SecureSphere Innovations

We maintain privacy.

How to detect malicious cyber behavior

YOUR DATA IS MINE

by Dustin Lischke, Nilgün Yesil, Stefanie Reimers and Tanvi Goel

2024-07-17

SecureSphere Innovations
We maintain privacy.

The Team

We teamed up to protect your data.

Dustin Lischke

Data Scientist

Thuringia
Germany
International

Nilgün Yesil

Data Scientist

NRW
Germany (Remote)
International (Remote)

Tanvi Goel

Data Analyst

Munich
Germany (Remote)
International (Remote)

Stefanie Reimers

Data Analyst

Berlin
Germany

About us

We are a cybersecurity company, driven to help our clients protecting themselves and their customers from threats in the modern, interconnected world.

Our Goals

We want to make your company a safe place for customer data.

Goal # 1

Use data analytics to understand patterns and irregularities.

Goal # 2

Use machine learning to detect/ predict malicious activities before they can do harm.

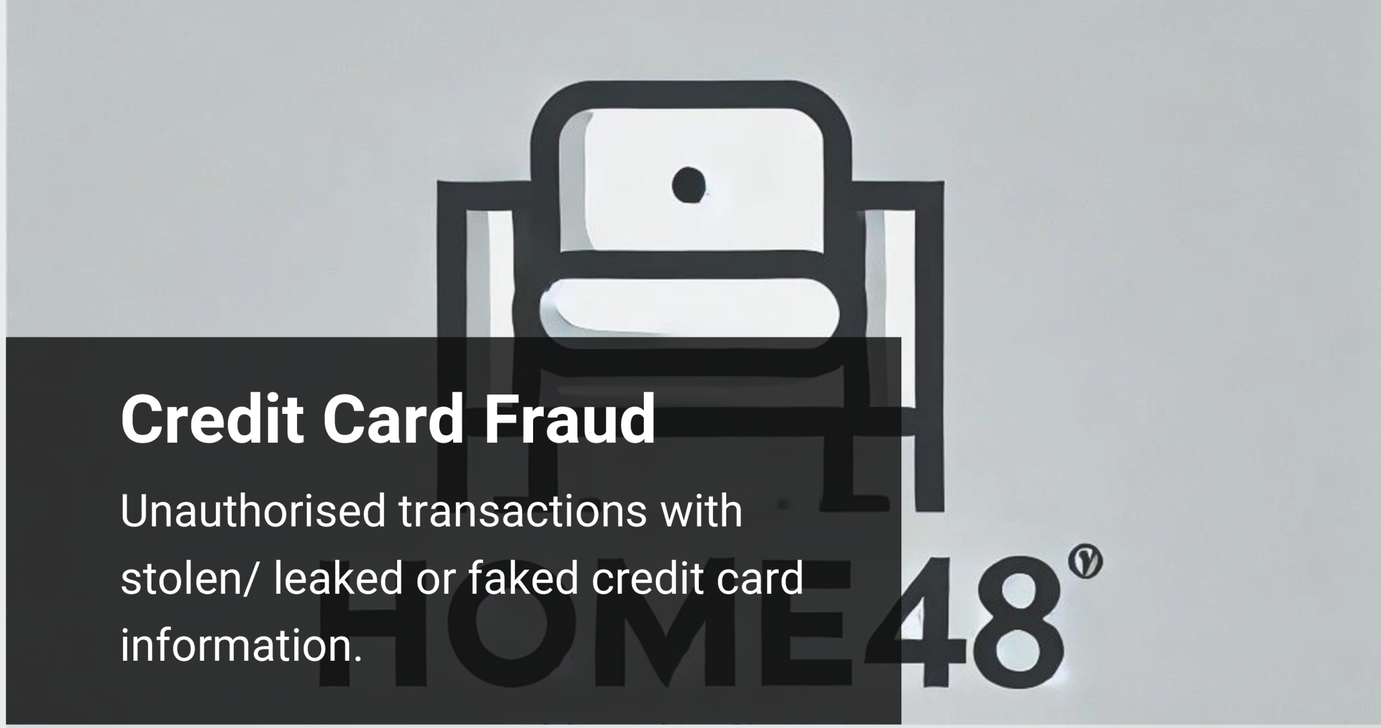
Goal # 3

Be your longterm security solution with iterative service cycles to ensure secure systems

SecureSphere Innovations

We maintain privacy.

Our Clients



Credit Card Fraud

Unauthorised transactions with stolen/ leaked or faked credit card information.

Network Attack

Customer suddenly faces a wave of attacks and intrusions on their network.

SecureSphere Innovations
We maintain privacy.

Credit Card Fraud

2030: predicted total loss from credit card fraud will be **$49.32 billion**

2022: **44%** of credit card users (USA) reported two or more fraudulent charges

estimation for 2024: **74%** of losses will be attributed to “Card-not-found” fraud type

2022: **915,000 children** in USA were victims of identity theft

3rd quarter of 2022: **15 million** data records exposed worldwide through data breaches

https://dataprot.net/statistics/credit-card-fraud-statistics/

SecureSphere Innovations
We maintain privacy.

Credit Card Fraud

Purchases have been made with stolen credit card information.

Products have been shipped without receiving the payment.

Customers raise concerns regarding the way their data is protected and processed by the company.

Overview of the Credit Card Fraud Detection Dataset

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431

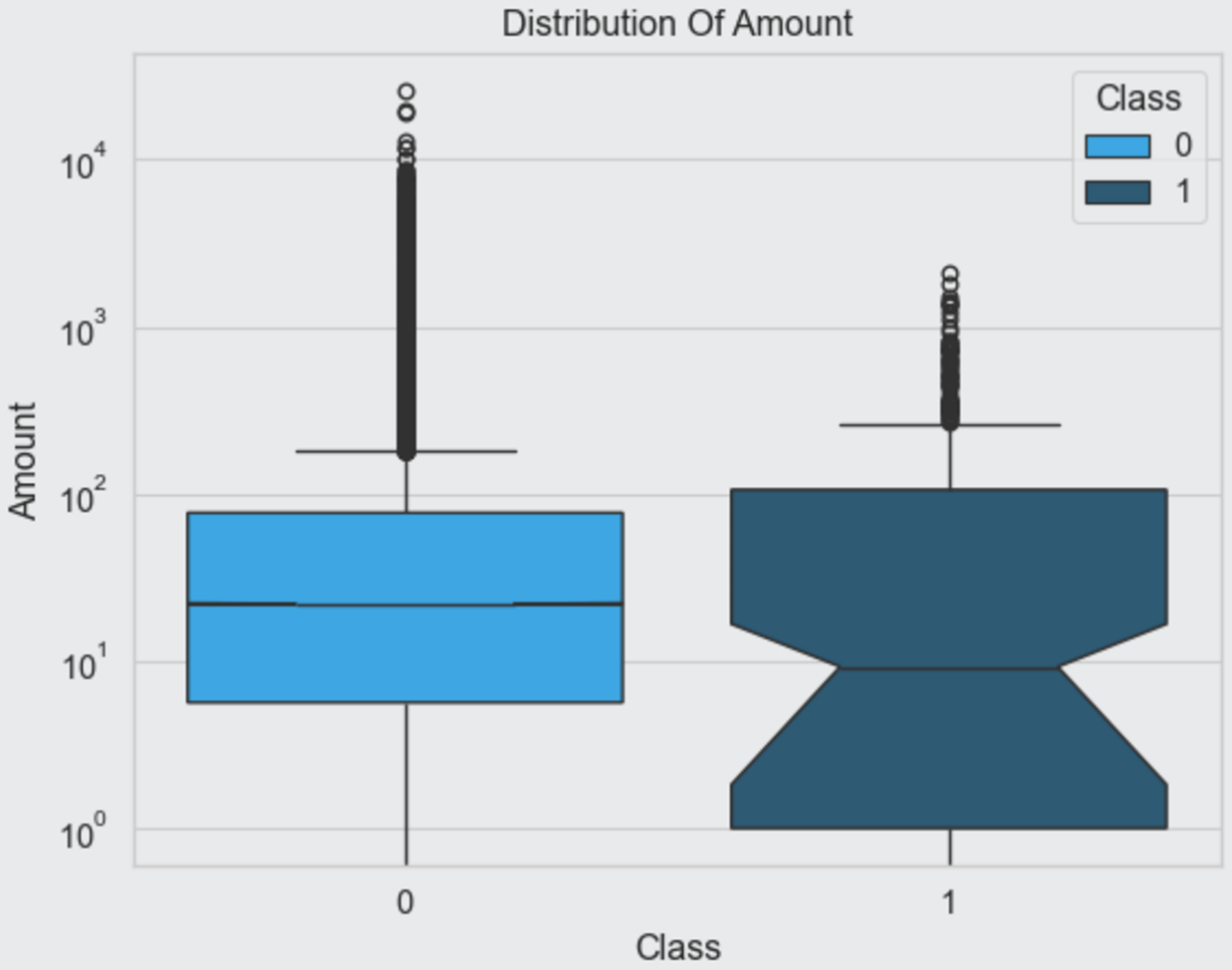
5 rows × 31 columns

28 encoded columns
time as seconds from reference time stamp
transaction amount
target (genuine/ fraud classification)

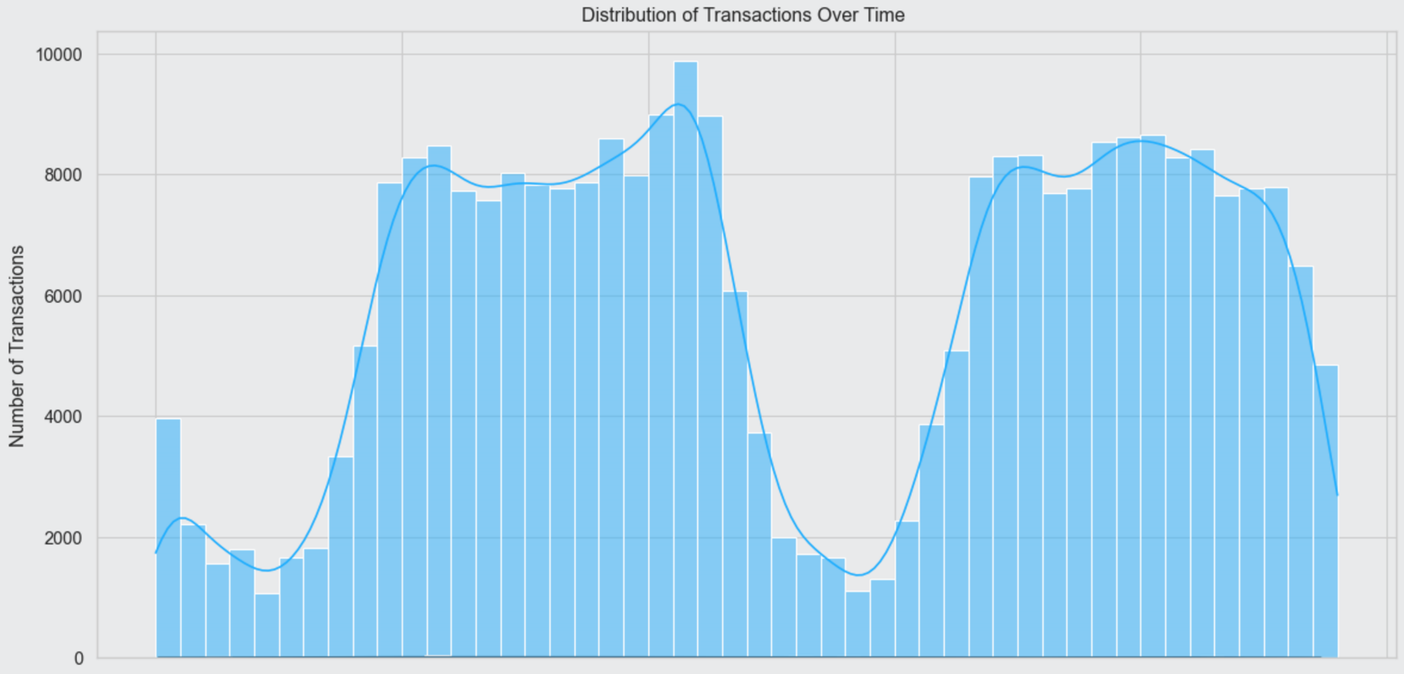
48h of recorded transactions, 284,807 records in total
284,315 genuine transactions, 492 frauds (**0.173%**)
no missing data, 1081 duplicates (of which 19 were classified as fraud)

