ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse54942020.pdf
https://doi.org/10.30534/ijatcse/2020/54942020

Investigating Text Mining Features and Classifiers: An Experimental Analysis

Loveleen KaurPabla[1], Prashant Jain[2], Prabhat Patel [3]
[1] Assistant Professor, IT Department, Jabalpur Engineering College, Jabalpur, India, loveleenkaurpabla@gmail.com
[2] HOD, IT Department, Jabalpur Engineering College, Jabalpur, India, pjain@jecjabalpur.ac.in
[3] Professor, EC Department, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, India, peekaypatel@gmail.com

ABSTRACT

The proposed work presents the detection of hate speech in social media. In this context, the text mining techniques are valuable. This, in turn, requires an accurate classification algorithm as well as text feature selection technique which works well with the sentiment classification. Therefore, a review of existing techniques is performed first. According to literature, amongst the techniques suggested by the researchers, the GI (Gini Index), DF (document frequency), POS (part of speech) tagging, and IG (information gain)are popular and frequently used techniques for sentiment classification. Thus, these techniques are chosen to implement the text feature selection. Further, the SVM (Support Vector Machine) and KNN (k-Nearest Neighbour) algorithms are applied to classify the selected features. The experimental outcomes show that the SVM is accurate and efficient algorithm for classifying measured features. Additionally, it is seen that the combination of SVM with the feature extraction techniques POS and IG is also time efficient. The proposed work may be extended by use of the prevalent algorithms.

Key words: Sentiment analysis, text mining, natural language processing, SVM, KNN

1. INTRODUCTION

Text is an essential part of daily life and we are surrounded by the different forms of text such as emails, SMS, social media post, articles, blogs and many more forms [1]. However, text reading and understanding is a lengthy task. Therefore, at the industry level, we need automated tools and techniques that help to find conclusion of huge text in a small amount of summarized data [2]. Basically the text is an unstructured form of information. Handling and recovering the valuable information from the raw text needs additional concentration and handling expertise. In this context, the text mining technique is very helpful [3].

Text mining is a sub-domain of data mining. For text mining the data mining algorithms are used with slight modifications, for dealing and handling of the text data [4]. The handling of such data involves some basic issues such as the length of different documents or data instances, semantics, the quantity of data and many more. In this context, the proposed work is focused on the two major complexities for text classification.

1. Obtaining the effective sentiment-based feature selection technique for text data analysis, and
2. Obtaining an accurate and efficient algorithm for classifying the text features

Therefore the proposed work establishes the contribution as the objective of the given paper:

1. Finding the frequently used text features which are used for text classification, categorization and sentiment analysis applications
2. Finding the effective classifier which can be used for recognizing the text patterns and able to produce accurate classification outcomes for the text data.

This paper is organized as follows: section II provides an overview of the work reviewed in this paper, further, a survey on different text mining and sentiment analysis techniques are provided. On the basis of the literature surveyed, the different text feature selection techniques are explored and used further for extracting text features in section III. Additionally, the used different classifiers are also investigated in this paper. In section IV, the implementation of all the chosen techniques is carried out and their performance is measured. Finally, conclusion is presented in section V.

2. BACKGROUND

This section provides the recently contributed work and research which help to find the different frequently used features and classifiers for text based sentiment analysis.

2.1 Text Analysis

The Internet is perceived as a wellspring of data and feelings for trade, just as purposeful publicity. An assumption investigation is offered by A. Abbasi et al [5] for grouping of multi-lingual web gathering. The expressive and syntactic highlights are assessed for English and Arabic substance. The Entropy Weighted Genetic Algorithm (EWGA) is industrialized, which expends the data gain (IG) for include

choice. The aftereffects of EWGA with SVM demonstrate elite, with exactness over 95% and 93% for both datasets.

S. Tan et al [6] presents an examination of assumption request on Chinese reports. Four-part decision procedures (MI, IG, CHI, and DF) and five learning methods (centroid classifier, K-nearest neighbor, winnow classifier, Naı¨ve Bayes, and SVM) are inspected with 1021 reports. The results exhibit that IG plays out the best for thoughtful terms decision and SVM shows the best execution for feeling requests.

In this paper, considering Fisher's discriminant extent, a fruitful component assurance procedure is proposed by S. Wang et al [7] for subjectivity content feeling arrangement. To approve that strategy, creators contrasted it and the strategies subject to Information Gain and SVM are gotten. The examinations are driven by joining different component decision methods with two contender features. Under 2739 reports and 1006 vehicle-related records, the results show that Fisher's discriminant extent subject to word repeat estimation has the best execution with precision 86.61% and 82.80% with two applicant highlights.

In this paper, A. Duric et al [8] propose a great deal of new part decision designs that use a Content and Syntax model to thusly learn remembers for a record by segregating the components from the theoretical enunciations.

Individuals express their perspectives and assessments via web-based networking media and, notion investigation worries about distinguishing and removing slant from the content. Slant based content grouping is not the same as topical content characterization. Highlight choice is huge for feeling investigation. A. Sharma et al [9] explores the tangibility of feature decision methodologies for assessment and examines their presentation in terms of audit, precision, and accuracy. Five segment decision strategies (Document Frequency, Information Gain, Gain Ratio, Chi-Squared, and Relief-F) and three notable presumptions incorporate word references (HM, GI, and Opinion Lexicon) are explored. The outcomes show that IG gave steady outcomes and Gain Ratio.

It is trying to comprehend the most recent patterns and sum up the state or suppositions about items because of the enormous assorted variety and size of information. Mining supposition is a type of notion examination that is treated as a grouping. E. Haddia et al [10] investigate the job of content pre-handling and report results that exhibit that with proper component determination and portrayal, assumption examination exactnesses utilizing bolster vector machines (SVM). The exactness accomplished is demonstrated to be similar to the ones accomplished in theme categorization.

S. Poria et al [11] present a novel method for removing highlights from short messages, in light of the initiation estimations of an inward layer of a profound Convolutional neural system. Creators utilize the highlights in multimodal opinion investigation of short video cuts. The consolidated highlights of printed, visual, and sound modalities are utilized to prepare a classifier. They get a 14% improvement over the cutting edge. That is a lot quicker, however somewhat less precise.

A colossal number of reviews are posted by the buyers on the Web. Such information is significantly basic for dynamic. This information is genuinely huge for purchasers just as for associations in envisioning the accomplishment and legitimacy. In this paper, a Gini Index based part decision methodology with Support Vector Machine (SVM) classifier is proposed for appraisal portrayal by A. S. Manek et al [12]. The results show that the Gini Index has better portrayal execution with respect to precision.

Highlight choice straightforwardly influences the depiction of reviews. In any case, little thought has been paid to feature the assurance of Chinese online reviews. Accordingly, L. Zheng et al [13] are propelled to research the effects of feature decisions on appraisal assessment. Directly off the bat, N-consume grams and N-POS-grams are picked as the potential insightful features. By then, the improved Document Frequency procedure is used to pick feature subsets, and the Boolean Weighting system is grasped to process incorporate weight. Assessments subject to the Chi-square test are finished to test the significance of exploratory. The results recommend that end examination obtains higher precision when taking 4-POS-grams as features. Likewise, low solicitation N-consume grams can achieve a predominant presentation than high solicitation N-scorch grams when taking N-cook grams as features.

Because of the multifaceted nature of human dialects, the vast majority of assessment characterization calculations are experienced a high component of vocabularies. Profound Belief Networks (DBN) handles this issue by learning valuable data in input corpus with their few concealed layers. Lamentably, DBN is tedious and computationally costly. P. Ruangkanokmas et al [14], a semi-managed learning calculation, called Deep Belief Networks with Feature Selection (DBNFS) is created. Utilizing our chi-squared based component determination, the multifaceted nature of the jargon input is diminished since some insignificant highlights are sifted which makes the learning period of DBN progressively productive. The test aftereffects of our proposed DBNFS shows that the proposed DBNFS can accomplish higher order exactness and can accelerate preparing time contrasted and other notable semi-managed learning calculations.

