National College of Ireland

Insider Threat Detection using Ensemble and Sequential Models

MSc Practicum part-2
MSc In Cyber Security

Arun Joy
Student ID: 23201592

School of Computing
National College of Ireland

Supervisor: Mr Vikas Sahni

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Arun Joy		
Student ID:	23201592		
Programme:	MSc In Cyber Security	**Year:**	2024-2025
Module:	MSc Practicum part-2		
Supervisor:	Mr Vikas Sahni		
Submission Due Date:	12/12/2024		
Project Title:	Insider Threat Detection using Ensemble and Sequential Models		
Word Count:	6236	**Page Count:**	20

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: ARUN JOY

Date: 12/12/2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
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.	□

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 Threat Detection using Ensemble and Sequential Models

Arun Joy
23201592

Abstract

Insider threats are the major cybersecurity risks to the organisations that causing damage. Current detection approaches rely on predetermined criteria struggle to recognise small behavioural deviations. This insider problem research approaches by analysing behavioural patterns and anomalies within the user activity data. In order to solve this problem, this research used the CERT Insider Threat dataset and sophisticated machine learning algorithms to find unusual email communication patterns. This study used advanced algorithms like Random Forest, Isolation Forest, LSTM, GRU and Stacking Ensembles with feature engineering techniques including time-based and textual evaluation. The comparative study demonstrated that ensemble learning approaches, particularly the Stacking Classifier, significantly increased the detection accuracy when compared to traditional methods. These findings support the body of research on machine learning's effectiveness in anomaly detection and highlight the value of hybrid models in enhancing insider threat identification. In practice, this method gives the businesses a strong foundation for anticipating and proactively identifying hazards.

1 Introduction

A major problem in cybersecurity is insider threats, which occur when someone with authorised access do careless or malicious activities that compromise the security. According to the Cybersecurity Insiders' 2023 Insider Threat Report[1], 74% of organisations are at least a little exposed to insider attacks, underscoring the essential need for effective detection systems. Traditional approaches frequently depend on human monitoring or rule-based systems, which have limited scalability and responsiveness to changing threat patterns. A realistic solution is provided by developments in the machine learning (ML), which make it possible to analyse behavioural data intelligently and automatically in order to spot any dangers. This study examines the effectiveness of ML in improving detection skills using a publicly accessible dataset from Kaggle[2] that includes actual insider threat scenarios.

Detecting insider threats is essential for safeguarding confidential company information and guaranteeing legal compliance. The limits of static systems are addressed by the use of machine learning techniques, which offer scalable and dynamic solutions that can adjust to a

[1] https://securityboulevard.com/2023/04/insider-threat-statistics-for-2023-reports-facts-actors-and-costs/

[2] https://www.kaggle.com/code/alabiobin/insider-threat-detection/

variety of attack vectors. By concentrating on feature engineering and algorithm optimisation, this work closes a gap in existing research about the use of machine learning in actual insider threat scenarios and expands on earlier studies. Reduced detection time, increased accuracy and proactive threat mitigation are some possible advantages of this work that might strengthen organisational security.

1.1 Research question:

How can anomaly detection algorithms be optimised to improve the accuracy and reduce the false positive rate in detecting insider threats within the organization?

The study aims to answer this question by doing the following objectives:

- Examine current research and techniques for detecting insider threats.
- Create and put into use machine learning models, which include categorisation algorithms, to detect insider threats.
- Evaluate measures like accuracy, precision, recall and F1-score to compare the models' performance.

The size of the dataset and the difficulties in extending results to other organisations with different security postures are the limitations of this study. The accurate representation of the simulated threat situations and the accuracy of the dataset's labels are examples of assumptions.

1.2 Report Structure

This is how the rest of the paper is organised:

- Literature Review: Examines earlier research and machine learning applications for detecting insider threats.
- Methodology: Describes the feature selection, data preprocessing, and machine learning techniques used.
- Design and Implementation: Describes the steps that carried out and the final stage of implementation.
- Findings and Discussion: Outlines and assesses the results.
- Conclusion: Provides an overview of the contributions and suggests areas for further study.

The goal of this research is to fill a key gap in current cybersecurity frameworks by providing a scalable and efficient machine learning-based method for insider threat identification.

2 Related Work

The following describes and critically assesses some of the core publications in terms of their relevance to the project and the research issue. This evaluation is based on how well the study addresses challenges of accuracy and the false positives in anomaly identification for insider threat detection.

The article by Sarhan, B.B. and Altwaijry, N. (2022) have looked at employing advanced machine learning approaches such as deep learning and ensemble models to detect these dangers. Deep Feature Synthesis (DFS) is used to automate feature extraction, while dimensionality reduction techniques whereas PCA assist handle the enormous datasets created. Models that including anomaly detection and classification algorithms like SVM have demonstrated great accuracy in detecting insider threats particularly in unbalanced datasets such as CERT. SMOTE oversampling enhances recollection, although frequent retraining is still required for system robustness.

Al-Mhiqani et al. (2020) discusses a review that says the insider threats have been characterised by access level, motive and attack tactics with real-world incidents analysed to better understand insider behaviour. Machine learning and behaviour analysis are often used in detection systems, but lack of real-world datasets and study biases continue to be barriers. The study identifies information gaps, highlights detection issues and recommends future research to improve detection methods and security measures.

