Journal of Engineering Sciences

Vol 11, Issue 4 , April/2020
ISSN NO:0377-9254

SOCIAL NETWORK MENTAL DISORDER DETECTION VIA ONLINE SOCIAL MEDIA MINING USING MACHINE LEARNING FRAMEWORK

Pathan Farjana[1],G. Neha Nikhila[2],K. Khyathi Nandini[3] ,Mohammed Afrin[4] ,Anusha Papasani [5]

[1,2,3,4] IV B.Tech, Department of Information Technology, Vignan's Nirula Institute of Technology & Science for Women, Peda Palakaluru, Guntur-522009, Andhra Pradesh, India.

[5]Asst.Professor, Department of Information Technology, Vignan's Nirula Institute of Technology & Science for Women, Peda Palakaluru, Guntur-522009, Andhra Pradesh, India.

anoosha.papasani@gmail.com

ABSTRACT

Present days attractiveness of social networking sites indications to the problematic habit. For this reason, researchers devised stress detection systems based psychological disorders in social networks. In this work, we propose a system of mental disorders detection (MDD) that can provide online social extraction. It offers an opportunity to identify disorder at an early stage. These MDD system are made a different and advanced for the preparation of disorder detection. Propose system a machine learning approach that is detection of psychological disorders in social networks and social interaction features from social network data for detect with precision possible cases of disorders detection. Development in social network communication encourages dangerous use. An increasing number of mental disorders in social networks, such as cybernetic dependency, information overload and network constraint, have recently been observed. Currently, the symptoms of these mental illness are passively observed, which causes late clinical intervention. In this paper, we argue that online social behaviour mining offers the chance to effectively recognize mental disorder at a beginning period. It is difficult to detect disorder because the mental state cannot be observed directly from the registers of online social activities. Our new and innovative approach to the act of disorder detection is not based on self-disclosure of these mental factors through psychology questionnaires. we argue that the potential Social Network Mental Disorder (SNMD) users can be automatically identified and classified into various categories like Virtual Relationship Addiction, Obsessive Online Gambling and Information Glut using SNMD based tensor model, with the data sets collected from data logs of various Online Social Networks (OSNs). Major findings revealed that the most frequently used method in mental health detection is machine learning techniques, with Support Vector Machine (SVM) as the most chosen algorithm. Meanwhile, Twitter is the major data source from OSN with English language used for mental health detection. The researcher found a few challenges from the previous studies and analyses, and these include limitations in language barrier, account privacy in OSN, single type of OSN, text analysis, and limited features selection. Based on the limitations, the researcher outlined a future direction of mental health detection using language based on user's geo-location and mother tongue. The use of pictorial, audio and video formats in OSN could become one of the potential areas to be explored in future research. Extracting data from multiple sources of OSNs with new features selection will probably improve mental health detection in the future by Logistic Regression algorithm . In conclusion, this research has a big potential to be explored further in the future.

Keywords: Online social networking sites (OSN), mental disorder detection(MDD), feature extraction, Logistic Regression algorithm.

1. INTRODUCTION

"Mental psychological disorder is becoming a threat to people's health today with the rapid pace of life, more and more people are mentally disturbed [1] [2].

It is not easy to detect the mental disorder of the user at an early age to protect it with the fame of web-based social networks, people are used to sharing their daily activities and interacting with friends through the web-based network media phases, making it possible to use online social network data for identification of mental disorders [3]. In our system, we have discovered that the state of user disruption is closely related to that of their friends in social networks and we use a large-scale set of real social stages to methodically examine the connection of client disturbance states [4] [5]. Interactions first of all we define a set of textual, visual and social attributes related to the mental disorder from various aspects [6]. Fast pace of life, progressively and more and more people feel stressed. Although mental disorder is not clinical and is common in our lives, excessive and chronic disorder can be very detrimental to people's physical and mental health [7]. The social interactions of users in social networks contain useful indications for detecting disorder [8]. With the explosive growth in popularity of social networks, messaging applications and online social networks (OSN) they have become part of the daily life of many peoples [9]. Psychological stress is turning into a risk to individual's well-being these days. The research on the extraction of social networks focuses on the discovery. The knowledge behind the data to improve people's life while OSNs seem to expand the capacity of their users increasing social contacts can actually diminish Interpersonal interactions face to face in the real world [10],[11].Psychological stress is a leading cause of several psycho physiological disorders. For example, it increases the likelihood of depression, stroke, heart attack and cardiac arrest [12] [13]. User social interactions on social networks contain valuable prompts for pressure identification. It is not easy to detect user psychological disorder in an early time to protect user.

Social psychological studies have made two interesting observations [14]. The first is contagion of the mental state: a bad mood can be transferred from one person to another during social interaction [15] [16]. The second social interaction: people are known for the social interaction of the user. The progress of social networks like Twitter and Facebook a growing number of people will share their events and moods every day and interact with friends through social networks [17] [18]. We can classify using the machine learning framework because of the use of the content attributes of Facebook publications and social interactions to improve the detection of mental disorders. Meanwhile, the researchers also should be aware of the changes of account privacy policy by the OSN service providers before data extraction to avoid any difficulty in the future [19]. The format of data extraction also needs to be widely explored and should not be limited to text format extraction. The contribution of this paper are as follows:

- Today online disorders are usually processed late. To actively identify possible cases of disorders, we propose an innovative approach, disorder detection practice, extracting data records from OSN users as an early detection system [20].
- We develop a machine learning framework to be detected Mental disorders detection(MDD). We also design and analyse many important features to identify OSN disorders, such as disinhibition, parasocality, self-revelation, etc. The proposed framework can be implemented to provide a timely alert for potential patients [21].
- To improve the accuracy using advanced Logistic Regression algorithm and it is simple to recognize [22]. The proposed framework provides a better results and make predictions for these reasons alone you should take a closer look at the algorithm [23].

