# Fraud detection in online product review system via heterogeneous graph Transformer

**P. Himani,[1] K. Mahesh[2], K Nageswarao[3], N. Madhav kumar[4], S. Sumanth kalyan[5]**

[1]Asst. Professor, Department of Computer Science and engineering

[2,3,4,5]Student, Department of Computer Science and engineering

[1,2,3,4,5]QIS College of Engineering & Technology

**Abstract**-Natural Language Processing (NLP) methods may be used to identify and remove fake reviews from a given dataset. In this article, two alternative machine learning (ML) models are used to train a false review dataset in order to predict how accurate the reviews in a given dataset are. Product reviews available online on various websites and apps are increasingly being used to fabricate customer reviews in the e-commerce business and on other platforms as well. Before making a purchase, the company's items were regarded as trustworthy. As a result, huge E-commerce companies like Flipkart, Amazon, and the like must deal with the issue of phoney reviews and spammers in order to avoid customers from losing faith in the platforms they use to purchase online. There are websites and apps with a few thousand users that can use this model to forecast the legitimacy of reviews so that website owners may take action. The Nave Bayes and random forest approaches are used to build this model. Using these models, it is possible to quickly determine the amount of spam reviews on a website or app. There must be a sophisticated algorithm that is trained on millions of reviews in order to combat spammers like this one. These models are trained using the "amazon Yelp dataset," which is a tiny dataset that may be expanded up to achieve great accuracy and flexibility.

**Keywords**— opini***on mining, sentiment analysis, text mining.***

## I. INTRODUCTION

More and more individuals are purchasing goods online and having them delivered to their doorsteps because of the increasing sophistication of online review posting. In order to avoid being deceived by phoney reviewers and spammers, purchasers must rely on the honest opinions of other customers when purchasing products over the internet, and this must be done as much as possible. Despite its simplicity, this task is exhausting and time-consuming, and must be done methodically in order to find the source of the issue. To solve this issue, a machine learning model that focuses on the review section may be trained to identify whether or not a given review is legitimate or spam. If you don't utilise the product, you may still be detected by this method.

It is possible to gain a favourable rating for a product by using a spam review or a different customer id. The usage of terms like "amazing," "so excellent," "great," and so on might be noted for further investigation. Due of their tendency to exaggerate the product's benefits or attempt to mimic real reviews by utilising the same phrases again and over again. As a result, spam filtering needs a large amount of data in order to train and be successful, as well as domain knowledge such as sarcastic words used by customers to express their dissatisfaction with the product. Such reviews are identified using an NLP approach rather than misclassification as in sentiment analysis. Data pre-processing is used to eliminate unnecessary or obsolete product reviews.

In order to create an online E-commerce industry where consumers can build trust in a platform where the products they purchase are genuine and feedbacks posted on these websites/applications are checked regularly by the company where the number of users is increasing day by day, companies like Twitter, WhatsApp and Facebook use sentiment analysis to check fake news, harmful/derogatory posts, and banning such users/apps. Parallel to that \sE-commerce (Flipkart, Amazon) industries, hotels booking \s(Trivago), logistics, tourism (Trip Advisor), job search \s(LinkedIn, Glass door), food (Swiggy, Zomato), etc. use \salgorithms to tackle fake reviews, spammers to deceive the \sconsumers in buying below average products/ services. Also, spammers like "not verified profiles" should be flagged so that people aren't concerned about them.

Time-consuming and ineffective: Manual labelling of the reviews is time-consuming and ineffective. As a result, labelling reviews and then predicting the label using a supervised learning model is not viable. According to Sunil Soumya, et al., it is difficult and time-consuming to label 2431 reviews manually for over eight weeks, hence automated labelling of reviews should be able to save time and energy. This practise is common in several businesses, where it is not easy to distinguish between a review that has been paid for and a review that has not. 30 to 40 percent of the reviews on Amazon's "Yelp" collection are fake. The process of choosing and training these models relies heavily on the process of feature selection. For the "Amazon's yelp" dataset, we compare two models to see which one performs better and if it is appropriate to implement these models into live applications. In the fake review data analysis, the RF model outperformed the Nave Bayes method by a wide margin. There is a reasonable discussion of the issue of detecting phoney reviews, as well as the legality and need of doing so. The goal is to find a suitable replacement.

The following is how the remainder of the article is laid out: For each of them, we've included a section that outlines the background work (Section II), methodology (Section III), and datasets (Section IV). Result and analysis are shown in Section V.

It comes to an end with Section VI, which sums up the findings and provides a look at the future.

## II. RELATEDWORKS

Sentiment analysis is a term used to describe the process of analysing the stated thoughts of people in the form of text, blogs, reviews, feedbacks, and so on.

SVM classifier was utilised in the exiting study to categorise tweets in two steps [1]; emoticons, smileys, and hashtags were also employed to classify labels into various attitudes [2].. Emoticons were utilised to train an SVM classifier by another researcher [3]. Methods based on a lexicon: Tweets are rated by how many good and negative words they include. Syntactic rules, such as [query] is pos-adj – Twitter feel. — Based on the classifier constructed on a training dataset of Twitter sentiment [14]. Fake review detection is a challenge that researchers have come up with solutions to solve. The accuracy of new models, such as the ICF++ model that incorporates honesty value, rose by 49% [7]. For the purpose of identifying/classifying bogus reviews and removing them, VADER and Polarity-based method was utilised to assign polarities of +1,- 1, and 0, as well as mark reviews as true, false, or suspicious [4]. All the methodologies and techniques utilised by researchers for sentiment analysis and detection of phoney reviews over the last decade have been assembled in a huge collection of extensive literature [5].

For this model, the approach produced an F1 score of 91 percent [8]. Using a singleton spam review linked temporal pattern, the identification of spam reviews in singleton reviews was followed [9]. As a result of the KL divergence algorithm's asymmetric properties, it is utilised to distinguish bogus reviews from the originals. [10]. In order to distinguish between spam and legitimate reviews, the review sequence employed feature extraction up to six times [6]. New concept time series prediction technique, which employs pattern recognition to identify suspicious time periods when spam review was written, was developed by another researcher [11]. A spam score was computed by using activeness, context similarity, and ratings of review behaviour. Deep neural networks were examined to learn how models behave in detecting spam opinions, and recurrent and convolution networks were also monitored to convert raw text into vectors that could be utilised to locate spam reviews [12].

## III. PROPOSED SYSTEM ARCHITECTURE

"amazon academic review" is utilised as a dataset, which includes user ids and many other information such as beneficial votes, ratings, and reviews. The parameters that will be of use in feature engineering may be accessed. Many real and fraudulent evaluations have been included into the dataset so that the model's correctness can be readily evaluated. 11,537 establishments are included in Yelp's data collection for the academic challenge. More than 43,000 Yelp users and 229,000 reviews are included in this dataset (www.yelp.com/dataset). There are a lot of different reviews and factors in the dataset, making it difficult to train any algorithm on it. The initial stage in evaluating any dataset is pre-processing, which removes extraneous characteristics, punctuation, stop words, missing words, duplicated words, etc. to clean the dataset for training. This guarantees that the model is properly trained. All of the techniques used to eliminate undesired data from a dataset are included in this function, which is also known as data cleaning. Finding the gaps and the relationships between the various qualities (columns) is a critical stage in coming to meaningful findings, therefore don't skip it. Corpora are built using the NLTK library's collection of words and phrases.

These functions are imported from OrderedDict: term frequency, tokenizer, stopwords Stopwords, which include words like is, then, to, why, and so on, are deleted from the English language.Frequency tracks how many times a term has been used and may be exploited by spammers to identify the spammer again and over again. There are several reviews in the dataset, therefore data sampling is necessary before it is given to the classifier.In order to reduce the classifier's workload, random sampling is used. The bogus reviews are authenticated using distinct labels, which are then concatenated into two columns and returned to the data frame.

This is the Nave Bayes algorithm. Using a Naive Bayes computation, a two-arrangement model was constructed to predict whether the survey's findings were favourable or negative. Given the class variable, a Naive Bayes classifier assumes that the estimate of a single component is independent of the estimation of any other element. It uses the information from the preparation to calculate the chance of each outcome based on the highlights. The Naive Bayes computation has the characteristic of harbouring doubts about the validity of the data. It assumes that the dataset's high points are all independent and of equal importance. Calculating probabilities for values that fall inside a certain range may be done using the classic Nave Bayes formulation (equations 1), (2), and (3), which are presented below (0, 1). There are many variables in this equation: the standard deviation, the number of observations (a), the number of observations (y), and the number of observations (yi).

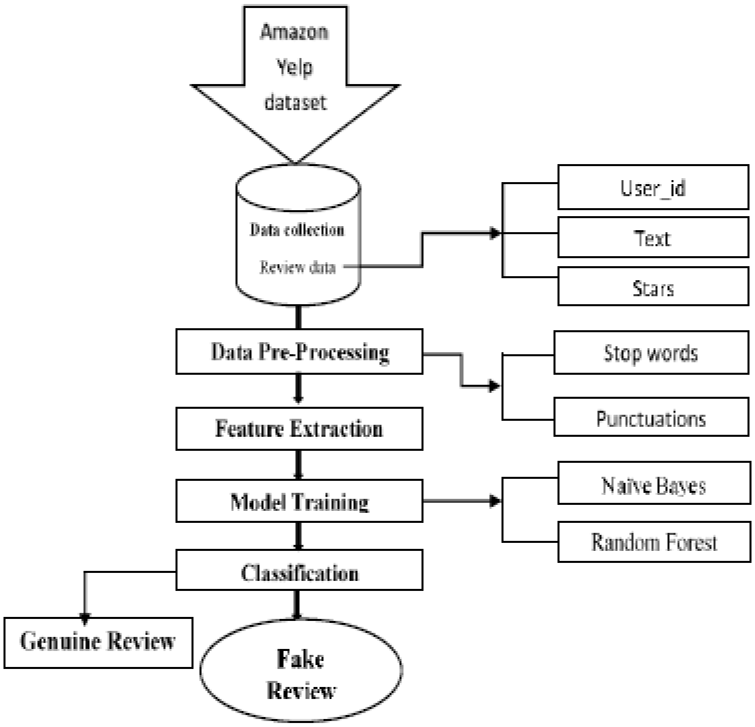


**Figure 1: Model diagram for fake review detection**

$$p\left(\frac{a}{b}\right) = \frac{p\left(\frac{b}{a}\right)p(a)}{p(b)} \qquad (1)$$

$$Posterior = \frac{prior * livelihood}{excellence} \qquad (2)$$

$$p\left(\frac{x_i}{y}\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2{}_y}}\right)\exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma^2{}_y}\right) \qquad (3)$$

Classifier: - Random forest Machine learning models are trained and tested using this supervised learning approach. Bagging is a decision tree training approach that produces a "forest" of decision trees. In this case, decision trees are integrated to improve the model's overall performance and learning ability. Multiple decision trees are combined to improve the random forest's performance and make better predictions [13].

a) Accuracy= TP+TN/ FP+FN+TN

b) Precision= TP/TP+FP

c) Recall (sensitivity) = TP/TP+FN

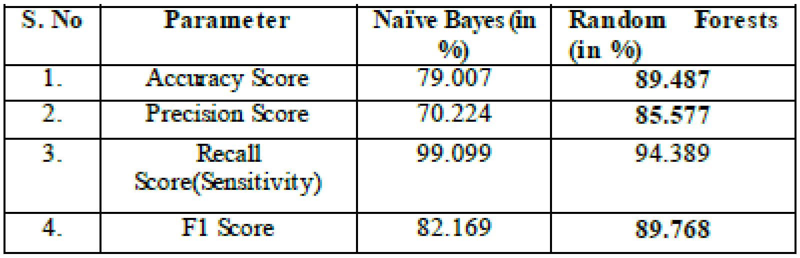
d) F1_score= 2*(Recall*Precision)/ (Recall + Precision)

Using these settings, the model's performance is given together with the corresponding confusion matrix. Figure 1's flowchart explains the issue solution process as follows: First, a dataset must be collected to determine if it is a binary or categorical dataset. I used the Yelp academic dataset review. Json file to load the reviews into the model's needed data format. In the end, only those characteristics were selected for further consideration that would be beneficial in future events, in order to save time. Attribute correlations are recorded in a feature extraction process and utilised to train Random forests and Naive Bayes for classification. Models that have been trained may then be supplied with additional data or test data in order to improve their accuracy and get better results from the training process, as shown in table 1. Confusion matrix, accuracy, precision, sensitivity, and F1 score are some of the measuring parameters of this model.

## IV. RESULTS AND DISCUSSION

Table 1 shows that the random forests classifier outperforms the other models, except when compared. As a result, random forests have a higher F1 accuracy, precision, and accuracy score. As a result, a random forest classifier may be used to detect and remove false product reviews. Because of the wide range of applications for which they may be used, they need considerable skill to get the most out of them.

*Table 1: Compiled Results of Both models*



| S. No | Parameter | Naïve Bayes (in %) | Random Forests (in %) |
|---|---|---|---|
| 1. | Accuracy Score | 79.007 | **89.487** |
| 2. | Precision Score | 70.224 | **85.577** |
| 3. | Recall Score(Sensitivity) | 99.099 | 94.389 |
| 4. | F1 Score | 82.169 | **89.768** |

## V. FUTURE SCOPE AND CONCLUSION

For the "Amazon's yelp" dataset, two models were constructed to explain the model performance and their relevance to deploy these models in real-time applications. Hence Compared to the Nave Bayes method, the random forests model outperformed the latter by a wide margin. Despite the fact that the issue of detecting and eliminating false reviews is a complex one, it is treated honestly and provides a clear understanding of its legality and need. In the future, hybrid and novel approaches for the identification of false reviews may be attempted. An NVIDIA graphics GPU coupled with a Google co-lab will speed up the study process.

## REFERENCES

[1] Barbosa, Luciano & Feng, Junlan. (2010). Robust Sentiment Detect ionon Twit ter from Biased and Noisy Data. Coling 2010 - 23rdInternational Conference on Computational Linguist ics, Proceedings ofthe Conference. 2. 36-44.

[2] Enhanced Sent iment Learning Using Twit ter Hashtags and SmileysDmit ry Davidov, Oren Tsur, ICNC / 2, Inst itute of Computer ScienceThe Hebrew University 2010.

[3] Go, Alec & Bhayani, Richa & Huang, Lei. (2009). Twit ter sent imentclassificat ion using distant supervision. Processing. 150.

[4] " Fake review det ect ion using opinion mining" by Dhairya P at el,Aishwerya Kapoor and SameetSonawane, International Research journalof Engineering and technology (IRJET) , volume 5, issue 12,Dec 2018.

[5] Ravi, k. Ravi., 2015. A survey on opinion mining and sent imentanalysis: Tasks, approaches and applications. Knowledge based systems,89.14-46.

[6] Khan, K. et al., "Mining opinion components from unstructured reviews:A review". Journal of King Saud Universit y – Computer andInformat ion Sciences (2014),ht tp://dx.doi.org/10.1016/j.jksuci.2014.03.009.

[7] " Fake review det ection from product review using modified met hod ofit erat ive comput ation framework", by EkaDyarWahyuni&ArifDjunaidy,MATEC web conferences 58.03003(2016) BISSTECH 2015.

[8] Saumya, S., Singh, J.P. Detection of spam reviews: a sent iment analysisapproach. CSIT 6, 137–148 (2018). https://doi.org/10.1007/s40012-018-0193-0.

[9] Xie S, Wang G, Lin S, Yu PS (2012) Review spam detect ion viatemporal pattern discovery. In: Proceedings of the 18th ACM SIGKDDinternat ional conference on Knowledge discovery and data mining.ACM, pp 823–831.

[10] Mukherjee A, Venkataraman V, Liu B, Glance N (2013a) Fake reviewdetect ion: classificat ion and analysis of real and pseudo reviews.Technical Report UIC-CS-2013–03, University of Illinois at Chicago.

[11] Heydari A, Tavakoli M, Salim N (2016) Detect ion of fake opinionsusing t ime series. Expert SystAppl 58:83–92.

[12] Ren Y, Ji D (2017) Neural networks for decept ive opinion spamdetect ion: an empirical study. InfSci 385:213–224.

[13] McCallum, Andrew. "Graphical Models, Lecture2: Bayesian NetworkRepresent ion" (PDF). Ret rieved 22 October 2019.

[14] Joseph, S. I. T. (2019). SURVEY OF DATA MINING ALGORITHM'SFOR INTELLIGENT COMPUTING SYSTEM. Journal of t rends inComputer Science and Smart technology (TCSST),1(01), 14-24.