

# Classification Report

RIDGECLASSIFIER

## Set Exploration

The exploration of the dataset led to the following dates

Dataset	HATESPEECH- EXTEND-DATASET
Size	32,374 sets
Features	1: c_text 2: c(c_text) 3: c(emojis) 4: sentiment_neg 5: sentiment_pos 6: sentiment_neu 6: sentiment_comp 6: profanity_score
Label	1: hatespeech 2: target
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	0.836± 0.004234
execution time	0.060s
TARGET METRICS, BINARY APPROACH	
5fold-CV-Score	0.740± 0.015
execution time	0.113s
TARGET METRICS, CONCATENATED APPROACH	
5fold-CV-Score	0.678± 0.010
execution time	139.191s

## Examined best performing model score

–HATESPEECH –

The RIDGECLASSIFIER classifier achieved a **placement of 3** in hatespeech classification, which ranks it **3/6 in the overall ranking**.

–TARGET –

The RIDGECLASSIFIER classifier achieved a **placement of 2** in target classification, which ranks it **2/6 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