

# A Survey of Depression Analysis in Human Behaviour Using Data Mining

[1] Rupinder Kaur, [2] Sandeep Kumar Singla

[1]PG Student,[2]Assistant Professor

Department of Information Technology, Guru Nanak Dev Engineering College, Ludhiana,Punjab, India,  
[1rupinderk1203@gmail.com 2][sandeepkumar.singla@gmail.com](mailto:sandeepkumar.singla@gmail.com)

**Abstract:** Depression is the main cause of many diseases caused by changes in autonomous nervous system. We look at the potential for utilizing web-based social networking postings as another sort of focal point in understanding causes of depression in humans. Utilizing the model, we present an online networking depression record that may serve to portray levels of depression among populations. The aim of study is to tackle the problem of depression among individuals. The proposed system comes out with the finding that how the behavior of person can be detected to predict the depression-indicative posts. The 3 techniques are used: Rule Based Method, Deep neural network and discriminate classification analysis (LDA and PCA) which gives approximately 86% and 67% accuracy. The experiment can be used to detect depression beforehand and it suggests taking proper treatment in case of depression.

**Index Terms:** Data mining, Depression, Emotions, DNN (Deep Neural Network), LDA (Linear Discriminant Analysis), SVM (Support Vector Machine), Hypertension.

## I.INTRODUCTION

The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Data mining is the efficient process of revealing useful and valid information from large scale databases. It transforms large amount of information into meaningful knowledge. The main purpose of data mining is to search homogeneous patterns and ordered correlation among data, validate the findings by implementing detected patterns to new subsets of data, predict new findings on new datasets. The important task of data mining is the automatic analysis of huge amount of data to reveal previously unknown, similar patterns, unusual records and dependencies. These patterns are considered as brief summary of the input data, and might be used to further analyze the data set.

**Table I**  
**Data mining tasks**

Sr. No.	Technique	Description
1	Prediction based	method used to forecast unknown or future values of other variables utilizing some other variables
2	Description based	method describe the data by finding human-interpretable pattern

## II. TECHNIQUES OF DATA MINING

Data mining consists of many up-to-date techniques generally divided into classification and clustering and association.

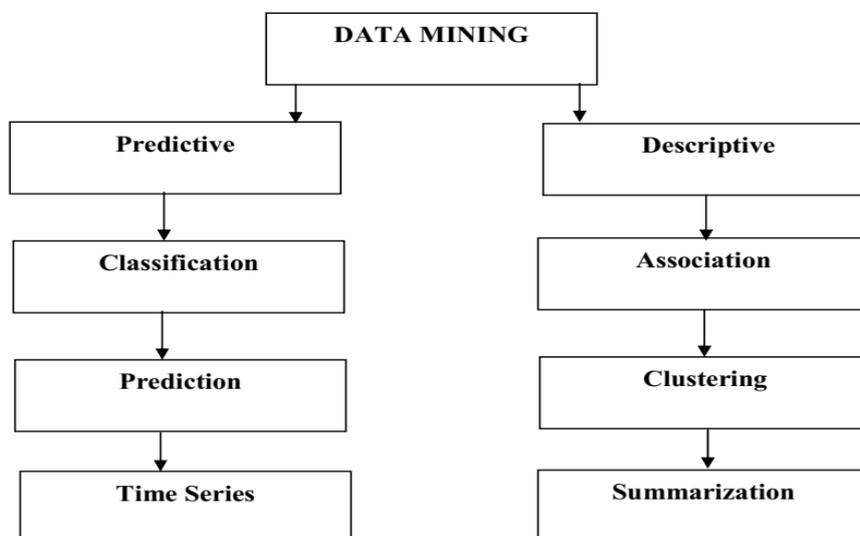


Fig. 1 Techniques of Data Mining

- Classification techniques: Classification techniques are composed to classify unidentified samples using information from already classified samples. Classification involves predicting a definite outcome on the basis of given input. The algorithm seeks to betray the existence of relation among the attributes of the entities by which it could be possible to determine the respective outcome. The techniques under classification are rule based, decision trees, naive Bayes classifier, neural network, memory based, KNN and SVM.
- Clustering techniques: A cluster is assembling together objects in one class which are alike and different to other classes. Clustering is the process of arranging those objects in the same group which have similar properties. The objects are more similar in the same group in some way than those in other groups. More the similarity within a group and there will be more difference between the groups and therefore the clustering will be more précised. k-means, hierarchical clustering, density-based clustering are some of the clustering types.
- Association techniques: This is a method to discover interested relation which is uncovered in large data set. It reveals the link between data. To apply association rule to market basket data there are two intakes that are keen to be addressed. First is determining patterns from huge database could be mathematically expensive. Secondly some of the predicted patterns are potentially counterfeit because they simply may occur by chance. Association mining could be one-dimensional, multidimensional, multilevel association, constraint-based association.

