



ML1010 Milestone 2

M.Vasiliou

Data Construction



Review Text



adjectives



verbs



nouns



nouns, adj,
verbs

X 5

lemma



spacy/TextBlob
polarity



spacy/TextBlob
subjectivity



spacy/TextBlob
tokens



spacy/TextBlob
length



BERT



Flair

X 6



Mapping Values

-1.....1



1,2,3,4,5



-1..1



0,1



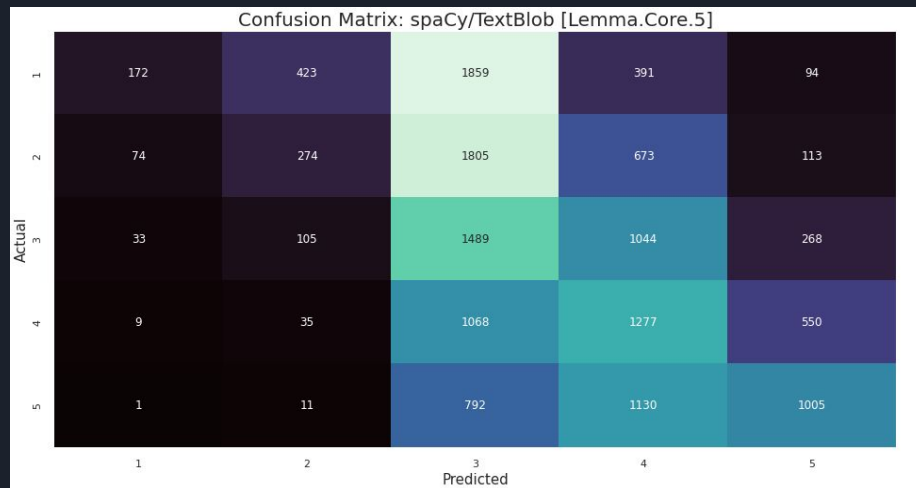
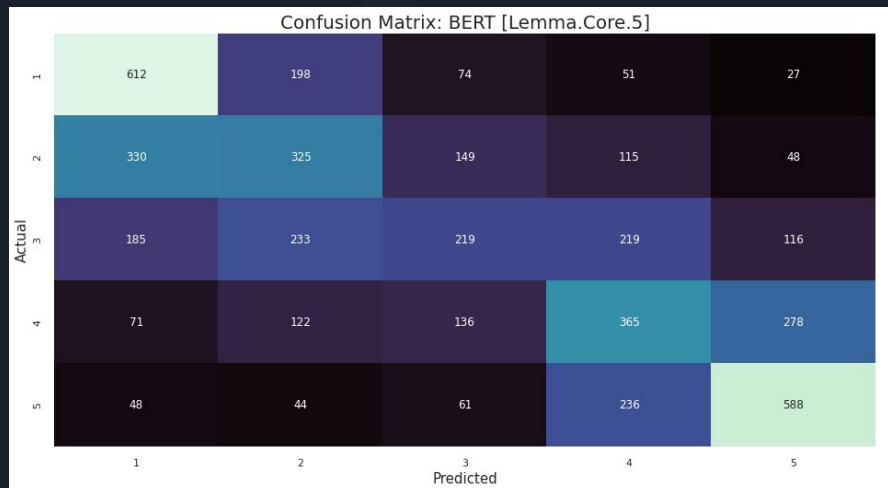
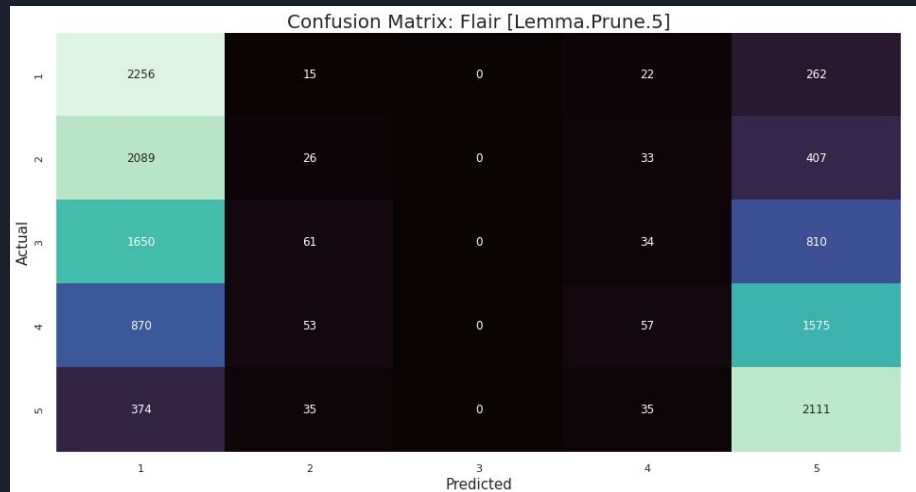
1,2,3,4,5



