Sentiment Analysis for Depression based on Social Media Stream

A. Punitha, M. Tech[1], K. Kiruthigaipriya[2], B. Vanitha[3] and, U.S. Harshanaa[4]
[1] Faculty, *Manakula Vinayagar Institute of Technology, Puducherry, India*
[2,3,4] Student, *Manakula Vinayagar Institute of Technology, Puducherry, India*

Abstract **- This project deals with depressive disorder. In the existing system challenging to process the unusual symbol, stop words and punctuation and major drawback of this system is aid of human (i.e. person need to extract and feed the inputs to the system). The time consumption of this system is more, the intimation of messages are not done at early stage of depressive disorder. In order to overcome this, we propose a real-time chat application, in which people interaction sentiments are stored, and Support Vector Machine a machine learning technique is used to predict the mental discomfort of person. With this technique, analyzing the posts like chats, unusual symbols, stop words, numeric and punctuation on depressive and stressful content. According to the prediction it sends the consolation quote to respective person and intimation of the person discomfort to their guardian by using SMTP and MIME protocol. The time consumption is reduced and the prediction prompt accuracy (95%).**

Index Terms **- MIME, Support Vector Machine, sentiment analysis, SMTP.**

I.INTRODUCTION

Artificial intelligence (AI) is a very broad field in computer science whose goal is to build an entity that can perceive and reason about the world as well as, the human thinking. Machine Learning concept is the massive extent in AI, tries to move away from this distinct programming of machines. In place of hard coding all of our computer system's actions, we provide our computer systems with many samples of what we want, and the computer will learn what to do when we give it new samples it has never seen before. To initiate the crop field, we'll briefly show the wide variety of topics to get a better result for what are they and how they work. Then, we will scour into more information to discuss really how and why these techniques, algorithms, concepts architecture works. First we will discuss something about the moderate ideas, how AI solved problems before the advent of machine learning. Although those algorithms are still used for other tasks, they have been replaced by more powerful algorithms, and it will be the only time we discuss classical AI. This should encourage the humans to proceed on modern approaches of machine learning. Finally, there will be an additional section towards the end briefly introducing neural networks. Later, we will take apart a neural network to see accurately how they work and write the code to construct, train, and test it from engrave. Sentiment Analysis is a technique widely used in text mining. Sentiments are feeling, opinions, emotions, dislikes or likes.

Depressive disorder is the more severe form of depression. It's characterized by persistent feelings of sadness, hopelessness, and worthlessness that don't go away on their own. Depression is classified as a mood disorder. It may be described as feelings of sadness, loss, or anger that interfere with a person's everyday activities. Depression and anxiety can occur in a person at the same time. In fact, research has shown that over 70 percent of people with depressive disorders also have symptoms of anxiety. In order to overcome this, we propose a real-time application to predict the sentiment from chat application, in which people interaction of messages are stored in chat-server, and also this model includes Support Vector Machine, a machine learning technique is used to predict the mental illness of person with fast processing time. Support Vector Machine is considered a state-of-the-art machine learning technique and is used in many real-life classification problems. SVM is finding the best hyper plane that separates the data points of two different classes. For sake of simplicity and because we are creating a sentiment tool, we call these classes 'positive' and

'negative', and denote them by respectively. This algorithm predicts the sentiment of person in effective way, increases accuracy level, speeds up the process of prediction. The real time input is achieved automatically by using signal R, a chat server, then the support vector machine technique is used for obtaining sentiment and consolation quotes is sent according to analyzed result.

II. LITERATURE REVIEW

A. A Multiclass Depression Detection in social media Based on Sentiment Analysis

This paper proposed a system which collects data from Twitter which is a social network site and it does NLP techniques to extract features out from the tweets. Word Sense Disambiguation and WordNet synsets are added to this feature vector. The various Ensemble methods of classification are finely applied to classify the sentiment of the dataset such as Positive, Negative and Neutral. Word Sense Disambiguation require manually annotated training data for each term that needs to be disambiguated. However, manual annotation is an expensive, difficult and time-consuming process. Tokenization is done for breaking large data into small units which will add additional cost to the processing of a transaction.

B. Sentiment analysis of social networking sites(SNS)Data for Measurement of depression

This paper they presented about how to find the depression level of a person by observing and extracting emotions from the text, using emotion theories, machine learning techniques and natural language processing techniques on different social media platforms. They used two types of sentiment classification they are binary classification and multi class sentiment classification technique. Naive Bayes (machine learning model) which is used in both classification technique. Whereas class conditional independence occurs in Nave Bayes which results in loss of accuracy. Practical dependencies exist among variables (Eg - profile: age, gender etc..) which cannot be modeled by Naive Bayes.

