Network Intrusion Detection System with Machine Learning Approach

Edwin Valdez
Data 606 Capstone Project

NETWORK INTRUSION DETECTION SYSTEM

- This is a security system that monitors the network activity to detect malicious activity that might be harmful to any company.
- The purpose of this project servers two main objectives, one is to be able to detect malicious activities vs normal activity, and Second is to be able to detect the abnormal activity between nine categories.
- The benefits of having a NIDS are numerous since, in many cases, the system alerts the security team about the possible malicious attack and uses that information to plan a risk management plan.

DATASET USED FOR THIS PROJECT

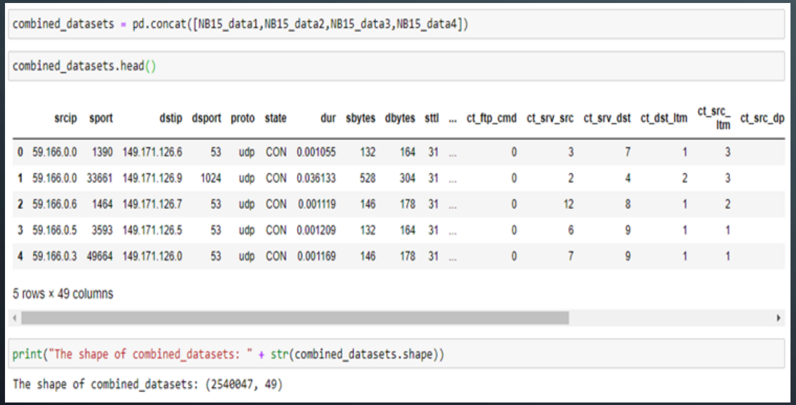
- I going to use for this project the dataset for UNSW-NB15 Dataset which comes from the IXIA PerfectStorm tool in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS).

SCIENTIA

MANU ET MENTE

UNSW

SYDNEY



```
combined_datasets = pd.concat([NB15_data1,NB15_data2,NB15_data3,NB15_data4])
```

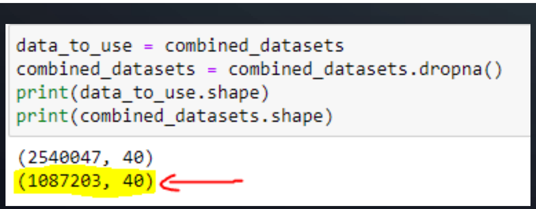
```
combined_datasets.head()
```

	srcip	sport	dstip	dsport	proto	state	dur	sbytes	dbytes	sttl	...	ct_ftp_cmd	ct_srv_src	ct_srv_dst	ct_dst_ltm	ct_src_ ltm	ct_src_dp
0	59.166.0.0	1390	149.171.126.6	53	udp	CON	0.001055	132	164	31	...	0	3	7	1	3	
1	59.166.0.0	33661	149.171.126.9	1024	udp	CON	0.036133	528	304	31	...	0	2	4	2	3	
2	59.166.0.6	1464	149.171.126.7	53	udp	CON	0.001119	146	178	31	...	0	12	8	1	2	
3	59.166.0.5	3593	149.171.126.5	53	udp	CON	0.001209	132	164	31	...	0	6	9	1	1	
4	59.166.0.3	49664	149.171.126.0	53	udp	CON	0.001169	146	178	31	...	0	7	9	1	1	

5 rows × 49 columns

```
print("The shape of combined_datasets: " + str(combined_datasets.shape))
```

```
The shape of combined_datasets: (2540047, 49)
```