0,1



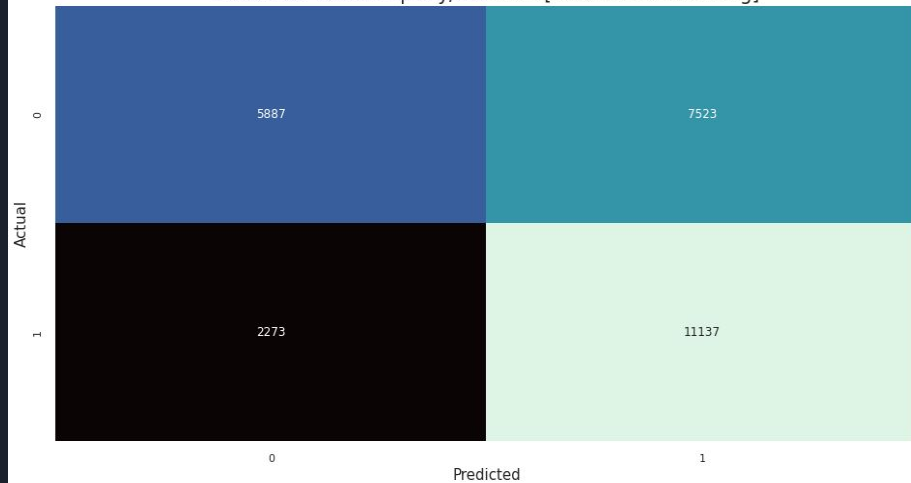
Performance (5 Star)



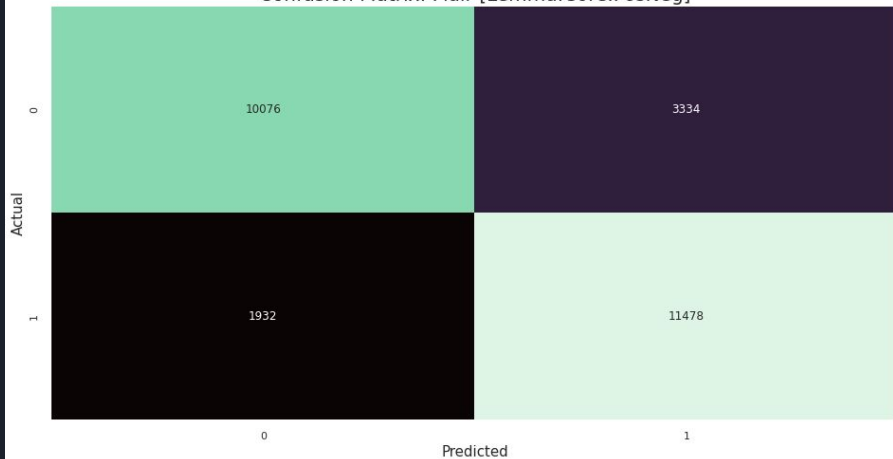
Performance (positive-negative)



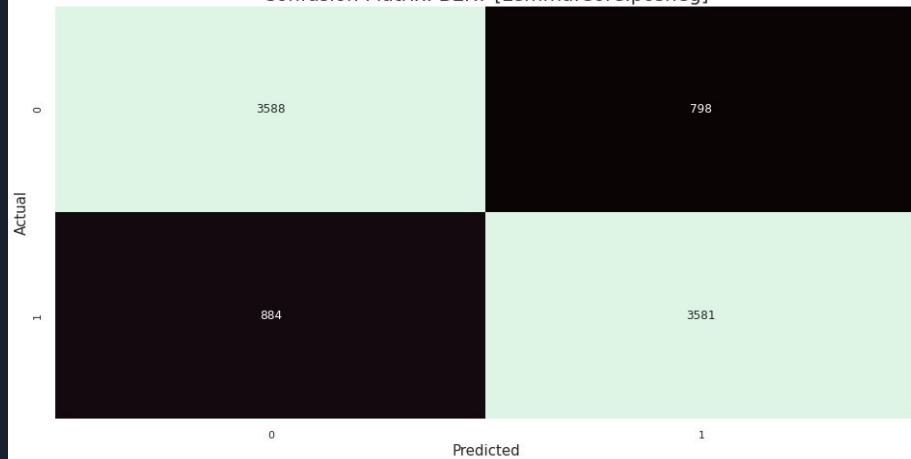
Confusion Matrix: spaCy/TextBlob [Lemma.Core.PosNeg]



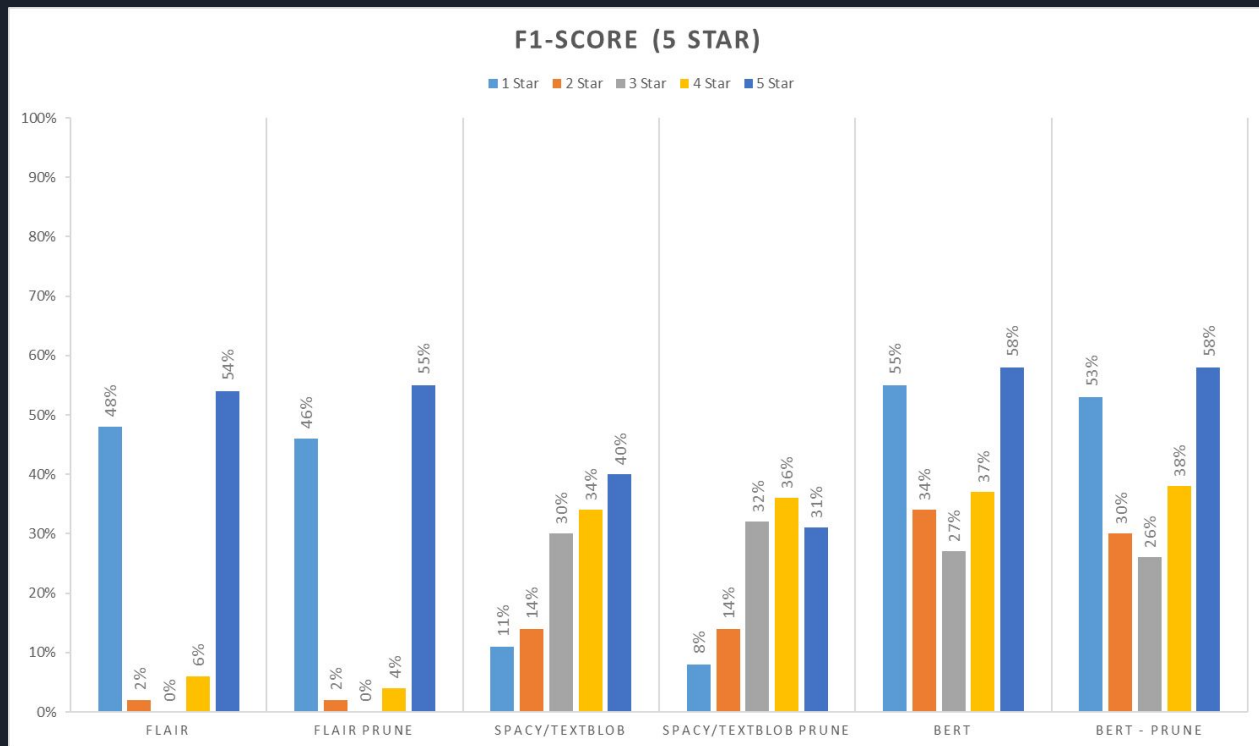
Confusion Matrix: Flair [Lemma.Core.PosNeg]



