Sentiment analysis - Base code

April 14, 2022

0.1 Importing Liberaries

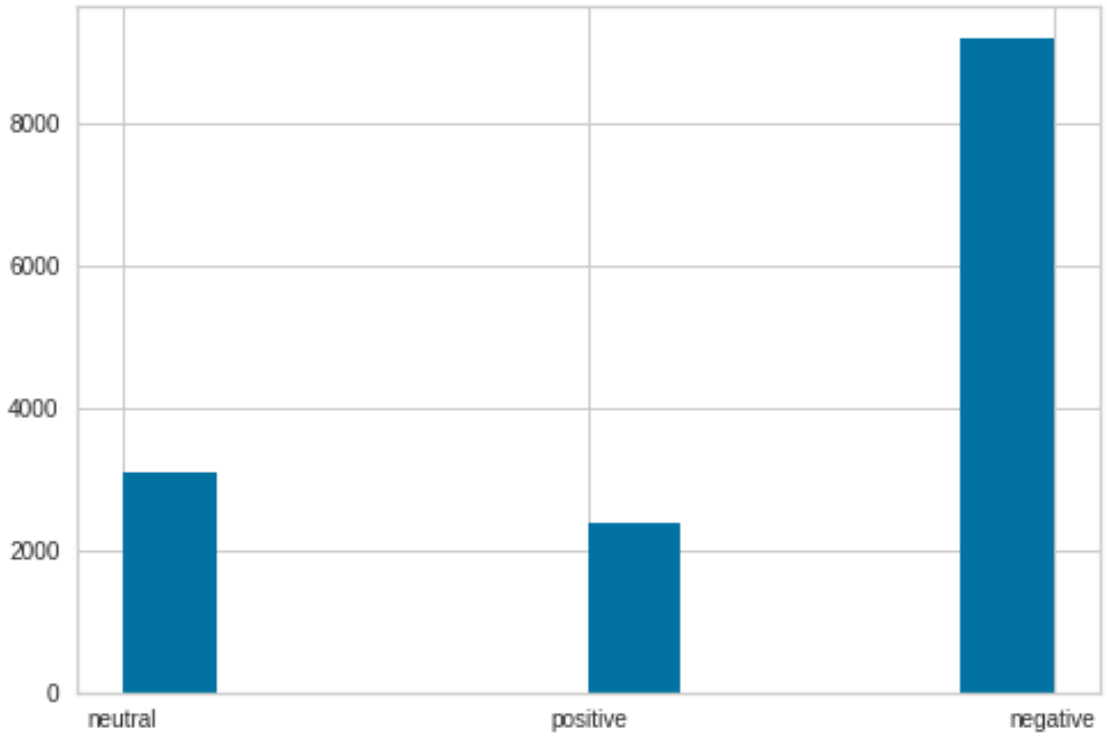
[1]:
```
import numpy as np
import pandas as pd # For reading CSV
import multiprocessing as mp
import string

## preprocessing liberaries

import nltk
from sklearn.feature_extraction.text import CountVectorizer

## Data visualising liberaries

import re

from matplotlib import pyplot as plt
from matplotlib import patches as mpatches
%matplotlib inline
import seaborn as sns

## For comparing models

from pycaret.classification import setup
from pycaret.classification import compare_models

## classification liberaries

from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import train_test_split
from sklearn.model_selection import KFold
from sklearn.feature_selection import SelectKBest, chi2
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.pipeline import Pipeline
from sklearn.metrics import accuracy_score, precision_score, recall_score,␣
 ↪f1_score, roc_auc_score, confusion_matrix, classification_report
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import LogisticRegression
```

```
from sklearn.svm import LinearSVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.preprocessing import OrdinalEncoder
```

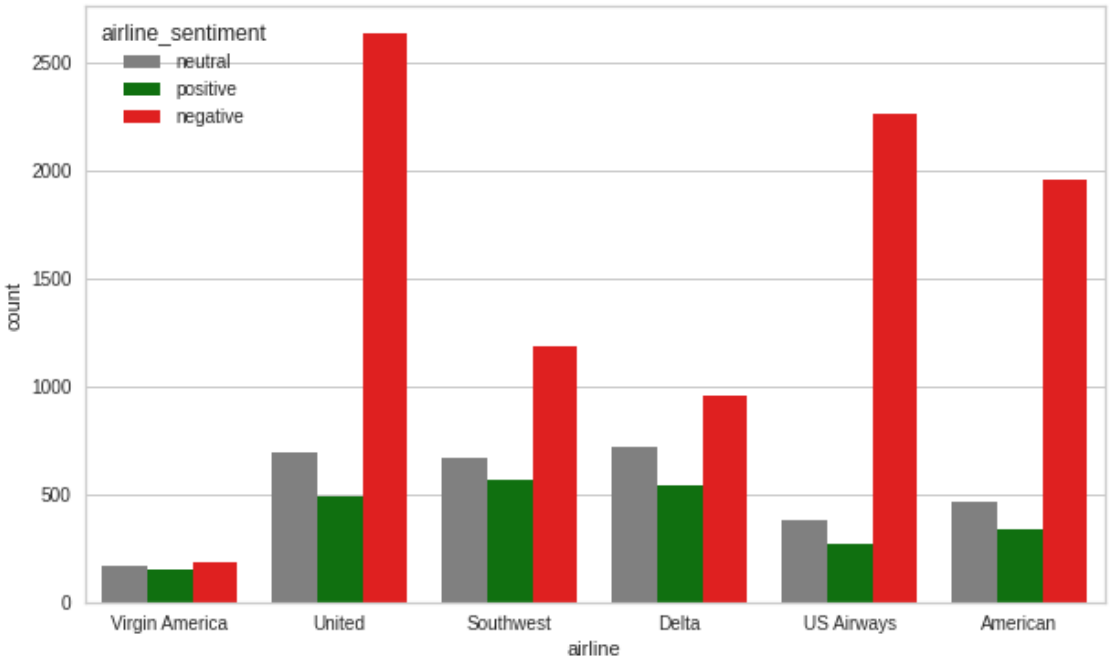
[2]:
```
tweets = pd.read_csv('Tweets.csv')

dataset = tweets.copy()
```

[3]:
```
dataset['airline_sentiment'].hist()
```

[3]: <AxesSubplot:>



[4]:
```
# No. of positive , neutral and negative tweets in dataset per airline

plt.rcParams["figure.figsize"] = (10,6)

colors ={"neutral": "gray", "positive": "green", "negative": "red"}
ax = sns.countplot(data = dataset, x ="airline", hue = "airline_sentiment",␣
 ↪palette=colors)
```



[5]:
```
dataset.shape
dataset.describe
```

[5]:
```
<bound method NDFrame.describe of                 tweet_id airline_sentiment
airline_sentiment_confidence  \
0      570306133677760513           neutral                        1.0000
1      570301130888122368          positive                        0.3486
2      570301083672813571           neutral                        0.6837
3      570301031407624196          negative                        1.0000
4      570300817074462722          negative                        1.0000
…                     …                 …                           …
14635  569587686496825344          positive                        0.3487
14636  569587371693355008          negative                        1.0000
14637  569587242672398336           neutral                        1.0000
14638  569587188687634433          negative                        1.0000
14639  569587140490866689           neutral                        0.6771

               negativereason  negativereason_confidence         airline  \
0                         NaN                        NaN  Virgin America
1                         NaN                     0.0000  Virgin America
2                         NaN                        NaN  Virgin America
3                  Bad Flight                     0.7033  Virgin America
4                  Can't Tell                     1.0000  Virgin America
…                         …                          …             …
14635                     NaN                     0.0000        American
```

```
14636  Customer Service Issue                     1.0000        American
14637                     NaN                        NaN        American
14638  Customer Service Issue                     0.6659        American
14639                     NaN                     0.0000        American

      airline_sentiment_gold             name negativereason_gold  \
0                        NaN          cairdin                 NaN
1                        NaN         jnardino                 NaN
2                        NaN       yvonnalynn                 NaN
3                        NaN         jnardino                 NaN
4                        NaN         jnardino                 NaN
…                        …              …                    …
14635                    NaN  KristenReenders                 NaN
14636                    NaN         itsropes                 NaN
14637                    NaN         sanyabun                 NaN
14638                    NaN       SraJackson                 NaN
14639                    NaN        daviddtwu                 NaN

       retweet_count                                                text  \
0                  0                @VirginAmerica What @dhepburn said.
1                  0  @VirginAmerica plus you've added commercials t…
2                  0  @VirginAmerica I didn't today… Must mean I n…
3                  0  @VirginAmerica it's really aggressive to blast…
4                  0  @VirginAmerica and it's a really big bad thing…
…                …                                                   …
14635              0  @AmericanAir thank you we got on a different f…
14636              0  @AmericanAir leaving over 20 minutes Late Flig…
14637              0  @AmericanAir Please bring American Airlines to…
14638              0  @AmericanAir you have my money, you change my …
14639              0  @AmericanAir we have 8 ppl so we need 2 know h…

      tweet_coord              tweet_created tweet_location  \
0             NaN  2015-02-24 11:35:52 -0800            NaN
1             NaN  2015-02-24 11:15:59 -0800            NaN
2             NaN  2015-02-24 11:15:48 -0800      Lets Play
3             NaN  2015-02-24 11:15:36 -0800            NaN
4             NaN  2015-02-24 11:14:45 -0800            NaN
…           …                          …             …
14635         NaN  2015-02-22 12:01:01 -0800            NaN
14636         NaN  2015-02-22 11:59:46 -0800          Texas
14637         NaN  2015-02-22 11:59:15 -0800  Nigeria,lagos
14638         NaN  2015-02-22 11:59:02 -0800     New Jersey
14639         NaN  2015-02-22 11:58:51 -0800     dallas, TX

                   user_timezone
0      Eastern Time (US & Canada)
1      Pacific Time (US & Canada)
```

```
2      Central Time (US & Canada)
3      Pacific Time (US & Canada)
4      Pacific Time (US & Canada)
…                             …
14635                        NaN
14636                        NaN
14637                        NaN
14638  Eastern Time (US & Canada)
14639                        NaN

[14640 rows x 15 columns]>
```

[6]:
```
dataset.drop(['tweet_id' , 'airline_sentiment_confidence' , 'negativereason' ,␣
↪'negativereason_gold' , 'airline_sentiment_gold' , 'name' , 'user_timezone'␣
↪, 'tweet_location' , 'tweet_created' , 'tweet_coord' ,'negativereason_gold',␣
↪'retweet_count' , 'negativereason_confidence'], axis=1 , inplace = True)
```

[7]:
```
dataset.shape
```

[7]:
```
(14640, 3)
```

[8]:
```
dataset["airline_sentiment"].value_counts()
```

[8]:
```
negative    9178
neutral     3099
positive    2363
Name: airline_sentiment, dtype: int64
```

0.2 Removing Puntuation

[9]:
```
def remove_pun(txt):
    text_nopun = "".join([c for c in txt if c not in string.punctuation])
    text_lower = "".join([c.lower() for c in text_nopun])
    return text_lower

dataset['data_no_Puntuation'] = dataset['text'].apply(lambda x: remove_pun(x))
dataset.head()
```

[9]:
```
  airline_sentiment         airline  \
0           neutral  Virgin America
1          positive  Virgin America
2           neutral  Virgin America
3          negative  Virgin America
4          negative  Virgin America

                                                 text  \
0                @VirginAmerica What @dhepburn said.
```

```
1  @VirginAmerica plus you've added commercials t…
2  @VirginAmerica I didn't today… Must mean I n…
3  @VirginAmerica it's really aggressive to blast…
4  @VirginAmerica and it's a really big bad thing…

                                  data_no_Puntuation
0                   virginamerica what dhepburn said
1  virginamerica plus youve added commercials to …
2  virginamerica i didnt today must mean i need t…
3  virginamerica its really aggressive to blast o…
4  virginamerica and its a really big bad thing a…
```

0.3 Tokeniseing

[10]:
```
def tokenize(txt):
    tokens = re.split('\W+' , txt)
    return tokens

dataset['text_tokenised'] = dataset['data_no_Puntuation'].apply(lambda x :␣
 ↪tokenize(x))
dataset.head()
```

[10]:
```
  airline_sentiment         airline  \
0           neutral  Virgin America
1          positive  Virgin America
2           neutral  Virgin America
3          negative  Virgin America
4          negative  Virgin America

                                                 text  \
0                @VirginAmerica What @dhepburn said.
1  @VirginAmerica plus you've added commercials t…
2  @VirginAmerica I didn't today… Must mean I n…
3  @VirginAmerica it's really aggressive to blast…
4  @VirginAmerica and it's a really big bad thing…

                                  data_no_Puntuation  \
0                   virginamerica what dhepburn said
1  virginamerica plus youve added commercials to …
2  virginamerica i didnt today must mean i need t…
3  virginamerica its really aggressive to blast o…
4  virginamerica and its a really big bad thing a…

                                    text_tokenised
0              [virginamerica, what, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, i, didnt, today, must, mean, i…
```

```
3  [virginamerica, its, really, aggressive, to, b…
4  [virginamerica, and, its, a, really, big, bad,…
```

0.4 Removing stopword

[11]:
```
stopword = nltk.corpus.stopwords.words('english')
def remove_stopWord(txt_tokenised):
    txt_clean = [word for word in txt_tokenised if word not in stopword]
    return txt_clean

dataset['text_no_SW'] = dataset['text_tokenised'].apply(lambda x :␣
 ↪remove_stopWord(x))
dataset.head()
```

[11]:
```
  airline_sentiment         airline  \
0           neutral  Virgin America
1          positive  Virgin America
2           neutral  Virgin America
3          negative  Virgin America
4          negative  Virgin America

                                                 text  \
0               @VirginAmerica What @dhepburn said.
1  @VirginAmerica plus you've added commercials t…
2  @VirginAmerica I didn't today… Must mean I n…
3  @VirginAmerica it's really aggressive to blast…
4  @VirginAmerica and it's a really big bad thing…

                                  data_no_Puntuation  \
0                  virginamerica what dhepburn said
1  virginamerica plus youve added commercials to …
2  virginamerica i didnt today must mean i need t…
3  virginamerica its really aggressive to blast o…
4  virginamerica and its a really big bad thing a…

                                       text_tokenised  \
0             [virginamerica, what, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, i, didnt, today, must, mean, i…
3  [virginamerica, its, really, aggressive, to, b…
4  [virginamerica, and, its, a, really, big, bad,…

                                          text_no_SW
0                   [virginamerica, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, really, aggressive, blast, obn…
```

```
4           [virginamerica, really, big, bad, thing]
```

[12]:
```
def remove_based_on_length(text_no_SW):
    text_lengh_based = [word for word in text_no_SW if len(word) in range(3 ,␣
 ↪21)]
    return text_lengh_based

dataset['text_lengh_based'] = dataset['text_no_SW'].apply(lambda x :␣
 ↪remove_based_on_length(x))
dataset.head()
```

[12]:
```
  airline_sentiment         airline  \
0           neutral  Virgin America
1          positive  Virgin America
2           neutral  Virgin America
3          negative  Virgin America
4          negative  Virgin America

                                                 text  \
0               @VirginAmerica What @dhepburn said.
1  @VirginAmerica plus you've added commercials t…
2  @VirginAmerica I didn't today… Must mean I n…
3  @VirginAmerica it's really aggressive to blast…
4  @VirginAmerica and it's a really big bad thing…

                                 data_no_Puntuation  \
0                  virginamerica what dhepburn said
1  virginamerica plus youve added commercials to …
2  virginamerica i didnt today must mean i need t…
3  virginamerica its really aggressive to blast o…
4  virginamerica and its a really big bad thing a…

                                      text_tokenised  \
0             [virginamerica, what, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, i, didnt, today, must, mean, i…
3  [virginamerica, its, really, aggressive, to, b…
4  [virginamerica, and, its, a, really, big, bad,…

                                            text_no_SW  \
0                   [virginamerica, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, really, aggressive, blast, obn…
4          [virginamerica, really, big, bad, thing]

                                  text_lengh_based
```

```
0                    [virginamerica, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, really, aggressive, blast, obn…
4           [virginamerica, really, big, bad, thing]
```

0.5 Stemming list

[13]:
```
ps = nltk.PorterStemmer()

def stemming(text_lengh_based):
    text = [ps.stem(word) for word in text_lengh_based]
    return text

dataset['text_stemized'] = dataset['text_lengh_based'].apply(lambda x :␣
 ↪stemming(x))
dataset.head()
```

[13]:
```
  airline_sentiment         airline  \
0           neutral  Virgin America
1          positive  Virgin America
2           neutral  Virgin America
3          negative  Virgin America
4          negative  Virgin America

                                                 text  \
0                @VirginAmerica What @dhepburn said.
1  @VirginAmerica plus you've added commercials t…
2  @VirginAmerica I didn't today… Must mean I n…
3  @VirginAmerica it's really aggressive to blast…
4  @VirginAmerica and it's a really big bad thing…

                                  data_no_Puntuation  \
0                  virginamerica what dhepburn said
1  virginamerica plus youve added commercials to …
2  virginamerica i didnt today must mean i need t…
3  virginamerica its really aggressive to blast o…
4  virginamerica and its a really big bad thing a…

                                       text_tokenised  \
0             [virginamerica, what, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, i, didnt, today, must, mean, i…
3  [virginamerica, its, really, aggressive, to, b…
4  [virginamerica, and, its, a, really, big, bad,…

                                          text_no_SW  \
```

```
0                    [virginamerica, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, really, aggressive, blast, obn…
4            [virginamerica, really, big, bad, thing]

                                    text_lengh_based  \
0                    [virginamerica, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, really, aggressive, blast, obn…
4            [virginamerica, really, big, bad, thing]

                                      text_stemized
0                    [virginamerica, dhepburn, said]
1  [virginamerica, plu, youv, ad, commerci, exper…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, realli, aggress, blast, obnoxi…
4            [virginamerica, realli, big, bad, thing]
```

0.6 Removing First word as it is flight name only

[14]:
```
flightNames = ["virginamerica" , "united" , "delta" , "southwestair" ,␣
 ↪"usairways" , "americanair"]
def remove_first_word(text_stemized):
    first_Remove = [word for word in text_stemized if word not in flightNames]
    return first_Remove

dataset['first_Remove'] = dataset['text_stemized'].apply(lambda x :␣
 ↪remove_first_word(x))
dataset.head()
```

[14]:
```
  airline_sentiment         airline  \
0           neutral  Virgin America
1          positive  Virgin America
2           neutral  Virgin America
3          negative  Virgin America
4          negative  Virgin America

                                                 text  \
0               @VirginAmerica What @dhepburn said.
1  @VirginAmerica plus you've added commercials t…
2  @VirginAmerica I didn't today… Must mean I n…
3  @VirginAmerica it's really aggressive to blast…
4  @VirginAmerica and it's a really big bad thing…

                                data_no_Puntuation  \
```

```
0                   virginamerica what dhepburn said
1  virginamerica plus youve added commercials to …
2  virginamerica i didnt today must mean i need t…
3  virginamerica its really aggressive to blast o…
4  virginamerica and its a really big bad thing a…

                                      text_tokenised  \
0              [virginamerica, what, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, i, didnt, today, must, mean, i…
3  [virginamerica, its, really, aggressive, to, b…
4  [virginamerica, and, its, a, really, big, bad,…

                                         text_no_SW  \
0                    [virginamerica, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, really, aggressive, blast, obn…
4           [virginamerica, really, big, bad, thing]

                                    text_lengh_based  \
0                    [virginamerica, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, really, aggressive, blast, obn…
4           [virginamerica, really, big, bad, thing]

                                       text_stemized  \
0                    [virginamerica, dhepburn, said]
1  [virginamerica, plu, youv, ad, commerci, exper…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, realli, aggress, blast, obnoxi…
4           [virginamerica, realli, big, bad, thing]

                                       first_Remove
0                                    [dhepburn, said]
1           [plu, youv, ad, commerci, experi, tacki]
2  [didnt, today, must, mean, need, take, anoth, …
3  [realli, aggress, blast, obnoxi, entertain, gu…
4                          [realli, big, bad, thing]
```

0.7 Handeling airline_sentiments

[15]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
dataset['airline_sentiment_encoded'] = le.
 ↪fit_transform(dataset['airline_sentiment'])
```

