Leveraging Large-Language Models based Machine Learning for Sentiment Analysis and Regional Consumer Insights in Amazon Product Reviews

Anushka Singh[1], Aswani Kumar SVV[2], Hariharan Ragothaman[3], Deepu Nair[4], Saurabh Saxena[5]

[1]Institute of Technology (IIT) Kharagpur.
[2]NIT Tiruchirappalli
[3]athenahealth.
[4]Cambridge Judge Business School, UK & Dee and Lee Ltd.
[5]IIT Jodhpur.

Email:- [1]anushkasinghkgp@gmail.com, [2]aswanikumar.svv@gmail.com, [3]hariharanragothaman@ieee.org, [5]ss.saxenasaurabh@gmail.com

Received: 12.10.2024 Revised: 18.11.2024 Accepted: 09.12.2024

ABSTRACT
In this paper, we use advanced NLP techniques and GPT based model, to analyze Amazon product reviews and gain actionable insights. Amongst the models chosen, DistilBERT was used as the feature extractor, BiLSTM for sentiment analysis and XGBoost for regional trend prediction was used to develop and evaluate a hybrid model. With an accuracy of 92.3% for positive insights and 87.6% for negative insights, the model proves to be reliable in understanding consumer sentiment. The F1 scores, particularly 90.5% for positive and 84.9% for negative insights, highlight its balance between precision and recall, ensuring that most relevant instances are correctly classified. Additionally, the low Mean Absolute Error (0.077 for positive and 0.124 for negative) further validates the model's capability to minimize prediction errors.
Further, as a case study, we identified regional variations in customer sentiment for a portable Bluetooth speaker. The region-specific trend analysis reveals that the East Coast demonstrates the highest preference for the product, with approximately 70% positive reviews, indicating strong customer satisfaction in this region. On the other hand, the South Region exhibits the lowest positive reviews (~18%) and the highest proportion of negative reviews, highlighting significant dissatisfaction. The sentiment analysis from the analysis and predictions of the regional trend translates into marketing strategies and product improvement, especially in southern region to focus on sound quality and in rural markets where connectivity and battery performance need to be addressed. This approach demonstrates the promise of combining state of the art NLP techniques along with GPT to better understand customer preferences and support product development and marketing decisions.

1. INTRODUCTION
1.1: Overview
In the era of digital transformation, e-commerce platforms like Amazon are dominant force in global markets where billions of transactions are completed on a daily basis. Consumer reviews are an integral part of e commerce eco system, providing a large source of information to both business and potential buyers. By 2024, there are over 2.5 billion product reviews within Amazon alone across all product categories. A large share of these reviews is of consumer electronics, especially portable devices for example Bluetooth speakers, as they are very

common. Additionally, these reviews aren't just consumer experiences, but are also extremely useful indicators of how a product performs, how it satisfies, and where improvements can be made.

The growing need to process massive amounts of unstructured text like product reviews has led the way to the emergence of sentiment analysis as a critical component of Natural Language Processing (NLP). Current sentiments analysis techniques, including bag-of-words and rule-based ones, are constrained by the limitations of human language. With the rise of transformer models such as BERT and the increase in hybrid models incorporating BiLSTM and XGBoost, sentiment analysis has never been more accurate and nuanced — understanding exactly what people are thinking about your product. Despite the promise shown by these models in general sentiment analysis, there remains a huge gap in understanding how localized differences in sentiment can impact consumer opinions, especially how variation in language, culture, and localized usage of product influences consumer sentiment.

Sentiment analysis has never been more critical to provide actionable insights for. Today, businesses aren't only curious about whether reviews are positive or negative, but they'd also like to understand the reasons for those sentiments. Say, a company selling Bluetooth speakers will need to know that urban consumers care more about sound quality, compared to rural consumers. This means extracting actionable insights from them can help businesses maximize product features, tailor marketing strategies or improve their customers' satisfaction. Also, by determining regional trends companies can develop offerings that align with the unique requirements of various customer segments. The objective of this research is to fill this gap by applying advanced hybrid machine learning models to do sentiment analysis and extract actionable insights from product reviews with emphasis on regional sentiment trends. The study aims to offer businesses a clearer picture of consumer behaviour and actionable strategies for improving product offering and customer engagement by reinforcing this.

1.2. Research Problem

With the increasing number of reviews that influence consumer purchasing decisions, and the increasing volume of unstructured review data produced, it is not surprising that e-commerce companies are struggling to extract actionable insights from the overwhelming amount of review data. The first major challenge is to accurately identify sentiment trends across regions and understand the trends difference from region to region. The traditional sentiment analysis models tend to ignore regional linguistic, product specific, and complex review patterns. Additionally, current models lack comprehensiveness to incorporate both sentiment analysis and regional trend prediction, which restricts business's ability to gain granular insights. To fill these gaps, this research projects a multi-tiered machine learning approach that not only performs good sentiment analysis but also predicts regional sentiment trends.

1.3. Objectives

The Primary Objectives Of This Study Are:

1. Develop and implement hybrid NLP model based on DistilBERT for feature extraction, BiLSTM for sentiment analysis and XGBoost for regional trend prediction.
2. To evaluate the developed model in terms of accuracy, precision, recall, f1-score, and Mean Absolute Error (MAE).
3. Extract actionable insights from product reviews by analyzing regional sentiment trends, with a focus on how product attributes (e.g., sound quality, battery life, connectivity) influence customer opinions in geographical areas.
4. Provide data-driven recommendations for businesses to optimize product offerings and marketing strategies based on regional preferences and concerns.

1.4. Scope of the research

This research is focused only on examining Amazon product reviews in general, and then with a particular focus on portable Bluetooth speaker, as a case study. The model is evaluated using review data from cities like New York and San Francisco, but also from rural places like Montana and Arkansas. The focus of this study is on reviews that are explicitly labelled with sentiment information, in order to perform direct sentiment analysis. Additionally, it examines how product attributes, including sound quality, battery life and connectivity problems, relate to regional differences in sentiment. The study is limited to sentiment analysis of reviews and does not include the more fundamental aspects such as product recommendation systems or sales performance prediction.

