

IET Digital Library

[Home](#) > [Conferences](#) > [International Conference on Green Energy, Computi...](#) > [Article](#)

Explainable Twitter bot detection model for limited features

Author(s): [D. Javed](#)¹; [N. Z. Jhanjhi](#)¹; [N. A. Khan](#)¹

[View affiliations](#)

Source: International Conference on Green Energy, Computing and Intelligent Technology (GEn-CITy 2023), 2023 p. 476 – 481

Conference: [International Conference on Green Energy, Computing and Intelligent Technology \(GEn-CITy 2023\)](#)

[Access Full Text](#)

[Recommend Title
Publication to
library](#)



- [« Previous article](#)
- [Table of contents](#)
- [Next article »](#)

Abstract

Twitter provides a rich field for open human conversation, yet it also attracts many fully automated or partially automated accounts "disguised" as human users. These accounts mostly encourage criminal actions, including manipulating ideas and disseminating abusive speech, to mention a few. The malicious information spreads in online discussions, particularly during election seasons, where, aside from lawful bots employed for propagation and communication, the intention is to influence public sentiment and the voters towards a specific direction, philosophy, or political group. Despite implementing various AI-driven bot detection

DOI:

[10.1049/icp.2023.1822](https://doi.org/10.1049/icp.2023.1822)

ISBN: 978-1-83953-931-2

Location: Hybrid

Conference, Iskandar

Puteri, Malaysia

Conference date: 10-12

July 2023

technologies, the basis for bot categorization and generalization remains murky, and AI judgments are deprived of ethical accountability. Most of these techniques use black-boxed algorithms, and their efficacy is frequently questioned. The purpose of this study is to present the creation of a new approach for recognizing Twitter bots using annotated Twitter data. To that purpose, an explainable AI-based machine learning (ML) approach is used, with the hyper-parameters tweaked through cross-validation. Shapley Additive Explanations (SHAP) are also used in our work to explain ML model predictions by assessing feature significance using game theory Shapley values. Experiments show that our proposed model provides excellent results with an accuracy of 87.9% and an AUC-ROC value of 0.94 for the test set.

Format: PDF

Inspec keywords: [social networking \(online\)](#); [learning \(artificial intelligence\)](#); [politics](#); [sentiment analysis](#); [game theory](#); [feature extraction](#)

Subjects: [Game theory](#); [Data handling techniques](#); [Machine learning \(artificial intelligence\)](#); [Information networks](#)

Related content

[Aspect based sentiment analysis using multi-criteria decision-making and deep learning under COVID-19 pandemic in India](#)

- [Rakesh Dutta ; Nilanjana Das ; Mukta Majumder ; Biswapati Jana](#)
- [View description Hide description](#)
- [Abstract](#)

The COVID-19 pandemic has a significant impact on the global economy and health. While the pandemic continues to cause casualties in millions, many countries have gone under lockdown. During this period, people have to stay within walls and become more addicted towards social networks. They express their emotions and sympathy via these online platforms. Thus, popular social media (Twitter and Facebook) have become rich sources of information for Opinion Mining and Sentiment Analysis on COVID-19-related issues. We have used Aspect Based Sentiment Analysis to

anticipate the polarity of public opinion underlying different aspects from Twitter during lockdown and stepwise unlock phases. The goal of this study is to find the feelings of Indians about the lockdown initiative taken by the Government of India to stop the spread of Coronavirus. India-specific COVID-19 tweets have been annotated, for analysing the sentiment of common public. To classify the Twitter data set a deep learning model has been proposed which has achieved accuracies of 82.35% for Lockdown and 83.33% for Unlock data set. The suggested method outperforms many of the contemporary approaches (long short-term memory, Bi-directional long short-term memory, Gated Recurrent Unit etc.). This study highlights the public sentiment on lockdown and stepwise unlocks, imposed by the Indian Government on various aspects during the Corona outburst.

Spammer detection in social networks using deep learning

- K. Sivanagi Reddy ; V. Harshini ; K. Maneesha ; T. Gyaneshwari
- [View description](#) [Hide description](#)
- In contrast to other social media platforms, Twitter is one of the most notorious and substantial social networking services, as we all know. In the course of the discussion, a user might engage in unauthorized behaviour and send unsolicited communications to disturb the dialogue process. As a result of the intricacy of spammers' communications, spam categorization has gotten increasingly difficult. These types of spam communications are tough to detect. In the previous According to published research, over the last 20 years, many efforts have been made to discover and eliminate spammers and fake accounts. SVR, Some of the deep learning algorithms used in this study are Random Forest and Decision tree. In the proposed technique, we use these algorithms on actual Twitter datasets to detect spam tweets. In this instance, spam is broken down into distinct types of spammers: positive, negative, and moderate.

