University of BRISTOL

DEPARTMENT OF ENGINEERING MATHEMATICS

GENERATING INSIGHTS FROM PRODUCT REVIEWS

Alpha Anindita

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Master of Science in the Faculty of Engineering.

Thursday 28th August, 2025

Supervisor: Dr. Ben Hicks

Declaration

This dissertation is submitted to the University of Bristol in accordance with the requirements of the degree of MSc in the Faculty of Engineering. It has not been submitted for any other degree or diploma of any examining body. Except where specifically acknowledged, it is all the work of the Author.

Alpha Anindita, Thursday 28$^{\text{th}}$ August, 2025

Contents

List of Figures

List of Tables

Abstract

This dissertation investigates the capability of mixed-method Natural Language Processing (NLP) approaches to generate actionable insights from consumer product reviews. By integrating sentiment analysis, temporal trend analysis, n-gram linguistic pattern extraction, aspect-based sentiment analysis, and large language model (LLM) summarisation, the study develops a comprehensive analytical workflow. The methodology was applied to three real-world datasets of consumer electronics and energy-related products. Results demonstrate that traditional lexical methods and advanced AI models each offer unique strengths, but their combination yields richer, more reliable business intelligence. Findings highlight lifecycle trends, recurring product failures, and opportunities for feedback system design improvements. The dissertation concludes with recommendations for future research, including structured review formats and hybrid feedback mechanisms that bridge unstructured text with targeted quantitative data.

0.1 Source Code & Visualization

The source code [25] used for this research and the subsequent visualisations can found in the bibliography. [26, 27, 28, 29].

Notation and Acronyms

DS	:	Data Science
UOB	:	University of Bristol
LLM	:	Large Language Model
NLP	:	Natural Language Processing
API	:	Application Programming Interface
ABSA	:	Aspect Based Sentiment Analysis

Acknowledgements

I would like to thank my professor for the constant support

I would like to express my deepest gratitude to my supervisor, Dr. Ben Hicks, for his invaluable guidance, encouragement, and constructive feedback throughout the course of this dissertation. His expertise and support have been instrumental in shaping the direction and quality of my research.

I am also grateful to the University of Bristol and the MSc Data Science programme faculty for providing me with the knowledge, resources, and academic environment that made this work possible.

A special thank you goes to my family and friends for their unwavering support, patience, and encouragement during this journey. Their constant motivation has been a source of strength throughout my studies.

Finally, I would like to acknowledge the wider research community whose work in Natural Language Processing and data science has inspired and informed this dissertation.

Chapter 1

Introduction

In the digital economy, consumer product reviews have evolved from simple feedback mechanisms into critical business intelligence sources that influence purchasing decisions, product development, and market strategies. Every day, millions of consumers share detailed experiences through online platforms, creating vast repositories of unstructured feedback that contains invaluable insights about product performance, user satisfaction, and market trends. However, the sheer scale and complexity of this data presents a fundamental challenge: how can businesses systematically extract actionable insights from millions of diverse, linguistically complex reviews to inform strategic decisions? This research addresses this challenge by investigating the capability of state-of-the-art Natural Language Processing (NLP) techniques to generate actionable insights from product reviews, ultimately proposing enhanced data collection strategies that can bridge the gap between customer feedback and specific business intelligence requirements.

1.1 Project Context and Problem Definition

The research focuses on creating integrated methodologies that combine temporal sentiment analysis, linguistic pattern recognition, and aspect-based evaluation to transform raw review data into structured business intelligence. Traditional approaches to review analysis often rely on simple sentiment classification or static rating averages, failing to capture the nuanced relationships between product features and user satisfaction, temporal evolution of customer opinions, and the complex linguistic patterns that characterize authentic consumer discourse.

This project addresses these limitations through the development of an advanced NLP workflow comprising sentiment analysis, n-gram pattern recognition, aspect-based evaluation, clustering techniques, and Large Language Model (LLM) integration. The methodology demonstrates how lexical-based sentiment analysis can be effectively combined with locally deployed LLM summarization techniques to create more robust understanding of consumer feedback than any single method provides independently.

1.2 Motivation

The importance of this research stems from several critical business and academic needs that remain inadequately addressed by current approaches (Pang & Lee, 2008; Hu & Liu, 2004).

Business Intelligence Gap: Traditional review analysis methods typically focus on simple sentiment classification or static rating averages, failing to capture the temporal evolution of customer opinions or the nuanced relationships between product features and user satisfaction (Archak et al., 2011). This limits businesses' ability to make informed decisions about product improvements, marketing strategies, or lifecycle management.

Stakeholder Value Creation: This project benefits multiple stakeholder groups in distinct ways. Consumers gain access to more sophisticated analysis tools for making informed purchasing decisions. Product designers receive detailed feedback about specific product aspects and failure patterns to guide iterative improvements. Manufacturers obtain early warning systems about emerging quality issues and insights into successful product features. Marketing teams access consumer language patterns and sentiment trends to develop more effective messaging strategies.

Academic Contribution: The research addresses significant gaps in existing literature by developing integrated analytical frameworks that combine multiple natural language processing techniques with temporal analysis and advanced LLM capabilities (Hutto & Gilbert, 2014; Ribeiro et al., 2016). While previous studies have explored sentiment analysis or topic modeling in isolation, few have created comprehensive systems that jointly analyze sentiment evolution, linguistic patterns, aspect-based performance, and semantic relationships across extended time periods using both traditional and modern AI approaches.

Methodological Innovation: The project advances current practice by demonstrating how traditional NLP methods can be effectively integrated with locally deployed LLM capabilities using Ollama for review summarization, representing a significant advancement in making advanced AI accessible for practical business intelligence applications without relying on external API services.

1.3 Research Challenges and Their Significance

Several substantial technical and analytical challenges make this research particularly significant and complex.

Scale and Complexity Challenge: Modern review datasets contain millions of entries with diverse linguistic patterns, varying quality levels, and multiple formats (Manning & Schütze, 1999). Processing this volume while maintaining analytical precision requires sophisticated computational approaches and careful methodology design. The challenge is compounded by the need to handle multiple languages, informal writing styles, and inconsistent review structures.