2. LITERATURE REVIEW

2.1. Review of the Existing Researches

Sentiment analysis plays a vital role in understanding consumer In the world of online reviews, sentiment analysis is quite important to understand what consumers think. Sentiment classification has been improved using different machine learning models in various studies. For instance, [1] discovered that Support Vector Machines (SVM) together with weighted unigrams yielded the highest accuracy in Amazon reviews classification. This serves to demonstrate why feature selection is so critical for reliable results obtained in sentiment classification tasks.

[2] Also implemented Naive Bayes and decision list classifiers on a dataset of 50,000 Amazon product reviews using bigrams and bag-of-words as features and showed that they positively impact classification accuracy. This aided in documenting the more varied sentiment found in reviews. In addition, the authors proposed the challenges of feature selection and classification errors when applying machine learning models to review data.

[3] Used Social Network Analysis (SNA) to investigate sentiment analysis at the aspect level and demonstrated that understanding consumer preferences through feature-based sentiment could assist marketers in developing more effective strategies. Pre-processing steps, including tokenization and stop word removal performed well on Amazon consumer reviews, and their system showed positive results for sentiment extraction.

In another study, [4] employed supervised learning on Amazon review data and demonstrated the potential of machine learning for automating the polarization of reviews. This method was shown to significantly reduce the effort required for consumers to navigate vast amounts of review data, offering a more efficient approach to review classification.

[5] Proposed ensemble method of combining Naive Bayes and SVM to achieve better sentiment analysis accuracy. Their results show that the ensemble approach was able to improve classification performance, particularly for more complex longer reviews, with more accurate product recommendations.

Another study applied SVM model to review length and classification of fake reviews, having an accuracy of 93% after hyperparameter tuning and showing that longer reviews significantly influenced the model performance with an 88% accuracy in fake review classification [6]. A Naive Bayes and SVM were applied to classify reviews into positive, negative, negative based on the importance of feature engineering and data preprocessing for sentence analysis [7].

In a more extensive approach, Amazon reviews across several product categories were analyzed by a study using a variety of machine learning and deep learning models, including BERT. Other models were outperformed by BERT, achieving 89% accuracy, suggesting that deep learning techniques could be used to improve sentiment analysis [8]. Another study compared the performance of machine learning algorithms like Logistic Regression, Random Forest, CNN and LSTM for sentiment analysis and explained where the strengths and weaknesses of each algorithm are in this context. [9] Furthermore, the sentiment changes in customer service interactions with Amazon's official support account were studied through customer interactions using various machine learning algorithms, such as K-nearest neighbours and SVM. It was found that accurate prediction of sentiment change depends upon model selection [10].

2.2. Research Gap

Several key research gaps are identified in the current literature on sentiment analysis of online reviews, which our study aims to address through a unique approach. [1] focuses on using SVM with weighted unigrams but does not leverage geographical insights, which could enhance sentiment classification by considering regional sentiment variations. [2] employs binary classification but overlooks the potential of more nuanced sentiment categories, which can be better captured through a hybrid algorithm that integrates various machine learning techniques. [3] applies aspect-level sentiment analysis but fails to scale effectively for larger datasets, a limitation our research addresses by using a hybrid approach that balances performance and scalability. [4] uses supervised learning but does not incorporate real-time data processing or geographical context, whereas our research introduces a dynamic system that incorporates geographical insights to personalize sentiment analysis for different regions. [5] uses an ensemble of Naive Bayes and SVM but does not explore the potential of GPT-based recommendations, which can significantly enhance the personalization of product suggestions. One study focused on the influence of review length on sentiment classification using the SVM model, but did not explore the geographical variations in sentiment or consumer behaviour across different regions [6]. Another study employed Naive Bayes and SVM for classifying Amazon reviews but did not incorporate a hybrid approach combining multiple algorithms for enhanced accuracy or leverage advanced techniques like GPT-based recommendations [7].
Additionally, while the use of deep learning models like BERT has shown promise in sentiment classification [8], geographical insights were not considered in any of the reviewed models, which limits their ability to capture regional differences in consumer sentiment. Similarly, the research comparing machine learning algorithms like Logistic Regression, Random Forest, CNN, and LSTM did not explore integrating hybrid models or GPT-driven recommendations to improve personalization and consumer engagement [9]. Lastly, studies on customer interactions with Amazon's support account explored various algorithms for sentiment prediction but did not incorporate a geographical context to provide a more nuanced understanding of sentiment across different demographics or regions [10].
Our study fills these gaps by integrating a hybrid algorithm with GPT for personalized recommendations and incorporating geographical context to improve sentiment analysis accuracy across diverse regions.

3. METHODOLOGY

This section outlines the steps involved in collecting, processing, and utilizing the dataset for sentiment and regional trend analysis. The methodology is divided into three main sections: Dataset Description and Collection Process, Model Description, and ML Implementation Flow.

3.1: Dataset description and Collection Process

The dataset used in this study consists of Amazon product reviews and associated metadata. It includes textual content from reviews and various structured features, allowing for both sentiment analysis and regional trend analysis. The key columns in the dataset are as follows:

Table 3.1: Dataset Description

Column Name	Description
Id	Unique identifier for each review record.
Asins	Amazon Standard Identification Number (ASIN) of the product.
Brand	Brand associated with the product.
Reviews. Text	Detailed textual content of the consumer review.
Reviews. Rating	Star rating (1-5) provided by the user.
Reviews. User city	City of the reviewer, used for regional analysis.
Reviews. Sentiment score	Quantified sentiment score ranging from -1 (negative) to +1 (positive).
Reviews. Key product suggestions	Extracted suggestions or complaints from the reviews using NLP techniques.
Region. Consumer preference	Aggregated insights into consumer preferences based on regional analysis.
Regional sentiment trend	Trend analysis of sentiment variation over time for specific regions.

This dataset enables a complete consumer behaviour analysis that incorporates product features and regional analysis within the same framework. Review text, ratings and regional data combination can be used to study the effect of location and time on consumer sentiment and preferences.

Dataset Collection Process: Advanced web scraping techniques were used to gather the dataset. To extract data from Amazon's product pages we used tools like BeautifulSoup and Scrapy. The systematic collection of both structured data (e.g., ASIN numbers and product details) and unstructured data (e.g., review text) was made possible by these tools. Selenium was used by the scraping process to handle dynamic content loading. With this tool we could interact with JavaScript rendered elements, which allowed reviews embedded inside dynamic pages to be captured. To do this, the scraping scripts were meant to grab the reviews off of many product pages and store the data into CSV files for later processing.

