iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** VIII **Month of publication:** August 2018

DOI:

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 VIII, August 2018- Available at www.ijraset.com

Trend Fusion: Mobile Personalized Recommendation of trends for social networks

Pooja Chaudhari[1], Prof. A. R. Deshpande[2]
[1, 2]Department of Computer Engineering, PICT, Pune

Abstract: Number of users and their information feed is in-creasing day by daqy on social networks. The users communicate through social networks, share and distribute information by sending short text messages in near real-time. As a result of the social networks, the users are often experiencing information overload as they interact with many users and keep on reading contents on large scale. Recommendation systems have been proposed to help users deal with information overload by prioritizing the items of user's interest. The user's preferences are shaped by personal interests. At the same time, users are affected by their surroundings, as determined by their geographically located communities. One of the approach takes into account both personal interests and local communities. These community preferences are generally reflected in the lo-calized trending topics. The proposed dynamic recommendation system provides better customized content described with the most important tweets on social media like Twitter according to his/her individual interests considering the location diffusion. Hence the effect of change in the geographical community preference on individual user's interests is observed through this system which gives top trending tweets based on the same.

Index Terms: Recommendation systems, social networks, topics modeling, trending topics.

I. INTRODUCTION

Social networks having large growth recently in number of users and their shared information. The social networks have different challenges in providing different types of information to the user. Twitter, Face-book is a social networking applications that allow people to brief about a broad range of topics.

Personalized recommendation systems can be a promis-ing solution for the information overload problem in social network sites. Three recommendation problems on social network sites are explored, being recommending people, recommending information, and recommending conversation.

Social networks became an important source for gen-erating recommendations. Using social networks to un-derstand the relations between users and their friends as well as the information obtained about them can improve the knowledge about users' behaviors and ratings. Also, integrating recommendation systems into social networks can provide new observations and thus decisions that can-not be achieved through using traditional recommendation systems. Research studies have also found that different properties of social networks encourage the integration of recommendation systems with social networks. In this paper, the study is varied and address areas such as network structure, trust, information credibility, event detection, social tagging, Geo fencing etc. The recommen-dation systems aim is returning items that are similar to the users' demand.

To provide the user with personalized recommendations for online social network information related to personal and community level using locations of the user .

The system can be show the number of users they can be login in the trend set location. This system can be useful for new friend creation. Also in this system it is easy to identify favourite category for tweets for the purpose of read or write the tweets.

The users' preferences are shaped by personal interests. At the same time, users are affected by their surroundings, as determined by their geographically located communi-ties. Ever since the dawn of civilization, human beings have always been a part of one tribe or another, brought together by their shared interests and a common way to communicate the same.

Capturing user's interest, which change over the time, is important nowadays. So, focusing on suggestions provided by the social media need to be improved. The social networks suggest the recent trends to the users based on their location will reflect positively on their online experience. User can show message that correspond to dynamic interest. Hence, it is important to mine user‘s interest from social network.

Although tweets may contain valuable information, many are not interesting to the users. A large number of tweets can overwhelm users since they interact with many other users and they have to read ever increasing content volume on their timeline. Thus, the difficulty in finding the "matching" users and recommending content that are of interest to users became a key challenge for social networks sites.

Recommendation systems have been proposed to help users cope with information overload by predicting the items that a user may be interested in. Recommender systems have become an important research area since the appearance of the first papers on collaborative filtering in the mid-1990s . There has been much work done both in the industry and academia on developing new approaches to recommender systems over the last decade. The interest in this area still remains high because it constitutes a problem-rich research area and because of the abundance of practical applications that help users to deal with information overload and provide personalized recommendations, content, and services to them. However, despite all of these advances, the current generation of recommender systems still requires further improvements to make recommendation methods more effective and applicable to an even broader range of real-life applications. A method is proposed to identify tweets that may be of potential interest to the user. Since the user's interests in different topics change over time, we focus on studying this change, and recommending to each user the most interesting tweets on the user's timeline at specific time.

The users' preferences are shaped by personal interests. At the same time, users are affected by their surroundings, as determined by their geographically located communities.

Ever since the dawn of civilization, human beings have always been a part of one tribe or another, brought together by their shared interests and a common way to communicate the same.

Capturing user's interest, which change over the time, is important nowadays. So, focusing on suggestions provided by the social media need to be improved. The social media can suggest the trending topics to the users based on their location will reflect positively on their online experience. User can show message that correspond to dynamic interest. Hence, it is important to mine user‘s interest from social network.

II. REVIEW OF LITERATURE

Study of information propagation in social networks along with recent trends is important to be considered on the basis of localities.

A. Recommendation Systems

The personalized recommendation systems recommend through a ranking procedure, useful content to the users using collaborative filtering method to generate personal-ized recommendations in Twitter.

In [2] paper, a model is based on different topics and the history of the user activities in each topic.The limitation is rigid recommendation methods are proposed.

B. Topic Modelling

Topic Modelling is important research field in the area of text mining and statistical modelling. Topic models are used to solve the problem of "information overload" in text and corpuses. for example latent Dirichlet allocation (LDA),used successfully on articles and documents.

In [3] paper, a numerous content-based, collaborative, knowledge and data engineering, and hybrid methods were proposed. Capabilities are improved and recommender systems are made applicable to wide range of applications. Limitation is Utilization of multi-criteria ratings is not used.

C. Information and Influence Propagation in Social Net-works

The information propagation consists of analysis of the messages propagation and the decrease time since the posting of the message. The second is the level of interactions on geographic, demographic levels and their contextual features.

In [1] paper, domain-specific features are studied with hemp of the the user's profile and text. Predefined set of generic classes such as News, Events, Opinions, Deals, and Private Messages are made more correctly. The limitation of this paper is less intrusive recommendation process.

D. Trends in Social Networks

Trends (words and phrases) appearing on the time line of social networking site are identified per hour, day, and week.The attempt to analyze the relation between trends and locations is not satisfactory.

In [4] paper, tweets and networks of their social graphs for Twitter is shown. Advantage is to demonstrate the potential for effective and efficient recommendation. Lim-itation of that noisy content are not properly removed.

In [5] paper, chronological order is followed for tweets and users view the followers' time lines to find their interests. A collaborative ranking to capture personal interests is proposed and is used to decrease the user efforts. Contextual information can be clubbed with this to find out the interests based on topics, social relations and contextual features.

Collaborative personalized tweet recommendation[2]:

Twitter has rapidly grown to a popular social network in recent years and provides a large number of real-time messages for users. Tweets are presented in chronological order and users scan the followees' timelines to find what they are interested in. However, an information overload problem has troubled many users, especially those with many followees and thousands of tweets arriving every day. In this paper, we focus on recommending useful tweets that users are really interested in personally to reduce the users' effort to find useful information. Many kinds of information on Twitter are available for helping recommendation, including the user's own tweet history, retweet history and social relations between users.

Experiments on recommending content from information streams[6]:

More and more web users keep up with newest information through information streams such as the popular microblogging website Twitter. In this paper we studied content recommendation on Twitter to better direct user attention. In a modular approach, one explored three separate dimensions in designing such a recommender: content sources, topic interest models for users, and social voting.

Analysing user modeling on Twitter for personalized news recommendations[7]:

How can micro-blogging activities on Twitter be lever-aged for user modeling and personalization? In this paper we investigate this question and introduce a framework for user modeling on Twitter which enriches the semantics of Twitter messages (tweets) and identifies topics and entities (e.g. persons, events, products) mentioned in tweets.

Following are some advantages of different methods used:

E. Media Connectivity

Easier or faster way to connect is media like Face book,
Twitter, LinkedIn etc. and interests are shared in different contexts.
This helps to:

1) Get new opportunities.
2) Locate assistance
3) Get product and service referrals
4) Receive support from similar posting users
5) Share advice and information
7. Access news or posts

F. Similarity of Interest

One can share , access and get idea about similar interests and hence can keep updating the overall information.

G. Information Sharing

Social media provides way of communications and better improvements in the information world which keep on updating the contents and users worldwide.

The problem can be stated as to develop system 'Trend Fusion', to enhance user's interaction and experi-ence in social networks through a dynamic personalized recommendation system that provides with the most im-portant news, events and actions based on location, using GPS like sensors. It analyses and predicts the localized diffusion of trends in social networks, and recommends the most interesting trends to the user.

Objectives are to enhance user's interaction and ex-perience in social network to the user through a dy-namic mobile personalized recommendation system. Also to analyse and predict the localized diffusion of news, events and actions in social networks, and recommend the most interesting trends to the user. And to update user by sensing the contexts (location and weather), for finding nearest events and actions taken by other users.

