Aspect-Based Sentiment Analysis with Sarcasm Detection on Amazon Product Reviews Using Classical and Transformer-Based Models

Deepanshu Joshi
MSc. In Computing (Majors in DA)
Dublin City University
Dublin, Ireland
deepanshuganesh.joshi2@mail.dcu.ie

Madhumathi Lingala
MSc. In Computing (Majors in DA)
Dublin City University
Dublin, Ireland
madhumathi.lingala2@mail.dcu.ie

Prof. Yalemisew Abgaz
School Of Computing
Dublin City University
Dublin, Ireland
Yalemisewm.abgaz@dcu.ie

Abstract- As online shopping continues to grow, customer reviews have become an important way to gather product feedback. However, many sentiment analysis tools still struggle to understand complex language especially when sarcasm or mixed opinions are involved. This study proposes a method that combines aspect-based sentiment analysis with sarcasm detection to improve how reviews are interpreted. The approach was applied to Amazon product reviews from both balanced and unbalanced datasets. Adding sarcasm detection made a clear difference—it boosted accuracy, particularly when the reviews were grouped simply as either positive or negative. By merging aspect-based analysis with sarcasm detection, we can better reflect genuine customer sentiment and apply this approach across a range of review types.

Index Terms- Sentiment Analysis, Sarcasm Detection, Aspect-Based Sentiment Analysis (ABSA), Customer Reviews, Machine Learning

I. INTRODUCTION

A. Background and Motivation

In practical applications like recommendation systems, brand monitoring, and automated customer support, misinterpreting sentiment—especially sarcasm—can result in flawed business decisions. Sarcasm is challenging because it often contradicts the literal meaning of words. Take a review like "Just perfect, stopped working in a day." On the surface, it sounds positive, but the real meaning is clearly negative. Systems that depend only on keywords might miss that. While aspect-based sentiment analysis (ABSA) helps by looking at specific product features, sarcasm still makes things tricky. That's why researchers are now combining ABSA with sarcasm detection to better understand what people truly think.This study looks at how combining these two methods makes sentiment analysis more accurate and meaningful in real-world reviews

B. Importance of Sentiment and Sarcasm Detection

In everyday use—like recommending products or handling customer support—misreading what a customer really feels, especially when sarcasm is involved, can lead to bad outcomes. Sarcasm twists the actual meaning of words, which can throw off systems that just look for certain keywords. Aspect-based sentiment analysis offers more accuracy by breaking down opinions based on features like battery life or delivery time instead of treating the whole review as one opinion. Still, sarcasm makes things tricky by hiding true feelings. That's why recent studies are looking into how to mix sarcasm detection with aspect-based methods. This approach aims to improve our understanding of subtle customer feelings.

C. Research Questions

This research addresses the following key questions:

- How does sarcasm affect how accurately we can detect sentiment in reviews?
- Do newer models like DistilBERT do a better job of figuring out how customers really feel compared to older ones
- Does cleaning the text or balancing the number of reviews for each sentiment affect how well the model works?

D. Objectives of the Study

- To compare how well traditional machine learning models and transformer-based models like DistilBERT perform when classifying sentiment, especially in the presence of sarcasm.
- To investigate the role of sarcasm in misclassifying sentiment, especially in negative and ambiguous reviews.
- To construct and compare benchmark datasets (balanced 3-class, binary, and full unbalanced) to assess model robustness under different class distributions.
- To identify the optimal configuration for sentiment analysis with the highest predictive accuracy.

E. Overview of Methodology

The study uses 50,000 product reviews from five Amazon categories: Electronics, Books, Home & Kitchen, Cell Phones & Accessories, and Sports & Outdoors. During preprocessing, emojis were converted to text, irrelevant tokens were removed, and sentiment labels were assigned based on star ratings. Aspect terms and polarities were extracted using a pretrained ABSA model (yangheng/deberta-v3-base-absa-v1.1). Sarcasm detection was applied with a transformer-based classifier (helinivan/english-sarcasm detector), and sarcasm annotations were added at both review and aspect levels. Three sentiment-labeled datasets were created: a balanced 3-class dataset, a binary sentiment dataset, and an unbalanced dataset with computed class weights. Classical models (Logistic Regression, SVM, Naive Bayes, Random Forest) were trained with TF-IDF features, while DistilBERT was fine-tuned on all datasets.Model performance was evaluated using accuracy, precision, recall, and F1-score, with a focus on how well it handled sarcasm-aware sentiment classification

