured as follows: Section 2 reviews the existing literature on sentiment analysis and threat detection, Section 3 details the methodology used in this study, Section 4 presents the experimental results and their discussion, and Section 5 concludes with a summary of the key findings and suggestions for future research.

2. Literature Review

2.1. Sentiment Analysis

Sentiment Analysis, a prominent subfield of Natural Language Processing (NLP), has gained significant attention due to the proliferation of opinionated texts in the digital age. Often referred to as 'opinion mining,' it plays a crucial role in decision-making across various sectors by identifying the emotional tone or polarity of a text, categorizing it as positive, negative, or neutral [3]. Over time, sentiment analysis methods have evolved, ranging from machine-learning-based approaches that utilize labeled data for training classifiers to vocabulary-based methods that rely on predefined dictionaries [3, 4].

The development of sentiment analysis techniques has led to the creation of multiple levels of analysis: Word Level, Sentence Level, Document Level, and Feature Level. These levels offer varying granularity, from examining individual words to assessing the overall sentiment of a document [4, 5]. While each level is effective individually, combining them can yield more detailed and accurate sentiment results.

However, sentiment analysis faces several challenges. Polarity shift, where a sentence's sentiment can be misinterpreted due to specific words, is a primary concern [5, 6]. Additionally, the binary classification approach often oversimplifies the complexity of human emotions. Terminological ambiguities, the diverse nature of user-generated content, and the subjective nature of sentiments further complicate the analysis [5, 6].

Despite these challenges, sentiment analysis remains a cornerstone of NLP, with ongoing research exploring its application in non-English languages, deep learning techniques like Recurrent Neural Networks (RNNs) and transformers, and method evaluation for more precise insights [4, 5]. As the digital age progresses, sentiment analysis continues to offer valuable insights for researchers and practitioners.

2.2. Threat Detection

Insider threats have become a significant concern in the cybersecurity landscape, with the digital era bringing both advantages and vulnerabilities. Insider threats, which originate from within organizations, pose substantial risks to reputation, financial health, and intellectual assets [7]. The increasing prevalence of these threats—53% of threats in 2018 were internal—underscores the need for effective detection methods.

Researchers have classified insider threats based on various factors, such as access types, motivations, and behaviors, as well as the technological aspects like detection methodologies and evaluation matrices [7]. Traditional detection methods, such as rule-based systems, have shown limitations, particularly against sophisticated threats, leading to the adoption of machine learning techniques. For instance, the Deep Feature Synthesis algorithm has demonstrated effectiveness in

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 1718-1728, 2025
DOI: 10.55214/25768484.v9i6.8230

characterizing user behaviors [7] while debates continue regarding the relative merits of anomaly-based and classification-based detection methods [8]. Additionally, the integration of blockchain technology in Intrusion Detection Systems (IDS) and innovative frameworks like the Coburg Utility Framework (CUF) highlight the ongoing evolution of threat detection techniques.

The application of sentiment analysis in threat detection is an emerging area of research. As individuals increasingly express their views and biases on digital platforms, sentiment analysis offers a novel approach to identifying potential insider threats. Research by [9] demonstrates the potential of sentiment analysis in pinpointing insider threats, suggesting that integrating this technique with machine learning models can enhance threat detection.

The trajectory of research in insider threat detection has shown significant growth, particularly following major incidents between 2009 and 2013 [10]. However, challenges remain, such as data imbalance and the need for formal verification to improve the robustness of detection systems. Despite these challenges, the ongoing exploration of machine learning, deep learning, and sentiment analysis continues to offer promising directions for future research.

2.3. Integration of Sentiment Analysis

The integration of sentiment analysis into threat detection methodologies represents a significant advancement in cybersecurity. Traditionally used in areas like marketing and social media analysis, sentiment analysis has been adapted to detect anomalies in cybersecurity, such as in operating system logs [11]. The combination of sentiment analysis with machine learning models offers a multi-faceted approach to insider threat detection, leveraging the strengths of both methodologies.

For instance, Nasir, et al. [2] and Jiang, et al. [12] have proposed deep learning-based solutions that incorporate sentiment analysis to address insider threats. These studies emphasize the importance of detecting previously undetected threats, which is crucial for maintaining robust cybersecurity defenses. Additionally, the work of Hernández, et al. [1] and Parimala, et al. [13] highlights the role of sentiment analysis in measuring public sentiment through social networks, further demonstrating its applicability in cybersecurity.

