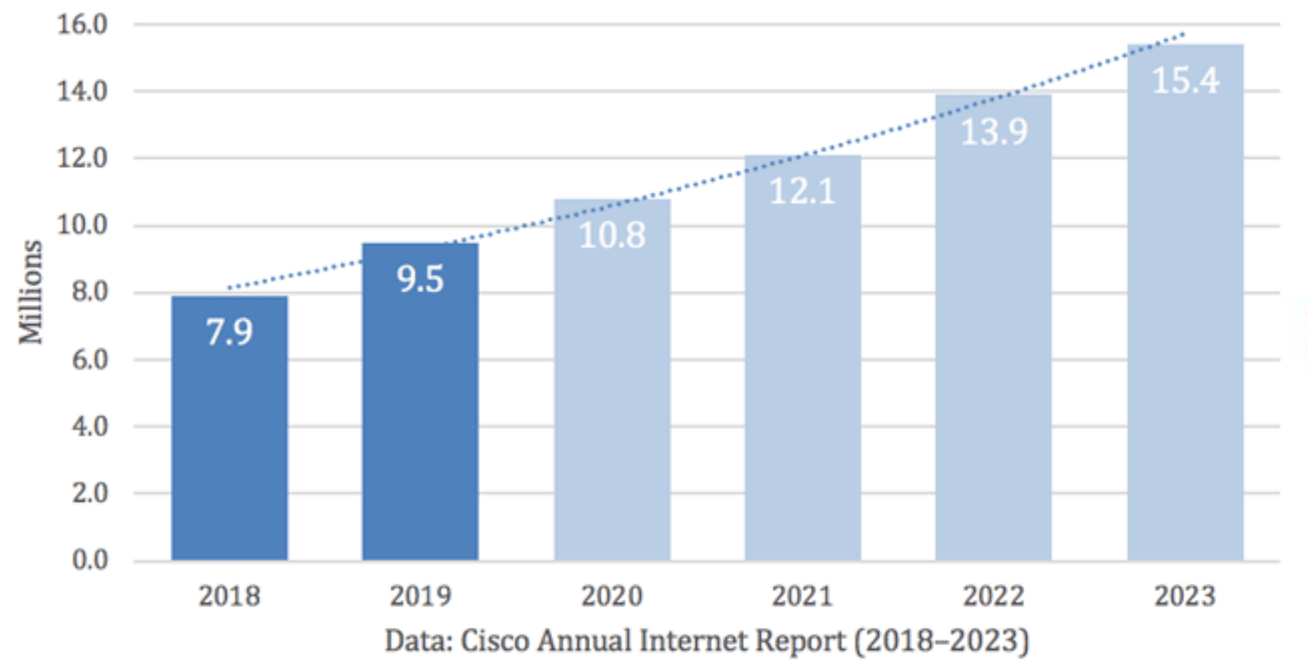
Evaluation and Development of Machine Learning based Algorithms for Predicting Distributed Denial of Service (DDoS) Attack

Qozeem Adeshina
Concordia University of Edmonton

Background

The first known DDoS occurred in 1996 when Panix, one of the oldest internet service providers, was knocked offline for several days by a SYN flood.



Data: Cisco Annual Internet Report (2018–2023)

Cisco predicts that DDoS attacks will double from the 7.9 million in 2018 to 15 million by 2023.

The Google DDoS Attack, 2017

Google's infrastructure absorbed a 2.5Tbps DDoS attack in 2017, the largest such attack in terms of its sheer volume.

