# DETECTING EXTREMIST REVIEWER GROUPS IN ONLINE PRODUCT REVIEWS

P.HemaVenkataramana[1], S. Divya[2], K.BhuvanaPriya[3], T.Jyothi[4]

*Department of Computer Science and Engineering,*

*Andhra Loyola Institute of Engineering and Technology,*

*Jawaharlal Nehru Technological University - Kakinada*

*pandirihema90@gmail.com[1],*

*sunkesuladivyalakshmi@gmail.com[2],*

*bhuvanakothamas@gmail.com[3],*

*jyothitejavath510@gmail.com[4]*

**Abstract**: Online business mostly depends on the reviews of the previous buyers. So, online marketplace is the place where the spam reviews can take place. The reviews may be positive or negative. The review should be genuine for the product rating. But some people either as individual or as a group tending to spread the spam reviews. They may have supposed focus on particular type of product or brand level. In this article we collected the reviews of one product from Amazon product review site. The groups are extracted from the site and they can have clustered together if they are mutually reviewed extremely.

## 1. Introduction

People usually review a product based on their experience, comfort and quality according to the price of the product. These people look like they are acting individual at group. While people giving individual replies they do not much so much effect on the opinion of the buyer but they may useful by giving their experience. Some people are trying to manipulate the buyers according to their favour side. Many people may have fed into their trap by seeing their reviews. Due to that the opinion of the buyers may change based on those spam reviewsthe spam reviewers used to give extreme reviews either positively or negatively.

## II.ABOUT THE PROPOSED WORK

### A. Literature Survey

Extremist reviews are similar to fake reviews. But here it is observed that some people targeting specific brand or product for promoting it or to diminish it. This is also called as collective fraud behavior, where several users are part of particular product or brand targeting to influence the people towards their brand. These type of reviews can be classifying through robust and by using analysis technique. Amazon India, to prevent spam opinion it limits the reviews per a day.

### B. Proposed Work

We develop a supervised model to classify reviewer candidate groups as extremist entities. We run different types of classifiers for the task of classifying a group based on the reviews written by the user of that group having extremity sign. A three-layer perceptron based classifier is the best classifier. In this we can determine the extremist reviews with their rating so the review written user is extremist reviewer. The multiple layer perceptron classifier brings the good accuracy and results.
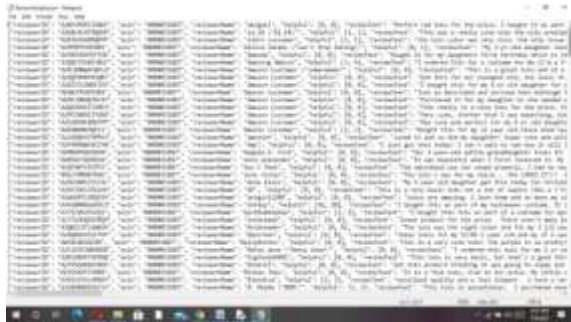
### 1. Design Methodology

1.1 This proposed system is based on the reviews taken from Amazon there will be a number of reviews in a product the reviews. They are further classified by using this method word embedding it is a methodology for gathering for grouping the words which are same or having the clause meaning of other words this will gather the words from the reviews and give a number to it technically. This can be known as follows by using vocab package the words in the reviews which are given by the user are given a mathematical real number these are stored in the vector space. It will give the count of words which are given in the reviews the words in the review may be more or less it will give the exact the count of the word it gives how many times the word is present in that particular data set are file or reviews which we have taken.

1.2 In the previous stage we used word embedding in that many of the words are grouped. The remaining which are left over by that method are not useful. Those useless words are removed by a method called

dropout method. It can also be known as dilution. It is use for text purpose; it comprises the over fitting on the data. It is one of the best useful methods to remove noisy data from the data set. The idiom dropout refers to dropping out the units of both hidden layer as well as in the visible layer. In simple terms the dropout means its skips the units of data which are remained in the training face of the data set which is done in the above process.
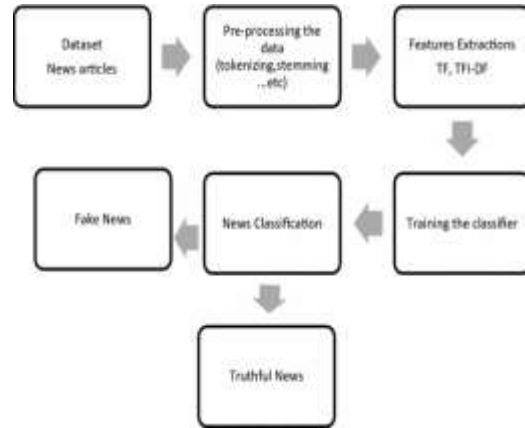
1.3 The other methodologies we used in this proposed system are training, validating and testing. Training means we are taking data from the various sites are reference. Now we are splitting the data using vocab package and we are finding the density to know how many times the word is used. Validating means classifying. The data will be classified according to the given instructions in this stage. After this validating we are having testing stage. The use of testing is to check whether the output from the validating is correct or not. It is totally based on our required output of this system. It will check whether there are any bugs in the proposed system. After this testing stage we will get know whether our system will work properly or not.

2.                                    Dataset



This is the data set which we are using in our project. The data set consist of reviewer ID reviver name, review and rating. The dataset is taking from amazon by web scraping method. We scraping means data scraping which is used for the extraction of data from different website this software can access world wide web by using HTTP protocol while this can be done normally by a user. This is totally done by as automated process which can be implemented using a bot.
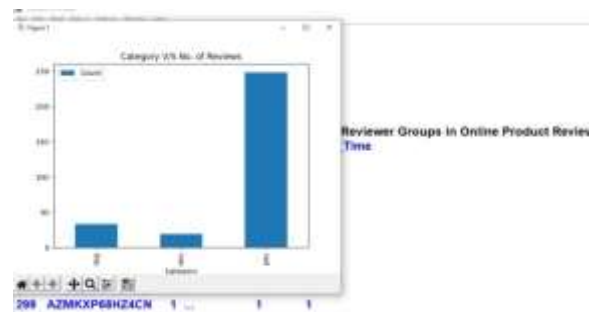
3.*Dataflow*



This is the data flow which we proposed in our system. First we are collecting the dataset from amazon. Then the dataset will go through pre-processing. This contains tokenizing, stemming extra. Then it goes through feature extraction. Next it under goes training the classifier. After this we will get new classification then we will come to know the news that is extreme positive or extreme negative.
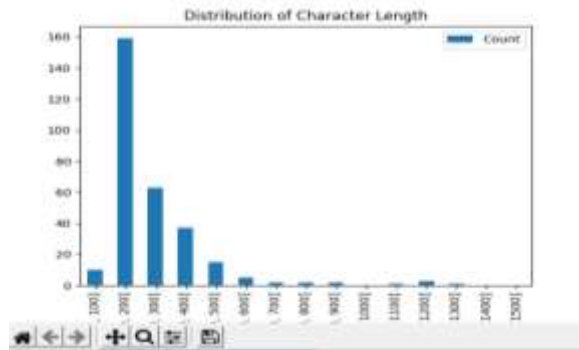
3.*Results*



This is the result of our project. It is the result came when we enter our own review to know the extremist and truth probability score. Here we need to enter a review then it classified and produce the results.

These are the number of reviews u/s different types categories like negative neutral and positive. In this graphs we come to know the number of different types of reviews.



This is used to know the character length over the number of reviews given in the dataset. It gives the number of characters per review.

*4. Advantages*

- This model can give the extreme review exactly
- By using neural networks, we can get high accuracy in the result.
- This will help us to know whether the product is getting extreme positive response or negative response.

*5.Disadvantages*

- This will not check for many products at a time.
- If the dataset is too large then it will take more time to process and to produce results

*6.Conclusion*

This system is used to give high and accurate results. It can be occurred by using forward and backward propagation recursively. There will be a less interaction of human because it is unsupervised learning. The future scope of the project is it can be extended using various platform reviewers and can be known.

*7.Acknowledgement*

8. References

[1]. A. Kim, "That review you wrote on Amazon? Price-less," https://www.usatoday.com/story/tech/news/2017/03/20/review-you-wrote-amazon-pricess/99332602/, 2017.

[2]. E. Gilbert and K. Karahalios, "Understanding deja reviewers," in CSCW, 2010, pp. 225–228. [Online]. Available: http://doi.acm.org/10.1145/1718918.1718961

[3]. Amazon.in, "Review Community Guidelines," https://www.amazon.in/gp/help/customer/display.html?nodeId=201929730, 2018, online.

[4]. A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," in WWW, 2012, pp. 191–200.

[5]. SK Kotamraju, PG Arepalli, SS Kanumalli (2021), Implementation patterns of secured internet of things environment using advanced blockchain technologies, Materials Today: Proceedings, 2021.

[6]. Gopi A.P., Patibandla R.S.M. (2021) An Efficient Methodology for Avoiding Threats in Smart Homes with Low Power Consumption in IoT Environment Using Blockchain Technology. In: Choudhury T., Khanna A., Toe T.T., Khurana M., Gia Nhu N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-65691-1_16.

[7]. RSML Patibandla, SN Mohanty (2021), Need of Improving the Emotional Intelligence of Employees in an Organization for Better Outcomes,Decision Making And Problem Solving: A Practical guide, 2021.

[8]. RSML Patibandla (2021), An Efficient Fog-Based Model for Secured Data Communication, Integration of Cloud Computing with Internet of Things …, 2021.

[9].    Y. Lu, L. Zhang, Y. Xiao, and Y. Li, "Simultaneously detecting fake reviews and review spammers using factor.