Yuan, S. and Wu, X. (2020) put forward research that indicates the deep learning models, such as RNNs and GNNs, can enhance detection by exploiting user behaviour patterns. However, obstacles are remaining such as insufficient labelled data, adaptive attacks and imbalanced datasets. Future research areas will focus on techniques such as few-shot learning, self-supervised learning and multimodal approaches to improve detection capabilities.

The article published by Nasir et al. (2021) says that the traditional approaches like access restrictions and behaviour monitoring are restricted, requiring the employment of advanced methods such as deep learning (DL). DL-based models, particularly those that use methods like as LSTM and autoencoders, provide higher accuracy and fewer false positives for identifying threatening behaviour. This study analyses many machine learning models, including LSTM-CNN and Random Forest and shows that the suggested DL technique performs well in detecting insider threats via behavioural analysis.

T.S Naicker and Brett V.N (2021) review a literature that recognises two main approaches: abuse detection, which uses known attack signatures and anomaly detection, which finds departures from typical behaviour. Random Forest, K-Nearest Neighbours and Multi-Layer Perceptron are among the most popular algorithms with Random Forest outperforming the others. Psychometric analysis, such as the Big Five Personality Traits has the potential to improve detection by include behavioural insights. Studies highlight the limits of static datasets, implying that live situations and bigger datasets might strengthen detection algorithms. Future research should focus on combining psychometric data and increasing computing resources.

The journal article by Alzaabi, F.R. and Mehmood, A. (2024) says that traditional machine learning approaches are limited in tackling the complex nature of insider threats that requiring the use of advanced methods like as deep learning and NLP. These current approaches, particularly in time-series analysis and big language models that improve anomaly identification by analysing textual and contextual data. Studies on the CMU CERT dataset indicate promising results for detecting opposed insiders. As insider threats grow, the ongoing innovation in AI-driven solutions is critical for successful detection and mitigation.

The conference paper published by Yuan et al. (2018) presents a variety of computational strategies for detecting insider threats, including both classical and machine learning methods. Although this article covers a wide range of approaches, its vast reach may limit its direct applicability to our specific study topic. The lack of emphasis on algorithmic improvement or false positive reduction may make it less valuable than earlier studies.

The research reviewed by Janjua et al. (2020) emphasises the rising issue provided by insiders who may bypass security measures and get access to sensitive data. Several studies highlight the usefulness of supervised machine learning techniques for detecting malevolent insiders, with language and behavioural analysis emerging as promising approaches. Salima et al. argue that supervised algorithms have higher recall than semi-supervised and unsupervised approaches. The use of text analysis and basic models, such as AdaBoost has shown great accuracy in detecting fraudulent emails. However, issues persist with tiny datasets and future research advises investigating deep learning classifiers for enhanced performance.

Caterina U and Antoine M (2021) says that the formal approaches provide strict accuracy for safety-critical systems and are widely used in sectors such as aviation. However, its application to machine-learned software, which is currently regarded vital is restricted and faces hurdles in terms of soundness, accuracy and scalability. SMT, optimisation and abstract interpretation are the most often used techniques for verifying neural networks, however some approaches also target SVMs, decision trees, training and data preparation. Industry need is driving research in this field, and future directions will focus on developing verification methodologies for machine-learned components.

The research by Aaron T and Samuel K and Brian H (2017) is on using unsupervised deep learning models like DNNs and LSTMs to detect abnormalities in user behaviour in streaming system data. These models constantly adapt to shifting data patterns and produce interpretable results, assisting human analysts. Studies with the CERT Insider Threat Dataset reveal that deep learning models outperform standard baselines in anomaly detection tasks.

Another article Gavai et al. (2015) examines that the security flaws in the devices that connected to the internet. It highlighting the weaknesses in the defences of the internet connected devices and the requirement for better network protocols. It assesses the weaknesses of current defences, its remedies and examines the typical attack pathways. The results show that connectivity issues and the limitation in the resource that make it difficult to secure the IoT devices. To improve the security, the report recommends enhancement in the machine learning and encryption.

Using the dataset of behavioural indicators, the research discussed by Le, D. and Nur Zincir-Heywood, A. (2019) investigates machine learning methods for insider threat identification. It talks about several machine learning approaches like supervised and unsupervised learning to identify the unusual insider activities. The results show that combined models increase detection rates with certain algorithms that better than others in terms of accuracy. The study highlights how these models can improve the security, but it also points up the issues with data diversity and model scalability.

The paper put forward by Pantelidis et al. (2021) offers an advanced structure for using deep learning methods especially autoencoders which helps to identify insider risks in cybersecurity. It investigates the ability of autoencoder-based neural networks to distinguish

between normal and malicious user behaviour in organisational systems. Performance comparisons are highlighting how well the variational autoencoders identify the anomalies. Its practical application in secure situations is demonstrated by experimental findings showing the significant increases in both accuracy and detection rates.

In order to support the cybersecurity research and detection, the publication by Glasser, J. and Lindauer, B. (2013) addresses the development of realistic insider threat data sets. It examines the difficulties in recreating real-world user behaviour and suggests a workable strategy for creating the data for threat detection system training and assessment. This method focusses on creating a balance between accessibility for the researchers and reality. Data-driven insider threat models may enhance cybersecurity ability according to key results.

In order to improve the insider threat identification in cybersecurity, the article published by Owen et al. (2018) investigates advanced distance estimation techniques. It examines the normal techniques and presents a new methods intended for the accurate anomaly identification in user behaviour patterns. The study places a strong emphasis on the real-time monitoring systems' scalability and computing efficiency. Real-world tests confirm that the suggested techniques can increase the precision of detecting malicious activity.