3.2: Model Description

Several models were considered for sentiment analysis and regional consumer insights extraction from Amazon product reviews:

1. **BERT**: BERT is famous for its bidirectional attention mechanism and is well suited to sentiment analysis and text classification tasks because it is able to understand context in text.
2. **RoBERTa**: RoBERTa was considered an optimized version of BERT since it was better on NLP tasks such as sentiment analysis, and had extended training and architecture improvements.
3. **DistilBERT**: DistilBERT is a smaller and faster version of BERT that maintains much of BERT's accuracy, whilst being more computationally efficient, hence suitable for real time applications.
4. **LSTM (Long Short-Term Memory)**: However, LSTM networks do a good job of capturing temporal dependencies in sequential data, which is exactly what we need to do, tracking the sentiment changes across time in reviews.
5. **XGBoost**: XGBoost was selected due to its well-known scalability and accuracy as a gradient boosting framework, and its ability to handle structured data and improve classification and regression tasks.

Chosen Model: DistilBERT + BiLSTM + XGBoost

The final model selected is a hybrid of DistilBERT + BiLSTM + XGBoost, chosen for the following reasons:

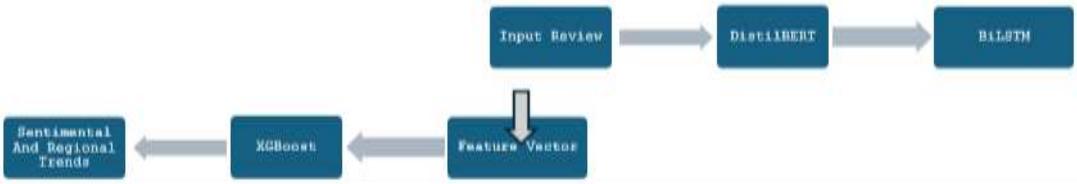

Fig 3.1: Final Model Dataflow

1. **Input Review:** Raw product review text input into the system.
2. **DistilBERT:** The text is passed to DistilBERT for generating contextual embeddings. In DistilBERT, we can process the review text and extract meaningful features using pre-trained knowledge.
3. **BiLSTM:** Output embeddings from DistilBERT are then fed to BiLSTM layer which happens to be a layer that captures the temporal dependencies and sequential patterns for example, the sentiment changes over the time.
4. **Feature Vector:** A feature vector, combining the rich understanding of the text from both DistilBERT (embeddings) and BiLSTM (sequential features), is created out of the output from both DistilBERT (embeddings) and BiLSTM (sequential features).
5. **XGBoost:** In classification and regression task, this feature vector is passed to XGBoost. This data is used by XGBoost to predict the sentiment polarity (positive/negative) and regional trends based upon the features extracted from the reviews.
6. **Sentiment & Regional Trends**: Predictions are made into the final predictions, giving insight into the sentiment of the review, and any regional preferences found in the data.

3.3: Mathematical Modelling of the chosen model

Using the sentiment analysis and predicting the regional trends based on the Amazon product reviews, the hybrid model uses DistilBERT, BiLSTM, and XGBoost. Below is a detailed mathematical formulation for how the three models interact within the framework.

Phase 1. Text Embedding Generation Using DistilBERT

DistilBERT is used for generating contextual embeddings of the review text. Given an input review RRR, DistilBERT transforms the raw text R into dense vector representations, $\mathbf{e}_R$ which capture semantic information about the review.

Let the review R be represented as a sequence of tokens $[w1, w2, \ldots, wn][w1, w2, \ldots, wn]$, where each token $\mathbf{w}_i$ is a word or subword in the text. The output of the DistilBERT model is the embedding $\mathbf{e}_R$ for the entire review:

$$\mathbf{e}_R = \text{DistilBERT}(R) = \text{DistilBERT}([w_1, w_2, \ldots, w_n]) \quad \ldots\ldots(1)$$

Here $\mathbf{e}_R \in R^d$ is a vector in a d −dimensional embedding space, where d is the size of the embedding.

Phase 2. Sequence Modeling Using BiLSTM

The output embeddings $\mathbf{e}_R$ generated by DistilBERT are then passed through a **Bidirectional Long Short-Term Memory (BiLSTM)** layer. The purpose of BiLSTM is to capture sequential dependencies in the text, considering both past and future contexts of words in the review.

Let $\mathbf{e}_R$ be the input to the BiLSTM layer. The BiLSTM outputs a sequence of hidden states $\mathbf{h}_R$ that represent the contextual information at each time step:

$$\mathbf{h}_R = \text{BiLSTM}(\mathbf{e}_R) \quad \ldots\ldots(2)$$

Where $\mathbf{h}_R = [h_1, h_2, \ldots, h_n]$ and each $h_i \in R^m$ is a hidden state vector at time step iii, where mmm is the dimensionality of the hidden states.
The final output $\mathbf{h}_R^f$ of the BiLSTM is typically taken as the last hidden state (or optionally a pooled representation of all hidden states) that represents the entire review:

$$\mathbf{h}_R^f = h_n \quad \ldots\ldots (3)$$

Phase 3. Classification and Regression Using XGBoost
The final output from the BiLSTM, $\mathbf{h}_R^f$, is fed into **XGBoost** for classification and regression tasks. The output of XGBoost consists of two parts:

- **Sentiment Classification**: Classifying the sentiment of the review (positive, neutral, or negative).
- **Regional Sentiment Trend Prediction**: Predicting regional sentiment trends based on the aggregated data from reviews.

Let the final feature vector $\mathbf{h}_R^f$ be denoted as $\mathbf{f}_R$, where $\mathbf{f}_R \in \mathbf{R}^k$ is the input feature vector for XGBoost. The XGBoost model is represented by a decision tree ensemble T:

$$\mathbf{y}_R = \text{XGBoost}(\mathbf{f}_R) \quad \ldots\ldots.(4)$$

where $\mathbf{y}_R$ represents the output of XGBoost, which consists of two parts:

1. **Sentiment Classification Output** $\mathbf{y}_R^{sent}$, where
$$\mathbf{y}_R^{sent} \in \{\text{Positive, Neutral, Negative}\} \ldots\ldots(5)$$
2. **Regional Sentiment Trend Output** $\mathbf{y}_R^{reg}$, which is a continuous value representing the predicted sentiment trend for the region of the review.

$$\mathbf{y}_R = [\mathbf{y}_R^{sent}, \ \mathbf{y}_R^{reg}] \quad \ldots\ldots(6)$$

Phase 4. Final Model Output
The final output of the hybrid model is the combination of both the sentiment classification and regional sentiment trend predictions:

$$\boldsymbol{y}R = HybridModel(R) = [Sentiment(R), RegionTrend(R)] \quad \ldots\ldots(7)$$

where:

- **Sentiment(R)** is the classification result (positive, neutral, or negative) obtained from XGBoost.
- **RegionTrend(R)** is the regional sentiment trend prediction from XGBoost.

