**The Islamia University Bahawalpur**



**Data Mining Project Submitted to:**

**Sir Muhammad Bux Alvi Submitted By:**

**Mubashra Munir**

**18CS01**

**Machine Learning Based Fake Review Prediction**

**Abstract**

Fake reviews are increasing as trend is changing towards online sale and purchase. To detect fake review is critical as well as challenging for both researchers and online retailers. As new techniques are introduced to catch the fake reviewer, so are their intruding approach. In this report we have combined different features along with sentiment score and design a model to check the performance under different classifiers. Supervised learning approach is used to train the data set and six classification algorithms are applied. Results shows that combination of selected features is a better approach in fake review detection.

# Introduction

Fake review is a false judgement or an opinion on a product or a service. According to statista statistics, e-commerce sales increases 6% in America from 2013-2020[[1]](#footnote-1). As online purchase is increasing so is the competition of online retailer giants. Fake reviews are serious threat for both researchers [1] and online retailers [2]. Fake reviews can either be positive to increase purchase on an online platform by manipulating users by fake customer’s review. In a survey, 80% users believe on reviews as much as their personal recommendations before purchasing any product[[2]](#footnote-2). Negative fake reviews are used to defame competitor’s reputation. There are freelancers and companies for writing fakes reviews.

Giant retailers like Amazon find these fake reviews of great threat to their reputation and filed a complaint against review spamming[[3]](#footnote-3). Another interesting example is of the Italian magazine that wrote fake positive review against a fictitious restaurant on TripAdvisor and as a result that restaurant gain popularity among all competitors[[4]](#footnote-4).

Fake review prediction can be performed manually or automatically. Research work has been carried out on manual opinion spam prediction for several years [5]. Early methods of fake review prediction were rudimentary and easily caught. Many text analysis based approaches are found in literature [6]. Based on the research work, commercial platforms developed opinion spam filtering systems to detect the deceptive reviews. But these systems make the fake reviewers to enhance the quality of their review and deceive the detecting systems [7]. As the time elapsed, those traditional approaches would not work capably because the fake reviewers started behaving like normal users. Therefore the trend of manual fake review prediction changed from text based analysis to pattern and feature analysis like time [8], topics [9], ranking pattern [10], activity volume [11] and geolocation [12]. However manual methods are slow, expensive and of low accuracy. Automated methods based on machine learning could also be used to identify the opinion spams and spammers by analyzing the review features. Text mining and Natural Language Processing (NLP) work together to generate the concept of content mining and review spam detection comes under this concept. Additional review characteristics like review timings, reviewer id, and deviation trend of the review from other reviews of same category are also considered in spam review detection. Jindal et al. [13] used machine learning technique and showed that the amalgam of features are more robust than single feature for fake review prediction. Li et al. [14] showed that combining bag of words with more general features perform better than bag of [1](https://www.statista.com/statistics/379112/e-commerce-share-of-retail-sales-in-us/)

<https://www.statista.com/statistics/379112/e-commerce-share-of-retail-sales-in-us/> Accessed 26 Feb 2020 2

https:// www.brightlocal.com/research/local-consumer-review-survey/. Accessed 26 Feb 2020 3

http://goo.gl/oz13iZ, Accessed 26 Feb 2020 4

http://goo.gl/NtSKpi Accessed 26 Feb 2020

words alone. Mukherjee et al. [15] used machine learning with abnormal behavioral features of the reviewers and depicted that this technique is better than linguistic features based technique.

All these machine learning techniques aim to make the fake review prediction more robust and reliable.

## Motivation

Reviews can greatly affect the decision of a buyers while shopping online. Therefore the retailers and manufacturers take these reviews on a serious note. Fake reviewers capitalize this opportunity to artificially devalue or promote products and services [16, 17]. Hence fake review prediction is a critical research area as online purchase is increasing. With the explosive growth of e-commerce and other online businesses, the quantity and importance of reviews continuous to increase. Manual review methods being very slow and expensive are not viable for large online businesses. Machine learning based techniques become handy in automated prediction of poison reviews. These techniques can minimize the manipulation of reviews for profit or gain.

## Research Contribution

In this research, we will use machine learning technique which will employ heuristic optimization algorithm for affecting features and will test its reliability and robustness against existing techniques.

# Literature Review

Advancements have been made in the field of fake review detection by introducing new techniques and methods by researchers. These techniques are playing their role by improving accuracy and performance. So far reviews are marked as spam on the basis of either review spam detection or reviewer spam detection. Both techniques are useful in fake review detection. Prior deals with content mining and natural language processing (NLP) whereas later technique applied on reviewer id and his behavior.

Jindal et al. [13] is the first researcher who studied opinion spamming using supervised learning. The author divided the reviews into 3 categories (fake opinion, brand only reviews and non-review) and detect opinion spamming by finding duplicate reviews by a method of w-shingling. The author used dataset from Amazon with more than 5 million reviews of product and applied his devised technique with logistic regression algorithm and achieved AUC of 78%.

Lim, Nguyen, Jindal, Liu and Lauw [18] proposed a behavioral methodology for the revealing of spammers for review. They've tried to figure out some spammer habits like they're targeting goods and trying to optimize their effect. And they suggested a model focused on certain patterns to identify rating spammers.

Ott et al.[19, 20] created a data set for analysis in the field of review spam detection. The data set comprises of positive opinion spam with truthful reviews and negative opinion spam with real review. The author applied the n-gram and linguistic features to find fake review under supervised learning mechanism and results are verified with human performance.

In their research, Feng et al. [1] framed a model based on normal distribution of opinion to detect fake reviews. According to their opinion a product or a service review involved this concept of normal distribution of opinion.

Shojaee et al. [6] suggested a novel technique for fake review detection by joining two features together, Lexical and Synthetic. Lexicals are word based and reviewer’s way of writing using at the sentence level is synthetic feature.

Elmurngi and Gherbi [21] proposed text classification and sentimental analysis approach for different machine learning algorithms with stopwords and without stopwords. They also applied decision tree algorithm to improve their results.

Shah, Ahsan, Kafi, Nahian and Hossain [22] combined Supervised & Active learning and created a model to detect spamming. Both fictitious and real life data is used for spam analysis.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Reference of Paper | Year | Data set | Learning type | Techniques/Algorithm | Results | Limitations |
| [13] Jindal N,  Liu B | 2007 | Data set of manufactur -ed product only | Supervised Learning | Logistic  Regression | average  AUC 78% | Lack of accuracy of real world data set |
| [18] Lim,  Ee-Peng | 2010 | Amazon  Data set | Supervised Learning | Behavioral features of  Spammer | Accuracy 78% | Limit set of data for supervised learning |
| [20] Jeffrey T.  Hancock, Ott, Claire Cardie, and Myle | 2013 | Ott et al. data set | Supervised Learning | Support Vector  Machine (SVM) | Accuracy 86 % | Human judgments can be imperfect and biased. |
| [1] Feng | 2012 | Ott et al. data set with modificatio n | Supervised Learning | LIBSVM  classifier/  Term frequency | Accuracy  72.5% | Specific kind of dataset |
| [6] Shojaee | 2013 | Ott et al. data set | Supervised Learning | SVM/ Naive  Bayes/ Stylometric  Feature | F-measure 84% | Limited to specific domain |
| [21] Elmurngi  E. and Gherbi  A | 2017 | Movie  review data set | Supervised Learning | Decision Tree  (DT-J48)/ SVM/  KNN | Accuracy without stopwords 81.75% | Feature selection methods are not used |
| [22]Ahsan,  Nahian, Kafi,  Hossain and  Shah | 2017 | Ott dataset | Active/  Supervised  Learning | Hybrid classifier using Naïve Bayes (NB)/ Support  Vector Machine  (SVC) / Decision  Tree (DT) /  Maximum Entropy  (Maxent) | Accuracy 95% | Small scale dataset  is used for specific domain |

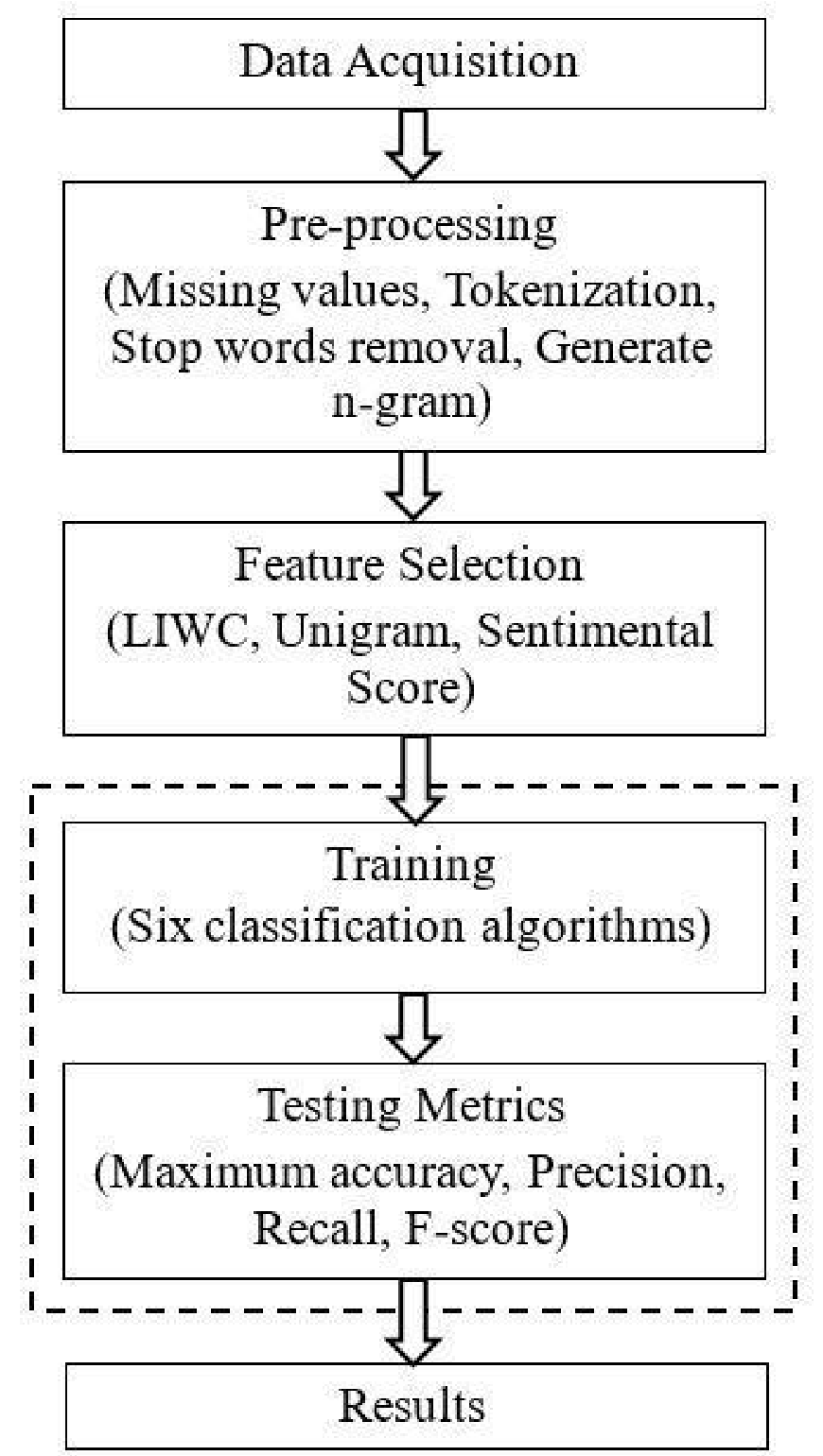
**Table 1: State of Art Spam Review Detection Techniques**

# Proposed approach

In this research work, two sets of features are used as a classification criteria with an additional feature called as sentiment score. These individual features and their different combinations are used to train various classifiers and tested against evaluation metrics. Reviews are classified into spam and truthful and supervised learning approach is used to attain accuracy. In this study, two classification algorithms Naive Bayes, support vector machine (SVM) are used.

## Data acquisition

The data set selected for this research contain 1600 reviews. This data set contains eight hundred truthful reviews of which four hundred are positive and four hundred are negative.

Similarly 800 spam reviews are also included in this data set of which half are positive and half are negative.

**Figure 1: Proposed methodology for fake review detection.**

## Pre-processing

Pre-processing of data set have a significant effect on the accuracy of results [23]. Therefore some pre-processing techniques like missing values, tokenization, stop words removal and generate n-gram are implemented on the data set.

## Classification algorithms

Six various classification algorithms are used in this paper in order to determine the effect of different features and their combinations on the classification accuracy and performance.

***3.4.1 Naïve Bayes (NB)***

NB is based on Bayes theorem [26]. It is a probabilistic multiclass classification algorithm with the assumption of features independency to foresee the output class. Following formula is used to check probability of a set of features to be categorized into class A:



where x = (x1,x2,…,xn) are a set of features. Individual probability of a feature in a given class A is calculated as:



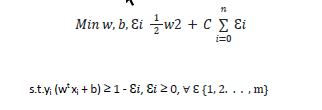
(3)

***3.4.4 Support Vector Machine (SVM):***

SVM is a binary classification algorithm based on finding the maximum margin hyperplane to classify signify the ith vector and **yi** Rn signify the target. Optimal𝑌𝑖, 𝑋𝑖 hyper plane finds by linear features between two classes (0 or 1). Let’s say training data = { }; i = 1, 2………, n where

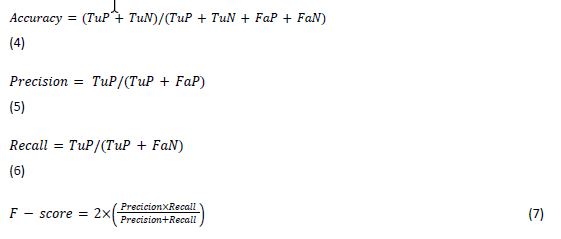
SVM𝑋𝑖 ∈𝑅𝑛in the form of equation in which∈ b consider as offset and we consider as dimensional coefficient vector. At the end this is completed by solving following equations:

Equation:



## Testing metrics

Accuracy, precision, recall and f-score are used for evaluation of performance. These metrics can be defined as:

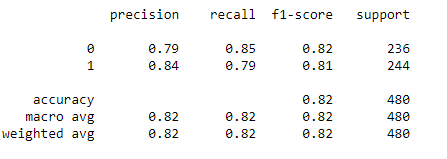


Where TuN, TuP, FaN and FaP are true negative, true positive, false negative and false positive

respectively.

# Experiment, Results and Discussion / Evaluation

The dataset described is used to obtain our experimental results. The results are compiled using jupyter notebook. The outcome from SVM(Support Vector Machine) classifier gives better results. 10% accuracy is achieved when (x,y\_train) is used but when (y,y\_test) is used, accuracy increases to 81.6%. Classifier Multinomial NB gives accuracy of 97.1% when we use (x,y\_train) but (y,y\_test) is used it is 85.62%.



# Conclusion

In this work, we tried to find out which combination of features performs well for spam review detection. The classification algorithms we prefer in our work are described in section 3.4. If classifiers performance is analyzed, NB gives better results as compare to SVM along. For future work, semi supervised learning can be used to check the accuracy and performance of SVM and NB on fake review detection method. In this way not only performance is tested but the limitation of labelled data set for supervised learning is also resolved.

# References

1. Feng S, Xing L, Gogar A, Choi Y (2012) Distributional footprints of deceptive product reviews. The International Conference on Weblogs and Social Media (ICWSM) 12:98–105.
2. Sussin J, Thompson E (2012) The consequences of fake fans, likes, and reviews on social networks.

Gartner, Inc., July 1–7

1. Stempel J (2015) Amazon sues to block alleged fake reviews on its website.

[4]Fenton S (2015) Tripadvisor denies rating system is flawed, after fake restaurant tops rankings in italy.

[5]. Spirin, N., Han, J. (2012) Survey on web spam detection: principles and algorithms. SIGKDD Explor. Newslett. 13, 50–64.

[6]. Shojaee, S., Murad, M.A.A., Azman, A.B., Sharef, N.M., Nadali, S. (2013) Detecting deceptive reviews using lexical and syntactic features. In: 13th International Conference on Intelligent Systems Design and Applications, Malaysia, pp. 53–58.

1. Yao, Y., Viswanath, B., Cryan, J., Zheng, H., Zhao, B.Y. (2017) Automated crowdturfing attacks and defenses in online review systems. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, USA, pp. 1143–1158.
2. KC, S., Mukherjee, A. (2016) On the temporal dynamics of opinion spamming: case studies on yelp. In: Proceedings of the 25th International Conference on World Wide Web, pp. 369–379. Republic and Canton of Geneva, Switzerland.
3. Nilizadeh, S., et al. (2017) Poised: spotting twitter spam off the beaten paths. In: Proceedings of the

2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, USA, pp.

1159–1174.

1. Chen, H., He, D., Zhu, S., Yang, J. (2017) Toward detecting collusive ranking manipulation attackers in mobile app markets. In: Proceedings of the 2017 ACM on Asia Conference on Computer and

Communications Security, Abu Dhabi, UAE, pp. 58–70.

1. Chino, D.Y.T., Costa, A.F., Traina, A.J.M., Faloutsos, C. (2017) VolTime: unsupervised anomaly detection on users’ online activity volume. In: Proceedings of the 2017 SIAM International Conference on Data Mining, Houston, USA, pp. 108–116.
2. Ruoyu Deng, Na Ruan, Ruidong Jin, Yu Lu, Weijia Jia, Chunhua Su, Dandan Xu (2019) SpamTracer: Manual Fake Review Detection for O2O Commercial Platforms by Using Geolocation Features. Proceedings of International Conference on Information Security and Cryptology, China, pp 384-403.
3. Jindal N, Liu B (2007) Review spam detection. In: Proceedings of the 16th international conference on World Wide Web (pp. 1189–1190). ACM, Lyon, France.
4. Li J, Ott M, Cardie C, Hovy E (2014) Towards a general rule for identifying deceptive opinion spam. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pages 1566–1576, Baltimore, Maryland, USA, June 23-25 2014. ACL.
5. Mukherjee A, Venkataraman V, Liu B, Glance NS (2013) What yelp fake review filter might be doing? Boston, International Conference on Weblogs and Social Media (ICWSM).
6. Lau RY, Liao SY, Kwok RCW, Xu K, Xia Y, Li Y (2011) Text mining and probabilistic language modeling for online review spam detecting. ACM Trans Manage Inf Syst 2(4):1–30.
7. Li F, Huang M, Yang Y, Zhu X (2011) Learning to identify review spam. In: IJCAI

Proceedings-International Joint Conference on Artificial Intelligence, vol 22, No. 3., p 2488

1. Lim, Ee-Peng, et al.: Detecting product review spammers using rating behaviors. In: Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010.
2. Ott, Myle, et al.: 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-Volume 1. Association for Computational Linguistics, 2011.
3. Ott, Myle, Claire Cardie, and Jeffrey T. Hancock.: Negative Deceptive Opinion Spam. In: HLTNAAC L. 2013
4. Elmurngi E. and Gherbi A.,(2017),“An Empirical Study on Detecting Fake Reviews”, proceeding of IEEE The 7th International Conference on Innovative Computing Technology, pp. 107-114.
5. Ahsan M.N.I., Nahian T., Kafi A.A., Hossain I., Shah F.M., (2017), “An Ensemble approach to detect review spam using hybrid machine learning technique”, IEEE Nineteenth International conf. on computer and Information Technology, pp. 381- 388.
6. Etaiwi, W., & Naymat, G. (2017). The impact of applying different preprocessing steps on review spam detection. Procedia computer science, 113, 273-279.

1. <https://www.statista.com/statistics/379112/e-commerce-share-of-retail-sales-in-us/> Accessed 26 Feb 2020 [↑](#footnote-ref-1)
2. https:// www.brightlocal.com/research/local-consumer-review-survey/. Accessed 26 Feb 2020 [↑](#footnote-ref-2)
3. http://goo.gl/oz13iZ, Accessed 26 Feb 2020 [↑](#footnote-ref-3)
4. http://goo.gl/NtSKpi Accessed 26 Feb 2020 [↑](#footnote-ref-4)