In [1]:

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split,cross_val_score
from sklearn.feature_selection import f_classif
from sklearn.preprocessing import LabelEncoder, StandardScaler, MinMaxScaler
from sklearn import metrics
from scipy import stats
import statsmodels.api as sm
import math
import re
import seaborn as sns
from statsmodels.stats.outliers_influence import variance_inflation_factor
import sys
import pandas.core.algorithms as algos
pd.pandas.set_option('display.max_columns',None)
```

In [7]:

```python
train=pd.read_csv("train.csv")
train
```

Out[7]:

	id	title	author	text	label
0	0	House Dem Aide: We Didn’t Even See Comey’s Let...	Darrell Lucus	House Dem Aide: We Didn’t Even See Comey’s Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1
...	...	...	...	...	...
20795	20795	Rapper T.I.: Trump a ’Poster Child For White S...	Jerome Hudson	Rapper T. I. unloaded on black celebrities who...	0
20796	20796	N.F.L. Playoffs: Schedule, Matchups and Odds -...	Benjamin Hoffman	When the Green Bay Packers lost to the Washing...	0
20797	20797	Macy’s Is Said to Receive Takeover Approach by...	Michael J. de la Merced and Rachel Abrams	The Macy’s of today grew from the union of sev...	0
20798	20798	NATO, Russia To Hold Parallel Exercises In Bal...	Alex Ansary	NATO, Russia To Hold Parallel Exercises In Bal...	1
20799	20799	What Keeps the F-35 Alive	David Swanson	David Swanson is an author, activist, journa...	1

20800 rows × 5 columns

In [8]:

```
train.shape
```

Out[8]:

```
(20800, 5)
```

In [10]:

```
features=[i for i  in train.columns if i not in ['label']]
```

In [11]:

```
features
```

Out[11]:

```
['id', 'title', 'author', 'text']
```

In [12]:

```
train.shape
```

Out[12]:

```
(20800, 5)
```

In [13]:

```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20800 entries, 0 to 20799
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   id      20800 non-null  int64 
 1   title   20242 non-null  object
 2   author  18843 non-null  object
 3   text    20761 non-null  object
 4   label   20800 non-null  int64 
dtypes: int64(2), object(3)
memory usage: 812.6+ KB
```

Handling Missing Values

In [14]:

```
train.fillna('unavailable',inplace=True)
```

In [16]:

```
test=pd.read_csv("test.csv")
test
```

Out[16]:

	id	title	author	text
0	20800	Specter of Trump Loosens Tongues, if Not Purse...	David Streitfeld	PALO ALTO, Calif. — After years of scorning...
1	20801	Russian warships ready to strike terrorists ne...	NaN	Russian warships ready to strike terrorists ne...
2	20802	#NoDAPL: Native American Leaders Vow to Stay A...	Common Dreams	Videos #NoDAPL: Native American Leaders Vow to...
3	20803	Tim Tebow Will Attempt Another Comeback, This ...	Daniel Victor	If at first you don’t succeed, try a different...

	id	title	author	text
4	20804	Keiser Report: Meme Wars (E995)	Truth Broadcast Network	42 mins ago 1 Views 0 Comments 0 Likes 'For th...
...	...	...	...	...
5195	25995	The Bangladeshi Traffic Jam That Never Ends - ...	Jody Rosen	Of all the dysfunctions that plague the world’...
5196	25996	John Kasich Signs One Abortion Bill in Ohio bu...	Sheryl Gay Stolberg	WASHINGTON — Gov. John Kasich of Ohio on Tu...
5197	25997	California Today: What, Exactly, Is in Your Su...	Mike McPhate	Good morning. (Want to get California Today by...
5198	25998	300 US Marines To Be Deployed To Russian Borde...	NaN	« Previous - Next » 300 US Marines To Be Deplo...
5199	25999	Awkward Sex, Onscreen and Off - The New York T...	Teddy Wayne	Perhaps you’ve seen the new TV series whose pi...

5200 rows × 4 columns

In [19]:

```
test.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5200 entries, 0 to 5199
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   id      5200 non-null   int64 
 1   title   5078 non-null   object
 2   author  4697 non-null   object
 3   text    5193 non-null   object
dtypes: int64(1), object(3)
memory usage: 162.6+ KB
```

In [21]:

```
test.fillna('unavailable',inplace=True)
```

In [22]:

```
df=pd.concat([train,test],axis='rows')
```

In [23]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 26000 entries, 0 to 5199
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   id      26000 non-null  int64  
 1   title   26000 non-null  object 
 2   author  26000 non-null  object 
 3   text    26000 non-null  object 
 4   label   20800 non-null  float64
dtypes: float64(1), int64(1), object(3)
memory usage: 1.2+ MB
```

In [24]:

```
df.shape
```

Out[24]:

```
(26000, 5)
```

In [25]:

```
df['comb']=df['author']+"_"+df['title']#+df['text']      # Combined all the features
```

clean text data

In [26]:

```
import nltk
nltk.download('stopwords')
nltk.download('wordnet')
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from nltk.stem import WordNetLemmatizer
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\ARVIND\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\stopwords.zip.
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\ARVIND\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\wordnet.zip.
```

In [27]:

```
import string
import re
string.punctuation
```

Out[27]:

```
'!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
```

In [28]:

```
wordnet=WordNetLemmatizer()
stemmer=PorterStemmer()
def clean(text):
  # text="".join([char for char in text if char not in string.punctuation])
  text="".join([re.sub('[^a-zA-Z]',' ',char) for char in text ])
  text=text.lower()
  text=text.split()
  text=[stemmer.stem(word) for word in text if word not in set(stopwords.words("english
  text=" ".join(text)
  return text
```

In [29]:

```
df['comb']=df['comb'].apply(clean)
```

In [30]:

```
df.head()
```

Out[30]:

	id	title	author	text	label	comb
0	0	House Dem Aide: We Didn’t Even See Comey’s Let...	Darrell Lucus	House Dem Aide: We Didn’t Even See Comey’s Let...	1.0	darrel lucu hous dem aid even see comey letter...

	id	title	author	text	label	comb
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0.0	daniel j flynn flynn hillari clinton big woman...
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1.0	consortiumnew com truth might get fire
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1.0	jessica purkiss civilian kill singl us airstri...
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1.0	howard portnoy iranian woman jail fiction unpu...

Word Embedding

In [32]:

```
!pip install tensorflow
```

```
Requirement already satisfied: tensorflow in c:\users\arvind\anaconda3\lib\site-packages
(2.8.0)
Requirement already satisfied: keras<2.9,>=2.8.0rc0 in c:\users\arvind\anaconda3\lib\sit
e-packages (from tensorflow) (2.8.0)
Requirement already satisfied: flatbuffers>=1.12 in c:\users\arvind\anaconda3\lib\site-p
ackages (from tensorflow) (2.0)
Requirement already satisfied: numpy>=1.20 in c:\users\arvind\anaconda3\lib\site-package
s (from tensorflow) (1.20.3)
Requirement already satisfied: keras-preprocessing>=1.1.1 in c:\users\arvind\anaconda3\l
ib\site-packages (from tensorflow) (1.1.2)
Requirement already satisfied: wrapt>=1.11.0 in c:\users\arvind\anaconda3\lib\site-packa
ges (from tensorflow) (1.12.1)
Requirement already satisfied: grpcio<2.0,>=1.24.3 in c:\users\arvind\anaconda3\lib\site
-packages (from tensorflow) (1.46.0)
Requirement already satisfied: astunparse>=1.6.0 in c:\users\arvind\anaconda3\lib\site-p
ackages (from tensorflow) (1.6.3)
Requirement already satisfied: opt-einsum>=2.3.2 in c:\users\arvind\anaconda3\lib\site-p
ackages (from tensorflow) (3.3.0)
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in c:\users\arvind\a
naconda3\lib\site-packages (from tensorflow) (0.25.0)
Requirement already satisfied: gast>=0.2.1 in c:\users\arvind\anaconda3\lib\site-package
s (from tensorflow) (0.5.3)
Requirement already satisfied: protobuf>=3.9.2 in c:\users\arvind\anaconda3\lib\site-pac
kages (from tensorflow) (3.20.1)
Requirement already satisfied: h5py>=2.9.0 in c:\users\arvind\anaconda3\lib\site-package
s (from tensorflow) (3.2.1)
Requirement already satisfied: six>=1.12.0 in c:\users\arvind\anaconda3\lib\site-package
s (from tensorflow) (1.16.0)
Requirement already satisfied: absl-py>=0.4.0 in c:\users\arvind\anaconda3\lib\site-pack
ages (from tensorflow) (1.0.0)
Requirement already satisfied: setuptools in c:\users\arvind\anaconda3\lib\site-packages
(from tensorflow) (58.0.4)
Requirement already satisfied: libclang>=9.0.1 in c:\users\arvind\anaconda3\lib\site-pac
kages (from tensorflow) (14.0.1)
Requirement already satisfied: typing-extensions>=3.6.6 in c:\users\arvind\anaconda3\lib
```

```
\site-packages (from tensorflow) (3.10.0.2)
Requirement already satisfied: tf-estimator-nightly==2.8.0.dev2021122109 in c:\users\arv
ind\anaconda3\lib\site-packages (from tensorflow) (2.8.0.dev2021122109)
Requirement already satisfied: google-pasta>=0.1.1 in c:\users\arvind\anaconda3\lib\site
-packages (from tensorflow) (0.2.0)
Requirement already satisfied: termcolor>=1.1.0 in c:\users\arvind\anaconda3\lib\site-pa
ckages (from tensorflow) (1.1.0)
Requirement already satisfied: tensorboard<2.9,>=2.8 in c:\users\arvind\anaconda3\lib\si
te-packages (from tensorflow) (2.8.0)
Requirement already satisfied: wheel<1.0,>=0.23.0 in c:\users\arvind\anaconda3\lib\site-
packages (from astunparse>=1.6.0->tensorflow) (0.37.0)
Requirement already satisfied: tensorboard-data-server<0.7.0,>=0.6.0 in c:\users\arvind
\anaconda3\lib\site-packages (from tensorboard<2.9,>=2.8->tensorflow) (0.6.1)
Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in c:\users\arvind\anaconda
3\lib\site-packages (from tensorboard<2.9,>=2.8->tensorflow) (1.8.1)
Requirement already satisfied: google-auth-oauthlib<0.5,>=0.4.1 in c:\users\arvind\anaco
nda3\lib\site-packages (from tensorboard<2.9,>=2.8->tensorflow) (0.4.6)
Requirement already satisfied: requests<3,>=2.21.0 in c:\users\arvind\anaconda3\lib\site
-packages (from tensorboard<2.9,>=2.8->tensorflow) (2.26.0)
Requirement already satisfied: google-auth<3,>=1.6.3 in c:\users\arvind\anaconda3\lib\si
te-packages (from tensorboard<2.9,>=2.8->tensorflow) (2.6.6)
Requirement already satisfied: werkzeug>=0.11.15 in c:\users\arvind\anaconda3\lib\site-p
ackages (from tensorboard<2.9,>=2.8->tensorflow) (2.0.2)
Requirement already satisfied: markdown>=2.6.8 in c:\users\arvind\anaconda3\lib\site-pac
kages (from tensorboard<2.9,>=2.8->tensorflow) (3.3.7)
Requirement already satisfied: pyasn1-modules>=0.2.1 in c:\users\arvind\anaconda3\lib\si
te-packages (from google-auth<3,>=1.6.3->tensorboard<2.9,>=2.8->tensorflow) (0.2.8)
Requirement already satisfied: rsa<5,>=3.1.4 in c:\users\arvind\anaconda3\lib\site-packa
ges (from google-auth<3,>=1.6.3->tensorboard<2.9,>=2.8->tensorflow) (4.8)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in c:\users\arvind\anaconda3\lib\s
ite-packages (from google-auth<3,>=1.6.3->tensorboard<2.9,>=2.8->tensorflow) (5.0.0)
Requirement already satisfied: requests-oauthlib>=0.7.0 in c:\users\arvind\anaconda3\lib
\site-packages (from google-auth-oauthlib<0.5,>=0.4.1->tensorboard<2.9,>=2.8->tensorflo
w) (1.3.1)
Requirement already satisfied: importlib-metadata>=4.4 in c:\users\arvind\anaconda3\lib
\site-packages (from markdown>=2.6.8->tensorboard<2.9,>=2.8->tensorflow) (4.8.1)
Requirement already satisfied: zipp>=0.5 in c:\users\arvind\anaconda3\lib\site-packages
(from importlib-metadata>=4.4->markdown>=2.6.8->tensorboard<2.9,>=2.8->tensorflow) (3.6.
0)
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in c:\users\arvind\anaconda3\lib\sit
e-packages (from pyasn1-modules>=0.2.1->google-auth<3,>=1.6.3->tensorboard<2.9,>=2.8->te
nsorflow) (0.4.8)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\arvind\anaconda3\lib\site-
packages (from requests<3,>=2.21.0->tensorboard<2.9,>=2.8->tensorflow) (2021.10.8)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\arvind\anaconda3\lib\si
te-packages (from requests<3,>=2.21.0->tensorboard<2.9,>=2.8->tensorflow) (1.26.7)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\arvind\anaconda3\li
b\site-packages (from requests<3,>=2.21.0->tensorboard<2.9,>=2.8->tensorflow) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\arvind\anaconda3\lib\site-packag
es (from requests<3,>=2.21.0->tensorboard<2.9,>=2.8->tensorflow) (3.2)
Requirement already satisfied: oauthlib>=3.0.0 in c:\users\arvind\anaconda3\lib\site-pac
kages (from requests-oauthlib>=0.7.0->google-auth-oauthlib<0.5,>=0.4.1->tensorboard<2.9,
>=2.8->tensorflow) (3.2.0)
```

In [38]:

```python
import tensorflow
from tensorflow.keras.layers import Embedding
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.models import Sequential
from tensorflow.keras.preprocessing.text import one_hot
from tensorflow.keras.layers import LSTM
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Dropout
from tensorflow.keras.layers import Bidirectional
```

In [42]:

```
voc_size=10000   # Vocabulary size
```

In [43]:

```
title=df['comb']
title[0]
```

Out[43]:

```
0    darrel lucu hous dem aid even see comey letter...
0    david streitfeld specter trump loosen tongu pu...
Name: comb, dtype: object
```

In [44]:

```
one_hot_r=[one_hot(words, voc_size) for words in title]
```

In [45]:

```
one_hot_r[0:2]
```

Out[45]:

```
[[7492, 8955, 8100, 3991, 9456, 9084, 460, 4028, 7703, 4016, 1168, 2544],
 [2342, 642, 5001, 5001, 6501, 3665, 3331, 1258, 9731, 8672]]
```

In [46]:

```
sent_len=30
embedded_docs=pad_sequences(one_hot_r,padding='post',maxlen=sent_len)
```

In [47]:

```
embedded_docs[:10]
```

Out[47]:

```
array([[7492, 8955, 8100, 3991, 9456, 9084,  460, 4028, 7703, 4016, 1168,
        2544,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0],
       [2342,  642, 5001, 5001, 6501, 3665, 3331, 1258, 9731, 8672,    0,
           0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0],
       [6987, 4706, 5771, 1026, 9906, 5068,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0],
       [8547, 6207, 5925, 4520, 8815, 2973, 7191, 1642,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0],
       [9537, 8986, 5501, 1258, 2399,  641, 9605,  611, 1258, 5614,  309,
        4996,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0],
       [2342, 8018, 9435, 5967, 6473, 3411, 5496, 9712, 7545, 6246,   54,
        3124, 4908, 6202, 1453, 8794, 8672,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0],
       [9993, 5498, 5498, 1014, 8421, 4432, 4127,  321, 2131, 4204,   25,
        3940, 5705,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0],
       [1053,  642, 3964, 7147, 3347, 8582, 7767,  690, 7721, 3405, 9572,
        5711, 1570, 2754,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0],
       [9993, 6191, 1512, 5695, 4518, 9712, 4587, 9889, 8436,  918, 2340,
        5711, 1570, 2754,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0],
       [1840, 1855, 5612, 2708, 2539, 7818, 6110, 4993, 7269,  399, 9712,
        2521, 5711, 1570, 2754,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0]])
```

In [48]:

```
sgd=tensorflow.keras.optimizers.SGD(learning_rate=0.01, decay=1e-6,momentum=0.9, nester
rms = tensorflow.keras.optimizers.RMSprop()
nadam=tensorflow.keras.optimizers.Nadam(
    learning_rate=0.001, beta_1=0.9, beta_2=0.999, epsilon=1e-07, name="Nadam"
)
```

In [49]:

```
embedding_vector_features=50
model=Sequential()
model.add(Embedding(voc_size,embedding_vector_features,input_length=sent_len))
# model.add(Dropout(0.1))
model.add(Bidirectional(LSTM(100)))  # used Bidirectional LSTM
model.add(Dropout(0.1))
model.add(Dense(1,activation='sigmoid'))
model.compile(loss='binary_crossentropy',optimizer=nadam,metrics=['accuracy'])
print(model.summary())
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #   
=================================================================
 embedding (Embedding)       (None, 30, 50)            500000    
                                                                 
 bidirectional (Bidirectiona  (None, 200)              120800    
 l)                                                              
                                                                 
 dropout (Dropout)           (None, 200)               0         
                                                                 
 dense (Dense)               (None, 1)                 201       
                                                                 
=================================================================
Total params: 621,001
Trainable params: 621,001
Non-trainable params: 0
_________________________________________________________________
None
```

In [50]:

```
len(embedded_docs)
```

Out[50]:

```
26000
```

In [52]:

```
X=embedded_docs[:train.shape[0]]
y=df['label'][:train.shape[0]]
x_test=embedded_docs[train.shape[0]:]
```

Split data for training and testing

In [53]:

```
X_train, x_valid, y_train, y_valid = train_test_split(X, y, test_size=0.2, random_state
```

In [55]:

```
model.fit(X_train,y_train,validation_data=(x_valid,y_valid),epochs=25,batch_size=124)
```

```
Epoch 1/25
135/135 [==============================] - 4s 29ms/step - loss: 8.6207e-06 - accuracy:
1.0000 - val_loss: 0.0979 - val_accuracy: 0.9875
Epoch 2/25
135/135 [==============================] - 4s 29ms/step - loss: 7.6496e-06 - accuracy:
1.0000 - val_loss: 0.0992 - val_accuracy: 0.9873
Epoch 3/25
135/135 [==============================] - 4s 29ms/step - loss: 7.0172e-06 - accuracy:
1.0000 - val_loss: 0.1004 - val_accuracy: 0.9870
Epoch 4/25
135/135 [==============================] - 4s 29ms/step - loss: 6.4416e-06 - accuracy:
1.0000 - val_loss: 0.1015 - val_accuracy: 0.9870
Epoch 5/25
135/135 [==============================] - 4s 29ms/step - loss: 5.7535e-06 - accuracy:
1.0000 - val_loss: 0.1027 - val_accuracy: 0.9870
Epoch 6/25
135/135 [==============================] - 4s 29ms/step - loss: 5.2410e-06 - accuracy:
1.0000 - val_loss: 0.1038 - val_accuracy: 0.9870
Epoch 7/25
135/135 [==============================] - 4s 29ms/step - loss: 4.9687e-06 - accuracy:
1.0000 - val_loss: 0.1048 - val_accuracy: 0.9870
Epoch 8/25
135/135 [==============================] - 4s 29ms/step - loss: 4.5191e-06 - accuracy:
1.0000 - val_loss: 0.1059 - val_accuracy: 0.9870
Epoch 9/25
135/135 [==============================] - 4s 29ms/step - loss: 4.0913e-06 - accuracy:
1.0000 - val_loss: 0.1070 - val_accuracy: 0.9870
Epoch 10/25
135/135 [==============================] - 4s 29ms/step - loss: 3.6825e-06 - accuracy:
1.0000 - val_loss: 0.1080 - val_accuracy: 0.9868
Epoch 11/25
135/135 [==============================] - 4s 29ms/step - loss: 3.6019e-06 - accuracy:
1.0000 - val_loss: 0.1091 - val_accuracy: 0.9868
Epoch 12/25
135/135 [==============================] - 4s 29ms/step - loss: 3.2232e-06 - accuracy:
1.0000 - val_loss: 0.1101 - val_accuracy: 0.9868
Epoch 13/25
135/135 [==============================] - 4s 29ms/step - loss: 3.0232e-06 - accuracy:
1.0000 - val_loss: 0.1111 - val_accuracy: 0.9868
Epoch 14/25
135/135 [==============================] - 4s 29ms/step - loss: 2.7582e-06 - accuracy:
1.0000 - val_loss: 0.1121 - val_accuracy: 0.9865
Epoch 15/25
135/135 [==============================] - 4s 29ms/step - loss: 2.5289e-06 - accuracy:
1.0000 - val_loss: 0.1132 - val_accuracy: 0.9865
Epoch 16/25
135/135 [==============================] - 4s 31ms/step - loss: 2.3830e-06 - accuracy:
1.0000 - val_loss: 0.1142 - val_accuracy: 0.9865
Epoch 17/25
135/135 [==============================] - 4s 31ms/step - loss: 2.1690e-06 - accuracy:
1.0000 - val_loss: 0.1151 - val_accuracy: 0.9865
Epoch 18/25
135/135 [==============================] - 4s 31ms/step - loss: 2.0314e-06 - accuracy:
1.0000 - val_loss: 0.1161 - val_accuracy: 0.9865
Epoch 19/25
135/135 [==============================] - 4s 28ms/step - loss: 1.9298e-06 - accuracy:
1.0000 - val_loss: 0.1171 - val_accuracy: 0.9865
Epoch 20/25
```

```
135/135 [==============================] - 4s 29ms/step - loss: 1.7455e-06 - accuracy:
1.0000 - val_loss: 0.1180 - val_accuracy: 0.9865
Epoch 21/25
135/135 [==============================] - 4s 29ms/step - loss: 1.5677e-06 - accuracy:
1.0000 - val_loss: 0.1190 - val_accuracy: 0.9865
Epoch 22/25
135/135 [==============================] - 4s 28ms/step - loss: 1.5050e-06 - accuracy:
1.0000 - val_loss: 0.1199 - val_accuracy: 0.9865
Epoch 23/25
135/135 [==============================] - 4s 28ms/step - loss: 1.4239e-06 - accuracy:
1.0000 - val_loss: 0.1209 - val_accuracy: 0.9865
Epoch 24/25
135/135 [==============================] - 4s 28ms/step - loss: 1.3028e-06 - accuracy:
1.0000 - val_loss: 0.1219 - val_accuracy: 0.9868
Epoch 25/25
135/135 [==============================] - 4s 28ms/step - loss: 1.2220e-06 - accuracy:
1.0000 - val_loss: 0.1228 - val_accuracy: 0.9865
```