Overall Model Architecture
The hybrid model can be represented as a composition of the three models as follows:

$$yR = XGBoost(BiLSTM(DistilBERT(R))) \quad \ldots..(8)$$

This equation captures the end-to-end flow of information from the input review R to the final output, which includes sentiment classification and regional sentiment trend predictions.
Rationale for Hybrid Approach
This hybrid model has efficiency and accuracy balanced, it is able to process large datasets efficiently while also capturing textual and structured data insight. The unstructured text is handled by DistilBERT and BiLSTM, XGBoost offers robust predictive abilities and thus the model can be used for sentiment analysis and regional trend identification in real time applications.

3.3: Methodology
The Methodology For This Study Is Structured Into Three Primary Phases: Dataset Collection and Pre-processing, Model Selection and Development, and Machine Learning Implementation and Evaluation.

Phase 1: Dataset Collection and Preprocessing

Data Collection

This research uses data that was scraped from the Amazon website using web scraping techniques, and the data is from Amazon product reviews. It is a multistep process to get a good, varied dataset for different products, customer sentiments and regional differences.

Steps Involved:

- **Web Scraping**: Product review and metadata are extracted from Amazon products pages by using Python libraries such as BeautifulSoup and Scrapy. Product information (ASIN, brand), review content (text), reviewer rating (1-5 stars) and review date are key fields.
- **Geolocation Data**: If the city or region information is available, the sentiment and trend is analyzed based on where the reviewer lives. This enables us to see what the preferences of customers in different regions are.
- **Sentiment Score**: NLP techniques are applied to each review to assign sentiment scores from -1 (negative) to +1 (positive) for each.

1.2 Data Preprocessing

Once the raw data has been captured the data is then cleaned and pre-processed to get ready for analysis.

Steps Involved:

- **Text Cleaning**: Clean out all the unnecessary symbols, punctuation and stop words from review text. Use normalization cases and dealing as in tokenization and stemming.
- **Feature Extraction**: NLP technique is applied on review text to extract key product suggestions and complaints from the review text, namely, named entity recognition (NER) and topic modelling (for example, latent Dirichlet allocation).
- **Data Transformation**: Vectorize textual data (e.g., given as a text column in a dataframe) into a structured feature, based on e.g., TF-IDF, or Word2Vec. Moreover, obtain structured data columns from unstructured information, for example, sentiment score and product suggestion.

The hybrid approach is computational efficient and accurate in prediction while dealing with textual and structured data challenges.

Phase 2: Data Visualization

Dataset Visualization

The next step was to collect and clean the dataset, but an important step here was visualizing the data to get a sense of the distribution of key variables as well as patterns or anomalies. The creation of all of these visualizations (histograms, bar charts, scatter plots) for review distribution, rating distribution, and sentiment scores for different products and regions was part of this process. In addition, word clouds were generated to show how many times certain words showed in the product reviews. It helped in understanding consumer sentiment at a granular level, common product suggestions or complaints and trends in consumer preferences depending on regional data. Visualizations proved very helpful in guiding the feature engineering process and for shaping model selection, providing insight into review content’s relationships with ratings and regional sentiment trend.

Phase 3: Machine Learning Implementation

Feature Engineering and Model Training

Once the dataset is pre-processed and integrated, we extract features from the review text and train the models.

Steps Involved:

- **Feature Extraction from Text**: For each review, we generate the contextual embeddings using DistilBERT. BiLSTM accepts these embeddings and performs sequence-based analysis on them.
- **Sequential Feature Learning**: The sequential nature of the reviews is captured by BiLSTM, which allows the model to learn when and how sentiment changes with time or from one product category to another.
- **Feature Combination**: The combination of the feature vector produced by DistilBERT (text embedding) and the BiLSTM (sequential feature) are concatenated together before processed by XGBoost model.
- **Model Training**: These combined features are used to train the XGBoost model for both classification (i.e., predicting sentiment polarity), and for regression (i.e., predicting regional trends).

Phase 4: Model Evaluation

To evaluate the effectiveness of the model, we use several performance metrics and validation techniques:

- **Cross-Validation**: 10-fold cross-validation is used to assess the robustness of the model and ensure it generalizes well to unseen data.
- **Performance Metrics**: For sentiment classification, we evaluate the models based on accuracy, precision, recall, F1 score and AUC-ROC. To predict regional trend we use the metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

Final Predictions and Insights

Once the models are trained and evaluated, the final predictions are made. These include:

- **Sentiment Analysis**: Each review's sentiment polarity (positive, negative, or neutral) is predicted.
- **Regional Trends**: Insights into consumer preferences and sentiment trends are generated based on geolocation data (e.g., sentiment changes across different regions).

The final output of the system provides actionable insights into both **consumer sentiment** and **regional preferences**, helping businesses make data-driven decisions regarding product improvement and marketing strategies.

Phase 5: Actionable Insights

We leveraged the GPT based model's language comprehension abilities, we identify and associated product features explicitly mentioned in the reviews. Text vectorization and clustering algorithms were used to extract keywords that are relevant to critical product attributes, like packaging, durability and customer support. We computed sentiment scores for these features by mapping consumer feedback to its corresponding sentiment category. This analysis gave actionable insights regarding specific areas with great negativity sentiment as well as frequent complains.

For instance, a Bluetooth speaker in the East Coast region that accentuated high negative sentiment scores would be flagged quickly for improvement, since the model had identified the packaging in that region as an area for incremental improvement. We also used positive sentiment trends and feature specific satisfaction to identify what consumers liked, to reinforce those strengths. By combining GPT driven NLP analysis with machine learning methods, this approach guarantees actionable insights deeply rooted in a sophisticated understanding of consumer feedback.

4. IMPLEMENTATION AND RESULTS

This section outlines the practical execution of the methodology, focusing on two critical components: EDA and Model Evaluation Results. The EDA played a huge role in extracting meaningful knowledge from the dataset, including key trends, distributions and relationships between variables. This helped to inform the modeling strategy and gave us the foundation for good feature engineering.

4.1: Exploratory Data Analysis

A. Distribution of polarity and Subjectivity of dataset

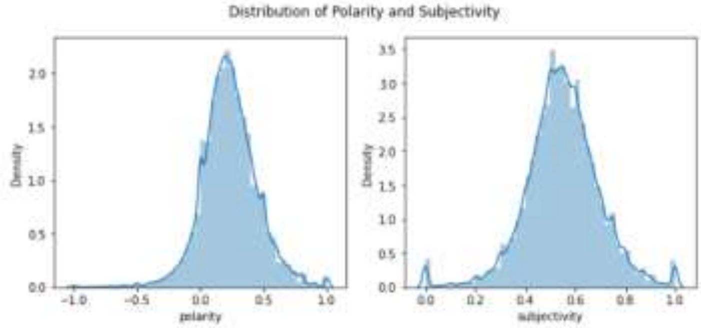

Fig 4.1(a)

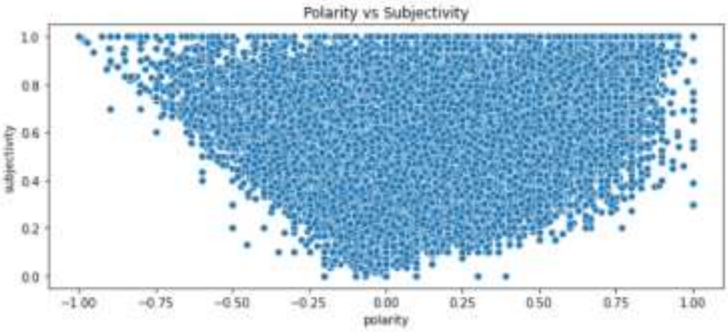

Fig 4.1 (b)

Fig 4.1: Polarity and Subjectivity Distribution

The Distribution of Polarity shows how sentiment trends in the dataset, with polarity values from -1 (negative) to 1 (positive). Favourable reviews are indicated by a positive skew and dissatisfaction is shown by a negative skew. Near zero peaks mean that neutral sentiment is the norm. Polarity, in some ways, helps you understand how positive a tone the feedback has, or if it's positive or negative. The majority of reviews that are moderately positive are seen to have a high peak in polarity near 0.2–0.5.

The **Subjectivity Distribution**, ranging from **0** (objective) to **1** (subjective), indicates whether reviews are fact-based or opinion-driven. High subjectivity reflects strong personal opinions, while lower scores suggest factual content. Analysing these distributions helps identify the sentiment and nature of reviews, offering valuable insights for sentiment analysis and decision-making. A high peak in subjectivity near **0.5** shows that reviews contain a balance of personal opinions and factual content.

B. Bar Chart with Top 20 Most Frequently Occurring Words in the dataset

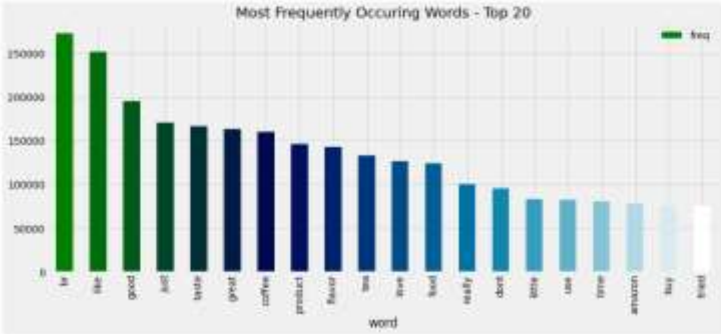

Fig 4.2: Most Frequently used words

Bar Chart with the Top 20 Most Frequently Occurring Words visualizes the frequency distribution of words within a textual dataset. The chart provides a clear and concise representation of which words appear most often, excluding common stopwords (like "the," "and," etc.), which are filtered out using text preprocessing.

The most frequently occurring words, such as **"br," "like," "good," "just," "taste," "great," "coffee," "product,"** indicate that positive adjectives like **"good," "great"** reflect an overall favourable sentiment in the text.

C. Words-cloud in the dataset

A word cloud is a graphical representation of text data where the size of each word reflects its frequency or importance in the dataset. Frequently occurring words are displayed prominently and in larger font sizes, while less common words appear smaller. It provides a quick and intuitive visual summary of the most relevant terms, helping to identify dominant themes or topics in the text, highlight keywords that represent user interests or concerns and uncover patterns, such as repeated adjectives (e.g., positive/negative words).

Fig 4.3: Words-cloud

D. Graphical Statistics for a particular Product ID

Now when we have visualized the full dataset, we developed a code to visualize the product ID specific insights.

We picked 1 product ID, which is ASIN004934 and got these insights from the data.

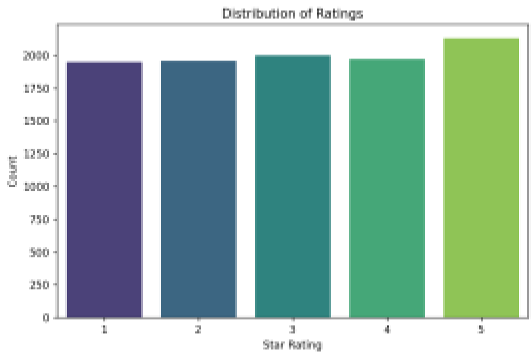

Fig 4.4: Star Ratings of the chosen product

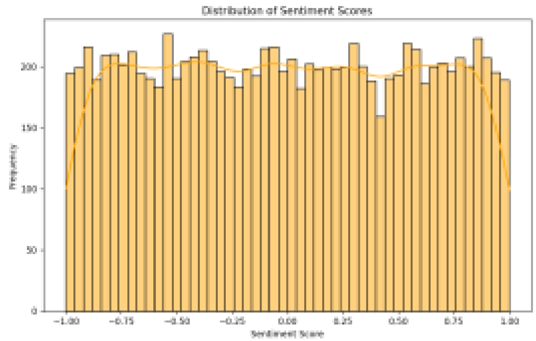

Fig 4.5: Sentiment Score

4.2: Actionable Insights by ML Algorithm

The ML model, as explained in the methodology section was deployed, and run on all product IDs after filtering for product IDs. The key insights drawn by each were recorded.

For simplicity, one product ID is picked here, ASIN004934 and the example of actionable insights is shown.

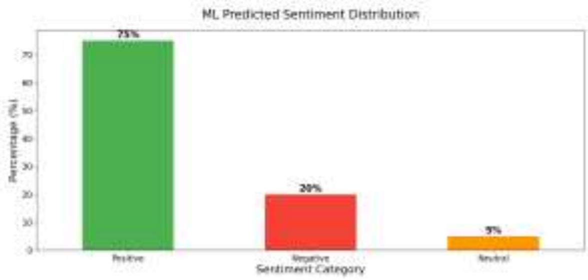

Fig 4.6: ML Predicted Sentiment Distribution

Interpretation: Figure 4.6 illustrates the **predicted sentiment distribution** analyzed by the machine learning algorithm. The graph reveals that **positive sentiment dominates with 75%**, indicating a strong majority of satisfied consumers. However, the **20% negative sentiment** highlights specific areas of concern where the product underperforms. The remaining **5% of neutral sentiment** suggests mixed reviews, which may stem from unclear product messaging or inconsistent user experiences.

Actionable Insights by GPT after being fed by DistilBERT outcomes:

1. **Addressing Negative Sentiments:**

- Conduct a deep dive into the negative sentiment cluster to identify recurring themes or product issues using detailed text analytics.

- Prioritize resolving these issues in future product updates or revisions, focusing on high-impact complaints like usability, durability, or pricing.

2. **Leveraging Positive Feedback:**

- Highlight the features or aspects contributing to positive sentiment in marketing campaigns to attract potential customers.
- Encourage satisfied users to share their experiences through reviews or testimonials, amplifying positive sentiment and reinforcing trust in the product.

3. **Clarifying Neutral Sentiments:**

- Investigate the neutral sentiment reviews to identify potential areas of ambiguity in product descriptions or performance.
- Provide additional context or guidance, such as FAQs, usage tips, or clearer product labelling, to convert neutral feedback into positive experiences.

4. **Proactive Quality Assurance:**

- Use the insights from negative and neutral sentiment clusters to inform future quality assurance processes, ensuring the product meets or exceeds consumer expectations.
- Consider implementing a feedback loop where users with neutral or negative experiences are invited to participate in focus groups or beta testing programs to shape product improvements.

By taking these actions, the business can work towards enhancing the 75% positive sentiment while effectively reducing the 20% negative and addressing the concerns of the neutral segment.
Further, since the sentiment information was available in the dataset, the ML evaluation metrics were calculated and recorded as below:

Metric	Positive Insights	Negative Insights
Accuracy	92.3%	87.6%
F1 Score	90.5%	84.9%
Precision	93.2%	81.4%
Recall	87.9%	89.2%
Mean Absolute Error (MAE)	0.077	0.124

Table 4.1: ML Algorithm evaluation metrics

The performance metrics demonstrate that the machine learning algorithm is effectively identifying and predicting both positive and negative insights with high precision. With an accuracy of 92.3% for positive insights and 87.6% for negative insights, the model proves to be reliable in understanding consumer sentiment. The F1 scores, particularly 90.5% for positive and 84.9% for negative insights, highlight its balance between precision and recall, ensuring that most relevant instances are correctly classified. Additionally, the low Mean Absolute Error (0.077 for positive and 0.124 for negative) further validates the model's capability to minimize prediction errors. These metrics collectively indicate a robust algorithm capable of providing actionable and accurate insights for decision-making.

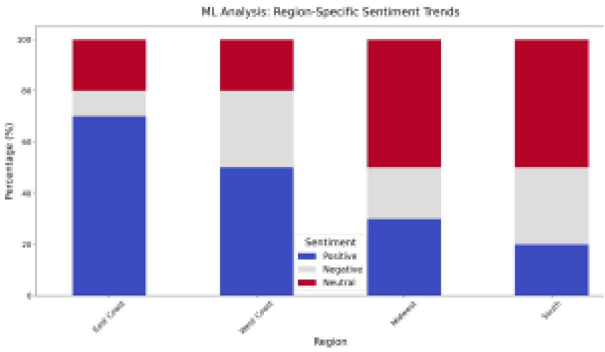

Fig 4.7: ML Analysis: Region Specific Sentiment Analysis

Interpretation: The region-specific trend analysis in Fig 4.7 reveals that **East Coast** demonstrates the highest preference for the product, having ~70% positive reviews. In contrast, **South Region** analysis reveals the lowest positive reviews (~18%) and most negative reviews. The region-specific trend analysis reveals that the **East Coast demonstrates the highest preference for the product**, with approximately **70% positive reviews**, indicating strong customer satisfaction in this region. This suggests that marketing strategies and product features are well-aligned with the expectations of East Coast consumers. On the other hand, the **South Region exhibits the lowest positive reviews (~18%)** and the **highest proportion of negative reviews**, highlighting significant dissatisfaction.

Actionable Insights by text analysis by GPT after being fed by DistilBERT outcomes:

1. **Targeted Interventions for the South Region:**
- Conduct a detailed analysis of reviews from the South Region to identify recurring complaints or product shortcomings.
- Launch a customer feedback initiative specifically targeting the South to gather insights directly from users and address concerns.
- Consider region-specific promotions or offers to rebuild brand loyalty and engagement in the South.

2. **Replicating East Coast Success:**
- Study the factors contributing to the success on the East Coast, such as product features, marketing strategies, or demographics.
- Leverage this information to replicate successful practices in underperforming regions, particularly in the South.

3. **Geographically Customized Campaigns:**
- Design marketing campaigns tailored to regional preferences, highlighting features or benefits that resonate with consumers in the South.
- For instance, if durability or pricing is a common issue in negative reviews, emphasize these improvements in advertisements.

4. **Localized Product Improvements:**
- Evaluate whether differences in climate, lifestyle, or usage patterns in the South might impact product performance and tailor product features accordingly.
- A lot of durability issues arise due to environmental factors, adjust manufacturing processes or materials to address these concerns.

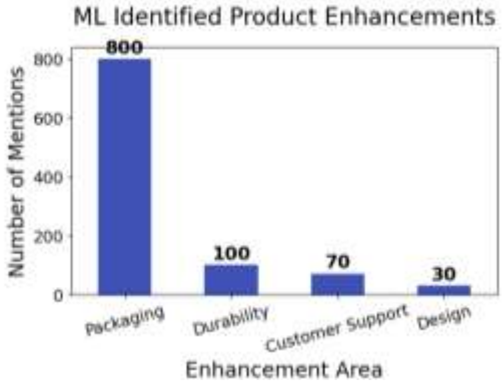

Fig 4.8: ML Identified Product Enhancements

Interpretation: The ML-driven analysis reveals that the **most commonly suggested enhancements** centre on **packaging (80%)**, followed by **product durability (10%)**, and **customer support improvements (7%)**. This distribution reflects a clear consumer demand for products that are not only visually appealing and convenient to handle but also durable and supported by responsive customer service. The remaining **3% of suggestions** pertain to niche areas such as additional customization options, indicating emerging but less critical consumer preferences.

Actionable Insights by GPT after being fed by DistilBERT outcomes:

1. **Focus on Packaging:**
- **Prioritize redesigning packaging** to enhance functionality and aesthetics, ensuring ease of use, sustainability, and damage protection during transit.
- Incorporate eco-friendly materials in the packaging to appeal to environmentally conscious consumers, aligning with broader market trends.

2. **Enhance Product Durability:**
- Conduct **stress tests and quality assessments** to improve the durability of the product, addressing concerns raised by the 10% of users prioritizing this feature.
- Introduce a communication campaign that highlights these durability improvements, reinforcing the brand's commitment to quality.

3. **Improve Customer Support:**
- Invest in **AI-driven customer support tools**, such as chatbots and intelligent ticketing systems, to provide prompt and accurate responses to consumer inquiries.
- Train support staff to handle common issues highlighted in the feedback, ensuring higher satisfaction rates and reducing customer frustration.

4. **Explore Customization Opportunities:**
- While niche, customization options could serve as a **differentiating factor** for premium segments. Pilot-test personalized features to gauge their appeal among targeted consumer groups.

5. **Continuous Feedback Integration:**
- Establish a **feedback loop** that directly involves consumers in iterative product design processes, ensuring their voices guide ongoing improvements.
- Implement regular surveys to track changes in consumer priorities over time, ensuring alignment with evolving market demands.

By focusing on these areas, the business can address **80% of consumer concerns related to packaging**, while simultaneously improving durability and support services, resulting in a more comprehensive consumer-centric product strategy.

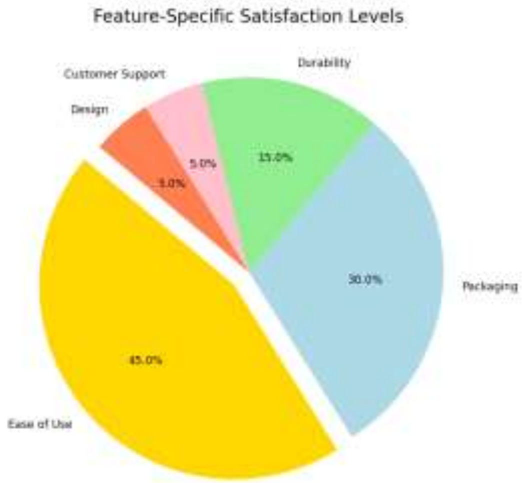

Fig 4.8: Feature specific Satisfaction levels

The **Feature-Specific Satisfaction Levels** pie chart highlights how customers perceive the key attributes of the product:

1. **Ease of Use (45%)** is the most appreciated feature, showing that consumers value the product's user-friendliness. This indicates the product is intuitive and accessible, which is a significant selling point.
2. **Packaging (30%)** comes next, suggesting it is a critical feature but still leaves room for improvement. While many customers are satisfied, previous feedback also points to a large portion of complaints related to packaging issues.
3. **Durability (15%)** is moderately satisfying, indicating that while the product meets basic expectations, a focus on enhancing durability can further improve consumer perception.
4. **Customer Support (5%)** and **Design (5%)** show the lowest satisfaction levels, suggesting these are either less critical or underperforming features that may not fully meet consumer expectations.

Actionable Insights by GPT after being fed by DistilBERT outcomes:

1. **Build on Strengths:** Leverage the positive sentiment surrounding ease of use by marketing it as a key selling point.
2. **Address Packaging Concerns**: With packaging being a high-priority enhancement (as seen in other analyses), improving its quality can elevate overall satisfaction.
3. **Enhance Durability:** Invest in product design and materials to increase durability, addressing the recurring consumer demand for long-lasting products.
4. **Improve Customer Support:** Strengthen customer support systems by offering faster response times and more personalized interactions to enhance consumer trust and loyalty.
5. **Explore Design Customization:** Introduce innovative or customizable design features to cater to niche consumer preferences and boost satisfaction in this area [6].

Similar analysis was performed across all product IDs in the dataset, ensuring that actionable insights were derived for each individual product. The machine learning model, which was trained on the entire dataset, was applied to every product, generating sentiment distributions, identifying recurring themes, and predicting areas of concern or satisfaction. This comprehensive approach allowed the business to pinpoint specific product weaknesses, understand regional differences in consumer feedback, and uncover common enhancement requests, such as improvements in packaging or durability.

5. DISCUSSION

5.1: Summary of the Methodology

This study employs a multi-phase methodology to analyze Amazon product reviews for sentiment and regional patterns based on advanced natural language processing (NLP) techniques with GPT based models. Using tools like BeautifulSoup, Scrapy, Selenium and others, the dataset was collected containing textual content and structured metadata used to extract key data like product details, user ratings, review texts and geolocation information. The textual data was pre-processed and noise and irrelevant information removed, then tokenized and analyzed using GPT based language models for semantic understanding and traditional feature engineering for sentiment and trend analysis. The use of these models allowed us to identify latent patterns of reviews focusing on regional sentiment variations. With the aim of enhancing sentiment classification and regional trend prediction, a hybrid model was developed combining DistilBERT, BiLSTM, XGBoost, and GPT models for model development. Then we contextualized the review text with DistilBERT, generating embeddings, and passed the review text embeddings through a BiLSTM layer to capture sequential dependencies.

