

# Sentiment Analysis on Product Reviews

## Using Machine Learning Techniques

Achargui Afkir Aymane, Tribak Mohammed, Yahiaoui Anass, Zaid Ilyass

**Abstract** Sentiment Analysis and Opinion Mining have emerged as popular fields for analyzing text data from various sources, including social media platforms like Facebook, Twitter, and e-commerce websites such as Amazon. These techniques are instrumental in enabling businesses to actively improve their strategies and gain in-depth insights into customer feedback regarding their products. By computationally studying an individual's behavior and mining their opinions about a company's business entity, be it an event, individual, blog post, or product experience, valuable insights can be derived. This paper focuses on a dataset obtained from Amazon, comprising reviews of cameras, laptops, mobile phones, tablets, TVs, and video surveillance products. Following data preprocessing, machine learning algorithms were applied to classify the reviews as either positive or negative. The findings of this study highlight the effectiveness of machine learning techniques in accurately classifying product reviews. this study demonstrates that Machine Learning Techniques yield the best results for classifying product reviews. Decision Tree achieved an accuracy of 62.06%, while K-NN achieved an accuracy of 34.48% specifically for laptops reviews. These findings highlight the effectiveness of employing Machine Learning Techniques in accurately categorizing product reviews.

**Keywords** Sentiment analysis - Natural language processing - Product reviews

Machine learning - K-Nearest Neighbors (K-NN) – Decision Tree.

## **1. Introduction :**

Sentiment analysis refers to the scientific study of text analysis, natural language processing, and computational linguistics to identify, extract, and analyze subjective information from textual data. Sentiments or opinions are the attitudes expressed by customers in various forms such as reviews, survey responses, online social media posts, healthcare media, and more. The primary objective of sentiment analysis is to determine the polarity or sentiment orientation of a speaker, writer, or subject regarding a particular topic or context, be it an event, discussion, forum, interaction, or any document.

The fundamental task of sentiment analysis is to determine the polarity of given text at different levels, including the feature, sentence, and document level. With the increasing number of internet users, individuals are inclined to share their opinions through various mediums, leading to the generation of opinionated data on the internet. Sentiment analysis plays a crucial role in analyzing this opinionated data and extracting valuable insights that can aid other users in making informed decisions.

Social media data encompasses various types, such as product reviews, movie reviews, airline reviews, cricket reviews, hotel reviews, employee interactions, healthcare reviews, news articles, and more. By leveraging sentiment analysis techniques, businesses and individuals can gain a deeper understanding of customer sentiments and opinions expressed across these different types of social media data. This analysis can provide valuable insights for decision-making processes and help enhance user experiences across various domains.

## 1.1 Data sources:

In the field of sentiment analysis and opinion mining, various data sources are utilized to gather extensive amounts of data from social media. Here are a few examples:

### 1.1.1 Blogs:

It is the online platforms where individuals or organizations can publish and share their thoughts, ideas, experiences, or information on various topics in a chronological format.

### 1.1.2 Datasets

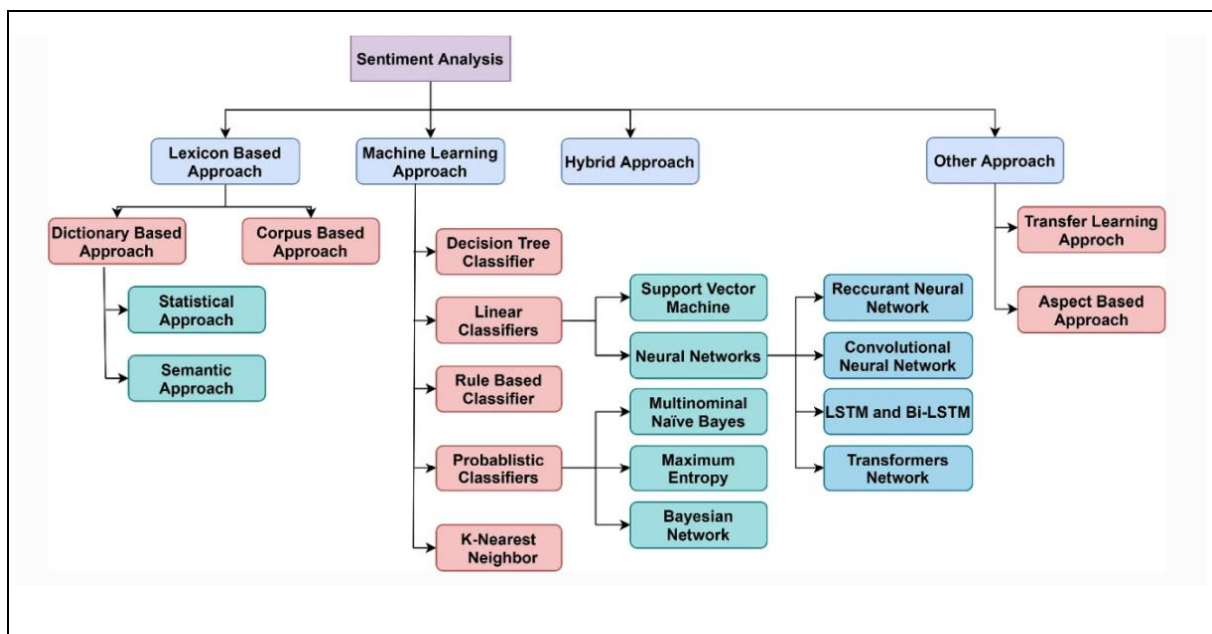
Many researchers have made various datasets available online for free, including movie reviews, product reviews, hotel reviews, and more, allowing easy access to these collections of data

### 1.1.3 Review Sites

Customers frequently express their opinions on e-commerce platforms such as Amazon, CNET, Epinion, ZDNet, ConsumerReview, IMDB, and others, sharing their feedback and experiences regarding the products they have purchased.

### 1.1.4 Micro-blogging:

It is a widely used service for sending short text messages, similar to platforms like Twitter, Tumblr, Dipity, and others. (Fig.1 )



## **2. Related Work**

Sentiment Analysis is a prominent research field within Natural Language Processing (NLP) that focuses on identifying and extracting sentiment or opinion from text, as well as categorizing the expressed sentiment. It involves analyzing people's opinions, evaluations, emotions, and attitudes towards individuals, organizations, products, movies, issues, events, and more.

### ***2.1. Document Level***

In the context of sentiment analysis, document-level analysis involves analyzing and classifying entire documents, such as reviews or articles, as expressing a positive or negative sentiment. This method is suitable for cases where a single product review is being processed to determine the opinion about that specific product. However, it is not applicable when a document contains multiple product reviews, as it is not relevant to the analysis of documents with diverse product reviews.

### ***2.2. Sentence Level***

At this level, the task involves analyzing individual sentences to determine whether they convey a positive, negative, or neutral opinion. This process is comparable to Subjectivity Classification, where sentences are categorized into objective and subjective groups. Objective sentences contain factual information, while subjective sentences express personal opinions or feelings.

### ***2.3.Aspect Level***

Previously known as Feature level sentiment analysis, Aspect level sentiment analysis focuses on feature-based opinion mining and summarization, aiming to identify specific aspects that people like or dislike. This level provides a more detailed sentiment analysis by directly examining opinions themselves, rather than considering entire documents or sentences. The output of this analysis includes the entity, aspect of that entity, opinion of the aspect, opinion holder, and time. For instance, in the sentence "Samsung j7 has the best camera quality," the aspect level analysis identifies "camera" as an aspect of the entity "Samsung j7" expressing a positive opinion. In movies, specific scenes, actors, acting, and actions are key points within aspect level sentiment analysis.

Various methodologies have been compared for sentiment analysis, including machine learning methods such as Support Vector Machines (SVM), Naive Bayes (NB), Maximum Entropy (ME), and N-gram approaches. Aspect classification, polarity detection, and lexical approaches have been discussed. Customer reviews, tweets, and online product reviews have been utilized as datasets for experimentation, achieving accuracies up to 78%. Hybrid approaches combining machine learning and lexicon-based methods have also been explored, incorporating techniques like pSenti, SAIL, NILC\_USP, and MSA-COSRs.

Opinion mining has been surveyed at different levels, analyzing the tools, architectures, techniques, and challenges involved. Sentence level sentiment analysis has been conducted on live tweets using tools like R, along with the utilization of lexicons such as SentiWordNet, WordNet-Affect, and MPQA. Different events, marked by hashtags like #Budget2016, #RailBudget2016, and #InternationalWomensDay, have been collected, with 10,000 tweets per event classified as positive or negative. Machine learning approaches have been employed, addressing issues like spam reviews and unauthenticated users, utilizing datasets from Amazon reviews.

These sentiment analysis studies have aimed to understand public perception of products over time and identify areas for improvement based on people's opinions.

### **3. Proposed Method**

In the proposed method, the following preprocessing tasks were performed to classify sentiment analysis from reviews.

#### ***3.1. Collection of Dataset***

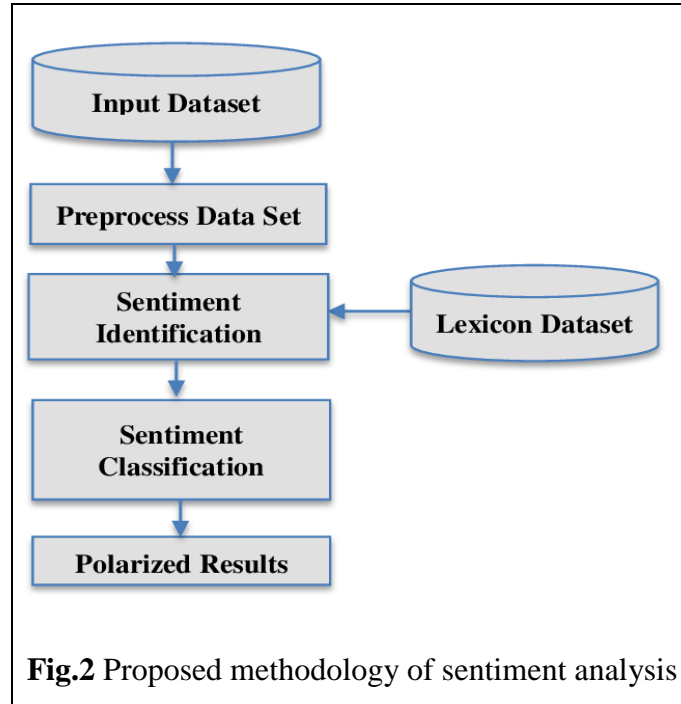
This dataset is collected from Amazon and it is in csv format. Each csv file contains number of reviews. Dataset had reviews of Camera, Laptops, Mobile phone, tablets, TVs, video surveillance.

#### ***3.2. Preprocessing***

In the preprocessing stage, tasks such as tokenization, stop word removal, stemming, and punctuation marks removal were carried out. The text was then converted into a bag of words representation. Preprocessing plays a crucial role in sentiment analysis and opinion mining.

### ***3.3. Sentiment Classification***

By employing various features, different machine learning algorithms were applied, and distinct accuracy measurements were calculated. The proposed method utilized the following tasks to classify sentiment analysis using machine learning techniques (Fig. 2)



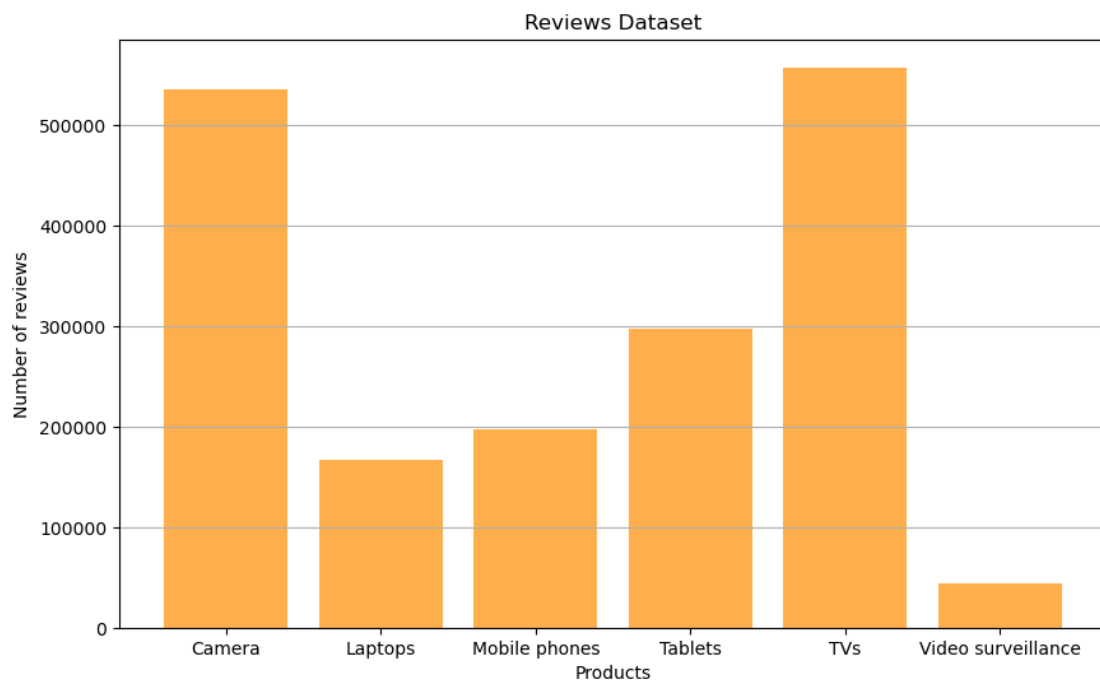
## **4. Experimental Results**

### ***4.1.Dataset Description***

Dataset contains reviews of Camera, Laptops, Mobile phones, tablets, TVs, video surveillance collected from Amazon. It is in the form of json files and each json file contains number of reviews (Fig. 3; Table 1).

### ***4.2. Classification Results***

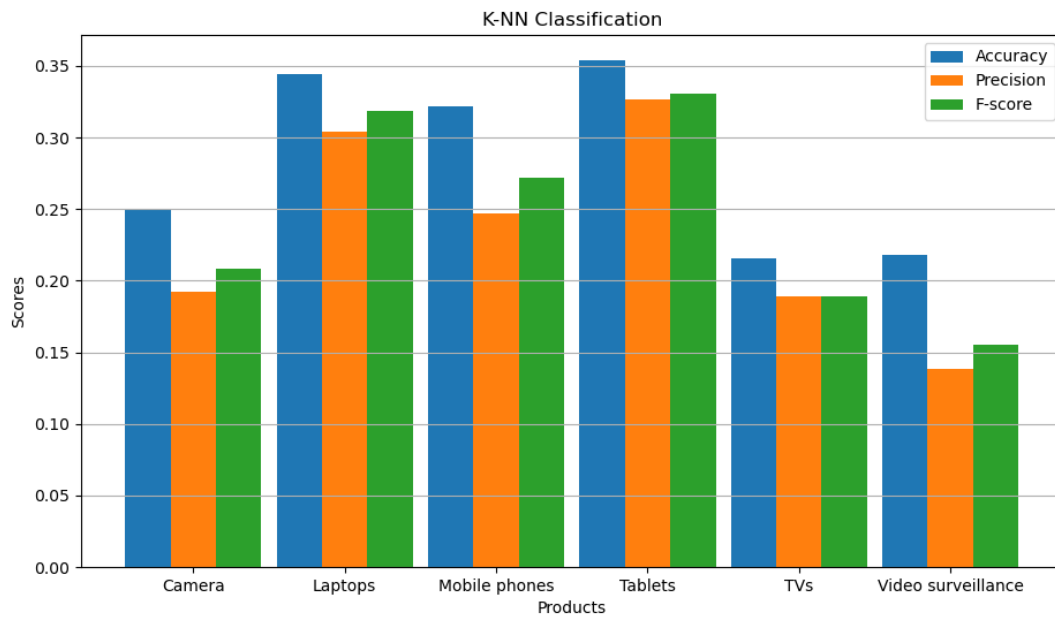
See (Fig. 4), (Fig. 5).



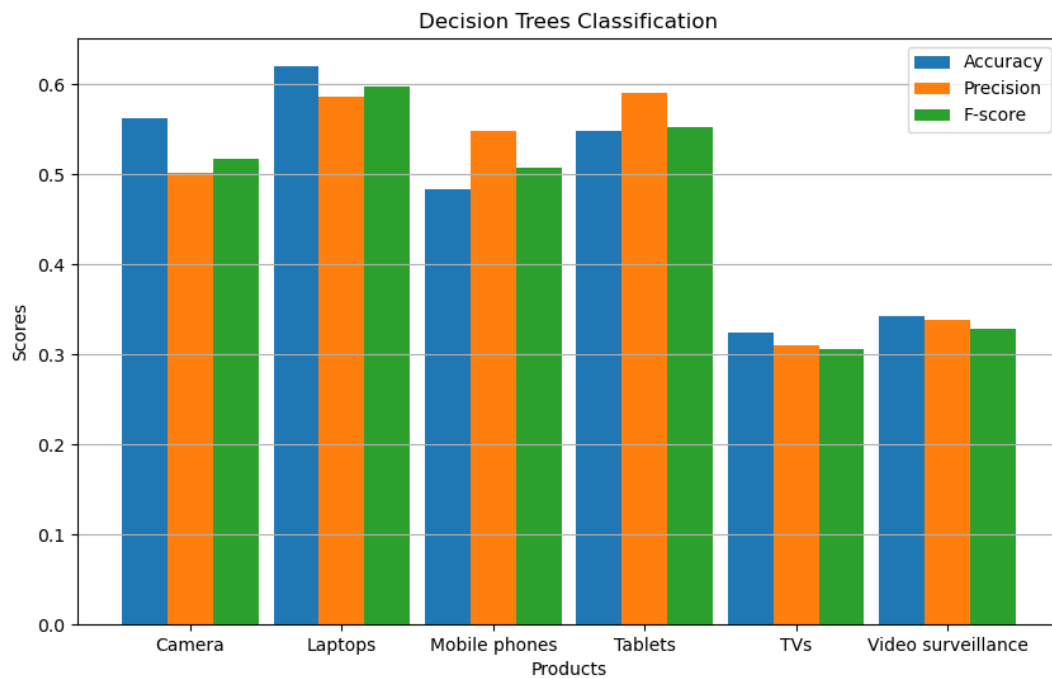
**Fig. 3** Graphical representation of number of reviews

**Table 1** Dataset and reviews count

Dataset name	Number of reviews
Camera	536317
Laptops	166853
Mobile phones	197469
Tablets	297830
TVs	557455
Video surveillance	44244



**Fig.4** Graphical representation of experimental results using k-NN model:



**Fig. 5** Graphical representation of experimental results using Decision tree model



**Table 2** Evaluation parameters for classifiers of datasets

Dataset	Classifiers					
	K-NN			Decision trees		
	Accuracy	Precision	Fscore	Accuracy	Precision	Fscore
Camera	25%	19.25%	20.83%	56.25%	50.26%	51.77%
Laptops	34.48%	30.45%	31.95%	62.06%	58.62%	59.77%
Mobile phones	32.25%	24.73%	27.20%	48.38%	54.83%	50.75%
Tablets	35.48	32.79%	33.17%	54.83%	59.13%	55.37%
TVs	21.62%	18.91%	18.91%	32.43%	31.08%	30.63%
Video surveillance	21.87%	13.80%	15.58%	34.37%	33.85%	32.81%

## 5. Conclusion

In today's digital age, sentiment analysis and opinion mining play a crucial role in various industries. The increasing volume of data generated by industries and social media platforms necessitates effective analysis to make informed decisions. This paper focuses on a dataset obtained from Amazon, comprising product reviews for cameras, laptops, mobile phones, tablets, TVs, and video surveillance devices. The proposed methodology utilizes a lexicon-based approach with machine learning techniques, specifically the k-NN and Decision Tree algorithms, to perform sentiment analysis on the reviews. The Decision Tree classifier achieved 62.06% accuracy for laptops reviews, while the k-NN classifier achieved an accuracy of 34.48%. For future work, conducting aspect-level sentiment analysis could further enhance the results, enabling a more detailed understanding of people's preferences or dislikes regarding specific aspects such as laptop quality, CPU performance, screen resolution, size, battery life, and more. Aspect-level analysis provides a fine-grained approach that yields promising results in sentiment analysis.

## **Bibliography**

- Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10. Association for Computational Linguistics, 2002
2. Turney, P.D.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002
3. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2004
4. Devika, M.D., Sunitha, Ca., Ganesh, A.: Sentiment analysis: a comparative study on different approaches. Proc. Comput. Sci. 87, 44–49 (2016)
5. Bhadane, C., Dalal, H., Doshi, H.: Sentiment analysis: measuring opinions. Proc. Comput. Sci. 45, 808–814 (2015)
6. Rosenthal, S., Farra, N., Nakov, P.: SemEval-2017 task 4: sentiment analysis in Twitter. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)
- 7 "Aspect and Sentiment Unification Model for Online Review Analysis" Authors: Bing Liu, Minqing Hu, Junsheng Cheng Published: 2005 Source: Proceedings of the 2005 ACM SIGIR International Conference on Research and Development in Information Retrieval
- 8 "Mining and Summarizing Customer Reviews" Authors: Minqing Hu, Bing Liu Published: 2004 Source: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
- 9 "Opinion Observer: Analyzing and Comparing Opinions on the Web" Authors: Bing Liu, Minqing Hu, Junsheng Cheng Published: 2005 Source: Proceedings of the 14th International World Wide Web Conference