

Original Article

# Supervised Machine Learning Classifiers: Computation of Best Result of Classification Accuracy

Himanshu Thakur<sup>1</sup>, Aman Kumar Sharma<sup>2</sup>

<sup>1,2</sup> Department Of Computer Science, Himachal Pradesh University, India.

Received Date: 07 September 2020

Revised Date: 09 October 2020

Accepted Date: 15 October 2020

**Abstract** - Sentiment Analysis is one of the fastest spreading research fields in computer science, originating its demand to observe the trace of all the activities in the region. The focus of sentiment analysis is to release data on the features of the author or speaker about an exclusive subject or the total variance of a record next to examine textual data assemble from the countless origin. The indicated paper is conferring an equivalent study to evaluate and formulate a list of three supervised machine learning techniques (Support vector machine, K-Nearest Neighbor, and Random Forest) based on a literature survey that has opted in this research work. To evolve and validate a mechanism to compute better classification accuracy results from among the selected best-supervised machine learning classifiers.

**Keywords** - Sentiment Analysis, Sentiment Classification, Opinion Mining, Feature selection, Machine Learning, Supervised Learning, Support Vector Machine, K- Nearest Neighbor, Random forest, Ensemble learning, Jupyter Notebook.

## I. INTRODUCTION

Sentiment Analysis, also called opinion mining, is the evolution of determining and detecting emotional information using natural language processing, text analysis. Opinion mining is an especially narrow NLP issue because the framework only needs to acknowledge the polarity of emotion, i.e., positive or negative opinion of each sentence and the target entities or aspects. The entity can symbolize individuals, events, or topics. These topics are usually acceptable to be covered by reviews. The two indications of SA or OM are equivalent. They signify the same meaning. However, some researchers stated that OM and SA have marginally singular perceptions [1]. The particular public ideas are acquired from certain network 2.0 and appearance-related blogs, online data sets, movie reviews, and brand review slots [2]. Its use of NLP and arithmetic strategies to automate the set of judgments from ordinarily unstructured

text.OM is a series of strategies and tools for finding and extracting specific information, such as perspective and conclusion, from language [3]. Usually, emotions have been

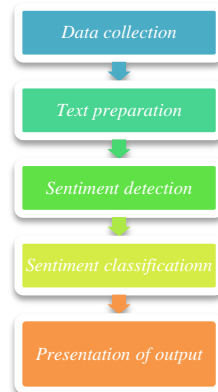


Fig. 1 Sentiment Analysis Process

along with opinion polarity, i.e., positive, neutral, or negative opinions regarding something [4]. Sentiment analysis is a varied transaction with five phases for indexing sentiment data.

Sentiment analysis can occur at different levels: document -level purpose to categorize review as positive, negative, or neutral. The main operation of the document level sentiment classification is to relate the similarity of User Generated Content (UGC). The sentence-level aims to identify the relationship between each sentence is calculated. Sentence-level objective to allocate ideas communicate in an individual sentence. Essentially the first step is to determine even if the sentence is biased or unbiased. Wilson et al. [5] obtain that learning statements are not fundamentally based on nature. After all, inside no symbolic characteristics between document and sentence level arrangement because sentences are brief equitable documents. Aspect-level objective to allocate opinions with respect to specialized aspects of existence. Firstly it classifies



the substance and then aspects because the judgment titleholder can allow the various suggestions for different aspects of the equivalent entity.

Opinion mining is not only enforced on a production survey but can also be activated on traditional business [6, 7], broadcast things [8], or official arguments [9]. Feature selection in sentiment classification. The first step in the SC problem is to take out and select text features. Some of the current features are Term presence and frequency: The present simulation consists of uni-gram or n-Grams and their frequency. *Parts of speech Information:* Identification words to the POS identity help express differently and assist in leading feature selection. Including POS identification, the whole world in order is indicated as a label, and the label implements the format of the word in the grammatical text. *Negation* argument performs an essential appearance to appropriate into functions because negations accept possible to reversal the opinion. *Opinion Words and phrases* are exchanged and codified, examining either a positive or negative perspective. Dictionary-based and numerical based is the fundamental approach to classifying the linguistic direction of sentiment conversation.

