CS725: Foundations of
Machine Learning

Intrusion Detection Through Machine Learning Algorithms in Cyber Security

Rohit Parida: 24M2136
Vinjamuri Aravind: 24M0758
Ramya Shri Shakthi K R: 24D0369
Gurujeet Singh Raghuwanshi: 24M0759

Problem Statement

With the increasing sophistication of cyber threats, manual intrusion response is no longer sufficient to ensure timely and effective mitigation.

Developing an automated intrusion detection system using on UNSW-NB15 data set to enhance cyber security.

Dataset

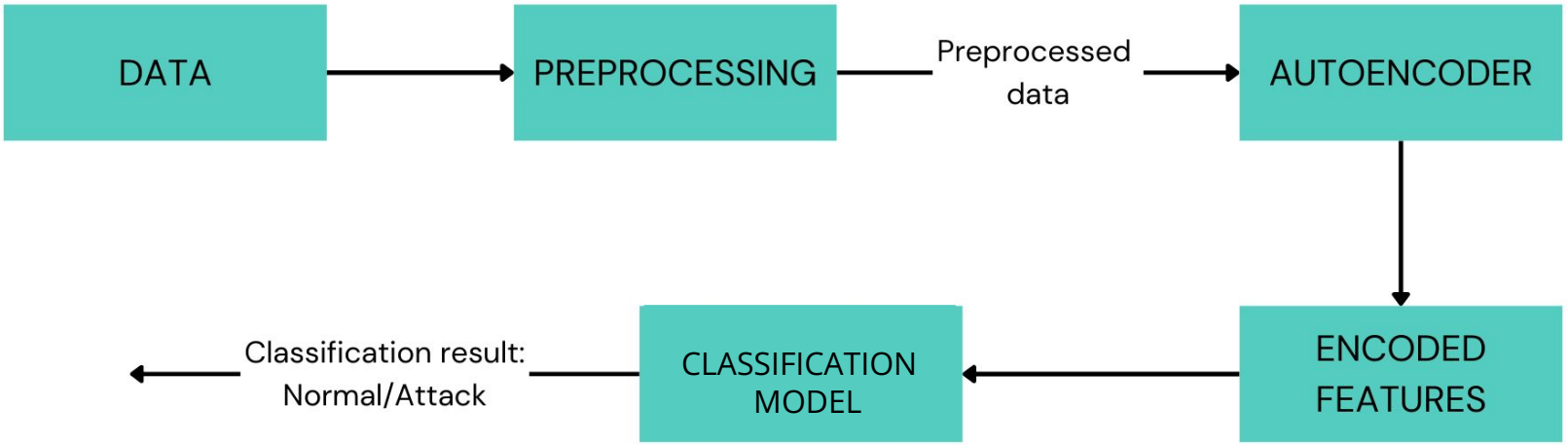
- **Data Source:** Generated by IXIA PerfectStorm at UNSW Canberra
- **Traffic captured:** 100 GB raw traffic (Pcap files) via Tcpdump
- **Number of attacks captured:** 9 (e.g., DoS, Worms, Exploits)
- **Feature set:** 45 features extracted using Argus and Bro-IDS
- **Training Data:** 175,341 records
- **Testing Data:** 82,332 records

Data Preprocessing

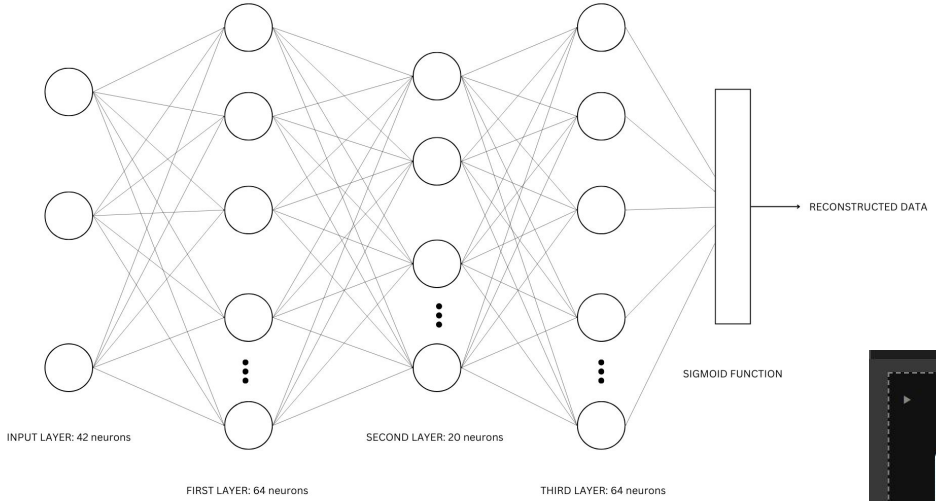
- **Drop irrelevant columns in data**
- **Splitting data:** Data is split into training set (95%) and validation set (5%)
- **Label Encoding:** Encoding categorical columns into numeric values
- **Feature Scaling:** Appling Min-Max scaling to normalize all columns, bringing them to a range of 0 to 1

Model-1

MODEL ARCHITECTURE



- Autoencoder architecture

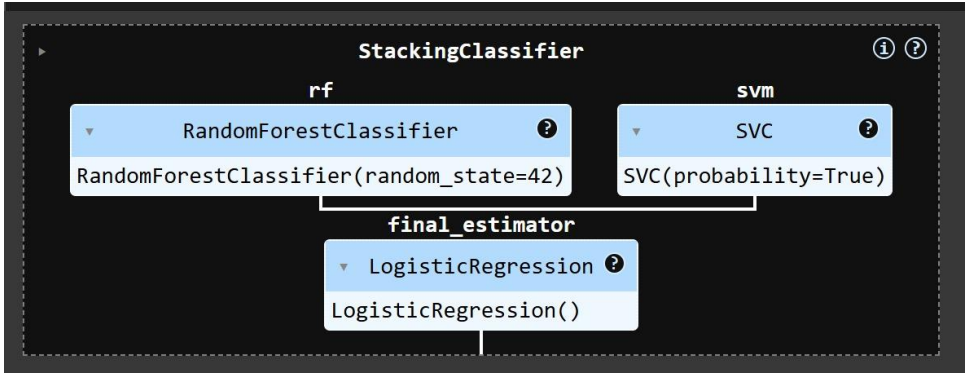


Activation function used: ReLU
Training configuration:

- Loss: MSE
- Optimizer: Adam
- No. of epoch: 10
- Batch size: 32

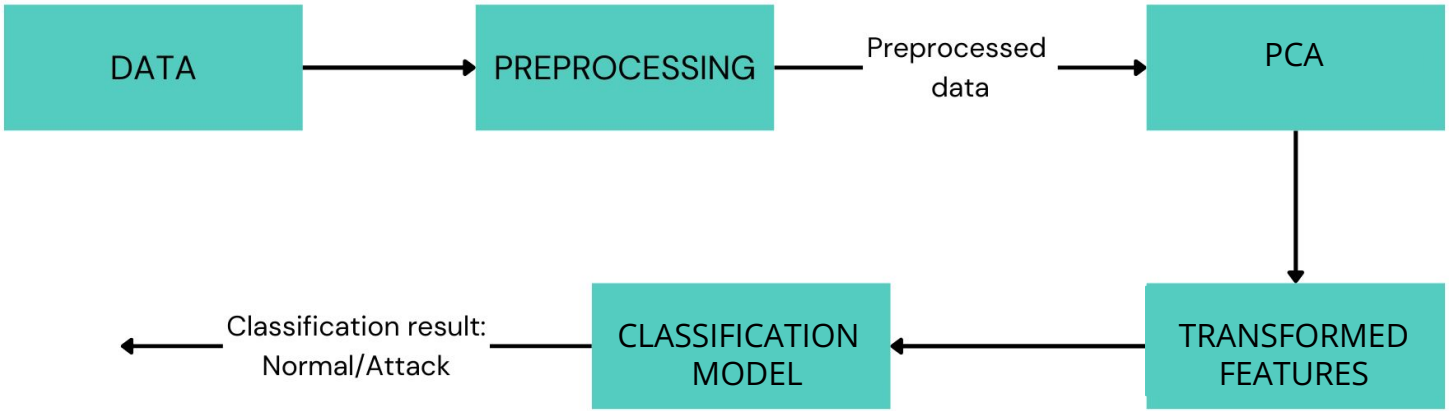
- Classification:

The 20 feature encodings from the bottleneck layer are taken out and added as the new set of features to train on the classification model. We tried two classification models:
Random forest and **Stacking classifier**



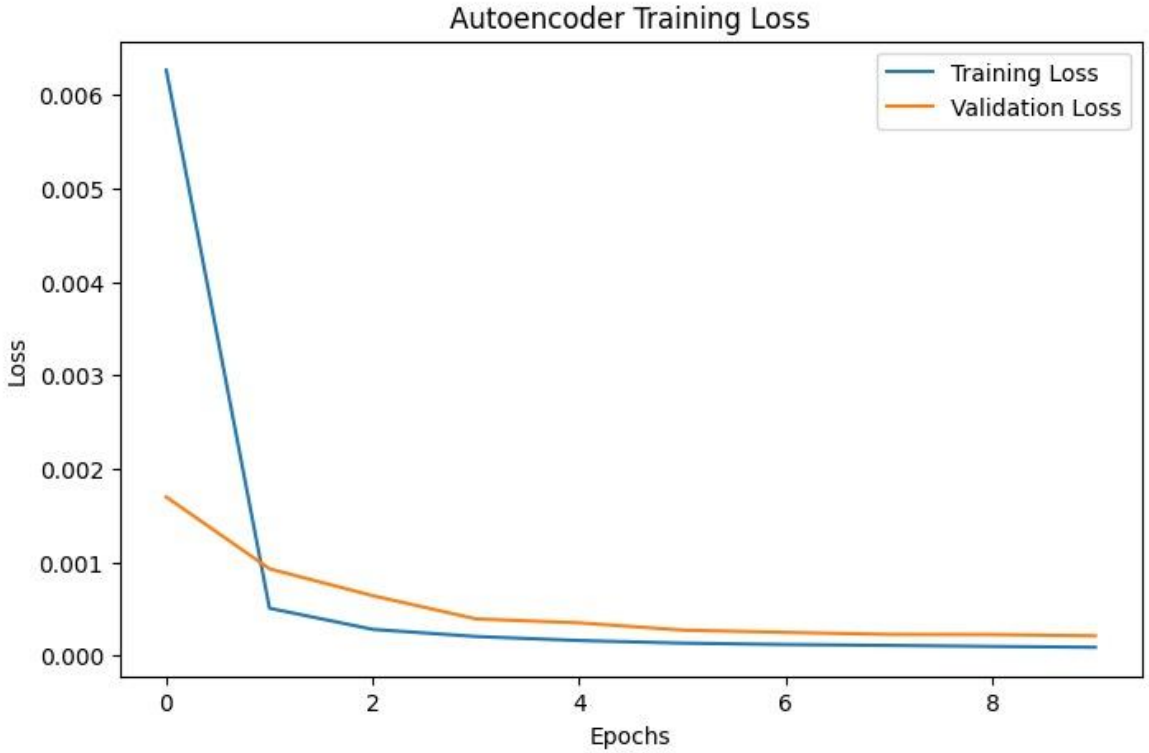
Model-2

MODEL ARCHITECTURE



The preprocessed data is given as input to PCA and the transformed features obtained from it are added as the new set of features to train to Random forest classifier.

RESULTS


Autoencoder Training Loss
Training Loss
Validation Loss
Loss
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
2
4
6
8
Epochs

MODEL	TOTAL ACCURACY	WEIGHTED PRECISION	WEIGHTED RECALL	F1 SCORE
AUTOENCODER + RANDOM FOREST	0.87	0.87	0.83	0.82
AUTOENCODER + STACKING CLASSIFIER	0.89	0.89	0.87	0.87
PCA + RANDOM FOREST	0.91	0.91	0.91	0.91

TEAM CONTRIBUTION

Rohit Parida: Autoencoder
Vinjamuri Aravind: Stacking classifier
Ramya Shri Shakthi K R: Random forest
Gurujeet Singh Raghuwanshi: Data collection and preprocessing

As a team, we contributed to PCA and other classification algorithms like SVM.

Thank you!