```
# dataset.head()

result = dataset['airline_sentiment_encoded']
result
```

[15]:
```
0        1
1        2
2        1
3        0
4        0
        ..
14635    2
14636    0
14637    1
14638    0
14639    1
Name: airline_sentiment_encoded, Length: 14640, dtype: int64
```

[16]:
```
def detokenise(first_Remove):
    text = ' '.join([str(word) for word in first_Remove])
    return text

dataset['detokenise_sentance'] = dataset['first_Remove'].apply(lambda x :␣
 ↪detokenise(x))

dataset.head()
```

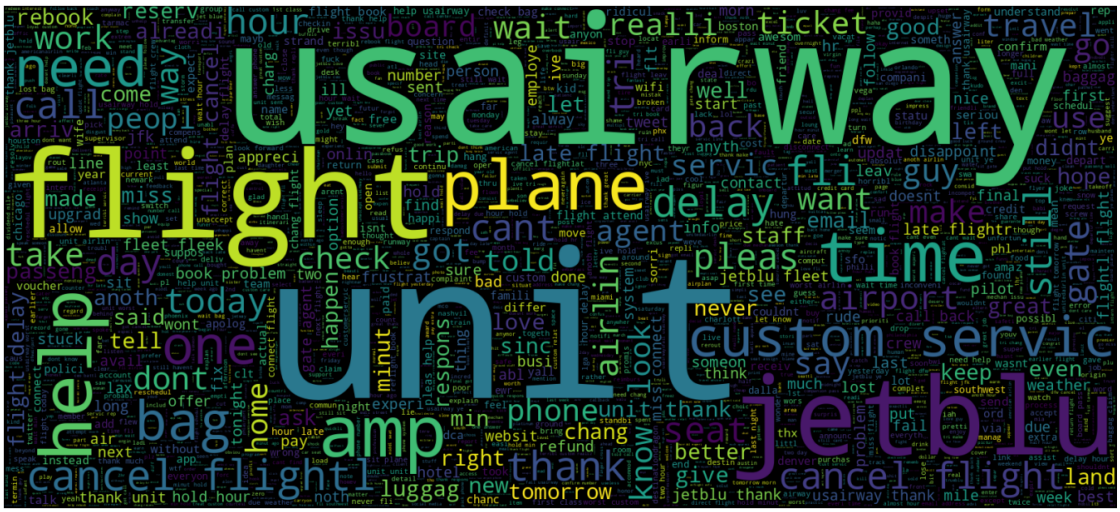
[16]:
```
  airline_sentiment         airline  \
0           neutral  Virgin America
1          positive  Virgin America
2           neutral  Virgin America
3          negative  Virgin America
4          negative  Virgin America

                                                 text  \
0                @VirginAmerica What @dhepburn said.
1  @VirginAmerica plus you've added commercials t…
2  @VirginAmerica I didn't today… Must mean I n…
3  @VirginAmerica it's really aggressive to blast…
4  @VirginAmerica and it's a really big bad thing…

                                data_no_Puntuation  \
0                   virginamerica what dhepburn said
1  virginamerica plus youve added commercials to …
2  virginamerica i didnt today must mean i need t…
3  virginamerica its really aggressive to blast o…
4  virginamerica and its a really big bad thing a…
```

```
                                      text_tokenised  \
0              [virginamerica, what, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, i, didnt, today, must, mean, i…
3  [virginamerica, its, really, aggressive, to, b…
4  [virginamerica, and, its, a, really, big, bad,…

                                         text_no_SW  \
0                    [virginamerica, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, really, aggressive, blast, obn…
4           [virginamerica, really, big, bad, thing]

                                    text_lengh_based  \
0                    [virginamerica, dhepburn, said]
1  [virginamerica, plus, youve, added, commercial…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, really, aggressive, blast, obn…
4           [virginamerica, really, big, bad, thing]

                                      text_stemized  \
0                    [virginamerica, dhepburn, said]
1  [virginamerica, plu, youv, ad, commerci, exper…
2  [virginamerica, didnt, today, must, mean, need…
3  [virginamerica, realli, aggress, blast, obnoxi…
4           [virginamerica, realli, big, bad, thing]

                                       first_Remove  \
0                                  [dhepburn, said]
1           [plu, youv, ad, commerci, experi, tacki]
2  [didnt, today, must, mean, need, take, anoth, …
3  [realli, aggress, blast, obnoxi, entertain, gu…
4                          [realli, big, bad, thing]

   airline_sentiment_encoded  \
0                          1
1                          2
2                          1
3                          0
4                          0

                                detokenise_sentance
0                                     dhepburn said
1                   plu youv ad commerci experi tacki
2          didnt today must mean need take anoth trip
```

```
3  realli aggress blast obnoxi entertain guest fa…
4                                realli big bad thing
```

1 Visualize the Maximum Repeated Words

[17]:
```
from wordcloud import WordCloud

text = " ".join(dataset['detokenise_sentance'].values)
plt.figure(figsize = (24,12))
wordcloud = WordCloud(min_font_size = 3,  max_words = 3200 , width = 1600 ,␣
 ↪height = 720).generate(text)

plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



1.1 APPLYING VECTORISATION

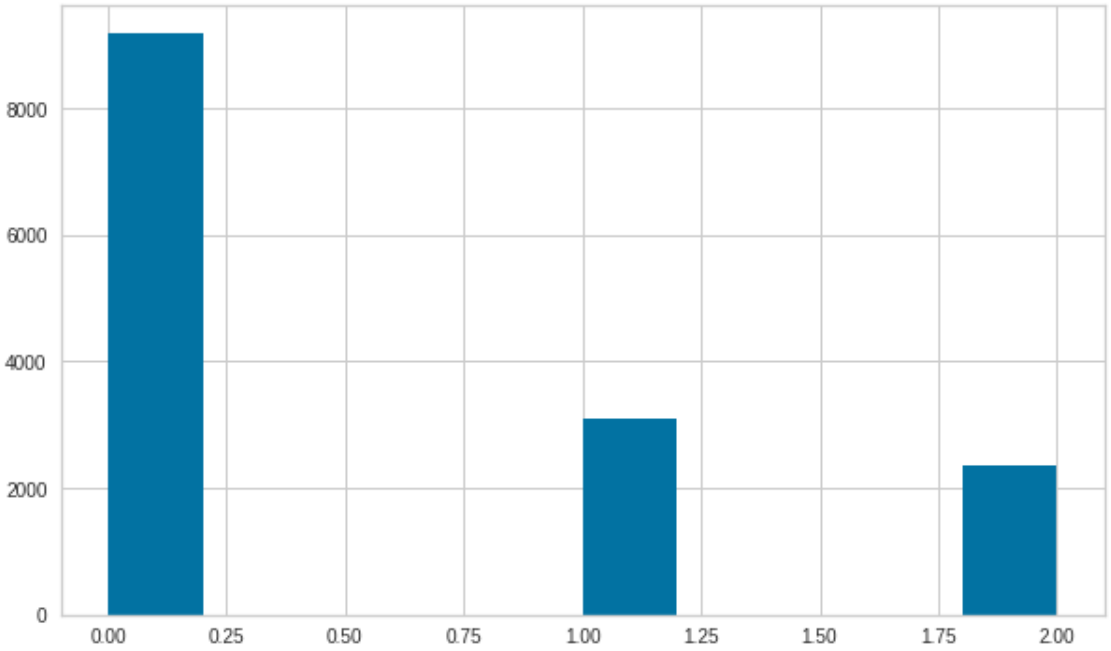
[18]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
tfidf_vect = TfidfVectorizer(lowercase=True)

corpus = dataset['detokenise_sentance']

X = tfidf_vect.fit(corpus)

X = tfidf_vect.transform(corpus)

# df = pd.DataFrame(X.toarray(), columns = tfidf_vect.get_feature_names())
# df
```

2 Handle Imbalanced Dataset

Negative data has been down sampled to balance the proportion of negative and positive tweets

[19]:
```
## 0 --> Negative
## 1 --> Neutral
## 2 --> Positive

Y_final = dataset['airline_sentiment_encoded']

dataset['airline_sentiment_encoded'].hist()
```

[19]: <AxesSubplot:>



[20]:
```
from imblearn.over_sampling import SMOTE # Handling Imbalanced

# Handling imbalanced using SMOTE
smote = SMOTE()
x_sm,y_sm = smote.fit_resample(X,Y_final)
```

2.1 Train - Test split

[21]:
```
X_train , X_test , Y_train , Y_test = train_test_split(x_sm , y_sm ,␣
↪test_size=0.33,random_state=3)
```

[22]:
```
X_train
```

[22]:
```
<18447x13328 sparse matrix of type '<class 'numpy.float64'>'
        with 168028 stored elements in Compressed Sparse Row format>
```

[23]:
```
print(X_train.shape)
print(X_test.shape)
print(Y_train.shape)
print(Y_test.shape)
```

```
(18447, 13328)
(9087, 13328)
(18447,)
(9087,)
```

3 Applying Models

3.1 Naive bayes

[24]:
```
nb = MultinomialNB(alpha=.7)   #try gridsearch
nb.fit(X_train, Y_train)

nb_pred = nb.predict(X_test)
```

[25]:
```
print(confusion_matrix(Y_test, nb_pred))
print(classification_report(Y_test, nb_pred))
accuracy_nb = accuracy_score(Y_test, nb_pred)
print(accuracy_score(Y_test, nb_pred))
```

```
[[2634  265  167]
 [ 613 2041  338]
 [ 157  100 2772]]
              precision    recall  f1-score   support

           0       0.77      0.86      0.81      3066
           1       0.85      0.68      0.76      2992
           2       0.85      0.92      0.88      3029

    accuracy                           0.82      9087
   macro avg       0.82      0.82      0.82      9087
weighted avg       0.82      0.82      0.82      9087

0.8195223946296908
```

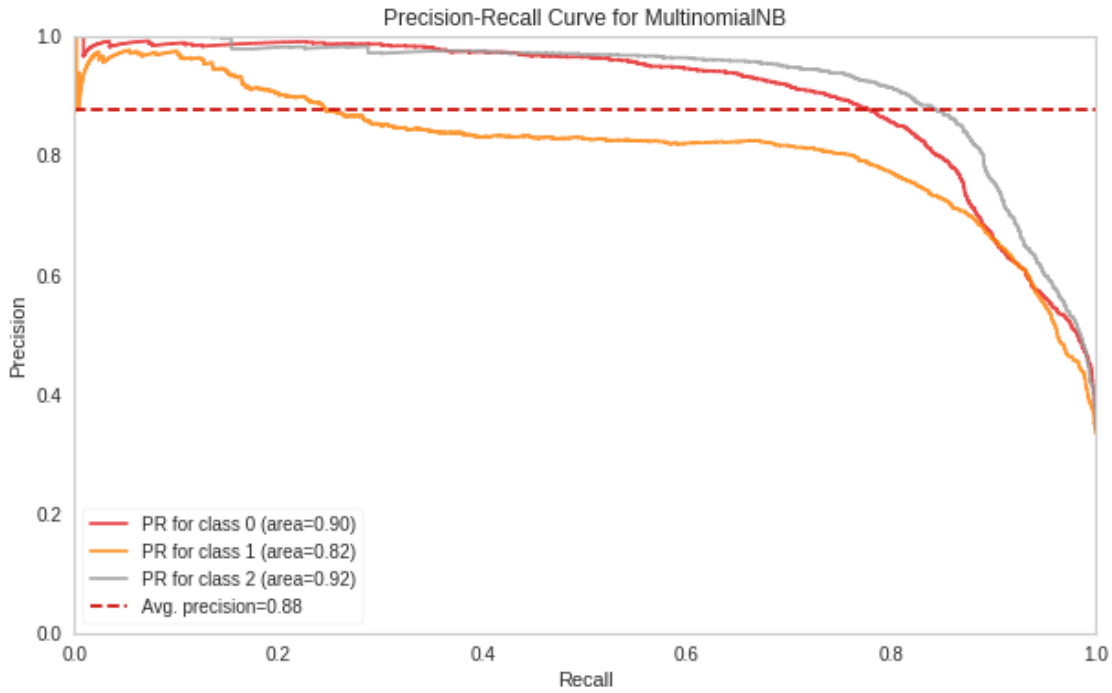
3.1.1 Precesion-Recall Curve for naive bayes

[26]:
```
from yellowbrick.classifier import PrecisionRecallCurve
```

[27]:
```
def precision_recall_curve(curve_model, trained_model, model_name):
    curve = PrecisionRecallCurve(
```

```
        curve_model,
        classes=trained_model.classes_,
        per_class=True,
        cmap="Set1"
    )
    curve.fit(X_train,Y_train)
    curve.score(X_test, Y_test)
#     reports_dict[model_name]["Precision-Recall Curve"] = curve.
 ↪score_["negative"]
    curve.show();
```

[28]:
```
precision_recall_curve(MultinomialNB(), nb, "Naive Bayes")
```



3.2 Logestic Regression

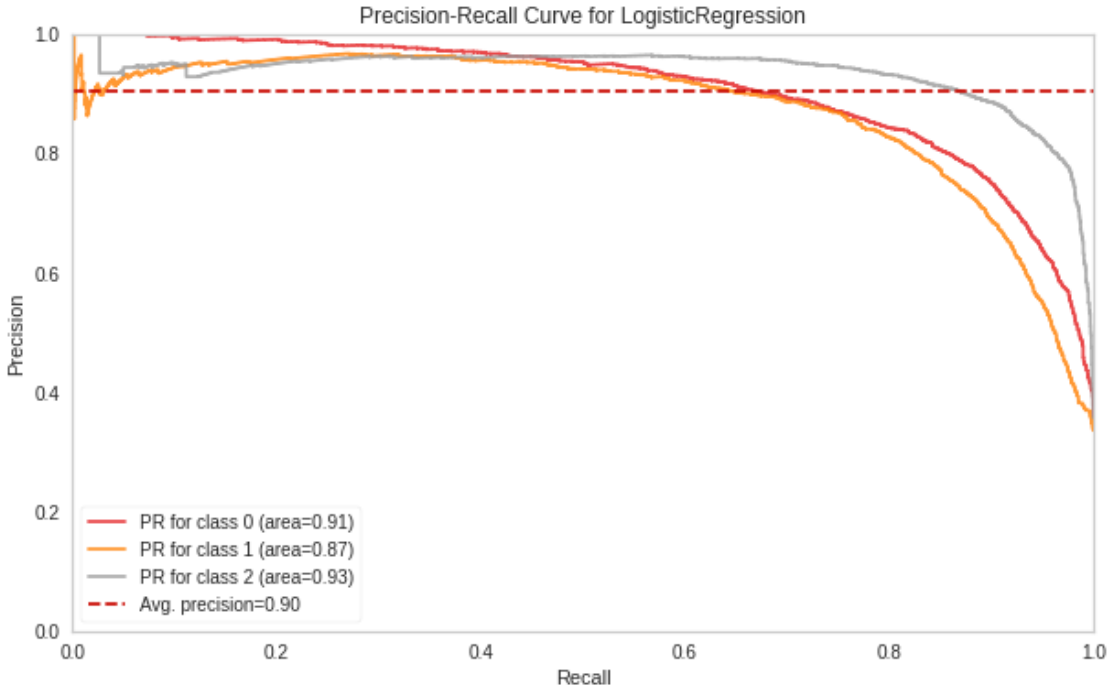
[29]:
```
logistic_regression = LogisticRegression(C=1.0, class_weight=None, dual=False,␣
 ↪fit_intercept=True,
                   intercept_scaling=1, l1_ratio=None, max_iter=1000,
                   multi_class='auto', n_jobs=None, penalty='l2',
                   random_state=7823, solver='lbfgs', tol=0.0001, verbose=0,
                   warm_start=False)
logistic_regression.fit(X_train, Y_train)
```

[29]:
```
LogisticRegression(max_iter=1000, random_state=7823)
```

[30]:
```
Y_pred_logistic_regression = logistic_regression.predict(X_test)
print(confusion_matrix(Y_test, Y_pred_logistic_regression))
print(classification_report(Y_test, Y_pred_logistic_regression))
accuracy_logestic = accuracy_score(Y_test, Y_pred_logistic_regression)
print(accuracy_score(Y_test, Y_pred_logistic_regression))
```

```
[[2483  462  121]
 [ 289 2551  152]
 [ 115  223 2691]]
              precision    recall  f1-score   support

           0       0.86      0.81      0.83      3066
           1       0.79      0.85      0.82      2992
           2       0.91      0.89      0.90      3029

    accuracy                           0.85      9087
   macro avg       0.85      0.85      0.85      9087
weighted avg       0.85      0.85      0.85      9087

0.8501155496863652
```

[31]:
```
precision_recall_curve(LogisticRegression(C=1.0, class_weight=None, dual=False,␣
 ↪fit_intercept=True,
                   intercept_scaling=1, l1_ratio=None, max_iter=1000,
                   multi_class='auto', n_jobs=None, penalty='l2',
                   random_state=7823, solver='lbfgs', tol=0.0001, verbose=0,
                   warm_start=False), logistic_regression, "Logestic␣
 ↪Regression")
```



3.3 Random Forest

[32]:
```
rf = RandomForestClassifier(100,
                            max_depth=40,
                            random_state=42,
                            n_jobs=-1)
rf.fit(X_train, Y_train)

rf_pred = rf.predict(X_test)
```

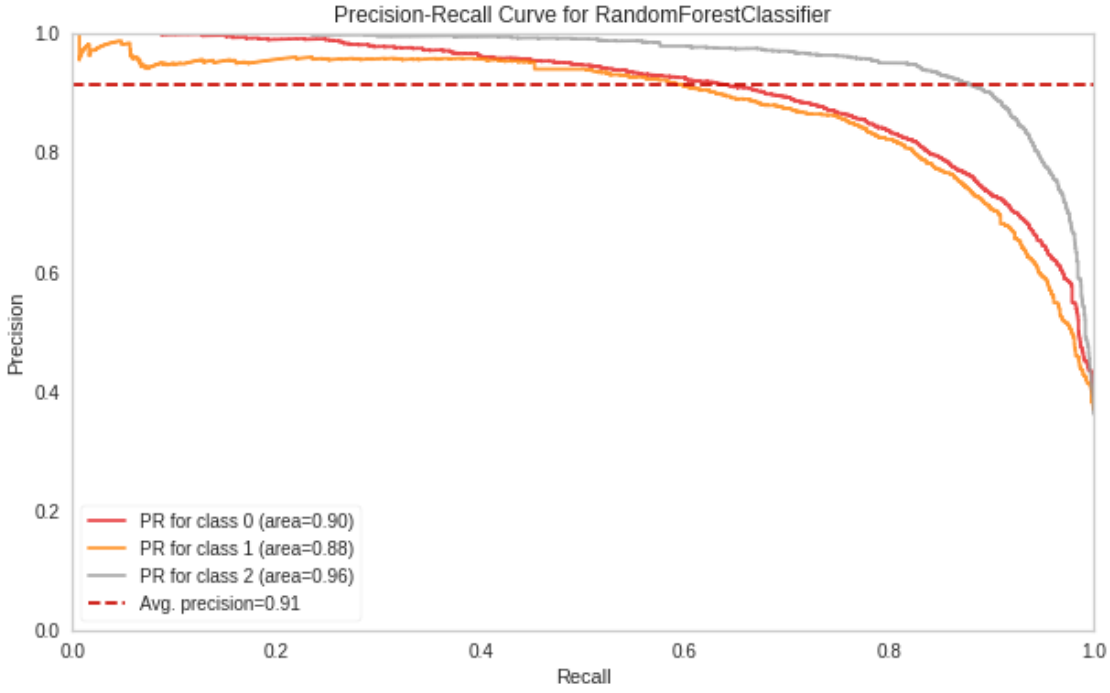
[33]:
```
print(confusion_matrix(Y_test, rf_pred))
print(classification_report(Y_test, rf_pred))
accuracy_rf = accuracy_score(Y_test, rf_pred)
print(accuracy_score(Y_test, rf_pred))
```

```
[[2448  499  119]
 [ 364 2439  189]
 [ 118  229 2682]]
              precision    recall  f1-score   support

           0       0.84      0.80      0.82      3066
           1       0.77      0.82      0.79      2992
           2       0.90      0.89      0.89      3029

    accuracy                           0.83      9087
```

```
   macro avg       0.83      0.83      0.83      9087
weighted avg       0.83      0.83      0.83      9087
```

```
0.8329481677121162
```

[34]:
```
precision_recall_curve(RandomForestClassifier(100,␣
↪max_depth=40,random_state=42,n_jobs=-1),
                       rf,
                       "Random Forest",
                      )
```



3.4 Result

[35]:
```
print("------------------- Accuracy Score ---------------------")
print(" ")
print(f"Naive Bayes  --  {accuracy_nb * 100}")
print(f"Logestic Regression  --  {accuracy_logestic * 100}")
print(f"Random Forest  --  {accuracy_rf * 100}")
```

```
------------------- Accuracy Score ---------------------

Naive Bayes  --  81.95223946296908
Logestic Regression  --  85.01155496863652
Random Forest  --  83.29481677121163
```