Fake Review Detection using Sentiment Analysis

Chandrabhanu Rawat
chandrabhanu.rawat2@mail.dcu.ie
Dublin City University
Dublin, Ireland
Student Number: 22261524

Harshita Tyagi
harshita.tyagi2@mail.dcu.ie
Dublin City University
Dublin, Ireland
Student Number: 22266026

Sagar Gatla
sagar.gatla2@mail.dcu.ie
Dublin City University
Dublin, Ireland
Student Number: 22262265

Udit Srivastava
udit.srivastava2@mail.dcu.ie
Dublin City University
Dublin, Ireland
Student Number: 22260528

Links:
1.Github: https://gitlab.computing.dcu.ie/tyagih2/ca683-group18-mainassignment/-/tree/main/
2.Video:https://drive.google.com/file/d/1R-2uvcbCQSL2W-z8I1O3dV-PhvNwyAYT/view?usp=share_link

Abstract—In order to identify fraudulent reviews on the Yelp Dataset using the feelings of the review, this research compares many machine learning models, including Logistic Regression, Random Forest, Naive Bayes, and LightGBM. We are comparing these models in order to select the model that is most appropriate for dealing with the difficulties given by fake review classification. When combined with ADASYN, LightGBM produced the best accuracy, with a forecast F1 score of 0.88.

I. Introduction

The e-commerce revolution has driven the online community to use review feature for posting feedbacks about goods and services, which in turn will help people to make buying choices and help businesses to analyse their trends and maintain their standards, but these review mechanisms are widely misused by vendors. This is the reason why product and services review play a crucial role in e-commerce.

Fake reviews have recently gained importance, as they can help businesses to gain customer attraction, promote a brand's reputation or can completed demote a brand's performance. For example, if a company creates a fake account and leaves negative feedbacks on a particular product or website, every customer visiting that website will see the negative comments and will leave without availing any services. This will in turn negatively impact the business.

Another challenge here is, same word can have different meaning in different context. For e.g. the word "long" in terms of a laptop's battery life is a positive opinion while the same word about the start time is a negative opinion, which is why we need a system that can check for the sentiments of the review and classify it as fake or real.

Sentiment analysis also called opinion mining is the study of digital text to identify whether the message's emotional tone is good, negative, or neutral. Using NLP, we can train computer software to understand text similarly to humans. Therefore, this research aimed to use supervised learning algorithms to increase the fake review identification system's accuracy. Extracting characteristics from the review text served the purpose and was a significant task.

Yelp is a commercial service and crowd-sourced review site where users may publish reviews about a specific business. This enables businesses to obtain free advertising from consumers who utilize a service and provide favorable comments about it. However, a problem develops when a tiny number of unscrupulous business owners attempt to manufacture bogus reviews in order to improve their businesses. They recruit people to write fake reviews about their businesses on the Yelp website. [1]

In this research, we are doing a comparative study of different classification techniques in Machine Learning like Logistic Regression, Naive Bayes, Random Forest and LightGBM on the Yelp dataset. When combined with ADASYN, LightGBM produced the best accuracy, with a forecast F1 score of 0.88.

II. Related work

Setieyvi et al. (2020) conduct a comparison study on various machine learning methods for fake review detection. The research compares three algorithms: Support Vector Machines (SVM), Logistic Regression, and Random Forest. A comparison of the findings before and after fine tweaking is also performed. These algorithms' performance is compared using several assessment measures such as Precision, Recall, and F1 score. The research finishes by stating that the algorithm used to detect false reviews is determined by a variety of criteria, including the size and features of the dataset, the required level of accuracy, and the computational resources available.[2]

S. M. Anas and S. Kumari (2021) conducted a similar analysis, comparing Random Forest and Naive Bayes using Amazon and Yelp datasets. According to the study, Random Forest outperforms Naive Bayes when it comes to classifying fake reviews.[3]

Chen et al. (2020) proposed a intrusion detection method using ADASYN and LightGBM for improving detection accuracy. ADASYN is used to improve the accuracy by solving the problem of imbalanced dataset. A comparison is done between the results of detection with and without using ADASYN, based on the scores on various metrics like FAR, Precision, Recall and Accuracy. The paper concludes by stating that when employed with ADASYN, LightGBM gives a very low false alarm rate. [5]

A study by Wang et al. (2020) proposes a method for detecting fake reviews using multiple feature fusion. The study highlights the importance of combining different features and leveraging collaborative training to improve the accuracy of fake review detection. The proposed method is evaluated using two public datasets, Yelp and TripAdvisor. The results show that the proposed method achieves higher accuracy and F1 score than existing methods such as Support Vector Machine (SVM) and Random Forest (RF). The study also shows that collaborative training improves the performance of the detection model over time.[6]

Another study by Iksan et al.(2021) performs a classification study on twitter data to predict public sentiments on Covid19 out break. This paper describes various pre-processing techniques used to improve the accuracy of Naive Bayes algorithm for identifying sentiments. The paper uses TF-IDF for extraction and generates a feature vector which was then used for the identification or sentiment analysis of public reaction. [7]

III. Methodology

A. *Data Description:*

We used the Yelp Dataset available on Kaggle[8]. This dataset is previously used by various researchers for classifying fake reviews. This is a labelled dataset and was split in two files -

- Yelp NYC Metadata
- Yelp

This dataset contained approx. 3,00,000 reviews of various products and both the files were linked by product id. While Yelp NYC Metadata contains Product ID, Rating and Labels, Yelp file had Product ID, Review ID and Reviews. For our experiment we merged both these datasets into one, using Product ID, including all the columns.

B. Procedure:

a) Data Analysis -: The data analysis process includes analyzing data and finding patterns in data. As part of the data analysis process, we reviewed number of labels and the number of ratings per label. As a result it was observed that our data was highly imbalanced, with a ratio of 1:10.

b) Data Cleaning -: The data cleaning process includes handling missing and noisy data. The dataset used for this experiment does not have any missing values. However, for label the values used were -1 for fake and 1 for real. As part of the data cleaning process, we updated these values to Real and Fake to make it more easily readable and understandable.

c) Data Pre-Processing -: As part of this experiment, preprocessing is mainly done on the review column. The data pre-processing includes various steps -

- Casing - Data is converted to a particular case, in order to reduce the vocabulary size. It could be converted to lower or upper case.
- Tokenization - Data Tokenization is the process of breaking a piece of text apart into pieces that a machine can understand. We removed special characters from review text as part of the tokenization process.
- Stopwords Removal - Stopwords are words that do not contribute to the meaning of a sentence. Hence, they can safely be removed without causing any change in the meaning of the sentence. We used NLTK library to remove stopwords from our text and return a list of word tokens.
- Lemmatization - Lemmatization converts a word to its root form and ensures that the root word belongs to the language. We used NLTK libraries to perform lemmatization.

d) Sentiment Analysis-: Sentiment analysis employs natural language processing (NLP) and text analysis technologies to discover and extract a writer’s feelings from text in the form of tweets. These emotions or sentiments can be good, negative, or neutral.[8] For this research we have used TextBlob. Textblob classifies the sentiment into Positive, Negative and Neutral Sentiment score and Polarity, and provides a comparatively better and faster result.[9]

e) Feature Engineering-: We translate queries into numbers/number of vectors because Machine Learning models do not function with language directly. We are using Term Frequency- Inverse Document Frequency (TF-IDF) as part of the feature Engineering process. TF-IDF is a Statistical measure that reflects how important a word is to a document in a set of documents.

f) Over Sampling-: Over-sampling is done to handle the imbalance in data. ADASYN is a technique for over-sampling. The ADASYN algorithm is a synthetic sampling approach that is adaptable. Its key idea is to apply weights to various minority class instances based on their learning difficulty levels. In contrast to easier-to-learn minority examples, these minority examples will generate more detailed data. The sample rate can have a somewhat balanced effect, which helps to alleviate the problem of data imbalance.[6] The dataset used for this experiment has an imbalance of 1:10 and so we are using ADASYN oversampling technique for minority classes.

g) Model Selection:: In this research, we have chosen 4 classification techniques, namely Logistic Regression, Random Forest, Naive Bayes and LightGBM.

- Logistic Regression: Logistic regression concerns with describing the relationship between explanatory variables and discrete predictors which is achieved by estimating probabilities using the underlying logistic function. Logistic regression assumes the explanatory variables are independent of each other, which can be an advantage or disadvantage depending on the dataset.[1] For fake review detection, it predicts value between ’Fake or Real reviews’ given ’the category, rating, and sentence review of a product.[2]

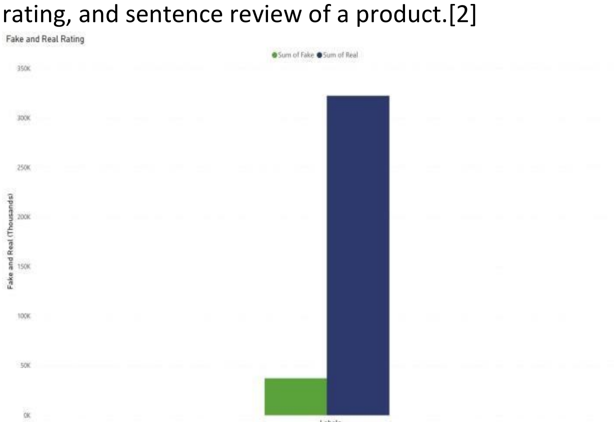

Fig. 1. Data Distribution

- Random Forest: Random Forest is a machine-learning technique that creates large ensembles of little decision trees. It employs a small number of inexperienced students to create a linear relationship. Output is the category chosen by the majority of the tree. The accuracy of the forest is proportional to the number of trees in it.[1]
- Naive Bayes: Given the class variable, a Naive Bayes classifier requires estimating one component to be independent of the estimation of another. It makes use of the preparation data to determine the likelihood of each result based on the highlights. One notable feature of the Naive Bayes calculation is its doubts about the information. It anticipates that all of the highlights in the dataset are autonomous and of comparable significance.[4]
- LightGBM: LightGBM is an open-source GBDT algorithm developed by Microsoft. It uses a histogram based algorithm to speed up the training process, reduce memory consumption and combine advanced network communication to optimize parallel learning, called the parallel voting DT algorithm. Also, LightGBM uses the leaf-wise strategy to grow trees and find a leaf with the largest variance gain to do the split. [5]

IV. Experiments

The experiment begins by using the labelled dataset consisting of 359052 records with 322167 Real and 36885 Fake reviews, as the data is imbalanced so we used ADASYN which is an extension of SMOTE.

The idea is to generate synthetic examples of the minority class so that the number of samples in the minority class is similar to the number of samples in the majority class. We choose to go with ADASYN because of its effectiveness in generating new samples in the harder-to-learn regions of the minority class.[10]

Based on all the extracted features, we implement four machine learning algorithm: logistic regression classification, Naive Bayes classification, Random Forest Classifier and LightGBM. The results are shown below:

TABLE I

Comparison between different Models

Model	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.924864	0.814217	0.866021	0.774115
Naive Bayes	0.929933	0.791542	0.855175	0.759619
Random Forest	0.932430	0.596238	0.727366	0.599240
LightGBM	0.916664	0.848370	0.881196	0.794892

Conclusions

In this research, we have compared the best-supervised machine learning models for fake review detection on the Yelp-NYC dataset. The data being imbalanced has been dealt with Over-sampling using ADASYN. We applied four classification algorithms in the experiment: Logistic Regression, Naive Bayes, Random Forest and LightGBM. The comparison concludes that LightGBM performs better and gives the best accuracy of 79.48% along with an f1-score of 0.8811.

For future work, there is a scope for improvement using complex feature engineering, by applying advanced Neural Network and better sentiment analysis method like Roberta can improve the performance even more.

References

[1] A. Sihombing and A. C. M. Fong, "Fake Review Detection on Yelp Dataset Using Classification Techniques in Machine Learning," 2019 International Conference on contemporary Computing and Informatics (IC3I), Singapore, 2019, pp. 64-68, doi: 10.1109/IC3I46837.2019.9055644.

[2] F. Setievi, J. Natalia, T. R. Tjhang, I. S. Edbert and D. Suhartono, "A Comparative Study of Supervised Machine Learning Algorithms for Fake Review Detection," 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 2022, pp. 306-312, doi: 10.1109/ISRITI56927.2022.10052860.

[3] S. M. Anas and S. Kumari, "Opinion Mining based Fake Product review Monitoring and Removal System," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 985-988, doi: 10.1109/ICICT50816.2021.9358716.

[4] M. R. Machado, S. Karray and I. T. de Sousa, "LightGBM: an Effective Decision Tree Gradient Boosting Method to Predict Customer Loyalty in the Finance Industry," 2019 14th International Conference on Computer Science Education (ICCSE), Toronto, ON, Canada, 2019, pp. 11111116, doi: 10.1109/ICCSE.2019.8845529.

[5] W. Chen, H. Wang, M. Fei, D. Du and A. Rakic, "An Intrusion Detection´ Method Using ADASYN and Bayesian Optimized LightGBM," 2022 34th Chinese Control and Decision Conference (CCDC), Hefei, China, 2022, pp. 4622-4627, doi: 10.1109/CCDC55256.2022.10033879

[6] J. Wang, H. Kan, F. Meng, Q. Mu, G. Shi and X. Xiao, "Fake Review Detection Based on Multiple Feature Fusion and Rolling Collaborative Training," in IEEE Access, vol. 8, pp. 182625-182639, 2020, doi: 10.1109/ACCESS.2020.3028588.

[7] N. Iksan, D. A. Widodo, B. Sunarko, E. D. Udayanti and E. Kartikadharma, "Sentiment Analysis of Public Reaction to COVID19 in Twitter Media using Na¨ıve Bayes Classifier," 2021 IEEE International Conference on Health, Instrumentation Measurement, and Natural Sciences (InHeNce), Medan, Indonesia, 2021, pp. 1-4, doi: 10.1109/InHeNce52833.2021.9537243.

[8] A. J. Nair, V. G and A. Vinayak, "Comparative study of Twitter Sentiment On COVID - 19 Tweets," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1773-1778, doi: 10.1109/ICCMC51019.2021.9418320.

[9] F. Fazrin, O. N. Pratiwi and R. Andreswari, "Comparison of K-Nearest Neighbor and Logistic Regression Algorithms on Sentiment Analysis of Covid-19 Vaccination on Twitter with Vader And Textblob Labeling," 2022 International Conference of Science and Information Technology in Smart Administration (ICSINTESA), Denpasar, Bali, Indonesia, 2022, pp. 39-44, doi: 10.1109/ICSINTESA56431.2022.10041609.

[10] Mitra, R., Bajpai, A. and Biswas, K., 2023. ADASYN-assisted machine learning for phase prediction of high entropy carbides. Computational Materials Science, 223, p.112142.

[11] Chiramdasu, Rupa, Gautam Srivastava, Sweta Bhattacharya, Praveen Kumar Reddy, and Thippa Reddy Gadekallu. "Malicious url detection using logistic regression." In 2021 IEEE International Conference on Omni-Layer Intelligent Systems (COINS), pp. 1-6. IEEE, 2021.

[12] Patel, Nidhi A., and Rakesh Patel. "A survey on fake review detection using machine learning techniques." In 2018 4th International Conference on Computing Communication and Automation (ICCCA), pp. 16. IEEE, 2018.