

SENTIMENT MINING OF CUSTOMER REVIEWS FROM E-COMMERCE WEBSITES

¹KUNDETI NAGA PRASANTHI, ²MVP CHANDRASEKHARARAO, ³SUDHASREE CHEKURI, ⁴SESHU BABU P

¹Professor, Department of CSE, Lakireddy Balireddy College of Engineering, Mylavaram, India.

^{2,3}Department of CSBS, RVR & JC College of Engineering, Chowdavaram, India.

⁴Professor, Department of Mathematics and Statistics, Kakaraparti Bhavanarayana College, Vijayawada, India

Email : ¹prasanthi.kundeti@gmail.com, ²manukondach@gmail.com, ³sudha.chekuri@gmail.com, ⁴psb.seshu@gmail.com

ABSTRACT

Over the past few decades e-commerce has increased manifolds. The e-commerce websites ask their customers to share their views about the products they have purchased. Therefore, millions of reviews are accumulated for the products in e-commerce websites. Customers view the reviews of the product before they purchase the product. If a product has more positive reviews then that will result in more customers buying the product. So, classifying the vast collections of reviews into different categories is the need of the hour. This paper describes one of such mechanisms for classifying reviews using text mining and extracting the sentiment of the review. The proposed mechanism involves extracting product reviews from e-commerce websites, identifying the terms that represent the sentiment of products, highlighting the positive terms that are more frequent and specifying the frequencies of different sentiment defining terms in the product reviews.

Keywords: - Classification, Amazon Reviews, Sentiment Analysis, Wordnet, Positive Word Cloud

1. INTRODUCTION

Covid-19 has changed the way customers purchase products. Earlier very few people used to purchase products using online sales applications. People are choosing online platforms to purchase products as social distancing is the need of the hour. These online sales applications or e-commerce websites request the customers to give their opinions about the products they have purchased, the services of the online sales site etc. These customers' reviews will give an overview for the new buyers about the product and the services offered. In this way a huge collection of customer reviews are accumulated in every online sales application for each product. Before purchasing a product using online sales platform, customers wish to see the product reviews.

As the collection of reviews for a product increase, there will be different opinions for same product by different customers. This leads to ambiguity for customer. There will be some positive reviews as well as negative reviews for the same product from different customers. So, there is a requirement to analyse these reviews and

provide overall customer opinion for a given product.

To perform sentiment analysis and classification, subjective information needs to be extracted from given review text in natural language. This subjective information can be opinions and sentiments. To extract opinions or sentiments from given customer review text, various approaches like natural language processing, text analysis, computational linguistics and biometrics are available. Different Machine learning methods can be used for this semantic and review analysis as they are very efficient and simple to implement.

Various e-commerce sites like Amazon, Flipkart, Snapdeal etc. are available for online purchasing for customers. Amazon is one of the e-commerce giants where thousands of purchases are performed every day. Customers give opinions about the product like its properties, quality, appearance, and recommendations. These recommendations will give an insight into almost all the features of the product for the new buyers. These are not only helpful for consumers but also for sellers to improve their services. Manufacturers of products can also understand

the requirements of the customers better and can make modifications to their products accordingly.

This paper discusses about the extraction of opinions or sentiments from the customer reviews and analysing them using machine learning algorithms.

Sentiment analysis techniques can be broadly categorized as (i) Lexicon-based, (ii) Machine learning based and (iii) Deep learning based.

In lexicon-based sentiment analysis, sentiment scores are assigned to words based on predefined sentiment lexicons or dictionaries. The overall sentiment of a document is then calculated based on the sentiment scores of its constituent words.

In machine-learning based sentiment analysis, supervised or unsupervised machine learning algorithms are used to train models on labelled data (where sentiments are annotated) or to cluster texts based on their sentiment patterns.

In deep-learning based sentiment analysis, deep learning architectures like Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), or Transformer models are applied to learn complex patterns in text data for sentiment analysis.

