Model Performance

Model Name: ScoreClassificationV15 **Test Date:** 21/03/2022 15:31:26 **Creator:** Giovanni Triulzi

Overview

ML Principle:
Linear Discriminant Analysis

References:

- LDA Doc.
- Stanford NLP Course
- Stanford NLP Lecture
- Engilsh Stopwords

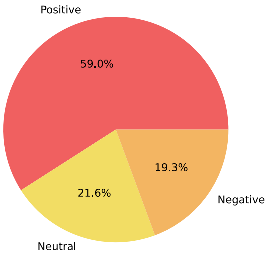
Algorithm Description:
The learning algorithm used in this classification is Linear Discriminant Analysis. This approach was chosen as it is easy to implement and is computational very efficient. The first step in the classification pipeline is removing all stop words for example 'i', 'me', etc. A list of English stop words is provided by the nltk module. Next the sentence is passed through a stemmer and a lemmatizer. Stemming just removes or stems the last few characters of a word, often leading to incorrect meanings and spelling. Lemmatization considers the context and converts the word to its meaningful base form, which is called Lemma. This is done with the SnowBallStemmer and WordNetLemmatizer class from the nltk module. The final preprocessing step is to vectorize the sentence. For this the Tf-idf vectorizer from sklearn is used. If a Tf-idf vectorizer is used the sentences don't have to be tokenized. The sentence is now represented in a numerical feature vector which now can be passed to the LDA classifier.

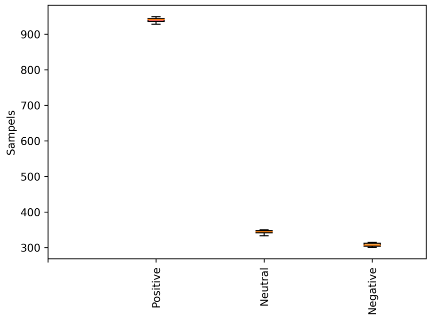
Metrics

Data: ClassifiedDataSetV1.2 with 10 folds cross validation
Split seed: 153.03125

Training Dataset

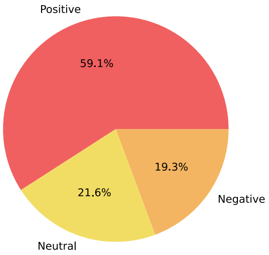
Classes	Number of samples
Positive	940
Neutral	344
Negative	308

