

PRODUCT RECOMMENDATION SYSTEM USING SENTIMENT ANALYSIS BASED ON MACHINE LEARNING

¹ Prof. Salve B. S., ²Ms. Gadade Anagha Savata, ³Ms. Kazi Sabiya Jakir,
⁴Ms. Bhitade Ashwini Bharat, ⁵Ms. Bhagat Punam Rajaram

¹Assistant Professor, ²Student, ³Student, ⁴Student, ⁵Student

¹Department of Computer Engineering,

¹S. B. Patil College of Engineering Indapur, India

Abstract : Nowadays, People place their trust in products based on product reviews and ratings. Customers are increasingly relying on online product reviews to help them make purchasing decisions. Unfortunately, the importance of reviews has been misused by some parties who have attempted to produce fake reviews in order to create the popularity of a product or to discredit the product. Recognizing these fake reviews and spam content is big debated issue of research. In this study, a new structure named Netspam which uses spam feature for demonstrating review data set as heterogeneous information framework to design spam identification method into a group of issue in this network. Extensive experiment on real dataset show the effectiveness of our method.

IndexTerms - Product Recommendation System, Sentiment Analysis, Semantic Analysis, fake reviews, fake reviews detection, spam detection.

I. INTRODUCTION

In today's, world every people trust on product on basis of product reviews & rating. In general, e-commerce sites allow customers to leave feedback about their services. Nowadays, the majority of customers express their feeling for items on online journal, web-based business, review destination. Related with financial benefit fraudster tried to play with existing system by writing fake reviews to discredit product or service[8]. This spam review had represented genuine danger to web-based business & organization.

In this study, we propose Netspam, a novel structure that employs spam highlights to demonstrate a review dataset as a heterogeneous information network and to develop a spam detection approach into the classification issue in such a network. This method is based on a survey dataset and allows users to build metapaths based on review metadata[10]. Spam feature help us to obtain better outcome regarding different metrics on review dataset. Netspam uses four feature including review behavioral, user behavioral, review linguistic, user linguistic. First type of feature perform better than other. The contributinal work is when user search query it will display all top-k products as well as recommendation of product.

II. RELATED WORK

Sr. No.	Paper title	Author	Year	Problems in existing system	Solution to existing system
1	Combating product review spam campaigns via multiple heterogeneous pairwise features.	Ch. Xu and J. Zhang.	14	Difficult problem to automate.	1. Pairwise features can be more robust model for correlating colluders. 2. To manipulate perceived reputations of the targets for their best interests. 3. To rank all the reviewers in the

					website globally so that top-ranked ones are more likely to be colluders.
2	Exploiting burstiness in reviews for review spammer detection..	G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh	2013	1. a generic framework is not used for detect spammers	<ol style="list-style-type: none"> 1. High accuracy. 2. The proposed method is effective. 3. To detect review spammers in review bursts. 4. To detect spammers automatically.
3	Trueview: Harnessing the power of multiple review sites.	A. j. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos.	2015	1. Difficult problem to automate.	<ol style="list-style-type: none"> 1. Develop novel features capable of identifying cross-site discrepancies effectively. 2. A hotel identity-matching method with 93% accuracy. 3. Enable the site owner to detect misbehaving hotels. 4. Enable the end user to trusted reviews.
4	Towards detecting anomalous user behavior in online social networks.	B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A.	2014	1. The attacker is trying to drain the budget of some advertiser by clicking on ads of that advertiser.	<ol style="list-style-type: none"> 1. Anomaly detection technique to effectively identify anomalous likes on Facebook ads. 2. Achieves a detection rate of over 66% (covering more than 94% of

		Mislove.			misbehavior) with less than 0.3% false positives.
5	Spotting fake reviews via collective PU learning.	H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao.	2014	<p>1. Fake reviews hiding in the unlabeled reviews that Dianping's algorithm did not capture.</p> <p>2. The ad-hoc labels of users and IPs used in MHCC may not be very accurate as they are computed from labels of neighboring reviews.</p>	<p>1. Proposed models can markedly improve the F1 scores of strong baselines in both PU and non-PU learning settings.</p> <p>2. Models only use language independent features; they can be easily generalized to other languages.</p> <p>3. Detects a large number of potential fake reviews hidden in the unlabeled set.</p>
6	Reducing Feature Set Explosion to Faciliate Real-World Review Spam Detection	M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa,"	2016	Detection Accuracy is low	In response to this growing problem, there have been many studies on the most effective ways of detecting review spam using various machine learning algorithms. One common thread in most of these studies is the conversion of reviews to word vectors, which can potentially result in hundreds of thousands of

					features.
7	Trust-Aware Review Spam Detection	H. Xue, F. Li, H. Seo, and R. Pluretti,”	2015	Trust aware review spam detection process difficult to automate.	The contributions of this paper are two-fold: (1) elaborate how social relationships can be incorporated into review rating prediction and propose a trust based rating prediction model using proximity as trust weight; and (2) design a trust-aware detection model based on rating variance which iteratively calculates user-specific overall trustworthiness scores as the indicator for spam city.
8	Fake Review Detection From a Product Review Using Modified Method of Iterative Computation Framework	E. D. Wahyuni , A. Djunaidy	2016	The result from the experiment shows that the proposed system has a low accuracy	The honesty value of a review will be measured by utilizing the text mining and opinion mining techniques.
9	Anomaly Detection in	R. Hassanzadeh	2014	This technology also opens the door	Detecting anomalies, in this

	Online Social Networks: Using Datamining Techniques and Fuzzy Logic			for unlawful activities.	new perspective of social life that articulates and reflects the off-line relationships, is an important factor as they could be a sign of a significant problem or carrying useful information for the analyzer.
10	Collective opinion spam detection: bridging review networks and metadata	R. Shebuti, L. Akoglu	2015	Not efficient	SpEagle employs a review-network-based classification task which accepts prior knowledge on the class distribution of the nodes, estimated from metadata. When it available then it enables seamless integration of labeled data.

III. WORK PROGRESS

We divide the whole project in to the three parts. Which are as follow as

1. Data Collection : In data collection, we collect reviews of products from online shopping website. In our project we collect reviews from [amazon.com](https://www.amazon.com) website. This data is stored in database.
2. Data Processing : After collecting the data or reviews, we do pre processing on that data and remove unwanted data from it by using Sentiment Analysis algorithm and Latent Semantic Analysis algorithm. After that we calculate weight of reviews and separate the real and fake/spam reviews with the help of two views of features(review-user and behavioral-linguistic).
3. Output : After this process, we recommend top products to customer or user and classify spam reviews.

IV. PROPOSED SYSTEM

Problem Statement :

To develop and identify the spammers and spam content using reviews and performs four categories of features including review-behavioral, user-behavioral, review-linguistic and user-linguistic features for showing only trusted reviews to the user's side.

Goals and Objective :

Goals:

1. People trust written reviews in their decision-making processes, with good and negative evaluations encouraging or discouraging them in their product and service selections.
2. Anyone can register for the app and leave comments as spam reviews to influence the opinions of other users.

Objectives:

1. To identify spam and spammers, and also conduct other types of analysis on the subject.
2. Written reviews can also assist service providers in improving the quality of their goods and services.
3. Using positive and negative evaluations in online social media to identify spam users.
4. Only show users reviews from people they can trust.

Overview of Proposed system:

1. A novel proposed framework comprises of describing a set of product review data as an HIN (Heterogeneous Information Network) and tackling the problem of spam identification in HIN classification.
2. To display their views dataset as an HIN, in which the reviews are linked by various sorts of nodes (such as functionality and users).
3. The relevance (or weight) of each function is then calculated using a weighing technique.
4. Using controlled and unsupervised techniques, these weights are used to create the most recent review labels.
5. In both semi-supervised and unsupervised approaches, the classified features as review behavioural have greater weights and give higher performance on recognising spam reviews, based on our observations, defining two viewpoints for features (review-user and behavioral-linguistic).
6. For labelling, feature weights can be added or subtracted, and thus time complexity can be tuned to a specified level of accuracy.
7. These features into four categories (review-behavioral, user-behavioral, review-linguistic, and user-linguistic) allows us to see how much each category contributes to spam identification.

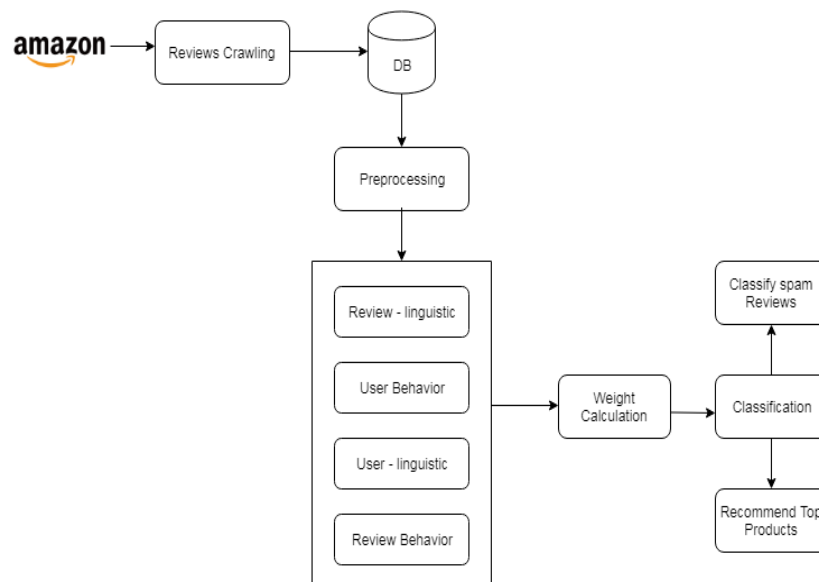


Figure 1: Proposed system architecture

V. APPLICATIONS :

1. Recommendation Applications.

VI. ADVANTAGES :

1. It detects spam and spammers, as well as many types of analysis in this field.
2. Written reviews can also assist service providers in improving the quality of their goods and services.
3. It detects spam users in online social media by analysing positive and negative evaluations.
4. This framework only shows users reviews from people they can trust.

VII. DISADVANTAGES :

1. Need of internet.
2. Not used images.

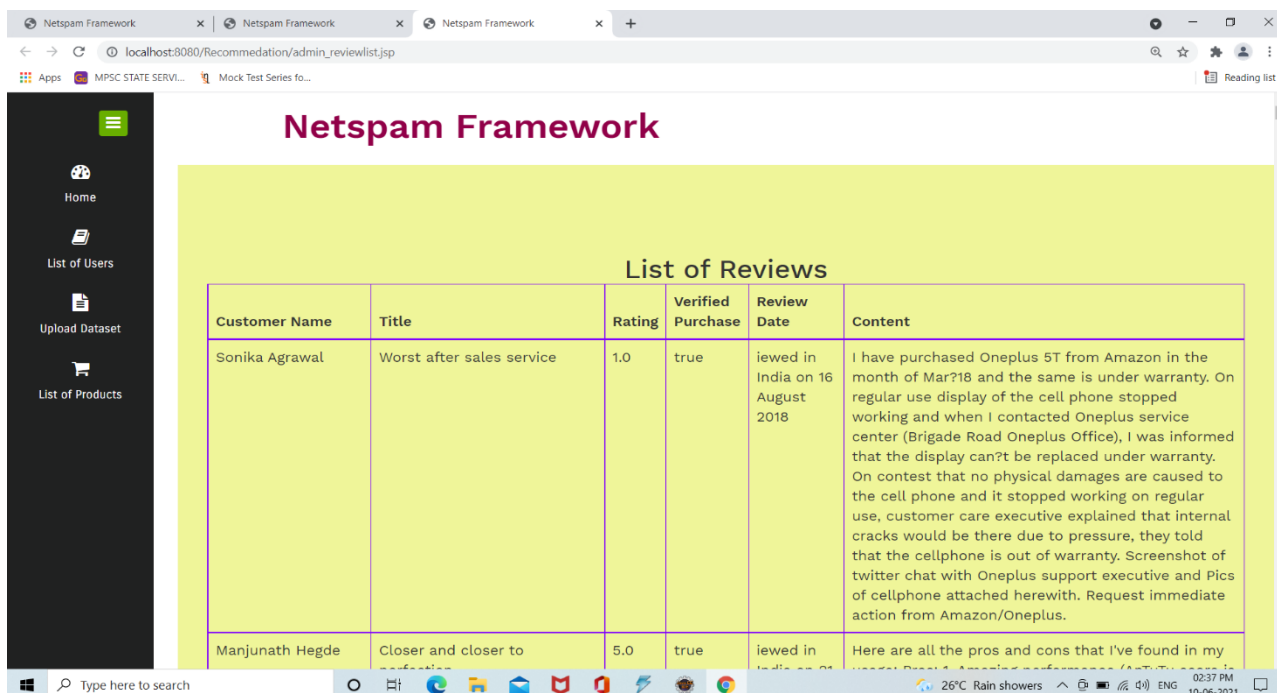
VIII. RESULT :

1. Product List Page

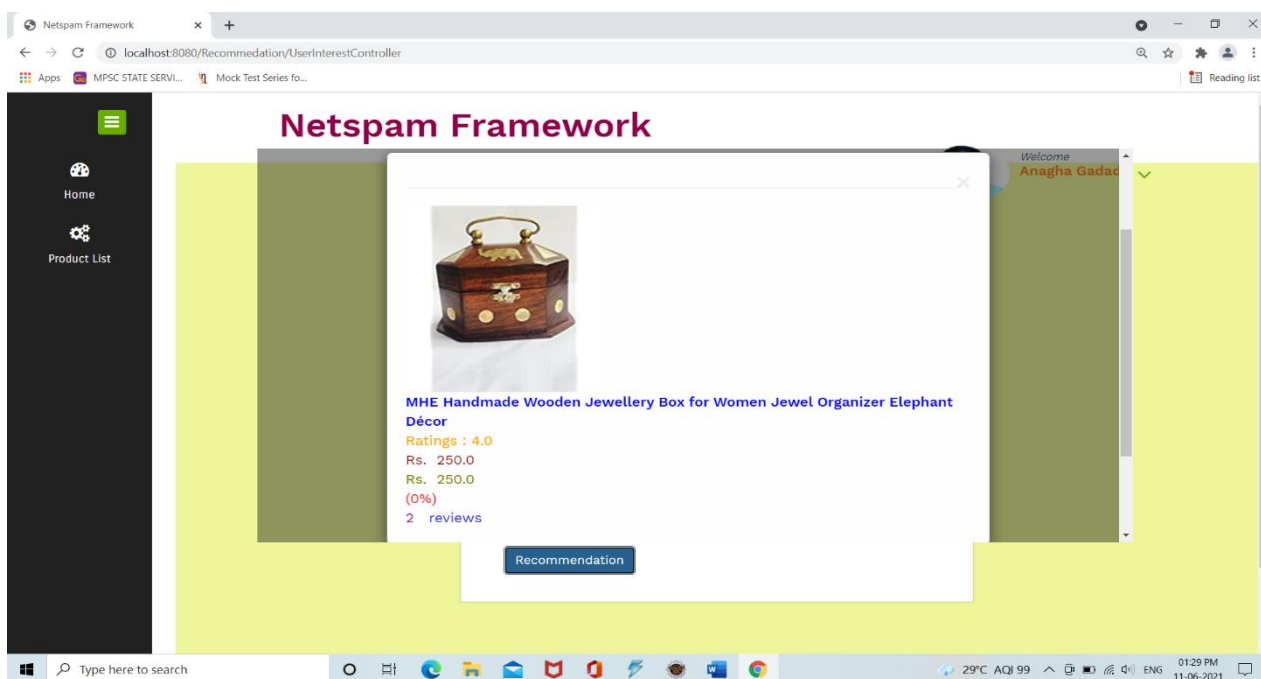
The screenshot displays the 'Netspam Framework' web application. The browser address bar shows 'localhost:8080/Recommendation/AdminListProductController'. The page title is 'Netspam Framework'. A sidebar on the left contains navigation links: Home, List of Users, and List of Products. The main content area is titled 'List of Products' and contains a table with the following data:

Id	Image	Item Id	Item Name	Price	Rating	Category	Reviews
1		B0756ZFXVB	OnePlus 5T (Midnight Black 6GB RAM + 64GB memory)	32999.0	4.5	Mobile	Extract
2		B01FVNS732	Lava A52 (Black, 4GB)	3599.0	2.5	Mobile	Extract
3		B07BHBH9KV	Vivo V9 (19:9 FullView Display, Pearl Black) with Offers	23990.0	4.5	Mobile	Extract
4		B071HWTHPH	Moto G5s Plus (Lunar Grey, 64GB)	16999.0	3.8	Mobile	Extract

2. Review List Page



3. Recommended Product Page



IX. CONCLUSION :

In this system investigation presents a novel spam detection system in particular NetSpam framework for product reviews based on Sentiment analysis (SA) and latent semantic analysis (LSA). This paper has used SA and LSA with netspam algorithm for spam detection. Furthermore, defining four feature for highlights our perceptions show that the behavioural classification reviews. LSA is used in the proposed system to reduce similar comments and try to improve spam detection accuracy.

REFERENCES

- [1] Ch. Xu and J. Zhang, "Combating product review spam campaigns via multiple heterogeneous pairwise features", In SIAM International Conference on Data Mining, 2014.
- [2] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting business in reviews for review spammer detection", In ICWSM, 2013.
- [3] A. j. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos, "True view: Harnessing the power of multiple review sites", In ACM WWW, 2015.
- [4] B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Towards detecting anomalous user behavior in online social networks", In USENIX, 2014.
- [5] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, "Spotting fake reviews via collective PU learning", In ICDM, 2014.
- [6] M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa, "Reducing Feature Set Explosion to Facilitate Real-World Review Spam Detection", In Proceeding of 29th International Florida Artificial Intelligence Research Society Conference, 2016.
- [7] H. Xue, F. Li, H. Seo, and R. Pluretti, "Trust-Aware Review Spam Detection", IEEE Trustcom/ISPA., 2015.
- [8] E. D. Wahyuni , A. Djunaidy, "Fake Review Detection From a Product Review Using Modified Method of Iterative Computation Framework", In Proceeding MATEC Web of Conferences, 2016.
- [9] R. Hassanzadeh, "Anomaly Detection in Online Social Networks: Using Datamining Techniques and Fuzzy Logic", Queensland University of Technology, Nov, 2014.
- [10] R. Shebuti, L. Akoglu, "Collective opinion spam detection: bridging review networks and metadata", In ACM KDD, 2015.