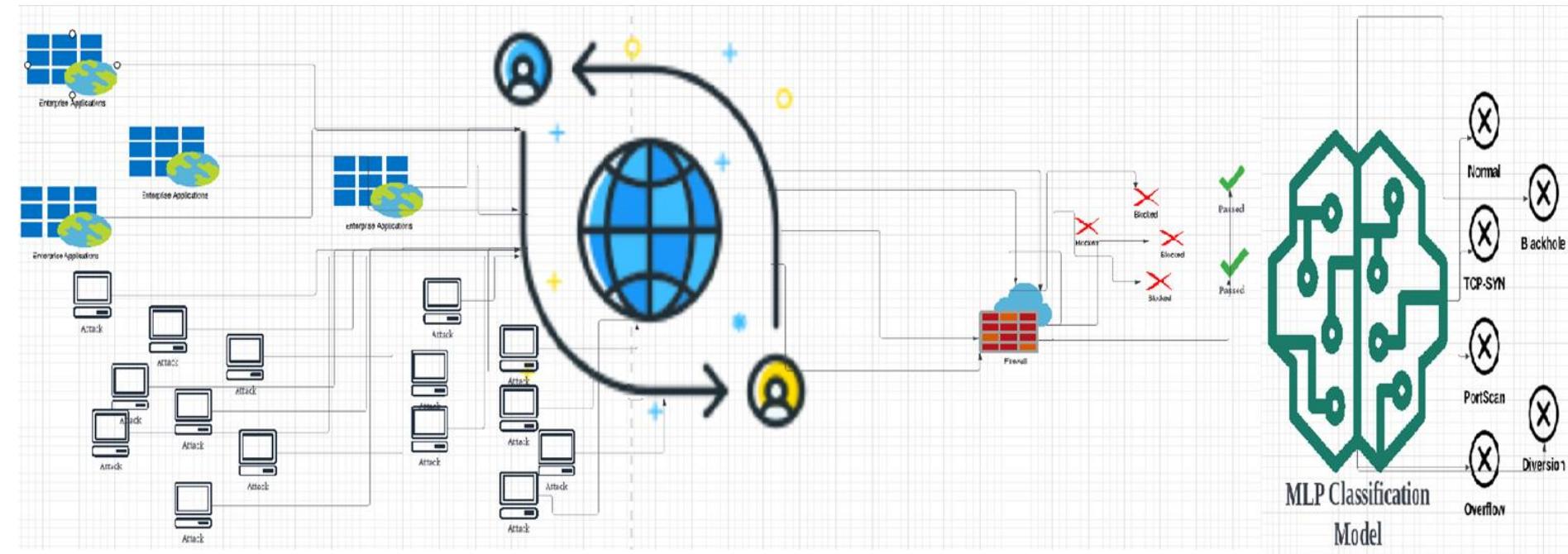




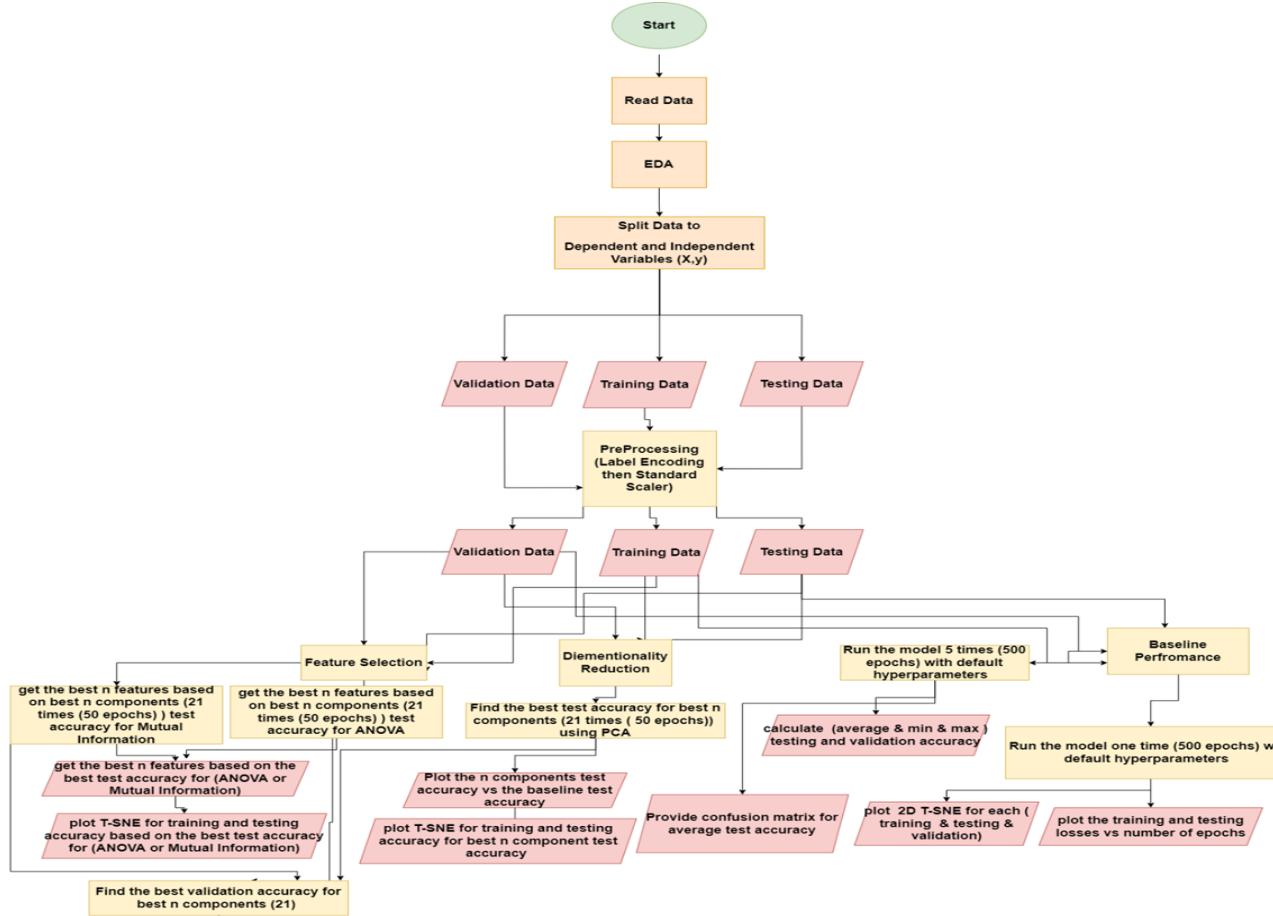
# Applied Machine Learning project

## Group 18

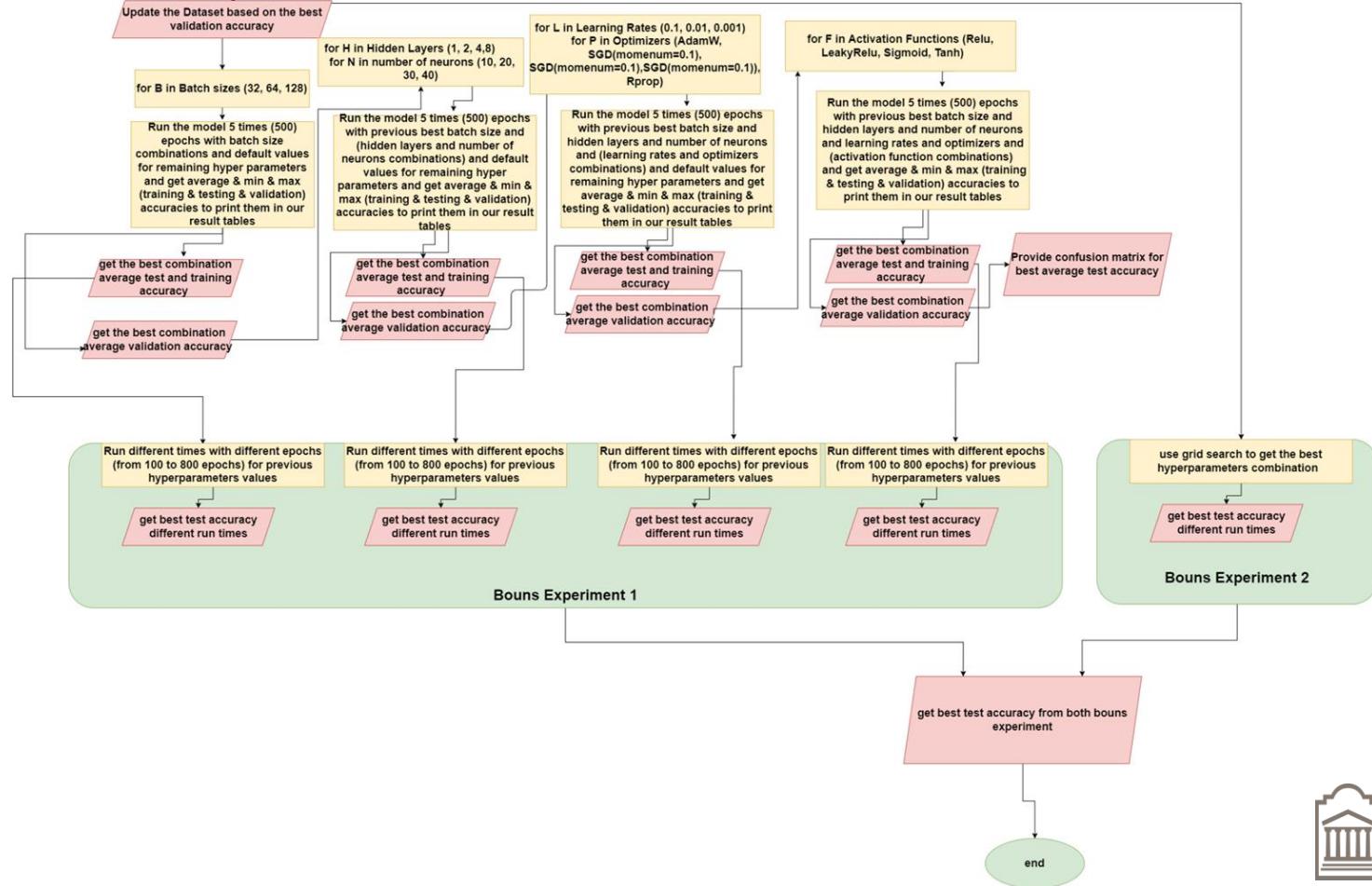
# Problem's overview



# Flowchart



# Flowchart. cont



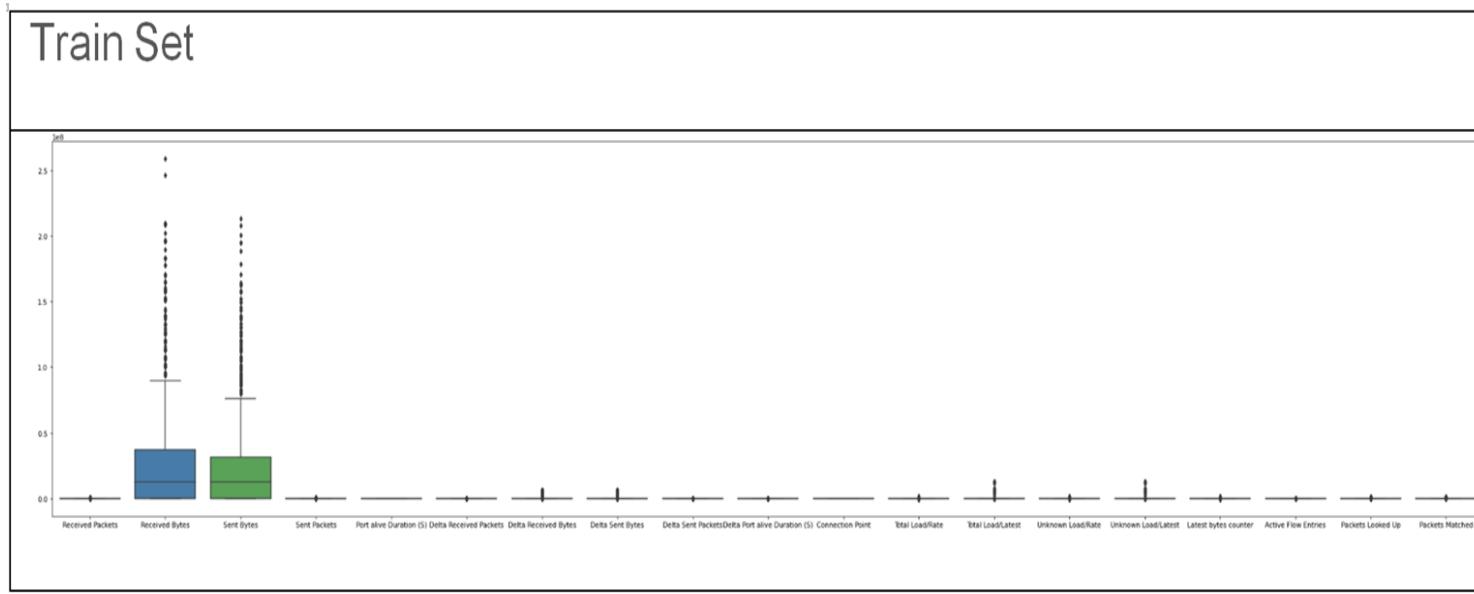
# Dataset's overview (EDA)

Number of instances per class:	Train Set	Test Set	Validation Set
	Counter({'TCP-SYN': 651, 'Normal': 254, 'PortScan': 666, 'Overflow': 73, 'Diversion': 414, 'Blackhole': 561})	Counter({'TCP-SYN': 156, 'Normal': 49, 'Blackhole': 122, 'Diversion': 90, 'PortScan': 127, 'Overflow': 17})	Counter({'Diversion': 92, 'TCP-SYN': 123, 'Blackhole': 143, 'PortScan': 141, 'Overflow': 15, 'Normal': 47})
null values per feature:	Switch ID Port Number Received Packets Received Bytes Sent Bytes Sent Packets Port alive Duration (s) Delta Received Packets Delta Received Bytes Delta Sent Bytes Delta Sent Packets Delta Port alive Duration (s) Connection Point Total Load/Rate Total Load/Latest Unknown Load/Rate Unknown Load/Latest Latest bytes counter Active Flow Entries Packets Looked Up Packets Matched Label	Switch ID Port Number Received Packets Received Bytes Sent Bytes Sent Packets Port alive Duration (s) Delta Received Packets Delta Received Bytes Delta Sent Bytes Delta Sent Packets Delta Port alive Duration (s) Connection Point Total Load/Rate Total Load/Latest Unknown Load/Rate Unknown Load/Latest Latest bytes counter Active Flow Entries Packets Looked Up Packets Matched Label	Switch ID Port Number Received Packets Received Bytes Sent Bytes Sent Packets Port alive Duration (s) Delta Received Packets Delta Received Bytes Delta Sent Bytes Delta Sent Packets Delta Port alive Duration (s) Connection Point Total Load/Rate Total Load/Latest Unknown Load/Rate Unknown Load/Latest Latest bytes counter Active Flow Entries Packets Looked Up Packets Matched Label
possible outliers	Active Flow Entries Connection Point Delta Port alive Duration (s) Delta Received Bytes Delta Received Packets Delta Sent Bytes Delta Sent Packets Label Latest bytes counter Packets Looked Up Packets Matched Port Number Port alive Duration (s) Received Bytes Received Packets Sent Bytes Sent Packets Switch ID Total Load/Latest Total Load/Rate Unknown Load/Latest Unknown Load/Rate	Active Flow Entries Connection Point Delta Port alive Duration (s) Delta Received Bytes Delta Received Packets Delta Sent Bytes Delta Sent Packets Label Latest bytes counter Packets Looked Up Packets Matched Port Number Port alive Duration (s) Received Bytes Received Packets Sent Bytes Sent Packets Switch ID Total Load/latest Total Load/Rate Unknown Load/Latest Unknown Load/Rate	Active Flow Entries Connection Point Delta Port alive Duration (s) Delta Received Bytes Delta Received Packets Delta Sent Bytes Delta Sent Packets Label Latest bytes counter Packets Looked Up Packets Matched Port Number Port alive Duration (s) Received Bytes Received Packets Sent Bytes Sent Packets Switch ID Total Load/Latest Total Load/Rate Unknown Load/Latest Unknown Load/Rate



