In [1]:
```
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Model
from tensorflow.keras.layers import Input
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import LeakyReLU
from tensorflow.keras.layers import BatchNormalization
from tensorflow.keras.models import load_model
```

In [2]:
```
import numpy as np
import pandas as pd
train = pd.read_csv('UNSW_NB15_training-set.csv')
test = pd.read_csv('UNSW_NB15_testing-set.csv')
from sklearn.preprocessing import OrdinalEncoder
ord_enc = OrdinalEncoder()
```

In [3]:
```
train['proto_code'] = ord_enc.fit_transform(train[['proto']])
train[['proto','proto_code']].head(175341)
train['state_code'] = ord_enc.fit_transform(train[['state']])
train[['state','state_code']].head(175341)
train_updated = train.replace('-',np.nan)
print(train_updated)
final_train = train_updated.fillna("nodata")
print(final_train)
final_train['service_code'] = ord_enc.fit_transform(final_train[['service']])
final_train[['service','service_code']].head(175341)
```

```
            id       dur proto service state  spkts  dpkts  sbytes  dbytes  \
0            1  0.121478   tcp     NaN   FIN      6      4     258     172
1            2  0.649902   tcp     NaN   FIN     14     38     734   42014
2            3  1.623129   tcp     NaN   FIN      8     16     364   13186
3            4  1.681642   tcp     ftp   FIN     12     12     628     770
4            5  0.449454   tcp     NaN   FIN     10      6     534     268
...        ...       ...   ...     ...   ...    ...    ...     ...     ...
175336  175337  0.000009   udp     dns   INT      2      0     114       0
175337  175338  0.505762   tcp     NaN   FIN     10      8     620     354
175338  175339  0.000009   udp     dns   INT      2      0     114       0
175339  175340  0.000009   udp     dns   INT      2      0     114       0
175340  175341  0.000009   udp     dns   INT      2      0     114       0

                rate  ...  is_ftp_login  ct_ftp_cmd  ct_flw_http_mthd  \
0          74.087490  ...             0           0                 0
1          78.473372  ...             0           0                 0
2          14.170161  ...             0           0                 0
3          13.677108  ...             1           1                 0
4          33.373826  ...             0           0                 0
...              ...  ...           ...         ...               ...
175336  111111.107200  ...             0           0                 0
175337     33.612649  ...             0           0                 0
175338  111111.107200  ...             0           0                 0
175339  111111.107200  ...             0           0                 0
175340  111111.107200  ...             0           0                 0

        ct_src_ltm  ct_srv_dst  is_sm_ips_ports  attack_cat  label  \
0                1           1                0      Normal      0
1                1           6                0      Normal      0
2                2           6                0      Normal      0
3                2           1                0      Normal      0
4                2          39                0      Normal      0
...            ...         ...              ...         ...    ...
175336          24          24                0     Generic      1
175337           1           1                0   Shellcode      1
175338           3          12                0     Generic      1
175339          30          30                0     Generic      1
175340          30          30                0     Generic      1

        proto_code  state_code
0            113.0         2.0
1            113.0         2.0
2            113.0         2.0
3            113.0         2.0
4            113.0         2.0
...            ...         ...
175336       119.0         3.0
175337       113.0         2.0
175338       119.0         3.0
175339       119.0         3.0
175340       119.0         3.0

[175341 rows x 47 columns]
            id       dur proto service state  spkts  dpkts  sbytes  dbytes  \
0            1  0.121478   tcp  nodata   FIN      6      4     258     172
1            2  0.649902   tcp  nodata   FIN     14     38     734   42014
2            3  1.623129   tcp  nodata   FIN      8     16     364   13186
3            4  1.681642   tcp     ftp   FIN     12     12     628     770
4            5  0.449454   tcp  nodata   FIN     10      6     534     268
...        ...       ...   ...     ...   ...    ...    ...     ...     ...
175336  175337  0.000009   udp     dns   INT      2      0     114       0
175337  175338  0.505762   tcp  nodata   FIN     10      8     620     354
175338  175339  0.000009   udp     dns   INT      2      0     114       0
175339  175340  0.000009   udp     dns   INT      2      0     114       0
175340  175341  0.000009   udp     dns   INT      2      0     114       0

                rate  ...  is_ftp_login  ct_ftp_cmd  ct_flw_http_mthd  \
0          74.087490  ...             0           0                 0
1          78.473372  ...             0           0                 0
2          14.170161  ...             0           0                 0
```

```
3          13.677108  ...             1          1                0
4          33.373826  ...             0          0                0
...              ...  ...           ...        ...              ...
175336  111111.107200  ...             0          0                0
175337     33.612649  ...             0          0                0
175338  111111.107200  ...             0          0                0
175339  111111.107200  ...             0          0                0
175340  111111.107200  ...             0          0                0

        ct_src_ltm  ct_srv_dst  is_sm_ips_ports  attack_cat  label  \
0                1           1                0      Normal      0
1                1           6                0      Normal      0
2                2           6                0      Normal      0
3                2           1                0      Normal      0
4                2          39                0      Normal      0
...            ...         ...              ...         ...    ...
175336          24          24                0     Generic      1
175337           1           1                0   Shellcode      1
175338           3          12                0     Generic      1
175339          30          30                0     Generic      1
175340          30          30                0     Generic      1

        proto_code  state_code
0            113.0         2.0
1            113.0         2.0
2            113.0         2.0
3            113.0         2.0
4            113.0         2.0
...            ...         ...
175336       119.0         3.0
175337       113.0         2.0
175338       119.0         3.0
175339       119.0         3.0
175340       119.0         3.0

[175341 rows x 47 columns]
```

Out[3]:

	service	service_code
0	nodata	6.0
1	nodata	6.0
2	nodata	6.0
3	ftp	2.0
4	nodata	6.0
...	...	...
175336	dns	1.0
175337	nodata	6.0
175338	dns	1.0
175339	dns	1.0
175340	dns	1.0

175341 rows × 2 columns

In [4]:

```
test['proto_code'] = ord_enc.fit_transform(test[['proto']])
test[['proto','proto_code']].head(175341)
test['state_code'] = ord_enc.fit_transform(test[['state']])
test[['state','state_code']].head(175341)
test_updated = test.replace('-',np.nan)
print(test_updated)
final_test = test_updated.fillna("nodata")
print(final_test)
final_test['service_code'] = ord_enc.fit_transform(final_test[['service']])
final_test[['service','service_code']].head(175341)
```

```
          id       dur proto service state  spkts  dpkts  sbytes  dbytes  \
0          1  0.000011   udp     NaN   INT      2      0     496       0
1          2  0.000008   udp     NaN   INT      2      0    1762       0
2          3  0.000005   udp     NaN   INT      2      0    1068       0
3          4  0.000006   udp     NaN   INT      2      0     900       0
4          5  0.000010   udp     NaN   INT      2      0    2126       0
...      ...       ...   ...     ...   ...    ...    ...     ...     ...
82327  82328  0.000005   udp     NaN   INT      2      0     104       0
82328  82329  1.106101   tcp     NaN   FIN     20      8   18062     354
82329  82330  0.000000   arp     NaN   INT      1      0      46       0
82330  82331  0.000000   arp     NaN   INT      1      0      46       0
82331  82332  0.000009   udp     NaN   INT      2      0     104       0

                rate  ...  is_ftp_login  ct_ftp_cmd  ct_flw_http_mthd  \
0       90909.090200  ...             0           0                 0
1      125000.000300  ...             0           0                 0
2      200000.005100  ...             0           0                 0
3      166666.660800  ...             0           0                 0
4      100000.002500  ...             0           0                 0
...              ...  ...           ...         ...               ...
82327  200000.005100  ...             0           0                 0
82328      24.410067  ...             0           0                 0
82329       0.000000  ...             0           0                 0
82330       0.000000  ...             0           0                 0
82331  111111.107200  ...             0           0                 0

       ct_src_ltm  ct_srv_dst  is_sm_ips_ports  attack_cat  label  proto_code  \
0               1           2                0      Normal      0       117.0
```

```
1              1           2                0      Normal      0       117.0
2              1           3                0      Normal      0       117.0
3              2           3                0      Normal      0       117.0
4              2           3                0      Normal      0       117.0
...          ...         ...              ...         ...    ...         ...
82327          2           1                0      Normal      0       117.0
82328          3           2                0      Normal      0       111.0
82329          1           1                1      Normal      0         6.0
82330          1           1                1      Normal      0         6.0
82331          1           1                0      Normal      0       117.0

       state_code
0             4.0
1             4.0
2             4.0
3             4.0
4             4.0
...           ...
82327         4.0
82328         3.0
82329         4.0
82330         4.0
82331         4.0

[82332 rows x 47 columns]
         id       dur proto service state  spkts  dpkts  sbytes  dbytes  \
0         1  0.000011   udp  nodata   INT      2      0     496       0
1         2  0.000008   udp  nodata   INT      2      0    1762       0
2         3  0.000005   udp  nodata   INT      2      0    1068       0
3         4  0.000006   udp  nodata   INT      2      0     900       0
4         5  0.000010   udp  nodata   INT      2      0    2126       0
...     ...       ...   ...     ...   ...    ...    ...     ...     ...
82327  82328  0.000005   udp  nodata   INT      2      0     104       0
82328  82329  1.106101   tcp  nodata   FIN     20      8   18062     354
82329  82330  0.000000   arp  nodata   INT      1      0      46       0
82330  82331  0.000000   arp  nodata   INT      1      0      46       0
82331  82332  0.000009   udp  nodata   INT      2      0     104       0

                rate  ...  is_ftp_login  ct_ftp_cmd  ct_flw_http_mthd  \
0       90909.090200  ...             0           0                 0
1      125000.000300  ...             0           0                 0
2      200000.005100  ...             0           0                 0
3      166666.660800  ...             0           0                 0
4      100000.002500  ...             0           0                 0
...              ...  ...           ...         ...               ...
82327  200000.005100  ...             0           0                 0
82328      24.410067  ...             0           0                 0
82329       0.000000  ...             0           0                 0
82330       0.000000  ...             0           0                 0
82331  111111.107200  ...             0           0                 0

       ct_src_ltm  ct_srv_dst  is_sm_ips_ports  attack_cat  label  proto_code  \
0               1           2                0      Normal      0       117.0
1               1           2                0      Normal      0       117.0
2               1           3                0      Normal      0       117.0
3               2           3                0      Normal      0       117.0
4               2           3                0      Normal      0       117.0
...           ...         ...              ...         ...    ...         ...
82327           2           1                0      Normal      0       117.0
82328           3           2                0      Normal      0       111.0
82329           1           1                1      Normal      0         6.0
82330           1           1                1      Normal      0         6.0
82331           1           1                0      Normal      0       117.0

       state_code
0             4.0
1             4.0
2             4.0
3             4.0
4             4.0
...           ...
82327         4.0
82328         3.0
82329         4.0
82330         4.0
82331         4.0

[82332 rows x 47 columns]
```

Out[4]:

	service	service_code
0	nodata	6.0
1	nodata	6.0
2	nodata	6.0
3	nodata	6.0
4	nodata	6.0
...	...	...
82327	nodata	6.0
82328	nodata	6.0
82329	nodata	6.0
82330	nodata	6.0
82331	nodata	6.0

82332 rows × 2 columns

In [5]:
```
x1 = final_train[['id','dur','spkts','dpkts','proto_code','state_code',
          'sbytes','dbytes','rate','sttl','dttl','sload','dload','sloss','dloss',
        'sinpkt','dinpkt','sjit','djit','swin','stcpb',
        'dtcpb','dwin','tcprtt','synack','ackdat','smean','dmean','trans_depth','response_body_len','ct_srv_src','ct_stat
       ,'ct_dst_ltm','ct_src_dport_ltm','ct_dst_sport_ltm','ct_dst_src_ltm','is_ftp_login',
        'ct_ftp_cmd','ct_flw_http_mthd','ct_src_ltm','ct_srv_dst','is_sm_ips_ports']]
y1 = final_train['label']
x2 = final_test[['id','dur','spkts','dpkts','proto_code','state_code',
          'sbytes','dbytes','rate','sttl','dttl','sload','dload','sloss','dloss',
        'sinpkt','dinpkt','sjit','djit','swin','stcpb',
        'dtcpb','dwin','tcprtt','synack','ackdat','smean','dmean','trans_depth','response_body_len','ct_srv_src','ct_stat
       ,'ct_dst_ltm','ct_src_dport_ltm','ct_dst_sport_ltm','ct_dst_src_ltm','is_ftp_login',
        'ct_ftp_cmd','ct_flw_http_mthd','ct_src_ltm','ct_srv_dst','is_sm_ips_ports']]
y2 = final_test['label']
```

In [6]:
```
model = MinMaxScaler()
model.fit(x1)
x1 = model.transform(x1)
x2 = model.transform(x2)
```

In [7]:
```
n_inputs = x1.shape[1]
n_inputs = x2.shape[1]
```

In [8]:
```
input_data_shape = Input(shape=(n_inputs,))
encoder = Dense(n_inputs*2)(input_data_shape)
encoder = BatchNormalization()(encoder)
encoder = LeakyReLU()(encoder)
encoder = Dense(n_inputs)(encoder)
encoder = BatchNormalization()(encoder)
encoder = LeakyReLU()(encoder)
n_bottleneck = round(float(n_inputs)/2.0)
bottleneck = Dense(n_bottleneck)(encoder)
decoder = Dense(n_inputs)(bottleneck)
decoder = BatchNormalization()(decoder)
decoder = LeakyReLU()(decoder)
decoder = Dense(n_inputs*2)(bottleneck)
decoder = BatchNormalization()(decoder)
decoder = LeakyReLU()(decoder)
```

In [9]:
```
output = Dense(n_inputs,activation = 'linear')(decoder)
model_AE = Model(inputs = input_data_shape,outputs = output)
model_AE.compile(optimizer = 'adam',loss='mse')
model_AE.summary()
```

```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 42)]              0

 dense (Dense)               (None, 84)                3612

 batch_normalization (BatchN  (None, 84)               336
 ormalization)

 leaky_re_lu (LeakyReLU)     (None, 84)                0

 dense_1 (Dense)             (None, 42)                3570

 batch_normalization_1 (Batc  (None, 42)               168
 hNormalization)

 leaky_re_lu_1 (LeakyReLU)   (None, 42)                0

 dense_2 (Dense)             (None, 21)                903

 dense_4 (Dense)             (None, 84)                1848

 batch_normalization_3 (Batc  (None, 84)               336
 hNormalization)

 leaky_re_lu_3 (LeakyReLU)   (None, 84)                0

 dense_5 (Dense)             (None, 42)                3570

=================================================================
Total params: 14,343
Trainable params: 13,923
Non-trainable params: 420
_________________________________________________________________
```

In [10]:
```
history = model_AE.fit(x1,x1,epochs = 6,batch_size = 16,verbose = 2,validation_data = (x2,x2))
```

```
Epoch 1/6
10959/10959 - 42s - loss: 0.0046 - val_loss: 9.8798e-04 - 42s/epoch - 4ms/step
Epoch 2/6
10959/10959 - 31s - loss: 0.0015 - val_loss: 8.9605e-04 - 31s/epoch - 3ms/step
Epoch 3/6
10959/10959 - 31s - loss: 0.0012 - val_loss: 8.7522e-04 - 31s/epoch - 3ms/step
Epoch 4/6
10959/10959 - 31s - loss: 9.6656e-04 - val_loss: 8.2763e-04 - 31s/epoch - 3ms/step
```

```
Epoch 5/6
10959/10959 - 31s - loss: 8.4750e-04 - val_loss: 9.0745e-04 - 31s/epoch - 3ms/step
Epoch 6/6
10959/10959 - 32s - loss: 7.9423e-04 - val_loss: 9.8309e-04 - 32s/epoch - 3ms/step
```

In [11]:

```
encoder = Model(inputs = input_data_shape,outputs = bottleneck)
encoder.save('encoder.h5')
```

```
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will
be empty until you train or evaluate the model.
```

In [12]:

```
from sklearn.metrics import accuracy_score
model = MinMaxScaler()
model.fit(x1)
x1 = model.transform(x1)
x2 = model.transform(x2)
from xgboost import XGBClassifier
import xgboost as xgb
params = {
            'objective':'binary:logistic',
            'max_depth': 4,
            'alpha': 10,
            'learning_rate': 0.1,
            'n_estimators':100
        }
xgb_clf = XGBClassifier(**params)
xgb_clf.fit(x1, y1)
y_pred=xgb_clf.predict(x2)
print(y_pred)
accuracy = accuracy_score(y2, y_pred)*100
print("Accuracy before feature selection:-", accuracy)
```

```
C:\Users\admin\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder in XGBClassifie
r is deprecated and will be removed in a future release. To remove this warning, do the following: 1) Pass option use_lab
el_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[20:35:43] WARNING: C:/Users/Administrator/workspace/xgboost-win64_release_1.5.1/src/learner.cc:1115: Starting in XGBoost
1.3.0, the default evaluation metric used with the objective 'binary:logistic' was changed from 'error' to 'logloss'. Exp
licitly set eval_metric if you'd like to restore the old behavior.
[0 0 0 ... 0 0 1]
Accuracy before feature selection:- 49.46557838993344
```

In [13]:

```
encoder = load_model('encoder.h5')
x1_encode = encoder.predict(x1)
x2_encode = encoder.predict(x2)
params = {
            'objective':'binary:logistic',
            'max_depth': 4,
            'alpha': 10,
            'learning_rate': 0.1,
            'n_estimators':100
        }
model_final = XGBClassifier(**params)
model_final.fit(x1_encode, y1)
y_pred_new = model_final.predict(x2_encode)
accuracy = accuracy_score(y2, y_pred_new)*100
print("Accuracy after feature selection:-", accuracy)
```

```
WARNING:tensorflow:No training configuration found in the save file, so the model was *not* compiled. Compile it manuall
y.
C:\Users\admin\anaconda3\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder in XGBClassifie
r is deprecated and will be removed in a future release. To remove this warning, do the following: 1) Pass option use_lab
el_encoder=False when constructing XGBClassifier object; and 2) Encode your labels (y) as integers starting with 0, i.e.
0, 1, 2, ..., [num_class - 1].
  warnings.warn(label_encoder_deprecation_msg, UserWarning)
[20:36:43] WARNING: C:/Users/Administrator/workspace/xgboost-win64_release_1.5.1/src/learner.cc:1115: Starting in XGBoost
1.3.0, the default evaluation metric used with the objective 'binary:logistic' was changed from 'error' to 'logloss'. Exp
licitly set eval_metric if you'd like to restore the old behavior.
Accuracy after feature selection:- 65.32211047952194
```

In [ ]: