A PROJECT MID–TERM PRESENTATION
ON

“USER BHAVIOR ANALYTICS FOR INSIDER THREAT USING TRANSFORMER BASED APPROACH”

Presented By:

Aarati Kumari Mahato [079MSISE01]
Department of Electronics and Computer Engineering
Thapathali Campus
Institute of Engineering

Supervised by:

Er. Roshan Pokharel

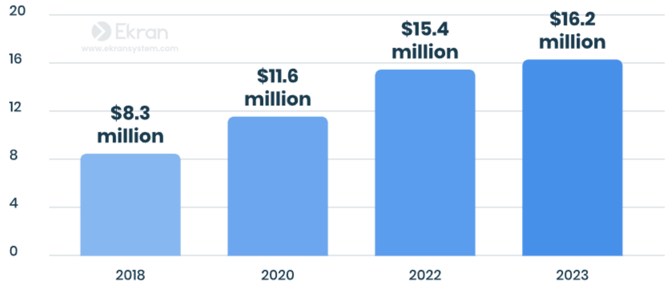
OUTLINE

- Motivation
- Real life data breaches caused by insider threat
- Background
- Problem Statement
- Modern Approach for insider threat
- Objectives of Project
- Scope of project
- Originality of project
- Potential Applications
- Literature Review
- Methodology
- Results
- Discussion and Analysis
- Future Enhancement
- Conclusion
- Project Schedule
- References

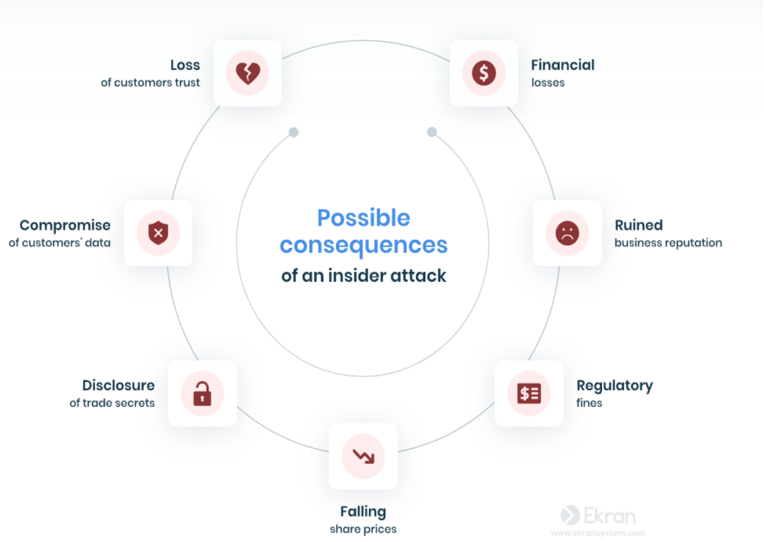
MOTIVATION

- As organizations critical assets have been digitized and access to information has increased, the nature and severity of threats have changed
- Insiders who work for an organization have more power than ever to abuse their access to crucial organizational resources

Total average cost of insider threat incidents



2023 Cost of Insider Risks Global Report by Ponemon Institute


Possible consequences of an insider attack
Loss of customers trust
Financial losses
Compromise of customers' data
Ruined business reputation
Disclosure of trade secrets
Regulatory fines
Falling share prices
Ekran
www.ekransystem.com

REAL LIFE DATA BREACHES CAUSED BY INSIDER THREAT

1. Intellectual property theft by a malicious insider at Yahoo

- Yahoo’s research scientist Qian Sang, who worked as a research scientist at Yahoo, stole the company’s intellectual property in Feb 2022 to use the stolen data for financial gain from Yahoo’s competitor, The Trade Desk.

- Prior to the incident, Sang had received a job offer from them.

- **Consequences**: downloaded 570,000 files containing sensitive information and the source code of AdLearn, Yahoo’s engine for real-time ad purchasing

- **Why did it happen:** Sang allegedly transferred the sensitive data from his corporate laptop to two personal external storage devices while he was still working at Yahoo.

REAL LIFE DATA BREACHES CAUSED BY INSIDER THREAT

2. Data theft by a former SGMC employee

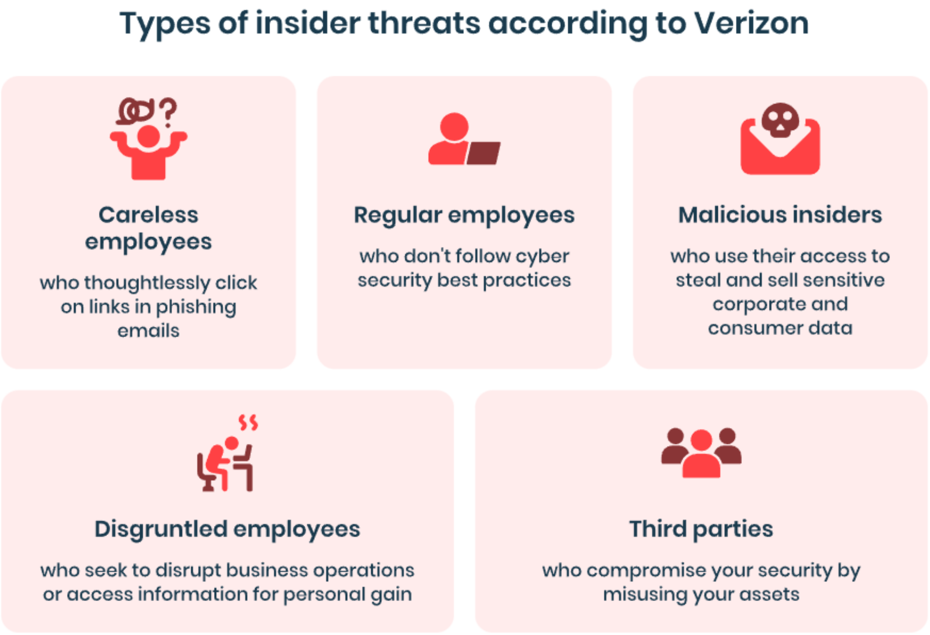
- downloaded private data from the medical center’s systems to his USB drive without obvious reason the day after quitting

- Patient test results, names, and birth dates were leaked.

- A former employee had legitimate access to the data he stole and had nothing preventing him from carrying through with his intentions

BACKGROUND

Insider Threats: Significant security concern where employees, contractors, or partners misuse their access to harm the organization



Types of insider threats according to Verizon

Careless employees
who thoughtlessly click on links in phishing emails

Regular employees
who don't follow cyber security best practices

Malicious insiders
who use their access to steal and sell sensitive corporate and consumer data

Disgruntled employees
who seek to disrupt business operations or access information for personal gain

Third parties
who compromise your security by misusing your assets

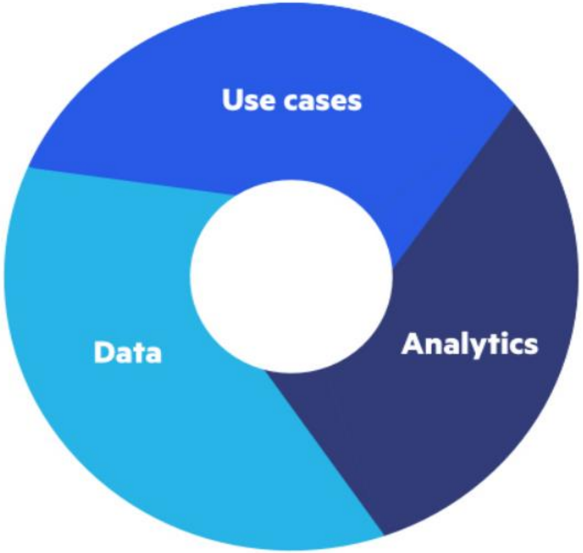
PROBLEM DEFINITION

- active attacks e.g. a sudden brute force attack, can be detected by modern firewalls, antivirus software, intrusion detection systems etc
- many data and security breaches has been done by the users within the organizations.
- traditional detection methods often miss subtle, sophisticated insider activities.
- to detect activities like this, we need to monitor behavior of users over a period of time

MODERN APPROACH FOR INSIDER THREAT

***UBA (User Behavior Analytics)*:**

- Monitors and analyzes user behavior to detect anomalies and potential threats.



Three pillars of UEBA

Use cases

- Malicious insider
- Compromised user
- APT and zero-day
- Known threats

Analytics

- Supervised machine learning
- Unsupervised machine learning
- Statistical modeling
- Rule-based system

Future:

- Generative adversarial networks
- Ensemble networks
- Deep learning

Data

- Events and logs
- Network flows and packets
- Business context
- HR and user context
- External threat intelligence

OBJECTIVES OF PROJECT

- To filter the raw log events for each user and generate natural language event using large language model(Llama3.1)
- To classify such events using SecureBERT model.

SCOPE OF PROJECT

- This model can be useful for the organizations to monitor their employees activities and detect anomaly behavior
- A modern approach for insider threat anomaly detection

ORIGINALITY OF PROJECT

- Transform the raw log events of CERT4.2 dataset into Natural Language context
- Fintune the Llama3.1 model for natural language contextual data generation
- Implement secureBERT model for the classification between normal and malicious instances

POTENTIAL APPLICATIONS

- Enterprise Security
- Healthcare
- Financial Services
- Government Agencies
- Educational Institutions
- Retail and E-commerce

LITERATURE REVIEW[1]

Paper	Year	Authors	Methodology	Results	Strengths	Weakness
User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder Insider Threat Detection	2020	Sharma, Balaram, et al.	-LSTM -RNN models -LSTM Autoencoder	Accuracy 90.17%, TPR 91.03%, FPR 9.84%	Better accuracy compared to traditional models	Missed out some of the features
MalBERTv2: Code Aware BERT-Based Model for Malware Identification	2023	Abir Rahali and Moulay A. Akhloufi	-BERT -Malware(MG) dataset	F1 score ranging from 82% to 99%	apply a classifier using bidirectional encoder representations from transformers (BERT) as a layer within the model pipeline	--lack of benchmarks for malware/goodware identification --Couldn’t effectively compare the model with existing methods

LITERATURE REVIEW[2]

Paper	Year	Authors	Methodology	Results	Strengths	Weakness
Lm-Hunter: An Nlp-Powered Graph Method for Detecting Adversary Lateral Movements in Apt Cyber-Attacks at Scale	2023	P´erez-Gomariz, Mario , et al.	-NLP -Transformer	Accuracy 65% to 85%	-provides a holistic view of the user's lateral movements in the network -use of graphs	-outputs of the models showed minimal fluctuation
Devising and Detecting Phishing Emails Using Large Language Models	2024	AHEIDING, FREDRIK, et al.	-used ChatGPT for generating emails	F1 score ranging from 82% to 99%	-good recommendations for reacting to phishing emails	-gateway to subsequent research rather than a final destination

Why large language model?

- **Text analytic models require machine readable input format**
  - Traditional models works based on occurrence frequency
  - One-hot encoding
- **Traditional models failed to adequately capture important text features**
  - **Semantic** relationships
  - **Context** understanding

METHODOLOGY[1]

Training Phase

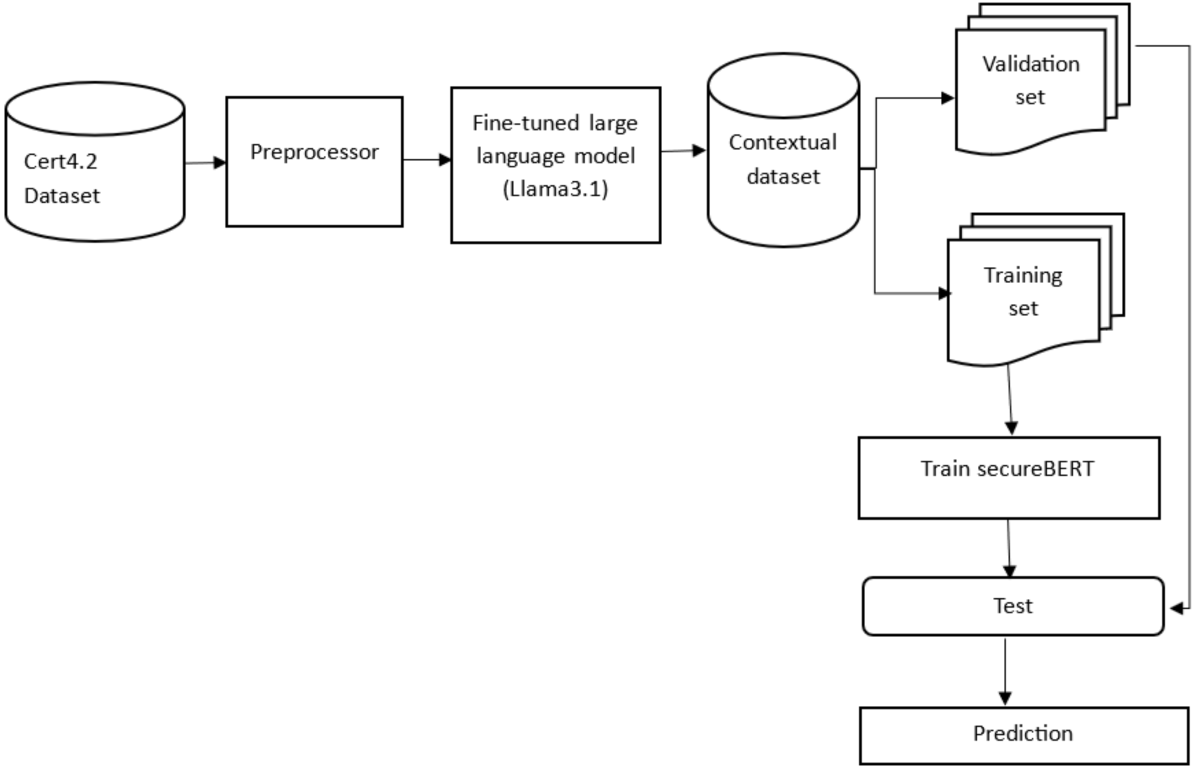
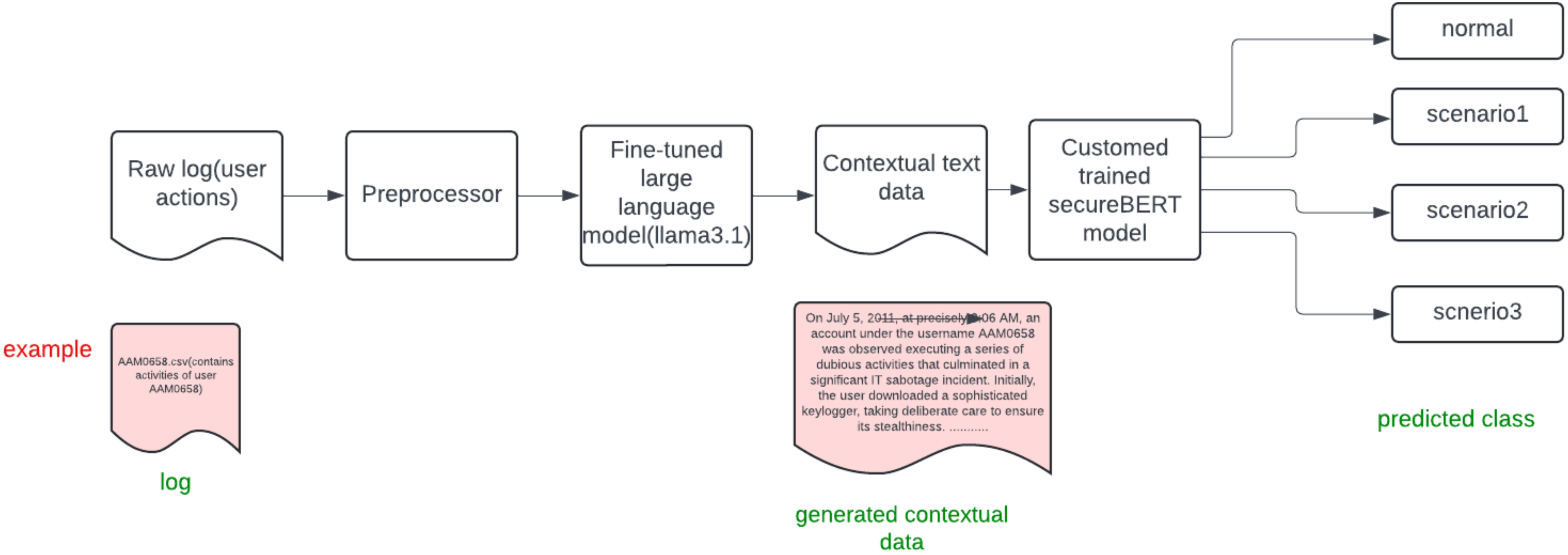


Fig: Block diagram of training phase of system

METHODOLOGY[2]


Raw log(user actions)
Preprocessor
Fine-tuned large language model(llama3.1)
Contextual text data
Customed trained secureBERT model
normal
scenario1
scenario2
scnerio3
example
AAM0658.csv(contains activities of user AAM0658)
log

On July 5, 2011, at precisely [illegible]:06 AM, an account under the username AAM0658 was observed executing a series of dubious activities that culminated in a significant IT sabotage incident. Initially, the user downloaded a sophisticated keylogger, taking deliberate care to ensure its stealthiness. ....

generated contextual data
predicted class

Fig: Block diagram of testing phase of system

DATASET [1]

- CERT 4.2, widely used in academia and industry for developing and testing insider threat detection algorithms.
- synthetic data representing normal user behavior and malicious insider activities
- developed by the CERT division of Carnegie Mellon University
- Stimulates behavior logs of 1000 users over 502 days
- Includes logon/logoff events, email logs, file accesses, HTTP requests, and device connect/disconnect events
- Features three specific insider threat scenarios, each with different motivations and methods.
- Modify the dataset by contextual data generated by the large language model

DATASET [2]

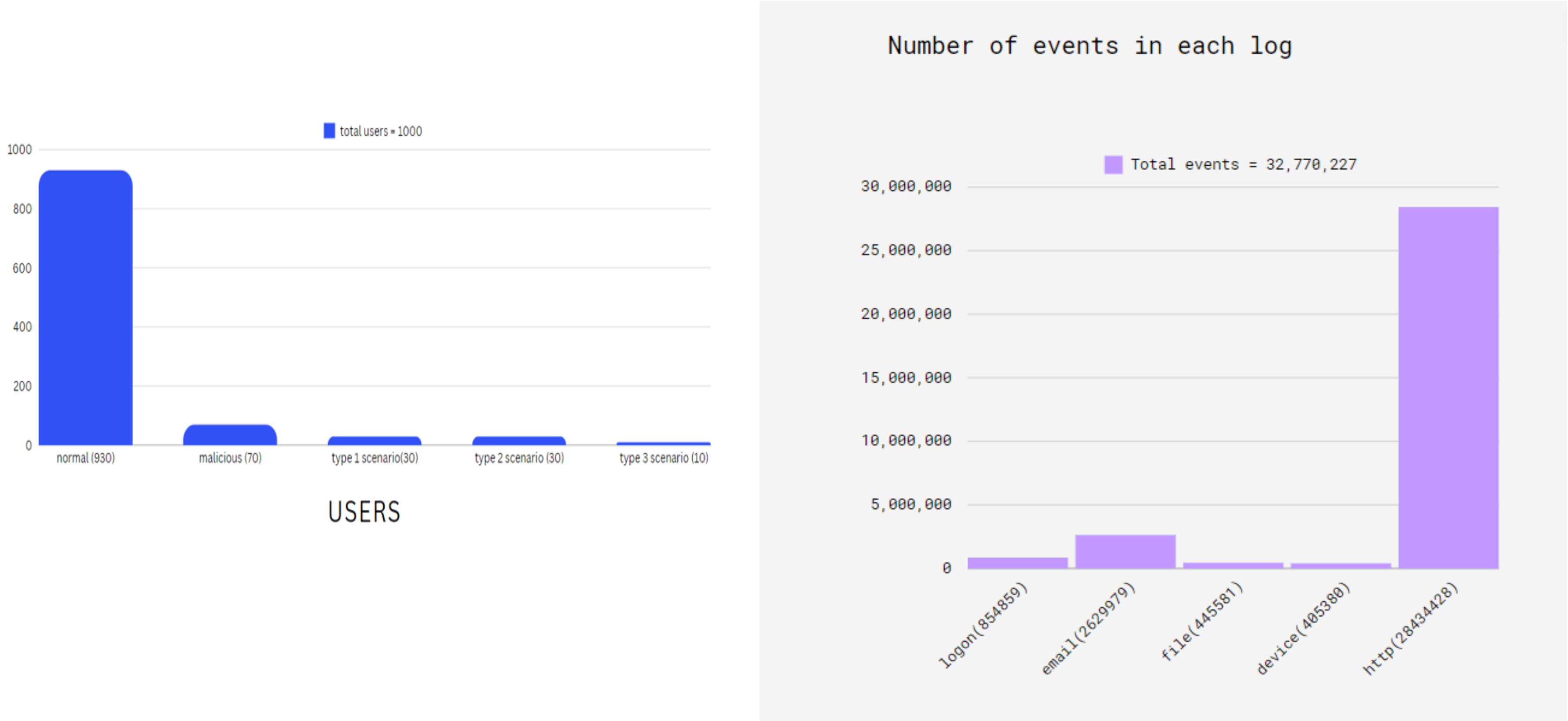
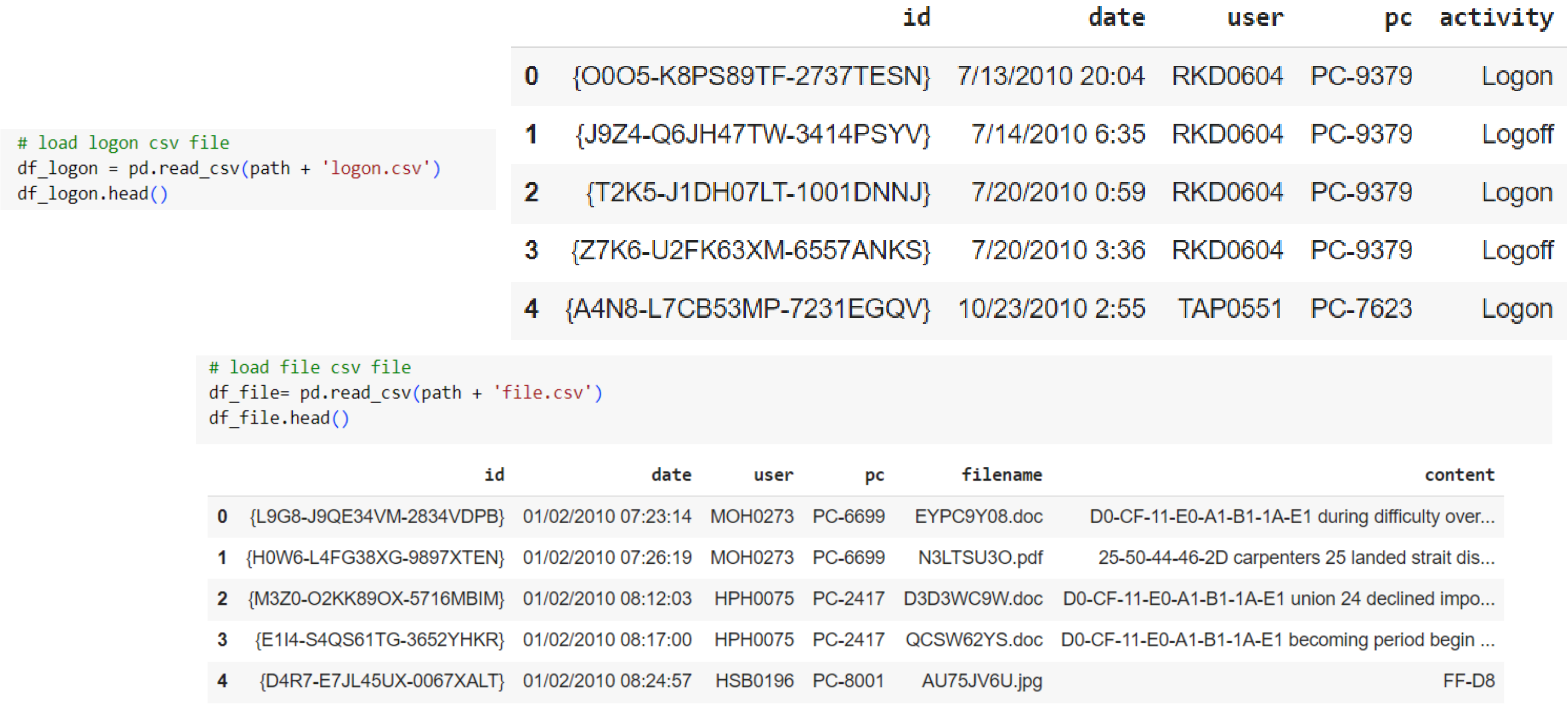


Fig: Illustration of CERT dataset

DATASET [3]



```python
# load logon csv file
df_logon = pd.read_csv(path + 'logon.csv')
df_logon.head()
```

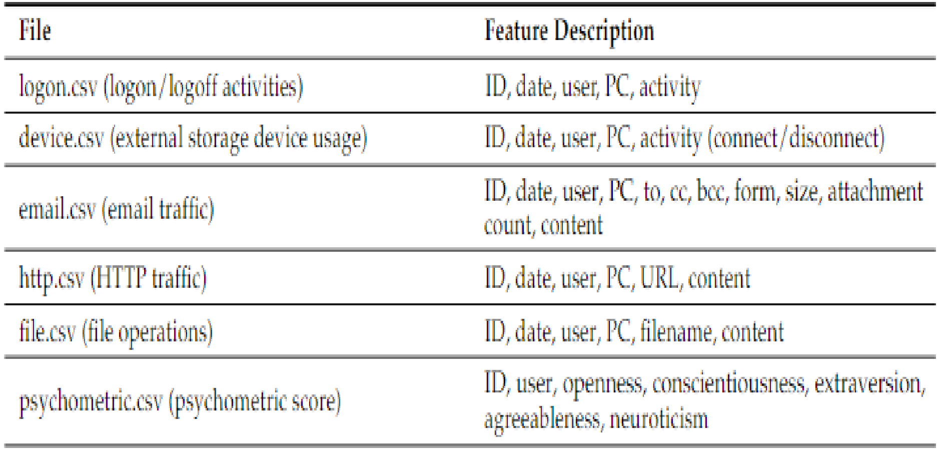
	id	date	user	pc	activity
0	{O0O5-K8PS89TF-2737TESN}	7/13/2010 20:04	RKD0604	PC-9379	Logon
1	{J9Z4-Q6JH47TW-3414PSYV}	7/14/2010 6:35	RKD0604	PC-9379	Logoff
2	{T2K5-J1DH07LT-1001DNNJ}	7/20/2010 0:59	RKD0604	PC-9379	Logon
3	{Z7K6-U2FK63XM-6557ANKS}	7/20/2010 3:36	RKD0604	PC-9379	Logoff
4	{A4N8-L7CB53MP-7231EGQV}	10/23/2010 2:55	TAP0551	PC-7623	Logon

```python
# load file csv file
df_file= pd.read_csv(path + 'file.csv')
df_file.head()
```

	id	date	user	pc	filename	content
0	{L9G8-J9QE34VM-2834VDPB}	01/02/2010 07:23:14	MOH0273	PC-6699	EYPC9Y08.doc	D0-CF-11-E0-A1-B1-1A-E1 during difficulty over...
1	{H0W6-L4FG38XG-9897XTEN}	01/02/2010 07:26:19	MOH0273	PC-6699	N3LTSU3O.pdf	25-50-44-46-2D carpenters 25 landed strait dis...
2	{M3Z0-O2KK89OX-5716MBIM}	01/02/2010 08:12:03	HPH0075	PC-2417	D3D3WC9W.doc	D0-CF-11-E0-A1-B1-1A-E1 union 24 declined impo...
3	{E1I4-S4QS61TG-3652YHKR}	01/02/2010 08:17:00	HPH0075	PC-2417	QCSW62YS.doc	D0-CF-11-E0-A1-B1-1A-E1 becoming period begin ...
4	{D4R7-E7JL45UX-0067XALT}	01/02/2010 08:24:57	HSB0196	PC-8001	AU75JV6U.jpg	FF-D8

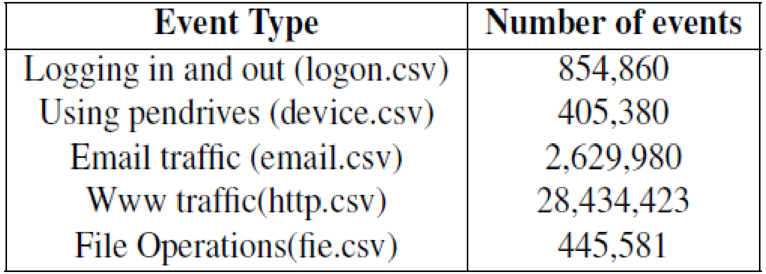
Fig: Illustration of different events and attributes of the dataset

DATASET [4]



File	Feature Description
logon.csv (logon/logoff activities)	ID, date, user, PC, activity
device.csv (external storage device usage)	ID, date, user, PC, activity (connect/disconnect)
email.csv (email traffic)	ID, date, user, PC, to, cc, bcc, form, size, attachment count, content
http.csv (HTTP traffic)	ID, date, user, PC, URL, content
file.csv (file operations)	ID, date, user, PC, filename, content
psychometric.csv (psychometric score)	ID, user, openness, conscientiousness, extraversion, agreeableness, neuroticism

Events and attributes



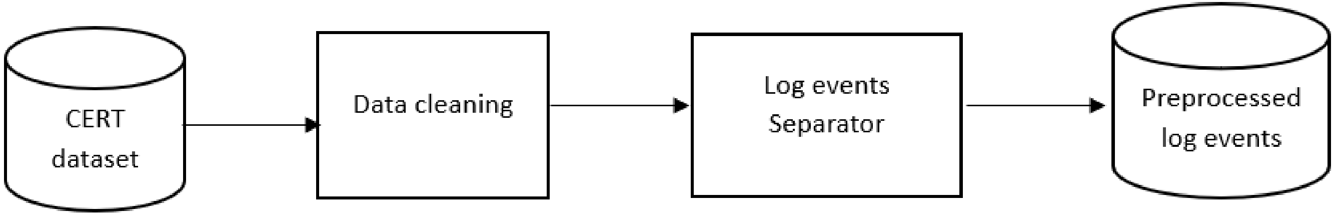
Event Type	Number of events
Logging in and out (logon.csv)	854,860
Using pendrives (device.csv)	405,380
Email traffic (email.csv)	2,629,980
Www traffic(http.csv)	28,434,423
File Operations(fie.csv)	445,581

Statistics of the dataset

Table: Illustration of different events and attributes of the CERT dataset

METHODOLOGY[3]

Data Preprocessing



Contains combined log events of 501 days

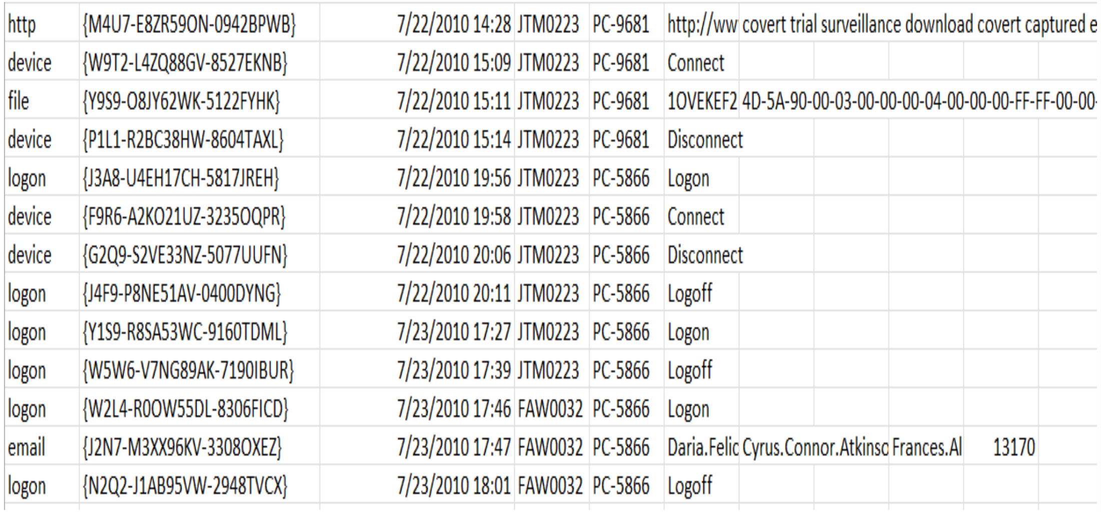
Contains the log events of each user based on per day

Fig: Block diagram of data preprocessing

METHODOLOGY[4]

Data Preprocessing

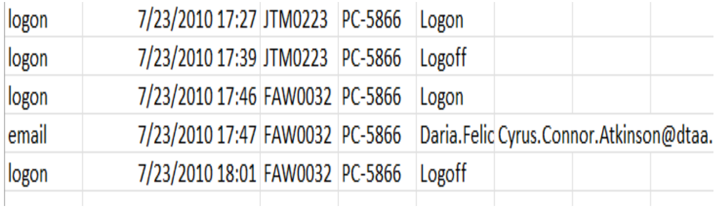
Before



http	{M4U7-E8ZR59ON-0942BPWB}	7/22/2010 14:28	JTM0223	PC-9681	http://ww	covert trial surveillance download covert captured e		
device	{W9T2-L4ZQ88GV-8527EKNB}	7/22/2010 15:09	JTM0223	PC-9681	Connect			
file	{Y9S9-O8JY62WK-5122FYHK}	7/22/2010 15:11	JTM0223	PC-9681	1OVEKEF2	4D-5A-90-00-03-00-00-00-04-00-00-00-FF-FF-00-00-		
device	{P1L1-R2BC38HW-8604TAXL}	7/22/2010 15:14	JTM0223	PC-9681	Disconnect			
logon	{J3A8-U4EH17CH-5817JREH}	7/22/2010 19:56	JTM0223	PC-5866	Logon			
device	{F9R6-A2KO21UZ-3235OQPR}	7/22/2010 19:58	JTM0223	PC-5866	Connect			
device	{G2Q9-S2VE33NZ-5077UUFN}	7/22/2010 20:06	JTM0223	PC-5866	Disconnect			
logon	{J4F9-P8NE51AV-0400DYNG}	7/22/2010 20:11	JTM0223	PC-5866	Logoff			
logon	{Y1S9-R8SA53WC-9160TDML}	7/23/2010 17:27	JTM0223	PC-5866	Logon			
logon	{W5W6-V7NG89AK-7190IBUR}	7/23/2010 17:39	JTM0223	PC-5866	Logoff			
logon	{W2L4-ROOW55DL-8306FICD}	7/23/2010 17:46	FAW0032	PC-5866	Logon			
email	{J2N7-M3XX96KV-3308OXEZ}	7/23/2010 17:47	FAW0032	PC-5866	Daria.Felic	Cyrus.Connor.Atkinsc	Frances.Al	13170
logon	{N2Q2-J1AB95VW-2948TVCX}	7/23/2010 18:01	FAW0032	PC-5866	Logoff			

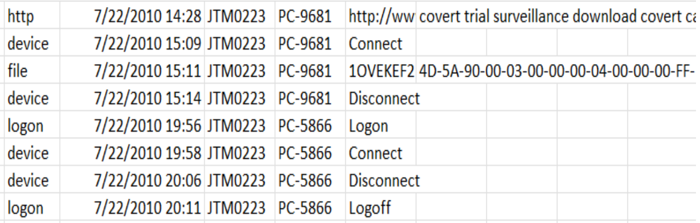
Contains activities of user of 2 days (7/22 and 7/23)

After
(Separated into 2 files each containing activities of a day)



logon	7/23/2010 17:27	JTM0223	PC-5866	Logon	
logon	7/23/2010 17:39	JTM0223	PC-5866	Logoff	
logon	7/23/2010 17:46	FAW0032	PC-5866	Logon	
email	7/23/2010 17:47	FAW0032	PC-5866	Daria.Felic	Cyrus.Connor.Atkinson@dtaa.
logon	7/23/2010 18:01	FAW0032	PC-5866	Logoff	

JTM0223_date_7/23/2010.csv



http	7/22/2010 14:28	JTM0223	PC-9681	http://ww	covert trial surveillance download covert ca
device	7/22/2010 15:09	JTM0223	PC-9681	Connect	
file	7/22/2010 15:11	JTM0223	PC-9681	1OVEKEF2	4D-5A-90-00-03-00-00-00-04-00-00-00-FF-
device	7/22/2010 15:14	JTM0223	PC-9681	Disconnect	
logon	7/22/2010 19:56	JTM0223	PC-5866	Logon	
device	7/22/2010 19:58	JTM0223	PC-5866	Connect	
device	7/22/2010 20:06	JTM0223	PC-5866	Disconnect	
logon	7/22/2010 20:11	JTM0223	PC-5866	Logoff	

JTM0223_date_7/22/2010.csv

Fig: Example of data preprocessing

DATASET [5]

Three Scenarios includes:

1. User who did not previously use removable drives or work afterhours begins logging in after hours, using a removable drive, and uploading data to wikileaks.org. Leaves the organization shortly thereafter.
2. User begins surfing job websites and soliciting employment from a competitor. Before leaving the company, they use a thumb drive (at markedly higher rates than their previous activity) to steal data.
3. System administrator becomes disgruntled. Downloads a keylogger and uses a thumb drive to transfer it to his supervisor's machine. The next day, he uses the collected keylogs to log in as his supervisor and send out an alarming mass email, causing panic in the organization. He leaves the organization immediately.

METHODOLOGY[5]

Fine-Tuning Llama3.1

- Model Preparation: Loading the 4-Bit Quantized Model

$$w_{quant} = \text{round}\left(\frac{w_{orig}}{\Delta}\right)$$

- Efficiency Optimization: Flash Attention with Xformers

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

- Parameter efficient Tuning with LoRA

$$W_{new} = W_{orig} - \eta \nabla L$$

Standard weight update in neural networks

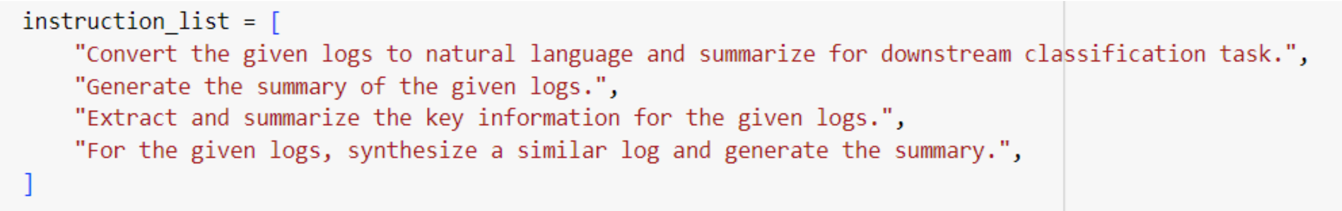
$$W_{new} = W_{orig} + A \cdot B^T$$

In LoRA, update is applied to low-rank matrices A and B

METHODOLOGY[6]

Dataset preparation for Fine-Tuning Llama3.1

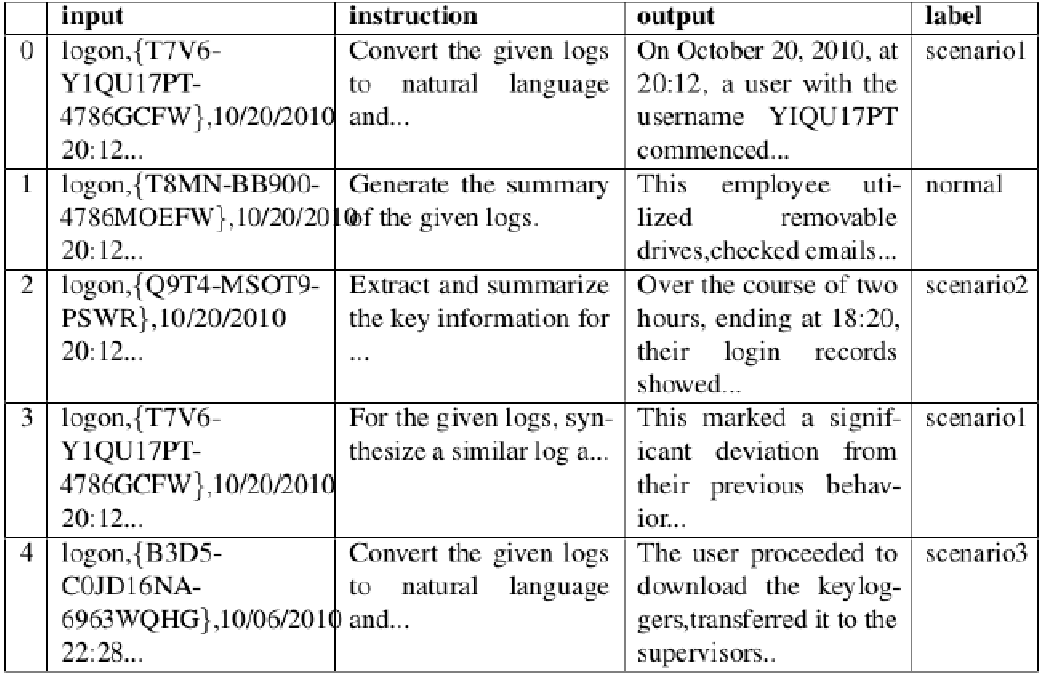
- synthesized data was generated from Openai’s API by using endpoint ”gpt-4-mini” model
- Data consist of attributes (input, output, label)
- Input refers to the log containing activities of the users
- Output refers to the contextual data in natural language
- Label refers to the type of events (normal, scenario 1, scenario 2, scenario 3)



```
instruction_list = [
    "Convert the given logs to natural language and summarize for downstream classification task.",
    "Generate the summary of the given logs.",
    "Extract and summarize the key information for the given logs.",
    "For the given logs, synthesize a similar log and generate the summary.",
]
```

METHODOLOGY[7]

Example of Data for Fine-Tuning Llama3.1



	input	**instruction**	**output**	**label**
0	logon,{T7V6-Y1QU17PT-4786GCFW},10/20/2010 20:12...	Convert the given logs to natural language and...	On October 20, 2010, at 20:12, a user with the username YIQU17PT commenced...	scenario1
1	logon,{T8MN-BB900-4786MOEFW},10/20/2010 20:12...	Generate the summary of the given logs.	This employee utilized removable drives,checked emails...	normal
2	logon,{Q9T4-MSOT9-PSWR},10/20/2010 20:12...	Extract and summarize the key information for ...	Over the course of two hours, ending at 18:20, their login records showed...	scenario2
3	logon,{T7V6-Y1QU17PT-4786GCFW},10/20/2010 20:12...	For the given logs, synthesize a similar log a...	This marked a significant deviation from their previous behavior...	scenario1
4	logon,{B3D5-C0JD16NA-6963WQHG},10/06/2010 22:28...	Convert the given logs to natural language and...	The user proceeded to download the keyloggers,transferred it to the supervisors..	scenario3

METHODOLOGY[7]

Process of Fine-Tuning Llama3.1

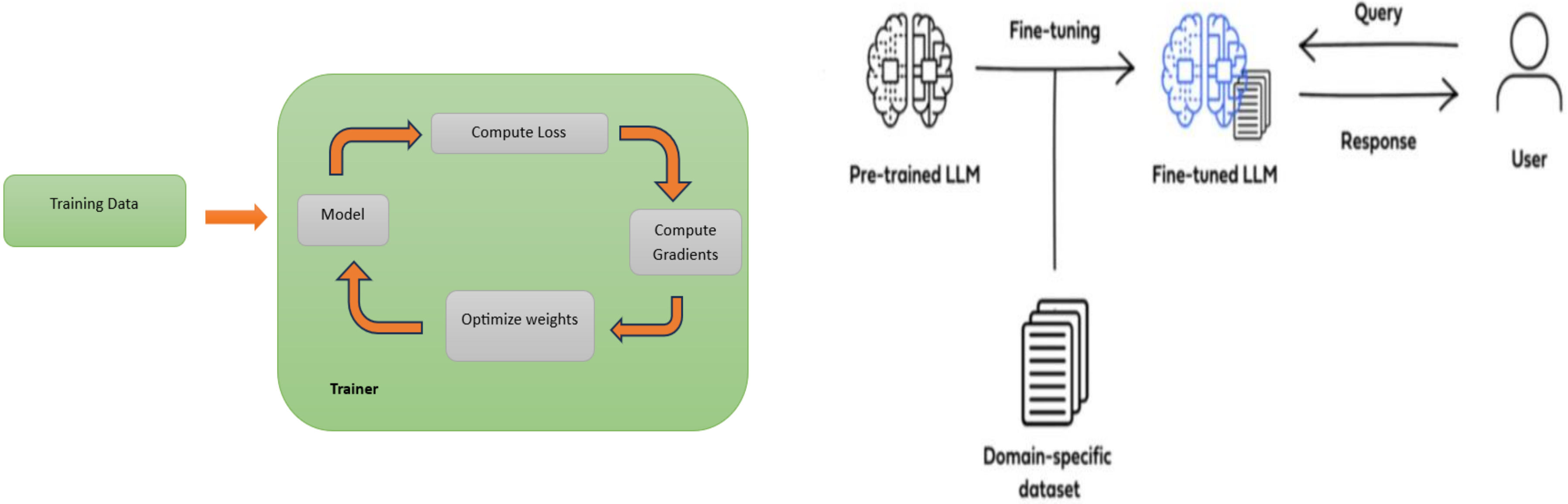


Figure: The diagram showing fine tuning process of Llama3.1 model

METHODOLOGY[6]