# Dataset's overview (EDA)

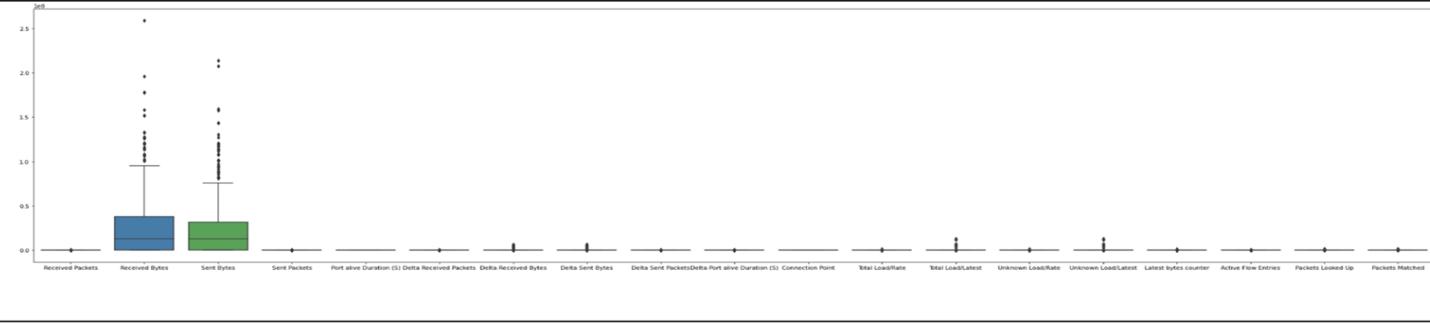
BoxPlot(Showing Outliers):



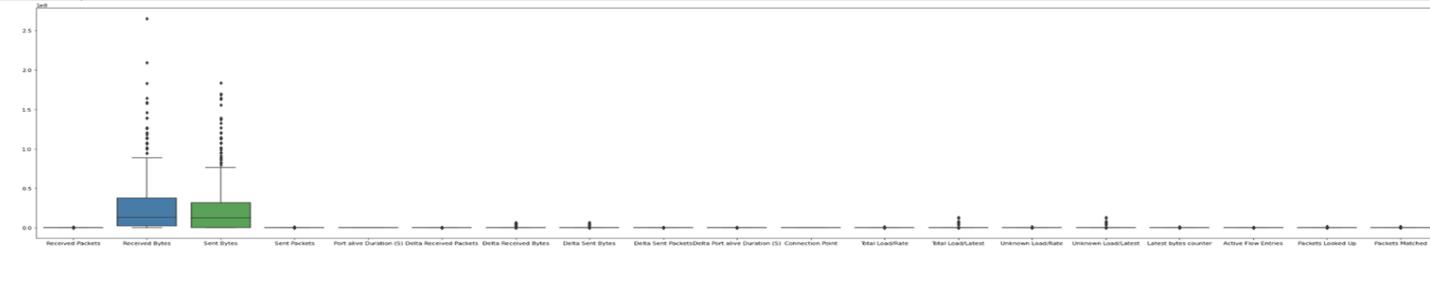
# Dataset's overview (EDA)

BoxPlot(Showing Outliers):

Test Set



Validation Set



uOttawa

# Dataset's overview (EDA)

## Train Set

	Received Packets	Received Bytes	Sent Bytes	Sent Packets	Port alive Duration (s)	Delta Received Packets	Delta Received Bytes	Delta Sent Bytes	Delta Sent Packets	Delta Port alive Duration (s)	Connection Point
count	2619.000000	2.619000e+03	2.619000e+03	2619.000000	2619.000000	2619.000000	2.619000e+03	2.619000e+03	2619.000000	2619.000000	2619.000000
mean	19929.095074	2.606639e+07	2.372645e+07	33676.567774	890.539137	174.948835	3.254661e+05	2.957171e+05	160.607866	4.831615	2.450935
std	61817.860816	3.670801e+07	3.309908e+07	90565.150583	970.916789	1028.907797	1.166624e+06	1.107859e+06	972.484039	0.374279	1.159286
min	10.000000	8.560000e+02	6.025000e+03	44.000000	26.000000	0.000000	0.000000e+00	2.780000e+02	2.000000	4.000000	1.000000
25%	321.500000	8.464900e+04	5.360900e+04	333.500000	136.000000	2.000000	2.780000e+02	2.800000e+02	2.000000	5.000000	1.000000
50%	1108.000000	1.262083e+07	1.262176e+07	1243.000000	254.000000	4.000000	5.560000e+02	5.560000e+02	4.000000	5.000000	2.000000
75%	3268.500000	3.736785e+07	3.170219e+07	3727.000000	1721.000000	6.000000	8.310000e+02	7.590000e+02	5.000000	5.000000	3.000000
max	352584.000000	2.589422e+08	2.130728e+08	420806.000000	3317.000000	15588.000000	6.171714e+06	6.302910e+06	15593.000000	5.000000	5.000000

Total Load/Rate	Total Load/Latest	Unknown Load/Rate	Unknown Load/Latest	Latest bytes counter	Active Flow Entries	Packets Looked Up	Packets Matched
2.619000e+03	2.619000e+03	2.619000e+03	2.619000e+03	2.619000e+03	2619.000000	2.619000e+03	2.619000e+03
2.044276e+04	5.033087e+05	2.044276e+04	5.033087e+05	2.044276e+04	94.972890	9.920463e+04	9.908445e+04
1.141013e+05	1.796231e+06	1.141013e+05	1.796231e+06	1.141013e+05	828.181086	2.363175e+05	2.363162e+05
-6.446240e+05	0.000000e+00	-6.446240e+05	0.000000e+00	-6.446240e+05	4.000000	1.050000e+02	5.000000e+01
0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	5.000000	2.263000e+03	2.147000e+03
0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	6.000000	6.968000e+03	6.843000e+03
0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	8.000000	2.163300e+04	2.140800e+04
1.260657e+06	1.275386e+07	1.260657e+06	1.275386e+07	1.260657e+06	9985.000000	1.012220e+06	1.012085e+06

Basic statistical analysis for every feature

(mean, std, min, max):



uOttawa

# Dataset's overview (EDA)

## Test Set

	Received Packets	Received Bytes	Sent Bytes	Sent Packets	Port alive Duration (s)	Delta Received Packets	Delta Received Bytes	Delta Sent Bytes	Delta Sent Packets	Delta Port alive Duration (s)	Connection Point
count	561.000000	5.610000e+02	5.610000e+02	561.000000	561.000000	561.000000	5.610000e+02	5.610000e+02	561.000000	561.000000	561.000000
mean	23141.636364	2.726677e+07	2.438093e+07	30155.998217	910.654189	149.048128	4.546660e+05	3.171120e+05	169.782531	4.841355	2.44385
std	69278.589077	3.627827e+07	3.439084e+07	82370.302150	981.303212	861.268543	1.381151e+06	1.207563e+06	1030.484551	0.365671	1.20006
min	10.000000	8.560000e+02	6.854000e+03	49.000000	26.000000	0.000000	0.000000e+00	2.780000e+02	2.000000	4.000000	1.00000
25%	353.000000	1.042010e+05	4.480100e+04	322.000000	136.000000	2.000000	2.780000e+02	2.800000e+02	2.000000	5.000000	1.00000
50%	1376.000000	1.267023e+07	1.262299e+07	1215.000000	256.000000	4.000000	5.560000e+02	5.560000e+02	4.000000	5.000000	2.00000
75%	3562.000000	3.809469e+07	3.176783e+07	4054.000000	1742.000000	19.000000	1.823000e+03	7.590000e+02	5.000000	5.000000	3.00000
max	352572.000000	2.589394e+08	2.138743e+08	420932.000000	3307.000000	11130.000000	6.323770e+06	6.647966e+06	13840.000000	5.000000	5.00000
	Total Load/Rate	Total Load/Latest	Unknown Load/Rate	Unknown Load/Latest	Latest bytes counter	Active Flow Entries	Packets Looked Up	Packets Matched			
5.610000e+02	5.610000e+02	5.610000e+02	5.610000e+02	5.610000e+02	5.610000e+02	561.000000	5.610000e+02	5.610000e+02			
1.992058e+04	6.084924e+05	1.992058e+04	6.084924e+05	1.992058e+04	61.912656	9.053312e+04	9.041352e+04				
1.105112e+05	1.985467e+06	1.105112e+05	1.985467e+06	1.105112e+05	601.548185	2.235338e+05	2.235324e+05				
-4.042080e+05	0.000000e+00	-4.042080e+05	0.000000e+00	-4.042080e+05	4.000000	1.790000e+02	9.800000e+01				
0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	5.000000	2.346000e+03	2.238000e+03				
0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	6.000000	7.066000e+03	6.929000e+03				
0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	8.000000	1.963300e+04	1.940900e+04				
1.194922e+06	1.264637e+07	1.194922e+06	1.264637e+07	1.194922e+06	9980.000000	1.009529e+06	1.009394e+06				

Basic statistical analysis for every feature

(mean, std, min, max):



uOttawa

# Dataset's overview (EDA)

## Validation Set

	Received Packets	Received Bytes	Sent Bytes	Sent Packets	Port alive Duration (s)	Delta Received Packets	Delta Received Bytes	Delta Sent Bytes	Delta Sent Packets	Delta Port alive Duration (s)	Connection Point
count	561.000000	5.610000e+02	5.610000e+02	561.000000	561.000000	561.000000	5.610000e+02	5.610000e+02	561.000000	561.000000	561.000000
mean	20265.901961	2.651052e+07	2.462073e+07	26309.124777	949.360071	238.839572	3.268137e+05	2.927781e+05	162.711230	4.846702	2.422460
std	59032.911054	3.505715e+07	3.420225e+07	75202.041514	977.609275	1149.069110	1.181636e+06	1.106077e+06	909.905479	0.360596	1.150215
min	10.000000	8.560000e+02	7.202000e+03	50.000000	36.000000	0.000000	0.000000e+00	2.780000e+02	2.000000	4.000000	1.000000
25%	379.000000	1.962210e+06	4.856800e+04	377.000000	141.000000	2.000000	2.780000e+02	2.800000e+02	2.000000	5.000000	1.000000
50%	1385.000000	1.266395e+07	1.263028e+07	1239.000000	317.000000	4.000000	5.560000e+02	5.560000e+02	4.000000	5.000000	2.000000
75%	3642.000000	3.785762e+07	3.176313e+07	3889.000000	1750.000000	5.000000	6.260000e+02	7.590000e+02	5.000000	5.000000	3.000000
max	350280.000000	2.652568e+08	1.837435e+08	419567.000000	3287.000000	11273.000000	6.249706e+06	6.302708e+06	11273.000000	5.000000	5.000000

Basic statistical analysis for every feature

(mean, std, min, max):

Total Load/Rate	Total Load/Latest	Unknown Load/Rate	Unknown Load/Latest	Latest bytes counter	Active Flow Entries	Packets Looked Up	Packets Matched
5.610000e+02	5.610000e+02	5.610000e+02	5.610000e+02	5.610000e+02	561.000000	5.610000e+02	5.610000e+02
1.349392e+04	4.418708e+05	1.349392e+04	4.418708e+05	1.349392e+04	62.071301	8.382923e+04	8.370799e+04
9.124726e+04	1.725488e+06	9.124726e+04	1.725488e+06	9.124726e+04	601.962008	2.056357e+05	2.056371e+05
0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	4.000000	1.400000e+02	8.400000e+01
0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	5.000000	2.462000e+03	2.350000e+03
0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	6.000000	8.105000e+03	7.979000e+03
0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	8.000000	2.073500e+04	2.060700e+04
1.260664e+06	1.273210e+07	1.260664e+06	1.273210e+07	1.260664e+06	9979.000000	1.009010e+06	1.008875e+06

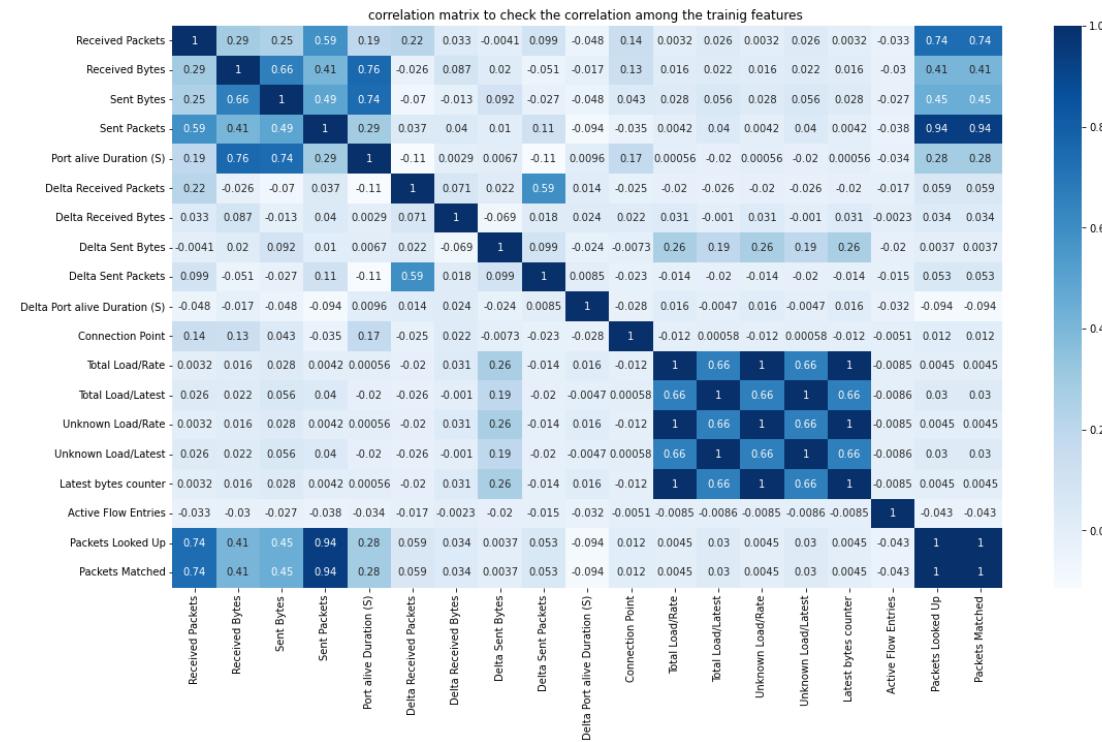
# Dataset's overview (EDA)

Information of train, test, and validation:

Train Set			Test Set			Validation Set						
#	Column	Non-Null Count	Dtype	#	Column	Non-Null Count	Dtype	#	Column	Non-Null Count	Dtype	
0	Switch ID	2619	non-null	object	0	Switch ID	561	non-null	object	561	non-null	object
1	Port Number	2619	non-null	object	1	Port Number	561	non-null	object	561	non-null	object
2	Received Packets	2619	non-null	int64	2	Received Packets	561	non-null	int64	561	non-null	int64
3	Received Bytes	2619	non-null	int64	3	Received Bytes	561	non-null	int64	561	non-null	int64
4	Sent Bytes	2619	non-null	int64	4	Sent Bytes	561	non-null	int64	561	non-null	int64
5	Sent Packets	2619	non-null	int64	5	Sent Packets	561	non-null	int64	561	non-null	int64
6	Port alive Duration (S)	2619	non-null	int64	6	Port alive Duration (S)	561	non-null	int64	561	non-null	int64
7	Delta Received Packets	2619	non-null	int64	7	Delta Received Packets	561	non-null	int64	561	non-null	int64
8	Delta Received Bytes	2619	non-null	int64	8	Delta Received Bytes	561	non-null	int64	561	non-null	int64
9	Delta Sent Bytes	2619	non-null	int64	9	Delta Sent Bytes	561	non-null	int64	561	non-null	int64
10	Delta Sent Packets	2619	non-null	int64	10	Delta Sent Packets	561	non-null	int64	561	non-null	int64
11	Delta Port alive Duration (S)	2619	non-null	int64	11	Delta Port alive Duration (S)	561	non-null	int64	561	non-null	int64
12	Connection Point	2619	non-null	int64	12	Connection Point	561	non-null	int64	561	non-null	int64
13	Total Load/Rate	2619	non-null	int64	13	Total Load/Rate	561	non-null	int64	561	non-null	int64
14	Total Load/Latest	2619	non-null	int64	14	Total Load/Latest	561	non-null	int64	561	non-null	int64
15	Unknown Load/Rate	2619	non-null	int64	15	Unknown Load/Rate	561	non-null	int64	561	non-null	int64
16	Unknown Load/Latest	2619	non-null	int64	16	Unknown Load/Latest	561	non-null	int64	561	non-null	int64
17	Latest bytes counter	2619	non-null	int64	17	Latest bytes counter	561	non-null	int64	561	non-null	int64
18	Active Flow Entries	2619	non-null	int64	18	Active Flow Entries	561	non-null	int64	561	non-null	int64
19	Packets Looked Up	2619	non-null	int64	19	Packets Looked Up	561	non-null	int64	561	non-null	int64
20	Packets Matched	2619	non-null	int64	20	Packets Matched	561	non-null	int64	561	non-null	int64
21	Label	2619	non-null	object	21	Label	561	non-null	object	561	non-null	object

