ML1010 - Group Project
Milestone 2 - Code

M.Boyd-Vasiliou

Dataset - Features

The main feature of data within the dataset is the “reviewText” column which is the actual review that the user has left and is associated with the rating they gave the product out of 5 stars.

I took a number of steps to generate new features for exploration within the project/model:

1. Using spaCy, clear stopwords from data and create new features from the reviewText split into separate columns:
    a. Lemmatized text (dataset label: lemma)
    b. Nouns only (dataset label: nouns)
    c. Adjectives only (dataset label: adjectives)
    d. Verbs only (dataset label: verbs)
    e. Nouns, Adjectives, and Verbs (dataset label: nav)
2. For each of the newly created features noted above, generate:
    a. spaCy/TextBlob polarity (dataset label: tb_pol)
    b. spaCy/TextBlob subjectivity (dataset label: tb_subj)
    c. spaCy/TextBlob num. tokens (dataset label: tb_tokens)
    d. spaCy/TextBlob length (dataset label: tb_length)
    e. BERT encoding (dataset label: bert)
    f. Flair polarity (dataset label: flairSent)
3. For each feature that is in: Flair polarity, spaCy/TextBlob polarity
    a. Normalize value to 0,1 for positive/negative comparison
    b. Normalize value to 1-5 for 5 star comparison to “overall” column
4. Normalize the 5 star rating to 0,1 for positive/negative comparison

The original data of “reviewText” and it’s associated “overall” (star rating) expanded to 49 columns of data to be used in our analysis.

Data dictionary of dataset features:

ID	Column	Description
1	uuid	Unique ID for joining
2	reviewText	Text of review
3	overall	Rating for review [1-5 stars]
4	reviewText_lemma	Lemmatized/Stopworded reviewText
5	reviewText_nouns	Lemmatized/Stopworded reviewText - nouns only
6	reviewText_adjectives	Lemmatized/Stopworded reviewText - adjectives only
7	reviewText_verbs	Lemmatized/Stopworded reviewText - verbs only
8	reviewText_nav	Lemmatized/Stopworded reviewText - nouns, adjectives, and verbs
9	reviewText_lemma_tb_pol	TextBlob polarity for lemmatized reviewText
10	reviewText_lemma_tb_subj	TextBlob subjectivity for lemmatized reviewText
11	reviewText_lemma_tb_tokens	TextBlob token count for lemmatized reviewText
12	reviewText_lemma_tb_length	TextBlob length for lemmatized reviewText
13	reviewText_lemma_bert	Bert encoding for lemmatized reviewText
14	reviewText_lemma_flairSent	Flair sentiment for lemmatized reviewText
15	reviewText_adjectives_tb_pol	TextBlob polarity for lemmatized reviewText - adjectives only
16	reviewText_adjectives_tb_subj	TextBlob subjectivity for lemmatized reviewText - adjectives only
17	reviewText_adjectives_tb_tokens	TextBlob token count for lemmatized reviewText - adjectives only
18	reviewText_adjectives_tb_length	TextBlob length for lemmatized reviewText - adjectives only
19	reviewText_adjectives_bert	Bert encoding for lemmatized reviewText - adjectives only

20	reviewText_adjectives_flairSent	Flair sentiment for lemmatized reviewText - adjectives only
21	reviewText_verbs_tb_pol	TextBlob polarity for lemmatized reviewText - verbs only
22	reviewText_verbs_tb_subj	TextBlob subjectivity for lemmatized reviewText - verbs only
23	reviewText_verbs_tb_tokens	TextBlob token count for lemmatized reviewText - verbs only
24	reviewText_verbs_tb_length	TextBlob length for lemmatized reviewText - verbs only
25	reviewText_verbs_tb_bert	Bert encoding for lemmatized reviewText - verbs only
26	reviewText_verbs_tb_flairSent	Flair sentiment for lemmatized reviewText - verbs only
27	reviewText_nav_tb_pol	TextBlob polarity for lemmatized reviewText - nouns, adj, and verbs
28	reviewText_nav_tb_subj	TextBlob subjectivity for lemmatized reviewText - nouns, adj, and verbs
29	reviewText_nav_tb_tokens	TextBlob token count for lemmatized reviewText - nouns, adj, and verbs
30	reviewText_nav_tb_length	TextBlob length for lemmatized reviewText - nouns, adj, and verbs
31	reviewText_nav_tb_bert	Bert encoding for lemmatized reviewText - nouns, adj, and verbs
32	reviewText_nav_tb_flairSent	Flair sentiment for lemmatized reviewText - nouns, adj, and verbs
33	overall_posneg	5 star rating adjust to 0/1
34	reviewText_lemma_flairSent_norm	Flair sentiment normalized to 5 categories for lemmatized reviewText
35	reviewText_lemma_flairSent_posneg	Flair sentiment adjusted to 0/1 for lemmatized reviewText
36	reviewText_adjectives_flairSent_norm	Flair sentiment normalized to 5 categories for lemmatized reviewText - adjectives only
37	reviewText_adjectives_flairSent_posneg	Flair sentiment adjusted to 0/1 for lemmatized reviewText - adjectives only

38	reviewText_verbs_flairSent_norm	Flair sentiment normalized to 5 categories for lemmatized reviewText - verbs only
39	reviewText_verbs_flairSent_posneg	Flair sentiment adjusted to 0/1 for lemmatized reviewText - verbs only
40	reviewText_nav_flairSent_norm	Flair sentiment normalized to 5 categories for lemmatized reviewText - for nouns, adjectives, and verbs
41	reviewText_nav_flairSent_posneg	Flair sentiment adjusted to 0/1 for lemmatized reviewText - for nouns, adjectives, and verbs
42	reviewText_lemma_tb_pol_norm	TextBlob sentiment normalized to 5 categories for lemmatized reviewText
43	reviewText_lemma_tb_pol_posneg	TextBlob sentiment adjusted to 0/1 for lemmatized reviewText
44	reviewText_adjectives_tb_pol_norm	TextBlob sentiment normalized to 5 categories for lemmatized reviewText - adjectives only
45	reviewText_adjectives_tb_pol_posneg	TextBlob sentiment adjusted to 0/1 for lemmatized reviewText - adjectives only
46	reviewText_verbs_tb_pol_norm	TextBlob sentiment normalized to 5 categories for lemmatized reviewText - verbs only
47	reviewText_verbs_tb_pol_posneg	TextBlob sentiment adjusted to 0/1 for lemmatized reviewText - verbs only
48	reviewText_nav_tb_pol_norm	TextBlob sentiment normalized to 5 categories for lemmatized reviewText - for nouns, adjectives, and verbs
49	reviewText_nav_tb_pol_posneg	TextBlob sentiment adjusted to 0/1 for lemmatized reviewText - for nouns, adjectives, and verbs

Analysis:

To start, I generated and analyzed the lemmatized text features under several permutations:

1. Using a 5 star classification, generate models/values for:
    a. Flair Sentiment
    b. spaCy/TextBlob Sentiment
    c. BERT
2. Normalizing the 5 star and polarity to 0,1 (positive,negative), generate models/values for:
    a. Flair Sentiment
    b. spaCy/TextBlob Sentiment
    c. BERT

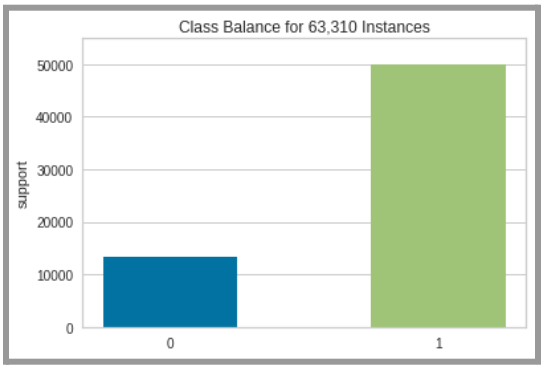
I generated the above permutations using two different methodologies for comparison:

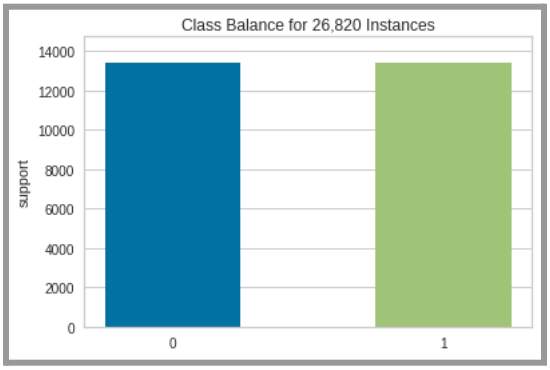
1. Balance the data to 5 classifications (star ratings), and 2 classifications (positive, negative). Balancing was performed by undersampling the data to the smallest of the categories.

2. Prune the data to exclude reviews with greater than 5 tokens and less than 2100 tokens. Then balance the data for both 5 star classification as well as 2 category (positive, negative) classification.

