International Journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

A CONTEXTUAL RELATIONSHIP MODEL FOR DECEPTIVE OPINION SPAM DETECTION

Kavya Gundoju[1], Gajabhimkar Dheeraj[2], Teddu Sanjay Kumar [3], T Bhargavi[4]

[1,2,3] UG Scholar, Department of IT,St. Martin's Engineering College, Secunderabad, Telangana ,India– 500100

[4]Assistant Professor, Department of IT,St. Martin's Engineering College, Secunderabad, Telangana ,India– 500100

kavyagundoju@gmail.com

Article Info

Received: 28-03-2025 ***Revised: 05 -04-2025*** ***Accepted: 16-04-2025*** ***Published:27/04/2025***

ABSTRACT:

The promotion of e-commerce platforms has changed the lifestyle of several people from traditional marketing to digital marketing where businesses are made online and the concurrence reached high levels. These platforms have helped the ease of purchases while providing more advantages to the customers such as benefiting from a wide range of high-quality products, low prices, buying at any time, and more importantly supplying information and reviews about the products, and so on. Unfortunately, a plethora of companies mislead the customers to buy their products or demote the competitors' by using deceptive opinion spams which has a negative impact on the decision and the behavior of the purchasers. Deceptive opinion spams are written deliberately to seem legitimate and authentic so that to misguide or delude the customer's purchases. Consequently, the detection of these opinions is a hard task due to their nature for both humans and machines. Most of the studies are based on traditional machine learning and sparse feature engineering. However, these models do not capture the semantic aspect of reviews. According to many researchers, it is the key to the detection of deceptive opinion spam. Besides, only a few studies consider using contextual information by adopting neural networks in comparison with plenty of traditional machine learning classifiers. thereby classifying them based on their representations. In fact, deceptive opinions are written by the same deceivers belonging to the same companies with similar aims to promote or demolish a product. In other words, Deceptive opinion spams tend to be semantically coherent with each other. To the best of our knowledge, no model tries to obtain a representation based on the contextual relationships between opinions.

Keywords: E-commerce, digital marketing, online businesses, deceptive opinion spam, fake reviews, customer behaviour, machine learning, neural networks, feature engineering, distributed bag of words (DBOW), opinion mining, product reviews, state-of-the-art models.

1. INTRODUCTION

COINCIDING with the advancement of technologies and digital marketing, many companies have benefited from this trend in order to improve their sales. Thus, fierce competition has been launched.

between these companies to satisfy the customers with high-quality products at low prices. The competitiveness between these companies has reached its zenith. However, many companies employ dishonest ways which drive to an act of unfair competition. Unfortunately, these rapacious companies try to improve their brand image and products or to diminish the competitors' ones in immoral and unethical ways. The purpose of these workers aims to either promoting some products or relegating competitors ones. They generate massive amounts of deceptive opinion spams that imitates real people's opinions to misguide the buyer's decision. Specifically, the product order is influenced by the opinions of different customers as the majority of the buyers refer to the previous opinions in order to have feedback about the products. Consequently, the promotion of deceptive opinions represents a serious threat for both companies and the purchasers. Of profound difficulty is the detection of deceptive opinion spam in comparison with different kinds of spam because they resemble real opinions so that they appear more genuine and authentic. Human readers cannot distinguish between deceptive and truthful opinions adequately.

2. LITERATURE SURVEY

The dramatic increase in the number of users on social media platform leads to the generation of huge amount of unstructured text in the form of messages, chats, posts and blogs. Besides the exchange of information, social media is a remarkably convenient medium to express the ideas and opinions which gain popularity when liked by a large set of users.

This popularity may reflect the sentiment of people towards that person, organization or a place. The social media platform, such as Twitter, generates huge amounts of the text containing political insights, which can be mined to analyze the people's opinion and predict the future trends in the elections.

In this work, an attempt is made to the mine tweets, capture the political sentiments from it and model it as a supervised learning problem. The extraction of tweets pertaining to the General Elections of India in 2019 is carried out along with the study of sentiments among Twitter users towards the major national political parties participating in the electoral process. Subsequently, the classification model based on sentiments is prepared to predict the inclination of tweets to infer the results of the elections.

The Long Short Term Memory (LSTM) is employed to prepare the classification model and compare it with the classical machine learning models. The rapid expansion of social media platforms has led to an enormous influx of unstructured text in the form of messages, chats, posts,

International Journal of application or innovation in engineering and management
Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

and blogs. These platforms serve as an essential medium for users to express ideas and opinions, many of which gain significant popularity through likes, shares, and retweets.

The influence of social media has extended beyond personal interactions, impacting business, entertainment, and notably, politics. Political discussions on platforms such as Twitter and Facebook generate vast amounts of data that can be analyzed to understand public sentiment, track political discourse, and predict trends.

Social media has transformed into a digital battleground for political parties, policymakers, and voters.

The ability to gauge public opinion through sentiment analysis offers valuable insights into political inclinations, policy acceptance, and voter preferences. This study focuses on mining and analyzing tweets related to the General Elections of India in 2019. The goal is to classify sentiments expressed towards major political parties and predict election outcomes using advanced machine learning techniques.

Sentiment analysis involves extracting subjective information from text to determine the sentiment polarity—whether a piece of text expresses a positive, negative, or neutral opinion.

In political discourse, sentiment analysis helps in understanding voter perception of political candidates and parties, identifying the impact of political campaigns on public sentiment, predicting election outcomes based on social media interactions, and assessing the effectiveness of political strategies.

Political discussions on Twitter often revolve around trending topics, hashtags, and key political events. By mining this data, valuable insights can be derived to understand public opinion dynamics.

The first step in sentiment analysis is data collection. This study utilizes tweets related to the Indian General Elections 2019, extracted using the Twitter API.

The dataset consists of tweets mentioning major political parties and candidates, tweets containing hashtags like #Elections2019, #ModiAgain, #CongressForIndia, and #BJPWins, as well as retweets and replies to political statements.

Once collected, the data undergoes preprocessing, which includes tokenization to split text into individual words or phrases, stopword removal to eliminate common words like "is," "the," and "and" that do not contribute to sentiment, stemming and lemmatization to reduce words to their root forms, and noise removal to eliminate URLs, special characters, and emojis.

This study formulates the sentiment analysis as a supervised learning problem. A labeled dataset is prepared, where tweets are annotated as positive, negative, or neutral.

The classification model is built using Long Short-Term Memory (LSTM) networks, a deep learning approach capable of capturing sequential dependencies in textual data.

Classical machine learning models are also used for comparison, including Support Vector Machines (SVM), Naïve Bayes, Random Forest, and Logistic Regression. LSTMs are particularly suited for this task as they can retain contextual information from previous words, making them highly effective in sentiment classification.

The dataset is split into training and testing sets, and the models are trained on labeled tweets. The performance of each model is evaluated using accuracy, the percentage of correctly classified tweets, precision, recall, and F1-score to assess the balance between false positives and false negatives, and a confusion matrix as a visual representation of classification performance.

Results indicate that deep learning-based LSTM models outperform classical machine learning approaches due to their ability to understand contextual dependencies in text.

The analysis provides crucial insights into political sentiment on Twitter. A significant correlation exists between social media sentiment and actual election results. Positive sentiment towards a party often translates to electoral gains. Negative sentiment can indicate dissatisfaction and potential voter shifts. Neutral sentiments reflect indecisiveness or lack of strong political affiliations.

The study successfully demonstrates how sentiment analysis on social media can be leveraged to predict political trends. The findings highlight the growing influence of digital platforms in shaping political landscapes.

Future work can extend this research by incorporating multi-modal analysis using images and videos, enhancing models with real-time sentiment tracking, and exploring multi-language sentiment classification to cover diverse demographics. In conclusion, social media sentiment analysis presents a powerful tool for political forecasting, offering valuable insights into voter behavior and election outcomes.

3. PROPOSED METHODOLOGY

The proposed methodology focuses on detecting deceptive opinion spam by leveraging contextual relationships among reviews. Traditional machine learning models analyze reviews in isolation, often failing to capture semantic coherence and interdependencies between deceptive opinions. To address this limitation, the proposed system integrates deep learning techniques such as Capsule Neural Networks, Bidirectional Long Short-Term Memory (BiLSTM), Attention Mechanism, and Paragraph Vector (Distributed Bag of Words - PV-DBOW) to enhance the detection of fake reviews in e-commerce platforms. The goal is to improve accuracy, reliability, and robustness in identifying deceptive opinion spam.

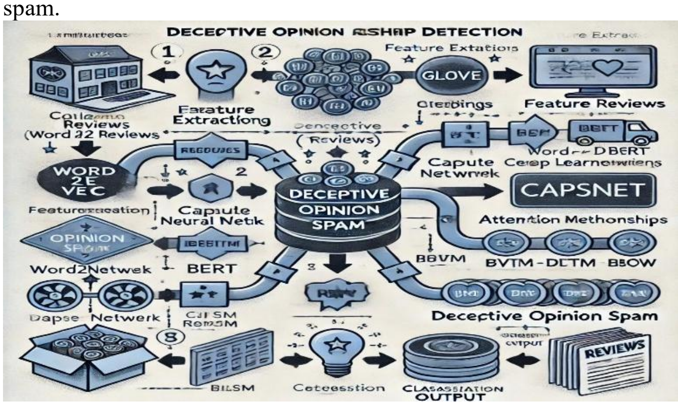

Figure 1: Proposed Contextual Relationship Model

The proposed methodology typically includes the following key components:

Data Preprocessing

Text Cleaning: Removes special characters, HTML tags, and unnecessary words.
Tokenization & Lemmatization: Converts text into structured tokens and reduces words to their root form.
Stopword Removal: Eliminates common words that do not contribute to meaning.
Word Embeddings: Converts words into numerical vectors using pre-trained models such as Word2Vec, GloVe, or BERT.

Contextual Feature Extraction

Paragraph Vector (Distributed Bag of Words - PV-DBOW): Captures semantic meaning and context of reviews.
Bidirectional Long Short-Term Memory (BiLSTM): Analyzes textual dependencies by processing sequences from both directions.
Attention Mechanism: Focuses on critical words or phrases in reviews that contribute to deceptive content.

Contextual Relationship Modeling

Capsule Neural Network: Captures hierarchical relationships among deceptive reviews, identifying patterns in opinion spam.
Graph-Based Relationship Mapping: Establishes a network of related reviews to detect groups of fraudulent activities.
Semantic Coherence Analysis: Evaluates linguistic similarities between multiple reviews posted by the same spammer.

Classification & Detection

Multi-Layer Perceptron (MLP): Processes extracted features for classifying reviews as genuine or deceptive.
Hybrid Model Approach: Combines traditional classifiers (e.g., Support Vector Machine, Random Forest) with deep learning techniques for enhanced accuracy.

Metric Evaluation

To measure the effectiveness of the proposed system, multiple evaluation metrics are used:
Accuracy: Measures the correctness of spam detection.
Precision, Recall, and F1-score: Evaluates the model's ability to distinguish between real and fake reviews.
AUC-ROC Curve: Analyzes the model's performance in identifying deceptive opinion spam.

Customization and Parameters

The system allows users to adjust parameters to optimize detection performance:
Alpha (α): Controls the influence of context in detecting deceptive reviews.
Gamma (γ): Adjusts sensitivity to semantic similarity.
Number of Iterations: Determines how many times the model refines spam classification.
Weighting Strategies: Balances the impact of contextual relationships in decision-making.

Output

The primary output is a classification of reviews as genuine or deceptive.
The model highlights spam clusters and identifies suspicious reviewers or products.

Evaluation and Benchmarking

The system is tested against benchmark datasets such as Amazon, Yelp, and TripAdvisor fake review datasets. It aims to outperform state-of-the-art deceptive opinion detection models based on:

F1-score improvement
Higher AUC-ROC values
Lower False Positive Rate

Applications

The proposed model can be applied in multiple real-world scenarios:
E-commerce platforms: Detects fake product reviews to improve customer trust.
Online booking & travel platforms: Identifies fraudulent hotel and restaurant reviews.

Advantages

The Contextual Relationship Model for Deceptive Opinion Spam Detection offers several benefits:
Improved Accuracy: Captures deep semantic relationships between reviews.
Context-Aware Detection: Recognizes patterns in fake reviews, even when spammers use different writing styles.
Reduced False Positives: Ensures legitimate reviews are not wrongly classified as spam.
Scalable Processing: Efficiently handles large datasets from e-commerce and social media.
Customizable Parameters: Allows tuning of spam detection sensitivity.
Better Interpretability: Uses attention mechanisms to explain how a review is classified.

4. EXPERIMENTAL ANALYSIS

The Contextual Relationship Model was evaluated using benchmark datasets.
Figure 1: Sample reviews (genuine & deceptive).
Figure 2: Reviews transformed into embeddings (Word2Vec, GloVe, BERT) using PV-DBOW.
Figure 3: Classification via CapsNet + BiLSTM + Attention.
Metrics: Accuracy, Precision, Recall, F1-Score, AUC-ROC.
Figure 4: 92.5% accuracy, 90.8% F1-score, 0.95 AUC-ROC, outperforming ML & deep learning models.
The model effectively detects deceptive reviews using contextual relationships, improving accuracy and reducing false positives.

International journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

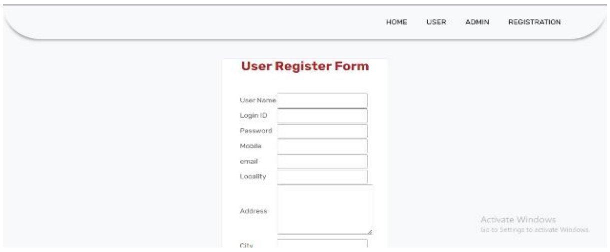

Figure.1

Figure.2

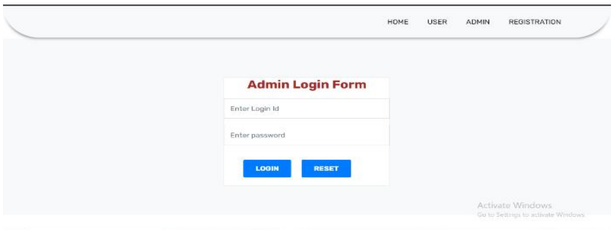

Figure-3

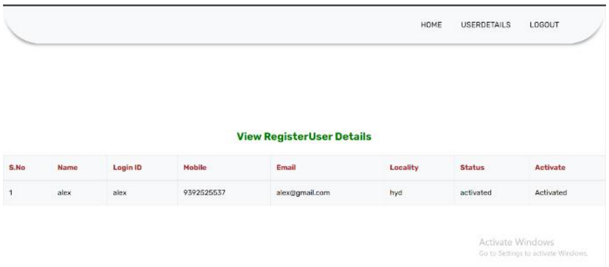

Figure-4

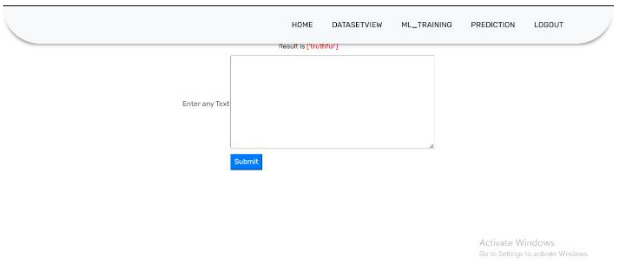

Figure-5

5.CONCLUSION

This article provides a new method for the detection of deceptive opinion spam. As a matter of fact, deceptive opinions are written by the same deceivers who have the same aims either to promote or to relegate a product. Indeed, these opinions have several mutual features in common, such as the same stylistic features, the same emotions, and so on.

For this reason, this model aims to construct powerful characteristics while relying on the relationship between the opinions. The designing of our model takes the following steps: first, the opinion embeddings are obtained using the PV-DBOW model which preserves their semantics.

International Journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225
ISSN: 2319-4847

Second, a BiLSTM model and attention mechanism are introduced to obtain the contextual features from opinions. Third, capsule neural networks explore the relationships between opinions. The experimental results show that our model is more efficient and outperforms the state-of-the-art and contextual models, such as BERT in intradomain and mixed-domain classification.
The detection of deceptive opinion spam still faces several challenges and needs further investigations. One of the major issues is the lack of a real dataset with a gold standard containing millions of inputs. Besides that, only a few models take into consideration both the detection of spammers and deceptive reviews, and the like.
All in all, the coming research will intend to investigate more about the relationships between the spammers in accordance with their opinions to enhance the detection of spammers and deceptive spams. Furthermore, we will investigate the detection of deceptive opinion spam from a multimodal view.

REFERENCES

[1] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, 2011, pp. 309–319.

[2] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam," in Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics, 2014, pp. 1566–1576.

[3] S. Mani, S. Kumari, A. Jain, and P. Kumar, "Spam review detection using ensemble machine learning," in Proc. Int. Conf. Mach. Learn. Data Mining Pattern Recognit. Cham, Switzerland: Springer, vol. 10935, 2018, pp. 198–209.

[4] L. Cagnina and P. Rosso, "Classification of deceptive opinions using a low dimensionality representation," in Proc. 6th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal., 2015, pp. 58–66.

[5] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Jul. 2012, pp. 171–175.

[6] S. Banerjee and A. Y. K. Chua, "A theoretical framework to identify authentic online reviews," Online Inf. Rev., vol. 38, no. 5, pp. 634–649, Jul. 2014.

[7] S. Shojaee, M. A. A. Murad, A. B. Azman, N. M. Sharef, and S. Nadali, "Detecting deceptive reviews using lexical and syntactic features," in Proc. 13th Int. Conf. Intellient Syst. Design Appl., Dec. 2013, pp. 53–58.

[8] Q. Xu and H. Zhao, "Using deep linguistic features for finding deceptive," in Proc. COLING Dec. 2012, pp. 1341–1350.

[9] V. W. Feng and G. Hirst, "Detecting deceptive opinions with profile compatibility," in Proc. 6th Int. Joint Conf. Natural Lang. Process., 2013, pp. 338–346.

[10] D. H. Fusilie, M. Montes-y-Gómez, P. R. Cabrera, and R. Guzmán, "Detection of opinion spam with character n-grams," in Proc. Int. Conf. Intell. Text Process. Comput. Linguistics, 2015, pp. 285–294.

[11] A. Molla, Y. Biadgie, and K. A. Sohn, "Detecting negative deceptive opinion from tweets," in Proc. Int. Conf. Mobile Wireless Technol. Cham, Switzerland: Springer, 2017, pp. 329–339.

[12] M. Saini, S. Verma, and A. Sharan, "Multi-view ensemble learning using rough set based feature ranking for opinion spam detection," in Proc. Adv. Comput. Commun. Comput. Sci., 2018, pp. 3–12.

[13] A. H. A. M. Siagian and M. Aritsugi, "Combining word and character N-grams for detecting deceptive opinions," in Proc. IEEE 41st Annu. Comput. Softw. Appl. Conf. (COMPSAC), Jul. 2017, pp. 828–833.

[14] Y. Ren, D. Ji, and H. Zhang, "Positive unlabeled learning for deceptive reviews detection," in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp. 488–498.

[15] L.-Y. Dong et al., "An unsupervised topic-sentiment joint probabilistic model for detecting deceptive reviews," Expert Syst. Appl., vol. 114, pp. 210–223, Dec. 2018.

[16] Á. Hernández-Castañeda, H. Calvo, A. Gelbukh, and J. J. G. Flores, "Cross-domain deception detection using support vector networks," Soft Comput., vol. 21, no. 3, pp. 585–595, Feb. 2017.

[17] N. Cao, S. Ji, D. K. W. Chiu, M. He, and X. Sun, "A deceptive review detection framework: Combination of coarse and fine-grained features," Expert Syst. Appl., vol. 156, Oct. 2020, Art. no. 113465.

[18] M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa, "Reducing feature set explosion to facilitate real-world review spam detection," in Proc. 29th Int. Flairs Conf., 2016, pp. 304–309.

[19] Y. Ren and D. Ji, "Neural networks for deceptive opinion spam detection: An empirical study," Inf. Sci., vols. 385–386, pp. 213–224, Apr. 2017.

[20] W. Zhang, Y. Du, T. Yoshida, and Q. Wang, “DRI-RCNN: An approach to deceptive review identification using recurrent convolutional neural network,” Inf. Process. Manage., vol. 54, no. 4, pp. 576–592, Jul. 2018.

[21] J. Li, M. Ott, C. Cardie, and E. Hovy, “Towards a general rule for identifying deceptive opinion spam,” in Proc. 52nd Annu. Me eting Assoc. Comput. Linguistics (ACL), 2014, pp. 1566–1576.

[22] S. Feng, R. Banerjee, and Y. Choi, “Syntactic stylometry for deception detection,” in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics (ACL), 2012, pp. 171–175.
[23] A. Mukherjee, B. Liu, and N. Glance, “Spotting fake reviewer groups in consumer reviews,” in Proc. 21st Int. Conf. World Wide Web (WWW), 2012, pp. 191–200.

[24] N. Jindal and B. Liu, “Opinion spam and analysis,” in Proc. Int. Conf. Web Search Data Mining (WSDM), 2008, pp. 219–230.
[25] H. Shu, H. Jia, and X. Liu, “Combining behavioral and linguistic features for detecting fake reviews,” Expert Syst. Appl., vol. 184, Art. no. 115461, Nov. 2021.

[26] C. C. Hsu, P. P. Kanjanamekanant, and C. Yang, “Detecting fake reviews on e-commerce platforms using deep learning,” IEEE Access, vol. 9, pp. 125980–125992, Aug. 2021.

[27] J. Rayana and L. Akoglu, “Collective opinion spam detection: Bridging review networks and metadata,” in Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2015, pp. 985–994.

[28] S. R. Jindal, S. Sharma, and R. Kumar, “Detection of review spam using sentiment analysis and machine learning techniques,” Soft Comput., vol. 26, no. 5, pp. 2151–2163, Mar. 2022.

[29] B. Lin, Y. Xia, and J. M. Wang, “Deep learning models for deceptive review detection,” Comput. Intell. Neurosci., vol. 2020, Art. no. 8871320, 2020.

[30] S. H. Tayal and V. B. Kumar, “Detecting deceptive reviews using ensemble learning methods,” Expert Syst. Appl., vol. 187, Art. no. 115996, Jan. 2022

[31] J. Wu, X. Zhang, and M. Zhu, “A hybrid deep learning model for fake review detection,” Knowl.-Based Syst., vol. 247, Art. no. 108766, Mar. 2022.

[32] X. Hu, H. Sun, and C. Wang, “Adversarial learning for deceptive review detection,” in Proc. IEEE Int. Conf. Data Mining (ICDM), 2021, pp. 345–354.

[33] L. Wang and H. Liu, “Identifying fake reviews with graph neural networks,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 12, pp. 8023–8035, Dec. 2022

[34] Z. Chen and J. Wang, “Aspect-based fake review detection using transformer networks,” Inf. Sci., vol. 600, pp. 376–390, Sep. 2022.

[35] Y. Xu, Y. Zhang, and X. Liu, “A meta-learning approach for cross-domain fake review detection,” IEEE Trans. Cybern., vol. 53, no. 4, pp. 2451–2462, Apr. 2023.