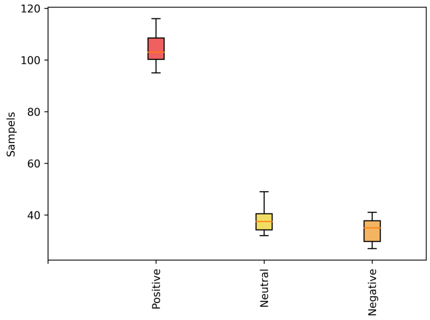


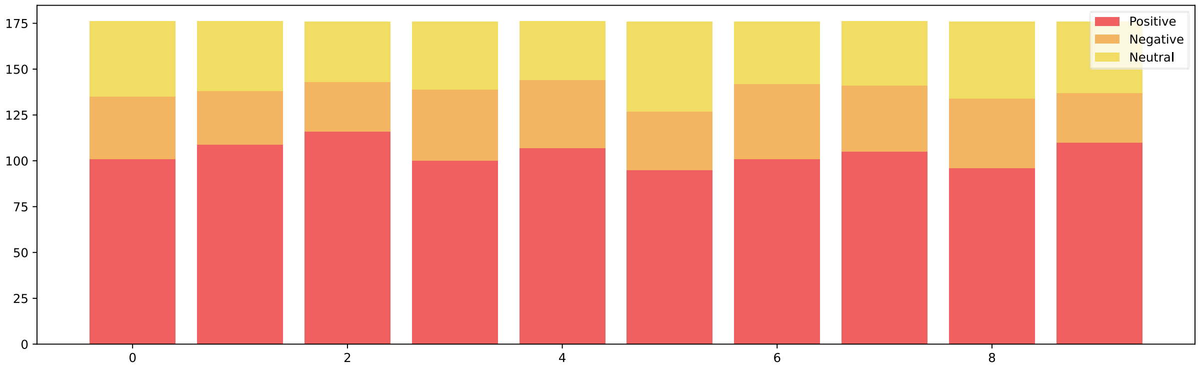


Test Dataset

Classes	Number of samples
Positive	104
Neutral	38
Negative	34

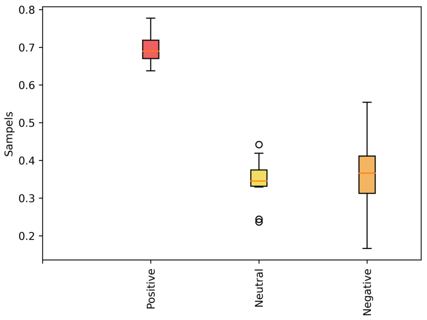




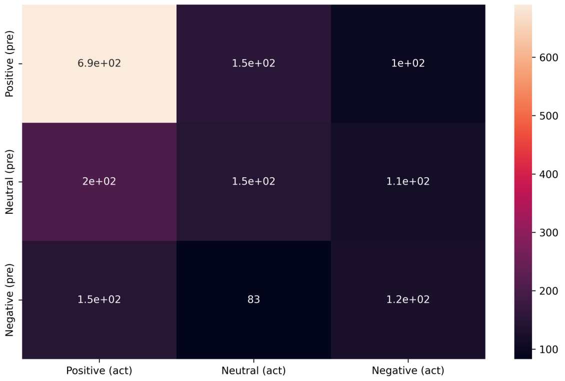


Classification Performance

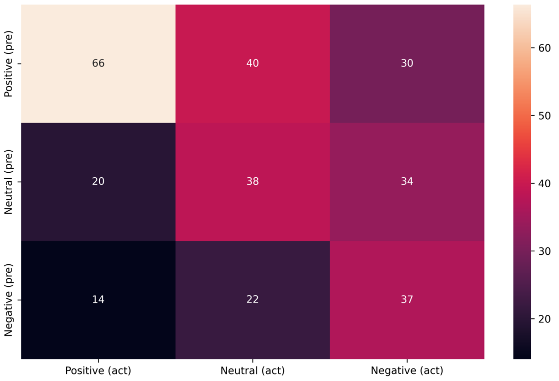
Classes	Precision	Recall	F1 Score
Positive	73.25%	66.35%	69.63%
Neutral	31.47%	38.42%	34.60%
Negative	35.31%	36.76%	36.02%
Accuracy			54.60%
Macro Average	46.67%	47.18%	46.75%
Weighted Average	56.90%	54.60%	55.57%



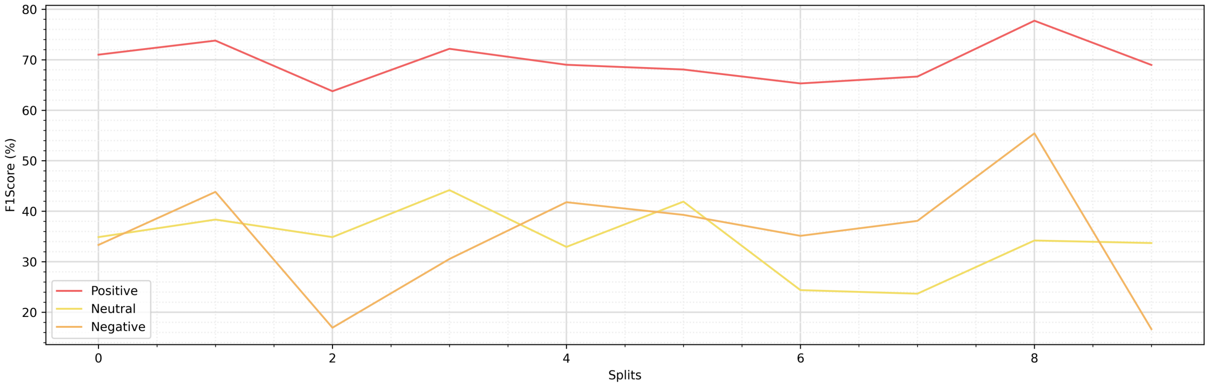
ConfusionMatrix:



Normalised ConfusionMatrix:



F1 Socre by split:


F1Score (%)
Splits
Positive
Neutral
Negative