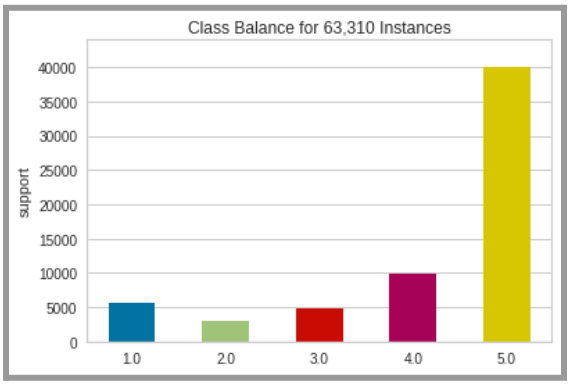
*Note: All data has gone through a cleaning process in the initial stages of the data wrangling
**Note: At this point, I have not yet run all new features through an analysis.

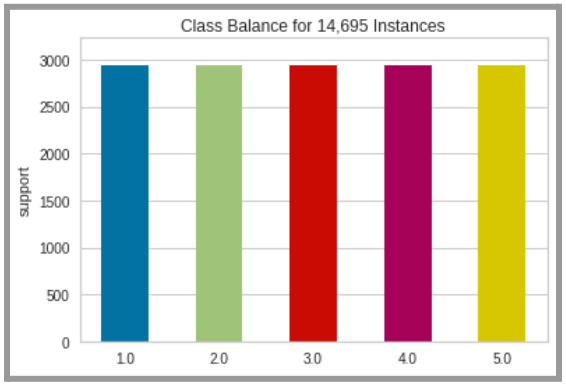
Example balancing (positive, negative):





Example balancing 5 star classification:





5 Star Classifications - Flair

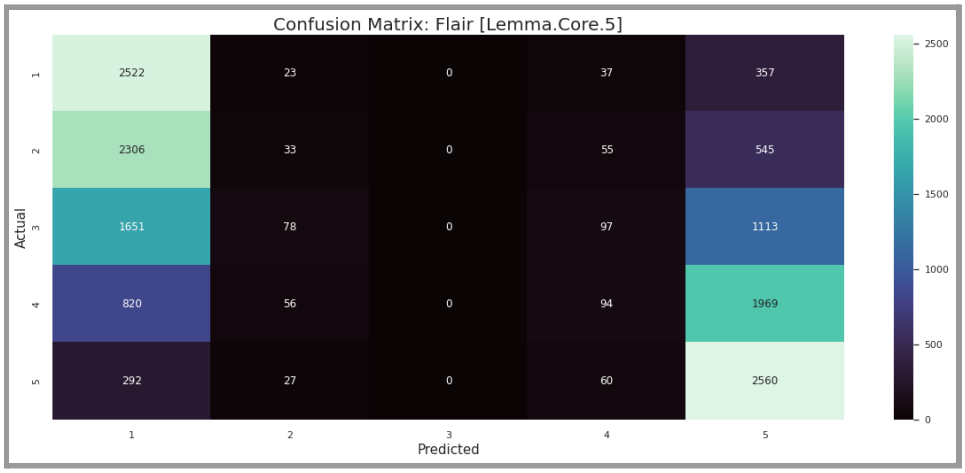
The Flair classification on 5 stars showed a distinct gap in the mid range for sentiment analysis. There is very little within the dataset that Flair believed to be of a neutral sentiment. This was one of the reasons I included a 2 star classification (positive, negative) in the analysis. Regardless of pruning unhelpful data , the model did not improve. The overall results were quite poor, and got worse with data pruning. Length of text may make this less suitable for Flair.

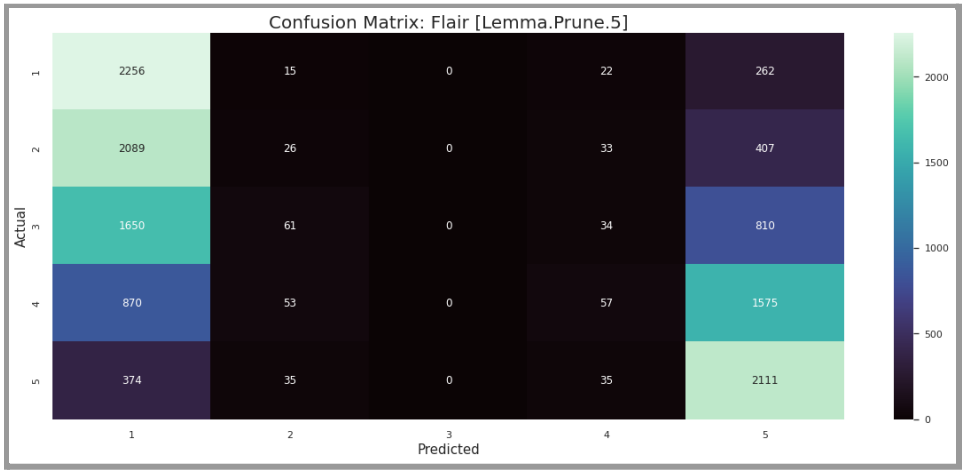
FLAIR Encoding: Five Classification
Column: ReviewText

	Precision	Recall	F1-Score	Support
1 Star	33%	86%	48%	2939
2 Star	15%	1%	2%	2939
3 Star	0%	0%	0%	2939
4 Star	27%	3%	6%	2939
5 Star	39%	87%	54%	2939
Accuracy			35%	14695
Macro Avg	23%	35%	22%	14695
Weighted Avg	23%	35%	22%	14695

FLAIR Encoding: Five Classification
Column: ReviewText with data pruning

	Precision	Recall	F1-Score	Support
1 Star	31%	88%	46%	2555
2 Star	14%	1%	2%	2555
3 Star	0%	0%	0%	2555
4 Star	31%	2%	4%	2555
5 Star	41%	83%	55%	2555
Accuracy			35%	12775
Macro Avg	23%	35%	21%	12775
Weighted Avg	23%	35%	21%	12775





Two Classifications (positive, negative) - Flair

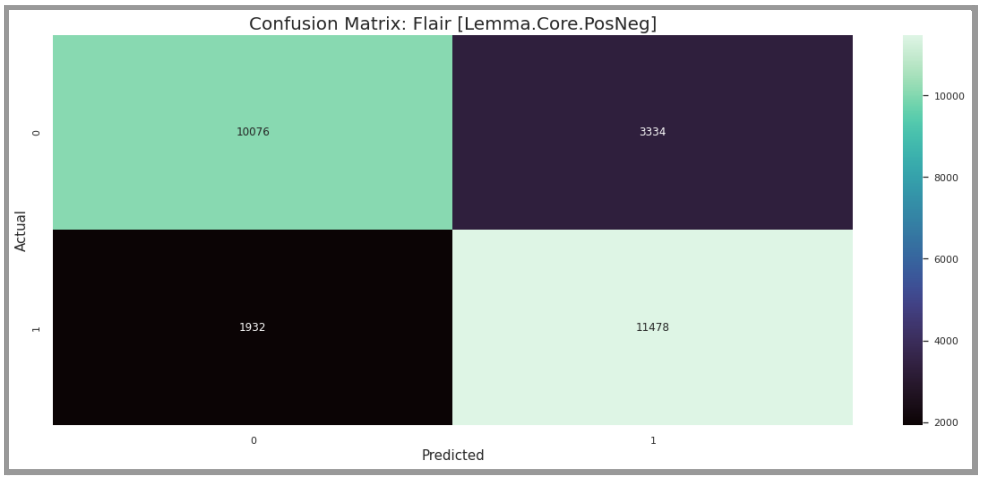
Changing the analysis to only include positive and negative produced much better results with most metrics around the 80% mark. While this is an improvement, it doesn't seem to capture the details we wished of the 5 star classification.

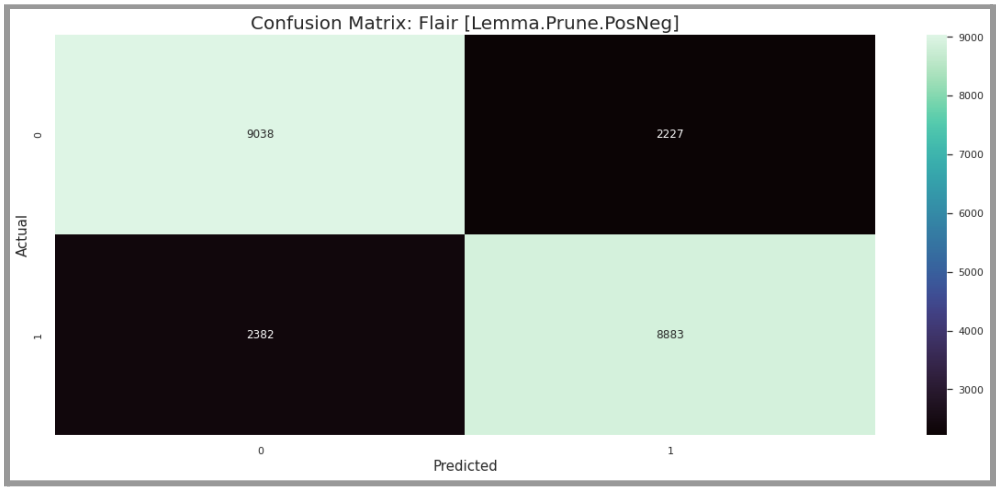
FLAIR Encoding: Two Classification (Positive/Negative
Column: ReviewText [Lemmatized, stopword]

	Precision	Recall	F1-Score	Support
Negative	84%	75%	79%	13410
Positive	77%	86%	81%	13410
Accuracy			80%	26820
Macro Avg	81%	80%	80%	26820
Weighted Avg	81%	80%	80%	26820