2. RELATED WORK

Literature survey is the most important step in any kind of research. Before start developing we need to study the previous papers of our domain which we are working and on the basis of study

we can predict or generate the drawback and start working with the reference of previous papers.

Sr No.	Author	Paper Name	Method	Result
1	B.Saha et al. [1]	Psychologic al stress detection from Cross-Medi a Microblog data using deep sparse neural network	Use cross-media microblog data with Deep sparse neural network	Accuracy- SVM-85.97, SAE-89.68, LAE-90.55
2	Chun-Hao Chang et al. [2]	Quantitative study of Individual emotional states in social networks	SVM and Naïve Bayes with Mood cast	Accuracy- 62.17
3	Lakshman Narayana et al. [3]	Social Media as a Measuremen t tool of Depression in population	SVM classifier, Center for Epidemiologic studies Depression scale(CES-D), social media depression index(SMDI),	SMDI can nearly reflect CDC characterized insights on depression.
4	Patibandla R.S.M.L et al. [16]	Psychologic al stress detection from Cross-Medi a Microblog data using deep sparse neural network	Use cross-media microblog data with Deep sparse neural network	Accuracy- SVM-85.97, SAE-89.68, LAE-90.55
5	K. Santhisri et al. [20]	Quantitative study of Individual emotional states in social networks	SVM and Naïve Bayes with Mood cast	Accuracy- 62.17
6	Vellalacheruvu Pavani et al. [37]	Social Media as a Measurement tool of Depression in population	SVM classifier, Center for Epidemiologic studies Depression scale(CES-D), social media depression index(SMDI),	SMDI can nearly reflect CDC characterized insights on depression.

Table 1: Comparative Study of Approaches Used For Detecting Depression

Lakshman Narayana Vejendla et al. [11] have introduce a research on mental disorders on social media networks. Author used predictive models for data collection process, using subconscious crowdsourcing. This study, Patibandla, R.S.M.L et al. [18], the proposed technique can help to classifying online mental health related communities. Author are extract two type of feature 1) STL (single task learning 2) LIWC (linguistic inquiry and word count) features from online social media for depression patients to analyze their outlines.

K.Santhi Sri et al. [21] have introduced about the study of evolution of users activities on facebook. In the social network of Facebook to capture this notion. Author find that links in the activity network tend to come and go. Quickly over time, and the strength of the bonds exhibits a reduce the decreasing tendency of the activity as a connection of a social network centuries. For example, only 30 percent of Facebook user pairs interact. It is interesting to note that also find this, even if the connections of the activity network many properties of graph theory change rapidly over time. The network of activities remains unchanged. Anveshini Dumala et al. [33] introduced about a programmed pressure identication technique from cross-media smaller scale blog information.

They implement three-level system for stress discovery from cross-media small scale blog information [24] [25]. This paper used by joining a Deep Sparse Neural Network to consolidate distinctive highlights from cross-media small scale blog information, the structure is very practical and effective for stress identication [26]. This structure, the proposed technique can help to consequently identify mental worry from informal communities. The future extension intend to research the social relationships in mental worry to additionally enhance the identication execution.

R S M Lakshmi Patibandla et al. [27] have introduced to contemplate about spanning the vocabulary hole between wellbeing searchers and medicinal services information with a worldwide learning approach [28]. This paper restorative wording task plan to connect the vocabulary hole between wellbeing searchers and human services information. Author includes two segments, nearby mining and worldwide learning. Extensive assessments on a genuine world dataset show that our plan can create promising execution when contrasted with the predominant coding techniques [29] [30]. This calculation sums up to inconspicuous labels, and is additionally enhanced joining tag-connection highlights got by means of ICR. Author used methods to all the more likely

1) consolidate multi-word terms.
2) out-of-vocabulary words.
3) Propelled NLP procedures for taking in word relations from freestyle content.
4) Assessment of idle idea connection proposal.
5) Anticipating the sort of relations.

Introduced a novel issue of feeling forecast in informal communities. This paper used a technique alluded to as Mood cast for demonstrating and anticipating feeling elements in the informal community[31] [32]. The proposed methodology can successfully demonstrate every clients feeling status and the expectation execution is superior to a few gauge strategies for feeling forecast [33]. It is utilized to because of the predetermined number of members. This framework trial results on two diverse genuine interpersonal organizations exhibit that the proposed methodology can successfully demonstrate every clients feeling status and the forecast execution is superior to a few standard techniques [34].

In cyber relationship obsession which includes the obsession with social media surfing to converse and share private information to the public where online relationships became more significant than friends and family circles [35] [36]. In Information Overload comprises obsessive scanning of user status tweets posts which leads minimal in person interaction and vast reduction in work level [37]. In Web addiction which includes obsessive online social gaming and gambling which affects ones profession.

3. EXISTING METHODOLOGY

In Existing system we develop new approaches for detecting psychological disorder cases of OSN

(online social network) users [38]. We claim that mining social network data of individuals, as an opposite another to the predictable psychological approach, delivers an outstanding opportunity to actively identify those cases at an early stage [39]. In this paper, we develop a machine learning framework for detecting PDD (psychological disorder detection) users, that is to say Social Network Mental Disorder (SNMD).This work aims to build a framework for detecting psychological disorders in social media users. We pursue to accomplish our complete method through:

- Collection of Data
- Cleaning and preprocessing of Data.
- Extracting Features

Through manipulating machine learning techniques with the ground truth found via the current analytical repetition in Psychology, we extract and analyse several features of different categories from OSNs, including parasocial relationships, online and offline interaction ratio, social capital, dis-inhibiting, self-disclosure, and bursting temporal behaviour. These features capture important factors or serve as proxies for disorder detection [40] [41].

A lot of work has been done in this field thanks to its extensive use and applications [42]. This section mentions some of the approaches that have been implemented to achieve the same purpose [43] [44]. These works are mainly differentiated from the algorithm for mental disorders detection systems . Existing jobs have shown that social networks can be used for health care and, in particular, for the detection of mental disorders [45] [46]. There are some limitations in Facebook content based on the detection of mental disorders. Users do not always express their stressful states directly in the Facebook post. Although no disturbance of the publication itself is revealed, from the interactive follow-up comments made by the user and his friends, we can discover that the user is really stressed at work.

General Architecture of Mental Health Detection

Numerous research works have applied a few steps for mental health detection in OSN. The researcher believes that this general architecture is a common step of future research implementation in mental health detection. The general architecture for mental health detection consists of several steps such as social network, data extraction using the keyword, data pre-processing, features selection, data classification using machine learning algorithms and early mental health detection [47] [48].

The most previous research found the OSN service providers as data input. Then, for the process part, a few steps were implemented such as data extraction based on keyword. After extracting the data, the pre-processing data must be conducted to eliminate the outliers before the features selection steps. The final process is modeling the data using the machine learning technique.

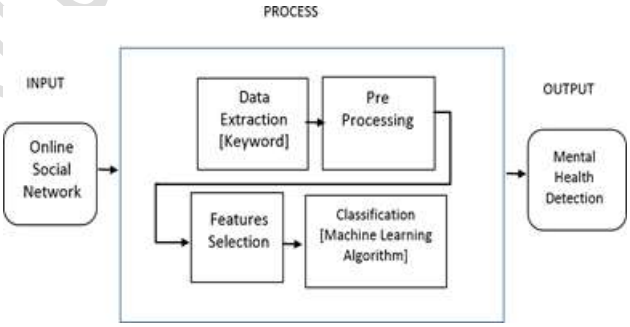

Fig 1 General architecture for mental health detection

Social network mental disorder Identification

In this we analyze data mining techniques to identify three types of SNMDs :

Virtual Relationship Addiction, which includes the addiction to social networking, checking and messaging to an extent where virtual and online friends become

more important than real-life relationships with family and friends.

Obsessive Online Gambling, which includes compulsive online social gaming or gambling, often leading to financial and job-related problems.

Information Glut, addresses how the information technology revolution would shape the world, and how the large amount of data available on the Internet would make it more difficult to sift through and separate fact from fiction. Accordingly, we formulate the detection of SNMD cases as a classification problem. We detect each type of SNMDs with a binary SVM [49] [50]. In this study, we propose a dual-phase framework, called Social Network Mental Disorder Identification (SNMDI). The first phase extracts various discriminative features of users, while the second phase presents a new SNMD-based tensor model to derive latent factors for training and use of classifiers built upon SVM .Two major challenges in the design of SNMD are:

We are not able to directly extract mental factors like those extracted via questionnaires in Psychology and hence need new features to learn the classification models.
We aim to exploit user data logs from multiple OSNs and thus need new techniques to integrate multi-source data based on SNMD characteristics.

Lack of mental features. Psychological studies have shown that many mental factors are related to SNMDs, e.g., low self-esteem [10], loneliness [11]. Thus, questionnaires are designed to reveal those factors for SNMD detection. Some parts of Psychology questionnaire for SNMDs are based on the subjective comparison of mental states in online and offline status, which cannot be observed from OSN logs. As it is difficult to directly observe all the factors from data collected from OSNs, psychiatrists are not able to directly assess the mental states of OSN users .

Heavy users vs. addictive users. To detect SNMDs, an intuitive idea is to simply extract the usage (time) of a user as a feature for training SNMD. But, this feature is not sufficient because i) the status of a user may be shown as “online” if she does not log out or close the social network applications on mobile phones, and ii) heavy users and addictive users all stay online for a long period, but heavy users do not show symptoms of anxiety or depression when they are not using social apps. To distinguish them by extracting discriminative features is critical.

Multi-source learning with the SNMD characteristics. As we intend to exploit user data from different OSNs in SNMD, extracting complementary features to draw a full portrait of users while considering the SNMD characteristics into the tensor model is a challenging problem.

To address the these challenges, we consider a number of factors to understand the mental states of users, e.g., self- esteem and loneliness. The goal is to distinguish users with SNMDs from normal users. Two types of features are extracted to capture the social interaction behavior and personal profile of a user. It is worth noting that each individual feature cannot precisely classify all cases, as research shows that exceptions may occur. Therefore, it is necessary to exploit multiple features to effectively remove exceptions.

Effective features as proxies to capture the mental states of users

A fundamental problem in text data mining is to extract meaningful structure from document streams that arrive continuously over time. Newsfeeds, messages exchanged, posts shared on an individual’s wall are all the natural examples of such streams, each characterized by topics that appear, grow in intensity for a period of time, and then fade away. The published literature in a particular research field can also be seen to exhibit similar phenomena over a much longer time scale. Underlying much of the text mining work in this area is the following intuitive premise that the appearance of a topic in a document stream is signaled by a “burst of activity," with certain features rising sharply in frequency as the topic emerges. The human appetitive system is in charge of the addictive behavior. A recent study has shown that social searching (actively reading news feeds from friends’ walls) creates more pleasure than social browsing (passively reading personal news feeds) [12].

This finding indicates that goal-directed activities of social searching are more likely to activate the appetitive system of a person as drug rewards do, and it is more related to SNMDs because the appetitive system is responsible for finding things in the environment that promote species survival(i.e., food, sexual mates) and thus is inclined to form addictive behavior after several rewards. While users with SNMDs perform social searching more frequently than non- SNMDs, it is not easy to distinguish these two behavior on social media. This example is just one such kind of a feature that could be used to analyse a user's social interaction and personal features. The new system will have many more similar features that are exploited to understand the mental status and habits of a SNMD user that considers online/offline interaction ratios, the temporal behavior, his self-obsessive characteristics hinting the possibility of SNMD.

The disadvantage of this system is that users may behave differently on different OSNs i.e if the user is not equally active in all OSNs, resulting in inaccurate SNMD detection. Also it has not explored the sentiment as feature which can contribute to the accuracy

4. PROPOSED SYSTEM

Sentiment Analysis

Utilizing the techniques from NLP, sentiment analysis field looks at users' expressions and in turn associate emotions with what the user has provided. The cultural norms add a different twist to this area. For example, the following statement could be interpreted very differently.

"This new gadget is bad!"

While the obvious meaning alludes to the user's dislike of the gadget, user community belonging to a certain age group would consider the above statement as a resounding endorsement of the gadget at hand. Furthermore, the sentiment analysis looks at the time at which the user expressed the sentiment or opinion. The same user can be under certain stressors which can cloud their judgment and hence gathering statements on a time continuum can provide better assurance of the sentiments expressed.

Moreover, the data can be gathered for a given interval that can prove vital in ensuring consistency. The data gleaned in this manner will offer a preponderance of evidence supporting the researcher's hypothesis and can provide a solid foundation for scientific deductions. Gathering data from the web has become the choice of many fields such as Marketing etc. Google, YouTube and Amazon are examples of how companies can provide customized content to the end-user. Many such fields can depend greatly on objective metrics such as number of likes, total number of items sold given an age range etc. However, the fields of psychology/psychiatry do not have such luxury as the data is in form of text written by users on the various media such as blogs, social media etc. This dimension adds more complexity due to a) use of different languages on a certain topic/blog, b) use of non-standard words that cannot be found in a dictionary and c) use of emojis and symbols. These questions are tackled by experts in the Natural Language Processing (NLP) domain along with those working in the sentiment analysis area.

There is a need of providing social scientists and psychiatrists the requisite vocabulary and the basic tools to scavenge data from the web, parse it appropriately and glean the contextual information. This work is intended to be a step in this direction. Specifically, the paper provides the following:

1. Provide a basic understanding on various prevalent theories in Natural Language Processing (NLP)
2. Explain the traditional and the statistical approaches to NLP
3. Look at the work done in the area of sentiment analysis and the challenges faced in the light of mental health issues
4. Present a brief synopsis of various applications in applying NLP concepts to mental health issues and sentiment analysis 2

Natural Language Processing

The field of NLP dates back to few decades and has matured quite significantly over the years. Initially confined to gathering data from a limited set of digitized documents, the advent of World Wide Web saw an explosion in information in many different languages. Significant amount of work was done in

the information Retrieval (IR) field which is considered an application of the Natural Language Processing domain. Before discussing the IR techniques, a bit more, let us delve into the theoretical and practical aspects of NLP.

Initially, the NLP approach followed the following discrete steps.

1. Text Preprocessing/Tokenization
2. Lexical Analysis
3. Syntactical Analysis
4. Semantic Analysis

Preprocessing algorithm

It is a technique that is used to translate the raw data into a clean data set. Each and every time the data is collected from different sources it is collected in fresh format which is not achievable for the study. For achieving better results from the realistic model in Machine Learning developments the format of the data has to be in a proper manner. So in data preprocessing is required because of the presence of unformatted real world data. In this algorithm, we will discuss the some steps involved in text processing.

1.Stop word Removal

In this process stop words are words that are mostly common in a text body and thus considered as rather un-informative (e.g., so, and, or, the...).One approach to stop word removal is to search against a language-specific stop word dictionary. Another approach is to create a stop list by sorting all words in the entire text body by regularity. This stop list after conversion into a set of non-redundant words is then used to remove all those words from the input documents that are classified between the top n words in this stop list.

The algorithm is implemented as below given steps.

1. In document text is tokenized and separable words are put in array.
2. A single stop word is read from stop word list.
3. The stop word is matched to goal text in form of array using sequential search technique.
4. If it equals, the word in array is removed, and the evaluation is continued checkout length of array.
5. After removal of stop word totally, an additional stop word is read from stop word list and again algorithm go to step

The algorithm runs continuously up to all the stop words are compared.

Follow-on text empty of stop words is displayed, also mandatory data like stop word removed, no. of stop words removed from goal text, total count of words in target text, count of words in resultant text, separate stop word count originate in goal text is displayed.

2. Tokenization

In tokenization defines the common process of breaking down a text body into separate features that help as input for various natural language processing algorithms. Follows, tokenization processing steps:

1. Segmenting act of breaking up a sequence of series into parts such as phrases, words, keywords, symbols and other features called tokens
2. Tokens or words are separated by whitespace, punctuation marks or line breaks.
3. White space or punctuation marks may or may not be included depending on the need.
4. The tokens become the input for another process like parsing and text mining.

3. Stemming

Stemming is a procedure where words are reduced to a root by removing inflection through reducing unnecessary characters, usually a suffix. The results can be used to identify relationships and commonalities across large datasets.

1. A stemming algorithm is a process of linguistic normalization, in which the variant forms of a word are reduced to a common form for e.g. - 1. Played-play 2. Clustering–cluster
2. After stemming find eventually is that you can be improving performance of the language, while producing a parallel degradation of performance in another area.

Classification Algorithm:

This section will introduce some of the main concepts and procedures that are needed to apply the classification tasks. We have to used Logistic Regression algorithm is easy to build and particularly useful for very large data sets. Logistic Regression machine learning algorithm used for classification of personal and social features.

