Journal of Tourism Economics and Applied Research

Volume: 5, Issue 2, 2021
ISSN 2602-8662 , ISBN: 978-955-703-080-7
Pages: 31-41

PERFORMANCE OF MACHINE LEARNING ALGORITHMS IN DETECTING FAKE ONLINE HOTEL REVIEWS

FHA. Shibly[1], Uzzal Sharma[2] and HMM. Naleer[3]

[1]South Eastern University of Sri Lanka
shiblyfh@seu.ac.lk
[2]Assam Don Bosco University, India
uzzal.sharma@dbuniversity.ac.in
[3]South Eastern University of Sri Lanka
drnaleer@seu.ac.lk

Abstract

Information Technology is becoming an inevitable one in all fields. It is a significant component in all vital areas such as education, political, production, science, business, Etc. Numerous activities and services are taking place online, especially in the tourism and hospitality industry. Consumers benefit from the tourist and hospitality industry's online efforts. Online customer reviews of hotel services in the tourism and hospitality industry have recently played a vital role. Online reviews of such hotels are sometimes evaluated as positive and sometimes negative. In this context, allegations arise that such online reviews are sometimes faked and lead to wrong decisions. Tourists face many inconveniences due to fake reviews thus posted without proper information of the user. This study aims to find out what machine learning algorithms can contribute to detecting online fake hotel reviews and which algorithm is more efficient. Researchers used multiple algorithms, namely Random Forest Accuracy, Random Forest Gini, KNN-5, KNN-7, Support Vector Machine (SVM), Stochastic Gradient Classifier, Extra tree classifier, AdaBoost, Decision Tree, Gradient Boost and Stacked Classifier to calculate the efficiency in detecting online fake hotel reviews. This dataset is accessible in Kaggle uploaded by Tannis Thamaiti. This dataset consists of 200 hotel reviews that are uniformly divided into 100 positive and 100 negative reviews. Based on the performance analysis, the stacked classifier has given the best accuracy, precision, recall and f1-score, which obtained 89%, 91%, 84% and 88%, respectively. Therefore, a stacked classifier will be an efficient algorithm in detecting fake hotel reviews.

Keywords: *Fake Reviews, Machine Learning, Algorithms, Detection and Classification.*

1.0 Introduction

Information technology has grown exponentially, and since then, it has been exerting its influence in many fields. The dominance of information technology is increasing day by day, especially in all sectors such as business, tourism, hospitality, education, health, transport, Etc. There is no denying that information technology is enormously contributing, especially to the tourism and hospitality sectors. Features of information technology help obtain a wide range of information such as where a tourist can visit, where he can stay and how he can travel. This technological development also creates an environment where tourists can write about their travel experiences through the website so that others can read and plan their tours accordingly. It allows other travellers to make quick decisions, save time and decide whether to visit tourist sites according to their economic status. To this end, various tourism and hospitality industry related websites have created many facilities to post other travellers' comments and mark their ratings. These internet facilities have many advantages as well as some disadvantages. The fierce competition in the tourism and hospitality industry allows many misconceptions and erroneous ratings to be uploaded on such websites.

In today's world, when consumers want to choose services or goods, reviews have surpassed all other sources of information for them. For example, when consumers decide to book a hotel, they research the opinions of previous customers about the hotel's services before making their decision. They determine whether or not to reserve a room based on the feedback they get from the reviews. After reading the reviews, they will most likely reserve the room if they received good comments. Because of this, historical evaluations become very trustworthy sources of information for most individuals in a variety of internet businesses. [1]. Online reviews are an important information source for decision making, for instance, before booking a hotel room [2]. However, how far such records are genuine and whether the real user is the one who posted them is something that cannot be determined for sure. Online reviews on sites like TripAdvisor, Yelp, and Google Places are not necessarily posted by actual consumers who have had a real hotel experience. As a result, some of the reviews are fictitious and, as a result, false. Tourists may make poor choices because of fake reviews containing misleading information. [2, 3]. Artificial intelligence technologies are beneficial in detecting such fake records. In that sense, machine learning technologies play an important role in detecting such fake reviews and comments on websites.