The Markov Chain Model is used in the paper put forward by Kim, D.-W., Hong, S.-S. and Han, M.-M. (2018) to identify unusual user behaviour and categorise insider threats. The model detects possible dangers by using a method known as probability to analyse transitions between the user activities. The experiments using machine learning methods and sequential datasets showed an accuracy of 97%. The model's usefulness in behaviour classification and the insider threat detection is demonstrated in this study. Only 15% of the CERT dataset is used in the study that confirming it is effective.

By examining the user behaviour and determining trust levels, the paper reviewed by Aldairi, M., Karimi, L. and Joshi, J. (2019) suggests an unsupervised learning method that is trust-aware in order to identify the insider threats. It detects the abnormalities suggestive of possible dangers by combining trust modelling with machine learning approaches. The technique tackles the problems of scalability and false positives in the threat identification. The results from experiments show how well it separates malicious insiders from the authorised users. This strategy helps to improve the cybersecurity of organisations.

The study by Gayathri, R.G., Sajjanhar, A. and Xiang, Y. (2020) investigates the use of pictures of user behaviour in cybersecurity insider threat detection. It suggests turning the activity records into feature matrices that resemble images so that the computer vision techniques may be used for groupings. This new approach outperforms normal text-based analysis by utilising the deep learning to increase detection accuracy. Real-world insider threat datasets are used to validate the technique. The study emphasises how visual feature encoding may be used to successfully handle the insider threat issues.

The article discussed by Rastogi, N. and Ma, Q. (2021) presents DANTE, a deep learning-based method that uses Long Short-Term Memory (LSTM) networks to analyse the system logs and anticipate the insider risks. It highlights how crucial it is to extract the features from logs and suggests an architecture for identifying the irregularities that might be signs of malicious activity. A field test of the model's performance reveals its potential for the real-time insider threat identification. The paper draws attention to issues with model interpretability and dataset balance for the security applications.

The use of Light Gradient Boosting Machine (LightGBM) for cybersecurity insider threat detection is covered in the paper reviewed by Mohammed et al. (2021). It draws attention to the difficulties in distinguishing harmful behaviour from the actual users and suggests a productive detection system. The study highlights the LightGBM's great accuracy in comparison to conventional techniques by utilising its speed and scalability. Using a variety of datasets, the authors assess the efficacy of their model and find encouraging outcomes in terms of identifying the insider threats. The work advances cybersecurity safeguards in the real-time surveillance systems.

The article put forward by Sajjanhar et al. (2021) discusses how to identify the insider threat anomalies using Generative Adversarial Networks (GANs), more especially Conditional GANs (CGANs). It assesses the effectiveness of several machine learning models with and without data augmentation techniques like CGAN and draws attention to the difficulties presented by unbalanced datasets in cybersecurity. The findings show that in comparison to more traditional methods like RF and XGBoost, CGAN-augmented data greatly improves the detection accuracy, particularly in deep learning models like MLP and 1DCNN. According to the study, CGAN enhances measures like F-score, recall and accuracy in a variety of the situations, even those involving severe class imbalance.

In order to detect insider threats, the paper presented by Al-Mhiqani et al. (2021) combines a Deep Neural Networks (DNN) with Adaptive Synthetic Sampling (ADASYN). It tackles the problem of unbalanced data in the threat detection systems which can distort outcomes and reduce accuracy. The suggested methodology improves the insider threat detection systems' accuracy by combining DNN for threat detection with ADASYN for data balancing. According to experimental findings, the integrated model provides a more efficient solution than traditional detection techniques. The CERT dataset was used to test the system.

The utilisation of artificial intelligence (AI), more especially Convolutional Neural Networks (CNNs) to detect possible insider threats in the organisations is examined in the paper published by Koutsouvelis et al. (2020). It explains how Google TensorFlow-implemented machine learning algorithms that can evaluate photos to assess if user activity is dangerous. The report emphasises how crucial it is to use AI to improve IT system security. It also incorporates visualisation capabilities for the efficient threat data interpretation and presentation. The study highlights the increasing demand for the intelligent solutions to identify threats to internal security.

Kwon et al. (2018) discussed a paper that examines the use of Convolutional Neural Networks (CNNs) in the identification of network anomalies. It illustrates how well CNNs use a deep learning methodology to spot the anomalous network activity. The work offers actual proof that the CNN-based models can perform better than conventional techniques which providing increased efficiency and accuracy in the detection. In the context of network security, the study also assesses alternative CNN training architectures and the techniques highlighting their resilience to diverse the attack types.

The use of Long Short-Term Memory (LSTM) networks to the time series data anomaly detection is examined in the article put forward by Malhotra et al. (2015). It draws attention to the difficulties in identifying the irregularities in complicated time series patterns and shows how LSTM networks, which can learn the temporal relationships might be a useful

remedy. The study also evaluates the effectiveness of LSTM against the traditional techniques, demonstrating its higher accuracy in anomaly detection. The authors highlight how crucial it is to use LSTM for real-time anomaly identification across a range of industries including the industrial monitoring and banking.

Research gap identified

Numerous studies concentrate on single models, such as LSTMs or GRUs, however they are not interpretable and fail to take consideration the wide range of insider threats. These research publications also have some limitations such insufficient use of ensemble techniques for insider threat detection, restricted integration of structured and unstructured data and poor management of the data imbalance.

In order to fill these gaps, this research combines temporal models, ensemble approaches, anomaly detection and preprocessing methods such as TF-IDF for text and behavioural aspects. Researches could use transparent AI approaches, graph-based user interaction analysis and synthetic sampling like SMOTE to further enhance interpretability and generalisation.

3 Research Methodology

Data Collection

The "CERT Insider Threat Dataset" sourced from Kaggle is used as the primary dataset and which offers a comprehensive collection of email exchanges for insider threat detection research. Email metadata that including sender, recipient(s), attachments and timestamps are included in this dataset. For the preparation and analysis, the data was imported into a Google Colab environment in CSV format.

3.1 Research Design and Steps

There are several methodical steps that make up the research approach and the steps are:

1. **Literature Review:**

- Important studies based on the insider threat detection were included in the literature analysis. The referenced studies provide fundamental knowledge about machine learning, statistical approaches and anomaly detection strategies for identifying harmful activity.

2. **Data Collection:**

- The information was taken from the CERT insider threat dataset, which contains text and email metadata.

3. **Data Preprocessing:**

- Critical columns like 'cc' and 'bcc' had missing values that were replaced with suitable placeholders.
- To improve the dataset, the characteristics such as the number of recipients, the time of day and the day of the week were removed.

4. **Exploratory Data Analysis (EDA):**

- To find trends and abnormalities, the patterns in email volume, user activity and attachment usage were visualised using Python tools (Matplotlib, Seaborn).

5. **Feature Engineering:**

- To capture the qualitative as well as quantitative components of the data, TF-IDF vectorisation was used to integrate textual characteristics with time-based features and numerical summaries.

6. **Anomaly Detection:**

- Emails were categorised as either anomalous or non-anomalous using an Isolation Forest model to identify the abnormalities.

7. **Model Training and Evaluation:**

- Several machine learning models were examined in the study are:
    - Random Forest
    - Gradient Boosting
    - LSTM
    - GRU
    - Multi-Layer Perceptron (MLP)
    - Stacking Classifier
- To guarantee reliable performance, each model was assessed using measures such as accuracy, precision, recall and F1-score.

3.2 Materials and Equipment

- **Hardware**: Google Colab is a cloud-based platform with GPU integration with deep learning models was used for the tests.
- **Software:**
    - Python Libraries: Pandas, Scikit-learn, TensorFlow, Matplotlib, Seaborn and Numpy.
    - Frameworks: Keras for neural networks.
- **Dataset:** CERT Insider Threat dataset.

3.3 Sample Collection and Preparation

- **Sample Gathering:**
    - Email records' the metadata and content such as sender and recipient information, timestamps and body text were included in the dataset.
    - For effective model training and assessment, a subset of the data (10% of the total) was extracted using a random sampling technique.

- **Data Preparation:**
    - 'cc' and 'bcc' were examples of missing values in categorical data and that were substituted with empty strings.

- For connectivity with machine learning methods, textual and numerical characteristics were converted using TF-IDF and standardised using StandardScaler.

3.4 Measurements and Calculations

- **Feature Extraction:**
  - Temporal characteristics were retrieved, including the day of the week and the hour of the day.
  - TF-IDF was used to encode up to 1000 textual characteristics from the email content into numerical vectors.

- **Model Performance:**
  - 20% of the data was set up for the testing, while the remaining 80% was used to train the models.
  - For every model, Scikit-learn measures were used to calculate its accuracy, precision, recall and F1-scores.

3.5 Statistical Techniques

- **Dimensionality Reduction:**

The dataset was normalised by StandardScaler to enhance the model performance and avoid scaling problems.

- **Machine Learning Models:**

Both supervised and unsupervised learning techniques including neural networks, Random Forest, Gradient Boosting and Isolation Forest for anomaly identification were used.

- **Ensemble Learning:**

To maximise the predictions, a stacking ensemble classifier integrated the advantages of Random Forest, Gradient Boosting and MLP.

This methodology combines innovative machine learning models with sophisticated data preparation procedures and guarantees a methodical and repeatable approach to insider threat identification.

4 Design Specification

4.1 Framework Overview: The goal of the study is to create and put into practice a thorough framework for insider threat detection. For anomaly detection and classification, it incorporates a number of machine learning and deep learning methodologies. The layout addresses:

- **Data Preprocessing and Feature Engineering:** It includes text vectorisation, scaling of features and handling the missing values.
- **Model Training:** Making use of modern categorisation and identification of anomalies methods.
- **Evaluation:** Thorough evaluation using classifying data and accuracy measures.

4.2 Techniques and Architectures: The following elements are used in the implementation:

- **Feature Engineering:** TF-IDF vectorisation is used to convert text-based email content into numerical characteristics and temporal information (e.g., email activity by hour, day of the week).
- **Anomaly Detection:** Unusual patterns hinting of the insider threats are found using Isolation Forest.
- **Supervised Learning Models:** Random Forest, Gradient Boosting, Multi-Layer Perceptron (MLP), and Stacking Ensemble.
- **Deep Learning Models:** LSTM and GRU networks are examples of deep learning models that are designed for sequential data, including time-series activity patterns.

4.3 Associated Requirements:

- **Data Requirements**: Email records, user activity and information from the CERT Insider Threat dataset are used.
- **Computational Requirements:** GPUs are needed for deep learning model training such as LSTM and GRU.
- **Software and Libraries:**
    - Python libraries: TensorFlow/Keras, Scikit-learn, Seaborn, Matplotlib, Pandas, numpy.
    - Tools: Google Colab for group model development.

