

M5L5 Solutions

May 1, 2023

Exercise 11

For the given two datasets, construct a combined dataframe with relevant features including subjectivity and polarity. Use Valence Aware Dictionary and Sentiment Reasoner lexicon tool to perform sentiment analysis using Linear Discriminant Analysis.

Solution

```
[1]: # Import required libraries
import re
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Import sklearn & classifier
from sklearn.model_selection import train_test_split
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.metrics import classification_report, RocCurveDisplay

# Import text and sentiment libraries
from textblob import TextBlob
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
```

```
[2]: # Read DJIA news
df1 = pd.read_csv('data/djia_news.csv')

# Read DJIA prices, tab separated
df2 = pd.read_csv('data/djia_prices.csv', sep='\t')
```

```
[3]: # Check the dataframe
df1.head(3)
```

```
[3]:
```

	Date	Label	Top1 \
0	2008-08-08	0	b"Georgia 'downs two Russian warplanes' as cou...
1	2008-08-11	1	b'Why wont America and Nato help us? If they w...
2	2008-08-12	0	b'Remember that adorable 9-year-old who sang a...

	Top2 \
0	b'BREAKING: Musharraf to be impeached.'

1 b'Bush puts foot down on Georgian conflict'
2 b"Russia 'ends Georgia operation'"

Top3 \

0 b'Russia Today: Columns of troops roll into So...
1 b"Jewish Georgian minister: Thanks to Israeli ...
2 b'"If we had no sexual harassment we would hav...

Top4 \

0 b'Russian tanks are moving towards the capital...
1 b'Georgian army flees in disarray as Russians ...
2 b"Al-Qa'eda is losing support in Iraq because ...

Top5 \

0 b"Afghan children raped with 'impunity,' U.N. ...
1 b"Olympic opening ceremony fireworks 'faked'"
2 b'Ceasefire in Georgia: Putin Outmaneuvers the...

Top6 \

0 b'150 Russian tanks have entered South Ossetia...
1 b'What were the Mossad with fraudulent New Zea...
2 b'Why Microsoft and Intel tried to kill the X0...

Top7 \

0 b"Breaking: Georgia invades South Ossetia, Rus...
1 b'Russia angered by Israeli military sale to G...
2 b'Stratfor: The Russo-Georgian War and the Bal...

Top8 ... \

0 b"The 'enemy combatent' trials are nothing but... ...
1 b'An American citizen living in S.Ossetia blam... ...
2 b'I'm Trying to Get a Sense of This Whole Geor... ...

Top16 \

0 b'Georgia Invades South Ossetia - if Russia ge...
1 b'Israel and the US behind the Georgian aggres...
2 b'U.S. troops still in Georgia (did you know t...

Top17 \

0 b'Al-Qaeda Faces Islamist Backlash'
1 b'"Do not believe TV, neither Russian nor Geor...
2 b'Why Russias response to Georgia was right'

Top18 \

0 b'Condoleezza Rice: "The US would not act to p...
1 b'Riots are still going on in Montreal (Canada...
2 b'Gorbachev accuses U.S. of making a "serious ...

```

Top19 \
0 b'This is a busy day: The European Union has ...
1 b'China to overtake US as largest manufacturer'
2 b'Russia, Georgia, and NATO: Cold War Two'

```

```

Top20 \
0 b"Georgia will withdraw 1,000 soldiers from Ir...
1 b'War in South Ossetia [PICS]'
2 b'Remember that adorable 62-year-old who led y...

```

```

Top21 \
0 b'Why the Pentagon Thinks Attacking Iran is a ...
1 b'Israeli Physicians Group Condemns State Tort...
2 b'War in Georgia: The Israeli connection'

```

```

Top22 \
0 b'Caucasus in crisis: Georgia invades South Os...
1 b' Russia has just beaten the United States ov...
2 b'All signs point to the US encouraging Georgi...

```

```

Top23 \
0 b'Indian shoe manufactory - And again in a se...
1 b'Perhaps *the* question about the Georgia - R...
2 b'Christopher King argues that the US and NATO...

```

```

Top24 \
0 b'Visitors Suffering from Mental Illnesses Ban...
1 b'Russia is so much better at war'
2 b'America: The New Mexico?'

```

```

Top25
0 b"No Help for Mexico's Kidnapping Surge"
1 b"So this is what it's come to: trading sex fo...
2 b"BBC NEWS | Asia-Pacific | Extinction 'by man...

```

[3 rows x 27 columns]

```
[4]: # Check the dataframe
df2.head()
```

```

[4]:      Date      Open      High      Low      Close \
0  2008-08-08  11432.089844  11759.959961  11388.040039  11734.320312
1  2008-08-11  11729.669922  11867.110352  11675.530273  11782.349609
2  2008-08-12  11781.700195  11782.349609  11601.519531  11642.469727
3  2008-08-13  11632.809570  11633.780273  11453.339844  11532.959961
4  2008-08-14  11532.070312  11718.280273  11450.889648  11615.929688

```

	Volume	Adj Close
0	212830000	11734.320312
1	183190000	11782.349609
2	173590000	11642.469727
3	182550000	11532.959961
4	159790000	11615.929688

```
[5]: df2.tail()
```

```
[5]:
```

	Date	Open	High	Low	Close \
1984	2016-06-27	17355.210938	17355.210938	17063.080078	17140.240234
1985	2016-06-28	17190.509766	17409.720703	17190.509766	17409.720703
1986	2016-06-29	17456.019531	17704.509766	17456.019531	17694.679688
1987	2016-06-30	17712.759766	17930.609375	17711.800781	17929.990234
1988	2016-07-01	17924.240234	18002.380859	17916.910156	17949.369141

	Volume	Adj Close
1984	138740000	17140.240234
1985	112190000	17409.720703
1986	106380000	17694.679688
1987	133030000	17929.990234
1988	82160000	17949.369141

Merge Dataframe

```
[6]: # Merge the two dataframe
merge = df1.merge(df2, how='inner', on='Date')
merge.set_index('Date', inplace=True)
merge.head(3)
```

```
[6]:
```

	Label	Top1 \
Date		
2008-08-08	0	b"Georgia 'downs two Russian warplanes' as cou...
2008-08-11	1	b'Why wont America and Nato help us? If they w...
2008-08-12	0	b'Remember that adorable 9-year-old who sang a...

	Top2 \
Date	
2008-08-08	b'BREAKING: Musharraf to be impeached.'
2008-08-11	b'Bush puts foot down on Georgian conflict'
2008-08-12	b"Russia 'ends Georgia operation'"

	Top3 \
Date	
2008-08-08	b'Russia Today: Columns of troops roll into So...
2008-08-11	b"Jewish Georgian minister: Thanks to Israeli ...

2008-08-12 b'"If we had no sexual harassment we would hav...

Top4 \

Date

2008-08-08 b'Russian tanks are moving towards the capital...

2008-08-11 b'Georgian army flees in disarray as Russians ...

2008-08-12 b'"Al-Qa'eda is losing support in Iraq because ...

Top5 \

Date

2008-08-08 b'"Afghan children raped with 'impunity,' U.N. ...

2008-08-11 b'"Olympic opening ceremony fireworks 'faked'"

2008-08-12 b'Ceasefire in Georgia: Putin Outmaneuvers the...

Top6 \

Date

2008-08-08 b'150 Russian tanks have entered South Ossetia...

2008-08-11 b'What were the Mossad with fraudulent New Zea...

2008-08-12 b'Why Microsoft and Intel tried to kill the XO...

Top7 \

Date

2008-08-08 b'"Breaking: Georgia invades South Ossetia, Rus...

2008-08-11 b'Russia angered by Israeli military sale to G...

2008-08-12 b'Stratfor: The Russo-Georgian War and the Bal...

Top8 \

Date

2008-08-08 b'"The 'enemy combatent' trials are nothing but...

2008-08-11 b'An American citizen living in S.Ossetia blam...

2008-08-12 b'"I'm Trying to Get a Sense of This Whole Geor...

Top9 ... \

Date

2008-08-08 b'Georgian troops retreat from S. Osettain cap... ...

2008-08-11 b'Welcome To World War IV! Now In High Definit... ...

2008-08-12 b'"The US military was surprised by the timing

Top22 \

Date

2008-08-08 b'Caucasus in crisis: Georgia invades South Os...

2008-08-11 b' Russia has just beaten the United States ov...

2008-08-12 b'All signs point to the US encouraging Georgi...

Top23 \

Date

2008-08-08 b'Indian shoe manufactory - And again in a se...

```
2008-08-11 b'Perhaps *the* question about the Georgia - R...
2008-08-12 b'Christopher King argues that the US and NATO...
```

Top24 \

Date

```
2008-08-08 b'Visitors Suffering from Mental Illnesses Ban...
2008-08-11 b'Russia is so much better at war'
2008-08-12 b'America: The New Mexico?'
```

Top25 Open \

Date

```
2008-08-08 b"No Help for Mexico's Kidnapping Surge" 11432.089844
2008-08-11 b"So this is what it's come to: trading sex fo... 11729.669922
2008-08-12 b"BBC NEWS | Asia-Pacific | Extinction 'by man... 11781.700195
```

	High	Low	Close	Volume	Adj Close
Date					
2008-08-08	11759.959961	11388.040039	11734.320312	212830000	11734.320312
2008-08-11	11867.110352	11675.530273	11782.349609	183190000	11782.349609
2008-08-12	11782.349609	11601.519531	11642.469727	173590000	11642.469727

[3 rows x 32 columns]

Clean Dataframe

```
[7]: # Combine headlines
headlines = []

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

```
[8]: # Verify the output
headlines[0]
```

```
[8]: 'b\'BREAKING: Musharraf to be impeached.\' b\'Russia Today: Columns of troops
roll into South Ossetia; footage from fighting (YouTube)\' b\'Russian tanks are
moving towards the capital of South Ossetia, which has reportedly been
completely destroyed by Georgian artillery fire\' b"Afghan children raped with
\'impunity,\' U.N. official says - this is sick, a three year old was raped and
they do nothing" b\'150 Russian tanks have entered South Ossetia whilst Georgia
shoots down two Russian jets.\' b"Breaking: Georgia invades South Ossetia,
Russia warned it would intervene on SO\'s side" b"The \'enemy combatent\' trials
are nothing but a sham: Salim Haman has been sentenced to 5 1/2 years, but will
be kept longer anyway just because they feel like it." b\'Georgian troops
retreat from S. Ossettain capital, presumably leaving several hundred people
killed. [VIDEO]\' b\'Did the U.S. Prep Georgia for War with Russia?\' b\'Rice
Gives Green Light for Israel to Attack Iran: Says U.S. has no veto over Israeli
```

military ops\' b\'Announcing:Class Action Lawsuit on Behalf of American Public Against the FBI\' b"So---Russia and Georgia are at war and the NYT\'s top story is opening ceremonies of the Olympics? What a fucking disgrace and yet further proof of the decline of journalism." b"China tells Bush to stay out of other countries\' affairs" b\'Did World War III start today?\' b\'Georgia Invades South Ossetia - if Russia gets involved, will NATO absorb Georgia and unleash a full scale war?\' b\'Al-Qaeda Faces Islamist Backlash\' b\'Condoleezza Rice: "The US would not act to prevent an Israeli strike on Iran." Israeli Defense Minister Ehud Barak: "Israel is prepared for uncompromising victory in the case of military hostilities."\' b\'This is a busy day: The European Union has approved new sanctions against Iran in protest at its nuclear programme.\' b"Georgia will withdraw 1,000 soldiers from Iraq to help fight off Russian forces in Georgia\'s breakaway region of South Ossetia" b\'Why the Pentagon Thinks Attacking Iran is a Bad Idea - US News & World Report\' b\'Caucasus in crisis: Georgia invades South Ossetia\' b\'Indian shoe manufactory - And again in a series of "you do not like your work?"\' b\'Visitors Suffering from Mental Illnesses Banned from Olympics\' b"No Help for Mexico\'s Kidnapping Surge" 11432.089844'

