

Article

Not peer-reviewed version

Advanced Sentiment Analysis and Topic Modeling on E-Commerce Reviews: A Comparison Study Using BERT and RoBERTa on Flipkart and Amazon

Himani Chauhan and [Teerath Kumar](#)*

Posted Date: 28 July 2025

doi: 10.20944/preprints202507.2312.v1

Keywords: sentiment analysis; topic modeling; e-commerce reviews; flipkart; Amazon; BERT; RoBERTa; natural language processing; transformer models; pre-trained models; fine-tuning; customer sentiment; LDA; latent dirichlet allocation; review analysis; model comp



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Advanced Sentiment Analysis and Topic Modeling on ECommerce Reviews: A Comparison Study Using BERT and RoBERTa on Flipkart and Amazon

Himani Chauhan and Teerath Kumar *

National College of Ireland, Dublin, Ireland

* Correspondence: teerath.menghwar2@mail.dcu.ie

Abstract

Our work delves into the realm of advanced sentiment analysis and topic modeling, focusing on e-commerce reviews from Flipkart and Amazon, two major platforms. Using state-of-the-art natural language processing models, namely BERT and RoBERTa, we aim to gain a deeper understanding of customer sentiments. By utilizing pre-trained transformer models and fine-tuned sentiment analysis techniques, we were able to improve the accuracy of classifying customer sentiments. Furthermore, we conducted a comparison between BERT and RoBERTa to evaluate which model better captures the nuances of e-commerce sentiments. We took our analysis a step further by implementing Latent Dirichlet Allocation (LDA) on the embeddings generated by BERT and RoBERTa. By doing so, we aimed to uncover underlying themes within the reviews, providing valuable insights into the factors that drive customer opinions. This comparative study enables us to gain a comprehensive understanding of the effectiveness of each model in capturing the diverse aspects of e-commerce sentiments.

Keywords: sentiment analysis; topic modeling; e-commerce reviews; flipkart; Amazon; BERT; RoBERTa; natural language processing; transformer models; pre-trained models; fine-tuning; customer sentiment; LDA; latent dirichlet allocation; review analysis; model comp

1. Introduction

Deep learning (DL) has demonstrated significant success across various domains, including image processing [1–7], audio analysis [8–12], and several other fields [13–18]. However, the application of DL in the e-commerce domain remains relatively underexplored. The modern e-commerce landscape is characterized by a dynamic interplay of user-generated content, such as online reviews, transaction records, and customer interactions, presenting unique challenges and opportunities for DL-based solutions. It is extremely necessary to comprehend the tone and subjects of online reviews as in this digital era, customers solely rely on online reviews to make decisions about what to buy. The two widely popular and two industry titans of e-commerce, Flipkart and Amazon have amassed a massive corpus of product reviews that bear witness to a range of customer opinions and experiences. In conjunction with this, the technologies which emerged as useful methodologies are Advanced Sentiment analysis and topic modeling providing businesses with in-depth understanding of customer feedback beyond what is typically seen as superficial. The recent times have witnessed the incorporation of advanced natural language processing models, redefining conventional frameworks for sentiment analysis and topic modeling. The emergence of models such as BERT and RoBERTa, that are known to leverage deep learning capabilities for extracting intricating language and contextual details that are beyond the scope of traditional techniques which signifies a paradigm shift in the area. Both the models possess the characteristics of bidirectional contextualized embeddings, proven to be remarkably effective at understanding the complex framework of human language. The extensive availability of

e-commerce platforms have resulted in the unparalleled influence of customer reviews. These reviews are not just an account of individual experiences but also serve as a repository of opinions that can influence the course of businesses. Therefore it is considered an advantage for businesses that opt to explore advanced cutting-edge methods which can reveal the hidden insights obscured in the abundance of user-generated material. The need for businesses to manage the dynamic world of e-commerce with accuracy and precision is higher than ever and it is the driving force behind this research. Businesses face extreme difficulties often to harness valuable insights from the stream of user-generated content as it continues and sees an exponential growth. We gravitate towards advanced and sophisticated models like BERT and RoBERTa as the conventional sentiment analysis techniques have shortcomings such as capturing the intricacy of language and the dynamic nature of themes. This study aims to go beyond the constraints of current approaches and methodologies to compare BERT and RoBERTa with regard to e-commerce reviews from Flipkart and Amazon. Following this approach, it endeavors to provide companies a solid foundation for deriving more complex and contextually rich insights from customer feedback. In this research, we perform a comparative study of BERT and RoBERTa in topic modeling and sentiment analysis of e-commerce reviews on Flipkart and Amazon. Research question states : To perform advance sentiment analysis and topic modeling on E-commerce reviews on Flipkart and Amazon dataset. The goals of the Research are as follows:

- We performed a thorough examination of the most recent developments in advanced sentiment analysis and topic modeling methods, particularly leveraging deep learning models.
- We create a flexible framework which is taking into account the unique features of Amazon and Flipkart datasets as we apply BERT and RoBERTa to e-commerce reviews.
- Implementing the constructed framework on both datasets by utilizing the contextual embeddings provided by BERT and RoBERTa.

The structure of this report is intended to facilitate a comprehensive investigation of the study subject. Following this introduction, we discuss Related work in Section 2 where we explore the existing body of literature and corpus of the research, with a focus on e-commerce platforms. Section 3 clarifies the research question and discusses the research methodology which serves as a guide for the next parts, highlights the justification for selected approach. Section 4 emphasises on design component where we develop the flexible framework for using BERT and RoBERTa in e-commerce reviews is developed. Section 5 focuses on implementing the framework to use in practice. It gives a thorough explanation of datasets which were utilized, outlining preprocessing steps and factors specific to e-commerce review features. Section 6 discusses the Results of applied framework and performance of BERT and RoBERTa are evaluated critically. The main conclusions are summarized in Section 7 and analyzes their importance in context of e-commerce analytics.

2. Related Work

This research uses latent discourse analysis (LDA) to identify the main conversational subject utilizing social media to ascertain public perceptions of two prominent Indian politicians. The authors gathered 18,000 tweets mentioning these political figures. After identifying the themes, they used a lexicon-based approach to match keywords in the tweets to a specified set of words in order to determine the moods or emotions that individuals were experiencing. They were able to determine whether people were happy, sad, angry, or apathetic. All of these analyses were carried out in real time using the SparkR program [19](P. Monish, Kumari, and C. Narendra Babu, 2018) Another study on people's reactions to increasing energy prices on Twitter was carried out. Once more, sentiment analysis is used to classify emotions as positive, neutral, or negative. Various deep learning models and conventional machine learning algorithms, such as BERT, DNN, and 1D-CNN, were used to analyze the data; however, BERT performed poorly because it was unable to learn domain-specific features, such as the energy-price and dataset nature domain. This work uses a machine learning model based on transformers and lexicon-based approaches to determine the optimum sentiment classification technique. They found that opinions ranged from largely negative to positive when

governments provided financial aid to cover high energy bills, and then to negative due to sharp increases linked to the conflict in Ukraine [20]Kastrati et al., 2023 To understand how customers feel about their products and services, businesses need to know what their customers have to say. The author evaluates and contrasts many contemporary topic modeling approaches with more established ones, such as FinBERT, RoBERTa, DistilBERT, and BERT, which were tailored especially to model the themes. BERTopic is a novel method that uses phrase embeddings to generate topics from Consumer Financial Protection Bureau (CFPB) data. The findings demonstrated that BERTTopics outperform more conventional topic modeling methods like LSA and LDA in terms of C_V scores. In contrast, BERT-based models interpret topics much more simply because no stemming or lemmatization is done [21]Vasudeva Raju et al., 2022. In a further study, the authors used a BERT-based model called the BERTopic algorithm to identify the major themes in a collection of publications on malaria research published between 2017 and 2022. In order to obtain new insights and understand the dynamics of past and future malaria research, this work employed a technique that included numerous themes [22]Alam Ahmad Hidayat et al., 2022. In conclusion, both studies demonstrate that BERT-based models have significantly enhanced topic modeling in many scenarios, particularly when used with BERTopic or changed on domain-specific data. By using these methods on our dataset of Flipkart product reviews, we may significantly increase our understanding of feelings and important topics. Both models are attractive methods for sentiment analysis since they have demonstrated strong performance in terms of accuracy and loss [23]Kundeti Naga Prasanthi et al., 2023. The BERT model was proposed in a different study for sentiment analysis. It was evaluated on two different datasets, and the outcomes were contrasted with those of the DistilBERT and RoBERTa models. BERT performed better than the other two models on the Sentiment 140 and Coronavirus tweets NLP datasets. On the Sentiment 140 dataset, training accuracy was 95.3%, validation accuracy was 93.13%, and testing accuracy was 92.76%. On the Coronavirus tweets NLP dataset, training accuracy was 94.1%, validation accuracy was 81.3%, and testing accuracy was 90.43%. Because of the distillation process, DistilBERT has fewer layers than BERT, which results in a difference in accuracy between the two methods. In these datasets, the NSP used to pre-train the model proved beneficial, as BERT performed better than the RoBERTa model [24]Archa Joshy and Sundar, 2022. The results of both tests demonstrate the effectiveness of transformer-based models in understanding sentence context, which is crucial while surfing social networking sites or reviews with complex language. Both studies used fine-tuning and transfer learning techniques to train their models. A pre-trained BERT model may be improved with our Flipkart review dataset. By being able to comprehend the unique subtleties of language and contextual information available in product evaluations, this can enhance the model's performance. This study explores the importance of internet reviews as useful tools for marketers, especially when examining consumer sentiment. It draws attention to the extreme bias found in customer evaluations while acknowledging the limits of conventional measurements like volume and ratings from online reviews. Regarding the effect of these measures on sales within the same product category, the research finds conflicting results. In addition, the study presents a unique method for analyzing customer views' excitement by taking the average word count and emotion score distribution into account in gleaning insightful information from online reviews for marketing objectives, this research adds to the body of knowledge. [25]R. Y. Kim, 2021. The use of Convolutional Neural Network (CNN) and Support Vector Machine (SVM) models for Customer Review Sentiment Analysis is covered in the article. Following the preprocessing and hyperparameter adjustments, the SVM model attained a remarkable 96% accuracy. The paper highlights how SVM and CNN are currently used for sentiment analysis and provides suggestions for further optimization. The authors want to improve model accuracy and shorten the SVM and CNN training times. Furthermore, they suggest going beyond positive/negative binary sentiment analysis to multilevel categorization, which includes categories like "Neutral Sentiment." [26]R. Uma et al., 2022. Four different strategies are used in this study to explore the categorization of online reviews for handicraft goods on Flipkart: Support Vector Machine (SVM), Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), and Logistic Regression (LR). When

the performance of different methods is compared, the study concludes that SVM performs more accurately than LR, KNN, and MLP. The study emphasizes the value of using Count Vectorizer, which ranks words in a corpus according to frequency of occurrence as statistically significant. It points out that although the tfidf technique understands word significance, it is unable to convey words' true meanings and their semantic links. This study makes a unique contribution by categorizing internet review information into three categories: negative, neutral, and positive. This classification helps craft entrepreneurs better understand their target market's preferences. [27]Kanakamedala, Singh & Talasani, 2023. The study focuses on implementing Kano model-inspired models in place of textual previews and reviews to implement strategic quality control techniques in e-commerce. The research use data crawling methodologies to get product descriptions and evaluations from Amazon.com. To identify quality characteristics, Latent Dirichlet Allocation (LDA), a topic modeling technique, is applied to the obtained data. Positive/negative quality analysis is then used to assess these quality parameters. [28]Im et al., 2021, p. 290 This article examines consumer happiness as it relates to Surabaya restaurant reviews. Recognizing the critical importance that customer happiness plays in the restaurant industry, the research uses TextBlob sentiment analysis and the Naïve Bayes classification approach to evaluate client sentiments. According to the study, Naïve Bayes outperforms TextBlob by 2.94%, with an accuracy of 72.06%. This implies that sentiments from customer evaluations may be effectively learned using Naïve Bayes. [29]Laksono et al., 2019, p. 51 This study introduces a novel approach by combining TF-IDF-weighted factors' relative importance with emotional polarity values extracted using a GRU model, aiming to understand the impact of emotional states of major factors on customer satisfaction, a theoretical method not explored in prior research While the study focuses on consumer perspectives, it underscores the need for future research to explore issues from the standpoint of small businesses, other user groups, and broader environmental and social aspects related to delivery applications. [30]Lee et al. (2022) In order to anticipate restaurant quality and make recommendations for future enhancements, this article presents an opinion mining approach based on Natural Language Processing (NLP) that is intended to be used for evaluating consumer views about restaurants. Sentiment analysis is performed using the SentiStrength classifier, which divides opinions into positive, negative, and neutral strengths. The suggested NLP-based method and manual observation both work well for forecasting client opinions. The efficiency rate of 85.714% in obtaining consumer views is indicated by the testing findings. In order to improve the accuracy of consumer preference classification, the article ends with recommendations for future approaches for expanding the analysis to incorporate additional customer perspectives. [31]Ara et al., 2020 Opinion mining, also known as sentiment analysis, is a case study that examines people's opinions, sentiments, and feelings regarding particular entities. Sentiment polarity classification is a key challenge in sentiment analysis that is addressed in this study. The internet product reviews gathered for this study are sourced from <http://Amazon.com>. Sentiment polarity classification and POS is a proposed approach that comes with step-by-step instructions. More creative and practical methods need to be developed in order to get over the obstacles that sentiment analysis and opinion mining are now facing. [32]Pankaj et al., 2019. In the context of mining online user-generated material, the study focuses on sentiment analysis's essential difficulty, paying special attention to customer evaluations. The study presents and applies a deep learning-based framework called Deep Learning-based Sentiment Analysis (DL-SA) to determine the semantic orientation of each line in customer reviews, acknowledging the critical role that customer reviews play in well-informed decision-making. Large-scale training data is necessary for these deep learning techniques to be successful, and this work attempts to provide this demand. [33]Seetharamulu, Reddy, and Naidu (2020) The study suggests using sentiment analysis to analyze product evaluations on the online retailer Tokopedia. The method integrates three machine learning classifiers into a Sentiment Selector and a Soft Voting Model. With a greater accuracy of 69%, the testing findings show that the suggested strategy performs better than current techniques. When it comes to precision and recall values, the Soft Voting Model outperforms the Sentiment Selector. The results indicate that this combined strategy works well to help e-commerce companies understand

the views and feedback of their customers and may be a useful tool for sentiment analysis. Among these, the Soft Voting Model stands out as the best option, demonstrating how it may improve the creation of new products and services as well as the general e-commerce platform consumer experience. [34]Zulfadli and Indrabayu, 2023, p. 295 Using supervised and unsupervised models, the research suggests a hybrid approach to sentiment analysis in the context of customer evaluations. TF-IDF vectorization-based Random Forest Model is selected after testing a number of supervised models. They draw attention to its usefulness in recommendation systems, imagining how customers may be grouped according to comparable evaluations left on sites like Amazon. This approach may be used to improve customers' comprehension and decision-making in the e-commerce area since it combines supervised and unsupervised models with browser plugin implementation. [35]Welgamage et al., 2022 The implementation of Aspect-level Sentiment Analysis utilizing Support Vector Machine (SVM) and Naïve Bayes (NB) machine learning methods is the main emphasis of this article. Comparing these algorithms' performances using the F1 measure, Precision, and Recall is the main goal. The outcomes show that, in terms of accuracy, the Naïve Bayes method performs better than the Support Vector Machine technique. The report ends with a summary of future work that includes comparing with Naïve Bayes and putting the C4.5 algorithm into practice. By shedding light on the efficacy of several machine learning methods for aspect-level sentiment categorization, this research advances the area of sentiment analysis. [36]vanaja20 The handling of three-star ratings is the main emphasis of the article, which tackles the problem of sentiment orientation discrepancy in customer evaluations. An existing sentiment orientation disparity classification model (SODCM) that initially classifies 3-star reviews as acting similarly to negative reviews is the source of the problem. This presumption is contested by the suggested algorithm, posSODCM, which views 3-star ratings as favorable. In order to better comprehend the disparities between star ratings and related reviews, the study ends by proving the usefulness of the suggested algorithm and outlining future research including the application of SODCM to product reviews from multiple marketplaces. The results pave the way for more research to improve the accuracy of sentiment classification models in various e-commerce scenarios and add to the continuing conversation about the difficulties of sentiment analysis, particularly when dealing with complex instances like three-star ratings. [37]Yao, Chatterjee, & Zhou, 2022 The study's main goal is to provide a sentiment analysis procedure for Twitter "tweets," with an emphasis on thoughts on different laptop models. Five primary qualities serve as the framework for the sentiment analysis: portability, durability, screen quality, battery life, and central processing unit performance. With an F-Measure of 94%, the study's performance was deemed adequate; its limits were related to the users' usage of idioms, slang, and sardonic statements. The researchers are dedicated to improving and expanding the suggested sentiment analysis technique, as seen by their acknowledgement of these difficulties and their intention to solve them in further publications. [38]Polsawat et al., 2018

3. Methodology

We follow a structured and comprehensive methodology which makes use of qualitative and quantitative approaches by utilizing cutting-edge natural language processing techniques and harnessing the power of Python programming through libraries like TensorFlow and Hugging Face Transformers.

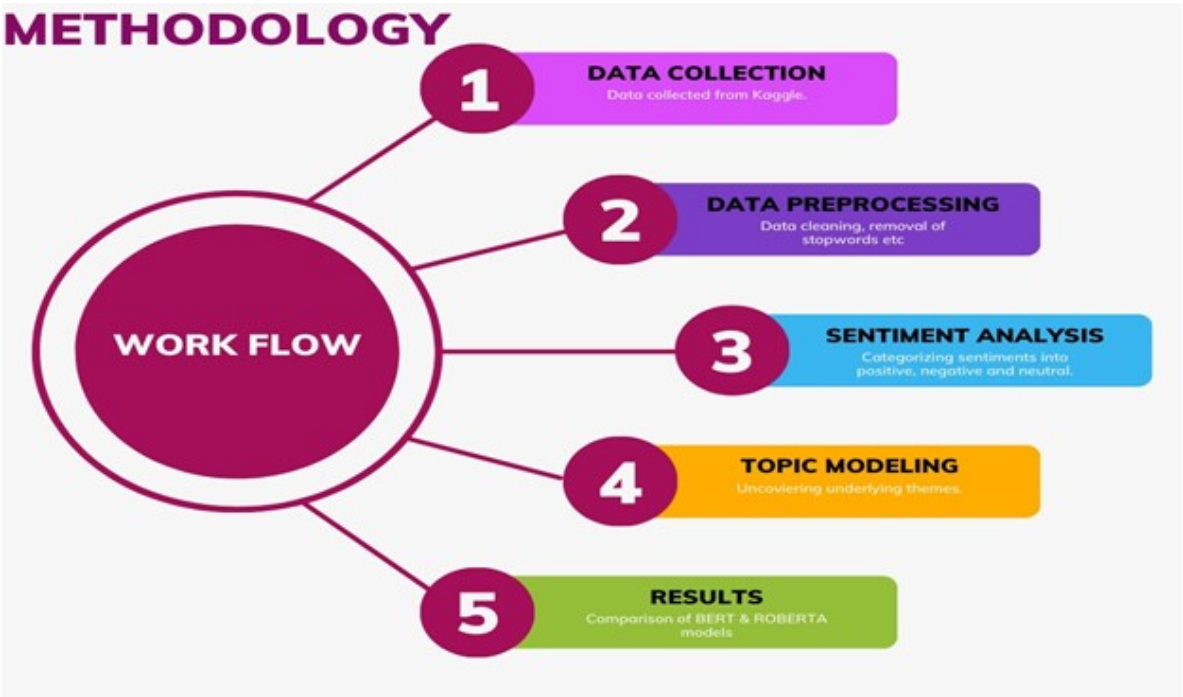


Figure 1. WORK FLOW DIAGRAM.

The first step, Data Collection involves finalizing dataset for this project. Our dataset comprises of a diverse set of consumer reviews from Amazon and Flipkart that covers a wide range of products. The dataset was downloaded from Kaggle, this platform is known to host datasets in domain of machine learning and data science. The second step, Data Preprocessing is an important step that keeps in check the quality and consistency of the dataset. In the first step of data preprocessing stage, this pipeline removes duplicate reviews, it achieves this objective by cross-checking each review’s textual content which makes sure that each review has a unique representation. Further, the dataset undergoes a filtering process where the dataset is checked for non-English reviews using a straightforward approach where the presence of English alphabets in the reviews are verified. Further in the preprocessing step, reviews undergoes normalization techniques, where a custom function is created called preprocesstext that is responsible for converting all text to lowercase, remove all the URLs and punctuations The third step, Model Selection involves selecting suitable models for sentiment analysis and topic modeling tasks. The two widely known natural processing language pre-trained models , BERT and RoBERTa known for their impressive quality of capturing context and being able to understand semantics. BERT and RoBERTa were effortlessly employed in this study as it was handled by Hugging Face Transformers library that was completed with pre-trained versions and ability to fine-tune. These models were fine-tuned on a labeled dataset based on sentiment analysis. The fourth step, Exploratory Data Analysis for sentiment analysis and topic modeling revolves around the process of analyzing and visualizing the data which is related to sentiment lables so as to gain insights and to achieve a deeper understanding of datasets.

- 1.) Data Overview: Flipkart and Amazon datasets information and shape were printed first to get a clear underswtanding of the composition and dimendions. To check if the datasets are complete the isnull(). Sum() function is employed so that we can find if there is any missing values in the dataset. The flipakart dataset has 1262 entries and amazon dataset has 393578 entries. Flipkart has 6 columns ProductName, Price, Rate, Review and Summary. Amazon has 10 columns named Id, ProductId, UserId, ProfileName, HelpfulnessNumerator, HelpfulnessDenominator, Score, Time, Summary, Text.
- 2.) Review Length Analysis: Here a custom function computelength is made which measures the length of each analyzed review in terms of that how many characters it contains. This function creates a histogram using sns.histplot, that showcxases distribution of text lengths of processed reviews and gives valuable insights.

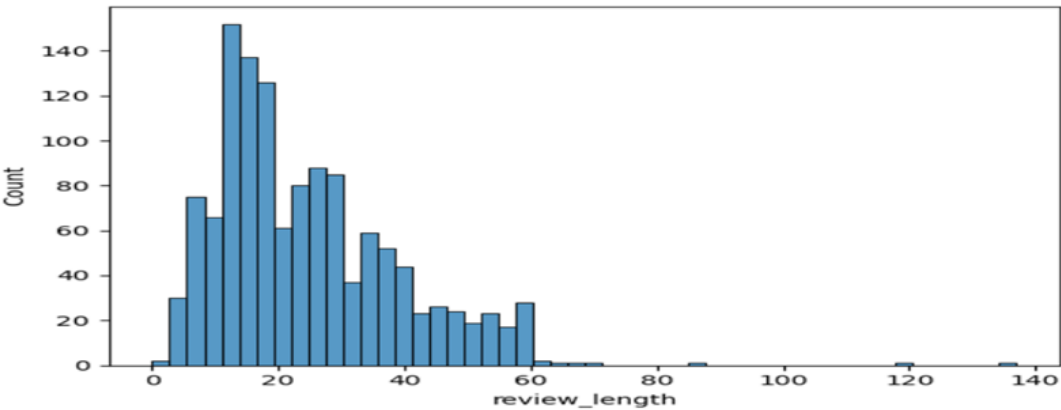


Figure 2. Flipkart Review Length.

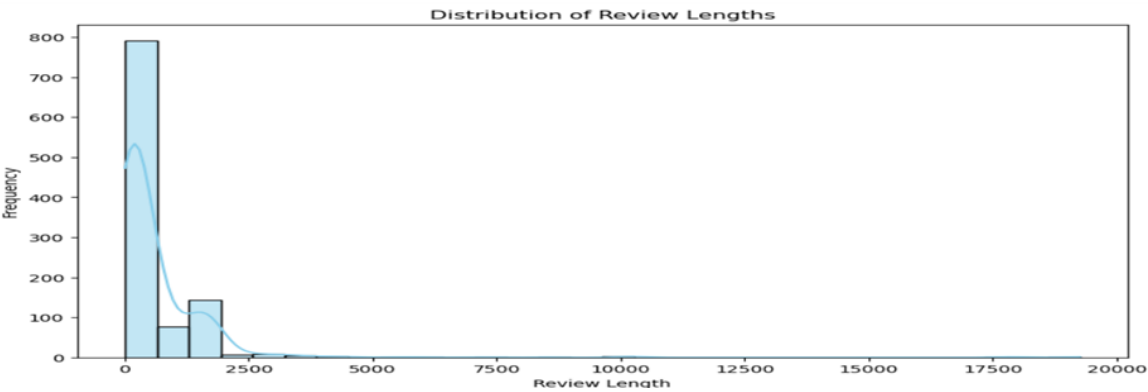


Figure 3. Amazon Reviews length.

3.) Sentiment Distribution: In this part we generate new columns by performing classification of reviews into positive, negative and neutral based on the rating bins. The sentiment distribution is then represented by a countplot which allows for a swift understanding of overall sentiment breakdown.

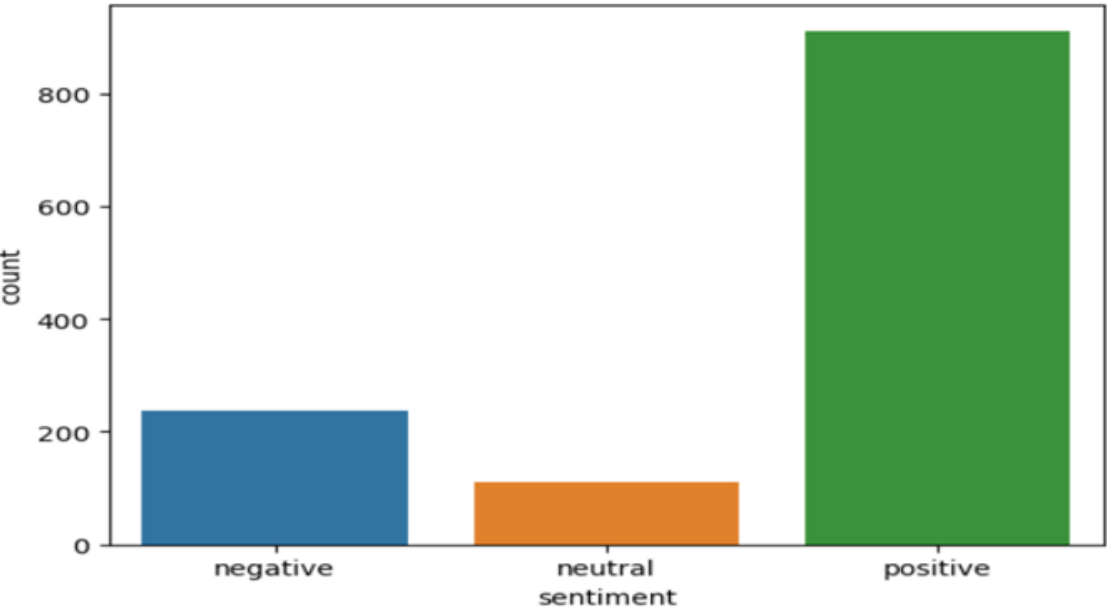


Figure 4. Flipkart Sentiment Distribution.

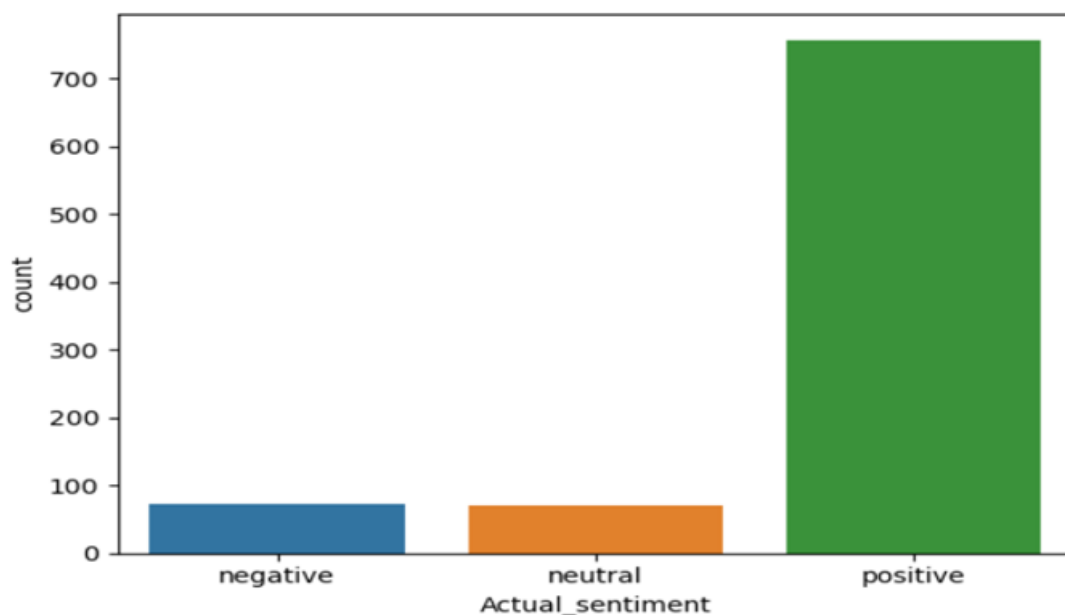


Figure 5. Amazon Sentiment Distribution.

4.) Word Frequency Analysis: Word Frequency Analysis is performed on the reviews of both the Flipkart and the Amazon dataset. We first make sure that the non-string reviews are removed and we drop the missing values in the 'processedreview' column. After having the clean dataset, we convert the reviews into strings.



Figure 6. Flipkart Wordcloud.



Figure 7. Amazon Wordcloud.

5.) N-gram Analysis (Bigrams): In this step `gettopngrams` function is utilized to get the most frequently occurring bigrams from the processed reviews. We generate a barplot in order to visually

represent frequency of bi-grams. The fifth step, Traditional Sentiment Analysis is studied so as to compare results of traditional ways and advanced sentiment analysis models which makes use of modern NLP techniques such as BERT and RoBERTa. We are using VADER which stands for Valence Aware Dictionary and Sentiment Reasoner to create a baseline model for sentiment analysis. This tool is widely known to work effectively in handling sentiment analysis for textual data and particularly for social media and usergenerated content.

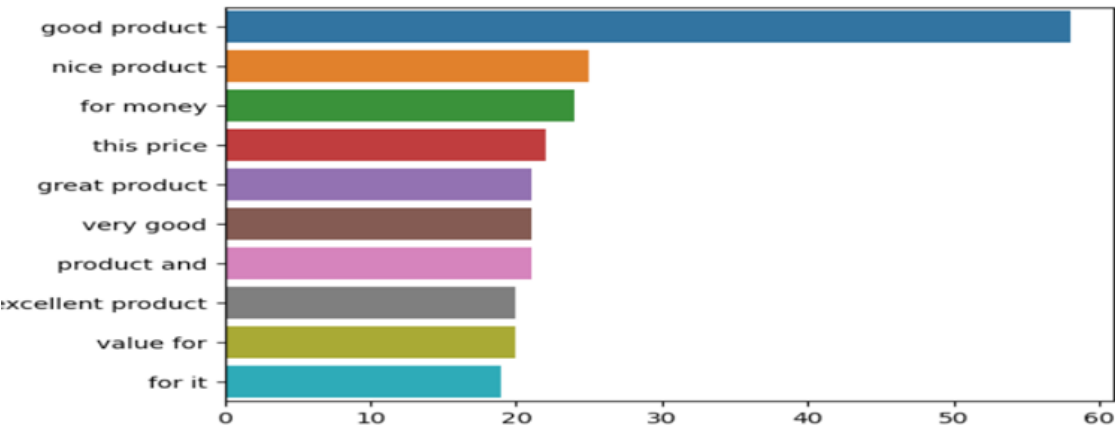


Figure 8. N-gram for Flipkart.

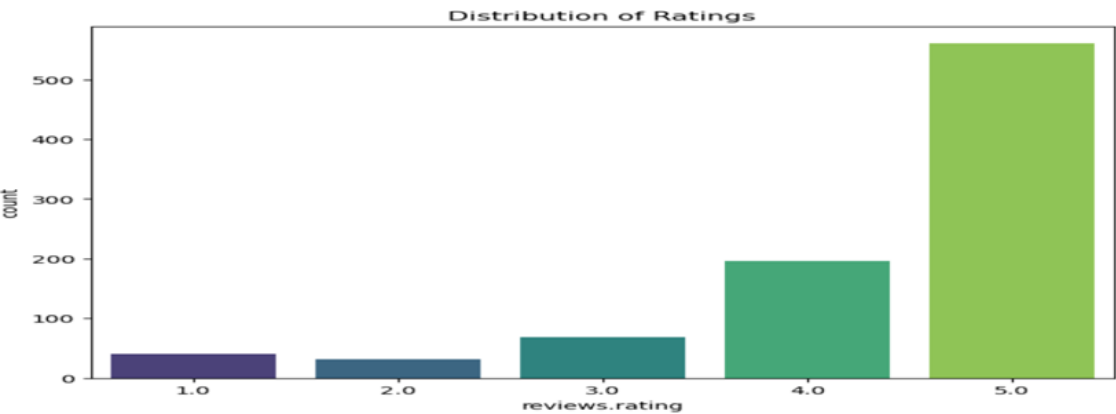


Figure 9. Amazon Reviews Rating.

The sixth step, Sentiment Analysis using BERT and Roberta, this task utilizes state -of-the -art BERT and RoBERTa models, which is later fine-tuned for optimal performance. On the basis of insights gathered from the training dataset these models were specifically trained to categorize reviews as positive, negative and neutral. The performance evaluations includes accuracy, precision, recall, and F1-score metrics. The seventh step, Topic Modeling, this step the second part of our study which is carried out after sentiment analysis. To find out the various topics present in the reviews, Latent Dirichlet Allocation (LDA) is utilized for base model and it’s results are compared to those by BERT and RoBERTa embeddings. The interesting fact about LDA is it’s unsupervised nature, its ability to identify dominating themes without any prior supervision.

4. Design Specification

In this section, we discuss in detail the design specification for our project. This project has dual goals to extract underlying issues from user reviews to piece together a complex web of customer behaviours across e-commerce industry. To achieve these two objectives, we use the most recent transformer-based models named BERT and RoBERTa.
SENTIMENT ANALYSIS ARCHITECTURE: The latest and state-of-the-art transformer-based models BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (Robustly optimized

BERT approach) are deployed for sentiment analysis. These two models have showcased impressive results in all the tasks involving the interpretation of natural language. The first step is to tokenize the review text in the architecture. The contextual information is then captured through transformer layers by encoding. In order to obtain smooth integration and application of pre-trained BERT and RoBERTa models, the Hugging Face Transformers library is utilized. Tokenization is the crucial stage in our sentiment analysis process where textual input is divided into smaller pieces known as tokens. Our sentiment analysis algorithm is better able to understand the nuances of sentiment expressed in different linguistic patterns for these contextual embeddings. The Hugging Face Transformers library provides extensive collection of pretrained transformer models and easy interfaces enabling a smooth integration into the architecture.

TOPIC MODELING ARCHITECTURE: In this study, the challenging problem in e-commerce of identifying latent topics is tackled by going beyond sentiment analysis. The idea behind investigating latent topics is necessary for learning more about the underlying themes and issues which customers have raised in their reviews. Bidirectional Encoder Representations from Transformers, is known to be remarkably effective in capturing semantic relationships and contextual information from textual data. The goal is to improve the granularity and quality of topics by integrating BERT embeddings into our topic modeling framework by using BERTopic library. The motivation for using BERT embeddings is rooted in its ability to extract complex semantic nuances and contextual information. In contrast to traditional methods, BERT considers a word's entire context inside a phrase which results in a more complex and context-aware representation of language. The goal is to identify more contextually relevant and nuanced topics in e-commerce reviews through topic modeling with BERT embeddings.

DATA PREPROCESSING: In our data preprocessing pipeline we focused on text normalization methods, stop words elimination method, and lemmatization methods. Also a close attention is paid to the difficulties posed by these architectures and techniques are put in place to get the textual data ready for best possible model performance, as special tokenization need of models BERT and RoBERTa are taken into consideration. In order to enhance the general uniformity of text and standardize the representation of words, text normalization is an important step in our preprocessing pipeline. This method involves changing all text to lowercase with the aim of decreasing the effect of case variances. By performing this, the words with diverse instances is prevented to be considered as independent entities and guarantees that this model generalizes across many situations successfully. Stopwords, are often used but have minimal semantic significance and might add noise to our study. We use stopword removal methods as a part of our preprocessing techniques to filter out these terms. This will improve the model's capacity to identify appropriate patterns by helping to minimize dimensionality of data and concentrate on more significant words. Another component of the pipeline is lemmatization which is trying to simplify words into their most basic and root form.

IMPLEMENTATION OF BERT AND ROBERTA: The implementation is carefully structured with an emphasis on using PyTorch and the transformers library. The light is shed on how these pre-trained models which are specifically tailored for sentiment analysis are fine-tuned. Furthermore, the adjustment of output layers is discussed for binary sentiment classification and embedding extraction for further analysis and clustering are covered.

EVALUATION METRICS: Our study relies mostly on metrics and this section sheds light on how to choose and utilize common assessment measures. This covers coherence scores for topic modeling and sentiment analysis's accuracy, precision, recall, and F1 score. A detailed explanation of the reasoning behind the selection of each metric and its critical function in assessing model performance provides a nuanced understanding of our assessment strategy.

EXPERIMENTAL CONFIGURATION: The tests were run on high-performance computer settings so as to handle the computational needs of big transformer models and guarantees stability. The solution is designed to be flexible and adaptive to changing requirements of our investigation, allowing investigations on a wide range of datasets and subsets and configurations.

COMPARISON METRICS: Going beyond conventional assessment scores, a thorough comparison is outlined between BERT and RoBERTa.

5. Implementation

Our project, "Advanced Sentiment Analysis and Topic Modeling on E-Commerce Reviews," follows a methodical approach to execution, emphasizing on four different models in particular. Using cutting-edge language models, namely BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (Robustly optimized BERT method), the main objective is to do sentiment analysis and topic modeling on the Flipkart and Amazon datasets.

TOOLS USED:

1. Python: The project's main programming language is Python, which offers a flexible and effective framework for data preprocessing, analysis, and model building.
2. Pandas: Pandas is used for analysis and data processing. Large dataset handling is made easier by its robust data structures, including DataFrames.
3. scikit-learn: scikit-learn is used for machine learning tasks, such as model training, performance evaluation, and data preprocessing. It offers a large selection of tools for model evaluation and classification.
4. Transformers: One essential tool for dealing with transformer-based models, such as BERT and RoBERTa, is the Hugging Face Transformers library. It makes it possible to incorporate pre-trained models into the project with ease.
5. PyTorch: The neural network model implementation and training framework for deep learning is PyTorch. It offers dynamic computation graphs and a user-friendly model development interface.
6. BERT and RoBERTa: Topic modeling and sentiment analysis rely on the pre-trained BERT and RoBERTa models that are available in the Hugging Face model hub. The efficiency of these transformer-based models in natural language processing tasks is well known.
7. Seaborn and Matplotlib: These are the visualization libraries used in our project to produce informative plots and charts that facilitate data interpretation and model performance analysis.
8. Jupyter Notebooks: A collaborative and interactive environment for writing, distributing, and documenting code is offered by Jupyter Notebooks. They are particularly helpful for exploring and analyzing data.
9. NumPy: NumPy is used to handle arrays and perform numerical operations, which is a crucial component of the project's mathematical computations.
10. LabelEncoder: To ensure compatibility with machine learning models, category labels are encoded into numerical representation using the scikit-learn LabelEncoder.
11. AdamW Optimizer: During the training stage, the BERT and RoBERTa models are adjusted with the help of the AdamW optimizer, which improves convergence and performance.
12. NLTK (Natural Language Tool Kit): The Natural Language Toolkit, or NLTK, is utilized when a lot of the data you have access to is unstructured and readable by people. This data has to be preprocessed before a computer program can analyze it. This library uses the tokenizing principle, which enables work on shorter text segments that are somewhat coherent and comprehensible even when taken out of the context of the entire text. Because the first portion of the research's data consists of unstructured tweets, tokenizing each word or phrase makes using NLTK more practical.
13. VADER (Valence Aware Dictionary for Sentiment Reasoning): It is a sentiment analysis tool that specifically adapts to sentiments expressed in social media. Its sentiment analysis is based on a vocabulary and rule-based system. A sentiment lexicon is a workable combination of words or other lexical items that have been assigned labels, either positive or negative, according to their underlying semantic orientation. As a result, it was essential to determining the sentiment analysis in this instance.

TRADITIONAL SENTIMENT ANALYSIS USING VADER:

The VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment analyzer is used to do the

sentiment analysis on the Flipkart dataset. VADER is a preconfigured lexicon and rule-based sentiment analysis tool used for text data processing, particularly in the context of brief communications and social media. First, the 'SentimentIntensityAnalyzer' class from the 'vaderSentiment' library is used to initialize the VADER sentiment analyzer. For every review, this analyzer calculates a compound sentiment score that represents the overall sentiment polarity. Using 'LabelEncoder' from scikit-learn, the real sentiment labels from the dataset were then converted into numeric representation, making it easier to compare and assess the VADER predictions. The VADER sentiment ratings are given a threshold in order to get the final predictions. Reviews were classified as positive if their compound sentiment score was higher than the threshold or negative if it was below. Ultimately, the 'classificationreport' function from scikit-learn is used to create an detailed classification report that included important metrics for both positive and negative sentiment classes, including precision, recall, and F1-score. This implementation provides a thorough breakdown of the model's predicted accuracy on both positive and negative emotions, enabling an informative evaluation of VADER's performance on sentiment analysis for the Flipkart dataset.

SENTIMENT ANALYSIS ON FLIPKART DATASET USING BERT:

A methodical approach to fine-tune the BERT model for multi-class sentiment classification is to use Sentiment Analysis on the Flipkart Dataset using BERT. The BERT tokenizer and sequence classification model are initially implemented using the Hugging Face Transformers library. To ensure that the training data is compatible with BERT's input requirements, it is tokenized and encoded, consisting of preprocessed Flipkart reviews. After that, the data is transformed into PyTorch tensors and the sequences are padded to the longest length possible within the encoded reviews. The sentiment labels in the 'sentiment' column are encoded using a LabelEncoder, turning them into numeric labels, to enable multi-class sentiment classification. To handle the batch processing efficiently during training, DataLoader instances are initialized when the data is divided into training, validation, and test sets. The model is fine-tuned through a training loop going over several epochs, and the optimization is performed using the AdamW optimizer. The model's performance is evaluated on the test set and predictions are compared with the true labels during the evaluation phase.

SENTIMENT ANALYSIS ON FLIPKART DATASET USING ROBERTA:

Utilizing state-of-the-art language models, a systematic and detailed method is used to execute sentiment analysis using RoBERTa on the Flipkart dataset. First, the code loads the pre-trained RoBERTa tokenizer and model using the Hugging Face Transformers library. Tokenization and encoding are performed on Flipkart dataset, the reviews have been preprocessed and stored in column 'processedreview', which contains the cleaned and tokenized reviews. To ensure consistent input dimensions, the sequences are padded to the longest possible length within the encoded reviews. A LabelEncoder is utilized to encode the sentiment labels which makes it easier to convert the sentiment categories into numerical values. Further, to start batch processing, the dataset is divided into training, validation, and test sets, also the associated DataLoader objects are constructed. The training loop loops over the given number of epochs, changing the model parameters to optimize for sentiment classification. The model is fine-tuned using the AdamW optimizer. Assessing the model's performance on the test set is the focus of the evaluation stage.

TOPIC MODELING ON FLIPKART DATASET USING BERT:

In this section we are using BERT embeddings and K-Means clustering to do topic modeling on the Flipkart dataset. Tokenizing the preprocessed reviews using the BERT tokenizer at the start of the procedure gives encoded sequences. The sequences are padded to the longest length possible in order to guarantee consistency. After loading the pre-trained BERT model, the representation corresponding to the [CLS] token is extracted to give the embeddings for each review. These embeddings are then subjected to K-Means clustering in an effort to isolate discrete themes within the dataset. There are five clusters, which represent the themes; however, this number can be changed if necessary. Ultimately, every review is allocated to a certain subject cluster, and the result displays an overview of the reviews

linked to every discovered topic. This approach offers a useful perspective on the latent themes found in the e-commerce reviews by utilizing the rich contextual information recorded by BERT embeddings to identify underlying subjects in the Flipkart dataset.

TOPIC MODELING ON FLIPART DATASET USING ROBERTA:

To extract relevant themes from processed reviews, the Flipkart Dataset's Topic Modeling implementation uses RoBERTa in a structured manner. To produce a consistent input format, the reviews are first tokenized using the RoBERTa tokenizer, which takes special tokens, truncation, and padding into account. After determining the maximum length from the encoded reviews, all sequences are padded to maintain consistency. For each review, embeddings are captured using the RoBERTa model, which has been pre-trained on large-scale data to capture contextual information that is nuanced. To make processing downstream easier, the resulting embeddings are flattened. An important step here is to make sure that the values are not negated by modifying them according to the value that is the absolute minimum of the set. Later, Latent Dirichlet Allocation (LDA), a popular topic modeling approach with its five components, are employed to identify underlying themes. The flattened embeddings goes through the LDA transformation, and the top words connected to each identified subject are shown to showcase the resulting themes. It combines sophisticated natural language processing techniques to uncover thematic patterns.

SENTIMENT ANALYSIS ON AMAZON DATASET :

Extracting sentiment ratings from processed reviews is made simple with the help of VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment analysis on the Amazon dataset. The 'SentimentIntensityAnalyzer' class from the NLTK (Natural Language Toolkit) library is the main focus of this implementation's importing of the required libraries. Every processed review in the Amazon dataset is given a compound sentiment score by this analyzer, which is intended for text sentiment analysis. The composite score takes into account both positive and negative polarity to capture the overall attitude. Creating a new column in the Amazon dataset called "Sentiment" to hold the computed compound sentiment scores is an important step in the implementation process. This column provides a quantitative assessment of the sentiment polarity for every review, acting as a numerical representation of sentiment. In order to improve interpretability, the continuous sentiment scores are transformed into categorical labels using the function 'categorizesentiment'. Reviews are classified as "Positive" if their score is higher than or equal to 0.05, and as "Negative" if it is less than or equal to -0.05. A score that falls between -0.05 and 0.05 is referred to as "Neutral." In conclusion, a practical method is used in the sentiment analysis utilizing VADER on the Amazon dataset, giving a numerical representation of sentiment scores and associated category labels. More in-depth studies and interpretations of customer sentiments are made possible by this implementation, which lays the groundwork for understanding the sentiment distribution inside the Amazon reviews.

SENTIMENT ANALYSIS ON AMAZON DATASET USING BERT:

Sentiment Analysis with BERT applied on the Amazon Dataset shows how to carefully fine-tune a pretrained BERT model for sentiment classification problem. The first step is to initialize the BERT tokenizer and model, which are particularly set up for sequence classification with output labels, by utilizing the robust Hugging Face Transformers library. To prepare the processed reviews from the Amazon dataset for input into the BERT model, the reviews go through tokenization, encoding, and padding. LabelEncoder is utilized to handle class imbalance by splitting the data into training, validation, and test sets and converting sentiment labels into a numeric representation. The AdamW optimizer is used in the training loop, and the model is fine-tuned across three epochs while paying close attention to data type conversions. Classification metrics are used in the evaluation stage to evaluate the model's performance, and the scikit-learn classificationreport function is used to create the sentiment analysis report in the end. This implementation offers a detailed method that demonstrates how BERT is applied for sentiment analysis on the Amazon dataset, offering insightful information about the sentiment distribution and model effectiveness.

SENTIMENT ANALYSIS ON AMAZON DATSET USING ROBERTA:

The utilization of RoBERTa for sentiment analysis on the Amazon dataset is carried out in a methodical and detailed manner. Correct tokenization, truncation, and padding are ensured as the reviews are first encoded using the RoBERTa tokenizer as part of the preprocessing step of the dataset. A label encoder is utilized to encode the labels, converting sentiment categories into integers. In the next step, the data is divided into testing and training sets, and then further subdivided for validation. An AdamW optimizer is used across three epochs to fine-tune the RoBERTa model, which was first trained for sequence classification. The training loop iterates over batches of tokenized input and labels. The model is evaluated on the test set. The trained RoBERTa model is used to make predictions once the test reviews are tokenized and encoded in a manner akin to that of the training set. The sentiment labels are then obtained by using the label encoder to decode the predictions. The classificationreport function in the scikitlearn library is used to produce classification metrics including precision, recall, and F1 score. The application demonstrates how to use the powerful transformer-based model RoBERTa for sentiment analysis applications. The model’s ability to learn from and generalize the given Amazon dataset is demonstrated by the training loop and assessment procedure. The application showcases how to use the powerful transformer-based model RoBERTa for sentiment analysis applications.

6. Evaluation

1. Sentiment Analysis on Flipkart dataset using VADER

The results of this VADER sentiment analysis show a clear bias in favor of predicting negative sentiment, with a recall of 1.00 and a precision of 0.28 for the negative class. However, the total accuracy is just 0.19, indicating that there may be difficulties in applying this method to reliably categorize sentiments. The poor accuracy, recall, and F1-score for the neutral and positive emotion classes indicate that the model suffers most with these sentiments. The macro-average F1-score, which takes into account the ratio of recall to precision for every class, is 0.16, suggests that the VADER sentiment analysis has trouble performing well overall. Taking into account class imbalances, the weighted average F1-score is likewise low at 0.08. These results imply that VADER might not be able to fully represent the complex sentiments portrayed in the dataset.

VADER Sentiment Analysis Report:				
	precision	recall	f1-score	support
negative	0.28	1.00	0.43	237
neutral	0.02	0.08	0.03	111
positive	0.00	0.00	0.00	911
accuracy			0.19	1259
macro avg	0.10	0.36	0.16	1259
weighted avg	0.05	0.19	0.08	1259

Figure 10. VADER Sentiment Analysis Report.

2. Sentiment Analysis on Flipkart Dataset U sing BERT

The BERT model’s sentiment analysis results for the Flipkart dataset show an overall accuracy of 83%. With an exceptional F1-score of 0.91, the model’s performance in recognizing neutral sentiments is impressive, with an accuracy score of 0.89 and a recall of 0.96. But as seen by an accuracy of 1.00 but a lower recall of 0.06, the model struggled to identify negative sentiments and resulted in a lower F1-score of 0.12. Additionally, positive sentiments has a strong recall of 0.88 and a good accuracy of 0.58, resulting a balanced F1-score of 0.70. Given that the macro-average F1-score is 0.58, it is crucial to treat all classes equally. After accounting for the class imbalance, the weighted average F1-score is 0.79.

BERT Sentiment Analysis Report:				
	precision	recall	f1-score	support
0	0.58	0.88	0.70	16
1	1.00	0.06	0.12	16
2	0.89	0.96	0.92	95
accuracy			0.83	127
macro avg	0.83	0.63	0.58	127
weighted avg	0.87	0.83	0.79	127

Figure 11. BERT Sentiment Analysis Report.

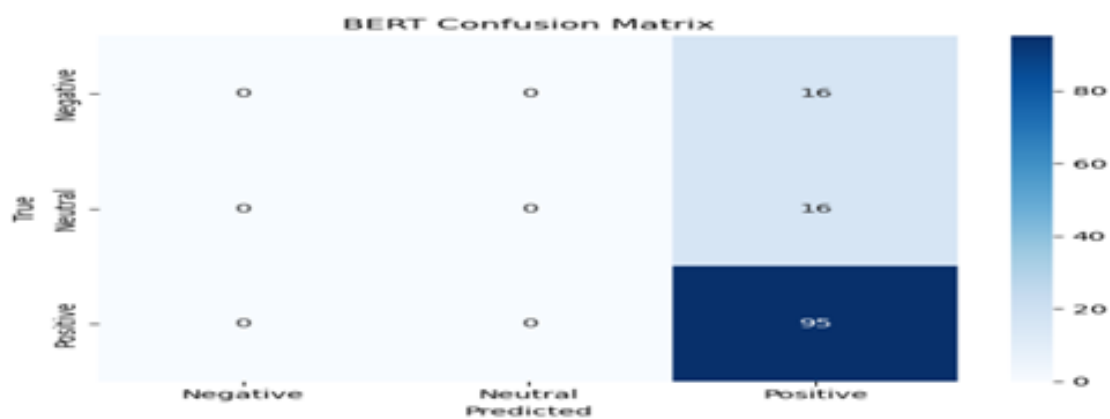


Figure 12. Confusion Matrix.

3. Sentiment Analysis on Flipkart Dataset Using RoBERTa

The following report presents the sentiment analysis results on the Flipkart dataset using RoBERTa. For each sentiment class—Negative (0), Neutral (1), and Positive (2)—the metrics for accuracy, recall, and F1-score are calculated. It’s important to remember that the model’s accuracy, recall, and F1-score values for the Negative (0) and Neutral (1) classes are 0.00 since it failed to predict any cases for these classes. Conversely, the model showed a recall of 1.00, an F1-score of 0.86, and a precision of 0.75 for the Positive (2) class, indicating a rather accurate classification. Using the Flipkart dataset, the RoBERTa sentiment analysis model’s overall accuracy is 0.75. The weighted average F1-score, which takes into account both precision and recall across classes, is 0.64, whereas the macro-average F1-score, which takes class imbalances into consideration, is 0.29. These findings add to the assessment of the model’s efficacy in sentiment prediction by showcasing its performance in several sentiment categories.

4.Topic Modeling on Flipkart Dataset using BERT In topic modeling processed reviews are grouped into themes using K-Means clustering and BERT embeddings, and then outputs the reviews for each subject.

5. Topic Modeling On Flipkart Dataset Using Roberta Based on the patterns found in the dataset, the algorithm identified unique themes in the topic modeling results using RoBERTa. A collection of keywords that emphasize the recurring elements within each topic serves as a representation of that topic. These subjects shed light on the underlying trends that run across the dataset. The algorithm found similarities relevant to a wide range of topics, including locations, events that have happened in real life, and maybe difficult issues.

6. Sentiment Analysis Using Vader On Amazon Dataset

A difficult situation is shown by the VADER sentiment analysis results on the Amazon dataset. The model’s overall accuracy of 0.00 indicates that there aren’t many effective sentiment predictions. The

RoBERTa Sentiment Analysis Report:				
	precision	recall	f1-score	support
0	0.00	0.00	0.00	16
1	0.00	0.00	0.00	16
2	0.75	1.00	0.86	95
accuracy			0.75	127
macro avg	0.25	0.33	0.29	127
weighted avg	0.56	0.75	0.64	127

Figure 13. RoBERTa Sentiment Analysis.



Figure 14. Topics Modeling using BERT.

Topic #1 Reviews:
good quality product
like an assembled one can get for 45k cheap quality
product is good but huge in size
worth the money desert cooler live up to the name
ok kind of coolernot effective in hot conditions
cools well but the user manual can be improved
fantastic air cooler in affordable price
ultimate cooler at this price range
item is good but delhivery courier service worst
good product recommended
good product in this price
suitable for lightweight stuff
great product for offered price
good camera for this price range
awesome product but no free installation
awesome both product service
good quality product

Figure 15. Topics Modeling using BERT.

Topic #2 Reviews:

super

awesome

fair

nice

brilliant

fabulous

just wow

excellent

terrific

simply awesome

wonderful

good not best

worthless

not good

horrible

just okay

delightful

Figure 16. Topics Modeling using BERT.

Topic #3 Reviews:
useless product
highly recommended
great product
expected a better product
nice product
very good
very poor
best in the market
bad quality
good choice
badly disappointed
utterly disappointed
must buy cool product
cool cooler for extreme summer heat
crompton ozone 75 desert air cooler
product changed
damaged

Figure 17. Topics Modeling using BERT.



Figure 18. Topics Modeling using BERT.

Topic #0: 160, Watts, real, Disc, supremacist, consequence, statistics, Byrne, Cond, separatist
Topic #1: DA, United, estead, real, arrangement, ignorance, statistics, disorders, grad, Oaks
Topic #2: railways, United, Lash, supremacist, estead, warning, sounds, grad, surgeries, injury
Topic #3: Watts, consequence, noting, separatist, ignorance, Lash, Sn, injury, sounds, Supreme
Topic #4: Oaks, semester, Patterson, arrangement, Lash, wheel, UG, disorders, stole, separatist

Figure 19. Topic Modeling using Roberta.

model’s inability to differentiate between negative, neutral, and positive sentiments is further shown in the classification report, which shows F1-score values of 0.00, precision, and recall for every class. The confusion matrix draws attention to the large number of cases classified as "positive" and "nan," which suggests a serious problem with misclassification. There aren’t many real positive predictions in any of the sentiment categories, suggesting that the algorithm has trouble producing meaningful sentiment labels. These results highlight the drawbacks of use VADER for sentiment analysis on the Amazon dataset and point to the necessity of more advanced models to manage the complexities of e-commerce reviews.

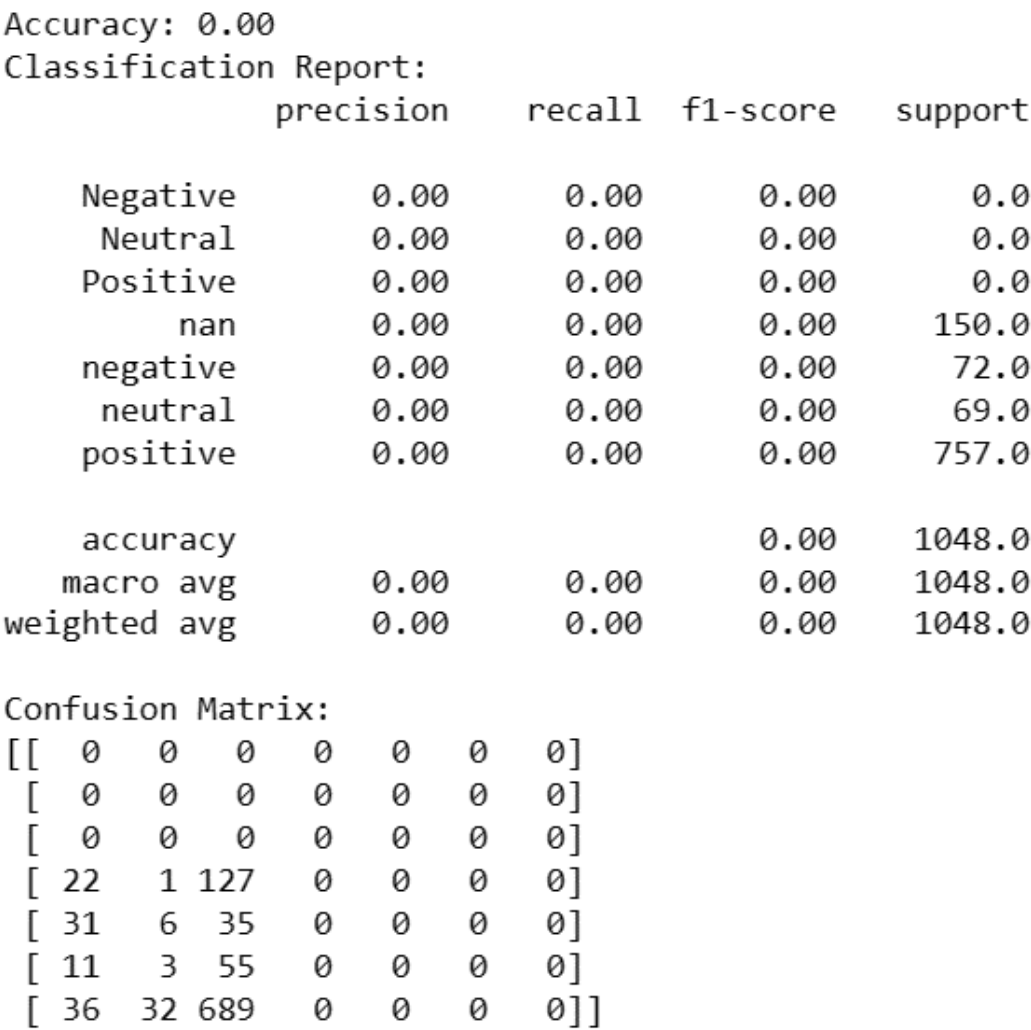


Figure 20. Classification Report.

7 Sentiment Analysis Using Bert On Amazon Dataset

The following report gives us an overview of the sentiment analysis outcomes using BERT on the Amazon dataset. For every sentiment class—Negative (0), Neutral (1), Positive (2), the precision, recall, and F1-score metrics are calculated. The model did not predict any occurrences for the Negative (0) and Neutral (1) classes, as evidenced by accuracy, recall, and F1-score values of 0.00. On the other hand, the model obtained an F1-score of 0.78, a recall of 0.95, and a precision of 0.67 for the Positive (2) class, indicating a relatively accurate classification. Using the Amazon dataset, the BERT sentiment analysis model’s overall accuracy is 0.57. The F1-score with a weighted average of 0.47 and a macro-average of 0.20 are obtained.

BERT Sentiment Analysis Report:				
	precision	recall	f1-score	support
0	0.00	0.00	0.00	28
1	0.00	0.00	0.00	11
2	0.67	0.95	0.78	63
3	0.00	0.00	0.00	4
accuracy			0.57	106
macro avg	0.17	0.24	0.20	106
weighted avg	0.40	0.57	0.47	106

Figure 21. BERT Sentiment Analysis Report.

8. Sentiment Analysis Using Roberta On Amazon Dataset

The following report provides an overview of the sentiment analysis outcomes using RoBERTa on the Amazon dataset. The RoBERTa sentiment analysis model’s overall accuracy on the Amazon dataset is 0.76. When class imbalances are taken into account, the macro-average F1-score is 0.25, and the weighted average F1-score is 0.70 when accuracy and recall are taken into account across classes. These outcomes offer a thorough evaluation of the model’s performance in identifying sentiment fluctuations in the Amazon dataset.

RoBERTa Sentiment Analysis Report:				
	precision	recall	f1-score	support
0	0.00	0.00	0.00	9
1	0.00	0.00	0.00	2
2	0.80	0.95	0.87	83
3	0.17	0.09	0.12	11
accuracy			0.76	105
macro avg	0.24	0.26	0.25	105
weighted avg	0.65	0.76	0.70	105

Figure 22. RoBERTa Sentiment Analysis Report.

7. Discussion

The comparative study shows different results for sentiment analysis tests on the Flipkart and Amazon datasets show different performances amongst the models. The Flipkart dataset’s VADER sentiment analysis resulted in a comparatively low accuracy of 19%, proving difficulties in correctly classifying sentiments. It is a known limitation with VADER when it comes to managing complex emotions, though. On the Flipkart dataset, however, BERT and RoBERTa showed much better accuracy rates of 83% and 75%, respectively, demonstrating their effectiveness in capturing the complexities of emotion expressions. VADER’s accuracy on the Amazon dataset was a surprisingly low 0%, highlighting its limitations in processing the complex and varied sentiments present in e-commerce reviews. In contrast, BERT and RoBERTa showed competitive results with 57% and 76% accuracy, respectively. These findings highlight how crucial it is to use sophisticated models for sentiment

analysis in e-commerce settings, such as BERT and RoBERTa. Even if the accuracies offer a numerical measurement, additional qualitative examination and investigation of misclassifications may increase our understanding of the advantages and disadvantages of the models and direct future developments and adjustments in experimental design. Using BERT for topic modeling, the results show different review clusters that correspond to different user moods and opinions. Regarding product quality, price, and suitability, Topic #0 presents a mixed picture with divergent viewpoints, illustrating the variety of customer experiences. Reviews pertaining to the general product quality, cost, and customer service encounters seem to be captured within Topic #1. Topic #2 features a collection of reviews that are highly positive and use terms like "super," "amazing," and "terrific," suggesting that the consumers are happy. Topic #3 displays a range of opinions, both favorable and unfavorable, showcasing the different experiences customers had using the items. Topic #4 highlights the variety of sentiments by displaying a wide range of opinions, including both positive and negative comments.

DATASET	MODEL	ACCURACY	PRECISION	RECALL	F-1 SCORE
Flipkart	VADER	19%	0.28	1.00	0.16
Flipkart	BERT	83%	0.89	0.96	0.91
Flipkart	RoBERTa	75%	0.0	0.00	0.00
Amazon	VADER	0%	-	-	-
Amazon	BERT	57%	0.67	0.95	0.78
Amazon	RoBERTa	76%	0	0.00	0.00

Figure 23. Sentiment Analysis Results.

Results of the RoBERTa Topic Modeling:

The study reveals particular topics that are highlighted by the RoBERTa-based topic modeling.

Topic #0: A variety of themes, such as "160," "Watts," "real," "Disc," and "supremacist," are suggested by the keywords. These topics have unclear coherence which is a sign of difficulty choosing important subjects.

Topic #1: Words like "DA," "United," and "ignorance" point to a possible social and political issues-related subject. It is difficult to determine the actual context or consistency of this topic given the absence of comprehensive material.

Topic #2: The phrases "railways," "Lash," and "surgeries" suggest a broad range of themes, which might include medical procedures, transportation, or other unrelated ideas.

Topic #3: It is difficult to understand the underlying concept of this topic since the terms "Watts," "consequence," and "Supreme" lack a strong sense of consistency.

Topic #4: Words like "Oaks," "semester," and "wheel" in this topic give just a limited understanding of a cohesive concept.

Problems with Coherence: The detected themes show difficulties with coherence and interpretability, particularly those produced by RoBERTa. It is unclear how well the model captures significant themes when there are phrases that appear to be unrelated within the same issue.

Model	Dataset	Topic #	Keywords
BERT-based Topic	Flipkart	0	recommended, product, buy, waste, money, working, fine
BERT-based Topic	Flipkart	1	good, quality, product, worth, money, cooler, price
BERT-based Topic	Flipkart	2	super, awesome, fabulous, excellent, value, money
BERT-based Topic	Flipkart	3	useless, highly, recommended, great, bad, choice
BERT-based Topic	Flipkart	4	unsatisfactory, worth, money, mind-blowing, classy
RoBERTa-based Topic	Flipkart	0	160, Watts, real, Disc, supremacist, consequence
RoBERTa-based Topic	Flipkart	1	DA, United, este ad, real, arrangement, ignorance
RoBERTa-based Topic	Flipkart	2	railways, United, Lash, supremacist, warning, sounds
RoBERTa-based Topic	Flipkart	3	Watts, consequence, noting, separatist, ignorance
RoBERTa-based Topic	Flipkart	4	Oaks, semester, Patterson, arrangement, Lash, wheel

Figure 24. Topic Modeling Results.

8. Conclusions and Future Work

Our research aimed to investigate and compare the efficiency of BERT and RoBERTa models in handling sentiment analysis and topic modeling tasks on two well-known e-commerce platforms. The study, which is titled "Advanced Sentiment Analysis and Topic Modeling on E-Commerce Reviews: A Comparison Study Using BERT and RoBERTa on Flipkart and Amazon," is extensively documented. Finding significant themes in customer reviews and comprehending the complexities of emotion expressions were the main goals of the study. The goals included assessing how well BERT and RoBERTa performed in topic modeling and sentiment analysis as well as outlining the advantages and disadvantages of each strategy. Using rigorous approaches, our study effectively answered the research issue by using the Flipkart and Amazon datasets for sentiment analysis and topic modeling. The main conclusions showed that BERT performed well in sentiment analysis for Flipkart dataset, especially when it came to correctly categorizing sentiments and RoBERTa performed well on Amazon dataset. RoBERTa did, however, have difficulties, particularly in anticipating neutral and negative thoughts. While RoBERTa’s results were harder to interpret and suggested further work, BERT’s topic modeling results showed clear clusters of reviews. With BERT being effective in capturing sentiment complexities, the research’s conclusions underscore the importance of model selection in sentiment analysis tasks. Future research seek to concentrate on enhancing the RoBERTa model, investigating substitute techniques for topic modeling, and tackling class imbalance in sentiment datasets. Future studies may focus on increasing the interpretability of topic modeling results and enhancing sentiment analysis models’ ability to handle a wider range of sentiments. All things considered, our study establishes the groundwork for future developments in topic modeling and sentiment analysis as they

relate to e-commerce reviews. **Acknowledgments:** This paper forms part of my thesis, and I would like to express my sincere gratitude to Mr. Teerath Kumar Menghwar for his outstanding supervision and mentorship throughout the Data Analytics course. His guidance has been invaluable to the successful completion of this work. I also wish to thank the faculty members of NCI for their expertise and support, which have greatly enriched my academic experience. Lastly, I am deeply thankful to my parents and friends for their unwavering encouragement and support throughout my academic journey.

References

1. Kumar, T., Mileo, A. & Bendeche, M. KeepOriginalAugment: Single Image-based Better Information-Preserving Data Augmentation Approach. *20th International Conference On Artificial Intelligence Applications And Innovations*. (2024)
2. Kumar, T., Mileo, A., Brennan, R. & Bendeche, M. RSMDA: Random Slices Mixing Data Augmentation. *Applied Sciences*. **13**, 1711 (2023)
3. Chandio, A., Gui, G., Kumar, T., Ullah, I., Ranjbarzadeh, R., Roy, A., Hussain, A. & Shen, Y. Precise single-stage detector. *ArXiv Preprint ArXiv:2210.04252*. (2022)
4. Kumar, T., Mileo, A., Brennan, R. & Bendeche, M. Advanced Data Augmentation Approaches: A Comprehensive Survey and Future directions. *ArXiv Preprint ArXiv:2301.02830*. (2023)
5. Kumar, T., Park, J., Ali, M., Uddin, A., Ko, J. & Bae, S. Binary-classifiers-enabled filters for semi-supervised learning. *IEEE Access*. **9** pp. 167663-167673 (2021)
6. Roy, A., Bhaduri, J., Kumar, T. & Raj, K. A computer vision-based object localization model for endangered wildlife detection. *Ecological Economics, Forthcoming*. (2022)
7. Kumar, T., Brennan, R., Mileo, A. & Bendeche, M. Image data augmentation approaches: A comprehensive survey and future directions. *IEEE Access*. (2024)
8. Chandio, A., Shen, Y., Bendeche, M., Inayat, I. & Kumar, T. AUDD: audio Urdu digits dataset for automatic audio Urdu digit recognition. *Applied Sciences*. **11**, 8842 (2021)
9. Turab, M., Kumar, T., Bendeche, M. & Saber, T. Investigating multi-feature selection and ensembling for audio classification. *International Journal Of Artificial Intelligence & Applications*. (2022)
10. Raj, K., Singh, A., Mandal, A., Kumar, T. & Roy, A. Understanding EEG signals for subject-wise definition of armoni activities. *ArXiv Preprint ArXiv:2301.00948*. (2023)
11. Kumar, T., Park, J. & Bae, S. Intra-Class Random Erasing (ICRE) augmentation for audio classification. *Proceedings Of The Korean Society Of Broadcast Engineers Conference*. pp. 244-247 (2020)
12. Park, J., Kumar, T. & Bae, S. Search of an optimal sound augmentation policy for environmental sound classification with deep neural networks. *Proceedings Of The Korean Society Of Broadcast Engineers Conference*. pp. 18-21 (2020)
13. Singh, A., Raj, K., Meghwar, T. & Roy, A. Efficient Paddy Grain Quality Assessment Approach Utilizing Affordable Sensors. *Artificial Intelligence*. **5**, 686-703 (2024)
14. Khan, W., Kumar, T., Cheng, Z., Raj, K., Roy, A. & Luo, B. SQL and NoSQL Databases Software architectures performance analysis and assessments—A Systematic Literature review. *arXiv 2022. ArXiv Preprint ArXiv:2209.06977*.
15. Turab, M., Kumar, T., Bendeche, M. & Saber, T. Investigating multi-feature selection and ensembling for audio classification. *arXiv 2022. ArXiv Preprint ArXiv:2206.07511*.
16. Kumar, T., Bhujbal, R., Raj, K. & Roy, A. Navigating Complexity: A Tailored Question-Answering Approach for PDFs in Finance, Bio-Medicine, and Science. (Preprints,2024)
17. Barua, M., Kumar, T., Raj, K. & Roy, A. Comparative Analysis of Deep Learning Models for Stock Price Prediction in the Indian Market. (Preprints,2024)
18. Raj, K. & Mileo, A. Towards Understanding Graph Neural Networks: Functional-Semantic Activation Mapping. *International Conference On Neural-Symbolic Learning And Reasoning*. pp. 98-106 (2024)
19. P. Monish, Kumari, S. and C. Narendra Babu (2018). Automated Topic Modeling and Sentiment Analysis of Tweets on SparkR. doi:https://doi.org/10.1109/iccnc.2018.8493973.
20. Kastrati, Z., Imran, A.S., Daudpota, S.M., Memon, M.A. and Kastrati, M. (2023). Soaring Energy Prices: Understanding Public Engagement on Twitter Using Sentiment Analysis and Topic Modeling With Transformers. *IEEE Access*, **11**, pp.26541–26553. doi:https://doi.org/10.1109/access.2023.3257283.

21. Vasudeva Raju, S., Kumar Bolla, B., Nayak, D.K. and Kh, J. (2022). Topic Modelling on Consumer Financial Protection Bureau Data: An Approach Using BERT Based Embeddings. [online] IEEE Xplore. doi:https://doi.org/10.1109/I2CT54291.2022.9824873.
22. Alam Ahmad Hidayat, Rudi Nirwantono, Arif Budiarto and Bens Pardamean (2022). BERT-based Topic Modeling Approach for Malaria Research Publication. doi:https://doi.org/10.1109/icimcis56303.2022.10017743.
23. Kundeti Naga Prasanthi, Rallabandi Eswari Madhavi, Degala and Battula Sravani (2023). A Novel Approach for Sentiment Analysis on social media using BERT & ROBERTA Transformer-Based Models. doi:https://doi.org/10.1109/i2ct57861.2023.10126206.
24. Archa Joshy and Sundar, S. (2022). Analyzing the Performance of Sentiment Analysis using BERT, DistilBERT, and RoBERTa. doi:https://doi.org/10.1109/iprecon55716.2022.10059542.
25. Kim, R. Y. (2021) "Using Online Reviews for Customer Sentiment Analysis," IEEE Engineering Management Review, 49(4), pp. 162-168. doi: 10.1109/EMR.2021.3103835.
26. Uma, R., H, A. S., Jawahar, P., & Rishitha, B. V. (2022). Support Vector Machine and Convolutional Neural Network Approach to Customer Review Sentiment Analysis. In 2022 1st International Conference on Computational Science and Technology (ICCST) (pp. 239-243). Chennai, India. doi: 10.1109/ICCST55948.2022.10040381.
27. Kanakamedala, V. C., Singh, S. H., & Talasani, R. (2023). Sentiment Analysis of Online Customer Reviews for Handicraft Product using Machine Learning: A Case of Flipkart. In 2023 International Conference for Advancement in Technology (ICONAT) (pp. 1-6). Goa, India. doi: 10.1109/ICONAT57137.2023.10080169.
28. Im, E. T., Tung, P. H., Oh, M. S., Lee, J. Y., & Gim, S. (2021) A Study on the Extraction of Customer Satisfaction Factors Based on the Customer Satisfaction Model Using Text Review and Preview. 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), Ho Chi Minh City, Vietnam, pp. 290-291. DOI: 10.1109/SNPDWinter52325.2021.00077
29. Laksono, R. A., Sungkono, K. R., Sarno, R., & Wahyuni, C. S. (2019). Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes. In 2019 12th International Conference on Information & Communication Technology and System (ICTS) (pp. 49-54). Surabaya, Indonesia. doi:10.1109/ICTS.2019.8850982.
30. Lee, S., Choi, W., & So, J. (2022). A Study on the Factors Affecting Customer Satisfaction in Delivery Applications: Focusing on Sentiment Analysis of Review Data. In 2022 IEEE/ACIS 7th International Conference on Big Data, Cloud Computing, and Data Science (BCD) (pp. 34-38). Danang, Vietnam. IEEE. doi: 10.1109/BCD54882.2022.9900519.
31. Ara, J., Hasan, M. T., Al Omar, A., & Bhuiyan, H. (2020). Understanding Customer Sentiment: Lexical Analysis of Restaurant Reviews. 2020 IEEE Region 10 Symposium (TENSYP), Dhaka, Bangladesh, pp. 295-299. doi: 10.1109/TENSYP50017.2020.9230712.
32. Pankaj, P., Pandey, Muskan, & Soni, N. (2019). Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) (pp. 320-322). Faridabad, India. DOI: 10.1109/COMITCon.2019.8862258.
33. Seetharamulu, B., Reddy, B. N. K., & Naidu, K. B. (2020). Deep Learning for Sentiment Analysis Based on Customer Reviews. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-5). Kharagpur, India. doi: 10.1109/ICCCNT49239.2020.9225665.
34. Zulfadli, A. A. Ilham and Indrabayu (2023) "Sentiment Analysis with Soft-Voting Method on Customer Reviews for Purchasing Transactions of E-Commerce," 2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), BALI, Indonesia, pp. 295-299. doi: 10.1109/IAICT59002.2023.10205954.
35. Welgamage, V.R., Senarathne, U.A.C., Madhubhashani, N.H.A.C., Liyanage, T.C., and Asanka, P.P.G.D. (2022) 'Overall and Feature Level Sentiment Analysis of Amazon Product Reviews Using Machine Learning Techniques and Web-Based Chrome Plugin', in: International Research Conference on Smart Computing and Systems Engineering (SCSE), Colombo, Sri Lanka, 2022, pp. 205-210. DOI: 10.1109/SCSE56529.2022.9905125.
36. Vanaja, S. and Belwal, M., 2018. Aspect-Level Sentiment Analysis on E-Commerce Data. In: 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, pp. 1275-1279. DOI: 10.1109/ICIRCA.2018.8597286.

37. Yao, D., Chatterjee, I., & Zhou, M. (2022). Conditioning Customers' Product Reviews for Accurate Classification Performance. In 2022 IEEE International Conference on Networking, Sensing and Control (ICNSC) (pp. 1-5). Shanghai, China. doi: 10.1109/ICNSC55942.2022.10004165.
38. Polsawat, T., Arch-int, N., Arch-int, S., & Pattanachak, A. (2018). Sentiment Analysis Process for Product's Customer Reviews Using Ontology-Based Approach. In 2018 International Conference on System Science and Engineering (ICSSE) (pp. 1-6). New Taipei, Taiwan. doi: 10.1109/ICSSE.2018.8520261.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.