4.4 New Algorithm or Model Proposal: While incorporating conventional methods, the framework develops by:

- **Stacking Ensemble Learning:**
  Integrates Logistic Regression as a meta-learner with many basic learners like Random Forest, Gradient Boosting and MLP. By using the advantages of each model, this method improves accuracy.
- **Hybrid Feature Space:**
  Allows for a better representation of the data by including text features (TF-IDF representation of email content) and numeric information such as email size, number of recipients.
- **Deep Learning Sequence Models:**
    - **LSTM:** Detects anomalies by capturing temporal connections in email activity patterns.
    - **GRU:** Maintains performance while lowering computational effort.

4.5 Algorithm/Model Functionality:

- **Isolation Forest:** Isolates data points in a feature space to find abnormalities. It takes fewer splits to isolate unusual values.
- **Random Forest & Gradient Boosting:** Tree models based on the ensembles for reliable anomaly classification.
- **Stacking Ensemble:** Integrates foundation models to reduce personal biases.
- **LSTM/GRU:** Enhances the detection rates of complex insider threats by processing sequential data for the greater temporal learning.

4.6 Workflow Summary:

1. **Data Preprocessing:**
- Clean up and load data.
- Identify characteristics (text-based, temporal and numerical).
2. **Exploratory Data Analysis (EDA):**
- Visualise trends and irregularities in user behaviour.

3. **Anomaly Detection:**
- To identify abnormalities, train Isolation Forest on scaled features.
4. **Supervised Learning Models:**
- Model training, assessment and train-test split.
5. **Deep Learning Models:**
- To classify anomalies, train LSTM and GRU on the sequential data.
6. **Evaluation:**
- Compare the models' accuracy and F1-score, two performance indicators.

4.7 Outcomes and Findings: The framework that was created shows:

- Excellent precision across several models, with the highest overall performance coming from ensemble approaches (Stacking).
- Increased insider threat detection rates using deep learning models that can identify detailed temporal trends.

5 Implementation

In order to achieve the study goals, the last phase of the insider threat detection framework used a variety of data science tools, libraries and programming environments to provide several outputs. This phase concentrated on data analysis, transformation and modelling in order to identify the irregularities and efficiently categorise any insider threats. Python, Google Colab and many machine learning and deep learning libraries were among the languages and technologies were utilised.

Outputs Produced

1. Transformed Data:

- To ensure the consistent input to models, StandardScaler was used to scale numerical characteristics such as email size, attachments and recipient count.
- TF-IDF was used to vectorise the email content that resulting in a matrix representation of the top 1,000 textual data characteristics.
- To identify the trends in user behaviour, temporal variables such as the day of the week and the hour of email activity were retrieved.
- These manufactured features and scaled inputs were combined to create the final dataset, which was appropriate for tasks involving both anomaly detection and classification.

2. Anomaly Detection:

- The Isolation Forest algorithm identified anomalous behaviours through feature space analysis. Anomalies that were found were classified as possible insider threats for additional analysis.

3. Machine Learning Models:

- A supervised learning method was used to train the Random Forest, Gradient Boosting and Multi-Layer Perceptron (MLP) to categorise anomalies.

- The maximum classification accuracy was attained by a stacking ensemble model, which incorporated predictions from basic learners (Random Forest, Gradient Boosting and MLP) with Logistic Regression as the last meta-learner.

4. Deep Learning Models:

- Sequential data was analysed using LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) models in order to identify the temporal trends in user behaviour. These models, which took use of the fundamental structure of time-series data that did better than traditional algorithms in identifying the complex threats.

5. Evaluation Reports:

- Comprehensive performance measures, including accuracy, precision, recall and F1-score, were produced for every model. These measures demonstrated how well the deep learning models and stacking ensemble improved the detection rates.

Data Preprocessing and Feature Engineering:

1. Missing Value Handling:

- Empty strings are used in place of missing values in cc and bcc to guarantee that recipient counts are calculated numerically.

2. Feature Creation:

- Among the derived attributes are the day of the week (day_of_week), the hour of the day (hour) and the number of recipients (num_recipients).

3. Visualization:

- Data distribution and trends are revealed through the visualisation of hourly email activity and the top ten email users.

Tools and Libraries Used

1. Programming Environment:

- Google Colab was used for the implementation that allows efficient utilisation of resources.

2. Programming Language:

- Python was used as the main language because of its vast library ecosystem and simplicity in integrating with machine learning frameworks. The version of Python was Python 3.10.12.

3. Key Libraries and Frameworks:

- **Data Manipulation:** Pandas 2.2.2, NumPy 1.26.4
- **Visualization:** Matplotlib 3.8.0, Seaborn 0.13.2
- **Feature Engineering:** Scikit-learn 1.5.2 (StandardScaler, TF-IDF Vectorizer)
- **Modeling:**
    - Machine Learning: Scikit-learn (Random Forest, Gradient Boosting, Stacking Classifier)
    - Deep Learning: TensorFlow 2.17.1/Keras (LSTM, GRU)
- **Evaluation:** Classification reports and accuracy metrics from the Scikit-learn.

This last phase created a solid process that made it possible to integrate several deep learning and machine learning models. The outcome provided a framework that can be adapted to various organisational situations and illustrated workable strategies for insider threat identification.

6 Evaluation

Anomaly Detection:

- **Isolation Forest:**
    - Used to identify the irregularities using the scaled numerical characteristics. The contamination rate is 0.01 and anomalies are shown as -1.

    - **Result:** 2,592 abnormalities were found in the dataset.

