

# **BIG DATA ANALYTICS ON SOCIAL NETWORKS TO DETECT MELANCHOLY IN REAL-TIME**

Dr.C.K.Gomathy, Assistant Professor, Department of CSE, SCSVMV Deemed to be University, Kanchipuram

Mr. V. Ram Charan Tej , Mr. R. Venkata Narayana, Mr. T. Vamsi Krishna, Mr. P. Pavan Sree Ram

UG Scholars, Department of CSE, SCSVMV Deemed to be University, Kanchipuram

## **ABSTRACT**

During the coronavirus pandemic, the wide variety of despair instances has dramatically increased. Several melancholy sufferers disclose their true feeling with the aid of social media. Thus, large records analytics on social networks for real-time despair detection is proposed. Machine mastering techniques were utilized as the detection mannequin construction. There are five computing device mastering strategies explored in this research which are Support Vector Machine, Decision Tree, Naïve Bayes, Random Forest, and Deep Learning. The experimental outcomes revealed that the Random Forest technique performed higher accuracy than other methods to become aware of the depression.

Keywords: Big data analytics, Depression detection, social networks

## **I. INTRODUCTION**

Depression ability “a nation of mind that expresses temper problems such as depressed, unhappy, bored, loss of appetite, lack of concentration, anxiety, etc.”. Depression can also adversely affect the nice of life and it can ultimately lead to suicide. Depression can reason of substantial health deterioration and affect human beings all over the world. Machine getting to know techniques were utilized as the detection mannequin construction. There are five machine learning techniques explored in this research which are Support Vector Machine, Decision Tree, Random Forest, Naïve Bayes, and Deep Learning. Support Vector Machine tries to pick out the highest quality selection boundary by way of maximizing the margin distance between lessons using Hinge Loss characteristics mannequin parameters. Decision tree divides the dataset based on the attribute that divides the dataset most effectively. The attribute which provides most Information Gain is selected for splitting. Thus, the selection bushes must be least correlated with every other, i.e., every selection tree predicts the use of special features. Naïve Bayes classifies facts based totally on Bayes’ theorem. Deep Learning is based totally on synthetic neural networks, stimulated by using facts processing and allotted conversation nodes in biological brain. It makes use of a couple of hidden layers to extract facets from inputs.

## **II.LITERATURE SURVEY**

The first crew pursuits to analyse feelings from social community data. Barhan and Shakhomirov developed a model to classify feelings from Twitter data using emoticons and sentiments based totally on N-gram, a technique used to calculate the possibilities of phrases based totally on the preceding n-1 phrases from the goal word.

The second crew intends to notice despair from social community records using a number approach. Park et al. proposed a framework for analysing sentiments from Twitter and searching for aspects that identified depression. The experimental consequences found that the usage of poor messages, messages expressing depression, poor Twitter emoticon used to be drastically associated with the prevalence of depression symptoms.

The third crew concentrates on establishing a despair detection model. Hu et al. created classification and regression models the usage of behavioural and linguistic facts from 10,102 social network users. The consequences printed that records from social networks could be used to predict the despair in boost up to two months.

### **III.METHODOLOGY**

Real-time despair detection is a technique where gathers streaming information from Twitter Application Programming Interface (API) by means of extracting, transforming and loading data into records storage in a Hadoop cluster. These statistics will be used as an unseen input data of the despair detection model. The model is used as a classifier to discover a despair stage of a person. There are four modes of users in the device known as yourself, parent, advisor, and employer. In the yourself mode, customers can use their very own Twitter ID as an input for the gadget to observe their personal depression level. In the father or mother mode, mother and father can song a repute of their infant via imparting child's twitter ID to the system.

### **IV.IMPLEMENTATION**

Big Data Analytics objectives to improve despair detection model the usage of demographic traits and sentiment analysis from tweets. The lookup methodology consists of 5 procedures which are data acquisition, data transformation, data storage, model construction, and model performance evaluation. The details of every process are described as follows.

#### ***A.DATA ACQUISTION***

The records acquisition for model construction in this research is bought from three sources which are the PHQ-9, a personal statistics questionnaire, and Twitter API.

1. Firstly, the information derived from the PHQ-9 are survey response date and a melancholy assessment score of every participant.
2. Secondly, the information gathered from the private fact questionnaire comprise Twitter ID and demographic characteristics such as gender, age, weight, education, congenital disease, career, income, wide variety of family members, self-couple status, and parent's marital status.
3. Finally, Twitter API is used to accumulate Twitter user's information in actual time from the Twitter page. The Twitter ID gathered from the first supply is used to access the Twitter user's information.

