Business Intelligence

Building a Customer Feedback Intelligence System using Web Scraping & Text Mining

Date: April 21st, 2025

Table of Contents

I. Introduction:

In today's digital landscape, customer evaluations have become an essential component of consumer decision-making and business strategy. These reviews not only indicate product satisfaction, but they also provide valuable insights into the user experience, product difficulties, and opportunities for improvement. However, raw reviews are frequently unstructured and lengthy, making it difficult to extract useful information manually.

This research focuses on developing an automated customer feedback intelligence system using web scraping, natural language processing (NLP), and sentiment analysis. Using BestBuy Canada's headphone ratings as an example, we scraped and analyzed thousands of real consumer evaluations to classify sentiment, uncover major product trends, and show sentiment-driven insights.

The goal of this project was to transform unstructured feedback into structured intelligence, allowing organizations to make more informed decisions about product development, marketing tactics, and consumer interaction. By evaluating phrase usage and sentiment polarity, this study reveals what customers genuinely value or detest, providing a clear path for enhancing customer happiness and brand reputation.

II. Executive Summary

This study describes an end-to-end system for collecting, processing, and analyzing product evaluations from BestBuy Canada using web scraping, text preprocessing, sentiment analysis, and keyword mining. The goal was to provide actionable business intelligence from customer feedback and improve data-driven decision-making in retail operations.

III. Phase 1: Data Collection - Web Scraping from BestBuy Canada

Objective: Gather genuine consumer feedback, item names, and links from the BestBuy Canada website.

Approach:

- We started with local HTML versions of Best Buy's headphone listings.
- The Selenium WebDriver was used to load and parse these websites.
- XPath was used to retrieve all applicable product hyperlinks.
- Each product page was then viewed, and customer reviews were extracted from their corresponding HTML review sites.

Outcomes:

- Scraped 710 product entries.
- Gathered 4346 customer reviews.
- All data was structured and placed in a MongoDB collection named reviews within a database called customer_reviews.

IV. Phase 2: Preprocessing and Data Cleaning

Goal: Prepare the scraped reviews for sentiment classification.

Steps:

- Promotional texts such as "[This review was collected as part of a promotion.]" / No reviews found, Null rows were removed.
- Each review was tokenized at both sentence and word levels.
- Reviews were cleaned to remove extra whitespace, special characters, and stored along with their tokens in MongoDB.

Result: A cleaned dataset of 4346 reviews was ready for NLP analysis.

V. Phase 3: Sentiment Classification

Tool Used: VADER Sentiment Analyzer (Lexicon and Rule-Based)

Process:

- Tokenized review texts were rejoined.
- Each review's sentiment score was calculated using VADER's compound scoring.
- Classification Rules:
    - Compound score > 0.2 → Positive
    - Compound score < -0.2 → Negative

- Else → Neutral

Final Sentiment Distribution:

- Positive: 87.7% (approx. 3813 reviews)
- Negative: 7.5% (approx. 326 reviews)
- Neutral: 4.8% (approx. 207 reviews)

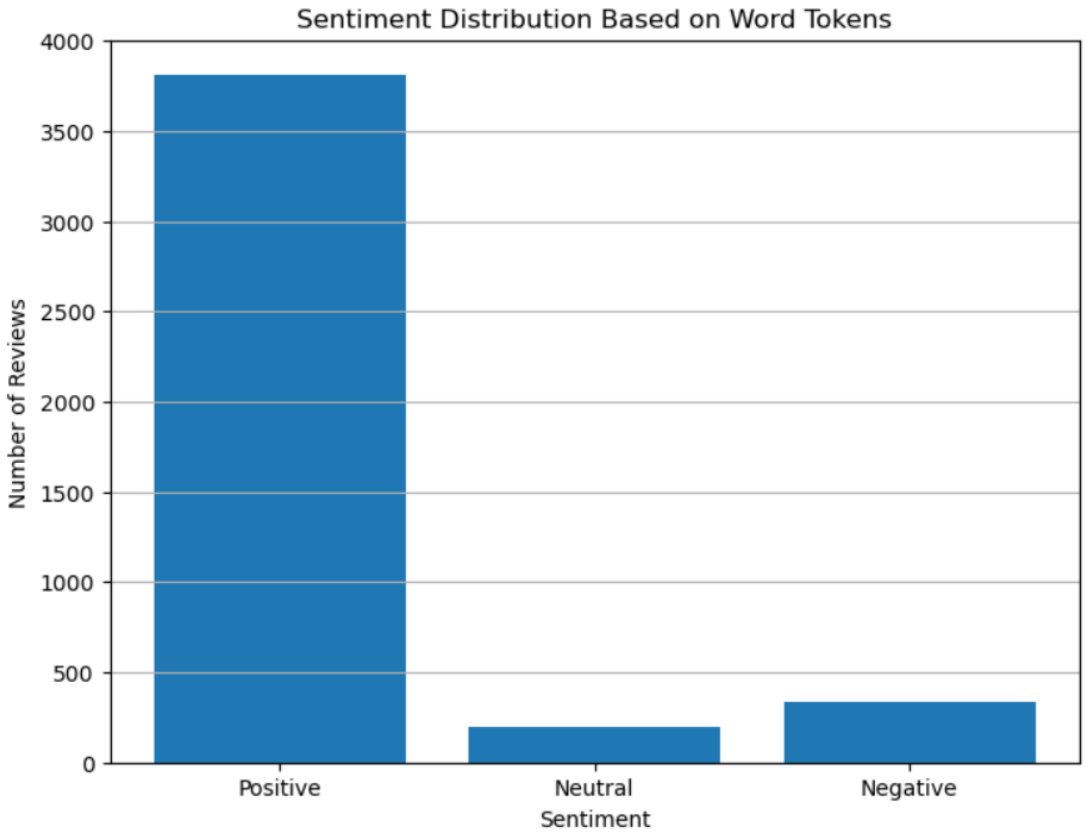


Figure 1: Sentiments Distribution Based on Word Tokens in numbers

VI. Phase 4: Visualizations and Insights

1. Sentiment Distribution (Pie Chart & Bar Chart)

A pie chart and bar chart were created using the sentiment counts:

- Positive: 87.7%
- Negative: 7.7%

- Neutral: 4.6%

This high level of positive shows that buyers are generally satisfied with the headphone products supplied on BestBuy Canada.

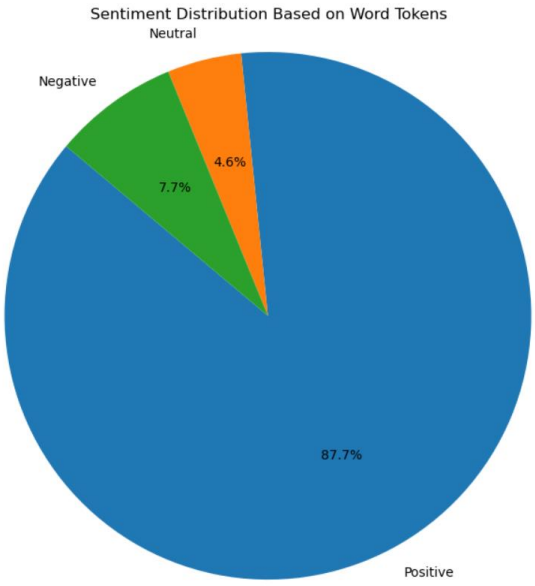


Figure 2: Sentiments Distribution Based on Word Tokens in Percentage

2. Keywords in Positive vs. Negative Reviews

Using frequency analysis:

- Positive Words: "comfortable", "sound", "quality", "noise", "battery", "design"
- Negative Words: "disconnect", "issues", "problem", "uncomfortable", "return", "expensive"

3. Percentage of Positive and Negative Reviews

Only the positive and negative reviews were considered in this visualization to understand polarity:

- Positive: 91.9% of this subset
- Negative: 8.1%

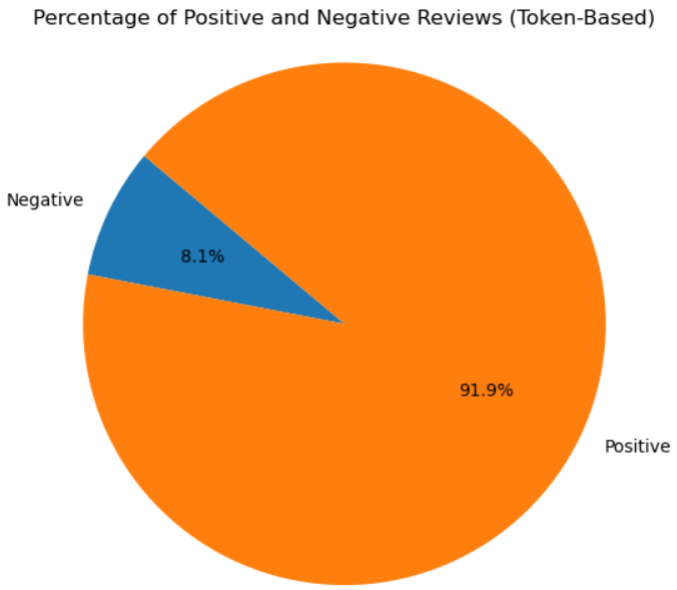


Figure 3: Percentage of Positive and Negative Reviews

4. Products with Most Positive and Negative Feedback

- **Top 5 Positively Reviewed Products:**

  According to the examination of customer sentiment, the following goods earned the most positive feedback (from highest to lowest).

  - Wireless Earbuds Bluetooth 5.3
  - Bluetooth Earbuds Invisible Small Mini Wireless Earbud
  - Skullcandy Grind In-Ear Wireless Earbuds
  - JBL Tune 320BT On-Ear Bluetooth Headphones
  - Bose QuietComfort Ultra In-Ear Wireless Earbuds

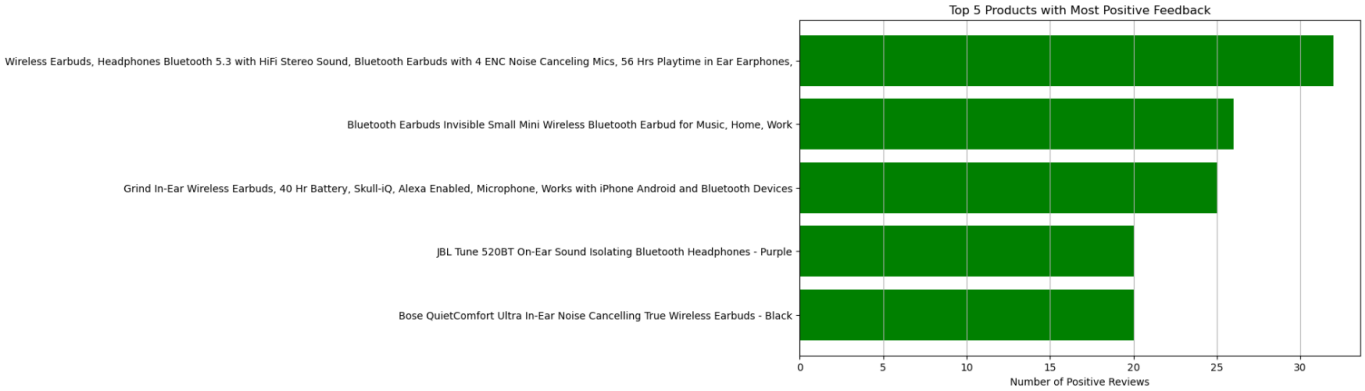


Figure 4: Top 5 Products with Most Positive Feedback

Products with qualities such as long battery life, small design, and good noise cancellation generate the most happiness. While premium brands like Bose and JBL are present, more budget-friendly or adaptable solutions are leading the list, indicating that value and everyday utility are significant drivers of good attitude.

- **Top 5 Negatively Reviewed Products:**

  These products received the most negative feedback:

  - Escape Platinum Bluetooth Sport Earbuds
  - Speck Presidio Pro AirPods Case (Gen 1/2)
  - MEE Audio Connect T1CMA Bluetooth System for TV
  - Samsung Galaxy Buds3 Noise Cancelling Earbuds
  - Apple AirPods (3rd Gen, Open Box)
- Negative evaluations may indicate use-case failures (e.g., TV connectivity or sports), perceived quality difficulties, or displeasure with open-box items. This emphasizes the necessity of setting explicit product objectives and delivering consistent performance.

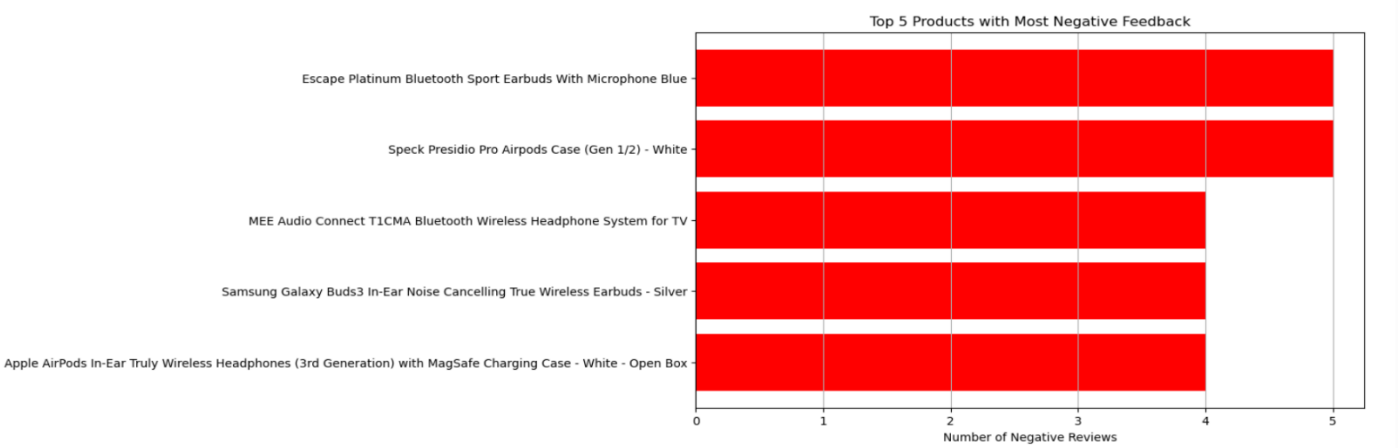


Figure 5: Top 5 Products with Most Positive Feedback

These rankings provide insights into customer satisfaction and dissatisfaction based on product performance**.**

5. Token-Based Word Frequency Analysis and Word Clouds

Goal: To highlight distinct terminology linked with good and negative feedback, filter out overlap and noise.

Process:

- Removed stopwords (both standard and custom).
- Token lists were flattened, and the top 50% of common terms across both feelings were removed.
- Only correct English terms were maintained.

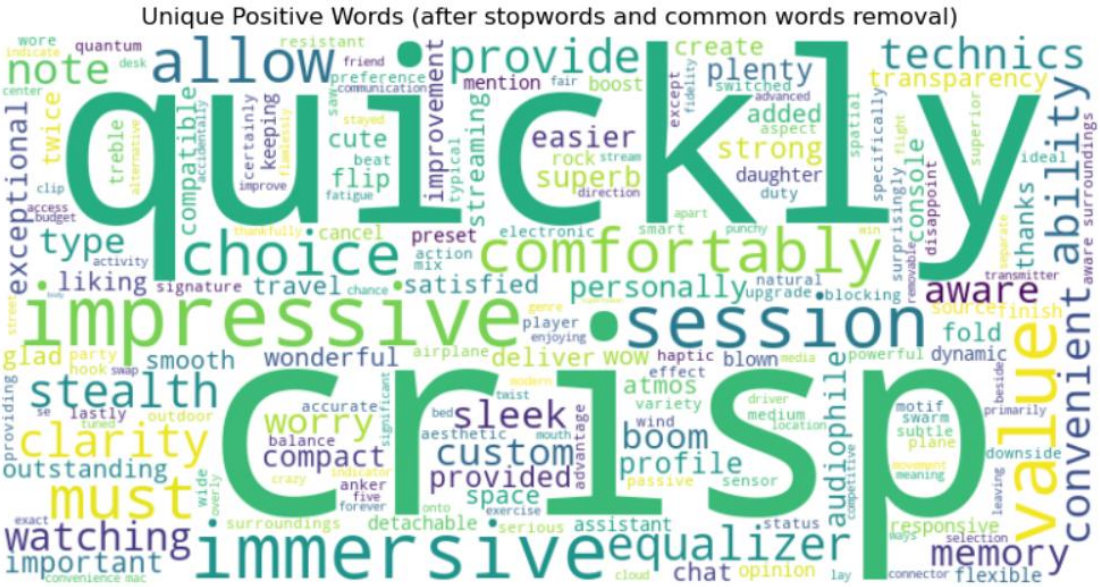
Results:

- Total frequent terms found: 2214.
- Removed the top 50% (1107) of frequent words.
- There are 38991 unique positive tokens and 1780 unique negative tokens.

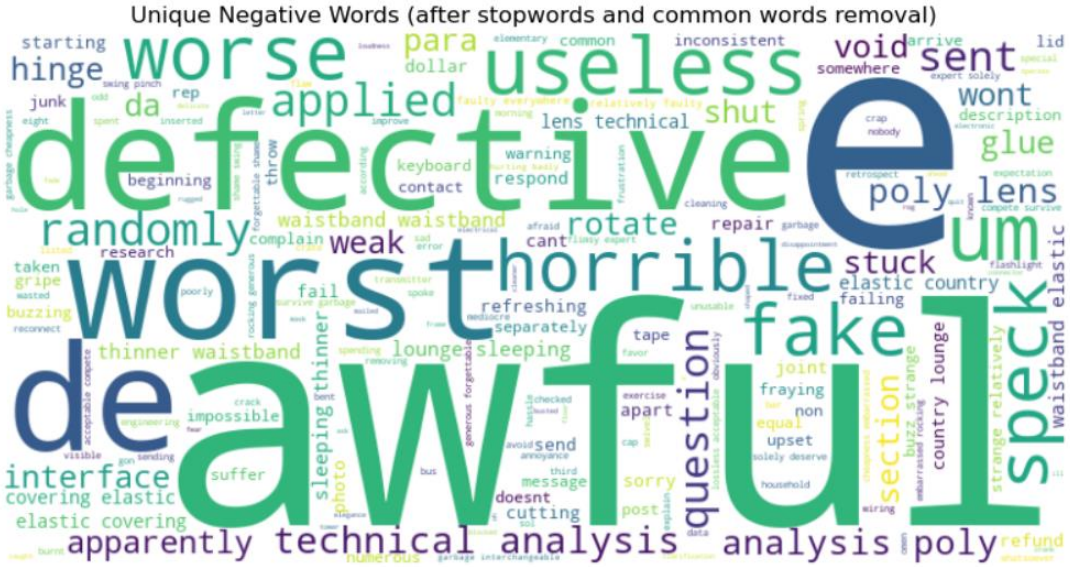
Visualization:

- Two Word Clouds were made to represent the tokens:

    - Positive Word Cloud emphasized features like "clarity", "battery", "convenient", "noise", "impressive".



- Negative Word Cloud showed recurring words like "awful", "inconsistent", "weak", "uncomfortable".



This provides a nuanced understanding of how customers describe products based on their experience.

VII. Key Insights

1. Positive evaluations (87.7%) indicate a solid market position and great product quality among leading headphone brands.
2. Negative reviews primarily focus on technical difficulties such as disconnection, poor fit, and sound quality.
3. Brand Leadership: Bose, Sony, and Apple receive the most positive feedback, whereas Samsung and Skullcandy products are usually criticized.
4. Consistent Sentiment: Products with frequent complaints received multiple bad ratings, indicating systematic flaws rather than individual events.
5. Marketing leverage: Customers prioritize comfort, sound quality, and battery life, making them ideal for targeted advertising.

VIII. Recommendations

1. Product Improvement: Samsung and Skullcandy should address connectivity and comfort difficulties that arise frequently.
2. Support Strategy: Prioritize replying to consumers who frequently report concerns in reviews, such as Bluetooth troubles.
3. Inventory Management: Increase inventory of top-rated models during peak sales periods to match customer demand.
4. Maintain the data pipeline by automating scraping, classification, and reporting.
5. Integrate Feedback Loop: Encourage verified users to post evaluations and establish feedback loops for continual improvement.

IX. Conclusion

This project successfully created a feedback intelligence system from scratch. It entailed scraping raw data, processing text, assessing sentiment, conducting token-based analysis, and extracting actionable insights. The technology demonstrates how text mining and sentiment analysis may help merchants better identify consumer satisfaction patterns and improve their product offerings and service quality.

X. References:

Best buy: Shop online for deals & save. (n.d.). https://www.bestbuy.ca/en-ca/search?search=headphones