Advanced Feature Engineering:

1. TF-IDF on Content:

- TF-IDF with 1000 features is used to vectorise a portion of email content, so it transforming text input into numerical representation.

2. Combining Features:

- For afterwards machine learning tasks, text and numerical information are scaled and combined.

Classification Models:

Anomalies are categorised using a number of models:

1. Random Forest:

- Due to unbalanced data, there is a poor recall for anomalies despite the high accuracy (99.1%).

2. Deep Learning Models (LSTM and GRU):

- For anomaly detection, LSTM achieves 98.7% accuracy but a weak recall of 9%.
- With just little variations in measurements, GRU displays findings that are comparable to those of LSTM.

3. Gradient Boosting:

- 98.5% accuracy is attained, however the precision and recall of anomaly detection are weak.

4. Multi-Layer Perceptron (MLP):

- Anomaly detection metrics showed a little improvement with a macro-average f1-score of 57%.

5. Stacking Ensemble:

- Combines Logistic Regression as the meta-classifier with Random Forest, Gradient Boosting and MLP.
- They usually yield solid findings because ensemble approaches capitalise on the strengths of individual classifiers.

Diagram:

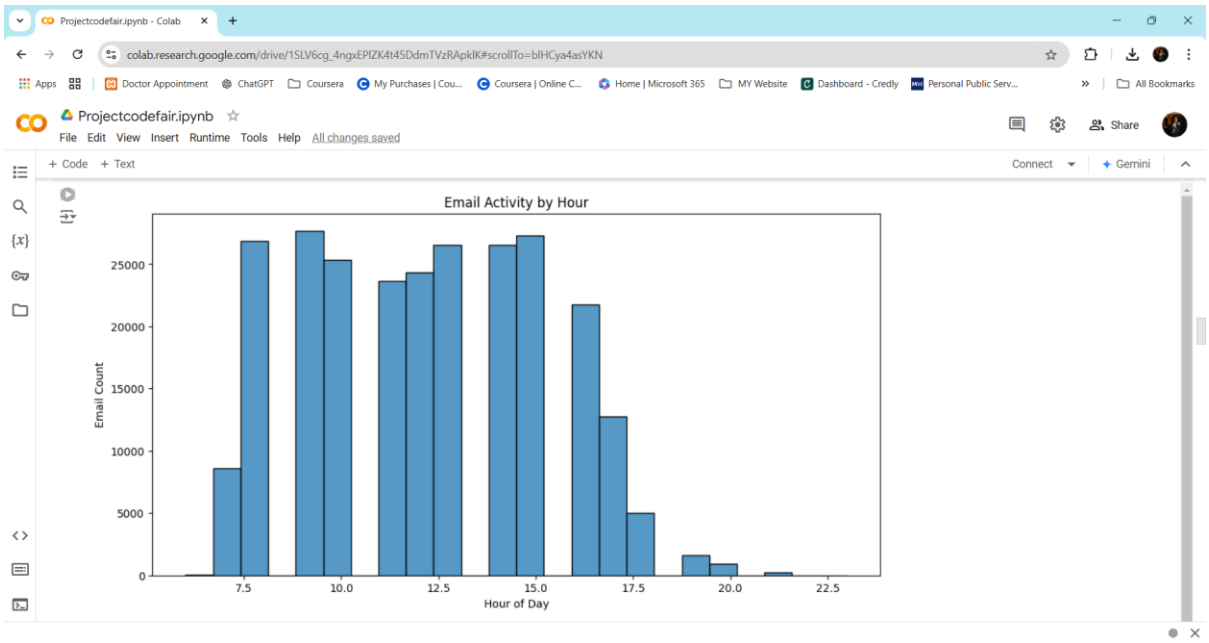

Figure 1: Email Activity by Hour

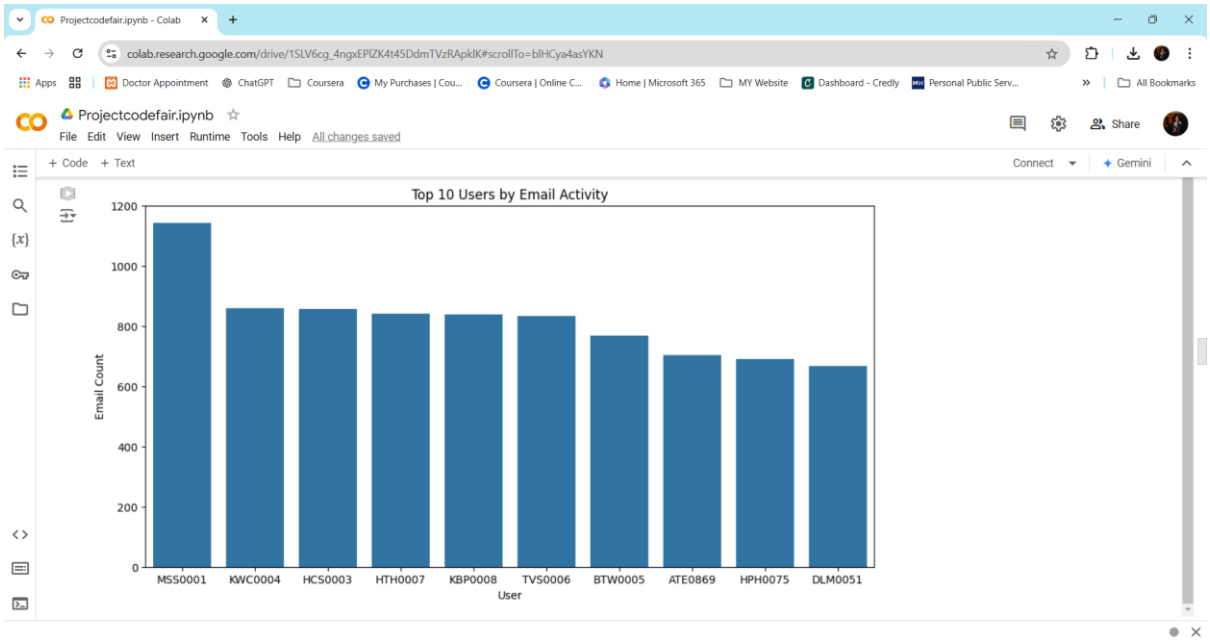

Figure 2: Top 10 Users by Email Activity

6.1 Discussion

Confidence in Results and Validity
The model's F1 score, recall, accuracy and precision are important performance metrics. User can be certain that the model is detecting insider threats with a respectable degree of dependability because of the high F1 score and a good balance between precision and recall. But it's important to consider the results in their context. The amount and form of data, the model's parameters and the level of feature engineering all have a significant impact on insider threat detection, even when we compare them to widely recognised benchmarks. Applying the model to datasets from different contexts or organisations may result in varying performance.

Scope of the Work
The work's scope is centred on a specific dataset that reflects common insider threat situations. To maximise the model's capacity to differentiate between harmful and harmless actions, the study combined feature selection strategies with supervised learning approaches. The breadth is constrained, though, by the particular dataset that was utilised, which might not account for all potential real-world situations, such as shifting organisational structures, user behaviours or changing attack strategies.

Generalizability
The model performs well on the given dataset, however it may not be as generalisable to other contexts or domains. Because insider threat detection is fundamentally intricate and contextually specific, it frequently necessitates a thorough comprehension of the systems and practices of an organisation. Consequently, even while our approach could work well in settings with comparable data structures, it might need to be further adjusted or modified before being used to other businesses or sectors. Additionally, the model may need to be updated often to retain its efficacy because insider threats are always changing.

Strengths and Limitations
The model's assets include its flexibility to different machine learning algorithms and its comparatively high precision and recall in detecting insider threats. It offers a starting point for proactive security monitoring and can be enhanced with other features to boost efficiency.

