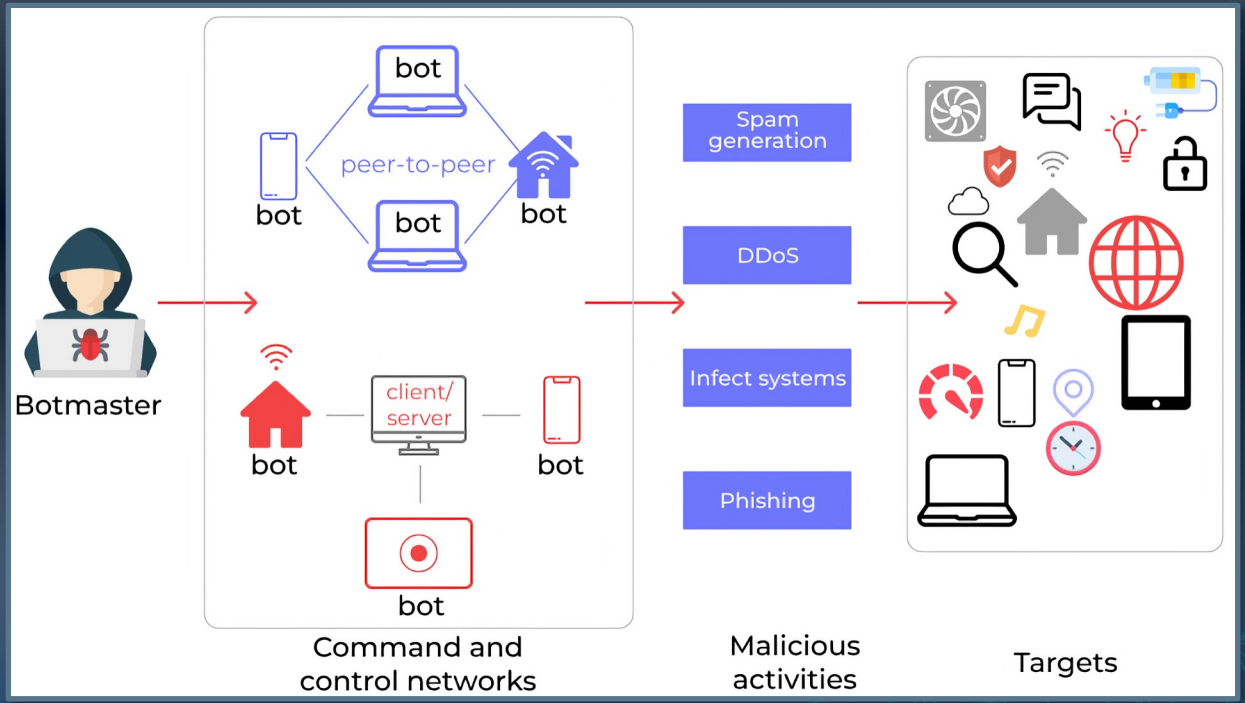
TACKLING THE BOTNETS
A MACHINE LEARNING APPROACH

Giuseppe Pericone
Giacomo Viaggi

Academic Year 2022/23

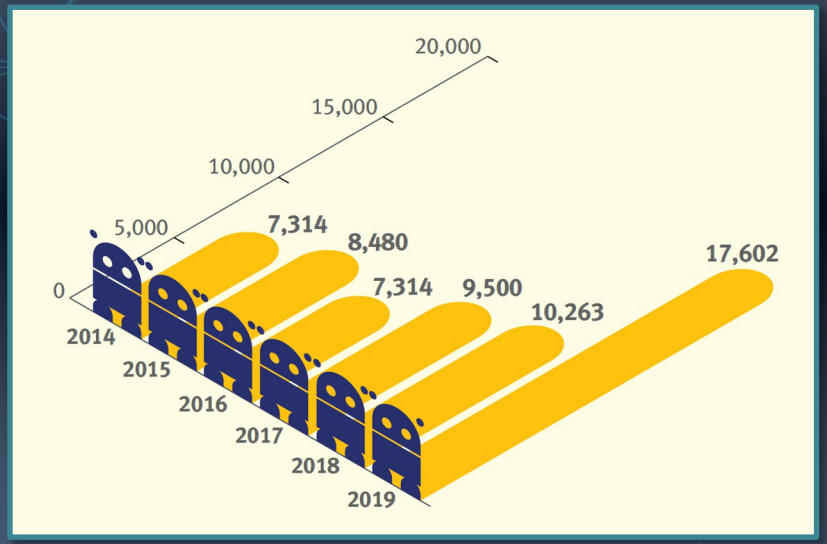
PROBLEM OVERVIEW

- How does a botnet work?



PROBLEM OVERVIEW

- State of play and future developments



Spamhaus Botnet Threat Report 2019: https://www.spamhaus.org/news/article/793/spamhaus-botnet-threat-report-2019

Mapping Mirai: A Botnet Case Study: https://www.malwaretech.com/2016/10/mapping-mirai-a-botnet-case-study.html

01 DATASET DESCRIPTION

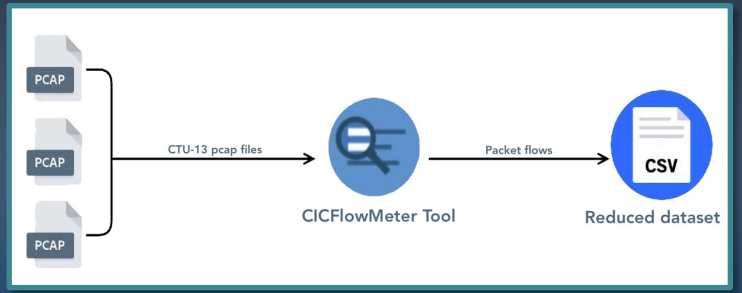
CTU-13 DATASET

- Gathering .pcap files into packet flows

From:

Id	Duration(hrs)	# Packets	#NetFlows	Size	Bot	#Bots
1	6.15	71,971,482	2,824,637	52GB	Neris	1
2	4.21	71,851,300	1,808,123	60GB	Neris	1
3	66.85	167,730,395	4,710,639	121GB	Rbot	1
4	4.21	62,089,135	1,121,077	53GB	Rbot	1
5	11.63	4,481,167	129,833	37.6GB	Virut	1
6	2.18	38,764,357	558,920	30GB	Menti	1
7	0.38	7,467,139	114,078	5.8GB	Sogou	1
8	19.5	155,207,799	2,954,231	123GB	Murlo	1
9	5.18	115,415,321	2,753,885	94GB	Neris	10
10	4.75	90,389,782	1,309,792	73GB	Rbot	10
11	0.26	6,337,202	107,252	5.2GB	Rbot	3
12	1.21	13,212,268	325,472	8.3GB	NSIS.ay	3
13	16.36	50,888,256	1,925,150	34GB	Virut	1

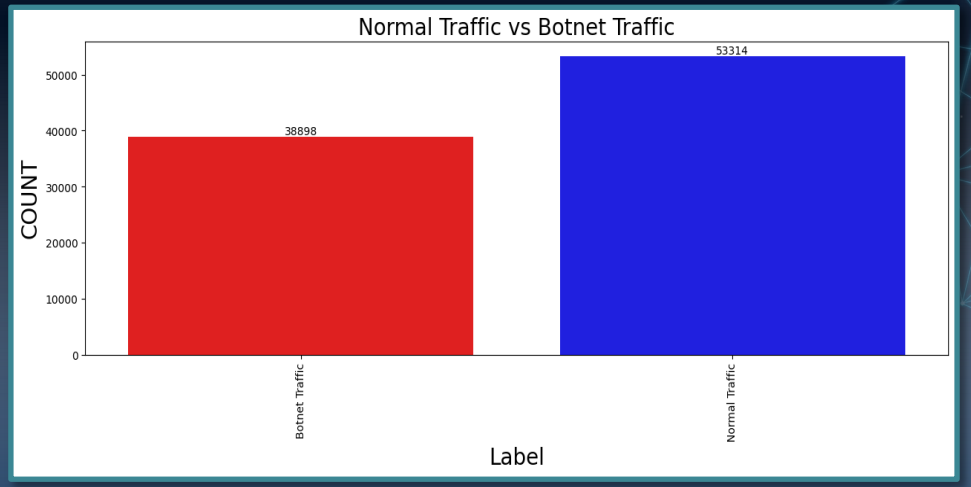
To:



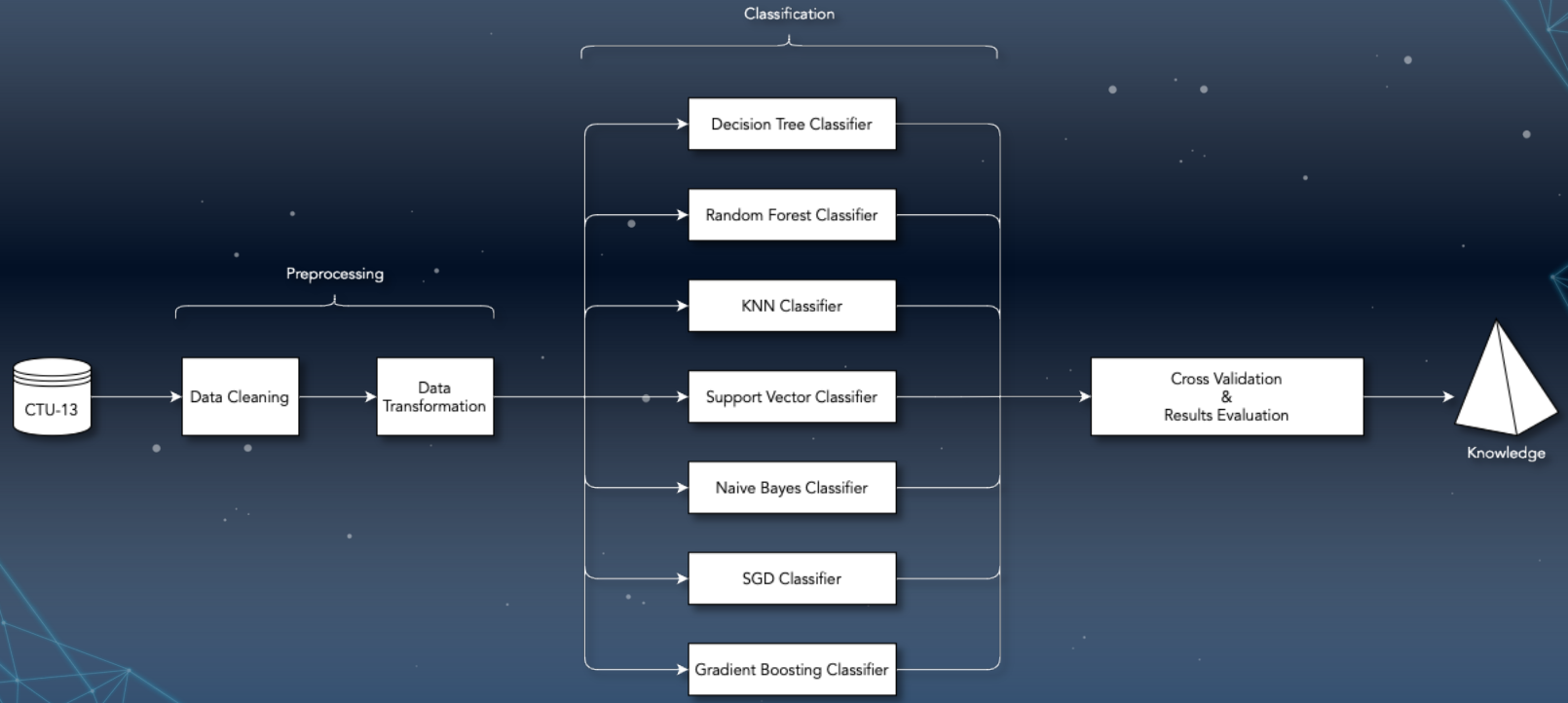
CTU-13 DATASET

➢ Let's deepen

- Number of tuples: **92.212**
- Number of attributes: **58**
  - Flow duration
  - Tot Fwd Pkts
  - Tot Bwd Pkts
  - SYN Flag Cnt
  - Down/Up Ratio
- Balancing of “Label”:
  - Botnet traffic: **38.898**
  - Normal traffic: **53.314**

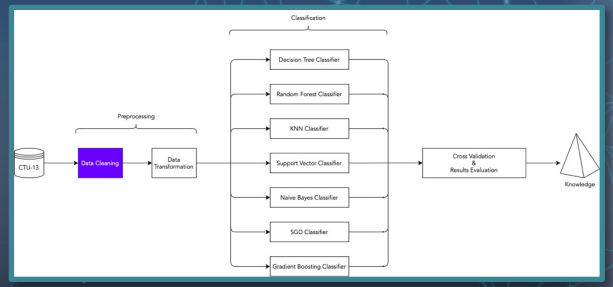


THE PIPELINE



02 PREPROCESSING

DATA CLEANING

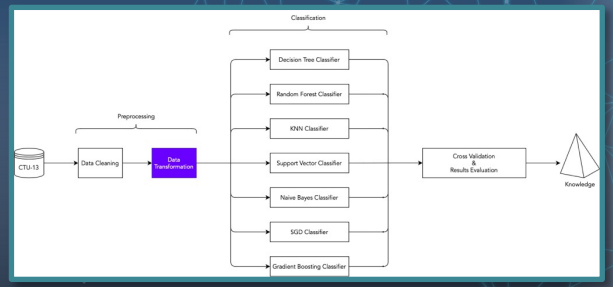


- Checked for the presence of null values
- Separation of binary attributes from numerical attributes
- Checked that the binary attributes have only the values 0 and 1

	count	mean	std	min	25%	50%	75%	max
Bwd PSH Flags	92212.0	0.021277	0.144307	0.0	0.0	0.0	0.0	1.0
FIN Flag Cnt	92212.0	0.041556	0.199574	0.0	0.0	0.0	0.0	1.0
SYN Flag Cnt	92212.0	0.317789	0.465620	0.0	0.0	0.0	1.0	1.0
RST Flag Cnt	92212.0	0.011853	0.108225	0.0	0.0	0.0	0.0	1.0
ACK Flag Cnt	92212.0	0.210851	0.407915	0.0	0.0	0.0	0.0	1.0
Label	92212.0	0.421832	0.493855	0.0	0.0	0.0	1.0	1.0

DATA TRANSFORMATION

- **Normalization of numerical values using MinMaxScaler**

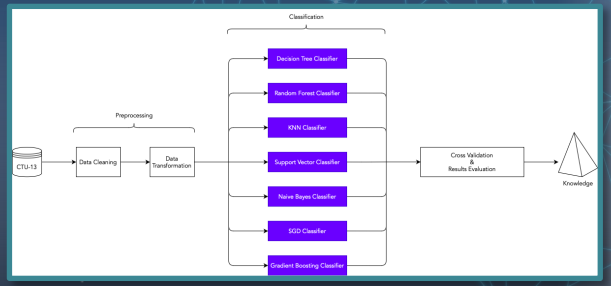


$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

03 CLASSIFICATION

CLASSIFICATION

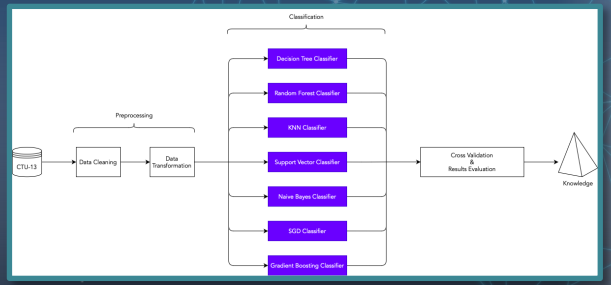
- Results of seven different classifiers

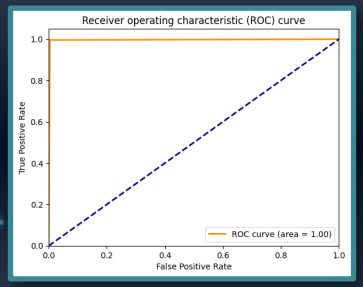


	Accuracy	Precision	Recall	F1 Score
Decision Tree	99.6349%	99.5489%	99.59128%	99.57008%
Random Forest	99.73973%	99.66809%	99.71901%	99.69354%
KNN	99.45416%	99.34451%	99.36989%	99.3572%
SVC	93.64517%	91.96504%	93.17098%	92.56408%
Naïve Bayes	73.54685%	89.71644%	42.56642%	57.73851%
SGD	85.56969%	90.02478%	74.23365%	81.37017%
Gradient Boosting	99.32042%	98.98271%	99.42098%	99.20136%