The scope of the proposed system is restricted to the geographical area within district(50km) to get the notifications of nearby events and actions (activity and status) of around 1000 other users, based on current location. User will be able to see top 10-15 recommended news on the timeline that are important based on previous inputs and interests.

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 VIII, August 2018- Available at www.ijraset.com

III. SYSTEM ARCHITECTURE

The system consists of 6 basic blocks as shown in fig 1. In system one can login with twitter. Once user logins the twitter account, the recent tweets can be shown. In this system user can write post or tweet. User can use Dynamic subject creation or LDA method for identify the favourite category based on user post. And also fencing nearby Geo location. User can do categorization of post according to different categories e.g. Social, Sport, Entertainment, NEWS etc.
These are shown in fig. below

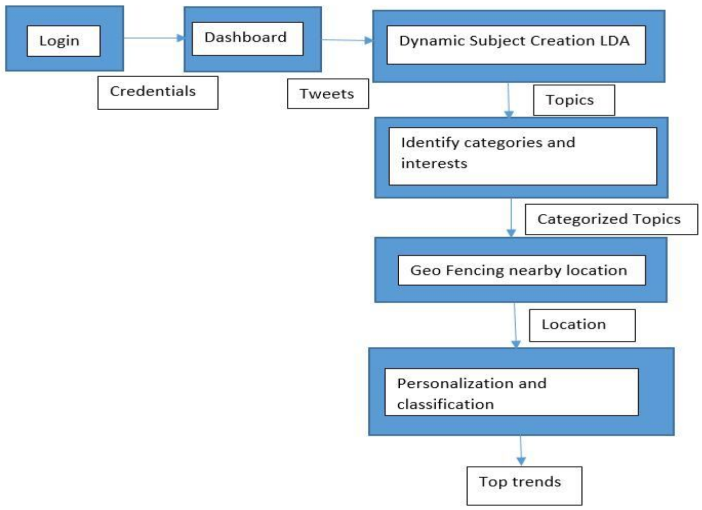


Fig. 1. Proposed System Architecture

A. *Latent Dirichlet Allocation (LDA)*

Latent Dirichlet allocation (LDA) is used widely for different micro blogs. Model involves a two-levels: LDA involves three levels.
T= fT 1; T 2::::::::T ng T- set of posts L= fl1; l2; ::::::::lng
L- set of topics Li= fwi1; wi2; ::::::::wing Li- set of probability values of words n for topic Li. Activity in each day d for a topic is denoted by Ad = fad1; :::; adkg, where ad[i]is the level of activity in topic li for post Oe of the user for a day. Ad[i]=Sum(Oe[i]) A new tweet Oenew and the user's LoI in the tweet can be calculated as = sum(Oenew[i].sum(Ad[i])) The cosine similarity over time for the user is given by: Ut.Pt/(MAG(Ut).MAG(Pt)) where, Ut is user vector and Pt is post vector. It is specified between vectors formed by the summation of the LoI in a topic and value of 1 means the exact distribution match, 0 means users have nothing in common.

B. *Dynamic Subject Creation:*

Important tweets are classified based on this. The model captures the user's interests over changing time and the messages are shown based on dynamic interests. Tis is used to analysis the tweet from the raw tweets and give the important tweets as per the user's requirement.

C. *System Stages Can Be Categorised As*

Group of trends from different locations: Trends should be identified from all the locations of user's interest. The trends are collected for each unit time.

1) Storing trending topics: Trending topics are labelled by the location/time and they were received from/at.
2) Building cascades: It is specified weather a trend is a beginning of a new cascade or a continuation of an old one.Hence it is important to categories those.
3) Extraction of Parameters: The parameters are measured for each location based on the diffusion models. There are mainly 4 types:

a) Diffusion Parameters
b) Geographical Parameters
c) Historical Parameters
d) Trend Parameters

4) *Model Learning:* Based on previous results and parameters set up model should be able to learn and give correct output based on different inputs and featured outputs.

Mathematical Model of a system can be stated as –

Let = S'be the twitter set, S = fT; F g, T =fT 1; T 2; T 3; T 4jT given tweets from userg, F= fF 1; F 2; F 3jF given feedback on newsg

Functions: F: F = fF 1(); F 2(); F 3(); F 4(); F 5(); F 6(); F 7()g, where,

F1 (S) = login with twitter (Once)

F2 (S) = Display recent tweets

F3 (S) = Dynamic subject creation

F4 (S) = Identify the favourite category based on user post

F5 (S) = Categorization of post according to different categories

F6 (S) = Get news feedback using geo fencing nearby location

D. *Algorithm Building Cascades*

Procedure: Build Cascades From Activations

Input: Activations List al begin

An activation a is a record a = (trend, location, time) Activations List alo = Order al by time for all Activation a in alo do

If a.trend appeared in (a.time - 24 hours) then

1) cas = last cascade of a .trend
2) If a.location appeared in cas then

Add a. time to instances of a. location in cas

3) Else if a.time equals time of last step in cas then Add a to last step of cas
4) else

Add new step to cas containing a. location end if

5) else

Create new cascade cas Add new step to cas containing a.location end if end for end

The record with trend, location and time is need to be categorized as new or old cascade based on its time to fetch accurate recent trend as per the user's interest.

Hence classification of trends is done using above algorithm.

Methodology LDA is prefered over Simple Dirichlet-multinomial clustering model. A classical clustering model involves two-levels -Dirichletis sampled once for a corpus. A multinomial clustering variable is selected once for each document in the corpus, and a set of words are selected for the document conditional on the cluster variable. A model restricts a document to being associated with a single topic.LDA involves three levels, and the topic node is sampled repeatedly within the document.

E. *Work Flow of system*

Flow of the system is stated in Figure 2:

1) Number of keywords: User will twits on twitter. System will search keywords. Based on keywords recommendation system is proposed.
2) Topic modeling: The unsupervised nature of topic modeling methods makes

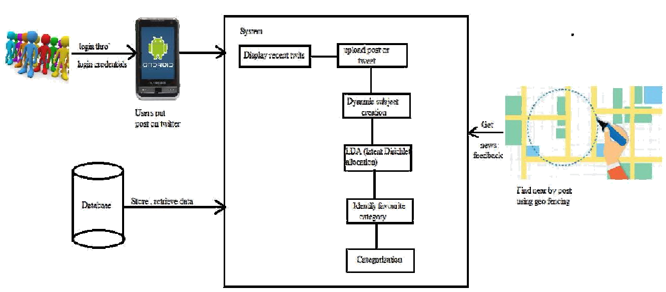


Fig. 2. System Flow

choosing one topic model over another a challenging task. Topic model quality tends to be evaluated by performance in a specific application. Topic models can be evaluated based on perplexity as a quantitative method. Perplexity is a well-known standard metric used in information retrieval field. It tries to quantify the accuracy of a model by measuring how well the trained model deals with an unobserved test data. Perplexity is defined as a reciprocal geometric mean of per word likelihood of a test corpus. A lower perplexity indicates a better generalization performance.

3) *Trained dataset:* Trained dataset is to be prepared. Twits related to keywords are stored in our dataset.
4) *Geo fencing:* Location of users are tracked using this functionality. Nearby users who tweet post are detected. This is a feature in a software program that uses the global positioning system (GPS) to define geographical boundaries.

Sequence is shown starting from user login to display of top trends on time line in Figure 3:

F. Experimental Setup

Software requirements:

XML (Extended Markup Language) for designing layouts screen of android application.

Java language for connecting xml design part with database and to perform basic operation of database. Google map API key| to use Google map services. Operating System: Windows 7 or above Technology: Java, J2EE ,Android

Web Server: Xammp server

Database: My SQL

Java Version: J2SDK 1.7 / 1.8, Android SDK 19 and above

Hardware requirements:

Windows PC/Android Mobile (above version 2.3)

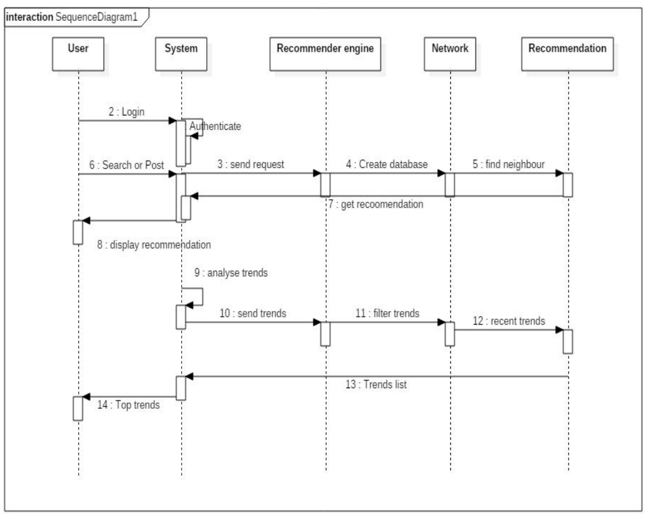


Fig. 3. Sequence Diagram

Functional Requirements

1) *Download The Mobile Application: A* user should be able to download the mobile application through any store or service and it should be free of cost. Download and notify users of new releases User should get notification of a new version and updates should be made easily within the application.
2) *User registration-Mobile application*: After download user should be able to register to the application and should easily recover the credentials if required.
3) *Mobile application- Search timeline*: User should be able to get the recent news after refreshing his/her timeline after login. Location, interests, tags should get considered while showing the trends to the user on the timeline.
4) *Recommended list to user:* User should get top 10 trends as a list on his/ her timeline which are relevant to user based on the specified constraints.
5) *Navigation though mobile application:* A user should be able to navigate through the application. User should be able to post, search, update and delete the feeds.

IV. SYSTEM ANALYSIS

After user twits on twitter, system will search keywords.
Based on keywords recommendation system is proposed.
Topic models are evaluated based on perplexity. Per-plexity gives the accuracy of a model on measuring unobserved test data.A lower perplexity indicates a better generalization performance. It is given by:

$$\text{Perplexity}(D_{\text{test}}|M) = e^{\frac{-\sum_{d \in D_{\text{test}}} \log P(w_d|M)}{\sum_{d \in D_{\text{test}}} N_d}}$$

Where, wd is the words in document d, M is the topic model, and Nd is the number of words in document d. Trained dataset is to be prepared. Twits related to keywords are stored in our dataset. Location of users are tracked using this functionality. Nearby users who tweet post are detected.
A new information cascade model, Snowball Cascade (SC) model is compared with General Threshold model (GT) with the help of recall and precision values based on three classifiers-logistic regression (LR), stochastic gradient descent (SGD) and random forest (RF).
The results shows that SC model performs better than GT model as shown in the below table, Fig 3.

	Snowball Cascade Model			General Threshold Model		
Classifiers	**Logistic Regression**	**Stochastic Gradient Descent**	**Random Forest**	**Logistic Regression**	**Stochastic Gradient Descent**	**Random Forest**
Average Recall	0.95	0.98	1	0.5	0.5	0.5
Average Precision	0.79	0.8	0.7	0.68	0.68	0.5
	Without Topics			Distance Only		
Average Recall	0.75	0.73	0.84	0.82	0.86	0.98
Average Precision	0.6	0.7	0.65	0.7	0.68	0.99

Fig. 4. Comparison of models

Implementation details:

A. Dataset

Twitter API
400k trends
Tuple form: (woeid, trend0, trend1, . . . , trend9, and date/time),
where woeid is Yahoo Where On Earth ID [50] trend0, . . . , trend9 are the top ten trends.

B. Results

After login user is able to see the options on dashboard. User can view the tweets uploaded by him or her on time line.
User can enter the topic of his or her choice and the location so that nearby tweets can be fetched.
Based on the fetched tweets classification is done according to the relevance of users's personal interests. Positive and negative tweets can be shown using graph alon with the location on the map.
Based on the location and the topic nearby and relevant tweets are fetched and shown to user as per Figure 5.

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 VIII, August 2018- Available at www.ijraset.com

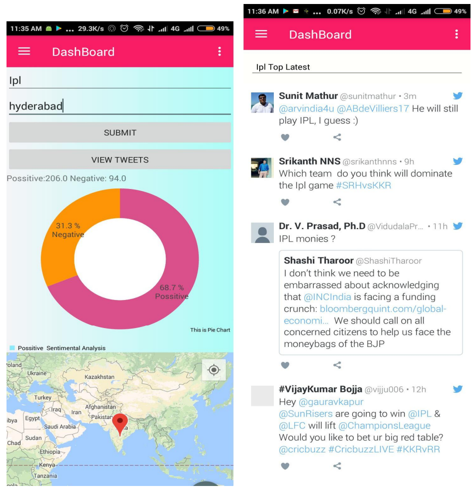


Fig. 5. Results

C. Performance evaluation parameters Top 10 trends

Topic	IPL	Dance	Paint	Food
Average precision	0.79	0.85	0.65	0.55
Average Recall	0.95	0.87	0.75	0.65
F measure	0.86	0.86	0.69	0.56
Perplexity	20	35	60	55
Location	Pune	Mum-bai	Kolkata	Ban-glore

TABLE I
Performance evaluation parameters(Top 10)

Top 20 trends:

Topic	IPL	Dance	Paint	Food
Average precision	0.72	0.65	0.55	0.50
Average Recall	0.85	0.76	0.67	0.56
F measure	0.78	0.70	0.60	0.53
Perplexity	17	30	55	50
Location	Pune	Mum-bai	Kolkata	Ban-glore

TABLE II
Performance evaluation parameters(Top 20)

In this way for different locations and topics based on their relevance with user, we get varied recall and precision values along with their perplexities.

Precision is the percentage true positives in the retrieved results.Recall is the percentage true negatives in the retrieved results.

Using the two values F measure is defined which is the harmonic mean of the two as:

= 2 precision recall precision + recall

Lower perplexity denotes more accuracy and more value of recall boosts the performance of the application.

V. CONCLUSION

The proposed solution categorizes trends using the author information and features within the trends that provide better way of recommendation. Users hence are able to filter feeds based on their interests and priorities. Dynamic LoI can be beneficial for users to recognize important feeds even with their short forms and in future if particular trend appearing in some city in any context along with its time. SC model used is more beneficial which is continuous and provide better results than other models.

VI. FUTURE WORK

In research attempt, I have focused on the localized diffusion of trends. However,the database is growing tremendously via use of the Internet. Thus,future work would be to extend the perplexity and improve effectiveness of trends suggestions based on influential locations of users.

REFERENCES

[1] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, “Short text classification in Twitter to improve in-formation filtering,” in Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., New York, NY, USA, 2010, pp. 841–842.

[2] Twitter Home Page, accessed on Jun. 23, 2017. [Online]. Avail-able: http://twitter.com

[3]] D. Laniado and P. Mika, “Making sense of Twitter,” in Proc. 9th Int. Semantic Web Conf. Semantic Web, vol. 1. Berlin, Germany, 2010, pp. 470–485.

[4] R. Mehrotra, S. Sanner, W. Buntine, and L. Xie, “Improving LDA topic models for micro-blogs via tweet pooling and automatic labeling,” in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., New York, NY, USA, 2013, pp. 889–892.

[5] Z. Ren, S. Liang, E. Meij, and M. de Rijke, “Personalized time aware tweets summarization,” in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., New York, NY, USA, 2013, pp. 513–522.

[6] Altingovde, I.S., Demir, E., Can, F., and Ulusoy, O. Site-based dynamic pruning for query processing in search engines. In Proc. SIGIR (Singapore, July 2008), 861-862.

[7] Banerjee, S., Ramanthan, K., and Gupta, A. Clustering short text using Wikipedia. In Proc. SIGIR (Amsterdam, the Nether-lands, July 2007), 787-788.

[8] H. Baars and H.-G. Kemper, “Management support with struc-tured and unstructured data—An integrated business intelligence framework,” Inf. Syst. Manage., vol. 25, no. 2, pp. 132–148, Mar. 2008.

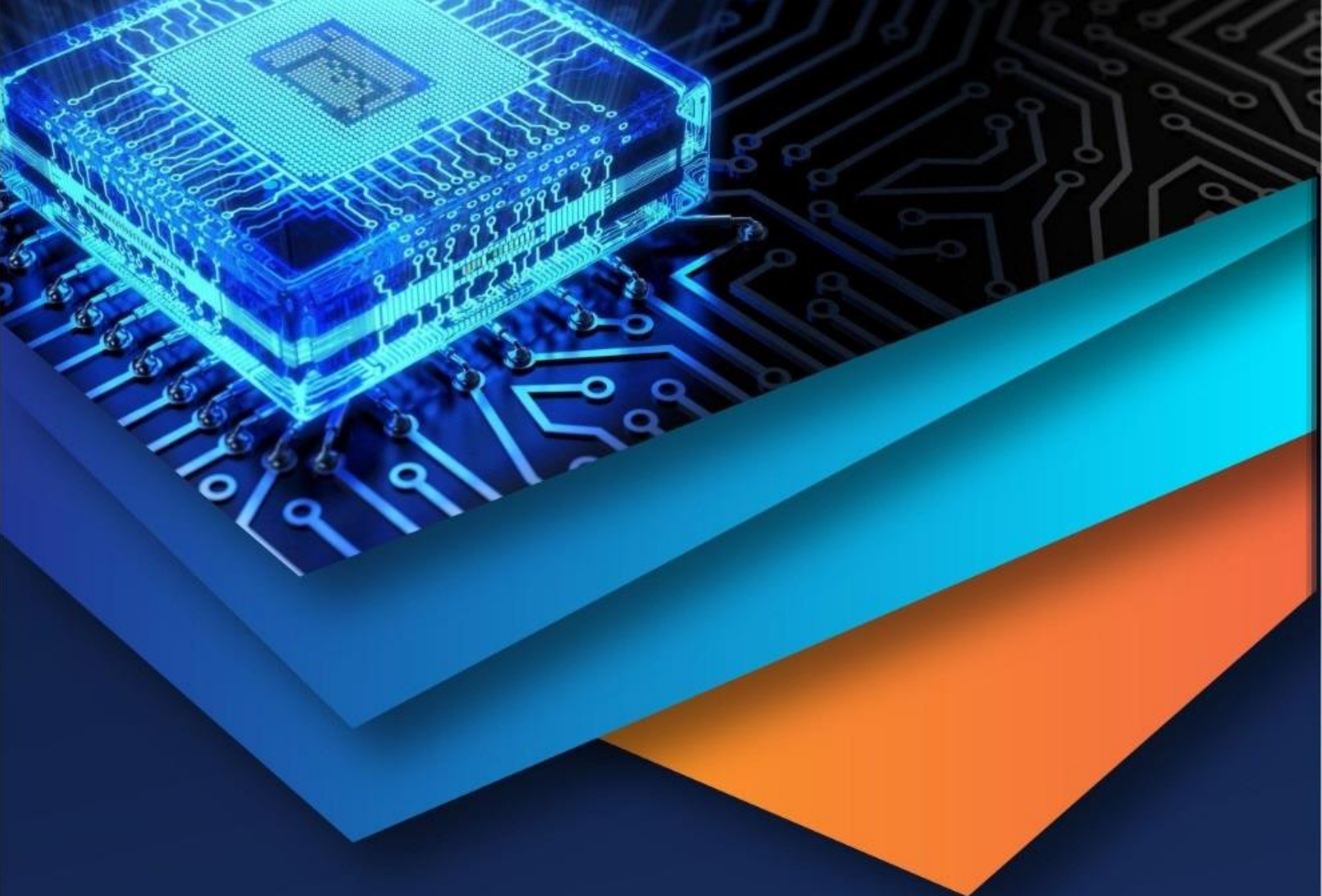
[9] C.L.Wang and Y.S.Chiou,“An adaptive Positioning Scheme Based on Radio Propagation Modeling for Indoor WLANs,” 63rd IEEE Vehicular Technolgy Conference, Melbourme Australia, May 2006: 2676-2680.

[10] J. Martinez-Romo and L. Araujo, “Detecting malicious tweets in trending topics using a statistical analysis of language,” Expert Syst. Appl., vol. 40, no. 8, pp. 2992–3000, 2013.

[11] J. Weng, E.-P. Lim, J. Jiang, and Q. He, “TwitterRank: Finding topicsensitive influential twitterers,” in Proc. 3rd ACM Int. Conf. Web Search Data Mining, New York, NY, USA, 2010, pp. 261–270.

[12] M. Steyvers, P. Smyth, M. Rosen-Zvi, and T. Griffiths, “Prob-abilistic author-topic models for information discovery,” in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, New York, NY, USA, 2004, pp. 306–315.

[13] W. X. Zhao et al., “Comparing Twitter and traditional media using topic models,” in Proc. Eur. Conf. Inf. Retr. (ECIR), Berlin, Germany, 2011, pp. 338–349.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429