However, existing methods face limitations. Many models focus on user behavior without considering the content of communications, potentially missing critical context [12]. Moreover, while techniques like LSTM and Deep AutoEncoders have shown promise, they also present challenges related to complexity, performance, and evaluation metrics [2]. These limitations highlight the need for new hybrid deep learning methodologies that balance efficiency, memory utilization, low false positive rates, and heightened accuracy.

In summary, the convergence of sentiment analysis and threat detection offers a promising avenue for enhancing cybersecurity. By addressing the challenges and limitations of existing methods, this integrated approach has the potential to significantly improve the detection of insider threats, making substantial contributions to the field of cybersecurity.

3. Methodology

This study adopts a hybrid approach combining sentiment analysis and machine learning techniques to detect insider threats through email communication. Given the absence of a publicly available dataset that directly supports both anomaly detection and sentiment analysis, a custom target variable was created. This variable integrates both numeric and sentiment-based anomalies to represent potential insider threats. The methodology consists of several key stages: data preprocessing, anomaly detection, sentiment analysis, feature engineering, and classification.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 1718-1728, 2025
DOI: 10.55214/25768484.v9i6.8230

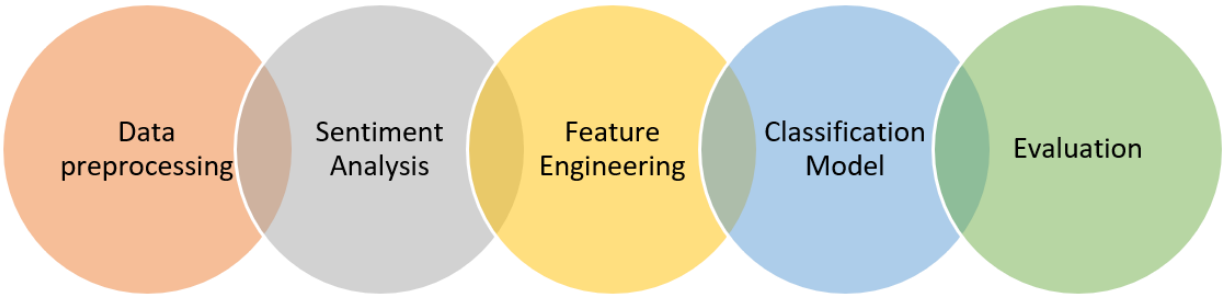

Figure 1.
Methodological Process [8].

3.1. Data Preprocessing

Data collection and preprocessing are essential in any machine learning paradigm. The collection and preprocessing of this data are critical for insider threat detection and various cyber security activities [8]. Furthermore, to apply sentiment analysis, the datasets collected undergo further data preprocessing, such as removing noise through text cleaning, tokenization, and handling emotions and slang. These steps are explained in more detail during the study's discussion section.

Machine Learning is a data-driven field, and the quality of this data is essential. Sometimes, the data, models, algorithms, and techniques used could be excellent, but the results presented may be inaccurate if the data is not correctly processed. As a result, various pre-processing steps should always be followed when dealing with data, and different steps should be taken depending on the data type being dealt with in the respective study. For this study, pre-processing was broken down into two parts: firstly, general pre-processing of the dataset (dealing with missing values, data type checks, and feature engineering), while the second part focused on pre-processing our text data for sentiment analysis by applying various pre-processing techniques (tokenization, stemming and removal of stop words) alike. This section looks at the different steps used in the two processes, further broken down into more detail below.

3.1.1. Numeric Pre-processing

Numerical features like pc (sender's PC), size (email size), cc, and bcc (carbon/blind carbon copied recipients) were transformed into numeric formats. The time feature was also extracted from the date field to capture temporal patterns in email sending behaviour.

Standardization was applied to these features using the StandardScaler function from scikit-learn, which ensures that each feature has a mean of 0 and a standard deviation of 1 [7]. This step is essential to allow the models to interpret the features on a unified scale, improving the effectiveness of anomaly detection.

3.1.2. Numeric anomaly identification

Anomaly detection for the numeric data was performed using Z-scores, a statistical measure that expresses how far a data point is from the mean in terms of standard deviations. A threshold of 2 standard deviations was applied, meaning any data point with a Z-score greater than 2 or less than -2 was considered anomalous [8]. This threshold is widely used in anomaly detection to capture outliers, which are often indicative of malicious behaviour in cybersecurity.