FLAIR Encoding: Two Classification (Positive/Negative)
Column: ReviewText [Lemmatized, stopword]

	Precision	Recall	F1-Score	Support
Negative	79%	80%	80%	11265
Positive	80%	79%	79%	11265
Accuracy			80%	22530
Macro Avg	81%	80%	80%	22530
Weighted Avg	81%	80%	80%	22530





5 Star Classifications - spaCy/TextBlob

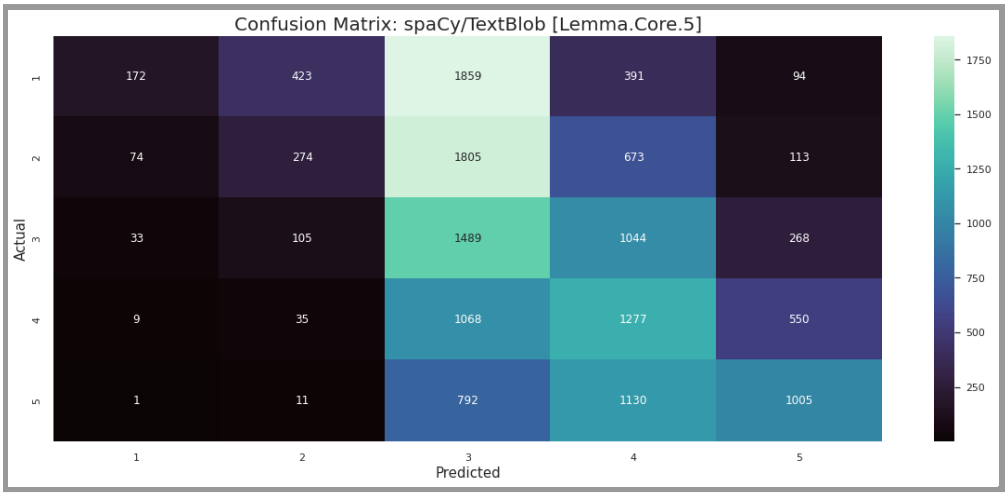
TextBlob, through spaCy produced equally poor results both with and without data pruning showing extremely low scores across the board. It's interesting to note that spaCy/TextBlob predicted neutral (3 star) ratings where Flair did not.

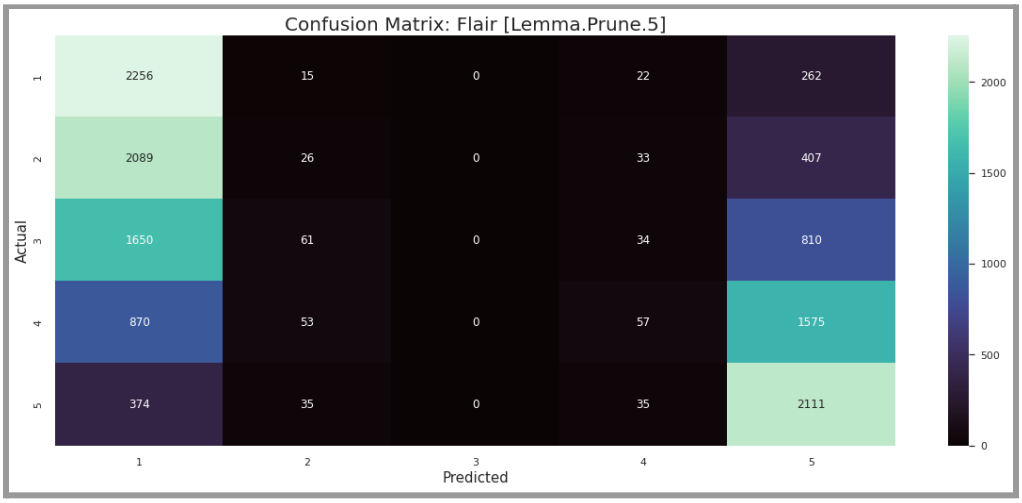
spaCy/TextBlob Encoding: Five Classification
Column: ReviewText [Lemmatized, stopword]

	Precision	Recall	F1-Score	Support
1 Star	60%	6%	11%	2939
2 Star	32%	9%	14%	2939
3 Star	21%	51%	30%	2939
4 Star	28%	43%	34%	2939
5 Star	50%	34%	40%	2939
Accuracy			29%	14695
Macro Avg	38%	29%	26%	14695
Weighted Avg	38%	29%	26%	14695

spaCy/TextBlob Encoding: Five Classification
Column: ReviewText [Lemmatized, stopword]