# Dataset's overview (EDA)

Check the correlation among the variables: (Train Set)

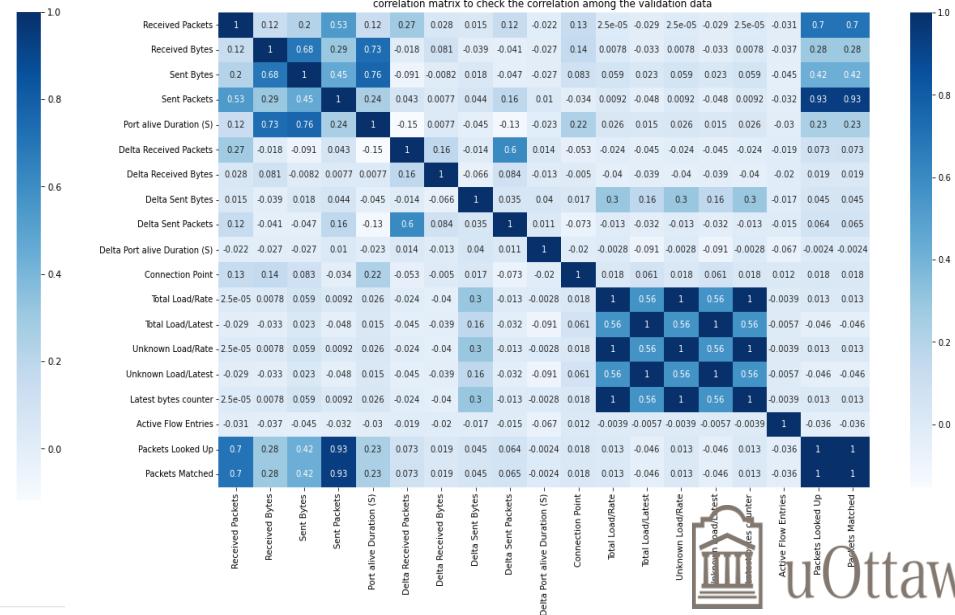


# Dataset's overview (EDA)

Check the correlation among the variables:

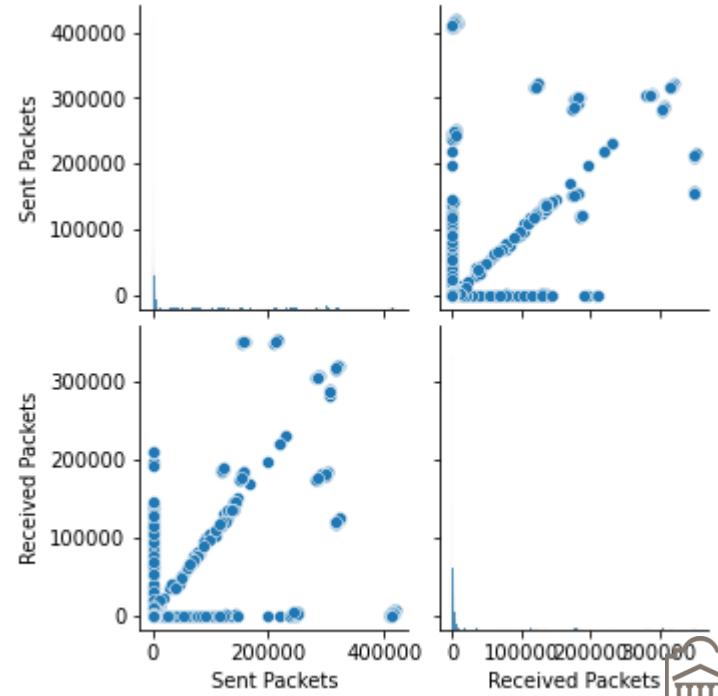
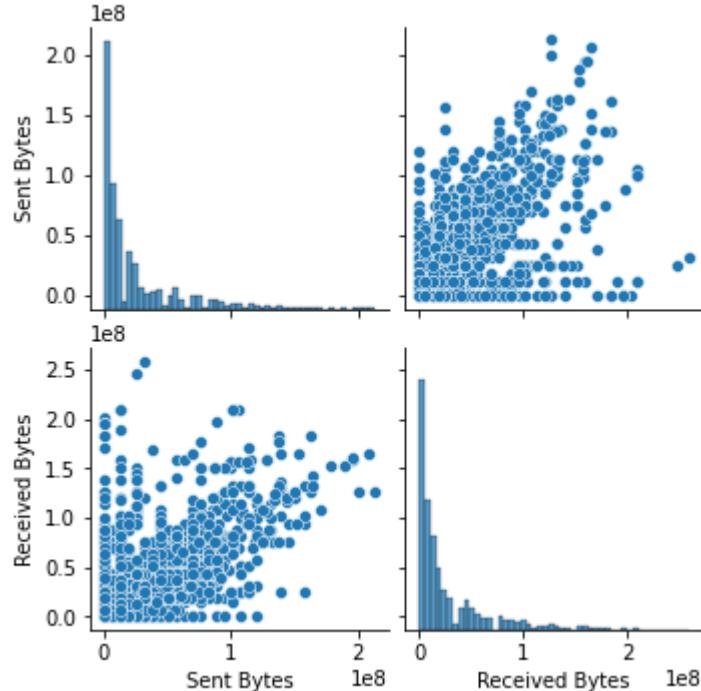
Test Set:

		correlation matrix to check the correlation among the testing data																		
		Received Packets	Received Bytes-	Sent Bytes-	Sent Packets-	Port alive Duration (S)-	Delta Received Packets-	Delta Received Bytes-	Delta Sent Packets-	Delta Sent Bytes-	Delta Port alive Duration (S)-	Connection Point-	Total Load/Rate-	Total Load/Latest-	Unknown Load/Rate-	Unknown Load/Latest-	Latest bytes counter-	Active Flow Entries-	Packets Looked Up-	Packets Matched-
Received Packets	1	0.2	0.21	0.66	0.14	0.21	0.021	0.033	0.12	-0.057	0.086	0.12	0.062	0.12	0.062	0.12	-0.029	0.78	0.78	
Received Bytes-	0.2	1	0.69	0.4	0.74	0.0054	0.14	0.095	-0.038	0.018	0.12	0.043	-0.013	0.043	-0.013	0.043	-0.025	0.37	0.37	
Sent Bytes-	0.21	0.69	1	0.5	0.77	-0.062	-0.0079	0.14	-0.048	-0.046	0.049	0.057	0.014	0.057	0.014	0.057	-0.013	0.43	0.43	
Sent Packets-	0.66	0.4	0.5	1	0.27	0.11	-0.0056	0.13	0.12	-0.12	-0.07	0.16	0.059	0.16	0.059	0.16	-0.032	0.94	0.94	
Port alive Duration (S)-	0.14	0.74	0.77	0.27	1	-0.11	0.029	0.06	-0.11	0.069	0.19	0.0011	-0.075	0.0011	-0.075	0.0011	-0.014	0.26	0.26	
Delta Received Packets-	0.21	-0.0054	-0.062	0.11	-0.11	1	0.13	0.00078	0.66	0.062	-0.012	-0.016	-0.031	-0.016	-0.031	-0.016	-0.015	0.082	0.082	
Delta Received Bytes-	0.021	0.14	-0.0079	-0.0056	0.029	0.13	1	-0.083	0.008	0.048	0.027	0.02	0.022	0.02	0.022	0.02	-0.019	0.025	0.025	
Delta Sent Bytes-	0.033	0.095	0.14	0.13	0.06	0.00078	-0.083	1	0.2	-0.12	0.021	0.26	0.14	0.26	0.14	0.26	-0.024	0.098	0.098	
Delta Sent Packets-	0.12	-0.038	-0.048	0.12	-0.11	0.66	0.008	0.2	1	-0.057	-0.021	-0.016	-0.031	-0.016	-0.031	-0.016	-0.015	0.058	0.058	
Delta Port alive Duration (S)-	-0.057	0.018	-0.046	-0.12	0.069	0.062	0.048	-0.12	-0.057	1	0.043	-0.022	-0.029	-0.022	-0.029	-0.022	0.029	-0.11	-0.11	
Connection Point-	0.086	0.12	0.049	-0.07	0.19	-0.012	0.027	0.021	-0.021	0.043	1	0.0029	0.0077	0.0029	0.0077	0.0029	0.047	-0.021	-0.021	
Total Load/Rate-	0.12	0.043	0.057	0.16	0.0011	-0.016	0.02	0.26	-0.016	-0.022	0.0029	1	0.6	1	0.6	1	-0.016	0.15	0.15	
Total Load/Latest-	0.062	-0.013	0.014	0.059	0.075	-0.031	0.022	0.14	-0.031	-0.029	0.0077	0.6	1	0.6	1	0.6	-0.022	0.052	0.052	
Unknown Load/Rate-	0.12	0.043	0.057	0.16	0.0011	-0.016	0.02	0.26	-0.016	-0.022	0.0029	1	0.6	1	0.6	1	-0.016	0.15	0.15	
Unknown Load/Latest-	0.062	-0.013	0.014	0.059	0.075	-0.031	0.022	0.14	-0.031	-0.029	0.0077	0.6	1	0.6	1	0.6	-0.022	0.052	0.052	
Latest bytes counter-	0.12	0.043	0.057	0.16	0.0011	-0.016	0.02	0.26	-0.016	-0.022	0.0029	1	0.6	1	0.6	1	-0.016	0.15	0.15	
Active Flow Entries-	-0.029	-0.025	-0.013	-0.032	-0.014	0.015	-0.019	-0.024	-0.015	0.029	0.047	-0.016	-0.022	-0.016	-0.022	-0.016	1	0.035	0.035	
Packets Looked Up-	0.78	0.37	0.43	0.94	0.26	0.082	0.025	0.098	0.058	-0.11	-0.021	0.15	0.052	0.15	0.052	0.15	-0.035	1	1	
Packets Matched-	0.78	0.37	0.43	0.94	0.26	0.082	0.025	0.098	0.058	-0.11	-0.021	0.15	0.052	0.15	0.052	0.15	-0.035	1	1	



# Dataset's overview (EDA)

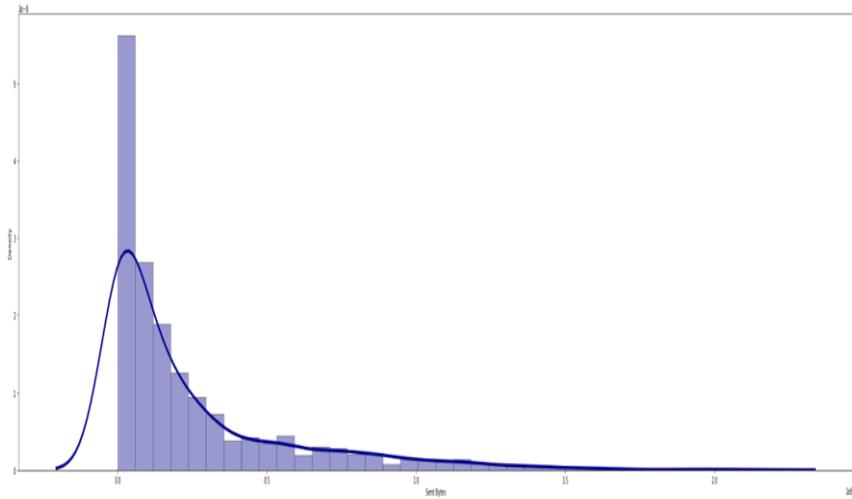
The distribution of variables and relationships between two variables:



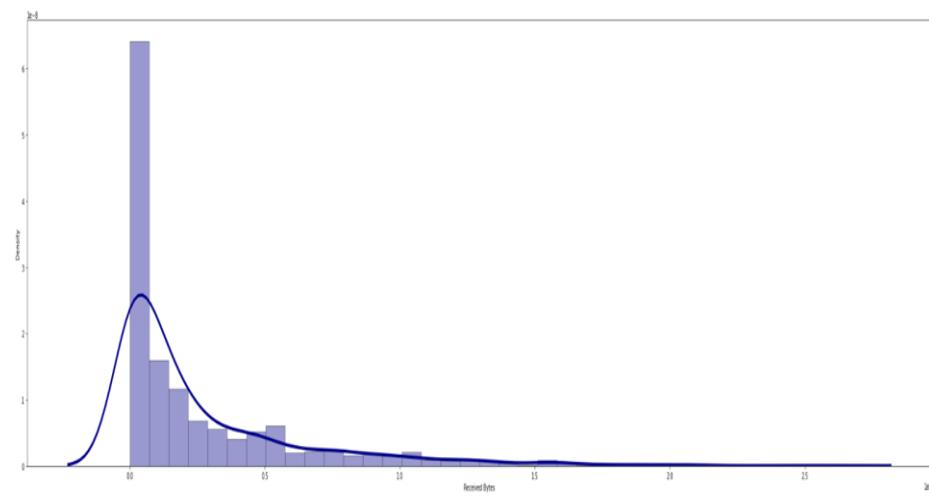
# Dataset's overview (EDA)

The distribution of dataset

(Sent Bytes)



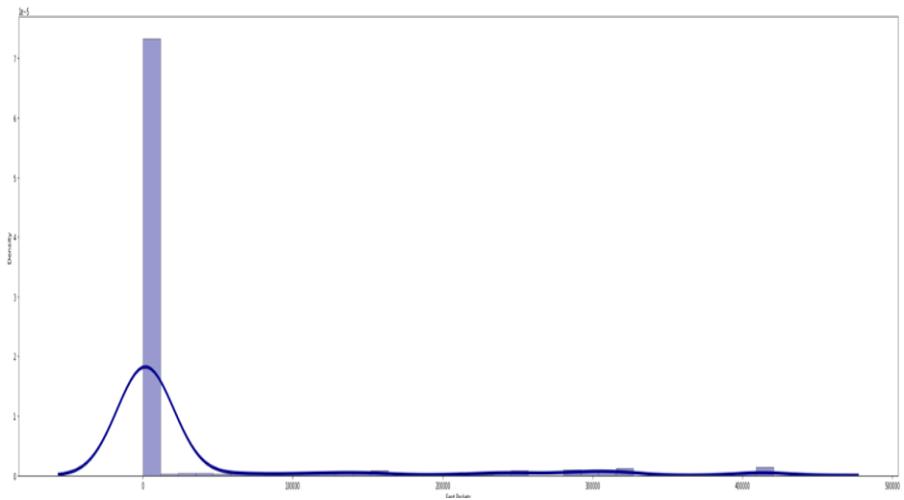
(Received Bytes)



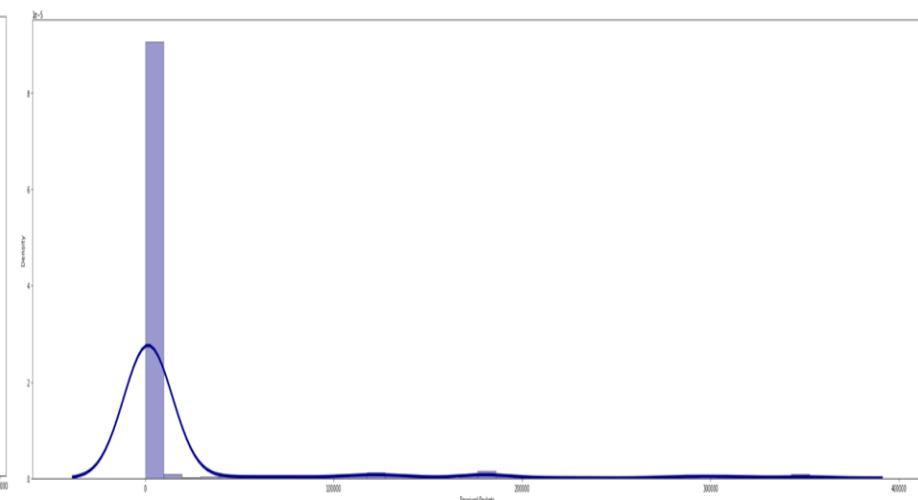
# Dataset's overview (EDA)

check the distribution of some features , each one alone , by using the distplot:

(Sent Packets)



(Received Packets)

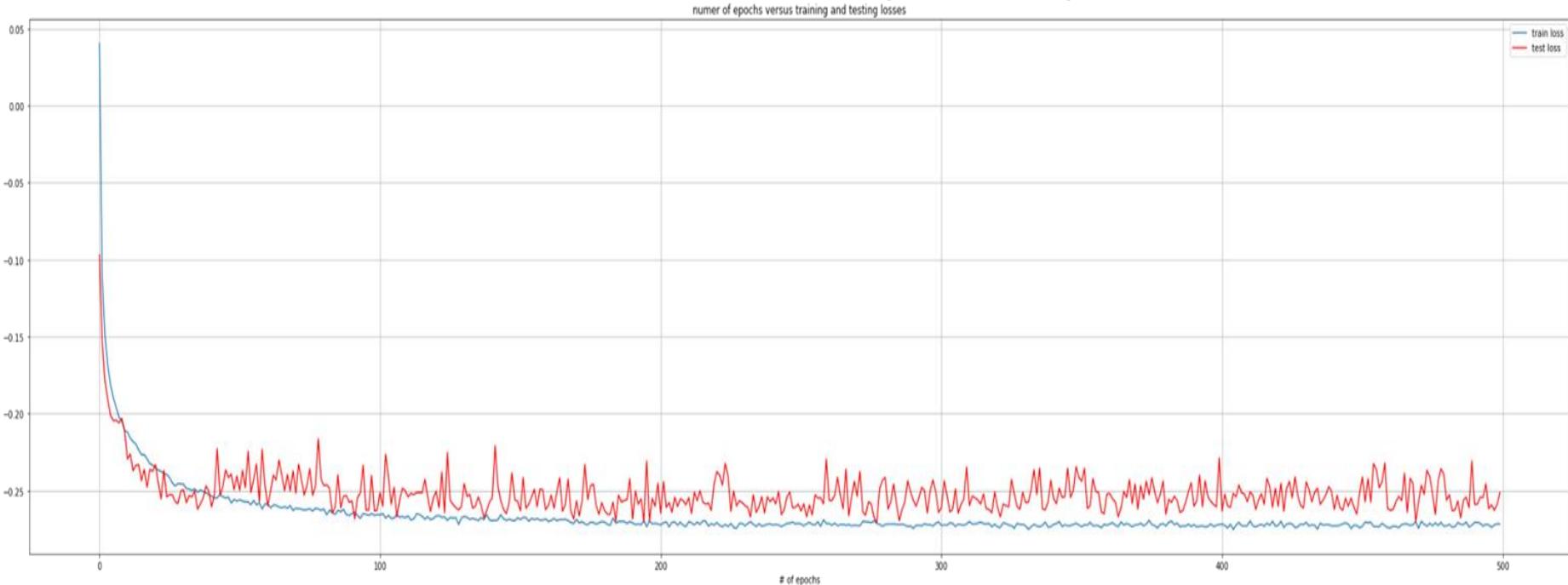


# Data Preprocessing

- **LabelEncoder()**  
(switch ID, and port number.)
- **to\_categorical()** ( labels ) to use in model.
- **StandardScaler()** on train Set, valSet, and testSet.

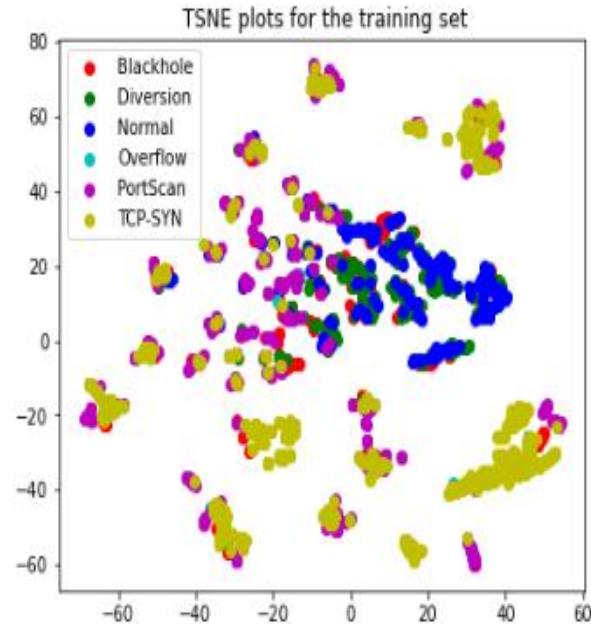
# Q1) Obtain a baseline performance

Number of epochs versus training and testing losses

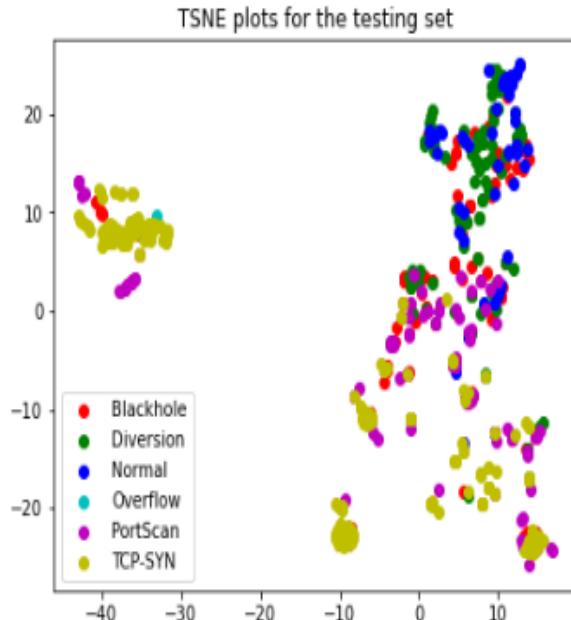


# Q1) Obtain a baseline performance (T-SNE Plots)

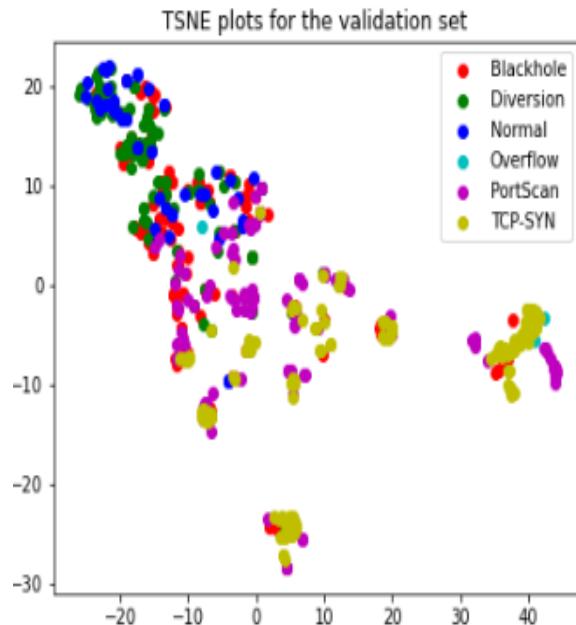
Training set T-SNE plot



Testing set T-SNE plot



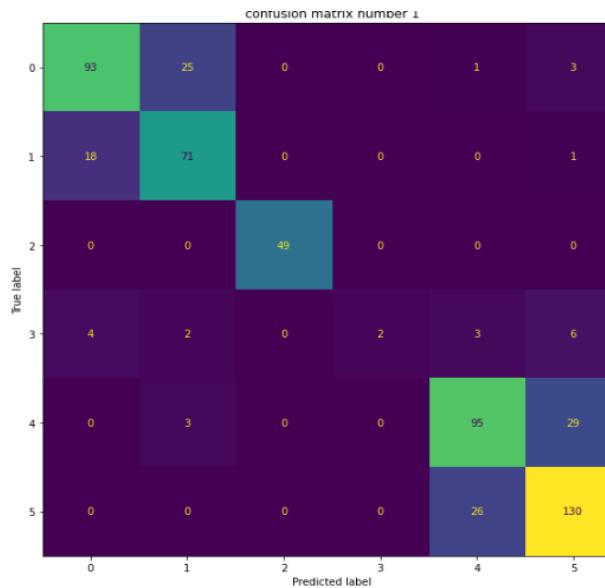
Validation set T-SNE plot



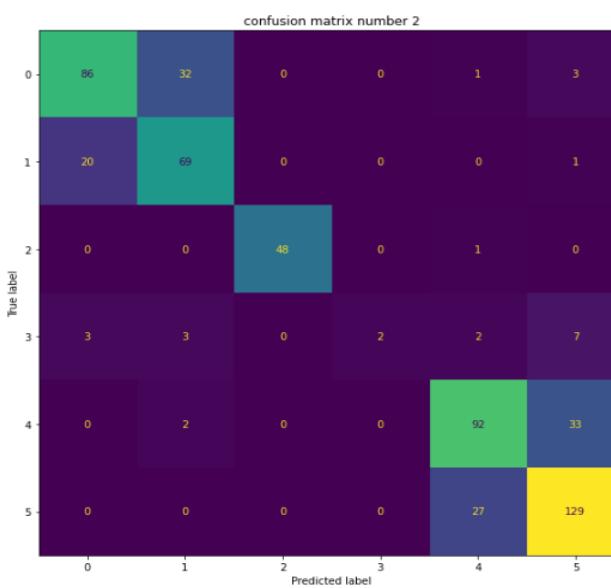
# Q1) Obtain a baseline performance ( Confusion Matrix)

	Max training acc ()%	Max test acc ()%	Min training acc ()%	Min test acc ()%	Avg training acc ()%	Avg test acc ()%
0	79.610538	78.609627	78.006876	75.935829	78.770524	77.611408

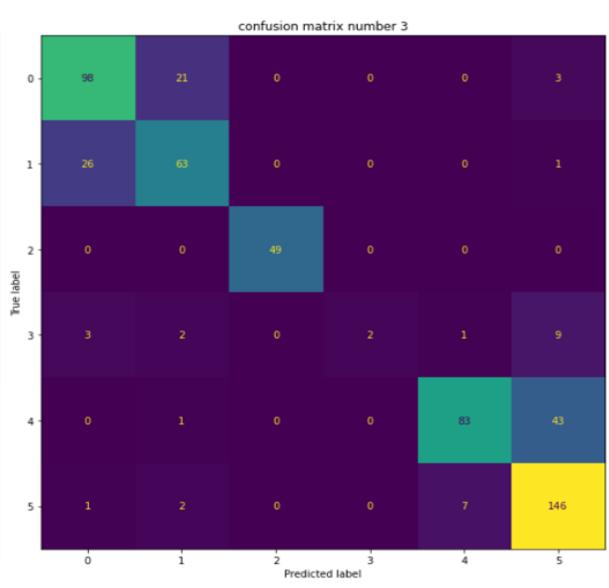
Confusion Matrix 1



Confusion Matrix 2

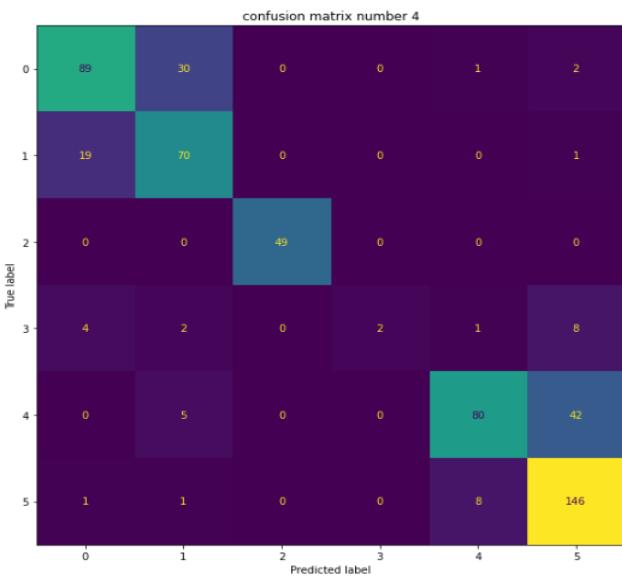


Confusion Matrix 3

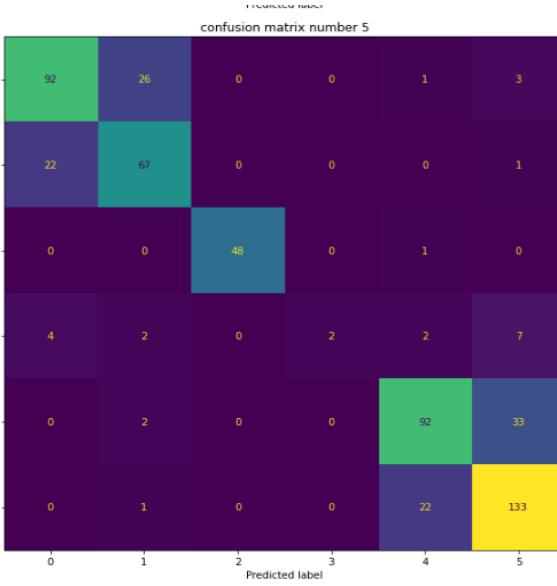


# Q1) Obtain a baseline performance ( Confusion Matrix)

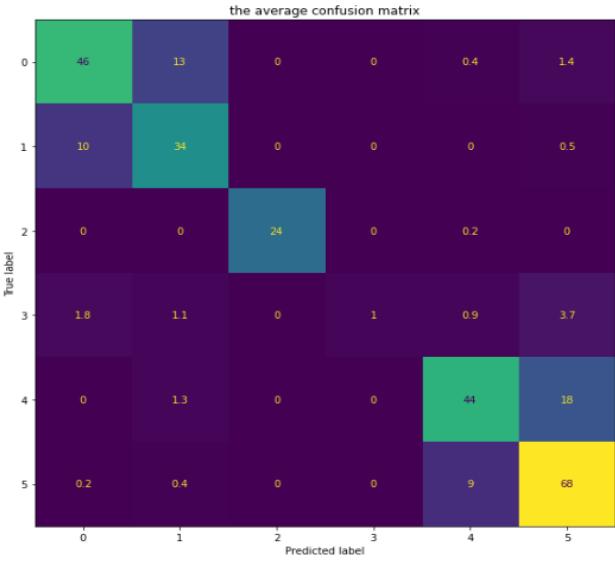
Confusion Matrix 4



Confusion Matrix 5



Average Confusion Matrix



## Q2) Compare dimensionality reduction to feature selection

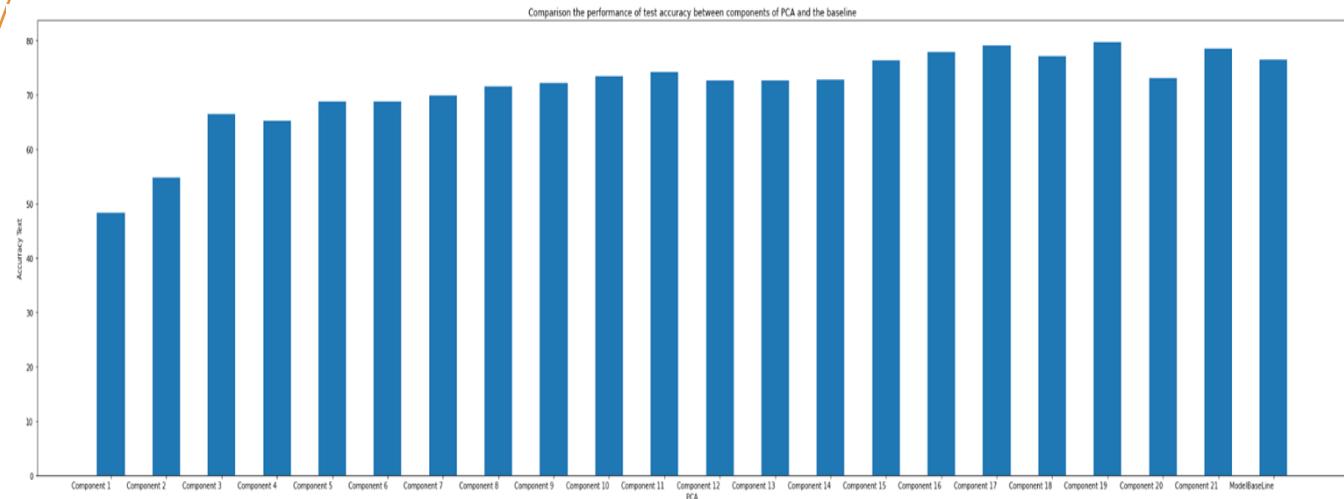
### Q2.1) Dimensionality reduction (based on the MLP test accuracy)

