Journal Homepage: - www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/14435
DOI URL: http://dx.doi.org/10.21474/IJAR01/14435

RESEARCH ARTICLE

TWITTER SENTIMENT ANALYSIS FOR PRODUCT REVIEWS TO GATHER INFORMATION USING MACHINE LEARNING TECHNIQUE

Madhura G.K and Puneet Shetteppanavar
Department of Computer Science and Engineering, East West College of Engineering, Yelahanka New Town, Bangalore, Karnataka, 560064, India.

Manuscript Info

Manuscript History
Received: 22 January 2022
Final Accepted: 24 February 2022
Published: March 2022

Key words:-
Machine Learning, Semantic Orientation, Sentiment Analysis, Twitter

Abstract

The concept of sentiment analysis of twitter data and semantic analysis with the augmentation of machine learning methodologies has become a hot topic in recent years. Many strategies have been presented in the area of sentiment analysis in the last few years to evaluate social media data and produce a graphical presentation towards a certain business. Sentiment analysis shows you how people feel about a product or brand when penning a social media message about it. This is crucial information if you know that one person's opinion of a firm or its products might impact the opinions of others. Like many other online data mining systems, sentiment analysis platforms are based on Support Vector Machine algorithms. This algorithm, in this situation, recognises specific terms as 'positive' or 'negative,' indicating whether or not your brand is being adored or floored. So, in this paper, we first pre-processed the dataset, then extracted the adjectives from the dataset that have some meaning (feature vector), then selected the feature vector list, and finally applied machine learning based classification algorithms such as Nave Bays, Maximum entropy, and SVM, as well as the Semantic Orientation based WordNet to extract synonyms and similarity for the content feature. Finally, we evaluated the classifier's performance in terms of recall, precision, and accuracy.

Introduction:-

The current research article examines the materials on the Internet in a variety of categories that are rapidly expanding in both quantity and volume as sites dedicated to specific types of products specialise in gathering user evaluations from various sites such as Amazon, eBay, and others. Sentiment analysis shows you how people feel about a product or brand when penning a social media message about it. This is crucial information if you know that one person's opinion of a firm or its products might impact the opinions of others. Like many other online data mining systems, sentiment analysis platforms are based on Support Vector Machine algorithms. This algorithm, in this situation, recognises specific terms as 'positive' or 'negative,' indicating whether or not your brand is being adored or floored. When someone uses words like 'terrible,' 'disappointing,' and so on in a tweet, the sentiment analysis classifies it as 'negative.' However, if someone speaks with sarcasm or irony in their tone, most tools will take everything at face value. Sentiment analysis study entails finding a better way to improve the Vector Support Machine Algorithm's classification of sentiments about a product brand. Sentiment analysis research uses a Support Vector Machine Learning Algorithm to classify customer sentiments and positive or negative attitudes toward a

Corresponding Author:- Madhura G.K
Address**:-** Department of Computer Science and Engineering, East West College of Engineering, Yelahanka New Town, Bangalore, Karnataka, 560064, India.

company's specific products via social media sites such as Facebook and Twitter. This study attempts to increase the algorithm's accuracy by using the right kernel and setting the SVM hyper-parameters as key aspects in SVM accuracy, as well as having a large number of training sets to broaden the hyper plane of vectors and create strong support vectors. Sentiments will be acquired using the Twitter API, then submitted to a pre-processor, which will remove any extraneous words. The pre-processed sentiments will then be transformed to SVM format, and the structured data will be used as the data Train. We know that sentiment is a complicated mix of emotions and viewpoints that serves as a foundation for action or judgement. In general, these are the feelings that users have had concerning a reputable company's products. The examination of comments and ideas left on social media sites such as blogs and social networks is known as sentiment analysis. Not only will the words be reviewed, but all of the user comments and attitudes toward a brand will be examined as well. Our study will take into account words as variables, such as context from Twitter or Facebook status. We shall attempt to propose a better machine learning algorithm solution.

