Sentiment Analysis on Reviews of Breweries using Text Summarization

Gayathri Reddy Sane
L.V. SaiCharan Kothapalli
Aishwarya Ramesh

Department of Computer Science, Old Dominion University

Abstract

Breweries have become increasingly popular, with a growing number of people sharing their thoughts and opinions about them in the form of textual comments and reviews. Understanding the sentiments expressed in these reviews is crucial for both brewery owners and consumers. In this project, we aim to tackle this challenge through the application of advanced Natural Language Processing (NLP) techniques, including Large Language Models and Deep Learning Models, with the ultimate goal of classifying text-based comments or feedback as positive, negative, or neutral. Our approach also involves the utilization of machine learning models to detect sentiment trends and address contextual concerns, thereby contributing valuable insights to the field of sentiment analysis.
Keywords-Sentiment analysis, Text Summarization, Product reviews, Natural language processing.

1 Introduction

1.1 Importance of the Problem

Sentiment analysis is a process of extracting and understanding the sentiments defined in the text document. The NLP with artificial intelligence capability and text analytics are used to determine whether the sentiment of the opinion is positive, negative and neutral. Sentiment analysis is also providing a business intelligence which can be used to make good impactful decision. Sentiment analysis and sentiment classification are the two methodologies used in opinion mining[3]. Sentiment analysis of brewery reviews serves as a valuable tool for various stakeholders. For brewery owners, it offers an opportunity to gain insights into customer satisfaction and identify areas for improvement. Consumers can benefit from sentiment analysis by making informed decisions when choosing a brewery to visit. Furthermore, the larger community interested in sentiment analysis and NLP stands to benefit from this project as it adds to the growing body of knowledge about sentiment analysis in specific domains, enhancing the ability to capture nuanced sentiment in text. In this paper several sentiment analysis methods have been discussed. Even though we find several papers proposed by different researchers in this area, there is a need to make sentiment analysis more accurate and easy to understand. Sentiment analysis is extremely useful in various situations[3].

1.2 Limitations of Current Research Works

In the domain of brewery review sentiment analysis, existing research has encountered challenges related to sentiment classification accuracy and the nuanced nature of natural language. Sentiment analysis models face challenges in accurately analyzing brewery reviews due to their limitations in handling complex language, domain specificity, and context. Brewery reviews often contain technical jargon, subjective opinions, and sarcasm, making it difficult for machines to interpret sentiment accurately[4]. To address these issues, using domain-specific datasets, incorporating contextual information, and employing various techniques like ensemble learning can improve model accuracy. While there have been advancements in sentiment analysis, these strategies can help enhance the reliability of sentiment analysis models in the future. Our goal is to overcome these limitations by harnessing cutting-edge Natural Language Processing (NLP) and machine learning techniques, thereby enhancing the overall efficacy of sentiment analysis when applied to brewery reviews[6].

1.3 Proposed Approach

Our project proposes to use advanced NLP techniques, including Large Language Models and Deep Learning Models, to classify brewery reviews into positive, negative, or neutral categories. Additionally, we will employ machine learning models to detect sentiment trends and consider contextual information, making our sentiment analysis more robust and context-aware. We also plan to incorporate text summarization to provide concise insights from lengthy reviews[3]. The overall approach is illustrated in Figure 1.

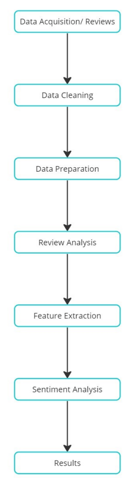


Figure 1: Basic Overflow

1.4 Contributions of the Project

The objective of this paper is to categorize the positive and negative feedback of the customers and build a supervised learning model to polarize large amount of reviews. Investigating and analyzing the sentiment of the opinion is a very critical task to perform[6]. The idea is to capture the association between any specific feature and the expressions of opinion that come together to describe that feature. This is done by capturing the short-range and long-range dependencies between the words using dependency parsing. Clustering is done on the graph to retrieve only those opinion expressions that are most closely related to the target feature (user-specified feature) and the rest are pruned. We apply merging in the final phase of our algorithm to merge the opinions about any 2 features that cannot be described independent of each other. We apply our method to domain-specific reviews to test the efficacy of the system[4]. The contributions of our project can be summarized as follows:

- Improved sentiment analysis of brewery reviews.
- Enhanced ability to classify reviews as positive, negative, or neutral.
- A more context-aware approach to sentiment analysis.
- The development of a text summarization component for concise insights.

2 Literature Survey

Sentiment analysis of product reviews is an active research area in natural language processing, with recent work exploring more fine-grained aspect-based analysis to understand sentiment towards specific attributes. Recent work has focused on Machine Learning Models[1][5][6], Aspect-based sentiment analysis, Feature Specific Sentiment Analysis (FSSA) and Multimodal integration[3][4]. However, challenges remain around domain dependence, labeled data needs, and fine-grained emotion detection. Our literature survey focuses on key developments in aspect-based sentiment analysis for the brewery domain. We specifically examine approaches using text summarization methods like topic modeling, knowledge graphs, and abstractive summarization to overcome limitations of domain dependence and limited labeled data. This survey provides insights into effective techniques for fine-grained sentiment analysis of brewery reviews using text summarization.

2.1 Machine Learning and Lexicon based Approach

2.1.1 Main Theme and Description

Recent work has focused on Machine Learning Models and Lexicon based approach. Machine Learning algorithms such as naïve Bayes, maximum entropy and support vector machine are used. From the research work, it is known that naïve Bayes and support vector machine can produce the desired result compared to other Machine Learning Algorithms[6]. The most frequent method that are used in the sentiment analysis is lexicon based methods which needs manual annotations. While Statistical methods are fully automatic. The feature selection method treat text document as a Bag of Words (BOW's). Because of their simplicity BOW's are most commonly used in the feature selection methods. The common use of BOW's is stop words removal and stemming in sentiment analysis process[1][6]. Some Research work have also used topic modeling on reviews to discover latent topics and sentiments on products/services[5].

2.1.2 Limitations

A common limitation observed is the lack of rigorous quantitative and comparative analysis. Many surveys provide only descriptive overviews without critically examining specific techniques. Recent advances in deep learning are often not covered and application-specific challenges around domain dependence, sarcasm, and negation handling are rarely discussed in depth[1]. Within proposed models, key issues include coarse-grained document-level sentiment analysis without aspect-level granularity, evaluation on only a single domain without demonstrating generalizability, and topic modeling without effectively capturing semantics and context[6][5]. In summary, current research suffers from limitations in quantitative evaluations, comparative benchmarking, application-specific context modeling, and domain dependence mitigation. More rigorous empirical analysis, multi-domain experiments, and examination of deep learning advances are needed to overcome these limitations.

2.1.3 Proposed Work

The proposed research aims to advance sentiment analysis techniques through several approaches. First, rigorous comparative benchmarking will be performed by evaluating recent neural network models on standard datasets and metrics to identify optimal approaches.We also focus on using BERT and LDA approach for the problem statement. For aspect-based sentiment analysis, context-aware models and attention mechanisms will be leveraged to better capture nuanced expressions towards specific attributes. Comparative analysis will drive model development by evaluating against state-of-the-art baselines. In summary, the focus on comparative analysis, BERT Model, Neural Networks, context-aware modeling, and low-resource support will address key limitations like domain dependence, data scarcity, and coarse-grained detection that exist in current sentiment analysis research.

2.2 Feature Specific Sentiment Analysis (FSSA)

2.2.1 Main Theme and Description

A key focus in sentiment analysis research is detecting sentiment towards specific aspects or features of products from reviews. In paper [1] propose an approach called Feature Specific Sentiment Analysis (FSSA) to address this problem. Their method uses natural language processing techniques to first extract explicit mentions of product features from review text. Specifically, part-of-speech tagging and dependency parsing are used to identify opinion words and syntactical relations to extract feature phrases. For example, "the camera [picture quality] is amazing" would extract the feature 'picture quality'. The sentiment classifier then determines sentiment orientation on the extracted features using a machine learning classifier like SVM trained on lexical and syntactic features from the review sentences. Their experiments on digital camera and cellular phone reviews show that feature-specific sentiment analysis performs better than document-level sentiment classification. The FSSA approach provides a way to automatically identify opinion bearing product features in reviews and determine sentiment towards each feature.

