Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, August : 2023

ONLINE REVIEW FRAUD DETECTION: SUPERVISED AND SEMI-SUPERVISED LEARNING

[#1]NEHA YASMEEN, ***MCA Student,***
[#2]B.SHARAN, ***Assistant Professor,***
[#3]Dr.V.BAPUJI, ***Associate Professor& HOD,***
Department of Master of Computer Application,
VAAGESWARI COLLEGE OF ENGINEERING, KARIMNAGAR, TELANGANA

ABSTRACT: With more and more people keeping an eye on social media, it's important to look at social data to figure out how people act. So, sentiment analysis is used to look at social data, especially Twitter Tweets, to see if the user's opinion about movie reviews is accurate. This study uses relevant keywords taken from social media and web reviews to build a full vocabulary and find hidden patterns of relationships. In recent years, there has been a big rise in the number of people who shop online. Online reviews have a big effect on how people decide what to buy when they shop online. Many people look at product or store reviews before deciding where to shop or what to buy. Because there are a lot of benefits to making fake reviews and doing other kinds of fraud, there has been a noticeable rise in the number of fake spam reviews on digital platforms that are used to review goods and services. Reviews that aren't true include those that are made up, reviews that aren't asked for, and reviews that aren't very good. Positive reviews of the product under consideration have the potential to attract more customers and boost sales, while negative reviews have the potential to lower demand and hurt sales. The above review is dishonest or fraudulent because it was written with the aim of tricking people or hurting the company's reputation by giving potential customers false information. The goal of our study is to find out how true the review is. In our work, we used three different classification algorithms: the Naive Bayes Classifier, the Logistic Regression, and the Support Vector Machines.'

Keywords: e-commerce, product recommender, product demographic, microblogs, recurrent neural

1. INTRODUCTION

Every day, there are new technological developments and advancements. Newer technology are always replacing older ones. People are able to do their jobs better with the help of this cutting-edge technology. One prominent manifestation of this technical development is the emergence of online markets. The internet marketplace allows us to make bookings and purchases at our convenience. Almost everyone reads product reviews before deciding whether or not to buy a given product or item. Because of this, it's possible that these reviews will have a major effect on the brands' reputations. The advertising and promotion of goods and services are also profoundly affected by these assessments. Therefore, there are more and more examples of fraudulent reviews on the web. People can post phony reviews to boost the reputation of their own items, which is harmful to the interests of legitimate businesses and consumers. Competitors' firms might also suffer damage to their reputations from fake bad reviews. Researchers looked explored several methods for identifying fabricated reviews on the web. Some depend on the nature of the review itself, while others are set by the author's choices and input during the review's creation. The text-based approach places more weight on the review's actual text or language than does the User behavior-based approach, which looks at factors like the reviewer's number of posts, location, and

Internet Protocol address. Supervised classification models are used in the great majority of the presented methods. Also, some studies have used semi-supervised models. Because the labels on the evaluations can't be trusted, they are being implemented. This research and application focus on supervised and semi-supervised classification methods for identifying fabricated online testimonials. The Expectation-Maximization algorithm is used in semi-supervised learning. Commonly used classifiers include Naive Bayes, SVM, and decision trees. In a content-based evaluation, that content is the main focus. We utilize review frequency, review length, and sentiment polarity as our attributes.

2. RELATED WORK

There are a plethora of strategies and methods published for spotting fake reviews. The following strategies have been demonstrated to be more effective at detecting fake online reviews empirically. The methods were divided into two categories. a) Content-based methodology focuses on dissecting and rating the evaluation's content as its first priority. The information provided is relevant to the review's language or claims made within it. In order to learn how to spot spam reviews, Heydari (2021) analyzed the language used in these ratings. Ott used three different approaches to classification. Text classification, genre identification, and the detection of psycholinguistic deception are the three approaches included in this investigation.

Ott uses the review's distribution of parts-of-speech (POS) to determine the genre it belongs to. Researchers utilized the frequency counts associated with POS tags as characteristics to evaluate classification efficiency.

The second method, known as psycholinguistic deception detection, employs the use of psycholinguistic interpretations to give weight to the most important aspects of a review. Review tools by Pennebaker were created in tandem with the LIWC (Linguistic Inquiry and Word Count) software.

