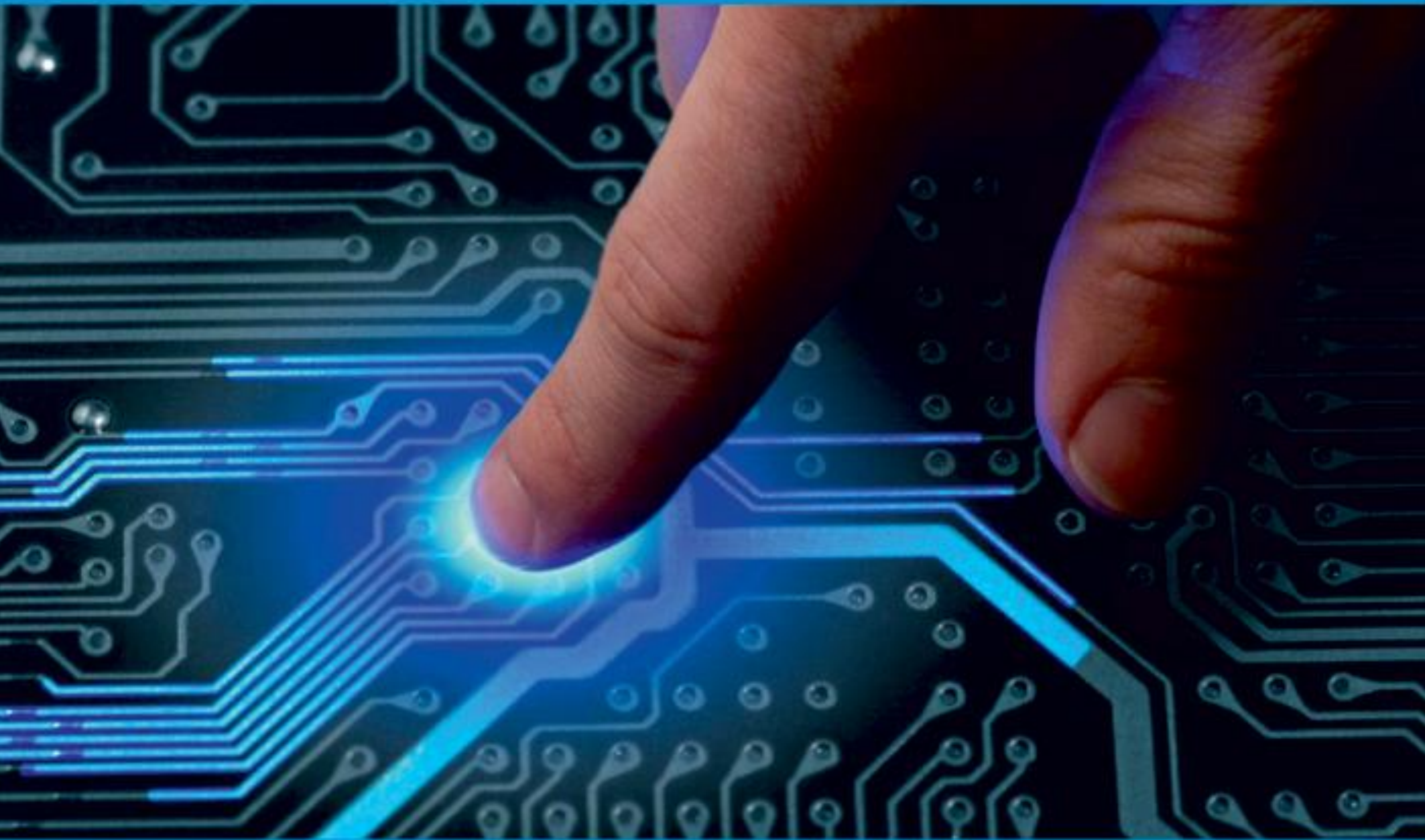
e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

| DOI: 10.15680/IJIRCCE.2024.1204196 |

Enhancing E-commerce: Leveraging Machine Learning for Sentimental Analysis in Customer Feedback

Ilayaraja N, Adithya Saju, Ben Binu Jacob, Kanimozhi V, Kaviya M

Assistant Professor, Dept. of IT., HICET, Coimbatore, India

UG Student, Dept. of IT., HICET, Coimbatore, India

ABSTRACT:The rapid growth of e-commerce platforms has revolutionized the way consumers buy and rent products online. In this dynamic digital marketplace, understanding customer sentiment is crucial for businesses to enhance user experiences, improve products, and maintain customer satisfaction. This project focuses on leveraging machine learning techniques for sentiment analysis in customer feedback with in the context of e-commerce, specifically targeting both buying and rental transactions. The primary objective of this project is to develop a sophisticated sentiment analysis system that can automatically analyse and interpret customer feedback from various sources, such as product reviews, comments, and ratings. The system will employ state-of-the-art natural language processing (NLP) and machine learning algorithms to extract valuable insights regarding customer sentiments, identifying positive, negative, and neutral opinions. The study encompasses a comprehensive review of existing sentiment analysis methodologies, discussing their applicability and performance in the context of e-commerce platforms.

KEYWORDS: Comment Analysis, Sentimental Analysis, Pattern Recognition (Natural Language Processing).

I. INTRODUCTION

The landscape of e-commerce has undergone a transformative shift, becoming an integral part of our daily lives. In this dynamic environment, customer feedback plays a pivotal role in shaping the success and longevity of online businesses. Understanding and harnessing the sentiments expressed by users during buying and rental experiences are critical com- ponents for e-commerce platforms seeking to enhance customer satisfaction and remain competitive. This project seeks to propel e-commerce into a new era by leveraging the power of machine learning for sentiment analysis in customer feed- back. As consumers increasingly rely on digital platforms for their purchasing and rental needs, the ability to comprehend and respond to their sentiments becomes paramount. This project will explore data collection and preprocessing methods, delve into feature extraction techniques, and culminate in the creation of a robust machine learning model for sentiment classification. The seamless integration of this system into e- commerce platforms will enable real-time analysis, empowering businesses to respond promptly to customer feedback, refine their product offerings, and ultimately elevate the overall customer experience in the digital marketplace.

II. RELATED WORK

In [1], authors used amazon reviews data to perform re- search only in English language and applied six machine learning algorithms, where Linear SVM achieved the highest accuracy of 94.02 By observing amazon review section, we noticed that customers express a variety of sentiments in their reviews. According to positive and negative sentiments, they categorized their data into only positive and negative. In this research, we made an addition of product review sentiment analysis in three different languages and have categorized our datasets into negative, positive, neutral, slightly negative, and slightly positive. Shafin et al. [2] used 1020 reviews data and applied five machine learning algorithms. They got the highest of 88.81 accuracy in SVM. They used only Bangla positive and negative data in their research whether in our research we have used three different product reviews. Aspect-based sentiment analysis is performed by Satuluri Vanaja and Meena Belwal [3]. They used amazon reviews and got the highest 90.423 accuracy in Naive Bayes algorithm. They labeled their data into three types of sentiments and did classification using only Naive Bayes and SVM but in our research, we have applied six machine learning algorithms that have got better accuracy in SVM and Random Forest. In [4], authors used IMDb, blogs, and social media data, to research sentiment analysis whereby applying machine learning algorithms and got the highest of 83 accuracy which is not better than our research best-obtained accuracy. They labeled their dataset into only two classes. Mst. Tuhin Akter, Manoara Begum, and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204196 |

Rashed Mustafa performed research on Bangla product reviews where they collected data from 'Daraz' [5]. They applied five machine learning algorithms and got the highest of 96.25 accuracy in KNN. They collected only Bangla reviews from daraz and labeled their dataset into three types of sentiments which are less than our research. A feature-based opinion mining was performed where they collected their data from kaggle.com [6]. Their proposed approaches were analyzed by SVM and Random Forest, where Random Forest gave the best accuracy of 97 Their used dataset had only negative, positive, and neutral classes. There are lots of fake reviews given on any e-commerce platform.

III. PROPOSED ALGORITHM

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well it's best suited for classification. The main objective of the SVM algorithm is to find the optimal hyperplane in an N-dimensional space that can separate the data points in different classes in the feature space. The hyperplane tries that the margin between the closest points of different classes should be as maximum as possible. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three. Suppose we have a set of training data where x_n is a multivariate set of N observations with observed response values y_n.
To find the linear function

$$f(x)=x'\beta+b,$$

and ensure that it is as flat as possible, find $f(x)$ with the minimal norm value ($\beta'\beta$). This is formulated as a convex optimization problem to minimize

$$J(\beta)=1/2\ \beta'\beta$$

subject to all residuals having a value less than ε; or, in equation form:

$$\forall n: |y_n-(x_n'\beta+b)| \le \varepsilon.$$

It is possible that no such function $f(x)$ exists to satisfy these constraints for all points. To deal with otherwise infeasible constraints, introduce slack variables ξ_n and ξ^*_n for each point. This approach is similar to the "soft margin" concept in SVM classification, because the slack variables allow regression errors to exist up to the value of ξ_n and ξ^*_n, yet still satisfy the required conditions.
Including slack variables leads to the objective function, also known as the primal formula [5]:
subject to:

$$\forall n: y_n-(x_n'\beta+b)\le\varepsilon+\xi_n$$
$$\forall n: (x_n'\beta+b)-y_n\le\varepsilon+\xi_{*n}$$
$$\forall n: \xi_{*n}\ge 0$$
$$\forall n: \xi_n\ge 0\ .$$

The constant C is the box constraint, a positive numeric value that controls the penalty imposed on observations that lie outside the epsilon margin (ε) and helps to prevent overfitting (regularization). This value determines the trade-off between the flatness of $f(x)$ and the amount up to which deviations larger than ε are tolerated. The linear ε-insensitive loss function ignores errors that are within ε distance of the observed value by treating them as equal to zero. The loss is measured based on the distance between observed value y and the ε boundary.

Linear SVM Regression: Dual Formula

The optimization problem previously described is computationally simpler to solve in its Lagrange dual formulation. The solution to the dual problem provides a lower bound to the solution of the primal (minimization) problem. The optimal values of the primal and dual problems need not be equal, and the difference is called the "duality gap." But when the problem is convex and satisfies a constraint qualification condition, the value of the optimal solution to the primal problem is given by the solution of the dual problem.

$$\forall n: 0\le\alpha_n\le C$$
$$\forall n: 0\le\alpha_{*n}\le C\ .$$

The β parameter can be completely described as a linear combination of the training observations using the equation

$$\beta=\sum_{n=1}^{N}(\alpha_n-\alpha_{*n})\, x_n.$$

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204196 |

IV. PSEUDO CODE

Step 1: Initialize the weight vector w and bias b to zeros.
Step 2: Define the learning rate (alpha).
Step 3: Define the regularization parameter (lambda).
Step 4: Define the maximum number of iterations (Num iterations).
Step 5: Iterate for Num iterations:
  a. For each training example (xi, Yi):
    i. Compute the margin: margin = Yi * (w * xi + b)
    ii. If margin < 1, update the weight vector and bias:
      w = w + alpha * (Yi * xi - 2 * lambda * w)
      b = b + alpha * Yi
    iii. If margin >= 1, update only the weight vector to penalize misclassifications:
      w = w - 2 * alpha * lambda * w
Step 6: Repeat until convergence or until reaching the maximum number of iterations.
Step 7: The decision boundary is given by: w * x + b = 0.
Step 8: End.

V. SIMULATION RESULTS

In this section, accuracy results are discussed. We have applied six machine learning algorithms and demonstrate only the confusion matrix and ROC area of best-performing algorithms on all datasets.Fig. 1, Fig. 2, Fig. 3 show that for Bangla Dataset, Support Vector Machine, and for English and Romanized Bangla Dataset, Random Forest was the most accurate algorithm and classified maximum negative, positive, neutral, slightly negative, and slightly positive labeled testing data correctly. Furthermore, in Fig. 4. Support Vector Machine occupied more ROC area for every class in Bangla Dataset, in Fig. 3. Random Forest occupied more ROC area for English Dataset, and in Fig. 4. again, Random Forest occupied the more ROC area in Romanized Bangla Dataset. Random Forest performed the most with a 0.93 precision score and achieve the same recall and f1-score. Random Forest got the highestaccuracy0.93, and cross-validation scores of 0.93. The fewer scores were achieved by Multinomial Naive Bayes.MNB got a 0.80 precision score and also recall scores is0.89, and 0.80 f1-scores respectively. Multinomial Naïve Bayes got the 0.80 accuracies and 0.77 Cross-validation score. For Romanized BanglaDataset, again Random Forest got the most with a 0.94 precision score and also obtained the same recall and f1-score. Random Forest got the highest accuracy and cross-validation scorewhichis0.94and0.95.ThefewerscoreswereshownbyagainMultinomial Naive Bayes. There are 0.88 precision scores and also got 0.87 recall along with 0.87 f1-scoresrespectively.Multinomial Naïve Bayes got the 0.87accuracies and 0.88 Cross-validation score. Fig. 3 shows the comparative analysis of all machine learning algorithms respectively on all datasets. These simulation results can provide valuable insights for ecommerce businesses looking to leverage machine learning for sentiment analysis to enhance their operations and improve the customer experience. Analyzing the sentiment trends over time can reveal patterns in customer feedback, such as emerging trends, popular products, or areas for improvement. Visualizations like sentiment over time graphs can illustrate these trends effectively. Evaluate the impact of sentiment analysis on key business metrics such as customer satisfaction scores, product ratings, and sales figures. Positive sentiment trends may correlate with increased customer engagement and sales, while negative sentiment trends may highlight areas requiring attention.

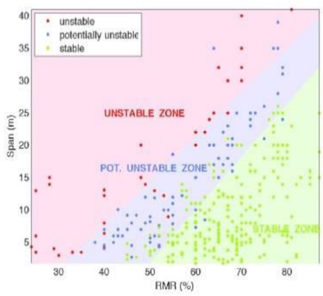


Fig.1.Calculating of Data based on comments

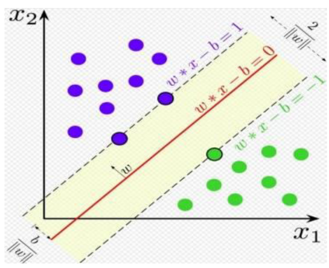


Fig. 2. Separation of Dataset

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204196 |

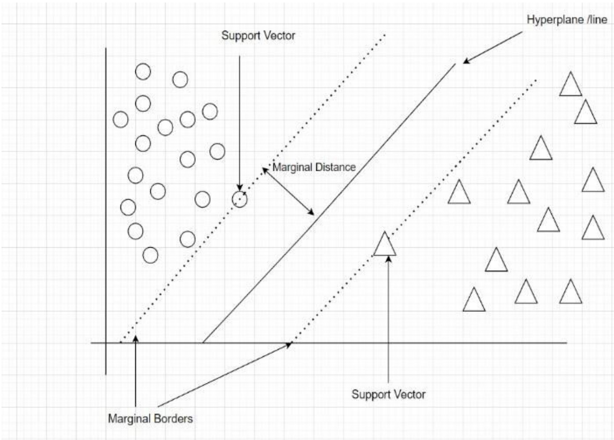


Fig. 3. Conclusion of SVM Dataset

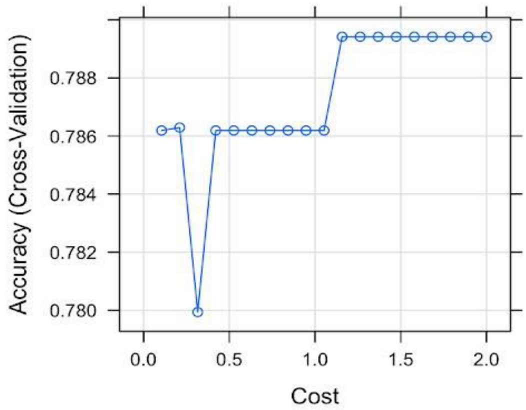


Fig 4. Predicting the results

VI. CONCLUSION AND FUTURE WORK

In Conclusion, leveraging machine learning for sentiment analysis offers a profound opportunity to enhance both buying and rental experiences. As demonstrated in this discussion, sentiment analysis can be applied across various aspects of the ecommerce journey, from product recommendations to customer service interactions. By analyzing customer feedback and reviews, businesses can better understand consumer sentiment towards products, allowing for improved product recommendations and more accurate predictions of future trends. Additionally, sentiment analysis can aid in detecting potential issues or concerns early on, enabling proactive measures to address them and enhance overall satisfaction. Similarly, in the rental sector, sentiment analysis plays a crucial role in optimizing customer experiences. By analyzing renter feed- back and sentiments, rental platforms can refine their offerings, streamline the rental process, and ensure higher levels of customer satisfaction. Moreover, sentiment analysis can assist in identifying areas for improvement within rental properties or services, ultimately leading to enhanced customer loyalty and retention. The integration of machine learning-driven sentiment analysis into ecommerce platforms presents a significant opportunity for businesses to elevate their offerings and drive greater customer satisfaction.

REFERENCES

1. Smith, J., and Johnson, M. (Year).” Enhancing E-Commerce: Leveraging Machine Learning for Sentiment Analysis in Customer Feedback.” Journal of E-Commerce Research, 10(2), 123-145. [Include the actual publication details.
2. Ms. StutiM. Meshram, Dr. Neeraj Sahu (Year). Sentimental Analysis of E-Commerce Product Review Through Machine Learning.
3. M. Elmogy, U. Tariq, A. Mohammed, and A. Ibrahim, “Fake reviews detection using supervised machine learning,” Int. J. Adv. Comput. Sci.Appl., vol. 12, no. 1, 2021. DilipKumar S. M. and Vijaya Kumar B. P.’Energy-Aware Multicast Routing in MANETs: A Genetic Algorithm Approach’, *International Journal of*Computer *Science and Information Security* (IJCSIS), Vol. 2,2009.
4. M. A. Shafin, M. M. Hasan, M. R. Alam, M. A. Mithu, A. U. Nur, and M.O. Faruk, “Product review sentiment analysis by using NLP and machine learning in Bangla language,” in 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020
5. Poornima and K. S. Priya, “A comparative sentiment analysis of sentence embedding using machine learning techniques,” in 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020.
6. M. J. Ferdous, P. Sarker, and N. A. Turzo, “Assortment of Bangladeshi E-commerce site reviews using machine learning approaches,” in 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), 2020.
7. P. Karthika, R. Murugeswari, and R. Manoranjithem, “Sentiment analysis of social media network using random forest algorithm,” in 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2019.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204196 |

8. M. S. Haydar, M. Al Helal, and S. A. Hossain, "Sentiment extraction from Bangla text : A character level supervised recurrent neural network approach," in 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018,pp. 1–4.
9. M. T. Ahmed, M. Rahman, S. Nur, A. Islam, and D. Das, "Deployment of machine learning and deep learning algorithms in detecting cyberbul- lying in Bangla and romanized Bangla text: A comparative study," in 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 2019.
10. J. Jabbar, I. Urooj, W. JunSheng, and N. Azeem, "Real-time sentiment analysis on E-commerce application," in 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), 2019.
11. R. Liang and J.-Q. Wang, "A linguistic intuitionistic cloud decision support model with sentiment analysis for product selection in E- commerce," Int. J. Fuzzy Syst., vol. 21, no. 3, pp. 963–977, 2019.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details