Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4056

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Dynamic Content based Behavior Analysis Model for Users on Social Media

S.Lakshmi[1], Jaswant S.J.A[2], Mageshwara Kumar C.R.L[3], Mohammed Adeen[4]
[1]*Department of Computer Science and Engineering , SRM University, Chennai.*
[2, 3, 4] *Computer Science and Engineering , SRM University, Chennai.*

Abstract: Social media is now widely used for uploading photos and sharing one's thoughts by posting them as status. This context collected from users is better than the conventional ones which requires tedious user involvement to study their behavioral characteristics. These posts made by individuals in their profile directly or indirectly reflects the interests, opinions and thoughts over a particular topic or an incident. By analyzing those posts, the correlation between the contents of the post and behavior characteristics can be understood. However it is difficult to find the individual's characteristics accurately since the posts have to be monitored over a certain period of time and all the posts within the period of time has to be analyzed together. The presented approach investigates the datasets containing the user posts over a period of time and predict the user's behavioral characteristics. We also define and classify a number of characteristics based on the posts made by an individual. The results obtained after the analysis reveal the behavioral characteristics of the user which can be used for service provision to a specific type of crowd.
Keywords: User, Posts, Behaviour, Characteristics, Analysis

I. INTRODUCTION

Data mining, also called knowledge discovery in database, in computer science, the process of discovering interesting and useful patterns and relationship in large volumes of data. Data mining is the practice of automatically searching large stores of data to discover patterns and trends that go beyond simple analysis. There are many types of data mining typically divided by the kind of information. Big data is data sets that are so voluminous and complex that traditional data processing application software are inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, query, updating, information and data source.

II. OBJECTIVE

To analyze the behavior of a user in social media with the help of the dataset containing the tweets posted by the user in the user's twitter account.

III. LITERATURE SURVEY

In the paper , "Design of A Universal User Model for Dynamic Crowd Preference Sensing and Decision-Making Behaviour Analysis."[1].It is based on the exploration of dynamic crowd preference and decision-making behavior analysis based on users' previous behavior/behaviors through the external factors. The INPUT is obtained from the IDS and the EDS and collectively stored in the Big Data Pool .In the Knowledge Discovery Engine, pre-processing is done to remove noise and ensure quality of data. The data is then analyzed on sentiment, emotion, behavior and preference. This approach using the Apriori and Association rule mining algorithm results in extremely slow candidate generation. The candidate generation could also generate duplicates depending on the implementation and also the counting method also makes the algorithm a lot heavier and resulting in huge memory consumption. In our paper, we suggest two algorithms namely content based filtering and collaborative filtering, which can overcome the above problem.

In the paper, "A Dataset for Psychological Human Needs Detection From Social Networks."[5], a theoretical based multi-layer framework for a psychological needs analysis that is guided by research in the field of motivational psychology is proposed. The framework's layers are constructed to identify the psychological needs, measure their satisfaction level, and assess the individual's surrounding environment in different aspects of life. A psychological needs corpus: a collection of Twitter posts annotated based on three universal needs proposed by the self-determination theory framework is created and Several techniques were employed to encourage high-quality annotations. A Descriptive statistics of the annotated corpus is provided. This corpus can be used in the development of automatic detection systems and predication models to detect individual needs and measure their satisfaction. It can also be used to better interpret and understand the individual's surrounding social contexts.

In the paper , "Influence analysis of emotional behaviors and user relationships based on Twitter data."[9] the influence of emotional behaviors to user relationships based on Twitter data using two dictionaries of emotional words is analyzed and Emotion scores are calculated via keyword matching. Moreover, three experiments with different settings: calculate the average emotion score of a user with random sampling, calculate the average emotion score using all emotional tweets, and calculate the average emotion score using emotional tweets, excluding users of few emotional tweets is designed. The influence of emotional behaviors is evaluated and the result shows that a positive user is more active than a negative user in constructing user relationships in a specific condition. In our paper, this methodology of calculating the scores for the content is incorporated.

IV. EXISTING SYSTEM

The input is obtained from the IDS and the EDS and collectively stored in the Big Data Pool, from which the behavior analysis is done using the Apriori and Association rule mining algorithm. The dataset used here in existing system is Airbnb by considering factors such as nationality, gender and age as internal factors and social media, device and time as external factors. These two factors reflect the preference of the crowd and its behavior. This paper targets on two main things such as crowd preference and decision making behavior.

V. PROPOSED SYSTEM

In the proposed system, we have implemented two effective algorithms namely, Content-based filtering and Collaborative filtering. We use the dataset from twitter. This paper targets the behavior analysis of the users based on the tweets in their Twitter profile .We ultimately aim to produce a graphical representation (charts, graphs, histograms etc.) to show how much the person is inclined towards each characteristics. So, the characteristics we wish to analyze are Emotional Quotient, Scientific Curiosity, Aggressive Nature and Social Awareness. We initially perform the pre-processing followed by the behavior analysis of the user.

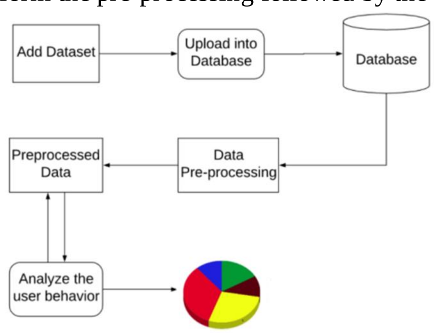


Fig. 1 Architecture of the Proposed System

A. Content-based Filtering

Content-based filtering, also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user profile. The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document. The user profile is represented with the same terms and built up by analyzing the content of items which have been seen by the user. In our paper , we use content based algorithm to compare the data with the words pertaining to certain traits and calculate the level of behavior traits present in the user.

B. Collaborative Filtering

Collaborative filtering also referred to as social filtering, filters information by using the recommendations of other people. It is based on the idea that people who agreed in their evaluation of certain items in the past are likely to agree again in the future. In our paper , collaborative filtering algorithm is used for identifying certain traits which can be identified only through a collaborative method of combining the user's posts and analyzing them.

C. Pre-Processing

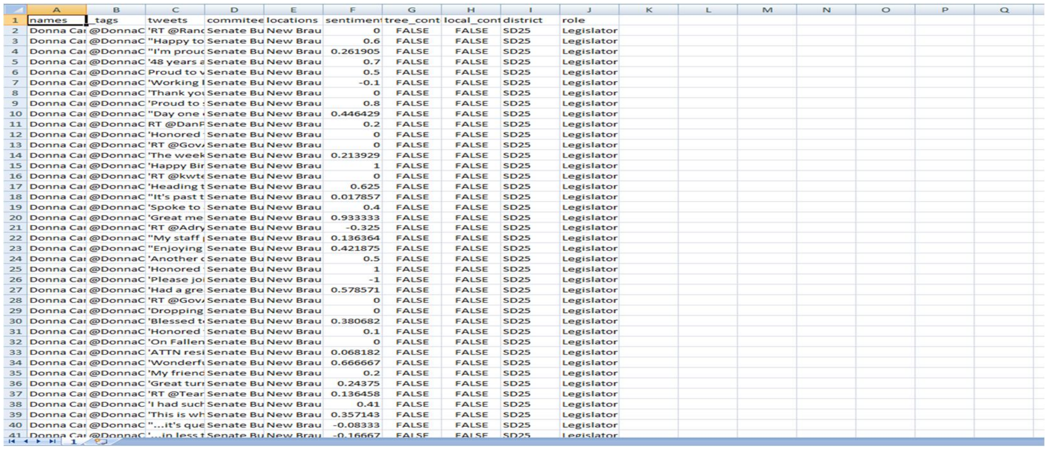
In this step, the stop words as removed from the user data by comparing each word with the words in the stop words library. After stop words removal, stemming process i.e. Converting each word to it root or stem form is done.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

D. Behavior Analysis

We use the two algorithms, content based filtering and collaborative filtering to analyze the behavioral characteristic of the user.

E. Experimental Results



	A	B	C	D	E	F	G	H	I	J
1	names	_tags	tweets	commitee	locations	sentiment	tree_cont	local_cont	district	role
2	Donna Ca	@DonnaC	'RT @Ranc	Senate Bu	New Brau	0	FALSE	FALSE	SD25	Legislator
3	Donna Ca	@DonnaC	"Happy to	Senate Bu	New Brau	0.6	FALSE	FALSE	SD25	Legislator
4	Donna Ca	@DonnaC	"I'm prouc	Senate Bu	New Brau	0.261905	FALSE	FALSE	SD25	Legislator
5	Donna Ca	@DonnaC	'48 years a	Senate Bu	New Brau	0.7	FALSE	FALSE	SD25	Legislator
6	Donna Ca	@DonnaC	Proud to v	Senate Bu	New Brau	0.5	FALSE	FALSE	SD25	Legislator
7	Donna Ca	@DonnaC	'Working l	Senate Bu	New Brau	-0.1	FALSE	FALSE	SD25	Legislator
8	Donna Ca	@DonnaC	'Thank you	Senate Bu	New Brau	0	FALSE	FALSE	SD25	Legislator
9	Donna Ca	@DonnaC	'Proud to :	Senate Bu	New Brau	0.8	FALSE	FALSE	SD25	Legislator
10	Donna Ca	@DonnaC	"Day one	Senate Bu	New Brau	0.446429	FALSE	FALSE	SD25	Legislator
11	Donna Ca	@DonnaC	RT @DanF	Senate Bu	New Brau	0.2	FALSE	FALSE	SD25	Legislator
12	Donna Ca	@DonnaC	'Honored	Senate Bu	New Brau	0	FALSE	FALSE	SD25	Legislator
13	Donna Ca	@DonnaC	'RT @Gov	Senate Bu	New Brau	0	FALSE	FALSE	SD25	Legislator
14	Donna Ca	@DonnaC	'The week	Senate Bu	New Brau	0.213929	FALSE	FALSE	SD25	Legislator
15	Donna Ca	@DonnaC	'Happy Bir	Senate Bu	New Brau	1	FALSE	FALSE	SD25	Legislator
16	Donna Ca	@DonnaC	'RT @kwte	Senate Bu	New Brau	0	FALSE	FALSE	SD25	Legislator
17	Donna Ca	@DonnaC	'Heading t	Senate Bu	New Brau	0.625	FALSE	FALSE	SD25	Legislator
18	Donna Ca	@DonnaC	"It's past t	Senate Bu	New Brau	0.017857	FALSE	FALSE	SD25	Legislator
19	Donna Ca	@DonnaC	'Spoke to	Senate Bu	New Brau	0.4	FALSE	FALSE	SD25	Legislator
20	Donna Ca	@DonnaC	'Great me	Senate Bu	New Brau	0.933333	FALSE	FALSE	SD25	Legislator