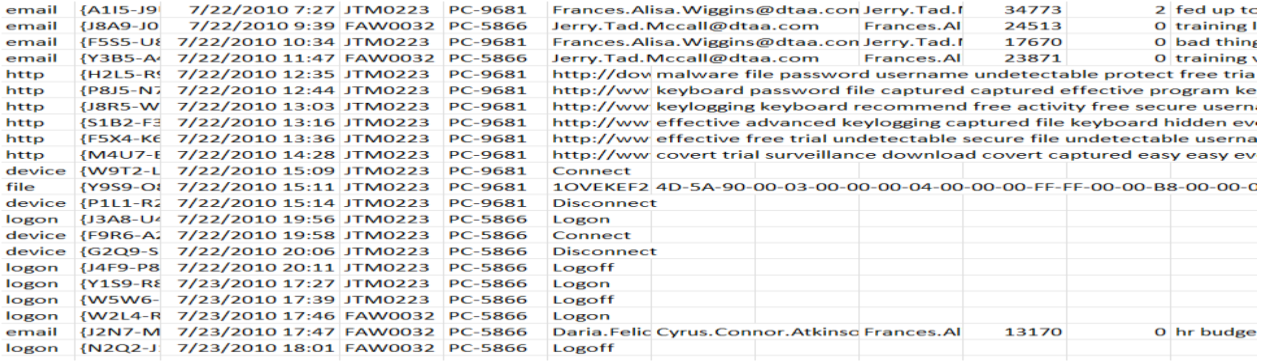
Example of log (scenario 1)

```
logon      {O0O5-K8PS89TF-2737TESN}    7/13/2010 20:04  RKD0604    PC-9379    Logon
device     {U2A2-K5UV96DW-1786RJHQ}    7/13/2010 20:59  RKD0604    PC-9379    Connect
http {A9S6-H6RM85XL-1447KFQN}          7/13/2010 21:05  RKD0604    PC-9379
           http://wikileaks.org/Julian_Assange/assange/The_Real_Story_About_DTAA/Gur_Erny_Fgbel_Nobhg_QGNN1528513805.php
device     {T8X1-F3BF45QK-7019GMQY}    7/13/2010 21:08  RKD0604    PC-9379    Disconnect
logon      {J9Z4-Q6JH47TW-3414PSYV}    7/14/2010 6:35   RKD0604    PC-9379    Logoff
logon      {T2K5-J1DH07LT-1001DNNJ}    7/20/2010 0:59   RKD0604    PC-9379    Logon
device     {Y5P9-T9XZ63DH-2197NQHW}    7/20/2010 3:10   RKD0604    PC-9379    Connect
http {V2A0-K2QI17TY-9117EZXD}          7/20/2010 3:18   RKD0604    PC-9379
           http://wikileaks.org/Julian_Assange/assange/The_Real_Story_About_DTAA/Gur_Erny_Fgbel_Nobhg_QGNN1528513805.php
device     {D4X5-E4OR78FJ-7133JDAG}    7/20/2010 3:30   RKD0604    PC-9379    Disconnect
logon      {Z7K6-U2FK63XM-6557ANKS}    7/20/2010 3:36   RKD0604    PC-9379    Logoff
```

Contextual Data generation from the log using Llama 3.1

On July 13, 2010, at 20:04, a user identified by the username RKD0604 initiated a series of suspicious activities that raised red flags for cybersecurity monitoring systems. Over the course of the next several hours, the user logged in 13 times, showcasing a concerning pattern of frequent access outside standard working hours. During this period, they connected to a removable drive a total of 4 times, which is particularly alarming as this action suggests an attempt to transfer sensitive information. The user also engaged in a significant amount of online activity, accessing 20 different websites, with a notable concentration on wikileaks.org, indicating a possible intention to share or leak confidential data. Notably, one of the websites accessed was a well known platform for whistleblowing, further intensifying concerns regarding the user's motives. The combination of these actions unprecedented logins, removable drive usage, and engagement with controversial websites— paints a troubling picture of potential insider threats. By the time the user logged out for the final time on July 20, 2010, at 3:36, the overall activity profile had clearly diverged from normal patterns, raising serious alarms about the integrity of sensitive information within the organization.

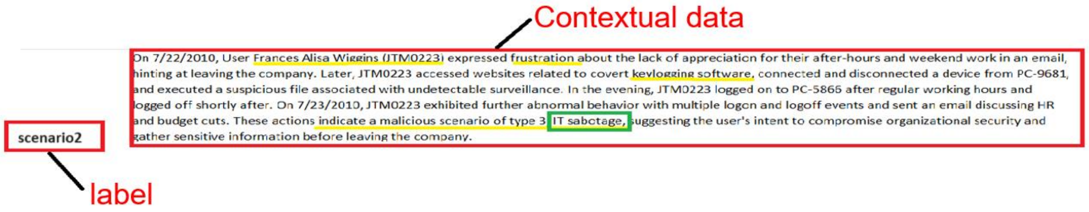
Contextual DATASET [6]



email	{A1I5-J9	7/22/2010 7:27	JTM0223	PC-9681	Frances.Alisa.Wiggins@dtaa.con			Jerry.Tad.I	34773	2	fed up tc
email	{J8A9-J0	7/22/2010 9:39	FAW0032	PC-5866	Jerry.Tad.Mccall@dtaa.com			Frances.Al	24513	0	training I
email	{F5S5-U8	7/22/2010 10:34	JTM0223	PC-9681	Frances.Alisa.Wiggins@dtaa.con			Jerry.Tad.I	17670	0	bad thing
email	{Y3B5-A	7/22/2010 11:47	FAW0032	PC-5866	Jerry.Tad.Mccall@dtaa.com			Frances.Al	23871	0	training
http	{H2L5-R	7/22/2010 12:35	JTM0223	PC-9681	http://dov	malware file password username undetectable protect free tria					
http	{P8J5-N7	7/22/2010 12:44	JTM0223	PC-9681	http://ww	keyboard password file captured captured effective program ke					
http	{J8R5-W	7/22/2010 13:03	JTM0223	PC-9681	http://ww	keylogging keyboard recommend free activity free secure userna					
http	{S1B2-F3	7/22/2010 13:16	JTM0223	PC-9681	http://ww	effective advanced keylogging captured file keyboard hidden ev					
http	{F5X4-K6	7/22/2010 13:36	JTM0223	PC-9681	http://ww	effective free trial undetectable secure file undetectable userna					
http	{M4U7-E	7/22/2010 14:28	JTM0223	PC-9681	http://ww	covert trial surveillance download covert captured easy easy ev					
device	{W9T2-L	7/22/2010 15:09	JTM0223	PC-9681	Connect						
file	{Y9S9-O8	7/22/2010 15:11	JTM0223	PC-9681	1OVEKEF2	4D-5A-90-00-03-00-00-00-04-00-00-00-FF-FF-00-00-B8-00-00-0					
device	{P1L1-R2	7/22/2010 15:14	JTM0223	PC-9681	Disconnect						
logon	{J3A8-U4	7/22/2010 19:56	JTM0223	PC-5866	Logon						
device	{F9R6-A2	7/22/2010 19:58	JTM0223	PC-5866	Connect						
device	{G2Q9-S	7/22/2010 20:06	JTM0223	PC-5866	Disconnect						
logon	{J4F9-P8	7/22/2010 20:11	JTM0223	PC-5866	Logoff						
logon	{Y1S9-R8	7/23/2010 17:27	JTM0223	PC-5866	Logon						
logon	{W5W6-	7/23/2010 17:39	JTM0223	PC-5866	Logoff						
logon	{W2L4-R	7/23/2010 17:46	FAW0032	PC-5866	Logon						
email	{J2N7-M	7/23/2010 17:47	FAW0032	PC-5866	Daria.Felic	Cyrus.Connor.Atkinso		Frances.Al	13170	0	hr budge
logon	{N2Q2-J	7/23/2010 18:01	FAW0032	PC-5866	Logoff						

Figure: Chain of user actions for threat scenario 3

- ➢ Convert it into contextual form(generated by Fine-tuned Llama3.1 model)


Contextual data
label

scenario2

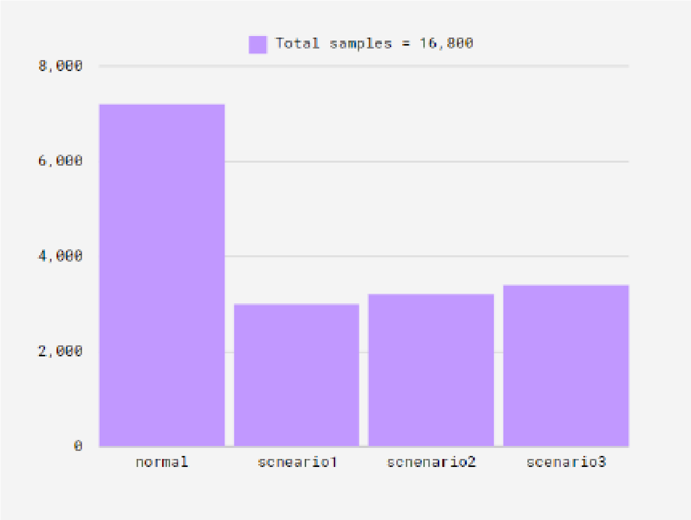
On 7/22/2010, User Frances Alisa Wiggins (JTM0223) expressed frustration about the lack of appreciation for their after-hours and weekend work in an email, hinting at leaving the company. Later, JTM0223 accessed websites related to covert keylogging software, connected and disconnected a device from PC-9681, and executed a suspicious file associated with undetectable surveillance. In the evening, JTM0223 logged on to PC-5866 after regular working hours and logged off shortly after. On 7/23/2010, JTM0223 exhibited further abnormal behavior with multiple logon and logoff events and sent an email discussing HR and budget cuts. These actions indicate a malicious scenario of type 3 IT sabotage, suggesting the user's intent to compromise organizational security and gather sensitive information before leaving the company.

