



NITTE
EDUCATION TRUST

N.M.A.M. INSTITUTE OF TECHNOLOGY

(An Autonomous Institution affiliated to Visvesvaraya Technological University, Belagavi)

Nitte – 574 110, Karnataka, India

(ISO 9001:2015 Certified), Accredited with 'A' Grade by NAAC

☎: 08258 - 281039 - 281263, Fax: 08258 - 281265

Department of Computer Science and Engineering

B.E. CSE Program Accredited by NBA, New Delhi from 1-7-2018 to 30-6-2021

Mini Project Synopsis

Project Title:

“News Category Classification using Naive Bayes Algorithm”

Submitted by:

4NM16CS108 Prem Stevenson Pereira

4NM16CS111 Rahul D Shetty

Submitted to:

Ms. Divya Jennifer D'Souza

Assistant Professor Gd. I

Department of Computer Science and Engineering

Abstract

The news category classification involves identifying the exact group to which an article or a post belongs. This task involves using Natural Language Processing to identify the relevant features and then using them to classify the article. We use the probabilistic approach to solve this problem by using Naive Bayes. The dataset for this problem is collected from Kaggle which is available under Public Domain License. The dataset consists of 2000 samples of text which specifies the title and body of the news article and also the category. There are five categories of news articles which include *sport*, *business*, *technology*, *entertainment* and *politics*.

Dataset Link: <https://www.kaggle.com/yufengdev/bbc-fulltext-and-category>

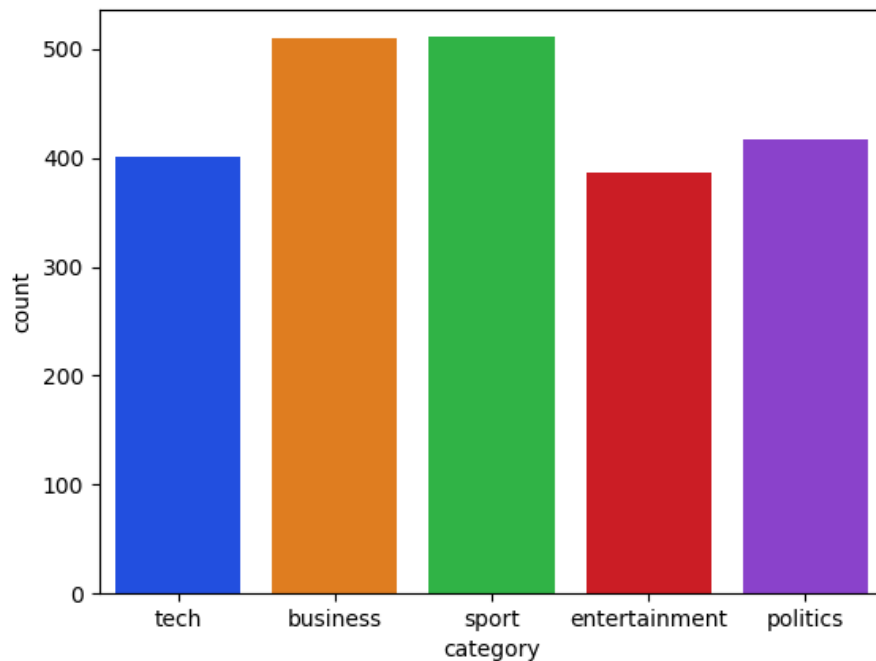


Figure 1: Distribution of Dataset

Figure 1 shows the distribution of dataset into various categories. On the X-axis there are different categories and on the Y-axis it shows the count of each articles.