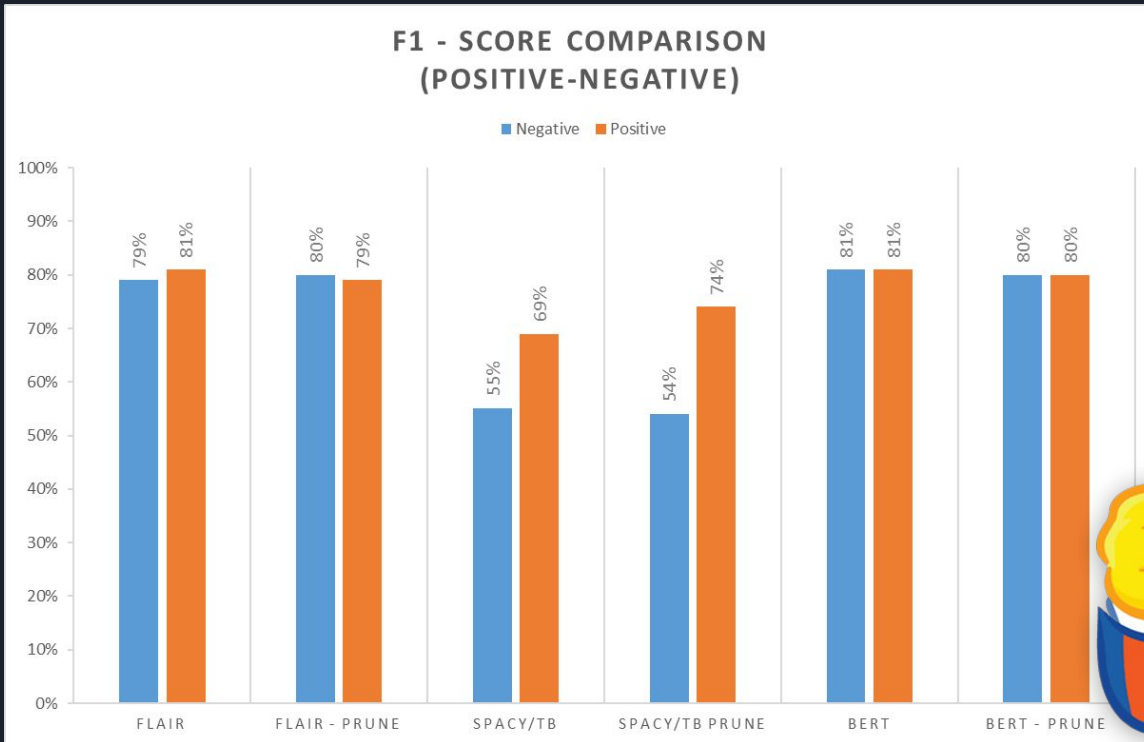
Confusion Matrix: BERT [Lemma.Core.posneg]



F1 Score Comparison (5 Star)



F1 Comparison (Positive-Negative)



Conclusion and Next Steps

More data - Got lowered dramatically once balanced in 5 star
63,310 documents -> < 1000 per classification

More columns analyzed

- verbs, adjectives, nouns, nav

More models applied

Sentence extraction for topic model

- Based on best sentiment



Class Balance -> Data Reduction