	Precision	Recall	F1-Score	Support
1 Star	57%	4%	8%	2555
2 Star	31%	9%	14%	2555
3 Star	23%	54%	32%	2555
4 Star	28%	50%	36%	2555
5 Star	52%	22%	31%	2555
Accuracy			28%	12775
Macro Avg	38%	28%	24%	12775
Weighted Avg	38%	28%	24%	12775





Two Classifications (positive, negative) - spaCy/TextBlob

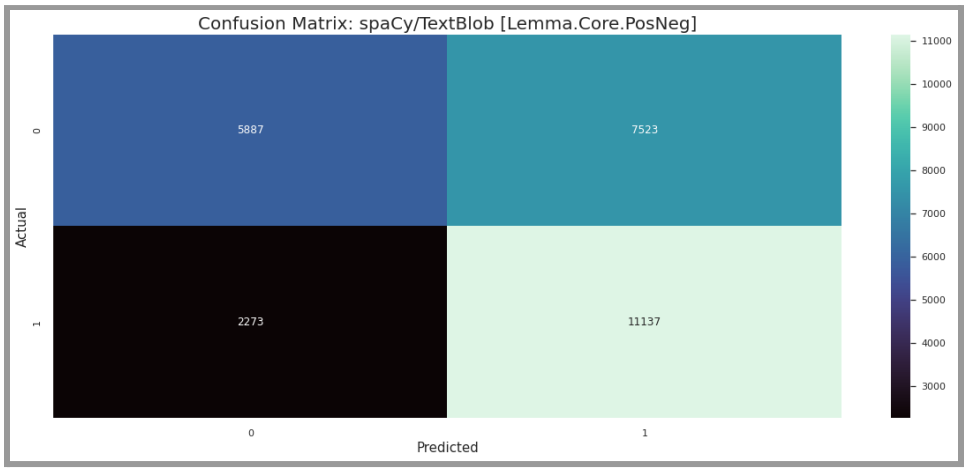
Similar to Flair, there were significant improvements when changing this to a two classification (positive, negative) problem. The results were not as strong as Flair but considering Flair did not predict any neutral ratings, there are pros and cons depending on use case and application.

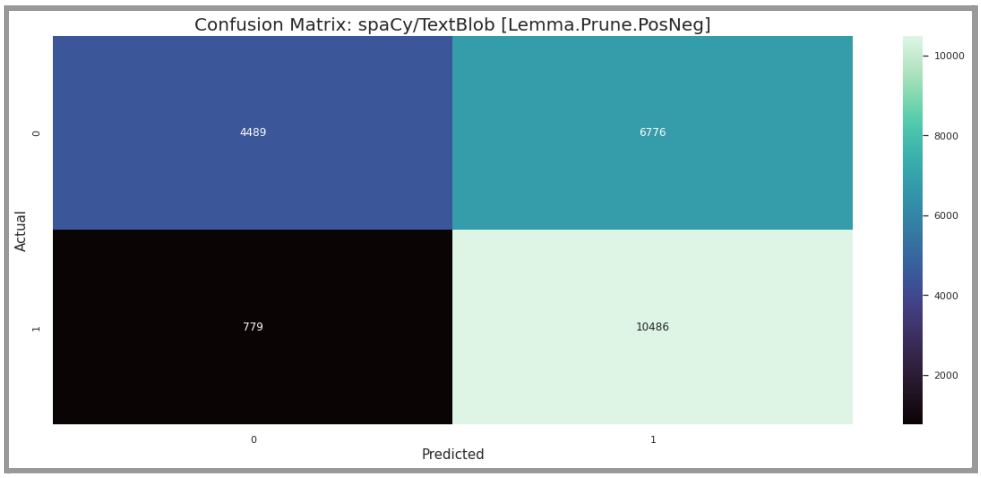
spaCy/TextBlob Encoding: Two Classification (Positive
Column: ReviewText [Lemmatized, stopword]

	Precision	Recall	F1-Score	Support
Negative	72%	44%	55%	13410
Positive	60%	83%	69%	13410
Accuracy			63%	26820
Macro Avg	66%	63%	62%	26820
Weighted Avg	66%	63%	62%	26820

spaCy/TextBlob Encoding: Two Classification (Positive/
Column: ReviewText [Lemmatized, stopword]

	Precision	Recall	F1-Score	Support
Negative	85%	40%	54%	11265
Positive	61%	93%	74%	11265
Accuracy			66%	22530
Macro Avg	73%	66%	64%	22530
Weighted Avg	73%	66%	64%	22530





5 Star Classifications - BERT

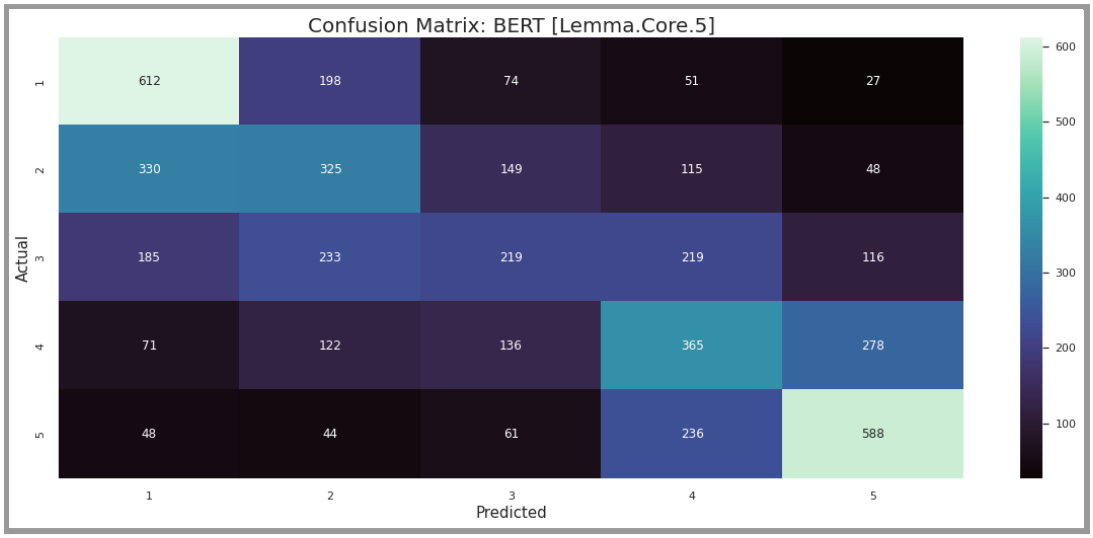
BERT encoding with an XGBoost model produced some strong results rivaling those in previous categories. There was stronger prediction in the middle (neutral) categories as well as decent performance in both positive and negative ranges.

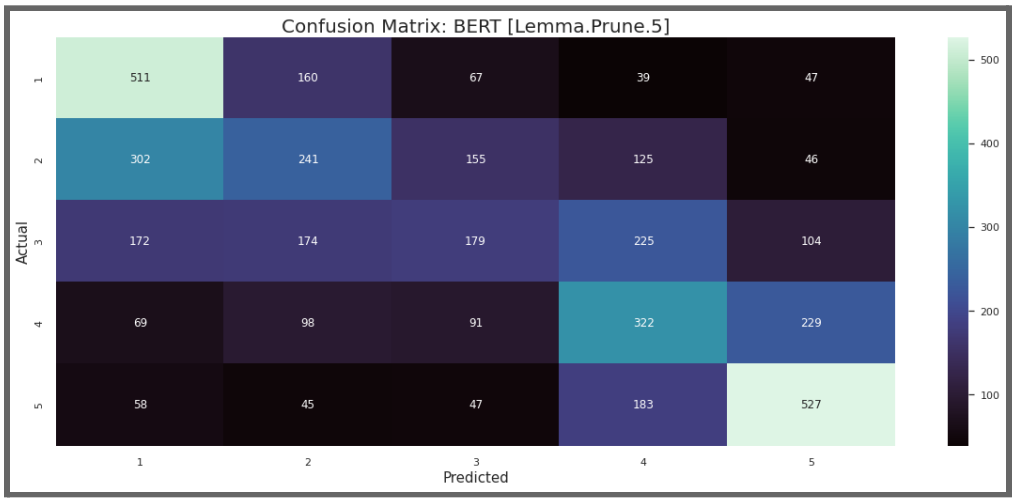
BERT Encoding: Five Classification
Column: ReviewText [Lemmatized, stopword]

	Precision	Recall	F1-Score	Support
1 Star	49%	64%	55%	962
2 Star	35%	34%	34%	967
3 Star	34%	23%	27%	972
4 Star	37%	38%	37%	972
5 Star	56%	60%	58%	977
Accuracy			43%	4850
Macro Avg	42%	43%	42%	4850
Weighted Avg	42%	43%	42%	4850

BERT Encoding: Five Classification
Column: ReviewText [Lemmatized, stopword, prune]

	Precision	Recall	F1-Score	Support
1 Star	46%	62%	53%	824
2 Star	34%	28%	30%	869
3 Star	33%	21%	26%	854
4 Star	36%	40%	38%	809
5 Star	55%	61%	58%	860
Accuracy			42%	4216
Macro Avg	41%	42%	41%	4216
Weighted Avg	41%	42%	41%	4216





Two Classifications (positive, negative) - BERT

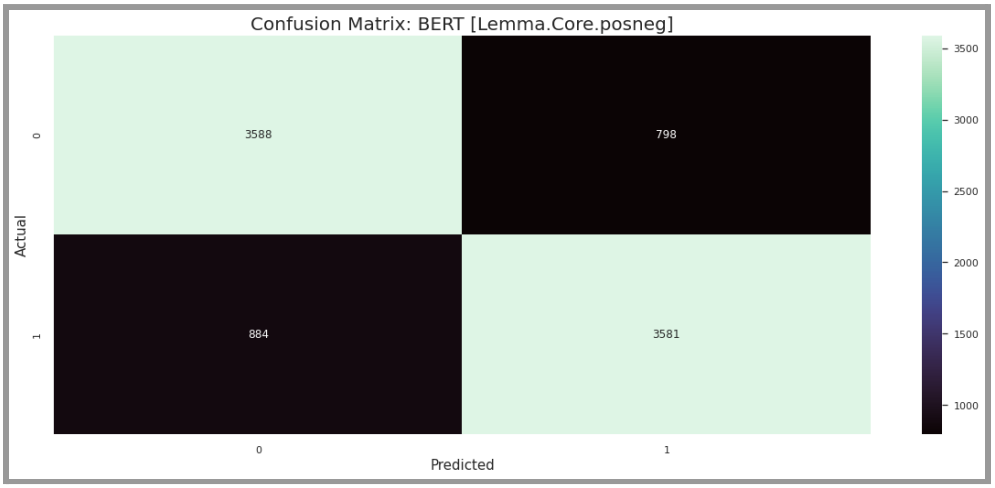
Similar to the other models, and understandably so based on pure statistics, but the 2 category (positive, negative) analysis produced strong results which surpassed the 5 star category. The BERT classifications performed on par with Flair and TextBlob in this current analysis.

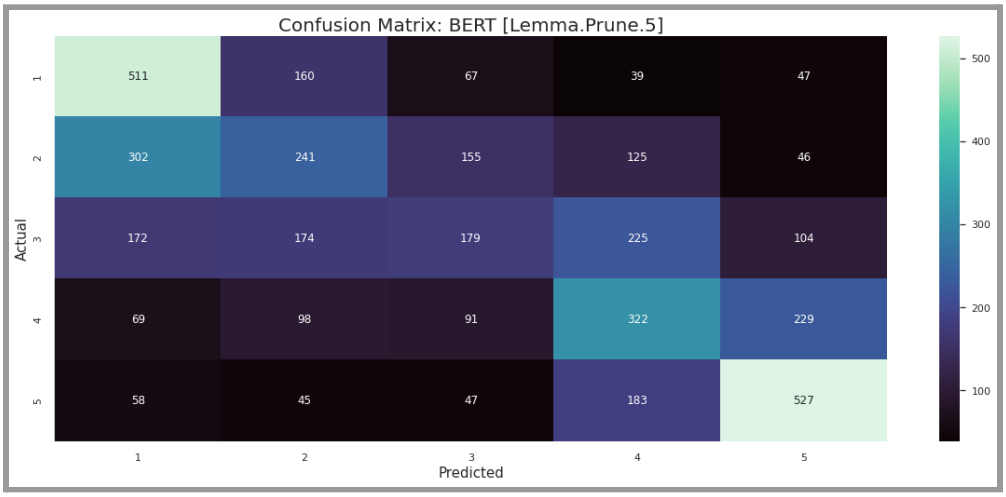
BERT Encoding: Two Classification (Positive/Negative)
Column: ReviewText [Lemmatized, stopword]

	Precision	Recall	F1-Score	Support
Negative	80%	82%	81%	4386
Positive	82%	80%	81%	4465
Accuracy			81%	8851
Macro Avg	81%	81%	81%	8851
Weighted Avg	81%	81%	81%	8851

BERT Encoding: Two Classification (Positive/Negative)
Column: ReviewText [Lemmatized, stopword, prune]

	Precision	Recall	F1-Score	Support
Negative	79%	80%	80%	3653
Positive	81%	80%	80%	3782
Accuracy			80%	7435
Macro Avg	80%	80%	80%	7435
Weighted Avg	80%	80%	80%	7435





Conclusion/Next Steps:

Additional analysis on other features are likely to produce additional insights that will assist in improving our results from the current analysis.

I have added the code to add some additional features by splitting the sentences of the “reviewText” field into separate columns based on the sentiment of each sentence. There were to be columns created for a 5 star rating as well as those for 2 classifications of positive, negative. There was insufficient memory available to complete the required task on Colab. Additional work in being done in this area so this feature may be utilized in keyword/topic analysis.

I had increased the data set to 63,000 but that still does not feel as though it is sufficient. When using a 5 star classification, where the star ratings are extremely skewed, the balancing removes a very significant portion of the data. Additional data and processing would be ideal.