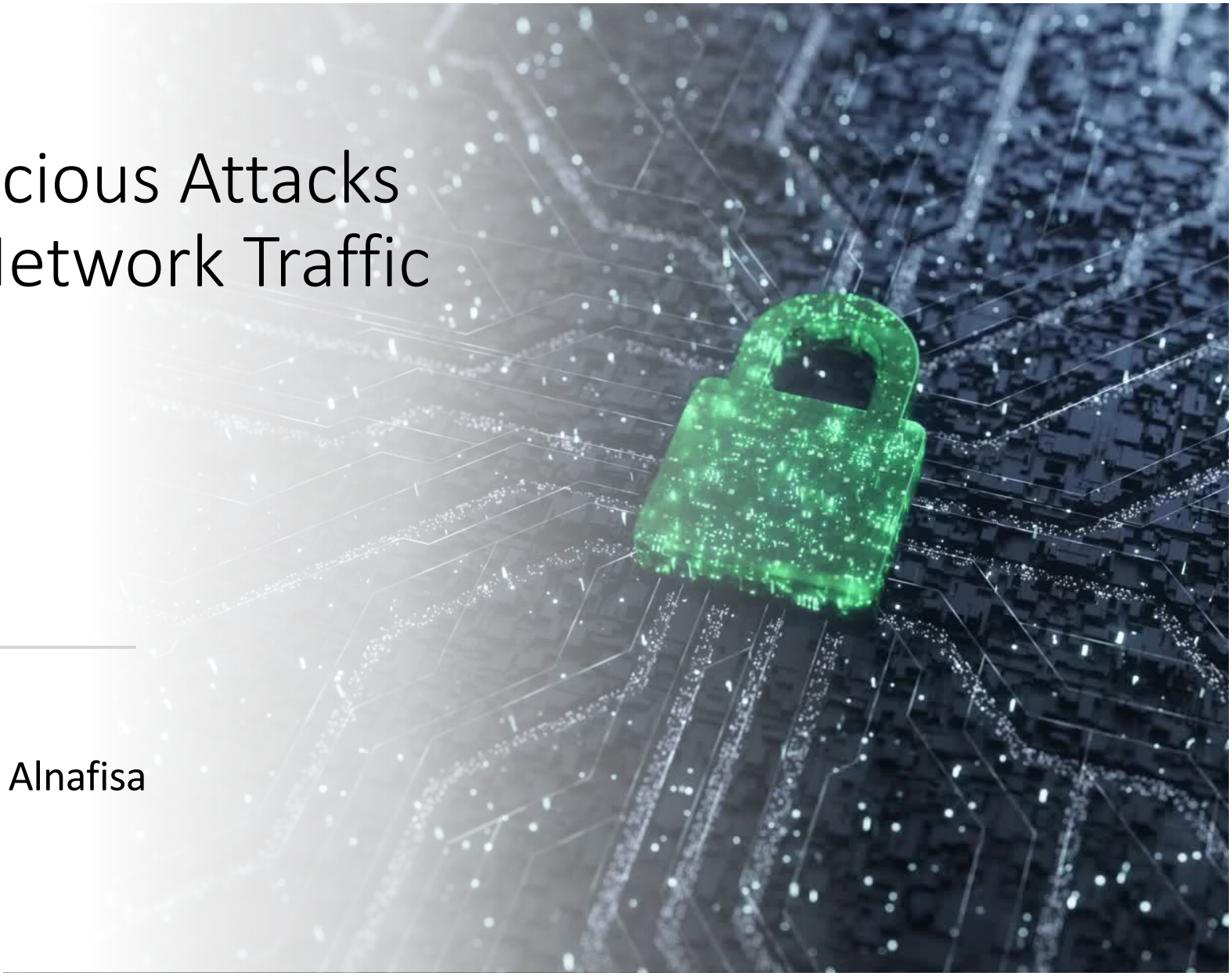
Detecting Malicious Attacks Through The Network Traffic



SDAIA BootCamp T5

Presented by:

Hatim Alshehri & Fahad Alnafisa

Methodolegy

- Problem Statement
- Data Analysis
- Data Pre-processing and feature engineering
- Classification models
- Conclusion

Problem Statement

With the rise of Internet usage, it is very important to protect Networks. The most common risk to a network's security is an intrusion such as brute force, denial of service or even an infiltration from within a network.

Problem Statement

Detecting Intrusion

network

Problem Statement

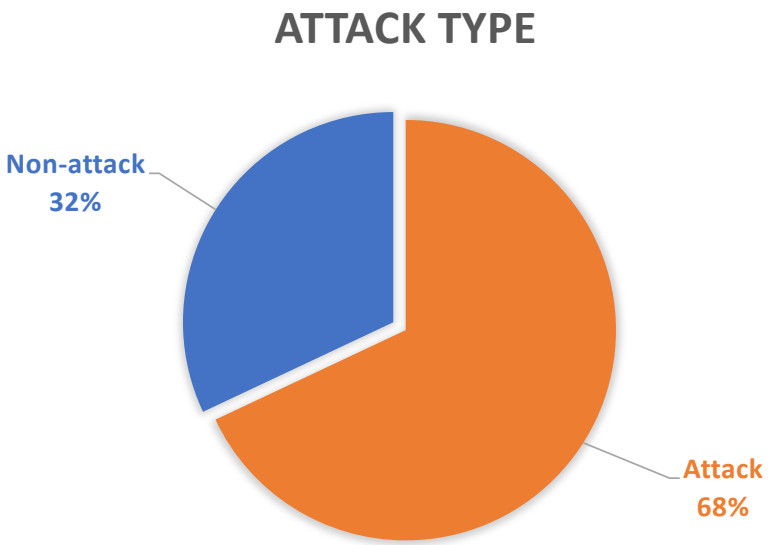
Data set

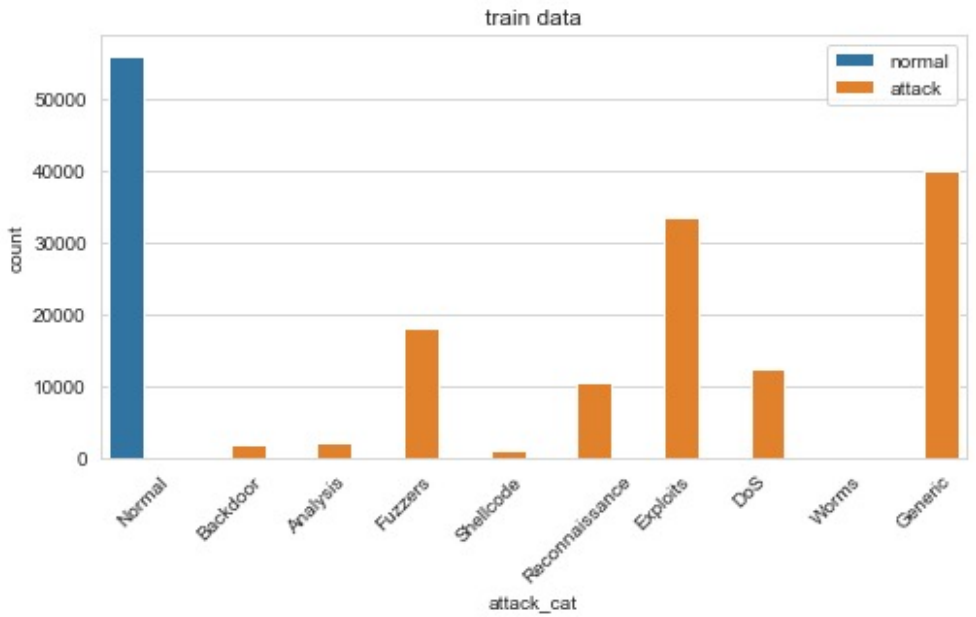
- UNSW-NB15
- UNSW Sydney
- +170k rows
- 45 columns

Tools

- Pandas
- Numpy
- Matplotlib
- Seaborn
- Sklearn
- XGBoost
- Pickle
- Prettytable

Data analysis





Data Analysis

Data Pre-processing and feature engineering

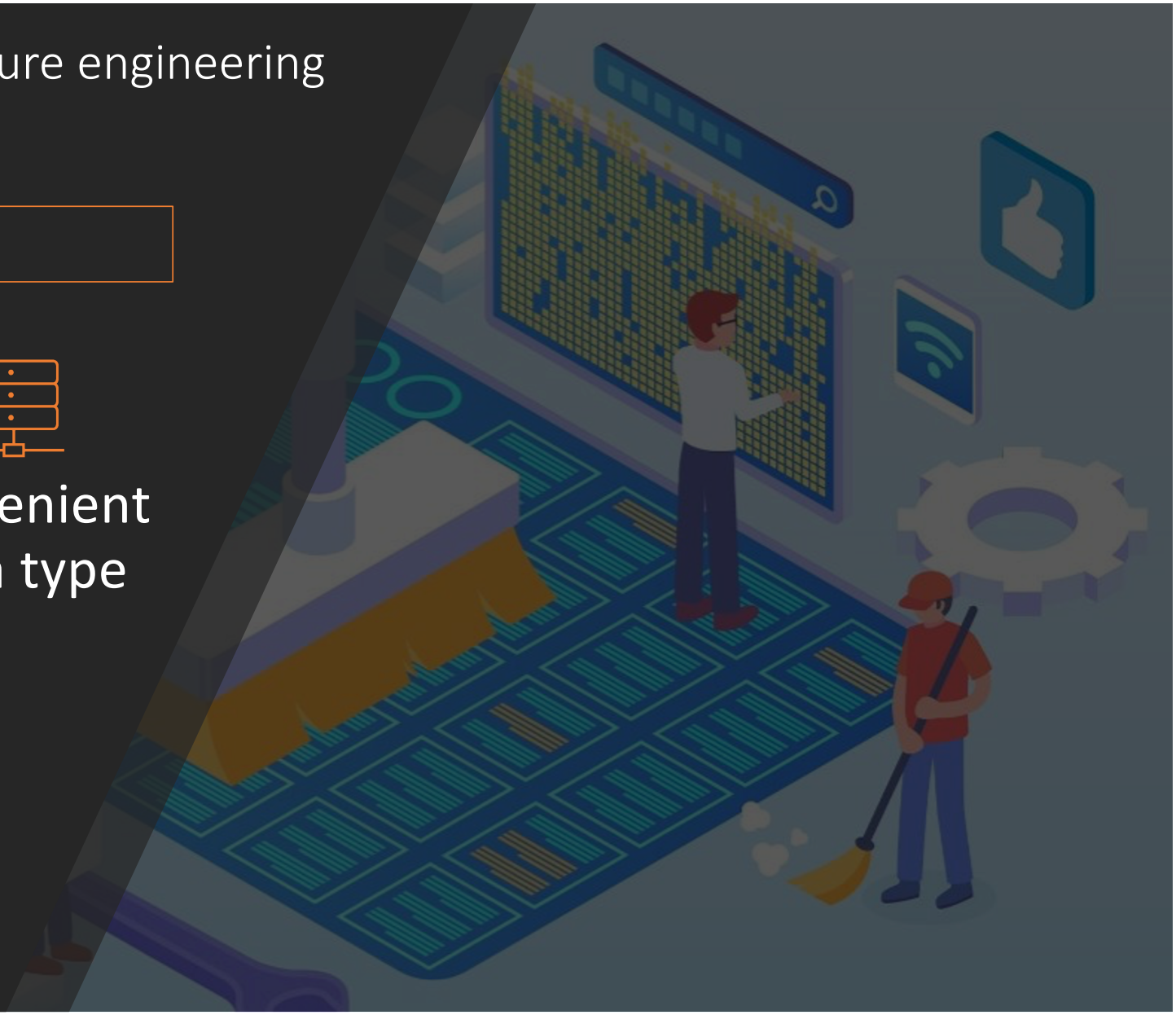
Data Pre-processing and feature engineering

Cleaning

Remove nan-values

Convenient data type

Data reformation



Data Pre-processing and feature engineering

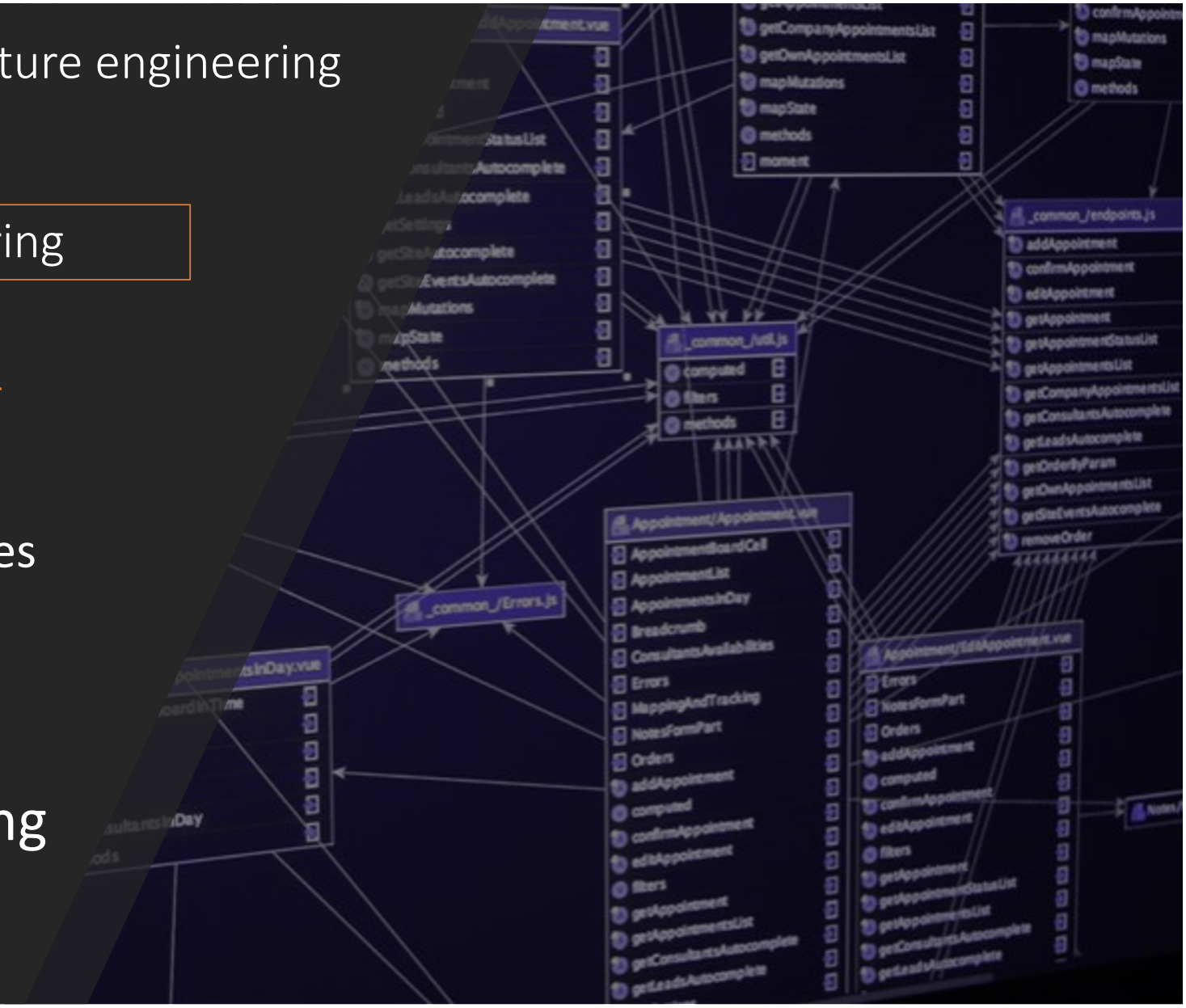
Feature Engineering

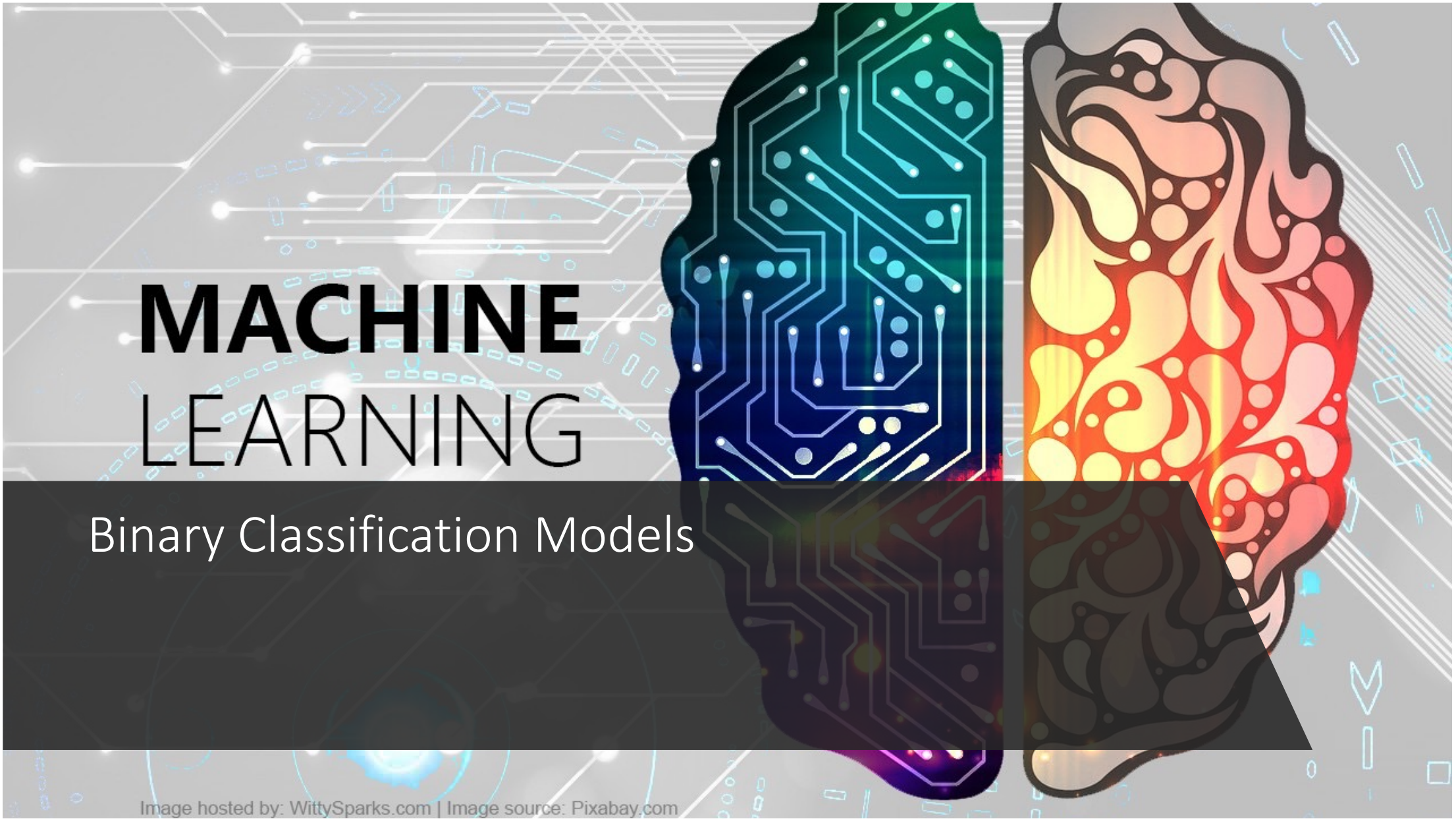
High correlation

New Features

Scaling

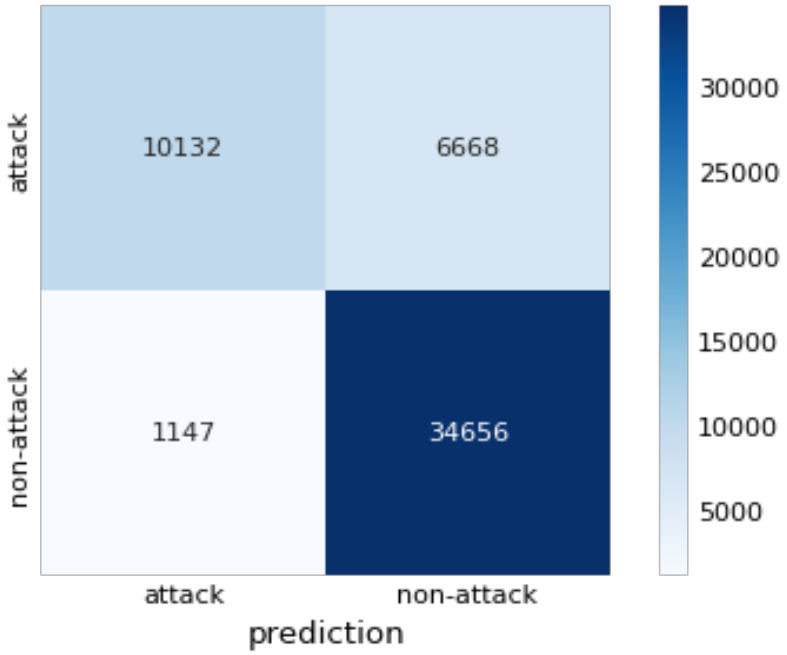
Encoding




MACHINE
LEARNING
Binary Classification Models
Image hosted by: WittySparks.com | Image source: Pixabay.com

Baseline Model

Logistic Regression



Accuracy		Precision		Recall	
Train	**Validation**	**Train**	**Validation**	**Train**	**Validation**
0.7865	**0.7855**	**0.8395**	**0.8386**	**0.9669**	**0.9679**

MODELS RESULT

	ACCURACY		F1-SCORE		FALSE ALARM RATE	
	TRAIN	VALIDATION	TRAIN	VALIDATION	TRAIN	VALIDATION
MODEL						
LOGISTIC REGRESSION	0.9052	0.9021	0.9544	0.9534	0.0947	0.0978
K-NEATEST NEIGHBORS	0.9173	0.9100	0.9591	0.9554	0.0826	0.0899
DECISION TREE	0.8926	0.8882	0.9517	0.9500	0.1073	0.1117
RANDOM FOREST	0.9143	0.9028	0.9612	0.9560	0.0856	0.0971
XGBOOST	0.9669	0.9391	0.9819	0.9665	0.0330	0.0608

ENSEMBELING

MODELS

- LOGISTIC REGRESSION
- RANDOM FOREST
- K-NEAREST NEIGHBORS
- EXTRA TREE

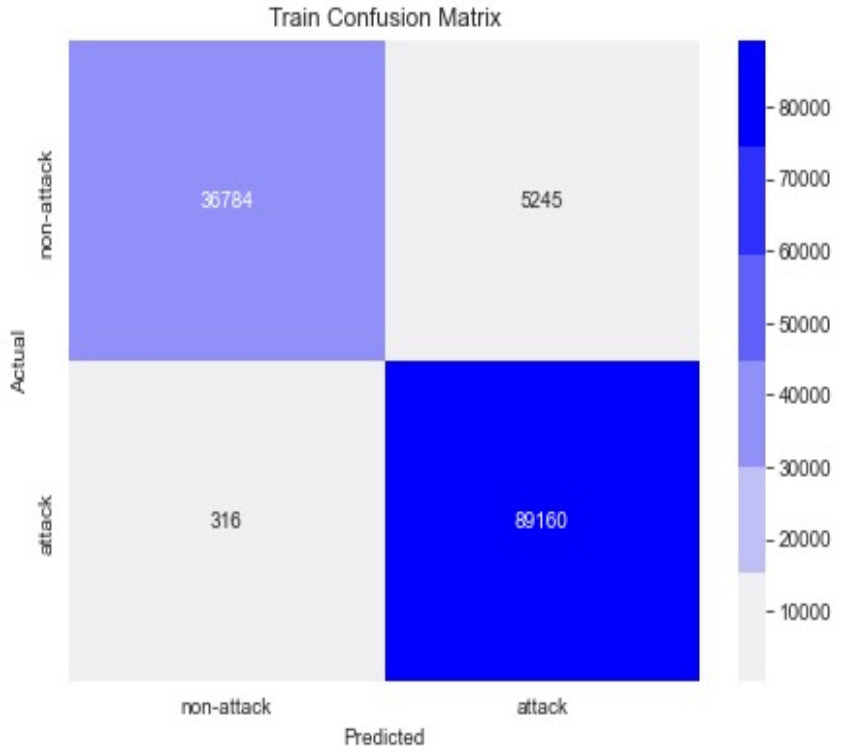
	ACCURACY		F1-SCORE		FALSE ALARM RATE	
MODEL	TRAIN	VALIDATION	TRAIN	VALIDATION	TRAIN	VALIDATION
VOTING CLASSIFIER (HARD)	0.9990	0.9500	0.9992	0.9641	0.0008	0.0656
VOTING CLASSIFIER (SOFT)	0.9990	0.9489	0.9992	0.9637	0.0014	0.0729
STACKING	0.9990	0.9502	0.9992	0.9644	0.0008	0.0664

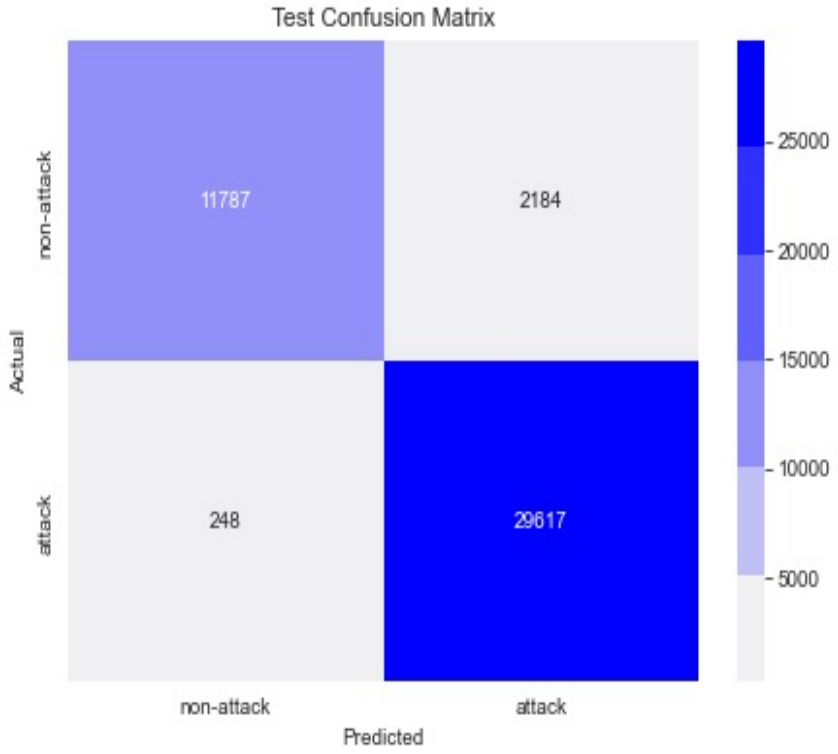
BEST CLASSIFICATION MODEL

VOTING CLASSIFIER (HARD)

ACCURACY		F1		PRECISION		RECALL	
TRAIN	TEST	TRAIN	TEST	TRAIN	TEST	TRAIN	TEST
0.9559	0.9431	0.9685	0.9596	0.9420	0.9291	0.9965	0.9921

VOTING CLASSIFIER (HARD)






Thank You