Chapter 12
Predicting Depression From Social Media Users by Using Lexicons and Machine Learning Algorithms

Santhi Selvaraj
https://orcid.org/0000-0002-3252-4728
Mepco Schlenk Engineering College, India

S. Selva Nidhyananthan
https://orcid.org/0000-0001-9131-8409
Mepco Schlenk Engineering College, India

ABSTRACT

Depression is one of the most common health issues among individuals. The rate of psychotic treatments is increasing day by day. Depression is created in many ways among the people, especially through work stress, financial burden, unemployment among the adults. Today, the emergence of social media into people's lives makes them expose their feelings and emotions on the social media platforms. The aim of this work is to predict the depressive features from social media users' comments by using machine learning techniques. Multinomial naïve bayes, non-linear support vector machine, and artificial neural network methods are used for classifying the features and comparing it using performance evaluation metrics and get the best classifier. This system includes data pre-processing, feature extraction, data splitting, classification, and performance evaluation. The results show that the proposed system has gradually improved performance accuracy. According to the results, ANN gives 99.19%, the best accuracy compared to other machine learning classifiers.

DOI: 10.4018/978-1-6684-8531-6.ch012

INTRODUCTION

Web 2.0, which is the conceptual and technological framework that allows users to share their ideas, opinions, and thoughts through virtual or social networks and communities, was made possible by the Web 2.0 collection of web-based applications known as social media (Kaplan, 2010). Various social media platforms, such as social networks, media sharing platforms, discussion forums, networks for user reviews, blogs and microblogs, and social commerce platforms are used based on data and applications (Sara Gancho, 2017). Based on Global Statistics (India Social Media Statistics, 2022) report, YouTube, Facebook, Instagram, and Twitter have the largest percentage of users and access rate. These social media platforms are used by millions of people to share information, photographs, news, live audio, and videos (Waseem Akram, 2018).

Trillions bytes of data are shared every second on social media, which can be in structured or un-structured data formats (Stefan Stieglitza et al., 2018) so Big data in social media could be analyzed (Prashant Sahatiya, 2018). Currently, social media is more important in health care applications than in other commercial ones. Health care organizations and medical management have looked into how social media may be used to connect different health informatics systems and develop interactions between patients and clinicians (How to Use Social Media in Healthcare, 2022). The majority of academics are now focusing on social networks analysis, which entails capturing relational or structured data in the form of social entities like people, groups, and organizations with some connections between them (Shazia Tabassum, 2018).

Depression is an extreme mental health problem among the people world-wide relevance of their ages, gender and religion. In this modern world of communication and technology, people feel more comfortable in sharing their thoughts and emotions on social media platforms / social networking sites on a daily basis. Anybody can view a person's mood through their comments, posts, videos and depression level is sensed easily by their comments and posts. This creates an environment that provides additional information regarding the depression disorder among the people. Social Media websites such as twitter and Facebook have become to express peoples' activity and thoughts by creating and posting the chats among the people.

In general, psychologists diagnose depressed humans through face-to-face counselling sessions followed by clinical depression criteria. But in modern world, most of the machine learning, deep learning and text mining methods have been improved to detect latent depression in initial stage itself. In day-to-day life the number of users in social media is increasing so social media platforms act as a reflecting mirror of people's mood. It is also more convenient to find a depressed user based on their comments, posts etc.

Currently, most of the researchers using social media platforms like Reddit, Twitter, Facebook to identify the depression among the people. Numerous research on depression detection has been done by identifying crucial depressive disorder of the social media users with the help of tweets posted by users. The aim of this work, to predict depressed feature through leveraging social media data. In order to find the depression using the user comments on social media, this work tend to use a Lexicon related model (Genghao, 2020).

The comments from the collected dataset contain both depressed and non-depressed words even neutral words, but our work tends to separate those words and train the machine with the correct labels then classify the comments and predict the depressed feature. Hence the overall objective of this work is to propose the model that predicts the depressive, non-depressive and neutral features. Text has been taken as input data for this model. This creates an awareness among the people to know the depressed