C. A text classification framework for simple and effective early depression tection over social media streams

SS3 was designed to be used as a general framework to deal with ERD (Early Risk Detection) problems. But, most standard and state- of-the-art supervised machine learning models are not well suited to deal with this scenario. In this paper we introduced SS3, a novel supervised learning model for text classification that naturally supports these aspects. This is due to the fact that they either act as black boxes or do not support classification/learning.

D. *Deep learning for automated sentiment analysis of social media*

In this paper they proposed a novel sentiment analysis framework based on deep learning models to extract sentiment from social media. They collected data from which they compiled a dataset and established a semantic dataset for the further research. For analysing comments extracted from social media platforms and to evaluate the effectiveness of semantic embedding for emotion detection they proposed three deep learning-based models to classify review sentiment. The deep learning models are LSTM, BLSTM, and GRU. But, the complexity and dynamic nature of social media data makes it difficult to accurately identify sentiment.

III. EXISTING SYSTEM

This existing model identifies which sentences present a stress or depression content using machine learning algorithms and the emotion of the sentence content. The monitoring system is able to send warning message to people that are previously registered. Later, the selected sentences are analyzed by the sentiment metric and the sentiment intensity is used as input of the recommendation engine. The KBRS server establishes a communication with the KBRS client application, in which the user receives a specific message according to his/her profile, ontology aspects, and the sentiment value calculated from his/her sentences extracted from social network.

The database built from the data captured from Online Social Network. There is a database with 360 messages, 90 messages for each kind to be suggested to the technique user by the recommendation engine. The sentences extracted from the social network they are filtered by machine learning techniques. The sentences are filtered and scored by the eSM2 sentiment metric, from -5 to+5. The sentiment

intensity of the sentence will determine the message intensity.

The machine learning techniques Convolution Neural Network model and soft-max algorithm were used in this work .The Convolution Neural Network computes the character-level representation with characters serving as inputs. The convolutional kernel of the CNN performs the convolutions for the characters of the word. In Hybrid Test Automation Framework, each network layer presents a bias node that is connected to all other nodes. The Hybrid Test Automation Framework is used for the hidden and output layers for calculating the back propagated error. The LSTM output performs the bottom-up and top-down computation, The output vectors of BLSTM are sent to the disease extraction layer to choose the label sequence; the DE represents the disease extraction by the Soft- max output layer. The Soft-max layer is responsible for finding the probability and for performance comparison purposes.

IV.PROBLEM STATEMENT

The set of sentences are challenging to find because people express thoughts differently and also they express their opinion differently. The data like stop words like (a, and the), special character, numeric values are creating difficulties in determining the expressed sentiment. The time consumption of existing system is more and aid of human is need to extract and feed the inputs to the system. It just sent consolation quotes to the registered users when application detect the depressed and stressed signs at late hour.

V. PROPOSED SYSTEM

We propose a real-time chat application, in which people interaction sentiments are stored in chat-server, and also this model includes Support Vector Machine, a machine learning technique is used to handle the mental discomfort of person in a better way. With this technique, analyzing the posts like chats, unusual symbols, stop words, numeric and punctuation on depressive and stressful content and it predict their sentiments like positive, negative and neutral signs. According to the prediction it sends the consolation quote to respective person and intimation of the person discomfort to the organization by using SMTP and MIME protocol.

MODULE -1
CHAT APPLICATION

In module 1, the chat application is developed by using asp.net platform; the first step is to fetches the user details to register. Then authentication is provided for further accessing the chat application to interact with convenient opponent. In this application we included Signal R, which provides APIs for bidirectional remote procedure calls (RPCs) between server and client and abstracts away real-time communication complexities. We exploited the database mongo dB to store the messages and to furnish client server communication setup.

MODULE -2
SENTIMENT CLASSIFICATION

The chats are stored in the chat server which is extract from the users in the previous process. In this process the data are classified by using support vector classification algorithm. We are using ML.NET, makes it possible for .NET developers to easily integrate machine learning into their applications, whether console,desktoporweb.ML.NET is a free-software machine learning library for the C# and F# programming languages. ML.N ET brings model-based Machine Learning analytic and prediction capabilities to existing .NET developers. MLNET gives you the ability to add machine learning to .NET applications, in either online or offline scenarios. With this capability, you can make automatic predictions using the data available to your application. Machine learning applications make use of patterns in the data to make predictions rather than needing to be explicitly programmed Central to ML.NET is a machine learning model. The model specifies the steps needed to transform your input data into a prediction. It covers the full life cycle of ML activity, from training and evaluation of models, to use and deployment. ML.NET makes it is quite easy to perform dynamic generation of learning models, based on information not known at compile time. The first step is to train a model with help of the dataset which was obtained from Github repository. The data contains 10,00,000 texts with three attributes. The input dataset has a string that is user comments label as Sentiment Text and a bool value of either 1 (positive) or 0 (negative) labels as Sentiment. Data may contain unwanted elements that slow down the learning process and reduce the efficiency of the

developed model. Therefore, data is subject to pre-process that partly depends on the data source. There are various processing technique like data cleaning, data integration, data transformation, data normalization and data reduction. Here using data normalization and data transformation.

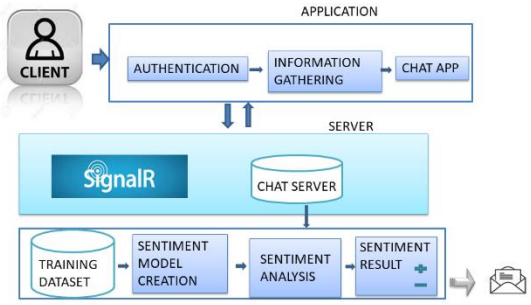

Fig. 2.0 Architecture Diagram for proposed System

DATA TRANSFORMATION

Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to use a common scale, without distorting differences in the ranges of values or losing information. Normalization is also required for some algorithms to model the data correctly. Here we uses Normalize text method which normalizes incoming text, change case, remove diacritical marks, punctuation marks, and numbers of text data may contain unwanted elements that slow down the learning process and reduce the efficiency of the developed model. Therefore, data is subject to preprocessing that partly depends on the data source data may contain unwanted elements that slow down the learning process and reduce the efficiency of the developed model.

Therefore, data is subject to preprocessing that partly depends on the data source. Data transformations are used to prepare data for model training, apply an imported model in Tensor-Flow or ONNX (Open Neural Network Exchange) format, post-process data after it has been passed through a model. Transform a text column into a float array of normalized n-grams and char-grams counts. The Feature Text method which transforms a text column into a feature vector of Single that represents normalized counts of n-grams and char grams, since transformation is automation process in ML.net no need of sequences of steps, transformation and normalization achieved through a single step. Further model trained by using machine learning algorithm support vector machine and binary classification. After the model is trained, the model performance is evaluated. Finally, the model is created the messages are fetched to predict the sentiment intensity of the text by using support vector machine (SVM), the trained model these are used to predicted results. The goal of the SVM algorithm is to create the decision boundary that can segregate n-dimensional space into classes so easily put the new data point in the correct category in the future.

MODULE-3

INTIMATION

The message is sent to the respective person according to the analyzed result. If the analyzed result is positive means (i.e.) when the person mental health is in normal condition, then the message will not be sent. When the result is negative means (i.e.) if result shows any sort of depression signs the message to warn the person against their mental illness is sent. This method is followed for the registered users. This will not disturb the privacy of the user. The messaging process is done through electronic mail. For sending the messages to the user we use Simple Mail Transfer Protocol (SMTP). SMTP is part of the application layer of the TCP/IP protocol and traditionally operates on port 25. It utilizes a process called “store and forward” which is used to orchestrate sending your email across different networks. Within the SMTP protocol, there are smaller software services called Mail Transfer Agents that help manage the transfer of the email and its final delivery to a recipient’s mailbox. It can also support delayed delivery of an email either at the sender site, receiver site, or at any intermediate server.

First the TCP connection is made with the client and server. The client who wants to send the mail opens a TCP connection to the SMTP server and then sends the mail across the connection. The SMTP server is always on listening mode. As soon as it listens for a TCP connection from any client, the SMTP process initiates a connection on that port (25). After successfully establishing the TCP connection the client process sends the mail instantly.

After this successful handshake process, the related message is sent from the user agent with MTA as an intermediate. The MTA maintains a small queue of

mails so that it can schedule repeat delivery of mail in case the receiver is not available. The MTA delivers the mail to the mailboxes and the information can later be downloaded by the user agents. This method is followed in messaging process.

VI. CONCLUSION

In this system is enabled to support the depressed people at earlier stage. This system collects data from social network site and does preprocessing to extract features and removal of stop words, punctuations, unusual symbol and numeric values. The hard terms of sentences are also recognized by transformation and sentiment intensity is predicted in efficient manner. In this system, we developed a chat application to fetch the real time texted data from the user which is feed as input to the trained model, overcomes the drawback of manual feed and regular feeding of data. In the real time detection of message intensity helps to analyze mental discomfort of a person at earlier stage.

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-yGomez, A text classification framework for simple and effective early depression detection over social media streams ,2019.

[5] Data Apoorv, Agarwal Boyi, Xie Ilia, Vovsha Owen, Rambow Rebecca Passonneau, Sentiment Analysis of Twitter Department of Computer ScienceColumbia 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. Rodr´ıguez, 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 sentimentanalysis," 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: Personalised 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 SentimentTreebank"876-880. https://nlp.stanford.edu/sentiment/index.html.