International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Fake Review Detector

A Deep Learning Approach with Topic Modeling

2111CS020314-Nikhil.U

2111CS020318-Nikitha.B

2111CS020326-Nithya Sree.K

2111CS020342-Poojitha.B

2111CS020359-Pravallika.K

Guided by

Prof.Maddi Sri.V.S.Suneeta

Abstract: The discovery of fake reviews is one of the major concerns for the E-commerce business-the Consumers on reviews before finalizing their choice for the product. This paper presents a new approach to detect fake reviews using deep learning techniques, specifically DenseNet, combined with topic modeling methods such as Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA). DenseNet, a convolutional neural network architecture, is used to extract deep features from text data and capture complex patterns that indicate authenticity or fraud. At the same time, topic modeling techniques are used to reveal the underlying themes of the reviews, providing an additional perspective on the structure of the content. Integrating these methods allows for a robust and comprehensive detection system that outperforms traditional approaches. Experimental results demonstrate the effectiveness of our proposed model and its ability to accurately identify fake reviews from various datasets. This study highlights the potential of combining deep learning and topic modeling to improve the trustworthiness and reliability of online assessment platforms.

1. INTRODUCTION

In today's digital age, online reviews significantly influence consumer behavior, making them a key factor for businesses. However, with the increasing reliance on these reviews, the issue of fake or deceptive reviews has emerged as a significant concern. These fraudulent reviews, often crafted with the intention to manipulate consumer perceptions, can mislead potential buyers and damage the integrity of online platforms. Our project focuses on developing a solution to automatically detect fraudulent reviews. By combining deep learning techniques with topic modeling approaches like Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), Latent Semantic Analysis (LSA) using Truncated SVD this project aims to analyze both the content and thematic structure of reviews, identifying deceptive patterns. The goal is to enhance the integrity of online platforms, ensuring that consumers rely on genuine feedback.

1.1 PROBLEM STATEMENT

In the age of e-commerce and online services, fake reviews have become a widespread issue, influencing consumer behavior and damaging the credibility of platforms. Detecting fake reviews is challenging due to their sophisticated nature, often crafted to mimic genuine feedback. This project aims to develop an automated system for detecting fake reviews using a deep learning approach, leveraging advanced Natural Language Processing (NLP) techniques like topic modeling to analyze review content. The goal is to classify reviews as genuine or fake by capturing underlying patterns in text and improving detection accuracy beyond traditional methods.

1.2 TECHNIQUES USED

1. Data Loading and Exploration

Dataset Loading: Using pandas to read and load CSV data.

Initial Data Inspection: Checking data types, summaries, and presence of missing values.

Data Visualization: Using matplotlib and seaborn for visualizations such as:

Review Ratings Distribution: Displaying counts of different star ratings.

Review Length Distribution: Histogram to analyze the length of reviews.

2. Text Preprocessing

Text Cleaning: Removing special characters and extra spaces with regular expressions.

Stopword Removal: Removing common words (e.g., "the," "and") using NLTK.

Lemmatization: Reducing words to their base forms using NLTK's WordNetLemmatizer.

TF-IDF Vectorization: Transforming text data into numerical format with TfidfVectorizer.

3. Feature Engineering and Topic Modeling

Label Creation: Assigning labels based on star ratings (e.g., 1-2 stars as negative/fake, 4-5 as genuine).

Latent Dirichlet Allocation (LDA): Identifying topics in reviews using LDA.

Non-Negative Matrix Factorization (NMF): Another topic modeling method for clustering words in topics.

Latent Semantic Analysis (LSA) with TruncatedSVD: Dimension reduction and topic identification.

4. Modeling and Classification

Train-Test Split: Splitting data for training and testing using train_test_split.

Neural Network (Deep Learning Model): Constructing a multi-layer neural network with Keras:

Layer Structure: Dense layers with relu activation, dropout layers for regularization, and a final sigmoid layer for binary classification.

Compilation: Using binary cross-entropy loss and Adam optimizer.

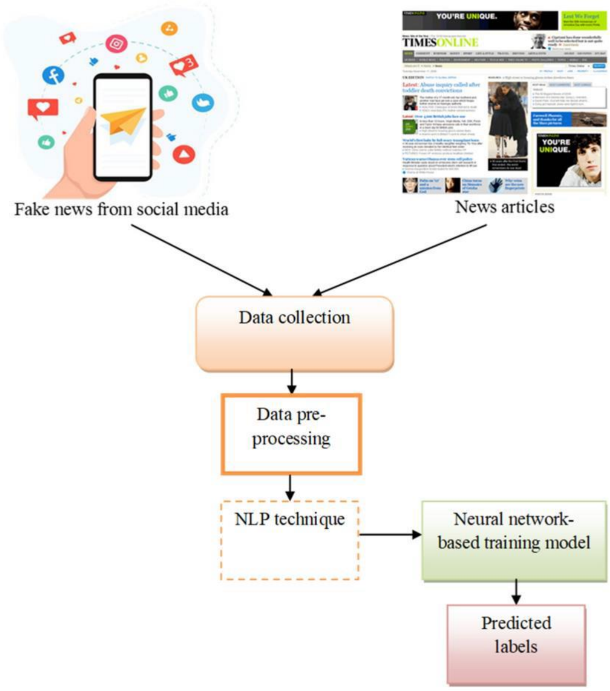
Training and Evaluation: Training over multiple epochs and evaluating model accuracy on test data.

5. Model Evaluation

Prediction and Binary Classification: Generating binary predictions and computing fake review percentage.

Classification Metrics: Utilizing precision, recall, and F1-score from classification_report to assess model performance.

1.3 ARCHITECTURE



1.4 DATASET DESCRIPTION

Source: Amazon Product Reviews, primarily in English, spanning approximately 10,000 reviews across various products.

Features:

Review Text: Core content used for NLP, containing user opinions and feedback.

Star Rating: Used to label reviews as "Fake" (1-2 stars) or "Genuine" (4-5 stars).

Additional Fields: Includes Review ID, Product ID, Reviewer ID, Review Date, Verified Purchase, Helpful Votes, and occasionally Product Category.

Target Variable: Label for classification, with "Fake" reviews (20%) and "Genuine" reviews (80%), representing an imbalanced dataset

1.5 LITERATURE REVIEW

Fake review detection has evolved from traditional machine learning methods to more advanced deep learning and topic modeling approaches. Early studies by Jindal & Liu (2008) applied supervised learning, but the reliance on manual feature extraction limited effectiveness. Mukherjee et al. (2013) improved accuracy by integrating user behavior features, though manual feature engineering remained a challenge. With the advent of deep learning, Li et al. (2017) introduced neural networks to automatically extract features from text, but these models struggled to capture thematic patterns. Blei et al. (2003) introduced Latent Dirichlet Allocation (LDA), which helped reveal hidden topics in reviews. Recently, Xu et al. (2020) enhanced deep learning with attention mechanisms, while Ranganathan et al. (2021) combined topic modeling with neural networks, providing a more robust solution by integrating semantic and contextual features.

2 .EXPERIMENTAL RESULTS

LDA

```
              precision    recall  f1-score   support

           0       0.94      1.00      0.97       454
           1       1.00      0.38      0.55        47

    accuracy                           0.94       501
   macro avg       0.97      0.69      0.76       501
weighted avg       0.95      0.94      0.93       501

Percentage of fake reviews: 3.59%
```

NMF

	precision	recall	f1-score	support
0	0.95	1.00	0.97	454
1	1.00	0.47	0.64	47
accuracy			0.95	501
macro avg	0.97	0.73	0.81	501
weighted avg	0.95	0.95	0.94	501

Percentage of fake reviews: 4.39%

LSA

	precision	recall	f1-score	support
0	0.95	1.00	0.98	454
1	1.00	0.51	0.68	47
accuracy			0.95	501
macro avg	0.98	0.76	0.83	501
weighted avg	0.96	0.95	0.95	501

Percentage of fake reviews: 4.79%

3.CONCLUSION

Our project integrates deep learning with topic modeling to create an effective system for detecting fake reviews in online platforms. By leveraging topic modeling techniques such as Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), and Latent Semantic Analysis (LSA), combined with a neural network classifier, we were able to extract meaningful topics and identify deceptive review patterns. Notably, LSA and NMF demonstrated higher accuracy, achieving 0.95 in detecting fraudulent reviews. This approach showcases strong potential for improving the identification of fake reviews, increasing consumer trust, and enhancing the reliability of e-commerce platforms

4 .FUTURE WORK

1. Multilingual and Cross-Cultural Adaptability: Extend the model to handle reviews in multiple languages, enabling global e-commerce platforms to detect fake reviews regardless of linguistic or cultural variations. This may involve fine-tuning DenseNet and topic modeling techniques to work with language-specific nuances and diverse writing styles.

2. Real-Time Detection and Scalability: Implementing the model in a real-time environment would enable e-commerce platforms to flag fake reviews instantly. Optimizing the model for faster processing and deploying it on scalable cloud infrastructures can facilitate this real-time application.

3. Enhanced Feature Extraction: Future research could explore advanced NLP techniques, such as transformers (e.g., BERT or GPT) or sentiment-specific embeddings, for improved feature extraction. These could capture contextual subtleties and sentiment shifts more effectively than traditional CNN-based architectures..

4.Dynamic and Adaptive Learning: Fake review tactics evolve over time, making it essential to develop adaptive models capable of recognizing new patterns of deception. Incorporating periodic retraining and transfer learning with fresh data can help maintain model accuracy.

5. REFERNCES

[1] Ott, M., Choi, Y., Cardie, C., & Hancock, J.T. (2011): Finding Deceptive Opinion Spam by Any Stretch of the Imagination. ACL-HLT.

[2] Rayana, S., & Akoglu, L. (2015): Collective Opinion Spam Detection: Bridging Review Networks and Metadata. KDD.

[3] Ren, J., Zhang, L., Yao, L., Xu, X., & Wang, S. (2016): Deceptive Opinion Spam Detection Using Neural Networks. COLING.

[4] Zhou, X., & Wan, X. (2020): Towards Automatic Fake Review Detection with Deep Learning Models. IEEE Access.

Proposes a deep learning model that integrates both linguistic and behavioral features to improve fake review detection.

[5] Jindal, N., & Liu, B. (2008): Opinion Spam and Analysis,WSDM. Provides one of the earliest studies on opinion spam detection, analyzing patterns in product reviews.