SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

SENTIMENT ANALYSIS OF PRODUCT REVIEWS USING NATURAL LANGUAGE PROCESSING

Ms. Sathyapriya V P[1], Dr.P.Deepika[2]

[1]*Student, Department of Artificial Intelligence and Machine Learning, Dr.N.G.P Arts and Science College, Coimbatore*

[2]*Associate Professor, Department of Artificial Intelligence and Machine Learning, Dr.N.G.P Arts and Science College, Coimbatore*

ABSTRACT

The project "Sentiment Analysis of Product Reviews" employs the natural language processing (NLP) process and machine learning algorithm to scan and classify customer reviews based on their sentiment as positive, negative, or neutral and how AI can trace consumer behaviour and contribute significantly to customer happiness and decisions. It begins with data collection and preprocessing to prepare textual data for analysis, then followed by feature extraction to identify sentiment patterns precisely. Sentiment classification is accomplished using highly robust machine learning models trained on carefully labelled datasets to ensure reliability and accuracy of classification. The project's performance is thoroughly tested using corresponding metrics to ensure its utility and accuracy. By employing advanced data analysis techniques, this project empowers businesses to improve their goods, services find the next big trends and meet customer's needs more effectively. In addition, it is a scalable and cost-effective means of automating the process of analysing large amounts of customer reviews, thus being a must-have application for e-commerce sites as well as other consumer-driven industries.

KEYWORDS: *Sentiment Analysis, Product Reviews, Natural Language Processing, Machine Learning, Consumer Behaviour, Customer Satisfaction, Data Analysis, Feature Extraction, Sentiment Classification, Labelled Datasets, Performance Evaluation, E-commerce, Data-Driven Decision Making, AI in Analytics, Automated Review Analysis.*

INTRODUCTION

A social media website is defined as a website that allows people to meet people, think, find and share content, and build a community. This type of website will enable or promote a variety of types of activities, including commercial, social, or a combination of both. Social media covers multiple kind of digital repositories, online-shopping, media and discussion boards. It acts as a dynamic communication channel that empowers users to interact, share insights and engage in discussions transforming how individuals and organizations connect.

Companies have always been concerned about their public perception. In the pre-Internet era, prominence regime depended heavily on media range. Today, with many newspapers and opinion articles available online, organizations can effortlessly track positive or negative remarks through perspective analysis of news articles. However, this analysis is challenging due to diverse ways people express opinions and the subtlety of news content. Techniques often involve identifying specific keywords to determine the emotions behind the articles, classifying sentiments as positive, negative, or neutral.

LITERATURE SURVEY

J. Serrano-HuerRero. [1] Today, digital library 2.0 claims that it relies on client compounds through adjusted efforts and wiki. Today, Google Wave's language transmission frame must talk to a professional inspired by the research field. This structure considers the formation of a typical space through thought transactions between many scientists who are interested in the waves of joint efforts and lonely themes. One of the fundamental results of this technology is difficult to predict. Zhang et al. Z.B. Sun et al. The relationship between the two clients and the mark value is used to determine the distortion of the actual expression in the network of the client object. In particular, we use two bundles to collect fitting to create a variety of tractions. According to the true number, research shows that normal technology achieves preferred results compared to existing strategies. Nevertheless, this strategy has its limitations and disadvantages to try. Another obstacle is a similar name and does not know whether two different things are considered. H.L. Zheng et al. [3] The theme of how to work in the reality of the store chain has changed from modern exams to an important issue. In many models they have changed to solve this problem, but often focuses on similar ideas. Unlike excellent tasks, we ask for a way to provide an input mixed with the buyer's interpersonal organization. The proposed technology uses numerous upper hands than typical arrangements. First of all, it provides a solid process with two substances that can distinguish customers, including direct trust and restriction decisions among customers in the informal community. Second, in light of Pierson, it creates general constitution technology. Coefficient for reducing predictive values. Third, for further development of technology, individual effects are combined with a CF structure. Here we use the network method. Despite other data, there is a way to solve this problem. This is because it is very difficult to understand the problems to be presented by customers who are starting to freeze. Lieutenant Colombo-Mensdos [5] explained that checking the Internet business is an intentional strategy for selling items and

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

receiving customers. This article explains how to cope with web consumption. This hybrid methodology was accustomed to the solution to the general counselling structure. If it's absurd, the lawyer is not an unexpected appropriate support.

PROBLEM STATEMENT

The project plan and objective of the project is clear and this project is implemented till application level without any problem. Also, the risk involved in this project is very minimal which is negligible. This project can further be carried out in next phase as well it has wide scope. As currently four classification algorithms are used in project. The accuracy and performance of each algorithm is used further to predict with the client-side data and it can also be improved in future. The applications are built with minimal risk. So, from this we can analyze that the project and feasible and it is submitted on time. The theme of the feasibility study is to find out the application is technical, economical, operational and market feasible so that applications sustain for longer time and gives return on investment with good results. It can be used to inform future decisions.

METHODOLOGY

There are different methodologies and process models that can be used for a project involving sentiment analysis of Amazon reviews. However, one commonly used approach is the following:

- Problem definition: Define the problem statement and objectives of the project, such as identifying the sentiment of Amazon reviews to gain insights into customer opinions and preferences.
- Data collection: Gather relevant data, such as Amazon reviews, from various sources and store them in a suitable format.
- Data preprocessing: Clean and preprocess the data to remove noise, such as irrelevant words, punctuation, and stop words, and convert it into a suitable format, such as a bag of words or a TF-IDF matrix.
- Model selection: Select an appropriate machine learning model or algorithm for sentiment analysis, such as Naive Bayes, Support Vector Machines (SVM), or Recurrent Neural Networks (RNN).
- Model training: Train the selected model on the pre-processed data and fine- tune its hyperparameters to achieve optimal performance.
- Model evaluation: Evaluate the performance of the trained model using appropriate metrics, such as accuracy, precision, recall, and F1 score.
- Model deployment: Deploy the trained model into a production environment to analyze new Amazon reviews and generate sentiment scores.
- Feedback and improvement: Continuously monitor and evaluate the performance of the deployed model and collect feedback from users to identify areas for improvement and refine the model accordingly.

Overall, this process model follows a typical machine learning pipeline for sentiment analysis and can be customized based on the specific requirements and constraints of the project.

DATASET

Datafiniti_Amazon_Consumer_Reviews_of_Amazon_Products.csv (99.56 MB)

Detail Compact Column

id	dateAdded	dateUpdat...	name	asins	brand	categories	primaryCa...	imageURLs	keys
						Tablets,Kindle E-readers,iPad Accessories,...		_sd.jpg,https://c1.newegg...	rs/b00zv9pxp2,allnewkindl...
AVqVGZNvQMlgsOJE6eUY	2017-03-03T16:56:05Z	2018-10-25T16:36:31Z	Amazon Kindle E-Reader 6" Wifi (8th Generation, 2016)	B00ZV9PXP2	Amazon	Computers,Electronics Features,Tablets,Electronics,iPad & Tablets,Kindle E-readers,iPad Accessories,...	Electronics	https://pisces.bbystatic.com/image2/BestBuy_US/images/products/5442/5442403_sd.jpg,https://c1.newegg...	allnewkindleereaderblack6glarefreetouchscreendisplaywifiincludesspecialoffers/b00zv9pxp2,allnewkindl...
AVqVGZNvQMlgsOJE6eUY	2017-03-03T16:56:05Z	2018-10-25T16:36:31Z	Amazon Kindle E-Reader 6" Wifi (8th Generation, 2016)	B00ZV9PXP2	Amazon	Computers,Electronics Features,Tablets,Electronics,iPad & Tablets,Kindle E-readers,iPad Accessories,...	Electronics	https://pisces.bbystatic.com/image2/BestBuy_US/images/products/5442/5442403_sd.jpg,https://c1.newegg...	allnewkindleereaderblack6glarefreetouchscreendisplaywifiincludesspecialoffers/b00zv9pxp2,allnewkindl...
AVqVGZNvQMlgsOJE6eUY	2017-03-03T16:56:05Z	2018-10-25T16:36:31Z	Amazon Kindle E-Reader 6" Wifi (8th Generation, 2016)	B00ZV9PXP2	Amazon	Computers,Electronics Features,Tablets,Electronics,iPad & Tablets,Kindle E-readers,iPad Accessories,...	Electronics	https://pisces.bbystatic.com/image2/BestBuy_US/images/products/5442/5442403_sd.jpg,https://c1.newegg...	allnewkindleereaderblack6glarefreetouchscreendisplaywifiincludesspecialoffers/b00zv9pxp2,allnewkindl...

Fig.No: 1 Sample Dataset

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

EXECUTION RESULTS & DISCUSSION

In order to obtain optimal result, each algorithm is analyzed based on five performance metrics they are accuracy, precision, recall, f1-score, confusion matrix.

Confusion Matrix

A confusion matrix is a table that summarizes the performance of a classification model by comparing the predicted labels with the true labels. It consists of four entries: true positives, false positives, true negatives, and false negatives. The diagonal of the matrix shows the number of correct predictions, while the off-diagonal entries show the misclassifications made by the model.

Accuracy

Accuracy measures the proportion of correct predictions made by the model.
Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision

Precision measures the proportion of true positive predictions among all positive predictions made by the model.
Precision = TP / (TP + FP)

Recall

Recall measures the proportion of true positive predictions among all actual positive samples in the dataset.
Recall = TP / (TP + FN)

F1 score

F1 score is the harmonic mean of precision and recall. It is a useful metric when both precision and recall are equally important. Mathematically, F1 score is defined as:

F1 score = 2 * (precision * recall) / (precision + recall)

We will ignore the first row and column because we already replaced all NAN with " ". This is the identical circumstance as was described in the classification of report. By looking at row 2-4 and column 2-4, which are labelled as negative, neutral, as well as positive, with scores of 246 and 104, respectively, we see that positive sentiment can occasionally be confused with neutral and negative sentiment. Even so, confusion scores of 246 and 104 for neutral as well as negative ratings, respectively, are considered negligible when compared to the total amount of significant positive sentiment at 6445.

This is also the result of a positively skewed dataset, which is consistent with our data exploration and sentiment analysis. As a result, we conclude that the products in this dataset are generally well-liked and should remain on the product roster.
Precision: determines how many objects selected were correct.

Table *6.1 Classification report of Algorithm*

	Precision	Recall	f1-score	support
	0.00	0.00	0.00	5
Negative	0.67	0.25	0.36	156
Neutral	0.47	0.11	0.18	292
Positive	0.95	1.00	0.97	6473
AVG/TOTAL	0.92	0.94	0.92	6926

Confusion Matrix: array([[0,0 ,0 ,4],
[0,32,7 ,119],
[0,14,21,266],
[0,4 ,21,6438]], dtype=int64).

CONCLUSION

In conclusion, while more data is required to balance out the lower rated products in order to consider their significance, we were able to successfully associate with positive, neutral, and negative sentiments for every product in Catalog.

FUTURE ENHANCEMENTS

The future work for a project on analyzing Amazon reviews using sentiment analysis would depend on the specific goals and objectives of the project, but here are some potential ideas: Improve the accuracy of the sentiment analysis, Expand the analysis to other marketplaces, incorporate user demographics, compare sentiment analysis to sales data, and Use sentiment analysis for product development. Overall, there are many different directions that a project on Amazon reviews using sentiment analysis could take in the future, depending on the specific goals of the project and the needs of the business or organization conducting the analysis.

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

REFERENCES

1. *Dataset:https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products*
2. *Anh, K.Q., Nagai, Y. and Nguyen, L.M., 2019. Extracting customer reviews from online shopping and its perspective on product design. Vietnam Journal of Computer Science, 6(01), pp.43-56.*
3. *Chong, A.Y.L., Li, B., Ngai, E.W.,Chang, E. and Lee, F., 2016. Predicting online product sales via online reviews,sentiments, and promotion strategies: A big data architecture and neural network approach. International Journal of Operations & Production Management.*
4. *Chauhan, U.A., Afzal, M.T., Shahid, A., Abdar, M., Basiri, M.E., and Zhou, X., 2020.*
5. *R. Collobert, J. Weston, L. Bottou, M. Karlen,K. Kavukcuoglu, and P. Kuksa. Natural language processing (almost) from scratch. Journal of Machine Learning Research, 12(Aug):2493–2537, 2011.*
6. *K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web, pages 519–528. ACM, 2003.*
7. *M. S. Elli and Y.-F. Wang. Amazon reviews, business analytics with sentiment analysis.*
8. *S. Hota and S. Pathak. Knn classifier based approach formulti-class sentiment analysis of twitter data. In International Journal of Engineering Technology, pages 1372–1375. SPC, 2018.*
9. *B. Liu and L. Zhang. A Survey of Opinion Mining andSentiment Analysis, pages 415–463. Springer US, Boston, MA, 2012.*
10. *C. Rain. Sentiment analysis in amazon reviews using probabilistic machine learning. Swarthmore College, 2013.*
11. *R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning,A. Ng, and C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedingsof the 2013 conference on empirical methods in natural language processing, pages 1631–1642, 2013.*
12. *Y. Xu, X. Wu, and Q. Wang. Sentiment analysis of yelps ratings based on text reviews 2015.*
13. *P.Deepika., Dr.S.Sasikala., Data Classification Pertaining To Heart Diseases Using Hybrid Chicken Swarm Optimization With Artificial Neural Network (HCSOANN), International Journal of Mechanical Engineering, Volume 7, 2022.*
14. *P.Deepika, Dr.S.Sasikala., Effective Anticipation Of Cardiovascular Disease Using Opt-Ann Deep Learning Techniques, Solid State Technology, Vol-63, Issue-6, 2020.*
15. *P.Deepika, S.Saranya, Dr.S.Sasikala, Comprehensive Review On Heart Disease Prediction Using Optimization Techniques, EPRA International Journal of Research and Development, 2020.*
16. *P.Deepika, S.Saranya, Dr.S.Sasikala., R.Sowparnika., S.P.Divyabharathi., Instigating Effective Cloud Computing Techniques For Elevating Agriculture Domain, Journal of Emerging Technologies and Innovative Research (JETIR), Vol-6, Issue 1, 2019.*
17. *P.Deepika, S.Saranya., Dr.S.Sasikala., An Effective Archetype Design Of Heart Disease Anticipation Using Optimization Techniques" EPRA International Journal of Research and Development, Vol-4, issue-11, pp 33-40, 2019.*
18. *P. Deepika, A.Kiruthika, S.Sasikala, S.Saranya., Predicting The Ailment Of Thyroid Using Classification And Recital Indicators, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol3, Issue-3, 2018.*
19. *P.Deepika., S.Sasikala, S.Saranya, A Quintessence For Focused Crawler Enhanced With Link Spam Detection, IOSR Journal of Engineering, 2018.*
20. *P.Deepika, Dr.S. Sasikala., S.Saranya., A.Kiruthika., A Novel Classification And Prediction Algorithm For Heart Disease Identification, International Journal of Engineering and Science Invention, ISSN: 2319 – 6734, Vol-7, Issue 2 - Version- 1, February 2018.*
21. *P.Deepika, S.Sasikala, S.Jansi., S.Saranya, A.Kiruthika, Anticipating The Chronic Kidney Disorder Using Performance Optimization In Adaboost And Multilayer Perceptron, Imperial Journal of Interdisciplinary Research, 2017.*
22. *P.Deepika, S.Saranya, S.Sasikala, Dr.S.Jansi, A.Kiruthika, Anticipating Heart Disease Using C4.5 Classification Augmented With Feature Selection, International Journal for Research and Development in Technology, Vol-6, issue-5, 2016.*
23. *P.Deepika, P.Vinothini., Heart Disease Analysis And Prediction Using Various Classification Models – A Survey in Indian journal of Research PARIPEX Volume :4 Issue :3 March 2015.*
24. *K.Thenmozhi., P.Deepika, M.Meiyappasamy, Different Datamining Techniques Involves In Heart Disease Prediction-A Survey, International Journal of Scientific Research, Vol-III, Issue-IX, 2014.*
25. *P.Deepika, K.Thenmozhi, Heart Disease Prediction Using Classification With Different Decision Tree Techniques, International Journal of Engineering and General Research, Vol-2, Issue-6, 2014.*