```
[9]: # Clean the data
_headlines = []

# Remove b', b" and \'
for i in range(0, len(headlines)):
    _headlines.append(re.sub("b[('\"'"]", '', headlines[i]))
    _headlines[i] = re.sub('b[("\'"]', '', _headlines[i])
    _headlines[i] = re.sub("\'", '', _headlines[i])
```

```
[10]: # Verify the cleaned headlines
_headlines[0]
```

```
[10]: 'BREAKING: Musharraf to be impeached. Russia Today: Columns of troops roll into South Ossetia; footage from fighting (YouTube) Russian tanks are moving towards the capital of South Ossetia, which has reportedly been completely destroyed by Georgian artillery fire Afghan children raped with impunity, U.N. official says - this is sick, a three year old was raped and they do nothing" 150 Russian tanks have entered South Ossetia whilst Georgia shoots down two Russian jets. Breaking: Georgia invades South Ossetia, Russia warned it would intervene on SOs side" The enemy combatent trials are nothing but a sham: Salim Haman has been sentenced to 5 1/2 years, but will be kept longer anyway just because they feel like it." Georgian troops retreat from S. Osettian capital, presumably leaving several hundred people killed. [VIDEO] Did the U.S. Prep Georgia for War with Russia? Rice Gives Green Light for Israel to Attack Iran: Says U.S. has no veto over Israeli military ops Announcing:Class Action Lawsuit on Behalf of American Public Against the FBI So---Russia and Georgia are at war and the NYTs top story is opening ceremonies of the Olympics? What a fucking disgrace and yet further proof of the decline of journalism." China tells Bush to stay out of other
```

countries affairs" Did World War III start today? Georgia Invades South Ossetia - if Russia gets involved, will NATO absorb Georgia and unleash a full scale war? Al-Qaeda Faces Islamist Backlash Condoleezza Rice: "The US would not act to prevent an Israeli strike on Iran." Israeli Defense Minister Ehud Barak: "Israel is prepared for uncompromising victory in the case of military hostilities." This is a busy day: The European Union has approved new sanctions against Iran in protest at its nuclear programme. Georgia will withdraw 1,000 soldiers from Iraq to help fight off Russian forces in Georgias breakaway region of South Ossetia" Why the Pentagon Thinks Attacking Iran is a Bad Idea - US News & World Report Caucasus in crisis: Georgia invades South Ossetia Indian shoe manufactory - And again in a series of "you do not like your work?" Visitors Suffering from Mental Illnesses Banned from Olympics No Help for Mexicos Kidnapping Surge" 11432.089844'

```
[11]: # Add clean headlines to merged dataframe
merge['Combined_news'] = _headlines
# Verify
merge['Combined_news']['2016-07-01']
```

