An Approach To Building Models For Sentiment Analysis Using Machine Learning

[1]Deepak Singh, [2] Dr. Anuradha Dandnaik

[1]Research Scholar, [2]Associate Professor
[1]Computer Management,
[1]NWIMSR, Pune, INDIA

Abstract: In this paper the author discusses about forming a model for business to create a framework. This involves number of steps from creation or collection of data, cleaning or pre-processing of data, training the data and testing the data. This paper is written for sentiment analysis of customer in various sectors like education, hospitality and telecommunication. The consumer expression verbally or on social platforms influence consumers and consumer buying patterns. E-commerce giants like shop clues, Flip-kart, Amazon etc. provide a platform or review section to consumers to share their experience and provide real insights about the performance of the product to future buyers also the consumers they express their views happiness or sadness or dissatisfaction or product support. The business houses use these reviews to analyze the consumer sentiment into negative and positive sentiment.

Index Terms - **Education, Hospitality, Telecommunication, Sentiment analysis, Machine learning, Models.**

I. INTRODUCTION

In Today's world the "customer is King" i.e. consumers define the decision and strategy making for the business houses and Managers. The consumer expression verbally or on social platforms influence consumers and consumer buying patterns. E-commerce giants like shop clues, Flip-kart, Amazon etc. provide a platform or review section to consumers to share their experience and provide real insights about the performance of the product to future buyers also the consumers they express their views happiness or sadness or dissatisfaction or product support. The business houses use these reviews to analyze the consumer sentiment into negative and positive sentiment.

Segregating of views on the basis of their sentiment of consumer can help future consumers and buyers to understand and evaluate positive and negative feedback constructively and reach to better decisions as per their requirements. This evaluation acts as a testimonial to the users who are looking to know the details and specifications of the smart phones; thereby increasing user credibility.[1]

Today's the business houses implement the AI in their decision making and strategy formation where the data collected is converted into a model to have faster sentiment analysis and immediate implementation in the business.

In the proposed work reviews have been classified into positive and negative sentiments using Sentiment Analysis.

This paper deals with steps involved in the model creation for the consumer sentiment analysis

II. GOAL AND METHOD

The literature review is done to recognize the applications and solutions of sentimental analysis for the analyzing and classifying of views using Machine Learning (ML) approaches. These techniques are used to find various solutions to the problems found. It also requires certain classification and method to evaluate the solutions to certain limitations. Finally, the review ends with the stability and significance of the proof. The following are few research questions:

- RQ1: What are sentimental analysis applications used Machine learning techniques?
- RQ2: How much activity was carried out recently?
- RQ3: What are the different classification, methodology and techniques that were used?
- RQ4: How do the different techniques prove themselves as best?

III. SENTIMENT CLASSIFICATION

Sentiment analysis is an automated method of determining whether a usage-produced text conveys a positive, negative or common view of an object (i.e. the item, the individual, the subject, the case, etc.). Sentiment classification can be achieved at the four levels such as Document level, Sentence level, and Aspect or Feature level [2].

A. Document-level: The Document-level uses the entire documents to categorize it into a positive or negative class as a simple information category.

B. Sentence level: In the Sentence level, the sentiment classification creates categories of any sentence as subjective or objective, and then it categorize into a positive, negative, or common class.

C. Aspect or Feature level: This type of sentiment classification discusses the identification and extraction of item features from source data.

IV. APPROACHES FOR SENTIMENT ANALYSIS

A. Lexicon based approach: A dictionary containing both positive and negative terms used by Lexicon is applied to assess the polarity of opinion.

(a) The count of optimistic and pessimistic words is discussed in the text. If the text is more positive, a positive score will be assigned to the text.

(b)The text is awarded a negative score if it has high amount of negative or pessimistic words. If the text contains the same number of good and bad terms, a neutral score is given.

(c) A lexicon of opinion (positive and negative opinions) is developed to finalize the word is positive or negative. [3]

1. Dictionary-based approach: A small number of words of opinion with established guidelines are gathered manually [3]. In features like WordNet or thesaurus, similar words and opposite from these words are then searched and appended to the group. The collection lowers down slowly until there are no new terms. This method has the inconvenience of depending on the dictionary, the intensity of the sentiment classification. [4].

2. Corpus-based approach: They depend on large corporations for syntactic and semantic opinion patterns. The created words are context-specific and it needs a larger dataset labeled. [4]

3. Corpus-based approach: They depend on large corporations for syntactic and semantic opinion patterns. The created words are context-specific and it needs a larger dataset labeled. [4]

B. Machine learning-based approach: Machine learning techniques are used in the classification of sentiment depends on the use of well-known machine learning technology on text data. The classification of the sentiment based on machine learning can be put into category of primarily into supervised and unsupervised methods of learning Sample Model. [5]

1. Supervised learning: Supervised learning: Supervised methods of learning rely on labeled training manuals. Supervised learning is an effective classification method and has been used with very promising results for classifying opinions. The regularly used supervised classification techniques in sentiment analysis are Support Vector Machine (SVM), Naïve Bayes (NB) Maximum Entropy (ME), and Artificial Neural Network (NN) and Decision Tree (DT) classifiers. Some less commonly used algorithms are Logistic Regression (LR), K-Nearest Neighbor (KNN), Random Forest (RF), and Bayesian Network (BN).[5].
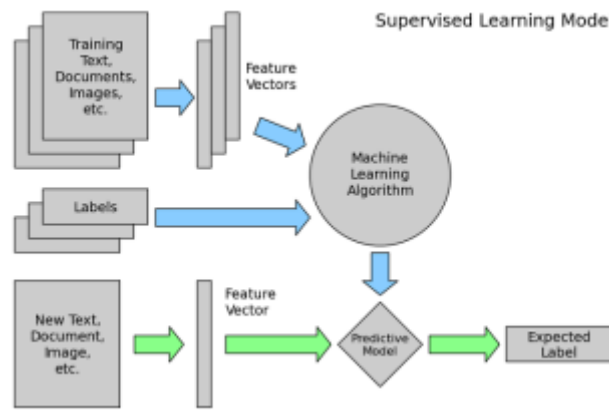
Fig 1. Supervised Learning Model

2. Unsupervised learning: This technique does not use pre-listed data to train the classifier, unlike supervised learning. The more common instance of unsupervised machine learning algorithms are K-means and Apriori Algorithms. Unsupervised machine learning may also be divided into clusters and associations [6].
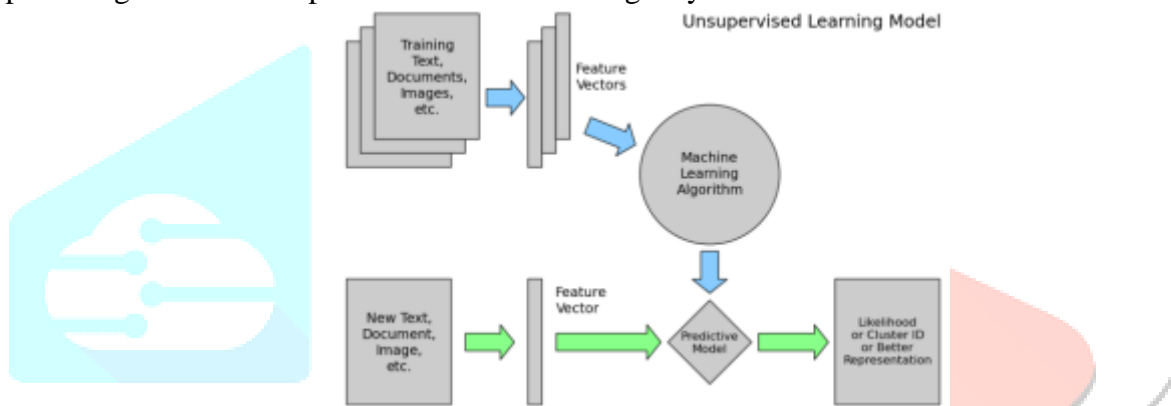

Fig 2. Unsupervised Learning Model

C. Hybrid based approach: The hybrid-based approach uses both ML and lexicon-based classification approach. Few research techniques propose a mixture of lexicon-based and automated learning techniques to enhance the classification of sentiment. This hybrid approach is primarily advantageous as it can achieve the best of both. The combination of Lexicon and Learning has demonstrated increased accuracy. [4]
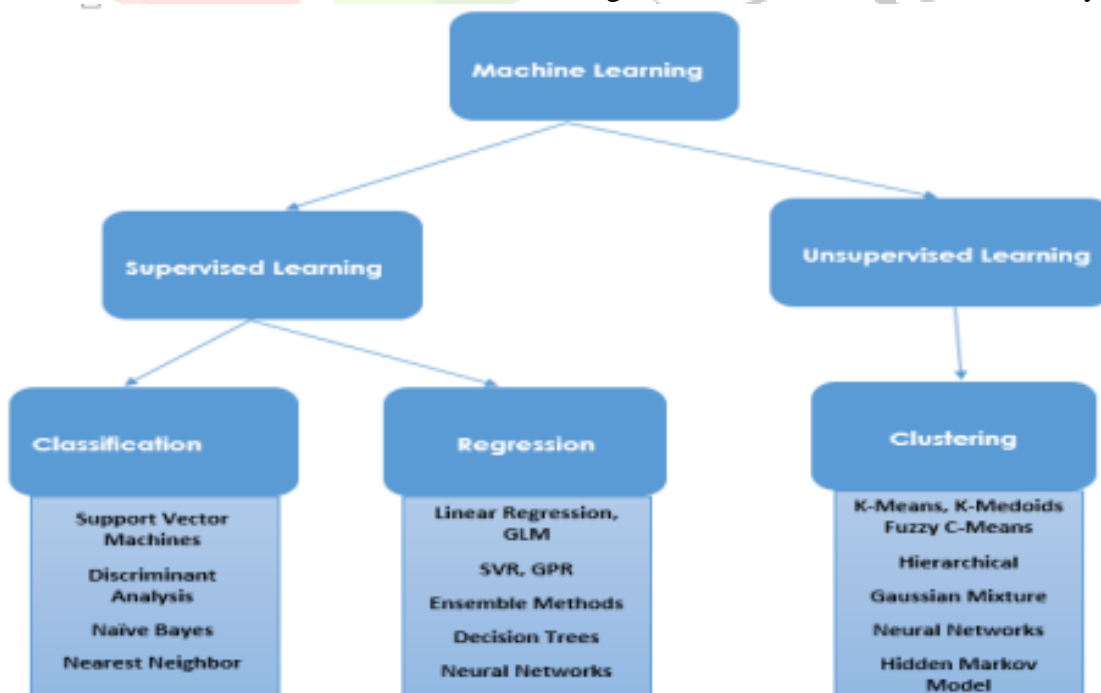

Fig 3. Machine Learning Based Approach (Medhat et al.,2014)[3]

V. MACHINE LEARNING BASED SENTIMENT ANALYSIS METHODOLOGY

The polarity of an analysis data is calculated by different methods. Machine learning Basic sentiment analysis technique is the most popular and efficient technique. As discussed, the polarity in analysis data and the most successful algorithm are calculated. [7]

1.   Data collection- For any kind of text classification task-specific in size as to the number of words, data sets can be used. Such data sets were used after slight preprocessing for sentiment analysis such as case folding, word deletion, etc.

2.   Data Pre-processing-This pre-processing phase seeks to prepare text data for further processing.

3.   Feature Selection and Feature Vector Construction-A computer is not able to process text data straight away, which is an inherent problem. Text data must also be numerically interpreted. Terms are usually used as the characteristics of the text. This gives the text representation a high dimension. Features need to be filtered to reduce dimensions and remove noise to improve classification performance and processing efficiency.

4.   Classification Algorithms for Sentiment Analysis-Several popular and commonly used classification algorithms such as the Multinomial Naïve Bayes Algorithm or the K-Nearest Neighbor Algorithm are commonly used to identify sentiment polarity of users' opinions based on given opinion data Support Vector Machines Algorithm**.**

5. Evaluation Metrics-Measuring every algorithm's output using parameters such as confusion matrix, efficiency, recall, and F-Score.
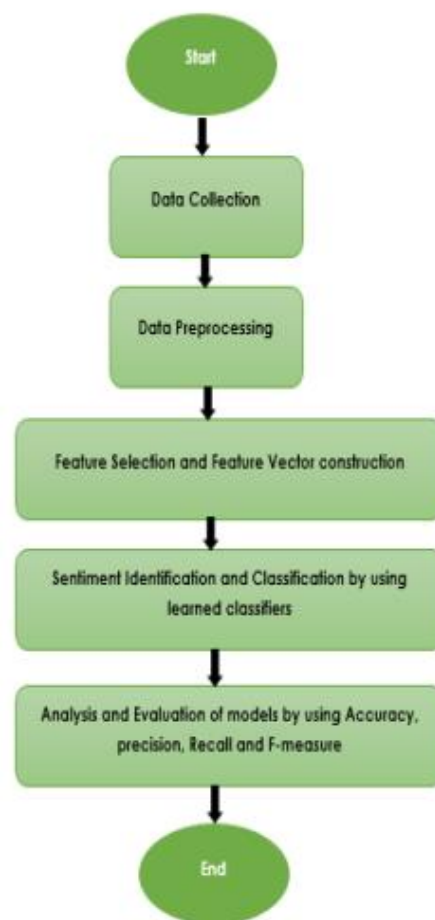

Fig 4: Flowchart of Machine Learning Based sentiment analysis Technique*Yogi & Paudel,2020) [7]

VI. ARCHITECTURE OF SENTIMENTAL ANALYSIS FOR SOCIAL MEDIA ANALYTICS

The area of feelings investigation which investigates feedback, customer opinion, feeling assessment, writing emotions, and attitudes. The review of the product as many is negative and positive is a decision-making process. The key method of the rejection norm and the classification of negative and positive feelings earned by the users or customers in the social community. [8]

Data Pre-processing: The preprocessing technique is more useful to identify and remove meaningless, noisy, and incompatible
data.

Eliminating URLs: URLs will not help to inspect the emotion in the non-formal text.

Questions: The terms question will not help to reduce the ambiguity of polarity such as when, when, who, how, etc.

Removing Special Characters: Special characteristics like, and, (), [] {}/' are separated to eliminate inconsistencies by the function

of polarity.

Removal of Retweets: The re-tweeting process doubles the tweet of another user and redistributes it to peers. This also happens

whenever a user tries to tweet another user. Re-tweets are normally reduced.
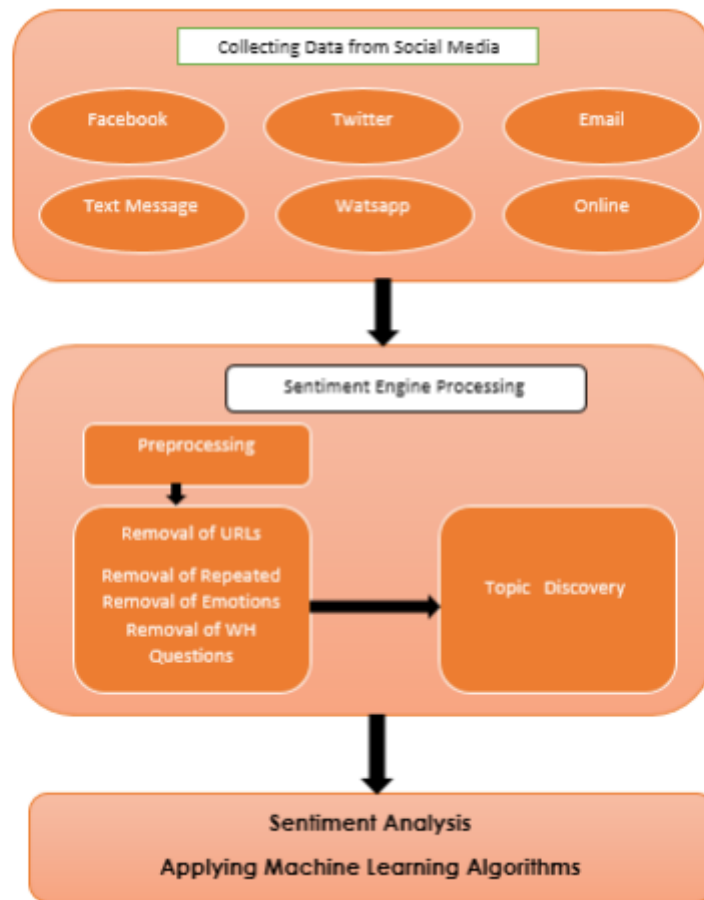

Fig 5: Architecture of Sentimental Analysis(Mahendran & Mekala,2018)[8]

REFERENCES

[1] Zeenia Singla, Sukhchandan Randhawa, Sushma Jain," Sentiment Analysis of Customer Product Reviews Using Machine Learning", 2017 International Conference on Intelligent Computing and Control (I2C2).

[2] Vohra, S., &Teraiya, J. (2013). A Comparative Study of Sentiment Analysis Techniques. International Journal of Information, Knowledge and Research in Computer Engineering, 2(2),313-317..

[3] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113. DOI: https://doi.org/10.1016/j.asej.2014.04.011

[4] Jain, A. P., & Dandannavar, P. (2016). Application of machine learning techniques to sentiment analysis. Second International Conference on Applied and Theoretical Computing and Communication Technology (ICATccT), 1(1). 628–632. DOI: https://doi.org/10.1109/ICATCCT.2016.7912076

[5] Aydogan, E., & Akcayol, M. A. (2016). A comprehensive survey for sentiment analysis tasks using machine learning techniques. 2016 International Symposium on Innovations in Intelligent Systems and Applications (INISTA,.1(1) 1–7. DOI: https://doi.org/10.1109/INISTA.2016.7571856

[6] Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. (2017). Machine learning techniques for sentiment analysis: A review. Iternational journal of Multi-disciplinary science and Engineering, 8(3), 27-35.

[7] Yogi, T. N., &Paudel, N. (2020). Comparative Analysis of Machine Learning Based Classification Algorithms for Sentiment Analysis. International Journal of Innovative Science, Engineering & Technology, 7(6),1-9

[8] Mahendran, N., & Mekala, T. (2018). A survey: Sentiment analysis using machine learning techniques for social media analytics. International Journal of Pure and Applied Mathematics, 118(8), 419–422.