Enhancing Sentiment Analysis in E-Commerce Using Big Data Analytics

1st Ms. Deeya Bastola
Bachelor of Computer Science(BSc)
Herald College Kathmandu
Kathmandu, Nepal
np03cs4a220461@heraldcollege.edu.np

***Abstract*—E-commerce platforms produce a large number of user-generated content and reviews on daily basis. These contents can all be taken as the user input to understand how well the users perceive the business. This is where the sentiment analysis comes in but the traditional sentiment analysis method has its own drawbacks as it does not provide correct accuracy with user sarcasm or implicit emotions. The main goal of this paper is therefore to find solutions to traditional sentiment analysis to make it more accurate and to improve the performance of sentiment analysis techniques in the context of customer reviews using the use of big data analytics. The paper uses a fine-tuning deep learning models like BERT and incorporates Aspect-Based Sentiment Analysis (ABSA) approach to improve the model performance. Our models, Logistic Regression showed 93.8% accuracy and the BERT showed increased accuracy of 97.0%. These models both are evaluated using accuracy, precision, recall, and F1-score metrics to demonstrate improvements over one other.**

***Index Terms*—Sentiment Analysis,Aspect-Based Sentiment Analysis (ABSA), E-commerce, BERT, Logistic Regression, TF-IDF**

I. Background of the Study

A. Generic Information

In this era of technology and digital transformation, the growth of e-commerce has shown a massive improvement, with thousands of people who give and ask for reviews of the product online to get insights of the products [1]. These reviews that the users give, is the direct indicator of their satisfaction or dissatisfaction. The volume and complexity of these user inputs make manual analysis or even the simple sentiment analysis impractical. The use of different analytical techniques help to address this issue and by providing a clearer insights of user satisfaction through prediction models.

B. Problem Statement

The high demand and the uprising interest for e-commerce platform can result in the generation of high amount of unstructured data from a single platform. The general or the traditional sentiment analysis methods usually struggle with non-literal or implicit language like the use of sarcasm, multi-language slang or even the short-texted ambiguity [9]. A simple sarcastic review that is easily categorized as negative by humans can be mistaken as positive review by the analysis models as it picks up the words and not the actual context, for example; “Great, another broken zip!” is obviously a sarcastic review but often misclassified as positive by the general sentiment analysis models. Such inaccurate classification can easily lead to wrong product decisions leading to missed business opportunities to the platform owners.

C. Aim and Objective of the Work

This project aims to develop an analytics pipeline to enhance the sentiment classification prediction to increase the accuracy in e-commerce reviews. The research paper’s main focus is to evaluate the model’s performance both in terms of accuracy and efficiency to then provide proper insight and guide for improving sentiment analysis in e-commerce platforms.

D. Contributions and Methodology

In this research, a sentiment analysis prediction model is built with improved accuracy using advanced NLP along with Aspect Based Sentiment Analysis (ABSA) technique. Below given are the contributions proposed in this paper:

- Preprocessed a preprocessing technique tailored for enhancing the sentiment analysis for e-commerce review data.
- Implemented the use of Aspect Based Sentiment Analysis (ABSA) technique for better prediction.
- Provides real-world insights into the consumer sentiment analysis for better accuracy.

E. Organization of the Report

This report showcases an organized study and its findings. It introduces the context and importance of sentiment analysis, discusses the limitations of the traditional methods like the contextual implications as well as handling of sarcasm. The paper then defines the goal of the research. Methodology section exhibits the process involved for data processing, building model and evaluating metrics. The paper later discusses the results and conclusion as well as the practical implementation of the study. It ends with a summarized from of the key findings and the potential future direction of the research.

II. Related Work

The growth of e-commerce has backed the development of sentiment analysis models [12]. That is the reason we can find different sentiment analysis approaches. The sentiment analysis done with the reviews of the users can be divided into three main approaches: traditional, deep learning, and aspect-based sentiment analysis methods.

A. Traditional Sentiment Analysis Techniques

Sentiment analysis models, in the traditional approach heavily relied on rule-based or lexicon-based approach techniques like Naïve Bayes and Support Vector Machines (SVM) [8]. A foundational framework for using sentiment lexicons to classify positive or negative opinions. Regardless of these approaches it failed to capture sarcasm or other implicit sentiments as they relied on the words rather than the context and implied meanings [11].