```
[11]: 'IMF chief backs Athens as permanent Olympic host The president of France says
if Brexit won, so can Donald Trump British Man Who Must Give Police 24 Hours
Notice of Sex Threatens Hunger Strike: The man is the subject of a sexual risk
order despite having never been convicted of a crime. 100+ Nobel laureates urge
Greenpeace to stop opposing GMOs Brazil: Huge spike in number of police killings
in Rio ahead of Olympics Austrias highest court annuls presidential election
narrowly lost by right-wing candidate. Facebook wins privacy case, can track any
Belgian it wants: Doesnt matter if Internet users are logged into Facebook or
not Switzerland denies Muslim girls citizenship after they refuse to swim with
boys at school: The 12- and 14-year-old will no longer be considered for
naturalised citizenship because they have not complied with the school
curriculum, authorities in Basel said China kills millions of innocent
meditators for their organs, report finds France Cracks Down on Factory Farms -
A viral video campaign has moved the govt to act. In footage shared widely
online, animals writhe in pain as they bleed to death or are dismembered, in
violation of rules requiring they be rendered unconscious before slaughter.
Abbas PLO Faction Calls Killer of 13-Year-Old American-Israeli Girl a Martyr
Taiwanese warship accidentally fires missile towards China Iran celebrates
American Human Rights Week, mocks U.S. rights record U.N. panel moves to curb
bias against L.G.B.T. people despite fierce resistance from Muslim and African
countries. The United States has placed Myanmar, Uzbekistan, Sudan and Haiti on
its list of worst human trafficking offenders. S&P revises European Union
credit rating to AA from AA+ India gets $1 billion loan from World Bank for
solar mission U.S. sailors detained by Iran spoke too much under interrogation:
Navy Mass fish kill in Vietnam solved as Taiwan steelmaker accepts
responsibility for pollution Philippines president Rodrigo Duterte urges people
to kill drug addicts | Duterte, 71, won power in a landslide after a campaign
dominated by threats to kill tens of thousands in a war on crime Spain arrests
```


three Pakistanis accused of promoting militancy Venezuela, where anger over food shortages is still mounting, continued to be roiled this week by angry protests and break-ins of grocery stores and businesses that have left five dead, at least 30 injured and 200 arrested, according to various news reports. A Hindu temple worker has been killed by three men on a motorcycle, local police have said. More than 40 people have been killed in attacks blamed on Islamist militants in Bangladesh since February 2013, including secular bloggers, academics, gay rights activists and members of religious minorities. Ozone layer hole seems to be healing - US & UK team shows its shrunk & may slowly recover. "If you had to have an ozone hole anywhere in the world, itd be Antarctica because its not teeming with life. It showed us if we didnt back off with these chemicals, wed have a crisis." 17924.240234'

Subjectivity & Polarity

- Subjectivity refers to a personal opinion or judgement. It can be viewed as subjectivity to being objective or as an objective scale ranging from 0 to 1. A 0 indicates objective and 1 indicates subjective.
- Polarity ranges from -1 to 1. A +1 indicates a positive statement and -1 indicates a negative statement.

We'll now create a user defined function to derive both subjectivity and polarity.

```
[12]: # Create a function to get the subjectivity
def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity

# Create a function to get the polarity
def getPolarity(text):
    return TextBlob(text).sentiment.polarity
```

```
[13]: # Add subjectivity and polarity to merged dataframe
merge['Subjectivity'] = merge['Combined_news'].apply(getSubjectivity)
merge['Polarity'] = merge['Combined_news'].apply(getPolarity)
```

```
[14]: merge.head(3)
```

```
[14]:
```

	Label	Top1 \
Date		
2008-08-08	0	b"Georgia 'downs two Russian warplanes' as cou...
2008-08-11	1	b'Why wont America and Nato help us? If they w...
2008-08-12	0	b'Remember that adorable 9-year-old who sang a...

	Top2 \
Date	
2008-08-08	b'BREAKING: Musharraf to be impeached.'
2008-08-11	b'Bush puts foot down on Georgian conflict'
2008-08-12	b"Russia 'ends Georgia operation'"

Top3 \

Date

2008-08-08 b'Russia Today: Columns of troops roll into So...
2008-08-11 b'"Jewish Georgian minister: Thanks to Israeli ...
2008-08-12 b'"If we had no sexual harassment we would hav...

Top4 \

Date

2008-08-08 b'Russian tanks are moving towards the capital...
2008-08-11 b'Georgian army flees in disarray as Russians ...
2008-08-12 b'"Al-Qa'eda is losing support in Iraq because ...

Top5 \

Date

2008-08-08 b'"Afghan children raped with 'impunity,' U.N. ...
2008-08-11 b'Olympic opening ceremony fireworks 'faked'"
2008-08-12 b'Ceasefire in Georgia: Putin Outmaneuvers the...

Top6 \

Date

2008-08-08 b'150 Russian tanks have entered South Ossetia...
2008-08-11 b'What were the Mossad with fraudulent New Zea...
2008-08-12 b'Why Microsoft and Intel tried to kill the X0...

Top7 \

Date

2008-08-08 b'"Breaking: Georgia invades South Ossetia, Rus...
2008-08-11 b'Russia angered by Israeli military sale to G...
2008-08-12 b'Stratfor: The Russo-Georgian War and the Bal...

Top8 \

Date

2008-08-08 b'"The 'enemy combatent' trials are nothing but...
2008-08-11 b'An American citizen living in S.Ossetia blam...
2008-08-12 b'"I'm Trying to Get a Sense of This Whole Geor...

Top9 ... \

Date

2008-08-08 b'Georgian troops retreat from S. Osettain cap... ...
2008-08-11 b'Welcome To World War IV! Now In High Definit... ...
2008-08-12 b'"The US military was surprised by the timing

Top25 Open \

Date

2008-08-08 b'"No Help for Mexico's Kidnapping Surge" 11432.089844
2008-08-11 b'"So this is what it's come to: trading sex fo... 11729.669922

```
2008-08-12 b"BBC NEWS | Asia-Pacific | Extinction 'by man... 11781.700195
```

	High	Low	Close	Volume	Adj Close \
Date					
2008-08-08	11759.959961	11388.040039	11734.320312	212830000	11734.320312
2008-08-11	11867.110352	11675.530273	11782.349609	183190000	11782.349609
2008-08-12	11782.349609	11601.519531	11642.469727	173590000	11642.469727

	Combined_news Subjectivity \
Date	
2008-08-08	BREAKING: Musharraf to be impeached. Russia To... 0.277104
2008-08-11	Bush puts foot down on Georgian conflict Jewis... 0.374806
2008-08-12	Russia ends Georgia operation" "If we had no s... 0.494074

	Polarity
Date	
2008-08-08	-0.050303
2008-08-11	0.121956
2008-08-12	-0.048329

```
[3 rows x 35 columns]
```

Sentiment Analysis

Sentiment analysis is performed using VADER Sentiment Analyzer. The compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive).

The standardized thresholds for classifying sentences as either positive, neutral, or negative is also used in research. Typical threshold values are:

- positive sentiment: compound score ≥ 0.05
- neutral sentiment: (compound score > -0.05) and (compound score < 0.05)
- negative sentiment: compound score ≤ -0.05

The compound score is the one most commonly used metric for sentiment analysis.

```
[15]: # Create function for sentiment scores
def getSentiment(text):
    sa = SentimentIntensityAnalyzer()
    sentiment = sa.polarity_scores(text)
    return sentiment
```

```
[16]: # Get the sentiment scores for each day

compound = []
negative = []
positive = []
```

```

neutral = []
sa = 0

for i in range(0, len(merge)):
    sa = getSentiment(merge['Combined_news'][i])
    compound.append(sa['compound'])
    negative.append(sa['neg'])
    neutral.append(sa['neu'])
    positive.append(sa['pos'])

```

[17]: *# Store the sentiment scores in the dataframe*

```

merge['Compound'] = compound
merge['Negative'] = negative
merge['Neural'] = neutral
merge['Positive'] = positive

```

[18]: *# Verify the output*

```
merge.head(3)
```

[18]: Label Top1 \

```

Date
2008-08-08      0  b"Georgia 'downs two Russian warplanes' as cou...
2008-08-11      1  b'Why wont America and Nato help us? If they w...
2008-08-12      0  b'Remember that adorable 9-year-old who sang a...

```

Top2 \

```

Date
2008-08-08      b'BREAKING: Musharraf to be impeached.'
2008-08-11  b'Bush puts foot down on Georgian conflict'
2008-08-12      b"Russia 'ends Georgia operation'"

```

Top3 \

```

Date
2008-08-08  b'Russia Today: Columns of troops roll into So...
2008-08-11  b"Jewish Georgian minister: Thanks to Israeli ..."
2008-08-12  b'"If we had no sexual harassment we would hav...'

```

Top4 \

```

Date
2008-08-08  b'Russian tanks are moving towards the capital...'
2008-08-11  b'Georgian army flees in disarray as Russians ...
2008-08-12  b"Al-Qa'eda is losing support in Iraq because ..."

```

Top5 \

```

Date
2008-08-08  b"Afghan children raped with 'impunity,' U.N. ..."
2008-08-11      b"Olympic opening ceremony fireworks 'faked'"

```

2008-08-12 b'Ceasefire in Georgia: Putin Outmaneuvers the...

Top6 \

Date

2008-08-08 b'150 Russian tanks have entered South Ossetia...

2008-08-11 b'What were the Mossad with fraudulent New Zea...

2008-08-12 b'Why Microsoft and Intel tried to kill the XO...

Top7 \

Date

2008-08-08 b"Breaking: Georgia invades South Ossetia, Rus...

2008-08-11 b'Russia angered by Israeli military sale to G...

2008-08-12 b'Stratfor: The Russo-Georgian War and the Bal...

Top8 \

Date

2008-08-08 b"The 'enemy combatent' trials are nothing but...

2008-08-11 b'An American citizen living in S.Ossetia blam...

2008-08-12 b'I'm Trying to Get a Sense of This Whole Geor...

Top9 ... \

Date

2008-08-08 b'Georgian troops retreat from S. Osettain cap... ..

2008-08-11 b'Welcome To World War IV! Now In High Definit... ..

2008-08-12 b"The US military was surprised by the timing

Close Volume Adj Close \

Date

2008-08-08 11734.320312 212830000 11734.320312

2008-08-11 11782.349609 183190000 11782.349609

2008-08-12 11642.469727 173590000 11642.469727

Combined_news Subjectivity \

Date

2008-08-08 BREAKING: Musharraf to be impeached. Russia To... 0.277104

2008-08-11 Bush puts foot down on Georgian conflict Jewis... 0.374806

2008-08-12 Russia ends Georgia operation" "If we had no s... 0.494074

Polarity Compound Negative Neural Positive

Date

2008-08-08 -0.050303 -0.9981 0.236 0.722 0.042

2008-08-11 0.121956 -0.9875 0.196 0.726 0.078

2008-08-12 -0.048329 -0.9702 0.127 0.821 0.052

[3 rows x 39 columns]

Fit & Predict

```
[19]: # Create new dataframe with required columns
df = merge(['Open', 'High', 'Low', 'Volume', 'Subjectivity', 'Polarity', 'Compound', 'Negative', 'Neural', 'Positive', 'Label'])
```

```
[20]: # Create feature set
X = np.array(df.drop(['Label'], axis=1))

# Create target
y = np.array(df['Label'])
```

```
[21]: # Verify array shape
X.shape, y.shape
```

```
[21]: ((1989, 10), (1989,))
```

```
[22]: # Split the train and test dataset
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=23, shuffle=False)
```

```
[23]: # Create and train the model
model = LinearDiscriminantAnalysis().fit(X_train, y_train)
# Prediction
y_pred = model.predict(X_test)
```

```
[24]: # Output prediction score
print(f'Train Accuracy: {model.score(X_train, y_train):0.4}')
print(f'Test Accuracy: {model.score(X_test, y_test):0.4}')
```

Train Accuracy: 0.8592

Test Accuracy: 0.8719

