INTRUSION PREVENTION SYSTEM

Presented by – SHUBHAM SINGH (MIT2021023 – I Year)

UNDERSUPERVISION OF – PROF. O.P. VYAS

CONTENT

- **Problem Statement**
- **Dataset Used**
  - UNSW-NB15
    - UNSW-NB15 features
    - UNSW-NB15 for Binary labelling
    - UNSW-NB15 for Multiclass labelling
  - TON IoT (UNSW-IoT20)
    - IoT Fridge, IoT Garage Door, IoT GPS Tracker, IoT Modbus, IoT Motion light, IoT Thermostat, IoT Weather
- **Methodology**
  - UNSW-NB15
    - Data Preprocessing & Normalization
    - Correlation Matrix
  - TON IoT (UNSW-IoT20)
    - Data Preprocessing
    - Normalization
- **Intrusion Prevention System**
  - Phase 1 : Model Formation & calculation of Threshold value
  - Phase 2 : Prevention – Dropping of packets
- **Models used in both datasets for analysis**
- **Timeline**
- **References**

ANALYSIS OF UNSW-NB15 & TON-IOT20 FOR DETECTION PHASE AND INTRUSION PREVENTION SYSTEM ON UNSW-NB15

DATASET USED

- UNSW-NB15[1]
- UNSW-IoT20 (TON-IoT)[2]

[1]. https://research.unsw.edu.au/projects/unsw-nb15-dataset
[2]. https://research.unsw.edu.au/projects/toniot-datasets

UNSW-NB15[1]

- UNSW-NB 15 data set is created by the IXIA Perfect-Storm tool in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS) for generating a hybrid of real modern normal activities and synthetic contemporary attack behaviors.
- This data set has nine families of attacks, namely, Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Re-connaissance, Shellcode and Worms.
- This dataset consists of 49 features, and are described in UNSW-NB15 features.csv file.
- Dataset used for binary and multiclass classification is UNSW NB15 training-set.csv.
- The number of records in the Dataset is 175,341 records from different the types of attack and normal.

[1]. https://research.unsw.edu.au/projects/unsw-nb15-dataset

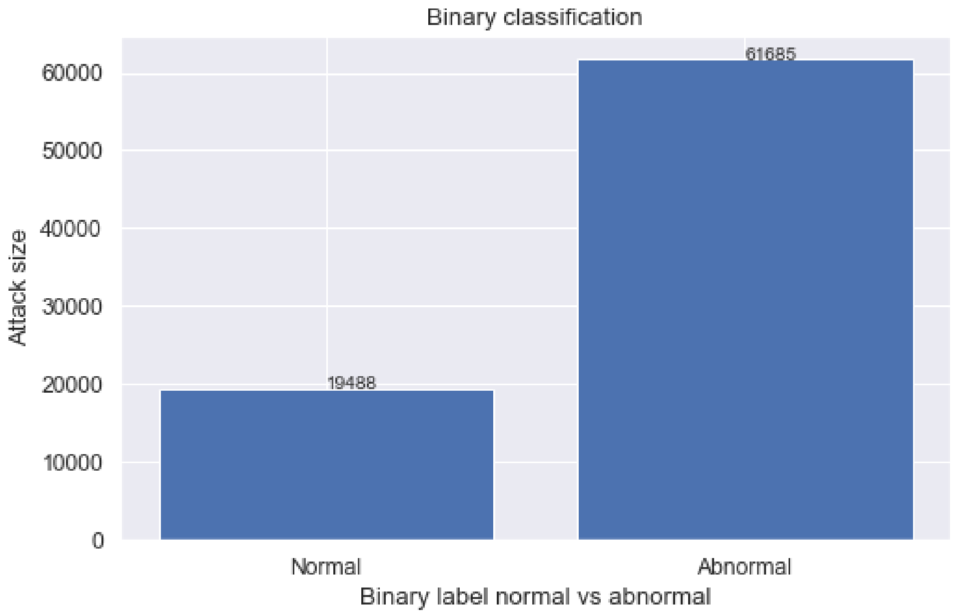
UNSW-NB15[1] FEATURES

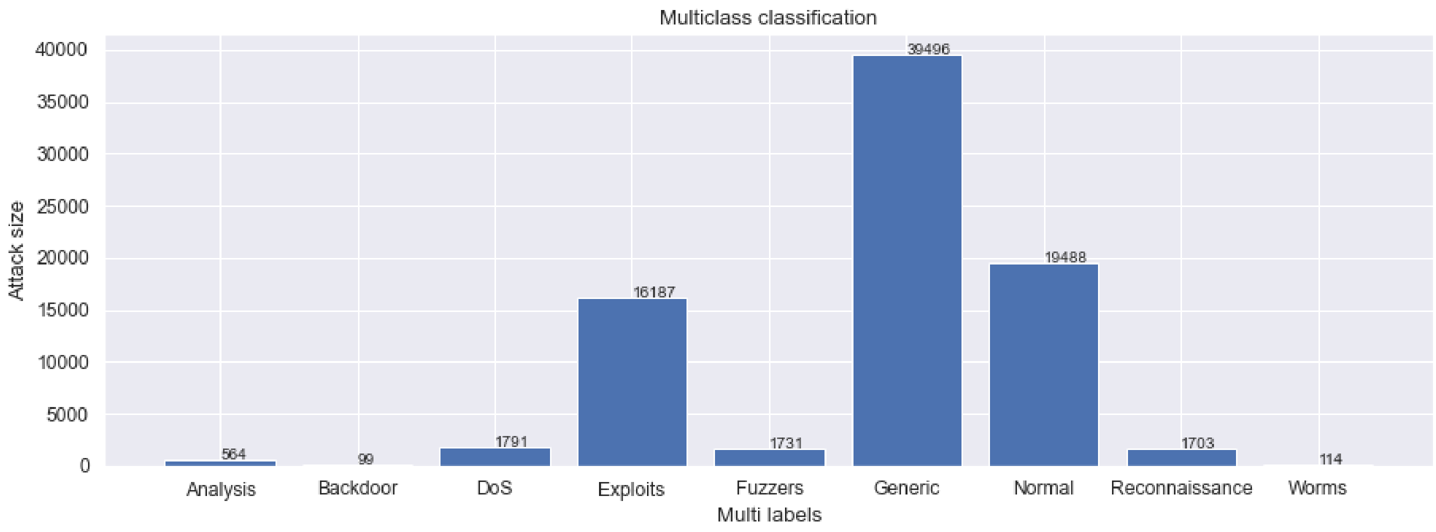
- Dataset **UNSW-NB15_features.csv** consists of 49 features.
- **Nominal** : 'proto', 'service', 'state', 'attack cat
- **Integer** : 'sbytes', 'dbytes', 'sttl', 'dttl', 'sloss', 'dloss', 'swin', 'stcpb','dtcpb', 'dwin', 'trans depth', 'ct srv src', 'ct state ttl', 'ct dst ltm', 'ct src dport ltm', 'ct dst sport ltm', 'ct dst src ltm', 'ct ftp cmd', 'ct flw http mthd', 'ct srv dst'
- **Float** : 'dur', 'tcprtt', 'synack', 'ackdat
- **Binary** : 'is ftp login', 'is sm ips ports'

[1]. https://research.unsw.edu.au/projects/unsw-nb15-dataset

UNSW-NB15[1] LABELLING

- For Binary Classification, feature label has two labels as 19,488 records as normal data and 61685 records as attack data.[2]
- For Multiclass Classification, feature attack cat has 9 labels Analysis(564), Backdoor(99), Dos(1,791), Exploits(16,187), Fuzzers Multi labels(1,731), Generic(39,496), Normal(19,488), Reconnaissance(1,703), Worms(114).[2]





[1]. https://research.unsw.edu.au/projects/unsw-nb15-dataset
[2]. https://matplotlib.org/3.5.0/api/_as_gen/matplotlib.pyplot.bar.html

TON IOT[1] (UNSW-IOT20)

- The **TON IoT (UNSW-IoT20)** datasets are new generations of Internet of Things (IoT) and Industrial IoT (IIoT) datasets.

- The datasets have been called ‘**ToN IoT**’ as they include heterogeneous data sources collected from Telemetry datasets of **IoT** and **IIoT sensors**, Operating systems datasets of **Windows 7** and **10** as well as **Ubuntu 14** and **18 TLS** and **Network traffic datasets**.

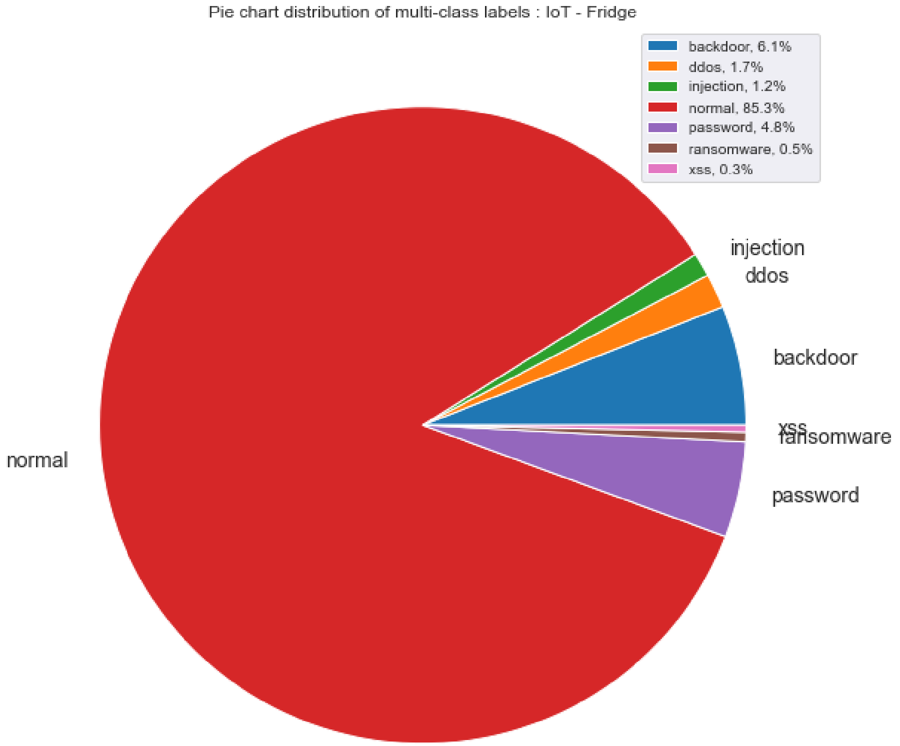
- The datasets were collected from a realistic and largescale network designed at the **IoT Lab** of the **UNSW Canberra Cyber (SEIT)**.

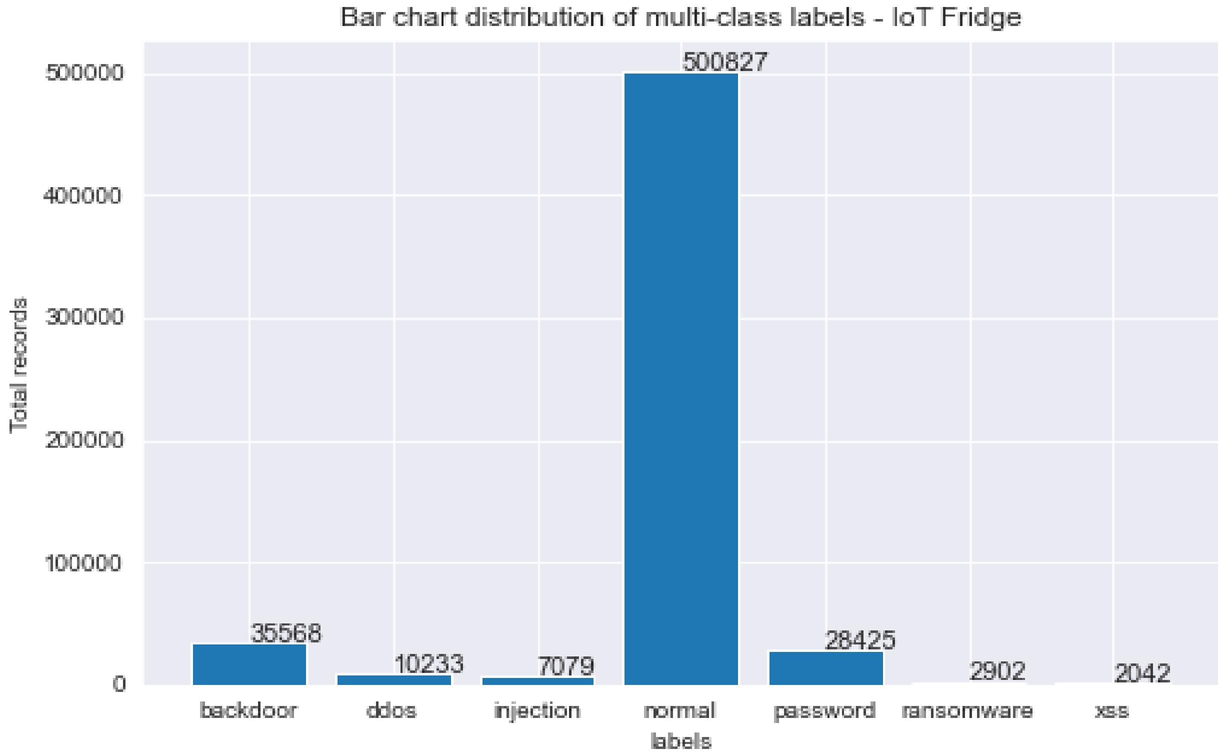
- Processed IoT dataset is being used for binary and multiclass classification. It consists of seven .csv files **IoT Fridge.csv**, **IoT Garage Door.csv**, **IoT GPS Tracker.csv**, **IoT Modbus.csv**, **IoT Motion Light.csv**, **IoT Thermostat.csv**, **IoT Weather.csv**.

[1]. https://research.unsw.edu.au/projects/toniot-datasets

IOT[1] FRIDGE

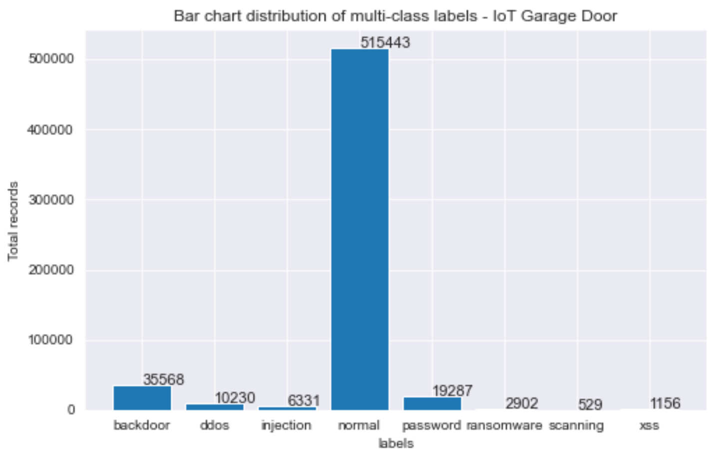
- **IoT Fridge** dataset features are **ts**, **date**, **time**, **fridge_temperature**, **temp_condition**, **label**, **type**. Below diagrams are for multi-class labels from '**type**' feature.[2,3]



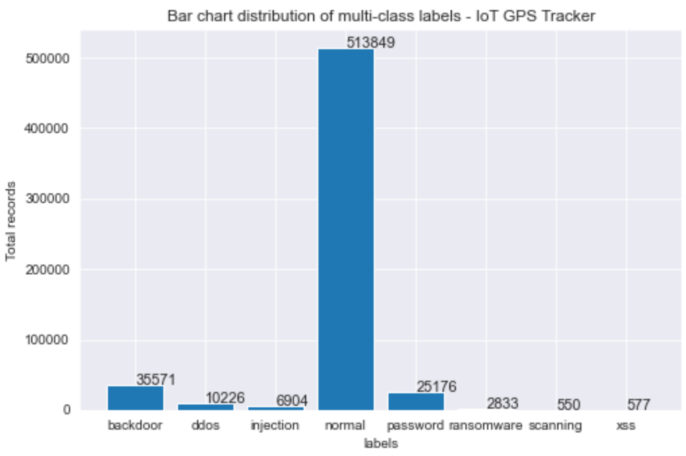


[1]. https://research.unsw.edu.au/projects/toniot-datasets
[2]. https://matplotlib.org/3.5.0/api/_as_gen/matplotlib.pyplot.bar.html
[3]. https://matplotlib.org/stable/gallery/pie_and_polar_charts/pie_features.html