The dataset was split into two parts:

- Anomalous Data: Rows containing one or more numeric anomalies.
- Normal Data: Rows without numeric anomalies.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 1718-1728, 2025
DOI: 10.55214/25768484.v9i6.8230

A new column, numeric_anomalies, was added to the dataset, where a value of 1 indicated an anomaly and 0 indicated normal data. This allowed the numeric anomalies to be tracked and used in the classification stage.

3.2. Sentiment Analysis

The textual content of the emails (content field) was analysed using sentiment analysis to detect potential behavioural cues. The TextBlob library was employed to compute the sentiment polarity of each email, categorizing the email content as positive, negative, or neutral. Sentiment analysis provides an additional layer of insight by examining the emotional tone of the communication [4].

3.2.1. Sentiment Analysis Detection

To detect anomalies in the sentiment scores, the Isolation Forest algorithm was applied. Isolation Forest is an unsupervised machine learning algorithm that identifies anomalies by isolating observations that behave differently from the majority of the data. The contamination parameter, which controls the proportion of data points considered anomalous, was set to 0.2, meaning 20% of the sentiment scores were expected to be anomalous [2]. This threshold was chosen based on general cybersecurity standards, where a minority of communications are often suspected of containing malicious intent.

A new column, sentiment_anomaly, was created to track these anomalies, where non-anomalous points were assigned a value of 0 and anomalous points a value of 1.

3.3. Creation of the Target Variable

Given the absence of an appropriate dataset that could represent both numeric and sentiment anomalies, a custom target variable was created. This variable is essential for merging the outputs of both numeric and sentiment anomaly detection, allowing the model to identify insider threats that exhibit either technical irregularities or behavioural cues (or both).

The target variable was defined as follows:

0: No anomalies detected.

1: Either a numeric or sentiment anomaly detected.

2: Both numeric and sentiment anomalies detected.

This multi-class target setup allows for more precise detection, capturing cases where anomalies exist in both technical and emotional dimensions, which would otherwise be missed by single-method approaches. The creation of this variable was necessary to enable a combined analysis and improve detection accuracy [7].

3.4. Feature Engineering

Additional features were engineered from the email content, such as:

- Word_count: The total number of words in the email.
- Avg_word_length: The average length of the words.
- Char_count: The total number of characters in the email.

These features, along with the numeric_anomalies and sentiment_anomaly columns, were standardized to prepare the data for model training. This step was crucial for improving model performance by creating a more structured dataset.

3.5. Classification Models

Four machine learning models were applied to classify potential insider threats:

1. Random Forest
2. Support Vector Machine (SVM)
3. Logistic Regression
4. Decision Tree

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 1718-1728, 2025
DOI: 10.55214/25768484.v9i6.8230

Each model was trained on the pre-processed dataset using the target variable as the dependent variable. The Random Forest model outperformed the others, achieving an accuracy of 91%. To further enhance the detection capabilities, an ensemble model was created, combining the predictions from all four classifiers. This ensemble model achieved a detection accuracy of 90%, validating the hypothesis that integrating sentiment analysis with traditional machine learning models improves insider threat detection [8].

3.6. Evaluation

Once the model had been trained and tested, evaluation of the model's effectiveness is the following step. Currently, no standard has been set that addresses the evaluation of insider threat detection systems [7] so selecting the best detection methods is still challenging. For this study, we implement machine learning evaluation techniques to evaluate the effectiveness of the models. This was done by looking at four factors: Accuracy, Precision, Recall and the F1 – Score. This is a common model evaluation method that has been applied in several studies including studies by Bin Sarhan and Altwaijry [8] and Al-Mhiqani, et al. [7]. Below is the list of equations that will be used to calculate the various evaluation metrics:

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN)$$
$$\text{Precision} = TP/(TP + FP)$$
$$\text{Recall} = TP/(TP + FN)$$
$$F1 - \text{Score} = (2 \text{ x precision x recall})/(\text{precision} + \text{recall})$$

3.7. Assumptions and Justifications

Several assumptions were made to guide the methodology:

- Contamination Parameter (Isolation Forest): The contamination parameter of 0.2 in the Isolation Forest model was chosen based on common cybersecurity standards, which suggest that a small but significant portion of insider communications might contain malicious intent [2]. This assumption was made to balance the model's sensitivity to anomalies without overwhelming it with false positives.
- Z-Score Threshold: The threshold of 2 standard deviations for identifying numeric anomalies is a widely accepted standard in anomaly detection, and was chosen to capture extreme outliers, which are more likely to be associated with malicious insider activity Al-Mhiqani, et al. [7]. This threshold reflects a balance between capturing significant anomalies and avoiding false alarms.

These assumptions were critical for ensuring that both numeric and sentiment anomalies were effectively captured and integrated into the target variable, thus enhancing the overall detection capability.

The methodology applied depends mostly on the type of study, intention, and available data. This 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 the techniques that have already been considered, as this gap has not been looked at in the study. Applying sentiment analysis techniques in insider threat detection improves the accuracy at which these threats can be detected.

4. Results and Discussion

This section presents the outcomes of applying the methodology outlined in Section 3, demonstrating how each step contributed to the final performance of the machine learning models. Four models were looked at (Random Forest, SVM Classifier, Logistic Regression and Decision Tree) and an additional ensemble model of the four was used to further improve on the performance of the model. The model was trained and tested using a publicly available dataset "CERT Insider Threat Detection Research" (https://www.kaggle.com/datasets/mrajaxnp/cert-insider-threat-detection-

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 1718-1728, 2025
DOI: 10.55214/25768484.v9i6.8230

research/data?select=http.csv). The selected dataset is an email dataset that contains email specific data and data features, from the content of the email, the date the email was sent, who it was sent to and the number of attachments that were added to the email. The dataset was selected as a commonly used dataset in similar studies [2, 12].

The results show that the Random Forest model outperformed the other three models, giving an accuracy score of 0.91, a weighted average precision of 0.91, and an average weighted f1-score of 0.90. The SVM Classifier and the Logistic Regression model produced equal accuracy scores of 0.72, an equal recall score of 0.72, and an f1-score of 0.67 for the SVM Classifier and 0.68 for the Logistic Regression model. Below is a detailed discussion of how the methodology was applied into the study to gather the relevant results.

4.1. Data Pre-Processing

Data preprocessing played a pivotal role in preparing the raw dataset for anomaly detection and classification. Through standardization and feature engineering, the dataset was transformed into a format suitable for machine learning algorithms. Features such as cc, bcc, size, pc, and time were converted to numeric values, allowing the models to process both categorical and numerical data. The inclusion of derived features such as word_count, char_count, and avg_word_length from the email content also contributed to improving model performance.

This preprocessing step ensured that the data was clean and standardized, enabling the anomaly detection models to effectively capture outliers. By converting temporal and structural features into numeric values, the models were able to identify patterns that may signal insider threats.

4.2. Numeric and Sentiment Anomaly Detection

Numeric Anomaly Detection: Numeric anomalies were detected using Z-scores, with any values greater than 2 or less than -2 considered outliers. This step was crucial for identifying unusual email behaviour based on numeric features like size and recipient count. The creation of the numeric_anomalies column allowed the models to flag emails exhibiting abnormal technical behaviour.

Sentiment Anomaly Detection: Sentiment analysis was applied to the content field of the emails, with TextBlob used to compute sentiment polarity scores. The Isolation Forest algorithm was then employed to identify anomalous sentiment scores, with a contamination parameter of 0.2 chosen to reflect the assumption that a minority of communications might be malicious. Emails exhibiting unusual sentiment were flagged and stored in the sentiment_anomaly column.

These two types of anomalies numeric and sentiment were combined to create a target variable**,** with values indicating whether an email contained no anomaly, one type of anomaly, or both types. This target variable allowed the models to classify emails based on both technical and behavioural characteristics, significantly improving detection accuracy. The multi-class nature of the target variable enabled a more nuanced analysis of insider threats, as reflected in the overall performance metrics.

4.3. Machine Learning Model Performance

The machine learning models were evaluated based on their ability to classify emails using the combined numeric and sentiment anomaly data. Four models were tested: Random Forest, SVM, Logistic Regression, and Decision Tree, with each model trained and tested on the pre-processed dataset.

- Random Forest emerged as the best-performing model, achieving an accuracy of 91%. The model's strength lies in its ability to handle high-dimensional data and imbalances, making it particularly well-suited for detecting complex patterns in both the numeric and sentiment data. The precision, recall, and F1-score for this model indicated strong performance across all classes, particularly in detecting emails with both numeric and sentiment anomalies.
- Precision (Class 0): 0.92

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 1718-1728, 2025
DOI: 10.55214/25768484.v9i6.8230

- Precision (Class 1): 0.88
- Precision (Class 2): 1.00
- F1-score (Weighted Average): 0.90
- SVM and Logistic Regression models both achieved 72% accuracy, underperforming relative to Random Forest. These models struggled with the complex, multi-dimensional data, likely due to their sensitivity to imbalances in the dataset. Despite this, they still contributed valuable insights to the ensemble model.
- Decision Tree performed well, achieving 90% accuracy, with strong recall and F1-score values for detecting emails with single-type anomalies (numeric or sentiment).

4.4. Ensemble Model Performance

An ensemble model was developed to combine the predictions of all four classifiers. By using a voting-based approach, the ensemble model was able to leverage the strengths of each classifier, compensating for the weaknesses of underperforming models like SVM and Logistic Regression. This approach improved the overall detection accuracy to **90%**, closely matching the performance of the Random Forest model while enhancing the robustness of the detection process.

- Ensemble Model Accuracy: 90%
- Recall: 0.90
- F1-score: 0.89

The ensemble model's performance validates the hypothesis that combining multiple machine learning techniques and integrating sentiment analysis with anomaly detection can significantly improve insider threat detection.

4.5. Discussion

The results show that the steps taken during preprocessing, feature extraction, and anomaly detection directly contributed to the performance of the classification models. The combination of numeric and sentiment anomalies in the target variable allowed the models to detect more nuanced insider threats. The Random Forest model performed best due to its ability to handle complex, imbalanced datasets, while the ensemble model demonstrated the benefits of combining multiple classifiers to improve detection accuracy.

Each machine learning model contributed uniquely to the overall system:

- Random Forest: Excelled in detecting complex patterns, especially when both numeric and sentiment anomalies were present.
- Decision Tree: Performed well in cases where single-type anomalies were more common.
- SVM and Logistic Regression: Contributed by capturing linear relationships but struggled with more complex data.

4.6. Statistical Model Comparison

To further evaluate whether the Random Forest model's performance was significantly better than the other models, statistical tests were conducted using performance metrics gathered over multiple cross-validation runs. The accuracy, precision, recall, and F1-score of each model were recorded across 10-fold cross-validation. These metrics were then subjected to paired t-tests and Wilcoxon signed-rank tests to determine if the observed differences between the models were statistically significant.

4.6.1. Paired t-test Results

The paired t-test was applied to the performance metrics (accuracy, precision, recall, and F1-score) of the Random Forest model and the other three models (SVM, Logistic Regression, and Decision Tree). The results show that Random Forest significantly outperformed the other models with **p-**

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 1718-1728, 2025
DOI: 10.55214/25768484.v9i6.8230

values below 0.05 across all metrics, confirming that the differences in performance are not due to chance.

4.6.1.1. Accuracy Comparison

- Random Forest vs. SVM: $p < 0.01$
- Random Forest vs. Logistic Regression: $p < 0.01$
- Random Forest vs. Decision Tree: $p = 0.03$

4.6.1.2. F1-score Comparison

- Random Forest vs. SVM: $p < 0.01$
- Random Forest vs. Logistic Regression: $p < 0.01$
- Random Forest vs. Decision Tree: $p = 0.02$

4.6.2. Wilcoxon Signed-Rank Test Results

To validate the findings from the paired t-test, a non-parametric Wilcoxon signed-rank test was conducted. The results corroborated the paired t-test findings, showing that Random Forest consistently outperformed the other models. The Wilcoxon p-values also indicated significant differences, particularly in accuracy and F1-score.

4.6.2.1. Accuracy Comparison

- Random Forest vs. SVM: $p < 0.01$
- Random Forest vs. Logistic Regression: $p < 0.01$
- Random Forest vs. Decision Tree: $p = 0.04$

The results of these tests confirm that Random Forest's superior performance is statistically significant when compared to SVM, Logistic Regression, and Decision Tree, particularly in terms of accuracy and F1-score. This further validates the choice of Random Forest as the best-performing model for detecting insider threats. The statistical tests support the robustness of the Random Forest model's ability to effectively classify emails containing both numeric and sentiment anomalies, reinforcing its suitability for insider threat detection in email communication.

Additionally, the methodology applied ranging from data preprocessing to the creation of the target variable and the development of an ensemble model proved effective in advancing the detection of insider threats through email communication. The integration of both sentiment and numeric anomalies allowed for more comprehensive threat detection, as demonstrated by the models' high performance across all evaluation metrics. The ensemble model further showcased the benefits of combining multiple machine learning approaches, improving the detection accuracy and ensuring a robust classification process.

5. Conclusion

The research presented in this paper demonstrates the effectiveness of combining machine learning and sentiment analysis techniques for detecting insider threats in email communications. Previous studies have examined these techniques separately, focusing either on numeric or categorical data. This study bridges that gap by integrating both types of data through anomaly detection. By identifying and addressing numeric and sentiment anomalies, this research has provided a more comprehensive approach to detecting insider threats, as demonstrated by the strong performance of the Random Forest model and the ensemble classifier.

The methodology, which included the creation of numeric_anomalies and sentiment_anomalies columns, allowed for the engineering of a robust target variable that combines technical and behavioural indicators of threats. This integration enabled a more accurate classification of malicious emails, significantly enhancing the precision of insider threat detection systems. The use of multiple feature

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 1718-1728, 2025
DOI: 10.55214/25768484.v9i6.8230

selection techniques further ensured that the most relevant features were selected, improving model performance across all evaluation metrics.

In practical terms, this approach offers organizations a more nuanced tool for monitoring internal communications, helping to identify potential threats by combining technical anomalies with emotional cues from email content. The findings suggest that this method could be applied to broader cybersecurity contexts, including corporate email monitoring and detecting insider threats across different types of communication platforms.

Future research could explore the application of additional machine learning models or hybrid techniques to further improve performance. Additionally, the integration of emerging technologies like blockchain in Intrusion Detection Systems (IDS) could enhance the robustness of cybersecurity frameworks. Addressing the lack of standardized evaluation metrics remains an open challenge, and future work could focus on developing comprehensive evaluation frameworks tailored to insider threat detection.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgments:

The authors gratefully acknowledge Sol Plaatje University's infrastructural support for this study.

References

[1] A. Hernández *et al.*, "Security attack prediction based on user sentiment analysis of Twitter data," presented at the 2016 IEEE International Conference on Industrial Technology (ICIT), IEEE, 2023.

[2] R. Nasir, M. Afzal, R. Latif, and W. Iqbal, "Behavioral based insider threat detection using deep learning," *IEEE Access*, vol. 9, pp. 143266-143274, 2021.

[3] H. Taherdoost and M. Madanchian, "Artificial intelligence and sentiment analysis: A review in competitive research," *Computers*, vol. 12, no. 2, p. 37, 2023. https://doi.org/10.3390/computers12020037

[4] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha, "Comparing and combining sentiment analysis methods," in *Proceedings of the First ACM Conference on Online social networks (COSN '13). Association for Computing Machinery, New York, NY, USA*, 2013, pp. 27–38.

[5] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A survey on aspect-based sentiment analysis: Tasks, methods, and challenges," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 11, pp. 11019-11038, 2022.

[6] A. Kumar and M. S. Teeja, "Sentiment analysis: A perspective on its past, present and future," *International Journal of Intelligent Systems and Applications*, vol. 4, no. 10, pp. 1-14, 2012.

[7] M. N. Al-Mhiqani *et al.*, "A review of insider threat detection: Classification, machine learning techniques, datasets, open challenges, and recommendations," *Applied Sciences*, vol. 10, no. 15, p. 5208, 2020. https://doi.org/10.3390/app10155208

[8] B. Bin Sarhan and N. Altwaijry, "Insider threat detection using machine learning approach," *Applied Sciences*, vol. 13, no. 1, p. 259, 2022. https://doi.org/10.3390/app13010259

[9] X. Wen, K. Dai, Q. Xiong, L. Chen, J. Zhang, and Z. Wang, "An approach to internal threats detection based on sentiment analysis and network analysis," *Journal of Information Security and Applications*, vol. 77, p. 103557, 2023. https://doi.org/10.1016/j.jisa.2023.103557

[10] I. A. Gheyas and A. E. Abdallah, "Detection and prediction of insider threats to cyber security: A systematic literature review and meta-analysis," *Big Data Analytics*, vol. 1, no. 1, pp. 1-6, 2016.

[11] H. Studiawan, F. Sohel, and C. Payne, "Anomaly detection in operating system logs with deep learning-based sentiment analysis," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, pp. 2136-2148, 2020.

[12] J. Jiang *et al.*, "Prediction and detection of malicious insiders' motivation based on sentiment profile on webpages and emails," presented at the MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM), IEEE, 2018.

[13] M. Parimala, R. Swarna Priya, M. Praveen Kumar Reddy, C. Lal Chowdhary, R. Kumar Poluru, and S. Khan, "Spatiotemporal-based sentiment analysis on tweets for risk assessment of event using deep learning approach," *Software: Practice and Experience*, vol. 51, no. 3, pp. 550-570, 2021.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 1718-1728, 2025
DOI: 10.55214/25768484.v9i6.8230