AG News Topic Classification Analysis

Project Description

AG is a collection of more than 1 million news articles that have been gathered from more than 2000 news sources by ComeToMyHead in more than 1 year of activity.

ComeToMyHead is an academic news search engine that has been running since July 2004. The dataset is provided by the academic community for research purposes in data mining (clustering, classification, etc), information retrieval (ranking, search, etc), xml, data compression, data streaming, and any other non-commercial activity.

Dataset Structure
Data Fields

The first column is the Class Id, the second column is Title and the third column is Description.

The class ids are numbered 1-4 where
- 1 represents World,
- 2 represents Sports,
- 3 represents Business
- 4 represents Sci/Tech.

Data Source
Link: Tweet Hate Speech Detection