```
[25]: # Classification Report
print(classification_report(y_test, y_pred))
```

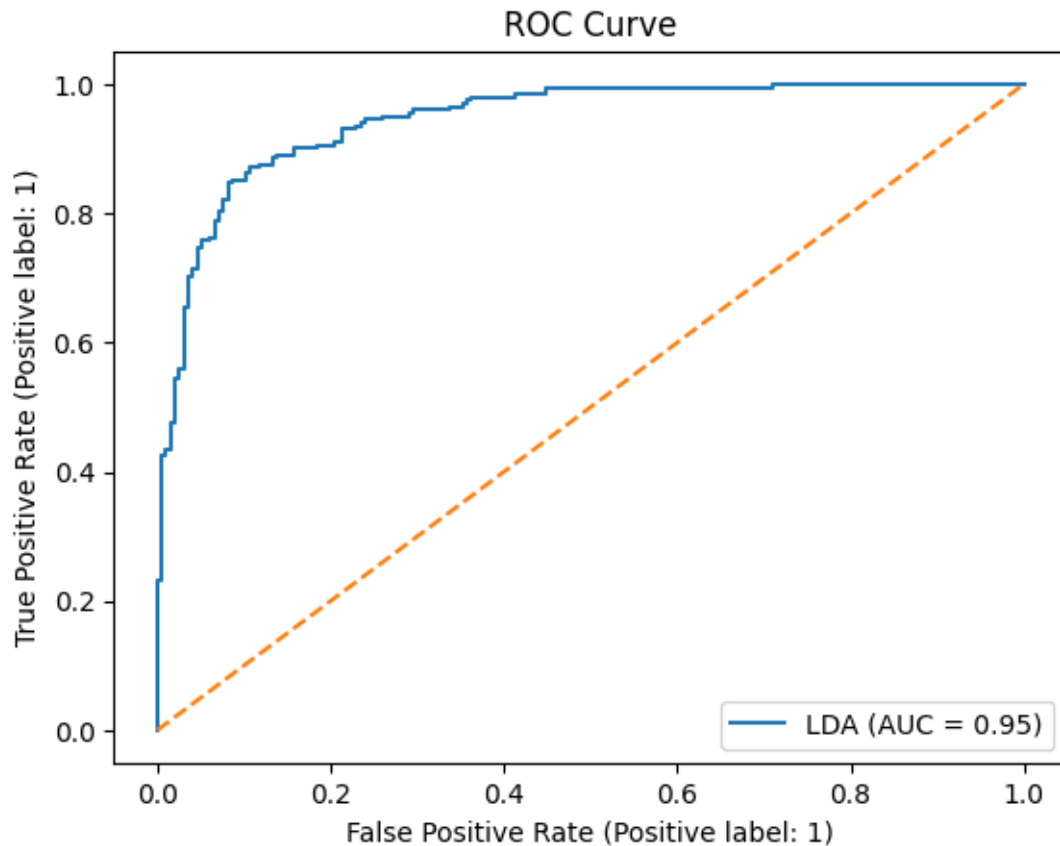
	precision	recall	f1-score	support
0	0.89	0.84	0.87	196
1	0.85	0.90	0.88	202
accuracy			0.87	398
macro avg	0.87	0.87	0.87	398
weighted avg	0.87	0.87	0.87	398

```
[26]: # Display ROC Curve
disp_roc = RocCurveDisplay.from_estimator(
    model,
```

```

X_test,
y_test,
name='LDA')
disp_roc.ax_.set_title('ROC Curve')
plt.plot([0,1],[0,1], linestyle='dashed')
plt.show()

```



The Linear Discriminant model is thus able to classify the labels based on the sentiments with an accuracy of 87%.

Note: The above datasets are sourced from a publically available data source. These datasets are pre-worked with labels for the respective dates. The dataset may not contain futuristic information given how labels are determined. The exercise is only for demonstration purposes on the treatment of sentiment analysis for financial time series.

References

- [VADER Sentiment](#)
- [Linear Discriminant Analysis](#)
- [Python resources](#)

* * *