Third, classifying texts: Ottl (year) established the idea of n-gram, a commonly used tool in the detection of dishonest assessments. In addition, other facets of language research are being conducted. Feng created phrase parse trees based on lexicalized and unlexicalized syntactic factors to detect fake reviews. The addition of deep syntactic information has been shown to improve prediction accuracy in experiments. It was found that there are a number of telltale signs of fraud that can be used to spot fake reviews. It was also found that the bag of words method is more effective when combined with broader features such as LIWC or POS, rather than used alone. Review meta-data, such as total number of words, publication date, and average rating, are also used as attributes by a number of researchers. Behaviorally-based approaches analyze reviews by looking at the reviewer's character traits. Lim (2021) looked into the difficulty of tracking down the real writers of spam reviews. Intentionally leaving false reviews is a very different user experience than for honest reviewers. The subsequent deceptive strategies used to manipulate reviews and ratings were uncovered by the researchers.

• The practice of passing extremely critical judgments As can be seen frequently in the actions of spammers, the number of fake reviews on the internet much outnumbers the number of real ones. Let's say that, on average, people rate a product a 9.0 out of 10. However, one critic gave it a perfect score. You may spot a spammer by looking through their review history and finding a consistent pattern of giving low ratings or leaving no ratings.

• Giving a high mark to a product made within one's own country. Some people give things from their home nation a positive review even when they know it isn't the best. The vast majority of posts in this subset of spam are responses to

requests for movie reviews. Now imagine that on a major international film site that mostly displays reviews from Indian reviewers, the film from India obtains a perfect score of 10.0. Examining the reviewers' email addresses is a surefire way to spot this kind of scam. • Creating product reviews for a wide range of consumer items: Every person has their own special set of hobbies and pursuits. In most cases, consumers show little inclination to buy everything presented to them. There is an assumption that a gamer doesn't care about or appreciate classic literature because of their hobby. However, if we see people routinely giving ratings that are much over the average range across multiple categories, we can infer that they are purposefully engaged in the falsification of their evaluations. The detection of fake reviews on the internet is commonly seen as a classification task, which is occasionally tackled by means of supervised text classification methods. Using large datasets labeled with examples from both the fake perspective (positive instances) and the actual opinion (negative examples) groups, these methods show good performance after extended training. Some studies have used semi-supervised classification methods. Helpfulness votes, rating-based behaviors, the use of seed words, and human observation are just a few of the methods that supervised classification systems rely on to build ground truth. The Product Word Composition Classifier (PWCC), the Trigram Support Vector Machine (TRIGRAMSVM), and the Bigram Support Vector Machine (BIGRAMSVM) classifiers make up the bagging model that Sun (year) proposed to present classification results. It's possible to anticipate the tone of reviews by using a classifier that parses the meaning of individual words in reviews about a product. The model was used to both incorporate previously disjointed product-review associations and turn the review's language into a continuous representation. To build the representation model, the researchers fed product word composition vectors into a Convolutional Neural Network (CNN). Their F-score was 0.77 after they had finished both the T RIGRAMSSV M and BIGRAMSSV M classifications. However, the guided approach has a number of drawbacks. Possible complications that may arise from using supervised procedures include the ones listed below.

Maintaining reliable reviews is difficult work. It is difficult to acquire labeled data points for the purpose of training a classifier. • It might be difficult for individuals to tell the difference between fake and genuine ratings. Jitendra thus came up with a semi-supervised method that allows for training on both annotated and unannotated data. In these cases, it was suggested that the semi-supervised method be used.

When there is a scarcity of available information that can be trusted.

2) The ebb and flow of feedback posted to websites.

Thirdly, difficulties can arise during the process of creating heuristic rules. Algorithms for semi-supervised learning such as co-training, expectation maximization, label propagation and spreading, and positive unlabeled learning were introduced. k-Nearest Neighbor, Random Forest, Logistic Regression, and Stochastic Gradient Descent were all used as classifiers in the research. The researchers improved accuracy to an impressive 84% by employing semi-supervised methods.

3. SYSTEM DESIGN

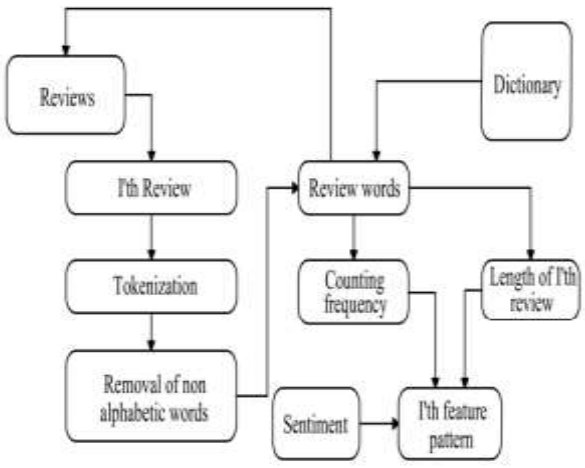
An overview of the dataset that was analyzed follows. In this investigation, we use the 'gold standard' dataset developed by Ott et al. (3, 8). Included in the collection are 1,600 textual evaluations of 20 hotels in and around Chicago. In a total of 1,600 ratings, 800 are found to be fake, while the other 800 are trusted. A '0' next to an evaluation indicates that the review was likely made up, whereas a '1' indicates that the review was genuine. There are 400 negative reviews and

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, August : 2023

400 positive reviews in the dataset, each with their own unique numeric value. Four hundred of the genuine testimonials are positive, while the other four hundred are critical, just like fake reviews. These evaluations stem from various places. Fraudulent reviews were generated with the help of Amazon Mechanical Turk (AMT), and additional content was culled from sites including Yelp, TripAdvisor, Expedia, and Hotels.com, among others. The evaluations use a predefined dataset split. The corpus is split into two sets of instances, the training set and the test set, with a total of 1,600 cases utilized for each. The corpus is split in two halves, with a 75:25 gender split and an 80:20 male/female split. The events in each set are chosen at random.

The suggested procedure

Step one involves making use of raw textual data to spot fake online reviews. The dataset we used has been annotated by other academics before we used it in our analysis. We get rid of all the fluff, like



Articles and prepositions are removed in a variety of ways, some of which are seen in Figure 1. The classification algorithm then takes these textual inputs and converts them into numerical ones. Classification efforts were launched after the most vital characteristics had been eliminated. Because we relied on Ott's "gold standard" collection, we avoided having to manually add missing data, correct inconsistencies, remove duplicates, etc. Preprocessing duties, on the other hand, necessitated collating texts, developing a vocabulary, and translating content into numerical form. Word count, emotional valence, and review length were all considered in our analysis. About 2,000 words have been utilized as highlights. As a result, our feature vector has 1,602,002 dimensions. Our feature set does not include n-grams or parts of speech because they are derived from a word corpus, which increases the risk of overfitting. The overall process of feature extraction is depicted in Figure 1. Figure 1 depicts the composition of the i-th review's correlated features.

Tokenization is the starting point for any audit. The next step is to eliminate superfluous words and generate potential key words.

Each candidate feature term is checked against its dictionary meaning here. If a match is made, the frequency of the word is computed and added to the frequency column of the feature vector.

The review's length and frequency count are both added to the feature vector.

Next, the mood score is added to the dataset, and the feature vector is updated accordingly. Having values other than "0" in the feature vector is indicative of a negative mood, whereas having positive values is indicative of a positive mood. Semi-supervised and fully-supervised models have both been implemented. Using the semi-supervised Expectation-Maximization (EM) technique, the dataset was partitioned. The concept of using Expectation Maximization to locate unlabeled data was initially proposed by Karimpour.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, August : 2023

Algorithm 2 EM Algorithm

INPUT: Labeled instance set L, and unlabeled instance set U.

OUTPUT: Deployable classifier, C.

```
C ← train(L);
PU = ∅;
while true do
    PU = predict(C,U);
    if PU same as in previous iteration then
        return C;
    end if
    C ← train(L ∪ PU);
end while
```

The plan to raise expectations as much as feasible The intended training procedure is depicted in Diagram 2. The procedure entails the following steps: When creating a classifier, the labeled dataset is the starting point. The dataset without labels is then labeled using the aforementioned algorithm. The set of candidates being considered is PU. Using a classifier trained on mixed datasets consisting of both labeled and unlabeled data, the unlabeled dataset is classed. Continue doing this until the selected working unit is steady. Once a consistent positive and unlabeled (PU) data set has been assembled, the classification system can be trained on both labeled and unlabeled data. After being trained, the system is used to make predictions on an unrelated test dataset. Here is how the process goes down. We employed the Expectation-Maximization (EM) technique with classifiers from the Support Vector Machine (SVM) family and the Naive Bayes (NB) family. Numerous classifiers may be found in Python's Scikit Learn tool. Therefore, Python was used to conduct the research, with the scikit-learn and numpy modules. The parameters of the Support Vector Machine (SVM) have been adjusted for better outcomes. Supervised classification has been accomplished using SVM and Naive Bayes methods. If conditional independence is assumed, it is well known that the Naive Bayes algorithm can be utilized. The unpredictable nature of the information generation process, which is tied to the user's mental state, makes it tough to anticipate what will come next in a given passage. Naive Bayes classifier is often employed in text analysis because to this. Because it is a probabilistic technique, it can be used for both classification and regression. Similarly, this occurrence is calculated at a lightning-fast clip.

4. PERFORMANCE ANALYSIS

Experimental Environment and Tools

We have applied our experiments on a machine with Processor: Intel (R) Core (TM) i5-4200U and CPU - 1.6GHz, RAM: 6 GB, System type: 64 bit OS, x64- based processor, Hard Disk: 1 TB. We have used Linux(Ubuntu 16.04) as our operating system. We have used Python programming language with Scikit-learn and numpy packages.

Results We have used Expectation maximization(EM) algorithm for semi-supervised classification. As classifier we have used

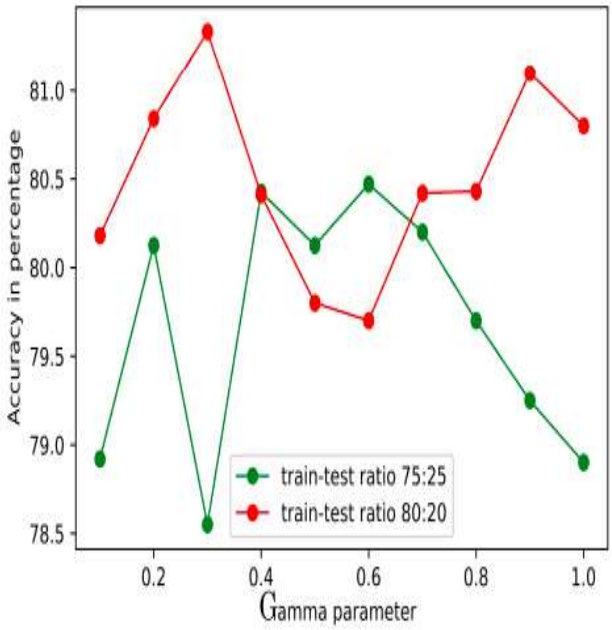


Fig. 3. Graph showing Gamma parameter vs Accuracy for EM with SVM

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 8, August : 2023

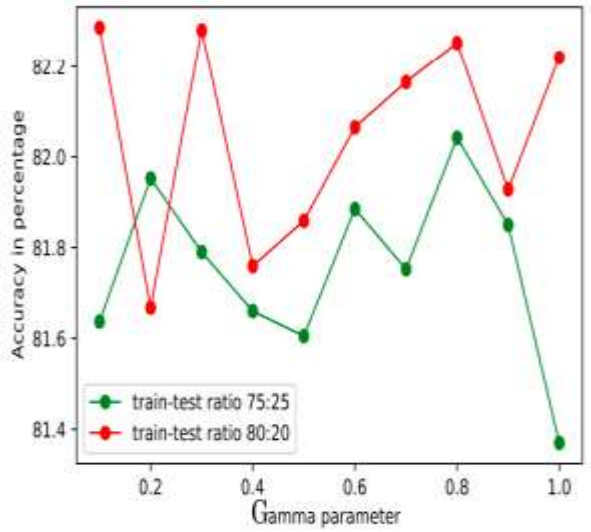


Fig. 4. Graph showing Gamma parameter vs Accuracy for supervised SVM classier

Naive Bayes classifier and Support Vector Machines (SVM) are two of the most used classification methods in machine learning. For each method of classification, the dataset was split 75:25 or 80:20 across the training and testing sets. In this research, we experimented with varying several gamma parameters within the context of semi-supervised classification using Support Vector Machines (SVM), all while holding C constant. The accuracy % curve is shown in Figure 3. The graph shows that an SVM classifier used for semi-supervised classification with an 80:20 split ratio achieved an accuracy of 81.34 percent. Similarly, 80.47 percent accuracy was achieved with a 75:25 split. The gammas of 0.3 and 0.6 were used to attain these results. Naive Bayes classifier-based semi supervised classification achieved an accuracy of 85.21% and 84.87% for 80:20 and 75:25 split ratios, respectively. With a train-test split of 80:20, Jiten's semi-supervised classification using the Expectation-Maximization (EM) algorithm yielded an accuracy of 83.01%, while Positively Unlabeled (PU) learning yielded an accuracy of 83.75%. In their research, they used a variety of classifiers, such as random forest, stochastic gradient descent, K-nearest neighbor, and logistic regression. Additionally, supervised classification algorithms' performance on the dataset was evaluated. In this research, both Support Vector Machines (SVM) and Naive Bayes were used as classifiers. We tweaked the gamma value of the support vector machine (SVM) classifier while holding the C parameter steady to increase the model's accuracy. The end effect can be seen in Figure 4. Using an SVM classifier for supervised classification, an 80:20 split resulted in an accuracy of 82.22%. Similarly, we obtained an 82.04% precision while using a 75:25 split. The values of gamma used to get these outcomes were 0.1 and 0.8. When the data was divided into training and testing sets at a ratio of 80:20 and a ratio of 75:25, respectively, the Naive Bayes classifier's supervised classification accuracy reached 86.32 and 86.21 percent, respectively.

Figure 5 is a histogram displaying the results of the applied approaches and previous studies conducted on the dataset. We gave considerable thought to the features we chose to study in order to reduce the risk of overfitting. We did not use any derived features in our analysis, such as those that are generated from other characteristics.

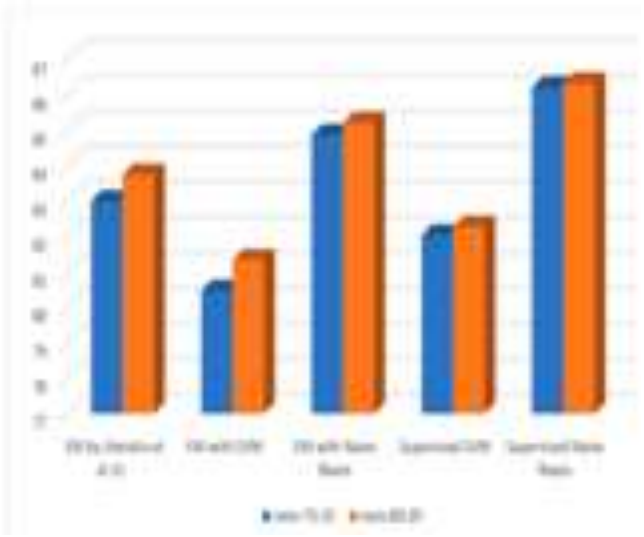


Fig. 5. Histogram showing performances of implemented techniques. Similar grammatical constructions to trigrams and bigrams. Due to its consequential nature, review length was considered worthy of inclusion as a feature. Naive Bayes was selected as the classifier of choice because of its suitability to our dataset. Using this

method, we were able to increase the maximum accuracy of semi-supervised classification from 83.75 percent to 85.21 percent, as reported by Jiten et al. in their study [8]. It was also shown that the maximum accuracy, 86.32 percent, was achieved by the use of supervised classification with a Naive Bayes classifier. Table I provides a summary of the results.

5. CONCLUSIONS

This research showcases a selection of semi-supervised and supervised text mining methods used to detect sham online evaluations. We added to the existing set of capabilities by incorporating findings from previous studies. In addition, we tested out some alternative classifiers that weren't used in the earlier research. This enhanced the precision of previous Jiten semi supervised methods. In addition, the supervised Naive Bayes classifier was found to be the most accurate. This method ensures that our dataset is correctly labeled.

TABLE I
COMPARATIVE SUMMARY OF SEMI-SUPERVISED AND SUPERVISED LEARNING TECHNIQUES

	Features	Algorithm type	Classifier used	Train-test ratio	Accuracy
Jitendra et al.[8]	Bigrams, sentiment score, POS, LIWC	Semi-supervised	k-NN	75:25 80:20	0.8300 0.8313
			Logistic Regression	75:25 80:20	0.8300 **0.8375**
Proposed Work	Word frequency count, sentiment score, review size	Semi-supervised	Naive Bayes	75:25 80:20	0.8487 **0.8521**
			SVM	75:25 80:20	0.8047 0.8134
		Supervised	Naive Bayes	75:25 80:20	0.8621 **0.8632**
			SVM	75:25 80:20	0.8204 0.8228

In cases where precise labeling is unattainable, the usefulness of semi-supervised models is well acknowledged. All of our research thus far has been based on feedback from actual customers.

REFERENCES

[1] Chengai Sun, Qiaolin Du and Gang Tian, "Exploiting Product Related Review Features for Fake Review Detection," Mathematical Problems in Engineering, 2016.

[2] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, "Detection of review spam: a survey", Expert Systems with Applications, vol. 42, no. 7, pp. 3634–3642, 2015.

[3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT), vol. 1, pp. 309–319, Association for Computational Linguistics, Portland, Ore, USA, June 2011.

[4] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic Inquiry and Word Count: Liwc," vol. 71, 2001.

[5] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, Vol. 2, 2012.

[6] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014.

[7] E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.

[8] J. K. Rout, A. Dalmia, and K.-K. R. Choo, "Revisiting semi-supervised learning for online deceptive review detection," IEEE Access, Vol. 5, pp. 1319–1327, 2017.

[9] J. Karimpour, A. A. Noroozi, and S. Alizadeh, "Web spam detection by learning from small labeled samples," International Journal of Computer Applications, vol. 50, no. 21, pp. 1–5, July 2012.