Architecture:-

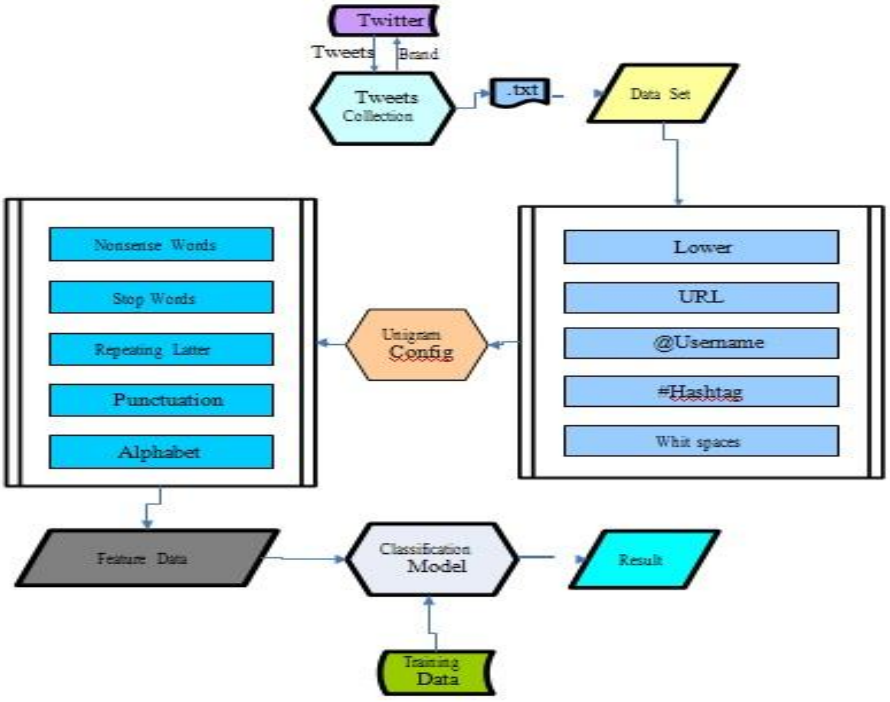

Fig:- Schematic Architecture.

Issues in Sentiment Analysis:-

Despite the fact that the number of classes in sentiment analysis is less than the number of classes in topic-based classification, research reveals that sentiment analysis is more challenging. The classes to which a piece of text is assigned in sentiment analysis are usually negative or positive. They can also be other binary or multi-valued classes, such as 'positive, negative, and neutral' categorization, but they are still fewer than the number of classes in topic-based classification. The fundamental reason that sentiment analysis is more complex than topic-based text classification is that topic-based text classification may be done using keywords, whereas sentiment analysis does not.

The Problem of Sentiment Analysis:

Object identification: -

"moto" (Motorola) and "Nokia" are the objects to be discovered in this blog. This is a significant issue since an opinion is useless unless the object about which it was stated is known. The challenge is comparable to the well-known problem of named entity recognition. There is, nevertheless, a distinction. In a typical opinion mining application, the user is looking for feedback on a set of competing items (e.g., products). As a result, the system

must distinguish between important and irrelevant objects. For example, "BestBuy" is the name of a store, not a competitive product.

Feature extraction and synonym grouping: -
The phone functionalities in the sample above are "voice," "sound," and "camera." Despite attempts to overcome the problem, it remains a significant obstacle. The majority of nouns and noun phrases are discovered in current research. Even if the recall is high, the precision may be low. Verb characteristics are also prevalent, but they're more difficult to spot. We also need to aggregate synonym features to get a summary comparable to the one since individuals often use different words or phrases to express the same feature (for example, "voice" and "sound" refer to the same feature in the above example). This is also a difficult problem. There is still a lot of research to be done.

Opinion orientation classification: -
This job assesses whether or not there is a positive or negative view on a feature in a text. Existing methods use both supervised and unsupervised techniques. One of the most important challenges is identifying opinion words and phrases (for example, good, bad, poor, and fantastic), which are crucial for sentiment analysis. The issue is that there appears to be a positive visual comparison of feature-based opinion summaries of two mobile phones. Image, Battery, Camera, Dimensions, and Weight Cellular Phone 1; Cellular Phone 2; Cellular Phone 3; Cellular Phone 4; Cellular Phone People utilise an infinite amount of terms to communicate their thoughts, and they might differ dramatically between fields. Even within the same domain, the same word might convey distinct meanings depending on the context [. In the line "The battery life is long," for example, "long" denotes a favourable assessment of the "battery life" feature. The word "long" expresses a negative judgement in the statement "This camera takes a long time to focus." There are still numerous issues to be resolved.

Integration: -
Because we need to match the five pieces of information in the quintuple, integrating the about tasks is also difficult.

Object and feature: -
The person or organisation who expresses an opinion is known as the holder of that opinion. In the case of product reviews and blogs, the authors of the entries are usually the ones who hold the opinions. In news pieces, opinion holders are more essential because they frequently specify who or what organisation has a specific viewpoint.

Opinion and orientation: -
An opinion on a feature f (or object o) is a person's positive or negative assessment of that feature (or item). Opinion orientations refer to the positive and bad aspects of a situation. With these notions in mind, we can define the feature-based sentiment analysis model, which includes an object model, an opinionated text model, and the Mining aim.

Comparative opinion:
A comparative view is a preferred relationship between two or more objects based on some of their common characteristics. "Coke tastes better than Pepsi," for example, is frequently stated using the comparative or superlative form of an adjective or adverb. Due to a lack of space

Related Work:-
A lot of work has been done in the topic of "Sentiment analysis" by a number of researchers in recent years. Work in the field has been going on since the turn of the century. It was originally designed for binary classification, which categorises thoughts or reviews into two categories: positive or negative. Paper [3] uses an unsupervised learning system to predict review based on the average semantic orientation of a phrase containing adjective and adverb, determining whether the phrase is positive or negative and classifying it as thumbs up or thumbs down. Some sentiment assessments, such as [4], are based on a review of the product's user summarising system. The product feature in [4] uses a filtering approach based on latent semantic analysis (LSA) to identify opinions. A comparison of positive and negative statements is used in Paper [5.] It pulls data from the web and labels the word set manually, which takes a lot of time and work. The author of [6] employed a rule-based system for sentiment analysis of Chinese documents based on Baseline and SVM, which extracts the overall document polarity of specific words using a sentiment word dictionary and adjusts it based on context information. The polarity of a word is computed in another work [7] by adding up the polarities of all the words in the phrase, which might be positive or negative

depending on the sentence structure. Lakshmi and Edward [8] recommended preprocessing the data to improve the original sentence's quality structure. For sentiment analysis, they used the LSA approach and cosine similarity. For sentiment categorization, Basant Agarwal et al. [9] used the phrase pattern technique. It extracts contextual and grammatical information from the document using part of speech-based rules and dependency relationships. The author of [10] aimed to present aspect-based opinion polls based on unlabeled free form textual customer reviews that do not compel customers to respond to the questions. M. Karamibekr and A.A. Ghorbani [11] suggested a method for sentiment classification of a document in the social domain based on verbs as a key opinion term. SentiFul is a sentiment lexicon created by Paper [12], which uses and expands it through synonyms, antonyms, hyponyms, derivation, and compounding. They suggested a strategy for identifying four types of affixes based on their role in sentiment features: propagation, weakening, reversal, and intensification. These methods attribute polarity to sentiments, which aids in the expansion of the lexicon and hence improves sentiment analysis. For sentiment analysis, a lot of work has been done where academics have researched and implemented soft-computing methodologies, primarily fuzzy logic and neural works. [13] and [14] are two instances of works that use fuzzy logic to solve problems. [13] makes a significant addition by using a fuzzy domain sentiment ontology tree extraction technique. This algorithm creates a fuzzy domain sentiment ontology tree based on reviews, which involves extracting sentiment words, product attributes, and feature relationships, and accurately predicting the polarity of reviews. The authors of [14] developed a membership function-based fuzzy inference system. They defined and standardised the procedure of quantifying the strength of reviewer's judgments in the presence of adverbial modifiers by constructing membership functions. They used the approach to study adverbial modifier trigram patterns.

Research Issues:-

Sentiment and subjectivity classification, which regarded the topic as a text classification problem, was the first step in the field's research. Sentiment categorization determines whether a document (for example, product reviews) or a sentence reflects a favourable or negative impression. A sentence's subjective or objective nature is determined by its subjectivity categorization. Many real-world applications, on the other hand, necessitate a more in-depth examination because the user frequently wants to know what perspectives have been voiced. For example, one would wish to discover what characteristics of a product have been appreciated and condemned by users from a product review. To introduce the overall problem, let's use the following iPhone review segment as an example (each sentence has a number for easier reference):

"(1) I recently purchased an iPhone. (2) It was a lovely phone. (3) I really liked the touch screen. (4) The audio quality was also excellent. (5) My mother, on the other hand, was furious with me since I didn't notify her before I bought it. (6) She also said the phone was overpriced and requested that I return it to the store."

What we want to get out of this review is the answer to that question. The first thing we notice is that this review contains a variety of viewpoints. Three favourable opinions are expressed in sentences (2), (3), and (4), whereas negative opinions are expressed in sentences (5) and (6). Then we see that each opinion has a certain objective on which it is expressed. The opinion expressed in line (2) is about the iPhone as a whole, whereas the opinions expressed in sentences (3) and (4) are about the iPhone's "touch screen" and "voice quality," respectively. The sentiment expressed in line (6) is about the price of an iPhone, but the sentiment expressed in phrase (5) is about "me," not iPhone. This is a crucial point. In an app, the user may be interested in opinions on specific targets but not all (for example, "myself" is improbable). Finally, we may notice the sources of information or the people who have the opinions. In sentences (2), (3), and (4), the source or bearer of the opinions is the review author ("I"), however in words (5) and (6), it is "my mum." We can define sentiment analysis or opinion mining with this case in mind.

Approaches to Sentiment Analysis:-

As part of their research, several companies use opinion mining and sentiment analysis. Companies, for example, employ opinion mining to construct and manage review and opinion aggregate websites automatically. Their systems are constantly collecting data from the Internet, such as product reviews, brand perceptions, and political issues. Opinion mining and sentiment analysis may be used as a subcomponent technology in other systems to improve customer relationship management and recommendation systems by analysing positive and negative customer comments. Similarly, opinion mining and sentiment analysis in social communication may discover and remove "flames" (overly heated or aggressive words). Sentiment analysis is used by businesses to establish marketing strategy by measuring and predicting public perceptions of their brand. Automatic programmes that trawl

online reviews and compress the information acquired are the focus of research and development. Several companies currently offer tools that track public attitudes on a wide scale by providing graphical summaries of blogosphere trends and opinions. The development of opinion-tracking tools is crucial for business. There are also various technologies available to assist organisations in extracting and analysing information from blogs concerning large-scale customer patterns.

Common Sentiment Analysis Tasks: -

Polarity classification is the most basic task in opinion mining. When a piece of text expressing an opinion on a particular problem is classed as one of two opposing attitudes, it is called polarity classification. Polarity classification includes terms like "thumbs up" vs "thumbs down," or "like" versus "dislike." Polarity classifications also help to distinguish pro and con expressions in online reviews, which improves the credibility of product reviews. Another type of binary emotion categorization is agreement detection. When pair of text documents are compared, agreement detection determines whether they should have the same or different sentiment-related labels. Following the identification of the polarity classification, the system may assign degrees of Positivity to the polarity—that is, it may place the view on a scale of positive to negative. It can also categorise multimedia materials based on their mood and emotional content, which can be useful for things like affective human-machine interaction, troll filtering, and cyber-issue detection. New obstacles, such as subjectivity detection and opinion-target identification, arise when the text does not contain strong viewpoints or covers more than one issue or object. It's easier to classify the sentiment by distinguishing between subjective and objective language. Furthermore, a text can have polarity without including a viewpoint; for example, a news report can be characterised as good or bad news without being subjective. Typically, a system uses the results of a topic-based search engine to perform sentiment analysis on on-topic documents. Several studies, however, suggest that combining these two activities may improve overall performance. For example, off-topic portions in a publication may contain irrelevant affective information, resulting in erroneous global sentiment polarity concerning the core issue. A document may also contain information on a variety of topics that the user is interested in. It is critical to distinguish subjects and differentiate the viewpoints related with each issue in such situations.

Evolution of Opinion Mining: -

Currently, vector extraction is used to represent the most essential text attributes in opinion mining and sentiment analysis. This vector can be used to classify the most important features. Term frequency and existence are two often utilised features. Presence is a binary-valued feature vector whose elements solely indicate if a phrase occurs (value 1) or does not (value 0). (value 0). The use of presence as a basis for reviewing polarity categorization indicates an important distinction: whereas recurrent keywords indicate a topic, repeated terms may not reflect the overall emotion. Other term-based features can be added to the features vector. The position of a token in a text unit can have an impact on the emotion of the text. Furthermore, presence n-grams—typically bigrams and trigrams—might be considered useful qualities. The distance between terms is also used in some approaches. As a simple kind of word-sense disambiguation, general textual analysis uses part of speech (POS) information (for example, nouns, adjectives, adverbs, and verbs). Certain words are good markers of sentiment and help classify the sentiment by guiding feature selection. Additionally, pre-defined POS Patterns pick phrases, which usually include an adjective or adverb, to aid in the detection of feelings. Other text mapping strategies have been created by some academics, which assign labels to predetermined categories or real numbers that describe the degree of polarity. Domain and topic are the only constraints on these approaches. Furthermore, most sentiment analysis research focuses on text written in English, and as a result, the majority of the resources developed (such as sentiment lexicons and corpora) are in English. Domain adaptation is an issue when it comes to applying this research to other languages.

Conclusion:-

We introduced a set of machine learning approaches with semantic analysis for identifying sentences and product reviews based on Twitter data in this research. The main goal is to analyse a huge number of reviews using a pre-labeled Twitter dataset. The nave byes approach provides us a better result than maximal entropy, and SVM is subjected to a unigram model, which gives us a better result than just utilising it. When the semantic analysis WordNet is combined with the above approach, the accuracy increases to 89.9% from 88.2%. The training data set can be expanded to improve the feature vector-related sentence identification process, as well as WordNet for review summarization. It may provide a better visual representation of the content, which will be beneficial to the users.

References:-

[1] J. Xie and B. K. Szymanski. Community detection using a neighborhood strength driven label propagation algorithm. In IEEE Network Science Workshop 2011, pages 188-195, 2011.

[2] J. Xie and B. K. Szymanski. Towards linear time overlapping community detection in social networks. In PAKDD, pages 25-36, 2012.

[3] J. Xie, B. K. Szymanski and X. Liu. SLPA: Uncovering Overlapping Communities in Social Networks via A Speaker-listener Interaction Dynamic Process. In Proc. of ICDM 2011 Workshop, 2011.

[4] S. Gregory. Finding overlapping communities in networks by label propagation. New J. Phys., 12:103018, 2010.

[5] K. Wakita and T. Tsurumi. Finding community structure in mega-scale social networks. In WWW Conference, pp. 1275-1276, 2007.

[6] M. E. J. Newman and M. Girvan. Finding and Evaluating Community Structure in Networks. Phys. Rev. E, 69, pp. 026113, 2004.

[7] V. Blondel, J. Guillaume, R. Lambiotte and E. Lefebvre. Fast Unfolding of Communities in Large Networks. J. Stat. Mech., 2008.

[8] S. White and P. Smyth. A spectral clustering approach to finding communities in graphs. Proc. of SIAM International Conference on Data Mining, pp. 76-84, 2005.

[9] I. Leung, P. Hui, P. Lio, and J. Crowcroft. Towards real-time community detection in large networks. Phys. Rev. E, 79:066107, 2009.

[10] Y.Singh, P.Kumar Bhatia, and Omprakash Sangwan. "A review of studies on machine learning techniques." International Journal of Computer Science and Security 1, no. 1, 70-84, 2007

[11] P.D.Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." In Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417-424. Association for Computational Linguistics, 2002.

[12] L.Ramachandranand Edward F. Gehringer. "Automated assessment of review quality using latent semantic analysis." In Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference on, pp. 136-138, 2011

[13] J. Fernando S_anchez-Rada1, Marcos Torres1, Carlos A. Iglesias1, Roberto Maestre2, and Esther Peinado, "A Linked Data Approach to Sentiment and Emotion Analysis of Twitter in the Financial Domain", Retrieved 2014

[14] Jalaj S. Modha* Prof & Head Gayatri S. Pandi Sandip J. Modha, " Jalaj S. Modha* Prof & Head Gayatri S. Pandi Sandip J. Modha", International Journal of Advanced Research in Computer Science and Software Engineering, 2013

[15] H.Saif, Y. He, and H. Alani. "Alleviating data sparsity for twitter sentiment analysis." CEUR Workshop Proceedings (CEUR-WS. org), 2012.

[16] H. Saif, Y.He, and H. Alani. "Semantic sentiment analysis of twitter." In The Semantic Web–ISWC 2012, pp. 508-524. Springer Berlin Heidelberg, 2012.

[17] A.Kumar and T. M. Sebastian, "Sentiment Analysis on Twitter", International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, 2012

[18] P. Bo, 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-Volume 10, pp. 79-86. Association for Computational Linguistics, 2002.

[19] A. Go, R. Bhayani, and L. Huang. "Twitter sentiment classification using distant supervision." CS224N Project Report, Stanford, pp.1-12, 2009

[20] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu. "Combining lexicon-based and learning-based methods for Twitter sentiment analysis." HP Laboratories, Technical Report HPL-2011 89, 2011

[21] A. Mudinas, D. Zhang, and M. Levene. "Combining lexicon and learning based approaches for concept-level sentiment analysis." In Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, p. 5. ACM, 2012.