Detection of IoT Botnet Attacks

Utkarsh Meshram (181IT250)
Bhagyashri Bhamare (181IT111)
Chinmayi C. Ramakrishna (181IT113)

Introduction

- A ‘bot’ is a computer program which enables the operator to remotely control the infected system where it is installed.
- A network that is compromised with the attack by such bots is called a botnet.
- It is essential to detect such bots in the network to ensure safety of a system.
- The proliferation of IoT devices which can be more easily compromised than desktop computers has led to an increase in the occurrence of IoT based botnet attacks.
- There is a need to differentiate between hour and millisecond long IoT based attacks.

Abstract

- A **network-based** anomaly detection method for the IoT
- Extracts behavior snapshots of the network
- Uses deep autoencoders to detect anomalous network traffic from compromised IoT devices
- More accurate than the traditional machine learning techniques

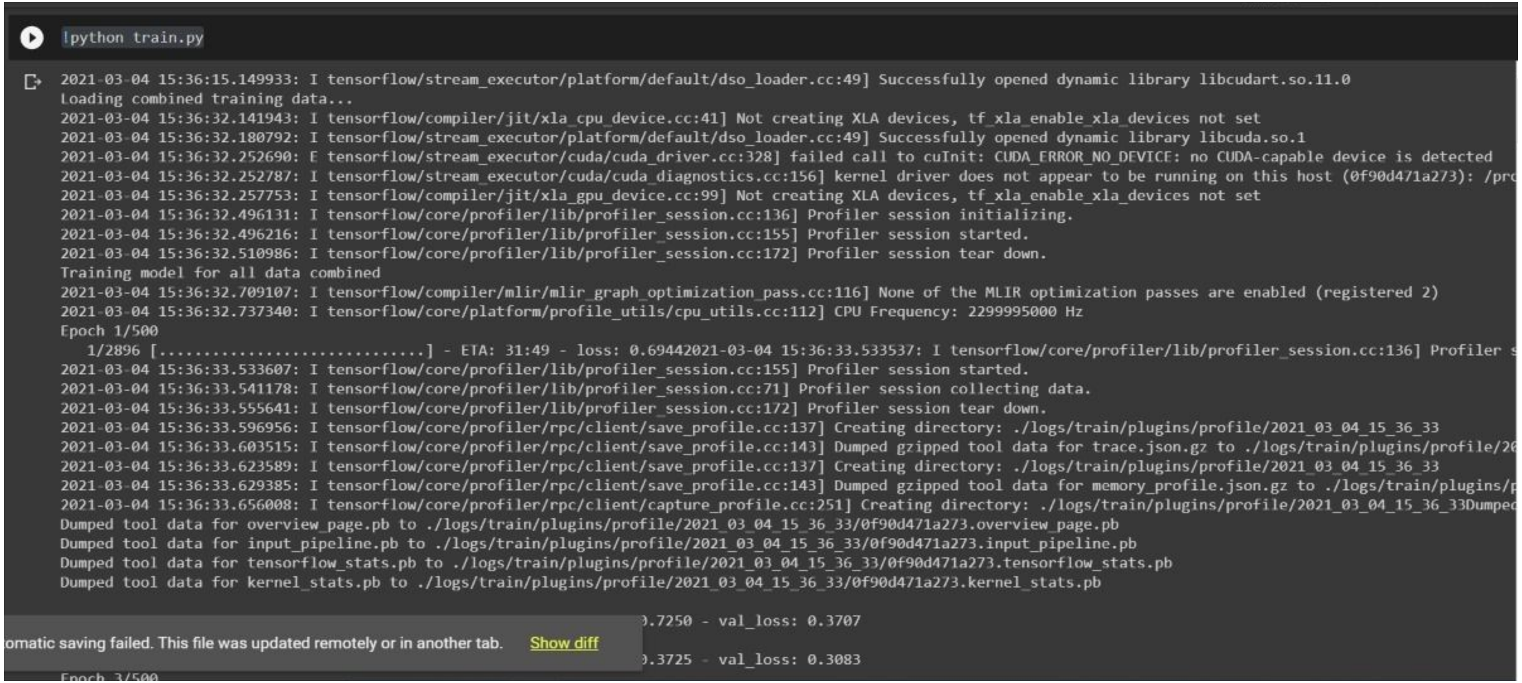
Objectives

- **Heterogeneity Tolerance:** Accommodates growing diversity of IoT devices.
- **Real world:** Detects abnormal behaviour rather than classification.
- **Efficiency:** Semi online training of autoencoders is used to improve storage efficiency.
- Use auto encoders as a complete means of botnet detection.
- Use real traffic to perform analysis

Methodology

- **Preparing the data:**
    - Splitting the datasets: train, optimise and test
    - Feature Scaling
    - Feature selection
- **Anomaly detection:**
    - Deep auto encoding
- **Attack classification:**
    - Deep neural network
- **Evaluation Metrics**

Screenshots



```
!python train.py
```

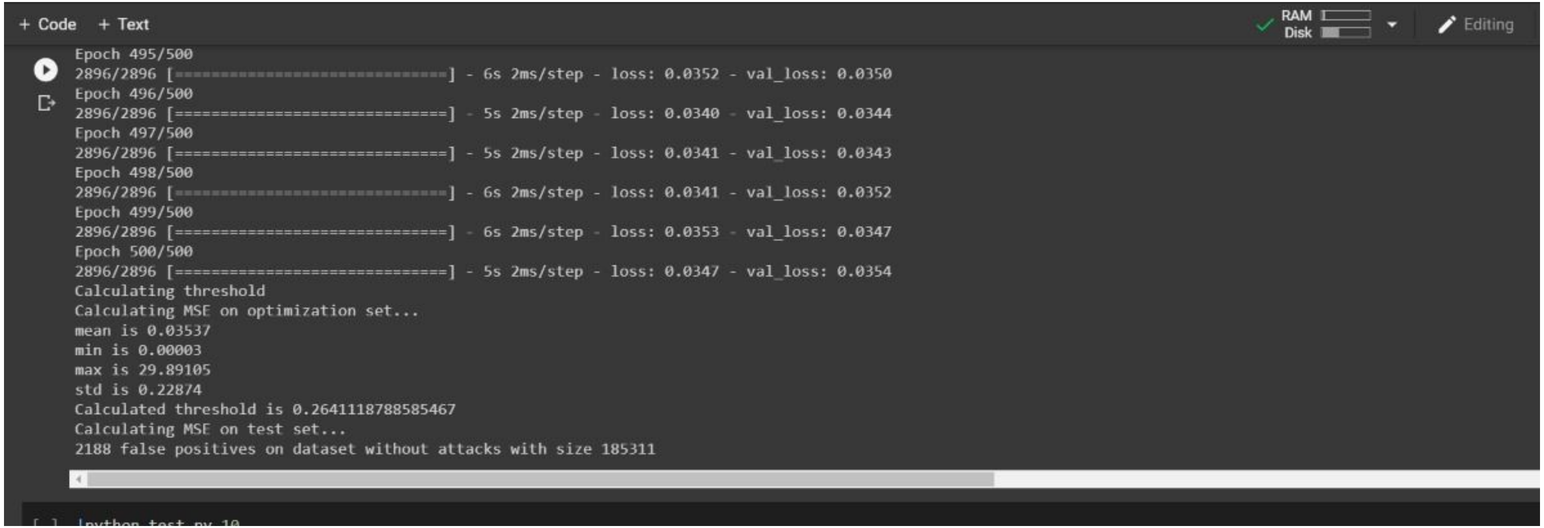
```
2021-03-04 15:36:15.149933: I tensorflow/stream_executor/platform/default/dso_loader.cc:49] Successfully opened dynamic library libcudart.so.11.0
Loading combined training data...
2021-03-04 15:36:32.141943: I tensorflow/compiler/jit/xla_cpu_device.cc:41] Not creating XLA devices, tf_xla_enable_xla_devices not set
2021-03-04 15:36:32.180792: I tensorflow/stream_executor/platform/default/dso_loader.cc:49] Successfully opened dynamic library libcuda.so.1
2021-03-04 15:36:32.252690: E tensorflow/stream_executor/cuda/cuda_driver.cc:328] failed call to cuInit: CUDA_ERROR_NO_DEVICE: no CUDA-capable device is detected
2021-03-04 15:36:32.252787: I tensorflow/stream_executor/cuda/cuda_diagnostics.cc:156] kernel driver does not appear to be running on this host (0f90d471a273): /pro
2021-03-04 15:36:32.257753: I tensorflow/compiler/jit/xla_gpu_device.cc:99] Not creating XLA devices, tf_xla_enable_xla_devices not set
2021-03-04 15:36:32.496131: I tensorflow/core/profiler/lib/profiler_session.cc:136] Profiler session initializing.
2021-03-04 15:36:32.496216: I tensorflow/core/profiler/lib/profiler_session.cc:155] Profiler session started.
2021-03-04 15:36:32.510986: I tensorflow/core/profiler/lib/profiler_session.cc:172] Profiler session tear down.
Training model for all data combined
2021-03-04 15:36:32.709107: I tensorflow/compiler/mlir/mlir_graph_optimization_pass.cc:116] None of the MLIR optimization passes are enabled (registered 2)
2021-03-04 15:36:32.737340: I tensorflow/core/platform/profile_utils/cpu_utils.cc:112] CPU Frequency: 2299995000 Hz
Epoch 1/500
   1/2896 [..............................] - ETA: 31:49 - loss: 0.69442021-03-04 15:36:33.533537: I tensorflow/core/profiler/lib/profiler_session.cc:136] Profiler s
2021-03-04 15:36:33.533607: I tensorflow/core/profiler/lib/profiler_session.cc:155] Profiler session started.
2021-03-04 15:36:33.541178: I tensorflow/core/profiler/lib/profiler_session.cc:71] Profiler session collecting data.
2021-03-04 15:36:33.555641: I tensorflow/core/profiler/lib/profiler_session.cc:172] Profiler session tear down.
2021-03-04 15:36:33.596956: I tensorflow/core/profiler/rpc/client/save_profile.cc:137] Creating directory: ./logs/train/plugins/profile/2021_03_04_15_36_33
2021-03-04 15:36:33.603515: I tensorflow/core/profiler/rpc/client/save_profile.cc:143] Dumped gzipped tool data for trace.json.gz to ./logs/train/plugins/profile/20
2021-03-04 15:36:33.623589: I tensorflow/core/profiler/rpc/client/save_profile.cc:137] Creating directory: ./logs/train/plugins/profile/2021_03_04_15_36_33
2021-03-04 15:36:33.629385: I tensorflow/core/profiler/rpc/client/save_profile.cc:143] Dumped gzipped tool data for memory_profile.json.gz to ./logs/train/plugins/p
2021-03-04 15:36:33.656008: I tensorflow/core/profiler/rpc/client/capture_profile.cc:251] Creating directory: ./logs/train/plugins/profile/2021_03_04_15_36_33Dumped
Dumped tool data for overview_page.pb to ./logs/train/plugins/profile/2021_03_04_15_36_33/0f90d471a273.overview_page.pb
Dumped tool data for input_pipeline.pb to ./logs/train/plugins/profile/2021_03_04_15_36_33/0f90d471a273.input_pipeline.pb
Dumped tool data for tensorflow_stats.pb to ./logs/train/plugins/profile/2021_03_04_15_36_33/0f90d471a273.tensorflow_stats.pb
Dumped tool data for kernel_stats.pb to ./logs/train/plugins/profile/2021_03_04_15_36_33/0f90d471a273.kernel_stats.pb

                                                                          0.7250 - val_loss: 0.3707

                                                                          0.3725 - val_loss: 0.3083
Epoch 3/500
```