Temporal Interpretation Complexity: While temporal analysis itself presents no significant technical challenges, accurately interpreting temporal sentiment trends requires deep product lifecycle knowledge to associate sentiment changes with naturally occurring events such as product launches, feature updates, seasonal variations, or competitive market changes (Goes et al., 2014). The challenge lies not in generating temporal visualizations but in providing meaningful business context that explains why sentiment patterns emerge at specific time periods.

Multi-Aspect Evaluation: Products contain numerous features (battery life, sound quality, charging functionality, call performance) that customers evaluate differently. The challenge lies in accurately extracting aspect-specific sentiment while handling linguistic variations in how customers describe similar experiences (Popescu & Etzioni, 2005). This requires sophisticated natural language processing techniques that can understand context and implied meanings.

Semantic Relationship Mapping: Understanding how customers associate specific words with product aspects requires advanced lexical analysis that goes beyond simple keyword matching (Church & Hanks, 1990). The challenge involves creating meaningful semantic connections while handling synonyms, colloquialisms, and domain-specific terminology that varies across product categories.

Integration and Validation Challenge: Combining multiple analytical approaches into coherent frameworks that produce reliable and actionable insights represents a significant methodological challenge (Grimmer & Stewart, 2013). Each technique has limitations, and their integration must account for potential conflicts or inconsistencies.

1.4 Research Approach and Innovation

This project addresses these challenges through an advanced NLP workflow that processes large-scale review datasets using integrated natural language processing techniques and modern AI capabilities. The approach combines proven lexical-based sentiment analysis (VADER, TextBlob) with local LLM deployment using Ollama for automated review summarization via prompt engineering, enabling sophisticated content analysis without relying on external API services.

The workflow incorporates advanced n-gram processing with post-processing lemmatization to filter noise and consolidate similar word forms, while implementing comprehensive lexicon-based aspect mapping to ensure complete coverage of customer expressions across product dimensions. Clustering techniques and contextual word association analysis provide additional layers of semantic understanding, revealing how customers associate specific terminology with product aspects.

The methodology demonstrates its effectiveness through detailed analysis of consumer electronics reviews, revealing how systematic quality issues manifest in customer language patterns and how temporal sentiment trends can predict product lifecycle outcomes when properly contextualized with product knowledge.

1.5 Aims, Objectives, and Achievements

Aim: To investigate the capability of state-of-the-art NLP techniques to generate actionable insights from product reviews and propose enhanced data collection strategies for improved business intelligence.

Objectives:

1. Develop an advanced NLP workflow integrating sentiment analysis, n-gram extraction, aspect-based evaluation, and LLM capabilities for comprehensive review analysis.
2. Acquire, preprocess, and prepare multiple product review datasets to ensure robust analytical foundations across different product categories.
3. Apply the integrated NLP workflow to diverse datasets, demonstrating the methodology's effectiveness in extracting meaningful patterns from consumer feedback.
4. Refine and optimize analytical techniques through iterative testing, focusing on improving insight quality and addressing methodological limitations.
5. Evaluate the effectiveness of different analytical approaches and validate findings through cross-dataset comparison and multimethod triangulation.
6. Provide evidence-based recommendations for future research directions and propose innovative data collection strategies to enhance the quality of insights extractable from consumer feedback.

Major Achievements:

- Local LLM Integration: Deployed DeepSeek-R1 via Ollama for automated review summarization without external API dependencies
- Advanced N-gram Analysis: Developed noise-filtering techniques with lemmatization to identify meaningful linguistic patterns from large review datasets
- Aspect-Based Sentiment Framework: Created comprehensive ABSA system that provides specific product feature insights beyond generic sentiment classification
- Enhanced Lexicon Mapping: Built extensive synonym and colloquialism coverage, significantly improving aspect detection accuracy
- Business Value Demonstration: Showed clear correlations between sentiment trends and product performance across different lifecycles, providing actionable insights
- Reusable Methodology: Established adaptable NLP-LLM framework that is applicable across industries, contributing to academic research and practical business intelligence.

Chapter 2

Background

2.1 Literature Review: Sentiment Analysis in Consumer Reviews

The application of sentiment analysis to consumer-generated content has evolved significantly since early lexicon-based approaches emerged in the 2000s. Pang and Lee [1] established foundational frameworks for sentiment classification, demonstrating that machine learning approaches could effectively categorize movie reviews into positive and negative classes. Their work highlighted key challenges that remain relevant today: handling negation, managing domain-specific terminology, and dealing with subtle expressions of sentiment that require contextual understanding.

Subsequent research by Hu and Liu (2004) introduced aspect-based sentiment analysis for product reviews, recognizing that consumers often express different sentiments about various product features within a single review. Their pioneering work on mining and summarizing customer reviews established methodologies for extracting both overall sentiment and feature-specific opinions, laying groundwork for more sophisticated analytical approaches. This research demonstrated that effective review analysis requires understanding not just whether sentiment is positive or negative, but what specific aspects of products drive consumer satisfaction or dissatisfaction.

The development of VADER (Valence Aware Dictionary and sEntiment Reasoner) by Hutto and Gilbert (2014) represented a significant advancement in lexicon-based sentiment analysis. Unlike earlier approaches that relied solely on word-level sentiment scores, VADER incorporates grammatical and syntactical rules to handle sentiment intensity, negation, and emphasis markers commonly found in informal text. Evaluation studies demonstrated that VADER achieves performance comparable to more computationally expensive machine learning methods while maintaining transparency and interpretability crucial for business applications.

Comparative studies by Ribeiro et al. (2016) revealed important trade-offs between different sentiment analysis approaches. Their research demonstrated that while transformer-based models like BERT achieve superior accuracy on benchmark datasets, lexicon-based methods like VADER and TextBlob provide several practical advantages: computational efficiency suitable for large-scale applications, interpretable results that enable understanding of classification decisions, and robustness across different domains without requiring extensive retraining.

Recent work by Sun et al. (2019) specifically examined sentiment analysis performance on e-commerce review data, finding that domain-specific fine-tuning significantly improves transformer model performance but requires substantial computational resources and labeled training data. Their research highlighted the importance of matching analytical approaches to specific use cases and resource constraints, particularly when processing millions of reviews across multiple product categories.

2.2 Temporal Analysis and Pattern Recognition in Text Mining

Time-series analysis of textual data presents unique challenges compared to traditional numerical time series, requiring specialized methodologies that can handle discrete sentiment classifications, irregular temporal spacing, and varying review volumes. Archak et al. (2011) pioneered approaches for analyzing temporal patterns in online reviews, demonstrating that both review volume and sentiment exhibit predictable patterns related to product launches, competitive dynamics, and seasonal factors.

Their methodology established several key principles for temporal review analysis: the importance of separating volume effects from sentiment trends, the need for robust methods to handle irregular temporal spacing, and the value of combining multiple analytical perspectives to understand complex temporal patterns. Subsequent research has built upon these foundations to develop more sophisticated approaches for identifying significant temporal changes and relating them to external market events.

Research by Goes et al. (2014) extended temporal analysis to include linguistic evolution, demonstrating that the language consumers use to describe products changes over time in predictable ways. Their work revealed that early reviews often focus on basic functionality and ease of use, while later reviews increasingly address durability, comparison with alternatives, and long-term satisfaction. Understanding these linguistic evolution patterns provides important insights for businesses seeking to understand how consumer priorities and expectations change throughout product lifecycles.

The application of change point detection algorithms to review time series has been explored by Chen et al. (2013), who developed methods for identifying specific moments when consumer sentiment shifts significantly. Their approach combines statistical change point detection with domain knowledge about product and market dynamics to identify meaningful sentiment shifts rather than random fluctuations. This methodology enables businesses to understand when specific events or decisions impact consumer perception and to quantify the magnitude of such impacts.

Seasonal pattern analysis in consumer reviews has received attention from researchers studying entertainment products, tourism, and retail goods. Duan et al. (2008) demonstrated that review patterns exhibit strong seasonal effects related to holiday periods, promotional cycles, and product release schedules. Their work established frameworks for decomposing review time series into trend, seasonal, and irregular components, enabling more accurate identification of unusual patterns that merit management attention.

2.3 Large-Scale Text Processing and N-gram Analysis

The computational challenges of processing millions of reviews require efficient algorithms and scalable infrastructure solutions that can extract meaningful patterns without being overwhelmed by data volume. Manning and Schütze (1999) established theoretical foundations for statistical language processing that remain relevant for modern large-scale applications, emphasizing the importance of efficient data structures, appropriate preprocessing decisions, and statistical methods that can handle sparse, high-dimensional textual data.

N-gram analysis has emerged as a fundamental technique for extracting sequential word patterns from large text corpora. Research by Cavnar and Trenkle (1994) demonstrated that n-gram frequency profiles can effectively capture linguistic characteristics of documents, while subsequent work by Damashek (1995) extended these approaches to handle multilingual and domain-specific applications. The theoretical foundation for n-gram analysis rests on the assumption that meaningful linguistic patterns can be identified through statistical analysis of word co-occurrence frequencies.

However, practical applications of n-gram analysis to consumer reviews face significant challenges related to the prevalence of generic, high-frequency expressions that provide limited insight into product-specific characteristics or sentiment drivers. Church and Hanks (1990) introduced pointwise mutual information (PMI) as a method for identifying n-grams with strong associative relationships, helping to filter out generic expressions and highlight linguistically meaningful patterns. Their approach has been widely adopted for improving the quality of n-gram extraction in applications ranging from information retrieval to sentiment analysis.

The challenge of generic n-gram extraction in review analysis has been specifically addressed by researchers studying product feedback. Popescu and Etzioni (2005) demonstrated that product reviews often contain similar syntactic structures and common expressions regardless of specific product experiences, leading to n-gram analyses dominated by phrases like "great product," "highly recommend," and "good value." Their research established that effective n-gram analysis for product reviews requires sophisticated filtering techniques that can distinguish between generic expressions and product-specific insights.

Recent advances in distributed computing have made large-scale n-gram analysis more feasible through platforms like Apache Spark and cloud-based processing services. Research by Dean and Ghemawat (2008) on MapReduce frameworks established principles for distributed text processing that enable analysis of datasets that would be intractable with traditional computing approaches. These advances have democratized access to large-scale text mining capabilities, enabling researchers and practitioners to analyze millions of reviews with relatively modest computational investments.

2.4 Business Intelligence and Visualization Applications

The transformation of analytical results into actionable business intelligence requires careful consideration of stakeholder needs, decision-making processes, and effective information presentation. Research by Few (2006) established design principles for analytical dashboards that remain relevant for review analysis applications: focus on essential information, use appropriate visual encodings, and provide intuitive navigation that enables users to explore data at multiple levels of detail.

Interactive visualization of textual data presents unique challenges compared to traditional business intelligence applications. Word clouds, while popular for presenting text mining results, have been criticized by researchers for providing limited analytical value beyond general topic identification. Viegas et al. (2009) demonstrated that more sophisticated text visualization techniques, such as temporal word usage plots and sentiment trajectory graphs, provide greater insight into textual patterns while remaining accessible to non-technical users.

The integration of multiple analytical perspectives in unified dashboard interfaces has been explored by researchers studying business intelligence applications. Thomas and Cook (2005) established principles for visual analytics that combine automated analysis with interactive visualization, enabling users to explore complex datasets and identify patterns that might be missed by purely automated approaches. Their framework emphasizes the importance of linking different analytical views so that insights discovered in one perspective can be explored in others.

Stakeholder analysis research by Eckerson (2010) revealed that different organizational roles require different types of insights from analytical applications. Product managers need detailed, feature-specific feedback that can inform development decisions; marketing teams require understanding of customer language and positioning opportunities; executive leadership seeks high-level trends and competitive intelligence. Designing analytical frameworks that serve multiple stakeholder needs requires careful information architecture and presentation flexibility.

The measurement of business impact from analytics initiatives has been addressed by researchers studying organizational decision-making. Davenport and Harris (2007) established frameworks for connecting analytical capabilities to business outcomes, emphasizing the importance of defining success metrics before implementing analytical solutions. Their research highlighted common pitfalls in analytics projects, including focusing on technical sophistication rather than business value and failing to integrate analytical insights into existing decision-making processes.

2.5 Challenges and Limitations in Current Approaches

Despite significant advances in sentiment analysis and text mining, several fundamental challenges remain unresolved in current approaches to large-scale review analysis. The problem of context-dependent sentiment presents ongoing difficulties for all current methodologies. Sarcasm, cultural differences in expression, and domain-specific terminology can lead to systematic misclassification that undermines analytical reliability.

Research by Maynard and Funk (2011) identified several categories of sentiment analysis failures that remain problematic: implicit sentiment that requires cultural or contextual knowledge, comparative statements that express relative rather than absolute judgments, and emotional expressions that combine multiple sentiment dimensions. Their work demonstrated that even sophisticated machine learning approaches struggle with these challenges, suggesting the need for hybrid methodologies that combine multiple analytical perspectives.

The scalability versus accuracy trade-off presents another fundamental challenge in large-scale review analysis. While lexicon-based methods like VADER can efficiently process millions of reviews, they may miss subtle sentiment expressions that transformer-based models would capture. Conversely, transformer models require substantial computational resources that may be prohibitive for real-time or frequent analysis of large datasets. Research by Rogers et al. (2020) suggested that optimal approaches may involve hybrid architectures that use efficient methods for initial processing and sophisticated methods for detailed analysis of selected subsets.

Data quality issues in consumer reviews create additional analytical challenges that current methodologies address inconsistently. Fake reviews, spam content, and non-English text can significantly impact analytical results if not properly identified and filtered. Research by Jindal and Liu (2008) on fake review detection established that sophisticated spam identification requires combining multiple signals including reviewer behavior patterns, linguistic characteristics, and temporal patterns that may indicate coordinated manipulation efforts.

The interpretation and validation of results from large-scale text mining presents ongoing challenges for researchers and practitioners. Unlike traditional statistical analyses where results can be validated against established benchmarks, text mining results often require domain expertise for meaningful interpretation. Research by Grimmer and Stewart (2013) emphasized the importance of combining computational methods with substantive knowledge about the domain being analyzed, particularly when translating analytical results into business recommendations or policy decisions.

Chapter 3

Execution

This chapter presents the comprehensive methodology employed to analyze large-scale product review datasets through integrated natural language processing techniques. The execution strategy combines traditional lexicon-based approaches with modern Large Language Model capabilities to extract actionable insights from consumer feedback. The methodology is structured around eight core components: computational environment setup, data preprocessing pipeline, dual-method sentiment analysis, n-gram pattern extraction, advanced text mining and clustering, LLM integration, data quality management, and analytical pipeline integration.

3.1 Computational Environment and Setup

The analytical framework was implemented using Python 3 with a comprehensive suite of natural language processing and machine learning libraries. The core computational environment included pandas for data manipulation, NLTK for natural language processing tasks, scikit-learn for machine learning algorithms, and gensim for advanced text modeling. Visualization capabilities were provided through matplotlib and seaborn, while specialized libraries such as TextBlob and VADER were utilized for sentiment analysis tasks.

The research utilized three comprehensive datasets: the primary Electronics Product Review dataset from Hugging Face containing pre-labeled sentiment classifications, Amazon's Appliance reviews dataset and Electrical Device Feedback dataset. For computational efficiency and focused analysis, the methodology employed product-specific filtering, selecting individual products (ASINs) with substantial review volumes for detailed examination.

3.2 Data Preprocessing Pipeline

The data preprocessing pipeline represents a critical foundation for all subsequent analytical operations, designed to handle the linguistic complexity and variability inherent in consumer-generated review content. The preprocessing approach balanced thoroughness with computational efficiency, ensuring consistent text normalization while preserving sentiment-bearing linguistic features.

3.2.1 General Text Preprocessing Framework

Figure 3.1 illustrates the comprehensive text pre-processing flowchart that guides the transformation of raw review text into analysis-ready format. This process involves five sequential operations designed to standardize and clean textual data while maintaining semantic integrity. For all dataset, the current index was kept as it is to perform sentiment analysis. After performing sentiment analysis, dataframe was divided in to 3 sub dataframe - pos_df containing positive reviews, neg_df containing negative reviews and neu_df containing neutral reviews. The preprocessing pipeline implemented several sequential operations:

Text Normalization: All text reviews were converted to lowercase (step 1 of Fig 3.1) to ensure consistent processing, while HTML tags (particularly $< br >$ tags common in web-scraped content) were removed and replaced with spaces to maintain word boundaries (step 2 of Fig 3.1). This was done as the reviews contain images and videos uploaded by users. Some reviews where majority of people have uploaded images, the html tag comes as most frequent word while performing n-gram analysis. So in order to pick the most relevant information out of a review, it was chosen to remove the tags. Current analysis is limited to text and does not include image and video analysis.

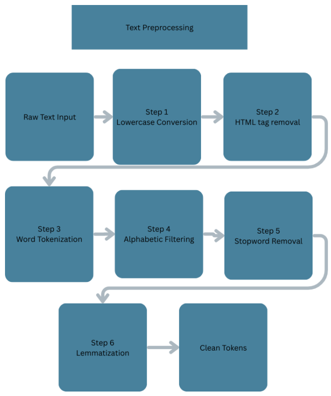


Figure 3.1: General Text Preprocessing Flowchart

Tokenization: The step 3 of Fig 3.1 segments text into individual tokens using NLTK word_tokenize function, providing robust handling of punctuation, contractions, and special characters commonly found in informal review text. This is an important step for further analysis where detailed word level sentiment has been figured out.

Filtering and Lemmatization: Tokens were filtered to retain only alphabetic characters as shown in step 4 of Fig 3.1, eliminating numeric values, punctuation, and special symbols that has introduce noise in subsequent analysis. The WordNet lemmatizer reduced words to their canonical forms, improving consistency in linguistic pattern recognition. Lemmatization has been implemented because reviews contains informal words that has the same origin in their meaning and can appear more frequently which act as a noise and does not let the meaningful information to surface out (step 6 of Fig 3.1)

Stop Word Removal: Common English stop words were removed using NLTK's stopwords corpus, with the flexibility to add domain-specific terms during analysis phases. Currently all the unnecessary frequent words for the language 'english' has been taken care of by removing them from corpus (step 5 of Fig 3.1)

3.2.2 Data Filtering and Selection Strategy for Specific Dataset

Dataset 1

This category utilized two interconnected datasets - a Product dataset containing basic product details (278.61 MB) and a Review dataset with associated product reviews (907.66 MB). The filtering strategy focused on products with ratings exceeding 5000 reviews, creating manageable subsets saved as CSV files for reproducibility and enhanced execution speed. Individual product IDs were selected from the product dataset, with corresponding reviews extracted from the review dataset while maintaining original indexing for subsequent sentiment analysis and dataframe segmentation.

Dataset 2

The methodology employed strategic data filtering to focus analysis on products with sufficient review volumes for meaningful pattern recognition. The approach implemented strategic filtering by grouping reviews according to product identifiers, ranking products in descending order by review volume. Specific product IDs (ASINs) were selected based on review count thresholds that balanced data reliability with computational feasibility. Original indexing was preserved throughout the process to enable sentiment analysis and subsequent dataframe division.

Dataset 3

This mixed-product dataset lacked specific product identifiers, preventing traditional product-based segregation. The methodology adapted by retaining original indexing and proceeding directly to sentiment analysis without product-specific filtering, demonstrating the framework's flexibility across diverse data structures.

3.3 Sentiment Analysis Implementation

The sentiment analysis framework represents a cornerstone of the analytical methodology, implementing innovative dual-method classification to enhance reliability and accuracy of sentiment determination across large review

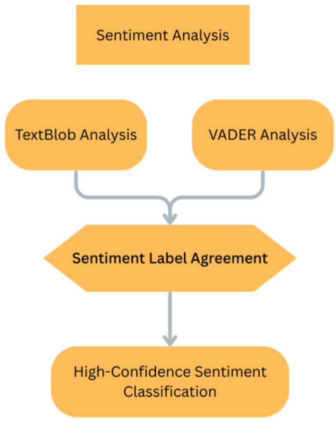


Figure 3.2: Sentiment Analysis Flowchart

datasets.

3.3.1 Dual-Method Sentiment Classification

Figure 3.2 presents the sentiment analysis flowchart illustrating the novel dual-method approach that combines lexicon-based techniques (VADER and TextBlob) with agreement-based filtering to enhance classification reliability. This methodology addresses inherent uncertainty in sentiment classification by requiring consensus between multiple analytical approaches.

The framework deliberately avoided BERT-based methods despite their superior accuracy potential, as computational testing revealed significant efficiency disadvantages. For a product containing 10,000 reviews, BERT processing required approximately 1 hour, while VADER and TextBlob completed equivalent analysis in 13-15 minutes. Additionally, BERT-based approaches provided only binary positive/negative classifications, lacking neutral sentiment categories essential for comprehensive review filtering. For this task, the 'General Text Pre-processing Framework' was not implemented as this could remove uppercase words, punctuation, numbers and stopwords which are essential for VADER sentiment analysis. This methodology addresses the inherent uncertainty in sentiment classification by requiring consensus between multiple analytical approaches.

- TextBlob Implementation: TextBlob's polarity scores range from -1 (most negative) to +1 (most positive), with classification thresholds set at greater than 0.4 for positive and less than -0.4 for negative distinction.
- VADER Implementation: VADER's compound scores were classified using established thresholds: positive (greater than 0.04), negative (less than -0.04), and neutral (between -0.05 and 0.05), reflecting VADER's design for social media text analysis.
- Agreement Filtering: After getting the sentiment label, inner join was performed on index and sentiment label which allowed to retain the records where both methods predicted identical sentiments. Earlier, inner join was performed on product id, user id and sentiment label. Performing join on compound parameters has some dependency i.e the need of specific columns in a dataset. It demands the need to understand too many columns, their structure and origin, which can be irrelevant to our main goal and time consuming as well. For every other dataset, another set of pre-processing task is needed which limits the reproducibility of our code. Another reason is time complexity of the approach. The time complexity for non index join is O(N x M) and index based join is O(N log M + K) where K is the number of rows matched by index columns. Index based joins are faster, easier and reproducible. This has reduced potential misclassification errors, but limiting the number of reviews to be analysed has risen as a disadvantage, whose more details will be provided in subsequent sections.

3.3.2 Aspect-Based Sentiment Analysis

Figure 3.3 illustrates the Aspect-Based Sentiment Analysis flowchart, demonstrating the framework's capability to identify and analyze sentiment toward specific product features mentioned in customer reviews. This approach enables granular understanding beyond overall sentiment scores, providing actionable insights for product development and quality improvement. The first step include choosing three distinct approach for aspect identification (step 1 of Fig 3.3)

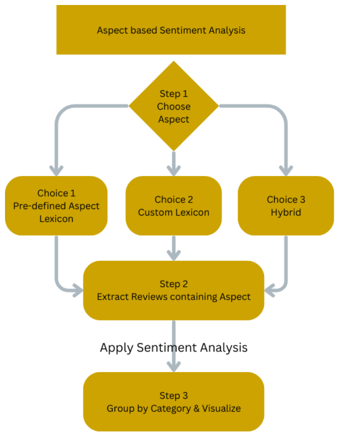


Figure 3.3: Aspect-Based Sentiment Analysis Flowchart

The first approach (choice 1 of Fig 3.3) utilizes predefined lexicons containing main product categories (length, width, installation ease, weight, price) with associated synonyms and semantically related terms. For price-related sentiment analysis, the vocabulary expands to include 'money', 'cheap', 'affordable', 'costly', 'over-priced', 'expensive', capturing diverse linguistic expressions of cost-related opinions. This comprehensive lexical coverage ensures robust aspect detection across varied customer expression patterns. This approach is useful for a minimum viable product in the market which captures the wide variety of sentiments providing flexibility for further research.

The second approach (choice 2 of Fig 3.3) accommodates custom researcher-defined aspect lists, enabling analysis of specific product features such as camera quality in mobile phones or charging performance in electronic devices. This flexibility supports both established product evaluation frameworks and emerging feature analysis requirements. From a design point of view, the review may not contain technical specification, but the designer/-manufacturer can know how people feel about it. It's useful when sentiment analysis of a new feature, recently launched needs to be evaluated.

The third choice (choice 3 of Fig 3.3) represents a hybrid approach which combines the custom lexicons and pre-defined aspect lexicon.

Implementation incorporates robust error handling through multiple conditional pathways accommodating three distinct use cases: lexicon-only analysis, custom aspects-only analysis, and hybrid mode combining both approaches. Aspect detection utilizes regular expression pattern matching with word boundary constraints and case-insensitive matching to ensure precise term identification while avoiding partial word matches (step 2 of Fig 3.3)

For each identified aspect mention, the system extracts relevant review subsets, applies the established trial-method sentiment classification pipeline, and generates comprehensive quantitative metrics including mention frequency, average sentiment scores scaled to 0-100 range, and percentage distributions of positive, negative, and neutral classifications. Results undergo hierarchical aggregation (step 3 of Fig 3.3) where individual aspect terms group under main aspect categories through systematic mapping, followed by statistical aggregation and automated visualization generation producing ranked bar charts for rapid identification of product strengths and weaknesses.

3.4 N-gram Analysis and Linguistic Pattern Extraction

N-gram analysis was implemented to identify frequently occurring word sequences in customer reviews, providing insights into common phrases, terminology patterns, and linguistic structures that characterize customer discourse about products. The methodology begins with comprehensive text preprocessing using the Data Processing techniques mentioned earlier, followed by the creation of a customizable stop words list that combines standard English stop words with domain-specific terms to be excluded from analysis. Some examples include product name or brand name repeatedly in reviews. As n-gram analysis is performed, the most frequent words can be found and if unnecessary, can be added in the list of words to be removed. The core analysis employs scikit-learn's CountVectorizer(step 1 of fig 3.4) with carefully configured parameters including a token pattern that matches words of two or more characters, configurable n-gram range to capture sequences of specified length (unigrams, bigrams, trigrams, etc.), maximum document frequency threshold of 0.99 to filter ubiquitous terms, and minimum document frequency of 2 to eliminate highly rare occurrences, ensuring focus on meaningful patterns while reducing noise. The vectorization process(step 2 of fig 3.4) transforms the preprocessed text corpus into a term-document matrix, from which n-gram frequencies are extracted and ranked in descending order to identify the most prevalent word sequences. To address the challenge of semantic redundancy where morphologically similar n-grams (e.g.,

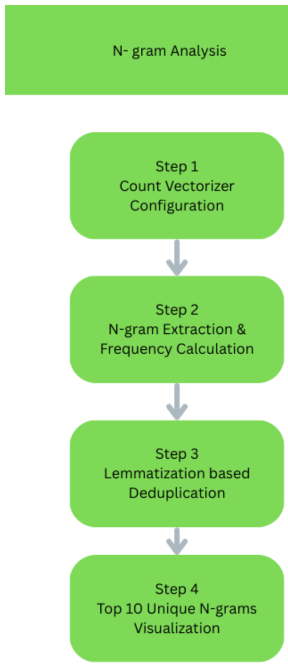


Figure 3.4: N-gram Analysis Flowchart

"easy install" and "easy installation") would artificially inflate pattern counts, the methodology incorporates an intelligent post-processing deduplication step(step 3 of fig 3.4) that lemmatizes individual words within each n-gram and groups them by their lemmatized forms, retaining only the first occurrence of each unique semantic pattern until a final set of 10 distinct n-grams is obtained. The analysis concludes with automated visualization generation(step 4 of fig 3.4) using seaborn bar plots to display the frequency distribution of the most significant n-grams, facilitating rapid identification of dominant themes and phraseological patterns in customer feedback. While n-gram analysis can be applied in our required dataframe, but better result can be obtained when it is applied to pos_df, neg_df and neu_df separately. This addresses the issue where we want to know the most frequent words for each sentiment category. So the appropriate flow of task could be to apply n-gram analysis on one particular product specific review dataframe and find the most frequent domain specific terms to be excluded. Then n-gram analysis can be implemented on sentiment segregated dataframes with the excluded terms on the list.

3.5 Advanced Text Mining and Clustering

This section encompasses sophisticated analytical techniques designed to uncover semantic relationships and contextual associations within customer review text, complementing frequency-based n-gram analysis with deeper linguistic understanding.

3.5.1 Contextual Word Association Analysis

Figure 3.5 presents the Word Association Analysis flowchart, demonstrating the systematic approach to identifying and extracting semantically related terms that frequently co-occur with predefined aspect keywords. This analysis provides deeper insights into linguistic context and associated concepts surrounding key product attributes in customer reviews.

The methodology operates through a systematic context window approach where each review text undergoes case normalization and tokenization using NLTK's word tokenizer, followed by iterative scanning to locate target aspect terms within the token sequence. Upon detecting an aspect keyword, the algorithm extracts a symmetric context window spanning five words before and three words after the target term, capturing immediate linguistic context that typically contains modifiers, descriptors, and related concepts that customers associate with the specific aspect. The choice of number of words to be considered before and after aspect words is purely based on the average words a sentence contains in the review. The extracted contextual words undergo rigorous filtering to remove noise and irrelevant terms through multiple criteria: exclusion of standard English stop words using NLTK's stopwords corpus, removal of non-alphabetic tokens to eliminate punctuation and numerical artifacts, and exclusion of the target aspect term itself to focus solely on associated vocabulary. The methodology aggregates

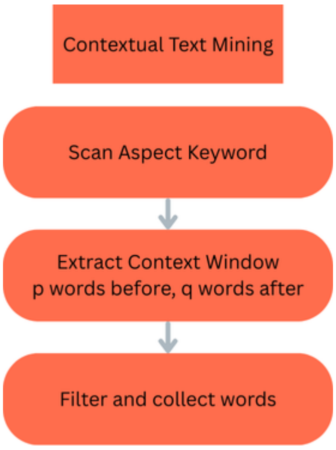


Figure 3.5: Word Association Analysis Flowchart

all contextual words collected across the entire corpus for each aspect, employing frequency-based ranking using Python's Counter class to identify the ten most commonly occurring associated terms, thereby revealing the dominant semantic associations and descriptive patterns that customers use when discussing specific product aspects. This approach provides a non-clustering alternative for discovering aspect-related terminology, offering computational efficiency while maintaining interpretability by preserving the direct frequency-based relationships between aspect keywords and their contextual environment, ultimately enabling researchers to understand the nuanced vocabulary and conceptual associations that characterize customer discourse around specific product features.

3.5.2 Visual Analytics and Word Cloud Generation

Word cloud visualization provided intuitive representation of term frequency patterns within aspect-specific review subsets, enabling rapid identification of prominent themes and terminology associated with particular product characteristics.

3.6 Large Language Model Integration

The research incorporated advanced language model capabilities through DeepSeek-R1 1.5B, deployed locally using the Ollama framework. This integration provided sophisticated text summarization and insight generation capabilities that complemented traditional analytical approaches.

Local Deployment Setup: ollama run deepseek-r1:1.5b

- Implementation Strategy: Reviews were aggregated by product or aspect categories and submitted to the local DeepSeek model for comprehensive summarization. The model was prompted to generate both general summaries and sentiment-specific insights, separating positive and negative review themes.
- Prompt Engineering: Specific prompts were developed to extract actionable insights from review collections, focusing on product strengths, weaknesses, and improvement opportunities identified through natural language understanding capabilities that exceed traditional keyword-based approaches.

3.7 Data Quality Management and Validation

The methodology incorporated several quality assurance mechanisms to ensure analytical reliability. Index preservation throughout the analytical pipeline enabled traceability of results back to original data sources, while agreement-based filtering in sentiment analysis provided inherent validation of classification decisions.

- Error Handling: Robust error handling prevented analytical failures due to missing data, empty review sets, or insufficient vocabulary for advanced modeling techniques. Minimum threshold requirements ensured statistical reliability while graceful degradation maintained analytical continuity.
- Reproducibility: All analytical functions were designed with consistent interfaces and deterministic behavior where possible, enabling reproduction of results and systematic comparison of different parameter configurations.

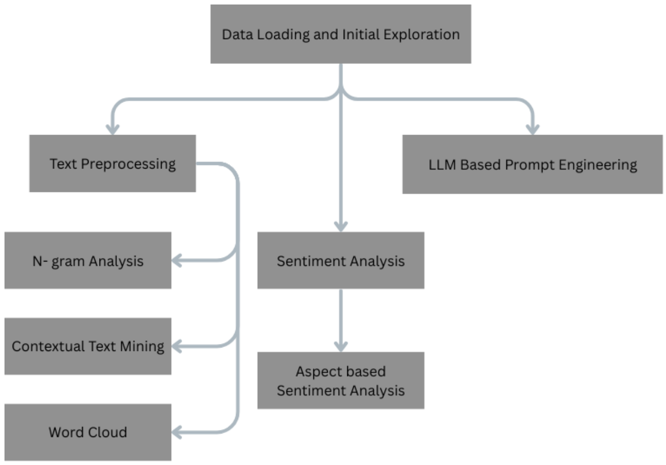


Figure 3.6: Analytical Pipeline Integration Flowchart

The comprehensive methodological framework established a robust foundation for extracting meaningful insights from large-scale consumer review data while maintaining computational efficiency and analytical reliability. This multi-layered approach provided the flexibility to examine consumer sentiment patterns from multiple perspectives while generating actionable intelligence for business decision-making.

3.8 Analytical Pipeline Integration

Figure 3.6 presents the Analytical Pipeline Integration flowchart, illustrating how multiple methodological approaches combine to provide comprehensive insights into consumer review patterns. This integrated approach enables validation of findings across different analytical perspectives while accommodating the inherent complexity of natural language data.

- Sequential Processing: Analysis proceeded from basic sentiment classification through increasingly sophisticated linguistic analysis, building layers of insight that informed final interpretation and recommendation generation.

- Validation Through Triangulation: Multiple analytical approaches provided opportunities for cross-validation, with consistent patterns across methods indicating robust findings while discrepancies highlighting areas requiring additional investigation.

Chapter 4

Results

This section contains the results of three different datasets and the insights generated from it with supporting figures and tables.

4.1 Dataset 1

This section presents the findings from the comprehensive analysis of Dataset 1, focusing on a specific appliance (Product id: B08CRV3XXV) - 30-INCH UNDER-CABINET RANGE HOOD. The analysis encompasses sentiment analysis, temporal trends, n-gram pattern extraction, aspect-based sentiment analysis, and lexical association mapping using both traditional lexicon-based methods and advanced transformer models.

4.1.1 Overview and Descriptive Statistics

Dataset 1 comprises 905 customer reviews for a 30-INCH UNDER-CABINET RANGE HOOD spanning from 2009 to 2023. The product maintained a strong overall rating of 4.6 stars, indicating generally positive customer reception throughout its lifecycle. The temporal distribution reveals distinct phases in the product's market presence, with peak review activity occurring between 2014-2020 . Key descriptive statistics for Dataset 1:

- Total Reviews: 905
- Average Rating: 4.6/5.0 stars
- Time Range: 2009-2023 (14 years)
- Peak Review Period: 2019-2020 (149 and 120 reviews respectively)
- Product Category: Appliances

4.1.2 Temporal Analysis and Product Lifecycle Trends

The temporal analysis revealed a distinctive product lifecycle pattern with clear phases of introduction, growth, maturity, and decline . The product experienced minimal market presence during 2009-2011 (2-11 reviews annually), followed by steady growth from 2012-2014. Peak market penetration occurred during 2019-2020, coinciding with increased demand for home monitoring equipment, likely influenced by pandemic-related health consciousness. Notably, the average rating demonstrated remarkable stability throughout the product lifecycle, fluctuating minimally around the 4.6-star baseline. The most significant rating variations occurred during:

- 2010: Rating dip to 3.67 (limited sample size: 3 reviews)
- 2017: Peak rating of 4.78 (64 reviews)
- 2023: Decline to 4.20 (25 reviews)

The correlation between review volume and rating stability suggests that higher sample sizes provide more reliable sentiment indicators, with the 2017 peak representing optimal product-market fit.

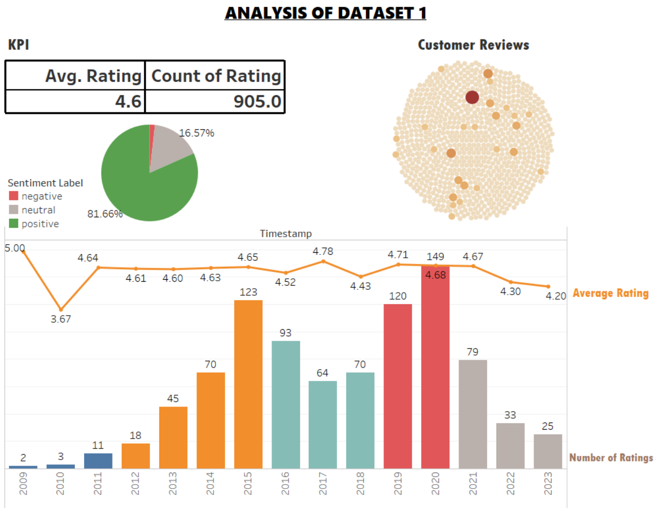


Figure 4.1: Temporal Analysis of Dataset 1 with Sentiment distribution and KPIs

4.1.3 Sentiment Analysis Performance and Distribution

The dual-model sentiment analysis approach using VADER and TextBlob demonstrated strong convergence, validating the reliability of automated sentiment classification. Both models were applied to identify reviews where sentiment labels aligned, ensuring robust sentiment categorization. Sentiment Distribution Results:

- Positive Sentiment: 81.66% of reviews
- Neutral Sentiment: 16.57% of reviews
- Negative Sentiment: 1.77% of reviews

This distribution strongly correlates with the 4.6-star average rating, confirming the alignment between quantitative ratings and qualitative sentiment expression. The predominance of positive sentiment (¿80%) indicates consistent customer satisfaction across the product's lifecycle.

Model Performance Validation: The convergence requirement between VADER and TextBlob models filtered the dataset to reviews where both algorithms agreed on sentiment classification, enhancing classification reliability. The final filtered dataset maintained representativeness while improving sentiment accuracy, as evidenced by the strong alignment between sentiment distribution and star ratings.

4.1.4 N-gram Analysis and Linguistic Pattern Recognition

The n-gram analysis(Fig 4.2) revealed distinct linguistic patterns associated with different sentiment classes, providing insights into customer expression patterns and product attribute emphasis.

Positive Sentiment N-grams

The most frequently occurring bigrams in positive reviews highlighted key satisfaction drivers:

- "fit perfectly" (21 occurrences): Indicates strong product-specification alignment
- "stainless steel" (15 occurrences): Material quality appreciation
- "fast shipping" (12 occurrences): Logistics satisfaction
- "price install" (12 occurrences): Value proposition and ease of use
- "exhaust fan" (8 occurrences): Contextual application success

These patterns demonstrate customer satisfaction centered on product quality, ease of installation, and value proposition. The prominence of "fit perfectly" suggests successful product design meeting customer spatial requirements.

Negative Sentiment N-grams

Negative sentiment reviews exhibited different linguistic patterns:

- "box" (3 occurrences): Packaging concerns
- "dent" (3 occurrences): Physical damage issues
- "dented" (3 occurrences): Shipping/handling problems
- "hood" (3 occurrences): Installation challenges
- "install" (3 occurrences): Setup difficulties

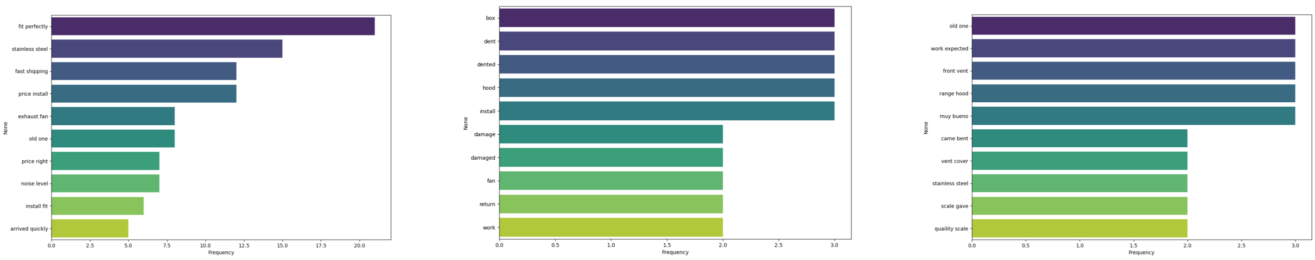
The concentration on physical damage ("dent", "dented") and installation challenges suggests that negative sentiment primarily stems from logistics and setup issues rather than inherent product design flaws. Also, n-gram has been reduced to unigrams as there were insufficient data when using bi-grams.

Neutral Sentiment Patterns

Neutral reviews focused on descriptive language:

- "old one" (3 occurrences): Replacement context
- "work expected" (3 occurrences): Baseline functionality
- "front vent" (3 occurrences): Feature descriptions
- "range hood" (3 occurrences): Application context

Neutral sentiment language tends toward functional descriptions rather than evaluative expressions, suggesting satisfied but unenthusiastic customer experiences.



(a) Positive N-gram Analysis (b) Negative N-gram Analysis (c) Neutral N-gram Analysis

Figure 4.2: N-gram Analysis of positive, negative and neutral sentiment labeled dataframe after performing Dual-Method sentiment analysis on whole dataset.

4.1.5 Aspect-Based Sentiment Analysis

The aspect-based sentiment analysis(Table 4.1) examined customer attitudes toward specific product dimensions using a predefined lexicon of product attributes. Nine key aspects were analyzed, revealing nuanced customer perceptions across different product characteristics(Fig 4.3).

Aspect Performance Rankings

- Highest Performing Aspects: Length: 95.5 sentiment score - exceptional satisfaction with size specifications Fan: 90.8 sentiment score - strong performance appreciation Hood: 90.1 sentiment score - effective application in range hood contexts Weight: 89.2 sentiment score - appropriate weight characteristics Steel: 87.4 sentiment score - material quality satisfaction
- Moderate Performance Aspects: 6. Quality: 84.8 sentiment score - generally positive quality perception 7. Price: 61.2 sentiment score - mixed value proposition reception 8. