MOVIE REVIEW SENTIMENT ANALYSIS

BY TIM HULAK – IST 664: NLP

ABOUT THE DATA

- THE DATA WAS TAKEN FROM THE ORIGINAL PANG AND LEE MOVIE REVIEW CORPUS BASED ON REVIEWS FROM THE ROTTEN TOMATOES WEB SITE.

- THE DATA WAS SPLIT INTO 2 TAB SEPARATED DATA SETS: TEST.TSV AND TRAIN.TSV
    - TEST DATA MADE UP ~30% OF THE TOTAL 222,352 TOTAL RECORDS (~30/70 SPLIT)

- SENTIMENT KEY:
    - 0 = NEGATIVE
    - 1 = SOMEWHAT NEGATIVE
    - 2 = NEUTRAL
    - 3 = SOMEWHAT POSITIVE
    - 4 = POSITIVE

DATA PREVIEW

Sentiment Key:

- 0 = negative
- 1 = somewhat negative
- 2 = neutral
- 3 = somewhat positive
- 4 = positive

Test Data

	PhraseId	SentenceId	Phrase
0	156061	8545	An intermittently pleasing but mostly routine ...
1	156062	8545	An intermittently pleasing but mostly routine ...
2	156063	8545	An
3	156064	8545	intermittently pleasing but mostly routine effort
4	156065	8545	intermittently pleasing but mostly routine

Test Shape (66292, 3)

Training Data

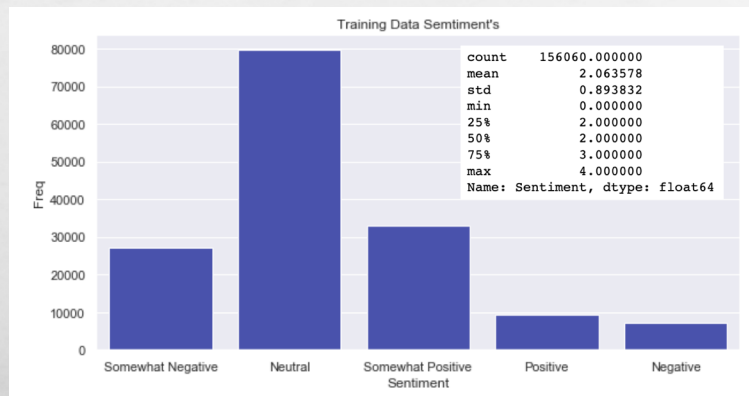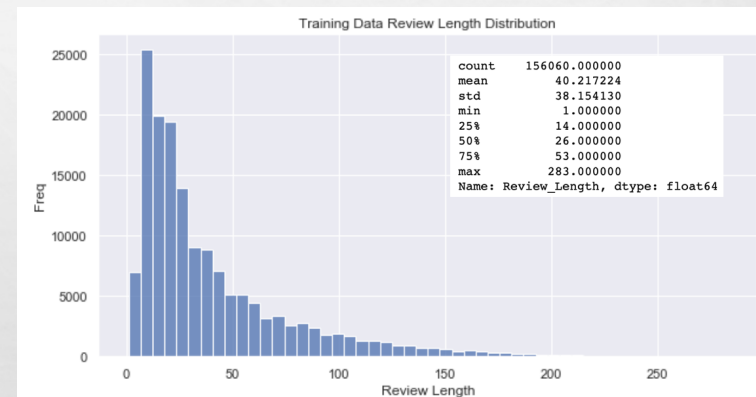|	PhraseId	SentenceId	Phrase	Sentiment
0	1	1	A series of escapades demonstrating the adage ...	1
1	2	1	A series of escapades demonstrating the adage ...	2
2	3	1	A series	2
3	4	1	A	2
4	5	1	series	2

Training Shape: (156060, 4)

EXPLORATORY DATA ANALYSIS

Sentiment Key:

- 0 = negative
- 1 = somewhat negative
- 2 = neutral
- 3 = somewhat positive
- 4 = positive

FREQUENCY OF SENTIMENT

Training Data Semtiment's

```
count    156060.000000
mean          2.063578
std           0.893832
min           0.000000
25%           2.000000
50%           2.000000
75%           3.000000
max           4.000000
Name: Sentiment, dtype: float64
```

DISTRIBUTION OF REVIEW LENGTH

Training Data Review Length Distribution

```
count    156060.000000
mean         40.217224
std          38.154130
min           1.000000
25%          14.000000
50%          26.000000
75%          53.000000
max         283.000000
Name: Review_Length, dtype: float64
```

NAÏVE BAYES CLASSIFIER

- AFTER TOKENIZATION AND VECTORIZATION, THE ACCURACY OF THE NAÏVE BAYES CLASSIFIER WAS ~ 58%

```
7  # Evaluate Model
8  print("Accuracy ",round(accuracy_score(y_test,predictions),3))
```

```
Accuracy  0.582
```

	precision	recall	f1-score	support
Negative	0.61	0.03	0.05	2113
Neutral	0.60	0.90	0.72	23588
Positive	0.65	0.05	0.09	2848
Somewhat Negative	0.52	0.24	0.33	8228
Somewhat Positive	0.53	0.38	0.44	10041
accuracy			0.58	46818
macro avg	0.58	0.32	0.33	46818
weighted avg	0.57	0.58	0.52	46818

PREDICTING ON THE TEST DATA

- **THE NAÏVE BAYES CLASSIFIER MODEL WAS APPLIED TO THE UNLABELED TEST DATA**

	PhraseId	SentenceId	Phrase	Review_Length	Predicted_Sentiment
0	156061	8545	An intermittently pleasing but mostly routine ...	54	Somewhat Positive
1	156062	8545	An intermittently pleasing but mostly routine ...	52	Somewhat Positive
2	156063	8545	An	2	Neutral
3	156064	8545	intermittently pleasing but mostly routine effort	49	Somewhat Positive
4	156065	8545	intermittently pleasing but mostly routine	42	Somewhat Positive
5	156066	8545	intermittently pleasing but	27	Somewhat Positive
6	156067	8545	intermittently pleasing	23	Somewhat Positive
7	156068	8545	intermittently	14	Somewhat Positive
8	156069	8545	pleasing	8	Somewhat Positive
9	156070	8545	but	3	Neutral
10	156071	8545	mostly routine	14	Neutral
11	156072	8545	mostly	6	Neutral
12	156073	8545	routine	7	Neutral
13	156074	8545	effort	6	Neutral
14	156075	8545	.	1	Neutral

NLTK: VADER

- VADER (VALENCE AWARE DICTIONARY FOR SENTIMENT REASONING) IS A MODEL USED FOR TEXT SENTIMENT ANALYSIS THAT IS SENSITIVE TO BOTH POLARITY (POSITIVE/NEGATIVE) AND INTENSITY (STRENGTH) OF EMOTION.

- IT IS AVAILABLE IN THE NLTK PACKAGE AND CAN BE APPLIED DIRECTLY TO <u>UNLABELED</u> TEXT DATA.


```
1  # VADER
2  from nltk.sentiment.vader import SentimentIntensityAnalyzer
3  sid = SentimentIntensityAnalyzer()
```

VADER EXAMPLES FROM NLTK

```
1  test_phrase = "This is a good movie"
2
3  sid.polarity_scores(test_phrase)
```
{'neg': 0.0, 'neu': 0.508, 'pos': 0.492, 'compound': 0.4404}

MOSTLY POSITIVE

```
1  test_phrase_2 = "This is was the best, most awesome movie EVER MADE!!!"
2  sid.polarity_scores(test_phrase_2)
```
{'neg': 0.0, 'neu': 0.458, 'pos': 0.542, 'compound': 0.8877}

VERY POSITIVE

VERY NEGATIVE

```
1  test_phrase_3 = "This is the WORST movie that has ever disgraced the screen"
2  sid.polarity_scores(test_phrase_3)
```
{'neg': 0.465, 'neu': 0.535, 'pos': 0.0, 'compound': -0.8331}

VADER PERFORMANCE AGAINST TRAINING DATA

	precision	recall	f1-score	support
Negative	0.20	0.25	0.22	7072
Neutral	0.67	0.70	0.68	79582
Positive	0.24	0.48	0.32	9206
Somewhat Negative	0.37	0.25	0.30	27273
Somewhat Positive	0.36	0.31	0.33	32922
micro avg	0.50	0.50	0.50	156055
macro avg	0.37	0.40	0.37	156055
weighted avg	0.51	0.50	0.50	156055

	precision	recall	f1-score	support
Negative	0.51	0.40	0.45	34345
Neutral	0.67	0.70	0.68	79582
Positive	0.54	0.60	0.57	42133
micro avg	0.60	0.60	0.60	156060
macro avg	0.57	0.56	0.57	156060
weighted avg	0.60	0.60	0.60	156060

	precision	recall	f1-score	support
Negative	0.80	0.65	0.72	21158
Positive	0.77	0.88	0.82	28629
micro avg	0.78	0.78	0.78	49787
macro avg	0.79	0.76	0.77	49787
weighted avg	0.78	0.78	0.78	49787

1ST MODEL (ALL 5 LABELS)

- ~ 50.6% ACCURACY

2ND MODEL (POS, NEG, NEU)

- "SOMEWHAT NEGATIVE" AND "SOMEWHAT POSITIVE" WERE CONVERTED TO "POSITIVE" AND "NEGATIVE": ~ 60.4% ACCURACY
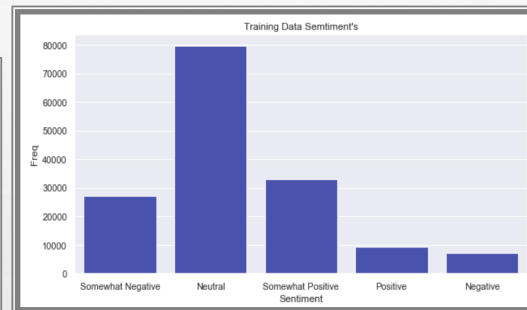
3RD MODEL (POS, NEG)

- TRAINING DATA FILTERED TO ONLY INCLUDE "POSITIVE" AND "NEGATIVE" (FEWER ROWS): ~78.2% ACCURACY
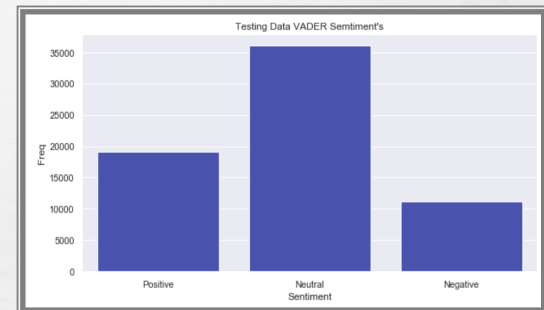
VADER CLASSIFICATION ON THE TEST DATA

	PhraseId	SentenceId	Phrase	Polarity_Score	VADER_Score
0	156061	8545	An intermittently pleasing but mostly routine ...	{'neg': 0.0, 'neu': 0.732, 'pos': 0.268, 'comp...	Positive
1	156062	8545	An intermittently pleasing but mostly routine ...	{'neg': 0.0, 'neu': 0.732, 'pos': 0.268, 'comp...	Positive
2	156063	8545	An	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...	Neutral
3	156064	8545	intermittently pleasing but mostly routine effort	{'neg': 0.0, 'neu': 0.694, 'pos': 0.306, 'comp...	Positive
4	156065	8545	intermittently pleasing but mostly routine	{'neg': 0.0, 'neu': 0.645, 'pos': 0.355, 'comp...	Positive

VADER CLASSIFIED THE PHRASES

TRAINING DATA

TESTING DATA

- SINCE "NEUTRAL" WAS SUCH A FREQUENT FEATURE, THE VADER MODEL WITH POS, NEG, AND NEU WAS APPLIED TO THE TEST DATA

- WHILE THE ACCURACY OF THE VADER MODEL(S) WERE BETWEEN 50% – 78%, THE SHAPE OF THE PREDICTIONS SEEMS TO MATCH THE SHAPE OF THE TRAINING DATA

- MOSTLY NEUTRAL, BUT SLIGHTLY MORE POSITIVES THAN NEGATIVES

VADER VS. NAÏVE BAYES

	PhraseId	SentenceId	Phrase	Review_Length	Predicted_Sentiment	VADER_Score
0	156061	8545	An intermittently pleasing but mostly routine ...	54	Somewhat Positive	Somewhat Positive
1	156062	8545	An intermittently pleasing but mostly routine ...	52	Somewhat Positive	Somewhat Positive
2	156063	8545	An	2	Neutral	Neutral
3	156064	8545	intermittently pleasing but mostly routine effort	49	Somewhat Positive	Somewhat Positive
4	156065	8545	intermittently pleasing but mostly routine	42	Somewhat Positive	Somewhat Positive
5	156066	8545	intermittently pleasing but	27	Somewhat Positive	Somewhat Positive
6	156067	8545	intermittently pleasing	23	Somewhat Positive	Positive
7	156068	8545	intermittently	14	Somewhat Positive	Neutral
8	156069	8545	pleasing	8	Somewhat Positive	Positive

```
                   precision    recall  f1-score   support

         Negative       0.00      0.25      0.00        32
          Neutral       0.89      0.59      0.71     54011
         Positive       0.00      0.24      0.01       132
Somewhat Negative       0.15      0.29      0.19      3704
Somewhat Positive       0.23      0.32      0.26      8413

         accuracy                           0.54     66292
        macro avg       0.25      0.34      0.24     66292
     weighted avg       0.76      0.54      0.63     66292
```
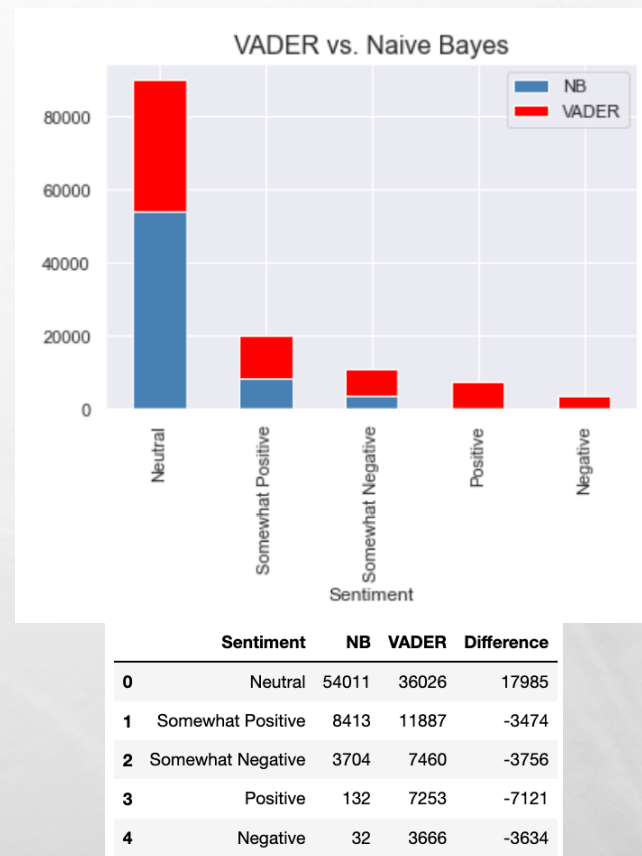

DATA PREVIEW

- VADER PREDICTED LABELS COMPARED TO NAÏVE BAYES PREDICTED LABELS

CLASSIFICATION REPORT
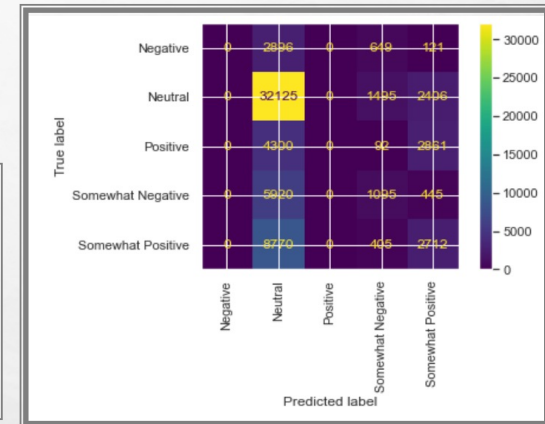
CONFUSION MATRIX

- SEEMED TO PREDICT "NEUTRAL" THE BEST, WHICH MAKES SENSE BECAUSE IT MADE UP OVER 50% OF THE DATA

VADER VS NAÏVE BAYES

VADER vs. Naive Bayes

- INTERESTINGLY, WHILE VADER'S ACCURACY WAS SLIGHTLY LOWER THAN NAÏVE BAYES (50.6% VS 58.2%), VADER SEEMED TO HAVE A MORE REALISTIC SET OF PREDICTIONS

- IN OTHER WORDS, OF THE 66,292 RECORDS IN THE TEST DATA, IT IS UNLIKELY THERE WOULD ONLY BE 32 NEGATIVE PHRASES OR 132 POSITIVE PHRASES.

	Sentiment	NB	VADER	Difference
0	Neutral	54011	36026	17985
1	Somewhat Positive	8413	11887	-3474
2	Somewhat Negative	3704	7460	-3756
3	Positive	132	7253	-7121
4	Negative	32	3666	-3634

THANK YOU

MOVIE REVIEW SENTIMENT ANALYSIS: BY TIM HULAK – IST 664: NLP