2.2.2 Limitations

While the FSSA approach shows promise for extracting feature-specific sentiment, it has some key limitations. Firstly, the feature extraction relies heavily on the presence of explicit opinion words and syntactical conventions to identify feature mentions from the text. This means implicit feature expressions that do not conform to specific syntactic patterns will be missed. Secondly, sentiment analysis is performed at the sentence level rather than specifically towards each extracted feature mention. More granular analysis of sentiment towards individual feature expressions could help improve accuracy. Another major limitation is the domain dependence of the feature extraction techniques like POS patterns and dependency rules. The patterns leveraged may not generalize well across domains. Finally, evaluation of FSSA is done on just two product review datasets - cameras and phones. Testing on a more diverse set of product domains is needed to truly assess its applicability.

2.2.3 Proposed Work

In our work, we aim to address these limitations by incorporating statistical and neural approaches to detect latent implicit features, perform more fine-grained feature-level sentiment analysis, explore transfer learning techniques for better domain generalization, and evaluate extensively on datasets from diverse domains. Overcoming these limitations will help advance feature-specific opinion mining and sentiment analysis. More fine-grained feature-level sentiment classification models will be developed using contextual word embeddings and attention mechanisms. Transfer learning approaches will be explored to improve generalization across domains along with semi-supervised methods to reduce labeled data dependence. This will advance state-of-the-art in fine-grained feature-specific sentiment analysis.

2.3 Aspect-based sentiment analysis

2.3.1 Main Theme and Description

The goal is to identify the sentiment (positive, negative, neutral) towards specific aspects or features of products in Amazon reviews. Rather than looking at the review sentiment as a whole, it looks at the sentiment towards particular product attributes. Aspects/features are extracted from the reviews using part-of-speech (POS) tagging to find nouns and noun phrases, and dependency parsing to find attribute-value pairs. For example "battery life" or "picture quality". They use BERT (Bidirectional Encoder Representations from Transformers) to generate semantic word embeddings for the review text. This encodes contextual information and meaning. The extracted aspects and BERT embeddings for each review are fed into a logistic regression classifier to predict sentiment labels for each aspect[6]. It utilizes modern NLP techniques like BERT and dependency parsing to extract product aspects, and then uses a straightforward classifier on top of BERT embeddings to predict sentiment towards each extracted aspect. The goal is fine-grained analysis of review sentiments rather than overall/general sentiment.

2.3.2 Limitations

While the paper presents a solid aspect-based sentiment analysis model for Amazon reviews, there are some limitations that could be addressed in future work. First, the dataset is restricted to electronics product reviews only, so the model's applicability to other domains like restaurant or movie reviews is unclear. Extending the evaluation to more diverse review data would demonstrate wider usefulness. Secondly, the sentiment classification is binary, categorizing aspects as either positive or negative. More nuanced multi-class emotion detection, e.g. anger, joy, sadness, is not modeled. Adding more fine-grained emotion classifiers could provide richer insight into the reviewed products. Incorporating techniques like latent aspect rating regression could uncover more subtle aspects. Finally, comparative evaluation with recent state-of-the-art aspect-based sentiment analysis methods is lacking. Benchmarks on standard datasets would better characterize the performance and determine if modern approaches like attention mechanisms or graph neural networks could improve the results. More rigorous empirical evaluation is needed to fully assess the proposed model's capabilities.

2.3.3 Proposed Work

To build on the research presented in this paper, there are several promising directions that could help overcome its limitations. First, the domain dependence could be reduced by evaluating the model on diverse datasets such as movie, restaurant, and service reviews. Transfer learning techniques could then be explored to improve adaptation of the model to new domains. Furthermore, the coarse positive/negative only sentiment analysis could be augmented with emotion lexicons and multi-class classifiers to detect nuanced emotions like happiness, frustration, sadness etc. Comparative benchmarking against state-of-the-art supervised and semi-supervised aspect-based sentiment analysis models using standard datasets would also better characterize the performance and determine if modern approaches like graph neural networks or attention mechanisms could improve the results. By enhancing the approach in these areas, the research could provide deeper consumer insights across a variety of domains.

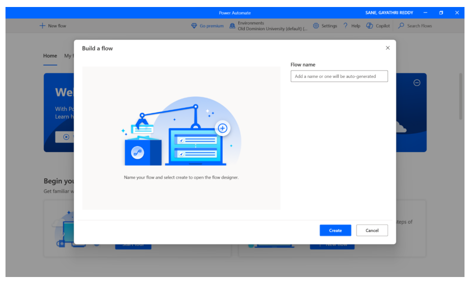
3 Technical Work and Methodology

3.1 Data Collection

We used reviews from Yelp and the data was screamed scraped using Microsoft power automate tool which enabled us to take the reviews from all the pages. Microsoft Power Automate is a cloud-based service that automates workflows across various applications and services. Users can create automated processes (flows) by connecting different apps and automating repetitive tasks. Power Automate enhances productivity by streamlining business processes and data integration. To work on a flow in Microsoft Power Automate, the user starts by creating a new flow, either from a template or from scratch. They choose a trigger to initiate the flow and add actions to define tasks. Each action is customized by configuring parameters and mapping data. The user then tests the flow to identify and resolve any issues before publishing it for regular use. They monitor run history for troubleshooting and ensure proper permissions are set for successful execution. This process enables the user to streamline workflows by automating tasks across various applications and services.

