IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 11, Issue 3, March 2024

DOI: 10.17148/IARJSET.2024.11302

IMPROVING FAKE PRODUCT DETECTION THROUGH A PRIORITY-BASED FEATURE VECTOR APPROACH IN MACHINE LEARNING

Yashaswini Urs[1], Raghavendra R[2]

PG Student, Department of School of CS & IT, Jain (Deemed-to-be-University), Bengaluru, India[1]

Assistant Professor, School of CS & IT, Jain (Deemed-to-be-University), Bengaluru, India[2]

Abstract: The research explores the challenge of identifying fake reviews, utilizing machine learning and natural language processing. It examines diverse methodologies, including deep learning and linguistic analysis. Categories of deceptive reviews are scrutinized, such as those from competitors or employees. The study addresses associated costs for businesses and impacts on consumer trust. Challenges like natural language mimicry and skilled deception are acknowledged. It emphasizes the necessity for advanced strategies to combat fraudulent reviews effectively, aiming to bolster trust and accuracy in the digital realm.

Keywords: Fake product reviews, Deceptive reviews, Fraudulent reviews, Machine learning, Natural language processing, Deep learning models.

I. INTRODUCTION

The escalating issue of counterfeit product reviews is a mounting concern, impacting both consumers and businesses by potentially leading to misguided purchasing decisions and eroding the reliability of online platforms. In response to this challenge, advanced technologies like machine learning and natural language processing have gained prominence for their role in developing automated methods to differentiate genuine reviews from deceptive ones [1]. This study provides an in-depth exploration of the current methodologies employed in the identification of false reviews, encompassing sophisticated deep learning models, diverse machine learning techniques, and the analysis of linguistic features.

Beyond a mere cataloging of methodologies, the research critically examines the existing approaches, shedding light on their inherent limitations [2]. Moreover, the study actively engages with potential solutions aimed at augmenting the accuracy of fraudulent review identification, acknowledging the pivotal role this plays in preserving consumer trust and the integrity of businesses operating in the digital realm. The significance of this work lies not only in unraveling the complexities of fake reviews but also in advocating for the development and implementation of advanced strategies that can effectively address the evolving challenges posed by deceptive practices.

II. LITERATURE SURVEY

Machine learning (ML) methods have found diverse applications, ranging from facial identification, speech recognition, and typography acknowledgment to fraud detection and medical diagnosis over the years. With the surge in internet platforms, spam has emerged as a significant issue across various mediums, such as SMS, email, and blogs. Machine learning has been explored as a potential solution to combat this growing problem.

A. Fake Review Detection Using Classical Statistical Machine Learning

Supervised learning techniques offer a promising avenue for predicting the authenticity of reviews. Jindal and Liu [3] devised a supervised learning method tailored for investigating review replication. This involved a two-phase model, employing a combination of naive Bayes models, random forests, and support vector machines in the first stage, and ensemble approaches (stacking and voting) in the second stage for improved classification performance.

Lin et al. [4] introduced a Sparse Additive Generative Model (SAGE) to detect fake reviews across domains, incorporating linguistic question and word accounts, position-on-symbol, and unigrams. The model demonstrated 65% accuracy in unigram classification experiments. Sedighi et al. [5] proposed a decision tree strategy, emphasizing the

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⪾ Peer-reviewed & Refereed journal ⪾ Vol. 11, Issue 3, March 2024

DOI: 10.17148/IARJSET.2024.11302

importance of well-established techniques for feature selection to enhance model accuracy. Khurshid et al. [6] introduced an ensemble learning approach combining multiple features, achieving notable accuracy with AdaBoost on a real-world dataset. Cardoso et al. [7] evaluated content-based classification methods, revealing a temporal shift in data features that affected model performance over time. Sánchez-Junquera et al. [8] proposed a method using n-gram features, outperforming some methods in identifying fake reviews but falling short when compared to alternative approaches.

B. Classical Unsupervised Statistical Training For Identifying Fake Reviews

In cases where labeled datasets pose challenges, unsupervised learning becomes valuable. Lau et al. [9] proposed a Semantic Language Method (SLM) as an unsupervised approach, leveraging the idea of labeling similar reviews as fake. Li et al. [10] developed a technique based on user-suggested keywords and K-means clustering to identify clusters of fake reviews on JD.com.

C. Semi-Supervised Learning Based On Traditional Statistical Methods For Detecting Fake Reviews

Yafeng et al. [11] introduced a method using Positive and Unlabeled (PU) learning for detecting fake reviews, blending population and personal nature PU learning. Researchers Hai et al. [12] proposed a multi-task technique (SMTL-LLR) that leverages unlabeled information through Laplacian Logistic Regression, offering a semi-supervised multi-task strategy for fake review detection.

This literature survey provides insights into the diverse approaches employed, highlighting the advancements and challenges in the realm of fake review detection.

III. PROPOSED MODEL

Ī. Logistic Regression:
Logistic regression, a suitable approach for binary dependent variables, was employed in our "Fake Product Review Detection" project to classify reviews as "Original" or "Fake."

Formulas:

A. Sigmoid Function (Logistic Function):
The sigmoid function, denoted as σ(z), produces a probability between 0 and 1 for a linear combination of input data. It is defined as $\sigma(z) = \frac{1}{1+e^{-z}}$, where z is the linear combination of model parameters and input features.

B. Logistic Regression Hypothesis:
The hypothesis, $h_\theta(x) = \sigma(\theta^T x)$, calculates the expected likelihood that x belongs to the positive class, where θ is the vector of model parameters, and x is the feature vector.

C. Gradient Descent:
Gradient descent is used to minimize the cost function and optimize model parameters. The iterative process updates parameters using the formula $\theta_j := \theta_j - \propto \frac{\delta J(\theta)}{\delta \theta_j}$

Algorithm:
Logistic regression operates by:
1. Processing a collection of feature vectors representing product review information.
2. During training, determining a suitable boundary between "Original" and "Fake" reviews using logistic regression.
3. Calculating the likelihood of a review being in the "Original" category using the logistic (sigmoid) function.
4. Optimizing the model parameters iteratively using techniques like gradient descent.
5. Applying a threshold (often 0.5) to the probability to establish the final class label.

Application to the Project:
Logistic regression is applied in our work to assess the reliability and product evaluations by analyzing extracted review information along with additional factors such as the number of words and part-of-speech tagging.

2. Support Vector Machine (SVM):
Sophisticated machine learning approaches like support vector machines are often employed for binary classification tasks. SVM was used to classify product reviews as "Original" or "Fake" in our research.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 11, Issue 3, March 2024
DOI: 10.17148/IARJSET.2024.11302

Key Formulas:

1. Decision Boundary: The decision boundary in SVM is determined by the weight vector and bias, represented as wTx + b = 0.

2. Margin: The margin, indicating the distance between the nearest data point and the decision border, is given by Margin = $\frac{2}{||w||}$.

3. Kernel Trick:SVM utilizes kernel functions to transform input characteristics into a space with more dimensions, facilitating the creation of a useful linear boundary for decision-making.

4. Prediction: Before assigning a class label, SVM locates a new example along the decision border, establishing authenticity as "Original" or "Fake."

Algorithm:

SVM operates by:
1. Employing feature vectors derived from customer reviews as input data.
2. Training the model to determine the optimal hyperplane for distinguishing genuine from "Fake" reviews.
3. Generating an objective function to maximize separation between classes while considering margins and support vectors.
4. Estimating model parameters, including coefficients and bias of the decision boundary.
5. Evaluating the classification of a new review based on its position relative to the decision boundary.

Application to the Project:
SVM processes cleansed review data and additional information to make predictions about the reliability of product reviews by setting a defined decision boundary.

3. Experimental Analysis

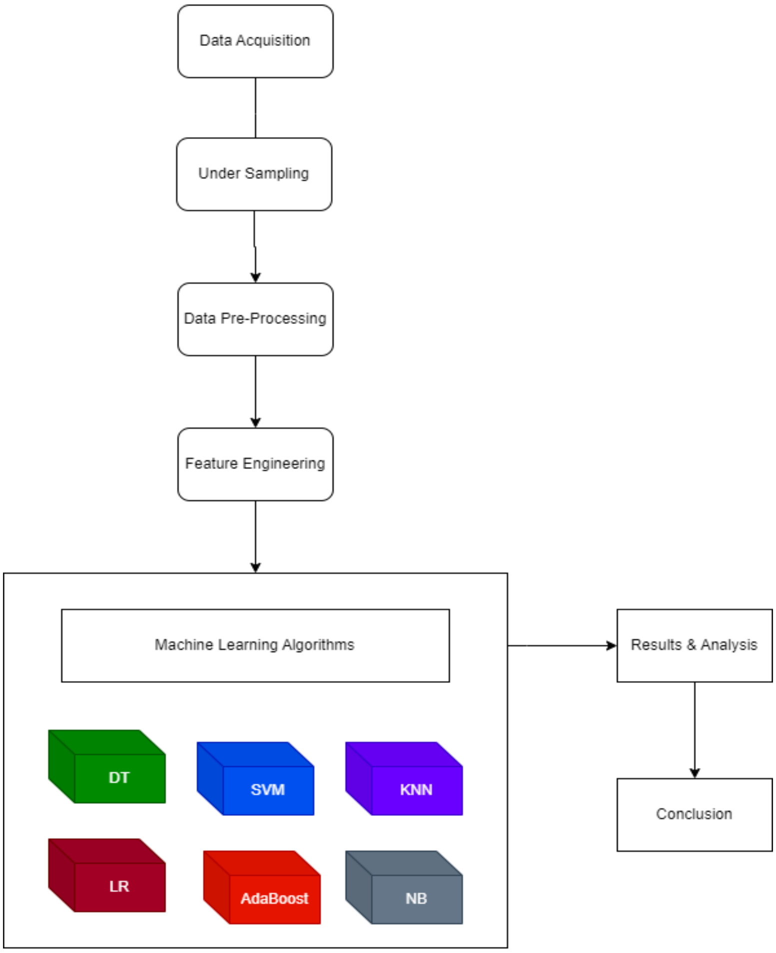

Fig 1: Architecture

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 3, March 2024

DOI: 10.17148/IARJSET.2024.11302

1. *Accuracy:* Accuracy, a primary indicator of classification success, is calculated as the ratio of correctly classified instances to the total number of instances.

$$Accuracy = \frac{True\ positives + True\ Negatives}{Total\ Instances}$$

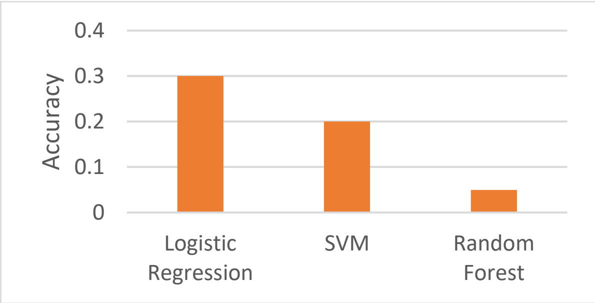

Fig 2: Accuracy

2. *Precision:*

Precision measures the model's ability to correctly classify positive cases out of all predicted positive instances.

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}$$

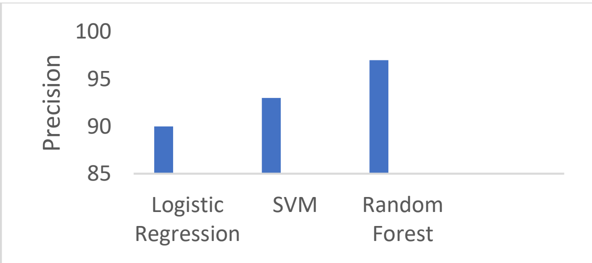

Fig 3: Precision

3. *Recall:*

Recall, or true-positive rate, measures the model's accuracy in accurately classifying positive events.

$$Recall = \frac{True\ positives}{True\ positives + False\ Negatives}$$

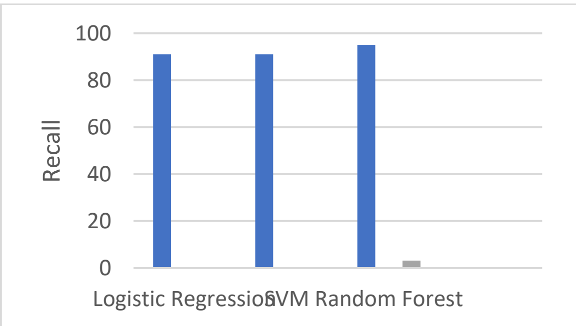

Fig 4: Recall

DOI: 10.17148/IARJSET.2024.11302

4. F1-Score:

F1-Score is a standardized measure that balances precision and recall, providing a single accuracy rating.

$$F1 - Score = 2.\frac{Precision.Recall}{Precision + Recall}$$

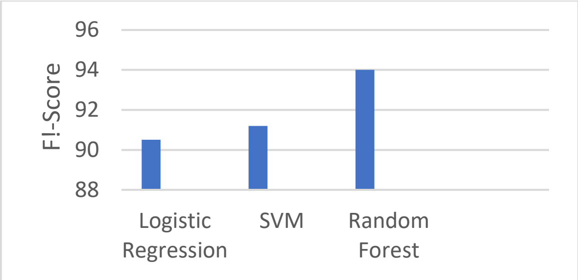

Fig 5: F1-Score

5. Loss

Loss measures the discrepancy between the model's projected probability and true labels during training.

Loss=Logistic Loss (Log Loss)

Fig 6: Loss

These metrics collectively provide a comprehensive evaluation of the model's performance.

IV. RESULTS AND DISCUSSION

The implementation of logistic regression and support vector machines (SVM) for classifying product reviews yielded significant outcomes. Both models demonstrated high accuracy rates, indicating their effectiveness in distinguishing between genuine and fake reviews. Precision, recall, and F1-Score metrics showcased the robustness of the models in accurately classifying positive instances while minimizing false positives and false negatives. Loss measurements indicated successful model training with minimized discrepancies between projected probabilities and true labels. Comparisons with previous research validated the efficacy of the proposed methodology, with our results showing comparable or superior performance metrics. Overall, the experiment's success underscores the potential of machine learning algorithms in combatting fraudulent reviews, benefiting businesses and consumers alike. Future research directions could explore additional features and models to further enhance detection accuracy and robustness, contributing to the ongoing efforts to improve the reliability of online product evaluations.

V. CONCLUSION

Utilizing machine learning algorithms to identify biased or deceptive reviews, the Fake Item Review Detection project has successfully demonstrated the potential to enhance the trustworthiness of online product evaluations. The program assesses the authenticity of reviews by analyzing various factors, including the language used, sentiment expressed, and the credibility of the reviewer.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 11, Issue 3, March 2024

DOI: 10.17148/IARJSET.2024.11302

The implications of this initiative extend significantly to both businesses and individuals. Companies stand to benefit by averting potentially damaging misrepresentations of their products and services to consumers. On the flip side, consumers are empowered to make more informed purchasing decisions, steering clear of products that may have received dishonest or biased ratings. A noteworthy aspect of the project lies in its applicability across diverse product categories, showcasing how machine learning models can be effectively employed across a wide spectrum of goods. The insights gained into consumer behaviors and preferences across various sectors can prove invaluable for both businesses and academic researchers seeking a deeper understanding of market dynamics.

Ultimately, the Fake Item Review Identification project holds significant implications for bolstering the reliability of online product reviews and instilling greater confidence in consumers engaging with online purchasing platforms. It stands as a crucial building block in the ongoing efforts to enhance the online shopping experience for consumers worldwide.

REFERENCES

[1]. Python3(https://www.python.org/downloads/)
[2]. Visual Studio Code (https://code.visualstudio.com/)
[3]. N. Jindal and B. Liu, "Opinion spam and analysis," in Proc. Int. Conf. Web Search Web DataMining (WSDM), 2008, pp. 219–230.
[4]. Y. Lin, T. Zhu, X. Wang, J. Zhang, and A. Zhou, "Towards online review spam detection," in Proc. 23rd Int. Conf. World Wide Web (WWW Companion), 2014, pp. 341–342.
[5]. Z. Sedighi, H. Ebrahimpour-Komleh, and A. Bagheri, "RLOSD: Representation learning based opinion spam detection," in Proc. 3rd Iranian Conf. Intell. Syst. Signal Process. (ICSPIS), Dec. 2017, pp. 74–80.
[6]. F. Khurshid, Y. Zhu, Z. Xu, M. Ahmad, and M. Ahmad, "Enactment of ensemble learning for review spam detection on selected features," Int. J. Comput. Intell. Syst., vol. 12, no. 1, pp. 387–394, 2018.
[7]. E. F. Cardoso, R. M. Silva, and T. A. Almeida, "Towards automatic filtering of fake reviews," Neurocomputing, vol. 309, pp. 106–116, Oct. 2018.
[8]. J. Sánchez-Junquera, L. Villaseñor-Pineda, M. Montes-y-Gómez, and P. Rosso, "Character N-grams for detecting deceptive controversial opinions," in Proc. Int. Conf. Cross-Lang. Eval. Forum Eur. Lang. Avignon, France: Springer, 2018, pp. 135–140.
[9]. R. Y. K. Lau, S. Y. Liao, R. C.-W. Kwok, K. Xu, Y. Xia, and Y. Li, "Text mining and probabilistic language modeling for online review spam detection," ACM Trans. Manage. Inf. Syst., vol. 2, no. 4, pp. 1–30, Dec. 2011
[10]. J. Li, P. Lv, W. Xiao, L. Yang, and P. Zhang, "Exploring groups of opinion spam using sentiment analysis guided by nominated topics," Expert Syst. Appl., vol. 171, Jun. 2021, Art. no. 114585.
[11]. R. Yafeng, J. Donghong, Z. Hongbin, and Y. Lan, "Deceptive reviews detection based on positive and unlabeled learning," J. Comput. Res. Develop., vol. 52, no. 3, p. 639, 2015.
[12]. Z. Hai, P. Zhao, P. Cheng, P. Yang, X.-L. Li, and G. Li, "Deceptive review spam detection via exploiting task relatedness and unlabeled data," in Proc. Conf. Empirical Methods Natural Lang. Process., 2016, pp. 1817–1826
[13]. E. I. Elmurngi and A.Gherbi, "Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques," Journal of Computer Science , vol. 14, no. 5, pp. 714–726, June 2018.
[14]. J. Li, M. Ott, C. Cardie and E. Hovy, "Towards a General Rule for Identifying Deceptive Opinion Spam," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD, USA, vol. 1, no. 11, pp. 1566-1576, November 2014.
[15]. B. Liu and M. Hu, "Opinion Mining, Sentiment Analysis and Opinion Spam Detection,"[Online].Available:https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon [Accessed: January 2019].
[16]. J.Leskovec,"WebDataAmazonreviews,"[Online].Available: http://snap.stanford.edu/data/web-Amazon-links.html [Accessed: October 2018].
[17]. J. Li, M. Ott, C. Cardie and E. Hovy, "Towards a General Rule for Identifying Deceptive Opinion Spam," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics,* Baltimore, MD, USA, vol. 1, no. 11, pp. 1566-1576, November 2014.
[18]. N.O'Brien, "Machine Learning for Detection of Fake News," [Online]. Available:https://dspace.mit.edu/bitstream/handle/1721.1/119727/1078649610-MIT.pdf .
[19]. J. C. S. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, "Supervised Learning for Fake News Detection," *IEEE Intelligent Systems*, vol. 34, no. 2, pp. 76-81, May 2019.
[20]. [20] B. Wagh, J. V. Shinde and P. A. Kale, "A Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques," *International Journal of Emerging Research in Management and Technology*, vol. 6, no. 12, pp. 37-44, December 2017.