PROJECT REPORT ON

Insider Threat Predictor System

Submitted by

Shaikh Aiman
AD23-ST#IS#7003

Under the Supervision of

UPENDRA

Senior Security Analyst

KRISHNA

Security Analyst

Registered And Head Office

D.NO: 11-9-18, 1st Floor,

Majjivari Street, Kothapeta,

Vijayawada - 520001.

+91 9550055338 / +91 7901336873

contact@suprajatechnologies.com

COMPANY INTRODUCTION:

Supraja Technologies is a leading Knowledge and Technical Solutions Provider and pioneer leader in IT industry, is operating based out of Vijayawada.

R&D at Supraja

With a 24X7 work in Research & Development, experts at Supraja Technologies work under our :

- **Supraja Technologies Cyber Security Cell**

About Supraja Technologies:

Supraja Technologies (a unit of CHSMRLSS Technologies Pvt. Ltd.) with its foundation pillars as Innovation, Information and Intelligence is exploring indefinitely as a **Technology Service Provider (Corporate Consulting)** and as a **Training Organization (Ed-Tech)** as well.

You may visit us at :
www.suprajatechnologies.com

The multi domains of trainings which Supraja Technologies operate include the following :

- **Workshops & Hackathons**
  - Engineering Colleges
  - Corporate Companies (Startups & MNC's)
  - Government Organizations

- **Classroom Trainings Cum Certification Courses**
  - Summer Training (30-45 Days)
  - Winter Training (10 - 15 Days)
  - Weekend Training (2 Days)
  - 1 Month / 3 Months / 6 Months Courses

- **On-site Trainings**
  - Value Added Courses / Two Credit Courses for Colleges
  - Faculty Development Programs
  - College Summer Training (15 Days, 30 Days, 45 Days & 60 Days)
  - Govt Agencies, Police Academies, Corporates etc

- **Cloud Campus**
  - (Distance Learning Program) *Coming Soon

- **Internships**
  - Internship for Engineering Students (30 Days, 45 Days & 60 Days)
  - Internship for Graduates (6 Months)

- **Lab Setup**
  - Cyber Lab

- **CoE**
  - Cyber Security Centre of Excellence

- **Supraja Technologies Security Assessment Product**
  - (Our SaS Product is currently under development) *Coming Soon

Why Supraja Technologies:

Be it Training or a workshop, the course content is always from R&D Cell of Supraja.

- A proven track record of delivering quality services.
- **68,500+** Students trained by our trainers till date.
- Training Partners of recognized institutions.
- Trainers with excellent research and teaching pedagogy illustrate their findings through corporate standard **practical demonstrations** during their sessions.
- Easy to learn and **hands-on sessions** are given, with additional benefits of Study Material, Tool kit and immediate query handling.
- Self-Prepared **Cyber Security Cell.**
- Supraja Technologies has the best, experienced and highly **skilled bunch of R&D Engineers, Security Analysts, Security Consultants & Trainers.**
- We provide training in Innovating and Trending Technologies to Govt. Officials, Corporate Houses and Colleges.

✓ Something we are proud of :

1. Supraja Technologies CEO Mr.Santosh Chaluvadi is an Alumni of Potti Sriramulu Chalavadi Mallikharjuna Rao College of Engineering and Technology, Vijayawada.

   With our CEO this college conducted/organised a 50 hours Nonstop Marathon Training Workshop on Ethical Hacking & Cyber Security for which this respective college and our CEO both holds their name in **“LIMCA BOOK OF RECORDS 2017”**

2. We are very happy to inform you all that our company, Supraja Technologies has been shortlisted for **"Top 50 Tech Companies" award 2019**, conferred at InterCon - Dubai, UAE.

   Supraja Technologies is one out of thousands companies that were initially screened by InterCon team of 45+ research analysts over a period of three months and the final shortlist includes 150+ firms and we are very proud to inform you all that our company Supraja Technologies also happens to be a part of the same.

✓ Life changing solution/service :

After working on R&D for around 2 years, finally in the mid 2019 we have successfully developed a service/solution of various techniques and strategies for the Film Industry through which he can kill piracy of any film in online up to 35% right now. This betaservice is being appreciated & adopted by various Tollywood Film Industry Producers & Hero's to safeguard their film from piracy in online and to gain more profits.

By the end of 2033 our vision is to rollout a complete full packed service/solution where we can kill piracy entirely 100% everywhere in online for sure.

Appreciation :

Received a great appreciation from our 1st Tollywood Film Industry client Mr.Saptagiri for providing our Anti-Piracy betaservice for his film VAJRA KAVACHADARA GOVINDA

✓ Achievements:

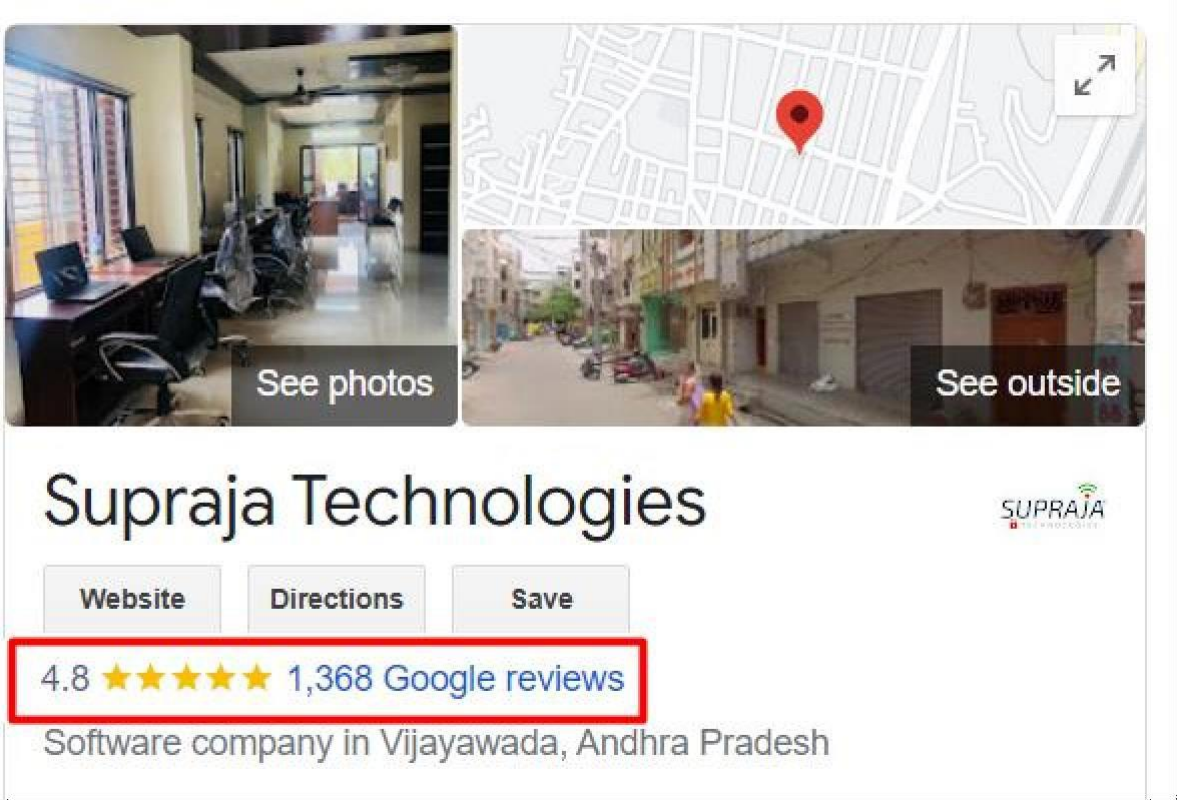
- On 17th August 2024, We (Supraja Technologies) launched our company's **Centre of Excellence in Cyber Security** (CoE) at Ramco Institute of Technology, Rajapalayam

- On 18th September 2024, We (Supraja Technologies) launched our company's **Centre of Excellence in Cyber Security** (CoE) at SRM University (Ramapuram Campus), Chennai

- On 20th November 2024, We (Supraja Technologies) launched our company's **Centre of Excellence in Cyber Security** (CoE) at St.Joseph's Institute of Technology, Chennai

Company Identification Number (CIN Number) : U72900AP2019PTC111909
PAN Number : AAICC1770E, TAN Number : HYDC10078F
ISO Certificate Number : 031122019107
MSME Registration Certificate No : UDYAM-AP-06-0035821
Supraja Technologies Trademark Application ID No for Class 41 : 4146139
Supraja Technologies Trademark Application ID No for Class 42 : 4146144

Our Company Supraja Technologies
is one of the emerging startup
in Andhra Pradesh
with 4.8 Google Ratings



Link : https://bit.ly/SuprajaGoogle

AND

also check our Company CEO Instagram Profile
for our recent more success stories:
https://www.instagram.com/chaluvadisantosh/

SUPRAJA TECHNOLOGIES
(a unit of CHSMRLSS Technologies Private Limited)
An MSME & ISO 9001:2015 Certified Company
Regd. & Head Office : Door No. 11-9-18, 1st Floor, Majjivari Street, Kothapeta, Vijayawada - 520001.
Branch Office : Prabhu Villas, Plot No. 124, 1st Floor, Church Road, Ayyappa Nagar, Vijayawada - 520007.
contact@suprajatechnologies.com | www.suprajatechnologies.com | +91 - 9550055338

Santosh Chaluvadi

Founder & CEO
Supraja Technologies

He is a 28-year-old entrepreneur, one of the India's efficient Cyber Security Analyst and also he is an expert Digital Marketer as well. He is a digital marketer by profession and security enthusiast by passion. He primarily focuses on content building, testing and monetization of blogs. He has successfully developed many websites and done the security testing himself to ensure that the user's data is in safe hands and their privacy is protected. He is very active on social media and shares lot of tech stuff with his followers. The young student hacker has solved many issues with the vulnerabilities present in various websites and databases, given a solution in clearing the loopholes in order to protect the data to be leaked from the databases. Besides Ethical Hacking & Cyber Security, he also has a passion in Blogging & Digital Marketing.

While pursuing his engineering itself, he has trained many young generation people/students of more than 3500+ from various parts across Andhra Pradesh through his workshops, seminars, courses in Cyber Security and this makes him one of the youngest student trainers in India.

At the age of 20 he conducted his first workshop on Blogging & Ethical Hacking which was the beginning to his success in this field and right now he has handful of workshops to train students, government and corporate organizations as well in Andhra Pradesh & Telangana. He is the only student trainer who started conducting workshop for his peers, professors and for corporates.

The year 2016 gave me the conviction and confidence to notch up whatever I was doing. I'd organized a 50-hour marathon event on Ethical Hacking and Cyber Security in PSCMR College, Vijayawada that went on to bag to achieve Limca Book of Records for non-stop longest duration workshop. The impact we were making was clearly visible by now. Diverse people from Jammu & Kashmir in the North to Kanyakumari in the South had attended the event. Some of the Government of India officials took notice of this record and invited our team for couple of conferences at New Delhi. The country's esteemed institutions were reaching out to me to conduct training, workshops, and events on a myriad of subjects related to online security. Multi-National Corporations (MNC's) had begun to consult me on the Cyber Security of their systems.

❖ Life changing solution/service :

After working on R&D for around 2 years, finally in the mid 2019 we have successfully developed a service/solution of various techniques and strategies for the Film Industry through which he can kill piracy of any film in online up to 35% right now. This betaservice is being appreciated & adopted by various Tollywood Film Industry Producers & Hero's to safeguard their film from piracy in online and to gain more profits.

By the end of 2033 our vision is to rollout a complete full packed service/solution where we can kill piracy entirely 100% everywhere in online for sure.

