

Text Classification: 20 Newsgroups

Overview:

The 20 Newsgroups data set is a collection of approximately 20,000 newsgroup documents, partitioned (nearly) evenly across 20 different newsgroups. The 20 newsgroups collection has become a popular data set for experiments in text applications of machine learning techniques, such as text classification and text clustering

In this Hackathon, your task is to build a Deep Learning text classifier capable of classifying text

Data:

Download dataset following from here https://ogrisel.github.io/scikit-learn.org/sklearn-tutorial/tutorial/text_analytics/working_with_text_data.html

Public and Private LeaderBoard

- Your initial responses will be checked and scored on the Public data.
- The final rankings would be based on your private score which will be published once the competition is over.

Evaluation Criteria

Your model performance will be evaluated on the basis of the **F1 Score**.

Rubric

Component	Weightage
Data Cleaning and Data Visualization	25%
Model Building and Evaluation	60%
Pipeline and Deployment (Dashboard/Webapp)	15%