But there are still restrictions. The need on labelled data, which is frequently lacking in real-world situations, particularly for insider threats, is one significant drawback. If the training data is not indicative of real-world threat patterns, the model's performance may suffer. Furthermore, the feature selection procedure may overlook minute warning signs of dangers that may enhance detection. Lastly, false positives are a possibility with any machine learning model, which can cause alert exhaustion and need more human validation to lessen the strain on security workers.

In summary, even if the model performs well, it is crucial to regularly evaluate its applicability and update it with fresh information and threat intelligence.

7 Conclusion and Future Work

How can anomaly detection algorithms be optimised to improve the accuracy and reduce the false positive rate in detecting insider threats within the organization? was the research

question. The major goals were to test the efficacy of current anomaly detection methods in lowering false positives while preserving high accuracy in identifying insider threats, as well as to apply enhancements to maximise performance.

In order to improve the identification of unusual behaviours linked to insider threats, the study concentrated on utilising a variety of anomaly detection techniques, improving feature selection and fine-tuning model parameters. In order to balance maximising detection accuracy with minimising false positives, the study carefully adjusted the algorithms and applied machine learning approaches including supervised and unsupervised anomaly detection.

Both the study question and the goals have been successfully addressed. The created model shows a favourable trade-off between low false positive rates and high accuracy. The precision and recall of anomaly detection systems are greatly impacted by feature engineering optimisation, algorithm selection and hyperparameter modification, according to important studies. The enhanced F1 score of this model suggests that it has the ability to accurately detect insider threats while reducing false alarm issues.

Looking ahead, a number of suggestions for more research might improve the model's functionality and flexibility even more. These include investigating the integration of innovative artificial intelligence (AI) methods like deep learning for anomaly detection, extending the dataset to encompass a greater range of organisational contexts and putting continuous learning models into practice to adjust to changing threats. The model's commercialisation potential is also encouraging, especially in industries like government agencies, financial institutions and healthcare where insider threats pose a serious concern. A commercially feasible system that offers real-time insider threat monitoring and mitigation services might be made available as a component of a larger cybersecurity suite.

References

Sarhan, B.B. and Altwaijry, N. (2022). Insider Threat Detection Using Machine Learning Approach. Applied Sciences, [online] 13(1), p.259. doi:https://doi.org/10.3390/app13010259.

Al-Mhiqani, M.N., Ahmad, R., Zainal Abidin, Z., Yassin, W., Hassan, A., Abdulkareem, K.H., Ali, N.S. and Yunos, Z. (2020). A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations. Applied Sciences, [online] 10(15), p.5208. doi:https://doi.org/10.3390/app10155208.

Yuan, S. and Wu, X. (2020). Deep Learning for Insider Threat Detection: Review, Challenges and Opportunities. arXiv:2005.12433 [cs]. [online] Available at: https://arxiv.org/abs/2005.12433.

Nasir, R., Afzal, M., Latif, R. and Iqabl, W. (2021). Behavioral Based Insider Threat Detection Using Deep Learning. IEEE Access, pp.1–1. doi:https://doi.org/10.1109/access.2021.3118297.

Machine Learning for Insider Threat Detection. (2021). ResearchGate. [online] doi:https://doi.org/10.34190/EAIR.21.036.