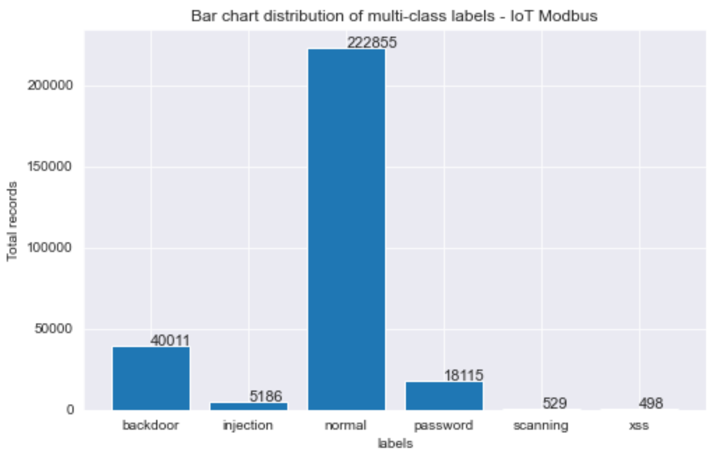
IOT Garage Door



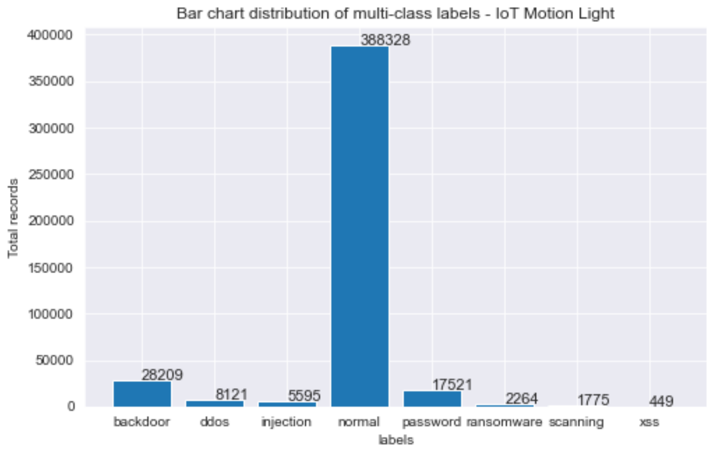
IOT GPS Tracker



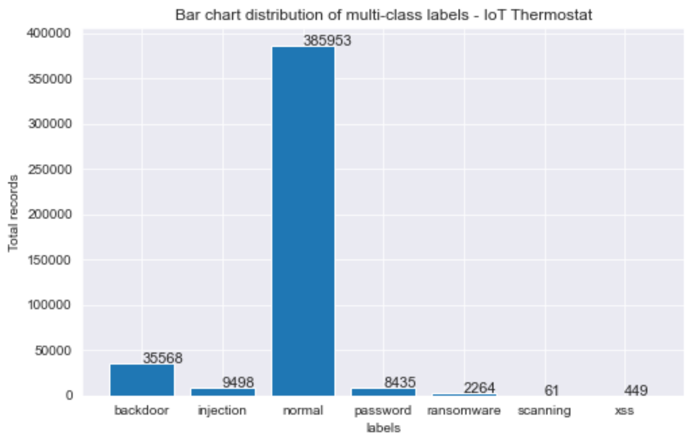
IOT Modbus



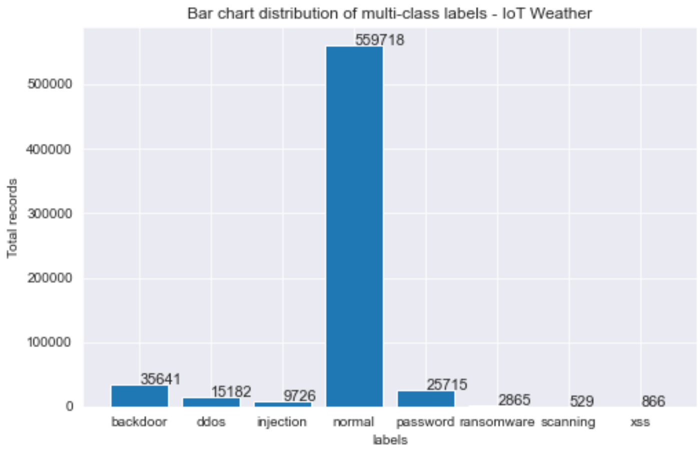
IOT Motion Light



IOT Thermostat



IOT Weather



METHODOLOGY – UNSW-NB15[1]

- **Data Pre-processing**:

  - Dataset's feature **'select'** consists of values **'-'** so entire rows are deleted from dataset.

  - Variant numeric data types are converted into single numeric data type.

  - Nominal/categorical data is dealt using **one-hot encoding**[4] i.e., features which lie in this category are '**proto**', '**service**', '**state**'.

  - Total features after encoding are 61.

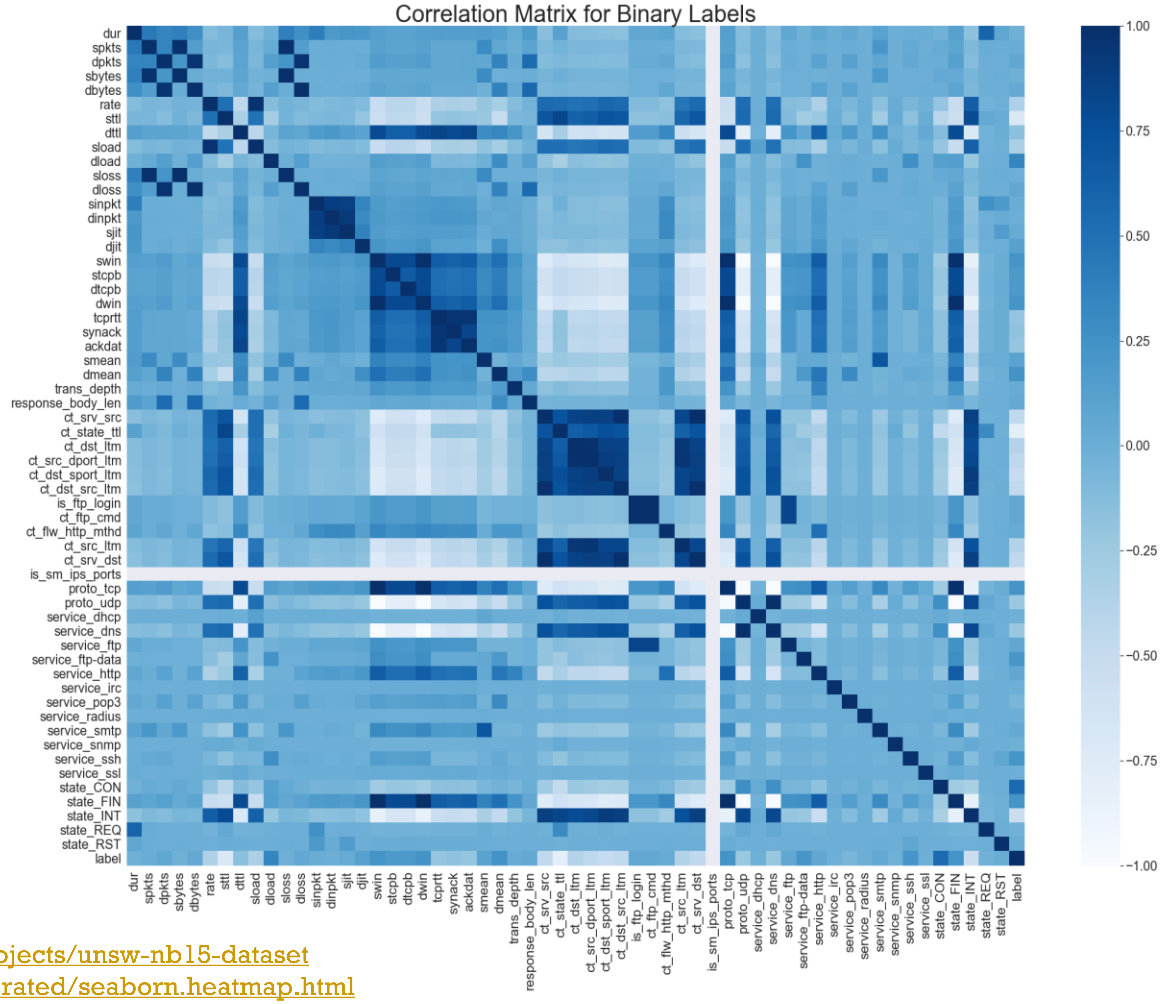
  - All numeric data type features are normalized using **MinMaxScaler()**[2] with range(0,1).