II. LITERATURE REVIEW

As online shopping continues to grow, making sense of customer feedback through reviews is becoming more essential. In the beginning, researchers relied on traditional machine learning models such as Naive Bayes and Support Vector Machines (SVMs) to analyse this feedback. One such study [1] applied these models to Amazon reviews, finding they worked well for general sentiment but struggled with sarcasm and mixed opinions. As the need for more accurate models increased, researchers began using newer techniques like deep learning and transformer models. A recent study [2] found that models such as BERT and RoBERTa did better because they could grasp context and subtle meanings in language, which older models often missed. A follow-up study [3] looked at how sentiment analysis has developed over time. It found that while newer models are generally more accurate, they often have trouble picking up on sarcasm or dealing with reviews that express emotions in subtle or complicated ways.. Sarcasm detection remains a challenge.

One study [4] focused on detecting sarcasm in online comments, showing that adding conversation context helped models recognize sarcasm, but it was treated as a separate task from sentiment analysis. A follow-up study [5] combined machine learning and deep learning for better sarcasm detection. While it improved recognition, it didn't address real-time processing or aspect-specific opinions, meaning it still couldn't fully capture sentiment on individual product features. Even transformer models like BERT face limitations.

A study [6] found BERT to be the most accurate for general sentiment analysis but didn't delve into specific product features or sarcasm. The Amazon Reviews 2023 dataset [7] is commonly used in such studies because it offers a wide variety of customer reviews. It serves as a solid base for bringing together aspect-based sentiment analysis and sarcasm detection, helping to make predictions that are both accurate and useful in real-world scenarios

III. RESEARCH METHODOLOGY

This section presents the overall research framework adopted to perform aspect-based sentiment analysis with sarcasm detection on Amazon product.

A. Overview of Research Workflow

To carry out this study, a pipeline that handles customer reviews in a clear and organized way was designed step-by-step. The overall workflow is shown in Figure 1(Appendix). It starts with collecting and cleaning the review data. Then, a pretrained model is used to pull out specific product aspects mentioned in the text. After that, a sarcasm detection step helps us identify where the tone of the review might clash with the actual meaning. Finally, a test is done to check how well different models—both traditional machine learning and transformer-based-perform in predicting sentiment.

B. Dataset Description

The dataset for this study comes from the publicly available **Amazon Reviews 2023** collection, hosted on HuggingFace (McAuley-Lab/Amazon-Reviews-2023). It includes a wide range of user-written reviews covering various product categories. For this research, a sample of 50,000 reviews was carefully selected across five different categories

- Electronics
- Home and Kitchen
- Cell Phones and Accessories
- Sports and Outdoors
- Books

C. Data Ingestion and Preprocessing Pipeline

The review data started off with user comments, star ratings, and some extra details. To get it ready for analysis a clear step by-step process was followed—cleaning up the text and assigning sentiment labels based on the star ratings to reflect how users felt Figure 2(Appendix).

1) Emoji Normalization: Since emojis often express emotions, all emojis in the review texts were replaced with descriptive words from the demoji library to better understand the intended meaning.

2) Text Cleaning: The textual content of each review was standardized using the following transformations:

- Lowercasing all characters
- Removing punctuation and special characters
- Stripping HTML tags
- Eliminating numbers and URLs

- Removing stopwords and excessive whitespace

3) Sentiment Labelling: Each review was labeled with a sentiment class using its associated star rating (overall column), following a widely accepted schema as **Star Rating- Sentiment Label (**1–2 - Negative), (3 - Neutral), (4–5 Positive). This mapping facilitated supervised learning using binary and multiclass classification.

4) Cleaned Dataset Summary: The dataset statistics after preprocessing are summarized below:

- **Initial review count:** 50,000
- **After removing duplicates:** 47,795
- **Average word count per review:** 90.62
- **Average number of tokens (after cleaning):** 45.84

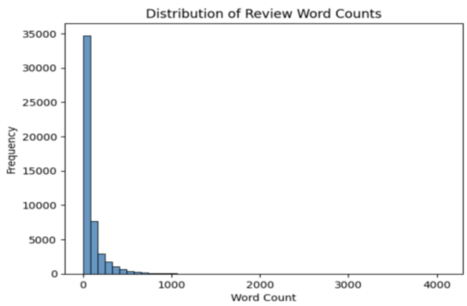

Figure 3: Distribution of Review Word Counts

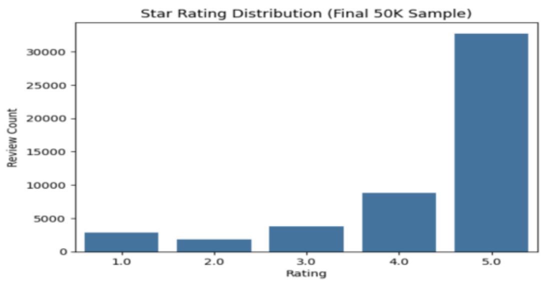

Figure 4: Star Rating Distribution

Most of the reviews in the dataset are 5-star, which creates an imbalance between positive and other sentiment classes. This issue is handled later through data balancing techniques and weighted loss functions. Overall, this section explains how the dataset was prepared and how the complete pipeline was built to combine aspect-based sentiment analysis with sarcasm detection.

