

Chapter 1

A Framework for Classifying Imbalanced Tweets Using Machine Learning Techniques

R. Srinivasan

Department of Computing Technologies, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, India

Rajeswari D.

 <https://orcid.org/0000-0002-2677-4296>

Department of Data Science and Business Systems, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, India

ABSTRACT

The research work presented focuses on utilizing social media platforms as a source of data to diagnose depression-related issues. The popularity of social platforms such as LinkedIn, Instagram, Twitter, YouTube, and Facebook, gave researchers an opportunity to analyse user experiences and gain insights into depression. Depression is a significant problem that affects individuals' lives, disrupts normal functioning, and impacts their perspectives. The primary objective of this research is to employ machine learning (ML) approaches for classifying tweets. Additionally, the research addresses the issue of data imbalance by using sampling techniques. This research work utilizes a sampling technique to normalize the dataset. The study explores four techniques that helps to extract meaningful information from the tweets. The research work conducts an empirical study to evaluate the performance of various ML techniques. Based on the experimental results, it is found that the AdaBoost classifier with the BoW feature extraction technique achieves the best results among all the classifiers tested.

DOI: 10.4018/978-1-6684-8306-0.ch001

INTRODUCTION

Depression is cause of disability in every region of the country (Dewa CS et al., 2010). Clinical depression is a lethal medical issue that is associated with a person's mental health (G. Geetha et al., 2020). United Nation (UN) health agency estimates approximately thirty crore people are affected by it, with the highest rates among women, young people, and the elderly (D. M. Shukla et al., 2020). Depression analysis evaluates depression based on social media textual activity, a Natural Language Processing (NLP) domain problem, affecting individuals with symptoms like melancholy, low self-esteem, and poor sleep habits (P. Arora & P. Arora, 2019). The severity of the depression is determined by the symptoms, but it is not applicable to everyone.

The best way to diagnose depression is to consult with a specialist in the field. Many people will wait until their symptoms become severe. The reason for the people become severe is unaware of the symptoms whether they are having depression or not. As a result, modern users use social media platforms to express their opinions, ideas, thoughts, and emotions, allowing them to empathize with their counterparts. The reason for using social media platforms is to express emotions in their native language (S., Soumya & K.V., Pramod, 2020).

The study focuses on utilizing information from social networking sites to evaluate an individual's mental well-being. Previous research suggests that social media platforms predominantly consist of human emotions and thoughts expressed through words like 'happy,' 'sad,' 'excited,' 'nervous,' 'depressed,' and others. These platforms also include self-disclosed phrases indicating mental disorders, lack of confidence, bipolarity, and depression (Islam et al., 2018). By analyzing tweets and short messages from social media, the researchers extract patterns to construct an effective classification model. Previous studies have employed various techniques for text classification, yielding satisfactory outcomes.

The present work offers a comprehensive exploration of different machine learning methods for classifying depression-related tweets. Twitter data, consisting of 80 percent non-depressive tweets, is utilized in this study. The principal conclusions drawn from this research indicates.

- 1) Incorporation of various approaches to preprocess the data extracted from twitter
- 2) Identifying the optimal feature from the dataset
- 3) Smote Techniques are applied to normalize the tweet imbalance problem
- 4) Predict the best ML model to classify this tweets

LITERATURE REVIEW

Many researchers have investigated various approaches to classifying tweets. Recent Studies have proved that depression can be predicted from speech using Machine Learning Techniques. From human speech, the author developed a depression dataset. The author obtained the data from Northwest Normal University (NNU) and used the self-reference effect (SRE) as a conventional paradigm (D. Shi et al., 2021). The author produced an achievable result of 65% using Support Vector Machine (SVM). Shubam et al had recognized the depression from the low level audio, visual and textual process (Dham et al., 2017). The results are obtained from the baseline features for audio, visual (head pose) and textual process. The author proved that notable results are obtained by applying the SVM techniques.

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/a-framework-for-classifying-imbalanced-tweets-using-machine-learning-techniques/327996

Related Content

The Development and Learning Effectiveness of a Teaching Module for the Algal Fuel Cell: A Renewable and Sustainable Battery

Li-Ling Chao, Yu-Der Wen, Pin-Chen Chen, Chung-Chi Lin, Shu-Hua Lin, Chorng-Jee Guo and Wei-Lung Wang (2012). *International Journal of Technology and Human Interaction* (pp. 1-15).

www.irma-international.org/article/development-learning-effectiveness-teaching-module/70759

Planning and Managing the Human Factors for the Adoption and Diffusion of Object-Oriented Software Development Processes

Magdy K. Serour (2009). *Human Computer Interaction: Concepts, Methodologies, Tools, and Applications* (pp. 1937-1955).

www.irma-international.org/chapter/planning-managing-human-factors-adoption/22361

Heterogeneity in Digital Innovation: A Study of the e-Newspaper Case

Jesper Lund (2017). *International Journal of Social and Organizational Dynamics in IT* (pp. 1-16).

www.irma-international.org/article/heterogeneity-in-digital-innovation-a-study-of-the-e-newspaper-case/186754

Web-STAR: Development of Survey Tools for Use with Requirements Gathering in Web Site Development

Jonathan Lazar, Adam Jones and Kisha-Dawn Greenidge (2004). *Issues of Human Computer Interaction* (pp. 37-49).

www.irma-international.org/chapter/web-star-development-survey-tools/24728

Informal Education of Energy Conservation: Theory, Promotion, and Policy Implication

Wang-Kun Chen, Yih-Ruey Juang, Sheng-Hua Chang and Ping Wang (2012). *International Journal of Technology and Human Interaction* (pp. 16-44).

www.irma-international.org/article/informal-education-energy-conservation/70760