Sentimental Analysis Using Deep Learning For Psychological Diagnosis

Sonam Kumari[1], Baskar Kasi[2], Rajesh E[3], Sakthi Govindaraju[4], Tushar[5]

{ sonamsingh7549@gmail.com [1], baskar_ka@yahoo.com[2], rajsugan20@gmail.com[3]}

School of Computing Science and Engineering, Galgotias University, India[1,2,3,4,5]

Abstract. Sentiment analysis is the process of managing feelings, opinions, and subjective language. The demand to analyze and organize unstructured data obtained from social media in the form of hidden information has increased the need for sentiment analysis. The research covered in this paper focuses on people with mental illnesses. After speaking with a large number of patients, it became clear why some of them were reluctant to meet with doctors and discuss their issues. Even patients' family did not feel very comfortable visiting doctors. Constraints are the main factors in the Asian context. the absence of nearby doctors and their inability to accept their disease. Patients are subject to social limitations, which are ones that society imposes on them. If the public learns that a patient is seeing a mental health professional, they label him as insane. Nearly all Asian nations frequently encounter situations like these. This forces patients and their family members to conceal their disease and avoid seeing the doctors unless circumstances become uncontrollable. The indications of a mental patient's out-of-control craziness lead society to accept what they already believed. Since only mentally ill patients have been observed in society visiting mental health professionals. So, they assume that only insane people seek out mental health professionals. The second and most significant factor is that patients refuse to acknowledge that they are dealing with a mental illness. They do not see it favorably if someone claims to have some sort of mental illness. In this study, researchers will put up a creative idea for assisting and enhancing the care of someone with psychological distress. The suggested remedy will build a learning pattern from the assessed input data and compare it to the tested model. Based on the prior treatment of the patient's repository in text form, it may advise and offer insightful information that helps psychologists diagnose their patients more effectively. Both the patient and the psychiatrist will benefit from the suggested remedy. Inputs: Doctor's evaluation form in digital format. Deep learning: Training model developed from repository of prior treatment records. The remedy is a diagnostic and an assistance report.

Keywords: Classification; Long Short-Term Memory; Natural Language Processing; Machine Learning; Deep Learning.

IACIDS 2023, November 23-25, Lavasa, India

DOI 10.4108/eai.23-11-2023.2343175

1 Introduction

Since many mental health conditions have a propensity to get worse over time, early detection of metal disorders is essential for effective treatment. There are many ways to identify mental problems, and even hand movements, voice intonation, and facial expressions are taken into account. In order to identify the sort of mental disease, seeking medical attention and examining the symptoms may be helpful. In order to determine whether another ailment is the source of the patient's symptoms, a doctor will examine the patient, inquire about their symptoms, and then advise a blood test. The use of questionnaires in assessments to identify mental illness is one common technique. The state of a person's social, psychological, and emotional welfare is one definition of mental health. Mental health issues that adversely affect an individual's intelligence level, emotions, and social relationships may have an impact on it. To differentiate between various illnesses, an accurate and prompt evaluation is crucial for treatment. Utilizing self-report questionnaires created to identify specific feelings or attitudes toward social interactions, mental health disorders are screened[1].

The use of Machine Learning (ML) in psychological therapies has been applied to many different fields, and it has great promise for diagnosing and treating mental health issues as well as related physical results. To find patterns and carry out classification tasks, these algorithms often need a lot of data. Supervised learning is one of the ML techniques that is most often used to predict mental diseases. By using supervised learning, which requires discovering an analogy between a set of input variables and an output variable, one can predict the outcomes of unseen data [3].

For classification and regression issues, the Support Vector Machine (SVM) is one of the most efficient supervised learning algorithms. This method locates the hyperplane, a perfect boundary or decision line, and divides n-dimensional space into various classes using the margin calculation strategy. Future data points must be appropriately categorised in order to do this. SVM has the flexibility to handle both structured and semi-structured data, which is one of its benefits. Because generality is used, it also reduces the likelihood of overfitting. SVM does, however, have a number of problems. More datasets mean longer training times. As a result, its performance starts to suffer. A noisy dataset also does not perform well with SVM. Decision trees, which may be applied to classification and regression issues, are another supervised learning technique. An image of a piecewise constant as a tree is possible. It builds models that forecast the result of the selected variables by learning common decision-making principles derived from data attributes. Because a categorical dependent variable is predicted using logistic regression, the result could be Yes or No, 0 or 1, etc.

The Bayes theorem is also used in naive Bayes categorization. It is presumptive that a specific dataset feature is unrelated to any other properties. Ensemble learning is a separate yet widely applied machine learning paradigm. This technique generates a huge number of experts that are integrated to produce a single model by having each learner replicate a different classical ML strategy. Ensemble learning is made up of the three courses bagging, boosting, and stacking. Bagging simultaneously produces a huge number of learners and aggregates all of the experts using either an average or the silent majority vote method by using random sampling to create a large number of datasets Boosting creates a large number of datasets and incrementally trains learners by using weighted data replacement and random selection. The weighted average

approach is then used to mix these students. In stacking, a different classic neural network ML technique that either starts with bagging or boosting (meta-model), the expert outputs are used as inputs. The outputs from the meta-model are then combined to get the outcomes. Two of the most well-liked ensembles learning approaches are Extreme Gradient Boosting (XGBoost) and Random Forest (RF). The Random Forest technique builds decision trees from subsets of data using the bagging technique, and the output from each decision tree is combined to construct the final decision tree. DGBDT (Distributed Gradient Boosted Decision Tree) is the name of the scalable distributed gradient-boosting approach for decision trees. Researchers in this discipline are also looking into how transfer learning may be used in conjunction with ML. In layman's words, it is the application of previously acquired information from a similar activity to a new task to enhance learning [4].

The widespread use of these algorithms raises ethical and legal questions concerning data anonymization, even if they provide new opportunities for psychological study [5]. Even the patients' families found it uncomfortable when doctors visited. The main factors in the Asian context are restrictions, resistance to acknowledge one's illness, and a lack of nearby physicians. Social restraints are limitations put on patients by society. If the public finds out that the patient is seeing a mental health professional, they declare him insane. This situation is rather common in almost all Asian countries. Because of this, patients and family members are driven to hide their illness and put off visiting the doctor until things get out of hand. These out-of-control lunatic indicators of a mental disease lead society to believe what they already believed. As a result, society has only observed people seeking medical care who are mentally sick. They consequently think that only the mad consult with mental health specialists. The second and most important factor is that patients do not recognise they are suffering from a mental disease; even if they are convinced they do, they do not view it favourably. Even they can fight with them. Family members share the same experience. People are hesitant to suggest that those suffering from mental illnesses consult mental health professionals because most patients in Asian countries behave in this way.

This exacerbates the issue, and occasionally a protracted course of therapy is necessary. Patients occasionally cause harm to themselves. Some of the most prevalent types of anxiety disorders include generalized anxiety disorder (GAD), obsessive compulsive disorder (OCD), panic disorder, post-traumatic stress disorder (PTSD), social mental illness, agoraphobia, and suicidal tendencies.

Those who suffer from generalised anxiety disorder may worry excessively about everyday matters. It differs from the typical apprehensive emotions. It's common to experience periodic anxiety. Uncontrolled, excessive worry can be upsetting and disruptive to regular activities. In addition to exhaustion, hot hands, a parched tongue, tense muscles, and other symptoms, pain is frequently accompanied by these. GAD is a fairly common disorder that affects about 3% of Americans. Obsessive compulsive disorder (OCD) is a chronic and recurrent condition. In this, one experiences the need to repeat an action. Someone with OCD may exhibit signs of compulsion, fixation, or both. Obsession is the repeated thoughts, whereas compulsion is the repetitive activity. A person experiencing a panic attack due to panic disorder may feel out of control and experience an excessively rapid heartbeat. They include a sudden feeling of dizziness, dread, and tightness in the chest. The name for the most extreme kind of anxiety is panic disorder. People with panic disorder may have a tendency to avoid specific situations out of fear that they'll trigger another attack in an effort to stop the cycle of living in terror. Although

