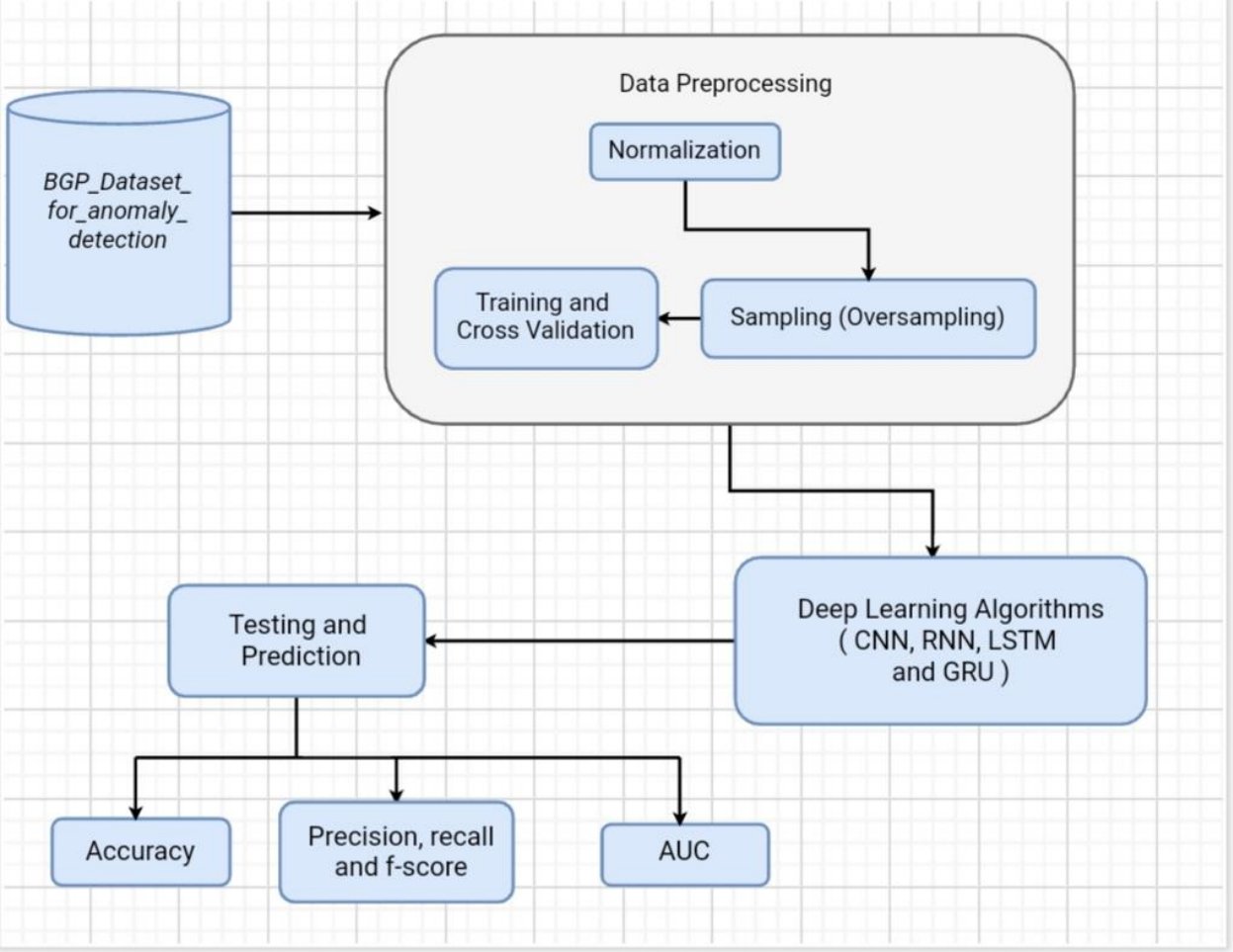
REVIEW 2

RAHUL ANAND

18BIT0436

TOPIC :INTRUSION DETECTION USING DEEP LEARNING(LSTM)

Design and description of system:



BGP datasets for anomaly detection:

BGP stands for Border Gateway Protocol. BGP is the protocol that makes the Internet work. It does this by enabling data routing on the Internet. when someone submits data across the Internet, BGP is responsible for looking at all of the available paths that data could travel and picking the best route, which usually means hopping between autonomous systems.

Three well-known Border Gateway Protocol (BGP) anomalies Slammer, Nimda, and Code Red I occurred in January 2003, September 2001, and July 2001, respectively.

We are using Border Gateway Protocol anomalies for training and testing our algorithm. We are using deep learning algorithm like CNN for anomaly detection.