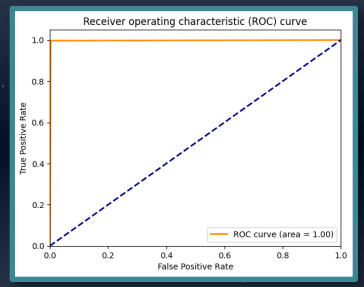
CLASSIFICATION

- Results of seven different classifiers

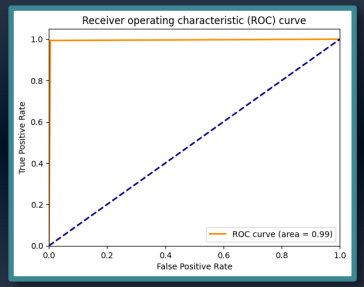




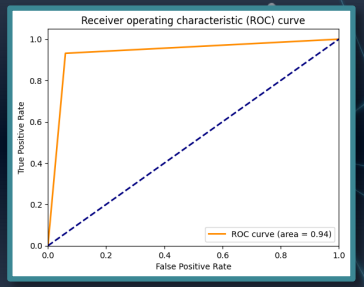
Decision Tree



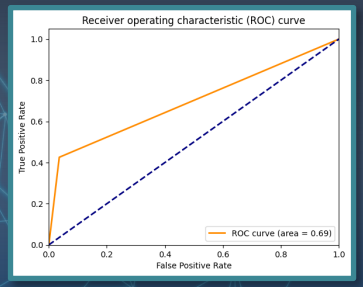
Random Forest



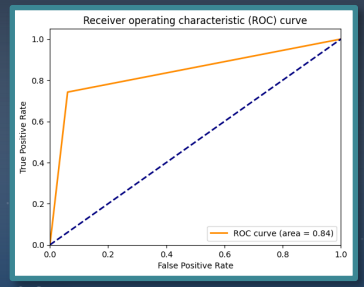
KNN



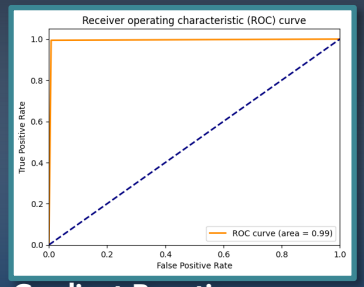
SVC



Naïve Bayes



SGD

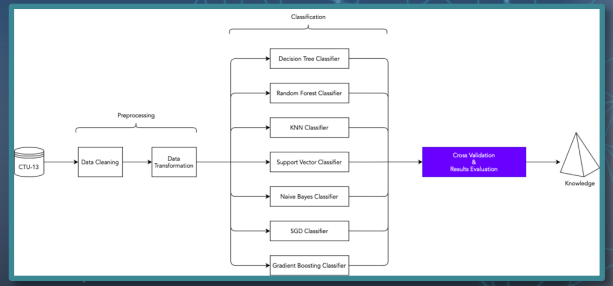


Gradient Boosting

04 CROSS VALIDATION & RESULTS EVALUATION

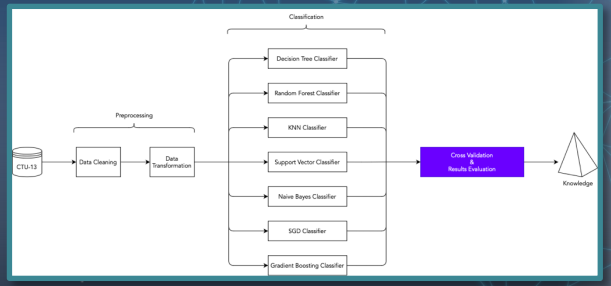
CROSS VALIDATION

- 10-fold cross validation results



	Accuracy	Precision	Recall	F1 Score	Avg Fit Time
Decision Tree	99.63128%	99.57834%	99.54754%	99.56291%	1.48619 s
Random Forest	99.76033%	99.69935%	99.73263%	99.71597%	11.00621 s