  - Binary labels are formed using **LabelEncoder()**[3] , where as Multiclass Labels are formed using **one-hot-encoding**[4] & **LabelEncoder()**[3] .

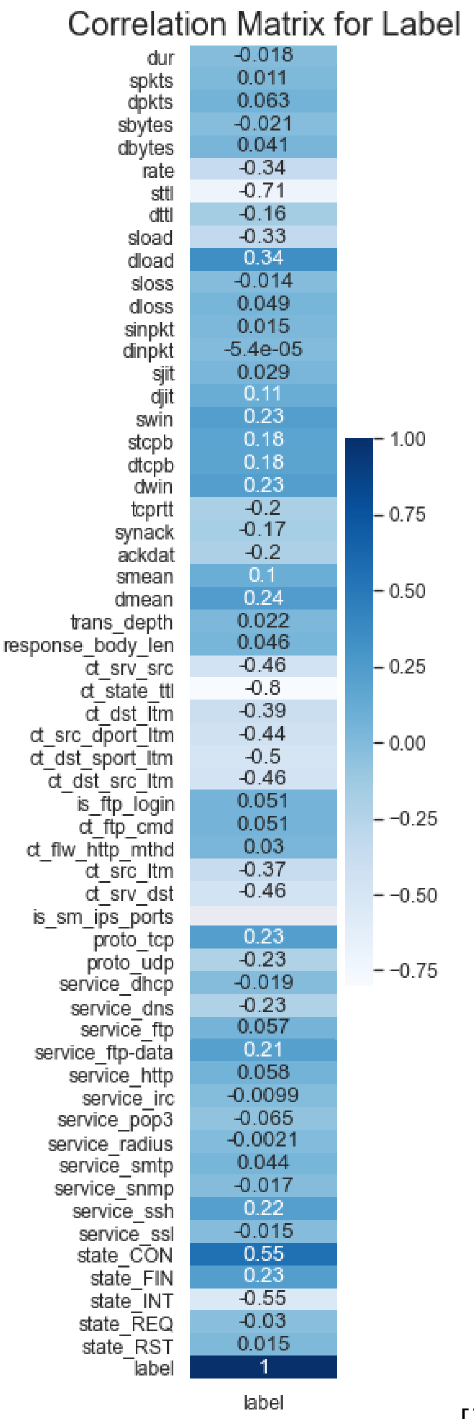
[1]. https://research.unsw.edu.au/projects/unsw-nb15-dataset
[2]. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html
[3]. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html
[4]. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html

METHODOLOGY – UNSW-NB15[1]

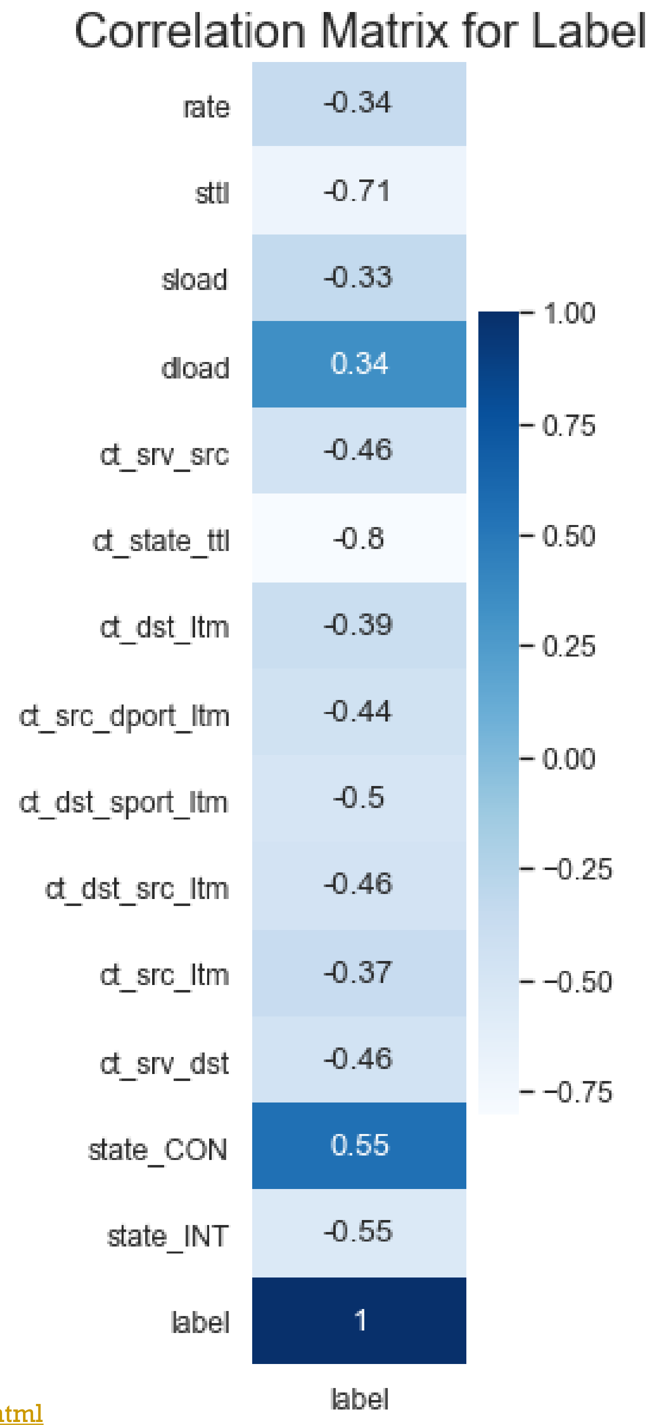
- **Feature Selection for Binary Labelled Data:**
- Correlation matrix[2] is formed and features with correlation value less than 0.3 are removed from dataset.



[1]. https://research.unsw.edu.au/projects/unsw-nb15-dataset
[2]. https://seaborn.pydata.org/generated/seaborn.heatmap.html