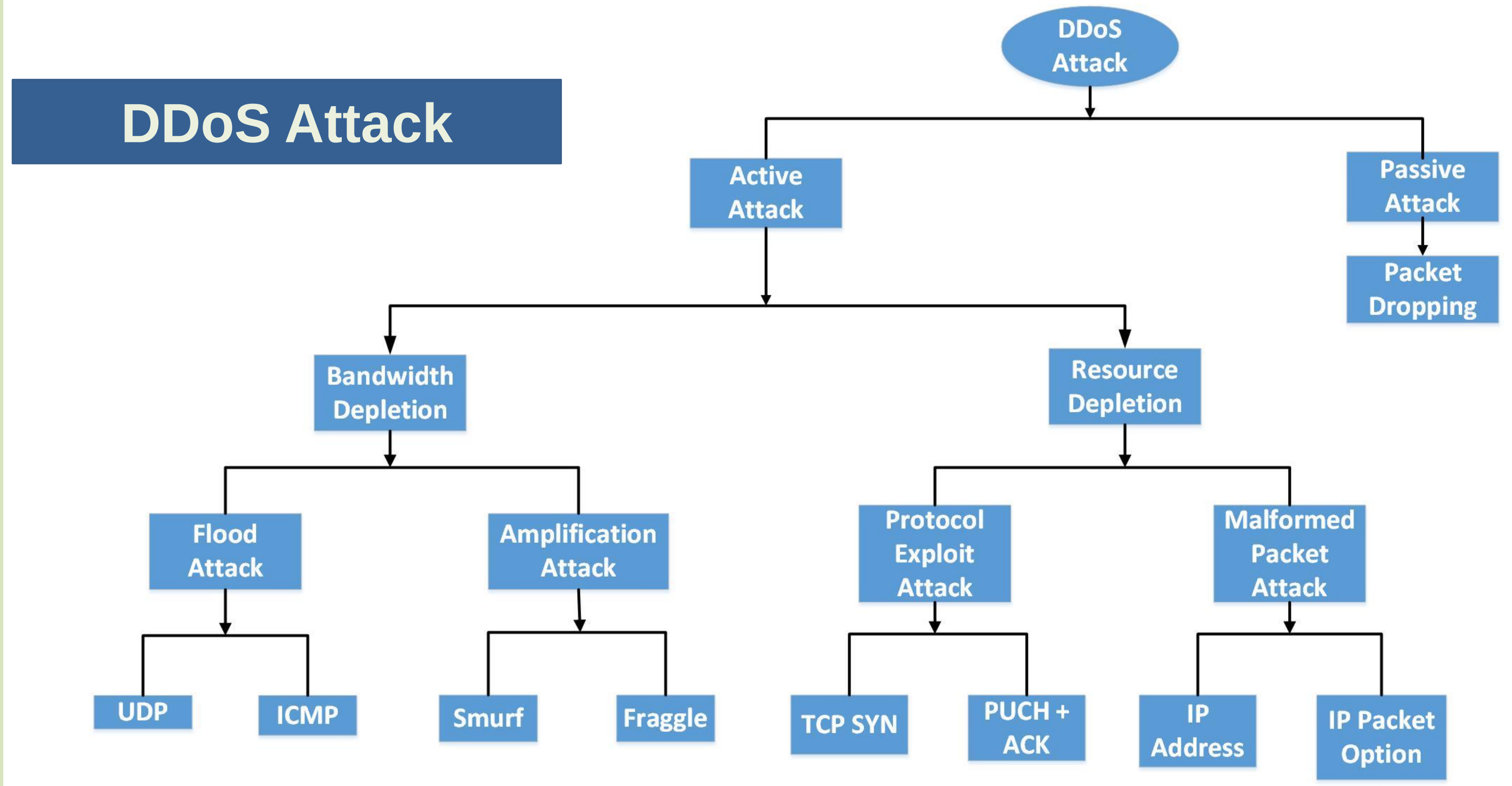
The AWS DDoS Attack, 2020

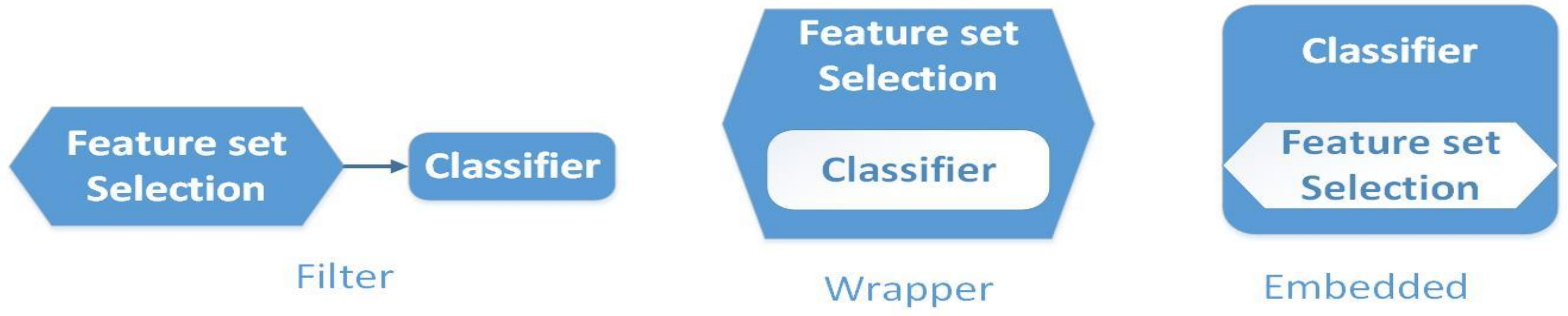
AWS was hit by a gigantic DDoS attack in February 2020 - was lasted for three days and peaked at 2.3 terabytes/ second.

Six US Bank Attack, 2012

On March 12, 2012, six U.S. banks were targeted by DDoS attacks - were carried out by Brobot with each attack generating over 60 gigabits of DDoS attack traffic per second.