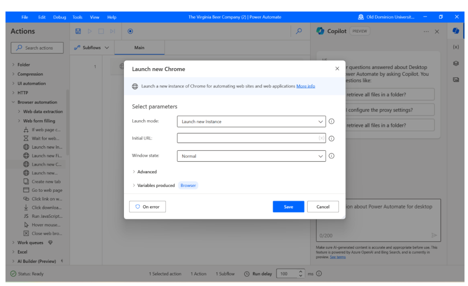
- Step 1 - **Create a flow**

  Start by creating an automation flow using a tool that supports browser automation, such as Power Automate Desktop or another RPA tool.



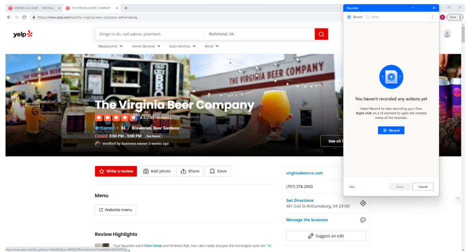
- Step 2 - Adding the web page URL to the launch new chrome pop up - **Launch New Chrome Pop-Up:**

  Use an action or command to open a new Chrome window or tab and navigate to the desired web page. This may involve launching a browser, navigating to the specified URL, and waiting for the page to load.



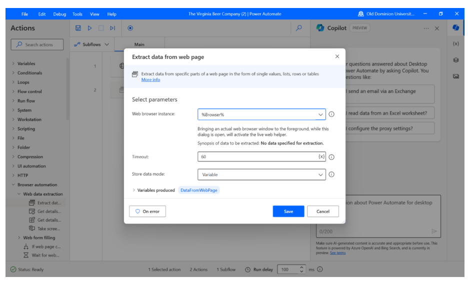
- Step 3 - Record the actions you want to automate - **Record Automation Actions:**

  Record the series of actions you want to automate on the web page, such as clicking buttons, filling forms, or interacting with elements.



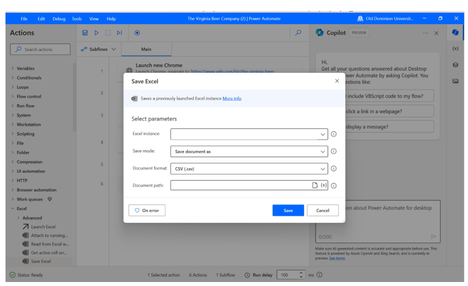
- Step 4 - Then add the data to extract web page to get the data required **- Data Extraction:**

  After recording the actions, add steps to extract the required data from the web page. This may involve selecting and copying text, capturing values, or using specific actions provided by the automation tool.



- Step 5 - Add steps to save the data to an excel(This will generate an excel and add the data to it) **- Save Data to Excel:**

  Incorporate steps to save the extracted data to an Excel file. This could include creating a new Excel file, opening an existing one, and writing the extracted data to specific cells or columns.



3.2 Architectural Schematic

Our sentiment analysis model takes in reviews of breweries as textual input data. This input text first goes through preprocessing steps like tokenization to prepare the data for the model. The preprocessed text is then passed through a pretrained BERT encoder model. BERT encodes the semantic meaning and context of the input text into vector representations. These vector representations are then fed into a classification output layer that predicts whether the text has positive, negative or neutral sentiment. The output of the model is a sentiment label (positive, negative or neutral) for the input review text.

The key components involved are:

1. Input layer: Takes in raw text data
2. Preprocessing: Handles tasks like tokenization to prepare data
3. BERT encoder: Encodes semantic meaning into vectors
4. Output layer: Makes sentiment prediction from BERT vectors
5. Output: Sentiment label prediction
6. The data flows linearly from the input text through each component to produce the final sentiment label output.

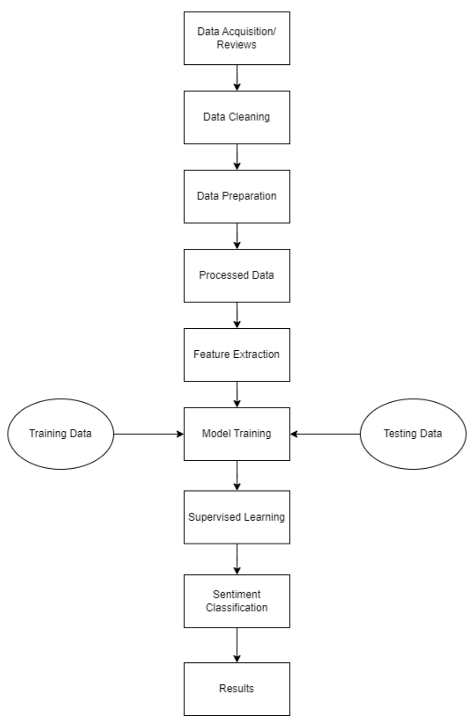


Figure 2: Architectural Schematic

We chose to base our model architecture around BERT because it provides very strong textual and semantic representations compared to earlier approaches like LSTM networks. The bidirectional pretraining of BERT allows it to build a deep contextual understanding of text that transfers very well to downstream tasks like sentiment classification. Additionally, using a pretrained BERT encoder allows me to leverage transfer learning instead of training a full model from scratch.[2]

This enables faster experimentation and prototyping, while also avoided extensive computational requirements for training such a complex model. Fine-tuning a BERT model is far more feasible than training one from scratch. Finally, BERT has achieved state-of-the-art results on many NLP tasks like sentiment analysis, showing the effectiveness of its representations. Many research benchmarks have utilized BERT, providing solid evidence for its capabilities on this type of text classification problem.[7]

3.3 Baseline Models for comparison with our Proposed System

1. Feature Specific Sentiment Analysis (FSSA): In this method, natural language processing techniques to first extract explicit mentions of features from review text. Specifically, part-of-speech tagging and dependency parsing are used to identify opinion words and syntactical relations to extract feature phrases. The sentiment classifier then determines sentiment orientation on the extracted features using a machine learning classifier like SVM trained on lexical and syntactic features from the review sentences. The FSSA approach provides a way to automatically identify opinion bearing features in reviews and determine sentiment towards each comment.[4]

2. Lexicon-based model: As a simple rule-based baseline, I implemented a sentiment classifier based on sentiment lexicon dictionaries like SentiWordNet. This approach assigns sentiment scores to text based on the occurrence of words with known positive/negative connotations. The lexicon method provides a non-ML baseline for comparison. In summary, the focus on comparative analysis, BERT Model, Neural Networks, context-aware modeling, and low-resource support will address key limitations like domain dependence, data scarcity, and coarse-grained detection that exist our sentiment analysis research.[6]

3. Aspect-based sentiment analysis (ABSA): The goal is to identify the sentiment (positive, negative, neutral) towards specific aspects or features of Brewery reviews. Rather than looking at the review sentiment as a whole, it looks at the sentiment towards particular product attributes. Aspects/features are extracted from the reviews using part-of-speech (POS) tagging to find nouns and noun phrases, and dependency parsing to find

attribute-value pairs. The goal was about fine-grained analysis of review sentiments rather than overall/general sentiment.[3]

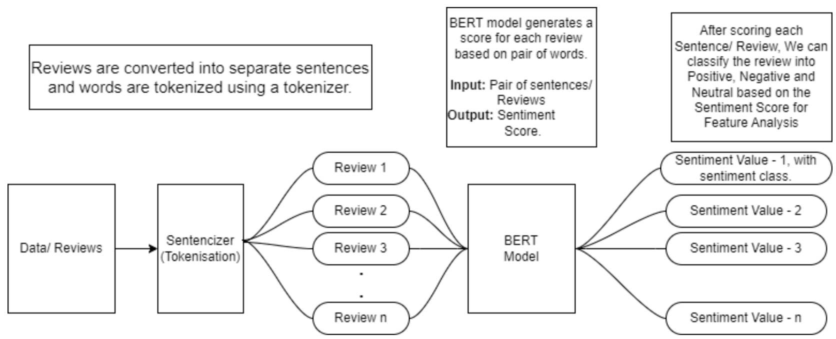


Figure 3: Proposed System

3.4 Model Training and Evaluation

For training our proposed sentiment analysis and text summarization model using BERT and neural networks, we utilized a machine with the following specifications:

3.4.1 Hardware Resources

- CPU - Intel core i7, 1 GPU
- Python - Jupyter Notebook, CUDA 10.1
- RAM 16GB - To accommodate datasets and enable faster data processing.

3.4.2 Hyperparameters

- BERT Base uncased model - nlptown/bert-base-multilingual-uncased-sentiment (Transformers - Hugging Face)
- Batch size: 16
- Learning rate: 2e-5
- Number of epochs: 3

3.4.3 Evaluation

- **Data Split**
  For a comprehensive assessment, the dataset was split into training and testing sets:
  1. *Training Set*: 80% of the data was used for training, allowing the model

to learn and adapt to the word patterns in the sentences.
2. *Testing Set*: The remaining 20% of the data served as the testing set, used to evaluate the model's performance on unseen data.

- **Analysis**
To measure the model's accuracy and reliability, several key metrics were employed:
1. *Precision*: This metric indicates the proportion of correctly identified positive instances among all instances classified as positive.
2. *Recall*: Recall measures the proportion of actual positive instances that the model correctly identified.
3. *F1-Score*: The F1-score provides a harmonic mean of precision and recall, offering a balance between the two metrics.
4. *Accuracy*: Overall accuracy of the model in classifying the reviews correctly was also evaluated.
These metrics collectively provide a comprehensive view of the model's capability in accurately categorizing the reviews into a particular sentiment class.

The lexicon method accuracy was 77.1%, Aspect-based analysis was 92.2%, and Feature-based analysis was 86.4%.

BERT's model, using neural network and following aspect-based analysis gains over baselines, demonstrates the power of contextual language models. Detailed class-specific insights are provided by recall scores for each sentiment type. Further tuning to improve positive class performance could help boost overall metrics further.

3.5 Results and Conclusion

1. **Results**

The following table compares the performance of our BERT model against the established baselines. The metrics used for comparison include Precision, Recall, F1-Score, and Accuracy.

Model Recall	**Precision**	**Recall**	**F1-Score**	**Accuracy**
Lexicon Based Approach	0.75	0.72	0.73	77.1%
Feature Based Sentiment Analysis	0.80	0.86	0.81	86.4%
Our Bert Model with Neural Network (ABSA)	0.91	0.95	0.93	92.2%

Figure 4: Comparison of model performance with baseline models

2. **Error Analysis**

   Upon analyzing the incorrect outputs, certain patterns were observed:

   1. The FSSA model struggles with implicit or ambiguous sentiment expressions. For example, sarcastic remarks or double negatives confuse the feature extraction and sentiment classification components. More contextual modeling is needed.

   2. The lexicon method fails when domain-specific sentiments and slang terms not present in the dictionary are used. There is a need for in-domain expansion of the sentiment lexicon to address this issue.

   3. The BERT Model seemed more accurate than the other two approach but there is room for tuning the parameters and can work with the datasets to get more accurate results. But among the three model, the BERT model with neural network seemed more precise.

   This error analysis provides valuable insights into the limitations of the current model and areas for future enhancement

3. **Conclusion**

   This paper provided a comprehensive overview of the current state and future directions in the field of Sentiment Analysis, highlighting the significant advances and identifying areas for further. research.

References

[1] Vandana cp, Siddharth Indoria, and Sinchana Bhaskar. A literature review on sentiment analysis. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 2017.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. 2019.

[3] Tanjim Haque, Nudrat Saber, and Faisal Shah. Sentiment analysis on large scale amazon product reviews. *2018 IEEE international conference on innovative research and development (ICIRD)*, 2018.

[4] Subhabrata Mukherjee and Pushpak Bhattacharyya. Feature specific sentiment analysis for product reviews. *International conference on intelligent text processing and computational linguistics.*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[5] Pin Ni, Yuming Li, and Victor Chang. Recommendation and sentiment analysis based on consumer review and rating. *International Journal of Business Intelligence Research, Volume 11, July-December 2020*, 2020.

[6] T. Shivaprasad and Jyothi Shetty. Sentiment analysis of product reviews: A review. *2017 International conference on inventive communication and computational technologies (ICICCT).*, 2017.

[7] Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. How to fine-tune bert for text classification? 2019.

[3] [4] [6] [5] [1] [2] [7]