COMBATING FAKE PRODUCT REVIEWS: DETECTION AND PREVENTION TECHNIQUES

Dr. K. Raghu Naga Prabhakar [1] , Dr.B.Bhavani [2] , A. Tulasi [3]

[1]Professor, Department of Business Management, Aurora's PG College (MBA), Uppal, Hyderabad
Email: prabhakalepu@gmail.com

[2] Assoc. Professor, Department of Business Management, Aurora's PG College (MBA), Uppal, Hyderabad
Email:bb3486577@gmail.com
[3] Assistant Professor, Department of Business Management, Aurora's PG College (MBA), Uppal, Hyderabad
Email: ananthatulasi6@gmail.com

ABSTRACT

Both the customer and the seller can benefit from customer reviews, ratings, and experience stories on e-commerce websites and online services. In addition to helping other consumers better understand their experience with the product, the reviewer can boost brand loyalty. In a similar vein, when customers submit favorable reviews for their products, vendors benefit from increased profile-building through product sales. Unfortunately, suppliers may abuse these review processes. For instance, one may fabricate favourable evaluations to boost a brand's reputation or attempt to denigrate rival items by posting fictitious negative reviews on them. Current supervised learning solutions involve utilizing various machine learning algorithms and technologies such as Weka. In contrast to previous research, I opted to work with a large range of vocabulary, such as multiple datasets merged into a single big data collection, rather than a restricted dataset. Emojis and text content in the reviews have been used to inform sentiment analysis. Reviews that are not real are identified and sorted. The Naïve Bayes, Linear SVC, Support Vector Machine, and Random Forest methods are used to obtain the test results.Sorting these reviews into authentic and fraudulent reviews is the implemented (recommended) approach. Adding a sentiment classifier to Naïve Bayes yields the maximum accuracy.

Keywords: Fake Product, Detection, Prevention, SVC, Random forest

1. INTRODUCTION

Everyone can freely express his/her views and opinions anonymously and without the fear of consequences. Social media and online posting have made it even easier to post confidently and openly. These opinions have both pros and cons while providing the right feedback to reach the right person which can help fix the issue and sometimes a con when these get manipulated These opinions are regarded as valuable. This allows people with malicious intentions to easily make the system to give people the impression of genuineness and post opinions to promote their own product or to discredit the competitor products and services, without revealing identity of themselves or the organization they work for. Such people are called opinion spammers and these activities can be termed as opinion spamming.
There are few different types of opinion spamming. One type is giving positive opinions to some products with intention to promote giving untrue or negative reviews to products to

damage their reputation. Second type consists of advertisements with no opinions on product. There is lot of research work done in field of sentiment analysis and created models while using different sentiment analysis on data from various sources, but the primary focus is on the algorithms and not on actual fake review detection. One of many other research works by E. I. Elmurngi and A. Gherbi [1] used machine learning algorithms to classify the product reviews on Amazon.com dataset [2] including customer usage of the product and buying experiences. The use of Opinion Mining, a type of language processing to track the emotion and thought process of the people or users about a product which can in turn help research work.

Opinion mining, which is also called sentiment analysis, involves building a system to collect and examine opinions about the product made in social media posts, comments, online product and service reviews or even tweets. Automated opinion mining uses machine learning, a component of artificial intelligence. An opinion mining system can be built using a software that can extract knowledge from dataset and incorporate some other data to improve its performance.

One of the biggest applications of opinion mining is in the online and e-commerce reviews of consumer products, feedback and services. As these opinions are so helpful for both the user as well as the seller the e-commerce web sites suggest their customers to leave a feedback and review about their product or service they purchased. One of the biggest applications of opinion mining is in the online and e-commerce reviews of consumer products, feedback and services. As these opinions are so helpful for both the user as well as the seller the e-commerce web sites suggest their customers to leave a feedback and review about their product or service they purchased.

2.LITERATURE SURVEY AND RELATED WORK

Data Sources:

Discuss various sources of data, such as e-commerce websites, social media, and review platforms.

Examine the types of data used in research, including text reviews, ratings, and user profiles.

Data Preprocessing:

Describe the steps taken to clean and prepare the data for analysis.

Discuss text preprocessing techniques like tokenization, stemming, and stop-word removal.

Feature Extraction:

Explore different feature extraction methods, including TF-IDF, word embeddings (e.g., Word2Vec, GloVe), and deep learning approaches (e.g., BERT embeddings).

Classification Algorithms:

Survey various machine learning and deep learning algorithms used for fake review detection, such as SVM, Random Forest, and neural networks.

Highlight ensemble methods and their effectiveness.

Lexical and Semantic Analysis:

Discuss how researchers use lexical and semantic analysis techniques to identify fake reviews based on language patterns and sentiment analysis.

Behavioral Analysis:

Examine studies that focus on user behavior, including reviewing patterns, posting frequency, and user history analysis.

Feature Engineering:

Explore novel features or feature combinations that have been proposed to improve detection accuracy.

Datasets:

List publicly available datasets for fake review detection and their characteristics.
Evaluate the limitations and challenges of using these datasets.

Evaluation Metrics:

Describe common evaluation metrics used in the literature, such as accuracy, precision, recall, F1-score, and AUC-ROC.

Challenges and Future Directions:

Summarize the challenges and limitations in existing research.
Suggest potential future research directions, including adapting to evolving fake review techniques.

Real-World Applications:

Discuss practical applications of fake review detection in industry, such as improving consumer trust and brand reputation.

Conclusion:

Summarize key findings from the literature survey and their implications.
Emphasize the importance of ongoing research in this area.

3. EXISTING SYSTEM

There are several existing approaches and systems for detecting fake product reviews:
NLP and Sentiment Analysis: Many systems use Natural Language Processing (NLP) and sentiment analysis techniques to analyze the language used in reviews. They look for inconsistencies, overly positive or negative language, and unnatural patterns that might indicate fake reviews.
User Behavior Analysis: Some systems focus on analyzing the behavior of users posting reviews. They may look at the frequency and timing of reviews, the history of the reviewer, and whether the reviewer has a pattern of posting similar reviews for different products.
Machine Learning Models: Machine learning models, such as supervised classifiers or deep learning models, can be trained on labeled datasets of fake and genuine reviews. These models can then predict the likelihood of a given review being fake based on various features.
Review Metadata Analysis: Analyzing metadata associated with reviews, such as the reviewer's location, the timing of the review, and the product's release date, can help in detecting anomalies and fake reviews.
Collaborative Filtering: Collaborative filtering techniques can be used to identify suspicious patterns in the relationships between reviewers and products. For example, if a group of reviewers consistently gives high ratings to each other's products, it could be a sign of fake reviews.
Human Moderation: Some platforms employ human moderators to manually review and verify product reviews, removing those that are suspected to be fake.
Blockchain and Tamper-Proof Systems: Blockchain technology is being explored to create tamper-proof review systems where once a review is submitted, it cannot be altered or deleted, reducing the chances of manipulation.
Advanced AI Models: State-of-the-art AI models, such as GPT-3 and its successors, can be used to generate fake reviews. Consequently, advanced AI models are also being used to detect fake reviews by identifying inconsistencies in the language and context of reviews

4. PROPOSED SYSTEM

To solve the major problem faced by online websites due to opinion spamming, this project proposes to identify any such spammed fake reviews by classifying them into fake and genuine. The method attempts to classify the reviews obtained from freely available datasets from various sources and categories including service based, product based, customer feedback, experience based and the crawled Amazon dataset with a greater accuracy using Naïve Bayes [7], Linear SVC, SVM, Random forest, Decision Trees algorithm. In order to improve the accuracy, the additional features like comparison of the sentiment of the review, verified purchases, ratings, emoji count, product category with the overall score are used in addition to the review details.

A classifier is built based on the identified features. And those features are assigned a probability factor or a weight depending on the classified training sets. This is a supervised learning technique applying different Machine learning algorithms to detect the fake or genuine reviews,

The high-level architecture of the implementation can be seen in Figure:1 and the problem is solved in the following six steps:

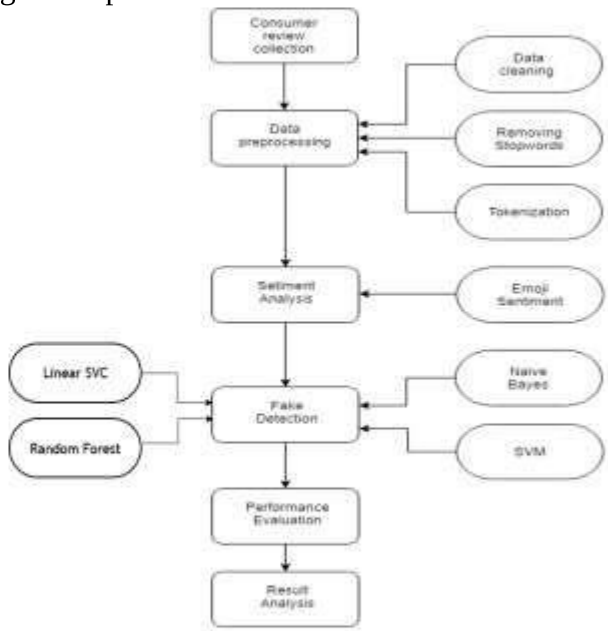

Figure 1: Implementation Architecture

Data Collection

Consumer review data collection- Raw review data was collected from different sources, such as Amazon, websites for booking Airlines, Hotel and Restaurant, CarGurus, etc. reviews. Doing so was to increase the diversity of the review data. A dataset of 21000 was created.

Data Preprocess

Processing and refining the data by removal of irrelevant and redundant information as well as noisy and unreliable data from the review dataset.

Step 1 Sentence tokenization

The entire review is given as input and it is tokenized into sentences using NLTK package.

Step 2 Removal of punctuation marks

Punctuation marks used at the starting and ending of the reviews are removed along with additional white spaces.

Step 3 Word Tokenization

Each individual review is tokenized into words and stored in a list for easier retrieval. Step 4 Removal of stop words

Affixes are removed from the stem. For example, the stem of "cooking" is "cook", and the stemming algorithm knows that the "ing" suffix can be removed. A few words from the frequent word list is shown below in Figure: 2.

Feature Extraction

Rating-Fake reviews in most scenarios have 5 out of 5 stars to entice the customer or have the lowest rating for the competitive products thus it plays an important role in fake detection.

Verified Purchase-Purchase reviews that are fake have lesser chance of it being verified purchase than genuine reviews.Thus these combination of features are selected for identifying the fake reviews.This in turn improves the performance of the prediction models.

Sentiment Analysis

Classifying the reviews according to their emotion factor or sentiments being positive, negative or neutral. It includes predicting the reviews being positive or negative according to the words used in the text, emojis used, ratings given to the review and so on. Related research [8] shows that fake reviews has stronger positive or negative emotions than true reviews. The reasons are that, fake reviews are used to affect people opinion, and it is more significant to convey opinions than to plainly describe the facts. The Subjective vs Objective ratio matters: Advertisers post fake reviews with more objective information, giving more emotions such as how happy it made them than conveying how the product is or what it does. Positive sentiment vs negative sentiment: The sentiment of the review is analyzed which in turn help in making the decision of it being a fake or genuine review.

Fake Review Detection

Classification assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. Each data in the review file is assigned a weight and depending upon which it is classified into respective classes - Fake and Genuine.

Performance Evaluation and Results

Comparison of the accuracies of various models and classifiers with enhancements for better results, as discussed in Accuracy Enhancements [chapter 6]

5. IMPLEMENTATION

Project Title: Fake Product Review Detection
Module Implementation:
1. Data Collection:
- Gather a diverse dataset of product reviews from various sources, including e-commerce websites and social media platforms.
- Collect both genuine and potentially fake reviews to train and test the detection model.

2. Data Preprocessing:
- Clean and preprocess the collected data to handle text data effectively.
- Perform text normalization, including lowercasing and removing punctuation.
- Tokenize the reviews and remove stop words.

3. Feature Extraction:
- Extract relevant features from the text data that can be used to identify fake reviews.
- Common features may include sentiment analysis, word frequency, and n-grams.

4. Labeling Data:
- Manually label a subset of the reviews as genuine or fake to create a labeled dataset for supervised learning.

5. Model Selection:

- Choose appropriate machine learning or natural language processing (NLP) models for fake review detection.
- Common models include Logistic Regression, Random Forest, Support Vector Machines, and deep learning models like LSTM or BERT.

6. Model Training:
- Train the selected model using the labeled dataset, using labeled reviews to teach the model to distinguish between genuine and fake reviews.

7. Model Evaluation:
- Evaluate the model's performance using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC on a validation dataset.
- Fine-tune the model if necessary to achieve desired performance.

8. Text Analysis Techniques:
- Implement various text analysis techniques such as sentiment analysis, semantic analysis, and topic modeling to gain insights into the content and context of reviews.

9. Anomaly Detection:
- Utilize anomaly detection techniques to identify unusual patterns or outliers in the reviews that may indicate fake reviews.

10. User Behavior Analysis (Optional):
- Analyze user behavior, including review history and posting frequency, to detect patterns associated with fake reviewers or review farms.

11. Ensemble Methods (Optional):
- Implement ensemble methods such as bagging or boosting to combine the predictions of multiple models for improved accuracy.

12. Real-time Detection (Optional):
- Develop a real-time monitoring system that can detect fake reviews as they are posted, providing immediate alerts to businesses.

13. Deployment:
- Deploy the trained model in a production environment to automatically detect fake reviews in real-time or batch processing.

14. Continuous Monitoring and Updates:
- Continuously monitor the performance of the detection system and update the model as new fake review patterns emerge.

15. Reporting and Visualization:
- Create reports and visualizations to present the results and insights gained from the fake review detection system.

16. User Interface (Optional):
- Develop a user-friendly interface for businesses and platforms to access and visualize the results of fake review detection.

By implementing these modules, you can create an effective Fake Product Review Detection system that helps businesses maintain the integrity of their online reviews and protect consumers from misleading information.

6. RESULTS AND DISCUSSION SCREEN SHOTS

Output Screen

```
In [3]: dataframe = pd.read_csv('datasest.csv') #reading our dataset which contains text and a label whether it is fake or real
        dataframe.head() #printing the first 5 columsn in our dataset
```

Out[3]:

	Unnamed: 0	category	rating	label	text_
0	0	Home_and_Kitchen_5	5.0	CG	love well made sturdi comfort i love veri pretti
1	1	Home_and_Kitchen_5	5.0	CG	love great upgrad origin i 've mine coupl year
2	2	Home_and_Kitchen_5	5.0	CG	thi pillow save back i love look feel pillow
3	3	Home_and_Kitchen_5	1.0	CG	miss inform use great product price i
4	4	Home_and_Kitchen_5	5.0	CG	veri nice set good qualiti we set two month

```
Classification Report:              precision    recall  f1-score   support

          CG       0.80      0.89      0.84      7032
          OR       0.88      0.78      0.83      7119

    accuracy                           0.84     14151
   macro avg       0.84      0.84      0.84     14151
weighted avg       0.84      0.84      0.84     14151
```

```
Classification Report:              precision    recall  f1-score   support

          CG       0.89      0.87      0.88      7032
          OR       0.87      0.89      0.88      7119

    accuracy                           0.88     14151
   macro avg       0.88      0.88      0.88     14151
weighted avg       0.88      0.88      0.88     14151
```

```
Classification Report:              precision    recall  f1-score   support

          CG       0.86      0.85      0.86      7032
          OR       0.86      0.87      0.86      7119

    accuracy                           0.86     14151
   macro avg       0.86      0.86      0.86     14151
weighted avg       0.86      0.86      0.86     14151

Confusion Matrix: [[5996 1036]
 [ 938 6181]]
Accuracy Score: 0.8605045579817681
Model Prediction Accuracy: 86.05%
```

7.CONCLUSION

The fake review detection is designed for filtering the fake reviews. In this research work SVM classification provided a better accuracy of classifying than the Naïve Bayes classifier for testing dataset. On the other hand, the Naïve Bayes classifier has performed better than other algorithms on the training data. Revealing that it can generalize better and predict the fake reviews efficiently.This method can be applied over other sampled instances of the dataset.The data visualization helped in exploring the dataset and the features identified contributed to the accuracy of the classification. The various algorithms used, and their accuracies show how each of them have performed based on their accuracy factors.
Also, the approach provides the user with a functionality to recommend the most truthful reviews to enable the purchaser to make decisions about the product.Various factors such as adding new vectors like ratings, emojis, verified purchase have affected the accuracy of classifying the data better.

REFERENCE

• 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.

• N. O'Brien, "Machine Learning for Detection of Fake News," [Online]. Available: https://dspace.mit.edu/bitstream/handle/1721.1/119727/1078649610-MIT.pdf [Accessed: November 2018].

• 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.

• 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.

• A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," in Proceedings of AAAI-98 Workshop on Learning for Text Categorization, Pittsburgh, PA, USA, vol. 752, no. 1, pp. 41-48, July 1998.

• 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].

• C. Hill, "10 Secrets to Uncovering which Online Reviews are Fake," [Online]. Available: https://www.marketwatch.com/story/10-secrets-to-uncovering-which-online-reviews- are-fake-2018-09-21 [Accessed: March 2019].

• J. Novak, "List archive Emojis," [Online]. Available: https://li.st/jesseno/positive-negative-and-neutral-emojis-6EGfnd2QhBsa3t6Gp0FRP9 [Accessed: June 2019].

• P. K. Novak, J. Smailović, B. Sluban and I. Mozeti, "Sentiment of Emojis," Journal of Computation and Language, vol.10, no. 12, pp. 1-4, December 2015.