KNN	99.52826%	99.46005%	99.42157%	99.44073%	0.04278 s
SVC	93.84138%	92.06792%	93.45468%	58.36326%	305.20856 s
Naïve Bayes	73.96543%	89.68313%	43.26185%	58.36326%	0.06237 s
SGD	86.04737%	90.35934%	74.9216%	81.91745%	0.21157 s
Gradient Boosting	99.31896%	99.00413%	99.38557%	99.19434%	47.83948 s

RESULT EVALUATION

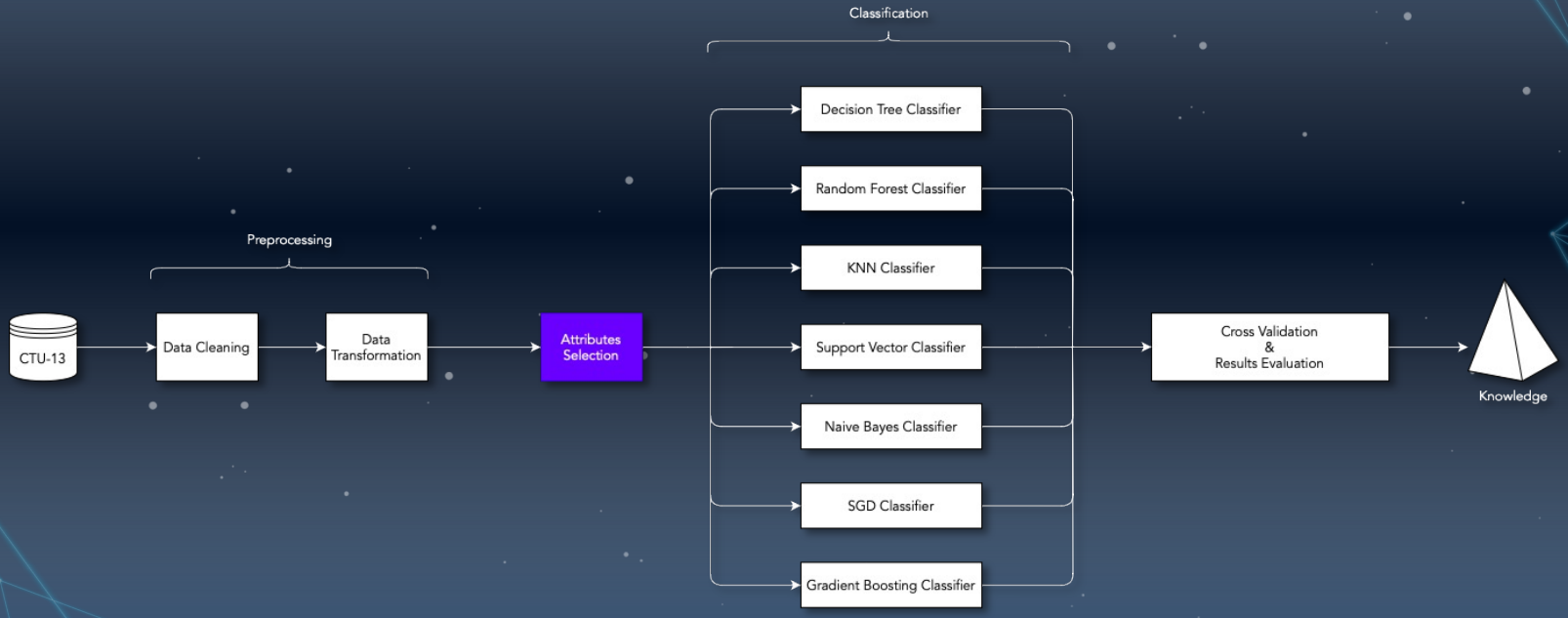


➢ Results & Wilcoxon Tests

Based on performance metrics such as recall, accuracy, and average model fitting time, Decision Tree and Random Forest have emerged as strong performers in our analysis. However, to compare all seven classifiers used in our study, we conducted the Wilcoxon test for every possible pair among them. For this comprehensive analysis, our chosen confidence level was set at α=0.05.

