THE PERFORMANCE OF VARIOUS SUPERVISED MACHINE LEARNING CLASSIFICATION ALGORITHMS IN SENTIMENT ANALYSIS OF ONLINE CUSTOMER FEEDBACK IN RESTAURANT SECTOR OF HOSPITALITY INDUSTRY

Nihara Yadav[1], Ms. Sonal Arora[2]
[1]M.TECH (CSE), [2]Asst. Prof.(CSE)
M.Tech (Database Mining and Information Technology)
D.P.G. Institute of Technology & Management, Gurugram, Haryana.

ABSTRACT: In the world of internet, connecting different businesses to businesses and businesses to customers offering a multitude of choices leading to a dilemma in the minds of the customers. The customers largely depend on the available online reviews for making purchase decisions as well it helps the business to make decisions. Sentiment Analysis has a wide range of commercial applications. One of the main applications of sentiment analysis is how to extract emotions from a customer review, and how to classify the customer reviews into positive or negative reviews, & find fake reviews. Sentiment Analysis and text classification methods are applied to two different datasets of restaurant reviews namely dataset 1 and dataset 2. The results show SVM algorithm outperforms not only in text classification but also can be deployed in detecting fake reviews. More specifically, A comparison is made five supervised machine learning algorithms: Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbours (KNN-IBK), KStar (K*) and Decision Tree (DT-J48) for sentiment classification of reviews

Keywords: Sentiment Analysis, Customer Reviews, Fake Reviews, Naïve Bayes, Decision Tree -J48, Support Vector Machine, k-Nearest Neighbour, K-Star.

I. INTRODUCTION

The customer adopts the easiest approach i.e. reading the online reviews, ratings, having a glimpse of uploaded images, etc. by other experienced customers, therefore, making an informed choice. Positive reviews aids in attaining significant profit and growth in the business. While negative reviews can cause a complete downfall. Manually sorting through all such data would be difficult, expensive, and extremely time-consuming. Sentiment Analysis thus allows one to make sense of the all the data through automation thus yielding actionable insights which are otherwise unattainable. Hence, Sentiment Analysis allows businesses to listen to the customers via internet and take necessary action. The scope of the study is to assess the performance of various supervised machine learning classification algorithms in sentiment analysis.

II. OBJECTIVES OF STUDY

The main objectives of the study

- To classify online customer feedback based on polarity i.e. positive or negative.
- To assess the performance of various supervised machine learning algorithms.
- To carry out secondary research to ascertain if sentiment analysis can be extended for detection of fake positive or negative reviews.

III. RESEARCH METHODOLOGY

The following steps show the research methodology in order to reach the objectives of study.

• Problem Identification:
To define the problem preciously a lot of searching was done on the original idea.

•Literature Review:
After identifying the exact problem to be solved, a searching process was done going back to the past papers, conferences, books, articles and any other material that involved studying either the whole problem or a part of the problem.

•Data Collection:
Data Collection is the process of gathering and measuring information in a systematic manner based on the variables of interest in order to achieve the objectives of the study.
The below table summarizes the two collected datasets.

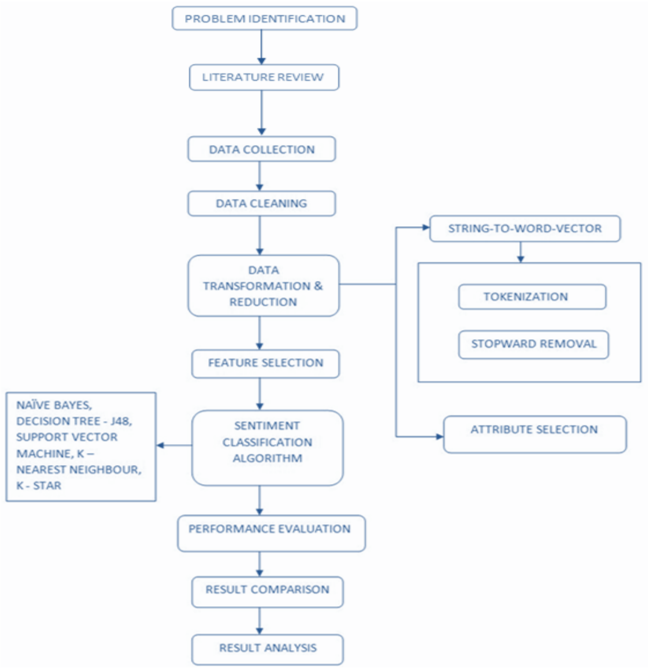

Figure : Research Methodology

•Data Cleaning:
Data cleaning is a vital aspect in data mining It is very important to clean the data before analysing. The skewed distribution of data can have a major negative impact on the classification accuracy.
The datasets before and after the cleaning process are summarized in the below tables:

Before Data Cleaning			
Dataset Source	**Attributes**	**Class Rating**	**No. of Online Customer Feedback**
TripAdvisor	Review, Overall Rating, Date of visit, Length, Image Shared, helpful, no. of reviews posted by the same user, Reviewer id	Poor, Terrible	1250
		Average, Very Good, Excellent	1564
Dineout	Review, Overall Rating, Name of reviewer, liked, Disliked, Image Count, Length, Image Shared, votes	1star,2star	912
		3star,4star, 5star	1439

Table: Datasets Before Data Cleaning

Dataset Collection Source	Number of Online Customer Feedback Collected	
TripAdvisor	Positive	1564
	Negative	1250
Dineout	Positive	1439
	Negative	912

Table: Datasets After Data Cleaning

•Data Transformation and Reduction:
It incorporates two primary operations that are carried out before the application of the various classification algorithms. These processes are:
i)String-To-Word-Vector
A string vector is defined as an ordered finite set of strings. The STWVs are structured data that represent the raw data. The data is transformed by using the String-To-Word-Vector filter present in WEKA. The STWV is the main filter for text analysis.
a) Configuring the tokenizer
This process makes the provided data classifiable by converting the content into a set of features using machine learning techniques.
b) Specifying a stopwords list
The stopwords are the unwanted words which are to be filtered out and completely eliminated before training the classifier. Some of the most common stopwords are "a", "of ", "the", "you", "I", "it", "and".
ii)Attribute Selection
Attribute selection is analogous to feature selection in ML. An attribute selection scheme has to be applied for training the classifier as it significantly increases the classification accuracy.

•Feature Selection
Feature Selectionhelps in increasing the classification accuracy by identifying the relevant attributes in the dataset. The feature selection method used is CfsSubsetEval with BestFirst which is the preferred choice for sentiment analysis classification process along with stop-word removal.
•Sentiment Classification through Supervised Machine Learning Algorithms:
The Sentiment Classification through Supervised Machine Learning Algorithms is an Automated Sentiment Approach.
The algorithm which implements the classification on a dataset is known as a classifier. The classifiers used in the study are explained below
Naïve Bayes is a very basic and simple probabilistic classifier based on Bayes Theorem. It calculates a set of probabilities based on the Bayes rule with independence assumptions between the features.
Decision Tree – J48 is a predictive classifier that decides the target value of a new sample based on several attribute values of the available data. It is the implementation of Ross Quinlan's Iterative Dichotomiser 3 algorithm.
Support Vector Machine is a discriminative classifier. It has related learning algorithms that analysedataset usedfor classification. It examines the data and then identifies the patterns and learns them.
K-NN is a non-parametric lazy classifier. It classifies an object by a majority vote of the object's neighbours in the space of input parameter. This 'k' is the number that decides how many neighbours will influence the classification process of the classifier.
K-star is an instance-based classifier. Each new instance is compared with the existing ones using a distance metric and the closest existing instance is used to assign the class to the new one.

•Performance Evaluation
In this step a confusion matrix is generated. The confusion matrix classifies the reviews as either real or fake. It provides a matrix as an output which describes the performance of the classifier used for the classification process. Confusion Matrix is also known as the 'Error Matrix'. This matrix allows visualization of the performance of a classifier.

		PREDICTED	
		NEGATIVE	**POSITIVE**
ACTUAL	**NEGATIVE**	True Negative Reviews (TN)	False Positive Reviews (FP)
	POSITIVE	False Negative Reviews (FN)	True Positive Reviews (TP)

Table: Confusion Matrix

True Negative Reviews (TN) = These are negative reviews which are correctly classified by the model as negative.
False Negative Reviews (FN) = These are positive reviews which are incorrectly classified by the model as negative.
True Positive Reviews (TP) = These are positive reviews which are correctly classified by the model as positive.
False Positive Reviews (FP) = These are negative reviews which are incorrectly classified by the model as positive.

A Confusion Matrix is generated for the performance evaluation of each of the five supervised machine learning algorithms/classifiers applied on both the datasets. The confusion matrix shows the counts of positive and negative predictions obtained with known data.

ALGORITHMS	SA	NEGATIVE	POSITIVE
NB	NEGATIVE	660	240
	POSITIVE	215	685
J48	NEGATIVE	686	214
	POSITIVE	301	599
SVM	NEGATIVE	694	206
	POSITIVE	194	703
KNN	NEGATIVE	617	283
	POSITIVE	252	648
K*	NEGATIVE	665	235
	POSITIVE	267	633

Table: Confusion Matrix for TripAdvisor (Dataset 1)

ALGORITHMS	SA	NEGATIVE	POSITIVE
NB	NEGATIVE	920	280
	POSITIVE	211	989
J48	NEGATIVE	899	301
	POSITIVE	313	887
SVM	NEGATIVE	991	209
	POSITIVE	201	999
KNN	NEGATIVE	913	287
	POSITIVE	396	804
K*	NEGATIVE	897	303
	POSITIVE	324	876

Table: Confusion Matrix for Dineout (Dataset 2)

The below tables and graphs summarise the results of the performance evaluation parameters on the two datasets:

ALGORITHM	ACCURACY %	PRECISION %	TRUEPOSITIVE REVIEW RATE %	TRUE NEGATIVE REVIEW RATE%	FALL OUT RATE%	MISS RATE %	F1-MEASURE%	ERROR RATE%
NB	79.54	77.9	82.4	76.7	23.3	17.6	80.1	20.46
J48	74.41	74.7	73.9	74.9	25.1	26.1	74.3	25.58
SVM	82.91	82.7	83.3	82.6	17.4	16.8	82.9	17.08
KNN	71.54	73.7	67.0	76.1	23.9	33.0	70.2	28.46
K*	73.87	74.3	73.0	74.8	25.3	27.0	73.6	26.13

Table: Performance Evaluation Parameters for TripAdvisor (Dataset 1)

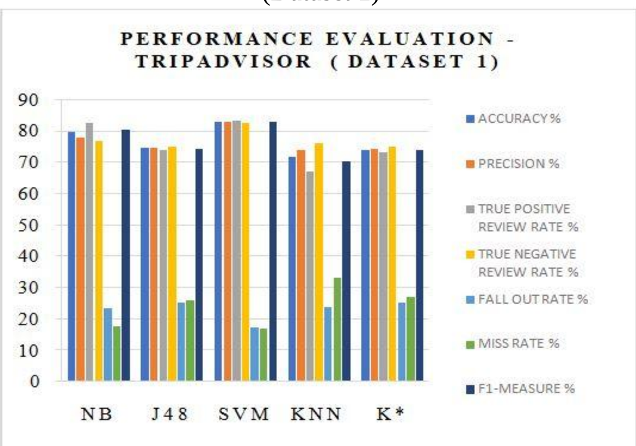

Figure: Performance Evaluation – TripAdvisor (Dataset 1)

Table: Performance Evaluation Parameters for Dineout (Dataset 2)

ALGORITHM	ACCURACY %	PRECISION %	TRUE POSITIVE REVIEW RATE %	TRUE NEGATIVE REVIEW RATE %	FALL OUT RATE%	MISS RATE%	F1-MEASURE%	ERROR RATE%
NB	74.72	74.1	76.1	73.3	26.7	23.9	75.1	25.28
J48	71.38	73.7	66.6	76.2	23.8	33.4	69.9	28.61
SVM	77.61	77.3	78.1	77.1	22.9	21.9	77.7	22.39
KNN	70.27	69.6	72.0	68.6	31.4	28.0	70.8	29.72
K*	72.11	72.9	70.3	73.9	26.1	29.7	71.6	27.89

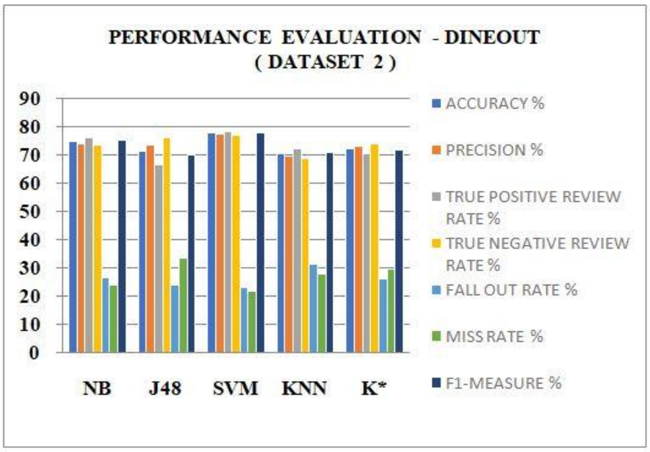

Figure: Performance Evaluation – Dineout (Dataset 2)

Accuracy

The accuracy of each of the algorithm for both the datasets is summarised in the table below

ALGORITHM	ACCURACY %	
	DATASET1	DATASET 2
NB	79.54	74.72
J48	74.41	71.38
SVM	82.91	77.61
KNN	71.54	70.27
K*	73.87	72.11

Table: Accuracy Percentage

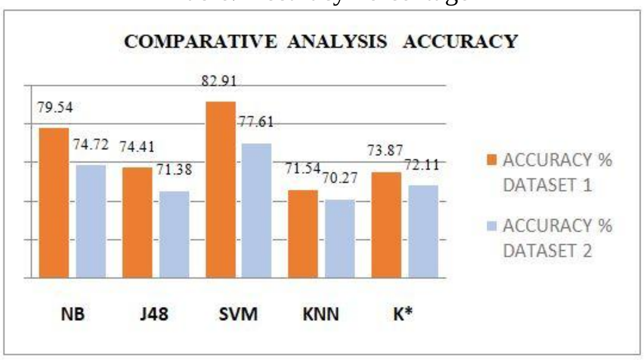

Figure: Comparative Analysis – Accuracy

From the above results it can be concluded that Support Vector Machine i.e. SVM algorithm has the highest accuracy rate for both the datasets and k- Nearest Neighbour i.e. KNN algorithm has the least accuracy rate.

Precision

The precision of each of the algorithm for both the datasets is summarised in the table below

ALGORITHM	PRECISION %	
	DATASET1	DATASET 2
NB	77.9	74.1
J48	74.7	73.7
SVM	82.7	77.3
KNN	73.7	69.6
K*	74.3	72.9

Table: Precision Percentage

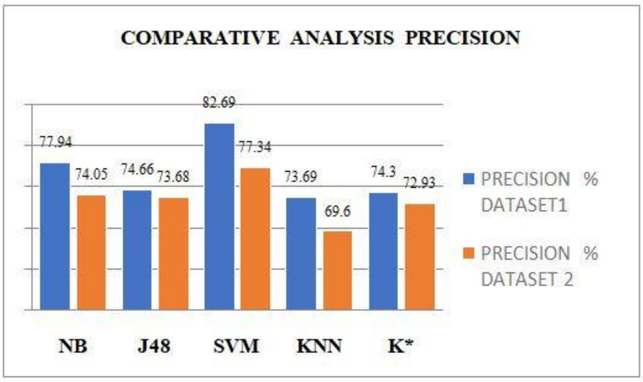

Figure: Comparative Analysis – Precision

From the above results it can be concluded that Support Vector Machine i.e. SVM algorithm has the highest precision rate for both the datasets and k- Nearest Neighbour i.e. KNN algorithm has the least precision rate.

True Positive Review Rate

The true positive review rate of each of the algorithm for both the datasets is summarised in the table below

ALGORITHM	TRUE POSITIVE REVIEW RATE %	
	DATASET1	DATASET 2
NB	82.4	76.1
J48	73.9	66.6
SVM	83.3	78.1
KNN	67.0	72.0
K*	73.0	70.3

Table: True Positive Review Rate

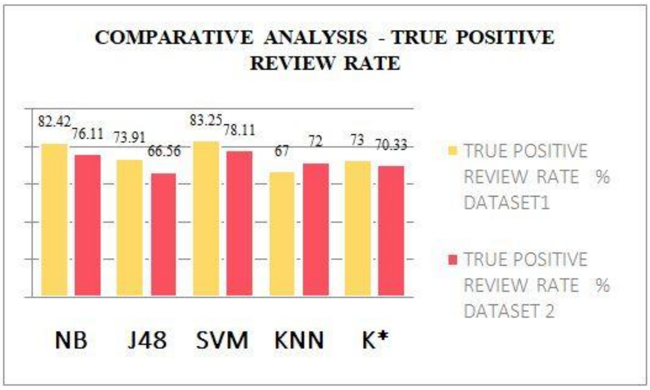

Figure: Comparative Analysis – True Positive Review Rate

From the above results it can be concluded that Support Vector Machine i.e. SVM algorithm has the highest positive review detection rate for both the datasets.

True Negative Review Rate

The true negative review rate of each of the algorithm for both the datasets is summarised in the table below

ALGORITHM	TRUE NEGATIVE REVIEW RATE %	
	DATASET1	DATASET 2
NB	76.7	73.3
J48	74.9	76.2
SVM	82.6	77.1
KNN	76.1	68.6
K*	74.8	73.9

Table: True Negative Review Rate

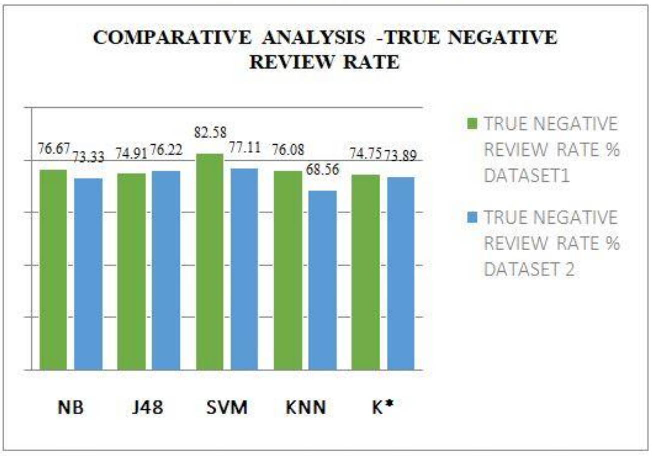

Figure: Comparative Analysis – True Negative Review Rate

From the above results it can be concluded that Support Vector Machine i.e. SVM algorithm has the highest negative review detection rate for both the datasets.

Fall Out Rate

The fall out rate of each of the algorithm for both the datasets is summarised in the table below

ALGORITHM	FALL OUT RATE %	
	DATASET1	DATASET 2
NB	23.3	26.7
J48	25.1	23.8
SVM	17.4	22.9
KNN	23.9	31.4
K*	25.3	26.1

Table: Fall Out Rate

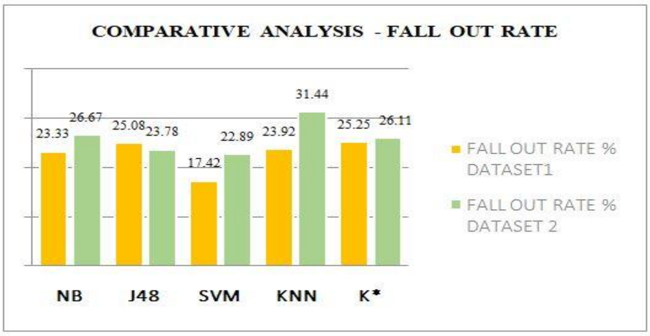

Figure: Comparative Analysis – Fall Out Rate

From the above results it can be concluded that Support Vector Machine i.e. SVM algorithm has the least fall out rate for both the datasets.

Miss Rate

The miss rate of each of the algorithm for both the datasets is summarised in the table below

ALGORITHM	MISS RATE %	
	DATASET1	DATASET 2
NB	17.6	23.9
J48	26.1	33.4
SVM	16.8	21.9
KNN	33.0	28.0
K*	27.0	29.7

Table: Miss Rate

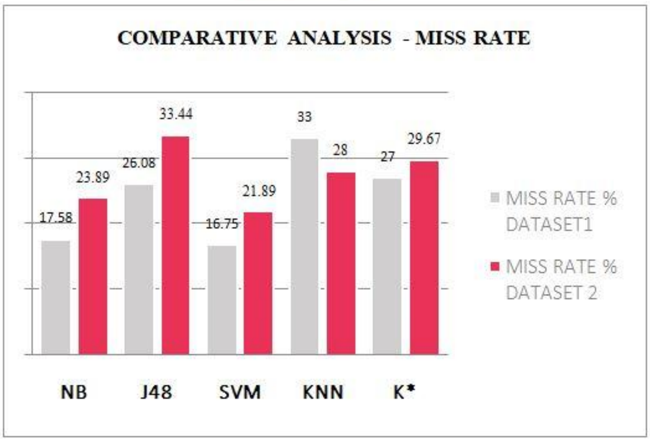

Figure: Comparative Analysis – Miss Rate

From the above results it can be concluded that Support Vector Machine i.e. SVM algorithm has the least miss rate for both the datasets.

F1- Measure

The F1- Measure of each of the algorithm for both the datasets is summarised in the table below

ALGORITHM	F 1 -MEASURE	
	DATASET1	DATASET 2
NB	80.1	75.1
J48	74.3	69.9
SVM	82.9	77.7
KNN	70.2	70.8
K*	73.6	71.6

Table: F 1- Measure

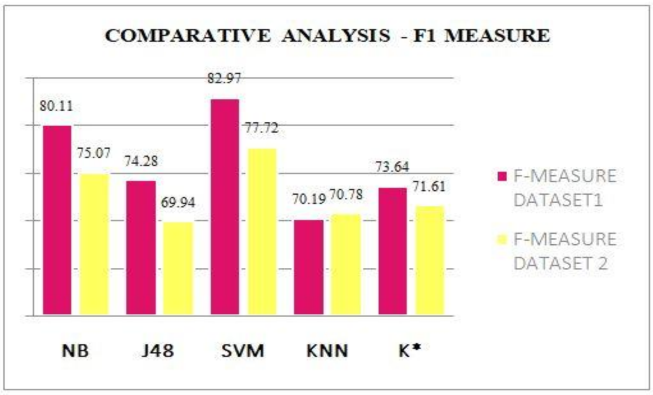

Figure: Comparative Analysis – F1 – Measure

From the above results it can be concluded that Support Vector Machine i.e. SVM algorithm has the highest F1-Measure for both the datasets.

Error Rate

The error rate of each of the algorithm for both the datasets is summarised in the table below

ALGORITHM	ERROR RATE %	
	DATASET1	DATASET 2
NB	20.46	25.28
J48	25.58	28.61
SVM	17.08	22.39
KNN	28.46	29.72
K*	26.13	27.89

Table: Error Rate

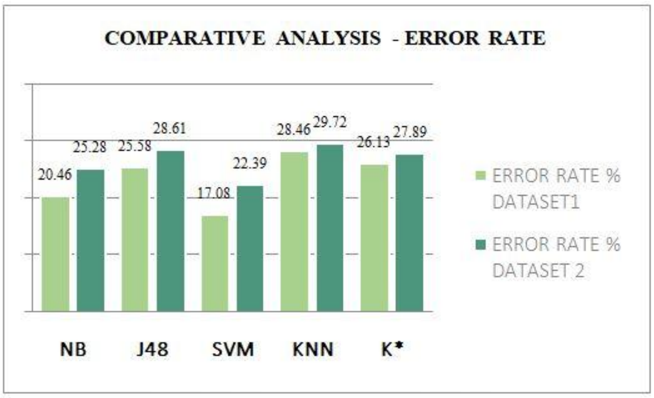

Figure: Comparative Analysis - Error Rate

From the above results it can be concluded that Support Vector Machine i.e. SVM algorithm has the least error rate for both the datasets.

Time Taken

The time taken to build the model by each of the algorithm for both the datasets is summarised in the table below

ALGORITHMS	TIME TAKEN (mm's)	
	DATASET1	DATASET 2
NB	0.13	0.09
J48	0.38	0.33
SVM	16.54	6.84
KNN	0.01	0.00
K*	0.00	0.01

Table: Time Taken

From the above results it can be concluded that Support Vector Machine i.e. SVM algorithm takes the longest time for both the datasets.

The following researches provided a vision for the further expansion of the undertaken study:

- " Prediction of Fake Profiles on Facebook using Supervised Machine Learning Techniques - A Theoretical Model " - A research by Suheel Yousuf Wani (IIIT Banglore) , Mudasir M Kirmani (SKUAST-K, J&K) ,Syed ImamulAnsarulla (MANUU, Hyderabad), International Journal of Computer Science and Information Technologies(IJCSIT), Vol. 7 (4) , 2016, 1735-1738

- " Fake Comment Detection Based on Sentiment Analysis " - A research by Su Chang (School of Foreign Languages, Harbin Institute of Technology, Harbin 150001, PR China), Xu Zhenzhong (College of Computer Science and Technology, Beihang University, Beijing 100191) , Gao Xuan (State Key Laboratory of Super hard Materials, Jilin University, Changchun 130000, PR China).
- "Detection of Fake Tweets Using Sentiment Analysis"- A research by C. Monica, N. Nagarathna (Department of Computer Science, B. M. S. College of Engineering, Bengaluru 560019, India), SN Computer Science (2020) 1:89 https://doi.org/10.1007/s42979-020-0110-0,Springer Nature Singapore Pte Ltd 2020
- "Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews" - A research by ArjunMukherjee, VivekVenkataraman, Bing Liu (University of Illinois, Chicago), Natalie Glance (Google Inc)
- "Opinion Spam and Analysis" - A research by NitinJindal and BingLiu (Department of Computer Science University of Illinois, Chicago).
- "A Framework for Fake Review Detection in Online Consumer Electronics Retailers"- Aresearch by Rodrigo Barbado, Oscar Araque, Carlos A. Iglesias (Intelligent Systems Group, Department of Telematic Engineering Systems, Universidad Polit´ecnica de Madrid)

IV. CONCLUSION AND FUTURE WORK

The primary research concluded that the classification of online customer feedback/reviews as positive or negative using sentiment analysis through supervised machine learning achieved.Taking into consideration it is concluded that Support Vector Machine i.e. SVM Algorithm outperforms all the other supervised machine learning classification algorithms taken for carrying out the study but it takes the longest time. From the secondary research studies, it can be implied that sentiment analysis can be deployed in fake positive or negative review detection with supervised machine learning which has been one of the main approaches for solving the problem of fake positive or negative reviews. However, obtaining a labeled dataset of fake reviews both positive and negative for the training purpose is quite difficult and it is impossible to manually label fake reviews. The existing studies have used several types of pseudo fake reviews for training. While the other existing studies propose only a theoretical model. Both can be summed up as the conclusive statements to validate the aimed statement of the thesis to provide insightful gatherings.
For futurework,the study can be extended to e-commerce online portals like Club Factory or Wish and different feature selection methods can be applied and tested as well as other classification algorithms like Logistic Regression can also be tested to see the performance of the algorithm in sentiment analysis. Moreover, sentiment classification algorithms can be applied using tools like Python.
As for the detection of fake customer feedback/reviews python could also be tested in combination with machine learning. This can further be extended to fake review detection in e-commerce, fake news detection, fake profile detection on Instagram. However, it is subjected to availability of a reliable labelled fake dataset

REFERENCES

[1] C.-H. Chu, C.-A. Wang, Y.-C. Chang, Y.-W. Wu, Y.-L. Hsieh, and W.-L. Hsu, "Sentiment analysis on Chinese movie review with distributed keyword vector representation," in Technologies and Applications of Artificial Intelligence (TAAI), 2016 Conference on. IEEE, 2016, pp.84–89.
[2] V. Singh, R. Piryani, A. Uddin, and P. Waila, "Sentiment analysis of movie reviews and blog posts," in Advance Computing Conference (IACC), 2013 IEEE 3rd International. IEEE, 2013, pp. 893–898.
[3] G. Vinodhini and R. Chandrasekaran, "Sentiment analysis and opinion mining: a survey," International Journal, vol. 2, no.6, 2012, pp. 282–292.
[4] P. Kalaivani and K. L. Shunmuganathan, "Sentiment classification of movie reviews by supervised machine learning approaches," Indian Journal of Computer Science and Engineering, vol. 4, no. 4, pp. 285-292, 2013.
[5] N. Jindal and B. Liu, "Opinion spam and analysis," in Proceedings of the 2008 International Conference on Web Search and Data Mining. ACM, 2008, pp. 219–230.
[6] R. Moraes, J. F. Valiati, and W. P. G. Neto, "Document-level sentiment classification: An empirical comparison between svm and ann," Expert Systems with Applications, vol. 40, no. 2, 2013, pp. 621–633.
[7] B. Liu, "Sentiment analysis and opinion mining," Synthesis lectures on human language technologies, vol. 5, no. 1, 2012, pp. 1–167.
[8] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, no. 4, 2014, pp. 1093–1113.
[9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in Proceedings of EMNLP,2002, pp. 79–86. [Online]. Available: http://www.cs.cornell.edu/People/pabo/movie%2Dreview%2Ddat a/
[10] B. Liu, M. Hu, and J. Cheng, "Opinion observer: analysing and comparing opinions on the web," in Proceedings of the 14th international conference on World Wide Web. ACM, 2005, pp. 342–351.
[11] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20.
[12] S. Hassan, M. Rafi, and M. S. Shaikh, "Comparing svm and naive bays classifier for text categorization with wikitology as knowledge enrichment," in Multi topic Conference (INMIC), 2011 IEEE 14th

International.IEEE, 2011, pp. 31–34.

[13] G. Xu, Y. Cao, Y. Zhang, G. Zhang, X. Li, and Z. Feng, “Term: Computing reputation score by mining reviews.” in AAAI Workshop: Incentives and Trust in Electronic Communities, 2016.

[14] N. Tian, Y. Xu, Y. Li, A. Abdel-Hafez, and A. Josang, “Generating product feature hierarchy from product reviews,” in International Conference on Web Information Systems and Technologies. Springer, 2014, pp. 264–278.

[15] N. Tian, Y. Xu, Y. Li, A. Abdel-Hafez, and A. Jøsang, “Product feature taxonomy learning based on user reviews,” in WEBISTE (2), 2014, pp.184–192.

[16] A. Abdel-Hafez and Y. Xu, “A survey of user modelling in social media websites,” Computer and Information Science, vol. 6, no. 4, 2013, p. 59.

[17] A. Abdel-Hafez, Y. Xu, and D. Tjondronegoro, “Product reputation model: an opinion mining-based approach,” in SDAD 2012 The 1st International Workshop on Sentiment Discovery from Affective Data, 2012, p. 16.

[18] A. Liu, M. Hu, and J. Cheng, “Opinion observer: analysing and comparing opinionson the web,” in Proceedings of the 14th international conference on World Wide Web. ACM, 2005, pp. 342–351.

[19] B. Fujii and T. Ishikawa, “A system for summarizing and visualizing arguments in subjective documents: Toward supporting decision making,” in Proceedings of the Workshop on Sentiment and Subjectivity in Text. Association for Computational Linguistics, 2006, pp. 15–22.

[20] L.-W. Ku, Y.-T. Liang, and H.-H. Chen, “Opinion extraction, summarization and tracking in news and blog corpora,” in Proceedings of AAAI, 2006, pp. 100–107.

[21] J. Malbon, “Taking fake online consumer reviews seriously,” Journal of Consumer Policy, vol. 36, no. 2, 2013, pp. 139–157.

[22] R. Xia, C. Zong, and S. Li, “Ensemble of feature sets and classification algorithms for sentiment classification,” Information Sciences, vol. 181, no. 6, 2011, pp. 1138 – 1152.

[23] Liu, Bing (2010). “Sentiment Analysis and Subjectivity” (PDF). In Indurkhya, N.; Damerau, F. J. (eds.). Handbook of Natural Language Processing (Second ed.).

[24] Galitsky, Boris; McKenna, Eugene William. “Sentiment Extraction from Consumer Reviews for Providing Product Recommendations”. Retrieved 18 November 2013.

[25] Ogneva, M. “How Companies Can Use Sentiment Analysis to Improve Their Business”. Mashable. Retrieved 2012-12-13.

[26] Roebuck, K. (2012-10-24). Sentiment Analysis: High-impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors.

[27] Tang, Huifeng; Tan, Songbo; Cheng, Xueqi (2009). "A survey on sentiment detection of reviews" (PDF). Expert Systems with Applications. 36 (7): 10760 10773.

