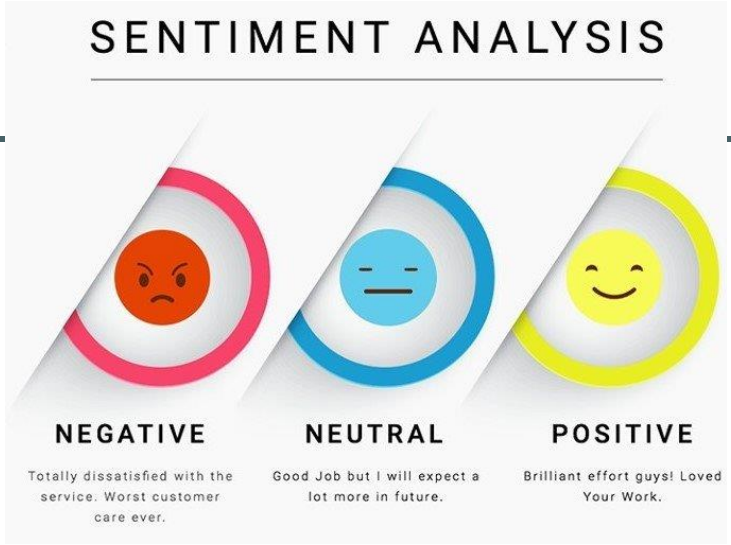
# Tweet Sentiment Prediction

**RASHID ALI**

INTERNSHIP TASK

## INSTRUCTIONS

Our goal is to Predict the Sentiment of Tweet using Machine Learning

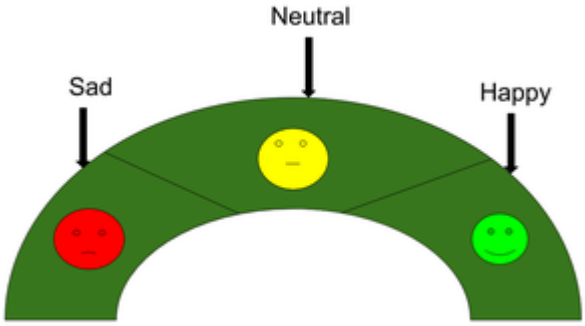


## THE PROCESS

What is Sentiment Analysis?

Sentiment analysis is the process of classifying whether a block of text is positive, negative, or, neutral. Sentiment analysis is contextual mining of words which indicates the social sentiment of a brand and also helps the business to determine whether the product which they are manufacturing is going to make a demand in the market or not. The goal which Sentiment analysis tries to gain is to analyse people's opinion in a way that it can help the businesses expand. It focuses not only on polarity (positive, negative & neutral) but also on emotions (happy, sad, angry, etc.). It uses various Natural Language Processing algorithms such as Rule-based, Automatic, and Hybrid.

For example, if we want to analyze whether a product is satisfying customer requirements, or is there a need for this product in the market? We can use sentiment analysis to monitor that product's reviews. Sentiment analysis is also efficient to use when there is a large set of unstructured data, and we want to classify that data by automatically tagging it.Net Promoter Score (NPS) surveys are used extensively to gain knowledge of how a customer perceives a product or service. Sentiment analysis also gained its popularity due to its feature to process large volumes of NPS responses and obtain consistent results quickly.




Why perform Sentiment Analysis?

According to the survey,80% of the world's data is unstructured. The data needs to be analysed and be in a structured manner whether it is in the form of emails, texts, documents, articles, and many more.

1. Sentiment Analysis is required as it stores data in an efficient, cost-friendly.
2. Sentiment analysis solves real-time issues and can help you solve all the real-time scenarios.