5) Handling Imbalance Dataset: To handle the class imbalance, each sentiment category was reduced to the same number of samples as the smallest group—Neutral reviews. This helped create a more balanced dataset, preventing the model from favouring any one class and leading to fairer and more reliable predictions.

```
overall  sentiment  label
5.0      Positive   2        32729
4.0      Positive   2         8837
3.0      Neutral    1         3787
1.0      Negative   0         2814
2.0      Negative   0         1833
Name: count, dtype: int64
```

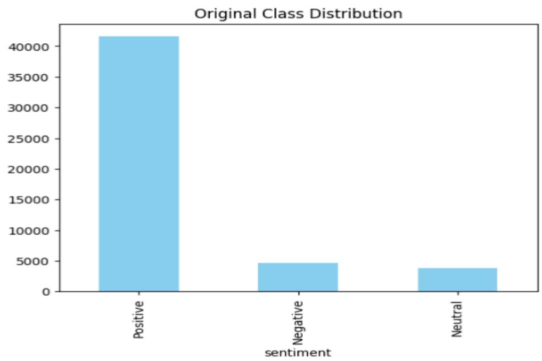

Figure 5: Dataset Summary.

```
sentiment
Negative    3787
Positive    3787
Neutral     3787
Name: count, dtype: int64
```

```
<Axes: title={'center': 'Balanced Class Distribution'}, xlabel='sentiment'>
```

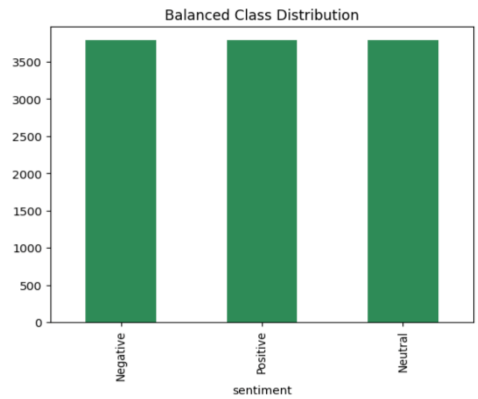

Figure 6: Distribution of Counts overall score 1 to 5(Balanced).

IV. ASPECT-BASED SENTIMENT ANALYSIS

A. Overview and Purpose

Aspect-Based Sentiment Analysis (ABSA) helps to identify the exact product features customers talk about in their reviews and picks up the emotion linked to each one. Unlike general sentiment analysis, ABSA takes it further by focusing on particular parts of a product.

B. ABSA Model and Pipeline

A pretrained transformer-based ABSA model**, yangheng/deberta-v3-base-absa-v1.1**, was used to extract specific aspect terms and determine theirsentiment—positive, negative, or neutral—from the cleaned customer reviews. The pipeline for ABSA involved the following steps: Figure 7(Appendix)

1. ***Input Preparation***: Cleaned review text was passed as input.
2. ***Aspect Extraction****:* Terms like "battery," "sound quality," and "price" were identified as aspect terms using the pretrained ABSA-BERT model.
3. ***Polarity Assignment***: Each aspect term received one of the following three classifications: Positive, Negative, or Neutral.
4. ***Aspect Dataset Construction***: The outcome was organized into a cohesive aspect-level dataset.

C. Key Findings

- **Rich Aspect Variety**: The model picked up on a broad mix of product details, from how easy it was to use or how well it worked, to more emotional reactions like "love" or "easy to use.
- **Sentiment Disparity**: Some aspects consistently attracted positive sentiment (e.g., “love”, “easy”), while others were more variable or negative.
- **Review Nuance Unveiled**: This level of analysis revealed subtleties missed by standard sentiment classifiers, especially in multi-topic reviews.

D. Contribution to the Study

By using a pretrained ABSA model, this study moved beyond overall sentiment scores and explored opinions tied to specific product features. It allowed sarcasm detection to be included in the sentiment analysis phase.

V. SARCASM DETECTION

A. Sarcasm Detection Pipeline

Catching sarcasm in product reviews really helps make sentiment analysis more accurate. People often say the opposite of what they mean—like calling a bad product "just amazing"—and older models can miss that. To address this, a structured sarcasm detection pipeline was implemented, as shown in Figure 8(Appendix). The pipeline begins with cleaned review text (50K samples) from the Amazon Reviews 2023 dataset. These texts were passed through a pretrained transformer model, **helinivan/english-sarcasm-detector**, to assign a binary sarcasm label (Yes/No) to each review. These labels were integrated into both review-level and aspect-level datasets for further comparative analysis. **Logistic Regression**, **Support Vector Machine (SVM)**, and **Naive Bayes** — three traditional machine learning models — were also employed to train with TF-IDF features derived from the reviews for sarcasm detection capability confirmation.

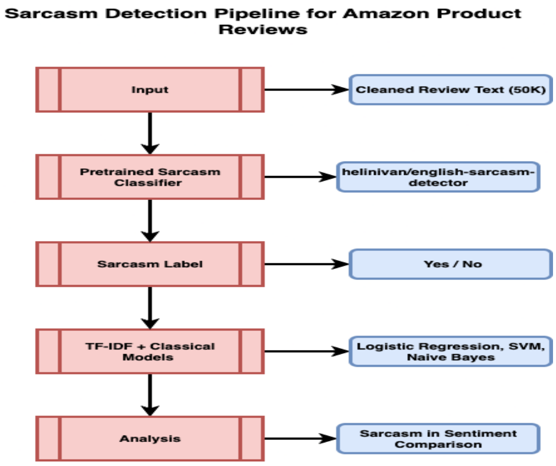

Figure 8: Sarcasm Detection Pipeline for Amazon Product Reviews

B. Sarcasm Distribution Overview

Only a small number of reviews about 3.8% were found to be sarcastic, which is quite typical since this kind of imbalance often shows up in real-world review data

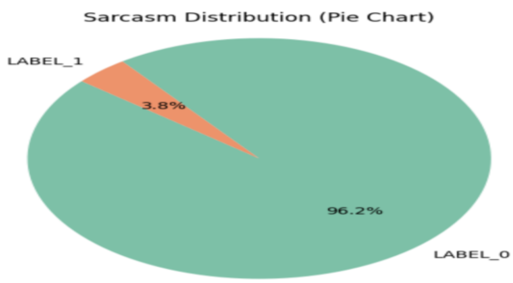

Figure 9: Sarcasm Distribution in the Dataset

Sarcasm was analyzed across sentiment classes and product categories:

- By **Sentiment**: Sarcasm was most prevalent in **negative reviews (8.42%)**, followed by **neutral (4.82%)**, and least frequent in **positive reviews (3.21%)**.
- By **Product Category**: Sarcasm was more common in reviews from **Cell Phones and Accessories** and **Electronics**, whereas **Books** had the lowest sarcasm ratio at **0.79%**. Figure 10,11 (Appendix)

The sarcasm-tagged dataset was split into 80% for training and 20% for testing, with care taken to ensure that the small portion of sarcastic reviews—around 3.8%—was properly included in both sets for balanced evaluation.

C. Classifier Evaluation

To see how well sarcasm could be detected, three traditional machine learning models were trained on a sarcasm-labeled dataset using TF-IDF features. Their performance was then evaluated using standard metrics.

- **Logistic Regression**: Achieved **94% accuracy**, with balanced precision and recall for both classes.
- **Support Vector Machine (SVM)**: Achieved the **highest accuracy of 96%**, with near-perfect recall for the sarcastic class.
- **Naive Bayes**: Performed moderately with **89% accuracy**, showing limitations in sarcastic detection recall.

D. Sentiment–Sarcasm Interplay

To better understand how sarcasm might lead to incorrect sentiment predictions, we checked how sarcasm was distributed across the different sentiment categories. Table 1 shows how often it popped up in each one, helping us see where the model might be misreading the sentiment. The results show that sarcasm is more concentrated in negative reviews. This reinforces the importance of detecting sarcastic expressions to prevent false sentiment classification.

TABLE 1

SENTIMENT VS SARCASM COUNT.

	sentiment	Not Sarcastic	Sarcastic
0	Negative	16091	89
1	Neutral	16148	83
2	Positive	16218	132

E. Implications for Sentiment Classification

Adding sarcasm detection to the sentiment analysis really made a difference—it didn't just boost accuracy, it also made the results clearer. Models that understood sarcasm were especially good at cutting down false positives when identifying negative sentiment. This proves how useful sarcasm detection can be in real-life scenarios, especially when dealing with subjective content like product reviews.

VI. SENTIMENT CLASSIFICATION MODELS

This study looks at two main approaches for sentiment classification: traditional machine learning models with TF-IDF features, and transformer-based models using DistilBERT. It uses three different versions of the dataset to examine how class balance and label structure affect the results.

- **df_balanced_3class**: Equal samples across Negative, Neutral, and Positive labels (n = 11,361)
- **df_binary**: Neutral and Positive combined into a single Positive class for binary classification (n = 46,213)
- **df_full_weighted**: Original class distribution preserved (n = 50,000), with class weights applied

The overall modeling workflow is illustrated in Figure 12(Appendix), which depicts the dual pipeline for classical and transformer models—from preprocessing through evaluation.

A. Classical Machine Learning Models

Four traditional machine learning models Logistic Regression, SVM, Naive Bayes, and Random Forest on the **df_balanced_3class** dataset using TF-IDF features (up to 5000, including bi-grams). To keep the sentiment categories balanced, we used an 80/20 stratified train-test split.

Results Summary:

- **Logistic Regression**: 68% accuracy; reliable on positive/negative classes but weaker on neutral
- **SVM**: 66% accuracy; struggled with neutral distinction
- **Naive Bayes**: 68% accuracy; good precision but low recall for neutral
- **Random Forest**: 64% accuracy; consistent but underperformed overall

B. DistilBERT Transformer Models

To boost classification performance, we fine-tuned the **DistilBERT** model (distilbert-base-uncased) on all three datasets. We used HuggingFace's DistilBertTokenizerFast for tokenization and trained the models with the Trainer API.

1) *Dataset Setups:*

- **Binary (df_binary)**: Positive vs Negative, no neutral class
- **Balanced (df_balanced_3class)**: Equal distribution of all three classes
- **Weighted (df_full_weighted)**: The natural class imbalance was handled by applying class weights

2) *Training Configuration:*

- **Epochs**: 3
- **Batch size**: 16 (train), 64 (eval)
- **Loss Function**: CrossEntropyLoss; class-weighted for df_full_weighted
- **Evaluation Strategy**: Metrics captured every 500 steps

3) *Results Overview:*

- **DistilBERT_Binary** showed excellent performance with Accuracy (97%) and F1-Score (0.97) due to reduced label complexity.
- **DistilBERT_Balanced** struggled with the Neutral class, consistent with classical models showing Accuracy (71%) and F1-Score (0.72).

- **DistilBERT_Weighted** effectively handled class imbalance, improving F1-score substantially resulting in Accuracy (89%) and F1-Score (0.89).

4) *Comparative Insights:*

- DistilBERT always performed better than all baseline models in every scenario.
- The best performance was seen in the binary sentiment tasks, showing how effective this method can be in real-world settings.
- For tasks where the classes were imbalanced, using class weights helped improve the resultsAmong the different sentiments, neutral sentiment is still the hardest one to predict for all models.
- Future work may apply class-weighting to classical models and incorporate explainability tools to interpret model decisions.

VII. RESULTS AND DISCUSSION

This section looks at how both classical and transformer-based models handle sentiment classification, how sarcasm leads to misclassification, and what patterns appear at the aspect level using ABSA. The experiments were carried out on three types of datasets: balanced 3-class, binary, and full-weighted.

A. Model Performance Comparison

Table 2 summarizes the evaluation metrics for all classifiers. The classical ML models implemented had moderate F1-scores with Logistic Regression and Naive Bayes at ~0.68 and Random Forest underperforming. On the other hand, all the DistilBERT models performed well, with **DistilBERT_Binary** achieving the highest F1 score of 0.966.

TABLE 2
PERFORMANCE METRICS FOR CLASSICAL ML AND DistilBERT MODELS

Model	Accuracy	Precision	Recall	F1
Logistic Regression	0.680	0.683	0.680	0.681
SVM	0.662	0.663	0.662	0.663
Naive Bayes	0.676	0.683	0.676	0.677
Random Forest	0.643	0.658	0.643	0.647
DistilBERT_Binary	0.966	0.965	0.966	0.966
DistilBERT_Balanced	0.713	0.722	0.713	0.716
DistilBERT_Weighted	0.891	0.898	0.891	0.894

Binary classification worked better because it was simpler—fewer labels meant less confusion. Multi-class models, however, struggled, especially with Neutral reviews, which were harder to interpret.