Logistic Regression

Logistic regression analysis, which is a probabilistic statistical classification model, is a simple but powerful linear algorithm for analyzing multidimensional data and generating predictions of clinical outcomes. Multivariate selection is an important step in the development of a multivariate model that can effectively separate patients with MDD from the healthy control subjects.

Logistic Regression (LR) is utilized to train our model since the dependent variable is nominal. Some of the independent variables categorical (questions) and others are continuous. Let J represent the number of mental disorders that need to be predicted (j = 1,2, … ,J), N denotes to the number of observations (n = 1,2, … . , N), and K represents the number of independent variables(k = 1,2, … , K). The form of the traditional multinomial logistic regression is as follows

$$(Yi = k) = e \sum X\beta kj\ K\ k{=}0\ 1 + \sum e \sum X\beta kj\ K\ k{=}0\ J{-}1\ j{=}1\ j < J \quad (1)$$

where, $X = (x1, x2, \ldots, xk)\ N$ is the matrix of independent variables (SCL-90 questionnaire), βkj is the coefficient of the model. The parameters βkj of the model can be obtained (with grid search approach) by minimizing the following log-likelihood function

$$(\boldsymbol{\beta}) = \sum\sum(\sum yij \sum X\beta kj\ K\ k{=}0\ m\ i{=}1\ )\ j{-}1\ j{=}1\ N\ n{=}1 - ni\ log\sum 1 + e \sum X\beta kj\ K\ k{=}0\ j{-}1\ j{=}1$$

where Y is the dependent variable matrix. Since our model includes a high number of independent variables (i.e., more than 90 independent variables), we utilize lasso as a regularization approach to select the best subset of variables and to avoid over-fitting. In LASSO, a penalty function is added $L = \lambda \sum |\beta j|$ K k=1 to the traditional multinomial logistic regression to shrink coefficients to exactly zero. Therefore, the lasso logistic regression coefficients can be obtained maximizing log-likelihood of the following function

$$L(\beta) = \sum\sum(\sum yij \sum X\beta kj\ K\ k{=}0\ m\ i{=}1\ )\ j{-}1\ j{=}1\ N\ n{=}1 - ni\ log\sum 1 + e \sum X\beta kj\ K\ k{=}0 - \lambda\sum|\beta kj|\ J\ j{=}1\ j{-}1\ j{=}1 \quad (3)$$

where λ is called a tuning parameter and it controls the tradeoff between the penalty and the amount of regularization.

Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable(target) is categorical.

For example,

- To predict whether an email is spam (1) or (0)
- Whether the tumor is malignant (1) or not (0)

Consider a scenario where we need to classify whether an email is spam or not. If we use linear regression for this problem, there is a need for setting up a threshold based on which classification can be done. Say if the actual class is malignant, predicted continuous value 0.4 and the threshold value is 0.5, the data point will be classified as not malignant which can lead to serious consequence in real time.

From this example, it can be inferred that linear regression is not suitable for classification problem. Linear regression is unbounded, and this brings logistic regression into picture. Their value strictly ranges from 0 to 1.

Linear regression uses mean squared error as its cost function. If this is used for logistic regression, then it will be a non-convex function of parameters (theta). Gradient descent will converge into global minimum only if the function is convex.

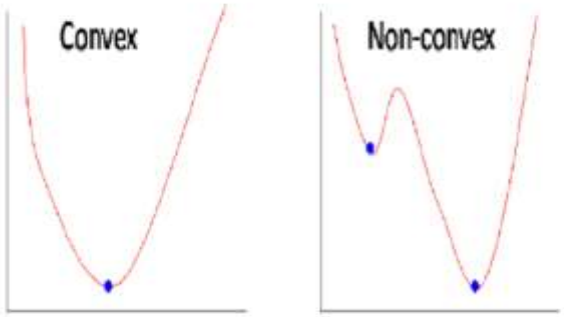

Fig 2: convex and non-convex cost function

5. EVALUATION

To evaluate the performance of a logistic regression model, we must consider few metrics.
Irrespective of tool (SAS, R, Python) you would work on, always look for:

1. AIC (Akaike Information Criteria) – The analogous metric of adjusted R^2 in logistic regression is AIC. AIC is the measure of fit which penalizes model for the number of model coefficients. Therefore, we always prefer model with minimum AIC value.

2. Null Deviance and Residual Deviance – Null Deviance indicates the response predicted by a model with nothing but an intercept. Lower the value, better the model.
Residual deviance indicates the response predicted by a model on adding independent variables. Lower the value, better the model.

3. Confusion Matrix: It is nothing but a tabular representation of Actual vs Predicted values. This helps us to find the accuracy of the model and avoid over fitting

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Fig 3: Confusion Matrix

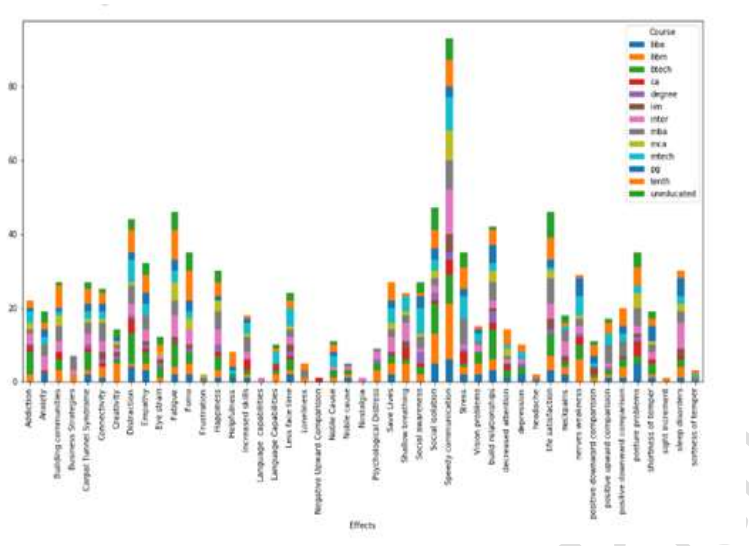

Fig 4: Course based Evaluation

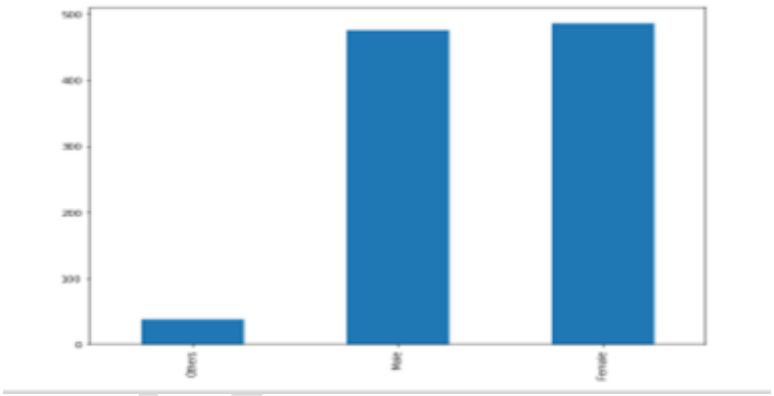

Fig 5: Gender based Evaluation

ROC Curve:

Receiver Operating Characteristic(ROC) summarizes the model's performance by evaluating the trade offs between true positive rate (sensitivity) and false positive rate(1- specificity). For plotting ROC, it is advisable to assume p > 0.5 since we are more concerned about success rate. ROC summarizes the predictive power for all possible values of p > 0.5. The area under curve (AUC), referred to as index of accuracy(A) or concordance index, is a perfect performance metric for ROC curve. Higher the area under curve, better the prediction power of the model. Below is a sample ROC curve. The ROC of a perfect predictive model has TP equals 1 and FP equals 0. This curve will touch the top left corner of the graph.

You can calculate the **accuracy** of your model with:

Accuracy is the percentage of the correct decision, i.e., **Accuracy** = $\frac{TP+TN}{TP+FP+TN+FN}$

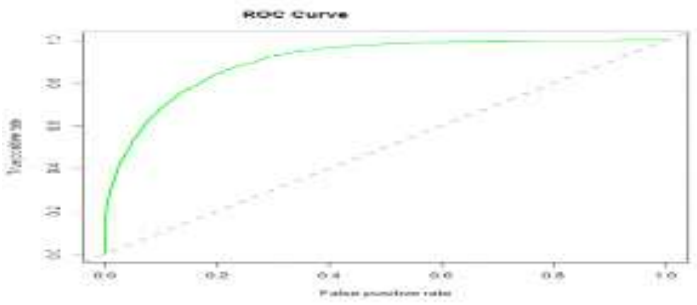

Fig 6: ROC curve based on Logistic regression algorithm

6. CONCLUSION

The primary conclusion of this research is that new model is developed by reducing the system can detect mental disorders automatically. At the same time, we proposed the LR model to detect mental disorders. Psychiatric tests are started to be applied in a computer environment, however, experts do evaluation and interpretation. In this sense, the first extended evaluation is done by an algorithm. It is unique in terms of quick and easy applicability with planned interfaces and cognitive architecture. It gives you the chance to practice and evaluate a test in thousands of days. As the evaluation of the test is done by artificial intelligence, results can be obtained in seconds, which is a first in the field of mental health.

References

[1]. B.Saha, T. Nguyen, D. Phung, and S. Venkatesh. "A framework for classifying online mental health-related communities with an interest in depression". IEEE Journal of Biomedical and Health Informatics, 2016.

[2]. Chun-Hao Chang, Elvis Saravia, Yi-Shin Chen "Subconscious Crowdsourcing: A Feasible Data Collection Mechanism for Mental Disorder Detection on Social Media". IEEE/ACM International Conference on Advance in Social Networks Analysis and Mining (ASONAM) 2016.

[3]. Lakshman Narayana Vejendla and A Peda Gopi, (2019)," Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology", Revue d'Intelligence Artificielle , Vol. 33, No. 1, 2019,pp.45-48.

[4]. Gopi, A.P., Jyothi, R.N.S., Narayana, V.L. et al. (2020), "Classification of tweets data based on polarity using improved RBF kernel of SVM" . Int. j. inf. tecnol. (2020). https://doi.org/10.1007/s41870-019-00409-4.

[5]. A Peda Gopi and Lakshman Narayana Vejendla, (2019)," Certified Node Frequency in Social Network Using Parallel Diffusion Methods", Ingénierie des Systèmes d' Information, Vol. 24, No. 1, 2019,pp.113-117.. DOI: 10.18280/isi.240117

[6]. Lakshman Narayana Vejendla and Bharathi C R ,(2018),"Multi-mode Routing Algorithm with Cryptographic Techniques and Reduction of Packet Drop using 2ACK scheme in MANETs", Smart Intelligent Computing and Applications, Vo1.1, pp.649-658. DOI: 10.1007/978-981-13-1921-1_63 DOI: 10.1007/978-981-13-1921-1_63

[7]. Lakshman Narayana Vejendla and Bharathi C R, (2018), "Effective multi-mode routing mechanism with master-slave technique and reduction of packet droppings using 2-ACK scheme in MANETS", Modelling, Measurement and Control A, Vol.91, Issue.2, pp.73-76. DOI: 10.18280/mmc_a.910207

[8]. Lakshman Narayana Vejendla , A Peda Gopi and N.Ashok Kumar,(2018)," Different techniques for hiding the text information using text steganography techniques: A survey", Ingénierie des Systèmes d'Information, Vol.23, Issue.6,pp.115-125.DOI: 10.3166/ISI.23.6.115-125

