

Product Reviews Using Machine Learning

¹Student, Chaitanya Kini, ²Student, Krishnamurthy Pai,
³Student, Nakul N Shet, ⁴Student, Sudhir Hegde

Canara Engineering College Mangalore, Karnataka

Canara Engineering College Mangalore, Karnataka

Canara Engineering College Mangalore, Karnataka

Canara Engineering College Mangalore, Karnataka

Submitted: 18-12-2022

Accepted: 31-12-2022

ABSTRACT: Sentiment analysis is used to examine and verify user comments in order to get opinions from them. Sentiment analysis is a machine learning(ML) technique that involves classifying and analyzing the sentiments, emotions, opinions, etc. that people have about products and express via text, rankings, thumbs up and down, and other means. Online product reviews that were gathered from the model website that we built were the source of the data for this study.

KEYWORDS: Sentiment, Machine Learning(ML), Convolutional Neural Network (CNN), Deep Learning, Opinions, Product Reviews

I. INTRODUCTION

Product reviews play a prominent role in the development of the product. By including customer feedback on the product website, you can assist your visitors in deciding whether or not to purchase a product. Reviews give visitors to your website the information they have to determine whether purchase something or not since the products could not be tested in real life. All of this information aids buyers in understanding the advantages and disadvantages of the product you're offering. Keep in mind that approximately 40% of shoppers claim they wouldn't purchase an electronic item without first reading internet reviews!

Reading reviews and keeping them in mind as you steer your company's route Sentiment analysis is a collection of methodologies, techniques, and tools for identifying and extracting subjective data from language, such as opinions and attitudes, to help determine how customers are feeling about buying a particular item or issue. It entails creating a system to gather and analyze product reviews submitted to numerous internet storefronts. By minimizing the human work required to read every single product review on the

website, this system offers to create a classifier to handle the aggregation of these mixed sentiments and provide a representation of these feedback in a manner that is simpler and easier to understand. You can determine where you stand in terms of overall client satisfaction by looking at your reviews..

II. LITERATURE SURVEY

In Xing Fang et al., [1] A sentiment analytical method for product reviews that makes use of the sentiment polarity categorization procedure is represented in this paper. Three phases make up the entire process. Naive Bayesian(NB), Support Vector Machine(SVM), and Random Forest(RM) have been selected as the categorization methods. Phase 1 of the evaluation process involves removing objective content and extracting subjective content from the data. Perform POS tagging on the extracted content after extraction. There are two phases namely Negation of adjective(NO) and negation of verb(NV) we can choose either of them to calculate the sentiment score for the feeling tokens. The feature vector for sentiment is constructed using the sentiment score formula. Phase 3's sentiment polarity categorization was the last step.

In Subhabrata Mukherjee, Pushpak Bhattacharyya et al., [2] Have encompassed method of feature-specific sentiment analysis, this paper looks at product reviews. The relationship between the characteristics and the opinions connected with them is captured using a dependency parsing technique. They created a system that gathers opinion expressions defining various aspects and extracts possible features from reviews. To either get solid experimental results, they used two datasets: For dataset 1 and dataset 2, the suggested system performs with an average accuracy of 80.98% and 70%, respectively.

In Jian jian, ping ji et al.,[3] present a technique for deriving particular customer intentions from product reviews obtained from the internet. The co-clustering approach that is being proposed summarises customers' concerns about various product features. Additionally, customers should give product designers a precise description of their needs. The method of random fields of conditions is used to simultaneously detect aspects of product attributes and thorough justifications, allowing for the early diagnosis of a dangerous illness.

According to Lada Banic and Ana Mihanovic et al., [4] use ML and NLP to investigate hotel review mining. They build a KNIME sentiment analysis framework-based product evaluation system configuration. The construction of a lexicon, data extraction from a database, and review rating are all steps in the sentiment analysis process.. They represent two evaluation systems that assigned a positive or negative rating to each term or phrase. Cumulative data regarding the quantity of good and negative reviews or phrases was produced at the level of the final review.. They keep a grade from 1 to 5 in the second evaluation. In which 1 is considered bad and 5 is considered excellent.

In Haruna Isah et al.,[5] represents lexicon-based & machine learning(ML) approach emotion categorization method using support vector machine(SVM) and naive bayes classifiers. For analysing the client experiences of well-known medication and cosmetic brands, a framework was developed. Before moving on to additional classifiers, a baseline result is obtained using the Nave Bayes classifier. The term "Naive Bayes classifiers" refers to a set of classification algorithms based on Bayes' Theorem. In most cases, the basic assumptions by Naive Bayes are incorrect in real-world situations. In fact, even though the independence assumption is never correct, it frequently works well in practise.

In Shilpa & Shereen, Rissa & Jacob, Susmi & Vinod, et al [6] Using Deep Learning Sentiment Analysis. They developed a system for sentiment classification of Posts on twitter in this paper. The tweets considered in the analysis are a mash-up of various words and emoticons. They produced a model for a classifier. using RNN and LSTM as examples of deep learning algorithms. To increase accuracy, they applied various feature selection techniques, including TF-IDF and oc2Vect. A vector is created during the feature extraction process and input into to the classification algorithm. When it came to categorising emotional tweets from Twitter, their

algorithm fared better than others. To make the system more individualised, future research should look at analysing the personalities of people based on their posts and retweets.

In Muhammet Sinan Bařarslan et al [7] The progress of the classification models was noted in the study using two datasets, one was considered from IMDB and the other one was in Twitter. ANN, SVM, and Nave Bayes were the classifier algorithms that were implemented in the experiments. After message processing the required vector space designing for sentiment classification, the datasets were divided into training set data and test data for sets using 5-fold cross validation for testing. The performance was assessed using the ACC, PREC, SENS, and F-measure methods. The classification experiments began with the Twitter dataset. After calculating all of the results from both experiments for all datasets, algorithms used for classification performances confirm each other. The performance of algorithms has values also had the same rates of success. ANN performed the best across all performance criteria.

In David Mathew Thomas, Sandeep Mathuret al [8] The goal of this work is to collect data from multiple sources using Python scripting language version 3.6 and software recognized also as web crawler Scrapy. To use its requirements, a database is built that gathers any unstructured data from multiple sources, assesses it, arranges, organises, cleans, re-analyzes, employs techniques and algorithms, and then provides the desired results. The strategies used in this online scraping methodology, which has existed for almost as old as the web, must first be understood in order to comprehend how well the data process of extracting has developed. Business online scraping has historically had the effect of gaining a quick competitive edge, including things like undercutting a competitor's unique valuation, stealing leads, hijacking marketing initiatives, funnelling APIs, and outright information theft.

In Pankaj, Prashant Pandey, Muskan, Nitasha Son et al,[9]This paper provides a schema for representing opinions. Additionally, this study offers an universal technique for identifying opinions characteristics in online reviews by comparing opinion feature statistics from two sets, one domain-specific and the other domain-independent. This product review paper study describes existing feature extraction techniques and approaches in sentimental analysis and opinion mining. A sentiment polarity categorization and POS process has been proposed, along with detailed descriptions of each step. These steps include pre-processing, pre-filtering, biasing, data

accuracy, and other machine learning-related features.

In Rajkumar S. Jagdale, Sentiment Analysis et al [10], is one of the most popular fields for analysing and deriving information and insight from text data from a wide variety of references, including Facebook, Amazon and Twitter. It is critical in allowing businesses to operate actively. A framework was created for

analysing customer experiences with well-known pharmaceutical and cosmetic brands. The Nave Bayes predictor is used to acquire a baseline result before going on to other classifiers. The dataset considered for this paper were extracted from reviews of different products in amazon.com such as laptops, mobiles, cameras, dresses etc. Sentiment Analysis benefits from a fine-grained approach at the aspect level.

Figure: Table Analysis

Author & Year	Dataset Used	Methodology	Drawback
Xing Fang et al., [1]	Reviews from amazon.com	Extraction using POS tagging	Ambiguity due to multiple meaning of some single words
Subhabrata Mukherjee, Pushpak Bhattacharyya et al., [2]	Datasets consisted of 500 reviews extracted from various domains	Linear Discriminant Analysis (LDA)	Since it is not trained on any domain-specific data, it cannot evaluate implicit sentiment that is domain-dependent.
Jian jian, ping ji et al., [3]	The data set consisted of positive classification which consists of 23938 tuples.	Supervised learning using naïve Bayes	It does not provide with the right accuracy for the labelled datasets
Lada Banic ,Ana Mihanovic et al., [4]	Datasets consisted of 750 reviews extracted from various domains through scraping the relative web pages	Web Crawler and Linear Discriminant Analysis	Websites have hidden data that can be manipulated to make the page appear irrelevant
Haruna isah and Paul Trundle et al., [5]	The collection of product reviews gathered from several e-commerce websites is the source of the data utilized in this paper.	Naïve Bayes	The data extracted will have a lot noise in it which makes the extraction inefficient.
Shilpa & Shereen, Rissa & Jacob, Susmi & Vinod, et al [6]	The data set was extracted from a larger dataset that contained one lakh tuples.	LSTM and RNN	Requirement of data is comparatively large for this model to perform better than the other techniques
Muhammet Sinan Başarslan et al [7]	4500 health-related tweets were collected for this study using the Twitter platform.	Artificial Neural Network(ANN)	This system's disadvantage is that it is unable accurately assess sentiment because it did not receive training on facts particular to the subject.

David Mathew Thomas, Sandeep Mathuret al [8]	Unstructured data from various sources	Web Crawler	It is not optimal for non-uniform structure of web with irregularity in organizations and structure
Pankaj, Prashant Pandey, Muskan, Nitasha Son et al,[9]	Reviews from flipkart.com	POS(Part of speech)tagging in machine learning	Not optimal as it provides with minimal accuracy during the pre-processing of the data
Rajkumar S. Jagdale, Sentiment Analysis et al [10],,	Product Reviews,Movie Reviews,Hotel Reviews	Machine Learning (ML)	Not a fine grained approach

III. DISCUSSION

Both in terms of categorization at the sentence and review levels, the experimental outcome shows promise. Since the averaged emotion score is able to attain an F1 score > 0.8 for such statement categorization with the entire set, it was shown to be a good feature on its own. The feature can generate an F1 score greater than 0.73 for such evaluation category using the entire set. There are, however, a few remaining issues with this study.

The first is that if we wish to categorise reviews according to their individual star ratings, it becomes challenging. At values lower than 0.5, the F1 scores derived from such studies are thus relatively low. The second drawback is that, because the study's suggested sentiment analysis method depends on the presence of sentiment tokens, it might not be effective for evaluations that only contain implicit attitudes. It might be challenging to determine the polarity of an implicit attitude because it is frequently expressed through neutral language. For instance, a common phrase in positive reviews like "Item as described." is composed entirely of neutral terms.

Our future work will concentrate on finding solutions to those problems in light of those constraints. To improve review-level categorizations, more features will specifically be extracted and arranged into feature vectors. Our next step in solving the implicit sentiment analysis problem is to be able to recognise the presence of implicit sentiment within the purview of a specific product. Future work will also involve using different datasets to test our categorization method.

IV. CONCLUSION

This experiment shows that sentiment analysis and opinion mining are crucial when deciding whether to purchase a certain good or

service. But when examining each review, it is crucial to take into account specific quality indicators like helpfulness, usefulness, and utility. About any new applications that adhere to the principles of data mining, sentiment classification or collaborative filtering can be used. Even while the algorithms and techniques used for sentiment classification are improving quickly and producing high-quality findings, many issues in this area of study are still open, and it might be challenging to spot false reviews simply by reading them. Sometimes phoney reviews are mistaken for genuine ones, and they are altered so that no one can tell what they were really trying to hide: Therefore, the detection of false reviews is another crucial area that calls for deep data mining approaches.

REFERENCES

- [1] Xing Fang ,Justin Zhan , "Sentiment Analysis using product review data", Springer: Journal of Big data", 2015, North Carolina A& T State university, Greensboro, NC, USA.
- [2] Subhabrata Mukherjee, Pushpak Bhattacharyya, "Feature Specific Sentiment Analysis for product Reviews", IET, 2015, IIT Bombay.
- [3] Jian Jin and Ping Ji, "Mining online product reviews to identify consumers FineGrinedConcerns", IET, 2015, Hong Kong SAR, China.
- [4] Lada Banic, Ana mihanovic, Marko Brakus, "Using Big Data and Sentiment Analysis in Product Evaluation "MIPRO, 2013, Croatia.
- [5] Haruna isah, Paul Trundle, Daneiel Neagu, "Social Media Analysis for Product Safety and using Text Mining and SA", IET, 2015, University of Bradford, UK.

- [6] P.c, Shilpa & Shereen, Rissa & Jacob, Susmi & Vinod, P.. (2021). Sentiment Analysis Using Deep Learning. 930-937. 10.1109/ICICV50876.2021.9388382.
- [7] Başarslan, Muhammet & Kayaalp, Fatih. (2021). Sentiment Analysis with Machine Learning Methods on Social Media. *Advances in Distributed Computing and Artificial Intelligence Journal*. 9. 5-15. 10.14201/ADCAIJ202093515.
- [8] D. M. Thomas and S. Mathur, "Data Analysis by Web Scraping using Python," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 2019, pp. 450-454, doi: 10.1109/ICECA.2019.8822022.
- [9] Pankaj, P. Pandey, Muskan and N. Soni, "Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 320-322, doi: 10.1109/COMITCon.2019.8862258.
- [10] Jagdale, R.S., Shirsat, V.S., Deshmukh, S.N. (2019). Sentiment Analysis on Product Reviews Using Machine Learning Techniques. In: Mallick, P., Balas, V., Bhoi, A., Zobia, A. (eds) *Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing*, vol 768. Springer, Singapore. https://doi.org/10.1007/978-981-13-0617-4_61