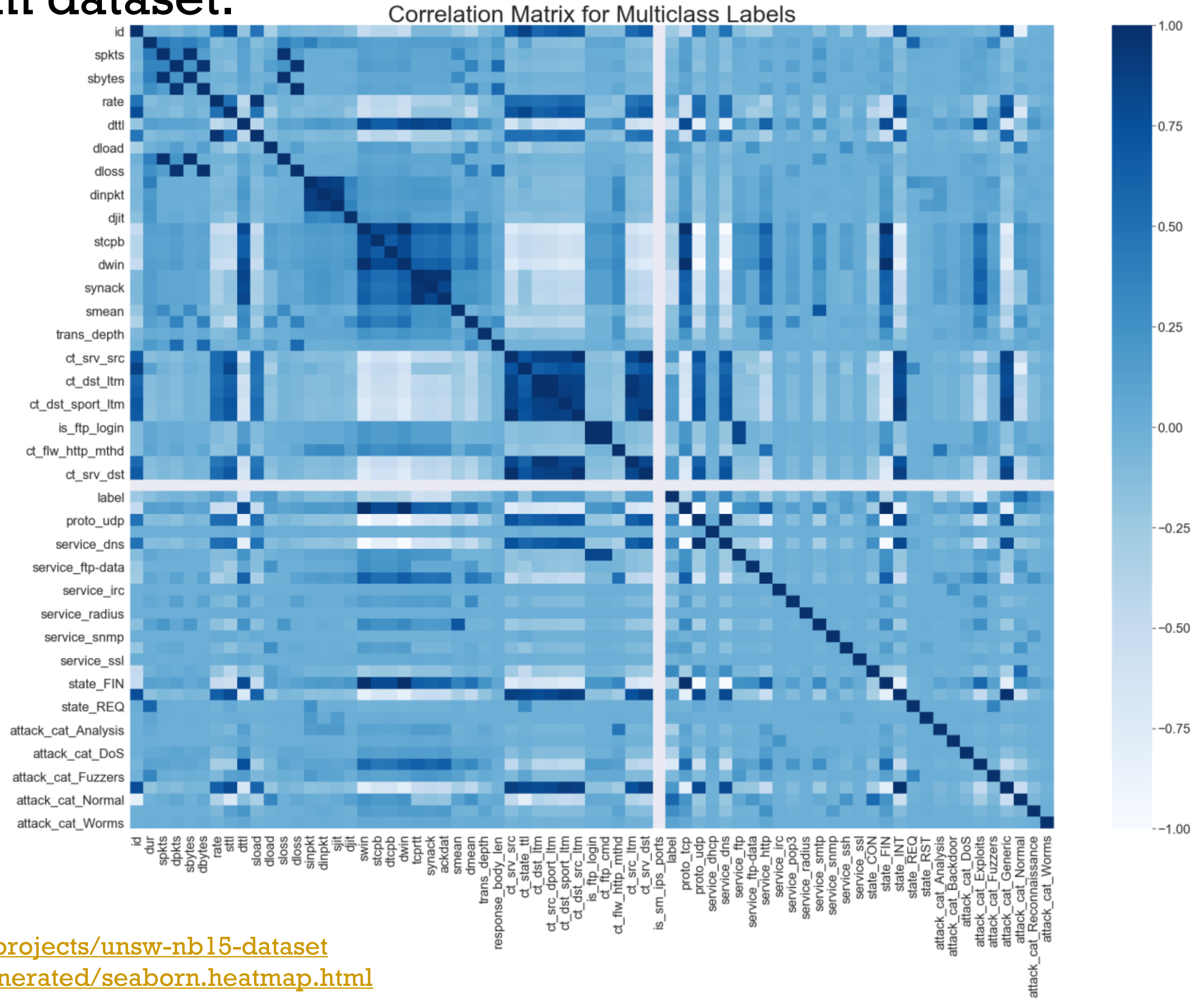
features gte 0.3



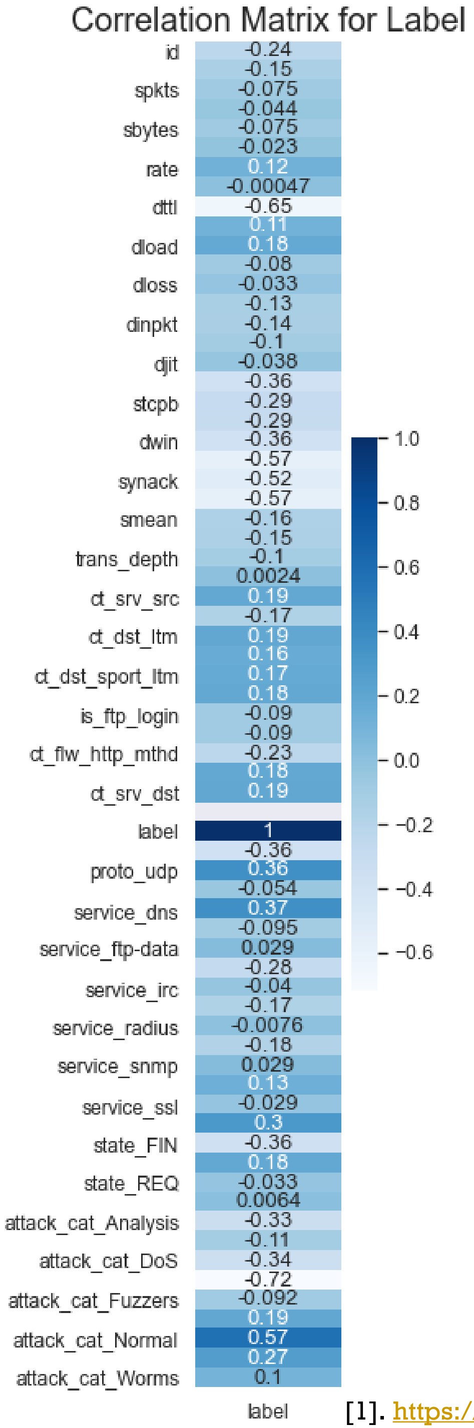
[1]. https://seaborn.pydata.org/generated/seaborn.heatmap.html

METHODOLOGY – UNSW-NB15[1]

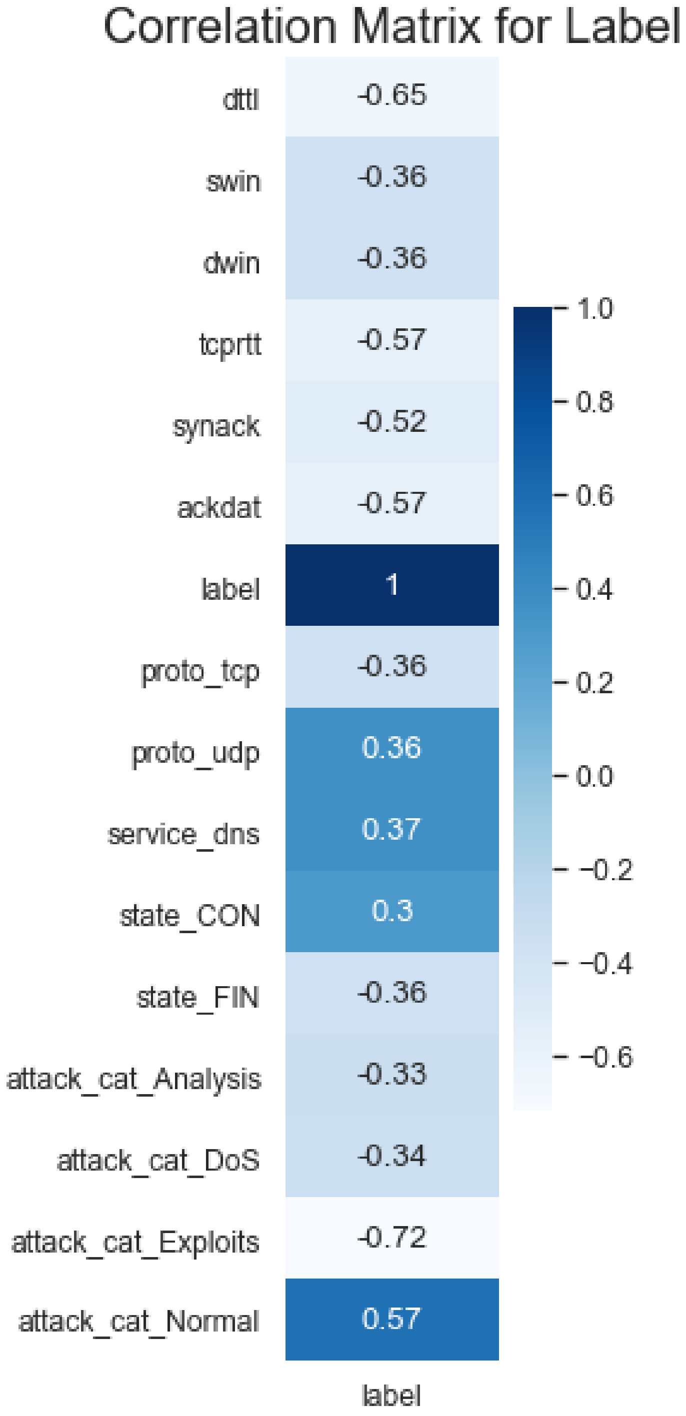
- **Feature Selection for Multiclass Labelled Data:**
- Correlation matrix[2] is formed and features with correlation value less than 0.3 are removed from dataset.



