

# Classification Report

## LOGISTIC REGRESSION

### Set Exploration

The exploration of the dataset led to the following dates

Dataset	HATESPEECH- EXTEND-DATASET
Size	32,374 sets
Features	1: <code>c_text</code> 2: <code>c(c_text)</code> 3: <code>c(emojis)</code> 4: <code>sentiment_neg</code> 5: <code>sentiment_pos</code> 6: <code>sentiment_neu</code> 6: <code>sentiment_comp</code> 6: <code>profanity_score</code>
Label	1: <code>hatespeech</code> 2: <code>target</code>
Label distribution (Label 1)	0: 13,728 1: 18,646
Label distribution (Label 2)	person: 4,834 group: 3,513 public: 2,018 NaN: 22,009

The data set was then examined for the distribution and frequency of occurring terms for the processing of the text data. Background noise was removed.

### Classification Metrics

After training the classifier in a predefined test-train split, the following metric was determined to evaluate the classifier performance.

Dataset	HATESPEECH- EXTEND-DATASET
training set size	0.8
testing set size	0.2
HATESPEECH METRICS	
5fold-CV-Score	0.840421± 0.006270
execution time	0.192s
TARGET METRICS, BINARY APPROACH	
5fold-CV-Score	0.732± 0.011
execution time	1.648s
TARGET METRICS, CONCATENATED APPROACH	
5fold-CV-Score	0.678± 0.014
execution time	44.967s

### Examined best performing model score

–HATESPEECH –

The LOGISTIC REGRESSION classifier achieved a **placement of 2** in hatespeech classification, which ranks it **2/6 in the overall ranking**.

–TARGET –

The LOGISTIC REGRESSION classifier achieved a **placement of 4** in target classification, which ranks it **4/6 in the overall ranking**.

HATESPEECH CLASSIFICATION RANKING		
1 st.	SVM	0.852
2 nd.	Logistic Regression	0.840
3 rd.	Ridge Classifier	0.836
4 th.	NaiveBayes	0.833
5 th.	RandomForest Classifier	0.826
6 th.	DecisionTree Classifier	0.768

TARGET CLASSIFICATION RANKING		
1 st.	RandomForest Classifier	0.760
2 nd.	Ridge Classifier	0.740
3 rd.	DecisionTree Classifier	0.737
4 th.	Logistic Regression	0.732
5 th.	SVM	0.731
6 th.	NaiveBayes	0.695