Data Preprocessing:

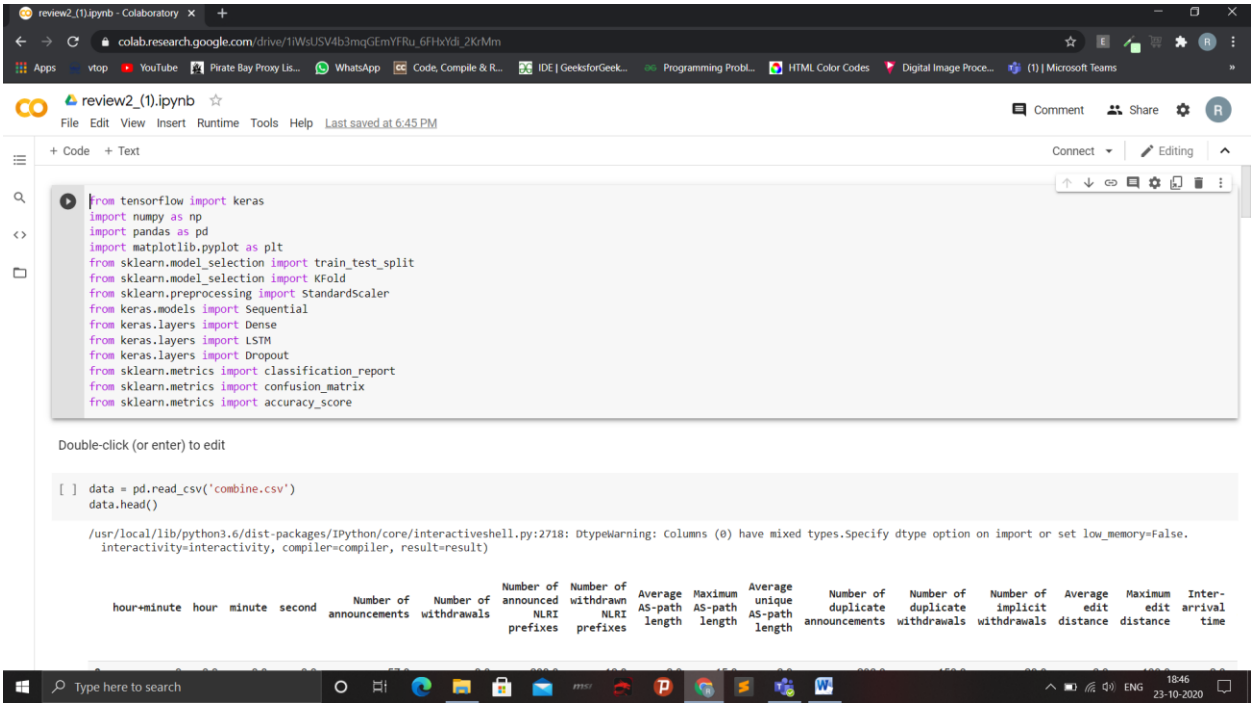
Data Preprocessing is a technique that is used to convert the raw data into a clean data set. Whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

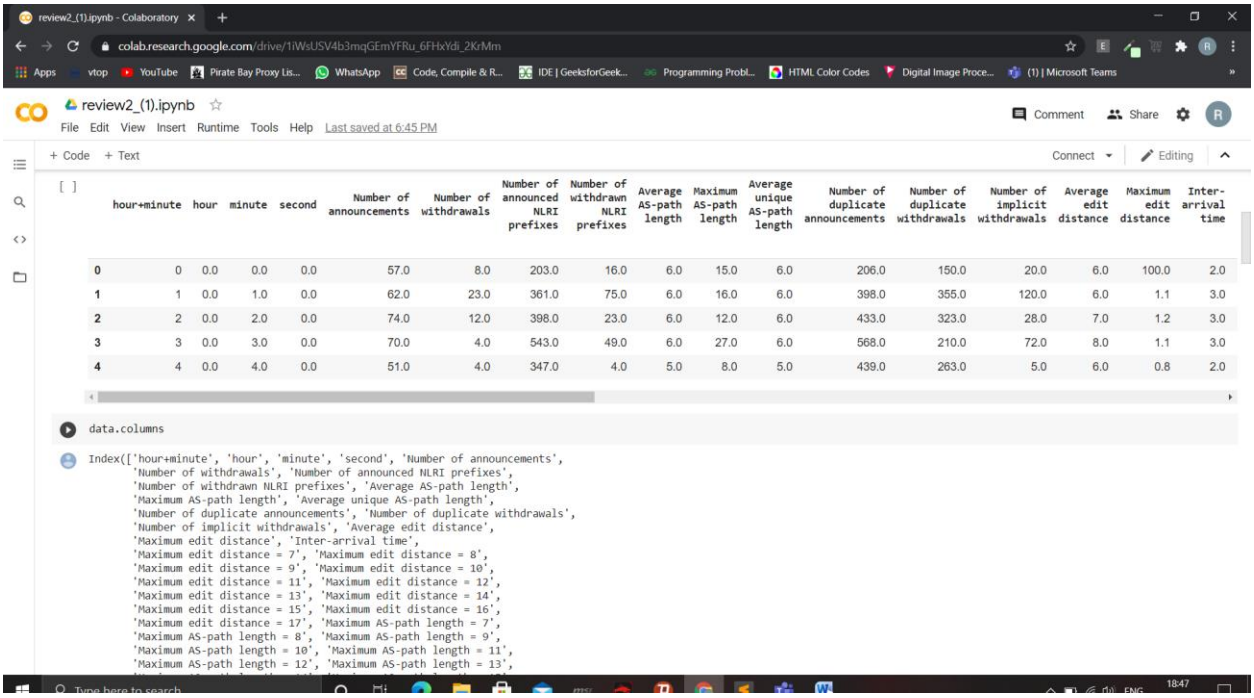
For achieving better results from the applied model in Deep learning projects the format of the data has to be in a proper manner. Data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in one data set and best out of them is chosen.