B. Binary vs Multi-Class Classification

The binary sentiment model (df_binary) outperformed the multi-class setup, achieving 96.57% accuracy and a 96.55% F1-score with DistilBERT. This boost came from simplifying the labels—by merging or removing neutral reviews—which clarified the sentiment and allowed the model to make sharper predictions

C. Effect of Class Imbalance

The DistilBERT_Weighted model, trained on the imbalanced df_full_weighted dataset with class-weighting, achieved 89.08% accuracy and an F1 score of 89.41%. This shows that addressing class imbalance can significantly improve performance on imbalanced data.

D. Impact of Sarcasm on Sentiment Prediction

Detecting sarcasm made sentiment predictions more accurate, especially in negative reviews where the real emotions were not obvious. Out of the traditional models tested, SVM gave the most reliable results on sarcasm-labeled data.

TABLE 3

SARCASM DETECTION MODEL PERFORMANCE

Model	Accuracy	Precision	Recall	F1
Logistic Regression	0.941	0.946	0.941	0.941
SVM	0.961	0.964	0.961	0.961
Naive Bayes	0.889	0.890	0.889	0.889

Sarcasm often showed up in reviews where people were unhappy but used language that sounded positive. This had a noticeable impact on how the model's classified sentiment, especially in multi-class setups—highlighting why it's important to include sarcasm detection in the sentiment analysis process.

VIII. CONCLUSION

This study introduces a smarter way to understand customer reviews by combining two key techniques: **Aspect-Based Sentiment Analysis (ABSA)** and **sarcasm detection**. Using 50,000 Amazon reviews from five product categories, the system digs deeper into what people are really saying—not just whether a review is positive or negative, but why and how they feel that way. The process begins with cleaning reviews and handling emojis, then picks out key product features, checks for sarcasm, and finally classifies sentiment using both traditional methods and advanced models like DistilBERT

- Best results, reaching an F1-score of 0.966.
- Using class weights helped improve the performance on less common sentiment classes.
- Looking at specific product features gave us insights that broad sentiment labels often miss.

In the end, a DistilBERT-based model with sarcasm and aspect-aware features worked best. It balances accuracy with real-world usefulness and is ideal for things like analyzing customer feedback or tracking brand reputation. This approach can also be expanded in future work using explainable tools and adapting to different domains.

Disclaimer: Use of Gen AI was in the initial stages for brainstorming and ideation. It was used to finalize the structure of the paper and to decide what is to be included in sub sections.

REFERENCES

[1] B. M. F. Harunasir, N. Palanichamy, S. Hawand K. Ng, "Sentiment Analysis of Amazon ProductReviews by Supervised Machine Learning Models," *JAdv. Inf. Technol.*, vol. 14, no. 1, pp. 23–31, 2023[Online]. Available https://www.semanticscholar.org/paper/86c009c12ea27c85a01fa1ae37d4628cc802144

[2] H. Ali, E. Hashmi, S. Yayilgan Yildirim, and S. Shaikh, "Analyzing Amazon Products Sentiment: A Comparative Study of Machine and Deep Learning, and Transformer-Based Techniques," Electronics, vol.13, no. 7, p. 1305, 2024. [Online]. Available: https://doi.org/10.3390/electronics13071305

[3] Y. Mamani-Coaquira and E. Villanueva, "A Reviewon Text Sentiment Analysis with Machine Learning and Deep Learning Techniques," in *IEEE* Access, vol. 12, pp. 193115-193130, 2024, doi:10.1109/ACCESS.2024.3513321

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10786014

[4] D. Šandor and M. B. Babac, "Sarcasm detection in online comments using machine learning," *Information Discovery and Delivery*, vol. 52, no. 2, pp. 213-226, 2024. doi: 10.1108/IDD-01-2023-0002.

[5] P. Yadav, I. Kashyap and B. S. Bhati, "UnravelingAmbivalence: A Novel Approach for Sarcasm Detection in Sentiment Analysis," 2023 3rd International Conference onTechnological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2023, pp. 638-647, doi:10.1109/ICTACS59847.2023.10390141. https://ieeexplore.ieee.org/document/10390141

[6] M. Ali, S. Hussain, F. Ahmad, and M.U. Khan, "A comprehensive analysis ofsentiment classification techniques onAmazon product reviews using classicaland transformer models," in Proc. ICMLA,2024.

[7] Amazon Reviews 2023 Dataset," HuggingFace, 2023. [Online]. Available: https://huggingface.co/datasets/McAuley-Lab/Amazon-Reviews-2023

[8] M. Alqahtani, B. AlShehri, and M. I.AlZain, "Aspect-based sentiment analysis for productreviews using BERT embeddings," in *Proc. ICCMIT*, 2021, pp. 167–172.

[9] S. Khandelwal and S. Sawant,"Sarcasm detection using contextualsentence embeddings," in *Proc. ACL*,2020, pp. 7577–7583.

[10] M. Ali, S. Hussain, F. Ahmad, and M.U. Khan, "A comprehensive analysis of sentiment classification techniques on Amazon product reviews using classicaland transformer models," in *Proc. ICMLA*,2024.

[11] R. Jindal, A. Malhotra, and N. Sharma, "Amazon product review analysisusing sentiment mining and machinelearning techniques," *Int. J. Adv. Comput.Sci. Appl.*, vol. 14, no. 1, pp. 157–164, 2023.

[12] H. Patel, M. Sinha, and N. Rathi,"Enhanced sentiment classification byintegrating sarcasm detection mechanismsin customer reviews," *Int. J. Comput. Intell. Res.*, vol. 17, no. 2, pp. 211–221,2021.

[13] V. D. Pavaloaia, "Clustering Algorithms in Sentiment Analysis Techniques in Social Media – A rapid literature review," International Journal of AdvancedComputer Science and Applications, vol. 15, no. 3, Jan.2024, doi: 10.14569/ijacsa.2024.0150314.https://thesai.org/Publications/ViewPaper?Volume=15&Issue=3&Code=IJACSA&SerialNo=14

[14] S. Saha, A. Sett and S. Karmakar, "Sentiment Analysis to Review Products based on Machine Learning," 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2022, pp. 816-821, doi: 10.1109/ICIRCA54612.2022.9985474.https://ieeexplore.ieee.org/document/9985474

[15] M. V. Rao and S. C., "Detection of Sarcasm on Amazon Product Reviews using Machine Learning Algorithms," in 2021 "Sixth International Conference on Wireless Communications,Signal Processing and Networking (WiSPNET), Chennai, India,2021, pp. 196-199, doi:10.1109/WiSPNET51692.2021.9419432.https://ieeexplore.ieee.org/document/9419432

[16] M. T. AL-Sharuee, F. Liu, & M. Pratama, "Sentiment analysis: dynamic and temporal clustering of product reviews". Applied Intelligence vol.51, pp.51–70 (2021).https://doi.org/10.1007/s10489-020-01668-6

[17] S. Gupta, R. Singh and J. Singh, "A Hybrid Approach for Enhancing Accuracy and Detecting Sarcasm in Sentiment Analysis," 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India 2020,pp.130-134, doi:10.1109/GUCON48875.2020.9231140.https://ieeexplore.ieee.org/document/9231140

[18]. T. Hua, X. Chen, Y. Wan, H. Wang, & X. Zhou, (2024). "A contextual-based approach for sarcasm detection". ScientificReports, vol14(no.1),p.2723, 2024.https://www.nature.com/articles/s41598-024-65217-8

[19]. W. Xue, & T. Li, (2024). "Aspect-based sentiment analysis: A comprehensive survey". Artificial Intelligence Review, vol 53(no.6), pp.4947-4998, 2024. https://arxiv.org/pdf/2412.02279.pdf

APPENDIX

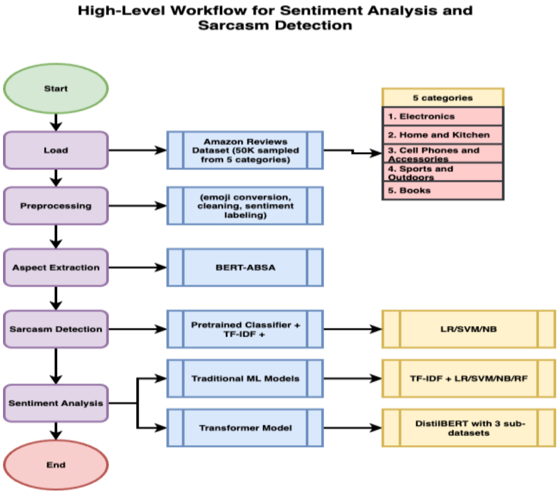

Figure 1: High-Level Workflow for Sentiment Analysis and Sarcasm Detection

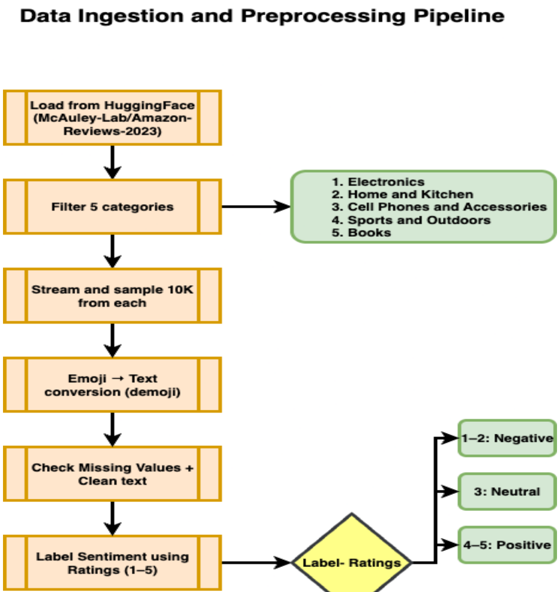

Figure 2: Data Ingestion and Preprocessing Pipeline

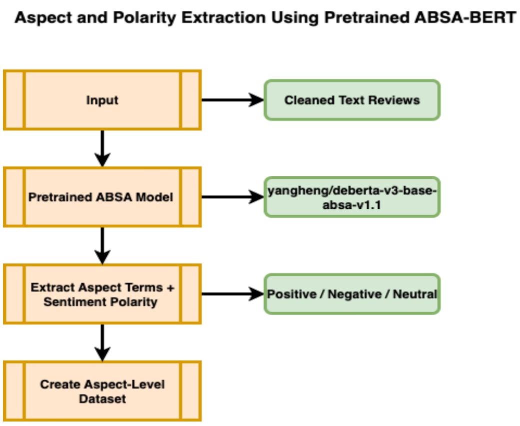

Figure 7: Aspect and Polarity Extraction Workflow Pipeline Diagram

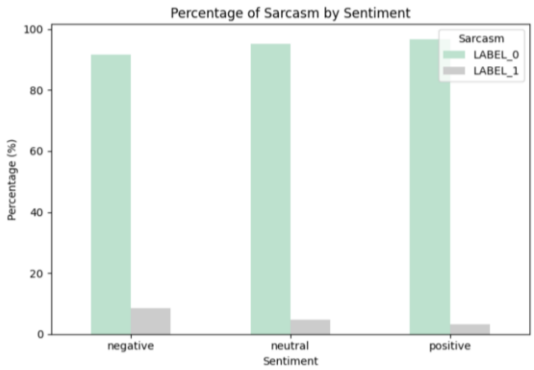

Figure 10: Percentage of Sarcasm by Sentiment

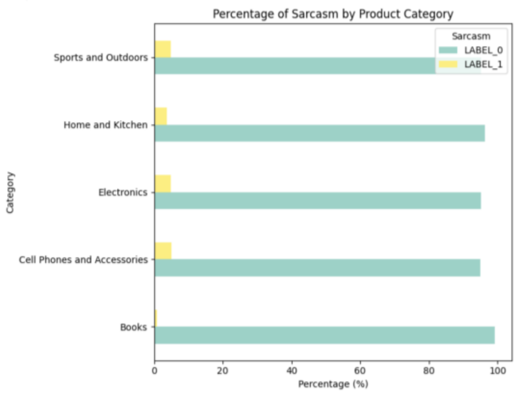

Figure 11: Percentage of Sarcasm by Product Category

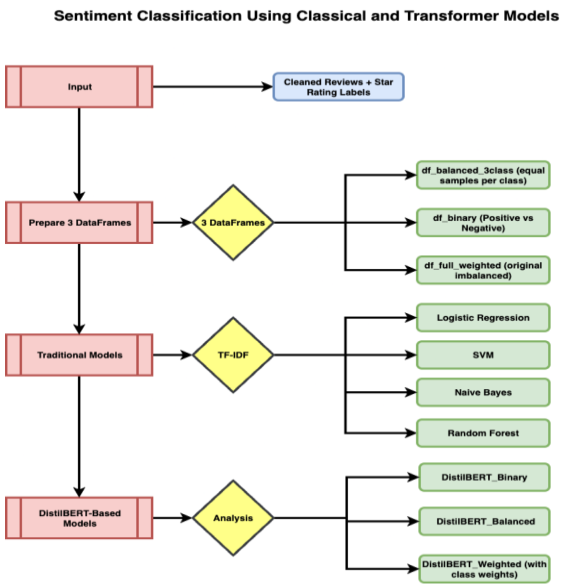

Figure 12: Sentiment Classification Pipeline Using Classical and Transformer Models