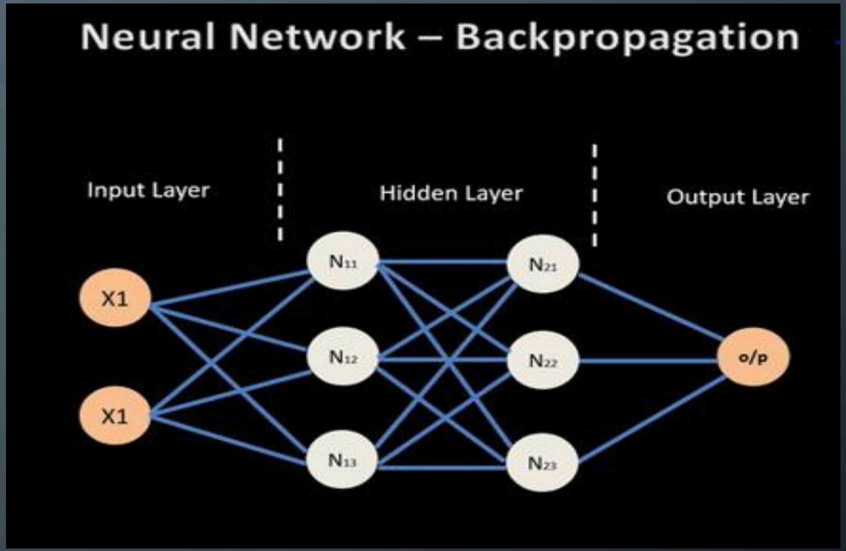


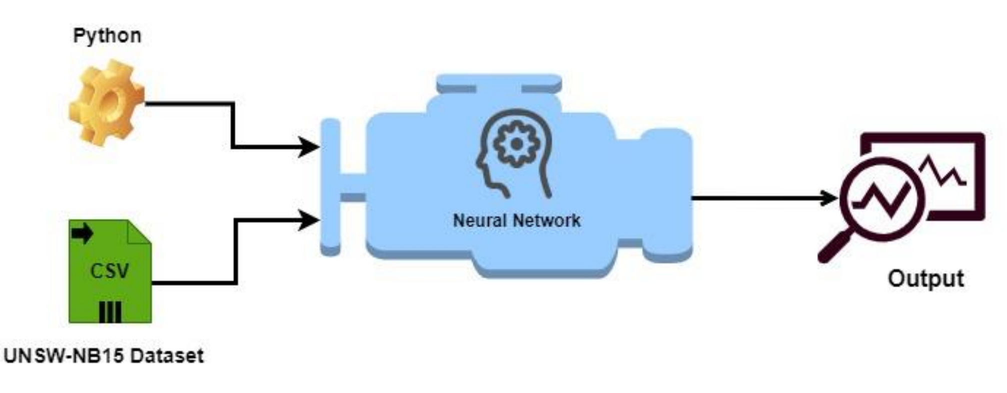
```
data_to_use = combined_datasets
combined_datasets = combined_datasets.dropna()
print(data_to_use.shape)
print(combined_datasets.shape)
```

```
(2540047, 40)
(1087203, 40)
```

METHODOLOGY TO USE IN THIS PROJECT

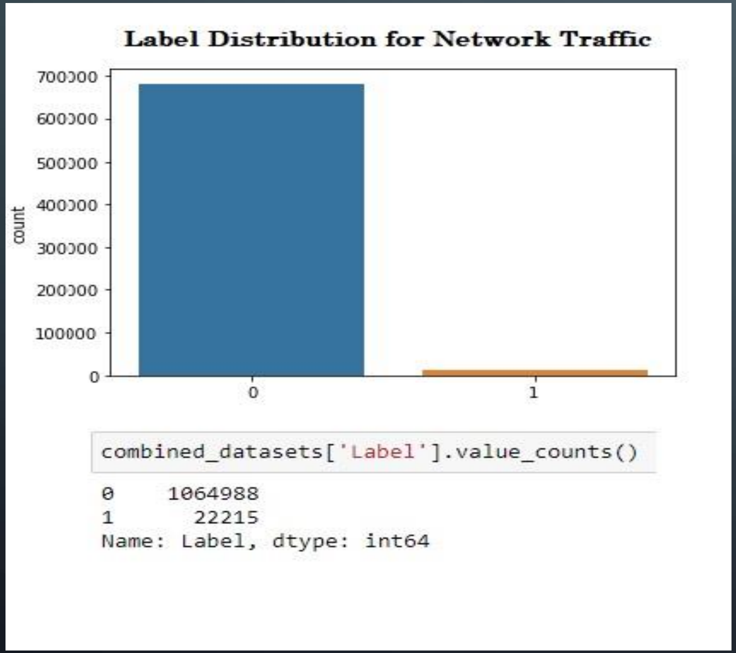
- One of the major reasons why I consider Neural Network is an excellent choice was the relationship between the features that suggest to me it would be an excellent choice. We can see the relationship between features thanks to the correlation matrix.
- The main objective of this project is to transform a traditional NIDS into a new system with the help of machine learning, in particular neural network algorithm.
- I used Keras and Keras classifier to train my model, and I tried different types of configuration such as bias value, add class weights, shuffle data, and playing with the sample size.





RESULTS FINDINGS – FIRST GOAL

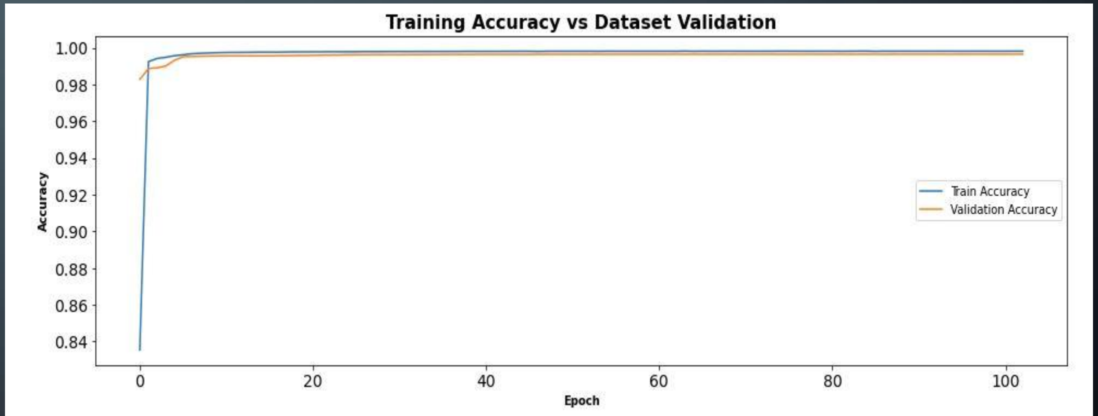
THE DISTRIBUTION BETWEEN NORMAL AND ABNORMAL TRAFFIC



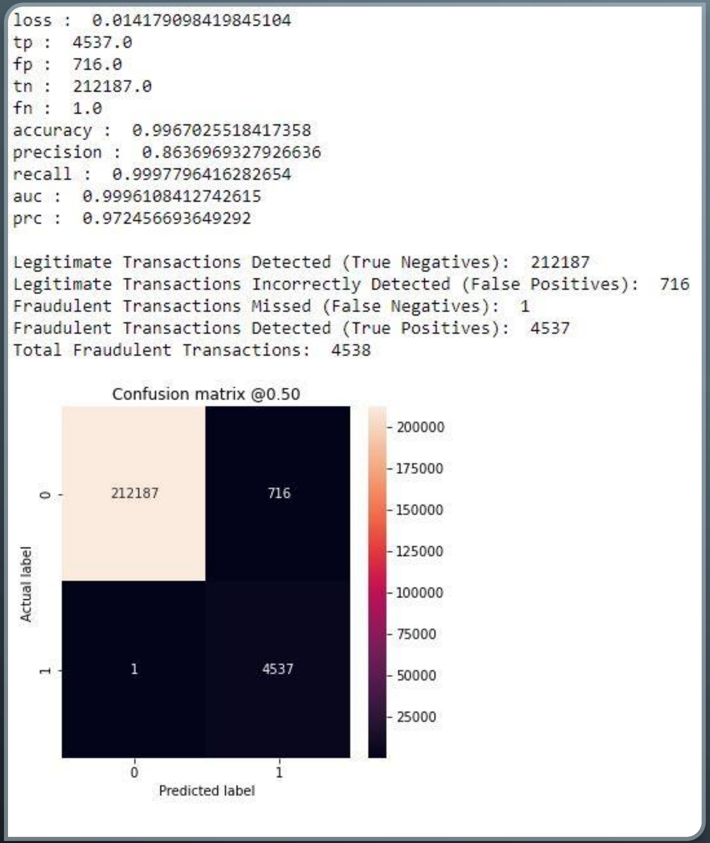
```
combined_datasets['Label'].value_counts()
```

```
0    1064988
1      22215
Name: Label, dtype: int64
```

COMPARISON BETWEEN TRAINING AND VALIDATION ACCURACY



RESULTS FINDINGS – FIRST GOAL



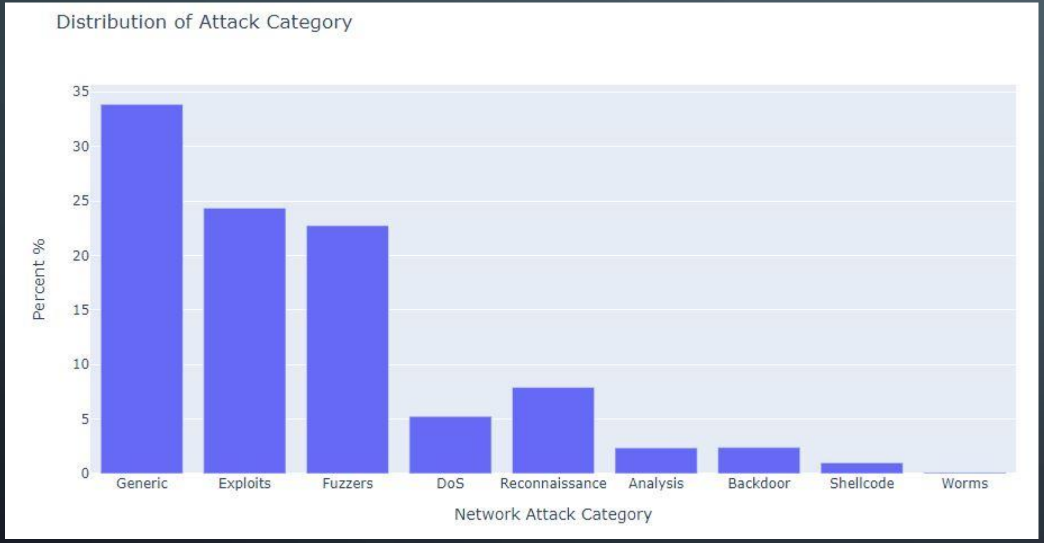
```
loss :  0.014179098419845104
tp :  4537.0
fp :  716.0
tn :  212187.0
fn :  1.0
accuracy :  0.9967025518417358
precision :  0.8636969327926636
recall :  0.9997796416282654
auc :  0.9996108412742615
prc :  0.972456693649292

Legitimate Transactions Detected (True Negatives):  212187
Legitimate Transactions Incorrectly Detected (False Positives):  716
Fraudulent Transactions Missed (False Negatives):  1
Fraudulent Transactions Detected (True Positives):  4537
Total Fraudulent Transactions:  4538
```

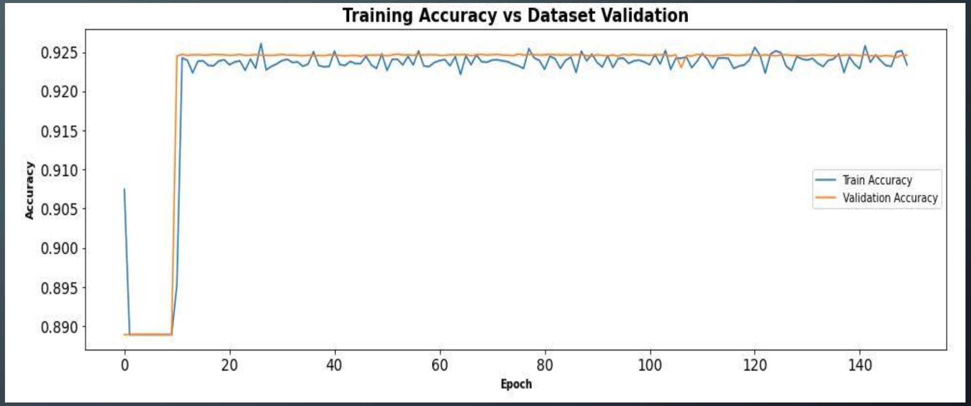
- Confusion matrix and other metrics’ results of our final model
- My model would be able to detect normal traffic, which good because it will not block any legitimate traffic on the network
- On the other hand, some malicious traffic would be able to pass any filter place on the network, which will not alert the security team about its behavior

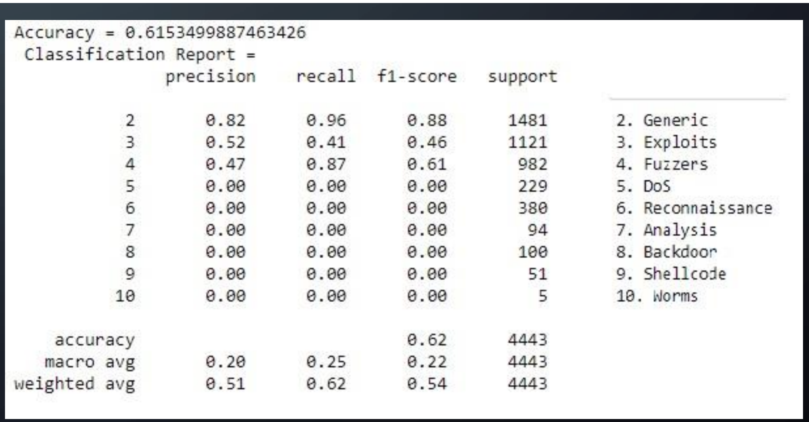
RESULTS FINDINGS- SECOND GOAL

DISTRIBUTION OF THE NETWORK ATTACK CATEGORY



TRAINING VS VALIDATION ACCURACY'S RESULTS AND OTHER METRICS





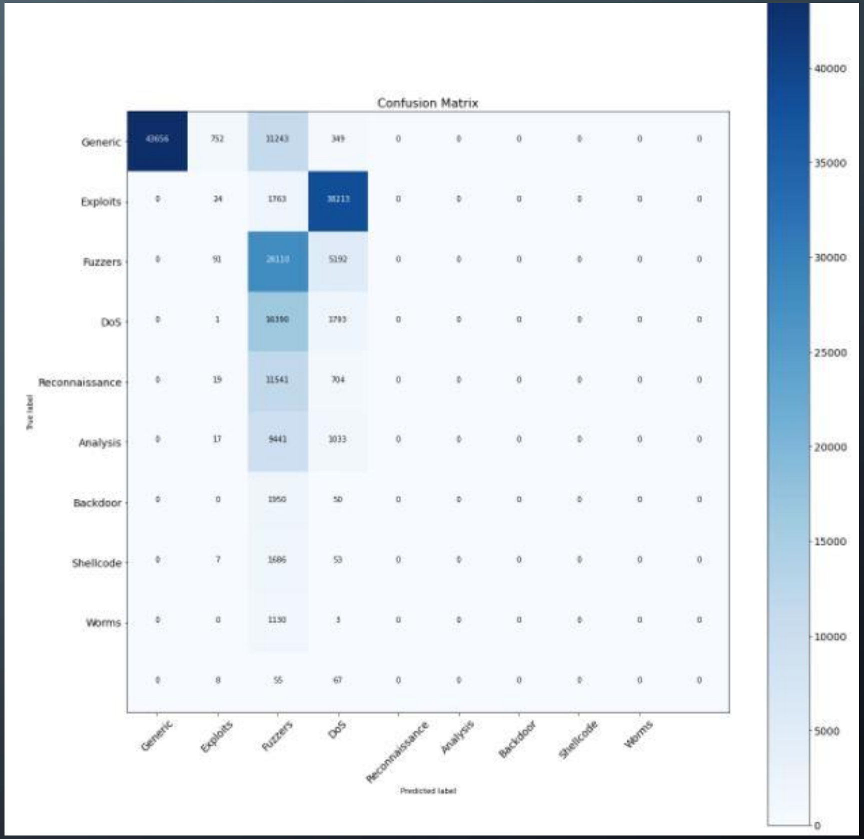
```
Accuracy = 0.6153499887463426
 Classification Report =
               precision    recall  f1-score   support

           2       0.82      0.96      0.88      1481     2. Generic
           3       0.52      0.41      0.46      1121     3. Exploits
           4       0.47      0.87      0.61       982     4. Fuzzers
           5       0.00      0.00      0.00       229     5. DoS
           6       0.00      0.00      0.00       380     6. Reconnaissance
           7       0.00      0.00      0.00        94     7. Analysis
           8       0.00      0.00      0.00       100     8. Backdoor
           9       0.00      0.00      0.00        51     9. Shellcode
          10       0.00      0.00      0.00         5     10. Worms

    accuracy                           0.62      4443
   macro avg       0.20      0.25      0.22      4443
weighted avg       0.51      0.62      0.54      4443
```

RESULTS FINDINGS- SECOND GOAL

- Confusion Matrix of the network attack categories.
- Some network attack Categories were difficult to identify from other categories like Generic.
- Needed more examples for the most specific network attack categories like worms and shellcode.

