2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT) | 979-8-3503-9648-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/InCACCT57535.2023.10141826

Deep Learning Based Model for Fake Review Detection

Digvijay Singh
Research Scholar, Dept of CSE, Uttaranchal University, Dehradun, Uttarakhand, India
digvijay.dbgi@gmail.com

Minakshi Memoria
Department of Computer Science and Engineering Uttaranchal Instituteof Technology, Uttaranchal University, Dehradun, India
minakshimemoria@gmail.com

Rajiv Kumar
Department of Computer Science and Engineering Uttaranchal Instituteof Technology, Uttaranchal University, Dehradun, India
rajiv.gill1@gmail.com

Abstract— **In present time, peoples are more inclined towards the e-commerce for their purchases and their choices are much influenced by the reviews available over there as review plays an important role in making their decision. If the reviews are more positive the possibility to buy the product is comparatively high. Here, the necessity arrives to develop a sustainable approach for the detection of malicious reviews to save the customers from the fraud. There are many sites or agencies are available which are hired by the merchandise to generate the positive reviews for them to increase their sales or damage the competitor's product sales. Deep learning methodologies for malicious review detection includes, Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are proposed in this paper. We have also compared the performance of these methods with state of arts techniques such as Naive Bayes (NB), K Nearest Neighbour (KNN) and Support Vector Machine (SVM) for the detection of fake reviews and ultimately, its efficiency is illustrated for both the traditional and the deep learning classifiers.**

Keywords— Fake reviews, Spam review detection, Deep learning, CNN, KNN, LSTM, SVM.

I. INTRODUCTION

During past few years the connectivity with the world wide web has become an important aspect of our day to day routine. People communicate and express their views or opinion in the form of blogs, ratings, reviews or posts on various online platforms like social media sites, online trading websites, virtual commerce, blogs, or review sites. Also, in present, peoples are more attracted towards the e-commerce for their purchases where this shared opinion plays a very important role and considered as a trusted information [2] that's why a customer believes that before making a purchasing decision, they should read the product reviews. If the reviews are more positive the possibility to buy the product is comparatively high [1]. According to the survey [3], less than 20% customers do not treat online reviews same as the personal recommendations. As a result, these reviews are considered as a fundamental component of business and a motivator for both customers and commercial organizations. But here a question arises "Are the opinions or reviews that people are expressing authentic or real?"

In fact, according to current statistics, one in three TripAdvisor reviews are false [4]. Additionally, because anybody may write reviews for free and anonymously, there have been instances of people ill-using this anonymity facility by publishing unreal or even false feedbacks and reviews with a motive to trick customers and gain significant commercial advantages [5]. These reviews are referred to as "opinion spam" or, more specifically, "false reviews," and the individuals who engage in these nefarious deeds are referred to as "opinion spammers". Several significant developments have been made to enhance automatic fake review identification over the past decade. In these techniques machine learning techniques have gained a good repute to automatically detect the fake reviews with a good accuracy. Generally, the fake negative reviews posted by the spammers is to defame the competitor product. But the reviews that include harsh comments, which accurately reflect the opinions of genuine customers, cannot be categorized as spam. Therefore, it has become crucial to identify legitimate evaluations from spam to make online reviews credible.

II. LITERATURE REVIEW

Still, recognition of the fake or spam reviews is considered as major challenge or concern for online shopping. The majority of fake review detection approaches extract useful information from the review text. Bag-of-words [12], psycholinguistic word lists, and part-of-speech tagging [13] are typical representations of these properties. In [14], aspect sentiment was employed to identify misleading users. Jindal and Liu [7] perform a study using logistic regression classifier to detect fake reviews of product based on fake reviewer's motive to replicate the product or service review. Persuaded by the reliability of reviewer a bidirectional NN with an attention mechanism is used by Liu et al. [8] to create the multimodal embedded representation of nodes in their suggested probabilistic graph classifier. To mine consumer reviews, Minqing Hu and Bing Liu [9] suggested a set of data mining and natural language processing- based approaches for summarising product reviews. Semantic clustering was used by the Wang P. et al. [10] by incorporating a new layer to the CNN architecture.

An enhanced four-layer OpCNN algorithm based on the Chinese word order problem was described by Zhao et al. [11]. Phrases including a specific word order are used as input for input layer. Authors optimized the OpCNN model parameters and employed the k-max pooling approach. Lin

ISBN: 979-8-3503-9648-5/23/$31.00 ©2023 IEEE

et. Al. [15] used sentimental classification algorithm, RNN and LSTM and show how LSTM could be utilized for solving the long-term reliance issue by adding a memory to the network that may have an impact on a document's meaning and polarity. A deep attention algorithm based on recurrent neural networks (RNN) was proposed by Chen T et al. [16] to selective learning of temporal characterization of sequential rumour detection reports. Examine more complex features from belief clustering output and user activity trends to enhance early detection efficiency. Tang et al. [17] conducted a sentiment classification experiment using four large datasets, including three Yelp.com restaurant review datasets and one IMDB dataset of movie reviews. When it comes to classifying reviews as positive or negative, the performance comparison demonstrates that LSTM performs better than other classifiers. Numerous research works had been developed using conventional techniques, but researchers are constantly working to increase the accuracy of detecting spam reviews.

III. PROPOSED SYSTEM

The major objective of our work is to identify spam textual reviews using deep learning techniques to improve the spam identification methodology with meaningful outcomes.

We have employed our suggested methodology for spam review identification in this part, which is depicted in figure 1. We used both labelled and unlabeled review data includes items like various words and phrases used in the reviews, numerous 1st personal pronouns identified, whether any linear feedback and rating system was use etc. In the initial analysis and data cleaning, any missing or superfluous data is removed by employing Natural Language Processes (NLP) tools. Through the Active Learning Algorithm moderately all the unlabeled data becomes labelled. Data transformation is the second stage of pre- processing that occurs after the data has been cleaned. After transformation feature identification and selection process was applied that ncludes TF- IDF, n-grams and Word2Vec techniques. For conventional ML algorithms, we used TF-IDF and n-grams techniques, and for deep learning approaches, we used Word Embeddings (Word2Vec) techniques for both CNN and LSTM to express texts as numeric value. Then, to categorize reviews and feedbacks as fake or spam and malicious, both conventional machine learning (SVM, KNN and NB) and deep learning classifiers (CNN and LSTM) are utilized. Finally, we carried out the comparison and results showed that deep learning methods performs better than tradition machine learning methods.

A. Exploratory Data Analysis

Exploratory Data Analysis and Evaluation (EDA) is initial stage of the followed process. We used two datasets in our study, first one is the gold- standard dataset developed by Ott et al. [18][25][26][27][28] and another from Yelp dataset. During active learning process, reviews of yelp dataset were labelled.

B. Data Pre-processing

Pre-processing is the transformation of data that has been gathered in its raw form before it is given to the algorithm for further processing. Natural Language Processing (NLP) operations like tokenization, punctuation removal, managing any missing data, stop-word removal, and stemming are performed as part of the pre-processing process.

We use the active learning method, also used by Istiaq et al. [19], for labelling the dataset of yelp.com. SVM decision function was used to select the unlabeled data samples and were trained using the highest and lowest average absolute confidence.

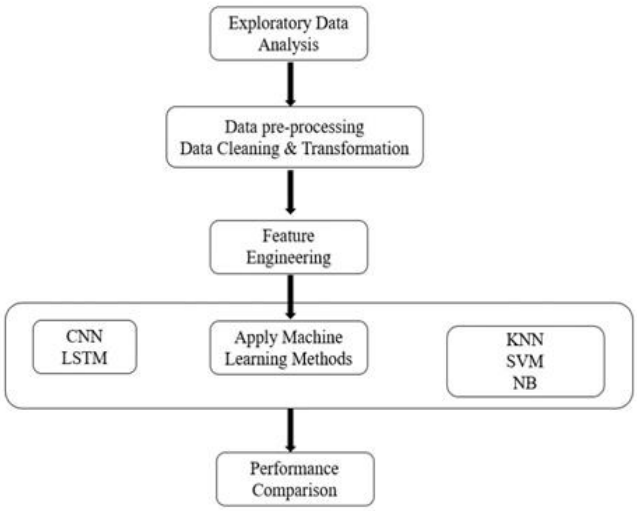

Fig. 1. Proposed module of Prediction Process

C. Feature Selection

All machine learning algorithms require some input data to generate desirable outcomes. These input data include a variety of attributes and are often presented as structured columns. The features in these input data must have some distinct qualities for the algorithm to work successfully. we used n-grams, TF-IDF and word embeddings feature selection techniques in our proposed model.

D. Prediction

The next stage after model training is to determine whether the model is correctly predicting and correlating with the actual values of the data set that was previously employed. The model is tested using the training datasets as the initial step in this stage, and the accuracy is assessed based on this. we have used conventional machine learning: SVM, KNN and NB; and deep learning classifiers: CNN and LSTM for predicting the fake review.

IV. EXPERIMENT RESULTS

This work has used the gold standard dataset known as Ott Dataset which can be found at [18] and from yelp.com. The Ott dataset has 1600 reviews out which 800 reviews are ham, and 800 reviews are spam. From yelp dataset [20] we used first 1600 reviews and after active learning, 400 reviews were labelled as spam and 1200 reviews were labelled as ham. We adjusted a few of the hyper-meters to train our model with CNN & LSTM, which improved performance. For the train- test splitting of our dataset, we employed the ratios 90:10, 80:20, 70:30, and 60:40. The starting weight value is 6, and the learning rate is .001. We

adjusted the hidden layer sizes to 50, 100, and 200. The batch size, epoch count, and drop out parameters were also adjusted.

Table 1. Performance of traditionalclassifiers over help dataset

Cross Validation	Classifier	Technique	Accuracy
10 - fold	SVM	Unigram	90.13%
5 - fold	KNN	Bi-grams	89.75%
10 - fold	NB	Unigram	90.85%

In experiment 2 CNN & LSTM methods were used over "Ott Dataset".

II. Result Analysis and Comparison:

Table 6 shows the outcomes of some previous research works, all of which used SVM, KNN, and NB conventional techniques.

Experiment 1 uses some conventional classifiers, including Naive Bayes (NB), K-Nearest Neighbour (KNN), and Support Vector Machine (SVM) To assess the performance of the "Yelp Dataset".

Table 2. Result of lstm over ott dataset

Train-to-Test Ratio	Dimensional embedding	Hidden dimension	Accuracy	Technique
90:10	100	200	92.15%	word2vec
80:20	100	200	92.39%	word2vec
70:30	100	50	93.89%	word2vec
60:40	200	50	92.35%	word2vec

Table 3. CNN result on ott dataset

Train-to-Test Ratio	Dimensional embedding	Accuracy	Technique
90:10	50	91.83%	word2vec
80:20	200	90.84%	word2vec
70:30	200	90.12%	word2vec
60:40	100	90.79%	word2vec

In experiment 3, we used CNN & LSTM methods over "Yelp Dataset".

Table 4. Lstm result on yelp dataset

Train-to-Test Ratio	Dimensional embedding	Hidden dimension	Accuracy	Technique
90:10	200	200	94.88%	word2vec
80:20	50	100	93.44%	word2vec
70:30	50	100	94.61%	word2vec
60:40	100	200	93.58%	word2vec

Table 5. Cnn outcome on the yelpdataset

Train-to-Test Ratio	Dimensional embedding	Accuracy	Technique
90:10	100	94.64%	word2vec
80:20	100	93.67%	word2vec
70:30	100	93.66%	word2vec
60:40	50	92.18%	word2vec

Table 6. Outcomes of some previousworks

Paper	Technique used	Classifier	Accuracy
[18]	Bigrams	SVM	89.60%
[21]	Unigram + Bigram	SVM	86%
[22]	features (n-gram)	SVM	86%
[23]	Unigram	KNN, NB, SVM, DT	82.00%
[24]	Unigrams, bigrams, trigrams and fourgrams	SVM	90.00%

The results demonstrate that our model's which uses deep learning techniques, produce outcomes that are more accurate.

V . CONCLUSION

In this paper, we showed the significance of reviews and how they alter almost every aspect of web-based information. As a result, detecting fake reviews is an active and ongoing area of investigation. In this study, Deep learning methodologies for malicious review detection includes, Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are proposed. We have also compared the performance of these methods with state of arts techniques such as Naive Bayes (NB), K Nearest Neighbour (KNN) and Support Vector Machine (SVM) for the detection of fake reviews. On comparing our results with some previous study results or outcomes, our model achieves the better accuracy.

VI. FUTURE WORKS

A lot of opportunities are available for the improvement of our work in the future. The size of the dataset can be increased. Deep learning methods can be used for data labelling to label the unlabeled data. reviewer and product-based features can also be included for better results. Other variation of deep learning methods like hybrid model can also be introduced.

REFERENCE

[1] N. Jindal and B. Liu, "Opinion spam and analysis," In Proceedings of the 2008 international conference on web search and data mining (pp. 219-230), 2008 February.

[2] J. K Rout, A. K Dash, and N. K Ray, "A framework for fake review detection: issues and challenges," In 2018 International Conference on Information Technology (ICIT) (pp. 7-10). IEEE, 2018, December.

[3] R. Kumar, M. Memoria, A. Gupta, and M. Awasthi, "Critical Analysis of Genetic Algorithm under Crossover and Mutation Rate," In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 976-980). IEEE, 2021, December

[4] Bright Local. *Local consumer review survey 2023*. Available at: https://www.brightlocal.com/research/local-consumer-review-survey/. Accessed on 8 Nov 2023.

[5] The Times. *A third of TripAdvisor reviews are fake' as cheats buy five stars*. Available at: https://www.thetimes.co.uk/article/hotel-and-caf-cheats-are-caught-trying-to-buy-tripadvisor-stars-027fbcwc8. Accessed on 22 Jan 2022.

[6] B. Liu, "Sentiment analysis and opinion mining. Synthesis lectures on human language technologies," 5(1), 1-167, 2012

[7] M. Ott, C. Cardie and J. T Hancock, "Negative deceptive opinion spam," In Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies (pp. 497-501), 2013, June.

[8] R. Kumar, "Efficient Genetic Operators Based on Permutation Encoding under OSPSP," Int. J. Latest Res. Sci. Technol., 1(1), 55-59, 2012.

[9] N. Jindal, and B. Liu, "Analyzing and detecting review spam," In Seventh IEEE international conference on data mining (ICDM 2007) (pp. 547-552). IEEE, 2007, October.

[10] Y. Liu, B. Pang and X. Wang, "Opinion spam detection by incorporating multimodal embedded representation into a probabilistic review graph," Neurocomputing, 366, 276-283, 2019

[11] H. Minqing and L. Bing, "Mining and summarizing customer reviews," Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004.

[12] P. Wang, J. Xu, B. Xu, C. Liu, H. Zhang, F. Wang and H. Hao, "Semantic clustering and convolutional neural network for short text categorization," In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers) (pp. 352-357), 2015, July.

[13] S. Zhao, Z. Xu, L. Liu, M. Guo, and J. Yun, "Towards accurate deceptive opinions detection based on word order-preserving CNN," Mathematical Problems in Engineering, 2018.

[14] N. A. Patel and R. Patel, "A survey on fake review detection using machine learning techniques," In 2018 4th International Conference on Computing Communication and Automation (ICCCA) (pp. 1-6). IEEE, 2018, December.

[15] N. A. Patel and R. Patel, "A survey on fake review detection using machine learning techniques," In 2018 4th International Conference on Computing Communication and Automation (ICCCA) (pp. 1-6). IEEE, 2018, December.

[16] Y. Liu, and B. Pang, "A unified framework for detecting author spamicity by modeling review deviation,"w Expert Systems with Applications, 112, 148-155, 2018.

[17] T. Lin, B. G Horne, P. Tino, and C. L Giles, "Learning long-term dependencies in NARX recurrent neural networks," IEEE Transactions on Neural Networks, 7(6), 1329-1338, 1996.

[18] T. Chen, X. Li, H. Yin, and J. Zhang, "Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection," In Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2018 Workshops, BDASC, BDM, ML4Cyber, PAISI, DaMEMO, Melbourne, VIC, Australia, June 3, 2018, Revised Selected Papers 22 (pp. 40-52). Springer International Publishing, 2018.

[19] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 1422-1432), 2015, September.

[20] E. R Kumar, and E. A Kaushik, "Premature convergence and genetic algorithm under operating system process scheduling problem," Journal of Global Research in Computer Science, 1(5), 2010.

[21] M. Ott, Y. Choi, C. Cardie, and J. T Hancock, "Finding deceptive opinion spam by any stretch of the imagination," arXiv preprint arXiv:1107.4557, 2011.

[22] D. Lin, Y. Matsumoto, and R. Mihalcea, "Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies," In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, June

[23] M. I Ahsan, T. Nahian, A. A Kafi, M. I Hossain and F. M Shah, "Review spam detection using active learning," In 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 1-7). IEEE, 2016, October.

[24] S. M Mohammad, and P. D Turney, "Crowdsourcing a word–emotion association lexicon," Computational intelligence, 29(3), 436-465, 2013.

[25] M. Ott, C. Cardie and J. T Hancock, "Negative deceptive opinion spam," In Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies (pp. 497-501), 2013, June.

[26] R. Kumar, S. Gill and A. Kaushik, "An impact of cross over operator on the performance of genetic algorithm under operating system process scheduling problem," In 2011 International Conference on Communication Systems and Network Technologies (pp. 704-708). IEEE, 2011, June.

[27] E. Elmurngi and A. Gherbi, "An empirical study on detecting fake reviews using machine learning techniques," In 2017 seventh international conference on innovative computing technology (INTECH) (pp. 107-114). IEEE, 2017, August.

[28] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification," Security and Privacy, 1(1), e9, 2018.