IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

Characterizing and Predicting Early Reviewers for Effective Product Marketing on E-Commerce Websites

Karthik Raja M[1] and V. Sumalatha[2]
PG Student, Department of Computer Applications[1]
Associate Professor, Department of Computer Applications[2]
Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, India
karthikmanohar2001@gmail.com and sumalathav.research@gmail.com

Abstract: *Online reviews have become an important source of information for users before making an informed purchase decision. Early reviews of a product tend to have a high impact on the subsequent product sales. In this project, we take the initiative to study the behaviour characteristics of early reviewers through their posted reviews on two real-world large e-commerce platforms, i.e., Amazon and Yelp. In specific, we divide product lifetime into three consecutive stages, namely early, majority and laggards. A user who has posted a review in the early stage is considered as an early reviewer. We quantitatively characterize early reviewers based on their rating behaviours, the helpfulness scores received from others and the correlation of their reviews with product popularity. We have found that (1) an early reviewer tends to assign a higher average rating score; and (2) an early reviewer tends to post more helpful reviews. Our analysis of product reviews also indicates that early reviewers' ratings and their received helpfulness scores are likely to influence product popularity. By viewing review posting process as a multiplayer competition game, we propose a novel margin-based embedding model for early reviewer prediction. Extensive experiments on two different e-commerce datasets have shown that our proposed approach outperforms a number of competitive baselines. In our project we have used algorithms like Decision Tree (DT) and Multi-Layer Perceptron (MLP). All are measured in terms of accuracy*

Keywords: E-commerce Websites, User Behaviour Analysis, Margin-based Embedding Model, Automated Spammers, Dataset Analysis, Detection of Suspicious Activities

I. INTRODUCTION

Web enable users to keep in touch with friends, relatives, family members, and people with similar interests, profession, and objectives. In addition, they allow users to interact with one another and form communities. A user can become a member of an Twitter by registering and providing details, such as name, birthday, gender, and other contact information. Although a large number of web exist on the web, Facebook and Twitter are among the most popular webs are included in the list of the top 10 websites[1] around the worldwide. Twitter allows the users to follow their favourite politicians, athletes, celebrities, and news channels, and to subscribe to their content without any hindrance. Through following activity, a follower can receive status updates of subscribed account. Although Twitter and other OSNs are mainly used for various benign purposes, their open nature, huge user base, and real-time message proliferation have made them lucrative targets for cyber criminals and socialbots.

II. PROBLEM STATEMENT

The main contributions of this study can be summarized as follows:
A novel study that uses community-based features with other feature categories, including metadata, content, and interaction, for detecting automated spammers. A detailed analysis of the working behaviour of automated spammers and benign users with respect to newly defined features. In addition, two-tailed Z-test statistical significance analysis is performed to answer the following question: "is the difference between the working behaviour of spammers and benign

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

users in terms of newly defined features a random chance?" A thorough analysis of the discriminating power of each feature category in segregating automated spammers from benign users.

III. PROPOSED SYSTEM

Multilayer Perceptron (MLP)

A **multilayer perceptron** (**MLP**) is a class of feed forward artificial neural network (ANN). The term MLP is used ambiguously, sometimes loosely to mean *any* feed forward ANN, sometimes strictly to refer to networks composed of multiple layers of perceptrons (with threshold activation); see § Terminology. Multilayer perceptrons are sometimes colloquially referred to as "vanilla" neural networks, especially when they have a single hidden layer.

An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.

If a multilayer perceptron has a linear activation function in all neurons, that is, a linear function that maps the weighted inputs to the output of each neuron, then linear algebra shows that any number of layers can be reduced to a two-layer input-output model. In MLPs some neurons use a *nonlinear* activation function that was developed to model the frequency of action potentials, or firing, of biological neurons.

IV. IMPLEMENTATION

System Description

Implementation is the process that actually yields the lowest-level system elements in the system hierarchy (system breakdown structure). The system elements are made, bought, or reused. Production involves the hardware fabrication processes of forming, removing, joining, and finishing; or the software realization processes of coding and testing; or the operational procedures development processes for operators' roles. If implementation involves a production process, a manufacturing system which uses the established technical and management processes may be required.

The purpose of the implementation process is to design and create (or fabricate) a system element conforming to that element's design properties and/or requirements. The element is constructed employing appropriate technologies and industry practices. This process bridges the system definition processes and the integration process.

System Implementation is the stage in the project where the theoretical design is turned into a working system. The most critical stage is achieving a successful system and in giving confidence on the new system for the user that it will work efficiently and effectively. The existing system was long time process.

Module Description

There are 5 components in the system. They are

1. Input dataset
2. Analysis of size of data set.
3. Oversampling.
4. Training and Testing.
5. Apply algorithms.
6. Predict results.

1. Input dataset:

Dataset can be taken from online data source provider from the internet sources. We have to collect a huge dataset in volume so as to predict the accuracy in an efficient manner.

2. Analysis of data set:

Here the analysis if dataset takes place. The size of data is taken into consideration for the data process.

3. Oversampling (Using SMOTE): we have created a detailed history of all reviews dataset that is been happened over a long duration.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

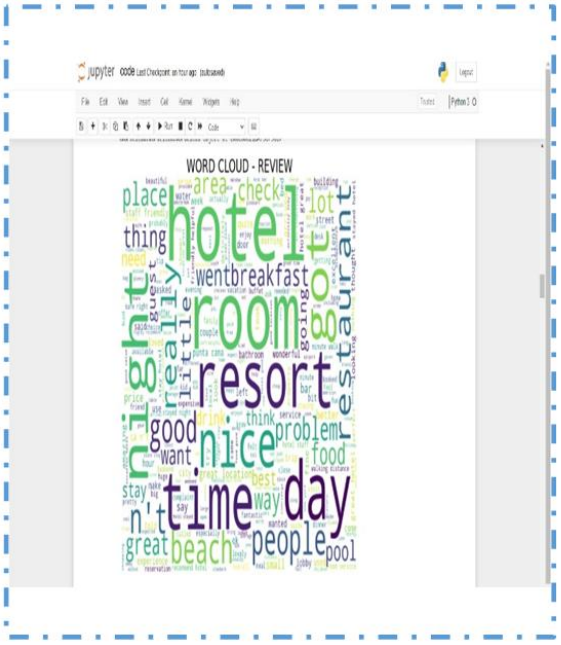
4. Training and Testing Subset: As the dataset is imbalanced, many classifiers show bias for majority classes. The features of minority class are treated as noise and are ignored. Hence it is proposed to select a sample dataset.

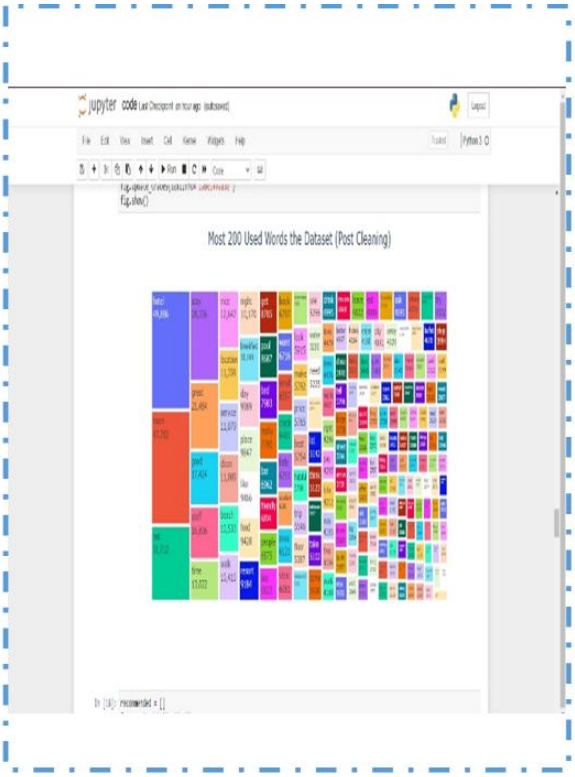
5. Applying algorithm: Following are the classification algorithms used to test the sub-sample dataset.

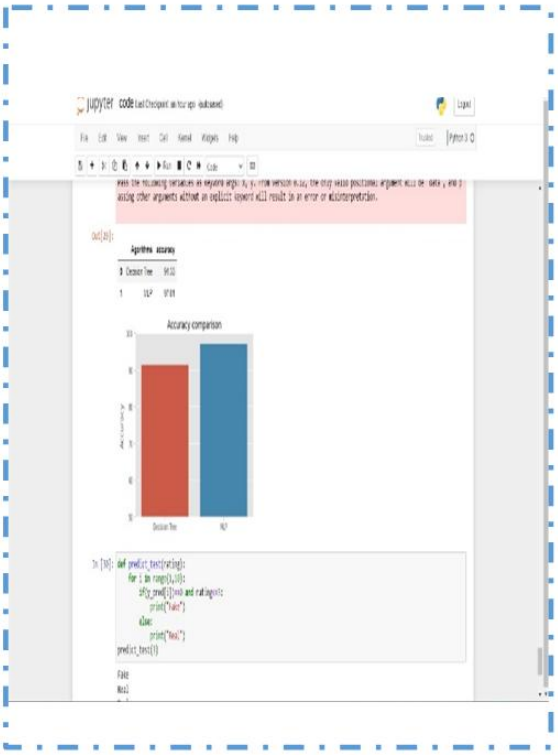
a. Decision Tree (EDT)

b. Multilayer Perceptron (MLP)

Predicting results: The test subset is applied on the trained model .The metrices used is accuracy. The ROC Curve is plotted and the desirable results are achieved.

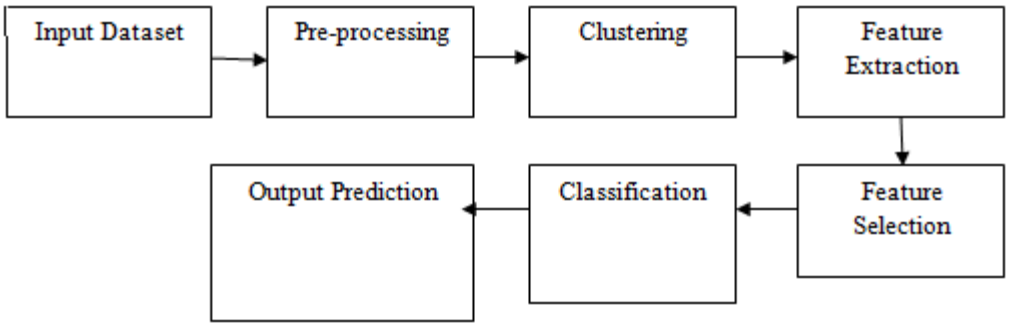






IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

V. SYSTEM ARCHITECTURE



VI. CONCLUSION

Purpose of our research work is to derive the right conclusion about the already rated reviews on various well defined parameters, In this case we may easily analyse the reviews given and can categories into groups and can predict the good or bad reviews usng sentimental classification methods. This is the case of supervised learning. In case of pre-specified parameters are not provided but some pre-defined clusters may be identified and hotel reviews may be fit in some group. The reviews that did not fit any group will be termed as the spam. This is the case of unsupervised learning. It is concluded that appropriate machine learning techniques can be used for analysis of online reviews of the products.

REFERENCES

[1]. C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@spam: The underground on 140 characters or less," in proc. ACM conf. Computer communication security, 2019, pp. 27-37.
[2] Y Boshmaf, I.Muslukhov, K Beznoson, and M. Ripeanu, "design and analysis of social botnet," computer networks, vol. 57, no. 2, pp. 556- 578, 2019.
[3] Arora, Harsha, GovindaMuraliUpadhyay, "A framework for the detection of Suspicious Discussion on online forum using integrated approach of support vector machine and particle Swarm Optimization", international Journal of Advanced research in computer science, 2019.
[4] J. McAuley and A. Yang, ―Addressing complex and subjective product-related queries with customer reviews,‖ in WWW, 2016, pp. 625–635. [2] N. V. Nielsen, ―E-commerce: Evolution or revolution in the fastmoving consumer goods world,‖ nngroup. com, 2014.
[5]. W. D. J. Salganik M J, Dodds P S, ―Experimental study of inequality and unpredictability in an artificial cultural market,‖ in ASONAM, 2016, pp. 529–532.
[6]. R. Peres, E. Muller, and V. Mahajan, ―Innovation diffusion and new product growth models: A critical review and research directions,‖ International Journal of Research in Marketing, vol. 27, no. 2, pp. 91 – 106, 2010.
[7]. L. A. Fourt and J. W. Woodlock, ―Early prediction of market success for new grocery products.‖ Journal of Marketing, vol. 25, no. 2, pp. 31 –38, 1960. [6] B. W. O, ―Reference group influence on product and brand purchase decisions,‖ Journal of Consumer Research, vol. 9, pp. 183–194, 1982.
[8] J. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, ―Imagebased recommendations on styles and substitutes,‖ in SIGIR, 2015, pp. 43–52.
[9]. D. Imamori and K. Tajima, ―Predicting popularity of twitter accounts through the discovery of link-propagating early adopters,‖ in CoRR, 2015, p. 1512.
[10]. X. Rong and Q. Mei, ―Diffusion of innovations revisited: from social network to innovation network,‖ in CIKM, 2013, pp. 499– 508.