Figure: Contextual data for user actions for threat scenario 3 for binary classification

DATASET for SecureBERT[7]

User_activity	Label
On July 4, 2024, at 03:37, the user with the username XYWJ0017 exhibited a significant shift in behavior that raised red flags for potential data exfiltration. Initially, this user had no history of utilizing removable drives; however, this day marked a drastic change as they began using a USB drive for the first time. Over the course of the day, from 03:37 to 11:37, XYWJ0017 logged in 13 times outside of regular working hours, a pattern that starkly contrasted with their previous login habits. During these late-night sessions, the user uploaded a total of 92 files to wikileaks.org, indicating a deliberate attempt to leak sensitive information. The combination of recent removable drive usage, coupled with the unusual frequency of after-hours logins, suggested a premeditated plan to exfiltrate data before ultimately leaving the organization. This alarming behavior not only deviated from XYWJ0017's past activity but also highlighted a potential insider threat that warranted immediate investigation.	Scenario1
On July 4th, 2012, from 19:46 to 04:46 the next day, a user under the username VAIT4615 exhibited a series of suspicious activities indicative of potential intellectual property theft. Over the course of the investigation, it was noted that this user accessed 21 job search websites during non-business hours, indicating a possible intention to explore employment opportunities elsewhere. Additionally, the user engaged in communications with competitors, sending 4 emails that raised red flags due to the sensitive nature of the conversations. The analysis revealed that VAIT4615 connected USB drives three times to their workstation, contradicting company policy regarding external storage devices, which allowed for the easy transfer of proprietary data. During the time frame of the investigation, the user conducted large data transfers, which typically flagged as unusual behavior, especially when paired with the 46 proprietary files that were deliberately copied and potentially exfiltrated. Furthermore, this user logged in 7 times outside standard operational hours, suggesting a premeditated effort to avoid detection while carrying out these activities. Collectively, these actions point towards a calculated approach to misappropriate critical intellectual property, highlighting the urgency for implementing stricter data protection and monitoring measures within the organization.	Scenario 2
On July 1, 2024, at 06:35, a user with the username LEWM3093 initiated a malicious operation that would have significant repercussions for the organization. Within the confines of an hour, LEWM3093 executed a series of deceptive actions beginning with the download of a sophisticated keylogger, aimed at capturing sensitive data. The user subsequently utilized a thumb drive to transfer the keylogger onto the supervisor's workstation, thereby bypassing internal security protocols. Once the keylogger was successfully installed, LEWM3093 commenced the collection of keylogs, discreetly monitoring the supervisor’s activities to gather critical information. After gaining all necessary data, the user took the bold step of logging into the supervisor’s account without authorization. Capitalizing on this unauthorized access, the user proceeded to dispatch a mass email to all employees, issuing alarming and potentially harmful misinformation that could lead to panic and distrust within the company. The way LEWM3093 executed each step—carefully planning the actions and timing—underscored the malicious intent of this sabotage scheme. Finally, at 07:35, the user abruptly exited the organization, leaving behind a trail of chaos and an urgent need for the IT department to respond to the breach and mitigate the fallout of the actions taken during this alarming hour. The sequence of events not only highlights the technical prowess of the insider threat but also reveals the profound potential for damage that such acts can inflict on organizational integrity and morale.	Scenario 3
On April 24, 2011, the user with the username KRSJ5575 logged an active work session starting at 00:51 and wrapping up exactly four hours later at 04:51. During this typical office routine, a total of 15 emails were checked, reflecting the user's dedication to staying updated on communications. In addition to managing emails, KRSJ5575 attended two meetings, contributing to discussions that shaped the direction of ongoing projects. Throughout this focused period, three reports were meticulously created, showcasing the user’s commitment to delivering quality work consistently. Furthermore, there was one instance of collaboration with a colleague, illustrating KRSJ5575’s engagement in teamwork, a vital part of their professional environment. This snapshot of user activity presents a familiar pattern of productivity, as the tasks completed—such as email management, report generation, and collaborative efforts—underscore a routine workflow seen in many office settings.	normal

Contextual DATASET [6]



<u>Statistics of dataset for SecureBERT</u>

Normal = 7200 samples

Abnormal = 9600 samples

- Scenario1 = 3000 samples
- Scenario2 = 3400 samples
- Scenario3 = 3200 samples

Figure:Histogram of dataset used for training secureBERT model

METHODOLOGY[8]

RoBERTa (Robustly Optimized BERT)

- Modified version of BERT
- Modifies key hyperparameters in BERT
- Improves on the masked language modeling objective compared with BERT

SecureBERT (Domain specific model based on RoBERTa

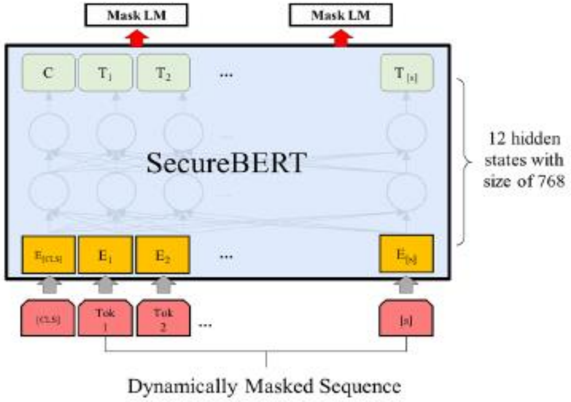
- Continual learning of RoBERTa using cyber data returns SecureBERT adjusting weights with smaller sized data is difficult
- First domain specific model, trained on cybersecurity corpus
- Showed high performance in pretraining task
- Trained on large corpus of data collected from different cybersecurity resources

METHODOLOGY[9]

SecureBERT

Classification head:

- hidden layer with dimensions 768 X 768.
- output layer with dimensions 768 X 4.
- probability vector with dimensions 1 X 4.



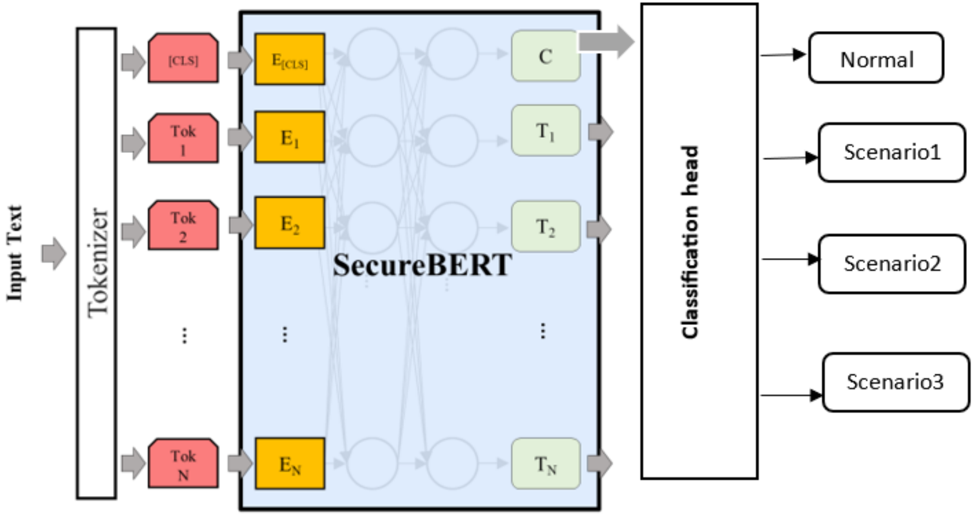


Figure: The architecture of SecureBERT model

METHODOLOGY[10]

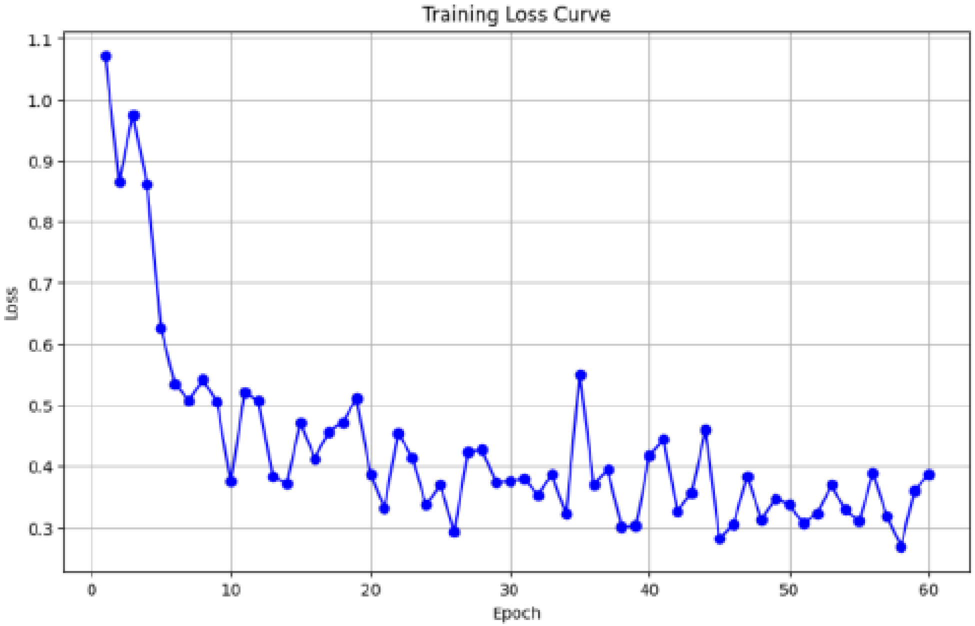
Training SecureBERT model for insider threat classification

- Fine tuned the pretrained SecureBERT model by training it into the contextual dataset(generated from the raw events of data)
- L2 regularization used to prevent overfitting
- Used AdamW optimizer to minimize the error of cross entropy loss function
- Used GridSearchCV for finding optimal hyperparameters
- Softmax as activation function in the classification layer for multi-class(scenario classification)

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

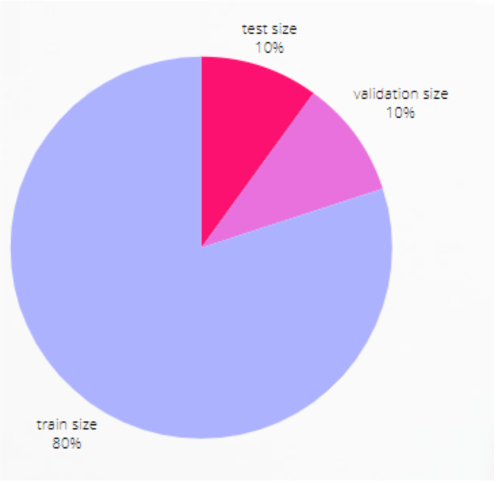
RESULTS[1]

Training- Testing loss curve obtained after fine-tuning secureBERT model

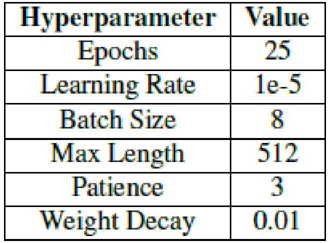


RESULTS[2]

