Stock_Sentiment_Analysis_using_News_Headlines

August 3, 2023

0.1 Scenario : Based on the headline will the stock price increase(1) or decrease/constant(0)

0.2 Stock Sentiment Analysis using News Headlines

```python
[1]: # import libraries
     import pandas as pd
     import numpy as np
```

```python
[2]: # read data
     data = pd.read_csv('Data.csv', encoding='ISO-8859-1')
     data
```

```
[2]:           Date  Label                                               Top1  \
     0     2000-01-03      0  A 'hindrance to operations': extracts from the…
     1     2000-01-04      0                                          Scorecard
     2     2000-01-05      0                     Coventry caught on counter by Flo
     3     2000-01-06      1                        Pilgrim knows how to progress
     4     2000-01-07      1                              Hitches and Horlocks
     …            …     …                                                 …
     4096  2016-06-27      0  Barclays and RBS shares suspended from trading…
     4097  2016-06-28      1  2,500 Scientists To Australia: If You Want To …
     4098  2016-06-29      1                 Explosion At Airport In Istanbul
     4099  2016-06-30      1  Jamaica proposes marijuana dispensers for tour…
     4100  2016-07-01      1  A 117-year-old woman in Mexico City finally re…

                                                       Top2  \
     0                                             Scorecard
     1                                    The best lake scene
     2                       United's rivals on the road to Rio
     3                                     Thatcher facing ban
     4                        Beckham off but United survive
     …                                                     …
     4096  Pope says Church should ask forgiveness from g…
     4097  The personal details of 112,000 French police …
     4098  Yemeni former president: Terrorism is the offs…
     4099  Stephen Hawking says pollution and 'stupidity'…
     4100   IMF chief backs Athens as permanent Olympic host
```

```
                                              Top3  \
0                      Hughes' instant hit buoys Blues
1                          Leader: German sleaze inquiry
2              Thatcher issues defence before trial by video
3                     McIlroy calls for Irish fighting spirit
4                              Breast cancer screening
…                                                          …
4096  Poland 'shocked' by xenophobic abuse of Poles …
4097  S&amp;P cuts United Kingdom sovereign credit r…
4098  UK must accept freedom of movement to access E…
4099  Boris Johnson says he will not run for Tory pa…
4100  The president of France says if Brexit won, so…

                                              Top4  \
0                       Jack gets his skates on at ice-cold Alex
1                                            Cheerio, boyo
2               Police help Smith lay down the law at Everton
3                           Leicester bin stadium blueprint
4                                              Alan Parker
…                                                          …
4096  There will be no second referendum, cabinet ag…
4097                  Huge helium deposit found in Africa
4098  Devastated: scientists too late to captive bre…
4099  Six gay men in Ivory Coast were abused and for…
4100  British Man Who Must Give Police 24 Hours' Not…

                                              Top5  \
0                      Chaos as Maracana builds up for United
1                                 The main recommendations
2              Tale of Trautmann bears two more retellings
3                              United braced for Mexican wave
4                  Guardian readers: are you all whingers?
…                                                          …
4096       Scotland welcome to join EU, Merkel ally says
4097  CEO of the South African state broadcaster qui…
4098  British Labor Party leader Jeremy Corbyn loses…
4099  Switzerland denies citizenship to Muslim immig…
4100  100+ Nobel laureates urge Greenpeace to stop o…

                                              Top6  \
0     Depleted Leicester prevail as Elliott spoils E…
1                                  Has Cubie killed fees?
2                                     England on the rack
3     Auntie back in fashion, even if the dress look…
4                                        Hollywood Beyond
…                                                          …
```

```
4096   Sterling dips below Friday's 31-year low amid …
4097   Brexit cost investors $2 trillion, the worst o…
4098   A Muslim Shop in the UK Was Just Firebombed Wh…
4099   Palestinian terrorist stabs israeli teen girl …
4100   Brazil: Huge spike in number of police killing…

                                            Top7   \
0                      Hungry Spurs sense rich pickings
1                                  Has Cubie killed fees?
2          Pakistan retaliate with call for video of Walsh
3                            Shoaib appeal goes to the top
4                                     Ashes and diamonds
…                                                       …
4096   No negative news about South African President…
4097   Hong Kong democracy activists call for return …
4098   Mexican Authorities Sexually Torture Women in …
4099   Puerto Rico will default on $1 billion of debt…
4100   Austria's highest court annuls presidential el…

                                            Top8  …  \
0                      Gunners so wide of an easy target  …
1                                  Has Cubie killed fees?  …
2                      Cullinan continues his Cape monopoly  …
3          Hussain hurt by 'shambles' but lays blame on e…  …
4                           Whingers - a formidable minority  …
…                                                       …  …
4096   Surge in Hate Crimes in the U.K. Following U.K…  …
4097   Brexit: Iceland president says UK can join 'tr…  …
4098              UK shares and pound continue to recover  …
4099   Republic of Ireland fans to be awarded medal f…  …
4100   Facebook wins privacy case, can track any Belg…  …

                                            Top16  \
0              Flintoff injury piles on woe for England
1                                   On the critical list
2                              South Melbourne (Australia)
3          Putin admits Yeltsin quit to give him a head s…
4                                  Most everywhere:  UDIs
…                                                       …
4096   German lawyers to probe Erdogan over alleged w…
4097   US, Canada and Mexico pledge 50% of power from…
4098   Escape Tunnel, Dug by Hand, Is Found at Holoca…
4099   Googles free wifi at Indian railway stations i…
4100   The United States has placed Myanmar, Uzbekist…

                                            Top17  \
0      Hunters threaten Jospin with new battle of the…
```

```
1                                    The timing of their lives
2                                              Necaxa (Mexico)
3                     BBC worst hit as digital TV begins to bite
4                               Most wanted:  Chloe lunettes
…                                                            …
4096  Boris Johnson says the UK will continue to "in…
4097  There is increasing evidence that Australia is…
4098  The land under Beijing is sinking by as much a…
4099  Mounting evidence suggests 'hobbits' were wipe…
4100  S&amp;P revises European Union credit rating t…


                                                     Top18  \
0                     Kohl's successor drawn into scandal
1                                              Dear doctor
2                                       Real Madrid (Spain)
3                              How much can you pay for…
4        Return of the cane 'completely off the agenda'
…                                                            …
4096  Richard Branson is calling on the UK governmen…
4097  Richard Branson, the founder of Virgin Group, …
4098  Car bomb and Anti-Islamic attack on Mosque in …
4099  The men who carried out Tuesday's terror attac…
4100  India gets $1 billion loan from World Bank for…


                                                     Top19  \
0                     The difference between men and women
1     Irish court halts IRA man's extradition to Nor…
2                             Raja Casablanca (Morocco)
3                                      Christmas glitches
4                     From Sleepy Hollow to Greeneland
…                                                            …
4096             Turkey 'sorry for downing Russian jet'
4097  37,000-yr-old skull from Borneo reveals surpri…
4098  Emaciated lions in Taiz Zoo are trapped in blo…
4099  Calls to suspend Saudi Arabia from UN Human Ri…
4100  U.S. sailors detained by Iran spoke too much u…


                                                     Top20  \
0                     Sara Denver, nurse turned solicitor
1     Burundi peace initiative fades after rebels re…
2                                    Corinthians (Brazil)
3     Upending a table, Chopping a line and Scoring …
4                     Blunkett outlines vision for over 11s
…                                                            …
4096  Edward Snowden lawyer vows new push for pardon…
4097  Palestinians stone Western Wall worshipers; po…
4098  Rupert Murdoch describes Brexit as 'wonderful'…
```

```
4099   More Than 100 Nobel Laureates Call Out Greenpe…
4100   Mass fish kill in Vietnam solved as Taiwan ste…


                                                  Top21  \
0          Diana's landmine crusade put Tories in a panic
1                     PE points the way forward to the ECB
2                                        Tony's pet project
3       Scientific evidence 'unreliable', defence claims
4       Embattled Dobson attacks 'play now, pay later'…
…                                                        …
4096    Brexit opinion poll reveals majority don't wan…
4097    Jean-Claude Juncker asks Farage: Why are you h…
4098         More than 40 killed in Yemen suicide attacks
4099    British pedophile sentenced to 85 years in US …
4100    Philippines president Rodrigo Duterte urges pe…


                                                  Top22  \
0       Yeltsin's resignation caught opposition flat-f…
1       Campaigners keep up pressure on Nazi war crime…
2                                   Al Nassr (Saudi Arabia)
3         Fusco wins judicial review in extradition case
4                                        Doom and the Dome
…                                                        …
4096    Conservative MP Leave Campaigner: "The leave c…
4097    "Romanians for Remainians" offering a new home…
4098    Google Found Disastrous Symantec and Norton Vu…
4099    US permitted 1,200 offshore fracks in Gulf of …
4100    Spain arrests three Pakistanis accused of prom…


                                                  Top23  \
0                                        Russian roulette
1                                          Jane Ratcliffe
2                                        Ideal Holmes show
3                              Rebels thwart Russian advance
4                              What is the north-south divide?
…                                                        …
4096    Economists predict UK recession, further weake…
4097    Brexit: Gibraltar in talks with Scotland to st…
4098    Extremist violence on the rise in Germany: Dom…
4099    We will be swimming in ridicule - French beach…
4100    Venezuela, where anger over food shortages is …


                                                  Top24  \
0                                                Sold out
1       Yet more things you wouldn't know without the …
2                        Pinochet leaves hospital after tests
3                        Blair orders shake-up of failing NHS
```

```
4                              Aitken released from jail
…                                                      …
4096  New EU 'superstate plan by France, Germany: Cr…
4097                      8 Suicide Bombers Strike Lebanon
4098  BBC News: Labour MPs pass Corbyn no-confidence…
4099  UEFA says no minutes of silence for Istanbul v…
4100  A Hindu temple worker has been killed by three…


                                                    Top25
0                                      Recovering a title
1                                  Millennium bug fails to bite
2                                             Useful links
3                                  Lessons of law's hard heart
4                                               Gone aloft
…                                                      …
4096  Pakistani clerics declare transgender marriage…
4097  Mexico's security forces routinely use 'sexual…
4098  Tiny New Zealand town with 'too many jobs' lau…
4099  Law Enforcement Sources: Gun Used in Paris Ter…
4100  Ozone layer hole seems to be healing – US &amp…

[4101 rows x 27 columns]
```

[3]:
```python
# extra info
print(data.columns)
print(data.dtypes)
```

```
Index(['Date', 'Label', 'Top1', 'Top2', 'Top3', 'Top4', 'Top5', 'Top6', 'Top7',
       'Top8', 'Top9', 'Top10', 'Top11', 'Top12', 'Top13', 'Top14', 'Top15',
       'Top16', 'Top17', 'Top18', 'Top19', 'Top20', 'Top21', 'Top22', 'Top23',
       'Top24', 'Top25'],
      dtype='object')
Date     object
Label     int64
Top1     object
Top2     object
Top3     object
Top4     object
Top5     object
Top6     object
Top7     object
Top8     object
Top9     object
Top10    object
Top11    object
Top12    object
Top13    object
Top14    object
```

```
Top15    object
Top16    object
Top17    object
Top18    object
Top19    object
Top20    object
Top21    object
Top22    object
Top23    object
Top24    object
Top25    object
dtype: object
```

[4]:
```
# total missing data
data.isna().sum()
```

[4]:
```
Date     0
Label    0
Top1     0
Top2     0
Top3     0
Top4     0
Top5     0
Top6     0
Top7     0
Top8     0
Top9     0
Top10    0
Top11    0
Top12    0
Top13    0
Top14    0
Top15    0
Top16    0
Top17    0
Top18    0
Top19    0
Top20    0
Top21    0
Top22    0
Top23    1
Top24    3
Top25    3
dtype: int64
```

[5]:
```
# drop missing data
data.dropna(inplace=True)
data.reset_index(drop = True, inplace = True)
```

```
[6]: # check again
     data.isna().sum()
```

```
[6]: Date      0
     Label     0
     Top1      0
     Top2      0
     Top3      0
     Top4      0
     Top5      0
     Top6      0
     Top7      0
     Top8      0
     Top9      0
     Top10     0
     Top11     0
     Top12     0
     Top13     0
     Top14     0
     Top15     0
     Top16     0
     Top17     0
     Top18     0
     Top19     0
     Top20     0
     Top21     0
     Top22     0
     Top23     0
     Top24     0
     Top25     0
     dtype: int64
```

```
[7]: # spliting data in train and test
     train = data[data['Date'] < '20150101']
     test = data[data['Date'] > '20141231']
```

```
[8]: # dropping date and label since it is not useful
     data = train.drop(columns = ['Date','Label'], axis = 0)
     data.head()
```

```
[8]:                                                  Top1  \
     0  A 'hindrance to operations': extracts from the…
     1                                       Scorecard
     2                Coventry caught on counter by Flo
     3                     Pilgrim knows how to progress
     4                             Hitches and Horlocks
```

```
                                    Top2  \
0                             Scorecard
1                    The best lake scene
2     United's rivals on the road to Rio
3                    Thatcher facing ban
4        Beckham off but United survive


                                         Top3  \
0               Hughes' instant hit buoys Blues
1                   Leader: German sleaze inquiry
2     Thatcher issues defence before trial by video
3           McIlroy calls for Irish fighting spirit
4                       Breast cancer screening


                                         Top4  \
0         Jack gets his skates on at ice-cold Alex
1                               Cheerio, boyo
2     Police help Smith lay down the law at Everton
3                 Leicester bin stadium blueprint
4                                 Alan Parker


                                         Top5  \
0         Chaos as Maracana builds up for United
1                     The main recommendations
2     Tale of Trautmann bears two more retellings
3               United braced for Mexican wave
4       Guardian readers: are you all whingers?


                                              Top6  \
0     Depleted Leicester prevail as Elliott spoils E…
1                          Has Cubie killed fees?
2                           England on the rack
3     Auntie back in fashion, even if the dress look…
4                          Hollywood Beyond


                                         Top7  \
0                 Hungry Spurs sense rich pickings
1                       Has Cubie killed fees?
2     Pakistan retaliate with call for video of Walsh
3                 Shoaib appeal goes to the top
4                          Ashes and diamonds


                                         Top8  \
0                 Gunners so wide of an easy target
1                       Has Cubie killed fees?
2                 Cullinan continues his Cape monopoly
3     Hussain hurt by 'shambles' but lays blame on e…
```

```
4                  Whingers - a formidable minority

                                            Top9  \
0        Derby raise a glass to Strupar's debut double
1        Hopkins 'furious' at Foster's lack of Hannibal…
2                    McGrath puts India out of their misery
3                         England's decade of disasters
4                           Alan Parker - part two


                                            Top10  …  \
0  Southgate strikes, Leeds pay the penalty  …
1                        Has Cubie killed fees?  …
2            Blair Witch bandwagon rolls on  …
3      Revenge is sweet for jubilant Cronje  …
4                 Thuggery, Toxins and Ties  …


                                            Top16  \
0          Flintoff injury piles on woe for England
1                            On the critical list
2                         South Melbourne (Australia)
3  Putin admits Yeltsin quit to give him a head s…
4                           Most everywhere:  UDIs


                                            Top17  \
0  Hunters threaten Jospin with new battle of the…
1                        The timing of their lives
2                             Necaxa (Mexico)
3      BBC worst hit as digital TV begins to bite
4                 Most wanted:  Chloe lunettes


                                            Top18  \
0          Kohl's successor drawn into scandal
1                                 Dear doctor
2                        Real Madrid (Spain)
3                 How much can you pay for…
4  Return of the cane 'completely off the agenda'


                                            Top19  \
0           The difference between men and women
1  Irish court halts IRA man's extradition to Nor…
2                     Raja Casablanca (Morocco)
3                        Christmas glitches
4              From Sleepy Hollow to Greeneland


                                            Top20  \
0          Sara Denver, nurse turned solicitor
1  Burundi peace initiative fades after rebels re…
```

```
2                              Corinthians (Brazil)
3      Upending a table, Chopping a line and Scoring …
4                   Blunkett outlines vision for over 11s


                                               Top21  \
0          Diana's landmine crusade put Tories in a panic
1                     PE points the way forward to the ECB
2                                        Tony's pet project
3        Scientific evidence 'unreliable', defence claims
4      Embattled Dobson attacks 'play now, pay later'…


                                               Top22  \
0      Yeltsin's resignation caught opposition flat-f…
1      Campaigners keep up pressure on Nazi war crime…
2                               Al Nassr (Saudi Arabia)
3        Fusco wins judicial review in extradition case
4                                    Doom and the Dome


                        Top23  \
0              Russian roulette
1                Jane Ratcliffe
2              Ideal Holmes show
3      Rebels thwart Russian advance
4      What is the north-south divide?


                                          Top24  \
0                                        Sold out
1      Yet more things you wouldn't know without the …
2              Pinochet leaves hospital after tests
3              Blair orders shake-up of failing NHS
4                        Aitken released from jail


                        Top25
0          Recovering a title
1      Millennium bug fails to bite
2                  Useful links
3        Lessons of law's hard heart
4                    Gone aloft

[5 rows x 25 columns]
```

```python
data.replace('[^a-zA-Z]',' ', inplace = True)
data
```

```
                                           Top1  \
0          A 'hindrance to operations': extracts from the…
1                                            Scorecard
```

```
2                    Coventry caught on counter by Flo
3                       Pilgrim knows how to progress
4                                 Hitches and Horlocks
…                                                     …
3967  Brazil declares emergency after 2,400 babies a…
3968  Germany recruits 8,500 teachers to teach Germa…
3969  Pentagon thwarts Obama's effort to close Guant…
3970  Tech companies face criminal charges if they n…
3971  Microsoft Corp said on Wednesday it will begin…


                                                 Top2  \
0                                            Scorecard
1                                    The best lake scene
2                           United's rivals on the road to Rio
3                                      Thatcher facing ban
4                           Beckham off but United survive
…                                                     …
3967  Scientists find big 1500-yr-old Viking settlem…
3968  Almost 23,000 people have signed an Italian po…
3969              Rare Footage of Giant Squid in Japan
3970  Mexican marijuana farmers see profits tumble a…
3971  Greek Orthodox Church: Homosexuals are God's c…


                                                 Top3  \
0                            Hughes' instant hit buoys Blues
1                              Leader: German sleaze inquiry
2            Thatcher issues defence before trial by video
3                    McIlroy calls for Irish fighting spirit
4                                Breast cancer screening
…                                                     …
3967  Paris attacks: Belgian police arrest ninth sus…
3968  Japan apologizes to Korea over sexual enslavem…
3969  Muslims tell ISIS they'd rather see Star Wars …
3970  A would-be suicide bomber and his wife have be…
3971  Vladmir Putin officially dissolves Roscosmos, …


                                                 Top4  \
0              Jack gets his skates on at ice-cold Alex
1                                         Cheerio, boyo
2            Police help Smith lay down the law at Everton
3                           Leicester bin stadium blueprint
4                                          Alan Parker
…                                                     …
3967  Wiretapping reveals communication between Turk…
3968  Milan and Rome ban cars as pollution rises. Mi…
3969  Syrian Journalist Who Documented ISIS Atrociti…
3970  The rich will get richer while two million mor…
```

```
3971  Donald Trump could be refused entry to the UK …

                                                   Top5  \
0                   Chaos as Maracana builds up for United
1                             The main recommendations
2            Tale of Trautmann bears two more retellings
3                         United braced for Mexican wave
4                Guardian readers: are you all whingers?
…                                                       …
3967  Russia airstrikes: Hundreds of Syrian civilans…
3968           Saudi posts record $98 bn deficit in 2015
3969  North Korea's second most-powerful man sent to…
3970  The North Pole will have above freezing temper…
3971  Saudi Arabia says its ready to meet any additi…

                                                   Top6  \
0       Depleted Leicester prevail as Elliott spoils E…
1                                   Has Cubie killed fees?
2                                     England on the rack
3       Auntie back in fashion, even if the dress look…
4                                        Hollywood Beyond
…                                                       …
3967  Pakistan changes neutral position on Syrian Ci…
3968  The United States, Russia, Iran, and Saudi Ara…
3969  Austria Turns Away Hundreds of Migrants for Ly…
3970  South Australia – New sexting laws to spare ch…
3971  New Year fireworks and festivities have been c…

                                                   Top7  \
0                          Hungry Spurs sense rich pickings
1                                   Has Cubie killed fees?
2          Pakistan retaliate with call for video of Walsh
3                             Shoaib appeal goes to the top
4                                      Ashes and diamonds
…                                                       …
3967  North Korean diplomat arrested in South Africa…
3968  Vladimir Putin has achieved his central goal o…
3969  Saudi Grand Mufti calls ISIS part of the Israe…
3970  ISIS militants burn 20 media activists to deat…
3971  New species of human may have shared our caves…

                                                   Top8  \
0                          Gunners so wide of an easy target
1                                   Has Cubie killed fees?
2                          Cullinan continues his Cape monopoly
3       Hussain hurt by 'shambles' but lays blame on e…
4                          Whingers – a formidable minority
```

```
…                                                       …
3967      German teachers want 'Mein Kampf' on syllabus
3968  Noted Syrian journalist who uncovered ISIS Ale…
3969  'What has war yielded, neither land nor heaven…
3970       Italy fines Apple over misleading tax conduct
3971  Egypt becomes second nation to shut down Faceb…

                                              Top9  \
0          Derby raise a glass to Strupar's debut double
1     Hopkins 'furious' at Foster's lack of Hannibal…
2                  McGrath puts India out of their misery
3                          England's decade of disasters
4                               Alan Parker - part two
…                                                       …
3967  North Korea slave force earns Kim Jong Un regi…
3968  Criminal gangs are contaminating democracy in …
3969  Ancient arch attacked by ISIS to be rebuilt in…
3970  Body of notorious Mexican drug cartel founder …
3971  Polish government takes control of public medi…

                                             Top10  …  \
0          Southgate strikes, Leeds pay the penalty  …
1                            Has Cubie killed fees?  …
2                    Blair Witch bandwagon rolls on  …
3                 Revenge is sweet for jubilant Cronje  …
4                         Thuggery, Toxins and Ties  …
…                                                   …  …
3967  50,000 Yemeni Forces Preparing to Attack Saudi…  …
3968           ISIS leader threatens to strike Israel  …
3969  Islamic State theologians have issued an extre…  …
3970  Morocco to deploy 100,000 armed officers to pr…  …
3971  From Jan 1, UK to stop financial aid to India …  …

                                             Top16  \
0          Flintoff injury piles on woe for England
1                               On the critical list
2                         South Melbourne (Australia)
3     Putin admits Yeltsin quit to give him a head s…
4                            Most everywhere:  UDIs
…                                                       …
3967  Report on sexual exploitation and abuse by pea…
3968  Saudi Arabia Plans Subsidy Cuts as King Unveil…
3969  Iran Hands Over Stockpile of Enriched Uranium …
3970  Norway to turn back asylum seekers on Swedish …
3971  Ramadi residents fleeing ISIS: 'They wanted to…

                                             Top17  \
```

```
0      Hunters threaten Jospin with new battle of the…
1                                 The timing of their lives
2                                          Necaxa (Mexico)
3              BBC worst hit as digital TV begins to bite
4                          Most wanted:  Chloe lunettes
…                                                      …
3967   Baby girl dies after X-president's son's staff…
3968   Syria anti-ISIS documentary maker 'assassinate…
3969    Registrar Shuts Down All Pirate Bay Domain Names
3970   Islamic State Suffers Double Blow As Ramadi Fa…
3971   Israeli Gov't approves NIS 15-billion plan to …


                                              Top18  \
0                    Kohl's successor drawn into scandal
1                                           Dear doctor
2                                     Real Madrid (Spain)
3                             How much can you pay for…
4          Return of the cane 'completely off the agenda'
…                                                      …
3967           China smog sparks red alerts in 10 cities
3968   Denmark wants Geneva Convention debate if Euro…
3969   The Irish Government has taken unprecedented a…
3970   Journalist reveals Erdoan imprisoned him for e…
3971   Russia denies 'absurd' claims that air strikes…


                                              Top19  \
0                    The difference between men and women
1      Irish court halts IRA man's extradition to Nor…
2                              Raja Casablanca (Morocco)
3                                     Christmas glitches
4                       From Sleepy Hollow to Greeneland
…                                                      …
3967   US planned East Berlin's 'systematic destructi…
3968   David Cameron is heckled during visit to flood…
3969   Saudi Arabia to raise domestic petrol prices b…
3970   Russian Moon base plan falls victim to budget …
3971           Web attack knocks BBC websites offline


                                              Top20  \
0                    Sara Denver, nurse turned solicitor
1      Burundi peace initiative fades after rebels re…
2                                    Corinthians (Brazil)
3      Upending a table, Chopping a line and Scoring …
4                    Blunkett outlines vision for over 11s
…                                                      …
3967   Qatar World Cup 2022 workers 'earning as littl…
3968   British Army Is Deployed as Flooding Submerges…
```

```
3969  Saudi Arabia hikes petrol prices by 40% at the…
3970  Belgium, Netherlands to exchange territory  wi…
3971   Iran says any U.S. sanctions on missiles illegal


                                              Top21  \
0         Diana's landmine crusade put Tories in a panic
1                        PE points the way forward to the ECB
2                                        Tony's pet project
3       Scientific evidence 'unreliable', defence claims
4       Embattled Dobson attacks 'play now, pay later'…
…                                                      …
3967  Thai Court Sentences Two to Death for Killings…
3968  Brazils reluctance to accept an Israeli ambass…
3969  World Health Organization declares Guinea Ebol…
3970  North Korea says top official Kim Yang-gon kil…
3971  Maternity Benefits: Indian Government plans to…


                                              Top22  \
0         Yeltsin's resignation caught opposition flat-f…
1         Campaigners keep up pressure on Nazi war crime…
2                                        Al Nassr (Saudi Arabia)
3         Fusco wins judicial review in extradition case
4                                        Doom and the Dome
…                                                      …
3967  Germany Pays to Halt Danish Wind Power to Prot…
3968  Russian GDP down 4.0 percent year-on-year in N…
3969  Hacker group Anonymous is helping the 'Making …
3970  Putin's Son-in-Law Boosted by $1.75 Bln Russia…
3971  'Rat on a plane' forces Air India flight to tu…


                                              Top23  \
0                                        Russian roulette
1                                          Jane Ratcliffe
2                                        Ideal Holmes show
3                               Rebels thwart Russian advance
4                               What is the north-south divide?
…                                                      …
3967  Egypts President Sisi Urges Islamic Scholars t…
3968  Syria and Iraq: Ethnic cleansing by Sunni and …
3969           Russia 'has removed uranium from Iran'
3970     An orgasm a day could cut prostate cancer risk
3971  Philippines to join China-backed AIIB infrastr…


                                              Top24  \
0                                          Sold out
1       Yet more things you wouldn't know without the …
2                        Pinochet leaves hospital after tests
```

```
3                    Blair orders shake-up of failing NHS
4                          Aitken released from jail
…                                             …
3967  Russia Rejects Amnesty Internationals Report o…
3968      Deadly car bomb explodes at Kabul airport
3969  China threatens to shoot down Australian aircr…
3970  Germans claim Belgian nuclear reactors are "fa…
3971  ESPN invites five FIFA presidential candidates…


                                              Top25
0                              Recovering a title
1                         Millennium bug fails to bite
2                                     Useful links
3                         Lessons of law's hard heart
4                                       Gone aloft
…                                             …
3967  Cameron, Osborne and Murdoch back together at …
3968  Experiment in Germany tests the idea of an 'em…
3969  Israel warns Brazil to accept its settler amba…
3970  UK Sperm Bank Doesn't Accept Donors with Dysle…
3971  Hamas to keep Palestinians from ringing in New…

[3972 rows x 25 columns]
```

```python
new_index = [str(i) for i in range(25)]
new_index
```

```
['0',
 '1',
 '2',
 '3',
 '4',
 '5',
 '6',
 '7',
 '8',
 '9',
 '10',
 '11',
 '12',
 '13',
 '14',
 '15',
 '16',
 '17',
 '18',
 '19',
```

```
      '20',
      '21',
      '22',
      '23',
      '24']
```

[11]:
```
data.columns = new_index
data.head()
```

[11]:
```
                                              0  \
0  A 'hindrance to operations': extracts from the…
1                                      Scorecard
2                 Coventry caught on counter by Flo
3                      Pilgrim knows how to progress
4                              Hitches and Horlocks


                               1  \
0                       Scorecard
1                 The best lake scene
2  United's rivals on the road to Rio
3                 Thatcher facing ban
4     Beckham off but United survive


                                         2  \
0                 Hughes' instant hit buoys Blues
1                   Leader: German sleaze inquiry
2  Thatcher issues defence before trial by video
3        McIlroy calls for Irish fighting spirit
4                          Breast cancer screening


                                        3  \
0        Jack gets his skates on at ice-cold Alex
1                              Cheerio, boyo
2  Police help Smith lay down the law at Everton
3                 Leicester bin stadium blueprint
4                                    Alan Parker


                                     4  \
0        Chaos as Maracana builds up for United
1                     The main recommendations
2  Tale of Trautmann bears two more retellings
3                 United braced for Mexican wave
4     Guardian readers: are you all whingers?


                                          5  \
0  Depleted Leicester prevail as Elliott spoils E…
1                          Has Cubie killed fees?
```

18

```
2                             England on the rack
3    Auntie back in fashion, even if the dress look…
4                               Hollywood Beyond


                                      6   \
0            Hungry Spurs sense rich pickings
1                         Has Cubie killed fees?
2    Pakistan retaliate with call for video of Walsh
3                     Shoaib appeal goes to the top
4                             Ashes and diamonds


                                      7   \
0            Gunners so wide of an easy target
1                         Has Cubie killed fees?
2             Cullinan continues his Cape monopoly
3    Hussain hurt by 'shambles' but lays blame on e…
4                   Whingers - a formidable minority


                                      8   \
0       Derby raise a glass to Strupar's debut double
1    Hopkins 'furious' at Foster's lack of Hannibal…
2               McGrath puts India out of their misery
3                     England's decade of disasters
4                         Alan Parker - part two


                                      9   …   \
0    Southgate strikes, Leeds pay the penalty   …
1                         Has Cubie killed fees?   …
2             Blair Witch bandwagon rolls on   …
3       Revenge is sweet for jubilant Cronje   …
4                   Thuggery, Toxins and Ties   …


                                     15   \
0            Flintoff injury piles on woe for England
1                             On the critical list
2                      South Melbourne (Australia)
3    Putin admits Yeltsin quit to give him a head s…
4                         Most everywhere:  UDIs


                                     16   \
0    Hunters threaten Jospin with new battle of the…
1                         The timing of their lives
2                               Necaxa (Mexico)
3       BBC worst hit as digital TV begins to bite
4                     Most wanted:  Chloe lunettes


                                     17   \
```

```
0           Kohl's successor drawn into scandal
1                                    Dear doctor
2                          Real Madrid (Spain)
3                     How much can you pay for…
4   Return of the cane 'completely off the agenda'


                                           18  \
0              The difference between men and women
1   Irish court halts IRA man's extradition to Nor…
2                        Raja Casablanca (Morocco)
3                               Christmas glitches
4                  From Sleepy Hollow to Greeneland


                                           19  \
0              Sara Denver, nurse turned solicitor
1   Burundi peace initiative fades after rebels re…
2                            Corinthians (Brazil)
3   Upending a table, Chopping a line and Scoring …
4               Blunkett outlines vision for over 11s


                                           20  \
0      Diana's landmine crusade put Tories in a panic
1               PE points the way forward to the ECB
2                               Tony's pet project
3    Scientific evidence 'unreliable', defence claims
4   Embattled Dobson attacks 'play now, pay later'…


                                           21  \
0   Yeltsin's resignation caught opposition flat-f…
1   Campaigners keep up pressure on Nazi war crime…
2                          Al Nassr (Saudi Arabia)
3     Fusco wins judicial review in extradition case
4                              Doom and the Dome


                        22  \
0                 Russian roulette
1                    Jane Ratcliffe
2                  Ideal Holmes show
3     Rebels thwart Russian advance
4   What is the north-south divide?


                                           23  \
0                                         Sold out
1   Yet more things you wouldn't know without the …
2                  Pinochet leaves hospital after tests
3               Blair orders shake-up of failing NHS
4                          Aitken released from jail
```

```
                                24
0               Recovering a title
1       Millennium bug fails to bite
2                      Useful links
3        Lessons of law's hard heart
4                         Gone aloft

[5 rows x 25 columns]
```

[12]: `data.index`

[12]:
```
Int64Index([    0,    1,    2,    3,    4,    5,    6,    7,    8,    9,
            …
            3962, 3963, 3964, 3965, 3966, 3967, 3968, 3969, 3970, 3971],
           dtype='int64', length=3972)
```

[13]:
```python
for index in new_index:
    data[index] = data[index].str.lower()
data.head(1)
```

[13]:
```
                                               0          1  \
0  a 'hindrance to operations': extracts from the…   scorecard


                               2                              3  \
0  hughes' instant hit buoys blues  jack gets his skates on at ice-cold alex


                              4  \
0  chaos as maracana builds up for united


                                5  \
0  depleted leicester prevail as elliott spoils e…


                            6                             7  \
0  hungry spurs sense rich pickings  gunners so wide of an easy target


                              8  \
0  derby raise a glass to strupar's debut double


                             9  …  \
0  southgate strikes, leeds pay the penalty   …


                             15  \
0  flintoff injury piles on woe for england


                                16  \
0  hunters threaten jospin with new battle of the…
```

```
                                         17                                          18  \
0   kohl's successor drawn into scandal   the difference between men and women

                                      19  \
0   sara denver, nurse turned solicitor

                                         20  \
0   diana's landmine crusade put tories in a panic

                                             21                        22  \
0   yeltsin's resignation caught opposition flat-f…   russian roulette

              23                    24
0   sold out   recovering a title

[1 rows x 25 columns]
```

```python
headlines = []
for i in data.iloc[1,0:25]:
    headlines.append(i)
' '.join(headlines)
```

[14]: "scorecard the best lake scene leader: german sleaze inquiry cheerio, boyo the
      main recommendations has cubie killed fees? has cubie killed fees? has cubie
      killed fees? hopkins 'furious' at foster's lack of hannibal appetite has cubie
      killed fees? a tale of two tails i say what i like and i like what i say elbows,
      eyes and nipples task force to assess risk of asteroid collision how i found
      myself at last on the critical list the timing of their lives dear doctor irish
      court halts ira man's extradition to northern ireland burundi peace initiative
      fades after rebels reject mandela as mediator pe points the way forward to the
      ecb campaigners keep up pressure on nazi war crimes suspect jane ratcliffe yet
      more things you wouldn't know without the movies millennium bug fails to bite"

```python
# list comprehension same output but diff command with compariosion to upper
  code
' '.join([str(i) for i in data.iloc[1,0:25]])
```

[15]: "scorecard the best lake scene leader: german sleaze inquiry cheerio, boyo the
      main recommendations has cubie killed fees? has cubie killed fees? has cubie
      killed fees? hopkins 'furious' at foster's lack of hannibal appetite has cubie
      killed fees? a tale of two tails i say what i like and i like what i say elbows,
      eyes and nipples task force to assess risk of asteroid collision how i found
      myself at last on the critical list the timing of their lives dear doctor irish
      court halts ira man's extradition to northern ireland burundi peace initiative
      fades after rebels reject mandela as mediator pe points the way forward to the
      ecb campaigners keep up pressure on nazi war crimes suspect jane ratcliffe yet

more things you wouldn't know without the movies millennium bug fails to bite"

```python
headlines = []
for row in range(0,len(data)):
    headlines.append(' '.join([str(i) for i in data.iloc[row,0:25]]))
```

```python
headlines[0:3]
```

```
["a 'hindrance to operations': extracts from the leaked reports scorecard
hughes' instant hit buoys blues jack gets his skates on at ice-cold alex chaos
as maracana builds up for united depleted leicester prevail as elliott spoils
everton's party hungry spurs sense rich pickings gunners so wide of an easy
target derby raise a glass to strupar's debut double southgate strikes, leeds
pay the penalty hammers hand robson a youthful lesson saints party like it's
1999 wear wolves have turned into lambs stump mike catches testy gough's taunt
langer escapes to hit 167 flintoff injury piles on woe for england hunters
threaten jospin with new battle of the somme kohl's successor drawn into scandal
the difference between men and women sara denver, nurse turned solicitor diana's
landmine crusade put tories in a panic yeltsin's resignation caught opposition
flat-footed russian roulette sold out recovering a title",
 "scorecard the best lake scene leader: german sleaze inquiry cheerio, boyo the
main recommendations has cubie killed fees? has cubie killed fees? has cubie
killed fees? hopkins 'furious' at foster's lack of hannibal appetite has cubie
killed fees? a tale of two tails i say what i like and i like what i say elbows,
eyes and nipples task force to assess risk of asteroid collision how i found
myself at last on the critical list the timing of their lives dear doctor irish
court halts ira man's extradition to northern ireland burundi peace initiative
fades after rebels reject mandela as mediator pe points the way forward to the
ecb campaigners keep up pressure on nazi war crimes suspect jane ratcliffe yet
more things you wouldn't know without the movies millennium bug fails to bite",
 "coventry caught on counter by flo united's rivals on the road to rio thatcher
issues defence before trial by video police help smith lay down the law at
everton tale of trautmann bears two more retellings england on the rack pakistan
retaliate with call for video of walsh cullinan continues his cape monopoly
mcgrath puts india out of their misery blair witch bandwagon rolls on pele turns
up heat on ferguson party divided over kohl slush fund scandal manchester united
(england) women in record south pole walk vasco da gama (brazil) south melbourne
(australia) necaxa (mexico) real madrid (spain) raja casablanca (morocco)
corinthians (brazil) tony's pet project al nassr (saudi arabia) ideal holmes
show pinochet leaves hospital after tests useful links"]
```

```python
# implement bag of words
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.ensemble import RandomForestClassifier
```

```
[19]: cv = CountVectorizer(ngram_range=(2,2))
      traindata_x = cv.fit_transform(headlines)
```

```
[20]: # build model and implement
      classifier = RandomForestClassifier(n_estimators= 200, criterion= 'entropy')
      classifier.fit(traindata_x, train['Label'])
```

```
[20]: RandomForestClassifier(criterion='entropy', n_estimators=200)
```

```
[21]: # bring the test data together
      test_transform = []
      for row in range(0, len(test)):
          test_transform.append(' '.join(str(i) for i in test.iloc[row,2:27]))
```

```
[22]: # transform the data
      test_data = cv.transform(test_transform)
```

```
[23]: predictions = classifier.predict(test_data)
```

```
[24]: from sklearn.metrics import confusion_matrix
```

```
[25]: cm = confusion_matrix(test['Label'], predictions)
      cm
```
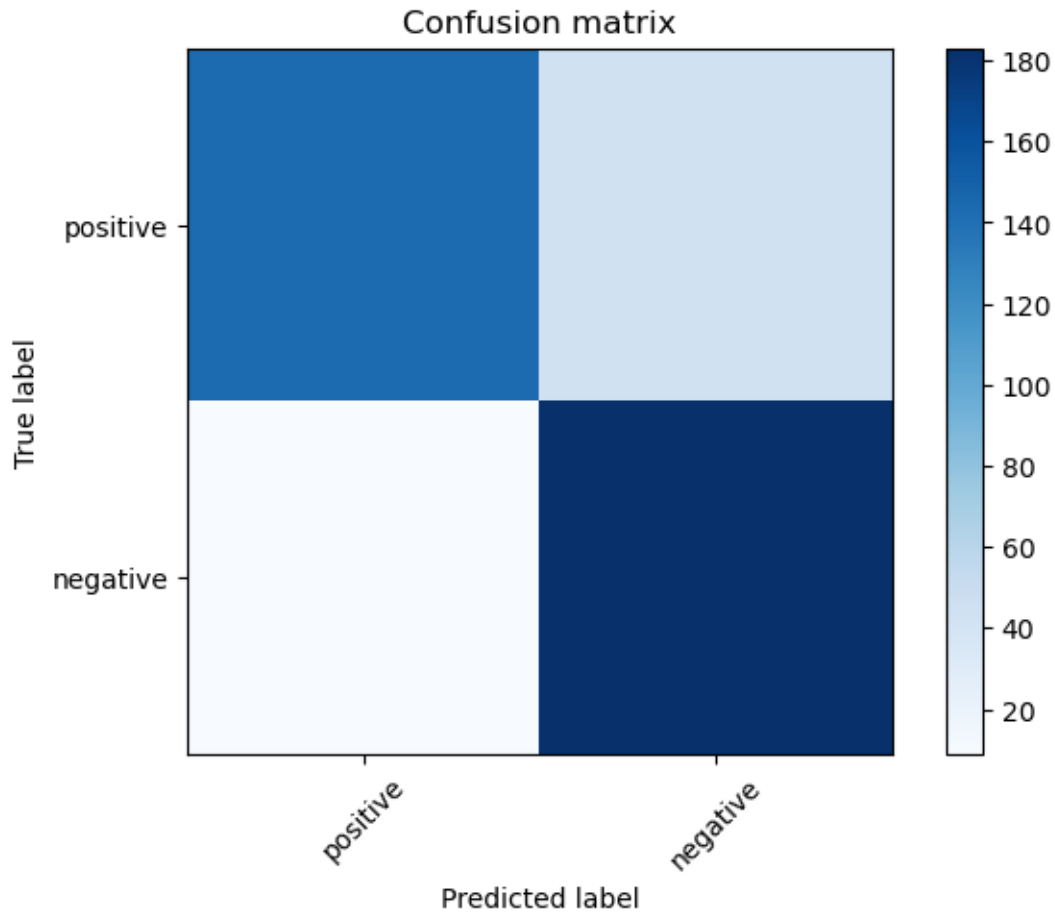
```
[25]: array([[143,  43],
             [  9, 183]], dtype=int64)
```

```
[26]: # plot
      from sklearn import metrics
      import matplotlib.pyplot as plt
      def plot_confusion_matrix(cm, title='Confusion matrix', cmap=plt.cm.Blues,␣
       ↪labels=["positive", "negative"]):
          plt.imshow(cm, interpolation='nearest', cmap=cmap)
          plt.title(title)
          plt.colorbar()
          tick_marks = np.arange(len(labels))
          plt.xticks(tick_marks, labels, rotation=45)
          plt.yticks(tick_marks, labels)
          plt.tight_layout()
          plt.ylabel('True label')
          plt.xlabel('Predicted label')
```

```
[27]: plt.figure()
      cm=confusion_matrix(test['Label'],predictions)
      print(cm)
      plot_confusion_matrix(cm)
      plt.show()
```

```
[[143  43]
 [  9 183]]
```

Confusion matrix

[28]:
```python
from sklearn.metrics import classification_report, confusion_matrix,
 ↪accuracy_score
```

[33]:
```python
acc = accuracy_score(test['Label'], predictions)
print(acc)
report = classification_report(test['Label'], predictions)
print(report)
```

```
0.8624338624338624
              precision    recall  f1-score   support

           0       0.94      0.77      0.85       186
           1       0.81      0.95      0.88       192

    accuracy                           0.86       378
   macro avg       0.88      0.86      0.86       378
```

```
weighted avg       0.87       0.86       0.86          378
```

[34]:
```python
from sklearn.naive_bayes import MultinomialNB
```

[36]:
```python
nb = MultinomialNB()
nb.fit(traindata_x, train['Label'])

predictions = nb.predict(test_data)
matrix = confusion_matrix(test['Label'], predictions)
print(matrix)
acc = accuracy_score(test['Label'], predictions)
print(acc)
report = classification_report(test['Label'], predictions)
print(report)
```

```
[[137  49]
 [  9 183]]
0.8465608465608465
              precision    recall  f1-score   support

           0       0.94      0.74      0.83       186
           1       0.79      0.95      0.86       192

    accuracy                           0.85       378
   macro avg       0.86      0.84      0.84       378
weighted avg       0.86      0.85      0.84       378
```

[ ]: