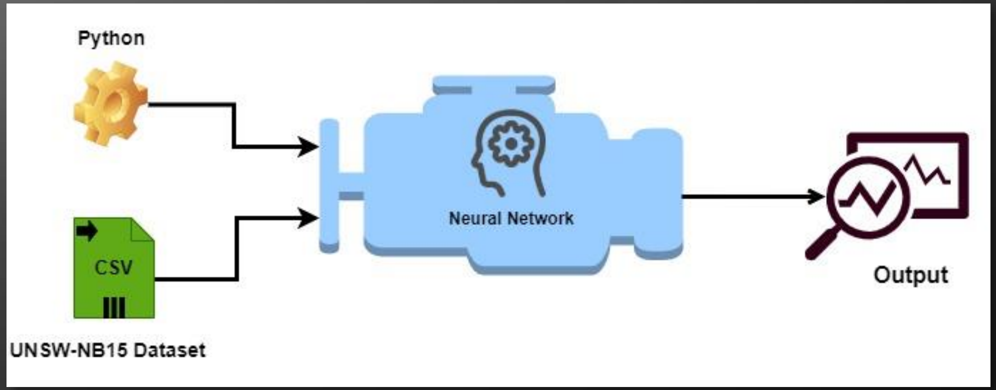
Network Intrusion Detection System with Machine Learning Approach

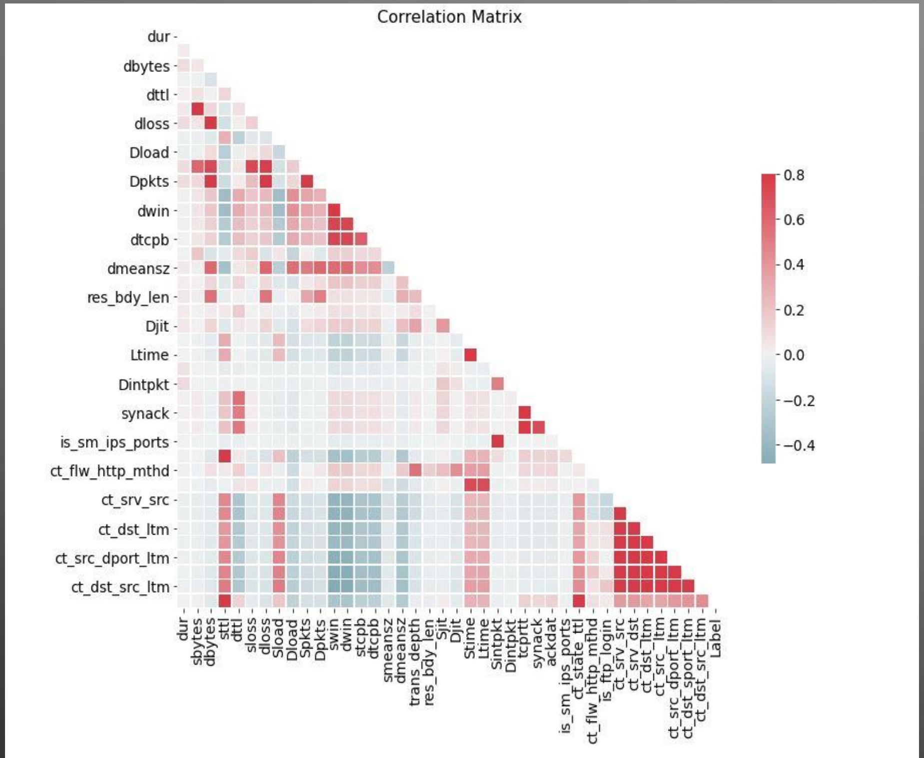
Edwin Valdez
Data 606

RECAP FROM THE PREVIOUS PHASE

- Define what was network Intrusion Detection System
- Use of Neural Network for our model
- Using the Dataset from UNSW-NB15 from the Cyber Range Lab of the Australian Center for Cyber Security (ACCS).
- The total of 49 features in the Datasets
- Nine different network attack classification



PHASE II: DATA EXPLORATION, CLEANING, TRANSFORMATION AND MODEL CONSTRUCTION

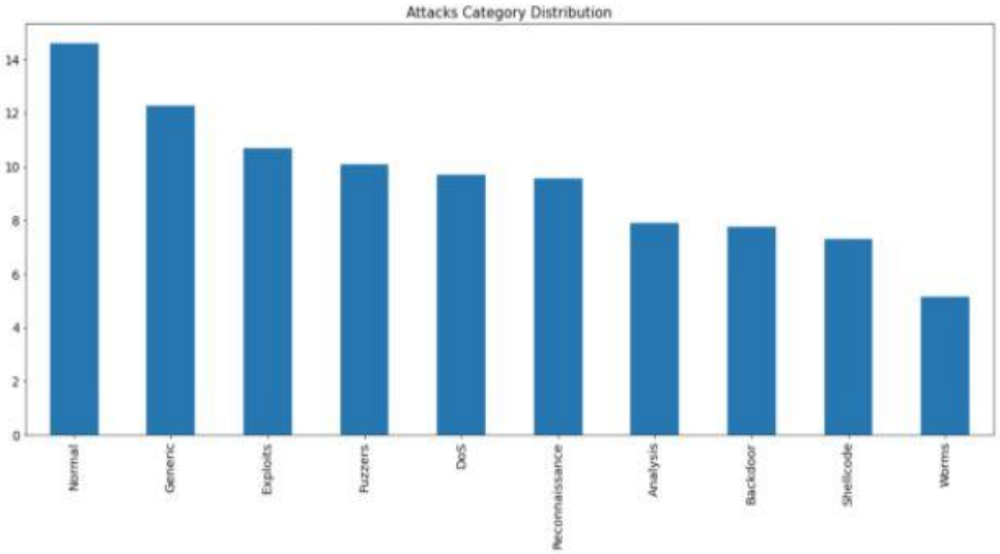


- We are giving a slightly check of what kind of data we are dealing with.
- First method implemented was the correlation Matrix.
- The matrix shows that we do not have a lot of correlation between features.

DATA EXPLORATION ANALYSIS

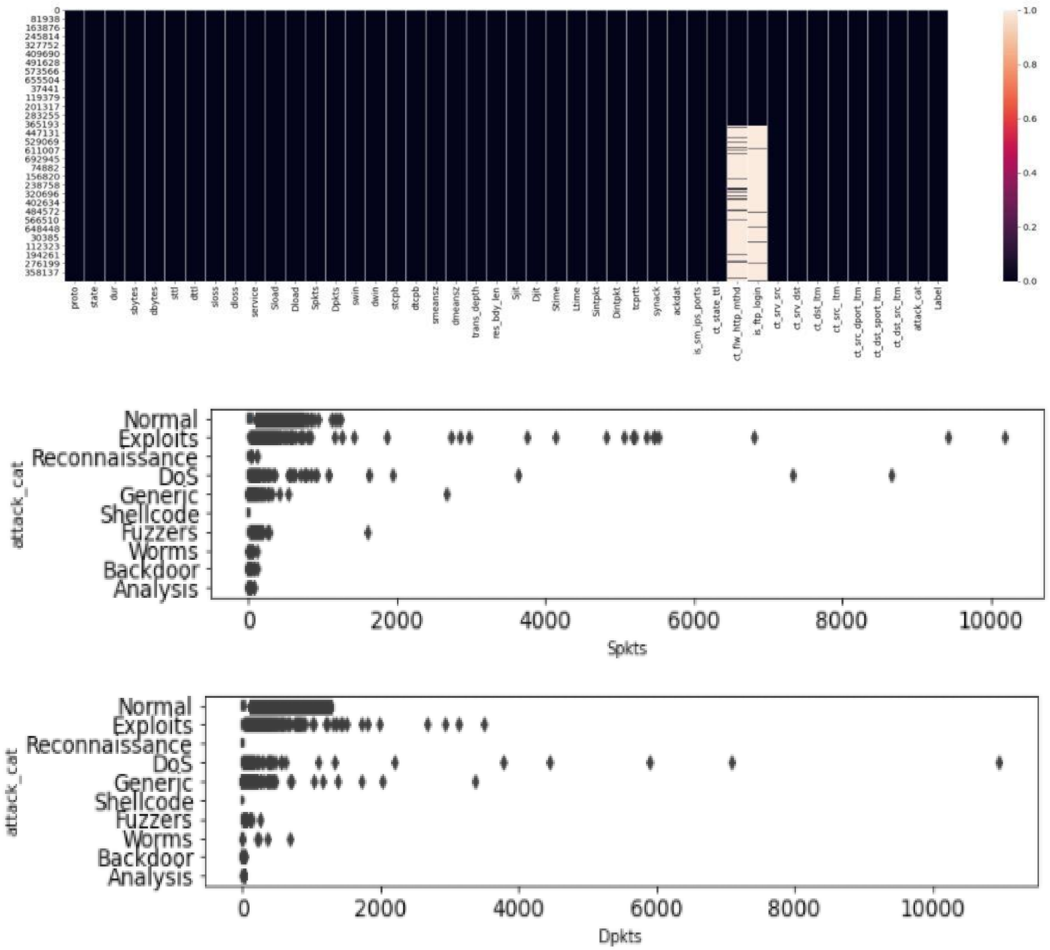
- The dataset contains the total of 49 unique features.
- The dataset can be break down into nine malicious network attack plus the addition of the “normal” network traffic.
- The distribution of the data type can be classified into float64, int64, nine category are object types.

```
Data columns (total 49 columns):
 #   Column             Non-Null Count    Dtype
---  ------             --------------    -----
 0   srcip              2540047 non-null  object
 1   sport              2540047 non-null  object
 2   dstip              2540047 non-null  object
 3   dsport             2540047 non-null  object
 4   proto              2540047 non-null  object
 5   state              2540047 non-null  object
 6   dur                2540047 non-null  float64
 7   sbytes             2540047 non-null  int64
 8   dbytes             2540047 non-null  int64
 9   sttl               2540047 non-null  int64
 10  dttl               2540047 non-null  int64
 11  sloss              2540047 non-null  int64
 12  dloss              2540047 non-null  int64
 13  service            2540047 non-null  object
 14  Sload              2540047 non-null  float64
 15  Dload              2540047 non-null  float64
 16  Spkts              2540047 non-null  int64
 17  Dpkts              2540047 non-null  int64
 18  swin               2540047 non-null  int64
 19  dwin               2540047 non-null  int64
 20  stcpb              2540047 non-null  int64
 21  dtcpb              2540047 non-null  int64
 22  smeansz            2540047 non-null  int64
 23  dmeansz            2540047 non-null  int64
 24  trans_depth        2540047 non-null  int64
 25  res_bdy_len        2540047 non-null  int64
 26  Sjit               2540047 non-null  float64
 27  Djit               2540047 non-null  float64
 28  Stime              2540047 non-null  int64
 29  Ltime              2540047 non-null  int64
 30  Sintpkt            2540047 non-null  float64
 31  Dintpkt            2540047 non-null  float64
 32  tcprtt             2540047 non-null  float64
 33  synack             2540047 non-null  float64
 34  ackdat             2540047 non-null  float64
 35  is_sm_ips_ports    2540047 non-null  int64
 36  ct_state_ttl       2540047 non-null  int64
 37  ct_flw_http_mthd   1191902 non-null  float64
 38  is_ftp_login       1110168 non-null  float64
 39  ct_ftp_cmd         2540047 non-null  object
 40  ct_srv_src         2540047 non-null  int64
 41  ct_srv_dst         2540047 non-null  int64
 42  ct_dst_ltm         2540047 non-null  int64
 43  ct_src_ ltm        2540047 non-null  int64
 44  ct_src_dport_ltm   2540047 non-null  int64
 45  ct_dst_sport_ltm   2540047 non-null  int64
 46  ct_dst_src_ltm     2540047 non-null  int64
 47  attack_cat         321283 non-null   object
 48  Label              2540047 non-null  int64
```




```
Normal            851960
Generic             6036
Exploits            4338
Fuzzers             4027
Reconnaissance      1440
DoS                  936
Analysis             422
Backdoor             404
Shellcode            182
Worms                 17
Name: attack_cat, dtype: int64
Total: 869762
Total percentange of attack label from the total of record is: 0.97953233183330
61
```

