JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Detecting and Characterizing Extremist Reviewer Groups inOnline Product Reviews

[1]Fardeen Khan,[2]Dr Sunita Sachin Dhotre

[1]Student,[2]Associate Professor
Department of Computer Engineering, Bharati Vidyapeeth(Deemed to be University)
Collegeof Engineering, Pune– 411043.

Abstract : The online market is rife with opinion spam that takes the appearance of reviews. People are frequently hired to write favourable or unfavourable evaluations for certain products in order to encourage or discourage sales of those products. A lot of times, groups work on this. Nothing has been done to uncover those who are targeting a complete brand rather than a specific product, despite earlier study attempts to identify and analyse this spam group.*Reviews* were collected from Amazon's product review website and painstakingly sorted into 923 different reviewer groups for this post. When groups are extracted using frequent itemset mining over brand similarities, users who have evaluated many different brands together are clustered together.It has been suggested that the composition of reviewer groups be determined by eight features unique to a (group, brand) pair. A feature-based supervised model has been developed to classify possible candidate groups as extremist entities. On the reviews that group members have supplied, a variety of classifiers have been run in order to determine whether a group demonstrates signs of extremism. A three-layer Perceptron system is discovered to be the highest accurate classifier. The inquiry has been carried out in order to thoroughly examine the actions of such organisations in order to better understand the dynamics of brand-level opinion fraud. Examples of these practises include consistency in rating, review sentiment, confirmed purchases, review dates, and favourable reviewer comments. Surprisingly, several authenticated reviews have been found to be expressing intense sentiments,Unexpectedly, it has been seen that a lot of qualified reviewers are voicing very strong sentiments. More investigation reveals ways to bypass Amazon's current measures to prevent unofficial incentives.

***IndexTerms* - Java,LSTM,rnn algorithm, NLTK (natural language ToolKit)**

I. Introduction

In the online marketplace-dominated digital age of today, review websites and portals play a critical role in influencing consumers' purchasing decisions. The owner of the online cosmetics company Elizabeth Mott, Alice, claims that this starts a positive feedback loop where more reviews encourage more purchases, which in turn encourage more reviews. Products that have more sales and favourable reviews consequently frequently have better search rankings and increasing sales. This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors. Digital Object Identifier 10.1109/TCSS.2020.2988098

It is clear from this situation, nevertheless, that some people or organisations might create reviews that aren't entirely truthful with the aim of swaying prospective customers' perceptions in their favour. These reviewers may act for their own reasons, expressing either joy or annoyance, but their influence on how others perceive the product as a whole is still quite small. However, when several people join together to create a complicated network of reviews, a more significant influence happens. These groups have a big impact on how people feel about a product because of their sheer numbers and use of certain tactics (described in Section VIII). Studies have indicated that a sizable fraction of reviews (10%–15%) simply repeat the feelings expressed in the first reviews, amplifying the impact of other factors in addition to opinion spam.

LITERATURESURVEY:

Understanding deja reviewersAUTHOR:

E. GILBERT AND K. KARAHALIOS,

People who invest a lot of time and effort in writing product evaluations online may be perplexed as to why some reviewers opt to simply restate what others have already mentioned. A mixed-method study that focused on "deja reviewers," or people who arrive late to the reviewing process and repeat prior comments, was done to answer this topic.About 10-15% of the reviews in this study's analysis of 100,000 Amazon.com reviews for signs of repetition were found to be quite similar to earlier ones. In order to better understand reviewers' motivations, interviews with reviewers were performed using these algorithmically identified reviews as the foundation for this inquiry.There have been many reviews that claim to explain why there are deja reviews, but deeper underlying issues connected to a person's standing in the community have also been noticed. This study finishes by offering an original idea motivated by our research: a self-aware community that motivates members to support shared goals.

Approximately 10-15% of all reviews strongly mimic earlier ones, according to the study. But after a certain point, these bad reviews are just squandered chances. Is it necessary, for instance, that the 130th review praise the Wii Fit's balance games? Instead, wouldn't learning about the Wii Fit's long-term performance be more beneficial? Even while the reviewer benefits from writing it, the community learns little new. Reading many viewpoints on a certain product might help prospective consumers have a more thorough knowledge of its features.

PROPOSED SYSTEM :

The AMAZON product reviews dataset was utilised to train the LSTM (a form of RNN algorithm) employed in this study to detect extremist reviews. The NLTK (natural language Tool Kit) approach has been used to clean the dataset by removing stop words and special symbols. We used the TF- IDF (term frequency inverse document frequency) approach to turn reviews into a numeric vector after text cleaning. Each word is replaced by its average frequency using TF-IDF, and this TF-IDF vector is then supplied to the LSTM to train the model.

This project uses Java to put this algorithm into practise.Any review can be input after model training, and LSTM will determine whether the review is extreme or moderate.

ADVANTAGES OF PROPOSED SYSTEM :

1) High accuracy 2)High efficiency

SYSTEM ARCHITECTURE :

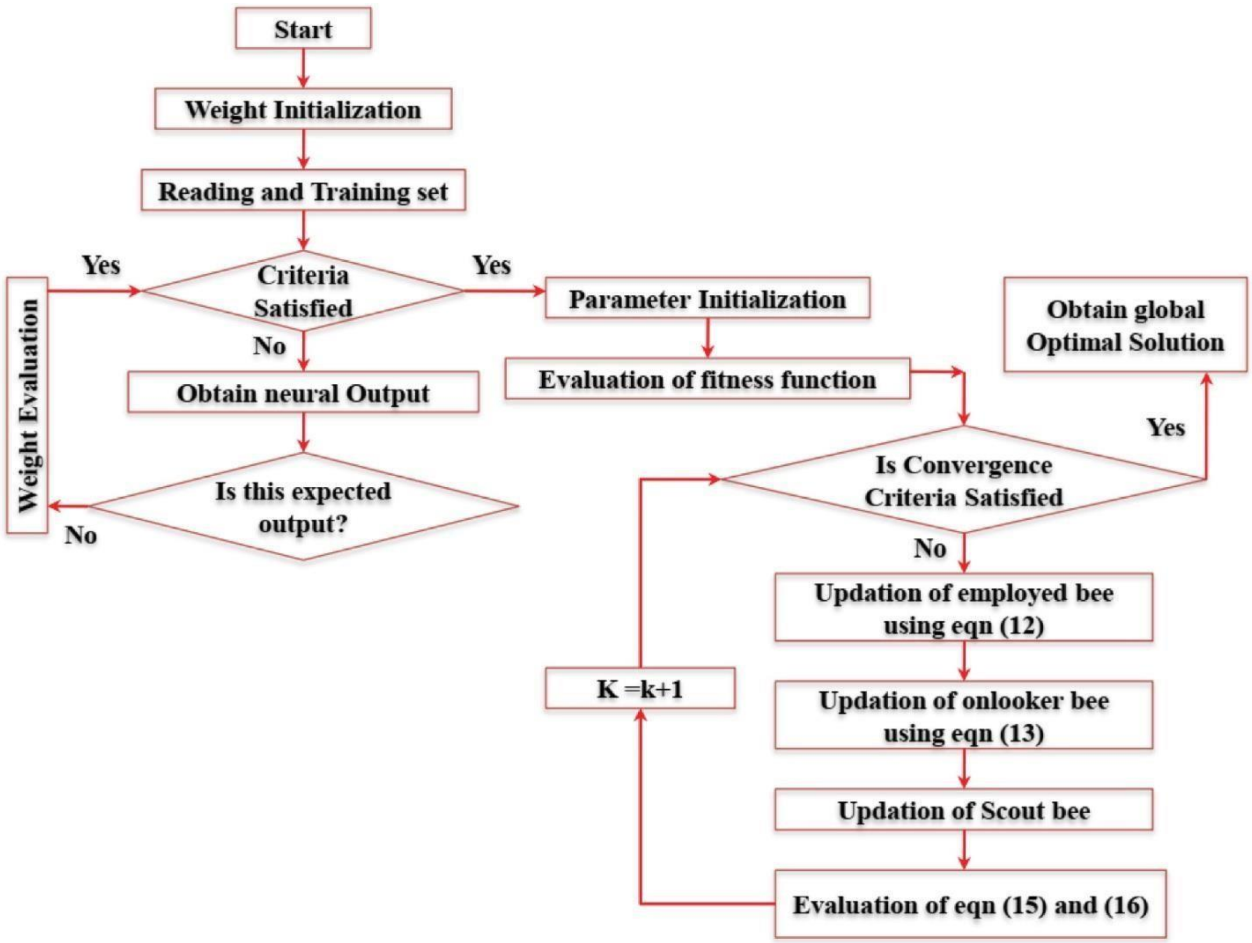


figure 1(System architecture)

SYSTEM STUDY

FEASIBILITY STUDY

In this stage, the project's viability is assessed, and a business proposal is presented with a very basic project plan and some cost projections. The proposed system's practicality must be investigated during system analysis. This will guarantee that the suggested solution won't burden the business. Understanding the main system requirements is crucial for the feasibility analysis.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is being done to see what kind of financial impact the system will have on the company. The corporation has a finite amount of money to invest in the system's research and development. The costs must be supported by evidence. As a result, the developed system came in under budget, which was made possible by the fact that most of the technology were public domain. Only the specialised goods needed to be bought.

TECHNICAL FEASIBILITY

This study is being done to evaluate the system's technical requirements, or technical feasibility. Any system created must not place a heavy burden on the technical resources at hand. The amount of technological resources available will be heavily strained as a result. As a result, the client will face high expectations. The created system must have a low demand because its implementation merely necessitates little or no adjustments.

SOCIAL FEASIBILITY

The goal of the study is to determine how much the user accepts the system. This includes the instruction needed for the user to operate the system effectively. The system shouldn't make the user feel threatened; instead, they should view it as a need. The techniques used to inform and acquaint the user with the system are the only factors that affect the level of acceptance by the users. As the system's ultimate user, his confidence must be increased so that he may offer some helpful criticism, which is encouraged.

SYSTEM DESIGNUML DIAGRAMS

Unified Modelling Language is known as UML. A general-purpose modelling language with standards, UML is used in the field of object-oriented software engineering. The Object Management Group oversees and developed the standard.

The objective is for UML to establish itself as a standard language for modelling object-oriented computer programmes. UML now consists of a meta-model and a notation as its two main parts. In the future, UML might also be coupled with or added to in the form of a method or process.

The Unified Modelling Language is a standard language for business modelling, non-software systems, and describing, visualising, building, and documenting the artefacts of software systems.

The UML is an amalgamation of best engineering practises that have been effective in simulating huge, complicated systems.

The UML is a crucial component of the software development process and the creation of objects-oriented software. The UML primarily employs graphical notations to convey software project design.

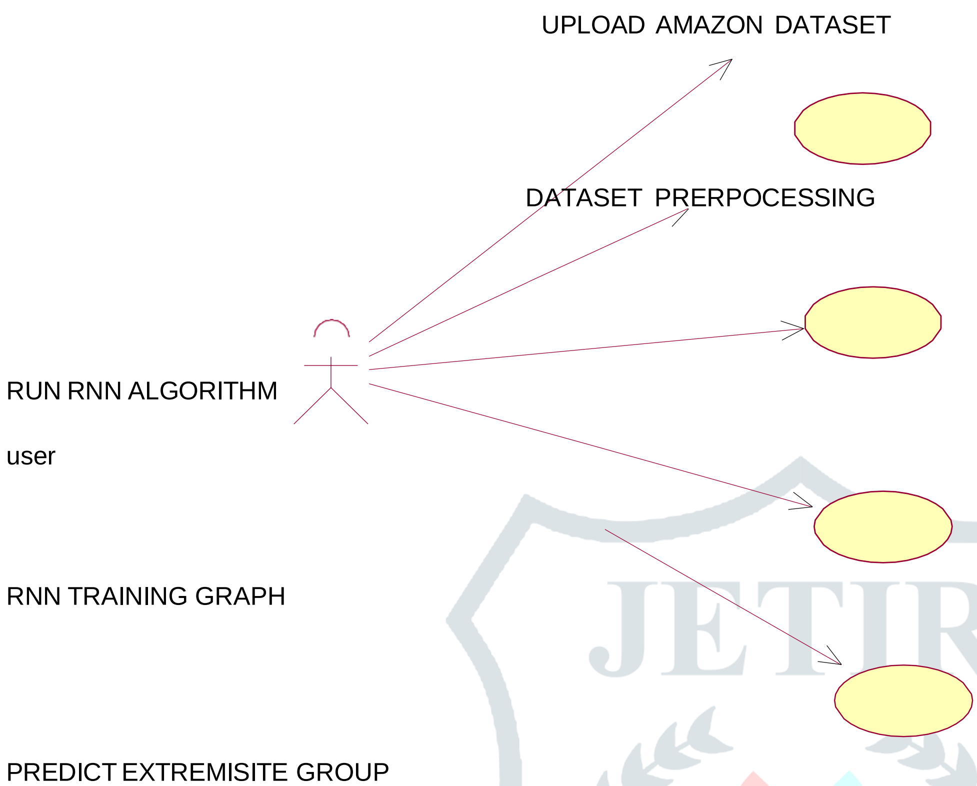
GOALS:

The following are the UML's primary design objectives:

1. Offer users an expressive visual modelling language that is ready to use so they can create and trade meaningful models.
2. Offer tools for specialisation and extendibility to expand the fundamental ideas.
3. Not depend on a certain development methodology or programming language.
4. Offer an official framework for comprehending the modelling language.
5. Promote the commercial expansion of OO tools.
6. Encourage the use of higher level development ideas like components, frameworks, patterns, and collaborations.
7. Include top techniques.

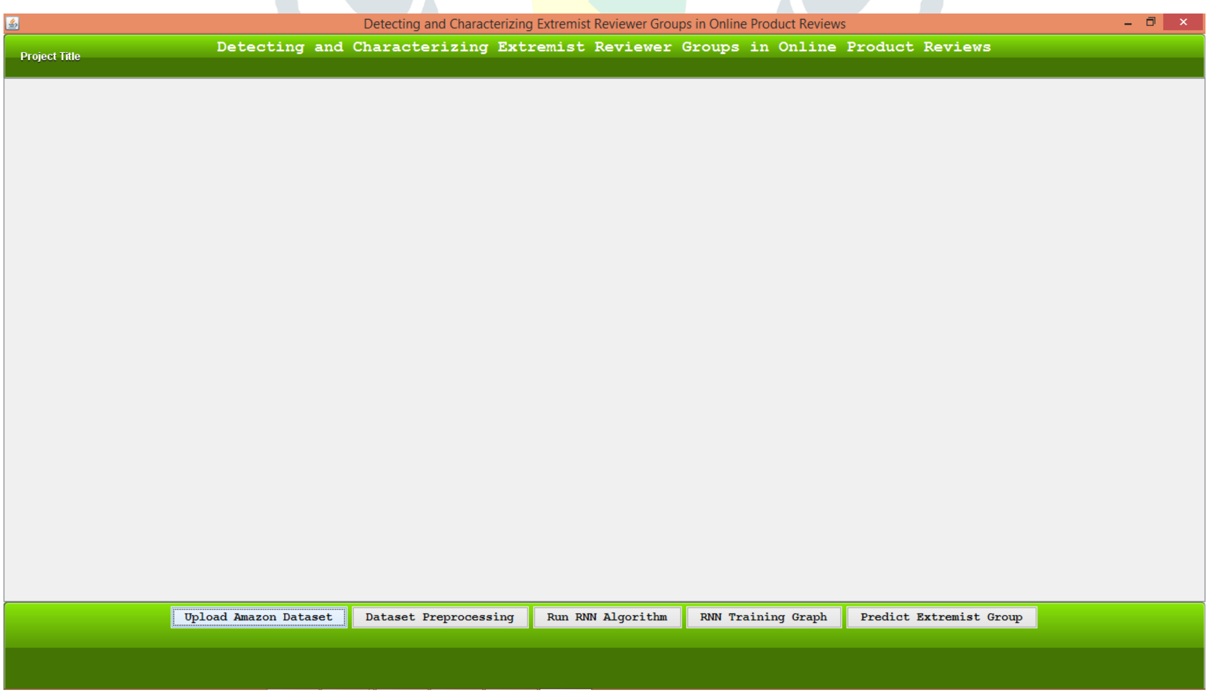
USE CASE DIAGRAM:

In the Unified Modelling Language (UML), a use case diagram is a specific kind of behavioural diagram that results from and is defined by a use-case analysis. Its objective is to provide a graphical picture of a system's functionality in terms of actors, their objectives (expressed as use cases), and any dependencies among those use cases. A use case diagram's primary objective is to identify which system functions are carried out for which actor. The system's actors can be represented by their roles.

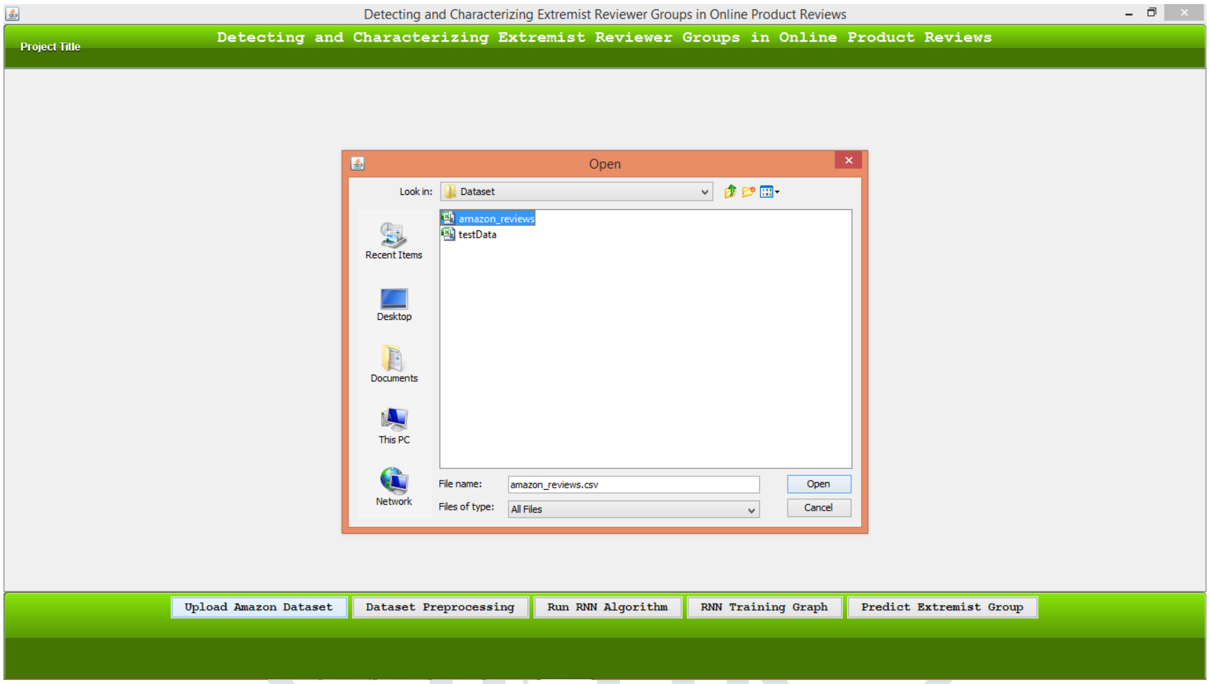


RESULT ANALYSIS

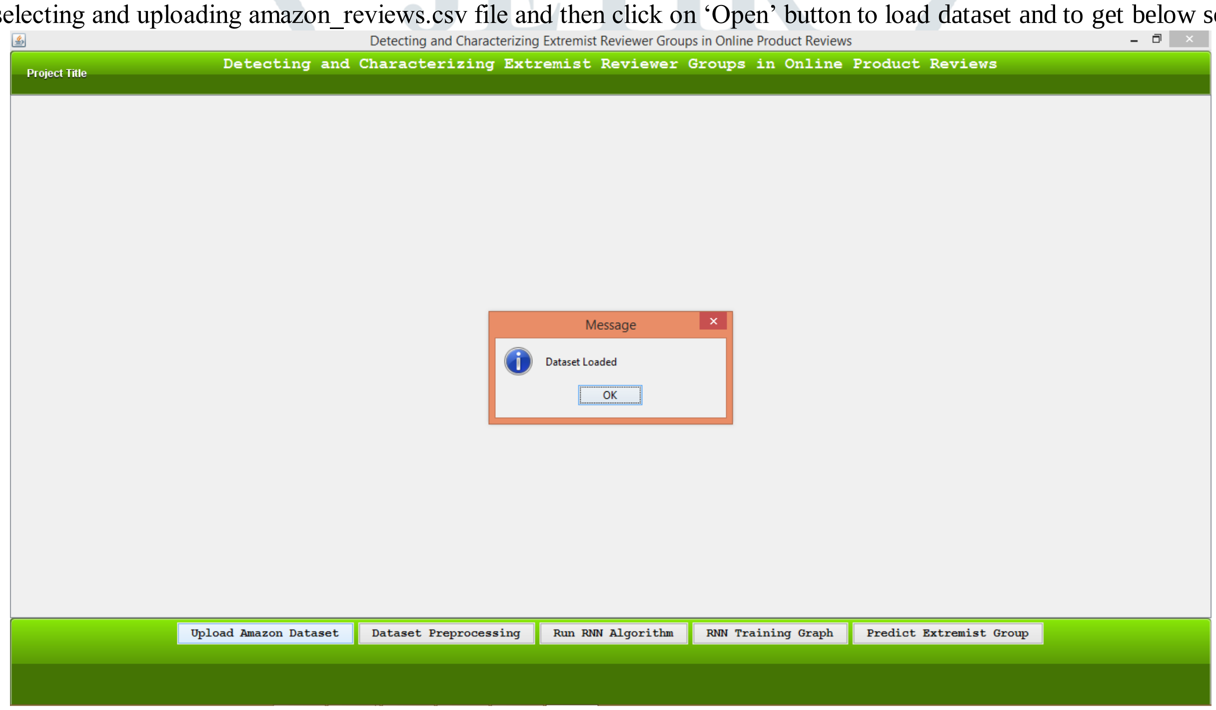
To run project double click on ‘run.bat’ file to get below screen



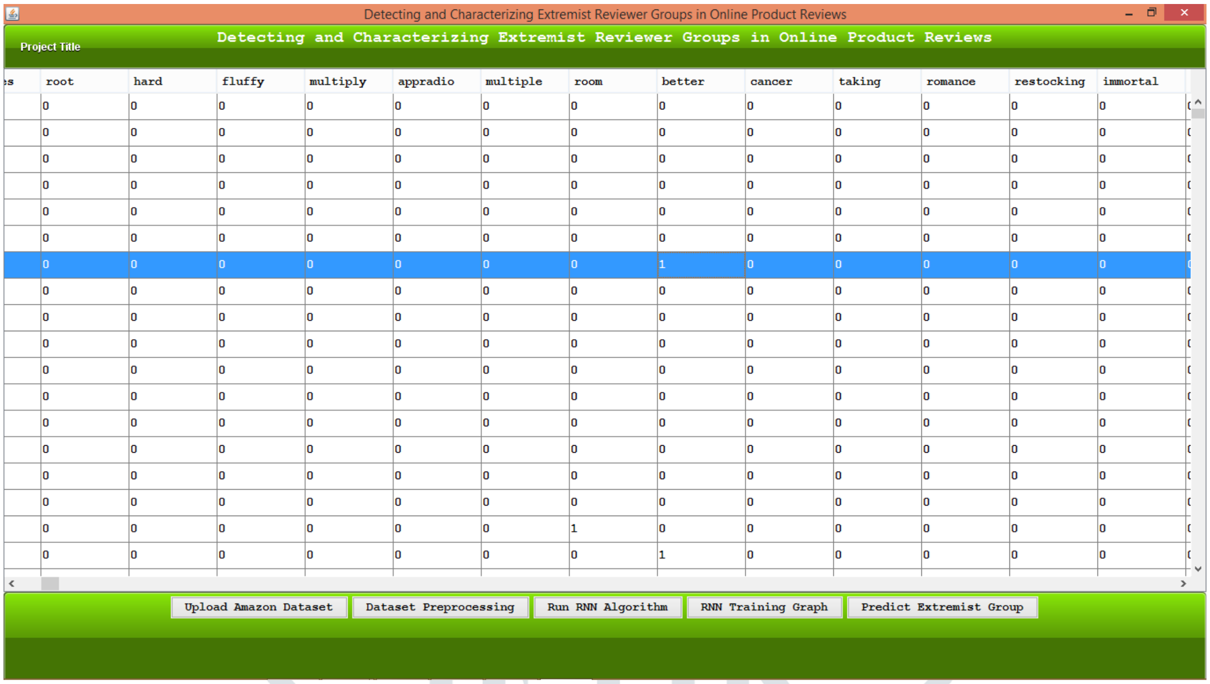
In above screen click on ‘Upload Amazon Dataset’ button to load dataset and to get below screen



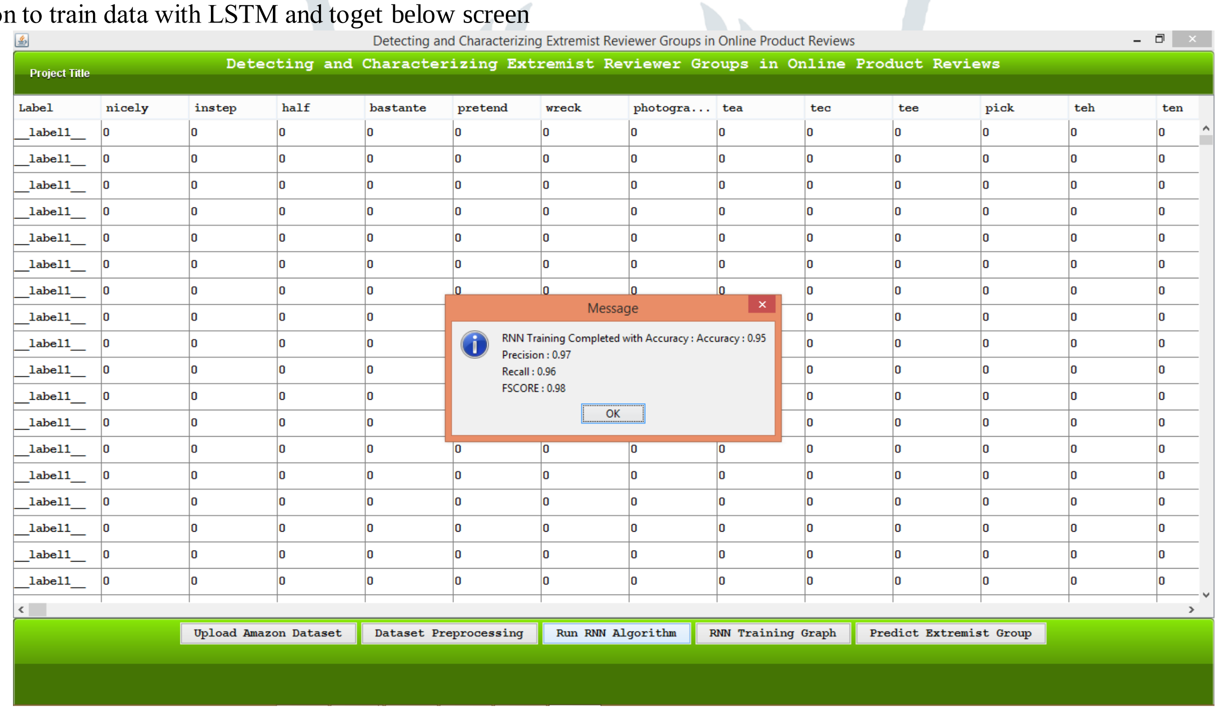
In above screen selecting and uploading amazon_reviews.csv file and then click on ‘Open’ button to load dataset and to get below screen



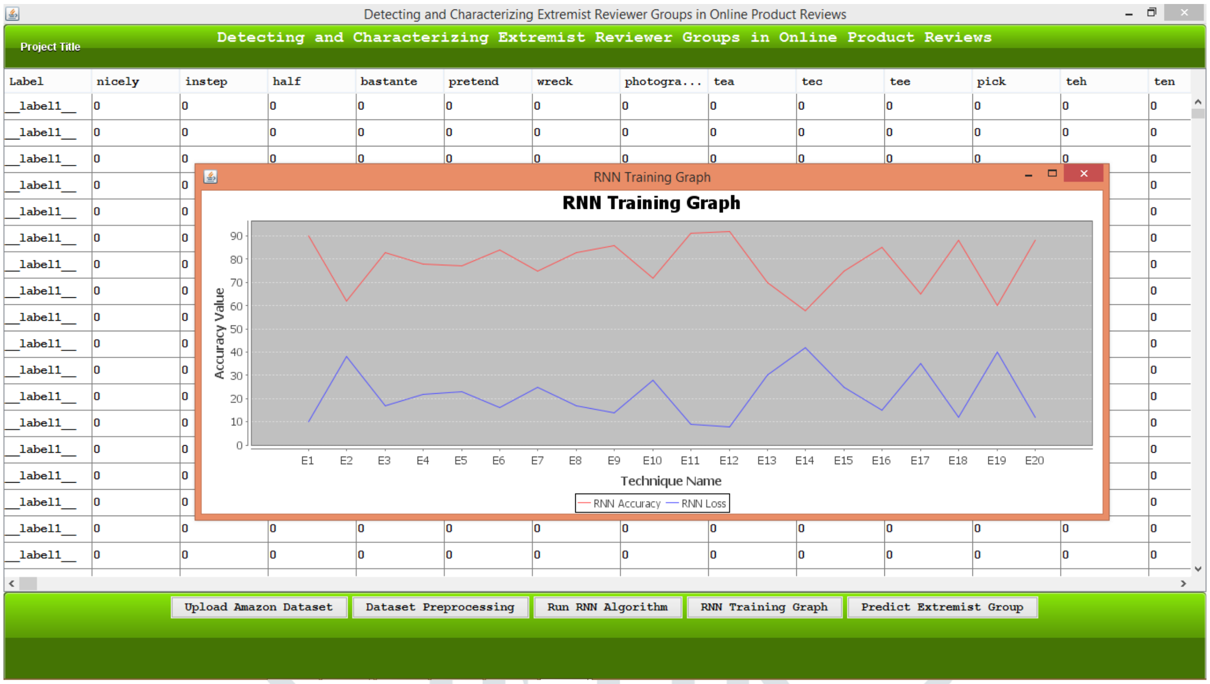
In above screen dataset loaded and reviews contains special symbols and stop words like (the and or where etc.) and we can remove such words by pre-processing reviews and then split dataset into train and test and then convert entire dataset into TFIDF vector by clicking on ‘Data Preprocessing’ button



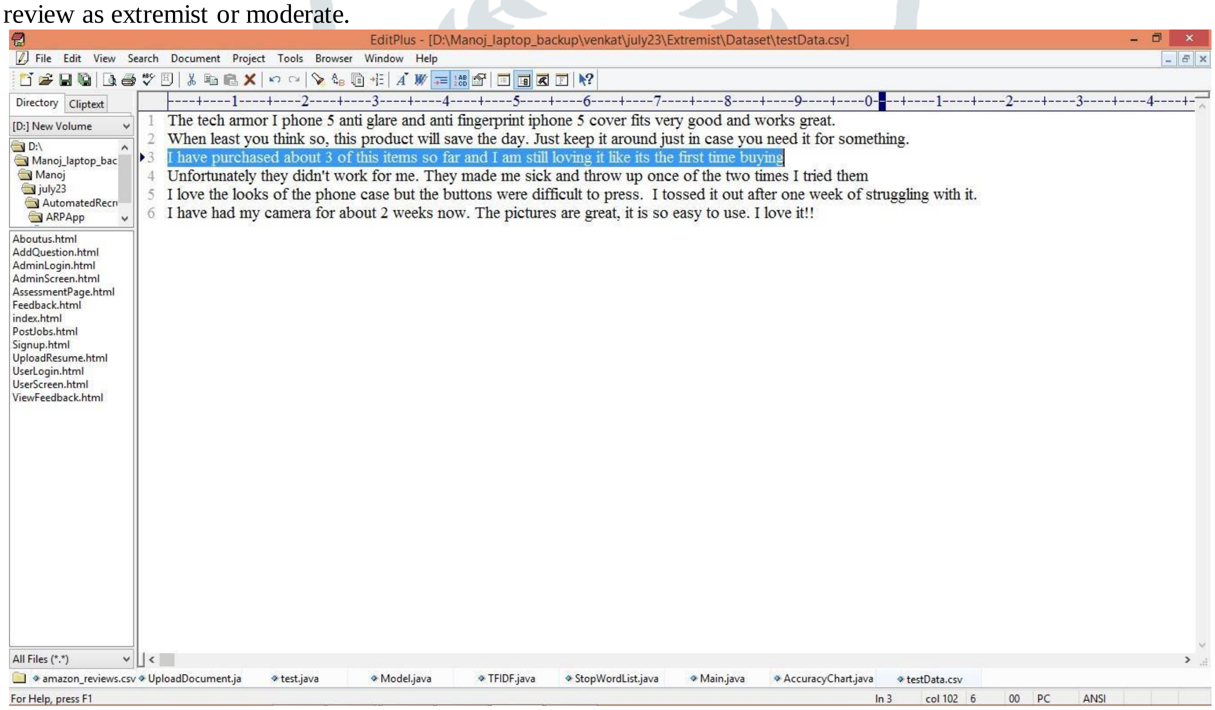
In above screen each review is converted to TFIDF vector where column header represents word names fromall reviews and column rows represents word occurrence count and if 0 means word not occur in that review and now cleaned reviews are and now click on ‘ Run RNN Algorithm’ button to train data with LSTM and toget below screen



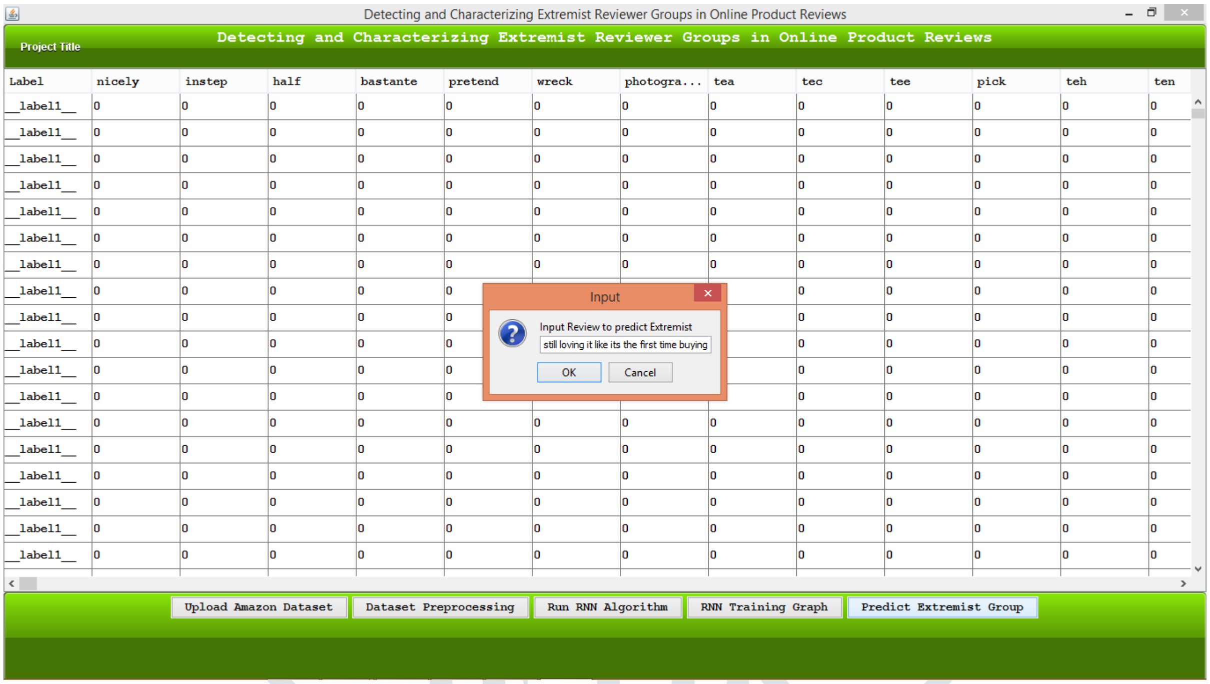
In above screen we got RNN accuracy as 95% and we can see other metrics like precision, recall and FSCORE and now click on ‘RNN Training Graph’ button to get below graph.



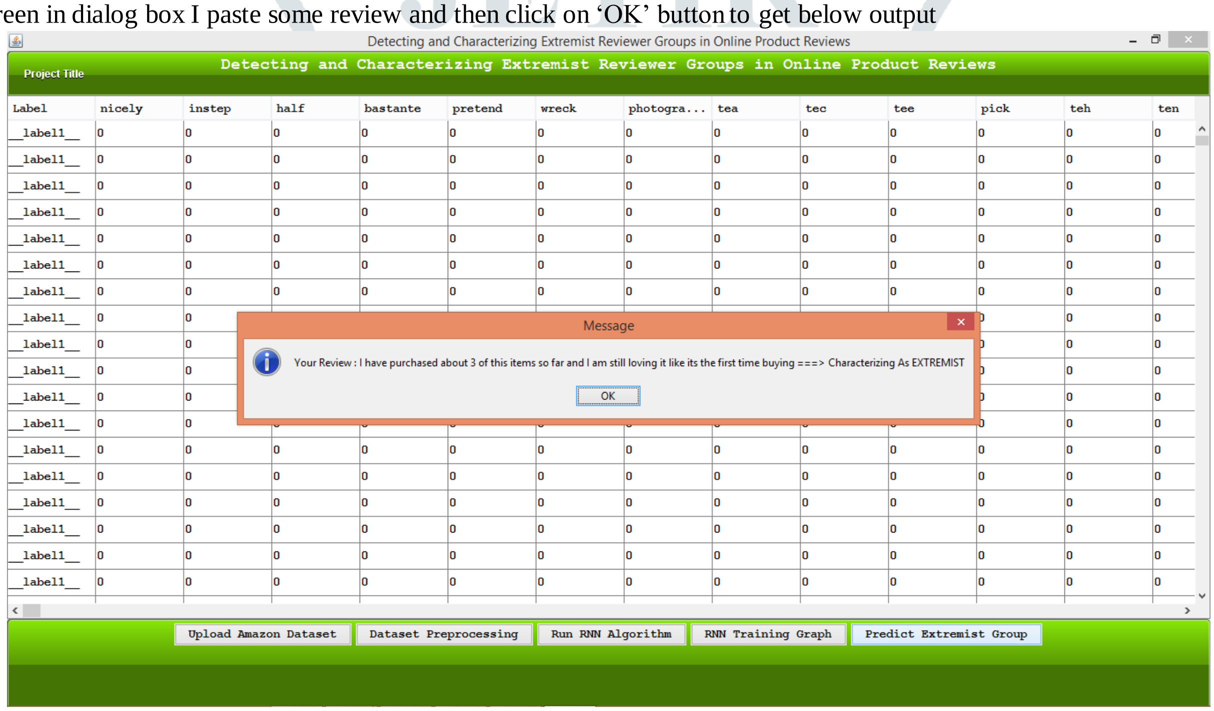
in above graph red line represents ACCURACY values and blue line represents loss values and to train LSTM we took 20 epoch and x-axis represents epoch and y-axis represents accuracy and loss values. In above graph with each increasing epoch accuracy got increase and loss got. Now close above graph and then enter some reviews by copying from 'TestData.csv' file and then click on 'Predict Extremist Group' button to predict review as extremist or moderate.



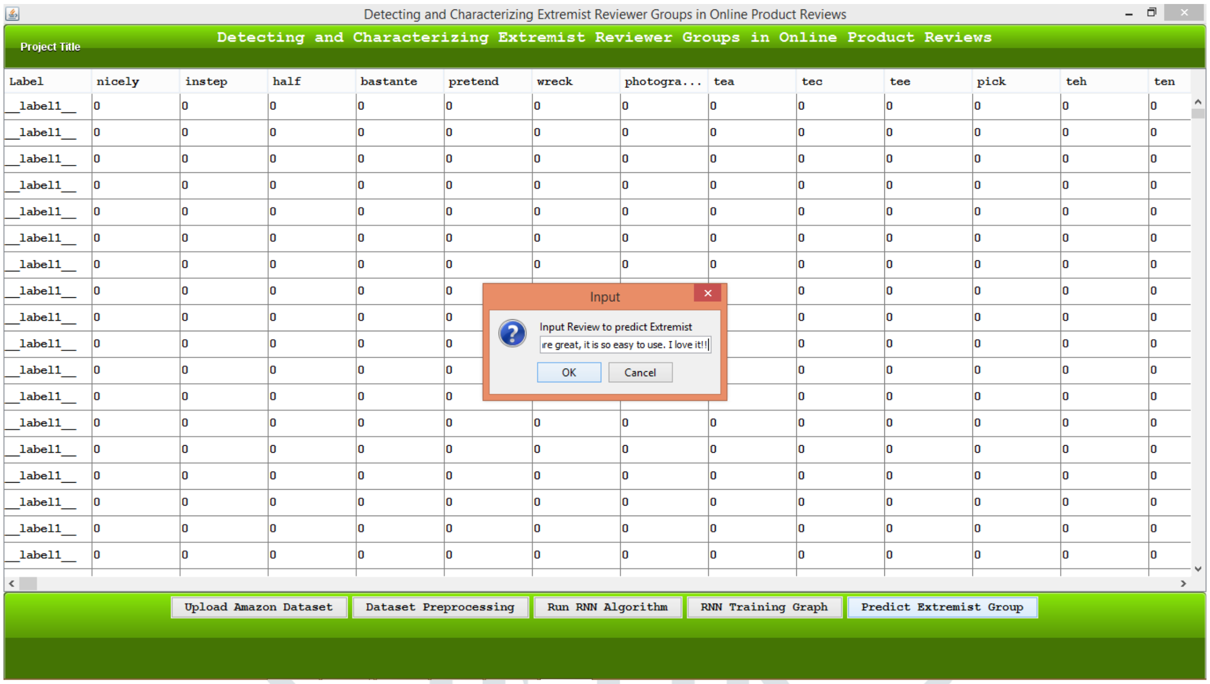
From above test reviews screen I am selecting and copying one review and paste in application dialog box byclicking 'Predict Extremist Group' button like below screen



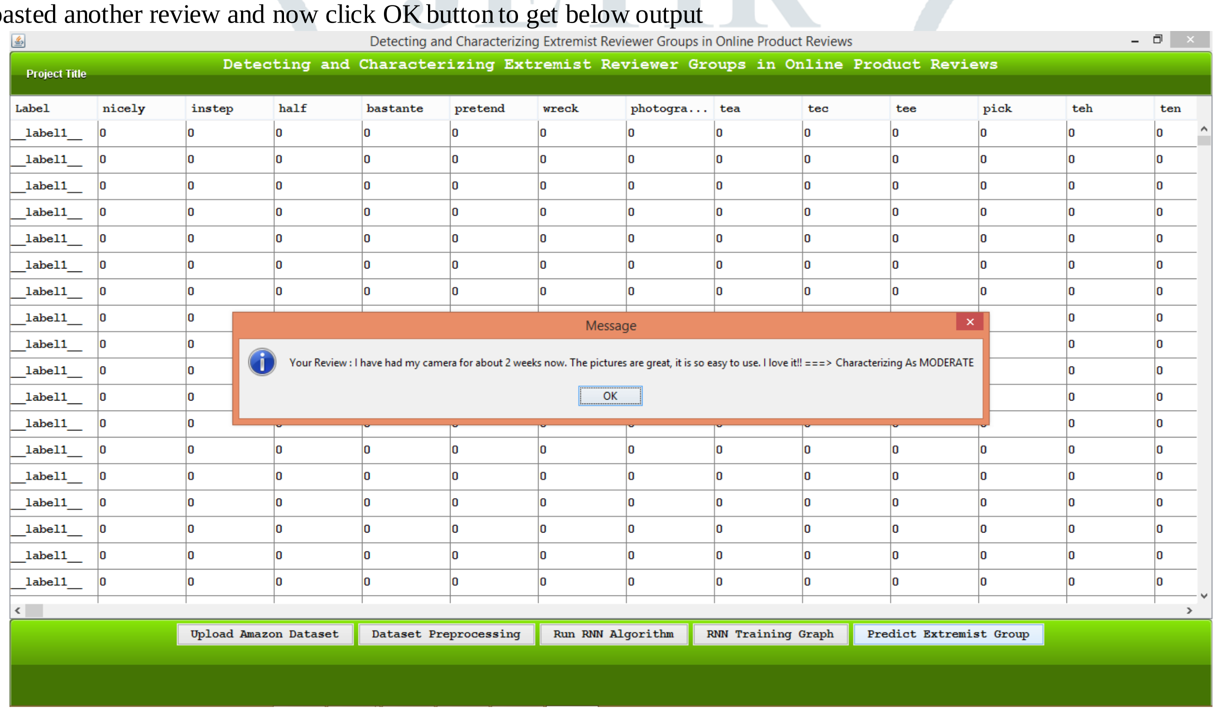
Now in above screen in dialog box I paste some review and then click on 'OK' button to get below output



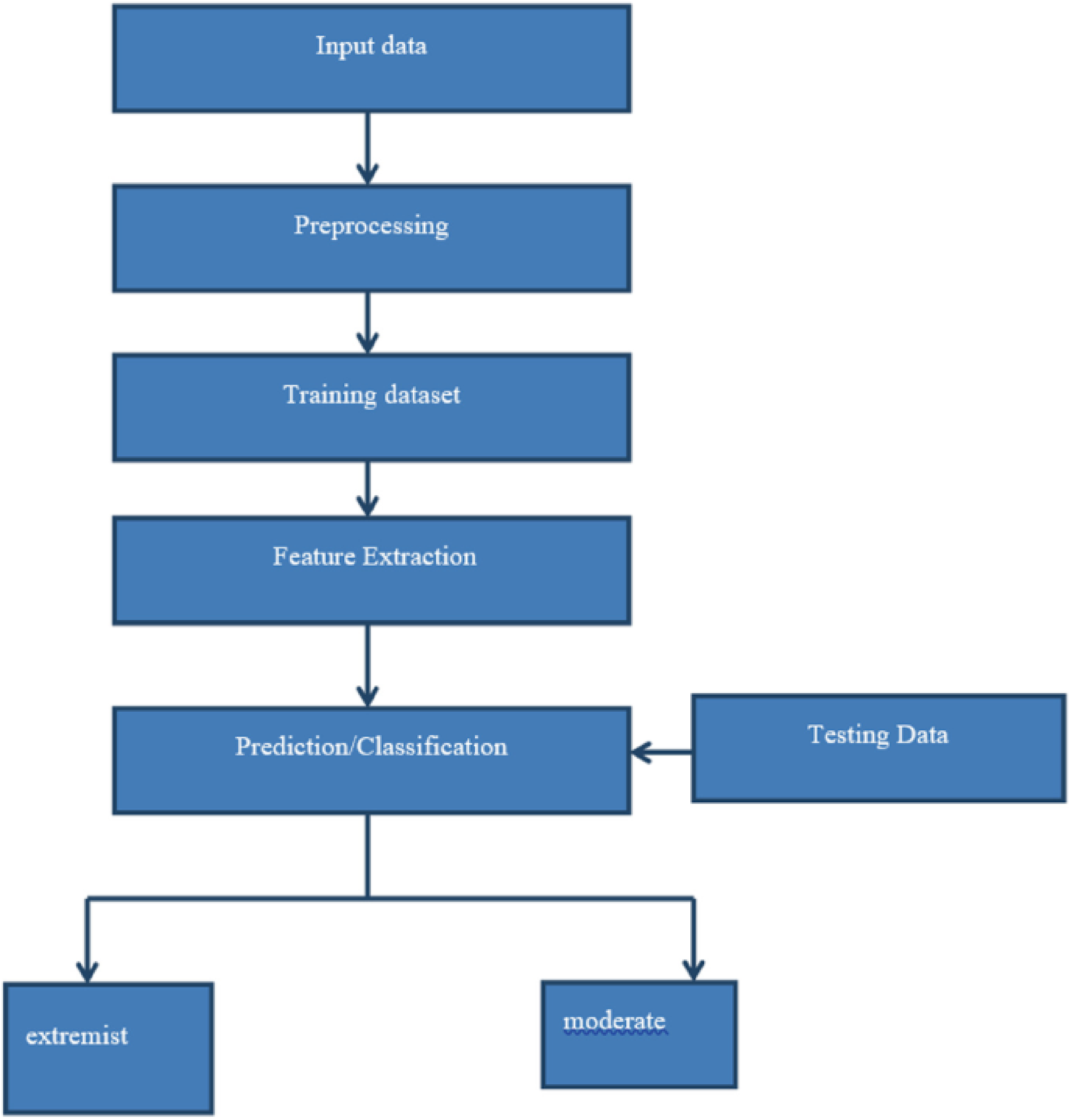
In above screen in output dialog box after arrow symbol given review predicted as 'EXTREMIST' and nowtest another review



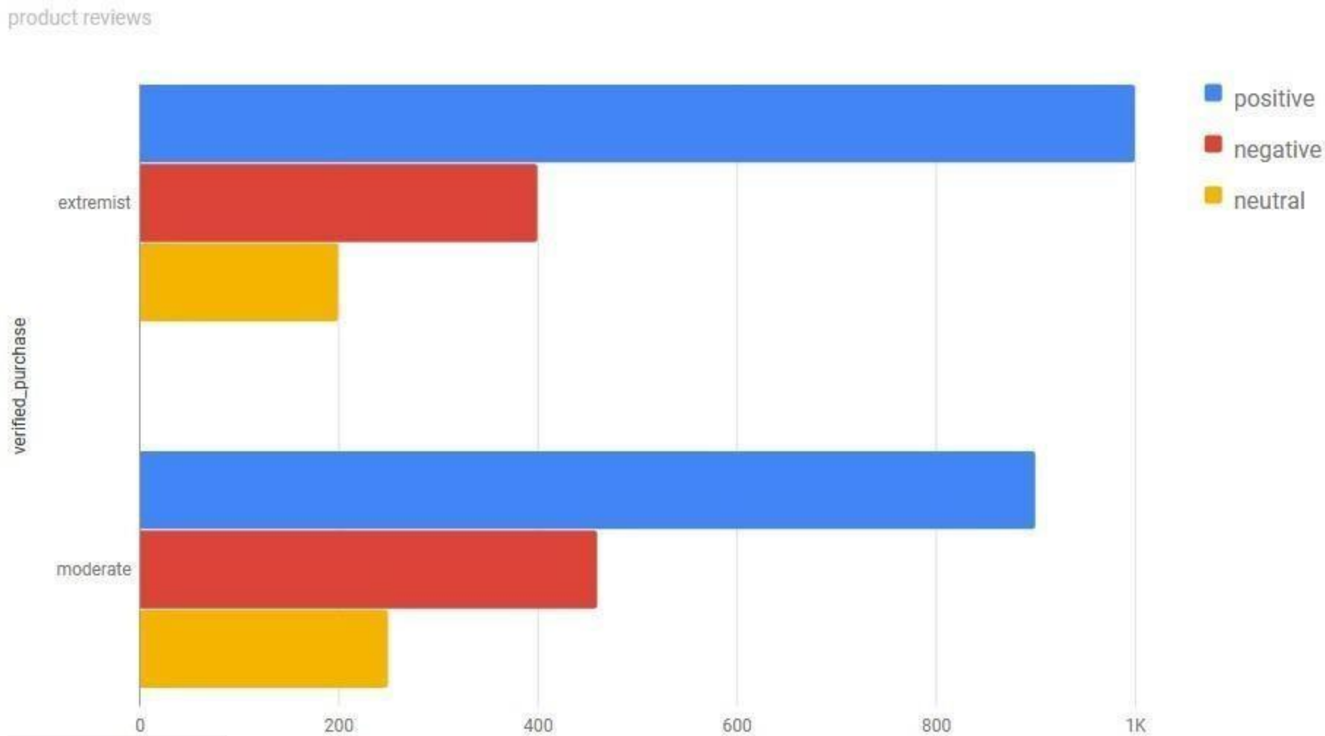
In above screen pasted another review and now click OK button to get below output



In above screen new review predicted as 'MODERATE' and similarly you can input any review and getprediction result Similarly you can enter any review and predict as Extremist or moderate


Input data
Preprocessing
Training dataset
Feature Extraction
Prediction/Classification
Testing Data
extremist
moderate

Performance Graph



DISCUSSION :

In this essay, we covered a new type of opinion spam when spammers submit irrational reviews of brands in an effort to alter public perceptions of those brands as a whole. These organisations are frequently a part of a sophisticated business web that has the power to affect the overall reputation and popularity of a number of businesses on review websites. This article is the first step in tying extremism in reviews to brand-level group activity, which reveals crucial information about market operations. With the aid of these insights, a better recommendation might be made using online reviews. Extremist groups were found by analysing their actions as a group based on multiple attributes, utilising a supervised learning technique based on a ground truth of manually annotated labels, and retrieving a set of candidate spam groups using FIM. Then, we divided groups into extremist and moderate categories and evaluated the accuracy of several classification techniques. After classifying these organisations, we closely examined the behaviours of extremist organisations to learn more about the phenomena and the general patterns of how these organisations target these brands. Additionally, we made the algorithms and annotated data set available for further research.

CONCLUSION :

This leads us to the conclusion that there is a previously unidentified form of opinion spam in which spammers target entire businesses and post nonsensical reviews in an effort to change how consumers regard the brand. These organisations frequently form a complex web of commercial relationships that can have an impact on the popularity and general reputation of a variety of companies on review websites. This article is the first to link extremism in evaluations to group behaviour at the brand level, which provides significant insights into how markets function. These information may help improve the recommendation provided utilising online reviews. Extremist organisations were identified by observing their collective behaviour based on a variety of characteristics and applying a supervised learning method based on FIM.

Conflict of Interest

On behalf of all authors, the corresponding author Mr Fardeen Khan states that there is no conflict of interest.

REFERENCES:

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

[2] E. Gilbert and K. Karahalios, "Understanding deja reviewers," in Proc. ACM Conf. Comput. supportedCooperat. Work (CSCW), 2010, pp. 225–228, doi: 10.1145/1718918.1718961.

[3] Amazon.in. (2018). Review Community Guidelines. [Online]. Available: https://www.amazon.in/gp/help/customer/display.html?nodeId= 201929730

[4] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," in Proc. 21st Int.Conf. World Wide Web (WWW), 2012, pp. 191–200.

[5] Y. Lu, L. Zhang, Y. Xiao, and Y. Li, "Simultaneously detecting fake reviews and review spammers usingfactor graph model," in Proc. 5th Annu. ACM Web Sci. Conf. (WebSci), 2013, pp. 225–233.

[6] S. Rayana and L. Akoglu, "Collective opinion spam detection: Bridging review networks and metadata," inProc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2015, pp. 985–994.

[7] S. Dhawan, S. C. R. Gangireddy, S. Kumar, and T. Chakraborty, "Spotting collective behaviour of online frauds in customer reviews," 2019, arXiv:1905.13649. [Online]. Available: http://arxiv.org/abs/1905. 13649

[8] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in Proc. 12th Int. Conf. World Wide Web (WWW), 2003, pp. 519–528.

[9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment clas- sification using machine learning techniques," in Proc. Conf. Empirical Methods Natural Lang. Process., 2002, pp. 79–86.

[10] K. Mouthami, K. N. Devi, and V. M. Bhaskaran, "Sentiment analysis and classification based on textual reviews," in Proc. Int. Conf. Inf. Commun. Embedded Syst. (ICICES), Feb. 2013, pp. 271–276.

[11] Q. Ye, Z. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches," Expert Syst. Appl., vol. 36, no. 3, pp. 6527–6535, Apr. 2009.

[12] M. Chelliah and S. Sarkar, "Product recommendations enhanced with reviews," in Proc. 11th ACM Conf.Recommender Syst., Aug. 2017, pp. 398–399.

[13] L. Chen and F. Wang, "Preference-based clustering reviews for aug- menting e-commercerecommendation," Knowl.-Based Syst., vol. 50, pp. 44–59, Sep. 2013.

[14] J. Feuerbach, B. Loepp, C.-M. Barbu, and J. Ziegler, "Enhancing an interactive recommendation system with review-based information filtering," in Proc. IntRS@RecSys, 2017, pp. 10–55.

[15] A. Almahairi, K. Kastner, K. Cho, and A. Courville, "Learning dis- tributed representations from reviewsfor collaborative filtering,"