Fake reviews are usually detected using a combination of characteristics that are not directly related to the content and the category of reviews. Text and natural language processing (NLP) is often used to develop review features. Fake reviews, on the other hand, may need the development of additional characteristics related to the reviewer, such as review time/date or writing styles. As a result, the creation of relevant characteristics extraction from reviewers is essential for effective fake review identification. [3]

Consequently, this research seeks to find the best machine learning algorithm or algorithms for detecting fake online hotel reviews submitted on tourist and hospitality sector websites by assessing their performance and comparing the results of each algorithm.

2.0 Literature Review

The fake reviews detection problem has been tackled since 2007 [4]. In the Fake review's detection study, two major characteristics were used: textual and behavioural features. The verbal aspect of the review activity is referred to as textual characteristics. To put it another way, textual characteristics are determined mainly by the substance of the evaluations. The nonverbal aspects of the evaluations are referred regarded as behavioural characteristics. They are determined mainly by the reviewers' actions, including their writing style, emotional expressions, and the frequency with which they submit the evaluations.

Machine learning is one of the most significant technical developments, and it is at the heart of a lot of crucial applications. Machine learning's primary strength is assisting computers in learning and improving themselves based on experience [5]. Machine learning algorithms may be classified into three categories: supervised, semi-supervised, and unsupervised [6]. Both input and output data are given in the surprising method, and the training data must be labelled and categorized [7]. Unsupervised learning uses data without categorization or labels to determine the input data's best fit grouping or classification. The method must identify every data in unsupervised learning since they are unlabelled. Finally, some data are labelled, but most are not in the semi-supervised method. This section summarizes the supervised learning algorithms, which are the subject of this paper.

In [8], the authors utilized supervised machine learning methods for the identification of fraudulent reviews. SVM, Naive-Bayes, KNN, k-star, and decision tree are the classifiers utilized in this application. The authors also employed Naive Bayes, Decision tree, SVM, Random forest, and Maximum entropy classifiers in [9] to identify fraudulent reviews on the gathered dataset. Even though it is still in its infancy, much work has been done concerning several different languages. [10] [11] [12] [13].

It has been found that there were only selected supervised machine learning algorithms studied in detecting fake reviews like SVM, KNN, NB, K-Star, DT and LR Etc., in the recent past. However, there are some more algorithms that can be used in sentiment analysis and detection issues. Therefore, researchers used multiple algorithms, namely Random Forest Accuracy, Random Forest Gini, KNN-5, KNN-7, Support Vector Machine (SVM), Stochastic Gradient Classifier, Extra tree classifier, AdaBoost, Decision Tree, Gradient Boost and Stacked Classifier to calculate the efficiency in detecting online fake hotel reviews. So, the performances of each algorithm can be calculated and compared to find out a decision related to this problem.

Table I presents comparative research on classification algorithms to demonstrate the most effective technique for identifying fake reviews using a variety of datasets.

Table 1: A comparison of several algorithm research.

Reference No	Size of the Dataset	Algorithm	Language	Efficient Method
14	2000 Reviews	IBK, DT, NB, SVM,	English	SVM
15	1000 Reviews	SVM, NB	English	NB
16	4000 Reviews	SVM, Delta, LLR, KNN, SVM, NB	Chinese	NB
17	1400 Reviews	SVM, NB	English	NB
18	2000 Reviews	SVM, KNN, NB, DT – J48	English	SVM
19	5853 Reviews	LR. NB, KNN, SVM, RF	English	KNN
20	205 Reviews	NB, LR	English	NB

Source: (Literature Review)

3.0 Methodology

This section exhibits the details of the proposed method shown in figure 1 in more depth. The suggested method is divided into six main stages that must be completed to get the best model utilized for fake review identification. The following sections explain these phases:

A. Dataset.

The experiment focused on assessing the sentiment value of the standard dataset was designed to offer an extensive examination of machine learning algorithms in a controlled environment. The original data set of the hotel fake and unfaked review is used to evaluate our categorization techniques for hotel reviews. This dataset is accessible in Kaggle uploaded by Tannis Thamaiti [21]. This dataset consists of 200 hotel reviews that are uniformly divided into 100 positive and 100 negative reviews.

