IDENTIFYING AND DESCRIBING EXTREMIST REVIEWER GROUPS IN OPERATIONAL PRODUCT REVIEW

B.R.Teja[1], R Venkat Raj Kumar[2], Dr.A.Manimaran[3]

[1,2]PG Scholar, [3]Assistant Professor
Department of Computer Applications
Madanapalle Institution of Technology and Science Madanapalle

***Abstract*: Nowadays, online marketplaces are frequently subjected to opinion spam in the form of reviews. Individuals are generally hired to encourage or sidetrack specific brands by writing highly positive or negative reviews. [1,2,3]This is regularly done in groups. Although previous studies attempted to identify and analyze such opinion spam groups, little has been conducted in order to identify those groups which thus target a brand as a whole, rather than always products. We collected reviews from the Amazon product review site and manually labeled a set of 923 candidate reviewer groups for this application. Users are clustered together if they have mutually reviewed (products of) a large number of manufacturers, as determined by request itemset mining over brand obvious parallels.**

***Keywords*: Behavior, electronic commerce, machine intelligence, machine learning, reviews, social computing, web mining.**

1. INTRODUCTION:

Users are clustered together if they have mutually reviewed (products of) a large range of manufacturers, as calculated by frequent item set mining over brand similarities. The nature of the reviewer groups, we reasonably deduce, is largely decided by eight features unique to a (group, brand) pair. To classify candidate groups as extremist entities, we develop a feature-based supervised model. We run multiple classifiers for the task of classifying a group based on user reviews to determine whether the group shows signs of extremism. The best classifier is a three-layer perceptron-based classifier.

We investigate the behavioral responses of such groups in considerable depth in order to better understand the dynamics of brand-level opinion fraud. Among these behavioral patterns are:

Product-of-experts model and a recurrent neural network. We demonstrate that the increased flexibility offered by the product-of-experts model allowed it to achieve state-of-the-art performance on the Amazon review dataset, outperforming the LDA-based approach. However, interestingly, the greater modeling power offered by the recurrent neural network appears to undermine the model's ability to act as a regularizer of the product representations.

PROPOSED METHOD

In proposed system, unlike other studies that majorly focus on fake review/reviewer detection, we here focus on extremist reviewer detection, which may not be fake. Moreover, we attempt to identify “groups” instead of detecting “individual user” by using machine learning algorithms

ADVANTAGES:

- It can detect the fake reviews easily.
- Accuracy is high while detecting fake reviews.
- It can identify groups rather than individual user.

2. IMPLEMENTATION:

System
User
1. SYSTEM

1.1 Take dataset:
System will take uploaded dataset.
1.2 Preprocessing:
If any null values are present in dataset that can removed in preprocessing step.
1.3 Training and testing:
System take data for training and testing based on user given test size.
1.4 Prediction:
System gives the predictions based on given characteristics.
1.5 predictions:
Using the machine leaning algorithms, we can predict the result

2. User:

2.1 Data collection:

Users collect the data from the different websites.

2.2 Upload dataset:

User uploads the dataset in this step.

2.2 View dataset:

In this step uploaded dataset is viewed by the user.

2.3 Model performance:

In model performance step machine learning supervised algorithms are performed.

2.4 View predictions:

Users view the predicted outputs**.**

LITERATURE SURVEY

[1]. H. S. Dutta, V. R. Dutta, A. Adhikary, and T. Chakraborty, "HawkesEye: Detecting fake retweeters using Hawkes process and topic modeling,"

Retweets are essential for increasing the popularity of a tweet, and a large number of fake retweeters can make a significant contribution to this aspect. We define a fake retweeter as a Twitter account that retweets spammy tweets, retweets an abnormally large amount of tweets in a short period, or misuses a trending hashtag to promote events irrelevant to the topic of discussion. To address the problem of fake retweeter detection, we introduce an up-to-date, temporally diverse, trend-oriented large dataset. In contrast to existing approaches, which require a graph-like relationship between tweet entities or the presence of a retweeter's entire reaction was initiated sequence of events, we develop a novel classifier called HawkesEye, which makes predictions based on a temporal window. HawkesEye means making use of both space - time and rate of incidence.

[2]. S. Dhawan, S. C. R. Gangireddy, S. Kumar, and T. Chakraborty, "Spotting collective behaviour of online frauds in customer reviews,"

Before purchasing any product, online reviews could perhaps help you determine it's own quality. Unfortunately, spammers quite often were using online review forums to promote/demote particular brands by writing outright fraud reviews. It may be more dangerous when such spammers work collaboratively and cumulatively incorporate spammers because then they can achieve complete control of user sentiment due to the volume of fraudulent reviews they inject. Due to misunderstanding of something like the definition of a group, the modification of inter-group dynamics, this same scarcity of clearly labelled group-level spam data, and perhaps other factors, group spam detection is more incredibly hard than individual-level fraud detection. DeFrauder, an unsupervised method for designed to detect online fraud reviewer groups, is proposed here. It needs to begin by designed to detect candidate fraud groups by leveraging the underlying product review graph and incorporating several cognitive and behavioural predictors.

[3]. Y. Wang, J. Wang, and T. Yao, "What makes a helpful online review? A meta-analysis of review characteristics,".

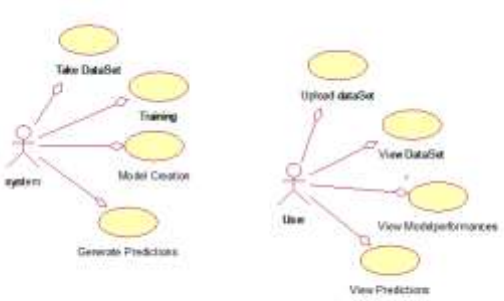
The purpose of this study is to identify the factors that influence online review usefulness in terms of review depth, extremity, and timeliness. We examine the effects of important review characteristics using a meta-analysis of 53 empirical studies yielding 191 effect sizes. With the exception of its sub-metric of review volume, which has a negative influence on review helpfulness, findings show that review depth has a greater impact on helpfulness than review extremity and timeliness. In particular, readability is the most important factor in determining review usefulness. Furthermore, we discuss important relationship moderators and discover interesting insights about the website and cultural background. In light of the findings, we propose several implications for researchers and E-business firms. Our research provides a much-needed quantitative synthesis of this information.

[4] Urcuqui, Christian, Andres Navarro, Jose Osorio, and Melisa Garcia. "Machine Learning Classifiers to Detect Malicious Websites." In SSN, pp. 14-17. 2017.).

We tackle the issue of automating network troubleshooting for a large-scale WiFi network. We are specifically interested in determining the causes of unnecessary active scans in WiFi networks, which are known to degrade WiFi performance. To train various machine learning models, we collect 340 hours of data from thousands of active scan episodes. Data is collected using 27 devices from various vendors in various network configurations in a controlled environment. We investigate unsupervised and supervised machine learning techniques and conclude that a multilayer perceptron is the most effective model for detecting the causes of active scanning. Furthermore, we validate the model in vivo in an uncontrolled real-world WiFi network [5]. "Learning Distributed Representations," A. Almahairi, K. Kastner, K. Cho, and A. Courville.

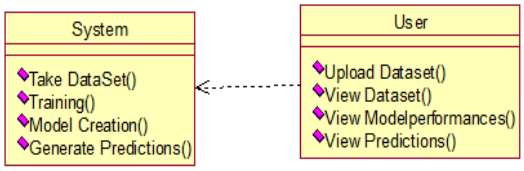
Use case diagrams:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



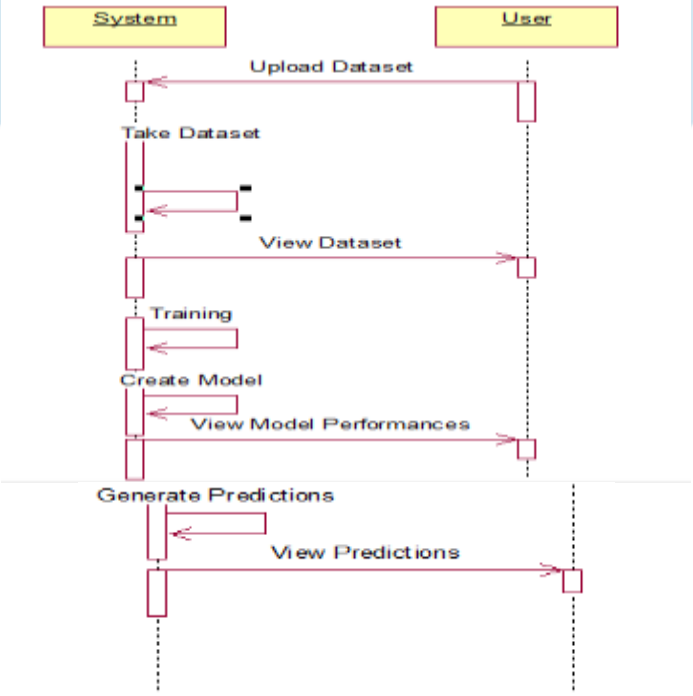
Class diagrams:

In software engineering, a class diagram in the Unified Modelling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information



SEQUENCE DIAGRAM:

A sequence diagram in Unified Modelling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



RESULTS:

Using SVM and after applying PCA and feeding the findings to train SVM, the model produced the following results. Only employing SVM improves prediction accuracy. The accuracy acquired by the first method was found to be 90.4 percent, whereas the accuracy gained by the second method was found to be 90.4 percent. PCA is used first, followed by SVM with a linear kernel function**.** The dataset has a 97.75 percent accuracy.

The following is the classification report:

Table 1: Confusion Matrix

Actual/ Predicted	Benign	Malware
Benign	1415	114
Malware	42	4313

Table 2: Classification Report

Classification Report	Precision	Recall	f-1 Score	Support
Benign	0.97	0.93	0.95	1592
Malware	0.97	0.99	0.98	4555
Micro Average	0.97	0.97	0.97	5884
Macro Average	0.97	0.96	0.96	5884
Weight Average	0.97	0.97	0.97	5884

6 .CONCLUSION:

In this article, we discussed an unexplored type of opinion spam in which spammers target entire brands and post extreme reviews in order to change the overall sentiment about the brand. These organisations are frequently part of a complex business Web that has the potential to influence 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 aid in the development of a better recommendation that incorporates online reviews. FIM was used to retrieve a set of candidate spam groups, and extremist groups were identified by observing their actions as a group based on various features, using a supervised learning technique based on

REFERENCES:

[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.