2. BACKGROUND

e-commerce has become very popular as it will allow the customers to leave reviews on different products. It is difficult for manufacturers to keep track of the opinions of the customers as they leave a voluminous collection for every product. So, it is important to process such large and complex data in order to derive useful information from it. To tackle this problem classification methods which are part of machine learning can be used. Classification is one of data mining techniques which divides the data into different categories based on their characteristics (Pandey et al. 2016; Rain 2013). Organizations want to automate this classification process while extracting data from large data sets (Liu et. al 2014).

Opinion mining or sentiment analysis is a natural language processing (NLP) problem which detects and mines subjective information of text sources. The main aim of sentiment analysis is to analyse the customer reviews written and classify them as positive or negative. When classification

is done as positive review or negative review then there is no need for the system to understand each phrase or document (Liu 2015; Pang et. al 2002; Turney & Littman 2003). Labelling the words as positive or negative is not sufficient. This process involves some challenges. For example, the word "excellent" has a positive polarity. But if this word is preceded by another word "not" then it gives opposite polarity like negative polarity (Singla et. al 2013). So, classification of words or phrases with prior polarity has some drawbacks.

Sentiment analysis is carried out in various fields like movie reviews, product reviews and travel reviews etc. (Liu et. al 2013; Pang et al. 2002; Ye et al. 2009). Methods based on lexical analysis and machine learning are most widely used methods for sentiment classification.

Sentiment Classification using Machine Learning Methods:

An algorithm is developed by machine learning methods to improve the performance of the system by learning from example data. There are two steps in the solution provided by machine learning for sentiment analysis. They are (i) learn or train the model from training data and (ii) classify unseen data using trained model (Khairnar & Kimikar 2013). Machine learning algorithms can be divided into different categories like

- (a) Supervised learning
- (b) Semi-supervised learning
- (c) Unsupervised learning

In Supervised Learning, a model is built using training data. It is similar to a teacher supervising the learning process of its students (Brownlee 2016). It trains the model to come up with some kind of output. The training data contains the class label. The class labels of output are also known. If more labelled data is provided as input, then output will be more precise. If output is deviating from expected result, then the model is built again using more labelled data. One of the limitations of supervised learning are it gives wrong result or unknown result if input data is not labelled. In unsupervised learning the input data is unlabelled with no corresponding output. The algorithm discovers similar patterns from data and provides output by grouping similar items into one category. Clustering is one of the unsupervised learning methods, which works efficiently. Clustering identifies similar groups of data in the data set (Kaushik 2016).

In semi-supervised learning, the benefits of both supervised and unsupervised learning are incorporated. Data sets that contain only some labelled data and remaining unlabelled data are generally applied with semi-supervised learning models. This method is generally used where data collection is cheaper and data labelling is costlier.

Yogeesh and Antoreep(2020) performed sentiment analysis on Twitter data using machine learning and deep learning models.

R.R. Kalangi et. al (2021) proposed machine learning based techniques for sentiment analysis of Airlines reviews.

S.Sindhu and S. Kumar(2023) presented survey of different machine learning and deep learning techniques used in sentiment analysis.

3.SYSTEM DESIGN AND IMPLEMENTATION

The process involved in sentiment mining of product reviews can be given as:

1. Data collection:- Gather product reviews from various sources such as e-commerce websites, social media platforms, forums or dedicated review websites.
2. Preprocessing:- Clean the text data by removing irrelevant information such as

HTML tags, punctuation and stop words. Tokenization, Stemming can also be applied to normalize text.

3. Sentiment Analysis Techniques: Apply one of sentiment analysis techniques.
4. Sentiment Classification: Classify each review into predefined sentiment categories such as positive, negative, or neutral.
5. Evaluation: Assess the performance of the sentiment analysis model using metrics such as accuracy, precision, recall, F1-score, or confusion matrix.
6. Post processing and visualization: Analyze the results, identify trends, and visualize the sentiment distribution using techniques like word clouds, bar charts, or sentiment heatmaps.

The following figure gives the basic workflow followed for opinion mining of customer reviews in e-commerce websites. Customer reviews from various e-commerce sites like amazon, flipkart etc. can be extracted using the API provided in their sites. Fig.1 shows the block diagram of system for sentiment mining.

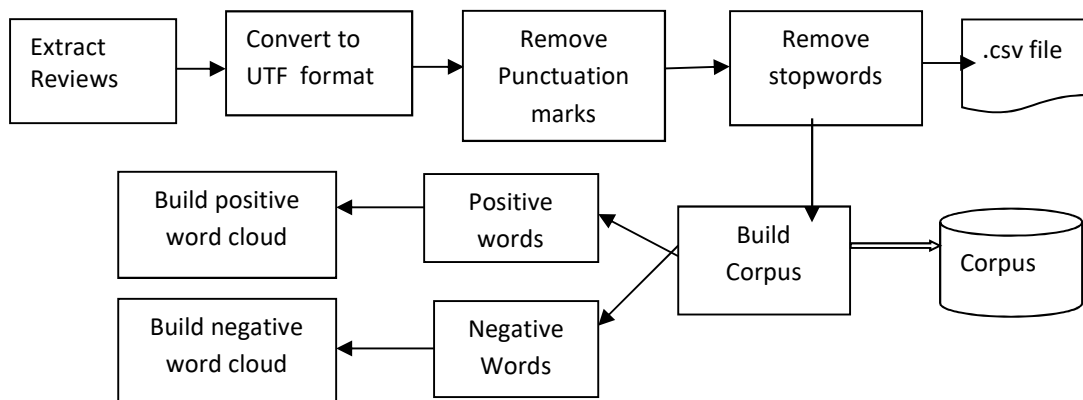


Fig 1. Block Diagram for Sentiment Mining

Input:- URL of customer reviews of the e-commerce website
 Output:- Summary of Sentiments extracted from customer reviews.

Steps

1. Crawl the ecommerce website url for the given product to extract the customer reviews. The customer reviews' text and rating for the product are extracted and stored in a frame.
2. Convert extracted text to UTF format and remove content present in other formats.
3. Remove punctuation marks and white spaces from extracted text.
4. Remove English stop words like a, an, the, this, that etc.
5. Build term document matrix.
6. Build a corpus of all the words in the reviews.
7. Compare the corpus with positive word list and identify all positive words of the product reviews.
8. Compare the corpus with negative word list and identify all negative words of the product reviews.
9. Build positive word cloud and negative word cloud.
10. A summary of sentiments is generated as a chart.

Fig 2. Algorithm for sentiment mining

4. RESULTS

As part of Implementation, amazon website is used for product "HP DeskJet Color Printer". The amazon website customer reviews page is accessed for extracting the reviews for product

"HP DeskJet Color Printer". First 5000 customer reviews are extracted and sentiment mining is performed. The algorithm for sentiment mining is shown in fig.2. The algorithm is implemented using code written in R programming Language.



Fig 3: Positive Word Cloud For The Reviews Of HP Deskjet Color Printer

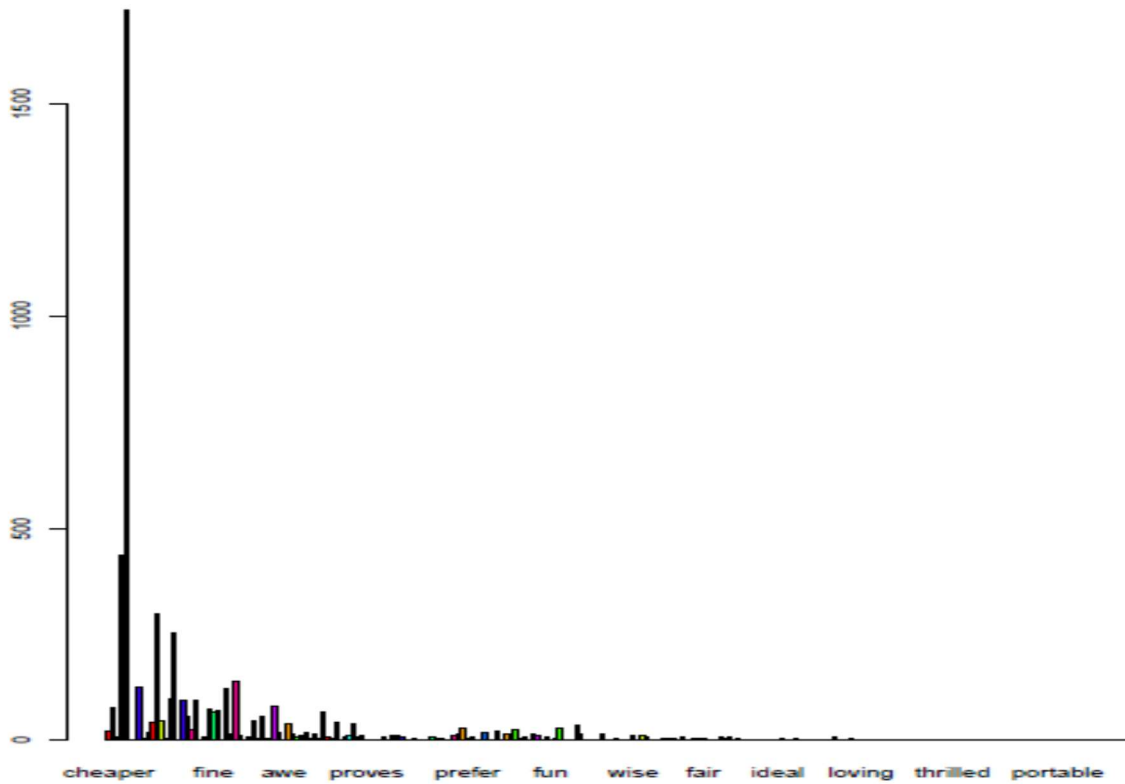


Fig 5: Frequency of Positive Words in Reviews

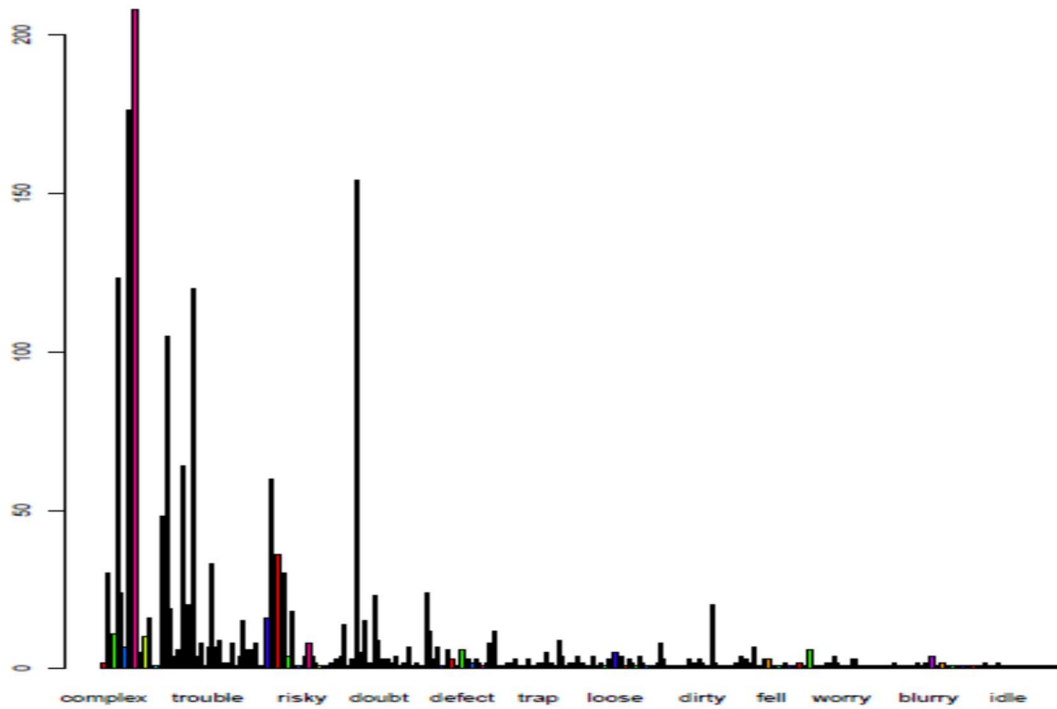


Fig 6: Frequency of Negative Words in Reviews

5.CONCLUSION

E-commerce sites have become very popular during this covid-19 pandemic era. Customers purchase products using e-commerce websites more. Customers wish to see feedback or reviews from previous customers of the product before making a purchase. The sentiment analysis system developed here helps customers to come to a conclusion about the customer reviews. The system developed here helps customers in taking decision before purchasing a product. E-commerce websites are faced with fake customer reviews. So, the proposed system can be extended in future to consider only real customer reviews and ignoring fake customer reviews while identifying overall sentiment of a given product. The current sentiment analysis techniques are unable to identify sarcasm present in customer reviews. Future research can focus on this area.

REFERENCES

- [1] Bo Pang, Lillian Lee and Shivkumar Vaithyanathan, "Sentiment Classification using Machine Learning Techniques", *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, vol- 10, pp:79-86, 2002.
- [2] Peter D Turney and Michael L Littman, "Measuring praise and criticism: Inference of semantic orientation from association". *ACM Transactions on Information Systems* ,vol-21(4), pp:315-346, 2003.
- [3] Qiang Ye, Ziqiong Zhang and Rob Law," Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches", *Expert Systems with Applications*, vol:36(3), pp:6527-6535, 2009.
- [4] Bing Liu, " Sentiment Analysis and Opinion Mining", *Synthesis Lectures on Human Language Technologies*, vol:5(1), pp: 1-167, 2012.
- [5] Callen Rain, " Sentiment analysis in amazon reviews using probabilistic machine learning", Swarthmore College.2013.
- [6] Jayashri Khairnar and Mayura Kinikar, "Machine Learning Algorithms for opinion mining and sentiment classification", *International Journal of Scientific and Research Publications*, vol:3(6), pp:1-6, 2013.
- [7] Bingwei Liu, Erik Blasch, Yu Chen, Dan shen and Genshe Chen, "Scalable sentiment classification for Big data analysis using naïve bayes classifier", *IEEE International Conference on Big Data*, pp:99-104, 2013.
- [8] Bing Liu, "Sentiment Analysis: Mining Opinions, sentiments and emotions", *Cambridge University Press*, 2015.
- [9] Priyank Pandey, Manoj Kumar and Prakhar Srivatsava, " Classification Techniques for Big Data: A Survey" , *3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp:3625-3629, IEEE 2016.
- [10] Jason Brownlee, " Supervised and Unsupervised Machine Learning algorithms", March, 2016.
- [11] Saurav Kaushik, " An Introduction to Clustering and Different Methods of Clustering", Nov. 2016.
- [12] Zeenia Singla, Sukhchandan Randhawa and Sushma Jain, "Statistical and sentiment analysis of consumer product reviews", *8th International conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, pp:1-6, 2017.
- [13] Y. Chandra and A. Jana, "Sentiment Analysis using Machine Learning and Deep Learning," *7th international Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, pp. 1-4, 2020.
- [14] R. R. Kalangi, S. Maloji, N. Tejasri, P. P. Chand and V. P. Ch, "Sentiment Analysis using Machine Learning," *2021 3rd International Conference on Advances in Computing, Communication Control and Networking*, Greater Noida, India, pp.116-121, 2021.
- [15] S. Sindhu, S. Kumar and A. Noliya, "A Review on Sentiment Analysis using Machine Learning," *2023 International Conference on Innovative Data Communication Technologies and Application* , Uttarakhand, India, pp. 138-142, 2023.