The value of components on the test accuracy of the MLP classifier, using PCA:

1	48.306596
2	54.723710
3	66.488415
4	65.240639
5	68.805707
6	68.805707
7	69.875222
8	71.479499
9	72.192514
10	73.440284
11	74.153298
12	72.549021
13	72.549021
14	72.727275
15	76.292336
16	77.896613
17	79.144382
18	77.005345
19	79.679143
20	73.083776
21	78.431374

**19 components** is the best value using PCA: 79.679%

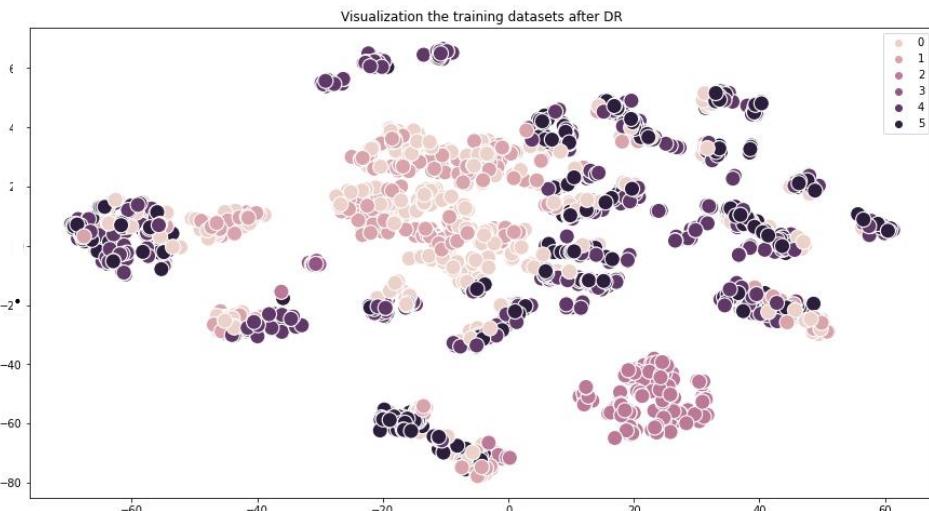
The (number of component-test accuracy) graph with the baseline performance:



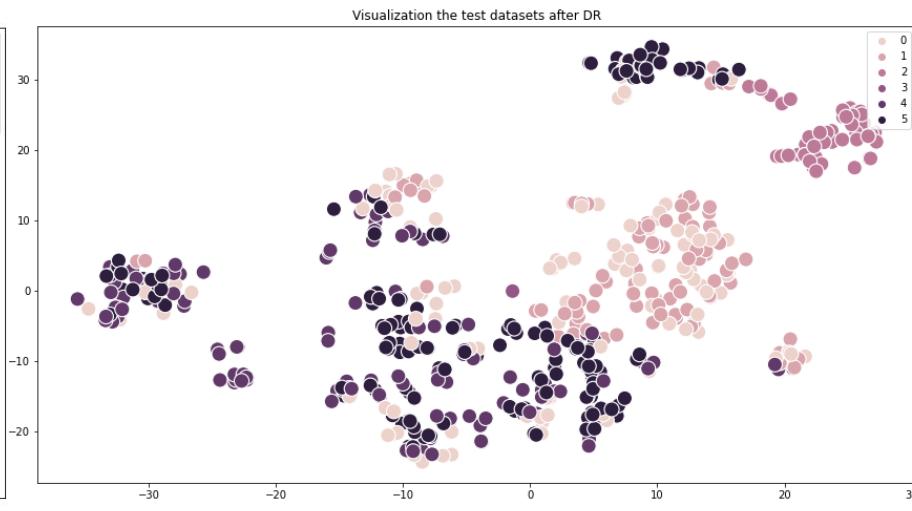
## Q2) Compare dimensionality reduction to feature selection

### Q2.1) Dimensionality reduction (based on the MLP test accuracy)

Visualise the training dataset after DR (TSNE)



Visualise the test dataset after DR (TSNE)



## Q2) Compare dimensionality reduction to feature selection

### Q2.2) Feature selection (based on the MLP test accuracy)

#### ANOVA

1	57.040995
2	57.754010
3	58.467025
4	71.479499
5	71.479499
6	70.944744
7	70.944744
8	74.331552
9	76.827097
10	75.757575
11	78.609627
12	77.183598
13	76.470590
14	77.183598
15	77.540106
16	76.648843
17	77.718359
18	73.975044
19	74.866313
20	73.975044
21	75.757575

11 features is the best number of features using ANOVA : 78.609%

#### Mutual Information

1	46.345812
2	71.657753
3	71.479499
4	72.905529
5	70.766491
6	71.479499
7	73.618537
8	73.262030
9	71.122992
10	74.331552
11	73.618537
12	71.479499
13	73.618537
14	71.122992
15	74.331552
16	68.449199
17	72.549021
18	77.005345
19	78.253120
20	75.757575
21	77.183598

19 features is the best number of features using MI: 78.253%



## Q2) Compare dimensionality reduction to feature selection

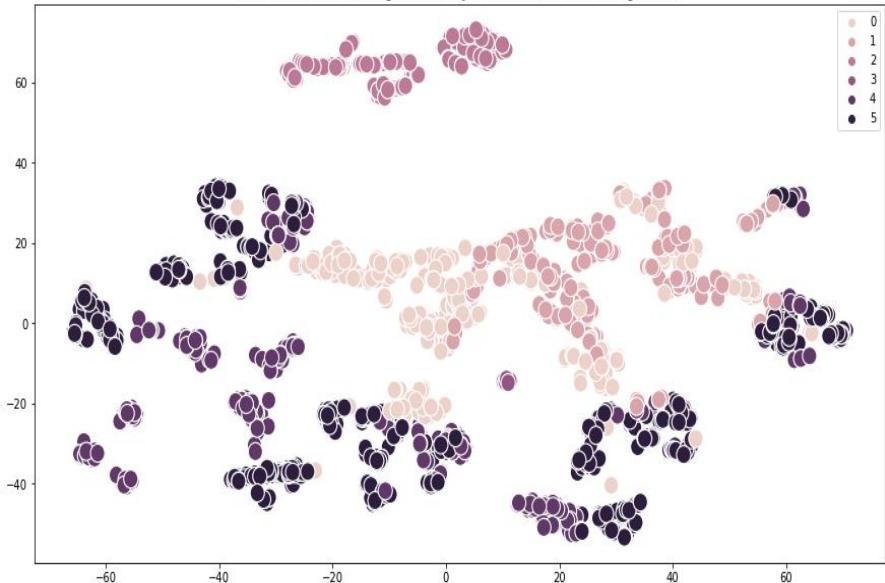
### Q2.2) Feature selection (based on the MLP test accuracy)

The method that achieves the best test accuracy results:

**ANOVA (78.609%) by using 11 features**

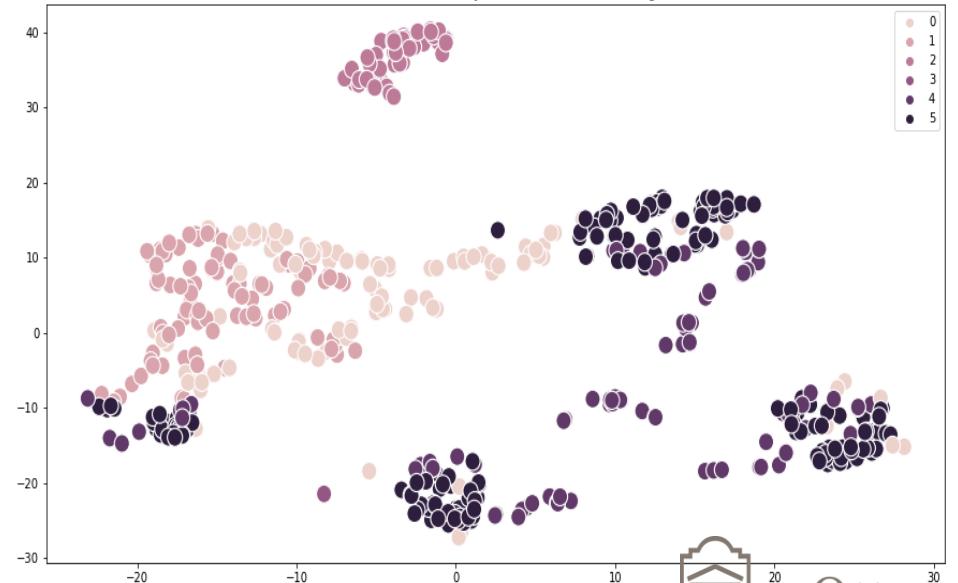
Visualization the training dataset using ANOVA (TSNE)

Visualization the training datasets by the best 11 features using ANOVA



Visualization the test dataset using ANOVA (TSNE)

Visualization test datasets by the best 11 features using ANOVA



## Q2) Compare dimensionality reduction to feature selection

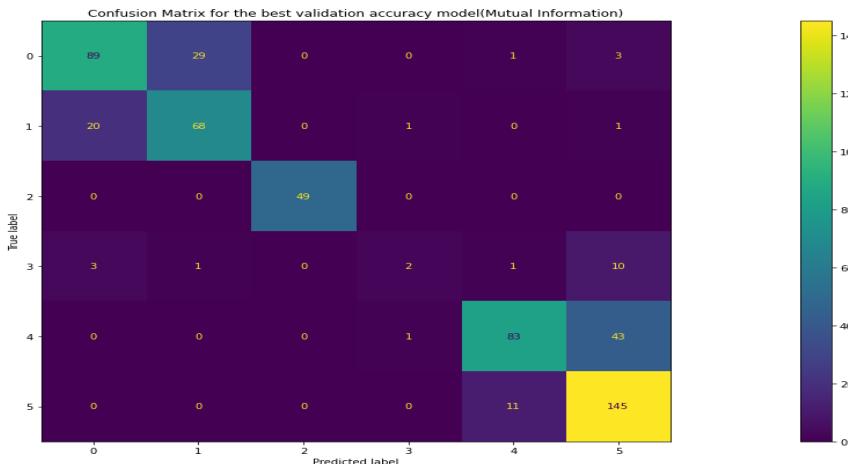
### Q2.2) Feature selection (based on the MLP test accuracy)

	Validation Accuracy	The Predication		
PCA	78.431374	[5, 2, 4, 0, 5, 1, 4, 5, 0, 4, 0, 4, 1, 4, 5, ...]		
ANOVA	77.361852	[5, 2, 5, 0, 5, 0, 5, 5, 0, 4, 0, 4, 1, 4, 5, ...]		
Mutual Information	78.787881	[5, 2, 4, 0, 5, 0, 4, 5, 0, 4, 0, 4, 1, 4, 5, ...]	The highest validation accuracy	The Predication

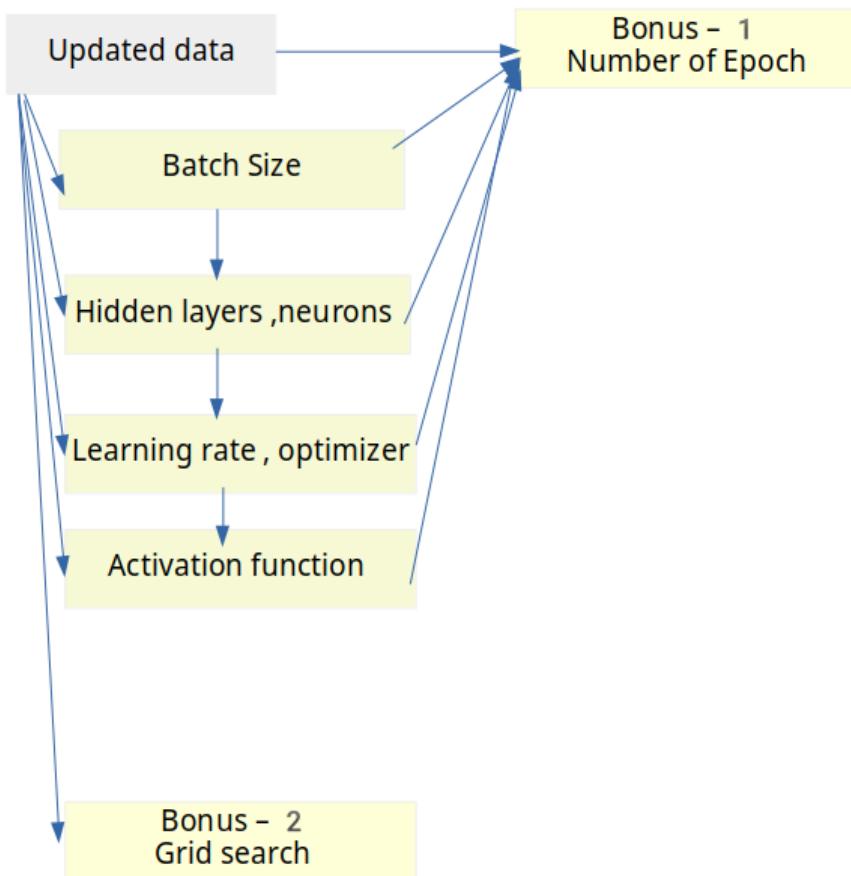
Mutual Information



The confusion matrix the highest validation accuracy:



### Q3) Vary the MLP parameters



**Experiment 1 :**  
examine the impact of using different Batch sizes

**Experiment 2 :**  
examine the impact of using different number of hidden layers and number of neurons in each layer

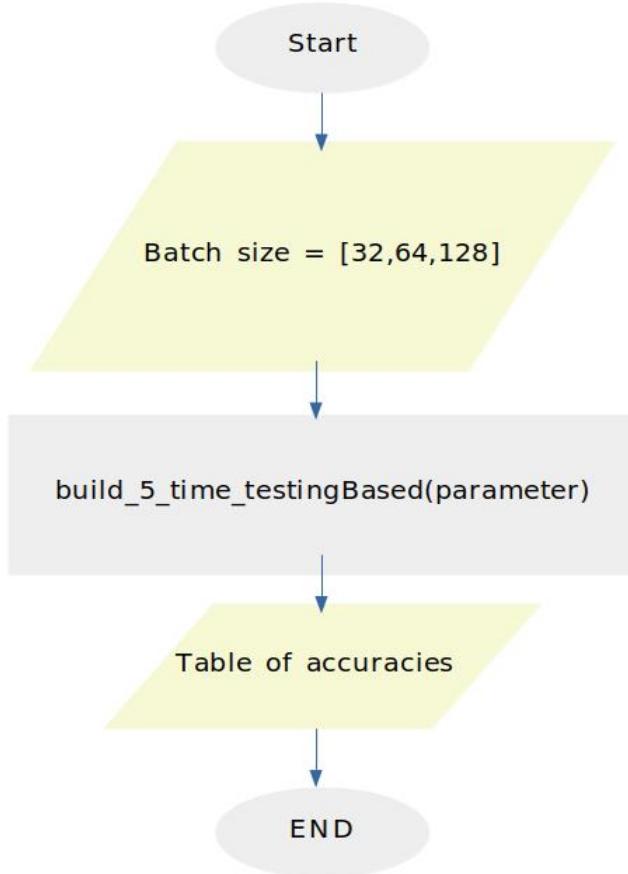
**Experiment 3 :**  
examine the impact of using different learning rates and different optimizers

**Experiment 4 :**  
examine the impact of using different activation functions

**Bonus 1 :**  
examine the impact of using different number of epochs with each best combination

**Bonus 2 : Grid search**

# Q3) Vary the MLP parameters [1/5]



**Experiment 1 :**  
examine the impact of using  
different Batch sizes

### Q3) Vary the MLP parameters [1/5] :

Table of results

Batch size= 32					
Max train acc (%)	Max test acc (%)	Min train acc (%)	Min test acc (%)	Avg train acc (%)	Avg test acc (%)
79.343259	78.966135	78.388697	74.866313	78.808706	77.290554

Batch size= 64					
Max train acc (%)	Max test acc (%)	Min train acc (%)	Min test acc (%)	Avg train acc (%)	Avg test acc (%)
79.572356	80.926919	78.465062	75.757575	78.930889	77.611408

Batch size= 128					
Max train acc (%)	Max test acc (%)	Min train acc (%)	Min test acc (%)	Avg train acc (%)	Avg test acc (%)
78.274149	78.431374	77.472317	76.292336	77.815961	77.290553

Highlight the best combination

Batch size = 32  
average training accuracy = 78.81%  
average testing accuracy = 77.29%

Batch size = 64  
average training accuracy = 78.93%  
average testing accuracy = 77.61%

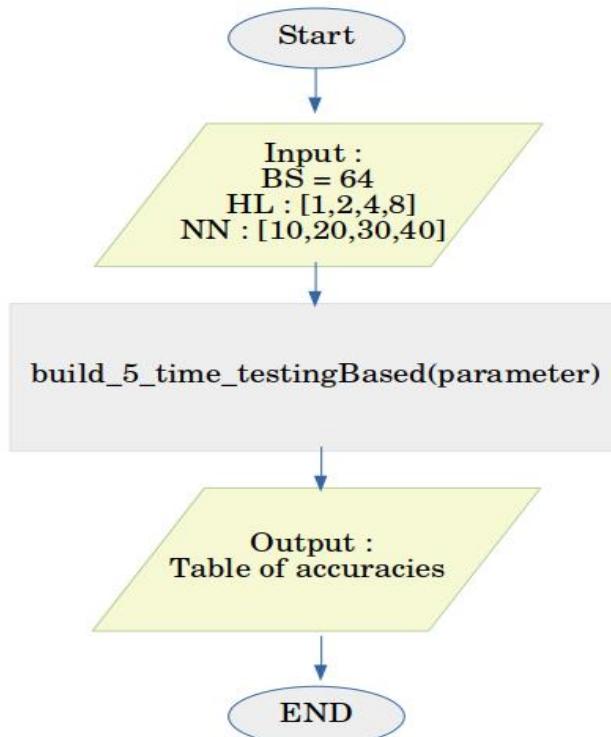
Batch size = 128  
average training accuracy = 77.82%  
average testing accuracy = 77.29%

===== the best combination is : with Batch size = 64



### Q3) Vary the MLP parameters [2/5]

Experiment 2 : examine the impact of using different number of hidden layers and number of neurons in each layer



Batch size = 32  
average validation accuracy = 76.54%

-----  
Batch size = 64  
average validation accuracy = 77.11%

-----  
Batch size = 128  
average validation accuracy = 77.11%

=====  
the best combination Based on the average validation is : with Batch size = 64

**The best combination that achieves the highest average validation accuracy  
From the previous question Q3[1 / 5]**

Figure: Flow chart for this Experiment

### Q3) Vary the MLP parameters [2/5]

	1 hidden layer						4 hidden layer								
	Max train acc (%)	Max test acc (%)	Min train acc (%)	Min test acc (%)	Avg train acc (%)	Avg test acc (%)	Max train acc (%)	Max test acc (%)	Min train acc (%)	Min test acc (%)	Avg train acc (%)	Avg test acc (%)			
of neurons for each layer	10	79.381442	79.679143	76.708668	73.262030	77.953417	76.541889	of neurons for each layer	10	82.550591	80.926919	81.252384	76.648843	81.618937	78.787879
	20	80.908746	77.896613	79.839635	75.222814	80.542192	76.648841		20	87.972510	81.283420	85.185188	75.935829	86.964490	78.467022
	30	83.237875	79.144382	81.748760	75.935829	82.436043	78.039216		30	91.370749	81.283420	90.263462	75.401068	90.698739	77.967914
	40	83.772433	79.500890	81.863308	76.470590	83.062238	78.573976		40	93.203515	81.996435	91.103476	76.827097	92.348225	79.037434
2 hidden layer															
of neurons for each layer	Max train acc (%)	Max test acc (%)	Min train acc (%)	Min test acc (%)	Avg train acc (%)	Avg test acc (%)	Max train acc (%)	Max test acc (%)	Min train acc (%)	Min test acc (%)	Avg train acc (%)	Avg test acc (%)			
	10	82.054222	78.787881	78.808707	74.688059	80.267279	76.399288	of neurons for each layer	10	84.230620	81.283420	81.061476	78.074867	82.581137	79.714794
	20	85.261548	78.609627	83.046967	73.440284	84.032073	76.791443		20	88.507062	78.609627	87.285221	76.114082	87.873232	77.361854
	30	86.330658	81.996435	85.299730	79.679143	85.933563	80.855615		30	92.821687	79.679143	91.408932	74.153298	91.920580	78.074867
	40	87.896144	84.135473	86.445206	77.005345	86.880488	80.071301		40	95.418096	80.213904	92.821687	74.153298	94.326078	77.361853

### Q3) Vary the MLP parameters [2/5]

```
1 hidden layer , 10 neurons,77.95% training accuracy , 76.54% testing accuracy  
=====  
1 hidden layer , 20 neurons,80.54% training accuracy , 76.65% testing accuracy  
=====  
1 hidden layer , 30 neurons,82.44% training accuracy , 78.04% testing accuracy  
=====  
1 hidden layer , 40 neurons,83.06% training accuracy , 78.57% testing accuracy  
=====  
2 hidden layer , 10 neurons,80.27% training accuracy , 76.40% testing accuracy  
=====  
2 hidden layer , 20 neurons,84.03% training accuracy , 76.79% testing accuracy  
=====  
2 hidden layer , 30 neurons,85.93% training accuracy , 80.86% testing accuracy  
=====  
2 hidden layer , 40 neurons,86.88% training accuracy , 80.07% testing accuracy  
=====  
4 hidden layer , 10 neurons,81.62% training accuracy , 78.79% testing accuracy  
=====  
4 hidden layer , 20 neurons,86.96% training accuracy , 78.47% testing accuracy  
=====  
4 hidden layer , 30 neurons,90.70% training accuracy , 77.97% testing accuracy  
=====  
4 hidden layer , 40 neurons,92.35% training accuracy , 79.04% testing accuracy  
=====  
8 hidden layer , 10 neurons,82.58% training accuracy , 79.71% testing accuracy  
=====  
8 hidden layer , 20 neurons,87.87% training accuracy , 77.36% testing accuracy  
=====  
8 hidden layer , 30 neurons,91.92% training accuracy , 78.07% testing accuracy  
=====  
8 hidden layer , 40 neurons,94.33% training accuracy , 77.36% testing accuracy  
=====
```

-----  
best training accuracy 94.33%  
best testing accuracy 80.86%  
-----

best number of hidden layers :2  
best number of neuron in each layer :30

**Highlight the best combination  
Based on the average training  
and average testing accuracies**



uOttawa

## Q3) Vary the MLP parameters [2/5]

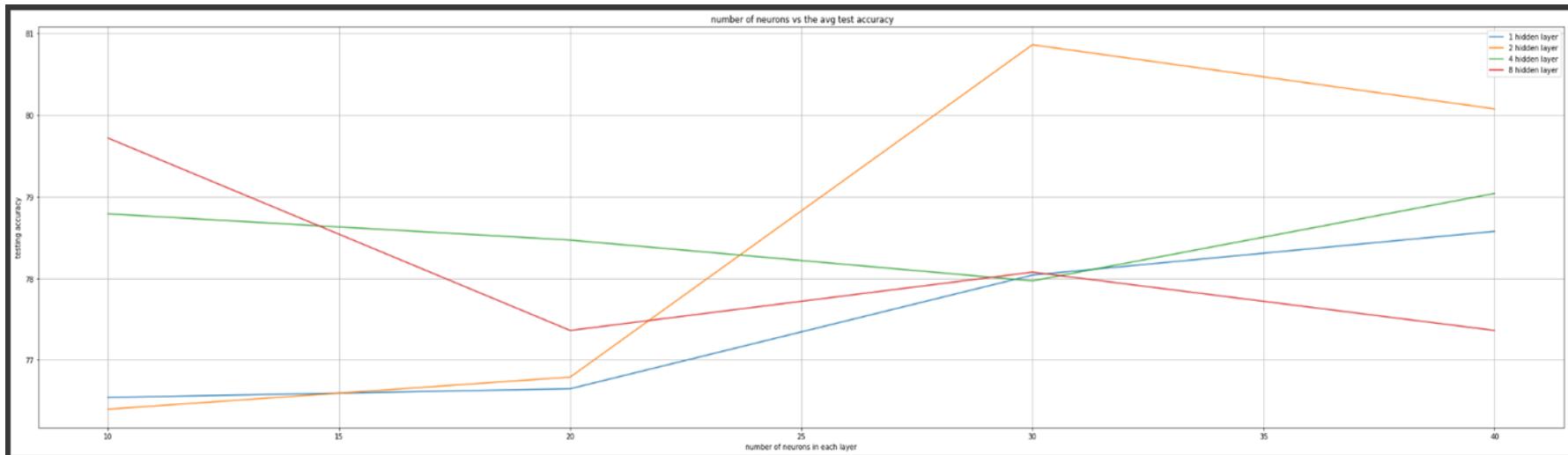


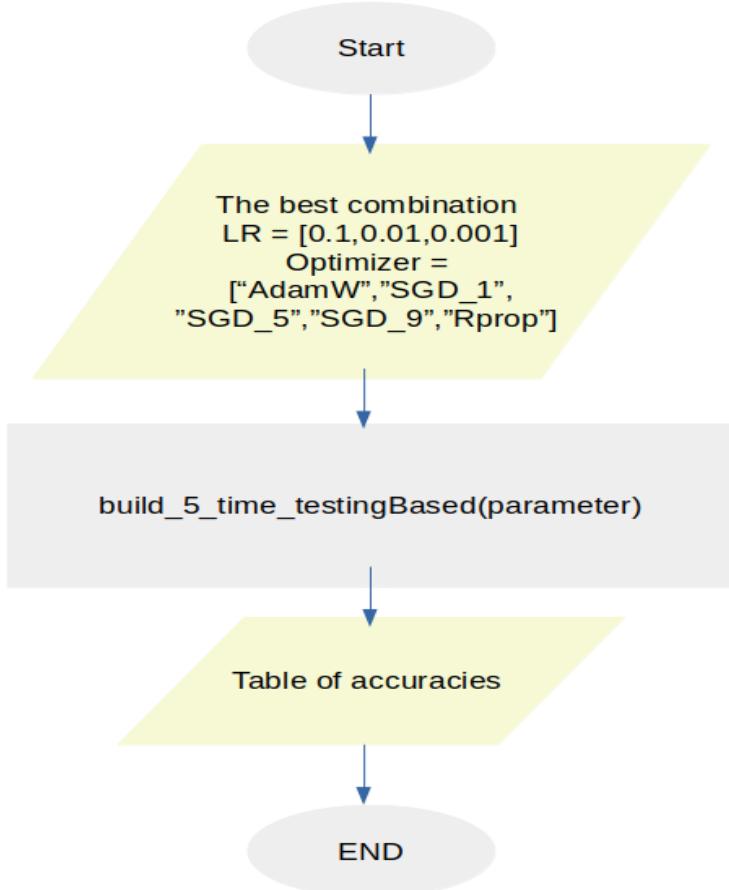
Figure : Plot the number of neurons in each layer versus the average testing accuracy

**X axis :** number of neurons in each layer

**Y axis :** average testing accuracy

The graph contains 4 lines :  
**1 hidden Layers**  
**2 hidden layers**  
**4 hidden layers**  
**8 hidden layers**

### Q3) Vary the MLP parameters [3/5]



**Experiment 3 :**  
**examine the impact of using  
different learning rates and  
different optimizers**

### Q3) Vary the MLP parameters [3/5]

1 hidden layer , 10 neurons, 75.76% validation accuracy

1 hidden layer , 20 neurons, 76.36% validation accuracy

1 hidden layer , 30 neurons, 76.51% validation accuracy

1 hidden layer , 40 neurons, 76.83% validation accuracy

2 hidden layer , 10 neurons, 76.65% validation accuracy

2 hidden layer , 20 neurons, 77.04% validation accuracy

2 hidden layer , 30 neurons, 78.40% validation accuracy

2 hidden layer , 40 neurons, 77.72% validation accuracy

4 hidden layer , 10 neurons, 77.90% validation accuracy

4 hidden layer , 20 neurons, 77.11% validation accuracy

4 hidden layer , 30 neurons, 78.04% validation accuracy

4 hidden layer , 40 neurons, 78.43% validation accuracy

8 hidden layer , 10 neurons, 78.22% validation accuracy

8 hidden layer , 20 neurons, 77.90% validation accuracy

8 hidden layer , 30 neurons, 78.47% validation accuracy

8 hidden layer , 40 neurons, 78.32% validation accuracy

-----  
highest average validation accuracy 78.47%

-----  
Based on the highest validation accuracy :

best number of hidden layers :8

best number of neuron in each layer :30

**Highlight the combination from the previous questions that achieves the highest average validation accuracy**

### Q3) Vary the MLP parameters [3/5]

		Learning rate=0.1						Learning rate=0.01					
		Max train acc (%)	Max test acc (%)	Min train acc (%)	Min test acc (%)	Avg train acc (%)	Avg test acc (%)	Max train acc (%)	Max test acc (%)	Min train acc (%)	Min test acc (%)	Avg train acc (%)	Avg test acc (%)
	AdamW	25.467736	27.807486	23.864070	21.746881	24.589538	24.527629	82.894236	80.748665	73.234057	69.518715	78.281788	74.723707
	SGD (momentum =0.1)	90.683466	81.639928	66.475755	71.657753	84.979000	77.290553	87.247038	80.213904	85.605192	78.431374	86.391752	79.358289
optimizer	SGD (momentum =0.5)	91.676211	79.322636	21.420389	21.746881	49.621993	44.420678	90.339822	82.174689	87.628865	75.401068	88.545246	78.823529
	SGD (momentum =0.9)	25.544101	27.807486	21.420389	21.746881	24.100802	23.493761	92.974418	80.926919	91.561663	75.757575	92.149676	78.823529
	Rprop	25.085911	28.163993	23.711340	22.638147	24.299351	26.844920	84.841543	81.461674	78.617793	76.648843	81.641848	79.964349

Table of results

### Q3) Vary the MLP parameters [3/5]

Learning rate=0.001						
	Max train acc ()%	Max test acc ()%	Min train acc ()%	Min test acc ()%	Avg train acc ()%	Avg test acc ()%
AdamW	93.241698	79.857397	91.714394	75.401068	92.386408	77.290552
SGD (momentum =0.1)	75.792289	75.757575	71.019471	71.122992	73.944254	73.654189
optimizer SGD (momentum =0.5)	80.526918	79.144382	76.937765	75.935829	78.823979	77.468804
SGD (momentum =0.9)	87.399769	80.570412	84.688812	76.114082	86.139748	78.431373
Rprop	93.661702	80.392158	91.332567	75.044560	92.386407	78.146167

Table of results.cont



### Q3) Vary the MLP parameters [3/5]

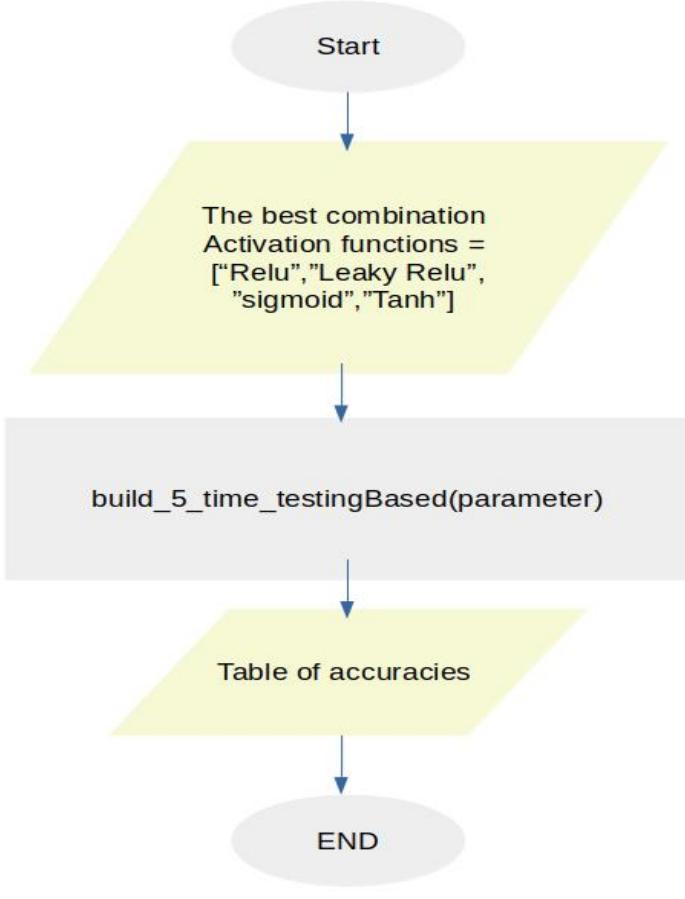
```
0.1 Learning Rate , adamW optimizer,24.59% training accuracy , 24.53% testing accuracy  
=====  
0.1 Learning Rate , SGD_1 optimizer,84.98% training accuracy , 77.29% testing accuracy  
=====  
0.1 Learning Rate , SGD_5 optimizer,49.62% training accuracy , 44.42% testing accuracy  
=====  
0.1 Learning Rate , SGD_9 optimizer,24.10% training accuracy , 23.49% testing accuracy  
=====  
0.1 Learning Rate , Rprop optimizer,24.30% training accuracy , 26.84% testing accuracy  
=====  
0.01 Learning Rate , adamW optimizer,78.28% training accuracy , 74.72% testing accuracy  
=====  
0.01 Learning Rate , SGD_1 optimizer,86.39% training accuracy , 79.36% testing accuracy  
=====  
0.01 Learning Rate , SGD_5 optimizer,88.55% training accuracy , 78.82% testing accuracy  
=====  
0.01 Learning Rate , SGD_9 optimizer,92.15% training accuracy , 78.82% testing accuracy  
=====  
0.01 Learning Rate , Rprop optimizer,81.64% training accuracy , 79.96% testing accuracy  
=====  
0.001 Learning Rate , adamW optimizer,92.39% training accuracy , 77.29% testing accuracy  
=====  
0.001 Learning Rate , SGD_1 optimizer,73.94% training accuracy , 73.65% testing accuracy  
=====  
0.001 Learning Rate , SGD_5 optimizer,78.82% training accuracy , 77.47% testing accuracy  
=====  
0.001 Learning Rate , SGD_9 optimizer,86.14% training accuracy , 78.43% testing accuracy  
=====  
0.001 Learning Rate , Rprop optimizer,92.39% training accuracy , 78.15% testing accuracy  
=====
```

-----  
best training accuracy 92.39%  
best testing accuracy 79.96%

-----  
best Learning Rate :0.01  
best optimizer :Rprop

**Highlight the best combination  
Based on the average training  
and average testing accuracies**

### Q3) Vary the MLP parameters [4/5]



**Experiment 4 :**  
**examine the impact of**  
**using different activation**  
**functions**

## Q3) Vary the MLP parameters [4/5]

```
0.1 Learning Rate , adamW optimizer, 23.92 % validation accuracy
=====
0.1 Learning Rate , SGD_1 optimizer, 77.79 % validation accuracy
=====
0.1 Learning Rate , SGD_5 optimizer, 46.74 % validation accuracy
=====
0.1 Learning Rate , SGD_9 optimizer, 24.56 % validation accuracy
=====
0.1 Learning Rate , Rprop optimizer, 22.64 % validation accuracy
=====
0.01 Learning Rate , adamW optimizer, 76.11 % validation accuracy
=====
0.01 Learning Rate , SGD_1 optimizer, 78.57 % validation accuracy
=====
0.01 Learning Rate , SGD_5 optimizer, 77.75 % validation accuracy
=====
0.01 Learning Rate , SGD_9 optimizer, 79.43 % validation accuracy
=====
0.01 Learning Rate , Rprop optimizer, 78.22 % validation accuracy
=====
0.001 Learning Rate , adamW optimizer, 77.65 % validation accuracy
=====
0.001 Learning Rate , SGD_1 optimizer, 72.44 % validation accuracy
=====
0.001 Learning Rate , SGD_5 optimizer, 75.04 % validation accuracy
=====
0.001 Learning Rate , SGD_9 optimizer, 78.43 % validation accuracy
=====
0.001 Learning Rate , Rprop optimizer, 78.89 % validation accuracy
=====
```

highest average validation accuracy 79.43%

Based on the highest validation accuracy :  
best number of Learning Rate :0.01  
best optimizer :SGD\_9

**Highlight the combination from the previous questions that achieves the highest average validation accuracy**



### Q3) Vary the MLP parameters [4/5]

Relu					
Max train acc ()%	Max test acc ()%	Min train acc ()%	Min test acc ()%	Avg train acc ()%	Avg test acc ()%
92.439860	80.926919	87.667048	71.479499	90.843834	78.217469

testing accuracy : 78.22% , activation function :relu  
testing accuracy : 78.93% , activation function :leaky\_relu  
testing accuracy : 22.64% , activation function :sigmoid  
testing accuracy : 76.43% , activation function :tanh

---

Leaky Relu					
Max train acc ()%	Max test acc ()%	Min train acc ()%	Min test acc ()%	Avg train acc ()%	Avg test acc ()%
93.279880	80.392158	85.567009	77.896613	90.080184	78.930482

=====  
the best activation function based on the testing accuracy : leaky\_relu  
=====

sigmoid					
Max train acc ()%	Max test acc ()%	Min train acc ()%	Min test acc ()%	Avg train acc ()%	Avg test acc ()%
25.353187	22.638147	24.360444	22.638147	24.994273	22.638147

**Highlight the combination that achieves the highest averages testing accuracy**

Tanh					
Max train acc ()%	Max test acc ()%	Min train acc ()%	Min test acc ()%	Avg train acc ()%	Avg test acc ()%
95.952654	77.896613	93.890798	75.401068	94.868270	76.434938

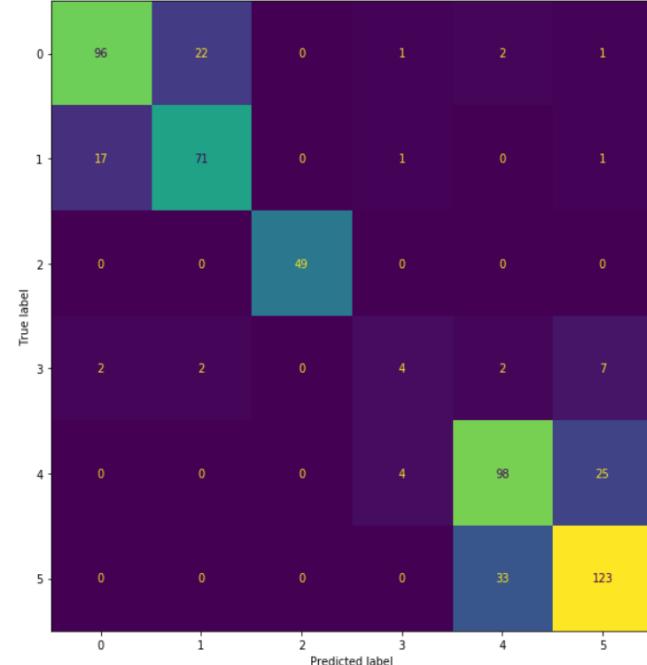


# Q3) Vary the MLP parameters [4/5]

Plot the confusion matrix for the combinations that achieves the highest average accuracy

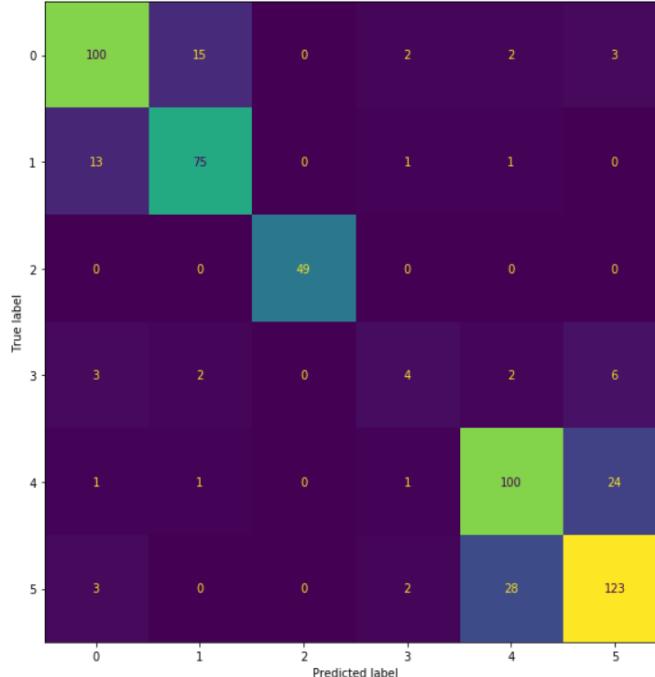
Confusion Matrix 1

confusion matrix number 1



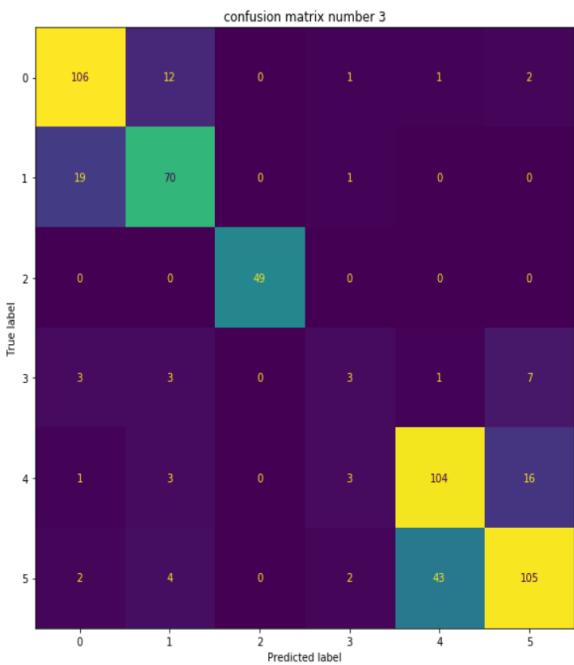
Confusion Matrix 2

confusion matrix number 2



Confusion Matrix 3

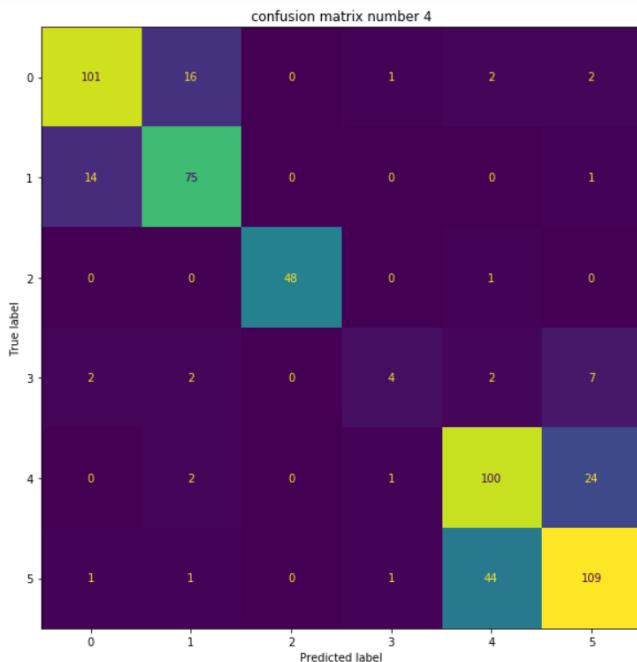
confusion matrix number 3



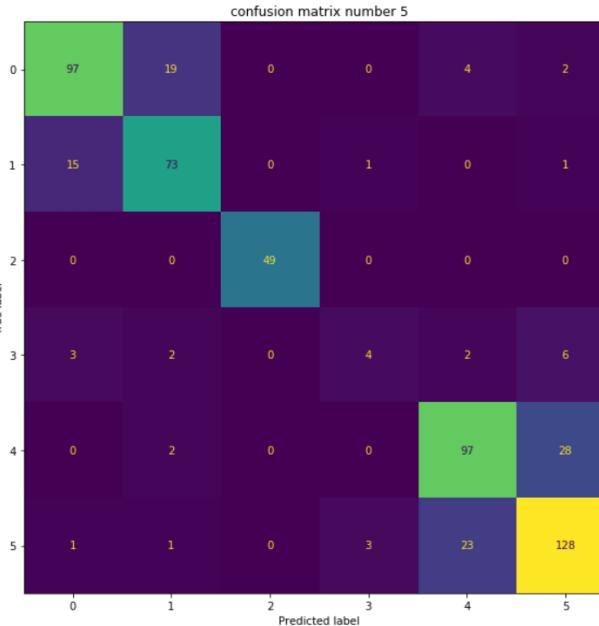
### Q3) Vary the MLP parameters [4/5]

Plot the confusion matrix for the combinations that achieves the highest average accuracy

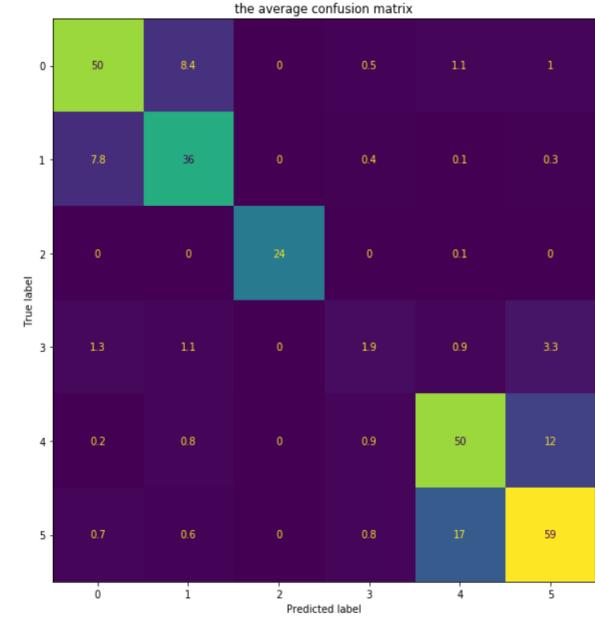
Confusion Matrix 4



Confusion Matrix 5



Average Confusion Matrix



### **Q3) Vary the MLP parameters [5/5] (Bonus)**

To enhance the average testing accuracy we have used two different approaches in other words two different experiments :

**Experiment 1 :**

try to enhance the average testing accuracy by examine the impact of using different number of epoch with the best combination from Q3.1 , Q3.2 , Q3.3 and Q3.4

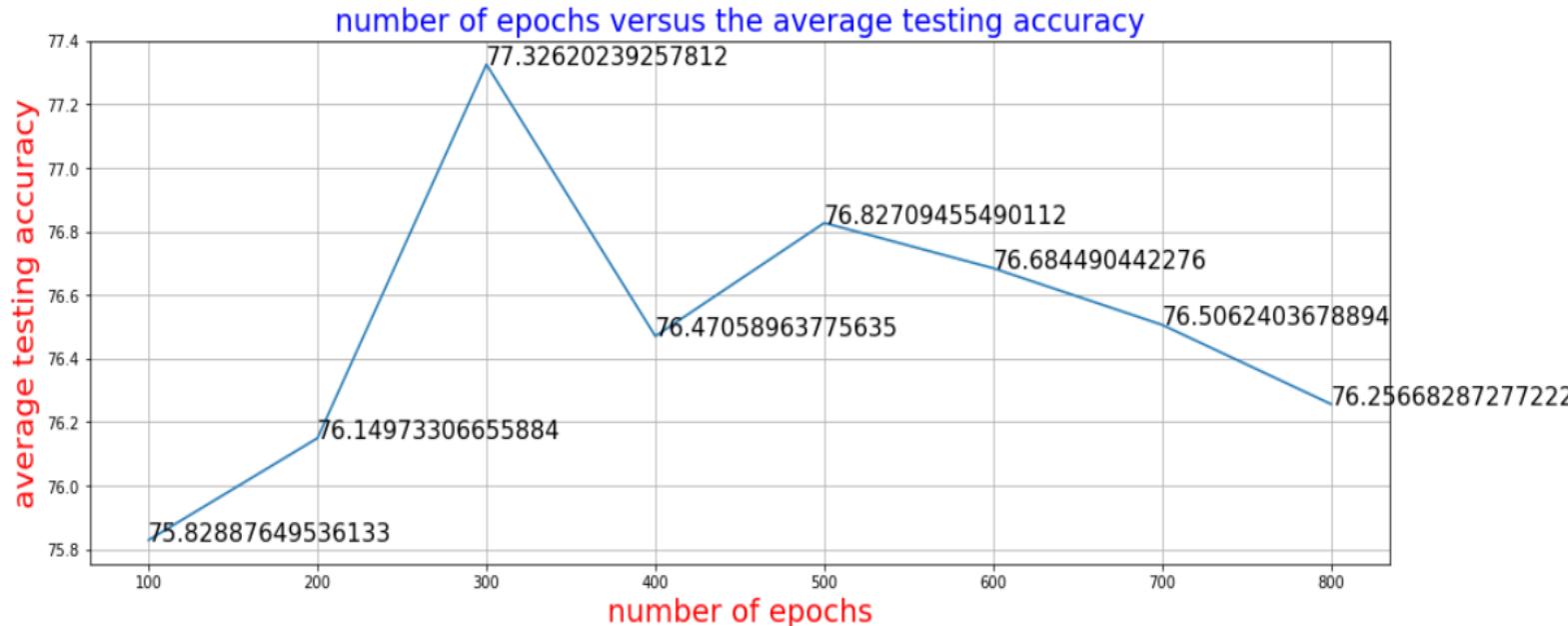
**Experiment 2 :**

try to enhance the average testing accuracy by using Grid search methods

### Q3) Vary the MLP parameters [5/5] (Bonus) : Experiment 1

```
=====
Experiment 1      : varying the Batch size
Batch size        : 64
Hidden Layers     : default
number of neurons : defeult
Learning Rate     : default
optimizer         : default
activation function : default
=====
```

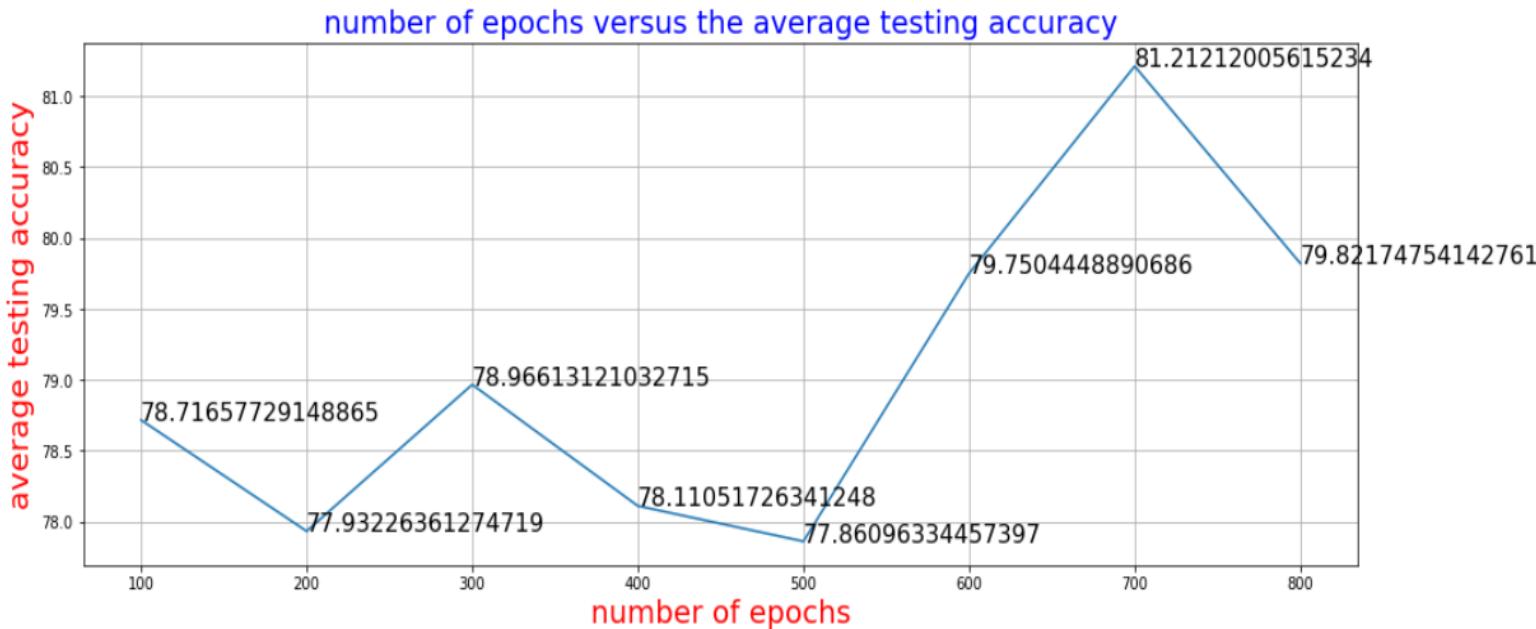
This experiment Based on the best combination from Q3.1



# Q3) Vary the MLP parameters [5/5] (Bonus) : Experiment 1

```
=====
Experiment 2      :varying the number of hidden layers and number of neurons
Batch size        : 64
Hidden Layers     : 2
number of neurons : 30
Learning Rate     : default
optimizer         : default
activation function: default
=====
```

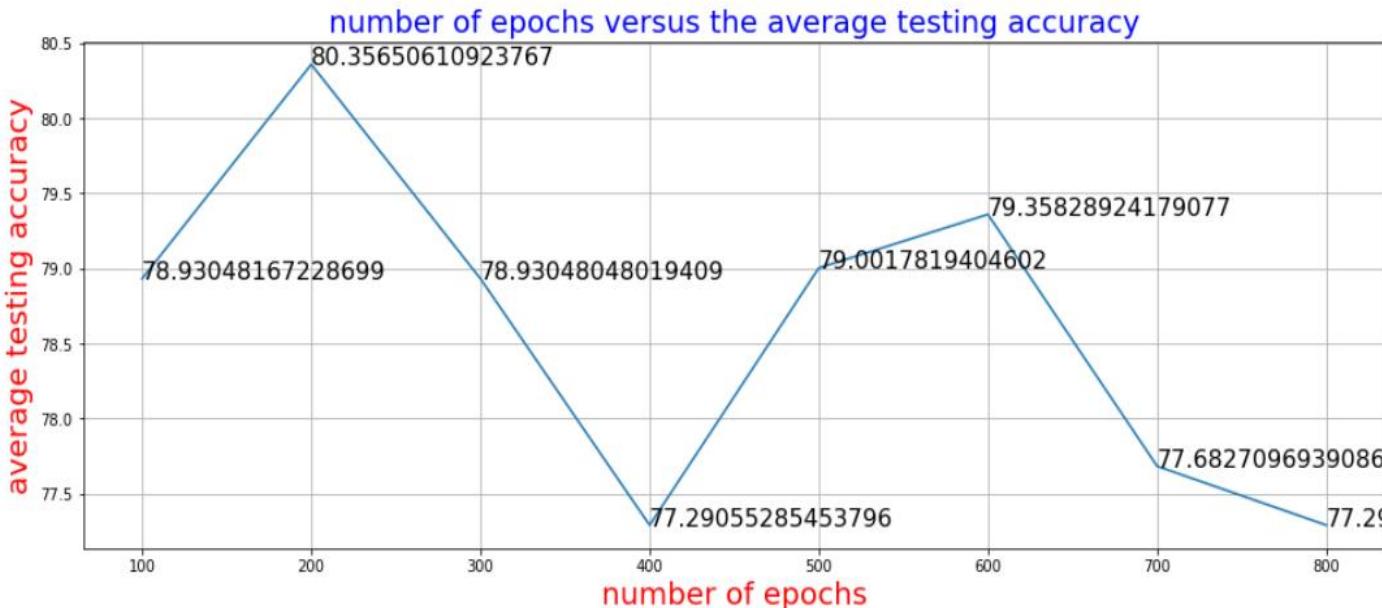
This experiment Based on the best combination from Q3.2



### Q3) Vary the MLP parameters [5/5] (Bonus) : Experiment 1

```
=====
Experiment 3      :varying the number of learning rates and optimizers
Batch size        : 64
Hidden Layers     : 2
number of neurons : 30
Learning Rate     : 0.01
optimizer         : Rprop
activation function: default
=====
```

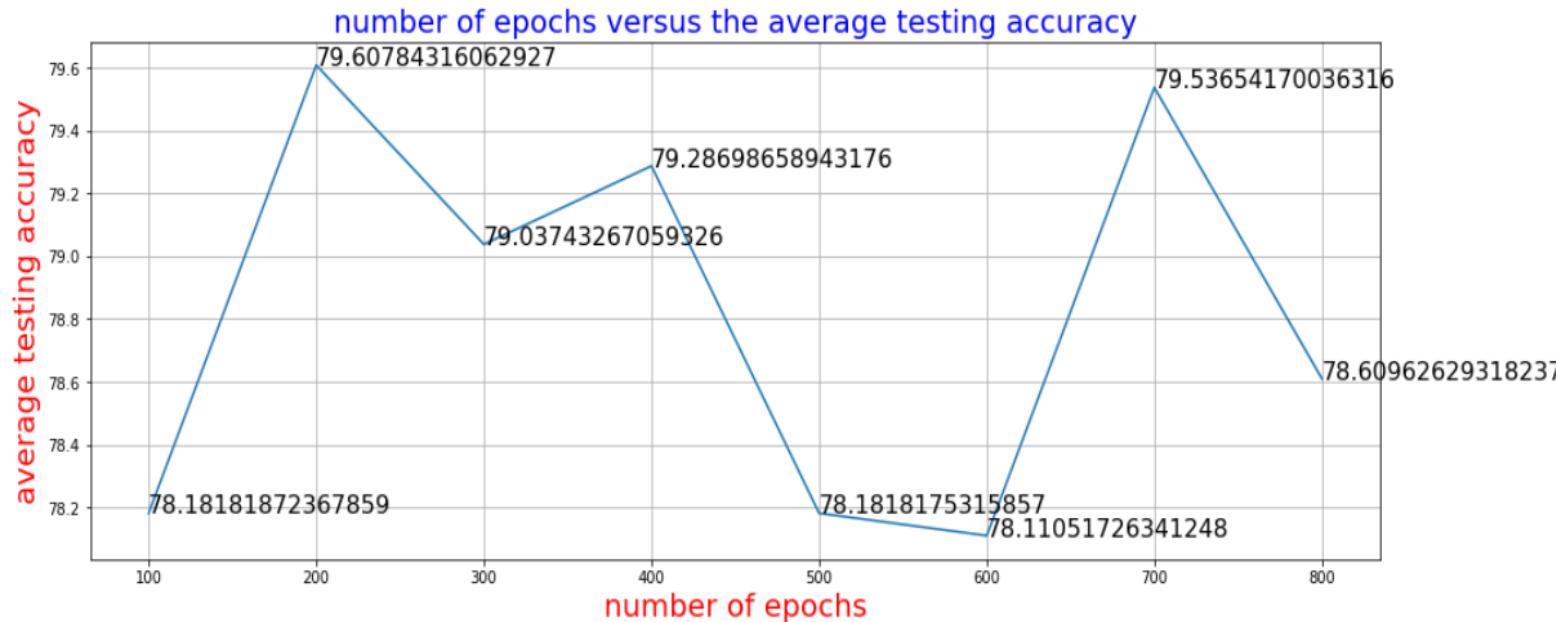
This experiment Based on the best combination from Q3.3



### Q3) Vary the MLP parameters [5/5] (Bonus) : Experiment 1

```
=====
Experiment 4      : vary the activation function
Batch size        : 64
Hidden Layers     : 2
number of neurons : 30
Learning Rate     : 0.01
optimizer         : Rprop
activation function: leaky_relu
=====
```

This experiment Based on the best combination from Q3.4



### Q3) Vary the MLP parameters [5/5] (Bonus) : Experiment 2

#### Grid Search

Batch Sizes = [128,250]  
Hidden Layers = [4,6]  
Number of neurons in each layer = [40,50]  
Learning rate = [0.01,0.001]  
Optimizers = ["adamW","SGD\_1","SGD\_5","SGD\_9","Rprop"]  
Activation functions = ["relu","leaky\_relu","sigmoid","tanh"]

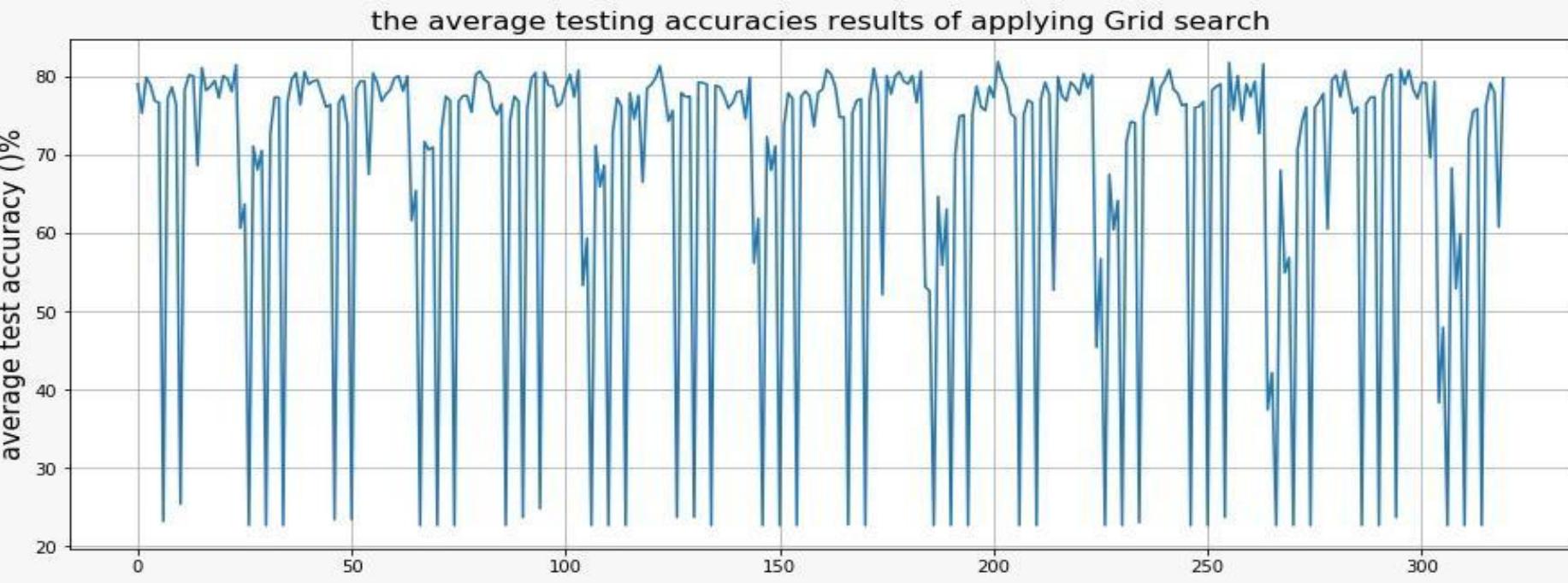
The maximum average accuracy after applying the Grid search is : **81.8538%**

Using :

**Batch size = 250, 4 hidden layers, 50 neurons in each layer , 0.01 learning rate , "adamW" optimizer and "leaky\_relu" activation function**

### Q3) Vary the MLP parameters [5/5] (Bonus) : Experiment 2

#### Analysis of the Grid search



The average accuracy decrease dramatically with some parameters ?!!!!!!

# Conclusion

- The Baseline performance was : **77.611 %**
- The highest average testing accuracy after varing the Batch size is : **77.61 %**
- The highest average testing accuracy after number of hidden layers and the number of neurons in each layer is : **80.86 %**
- The highest average testing accuracy after varing the learning rate and the optimizers is : **79.96%**
- The highest average testing accuracy after varing the activation function is : **78.93 %**
- The highest average testing accuracy after varing the number of epochs is : **81.2121 %**
- The highest average testing accuracy after using Grid search technique is : **81.8538 %**

So as a conclusion the batch size , the number of hidden layers , the number of neurons in each layer , the learning rate , the optimizer , the activation function , and the number of epoch to be run all of that affects on the testing accuracy .

Note that some of the combination had a bad effect on the accuracy as shown in the previous slide.

# References

[1] <https://www.tapadhirdas.com/unr-idd-dataset>