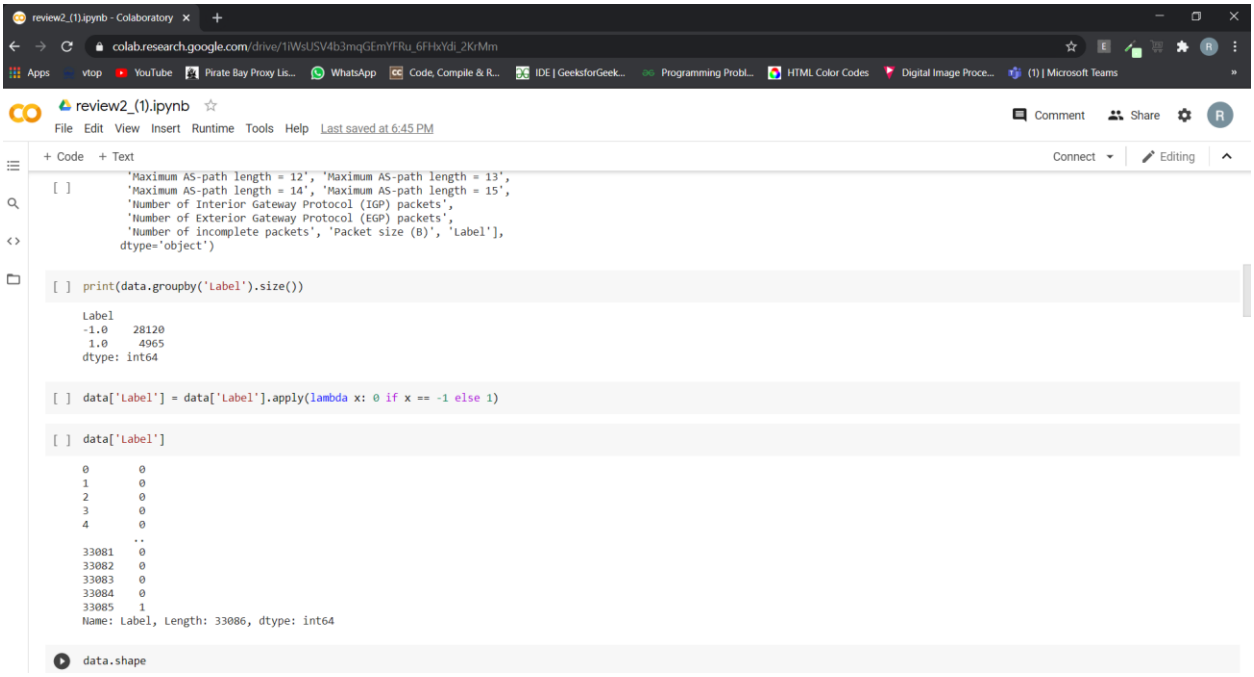
For data preprocessing we have used various techniques like Normalization, Random Oversampling and Hyper parameter tuning of individual algorithm parameters.

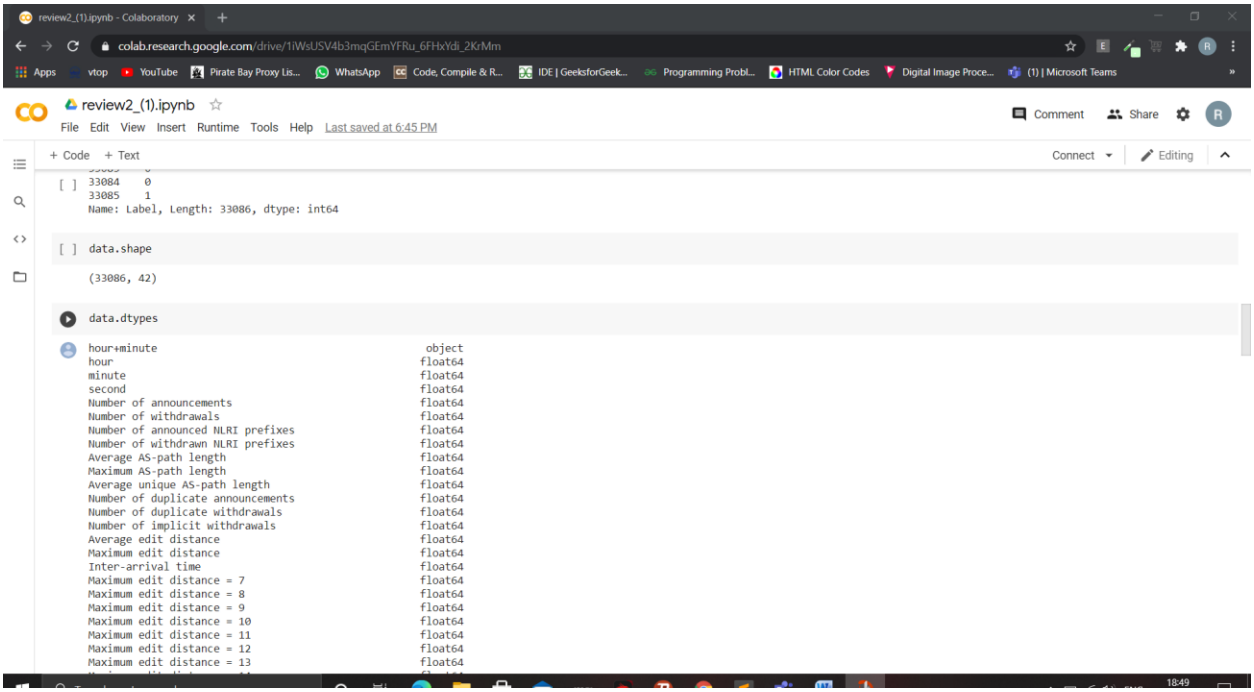
Deep learning Algorithm like CNN is used for detecting anomalous data. BGP datasets for anomaly detection is used for training and testing of the algorithm that we have used in this project.