Appreciation:

Received a great appreciation from our 1st Tollywood Film Industry client Mr.Saptagiri for providing our Anti-Piracy betaservice for his film VAJRA KAVACHADARA GOVINDA

❖ Our Company Achievements:

- On 17th August 2024, We (Supraja Technologies) launched our company's **Centre of Excellence in Cyber Security** (CoE) at Ramco Institute of Technology, Rajapalayam
- On 18th September 2024, We (Supraja Technologies) launched our company's **Centre of Excellence in Cyber Security** (CoE) at SRM University (Ramapuram Campus), Chennai
- On 20th November 2024, We (Supraja Technologies) launched our company's

Centre of Excellence in Cyber Security (CoE) at St.Joseph's Institute of Technology, Chennai

❖ Records, Appreciations, Awards & Recognitions etc at a glance :

1. Holds a National Record in "Limca Book of Records – 2017"
2. Ex-Associate Member for National Cyber Safety and Security Standards (NCSSS)
3. Steering Committee Member for United Conference on Cyber Space (UNITEDCON 2020)
4. Judge for the Grand Finale of SIH (Smart India Hackathon 2024) Software Edition for the Nodal Centre at Sri Sai Ram Engineering College, Chennai which is an initiative by Ministry of Education (Govt. of India) and AICTE
5. Received Appreciation from Mr.Amit Narayan, Executive Director at Rajahmundry Asset of Oil and Natural Gas Corporation Limited (ONGC) on 16th December, 2022 for training their employees on Cyber Security
6. Awarded as a "Karmaveer Chakra - 2019", on 12th October 2019 at IIT Delhi, which was instituted by iCONGO in partnership with the United Nations
7. Received Appreciation from Mr.Sandeep Rathore, Commissioner of Police, Chennai on 2nd March 2024 for the exceptional commitment and invaluable contribution as a Member of JURY at Greater Chennai Police Cyber Hackathon 3.0
8. Evaluator for the "Innovative Bharat" which is organized by AICTE and Ministry of Education held on 6th January 2024
9. Judge for the Grand Finale of SIH (Smart India Hackathon 2023) Senior Software Edition for the Nodal Centre at PVPSIT, Vijayawada which is an initiative by Ministry of Education (Govt. of India) and AICTE
10. Appointed as a Member for Board of Studies on 19th April 2025 for the Departments of Cyber Security, Data Science and AIML at Bapatla Engineering College, Bapatla
11. Appointed as a Member for Board of Studies on 12th March 2025 for the Department of Cyber Security at St. Joseph's Institute of Technology, Chennai
12. Appointed as one of the Industry Academia Advisory Council Member on 30th September 2023 for the Department of Cyber Security at VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad
13. Appointed as a Member for Board of Studies on 4th May 2023 for the Department of Cyber Security at Madanapalle Institute of Technology & Sciences, Madanapalle
14. Appointed as a Member for Board of Studies for the Department of MCA at NRI Institute of Technology, Perecherla (Guntur)
15. Awarded as a "Social Media Influencer - 2019", on 30th June 2019 by Jignasa in association with Government of Andhra Pradesh
16. Nominated for "INDIA 500 CEO AWARD 2019"
17. Invited & Interviewed by ETV Andhra Pradesh news channel on 27th July, 2019 for a Special Story Interview on "Spy Apps"

18. Appreciated by Mr.Sridhar, Sub-Inspector of Police at Central Crime Branch, Vijayawada on 23rd October, 2018 for exclusively training him on Special Investigation Course, which will help him to solve the cyber crime cases easily
19. Received a great appreciation from our 1st Tollywood Film Industry client Mr.Saptagiri, for providing Anti-Piracy betaservice for his movie VAJRA KAVACHADARA GOVINDA

❖ Something we are proud of :

We are very happy to inform you all that our company, Supraja Technologies has been shortlisted for **"Top 50 Tech Companies" award 2019**, conferred at InterCon - Dubai, UAE.

Supraja Technologies is one out of thousands of startup companies that were initially screened by InterCon team of 45+ research analysts over a period of three months and the final shortlist includes 150+ firms and we are very proud to inform you all that our company Supraja Technologies also happens to be a part of the same.

Some Glimpses of our Journey

Mr.Santosh Chaluvadi – CEO, Supraja Technologies
Giving hands-on Cyber Security training workshop to the CSE students
at IIT Kharagpur

Mr.Santosh Chaluvadi – CEO, Supraja Technologies was
Invited & Interviewed by ETV Andhra Pradesh news channel on 27th July, 2019
for a Special Story Interview on “Spy Apps”

Mr.Santosh Chaluvadi – CEO, Supraja Technologies was awarded as a **"Social Media Influencer 2019"** in recognition of his remarkable achievements in the social media as a part of First International Social Media Festival on 30th June 2019 by Jignasa in association with Government of Andhra Pradesh

On 23rd October 2018 Mr.Santosh Chaluvadi, CEO - Supraja Technologies and Mr.Krishna Chaitanya, CTO - Supraja Technologies had successfully completed delivering Special

Investigation Course training in Cyber Security to Mr.Sridhar Garu, Sub-Inspector of Police at Central Crime Branch, Vijayawada which will help him to solve the cyber crime cases easily

ETV Andhra Pradesh News Channel interviewed Mr.Santosh Chaluvadi, CEO - Supraja Technologies for his achievements in the domain of Cyber Security & Digital Marketing

Supraja Technologies was invited by Indian Air Force (Air Wing NCC) to deliver a session on Latest Cyber Crimes & Awareness for the NCC cadets, staff and officers on 4th July, 2019

Supraja Technologies – CEO, CTO & CMO with
Indian Air Force (Air Wing NCC) Group Captain Sandeep Gupta.
We thank Mr.Sandeep Gupta for inviting us on 4th July, 2019 to deliver a session on Latest Cyber Crimes & Awareness for the Indian Air Force (Air Wing NCC) cadets, staff & officers

TABLE OF CONTENTS

1. Introduction

1.1 Overall Description

Insider threats are one of the toughest problems in cybersecurity right now. Unlike hackers breaking in from the outside, insiders—think employees, contractors, or anyone with special access—already have the keys to the castle. That trust means they can slip past the usual security barriers, making them much harder to spot. Sometimes they steal data or sabotage systems on purpose. Other times, they make innocent mistakes that end up causing just as much harm. Either way, the fallout can be expensive and embarrassing for organizations. The usual security tools—firewalls, intrusion detection, signature-based monitoring—do a decent job keeping out known threats from outside. But they don't cut it with insiders. Why? Because insiders' actions usually look normal on the surface, so they fly under the radar. That's why more companies are turning to behavior-based security. Instead of just looking for known attack patterns, these systems watch how users behave and flag anything that seems off.

This project introduces an Insider Threat Prediction System using User and Entity Behavior Analytics (UEBA). The system digs through user activity logs to build up profiles of normal behavior, then hunts for anything suspicious as time goes on. It uses several layers of detection: rule-based risk scoring, machine learning with Isolation Forest to spot anomalies, and deep learning with LSTM networks for predictions. By putting all these pieces together, the system catches threats in real time and even predicts them before they strike. The system also taps into the MITRE ATT&CK framework, mapping what users do to established tactics and techniques. That makes it a lot easier to understand what's going on when a threat pops up. There's an interactive dashboard built with Python and Tkinter, showing risk scores, anomalies, alerts, and MITRE patterns in a format that's easy for a security team to work with. All in all, this system offers a practical and scalable way to spot insider threats and adds some solid academic value to the cybersecurity field.

2. Existing System

Most organizations still lean on the usual security tools for spotting insider threats—stuff like access controls, log monitoring, intrusion detection, and SIEM platforms. These systems stick to enforcing set policies and catching attacks that fit known patterns. They log what users do, but people usually review those logs only now and then, or after something's already gone wrong. So, most of the time, detection happens after the fact—not before.

The problem is, these tools follow static rules and use simple alerts, like flagging too many failed logins or someone poking around where they shouldn't. Sure, that catches obvious bad behavior, but it misses the small, early signs that someone on the inside is going rogue. Too often, organizations find out about insider threats after the damage is done. Plus, these systems treat everyone the same, without looking at how each person usually acts or what their normal activity looks like.

What's wrong with the old way? For starters, it doesn't really pay attention to long-term patterns or subtle shifts in how people behave, so it's easy to miss slow, sneaky attacks. The reliance on rigid rules and known signatures means you get flooded with false alarms, and security teams burn out sifting through noise. It's all reactive—by the time an alert pops up, the suspicious activities already happened. There's no real way to get ahead of problems.

Another issue: most of these tools haven't caught up with machine learning or deep learning, so they struggle to keep pace as insider tactics change. And because they don't use frameworks like MITRE ATTACK to put things in context, analysts spend even more time trying to make sense of alerts. Finally, the dashboards and visualizations are often clunky or missing, which just slows down investigations and decision-making for security teams.

3. Proposed System

This system takes insider threat detection to another level. Instead of just relying on old-school security tools, it digs deep into user behavior. It constantly watches user activity logs and uses that data to create detailed behavioral profiles for everyone. It doesn't just look at raw logs, either—it normalizes the data, groups it by user and by day, and engineers features that actually matter. Stuff like how often someone logs in, how many files they access, how much data they move around, and even when they tend to be active. By focusing on these patterns, the system spots both sudden and slow changes in how people behave.

The detection approach isn't just one-dimensional, either. There's a rule-based risk scoring engine that keeps an eye on sensitive actions. When someone does something risky and crosses a set threshold, the system immediately sends out an alert and explains why. At the same time, it uses an Isolation Forest model to sniff out weird behavior that doesn't fit the usual patterns—even when it's never seen that kind of thing before. And for the long game, it brings in an LSTM deep learning model. This one learns how each person typically acts over time and predicts how likely someone is to become a risk in the future. To tie it all together, every suspicious activity gets mapped to the MITRE ATT&CK framework, so security teams have clear, standard context for what's happening. All this feeds into a dashboard that looks and feels like a real SOC tool—risk scores, anomalies, alerts, and visual breakdowns, all easy to spot and understand.

So, what's actually better about this system? For starters, it doesn't just catch obvious threats. It notices when someone's behavior shifts in subtle ways, even if they still look "legit" on the surface. By blending rule-based logic, anomaly detection, and deep learning, it cuts down on false alarms and gets more accurate at finding real threats. The LSTM model means it can flag risky users earlier, giving security teams a chance to act before things go sideways.

Mapping everything to MITRE ATT&CK makes alerts clearer and helps everyone speak the same language about threats. The dashboard gives analysts a live, clear view of what's going on, speeding up investigations and responses. Plus, the whole thing scales up or down, explains its actions, and adapts to any organization—whether you're running research or locking down a real business.

4. System Analysis

4.1 Feasibility Study

The Insider Threat Prediction System runs on a modular, layered setup, built for easy scaling and straightforward maintenance. It starts with a data ingestion layer that takes in user activity logs as CSV files, then cleans up the timestamps and standardizes the data. Next, a feature engineering module steps in, grouping user activities by day and by user, and pulling out all sorts of features—behavioral, contextual, and time-based. These details feed into several detection engines: a rule-based risk engine, an anomaly detector that uses Isolation Forest, and a deep learning model running on LSTM networks. The system doesn't stop there—it ties in the MITRE ATT&CK framework to link activity patterns to known adversary tactics. Everything hooks into a central controller, and the whole thing is visualized through a Tkinter dashboard. Security analysts get real-time alerts, risk scores, and detailed analytics, all in one place.

Three key considerations in feasibility analysis are:

Economic Feasibility

Technical Feasibility

Social Feasibility

4.1.1 Economic Feasibility

This system makes sense from a cost perspective. It's built with open-source tools like Python, Pandas, Scikit-learn, TensorFlow, and Tkinter, so there's no need to worry about licensing fees or subscriptions. You can run it on regular computers—no fancy hardware required—which keeps both deployment and ongoing maintenance affordable. Plus, because it helps spot insider threats early, it helps cut down on expensive problems like data breaches, stolen intellectual property, or sudden disruptions. All in all, it's a security solution that actually saves organizations money.

Besides, the utilization of open-source libraries that are widely adopted, together with a desktop-based deployment model, diminishes the need for specialized technical competencies at installation and operation. This limits the training and support costs while enabling the organization to make use of existing IT resources. The modular design of the system will also ensure that enhancements or updates in the future can similarly be made on an incremental basis without requiring substantial redevelopment costs.

4.1.2 Technical Feasibility

This system actually works because it uses tech and machine learning methods people trust and rely on. Python makes things easier—it handles data crunching, machine learning, and building interfaces all in one place. Algorithms like Isolation Forest and LSTM, great for picking up patterns in behavior and time-series

data. The modular design is a big plus too. You can work on each part by itself, test it, improve it, and nothing gets tangled up. It chews through huge piles of log data fast, and since it runs heavy stuff in the background, the main interface stays quick and responsive. If you ever want more—like new data sources, different models, or fancier graphs—you can add those in without much trouble.

4.1.3 Social Feasibility

Looking at this from a social angle, the system actually works in favor of organizations. Instead of snooping or crossing privacy lines, it sticks to monitoring behaviors, using activity logs that already exist from daily operations. That means privacy stays intact. Plus, the system doesn't just spit out warnings—it explains them, and it uses familiar MITRE ATT&CK standards, so everyone's on the same page. This kind of transparency helps security teams and managers trust what the system tells them. On top of that, it boosts security awareness across the board. Analysts get a clearer picture of insider threats and the risks that come with them. Bottom line: the system encourages ethical security monitoring and gives organizations a stronger grip on protecting their sensitive data and resources.

4.1.4 Operational Feasibility

The Insider Threat Prediction System actually works in practice. It's built to be straightforward, so you don't have to jump through hoops to set it up or use it day-to-day. The dashboard looks and feels like what you'd find in a typical Security Operations Center, with everything security analysts need right in front of them—user behavior, risk scores, anomalies, alerts. You get clearly labeled controls for all the important stuff like loading data, running models, and seeing results. No need for a long training session.

It runs on the log data you already have, so you don't need to change how your organization does things. Heavy tasks—like finding anomalies or training models—happen in the background, so the system stays quick and smooth. The modular setup also means you can tweak or upgrade parts of it later without causing problems. In short, you can drop this system into your existing security workflow without much hassle.

