Characterizing And Predicting Reviews For Effective Product Marketing And Advancement

Pushpendra Tiwari[1], Chandan Mani Tripathi[2]
[1]*M. Tech Scholar, Dept. of CSE, S R Institute of Management & Technology, (AKTU), Lucknow, India*
[2] *Assistant Professors, Dept. of CSE, S R Institute of Management & Technology, (AKTU), Lucknow, India*

***Abstract*— In today's competitive market, understanding customer feedback is essential for refining product strategies and driving business success. This study focuses on characterizing and predicting reviews to aid in effective product marketing and advancement. By utilizing natural language processing (NLP) techniques, machine learning algorithms, and sentiment analysis, the research identifies key patterns and factors that influence consumer opinions. The study analyzes a diverse range of product reviews to predict future customer sentiments, providing valuable insights for targeted marketing efforts, product improvements, and customer engagement strategies. The findings offer actionable recommendations for businesses looking to enhance their marketing efforts, improve customer satisfaction, and streamline their product development processes.**

Keywords: Product Marketing, Customer Reviews, Sentiment Analysis, Natural Language Processing, Machine Learning, Predictive Analytics, Customer Feedback, Product Advancement, Marketing Strategy, Consumer Behavior.

1. INTRODUCTION

The development of web based business sites has empowered clients to distribute or share buy encounters by posting item audits, which for the most part contain helpful conclusions, remarks and criticism towards an item. In that capacity, a lion's share of clients will peruse online surveys prior to settling on an educated buy choice . It has been accounted for about 71% of worldwide online customers read online surveys prior to buying an item . Item surveys, particularly the early audits (i.e., the surveys posted in the beginning phase of an item), profoundly affect ensuing item deals . We call the clients who posted the early surveys early commentators. Albeit early analysts contribute just a little extent of audits, their sentiments can decide the achievement or disappointment of new items and administrations. It is significant for organizations to distinguish early commentators since their criticisms can assist organizations with changing promoting techniques and improve item plans, which can in the long run lead to the achievement of their new items. Thus, early commentators become the accentuation to screen and draw in at the early advancement phase of an organization. The essential job of early audits has drawn in broad consideration from promoting experts to actuate shopper buy expectations . For instance, Amazon, one of the biggest online business organization on the planet, has supported the Early Reviewer Program1, which assists with getting early surveys on items that have not many or no audits. With this program, Amazon customers can study items and settle on more brilliant purchasing choices. As another connected program, Amazon Vine2 welcomes the most confided in analysts on Amazon to post suppositions about new and prerelease things to help their kindred clients settle on educated buy choices.

Past examinations have profoundly stressed the wonder that people are emphatically impacted by the choices of others, which can be clarified by group conduct. The impact of early audits on ensuing buy can be perceived as a unique instance of crowding impact. Early audits contain significant item assessments from past adopters, which are important reference assets for resulting buy choices. As demonstrated in, when buyers utilize the item assessments of others to gauge item quality on the Internet, crowd conduct happens in the web based shopping measure. Not the same as existing examinations on crowd conduct, we center around quantitatively investigating the general qualities of early commentators utilizing huge scope genuine world datasets. Moreover, we formalize the early analyst expectation task as a contest issue and propose a novel installing based positioning way to deal with this undertaking. As far as anyone is concerned, the assignment of early commentator forecast itself has gotten almost no consideration in the writing. Our commitments are summed up as follows:

We present a first report to portray early analysts on a web based business site utilizing two certifiable huge datasets. We quantitatively dissect the attributes of early analysts and their effect on item prevalence. Our exact examination offers help to a progression of hypothetical ends from the social science and financial aspects. We see survey posting measure as a multiplayer contest game and foster an installing based positioning model for the forecast of early analysts. Our model can manage the chilly beginning issue by consolidating side data of items. Broad examinations on two true enormous datasets, i.e., Amazon and Yelp have exhibited the viability of our methodology for the forecast of early commentators.

In this paper section I contains the introduction, section II contains the literature review details, section III contains the details about methodologies, section IV shows architecture details, V describe the result and section VII provide conclusion of this paper.

