FAKE PRODUCT REVIEW DETECTION SYSTEM USING MACHINE LEARNING

[1]**S.A.Sagar,** [2]**Dharati Jawale**, [3]**Harshada Jedhe,** [4] **Rutuja Kakade,** [5]**Akanksha Durugkar**

[1]*Assistant Professor, Department of Information Technology,Bharati Vidyapeeth's College Of Engineering For Women, Savitribai Phule University,India.*

[2]*UG Student, Department of Information Technology, Bharati Vidyapeeth's College Of Engineering For Women, Savitribai Phule University,India.*

[3]*UG Student, Department of Information Technology,Bharati Vidyapeeth's College Of Engineering For Women, Savitribai Phule University,India.*

[4]*UG Student, Department of Information Technology,Bharati Vidyapeeth's College Of Engineering For Women, Savitribai Phule University,India*

[5]*UG Student, Department of Information Technology,Bharati Vidyapeeth's College Of Engineering For Women, Savitribai Phule University,India*

Abstract : Online systems people most of the time trust in products based on product reviews and ratings. Reviews have a great impact on a company or a brand profile. The company must examine market reactions to its products. Popular reviews, on the other hand, are difficult to track and arrange. In social media, many people viewpoints are difficult to manually process. The next step is to develop a mechanism for automatically categorizing favorable and negative public feedback. Customers will be able to see how the product performs in terms of consistency, efficiency, and guidance, which will provide prospective buyers a better knowledge of the product. The applicability of web contents from suppliers in order to fulfill client requirements by evaluating beneficial input is one such unfulfilled possibility. Good and bad reviews are important in determining customer demands and gathering product feedback from customers. Sentiment Analysis is a type of text analysis that collects contextual information from text. A large number of internet mobile phone ratings are studied in this study. We divided the text into positive and negative categories, as well as feelings of disappointment, expectation, disgust, trepidation, delight, regret, surprise, and confidence. This clearly defined category of feedback aids in a comprehensive evaluation of the product, allowing consumers to make better decisions.

IndexTerms - Machine Learning, Social Media, Text Mining, Sentiment Analysis, Semantic Analysis, Online Reviews

1. INTRODUCTION

Online Social Media sites play an important role in information propagation which is   taken   into   account as a main source for producers in their advertising campaigns as for customers in selecting products   and   services. Within the past few years, people be dependent on the constructed audits in their dynamic procedures, and positive/negative reviews promising/unpromising them in their selection   of   products and services. Additionally, written reviews also help service providers to highlight the quality of their products and services. These reviews   thus became important for the success of business. While positive reviews can bring benefits for a business, negative reviews can potentially impact business and cause economic losses. The very reality that anybody with any personality can leave remarks as audit gives a chance to spammers to record counterfeit surveys intended to delude clients conclusion.

2. LITERATURE SURVEY

Detailed pair-wise features are first explicitly utilized to detect group colluders in online product review spam campaigns, which can reveal collisions in spam campaigns from a more fine-grained perspective.

In [1] paper, Spam campaigns spotted in popular product review websites (e.g., amazon. com) have attracted both industry and academia, where a group of online posters are hired to collaboratively craft deceptive reviews for some target products. The goal is to manipulate perceived reputations of the targets for their best interests.

In [2] paper, Online product reviews have become an important source for user opinions. For profit or fame, imposters have been writing deceptive or fake reviews to promote or to demote some target products or services. In the last few years, several approaches have been proposed to deal with the problem. In this work, take a different approach, which exploits the fraud nature of reviews to identify review spammers.

In [3] paper, Online reviews on products and services can be very useful for customers, but they need to be protected from manipulation.So far, most studies have focused on analyzing online reviews from a single hosting site. How could one leverage information from multiple review hosting sites? This is the key question in our work. In response, develop a systematic methodology to merge, compare, and evaluate reviews from multiple hosting sites. focus on hotel reviews and use more than 15 million reviews from more than 3.5 million users spanning three prominent travel sites.

In [4] paper, almost all users rely on crowd sourced information, such as reviews on Yelp and Amazon, and liked posts and ads on Facebook. This has led to the market for black hat promotion techniques via fake (e.g., Sybil) and compromised accounts, and collusion networks. Existing approaches to detect such behavior relies mostly on supervised (or semi-supervised) learning over

known (or hypothesized) attacks. They are unable to detect attacks missed by the operator while labeling, or when the attacker changes strategy.

In [5] paper, Online reviews have become an increasingly important resource for decision making and product designing. But review systems are often targeted by opinion spamming. Although fake review detection has been studied by researchers for years using supervised learning, ground truth of large scale datasets is still unavailable and most of existing approaches of supervised learning are based on pseudo fake reviews rather than real fake reviews. Working with Dianping1, the largest Chinese review hosting site, present the first reported work on fake review detection in Chinese with filtered reviews from Damping's fake review detection system.

Disadvantages:

There is not any information filtering concept in online shopping sites.
Individuals accept the composed audits in their official procedures.
Less accuracy.
More time complexity.

3. PROPOSED APPROACH

The first step is to identify and calculate spammer behavioral features in an unlabeled Amazon review dataset. This calculation is carried out on all dataset reviews based spam review detection using behavioral features method.A new proposed framework consists in representing a set of reviews data provided as HIN (Heterogeneous Information Network) and solving the issue of spam detection in a problem of HIN classification. In particular, to show their views dataset as a HIN where the reviews are linked through different types of nodes (such as functionality and users). Then a weighting algorithm is used to calculate the importance (or weight) of each function. These weights are used to calculate the latest review labels using supervised and unsupervised procedures.

4. ARCHITECTURE

The Fig.1 shows the proposed system architecture:

1. Spam framework that is a novel network based approach which models review networks as heterogeneous information networks.

2. A new weighting method for spam features is proposed to determine the relative importance of each feature and shows how effective each of features are in identifying spams from normal reviews.

3. Spam framework improves the accuracy against the state-of-the art in points of time complexity, which extremely depends to the number of features utilized to detect a spam review.

   Our suggested framework's fundamental notion is to describe a given review dataset as a Heterogeneous Information Network (HIN) and transfer the challenge of spam detection into a FIN classification task. In particular, a model review dataset in which reviews are linked together using various node kinds.
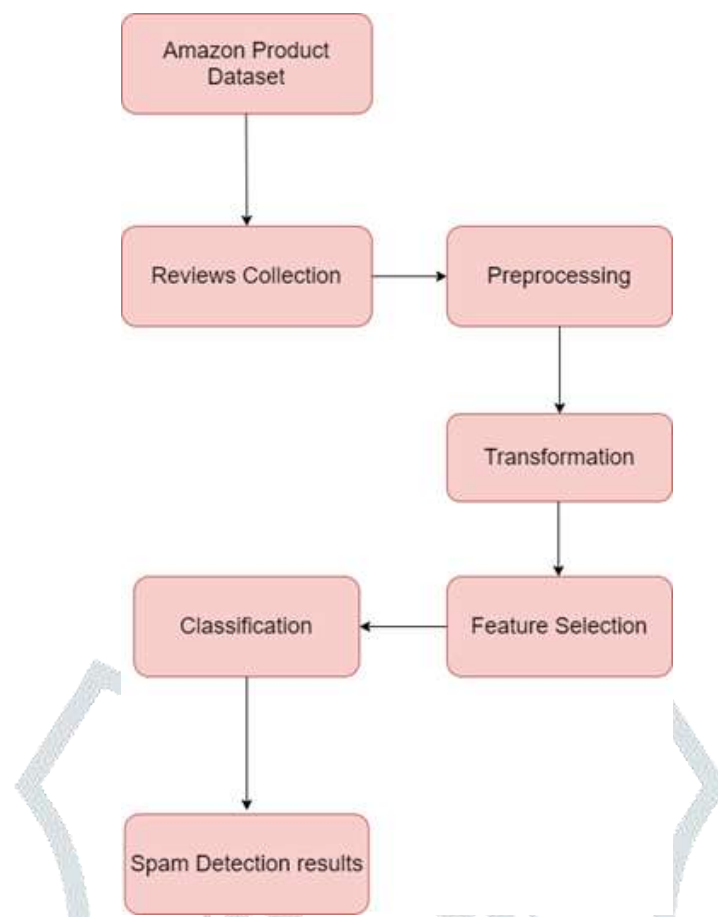
FIG. 1: PROPOSED SYSTEM ARCHITECTURE

ADVANTAGES OF PROPOSED SYSTEM

1. It identifies spam and spammers as well as different type of analysis on this topic.
2. Written reviews also help service providers to enhance the quality of their products and services.
3. It identifies the spam user using positive and negative reviews in online social media.
4. This framework displays only trusted reviews to the users.

5.   CONCLUSIONS

This investigation presents a completely unique fake reviews detection system especially spam detection in sight of a metapath idea and another graph based strategy to call reviews counting on a rank- based naming methodology.This paper has used SA and LSA with netspam algorithm for spam detection. Our perceptions appear that computed weights by utilizing this metapath idea can be exceptionally successful in distinguishing spam reviews and prompt superior performance. Four fundamental classifications for highlights our observations demonstrate that the review behavioral classification performs superior to anything different classification.

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 bustiness 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.