Extracting Customer Perception of iPhone SE from e-commerce Website Reviews to Propose Future Development Strategies

1.Executive Summary

Apple’s iPhone revenue in India rose 42% year-over-year and is expected to continue the growth as the middle-class demographics in India are increasing their “appetite for premium devices” (Kaustubh, 2024). As Apple’s fifth-largest smartphone market, the Indian market is significant strategic importance to the company. To take advantage of this opportunity, our project utilized Indian consumer reviews on Flipkart, the biggest Indian e-commerce platform, to conduct Sentiment Analysis and Topic Identification. By analyzing sentiments with the VADER Lexicon and topics through Latent Dirichlet Allocation, we identified key features that impacted user satisfaction and gained direction and guidance for product improvement. Despite predominantly positive sentiment our analysis revealed disparities, particularly concerning battery features of the smartphone, suggesting areas for improvement. We recommend marketing the product based on phone performance, camera quality, and value for money, while addressing negative sentiments on battery performance to better meet customer’s expectations and improve overall satisfaction.

2. Project Objectives

Our project is focused on extracting information from Flipkart regarding iPhone SE reviews and utilizing the information to derive future development strategies for Apple. This project is designed to achieve following key objectives: discern the sentiments articulated within the reviews, discover primary topic of interest most among iPhone SE users in India, identify topics correlated with positive, negative, and neutral sentiments, and evaluate the effectiveness of ratings as indicators of sentiment, compared to textual reviews.

These objectives are targeted to the goal of achieving better product development strategies for the Indian market. By analyzing customer needs, we can effectively identify smartphone features that need enhancement. The insights can also guide targeted marketing campaigns that resonate with consumer preferences, leading to increased personalization and competitiveness. In short, our project objective is to gain analysis that the company can utilize to quickly respond to market trends and maintain a proactive stance.

3. Data Description

The data was originally gathered from Flipkart website using Selenium and BeautifulSoup. There are 9,713 rows and 3 columns: Ratings, Comment, and Reviews. There is no missing data in the dataset. Ratings are on a scale of 1 to 5, where 1 is “poor” and 5 is the “best”. The mean and standard deviation for the “Ratings” is 4.46 and 1.03, respectively. “Comment” is the title of the user’s review and “Reviews” are textual data containing the body of the review. The original data of the reviews were constrained by “read more” function, which was intended to provide preview of the review and offer additional context upon clicking. Yet during scraping from the website, only the preview of the data was captured.

4. Methodology

We first preprocessed the data to transform unstructured text into structured dataset. The preprocessing phase involved tokenization, annotation, lemmatization, removal of special characters and stop words, elimination of non-text characters, cleaning up HTML markups, accent removal, and contraction expansion.

The dataset was mostly positive-skewed, so isolating each sentiment group was essential to prevent predominantly positive reviews from overshadowing neutral and negative sentiment groups. Therefore, we first conducted sentiment analysis of the reviews and dealt with negative and neutral reviews separately. The approach of including neutral sentiment category ensured that the rating of “3,” indicative of neutrality, was properly taken into consideration. We chose the VADER Lexicon for sentiment analysis, a tool tailored for social media text. VADER's strengths lie in its ability to handle both sentiment polarity and intensity, as well as recognizing various emotional expressions and idioms, making it exceptionally practical for analyzing online reviews filled with slang, emoticons, abbreviations, and unconventional punctuation. However, its reliance on static dictionary of words and sentiment scores may lead to incomplete analysis if the text contains words or phrases not in the dictionary. VADER may also struggle with complex linguistic features like sarcasm, irony, or jokes.

Despite the challenges, VADER has a highly relevant dictionary for social media text, including informal language that enables more accurate sentiment detection than other dictionaries. VADER's nuanced understanding of sentiment polarity and intensity offers a detailed view of customer emotions, which is crucial for depth in sentiment analysis.

Due to the size of the data, and extensive corpus, the automatic identification of common theme is essential. Therefore, we employed topic modeling to efficiently identify different topics in positive, negative, and neutral sentiment reviews. We first used the "Bag-of-Words" method to extract features and used Latent Dirichlet Allocation (LDA) function to identify 4 topics for each sentiment review group. There are limitations to the method. Because the overall topic of all the reviews are about different features of the smartphone, topic modeling may be less successful in distinguishing different topics, compared to manually categorizing topics. Another challenge would be interpretation of the generated topic, as it may be difficult to extract meaningful insight without additional context. Despite the limitations, topic modeling is essential for its ability to provide systematic insight into large text dataset with manual categorization facilitating deeper understanding of underlying themes.

Bi-gram was conducted as a further analyzation for interpretability of the result. Specifically, "Battery" was a prevalent term in both positive and negative sentiment, so Bi-gram used to analyze co-occurrence of word pairs and find association between adjacent words to provide a more insightful result relating to "Battery".

5. Results and Discussion

In our comprehensive analysis of customer reviews for the Apple iPhone SE on Flipkart, we implemented a text mining approach that combined Lexicon-based sentiment analysis and advanced topic modeling to classify the textual data into three categories: positive, neutral, and negative.

Sentiment Analysis**:** Using the VADER lexicon, we categorized the sentiments of the reviews into positive, neutral, and negative. The distribution of sentiments showed a predominance of positive sentiments, with 6,869 reviews classified as positive compared to only 552 as negative and 2,292 as neutral. The average sentiment scores were as follows: positive reviews had an average VADER score of 0.647758, neutral reviews had 0.000495, and negative reviews were -0.480504, indicating a clear skew towards positive feedback in the dataset.

Topic Modeling**:** We visualized the topic distribution through inter-topic distance map, which shows topic focus and separation of topics within each category. The semantic consistency of the topics was quantified using coherence scores, which revealed semantic consistency within the topic. Positive comments show a high coherence score (-3.215), indicating that the topic of discussion is well defined and clear. In contrast, negative reviews (-4.515) and neutral reviews (-13.9118) show lower coherence scores, suggesting that individuals express neutral and negative sentiments across a diverse range of topics, resulting in a less focused and coherent discussion.

Key topics highlighted in positive reviews included: battery life, camera quality, and value for money (Table 1.1). The terms frequently associated with positive sentiments included 'good', 'great', and 'love'. The titles for each of the four dominant topics are as follows: "Battery, Camera, Performance, and User Experience" for Topic 1, "Size and Performance" for Topic 2, "Best Deals and Value for Money" for Topic 3, and "Quality, Performance, and Satisfaction" for Topic 4. In the negative sentiment group, key concerns were around issues such as battery performance and customer service, with prevalent terms being 'bad', 'problem', and 'worst' (Table 1.2). Negative intertopic distance map highlights four dominant topics which can be categorized as follows: Topic 1 explores "Battery Life, Usage, and Screen Display," Topic 2 focuses on "Customer Service Experience: Purchases and Issues with Chargers," Topic 3 discusses "Battery Performance Concerns, Quality, and Price," while Topic 4 addresses "iPhone Issues: Network Connectivity and Heating Problems." The neutral reviews group often discussed both positive and negative aspects without strong emotional expressions, with common terms like 'average', 'okay', and 'fine'. The finding displayed in the intertopic distance map shows that the "read more" function of the review has miscategorized words such as "goodread," "niceread," or "awesomeread" into the neutral group, despite the word conveying positive sentiment once the word "read" is removed from the end (Table 1.3). This could explain why the neutral sentiment group exhibits the lowest coherence score. For neutral sentiment group, we titled the topics as follows: "Charger, Screen, and Size" for Topic 1, "Battery Performance and Heating Issue" as Topic 2, "Miscategorized: Customer Satisfaction" as Topic 3 and "Miscategorized: Positive Product Experiences" as Topic 4.

Bi-grams: One of the noticeable themes that overlaps across topic was "Battery," which consistently appeared throughout the positive, negative, and neutral sentiment groups. We conducted Bi-grams of the word "Battery" within positive and negative sentiment group. The words overlapped in both groups, with the top three words in the positive sentiment group related to "Battery" being "battery life," "battery backup," and "good battery," while the top three words in the negative sentiment group were "battery life," "battery backup," and "poor battery" (Table 2.1, Table 2.2). We can conclude that there is a disparity in customer's sentiment regarding battery. One observation we can make is that the proportion of customers mentioning the battery in the positive sentiment group is significantly lower than in the negative sentiment group, where "Battery" is predominantly mentioned.

***Accuracy of Sentiment Analysis vs. Star Ratings*:** Our evaluation also compared the effectiveness of sentiment analysis against traditional star ratings. We found that while the overall sentiment accuracy derived from textual analysis stood at 66.97%, it was slightly less than the 71.39% accuracy from star ratings. Notably, the accuracy for identifying neutral sentiments through ratings was particularly low. This lower accuracy may be attributed to two primary factors: the sparse data available for neutral ratings, which hampers the model's ability to accurately learn and classify neutral sentiments, and the inherent ambiguities present in textual data that make sentiment analysis challenging. Automated tools like VADER can misinterpret seemingly neutral phrases due to specific keywords, highlighting the complexities of relying solely on text for sentiment analysis.

