In [9]:
```
#Importing Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
df=pd.read_csv('Tweet_NFT.xlsx.csv')
df
```

Out[2]:

	id	tweet_text	tweet_created_at	tweet_intent
0	1212762	@crypto_brody @eCoLoGy1990 @MoonrunnersNFT @It...	2022-08-06T16:56:36.000Z	Community
1	1212763	Need Sick Character artâ “#art #artist #Artist...	2022-08-06T16:56:36.000Z	Giveaway
2	1212765	@The_Hulk_NFT @INagotchiNFT @Tesla @killabears...	2022-08-06T16:56:35.000Z	Appreciation
3	1212766	@CryptoBatzNFT @DarekBTW The first project in ...	2022-08-06T16:56:35.000Z	Community
4	1212767	@sashadysonn The first project in crypto with ...	2022-08-06T16:56:34.000Z	Community
...	...	...	...	...
127448	1278721	@Jason My company Universe is solving this, an...	2022-08-08T03:32:01.000Z	NaN
127449	1278723	@BosoTokyo Many will get rekt in this NFT	2022-08-08T03:32:00.000Z	NaN
127450	1278724	@leeneedham81 @KibaInuWorld @HODL_OFFICIAL @Ko...	2022-08-08T03:32:00.000Z	NaN
127451	1278725	@1kz3ro @UNLEASHED_NFT I love this job	2022-08-08T03:31:59.000Z	NaN
127452	1278729	@Shill_Ronin @UNLEASHED_NFT my proof https://t...	2022-08-08T03:31:58.000Z	NaN

127453 rows × 4 columns

In [3]:
```
df.shape
```

Out[3]:
```
(127453, 4)
```

In [4]:
```
#Dropping columns which are not required to identify
df.drop(['id','tweet_created_at'],axis=1,inplace=True)
df
```

Out[4]:

	tweet_text	tweet_intent
0	@crypto_brody @eCoLoGy1990 @MoonrunnersNFT @It...	Community
1	Need Sick Character artâ “#art #artist #Artist...	Giveaway
2	@The_Hulk_NFT @INagotchiNFT @Tesla @killabears...	Appreciation
3	@CryptoBatzNFT @DarekBTW The first project in ...	Community
4	@sashadysonn The first project in crypto with ...	Community
...	...	...
127448	@Jason My company Universe is solving this, an...	NaN
127449	@BosoTokyo Many will get rekt in this NFT	NaN
127450	@leeneedham81 @KibaInuWorld @HODL_OFFICIAL @Ko...	NaN
127451	@1kz3ro @UNLEASHED_NFT I love this job	NaN
127452	@Shill_Ronin @UNLEASHED_NFT my proof https://t...	NaN

127453 rows × 2 columns

In [7]:
```
df.tweet_text[61072]
```

Out[7]:
```
'@motoversegame is a Web 3.0 lifestyle app with built-in Game-Fi &amp; Social-F
i elements A new game concept of â€œDrive to Earnâ€œ users get tokens reward th
rough daily Mileage . #Airdrop #NFT #Motoverse #drivetoearn https://t.co/JqqqSb
fC0T' (https://t.co/JqqqSbfC0T')
```

In [11]:
```
#DATA PREPARING 1 :HTML decoding texted fields like &amp has been removed
from bs4 import BeautifulSoup
example1 = BeautifulSoup(df.tweet_text[61072], 'lxml')
print(example1.get_text())
```

```
@motoversegame is a Web 3.0 lifestyle app with built-in Game-Fi & Social-Fi ele
ments A new game concept of â€œDrive to Earnâ€œ users get tokens reward through
daily Mileage . #Airdrop #NFT #Motoverse #drivetoearn https://t.co/JqqqSbfC0T
 (https://t.co/JqqqSbfC0T)

C:\Users\Pruthvi Koli\anaconda3\lib\site-packages\bs4\__init__.py:435: MarkupRe
semblesLocatorWarning: The input looks more like a filename than markup. You ma
y want to open this file and pass the filehandle into Beautiful Soup.
  warnings.warn(
```

In [12]:
```
df.tweet_text[61073]
```

Out[12]:
```
'@emma_nft_1 These cutie 2D NFTs from @MicroPetsBSC is still on the TOP Trading
list of @LootexIO\n\nFor more infos, check our website\nand join our community
ðŸ\x8d»\nWebsite: https://t.co/jKLInSQJvI\nTG: (https://t.co/jKLInSQJvI\nTG:) h
ttps://t.co/Yy3ItxkCcr\n\n#SophiaEmily20April' (https://t.co/Yy3ItxkCcr\n\n#Sop
hiaEmily20April')
```

In [14]:
```
#DATA PREPARING 2: '@' mention  values has been removed
import re
re.sub(r'@[A-Za-z0-9]+','',df.tweet_text[61073])
```

Out[14]:
```
'_nft_1 These cutie 2D NFTs from  is still on the TOP Trading list of \n\nFor m
ore infos, check our website\nand join our community ðŸ\x8d»\nWebsite: https://
t.co/jKLInSQJvI\nTG: (https://t.co/jKLInSQJvI\nTG:) https://t.co/Yy3ItxkCcr\n\n
#SophiaEmily20April' (https://t.co/Yy3ItxkCcr\n\n#SophiaEmily20April')
```

In [17]:
```
df.tweet_text[12]
```

Out[17]:
```
'Just registered for the @HarakaiNFT raffle! Free mint and SUPER LOW SUPPLY! #F
reeMint via @PREMINT_NFT https://t.co/JjjZFPtsRO' (https://t.co/JjjZFPtsRO')
```

In [18]:
```
#DATA PREPARING 3: URL links has been removed
re.sub('https://[A-Za-z0-9./]+','',df.tweet_text[12])
```

Out[18]:
```
'Just registered for the @HarakaiNFT raffle! Free mint and SUPER LOW SUPPLY! #F
reeMint via @PREMINT_NFT '
```

In [23]:
```
df.tweet_text[25]
```

Out[23]:
```
"@nftbunkr To Celebrate New #NFT community. ðŸ¥³\nWe are giving away #NFT raffl
e to first 300 Members.ðŸ”¥\nDon't be late ðŸ'¥ðŸ¤—\nTo Enter: ðŸš€\nðŸ"-Follow
ðŸ‘‰ @italiaNft_art\nðŸ"-Join Telegram ðŸ‘‰https://t.co/xywDAzbKC4\nðŸ"-Rt this
Post \nðŸ‘‡\nhttps://t.co/QZfIrRiypu\nðŸ"- Visit: https://t.co/ln8bINigwv" (htt
ps://t.co/ln8bINigwv")
```

In [24]:
```
#DATA PREPARING 4 : HASHTAG # has been removed
re.sub("[^a-zA-Z]", " ",df.tweet_text[25])
```

Out[24]:
```
' nftbunkr To Celebrate New  NFT community       We are giving away  NFT raffle
to first     Members      Don t be late          To Enter           Follow
italiaNft art     Join Telegram     https   t co xywDAzbKC      Rt this Post
https   t co QZfIrRiypu      Visit  https   t co ln bINigwv'
```

In [28]:
```
df.tweet_text[5]
```

Out[28]:
```
'ðŸŽ‰ Just registered for the saphire on @PREMINT_NFT! https://t.co/lNIXaPFL4j'
(https://t.co/lNIXaPFL4j')
```

In [43]:
```
text_unicode = df.tweet_text[5]
# encoding the text to ASCII format
text_encode = text_unicode.encode(encoding="ascii", errors="ignore")
# decoding the text
text_decode = text_encode.decode()
# cleaning the text to remove extra whitespace
clean_text = " ".join([word for word in text_decode.split()])
print(clean_text)
```

```
Just registered for the saphire on @PREMINT_NFT! https://t.co/lNIXaPFL4j (http
s://t.co/lNIXaPFL4j)
```

In [76]:
```
#DATA CLEANING FUNCTION
from nltk.tokenize import WordPunctTokenizer
tok = WordPunctTokenizer()
pat1 = r'@[A-Za-z0-9]+'
pat2 = r'https://[A-Za-z0-9./]+'
combined_pat = r'|'.join((pat1, pat2))
def tweet_cleaner(text):
    soup = BeautifulSoup(text, 'lxml')
    souped = soup.get_text()
    stripped = re.sub(combined_pat, '', souped)
    try:
        clean = stripped.decode()
    except:
        clean = stripped
    letters_only = re.sub("[^a-zA-Z]", " ", clean)
    lower_case = letters_only.lower()
    # During the letters_only process two lines above, it has created unnecessay
    # I will tokenize and join together to remove unneccessary white spaces
    words = tok.tokenize(lower_case)
    return (" ".join(words)).strip()
testing = df.tweet_text[:100]
test_result = []
for t in testing:
    test_result.append(tweet_cleaner(t))
test_result
```

```
 'the first project in crypto with move to earn astrobirdz big things happeni
ng with this project with release of marketplace and nest feedtoearn birdzarm
y luna btc web gem nft nftcommunity',
 'just registered for the saphire on nft',
 'the bridged sold price s l rank m rank hr n r fp s l owner nfl gvp rfnh e s
sjgcjeyb hhgtyxst nxk pdr link nftnyc nftgiveaway nftcommunity nft',
 'project the best gamefi project on bsc multistage deflation tool pro multi
feature nft app on appstore and google play presale in august p is going to t
he moon whitelist tg project p gamefi',
 'feature it on we have a great community of nft artists and nft collectors',
 'project the best gamefi project on bsc multistage deflation tool pro multi
feature nft app on appstore and google play presale in august p is going to t
he moon whitelist tg project p gamefi',
 'nft depends',
 'promote it on community',
 'just registered for the raffle free mint and super low supply freemint via
nft',
 'definitely not gonna missed this more than what nft lfgggggg incubexwetrust
cubextakeover nftcommunity nfts',
 'sol project the best gamefi project on bsc multistage deflation tool pro mu
```

In [90]:
```
nums = [0,127453]
print ("Cleaning and parsing the tweets...\n")
clean_tweet_texts = []
for i in range(nums[0],nums[1]):
    if( (i+1)%10000 == 0 ):
        print ("Tweets %d of %d has been processed" % ( i+1, nums[1] ) )
    clean_tweet_texts.append(tweet_cleaner(df['tweet_text'][i]))
```

```
Cleaning and parsing the tweets...


C:\Users\Pruthvi Koli\anaconda3\lib\site-packages\bs4\__init__.py:435: MarkupRe
semblesLocatorWarning: The input looks more like a filename than markup. You ma
y want to open this file and pass the filehandle into Beautiful Soup.
  warnings.warn(

Tweets 10000 of 127453 has been processed
Tweets 20000 of 127453 has been processed
Tweets 30000 of 127453 has been processed
Tweets 40000 of 127453 has been processed
Tweets 50000 of 127453 has been processed
Tweets 60000 of 127453 has been processed
Tweets 70000 of 127453 has been processed
Tweets 80000 of 127453 has been processed
Tweets 90000 of 127453 has been processed
Tweets 100000 of 127453 has been processed
Tweets 110000 of 127453 has been processed
Tweets 120000 of 127453 has been processed
```

In [91]:
```
# CLEAN DATA HAS BEEN SAVED
clean_df = pd.DataFrame(clean_tweet_texts,columns=['tweet_text'])
clean_df['target'] = df.tweet_intent
clean_df.head(10)
```

Out[91]:

	tweet_text	target
0	brody nft bowie el krypto a new cryptocurrency...	Community
1	need sick character art art artist artists ani...	Giveaway
2	hulk nft great choice on a tesla good luck	Appreciation
3	the first project in crypto with move to earn ...	Community
4	the first project in crypto with move to earn ...	Community
5	just registered for the saphire on nft	Presale
6	the bridged sold price s l rank m rank hr n r ...	Giveaway
7	project the best gamefi project on bsc multist...	Whitelist
8	feature it on we have a great community of nft...	Community
9	project the best gamefi project on bsc multist...	Whitelist

In [92]:
```
clean_df.shape
```

Out[92]:
```
(127453, 2)
```

In [93]:
```
clean_df.isnull().sum()
```

Out[93]:
```
tweet_text        0
target        31089
dtype: int64
```

In [94]:
```
clean_df
```

Out[94]:

	tweet_text	target
0	brody nft bowie el krypto a new cryptocurrency...	Community
1	need sick character art art artist artists ani...	Giveaway
2	hulk nft great choice on a tesla good luck	Appreciation
3	the first project in crypto with move to earn ...	Community
4	the first project in crypto with move to earn ...	Community
...	...	...
127448	my company universe is solving this and we ve ...	NaN
127449	many will get rekt in this nft	NaN
127450	official classic of hodl hodl hands nft funds ...	NaN
127451	nft i love this job	NaN
127452	ronin nft my proof	NaN

127453 rows × 2 columns

In [95]:
```
df.tweet_intent[96364]
```

Out[95]:
```
nan
```

In [96]:
```
#SEPERATING NON NAN AND NAN VALUES FOR TRAINING THE MODEL
df_1 = clean_df.iloc[:96364,:]
df_2 = clean_df.iloc[96365:,:]
print("Shape of new dataframes - {} , {}".format(df_1.shape, df_2.shape))
```

```
Shape of new dataframes - (96364, 2) , (31088, 2)
```

In [98]:
```
df_1
```

Out[98]:

	tweet_text	target
0	brody nft bowie el krypto a new cryptocurrency...	Community
1	need sick character art art artist artists ani...	Giveaway
2	hulk nft great choice on a tesla good luck	Appreciation
3	the first project in crypto with move to earn ...	Community
4	the first project in crypto with move to earn ...	Community
...	...	...
96359	cryptogang x fe d a e e ea a d ac e c b fc tha...	Appreciation
96360	just registered to win with let s go freemint ...	Presale
96361	dyor on the nft collection dictators they re b...	Appreciation
96362	nft nftcommmunity deltaflare join me on discor...	Community
96363	nft finley eth nice	Appreciation

96364 rows × 2 columns

In [130]:
```
#CONVERTING CATEGORICAL DATA INTO BINARY BY LABEL ENCODING
from sklearn. preprocessing import LabelEncoder
le=LabelEncoder()
```

In [131]:
```
labels=le.fit_transform(df_1.target)
df_1['target']=labels
```

```
C:\Users\Pruthvi Koli\AppData\Local\Temp\ipykernel_7452\126959495.py:2: Setting
WithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/sta
ble/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pyd
ata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c
opy)
  df_1['target']=labels
```

In [132]:
```
df_1
```

Out[132]:

	tweet_text	target
0	brody nft bowie el krypto a new cryptocurrency...	1
1	need sick character art art artist artists ani...	3
2	hulk nft great choice on a tesla good luck	0
3	the first project in crypto with move to earn ...	1
4	the first project in crypto with move to earn ...	1
...	...	...
96359	cryptogang x fe d a e e ea a d ac e c b fc tha...	0
96360	just registered to win with let s go freemint ...	6
96361	dyor on the nft collection dictators they re b...	0
96362	nft nftcommmunity deltaflare join me on discor...	1
96363	nft finley eth nice	0

96364 rows × 2 columns

In [ ]:
```
```

In [133]:
```
import tensorflow as tf
import tensorflow_hub as hub
import tensorflow_text as text
```

In [134]:
```
from sklearn.model_selection import train_test_split
```

In [135]:
```
#ASSIGNING INDEPENDENT AND DEPENDENT VARIABLE
x=df_1.tweet_text
y=df_1.target
```

In [136]:
```
#SPLITTING THE MODEL FOR TRAIN TEST SPLIT
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)

print("Shape of X_train: ", x_train.shape)
print("Shape of X_test: ", x_test.shape)
```

```
Shape of X_train:  (67454,)
Shape of X_test:  (28910,)
```

In [137]:
```
x_train.head()
```

Out[137]:
```
72852    this is definitely going to be huge as the tea...
66848    just registered for whitelist via metaforest g...
44393                            solaverse nft good soul xg
94685    just registered for the murakami lucky cat coi...
53032    nft i get wl it moons i don t well better let ...
Name: tweet_text, dtype: object
```

In [112]:
```
#BERT INBUILY TRAINED MODEL FOR TEXT CLASSIFICATION
bert_preprocess = hub.KerasLayer("https://tfhub.dev/tensorflow/bert_en_uncased_pr
bert_encoder = hub.KerasLayer("https://tfhub.dev/tensorflow/bert_en_uncased_L-12_
```