#### ***B. DATA TRANSFORMATION***

The statistics processing for mannequin building the usage of computing device studying consists of three steps which are distinct as follows.

1. Extracting Twitter's User Information

The records acquired from Twitter may be processed to locate sentiment attributes and their values. The manner of extraction is illustrated underneath and may be defined as follows. Firstly, the Twitter ID is used to get right of entry to the Twitter user's records through Twitter API. The extracted records include tweets, retweets, hashtags, the quantity of friends, the quantity of followers, and durations of tweets. Secondly, phrases within the tweets, retweets, and hashtags are extracted the usage of an NLTK library.

## 2. Transforming the depression assessment score

The rankings acquired from the PHQ-nine are transformed from numeric to nominal. The rankings are transformed in to 4 lessons as follows: 5-eight factors are transformed to Level zero or no depression, nine-14 factors are transformed to Level 1 or mild depression, and extra than 15 factors are transformed to Level 2 or slight depression.

## 3. Loading relevant features

Before loading functions into statistics storage, attributes acquired from the preceding step should be applicable functions. Moreover, the desk consists of extra info and statistics forms of every feature. Irrelevant statistics together with Twitter ID and the survey reaction date are removed to save you needless computation within the device getting to know technique. These applicable functions could be loaded into the following manner.

### ***C. DATA STORAGE***

Data from the preceding method are loaded into a Hadoop cluster, a specialized pc cluster designed for storing and analysing a large quantity of data. This research employed a Hadoop's opensource software suite called Cloudera running on commodity computers. Data gathered from Twitter are kept in Hadoop Distributed File System (HDFS).

### ***D. MODEL CONSTRUCTION***

Machine learning is used to create the automated despair detection model. The type of laptop mastering applied for the mannequin building is a classification technique. More specifically, this research employs supervised desktop gaining knowledge of strategies for model creation, including Support Vector Machine, Naïve Bayes, Decision Tree, Random Forest and Deep Learning techniques. Model development is applied with Spark MLlib, Spark's computing device learning library using Python interface.

### ***E. MODEL PERFORMANCE EVALUATION***

The last technique is to consider the performance of despair detection model to locate the most excellent laptop gaining knowledge of method for detecting the depression. The Support Vector Machine method is used as a baseline of accuracy evaluation in view that the majority of preceding lookup work has relied on it. Comparing the results of special methods is measured by way of trendy measure rankings which are f-measure, accuracy, precision, and recall.

## **V.RESULT**

We have supplied the gadget that take real-time information from twitter and carry out sentiment evaluation to locate the different tiers of despair. The statistics of various tiers of despair are displayed within the shape of percent and pie chart. In this paper we used Twitter for tracking and detecting the despair amongst its users. Challenges in regions consisting of sampling, optimization of tactics to prediction and their features, generalizability, privacy, and different moral problems name for in addition research.

## VI.CONCLUSION

This article introduces a melancholy detection mannequin on social networks using huge statistics analytics. The foremost theoretical implication of this research is a novel mannequin primarily based on analysing demographic traits and textual content sentiment of Twitter users. Machine getting to know methods are applied for the despair detection model construction. There are five computer studying strategies explored in this research which are Support Vector Machine, Decision Tree, Naïve Bayes, Random Forest, and Deep Learning. The experimental effects published that the Random Forest technique completed greater accuracy than other desktop mastering methods to realize the depression.

