

Model Performance

Overview

ML Principle:
Support Vector Machine

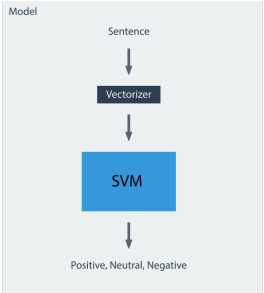
References:

- [Sentiment Analysis SVM](#)
- [Scikit SVM Kernels](#)
- [Scikit feature extraction](#)
- [Scikit Vectorizer](#)

Algorithm Description:

Support vector machines are a robust supervised learning model based on statistical learning. The idea is to find a Hyperplane separating the different classes with the most separation between the closest points. Before the SVM can classify a sentence, the sentence needs to be vectorised. To accomplish the Scikit learn, Tfidf Vectorizer is used. The Vectorizer converts the sentence to a fixed feature vector. With the vectorised sentences, the model can be trained. The best hyperplanes are found in the training step based on the training data. The flexibility of the hyperplane can be defined by the Kernel (linear, sigmoid, RBF). RBF is used for non-linear problems and is also a general-purpose kernel. This model uses a linear kernel.

Model



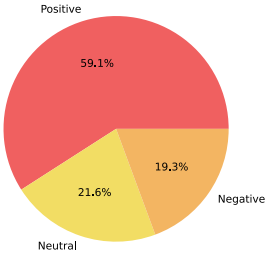
Classification Pipeline

Metrics

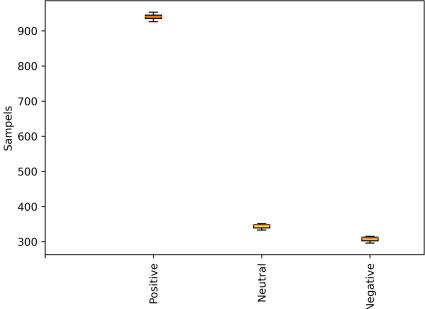
Data: ClassifiedDataSetV1.2 with 10 folds cross validation
Split seed: 4.83819

Training Dataset
(average)

Classes	Number of samples
Positive	940
Neutral	343
Negative	307



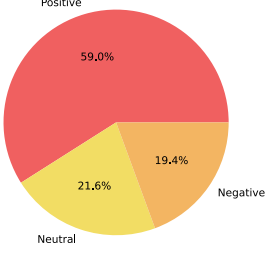
Average distribution of the samples



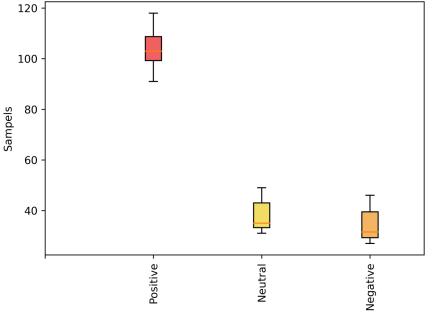
Distribution of the samples contained in each test split

Test Dataset
(average)

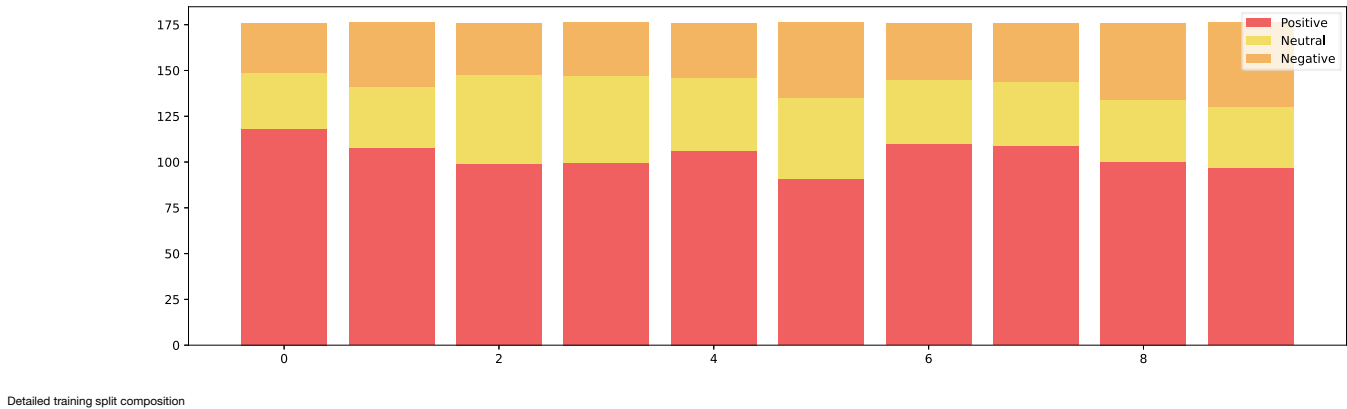
Classes	Number of samples
Positive	103
Neutral	38
Negative	34



Average distribution of the samples

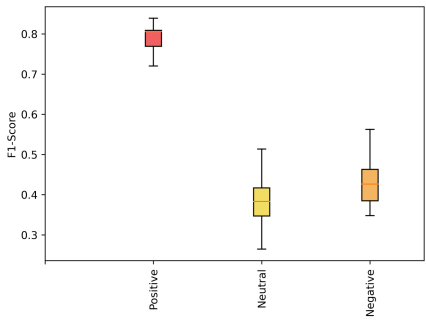


Distribution of the samples contained in each test split



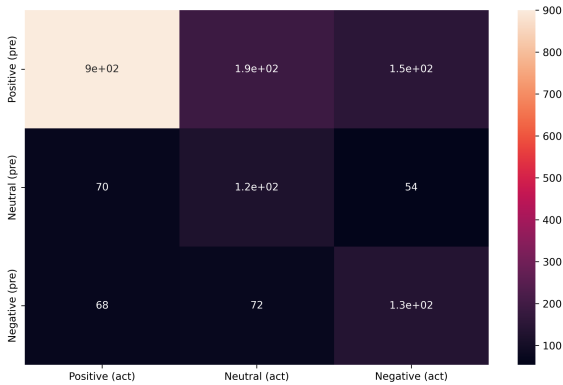
Classification Performance

Classes	Precision	Recall	F1 Score
Positive	72.58%	86.71%	79.02%
Neutral	49.59%	32.02%	38.92%
Negative	48.91%	39.30%	43.58%
Accuracy			65.68%
Macro Average	57.03%	52.67%	53.84%
Weighted Average	63.02%	65.68%	63.47%

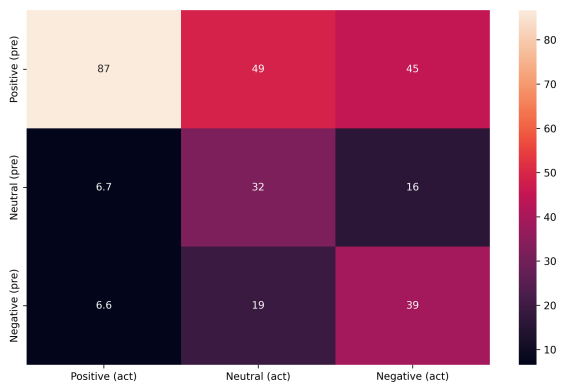


Distribution of the F1-Score

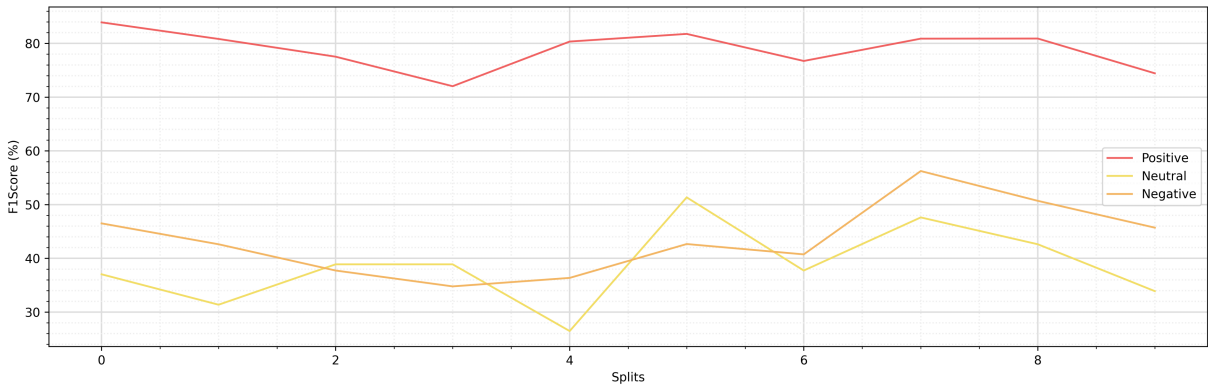
ConfusionMatrix:



Normalised ConfusionMatrix:



F1 Socre by split:



F1-Score per split