there is no established cause for panic disorder, there may be a number of contributing variables, including heredity, traumatic events, or significant accidents. Post-traumatic disorder (PTSD) may manifest in someone who has a traumatic or unpleasant event that leaves them feeling helpless, afraid, or shocked. For someone with PTSD, it frequently takes an unexpected turn and may get better or worse over time. They frequently suffer flashbacks or nightmares that are connected to the events. The common disorder known as social anxiety disorder includes the fear of being watched or judged. It has a big impact on a person's life and makes working difficult. Since many mental health conditions have a propensity to get worse over time, it is essential to recognise them early so that they can be successfully treated. Even hand gestures, vocal inflections, and facial expressions are taken into consideration when diagnosing mental illnesses. Consulting a doctor and looking at the symptoms may help with early disease diagnosis as well as determining the type of mental illness. The doctor recommends a blood test after seeing the patient and learning about their symptoms in order to determine whether these symptoms are being caused by another illness. A questionnaire-based evaluation is a typical method for detecting mental disorders. These methods will help those who are dealing with these situations and offer the best solution in the form of a report or thoughts. These technologies will also help people who are unwilling to verbally convey their problems. These systems are capable of self-learning and knowledge recovery with the aid of patients, therapists, or online resources.

2 Literature Review

Authors Patwa et al. (2020) discuss the results of SemEval-2020 Task 9 utilizing sentiment analysis of code-mixed Twitter posts. The word hing (Hindi-English) as well as Spanglish (Spanish-English) businesses, tagging them along with sentence-level emotion labels and word-level language recognition. These knowledge corpora contain 20 K and 19 K occurrences, respectively. Positive, negative, and neutral sentiment are the three types. For the Hinglish and Spanglish competitions, Senti Mix acquired a total of 89 applications, involving 61 teams and 28 systems, respectively. Hinglish F1 and Spanglish F1 earned the highest scores, with overall scores of 75.0 and 80.6 percent, accordingly. Researchers discover that participants preferred and perform better using ensembles techniques and models comparable to BERT[1].

De Choudhary et al. (2013) PCA (Principal Component Analysis), SVM (Support Vector Machine) classification method, SMDI (Social Media Depression Index), and CES-D measure were used to assess depression. The same authors' first prescriptions for the PHQ-9 and LIWC (Linguistic Inquiry and Word Count) surveys from 2014 were applied in predicting postpartum depression using data collected from Facebook [2][3]. Around 97 million individuals are expected to experience psychological disorders worldwide in 2019, with depression and anxiety being the two most common. This prediction comes from the World Health Organization (WHO). Nevertheless, the COVID-19 pandemic, which began in 2020, caused this figure to rise considerably. It is more important than ever that people have access to healthcare and treatment for mental problems as a result of this epidemic. When there are alternatives out there, a majority of individuals cannot use them, and many of them suffer prejudice, embarrassment, and legal violations[6]. A comprehensive psychiatric interview is necessary to diagnose mental health problems. This interview frequently covers the suspected signs and symptoms,

psychological history, and physical exam findings. When detecting psychiatric symptoms, psychological exams and diagnostic techniques are often beneficial [7].

There has been numerous research on how to recognize mental diseases. We focused our search on illnesses that have an impact on ML and DL, like schizophrenia, bipolar disorder, anxiety, depression, and attention deficit hyperactivity disorder (ADHD), in order to undertake this investigation. A severe mental condition called schizophrenia alters reality perceptions, leading to incorrect reality interpretations. According to a World Health Organization estimate, schizophrenia affects 300 million persons globally on average. Additionally, since they are more likely to develop cardiovascular, metabolic, and viral disorders, it roughly doubles or triples the likelihood that patients will pass away [8].

Delusions, hallucinations, slurred speech, erratic conduct, and other symptoms might develop. In the end, this could result in social and professional problems [9]. One of the common mental illnesses that is routinely evaluated using the Patient Health Questionnaire (PHQ) is depression (major depressive disorder). Intense melancholy and a lack of enthusiasm in activities are common symptoms, which can make it harder for a person to perform everyday responsibilities.

Shorey et al.'s study found that the risk of getting clinical depression is higher for teenagers between the ages of 10 and 19 (34%) than it is for those between the ages of 18 and 25 (25%)[10]. According to published study data, the Middle East, Africa, and Asia are home to the greatest number of instances with heightened depression symptoms, even though it is thought that female teenagers experience these symptoms more frequently than male teenagers. If depressive disorders get neglected, it could result in thoughts of suicide and behaviour[11].

Anxiety results in a worry or fear emotion, which could be mild or strong. Anxiety is a symptom of numerous illnesses, including phobias, panic disorders, and social anxiety disorder (also known as social phobia). Everybody experiences anxiety occasionally, but it only becomes a problem when it interferes with daily life and makes it difficult for the person to control their emotions. Symptoms of general anxiety include sweating or palpitations, difficulty falling asleep, difficulties concentrating, restlessness, and stress. According to a 2020 study, nearly sixty percent of the participants reported having some level of anxiety, while more women than males have the condition. [12].

Around the world, 264 million people are thought to struggle from anxiety. A specific type of mental illness called bipolar disorder, also referred to as "mania-depressive disease" or "mania-depression," is characterized by a sudden change in mood, a lack of energy, and lower levels of activity and concentration. Milder episodes as well as periods of extreme highs (mania or hypomania) and lows (depression) are all included in the mood swings that are discussed. According to studies, bipolar disorder affects over 46 million people globally in varying degrees. Given that 60 percent of those with this disease exhibit symptoms of drug abuse, they may also be at risk of suicide [13].

Based on a person's symptoms, experiences, life history, and, in rare instances, family history, it is frequently diagnosed throughout the course of their lifetime. Flashbacks, nightmares, severe anxiety, and persistent, irrational thoughts that are brought on by horrible events that a person either went through or witnessed are some of the symptoms of PTSD. Medical professionals physically examine the suspected patient to rule out any underlying illnesses that could be the cause of the typical symptoms in order to appropriately diagnose PTSD. They carry out a

psychological assessment to talk about the causes of the symptoms, and they use the DSM-5 criteria to appropriately identify the disease[9]. PTSD is incurable, like the majority of mental disorders, but curable with proper treatment (most often psychotherapy), which may assist those who have it in recovering control over their lives. Also based on statistics, 3.5% of American adults report having PTSD annually, with an average incidence of 8% among teenagers between the ages of 13 and 18[14].

For the purpose of examining sentiment on Twitter, the authors created a deep learning technique. The major objective of this study was to demonstrate the importance of parameter weighting in a convolutional neural network for rapid model training without the need for extra features. The word embedding procedure starts with the usage of a neural language that has been trained using a large amount of data. A traditional neural network is utilized to enhance the embedding using a random sample of tweets from a sizable supervised corpus.

Table 1. The Existing Results Obtained by Some Relevant Model

Year	Authors	Type of Feature	Sources of Dataset	Methods	Result
2017	Shen et al.[17]	Descriptive	Twitter	MDL , MSNL , WDL , NB	Acc: 85%; Pre: 85%; Rec: 85%; F1: 85%
2019	Burdisso et al.[18]	Textual	Reddit	SS3 , KNN, LR, SVM, NB	Pre: 63%; Rec: 60%; F1: 61%
2020	Alsagri and Mourad[20]	Textual	Twitter	SVM, NB, DT	Acc: 82.5%; Pre: 73.91%; Rec: 85%; F1: 79.06%; AUC: 0.78
2020	Kim et al.[19]	Textual	Reddit	CNN, XGBoost	Acc: 75.13%; Pre: 89.1%; Rec: 71.75%; F1: 79.49%
2020	Srimadhur and Lalitha[21]	Spectrogram & End-to-end	DAIC-WOZ	CNN (Nondepressed class)	Pre: 65%; Rec: 92%; F1: 76%
2021	Filho et al.[22]	Descriptive	Patients' clinical evaluation	RF, LR, KNN, DT, AB, SVM, GB	Acc: 89%

The network was initially initialized with embedded words and parameters using the same architecture as Semeval-2015 and supervised corpus training. Activations, sentence matrix

pooling, SoftMax, and convolutional layers are some of the techniques used. are all examples of these techniques. the components employed in the proposed study. In order to get the network ready, the methods used were stochastic non-convex function optimization with gradient descent (SGD), with the gradients being determined using a back propagation method. The dropout method was used to enhance neural network regularization. The deep learning model is applied to two challenges from Semeval-2015: message level task and phrase level work, in order to predict polarity and provide satisfactory results. The suggested model ranks best in terms of accuracy after applying six test sets[15].

Table 1's technique field defines the ideal model more precisely in each author's particular works. Where, 'MDL' is for Multimodal Depressive Dictionary Learning; 'MSNL' is for Multiple Social Networking Learning; 'WDL' is for Wasserstein Dictionary Learning; 'SS3' refers for Sequential S3 (Smoothness, Significance, and Sanction); 'KNN' indicates K- nearest neighbors; 'LR' refers for Linear Regression; 'SVM' indicates Support Vector Machine; 'NB' refers Naïve Bayes; 'DT' refers Decision Tree ; 'CNN' refers Convolutional Neural Network; 'XGBoost' refers a gradient boosting algorithm for supervised learning ; 'RF' refers Random Forest ; 'GB' refers Gradient Boosting.; 'DAIC-WOZ' refers Distress Analysis Interview Corpus Wizard of Oz. 'Acc' is for Accuracy; 'Pre' for Precision;' Rec' for Recall; 'F1' for F-measure; 'AUC' indicates Area Under the Curve.

3 Implementation

Psychological Diagnosis Using Deep Learning Model

3.1 Data Source

The data source is a specific type of data repository where data is stored and used to publish reports or compile data. The Mental Health Corpus, which is published on the Kaggle platform, collects the data for psychological diagnosis. The Mental Health Corpus includes information about depression, anxiety, and other mental health conditions. The corpus consists of two columns: labels designating whether or not the statements are detrimental, and the comments themselves.

The corpus can be used for sentiment analysis, language analysis for mental health, and dangerous language identification, among other uses. The content in the corpus may be of interest to researchers, mental health professionals, and others who are interested in examining the language and attitudes around mental health issues.

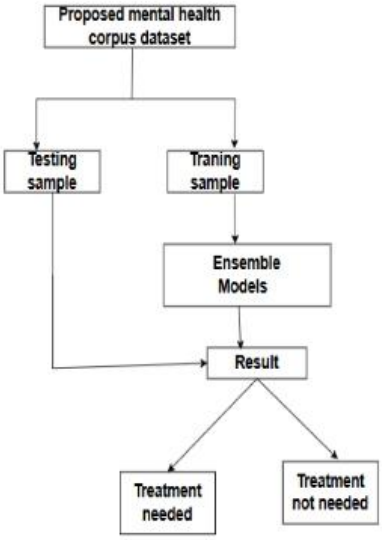

Fig.1 System Architecture of Psychological Diagnosis

'Stress.csv,' the dataset used for this project, is a fascinating collection of postings from Reddit's mental health-focused subreddits. The dataset is organized into five columns: "subreddit," "post_id," "sentence_range," "text-label," and "confidence-social_timestamp." The following details are contained in each column that reflects a user's post in the dataset:

‘subreddit: refers’ to the specific subreddit where the post was made.

‘post_id’ is a unique identifier for each post.

‘sentence_range’ depicts the span of sentences within a post considered for analysis.

‘text-label’: represents the main content of the user's post.

'confidence-social_timestamp' contains the timestamp of the post.

This dataset also provides an intriguing approach to the problem of stress identification. In addition to offering a glimpse into the myriad mental health problems individuals encounter, it also assigns a binary number to each record, where '0' indicates no stress and '1' indicates extreme stress. We can build and train a machine learning model that can recognize stress in textual data with the use of this tagging technique.

3.2 Pre-Processing

The dataset used in this study was sourced from Reddit's subreddits pertaining to mental health. The dataset is loaded into a Pandas Data Frame using the read_csv function from the Pandas package. Duplicate rows and missing values are handled to ensure the quality of the data. This phase includes utilizing the nltk library to remove stop words and other special characters, the Word Cloud library to create a word cloud, the Porter Stemmer and lambda function for tokenization and stemming, and the TfidVectorizer for vectorizing for feature normalization.

i. Exploratory Data Analysis

Data visualisation techniques such as pie charts and bar charts are used to understand the distribution of labels in the dataset. Text length analysis is performed to determine the optimal sequence length for the models. To make inferences about the dataset, a random sample of text is selected for each category.

ii. Model Development

A number of models are constructed and trained using various designs, such as Simple Dense Networks, LSTM (Long Short-Term Memory), LSTM (Bidirectional), GRU (Gated Recurrent Units), and Convolutional Neural Networks (CNN). These algorithms are trained using pre-processed text data and can predict stress levels based on textual posts.

iii. Model Evaluation

A variety of criteria are used to assess each model's effectiveness, including accuracy, precision, recall, and F1-score. Confusion matrices are created to examine the classification's effectiveness. Random guesses are created on the validation set in order to gain more knowledge about model predictions.

iv. Ensemble Models

The psychological diagnosis system architecture is shown in Fig.1. An ensemble of distinct models is produced by combining the predictions of the LSTM, bidirectional GRU, and GRU models. To get the final forecast, the ensemble model aggregates the prediction probabilities from several models. Assessment of the ensemble model's performance and comparison with individual models.

4 Result Analysis

The results acquired by relevant model are as shown in Table2. The results obtained from the model showcase their performance uses textual data to forecast depression. The GRU and Bidirectional LSTM models came in second and third, with accuracy values of 92.9% and 92.8%, respectively, behind the ensemble model, which had the best accuracy of 93.9%. Additionally, these models exhibit good precision, recall, and F1-scores, demonstrating their efficacy in predicting depression. The benchmark model, the baseline model, has an accuracy of 85.4%. The success of utilizing deep learning models for stress prediction based on text data is demonstrated by the fact that subsequent models continually exceeded the baseline. Accuracy levels of 92.3% to 92.7% were reached by the Simple Dense, LSTM, and usage models, which all demonstrated noteworthy performance. The accuracy values of 92.9% and 92.8%, respectively, behind the ensemble model, which had the best accuracy of 93.9%. Additionally, these models exhibit good precision, recall, and F1-scores, demonstrating their efficacy in predicting depression. The benchmark model, the baseline model, has an accuracy of 85.4%. The success of utilizing deep learning models for stress prediction based on text data is demonstrated by the fact that subsequent models continually exceeded the baseline. Accuracy levels of 92.3% to 92.7% were reached by the Simple Dense, LSTM, and usage models, which all demonstrated noteworthy performance.