[1]. https://research.unsw.edu.au/projects/unsw-nb15-dataset
[2]. https://seaborn.pydata.org/generated/seaborn.heatmap.html



features gte 0.3



[1]. https://seaborn.pydata.org/generated/seaborn.heatmap.html

METHODOLOGY – TON IOT[1] (UNSW-IOT20)

- **Data Pre-processing**:
  - From all seven datasets, **timestamp**, **date**, **time** has been removed.
  - In dataset **IoT Fridge**, feature '**temp_condition**' has six unique values 'high', 'high ', 'high ', 'low', 'low ', 'low ' and further processed by trimming down the space making it '**high**' and '**low**'.
  - In dataset **IoT Garage Door,** feature '**sphone_signal**' consists of six unique labels '0', 'false ', '0.0', '1', 'true ', '1.0' and further processed by omitting it to '**false**' and '**true**'.

[1]. https://research.unsw.edu.au/projects/toniot-datasets

METHODOLOGY – TON IOT[1] (UNSW-IOT20)

- **Normalization:**
  - All numeric data type features are normalized using **MinMaxScaler()**[2] with range(0,1).
  - In Dataset IoT Fridge, feature '**fridge_temperature**' has been normalized.
  - In Dataset IoT GPS Tracker, feature '**latitude**' & '**longitude**' has been normalized.
  - In Dataset IoT Modbus, features '**FC1_Read_Input_Register**', '**FC2_Read_Discrete_Value**', ' **FC3_Read_Holding_Register**', ' **FC4_Read_Coil**' has been normalized.
  - In Dataset IoT Thermostat, feature '**current_temperature**' has been normalized.
  - In Dataset IoT Weather, feature '**temperature**', '**pressure**' & '**humidity**' has been normalized.

[1]. https://research.unsw.edu.au/projects/toniot-datasets
[2]. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

MODELS USED IN BOTH DATASETS FOR ANALYSIS

- Logistic Regression.
- Naïve Bayes
- KNN
- Decision Tree
- Random Forest
- AdaBoost
- SVM-linear , rbf , sigmoid

OBSERVATION – UNSW-NB15[1]

	ML Model	Accuracy	Precision	Recall	F1-Measure	Execution Time(s)
UNSW - NB15 **Binary Labelled Data**	**Logistic Regression**	0.97	0.98	0.98	0.98	0.00267
	Naïve Bayes	0.74	0.87	0.75	0.76	0.01101
	KNN	0.98	0.98	0.98	0.98	4.27056
	Decision Tree	**0.98**	**0.98**	**0.98**	**0.98**	**0.00444**
	Random Forest	**0.98**	**0.98**	**0.98**	**0.98**	**0.04627**
	AdaBoost	0.98	0.98	0.98	0.98	0.07368
	SVM-linear	0.97	0.98	0.98	0.98	0.75300
	SVM-rbf	0.97	0.98	0.98	0.98	1.50678
	SVM-sigmoid	0.94	0.94	0.94	0.94	2.85291

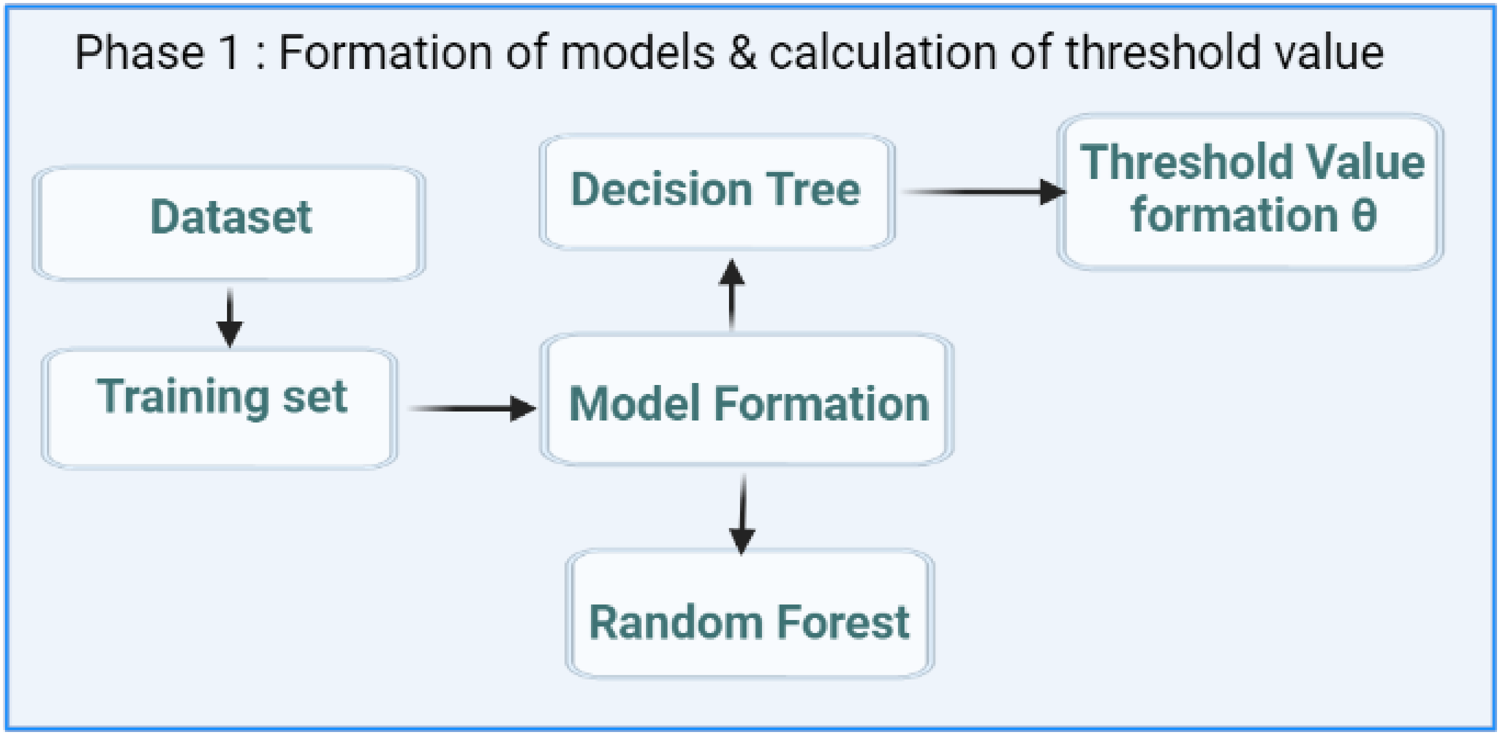
	ML Model	Accuracy	Precision	Recall	F1-Measure	Execution Time(s)
UNSW - NB15 **Multiclass Labelled Data**	**Logistic Regression**	0.97	0.97	0.97	0.97	0.00763
	Naïve Bayes	0.95	0.95	0.95	0.95	0.04636
	KNN	0.97	0.97	0.97	0.97	17.2074
	Decision Tree	**0.97**	**0.97**	**0.97**	**0.97**	**0.00661**
	Random Forest	**0.97**	**0.97**	**0.97**	**0.97**	**0.07719**
	AdaBoost	0.75	0.63	0.75	0.67	0.22904
	SVM-linear	0.97	0.97	0.98	0.97	1.26248
	SVM-rbf	0.97	0.97	0.98	0.97	2.14475
	SVM-sigmoid	0.97	0.96	0.97	0.96	2.64214

[1]. https://research.unsw.edu.au/projects/unsw-nb15-dataset

INTRUSION PREVENTION SYSTEM

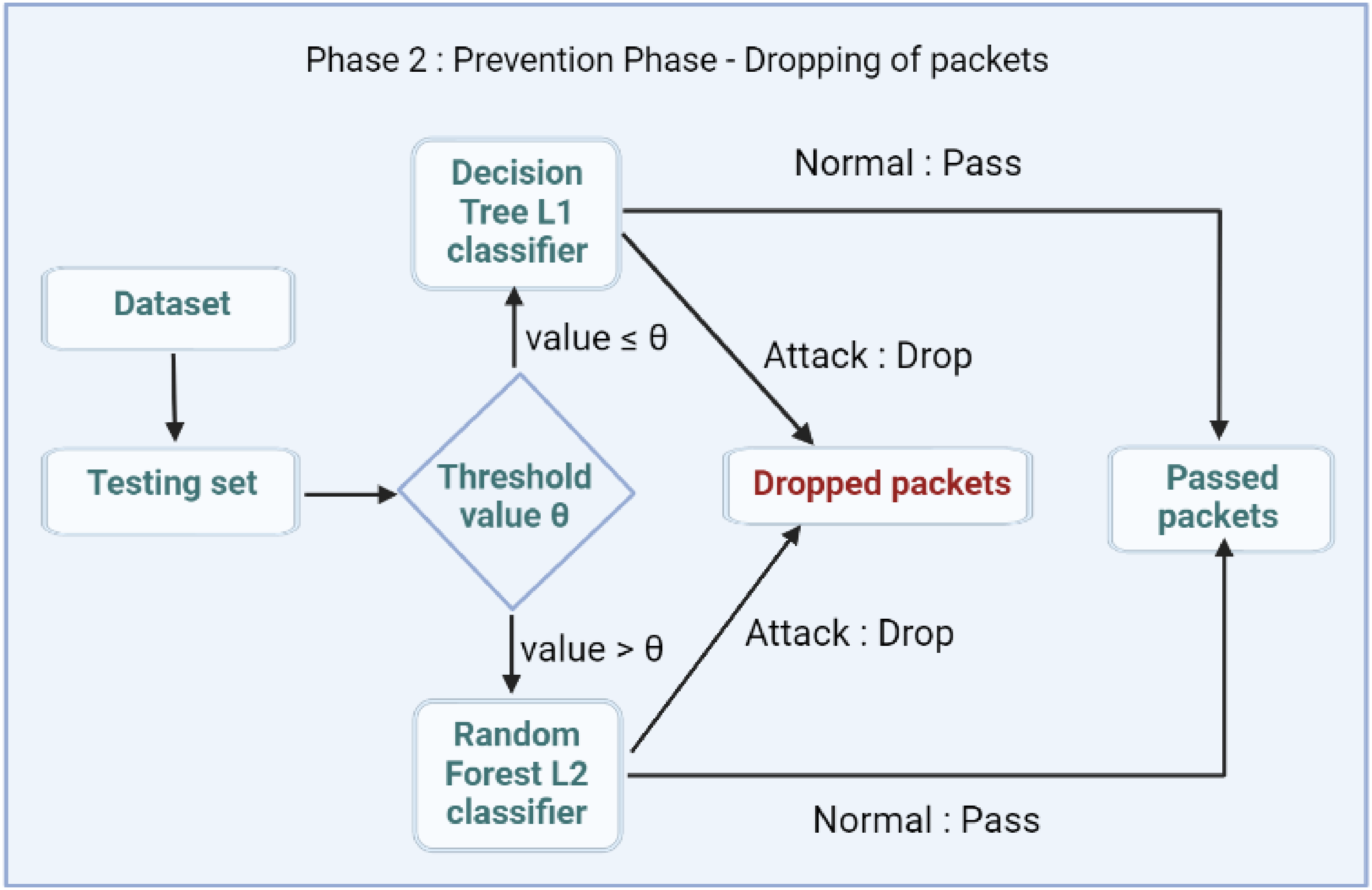
- Intrusion Prevention System is divided into two phases.
- UNSW dataset is used for training of models - Decision Trees & Random Forest.
- Same dataset is used to train both the models.
- Threshold values is calculated by calculating using time factor as major factor for elimination of packets.
- Now, Decision Tree Model is used for Level 1 classifier whereas Random Forest is used for Level 2 classifier.
- If the value is less than or equal to threshold value than L1 classifier is used otherwise stream is forward to L2 classifier.
- Both L1classifier , L2 classifier are used for prevention of attack.

PHASE 1 - IPS



[1]. Created with BioRender.com

PHASE 2 - IPS



[1]. Created with BioRender.com

RESULTS : L1 & L2 CLASSIFIER - BINARY

L1 Classifier – Decision Tree

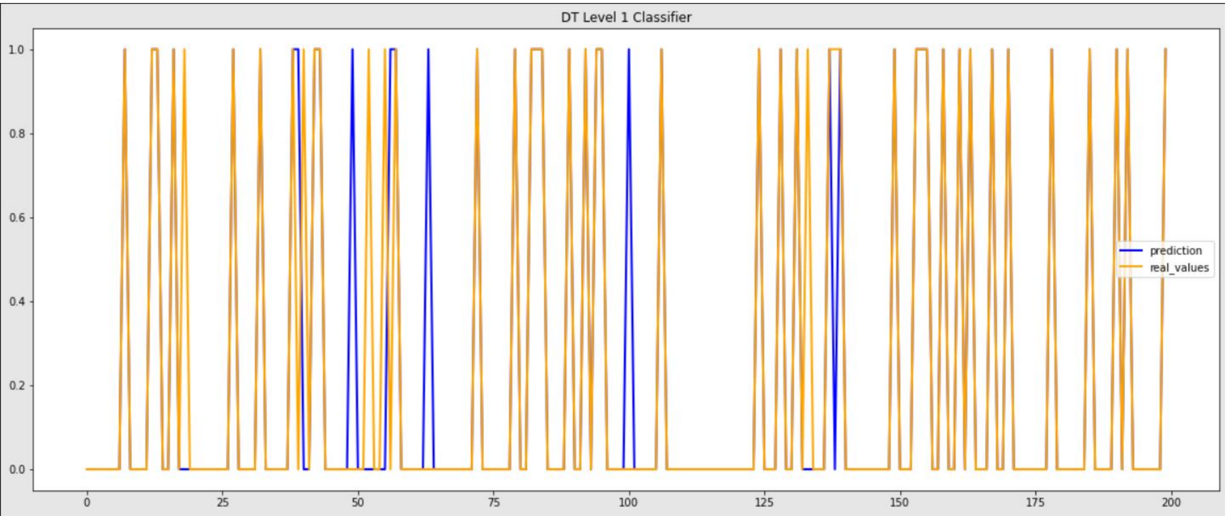
Accuracy - 98.15058159341996

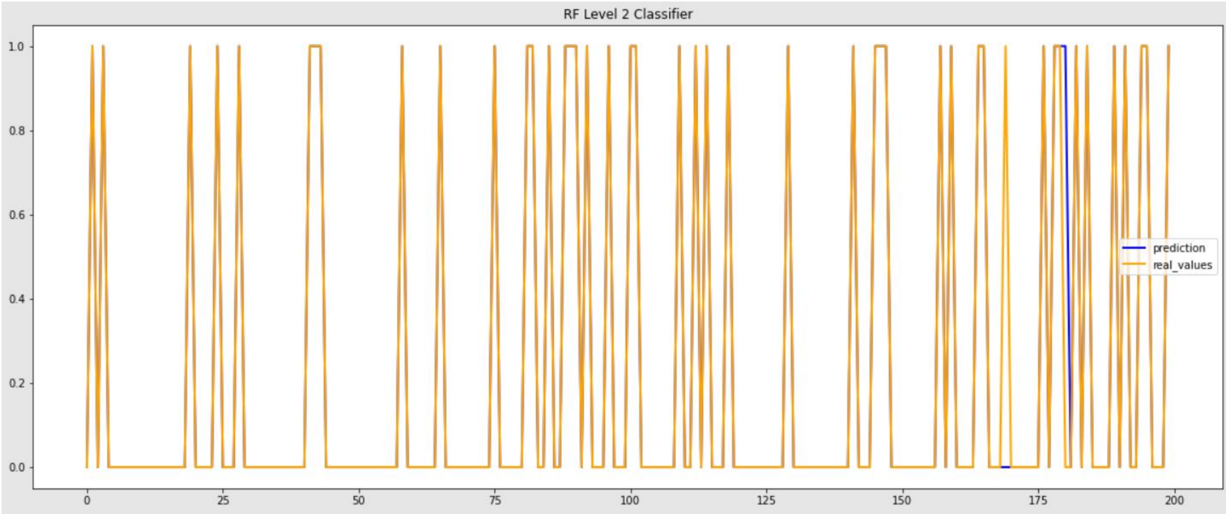
	precision	recall	f1-score	support
abnormal	0.99	0.99	0.99	15620
normal	0.96	0.96	0.96	4927
accuracy			0.98	20547
macro avg	0.97	0.98	0.97	20547
weighted avg	0.98	0.98	0.98	20547

