

¹Devendra Kumar²Faiyaz Ahamad

Opinion Extraction using Hybrid Learning Algorithm with Feature Set Optimization Approach



Abstract: - Evolution in engineering and technology added large size of data storage and transmission through the web application over the internet. This huge amount of data primarily used for exchange of information in between users and devices and in secondary aspects it has utilization as feedback, ratings and reviews that is supporting in generation of useful information of products, services, incidents etc. The data as opinion, feedback, view & suggestion is explored, organized & analyzed for selection of appropriate options. Sentiment analysis using the opinion extraction is a challenging task that is based on feature extraction and the concepts of Natural Language Processing that is applied in identification of the opinions of a user in terms of positive, neutral or negative ratings hidden in the form of comments typed as the text. Presently many data-processing based feature evaluation techniques for opinion extraction are used for solving the issues faced under sentiment classification applications. This article is based on development and application of algorithms for opinion extraction from text data available on web resources by K-Nearest Neighbor (KNN), Support vector machine (SVM) and hybrid of both named as SVM+KNN for classification of multi-label opinions from extracted text from review data of Twitter and Amazon. The performance of all the classification models (KNN, SVM and SVM+KNN) on both datasets is evaluated in terms of different parameters.

Keywords: Feature Extraction, KNN, SVM, Opinion Mining, Sentiment Classification, Text Mining.

I. INTRODUCTION

Social networking sites and online shopping platforms are performing as a huge and global platform for public to freely express the opinions, feelings, discussions on wide number of social, financial and technical issues. Users are free to provide reviews & feedbacks on products, services on online social networking platforms like instagram, twitter, facebook and e-commerce sites like amazon, flipkart etc. Ratings given by the customer in terms of stars or number of likes or shares along-with the comments or reviews on these web applications in form of text helps the commercial organizations and public to know about quality of goods and services. Presently digital marketing circulated by all the small or big organization and all of them are facing a transition from offline to online marketing domain; it has passed through a challenge of handling large amount of unstructured data. It is becoming an important target to transform unstructured data in helpful patterns that may be integrated to task of drawing useful decisions about product and services [1]. Commercial service circulates advertise or survey on social networking sites (Twitter, Instagram & Facebook) for collecting user opinions in efficient manner [2]. User opinion helps in estimating the level of satisfaction & tracking the competition. It is also useful to public in survey of goods and service available in market. Opinion extraction used to perceive the market trends in efficient manner that is used in updating the quality [3, 4]. The service providers, manufacturers, distributor and consumers views undergo through impact of response extracted from opinion of users. These opinions finally impact the stock market and economy of a country. Social networking platforms are growing rapidly & getting easier in terms of interaction of users. As a result, the opinion of public is spreading swiftly over the world and changing the views of followers significantly [1]. Hence decisions extracted from opinion as ratings or sentimental bringing transformation in associated planning of organizations [5].

Services provided by different sectors are reflected on comments publically available on Twitter, facebook etc. and works as significant information about service provided to citizens by the professional systems. Opinions of public shared as feeling or thoughts on several instances that has prevalent impact on social life [6]. For example, in health sector the patients sometimes face isolation from society and undergoes through seriously transformation in mental state that attains the stage of depression. In such cases experts may use automated extraction of sentiment using the text-based information shared by patient on web applications for systematic analysis [7]. Social networking sites openly share the opinions as the text data in terms of sentiments, emotions and feelings in the user's posts and if

¹Department of Computer Science and Engineering, Integral University, Lucknow, Uttar Pradesh, India

¹devendrakumar17@gmail.com

² Department of Computer Science and Engineering, Integral University, Lucknow, Uttar Pradesh, India

Copyright © JES 2024 on-line: journal.esrgroups.org

somebody has faced issues from a particular event, service or product may be tracked by estimating his level of satisfaction state.

Opinion extraction based on text data is useful because user comment plays also plays an important role in academic sector. The teacher performance may be estimated by in terms of enthusiasm, dedication & talent using the text-based comments provided by student. It is an effective approach for improving the teaching skills on online and offline class platform [8]. Unstructured text data-based feedback is difficult to observe and a lot of problems are faced in drawing useful conclusions manually. The automated approach for drawing sentiments from text data using opinion extraction by machine learning methods assist the domain covered under academic sectors and healthcare sectors for opting corrective steps. The social media platforms like Twitter & Facebook know incorporating the promotions activities associated to marketing and advertising for different organizations. Consumer and service provider both follow the considerable online analysis for knowing the potential of concerned service or product. Blogs and forums followed by users of similar domain of interests & help in assessment of the standard of various schemes circulated about products and services under different types of promotion strategy. In such application the opinion extraction-based analysis in terms of features extracted from unstructured text data helpful everyone to select the best option [9, 10]. Opinion extraction analysis is incorporating different methodologies and techniques on the basis of artificial intelligence and optimization schemes. All schemes have particular advantages as well as drawbacks. These feature-based techniques for opinion extraction using artificial intelligence and machine learning faces significant challenges that are including the ridicule, context, statements that are conveying emotions, spread of lexical, slang and syntax-based ambiguity [11]. Since no standard rules are set up in communication to express the opinion. Some methods focus on effect of post, some analyze sentiments, and some are performing on analysis over logical structure of message. Such variety of challenges in opinion extraction process under natural language processing requires an approach that gives performs efficiently. Remaining of the paper describe as follows: section II involves the related work, section III involves the method and material, section IV contains result and discussion and section V contains conclusion.

II. RELATED WORK

Worked on Facebook comments performed for multilingual texts using Machine Learning (ML) the F1-score parameter shows high value but the performance on non-English text was low. Author suggested for use of hybrid scheme to avoid this limitation [12]. An approach based on Lexicon was followed on text data in Chinese language-based opinion extraction on feedback over online website of poetry. Further improvement in accuracy is suggested using machine learning techniques [13]. A CNN and Bi-LSTM based ML approach proposed for achieving high F1-score using small dataset to focus on the problem related to feature extraction issues. The improvement may be further incorporated on analysis with different deep learning algorithm [14]. An emotion extraction work was performed for ISEAR dataset using hybrid learning model but the achieved accuracy was low due to poor classified of neutral emotion. Further improvements may expect on using different deep learning approaches [15]. A deep machine learning scheme proposed for opinion extraction for data collected from WhatsApp Chat. The performance in terms accuracy was medium due to presence of emojis in text [16]. Naive Bayes (NB) and KNN applied on extraction of sentiment on Twitter Dataset. The NB performed better than KNN but it had a drawback of improper working for unsupervised learning mechanisms [17]. A machine learning based approach followed on SVM on ISEAR dataset to improve performance in opinion extraction by natural language processing [18]. A machine learning algorithm developed for Emotion Line dataset the demonstrated result shows high F1-score value for extraction on two different datasets but data considered in this work was small [19].

III. METHODOLOGY

In this proposed work an implementation is performed for extracting information from text data to generate features in classification of two datasets belongs to Twitter and Amazon. The datasets consist of text-based comments as opinion of users. In the twitter dataset opinion expressed as four different types of sentiments and in the Amazon dataset the reviews are associated with star rating. Both datasets are large I size (8040 posts (text message) of different users. This dataset is first of all passed through the preprocessing steps that include the tokenization of posts that breaks the sentence into individual words [27]. The second step includes removing the punctuation from the tokenized words from each post. In the third step "bag of words created". The Bag contains the vocabulary that has unique words and counts of the number of times a word is repeated in each post [28]. After creating the word, the next step under data preprocessing involves the process of removing the stop words. The term stop words

represent the words as token like 'to', 'and', 'of' etc. They do not reflect any kind of sentiments or emotions. After removal of stop words the words that are rarely used are also removed. This step is called as removal of infrequent words [29].

After the preprocessing of data finally a bag is created that consist of those words that frequently used in the data set. Thus, the Term frequency of each word is calculated (TF) and Inverse document frequency (IDF) is also calculated and saved in a sparse matrix. In this way TF-IDF feature array is generated.

A. Data Processing: After generation of TF-IDF feature the dataset is used for development of machine learning model using KNN. The dataset is split into training and testing sets. About 80% of data used for training purpose [25]. The training is performed by random selection of feature vectors at different combination of number of top words and word frequency. Several K-NN models are developed with different training sets and performance is evaluated in terms of model accuracy. The text data as comments are imported from CSV file to MATLAB software. The twitter data has labels as 4 types of sentiments {happiness, relief, enthusiasm, surprise} the data size: 8040 tweets. The AmazonCellPhone.csv file has class label as 5 different user rating [1, 2, 3, 4, 5] the data size is 20000 reviews. After importing the data, the step applied for tokenization of the text document. In this step the document represented as a collection of words (also known as tokens).

B. Remove punctuation from the tokenized documents. After removing punctuation next step of create **Bag-of-words** is followed. A bag-of-words is a data structure stores the calculated term-frequency. It records the number of times that words appear in each document of a collection.

Number of tweets=Number of documents = [NumDocuments]_{1x1}=D

Number of unique words in all documents: [NumWords]_{1x1}=W

Collection of all unique words =[Vocabulary]_{1xW}

Counts_{DxW} = Number of counts a word (term) repeated in each document.

Table 1: Count of word repeated in a document

Document 1	Document 2	Document 3	Document 4	Document 5
<i>finally made it to phoenix! I am home.</i>	<i>finally made it to the QC...happy mother's day! got my car!!</i>	<i>Finally off work!!!</i>	<i>Finally on my way home</i>	<i>Finally, out of school for today! HAD TO SIT IN LIBRARY FOR LONG TIME</i>
8 words in	12 words in document	3 words	5 words in	15 words in document
1. (1,1)	1. (1,1) 1	1. (1,	1. (1,	1. (1,1) 1
2. (1,2)	2. (1,2) 1	2. (1,	2. (1,	2. (1,4) 1
3. (1,3)	3. (1,3) 1	3. (1,	3. (1,	3. (1,21) 1
4. (1,4)	4. (1,4) 1		4. (1,	4. (1,22) 1
5. (1,5)	5. (1,9) 1		5. (1,	5. (1,23) 1
6. (1,6)	6. (1,10) 1			6. (1,24) 2
7. (1,7)	7. (1,11) 1			7. (1,25) 1
8. (1,8)	8. (1,12) 1			8. (1,26) 1
	9. (1,13) 1			9. (1,27) 1
	10. (1,14) 1			10. (1,28) 1
	11. (1,15) 1			11. (1,29) 1
	12. (1,16) 1			12. (1,30) 1
				13. (1,31) 1
				14. (1,32) 1

From the bag of words, the "stop words" are removed. Words like "a", "and", "to", and "the" (known as stop words) can add noise to data. This step is applied to remove stop words before analysis. After this the Infrequent Words are removed. Remove words with low counts from bag-of-words. Words with counts less than 'T' are removed from total unique words. Here T is called as number of top words.

C. Hybrid Learning: The SVM+KNN as hybrid learner used in this paper applied Error correcting output codes (ECOC) for performing classification task. It is motivated by coding theory where transmitted information is encoded by binary strings. Presently this method upgraded for handling learning problem to increase ensemble diversity as hybrid learner. In this SVM_KNN classifier, each class assigned to codeword and a L learner trained in hybrid as binary classifiers constructed by columns of ECOC matrix such that rows represents class codewords and column represents partition of the dataset by classes merging of similar bit value. Decoding is referred as classification perform by matching codeword predicted by L with the class codeword nearest in Hamming distance. This SVM+KNN is generalized form of one-vs-one and one-vs-all classifier and as a hybrid algorithm, it effectively performs classifiers independent of errors in random sampling.

Algorithm

Set data distribution: Training_{percent}
 Set number of top words: N_{top}
 Set number for frequency of words: $N_{frequency}$

1: Data preparation

```
import data
read data file
alltweets <= extract text data
allsentiments <= extract sentiment labels
data1 <= break text sentence into token of words (tokenization)
data2 <= erase Punctuation from data1
```

2: Build a Bag of Words containing all tokenized tweets (ignore punctuation)

```
 $N_W$  <= count total number of words
 $N_D$  <= count total number of documents
 $V_{1 \times N_W}$  <= create vocabulary of unique words
Counts $N_W \times N_D$  <= assign address to each unique word in each document
bag <= form bag of Words of data2 as {Number of words, number of documents, vocabulary, counts};
bag1 <= remove stop words from bag of words "bag"
bag2 <= remove infrequent words that are repeated less than  $N_{frequency}$  time.
topwords <= extract top words having repetition frequency >  $N_{top}$ 
```

3: Features and labels

```
 $TF_{t,d}$  <= calculate number of occurrences of term t in document d
 $DF_t$  <= Calculate number of documents containing the term t.
 $W_{t,d}$  <= weight of term t in document d <=  $TF_{t,d} * \log(N/DF_t)$ 
 $tf(t,d)$  <= calculate term frequency <=  $(TF(t,d))/\sum(TF(t,d))$ 
 $idf(t,d)$  <=  $\log(N_D/|\{d : d \in D \text{ and } t \in T\}|)$ 
 $M_1$  <=  $tfidf(t,d,D)$  <=  $tf(t,d) * idf(t,d)$ 
```

4: Data Distribution

```
define: m <= percent train <= [80% ,70% or 60%] percent training data
n <= size of all tweets % size of all data in the document
n-m <= size of testing Tweets
%% Create a feature matrix for training by selecting the m rows of the TF-IDF matrix and all columns
m1 <= generate m random permutation of tweet address in between 1 to n integer value
trainx: training_features <= randomly select m tfidf value as features from bag  $M_1$  w.r.t m1 address locations
trainy: training_labels <= create a corresponding label vector with the first m from allsentiments vector
create one feature matrix for testing by selecting all rows of the TF-IDF matrix after row n (i.e. the remaining rows)
test_x: testing_features >= Bag  $M_1$  (from m to n data at remaining random address of m1)
%% from m to n data at remaining random address of m1 create a corresponding sentiment class label vector
test_y <= testing_labels <= allsentimentsyy
```

5: K-Nearest Neighbor

```
%% call KNN function fitKNN and generate KNN based classification model.
knnmodel <= fitknn(trainx, trainy)
predictionSA <= predict output labels using knnmodel for testing data test
```

6: Call SVM function fit_{SVM} and generate SVM based classification model.

