



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: <https://doi.org/10.22214/ijraset.2022.44902>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Opinion Mining For Text Data: An Overview

Ms. Shreya Sharadkumar Shah¹, Dr. Jagdish W Bakal²

^{1,2}Department of Computer Engineering, School of Engineering and Applied Sciences, University of Mumbai, Kalyan, India

Abstract: Nowadays there is huge growth in data. People post their views and opinions through the web on different apps, blogs, articles, etc. Customers post their reviews on shopping sites about the product or service. So, it becomes beneficial for companies, manufactures, business owners and sellers to understand customers, product users or buyers but due to huge data/feedbacks or posted opinions manually analyzing text data, is impossible to do. So, opinion mining is very important so as to analyze all the data and know the sentiments from that data without much human effort and in less time huge data can be analyzed. Many researches have made the base in this field of opinion mining. Here opinion mining will be discussed starting with what is opinion mining, how opinion mining is performed, levels, types and approaches for opinion mining, and applications. Also, methods for Text Preprocessing, Feature Extraction, Evaluation and Classification Approaches that are Machine Learning approaches and Lexicon Based approaches also, various opinion mining methods such as Support Vector machines (SVM), Neural Network, Naïve Bayes, Bayesian Network, Maximum Entropy, Corpus and Dictionary based methods are discussed here.

Keywords: Opinion Mining, Sentiments, Opinions, Machine Learning, Lexicon Based

I. INTRODUCTION

In today's era, everyone is connected to internet. We express ourselves freely on social media by putting posts, status, writing in blogs, articles, news sources, feedbacks and our opinions through apps and sites like Twitter, Facebook, Instagram, we write our reviews on product pages/sites etc. Such a huge amount of data we generate every day, using these reviews and feedbacks sellers, companies, manufacturers, service providers get to know about people feel about their product, service or what opinion people have about situations happening around. So, opinion mining is very important so as to analyze all the data and know the sentiments from that data without much human effort and in less time huge data can be analyzed. Thus, Opinion mining came into picture so as to analyze the opinions or sentiments and take action according to that.[1] Analyzing feedbacks can be beneficial for sellers in other ways also, they can show case this feedback analysis to attract new customers also. Using opinion mining peoples thinking can be understood about things happening around them.[9]. One of the key aspects of any successful business is knowing how customers feel about your brand or products. They often freely express their views and opinions on social media, or in product reviews, surveys, and beyond, providing lots of information about their thoughts and feelings. But with the growing data day by day, it is impossible to manually analyze this huge information. So here opinion mining comes in. This Natural Language Processing (NLP) technique allows you to go beyond mere numbers and statistics, to automatically understand the feelings and emotions of your customers. is the key usage or background for Opinion Mining Opinion mining is used for analyzing textual data and extracting the opinions and sentiments from the data using various natural language processing techniques. Opinion can be positive, negative, neutral, etc. It allows you to understand customer's thoughts and find out what are their likes and dislikes, so according to their needs you can create products and services.

II. APPROACHES FOR OPINION MINING

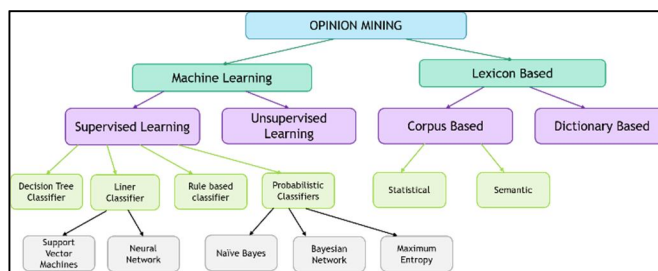


Figure 1: Approaches for Opinion Mining

The above “Figure 1” shows approaches for opinion mining. There are two types of approaches [1] they are Machine learning and Lexicon based approach.[13]

III. MACHINE LEARNING APPROACH

Machine Learning is one of the methods to analyze the data that automates analytical model building.[2] Machine learning could be a branch of AI supported the thought that systems will learn from knowledge, distinguishing patterns and create choices with less human intervention. For sentiment analysis machine learning tools are trained with samples of emotions in text, afterwards machines learn to detect sentiment without human input automatically. So, we can say simply that machine learning [4] allows computers to learn new tasks without being specially programmed to perform them. We can also train sentiment analysis models to understand things like, context, sarcasm and misapplied words. There are various techniques and complex algorithms which are used to train machines and perform sentiment analysis they can be classified under Supervised Learning and Unsupervised Learning.

A. Supervised Learning

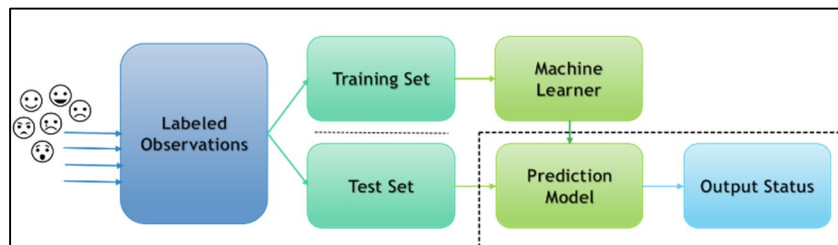


Figure 2: Supervised Learning

Above “Figure 2” shows general supervised learning model. The algorithm in supervised learning model learns with the help of labelled data and using that algorithm performance matrices of test data can be acquired, i.e., as having input and the desired output is known and using that algorithm learns. For example, if we want to recognize fruits then we have already labelled all the data with the fruit names and the learning algorithm receives a set of inputs along with the corresponding correct outputs. The algorithm learns by itself i.e., the algorithm learns by comparing its actual output with correct outputs to find errors. It then modifies the model accordingly. Supervised learning use patterns to predict the values of the label on future unlabeled dataset, through methods like regression, classification, gradient boosting, prediction. Supervised learning is usually used in applications wherever historical data predicts seemingly future events.

1) Support Vector Machine

Support Vector Machine (SVM) is a type of Supervised Learning it comes under Linear Classifier which is used for both classification and regression problems. But generally, SVMs are used for classification problems. In SVM each data is plot as n-dimensional space where n is no of features, with the value of each co-ordinate is the value of each feature. We need to gather perfect data if the data is unbalanced it affects the accuracy. Then classification is performed by finding the hyperplane which differentiates two classes.

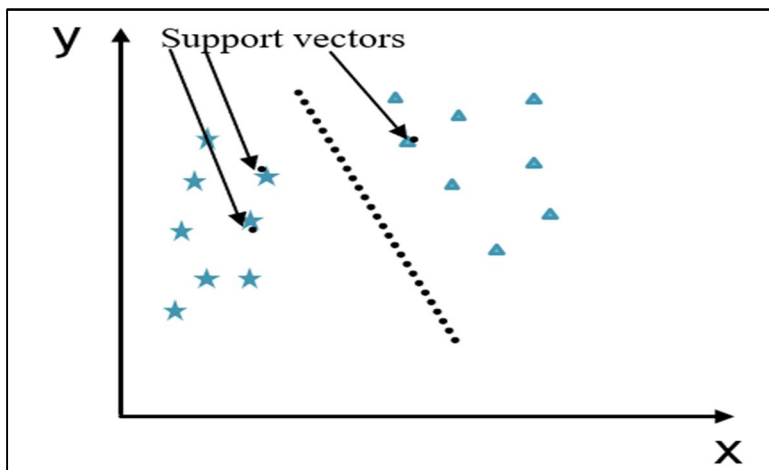


Figure 3: Example of Hyperplane

Followings are the important concepts in SVM. Figure 3 can be referred to understand the concepts discussed below

- a) *Support Vectors* – Support vectors are Datapoints closest to the hyperplane. With these data points separating line will be defined.
- b) *Hyperplane* – Hyper-plane is called the space or decision plane which is divided between a set of objects which are having different classes. Hyper-plane can be seen in the above figure.
- c) *Margin* – Margin is defined as gap between the two lines on the nearest data points of different classes. Margin is calculated as the perpendicular distance from the line to support vectors. If the margin is Large then it is considered as a good margin and if the margin is small it is considered as a bad margin.

The SVM focuses to divide the datasets into different classes and finds maximum marginal hyperplane (MMH). Following steps are performed to divide dataset:

- First, SVM generates hyperplanes one by one iteratively that segregates the classes best way.
- Second, it chooses the hyperplane that separates the classes properly and correctly

2) Neural Network

Neural Network (NN) is a type of supervised Learning it comes under a Linear Classifier. Here classification is done using artificial neural networks. We may not get good results using basic NN. So, we think of Recurrent Neural Networks (RNN) but there is another issue with RNN that is memory so Long Short-Term Memory RNNs (LSTM-RNN) can be a good option for Sentiment Analysis (SA). But for doing sentiment Analysis using Natural Language Processing (NLP) we cannot use the words directly and feed to our NN. There is a need of converting words to vectors for SA using NN.

Data needs to be pre-processed then Tokenize and next WordEmbedding can be performed. WordEmbedding can be used to transform words to vectors that will represent words in multidimensional space. After Vectorization we can apply our data to our NN Model.

Deep Learning also can be used for sentiment analysis. Deep learning is hierarchical machine learning that uses multiple algorithms in a very progressive chain of events to resolve advanced issues and permits you to tackle huge data, accurately and with very little human intervention. But deep learning models are very difficult and time taking to train.

3) Naïve Bayes

Naïve Bayes is Type of supervised learning it comes under Probabilistic classifiers. We need to pre-process the data which we want to classify since Naïve Bayes uses probability, we can use Bag of Words (BOW) for feature extraction. BOW is concerned with how many times the word is repeated instead of keeping the sequence of words. If we used single word for classification, it is mono-gram model if two words taken at a time it is Bi-gram model and for more it is n-gram model then Document Term Matrix is generated (DTM) and later we define, compile and fit the model.

Naïve Bayes model uses bayes theorem with Naïve assumptions of no relationship between the features.