In [116]:
```
def get_sentence_embeding (sentences):
    preprocessed_text = bert_preprocess (sentences)
    return bert_encoder(preprocessed_text) ["pooled_output"]

get_sentence_embeding([
    "Wow-sees discount, hurry up",
    "Bhavin, are you up for a volleybal game tomorrow?"
])
```

Out[116]:
```
<tf.Tensor: shape=(2, 768), dtype=float32, numpy=
array([[-0.849521  , -0.4166257 , -0.8501835 , ..., -0.7018479 ,
        -0.68762875,  0.90823793],
       [-0.8720834 , -0.50543964, -0.9444668 , ..., -0.8584752 ,
        -0.7174533 ,  0.88082963]], dtype=float32)>
```

In [119]:
```
# Bert Loyers

text_input=tf.keras.layers. Input (shape=(), dtype=tf.string, name="text")
preprocessed_text=bert_preprocess(text_input)
outputs = bert_encoder(preprocessed_text)
```

In [122]:
```
#NEURAL NETWORK LAYERS
l=tf.keras.layers.Dropout(0.1,name='dropout')(outputs['pooled_output'])
l=tf.keras.layers.Dense(1,activation='sigmoid',name='output')(l)
#FINAL MODEL
model=tf.keras.Model(inputs=[text_input], outputs=[l])
```

In [124]:
```
model.summary()
```

```
Model: "model_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #   
=================================================================
 text (InputLayer)           [(None,)]                 0         
                                                                 
 keras_layer_2 (KerasLayer)  {'input_mask': (None, 128),  0         
                              'input_type_ids': (None, 128),
                              'input_word_ids': (None, 128)}
                                                                 
 keras_layer_3 (KerasLayer)  {'default': (None, 768),  109482241 
                              'encoder_outputs':
                              [(None, 128, 768),
                              (None, 128, 768),
                              (None, 128, 768),
                              (None, 128, 768),
                              (None, 128, 768),
                              (None, 128, 768),
                              (None, 128, 768),
                              (None, 128, 768),
                              (None, 128, 768),
                              (None, 128, 768),
                              (None, 128, 768),
                              (None, 128, 768)],
                              'pooled_output': (None, 768),
                              'sequence_output': (None, 128, 768)}
                                                                 
 dropout (Dropout)           (None, 768)               0         
                                                                 
 output (Dense)              (None, 1)                 769       
                                                                 
=================================================================
Total params: 109,483,010
Trainable params: 769
Non-trainable params: 109,482,241
_________________________________________________________________
```

In [126]:
```
METRICS=[
    tf.keras.metrics.BinaryAccuracy(name='Accuracy'),
    tf.keras.metrics.Precision(name='Precision'),
    tf.keras.metrics.Recall(name='Recall')
]

model.compile(optimizer='adam',
              loss='binary_crossentropy',
              metrics=METRICS)
```

In [*]:
```
model.fit(x_train,y_train)
```

```
 610/2108 [=======>......................] - ETA: 11:39:35 - loss: -130.8969 -
Accuracy: 0.4271 - Precision: 0.8267 - Recall: 1.0000
```

In [ ]:
```
# AS ABOVE MODEL IS TAKING HOURS TO COMPLETE THE TRAINING I HAVE WRITTEN DOWN THE
#FOLLOWING CODES FROM WHICH I WAS GOING TO PREDICT THE MISSING INTENT
model.evaluate(x_test,y_test)
```

In [ ]:
```
y_predicted=model.predicted(x_test)
y_predicted
```

In [ ]:
```
from sklearn.metrics import confusion_metrics, classification_report
cm= confusion_matrix(y_test, y_predicted)
cm
```

In [ ]:
```
from matplotlib import pyplot as plt
import seaborn as sn
sn.heatmap(cm,annot=True,fmt='d')
plt.xlabel('Predicted')
plt.ylabel('Truth')
```

In [ ]:
```
print(classification_report(y_test,y_predicted))
```

In [ ]:
```
new_x=df_2.tweet_text
mode.predict(new_x)
```

In [ ]:
```
```