SVMmodel <= fits_{SVM}(train_x,train_y)
 predictions_B <= predict output labels using SVMmodel for testing data test_x

7: Call SVM_KNN hybrid function fit_SVMKNN and generate hybrid SVM_KNN based classification model.

SVMmodel <= fits_{SVMKNN}(train_x, train_y)
 predictions_C <= predict output labels using SVMmodel for testing data test_x
 label_1<=find test label = 'relief'
 label_2<=find test label = 'surprise'
 label_3<=find test label = 'enthusiasm'
 label_4<=find test label = 'happiness'
 Confusionchart <= generate confusion matrix from {label_1, label_2, label_3, label_4}

[TP, TN, FP, FN] <=calculate true positive, true negative, false positive, false negative value from confusion matrix

precision<=TP/(TP+FP)

recall<=TP/(TP+FN)

f1score<=(2*precision*recall)/(precision+recall)

accuracy<=(TP+TN)/(TP+TN+FP+FN)

Go to Step 1 and repeat step 1 to step 48

Steps involved in development of opinion extraction from text data:

Step 1: Set data distribution: Training_{percent}

Step 2: Set number of top words :N_{top}

Step 3: Set number for frequency of word :N_{frequency}

Step 4: Data preparation

Step 5: Build a Bag of Words containing all tokenized tweets (ignore punctuation)

Step 6: Calculate Features and define sentiment labels

Step 7: Perform Data Distribution

Step 8: Create a feature matrix for training data

Step 9: Create a corresponding sentiment class label vector from remaining random address

Step 10: Call KNN function fit_{KNN} and generate KNN based classification model.

Step 11: Call SVM function fits_{SVM} and generate SVM based classification model.

Step 12: Call SVM_KNN hybrid function to generate hybrid SVM_KNN based classification model.

Step 13: Predict output labels using SVMmodel for testing data test_x

Step 14: Generate confusion matrix from {label_1, label_2, label_3, label_4}

Step 15: Calculate true positive, true negative, false positive, false negative value from confusion matrix

