Dataset:

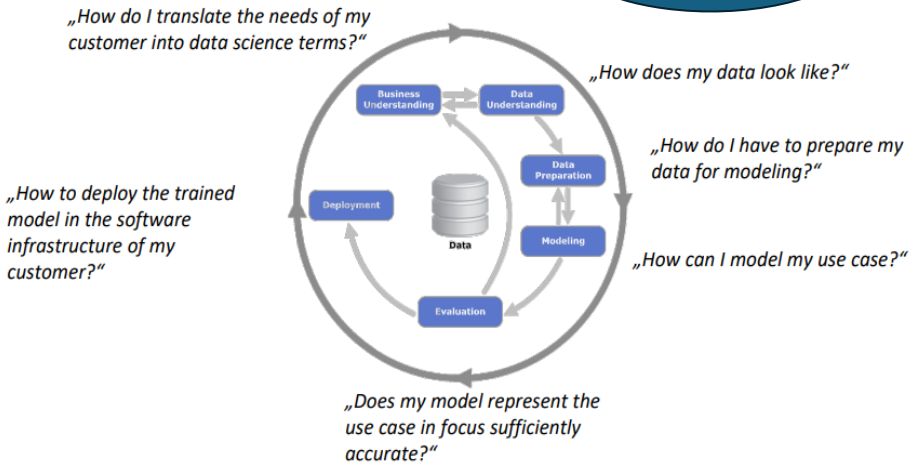
Mehr als 80 Features. (CICFlowmeter V3)

CSE-CIC-IDS2018 and Edge-IIoTset

63 Features. (Zeek Tool und TsharkTool)

Commented [HM1]: Durch AWS Client. **Funktioniert nicht

Commented [HM2]: # 0 - Indicates normal | 1 - attack
Zeek Tool and Tshark Tool to extrapolate Features from PCAP Files.



Methods:(Model and Feature Reduction)

KNN (K Nearest Neighbors)

Enhanced KNN

LOF (Local Outler Factor)

DT (Decision Tree)

SVM

XG BoosT

Random Forest

PCA

Fischer Scoring Algorithm

Pearson Correlation Test??

Recursive Feature Elimination??

Library:

PyOD ->Uses *Semi-supervised Novelty Detection* and *Unsupervised Detection Methods*

- ADBench_Algorithms.pdf (Page 20 – 22) has all the Algorithms used to detect the Anomalies.
- Contains 50 detection Algorithms. (Zhao, 2019)
- Can be classified further to Multivariate Data (20), Time series Outlier Detection (10) and Graph Outlier Detection Algorithms (10).

Numpy, pandas, Seaborn, scikit learn.

Types of Attacks

Dataset	Attacks
Edge IIoTset	DoS/DDoS attacks, Information gathering, Man in the middle attacks, Injection attacks, and Malware attacks – *7 Attacks*
CICIDS 2018	Brute-force, Heartbleed, Botnet, DoS, DDoS, Web attacks, and infiltration of the network from inside – *5 Attacks*
CICIDS 2017	DDOS (1), Botnet (2), Infiltration (3), Parator (4), Heartbleed (5), Portscan (6), Webattacks (7)

Results:

- Dataset: CICIDS 2017 (trained only with normal samples) (Feature reduction Technique is PCA)
  a. Accuracy: LOF better with 1,2,3,4,5 and similarity KNN better with 6,7. Simple KNN fell behind in most of the attacks.
  b. Detection Rates: LOF und similarity KNN had high detection Rates. LOF has potential to be more.

- Dataset: CICIDS 2017 (trained with -) (Feature Reduction – Fisher Score Algorithm) (Models: KNN, SVM, DT)

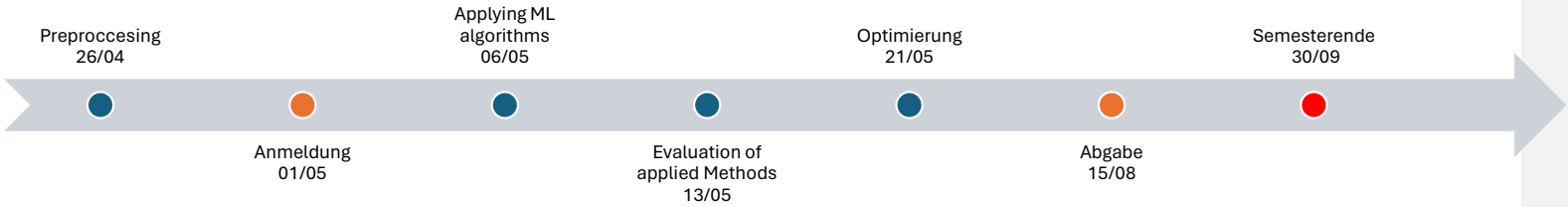
Commented [HM3]: High Detection Rate comes at a stake of accuracy.
Accuracy is not same as Precision.
Accuracy != Precision != Sensitivity(Recall)

a. Accuracy: KNN>DT>>SVM (Although Model is designed only for detecting DDOS attacks)

- Dataset: CICIDS 2018 (trained with normal and attack) (Feature Reduction –) (Models: DT, SVM, RF, KNN)
  a. Accuracy: 2-Klass: SVM=KNN=RF=DT
  b. Accuracy: 6-Klass: SVM>KNN>RF>DT
  c. Accuracy: 15-Klass: RF>KNN>SVM>>DT

- Read: If trained with only Normal Data samples (not together with attack Data Samples), The Model has better Potential to detect “Zero Day Attacks”
- ADBench: with merely 1% labeled anomalies, most semi-supervised methods can outperform the best unsupervised method, justifying the importance of supervision.

Timeline:



Hier habe ich großzügig die Zeit für die Bachelorarbeit eingeplant. Der Vorgang wird zunächst mit dem EDGE-Datensatz durchgeführt. Sobald ich mit den Funktionen/Anwendungen vertraut bin, möchte ich den gleichen Vorgang auch mit dem CICIDS-Datensatz durchführen.

Laut APO habe ich 5 Monate Zeit, der Bachelorarbeit fertigzustellen. Daher möchte ich die Anmeldung in der ersten woche der Mai erledigen, damit ich bis zum Ende dieses Semesters Zeit habe.

Der Abgabe habe ich im august geplant, damit haben Sie Zeit zu korrigieren und mich zu präsentieren.