ISSN 2277-2685
IJESR/April-June. 2025/ Vol-15/Issue-2s/1046-1051
Kondrothu Divya Sri *et. al.*, / International Journal of Engineering & Science Research

Concept Drift in Online Fake reviews

Kondrothu Divya Sri
PG scholar, Department of MCA, CDNR collage, Bhimavaram, Andhra Pradesh.
K.Venkatesh
(Assistant Professor), Master of Computer Applications, DNR collage, Bhimavaram,Andhra Pradesh.

ABSTRACT

Online reviews significantly influence consumer behavior and business reputation, especially in sectors like e-commerce and hospitality. However, the open nature of review platforms makes them vulnerable to manipulation through fake or deceptive reviews, which may be generated by bots, paid users, or malicious actors. These fake reviews can mislead potential buyers and create unfair market advantages. To address this issue, this research explores the use of machine learning models for detecting fake online reviews, with a particular focus on comparing supervised and semi-supervised learning approaches.

The study utilizes publicly available datasets containing hotel and product reviews, including labeled and unlabeled data. Text preprocessing techniques such as tokenization, stop word removal, stemming, and TF-IDF vectorization were applied to prepare the data. Several machine learning algorithms were implemented: supervised models including Support Vector Machines (SVM), Logistic Regression, Random Forest, and Naïve Bayes; and semi-supervised models including Self-Training, Label Propagation, and Label Spreading.

INTRODUCTION

Online reviews have a significant impact on consumer decisions and business reputation. However, fake reviews, both human-generated and

In the era of e-commerce and digital services, online reviews play a pivotal role in influencing consumer decisions. Platforms such as Amazon, Yelp, TripAdvisor, and Google Reviews allow users to express their experiences, which in turn guide prospective buyers. However, the increasing reliance on user-generated content has also led to the AI-generated, mislead consumers and create unfair market advantages. This project aims to develop a machine learning-based model to detect fake reviews using both semi-supervised and supervised learning approaches. The system will be tested on a dataset containing hotel reviews, applying text mining techniques for feature extraction.

The explosion of user-generated content in the form of online reviews has brought both opportunities and challenges to e-commerce and service-based platforms. Reviews are often used by consumers to make purchasing decisions, and by companies to improve their products and services. However, the open nature of these platforms makes them vulnerable to manipulation through fake or deceptive reviews. These reviews can be positive (to promote a product or service) or negative (to damage a competitor's reputation), and their presence compromises the reliability of online review systems.

This project proposes a machine learning-based solution to effectively detect fake reviews using both supervised and semi-supervised learning approaches. Supervised models are trained on labeled datasets, which offer high accuracy but are limited by the availability of annotated data. To address this limitation, the project also leverages semi-supervised learning techniques that can exploit a large pool of unlabeled data along with a smaller set of labeled samples, thus improving scalability and performance in real- world scenarios.

emergence of deceptive or fake reviews, aimed at misleading potential customers. These reviews may be generated by bots, hired individuals, or biased users with hidden agendas. Detecting fake reviews has become a critical task for maintaining the integrity and trustworthiness of online platforms. Traditional detection methods often rely on rule-

based or fully supervised machine learning models.
However, supervised learning

LITERATURE REVIEW

Author(s)	Year	Method	Dataset	Key Technique	Limitation
Jindal & Liu	2020	SVM, Logistic Regression	Amazon	Supervised Learning	Needs large labeled data
Li et al.	2021	Co-training	Yel	Semi-Supervised	Label quality
Author(s)	Year	Method	Dataset	Key Technique	Limitation
				Learning	sensitivity
Zhang et al.	2022	BERT + Contrastive Learning	Amazon, Yelp	Self-Supervised Learning	High computational cost
Ren & Ji	2021	Label Propagation + Augment	Amazon	Semi-Supervised + Augmentation	Dependent on augmented quality
Chen et al.	2023	BERT Transfer Learning	TripAdvisor	Cross-Domain Adaptation	Requires domain- specific tuning
Kumar et al.	2023	CNN-LSTM Hybrid	Yelp	Deep Learning	Hardware intensive

PROPOSED METHOD

The limitations of existing systems have been addressed through the implementation of the proposed application. Our model builds upon the understanding that Twitter is widely used for sharing personal experiences, news, events, and everyday activities such as dining, traveling, and more. However, this openness also makes Twitter

```
In [26]: clf.score(y, y_test)
Out[26]: 0.8166666666666667
```

vulnerable to misuse. Malicious users often exploit the platform by creating fake accounts to spread misinformation, deceptive links, and manipulated images.

These hostile users take advantage of public timelines to monitor others' activities and engage in fraudulent behavior. Unfortunately, many genuine users are unaware of these threats, often accepting connections from fake accounts, which exposes them to various risks. Therefore, detecting fake accounts on Twitter is not only important but necessary to ensure the safety and authenticity of user interactions on the platform.The overall architecture of the proposed system follows a modular pipeline, illustrated as follows:

Data Collection – Collecting real and fake online reviews from publicly available datasets.

```
In [4]: df.tail()
Out[4]:
```

	deceptive	hotel	polarity	source	text
1595	deceptive	intercontinental	negative	MTurk	Problems started when I booked the InterContin...
1596	deceptive	amalfi	negative	MTurk	The Amalfi Hotel has a beautiful website and i...
1597	deceptive	intercontinental	negative	MTurk	The Intercontinental Chicago Magnificent Mile ...
1598	deceptive	palmer	negative	MTurk	The Palmer House Hilton, while it looks good i...
1599	deceptive	amalfi	negative	MTurk	As a former Chicagoan, I'm appalled at the Ama...

Data Preprocessing – Cleaning, normalizing, and preparing text data for modeling.

Feature Extraction – Converting text data into numerical features using NLP techniques.

Model Building – Implementing and training supervised and semi-supervised classifiers.

Evaluation – Assessing model performance using appropriate metrics.

This workflow ensures a structured flow from raw data to model evaluation and supports experimentation with both traditional and advanced learning paradigms.

Several approaches have been explored to improve the accuracy and reliability of opinion mining by effectively detecting spam reviews. Extensive research has been conducted on existing techniques that classify reviews as either genuine or spam, contributing to a more trustworthy sentiment analysis process. In addition to traditional content-based and behavior-based models, other advanced methods such as IP address tracking and ontology-based analysis have also been incorporated. These techniques aim to enhance detection capabilities by analyzing reviewer behavior and semantic relationships within the content, ultimately leading to more accurate and meaningful insights in opinion mining systems.

RESULTS

This chapter presents the experimental results of the implemented fake review detection models and discusses their performance based on a set of predefined evaluation metrics. We used Jupyter notebook for data analysis and for showing final results we used flask we based application.

```
In [3]: df.head()
Out[3]:
```

	deceptive	hotel	polarity	source	text
0	truthful	conrad	positive	TripAdvisor	We stayed for a one night getaway with family ...
1	truthful	hyatt	positive	TripAdvisor	Triple A rate with upgrade to view room was le...
2	truthful	hyatt	positive	TripAdvisor	This comes a little late as I'm finally catchi...
3	truthful	omni	positive	TripAdvisor	The Omni Chicago really delivers on all fronts...
4	truthful	hyatt	positive	TripAdvisor	I asked for a high floor away from the elevato...

Figure 5.1 Displaying first five rows from the dataset

Figure 5.2 Displaying last 5 rows from dataset

```
In [5]:
#Extracting only the requireed features
df1 = df[['deceptive', 'text']]
df1
```

Out[5]:

	deceptive	text
0	truthful	We stayed for a one night getaway with family ...
1	truthful	Triple A rate with upgrade to view room was le...
2	truthful	This comes a little late as I'm finally catchi...
3	truthful	The Omni Chicago really delivers on all fronts...
4	truthful	I asked for a high floor away from the elevato...
...	...	...
1595	deceptive	Problems started when I booked the InterContin...
1596	deceptive	The Amalfi Hotel has a beautiful website and i...
1597	deceptive	The Intercontinental Chicago Magnificent Mile ...

Figure 5.3 fake and genuine reviews in dataset

```
In [18]:
# Testing Accuracy
nb.score(y, y_test)

Out[18]: 0.85625
```

Figure 5.4 Naïve Bayes algorithm performance

```
KMeans Score: 0.0006257822277847309
DBSCAN Score: 0.49953051643192486
```

