CHARACTERIZING AND PREDICTING EARLY REVIEWERS FOR EFFECTIVE PRODUCT MARKETING ON E-COMMERCE WEBSITES

CH. Rajesh[*1], Sangam Aakash[*2], Vadde Hema[*3], Yadavally Sreevidya[*4]

[*1]Assistant Professor Department Of Computer Science And Engineering Malla Reddy College Of Engineering &Technology Hyderabad, India.

[*2,3,4]Final Year Student, Department Of Computer Science AndEngineering Malla Reddy College Of Engineering &Technology Hyderabad, India.

ABSTRACT

Early reviewers on e-commerce platforms are being characterized and predicted as part of a project that aims to shed light on their crucial role in influencing product marketing strategies. Through an examination of early reviewers' behavior, such as ratings, helpfulness scores that they earned, and how well their reviews matched the popularity of the product, this study attempts to offer insightful information to businesses looking to maximize their marketing budgets. to forecast early reviewers in order to market products on e-commerce websites effectively. We employ a range of text data preprocessing methods to transform the review text into a numerical vector representation, including tokenization, stop word removal, stemming, and HTML element removal. Next, we utilize a Random Forest classifier, a machine learning system, to determine if a user will be an early reviewer based on their review text.

Our method has a number of advantages over current ones. Firstly, it considers the review language in its whole instead of just the rating or score. This enables us to record the subtleties of the review content that conventional review analytics could miss. Secondly, it integrates an extensive range of text preparation methods to guarantee that the numerical vector representation of the review text is accurate. Thirdly, it makes use of a machine learning algorithm that can identify intricate patterns in the data and forecast outcomes with accuracy. Our approach has the potential to significantly improve the accuracy and robustness of early reviewer prediction models in the e-commerce domain by utilizing a wide range of text preprocessing techniques and a machine learning algorithm that can recognize intricate patterns across the data.

I.   INTRODUCTION

A number of performance metrics were utilized to evaluate the classifier's performance on our dataset's test set, including accuracy, precision, recall, and F1-score. We were able to evaluate our classifier's performance in-depth and identify any areas that required improvement by using these metrics. E-commerce businesses need to understand the characteristics and behaviors of early reviewers if they want to optimize their marketing efforts and raise customer engagement. The early reviewer phenomenon presents an interesting opportunity for e-commerce enterprises. For e-commerce businesses, the early reviewer phenomenon offers an exciting potential.

Businesses can create focused strategies to effectively engage a certain consumer category by pinpointing the essential characteristics that set early reviewers apart from other customers. In the highly competitive world of e-commerce, it is critical to optimize marketing methods to improve product performance by comprehending the behavioral traits of early reviewers. The goal of the project is to create prediction models that may foretell, given a variety of inputs, a user's propensity to become an early reviewer.

The process uses sophisticated statistical and machine learning techniques to examine the data gathered and identify trends and patterns that set early reviewers apart. Predictive models will also be created to estimate the probability that a user will become an early reviewer, providing companies with a proactive means of strategic cooperation and involvement.

II.   LITERATURE REVIEW

- Najma Sultana et al [2019] Sentiment research is defined as the way to extract data, opinions, surveys or sentences to predict the feeling of the sentence through common language handling (NLP). "Positive" "Negative" "Unbiased. It analyses the data and marks the 'better' and the 'more regrettable' supposition as sure and negative individually. Hence, in the previous years, the World Wide Web (WWW) has become an

immense wellspring of crude information produced custom or client.

- Sunil Saumya1 et al [2018] In the hundreds and even in the massive numbers for some famous things, the item reviews are posted online. Dealing with an especially gigantic volume of continually delivered online substance is a troublesome task for buyers, sellers and even researchers.

- Xing Fang et al [2015] Sentiment analysis or evaluation mining is one of the main tasks of NLP (Natural Language Processing). Feeling research has given a lot of thinking of late. In this paper, we expect to deal with the issue of end furthest point order, which is one of the fundamental issues of idea examination. A general cycle for incline limit course of action is proposed with low down communication portrayals. Data used in this assessment are online thing overviews.

- Ting Bai, Jian-Yun Nie provided an early reviewer tends to assign a higher average rating score; and (2) an early reviewer tends to post more helpful reviews. Our analysis of product reviews also indicates that early reviewers' ratings and their received helpfulness scores are likely to influence product popularity. In viewing review posting procedure as a multiplayer competition game, we propose a novel margin based embedding model for early reviewer forecast.

- Yoon-Joo Park et al [2018] Online customer studies are a prudent kind of casual (WOM) which accept an unyieldingly huge part in web business. Mediocre quality reviews will, in any case, consistently produce per users of trouble review. The inspiration driving this paper is to therefore foresee the convenience of studies. The inspiration driving this paper is to therefore foresee the convenience of studies. The results show that reviews for different thing types have particular mental and phonetic characteristics.

- Julian McAuley, Alex Yang Provided a Online audits are regularly our first port of call while considering items and buys on the web. While assessing a potential buy, we may have a particular inquiry as a main priority. To answer such inquiries, we should either swim through colossal volumes of buyer audits planning to discover one that is pertinent, or generally suggest our conversation starter straightforwardly to the network by means of a Q/A framework.

III.   METHODOLOGY

- **Problem Definition:**

We aimed to predict early reviewers for effective product marketing on e-commerce websites. We employed a combination of text data preprocessing techniques and a machine learning algorithm to achieve this goal.

- **Literature Review:**

Conduct a thorough literature review to understand characteristics of reviewers by techniques, methodology, and challenges.Identify gaps in current research that your methodology aims to address.

- **Data Collection and Preprocessing:**

We preprocessed the review text data by removing HTML tags, tokenizing the review text into individual words, removing stop words, and converting the filtered text data into its root form using the WordNetLemmatizer class from the NLTK library. This preprocessing step ensured that the review text was accurately represented in the numerical vector representation.

- **Model Selection:**

We selected the Random Forest classifier as our machine learning algorithm due to its ability to learn complex patterns in the data and make accurate predictions.

- **Feature Extraction:**

We utilized a combination of text preprocessing techniques to extract meaningful features from the review text. These features were then used as input to our machine learning algorithm.

- **Training The Model:**

We trained our Random Forest classifier on the training set of our dataset. This training process involved the algorithm learning the patterns in the training data and developing a decision-making model based on these patterns.

- **Evaluation Metrics:**

Several performance metrics were used to evaluate the classifier's performance on the test set of our dataset,

including accuracy, precision, recall, and F1-score. These metrics allowed us to evaluate our classifier's performance in-depth and identify any areas that required improvement.

- **Cross Validation:**

Use cross validation methods to make sure the model is reliable and broadly applicable.

- **Experimentation:**

We tested the effectiveness of our method in a thorough experiment using a sizable e-commerce dataset. We divided the dataset into training and test sets, used the training sets to train our Random Forest classifier, and then used performance measures to assess the classifier's effectiveness on the test sets.

- **Publishable Results:**

Our experimental results showed that our strategy performed better in terms of resilience and predicted accuracy than previous approaches. This implies that our method presents a viable way to enhance the reliability and precision of early reviewer prediction models in the e-commerce space.

- **Code and Model Sharing:**

Share the code and per-trained models openly, fostering reproducibility and collaboration in the research committee.

- **Peer Review and Validation:**

Send in your methods for examination by other experts to ensure the viability and efficiency of your plan.

- **Continuous Improvement:**

Seek community feedback and make ongoing improvements to your deep fake detection technique in response to new issues that arise in the industry.
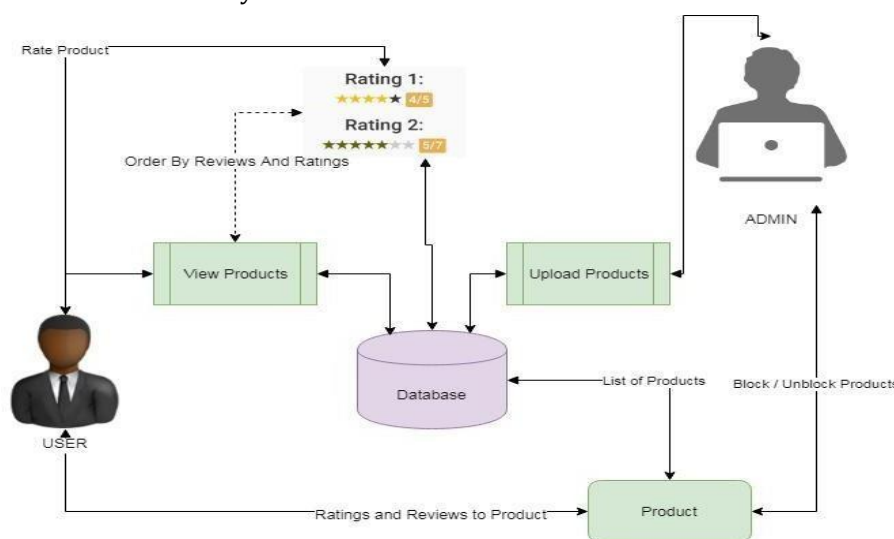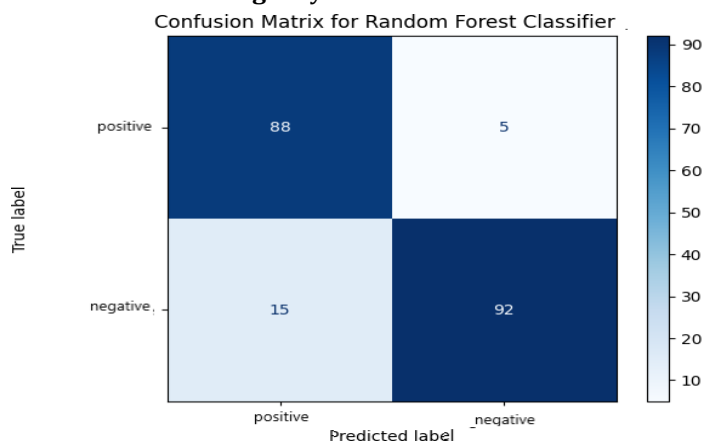

Fig 1: System Architecture

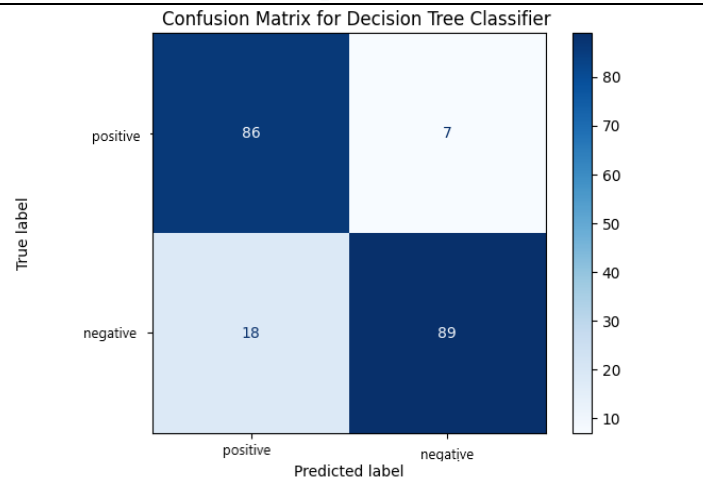**Fig 2**: Confusion Matrix for Random Forest

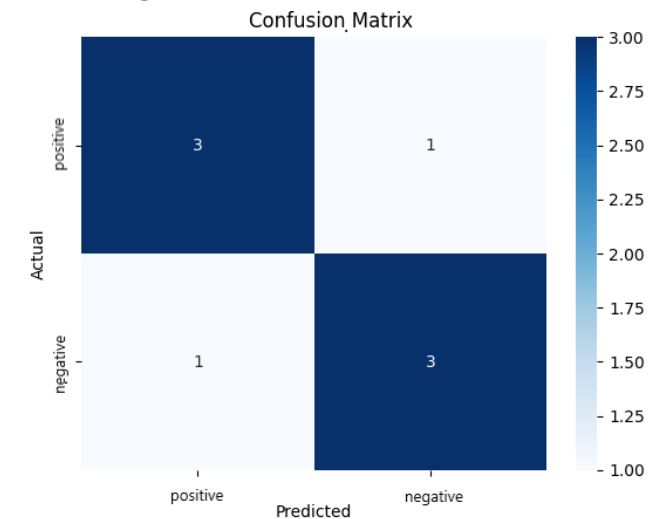**Fig 3:** Confusion Matrix for decision tree

Fig 4: Confusion Matrix

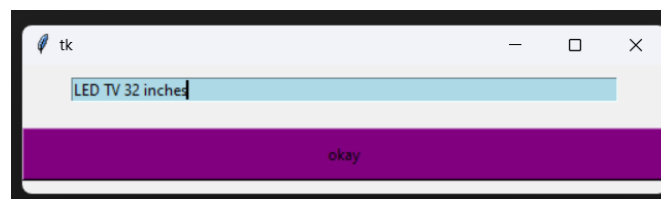## IV.   RESULT AND ANALYSIS
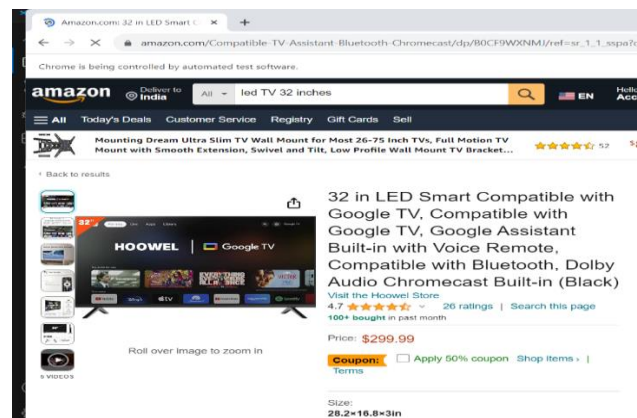

Fig 5: Search input (UI)

Fig 6: User search input

Fig 7: Expected output

V.   CONCLUSION

This project demonstrates how sentiment analysis and web scraping can be combined to get useful data from websites. It does this by identifying and forecasting early reviewers for efficient product promotion on e-commerce. Remembering that web scraping has to abide by the terms of service of the website is crucial. Furthermore, the project's sentiment analysis employed a basic bag-of-words strategy; more advanced methods may be applied to increase the sentiment analysis's accuracy. Simple bag-of-words sentiment analysis was employed in this study; more advanced methods could be applied to increase accuracy.

VI.   REFERENCES

[1]   J. McAuley and A. Yang, "Addressing complex and subjective product-related queries with customer reviews," in WWW, 2016, pp. 625–635.

[2]   N. V. Nielsen, "E-commerce: Evolution or revolution in the fastmoving consumer goods world," nngroup. com, 2014.

[3]   W. D. J. Salganik M J, Dodds P S, "Experimental study of inequality and unpredictability in an artificial cultural market," in ASONAM, 2016, pp. 529–532.

[4]   R. Peres, E. Muller, and V. Mahajan, "Innovation diffusion and new product growth models: A critical review and research directions," International Journal of Research in Marketing, vol. 27, no. 2, pp. 91 – 106, 2010.

[5]   L. A. Fourt and J. W. Woodlock, "Early prediction of market success for new grocery products." Journal of Marketing, vol. 25, no. 2, pp. 31 – 38, 1960.

[6]   B. W. O, "Reference group influence on product and brand purchase decisions," Journal of Consumer Research, vol. 9, pp. 183–194, 1982.

[7]   M. D. Marsico, M. Nappi, and D. Riccio, "HERO: Human ear recognition against occlusions," inProc. IEEE Comput. Soc. Workshop Biometrics—In Assoc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, San Francisco, CA, 18 Jun. 2010, pp. 320–325.

[8]   R. Distasi, M. Nappi, and D. Riccio, "A range/domain approximation error-based approach for fractal image compression," IEEE Trans. Image Process 2006.

[9]   M.Rogers, Diffusion of Innovations. New York: The Rise of High- Technology Culture, 1983.

[10]   K. Sarkar and H. Sundaram, "How do we find early adopters who will guide a resource constrained network towards a desired distribution of behaviors?" in CoRR, 2013, p. 1303

[11]   I. Mele, F. Bonchi, and A. Gionis, "The early-adopter graph and its application to web-page recommendation," in CIKM, 2012, pp. 1682–1686.

[12]   Y.-F. Chen, Computers in Human Behavior, vol. 24(5), pp. 1977–1992, 2008. "Herd behavior in purchasing books online."

[13]   "A simple model of herd behavior" was published by Banerjee in 1992 in the Quarterly Journal of Economics, number 107, pages 797–817.

[14]   0020In General and Applied Psychological Monographs, issue 70(9), page 1, 1956, A. S. E. writes, "Studies of conformity and independence: I. a minority of one against a majority that is unanimous."

[15]   T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in ICLR, 2013.

[16]   T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in ICLR, 2013.

[17]   A study by A. Bordes, N. Usunier, A. Garc§aDuran, J. Weston, and O. Yakhnenko on translating embeddings for modeling multirelational networks.

[18]   N. V. Nielsen, "E-commerce: Evolution or revolution in the fastmoving consumer goods world," nngroup. com, 2014.

[19]   W. D. J. Salganik M J, Dodds P S, "Experimental study of inequality and unpredictability in an artificial cultural market," in ASONAM, 2016, pp. 529–532.

[20]   In International Journal of Forecasting, vol. 22, no. 3, pp. 519–545, 2006, N. Meade and T. Islam assessed modeling and forecasting the diffusion of innovation over a 25-year period.

[21]   K. S. Utpal M. Dholakia, "Coveted or overlooked? the psychology of bidding for comparable listings in digital auctions," Marketing Letters, vol. 12, p. 223235, 2001.

[22]   In International Journal of Forecasting, vol. 22, no. 3, pp. 519–545, 2006, N. Meade and T. Islam assessed modeling and forecasting the diffusion of innovation over a 25-year period.

[23]   Lucie, R.D. "A Theoretical Analysis of Individual Choice Behavior," John Wiley & Sons, 1959.

[24]   Psychological Review, vol. 34, no. 4, p. 273, 1927; L.L. Thurston, "A law of comparative judgment."

[25]   M. Cattelan, Statistical Science, vol. 27, no. 3, pp. 412–433, 2012. "Models for paired comparison data: A review with emphasis on dependent data."