DATA 240

Group2

Dataset was taken from the site of University of New brunswick for CIC-IDS2017 [<https://www.unb.ca/cic/datasets/ids-2017.html>]

After registering you can access the data at [<http://205.174.165.80/CICDataset/CIC-IDS-2017/>]. Data is also loaded into SJSU account accessible google drive with link [<https://drive.google.com/drive/folders/17KXI614WYMVTAPQrhf1hk--MfGiWUjxQ?usp=sharing>]

To run the code.

* Download the data and place the data files under Data/MachineLearningCVE
* Create a conda environment using `conda create -n env\_name python`
* Open the conda environment using `conda activate env\_name`
* Install condarequirements.txt by `conda install --file condarequirements.txt`
* Install piprequirements.txt by `pip install -r piprequirements.txt`
* Go into code directory and Run the main.py using `python main.py`

**Team Contribution**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Task** | **Dhrumil** | **Karthik** | **Shashank** | **Shresta** |
| Background research | Yes | Yes | Yes | Yes |
| Literature review | Yes | Yes | Yes | Yes |
| Data Collection | Yes | Yes | Yes | Yes |
| Data Processing | Yes | Yes | Yes | Yes |
| Modeling | Logistic Regression, Random Forest, Naive Bayes. | SGD, Naive Bayes, Logistic Regression | Naive Bayes,Logistic Regression, Random Forest | Adaboost, Logistic Regression, Random Forest, |
| Documentation | Yes | Yes | Yes | Yes |