Testing and Prediction is done using the deep learning algorithm used and various performance metrics like precision, recall, f-score is calculated for the algorithm.


review2_(1).ipynb - Colaboratory
colab.research.google.com/drive/1iWsUSV4b3mqGEmYFRu_6FHxYdi_2KrMm
review2_(1).ipynb
File Edit View Insert Runtime Tools Help Last saved at 6:45 PM
Comment Share
+ Code + Text
Connect Editing
from tensorflow import keras
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.model_selection import KFold
from sklearn.preprocessing import StandardScaler
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras.layers import Dropout
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
Double-click (or enter) to edit
[ ] data = pd.read_csv('combine.csv')
data.head()
/usr/local/lib/python3.6/dist-packages/IPython/core/interactiveshell.py:2718: DtypeWarning: Columns (0) have mixed types.Specify dtype option on import or set low_memory=False.
interactivity=interactivity, compiler=compiler, result=result)
hour+minute hour minute second Number of announcements Number of withdrawals Number of announced NLRI prefixes Number of withdrawn NLRI prefixes Average AS-path length Maximum AS-path length Average unique AS-path length Number of duplicate announcements Number of duplicate withdrawals Number of implicit withdrawals Average edit distance Maximum edit distance Inter-arrival time
Type here to search
18:46
23-10-2020


review2_(1).ipynb - Colaboratory
colab.research.google.com/drive/1iWsUSV4b3mqGEmYFRu_6FHxYdi_2KrMm
review2_(1).ipynb
File Edit View Insert Runtime Tools Help Last saved at 6:45 PM
Comment Share
+ Code + Text
Connect Editing
[ ]
hour+minute hour minute second Number of announcements Number of withdrawals Number of announced NLRI prefixes Number of withdrawn NLRI prefixes Average AS-path length Maximum AS-path length Average unique AS-path length Number of duplicate announcements Number of duplicate withdrawals Number of implicit withdrawals Average edit distance Maximum edit distance Inter-arrival time
0 0 0.0 0.0 0.0 57.0 8.0 203.0 16.0 6.0 15.0 6.0 206.0 150.0 20.0 6.0 100.0 2.0
1 1 0.0 1.0 0.0 62.0 23.0 361.0 75.0 6.0 16.0 6.0 398.0 355.0 120.0 6.0 1.1 3.0
2 2 0.0 2.0 0.0 74.0 12.0 398.0 23.0 6.0 12.0 6.0 433.0 323.0 28.0 7.0 1.2 3.0
3 3 0.0 3.0 0.0 70.0 4.0 543.0 49.0 6.0 27.0 6.0 568.0 210.0 72.0 8.0 1.1 3.0
4 4 0.0 4.0 0.0 51.0 4.0 347.0 4.0 5.0 8.0 5.0 439.0 263.0 5.0 6.0 0.8 2.0
data.columns
Index(['hour+minute', 'hour', 'minute', 'second', 'Number of announcements',
'Number of withdrawals', 'Number of announced NLRI prefixes',
'Number of withdrawn NLRI prefixes', 'Average AS-path length',
'Maximum AS-path length', 'Average unique AS-path length',
'Number of duplicate announcements', 'Number of duplicate withdrawals',
'Number of implicit withdrawals', 'Average edit distance',
'Maximum edit distance', 'Inter-arrival time',
'Maximum edit distance = 7', 'Maximum edit distance = 8',
'Maximum edit distance = 9', 'Maximum edit distance = 10',
'Maximum edit distance = 11', 'Maximum edit distance = 12',
'Maximum edit distance = 13', 'Maximum edit distance = 14',
'Maximum edit distance = 15', 'Maximum edit distance = 16',
'Maximum edit distance = 17', 'Maximum AS-path length = 7',
'Maximum AS-path length = 8', 'Maximum AS-path length = 9',
'Maximum AS-path length = 10', 'Maximum AS-path length = 11',
'Maximum AS-path length = 12', 'Maximum AS-path length = 13',
18:47


review2_(1).ipynb - Colaboratory
colab.research.google.com/drive/1iWsUSV4b3mqGEmYFRu_6FHxYdi_2KrMm
review2_(1).ipynb
File Edit View Insert Runtime Tools Help Last saved at 6:45 PM
Comment
Share
+ Code + Text
Connect
Editing
'Maximum AS-path length = 12', 'Maximum AS-path length = 13',
'Maximum AS-path length = 14', 'Maximum AS-path length = 15',
'Number of Interior Gateway Protocol (IGP) packets',
'Number of Exterior Gateway Protocol (EGP) packets',
'Number of incomplete packets', 'Packet size (B)', 'Label'],
dtype='object')
print(data.groupby('Label').size())
Label
-1.0 28120
1.0 4965
dtype: int64
data['Label'] = data['Label'].apply(lambda x: 0 if x == -1 else 1)
data['Label']
0 0
1 0
2 0
3 0
4 0
33081 0
33082 0
33083 0
33084 0
33085 1
Name: Label, Length: 33086, dtype: int64
data.shape


review2_(1).ipynb - Colaboratory
colab.research.google.com/drive/1iWsUSV4b3mqGEmYFRu_6FHxYdi_2KrMm
review2_(1).ipynb
File Edit View Insert Runtime Tools Help Last saved at 6:45 PM
Comment
Share
+ Code + Text
Connect
Editing
33084 0
33085 1
Name: Label, Length: 33086, dtype: int64
data.shape
(33086, 42)
data.dtypes
hour+minute object
hour float64
minute float64
second float64
Number of announcements float64
Number of withdrawals float64
Number of announced NLRI prefixes float64
Number of withdrawn NLRI prefixes float64
Average AS-path length float64
Maximum AS-path length float64
Average unique AS-path length float64
Number of duplicate announcements float64
Number of duplicate withdrawals float64
Number of implicit withdrawals float64
Average edit distance float64
Maximum edit distance float64
Inter-arrival time float64
Maximum edit distance = 7 float64
Maximum edit distance = 8 float64
Maximum edit distance = 9 float64
Maximum edit distance = 10 float64
Maximum edit distance = 11 float64
Maximum edit distance = 12 float64
Maximum edit distance = 13 float64

review2_(1).ipynb - Colaboratory

colab.research.google.com/drive/1iWsUSV4b3mqGEmYFRu_6FHxYdi_2KrMm

review2_(1).ipynb

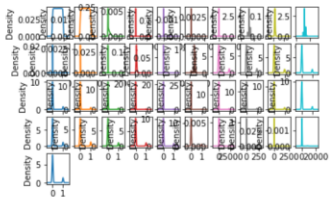
File Edit View Insert Runtime Tools Help Last saved at 6:45 PM

Comment Share

+ Code + Text

Connect Editing

```
[ ] data.plot(kind='density', subplots=True, layout=(5,10), sharex=False, legend=False, fontsize=10)
    plt.show()
```



```
from matplotlib import cm as cm

fig = plt.figure()
ax1 = fig.add_subplot(111)
cmap = cm.get_cmap('jet', 30)
cax = ax1.imshow(data.corr(), interpolation="none", cmap=cmap)
ax1.grid(True)
plt.title('Attributes Correlation')
fig.colorbar(cax, ticks=[.75,.8,.85,.90,.95,1])
plt.show()
```

Type here to search

ENG 18:50 23-10-2020

review2_(1).ipynb - Colaboratory

colab.research.google.com/drive/1iWsUSV4b3mqGEmYFRu_6FHxYdi_2KrMm

review2_(1).ipynb

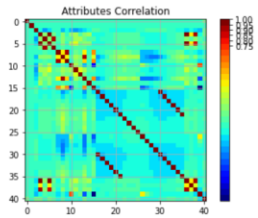
File Edit View Insert Runtime Tools Help Last saved at 6:45 PM

Comment Share

+ Code + Text

Connect Editing

```
[ ] fig.colorbar(cax, ticks=[.75,.8,.85,.90,.95,1])
    plt.show()
```



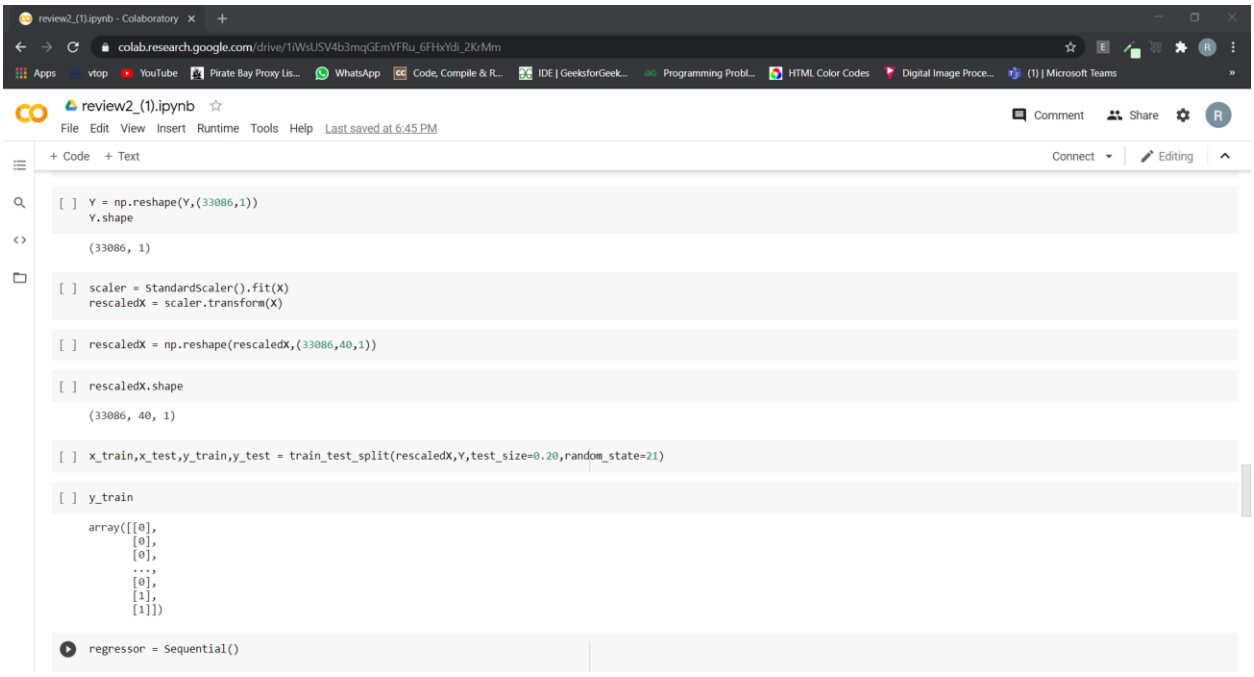

```
[ ] Y = data['Label'].values
    X = data.drop('Label', axis=1).values
```

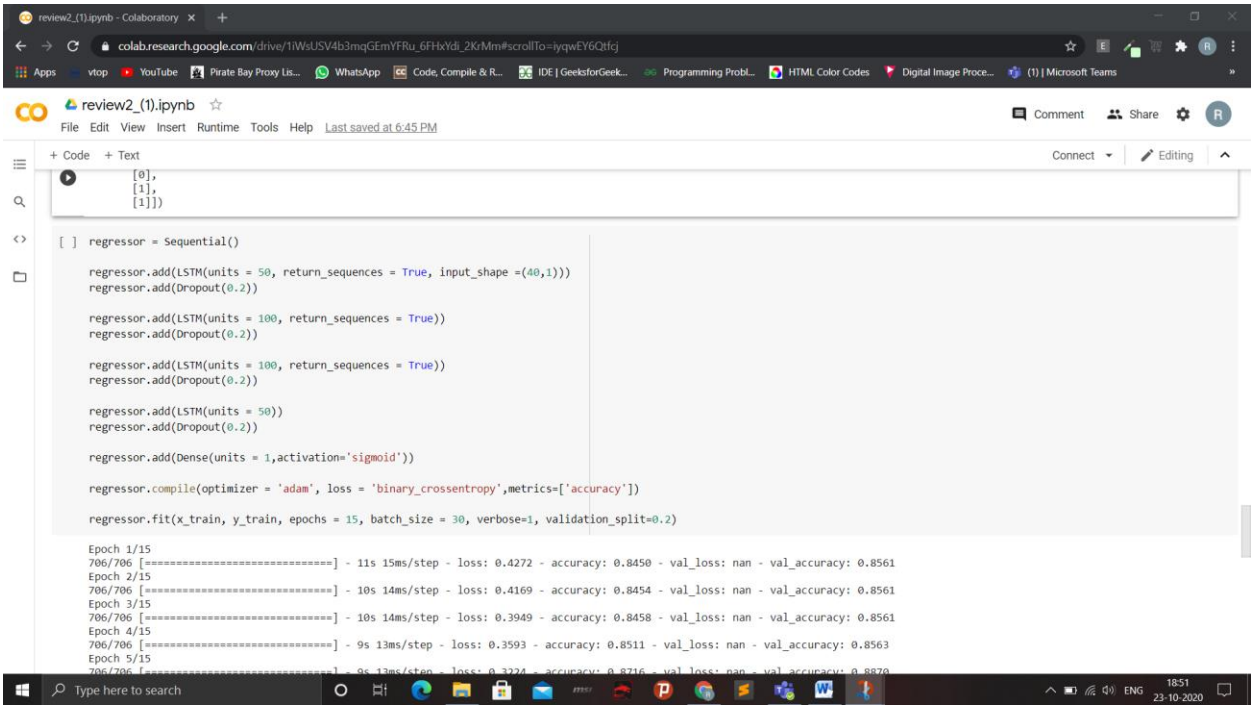
```
[ ] Y = np.reshape(Y,(33086,1))
    Y.shape
```