2. LITERATURE REVIEW

A developmental shift from disconnected business sectors to advanced business sectors has expanded the reliance of clients on online audits generally. Online surveys have become a stage for building trust and affecting purchaser purchasing behaviours. With such reliance there is a need to deal with such enormous volume of surveys and present believable audits before the customer. In this, actually progressed decade, the meaning of dynamic of market system relies exceptionally upon the investigation of advertising studies and item surveys. Hence, in this writing we have attempted to think about and get familiar with the absolute best models of assessment investigation. Greatest analysts have attempted to discover the general investigation of the surveys yet scarcely anybody utilized that examination for item advertising and improvement.

Table 1: Different technique used for predicting reviews

YEAR	AUTHOR	PURPOSE	TECHNIQUE
2019	N. Sultana, P. kumar	Sentiment analysis for product review	NB,SVM & Linear model algorithm
2019	Alpna Patel & Arvind Kumar	Sentiment analysis by using RNN	Used RNN
2018	Yoon-Joo Park	Online review helpfulness across different product type	Linear regression, SVM, Random Forest,M5P
2018	S. Saumya, J. Prakash	Ranking Online customer reviews	Cosine similarity, SVM and Random Forest
2017	Liao Shiyang, Junbo Wang.	SA of twitter data	NLP and CNN for classification
2015	D. Imamori & K Tajima	Predicting popularity of twitter account.	Cosine similarity and SVM
2015	Xing Fang & Justin Zhan	Sentiment analysis using product review data	NB, SVM, Random Forest

3. METHODOLOGIES

- **Preprocessing**

In this algorithm, the tweets which are foreign made to database from the twitter API, these tweets comprise of pointless words, whitespaces, hyperlinks and unique characters. First we have to do separating process by evacuating every single superfluous word, whitespaces, hyperlinks and extraordinary characters.

The preprocessing steps aim to begin the feature extraction process and start extracting bags of words from the samples. One of the main focus is to reduce the final amount of features extracted. Indeed, features reduction is important in order to improve the accuracy of the prediction for both topic modeling and sentiment analysis. Features are used to represent the samples, and the more the algorithm will be trained for a specific feature the more accurate the results will be. Hence, if two features are similar it is convenient to combine them as one unique feature. Moreover, if a feature is not relevant for the analysis, it can be removed from the bag of words.

- Lower uppercase letters: The first step in the preprocessing is to go through all the data and change every uppercase letter to their corresponding lowercase letter. When processing a word, the analysis will be case sensitive and the program will consider “data” and “Data” as two totally different words. It is important that, these two words are considered as the same features. Otherwise, the algorithms will affect sentiments which may differ to these two words. For example, on these three sentences: “data are good”, “Awesome data”, and “Bad Data”. The first and second sentences both contain “data” and are positive, the third sentence contains “Data” and is negative. The algorithm will guess that sentences containing “data” are more likely to be positive and those containing “Data” negative. If the uppercases had been removed the algorithm would have been able to guess that the fact that the sentence contains “data” is not very relevant to detect whether or the sentence is positive. This preprocessing step is even more important since the data are retrieved from Twitter. Social media users are often writing in uppercase even if it is not required, thus this preprocessing step will have a better impact on social media data than other “classical” data.
- Remove URLs and user references: Twitter allows user to include hashtags, user references and URLs in their messages. In most cases, user references and URLs are not relevant for analyzing the content of a text. Therefore, this preprocessing step relies on regular expression to find and replace every URLs by “URL” and user reference by “AT_USER”, this allows to reduce the total amount of features extracted from the corpus [2]. The hashtags are not removed since they often contain a word which is relevant for the analysis, and the “#” characters will be removed during the tokenization process. • Remove digits: Digits are not relevant for analyzing the data, so they can be removed from the sentences. Furthermore, in some cases digits will be mixed with words, removing them may allow to associate two features which may have been considered different by the algorithm otherwise. For example, some data may contain “iphone”, when other will contain “iphone7”. The tokenization process, which will be introduced later.
- Remove stop words: In natural language processing, stop words are often removed from the sample. These stop words are words which are commonly used in a language, and are not relevant for several natural language processing methods such as topic modeling and sentiment analysis [10]. Removing these words allows to reduce the amount of features extracted from the samples.
- **Self-Learning and word standardization System**

In this algorithm, first we have to instate the word reference (first emphasis dictionary).In the lexicon for the most part we have to introduce the positive, negative nonpartisan and things. Every single huge datum and information mining ventures in view of the prepared information, without prepared information (introduction of words).So instatement of the prepared information is vital. In the self-learning framework, we are doing word institutionalization, here we are not considering past, present and future status of the words, just we are thinking about the word.

- **Sentiment Analysis**

In this calculation, pre-processed tweets are brought from the data set individually. In any case we require check individually watchword whether that expression is thing are not, if thing we will oust it from the particular audit. After that the remainder of the watchwords checked with evaluation create, whether or not those expressions are sure assessment or adverse end or unbiased inclination. The remainder of the watchwords in the tweet which doesn't has a spot with any of the assumption will be consigned fleeting end considering the more check of positive, negative and fair. In the subsequent cycle if the reaming word crosses the restriction of positive, negative or impartial, that watchword everlastingly included as improvement in the vocabulary.

Algorithm Step in Sentiment Analysis

Step 1: Get some sentiment examples
As for every supervised learning problem, the algorithm needs to be trained from labeled examples in order to generalize to new data.

Step 2: Extract features from examples
Transform each example into a feature vector. The simplest way to do it is to have a vector where each dimension represents the frequency of a given word in the document.

Step3: Train the parameters
This is where your model will learn from the data. There are multiple ways of using features to generate an output, but one of the simplest algorithms is logistic regression. Other well-known algorithms are Naive Bayes. In the simplest form, each feature will be associated with a weight. Let's say the word "love" has a weight equal to +4, "hate" is -10, "the" is 0 ... For a given example, the weights corresponding to the features will be summed, and it will be considered "positive" if the total is > 0, "negative" otherwise. Our model will then try to find the optimal set of weights to maximize the number of examples in our data that are predicted correctly.
If you have more than 2 output classes, for example if you want to classify between "positive", "neutral" and "negative", each feature will have as many weights as there are classes, and the class with the highest weighted feature sum wins.

Step 4: Test the model
After we have trained the parameters to fit the training data, we have to make sure our model generalizes to new data, because it's really easy to over fit. The general way of regularizing the model is to prevent parameters from having extreme values.

- **Naive Bayes**
- Bayes classifiers are a group of straightforward probabilistic classifiers dependent on applying Bayes' hypothesis with solid (gullible) freedom presumptions between the highlights.

- Naive Bayes classifiers are exceptionally adaptable, requiring various boundaries direct in the quantity of factors (highlights/indicators) in a learning issue. Greatest probability preparing should be possible by assessing a shut structure articulation, which takes direct time, as opposed to by costly iterative guess as utilized for some different sorts of classifiers.

- In the insights and software engineering writing, credulous Bayes models are known under an assortment of names, including basic Bayes and autonomy Bayes. Every one of these names reference the utilization of Bayes' hypothesis in the classifier's choice guideline, yet innocent Bayes isn't (really) a Bayesian technique

- Naive Bayes is a straightforward method for building classifiers: models that appoint class marks to issue cases, addressed as vectors of highlight esteems, where the class names are drawn from some limited set. It's anything but a solitary calculation for preparing such classifiers, yet a group of calculations dependent on a typical guideline: all gullible Bayes classifiers expect that the worth of a specific element is free of the worth of some other component, given the class variable. For instance, an organic product might be viewed as an apple on the off chance that it is red, round, and around 10 cm in breadth. An innocent Bayes classifier thinks about every one of these highlights to contribute autonomously to the likelihood that this organic product is an apple, paying little heed to any potential connections between's the shading, roundness, and breadth highlights.

- For a few sorts of likelihood models, credulous Bayes classifiers can be prepared productively in a managed getting the hang of setting. In numerous down to earth applications, boundary assessment for innocent Bayes models utilizes the strategy for most extreme probability; all in all, one can work with the credulous Bayes model without tolerating Bayesian likelihood or utilizing any Bayesian strategies.

- Despite their guileless plan and obviously distorted suspicions, credulous Bayes classifiers have functioned admirably in numerous intricate genuine circumstances. In 2004, an examination of the Bayesian order issue showed that there are sound hypothetical purposes behind the clearly doubtful viability of credulous Bayes classifiers. In any case, a complete examination with other order calculations in 2006 showed that Bayes arrangement is outflanked by different methodologies, like supported trees or irregular woods.

4. SYSTEM ARCHITECTURE

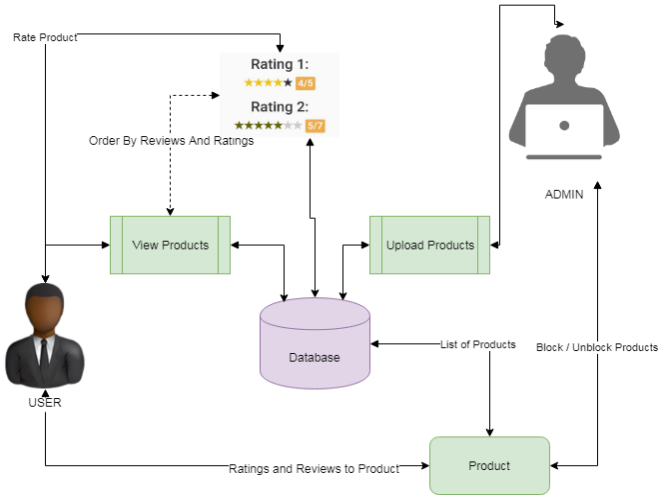

Figure 1 System Architecture

5. RESULTS

In this outcome part, we step up and study the conduct qualities of posted audits on delegate online business stages. We plan to lead powerful investigation and make precise forecast towards item improvement. With the blasting of internet business, individuals are becoming accustomed to burning-through on the web and composing remarks about their buy encounters on vendor/audit Websites. These stubborn substance are important assets both to future clients for dynamic and to shippers for improving their items as well as administration.

These are the modules implemented in this research paper result part:

- **UPLOAD PRODUCTS**

Uploading the products is done by admin. Authorized person is uploading the new arrivals to system that are listed to users. Product can be uploaded with its attributes such as brand, color, and all other details of warranty. The uploaded products are able to block or unblock by users.

- **PRODUCT REVIEW BASED ORDER**

The suggestion to user's view of products is listed based on the review by user and rating to particular item. Naïve bayes algorithm is used in this project to develop the whether the sentiment of given review is positive or negative. Based on the output of algorithm suggestion to users is given. The algorithm is applied and lists the products in user side based on the positive and negative.

- **RATINGS AND REVIEWS**

Ratings and reviews are main concept of the project in order to find effective product marketing. The main aim of the project is to get the user reviews based on how they purchased or whether they purchased or not. The major find out of the project is when they give the ratings and how effective it is. And this will helpful for the users who are willing to buy the same kind of product.

- **DATA ANALYSIS**

The main part of the project is to analysis the ratings and reviews that are given by the user. The products can be analysis based on the numbers which are given by user. The user data analysis of the data can be done by charts format. The graphs may vary like pie chart, bar chart or some other charts.

USER NAME	MOBILE NUMBER	USER EMAIL ID	PRODUCT NAME	VENDOR NAME	PRODUCT RATING	PRODUCT
vinay	9685741230	gokul@gmail.com	shirt	Xiomi Peter England	3	It is very goo
vinay	9685741230	gokul@gmail.com	shirt	Xiomi Peter England	2	it is g
siva	9685741230	gokul@gmail.com	shirt	Van Heusen	2	goo
vinay	9685741230	gokul@gmail.com	T-shirt	Parx	4	It is w
vinay	9685741230	gokul@gmail.com	T-shirt	Parx	1	poor product i h
vinay	9685741230	gokul@gmail.com	shirt	Van Heusen	5	worst product purcha
vinay	9685741230	gokul@gmail.com	shirt	Van Heusen	3	not g
sabari	9685741230	gokul@gmail.com	cotten pant	LEVIS	3	poor pur
vinay	9685741230	gokul@gmail.com	shirt	Xiomi Peter England	2	great pur
vinay	9685741230	gokul@gmail.com	mobile accessories	Huwai	5	supe
vinay	9685741230	gokul@gmail.com	mobile accessories	Huwai	2	It is very goo

Figure 2: Product reviews

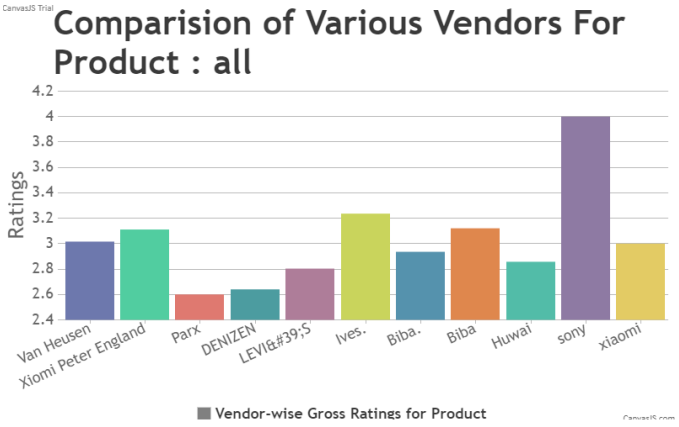

Figure 3: Comparison of various vendors for product

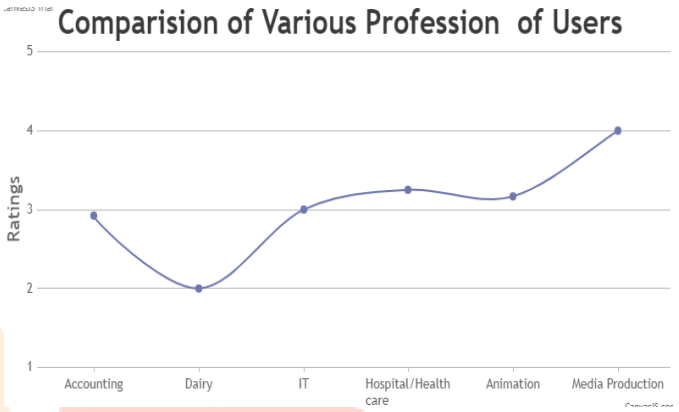

Figure 4: Comparison of various profession of users

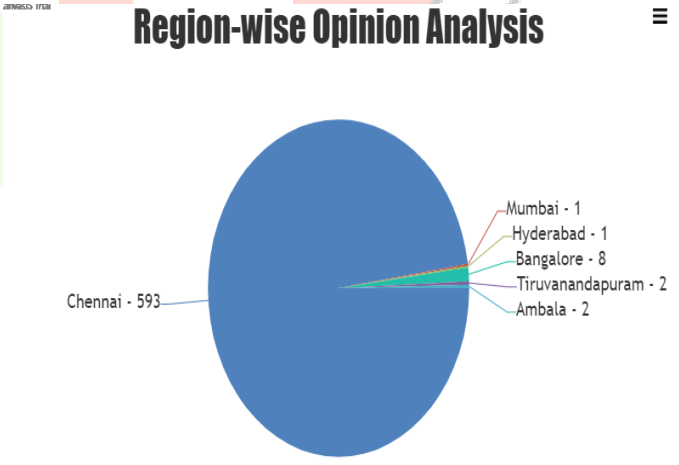

Figure 5: Region-wise opinion analysis

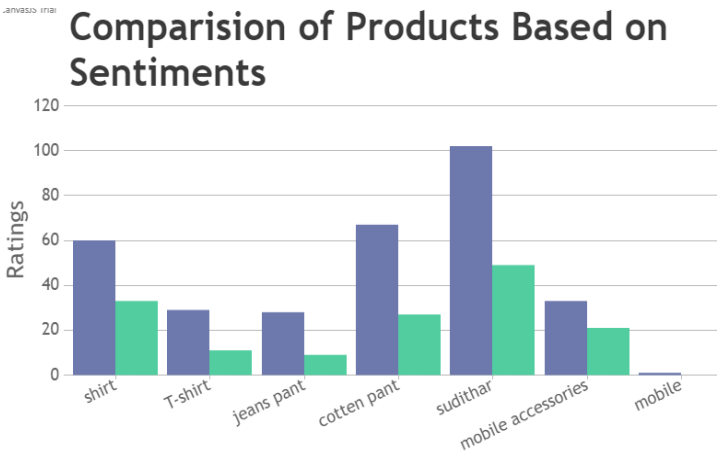

Figure 6: Comparison of product based on sentiments

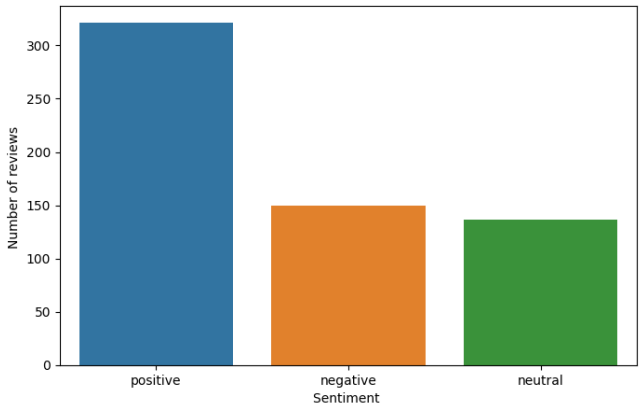

Figure 7: Merm analysis

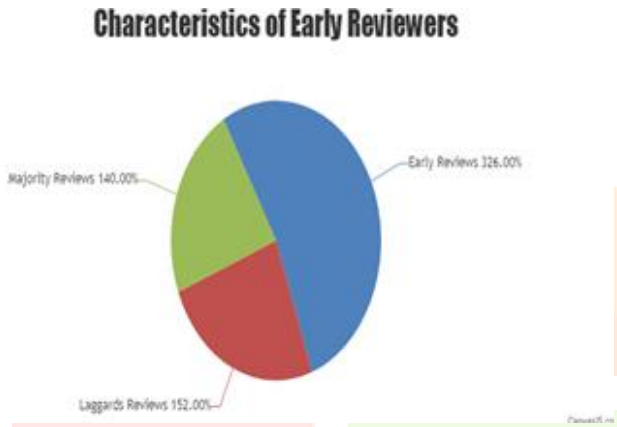

Figure 8: shows the characteristics of early reviews

6. CONCLUSION

In this research paper, we have considered the novel concept of early reviewer portrayal and expectation on two certifiable online audit datasets. Our exact examination reinforces a progression of hypothetical ends from social science and financial aspects. We tracked down that (1) an early analyst will in general appoint a higher normal rating score; and (2) an early reviewer will in general post more accommodating audits. Our tests likewise show that early reviewers' evaluations and their got supportiveness scores are probably going to impact item ubiquity at a later stage. We have embraced a rivalry based perspective to display the survey posting measure, and fostered an edge based implanting positioning model (MERM) for foreseeing early reviewers in a beginning setting.

REFERENCE

[1] Najma Sultana , Pintu Kumar , Monika Rani Patra , Sourabh Chandra And S.K. Safikul Alam" Sentiment Analysis For Product Review" ICTACT Journal On Soft Computing, April 2019, Volume: 09, Issue: 03

[2] Alpna P. , Arvind K. T. , “Sentiment Analysis by using Recurrent Neural Network” proceedings of 2nd ICACSE, 2019

[3] Yoon-Joo Park "Envisioning the Helpfulness of Online Customer Reviews across Different Product Types" MDPI, Sustainability 2018, 10, 1735

[4] Sunil Saumya, Jyoti Prakash Singh, Nripendra P. Rana,Yogesh k. Dwivedi, Swansea University Bay Campus, Swansea "Situating Online Consumer Reviews" Article in Electronic Commerce Research and Applications, March 2018

[5] Liao, Shiyang, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng, “CNN for situations understanding based on sentiment analysis of twitter data”, Procedia Computer Science, vol. 111, pp.376-381, 2017.

[6] Xing Fang* and Justin Zhan "Sentiment Analysis using product review data” Journal of Big Data , 2015

[7] D. Imamori and K. Tajima, "Foreseeing notoriety of twitter accounts through the revelation of link-propagating early adopters," in CoRR, 2015, p. 1512.

[8] Yeole V., P.V. Chavan, and M.C. Nikose, “Opinion mining for emotions determination”, ICIIECS 2015-2015 IEEE Int. Conf. Information, Embed. Commun. Syst., 2015.

[9] F. Luo, C. Li, and Z. Cao, Affective-feature-based sentiment analysis using SVM classifier, 2016 IEEE 20th Int. Conf. Comput. Support. Coop. Work Des., pp.276281, 2016.

[10] Kalaivani A., Thenmozhi D, Sentiment Analysis using Deep Learning Techniques. , IJRTE,2018

[11] Bingwei Liu, Erik Blasch, Yu Chen, Dan Shen, and Genshe Chen. Scalable sentiment classification for big data analysis using naïve bayes classifier. In big data, 2013 IEEE International Conference on, pages 99-104. IEEE,2013

[12] D. N. Devi, C. K. Kumar, and S. Prasad, “A feature based approach for sentiment analysis by using support vector machine,” in Advanced Computing (IACC), 2016 IEEE 6thInterna-tional Conference on. IEEE, 2016, pp. 3–8.

[13] McAuley, J.; Leskovec, J. Hidden factors and hidden topics: Understanding rating dimensions with review text. In Proceedings of the 7th ACM Conference on Recommender Systems, RecSys’, Hong Kong, China,12–16 October 2013; pp. 165–172.

[14] Ghose, A.; Ipeirotis, P.G. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. IEEE Trans. Knowl. Data Eng. 2011, 23, 1498–1512.

.