## III. APPROACHES TO DEAL WITH THE PROBLEM OF DEPRESSION

Depression is the main cause of many diseases. Depression is a state low-spirited habit (may be temporary) caused due to other's dislike or neglecters which could affect person's thoughts, behavior, feelings and sense of stability. It is a critical sickness caused due to changes in neural system of humans. Instead of straightaway issue depression normally results due to the combination of latest events and other personal or long term factors. Based on the principle of "Chinta chita saman hai" . "Chinta is what is known as worry which creates disturbance in the stability of mind. Good and bad days are the part of one's life like the two sides of coins. Parents, school, friends, relationships, surroundings can be frustrating or irritating.

Things can be right at one minute and horrific at the other. In everyday life sadness can be caused due to loss of someone special, accident, breaking of relations. It becomes more serious problem when this sorrow lasts for over more than 2 weeks and interfere your life. When someone talk about feeling tensed or depressed, they might mean they have a sad day or they might talk about clinical depression.

Hypertension is also one of the main cause. No particular medical test is available for depression rather it could only be detected through symptoms which lead to depression. Medicines may help to get relieve from some of the signs of depression, but it couldn't fully cure the basic problem. Our main target was to build a system to personify user activity on social health networks using data mining techniques. This will allow us to efficiently represent consumer relationships by depending on the data's semantic content. More than 20% of the adults are affected by depression.

**Culjak** (2006) examined the Theory of Behavior Change how the Internet render different paths to bother through broad approach and information of depression and relevant mental disorders than earlier attained through conventional media. The outcome represents that for depression the service of different sites on Internet does effect mental behavior and the influence is affirmative. It recognizes how transition in behavior bulge to decrease in the severity and lastingness of depression.

**Rosenquest et al.** (2010) analysed thickly interconnected informal community of 12067 individuals surveyed several times over 32 years as a feature of the Framingham Heart Study. Longitudinal factual models were utilized to look at whether depressive manifestations in one individual were related with comparable scores in companions, associates, spouse, kin, and neighbors. Symptoms of depression were surveyed utilizing CES-D scores that were accessible for subjects in three waves measured in the vicinity of 1983 and 2001. Female companions seem to be peculiarly prestigious in spreading depression from one individual to the next. The results also show that friends over time are known to influence more than spouses.

**Wu et al.** (2012) introduced that the inter-sentential language patterns can catch relationship among numerous words inside and between sentences, therefore can give more exact data than word sets. Experimental outcomes demonstrate that the utilization of inter sentential structures beat the utilization of word sets formulated in previous studies.

**Choudhury et al.,** (2013) displayed work at utilizing a crowd sourcing philosophy to fabricate a vast corpus of postings on twitter that have been shared by people determined to have clinical dejection. Next, they built up a probabilistic model prepared on this corpus to decide whether posts could demonstrate depression. The model use signs of social activity, feeling and language showed on twitter to develop the SVM classifier. The classifier anticipated with high exactness (73%) regardless of whether a post on twitter could be depression indicative.

**Mok** (2014) investigated the relation between social media and different age groups. They surveyed at Singapore and observed that among the total participants 48.9% used social media from 5-7 days per week. 9.8% used gaming platforms, 69.2% used video sites and furthermore 82.3% among them used social site. They observed that senior citizens above the age of 55 years were not using social media.

**Akay et al.,** (2016) build up a weighted network model to reflect user's actions on social wellbeing systems which empowered them to precisely reflect user associations by depending on the information's semantic content. The three-stage strategy utilizes the weighted system model to show user interactions, and system clustering and domain detection to portray user communications and concentrate promote learning from users posts.

## IV. DESIGN AND IMPLEMENTATION

To design the framework there are diverse steps which are applied to analyze depression from the input text. In the pre-handling stage, the data is collected from online social sites. For predicting depression the database is uploaded and features are extracted. In the next step the stop words are expelled from input text. In the processing stage, 2 methods are applied to determine the depression and give accurate results.