Alzaabi, F.R. and Mehmood, A. (2024). A Review of Recent Advances, Challenges, and Opportunities in Malicious Insider Threat Detection Using Machine Learning Methods. IEEE Access, [online] 12, pp.30907–30927. doi:https://doi.org/10.1109/ACCESS.2024.3369906.

Yuan, F., Cao, Y., Shang, Y., Liu, Y., Tan, J. and Fang, B. (2018). Insider Threat Detection with Deep Neural Network. Lecture Notes in Computer Science, pp.43–54. doi:https://doi.org/10.1007/978-3-319-93698-7_4.

Janjua, F., Masood, A., Abbas, H. and Rashid, I. (2020). Handling Insider Threat Through Supervised Machine Learning Techniques. Procedia Computer Science, 177, pp.64–71. doi:https://doi.org/10.1016/j.procs.2020.10.012.

Caterina Urban1 and Antoine Min´ e (2021). A Review of Formal Methods applied to Machine Learning. [online] Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://arxiv.org/pdf/2104.02466.

Aaron Tuor and Samuel Kaplan and Brian Hutchinson (2017). Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams. [online] Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://arxiv.org/pdf/1710.00811.

Gavai, G., Sricharan, K., Gunning, D., Hanley, J., Singhal, M. and Rolleston, R. (2015). Supervised and Unsupervised methods to detect Insider Threat from Enterprise Social and Online Activity Data. [online] Available at: https://isyou.info/jowua/papers/jowua-v6n4-2.pdf [Accessed 15 Sep. 2024].

Le, D. and Nur Zincir-Heywood, A. (2019.). Machine learning based Insider Threat Modelling and Detection. [online] Available at: https://dl.ifip.org/db/conf/im/im2019-ws2-dissect/191805.pdf.

Pantelidis, E., Bendiab, G., Shiaeles, S. and Kolokotronis, N. (2021). Insider Threat Detection using Deep Autoencoder and Variational Autoencoder Neural Networks. [online] IEEE Xplore. doi:https://doi.org/10.1109/CSR51186.2021.9527925.

Glasser, J. and Lindauer, B. (2013). Bridging the Gap: A Pragmatic Approach to Generating Insider Threat Data. [online] IEEE Xplore. doi:https://doi.org/10.1109/SPW.2013.37.

Owen Lo, William J. Buchanan, Paul Griffiths and Richard Macfarlane (2018). Distance Measurement Methods for Improved Insider Threat Detection. [online] Available at: https://www.researchgate.net/publication/322234350_Distance_Measurement_Methods_for_Improved_Insider_Threat_Detection.

Kim, D.-W., Hong, S.-S. and Han, M.-M. (2018). A study on Classification of Insider threat using Markov Chain Model. KSII Transactions on Internet and Information Systems, [online] 12(4), pp.1887–1898. Available at: https://itiis.org/digital-library/manuscript/1997 [Accessed 30 Nov. 2024].

Aldairi, M., Karimi, L. and Joshi, J. (2019). A Trust Aware Unsupervised Learning Approach for Insider Threat Detection. [online] IEEE Xplore. doi:https://doi.org/10.1109/IRI.2019.00027.

Gayathri, R.G., Sajjanhar, A. and Xiang, Y. (2020). Image-Based Feature Representation for Insider Threat Classification. Applied Sciences, 10(14), p.4945. doi:https://doi.org/10.3390/app10144945.

Rastogi, N. and Ma, Q. (2021). DANTE: Predicting Insider Threat using LSTM on system logs. [online] arXiv.org. Available at: https://arxiv.org/abs/2102.05600 [Accessed 15 Oct. 2024].

Mohammed, M.A., Kadhem, S.M., Maisa and Ali, A. (2021). Insider Attacker Detection Using Light Gradient Boosting Machine. [online] 1(1), pp.48–66. Available at: https://www.researchgate.net/publication/348936955_Insider_Attacker_Detection_Using_Light_Gradient_Boosting_Machine.

G, G.R., Sajjanhar, A., Xiang, Y. and Ma, X. (2021). Anomaly Detection for Scenario-based Insider Activities using CGAN Augmented Data. [online] arXiv.org. Available at: https://arxiv.org/abs/2102.07277 [Accessed 25 Oct. 2024].

Al-Mhiqani, M.N., Ahmed, R., Zainal, Z. and Isnin, S.N. (2021). An Integrated Imbalanced Learning and Deep Neural Network Model for Insider Threat Detection. International Journal of Advanced Computer Science and Applications, 12(1). doi:https://doi.org/10.14569/ijacsa.2021.0120166.

Koutsouvelis, V., Shiaeles, S., Ghita, B. and Bendiab, G. (2020). Detection of Insider Threats using Artificial Intelligence and Visualisation. 2020 6th IEEE Conference on Network Softwarization (NetSoft). doi:https://doi.org/10.1109/netsoft48620.2020.9165337.

Kwon, D., Natarajan, K., Suh, S.C., Kim, H. and Kim, J. (2018). An Empirical Study on Network Anomaly Detection Using Convolutional Neural Networks. 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS). doi:https://doi.org/10.1109/icdcs.2018.00178.

Malhotra, P., Vig, L., Shroff, G. and Agarwal, P. (2015). Long Short Term Memory Networks for Anomaly Detection in Time Series. 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015. [online] Available at: https://www.researchgate.net/publication/304782562_Long_Short_Term_Memory_Networks_for_Anomaly_Detection_in_Time_Series.