JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# Analysing Depression-Related Sentiments in Social Media Streams using Machine Learning

**[1] Mr. Krishna Kishore Thota, [2] Mrs. Thokala Srivalli**

[1]Assistant Professor, [2]Assistant Professor
[1] Bapatla Engineering College (Autonomous), Bapatla, Andhra Pradesh, INDIA.
[2] Bapatla Engineering College (Autonomous), Bapatla, Andhra Pradesh, INDIA.

***Abstract :*** This project addresses the issue of depressive disorders and aims to overcome the challenges of processing unconventional symbols, stop words, and punctuation. The existing system relies on human intervention for input extraction, resulting in time-consuming and delayed detection of depressive disorders. To mitigate these challenges, we propose a real-time chat application that captures and analyses user interactions, including sentiments. We employ Support Vector Machine, a machine learning technique, to predict individuals' mental well-being by analysing chat content, unusual symbols, stop words, numeric elements, and punctuation marks. Based on the predictions, the system delivers comforting quotes to the individuals and notifies their guardians of any signs of discomfort using SMTP and MIME protocols. This approach significantly reduces processing time and enhances prediction accuracy (95%).

***IndexTerms*** **- MIME, Support Vector Machine, sentiment analysis, SMTP.**

## I. INTRODUCTION

Artificial intelligence (AI) is an expansive field in computer science aimed at creating entities capable of perceiving, reasoning about the world, and emulating human thinking. Within the realm of AI, Machine Learning is a pivotal concept that departs from traditional programming paradigms. Rather than hard-coding every action in our computer systems, we provide them with ample examples of desired behaviour. The computer system then learns to respond appropriately to new, unseen samples.

This paper commences by offering an overview of the diverse topics within this domain, exploring their fundamentals and mechanisms. We will delve into classical AI methods, which predate the emergence of machine learning and are still used for various purposes. However, these classical approaches are gradually being supplanted by more potent machine learning algorithms, which will be the primary focus of this discussion.

Subsequently, we provide an introductory section on neural networks. Later in the paper, we dissect the inner workings of neural networks, elucidate how to construct, train, and test them via code implementation. Notably, we employ Sentiment Analysis, a prominent text mining technique. Sentiments encompass emotions, opinions, likes, and dislikes.

Depressive disorder, a severe form of depression, is characterized by persistent feelings of sadness, hopelessness, and worthlessness. This debilitating condition falls under the category of mood disorders, marked by profound feelings of sadness, loss, or anger that hinder daily functioning. It is not uncommon for individuals to experience both depression and anxiety concurrently.

To address this issue, we propose a real-time application for sentiment prediction in a chat environment. This model captures and stores chat interactions on a server. It leverages Support Vector Machine (SVM), a robust machine learning technique, for swift mental health prediction. SVM excels in real-world classification tasks, aiming to find the optimal hyperplane that separates data points belonging to two distinct classes.

For the purpose of this sentiment analysis, we designate the classes as 'positive' and 'negative'. SVM efficiently predicts the sentiment of individuals, elevating accuracy and expediting the prediction process. Real-time input is seamlessly integrated using Signal R, a chat server, while the SVM technique ensures precise sentiment analysis. The model further facilitates sending consolation quotes based on the analysed results, contributing to improved mental health support.

## II. LITERATURE REVIEW

### *A. Multiclass Detection of Depression in social media Through Sentiment Analysis :*

This study introduces a system that aggregates data from Twitter, a prominent social networking platform, and employs Natural Language Processing (NLP) techniques to extract pertinent features from the tweets. To enhance feature vectors, it incorporates Word Sense Disambiguation and WordNet sunsets. The paper explores the application of various ensemble classification methods to categorize sentiments within the dataset, differentiating between Positive, Negative, and Neutral emotions. A noteworthy challenge addressed is the manual annotation required for Word Sense Disambiguation, a resource-intensive and time-consuming process. Additionally, tokenization is performed to break down extensive data into smaller units, incurring added processing costs.

### *B. Sentiment Analysis of Social Networking Site (SNS) Data for Depression Measurement:*

This research investigates methods for evaluating an individual's level of depression by analysing and extracting emotions from textual content. Employing emotion theories, machine learning techniques, and natural language processing on diverse social media platforms, the study utilizes two sentiment classification techniques: binary classification and multi-class sentiment classification. Naive Bayes, a machine learning model, is applied to both classification methods. However, it is noted that Naive Bayes is limited by class conditional independence, which can lead to reduced accuracy. Practical dependencies exist among variables, such as user profiles (e.g., age, gender), that cannot be adequately modeled by Naive Bayes.

### *C. A Text Classification Framework for Effective Early Depression Detection in social media Streams:*

The paper introduces SS3, a novel supervised learning model for text classification, designed to address Early Risk Detection (ERD) issues. It is highlighted that standard supervised machine learning models are not well-suited for scenarios involving ERD. These models often function as black boxes or lack the requisite support for classification and learning in such contexts. SS3 serves as a flexible framework tailored to handle ERD challenges effectively.

### *D. Deep Learning for Automated Sentiment Analysis of social media:*

This study proposes an innovative sentiment analysis framework that harnesses deep learning models to extract sentiment from social media content. The researchers compile a comprehensive dataset by collecting data from social media platforms, establishing a semantic foundation for further investigation. The paper introduces three deep learning-based models, namely Long Short-Term Memory (LSTM), Bidirectional LSTM (BLSTM), and Gated Recurrent Unit (GRU), to classify sentiment within reviews. Notably, the complexity and dynamic nature of social media data pose challenges in accurately discerning sentiment. This work aims to address these complexities and offer valuable insights into sentiment analysis for social media content.

## III. EXISTING SYSTEM

The current model is designed to identify sentences containing stress or depression-related content through the utilization of machine learning algorithms and an analysis of the sentence's emotional context. It operates as a monitoring system that has the capability to send warning messages to pre-registered individuals. After the selection of relevant sentences, a sentiment metric is employed to gauge the intensity of sentiment, which subsequently serves as input for a recommendation engine.

Communication between the KBRS server and the KBRS client application is established to deliver personalized messages based on the user's profile, ontology aspects, and the sentiment value derived from sentences originating from social networks.

The system's database is populated with data obtained from various Online Social Networks, encompassing 360 messages, with 90 messages for each specific type recommended by the technique user through the recommendation engine. Machine learning techniques are deployed to filter and score sentences, while the eSM2 sentiment metric assigns scores ranging from -5 to +5, determining the message's intensity.

The existing model leverages the Convolutional Neural Network (CNN) model and the soft-max algorithm. The CNN extracts character-level representations from sentences, treating characters as inputs. A convolutional kernel processes the characters within words, while bias nodes in each network layer connect to other nodes, enabling comprehensive computations for error propagation. The LSTM output facilitates both bottom-up and top-down computations, leading to the extraction of disease labels by the BLSTM and the Soft-max output layer, responsible for probability determination and performance comparisons.

## IV. PROBLEM STATEMENT

The challenge lies in the diversity of sentence expressions, with individuals conveying thoughts and opinions in distinct ways. The presence of data elements such as stop words (e.g., "a," "and," "the"), special characters, and numeric values adds complexity to sentiment analysis. Furthermore, the current system suffers from extended processing times and relies on manual intervention to extract and input data. It only dispatches consolation messages to registered users when it detects signs of depression and stress, often at a late stage in their experience.

## V. PROPOSED SYSTEM

Our proposed system introduces a real-time chat application designed to capture user sentiments during interactions, analyse them for depressive and stressful content, and predict their sentiment as either positive, negative, or neutral. This application leverages Support Vector Machine, a powerful machine learning technique, to enhance the understanding of a person's mental state. Once the sentiment is predicted, the system takes appropriate actions, sending consolation quotes to users and notifying relevant organizations of potential mental discomfort through SMTP and MIME protocols.

***Module 1: Chat Application:***

The chat application, built on the asp.net platform, enables user registration and authentication, facilitating interactions with other users. Signal R is integrated into this application, simplifying real-time communication between the server and clients. Data, including chat messages, is stored in a MongoDB database, ensuring seamless communication.

***Module 2: Sentiment Classification:***

Chat messages collected from users are subject to sentiment classification using Support Vector Classification. Our choice for machine learning is ML.NET, a powerful library that seamlessly integrates with .NET applications. It allows for dynamic generation of learning models based on data patterns. Data preprocessing, including normalization and transformation, is conducted to enhance model efficiency. Data is cleaned to remove unwanted elements, such as case variations, diacritical marks, punctuation, and numbers. The transformation process employs automation within ML.NET, streamlining preprocessing steps.