- i. **Data Gathering:** Data has been gathered from social media. From the available dataset posts are loaded and populated in to a database for the possible queries.
- ii. **Load Database:** To perform an action, sentiments from data set are taken as input. The empirical research has been carried out to perform the action on the sentiments.
- iii. **Preprocessing Phase:** This stage involves Tokenization, extract features and remove stop words.
  - a. **Tokenization:** Tokenization is a process used to partition the input text and make tokens. The tokens generally represents terms or words. Tokenization is based on the strategy to split all the characters in a sentence into pieces such as words, punctuation marks, numbers, and symbols. The punctuation marks and other unuseful characters are thus eliminated. The tokens like words, keywords act as input for other processes such as text mining.
  - b. **Stop word removal:** Stop words are the common words which are of no sense are eliminated from the input text. During the pre-processing stage a list of stop words has been created in the database. The stop words are removed from the text before the processing of text.

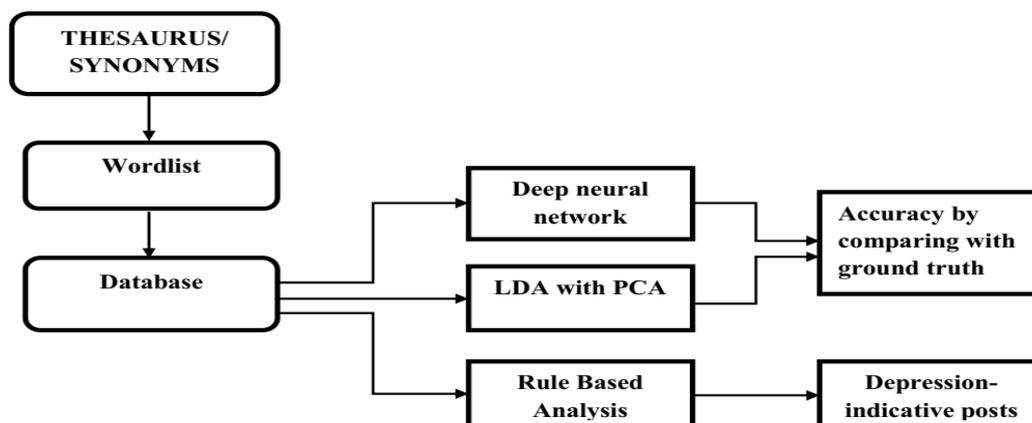


Fig. 2 Overview of the Research

## V. DISCUSSION

Data gathered from various resources is used to examine the system. The accuracy of the system may vary for different databases. The proposed system works on 6 features and uses 3 techniques rule based prediction, DNN and LDA method. Deep neural network plots a confusion matrix to represent depression class, non-depression class and misclassifications. The proposed system comes out with the finding that how the behavior of person can be detected to predict the depression-indicative posts. The experiment can be used to detect depression beforehand and it suggests to take proper treatment in case of depression.

## VI. CONCLUSION

Through our experimental findings, we have demonstrated how the social media posts could be extracted to reflect the behavior of person and also to predict whether the posts are depression-indicative or not. We come to the conclusion that mostly the people suffering from depression are active during night. This system automatically detects the signs of depression from postings and estimates the depression. We collected ground truth data from twitter and face book and devised various features to distinguish the posts such as positive craziness, negative craziness, egoistic and others to build Deep Neural Network and LDA. The DNN classifier predicts approximately 86% accuracy and LDA gives 67% accuracy whereas the single individual prediction estimates whether the individual is depression-active or not.

The method can be used to track the mental illness of those users who are online, where as the whole population is not analyzed. What is happening behind the scenes, we missed that while analyzing online users. The system would give different output for variant databases.

## VII. FUTURE SCOPE

We look at the potential for utilizing web-based social networking postings as another sort of focal point in understanding causes of depression in humans. Although the system predicts depression from various features but in future the features can be improved such as gender, friends list and likes can be added and the advanced predictive model can be developed to automatically gather online data from social sites. Automatically gathering posts would detect signs of depression which will result in better results before hand. The advanced factors can be considered to predict depression with better efficiency.

## REFERENCES

- [1] Akay, A., Dragomir, A. and Erlandsson, B.E., "A Novel Data-Mining Approach Leveraging Social Media to Monitor Consumer Opinion of Sitagliptin," *Journal of Biomedical Health Informatics*, 2015.
- [2] Akay, A., Dragomir, A. and Erlandsson, B.E., "Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care", 2013.
- [3] Akay, A., Dragomir, A. and Erlandsson, B.J., "Assessing anti-depressants using intelligent data monitoring and mining of online fora", *IEEE Journal of Biomedical and Health Informatics*, 2016.
- [4] Bremner, J. D., "Structural changes in the brain in depression and relationship to symptom recurrence," *CNS Spectrums*, vol. 7, no. 2, pp. 129–139, 2002.
- [5] Choudhury, M. D., Counts, S. and Horvitz, E., "Social media as a measurement tool of depression in populations", 2013.
- [6] Choudhury, M. D., Gamon, M., Counts, S. and Horvitz, S., 2013, "Predicting depression via social media," presented at the *Int. Conf. Weblogs Social Media*, Cambridge, MA, USA.
- [7] Choudhury, M.D., "Role of Social Media in Tackling Challenges in Mental Health", 2013.
- [8] Culjak, G. and Spranca, M., "Internet Improves Health Outcomes in Depression", *International Conference on System Sciences*, 2006.
- [9] Faedda<sup>1</sup>, N. Cerutti, R. Verdecchia<sup>1</sup>, P. Migliorini, D. Arruda, M. and Guidetti<sup>1</sup>, V. "Behavioral management of headache in children and adolescents", *The Journal of Headache and Pain*, 2016.
- [10] Frommholz and Lechtenfeld, M., "Determining the Polarity of Postings for Discussion Search," in *LWA-Workshop-Woche*, 2008.
- [11] Getoor, L. and Diehl, C.P., "Link mining: A survey," *SIGKDD Explor. Newslett.*, vol. 7, pp. 3- 12, 2005.

- [12] Getoor, Q.L.L., "Link Based Classification", *International Conference of Machine Learning*, 2003.
- [13] Han, J., Kamber, M., "Data Mining: Concepts and Techniques", Morgan Kaufmann, 2006.
- [14] Kaur, S. and Singh, L., "Improving a Novel Data Mining Approach Leveraging Social Media to Monitor Consumer Opinion of Sitagliptin", *IJIR*, 2016.
- [15] Li, X., Yu, W. and Villegas, S., "Structural Health Monitoring of Building Structures With Online Data Mining Methods", *IEEE*, vol.10, no. 3, 2016.
- [16] Mok, W.T., Sing, R., Jiang, X. and See, S.L., "Investigation of Social media on depression", *IEEE*, 2014.
- [17] Ng, A.Y., Zheng, A.X. and Jordan, M.I., "Stable algorithms for link analysis", in *Proc. SIGIR Conf. Inf. Retrieval, New Orleans, LA, USA*, pp. 258–266, 2001.
- [18] Nguyen, T., Phung, D., Dao, B., Venkatesh, S. and Berk, , "Affective and Content Analysis of Online Depression Communities", *IEEE*, 2013.
- [19] Noessner, J., Niepert, M., Meilicke, C. and Stuckenschmidt, H., "Leveraging Terminological Structure for Object Reconciliation", *ESWC*, pp 334-348, 2010.
- [20] Nowell, D. L. and Kleinberg, J. M., "The link prediction problem for social networks," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 57, pp. 556–559, 2004.
- [21] Paul, M. J. and Dredz̧e, M., "Discovering health topics in social media using topic models," *PLOS One*, vol. 9, Issue 8, 2014.
- [22] Rajagopals, S., "Customer data clustering using data mining technique", *IJDMS*, Vol.3, No.4, 2011.
- [23] Rosenquist, J.N., Fowler, J.H. and Christakis, N.A., "Social network determinants of depression", 2010.
- [24] Shokri, R. and Kocabas, I., "Big Data Analytics in Healthcare: Social & Biological Data", *International Journal of Engineering Inventions*, 2016.
- [25] Solanki, J. and Mulge, Y., "Different Techniques Used in Data Mining in Agriculture", *IJARCSSE*, Volume 5, Issue 5, 2015.
- [26] Taskar, B. Wong, M. F., Abbeel, P. and Koller, D., "Link prediction in relational data," *Advances Neural Information Processing Systems*, Vancouver, BC, USA, 2003.
- [27] Tsugawa, S., Kikuchi, Y., Kishino, Kosuke, F., Nakajima, K., Itob, Y. and Obsaki, H., "Recognizing Depression from Twitter Activity", 2015.
- [28] Umesb, D. R. and Ramachandra, B., "Association Rule Mining Based Predicting Breast Cancer Recurrence on SEER Breast Cancer Data", *IEEE*, 2015.
- [29] Watada J. and Yamashiro K, "A Data Mining Approach to Consumer Behavior," *First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC'06)*, Beijing, 2006, pp. 652-655.
- [30] Wu, J.L., Yu, L.C. and Chang, P.C., "Detecting causality from online psychiatric texts using inter-sentential language patterns", *BMC Medical Informatics and Decision Making*, 2012.
- [31] Zhao, K., Yen, J., Greer, G., Qiu, B., Mitra, P. and Portier, K., "Finding influential users of online health communities: A new metric based on sentiment influence," *J. Amer. Med. Inform. Assoc.*, vol. 21, pp. 212–218, 2014.