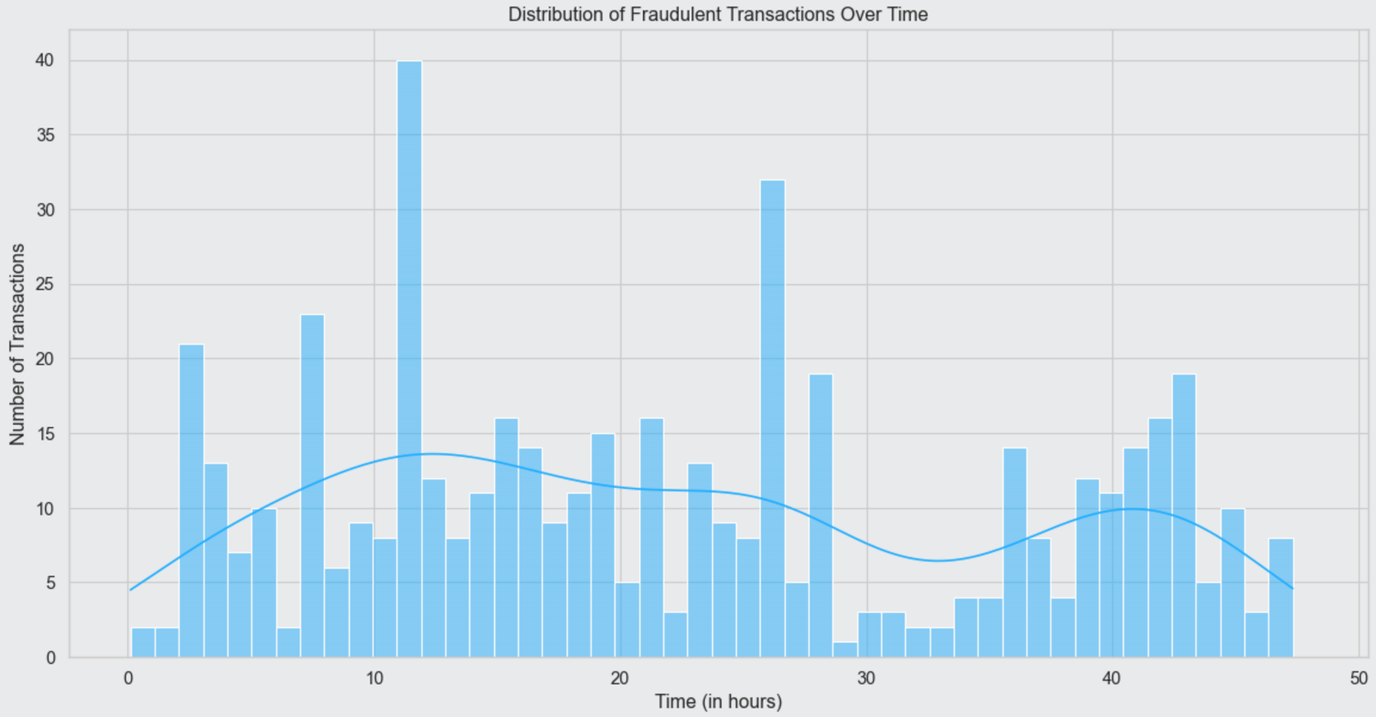
https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

Is Fraud typical for particularly high/ low Amounts of Money transferred?



0 - genuine
1 - fraud





Is there a window or pattern in the time-related appearence of fraud?

Machine Learning Strategies

Goal:

tool that is able to detect fraudulent credit card transactions

Metric/ Evaluation:

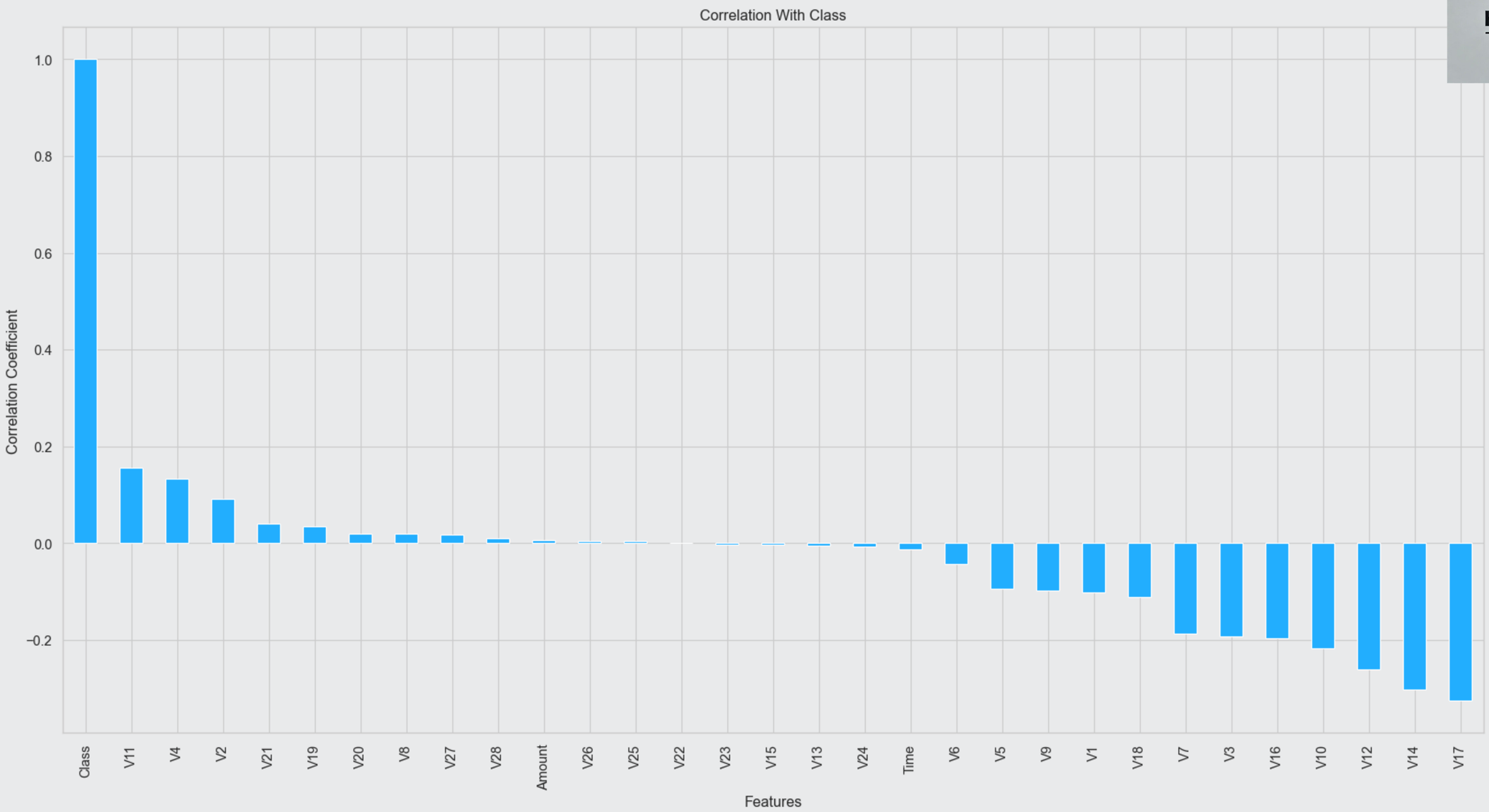
Fbeta score; number of false negative predictions

Considerations/ Strategy:

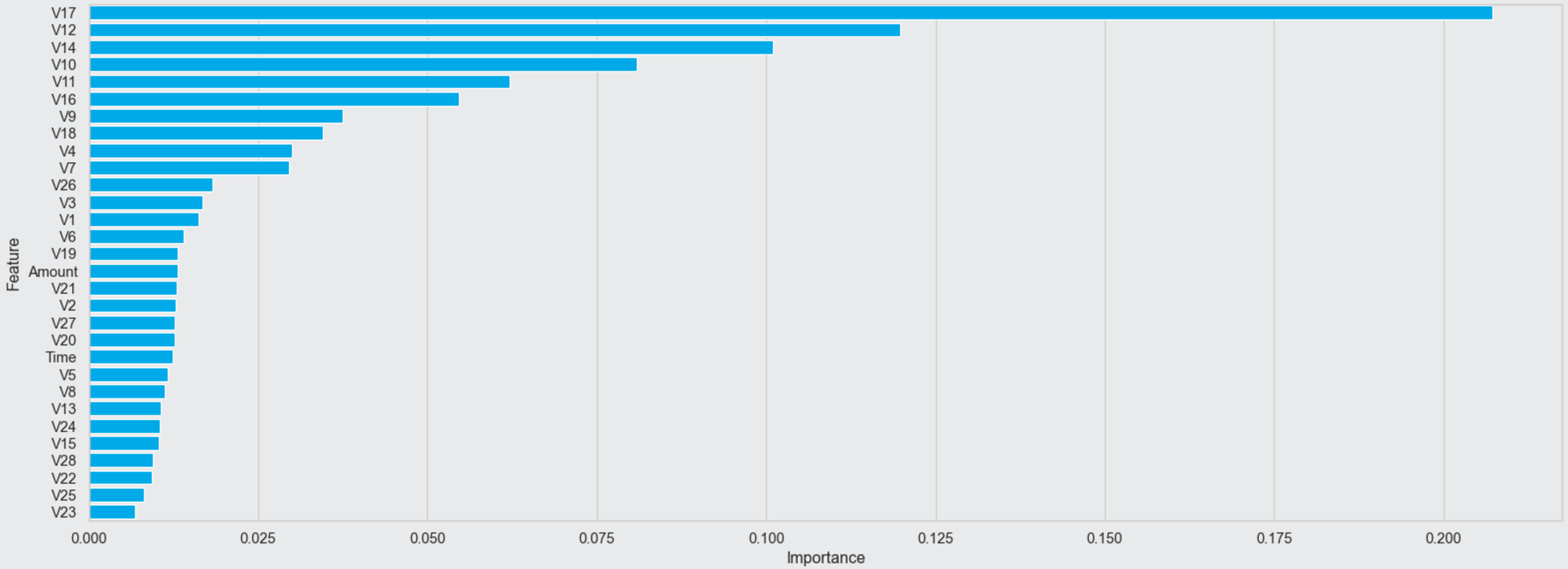
- dataset preparation
  - considering imbalance
  - selecting subset of features
- find suitable candidates for the given task
- use hyperparameter boosting to increase scores
- combine best candidates into complete detection system

F1 score:

Fbeta score:


Correlation With Class
Correlation Coefficient
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
Class
V11
V4
V2
V21
V19
V20
V8
V27
V28
Amount
V26
V25
V22
V23
V15
V13
V24
Time
V6
V5
V9
V1
V18
V7
V3
V16
V10
V12
V14
V17
Features

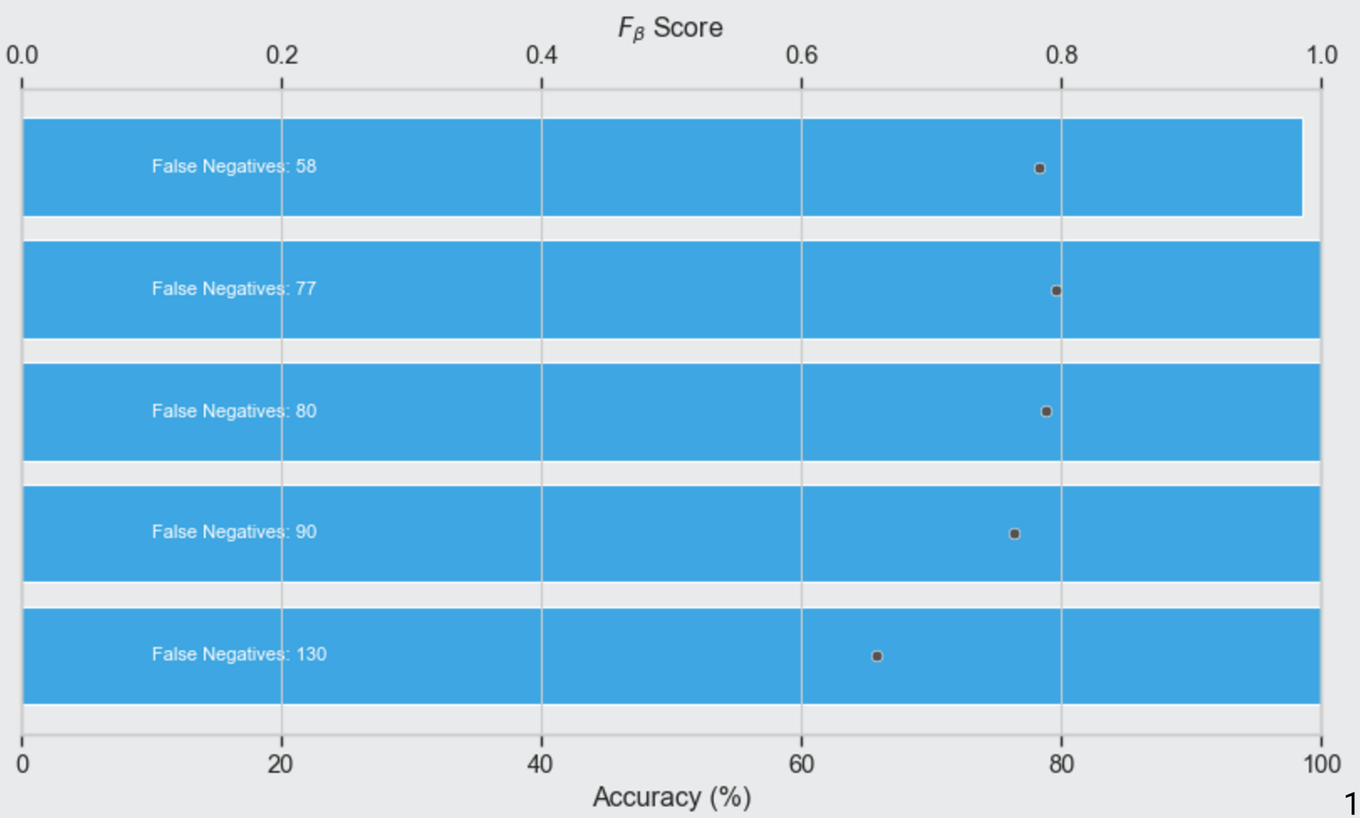
Feature Importance Credit Card Fraud



Comparing Boosted Models

☐ Full Dataset ☑ Hyperparameter Tuning

Model	Accuracy	Fbeta Score	False Negatives
Naive Bayes	98.596%	0.783534	58 (15.344 %)
Support Vector Classification	99.937%	0.796505	77 (20.370 %)
K Nearest Neighbours	99.936 %	0.788670	80 (21.164 %)
Random Forest	99.945%	0.762984	90 (23.810 %)
Logistic Regression	99.925%	0.657635	130 (34.392 %)

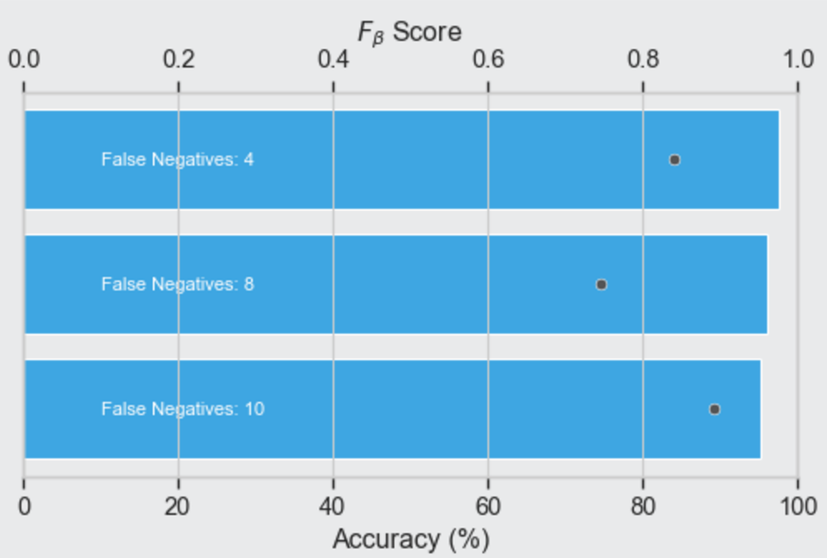


Best Results

Attack Detection

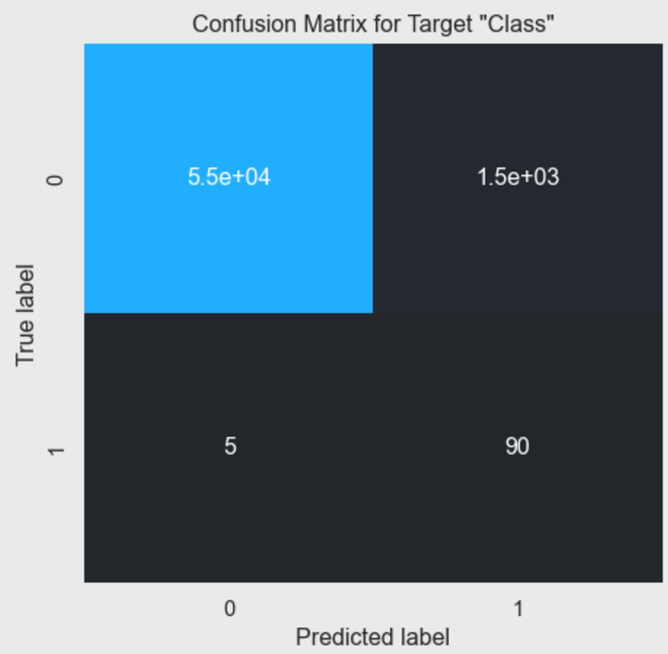
☑ Full Dataset ☑ Hyperparameter Tuning

Model	Accuracy	Fbeta Score	False Negatives
Random Forest	97.637 %	0.841051	4 (4.211 %)
Logistic Regression	96.095 %	0.744977	8 (8.421 %)
Neural Network	95.221 %	0.892209	10 (10.526 %)

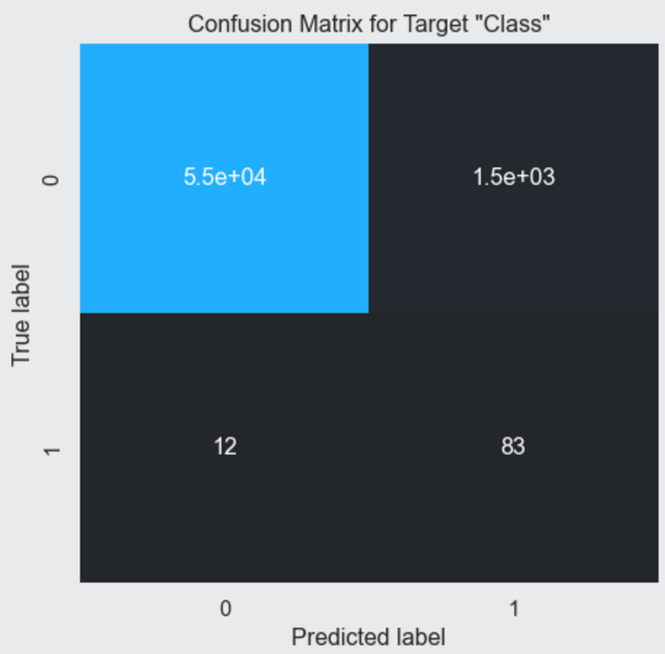


Confusion Matrices

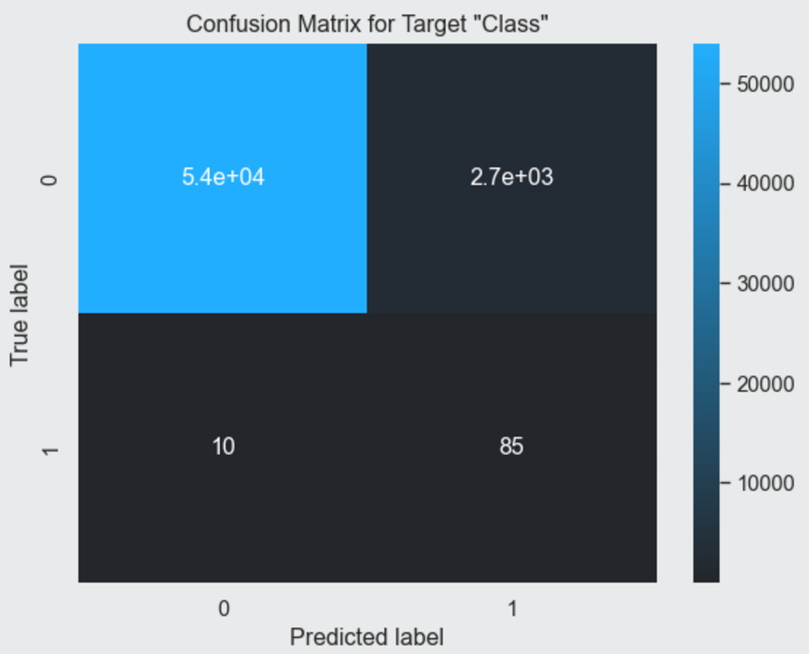
Random Forest (97.637 %)



Logistic Regrression (96.095%)



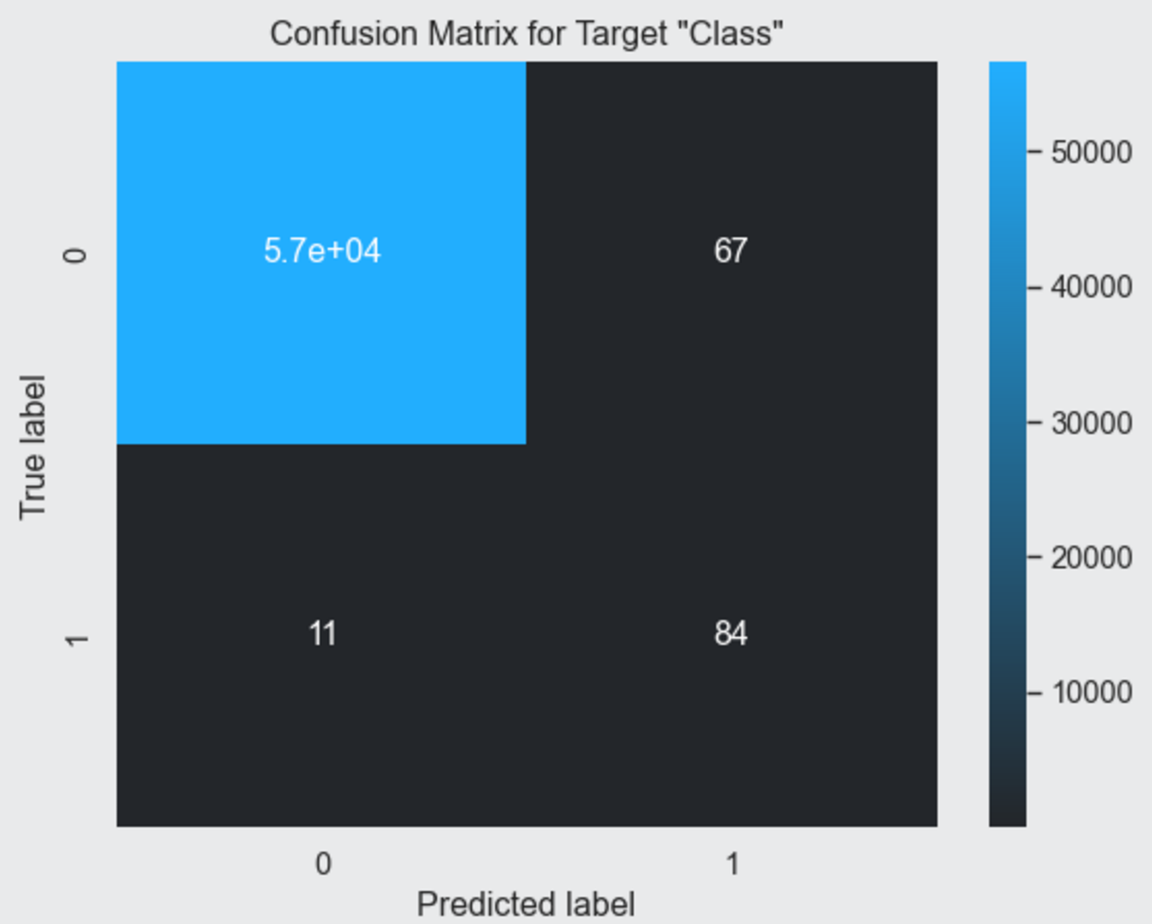
Neural Network (95.221 %)



0 - genuine

1 - fraud

Voting Classifier



Accuracy: 99.863 %

Fbeta: 0.879080

0 - genuine

1 - fraud

SecureSphere Innovations
We maintain privacy.

Conclusion

Best result: combination of 3 different models, **over 99% accuracy**

Feedback from customer regarding tolerance for false positives required; it is possible to push them to **0** if desired

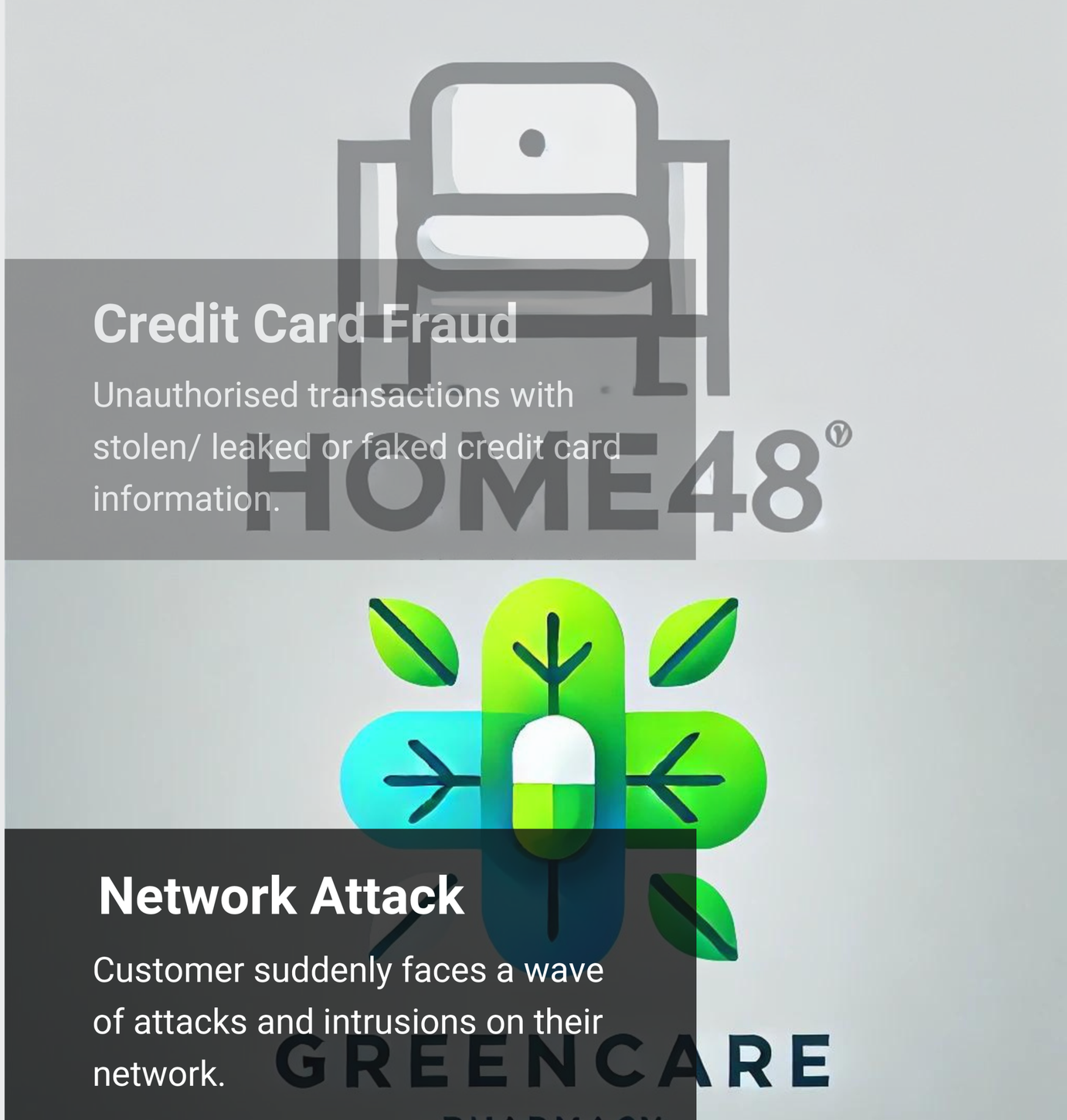
To further challenge and verify model’s accuracy, more data from beyond just 48h of recording is required

We can use data provided so far to build a defence tool that **will** lower their loss due to credit card fraud

SecureSphere Innovations

We maintain privacy.

Our Clients



Credit Card Fraud

Unauthorised transactions with stolen/ leaked or faked credit card information.

Network Attack

Customer suddenly faces a wave of attacks and intrusions on their network.

SecureSphere Innovations
We maintain privacy.

Network Attack

2020: governmental health ministries, information services and websites as well as delivery services (e.g. Lieferando) and others targeted by DDOS attacks

2023: DDOS attacks on financial sector grew by 121% compared to previous year

2023: ransomware make up 70.13% of cyberattacks worldwide, followed by network breach with 18.83%; DDOS accounts for 0.65%

https://pixelprivacy.com/resources/ddos-attack-statistics-report/
https://www.statista.com/statistics/1382266/cyber-attacks-worldwide-by-type/

SecureSphere Innovations
We maintain privacy.

Network Attack

Experienced several intrusion attempts, probing etc. before DOS attack almost shut down servers

Realisation: Greencare Pharmacy's network defences just barely capable of protecting the company and its associates

Hiring SecureSphere Innovations to assess their current situation and assist them in reinforcing their tools and strategies

Overview of the Network Traffic Attack Dataset

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_same_srv_rate	dst_host_diff_srv_rate
0	0	tcp	http	SF	215	45076	0	0	0	0	...	0.0	0.0
1	0	tcp	http	SF	162	4528	0	0	0	0	...	1.0	0.0
2	0	tcp	http	SF	236	1228	0	0	0	0	...	1.0	0.0
3	0	tcp	http	SF	233	2032	0	0	0	0	...	1.0	0.0
4	0	tcp	http	SF	239	486	0	0	0	0	...	1.0	0.0

5 rows × 43 columns

42 features plus attack type column

training: 22 unique attack types, 4 attack categories

verification: 37 unique attack types, 4 attack categories

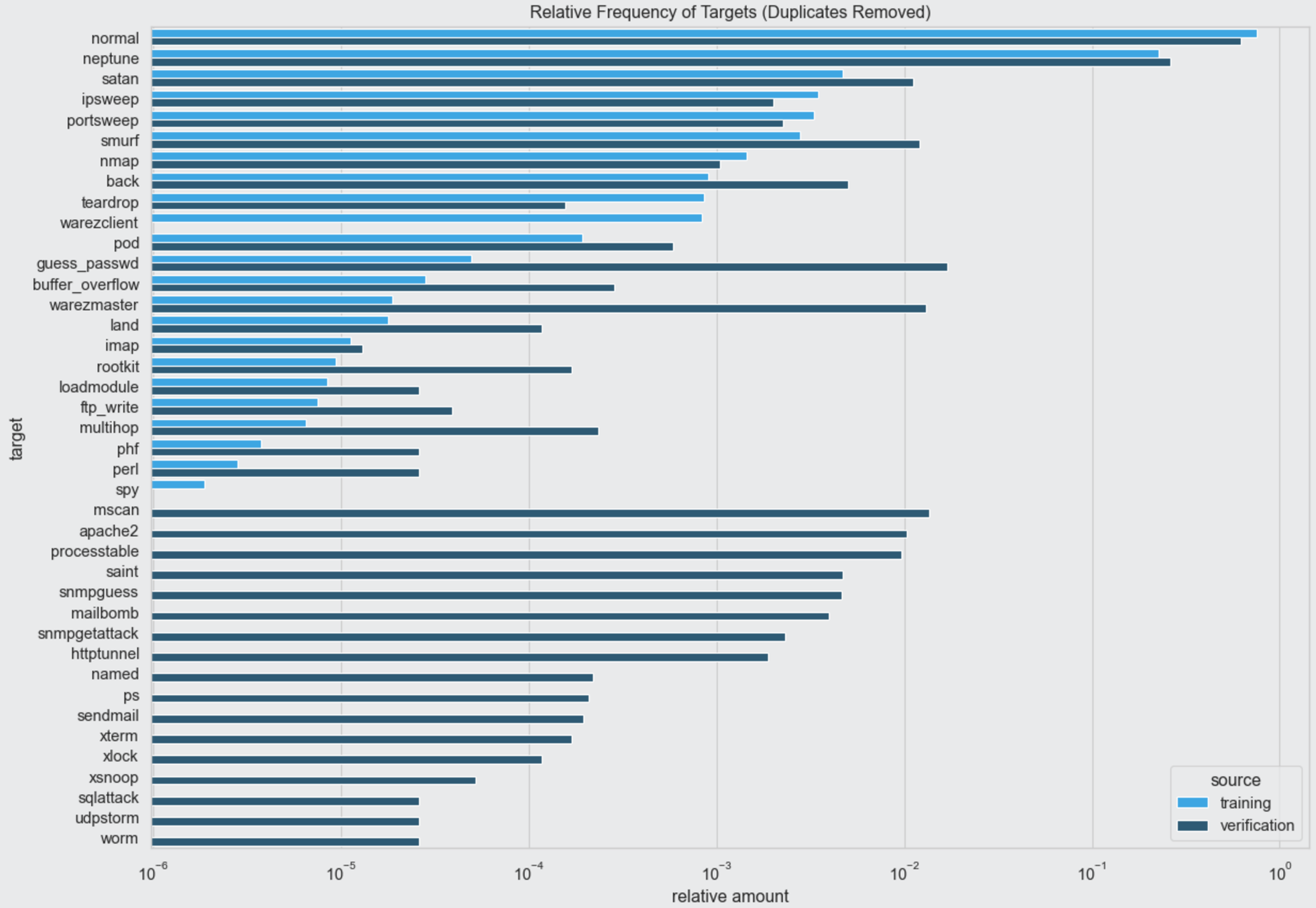
4,898,431 training and 311,029 verification records (including duplicates)