With the sole exception of the KNN-Gradient Boosting pair, the other Wilcoxon Tests reject the null hypothesis (p-value less than 0.05). The difference between the models, therefore, is statistically significant.

A NEW PIPELINE



ATTRIBUTES SELECTION

➢ Multiple selection methods

We want to understand which attributes are most significant for model development and use only those for training. For this purpose, we use multiple attribute selection techniques to find the attributes that are considered important in more than one method.

- Correlation-based
- Random Forest
- Singular Value Decomposition (SVD)
- Principal Component Analysis (PCA)
- Heuristic Search

CROSS VALIDATION

- With Attributes Selection

	Accuracy	Precision	Recall	F1 Score	Avg Fit Time
Decision Tree	99.62261%	99.54%	99.56553%	99.55274%	0.65559 s
Random Forest	99.77335%	99.71991%	99.74292%	99.73138%	6.40766 s
KNN	99.47946%	99.39325%	99.37272%	99.38291%	0.01215 s
SVC	93.55508%	91.67807%	93.18217%	51.91472%	104.01777 s
Naïve Bayes	71.41804%	89.41561%	36.5777%	51.91472%	0.02649 s
SGD	82.59772%	89.40059%	66.6538%	76.3669%	0.12559 s
Gradient Boosting	99.19099%	98.77801%	99.31101%	99.04368%	21.21699 s

A DIFFERENT CROSS VALIDATION

- **Stratified K-Fold Cross Validation**

	Accuracy	Precision	Recall	F1 Score	Avg Fit Time
Decision Tree	99.6443%	99.59377%	99.56296%	99.57834%	1.21188 s
Random Forest	99.75057%	99.6814%	99.72749%	99.70442%	8.96705 s
KNN	99.52826%	99.46005%	99.42157%	99.44073%	0.02177 s
SVC	93.84138%	92.06792%	93.45468%	58.36326%	165.68015 s
Naïve Bayes	73.96543%	89.68313%	43.26185%	58.36326%	0.042 s
SGD	86.12654%	90.17767%	75.31752%	82.07879%	0.2221 s
Gradient Boosting	99.31896%	99.00413%	99.38557%	99.19434%	38.16156 s

WHAT OTHERS DO

- **Methods and results from other papers**

From researching papers dealing with the same topic, we note that the most suitable classifiers for this type of problem appear to be Decision Tree and Random Forest, as we also inferred. The results in terms of performance seem to be very similar to ours.

Some papers also use other classification methods not found in our project, such as Neural Network, RNN, and BClus. Only the latter has statistics like ours.

REFERENCES

https://www.wallarm.com/what/what-is-a-botnet

https://www.malwaretech.com/2016/10/mapping-mirai-a-botnet-case-study.html

https://www.enisa.europa.eu/topics/cyber-threats/threats-and-trends/etl-review-folder/etl-2020-botnet

https://www.spamhaus.org/news/article/793/spamhaus-botnet-threat-report-2019

https://www.stratosphereips.org/datasets-ctu13

https://github.com/imfaisalmalik/CTU13-CSV-Dataset

García, Sebastián & Grill, Martin & Stiborek, Jan & Zunino, Alejandro. (2014). An Empirical Comparison of Botnet Detection Methods. Computers & Security. 45. 100-123. 10.1016/j.cose.2014.05.011.

Bansal, Ankit & Mahapatra, Sudipta. (2017). A comparative analysis of machine learning techniques for botnet detection. 91-98. 10.1145/3136825.3136874.

C. Maudoux, S. Boumerdassi, A. Barcello and E. Renault, "Combined Forest: a New Supervised Approach for a Machine-Learning-based Botnets Detection," 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 2021, pp. 01-06, doi: 10.1109/GLOBECOM46510.2021.9685261.

Velasco-Mata, J., González-Castro, V., Fernández, E.F., & Alegre, E. (2021). Efficient Detection of Botnet Traffic by Features Selection and Decision Trees. IEEE Access, 9, 120567-120579.

THANKS
FOR YOUR ATTENTION