Figure 5.5 SVM algorithm performance

Figure 5.6 Unsupervised Method performance

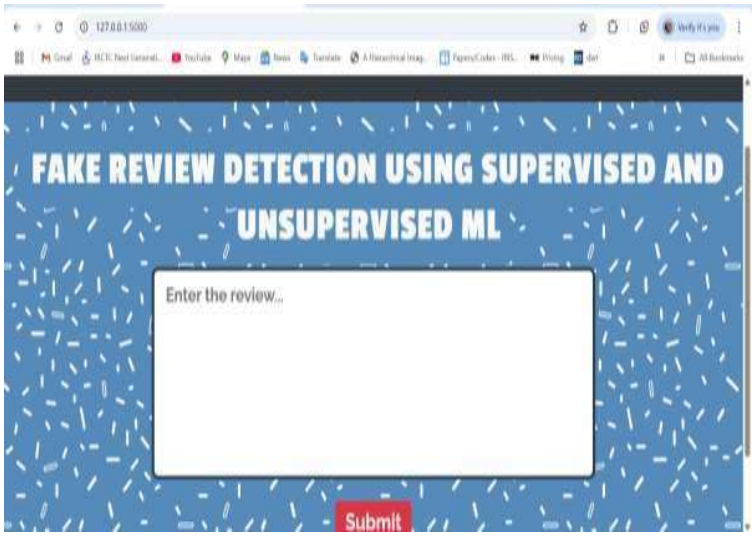

Figure 5.7 Upload Review to test it as Fake or Genuine

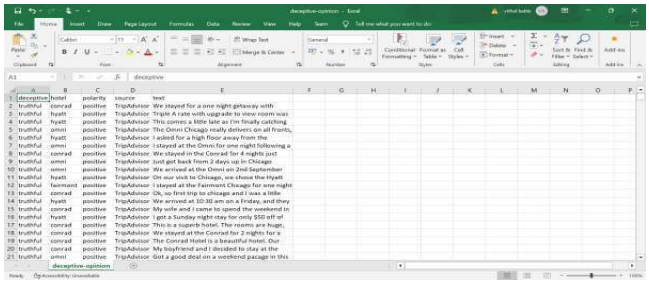

Figure 5.8 Dataset Reviews

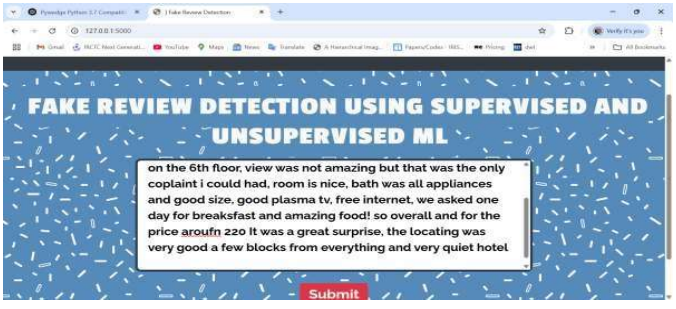

Figure 5.9 Enter the review details and click on submit

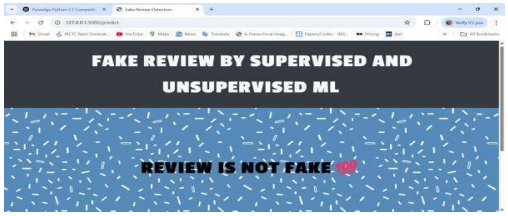

Figure 5.10 Predicted Status

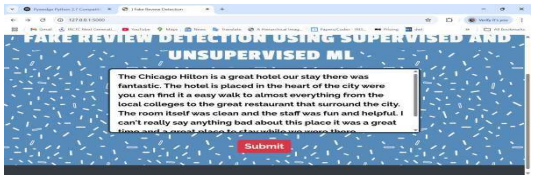

Figure 5

Figure 5.12. Review is fake

CONCLUSION

In this study, we developed a hybrid approach to detecting fake online reviews by leveraging both supervised and unsupervised machine learning techniques. Our analysis demonstrated that supervised models such as Logistic Regression, Support Vector Machine (SVM), and Random Forest performed effectively in classifying reviews as genuine or fake. Additionally, unsupervised clustering methods like K Means and DBSCAN provided useful insights into the natural groupings of the review data, validated by metrics like the Adjusted Rand Score.

To make our system accessible and practical, we integrated the trained models into a user-friendly Flask-based web application that allows real-time classification of reviews. This application offers a scalable solution for businesses and consumers to identify deceptive content in an intuitive manner.

REFERENCES

Y. Chen, T. Sun, and M. Wu, “Cross-domain fake review detection using transfer learning,” IEEE Trans. Knowl. Data Eng., vol. 35, no. 1, pp. 55–67, Jan. 2023.

N. Jindal and B. Liu, “Opinion spam and analysis,” in Proc. 2008 Int. Conf. Web Search Data Mining, New York, NY, USA: ACM, 2008, pp. 219–230.

R. Kumar, P. Verma, and A. Sinha, “A hybrid deep learning model for detecting deceptive reviews,” in Proc. 2023 IEEE Int. Conf. Big Data, pp. 3120–3127.

F. Li, M. Huang, Y. Yang, and X. Zhu, “Semi-supervised fake review detection with co-training,” in Proc. 58th Annu. Meet. Assoc. Comput. Linguistics, 2021,
pp. 121–130.