Bayes theorem states that: $P(X|Y) = P(Y|X) * P(X)/P(Y)$

Naïve bayes model works well with text classification. It can train and learn with small amount of data and requires less time to train. But it cannot be used for regression problems. It is a frequently used method.

4) Bayesian Network

Bayesian Network is Type of supervised learning it comes under Probabilistic classifiers. It is a graphical representation of probabilistic relationships between some random variables. Every node in a Bayesian Network represents a random variable, and each edge represents the dependency between the random variables.

The probability distribution of every node in Bayesian Network is estimated by parameter learning based on data. To learn Structure learning in Bayesian Network Structure learning is used.

Bayesian Network model types are Full Bayesian Network (FBN), Bayesian Network with Mood Indicator (BNM). In Full Bayesian Network, words are used as nodes with the addition of one class node. Bayesian Network with Mood Indicator is an improvement for Multinomial Naive Bayes by connecting indicator nodes to the class node.

5) *Maximum Entropy*

Maximum Entropy is Type of supervised learning it comes under Probabilistic classifiers which belongs to class of exponential models. Maximum entropy does not assume that features are conditionally independent of each other. Due to less Assumptions Maximum entropy makes we can use it when we don't know about previous distributions and when it is unsafe to make any assumptions. It requires more time to train because of optimization problem that takes more time to solve for estimating parameters of the model. After computation of these parameters the method provides robust results.

IV. UNSUPERVISED LEARNING

In unsupervised model, unlabeled data is provided which the algorithm tries to understand by extracting features and patterns by itself. Unsupervised learning can be used for data that do not have historical labels. The system is not given the right answer. The algorithm should figure out on its own what is being showed. The aim is to examine the dataset and find some structure within. Unsupervised learning works well on transactional data. For instance, it will identify segments of shoppers with similar attributes then they can be treated similarly in marketing and promoting campaigns. Or it will notice the main attributes that separate client segments from each other. Unsupervised algorithms can also be used to identify data outliers, segmentation of text topics, recommend items.

Popular unsupervised learning techniques include:

- self-organizing maps
- nearest-neighbor mapping
- k-means clustering
- singular value decomposition

Lexicon Based Approach

List of predefined words are used in lexicon-based approach, where each word is related to a particular sentiment. Lexicon-based [14] methods are simple and efficient. In this approach using a basic algorithm sentiment lexicon assigns a polarity value to each text document. A sentiment lexicon is made up of list of lexical features (example: words, phrase, etc.) which are labelled as either positive or negative according to their semantic orientation/polarity. There are three ways of making sentiment Lexicon: Hand-Craft Elaboration, Automatic Expansion from initial list of seed words, Corpus Based Approaches.[4]

For Implementation of lexicon-based approach no prior training is required. Its execution is faster. It can also work with small dataset. Limitation of lexicon-based approach is sometimes it can incorrectly do sentiment scoring of opinion words [3] using existing lexicons such as SentiWordNet. Parts of Speech (POS) tagging is important part for lexicon-based approach.

A. *Dictionary Based Approach*

Dictionary-based approach for sentiment analysis is a computational approach for measuring the feeling that a text conveys. In the simplest way, sentiment has a binary classification: positive sentiments or negative sentiments, but it can be extended to multiple dimensions such as fear, sadness, anger, joy, etc. This method relies heavily on a pre-defined list (or dictionary) of words full of sentiments. Firstly, we need to do all the data pre-processing then we need to download the dictionary for performing sentiment analysis, we can use NLTK for that.

B. *Corpus Based Approach*

Corpus based Approach uses seed sentiment words and find another sentiment word then their polarity from the given corpus. The Corpus-based approach helps in solving the problem to find opinion words with context specific orientations. Corpus based methods relies upon syntactic patterns or patterns that occur along at the side with a seed list of opinion words to seek out different opinion words in an exceedingly large corpus. Corpus-based approach alone is much effective as the dictionary-based approach because it is difficult to prepare a huge corpus for covering all the English words, but this approach has an important advantage that can help to find domain and context specific opinion words and their orientations using a domain corpus. The corpus-based approach can be performed using two ways they are: statistical approach or semantic approach.

1) *Statistical Approach*

Co-occurrence patterns or seed opinion words can be found by doing use of statistical techniques. It is attainable to use the complete set of indexed documents available on the web as the corpus for construction of dictionary. Thus, overcomes the problem of the unavailability of few words if the used corpus is not huge enough.

2) Semantic Approach

The Semantic approach relies on different principles for finding the similarity between words and gives sentiment values directly. Using this principle similar sentiment values is given to semantically close words. WordNet for instance provides different kinds of semantic relationships between words that are used to calculate the sentiment polarities. WordNet can also be used for obtaining a list of sentiment words by iteratively growing the initial set with synonyms and antonyms and then finding the sentiment polarity.

V. PHASES OF OPINION MINING

In the “Figure 3” below shows general steps for opinion mining. Data is collected in the form of text. But the original data contains lots of unimportant things so the data must be filtered, pre-processed. Originally the input text data contains lots of unimportant and unrequired words such as stopwords, repeating words, emojis etc which are filtered i.e., removed. Then the filtered text data is pre-processed using stemming, lemmatization. Then feature selection and extraction techniques used on the text data. Also, negation handling can be considered for views such as “not good”, “not bad” etc. And after that we get proper pre-processed data which is used for further data analyzing operations. Firstly, data must be divided into test and train then we do sentiment classification on train data. After model is built, apply the model on our train data to validate the model. Finally, evaluation of obtained results is performed to determine overall accuracy of the techniques used for Opinion Mining. There are various techniques and methods for pre-processing and classification we can use any of them to build our model.

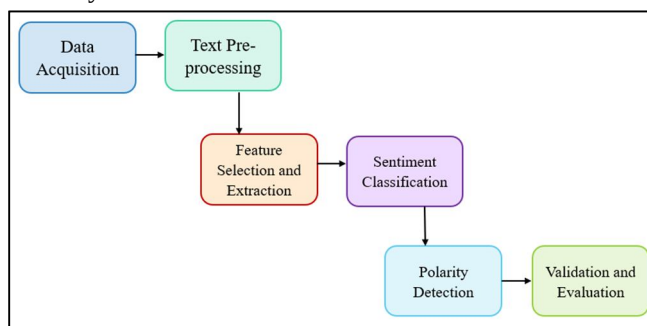


Figure 3: General Steps for Opinion Mining

A. Data Acquisition

Data acquisition means collecting data. The data collection is an important phase because there is need of a proper dataset for analyzing and classifying the text in the dataset. For opinion mining data set can be words, sentences, paragraphs. This data can be gathered from various resources such as surveys, comments, tweets, reviews, feedbacks. All these data source have unstructured data which must be pre-processed in order to use the data for sentiment analysis or opinion mining.

B. Text Pre-Processing

After collecting the data, the data is pre-processed. Data pre-processing reduces noise in data. This is done by removing the unnecessary stop words, repeated words, stemming, removal of emoticons, removal of URLs, etc from our data. Data pre-processing is a very important step in building a machine learning model. Whenever we have data in textual format, we need to apply various pre-processing methods to the text data before using. Text pre-processing refers to the process of converting human textual language into machine understandable text. Not all text pre-processing methods are needed to be applied on the data, it depends on the problem. There are various methods for pre-processing stated below.

- 1) *Lower Casing*: Converting words in the dataset to lower case (example- Buy → buy, NLP → nlp). The words like Buy and buy mean the same but when they are not converted to lower case, they both represent two different words in vector space model. To reduce the vocabulary size of our text data we lower case all the text in dataset.
- 2) *Tokenization*: Tokenization means dividing paragraphs or large text chunks into sentences and also, we can split longer sentences to small words (example- “this book is so good to read” after tokenization output is → [‘this’, ‘book’, ‘is’, ‘so’, ‘good’, ‘to’, ‘read’]). After tokenization the words get stored in list.
- 3) *Removing Punctuation*: Punctuations are removed from our document to avoid different forms of same word (example- ‘amazing!’ and ‘amazing’ these will be treated different if punctuations are not removed).

- 4) *Removing Stopwords*: Stopwords are the commonly used words in the document such as – ‘is’, ‘and’, ‘the’, ‘a’. These stopwords are removed from data because they do not significantly contribute for analyzing of the textual data. Before removing stopwords from our data we need to lower case the data because stopwords are case sensitive.
- 5) *Stemming*: Stemming is the process of transferring a word to its root form of word. Stem or root is the part to which “-ed, -ize, -de, -s, -es, -ing, etc” are added. Example- books → book, looked → look, flies → fly, denied → deny, gaming → game etc. In stemming the root of word is created by removing suffix of a word so the root word may not always result in actual or meaningful word.
- 6) *Lemmatization*: Similar to stemming, Lemmatization is the process of transferring a word to its root form of word but unlike stemming the result after lemmatization we always get the actual or meaningful base word belonging to language. Stemmer is easy to build than the lemmatizer as it requires the deep and detailed knowledge of language. Parts of speech of word is included to attain lemma. Example- books →book, looked → look, flies → fly, caring → care etc.
- 7) *Removing Emojis*: Nowadays in this era of social media we find all reviews and comments on different platforms, some of these comments contain emojis. Emojis are nothing but special coded characters. So, with increase in use of emojis we need to pre-process these emojis also.
- 8) *Removing URLs*: Sometime the text data contains URLs, Hyperlinks which need to be removed from the data which is to be analyzed.
- 9) *Negation Handling*: There is need for negation handling because in some statements we use “not” i.e., “not good”, “not bad”. Here not negates good, not negates bad. So, to handle these types of reviews we do negation handling [5]

C. Feature Extraction

Proper selection and extraction of features plays an important role in determining the accuracy of the model. Thus, the appropriate feature extraction technique must be chosen for extracting the features. There are many feature extraction techniques such as TF-IDF, Bag of Words, Word Embedding etc. The document for analysis is in the textual form of data but for further processing we need to convert this text data to numerical form of data such as vector space model.

1) Bag of Words with TF-IDF

Bag-of-Words with TF-IDF is a simple feature extraction technique in natural language processing. BOW is a representation model of text data and TF-IDF is a calculation method to score an importance of words in a document.

- a) *Bag of Words*: Bag of Words is a representation model of data document, which simply counts how many times a word has appeared in a document. Bag-of-Words only represents number of occurrences without any relationships and contexts or sequence in sentence. Bag-of-Words can be commonly used for clustering, classification, and topic modeling by weighing special words, relevant terminologies. The method to perform Bag-of-Words transformation is as follows.
 - Step 1: Tokenization, which transform sentence to words or tokens.
 - Step 2: Creating dictionary, which removes word duplication and makes set of words which is called dictionary or vocabulary.
 - Step 3: Counting occurrences of each word and make it Bag-of-Words model.
- b) *TF-IDF*: TF-IDF (Term Frequency / Inverse Document Frequency) is one of the most popular IR (Information Retrieval) technique used to analyze importance of a word in a document. TF-IDF will weigh the importance of words in a document. For example, “a” is common word used in any documents so TF-IDF will not consider “a” an important to characterize documents. On contrary, “occurrences” is used in probability relevant topic so TF-IDF will consider “occurrences” an important feature word to recognize topic and category.

2) Word Embedding

Word embedding is one of the document representations in vector space model. Word embedding captures contexts and semantics of word. Detection of similar words is performed more accurately because word embedding preserves contexts and relationships of words. Word embedding has several different implementations such as word2vec, GloVe, FastText and etc

a. *Word2vec*: It describes word embedding with two-layer shallow neural networks so as to recognize context meanings. Word2vec is good at grouping similar words and making extremely accurate guesses concerning meaning of words based on contexts. It has two different algorithms inside: CBoW (Continuous Bag-of-Words) and skip gram model. Skip Gram model is used to predict context from target word. CBoW(Continuous Bag-of-Words) can be used to predict a target word from data.

D. Sentiment Classification

In this step, various sentiment classification techniques are applied to classify the text. There are two types of Approaches for Sentiment analysis, they are Machine Learning and Lexicon based. Some popular sentiment classification techniques are Naive Bayes (NB) and Support Vector Machines (SVM).

Comparison [6] of different Opinion Mining approaches shown as follows in the below “Table 1”.

	Machine Learning	Lexicon Based
Approach	Supervised Learning, Unsupervised Learning.	Dictionary Based, Corpus Based.
Opinion	ML Give better results than Lexicon based.	ML Give good result than Lexicon based.
Advantages	Give higher Precision. High Adaptability and Accuracy. The ability to adapt and create trained models for specific purpose.	Have a strong linguistic resource. Data training is not required. Labelled data is not needed. Wider term coverage.
Disadvantages	The results may not be better if the data are bi-ased during training. Well-designed classifier is needed. La-belled data is needed.	Do not adapt well in different Domain or languages. Low accuracy when data is big. Require Dictionary.

Table 1: Different Opinion Mining Approaches

E. Polarity Detection

After sentiment classification, the polarity of the sentiment is determined. The main aim of polarity detection is to decide whether a text expresses positive, negative or neutral sentiment.

F. Validation and Evaluation

After polarity detection, validation and evaluation is performed for obtained results and overall accuracy of techniques used for sentiment analysis is determined. SA uses various evaluation metrics for Recall, F-score, Precision, and Accuracy. Also, uses average measures such as macro, micro, and weighted F1-scores are useful for multi-class problems. Appropriate metric should be used depending on the balance of classes of the dataset used.

VI. OPINION MINING TYPES

Purpose of opinion mining models is classification of opinion polarities mainly as positive, negative, neutral. Also, different feelings such as anger, happiness, sadness, excitement etc and intentions/objectives such as interested or not interested. Some common types of opinion mining are given below:

- 1) Fine-grained
- 2) Emotion detection
- 3) Aspect-based
- 4) Multilingual
- 5) Intent Based

A. Fine Grained Opinion Mining

“Figure 4” below is the simple flow for fine grained opinion mining. The reviews are taken from web or various social media. The common use of opinion mining is to categorize remarks, comments and statements on a scale of opinion polarity. Basically, it can be positive, negative, or neutral, or beyond that into fine-grained opinions which has a larger scale of categories of opinions that include:

- 1) Very positive
- 2) Positive
- 3) Neutral
- 4) Negative
- 5) Very negative

This is generally used in opinion polls or surveys, giving:

- Very Positive = 5 stars
- Very Negative = 1 star

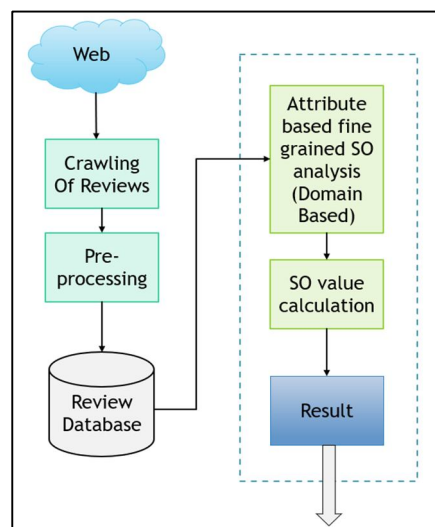


Figure 4: Fine Grained Opinion Mining

Application: This textual analysis tools allows to analyze open-ended survey responses or examine opinions and categorize texts into pre-defined opinions. Fine Grained Opinion Mining is for quantifying qualitative data.

B. Emotion Detection

Emotion detection opinion mining aims at finding and extracting emotions in details such as anger, disappointment, irritation, happiness, etc from text data. The “Figure 5” below shows the basic flow for emotion detection. Some of the emotion detection tools use lexicons, or lists of words defined by the feelings they express. There can be an issue as some words conveying negative emotions, like bad, kill, awe, terrific etc can also be used to express happiness or approval for example “waffle with ice-cream tastes terrifically yum”.

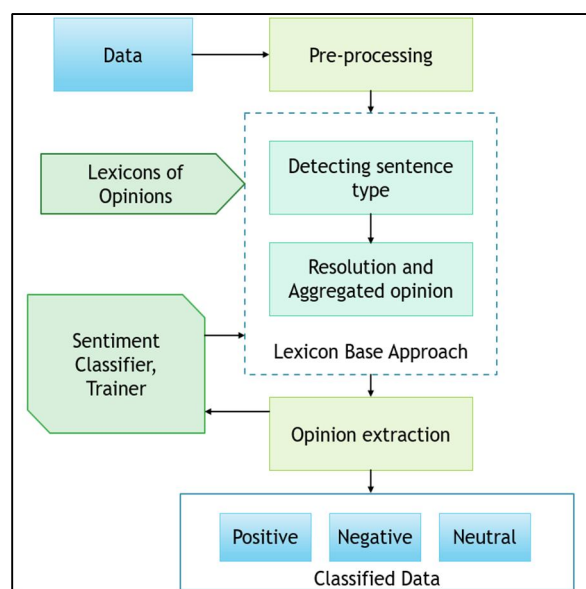


Figure 5: Emotion Detection Opinion Mining

Text analysis programs are allowed to learn directly from sample text, with the help of advanced machine learning algorithms. So as to understand the human language and emotion in detail, also detecting irony and sarcasm.

C. Aspect Based

In aspect-based opinion mining product features are identified and feedbacks or reviews are categorized into different aspects/ features of brand, product. For instance, if all the feedbacks for mobile is to be categorized then the reviews can be classified into aspects like camera, memory, usability, other features, shipping etc. After that each statement is analyzed for positive, negative, or neutral. Above “Figure 6” below gives idea about aspect-based opinion mining. Aspect-based sentiment analysis would read the comment: “The new GUI is amazing,” as a positive opinion about the feature, “Graphical User Interface”.

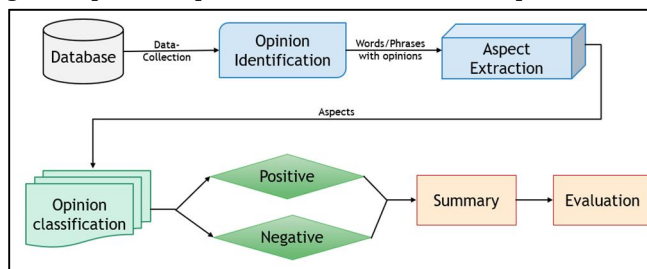


Figure 6: Aspect Based Opinion Mining

D. Multilingual

In the “Figure 7” below multilingual opinion mining is shown. Multilingual sentiment analysis is often very difficult, as it involves a lot of pre-processing and resources. Few resources, such as sentiment lexicons are available also some, like translated corpora, noise detection algorithms, have to be built.

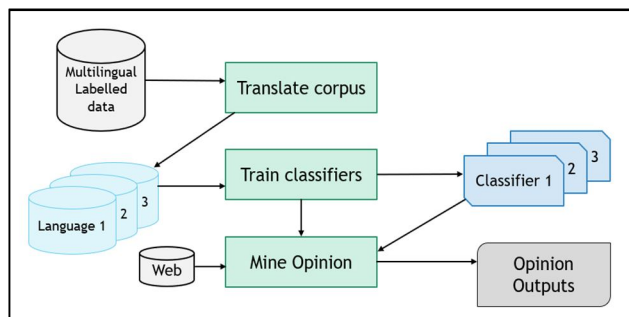


Figure 7: Multilingual Opinion Mining

E. Intent Based

Intent based opinion mining is for deeper understanding of customer Intentions. So, customer intends to use product or not can be tracked and then pattern can be made. Then this can be used for marketing and advertising.

VII. OPINION CLASSIFICATION LEVELS

There are 3 opinion classification levels, they are Document based, Sentence Based, Phrase/Aspect Based.[7]

A. Document Based

Document based is the first level of sentiment Analysis at this level document is considered as whole. So, it is classified based on overall sentiment of whole document. Often, in this case opinion holder is assumed as single source or individual. There is an issue to perform sentiment analysis at this level that is there may not be any sentiments in many sentences of the document thus result may not be accurate. So, to get more accuracy sentence level classification is required.

B. Sentence Based

Various researches have been done for sentence level classification. For classification of sentences firstly the subjective and objective sentences are detected and then subjective sentences are analyzed for positive, negative sentiments. But due to these the objective sentence containing opinions are not analyzed so to overcome this limitation there is need for word level sentiment analysis.

C. Phrase/Aspect Based

It is important to consider words or phrases for doing sentiment analysis because smaller words in sentences show persons opinion. So, Aspect based opinion mining is considered while doing opinion mining at sentence level or document level. Generally, word lexicons include verbs, adverbs, adjectives etc. we take into consideration the polarity of both phrases and words to get more accuracy for sentiment analysis.

VIII. OPINION MINING APPLICATIONS

Some of the popular opinion mining applications are Social Media Analysis, Brand Awareness, Information Extraction, Making Prediction, Decision Making, Customer Feedback, Market Research, Customer Service, Analysis from Surveys.[12]

A. Social Media Analysis

There are huge number of tweets, status, posts with large numbers of active users on facebook, twitter, instagram and many more, they write their opinions and feedbacks at vast range. So, it becomes impossible to monitor and analyze all social media data in real time. Thus, there is need for tool to do the same so as to analyze data in real time constantly so opinion mining can be used to analyze user opinions.

B. Brand Awareness

Opinion Mining can be used for brand awareness since opinion of customers about brand can be obtained and their opinions and feedbacks can be showcased to attract more customers. Also, customers opinions over time can also be monitored so as to maintain brand image.

C. Customer Feedback

Customers feedbacks can be analyzed and there thinking about brands, products can be known. So that prediction can be made about their support to brands or products. Feedbacks can be collected from surveys, customer reviews etc.

IX. CONCLUSION

Opinion mining now a days is need of the hour, where opinion mining has various applications. When we have the proper tools then opinion mining can be performed on various forms of unstructured text data automatically, with less human intervention. With help of opinion mining lots of real time data, comments, tweets, surveys, pages can be processed in few minutes. Also, opinion mining can be performed over time to see how the opinions are changed. The data required for opinion mining can be collected from various sources such as social media, surveys, tweets etc.

Opinion mining is performed by following various stages. Before using any approach for opinion mining lots of pre-processing of text data is required as the data is unstructured, we need to process the data as per the required by the opinion mining approach. There are two approaches for opinion mining Machine learning and Lexicon based. Machine learning approaches give better outcomes than lexicon-based approaches. Machine learning methods train the models based on supervised and unsupervised learning approaches. Whereas Lexicon based methods use the dictionary or linguistic resource. So, here we have discussed the Methods to perform Machine Learning and Lexicon based opinion mining. The more commonly used methods for opinion mining are Naïve bayes and SVM.

There are five types of opinion mining fine grained, emotion detection, aspect based, multilingual, intent based. Opinion classification can be performed at document-based level, sentence-based level, phrase/ aspect-based level. We can perform any level of opinion classification depending upon our application requirement.

X. ACKNOWLEDGMENT

I would like to take the opportunity to express our heartfelt gratitude to the people whose help and co-ordination has made this seminar a success. I thank my guide Dr. J. W. Bakal for knowledge, guidance and co-operation in the process of making this project. Also, I thank Dr. Savita Sangam for motivation and support. I would like to thank our principal for the conducive environment in the institution.

REFERENCES

- [1] Walaa Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, Volume 5, Issue 4, December 2014, Pages 1093-1113
- [2] Wangchuchu Zhao, Keng L. Siau, "Machine Learning Approaches to Sentiment Analytics", Association for Information Systems AIS Electronic Library (AISeL), Proceedings of the Twelfth Midwest Association for Information Systems Conference, Springfield, Illinois May 18-19, 2017, p 1-6
- [3] Xiaowen Ding, Bing Liu, Philip S. Yu, "A holistic lexicon-based approach to opinion mining", Conference: Proceedings of the International Conference on Web Search and Web Data Mining, WSDM 2008, Palo Alto, California, USA, February 11-12, 2008.
- [4] T. Nikil Prakash, A. Aloysius, "A Comparative study of Lexicon based and Machine learning based classifications in Sentiment analysis", International Journal of Data Mining Techniques and Applications Volume: 08, Issue: 01, June 2019, Page No.43-47.
- [5] Umar Farooq, Hasan Mansoor, Antoine Nongailard, Yacine Ouzrout, Muhammad Abdul Qadir, "Negation Handling in Sentiment Analysis at Sentence Level", Journal of Computers, Volume 12, Number 5, September 2017, p 470-478.
- [6] Alessia D'Andrea, Fernando Ferri, Patrizia Grifoni Italy, Tiziana Guzzo, "Approaches, Tools and Applications for Sentiment Analysis Implementation", International Journal of Computer Applications, Volume 125 – No.3, September 201, p 0975 – 8887.
- [7] Bilal Saberi, Saidah Saad, "Sentiment Analysis or Opinion Mining: A Review", International Journal of Advance Science Engineering Information Technology, Vol.7 (2017) No. 5 ISSN: 2088-5334
- [8] S. Fouzia Sayeedunnisa, Dr.Nagaratna P Hegde, Dr. Khaleel UR Rahman Khan, "Sentiment Analysis: Cotemporary Research Affirmation of Recent Literature", International Journal of Pure and Applied Mathematics, Volume 119 No. 15 2018, 1921-1951.
- [9] Mrs. A Hema, M. Abhayadev, "An Overview of Opinion Mining", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 5, May – 2013, p 1822-1825.
- [10] Bakhtawar Seerat, Farouque Azam, "Opinion Mining: Issues and Challenges (A survey)", International Journal of Computer Applications Volume 49– No.9, July 2012, p 0975 – 8887.
- [11] Dipti Sharma, Dr. Munish Sabharwal, Dr. Vinay Goyal, and Dr. Mohit Vij, "Sentiment Analysis Techniques for Social Media Data: A Review", Conference Paper, Sept 2019
- [12] David Osimo and Francesco Mureddu, "Research Challenge on Opinion Mining and Sentiment Analysis", Wired Magazine, 16(7), 16–07
- [13] Maite Taboada, "Sentiment Analysis: An overview from Linguistic", Annual Review of Linguistics 2(1), Feb. 2016.
- [14] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Stede, "Lexicon-Based Methods for Sentiment Analysis", Computational Linguistics Volume 37, Number 2, p 267-30
- [15] Franco Chiavetta, Giosue Lo Bosco and Giovanni Pilato, "A Lexicon-based Approach for Sentiment Classification of Amazon Books Reviews in Italian Language", WEBIST 2016 - 12th International Conference on Web Information Systems and Technologies, p 159-170.
- [16] Kudakwashe Zvarevashe, Oludayo O Olugbara, "A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews", 2018 Conference on Information Communications Technology and Society (ICTAS)
- [17] Eftekhar Hossain, Omar Sharif, Mohammed Moshui Hoque and Iqbal H. Sarker, "SentiLSTM: A Deep Learning Approach for Sentiment Analysis of Restaurant Review" p 1-13.
- [18] Justin Martineau, and Tim Finin, "Delta TFIDF: An Improved Feature Space for Sentiment Analysis", preprint, Third AAAI International Conference on Weblogs and Social Media, May 2009, San Jose CA
- [19] Imane El Alaoui, Youssef Gahi, Rochdi Messoussi, Youness Chaabi, Alexis Todoskof and Abdessamad Kobi, "A novel adaptable approach for sentiment analysis on big social data", El Alaoui et al. J Big Data (2018) 5:12, p 1-18.
- [20] Sonakshi Vij, Amita Jain, Devendra Tayal, "Performing opinion mining and analytical study for cashless transactions", Int. J. Forensic Software Engineering, Vol. 1, No. 1, 2019, p21-31
- [21] Hanif Sudira, Alifiannisa Lawami Diar, Yova Ruldeviyani, "Instagram Sentiment Analysis with Naive Bayes and KNN: Exploring Customer Satisfaction of Digital Payment Services in Indonesia", 978-1-7281-5347-6/19 2019 IEEE, p 21-26.
- [22] Wangchuchu Zhao, Keng L. Siau, "Machine Learning Approaches to Sentiment Analytics", Association for Information Systems AIS Electronic Library (AISeL), Proceedings of the Twelfth Midwest Association for Information Systems Conference, Springfield, Illinois May 18-19, 2017, p 1-6
- [23] Rohith.V, D. Malathi, "Sentiment Analysis on Twitter: A Survey", International Journal of Pure and Applied Mathematics, Volume 118 No. 22 2018, 365-375.
- [24] Yung-Chun Chang, Fang Yi Lee, Chun Hung Chen, "A Public Opinion Keyword Vector for Social Sentiment Analysis Research", 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI) March 29–31, 2018, Xiamen, China.
- [25] Federico Neri Carlo Aliprandi Federico Capeci Montserrat Cuadros Tomas, "Sentiment Analysis on Social Media", 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, p 951-958
- [26] Shanshan Yi, Xiaofang Liu, "Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review", Springer, Complex & Intelligent Systems (2020) 6:621–634.
- [27] Alexander Pak, Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", p 1320-1326.
- [28] R. A. S. C. Jayasanka, M. D. T. Madhushani, E. R. Marcus, I. A. A. U. Aberathne and S. C. Premaratne, "Sentiment Analysis for Social Media"
- [29] Renata Maria Abrantes Baracho, Marcello Bax, "Marcello Bax, "Sentiment analysis in social networks" See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/299435950>
- [30] Xing Fang and Justin Zhan, "Sentiment analysis using product review data", Springer, n Journal of Big Data (2015) 2:5



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)