Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

# Fake Review Detection Using Supervised Machine Learning Techniques

B Chakradhar[1], Kumar Swamy Pydi[2], Naveen Thota[3], Jani Basha[4], Yerra Harshita[5]
[1]Asst.Professor, [2, 3, 4, 5]B.Tech Student
[1, 2,3,4,5] Department of Computer Science & Engineering
[1, 2,3,4,5] Raghu Engineering College, Visakhapatnam, India
chakri.589@gmail.com, kumarswamypydi@gmail.com,
naveenthota163@gmail.com, harshitha.y2022@gmail.com, janibasha070503@gmail.com

**Abstract: - The proliferation of online reviews has become a pivotal factor influencing consumer decisions. However, the prevalence of fake reviews poses a significant challenge to the authenticity of these platforms. We aim to address this issue through the implementation of a Fake review detection using Supervised Machine Learning Techniques. The collections of dataset encompass both genuine and deceptive reviews. Various features, including language patterns, and review metadata, are extracted to form a comprehensive set of input variables. A Supervised Learning algorithm, such as logistic regression, support vector machines, or neural networks, is employed to train the model using the labelled dataset. The trained model undergoes rigorous evaluation to gauge its effectiveness in discerning between authentic and fake reviews. Fine-tuning and optimization processes are carried out based on the evaluation results to enhance the model's accuracy and generalization capabilities. Upon successful development and validation, the ultimate goal is to provide a practical tool for online platforms and businesses to automatically identify and mitigate the impact of fake reviews. This research contributes to the field of online reputation management and consumer trust by leveraging advanced machine learning techniques to foster a more reliable online review ecosystem.**

***Keywords:*** *Online reviews, Fake reviews, Supervised Machine Learning, Language patterns, Logistic regression, Support Vector Machines (SVM).*

## I. Introduction

In the contemporary digital landscape, online reviews play a pivotal role in shaping consumer decisions. As individuals increasingly rely on the various online platforms, the authenticity and reliability of these reviews become paramount. However, the surge in deceptive practices, including the creation of fake online reviews, presents a formidable challenge to the credibility of such platforms. Recognizing the importance of maintains trust in online review systems, so we introduce a solution – Fake Review Detection Using Supervised Machine Learning Techniques. The ubiquity of online reviews across e-commerce, hospitality, and service sectors their influence on consumer behaviour. Yet, the malicious manipulation of these platforms through the submission of fake reviews has become a prevalent issue. Fake reviews not only mislead consumers but also undermine the integrity of businesses and online platforms. Therefore, the need for robust mechanisms to identify and filter out fake reviews is imperative for sustaining a trustworthy online review ecosystem. This project sets out to address this challenge by harnessing the power of supervised machine learning techniques. The primary objective is to develop a model capable of distinguishing between genuine and deceptive online reviews through the analysis of various linguistic and metadata features. By employing a diverse dataset containing labelled instances of authentic and fake reviews, the project aims to train a machine learning algorithm to discern subtle patterns and characteristics indicative of deceptive practices. The significance of this lies in its potential to provide an automated and scalable solution for the identification of fake online reviews. As the model learns from a diverse range of features and patterns inherent in both authentic and deceptive reviews, it holds the promise of being a versatile tool applicable across different domains and platforms. In the subsequent sections of the methodology, dataset collection process, feature extraction techniques, and the chosen supervised learning algorithm will be discussed in detail. The success will be measured by its ability

to accurately identify fake reviews, contributing to the broader discourse on online reputation management and bolstering consumer trust in the digital marketplace.

## II. Literature Survey

The proposed model accepts products and reviews as its input and generates classification results as its output. The proposed method offers classification results through a bagging model which bags three classifiers including product word composition classifier [1].

The spam detection problems as a network classification task on the user-review-product network. In this task, users are to be classified as spammer or benign, products as targeted or non-targeted, and reviews as fake or genuine. To aid the network classification, they utilize additional metadata (i.e., ratings and review text) to extract indicative features of spam, which they incorporate into the inference procedure. The proposed method works in a completely unsupervised fashion; however, it can easily accommodate labels when available. As such, it is amenable to semi-supervised detection [6].

Extract relevant attributes from review text. Train classifiers using labelled data. Investigate the impact of different features on model performance. Their work contributes to the development of effective machine learning models for spam detection [8].

Yelp has a filtering algorithm in place that identifies fake/suspicious reviews and separates them into a filtered list. Yelp datasets contain both recommended and filtered reviews. Consider them as genuine and fake, respectively and also separate the users into two classes; spammers: authors of fake (filtered) reviews, and benign: authors with no filtered reviews. Evaluate method on three real-world datasets collected from Yelp.com, containing filtered (spam) and recommended (non-spam) reviews. To the best of knowledge, work provides the largest scale quantitative evaluation to date for the opinion spam problem [4].

Explored feature engineering for review spam detection. Extract relevant attributes from review text. Trained classifiers Focused on burst behaviour in reviews. Investigate temporal patterns and frequency of posting. Combine temporal features with content-based features [5].

## III. Proposed System

Fake Review Detection Using Supervised Machine Learning Techniques can be divided into several key modules, each contributing to the overall goal of developing an effective detection system. Here is a detailed explanation of the methodology, organized module-wise.

**1. Dataset Collection and Pre-processing:** Gather a diverse dataset containing labelled instances of genuine and fake online reviews. Identify and select relevant datasets from various domains and online platforms. Manually label a subset of reviews as genuine and fake for training purposes. Pre-process the dataset by removing irrelevant information, handling missing data, and ensuring consistency.

**2. Feature Extraction:**

Extract relevant features from the reviews to represent linguistic, sentiment, and metadata characteristics. Conduct sentiment analysis to capture the emotional tone of the reviews. Extract metadata features such as review length, timestamps, and user activity.

**3. Supervised Machine Learning Model Training:**

Train a robust machine learning model using the labelled dataset to distinguish between genuine and fake reviews. Split the dataset into training and validation sets. Choose a suitable supervised learning algorithm (e.g., logistic regression, support vector machines, neural networks). Train the model on the training set, adjusting hyper-parameters for optimal performance. Validate the model on the separate validation set to using labelled data.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

**4. Cross-Domain Generalization:**
Ensure that the trained model can adapt well across different domains and online platforms. Explore techniques for transfer learning and domain adaptation during the training phase. Test the model's performance on datasets from diverse domains to assess its cross-domain generalization capabilities.

**5. Evaluation and Fine-Tuning:**
Evaluate the model's performance and refine its parameters to enhance accuracy and reliability. Use metrics such as precision, recall, F1 score, and accuracy to evaluate the model. Identify and analyse false positives and false negatives for further insights. Fine-tune the model based on evaluation results to improve its overall performance.

**6. Integration with Online Platforms:**
Develop a practical tool that can be seamlessly integrated into online platforms and review systems. Implement an interface for easy integration, allowing the model to process and analyse incoming reviews. Provide functionalities for automatically flagging or filtering potentially deceptive reviews. Ensure compatibility with various platforms and industries.

**7. Documentation and Reporting:**
Document the entire process, methodologies, and outcomes for future reference and presentation. Create detailed documentation for each module, including code documentation, data pre-processing steps, and model training procedures. Generate reports summarizing the project's methodology, key findings, and the performance of the developed system.

## IV. Result Analysis

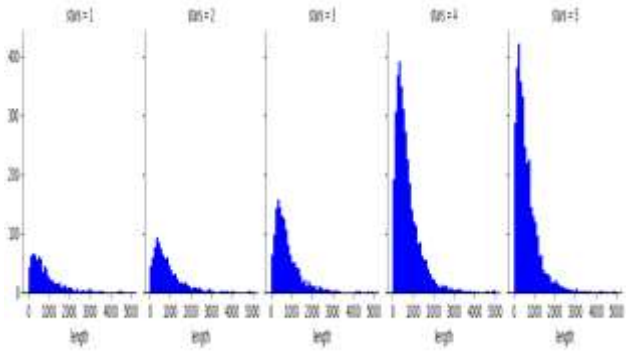

*Fig – 1.1 Show words involved in each star review.*

Visualize if there is any correlation between stars and the length of the review.

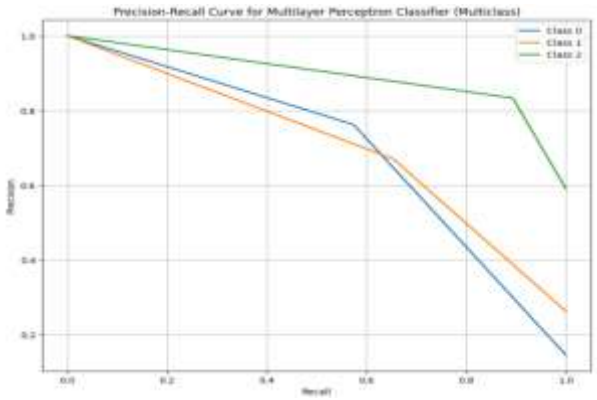


*Fig-1.2 Precision-Recall Curve for Multilayer Perception Classifier (Multiclass)*

The precision-recall curve is used to evaluate the performance of a classification model, especially in scenarios where class imbalance exists or where the positive class is of particular interest.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

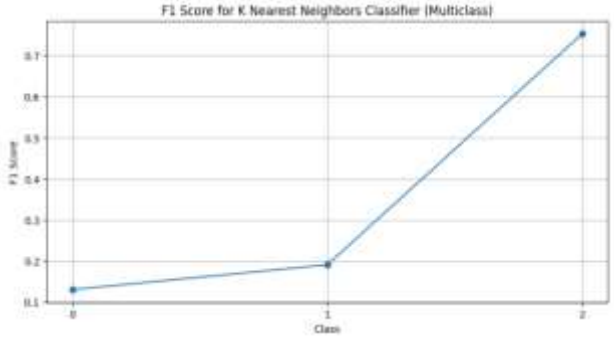


*Fig – 1.3 F1 Score for K-Nearest Neighbours Classifier (Multiclass)*

The F1 score is a metric commonly used in binary classification tasks, but it can also be extended to multiclass classification by averaging the scores across different classes. It is the harmonic mean of precision and recall and is calculated using the following formula:

$$F1=2\times (\text{Precision} + \text{Recall})/(\text{Precision} \times \text{Recall})$$

By following this modular methodology, the project aims to systematically address the challenges of fake online review detection and provide a robust,

```
Accuracy: 71.08
Confusion Matrix for Support Vector Machine:
[[ 31  23 108]
 [  5 122 165]
 [  1  19 636]]
Score: 71.08
Classification Report:              precision    recall  f1-score   support

           1       0.84      0.19      0.31       162
           3       0.74      0.42      0.54       292
           5       0.70      0.97      0.81       656

    accuracy                           0.71      1110
   macro avg       0.76      0.53      0.55      1110
weighted avg       0.73      0.71      0.67      1110
```

*Fig – 1.4 Accuracy and score of Support Vector Machine Algorithm*

Accuracy and score of every algorithm is shown to know about the efficiency of algorithms. The confusion matrix is also shown for better understanding.

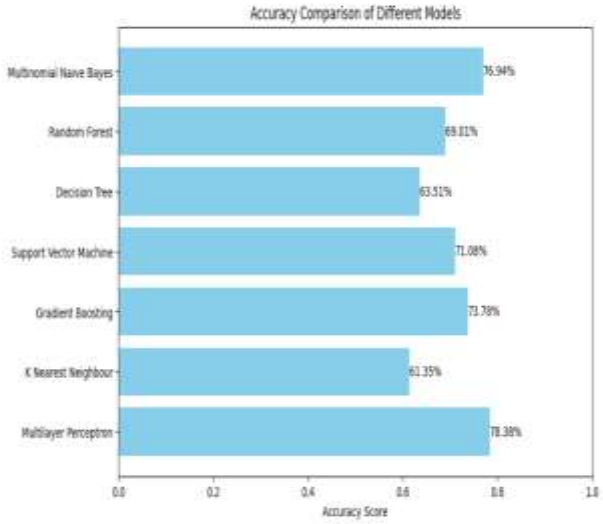


*Fig – 1.5 Graph showing accuracy of different algorithms.*

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 4, April : 2024

Accuracies of different algorithms are displayed to know which algorithm works better for the given data set as the UI is developed to show the probability of review being fake or genuine depending on it.

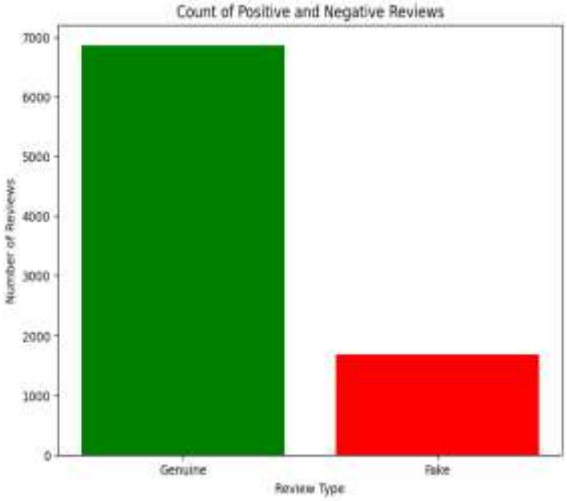


*Fig 1.6 Comparing Genuine to Fake Reviews*

After going regress training and evaluations by different algorithms it shows the amount of genuine and fake reviews involved in the given data set.

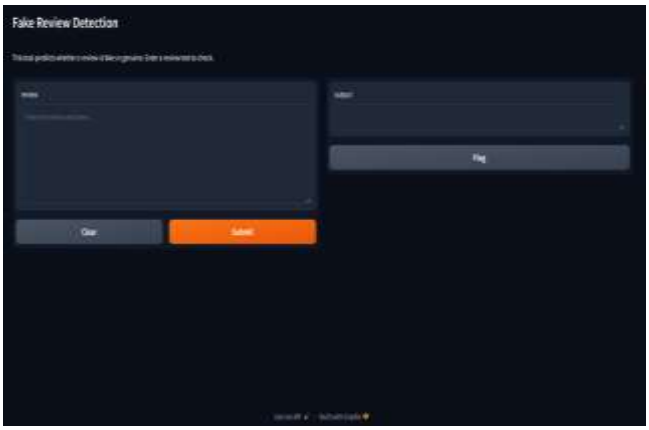


*Fig – 1.7 Check review manually.*

The UI designed helps the user to type the reviews manually and check the nature of the reviews. Here 1 indicates a fake review and 5 indicates a genuine review. The results are given and based on the trained dataset and the algorithm that gave the best score for the corresponding dataset.

## V. Conclusion

The Fake Review Detection Using Supervised Machine Learning Techniques represents a significant step towards addressing the pervasive issue of deceptive reviews on online platforms. Through the implementation of a supervised machine learning model, the system aims to provide users and platforms with a reliable tool to identify and mitigate the impact of fake reviews. The project identifies several avenues for future development, including the exploration of advanced machine learning models, integration with emerging technologies, and continuous dataset enhancement. Collaborate with the industry stakeholders, researchers, and users.

**References**

[1] Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding Deceptive Opinion Spam by Any Stretch of the Imagination. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics.
[2] Jindal, N., & Liu, B. (2008). Opinion Spam and Analysis. In Proceedings of the International Conference on Web Search and Web Data Mining.
[3] Mukherjee, A., & Liu, B. (2012). Spotting Fake Reviewer Groups in Consumer Reviews. In Proceedings of the 21st International Conference on World Wide Web.
[4] Rayana, S., & Akoglu, L. (2015). Collective Opinion Spam Detection: Bridging Review Networks and Metadata. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
[5] Feng, S., Banerjee, S., & Choi, Y. (2012). Syntactic Stylometry for Deception Detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics.
[6] Fei, Y., & Mukherjee, A. (2016). Exploiting Burstiness in Reviews for Review Spammer Detection. In Proceedings of the 25th International Conference on World Wide Web.
[7] Jindal, N., & Liu, B. (2007). Review Spam Detection. In Proceedings of the International Conference on Web Search and Web Data Mining.
[8] Li, F., Huang, M., Yang, Y., & Zhu, X. (2011). Learning to Identify Review Spam. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management.
[9] Mukherjee, A., Liu, B., Glance, N., & Jindal, N. (2013). Spotting Fake Reviews via Collective Classification. In Proceedings of the 21st ACM International Conference on Information and Knowledge Management.
[10] Ott,M., Cardie, C., & Hancock, J. T. (2013). Negative Deceptive Opinion Spam. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Comuptational Linguistics.