21	Donna Ca	@DonnaC	'RT @Adry	Senate Bu	New Brau	-0.325	FALSE	FALSE	SD25	Legislator
22	Donna Ca	@DonnaC	"My staff	Senate Bu	New Brau	0.136364	FALSE	FALSE	SD25	Legislator
23	Donna Ca	@DonnaC	"Enjoying	Senate Bu	New Brau	0.421875	FALSE	FALSE	SD25	Legislator
24	Donna Ca	@DonnaC	'Another c	Senate Bu	New Brau	0.5	FALSE	FALSE	SD25	Legislator
25	Donna Ca	@DonnaC	'Honored	Senate Bu	New Brau	1	FALSE	FALSE	SD25	Legislator
26	Donna Ca	@DonnaC	'Please joi	Senate Bu	New Brau	-1	FALSE	FALSE	SD25	Legislator
27	Donna Ca	@DonnaC	'Had a gre	Senate Bu	New Brau	0.578571	FALSE	FALSE	SD25	Legislator
28	Donna Ca	@DonnaC	'RT @Gov	Senate Bu	New Brau	0	FALSE	FALSE	SD25	Legislator
29	Donna Ca	@DonnaC	'Dropping	Senate Bu	New Brau	0	FALSE	FALSE	SD25	Legislator
30	Donna Ca	@DonnaC	'Blessed t	Senate Bu	New Brau	0.380682	FALSE	FALSE	SD25	Legislator
31	Donna Ca	@DonnaC	'Honored	Senate Bu	New Brau	0.1	FALSE	FALSE	SD25	Legislator
32	Donna Ca	@DonnaC	'On Fallen	Senate Bu	New Brau	0	FALSE	FALSE	SD25	Legislator
33	Donna Ca	@DonnaC	'ATTN resi	Senate Bu	New Brau	0.068182	FALSE	FALSE	SD25	Legislator
34	Donna Ca	@DonnaC	'Wonderfu	Senate Bu	New Brau	0.666667	FALSE	FALSE	SD25	Legislator
35	Donna Ca	@DonnaC	'My friend	Senate Bu	New Brau	0.2	FALSE	FALSE	SD25	Legislator
36	Donna Ca	@DonnaC	'Great turr	Senate Bu	New Brau	0.24375	FALSE	FALSE	SD25	Legislator
37	Donna Ca	@DonnaC	'RT @Tear	Senate Bu	New Brau	0.136458	FALSE	FALSE	SD25	Legislator
38	Donna Ca	@DonnaC	'I had such	Senate Bu	New Brau	0.41	FALSE	FALSE	SD25	Legislator
39	Donna Ca	@DonnaC	'This is wh	Senate Bu	New Brau	0.357143	FALSE	FALSE	SD25	Legislator
40	Donna Ca	@DonnaC	"...it's que	Senate Bu	New Brau	-0.08333	FALSE	FALSE	SD25	Legislator
41	Donna Ca	@DonnaC	'...in less t	Senate Bu	New Brau	-0.16667	FALSE	FALSE	SD25	Legislator

Fig. 2 Dataset containing the user's tweets

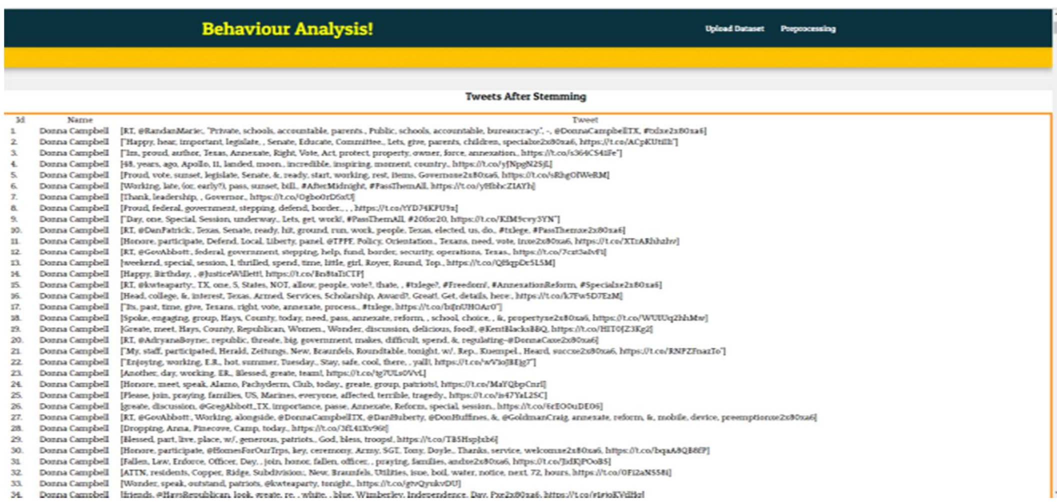


Fig. 3 User's tweets after the Stemming process is done

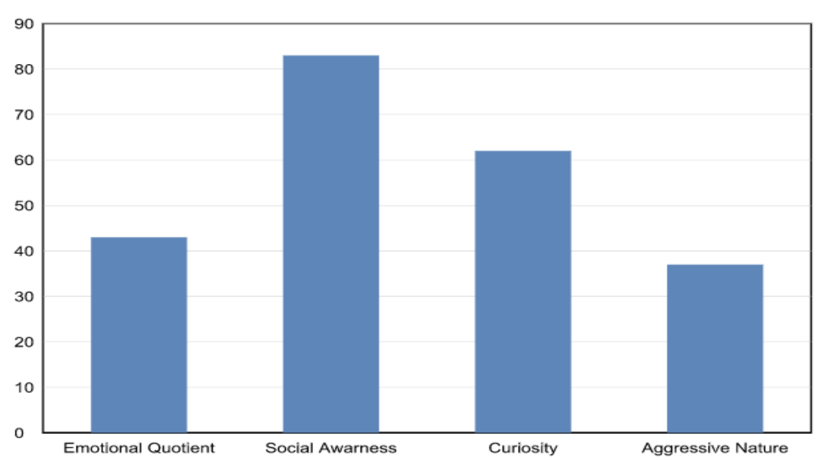


Fig. 4 Graphical representation of the user's behaviour

F. Admin

The admin logs into the admin homepage and the admin uploads the user data from twitter in the form of a dataset. Once the admin uploads the dataset, a message indicating the successful upload of the dataset is displayed.

G. Database

The user data containing the Username, Twitter ID, Tweets are stored in the database.

VI. CONCLUSION

In this paper, we have shown that the methodology used by us in analyzing the user's behavior through the user's tweets is quicker and more efficient than the existing system where the behavior analysis is done using the Apriori and Association rule mining algorithm. This can be extended to a higher level for a large number of users.

REFERENCE

[1] Moloud Abdar and Neil Y. Yen , "Design of A Universal User Model for Dynamic Crowd Preference Sensing and Decision-Making Behavior Analysis." IEEE August 2017.

[2] P. M. Falcone and E. Imbert, ``Bringing a sharing economy approach into the food sector: The potential of food sharing for reducing food waste,"in Food Waste Reduction and Valorisation. Cham, Switzerland: Springer,2017, pp. 197_214.

[3] R. Belk, ``Sel_sh sharing," in NA_Advances in Consumer Research,vol. 44, P. Moreau and S. Puntoni, Eds. Duluth, MN, USA: Association Consumer Research, 2016, pp. 37_41.[Online]. Available:http://acrwebsite.org/volumes/1021384/volumes/v44/NA-44.

[4] A. Fagerstrøm, S. Pawar, V. Sigurdsson, G. R. Foxall, and M. Yani-de-Soriano, ``That personal pro_le image might jeopardize your rental opportunity! On the relative impact of the seller's facial expressions upon buying behavior on AirBnb," Comput. Human Behavior, vol. 72,pp. 123_131, Jul. 2017.

[5] Rajwa Alharthi , Benjamin Guthier , Camille Guertin , Abdulmotaleb El Saddik , "A Dataset for Psychological Human Needs Detection From Social Networks." IEEE May 2017.

[6] K. Starbird, L. Palen, A. L. Hughes, S. Vieweg, "Chatter on the red: what hazards threat reveals about the social life of microblogged information", *Proc. ACM Conf. Comput. Supported Cooperat. Work*, pp. 241-250, 2010.

[7] X. Wang, M. S. Gerber, D. E. Brown, "Automatic crime prediction using events extracted from twitter posts", *Proc. Int. Conf. Social Comput. Behavioral-Cultural Modeling Predict.*, pp. 231-238, Sep. 2012.

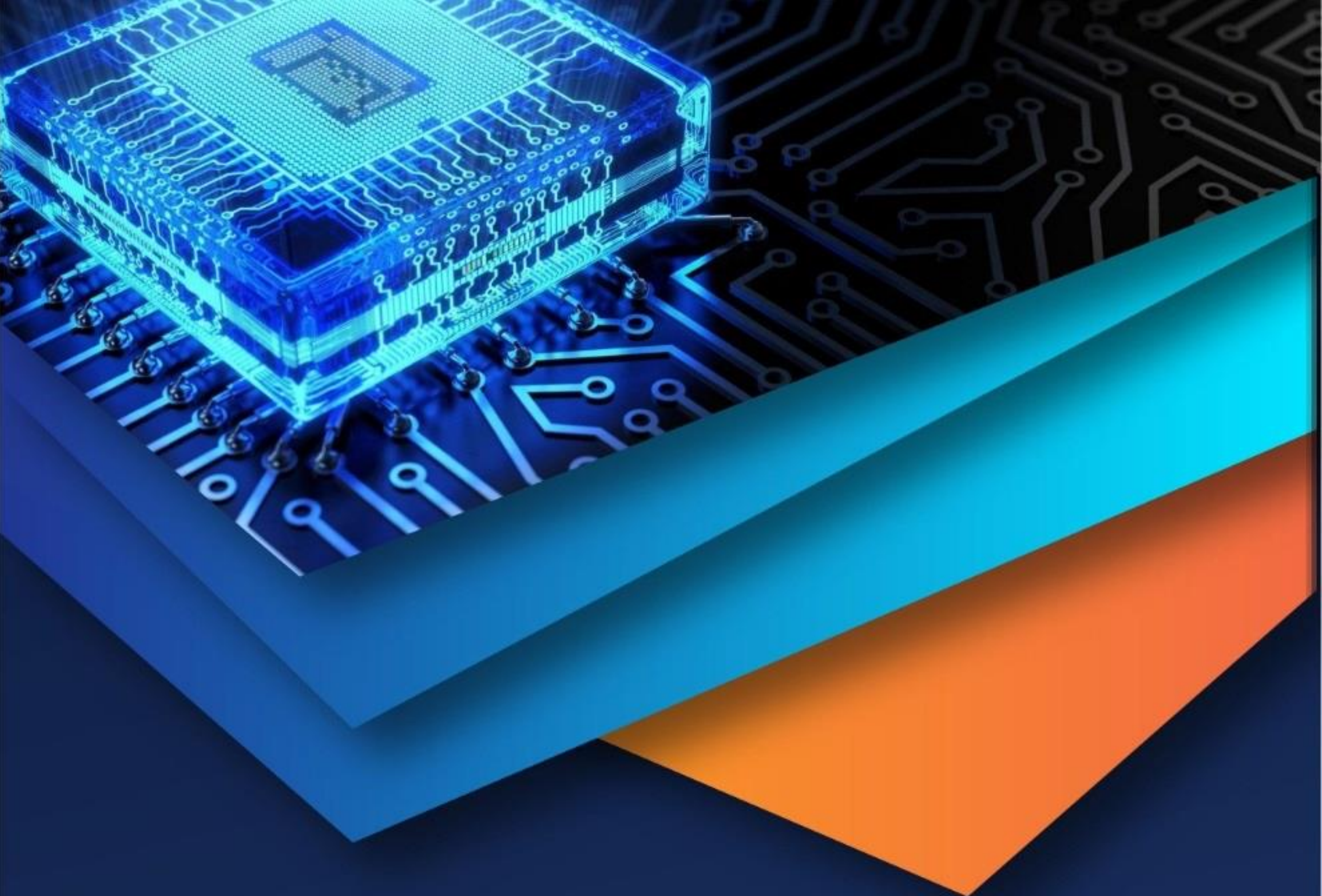
[8] A. Tumasjan, T. O. Sprenger, P. G. Sandner, I. M. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment", Proc. Int. AAAI Conf. Web Social Media (ICWSM), vol. 10, pp. 178-185, 2010.

[9] Kiichi Tago , Qun Jin , "Influence analysis of emotional behaviors and user relationships based on Twitter data." IEEE February 2018.

[10] C. N. Dos Santos and M. Gatti, Deep convolutional neural networks for sentiment analysis of short texts, in Proc. COLING 2014, the 25th Int. Conf. Computational Linguistics: Technical Paper, Dublin, Ireland, 2014, pp. 69–78.

[11] M. Hasan, E. Rundensteiner, and E. Agu, EMOTEX: Detecting emotions in twitter messages, in Proc. 2014 ASE BIGDATA/SOCIALCOM/CYBERSECURITY Conf., Stanford University, CA, USA, 2014.

[12] S. J. Liu, N. Yang, M. Li, and M. Zhou, A recursive recurrent neural network for statistical machine translation, in Proc. 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD, USA, 2014, pp. 1491–1500.



10.22214/IJRASET

45.98

IMPACT FACTOR:
7.129

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY