Global Social Sciences Review (GSSR)
Vol. IV, **No.** IV (Fall 2019) | **Page:** 146 – 156

The Classification of Customers' Sentiment using Data Mining Approaches

Dost Muhammad Khan — Assistant Professor, Department of Computer Science and Information Technology, The Islamia University of Bahawalpur, Punjab, Pakistan.
Email: khan.dostkhan@gmail.com

Tariq Aziz Rao — Visiting Lecturer, Department of Computer Science, Virtual University of Pakistan, Lahore, Punjab, Pakistan.

Faisal Shahzad — Lecturer, Department of Computer Science and Information Technology, The Islamia University of Bahawalpur, Punjab, Pakistan.

Abstract *Data mining is a procedure of extracting the requisite information from unprocessed records by using certain methodologies and techniques. Data having sentiments of customers is of utmost importance for managers and decision-makers who intend to monitor the progress, to maintain the quality of their products or services and to observe the latest market trends for business support. Billions of customers are using micro-blogging websites and social media for sharing their opinions about different topics on daily basis. Therefore, it has become a source of acquiring information but to identify a particular feature of a product is still an issue as the information retrieves from varied sources. We proposed a framework for data acquisition, preprocessing, feature extraction and used three supervised machine-learning algorithms for classification of customers' sentiments. The proposed framework also tested to evaluate the system's performance. Our proposed methodology will be helpful for researchers, service providers, and decision-makers.*

Key Words
Data Mining; Sentiment Analysis; Classification

Introduction

Today, social media has become the most popular information-sharing platform with high interactivity as its spread over small villages where Internet access is available. Billions of users are posting millions of messages on daily basis on micro-blogging websites such as Facebook, Twitter and many more.

According to the report of Statista (The Statistics Portal), in 2017, a number of social media users were approximate 2.46 billion, whereas, it will cross 3.02 billion around the globe till 2021 as shown in Figure 1.

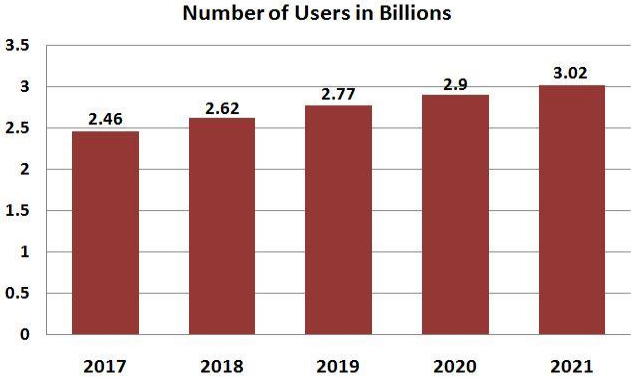

Figure 1. Number of Social Media Users (in billions) Worldwide from 2017 to 2021

p-ISSN 2520-0348 | **e-ISSN** 2616-793X | **L-ISSN** 2616-793X | **DOI:** 10.31703/gssr.2019(IV-IV).19 | **URL:** http://dx.doi.org/10.31703/gssr.2019(IV-IV).19

Information disseminates through social media contains valuable opinions and reviews about different topics that provides rich knowledge about real-world events occurred in our daily life. Marketers and decisions-makers in a company can efficiently use this information to improve their products quality and sale. However, due to the massive amount of information and complexity of data, it is nearly unworkable for a decision-maker or service providers to read manually all the data extracted from popular social-media like Twitter and Facebook and from different web sources, such as blogs and discussion forums. Resultantly, valuable information often ignored. Therefore, the opinion mining has turned into a major research area for researchers to come forward and play their beneficial role to tackle this challenge. Therefore, there is a dire need to develop a system that will automatically or semi-automatically extracts and analysis the users' sentiments about any service/product provided by any company/organization from the massive volume of data.

In our proposed architecture, we extracted Twitter data having customers' sentiments regarding service provided by an airline company. Analyzed this data by using a data mining technique i.e. classification with the help of three supervised machine learning algorithms, such as Support Vector Machine (SVM), Naive Bayes (NB) and Decision Tree (DT) and produces a final outcome based on experimental results.

The rest of the paper is organized as: Literature is reviewed in Section 2, Section 3 is about the Proposed Methodology, the Experimental Results are discussed in Section 4, and finally Conclusion is drawn in Section 5.

Literature Review

Social media like Facebook and Twitter have become a more popular platform that permits the consumers to express their personal opinion about any service or product provided by the company. Their opinions are helpful for assessment to a person who would intend to know others' sentiments prior to avail any service or purchase a product and helpful for companies' executives.

Mani et al (2018) made an evaluation of different data mining classification techniques such as Random Forest (RF), Radial Basis Function (RBF), Decision Tree (DT), Multilayer Perception (MLP), Sequential Minimal Optimization (SMO), Naive Bayes (NB), Ada Boost, and Decision Stump to sort the crime velocity with elevated accuracy and concluded that DT produced 96.4% precision with negligible false-positive rhythm.

A classifier model proposed by Songpan (2017) that used a case study of 400 Thai customer reviews about hotels from a website to categorize the comments as positive or negative. This model has estimated probability which demonstrates the value of craze to present the score by utilizing NB and DT techniques and concluded that the Naive Bayes gives the better result with an average of 93.61% accuracy.

Mars and Salah Gouider (2017) used an approach that composed of four phases, such as opinion identification, feature extraction, sentiment classification, and result visualization & summarization. Their proposed framework used big data technologies merged with text mining tools and machine learning that enables to detect the opinions of customers about merchandise characteristics from social media. They also developed a method for retrieving 100000 customers' tweets about five different electronic devices from Twitter by using some hash-tag and extract opinions regarding characteristics of these devices and their sentiments polarity based on powerful programming model i.e. MapReduce. They also suggested that their system may be tested with other benchmarks to quantify its performance.

Qadri et al (2017) used Artificial Neural Network (ANN), J48 (extended version of C 4.5), Naive Bayes (NB) and Random Forest (RF) data mining techniques for the classification of Multispectral and texture datasets. The acquired results were 96.40 % for Multispectral and 91.334 % for texture data.

Kharde and Sonawane (2016) made a survey on sentiment analysis of Twitter data where opinions are highly unstructured and heterogeneous. They mainly focused on lexicon-based approaches such as dictionary based and corpus-based and machine-learning algorithms like SVM and NB for opinion analysis. According to the authors, the words sentiment, opinion, belief, and view are different in meaning. Sentiment analysis is an interdisciplinary task that includes machine learning, web mining and natural language processing. It can be rotted into sentiment classification, subjectivity classification, and complementary tasks. These tasks can be made at four levels, such as word, sentence, document, and feature-based sentiment analysis. At last, they concluded that Naive Bayes (NB) algorithm with bigram model give the high accuracy and extra cleaner data provides more accurate results.

Jadav and Vaghela (2016) described that sentiment analysis is the classification of users' reviews about anything either in positive or negative. They made an analysis of the movie, Twitter, and gold datasets by using two supervised learning algorithms such as NB and SVM classifiers. Also made an assessment of SVM and NB algorithm and found SVM with RBF kernel hyper-parameter (C, γ) gives better results as they acquired 73.56%, 74.74% and 78.18% accuracy on gold, movie and Twitter datasets respectively.

Noor Injadat et al (2016) conducted a comprehensive survey of 19 data mining techniques by selecting 66 articles after filtrating the 1187 articles published in digital libraries of Science Direct, IEEE explorer, ACM, Google

Scholar, etc. between the period from 2003 to 2015 to answer the five research questions. After a thorough survey, they concluded that SVM, Bayesian Network (BN) and Decision Tree (DT) are the most frequently used data mining approaches to extract the social media information. During aforesaid period, quality improvement and sentiment analysis were the most vigorous research objectives in six general domains, such as Social Networks, Business and Management, Finance, Education, Medical and Health, Government and Public sectors. According to their survey report, both machine and non-machine learning data mining approaches are essential for data extracting tasks. Furthermore, the researchers have not yet investigated some domains like Human Resource Management and Customer Relationship Management and others but they encouraged the researchers to come forward and play their valuable role in these areas. During the investigation, they also transpired that large numbers of studies have not applied any statistical test, such as t-test, ANOVA, and MANOVA. They also identified that there are 9 research objectives (Sentiment Analysis, Cyber Crime, Semantic Analysis, Geo-locating, Biometric, Disease Awareness, Content Analysis, and Risk Management) due to which these data mining techniques are adopted by the researchers.

About 400 million people of 22 countries are speaking the Arabic language, so, its importance cannot be ignored. 120 patterns and 10,000 roots make this language more complex. Atia and Shaalan (2015) presented an approach for mounting the precision of opinion mining in Arabic by using text processing, machine learning & text mining and performance evolution method. Various machine-learning approaches are available for opinion mining but according to the authors, NB and SVM are the most valuable classifiers. During performance evaluation, they acquired 8% accuracy when used NB and 7% accuracy when used SVM. They also concluded that supervised learning algorithm i.e. Naive Bayes attain the premier level of accuracy when the BTO is utilized, whereas, SVM attained the maximum level of accuracy when TF-IDF is utilized.

Cluster analysis played a pivotal role in market research during working with multivariate records from test panels and surveys. It used by the market researchers to partition the general inhabitants of customers into inter stock patterns, market segmentation and to know the relationship between them. Joshi et al (2015) explained that classification and patterns detection from customer's reviews about any product is a very significant aspect of administrative and business support. They planned a system that consists of two stages. In Stage-I, split the data into three different bunches based on product classes and retailed quantities such as Slow-Moving Stock, Fast-Moving Stock, and Dead Stock by utilizing an unsupervised learning clustering algorithm i.e. K-means algorithm. In Stage-II, find out the repeated patterns of product's features of every category of goods and give trade trends in a compressed shape by using Most Frequent Pattern algorithm.

The fundamental aim of reviews analysis is to evaluate the mind of a speaker/writer about some matter. Word of mouth (WOM) involves end users' sharing attitudes and opinions about any product or services that played a vital role in decision making by the users as well as decision makers. For sentiment analysis, Patil et al (2014) studied Support Vector Machine for opinion analysis of user sentiments towards political candidates through comments and tweets and concluded that SVM provides better performance on text classification. During experiments, they came on the conclusion that simply uses of an adjective as features lead to the inferior performance of the system.

Zainuddin and Selamat (2014) obtained remarkable results by applying SVM on target datasets to train an opinion classifier that uses N-gram & various weighting methods as an input to the classifier to mine the most classical characteristics for Taboada and Pang Corpus. In order to train the model and to evaluate its accuracy, they used the training dataset and detach testing dataset and each set split into two parts, 70% part is used for training and remaining 30% utilized for testing purpose. To classify the testing dataset (as positive or negative), they used the supervised learning algorithm i.e. Support Vector Machine. They achieved 0.892 value of F-measure and 89.17% accuracy, which depicts that these results are much better than previous where results were obtained without feature selection and acquired only 0.87 value of F-measure and 87.33% accuracy.

Due to the massive development of social media like Twitter and Facebook, microblogs have become quick and effortless online information sharing platform. Similarly, Weibo a microblogging service uses in China. Lin et al (2014), described that researchers are mainly focused on sentiments classification which neither capable to combine tree-like re-tweeting structure nor analyze opinion evolutions with a holistic analysis. They constructed an opinion descriptive model and designed a lexicon-based sentiments orientation analysis algorithm to classify the sentiments. They design and implement a prototype that can extract opinions with respect to re-tweeting tree structure as well as comments. In order to conduct the sentiment classification, they used the two methods such as lexicon based and SVM and achieved 63.52% and 72.90% precision respectively.

Kumari and Narayan Singh (2014) presented a review paper in which they described that the corporate intend to observe the pulse of customers on social media regarding their brand to take an appropriate action. They proposed an opinion-mining method consisting of three phases, such as pre-processing, extract association rule and summarization.

Fegade and Patil (2014) proposed a system by using unsupervised machine learning Apriori algorithm and optimization method (genetic algorithm) to learn the customer behavior. According to the authors, this system requires less time and memory as compared to the other existing techniques. This method is valuable to progress the efficiency of sales and marketing as identification of customer behavior is essential to know that which product purchase by the customer on a frequent basis.

Table 1. Summaries the Various data Mining Approaches Discussed in Section 2.

Authors	Algorithm	Results
Mani et al	RF, RBF, DT, MLP, SMO, NB, AB, DS	Precision of DT was 96.4%
Ahmad et al	SVM	Accuracy of SVM on two datasets was 59.91% & 71.2% respectively
Songpan	NB, DT	Accuracy of NB was 93.61% on small dataset i.e. 400 Thai customers reviews
Qadri et al	ANN, J48, NB, RF	Acquired results were 96.40% for multispectral and 91.33% for texture data
Kharde and Sonawane	NB, SVM	Accuracy of SVM with unigram and bigram was 76.68% & 77.73% respectively
Jadav and Vaghela	SVM, NB	Accuracy of SVM with RBF kernel was 73.56%, 74.74% & 78.18% on gold, movie and Twitter datasets respectively
Atia and Shaalan	NB, SVM	Accuracy of NB and SVM were 96.00% & 96.97% respectively
Zainuddin and Selamat	SVM	Accuracy of SVM was 89.17 %
Lin et al	SVM	SVM provided 72.90% precision

Comparison of Results Obtained by Various Researchers

Proposed Methodolgy

Two main approaches are being used for sentiment analysis, such as Machine-learning approaches (supervised and unsupervised learning) and Lexicon-based approaches (corpus-based and dictionary based). In our proposed architecture, a supervised machine learning approach is used. An overview of sequential steps and modus operandi used in customers' sentiments analysis is shown in Figure-2.

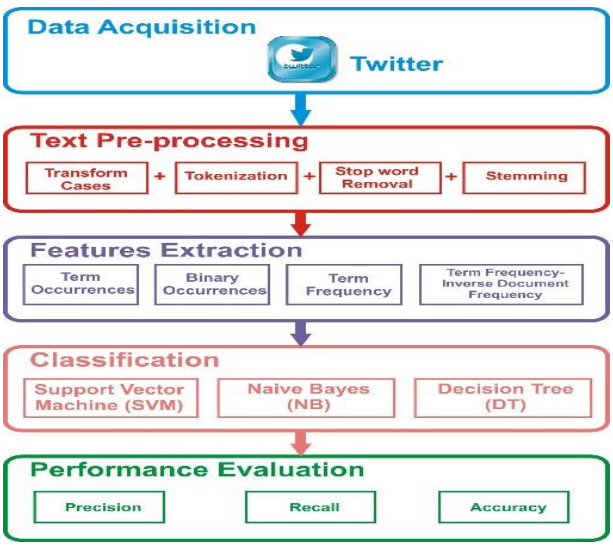

Figure 2: Flowchart of Proposed Framework

The proposed framework consists of data acquisition, text pre-processing, feature extraction, classification, and performance evaluation. Rapid Miner Studio is used for text-pre-processing, feature extraction, and classification. In last, a result-based comparison between three data mining algorithms, such as SVM, NB, and DT is also be mentioned that will be helpful for decision-makers and companies' executives in boosting their product quality and sale, and service. A course of action of each stage of the proposed architecture is described below:

Data Acquisition

This stage engages the collection of raw data from Twitter and an efficient transformation mechanism is utilized to store it in the storage system. For sentiment analysis, a pre-labeled dataset having 4010 tweets about difficulties experienced by passengers in an airline is used. These tweets are classified as under:

Table 2. Twitter dataset of airline

Class	Number of Tweets
Positive	1802
Negative	2208
Total	4010

Text Pre-processing

It is a process of organizing and cleaning data for classification. Customers' sentiments obtained from Twitter are not completely clear and contained many grammatical and spelling errors, which require its necessary pre-processing before transformation. In order to minimize the noise in the text and to progress the feat of the classifier, text pre-processing include the following:

a. Transform Cases

Document having customers' reviews about airline contain capital as well as lower letters, so during text pre-processing, it transform all the characters into either lower or upper case.

b. Tokenization

Customers' reviews about service provided by an airline company are in sentences form that is divided into meaningful words separately and removed certain characters like punctuation marks, Twitter hashtags (e.g. #topic) by tokenization.

c. Stop word removal

These words commonly come across the texts such as "in", "this", "i", "an", "a", "and", "to", "be", etc. which are meaningless in sentiments analysis, so, stop word are removed.

d. Stemming

It converts word into its root form before indexing such as "computation", "computer", "computing", all trim down to compute.

Feature Extraction

It is a method of renovating the key data into a set of characteristics, which played a vital role in the performance of the machine learning process, so, it is crucial. N-gram model provides statistical information for calculating the importance of words. Several n-gram models, such as unigrams, bigrams, and trigrams are applied to evaluate the power of using these n-gram schemes efficiently.

To get the best presentation of the classifier, the computation of the term weighting scheme played a vital job. N-gram is shown as a feature vector which can be generated in the following traditions:

a. Term Occurrences (TO)

It is the absolute number of incidence of a word in the text.

b. Binary Occurrences (BO)

It is the binary representation or occurrence as a binary value, which indicates the presence of word as 1 and 0.

1 if it exists and 0 otherwise

c. Term Frequency (TF) – Inverse Document Frequency (TF-IDF)

It predicts the significance of words/phrase inside the given text in the corpus using following equations from (1) to (4):

$$\Pr ecision(P) = \frac{TP}{(TP + FP)} \quad (1)$$

$$\mathrm{Re}call(R) = \frac{TP}{(TP + FN)} \quad (2)$$

$$Accuracy(A) = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (3)$$

$$F - measure = \frac{2(\Pr ecision \times \mathrm{Re}call)}{(\Pr ecision + \mathrm{Re}call)} \quad (4)$$

Where:
TP = Number of positive data that are correctly classified.
FN = Number of positive data that the incorrectly classified.
FP = Number of negative data that are incorrectly classified.
TN = Number of negative data that are correctly classified.
TP, FP, FN and TN also defined in below mentioned confusion matrix.

Table 3. Confusion Matrix

Classification	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Positive (FP)
Actual Negative	False Negative (FN)	True Negative (TN)

Classification

After the literature review, we knew that various data mining approaches have been utilized by the researchers and found that the NB, SVM, and DT classification algorithms provided the best result as compared to others classifiers; therefore, we have utilized these techniques to classify the dataset.

a. Naive Bayes (NB)

It is a probabilistic representation that based on Bayes' Theorem and statistical independence assumption of random variables instead of calculating full covariance matrix [5]. It provides a great result when using it in for textual information analysis such as Natural Language Processing. This theorem is described as in equation (5).

$$P(c \mid f) = \frac{P(c)P(f \mid c)}{P(f)} \quad (5)$$

Here, f = (f_1, f_2, ...,f_n) representing some 'n' features (independent variables) and 'c' representing the class.

It is used for binary and multi-cases classification. Naive Bayes algorithm has different types like Gaussian, Bernoulli and Multinomial Naive Bayes.

b. Support Vector Machine (SVM)

It is also a supervised machine-learning algorithm which is invented by Vapnik and Chervonenkis. It has become one of the more powerful techniques for classification as well as regression, which are helpful for arithmetical learning theory and supportive in identifying the factor accurately. It has described in input and output format [13], where the input is vector space and the output is positive or negative (0 or 1).
It is a non-probabilistic algorithm, which is being utilized for the separation of data linearly and non-linearly.

c. Decision Tree (DT)

It is a supervised machine-learning model used for both regression and classification. Decision Tree is typically used for attribute selection. It has two nodes i.e. internal and leaf node. An internal node represents an attribute and every attribute has its individual value i.e. true or false. Leaf node represents as class label i.e. positive or negative.

Performance Evaluation

The performance of supervised machine learning algorithms will be measured in term of precision, recall, accuracy.

Experimental Results and Discussion

For model evaluation, a random sample online Twitter data containing 4010 positive and negative reviews about an airline service is taken to train and test the dataset by a ratio of 70:30. Term Occurrences (TO), Binary Occurrences (BO), Term Frequency (TF) and Term Frequency/Inverse Document Frequency (TF/IDF) weighting schemes are used to generate the word vector. To classify the testing dataset as positive and negative, three classifiers such as SVM, NB, and DT is used. In order to make the new predictions accurately, we used cross-validation model, which demonstrate the ability of the proposed system where k-fold cross-validation is implemented in order to find out the efficiency of the model.

i. Performance Evaluation using Support Vector Machine

This algorithm is applied to a sample dataset by utilizing 10-fold cross validation with various kernel types, such as Dot, Radial, Polynomial, Neural, Anova, Epachnenikov, Gaussian-comparison and Multiquadric, and value of C is set to 0.1. TF/IDF weighting scheme is used to generate the word vector. The basic aim of utilizing various kernels is to analyze its performance. After experiments, it observed that Dot, Polynomial, and Anova give the better results as compared to other kernels. However, Anova kernel provides the best performance in c=0.1, results of which are given in Table 4 below:

Table 4. Testing Results of SVM

Word Vector	K	Kernel Type	C=0.1 % of Precision	% of Recall	% of Accuracy
TF/IDF	10	Dot	88.38	65.87	80.7
		Radial	100	4.22	56.96
		Polynomial	81.37	41.45	69.43
		Neural	00.00	00.00	55.06
		Anova	92.69	85.13	90.30
		Epachnenikov	100	4.22	56.96
		Gaussian-combination	00.00	00.00	55.06
		Muliquadric	00.00	00.00	55.06

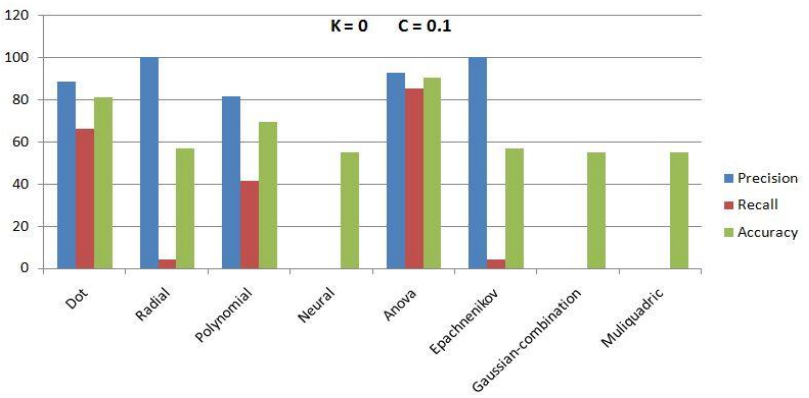

Figure 3: Performance level of different Kernels using TF-IDF

Figure 3 illustrates the performance level of various Kernels such as, Dot, Radial, Polynomial, Neural, Anova, Epachnenikov, Gaussian-comparison and Multiquadric using TF-IDF.

ii. Performance Evaluation using Naïve Bayes

Naïve Bayes algorithm is applied on some dataset by utilizing various k-fold cross-validations. This algorithm gives the utmost accuracy level when TF is used as shown in Table 5 below:

Table 5. Testing Results of Naïve Bayes

Word Vector	K	% of Precision	% of Recall	% of Accuracy
	10	65.13	84.46	72.69
TO	15	65.91	85.29	73.57
	20	65.81	84.79	73.37
	10	64.54	84.63	72.19
BTO	15	65.56	85.35	73.27
	20	65.34	84.22	73.00
	10	69.81	81.74	75.91
TF	15	69.95	82.69	76.26
	20	70.14	81.85	76.19
	10	68.26	79.13	74.09
TF-IDF	15	68.71	79.8	74.59
	20	68.92	79.36	74.64

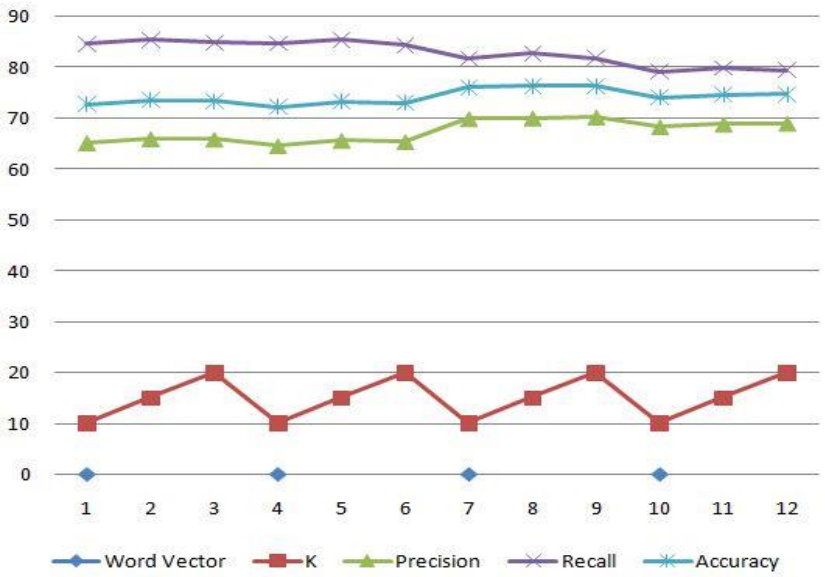

Figure 4: Word vector wise results of each measure

Figure 4 depicts TO, BTO, TF and TF-IDF weighting schemes used to generate the word vector. Word vector wise result of each measure (precision, recall & accuracy) at various value of k.

iii. Performance Evaluation using Decision Tree

Table 6 represents the results obtained by applying this algorithm on the same dataset by utilizing various k-fold cross-validations. Decision Tree classifier gives the best accuracy level when BTO is used.

Table 6. Testing Results of Decision Tree

Word Vector	K	% of Precision	% of Recall	% of Accuracy
	10	93.86	79.8	88.58
TO	15	93.81	79.91	88.61
	20	94.01	80.08	88.75
	10	93.99	79.8	88.63
BTO	15	93.93	79.91	88.66
	20	94.13	80.08	88.8
	10	96.31	73.92	87.01
TF	15	96.59	73.86	87.08
	20	96.23	73.58	86.83
	10	96.71	71.75	86.21
TF-IDF	15	97.00	71.85	86.36
	20	97.71	71.7	86.19

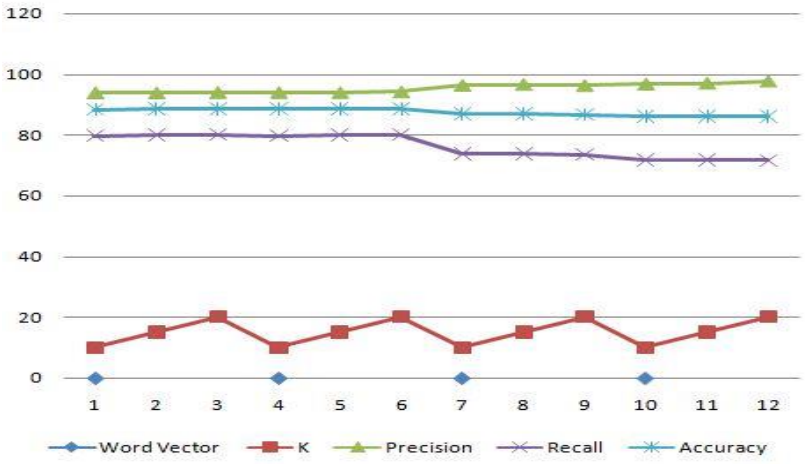

Figure 5: Word vector wise results of each measure

Figure 5 shows TO, BTO, TF and TF-IDF weighting schemes used to generate the word vector. Word vector wise result of each measure (precision, recall & accuracy) at various value of k.

iv. Performance Comparison

Comparison of performance amongst SVM, NB and DT is shown in given Table 7:

Table 7. Comparison of SVM, NB & DT

Classifier	% of Accuracy
Support Vector Machine	90.30
Naïve Bayes	76.26
Decision Tree	88.63

Experimental results showed that Support Vector Machine classifier has the best-predicted accuracy level as compare to other two classifiers such as Naïve Bayes (NB) and Decision Tree (DT).

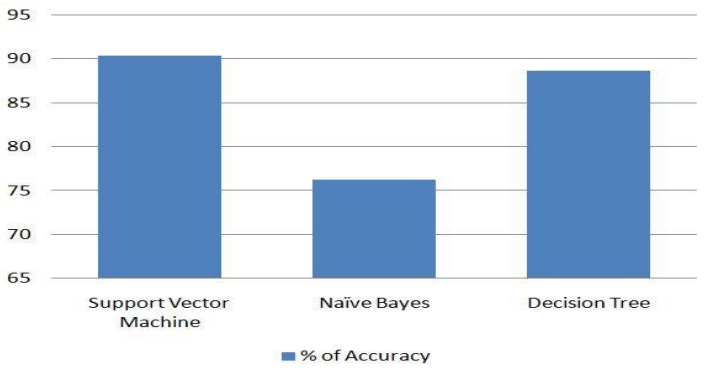

Figure 6: Percentage of Accuracy of SVM, NB & DT

Figure 6 demonstrates the experimental results that Support Vector Machine classifier has the best-predicted accuracy level i.e. 90.30%, as compared to the other two classifiers, such as Naïve Bayes (NB) and Decision Tree (DT).

Conclusion

Nowadays, the corporate sector is giving importance on the sentiments of their customers regarding services or products due to evolutionary changes in expressing their opinion from offline to online. However, extracting particular information from unprocessed and gigantic data is a great challenge for companies' decision-makers. In this research, we propose a framework to analyze customers' sentiments using different data mining techniques. The framework is verified on the data of an airline company. The results show that accuracy of Support Vector Machine (SVM) is greater than other techniques which is 90.30. The obtained results are satisfactory and in future, proposed framework will be tested on diverse and big datasets and will be helpful for researchers and beneficial for market-experts in decision-making.

References

Ahmad, M., S. Aftab & I. Ali. (2017). Sentiment Analysis of Tweets using SVM. *International Journal of Computer Application*. (NY, USA). 177(5): 25-29.

Atia, S. and K. Shaalan. (2015). Increasing the Accuracy of Opinion Mining in Arabic. *IEEE 1st International Conference on Arabic Computational Linguistics (ACLing)*. (Cairo, Egypt). 106-113.

Fegade, P. & D. D. Patil. (2014). Mining Frequent Item sets for Improving the Effectiveness of Marketing and Sales. *International Journal of Computer Science and Information Technologies.* (USA). 5(3): 3736-3738.

Injadat, M.N., F. Salo & A. B. Nassif. (2016). Data Mining Techniques in Social Media: A Survey. *Neuro computing, Elsevier*. (Amsterdam, Netherlands). 214: 654-670.

Jadav, B. M. and V. B. Vaghela. (2016). Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis. *International Journal of Computer Applications.* (NY, USA). 146(13): 26-30.

Joshi, A., N. Pandey, R. Chawla & P. Patil. (2015). Use of Data Mining Techniques to Improve the Effectiveness of Sales and Marketing. *International Journal of Computer Science and Mobile Computing.* (Andhra Pradesh, India) 4(4): 81-87.

Kharde, V. A. & S. S. Sonawane. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*. (NY, USA). 139(11): 5-15.

Kumari, N. & S. Naryan Singh. (2016). Sentiment Analysis on E-commerce Application by using Opinion Mining. *IEEE 6th International Conference on Cloud System and Big Data Engineering (Confluence)*. (Noida, India). 320-325.

Lin, L. L., J. Li, R. Zhang, W. Yu & C. Sun. (2014). Opinion Mining and Sentiment Analysis in Social Network: A Retweeting Structure-aware Approach. *IEEE/ACM 7th International Conference on Utility and Cloud Computing*. (London, UK). 890-895.

Mani, B. Suri & M. Kumar. (2018). Performance Evaluation of Data Mining Techniques. In: Mishra D., Nayak M., Joshi A. (eds) *Information and Communication Technology for Sustainable Development. Lecture Notes in Networks and Systems*, *Vol 9. Springer*, *Singapore*. 375-383. ISBN: 978-981-10-3932-4.

Mars, A. & M. S. Gouider. (2017). Big Data Analysis to Features Opinions Extraction of Customer. *Procedia Computer Science, Elsevier*. (Amsterdam, Netherlands). 112: 906-916.

Patil, G., V. Galande, V. Kekan & K. Dange. (2014). Sentiment Analysis using Support Vector Machine. *International Journal of Innovative Research in Computer and Communication Engineering*. (Chennai, India). 2(1): 2607-2612.

Qadri, Salman., Khan, Dost Muhammad et al. (2017) "A Novel Optimized Land Cover Classification Framework Using Data Mining Techniques", *Sindh Univ. Res. Jour. (Sci. Ser.)*, ISSN: 1813-1743, 49(2). 433-440.

Qadri, Salman., Khan, Dost Muhammad, et al. (2016). "A Comparative Study of Land Cover Classification by Using Multispectral and Texture Data", *Bio Med Research International, (Hindwai Publishing Corporation)*, *ISSN*: 2314-6133, Volume 2016, Article ID 8797438, 12 pages, http://dx.doi.org/10.1155/2016/8797438.

Qadri, Salman., Khan, Dost Muhammad, et al. (2017). "Multisource Data Fusion Framework for Land Use / Land Cover Classification Using Machine Vision", *Journal of Sensors,* ISSN: 1687-725X (Print)ISSN: 1687-7268 (Online), Volume 2017, Article ID 3515418, 8 pages, https://doi.org/10.1155/2017/3515418.

Songpan, W. (2017). The Analysis and Prediction of Customer Review Rating Using Opinion Mining. *IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*. (London, UK). 71-77.

Zainuddin, N. & A. Selamat. (2014). Sentiment Analysis using Support Vector Machine. *IEEE International Conference on Computer and Control Technology (I4CT 2014).* (Langkawi, Malaysia). 333-337.