



Using Different Machine Learning Algorithms, Depression is Detected

*¹Hatoon AlSagri

*¹Assistant Professor Department of Computer Science and Engineering, Scms School of Engineering and Technology, Karukutty, Angamaly, Ernakulam, Kerala, India.

Abstract

As more and more individuals experience it, the problem of depression in our society is taking on a greater significance. It's a crippling condition that may afflict people of all ages, and while some people may be aware that they have it, others may not. People have begun to use social media as a venue to document their emotional states, and research has been done using machine learning algorithms to identify depression by examining social media posts. It is possible to tell whether a person is depressed or not by examining the data that is available on social media (Twitter). To categorise the data and separate depressive from non-depressive posts, machine learning methods are applied. The suggested system identifies depression using Twitter data and a variety of techniques, including XGBoost, SVM, Logistic Regression, and Random Forest Classifier. The XGBoost algorithm was found to have the best accuracy level when the results were compared based on the highest accuracy level. Real-time classification is introduced with a user interface.

Keywords: Depression, Twitter, XGBoost

Introduction

Systems can improve and improve predictions via experience without being explicitly programmed thanks to machine learning, a subset of artificial intelligence. The algorithms use historical data to forecast future events. The technology is now widely employed in the automation industry by several businesses, organisations, and people. This is because improvements in data processing and access have made it possible for practitioners to produce useful results across a range of industries. Sentiment analysis, or opinion mining, is a branch of natural language processing that aims to identify the sentiment or point of view a writer or speaker is trying to get over about a certain subject. It is used to draw conclusions, judgements, and emotions from text data and can be applied to a variety of tasks, such as text categorization, sentiment analysis, and emotion identification.

The serious functional impairment that depression causes, along with its detrimental effects on interpersonal relationships and general quality of life, make it a huge social problem. People of all ages can be afflicted by this crippling condition, which can also cause physical symptoms like persistent back pain, low mood, guilt-related sentiments, and insomnia. Depression can be fatal if left untreated, and it also raises the possibility of dangerous behaviours like substance addiction and a loss in physical health. In addition, untreated depression increases the chance of suicide.

Machine learning can be used to categorise data into specified categories, allowing computers to teach themselves to get better at classifying data. The objective in this situation is to

utilise machine learning to identify if a person is depressed or not based on their texts. It is possible to tell whether a person is depressed or not by examining the data that is available on social media (Twitter). Machine learning techniques are used to categorise the data and separate out posts that are depressive from those that are not in order to make this distinction. This system under consideration uses Twitter data and a number of algorithms, including XGBoost, SVM, Logistic Regression, and Random Forest Classifier, to diagnose depression.

Objective

Since depression can have a significant impact on a person's interpersonal connections and general quality of life, it is crucial to identify depression in people. The goal of this project is to develop a system and user interface that can analyse social media communications to detect depression in people. Social media is an important part of life in the present digital age and can be a useful tool for assessing someone's emotional health. The technology can identify patterns and indicators that may point to sadness by carefully examining social media posts. Numerous machine learning techniques are utilised for prediction in order to achieve this goal. The pre-processing of the text messages starts the procedure. To train and assess the models, the messages are divided into training and testing data sets. The models are modified during training to correctly distinguish between people who are experiencing depression and those who aren't. The model is evaluated on the testing data after it has been trained to make

sure it is accurate. The system also seeks to identify the algorithm with the highest accuracy. The ultimate result is a system that, based on a person's communications, can reliably assess whether or not they are depressed. This technology has the potential to offer a quick, painless, and economical technique to detect depression in people. The excellent classification accuracy and quick processing times of our solution are its main benefits.

Scope

The System can not only identify people who may be at risk of depression by examining social media posts and looking for trends and signs, but it can also identify depression in its early stages, allowing for early intervention and better outcomes. This kind of technology can also be non-invasive and economical because it depends on readily available data and does not call for physical examination. Additionally, it can be used to track the development of people getting therapy for depression, enabling medical providers to modify treatment regimens as required. Additionally, it can be used to test a sizable population for depression, enabling medical experts to find and treat individuals who require it most.

The diagnosis, detection, and treatment of depression could be enhanced with the use of a machine learning-based depression detection system, leading to better results for those who are suffering from the illness. For better depression diagnosis, this method can be combined with other data sources, such as physiological data. It can also be used in predictive analytics to find those who are more likely to experience depression in the future.

Results

The depression detection system project was effective in creating models that can correctly categorise people as depressed or not. In this study, models were created using a variety of machine learning approaches, such as Logistic Regression, XGBoost Classifier, Support Vector Machine, and Random Forest Classifier. 16,000 text data points from depressed and non-depressed people made up the dataset used to train the algorithms. The algorithms examine the text after receiving a text input for any signs of depressive behaviour and then categorise the user as depressed or not. The models have undergone training, testing, and evaluation. When compared to other algorithms, the XGBoost classifier has the highest accuracy of 97.105%, and this algorithm is used in real-time to assess whether or not a person is sad.

Conclusion

In conclusion, the goal of this study was to create a machine learning model that could recognise depression through the examination of tweets from Twitter. We used a dataset with appropriate classifications that included both sad and non-depressed tweets. Unwanted symbols and emojis were removed, usernames were removed, null rows were removed, and GloVe, lemmatization, and oversampling were among the pre-processing methods. The classification algorithms Logistic Regression, SVM, XGBoost, and Random Forest Classifier were all used. Utilising criteria including the Confusion Matrix, Accuracy Score, and Classification Report, the model's performance was assessed. The XGBoost method was selected as the final model since it was discovered to have the highest accuracy of all the models. The construction of a user interface for the model using the Gradio platform made it simpler to use be improved in numerous ways.

References

1. Park H. An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*. 2013; 43(2):154-164.
2. Kanakaraj, Monisha, and Ram Mohana Reddy Guddeti. "Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques." Proceedings of the 2015 IEEE 9th international conference on semantic computing (IEEE ICSC 2015). IEEE, 2015.
3. Pennington J, Socher R, Manning, Glove: Global vectors for word representation, in: Empirical Methods in Natural Language Processing, 2014, 1532-1543.
4. A gentle introduction to xgboost for applied machine learning," Sep 2016. [Online]. Available: <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/>
5. Debasish Bhattacharjee Victor; Jamil Kawsher; Md Shad Labib; Subhenur Latif; (2020). Machine Learning Techniques for Depression Analysis on Social Media-Case Study on Bengali Community . 2020. Machine Learning Techniques for Depression Analysis on Social Media-Case Study on Bengali Community | IEEE Conference Publication | IEEE Xplore(random forest)
6. Python-Creates UIs for prototyping Machine Learning model with Gradio. Python-Creates UIs for prototyping Machine Learning model with Gradio-GeeksforGeeks
7. Predicting depression using deep learning and ensemble algorithms on raw twitter data Nisha P. Shetty, Balachandra Muniyal, Arshia Anand, Sushant Kumar, Sushant Prabhu.
8. Hatton, Christopher M., *et al.* "Predicting persistent depressive symptoms in older adults: a machine learning approach to personalized mental healthcare." *Journal of affective disorders* 246 (2019): 857-860.