Machine Learning-based Sentiment Analysis Web Application using Django

Zahoora Abid¹*, Tahera Abid² , K.G.S.Venkatesan³

¹Department of Computer Science & Engineering, Nawab Shah Alam Khan College of Engineering and Technology, Hyderabad, Telangana, India.

²Department of Information Technology, Nawab Shah Alam Khan College of Engineering and Technology, Hyderabad, Telangana, India.

³Department of Computer Science & Engineering, MEGHA Institute of Engineering & Technology for Women, Hyderabad, Telangana, India.

Received: 15.08.2023  •  Accepted: 21.08.2023  •  Published Online: 31.08.2023

Abstract: The vast majority of customers will investigate several aspects of a product, such as its price, warranty, quality, and others, before making a purchasing decision. Customers who are satisfied with the aforementioned aspects have a greater propensity to attempt to acquire a product based on the level of service they have experienced from a company. The use of sentiment analysis (SA) is essential for making an informed purchasing choice since it requires a significant time investment and increases the likelihood of being misled by the dealer. Analysis of people’s feelings about a product or service via the use of text evaluations and comments presents a number of challenges that need to be overcome before the relevant information can be made available to the general public. These challenges need to be overcome before the information in question can be provided to the public. In addition, SA is an essential study avenue in the discipline of NLP (which is an abbreviation for "natural language processing"). In this essay, we create an original model for sentiment analysis by making use of the ML Algorithm. This method is capable of accurately conveying feelings in writings that include the opinions of several writers. In the first stages of the data processing process, the stop-word approach is used.

*Correspondence: Assistant Professor, Department of Computer Science & Engineering, Nawab Shah Alam Khan College of Engineering and Technology, Hyderabad, Telangana, India. Email:zahooraabid@nsakcet.ac.in

https://doi.org/10.58599/IJSMEM.2023.1804
Volume-1, Issue-8, PP:26-36 (2023)

This work is licensed under a Creative Commons Attribution 4.0 International License CC BY-NC-ND 4.0.
In order to extract the necessary attributes from the text input, a piece of software known as count vectorizer is used. The information will be transformed into vector form as a consequence of this action. In the end, we decide whether the sentiment is good, negative, or neutral by using a machine learning (ML) classifier known as the Naive Bayes classifier. Those in need of sentiment analysis may get exact sentiment categorization via the use of this model, which was built on top of the Django web platform.

**Key words:** Natural Language Processing, Text Analysis Tool, Sentiment Visualization, Opinion Analysis.

1. **INTRODUCTION**

Massive amounts of textual material already saved on the Internet are largely responsible for its rising significance as a medium for the distribution of information about ideas and goods. Better choices might be made with the aid of analysis of this text data to extract ideas or viewpoints utilising tools like Natural Language Processing (NLP) and big data [1, 2]. These tools may help people make better choices. The purpose of sentiment analysis is to determine if a review or remark is favourable, negative, or neutral in tone. In order to create models that can explain why specific bits of content belong in certain categories, natural language processing is employed to uncover hidden terms in texts. Example: "The cost of swallow nest is too much but it is so delicious," where the negative valence of the emotion is grounded on the provision quality and the positive valence of the sentiment is based on the food quality. When a SA employs this kind of wording, it indicates that they are unable to predict the result and have instead opted to make a decision based on their own criteria. The SA’s primary purpose is to detect bipolar perspectives in the judgement or comment of a statement [3].

The traditional methods of SA misrepresent the text’s polarity. Many businesses may improve by learning what customers think of their offerings. Aspect level sentiment analysis, which looks at multiple feelings in relation to several factors, is hence gaining in popularity. The SA may be carried out on many different levels, from the word to the document to the phrase to the aspect. Nowadays, SA is often used in the background in many industries and enterprises. Corporate transformation, public opinion management, box office and election prediction, and many more are only some of SA’s numerous applications [4, 5].

The primary objective of SA is to examine data from many sources, evaluate it, and then draw conclusions. Several aspects influencing the SA process must be controlled adequately to get the final categorization result. The next section [6] will address some of these concerns in further detail. Repairing internal references is the main issue. This issue arises if the reviewer fails to distinguish between two related terms. One cannot tell from the statement "After watching the movie, we went out for dinner; it was amazing" whether or if the talk occurred before or after the meal. This kind of issue is especially prevalent. The second consideration is the era to which it is most closely related. The best results may be achieved by reviewing SA or asking for comments at the optimal time. Positive and negative reactions to a product at the same time. Opinion analyzer will inevitably have difficulties as a result of this. Comparative SA is where you’ll find the vast majority of issues of this kind [7]. The sarcastic tone is the third issue. The handling of the situation is the most important concern at
the moment, thus great care must be taken. Sarcasm is the hilarious ironic expression of important ideas via the use of sarcastic or mocking language. Because of how well he does academically, he consistently receives only A’s and B’s on his monthly tests. In this context, the adjective ”excellent” is used in a derogatory way. Writing like this may affect SA [8]. Fourthly, ”Negations,” which refers to the use of negative phrases in writing that may change the meaning of a sentence entirely. Even when two sentences, such ”This is a good article.” and ”This is not a good article,” seem to represent polar opposite meanings, a word-by-word examination may offer different answers. N-gram examination is the go-to technique for production with problems of this kind. Spam Detection is the fourth issue, and it entails determining whether a review or remark was posted by a human person or a bot. Many individuals who have never tried the company’s wares provide their thoughts, positive or unfavourable. As a result, it may be difficult to tell which reviews are authentic and which are fake, which has serious consequences for SA [9].

2. RELATED WORKS

This study introduces an original SA application with the potential to address these problems. This software can do things like preprocess data, identify features, and classify emotions. Stop words are used in the initial stage of data pre-processing in sentiment analysis to remove noise phrases. Count After the extraneous words have been removed, Vectorizer transforms the remaining text so that it assumes the shape of a vector. Extracting the characteristics from the text is the first step. When all else fails, a machine learning classifier.

proposed using Filtered-LDA (Latent Dirichlet Allocation) to comprehend shifts in Twitter sentiment. In order to capture potential causes for shifts in attitude, several series-connected LDA models with many hyper parameters have been explored. Using Topic Model, we can filter out old tweets on irrelevant topics. Tweets in Arabic or any other language not supported by this framework will be ignored. Reviews from customers were subjected to a multi-strategy SA by reason for mental confusion. Probability, rather than a hard-and-fast number, is used in this method to determine the intensity and polarity of Chinese emotion phrases. We used a mixture of the SVM and NB multi-strategy SA algorithms. Complex expressions of emotion and syntactic structural modifiers are not adequately evaluated by these approaches. There is also no evaluation of hazy articles. A multichannel SA model was proposed which uses CNN, BiGRU, and the Variational Information Bottleneck. The multigrain sentiment characteristics are extracted using BiGRU for context extraction and CNN for local features. This approach deals with out-of-the-ordinary problems in SA that have suitable emotional underpinnings. This approach cannot recall nuanced feelings [10] . Sentence-level competition between an Aspect-Level SA Approach using CE-HEAT and SemEval was first developed.

Emotion and aspect traits are extracted by the CE-HEAT’s two hierarchical attention units. Aspect-level sentiment classification is a strength of the CE-HEAT approach. It quickly picks up on connections between emotions and aspects. Aspect-level sentiment classification accuracy might be enhanced by taking use of sentence structure, semantic properties, and logical relations; however, this was not
investigated in this work. To enhance Tweet SA with Fuzzy Sentiment, it presented the Feature Ensemble Model. In addition, we use CNN models. This model improves the efficiency of tweet SA. This method was limited to analyzing tweets containing vague emotions, and not sarcasm or language. The innovative Refined Global Word Embeddings (RGWE) by extends the word representation by combining two refined word embedding algorithms based on a sentiment idea. But it doesn’t work with SA verbs, adjectives, or adverbs, so it’s not a perfect solution. Despite the fact that scholars have developed a wide variety of SA approaches, none of them have shown to be very successful. In order to solve this problem, our research suggests using a Django-based SA application approach [11, 12].

3. PROPOSED WORK

As ML methods become more popular in SA, we show how they may be used to create a unique SA application at the aspect level. To do this, we use Naive Bayes, a machine learning classification approach, in conjunction with Stop Words and the Count Vectorizer. Textual information may be entered into documents and files. In order to properly classify how people feel about anything, multiple stages must follow the submission of this file or document. Processes like preparing data, feature extraction, and sentiment classification are all examples. User reviews and comments often take the form of text, which might be difficult to read and understand for certain persons. Therefore, the right method must be used to the data sets in order to filter out the irrelevant information. The process of getting data ready for analysis is called "pre-processing," a format that a computer can read and interpret. Data filtering is a common preprocessing procedure. Stop words are essential for this kind of pre-processing work. The next step is to turn the feedback into a numerical format that can be used in ML. In order to get accurate findings, the process of converting textual criticisms into quantitative data is complex and must be carried out with great care. For this reason, we use a count vectorizer to carry out the vector translation. amount to After the data has been cleaned of stop words, the properties of interest (unique words) may be extracted using a vectorizer. To account for how often each phrase appears, the input text is also transformed into a vector format. The text is fed into Naive Bayes, a machine learning classification algorithm, and the results are categorized according to whether they have a positive, negative, or neutral bias. The suggested procedure for organizing feelings is shown in Figure 1.

A. The Use of 'Stop Words' in Pre-Processing Data

Data During the pre-processing stage, extraneous information is removed from the texts. In order to get rid of unnecessary data, the SA model that has been developed makes use of stop words. "Stop words" are commonly used words that search engines have been programmed to disregard in favour of more precise phrases. Some examples of stop words are "the," "in," "a," and "an." Because this is what the vast majority of people are seeking for, the phrase "how to enhance the malware detection approaches" delves rather deeply into this particular sphere of concern and concern. The search engine will employ terms and phrases such as "how," "to," "enhance," "the," "malware," "detection," and "approaches" to locate results that are relevant to the query. However, the search engine will make
every attempt to provide results that include the phrases that you have chosen, "how," "to," and "the" in comparison to websites that summarise information on a variety of methods for more efficient virus identification. As a result of the fact that "how," "to," and "the" are three of the most often used phrases in the English language. The study of polarity may benefit from the use of a number of stop words, which are found in the English language. Examples of "stop words" in the English language include the words "a," "the," "is," "are," "for," "an," "nor," "but," "or," "yet," and "so," as well as a number of other terms. In addition to those described above, a condensed version of the list of stop words may be found in Figure 2. The search engine will be able to concentrate on giving results that include the necessary keywords if phrases like these that are irrelevant are excluded.

**B. Feature Extraction Using Count Vectorizer**

After the irrelevant words have been removed from the input text, the remaining text is converted into a token matrix. To do this, we utilise the Count Vectorizer programme to convert the data to a vector format. Word frequency counts are used by the Count Vectorizer to produce vectors from the input text. The collection of text documents are sent into the count vectorizer, which generates a token count matrix. This approach can be used to produce a sparse count matrix. In the next example, we will build the Count Vectorizer matrix from scratch. The following expressions are provided in the style of a written document. One automobile can be described as beautiful by one person, filthy by another, and speedy by a third, all without contradiction. Due to the existence of three distinct documents and six distinct qualities (including "This", "car", "is", "beautiful", "dirty", and "speedy"), the aforementioned phrases may be structured as a 3 by 6 Count Vectorizer matrix.

In the aforementioned table, the existence of a feature is represented by the number "1," while the lack of the feature is represented by the value "0." Sentences 2 and 3 have an analogous feature,
Table 1. A Matrix Used as a Count Vectorizer

<table>
<thead>
<tr>
<th>Features Sentences</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

however it is labelled "0," whereas features 5 and 6 are all labelled "1" in those sentences in table 1. Four features/words in the first sentence are marked with a "1," whereas their counterparts in the second and third sentences are both marked with a "0." So, to make vectors out of the words in the text input, Count Vectorizer is the tool utilised. No matter the length of the document, the count vectorizer will use the same strategy to transform the data into a matrix. Checking for accuracy ensures that the SA application may be used as intended.

4. RESULTS AND DISCUSSION

The proposed SA application’s accuracy may be represented as a percentage by dividing the number of correct predictions by the total number of correct predictions. Since the suggested SA app has a greater percentage of accuracy, this suggests that it effectively evaluates sentiment.

Figure 2. Home Page

The 'about' page is seen in Figure 4. This article provides details on SA, such as a definition of the term and an examination of its salient features.

The "Contact" page is shown in Figure 4. This page lists the contact information for the organisation or administrator, including postal address, email address, and phone number. Users
may contact the program’s developers through in-app messaging to pose questions or request further information. Users may provide feedback on the app by filling out the fields labelled "Name," "Mail ID," "Subject," and "Message" and clicking the "contact" button.

When a user first accesses this programme, they will see a screen labelled "Registration." On this page, they will be prompted to register their account. After registering, users are required to log in before they may use the site again. Filling out the required information for registration, which includes the
User name, Password, and Password Confirmation fields, and then clicking the Sign up button will result in the successful completion of the registration process.

![Registration Page](image)

**Figure 5.** Registration Page

The 'Login' screen that will display when this is completed is shown in Figure 6. After a user has registered for this SA application, they are able to access it whenever they want by navigating to the login page and entering the person name and password that they generated during the registration process. In the event that a user enters a wrong password or one that does not correlate to their user name, they will be locked out of the system. You have the option of selecting the "Forgotten Password" choice in order to get around the difficulties I just described. If the user chooses this option, he will be allowed to change the password and continue using the account with the new password after making the change.

The 'Upload' interface, seen in Figure 7, is used to upload files containing training data. The file that was supplied has several different types of reviews that may be used as examples. The user must click a box indicating their acceptance of the terms and conditions before uploading a training material.

![Upload Interface](image)

Last but not least is Figure 8, which depicts the page viewed to preview or assess the review. Users can leave their own thoughts and opinions about a service or product in a comment box labelled "Enter text here..." on this page. In this study, participants were asked to rate the experience positively and enter the word "PREDICT" before pressing the button. Positive prognosis is the logical inference from this remark.
5. CONCLUSION
In this research, we implement the Django web framework to create a sentiment analysis app. This model uses a supervised classification method to better classify the reviews. In a variety of settings, our new approach to sentiment analysis is designed to provide reliable information on the emotional weight of individual words. Processing of data in its preliminary stages, involving the usage of stop words. A count vectorizer is used to accomplish the goal of extracting features from the text data. In this inquiry, the Naive Bayes classifier is used to determine how positive, negative, or neutral each phrase is. The following are areas where more research is needed that are related to this work: While tagged data may be available in some circumstances, unlabeled data is more common. In such a case, semi-supervised learning, which entails first transforming unlabeled input into labelled data and then assessing the results of the transformation, could be useful. The use of emoticons in online reviews and comments has grown rapidly, necessitating the development of sophisticated tools for interpreting
the images.

References