## VII.REFERENCES

- [1] DR.C.K.Gomathy , V.Geetha , S.Madhumitha , S.Sangeetha , R.Vishnupriya Article: A Secure With Efficient Data Transaction In Cloud Service, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, March 2016, ISSN: 2278 – 1323.
- [2] Dr.C.K.Gomathy,C K Hemalatha, Article: A Study On Employee Safety And Health Management International Research Journal Of Engineering And Technology (Ijjet)- Volume: 08 Issue: 04 | Apr 2021
- [3] Dr.C K Gomathy, Article: A Study on the Effect of Digital Literacy and information Management, IAETSD Journal For Advanced Research In Applied Sciences, Volume 7 Issue 3, P.No-51-57, ISSN NO: 2279-543X,Mar/2018
- [4] Dr.C K Gomathy, Article: An Effective Innovation Technology In Enhancing Teaching And Learning Of Knowledge Using Ict Methods, International Journal Of Contemporary Research In Computer Science And Technology (Ijrcst) E-Issn: 2395-5325 Volume3, Issue 4,P.No-10-13, April '2017
- [5] Dr.C K Gomathy, Article: Supply chain-Impact of importance and Technology in Software Release Management, International Journal of Scientific Research in Computer Science Engineering and Information Technology ( IJSRCSEIT ) Volume 3 | Issue 6 | ISSN : 2456-3307, P.No:1-4, July-2018.
- [6] C K Gomathy and V Geetha. Article: A Real Time Analysis of Service based using Mobile Phone Controlled Vehicle using DTMF for Accident Prevention. International Journal of Computer Applications 138(2):11-13, March 2016. Published by Foundation of Computer Science (FCS), NY, USA,ISSN No: 0975-8887
- [7] C K Gomathy and V Geetha. Article: Evaluation on Ethernet based Passive Optical Network Service Enhancement through Splitting of Architecture. International Journal of Computer Applications 138(2):14-17, March 2016. Published by Foundation of Computer Science (FCS), NY, USA, ISSN No: 0975-8887
- [8] C.K.Gomathy and Dr.S.Rajalakshmi.(2014), "A Software Design Pattern for Bank Service Oriented Architecture", International Journal of Advanced Research in Computer Engineering and Technology(IJARCET), Volume 3,Issue IV, April 2014,P.No:1302-1306, ,ISSN:2278-1323.
- [9] C. K. Gomathy and S. Rajalakshmi, "A software quality metric performance of professional management in service oriented architecture," Second International Conference on Current Trends in Engineering and Technology - ICCTET 2014, 2014, pp. 41-47, doi: 10.1109/ICCTET.2014.6966260.
- [10] Dr.C K Gomathy, V Geetha ,T N V Siddartha, M Sandeep , B Srinivasa Srujay Article: Web Service Composition In A Digitalized Health Care Environment For Effective Communications, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, April 2016, ISSN: 2278 – 1323.
- [11]C.K.Gomathy.(2010),"Cloud Computing: Business Management for Effective Service Oriented Architecture" International Journal of Power Control Signal and Computation (IJPCSC), Volume 1, Issue IV, Oct - Dec 2010, P.No:22-27, ISSN: 0976-268X .

- [12] Dr.C K Gomathy, Article: A Study on the recent Advancements in Online Surveying , International Journal of Emerging technologies and Innovative Research ( JETIR ) Volume 5 | Issue 11 | ISSN : 2349-5162, P.No:327-331, Nov-2018
- [13] Dr.C.K.Gomathy,C K Hemalatha, Article: A Study On Employee Safety And Health Management International Research Journal Of Engineering And Technology (Irjet)- Volume: 08 Issue: 04 | Apr 2021
- [14] Dr.C K Gomathy, V Geetha , T.Jayanthi, M.Bhargavi, P.Sai Haritha Article: A Medical Information Security Using Cryptosystem For Wireless Sensor Networks, International Journal Of Contemporary Research In Computer Science And Technology (Ijrcrst) E-Issn: 2395-5325 Volume3, Issue 4, P.No-1-5, April '2017
- [15] C.K.Gomathy and Dr.S.Rajalakshmi.(2014), "Service Oriented Architecture to improve Quality of Software System in Public Sector Organization with Improved Progress Ability", Proceedings of ERCICA-2014, organized by Nitte Meenakshi Institute of Technology, Bangalore. Archived in Elsevier Xplore Digital Library, August 2014, ISBN:978-9-3510-7216-4.
- [16] Parameshwari, R. & Gomathy, C K. (2015). A Novel Approach to Identify Sullied Terms in Service Level Agreement. International Journal of Computer Applications. 115. 16-20. 10.5120/20163-2253.
- [17] C.K.Gomathy and Dr.S.Rajalakshmi.(2014),"A Software Quality Metric Performance of Professional Management in Service Oriented Architecture", Proceedings of ICCTET'14, organized by Akshaya College of Engineering, Coimbatore. Archived in IEEE Xplore Digital Library, July 2014,ISBN:978-1-4799-7986-8.
- [18] C.K.Gomathy and Dr.S.Rajalakshmi.(2011), "Business Process Development In Service Oriented Architecture", International Journal of Research in Computer Application and Management (IJRCM) , Volume 1, Issue IV, August 2011, P.No:50-53, ISSN : 2231-1009

#### **AURTHORS PROFILE**



V. Ram Charan Tej Student, B.E. Computer Science and Engineering, Sri Chandrasekharana Saraswathi Viswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.



T. Vamsi Krishna Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.



R. Venkata Narayana Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.



P. Pavan Sree Ram Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.



Dr.C.K.Gomathy is Assistant Professor in Computer Science and Engineering at Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. Her Area of Interest in Big Data Analytics.