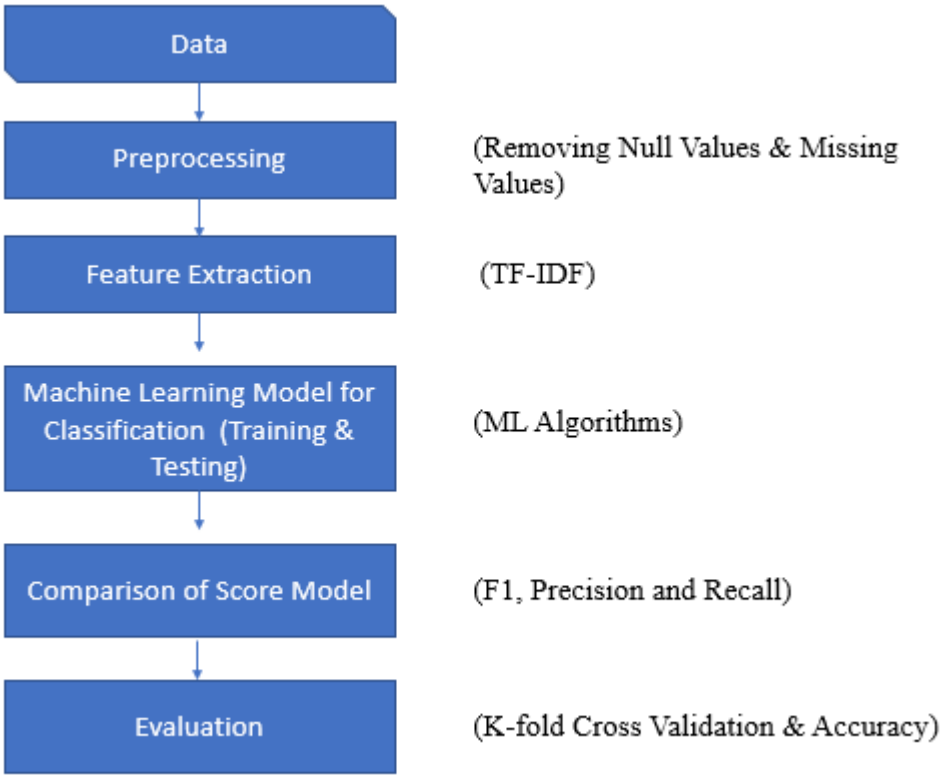
B. Pre-processing

The pre-processing phase includes preparatory activities that aid in the transformation of data before the real analysis. To illustrate the impact of pre-processing on classification models, researchers applied to remove null values, remove missing values, and check data duplication.

C. Feature Extraction

In this paper, we consider all the features which were used to classify the review available in the dataset and their impact on the performance of the fake review's detection process.

Figure 1: Relationship between sustainable and ecotourism



Source: (Literature Review)

D. Machine Learning Model for Classification

To identify fake and unfaked reviews, multiple algorithms were applied into the dataset namely Random Forest Accuracy, Random Forest gini. KNN-5, KNN-7, SVM, Stochastic Gradient Classifier, Extra tree classifier, AdaBoost, Decision Tree, Gradient Boost and Stacked Classifier to find out the most efficient algorithm in detecting fake reviews.

E. Comparison of Score Model

Using the above selected machine learning algorithms, it is a must to compare the performance of each algorithm in terms of F score, precision and recall identifying the most efficient one.

F. Evaluation

The performance of algorithms and the research analysis should be evaluated to validate the results and findings. Otherwise, the reliability of the research will be low. Therefore, researchers used 10-fold cross validation technique and calculate the accuracy of each algorithm.

4.0 Findings

This section presents the experimental results of eleven different supervised machine learning approaches to find out the best algorithm. Accuracy, precision, and recall are three metrics used to assess the performance of fake review detection. In this case, accuracy refers to the total accuracy of several sentiment models. True positive reviews have a high recall (Pos) and precision (Pos) ratio, the ratio of true positive reviews to false positive reviews. The recall (Neg) and precision (Neg) ratios for true negative reviews are the recall and precision ratios, respectively.

One of the most significant stages in data mining and data retrieval is evaluating the accuracy of machine learning classifiers. It is common practice to use the error rate and the F-measure to assess a classifier's ability to identify the appropriate category or class of unknown instances correctly. The error rate is the proportion of occurrences in the test set that were incorrectly classified. This data collection will be referred to as "X," and the number "m" will indicate the number of occurrences incorrectly categorized by a classification model C. It is possible to assess the accuracy of C in choosing the proper classes of X instances by applying the following formula to the data:

$$Accuracy(C) = \frac{m}{n} \quad (1)$$

In machine learning, the error rate method does not account for the cost of making incorrect predictions. Most of the time, the F-measure is employed to address this issue. Precision and recall are two fundamental measures that are utilized to evaluate the value of the F-measure. Consider the scenario in which part of the data in the test set belongs to a particular class or category S of data. It categorizes each piece of test data and gives a label to it. There will be four different types of forecasts for the test set S: Precision is defined as the percentage of data for category S that is correctly forecasted. The recall rate for category S is defined as the percentage of adequately projected actual data that were correctly predicted. When accuracy and recall are considered, it is feasible to compute the F-measure (2-4).

$$Precision = \frac{|TP|}{|TP|+|FP|} \quad (2)$$

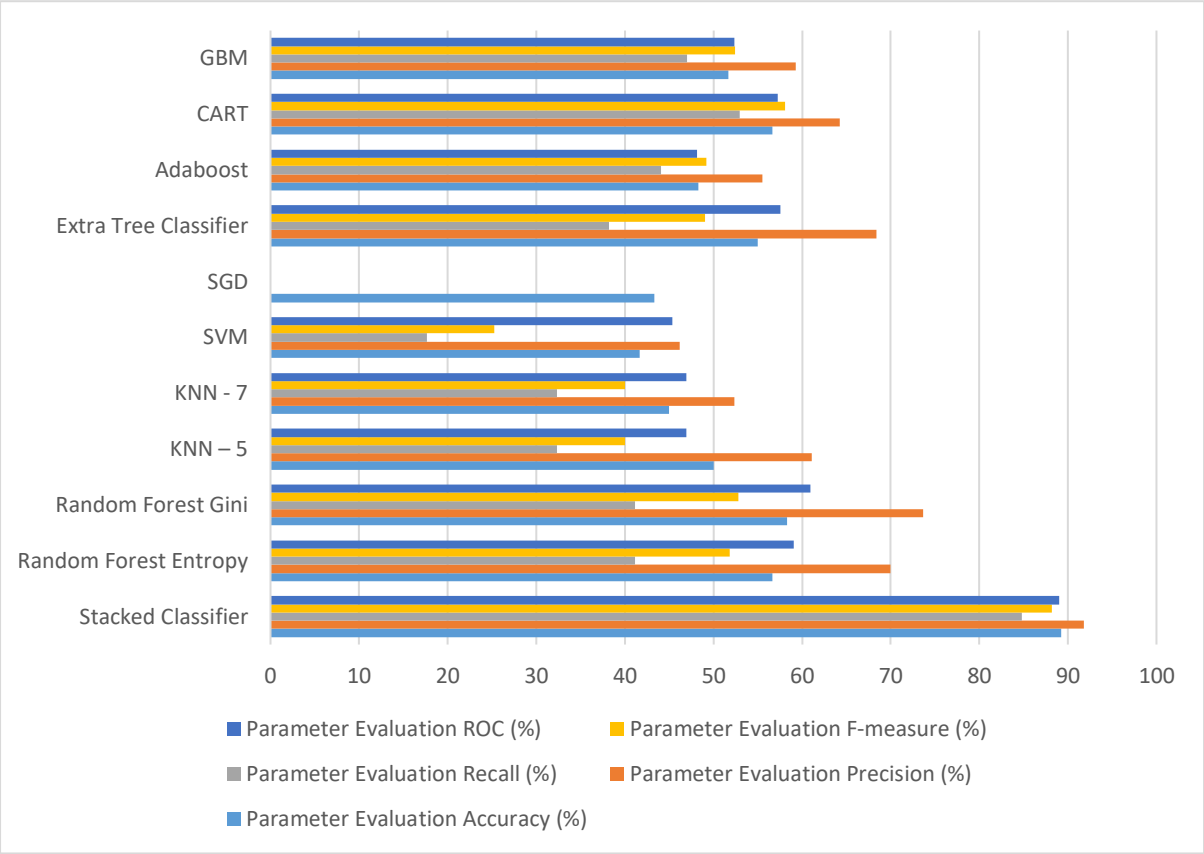
$$Recall = \frac{|TP|}{|TP|+|FN|} \quad (3)$$

$$F - measure = 2 \cdot \frac{Precision.Recall}{Precision+Recall} \quad (4)$$

Performance Analysis and Comparison

In this analysis, all eleven algorithms which were used in this research have been analyzed in terms of accuracy, recall, precision, and F-measure. Figure 2 and Table 2 show the comparative analysis of proposed shows the graphical representation of the comparative analysis.

Figure 2: Performance Analysis and Comparison



Source: (Data analysis in Google Colab)

Table 2: Analysis and comparison of the performances of machine learning algorithms

Machine Learning Algorithm	Parameter Evaluation					
	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)	ROC (%)	Log_Loss
Stacked Classifier	**89.28**	**91.80**	**84.84**	**88.18**	**89.04**	**3.70**
Random Forest Entropy	56.66	70	41.17	51.85	59.04	14.96
Random Forest Gini	58.33	73.68	41.17	52.83	60.97	14.39
KNN – 5	50	61.11	32.35	40	46.94	18.99
KNN - 7	45	52.38	32.35	40	46.94	18.99
SVM	41.66	46.15	17.64	25.23	45.36	20.14
SGD	43.33	0	0	0	0	19.57
Extra Tree Classifier	55	68.42	38.23	49.05	57.57	15.54
Adaboost	48.33	55.55	44.11	49.18	48.18	17.84
CART	56.66	64.28	52.94	58.06	57.23	14.96
GBM	51.66	59.25	47.05	52.45	52.37	16.69

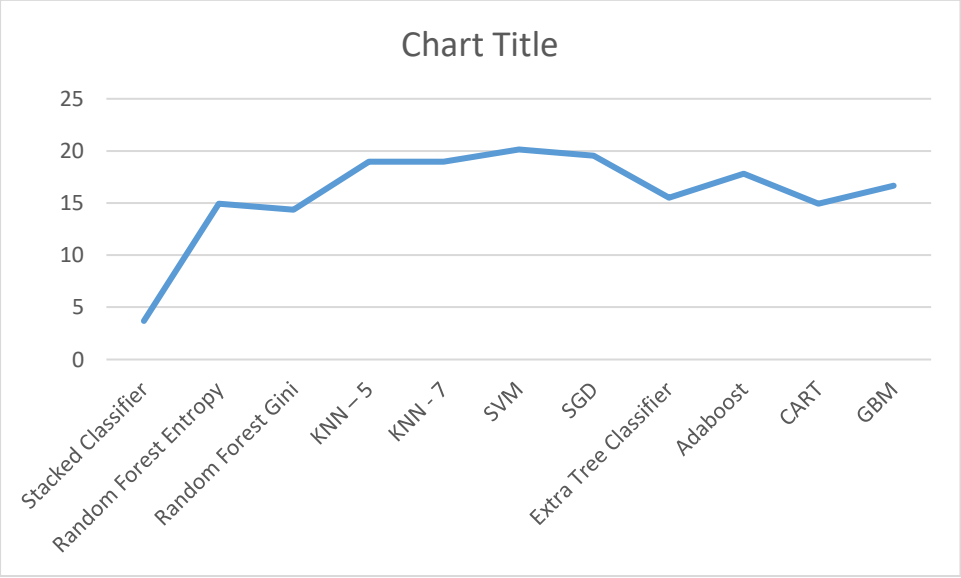
Source: (experimental analysis using google colab)

Multiple algorithms were implemented to identify fake and unfaked reviews, namely Random Forest Accuracy and Random Forest Gini. KNN-5, KNN-7, SVM, Stochastic Gradient Classifier, Extra tree classifier, AdaBoost, Decision Tree, Gradient Boost and Stacked Classifier.

From the figure 2 and table 2, it is observed that the stacked classifier achieved better performance than other algorithms. For accuracy analysis, the Random Forest Entropy, Random Forest Gini, and CART algorithms have achieved nearly 58%, but the stacked classifier achieved 89.28 % of accuracy. When compared with all techniques, SGD achieved low precision (0 %), whereas Random Forest Gini and Random Forest Entropy achieved nearly 70% precision. However, the stacked classifier achieved 91.80 % of precision. In addition, the stacked classifier achieved 84.84% of recall and 88.18 % of F-measure, where the other algorithms achieved less than 60 % of recall and F-measure.

For a more accurate assessment of performance and to justify the most efficient method from various perspectives, the researchers looked at the Receiver Operating Characteristic (ROC) curve and Log Loss curves. In general, a ROC of 0.5 indicates no discrimination, 0.7 to 0.8 indicates good discrimination, 0.8 to 0.9 indicates excellent discrimination, and more than 0.9 indicates exceptional discrimination [22]. According to Table 2, the stacked classifier achieved nearly 90%, and undoubtedly it outperformed other algorithms in ROC.

Figure 3: Log Loss



Source: (Data analysis)

The stacked classifier (Figure 3) scored 3.7 based on Log Loss once again. The most significant probabilities-based classification measure is log loss. Although raw log-loss numbers are difficult to understand, log-loss is a helpful measure for comparing models. Lower log loss values indicate better forecasts for any given issue [23]. Therefore, the stacked classifier has better prediction skills in terms of log loss as well.

5.0 Conclusions & Recommendations

In this paper, researchers applied several methods to analyze a dataset of hotel reviews and supervised learning used in our work. This research looked at the accuracy of all feasible classification algorithms and how to figure out which one is the most accurate. Our findings also indicate that we were able to identify positive and negative reviews using detection methods.

Reviews influence people's decisions. As a result, detecting false reviews is an active and ongoing research topic. A machine learning strategy for detecting fake reviews is provided in this research. The Yelp dataset is utilized to assess the suggested method. The developed technique employs a variety of classifiers. This article compares eleven supervised learning methods for evaluating the performance of identifying fake reviews in the dataset. Multiple algorithms were implemented, namely Random Forest Accuracy, Random Forest Gini. KNN-5, KNN-7, SVM, Stochastic Gradient Classifier, Extra tree classifier, AdaBoost, Decision Tree, Gradient Boost and Stacked Classifier.

In this paper, our experiments have shown the accuracy of results through sentiment classification algorithms, and we have found that the stacked classifier is outperformed all other algorithms in terms of accuracy, precision, recall, f1 score, ROC, and log loss. Previous studies have recommended SVM [14, 18], NB [15,16,17,20] and KNN [19] are the efficient algorithms in different methodological analysis, but this research proved that stacked classifier can be a most efficient algorithms which can be used in fake review detection rather than other supervised algorithms of machine learning. Saso D Zeroski and Bernard Zenko [24] also endorsed that combining classifier is better than others. There are some researchers [25, 26] measures the performance of stacked classifier and concluded it’s a best classifier in text classifications. Therefore, the hotel review system incorporates the machine learning techniques with a stacked classifier; the fake reviews can be identified and removed to maintain a simple review system. So, Tourists can take perfect decisions regarding the hotels, and they can enjoy their travel and tour without any hesitations.

Researchers recommend to the hospitality and tourism industry that there are many features and aspects available in machine learning which can enhance the operation of tourism and hospitality in many ways. Specially sentiment analysis [27] and machine learning models [28] will be very useful techniques to rebuild the industry in post pandemic actions. Also, researchers recommend to apply the machine learning applications in automatic travel and hotel suggestions based on customers’ past history, automatic comparison and real time ratings by users reviews,

References:

Ayodele, T.O. "Types of machine learning algorithms," in New advances in machine learning. InTech, 2010.

Bhandari, A. (2020, July 20). AUC-ROC Curve in Machine Learning Clearly Explained. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/

Casalo LV, Flavian C, Guinaliu M, Ekinci Y (2015) Do online hotel rating schemes influence booking behaviors? Int J Hosp Manag 49:28–36 Casalo LV, Flavian C, Guinaliu M, Ekinci Y (2015) Do online hotel rating schemes influence booking behaviors? Int J Hosp Manag 49:28–36

Choi S, Mattila AS, Van Hoof HB, Quadri-Felitti D (2017) The role of power and incentives in inducing fake reviews in the tourism industry. J Travel Res 56(8):975–987

Chu, Chun-Han, Chen-Ann Wang, Yung-Chun Chang, Ying-Wei Wu, Yu-Lun Hsieh, and Wen-Lian Hsu. "Sentiment analysis on Chinese movie review with distributed keyword vector representation." *In Technologies and Applications of Artificial Intelligence (TAAI),* 2016 Conference on, pp. 84-89. IEEE, 2016.

Dewangan, A., & Chatterjee, R. (2018). Tourism recommendation using machine learning approach. In Progress in Advanced Computing and Intelligent Engineering (pp. 447-458). Springer, Singapore.

Džeroski, S., & ŽEnko, B. (2004). Is Combining Classifiers with Stacking Better than Selecting the Best One? Machine Learning, 54(3), 255–273. https://doi.org/10.1023/b:mach.0000015881.36452.6e

Elmogy, A., Tariq, U., Mohammed, A., & Ibrahim, A. (2021). Fake Reviews Detection using Supervised Machine Learning. *International Journal Of Advanced Computer Science And Applications*, 12(1). doi: 10.14569/ijacsa.2021.0120169

Elmurngi.E, and Gherbi. A., Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques. IARIA/DATA ANALYTICS, 2017.

Elmurngi, E., & Gherbi, A. (2017, August). 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.

Elmogy, A. M., Tariq, U., Mohammed, A., & Ibrahim, A. (2021). Fake Reviews Detection using Supervised Machine Learning. Int. J. Adv. Comput. Sci. Appl, 12.

Fujii, Atsushi, and Tetsuya Ishikawa. "A system for summarizing and visualizing arguments in subjective documents: Toward supporting decision making." *In Proceedings of the Workshop on Sentiment and Subjectivity* in Text, pp. 15-22. Association for Computational Linguistics, 2006.

Hassan, Sundus, Muhammad Rafi, and Muhammad Shahid Shaikh. "Comparing svm and NBclassifiers for text categorization with wikitology as knowledge enrichment." *In Multitopic Conference (INMIC), 2011 IEEE 14th International*, pp. 31-34. IEEE, 2011.

Jindal. N., and Liu, B. "Review spam detection*," in Proceedings of the 16th International Conference on World Wide Web*, ser. WWW '07, .2007.

Kulkarni, A., Samant, R. M., Barve, P., & Phade, A. A Machine Learning Approach to Building a Tourism Recommendation System using Sentiment Analysis. *International Journal of Computer Applications*, 975, 8887

Ku, Lun-Wei, Yu-Ting Liang, and Hsin-Hsi Chen. "Opinion Extraction, Summarization and Tracking in News and Blog Corpora." In AAAI spring symposium: Computational approaches to analyzing weblogs, vol. 100107. 2006.

Kalaivani, P., and K. L. Shunmuganathan. "Sentiment classification ofmovie reviews by supervised machine learning approaches." *Indian Journal of Computer Science and Engineering* 4, no. 4 (2013): 285-292.

Liu, Bing, Minqing Hu, and Junsheng Cheng. "Opinion observer: analyzing and comparing opinions on the web." *In Proceedings of the 14th international conference on World Wide Web*, pp. 342-351. ACM, 2005.

Li, H. (2005b). Combined Multiple Classifiers Based on a Stacking Algorithm and Their Application to Chinese Text Chunking. Journal of Computer Research and Development, 42(5), 844. https://doi.org/10.1360/crad20050519

Liu, X. J., Gong, X. J., Yu, H., & Xu, J. H. (2018). A Model Stacking Framework for Identifying DNA Binding Proteins by Orchestrating Multi-View Features and Classifiers. Genes, 9(8), 394. https://doi.org/10.3390/genes9080394

Michie.D, Spiegelhalter D.J.& Taylor C.et al., "Machine learning,"Neural and Statistical Classification, vol. 13, 1994.

Molla, Y. Biadgie, and K.-A. Sohn, "Detecting Negative Deceptive, Opinion from Tweets." in International Conference on Mobile and Wireless Technology. Singapore: Springer, 2017.

M. (2020, November 20). Log Loss - Logistic Regression's Cost Function for Beginners. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/11/binary-cross-entropy-aka-log-loss-the-cost-function-used-in-logistic-regression

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." *In Proceedings of the ACL-02 conference on Empirical methods in natural language processing*-Volume 10, pp. 79-86. Association for Computational Linguistics, 2002

Singh, V. K., R. Piryani, A. Uddin, and P. Waila. "*Sentiment analysis of Movie reviews and Blog posts." In Advance Computing Conference (IACC),* 2013 IEEE 3rd International, pp. 893-898. IEEE, 2013.

Sandifer, A. V., Wilson, C., & Olmsted, A. (2017, December). *Detection of fake online hotel reviews. In 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)* (pp. 501-502). IEEE.

Sebastiani, F. "Machine learning in automated text categorization," ACM computing surveys (CSUR), vol. 34, no. 1, pp. 1–47, 2002.

Yelp dataset based on fake reviewers. (2018, November 20). [Dataset]. Tannis Thamaiti. https://www.kaggle.com/tannistha/yelp-dataset-based-on-fake-reviewers/code