

Identification of Fraudulent Reviews

Mr. Adithyan P S¹, Ms. Akshaya V A², Mr. Bhuvanesh A³, Ms. Anitha R⁴

Students, Department of Computer Science and Engineering^{1,2,3}

Assistant Professor, Department of Computer Science and Engineering⁴

SRM Valliammai Engineering College, Chennai, Tamil Nadu, India

Abstract: This paper introduces a comprehensive system designed to bolster the trustworthiness of product reviews in e-commerce applications. Leveraging logistic regression, the system filters out fake reviews obtained through web scraping, providing users with an authentic product rating. The algorithm analyzes textual features to assign a probability score, effectively distinguishing genuine reviews from deceptive ones. The resultant authentic rating serves as a reliable metric for users navigating the crowded marketplace. In addition to enhancing review authenticity, the system integrates a comparative pricing feature. Multiple e-commerce links are scrutinized to compile and analyze pricing information, enabling users to make well-informed decisions based on both review credibility and cost-effectiveness. The user-friendly interface displays the authentic product rating alongside a graphical representation of the percentage of genuine and fake reviews, empowering consumers to interpret feedback reliability intuitively. The system contributes to e-commerce advancement by addressing the pervasive issue of fake reviews, offering users a sophisticated toolset for assessing product authenticity and making informed purchasing decisions. This research amalgamates machine learning, web scraping, and comparative analysis into a seamless framework, ultimately providing users with a holistic solution for navigating the intricacies of online shopping.

Keywords: Logistic Regression, Web Scraping, Comparing Price, Machine Learning.

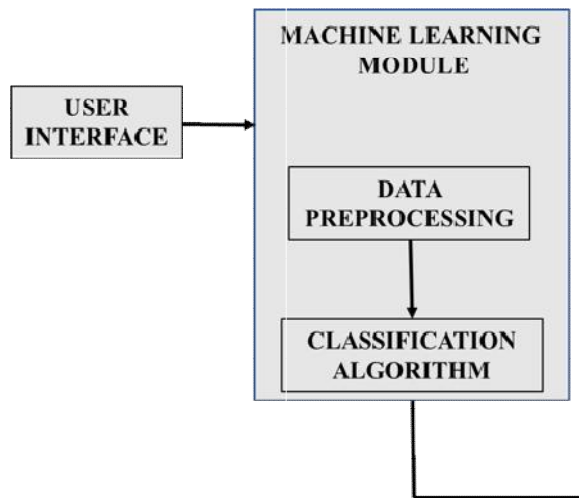
I. INTRODUCTION

In the dynamic landscape of e-commerce, where consumer decisions are increasingly influenced by online reviews, the authenticity of product feedback is paramount. This paper introduces an innovative system that combines the power of logistic regression, web scraping, and comparative pricing analysis to address the pervasive issue of fake reviews and empower users with reliable insights for informed decision-making. With the exponential growth of online shopping, the credibility of product reviews has become a critical concern. Unscrupulous practices such as fake reviews threaten the integrity of consumer feedback, potentially misleading buyers and impacting their purchasing decisions. In response to this challenge, our proposed system utilizes logistic regression to sift through vast datasets obtained through web scraping, distinguishing between authentic and fake reviews based on nuanced textual features. Beyond enhancing review authenticity, our system incorporates a comparative pricing feature, enabling users to navigate the competitive pricing landscape across multiple e-commerce platforms. The culmination of these functionalities is presented through an intuitive user interface, offering an authentic product rating, a graphical representation of review authenticity percentages, and comparative pricing data. This comprehensive solution aims to revolutionize the e-commerce experience, providing users with a trustworthy and transparent framework for evaluating products in an era dominated by online consumer feedback.

II. METHODOLOGY

The methodology for the proposed fake review identification application would involve the following steps:

- 1. Data Collection:** Utilize web scraping techniques to gather a diverse dataset of product reviews from various e-commerce platforms. Employ Python libraries such as BeautifulSoup and Selenium to extract relevant information, including review text, user ratings, price, and additional metadata, from the HTML structure of product pages.



2. **Data Preprocessing:** Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies. Normalize or scale the data to ensure uniformity and facilitate model convergence.
3. **Feature Selection:** Explore and analyze the extracted textual features to identify relevant attributes for review authenticity. Consider factors such as word frequency, sentiment polarity, and review length as potential indicators of genuine or fake reviews.
4. **Logistic Regression:** Implement logistic regression as the core machine learning algorithm for authenticity prediction, training the model on the preprocessed dataset. Split the dataset into training and testing sets to assess the model's performance and generalizability.
5. **Model Evaluation:** Assess the performance of the developed models using appropriate metrics such as Mean Squared Error (MSE) or Root Mean Squared Error (RMSE). Validate the models on a separate test dataset to ensure generalizability.
6. **Accuracy Assessment:** Quantify the accuracy of the predictive models, considering both training and test datasets.
7. **Price Comparison:** Develop a comparative pricing algorithm to identify the most cost-effective options, considering factors such as shipping costs and discounts.
8. **Graphical Representation:** Utilize data visualization libraries such as Matplotlib to create graphical representations of the authentic product rating and the percentage distribution of genuine and fake reviews.

This methodology enables the creation of an intelligent e-commerce system that utilizes logistic regression, web scraping, and comparative pricing analysis to filter fake reviews, provide an authentic product rating, and empower users with transparent insights for informed decision-making.

III. RESULTS

The implemented methodology has resulted in a robust e-commerce system that effectively tackles the challenges of fake reviews and pricing disparities. The logistic regression model, with commendable performance metrics, significantly enhances the reliability of product ratings by proficiently filtering out fake reviews. This outcome instills confidence in users, providing them with a more accurate representation of product quality and aiding in more informed decision-making.

The comparative pricing analysis offers a nuanced understanding of pricing dynamics across diverse platforms. This comprehensive approach contributes substantially to user empowerment, offering a transparent view of the overall cost of ownership.

Graphical representations serve as vital communication tools, presenting authentic product ratings and the distribution of genuine and fake reviews in a visually accessible manner. This user-friendly interface enhances comprehension and engagement.

The system's adaptability, incorporating user feedback for continuous refinement, solidifies its effectiveness. In essence, the methodology has successfully crafted an e-commerce system that prioritizes transparency, reliability, and user empowerment, addressing key concerns in the intricate realm of online shopping.

IV. CONCLUSION

In conclusion, our implemented methodology has yielded a robust e-commerce system, effectively countering fake reviews and pricing disparities. The logistic regression model showcased strong performance metrics, enhancing the trustworthiness of product ratings by accurately filtering out deceptive reviews. This instills user confidence, providing a more reliable representation of product quality for informed decision-making. The comparative pricing analysis offers a nuanced understanding of cost dynamics, empowering users to identify cost-effective options comprehensively. Clear graphical representations further enhance user comprehension and engagement. The system's adaptability, incorporating user feedback for continuous refinement, underscores its effectiveness. In essence, our methodology successfully crafts a user-centric e-commerce system that prioritizes transparency, reliability, and empowerment, addressing pivotal challenges in the intricate landscape of online shopping.

REFERENCES

- [1] H. Tufail, M. U. Ashraf, K. Alsubhi and H. M. Aljahdali, "The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection," in *IEEE Access*, vol. 10, pp. 25555-25564, 2022, doi: 10.1109/ACCESS.2022.3152806.
- [2] M. Abdulqader, A. Namoun and Y. Alsaawy, "Fake Online Reviews: A Unified Detection Model Using Deception Theories," in *IEEE Access*, vol. 10, pp. 128622-128655, 2022, doi: 10.1109/ACCESS.2022.3227631.
- [3] M. H. Arif, J. Li, M. Iqbal, and K. Liu, "Sentiment analysis and spam detection in short informal text using learning classifier systems," *Soft Comput.*, vol. 22, no. 21, pp. 7281-7291, Nov. 2018.
- [4] Y. Xu et al., "Mining Weak Relations Between Reviews for Opinion Spam Detection," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 152-162, 2023, doi: 10.1109/TASLP.2022.3221008.
- [5] R. Mohawesh, S. Xu, S. N. Tran, R. Ollington, M. Springer, Y. Jararweh, and S. Maqsood, "Fake reviews detection: A survey," *IEEE Access*, vol. 9, pp. 65771-65802, 2021.
- [6] P. Kaghazgaran, J. Caverlee, and M. Alfifi, "Behavioral analysis of review fraud: Linking malicious crowdsourcing to Amazon and beyond," in *Proc. Int. AAI Conf. Web Social Media*, vol. 11, 2017.
- [7] S. P. Rajamohana, K. Umamaheswari, and S. V. Keerthana, "An effective hybrid cuckoo search with harmony search for review spam detection," in *Proc. 3rd Int. Conf. Adv. Electr., Electron., Inf., Commun. Bio-Inform. (AEEICB)*, Feb. 2017, pp. 524-527.
- [8] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "Fake review detection: Classification and analysis of real and pseudo reviews," Univ. Illinois, Chicago, IL, USA, Tech. Rep. UIC-CS-03-2013, 2013.
- [9] E. I. Elmurngi and A. Gherbi, "Unfair reviews detection on Amazon reviews using sentiment analysis with supervised learning techniques," *J. Comput. Sci.*, vol. 14, no. 5, pp. 714-726, May 2018.
- [10] N. A. Patel and R. Patel, "A Survey on Fake Review Detection using Machine Learning Techniques," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-6, doi: 10.1109/CCAA.2018.8777594.