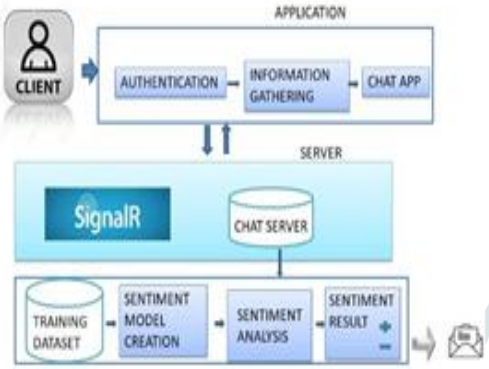


**Fig.1. Architecture Diagram for proposed System**

***Data Transformation***

Data normalization ensures numeric values are represented on a common scale without distorting value ranges. For our text data, text normalization is applied, which includes case conversion, diacritical mark removal, punctuation removal, and number elimination**.**

***Module 3: Intimation***

The system sends appropriate messages based on the analysed sentiment. If the sentiment is positive, indicating normal mental health, no message is sent. However, if the sentiment suggests signs of depression, a warning message is dispatched to alert the user. This process respects user privacy and uses Simple Mail Transfer Protocol (SMTP) to deliver messages via electronic mail. SMTP is part of the application layer of the TCP/IP protocol and operates on port 25, facilitating email transfer. It utilizes a "store and forward" process to manage email transfer between networks, and it supports delayed delivery. A TCP connection is established between the client and server, with the client sending the mail to the SMTP server. The SMTP server listens for incoming connections, and after a successful handshake process, the related message is sent, enabling communication between users and organizations.

## VI. CONCLUSION

This system represents a vital tool for the early support of individuals experiencing depression. By collecting data from social network sites and meticulously preprocessing it to extract meaningful features, the system eliminates stop words, punctuations, unusual symbols, and numeric values. Additionally, complex sentence structures are accurately recognized through a transformative process, enabling the efficient prediction of sentiment intensity. The integration of a real-time chat application allows for the immediate input of user-generated text data, effectively addressing the limitations associated with manual data entry and continuous data feeding. Overall, this system's real-time message intensity detection significantly enhances its capability to identify mental discomfort in individuals at an earlier stage, thereby contributing to timely intervention and support.

**REFERENCES**


**[1]** Y. Tang, Y. Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 39, no. 1, pp. 281-288, Feb 2009.

**[2]** H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, J. Tang, L. Feng, and T. S. Chua, "Detecting stress based on social interactions in social networks," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 9, pp. 1820-1833, Sept 2017.

**[3]** R. Glavan, A. Mirica, and B. Firtescu, "The use of social media for communication." Official Statistics at European Level. Romanian Statistical Review, vol. 4, pp. 37-48, Dec. 2017.

**[4]** Sergio G. Burdisso, Marcelo Errecalde, Manuel Montes-y-Gomez, "A text classification framework for simple and effective early depression detection over social media streams," 2019.

**[5]** Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau, "Sentiment Analysis of Twitter," Department of Computer Science, Columbia University, New York**.**

**[6]** Apoorv Agarwal, Fadi Biadsy, and Kathleen McKeown, "Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams." Conference of the European Chapter of the ACL (EACL 2009), pages 24-32.

**[7]** R. L. Rosa, D. Z. Rodríguez, and G. Bressan, "Sentimeter-br: A new social web analysis metric to discover consumers' sentiment," in IEEE International Symposium on Consumer Electronics (ISCE), Jun 2013, pp. 153-154.

**[8]** M. Al-Qurishi, M. S. Hossain, M. Alrubaian, S. M. M. Rahman, and A. Alamri, "Leveraging analysis of user behavior to identify malicious activities in large-scale social networks," IEEE Transactions on Industrial Informatics, vol. 14, no. 2, pp. 799-813, Feb 2018.

**[9]** H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, J. Tang, L. Feng, and T. S. Chua, "Detecting stress based on social interactions in social networks," IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 9, pp. 1820-1833, Sept 2017.

**[10]** R. Rosa, D. Rodr, G. Schwartz, I. de Campos Ribeiro, G. Bressan et al., "Monitoring system for potential users with depression using sentiment analysis," in 2016 IEEE International Conference on Consumer Electronics (ICCE). Sao Paulo, Brazil: IEEE, Jan 2016, pp. 381-382.

**[11]** G. Y. Tang, Y. Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 39, no. 1, pp. 281-288, Feb 2009.

**[12]** R. Paradiso, A. M. Bianchi, K. Lau, and E. P. Scilingo, "Psyche: Personalized monitoring systems for care in mental health," in 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Aug 2010, pp. 3602-3605.

**[13]** S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Hyett, G. Parker, and M. Breakspear, "Multimodal depression detection: fusion analysis of paralinguistic, head pose, and eye gaze behaviors," IEEE Transactions on Affective Computing, vol. PP, no. 99, pp. 1-1, 2017.

**[14]** R.A. Scholtz, Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher Manning, Andrew Ng, and Christopher Potts, "Recursive Models for Semantic Compositionality Over a Sentiment Treebank," 876-880. [Online]. Available: https://nlp.stanford.edu/sentiment/index.html.