[9]. A Peda Gopi and Lakshman Narayana Vejendla (2018), "Dynamic load balancing for client server assignment in distributed system using genetic algorithm", Ingénierie des Systèmes d'Information, Vol.23, Issue.6, pp. 87-98. DOI: 10.3166/ISI.23.6.87-98

[10]. Lakshman Narayana Vejendla and Bharathi C R,(2017),"Using customized Active Resource Routing and Tenable Association using Licentious Method Algorithm for secured mobile ad hoc network Management", Advances in Modeling and Analysis B, Vol.60, Issue.1, pp.270-282. DOI: 10.18280/ama_b.600117

[11]. Lakshman Narayana Vejendla and Bharathi C R,(2017),"Identity Based Cryptography for Mobile ad hoc Networks", Journal of Theoretical and Applied Information Technology, Vol.95, Issue.5, pp.1173-1181. EID: 2-s2.0-85015373447

[12]. Lakshman Narayana Vejendla and A Peda Gopi, (2017)," Visual cryptography for gray scale images with enhanced security mechanisms", Traitement du Signal,Vol.35, No.3-4,pp.197-208. DOI: 10.3166/ts.34.197-208

[13]. A Peda Gopi and Lakshman Narayana Vejendla, (2017)," Protected strength approach for image steganography", Traitement du Signal, Vol.35, No.3-4,pp.175-181. DOI: 10.3166/TS.34.175-181

[14]. Lakshman Narayana Vejendla and A Peda Gopi, (2020)," Design and Analysis of CMOS LNA with Extended Bandwidth For RF Applications", Journal of Xi'an University of Architecture & Technology, Vol. 12, Issue. 3,pp.3759-3765. https://doi.org/10.37896/JXAT12.03/319.

[15]. Chaitanya, K., and S. Venkateswarlu,(2016),"DETECTION OF BLACKHOLE & GREYHOLE ATTACKS IN MANETs BASED ON ACKNOWLEDGEMENT BASED APPROACH." Journal of Theoretical and Applied Information Technology 89.1: 228.

[16]. Patibandla R.S.M.L., Kurra S.S., Mundukur N.B. (2012), "A Study on Scalability of Services and Privacy Issues in Cloud Computing". In: Ramanujam R., Ramaswamy S. (eds) Distributed Computing and Internet Technology. ICDCIT 2012. Lecture Notes in Computer Science, vol 7154. Springer, Berlin, Heidelberg

[17]. Patibandla R.S.M.L., Veeranjaneyulu N. (2018), "Survey on Clustering Algorithms for Unstructured Data". In: Bhateja V., Coello Coello C., Satapathy S., Pattnaik P. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol 695. Springer, Singapore

[18]. Patibandla, R.S.M.L., Veeranjaneyulu, N. (2018), "Performance Analysis of Partition and Evolutionary Clustering Methods on Various Cluster Validation Criteria", Arab J Sci Eng ,Vol.43, pp.4379–4390.

[19]. R S M Lakshmi Patibandla, Santhi Sri Kurra and N.Veeranjaneyulu, (2015), "A Study on Real-Time Business Intelligence and Big Data",Information Engineering, Vol.4,pp.1-6.

[20]. K. Santhisri and P.R.S.M. Lakshmi,(2015), " Comparative Study on Various Security Algorithms in Cloud Computing", Recent Trends in Programming Languages ,Vol.2,No.1,pp.1-6.

[21]. K.Santhi Sri and PRSM Lakshmi,(2017), "DDoS Attacks, Detection Parameters and Mitigation in Cloud Environment", IJMTST,Vol.3,No.1,pp.79-82.

[22]. P.R.S.M.Lakshmi,K.Santhi Sri and Dr.N. Veeranjaneyulu,(2017), "A Study on Deployment of Web Applications Require Strong Consistency using Multiple Clouds", IJMTST,Vol.3,No.1,pp.14-17.

[23]. P.R.S.M.Lakshmi,K.Santhi Sri and M.V.Bhujanga Ra0,(2017), "Workload Management through Load Balancing Algorithm in Scalable Cloud", IJASTEMS,Vol.3,No.1,pp.239-242.

[24]. K.Santhi Sri, P.R.S.M.Lakshmi, and M.V.Bhujanga Ra0,(2017), "A Study of Security and Privacy Attacks in Cloud Computing Environment", IJASTEMS,Vol.3,No.1,pp. 235-238.

[25]. R S M Lakshmi Patibandla and N. Veeranjaneyulu, (2018), "Explanatory &

Complex Analysis of Structured Data to Enrich Data in Analytical Appliance", International Journal for Modern Trends in Science and Technology, Vol. 04, Special Issue 01, pp. 147-151.

[26]. R S M Lakshmi Patibandla, Santhi Sri Kurra, Ande Prasad and N.Veeranjaneyulu, (2015), "Unstructured Data: Qualitative Analysis", J. of Computation In Biosciences And Engineering, Vol. 2,No.3,pp.1-4.

[27]. R S M Lakshmi Patibandla, Santhi Sri Kurra and H.-J. Kim,(2014), "Electronic resource management using cloud computing for libraries", International Journal of Applied Engineering Research, Vol.9,pp. 18141-18147.

[28]. Ms.R.S.M.Lakshmi Patibandla Dr.Ande Prasad and Mr.Y.R.P.Shankar,(2013), "SECURE ZONE IN CLOUD", International Journal of Advances in Computer Networks and its Security, Vol.3,No.2,pp.153-157.

[29]. Patibandla, R. S. M. Lakshmi et al., (2016), "Significance of Embedded Systems to IoT.", International Journal of Computer Science and Business Informatics, Vol.16,No.2,pp.15-23.

[30]. AnveshiniDumala and S. PallamSetty. (2020),"LANMAR routing protocol to support real-time communications in MANETs using Soft computing technique", 3rd International Conference on Data Engineering and Communication Technology (ICDECT-2019), Springer, Vol. 1079, pp. 231-243.

[31]. AnveshiniDumala and S. PallamSetty. (2019),"Investigating the Impact of Network Size on LANMAR Routing Protocol in a Multi-Hop Ad hoc Network", i-manager's Journal on Wireless Communication Networks (JWCN), Volume 7, No. 4, pp.19-26.

[32]. AnveshiniDumala and S. PallamSetty. (2019),"Performance analysis of LANMAR routing protocol in SANET and MANET", International Journal of Computer Science and Engineering (IJCSE) – Vol. 7,No. 5, pp.1237-1242.

[33]. AnveshiniDumala and S. PallamSetty. (2018), "A Comparative Study of Various Mobility Speeds of Nodes on the Performance of LANMAR in Mobile Ad hoc Network", International Journal of Computer Science and Engineering (IJCSE) – Vol. 6, No. 9, pp. 192-198.

[34]. AnveshiniDumala and S. PallamSetty. (2018),"Investigating the Impact of IEEE 802.11 Power Saving Mode on the Performance of LANMAR Routing Protocol in MANETs", International Journal of Scientific Research in Computer Science and Management Studies (IJSRCSMS) – Vol.7, No. 4.

[35]. AnveshiniDumala and S. PallamSetty. (2016),"Analyzing the steady state behavior of RIP and OSPF routing protocols in the context of link failure and link recovery in Wide Area Network", International Journal of Computer Science Organization Trends (IJCOT) – Vol. 34 No 2, pp.19-22.

[36]. AnveshiniDumala and S. PallamSetty. (2016),"Investigating the Impact of Simulation Time on Convergence Activity & Duration of EIGRP, OSPF Routing Protocols under Link Failure and Link Recovery in WAN Using OPNET Modeler", International Journal of Computer Science Trends and Technology (IJCST) – Vol. 4 No. 5, pp. 38-42.

[37]. VellalacheruvuPavani and I. Ramesh Babu (2019) ,"Three Level Cloud Storage Scheme for Providing Privacy Preserving using Edge Computing",International Journal of Advanced Science and Technology Vol. 28, No. 16, pp. 1929 – 1940.

[38]. VellalacheruvuPavani and I. Ramesh Babu,"A Novel Method to Optimize the Computation Overhead in Cloud Computing by Using Linear Programming" ,International Journal of Research and Analytical Reviews May 2019, Volume 6, Issue 2,PP.820-830..

[39]. Anusha Papasani and Nagaraju Devarakonda,(2016),"Improvement of Aomdv Routing Protocol in Manet and Performance Analysis of Security Attacks", International Journal Of Research in Computer Science & Engineering ,Vol.6,No.5, pp.4674-4685.

[40]. Sk.Reshmi Khadherbhi,K.Suresh Babu , Big Data Search Space Reduction Based On User Perspective Using Map Reduce ,International Journal of Advanced Technology and

Innovative Research Volume.07, IssueNo.18, December-2015, Pages: 3642-3647

[41]. B.V.Suresh kumar,Sk.Reshmi Khadherbhi ,BIG-IOT Framework Applications and Challenges: A Survey Volume 7, Issue VII, JULY/2018 pg.no 1257-1264

[42]. P.Sandhya Krishna,Sk.Reshmi Khadherbhi,V.Pavani, Unsupervised or Supervised Feature Finding For Study of Products Sentiment ,International Journal of Advanced Science and Technology, Vol 28 No 16 (2019).

[43]. K.Santhi Sri, Dr.Ande Prasad (2013), "A Review of Cloud Computing and Security Issues at Different Levels in Cloud Computing" , International Journal on Advanced Computer Theory and Engineering Vol. 2,pp 67-73.

[44]. K.Santhi Sri, N.Veeranjaneyulu(2018), "A Novel Key Management Using Elliptic and Diffie-Hellman for Managing users in Cloud Environment", Advances in Modelling and Analysis B,Vol.61,No.2,pp 106-112.

[45]. K.Santhi Sri, N.Veeranjaneyulu(2019), "Decentralized Key Management Using Alternating Multilinear Forms for Cloud Data Sharing with Dynamic Multiprivileged Groups", Mathematical Modelling of Engineering Problems,Vol.6,No.4,pp511-518.

[46]. S.Sasikala, P.Sudhakar, "interpolation of CFA color Images with Hybrid image denoising", 2014 Sixth International Conference on Computational Intelligence and Communication Networks, DOI 10.1109/.53 193 DOI 10.1109/CICN.2014.53, pp. 193-197.

[47]. Me. Jakeera Begum and M.Venkata Rao, (2015), "Collaborative Tagging Using CAPTCHA" International Journal of Innovative Technology And Research, Volume No.3, Issue No.5,pp,2436 – 2439.

[48]. L.Jagajeevan Rao, M. Venkata Rao, T.Vijaya Saradhi (2016), "How The Smartcard Makes the Certification Verification Easy" Journal of Theoretical and Applied Information Technology, Vol.83. No.2, pp. 180-186.

[49]. Venkata Rao Maddumala, R. Arunkumar, and S. Arivalagan (2018)"An Empirical Review on Data Feature Selection and Big Data Clustering" Asian Journal of Computer Science and Technology Vol.7 No.S1, pp. 96-100.

[50]. Singamaneni Kranthi Kumar, Pallela Dileep Kumar Reddy, Gajula Ramesh, Venkata Rao Maddumala, (2019), "Image Transformation Technique Using Steganography Methods Using LWT Technique" ,Traitement du Signalvol 36, No 3, pp. 233-237.