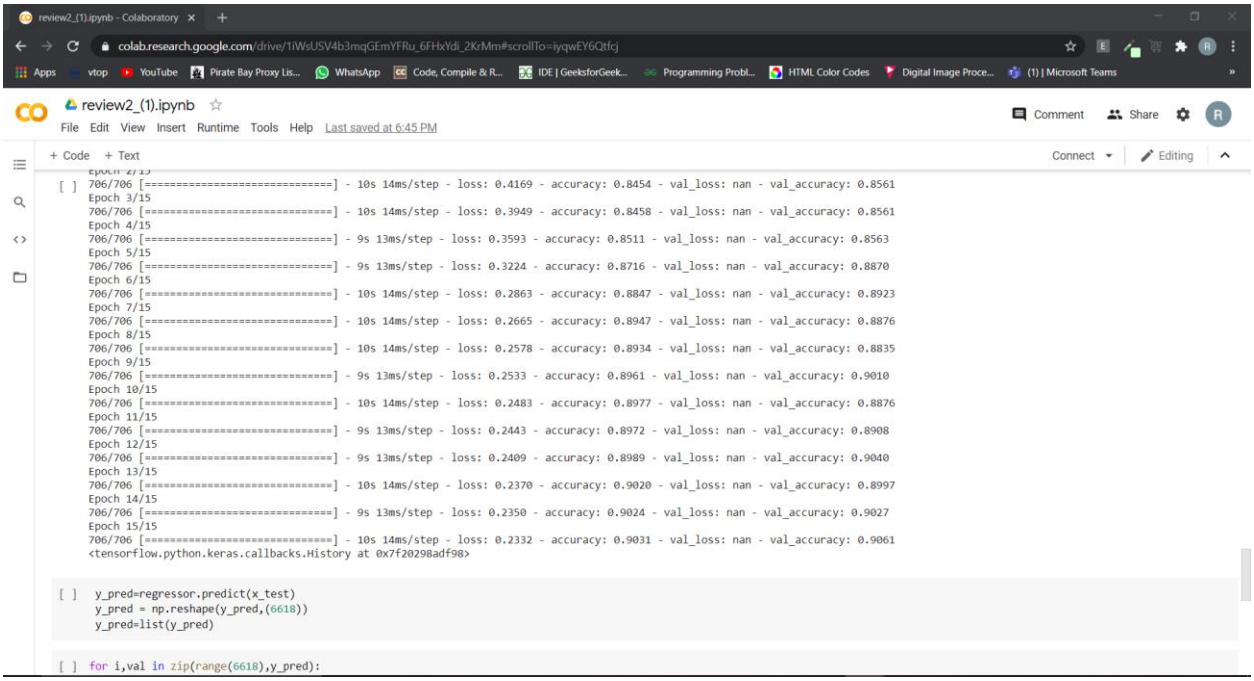
```
(33086, 1)
```

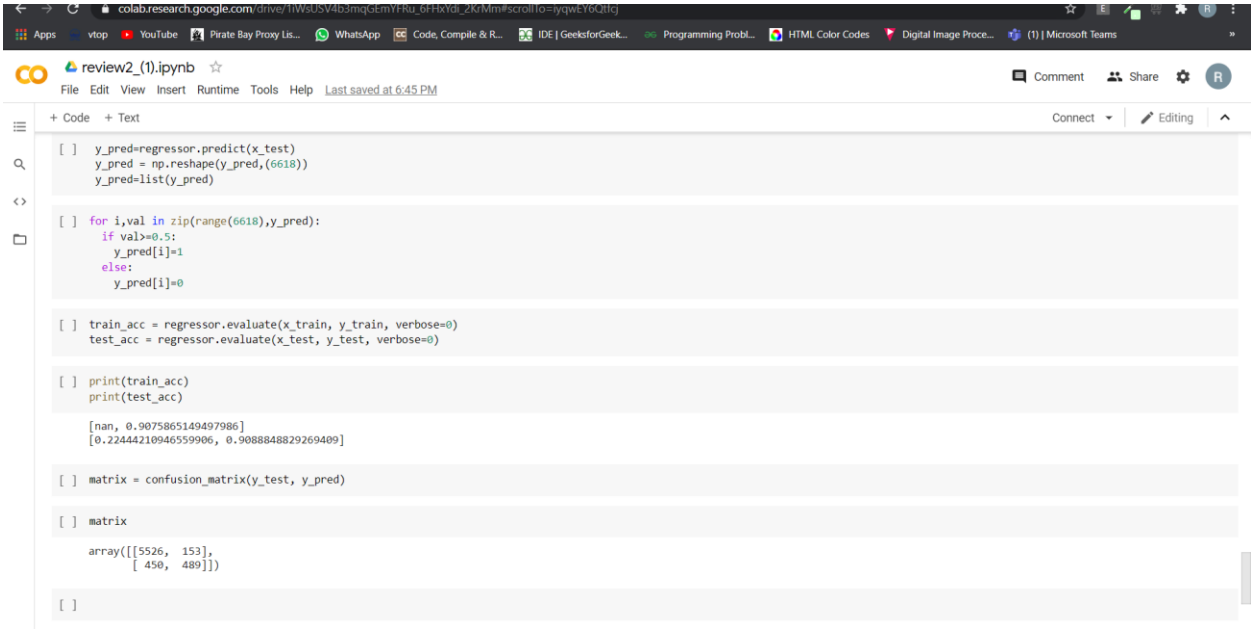
```
[ ] scaler = StandardScaler().fit(X)
    rescaledX = scaler.transform(X)
```

```
[ ] rescaledX = np.reshape(rescaledX,(33086,40,1))
```


review2_(1).ipynb - Colaboratory
colab.research.google.com/drive/1iWsUSV4b3mqGEmYFRu_6FHxYdi_2KrMm
review2_(1).ipynb
File Edit View Insert Runtime Tools Help Last saved at 6:45 PM
Comment
Share
+ Code + Text
Connect
Editing
[ ] Y = np.reshape(Y,(33086,1))
Y.shape
(33086, 1)
[ ] scaler = StandardScaler().fit(X)
rescaledX = scaler.transform(X)
[ ] rescaledX = np.reshape(rescaledX,(33086,40,1))
[ ] rescaledX.shape
(33086, 40, 1)
[ ] x_train,x_test,y_train,y_test = train_test_split(rescaledX,Y,test_size=0.20,random_state=21)
[ ] y_train
array([[0],
[0],
[0],
...,
[0],
[1],
[1]])
regressor = Sequential()


review2_(1).ipynb - Colaboratory
colab.research.google.com/drive/1iWsUSV4b3mqGEmYFRu_6FHxYdi_2KrMm#scrollTo=iyqwEY6Qtfcj
review2_(1).ipynb
File Edit View Insert Runtime Tools Help Last saved at 6:45 PM
Comment
Share
+ Code + Text
Connect
Editing
[0],
[1],
[1]])
[ ] regressor = Sequential()
regressor.add(LSTM(units = 50, return_sequences = True, input_shape =(40,1)))
regressor.add(Dropout(0.2))
regressor.add(LSTM(units = 100, return_sequences = True))
regressor.add(Dropout(0.2))
regressor.add(LSTM(units = 100, return_sequences = True))
regressor.add(Dropout(0.2))
regressor.add(LSTM(units = 50))
regressor.add(Dropout(0.2))
regressor.add(Dense(units = 1,activation='sigmoid'))
regressor.compile(optimizer = 'adam', loss = 'binary_crossentropy',metrics=['accuracy'])
regressor.fit(x_train, y_train, epochs = 15, batch_size = 30, verbose=1, validation_split=0.2)
Epoch 1/15
706/706 [==========] - 11s 15ms/step - loss: 0.4272 - accuracy: 0.8450 - val_loss: nan - val_accuracy: 0.8561
Epoch 2/15
706/706 [==========] - 10s 14ms/step - loss: 0.4169 - accuracy: 0.8454 - val_loss: nan - val_accuracy: 0.8561
Epoch 3/15
706/706 [==========] - 10s 14ms/step - loss: 0.3949 - accuracy: 0.8458 - val_loss: nan - val_accuracy: 0.8561
Epoch 4/15
706/706 [==========] - 9s 13ms/step - loss: 0.3593 - accuracy: 0.8511 - val_loss: nan - val_accuracy: 0.8563
Epoch 5/15
Type here to search
ENG
18:51
23-10-2020


review2_(1).ipynb - Colaboratory
colab.research.google.com/drive/1iWsUSV4b3mqGEmYFRu_6FHxYdi_2KrMm#scrollTo=iyqwEY6Qtfcj
review2_(1).ipynb
File Edit View Insert Runtime Tools Help Last saved at 6:45 PM
Comment
Share
+ Code + Text
Connect
Editing
[ ] 706/706 [==========] - 10s 14ms/step - loss: 0.4169 - accuracy: 0.8454 - val_loss: nan - val_accuracy: 0.8561
Epoch 3/15
706/706 [==========] - 10s 14ms/step - loss: 0.3949 - accuracy: 0.8458 - val_loss: nan - val_accuracy: 0.8561
Epoch 4/15
706/706 [==========] - 9s 13ms/step - loss: 0.3593 - accuracy: 0.8511 - val_loss: nan - val_accuracy: 0.8563
Epoch 5/15
706/706 [==========] - 9s 13ms/step - loss: 0.3224 - accuracy: 0.8716 - val_loss: nan - val_accuracy: 0.8870
Epoch 6/15
706/706 [==========] - 10s 14ms/step - loss: 0.2863 - accuracy: 0.8847 - val_loss: nan - val_accuracy: 0.8923
Epoch 7/15
706/706 [==========] - 10s 14ms/step - loss: 0.2665 - accuracy: 0.8947 - val_loss: nan - val_accuracy: 0.8876
Epoch 8/15
706/706 [==========] - 10s 14ms/step - loss: 0.2578 - accuracy: 0.8934 - val_loss: nan - val_accuracy: 0.8835
Epoch 9/15
706/706 [==========] - 9s 13ms/step - loss: 0.2533 - accuracy: 0.8961 - val_loss: nan - val_accuracy: 0.9010
Epoch 10/15
706/706 [==========] - 10s 14ms/step - loss: 0.2483 - accuracy: 0.8977 - val_loss: nan - val_accuracy: 0.8876
Epoch 11/15
706/706 [==========] - 9s 13ms/step - loss: 0.2443 - accuracy: 0.8972 - val_loss: nan - val_accuracy: 0.8908
Epoch 12/15
706/706 [==========] - 9s 13ms/step - loss: 0.2409 - accuracy: 0.8989 - val_loss: nan - val_accuracy: 0.9040
Epoch 13/15
706/706 [==========] - 10s 14ms/step - loss: 0.2370 - accuracy: 0.9020 - val_loss: nan - val_accuracy: 0.8997
Epoch 14/15
706/706 [==========] - 9s 13ms/step - loss: 0.2350 - accuracy: 0.9024 - val_loss: nan - val_accuracy: 0.9027
Epoch 15/15
706/706 [==========] - 10s 14ms/step - loss: 0.2332 - accuracy: 0.9031 - val_loss: nan - val_accuracy: 0.9061
<tensorflow.python.keras.callbacks.History at 0x7f20298adf98>
[ ] y_pred=regressor.predict(x_test)
y_pred = np.reshape(y_pred,(6618))
y_pred=list(y_pred)
[ ] for i,val in zip(range(6618),y_pred):


review2_(1).ipynb
File Edit View Insert Runtime Tools Help Last saved at 6:45 PM
Comment
Share
+ Code + Text
Connect
Editing
[ ] y_pred=regressor.predict(x_test)
y_pred = np.reshape(y_pred,(6618))
y_pred=list(y_pred)
[ ] for i,val in zip(range(6618),y_pred):
if val>=0.5:
y_pred[i]=1
else:
y_pred[i]=0
[ ] train_acc = regressor.evaluate(x_train, y_train, verbose=0)
test_acc = regressor.evaluate(x_test, y_test, verbose=0)
[ ] print(train_acc)
print(test_acc)
[nan, 0.9075865149497986]
[0.2244421094655990 6, 0.9088848829269409]
[ ] matrix = confusion_matrix(y_test, y_pred)
[ ] matrix
array([[5526, 153],
[450, 489]])
[ ]