Introduction

Grand Reopening. To most of the community the sound of those words is a magical event. After the long 9 weeks of "shelter in place" comes the "grand re-opening!" Most have dreamed of this moment! There is a vast measure of love for favorite meals.

So, where will society go to eat? It is at exactly at this time many will pull out mobile phones and check the reviews of a favorite restaurant. Restaurants themselves have also been managing around national and local restrictions on business, changes in lifestyle, and concerns over safety of workers and customers. It is no wonder an opportunity to get out of the house with others, at a safe distance is welcomed by both customers and establishments.

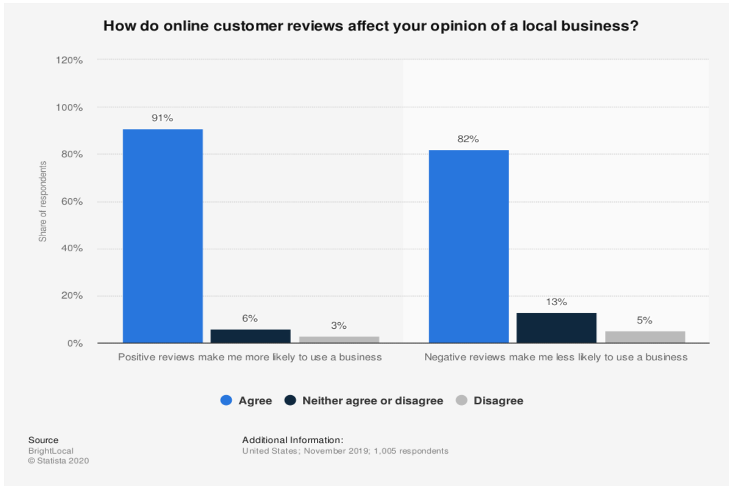
To select that "first place" is fun and brings a dilemma. Can reviews be trusted? Are collective opinions valued? "Fake News" and "Positive/Negative" will be evaluated in the following analysis and discussion to provide more confidence in your bon Appetit.

Analysis and Models

About the Data

Time is precious. There is a limited amount of it. Nothing like a pandemic to sink this into memory. After being at home there is a natural desire to be out of the house and in community. Selecting a meal location is part of making this a reality. But can new and reviews be trusted? For purposes of understanding data consisting restaurant reviews and news stories matched with a summary indication ("Is this True or False?{t,f]" or "Is this a Positive Review or Negative? [p,n]" will be used to evaluate the accuracy of detecting fake news and collective wisdom regarding restaurants.

Even prior to our exercise it is notable how much value the public brings to reviews regardless of their validity according to Bright Local and 1,005 respondents to their question. To classify customer reviews by sentiment and authenticity data was collected on restaurant reviews. The restaurant review



data consists of 90 distinct reviews measured with sentiment (positive/negative) and factual (true/false) rating (performed independently) attached to each review. A view of the original data is shown below:

lie	sentiment	review				
f	n	'Mike\'s Pi	NY Service	not. Stick to pre-made dishes lik		
f	n	'i really like	japanese	and chinese dishes. we also got		
f	n	'After I we	we went to DODO restaurant for dinner. I f			
f	n	'Olive Oil G	and the waitor had no manners whatsoeve			
f	n	'The Seven	never more. '			

Each line in the data is a separate and independent review. To prepare this data for our use it is read into Python, and split into train and test segments. It was then vectorized by Naïve Bayes and Support Vector Machines (SVM) using, Count, and Tfidf vectorzors. A fit/transform was run and then both the Naïve Bayes and SVM models were trained against the data. The results were evaluated for precision, and confusion matrix, The results were graphed below.

SENTIMENT RUN

Vectorization

label text

0 p do i eve pick a best experience at joe oodle h...

["p,'My", 'sister', 'and', 'I', 'ate', 'at', 'this', 'restaurant', 'called', 'Matador.', 'The', 'overall', 'look', 'and', 'ambiance', 'of', 'the', 'restaurant', 'was', 'very', 'appealing.', 'We', 'first', 'ordered', 'strawberry', 'margaritas--which', 'were', 'really', 'good.Then', 'my', 'sister', 'ordered', 'a', 'spinach', 'lasagna', 'with', 'Alfredo', 'sauce', 'and', 'I', 'ordered', 'Pasta', 'ravioli', 'with', 'marinara', 'sauce.', 'My', 'sister', 'and', 'I', 'unanimously', 'agreed', 'they', 'were', 'the', 'best', 'pastas', 'we', 'had', 'ever', 'had.', 'It', 'was', 'a', 'beautiful', 'blend', 'of', 'flavors', 'which', 'complimented', 'each', 'other.', 'I', 'would', 'totally', 'recommend', 'Matador', 'and', 'it', 'was', 'an', 'overall', 'amazing', "experience.',,,,,,,,,,,,,,,,,,,,"]

the new word is: p,'My

the new word is: sister

the new word is: and

the new word is: I

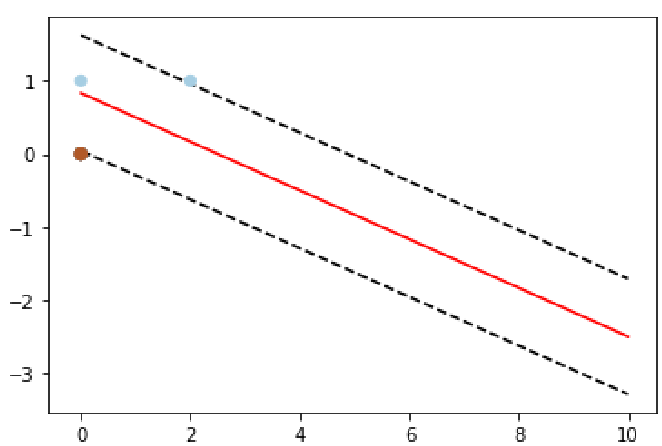
the new word is: ate

the new word is: at

the new word is: this

the new word is: restaurant

SVM GRAPH - SENTIMENT



Name: Label, Length: 65, dtype: object

[0 1 0 0 0 1 0 0 0 1 1 0 1 1 1 1 1 0 0 0 1 0 1 1 0 0 1 1 0 1 0 0 1 0 1 1 1

1 0 1 0 0 1 1 1 0 1 0 0 0 1 1 1 0 0 0 0 0 1 1 0 0 1 1 1]

Out[12]:

(-0.6040322679360337,

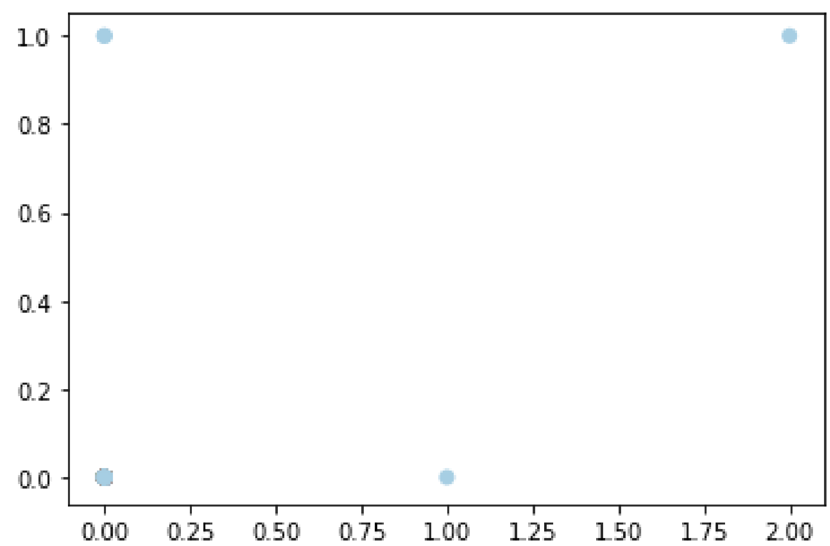
10.504953917520764,

-19.72983738762489,

4.729837387624885)

OBJ

SVM GRPAH - LIES



__main__:5: RuntimeWarning: divide by zero encountered in double_scalars

__main__:13: RuntimeWarning: invalid value encountered in double_scalars

__main__:24: RuntimeWarning: divide by zero encountered in double_scalars

Out[46]:

(-0.10660631689302408,

2.1004444014006816,

-0.06017712693684174,

1.0507923881602228)

SENTIMENT DATA

DATAFRAMES:

FinalDF_STEM

	Label	abc	abruptli	absolut	accept	...	yeah	yelp	york	youll	yuena
0	SENT_POS	0.0	0.0	0.0	0.0	...	0	0.0	0	0.0	0.0
1	SENT_POS	0.0	0.0	0.0	0.0	...	0	0.0	0	0.0	0.0
2	SENT_POS	0.0	0.0	0.0	0.0	...	0	0.0	0	0.0	0.0
3	SENT_POS	0.0	0.0	0.0	0.0	...	0	0.0	0	0.0	1.0

FinalDF_TFIDF

	Label	abc	abruptly	...	youll	youre	yuena
0	SENT_POS	0.000000	0.0	...	0.0	0.000000	0.000000
1	SENT_POS	0.000000	0.0	...	0.0	0.000000	0.000000
2	SENT_POS	0.000000	0.0	...	0.0	0.000000	0.000000
3	SENT_POS	0.000000	0.0	...	0.0	0.000000	0.362783

[93 rows x 1254 columns]

FinalDF_TFIDF_STEMF

	Label	abc	abruptli	absolut	...	yelp	york	youll	yuena
0	SENT_POS	0.000000	0.0	0.000000	...	0.000000	0.00000	0.0	0.000000
1	SENT_POS	0.000000	0.0	0.000000	...	0.000000	0.00000	0.0	0.000000
2	SENT_POS	0.000000	0.0	0.000000	...	0.000000	0.00000	0.0	0.000000
3	SENT_POS	0.000000	0.0	0.000000	...	0.000000	0.00000	0.0	0.358325

[93 rows x 1109 columns]

The prediction from NB is: STEM

```
['SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS'
 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_POS'
 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS'
 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS'
 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_POS']
```

The actual labels are:

23 SENT_NEG

4 SENT_POS

36 SENT_POS

The prediction from NB is: TFIDF

['SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG'
'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_POS'
'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_NEG'
'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG'
'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_POS']

The actual labels are:

37 SENT_NEG

45 SENT_NEG

6 SENT_POS

The prediction from NB is: TFIDF STEM

['SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG'
'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_NEG'
'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG'
'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_POS'
'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_POS']

The actual labels are:

43 SENT_NEG

34 SENT_POS

13 SENT_NEG

```
print(np.round(MyModelNB1.predict_proba(TestDF1),2))
print(np.round(MyModelNB2.predict_proba(TestDF2),2))
print(np.round(MyModelNB3.predict_proba(TestDF3),2))
```

The confusion matrix is:

```
[[13  1]
 [ 0 14]]
```

The confusion matrix is:

```
[[10  4]
 [ 3 11]]
```

The confusion matrix is:

```
[[ 8  9]
 [ 1 10]]
[[1.   0.  ]
 [0.   1.  ]
 [0.08 0.92]
 [0.   1.  ]
 [1.   0.  ]
 [0.   1.  ]
 [1.   0.  ]
 [0.   1.  ]
 [0.01 0.99]
 [1.   0.  ]
 [0.   1.  ]
 [0.   1.  ]
 [1.   0.  ]
```

```
 [0.   1.  ]
 [0.5  0.5 ]
 [0.01 0.99]
 [1.   0.  ]
 [0.03 0.97]
 [1.   0.  ]
 [1.   0.  ]
 [1.   0.  ]
 [0.   1.  ]
 [1.   0.  ]
 [0.   1.  ]
 [0.   1.  ]
 [1.   0.  ]
 [1.   0.  ]
 [0.01 0.99]]
[[0.46 0.54]
 [0.71 0.29]
 [0.51 0.49]
 [0.4  0.6 ]
 [0.55 0.45]
 [0.56 0.44]
 [0.46 0.54]
 [0.52 0.48]
 [0.47 0.53]
 [0.5  0.5 ]
 [0.31 0.69]
 [0.46 0.54]
 [0.56 0.44]
 [0.54 0.46]
```

```
 [0.45 0.55]
 [0.32 0.68]
 [0.6  0.4 ]
 [0.52 0.48]
 [0.6  0.4 ]
 [0.49 0.51]
 [0.44 0.56]
 [0.65 0.35]
 [0.49 0.51]
 [0.51 0.49]
 [0.52 0.48]
 [0.49 0.51]
 [0.43 0.57]
 [0.49 0.51]]
[[0.51 0.49]
 [0.39 0.61]
 [0.54 0.46]
 [0.44 0.56]
 [0.45 0.55]
 [0.55 0.45]
 [0.48 0.52]
 [0.43 0.57]
 [0.38 0.62]
 [0.38 0.62]
 [0.3  0.7 ]
 [0.52 0.48]
 [0.55 0.45]
 [0.39 0.61]
 [0.52 0.48]
```

[0.46 0.54]

[0.48 0.52]

[0.52 0.48]

[0.52 0.48]

[0.39 0.61]

[0.41 0.59]

[0.4 0.6]

[0.49 0.51]

[0.44 0.56]

[0.51 0.49]

[0.47 0.53]

[0.29 0.71]

[0.44 0.56]]

SVM MODELS1-3

SVM prediction:

['SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS'
'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_POS'
'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS'
'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS'
'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG']

Actual:

23   SENT_NEG

4   SENT_POS

36   SENT_POS

The confusion matrix is:

[[14 0]

[0 14]]

[LibSVM]SVM prediction:

['SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS'
'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_POS'
'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS'
'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS'
'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG']

Actual:

23 SENT_NEG

4 SENT_POS

36 SENT_POS

The confusion matrix is:

[[14 0]

[0 14]]

[LibSVM]SVM prediction:

['SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_POS'
'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_POS'
'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_POS'
'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_POS'
'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_POS']

Actual:

23 SENT_NEG

4 SENT_POS

36 SENT_POS

The confusion matrix is:

[[6 8]

[0 14]]

LIES DATA

(SKIPPING DATAFRAMES)

The prediction from NB is: STEM

['SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS'
 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_NEG'
 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_POS'
 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG'
 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS'
 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_POS'
 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG']

The actual labels are:

4 SENT_NEG

25 SENT_POS

26 SENT_NEG

The prediction from NB is: TFIDF

['SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS'
 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG'
 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG'
 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG']

The actual labels are:

16 SENT_NEG

37 SENT_POS

2 SENT_NEG

The prediction from NB is:Tfidf Stem

['SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS'
'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG'
'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG'
'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG']

The actual labels are:

48 SENT_NEG

11 SENT_POS

12 SENT_NEG

Lies Confusion Matrix

The confusion matrix is:

[[13 13]
[8 7]]

The confusion matrix is:

[[17 13]
[11 0]]

The confusion matrix is:

[[23 5]
[13 0]]

[[0.98 0.02]
[0.93 0.07]]

[1. 0.]

[1. 0.]

[0.57 0.43]

[0. 1.]

[1. 0.]

[0.84 0.16]

[0. 1.]

[0. 1.]

[1. 0.]

[1. 0.]

[1. 0.]

[0.99 0.01]

[1. 0.]

[0. 1.]

[0. 1.]

[0. 1.]

[1. 0.]

[0.06 0.94]

[0.09 0.91]

[0.66 0.34]

[0.34 0.66]

[1. 0.]

[0. 1.]

[0. 1.]

[1. 0.]

[0.34 0.66]

[0.88 0.12]

[0. 1.]

[0.98 0.02]

```
 [0.01 0.99]
 [1.   0.  ]
 [0.47 0.53]
 [0.   1.  ]
 [0.   1.  ]
 [0.05 0.95]
 [0.   1.  ]
 [1.   0.  ]
 [0.   1.  ]
 [0.99 0.01]]
[[0.71 0.29]
 [0.74 0.26]
 [0.63 0.37]
 [0.41 0.59]
 [0.66 0.34]
 [0.47 0.53]
 [0.79 0.21]
 [0.67 0.33]
 [0.8  0.2 ]
 [0.43 0.57]
 [0.45 0.55]
 [0.64 0.36]
 [0.67 0.33]
 [0.78 0.22]
 [0.42 0.58]
 [0.69 0.31]
 [0.47 0.53]
 [0.66 0.34]
 [0.67 0.33]
```

```
 [0.44 0.56]
 [0.78 0.22]
 [0.65 0.35]
 [0.81 0.19]
 [0.56 0.44]
 [0.78 0.22]
 [0.41 0.59]
 [0.45 0.55]
 [0.67 0.33]
 [0.72 0.28]
 [0.64 0.36]
 [0.4  0.6 ]
 [0.39 0.61]
 [0.72 0.28]
 [0.76 0.24]
 [0.76 0.24]
 [0.74 0.26]
 [0.41 0.59]
 [0.72 0.28]
 [0.63 0.37]
 [0.41 0.59]
 [0.63 0.37]]
[[0.45 0.55]
 [0.66 0.34]
 [0.76 0.24]
 [0.63 0.37]
 [0.73 0.27]
 [0.64 0.36]
 [0.78 0.22]
```

[0.75 0.25]

[0.83 0.17]

[0.62 0.38]

[0.66 0.34]

[0.71 0.29]

[0.73 0.27]

[0.73 0.27]

[0.68 0.32]

[0.57 0.43]

[0.65 0.35]

[0.31 0.69]

[0.77 0.23]

[0.75 0.25]

[0.79 0.21]

[0.47 0.53]

[0.76 0.24]

[0.63 0.37]

[0.78 0.22]

[0.56 0.44]

[0.79 0.21]

[0.33 0.67]

[0.64 0.36]

[0.53 0.47]

[0.61 0.39]

[0.68 0.32]

[0.65 0.35]

[0.8 0.2]

[0.67 0.33]

[0.62 0.38]

[0.74 0.26]

[0.47 0.53]

[0.62 0.38]

[0.51 0.49]

[0.7 0.3]]

Main SVM prediction:

['SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_POS'
'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG'
'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG'
'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS'
'SENT_NEG' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_POS']

Actual:

16 SENT_NEG

37 SENT_POS

2 SENT_NEG

The confusion matrix is:

[[12 18]

[10 1]]

Three SVM Models

SVM prediction:

['SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG'
'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_POS'

'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG'

'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG'

'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'

'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG']

Actual:

4 SENT_NEG

25 SENT_POS

26 SENT_NEG

The confusion matrix is:

[[26 0]

[0 15]]

[LibSVM]SVM prediction:

['SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'

'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG'

'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_POS'

'SENT_POS' 'SENT_POS' 'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG'

'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG'

'SENT_POS' 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'

'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG']

Actual:

4 SENT_NEG

25 SENT_POS

26 SENT_NEG

The confusion matrix is:

[[26 0]
 [0 15]]

[LibSVM]SVM prediction:

 ['SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG'
 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG' 'SENT_POS'
 'SENT_NEG' 'SENT_POS' 'SENT_NEG' 'SENT_NEG' 'SENT_NEG']

Actual:

4   SENT_NEG

25   SENT_POS

26   SENT_NEG

The confusion matrix is:

[[24 2]
 [15 0]]

], s=10,

__main__:5: RuntimeWarning: divide by zero encountered in double_scalars

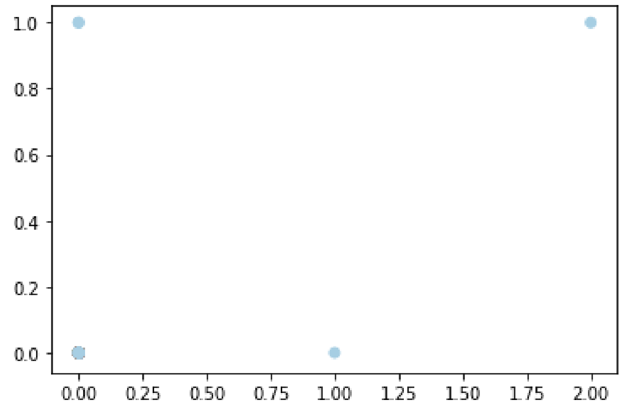
__main__:13: RuntimeWarning: invalid value encountered in double_scalars

__main__:24: RuntimeWarning: divide by zero encountered in double_scalars

Out[46]:

(-0.10660631689302408,
 2.1004444014006816,
 -0.06017712693684174,
 1.0507923881602228)

OBJ



Results

Examination of the model output reveal a much higher accuracy on Naïve Bayes with a much more accurate confusion matrix. This was surprising, and unexpected. The graph for SVM clearly shows the best plane for separation of sentiment data. The lie separation ran into issues which were not resolved.

The SVM poly kernel was superior to the other SVM kernels and achieved 100% accuracy which was far the others.

Sentiment analysis appears to be a more accurate predictor of good restaurants based on analysis of restaurant reviews.

Conclusion.

These are unprecedented times. A grand reopening of American society is occurring in real time. This includes some of our favorite restaurants. More than ever as people venture back out into public they are seeking out their favorite foods after being denied for over two months. A review of current restaurants which are open is now a habit for most people. But can these reviews be trusted. Results here show a greater likelihood that Positive/Negative Sentiment reviews have higher accuracy across multiple across multiple sources than True/False reviews.

Much of the community bases their perception on such reviews. It appears the asking the question "is this a positive review?" or "that seems obviously negative" are questions and emotions to listen to and resonate well with consumers. In this time of “fake news” there is another standby, that of “positive/negative”. These can be depended upon.