Tweet sentiment analysis using logistic regression

- S. Kumar ; N. Kaur ; Kavita ; A. Joshi
- [View description](#) [Hide description](#)

- Twitter is like a blogging platform where any users can send status and messages, called "tweets," to other people. Provided a large database of so-called emotions, its separation can be done by supervised learning. In order to making supervised learning models, segmentation algorithms need a large set of data with proxy labels. However, categorized data is often hard and costly to get, promoting an unrevealed reading interest. This type of reading uses label less data to be compatible with details given with labeled data in the training program; therefore, it is especially meaningful in programs that include analyzing tweet sentiments, in which a large amount of unspecified data is accessible. While attractive, the slowpaced reading of the tweet emotional analysis is new. We present an exhaustive survey of slow-tracking methods used to tweet segregation. Such methods contain graph-based methods, conclusions and subject-based methods. Comparative study of supported algorithms in self-training, co-training, title modelling, and remote monitoring expose their bias and shed light the factors an expert should consider in real-world systems.

Language-independent features for detecting fake news: a case study of COVID-19 twitter news feed

- W. K. Wong ; J. T. H. Kong ; F. H. Juwono ; R. Reine
- [View description](#) [Hide description](#)
- Social media platforms allow users to create and share information instantly. Due to the large amount of information on the social media platforms, some users may not be able to identify fake news. Fake news may be in a form of misinformation or disinformation and can lead to the loss of public trust on certain issues. It is crucial for social media users to detect the presence of fake news efficiently and in a timely manner. This paper aims to investigate the effectiveness of Support Vector Machine (SVM) classification approach to detect Twitter fake news. It is shown that SVM with linear kernel results in good accuracy of around 84% on independent test dataset. Furthermore, linear kernel function outperforms non-linear and polynomial kernel functions.

Study of Sentiment Classification for Chinese Microblog Based on Recurrent Neural Network

- Yangsen Zhang ; Yuru Jiang ; Yixuan Tong
- [View description](#) [Hide description](#)
- The sentiment classification of Chinese Microblog is a meaningful topic. Many studies has been done based on the methods of rule and word-bag, and to understand the structure information of a sentence will be the next target. We proposed a sentiment classification method based on Recurrent neural network (RNN). We adopted the technology of distributed word representation to construct a vector for each word in a sentence; then train sentence vectors with fixed dimension for different length sentences with RNN, so that the sentence vectors contain both word semantic features and word sequence features; at last use softmax regression classifier in the output layer to predict each sentence's sentiment orientation. Experiment results revealed that our method can understand the structure information of negative sentence and double negative sentence and achieve better accuracy. The way of calculating sentence vector can help to learn the deep structure of sentence and will be valuable for different research area.

Extracting and understanding user sentiments for big data analytics in big business brands

- Jaiteg Singh ; Rupali Gill ; Gaurav Goyal
- [View description](#) [Hide description](#)
- Consumer behavior has become the niche of the market for every user from a manufacturer to a customer. People are fairly good at expressing what they want, what they like, or even how much they will pay for an item. But they are not very good at accessing where that value comes from. Behavior is triggered from sentiments generated in response to an external stimulus. Sentiments and emotions are the subjects of study of sentiment analysis and opinion mining, and this field of study coincides with rapid growth of social media on the web, e.g. social networks, blogs and Twitter, and for the first time, we have huge volume (big data) of data in digital form with us to analyze. Developing algorithms for computers to recognize emotional expression is a widely studied area, and the study of big data analytics and neuromarketing techniques acts as the most powerful tool to develop these algorithms

for better understanding of consumer preferences, purchase behavior and decision patterns. The research aims to extract/read user behavior/sentiment to predict future preferences and to plan the business branding policies. The major objective of this chapter is to perform data analytics of the sample data using Hadoop framework based on crucial metrics related to consumer behavior: (1) customer acquisition cost; (2) customer retention cost; (3) lifetime value; (4) customer satisfaction and happiness; and (5) average purchase amount and behavior. The understanding of these metrics helps in extraction of customer buying trends leading to match the specific customer personas, hence meeting business strategies. The chapter provides a study of user sentiment using neuromarketing techniques and providing data analytics on the user-recorded sentiments based on consumer behavior metrics. The chapter provides an understanding of (1) user sentiments, (2) consumer behavior and neuromarketing process and (3) big data analytics.