6. Conclusion

***Recommendations*:** Based on the intricate analysis of the Apple iPhone SE reviews on Flipkart, we recommend the following strategic actions. The first step is to focus on the product features that stand out in positive sentiment. From the analysis, we recommend continuing to emphasize and improve the features that receive positive feedback, such as phone performance, camera quality, and overall value for money. The fact that these features are frequently mentioned and appreciated in positive reviews suggests that this has become a significant selling point for the iPhone SE and should be emphasized in marketing and further developed in product iterations. Despite battery being part of the positive review, the proportion of positive reviews mentioning the battery is significantly lower compared to the proportion of negative reviews mentioning the battery. Therefore, there is also a need to address the issues highlighted in the negative reviews, especially regarding battery heating and fast power loss, which has widely caused consumer dissatisfaction. We suggest that the R&D team could focus on strengthening and improving the battery technology to better adapt to warmer climates in India.

***Shortcomings**:* The inherent shortcoming of the analysis comes from the original data. The review data extracted from the website was truncated due to the "read more" function, resulting in an incomplete representations of user opinions and mis-categorization of positive sentiment to neutral sentiment group. Despite this limitation, we anticipated that the extracted reviews would still offer valuable insights, due to the size of the data and assumption that the most salient opinions and impressions are articulated at the outset of the review. If we were given the opportunity to conduct the analysis again, we hope that the methodology could be improved by developing the capability to scrape and compile our own dataset of reviews, ensuring a more comprehensive and unrestricted analysis, as our team truly believe our project could derive meaningful and insightful business strategy for Apple's smartphone development plan for targeting Indian market.

