

# Graduation Project Proposal

## Rating user reviews

Yara Ahmed Nader, Marim Mohamed, Amira Ashraf, Yasmine Maged  
Supervised by Dr.Ammar Mohammed, Eng.Huda Mohammed

January 10, 2019

### **Abstract**

Sales of products and services are affected by customer opinion. If it is positive selling rate will increase on the contrary if the reviews are negative, the selling rate will decrease. Nowadays, people are expressing their opinion by writing a review about the product so we can decide whether this product is good for purchase or not. Therefore, it is clear that a customer review has a significant impact on the reputation of the product and its suppliers, so fake reviews can damage a good product or raise a bad product. The aim of this project is to detect fake reviews from real one.

## **1 Introduction**

### **1.1 Background**

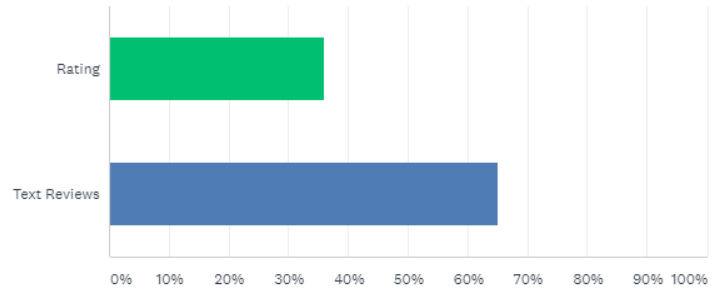
Fake reviews are the reviews which is designed specifically to give a false impression to consumers on the point of purchasing. In 2013 European Consumer Centres' Network web survey showed that 82% of respondents read consumer reviews before shopping. But the problem of fake online reviews not only concerns individual consumers but it can also lead to an erosion of consumer confidence in the online market, which can reduce competition. There are two types of fake reviews. A positive fake review which can give you a false image to persuade you to buy a product or a service and a negative fake review which is written for to damage the company's reputation and go on to recommend a different product from a competitor. This document proposes Detecting user fake reviews to provide a real positive and real negative reviews system.

### **1.2 Motivation**

Reviews are very important for both customers and service providers. Fake reviews should be detected since they might aect the product and the reputation of the service provider. The aim of this project is to detect fake reviews.

### what is more accurate rating or texting reviews ?

Answered: 100 Skipped: 0

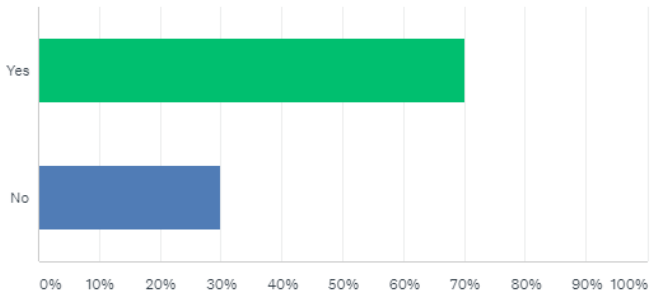


ANSWER CHOICES	RESPONSES
Rating	36.00% 36
Text Reviews	65.00% 65
Total Respondents: 100	

Figure 1: Statics 1

### Have you ever been misled by reviews?

Answered: 100 Skipped: 0

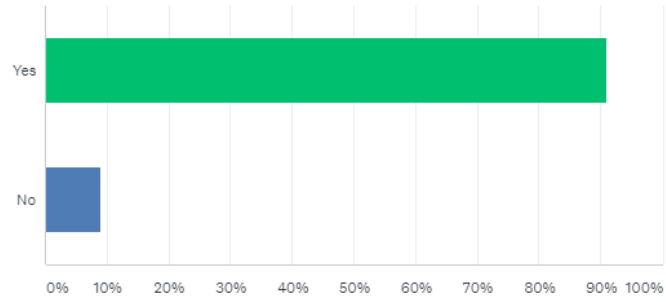


ANSWER CHOICES	RESPONSES
Yes	70.00% 70
No	30.00% 30
TOTAL	100

Figure 2: Statics 2

## Do you read product reviews before buying?

Answered: 100 Skipped: 0

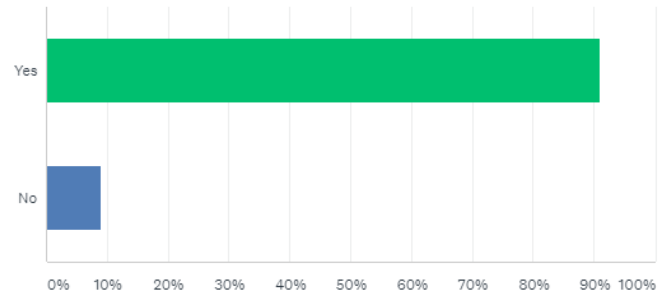


ANSWER CHOICES	RESPONSES
▼ Yes	91.00% 91
▼ No	9.00% 9
Total Respondents: 100	

Figure 3: Statics 3

Do you think it would be more accurate if there is a numerical ratings based on your real text reviews?

Answered: 100 Skipped: 0



ANSWER CHOICES	RESPONSES
Yes	91.00% 91
No	9.00% 9
<b>TOTAL</b>	<b>100</b>

Figure 4: Statics 4

### 1.3 Problem Definitions

In one of similar papers to this system the problem that was faced in it is how to extract emotions inside opinion and how to detect fake positive reviews and fake negative reviews from opinion reviews. they compare 5 supervised machine learning algorithm ( Nave Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN-IBK), KStar (K\*) and Decision Tree (DT-J48) )for sentiment classification of reviews and then apply the confusion matrix for detecting fake reviews using movie review dataset V2.0 and movie reviews dataset V1.0. The SVM classifier give the heights accuracy with is 81.5% .They faced a problem which affected the accuracy they get which is very small dataset so they would like to extend their dataset and try to use different feature selection methods. Also they may apply sentiment classification algorithms to detect fake reviews using various tools such as Python and R or R studio to evaluate the performance of their work.

## 2 Project Description

Detecting Fake reviews by filtering it from authentic data review and automate from this real reviews.

## **2.1 Objective**

This system filters reviews from fake ones. Since reviews on products have a critical impact on the product's reputation and service providers. So we detect the fake positive and fake negative review from the real reviews

## **2.2 Scope**

System will detect fake reviews by using Deep learning and some machine learning algorithms according to:

1. Text pre processing.
2. Extract the feature from user account.
3. Classify the reviews.

## 2.3 Project Overview

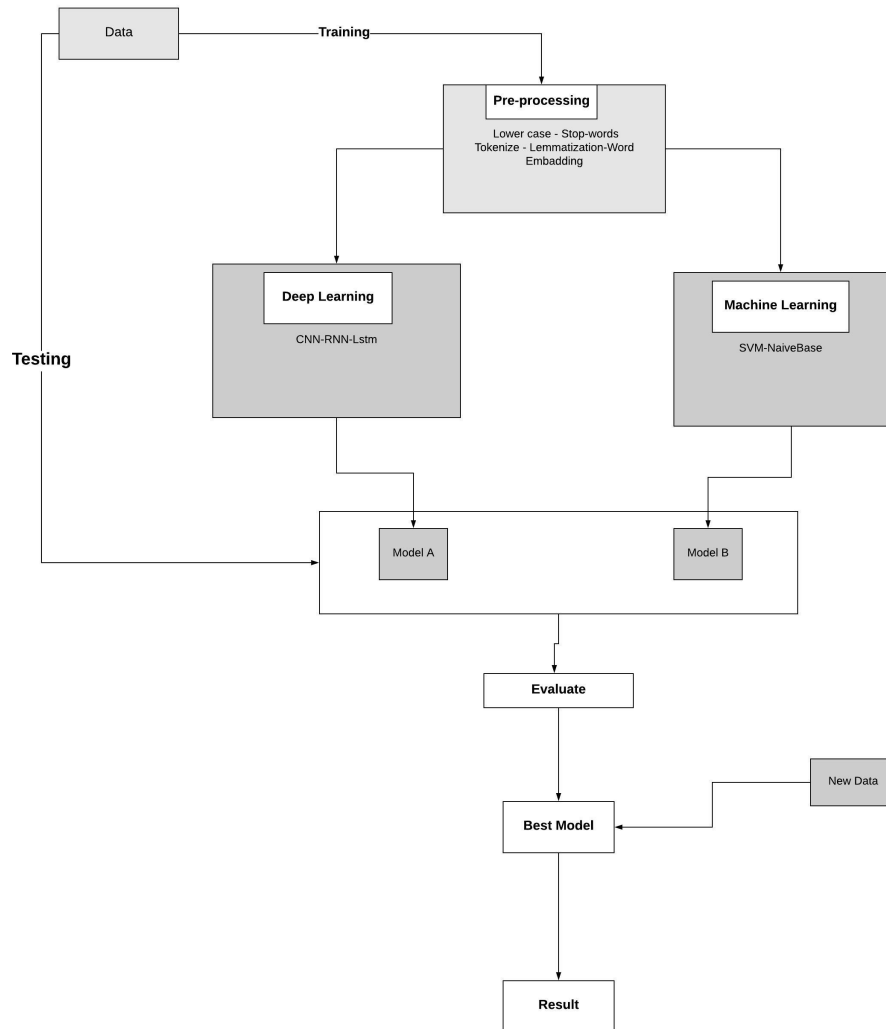


Figure 5: System Overview

1. We divide the text reviews data set into two categories one for training and the other for testing.

(a) The training data set category gets inside the pre-processing.

(b) Pre-processing :

- i. Lower case used to avoid having multiple copies of the same words.
  - ii. Removal of Stop Words which is the commonly occurring words .
  - iii. Tokenization refers to dividing the text into a sequence of words .
  - iv. Lemmatization it converts the word into its root word, rather than just stripping the suffices.
- (c) In classification their are two models deep learning and machine learning.
- (d) classification:
- i. Naive Base Classifier predicts membership probabilities for each class . And the class with the highest probability is considered as the most likely class.
  - ii. Support vector machines(SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis .
  - iii. convolutional neural network is a class of deep, feed-forward artificial neural networks, most commonly applied to analyzing visual imagery. CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing.
  - iv. A recurrent neural network (RNN) is a class of artificial neural network where connections between nodes form a directed graph along a sequence. This allows it to exhibit temporal dynamic behavior for a time sequence.
- (e) The testing data set category get inside the two models to evaluate them to the best one.
- (f) We compare between accuracy provided by the dataset with various classification algorithms and identified the best model for detecting Fake positive and negative Reviews.

### 3 Similar System Information

1. Debunking Fake News and False Claims using Evidence-Aware Deep Learning: In this paper they proposed an automated fact-checking end-to-end neural network model that collect signals from external evidence articles( language of these articles and the trustworthiness of their sources).In experiments they used four data sets(Snopex, PolitiFact, NewsTrust and SemEval ) to show the strength of their method. also use DeClarE method which includes external or counter evidence from Web.They compare their approach with LSTM-text, CNN-text , Distant Supervision approach for Credibility Classification
2. Fake News Detection Using Deep Learning: The aim in this paper is to build a classifier that can predict if this piece of news is fake based on its content.The problem that was faced in this paper was that fake news became increasingly menace to our society as it was believed that hat circulation of fake news had material impact on the outcome of the 2016 US Presidential Election.In this paper this problem was approached as an instance of text classication, using only the content of the article as the source of features.The data used for this project was drawn from two different sources. One from open Kaggle dataset that contains Fake news articles.The other one from Signal Media News dataset it contains Authentic news articles (negative examples for the classier).
3. Fake News Detection:  
This paper was written as Automatic fake news detection is a challenging problem in deception detection, and it has tremendous real-world political and social impacts. However there was a problem faced in this paper which is statistical approaches to combating fake news has been dramatically limited by the lack of labeled benchmark datasets.So in this paper they presented LIAR: a new, publicly available dataset for fake news detection. A 12.8K manually collected labeled short statements in various contexts from POLITIFACT.COM, which provides detailed analysis report and links to source documents for each case. This dataset can be used for fact-checking research as well. This paper helped us by providing a strong dataset for our project.
4. Fake Reviews Detection Under Belief Function Framework[1]:  
Researchers had written this paper to protect e-commerce from fake reviews to ensure customer’s confidence. There was some problems that was faced in this paper on of the main problem is to distinguish between fake reviews and truthful ones.The researchers use the belief fake reviews detection[BFRD] which deals with suspicion in rating reviews and it takes into account the similarity with other provided votes to detect the misleading.Researchers had discussed 4 steps to detect fake review :
  - (a) Modeling Reviewers Vote by Mass Functions



- (b) Distance Between the Current Reviewers Vote and All the Other Votes Aggregation
- (c) Construction of a New basic belief assignment (bba) Modeling the Vote into Fake or Not Fake
- (d) Decision Making as we get the final result after applying the pignistic probability on the bba

This paper helps us to show a new effective in distinguishing between the fraudulent reviews and the honest ones

5. spotting fake reviews via collective positive unlabeled learning [2]:

The aim of this paper to detect fake reviews and spam based on a real life experience on how fake reviews and spamming affected more than 500 restaurants in china. The main problem in the paper is although fake reviews detection was studied by researches for many years yet there is very small scale of data set available it is rather pseudo reviews rather than real one's. So researchers take the advantage of the correlations between users by behavioral features , reviews by text features and IP addresses would achieve better prediction results. The researches proposed a collective classification algorithm MHCC (Multi-typed d Heterogeneous Collective Classification) to identify fake reviews in the paper defined heterogeneous network over users, reviews and IP addresses. This paper helped us in detecting large number of potential fake reviews hidden in unlabeled data.

6. FAKE REVIEW DETECTION FROM A PRODUCT REVIEW USING MODIFIED METHOD OF ITERATIVE COMPUTATION FRAMEWORK:

Researchers had written this paper to detect a fake reviews for online products as they are now used as a source of information for these products. The main problem is detecting fake reviews and offering an unfair assessment to promote or defame a product or service. The researchers proposed iterative computation framework plus plus (ICF++) system to measure the honesty value of a review and the reliability value of the product. This paper will help us in finding a better accuracy compared with the result from iterative computation framework (ICF) method[R].

7. Spotting Fake Reviewer Groups in Consumer Reviews:

This paper works on opinion spam focused on detecting fake reviews either individual or group fake reviewers. The main problem that was faced in this paper that the number of a group of fake reviewers is even more damaging as they can take total control of the sentiment on the target product due to its size. The proposed paper first uses a frequent item by data mining method to find a set of candidate groups. It then uses

behavioral models and relation models to discover the relationships among groups, individual reviewers, and products they reviewed to detect fake reviewer groups. There are several steps they took to detect group fake reviews:

- (a) a labeled group spam data set.
- (b) a novel relation-based approach to detecting spammer groups.

This paper helps us in figuring out fake reviews that doesn't have same IP address, not related to each other and with large numbers.

8. Using Supervised Learning to Classify Authentic and Fake Online Reviews:

This paper was written to insure accuracy and credibility of product reviews as some entries could be fake yet written to appear authentic. The main problem that was faced in this paper is fake reviews are deliberately written to appear authentic, it is challenging for users to differentiate between those two things. So this paper uses supervised learning algorithms to analyze the extent to which authentic and fake reviews could be distinguished based on four linguistic clues, namely, understand ability, level of details, writing style, and cognition indicators.

- (a) The first aim by using several supervised learning algorithms to classify authentic and fake reviews based on the linguistic clues.
- (b) The second objective is to compare the classification performance of the proposed approach across all the algorithms against baselines from the literature

9. Detecting Deceptive Reviews Using Lexical and Syntactic Features

This paper help in detecting misleading opinion as it can be very difficult for human readers to detect spammers as they make their review look so true. The main problem in the paper is to differentiate between a spam and a real review. One of the approaches that was taken to solve this is by using stylometric features that help to distinguish the spammer writing style to find misleading reviews. In this paper supervised machine learning classifiers was used as Support Vector Machine (SVM) with Sequential Minimal Optimization (SMO) and Naive Bayes, is used to detect deceptive opinion. This paper helps in making the very small difference between a very well written spam or a real review appear.

10. Trust-Aware Review Spam Detection:

In this paper the researches aim to provide an efficient and effective method to identify review spammers. The problem that was faced in this paper

that spammers over the time are evolving to avoid being detected. So this makes the rule based and behavior based methods less effective. The suggested solution for this problem is that some online review systems encourage interactions among their users. for example :

- (a) Yelp.com and Last.fm allow registered users to form friendships
- (b) Amazon supports sending helpfulness tags to reviews that a user finds useful.

In this paper it was proposed a method based on random walk with restart to utilize two social relationships in rating prediction ,an iterative model to calculate the trustworthiness of each user based on rating behaviors and trustworthy predictions, and analyze the relationship between users trustworthiness and social relationships in the rating system.

### 3.1 Similar System Description

Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques. In this paper it was aimed to classify movie reviews into groups of positive or negative polarity by using machine learning algorithms. The problem that was faced in this paper was extract emotions inside the opinion, and how to detect fake positive reviews and fake negative reviews from opinion review. They used five supervised machine learning algorithms in detection : Nave Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN-IBK), KStar (K\*) and Decision Tree (DT-J48) for sentiment classification of reviews using two different datasets, which is movie review dataset V2.0 and movie reviews dataset V1.0. In this paper SA is used to find opinions from reviews and then classify these opinions based upon polarity. there are three major classifications in SA, namely: document level, sentence level, and aspect level. This paper helped in detecting fake reviews and finding the better algorithm with the highest accuracy.

### 3.2 Comparison with Proposed Project

	User Review Rating Prediction	Fake Review Detection From a Product Review Using Modified Method Of ITERATIVE COMPUTATION FRAMEWORK	Our Project
<b>Main purpose:</b>	Manual rating	Detecting fake views	Automated rating filtered reviews
<b>Algorithm:</b>	Learning algorithms : Logistic Regression Naïve Bayes SVM	ICF++ algorithm	Learning algorithms : Naïve Base Detection fake review algorithm by: ICF++
<b>Drawbacks:</b>	Their algorithm has a high accuracy but it takes long time	Low accuracy	We will use a naive base to decrease run time - sentiment analysis and eliminate least frequent word
<b>Data set:</b>	real data (yelp academy)	Non real data set	real data/kaggle (E-commerce )
<b>Accuracy</b>	90.3476%	63%	-----

Figure 6: comparison

## 4 Project Management and Deliverables

### 4.1 Tasks and Time Plan

Task Name	Start	Finish
Ideas of graduation projects	7/2/2018	7/15/2018
Idea Research	7/16/2018	9/1/2018
Survey and Proposal	9/1/2018	9/18/2018
Rehearsal of proposal	9/19/2018	9/19/2018
Proposal Presentation	9/26/2018	9/26/2018
Designing system	9/27/2018	10/1/2018
Designing Class Diagram	10/11/2018	10/18/2018
Submitting Survey Paper	10/20/2018	10/20/2018
SRS Writing	10/22/2018	11/9/2018
SRS presentation	11/14/2018	11/14/2018
External Examiner	12/3/2018	12/11/2018
SDD Writing	1/25/2019	1/30/2019
SDD Presentation	2 week of Feb	2 week of Feb
IMPLEMENTATION	2/25/2019	3/15/2019
IMPLEMENTATION Evaluation	After spring break	After spring break
Writing Paper		
Delivering Paper	3 days after Spring Vacation	
TECHNICAL EVALUATION	1st week of may	
Writing Thesis	5/9/2019	5/15/2019
Final Presentation	25-Jun-19	

Figure 7: Time plan

### 4.2 Budget and Resource Costs

No Budget is required

## 5 Reference

- [1] Popat, Kashyap, et al. "DeClarE: Debunking Fake News and False Claims using Evidence-Aware Deep Learning." arXiv preprint arXiv:1809.06416 (2018).
- [2] Ben Khalifa M., Elouedi Z., Lefvre E. (2019) Fake Reviews Detection Under Belief Function Framework. In: Hassanien A., Tolba M., Shaalan K., Azar A. (eds) Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018. AISI 2018. Advances in Intelligent Systems and Computing, vol 845. Springer, Cham
- [3] Li, Huayi, et al. "Spotting fake reviews via collective positive-unlabeled learning." Data Mining (ICDM), 2014 IEEE International Conference on. IEEE, 2014.
- [4] Wahyuni, Eka Dyar, and Arif Djunaidy. "Fake review detection from a product review using modified method of iterative computation framework." MATEC Web of Conferences. Vol. 58. EDP Sciences, 2016.
- [5] Mukherjee, Arjun, Bing Liu, and Natalie Glance. "Spotting fake reviewer groups in consumer reviews." Proceedings of the 21st international conference on World Wide Web. ACM, 2012.
- [6] Mukherjee, Arjun, Bing Liu, and Natalie Glance. "Spotting fake reviewer groups in consumer reviews." Proceedings of the 21st international conference on World Wide Web. ACM, 2012.
- [7] Shojaee, Somayeh, et al. "Detecting deceptive reviews using lexical and syntactic features." Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on. IEEE, 2013.
- [8] Leong Sammy and Bhatia Ashish, User Review Rating Prediction. Stanford. Dec 14, 2012.
- [9] Tang, Duyu, et al. "User Modeling with Neural Network for Review Rating Prediction." IJCAI. 2015.
- [10] Xue, Hao, et al. "Trust-aware review spam detection." Trustcom/BigDataSE/ISPA, 2015 IEEE. Vol. 1. IEEE, 2015.
- [11] Lei, Xiaojiang, Xueming Qian, and Guoshuai Zhao. "Rating prediction based on social sentiment from textual reviews." IEEE Transactions on Multimedia 18.9 (2016): 1910-1921.
- [12] indal, Nitin, and Bing Liu. "Opinion spam and analysis." Proceedings of the 2008 international conference on web search and data mining. ACM, 2008.

[13]Cambridge Dictionary.<https://dictionary.cambridge.org/dictionary/english/rating>.

[14]<https://thehill.com/policy/cybersecurity/381449-researchers-say-fake-news-had-substantial-impact-on-2016-election>