Training SecureBERT for multiclass Classification

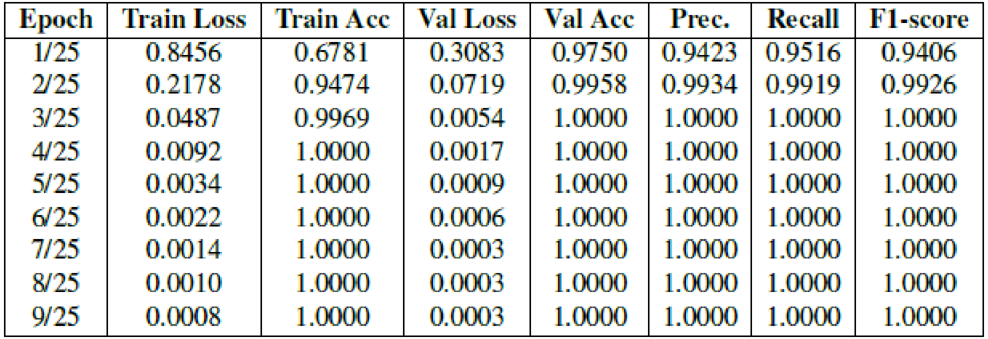


Splitting the data



Hyperparameter	Value
Epochs	25
Learning Rate	1e-5
Batch Size	8
Max Length	512
Patience	3
Weight Decay	0.01

Result obtained



Epoch	Train Loss	Train Acc	Val Loss	Val Acc	Prec.	Recall	F1-score
1/25	0.8456	0.6781	0.3083	0.9750	0.9423	0.9516	0.9406
2/25	0.2178	0.9474	0.0719	0.9958	0.9934	0.9919	0.9926
3/25	0.0487	0.9969	0.0054	1.0000	1.0000	1.0000	1.0000
4/25	0.0092	1.0000	0.0017	1.0000	1.0000	1.0000	1.0000
5/25	0.0034	1.0000	0.0009	1.0000	1.0000	1.0000	1.0000
6/25	0.0022	1.0000	0.0006	1.0000	1.0000	1.0000	1.0000
7/25	0.0014	1.0000	0.0003	1.0000	1.0000	1.0000	1.0000
8/25	0.0010	1.0000	0.0003	1.0000	1.0000	1.0000	1.0000
9/25	0.0008	1.0000	0.0003	1.0000	1.0000	1.0000	1.0000

RESULTS[3]

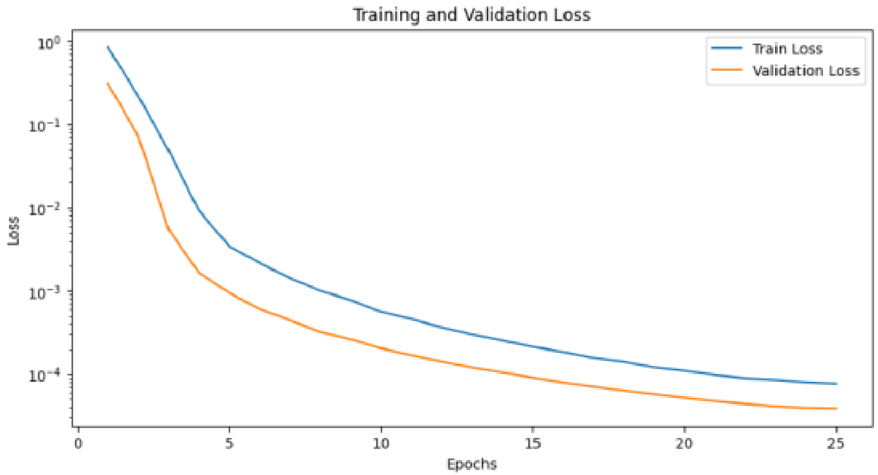
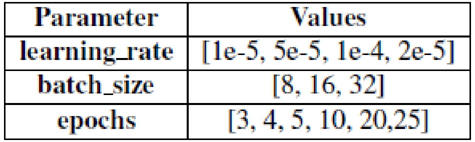


Figure: Training and validation loss and ROC curve for multiclass classification

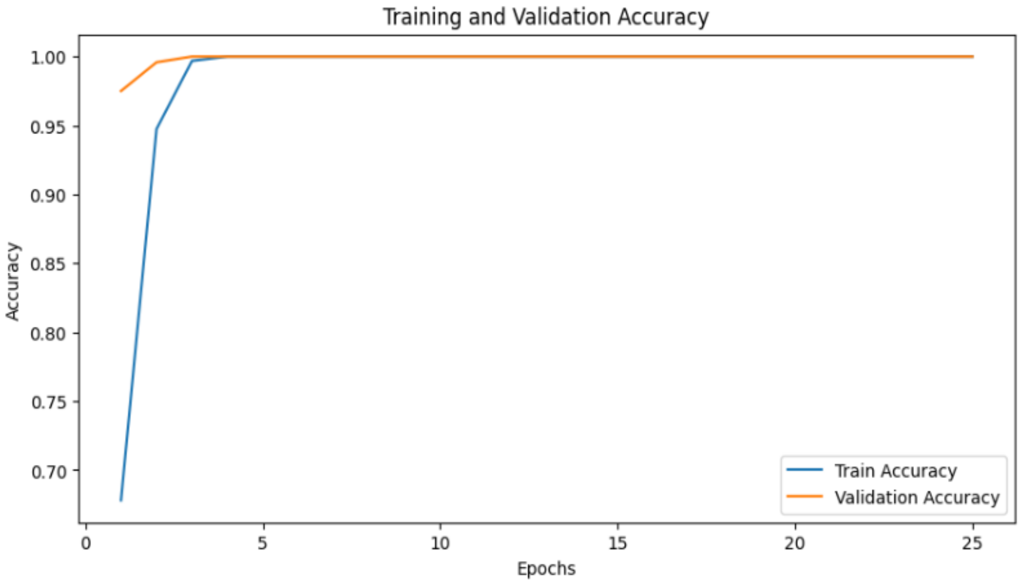
RESULTS[4]



Parameter	Values
learning_rate	[1e-5, 5e-5, 1e-4, 2e-5]
batch_size	[8, 16, 32]
epochs	[3, 4, 5, 10, 20,25]

Best parameters obtained after grid analysis

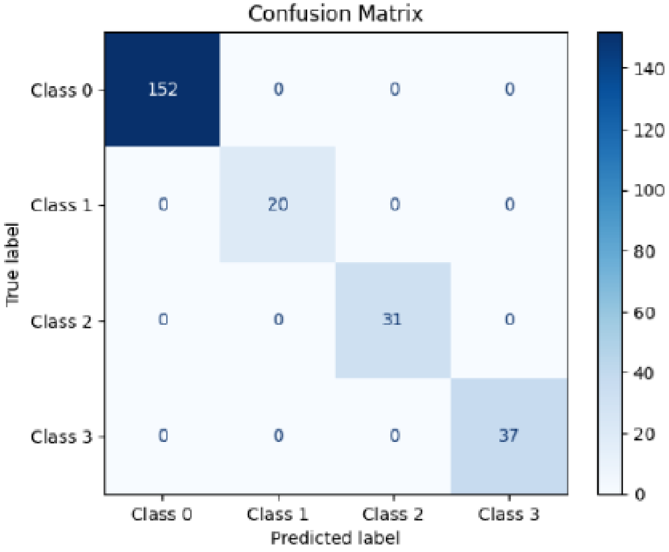
- learning rate = 1e-5
- batch size = 16
- number of epochs = 25



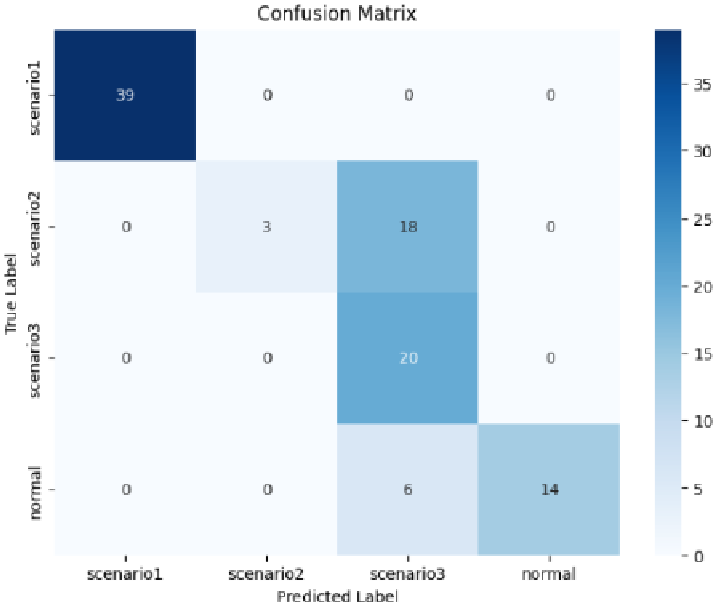
Parameters grid for hyperparameter tuning by GridSearchCV

RESULTS[5]

Testing on different sets of test data


Confusion Matrix

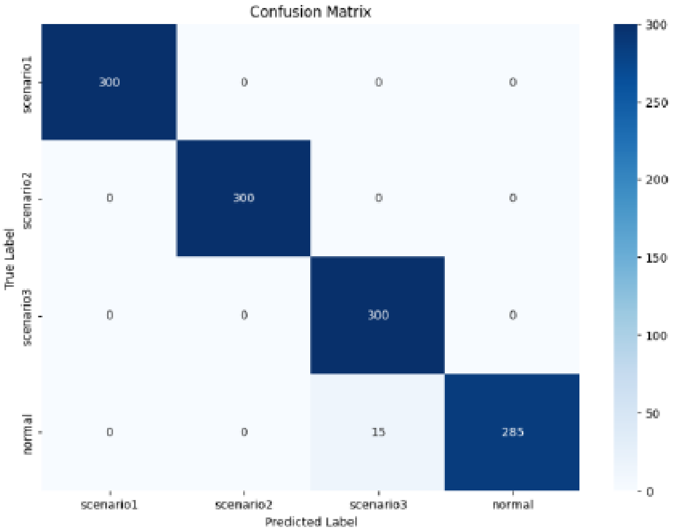
True label \ Predicted label	Class 0	Class 1	Class 2	Class 3
Class 0	152	0	0	0
Class 1	0	20	0	0
Class 2	0	0	31	0
Class 3	0	0	0	37


Confusion Matrix

True Label \ Predicted Label	scenario1	scenario2	scenario3	normal
scenario1	39	0	0	0
scenario2	0	3	18	0
scenario3	0	0	20	0
normal	0	0	6	14

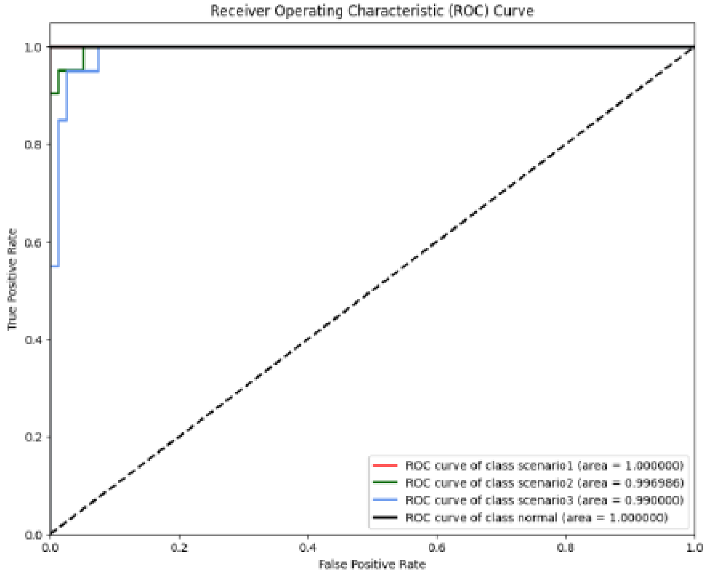
RESULTS[6]

Testing on different sets of test data



Confusion Matrix

True Label \ Predicted Label	scenario1	scenario2	scenario3	normal
scenario1	300	0	0	0
scenario2	0	300	0	0
scenario3	0	0	300	0
normal	0	0	15	285



Receiver Operating Characteristic (ROC) Curve

Evaluation Metrices

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

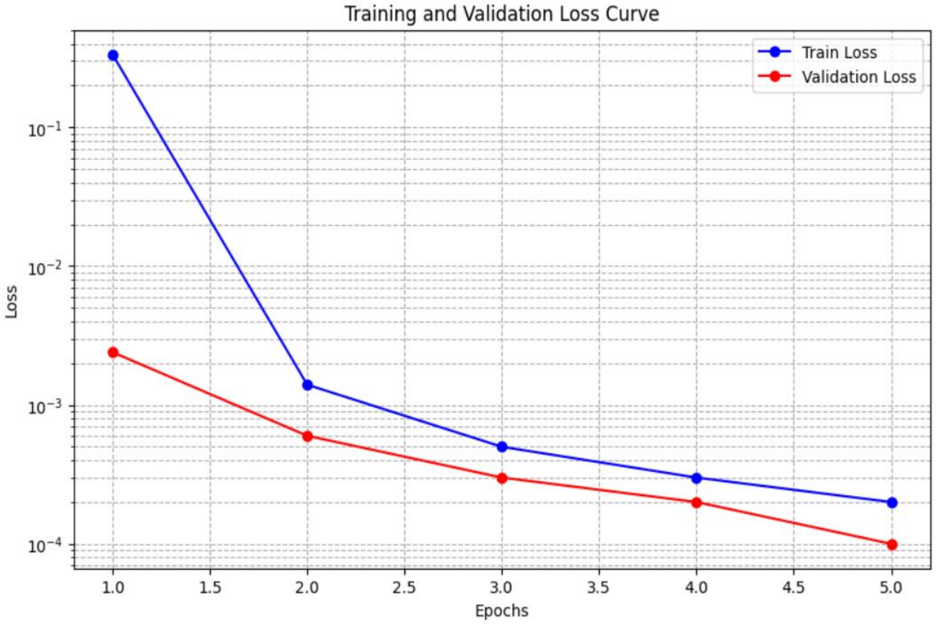
$$Precision = \frac{TP}{TP + FP}$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

$$Recall = \frac{TP}{TP + FN}$$

DISCUSSION AND ANALYSIS

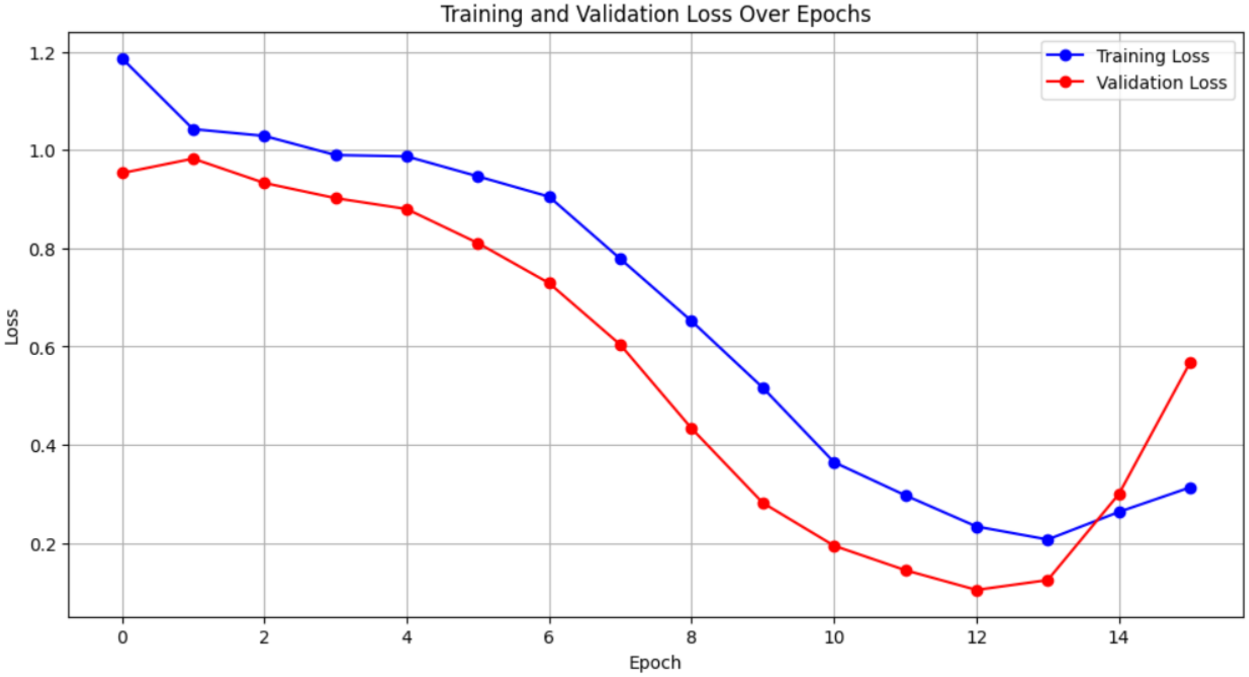
Best Case Scenario



Batch Size = 16, learning rate = 1e-5 , epochs =5

DISCUSSION AND ANALYSIS

Worst Case Scenario



Batch Size = 8, learning rate = 5e-6 , epochs = 40 (early stopping at 15th epoch)

DISCUSSION AND ANALYSIS

Comparison with results of other researchers

	Dataset	**Result Obtained**	**Methods used**
User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder Insider Threat Detection	CERT 4.2 - Used structured data	Accuracy 90.17 % (only binary classification)	-LSTM -RNN models -LSTM Autoencoder
User Behavior Analytics for Insider Threat using Transformer Based Approach	CERT 4.2 - Converted to the contextual data	Accuracy 82.45% in 1st epoch, 99.99 % in after 2nd epoch (binary and multi scenario classification)	-Llama3.1 for contextual data generation -SecureBERT for classification

FUTURE ENHANCEMENTS

- Evaluation on broader dataset
- Optimization of Llama 3.1 for more complex language scenarios
- Real time testing and deployment

CONCLUSION

This project has provided its contribution to the area of cybersecurity and natural language processing in the given ways in order to achieve the goal

- Contextual Data generation
- Multi class classification
- Performance and stability

PROJECT TIMELINE (GANTT CHART)

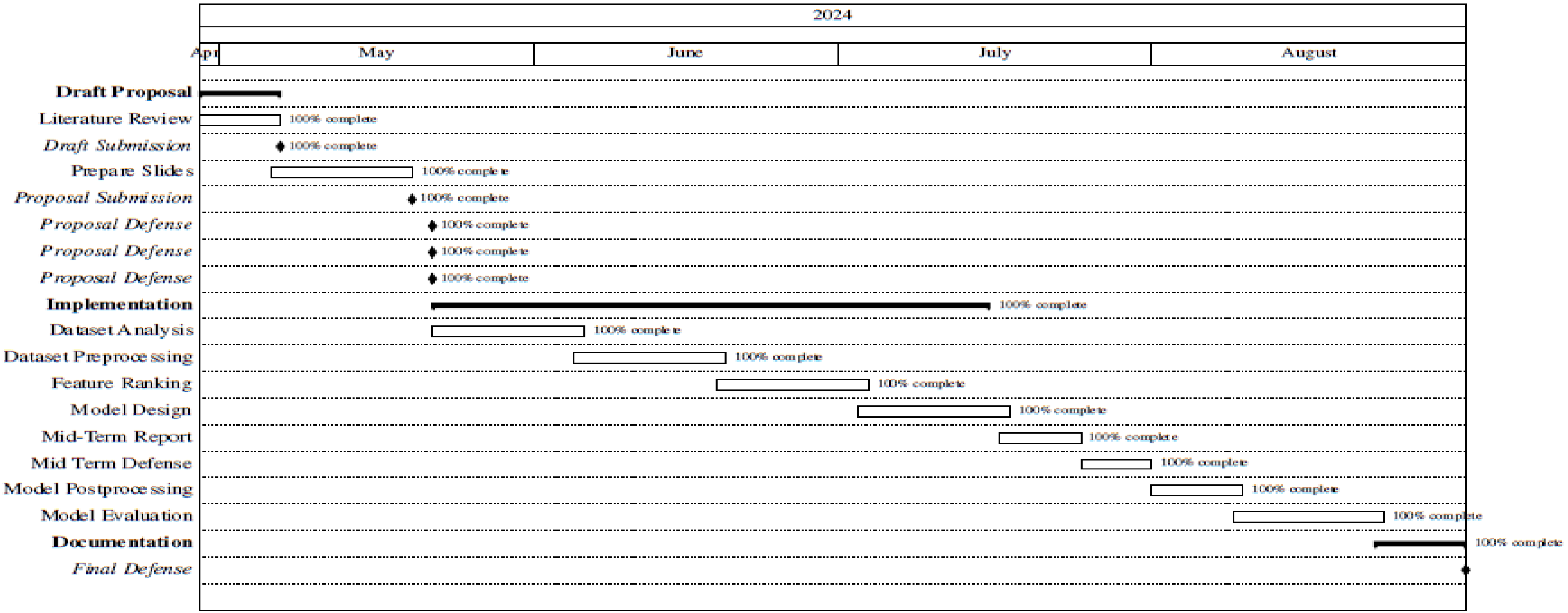


Fig: Tentative timeline of the project

REFERENCES

[1] Mohammed Nasser Al-Mhiqani, Rabiah Ahmad, Z. Zainal Abidin, Warusia Yassin, Aslinda Hassan, Karrar Hameed Abdulkareem, Nabeel Salih Ali, and Zahri Yunos. A review of insider threat detection: Classification, machine learning techniques, datasets, open challenges, and recommendations. Applied Sciences, 10(15), 2020.

[2] Fredrik Heiding, Bruce Schneier, Arun Vishwanath, Jeremy Bernstein, and Peter S. Park. Devising and detecting phishing: Large language models vs. smaller human models, 2023.

[3] Dahye Kim, Dongju Park, Honghyun Cho, and Kang. Insider threat detection based on user behavior modeling and anomaly detection algorithms. Applied Sciences, 9:4018, 09 2019.

[4] Abir Rahali and Moulay A. Akhloufi. Malbertv2: Code aware bert-based model for malware identification. Big Data and Cognitive Computing, 7(2), 2023.

REFERENCES

[5] Madhu Raut, Sunita Dhavale, Amarjit Singh, and Atul Mehra. Insider threat detection using deep learning: A review. In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pages 856–863, 2020.

[6] Balaram Sharma, Prabhat Pokharel, and Basanta Joshi. User behavior analytics for anomaly detection using lstm autoencoder - insider threat detection. In Proceedings of the 11th International Conference on Advances in Information Technology, IAIT 20, New York, NY, USA, 2020. Association for Computing Machinery.

[7] Aaron Tuor, Samuel Kaplan, Brian Hutchinson, Nicole Nichols, and Sean Robinson. Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. CoRR, abs/1710.00811, 2017.

THANK YOU!!