untitled5

July 9, 2024

1 Financial Market News - Sentiment Analysis

This is the data(dummy) of Financial Market Top 25 News for the Day and Task is to Train and Predict Model for overall Sentiment Analysis

Import Library

```python
[ ]: import pandas as pd
```

```python
[ ]: import numpy as np
```

Import Dataset

```python
[18]: df = pd.read_csv('/content/Financial Market News.csv', encoding='latin-1')
```

```python
[19]: df.head()
```

```
[19]:          Date  Label                                          News 1  \
      0  01-01-2010      0           McIlroy's men catch cold from Gudjonsson
      1  02-01-2010      0                Warning from history points to crash
      2  03-01-2010      0       Comment: Why Israel's peaceniks feel betrayed
      3  04-01-2010      1                  £750,000-a-goal Weah aims parting shot
      4  05-01-2010      1   Leeds arrive in Turkey to the silence of the fans

                                         News 2  \
      0                         Obituary: Brian Walsh
      1                   Investors flee to dollar haven
      2   Court deals blow to seizure of drug assets
      3                  Newcastle pay for Fletcher years
      4      One woman's vision offers loan lifeline

                                         News 3  \
      0        Workplace blues leave employers in the red
      1                     Banks and tobacco in favour
      2                       An ideal target for spooks
      3   Brown sent to the stands for Scotland qualifier
      4            Working Lives: How world leaders worked

                                         News 4  \
```

```
0                   Classical review: Rattle
1                       Review: Llama Farmers
2   World steps between two sides intent on war
3          Tourists wary of breaking new ground
4          Working Lives: Tricks of the trade


                                    News 5  \
0                   Dance review: Merce Cunningham
1                       War jitters lead to sell-off
2                       What the region's papers say
3           Canary Wharf climbs into the FTSE 100
4   Working Lives: six-hour days, long lunches and…


                                    News 6  \
0   Genetic tests to be used in setting premiums
1                   Your not-so-secret history
2           Comment: Fear and rage in Palestine
3                           Review: Bill Bailey
4                       Pop review: We Love UK


                                    News 7  \
0                   Opera review: La Bohème
1               Review: The Northern Sinfonia
2   Poverty and resentment fuels Palestinian fury
3                           Review: Classical
4               World music review: Marisa Monte


                                    News 8  …  \
0               Pop review: Britney Spears  …
1                           Review: Hysteria  …
2   Republican feud fear as dissident is killed  …
3               Review: New Contemporaries 2000  …
4           Art review:  Hollingsworth/Heyer  …


                    News 16                                News 17  \
0                   Finland 0 - 0 England           Healy a marked man
1   Why Wenger will stick to his Gunners  Out of luck England hit rock bottom
2                   FTSE goes upwardly mobile           At this price?  BP Amoco
3           More cash on way for counties       Cairns carries Kiwis to victory
4           Duisenberg in double trouble           Pru to cut pension charges


                                    News 18  \
0                   Happy birthday Harpers & Queen
1                       Wilkinson out of his depth
2                                           Go fish
3   Year of Blanchflower's flourish when Spurs sto…
4                       Art review: Paul Graham
```

```
                                          News 19  \
0   Win unlimited access to the Raindance film fes…
1                    Kinsella sparks Irish power play
2          Bosnian Serb blows himself up to evade law
3           New direct approach brings only pay-per-blues
4               Shearer shot sparks Boro humiliation

                                          News 20  \
0   Labour pledges £800m to bridge north-south divide
1                       Brown banished as Scots rebound
2                       Orange float delayed to 2001
3                       Third Division   round-up
4   Ridsdale's lingering fears as Leeds revisit Tu…

                                          News 21  \
0       Wales: Lib-Lab pact firm despite resignation
1                     Battling Wales cling to lifeline
2   Angry factory workers root out fear, favours a…
3                       Second Division   round-up
4           Champions League: Rangers v Galatasaray

                              News 22  \
0                       Donald Dewar
1    Ehiogu close to sealing Boro move
2   Smith defied advice on dome payout
3           First Division   round-up
4    Champions League: Lazio v Arsenal

                                          News 23  \
0   Regenerating homes   regenerates well-being in …
1                             Man-to-man marking
2                     Xerox takes the axe to jobs
3           McLean ends his career with a punch
4                             Lazio 1 - 1 Arsenal

                              News 24  \
0    Win £100 worth of underwear
1                     Match stats
2   Comment: Refugees in Britain
3     Heskey grabs triple crown
4           England in Pakistan

                                          News 25
0                       TV guide: Random views
1   French referee at centre of storm is no strang…
2               Maverick who sparked the new intifada
```

```
3                   Weah on his way as City march on
4              England given olive-branch reception

[5 rows x 27 columns]
```

[20]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4101 entries, 0 to 4100
Data columns (total 27 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   Date     4101 non-null   object
 1   Label    4101 non-null   int64
 2   News 1   4101 non-null   object
 3   News 2   4101 non-null   object
 4   News 3   4101 non-null   object
 5   News 4   4101 non-null   object
 6   News 5   4101 non-null   object
 7   News 6   4101 non-null   object
 8   News 7   4101 non-null   object
 9   News 8   4101 non-null   object
 10  News 9   4101 non-null   object
 11  News 10  4101 non-null   object
 12  News 11  4101 non-null   object
 13  News 12  4101 non-null   object
 14  News 13  4101 non-null   object
 15  News 14  4101 non-null   object
 16  News 15  4101 non-null   object
 17  News 16  4101 non-null   object
 18  News 17  4101 non-null   object
 19  News 18  4101 non-null   object
 20  News 19  4101 non-null   object
 21  News 20  4101 non-null   object
 22  News 21  4101 non-null   object
 23  News 22  4101 non-null   object
 24  News 23  4100 non-null   object
 25  News 24  4098 non-null   object
 26  News 25  4098 non-null   object
dtypes: int64(1), object(26)
memory usage: 865.2+ KB
```

[22]: `df.shape`

[22]: `(4101, 27)`

[23]: `df.columns`

```
[23]:  Index(['Date', 'Label', 'News 1', 'News 2', 'News 3', 'News 4', 'News 5',
              'News 6', 'News 7', 'News 8', 'News 9', 'News 10', 'News 11', 'News 12',
              'News 13', 'News 14', 'News 15', 'News 16', 'News 17', 'News 18',
              'News 19', 'News 20', 'News 21', 'News 22', 'News 23', 'News 24',
              'News 25'],
            dtype='object')
```

Get Feature Selection

```
[24]:  ' '.join(str(x) for x in df.iloc[1,2:27])
```

```
[24]:  "Warning from history points to crash Investors flee to dollar haven Banks and
        tobacco in favour Review: Llama Farmers War jitters lead to sell-off Your not-
        so-secret history Review: The Northern Sinfonia Review: Hysteria Review: The
        Guardsman Opera: The Marriage of Figaro Review: The Turk in Italy Deutsche
        spells out its plans for diversification Traders' panic sends oil prices skyward
        TV sport chief leaves home over romance Leader: Hi-tech twitch Why Wenger will
        stick to his Gunners Out of luck England hit rock bottom Wilkinson out of his
        depth Kinsella sparks Irish power play Brown banished as Scots rebound Battling
        Wales cling to lifeline Ehiogu close to sealing Boro move Man-to-man marking
        Match stats French referee at centre of storm is no stranger to controversy"
```

```
[25]:  df.index
```

```
[25]:  RangeIndex(start=0, stop=4101, step=1)
```

```
[26]:  len(df.index)
```

```
[26]:  4101
```

```
[27]:  news = []
       for row in range(0,len(df.index)):
         news.append(' '.join(str(x) for x in df.iloc[row,2:27]))
```

```
[28]:  type(news)
```

```
[28]:  list
```

```
[29]:  news[0]
```

```
[29]:  "McIlroy's men catch cold from Gudjonsson Obituary: Brian Walsh Workplace blues
        leave employers in the red Classical review: Rattle Dance review: Merce
        Cunningham Genetic tests to be used in setting premiums Opera review: La Bohème
        Pop review: Britney Spears Theatre review: The Circle Wales face a fraught night
        Under-21  round-up Smith off to blot his copybook Finns taking the mickey Praise
        wasted as Brown studies injury options Ireland wary of minnows Finland 0 - 0
        England Healy a marked man Happy birthday Harpers & Queen Win unlimited access
        to the Raindance film festival Labour pledges £800m to bridge north-south divide
```

```
Wales: Lib-Lab pact firm despite resignation Donald Dewar Regenerating homes
regenerates well-being in people Win £100 worth of underwear TV guide: Random
views"
```

[30]:
```python
x = news
```

[31]:
```python
type(x)
```

[31]:
```
list
```

Get Feature Text Conversion to Bag of Words

[32]:
```python
from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(lowercase=True, ngram_range=(1,1))
x = cv.fit_transform(x)
```

[33]:
```python
x.shape
```

[33]:
```
(4101, 48527)
```

[34]:
```python
y = df['Label']
```

[35]:
```python
y.shape
```

[35]:
```
(4101,)
```

Get Train Test Split

[36]:
```python
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3,
 ↪random_state=0)
```

[38]:
```python
from sklearn.ensemble import RandomForestClassifier
```

[39]:
```python
rf = RandomForestClassifier(n_estimators=200)
```

[40]:
```python
rf.fit(x_train, y_train)
```

[40]:
```
RandomForestClassifier(n_estimators=200)
```

[41]:
```python
y_pred = rf.predict(x_test)
```

[42]:
```python
from sklearn.metrics import classification_report, confusion_matrix,
 ↪accuracy_score
```

[43]:
```python
confusion_matrix(y_test, y_pred)
```

```
[43]: array([[134, 446],
             [156, 495]])
```

```
[44]: print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.46      0.23      0.31       580
           1       0.53      0.76      0.62       651

    accuracy                           0.51      1231
   macro avg       0.49      0.50      0.46      1231
weighted avg       0.50      0.51      0.47      1231
```