OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

SVM FOR CUSTOMER REVIEW ASSESSMENT

Suhita Kapai [1], Smita Tanmor[2], Jyoti Gaikwad[3], Shubhangi Jagtap[4], Prof. Ganesh Bandal[5]

Student, G. H. Raisoni College of Engineering and Management, Pune India[1 2 3 4]

Assistance Professor G. H. Raisoni College of Engineering and Management, Pune, India[5]

suhitakapai@gmail.com [1], smita.tanmor@gmail.com [2], jyotigaikwad239@gmail. com [3] ,shubhangijagtap4@gmail.com [4], ganesh.bandal@raisoni.net[5]

Abstract: **This paper presents the concept of using sentimental analysis for discovering the perspective of the customer towards a product or service. Text categorization and Data Mining work parallel for categorization of the customer review. The Sentimental analysis which is component of the NLP(Natural language processing) is used to discover if the text review is established according to the thoughts of the person and what facts does the text provide i.e., is the text liked, disliked or neutral. The machine learning technique like SVM (Support Vector Machine) and KNN are components of the supervised learning technique that can be used to discover if the text is liked, disliked or neutral. The model build time, searching time, accuracy and memory used is prominent in SVM than KNN.**

Keywords *– Text categorization, Sentimental analysis, SVM, supervised learning, KNN.*

----------∴∴∴----------

I INTRODUCTION

Text categorization [8] is very important in today's fast growing world. The text data has to be properly arranged and managed. Thus text categorization helps in the data management. The SVM (Support Vector Machine) is used to classify the text. In SVM the text is represented into the vector format. The SVM is used to distinguish the sentiments [2][6] of the people that are written in form of their reviews[9] as positive, negative and neutral. As it is supervised the names of the classes are predefined. SVMs don't depend on the number of features so they are favorable for large feature sets. SVMs can be applied for classification, regression, and outlier detection and clustering. This paper mostly targets the classification of the data. The main step in SVM is to discover the best line separator i.e., the hyper plane between the two categories of data. The hyper plane has to be built in such a way that there is the maximum margin between the hyper plane[11] and the categories of the classes i.e., the support points should be in close proximity to the dividing line. Whenever there is an entry of a new support point then it can be mapped to the corresponding categorical space with respect to the decision boundary and the predefined support points.

II PROBLEM STATEMENT

Knowing whether the sentiment[10] is positive or negative from the plaintext was a scrape because the machine cannot distinguish it, hence there arises a need for machine learning i.e. SVM (Support Vector Machine) algorithm[1], which will find a line of parting between the positive, negative and neutral [7]response from the text.

III METHODOLOGY

The following are the parts of the SVM framework[3]:

Data collection: Collecting data from reviews which is in disorganized form.

Text Preparation: Before analysis removing nonrelevant and non -textual contents[4].

Sentiment Detection: Here only the sentences with the subjective expression are included and the other expressions are discarded.

Classification: These subjective expressions are classified as positive and negative. It can also have multiple other points.

Presentation of output: Structured output is represented in form of pie charts, line graphs, and other graphical representation.

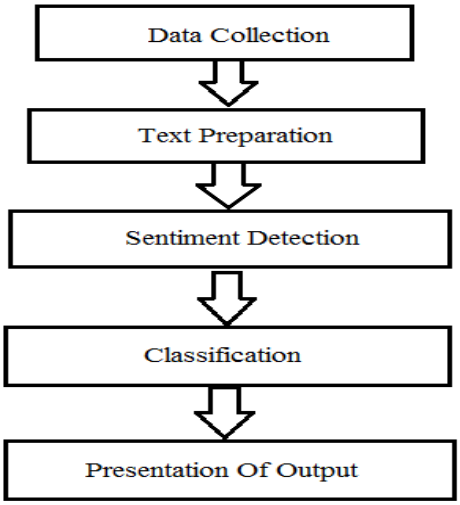

Figure 1 Framework of SVM

IV FEATURES

Accuracy:

Accuracy is the major of the correctness of guess i.e. if in any task having 10 rounds we are correct 8 times, that then that means accuracy is 80%. The formula for accuracy is given as

$$\frac{correct\ guess}{total\ instances} \times 100$$

We have performed some analysis, in that we compared the correctness of SVM with KNN.

Table 1 Shows the Accuracy

Data set size	SVM	k-NN
1000	82.542	79.225
500	76.279	76.538
200	81.528	86.151
100	80.73	85.245
50	78.282	86.864

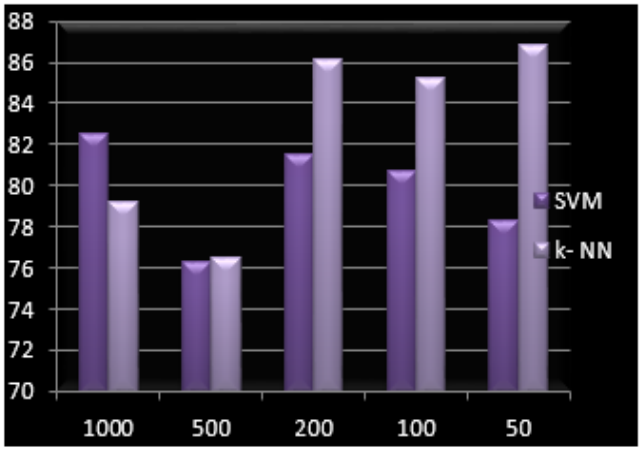

Figure 2 Graph shows the Accuracy

The graph above tells that as the amount of dataset enlarges the accuracy decreases of KNN.

Model Build Time:

The model build time is the total time required to build the model from the given dataset. Also, it is an elapsed time for the building the system.

Table 2 Shows the Build Time

Data set size	SVM(seconds)	k-NN(seconds)
1000	3.273	9.155
500	1.629	6.582
200	1.284	6.114
100	0.732	3.25
50	0.539	1.84

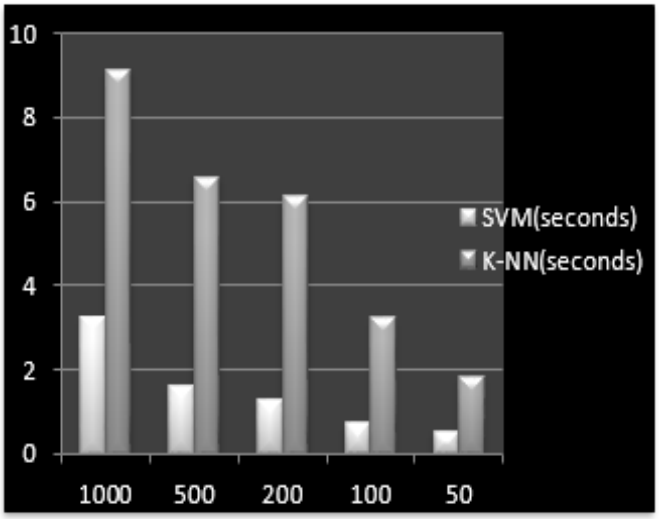

Figure 3 Graph shows the Build Time

The graph above clearly shows that there is a vast difference between the elapse time require for SVM and that for KNN. KNN takes very large time for building the model.

Search Time:

Search time is the time needed for guessing the values as well as the time for examining the model is also the search time. We performed some tasks on single value guessing in the second term.

Table 3 Shows the Search Time

Data set size	SVM(seconds)	k-NN(seconds)
1000	0.0642	0.261
500	0.0662	0.527
200	0.0642	0.527
100	0.0642	0.229
50	0.0642	0.103

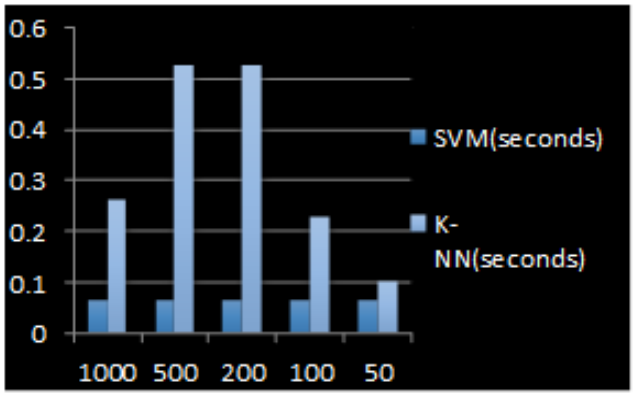

Figure 4 Graph shows the Search Time

Search time of the KNN increases as dataset size gets increases but in SVM the time is constant always and is comparatively very less than KNN.

Memory Used:

Memory used is a term which calculates the total main memory that is used by the system for the exact and successful working. Also memory i.e. main memory used by algorithms for implementation is again called as memory used. It is calculated in KB's.

Table 4 Shows the Memory Used

Data set size	SVM	k-NN
50	23836	24032
100	24776	25168
200	26464	26464
500	27108	27536
1000	29348	29348

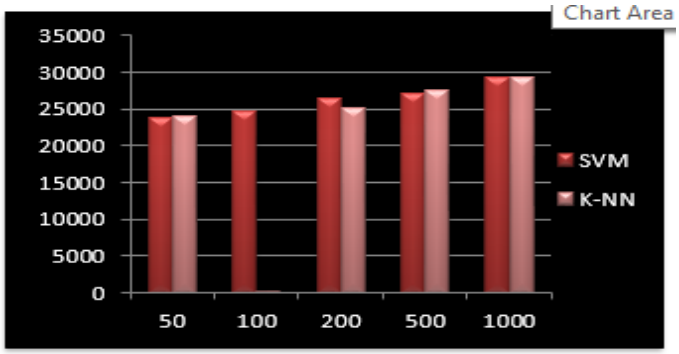

Figure 5 Graph shows the memory used

The graph shows both the algorithms i.e. SVM and KNN uses an approximately same size of memory.

Assumption And Dependency:

The actual scope comes into the picture when the raw data is not totally structured as well as the listings which are individual are not always as per the requirement in case of above-concerned fields. Hence for an experimental purpose, we select data in CSV pattern.

IV CONCLUSION

The paper represents extracting the unstructured customer review and processing it into effective data. In order to ease the customer's satisfaction, their opinion has to be taken into consideration. There are many actions to analyze the opinions and requirements of the customer. Sentiment analysis is one of the many ways that can be used to analyze the customer's point of view. KNN is good for classification but it cannot be implemented on textual data as its performance fluctuates according to the dataset sizes. As the performance of the SVM is not relying on the dataset it becomes favorable for the sentiment analysis. Our future work will focus on automatic summarization of the reviews extracted from the customer

REFERENCES

[1]Alexandra Cernian *et al.*, "Sentiment analysis from product reviews using SentiWordNt as lexical resource", Publisher IEEE, 2015.

[2]Geetika Gautam *et al.*, "Sentiment Analysis of Twitter data usingMachine Learning Approaches and Semantic Analysi", 20 14,IEEE.

[3]Farhan Hassan Khan *et at.*, "SentiView: A Visual Sentiment Analysis Framework", Publisher IEEE, 2014, pp. 291-296.

[4] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," Procedia Computer Science, vol. 17, no. 0, pp. 26– 32, 2013, first International Conference on Information Technology and Quantitative Management

[5] R. Feldman, "Techniques and applications for sentiment analysis," Commun. ACM, vol. 56, no. 4, pp. 82–89, Apr. 2013.

[6]M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments," Expert Systems with Applications, vol. 40, no. 10, pp. 4241 – 4251, 2013.

[7]Min Wang, Hanxio Shi., "Research on Sentiment Analysis Technology and Polarity Computation of Sentiment words", Publisher IEEE, 2010,
pp.331-334.

[8]T. S. Zakzouk and H. I. Mathkour, "Comparing text classifiers for sports news," Procedia Technology, vol. 1, no. 0, pp. 474 – 480, 2012, first World Conference on Innovation and Computer Sciences (INSODE 2011).

[9]H. Tang, S. Tan, and X. Cheng, "A survey on sentiment detection of reviews," Expert Systems with Applications, vol. 36, no. 7, pp. 10 760 – 10 773, 2009.

[10]S. Blair-Goldensohn, K. Hannan, R. McDonald,T. Neylon, G. A. Reis,J. Reynar, "Building a sentiment summarizer for local service reviews," WWW Workshop on NLP in the Information Explosion Era (NLPIX
08), Apr. 2008, Beijing, China.

[11] B. Pang and L. Lee, "A sentimental education:Sentiment analysis using subjectivity summarization based on minimum cuts," in Proceedings of the ACL,