Proceedings of the International Conference on Sustainable Computing and Smart Systems (ICSCSS 2023)
IEEE Xplore Part Number: CFP23DJ3-ART; ISBN: 979-8-3503-3360-2

2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS) | 979-8-3503-3360-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSCSS57650.2023.10169838

Fake Review Detection using Naive Bayesian Classifier

Ms.P.Kalaivani
Assistant Professor, Department of IT
Dr.Mahalingam College of Engineering and Technology
Coimbatore, India
kalaipadman@gmail.com

V.Dinesh Raj
Student, Department of IT
Dr.Mahalingam College of Engineering and Technology
Coimbatore, India
dineshrajvdr01@gmail.com

R.Madhavan
Student, Department of IT
Dr.Mahalingam College of Engineering and Technology
Coimbatore, India
maddyy1472002@gmail.com

A.P.Naveen Kumar
Student, Department of IT
Dr.Mahalingam College of Engineering and Technology
Coimbatore, India
naveenkuma1045@gmail.com

***Abstract*— The issue of fake reviews is becoming an increasingly prevalent one on the internet. The purpose of these reviews is to deliberately deceive potential customers and influence their purchasing decisions. Businesses and customers alike are looking for ways to spot and filter out these fake reviews as a result. The Naive Bayes algorithm is one effective method for identifying fake reviews. A well-known machine learning algorithm for classification tasks is Naive Bayes. It is based on the probability theorem of Bayes, which enables us to determine the probability of an event with some evidence. The Naive Bayes algorithm can be trained on a dataset of reviews that are known to be real or fake in the context of fake reviews. The characteristics of genuine and fake reviews are then learned with the algorithm by utilizing this training data. After the algorithm has been trained, it can use the characteristics of new reviews to determine whether they are genuine or fake. The fact that Naive Bayes is a relatively straightforward algorithm that can be trained quickly and easily is one advantage of using it to detect fake reviews. Additionally, it works well with text data, which is the format used by the majority of reviews. Having said that, it's critical to keep in mind that Naive Bayes isn't perfect and may not be able to spot all fake reviews. Cleaning and normalizing data, dealing with missing data, and dealing with outliers are all potential obstacles in data pre-processing.**

Keywords— Fake Review, Naïve Bayes Classifier, POS Tagging, Negative Datasets, Natural Language Processing,

I. Introduction

The problem statement of our project is to detect fake reviews from the million or any number of reviews posted. This can be resolved by our project. In order to solve the process of finding and removing reviews that are not genuine and were written with the intention of manipulating the overall rating or reputation of a product, service, or business is known as Fake Review Detection. And for the Classification and Regression Analysis, Support Vector Machine (SVM) is used.

The main reason for choosing this approach is its accuracy and efficiency. And also the other advantage is that the SVM can be applied to both linear and non-linear data. Additionally, they can handle noisy and high-dimensional data.

Finally, another algorithm is used for the Classification which is the “Naive Bayes Algorithm”. The main specialty of this algorithm is to state the likelihood of a hypothesis. which means it is good to predict the Probability of the evidence in terms of the hypothesis. And also it is used for text classification including spam filtering, sentimental analysis, and topic classifying as well as picture recognition, medical diagnosis, and fraud detection. All-in-all, Naive Bayes is an effective yet simple algorithm that’s extensively used in ML and DS due to its efficiency in classification tasks

The review dataset is taken from the Kaggle website and loaded into the proposed “review classifier model”. The model will load 20 thousand reviews as training data. The model will classify the training data into 3 voting columns cool useful and funny. The proposed model will split the review based on the star ratings. The 1-star rating 2 and 3-star combined ratings and 4 and 5-stars combined.

Later on, data cleaning is used to remove the unnecessary words in the training data. The stopwords, and punctuation all are removed. By using vectorization we convert the data into vectors. The model will split the data into training and test data. The testing will b e only 2000 data. Now the test data will be used to detect fake reviews using many ML methods.

The dataset needs to include reviews from reliable sources, including features like length, language, and emotive expression. The dataset's size is also crucial to take into account.The dataset has to contain standards for classifying reviews as real or fraudulent."The dataset should be evaluated" should follow. This study uses SVM and the naive bayes algorithm to detect fake reviews. For display, the accuracy of the algorithm the model uses a confusion matrix to plot the accuracy. The classification report tells us about the precision and weighted average of the test data.

The Naive Bayes provides a clear understanding of how the classification is made based on the probabilities of the features. It requires less training data compared to other more complex algorithms. One of the main advantages of Naive Bayes is that it is simple, fast, and easy to implement.

II. Literature Survey

A. *A New Approach to measuring Conflict in Legislative Speeches using Multi Sentiment Analysis*

This work looks at how to incorporate user perception of social network influence into a probabilistic model for temporal activity (such as posting and tweeting) on social networks. Previous models have been developed for this, but often lack transparency on an individual level; thus in which a user's activity is modeled as a linked hidden Markov chain [1] with a secret state influenced by their peers' combined activity. For the purpose of assisting with parameter learning and state determination, we also developed generalized Baum-Welch and Viterbi algorithms. To substantiate this suggested structure, we used a substantial Twitter user activity corpus. Our numerical studies show that the proposed framework works better than models based on the renewal procedure or standard (uncoupled) HMMs when attempting to describe a set of facts when there are sufficient observations to ensure accurate model learning.

B. *Analysis of Social Networks*

This article discusses the application of social network Analysis methods in the area of data fusion. From fused data that merges various intelligence reports from the same setting, social network extraction and high-value person (HVI) recognition are of interest. Research on the feasibility of such activities may contribute to the testing and assessment of fusion quality as well as methodological developments in network science. This study describes a method for extracting a social network of people derived from combined data and storing it as a cumulative related data graph using parallel computing. To build the ideal social network, two methods—a step count weighted method and a route salience method—are devised and contrasted. A supervised learning approach is used in order to parameterize the extraction methods. The parameters of the extraction algorithm consider the pathways of the social network and weight connections between them using the count weighted and path salience techniques. Path saliences and path hop count weights are merged to create a total link strength number for the hop count weighted and path salience methods, respectively. Ordered centrality-based HVI lists are derived from CDGs built from the Sunni criminal and Bath'est resurgence strands of the SYNCOIN data set under a range of fusion system parameters. The findings show the value of HVI identity as a tool for fusion process optimization testing and evaluation as well as the susceptibility of accuracy, and degree centrality measures to fused graph inputs. The computational findings demonstrate that the route salience technique is more effective at locating HVIs.

C. *Text Classification using Wikitology as Knowledge Enrichment: A Comparison*

Text categorization is the process of labeling papers based on their substance. The addition of background information to the document using knowledge sources like Word Net, Open Project Directory (OPD), Wikipedia, and Wikitology has been the subject of numerous trials to improve text categorization. In our earlier work, we conducted in-depth tests using information taken from Wikitology and evaluated the results using Support Vector Machines and 10-fold cross-validations. The findings make it abundantly obvious that Wikitology is superior to other information bases. In this study, we compare the Naive Bayes (NB) and Support Vector Machine (SVM) models for text augmentation using Wikitology. With 10-fold cross-validation, we verified the findings and demonstrated that, when compared to baseline results, NB provides an increase of +28.78%, while SVM provides an improvement of +6.36%. When external enrichment is used through any external knowledge source, the Naive Bayes classifier is a superior option. By contrasting trial findings on SVM and Naive Bayes Classifier, we offer an improved version of our prior study in this article.

D. *Using Supervised Machine Learning Techniques to Analyse Sentiment in Movie reviews*

Nowadays, societal proposals are increasingly being made with a social democratic perspective. It helps make precise suggestions by using variables like social confidence and other factors. We introduce Matrix factorization (MF) and nearest neighbor-based recommender systems that consider user actions and evaluate them against peer evaluators in order to provide an accurate suggestion. Our research brought us to the conclusion that the affiliation variables are crucial to improving the reliability of recommender systems. The study showed that cold users found the suggestion system to be more useful than heavy users, which knowledge helps us overcome the recommendation system's cold start problem. In our research, the basic neighborhood model did better in both the hot voting and non-hot voting recommendations than computerized matrix factorization models.A hybrid recommender system that merged various separate methods to produce a top-k suggestion was also one of the ideas we put forth. Users who shared comparable traits were grouped and filtered as part of the first approach. Only users who exhibit comparable traits are regarded as informational pals in this method. The second tactic [2], referred to as a static threshold, includes accounting for a set number of best peers for resemblance. Having informational pals with a changing barrier is the third tactic. The flexible limit of users in this technique is set by the friends of the active user, but only if those friends have the most things in common with the targeted user.

E. *A Repulsive Voter Model's Dynamics*

The kinetics of repellent interactions between network nodes are examined in this article, in contrast to the majority of conventional social interaction models. This model could be used to determine a first-order approximation of many anti-conformist societal policies, such as drivers deciding to shift lanes when faced with other vehicles in the same lane. In any social situation where there is a lot of novelty or option, these types of policies are also evident, which can indicate a lack of interest in or resistance to current trends. Examples of this kind of anti-conformist social behavior include a variety of styles of anti-fashionable dresses that explicitly go against the current fashion. First, a complete graph with and without biased nodes is used to develop and solve the model given in this paper. To confirm the model's formulation's accuracy, Monte Carlo simulations were used. The next two topics of study are the behavior of the model on a regular lattice and an Erdos Rényi random graph. For the aforementioned combinations, analytical answers have been found. The model exhibits intriguing behavior with regard to the effects of biased nodes, convergence time, and population size on the equilibrium distribution. where each node repels rather than draws its neighbors. A complete graph of the model is first created, and it is then solved using the impact of influences at equilibrium. The equilibrium density pe(m) is computed as a function of population size N, overall impact size I, and control input u = (I+ I/I). After that, the model is tested on a random network with p connections. Analytical answers for such a network are also derived at a balance point.

III. Proposed Methodology

This study intends to use a variety of feature selection techniques and extend this study to include other datasets like the eBay or Amazon dataset in the proposed system. Additionally, using a variety of tools from NAIVE BAYES Techniques, we may use sentiment classification algorithms to identify fake reviews. We will then assess how well our work with some of these tools performed. However, consumers have found it challenging to trust the veracity of these reviews due to the rise of fake reviews. Fake reviews can be detected using machine learning algorithms like Naive Bayes to Address this issue. A probabilistic algorithm known as Naive Bayes uses the words used in a review to determine whether it is genuine or fake. It expects that the presence Address this issue. A probabilistic algorithm known as Naive Bayes uses the words used in a review to determine whether it is genuine or fake. It expects that the presence of specific words in a survey expands the likelihood of the survey being phony, while the shortfall of these words builds the likelihood of the audit being real. Customer review mining and summarization is extremely pertinent to the suggested task. The "Document Based Sentiment Orientation System" incorporates feedback from customers and reviewers. It divides the overall production into positive, negative, and neutral papers before classifying each one individually. Users can easily make choices using this information by counting the number of favorable and bad evaluations.SVM is a more complicated algorithm that tries to find the best hyperplane to divide a high-dimensional space into classes. However, because of these factors, SVM may require more resources and take longer to train than Naive Bayes.Because of these factors, accuracy suffers.

We load the opinion data set into our model .the dataset was taken from Kaggle which contains 20 thousand reviews with ratings and user details. we store the dataset in an Excel sheet we load the whole dataset as training data into our model. Now the model will display the no of reviews with the user details. We add a new column length to store the length of the review.SVM is a discriminative algorithm, whereas Naive Bayes is a probabilistic algorithm. Naive Bayes , a fast and easy algorithm.In general, Naive Bayes is quicker than SVM and uses less memory. Naive Bayes is many times utilized for Classification

Let us now visualize the training data if there is any correlation between the length and the star ratings. the star's ratings are classified into 5 categories 1-star rating to 5-star rating. we use the graph to plot all the reviews based on the reading in a graph. the star rating fig is shown below.

After analyzing the dataset, different feature selection techniques are used to select the most relevant features for the proposed model. By applying these feature, the performance of the proposed model can be improved and make it more efficient. Once the most relevant features are selected, the datasets are trained using Naive Bayes algorithm which calculates the probability of a review being positive, negative reviews. Cross-validation techniques are used in the evaluation process.

Fig. 1. Star Rating Diagram

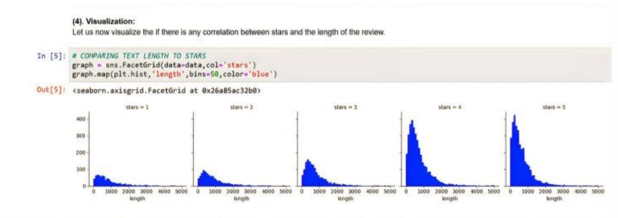
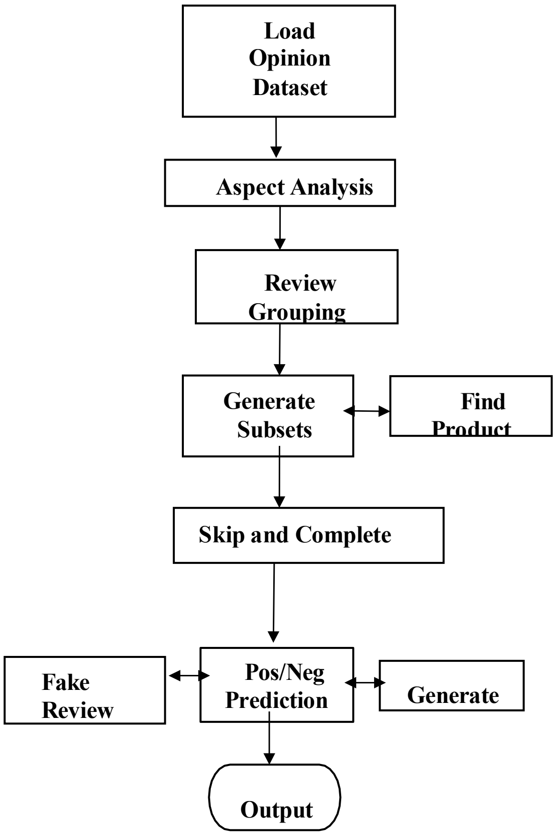


Fig. 2. Flow Diagram



Based on the star ratings we classify the reviews into 3 voting columns cool useful and funny. And the model will find the correlation between the voting columns. The correlation between the voting columns tells us that there is a negative correlation between cool and funny, cool and length and, cool and useful and there is a positive correlation between useful and funny, funny and length, useful and length. Thus the longer reviews tend to be useful and funny. Now the model will classify the dataset and split it into reviews and stars. Next by using a function it will remove all the grammatical errors and stopwords from the training data. we vectorize the whole dataset which means converting the review words into vectors it will be easy to use algorithm methods. The dataset is then split into the training data and testing data. The testing will be 2000 reviews. Now the model uses an ML algorithm Naive Bayes and SVM to detect the fake reviews from the classified vectors. by using the confusion matrix it will plot the accuracy of the ML algorithm. By analysing datasets in the Multidimensional Naive Bayes Algorithm, Jupyter Notebook is used to determine the accuracy of fake review detection using the Naive Bayes approach. The precision can then be seen accordingly.

Fig. 2. Flow Diagram Fake Review Detection

A. *Data Preprocessing*

As Shown in the flow diagram Fig.1.The Data preprocessing is an essential component of fake review detection, as it involves transforming raw data into something that machine learning algorithms can use to identify patterns and relationships. To ready data for detecting fake reviews, it has to be scrubbed of irrelevant information, tokenized into individual words/phrases, converted to all lowercase, stemmed/lemmatized to reduce words to their root form, have common stop words taken out, and relevant features extracted. Furthermore, balancing the dataset is important for better accuracy of the model. With proper preprocessing of data, ML algorithms [4] can be taught to detect fraudulent reviews and protect consumers from deceitful practices.

B. *Data Extraction*

Data extractions are carried out using a variety of mining methods; It can be at the phrase, sentence, or document level. The tool that is used for data extraction can be supervised or unsupervised. The machine learning algorithm Naive Bayes (NB) is part of the supervised method. Since the current review information isn't taken into account for analysis, the proposed system uses an unsupervised method. This process consists of the following steps.

C. *Stop Words*

Some words have been identified and tagged, and since some of them are present in the review, they are useless when performing sentiment analysis. Stop words, such as "is, a, the," which are commonly used, along with '_IN', '_DT', '_CC', '_TO' and '_VBZ' should be removed to simplify things further.Each block is made up of an activation, convolutional, and max pooling layer. Fully connected layers, softmax of activation, and three of these blocks are utilized in this architecture. Classification is done with the fully connected layers[6,7], while feature extraction is done with the convolutional layer and pooling layer. Through activation layers, nonlinearity is introduced into the network.

D. *Stemming*

Stemming is a technique for separating words' basic forms from affixes. It's like taking the limbs off a tree and just leaving the roots. The words are indexed by search engines using stemming. Because of this, a search engine can only store stems rather than all word forms. Stemming does this by making the index smaller and making retrieval more accurate.

E. *Post Tagging*

Using Natural Language Processing (NLP), we label the words in a sentence based on their parts of speech. To do this, we break the sentence up grammatically and attach a tag to each word indicating if it is a noun, adjective, verb, preposition or any other type of word.

F. *Data Comparison*

The negative and positive datasets are compared to the POS tagging output. Positive values are either positive or negative. Intermediate is shown if both positive and negative values are the same.

IV. RESULT ANALYSIS

Various fake reviews have been identified in a number of research papers. The relationships between proposed and prior work on the detection of fake reviews are shown in Table I. To improve the model, use domain-specific characteristics like the frequency of keywords, review length, and specific phrases have to be used in our project.

TABLE I. Accuracy and Algorithms

Algorithm	Accuracy
NB	84%
SVM	76%

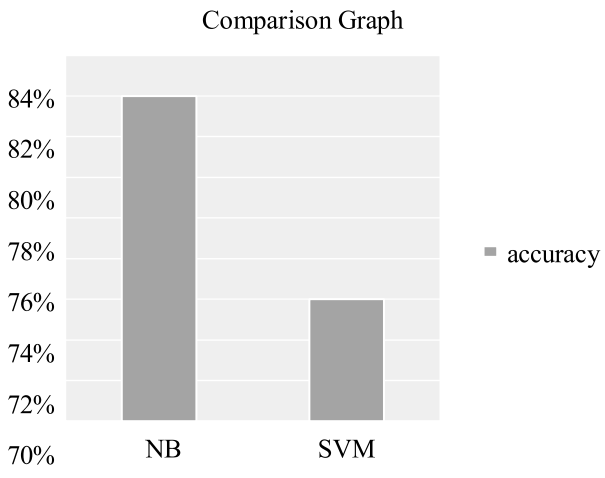


Fig. 3. Accuracy – Comparison Graph

By Summarizing the accuracy results of user behaviour without extracting features as shown in Fig.3. we see that the NB classifier yields an average accuracy of 76%, while the SVM classifier provides an accuracy of 71%. Taking into account both language models, it's evident that NB has the highest average accuracy with 84%. In making assumptions, Naive Bayes strongly assumes that features are conditionally independent. In terms of training speed, Naive Bayes is faster to train than SVM because it involves simple probability calculations. In terms of robustness, Naive Bayes is known to be more robust to noisy data than SVM. When it comes to interpretability, Naive Bayes is a probabilistic model, so it can provide insight into the probability of belonging to each class.

CONCLUSION

This study has demonstrated the significance of reviews and how they influence almost all web-related matters. It is clear that reviews are essential to people's choices. Consequently, detecting fake reviews is a lively and continuous field of research. In the proposed approach, both characterizing features of the reviews and behavioral features of the reviewers are taken into consideration. The results show that NB classifier has an edge over all other classifiers when it comes to detecting fake reviews. Also, it is evident that considering reviewers' behavior boosts the f-score, even though not all reviewers' behavior has been taken into account in this work. Further research could consider including other behavioral attributes such as how often reviewers do their reviews, how quickly they finish reviewing, and how frequently they give positive or negative reviews. It is likely that accounting for more behavior boosts the efficiency of this fake review detection approach. The accuracy of Naive Bayes can be evaluated by comparing the true class labels of the predicted class labels for a set of test data. The classification accuracy, which is the ratio of instances correctly classified to the total number of instances, is a common metric for measuring accuracy. By this presentation of the classifier is estimated. The advantages include increased sales as a result of positive reviews because negative reviews are shown to be fake. Another advantage is a better user experience as a result of the user not having to worry about the review. However, the disadvantages include high costs, especially when dealing with large datasets, making it challenging to scale up to handle large volumes of reviews. Additionally, the utilization of fake review detection raises ethical concerns regarding privacy and fairness

REFERENCES

[1] Karunananda, A. S., Silva, T. P., and Vidanagama, D. U. (2020). Consumer review fraud detection: a study. 53(2) of the Artificial Intelligence Review,13231352.http://dx.doi.org/10.1007/s10462019-09697-5.

[2] Du, Q., Tian, G., and Sun, C. (2016). exploiting characteristics of reviews that are linked to products to identify fake reviews. Engineering Mathematical Problems, Article e4935792.http://dx.doi.org/10.1155/2016/4935792.

[3] Machine & deep learning methods for the identification of fake reviews: A survey, J. C. Rodrigues, J. T. Rodrigues, V. L. K. Gonsalves U. Naik, P. Shetgaonkar, and S. Aswale, Proc. Int. Conf. Emerg.Trends Inf. Technol. Eng. (ic-ETITE), Feb. 2020, pp. 1-8.

[4] Mars, A., and M. S. Gouider. (2017). Analysis of big data includes consumer opinion mining. Procedia Computer Science 112,906914.http://dx.doi.org/10.1016/j.procs.2017.08.114

[5] A review structure-based ensemble model for misleading review spam was developed by Z.-Y. Zeng, J.-J. Lin, M.-S. Chen, M.-H. Chen, Y.-Q. Lan, and J.-L. Liu. Information, July 2019, vol. 10, no. 7, p. 243

[6] Deep graph neural network-based spammer detection under the The viewpoint of diverse cyberspace, Future Generation Computer Systems, vol. 117, pp. 205-218, April 2021. Z. Guo, L. Tang, T. Guo, K. Yu, M. Alazab, and A. Shalaginov.

[7] A new model for opinion spam identification based on multi-iteration network structure, Adv. Sci. Lett., vol. 24, no. 2, pp. 1437-1442, February 2018. S. Noekhah, N. B. Salim, and N. H. Zakaria.

[8] Deep networks and transfer learning are used by N. Dhamani, P. Azunre J. L.Gleason, C. Corcoran, G. Honke, S. Kramer,and J. Morgan to combat misinformation. 2019, arXiv:1905.10412. [Online]. Available: http://arxiv.org/abs/1905.10412.

[9] Leveraging deep learning models for ransomware discovery in the industrial Internet of Things ecosystem, M. Al-Hawawreh and E. Sitnikova, in Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS), November 2019, pp. 1-6.

[10] "DistilBERT, a condensed form of BERT: Smaller, quicker, cheaper, and lighter," by V. Sanh, L. Debut, J. Chaumond, and T. Wolf 2019, arXiv:1910.01108. [Online]. Accessible at: arXiv.org/abs/1910.0110

[11] He, D., Qiu, M., Xiong, You, L., Peng, Q., and Zhang, X. (2020). Aspect analysis and the local anomaly factor are combined for effective review spam identification. 102, 163–172, Future Generation Computer Systems. http://dx.doi.org/10.1016/j.future.2019.07.044

[12] In September 2017, Y. Zhu, D. Li, R. Yan, W. Wu, and Y. Bi published a paper in the IEEE Transactions on Computational Social Systems titled "Maximize the Influence and profits in social networks."

[13] Explainable artificial intelligence: A comprehensive study, G. Vilone and L. Longo 2020, arXiv:2006.00093. [Online]. Available: http://arxiv.org/abs/2006.00093

[14] Roberta: A highly optimized BERT pretraining method, by Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, 2019, arXiv:1907.11692. [Online]. Accessible at: arXiv.org/abs/1907.11692

[15] Ren, F.;Quan, C. (2016). Enterprise Information Systems, 10(5), 505-522. http://dx.doi.org/10.1080/17517575.2014.985613 Feature-level mood analysis using comparison topic corpora.