Table 2. Table of the Results Obtained by the Proposed Model

Model	Accuracy	Precision	Recall	F-1 Score
Baseline	0.854350	0.881010	0.854350	0.851979
Simple dense	0.923767	0.923794	0.923767	0.923762
LSTM	0.924843	0.925147	0.924843	0.924818
Bidirectional LSTM	0.928430	0.928551	0.928430	0.928419
GRU	0.928610	0.928631	0.928610	0.928606
Bidirectional GRU	0.926099	0.926134	0.926099	0.926093
Conv1D	0.924664	0.925421	0.924664	0.924613
USE	0.926816	0.926843	0.926816	0.926818
Ensemble Results	0.939372	0.939417	0.939372	0.939367

5 Conclusion

As a result of textual data from Reddit's mental health-related subreddits, researchers used a variety of deep learning models, including LSTM, GRU, Bidirectional LSTM, and an ensemble of these models, to predict stress levels. The models performed well in terms of accuracy, recall, and F1-scores, achieving high accuracies between 92.3% and 93.9%. By offering scalable and remote methods for stress level monitoring, the results demonstrate the promise of AI and deep learning in tackling mental health issues. These models can help with early interventions, individualized support, and the creation of stress monitoring apps or mental health improvement platforms by analyzing text data from social media platforms like Reddit.

References

[1] M. Velayanikal, 13 January 2018. [Online]. Available: https://www.business-standard.com/article/economy-policy/bhim-app-market-share-falls-to-6-as-tez-phonepe-and-paytm-make-gains-118010500063_1.html.

[2] Statistics, 8 March 2018. [Online]. Available: http://rbidocs.rbi.org.in/rdocs/content/docs/ELECT07022016_A.xls

[3] Product Overview, 2016. [Online]. Available: https://www.npci.org.in/product-overview/upi-product-overview

[4] Digital payments in India to reach $1 trillion by 2023: Credit Suisse," 15 February 2018. [Online]. Available: https://economictimes.indiatimes.com/small-biz/startups/newsbuzz/digital-payments-in-india-to-reach-1-trillion-by-2023-credit-suisse/articleshow/62935890.cms

[5] D. S. Denis Dennehy, „Trends in Mobile Payments Research : A Literature Review," March 2015`. [Online]. Available: https://www.researchgate.net/publication/275155059

[6] X. Song, „Mobile Payment and Security," rev. Seminar on Network Security, 2001

[7] K. P. a. K. T. Nina Kreyer, „Standardized Payment Procedures as Key Enabling Factor for Mobile Commerce," Munich Personal RePEc Archive, 2002

[8] D. J. K. Basudeo Singh, „Enabling P2P Mobile Payment through P2P Network," IRACST – International Journal of Computer Networks and Wireless Communications (IJCNWC), 2012

[9] N. M. Tomi Dahlberg, „MOBILE PAYMENT SERVICE DEVELOPMENT – MANAGERIAL IMPLICATIONS OF CONSUMER VALUE PERCEPTIONS," ECIS, 2002

[10] B. N. a. S. Sun, „Using Text Mining Techniques to Identify Research Trends: A Case Study of Design Research," MDPI, 2017

[11] M. K. H. S. A. a. F. F. Ramzan Talib, „Text Mining: Techniques, Applications and Issues," (IJACSA) International Journal of Advanced Computer Science and Applications, 2016.

[12] A. K. a. S. Naithani, „A Comprehensive Study of Text Mining Approach," IJCSNS International Journal of Computer Science and Network Security, 2016.

[13] D. K. a. J. W. Kim, „Public Opinion Mining on Social Media: A Case Study of Twitter Opinion on Nuclear Power," Advanced Science and Technology Letters , 2014

[14] M. Mulins, „Information extraction in text mining," Computer Science Graduate and Undergraduate Student Scholarship, 2008

[15] S. R.Sagayam, „A Survey of Text Mining: Retrieval, Extraction and Indexing Techniques," International Journal Of Computational Engineering Research (ijceronline.com) Vol. 2 Issue. 5, 2012