3,823,439 training and 233,738 verification duplicates (primarily DOS)

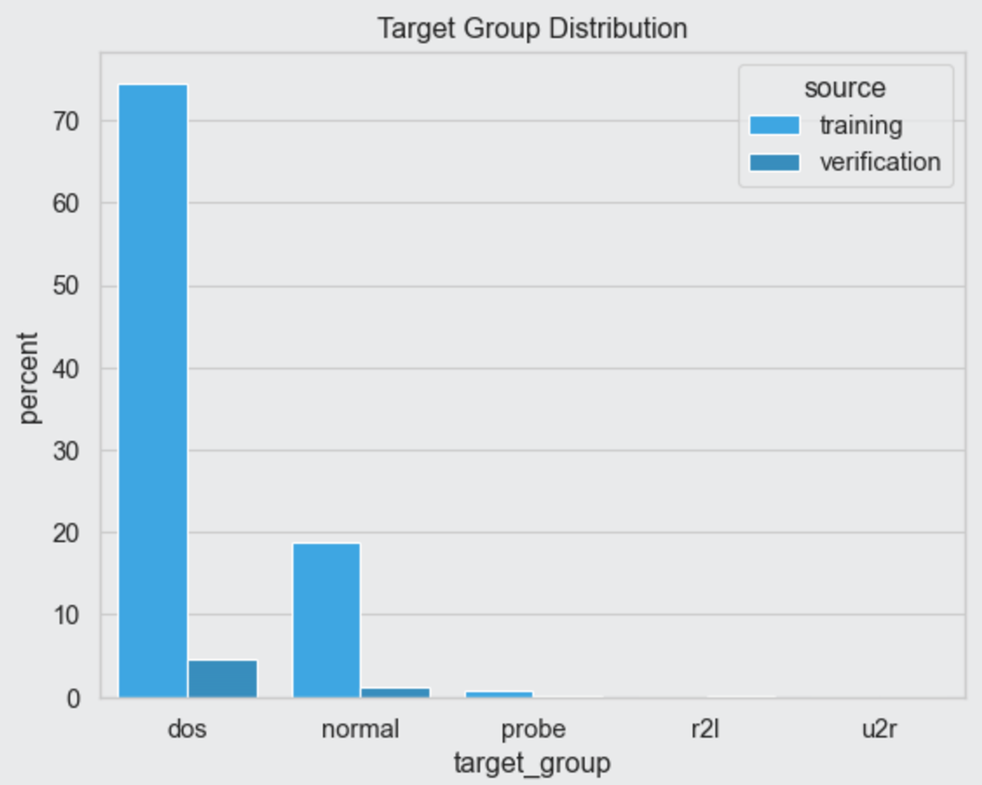
1,074,992 training and 77,291 unique verification records

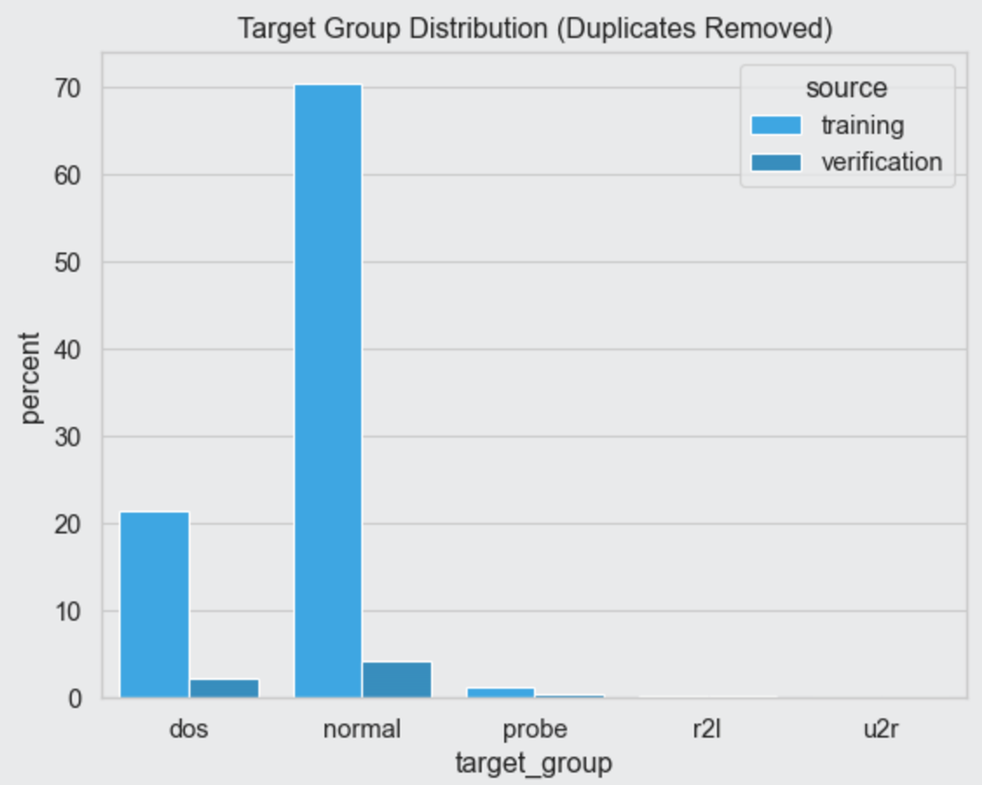
no missing data

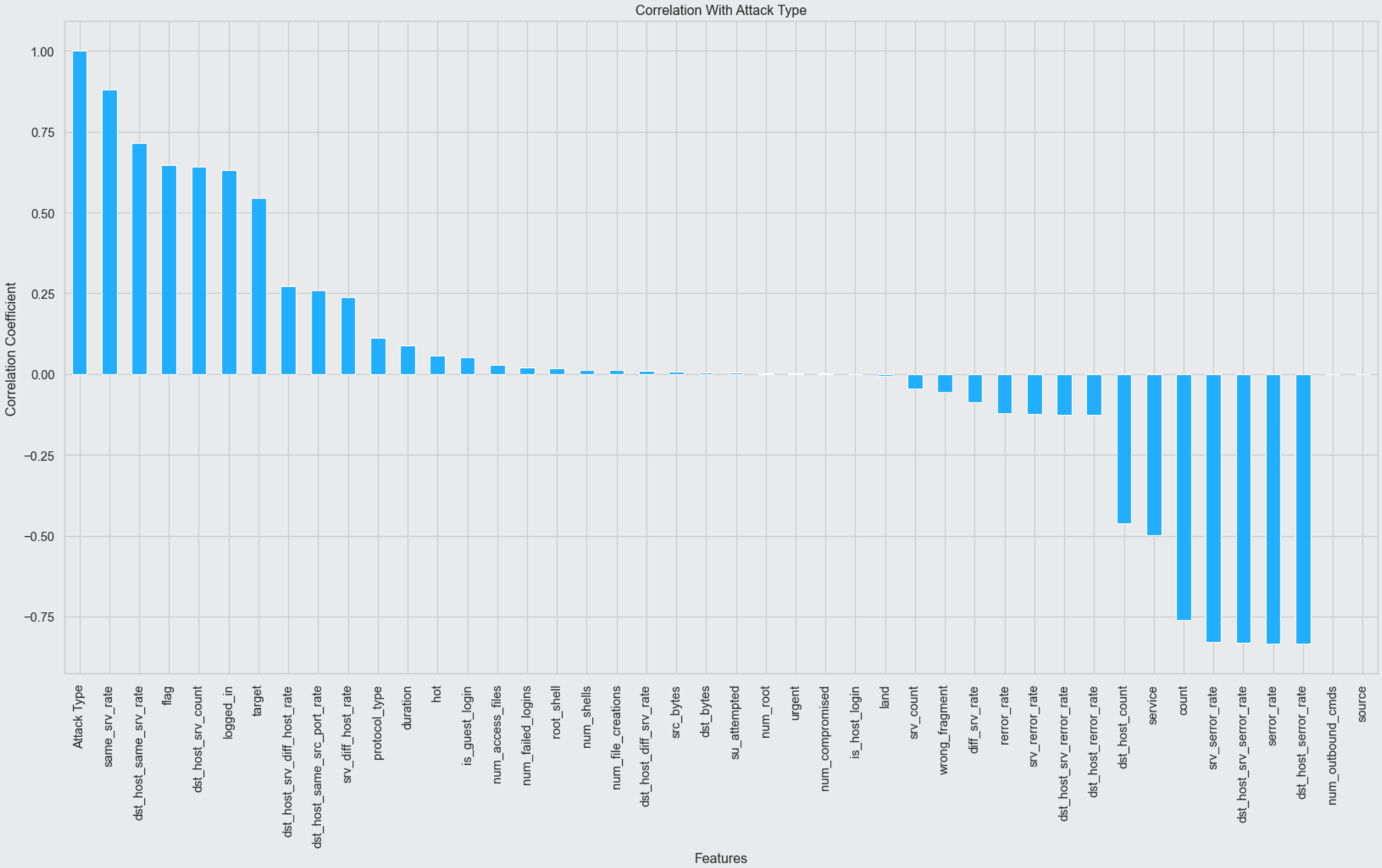
http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html


Relative Frequency of Targets (Duplicates Removed)
target
normal
neptune
satan
ipsweep
portsweep
smurf
nmap
back
teardrop
warezclient
pod
guess_passwd
buffer_overflow
warezmaster
land
imap
rootkit
loadmodule
ftp_write
multihop
phf
perl
spy
mscan
apache2
processtable
saint
snmpguess
mailbomb
snmpgetattack
httptunnel
named
ps
sendmail
xterm
xlock
xsnoop
sqlattack
udpstorm
worm
source
training
verification
10^{-6}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
relative amount

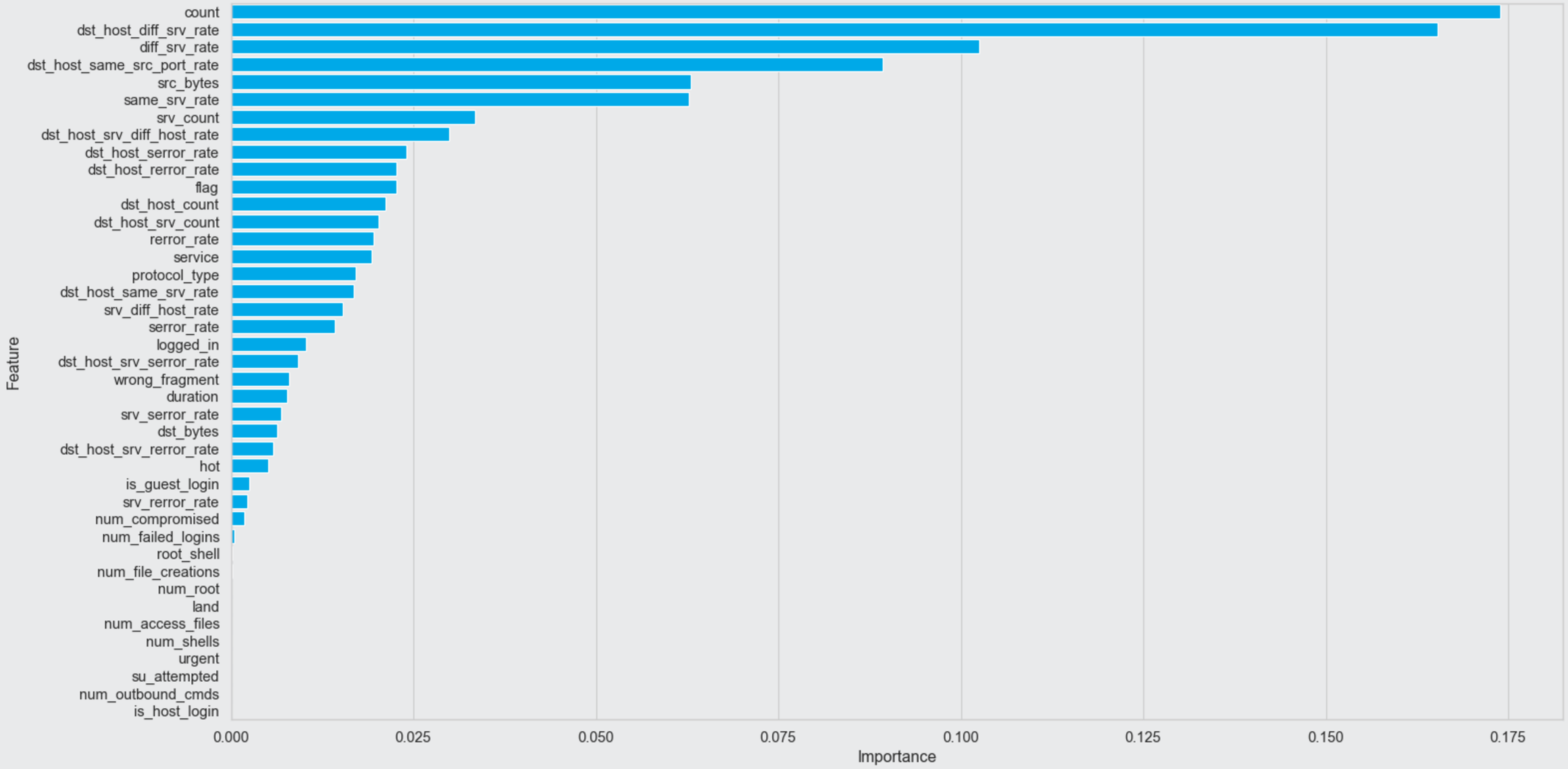
Target Group Distribution






Correlation With Attack Type
Correlation Coefficient
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
Attack Type
same_srv_rate
dst_host_same_srv_rate
flag
dst_host_srv_count
logged_in
target
dst_host_srv_diff_host_rate
dst_host_same_src_port_rate
srv_diff_host_rate
protocol_type
duration
hot
is_guest_login
num_access_files
num_failed_logins
root_shell
num_shells
num_file_creations
dst_host_diff_srv_rate
src_bytes
dst_bytes
su_attempted
num_root
urgent
num_compromised
is_host_login
land
srv_count
wrong_fragment
diff_srv_rate
rerror_rate
srv_rerror_rate
dst_host_srv_rerror_rate
dst_host_rerror_rate
dst_host_count
service
count
srv_serror_rate
dst_host_srv_serror_rate
serror_rate
dst_host_serror_rate
num_outbound_cmds
source
Features

Feature Importance Network Traffic



Machine Learning Strategies

Goal:

two models: one to *detect* an attack, another to *classify* the attack

Metric/ Evaluation:

attack detection: Fbeta score; number of false negative predictions

attack classification: Accuracy

verification on intended verification dataset exclusively (no separate train/ test splitting)

Considerations/ Strategy:

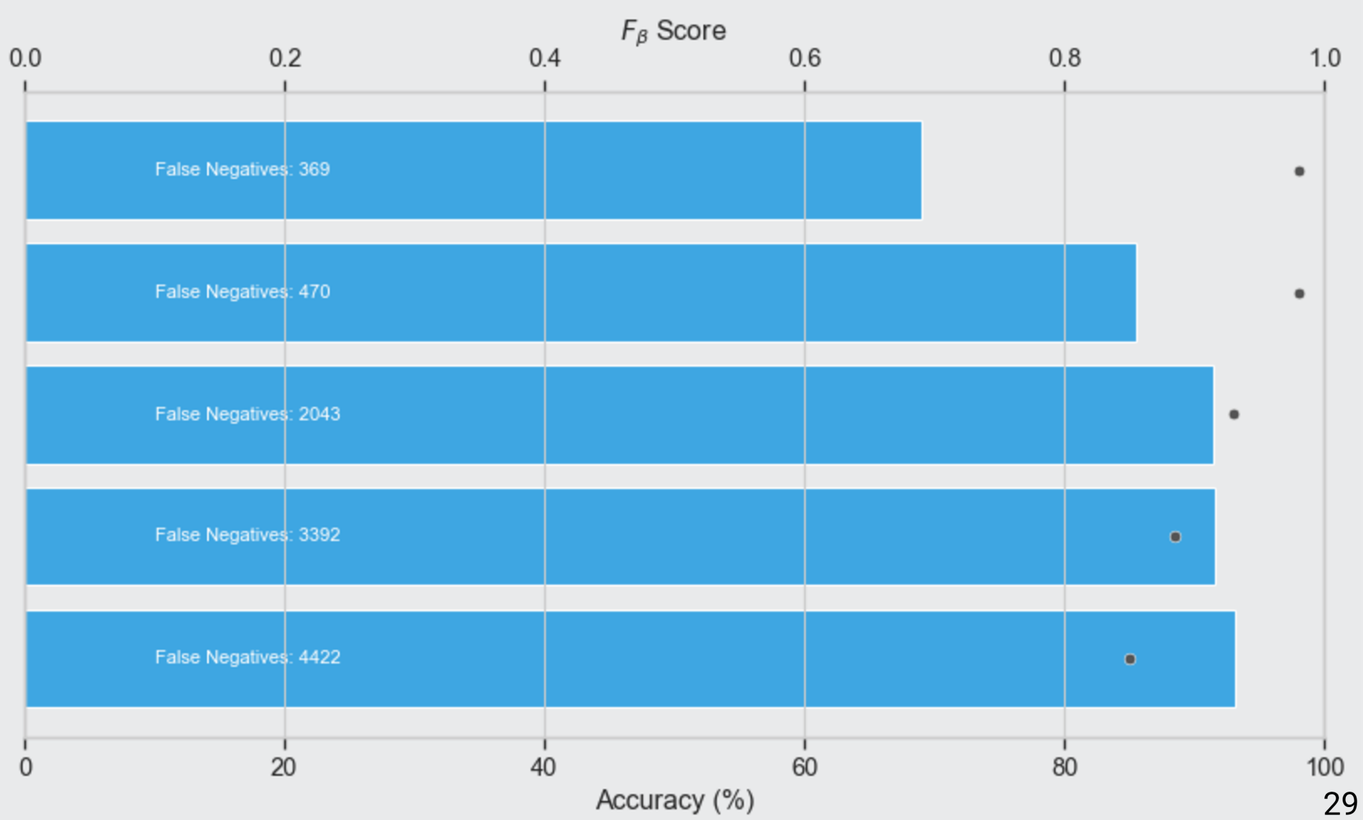
- dataset preparation
  - considerinig imbalance of learning data
  - selecting subset of features for higher efficiancy and accuracy
  - different subsets for both models
- find suitable candidates for the given task
- use hyperparameter boosting to increase scores
- combine best candidates into complete detection system
  - 1st layer: attack detection
  - 2nd layer: classification in cases where first layer fires up

Comparing Boosted Models

Attack Detection

☐ Full Feature Range ☑ Hyperparameter Tuning

Model	Accuracy	Fbeta Score	False Negatives
AdaBoost	69.021 %	0.979777	369 (1.256%)
Gradient Boost	85.557 %	0.980623	470 (1.600 %)
Random Forest	91.427 %	0.929662	2043 (6.954 %)
Logistic Regression	91.601 %	0.884627	3392 (11.546 %)
K Nearest Neighbours	93.122 %	0.850490	4422 (15.052 %)

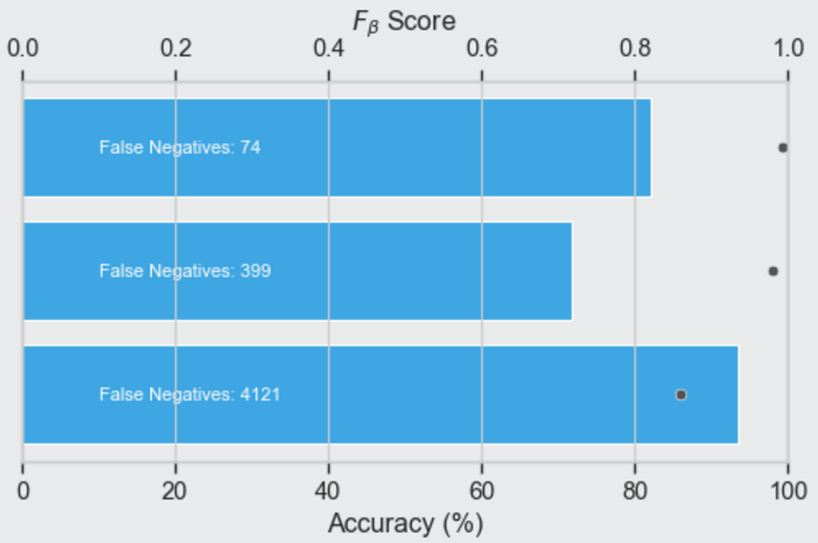


Best Results

Attack Detection

☑ Full Feature Range ☑ Hyperparameter Tuning

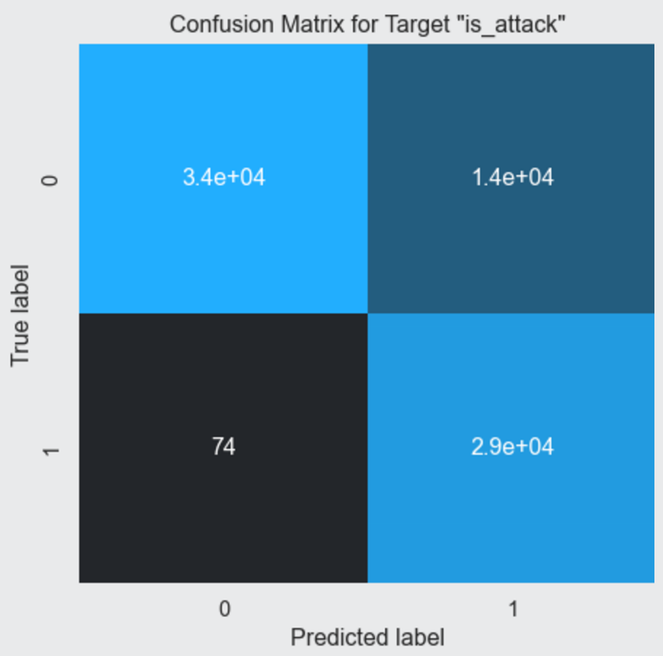
Model	Accuracy	Fbeta Score	False Negatives
Ada Boost	82.139 %	0.992911	74 (0.252 %)
Gradient Boost	71.690 %	0.979459	399 (1.358 %)
Neural Network	93.478 %	0.859734	4121 (14.028 %)



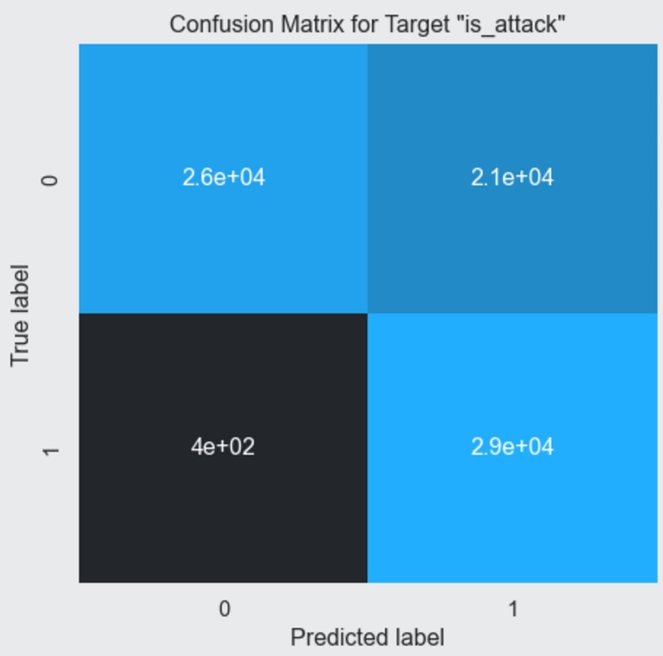
Confusion Matrices

Attack Detection

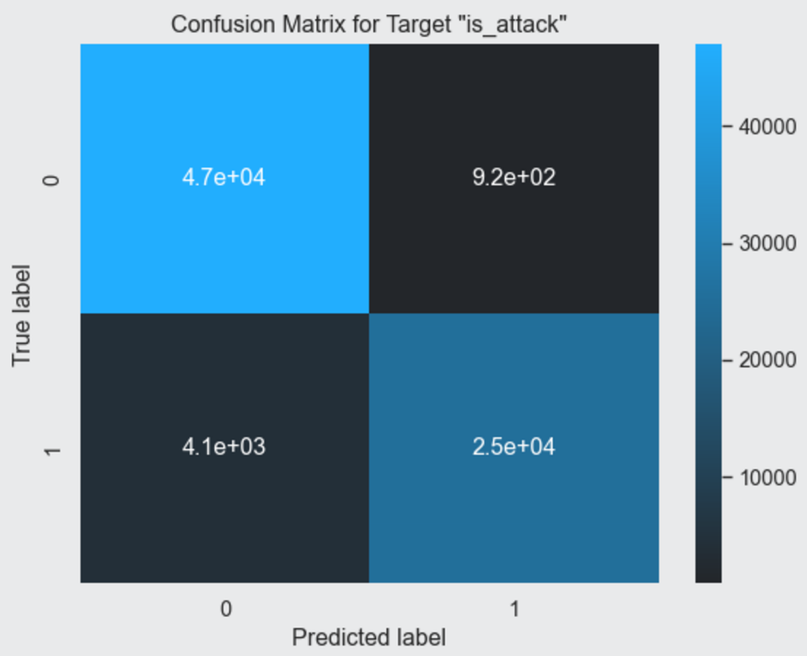
Ada Boost (82.139 %)



Gradient Boost (71.690 %)



Neural Network (93.478 %)



0 - normal

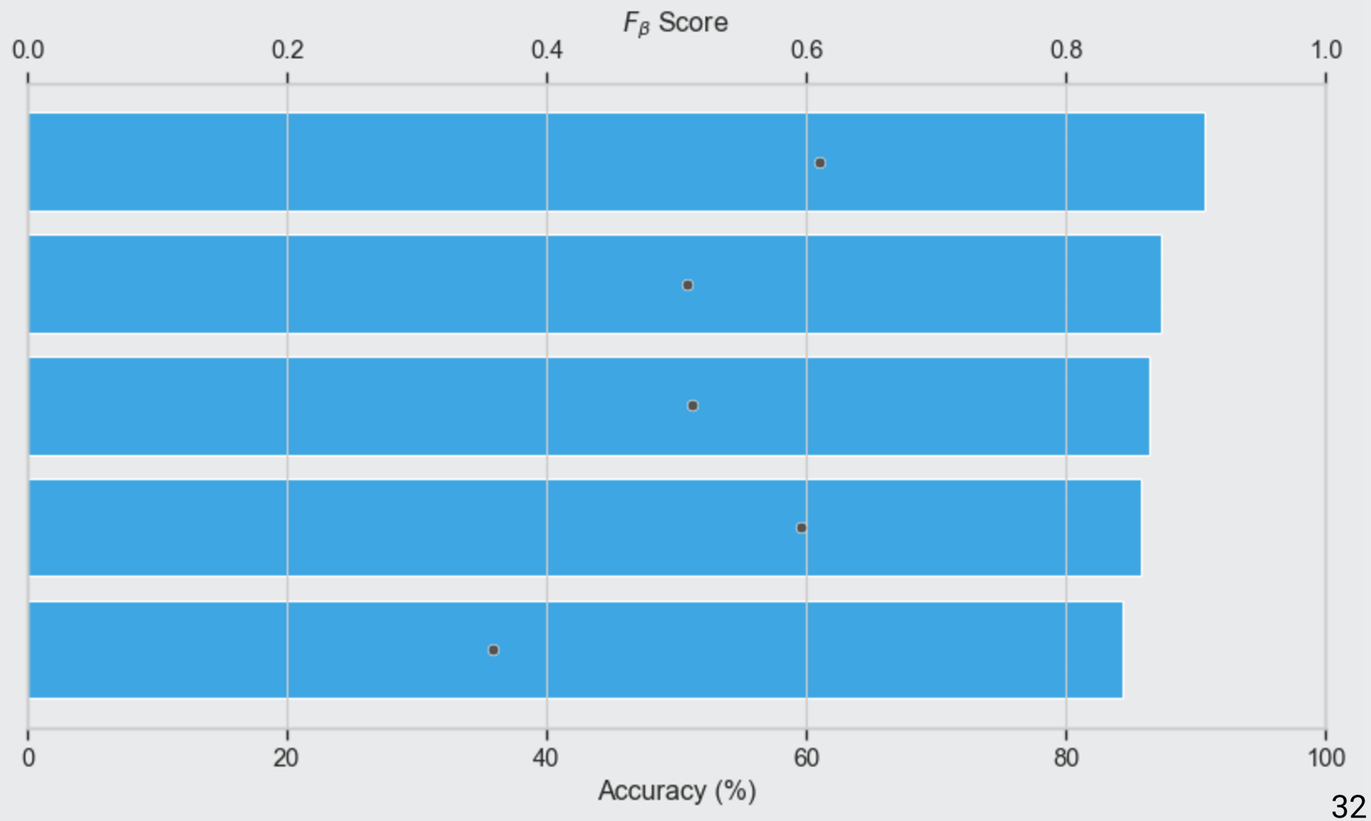
1 - attack

Comparing Boosted Models

Attack Classification

☐ Full Feature Range ☑ Hyperparameter Tuning

Model	Accuracy	Fbeta Score
Random Forest	90.677 %	0.610170
AdaBoost	87.392 %	0.507990
Gradient Boost	86.487 %	0.511909
K Nearest Neighbours	83.797	0.521141
Logistic Regression	84.325 %	0.358221

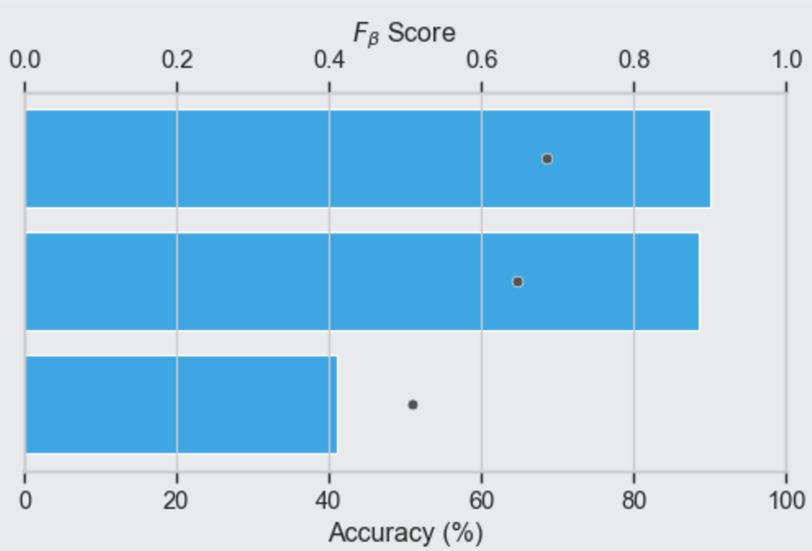


Best Results

Attack Classification

☑ Full Feature Range ☑ Hyperparameter Tuning

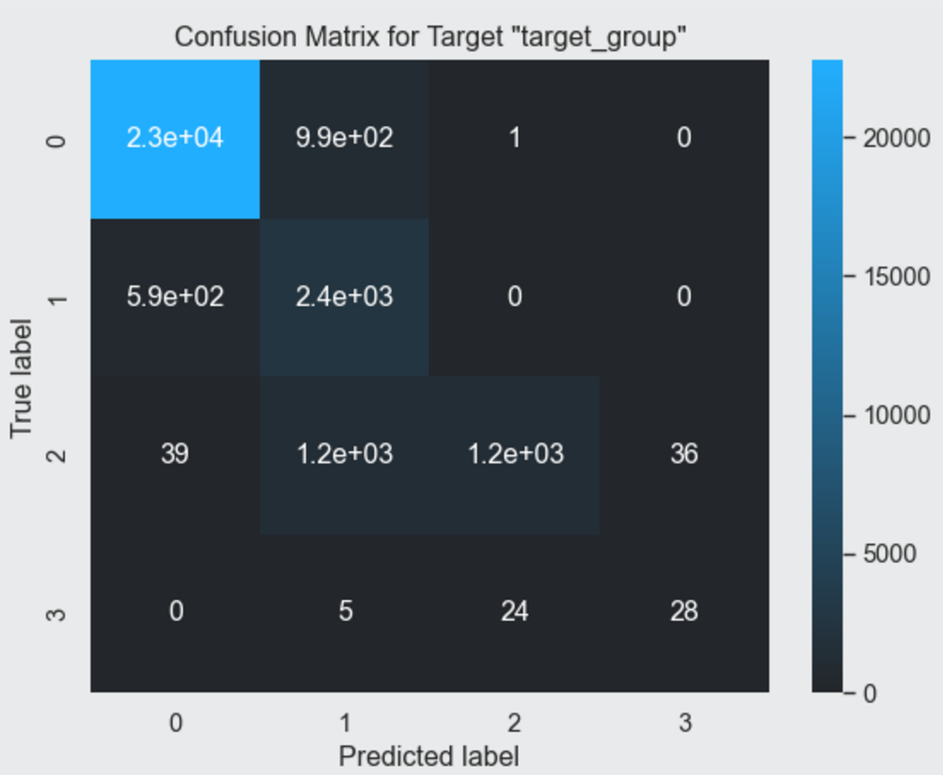
Model	Accuracy	Fbeta Score
Random Forest	90.112 %	0.685390
Gradient Boosting	88.505 %	0.646515
Neural Network	41.112 %	0.509681



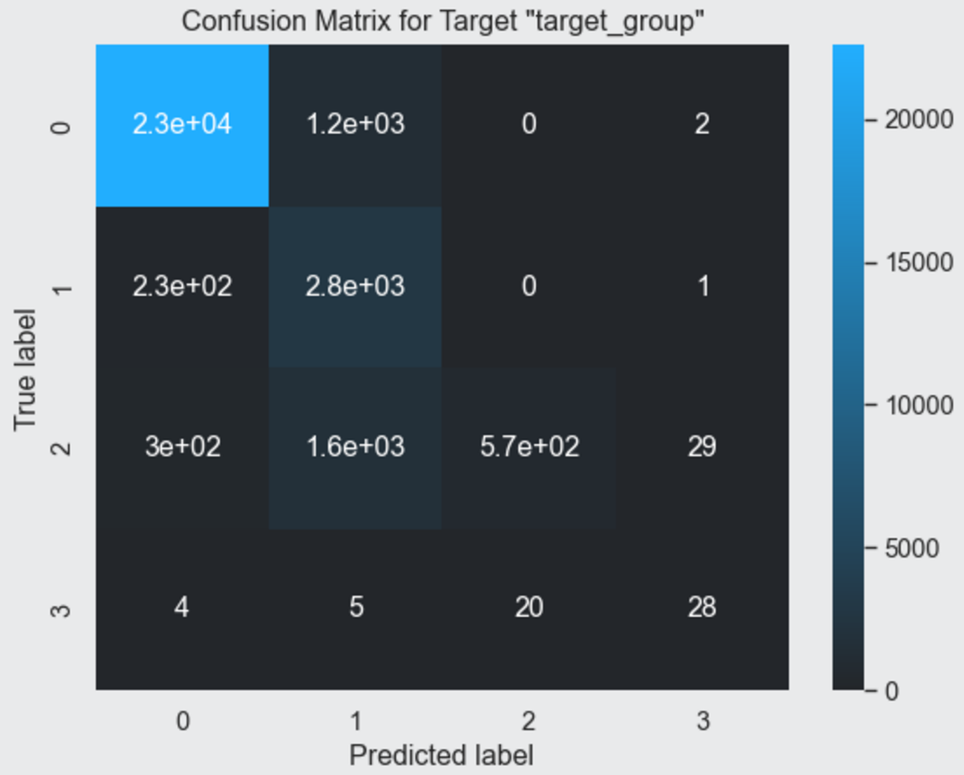
Confusion Matrices

Attack Classification

Random Forest (90.112 %)



Gradient Boost (88.505%)



0 - DOS
1 - Probe
2 - R2L
3 - U2R

SecureSphere Innovations
We maintain privacy.

Conclusion

Attack Detection

Simple models (Ada and Gradient Boost) perform well for **avoiding attacks**, but raise significant amount of **false alarms**

Neural Network more sensitive regarding false alarms (overall **highest accuracy**), but has long learning time and comparably **poor attack prevention capabilities**

SecureSphere Innovations
We maintain privacy.

Conclusion

Attack Classification

Imbalance of class distributions and new attacks in verification dataset **challenge all** models

Groups “R2L” (3,532 records) and “U2R ”(109 records) are **critically underrepresented**; either **more data** on those need to be provided, or their detection should be **removed altogether** from the classification models’ targets

Alternatively, similar to creating attack detection as **binary problem**, those two cases could also be treated individually in a **multi-layered** defence-and-detection system

SecureSphere Innovations
We maintain privacy.

Conclusion

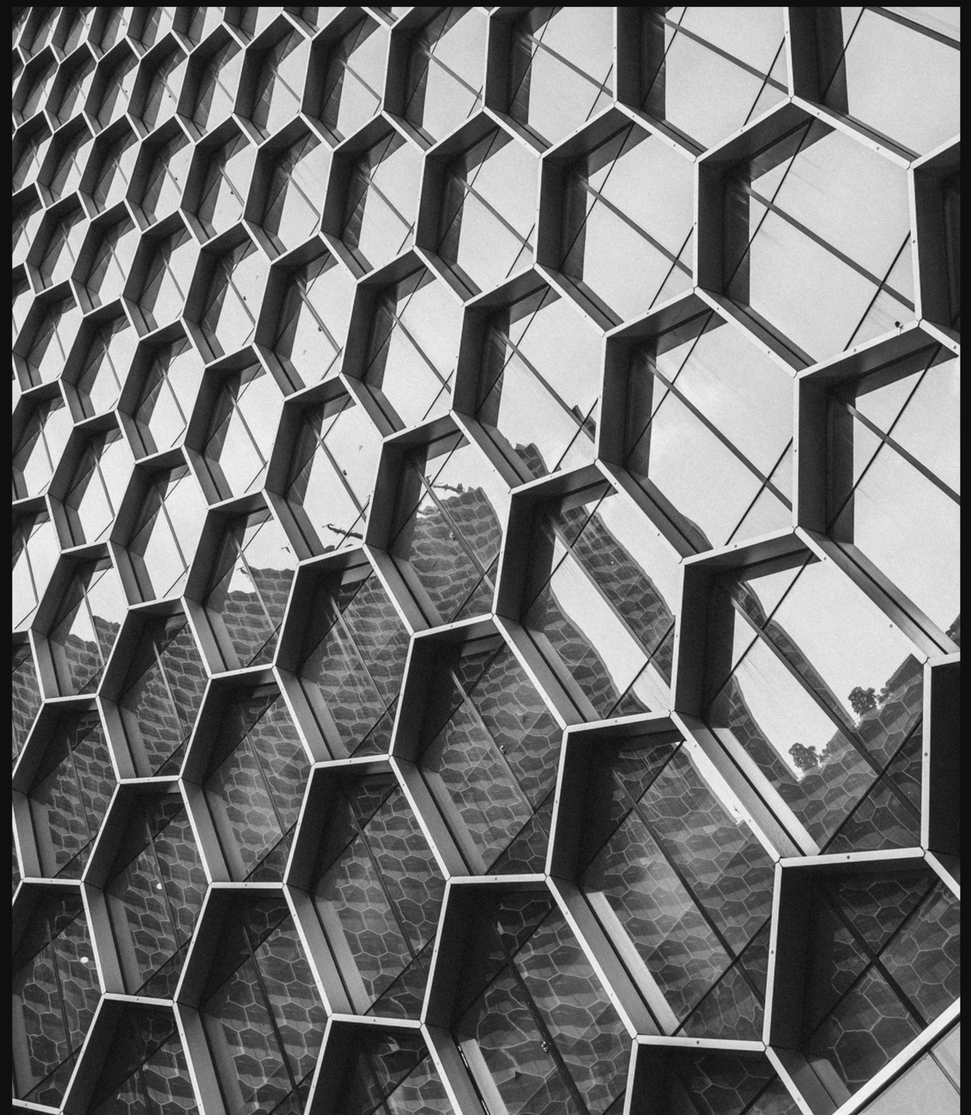
Advice for Greencare Pharmacy

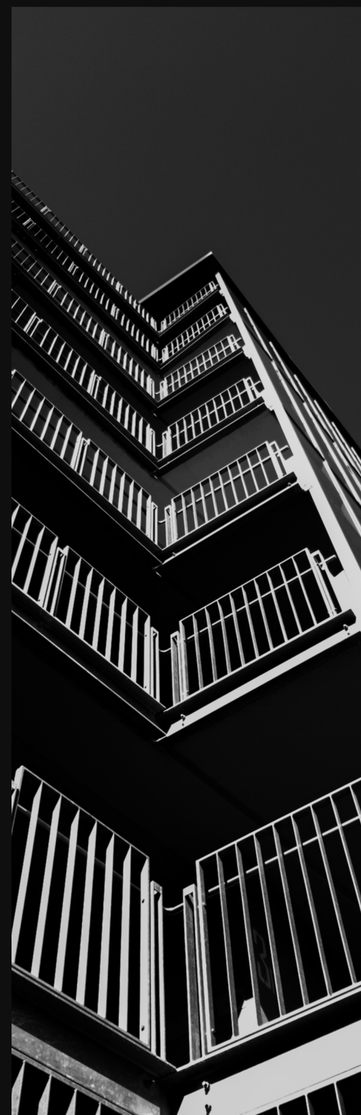
As cyber threats develop like **viruses**, defence mechanisms need to be trained, improved and verified **continuously**

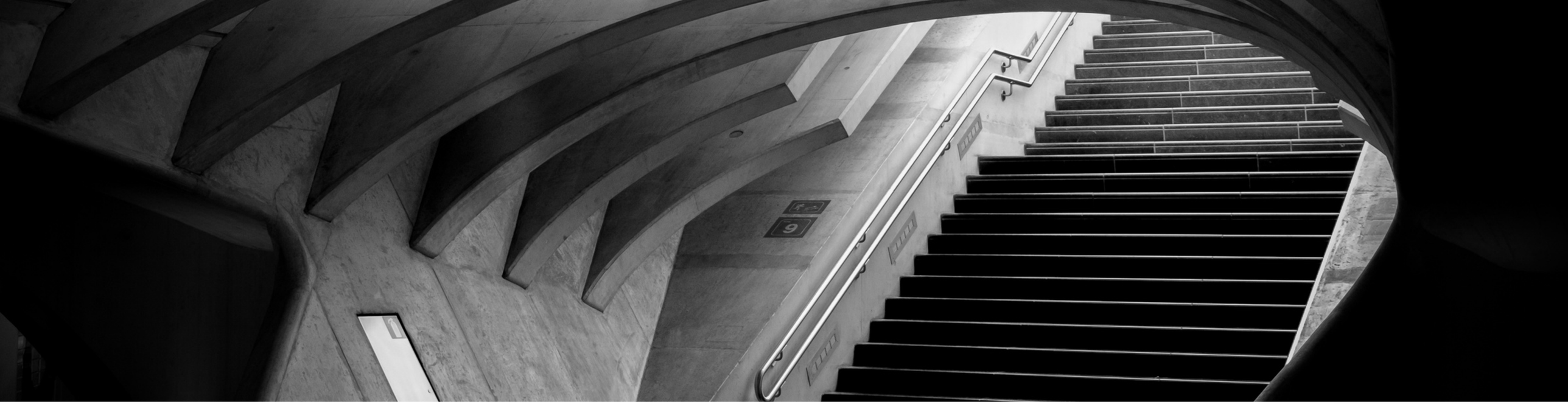
For GP’s current problem, a **complex defence architecture** is mandatory

A “fast” solution could be offered which may **boost security**, but will inevitably inhibit their network’s communication abilities and pose a **risk** to its **functionality** or the company’s internal communication/ processes

Thanks for your attention.







Connect with us.

Dustin Lischke
dustin@lischke-online.de

Nilgün Yesil
n.yesil86@gmx.com

Tanvi Goel
dreamz.tanvi@gmail.com

Stefanie Reimers
stefanie.studium@gmail.com

One klick to Github

SecureSphere Innovations
We maintain privacy.

The Team

We teamed up to protect your data.

Dustin Lischke

Data Scientist

Thuringia
Germany
International

Nilgün Yesil

Data Scientist

NRW
Germany (Remote)
International (Remote)

Tanvi Goel

Data Analyst

Munich
Germany (Remote)
International (Remote)

Stefanie Reimers

Data Analyst

Berlin
Germany