International Journal of Engineering and Management Research
e-ISSN: 2250-0758 | p-ISSN: 2394-6962
https://ijemr.vandanapublications.com

Peer Reviewed & Refereed Journal
Volume-13, Issue-4 (August 2023)
https://doi.org/10.31033/ijemr.13.4.5

Sentiment Analysis using Opinion Mining onCustomer Review

Prachiti Akre[1], Rushikesh Malu[2], Aparna Jha[3], Yash Tekade[4] and Wani Bisen[5]
[1]Student, Department of CSE, Shri Ramdeobaba College of Engineering and Management,Nagpur, INDIA
[2]Student, Department of CSE, Shri Ramdeobaba College of Engineering and Management,Nagpur, INDIA
[3]Student, Department of CSE, Shri Ramdeobaba College of Engineering and Management,Nagpur, INDIA
[4]Student, Department of CSE, Shri Ramdeobaba College of Engineering and Management,Nagpur, INDIA
[5]Assistant Professor, Department of CSE, Shri Ramdeobaba College of Engineering and Management,Nagpur, INDIA

[5]Corresponding Author: bisenwh@rknec.edu

Received: 03-07-2023 Revised: 18-07-2023 Accepted: 02-08-2023

ABSTRACT

Opinion mining is the process of discovering user opinions about a subject, a product, or an issue. Sentiment analysis is the process of separating opinions' feelings from those opinions. This work presents a comprehensive review of techniques in sentiment analysis using opinion mining. We start by discussing the basic concepts of opinion mining and then delve into various techniques and approaches used for sentiment analysis. This work focuses on analyzing the user reviews on products from e-commerce websites such as amazon.com and compares products of similar specifications based on the polarity of user reviews for the products calculated using nlp and naive bayes classification.

Keywords-- Opinion Mining, Opinion Summarization, Natural Language Processing

I. INTRODUCTION

The Internet has become the primary source of global knowledge thanks to its rising popularity. Many individuals communicate their views and opinions through numerous online resources. Corporations, governments, and people can all benefit from public opinion when gathering data and making judgment calls. In recent years, e-commerce platforms have become more popular than traditional marketplaces.

Customers frequently use user ratings and feedback to inform their decisions when making purchases on ecommerce platforms. Customer purchase decisions may be influenced by the user reviews since they provide intelligent input on products and services. However, as more reviews are added, it gets harder to manually go through them, comprehend the tone they're written in, and decide whether or not to buy the product.

Sentiment analysis is useful in this situation. It can be used to automatically categorize user reviews' sentiments into three categories: positive, negative, and neutral. This helps customers better grasp how people feel about a given product as a whole. Customers may immediately assess the benefits and drawbacks of a certain product using sentiment analysis. They can also compare the product ratings with those of other items with comparable specs. Using this data, consumers can choose items that have received generally more favorable reviews than those that have, and make informed purchasing decisions. Sentiment analysis can assist users in spotting fraudulent or deceptive reviews that may influence their purchasing choices.

II. RELATED WORK

The current research work on the topic of opinion mining and sentiment analysis is provided in this section.

In [1], one of the core issues with sentiment analysis was solved, i.e., the difficulty in categorizing sentiment according to its polarity. With thorough process descriptions, a general method for categorizing sentiment polarity is suggested.

Online product reviews gathered from Amazon.com were the source of the data for this study.

In [2], the approaches for carrying out this task and the uses of sentiment analysis are discussed. In order to fully comprehendthe benefits and drawbacks of each approach, it is then evaluated, compared, and investigated. Finally, in order to determine future directions, the difficulties of sentiment analysis are examined.

In [3], it is determined which words and phrases people use on social media to express their opinions about various goods, services, organizations, and events. For the purpose of optimizing sentiment analysis, four machine learning classifiers i.e J48, OneR, Naive Bayes and BFTree are used. Three manually compiled datasets are used in the experiments out of which two are retrieved from amazon, and one is compiled from IMDB movie reviews. The effectiveness of the mentioned four methods are tested and contrasted.

In [4], We focus on the class of simple "probabilistic classifiers" known as naive Bayes classifiers, which employ the Bayes theorem while making strong (naive) assumptions about the independence of the correlations between the features. A simple strategy for creating classifiers is the Naive Bayes approach. It involves models that select class labels from a finite set and apply them to problem instances, which are represented as vectors of feature

Peer Reviewed & Refereed Journal
Volume-13, Issue-4 (August 2023)
https://doi.org/10.31033/ijemr.13.4.5

values.

III. METHODOLOGY

A. Data Collection

Identify sources of product reviews, such as e-commerce websites or social media. Collect a large and diverse dataset of product reviews for various products, ensuring that it covers a wide range of opinions and sentiments.

B. Data Preprocessing

Tokenize the text data by splitting it into individual words or phrases. Remove stopwords (common words that do not add meaning to the text) and punctuation. Normalize the text data by converting all words to lowercase and removing any non-alphabetic characters. Stem or lemmatize the words to reduce them to their base form.

C. Feature Extraction

Choose appropriate features to represent the text data, such as word frequency, n-grams (sequences of adjacentwords), and part-of-speech tags. Generate a feature vector for each text instance, which represents the occurrence or frequency of each feature in the text.

D. Sentiment Classification

Split the dataset into training and testing sets, using a random sampling or cross-validation approach. Choose an appropriate classification algorithm, such as Support Vector Machines (SVM), Naive Bayes, or Random Forests. Train the classification algorithm on the training set, using the feature vectors and corresponding sentiment labels as input. Evaluate the performance of the trained model on the testing set, using metrics such as accuracy, precision, recall, and F1-score. Tune the hyperparameters of the classification algorithm to optimizethe performance of the model.

E. Product Analysis

Implement a product search function that takes user input and retrieves all product reviews that match the input product description. Apply the trained sentiment classification model to the retrieved product reviews, to obtain their sentiment labels. Compute the average sentiment score for each product, by taking the mean of the sentiment labels of allits reviews.

F. Sorting

Implement a sorting function that takes user input for preferred keywords and sorts the products based on their average sentiment score for the given keywords. Display the sorted list of products to the user.

G. Model Evaluation

Monitor the performance of the deployed model over time, using metrics such as accuracy, precision, recall, and F1-score. Continuously update the model by retraining it on new data or tuning its hyperparameters, to improve its performanceon new product reviews.

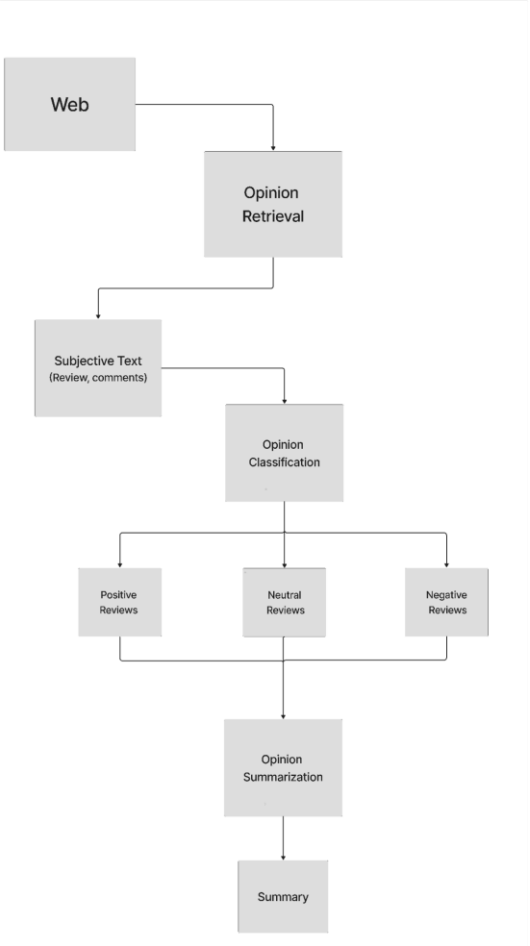

Figure 1: Flow chart of proposed method

IV. TECHNOLOGIES AND METHODS

A. Beautiful Soup

A Python library called Beautiful Soup is used to parse HTML and XML texts. It offers a means of data extraction, document structure transformation, and document tree navigation and search. Web scraper programs frequently use Beautiful Soup to gather data from websites.

B. Natural Language Toolkit (NLTK)

NLTK is a Python library that offers materials and tools for processing natural language. It has components for text categorization, tokenization, stemming, tagging, and parsing. In both study and instruction, NLTK is frequently used for natural language processing.

C. PorterStemmer

PorterStemmer is a popular English word stemming method. Words are stripped of common suffixes to return to their original shape. Applications for text extraction and information retrieval frequently use PorterStemmer.

Peer Reviewed & Refereed Journal
Volume-13, Issue-4 (August 2023)
https://doi.org/10.31033/ijemr.13.4.5

D. WordNetLemmatizer

Lemmatizing English words is made possible by the WordNetLemmatizer function in the NLTK library. Reducingwords to their most basic or dictionary version is known as lemmatization. Lemmatization is carried out by WordNetLemmatizer using WordNet, a lexical database of English terms.

E. Scikit-learn

Scikit-learn is a well-known Python machine learning framework. It offers tools for data preprocessing, feature selection, model assessment, classification, regression,and clustering. Various machine learning methods such as SVM(support vector machines) , naive Bayes and random forests are included in the Scikit-learn library.

F. CountVectorizer

CountVectorizer is a scikit-learn library tool that enables the transformation of text documents into a numerical representation. It generates a grid of word frequencies by tabulating the number of times each word appears in the text. The output of these methods for machine learning can then bethis matrix.

G. TensorFlow

Google created the well-known open-source machine learning platform known as TensorFlow. It offers resources for creating and refining machine learning models, such as deep learning and neural networks. TensorFlow is capable of handling a variety of machine learning jobs and supports both CPU and GPU processing.

H. Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are supervised machine learning algorithms that are used for regression and classification. SVMs divide data into various groups by locating the best hyperplane. Text classification jobs like sentiment analysis frequently make use of SVMs.

I. Naive Bayes

The Bayes theorem is the foundation of the Naive Bayes classification method. It is a probabilistic classifier that works under the premise that the characteristics are unrelated to one another. Based on the feature values, the algorithm determines the likelihood of each class and selects the most likely class as the predicted output. Applications for text categorization and spam filtering frequently use it

For a given dataset D of training data instances X = (X1, X2 … Xn) which has n attributes and m classes C1…..Cm the classifier predicts text X belongs to class which has a higher probability value for given condition.

$$P(C_i/X) > P(c_j/X) for 1 \leq j \leq m \neq i$$

$$P(C_i/X) = \frac{P(X/C_i)P(C_i)}{P(X)}$$

J. Pandas

Pandas is a free and open-source Python frameworkthat offers tools for manipulating and analyzing data. It offers a data frame structure to deal with structured data and is built on top of the NumPy library. Pandas is a crucial instrument for data analysis and visualization because it offers a variety of functions to clean, manipulate, and merge datasets.

K. Django

Django is a model-view-controller (MVC) architecturally sound web platform written in high-level Python. In order to create scalable web apps, it offers a complete set of tools and libraries. Working with databases is made simple by Django's object-relational mapper (ORM), which converts database records into Python objects. Additionally, it has built-in support for caching, account authentication, and templating. Django is a popular web programming tool for creating complex web.

V. CONCLUSION

Our paper proposes an approach where we use data mining, web crawling, parsing, and parts of speech tagging for the opinion mining process. By analyzing the sentiments expressed in online feedback, we were able to identify the key areas of concern for customers and help them make informed decisions about their purchases. Our approach has shown promising results in recommending products to customers based on their feedback. We believe that our research will have significant implications for the e-commerce industry, as it highlights the importance of leveraging customer feedback to enhance the online shopping experience.

FUTURE WORK

The review might be provided solely by text in the proposed paper. There is therefore potential for the future where the input can be provided via means such as speech-to- text, facial expressions and hand gestures, which can then be used by computers to interpret human emotions.

ACKNOWLEDGEMENTS

We would like to thank Prof. Wani Bisen, our research guide, for her invaluable support and guidance throughout our research. Her insightful contributions, encouragement, and critical analysis have been critical in shaping the direction and outcomes of our research.

We also want to thank Dr. A. J. Agrawal, the Department Head, for providing us with the resources and facilities we needed to carry out our research. His consistent motivation and support were critical to the project's success.

We'd like to thank them both for their unwavering support and guidance throughout this research project, which has assisted us in meeting our goals.

Peer Reviewed & Refereed Journal
Volume-13, Issue-4 (August 2023)
https://doi.org/10.31033/ijemr.13.4.5

REFERENCES

[1] Fang, X. & Zhan, J. (2015). Sentiment analysis data using product review. *Journal of Big Data, 2*(5). https://doi.org/10.1186/s40537-015-0015-2.

[2] Wankhade, M., Rao, A.C.S. & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. *Artif Intell Rev, 55*, 5731–5780. https://doi.org/10.1007/s10462-022-10144-1.

[3] Singh, J., Singh, G. & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. *Hum. Cent. Comput. Inf. Sci., 7*, 32. https://doi.org/10.1186/s13673-017-0116-3.

[4] Wikipedia contributors. (2023, March). Naive Bayes classifier in Wikipedia, The free encyclopedia. Available at: https://en.wikipedia.org/w/index.php?title=Naive_Bayes_classifier&oldid=11426027405.