DDoS Attack

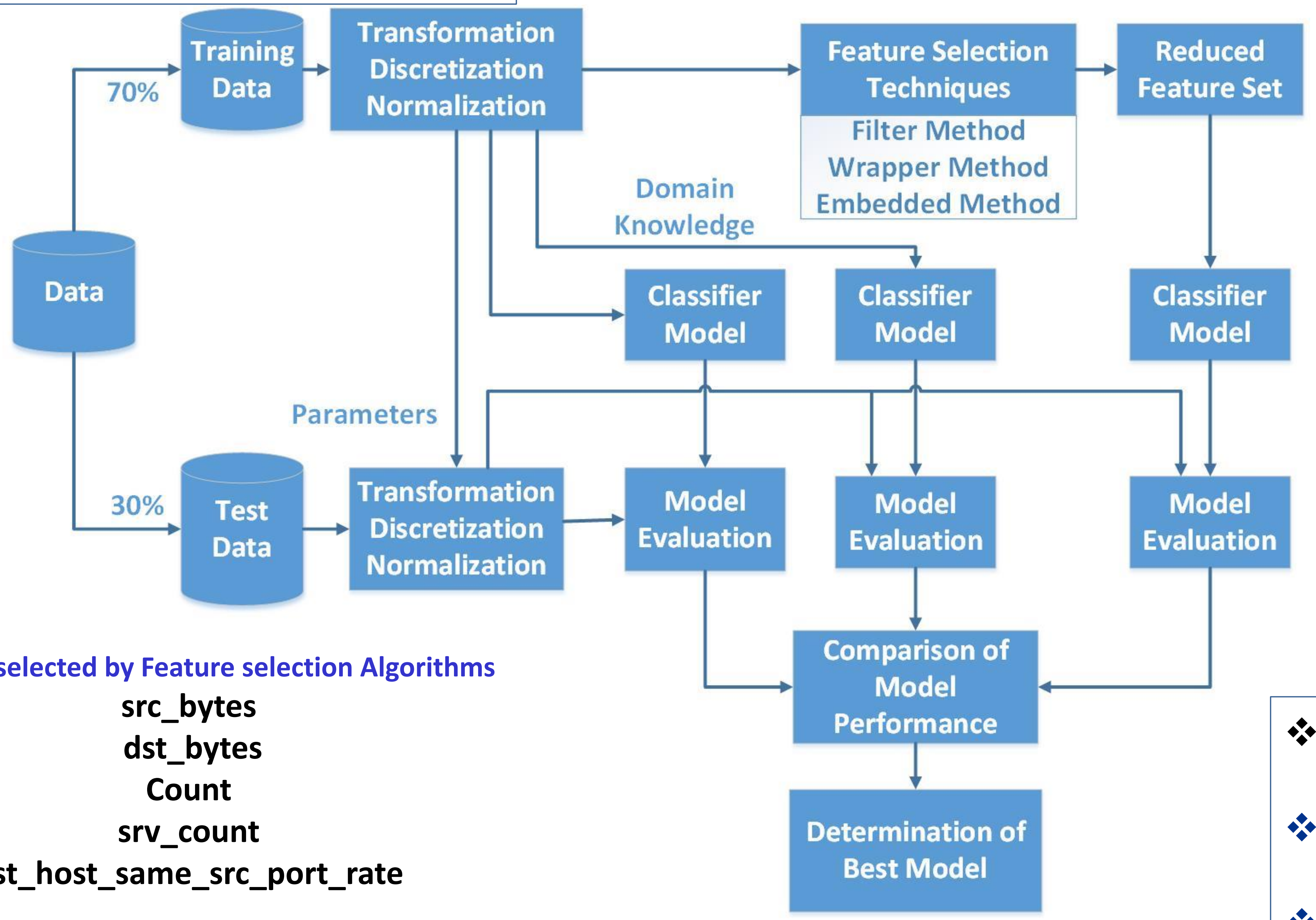




Variance Threshold
Mutual Information
Anova f-test
Select KBest
Pearson Correlation

Recursive Feature Elimination (RFE)

Logistic Regression
Support Vector Machine
Random Forest



Features selected by Feature selection Algorithms

src_bytes
dst_bytes
Count
srv_count
dst_host_same_src_port_rate

Flowchart of ML based Algorithms for Predicting DDoS Attack

Feature Types	Feature Names	Number of Features
Basic features	duration, protocol_type, etc.	9
Content features	logged_in, num_root, etc.	13
Traffic features	count, srv_count, etc.	9

Performance Evaluation

Classifier	Train Score	Test Score	Train Time (seconds)
Random Forest	0.99	0.99	19.34
Decision Tree	1.00	0.98	0.64
Gradient Boosting	0.98	0.98	133.36
Nearest Neighbor	0.98	0.97	16.91
Logistic Regression	0.83	0.83	33.65
Linear SVM	0.82	0.82	3154.66
Ada Boost	0.77	0.77	9.61
Multinomial Naïve Bayes	0.67	0.67	0.07
Naïve Bayes	0.65	0.65	0.15
Neural Net	0.53	0.53	146.72

Conclusions and Future Works

- ❖ **Gradient Boosting**, **Random Forest**, **Nearest Neighbors** and **Decision Tree** are recommended for DDoS attack prediction (consistent accuracy over 95%).
- ❖ SVM has proven to be the slowest when classifying DDoS and normal traffics. It is therefore not recommended during DDoS traffic classification when speed is paramount.
- ❖We would like to explore deep learning Algorithms and incorporate domain knowledge for DDoS prediction

Contact

Qozeem Adeshina
MSc.IT, Concordia University of Edmonton
Email: qadeshin@student.concordia.ab.ca

References

1. S. Pande et al. (2021) DDOS Detection Using Machine Learning Technique. Recent Studies on Computational Intelligence.
2. P. S. Saini et al. (2020) Detection of DDoS Attacks using Machine Learning Algorithms. International Conference on Computing for Sustainable Global Development
3.R. Doshi, N. Apthorpe and N. Feamster. (2018) Machine Learning DDoS Detection for Consumer Internet of Things Devices. IEEE Security and Privacy Workshops.