4.2 Input And Output Design

4.2.1 Input Design

Input and output design really matter when it comes to how well the Insider Threat Prediction System works. Good input design means the system grabs and processes user activity data the right way. As for output, it's all about making the results easy to understand and useful. This system deals with actual log data smoothly, giving security analysts clear visualizations and alerts they can act on right away.

4.2.2 Objectives

The system's input design takes in user activity logs in CSV format. These logs can come from all sorts of places like authentication systems, file servers, network monitoring tools, you name it. Every organization has its own way of formatting logs, so the system steps in to make sense of the mess. It automatically hunts down and normalizes important details like timestamps, user IDs, event types, resource names, and data transfer sizes.

The input module doesn't get tripped up by missing or inconsistent data. If something's missing, it fills in the gaps and even creates timestamps if needed. This kind of flexibility keeps the system steady, even when it's dealing with messy or half-baked datasets. By turning all those raw logs into a clean, standard format, the system makes it possible to pull out useful features, spot patterns, and analyze behavior. In the end, this solid input design gives insider threat detection a strong place to start.

4.2.3 Output Design

The system's output design keeps things simple and clear. It gives security analysts what they need: risk scores for every user, alerts about insider threats, results from anomaly detection, and activity mapped to MITRE ATT&CK. All of this shows up in a way that's easy to understand and quick to use.

Everything comes together on an interactive dashboard. You get tables, charts, heatmaps, and alert logs, all in one place. Visuals like risk score bar charts and MITRE ATT&CK heatmaps make it easier to spot what's important and act fast. If you need to share or dig into the data, you can export results as a CSV file.

5. Module Description

Module 1: Data Ingestion Module

Here's where everything starts. This module loads user activity logs into the system, working with CSV files. It takes care of the basics—figuring out timestamps, normalizing columns, cleaning up missing data, and making sure everything checks out. By the end, raw logs turn into clean, structured data that's ready for analysis.

Module 2: Feature Engineering Module

Next up, this module digs into the data to pull out meaningful behavioral features. It sums up what each user does every day, like counting events, tallying up unique resources accessed, measuring data transfer, tracking failed logins, and calculating average gaps between actions. It also pulls out time-based features to spot changes in how users behave over time.

Module 3: MITRE ATT&CK Mapping Module

This part links user activity to MITRE ATT&CK tactics and techniques. Basically, it turns low-level log events into standard threat categories, making it easier to understand what's going on and helping analysts connect the dots during investigations.

Module 4: Rule-Based Risk Scoring Module

Here, the system assigns set risk scores to different user actions and adds them up for each person. If someone's score crosses a certain line, the module raises an incident alert. The whole process is straightforward and explainable, so analysts always know why a user was flagged.

Module 5: Anomaly Detection Module

This module uses the Isolation Forest algorithm to spot users acting out of the ordinary. By analyzing the Behavioral features, it detects anyone whose activity stands out from the crowd. That way, it picks up on insider threats that might otherwise slip through the cracks.

Module 6: LSTM-Based Prediction Module

Now the system digs deeper with time-series analysis, using an LSTM neural network to learn long-term Patterns from historical data. It predicts insider risk, giving security teams an early heads-up about potential threats.

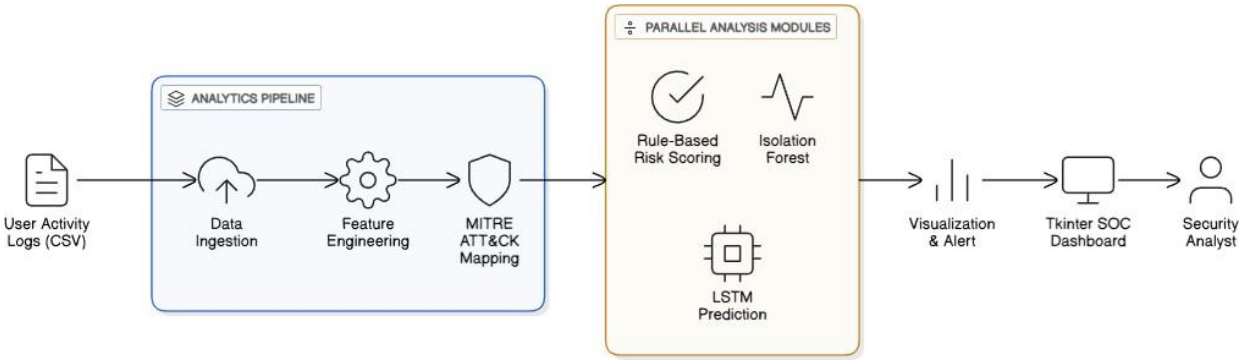
Module 7: Visualization and Alert Module

All the results and insights show up here, on an interactive dashboard. You'll see risk scores, lists of anomalies, alerts, tables, charts, and even MITRE ATT&CK heatmaps all in one place for easy monitoring and fast follow-up.

Module 8: Dashboard and Control Module

Finally, this is the main hub for security analysts. Here, they can load data, run analysis, train models, and look at the results. Background processing keeps everything running smoothly, so the system stays quick and responsive.

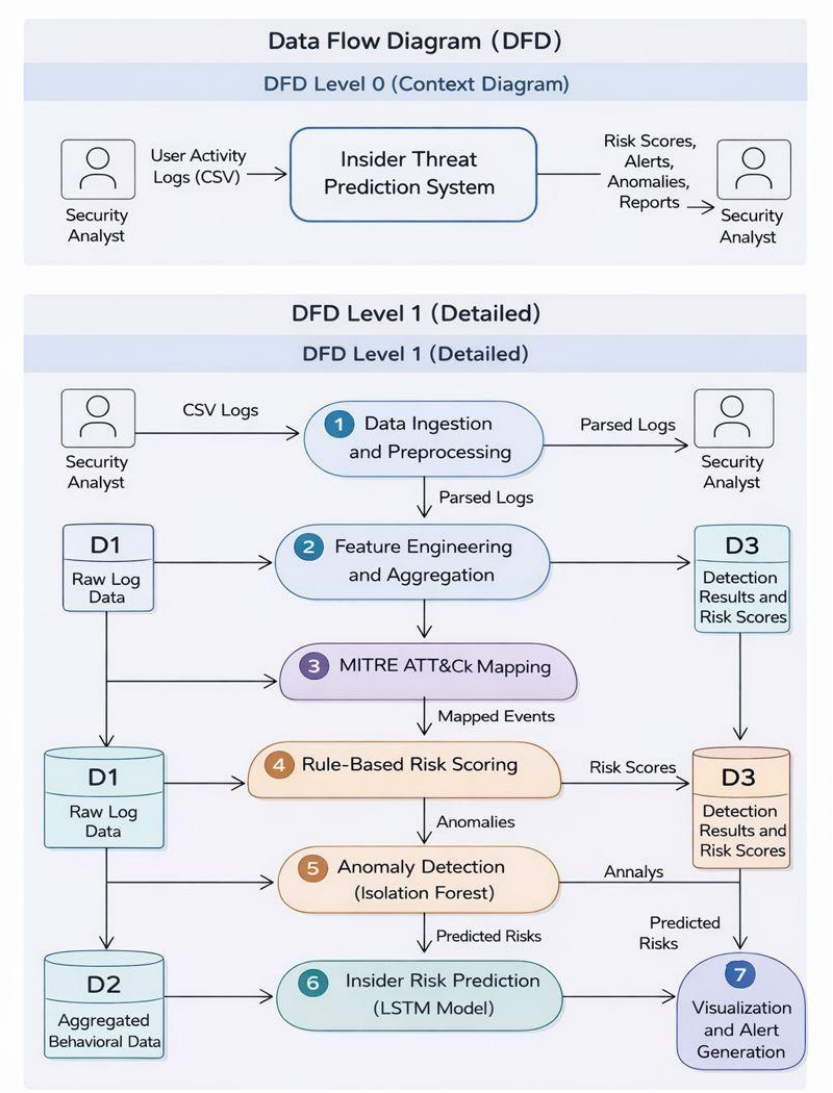
5.1 System Architecture



The system begins by collecting user activity logs in CSV format, which include events such as logins, file access, and data transfers. These logs are first passed through a structured analytics pipeline, where data ingestion cleans and organizes the raw logs, followed by feature engineering that extracts meaningful behavioral patterns from user activities. The processed events are then enriched using MITRE ATT&CK mapping to associate each action with a recognized security tactic or technique. After this, the data flows into parallel analysis modules, where rule-based risk scoring evaluates the severity of user actions, Isolation Forest detects unusual behavior by comparing users against normal patterns, and an LSTM model analyzes historical behavior to estimate insider risk trends. The combined results are then used to generate visualizations and alerts, which are displayed through a Tkinter-based SOC dashboard, enabling security analysts to easily monitor risks, investigate anomalies, and take timely action to mitigate potential insider threats.

6. System Design

6.1 Data Flow Diagram



The Insider Threat Prediction System has a diagram that shows how data moves around. This diagram is called the Data Flow Diagram. It shows what happens to the data from the start to the end. The Insider Threat Prediction System collects logs of what the users doing. Then it processes these logs analyzes them and turns them into information about security.

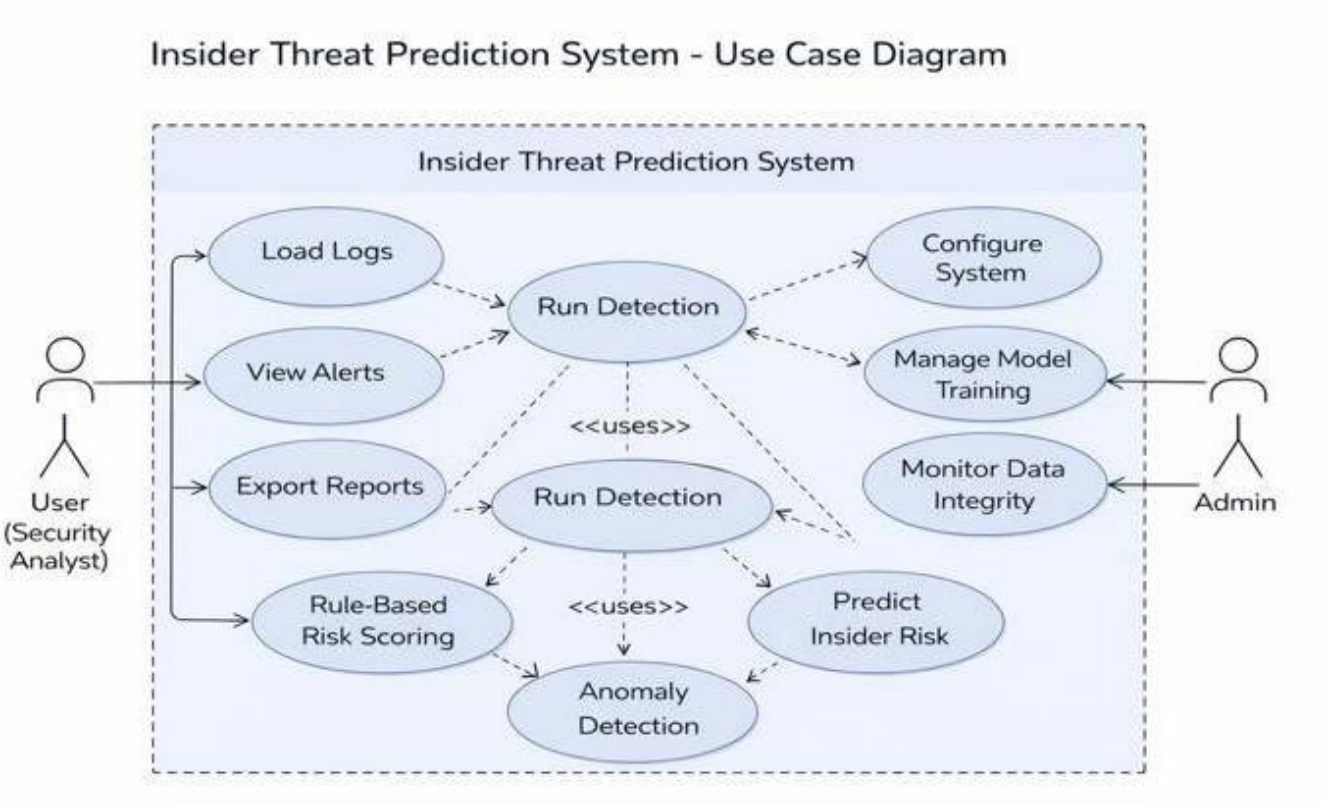
The Data Flow Diagram also shows how different parts of the Insider Threat Prediction System work together. It shows how things outside the system the parts inside the system that do the work and the places where data is stored all interact, with each other to detect insider threats.

- The Security Analyst gives the Data Ingestion module user activity logs, in CSV format at the beginning.
- The system looks at the input data cleans it up and makes sure everything is consistent and correct.
- The Data Ingestion module then stores the cleaned-up data. Sends it to the Feature Engineering module.
- At the Feature Engineering module user activities are put together. Behavioral features of the user activities are found.
- The user activities are what the system focuses on to get the features that show how the users behave.
- The MITRE ATT&CK Mapping module gets the enriched data to figure out what is going on with the events. It does this by using adversary tactics and techniques.
- The data then goes to a few places at the same time. It goes to the Rule-Based Risk Scoring module. It also goes to the Anomaly Detection module which uses something called Isolation Forest. It goes to the LSTM-based Prediction module.

Each of these modules looks at the MITRE ATT&CK Mapping module data, on its own to find things that seem suspicious things that're not normal and potential insider risks that the LSTM based Prediction module predicts. The MITRE ATT&CK Mapping module data is used by each module to identify these things.

Finally, the detection results, risk scores, and alerts are consolidated and sent to the Visualization and Alert module. The system presents the outputs through an interactive dashboard in the form of charts, tables, alerts, and reports, enabling the Security Analyst to effectively monitor and respond to potential insider threats.

6.2 Use Case Diagram



The Insider Threat Prediction System is shown in a kind of diagram. This diagram shows what the system can do from the users point of view. It shows how the user works with the system to find and watch insider threats. The user does this by using the dashboard. The dashboard has tools that help the user do their job. The diagram gives us a look, at how the system works and how all the parts are connected. The Insider Threat Prediction System is what we are focusing on here.

The User, who is the Security Analyst is the person in the system. They start everything. Control what the system does. It all begins when the User starts the application. This opens up the window, which has a sidebar with navigation and a status bar. The User uses the sidebar to pick what they want to do. They can choose from things like Aggregates, Anomaly Detection, Prediction, Risk Scores and Model Training. The User is, in charge of what happens in the system.

- The user can load user activity logs, in CSV format. They can make some sample data to try things out.
- When the user has the data the system makes sure the data is good and ready to use.
- The system does this by doing some things to the data to make it all the same and then it does something called MITRE ATT&CK mapping.
- After that the user can do some things with the user activity logs like figure out what people do every day or look at how things change over time or make a graph to show how things are connected.
- The user can compute these things like what each user does every day and the system will show the user activity logs in a way that's easy to understand.
- The user can do a lot of things to detect and predict problems. They can use different analysis modules.

These are things like figuring out if something is a risk, by using rules finding things that're not normal with something called Isolation Forest and training a thing called an LSTM model to predict if someone inside is a risk.

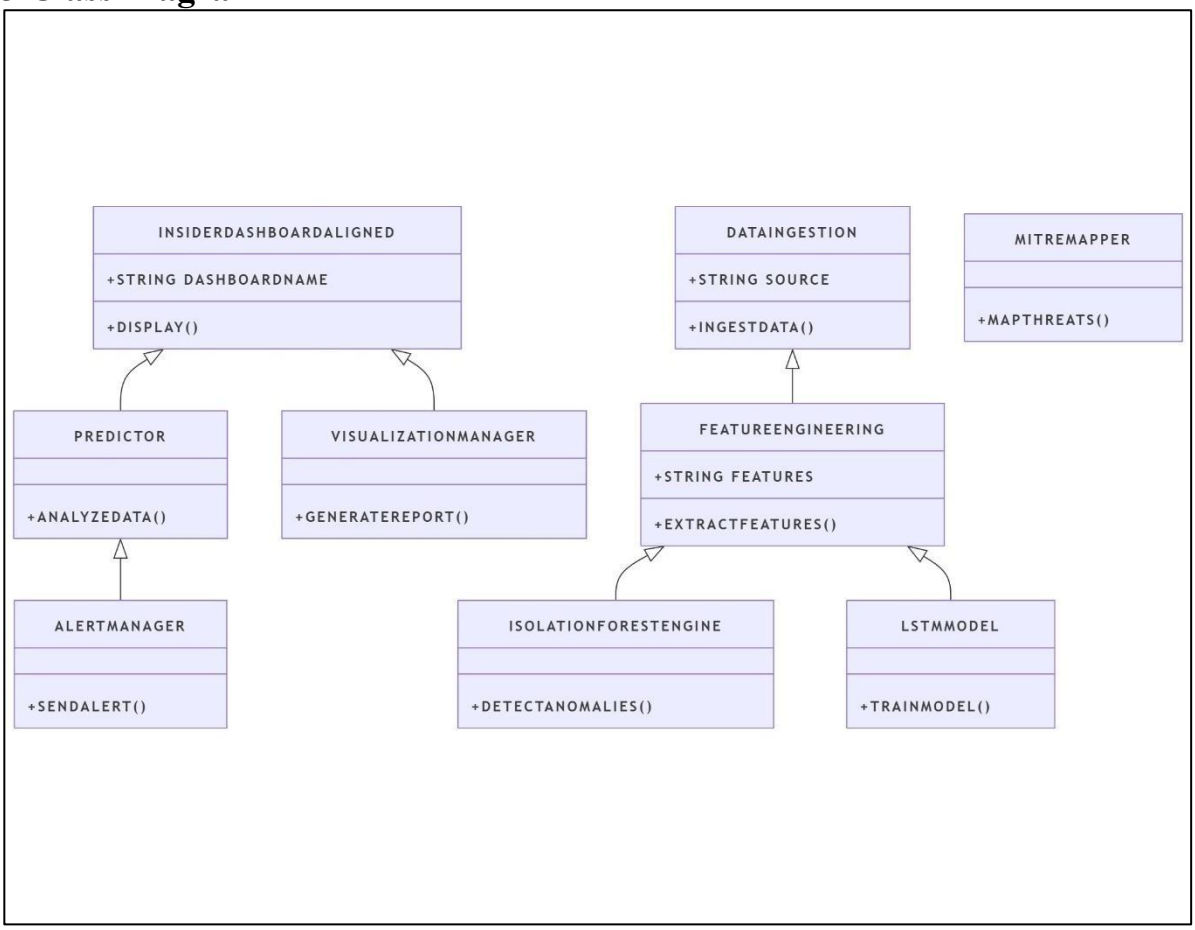
The system lets the user see what the LSTM model predicts and any warnings it gives from the deep learning model. The user can look at the LSTM predictions and the alerts that come from the deep learning model, which's really the LSTM model to see what is going on with the insider risk.

The system gives us ways to look at the information it provides. We can see charts that show us the risks, scores that tell us how big the risks are and maps that use something called MITRE ATT&CK to show us where the problems are. We can also see what has been happening recently what incidents have occurred what is not normal and tables that put all the information together.

We can even take the results like the risk scores and reports of things that're not normal and save them to our computer in a format called CSV so we can look at them more closely or put them in a document.

The use case diagram also includes system maintenance actions such as clearing data and closing the application. Overall, the diagram captures the complete interaction flow between the user and the Insider Threat Prediction System, highlighting how data is processed, analyzed, and presented to support effective insider threat monitoring and decision-making.

6.3 Class Diagram



The Class Diagram shows what the Insider Threat Prediction System looks like when it is not working. It shows the parts of the Insider Threat Prediction System what they are made of what they can do and how they work together. The Class Diagram gives us an overview of how all the different parts of the Insider Threat Prediction System are set up and how they work together to find and show insider threats.

The Insider Dashboard Aligned class is really the part of the system. It has all the settings, for the dashboard. It has ways to show and manage the interface that you see. This class works with the Predictor class and the Visualization Manager class to get the data analyzed and to show the results in a way. The Insider Dashboard Aligned class is important because it helps the Predictor class and the Visualization Manager class work together.

The Predictor class is what we use to look at the data we have processed to see if there are any insider threats. It has methods that help us analyze what the users are doing and come up with some results that we need to detect things.

The Alert Manager class works with the Predictor class. Is in charge of generating alerts, by sending out notifications when the Predictor class finds high-risk or anomalous activities that the Predictor class has detected.

The Data Ingestion class is in charge of the input part of the system. It reads logs of what the user sre doing from different places and puts this data into the system so it can be used later. The Data Ingestion class then sends the data it has ingested to the Feature Engineering class. The Feature Engineering class looks at the data from the Data Ingestion class. Finds the important things, about how the users are behaving that we need to know for our analysis. The Feature Engineering class has things that describe the features it finds and ways to find and combine these features.

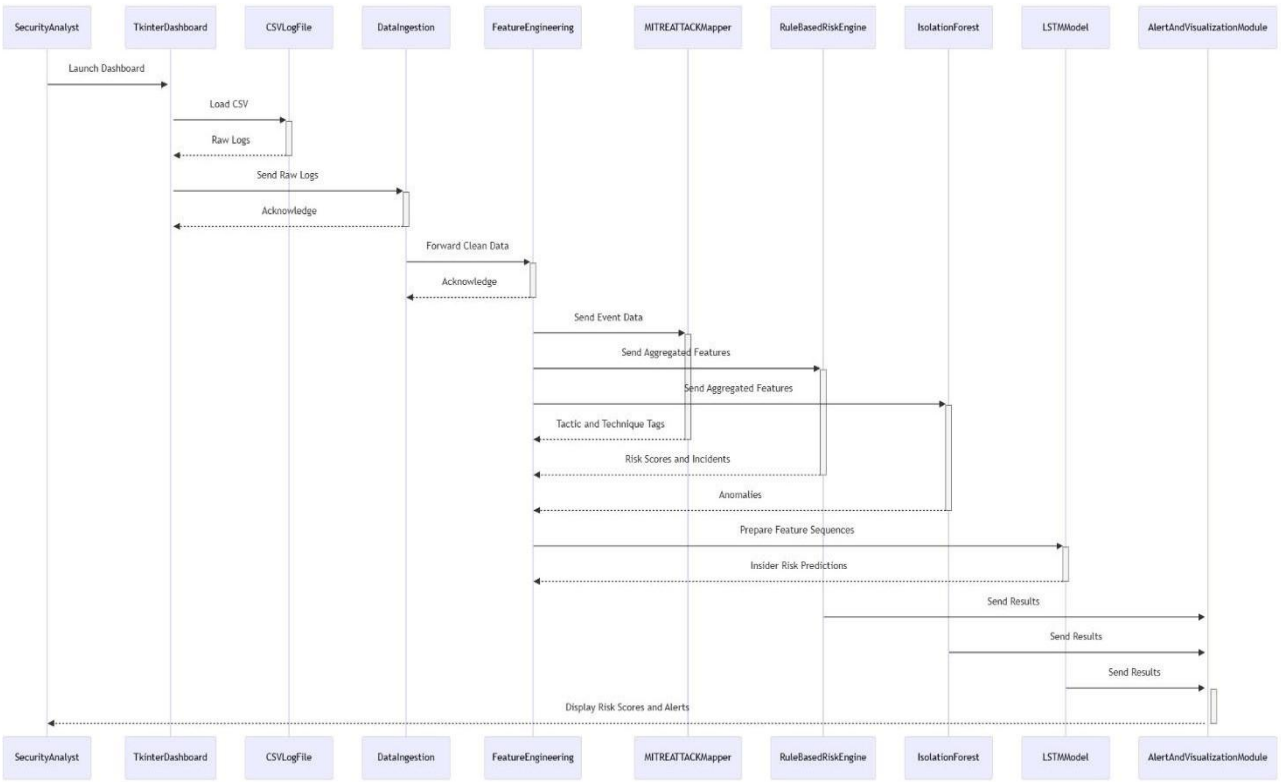
The MITRE Mapper class is used to connect what happens after an event is processed to the MITRE ATT&CK tactics and techniques. This helps people understand threats better and makes security analysis more standard. The MITRE Mapper class does this by making a map of these events to the MITRE ATT&CK tactics and techniques.

The Isolation Forest Engine and LSTM Model classes are the parts of the system that use machine learning and deep learning. The Isolation Forest Engine looks for behavior that's not normal using techniques that do not need any help. The LSTM Model is in charge of teaching the system and guessing the risk of someone inside the system based on what they do, over time.

The Isolation Forest Engine and the LSTM Model classes work with the Feature Engineering class because they need the information that is taken out of the data to find and guess things. The Isolation Forest Engine and the LSTM Model classes use this information to do their jobs.

The Visualization Manager class is what handles making reports and showing results in a way that's easy to understand. It shows things like risk scores and alerts and anomalies and reports that have been analyzed through the dashboard interface. This helps security analysts do their job of monitoring things. The Visualization Manager class is really important, for this because it helps security analysts see what is going on with the risk scores and alerts and anomalies and analytical reports.
Overall, the class diagram demonstrates a modular and well-structured design where data flows from ingestion and feature engineering to detection, alerting, and visualization. This separation of responsibilities improves system maintainability, scalability, and clarity.

6.4 Sequence Diagram



The Sequence Diagram shows how the Insider Threat Prediction System and the user interact with each other over time. It shows what happens step by step from when the user uses the system to when they see the final alert. The Insider Threat Prediction System has parts that work together and the Sequence Diagram shows how data moves from one part, to another. It helps us see the order in which messages are sent and how the Insider Threat Prediction System responds to the user.

The Security Analyst starts by opening the Tkinter Dashboard. When the Security Analyst picks a CSV log file the dashboard starts loading the data. It reads the raw log data from the CSV file. Sends it to the Data Ingestion module. The system says the data was sent successfully then it gets ready for the steps, which are preprocessing and normalization of the log data. The Security Analyst can then work with the log data, in the Tkinter Dashboard.

