IJIRCCE

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

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906268 |

Spam Posting Twitter Account Detection and Mitigation

Mr. Dayanand Argade, Prof. Hydar Ali Hingoliwala

PG Student, Department of Computer Engineering, JSCOE Hadapsar, Pune, Maharashtra, India

Assistant Professor, Department of Computer Engineering, JSCOE Hadapsar, Pune, Maharashtra, India

ABSTRACT: Twitter is most popular micro blogging services administrations, or, generally which is used to share news and updates through short messages limited to 280 characters. In any case, it's open nature and huge client base are as often as possible misused by automated spammers, content polluters, ill-intended users to commit various cyber-crimes, for example, cyber bullying, trolling, talk scattering, and stalking. As per necessity, a number of methodologies have been proposed by researchers to address these issues. Be that as it may, the vast majority of these methodologies are based on user characterization and totally ignoring mutual interactions. A cross breed approach for identifying automated parameters by amalgamating community based highlights with other component classes, to be specific metadata-, content-, and association based highlights. Nineteen different features, including six recently characterized features and two redefined features, are distinguished for learning three classifiers, to be specific, random forest, decision tree, and Bayesian network, on a real dataset that includes benign clients and spammers. The discrimination power of various element classifications is also analyzed, and interaction and community-based features most effective for spam detection, whereas metadata-based features are proven to be the least effective.

KEYWORDS: Aspect-based Opinion Mining, Maximum Entropy Model, Word Embedding

I. INTRODUCTION

Online social media is one of the defining phenomena in this technology-driven era. Platforms, such as Face book and Twitter, are instrumental in enabling global connectivity. 2.46 billion Users are estimated to be now connected and by the year 2020 one third of the global population will be connected. Users of these platforms freely generate and consume information leading to unprecedented amounts of data. Several domains have already recognized the crucial role of social media analysis in improving productivity and gaining competitive advantage. Information derived from social media has been utilized in health-care to support effective service delivery, in sport to engage with fans, in the entertainment industry to complement intuition and experience in business decisions and in politics to track election processes, promote wider engagement with supporters and predict poll outcomes. However, alongside the benefits, the rapid increase in social media spam contents questions the credibility of research based on analyzing this data. A report by Nexgate estimates that on average one spam post occurs in every 200 social media posts and a more recent study reports that approximately 15% of active Twitter users are automated bots. The growing volume of spam posts and the use of autonomous accounts (social bots) to generate posts raise many concerns about the credibility and representativeness of the data for research.

II. RELATED WORK

Spam entails any form of activity that causes harm or disrupts other online users. The increasing amount of spam tweets can be attributed to humans' inclination to spread misleading information, even if such information originated from unreliable sources, such as a social bot account. Recently, Vosoughi et al. [1] discover that both genuine and false news spread at equal rate. False news on Twitter spread rapidly. Social bots are deployed to accelerate the process and human users further amplify the content. To detect spam tweets, numerous detection systems have been proposed, using various techniques that are reviewed in this section. Thomas et al. [2] and Lee and Kim [3] analysed streams of URLs used by spam users and studied how spammers exploit URLs obfuscation to redirect users to malicious sites.

Grier et al. [4] analysed a large number of distinct URLs pointing to blacklisted sites due to their involvement in scam, phishing and malware activities. Although the approach is effective, it is often slow and fails to detect URLs that point to malicious sites but have not been blacklisted previously.

Gao et al. [5] also studied URL usage on Facebook to detect spamming activity and observed that this form of spamming is mostly associated with compromised accounts rather than accounts created solely for spam activity.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906268 |

Benevenuto et al. [6] studied the statistical properties of user accounts and how URL shortening services affect spam detection mechanisms. However, the universal use of URLs and URL shortening by the vast majority of Twitter users makes it difficult to directly identify potentially nefarious links on a large scale. In general, the use of URLs relies on historical information, limiting the possibilities for real-time detection. Danezis and Mittal [7] utilised a social network model to infer legitimate user accounts that are being controlled by an adversary. Lee et al. [8] created social honeypot accounts mimicking naive Twitter users to entice spam posting users. Users who fall prey by engaging with these accounts are assumed to be in violation of usage policy. Users identified using this method were analysed to distinguish different user types focusing on link payloads and features that can capture the dynamics of follower-following networks of users. Varol et al. [9] employed many features related to users, content and the network to develop a system for social bot account detection. Chen et al. [10] provides an in depth analysis of deceptive words used by spammers on Twitter. The work of Chen et al. is motivated by Twitter Spam Drift, i.e. the property of statistical features of spam tweets to change over time. Twitter Spam Drift is caused because spammers continuously adopt and abolish various evasive tricks. Features related to this phenomenon were utilised in training machine learning classifiers.

Li and Liu [11] analysed how the effect of unbalance datasets can be mitigated in detection tasks. Standard machine learning methods are sometimes considered as inadequate in capturing the variability of spamming behaviour. Wu et al. [12] utilised a deep learning technique based on Word2Vec [31] to capture the variation of spam-related challenges. While it is essential to allow detection models to continuously learn features strong enough to distinguish spam from nonspam, methods that solely rely on textual information are be inadequate to draw the distinction between a habitual spam posting account and a non-spam posting account. Hand-crafted features related to the account and the user need to be considered. In this study, a set of hand-crafted features are leveraged in tandem with features learn by deep neural networks. Features studied by humans and encoded to classifiers can achieve better performance and low false positive rates [13]. The use of a large number of features introduces extra overheads to the detection system, some of which may be unavailable for real-time use. Subrahmanian et al. [14] offer insights into techniques utilised in identifying influence bots, i.e. autonomous entities determined to influence discussions on Twitter. Influence bots comprise a category of social bot accounts that seek to assert influence on topical or new discussions thereby generating unrepresentative or fake data. The surveyed studies on spam detection largely rely on either historical tweets of a user to extract features which contribute to an extra overhead for the detection system [15] or limited features learnt by unsupervised techniques. Our proposed approach relies on readily available features in real-time for better performance and wider applicability. The Semantic [16] Web has developed tremendously with increasingly more data on the web being available as the Resource Descriptor Framework (RDF). The point of this paper is to fill in as guide for future innovative work to advance connected open information which can be distributed as information solid shapes on the Semantic Web. The paper [17] talked about how Semantic Web innovations advanced from the customary Extract, Transform and Load (ETL) based to the more programmed mapping of multidimensional information. This paper gives a forward outline of examines that mean to improve the proficiency of SW advances. In this paper [18], web crawlers: their engineering, procedure of semantic centered creeping innovation, cosmology learning, design coordinating, types and different difficulties being confronted when web indexes utilize the web crawlers, have been explored. Framework adequately demonstrates the better exactness for the extraction of intriguing examples shape the website pages. The proposed research [19] work is done on streamlining the execution of the Search Engine. Research investigates dynamic crept dataset by figuring the cosine comparability and Shannon data gain edge esteem. They proposed [20]a bunch based information spread methodology which is valuable for accomplishing most extreme QoS utilizing CSMA and TDMA. At long last, they presumed that many steering conventions are intended to decrease vitality utilization and End-to-End delay.

III. PROPOSED SYSTEM

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906268 |

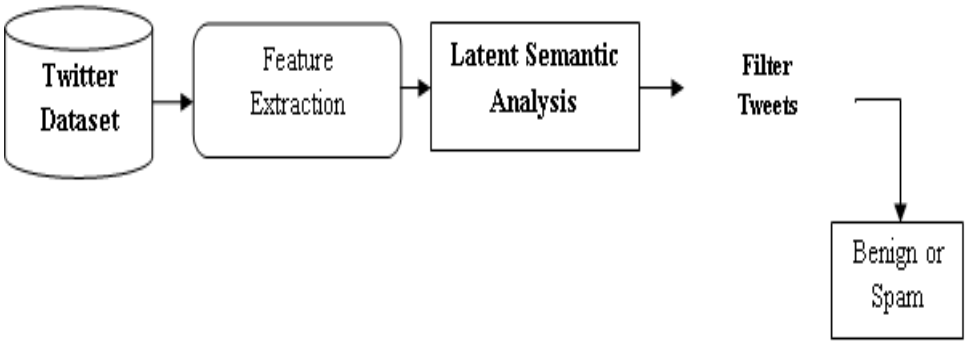

Fig 1: Proposed System

III. MODULES

1. Load Twitter Dataset
2. Feature Extraction
3. Latent Semantic Analysis

1) Load Twitter Dataset:
In this module, a user load twitter dataset for detect automatic spammer posting account.
This dataset contains users posting and his account details.
This post has any contents.

2) Feature Extraction:
This module has 4 features.
1. Community-based
2. Metadata-based
3. Content-based
4. Interaction-based.

These 4 features detecting automated spammers in Twitter.

3) Latent Semantic Analysis:
This module takes input dataset as Feature extraction module spam detection results.
This module once again detects spam using latent semantic analysis.
It provides accurate results.
Using spam posting, this module identifies spam posting user account.

IV. ALGORITHM

Algorithm 1: Spam Posting Account Detection:
Input: Twitter Dataset (TD)
Output: Detect Spam Account
Step 1: Extract Tweet T from TD
Step 2: Extract Feature from T
Step 3: Apply Latent Semantic Analysis in each T.
Step 4: Classify Tweet into spam or not using LSA Algorithm
Step 5: Find Spam Account.
Step 6: End

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906268 |

V. IMPLEMENTATION RESULTS

	No	Text	Target	Passingjud	religion	abusive	Comparison
0	1	leaders anilingus arsehole squandered the money and abandoned the parthic hospital which iz just now a toilet the community shame	1	no	no	yes	no
1	2	rt why do million us residents in their trillion wealthy economy hoarding	1	no	no	no	no
2	3	baloch student are being pushed away from the education system under a well known policy so to divert their attent	1	no	no	no	no
3	4	rt this is already happening under the rule of rss shame farmersprotests delhi	1	no	no	no	no
4	5	inspite of clear words of calling dint recieve any call at pm anilingus arsehole	1	no	no	yes	yes
5	6	shameshameshame jazzcash anilingus arsehole	1	no	no	yes	no
6	7	rt the bedrock of fascism is misogyny the vile sanghis going after are now making memes of her abuse and celebrati	1	yes	no	no	no
7	8	rt never expected this from you i still dont beleive we gave you so much love shame	1	yes	no	no	no
8	9	rt swamivivekananda bengal bjp insultsltsldolsshame onnadda bengal rejects bjp	1	no	no	no	no
9	10	rt with several internal conflicts going on pakistan focusing on its neighbor just proves its lack governance and its immaturit	1	no	no	no	no
10	11	ngl the new coming to america looks shit shame anilingus arsehole	1	no	no	yes	no
11	12	where is the science toolkit which was deleted by you have any answers to it its democracy if it	1	no	no	no	no
12	13	rt lucknow lucknow ietlucknow if the interview is being conducted virtuall	1	no	no	no	no
13	14	in the issue is not about assisting issue is your executive told me specifically talking in marathi is not	1	yes	no	no	no
14	15	rt in the issue is not about assisting issue is your executive told me specifically talking in marathi is not allowed in	1	yes	no	no	no
15	16	rt in i applied for credit card your executive told me speaking in marathi is not allowed in hsbc with due respect i don	1	yes	no	no	no
16	17	you mf let me now a single line you understand of this person i started watchi	1	no	no	no	no
17	18	scoring runs is batsmen responsibility root scores and our captain vice captain fucking legend	1	yes	no	yes	no
18	19	i cannot understand when we had placed an replacement on what basis did the seller can	1	no	no	no	no
19	20	geopolitics chinaeverywhere olympics minorityliberalgovernment uighurs a gov t that claims to support	1	no	no	no	no
20	21	and this man calls himself the mentor of the so called alliance shame sharadpawar indiawithmodi indiawithsachin	1	no	no	no	no
21	22	rt and this just breaks our heart so ashamed of you hope you can sleep peacefully knowing that you have g	1	no	no	no	no
22	23	rt tweets denormandie for the abolition of bullfig	1	no	no	no	no
23	24	this is our voice for justice give back our mother daw aung san su kyi and our president u win myint shame on you	1	no	no	no	no
24	25	of the ministerial zoning orders issued by the ford government since march of last year have benefited develo	1	no	no	no	no
25	26	rt cbdl has major updates coming out cbd life sciences inc cbdl announces new pain relief roll on	1	yes	no	no	no
26	27	rt bubblebutt mmmm we re liking everything in here how about you switch booty body muscle teammalebubblebutt hot	1	no	no	no	no
27	28	rt cbdl has huge updates coming news worldnews stockstowatch stocks investors stocktrading wa	1	no	no	no	no

Fig 2: Dataset

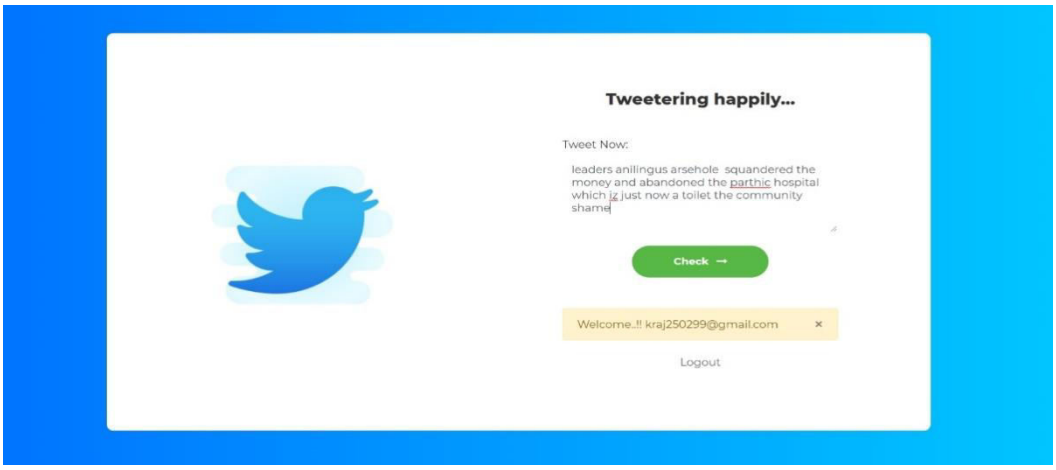

Fig 3: Input Tweet

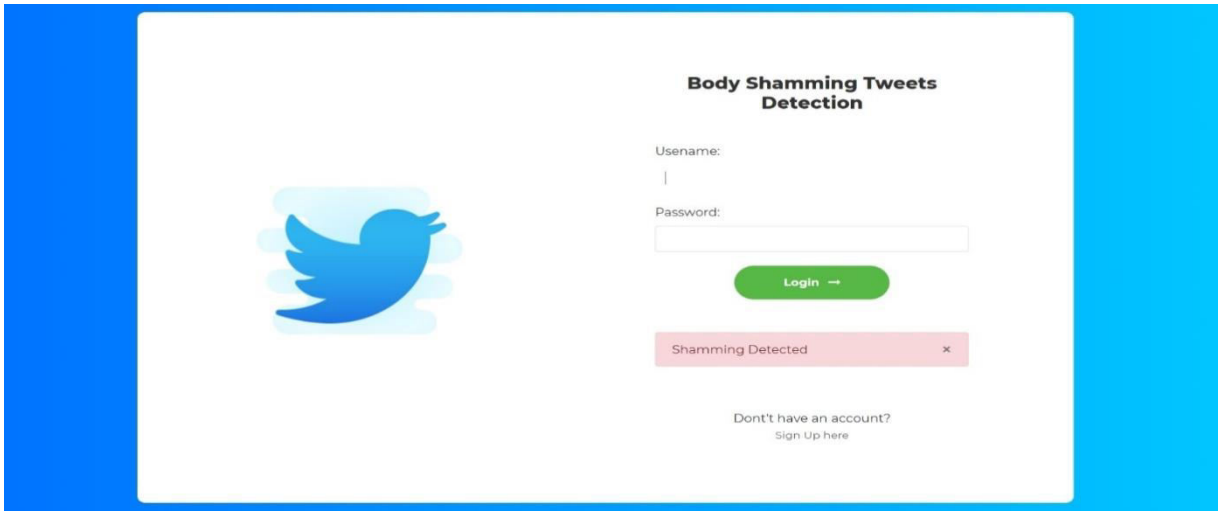

Fig 4: Spam Tweet detection

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906268 |

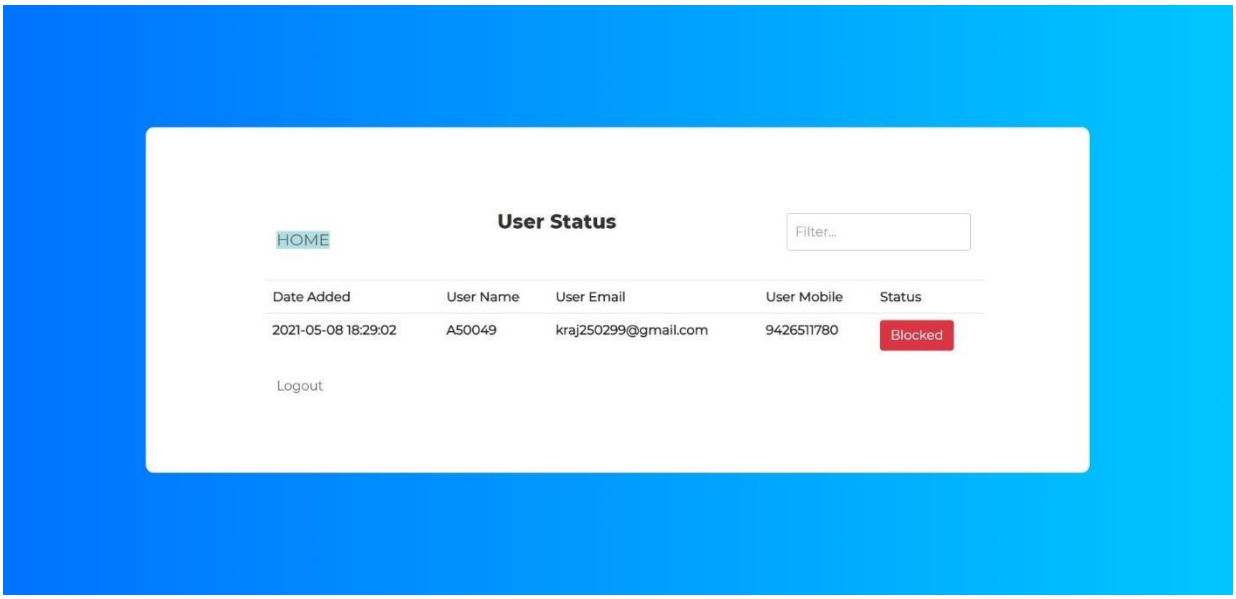

Fig 6: User Blocked

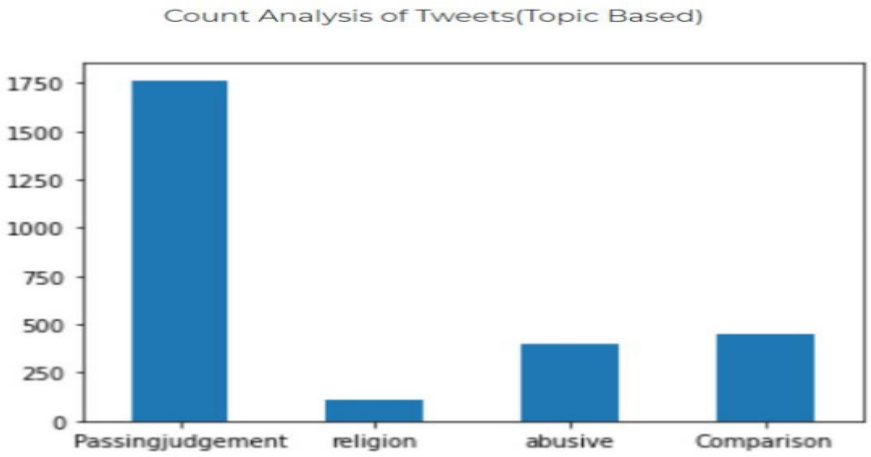

Fig 7 : Data Count Analysis

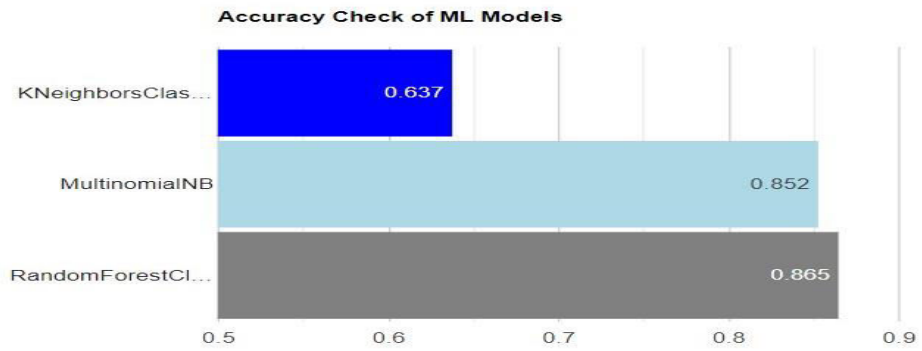

Fig 8: Model Accuracy Analysis

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906268 |

VI. CONCLUSION

This study offers an effective method for spam detection and new insights into the sophisticatedly evolving techniques for spamming on Twitter. The proposed spam detection method utilized an optimized set of readily available features. Being independent of historical tweets which are often unavailable on Twitter makes them suitable for real-time spam detection. The efficacy and robustness of the proposed features set is shown by testing a number of machine learning models and on dataset collected orthogonally from the study data. Performance is consistent across the different models and there is significant improvement over the baseline. It was also shown that automated spam accounts follow a well-defined pattern with surges of intermittent activities. The proposed spam tweet detection approach can be applied in any real-time filtering application. For example, it is applicable to data collection pipelines to filter out irrelevant content at an early pre-processing stage to ensure the quality and representativeness of research data. The combination of handcrafted features and features learnt in an unsupervised manner using word embedding is shown to significantly improve baseline performance and to perform comparably to the best performing feature set using a smaller number of features.

During the analysis of the data, we observed that spam users tend to be selective in following other users thereby forming enclaves of spammers. This is a high-level observation that we aim to explore further in the future. Additionally, both the two broad user groups, i.e. human users and social bot (autonomous entity) users contain spammers, whose spamming behaviour tends to be similar. The distinction between legitimate human users vs. legitimate social bots as well as human spammers vs. social bot spammers needs to be investigated further. Another interesting dimension for future work is to study the effect of the recent increase in the maximum length of tweets on spamming activity. Intuitively, automated spam accounts will face difficulties in generating lengthier tweets intelligently, thereby making these tweets easier to identify.

REFERENCES

[1] Liu, Bing. 2010. “Sentiment Analysis and Subjectivity.” Handbook of Natural Language Processing: 1–38.
[2] Sarvabhotla, Kiran, Prasad Pingali, and Vasudeva Varma. 2011. “Sentiment Classification: A Lexical Similarity Based Approach for Extracting Subjectivity in Documents.” Information Retrieval 14: 337–53.
[3] Turney, Peter D. 2001. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews.” in Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL ‘02 417-424.
[4] Goldberg, Andrew, Andrew Goldberg, Xiaojin Zhu, and Xiaojin Zhu. 2006. “Seeing Stars When There Aren‟t Many Stars: Graph-Based Semi-Supervised Learning for Sentiment Categorization.” In Proceedings of TextGraphs: The First Workshop on Graph Based Methods for Natural Language Processing, 45–52.
[5] Zhang, Zhu, and Balaji Varadarajan. 2006. “Utility Scoring of Product Reviews”. in Proceedings of the 15th ACM International Conference on Information and Knowledge Management, 51–57.
[6] J. Zhu, H. Wang, B. K. Tsou, and M. Zhu, “Multi-aspect opinion polling from textual reviews,” in Proc. 18th ACM Conf. Inf. Knowl. Manage., Hong Kong, 2009, pp. 1799–1802.
[7] A. Mukherjee and B. Liu, “Modeling review comments,” in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Jeju, Korea, Jul. 2012, pp. 320–329.
[8] W. X. Zhao, J. Jiang, H. Yan, and X. Li, “Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid,” in Proc. Conf. Empirical Methods Natural Lang. Process., Cambridge, MA, USA, 2010, pp. 56–65.
[9] I. Titov and R. McDonald, “A joint model of text and aspect ratings for sentiment summarization,” in Proc. 46th Annu. Meeting Assoc. Comput. Linguistics, Columbus, OH, USA, 2008, pp. 308–316.
[10] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai, “Topic sentiment mixture: modelling facets and opinions in weblogs,” in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 171–180.
[11] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” J. ACM, vol. 46, no. 5, pp. 604–632, Sep. 1999.
[12] Q. Zhang, Y. Wu, T. Li, M. Ogihara, J. Johnson, and X. Huang, “Mining product reviews based on shallow dependency parsing,” in Proc. 32nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, Boston, MA, USA, 2009, pp. 726–727.
[13] T. Ma and X. Wan, “Opinion target extraction in chinese news comments.” in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 782–790.
[14] Y. Wu, Q. Zhang, X. Huang, and L. Wu, “Phrase dependency parsing for opinion mining,” in Proc. Conf. Empirical Methods Natural Lang. Process., Singapore, 2009, pp. 1533–1541.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 6, June 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0906268 |

[15] F. Li, C. Han, M. Huang, X. Zhu, Y. Xia, S. Zhang, and H. Yu, "Structure-aware review mining and summarization." in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 653–661.
[16] K. Gupta and P. Lambhate Doshi, "Processing Linked Multidimensional Data On The Semantic Web," International Conference On Computing, Communication And Energy Systems, Jan. 2016.
[17] Poonam P. Doshi, "Feature Extraction Techniques Using Semantic Based crawler for Search Engine," International Conference on Computing, Communication and Energy Systems. Jan 2016.
[18] Poonam P. Doshi, Emmanuel M., "Web Pattern Mining using ECLAT," International Journal of Computer Applications, Volume 179 – No.8, December 2017.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN 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

www.ijircce.com

Scan to save the contact details