

PROJECT REPORT ON MALIGNANT COMMENTS CLASSIFIER

Problem statement: The proliferation of social media enables people to express their opinions widely online. However, at the same time, this has resulted in the emergence of conflict and hate, making online environments uninviting for users. Although researchers have found that hate is a problem across multiple platforms, there is a lack of models for online hate detection.

DATA SET Description: The data set contains the training set, which has approximately 1,59,000 samples and the test set which contains nearly 1,53,000 samples. All the data samples contain 8 fields which includes 'Id', 'Comments', 'Malignant', 'Highly malignant', 'Rude', 'Threat', 'Abuse' and 'Loathe'.

EDA(Exploratory Data Analysis):

The data science project on malignant comments classifier is based on the abuse on social media apps.

In the data set there is no missing values.

I did some normalization on Data set.

Model Evaluation:

First Model evaluation:

Logistic Regression:

Accuracy on training set: 0.9592744

Accuracy on Test set: 0.9552765

Decision Tree:

Accuracy on Training Set: 0.99893463

Accuracy on Testing set: 0.9397769

Random Forest Classifier:

Accuracy on Training set: 0.99388

Accuracy on Testing set: 0.951537

KNeighbours Classifier:

Accuracy on Training set: 0.9208945

Accuracy on Testing set: 0.91640207