Scope Database Link: https://sdbindex.com/documents/00000266/00001-01411
Article Link:
https://www.ijircst.org/DOC/2_twitter_data_classification_by_applying_and_comparing_multiple_machine_learning_techniques.pdf

Manuscript ID : 00001-01411

Source ID : 00000266

International Journal of Innovative Research in Computer Science & Technology

Volume 7, Issue 6, November 2019, Pages 147-152, Page Count - 6



Twitter Data Classification by Applying and Comparing Multiple Machine Learning Techniques

Ananya Sarker [(1)] Md. Shahid Uz Zaman [(2)] Md. Azmain Yakin Srizon [(3)]

[(1)] Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Rajshahi–6204, Bangladesh.
[(2)] Department of Computer Science & Engineering, Rajshahi University of Engineering and Technology, Rajshahi–6204, Bangladesh.
[(3)] Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Rajshahi–6204, Bangladesh.

Abstract

Having an average of five hundred million tweets sent out per day, twitter has become one of the largest platforms of data analysis for the researchers. Previously, various researches have been conducted on twitter data i.e., sentimental analysis. However, not much research has been done to classify the tweets in terms of categories so that tweets can be distributed as per user preferences. In this research we started by creating four broad categories: politics, sports, crime and natural. After that, we applied different machine learning techniques (Random Forest, K-Nearest Neighbors, Naïve Bayes, Logistic Regression, Decision Tree and Support Vector Machine) to classify the twitter data. Finally, we compared the results in terms of sensitivity, specificity, precision, false positive rate and accuracy. We found that Support Vector Machine (SVM) produced the best results in terms of sensitivity, specificity, precision, false positive rate and accuracy. Hence, we concluded that a machine learning approach (Support Vector Machine) can certainly be used to classify twitter data. Constructed dataset, all the programs, figures and snippets can be found at https://github.com/ananyasarkertonu/Twitter-Dataset

Author Keywords

Classification, Machine Learning, Social Media, Twitter Data

ISSN Print:
Source Type: Journals
Publication Language: English
Abbreviated Journal Title: IJIRCST
Publisher Name: Innovative Research Publication
Major Subject: Physical Sciences
Subject area: Computer Networks and Communications

ISSN Online: 2347-5552
Document Type: Journal Article
DOI: https://doi.org/10.21276/ijircst.2019.7.6.2
Access Type: Open Access
Resource Licence: CC BY-NC
Subject Area classification: Computer Science
Source: SCOPEDATABASE