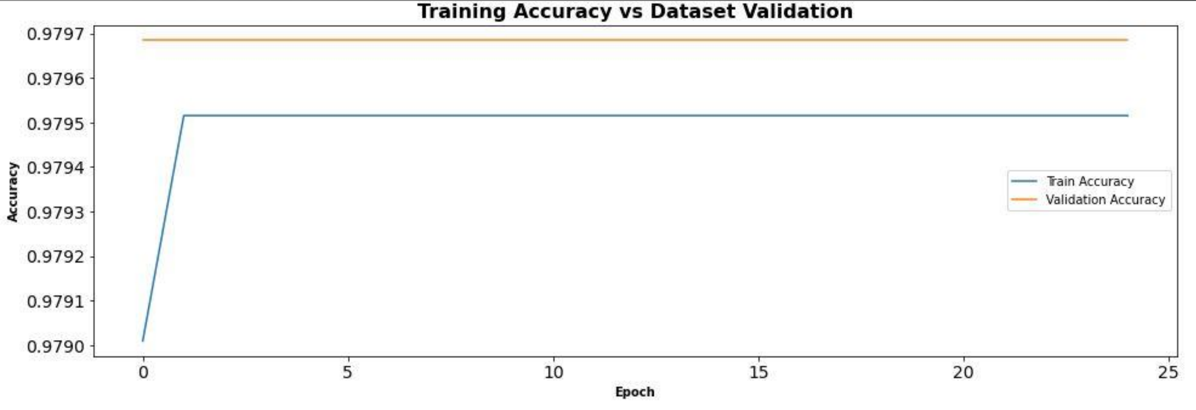
DATA CLEANING AND TRANSFORMATION



- By creating a graph that would show us now many nulls values we have, it was necessary to make a transformation on the data and some cleaning.
- Some null values under the "attack_cat" columns belongs to the category "Normal"
- There are a lot of noise in the data that we will need to clean and transform the data.

MODEL CONSTRUCTION USING KERAS CLASSIFIER

- The initial setup for our model would be with 25 epochs and a batch size of 15.
- Our neural network model would contain 6 hidden layer.
- As expected for our initial model, our lose value is low and accuracy score is high due to overfitting.



WHAT IS COMING NEXT …

- MORE DATA EXPLORATION, CLEANING AND TRANSFORMATION IS NEEDED.
- INCREASE THE NUMBER OF EPOCHS AND MODIFY THE BATCH SIZE.
- PLAY MORE WITH THE NUMBER OF HIDDEN LAYER.
- APPLYING DIFFERENT CONFIGURATION TO INCREASE BETTER PREDICTION.

REFERENCE

- 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