Section II distributes a review of related work executed on a comparative study of SA approaches. Section III illustrate different SC techniques like Machine Learning, Lexicon Based Approach. The indicated paper only spotlights the supervised machine learning approach used for sentiment analysis and also grants a short introduction to other approaches. Section IV explains the methodology adopted. Section V portrays the simulation results of the experimental work. Section VI concludes the present research work, and Section VII is rapid describes the future scope of the study.

## II. RELATED WORK

Opinion Mining originated its roadmap across the year 2000 & proceeding improvement in the research area over the last ten to fifteen years. OM is a sub-area of NLP, but it acquires that SA gets a more advancing trend than NLP in the earlier years. There are multiple articles introduced every year in the SA field, and the number of articles has been increasing over the years. The reader can find some refined and detailed surveys, including [10, 11]. Many research papers and survey papers have been published relevant to SA research work, and a few of the extensive papers from which the current study gets inspired are examined below:

Rehab M. Duwairi and Islam Qarqaz [12] carry out an identical inquiry of three classifiers: Naive Bayes, SVM, and K-NN. They organize their intimate training dataset by assembling tweets and Facebook comments taken away from the internet. They form the use of the Rapid miner tool for this approaching study and regulate that SVM delivers the highest Precision, although K-NN contributes the highest recall.

Vinodhini and RM Chandrasekaran [13] provide a complete survey on SA and OM in their paper. The present

survey paper talks about several opinion classification strategies and their challenges. The indicated study compared certain other papers written on SA and OM approaches based on four independent features accuracy, Precision, recall, and F1.

Nasukawa and Yi [14] represent an opinion mining approach to take out sentiments associated with polarities of positive or negative for special subjects against a document, quite of classifying the entire document nearing positive or negative. The mandatory topics in opinion mining recognise how sentiments are indicated in texts and whether the expressions indicate favourable or unfavourable sentiment about the subject.

Ding et al. [15] prefer an essential technique for recognizing semantic orientations of emotion extract by reviewers on outcome features. It is accomplished to examine two major problems with the existing methods, (1) sentiment words that semantic directions are situation dependent, and (2) combined multiple opinion words in a similar sentence. For (1), a holistic technique is suggested that can perfectly infer the semantic orientation of emotion based on the rating situation. For (2), a new function to combine multiple sentiments in the same sentence is present.

Van de Camp and van den Bosch. [16] introduce SVM and NN also used to categorise particular relations in biographical texts and establish historical biographical information expressing people in a specific state, zone, and time frame. They specify that their classifiers can label certain relations over a majority class baseline score. They show that SVM and one layer NN (1-NN) algorithm attain the highest scores.

Rui and Liu. [17] In their work, they were investigating whether and how twitter word of mouth (WOM) attacks movie sales by evaluating a dynamic panel data model. They used NB and SVM for the classification objective. Their results suggest that the effect of WOM on product sales from Twitter users with more followers is significantly larger than that from Twitter users with fewer followers and establish that the effect of pre-consumption WOM on movie sales is wise than that of post-consumption WOM.

Reyes and Rosso [18] aimed to define a feature model to mean part of the personal knowledge, which is key to such reviews and efforts to narrate salient characteristics of irony. They used NB, SVM, and DT for classification resolve, and the results with the three classifiers are adequate, both in terms of accuracy, as well as Precision, recall, and F-measure.

Cover and Hart [19] show the fault of the nearest neighbour rule is having bounds above by twice the Bayes

error below certain logical undertaking. Also, the error of the general KNN method asymptotically methods that of the Bayes error and can be nearly new to approximate it.

Pang et al. [20] used three supervised machine learning procedures to classify the text named SVM, Naive Bayes, and Maximum Entropy. In this paper, an individual feature selection technique was used, and results pinpoint that achievement of Naive Bayes classifier is gained on small feature set while the performance of SVM is adequate for the large feature set and moreover originate that Maximum Entropy has a better result than Naive Bayes for the large feature set.

Ashmeet Singh and R Sathyaraj [21] correlated three supervised machine learning algorithms that employ Rapid Miner named Naive Bayes, Random Forest, and Decision Tree based on six parameters such as accuracy, precision, recall, true positive, false positive, and f-measure. Two datasets, particularly a short and huge number of the pattern, were used for each classifier. In this paper, it has been established that Naive Bayes results in better achievement alternative two in the smaller dataset, although the Decision Tree is the best match for the large dataset, and Random forest executes as an average in both the cases.

Ashok Badresiya et al. [22] differentiate five review spam detection supervised learning capabilities for their execution. The five techniques were Naive Bayes, SVM, K-NN, Logistic Regression, and Decision Tree. In this relative study, they originate that SVM has greater accuracy (83.19%) than the substitute way and the Decision tree has very little accuracy, only 51.00 %.

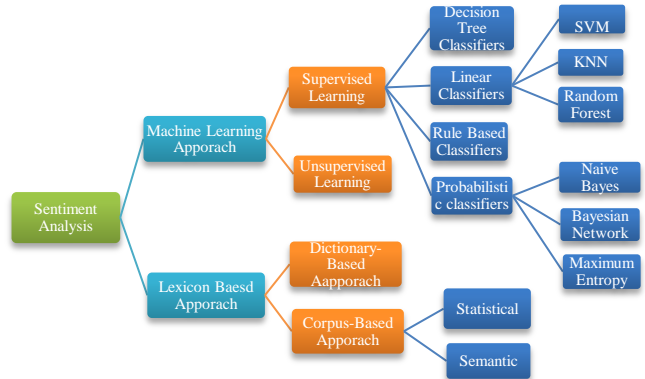
Ghosh, Fassnacht [23] SVM is often claimed to be the finest at dealing with complex classification issues such as tree species differentiation, followed by RF using the information in a broader electromagnetic spectrum (450–2500 nm) focused on classifying five tree species in managed forests of central Germany using SVM and RF on Hyperion and HyMap data. Apply only spectral information. The overall classification accuracies on a 4-m pixel hyperspectral image were 71% for SVM and 72% for RF. Both classifiers showed similar results for this scale.

L. I. Kuncheva and J. J. Rodriguez et al. [24], among the few popular combination schemes, majority voting and weighted voting for classification are generally used. Simple majority voting is a decision rule that selects one of many alternatives based on the predicted classes with the most votes. The majority voting does not require any parameter tuning once the individual classifiers have been trained.

**III. SENTIMENT CLASSIFICATION TECHNIQUES**

Sentiment allocation procedure can be branched into

machine learning techniques, dictionary-based techniques, and hybrid techniques [25]. Classification of text is done towards noticeable levels such as sentence based, document-based, and aspect or sentiment-based [26]. All the approaches and techniques this paper shows a comparison of supervised machine learning techniques, which are illustrated in this region down:



**Fig. 2 Sentiment Classification Techniques**

**A. Machine Learning**

Normally it depends upon the important machine learning approaches to figure out the SA as rectify text distribution drawback that makes use of static or pattern arrival. *Text distribution Problem Definition:* We have a set of training documents  $D = \{Y_1, Y_2, \dots, Y_n\}$  where the individual file is classified into a class. In a machine, learning-based classification requires two sets: One is the Training set, and another one is the Test set. A training set is used to set a mechanical classifier for learning the distinct features. A test set is employed for validating the presence of the classifier. The former is utilized by automated classifiers for learning the differentiable features of documents, while the latter is utilized for validating the performance of the automated classifier. Machine learning facilitates data processors to increase, customize, and determine, although they are identified as a unique input [27]. There are various groups of machine learning techniques classified as supervised and unsupervised learning described below: In supervised learning, the training set  $(X_n, Y_n)$ , an algorithm, train a model that is efficient to predict the output for every new input, as displayed in fig.4. Supervised learning helps classification and regression facilities to develop a predictive model and depends on the labelled training documents containing words related to the topic as a key feature.

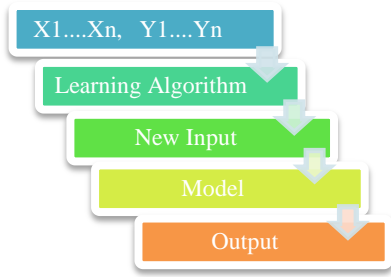


Fig. 3 Supervised Learning

Numerous supervised learning executions that acquire correlated in this paper are characterized below:

A linear classifier classifies an object into specific classes to whatever they apply according to the arrival or attributes. A linear classifier adopts a narrow arrangement of X and Y to create the distribution results. Three linear classifiers used in this study are SVM (Support vector machine), K-NN (K Nearest Neighbors), and RF (Random Forest):

- a) **SVM** In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with correlated learning algorithms that analyze data used for classification and regression analysis. The objective of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH). Support Vector Machines (SVM) [28] examine a must-try, and it offers one of the most robust and accurate methods among all well-known algorithms.
- b) **K-NN** K-nearest neighbour is one of the simplest machine learning algorithms established on a supervised learning strategy and is mostly used for classification problems. K-classification [29,30] finds a group of k objects in training set near the test object and bases the allocation of a label on the predominance of a particular class in this neighbourhood. KNN is specifically well suited for multi-modal classes and applications in which an object can have many class labels. The K Nearest Neighbors Algorithm can feel the necessity for the recall division to inventory group of the knowledge, but only put away prediction (or study) when forecasting is essential.
- c) **Random Forest** (RF) is one of the great machine learning algorithms used for supervised learning; this means learning from labelled data and making predictions established on the learned arrangements. A Random Forest is an ensemble technique efficient to accomplish both regression and classification missions with the use of multiple decision trees and a technique called Bootstrap and Aggregation, frequently known as **bagging**. However, it is mostly used for classification problems. The random Forest algorithm generates decision trees on data samples and then gets the prediction from each of them, and finally determines the

best solution by means of voting. It is an ensemble method that is better than a single decision tree because it reduces over-fitting by averaging the result. In unsupervised learning, this technique is mainly used in the creation or design of trained class labelled documents that are investigated to the most complex method. Unsupervised learning is no pre-classified training samples or training datasets. This technique is used for the document-based clustering analysis as it does not depend on pre-defined labelled training documents.



Fig. 4 Unsupervised Learning

Clustering is the mechanism of combining similar entities and is also used in the classification text. K-means clustering is an unsupervised machine learning algorithm. The goal of this unsupervised machine learning procedure is to find the correlation in the data point and group corresponding data points together.

**B. Lexicon-Based Approach**

The lexicon-based approach depends on finding the opinion lexicon, which is used to analyze the text. There are also opinion phrases and idioms, which are called opinion lexicons. There are two methods in this approach. The dictionary-based approach depends on finding opinion seed words and then searching the dictionary for synonyms and antonyms. The corpus-based approach begins with a seed list of opinion words and then finds other opinion words in a large corpus to help find opinion words with context-specific orientations. This could be done by using statistical or semantic methods.

**IV. METHODOLOGY**

The indicated paper examines three supervised machine learning classifiers, especially SVM (support vector machine), K-NN (K-nearest neighbour), and RF (random forest), based on six parameters (Training Accuracy, Testing Accuracy, Total Accuracy, Recall, Precision, and F1 score). In this consideration, the Jupyter Notebook tool is used to examine these techniques. Jupyter is a part of Anaconda (prepackaged Python distribution).

Evaluation metrics involves using a combination of these individual evaluation metrics to test a model or algorithm. Using evaluation metrics are critical in ensuring that your

model is operating correctly and optimally. Entire parameters are established on Evaluation metrics, resulting in the execution of individual classifiers. These parameters are explained as follows:

- **Training accuracy:** It is usually the accuracy that applies the model to the training dataset.
- **Testing accuracy:** It is the accuracy of testing the data. Normally it is sometimes useful to compare these identify overtraining.
- **Total accuracy:** This parameter defines the number of classifications divided by the total number of documents.

$$Accuracy = \text{Number of correct predictions} / \text{Total number of predictions.}$$

- **Recall:** This parameter defines how many true positives were recalled (found). High Recall indicates the class is correctly recognized (a small number of FN).

$$Recall = \text{Total no. of correctly classified positive examples} / \text{Total no. of positives examples.}$$

- **Precision:** It is also called positive predictive value. We divide the total number of correctly classified positive examples by the total number of predicted positive examples. High Precision indicates that an example labelled as positive is positive (a small number of FP).

$$Precision = \text{Positive correctly classified} / \text{Total predictive positive.}$$

- **F1 Score:** This parameter defines a weighted average or harmonic mean of precision & recall values and calculated as:

$$F = 2pr / (p+r), \text{ where } p = \text{precision} \& r = \text{recall.}$$

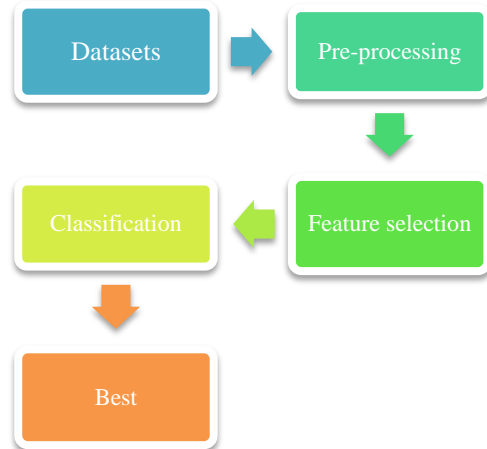
**A. The stepping procedure for identification based on the above-mentioned performance is inclined in fig.5**

In this study, we use seven datasets, and it consists of a collection of documents or files. All the datasets are accessed from Kaagle (ML). The names of the datasets are:

1. Iris dataset {Range Index: 150 entries (0 to 149) & Data column: 6}. [31]
2. Heart Disease {Range Index: 303 entries (0 to 302) & Data column: 14}. [32]
3. Breast Cancer Analysis and Prediction {Range Index: 599 entries (0 to 598) & Data column: 33}. [33]
4. Bank Marketing + Classification {Range Index: 41188 entries (0 to 41187) & Data column: 21}. [34]
5. Fake News Detection {Range Index: 10240 entries (0 to 10239) & Data column: 2}. [35]

6. COVID-19 {Range Index: 368 entries (0 to 367) & Data column: 8}. [36]
7. Instagram fake spammer genuine accounts {Range Index: 120 entries (0 to 119) & Data column: 12}. [37]

- All datasets consist of a collection of documents or files.
- Data preprocessing (clean and alter) the raw data in a useful and efficient format. Now use feature selections to enable the machine learning algorithms to train faster. It reduces the complexity of a model and improves its accuracy of the model.
- After this, use Label Encoder to convert the labels into a numeric form into a machine-readable form.
- In classifications, for each dataset, training the model by using machine learning techniques to find better accuracy.
- Use ensemble methods that combine the prediction of multiple machine learning models to improve predictions based on their highest probability.



**Fig. 5 Task Accomplishment Methodology**

**V. SIMULATION RESULTS**

Simulation Results of Supervised Learning Techniques on various Datasets show the performance by utilizing their merits using the highest probability of being predicted by each classifier using ensemble methods.

**Table 1. Simulation Results of Supervised Learning Techniques on various Datasets.**

Algorithms	Iris	Heart Disease	Breast Cancer	Banking Marketing	Fake News Detection	COVID-19	Instagram fake spammer
SVM	1	81.97	97.87	85.56	54.58	71.31	50.00
KNN	1	90.16	96.27	90.48	55.08	95.90	80.00
RF	1	85.25	95.21	90.97	1	95.90	92.50
<b>ENSEMBLE / VC</b>	<b>1</b>	<b>90.16</b>	<b>97.87</b>	<b>90.97</b>	<b>1</b>	<b>95.90</b>	<b>92.50</b>

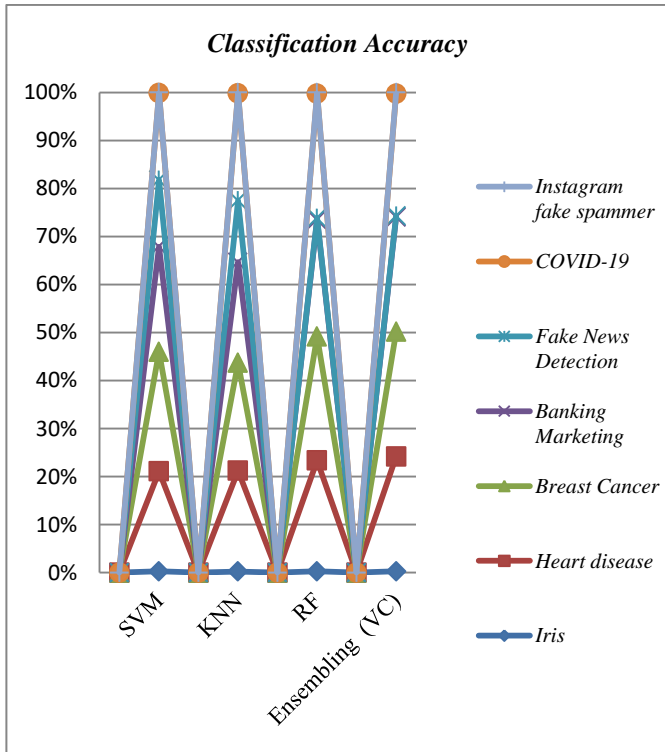


Fig. 6 Classification Accuracy on the line graph

Figure 6 shows the following results:

- Using simulation on various datasets, we have different Classification accuracy of SVM, KNN, and RF.
- They used ensemble learning to predict an output (class) based on the highest probability of chosen class as the output.
- It simply aggregates the findings of each classifier passed into Voting Classifier and compute the output class based on the highest majority of voting.

## VI. CONCLUSION

Sentiment Analysis is a technique meanwhile that the polarity of unregulated textual data is resolute. Sentiment analysis has many applications in information systems, including classifying reviews, summarizing reviews, and other real-time applications. Sentiment Analysis is a part of Natural Language Processing. This research work describes a specific performance on sentiment analysis and various supervised machine learning algorithms used in Opinion Mining. This research work covers every aspect to understand the concept of sentiment analysis and its approaches. This work has considered three sentiment classification techniques, such as Support vector machine, k-Nearest Neighbor, and Random Forest, and has compared them using simulation in python. All the techniques are compared using simulation based on six parameters (Training Accuracy, Testing Accuracy, Total Accuracy, Recall, Precision, and F1 score) for different training datasets, which are used to train and test the various classification models that have opted. This paper gives a detailed stepwise step experimental implementation of the research work. Various results have been analyzed for all the sentiment classification techniques and conclude that:

- I am using ensemble learning to enhance machine learning models to produce better predictive performance compared to a single model, i.e., Improvement in predictive accuracy.
- It has been discovered that ensembles are often much more accurate than the individual classifiers, which make them up and also known as committee-based learning or learning multiple classifier systems, train multiple assumptions to clarify the close trouble.
- Generating models from data is called learning or training, and the learned model can be called a hypothesis or learner. The training algorithms that formulate a set of classifiers and later classify new data points by catching a choice of their predictions are known as Ensemble methods.
- Ensemble methods are well known for their ability to boost weak learners.

## VII. FUTURE SCOPE

This study compares the Supervised Learning techniques that are used in Sentiment analysis. All these techniques are compared for their performance using simulation, but more future work is needed to improve the performance of such techniques further. This is a magnificent demand for the utilization of sentiment analysis because of the aim to reach and truly understand the significance of social media interactions. In the future, different kinds of approaches shall be combined to overcome their drawbacks and enhance their enforcement by utilizing their merits. In the future, a multiclass of a sentiment classification such as positive, negative, neutral, and so on might be considered. This will lead to a superior understanding of natural language

sentiments. The concept of Sentiment Analysis can also be added to the computer system being used nowadays like PDA (Personal Digital Assistants), computer systems, automatic cars, etc., which could detect the negative mood of an individual and respond positively.

**REFERENCES**

[1] M. Tsytsarau and T. Palpanas, Survey on mining subjective data on the web, *Journal of Data Mining and Knowledge Discovery*, 24(3), (2012)478-514,  
 [2] A. Kumar and M.T Sebastian, Sentiment Analysis: A Perspective on its Past, Present, and Future *International Journal of Intelligent Systems and Applications* 4(10) (2012)  
 [3] B. Liu, Handbook Chapter: Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing*, Handbook of Natural Language Processing. Marcel Dekker, Inc. New York, NY, USA, (2009)  
 [4] K. Dave, S. Lawrence, and D. M. Pennock, Mining the peanut gallery: Opinion extraction and semantic classification of product reviews, in *Proceedings of the 12th international conference on World Wide Web*, (2003)519–528.  
 [5] T. Wilson, J. Wiebe, and P. Hoffman, Recognizing Contextual Polarity in phrase-level Sentiment Analysis, in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, (2005).  
 [6] L. Yu, J. Wu, P. Chann, and H. Chu, Using a Contextual Entropy Model to Expand Emotion Words and Their Intensity for the Sentiment Classification of Stock Market News, *Knowledge-Based System*, 41 (2013)89–97.  
 [7] M. Hagenau, M. Liebmann and D. Neumann, Automated News Reading: Stock Price Prediction Based on Financial News Using Context-Capturing Features, *Decision Support System*, (2013).  
 [8] T. Xu, Q. Peng and Y. Cheng, Identifying the Semantic Orientation of Terms using S-HAL for Sentiment Analysis, *Knowledge-Based System*, 35 (2012) 279–89.  
 [9] I. Maks and P. Vossen, A Lexicon Model for Deep Sentiment Analysis and Opinion Mining Applications, *Decision Support System*, 53(2012) 680–688.  
 [10] B. Pang and L. Lee, Opinion Mining and Sentiment Analysis, *Found Trends Inform Retrieval*, 2 (2008) 1-2.  
 [11] A. Montoyo, P. Martinez-Barco, and A. Balahur, Subjectivity and Sentiment Analysis: An Overview of The Current State of The Area and Envisaged Developments, *Decision Support System*, 53(4) (2012).  
 [12] R. M. Duwairi and I. Qarqaz, Arabic Sentiment Analysis using Supervised Classification, *Future Internet of Things and Cloud (FiCloud)*, in *International Conference on IEEE*, (2014).  
 [13] G. Vinodhini and R. M. Chandrasekaran, Sentiment Analysis and Opinion Mining: A Survey, *International Journal of Computer Applications*, (2012) 282-292,  
 [14] T. Nasukawa, Sentiment Analysis: Capturing Favorability Using Natural Language Processing, *Definition of Sentiment Expressions*, (2003)70–77.

**Web References**

[1] <https://www.kaggle.com/jchen2186/machine-learning-with-iris-dataset>  
 [2] <https://www.kaggle.com/cdabakoglu/heart-disease-classifications-machine-learning>  
 [3] <https://www.kaggle.com/buddhiniw/breast-cancer-prediction>  
 [4] <https://www.kaggle.com/prakharrathi25/banking-dataset>

[15] X. Ding, S. M. Street, B. Liu, S. M. Street, P. S. Yu, and S. M. Street, A Holistic Lexicon-Based Approach to Opinion Mining, *ACM*, (2008)231–239.  
 [16] Van de Camp Matje, van den Bosch Antal, The socialist network, *Decision Support System*, 53(2012)761–769.  
 [17] H. Rui, Y. Liu, and A. Whinston, Whose and What Chatter Matters? The Effect of Tweets on Movie Sales, *Decision Support System*, 55(4)(2013).  
 [18] A. Reyes and P. Rosso, Making Objective Decisions from Subjective Data: Detecting Irony in Customer Reviews, *Decision Support System*, 53(2012)754–760.  
 [19] T. Cover and P. Hart, Nearest Neighbor Pattern Classification, *IEEE Transaction Information Theory*, 13(1)(1967) 21–27.  
 [20] B. Pang, L. Lee, and S. Vaithyanathan, Thumbs Up? Sentiment Classification using Machine Learning Techniques, in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, (2002).  
 [21] A. Singh and R. Sathyaraj, A Comparison between Classification Algorithms on Different Datasets Methodologies Using Rapid Miner, *International Journal of Advanced Research in Computer and Communication Engineering*, 5(2016).  
 [22] A. Badresiya, S. Vohra and J. Teraiya, Performance Analysis of Supervised Techniques for Review Spam Detection, *International Journal of Advanced Networking Applications*, (2014).  
 [23] A. Ghosh, E. F. Fassnacht, P. K. Joshi, and B. Koch, A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales, *International Journal of Applied Earth Observation and Geoinformation*, 26(2014)49– 63.  
 [24] L. I. Kuncheva and J. J. Rodriguez, A weighted voting framework for classifiers ensembles, *Knowledge and Information Systems*, 38(2)(2014) 259–275.  
 [25] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, Indexing by Latent Semantic Analysis, *Journal of the American Society for Information Science*, (1990).  
 [26] A. D’Andrea, F. Ferri, P. Griffoni, and T. Guzzo, Approaches, Tools and Applications for Sentiment Analysis Implementation, *International Journal of Computer Applications*, 125(3) (2015).  
 [27] <https://www.simplilearn.com/what-is-machine-learning-and-why-it-matters-article>  
 [28] V. Vapnik, *The nature of statistical learning theory*. Springer, New York., (2012).  
 [29] E. Fix and J. L. Hodges, Discriminatory Analysis, *Nonparametric Discrimination*, USAF School of Aviation Medicine, Randolph Field, Tex., Project 21-49-004, Rept. 4(1951).  
 [30] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Pearson Addison-Wesley, (2006).  
 [31] PriyankaNamdev, Prof. Lakhan Singh A Survey of Sentiment Analysis Process and Technologies *International Journal of Engineering Trends and Technology* 67(11) (2019)153-156.  
 [32] Srinivasan Suresh, Prediction of Average Annual Daily Traffic Using Machine Learning Methods *SSRG International Journal of Computer Science and Engineering* 6(11) (2019) 51-54.

marketing-targets

[5] <https://www.kaggle.com/ksaivenketpatro/fake-news-detection-dataset>  
 [6] <https://www.kaggle.com/brendaso/2019-coronavirus-dataset-01212020-01262020>  
 [7] <https://www.kaggle.com/free4ever1/instagram-fake-spammer-genuine-accounts>