Juni Khyat
(UGC Care Group I Listed Journal)
ISSN: 2278-4632
Vol-14 Issue-02 Aug 2024

DETECTING AND CHARACTERIZING EXTREMIST REVIEWER GROUPS IN ONLINE PRODUCT REVIEWS

Mrs.T.Sruthi [1], Singam Vidya Vani (20S11A1244) [2], Upperla Akhila (21S11A1203) [3], Bandaru Harika (21S15A1201) [4], Amballa Anish Reddy (20S11A1207) [5],

ASSISTANT PROFESSOR [1], UG STUDENTS [2,3,4,5,]

DEPARTMENT OF INFORMATION TECHNOLOGY

MALLA REDDY INSTITUTE OF TECHNOLOGY & SCIENCE,

Maisammaguda, Medchal (M), Hyderabad-500100, Telangana.

ABSTRACT

Online marketplaces often witness opinion spam in the form of reviews. People are often hired to target specific brands for promoting or impeding them by writing highly positive or negative reviews. This often is done collectively in groups. Although some previous studies attempted to identify and analyze such opinion spam groups, little has been explored to spot those groups who target a brand as a whole, instead of just products. In this article, we collected the reviews from the Amazon product review site and manually labeled a set of 923 candidate reviewer groups. The groups are extracted using frequent itemset mining over brand similarities such that users are clustered together if they have mutually reviewed (products of) a lot of brands. We hypothesize that the nature of the reviewer groups is dependent on eight features specific to a (group, brand) pair. We develop a feature-based supervised model to classify candidate groups as extremist entities. We run multiple classifiers for the task of classifying a group based on the reviews written by the users of that group to determine whether the group shows signs of extremity. A three-layer perceptron-based classifier turns out to be the best classifier. We further study behaviors of such groups in detail to understand the dynamics of brand-level opinion fraud better. These behaviors include consistency in ratings, review sentiment, verified purchase, review dates, and helpful votes received on reviews. Surprisingly, we observe that there are a lot of verified reviewers showing extreme sentiment, which, on further investigation, leads to ways to circumvent the existing mechanisms in place to prevent unofficial incentives on Amazon.

Introduction

In Today's world dominated by online marketplaces, review portals and websites play a crucial role in the buyer's decision for their next purchase. "It is a virtuous cycle–the more reviews, the more buys. The more buys, the more reviews. The more buys, the higher your rank in search and the more sales you get," says Alice , the owner of online cosmetic brand Elizabeth Mott. Undoubtedly, it is highly likely that some people write reviews that are less than truthful to manipulate widespread decision of buyers in their favor. These people act either individually or in groups. While individual reviewers write such reviews in a matter of frustration or joy, they do not influence the overall opinion on a product to a large extent but help other buyers by stating their experiences. However, a more compelling case is when multiple individuals form an intricate web, and due to sheer higher number of people reviewing (and certain other techniques, discussed in Section VIII), they end up being a major influence on the overall sentiment of the product. The extent of such influence is not just limited to the reviews by opinion spam. Previous work has shown that 10%–15% reviews are essentially echoing the earliest reviews, and thus, a misleading early review has an even higher influential potential. This is widespread opinion spam, and every review website must be aware of this activity and take appropriate measures for the identification and/or prevention of this phenomenon. This is a classic example of collective fraud behavior, where several users are part of a business network and work together to target and influence a particular product. This is a lesser known phenomenon, and most groups work following certain techniques to not make their collaboration obvious. However, since such groups are economically or otherwise incentivized, and several of these are generally run by a given organization, they have several targets for 3

opinion spam, which often share certain common characteristics in their nature of reviews. These characteristics can be exploited to classify them better using a robust and thorough an analysis technique. Amazon India, to prevent opinion spam, has brought about a new policy that limits the number of reviews on a product in a day, as stated in . In order to still be effective, we claim that certain groups target brands in general and postextreme reviews across multiple products for a given target brand. A detailed discussion is required for these brand-related activities because these practices are

ISSN: 2278-4632
Vol-14 Issue-02 Aug 2024

against the code of conduct of these review websites since they negatively skew the brand-based competition, giving innate (dis)advantages to certain brands. Since only the nonverified reviews are limited by the policies,1 reviewers from these groups can often purchase the product via Amazon in exchange for unofficial discounts (e.g., cash backs) and postverified reviews since they did not receive a discount via amazon's mechanisms (e.g., coupons) (see further discussed in Section VIII). Fig. 1 shows an example of such extremist groups (taken from our annotated data set as mentioned in Section III). Four rows correspond to the products belonging to four different brands. Four columns represent four different reviewers who, according to our annotation, are part of the same group. Each box represents the review information. This is an example of reviewers showing extreme likeliness toward these products/brands as can be seen from the extreme ratings, similar comments, and almost the same date . It is clear that this group of reviewers had extreme sentiments toward the brands reviewed, both in terms of the ratings and the review content. It is worth noting that such a kind of characterization is different from just combining the groups of people who provide extreme reviews on a product, because while the groups focusing on a product may be extreme in their opinion, they do not necessarily intend to influence the brand image Sellers may not have any inclination toward promoting any particular brand's products; rather, they would prefer to gain a better revenue on all products (may belong to different brands) by their promotional campaigns.In this article, we identify and study the behavioral characteristics of extremist reviewer groups. We also build a feature-based classifier based on the brand-specific activities of reviewer groups to identify the extremist groups on the Amazon India marketplace. We then further analyze our methodology to unfold behaviors that best signify such activities and compare and analyze the overall trend of these groups viz-a-viz their behaviors. The major contributions of this article are fourfold: 1) a manually labeled data set of 923 reviewer groups that are classified into "extremist" and "moderate" categories; 2) the first-ever characterization and study of the novel problem of identifying brand-level extremism; 3) detailed characterization of extremist reviewer groups; 4) design supervised approach to detect extremist reviewer groups.

LITERATURESURVEY

A.Kim

This study aimed to dissect factors thatengender Coronary road criticism exploitation Random forest Classifier. Itshows that random forests rule may be accustomed the process and bracket of knowledge similar as CAD.

E.Gilbert and K.Karahalios

In this paper, the Particle Swarm Optimization(PSO) rule is employed to induce rules for heartcriticism. Rules area unit optimized grounded on their accuracy exploitation PSO rule. militarizationmedical knowledge with intelligent tools for designation and treating health problem will cut back miscalculations.

A.Mukherjee, B.Liu and N.Glance

Experimenters have latterly planned many software package, tools, and algorithms for developing effective call support systems. one in all the foremost vital problems is that the opinion of heart criticism, and diverse experimenters have worked to developintelligent call support systems to assist physicians. A network may be a fashionable tool for prognosticating the opinion of heart criticism. A network- grounded heart criticism anticipation system is developed during this exploration paper. For heart criticism anticipation, the planned system utilized thirteen attributes. The trials distributed during this work incontestible that the planned rule out performed analogous state- of- the- art approaches.

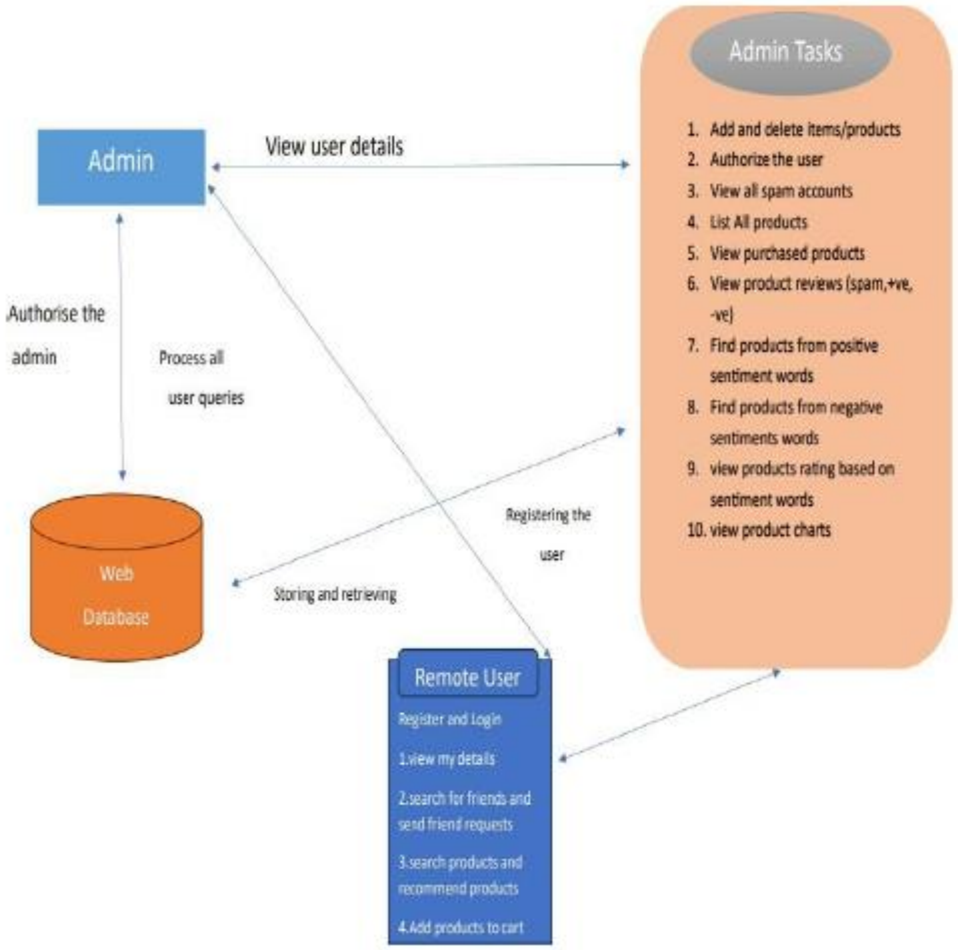
Y.Lu, L.Zhang, Y.Xiao and Y.Li

The planned clinical call network for threat anticipation in cases is split into 2 stages(1) an automatic approach for generating weighted fuzzy rules and call tree rules and(2) the event of a fuzzy rule- grounded call network. to realize weighted fuzzy rules within the initial section, we tend to used the mining fashion, attribute choice, and therefore the attribute weightage system. The fuzzysystem is additionally erected exploitation the weighted fuzzy rules and attributes that are named.

Eventually, the planned system is tested exploitation datasets earned from the UCI deposit, and its performance is compared to a neural network- grounded system exploitation delicacy, sensibility, andquality.

SYSTEM ANALYSIS

ISSN: 2278-4632
Vol-14 Issue-02 Aug 2024



Existing System

 Sentiment analysis can be conducted on three different levels: review-level, sentence- level, and phrase-level.

 Review-level analysis and sentence-level analysis attempt to classify the sentiment of a whole review to one of the predefined sentiment polarities, including positive, negative and sometimes neutral.

 While phrase-level analysis attempt to extract the sentiment polarity of each feature that a user expresses his/her attitude to the specific feature of a specific product.

Disadvantages of Existing System

 The existing work mainly focuses on classifying users into binary sentiment (i.e. positive or negative), and they do not go further in mining user's sentiment.

 The existing approaches mainly leverage product category information or tag information to study the interpersonal influence.

 These methods are all restricted on the structured data, which is not always available on some websites. However, user reviews can provide us ideas in mining interpersonal inference and user preferences.

Proposed System

 We propose a sentiment-based rating prediction method in the framework of matrix factorization. In our work, we make use of social users' sentiment to infer ratings.

 First, we extract product features from user reviews. Then, we find out the sentiment words, which are used to describe the product features. Besides, we leverage sentiment dictionaries to calculate sentiment of a specific user on an item/product.

 The main contributions of our approach are as follows:

 We propose a user sentimental measurement approach, which is based on the mined sentiment words and sentiment degree words from user reviews.

 We make use of sentiment for rating prediction. User sentiment similarity focuses on the user interest preferences. User sentiment influence reflects how the sentiment spreads among the trusted users. Item reputation similarity shows the potential relevance of items.

 We fuse the three factors: user sentiment similarity, interpersonal sentimental influence, and item reputation similarity into a probabilistic matrix factorization framework to carry out an accurate recommendation. The experimental results and discussions show that user's social sentiment that we mined is a key factor in improving rating prediction performances.

Advantages of Proposed System

 In our paper, we not only mine social user's sentiment, but also explore interpersonal sentimental influence and item's reputation. Finally, we take all of them into the recommender system.

 The purpose of our approach is to find effective clues from reviews and predict social users' ratings.

 We fuse user sentiment similarity, inter personal sentiment influence, and item reputation similarity into a unified matrix factorization frame work to achieve the rating prediction task.

Software Requirements:

Operating System: Windows

Coding Language: Java

Server: Tomcat

Database : Sqlyog

Hardware Requirements:

Processor - inter core i3

Speed – 2.4 GHz

RAM - 512 MB (min)

INPUT AND OUTPUT DESIGN

Input Design

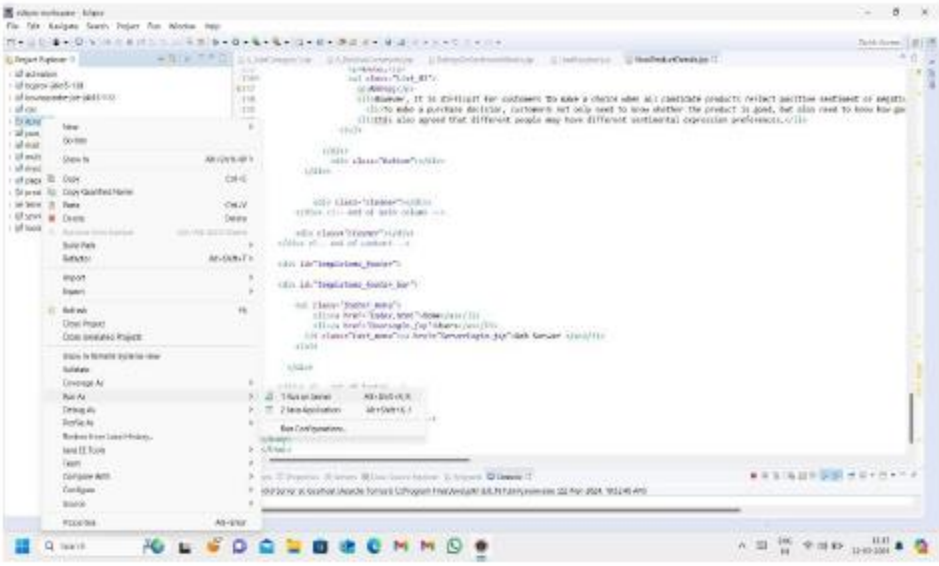
Input Design plays a vital role in the life cycle of software development, it requires very careful attention of developers. The input design is to feed data to the application as accurate as possible. So inputs are supposed to be designed effectively so that the errors occurring while feeding are minimized. According to Software Engineering Concepts, the input forms or screens are designed to provide to have a validation control over the input limit, range and other related validations.
This system has input screens in almost all the modules. Error messages are developed to alert the user whenever he commits some mistakes and guides him in the right way so that invalid entries are not made. Let us see deeply about this under module design.
Input design is the process of converting the user created input into a computer-based format. The goal of the input design is to make the data entry logical and free from errors. The error is in the input are controlled by the input design. The application has been developed in user-friendly manner. The forms have been designed in such a way during the processing the cursor is placed in the position where must be entered. The user is also provided with in an option to select an appropriate input from various alternatives related to the field in certain cases.
Validations are required for each data entered. Whenever a user enters an erroneous data, error message is displayed and the user can move on to the subsequent pages after completing all the entries in the current page.

Output Design

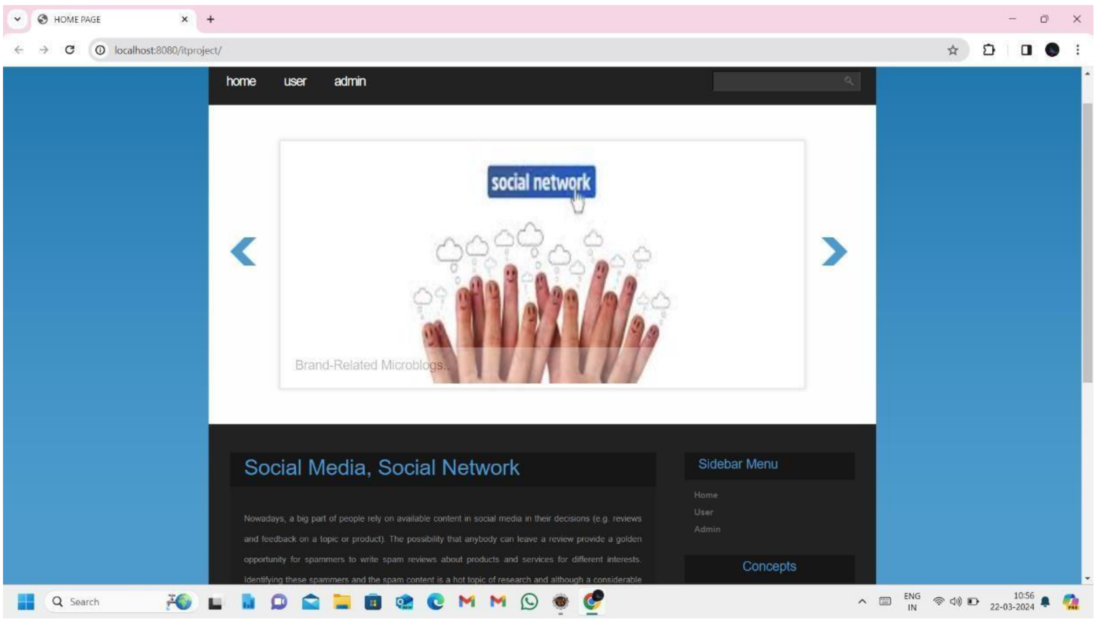
ISSN: 2278-4632
Vol-14 Issue-02 Aug 2024

The Output from the computer is required to mainly create an efficient method of communication within the company primarily among the project leader and his team members, in other words, the administrator and the clients. The output of VPN is the system which allows the project leader to manage his clients in terms of creating new clients and assigning new projects to them, maintaining a record of the project validity and providing folder level access to each client on the user side depending on the projects allotted to him. After completion of a project, a new project may be assigned to the client. User authentication procedures are maintained at the initial stages itself. A new user may be created by the administrator himself or a user can himself register as a new user but the task of assigning projects and validating a new user rests with the administrator only.
The application starts running when it is executed for the first time. The server has to be started and then the internet explorer in used as the browser. The project will run on the local area network so the server machine will serve as the administrator while the other connected systems can act as the clients. The developed system is highly user friendly and can be easily understood by anyone using it even for the first time.

RESULTS

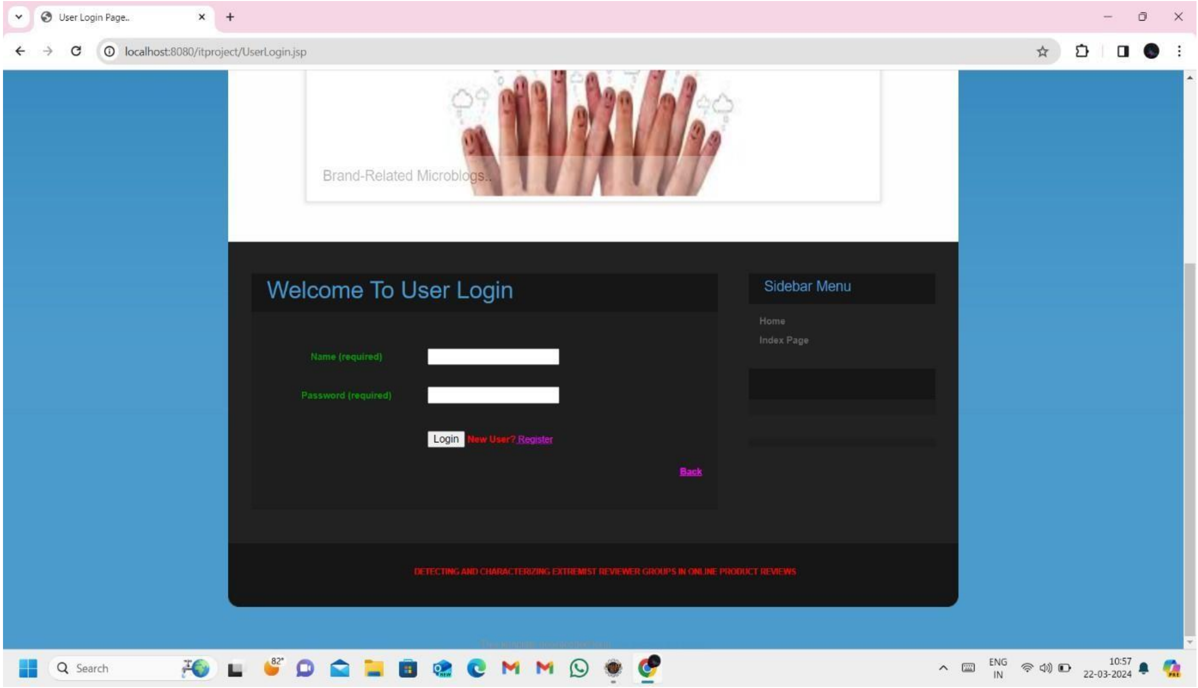


Running the code

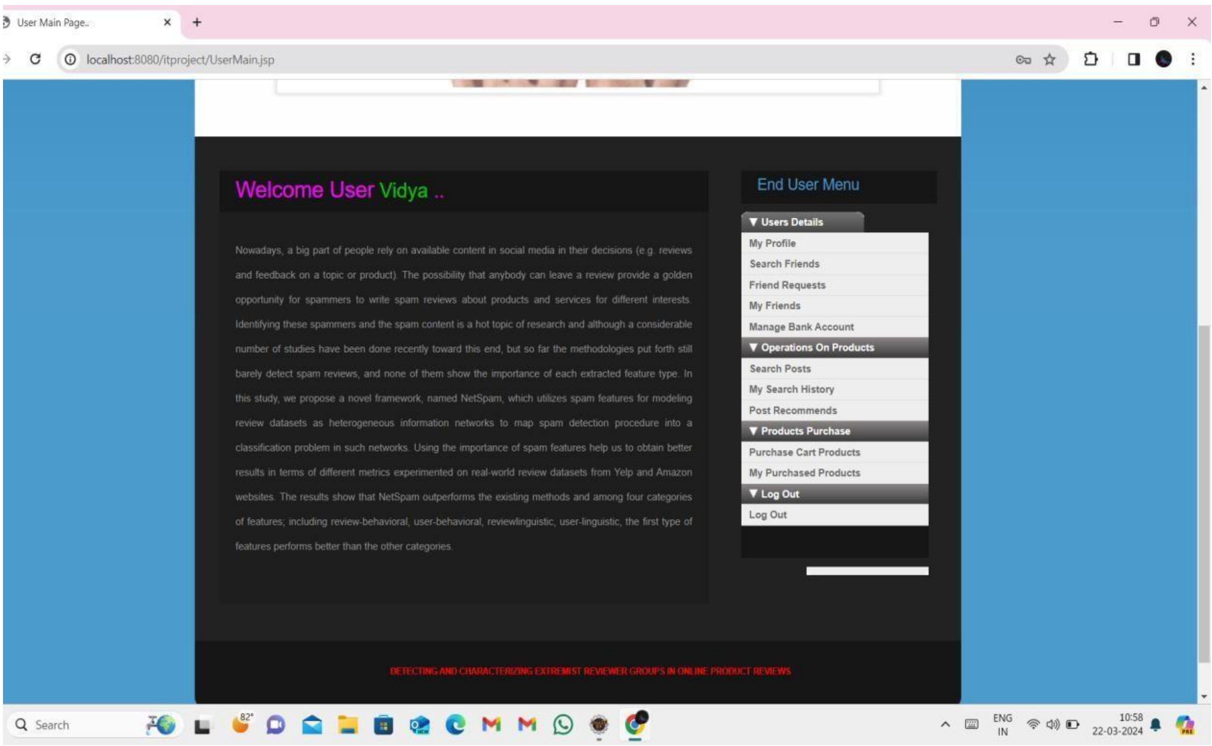


Home Page

ISSN: 2278-4632
Vol-14 Issue-02 Aug 2024

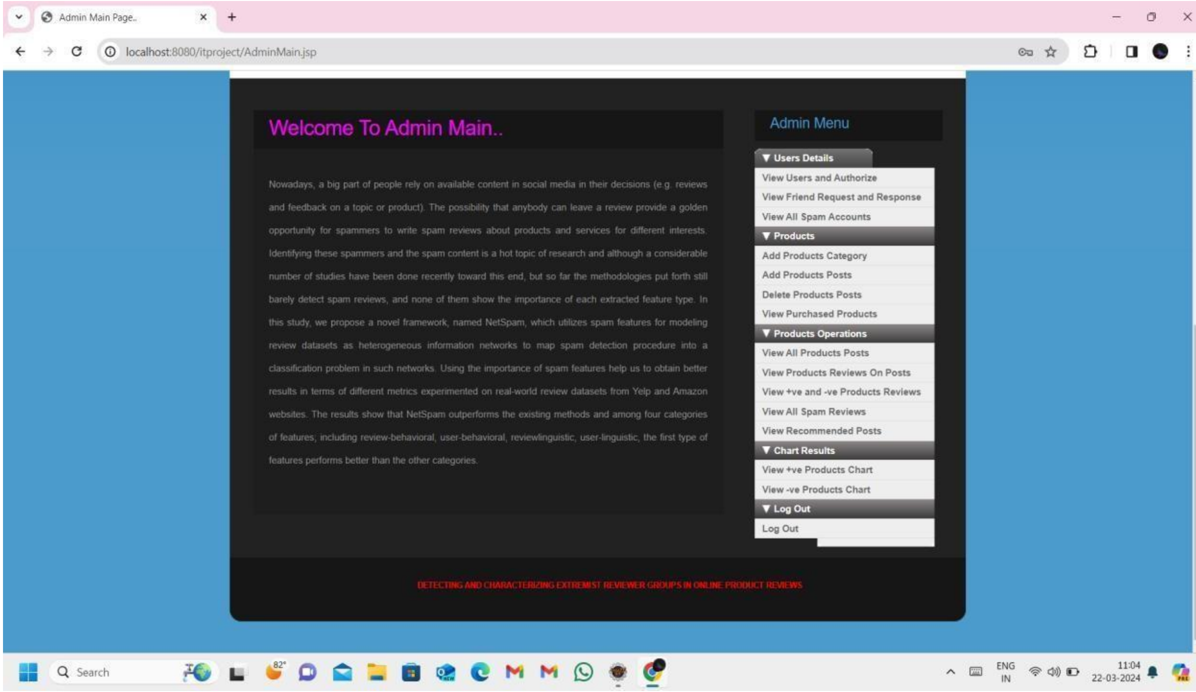


User Login Page



User’s Welcome page

ISSN: 2278-4632
Vol-14 Issue-02 Aug 2024

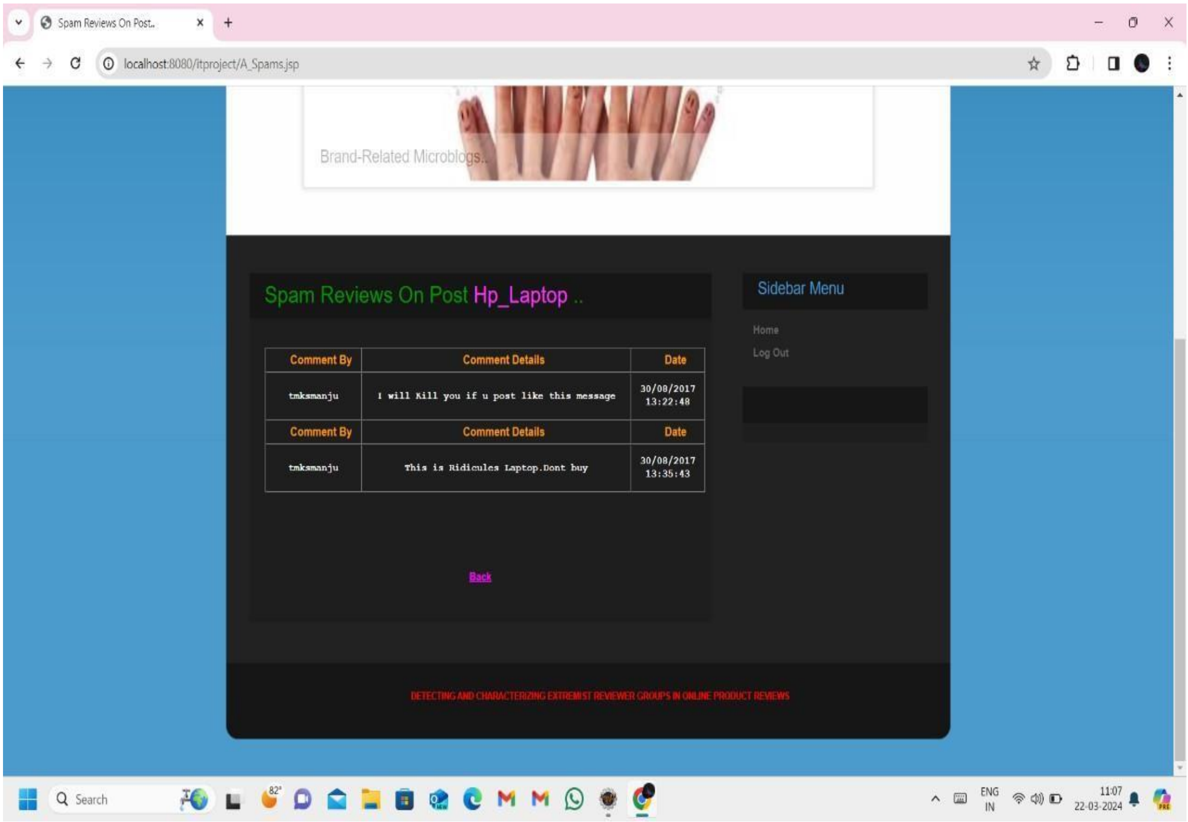


Admin Welcome page

Admin Authorizes Users

ISSN: 2278-4632
Vol-14 Issue-02 Aug 2024

Chat representation of Sentiment of Reviews



Identification of Spam Reviews

ISSN: 2278-4632
Vol-14 Issue-02 Aug 2024

CONCLUSION AND FUTURE ENHANCEMENT

In this article, we discussed an unexplored form of opinion spam, where spammers target brands as a whole, posting extreme reviews, to change the overall sentiment about the brand. These groups are often part of a complex business Web that is capable of influencing the overall popularity and reputation of several brands on review websites. This article is the first step toward linking brand-level group activities and extremism in reviews, which uncovers important insights about marketplace activities. These insights would help in developing a better recommendation that makes use of online reviews. A set of candidate spam groups was retrieved using FIM, and extremist groups were identified by observing their actions as a group based on various features, using a supervised learning technique based on a ground truth of manually annotated labels. We then classified extremist and moderate groups and compared the accuracy across multiple classification methods. After classifying these groups, we observed the behaviors for extremist groups in detail to gain further insights about the phenomenon and the overall trends of how these groups target these brands. We have also released the codes and annotated data set for further studies.

BIBLIOGRAPHY

[1] A. Kim. (2017). That review you wrote on Amazon? Priceless. [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. Supported Cooperat. 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 using factor 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," in Proc. 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 classification 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 augmenting e-commerce recommendation," 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 distributed representations from reviews for collaborative filtering," 2018, arXiv:1806.06875. [Online]. Available: http://arxiv.org/abs/1806.06875

[16] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in Proc. HLT, 2005, pp. 339–346.

[17] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the Web," in Proc. 14th Int. Conf. World Wide Web (WWW), 2005, pp. 342–351.

[18] M. Hu and B. Liu, "Mining opinion features in customer reviews," in Proc. AAAI, 2004, pp. 755–760.

[19] T. Donkers, B. Loepp, and J. Ziegler, "Explaining recommendations by means of user reviews," in Proc. 1st Workshop Explainable Smart Syst. (ExSS), 2018. [Online]. Available: http://ceur-ws.org/Vol- 2068/exss8.pdf [20] B. Pang et al., "Opinion mining and sentiment analysis," Found. Trends Inf. Retr., vol. 2, nos. 1–2, pp. 1–135, 2008.

[21] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2004, pp. 168–177.

[22] G. Somprasertsri and P. Lalitrojwong, "Mining feature-opinion in online customer reviews for opinion summarization," J. UCS, vol. 16, no. 6, pp. 938–955, 2010.

[23] L. Zhuang, F. Jing, and X.-Y. Zhu, "Movie review mining and summarization," in Proc. 15th ACM Int. Conf. Inf. Knowl. Manage. (CIKM), 2006, pp. 43–50.