Step 16: Calculate precision, recall, F1score, accuracy

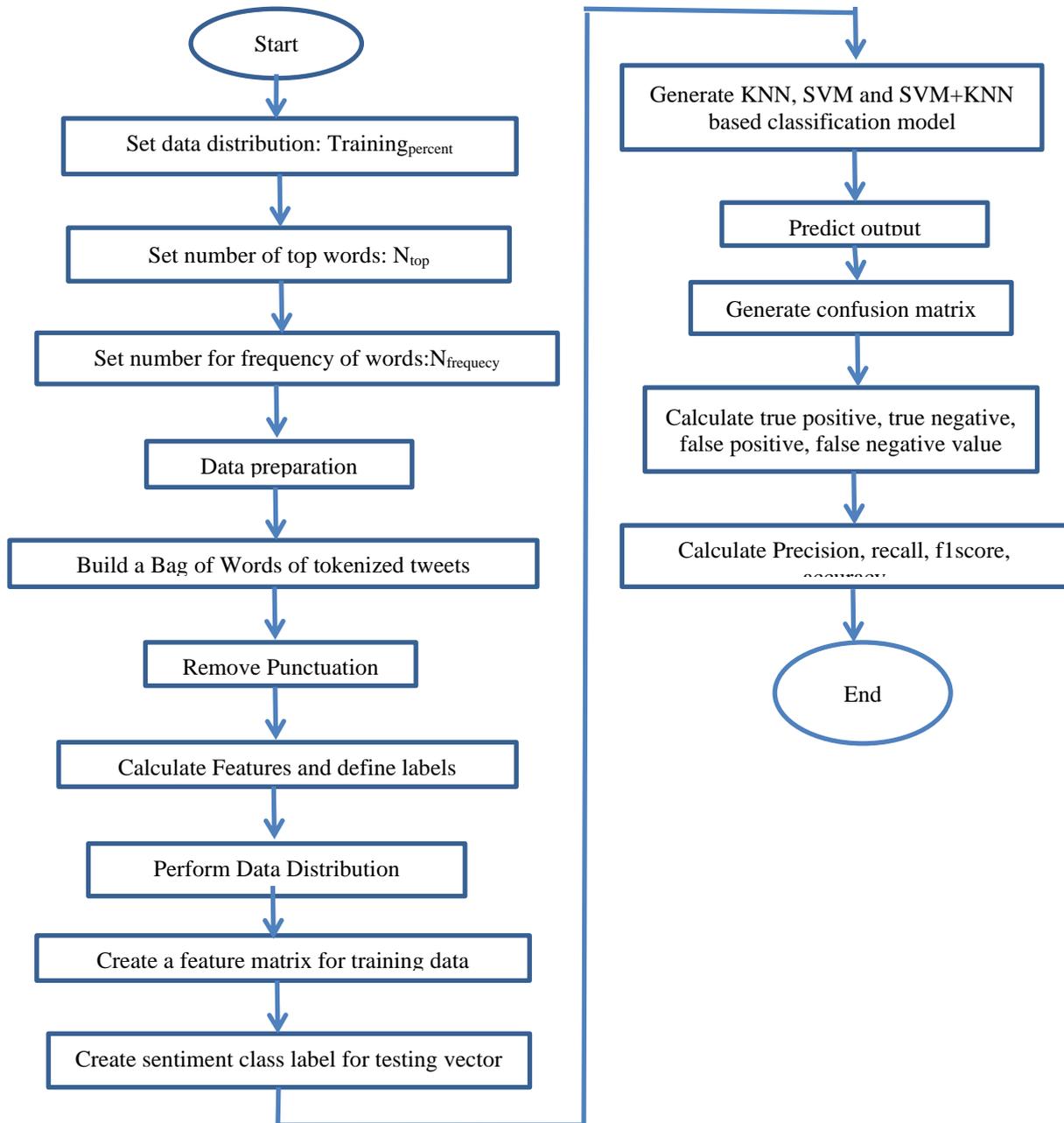


Figure 1: Flowchart of proposed work

IV. RESULTS

Before The simulation is performed on MATLAB software using Natural Language processing and Machine learning toolbox. The algorithm run for classification using KNN, SVM and SVM+KNN for different combinations of number of top words and word frequency.

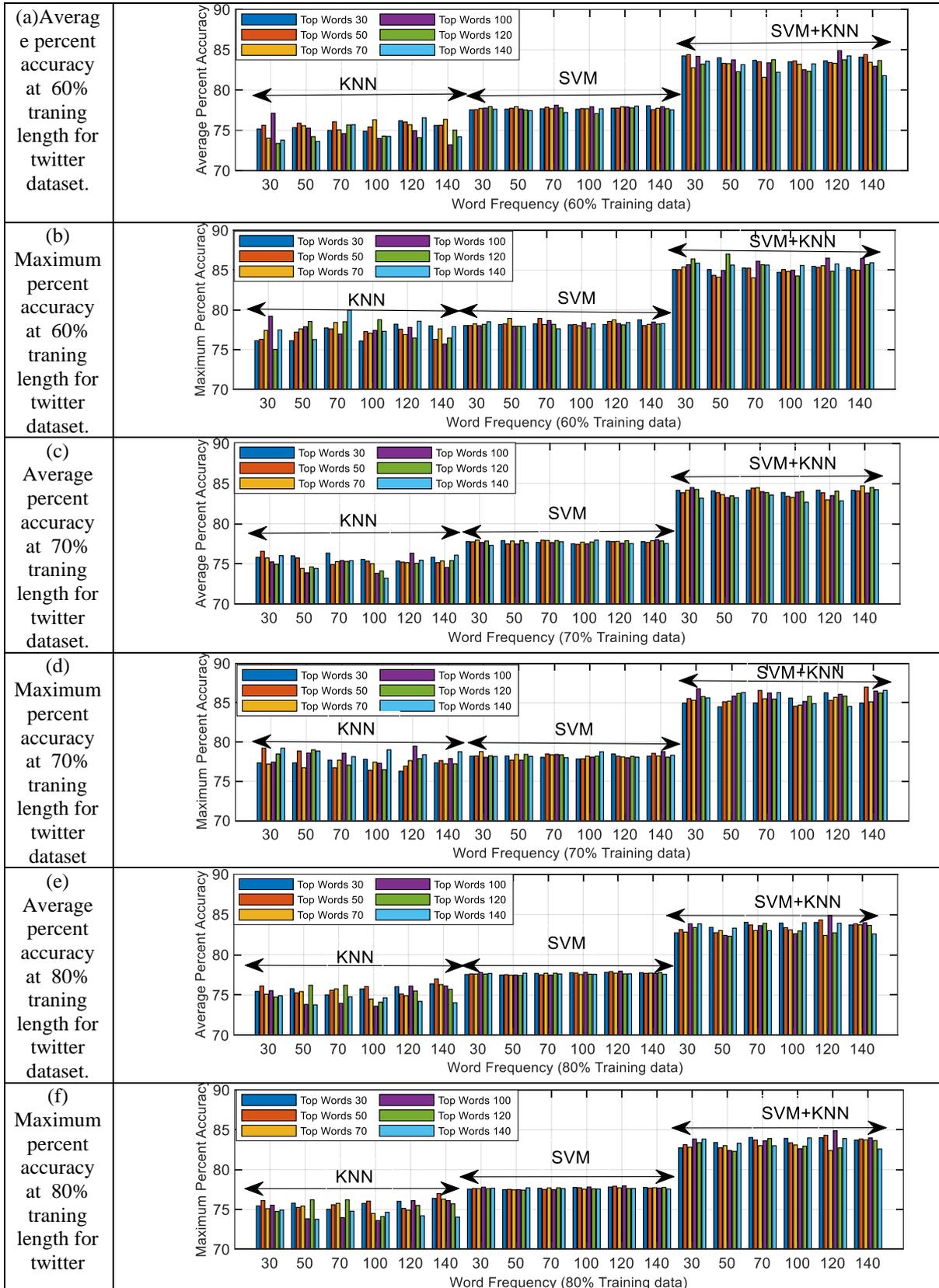


Figure 2 Results in terms of average percent accuracy using KNN, SVM and SVM+KNN for twitter dataset.

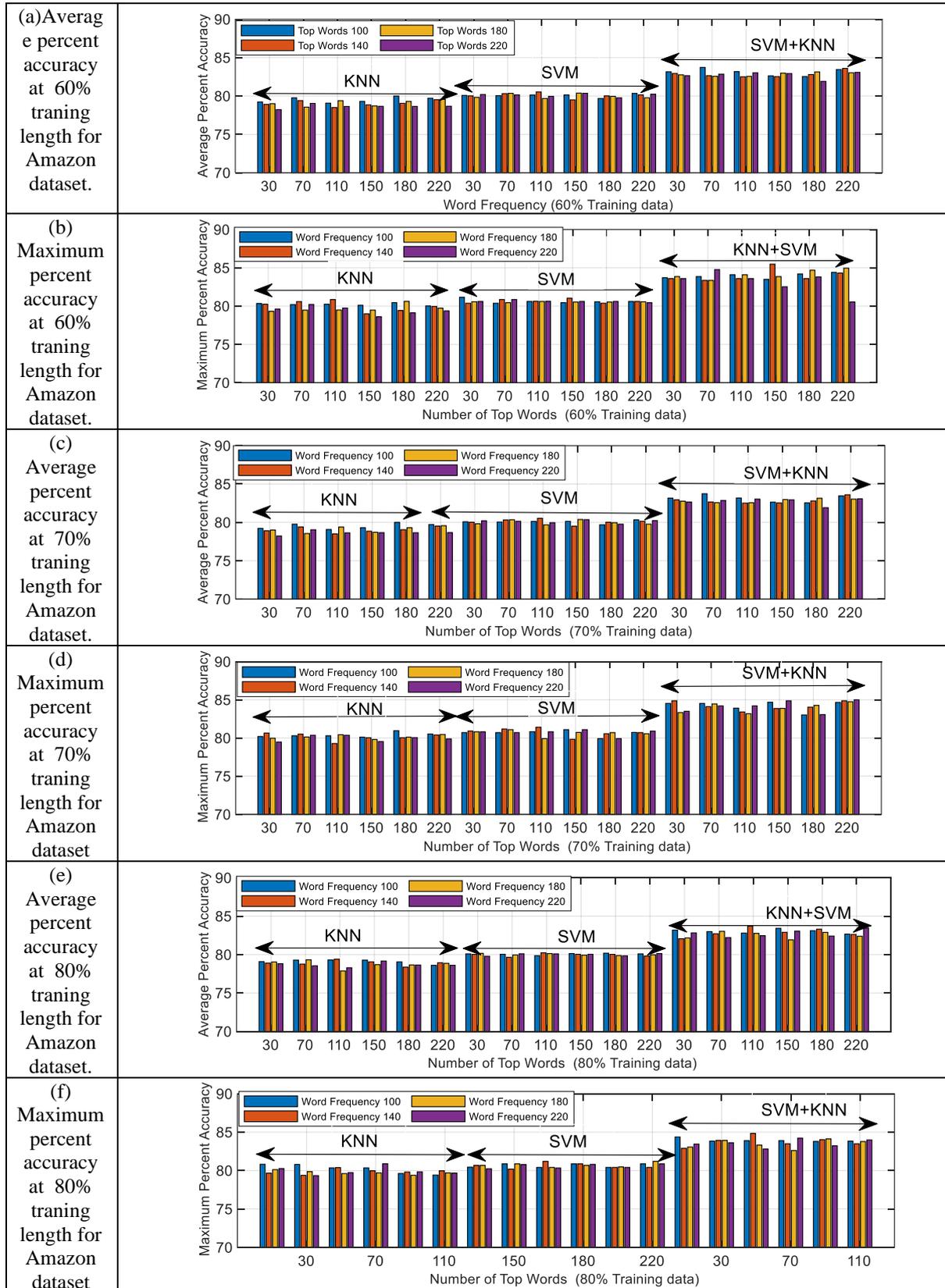


Figure 3 Results in terms of average percent accuracy using KNN, SVM and SVM+KNN for Amazon dataset.

Table 2 Maximum and average percent accuracy for Twitter and Amazon dataset

Twitter Dataset				Amazon Dataset			
Training Length	Max			Training Length	Max		
	KNN	SVM	KNN&SVM		KNN	SVM	KNN&SVM
4824	0.7998 (140-70)	0.7894 (50-70)	0.8701 (120-50)	12000	0.8086 (140-110)	0.8115 (100-30)	0.8548 (140-150)
5628	0.7947 (100-120)	0.7878 (70-30)	0.8699 (50-140)	14000	0.8096 (100-180)	0.8141 (140-110)	0.8501 (220-220)
6432	0.7993 (50-30)	0.7864 (100-30)	0.8699 (120-50)	16000	0.8088 (220-150)	0.812 (140-110)	0.8484 (140-110)
Training Length	Average			Training Length	Average		
	KNN	SVM	KNN&SVM		KNN	SVM	KNN&SVM
4824	0.7510	0.7772	0.8340	12000	0.7888	0.8015	0.8290
5628	0.7523	0.7774	0.8385	14000	0.7907	0.8007	0.8289
6432	0.7527	0.7766	0.8339	16000	0.7886	0.8001	0.8279

Table 3 Performance in terms of Precision, Recall and F1score for Twitter and Amazon dataset

	Twitter data Set				Amazon Data Set		
	Algorithm				Algorithm		
	KNN	SVM	SVM+KNN		KNN	SVM	SVM+KNN
Top words & word frequency	140-70	50-70	120-50		100-180	140-110	140-150
Training Length	60%	60%	60%		70%	70%	60%
Precision	0.8061	0.7759	0.8462		0.9303	0.9265	0.9469
Recall	0.8375	0.8625	0.903		0.8981	0.9071	0.9183
F1 Score	0.8215	0.8169	0.8736		0.9139	0.9167	0.9324

Table 4 Comparison of SVM, KNN, SVM+KNN Model with State-of-the-Art Methods

Technique	Precision	Recall	F1-measure	Accuracy
SABSA [26]	0.858	0.839	0.8485	0.837
SentiVec [21]	0.877	0.858	0.8675	0.861
Ngram +TF-IDF + SVM [23]	0.866	0.846	0.856	0.844
SEML [25]	0.854	0.837	0.8455	0.838
MTMVN [24]	0.817	0.789	0.803	0.792
SVM+KNN (Twitter Dataset)	0.846	0.903	0.873	0.869
SVM+KNN (Amazon Dataset)	0.946	0.918	0.932	0.854

V. CONCLUSIONS

The SVM+KNN based hybrid scheme gave highest accuracy with respect to applying only SVM or KNN only. High accuracy for classification based on opinion extraction from posts of twitter or reviews of user on Amazon are used to for generating TF-IDF based feature. The highest accuracy is 87 % for twitter dataset and 85% on Amazon data set is observed for task of deciding opinion in terms of sentiments/rating by using textual content. The analysis is focusing the ambiguity covered in observing sentiments/rating is the main concerning challenge that is reduced by using hybrid learning scheme. The results are reflecting potential of natural language processing for opinion extraction in the post (Tweets/reviews) as text data. The approach may help to provide valuable applications on dataset processing prior to classification. In future advanced refinement steps for dataset may be used for further improving the performance. Detection of inaccurate or mislabeled classes may be focused for further advancements for understanding linguistic and sentiment analysis.

ACKNOWLEDGMENT

We would like to thank the university's research departments for providing the Manuscript number IU/R&D/2024-MCN0002660 according to their guidelines. This identification makes it easier to track and communicate about our research as it moves through the publication process. We would also like to express our appreciation to everyone who helped to build this work.

REFERENCES

- [1] Ahire, V., & Borse, S. (2022). Emotion detection from social media using machine learning techniques: a survey. In *Applied Information Processing Systems: Proceedings of ICCT 2021* (pp. 83-92). Springer Singapore.
- [2] Agbehadji, I. E., & Ijadeniyi, A. (2021). Approach to sentiment analysis and business communication on social media. *Bio-inspired Algorithms for Data Streaming and Visualization, Big Data Management, and Fog Computing*, 169-193.
- [3] Jang, H. J., Sim, J., Lee, Y., & Kwon, O. (2013). Deep sentiment analysis: Mining the causality between personality-value-attitude for analyzing business ads in social media. *Expert Systems with applications*, 40(18), 7492-7503.
- [4] Al Ajrawi, S., Agrawal, A., Mangal, H., Putluri, K., Reid, B., Hanna, G., & Sarkar, M. (2021). WITHDRAWN: Evaluating business Yelp's star ratings using sentiment analysis.
- [5] Anurag, R. S. (2020). Load Forecasting by using ANFIS. *International Journal of Research and Development in Applied Science and Engineering*, 20(1).
- [6] Garcia, K., & Berton, L. (2021). Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied soft computing*, 101, 107057.
- [7] Singh, M., Jakhar, A. K., & Pandey, S. (2021). Sentiment analysis on the impact of coronavirus in social life using the BERT model. *Social Network Analysis and Mining*, 11(1), 33.
- [8] Sangeetha, K., & Prabha, D. (2021). Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM. *Journal of Ambient Intelligence and Humanized Computing*, 12, 4117-4126.
- [9] Arcane Rao PN, Baglodi K (2017) Role of sentiment analysis in education sector in the era of big data: a survey. *Int J Latest Trends Eng Technol* 22–24
- [10] Ahmed, W. M. (2020). Stock market reactions to domestic sentiment: Panel CS-ARDL evidence. *Research in International Business and Finance*, 54, 101240.
- [11] Khan, M. U., & Ahamad, F. (2024). An Affective Framework for Multimodal Sentiment Analysis to Navigate Emotional Terrains. *Telematique*, 23(01), 70-83.
- [12] Seman, N., & Razmi, N. A. (2020). Machine learning-based technique for big data sentiments extraction. *IAES International Journal of Artificial Intelligence*, 9(3), 473.
- [13] Ahmad, Z., Jindal, R., Ekbal, A., & Bhattacharyya, P. (2020). Borrow from rich cousin: transfer learning for emotion detection using cross lingual embedding. *Expert Systems with Applications*, 139, 112851.
- [14] Wlodarczyk, P., Ally, M., & Soar, J. (2015). Opinion mining in social big data. *Available at SSRN 2565426*.
- [15] El Alaoui, I., Gahi, Y., Messoussi, R., Chaabi, Y., Todoskoff, A., & Kobi, A. (2018). A novel adaptable approach for sentiment analysis on big social data. *Journal of Big Data*, 5(1), 1-18.
- [16] Tsapatoulis, N., & Djouvas, C. (2019). Opinion mining from social media short texts: Does collective intelligence beat deep learning? *Frontiers in Robotics and AI*, 5, 138.
- [17] Saidi, W., El Abderahmani, A., & Satori, K. (2021). New approach of opinion analysis from big social data environment using a supervised machine learning algorithm. In *E3S Web of Conferences* (Vol. 319, p. 01037). EDP Sciences.
- [18] Khan, K., Baharudin, B., Khan, A., & Ullah, A. (2014). Mining opinion components from unstructured reviews: A review. *Journal of King Saud University-Computer and Information Sciences*, 26(3), 258-275.
- [19] Wu, C., Wu, F., Wu, S., Yuan, Z., & Huang, Y. (2018). A hybrid unsupervised method for aspect term and opinion target extraction. *Knowledge-Based Systems*, 148, 66-73.
- [20] Nguyen, H. D., Khan, M. S., Kaegi, N., Ho, S. S., Moore, J., Borys, L., & Lavalva, L. (2021). *Ensemble learning using error correcting output codes: New classification error bounds*. In 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 719-723). IEEE.
- [21] Zhu L, Li W, Shi Y, Guo K. SentiVec: learning sentiment-context vector via kernel optimization function for sentiment analysis. *IEEE Trans Neural Netw Learn Syst*. 2021;32(6):2561–72. <https://doi.org/10.1109/tnnls.2020.3006531>.
- [22] Schouten K, van der Weijde O, Frasinca F, Dekker R. Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. *IEEE Trans Cybern*. 2018;48(4):1263–75. <https://doi.org/10.1109/tcyb.2017.2688801>.

- [23] Ayyub K, Iqbal S, Munir EU, Nisar MW, Abbasi M. Exploring diverse features for sentiment quantification using machine learning algorithms. *IEEE Access*. 2020;8:142819–31. <https://doi.org/10.1109/access.2020.3011202>
- [24] Bie Y, Yang Y. A multitask multiview neural network for end-to-end aspect-based sentiment analysis. *Big Data Mining Analytics*. 2021;4(3):195–207. <https://doi.org/10.26599/bdma.2021.9020003>
- [25] Li N, Chow CY, Zhang JD. SEML: a semi-supervised multi-task learning framework for aspect-based sentiment analysis. *IEEE Access*. 2020;8:189287–97
- [26] Schouten, K., Van Der Weijde, O., Frasincar, F., & Dekker, R. (2017). Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. *IEEE transactions on cybernetics*, 48(4), 1263-1275.
- [27] Khan, M. U., & Ahamad, F. (2024). An Affective Framework for Multimodal Sentiment Analysis to Navigate Emotional Terrains. *Telematique*, 23(01), 70-83.
- [28] Ahamad, F., Darbari, M., & Asthana, R. (2016). Service Mechanism for Diagnosis of Respiratory Disorder Severity Using Fuzzy Logic for Clinical Decision Support System. In *Emerging Research in Computing, Information, Communication and Applications: ERCICA 2015, Volume 3* (pp. 309-317). Springer Singapore.
- [29] Kumar, D., & Ahamad, F. (2024). Application of Machine Learning Algorithm for Optimal Model Design for Opinion Extraction. *Telematique*, 23(01), 215-227.