Classification Course
Assignment 2
NB and Logistic Regression Project

Follow the below instructions and perform the tasks on the given dataset: Network Attack – NB15

The raw network packets of the UNSW-NB 15 dataset was created by the IXIA PerfectStorm 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. Tcpdump tool is used to capture the raw traffic (e.g., Pcap files). This dataset has nine types of attacks, namely, Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms. The Argus, Bro-IDS tools are used and twelve algorithms are developed to generate totally 49 features with the class label. These features are described in the UNSW-NB15_features.csv file.

Build a classification model which will classify the 9 different attacks.

Tasks:

1. EDA and basic data pre-processing and preparation [02]
    a. Null/ outlier treatment
    b. Remove non-important features
    c. Split training and testing data set
    d. Standardize/normalize the variables whenever required.
2. Train model using suitable variation of NB and logistic regression (you may use sklearn) [02]
3. Compare accuracy of different approaches [02]
4. Apply feature reduction techniques [01]
5. Train model on the reduced feature subset using NB and Logistic regression [01]
6. Comment on performance of these approaches (NB and Logistic Regression) on reduced dataset (after applying PCA/feature reduction) [01]