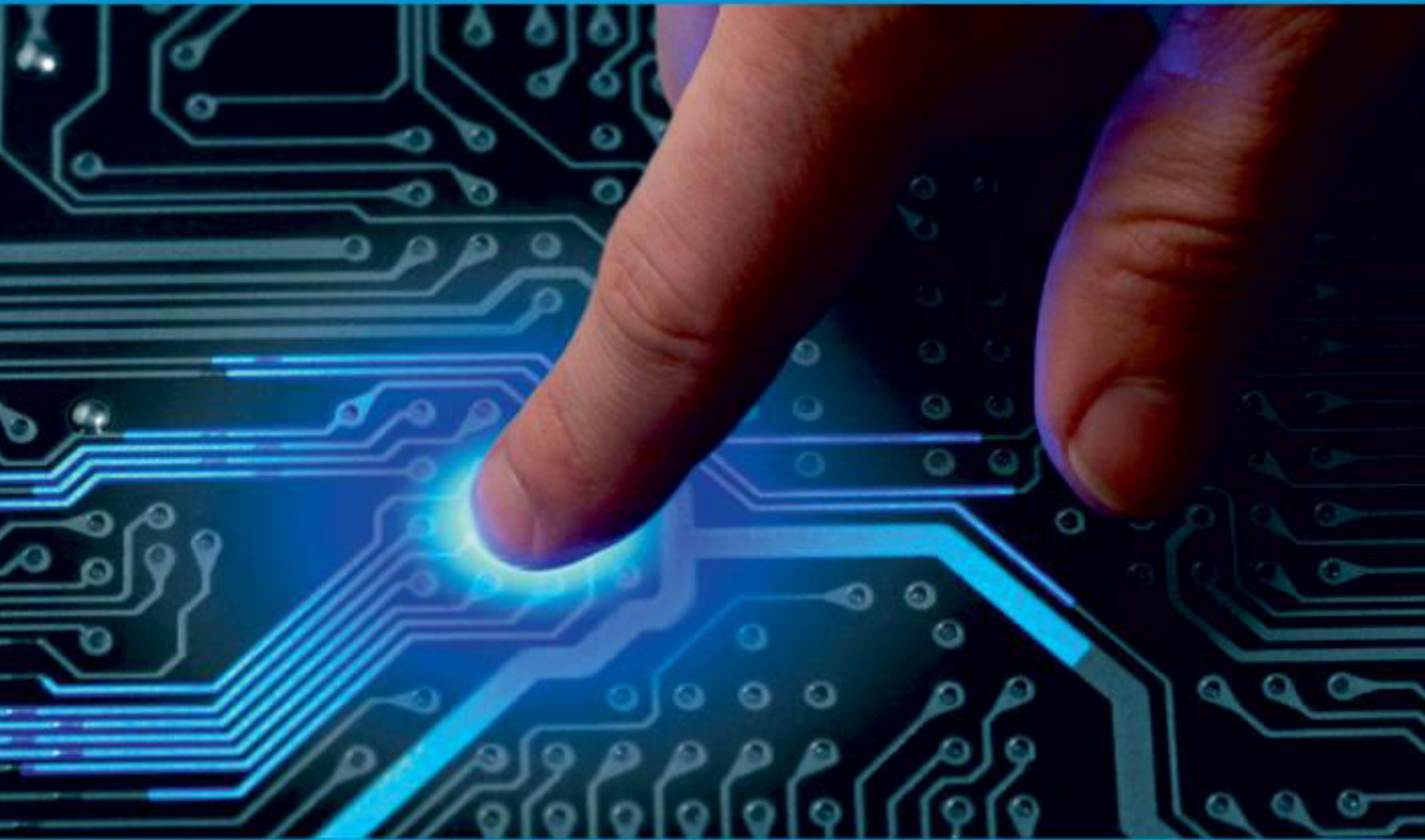
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006114 |

Tweets Based Emotion Detection Using Machine Learning Algorithms

Gowshika J, Dr. S Jayasundar

PG Scholar, Department of Computer Science and Engineering, Idhaya Engineering College for women, Chinnasalem, Kallakurichi, India

Professor, Department of Computer Science and Engineering, Idhaya Engineering College for Women, Chinnasalem, Kallakurichi, India

ABSTRACT: An Improved RNN language model is put forward using LSTM, which successfully covers all history sequence information and performs better than conventional RNN. It is applied to achieve multiclassification for text emotional attributes, and identifies text emotional attributes more accurately. Even Today Sentiment Analysis is still a difficult and Complex problem in computer Science. Sentiments are express by Humans in Different Ways. The objective of this work is to create a system for finding the emotion from tweets and emoticons for users in a user friendly way. In future work, we could try these improvement programs or use different models combination to improve the performance of text sentiment analysis. The models will be trained and validated against a test dataset.

KEYWORDS: RNN, MACHINE LEARNING ALGORITHMS, EMOTION DETECTION, TWEETS

I. INTRODUCTION

Recent advances in social networks increase the ways of explaining ideas on diverse subjects. Moreover users can share their opinions with their online friends in a collaborative manner. All that rich information sources make the social networks a suitable working base for researchers. Effective methodologies and techniques are required to extract various kinds of information from social networks automatically. Among them, identifying users' sentiments on a product or service has turned into a valid indicator of marketing success. Apart from the classical sentiment analysis algorithms, networked data include valuable relationship information that can contribute to this analysis process beside textual contexts that are produced by the users. Such data is can be useful in emotion analysis, in which, instead of detection of sentimentality, the type of sentimentality is further detected in terms of different kinds of emotions.

In this study, collective classification algorithms, which constitute a sub-field of link mining field, are applied within the context of emotion analysis in microblogs. As the microblog, Twitter is used as the data source. In our setting, Twitter users are nodes and their relationships are edges, which are extracted from retweets or user mentions (@) in tweets. Giving graph structure as input to collective classification framework, unknown emotion labels for users are predicted by utilizing their labeled neighbors. The performance of relational classifiers are experimented under different configurations.

Since the collected tweets are in Turkish, in addition to tokenization, Turkish morphological analysis and stemming are applied as well. However, apart from this, all of the remaining methods are equally applicable to the texts in other languages as well. With the aim of applying collective classification techniques on the context of emotion analysis in social networks, to the best of our knowledge, this is the first work in the literature.

Textual emotion mining has quite lot of applications in today's world. The applications include modern devices which sense person's emotion and suggest music, restaurants, or movies accordingly, product marketing can be improved based on user comments on products which in turn helps boost product sales. Other applications of textual emotion

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006114 |

mining are summarized by Yadollahi et.al [30] and include: in customer care services, emotion mining can help marketers gain information about how much satisfied their customers are and what aspects of their service should be improved or revised to consequently make a strong relationship with their end users [7].

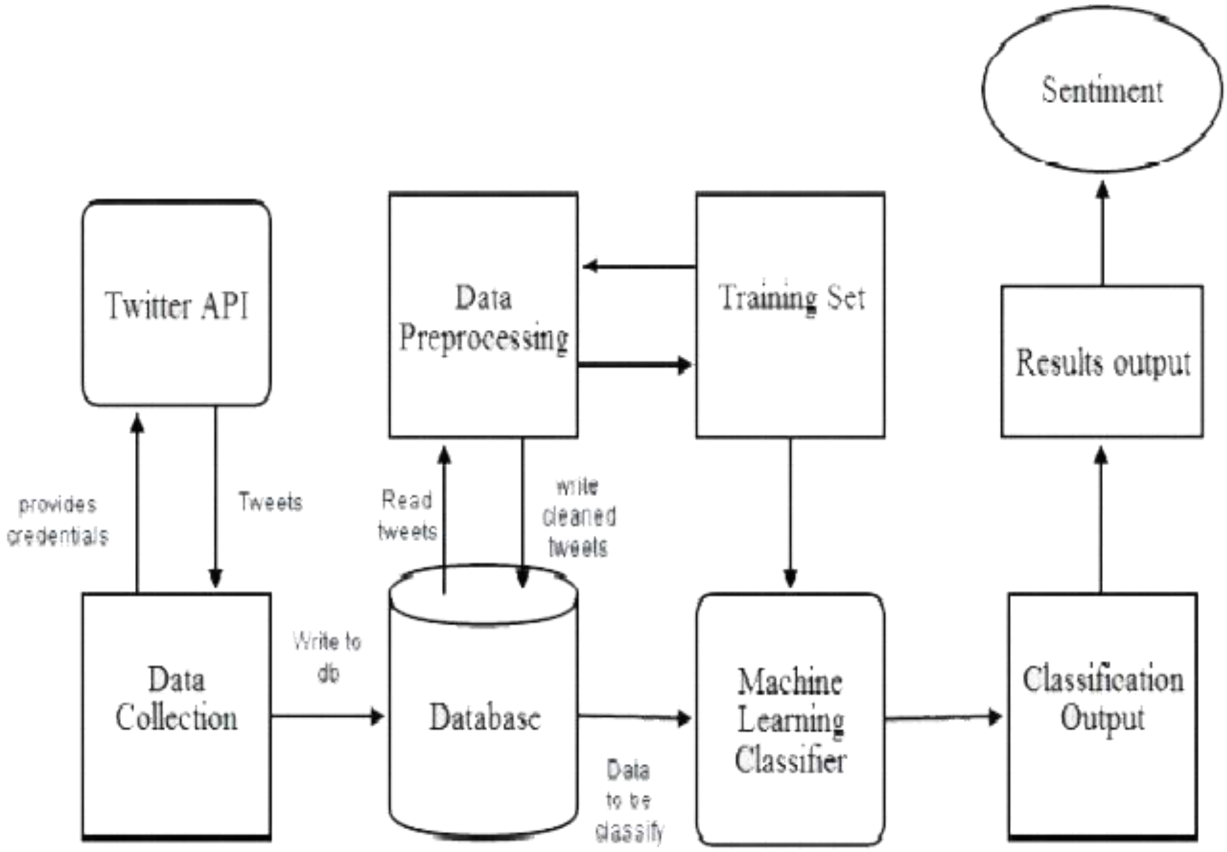


Fig1: Workflow diagram

User's emotions can be used for sale predictions of a particular product. In e-learning applications, the intelligent tutoring system can decide on teaching materials, based on user's feelings and mental state. In Human Computer Interaction, the computer can monitor user's emotions to suggest suitable music or movies [26]. Having the technology of identifying emotions enables new textual access approaches such as allowing users to filter results of a search by emotion. In addition, output of an emotion-mining system can serve as input to other systems. For instance, Rangel and Rosso [22] use the emotions detected in the text for author profiling, specifically identifying the writer's age and gender. Last but not least, psychologists can infer patients' emotions and predict their state of mind accordingly. On a longer period of time, they are able to detect if a patient is facing depression or stress [3] or even thinks about committing suicide, which is extremely useful, since he/she can be referred to counseling services [12]. Though this automatic method might help in detecting psychology related issues, it has some ethical implications as it is concerned with human emotion and their social dignity. In such cases it is always ethical to consult human psychiatrist along with the automatic systems developed.

II. RELATED WORK

Basically, there are two main approaches in the literature for emotion detection on texts. The first one is the text classification based methods that build classifiers from labeled text data as in traditional supervised learning. The second method that detects the emotional states especially in tweets is the lexicon-based approach. For this approach, an emotion lexicon is constructed.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006114 |

Regarding the field of psychology, Ekman (Ekman, 1999) defined 7 emotions that are categorized by observable human facial expressions. Kozareva et al. (Kozareva et al., 2007) classified news headlines using these emotion classes. They averaged different web search engines hit counts' on the emotional classes and news headlines as query words.

Alm et al. (Alm et al., 2005) presented empirical results of applying supervised machine learning techniques to categorize English fairy tale sentences into different emotions. They proposed their own textbased classifier algorithm (SNoW) and it achieved significant accuracy results. Go et al. (Go et al., 2009) applied supervised learning methods to classify collected Twitter data into binary sentiments as positive or negative.

Boynukaln (Boynukalin, 2012) worked on two data sets. One of them is the Turkish translation of ISEAR 1 data and the other is the manually labeled Turkish fairy tales. Emotion levels are predicted using different n-gram feature constructions and weighted log likelihood algorithm (Nigam et al., 2000) is utilized to determine most significant features.

Akba et al. (Akba et al., 2014) investigated the feature selection methods on Turkish movie reviews. They labeled their corpus by dividing emotions into three categories as positive, negative and neutral. In their experiments, supervised methods had been employed for the classification of movie reviews into two or three categories. Tocoglu et al. (Tocoglu and Alpkocak, 2014) proposed an emotion extraction system from Turkish texts, which is based on text classification approach. Applying Naive Bayes classifier in Weka achieved promising accuracy result.

Demirci (Demirci, 2014) classified Turkish tweets into six emotion categories (anger, surprised, fear, sadness, joy and disgust) with supervised learning. Beside the classical text preprocessing operations, morphological analysis is applied as well. Finally, several supervised classification methods are compared with the baseline algorithm of Boynukaln (Boynukalin, 2012).

Chakrabarti et al. (Chakrabarti et al., 1998) work on the problem of categorizing related news objects in the Reuters dataset. They are the first to leverage class labels of related instances and also their attributes. Although using class labels improves classification accuracy, the same thing does not apply for considering attributes. Neville and Jensen (Neville and Jensen, 2000) propose a simple link based classification method that classifies corporate datasets involving heterogeneous graphs with different set of features.

Lu and Getoor (Lu and Getoor, 2003) aim to enhance traditional machine learning algorithm by introducing new features that are built out of correlations between objects. As a result, a new link based classification algorithm that uses probability terms such as Markov blanket of related class labels, is developed.

Pang and Lee (Pang and Lee, 2004) seek to determine sentiment polarities of movie reviews by extracting subjective portions of the sentences. For this purpose, they use a graph-based technique that finds the minimum cuts. By this way, contextual information is added in polarity classification process and significant accuracy improvement is achieved. In the literature, the most similar study to our work is the one by Rabelo et al. (Rabelo et al., 2012), which proposes an user centric approach on the context of sentiment analysis. They have classified Twitter user's political opinions into binary classes by using collective classification. Their algorithm takes a partially labeled graph, applies a graph pruning process and runs the collective classification. Preliminary experiments have shown promising results.

Authors Wang et al. [27] built a dataset from Twitter, containing 2,500,000 tweets and use hashtags as emotion labels. In order to validate the hashtag labeling, they randomly select 400 tweets to label them manually. Then they compared manual labels and hashtag labels which had acceptable consistency. They explored the effectiveness of different features such as n-grams, different lexicons, part-of-speech, and adjectives in detecting emotions with accuracy close to 60%. Their best result is obtained when unigrams,bigrams, lexicons, and part-of-speech are used together.

Authors Xia et al. [29], propose distantly supervised lifelong learning framework for Sentiment Analysis in social media text. They use following two large-scale distantly supervised social media text datasets to train the lifelong

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006114 |

learning model: Twitter corpus (English dataset) [25], and Chinese Weibo dataset collected using Weibo API. This work focuses on continuous sentiment learning in social media by retaining the knowledge obtained from past learning and utilize the knowledge for future learning. They evaluate the model using nine standard datasets, out of which 5 are English language datasets and 4 are Chinese datasets. The main advantage of this approach is that it can serve as a general framework and compatible to any single task learning algorithms like naive bayes, logistic regression and support vector machines

Authors Hasan et al. [8] also validate the use of hashtags as emotion labels on a set of 134,000 tweets. They compared hashtag labels with labels assigned by crowd-sourcing and by a group of psychologist's. It is found that crowd labels are not consistent within themselves; On the other-hand psychologist's labels are more consistent with hashtags. They developed a supervised classifier, named "EmoTex" which uses the feature set of unigrams, list of negation words, emoticons, and punctuation's and runs k-nearest neighbors (KNN) and support vector machines on the training data achieved 90% accuracy for multi-class emotion detection.

III. METHODS

Real-time data collection

The posts on Twitter aka tweets are obtained in real-time by creating an app via the Twitter API. For accessing the real-time tweets, we use the open-source java library Twitter4J. Via Twitter4J, a user can do the following tasks:

• Post tweets

• View user timeline

• Access real-time latest tweets

• Send and receive messages Along with these tasks, this library also ensures the security and privacy of the user. The obtained tweets are stored in the database. The advantage of using real-time data is that analysis is done using the latest data whenever required.

Sentiment Analysis

It is a process of classifying a given text as expressing a positive, negative or neutral sentiment. For that purpose, we are using the SentiWordNet dictionary in which each word is given a score according to positive or negative information displayed. Using that scores, a tweet is probabilistically classified as positive, negative or neutral Feature extraction using NLP In this step, repetitive tweets' content, that occur mostly due to retweets by other users to a particular user's tweet, are removed using n-gram generation by using tweet-id. For extracting textual features to train our classification model, we used two NLP techniques viz. PoS-Tagging and Topic Modeling. Part-Of-Speech tagging is also called as grammatical tagging in which the words of a particular text are tagged with the corresponding Part-of-Speech. This process is very helpful for identifying the most useful features which in turn are useful for identifying the emotions.

Emotion classification

In this step, two classification models are built and trained with emotion-word features obtained from [6]. The first is the Unison model which is basically the traditional emotion classification model based on Bag-of-Words model. POMS is used for its implementation. The second model is our proposed Random Forest (RF) model. It is built by training it with emotionlabelled tweets. The random forest algorithm is used for solving Quadratic Programming (QP) problems and hence it is chosen to build a multiclass classification model for emotion detection.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006114 |

IV. RESULTS AND DISCUSSIONS

An emotion recognition web application is developed using Eclipse IDE, MySQL and Apache Tomcat server. The coding is done in Java. Java servlets are created. The Twitter API is used to obtain real-time tweets for analysis by developers via the Twitter4J library. The retrieved tweets are preprocessed in suitable format. The SentiWordNet dictionary is used for sentiment analysis of the tweets.

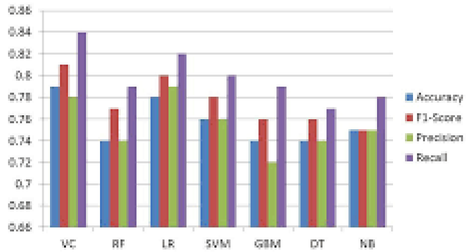


Fig2: Emotion Recognition by Textual Tweets

NLP algorithms are then used for extracting informative features. The Unison model is the previously developed model for emotion recognition based on POMS [6]. Proposed model for emotion recognition is developed as explained in section 3.1. The classification is performed using both the classifiers. The labels of real-time test data are assigned according to the probability of occurrence of emotion words predefined in [6] which later helps in performance evaluation of the classifiers. Their performances are expressed in terms of performance measures. The metrics of classification performance evaluation used are accuracy, precision, recall and f1-score

V. CONCLUSION

We have addressed the problem of classifying text into the six basic Emotion-Categories, rather than just labeling them as positive or negative. Through our research and a self-generated reliable bag of emotional words (EWS), we can now effectively quantify various emotions in any block of text. We have also automatically generated a labeled training-set (without manually labeling the tweets) of emotionally-biased tweets using a keyword-matching approach, which was then used to train various classifiers. Moreover, we have also introduced the concept of Surety Factor to suggest the reliability of our output and the degree of usefulness and correctness of our results. Finally, we visualized our results using pie-charts, bargraphs and maps, and demonstrated the various applications of our analysis. In future, a system could be established for automatically updating the bag-of-words which we created, on the basis of new tweets and data analysed. Using our approach, many interesting apps can be created, such as an add-on to a social-networking site displaying the recent mood of each of your friends. Also, our analysis of Twitter can be extended to the development of a real-time system, analyzing mood-swings and emotions on Twitter.

REFERENCES

[1] Johan Bollen, Huina Mao, and Alberto Pepe. 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. Icwsm 11 (2011), 450–453.
[2] Leo Breiman. 2001. Random forests. Machine learning 45, 1 (2001), 5–32.
[3] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. ICWSM 13 (2013), 1–10.
[4] Cicero dos Santos and Maira Gatti. 2014. Deep convolutional neural networks for sentiment analysis of short texts. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 69–78.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006114 |

[5] Elaine Fox. 2008. Emotion science cognitive and neuroscientific approaches to understanding human emotions. Palgrave Macmillan.
[6] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford 1, 12 (2009).
[7] Narendra Gupta, Mazin Gilbert, and Giuseppe Di Fabbrizio. 2013. Emotion detection in email customer care. Computational Intelligence 29, 3 (2013), 489– 505.
[8] Maryam Hasan, Emmanuel Agu, and Elke Rundensteiner. 2014. Using hashtags as labels for supervised learning of emotions in twitter messages. In ACM SIGKDD Workshop on Health Informatics, New York, USA.
[9] Robert C Holte. 1993. Very simple classification rules perform well on most commonly used datasets. Machine learning 11, 1 (1993), 63–90.
[10] Sunghwan Mac Kim, Alessandro Valitutti, and Rafael A Calvo. 2010. Evaluation of unsupervised emotion models to textual affect recognition. In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. Association for Computational Linguistics, 62–70.
[11] Ron Kohavi. 1995. The power of decision tables. In European conference on machine learning. Springer, 174–189.
[12] Kim Luyckx, Frederik Vaassen, Claudia Peersman, and Walter Daelemans. 2012. Fine-grained emotion detection in suicide notes: A thresholding approach to multi-label classification. Biomedical informatics insights 5 (2012), BII–S8966.
[13] Chunling Ma, Helmut Prendinger, and Mitsuru Ishizuka. 2005. Emotion estimation and reasoning based on affective textual interaction. In International Conference on Affective Computing and Intelligent Interaction. Springer, 622–628.
[14] Kevin Makice. 2009. Twitter API: Up and running: Learn how to build applications with the Twitter API. " O'Reilly Media, Inc.".
[15] Xiangrui Meng, Joseph Bradley, Burak Yavuz, Evan Sparks, Shivaram Venkataraman, Davies Liu, Jeremy Freeman, DB Tsai, Manish Amde, Sean Owen, et al. 2016. Mllib: Machine learning in apache spark. The Journal of Machine Learning Research 17, 1 (2016), 1235–1241.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details

www.ijircce.com