Out[55]:

```
<keras.callbacks.History at 0x23765189e80>
```

In [61]:

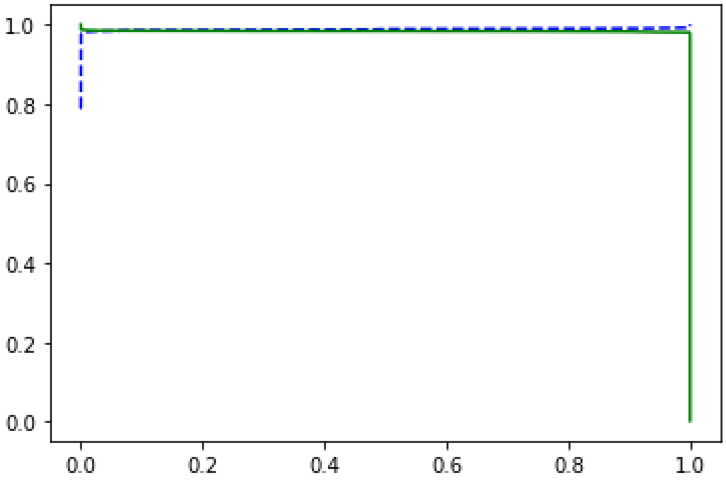
```
y_pred_prob=np.array(model.predict(X_train))[:,0]
y_pred_valid=np.array(model.predict(x_valid))[:,0]
y_pred=np.array(model.predict(x_test))[:,0]
```

In [62]:

```
from sklearn.metrics import precision_recall_curve
precisions, recalls, thresholds = precision_recall_curve(y_valid, y_pred_valid)
```

In [64]:

```
def plot_precision_recall_vs_threshold(precisions, recalls, thresholds):
  plt.plot(thresholds, precisions[:-1], "b--", label="Precision")
  plt.plot(thresholds, recalls[:-1], "g-", label="Recall")
[...]
plot_precision_recall_vs_threshold(precisions, recalls, thresholds)
plt.show()
```



Decision boundary to identify classes

In [65]:

```
def cutoff_youdens_j(fpr,tpr,thresholds):
  a = tpr-fpr
  b = sorted(zip(a,thresholds))
  return b[-1][1]
```

In [66]:

```
fpr, tpr, threshold = metrics.roc_curve(y_valid, y_pred_valid)
```

In [67]:

```
cutoff = cutoff_youdens_j(fpr, tpr, threshold)
```

In [68]:

```
A=[]
B=[]
C=[]
for i in range(len(threshold)):
  predicted = pd.DataFrame()
  predicted["label"] = y_pred_prob
  predicted["label"] = np.where(predicted["label"] > float(threshold[i]), 1, 0)
  pred_valid = pd.DataFrame()
  pred_valid["label"] = y_pred_valid
  pred_valid["label"] = np.where(pred_valid["label"] > float(threshold[i]), 1, 0)
  A.append(metrics.accuracy_score(y_train, predicted))
  B.append(metrics.accuracy_score(y_valid, pred_valid))
  C.append(threshold[i])
acc=pd.DataFrame(C,columns=['threshold'])
acc['train_acc']=A
acc['test_acc']=B
acc.sort_values(by='test_acc',ascending=False,inplace=True)
```

In [69]:

```
# cutoff=acc.iloc[0,0]
# cutoff
```

In [70]:

```
predicted = pd.DataFrame()
predicted["label"] = y_pred_prob
predicted["label"] = np.where(predicted["label"] > float(cutoff), 1, 0)
predicted

pred_valid = pd.DataFrame()
pred_valid["label"] = y_pred_valid
pred_valid["label"] = np.where(pred_valid["label"] > float(cutoff), 1, 0)
pred_valid

predictions = pd.DataFrame()
predictions['label'] = y_pred
predictions['label'] = np.where(predictions['label'] > float(cutoff), 1, 0)
predictions
```

Out[70]:

```
   label
0      0
1      1
2      1
3      0
4      1
```

	label
...	...
5195	0
5196	0
5197	0
5198	1
5199	0

5200 rows × 1 columns

In [71]:

```
from sklearn.metrics import roc_auc_score
```

Performance on Train Data

In [72]:

```
conf_matrix = metrics.confusion_matrix(y_train, predicted)
print(conf_matrix)
acc_train = metrics.accuracy_score(y_train, predicted)
print(acc_train)
precision_train = metrics.precision_score(y_train, predicted)
print(precision_train)
sensitivity_train = metrics.recall_score(y_train, predicted)
print(sensitivity_train)
specificity_train = conf_matrix[0,0] / (conf_matrix[0,0] + conf_matrix[0,1])
print(specificity_train)
roc_auc_score(y_train, predicted)
```

```
[[8255    0]
 [   0 8385]]
1.0
1.0
1.0
1.0
```

Out[72]:

```
1.0
```

Performance on Validation set Data

In [73]:

```
conf_matrix = metrics.confusion_matrix(y_valid, pred_valid)
print(conf_matrix)

acc_train = metrics.accuracy_score(y_valid, pred_valid)
print(acc_train)
precision_train = metrics.precision_score(y_valid, pred_valid)
print(precision_train)
sensitivity_train = metrics.recall_score(y_valid, pred_valid)
print(sensitivity_train)
specificity_train = conf_matrix[0,0] / (conf_matrix[0,0] + conf_matrix[0,1])
```

```
print(specificity_train)
roc_auc_score(y_valid, pred_valid)
```

```
[[2119   13]
 [  40 1988]]
0.9872596153846154
0.993503248375812
0.980276134122288
0.9939024390243902
```

Out[73]:

```
0.987089286573339
```

In [74]:

```
from sklearn.metrics import f1_score
print(f1_score(y_train, predicted,average='macro'))
f1_score(y_valid, pred_valid,average='macro')
```

```
1.0
```

Out[74]:

```
0.9872469689137349
```

In [76]:

```
submit=pd.read_csv('submit.csv')
```

In [77]:

```
submit
```

Out[77]:

	id	label
0	20800	0
1	20801	1
2	20802	0
3	20803	1
4	20804	1
...	...	...
5195	25995	0
5196	25996	1
5197	25997	0
5198	25998	1
5199	25999	0

5200 rows × 2 columns

In [78]:

```
submit['label']=predictions['label']
```

In [80]:

```
submit.to_csv('sub23',index=False)
```

In [ ]: