Universal Library of Engineering Technology | Research Article

ISSN: 3064-996X

Open Access | PP: 10-17

DOI: https://doi.org/10.70315/uloap.ulete.2022.002

Leveraging Big Data Analytics and Machine Learning Techniques for Sentiment Analysis of Amazon Product Reviews in Business Insights

Chethan Sriharsha Moore[1], Purna Chandra Rao Chinta[2], Niharika Katnapally[3], Krishna Ja[4], Kishan Kumar Routhu[5], Vasu Velaga[6]

[1]Papa Johns, Sr Devops Engineer.

[2]Microsoft, Support Escalation Engineer.

[3]AWS – QuickSight BI Engineer.

[4]Topbuild Corp, Sr Business Analyst.

[5]AT&T, MGM Systems, Senior Deployment Engineer.

[6]Cintas Corporation, SAP Functional Analyst.

Abstract

Sentiment research is essential for comprehending consumer input and raising the calibre of goods and services. This study looks at the dataset of Amazon product evaluations and how sentiment analysis using ML approaches was done. Several ML methods, such as Gradient Boosting (GB), Logistic Regression (LR), Naïve Bayes (NB), and Recursive Neural Network for Multiple Sentences (RNNMS), are used in this research to analyse the sentiment of Amazon product reviews. The approach begins with preprocessing the dataset by removing punctuation, filtering stop words, and tokenising the text, followed by feature extraction using techniques like Bag of Words (BoW). The models are evaluated using the F1-score, recall, accuracy, and precision once the data is separated into training and testing sets. Among a models tested, Gradient Boosting outperforms the others with a consistent 82% in all metrics, demonstrating its strong classification ability. The outcomes show that while GB provides a highest performance, future work could explore advanced models and techniques to further enhance sentiment classification accuracy across diverse product categories.

Keywords: *Sentiment Analysis, Big Data, Business, Machine Learning, Amazon Product Review Dataset, BOW.*

INTRODUCTION

Consumers rely heavily on online evaluations when making purchases in today's fast-paced digital age, due to the proliferation of e-commerce platforms. Businesses may learn a lot about what consumers want and how to improve their goods and services from these evaluations, which in turn affect the purchasing decisions of prospective consumers[1]. Among e-commerce platforms, Amazon stands as a global leader, hosting millions of product reviews that capture diverse customer sentiments. Extracting meaningful insights from this vast pool of unstructured data, however, poses a significant challenge[2].

An essential part of NLP is sentiment analysis, which is used to make sense of the views expressed in text [3]. It leverages machine learning algorithms, lexicon-based techniques, and hybrid models to classify customer sentiments as positive, negative, or neutral[4]. By analysing these sentiments, businesses gain deeper insights into customer satisfaction, product quality, and market trends, enabling data-driven decision-making and enhanced customer experience.

An integration of big data analytics with sentiment analysis has further revolutionised the field, enabling businesses to process and analyse vast datasets with high velocity and variety[5]. The explosive growth of data generated through reviews, social media interactions, and other digital platforms underscores the need for robust analytical frameworks. Big data technologies not only handle the scale of this information but also facilitate deeper insights into customer behaviour and emerging trends. By coupling big data with sentiment analysis, organizations can uncover patterns, identify pain points, and predict customer needs with unprecedented precision[6].

Advancements in AI and ML have made this even more practical with predictive analytics, which use past data to make predictions [7]. In the context of sentiment analysis, AI and ML algorithms enable automated learning from complex datasets, identifying subtle patterns and relationships that are often imperceptible to traditional methods[8]. By leveraging predictive models, businesses can anticipate customer preferences, forecast product demand, and enhance decision-making processes[9]and identifies the best part sequence available in the part-mix. A mathematical model has been formulated to minimize the broad objectives of set-up cost and time simultaneously. The proposed approach has more realistic attributes as fixture related intricacies are also taken into account for model formulation. It has been solved by a new variant of particle swarm optimization (PSOThis synergy between sentiment analysis, big data, and predictive analytics not only empowers e-commerce businesses to stay competitive but also fosters a customer-centric approach to innovation.

Significance and Contribution

This paper offers a thorough method for analysing Amazon product evaluations for sentiment, with a focus on using machine learning methods to make sentiment categorisation more accurate. The work is significant as it explores anemploy of widely-used NLP methods and advanced ML algorithms, ensuring the results have practical implications for e-commerce platforms and customer experience management. Here are some important points from this study:

- Collect Amazon product review dataset for sentiment analysis.
- Developed a preprocessing for including tokenisation, stop word removal, and punctuation filtering.
- Utilized BoW for feature extraction to convert textual data into numerical form.
- Implemented and compared multiple ML models, including GB, LR, NB, and RNNMS for sentiment classification.
- Assess the efficacy of the model by calculating its recall, accuracy, F1-score, and precision.

Structure of the Paper

The study is structured as follows: Section II presents relevant work on sentiment analysis. Section III details the procedures and materials used. section IV presents the experimental findings of a proposed system. The inquiry and its outcomes are summarised in Section V.

LITERATURE REVIEW

This section discusses the surveys and reviews articles on Sentiment Analysis of Amazon Product Reviews in Business Insights.

In this paper, Singla, Randhawa and Jain et.al. (2018) The goal of the computer research known as "sentiment analysis" is to glean subjective information from written texts. Using Sentiment Analysis, the proposed work sorts over 4,000 reviews into positive and negative attitudes. DT, NB, and SVM are the categorisation models that have been used for reviews. They use 10-fold cross-validation to evaluate the models[10]digital reviews play a pivotal role in enhancing global communications among consumers and influencing consumer buying patterns. E-commerce giants like Amazon, Flipkart, etc. provide a platform to consumers to share their experience and provide real insights about the performance of the product to future buyers. In order to extract valuable insights from a large set of reviews, classification of reviews into positive and negative sentiment is required. Sentiment Analysis is a computational study to extract subjective information from the text. In the proposed work, over 4,000,00 reviews have been classified into positive and negative sentiments using Sentiment Analysis. Out of the various classification models, Naïve Bayes, Support Vector Machine (SVM.

In This study, Rain (2013) aims to use and expand upon existing research in sentiment analysis and NLP by applying it to Amazon data. Classifying reviews as favourable or negative is done using decision list classifiers and NB. Supervised ML makes use of user-rated product ratings as training data [11].

In this study, Bali et.al. (2016) the proposal for consumer sentiment analysis involves mining user attitudes to generate product popularity, which in turn will lead to "personalised" outcomes; these personalised results are essential in today's client-centric environment. Thus, this is useful for firms who are looking to broaden their user base and develop more effective retail techniques to sell their goods [12].

In this study, Haque, Saber and Shah et.al. (2018) It is easy to grasp since it has employed simpler algorithms. Unfortunately, the system's great accuracy on svm prevents it from performing well on massive datasets. They accomplished this using DT, LR, and SVMs. The tfidf is used in this context as an auxiliary experiment. Utilising a word bag, it is able to derive ratings. However, we are just using a small number of classifiers here. A linear regression model was used, consisting of RMSE. In an effort to streamline our processes, we drew inspiration from the aforementioned connected works and implemented their finest concepts harmoniously[13].

In this study, Wladislav et.al. (2018) we present Sentilyzer, an advanced system that analyses Amazon product evaluations for sentiment and aspects using dynamically generated dictionaries. It achieves advanced aspect-oriented sentiment analysis[14].

In this study, Singla Randhawa and Jain et.al. (2017) analyse data from a big collection of mobile phone evaluations that people have posted online. They have incorporated not just positive and negative attitudes but also anger, anticipation, contempt, fear, pleasure, grief, surprise, and trust in our classification system. By categorising reviews in this way, we can better assess the product as a whole, which in turn helps customers make more informed purchases[15].

Table I provides the comparative analysis of Sentiment Analysis of Amazon Product Reviews in Business Insights based on the datasets, findings, limitations, and future work.

Table I. Summary of Background Study on Machine Learning Techniques for Sentiment Analysis of Amazon Product Reviews

Study	Methods	Key Findings	Dataset	Limitations & Future Work
Singla, Randhawa and Jain et.al.	NB, SVM, DT; 10-Fold Cross Validation	Classification of over 400,000 reviews into positive and negative sentiments.	400,000 Amazon reviews	Limited classification methods; future work could include advanced deep learning models for improved accuracy.
Rain et.al.	Naive Bayes, Decision List Classifiers; Supervised Machine Learning	Utilised the star ratings of products to tag reviews as positive or negative with supervised learning.	Data retrieved from Amazon	Relies on user star ratings, which may not always accurately reflect review sentiment; potential for incorporating context-sensitive models in future.
Bali et.al.	Sentiment Analysis, Popularity Mining	Proposed consumer sentiment mining to generate popularity insights for personalised recommendations and retail strategies.	User sentiment data	Focused on consumer-centric strategies without detailed model performance metrics; future work could apply state-of-the-art sentiment models for better insights.
Haque, Saber and Shah et.al.	SVM, Logistic Regression, Decision Trees; TF-IDF, Bag of Words (BoW)	Achieved high accuracy with SVM but struggled with large datasets.	Amazon reviews (unspecified size)	Limited scalability and classifiers; future work could improve computational efficiency and explore ensemble models for better performance on large datasets.
Wladislav et.al.	Sentilyzer System, Aspect-Oriented Sentiment Analysis	Performed advanced aspect-oriented sentiment analysis using category-specific sentiment and aspect dictionaries from reviews.	Amazon Product Reviews	Focuses only on aspect-oriented sentiment; future directions may include integrating broader sentiment paradigms.
Singla, Randhawa and Jain et.al.	Multi-class Sentiment Classification; Extended Sentiments (e.g., anger, joy, trust)	Classifies reviews into detailed sentiments for better evaluation and decision-making processes for consumers.	Online reviews for mobile phones	Requires extensive labelled data for multi-class sentiment tasks; future work could incorporate unsupervised learning for reduced dependency on labelled datasets.

METHODOLOGY

There are a number of steps to the process of utilising AI and ML algorithms to determine the tone of Amazon customer reviews. First, the Amazon product review dataset is preprocessed to enhance data quality, including removing punctuation, filtering out stop words, tokenising text. Next, features are extracted employing techniques such as BoW. The dataset is then split into an 80-20 ratio for training and testing, respectively which transform textual data into numerical representations suitable for analysis. These features are then input into a classification algorithm, which categorises the reviews into Positive, Negative, or Neutral sentiments. Then implement ML models like GB, RNNMS, LR, and NB. Lastly, to guarantee the efficacy and dependability of the sentiment classification process, a performance is assessed employing conventional measures like recall, accuracy, precision, and F1-score. A flowchart of Amazon product reviews is in Figure 1.

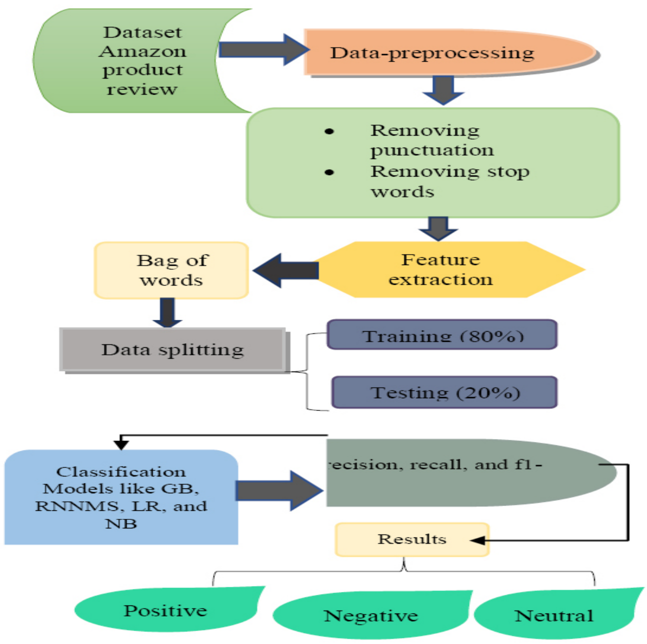

Fig. 1. Flowchart for Amazon product review

The following steps of the flowchart are briefly explained in below:

Data Collection

Customer reviews for Amazon product reviews are available dataset. It contains information that was scraped from the Amazon site, such as the product name, reviewer name, review content, rating, and timestamps with 6823 samples. The Amazon product review dataset visualisation graphics are provided below:

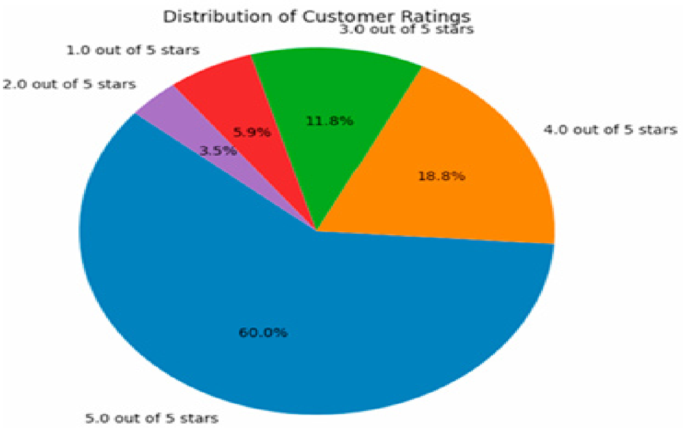

Fig. 2. Distribution of customer rating in the input data

The pie chart in Figure 2 illustrates the distribution of customer ratings, revealing that 5.0 out of 5 stars is the most common rating, accounting for 60.0% of all ratings, followed by 4.0 stars at 18.8%. Ratings of 3.0 stars represent 11.8% of the total, while 2.0 and 1.0 stars are the least frequent, comprising 5.9% and 3.5%, respectively. This distribution highlights a tendency for customers to provide higher ratings (4 or 5 stars) more frequently than lower ratings (1 or 2 stars).

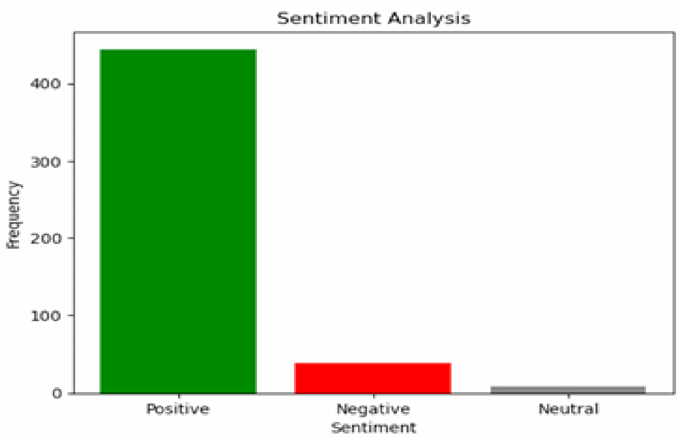

Fig. 3. Frequency of sentiment in the input data

The sentiment analysis of the input data reveals a strong positive bias. A majority of the data points express positive sentiments, as evidenced by the significantly taller bar representing positive sentiment in the figure 3. The data shows a predominance of positive sentiment, with negative sentiment being less frequent and neutral sentiment the least common, suggesting an overall positive sentiment.

Fig. 4. Heatmap for Amazon product review

This heatmap in Figure 4 visualises the correlations between various features in Amazon product reviews. Sentiment has a moderate positive correlation (0.31) with overall rating, indicating a connection between sentiment analysis scores and user ratings. Review length (review Len) and sentiment show a weak negative correlation (-0.15), suggesting longer reviews might slightly dampen sentiment scores. Word count and review length are perfectly correlated (1), as expected due to their direct relationship. Overall, the chart highlights weak or negligible relationships among most features except for word count and review length.

Data Preprocessing

Data preparation involves eliminating columns with unneeded values, such as those with missing, false, or null values [16]. This step is performed to eliminate these superfluous and missing columns since our dataset might be noisy at times. The following pre-processing steps are as follows:

Removing punctuation: Removing the punctuation in the review text column for simplicity.

Removing stop words: "Stop words" are frequent terms that readers may assume don't convey anything about the review's tone. So, "I" and "is" are two examples.

Tokenization: Tokenization is the act of converting a string of text into a collection of discrete tokens [17]. The building blocks of text segments are called tokens. They could take the form of individual words, whole sentences, or part phrases.

Bag of Words (BoW) for Extract Features

The bag-of-word method is used in NLP and information retrieval to represent reduced text or data in order to extract characteristics [18] . This approach depicts a document or text as a bag, or multiple set, of words. Consequently, sentiment analysis's "bag of words" represents the process of compiling a set of relevant terms [13]. The feature sets were extracted using the bag of words method.

Data Splitting

This stage involves dividing the dataset into two sections:

testing and training. Building the model is done using the training subset, and assessing the model's performance is done with the testing subset. In this paper, splitting the dataset into 80% for training and 20% for testing.

Classification with Gradient Boost Classifier Model

The ML method known as Gradient Boosting is quite effective for classification and regression jobs. Building on the idea of boosting, it combines a number of weak models (usually [19] DT) to produce a powerful, accurate prediction model[20] environmental and economical concerns draw considerable attention from both practitioners and researchers towards remanufacturing practices. The success of remanufacturing firms depends on how efficiently the recovery process is executed. Radio Frequency Identification (RFID. Everything from the loss function to the underlying learner models is up for grabs. Discovering the solution to the parameter estimates provided a certain base-learner h(x, θ) and/or a customised loss function φ (y,f) might really be difficult. This was addressed by suggesting that a new function h(x, θt) be selected as the one that is most parallel to the negative gradient along the observed data (1):

$$g_t(x) = E_y[\frac{\partial \varphi(y,f(x))}{\partial f(x)} x]_{f(x)=f^{t-1}(x)} \quad (1)$$

The new function increment that has the highest correlation with −gt(x) may be selected rather than searching the function space for the general solution for a boost increment [7]. This makes it possible to substitute the traditional least-squares minimisation work with a potentially very challenging optimisation challenge (2):

$$(\rho_t, \theta_t = \arg min \sum_{i=1}^{N}[-g_t(x_i) + \rho h(x_i, \theta)]^2 \quad (2)$$

The design decisions of φ (y,f) and (x, θ) will have a significant impact on the precise by of a generated algorithm with all the relevant formulae.

Evaluation Measures

This study assesses sentiment analysis by calculating the positivity and negativity of reviews and then utilising classification measures such as Accuracy, Precision, Recall, and F-Measure [21][22]. A confusion matrix is utilized to assess model performance by comparing actual and predicted values, consisting of TP, FP, TN, and FN. The following matrix is explained below:

Accuracy: Accuracy, the ratio of correctly anticipated observations to total observations, is the most basic and straightforward natural performance measure. The following eq (3):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Precision: The precision of a test is proportional to the proportion of valid results relative to the total number of outcomes (including both true and false positives). As seen in Eq. (4), the typical formula used for this purpose is:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (4)$$

Recall: The ratio is defined as TP divided by the total of TP and FN. Sensitivity is another name for it. Here is the following equation (5) for recall:

$$Recall = \frac{\mathrm{TP}}{TP+FN} \quad (5)$$

F1-score: Accuracy in both recall and precision are weighted. Retrieving relevant hot spots and accurately detecting the kind of malignancy are shown by the classifier. Its computation is described in Eq. (6).

$$F - Measure = \frac{2(Precision \ *Recall)}{\mathrm{Precision\ +Recall}} \quad (6)$$

ROC Curve: This is a probability curve that represents several classes. It provides insight into the model's performance in classifying inputs.

RESULT ANALYSIS AND DISCUSSION

Evaluating the performance of a Machine learning model requires the calculation of the accuracy, precision, recall, and F-score. The evaluation metrics for the Gradient boost classifier model on the Amazon product review dataset result in Table 2. In the below table 4, the accuracy for the models like LR[23], Naïve Bayes[23], and RNNMS[24] is compared with gradient boost and achieves the highest accuracy.

Table II. Gradient Boost model Performance for sentiment analysis on the Amazon product review dataset

Measures	Gradient boost (GB)
Accuracy	82
Precision	82
Recall	82
F1-score	82

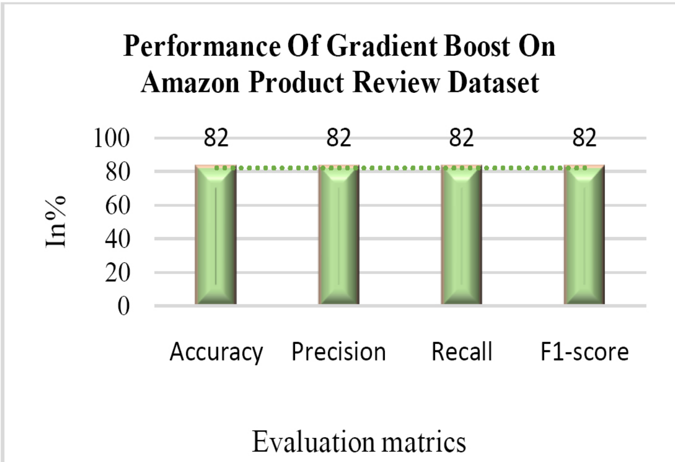

Fig. 5. Performance of Gradient Boost Mod

An above Table II and Figure 5 show the model performance sentiment analysis for Amazon product reviews. Gradient boost classifier model achieves excellent performance with accuracy, precision, recall and F1-score all are same 82%, indicating strong classification ability with minimal false positives and high detection of relevant instances.

Table III. Classification report for gradient boost classifier model

Class	Precision	Recall	F1-score	Support
0	0.83	0.82	0.83	891
1	0.81	0.81	0.81	811
Accuracy	0.82			1702
Macro Avg	0.82	0.82	0.82	1702
Weighted Avg	0.82	0.82	0.82	1702

The following Table III displays a Classification report for the Gradient boost classifier model. The report indicates that the model attains a precision83%, recall82%, and f1-score83% and Support891 for class0, while achieving a precision81%, recall81%, and f1-score83% and support811 for class1.

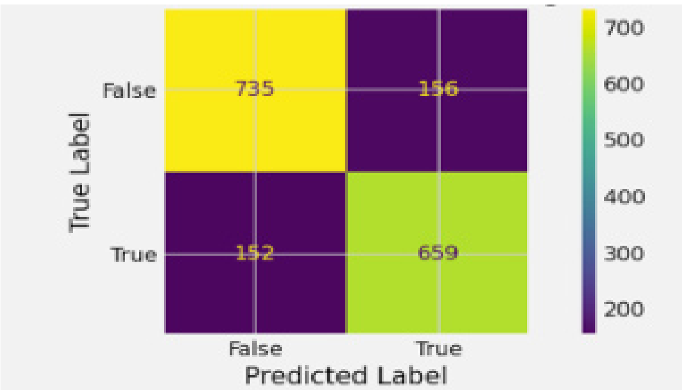

Fig. 6. Confusion matrix for Gradient boost classifier

Figure 6 demonstrated its overall performance by correctly predicting 659 instances as True (True Positives) and 735 instances as False (True Negatives), showcasing its ability to classify data accurately. Though it did correctly identify 152 True occurrences and 156 False Positives, it also made 156 False Negatives.

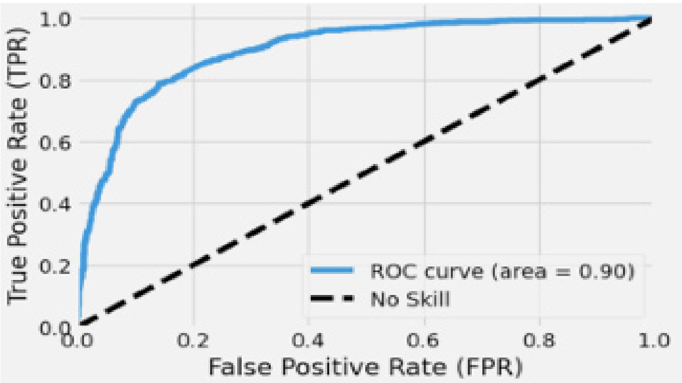

Fig. 7. ROC-AUC Curve of Gradient Boost Classifier

Figure 7 depicts a ROC curve that measures how well a binary classification model performs. The TPR is plotted against the FPR to visualise the trade-off between sensitivity and specificity. The ROC curve (blue line) shows strong predictive performance with an AUC of 0.90, indicating a high ability to distinguish between positive and negative classes. The dashed black line represents the "No Skill" baseline with random guessing (AUC = 0.5).

Table IV. Comparison between ML models on the Amazon product review dataset

Models	Accuracy	Precision	Recall	F1-score
GB	82	82	82	82
LR	80.86	81.16	1.06	81.06
NB	79.66	79.66	9.66	79.66
RNNMS	75	75	75	74

Th Table IV illustrate the performance of ML models across performance parameters. Gradient Boost has the best accuracy, along with very good precision, recall, and F1-score values (82%), when compared to all other models. Nearly as good as logistic regression is its accuracy (80.86%), which is accompanied by similarly excellent precision (81.16%), recall (81.06%), and F1-score (81.06%). Accuracy, precision, recall, and F1 scores for Naïve Bayes are around 79.66%, indicating somewhat worse performance. RNNMS, has the lowest overall performance, achieving75% accuracy and relatively lower precision, recall, and F1-scores (75%, 75%, 74%), making it less effective compared to the other models for this task. Gradient Boost delivers the best overall performance for Amazon product reviews.

CONCLUSION AND FUTURE WORK

Research in the area known as sentiment analysis or opinion mining seeks to understand how people feel about certain topics. The classification of sentiment polarity is a crucial issue in sentiment analysis, which is addressed in this working paper. This research utilises data derived from online product reviews found on Amazon.com. Positive or negative feelings conveyed by Amazon product reviews written by actual customers. Within the realm of ML models, sentiment analysis was used in this work. Ultimately, when compared to other models like LR, NB, and RNNMS, the Gradient Boost classifier outperforms them all according to accuracy, precision, recall, and score, which amounts to 82%. Despite its strong performance, the model's effectiveness could be limited by the quality and preprocessing of the dataset, particularly in handling noisy or ambiguous reviews. Future work could focus on improving model performance by incorporating advanced techniques like deep learning models or ensemble methods, experimenting with additional features for better feature extraction, and expanding the dataset to enhance generalization across different product categories and sentiments.

REFERENCES

1. A. Ejaz, Z. Turabee, M. Rahim, and S. Khoja, "Opinion mining approaches on Amazon product reviews: A comparative study," in *2017 International Conference on Information and Communication Technologies, ICICT 2017*, 2017. doi: 10.1109/ICICT.2017.8320185.
2. V. V. Kumar, S. R. Yadav, F. W. Liou, and S. N. Balakrishnan, "A digital interface for the part designers and the fixture designers for a reconfigurable assembly system," *Math. Probl. Eng.*, 2013, doi: 10.1155/2013/943702.

3. V. Kumar, V. V. Kumar, N. Mishra, F. T. S. Chan, and B. Gnanasekar, "Warranty failure analysis in service supply Chain a multi-agent framework," in *SCMIS 2010 - Proceedings of 2010 8th International Conference on Supply Chain Management and Information Systems: Logistics Systems and Engineering*, 2010.
4. X. Fang and J. Zhan, "Sentiment analysis using product review data," *J. Big Data*, 2015, doi: 10.1186/s40537-015-0015-2.
5. S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data," in *Proceedings of the International Conference on Inventive Research in Computing Applications, ICIRCA 2018*, 2018. doi: 10.1109/ICIRCA.2018.8597286.
6. S. Paknejad and A. Gravano, "Sentiment classification on Amazon reviews using machine learning approaches," *Degree Proj. Technol.*, 2018.
7. M. Z. Hasan, R. Fink, M. R. Suyambu, and M. K. Baskaran, "Assessment and improvement of intelligent controllers for elevator energy efficiency," in *IEEE International Conference on Electro Information Technology*, 2012. doi: 10.1109/EIT.2012.6220727.
8. T. Shaikh and D. Deshpande, "Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews," *Int. J. Comput. Trends Technol.*, 2016, doi: 10.14445/22312803/ijctt-v36p139.
9. V. V. Kumar, M. K. Pandey, M. K. Tiwari, and D. Ben-Arieh, "Simultaneous optimisation of parts and operations sequences in SSMS: A chaos embedded Taguchi particle swarm optimization approach," *J. Intell. Manuf.*, 2010, doi: 10.1007/s10845-008-0175-4.
10. Z. Singla, S. Randhawa, and S. Jain, "Sentiment analysis of customer product reviews using machine learning," in *Proceedings of 2017 International Conference on Intelligent Computing and Control, I2C2 2017*, 2018. doi: 10.1109/I2C2.2017.8321910.
11. C. P. B. . F. K. . K. B. Rain, "Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning," *Computational Linguist. Intell. Text Process.*, vol. 4, no. 1, pp. 475–487, 2013.
12. A. Bali, P. Agarwal, G. Poddar, D. Harsole, and N. M., "Consumer's Sentiment Analysis of Popular Phone Brands and Operating System Preference," *Int. J. Comput. Appl.*, 2016, doi: 10.5120/ijca2016912288.
13. T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," in *2018 IEEE International Conference on Innovative Research and Development, ICIRD 2018*, 2018. doi: 10.1109/ICIRD.2018.8376299.
14. S. Wladislav, Z. Johannes, W. Christian, K. Andre, and F. Madjid, "Sentilyzer: Aspect-oriented sentiment analysis of product reviews," in *Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018*, 2018. doi: 10.1109/CSCI46756.2018.00059.
15. Z. Singla, S. Randhawa, and S. Jain, "Statistical and sentiment analysis of consumer product reviews," in *8th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2017*, 2017. doi: 10.1109/ICCCNT.2017.8203960.
16. V. Kolluri, "a Pioneering Approach To Forensic Insights: Utilization Ai for Cybersecurity Incident Investigations," *Int. J. Res. Anal. Rev.*, vol. 3, no. 3, 2016.
17. M. Z. Hasan, R. Fink, M. R. Suyambu, M. K. Baskaran, D. James, and J. Gamboa, "Performance evaluation of energy efficient intelligent elevator controllers," in *IEEE International Conference on Electro Information Technology*, 2015. doi: 10.1109/EIT.2015.7293320.
18. V. V Kumar, "An interactive product development model in remanufacturing environment : a chaos-based artificial bee colony approach," *Eng. Int.*, vol. 6, no. 2, pp. 211–222, 2014.
19. V. KOLLURI, "A COMPREHENSIVE ANALYSIS ON EXPLAINABLE AND ETHICAL MACHINE: DEMYSTIFYING ADVANCES IN ARTIFICIAL INTELLIGENCE," *Int. Res. J.*, vol. 2, no. 7, 2015.
20. V. V. Kumar, F. W. Liou, S. N. Balakrishnan, and V. Kumar, "Economical impact of RFID implementation in remanufacturing: a Chaos-based Interactive Artificial Bee Colony approach," *J. Intell. Manuf.*, 2015, doi: 10.1007/s10845-013-0836-9.
21. Y. Al Amrani, M. Lazaar, and K. E. El Kadirp, "Random forest and support vector machine based hybrid approach to sentiment analysis," in *Procedia Computer Science*, 2018. doi: 10.1016/j.procs.2018.01.150.
22. V. V. Kumar and F. T. S. Chan, "A superiority search and optimisation algorithm to solve RFID and an environmental factor embedded closed loop logistics model," *Int. J. Prod. Res.*, 2011, doi: 10.1080/00207543.2010.503201.
23. E.-S. M. E.-H. and A.-B. M. S. Donia Gamal , Marco Alfonse, "Analysis of Machine Learning Algorithms for Opinion Mining in Different Domains," *MDPI*, 2018.
24. J. Wu and T. Ji, "2016_Deep Learning for Amazon Food Review Sentiment Analysis," pp. 1–8, 2016.
25. Gagan Kumar Patra, Shravan Kumar Rajaram, & Venkata Nagesh Boddapati. (2019). Ai And Big Data In Digital Payments: A Comprehensive Model For Secure Biometric Authentication. *Educational Administration: Theory and Practice*, *25*(4), 773–781. https://doi.org/10.53555/kuey.v25i4.7591
26. Chandrababu Kuraku, Hemanth Kumar Gollangi, & Janardhana Rao Sunkara. (2020). Biometric Authentication In Digital Payments: Utilizing AI And Big Data For Real-Time Security And Efficiency. *Educational Administration: Theory and Practice*, *26*(4), 954–964. https://doi.org/10.53555/kuey.v26i4.7590

27. Galla, E. P., Kuraku, C., Gollangi, H. K., Sunkara, J. R., & Madhavaram, C. R. AI-DRIVEN DATA ENGINEERING.
28. Rajaram, S. K. (2022). AI-Driven Threat Detection: Leveraging Big Data For Advanced Cybersecurity Compliance. *Educational Administration: Theory and Practice*, *28*(4), 285-296.
29. Bauskar, S. (2020). View of Unveiling the Hidden Patterns AI-Driven Innovations in Image Processing and Acoustic Signal Detection. *Journal of Recent Trends in Computer Science and Engineering*, *8*(1), 10-70589.
30. Eswar Prasad Galla.et.al. (2021). Big Data And AI Innovations In Biometric Authentication For Secure Digital Transactions Educational Administration: Theory and Practice, 27(4), 1228 –1236Doi: 10.53555/kuey.v27i4.7592
31. Janardhana Rao Sunkara, Sanjay Ramdas Bauskar, Chandrakanth Rao Madhavaram, Eswar Prasad Galla, Hemanth Kumar Gollangi, Data-Driven Management: The Impact of Visualization Tools on Business Performance, International Journal of Management (IJM), 12(3), 2021, pp. 1290-1298. https://iaeme.com/Home/issue/IJM?Volume=12&Issue=3.
32. Mohit Surender Reddy, Manikanth Sarisa, Siddharth Konkimalla, Sanjay Ramdas Bauskar, Hemanth Kumar Gollangi, Eswar Prasad Galla, Shravan Kumar Rajaram, 2021. "Predicting tomorrow's Ailments: How AI/ML Is Transforming Disease Forecasting", ESP Journal of Engineering & Technology Advancements, 1(2): 188-200.
33. K. Gollangi, S. R. Bauskar, C. R. Madhavaram, P. Galla, J. R. Sunkara, and M. S. Reddy, "Echoes In Pixels : The Intersection Of Image Processing And Sound Open Access Echoes In Pixels : The Intersection Of Image Processing And Sound Detection," Int. J. Dev. Res., vol. 10, no. 08, pp. 39735–39743, 2020, doi: 10.37118/ijdr.28839.28.2020.
34. Gollangi, H. K., Bauskar, S. R., Madhavaram, C. R., Galla, E. P., Sunkara, J. R., & Reddy, M. S. (2020).Unveiling the Hidden Patterns: AI-Driven Innovations in Image Processing and Acoustic Signal Detection. (2020). Journal Of Recent Trends In Computer Science And Engineering (JRTCSE), 8(1), 25- 45. https://doi.org/10.70589/JRTCSE.2020.1.3.
35. Gollangi, H. K., Bauskar, S. R., Madhavaram, C. R., Galla, E. P., Sunkara, J. R., & Reddy, M. S. (2020). Exploring AI Algorithms for Cancer Classification and Prediction Using Electronic Health Records. Journal of Artificial Intelligence and Big Data, 1(1), 65–74. Retrieved from https://www.scipublications.com/journal/index.php/jaibd/article/view/1109
36. Chandrakanth R. M., Eswar P. G., Mohit S. R., Manikanth S., Venkata N. B., & Siddharth K. (2021). Predicting Diabetes Mellitus in Healthcare: A Comparative Analysis of Machine Learning Algorithms on Big Dataset. In Global Journal of Research in Engineering & Computer Sciences (Vol. 1, Number 1, pp. 1–11). https://doi.org/10.5281/zenodo.14010835
37. Krutthika, H. K. (2019). Modelling of data delivery modes of next-generation SOC-NOC router. *2019 IEEE Global Conference for Advancement in Technology (GCAT)*. Bangalore, India. https://doi.org/10.1109/GCAT47503.2019.8978290.
38. Pavitha US, Nikhila S, Krutthika HK. design and implementation of image dithering engine on a spartan 3AN FPGA. Intern J Future Compt Comm. 2012;1(4):361.
39. S Nikhila, U. S. Pavitha and H. K. Krutthika, "Face recognition using wavelet transforms", International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering, vol. 3, no. 1, pp. 6740-6746, 2014.
40. H. K. Krutthika and Rajashekhara, "Modeling of Data Delivery Modes of Next Generation SOC-NOC Router," 2019 Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2019, pp. 1-6, doi: 10.1109/GCAT47503.2019.8978290.
41. Kalla, D., Kuraku, D. S., & Samaah, F. (2021). Enhancing cyber security by predicting malwares using supervised machine learning models. *International Journal of Computing and Artificial Intelligence*, *2*(2), 55-62.
42. Kalla, D., Smith, N., Samaah, F., & Polimetla, K. (2021). Facial Emotion and Sentiment Detection Using Convolutional Neural Network. *Indian Journal of Artificial Intelligence Research (INDJAIR)*, *1*(1), 1-13.
43. Kuraku, S., & Kalla, D. (2020). Emotet malware—a banking credentials stealer. *Iosr J. Comput. Eng*, *22*, 31-41.
44. Kalla, D., & Samiuddin, V. (2020). Chatbot for medical treatment using NLTK Lib. *IOSR J. Comput. Eng*, *22*, 12.
45. Bauskar, S., Boddapati, V. N., Galla, E. P., Bauskar, S. R., Patra, G. K., Kuraku, C., & Madhavram, C. (2021). Harnessing the Power of Big Data: The Evolution of AI and Machine Learning in Modern Times. *Available at SSRN 4988863*.

Citation: Chethan Sriharsha Moore, Purna Chandra Rao Chinta, et al., "Leveraging Big Data Analytics and Machine Learning Techniques for Sentiment Analysis of Amazon Product Reviews in Business Insights", Universal Library of Engineering Technology, 2022; 10-17. DOI: https://doi.org/10.70315/uloap.ulete.2022.002.