omatic saving failed. This file was updated remotely or in another tab. Show diff

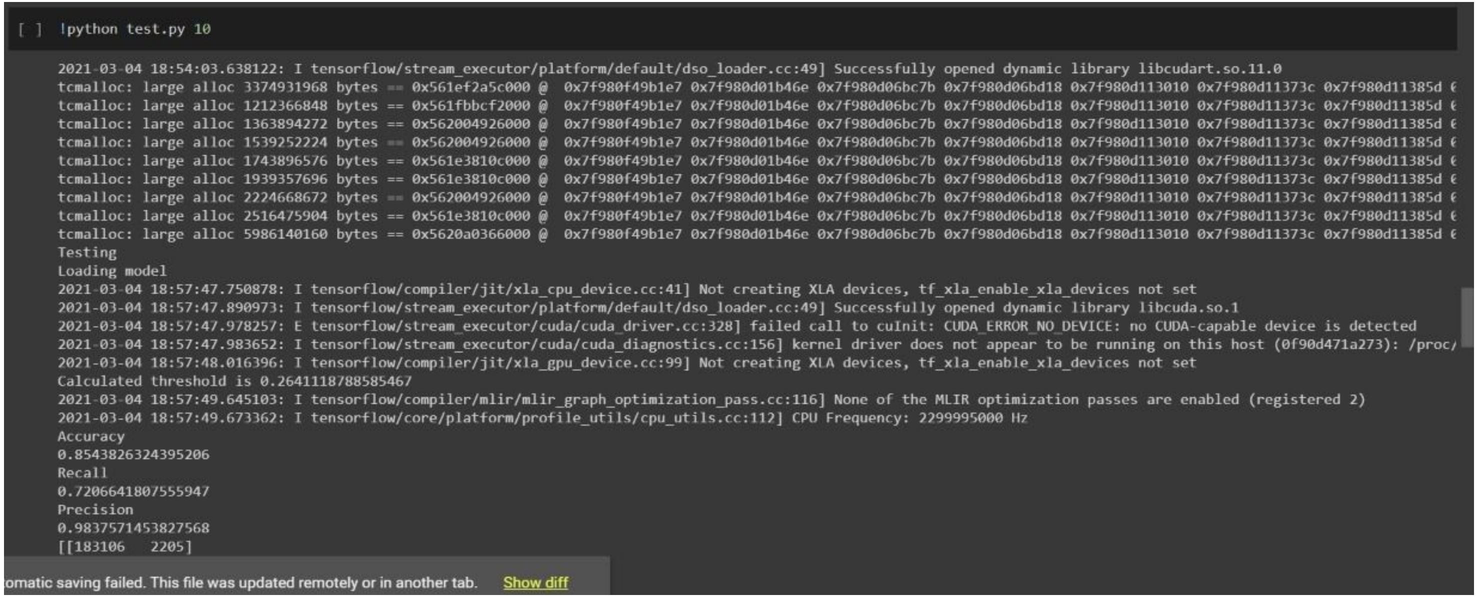
Screenshots



```
Epoch 495/500
2896/2896 [==============================] - 6s 2ms/step - loss: 0.0352 - val_loss: 0.0350
Epoch 496/500
2896/2896 [==============================] - 5s 2ms/step - loss: 0.0340 - val_loss: 0.0344
Epoch 497/500
2896/2896 [==============================] - 5s 2ms/step - loss: 0.0341 - val_loss: 0.0343
Epoch 498/500
2896/2896 [==============================] - 6s 2ms/step - loss: 0.0341 - val_loss: 0.0352
Epoch 499/500
2896/2896 [==============================] - 6s 2ms/step - loss: 0.0353 - val_loss: 0.0347
Epoch 500/500
2896/2896 [==============================] - 5s 2ms/step - loss: 0.0347 - val_loss: 0.0354
Calculating threshold
Calculating MSE on optimization set...
mean is 0.03537
min is 0.00003
max is 29.89105
std is 0.22874
Calculated threshold is 0.2641118788585467
Calculating MSE on test set...
2188 false positives on dataset without attacks with size 185311
```

```
[ ] !python test.py 10
```

Screenshots

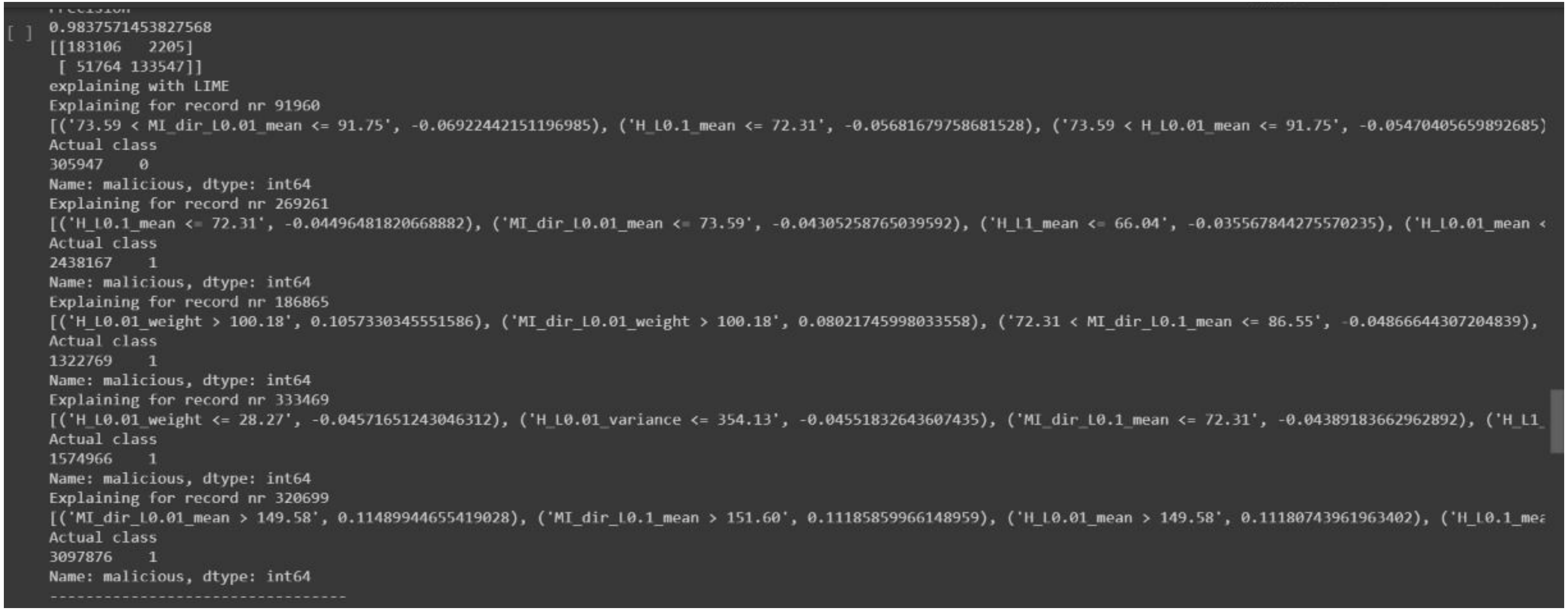


```
[ ]  !python test.py 10

     2021-03-04 18:54:03.638122: I tensorflow/stream_executor/platform/default/dso_loader.cc:49] Successfully opened dynamic library libcudart.so.11.0
     tcmalloc: large alloc 3374931968 bytes == 0x561ef2a5c000 @  0x7f980f49b1e7 0x7f980d01b46e 0x7f980d06bc7b 0x7f980d06bd18 0x7f980d113010 0x7f980d11373c 0x7f980d11385d
     tcmalloc: large alloc 1212366848 bytes == 0x561fbbcf2000 @  0x7f980f49b1e7 0x7f980d01b46e 0x7f980d06bc7b 0x7f980d06bd18 0x7f980d113010 0x7f980d11373c 0x7f980d11385d
     tcmalloc: large alloc 1363894272 bytes == 0x562004926000 @  0x7f980f49b1e7 0x7f980d01b46e 0x7f980d06bc7b 0x7f980d06bd18 0x7f980d113010 0x7f980d11373c 0x7f980d11385d
     tcmalloc: large alloc 1539252224 bytes == 0x562004926000 @  0x7f980f49b1e7 0x7f980d01b46e 0x7f980d06bc7b 0x7f980d06bd18 0x7f980d113010 0x7f980d11373c 0x7f980d11385d
     tcmalloc: large alloc 1743896576 bytes == 0x561e3810c000 @  0x7f980f49b1e7 0x7f980d01b46e 0x7f980d06bc7b 0x7f980d06bd18 0x7f980d113010 0x7f980d11373c 0x7f980d11385d
     tcmalloc: large alloc 1939357696 bytes == 0x561e3810c000 @  0x7f980f49b1e7 0x7f980d01b46e 0x7f980d06bc7b 0x7f980d06bd18 0x7f980d113010 0x7f980d11373c 0x7f980d11385d
     tcmalloc: large alloc 2224668672 bytes == 0x562004926000 @  0x7f980f49b1e7 0x7f980d01b46e 0x7f980d06bc7b 0x7f980d06bd18 0x7f980d113010 0x7f980d11373c 0x7f980d11385d
     tcmalloc: large alloc 2516475904 bytes == 0x561e3810c000 @  0x7f980f49b1e7 0x7f980d01b46e 0x7f980d06bc7b 0x7f980d06bd18 0x7f980d113010 0x7f980d11373c 0x7f980d11385d
     tcmalloc: large alloc 5986140160 bytes == 0x5620a0366000 @  0x7f980f49b1e7 0x7f980d01b46e 0x7f980d06bc7b 0x7f980d06bd18 0x7f980d113010 0x7f980d11373c 0x7f980d11385d
     Testing
     Loading model
     2021-03-04 18:57:47.750878: I tensorflow/compiler/jit/xla_cpu_device.cc:41] Not creating XLA devices, tf_xla_enable_xla_devices not set
     2021-03-04 18:57:47.890973: I tensorflow/stream_executor/platform/default/dso_loader.cc:49] Successfully opened dynamic library libcuda.so.1
     2021-03-04 18:57:47.978257: E tensorflow/stream_executor/cuda/cuda_driver.cc:328] failed call to cuInit: CUDA_ERROR_NO_DEVICE: no CUDA-capable device is detected
     2021-03-04 18:57:47.983652: I tensorflow/stream_executor/cuda/cuda_diagnostics.cc:156] kernel driver does not appear to be running on this host (0f90d471a273): /proc/
     2021-03-04 18:57:48.016396: I tensorflow/compiler/jit/xla_gpu_device.cc:99] Not creating XLA devices, tf_xla_enable_xla_devices not set
     Calculated threshold is 0.2641118788585467
     2021-03-04 18:57:49.645103: I tensorflow/compiler/mlir/mlir_graph_optimization_pass.cc:116] None of the MLIR optimization passes are enabled (registered 2)
     2021-03-04 18:57:49.673362: I tensorflow/core/platform/profile_utils/cpu_utils.cc:112] CPU Frequency: 2299995000 Hz
     Accuracy
     0.8543826324395206
     Recall
     0.7206641807555947
     Precision
     0.9837571453827568
     [[183106   2205]
```

omatic saving failed. This file was updated remotely or in another tab.

Screenshots



```
Precision
0.9837571453827568
[[183106   2205]
 [ 51764 133547]]
explaining with LIME
Explaining for record nr 91960
[('73.59 < MI_dir_L0.01_mean <= 91.75', -0.06922442151196985), ('H_L0.1_mean <= 72.31', -0.05681679758681528), ('73.59 < H_L0.01_mean <= 91.75', -0.05470405659892685)
Actual class
305947    0
Name: malicious, dtype: int64
Explaining for record nr 269261
[('H_L0.1_mean <= 72.31', -0.04496481820668882), ('MI_dir_L0.01_mean <= 73.59', -0.04305258765039592), ('H_L1_mean <= 66.04', -0.035567844275570235), ('H_L0.01_mean <
Actual class
2438167    1
Name: malicious, dtype: int64
Explaining for record nr 186865
[('H_L0.01_weight > 100.18', 0.1057330345551586), ('MI_dir_L0.01_weight > 100.18', 0.08021745998033558), ('72.31 < MI_dir_L0.1_mean <= 86.55', -0.04866644307204839),
Actual class
1322769    1
Name: malicious, dtype: int64
Explaining for record nr 333469
[('H_L0.01_weight <= 28.27', -0.04571651243046312), ('H_L0.01_variance <= 354.13', -0.04551832643607435), ('MI_dir_L0.1_mean <= 72.31', -0.04389183662962892), ('H_L1_
Actual class
1574966    1
Name: malicious, dtype: int64
Explaining for record nr 320699
[('MI_dir_L0.01_mean > 149.58', 0.11489944655419028), ('MI_dir_L0.1_mean > 151.60', 0.11185859966148959), ('H_L0.01_mean > 149.58', 0.11180743961963402), ('H_L0.1_mea
Actual class
3097876    1
Name: malicious, dtype: int64
---------------------------------
```

Conclusion

- The data is obtained by extracting a total of 115 traffic statistics.
- Data set is split into train and test subsets as 80:20.
- Use of auto encoders for two different networks: encoder and decoder.
- Minimise loss function by minimising mean squared error between the original input and the reconstruction.
- Set a threshold to consider errors.

Conclusion

- Autoencoder uses 5 hidden layers of sizes 0.75, 0.5, 0.25, 0.5, 0.75 of the input feature vector size.
- Hyperbolic tangent is used as an activation function for our hidden unit neuron.