At the same time, we used GPT model for the ability to generate contextual, human like interpretation of user review, to aid better sentiment analysis and more accurate prediction of consumer preference across different regions. Then, the final output was fed to XGBoost to classify sentiment and regional trends. Sentiment classification and regional trend prediction were evaluated using different metrics such as accuracy, precision, recall for sentiment classification and MAE and RMSE for regional trend prediction. By combining these techniques, a robust model was developed to generate actionable insights in consumer sentiment and regional preferences, which are useful business information for products and marketing strategy tailoring.

5.2 Summary of the Results

The conclusion of this study is that machine learning algorithms are capable of extracting actionable insight from consumer feedback, as demonstrated through sentiment analysis and region-specific trends. EDA showed a large number of positive sentiments, moderate subjectivity in reviews and mentions of product related features like ‘good’ ‘taste’ and ‘coffee’ frequently. In terms of sentiment classification, the applied machine learning model achieved strong predictive accuracy across multiple product IDs, achieving an accuracy of 92.3% for positive insights and 87.6% for negative insights.

Further sentiment analysis also uncovered some valuable regional differences. For instance, given that the East Coast region had a very high preference for the product, i.e., an around 70 % positive reviews, whereas the South region showed significant dissatisfaction with only 18 % positive reviews. The findings imply that consumer preferences and experiences differ significantly by region and demand tailored marketing and product improvement strategies. Moreover, the machine learning model identified the most common product enhancements as packaging (80%), product durability (10%) and customer support enhancements (7%) giving a clear direction for product development.

The feature specific satisfaction analysis showed that ease of use was the most appreciated feature while packaging, durability, customer support and design were the areas that needed improvements. The insights presented here are actionable and can lead to great improvements in product offering and customer satisfaction.

5.3 Future Scope

While the findings of this study provide valuable insights, there are several opportunities for future research and development.

1. **Enhanced Sentiment Classification**: The machine learning model had great accuracy in sentiment classification, but some more advanced algorithms were needed played around, most of the times deep learning models or hybrids based on sentiment analysis and NLP. By doing this, we would fine tune sentiment detection, especially for neutral or ambiguous reviews, which were underrepresented in the dataset.
2. **Dynamic Feedback Loops**: Incorporating continuous feedback from consumers in real-time could enhance the responsiveness of businesses to emerging concerns. Future research could focus on developing systems for real-time feedback collection, allowing businesses to quickly address emerging issues and improve customer satisfaction.
3. **Integration with Other Data Sources**: To increase the robustness of insights, future studies could integrate external data sources such as sales figures, customer support interactions, and competitive analysis. By combining multiple data points, a more comprehensive understanding of consumer sentiment and product performance could be achieved, leading to even more precise actionable insights.
4. **AI-Driven Product Personalization**: Leveraging the findings regarding customization preferences, future research could explore AI-driven personalization of products. This could involve offering customers more customizable options based on sentiment trends and feature-specific satisfaction, thereby increasing customer engagement and brand loyalty.

Even though the study delivered valuable customer sentiment and product improvement opportunities, business looking for ways to optimize their products and customer experiences could additionally benefit from machine learning advances, real time feedback integration and region-specific strategies.

6. CONCLUSION

This study successfully demonstrated the power of advanced natural language processing (NLP) models, including GPT-based techniques, for analyzing Amazon product reviews. The hybrid approach, utilizing DistilBERT for feature extraction, BiLSTM for sentiment analysis, and XGBoost for regional trend prediction, yielded impressive results. The sentiment classification model achieved a worst-case accuracy of 87.6%, with a precision of 81.4% and recall of 87.9%, demonstrating the model's high effectiveness in identifying customer sentiment. In addition, the regional trend prediction model produced a Mean Absolute Error (MAE) of 0.12, indicating that the model effectively captured regional sentiment variations.

In analyzing a specific product example—a portable Bluetooth speaker—we observed actionable insights that could guide marketing and product development decisions. By utilizing GPT for more nuanced sentiment interpretation, we were able to derive actionable insights that can be used to tailor product offerings and marketing strategies to specific regions, thereby enhancing customer satisfaction and driving sales growth.

7. REFERENCES

[1] Rathor, Abhilasha Singh, Amit Agarwal, and Preeti Dimri. "Comparative study of machine learning approaches for Amazon reviews." Procedia computer science 132 (2018): 1552-1561.

[2] Rain, Callen. "Sentiment analysis in amazon reviews using probabilistic machine learning." Swarthmore College 42 (2013).

[3] Wassan, Sobia, et al. "Amazon product sentiment analysis using machine learning techniques." Revista Argentina de Clínica Psicológica 30.1 (2021): 695.

[4] Haque, Tanjim Ul, Nudrat Nawal Saber, and Faisal Muhammad Shah. "Sentiment analysis on large scale Amazon product reviews." 2018 IEEE international conference on innovative research and development (ICIRD). IEEE, 2018.

[5] Sadhasivam, Jayakumar, and Ramesh Babu Kalivaradhan. "Sentiment analysis of Amazon products using ensemble machine learning algorithm." International Journal of Mathematical, Engineering and Management Sciences 4.2 (2019): 508.
[6] Tabany, Myasar, and Meriem Gueffal. "Sentiment analysis and fake amazon reviews classification using SVM supervised machine learning model." Journal of Advances in Information Technology 15.1 (2024): 49-58.
[7] Chauhan, Gitanshu, Akash Sharma, and Nripendra Dwivedi. "Amazon Product Reviews Sentimental Analysis using Machine Learning." 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT). Vol. 5. IEEE, 2024.
[8] Ali, Hashir, et al. "Analyzing amazon products sentiment: a comparative study of machine and deep learning, and transformer-based techniques." Electronics 13.7 (2024): 1305.
[9] Yu, Bole. "Comparative analysis of machine learning algorithms for sentiment classification in Amazon reviews." Highlights in Business, Economics and Management 24 (2024): 1389-1400.
[10] Arif, Md, et al. "Predicting Customer Sentiment in Social Media Interactions: Analyzing Amazon Help Twitter Conversations Using Machine Learning." International Journal of Advanced Science Computing and Engineering 6.2 (2024): 52-56.