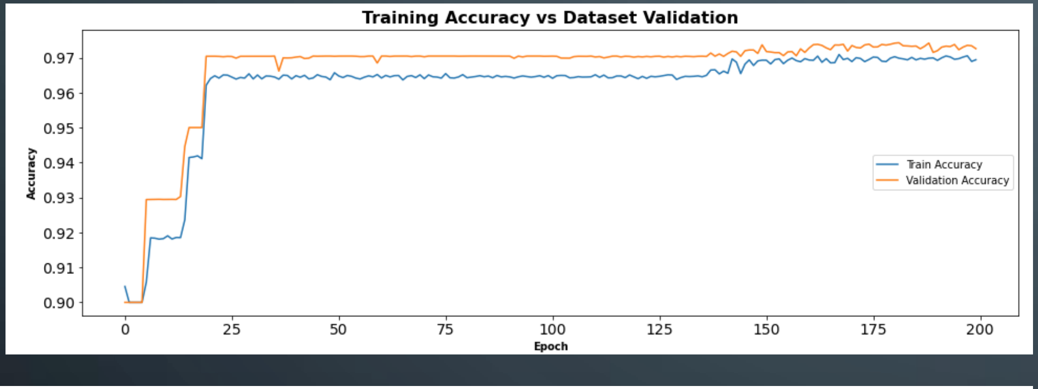


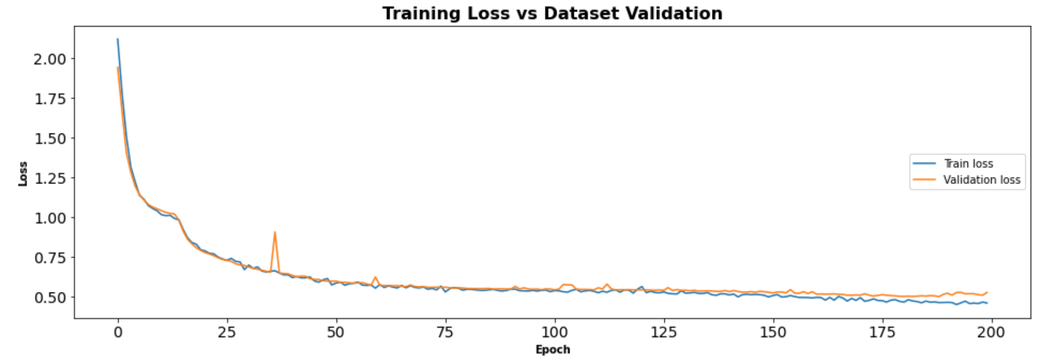
Confusion Matrix

	Generic	Exploits	Fuzzers	DoS	Reconnaissance	Analysis	Backdoor	Shellcode	Worms	
Generic	43656	752	11243	349	0	0	0	0	0	0
Exploits	0	24	1763	38213	0	0	0	0	0	0
Fuzzers	0	91	26110	5192	0	0	0	0	0	0
DoS	0	1	16390	1793	0	0	0	0	0	0
Reconnaissance	0	19	11541	704	0	0	0	0	0	0
Analysis	0	17	9441	1033	0	0	0	0	0	0
Backdoor	0	0	1950	50	0	0	0	0	0	0
Shellcode	0	7	1686	53	0	0	0	0	0	0
Worms	0	0	1130	3	0	0	0	0	0	0
	0	8	55	67	0	0	0	0	0	0

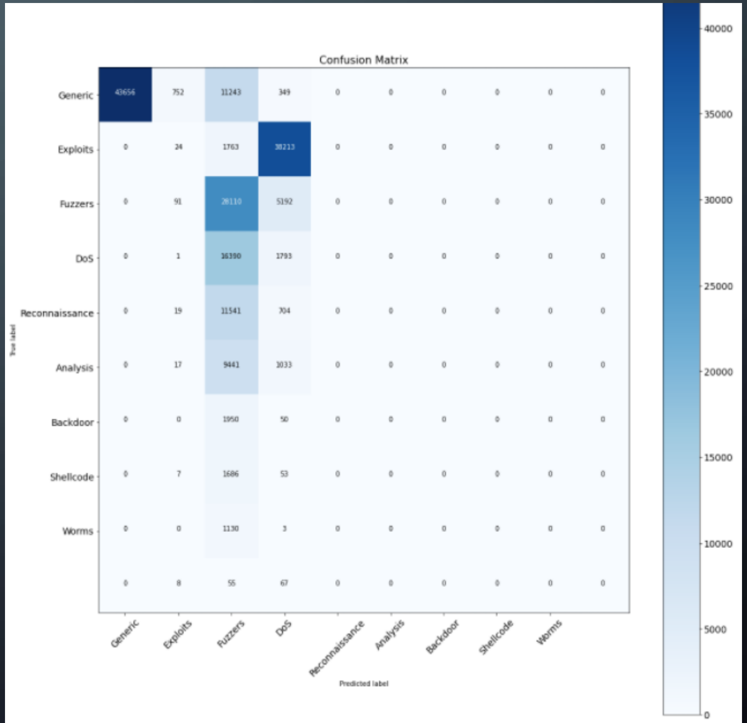
RESULTS FINDINGS- OVERALL RESULTS

COMBINE DATASET, TRAINING AND VALIDATION ACCURACY





CONFUSION MATRIX RESULTS WITH THE MODEL CREATED FOR BOTH GOALS



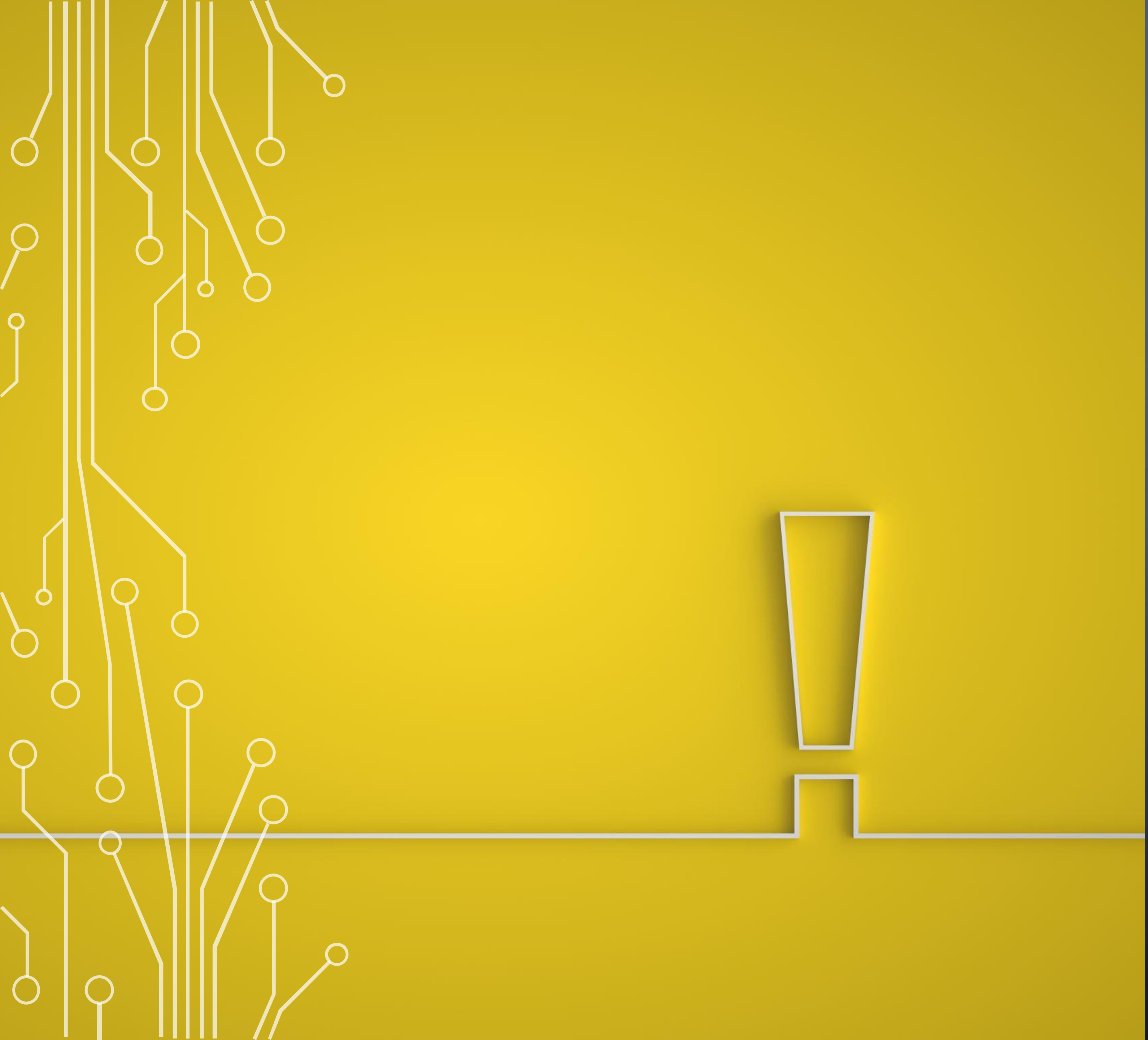
	Generic	Exploits	Fuzzers	DoS	Reconnaissance	Analysis	Backdoor	Shellcode	Worms	
Generic	43656	752	11243	349	0	0	0	0	0	0
Exploits	0	24	1763	38213	0	0	0	0	0	0
Fuzzers	0	91	28110	5192	0	0	0	0	0	0
DoS	0	1	16390	1793	0	0	0	0	0	0
Reconnaissance	0	19	11541	704	0	0	0	0	0	0
Analysis	0	17	9441	1033	0	0	0	0	0	0
Backdoor	0	0	1950	50	0	0	0	0	0	0
Shellcode	0	7	1686	53	0	0	0	0	0	0
Worms	0	0	1130	3	0	0	0	0	0	0
	0	8	55	67	0	0	0	0	0	0

LIMITATIONS AND FURTHER WORK

- Find a dataset that contains balance records between the different types of network attack categories. Collect network traffic from different sources and create one effective network dataset.
- To compare my model against other machine learning algorithms like random forests.
- Be able to design this model to perform live network traffic analysis by using direct feeding for tools like Wireshark, Cisco, Palo Alto, etc.
- Time to develop this project.

REFERENCE

- Abdelhameed M, Dr. Nour M. (Nov. 14, 2018) The UNSW-NB15 Dataset Description [Data set] . The University of New South Wales. https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/
- Cuelogic Technologies. (May 13, 2019). Evaluation of Machine Learning Algorithms for Intrusion Detection System. Medium. https://medium.com/cuelogic-technologies/evaluation-of-machine-learning-algorithms-for-intrusion-detection-system-6854645f9211
- Almseidin, M., Alzubi, M., Kovacs, S., & Alkasassbeh, M. (2020). Evaluation of Machine Learning Algorithms for Intrusion Detection System. Mutah University, Amman,Jordan. https://arxiv.org/ftp/arxiv/papers/1801/1801.02330.pdf
- Ahmad, Z., Khan, A., Shiang, C., Abdullah, J & Ahmad, F. (Oct. 16, 2020) Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1002/ett.4150
- MLK. (Oct. 30, 2019) Animated Explanation of Feed Forward Neural Network Architecture. MLK making AI simple. https://machinelearningknowledge.ai/animated-explanation-of-feed-forward-neural-network-architecture/
- Sharma, Bikash. (Sept. 4, 2019) Evaluating a Machine Learning Model. Skyl.ai. https://blog.skyl.ai/evaluating-a-machine-learning-model/
- Stone, Mark. (2021, April 9). Intrusion Detection Systems (IDS) explained. AT&T. https://cybersecurity.att.com/solutions/intrusion-detection-system/ids-explained
- Brownlee, Jason. (2016, Sept. 21). How To Improve Deep Learning Performance. Machine Learning Mastery. https://machinelearningmastery.com/improve-deep-learning-performance/
- Wang, Chi-Feng. (2018, Aug.16). Different Ways of Improving Training Accuracy. Towards Data Science. https://towardsdatascience.com/different-ways-of-improving-training-accuracy-c526db15a5b2
- Sharma, Pulkit. (2019, Nov. 7). 4 Proven Tricks to Improve your Deep Learning Model's Performance. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/11/4-tricks-improve-deep-learning-model-performance/



THANKS FOR YOUR ATTENTION

UMBC DATA 606