L2 Classifier – Random Forest

Accuracy - 98.29172141918528

	precision	recall	f1-score	support
abnormal	0.99	0.99	0.99	2879
normal	0.97	0.96	0.96	926
accuracy			0.98	3805
macro avg	0.98	0.97	0.98	3805
weighted avg	0.98	0.98	0.98	3805





RESULTS : L1 & L2 CLASSIFIER - MULTICLASS

L1 Classifier – Decision Tree

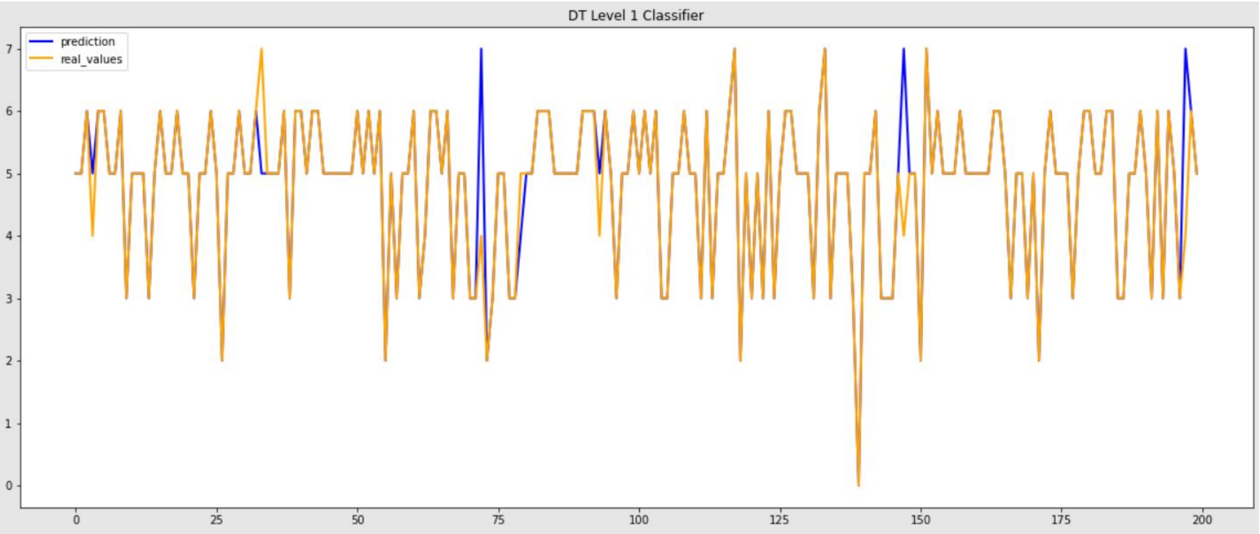
Accuracy - 97.04352054461464

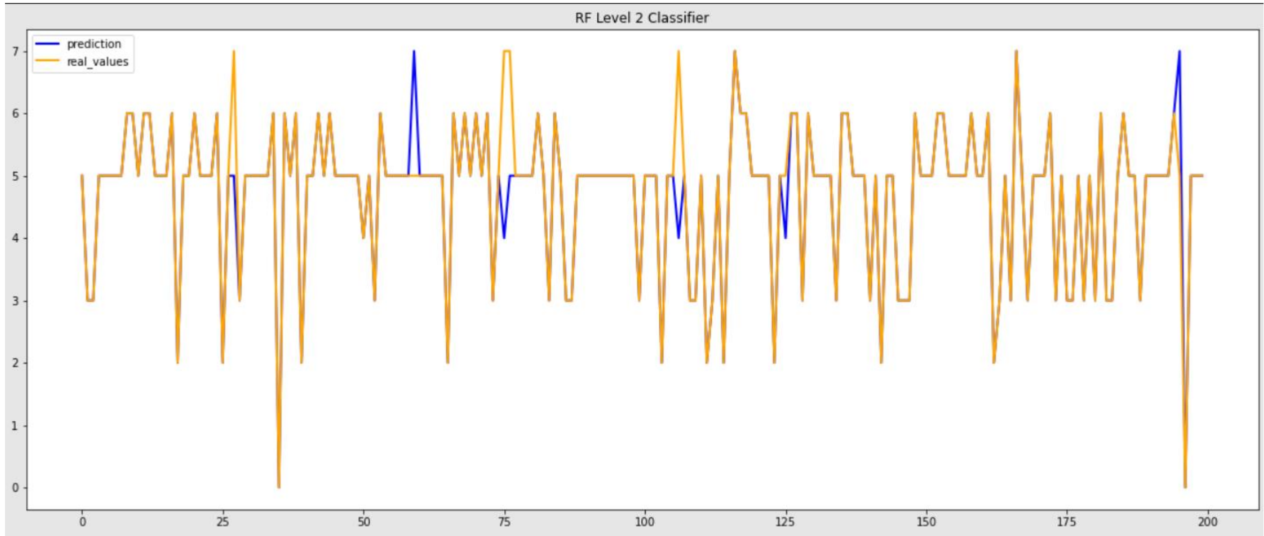
	precision	recall	f1-score	support
Analysis	1.00	1.00	1.00	147
Backdoor	0.08	0.09	0.08	22
DoS	1.00	1.00	1.00	477
Exploits	1.00	1.00	1.00	4060
Fuzzers	0.48	0.38	0.42	430
Generic	0.98	0.99	0.99	9926
Normal	1.00	1.00	1.00	5043
Reconnaissance	0.55	0.52	0.53	436
Worms	0.10	0.17	0.12	24
accuracy			0.97	20565
macro avg	0.69	0.68	0.68	20565
weighted avg	0.97	0.97	0.97	20565

L2 Classifier – Random Forest

Accuracy - 97.20095062054396

	precision	recall	f1-score	support
Analysis	1.00	1.00	1.00	26
Backdoor	0.00	0.00	0.00	3
DoS	1.00	1.00	1.00	82
Exploits	1.00	1.00	1.00	776
Fuzzers	0.51	0.52	0.52	77
Generic	0.99	0.99	0.99	1846
Normal	1.00	1.00	1.00	882
Reconnaissance	0.53	0.57	0.55	87
Worms	0.50	0.12	0.20	8
accuracy			0.97	3787
macro avg	0.73	0.69	0.70	3787
weighted avg	0.97	0.97	0.97	3787





TIMELINE

- C1 Evaluation involves literature study, research gap, possible solutions, proposed methodology.

- C2 Evaluation involves implementation and results of proposed method.

- C3 Evaluation involves final report of proposed method.

REFERENCES -

[1]. https://research.unsw.edu.au/projects/unsw-nb15-dataset

[2]. https://research.unsw.edu.au/projects/toniot-datasets

[3]. N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," 2015 Military Communications and Information Systems Conference (MilCIS), 2015, pp. 1–6, DOI: 10.1109/MilCIS.2015.7348942.

[4]. Khan, M.A. *et al.* (2022). Voting Classifier-Based Intrusion Detection for IoT Networks. In: Saeed, F., Al-Hadhrami, T., Mohammed, E., Al-Sarem, M. (eds) Advances on Smart and Soft Computing. Advances in Intelligent Systems and Computing, vol 1399. Springer, Singapore. https://doi.org/10.1007/978981-16-5559-3_26

[5]. A. R. Gad, A. A. Nashat and T. M. Barkat, "Intrusion Detection System Using Machine Learning for Vehicular Ad Hoc Networks Based on ToN-IoT Dataset," in *IEEE Access*, vol. 9, pp. 142206-142217, 2021, doi: 10.1109/ACCESS.2021.3120626.

THANK YOU