Sentiment Analysis of Amazon reviews with Natural Language Processing using Machine Learning Algorithms

Fauja Singh, Research Scholar, MGS University, Bikaner

Dr. Vishal Gaur, Research Supervisor, Govt. Engineering College, Bikaner

Abstract

Sentiment analysis utilises natural language processing (NLP) techniques to create software that interprets text similarly to human comprehension. Sentiment analysis is a crucial business intelligence tool that allows organisations to improve products and services by analysing digital text to determine the author's viewpoint on a subject. This information is employed to augment customer service and elevate brand reputation. Naive Bayes, Support Vector Machines, and Logistic Regression are commonly utilised in sentiment analysis to predict sentiment in Amazon reviews by analysing textual features and training the model on labelled data. These models aid organisations in understanding customer satisfaction, enabling data-driven decision-making, and leveraging the extensive collection of Amazon reviews to foster corporate growth and success. The primary objective of this study was to conduct a thorough examination of the sentiments expressed in Amazon reviews for a variety of product categories.

I. Introduction

A sentiment analysis of digital text figures out whether the emotional tone of a message is positive, negative, or neutral. The textual data that organisations have today includes emails, chat transcripts for customer service, comments on social media, and user reviews. An automated sentiment analysis tool can search this text to find out what the author thinks about a subject. Companies use the information they get from sentiment analysis to improve customer satisfaction and build the reputation of their brands.

A sentiment analysis model helps businesses make their products better by showing them which strategies work and which ones don't. Marketers look at comments on review sites, survey answers, and social media posts to get an insight into how certain features of a product work. The team can look at how people are acting in general to help either cut down on complaints or build on positive things that are happening.

Types of sentiment analysis

Business people use various types of sentiment analysis to get knowledge of how their customers feel when they use their products or services.

Fine-grained

Fine-grained sentiment analysis figures out what someone is trying to say and puts that meaning into different levels of emotional intensity. Typically, the method employs a numerical scale that spans from 0 to 100 to assess user sentiment, with each equal segment representing varying degrees of positivity, neutrality, negativity, and extreme negativity. Using a 5-star rating system as a thorough scoring system, e-commerce stores evaluate the quality of the purchasing experience.

Aspect-based approach

The aspect-based analysis method concentrates on specific features of a product or service. They use tools for sentiment analysis to find a link between what customers want and keywords related to the product.

Intention based

In market research, intention-based analysis is a useful way to figure out how customers feel. It helps marketers figure out which stage of the buying behaviour cycle a certain group of customers is in.

Emotion-based

Detection is the forensic study of how someone's mind was thinking while they were writing a text.

Methods for sentiment analysis

Sentiment analysis systems use three main approaches.

Rule-based

The rule-based approach assigns scores to targeted keywords by means of predefined lexicons, identifying and classifying them. Lexicons are collections of words meant to reflect the writer's goal, sentiment, and style.

Machine Learning

This approach generates accurate sentiment predictions of sentiment in unknown data by iteratively training a sentiment analysis model on known data. Data scientists use sentiment analysis datasets with numerous examples throughout the training process. Using the datasets as input, the machine learning program independently learns to reach the predefined conclusion. Sentiment analysis in machine learning has the advantage of exactly managing a large volume of textual data. When trained using enough data, machine learning sentiment analysis exactly predicts the emotional tone of the messages.

Hybrid

The hybrid sentiment analysis approach combines machine learning using rule-based systems. The process of putting the two separate systems together does take time and technical skill.

Sentiment Analysis of Amazon Reviews

Companies can look at Amazon reviews thanks to machine learning, which has helped them a lot. A small business located far away can still advertise its goods globally by registering on Amazon. Machine learning can guess how customers feel about things by using textual features and training models with labeled data.

By looking at how people feel in Amazon reviews, businesses can learn a lot about how satisfied their customers are and use that information to make their products or services better. By using the huge number of reviews on Amazon, sentiment analysis provides a scalable and useful way to get information that helps a business grow and succeed. You can find the specific improvements that your customers are looking for by looking at your products on Amazon and identifying a clear trend.

By means of a thorough assessment of these elements, companies can guarantee the dependability of their sentiment analysis instruments and obtain insightful knowledge on customer comments. Furthermore, the application of cross-valuation methods ensures the dependability of sentiment analysis models and helps validate their accuracy in guiding company decisions based on customer feedback. In essence, the success of a sentiment analysis model hinges on a meticulous evaluation of numerous elements, ranging from preprocessing techniques to algorithm selection, ultimately yielding significant insights that drive corporate development.

II. Literature review

Authors [1] evaluate the effectiveness of four machine learning models in gauging customer satisfaction with Amazon products. These results emphasize the challenges of large datasets, the limitations of traditional methods, and the unique capabilities of LDA to reveal emotional nuances. This elucidates all of these phenomena.

Researchers have specifically developed a modified CNN model for conducting sentiment analysis on e-commerce platforms using user reviews [2]. When applied to the dataset consisting of Amazon Product Reviews, the predictive model achieved a mean accuracy rate of 97.40%.

This method [3] can effectively identify complex patterns and relationships in the text. In contrast to traditional methods, this technique provides several notable advantages, such as improving the precision and efficiency of sentiment analysis.

We can analyze textual content, such as reviews, tweets, or customer feedback, to identify the emotions or attitudes expressed. We successfully meet this objective by using methods [4] from computational linguistics, machine learning, and natural language processing.

The application of sentiment analysis to customer reviews and tweets from Amazon, Flipkart, and Twitter gives valuable insights into employee conduct and customer sentiment, enabling

businesses to enhance customer service [5].

III. Methodology

Sentiment analysis is the application of natural language processing (NLP) methods to equip computer software with text understanding that quite closely approximates human comprehension. Usually, the research moves through several phases before showing the final result.
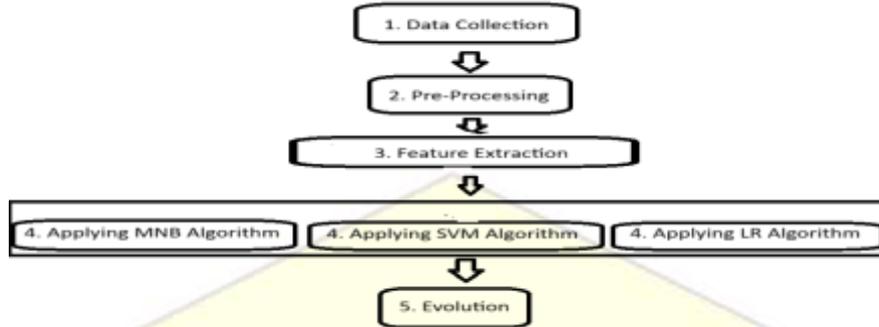

Figure-1

Data collection:

We obtained the dataset of Amazon reviews from Kaggle. It has 20,000 reviews, with data labelled 0 and 1. In this study, we classified the ratings into 1 and 0 values as follows: We assigned an indicative rating of 0 as a negative evaluation and 1 as a positive evaluation. Finally, we divided the data in this work into a larger training dataset of 15000 reviews and a smaller testing dataset of 5000 reviews.

Data Preprocessing:

To guarantee consistency and neatliness for analysis, we used the NLTK Python library to remove HTML tags, punctuation, and numerical values from the textual data. Standardise writing to be lowercase. We also removed stopwords to weed out often used words in sentiment analysis that have little bearing. An important first step in careful study is tokenizing text—that is, breaking it up into individual words or tokens. Tokenizing the text—that is, separating it into individual words or tokens—allows one to treat each unit separately for later study. Stemming and lemmatisation are techniques used to simplify the text by cutting words to their basic forms or root words. Through these techniques, the model can minimize noise and extraneous information in the text data, facilitating a focus on the genuine sentiment that the reviewers are trying to transmit. Finally, these text cleaning techniques are essential for preparing data for sentiment analysis and enhancing the model's ability to accurately classify consumer reviews.

Feature Extraction:

For analytical goals, it provides a relative estimate of the feeling expressed in written form. Researchers in customer review analysis use a 10-point rating system to indicate satisfaction and a 0-point rating system to indicate dismay.

Model selection:

Often, sentiment analysis uses the Naive Bayes classifier as an example. This classifier examines the words used to determine the likelihood of a review falling into a specific sentiment category. Support Vector Machines (SVMs) are an alternative model that does the best job of separating positive and negative reviews in a space with many dimensions. Here, we use logistic regression as the third model for sentiment analysis.

These models help companies to keep track of changes in public opinion over time, which lets them adapt their strategies as needed. By using these different models, businesses can accurately look at what customers say about Amazon apps and make beneficial choices to make them better.

IV. Results

The sentiment analysis tool used the Multinomial Naive Bayes classifier (MNB) to look at customer reviews and feedback from the Amazon reviews dataset. It accomplished this by

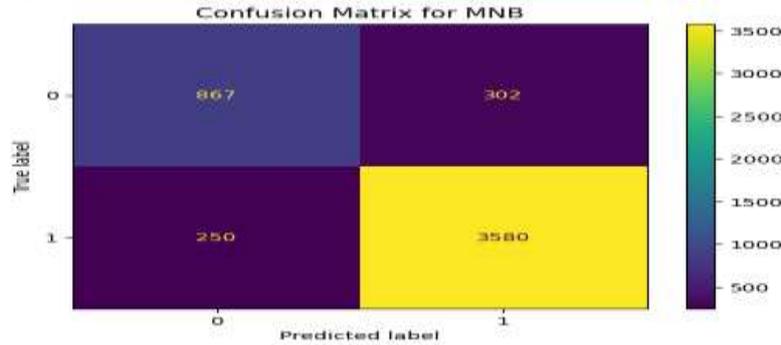precisely detecting emotions in textual data using the 88.95% accuracy score.

Figure-2

The Support Vector Machine (SVM) model gave an accuracy of 89.33% which is better than MNB classifier.
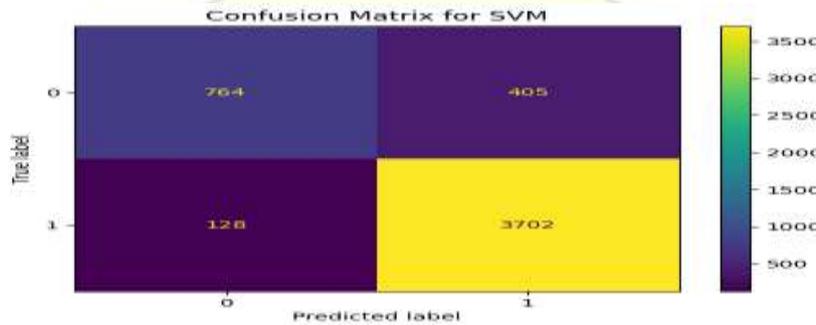

Figure-3

The Logistic Regression(LR) model obtained an accuracy statistic of 90.25%. The LR model has shown its higher efficiency in sentiment classification since it performed better than both previous models.
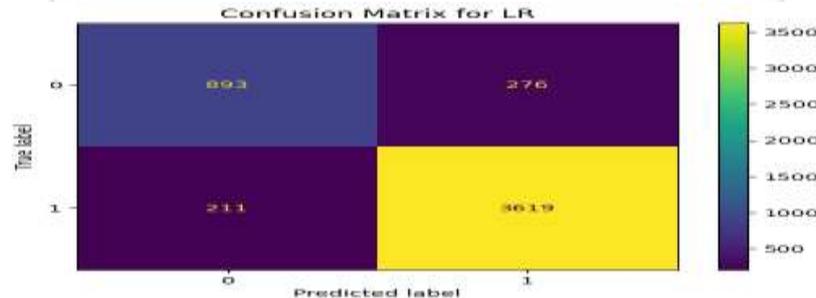

Figure -4

The Logistic Regression model outperformed other models, such as Multinomial Naive Bayes and Support Vector Machine models. They were able to correctly identify positive and negative emotions 88% to 91% of the time.

V. Conclusion

Textual materials, including reviews, tweets, or client comments, contain expressed attitudes classified and identified using sentiment analysis. By providing important new perspectives on consumer attitudes and employee behaviors, this method helps companies enhance customer service. Naive Bayes, Support Vector Machine, and Logistic Regression models accurately identify positive and negative emotions in sentiment analysis activities 88% to 91% of the time. This paper underlines the need for sentiment analysis in monitoring social media sentiment and enhancing tailored customer service contacts. Future directions for sentiment analysis could involve investigating new components, such as sentiment representations or sentiment dictionaries.

References:

1. Zijian, Qin., Kefan, Dong., Bing, Xie. (2024). (1) What affects customers online shopping behaviour?, research that applied machine learning to Amazon product

reviews. Applied and Computational Engineering, doi: 10.54254/2755-2721/76/20240564

2. Jiaqi, Li., Qi, Pan., Yihao, Wang. (2024). (6) Sentiment analysis applied to Amazon reviews. Applied and Computational Engineering, doi: 10.54254/2755-2721/44/20230079

3. Haili, Wang. (2024). (8) Word2Vec and SVM Fusion for Advanced Sentiment Analysis on Amazon Reviews. Highlights in Science Engineering and Technology, doi: 10.54097/sw4pft19

4. Ramdas, D., Gore., Vipin, Y., Borole., Rohit, S., Gupta., Viabhav, Sonaje., Prafulla, Chaudhari., Bharti, W., Gawali. (2023). (13) Sentiments Analysis on Amazon Product Reviews Using supervised machine learning algorithms. doi: 10.1109/icicis56802.2023.10430298

5. Padma, Adane., Avanti, Dhiran., Shruti, Kallurwar., and Sushmita, Mahapatra. (2024). (14) Sentiment Analysis of Product Reviews from Amazon, Flipkart, and Twitter. doi: 10.1007/978-981-99-9489-2_33

6. D., Geethanjali., P., Suresh. (2023). (20) Sentiment Analysis and Classification on Amazon Products Using Improved Support Vector Machine for Multiclass Classification. International Journal on Recent and Innovation Trends in Computing and Communication, doi: 10.17762/ijritcc.v11i10s.7619

7. En Nyaung, D., & Lai Lai Thein, T. (n.d.). Improved feature-based summarising and mining from hotel customer reviews.

8. Li, J., Pan, Q., & Wang, Y. (2024). Sentiment analysis was applied to Amazon reviews. Applied and Computational Engineering, 44(1), 26–32. https://doi.org/10.54254/2755-2721/44/20230079

9. Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. Artificial Intelligence Review, 54(7), 4997–5053. https://doi.org/10.1007/s10462-021-09973-3

10. Onan, A. (2021). Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. Concurrency and Computation: Practice and Experience, 33(23). https://doi.org/10.1002/cpe.5909

11. S. M. AlQahtani, A. (2021). Product Sentiment Analysis for Amazon Reviews. International Journal of Computer Science and Information Technology, 13(3), 15–30. https://doi.org/10.5121/ijcsit.2021.13302

12. Smeureanu, I., & Bucur, C. (2012). Applying supervised opinion mining techniques to online user reviews. In Informatica Economică (Vol. 16, Issue 2).

13. Sultan, N. (2022). Sentiment Analysis of Amazon Product Reviews using Supervised Machine Learning Techniques. Knowledge Engineering and Data Science, 5(1), 101. https://doi.org/10.17977/um018v5i12022p101-108

14. Verma, A., Rawat, C., & Gupta, Mrs. S. (2022). Sentiment Analysis for Amazon Product Reviews. International Journal of Recent Technology and Engineering (IJRTE), 11(2), 109–112. https://doi.org/10.35940/ijrte.B7099.0711222

15. Yu, B. (2024). Comparative Analysis of Machine Learning Algorithms for Sentiment Classification in Amazon Reviews. In Business, Economics, and Management, EMFRM (Vol. 2023).

16. Jitendra, Soni. (2024). (2) Sentiment Analysis of Amazon Reviews using NLTK Vader and Robert. Indian Scientific Journal Of Research In Engineering And Management, doi: 10.55041/ijsrem35700

Authors:
1. Fauja Singh (Research Scholar, MGS University, Bikaner)
2. Dr. Vishal Gaur (Govt. Engineering College, Bikaner)