

---

Open Access Article

## AN CONTEMPLATED APPROACH FOR SENTIMENTAL ANALYSIS WITH COMBINED TECHNIQUE IN SOCIAL APPLICATIONS

**Ravleen Singh**

Research Scholar, Department of Computer Science & Engineering, Madhav University, Abu Road,  
Tehsil - Pindwara Distt. Sirohi, Rajasthan, India. E-mail: ravleen1234@gmail.com

**Dr. Ganpat joshi**

Associate Professor, Department of Computer Science & Engineering, Madhav University, Abu  
Road, Tehsil - Pindwara Distt. Sirohi, Rajasthan, India. E-mail: shiv.joshi322@gmail.com

**Dr. Tariq Hussain Sheikh**

Lecturer, Department of Computer Science and Application. Govt. Degree College Poonch,  
Jammu & Kashmir, India E-mail: tariqsheakh2000@gmail.com

**Abstract**—In Recent years, the Digital world dominating the human lives in all the circumstances of their day to day life. One of the most prominent technology backs the social networks is the Natural Language Processing. In this Morphological domain the Competent Sentimental analysis is of high claim in many applications' platform for accurate and predictive classification. The research is concentrating to study of public opinion for product reviews to get the valuable information. The analysis of sentiment on social networks, such as Twitter or Facebook, has become a powerful means of learning about the users' opinions and has a wide range of applications. Recently, it has been demonstrated that Machine learning models area promising solution to the challenges of NLP. This Paper understands the above mentioned emotions in software side as well as it is essential to identify with the producers and the director of business. The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0

**Keywords**— sentiment analysis, Machine Learning, Opinion Mining, Sentiment classification, feature selection, NLP.

**摘要**——近年来, 数字世界在日常生活的所有环境中主导着人类的生活。支持社交网络的最突出的技术之一是自然语言处理。在这个形态学领域, Competent Sentimental 分析在许多应用程序平台中具有很高的要求, 以实现准确和预测性的分类。该研究集中于研究产品评论的舆情, 以获取有价值的信息。在 Twitter 或 Facebook 等社交网络上进行情绪分析已成为了解用户意见的有力手段, 应用范围广泛。最近, 已经证明机器学习模型是应对 NLP 挑战的有希望的解决方案。本文理解了上述软件方面的情绪, 也很需要认同制作方和业务总监。F1 分数

Received: September 14, 2021 / Revised: October 09, 2021 / Accepted: October 27, 2021 / Published: November 03, 2021

About the authors : Ravleen Singh

Corresponding author- Email: ravleen1234@gmail.com

可以解释为准确率和召回率的加权平均值，其中 F1 分数在 1 时达到最佳值，在 0 时达到最差分数

关键词——情感分析、机器学习、意见挖掘、情感分类、特征选择、NLP。

Introduction

Sentimental Analysis is an essential social media observing metric that investigates social and web collaborations rapidly and adequately. SA alludes to a method dependent on computational and NLP, applied to recognize, extract, or describe subjective data, for example, sentiments, communicated in a given text. Here NLP is referred as the Neural Network Processing. The principle goal of SA is to sort the creator's demeanor towards different themes into positive, negative or neutral classes. SA has numerous applications including, yet not restricted to, commercial knowledge, sociology, politics, and so forth. Including different fields [1]. Late years, then again, have seen the appearance of long range social networking sites, micro blogs, wikis and web applications and thusly. An uncommon development in client produced information is set for opinion mining. There are two type of methods available in the Sentimental Analysis which is below :

1. Lexicon based method: It is the first method of the sentiments analysis. It is the combination of the negative and positive opinion words. These negative and positive opinion words is used for the create opinion sentences. These sentences are the combination of the negative, neutral or positive words. It is the most useable method. This method need a scoring function to score every sentence. This score is depended on the positive and negative words. This method is used the lexicon words. It is the collection of the two opinion words

which is the negative and positive. It is also combined with the scoring functions. This functions is used for the determine the polarity sent. This method is also called as an untrained teaching method. It is not need any type of the training data. It is based only on the dictionary.

2. Machine Learning method: It is the second method of the sentiments analysis. It is the used for the training a classifier through a collection of predetermined opinions as training set. After completed this step then, use trained classifier to group novel ideas as negative, neutral or positive. Pak, A. et. al used one method which is known as the supervised method. This method is used for the construct a classifier through section of the speech tagger and n-gram approaches. The classifier is used for the classify opinions . It is need any type of the training data. It is not based on the dictionary [2][3].

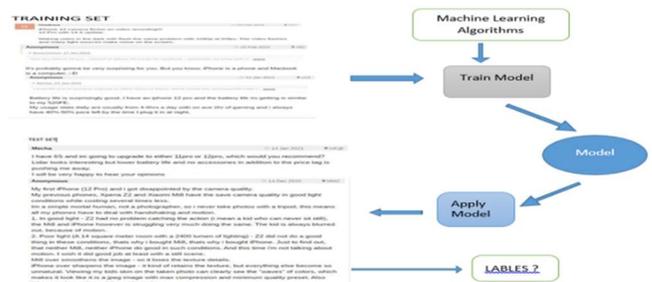


Fig 1. Machine learning Model.

SA is the methodology of determining and computing the emotional state of attitude, tone, responses and opinions. More Certainty, This is the belief of deciding whether a describe communication is negative, positive, or neutral.

SA has varied classifies and applications polarity, subjectivity, emotion mining, tonality, confusion analysis, opinion mining, and efficient computing. This is the assets which permit businesses to get knowledge about what their clients are review about their items and services. To estimate their client attitudes and opinions towards items and services in relation to their competitors. Monitoring too. Concerning the use of SA, it is reported that this has been finished in the context of marketing and business, politics and public activity. There are some example is available which is world events, voting application, and E-Commerce. The vast majority of the information extracted for the examination was extracted from social media. The social media contain a massive measure of information from online clients and we can get any data on an services, item, events or place which suitable it suitable for SA study[4].

There are some types of the online communication services available which is below :

### 1. **Blogs:**

It is the first communication services. It is the used for describe the information in reverse chronological order. It is the personal diary. This is the public commentary. It is the tool of the organize about thought. It is also used for the publish latest news. It may be communication tool to give information about the clients or staffs.

Business blogs is used for external interaction. There is some example available for business blogs which is media relations, crisis communications, and business development [5]. One purpose behind beginning a business blog might be marking, particularly for administration arranged organizations. How would you show quality, dependability and incentive in your

support of our client? Utilizing a blog is one approach to build familiarity with organization and increment traffic to site.

An vital factor of such analysis is to mark the sentiment expressed about specific brands and prod in the blog. SA focuses on this work of directly identifying whether a piece of text expresses a positive or negative opinion about the subject matter. Most past works around there utilize the lexical information on the previous to allude to the feeling of the words. Interestingly, some ongoing methodologies consider the errand as a text grouping issue, where they just figure out how to order feeling dependent on mark preparing information.

### 2. **Forums:**

It is the second services of online communication. A Forum is a compulsory classification or gathering for subjects. Subjects, which are made inside discussions, are the place members can post discussions. A Conversation is the string of messages wherein members post their commitments. It has become a huge collection of exceptionally significant commentaries and opinions. Feature extraction incorporates text assessment investigation that A whole number worth must be determined for each bit of text. this is a Semantic direction based methodology where supposition esteems All watchwords are added to get the feeling an incentive for Full article. String answers are decayed into a set Catchphrase. A statement is given for every watchword. Returns the total of notion esteems for all watchwords Estimation esteem for string. Assume for string T, the appropriate response is The keys are deteriorated into a lot of words [5][6].

### 3. **Online Communities:**

It is the third services of online communication. It is the community which forms on the internet.

It is the collection of people communication. It is also used for sharing the views about the items and working toward a normal aim. There is some example available of online communities which are the Facebook, twitter and Google+. it is also used the email group and forms.

### **Background**

Sentiment analysis is the process of extracting subjective information from natural language text, and it also expresses the opinion or view of a user toward a topic. Sentiment means feelings, and it includes attitudes, emotions, and opinions, and sentiment is a subjective impression, not a fact. Sentiment analysis can be done by applying natural language processing, statistical methods, or machine learning techniques to identify, extract, or characterize the sentiment content of a text unit. The insights and applications from sentiment analysis were useful in other areas, like politics, law, policy making, sociology, and psychology. In general, humans are subjective creatures, and opinions are important, and being able to interact with people on that level has many advantages for the information system [7]. With the speedy addition in the amount of web customers, the Internet has a huge proportion of information delivered by the customers. Numerous individuals share their perspectives with respect to a theme via online media stages, for instance, Twitter and Facebook and give their criticism or audit about an item on web based business sites, for example, Amazon and Flipkart which prompts an enormous measure of information. The ID of emotional articulations from the information is known as subjectivity location. To robotize the investigation of such information, opinion examination is utilized. The point is to locate the opinionative information and order it as per its extremity, for example

positive, negative or nonpartisan input, known as assumption characterization and afterward dissecting it which is known as slant investigation. In any case, prior to performing slant assessment, the data is presented to various pre-handling methodology which at long last give the ideal upgraded yield. This permits us to become more acquainted with about the public's mind-set or assessment on a specific point. This researchs helps the concerned association or public to improve their item or administration dependent on the feedback received[8][9].

### **Polarity :**

Emotions are firmly identified with assessments. The quality of a opinion or sentiment is regularly connected to the intensity of specific feelings, e.g., anger and anger. In this term the SA is recognized the sentiment orientation. There are three types of the sentiment orientation which is the neutral, negative and positive. Different sorts of SA incorporate fine-grained SA which gives more exactness in the degree of polarity (for example positive, very positive, negative, very negative and neutral)[22]. The goal of the emotion analysis is the recognized emotions in expression for instance sadness, surprise, happiness and frustration and so on.

In web-based media analysis for any area, SA is one of the generally utilized applications. The critical part of SA is to examine a collection of text for understanding the specific situation and feeling demonstrated by it. As PCs just comprehend mathematical information, we can evaluate the sentiment towards the positive and negative estimation by utilizing Polarity.

Opinions or Sentiments from web-based media give the most cutting-edge and comprehensive data, because of the multiplication of online media and the low hindrance for posting the

message. Regardless of the developing significance of estimation examination, this territory does not have a brief and precise plan of earlier endeavors [10]

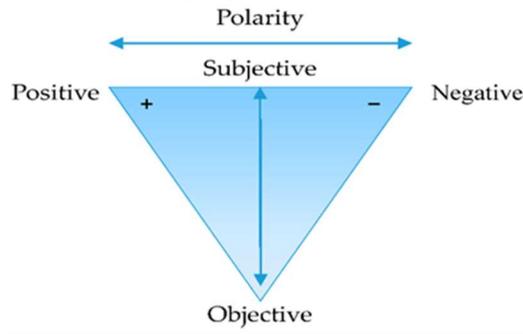


Fig 2. sentiment orientation

A few fundamental angles, accordingly, are tended to inside the extent of this review. From one viewpoint, this paper centers around introducing commonplace strategies from three alternate points of view (task-arranged, granularity-situated, approach situated) in the territory of notion investigation.

**Problem Formulation And Proposed Solution Problem Statement**

The sentiment analysis is the approach to analyze sentiments of the input data. This research work is related to sentiment analysis of product reviews. The sentiment analysis steps have three steps which are pre-processing, feature extraction and classification. In the pre-processing phases, the data set of the product review which is collected from the twitter which processed to remove missing and redundant values. In the second phase, N-gram algorithm will be applied which will extract features of the dataset. In the last phase, classification algorithm which takes input as training and test set. It is analyzed that classification techniques naïve bayes which is proposed in the previous research work has low accuracy. In this research work,

novel classification method needs to propose for the sentiment analysis of product reviews.

**Decision trees** – the skilled information space will be symbolized in a hierarchical framework because to detachment the data, featured charge has been used.

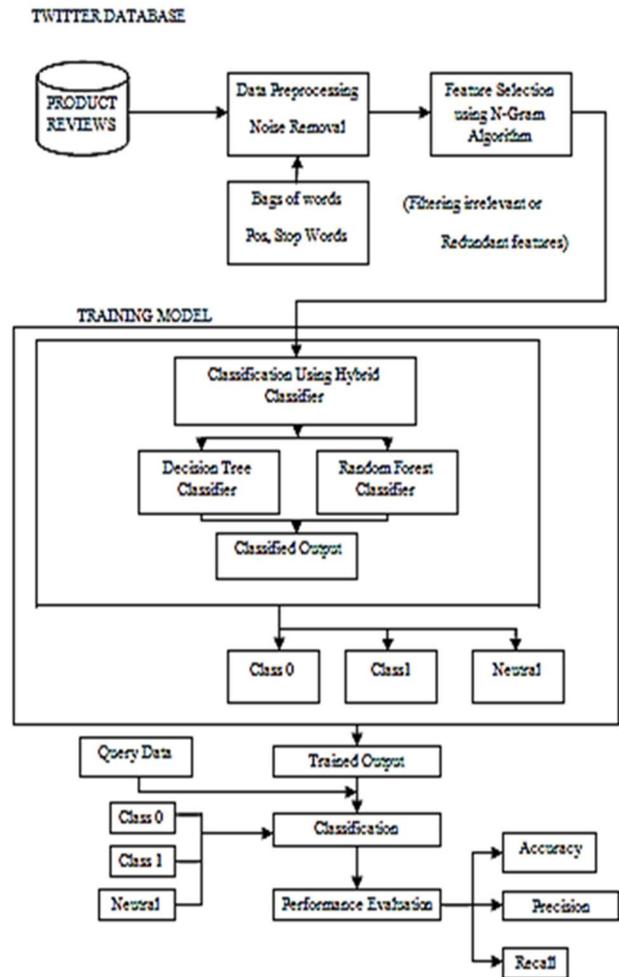


Fig 3. Block Diagram of the proposed approach

Block Diagram of the proposed approach

**Input:**

Data partition, D, which is a set of training tuples and their associated class labels.

Attribute\_list, the set of candidate attributes.

Attribute\_selection\_method, a procedure to determine the splitting criterion that “best” partitions the data tuples into individual classes.

**Output:** A decision tree.

### Decision Tree Rules

There are a number of advantages in converting a decision tree to rules. Decision tree make it easier to make pruning decisions. Since it is easier to see the context of each rule. Also, converting to rules removes the distinction between attribute tests that occur near the root of the tree and those that occur near the leaves. These rules are easier to read and to understand for people. The basic rules for decision tree are as below.

1. Each path from the root to the leaf of the decision tree therefore consists of attribute tests, finally reaching a leaf that describes the class.
2. If-then rules may be derived based on the various paths from the root to the leaf nodes.
3. Rules can often be combined to produce a smaller set of rules. For example:
4. If result = “distinction %” then credit rating = excellent
5. If stream = “arts” and result = “70 %” then credit rating = average.
6. Once all the rules have been generated, it may be possible to simplify the rules.
7. Rules with only one antecedent (e.g. if result = “distinction”) cannot be further simplified. So we only consider those with two or more antecedents.
8. Eliminate unnecessary rule antecedents that have no effect on the conclusion reached by the rule.
9. In some cases, a number of rules that lead to the same class may be combined.

We construct a decision tree based on a set of training examples  $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ , where  $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$  are the feature vectors and  $y_1, \dots, y_n \in \{1, \dots, c\}$  are the labels. Every internal node in the tree possesses two ordered child nodes and a decision rule of the form  $\mathbf{x}(i) < a$ , where  $\mathbf{x}(i)$  is the  $i$ th attribute and  $a$  is a real number. Feature vectors that satisfy the decision rule are directed to the node’s left child node, and the other vectors are directed to the right child node. Thus, every example  $\mathbf{x}$  has a path from the root to one of the leaves, denoted  $l(\mathbf{x})$ . Every leaf has a label  $t$ , so that an example  $\mathbf{x}$  is assigned the label  $t(l(\mathbf{x}))$ . Each time that line 3 is executed, we say that a new iteration has begun. If there are too many samples (possibly infinite in number), we read a predefined number of samples; otherwise, we use the complete data set. A new level of nodes is appended to the tree in each iteration. In line 5 we decide whether a leaf  $v$  is to be split or labeled, according to a stopping criterion. Possible stopping criteria can be some threshold on the number of samples reaching the node, or on the node’s impurity. A node’s impurity is a function  $G$  that measures the homogeneity of labels in samples reaching the node. The above algorithm provides an overview of the tree construction algorithm.

The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal. The formula for the F1 score is:

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$F_1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

Where

TP – true positive

TN – true negative

FP – false positive and

FN – false negative

### Simulation And Results

Table 1 shows the some of the observation of Product Reviews Dataset; the outcome of the classification has been estimated from the Product reviews datasets instances, then classifying the instances with the same observation, then the performance measures of various techniques of Naives Bayes, Linear, SVM and is compared with the proposed techniques Pro\_Hybrid. The Table-1 shows the Comparison of performance in F1-score. It has been analyzed from the Actual and predicted value from the objective of three classes class 0 , class 1 and Neutral in confusion matrix and it is represented in percentage.

Techniques	F1-score (%)		
	Class 0	Class 1	Neutral
<b>Naives Bayes</b>	40	28	94.8
<b>Linear</b>	42	30	95.4
<b>SVM</b>	43	32	96.3
<b>Pro_Hybrid</b>	45	32	97

Table 1: F1 – Score Calculation

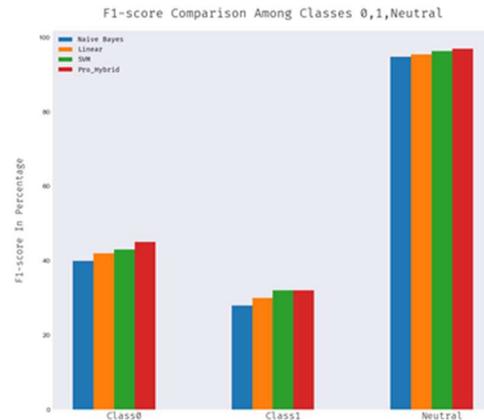


Figure 4 – Graphical Representation For F1 - Score Calculation For Classes 0, 1 And Neutral

The above figure is graphical representation for table 1. It shows the comparison for various methods in terms of F1-score .It is a F1-score comparison for Product review dataset between existing and proposed techniques. As shown in the above figure the proposed Hybrid achieves F1-score with maximum percentage in both the classes 0, 1 and Neutral than existing techniques

### Conclusion And Future Work

This Paper understands the above mentioned emotions in software side as well as it is essential to identify with the producers and the director of business. The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal. F1 Score is needed when you want to seek a balance between Precision and Recall. Right. We have previously seen that accuracy can be largely contributed by a large number of True Negatives which in most business circumstances, we do not focus on much whereas False Negative and False Positive usually has business costs (tangible & intangible) thus F1 Score might be a better measure to use if we need to seek a balance

between Precision and Recall AND there is an uneven class distribution (large number of Actual Negatives).

## References

- Doaa Mohey El-Din Mohamed Hussein, April 2016, A survey on sentiment analysis challenges, Journal of King Saud University – Engineering Sciences (2018) 30, 330–338
- Ling, P., Geng, C., Menghou, Z., Chunya, L., 2014. What Do Seller Manipulations of Online Product Reviews Mean to Consumers? (HKIBS Working Paper Series 070-1314) Hong Kong Institute of Business Studies, Lingnan University, Hong Kong.
- B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? sentiment classification using machine learning techniques, in: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2002.
- Walaa Medhat, Ahmed Hassan, Hoda Korashy, “Sentiment analysis algorithms and applications: A survey”, Ain Shams Engineering Journal (2014) 5, 1093–1113
- D. MALI, M. ABHYANKAR, P. BHAVARTHI, K. GAIDHAR, M. BANGARE, “SENTIMENT ANALYSIS OF PRODUCT REVIEWS FOR ECOMMERCE RECOMMENDATION”, International Journal of Management and Applied Science, Volume-2, Issue-1, Jan.-2016
- Nisha Jebaseeli, E. Kirubakaran, “A Survey on Sentiment Analysis of (Product) Reviews”, International Journal of Computer Applications (0975 – 888), Volume 47– No.11, June 2012
- Milan Gaonkar, Prof. Amit Patil, “Sentiment Classification Using Product Reviews”, 2017, IOSR Journal of Computer Engineering (IOSR-JCE)
- SupriyaSoni, Shivendra Dubey, Rakesh Tiwari, Mukesh Dixit, “FEATURE BASED SENTIMENT ANALYSIS OF PRODUCT REVIEWS USING DEEP LEARNING METHODS”, International Journal of Advanced Technology & Engineering Research (IJATER), Volume 8, Issue 3, May 2018
- Melville, Wojciech Gryc, “Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification”, KDD’09, June 28–July 1, 2009, Paris, France. Copyright 2009 ACM 978-1-60558-495-9/09/06.
- Qiang Ye, Ziqiong Zhang, Rob Law, “Sentiment classification of online reviews to travel destinations by supervised machine learning approaches”, Expert Systems with Applications 36 (2009) 6527–6535.
- Gang Li, Fei Liu, “A Clustering-based Approach on Sentiment Analysis”, 2010, 978-1-4244-6793-8/10 ©2010 IEEE.
- Khairullah Khan, Baharum B. Baharudin, Aurangzeb Khan, and Fazal\_e\_Malik, “Automatic Extraction of Features and Opinion Oriented Sentences from Customer Reviews”, World Academy of Science, Engineering and Technology 62 2010
- Ting-Chun Peng and Chia-Chun Shih, “An Unsupervised Snippet-based Sentiment Classification Method for Chinese Unknown Phrases without using

- Reference Word Pairs”, 2010 IEEE/WIC/ACM International Conference on Web Intelligence and intelligent Agent Technology JOURNAL OF COMPUTING, VOLUME 2, ISSUE 8, AUGUST 2010, ISSN 2151-9617 .
- Yongyong Zhail, Yanxiang Chenl, Xuegang Hu, “Extracting Opinion Features in Sentiment Patterns” , International Conference on Information, Networking and Automation (ICINA),2010.
  - ZHU Jian , XU Chen, WANG Han-shi, “” Sentiment classification using the theory of ANNs”, The Journal of China Universities of Posts and Telecommunications, July 2010, 17(Suppl.): 58–62 .
  - Long-Sheng Chen , Cheng-Hsiang Liu, Hui-Ju Chiu , “A neural network based approach for sentiment classification in the blogosphere”, Journal of Informetrics 5 (2011) 313–322.
  - Rui Xia , Chengqing Zong, Shoushan Li, “Ensemble of feature sets and classification algorithms for sentiment classification”, Information Sciences 181 (2011) 1138–1152.
  - P. Saloun, M. Hruzk and I. Zelinka, “Sentiment Analysis e-Bussines an e-Learning Common Issue,” ICETA 2013 ,11th IEEE International Conference on Emerging eLearning Technologies and Applications, Stary Smokovec, The High Tatras, Slovakia, October 24-25, 2013.
  - Hemnaath, R., and Low, B. W. 2010. Sentiment Analysis Using Maximum Entropy and Support Vector Machine. Semantic Technology and Knowledge Engineering in 2010. Kuching,Sarawak
  - Hu Minging, Liu Bing. Mining and summarizing customer reviews. In: Proceedings of ACM SIGKDD international conference on Knowledge Discovery and Data Mining (KDD’04); 2004.
  - Miller G, Beckwith R, Fellbaum C, Gross D, Miller K. WordNet: an on-line lexical database. Oxford Univ. Press; 1990.
  - H. Yu and V. Hatzivassiloglou. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. Proc. 2003 Conf. Empirical Methods in Natural Language Processing (Emnlp 03), ACM Press, 2003
  - RaheesaSafrin, K.R.Sharmila, T.S.ShriSubangi, E.A.Vimal, “SENTIMENT ANALYSIS ON ONLINE PRODUCT REVIEW”, International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 04 | Apr -2017