International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

AN ANALYTICAL STUDY ON VARIOUS DATA MINING METHODOLOGIES ON SOCIAL MEDIA APPLICATIONS

M. Selvapriya[1], Dr.P. Kalyani[2]

[1]Research Scholar, Student Member, IEEE, Department of Computer Science, S.N.R SONS College (Autonomous), Coimbatore, Tamilnadu, (India)

[2]Assistant Professor, Senior Member, IEEE, Department of Computer Science, S.N.R SONS College (Autonomous), Coimbatore, Tamil Nadu, (India)

ABSTRACT

Nowadays, Social Media has obtained note worthy attention. This is accredited to the affordability of gaining access to the social network websites for instance Facebook, Google+, Twitter and other social network sites via the internet as well as the web 2.0 technologies. Numerous people turn out to be attentive in and depending on the social media for info and view of other users on various subject matters. Social Media has gained reputation as it is an inexpensive as well as quicker communication provider. In addition, social media has let us to decrease the spaces of physical distance, it as well produces as well as conserves vast volume of data. The data are extremely beneficial and it provides association degree amid people and their views. In this paper, the complete examination of the techniques that are utilized on user behavior identification is presented. This evaluation would give a thorough info, advantages and disadvantages in the area of sentiment as well as opinion mining. This study presented experimental assessment of numerous advanced data mining techniques that is stated in the literature. The experimental outcomes are carried out and it is matched up in contradiction of one another in order to identify the superior method underneath numerous performance measures such asF1-score and classification accuracy.

Keywords: social media, data mining, opinion mining, user behaviour prediction.

I.INTRODUCTION

Presently, Social networks have turn out to be extremely widespread owing to the raising proliferation as well as affordability of internet empowered devices for instance personal computers, mobile devices and other more current hardware innovations for instance internet tablets. Usually, a social network is a network of relationships or interactions, here the nodes encompasses actors, and the edges comprise the relationships or interactions amid these actors. A generality of the notion of social networks is that of info networks, wherein the nodes encompass actors or entities, and the edges represent the associations amid them. Social media sites for instance Facebook, Twitter and YouTube gives superior places for students to share happiness as well as struggle vent sentiment and pressure and search for social support. Present researches and surveys have exposed a need to unremittingly gather, observe, examine, condense, as well as visualize related and appropriate info from social communications and user generated content in numerous are as for instance politics, business, consumer decision-making or public administration, [1-3]. These activities, on the other hand, are regarded problematic tasks because of the huge amount of diverse social media platforms and the huge volume, dynamics, and density of social media data.

Students converse as well as share their routine encounters in a casual manner on numerous social media sites. Students' digital footprints give huge extent of implicit knowledge and a complete novel viewpoint for researchers and practitioners to recognize students' skills outdoor the classroom environment [3]. This acquaintance could notify institutional decision-making on intrusions for threatened students, enhancement of the quality of education, and therefore improve student recruitment, withholding as well as success. The plenty of social media data gives chances to know students’ experiences however as well increases methodological problems in creating sense of social media data for educational reasons. Conventionally, researchers are utilizing approaches for instance interviews, surveys, class room activities, focus groups with the aim of gathering data associated to students learning skills. These approaches are time taking, therefore could not be replicated or repetitive with higher frequency. The measure of such researches is typically limited. Furthermore, while provoked regarding their understandings, students must reproduce on what they were thinking and doing formerly that might have turn out to be hidden over time.

Data mining is the process of taking outthought-provoking, interpretable, valuable as well as new info from data. It is utilized for a lot of years by scientists, businesses, and governments to sieve via massive amount of data such as census data, airline passenger records, and the supermarket scanner data, which generates market research reports. The aim of data mining in every application field is diverse. E.g., in case of business the key goal is to rise profit that is noticeable and gauged in regard to the money, amount of customers as well as customer loyalty. The forthcoming research objectives are 1) to exhibit a workflow of social media data sense-making for educational reasons, incorporating qualitative analysis as well as extensive data mining methods, 2) to search engineering student’s casual conversations on Twitter, with the purpose of knowing problems that students meeting in their learning practices.

The behaviour identification as well as emotional based data mining techniques are studied in this research. The pros and cons of the present methods in mining social media data are elucidated in the forthcoming segments. The performance of previous methods is examined. The respite of the other segments is arranged in this manner: the previous mining techniques of social media data are conversed in part 2; the dares and defies of previous methods in social media data is enlightened in part 3. The performance outcomes are depicted in part 4, and the conclusion is conversed in part 5.

II.LITERATURE SURVEY

An enhanced user-interest model based event evolution model, called HEE (Hot Event Evolution) model was presented by Shi et al., (2017) [4]. This method takes the user's interest distribution as well as utilizes the short text data in micro blogging network with the aim of modelling the data, stating the data sparsity issue, enhancing the topic definition quality as well as the interest level of every user for the period of the progression of hot events. An instinctive hot event filter is utilized rank popularity, eliminating the effect of common events and enhancing the mining event quality. A topic clustering technique is utilized to order the associated short texts into a solitary text document for resolving the sparse data issue. Lastly, in keeping with the users in the document as well as the scored topics, the topics of every document are modelled by LDA (Latent Dirichlet Allocation) topic model for obtaining the topics of the entire document and the users' interests. Experimental outcomes on micro blogging networks indicate that the HEE model can give wealthier information for the community structure of the identified events, and specify the accurateness and effectiveness of our presented system for user interest discovering as well as event identification for the period of the evolution of hot events.

A two-step analysis model was presented by Akay et al., (2015) [5], which concentrates on positive as well as negative sentiment, in addition to the side-effects of treatment, in users' forum posts, and finds the user communities (modules) and prominent users to ascertain the user choice of cancer treatment. In order to analyse the word frequency data derived from the users' forum posts, a Self-Organizing Map is utilized. After that a new network-based method was presented for modeling users' forum interactions as well as applied a network partitioning technique dependent upon enhancing a constancy quality measure. This let us to identify consumer view and determine prominent users in the recovered modules by means of info derived from network-based properties as well as word-frequency data. This method could extend research into logically mining social media data for consumer view of numerous treatments for giving fast, latest info for the hospitals, pharmaceutical industry, and medical staff, on the efficiency (or ineptitude) of forthcoming treatments.

A novel method called the integrated mixed methods technique (IMiME) was presented by Guendalina et al., (2015) [6] for observing the social media content by explaining a case study 'hypoactive sexual desire disorder (HSDD). This method is the mixture of quantitative as well as qualitative methods. Though researcher has proposed the outcome of IMiME as balanced amid environmental and located nature of qualitative research approaches with codification of quantitative measurement. This technique should be assessed with other case studies for solid gratification.

The design of personalized intervention system (PIS) was introduced by Tkalcicet al., (2014) [7] and has selected theory of planned behavior (TPB) technique for identifying if user would be present for the traditional musical concert or not via social media mining. Since, theory of planned behavior (TPB) doesn't think through the emotional elements such as feelings, mode; emotion element that contains direct consequence on personalized intervention system (PIS) predictions is not taken into account.

The social media (micro-blogging sites) as well as web (Musical content sites) for Musical Information Retrieval (MIR) was utilized by Sergio Oramas., (2014) [8]. Knowledge Extraction, Social Media Mining, and Natural Language Processing methods are united in Musical Information Retrieval (MIR), which would perform on structured as well as unstructured data that would obtain from web as well as micro-blogging sites correspondingly.

An evaluation of execution as well as performance with twelve (12) open source graph databases was carried out by Robert McColl et al., (2014) [9]. The researcher selected similar set of nodes, hardware, and edges for execution as well as performance. Every open source graph database is assessed with every algorithm that are

Connected Components, Single Source Shortest Path (SSSP), Page Rank and Update Rate.

Several approaches were assessed by Praveshet al., (2014) [10] that are utilized for opinion and sentiment. Support Vector Machine (SVM), Naïve Bays Classifier, Clustering techniques, Multilayer Perceptron were utilized with the intention of examining and comparing the outcomes of every methods, which provide numerous advantages and boundaries, every technique could be used in keeping with the state of affairs for features as well as text extractions. There are numerous approaches that is used for extraction of sentiments from micro-blogging sites.

Distance, influence, similarity, and adjustments-based methods were presented by Simoes et al. [11] for recognizing and identifying human behaviour for social communities. A model was designed by Zhang et al. [12] which are also known as socio scope for identifying human behavior in social networks. The public part of users' profile pages was crawled by Gyarmati et al. [13, 14], that comprised of online status info of the users. A social network based human dynamics model was presented by Yan et al. [15] to analyse the association amid the social network attributes of micro blog users and their behavior. On the other hand, owing to the complexity as well as diversity of human social behavior, no other method would identify each attributes, which rises while humans involve in social behaviours.

Sentiment propagation in social network was studied by Zafarani etal. [16] By utilizing a case study of Live Journal website. A technique was presented to handle the issue of product facets that are nouns and indicate choices by means of a big corpus in [17]. In [18], researchers have studied regarding numerous dares in designing opinion mining tools for social media. A hybrid method was presented by Ortigosa et al. [19] for carrying out sentiment analysis in Facebook with greater accurateness.

Numerous previous researches on tweet classification are binary classification on relevant as well as irrelevant content[20], or multi-class classification on generic classes for instance events, news, deals, opinions, and private messages [21].Sentiment analysis is also a wide spread three-class classification on positive, negative, or neutral emotions/opinions[22], which is valuable for mining customer views on products or companies via their online posts or reviews. It identifies extensive adoption in marketing as well as customer relationship management (CRM). Numerous approaches were designed to mine sentiment from texts. E.g., Davidov et al. [23] as well as Bhayaniet al. [24] utilize emoticons as pointers to give noisy labels to the tweets therefore reducing human effort required for labeling. On the other hand, according to this research, just recognizing the feeling of student-posted tweets doesn't give more actionable knowledge on appropriate intrusions and services for students. Our goal is to attain in-depth and better recognition of students' experiences particularly their learning-associated problems. However, for a human judge for identifying the student tweet is a difficult task compared to identify the feeling of a tweet. So, our research needs a qualitative analysis, and is difficult to perform in an entirely unsupervised manner. Sentiment analysis is, consequently, not appropriate to our research.

A workflow to incorporate large-scale data mining techniques as well as qualitative analysis was developed by Chenet al., (2014) [25]. We concentrated on engineering students' Twitter posts for knowing the problems in their educational practices. A qualitative analysis has been carried out on samples got from around 25,000 tweets associated with engineering students' college life. We identified engineering students come across issues for instance deficit of social engagement, substantial study load, and sleep deficiency. We developed a multi-label classification algorithm dependent upon these outcomes, to categorize tweets replicating students' issues. We utilized the algorithm to train an indicator of student issues from around 35,000 tweets streamed at the geo-location of Purdue University. This research, initially, proposes a method and outcomes, which indicate how casual social media data could give perceptions into students' practices.

Table 1: State-of-the-Art among Various Data Mining Schemes

Authors	Method	Advantages	Disadvantages
Graffigna, G., & Riva, G. (2015) [6]	Integrated mixed Methods(IMiMe) approach	Mixed-methods approach is combination of different qualitative and quantitative analytical strategies to overcome the drawbacks of available methods.	IMiMe approach is case specific and needed to be checked and verified with other cases
Tkalcic, M. B. (2014) [7]	Theory of Planned Behavior Model (TPB)	TPB model is used for predicting the behavior of users	theory of planned behavior overlooks emotional variables such as threat, fear, mood and negative or positive feeling
Oramas, S. (2014) [8]	Music Information Retrieval (MIR).	methodology that combines Social Media Mining, Knowledge Extraction and Natural Language Processing techniques, to extract meaningful context information for music from social data.	Application will not distinguish Between comparative sentiments and regular sentiments.
McColl, R. C. (2014) [9]	1. Single Source Shortest Path (SSSP) 2. Connected Components 3. Page Rank 4. Update Rate	Evaluate different algorithms of Opinion and sentiment mining and compare them	Precision and Recall technique can also compare for sentiment and opinion mining
Singh, P. K. (2014) [10]	1. Naïve Bays Classifier 2. Support Vector Machine (SVM) 3. Multilayer Perceptron	Discuss the execution power and comparison between open Source Graph databases with same hardware and collection of	Support vector machine can also be utilized for comparison between open source graph databases

	4. Clustering	edges.	
Chen et al., (2014) [25]	Naive Bayes Multi-Label Classifier scheme	It was attained better results for learning students experience	But, the information was lost and precision and recall values are low due to the high error rate.

III.THE CHALLENGE OF MININGSOCIAL MEDIA DATA

The inference of previous mining social media data techniques were conversed in this part.

- The developing areas of Educational Data Mining (EDM) as well as learning analytics have concentrated on examining structured data got from classroom technology usage, Course Management Systems (CMS), or online learning environments to notify educational decision-making. On the other hand, there is no research identified to unswervingly mine and examine student posts from social web comprising the flawless objective of knowing students' learning practices.
- Usually, examining social media data for educational reasons is problematic in educational data mining, in learning analytics, and learning methods because of the physical qualitative analysis as well as extensive computational analysis of user-generated textual content. Consequently, this study concentrates on obtaining additional knowledge of college experiences of the engineering students and could notify practitioners, educational administrators and other appropriate decision makers.
- Gathering of social media data associated with the experiences of the students was a thought-provoking task owing to the abnormality as well as diversity of the language utilized.
- In previous research, via a qualitative content analysis, we identified that engineering students are mainly stressed with the extensive study load, and are not capable of handling it effectively. Extensive study load brings about numerous effects comprising sleep difficulties, absence of social engagement, and other mental as well as physical health issues. Numerous students think engineering is very uninteresting as well as difficult that brings about deficit of inspiration to study and negative feelings. Diversity problems as well expose culture conflicts and culture stereotypes present amongst engineering students. On the other hand, the training time of this technique was high.

The previous method was not appropriate for extensive data because of the high computational time.

IV.RESULTS AND DISCUSSION

The performance of the previous social media mining techniques is assessed in this part. The performance of previous methods such as Support Vector Machine (SVM) [10], Naive Bayes Multi-Label (NBML) classifier [25] and Max-Margin Multi-Label (M3L) classifier [26] are assessed.

In Fig. 1, the 2,785 tweets are taken into account and the number of tweets in all type is depicted. When one tweet falls underneath several types, it was calculated numerous times; therefore the total amount of tweets seems above 2,785.

Explanation of every theme is along these lines. Consider, sample tweets given in every theme might as well fall innumerous other types all at once, excluding for the ones in "others".

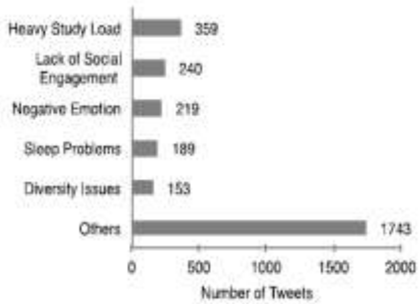

Figure 1: Number of Tweets in Each Category of the 2,785 Tweets Analyzed

The Naive Bayes multi-label classifier is utilized in this part for identifying engineering student issues from the Purdue data set. In the Purdue tweet collection, there were 35,598 distinctive tweets. We considered a random sample of 1,000 tweets, and identified just 5 percent of these tweets were conversing engineering issues. The main objective was to identify the minimum amount of tweets, which replicate the issues of the engineering students. The dissimilarities amid Purdue data set and #engineering-Problems data set is that the latter encompasses minimum amount of positive samples to be identified, and its “others” type contains different content. So, to form the training set well acclimatize to the Purdue data set, we considered a random sample of 5,000 tweets from the Purdue data set, included them into the 2,785 #engineering issues tweets, and considered them as “others”. Below 5 percent positive samples in this type don’t impact the efficiency of the trained model. So as to train the multi-label Naive Bayes classifier, we thereforeutilized7, 785 tweets as input. As no additional human power is required, and Naïve Bayes classifier is effective in regard to computation time, the model training here experienced nearly no additional cost.

The greatest possible words in every type ranked utilizing the conditional probability was depicted in Fig 2.The goal is to identify the minimum amount of the five issues from the huge Purdue data set, therefore, in this part, the “others” is not conversed.

Category	*Top 25 words*
Heavy Study Load	hour, homework, exam, day, class, work, negtoken, problem, study, week, toomuch, all, lab, still, out, time, page, library, spend, today, long, school, due, engineer, already
Lack of Social Engagement	negtoken, Friday, homework, out, study, work, weekend, life, class, engineer, exam, drink, break, Saturday, people, social, lab, spend, tonight, watch, game, miss, party, sunny, beautiful
Negative Emotion	hate, f***, shit, exam, negtoken, week, class, hell, engineer, suck, study, hour, homework, time, equate, FML, lab, sad, bad, day, feel, tired, damn, death, hard
Sleep Problems	sleep, hour, night, negtoken, bed, allnight, exam, homework, nap, coffee, time, study, more, work, class, dream, ladyengineer, late, week, day, long, morning, wake, awake, no-sleep
Diversity Issues	girl, class, only, negtoken, guy, engineer, Asia, professor, speak, English, female, hot, kid, more, toomuch, walk, people, teach, under-stand, chick, China, foreign, out, white, black

Figure 2: Top 25 Most Probable Words in Each Category

IJARSE
ISSN: 2319-8354

Every type contains top words for the particular content of this type. E.g., “life, weekend, social, drink, break” for the type Lack of Social Engagement, “all night, sleep, nap, dream, coffee,” for Sleep Problems, “suck, hate, bad, sad,” for the type Negative Emotion, and “guy, girl, Asia, China, female, foreign” for Diversity problems. This naturally shows the efficiency of the classification model.

In Fig. 3, the trained model utilized to the remaining 30,598 Purdue tweets, and identified a total of 940 tweets replicating the five issues students come across was depicted. Once more, one tweet might be innumerous diverse types, consequently the total amount of tweets of each and every types seems in excess of 940.

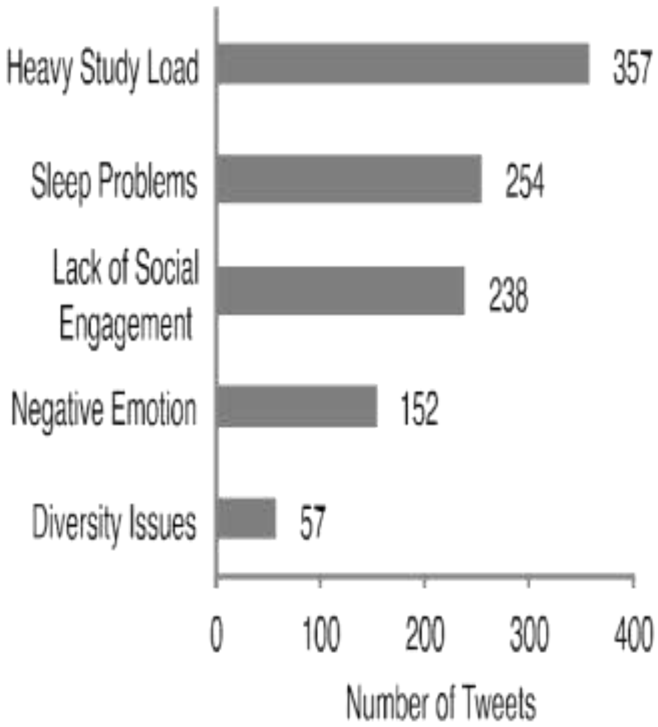

Figure 3: Number of Tweets for Each Issue Detected from the Purdue Tweet Collection

A comparatively minimum amount of tweets replicating the diversity problems that is not as serious as replicated by the engineering Problems tweets. The Engineering College at Purdue contains various international students, and is continuously making attempts to raise the admission of female and students in other understated categories. It is a supposition that there is a correlation amid Purdue’s effort and what the data indicate.

Every 940 tweets are related to a Twitter user account. The top 15 users, who post the top five engineering issues was depicted in Fig. 4. So as to safeguard their privacy, the twitter usernames were anonymized. 940 tweets consider 3.07 percentages of the 30,598 tweets.

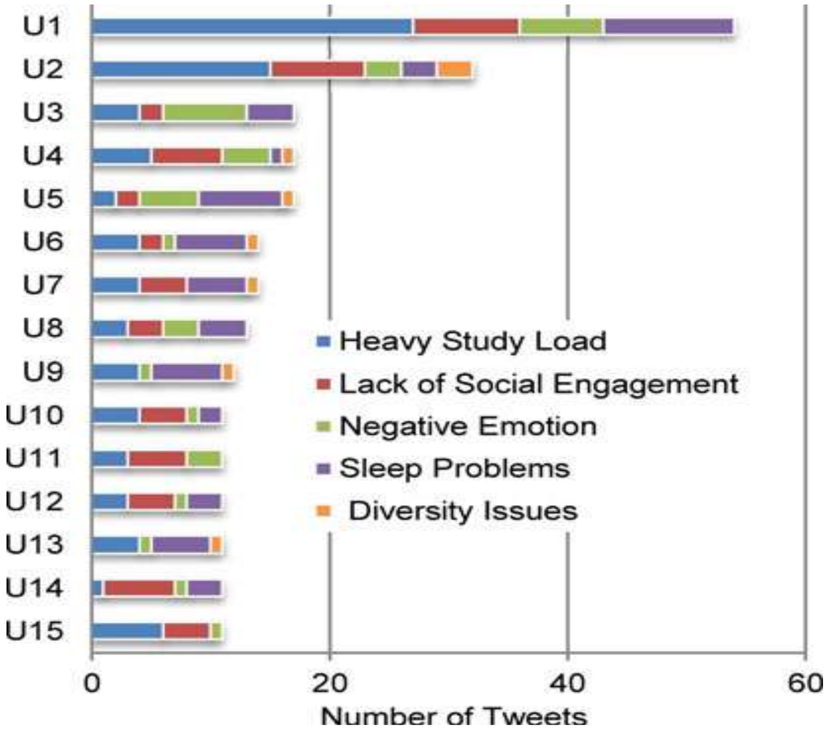

Figure 4: Top 15 Users in the Purdue Tweet Collection who posted the Most on the Five Engineering Problems

Performance Evaluation Matrices

With the aim of assessing the performance of classification models in regard to precision, accuracy, recall, and the harmonic average amid precision and recall—the F1 score, the provided matrices are utilized.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (1)$$

$$Precision\ p = \frac{tp}{tp + fp} \quad (2)$$

$$Recall\ r = \frac{tp}{tp + fn} \quad (3)$$

$$F_1 = \frac{2.p.r}{p + r} \quad (4)$$

Here the $tp -$ positivelyidentified true value, fp −positively identified false value, $tn -$ negatively identified true value and fn- negatively identified false value.

Accuracy Comparison

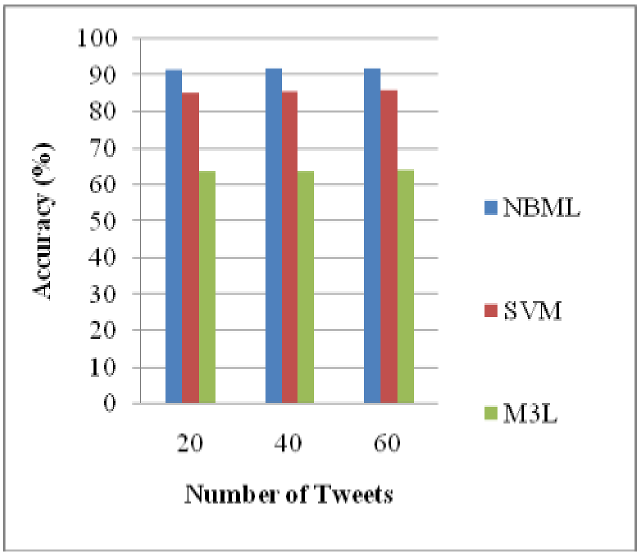

Figure 5: Accuracy Comparison among Various Data Mining Schemes

According to Fig 5, it is noticed that the evaluation of accurateness among stall data mining techniques. The number of tweets is taken in x-axis and accurateness is considered in y-axis. The SVM, NBML, and M3Lapproaches are assessed. These techniques attained greater accurateness outcome of 92.05%, 86.14% and 64.25%correspondingly. It proves that the NBML technique has depicted superior outcomes for all specified tweets.

F1-Score Comparison

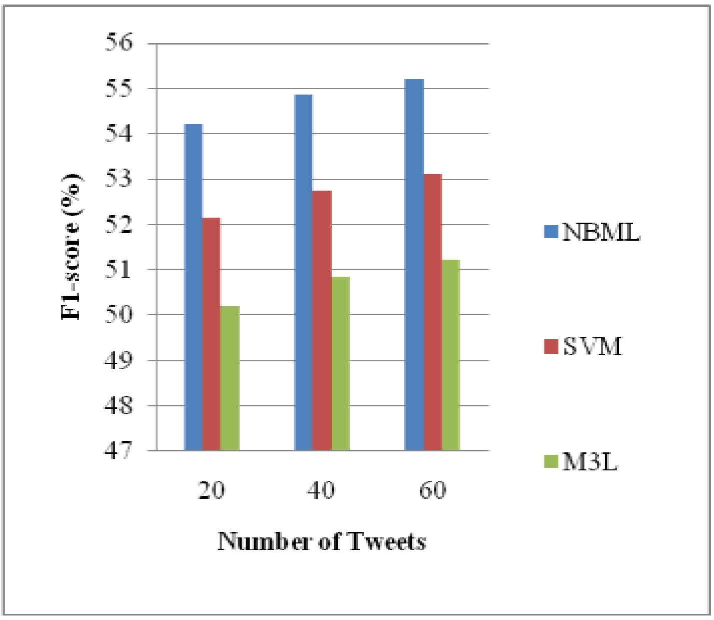

Figure 6: F1-Score Comparison among Various Data Mining Schemes

According to Fig 6, it is clear that the evaluation of F1-scorein contradiction of all data mining techniques. The number of tweets is considered in x-axis and F1-scoreis considered in y-axis. The NBML, SVM and M3Lapproaches are assessed and these techniques attainedgreaterF1-scoreoutcome of 55.21%, 53.12% and

51.21%correspondingly. It proves that the NBML technique displayed greater outcomes for all specified tweets.

V.CONCLUSION

People follow numerous events for instance political problems, natural disasters, sports events and posts their comments or sentiments (i.e. emotion) regarding the specific event in Social media. In this research, a thorough valuation of diverse articles is provided. There are numerous researches that was carried out on sentiment mining of social media on the other hand this mining method is presently imprecise. There are numerous research gaps in sentiment and data mining that are elucidated in the research. Numerous researchers are till now working on sentiment and data mining and how to progress these specific mining methods. These researches are dependent upon a complete research on sentiment mining, and behaviour prediction and it is conversed the restrictions and to state the novel resolution.

The educational learning contains a smaller amount of consideration in social media. Consequently, in future, to enhance the student learning in educational data mining would be concentrated. The emotional and behaviour based quantitative analysis were focused by previous Naive Bayes Multi-Label Classifier technique. On the other hand, it contains limited restrictions such as convergence could be slow, local minima could have impact on the training process and Hard to measure. Consequently, henceforth, the neural network based classification scheme as well as ranking technique would concentrate to enhance the students learning classification accurateness.

REFERENCES

[1] G. Siemens, and P. Long, Penetrating the Fog: Analytics in Learning and Education, Educause Rev., 46 (5) 2011, 30-32.

[2] R. Ferguson, The State of Learning Analytics in 2012: A Review and Future Challenges, Technical Report KMI-2012-01, Knowledge Media Inst., 2012.

[3] S. Cetintas, L. Si, H. Aagard, K. Bowen, and M. Cordova-Sanchez, Micro blogging in Classroom: Classifying Students' Relevant and Irrelevant Questions in a Micro blogging-Supported Classroom, IEEE Trans. Learning Technologies 4 (4) 2011, 292-300.

[4] L.L. Shi, L. Liu, Y. Wu, L. Jiang, and J. Hardy, Event Detection and User Interest Discovering in Social Media Data Streams, IEEE Access, 2017.

[5] A. Akay, A. Dragomir, and B.E. Erlandsson, Network-based modeling and intelligent data mining of social media for improving care, IEEE journal of biomedical and health informatics 19(1) 2015, 210-218.

[6] G. Graffigna, and G. Riva, Social Media Monitoring and understanding: an integrated mixed methods approach for the analysis of social media, International Journal of Web Based Communities 2015, 57-72.

[7] M.B. Tkalcic, Using social media mining for estimating theory of planned behaviour parameters, 2014.

[8] S. Oramas, Harvesting and structuring social data in music information retrieval. In The Semantic Web: Trends and Challenges, Springer International Publishing 2014, 817-826.

[9] R.C. McColl, A performance evaluation of open source graph databases, ACM in Proceedings of the first workshop on Parallel programming for analytics applications, 2014, 11-18.

[10] P.K. Singh, Methodological study of opinion mining and sentiment analysis techniques, International Journal on Soft Computing 5(1), 2014.

[11] J. Simoes, and T. Magedanz, Understanding and Predicting Human Behavior for Social Communities, in the Handbook of Social Network Technologies and Applications, Springer, 2010, 427-445.

[12] H. Zhang, R. Dantu, and J. W. Cangussu, Socio scope: Human Relationship and Behavior Analysis in Social Networks, Journal of IEEE transactions on systems, man, and cybernetics - part a: Systems and Humans 41 (6) (2011),1122-1143.

[13] L. Gyarmati, and T.A. Trinh, "Measuring user behavior in online social networks, Journal of IEEE Network: The Magazine of Global Internetworking 24 (5), 2010, 26–31.

[14] L. Gyarmati and T.A. Trinh, Measurement Methods of User Behavior in Online Social Networks, in Computational Social Networks: Security and Privacy, Chapter 7, Springer, 2012, 157-175.

[15] Q. Yan, L. Wu, L. Zheng, Social network based micro blog user behavior analysis, In the Journal of Physica A: Statistical Mechanics and its Applications 392 (7), 2013, 1712–1723.

[16] R. Zafarani, W.D. Cole, and H. Liu, Sentiment Propagation in Social Networks: A Case Study in Live Journal, in Advances in Social Computing, Springer, 2010, 413–420.

[17] L. Zhang, and B. Liu, Identifying noun product features that imply opinions, In Proc. of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011), 2011, 575-580.

[18] D. Maynard, K. Bontcheva, and D. Rout, Challenges in developing opinion mining tools for social media, Proc.of @NLP can u tag #usergeneratedcontent?! Workshop at LREC 2012, 15–22.

[19] A. Ortigosa, J.M. Martín, and R. M. Carro, Sentiment analysis in Facebook and its application to e-learning, In the Journal of Computers in Human Behavior, 2013.

[20] K. Nishida, R. Banno, K. Fujimura, and T. Hoshide, Tweet Classification by Data Compression, Proc. Int'l Workshop Detecting and Exploiting Cultural Diversity on the Social Web, 2011, pp. 29-34.

[21] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, Short Text Classification in Twitter to Improve Information Filtering, Proc. 33rd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, 2010, 841-842.

[22] B. Pang, L. Lee, and S. Vaithyanathan, Thumbs Up?: Sentiment Classification Using Machine Learning Techniques, Proc. ACL-02Conf. Empirical Methods in Natural Language Processing, 10, 2002, 79-86.

[23] D. Davidov, O. Tsur, and A. Rappoport, "Enhanced Sentiment Learning Using Twitter Hashtags and Smileys, Proc. 23rd Int'l Conf. Computational Linguistics: Posters, 2010, 241-249.

[24] A. Go, R. Bhayani, and L. Huang, Twitter Sentiment Classification Using Distant Supervision, CS224N Project Report, Stanford, 2009, 1-12.

[25] X. Chen, M. Vorvoreanu, and K. Madhavan, Mining social media data for understanding students' learning experiences, IEEE Transactions on Learning Technologies 7(3), 2014, 246-259.

[26] B. Hariharan, L. Zelnik-Manor, S.V.N. Vishwanathan, and MVarma, Large Scale Max-Margin Multi-Label Classification with Priors, Proc. Int'l Conf. Machine Learning, 2010.