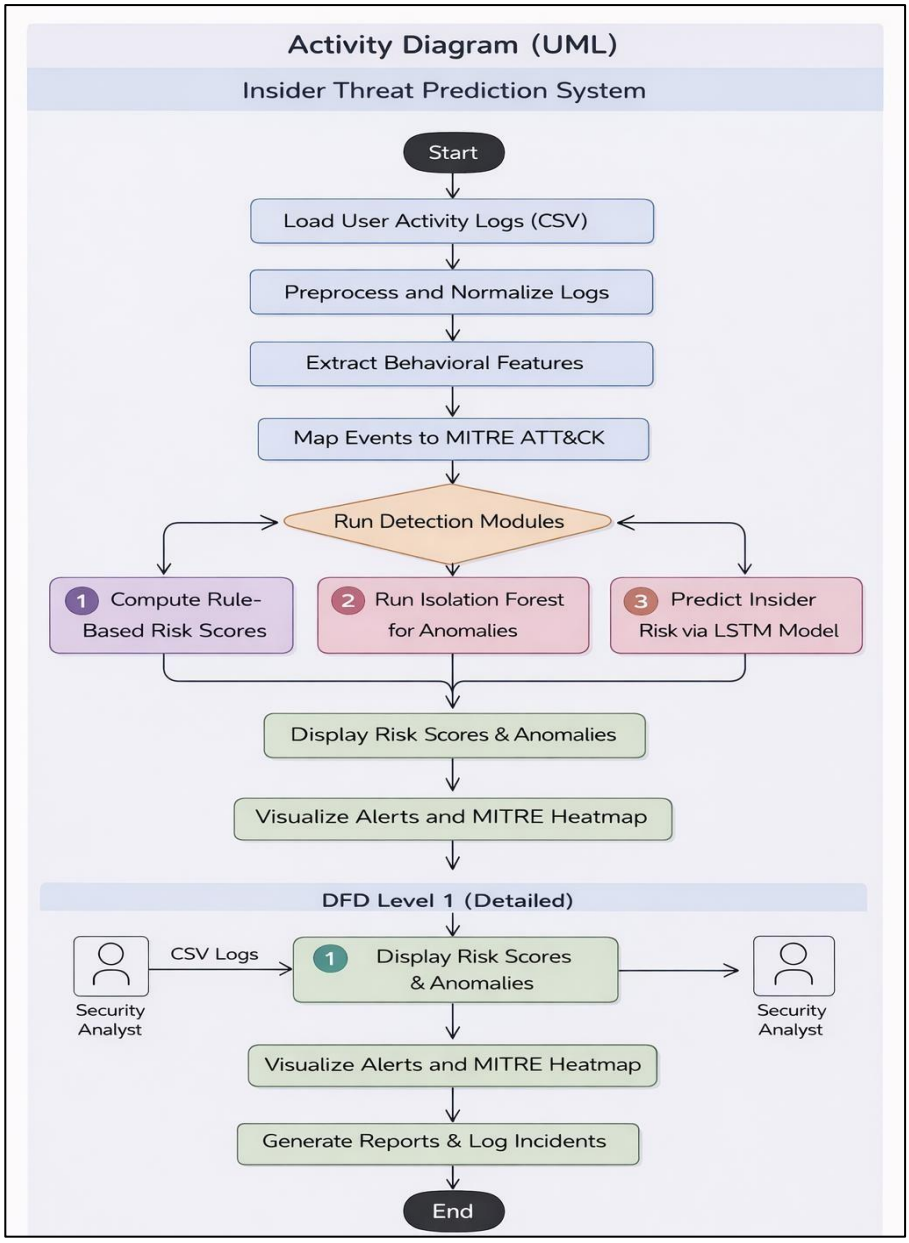
The cleaned data is sent to the Feature Engineering module. This is where the Feature Engineering module pulls out the features and aggregated metrics from the data. The processed event data goes to the MITRE ATT&CK Mapper. The MITRE ATT&CK Mapper then adds the tactics and techniques to the events.

The MITRE ATT&CK Mapper shares these tagged features with the Rule-Based Risk Engine. The Rule-Based Risk Engine uses these tagged features to calculate the risk scores and create incident alerts that are based on the rules that were set up beforehand. The Rule-Based Risk Engine and the MITRE ATT&CK Mapper and the Feature Engineering module all work with the data to get the results, from the Feature Engineering module and the MITRE ATT&CK Mapper.

At the time the combined features are sent to the Isolation Forest module. This module looks at how the user behaves to find things that're not normal. At the time the Feature Engineering module gets the features ready and sends them to the LSTM Model to predict if there is a risk from inside. The LSTM Model looks at these features. Makes predictions about the risk, from inside based on how likely it is to happen. The LSTM Model is used for insider risk prediction. It does this by looking at the features that the Feature Engineering module sends to it.

The results from the Rule-Based Risk Engine, Isolation Forest, and LSTM Model are then transmitted to the Alert and Visualization Module. This module consolidates risk scores, anomalies, and predictive alerts and displays them through the Tkinter Dashboard. Finally, the Security Analyst views the visualized risk scores, alerts, and reports, completing the interaction sequence.

6.5 Activity Diagram



The Activity Diagram shows us the workflow of the Insider Threat Prediction System. It tells us what happens from the moment we put in the data to the moment we get the alert and the report.

- The Insider Threat Prediction System takes the user activity logs. Processes them. Then it uses different techniques to analyze these logs.

- After that the results are shown to the security analyst so they can make a decision about the Insider Threat Prediction System. The Insider Threat Prediction System is really important, for this process.
- The activity flow starts at the Start node. This is where the system gets set up.
- The first thing that happens is that the system loads user activity logs. These logs are in CSV format.
- The user activity logs have a lot of information. They have things like what the user did when they did it what they looked at and what they moved around.
- After the system loads the logs, it moves on to making sure the data is good. This is called the preprocessing and normalization stage.
- At this stage the system fixes problems with the logs. It deals with things that are missing or not consistent. It also makes sure everything is in the format.

The system does something with the information it has after it is cleaned up. It looks at how the users of the system behave. This means it looks at things like how often they log in what files they look at how data they move around and when they do these things. The system uses these details about the user's behavior to figure out if someone is doing something they should not be doing with the system, which is called insider threat detection and the details about the user's behavior are really important, for insider threat detection.

The next thing we are going to do is map events to the MITRE ATT&CK framework. We take each thing a user does. Put it into the right category of tactics and techniques. This gives us a way to understand threats and makes it easier to figure out what is going on with the behaviors we detect. The MITRE ATT&CK framework is really helpful, for this because it gives us a language to talk about the tactics and techniques that the MITRE ATT&CK framework uses.

- The system is now at the Run Detection Modules stage like the MITRE mapping says. This is where the system has to make a decision and do a thing at the same time.
- First the system figures out how risky something is by giving points to actions that are sensitive and adding up these points for each user.
- Second the system uses something called the Isolation Forest algorithm to find behavior that's not normal for a user.
- Third the system uses the LSTM model to look at what a user has done in the past to guess if they will be a risk, in the future. The system is looking at the MITRE mapping and the Run Detection Modules phase to do all of this. The Run Detection Modules phase is a part of the system.

The system takes the results from all three detection modules. Combines them. Then it shows the risk scores and anything that is not normal. This helps people see the points of what the detection modules found. It points out users who're at high risk and behavior that seems suspicious. The system also uses pictures to show alerts and the MITRE ATT&CK heatmap. This helps security analysts understand where the threats are coming from and how bad they are, by looking at graphs and charts.

The security analyst looks at the system to see the risk scores and other important things, like anomalies and alerts. They also look at the MITRE heatmaps. When they do this the system makes reports. Writes down incidents that need to be looked into more closely. This is done so that everything can be checked and made sure it is okay.

Finally, the workflow reaches the End node, indicating the completion of the analysis cycle. The activity diagram clearly demonstrates how the system integrates data preprocessing, multiple detection techniques, and visualization into a unified insider threat monitoring process.

7. Requirement Specification

The Insider Threat Prediction System has a list of things it needs to do. This list is called the Requirement Specification. It talks about what the Insider Threat Prediction System should be able to do and what it should not do. It also talks about the rules the Insider Threat Prediction System has to follow when it is working.
The Requirement Specification also mentions how good the Insider Threat Prediction System should be.

7.1 Functional Requirements

The system has to do things and functional requirements tell us what these things are. These are the functions that the system must be able to do. The functional requirements are really, about what the system needs to be able to do for us.

It will let the user load logs of what the user has been doing in a format called CSV. The user can load user activity logs in CSV format. This way the user can see what the user has been doing. The system allows the user to load user activity logs, in CSV format.

It needs to be able to get the input logs ready and make sure they are all, in the format. This means it has to handle input logs that are missing some information or have data. The system has to support this preprocessing and normalization of input logs.

It will look at how people use it like how often they log in what files they look at how data they move around and when they are active. The system will extract these features, including login frequency, file access patterns, data transfer volume and temporal activity metrics to get a better idea of what people are doing.

It will gather all the things that each user does. It will do this for each user and for each day. The system will keep track of user activities for every user and every single day. This means the system will collect all the user activities and sort them out by user and, by day.

It will show how user events are connected to the things that attackers do which are called MITRE ATT&CK tactics and techniques. The system will help us understand what the MITRE ATT&CK tactics and techniques are and how they are used by matching them to user events. This will make it easier to see what is happening with the MITRE ATT&CK tactics and techniques when user events occur.

It will figure out risk scores based on rules. It will use predefined weights to decide how much risk is involved. The system will use these weights to calculate the risk scores, for the risk.

It will find things that users do by using something called the Isolation Forest algorithm. This algorithm is used to detect user behavior that's not normal. The system will use the Isolation Forest algorithm to figure out when users are doing things that they do not usually do.

It will use a kind of computer program called an LSTM model to predict when someone might be a threat from the inside. It does this by looking at what people have done in the past which is called historical behavior data. The system is going to train the LSTM model using this historical behavior data to make it better, at predicting insider threats.

It will use a kind of computer program called the LSTM model to figure out how likely it is that someone will be a security risk from inside the company. This LSTM model was taught with a lot of information so it can make predictions, about insider risk probabilities. The system is going to use this trained LSTM model to predict insider risk probabilities.

It will send out warnings when the risk levels get too high or when something strange happens. This is what the system does when it finds risk thresholds that're too high or anomalies that are detected by the system. The system is always looking for risk thresholds that are exceeded or anomalies that the system can detect.

It will show you the risk scores and the anomalies and the alerts and the incidents on a screen that you can interact with. This screen is like a dashboard where you can see the risk scores and the anomalies and the alerts and the incidents all, in one place.

It will show threat activity in a way using charts and MITRE ATT&CK heatmaps. It will help us see what the threat activity is doing. The system will use these MITRE ATT&CK heatmaps to make it easy to understand the threat activity.

It will let users get risk scores and anomaly reports out in a format that they can use in a spreadsheet. This format is called CSV. The system will allow users to export these risk scores and anomaly reports in CSV format so they can look at them easily.

The system needs to be able to clear system data and reset system data when we need to do that. We have to be able to clear system data and reset system data in the system.

7.2 Non-Functional Requirements

Non-functional requirements are really about how the system should work. They define things, like how the system performs how easy it is to use how reliable the system is and how secure the system is. The non-functional requirements are important because they talk about the performance of the system the usability of the system the reliability of the system and the security of the system.

7.2.1 Performance Requirements

The system will take care of log datasets that're not too big within a time that is okay for people to wait. The system has to make sure that these log datasets are handled enough so that people do not get frustrated. The log datasets that the system processes are moderate in size. This means the system is good, at handling log datasets that're this size within a time that people think is acceptable.

The computer will do some things in the background so that the user interface does not get slow. This is really helpful for tasks that need a lot of computer power. We want the user interface to always be responsive. Background processing is good, for these kinds of intensive tasks.

7.2.2 Usability Requirements

The system needs to have a simple and easy to use interface that looks like a Security Operations Center dashboard. The interface should be graphical and easy to understand like a SOC dashboard. This will help the users to easily navigate and use the system. The system will have a user- interface that is similar, to a SOC dashboard.

The system operations need to be easy to use. They should have clearly labeled controls and menus. This way the system operations are simple to understand and work with. The system operations should be accessible to everyone, through these labeled controls and menus.

The security analysts will not need a lot of training to use the system in a way. They can learn to use it and it will be easy for them to do their job with the system. The system is made so that security analysts can start using it away with minimal training.

7.2.3 Reliability Requirements

The system needs to be able to deal with incomplete information that people put in. The system should not stop working when it gets incomplete information. The system has to handle this kind of thing without crashing. When people put in incomplete input data the system should still work properly. The system has to be able to handle incomplete input data.

The system will give us error messages that make sense when something goes wrong with the system. This means the system will tell us what is wrong, with the system so we can fix the problem with the system.

7.2.4 Security Requirements

The system will only look at log files that the user is allowed to give it. The system has to get these log files from the user. The user has to give the system the log files that the system is supposed to process. The system is only going to process the authorized log files that the user provides.

We need to make sure that only people who are supposed to see the sensitive results can actually see them. This includes things like risk scores and alerts. Access to these things should be limited to authorized users, like the people who're, in charge of dealing with this kind of information. Authorized users should be able to see risk scores and alerts.

7.2.5 Scalability Requirements

The system needs to be able to handle more and more user activity logs as the number of users increases. This means the system has to be scalable to deal with the rising volume of user activity logs. The user activity logs will keep growing. The system must be able to handle this growth.

The design of this thing is made to be flexible so we can add models to detect things later on. The modular design will make it easy to add detection models in the future. This way the modular design of the system allows us to integrate detection models.

Additionally, the system promotes scalability on the processing dimension by adopting decomposition on functional modules related to data ingestion, feature creation, detection, and data visualization. With such decomposition, computationally intensive processes such as aggregation, anomaly identification, and modeling can remain optimized or parallelized depending on the data size. Hence, by offering well-defined module interfaces, there can be scalability on the dimension of the processing aspect, which can be achieved either by optimizing the efficiency of processing or by expanding the functional modules involved in the processing aspect.

7.3 Hardware Requirements

- Processor: Intel i5 or higher
- RAM: Minimum 8 GB
- Storage: Minimum 10 GB free disk space
- Display: Standard monitor with minimum resolution 1366×768

7.4 Software Requirements

- Operating System: Windows / Linux
- Programming Language: Python
- Libraries and Tools:
    - Pandas, NumPy
    - Scikit-learn
    - TensorFlow (for LSTM model)
    - Tkinter
    - Matplotlib
- Development Environment: Python IDE (VS Code / PyCharm)

8. System Test

The Insider Threat Prediction System is completely tested at the end to make sure it works properly. This is the step to check the whole Insider Threat Prediction System. The goal of this test is to see if all parts of the Insider Threat Prediction System work well together and if the Insider Threat Prediction System does what it is supposed to do. This test checks the Insider Threat Prediction System from start, to finish from when you put in the data to when you see the results.

Objectives of System Testing

- To verify correct ingestion and processing of user activity logs
- To ensure accurate feature extraction and MITRE ATT&CK mapping
- To validate risk scoring, anomaly detection, and insider risk prediction
- To confirm proper generation and visualization of alerts and reports
- To ensure system stability under different input conditions

System Testing Strategy

The system is checked to see if it works properly by using a method where we only look at what we put in and what we get out. We do not look at how the system's built inside. We make test plans based on what the Software Requirement Specification says. We test the system with not normal situations to see if it is correct if it can handle problems and if it can deal with errors. The system and the Software Requirement Specification are used to guide our testing of the system.

We do test in a setting where we can control things. We use some sample information and real log data to make it seem like we are dealing with insider threat situations. This helps us see how things would work in the world with the insider threat, like a real person who is a threat. We use this method to test the insider threat and the sample datasets and the log data to make it real.

Types of System Testing Performed

8.1 Functional System Testing

Functional testing makes sure that every single function that is supposed to be, in the program works the way it is supposed to. This includes:

- Loading valid and invalid CSV log files
- Preprocessing and normalization of logs
- Feature extraction and aggregation per user
- Mapping events to MITRE ATT&CK
- Computing rule-based risk scores
- Detecting anomalies using Isolation Forest
- Predicting insider risk using the LSTM model
- Displaying results on the dashboard
- Exporting reports in CSV format

8.2 Integration Testing

Integration testing checks how different parts of a system work together. The system is tested to make sure that the integration testing works properly and that these parts of the system talk to each other correctly. It is done to ensure that the data gets to where it needs to go all the way from when we put it into the system to when we make it useful, for our project, which is the feature engineering part of the data. The data flows correctly from ingestion to feature engineering.

8.3 Performance Testing

The detection modules get the information, about the features. This means the detection modules receive feature data so they can do their job properly with the accurate feature data. The results from the detection modules are put together correctly. The detection modules all work and their results are combined in a proper way. This means that the results, from all the detection modules are consolidated properly. The visualization modules show the detection results in a way. Performance testing checks how the system works when it is really busy. The system is tested with a lot of log data to make sure that the system performs well. Performance testing of the system is very important to ensure that the system works properly when it has to deal with a lot of data.

8.4 Usability Testing

Usability testing is about how people feel when they use something. The system is checked to see how it works for the user experience.
This includes things, like:

- How easy the system is to use for the user experience
- If the system is nice to look at for the user experience
- The system is evaluated for the user experience to make sure it is good.
- Ease of navigation through the dashboard
- Clarity of alerts, charts, and tables
- Minimal training requirement for security analysts

9. Working

The Insider Threat Predictor tool is facilitated by a structured and interactive process flow with a graphical user interface (GUI) designed to replicate a Security Operation Center (SOC) environment. The GUI is primarily used as a platform for ingesting, analyzing, and visualizing data for carrying out a process flow of each phase involved in insider threat detection tasks.
The process begins at the Aggregate Controls stage, wherein either user logs are loaded through reading csv files or simulation logs are used. Logs contain information regarding user operations like log-in, file accesses, data transfers, and system calls. The logs are further aggregated on per-user, per-day bases to obtain behavioral summaries. Enriched features based on behavioral information help extract important usage features, which the system needs for further processing.

After completing the data preparation task, Prediction Controls enable event aggregation processing with rule-based logic. Security-related actions are assessed on the basis of predetermined weightage parameters to determine user risk scores for possible security events. The most recent events, identified security threats, and rapid risk overviews can be accessed from the GUI itself.

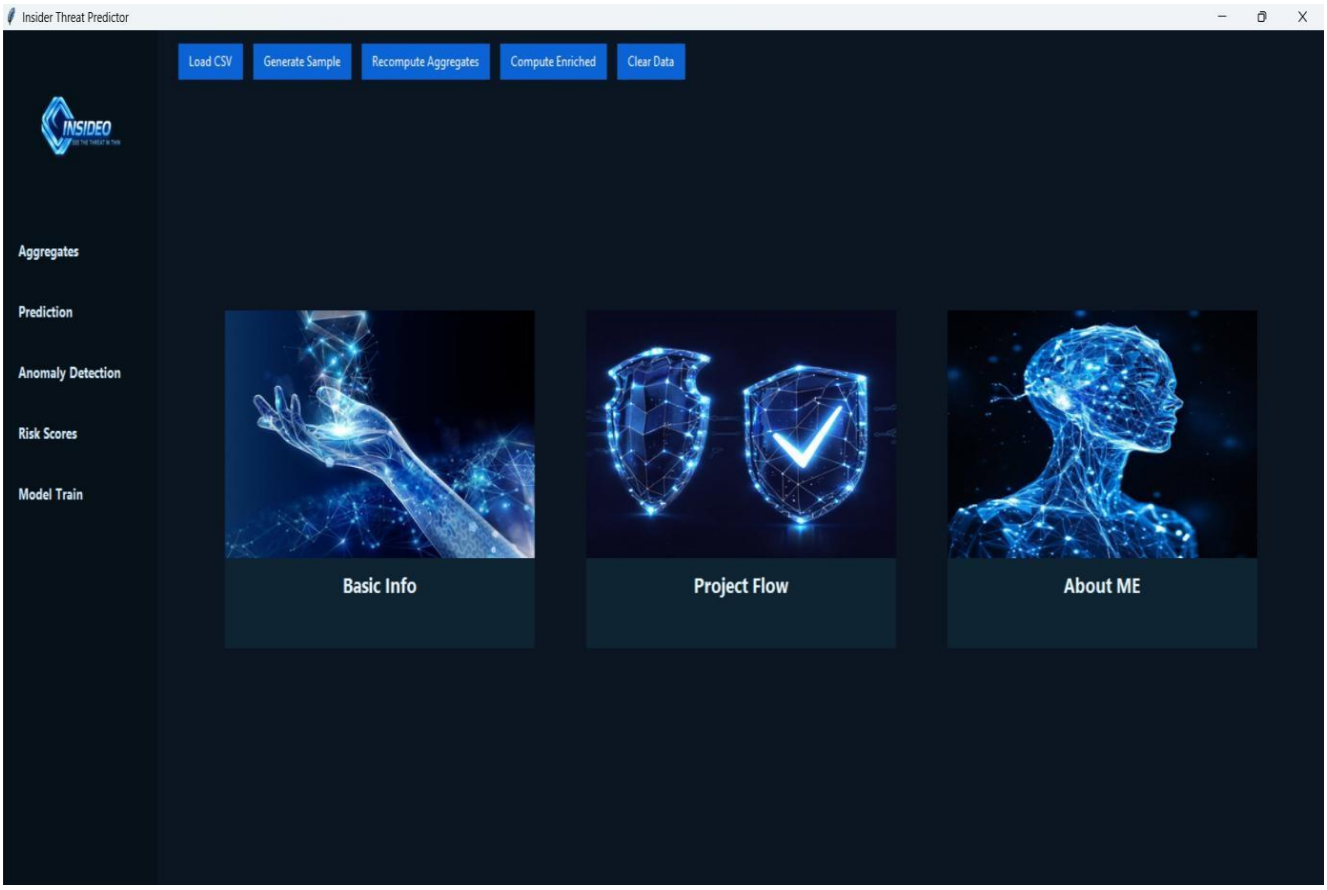
The controls of anomaly detection allow for the detection of abnormal behavior by using the Isolation Forest algorithm based on machine learning. This module analyzes aggregated behavioral features to identify those users whose activities are highly deviated from the established norms. The anomalies detected could be reviewed within the interface and were allowed to export as CSV reports for further investigation or documentation.

Risk Scores Controls also offer visualizations of user risk levels through various charts and tables, including a MITRE ATT&CK heat map. This tab enables the analyst to focus on high-risk users, view the top threats, and correlate suspicious activities against a set of standardized attack tactics and techniques. Risk tables and visual outputs can be exported for reporting and analysis purposes.

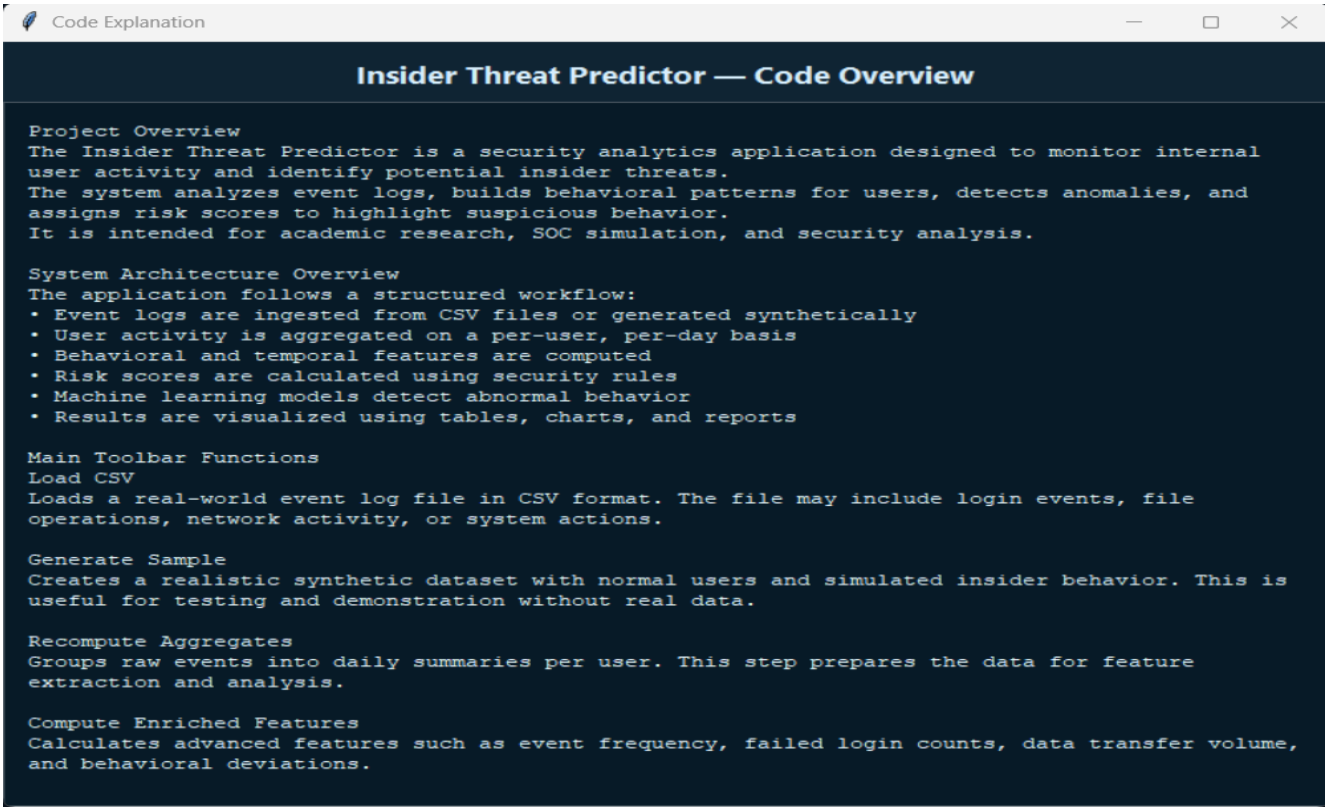
Advanced learning capabilities supported by Model Train Controls employ the LSTM-based model. The user defines the lookback window in which the model will be trained to learn the normal and riskier behavior patterns from historical behavioral sequences. This would then be used for prediction of insider risk levels, classification of users, and even generation of alerts on high-risk behavior that can be further reviewed through dedicated GUI views.

10. Graphical User Interface (GUI)

- **The .exe file of the python application.**



- **Main GUI of the application.**



- **Instruction Manual of the application.**

- **About me**

- **The HTML file which contains the Project Details.**

INSIDEO - Insider Threat Predictor

Definition

Insider Threat Predictor is a data-driven security system designed to identify and analyze potentially malicious or risky activities performed by internal users. It leverages log aggregation, behavioral feature engineering, anomaly detection, and risk scoring techniques to detect deviations from normal behavior. The system supports predictive analytics and visualization to proactively mitigate insider threats in organizational environments.

Core Features

1. **User Activity Aggregation**
   Collects and aggregates user activity logs such as login attempts, file access, and data transfers on a per-user, per-day basis. This aggregation helps establish individual behavioral baselines for accurate insider threat analysis.
2. **Behavioral Feature Engineering**
   Extracts meaningful statistical and temporal features including event frequency, resource diversity, failed login counts, and deviations from historical behavior to characterize user activity patterns.
3. **Anomaly Detection using Machine Learning**
   Applies unsupervised machine learning models such as Isolation Forest to identify abnormal user behavior that deviates significantly from established norms without requiring labeled attack data.

Suspicious Activity Risk Scoring

Suspicious Activity	Risk Score	Description
file_read	1	Normal data access
login_fail	2	Repeated authentication failures
file_write	3	File modification
file_download	5	Possible data staging
after_hours_login	5	Login outside business hours
login_from_unusual_ip	8	Credential misuse suspicion
file_delete	10	Data destruction risk
system_config_change	12	Defense evasion attempt
data_copy_to_usb	15	High-risk data exfiltration

MITRE ATT&CK Mapping

Event	Tactic	Technique
login_fail	Credential Access	Brute Force (T1110)
login_success	Credential Access	Valid Accounts (T1078)
file_read	Collection	Data from Local System (T1005)
file_download	Exfiltration	Exfiltration Over Network (T1041)
data_copy_to_usb	Exfiltration	Exfiltration to Removable Media
system_config_change	Defense Evasion	Modify System Configuration

How To Operate

1. **Launch the Application**
   Run `main.py`. The Insider Threat Predictor dashboard opens with the home screen and sidebar.

2. **Load or Generate Data**
   Load a real-world CSV event log or generate a synthetic sample dataset to simulate insider activity for testing and validation.

3. **Recompute Aggregates**
   Aggregate user activities on a per-user, per-day basis to build behavioral baselines.

4. **Compute Enriched Features**
   Temporal, behavioral, statistical, and graph-based features are calculated to enhance detection accuracy.

5. **Run Risk Scoring and Anomaly Detection**
   Rule-based risk scoring is applied, and the Isolation Forest model detects anomalous user behavior. Detected events are mapped to the MITRE ATT&CK framework and visualized using a heat map.

6. **Train LSTM Model (User-Defined Lookback)**
   Select the number of lookback days and train the LSTM model to learn historical behavior patterns.

7. **Insider Risk Prediction using LSTM**
   The trained LSTM model predicts insider risk probabilities, classifies users, and generates alerts for high-risk behavior.

8. **Results Visualization and Export**
   Risk scores, anomaly reports, and LSTM alerts are displayed through tables and charts. Reports can be exported as CSV files for further analysis or documentation.

9. **Clear Data and Reset System**
   Clear all loaded data, model outputs, and internal states to reset the system for a fresh analysis cycle.

Technologies & Libraries Used

1. **threading**
   Used to run time-consuming tasks such as event processing, Isolation Forest execution, and LSTM training in parallel, ensuring the Tkinter GUI remains responsive without freezing during heavy computations.

2. **traceback**
   Captures and displays detailed error stack traces during runtime exceptions, allowing graceful debugging without crashing the application.

Details	Information
Company Name	Supraja Technologies
Location	Hyderabad, India
Contact	contact@suprajatechnologies.com
Linkedin	www.linkedin.com/company/suprajatechnologies/

Developer Details

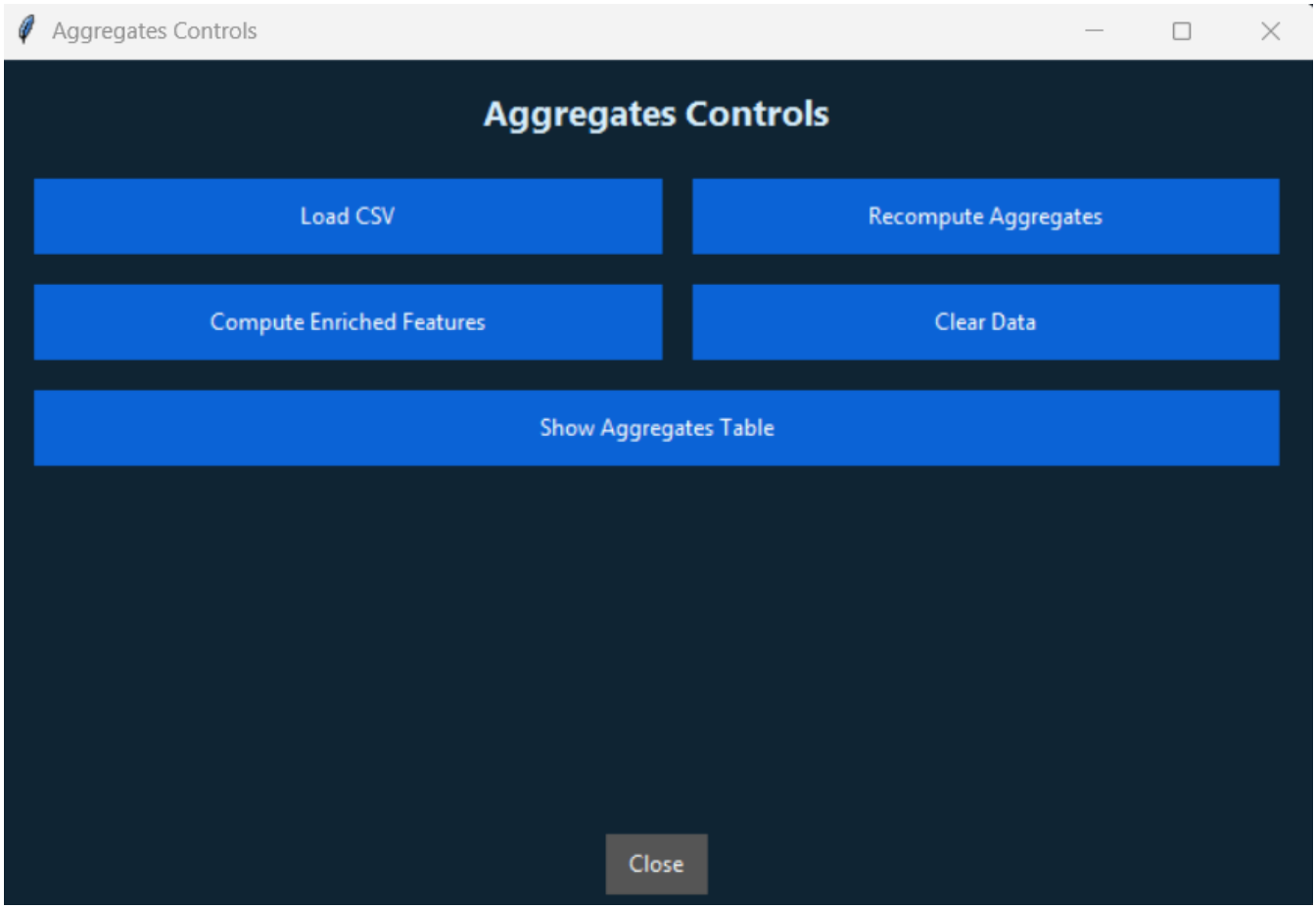
Name: Shaikh Aiman

Employee ID: AD23#ST#IS#7003

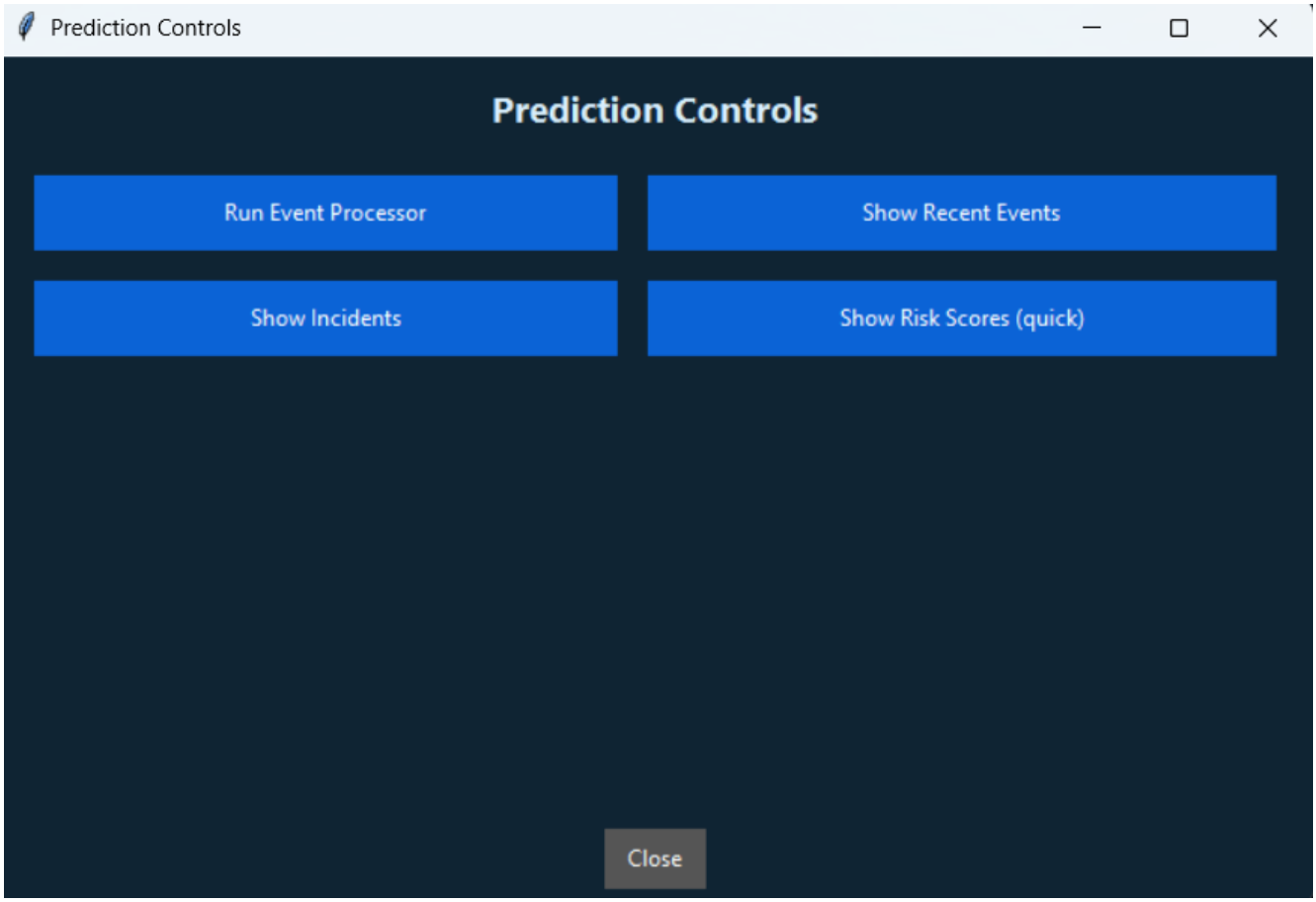
Email: shaikh.aiman2712@gmail.com

Github: https://github.com/Shaikh-Aiman

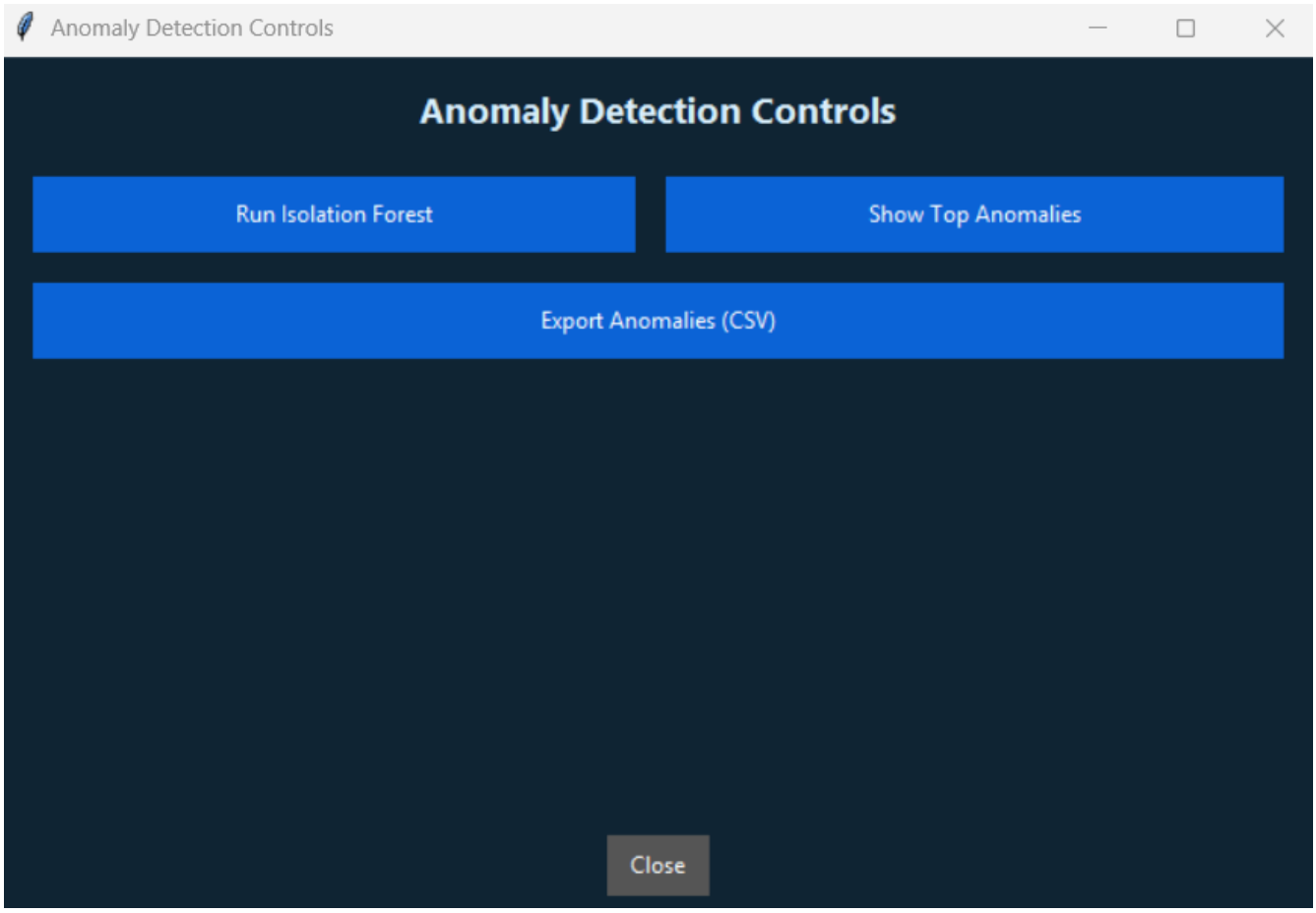
Linkedin: www.linkedin.com/in/shaikh-aiman



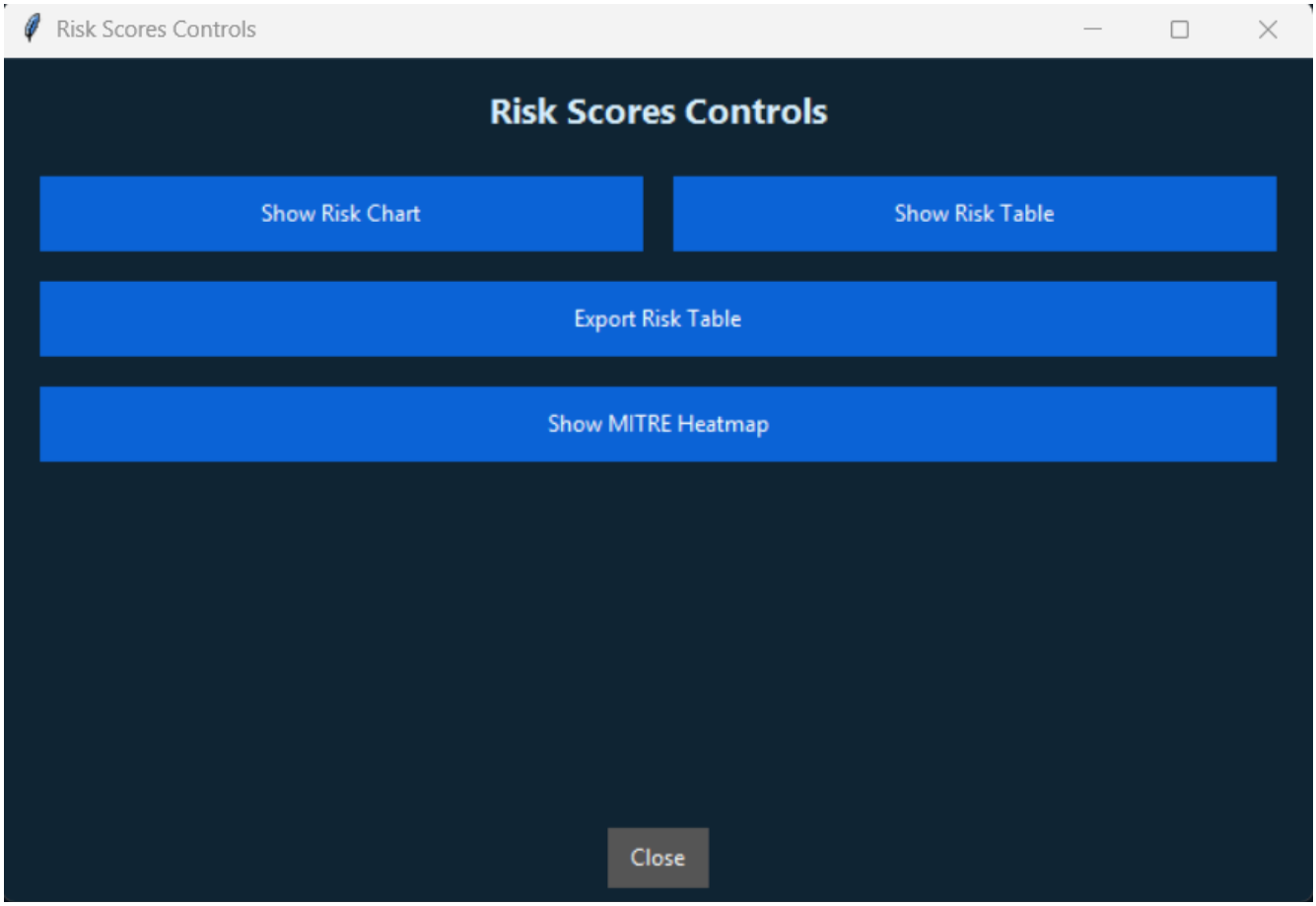
- **Loads user activity logs and recomputes per-user, per-day aggregates to create structured behavioural summaries.**
- **Computes enriched features and displays aggregate tables while allowing data reset for a fresh analysis cycle.**



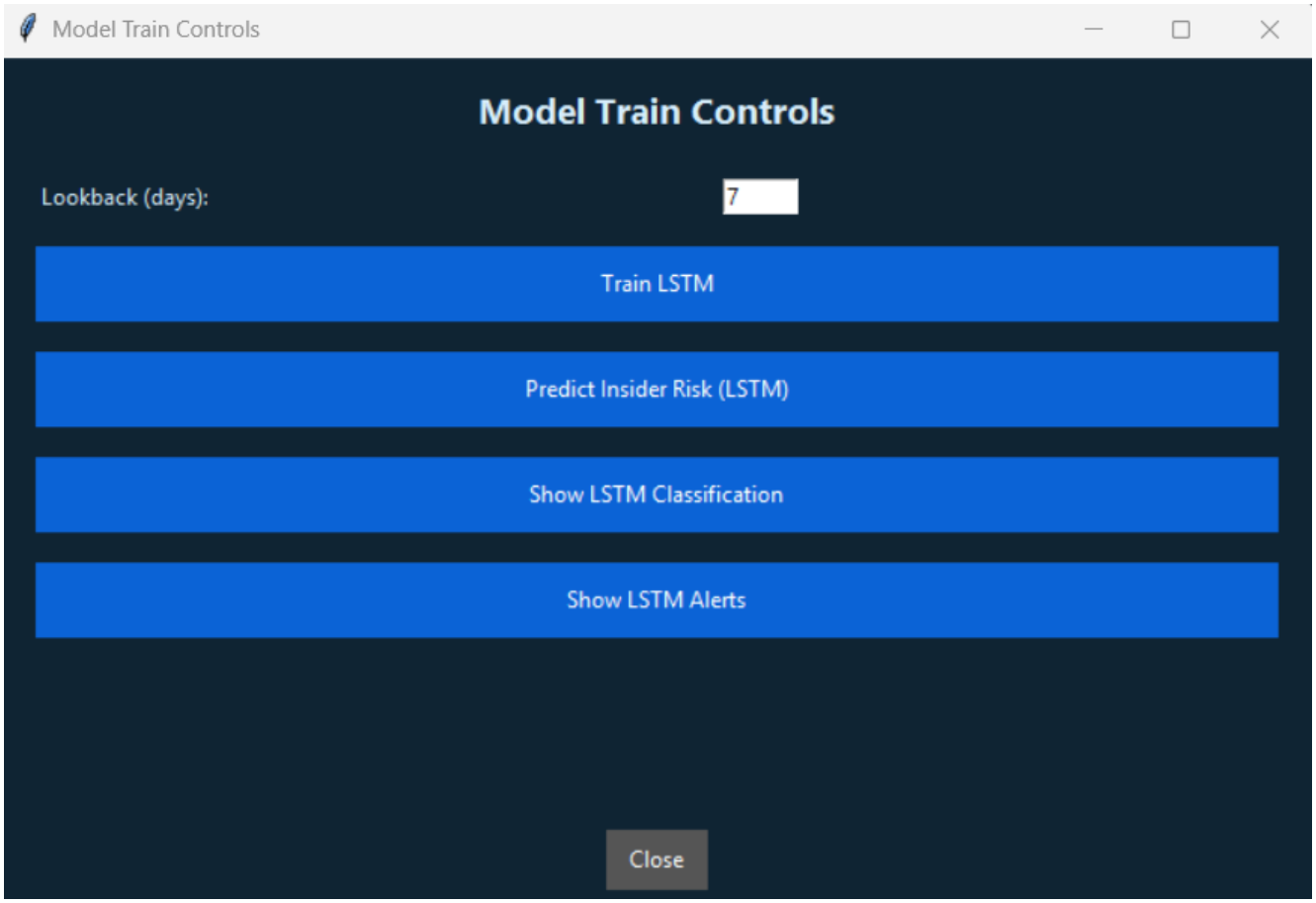
- **Processes user events using rule-based logic to calculate risk scores and identify potential security incidents.**
- **Displays recent events, detected incidents, and quick risk summaries for immediate review.**



- **Executes the Isolation Forest model to identify users whose behaviour deviates from established patterns.**
- **Displays top anomalies and allows exporting anomaly results as CSV for further investigation.**



- **Visualizes user risk levels through charts and tables to support prioritization and analysis.**
- **Provides export options and MITRE ATT&CK heatmap visualization for threat context mapping.**



- **Trains the LSTM model using a user-defined lookback window to learn behavioural patterns over time.**
- **Predicts insider risk, shows classification results, and generates alerts for high-risk users.**

11. Conclusion

This project has successfully shown how one could design and implement an Insider Threat Predictor that is capable of handling the increasing problem of malicious/risky activities being carried out by insiders. Through log aggregation, behavior feature extraction, rule-based risk assessment, and machine learning methods, one has a systematic and effective way of analyzing behavior and pinpointing insider threats. The addition of the use of the MITRE ATT&CK framework adds credence to this process by providing normalized threat knowledge, which allows for better interpretation of identified risks.

The entire system is facilitated by an interactive SOC-style GUI that enables users to perform seamlessly all the following tasks in one setting: data ingestion, aggregation, anomaly detection, risk scoring, model training, and everything else in between. The entire workflow is conducted using the GUI, which enables easier understanding of all the complicated operations necessary in the world of analytics. The use of tables, charts, heat maps in the GUI enables, among other things, risk prioritization.

Machine learning methods are important in improving the accuracy of the system's detections. The Isolation Forest algorithm is useful in pointing out abnormal user activity without the need to use labeled information, whereas the use of the LSTM module is useful in allowing the system to examine past activity to determine the risk levels.

Further, this project provides a great learning opportunity as theoretical ideas related to cybersecurity are implemented as a real-world-inspired system. The project illustrates experience in log analysis, behavior modeling, anomaly identification, or risk evaluation along with a focus on explainability and usefulness. The Insider Threat Predictor system is designed to be extendable to incorporate future ideas like real-time log functionality, more advanced predictive systems, or automated systems for future learning opportunities related to enterprise security analytics or SOC systems.