Appendix

Table 1.1: Positive Sentiment Intertopic Distance Map

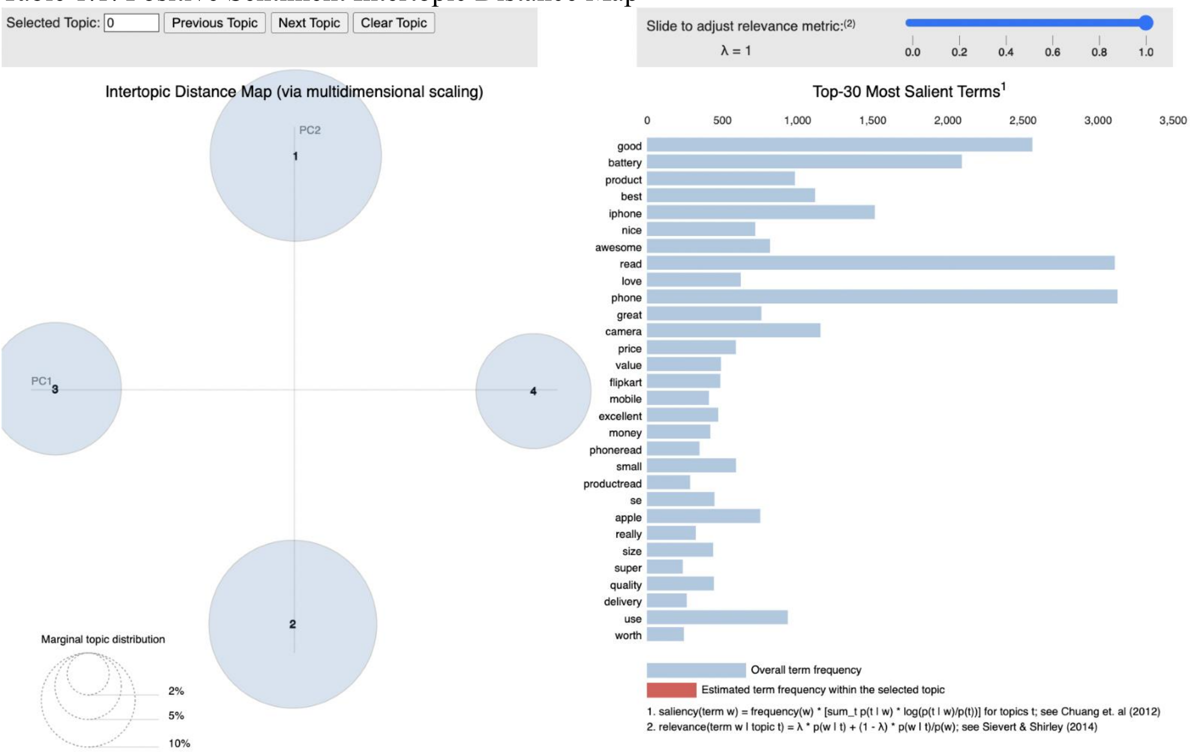

Table 1.2: Negative Sentiment Intertopic Distance Map

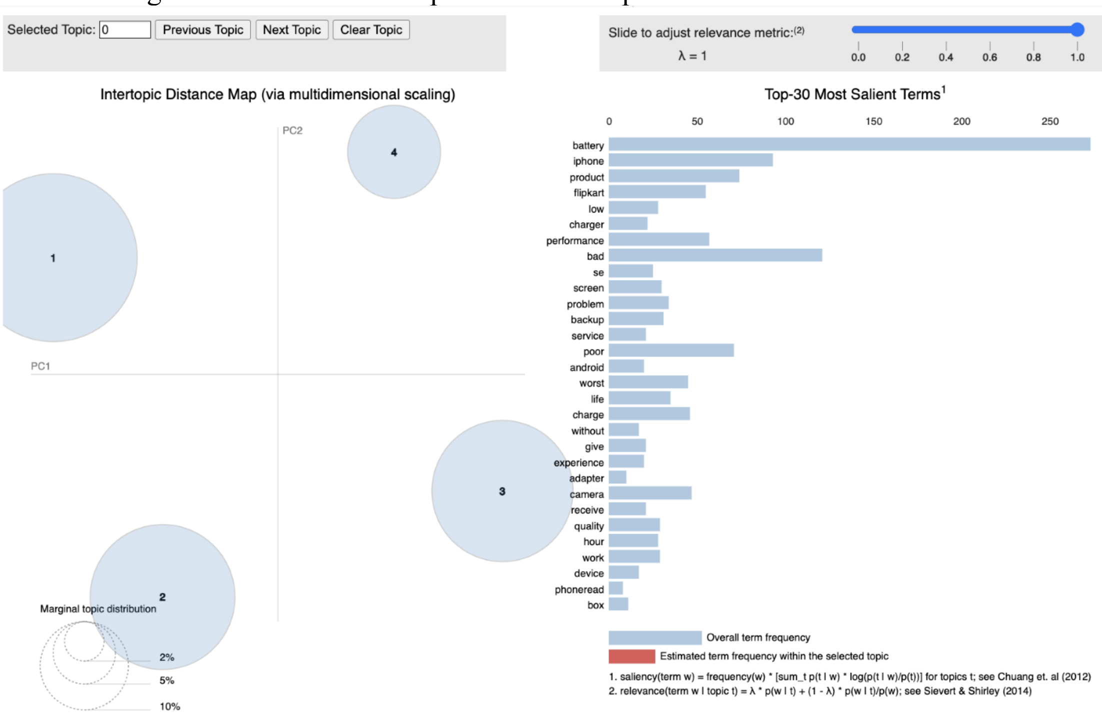

Table 1.3: Neutral Sentiment Intertopic Distance Map

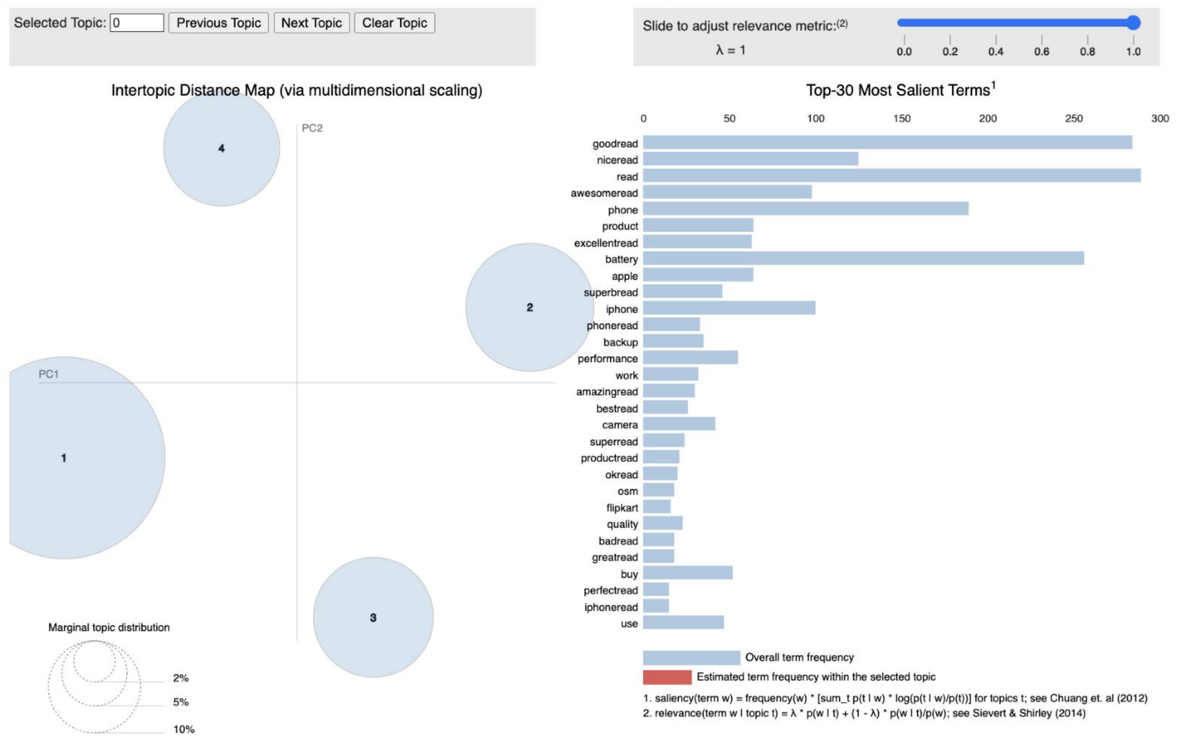
Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
Marginal topic distribution
2%
5%
10%
Top-30 Most Salient Terms1
0
50
100
150
200
250
300
goodread
niceread
read
awesomeread
phone
product
excellentread
battery
apple
superbread
iphone
phoneread
backup
performance
work
amazingread
bestread
camera
superread
productread
okread
osm
flipkart
quality
badread
greatread
buy
perfectread
iphoneread
use
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Table 2.1: Bi-gram of Positive Sentiment Group

```python
# Extract 'battery'-related pairs from the bi-gram table of positive sentiments, and then add up the number of counts by each column.
# We can notice that which word is relevant to 'battery'.
vectorizer_Bi_Uni_Grams = CountVectorizer(max_features=1000, ngram_range=(2,2))
Bi_Uni_Grams_matrix = vectorizer_Bi_Uni_Grams.fit_transform(positive_reviews).toarray()
df = pd.DataFrame(np.round(Bi_Uni_Grams_matrix,2),columns=vectorizer_Bi_Uni_Grams.get_feature_names_out())

# Filter columns containing 'battery'
battery_columns = [col for col in df.columns if 'battery' in col]
battery_related_data = df[battery_columns]

# Sum up the counts by each column, and then sort the results out in a descending order.
sums_by_column = battery_related_data.sum()
sorted_sums = sums_by_column.sort_values(ascending=False)

print(sorted_sums)
```

```
battery life          303
battery backup        244
good battery          153
except battery        130
phone battery         111
                     ...
battery management      8
range battery           8
people battery          8
drain battery           8
amazing battery         8
Length: 100, dtype: int64
```

Table 2.2: Bi-gram of Negative Sentiment Group

```python
# Extract 'battery'-related pairs from the bi-gram table of negative sentiments, and then add up the number of counts by each column.
# We can notice that which word is relevant to 'battery'.
vectorizer_Bi_Uni_Grams = CountVectorizer(max_features=1000, ngram_range=(2,2))
Bi_Uni_Grams_matrix = vectorizer_Bi_Uni_Grams.fit_transform(negative_reviews).toarray()
df = pd.DataFrame(np.round(Bi_Uni_Grams_matrix,2),columns=vectorizer_Bi_Uni_Grams.get_feature_names_out())

# Filter columns containing 'battery'
battery_columns = [col for col in df.columns if 'battery' in col]
battery_related_data = df[battery_columns]

# Sum up the counts by each column, and then sort the results out in a descending order.
sums_by_column = battery_related_data.sum()
sorted_sums = sums_by_column.sort_values(ascending=False)

print(sorted_sums)
```

```
battery life      38
battery backup    34
poor battery      20
battery drain     19
battery poor      18
                  ..
min battery        1
poor batterybad    1
point battery      1
pm battery         1
modal battery      1
Length: 87, dtype: int64
```

Works Cited

Das, K. (2021). Apple iPhone SE reviews & ratings, Version 1. Retrieved May 2, 2024 from https://www.kaggle.com/datasets/kmldas/apple-iphone-se-reviews-ratings/data

Kaustubh, A. (2024, April 23). *Move over Europe, India is Apple's brightest new iPhone Star: - Times of India*. Move over Europe, India is Apple's brightest new iPhone star. https://timesofindia.indiatimes.com/gadgets-news/move-over-europe-india-is-apples-brightest-new-iphone-star/articleshow/107944056.cms