Estimation examination is to characterize surveys dependent on its feeling as positive or negative. First B. M. Jadav et al [15] have pre-handled the dataset to change over unstructured information into organized. They have utilized a vocabulary based way to deal with convert organized audits into numerical scores. In this methodology, a pre-prepared dataset incorporates Stop word evacuation, stemming, POS labeling, and ascertaining the conclusion score utilizing SentiWordNet. At that point, they applied the order calculation to arrange a

conclusion. SVM is utilized to arrange audits where RBF pieces than SVM and innocent Bayes.

A. Nasser et al [16] execute a document level oversaw evaluation examination system for the Arabic setting. They use three assorted part extraction methods to make three datasets. To find the perfect number of features and to hear the best time in point of view assessment, the maker uses two-part situating procedures (Information Gain based and Chi-Square based) and discovers the score of every component concerning the classes. This part situating picks only the features that are relevant to the class stamps and removes the pointless features. Therefore, it helps with growing the portrayal execution and decrease the taking care of time. Finally, they evaluate the show of three standard classifiers on the as of late made unigram and bigram based enlightening files, to be explicit SVM, KNN, and Decision Tree. SVM classifier has demonstrated preferred execution as taken a gander at over various classifiers. The results show the sufficiency of the segment assurance methods.

Mining helpful information from video is a basic need that will develop exponentially, in pace with the worldwide development of substance. This is especially significant in the feeling examination, as both assistance and item surveys are step by step moving from uni-modal to multimodal. S. Poria et al [17] present a novel strategy to extricate highlights from visual and literary modalities utilizing profound convolutional neural systems. By taking care of highlights to numerous piece of learning classifier, it outflanks the cutting edge of multimodal feeling acknowledgment and estimation investigation. In [18] s. singh et al proposed special text mining method for goggle search string. Naive bays method is used here to classify search string and provide different category of internet user to find terrorists. But it is not finalized for summary report. In [19], [20] s. singh et al proposed well known multi criteria decision making methods for feature selection of graph. In [21] Hasan, AM proposed semantic feature selection method for knowledge based semantic relatedness. In [22] Perumal K described stability related feature selection.

2.2 Summary

In order to understand the frequently used techniques and methods for text processing and sentiment based text analysis. The table 1 shows the summary of the collected literature in the domain of the text processing and sentiment analysis.

Table 1: Literature Summary

Authors	**Contributions**	**Finding**
A. Abbasi et al [5]	multi-lingual web forum, stylistic and syntactic features	Entropy Weighted Genetic Algorithm (EWGA) is developed, that incorporates the information gain for feature selection.
S. Tan et al [6]	Four feature selection methods MI,IG, CHI and DF, Five learning methods : centroid classifier, K-nearest neighbor, winnow classifier, Naı¨ve Bayes and SVMare investigated	IG performs the best for sentimental terms selection and SVM
S. Wang et al [7]	Information Gain and SVM is adopted	Best performance with accuracy 86.61% and 82.80% with two candidate features.
A. Duric et al [8]	propose a set of new feature selection schemes that use a Content and Syntax model to automatically learn features	separating the entities from the subjective expressions
A. Sharma et al [9]	feature selection methods(Document Frequency, Information Gain, Gain Ratio, ChiSquared, and Relief-F) and three popular sentiment feature elexicons (HM, GI and Opinion Lexicon) are investigated	IG gave consistent results and Gain Ratio
E. Haddia et al [10]	explore the role of text pre-processing	sentiment analysis accuracies using support vector machines (SVM)
S. Poria et al [11]	extracting features from texts, based on the activation values of an inner layer of a deep CNN	The combined features of textual, visual, and audio modalities are used to train a classifier
A. S. Manek et al	a Gini Index based feature selection	Gini Index has better

[12]	method with Support Vector Machine (SVM) classifier is proposed	classification performance in terms of accuracy.
L. Zheng et al [13]	N-char-grams and N-POS-grams are selected as the potential sentimental features	Better performance than high order N-char-grams when taking N-char-grams as features.
P. Ruangkanokmas et al [14]	a semi-supervised learning algorithm, called Deep Belief Networks with Feature Selection(DBNFS) is developed	chi-squared based feature selection
B. M. Jadav et al [15]	this approach includes Stop word removal, stemming, POStagging and calculating sentiment score using SentiWordNet	SVM is used to classify reviews where RBF kernel than SVM and naïve bayes.
A. Nasser et al [16]	A document-level supervised sentiment analysis system for Arabic context.	Unigram and bigram based data sets, namely SVM, KNN and Decision Tree. SVM classifier has showed superior performance
S. Poria et al [17]	Present a novel method to extract features from visual and textual modalities using deep CNN.	it outperforms the state of the art of multimodal emotion recognition

2.3 Research pattern

Table 2 shows the algorithms used for extracting essential features from the text for sentiment classification purpose. Additionally the respective classifiers are also concluded which works effectively for classification of text features.

Table 2: Text Features and classifiers for sentiments

Reference	Text feature	Classifier
[5]	information gain	Genetic algorithm
[6]	MI,IG, CHI and DF	centroid classifier, K-nearest neighbor, winnow classifier, Naıtive Bayes and SVM
[7]	Information Gain	SVM
[8]	new feature selection scheme Content and Syntax model	separating the entities from the subjective expressions
[9]	Document Frequency, Information Gain, Gain Ratio, Chi Squared, and Relief-F) and HM, GI and Opinion Lexicon	IG and Gain Ratio
[10]	TF-IDF	SVM
[11]	DeepCNN	CNN
[12]	Gini Index	SVM
[13]	N-char-grams, N-POS-grams, Chi-square	KNN
[14]	chi-squared	Deep Belief Networks with Feature Selection(DBNFS)
[15]	POStagging	SVM and Naïve bayes
[16]	Unigram and Bigram	KNN, SVM, Decision Tree
[17]	CNN	CNN

The different text feature selection techniques are reported in figure 1 according to their utilization frequency in the domain of text feature selection and classification.

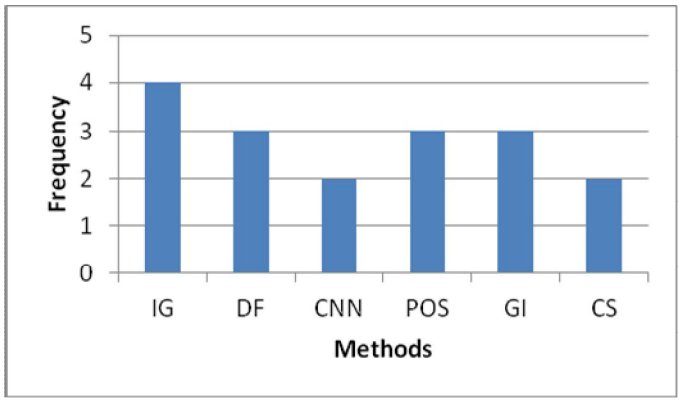

Figure 1: Text Feature Selection Techniques

A significant amount of different techniques are available for extracting the valuable features form text for analyzing the sentiments. Among them IG (information gain), DF (document frequency), CNN (convolutional neural networks), POS (part of speech tagging), GI (Gini Index) and CS (Chi-square test) are the frequently used approaches. In all these techniques we are selected IG, DF, POS and GI for further comparative study.

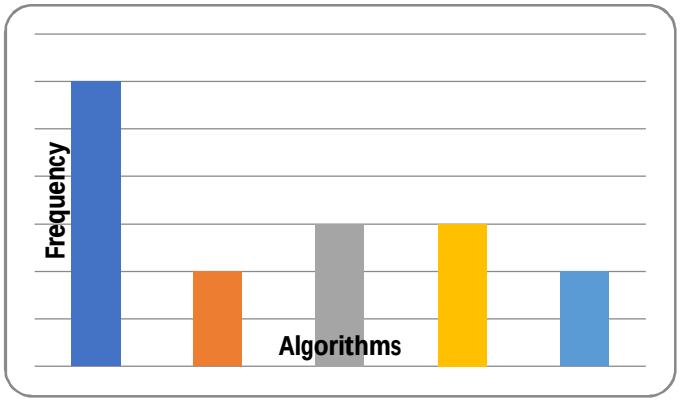

Figure 2: Classifiers

Similarly, figure 2 shows different classifiers are also identified which are frequently used for text processing task. Among them SVM (support vector machine), CNN (convolutional neural networks), KNN (k-nearest neighbor) classifier are frequently utilized now in these days.

3. PROPOSED WORK

This section explains the work carried out in this paper for the purpose of simulation. Therefore the work overview and the proposed simulation system is described in this section.

3.1 System overview

This paper provides the comparative performance study for different techniques and methods that are used for processing of text in other words for classifying the text data in terms of sentiment classes. Therefore using the collected literature different text feature selection techniques is obtained in addition of that the different classifiers are also available which are selected on the basis of their popularity. Using these concluded techniques and methods different combinations are prepared for experimentation. The combination of classifiers and feature selection techniques are provided in table 3.

Table 3: Experimental Scenarios

S. No.	Algorithm combination	Feasibility
1	GI + SVM	Ok
2	GI + KNN	Ok
3	GI + CNN	Not feasible
4	DF + SVM	Ok
5	DF + KNN	Ok
6	DF + CNN	Not feasible
7	POS + SVM	Ok
8	POS + KNN	Ok
9	POS + CNN	Not feasible
10	IG + SVM	Ok
11	IG + KNN	Ok
12	IG + CNN	Not feasible

According to the obtained literature the CNN mostly used with the self extracted features therefore these combinations are not implemented in this project work. Only the valid 9 feature extraction techniques and their classifier combinations are implemented with WEKA tool and JAVA are used here for experimentation.

3.2 Simulation Setup

The proposed system is simulated in figure 3. These system components are explained in this section with their functional aspects.

Input dataset: the proposed work is motivated to explore the domain of sentiment based text classification. In this context the twitter dataset is downloaded from online source. The dataset consist of user and twit information. In addition to it includes the predefined classes for training and testing. The dataset contains 2467 instances of data.

Data Pre-processing: the data pre-processing helps to improve learning process of algorithms by improving quality of data and by reducing noise. Here the incomplete data, special characters, and stop words are reduced from dataset. After pre-processing data is used further to obtain the features from data.

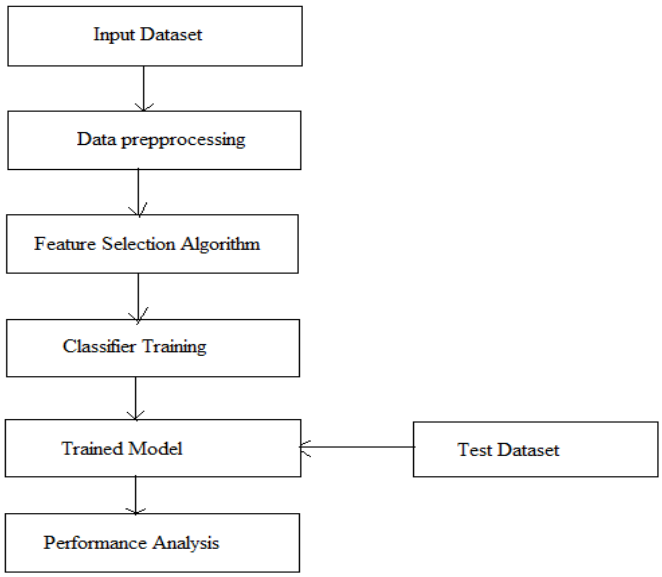

Figure 3: Proposed Model

Feature selection: the GI (gini index), DF (document frequency), POS (part of speech) tagging, and IG is implemented in this phase for extracting features from the image.

Classifier training: in this phase two popular text classifiers namely KNN (k nearest neighbour) algorithm and SVM (support vector machine) is implemented. In this context the WEKA library is used for implementing JAVA based UI (user interface) and experimental system. The training data in pre-processed form used here with both the algorithms and produces the trained model.

Trained model: the trained data model is used here for accepting the test dataset as input and produces the class labels for each instance of test data. The classified data is used in next phase for evaluation of the performance. Thus accuracy, memory usages and time consumption is measured.

Test dataset: in order to validate the trained model the test dataset is used. That data is random selected 30% of the data which is initially used as the input to the system.

Performance analysis: the classification outcomes of the test dataset are used for obtaining the accuracy of the trained model. Additionally the memory and time requirements of the algorithm for classifying the data are also measured.

4. RESULTS ANALYSIS

The aim of the proposed work is to find two fruitful outcomes from the experiments. First is to obtain an effective sentiment analysis feature set and second is the classifier which can be used in further studies.

4.1 Accuracy

The accuracy is an important parameter for efficiency measurement. That is the ratio of accurately recognized data patters and the total patterns provided for recognition. The following equation can be used for computing accuracy of the proposed system.

$$accuracy\,(\%) = \frac{total\ correctly\ classified}{total\ patterns} X100$$

Table 4: Accuracy (%)

Algorithm combination	SVM	KNN
GI	87.7	79.5
DF	82.4	77.9
POS	89.4	81.3
IG	86.5	80.9

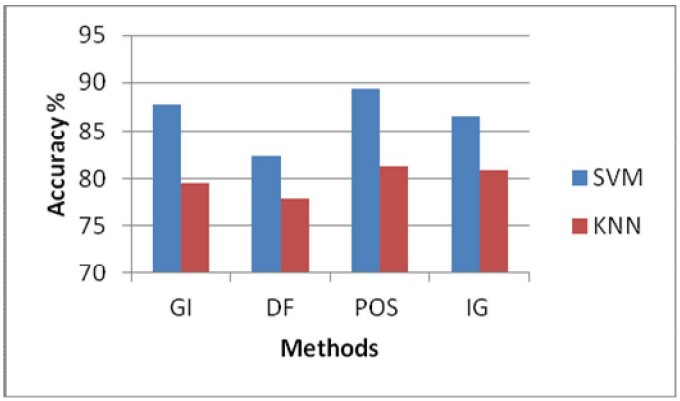

Figure 4: Accuracy (%)

The accuracy of algorithms and their feature combinations are provided in table 4 and in the figure 4 the X axis of this bar graph shows the feature selection algorithms involved additionally the Y axis shows the performance of algorithm in terms of percentage (%). According to the obtained performance the SVM classifier perform superior then KNN in all the implemented scenarios. In addition of that the combination of SVM and POS, SVM with GI and SVM with IG are much promising for feature selection for sentiment based text analysis.

4.2 Time consumption

Time consumption is also frequently known as the time complexity also. That is defined as the amount of time required for execution of the algorithm with the input amount of data. In our implementation the time consumption is measured in the following manner.

$$time\ consumption = end\ time - start\ time$$

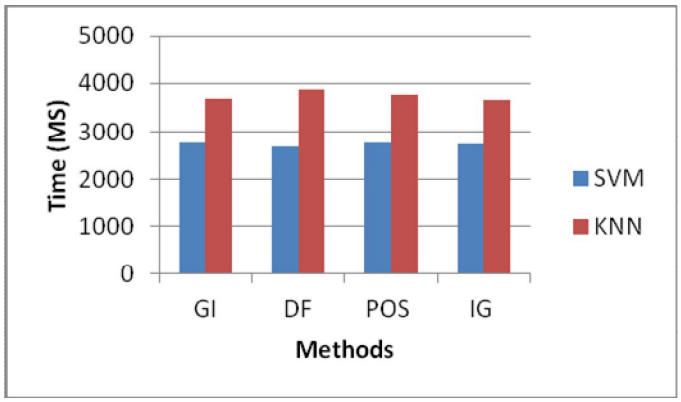

Figure 5: Time Consumption

The performance in terms of time consumption of all the implemented techniques is reported in table 5 and in figure 5 the X axis shows the feature selection methods implemented and Y axis demonstrates the time requirements of the algorithm. The KNN consumes higher amount of time as compared to SVM classifier. The time is measured here in terms of milliseconds (MS).

Table 5: Time Consumption

Algorithm combination	**SVM**	**KNN**
GI	2761	3672
DF	2676	3879
POS	2774	3764
IG	2741	3662

According to the given diagram the SVM is promising with the POS and IG. Because these two approaches are consuming less time and provides better accuracy in all the implemented experimental scenarios.

4.3 Memory Usages

The memory is an essential parameter for measuring the performance of an algorithm. Classically it is also known as the memory usages or space complexity. According to the available literature it is the part of main memory which acquired during the process execution. In Java technology that is calculated in the following manner.

$$memory\ usgaes = total\ assigned\ memory - free\ memory$$

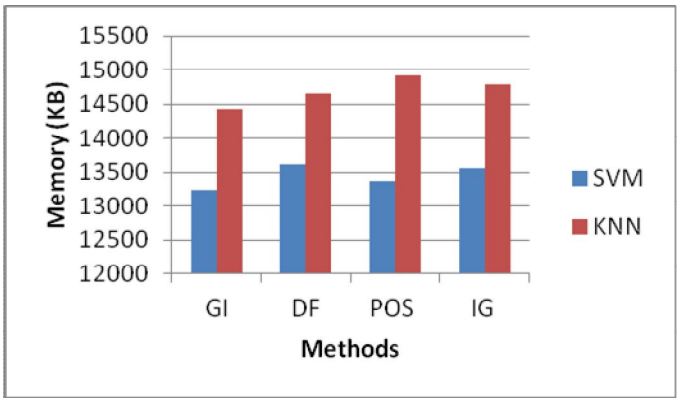

Figure 6: Memory Usages

Table 6: Memory Usages

Algorithm combination	**SVM**	**KNN**
GI	13245	14424
DF	13626	14662
POS	13372	14927
IG	13562	14782

The performance of the implemented experimental scenarios is demonstrated in figure 6 as well as in table 6. The X axis of the figure 4.3 shows the methods involved in experimentation and the Y axis shows the memory requirements of the algorithms. The memory is measured here in terms of KB (kilobytes). According to the investigative results the combination of SVM and POS, IG and GI is the promising techniques for text processing in the domain of the sentiment analysis or classification.

5. CONCLUSION AND FUTURE WORK

The aim of the proposed work is to obtain an efficient classification algorithm and a feature selection technique for proposed sentiment analysis of social media text. This section summarizes the entire efforts made for deriving these two consequences. Thus the conclusion includes the experimental observations and the future extension of the work also proposed.

5.1 Conclusion

The proposed work investigates the techniques of text feature extraction for sentiment classification. In order to understand the different techniques available for sentiment classification

and feature selection we collect a significant amount of literature among 13 of them are selected which are favoring our proposed investigation. In this literature we were found the GI (Gini-Index), DF (Document Frequency), POS (part of speech) tagging and IG (Information Gain) based feature selection approaches are much popular. Additionally the KNN (k-nearest neighbor) and SVM (support vector machine) is popular classifier for the target task.

Thus different combinations of the classifier and feature selection techniques are prepared; additionally a common dataset is taken for experimentations. That dataset is taken from online resource for twitter sentiment classification. The experiments are carried out with the help of this data. The three class labels are predicted namely positive, negative and neutral. The KNN algorithm is perform well with these feature selection techniques but the accuracy is low as compared to the SVM classifier. Therefore we extend this work with the help of SVM classifier. The SVM based classifier is works well with the GI, POS and IG feature selection method. Among them POS and IG is much effective techniques for classifying the text. These combinations are able to produce high accurate outcomes with less time and resource consumption.

5.2 Future work

The proposed work is aimed to classify the hate speech in social media text. In this context we require feature selection technique and a suitable classifier. In this experimentation we obtain some effective techniques of text feature selection, and using these concept we extend our work in the following manner.

1. A new feature selection technique using GI, IG and POS.
2. Extending the SVM classifier for multiclass class classification
3. Obtaining the dataset for hate speech

REFERENCES

1. T. Dodds, "**Reporting with WhatsApp: Mobile Chat Applications**' Impact on Journalistic Practices", Digital Journalism, 2019, VOL. 7, NO. 6, 725–745
2. R. O. Spínola, N.Zazworka, A.Vetro, F. Shull, C. Seaman, "**Understanding automated and humanbased technical debt identification approaches-a two-phasestudy**", Journal of the Brazilian Computer Society (2019) 25:5
3. U.Khadam, M. M.Iqbal, M. A.Azam, S. Khalid, S. Rho, N.Chilamkrti, "**Digital Watermarking Technique for Text Document Protection Using Data Mining Analysis**", VOLUME 7, 2019, 2169-3536, 2019 IEEE.
4. R.Agrawal, M.Batra, "**A Detailed Study on Text Mining Techniques**", International Journal of Soft Computing and Engineering (IJSCE),ISSN: 2231-2307, Volume-2, Issue-6, January 2013
5. A. Abbasi, H. Chen, Arab Salem, "**Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums**", © 2007 ACM 1073-0516/01/0300-0034 $5.00
6. S. Tan, J. Zhang, **"An empirical study of sentiment analysis for Chinese documents**", Expert Systems with Applications xxx (2007) xxx–xxx
7. S. Wang, D. Li, X. Song, Y. Wei, H. Li, "**A feature selection method based on improved fisher's discriminant ratio for text sentiment classification**", Expert Systems with Applications, 38, (2011), 8696-8702
8. A. Duric, F. Song, "**Feature Selection for Sentiment Analysis Based on Content and Syntax Models",**Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, ACL-HLT 2011, pages 96–103, 24 June, 2011,
9. A. Sharma, S. Dey, "**Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis"**, Special Issue of International Journal of Computer Applications (0975 – 8887) on Advanced Computing and Communication Technologies for HPC Applications - June 2012
10. E. Haddia, X. Liua, Y. Shib, "**The Role of Text Pre-processing in Sentiment Analysis**", 1877-0509 © 2013 Published by Elsevier
11. S. Poria, E. Cambria, A. Gelbukh, "**Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-Level Multimodal Sentiment Analysis**", Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 2539–2544, 17-21 September 2015. c 2015
12. A. S. Manek, P. D. Shenoy, M C. Mohan, K R. Venugopal,"**Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier**", World Wide Web, Springer Science+Business Media New York 2016 DOI 10.1007/s11280-015-0381-x
13. L. Zheng, H. Wang, S. Gao, "**Sentimental feature selection for sentiment analysis of Chinese online reviews**", Int. J. Mach. Learn. & Cyber. © Springer-Verlag Berlin Heidelberg 2015, DOI 10.1007/s13042-015-0347-4
14. P. Ruangkanokmas, T. Achalakul, K. Akkarajitsakul, "**Deep Belief Networks with Feature Selection for Sentiment Classification**", 2016 7th International Conference on Intelligent Systems, Modelling and Simulation, 2166-0670/16 $31.00 © 2016 IEEE
15. B. M. Jadav, V. B. Vaghela, "**Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis**", International Journal of Computer Applications (0975 – 8887), Volume 146 – No.13, July 2016
16. A. Nasser, K. Dinçer, H. Sever, "**Investigation of the Feature Selection Problem for Sentiment Analysis in Arabic Language**", pp. 41–54; rec. 2016-02-08; acc.

2016-03-21 41 Research in Computing Science 110 (2016)

17. S. Poria, I. Chaturvedi, E. Cambria, A. Hussain, **"Convolutional MKL Based Multimodal Emotion Recognition and Sentiment Analysis**", © 2016 IEEE Personal use of this material is permitted
18. Singh S., Verma, S., Tiwari, A., Tiwari, A,.**"A novel way to Classify Passenger data using Naïve Bayes algorithm"**, In Proceedings on 2016 2nd International Conference Next Generation Computing Technologies, NGCT 2016 (pp. 312– 316) Institute of Electrical & Electronics Engineers. https:// d o i .o r g / 1 0 . 1 1 0 9 / NGCT.2016.7877433
19. Singh, S., Verma, S. K., & Tiwari, A.,**"A novel approach for finding crucial node using ELECTRE method**". International Journal of Modern Physics B, 34(09), 2050076. doi:10.1142/s0217979220500769
20. Singh, S., Verma, S. K., & Tiwari, A.,**"A novel method for destabilization of terrorist network",** Modern Physics Letters B, 2050298. doi:10.1142/s021798492050298x
21. Hasan, A. M.,**"Knowledge-Based Semantic Relatedness measure using Semantic features",** International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 914-924. doi:10.30534/ijatcse/2020/02922020
22. Mohana P, Perumal K., **"On Feature Selection Stability and Privacy Preserving Data Mining: A Data Perspective",** International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1218-1233. doi:10.30534/ijatcse/2020/50922020.