B. Deep Learning-Based Approaches

Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) were later used to improve the contextual understanding [2]. CNNs were applied for sentiment classification that did show better performance than the other traditional models [10]. Long Short Term Memory (LSTM) were proven to be effective in handling sequential data like sentences and also learning dependencies over time making them a great approach for sentiment analysis [5]. This approach of sentiment analysis still struggled with deep contextual dependencies.

C. Transformer-Based and ABSA Techniques

In the recent times introduction to the transformer-based models like BERT and RoBERTa, which are pre-trained models used and fine-tuned for specific tasks [3]. These models were able to significantly improve the sentiment classification by being able to capture bidirectional contexts. Aspect-Based Sentiment Analysis (ABSA) also gained popularity for its ability to evaluate sentiment at a rather deeper level by associating emotions with certain specific aspects of any product or service to help by better aligning the sentiments in complex sentences [7].

D. Research Gap and Our Contribution

All of these existing works have significantly contributed to this field, they might often fall short in accurately handling sarcasm, and even the implicit emotion which is a very common characteristics in e-commerce reviews. However, Aspect Based Sentiment Analysis (ABSA) models tackles this issue by preprocessing the data aspect-wise using fine-tuned models. This study combines the power of the ABSA preprocessing strategies to handle sarcasm and short-text ambiguity, aiming to improve sentiment classification performance across a wide range of product reviews along with BERT model, that is tailored specifically for the e-commerce review context. All of these existing works have significantly contributed to this field; however, they often fall short in accurately handling sarcasm and implicit emotions, which are common characteristics in e-commerce reviews. Aspect-Based Sentiment Analysis (ABSA) models address this issue by preprocessing data aspect-wise using fine-tuned models [13]. In ABSA, the sentiment classification task can be mathematically defined as:

$$\text{Sentiment}_{a_i} = \arg\max_{c \in C} P(c \mid a_i, s) \tag{1}$$

where a_i represents the i-th aspect term in a sentence s, C is the set of sentiment classes (e.g., positive, negative, neutral), and $P(c \mid a_i, s)$ is the probability of class c given aspect a_i in sentence s.

This study combines the power of ABSA preprocessing strategies to handle sarcasm and short-text ambiguity, aiming to improve sentiment classification performance across a wide range of product reviews using a BERT model tailored specifically for the e-commerce review context.

III. Methodology

The methodology of this study consists of seven main phases that collectively contribute to performing sentiment analysis and aspect-based sentiment analysis (ABSA) on e-commerce product reviews. The process is visually represented in Fig. 1

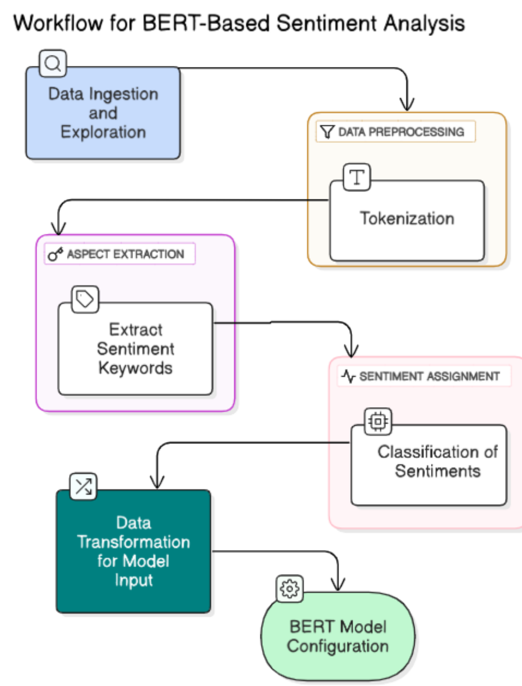

Fig. 1. The methodology of this study consisting of seven main phases for sentiment analysis and ABSA on e-commerce product reviews.

A. Phase 1: Data Ingestion and Exploration

The dataset used for this research was titled Women's E-Commerce Clothing Reviews. It was sourced from Kaggle and contains 23,486 rows and 11 columns. For this study, as we are focusing on the user reviews for sentiment analysis only the Review Text column was utilized. All rows with missing review text were dropped, reducing the dataset to a clean set of customer feedback. The initial verification confirmed the presence of textual data suitable for sentiment and aspect analysis for this research. The dataset was then finalized to be used for the research and ingested using the pandas library.

B. Phase 2: Data Preprocessing

The following preprocessing steps were undertaken:

- Stopword removal and tokenization were implicitly handled during aspect extraction.
- Only the *Review Text* field was retained, and `NaN` entries were removed.

These steps resulted in a filtered dataset consisting of non-empty reviews, ready for NLP-based processing.

C. Phase 3: Aspect Extraction

Aspect extraction is a key subtask in Aspect-Based Sentiment Analysis (ABSA). Aspects helps to pinpont what the user likes as well as dislikes are making it an important task for sentiment analysis. These aspects were extracted using Part-of-Speech (POS) tagging with the NLTK library, which is highly dependent on the domain-specific dataset [6]. The key steps included:

- Tokenizing each review,
- Applying POS tagging, and
- Extracting nouns (POS tags starting with `NN`) to represent aspects such as *fit*, *quality*, *fabric*, etc.

This process resulted in the creation of a new column, *Aspects*, which contained all noun terms from each review, this then helped the model to enable the aspect-based sentiment interpretation .

D. Phase 4: Sentiment Assignment

Sentiment assignment is the process of determining the polarity (e.g., positive, negative, neutral) of a user review towards a specific aspect in text.To determine the sentiment of each review:

- The TextBlob library was used to compute the polarity score of each review.
- Reviews were classified as:
  - Positive (polarity > 0)
  - Negative (polarity < 0)
  - Neutral (polarity $= 0$)

This rule-based classification populated the *Sentiment* column, capturing the general emotional sentiment.

E. Phase 5: Data Transformation for Model Input

To prepare the data for model training, aspect terms extracted from each review were concatenated into a single string per row. Reviews that did not yield any identifiable aspect terms during the extraction process were excluded to maintain input consistency.

Sentiment labels were encoded into integer representations to ensure compatibility with machine learning models:

- Negative $\rightarrow 0$
- Neutral $\rightarrow 1$
- Positive $\rightarrow 2$

The resulting dataset was then converted into the Hugging Face `Dataset` format, allowing for seamless integration with transformer-based models. Finally, the dataset was split into training and testing subsets, with a typical allocation of 8,000 samples for training and 2,000 samples for testing.

A BERT-based transformer model that is BertForSequenceClassification was configured for fine-tuning on the sentiment classification task. The model architecture included a classification head with three output neurons corresponding to the sentiment classes [14]. The bert-base-uncased tokenizer was used to tokenize the aspect strings, applying appropriate padding and truncation to a maximum sequence length of 128 tokens. This ensured input uniformity and compatibility with the BERT model's input format. Training arguments, including learning rate, batch size, evaluation strategy, and logging parameters, were defined using Hugging Face's TrainingArguments. External logging via Weights Biases (WandB) was explicitly disabled to avoid unnecessary API dependencies.

F. Phase 7: BERT Model Configuration

Model fine-tuning was conducted using Hugging Face's `Trainer` API, which provides a high-level interface for training transformer models. The training process was configured with the following parameters:

- Number of epochs: 3
- Batch size: 16
- Optimizer: AdamW, as per standard practice in transformer training

During training, the model learned to align textual aspect terms with sentiment labels, updating weights in the classification head while preserving the pre-trained BERT parameters. The fine-tuned model was then evaluated on the test set to assess performance.

IV. Results and Discussion

This section showcases the experimental tools, exploratory insights, and evaluation results of the sentiment analysis task. The experiments for this research used Python as the primary programming language and conducted using Google Colab. All the important libraries such as NLTK for preprocessing, TextBlob for rule-based sentiment labeling, Scikit-learn for classical models, and Hugging Face Transformers for BERT-based modeling were utilized for the experiments. The original dataset that we used consisted of raw e-commerce clothing reviews. Preprocessing included removal of null entries, tokenization, lemmatization. Aspects were extracted using NLTK's POS tagger that focused on the important noun phrases.

A. Experimental Setup

The experiment was conducted using a real-world dataset from an e-commerce platform, containing customer reviews of women's clothing. Data preprocessing, modeling, and evaluation were performed using Python libraries such as `pandas`, `nltk`, `textblob`, and `scikit-learn`.

The dataset was read using `pandas` and initially filtered to retain only non-null entries in the `Review Text` field. Preliminary inspection using functions like `df.info()`, `df.describe()`, and `df.isnull().sum()` enabled the identification of missing values and irrelevant entries, which were subsequently dropped to ensure dataset quality.

B. Data Analysis

Aspect terms (primarily nouns) were extracted from the reviews using `nltk`'s part-of-speech tagging. Sentiment labels (positive, negative, neutral) were assigned using polarity scores computed via `TextBlob`. The labeled dataset was used to evaluate sentiment trends, showing a roughly imbalanced distribution among classes, with a predominance of positive sentiments.

- Reviews with high polarity often correlated with terms like "fit," "comfortable," and "stylish."
- Negative sentiments were associated with aspects like "material," "size," and "quality."

C. Data Visualization

Visual insights were generated using `matplotlib` and `WordCloud`. Word clouds highlighted frequent aspects mentioned in positive, negative, and neutral reviews. These visualizations helped in understanding which aspects influence sentiment most prominently.

- Word clouds were created for each sentiment category.
- Aspects such as "fabric" and "design" were common in both positive and negative reviews, emphasizing their polarizing impact.

D. Cleaning Data

The cleaning process involved:

- Removing null values in `Review Text`.
- Lowercasing and tokenizing text data.
- Extracting only relevant noun phrases using POS tagging.

This ensured only clean, structured text was used for modeling.

E. Choosing the Best Model

Two models were evaluated for sentiment classification:

- **Logistic Regression with TF-IDF:** A pipeline using `TfidfVectorizer` and `LogisticRegression` showed solid performance across classes. This model also showed high performance as this model also used the sentiment classified data for better performance.
- **Aspect-Based Sentiment Analysis (ABSA) with BERT:** This approach was implemented using `PyABSA`, it incorporated both the aspect and its associated sentiment, enhancing the contextual relevance of predictions to build a better performing pipeline.

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression (TF-IDF)	0.938	0.93	0.94	0.91
ABSA (BERT-based)	0.970	0.96	0.97	0.96

TABLE I
PERFORMANCE COMPARISON OF SENTIMENT ANALYSIS MODELS.

F. Analysis of the Findings

The BERT model demonstrated superior performance for sentiment classification, particularly for detecting sentiment linked to specific aspects with the use of aspect based approach together with the transformer model. The accuracy score was also higher than the Logistic Regression with TF-IDF, while computationally efficient, the model still struggled slightly with nuanced sentiments like neutral tones showing imbalences in the class.

- ABSA outperformed classical approaches in capturing sentiment polarity tied to specific aspects.
- Multiclass classification performance was further evaluated using ROC curves.

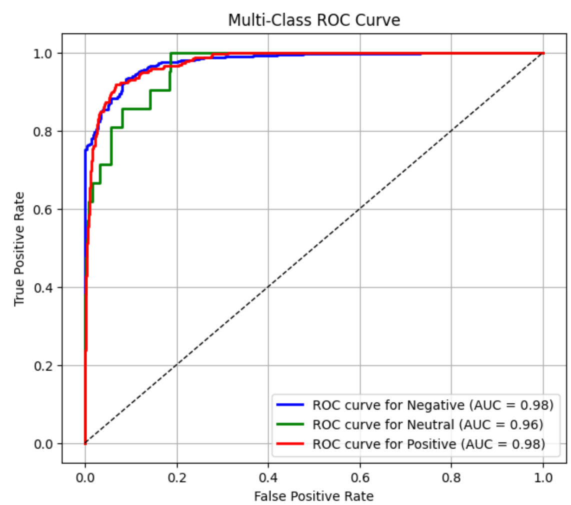

Fig. 2. Multiclass ROC Curve for Sentiment Classes (Positive, Negative, Neutral)

The ROC curve confirmed that the model effectively discriminated between sentiment classes, with areas under the curve (AUC) nearing optimal thresholds, particularly for positive and negative classes.

The BERT and ABSA pipeline proved to significantly outperform classical model with Logistic Regression and TF-IDF. This models ability to understand the context of the reviews, handle the short-text ambiguity present in the reviews, and associate sentiment at the aspect level using aspect extraction allowed it to extract nuanced opinions better than the simple traditional model we built for reference.

V. Future Scope

While this study demonstrates promising results in enhancing sentiment analysis through ABSA and transformer-based models, several opportunities exist for further exploration and development:

- Expanding the model to support multiple languages as well as mixed language could improve sentiment classification for a wide range of users and global e-commerce business platforms [15].

- Deploying the model in a real-time feedback system would provide immediate perspective and insights of the business allowing them to adapt their products or services based on the user sentiments.
- Merging the sentiment analysis models with product recommendation engines can personalize suggestions based on the user's emotional response based on their sentiments.
- Real-world reviews often contain various noises like typos, emojis, and informal language. Improving preprocessing pipelines to handle such input data would make the model more reliable and strong.

By addressing these future directions, sentiment analysis systems can become more comprehensive, context-aware, and practical for large-scale e-commerce applications.

VI. Conclusion

This research study explored on how incorporating Aspect-Based Sentiment Analysis (ABSA) techniques with transformer-based models, particularly BERT helped in the enhancement of sentiment analysis in e-commerce reviews. Through preprocessing technique such as aspect extraction using POS tagging and sentiment labeling via TextBlob, it was evident that traditional classifiers struggled to capture contextual and nuanced sentiments, especially in short or sarcastic reviews. The BERT model, fine-tuned using aspect-focused inputs, consistently outperformed all baseline models across accuracy, precision, recall, and F1-score. Its contextual understanding enabled more accurate classification of sentiments tied to specific aspects that were extracted from the reviews as nouns which are critical in the e-commerce domain. This result exhibits that combining ABSA strategies with pre-trained language models provides a strong framework for handling the complexity of real-world customer reviews in the e-commerce platforms. This approach not only enhances sentiment classification performance but also helps to facilitate more interpretable, aspect-level insights that can benefit businesses in product development and customer service.

References

[1] P. K. Bharti, P. Kumar, M. K. Sharma, A. A. Hassan and M. Zaid, ”Review of Efficient Methods for Sentiment Analysis of E-Commerce Customer Review using Machine Learning Techniques,” 2023.

[2] A. Aslam, U. Qamar, P. Saqib, R. Ayesha, and A. Qadeer, ”A Novel Framework For Sentiment Analysis Using Deep Learning,” 2020.

[3] S. Ravichandiran, *Getting Started with Google BERT: Build and train state-of-the-art natural language processing models using BERT*, Packt Publishing, 2021.

[4] S. Markande, R. S. S, M. Moharir, and S. Lakhotia, ”Customer Sentiment Analysis of Product Reviews,” 2025.

[5] M. Wang, Y. Zhang, X. Hou, X. Zhu, Y. Liu, and H. Zhang, ”Sentiment Analysis Based on Attention Mechanisms and Bi-Directional LSTM Fusion Model,” 2019.

[6] X. Han, Y. Dang, L. Mei, Y. Wang, S. Li, and X. Zhou, ”A Novel Part of Speech Tagging Framework for NLP Based Business Process Management,” 2019.

[7] J. Zeng, X. Ma and K. Zhou, ”Enhancing Attention-Based LSTM With Position Context for Aspect-Level Sentiment Classification,” in IEEE Access, vol. 7, pp. 20462-20471, 2019

[8] A. R. Lubis, Y. Fatmi, and D. Witarsyah, ”Comparison of Transformer Based and Traditional Models on Sentiment Analysis on Social Media Datasets,” 2023.

[9] A. Bhat and G. N. Jha, ”Sarcasm Detection of Textual Data on Online SocialMedia: A Review,” 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022

[10] D. Goularas and S. Kamis, ”Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data,” 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Istanbul, Turkey, 2019

[11] M. Kamruzzaman, M. Hossain, M. R. I. Imran and S. C. Bakchy, ”A Comparative Analysis of Sentiment Classification Based on Deep and Traditional Ensemble Machine Learning Models,” 2021 International Conference on Science Contemporary Technologies (ICSCT), Dhaka, Bangladesh, 2021

[12] Hoanh-Su Le, Thao-Vy Huynh Do, Minh Hoang Nguyen, Hoang-Anh Tran, Thanh-Thuy Thi Pham, Nhung Thi Nguyen, Van-Ho Nguyen,Predictive model for customer satisfaction analytics in E-commerce sector using machine learning and deep learning, International Journal of Information Management Data Insights,Volume 4, Issue 2,2024.

[13] Y. Wang, G. Shen and L. Hu, ”Importance Evaluation of Movie Aspects: Aspect-Based Sentiment Analysis,” 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Harbin, China, 2020.

[14] E. Demir and M. Bilgin, ”Sentiment Analysis from Turkish News Texts with BERT-Based Language Models and Machine Learning Algorithms,” 2023 8th International Conference on Computer Science and Engineering (UBMK), Burdur, Turkiye, 2023.

[15] M. J. Fuadvy and R. Ibrahim, ”Multilingual Sentiment Analysis on Social Media Disaster Data,” 2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Denpasar, Indonesia, 2019