# Process

1. Importing Libraries
2. Load the Data
3. Text Preprocessing
4. EDA
5. Feature Engineering
6. Model Selection
7. Hyperparameter Tuning

## Importing Libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import re
import string
import nltk
import warnings
%matplotlib inline

warnings.filterwarnings('ignore')
```

## Load Dataset

```
data = pd.read_excel('C:/Users/Acer/Downloads/Tweet_NFT.xlsx')
```

```
data.head()
```

| | id | tweet_text | tweet_created_at | tweet_intent |
|---|---|---|---|---|
| 0 | 1212762.0 | @crypto_brody @eCoLoGy1990 @MoonrunnersNFT @It... | 2022-08-06T16:56:36.000Z | Community |
| 1 | 1212763.0 | Need Sick Character artâ_x009d_"#art #artist #... | 2022-08-06T16:56:36.000Z | Giveaway |
| 2 | 1212765.0 | @The_Hulk_NFT @INagotchiNFT @Tesla @killabears... | 2022-08-06T16:56:35.000Z | Appreciation |
| 3 | 1212766.0 | @CryptoBatzNFT @DarekBTW The first project in ... | 2022-08-06T16:56:35.000Z | Community |
| 4 | 1212767.0 | @sashadysonn The first project in crypto with ... | 2022-08-06T16:56:34.000Z | Community |

# Text Preprocessing

# Code for remove Pattern from Text

```
# removes pattern in the input text
def remove_pattern(input_txt, pattern):
    r = re.findall(pattern, input_txt)
    for word in r:
        input_txt = re.sub(word, "", input_txt)
    return input_txt
```

# Remove tag from tweet

```
#remove tag from each tweet

data['clean_tweet'] = np.vectorize(remove_pattern)(data['tweet_text'], "@[\w]*")
```

# Remove special Character, number and Punctuations

```
# remove special characters, numbers and punctuations

data['clean_tweet'] = data['clean_tweet'].str.replace("[^a-zA-Z#]", " ")
data.head()
```

# Remove Short words

```
# remove short words

data['clean_tweet'] = data['clean_tweet'].apply(lambda x: " ".join([w for w in x.split() if len(w)>3]))
data.head()
```

#Lemmatize the words in Tweet

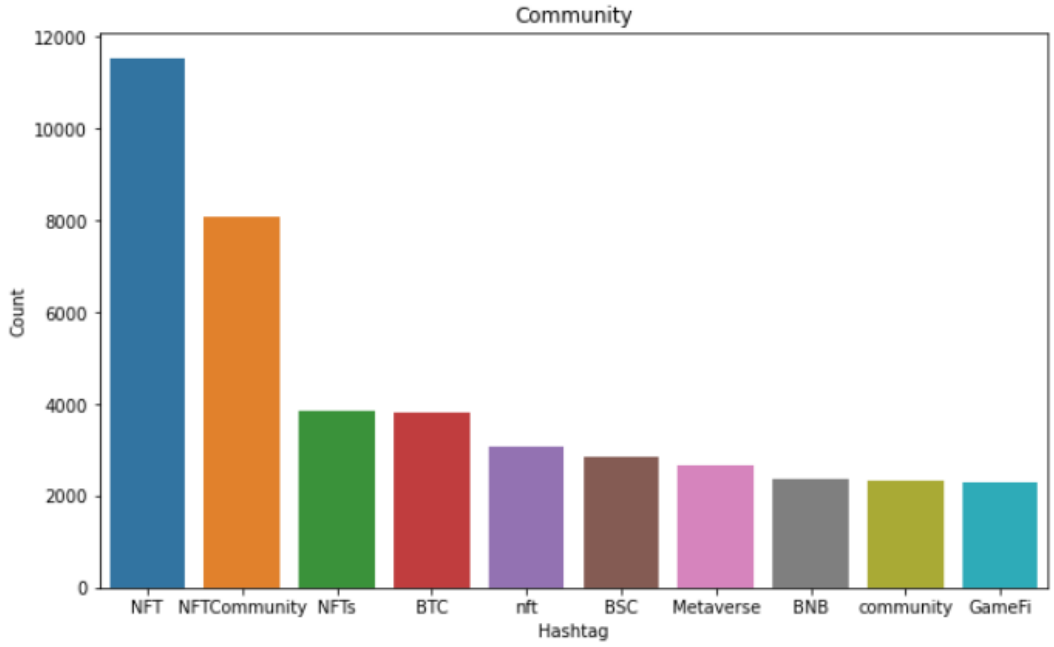
```
#Lemmatize the words
import nltk
nltk.download('wordnet')
nltk.download('omw-1.4')
from nltk.stem import WordNetLemmatizer
wnl = WordNetLemmatizer()

tokenized_tweet = tokenized_tweet.apply(lambda sentence : [wnl.lemmatize(words) for words in sentence])
tokenized_tweet.head()
```

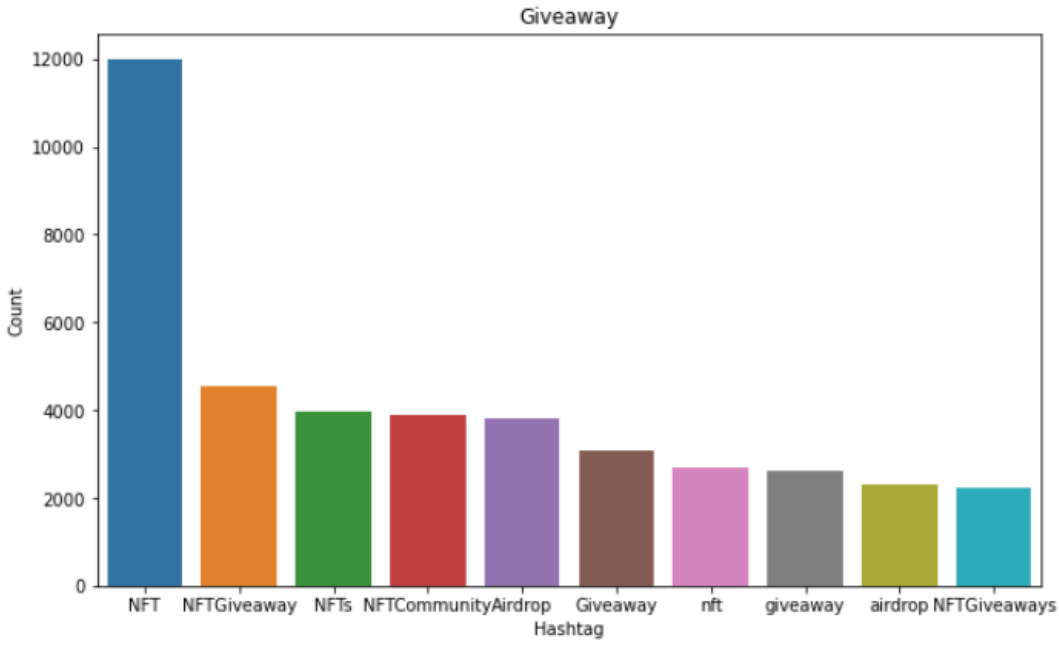
# Final Data

| | tweet_text | tweet_intent | clean_tweet |
|---|---|---|---|
| 0 | @crypto_brody @eCoLoGy1990 @MoonrunnersNFT @It... | Community | cryptocurrency born fan Chihuahua meme communi... |
| 1 | Need Sick Character artâ_x009d_˜#art #artist #... | Giveaway | Need Sick Character #art #artist #Artists #ani... |
| 2 | @The_Hulk_NFT @INagotchiNFT @Tesla @killabears... | Appreciation | Great choice Tesla Good luck |
| 3 | @CryptoBatzNFT @DarekBTW The first project in ... | Community | first project crypto with move earn #AstroBird... |
| 4 | @sashadysonn The first project in crypto with ... | Community | first project crypto with move earn #AstroBird... |
| 5 | ðŸŽ‰ Just registered for the saphire on @PREMI... | Presale | Just registered saphire http INIXaPFL |
| 6 | ðŸš¨ THE BRIDGED #4660/9999 SOLD!!! => PRIC... | Giveaway | BRIDGED SOLD PRICE RANK RANK OWNER rFNh sSjGcJ... |
| 7 | @mtnDAO PROJECT 21 - THE BEST GAMEFI PROJECT O... | Whitelist | PROJECT BEST GAMEFI PROJECT Multistage deflati... |
| 8 | @Ra8bitsNFT Feature it on @GlobalNft07\nWe hav... | Community | Feature have great community artist collector |
| 9 | @SpaceBrosBSC PROJECT 21 - THE BEST GAMEFI PRO... | Whitelist | PROJECT BEST GAMEFI PROJECT Multistage deflati... |

# Exploratory Data Analysis (EDA)

# Ten Most used Hashtags of Community and Giveaway Intend




# Feature Engineering

**TF-IDF** stands for Term Frequency Inverse Document Frequency of records. It can be defined as the calculation of how relevant a word in a series or corpus is to a text. The meaning increases proportionally to the number of times in the text a word appears but is compensated by the word frequency in the corpus (dataset).

```
# feature extraction
from sklearn.feature_extraction.text import TfidfVectorizer
tfidf = TfidfVectorizer(max_df=0.9,min_df=2,max_features=1000, stop_words='english')
vector = tfidf.fit_transform(data['clean_tweet'])
```

# Model Selection

**By calculating** best score for each classification model. I have created

| | model | best_score | best_params |
|---|---|---|---|
| 0 | random_forest | 0.949220 | {'criterion': 'gini', 'max_depth': 200, 'n_est... |
| 1 | logistic_regression | 0.944862 | {'C': 5, 'penalty': 'l1', 'solver': 'liblinear'} |
| 2 | decision_tree | 0.944018 | {'criterion': 'gini', 'max_depth': 300} |
| 3 | Naive Bayes | 0.789354 | {} |

DataFrame. Which strongly depicts that Random Forest is behaving well with this Data. Apart from that testing Data accuracy is also about 95% which shows our best Model is Random Forest. And we further tune the Random Forest for further accuracy. Classification Report of the model :

```
Classification Report:
                precision    recall  f1-score   support

  Appreciation       0.94      0.96      0.95      4151
     Community       0.97      0.99      0.98     10320
          Done       0.67      0.86      0.75       773
      Giveaway       0.98      0.94      0.96      5273
    Interested       0.94      0.92      0.93        66
Launching Soon       1.00      0.86      0.93        36
       Presale       0.97      0.95      0.96      1243
     Whitelist       0.91      0.80      0.85      2116
      pinksale       0.97      1.00      0.99       113

      accuracy                           0.95     24091
     macro avg       0.93      0.92      0.92     24091
  weighted avg       0.95      0.95      0.95     24091

Accuracy: 0.9516001826408202
```

## Predicted Intend of the Blank Data

| | tweet_text | tweet_intent | clean_tweet | Predicted Int |
|---|---|---|---|---|
| **124270** | @jalantathomas_ @damadril @LifesAJoke_NFT @Fkn... | NaN | Here this next http XIHV | D |
| **116053** | The Nfts Will Reveal Soon @420x69IQ @Saadsaqib... | NaN | Nfts Will Reveal Soon http CkVQhMV | Whit |
| **117820** | @Agnes_rose3 New Project!! \nMetaHero is a col... | NaN | Project MetaHero collection hero native MetaHe... | Commu |
| **119491** | @UNLEASHED_NFT Hey DM me Letâ€™s Collab ðŸ'€ | NaN | Collab | D |
| **107155** | ðŸ¢ Buy $ALPACA [Spot Hourly]\n#ALPACAUSDT\nF... | NaN | ALPACA Spot Hourly #ALPACAUSDT Filled #ALPACA ... | Whit |
| **100686** | Every time a piece of #NFT doesn't sell insta... | NaN | Every time piece #NFT doesn sell instantly thi... | Apprecia |
| **106102** | @SpaceRiders_NFT @AkhilSesh Welcome @Akhilsesh... | NaN | Welcome #LetsRide http Bnuf EdHK | Apprecia |
| **96377** | Just remember this is the beginning. You never... | NaN | Just remember this beginning never late Spread... | Commu |
| **104805** | This zesty faucet from @_bitcoiner is making m... | NaN | This zesty faucet from making tweet this claim... | Whit |
| **101521** | @nft_finley I'm pediatric nurse, I create NFT ... | NaN | pediatric nurse create with sick child profit ... | Apprecia |

# THANK YOU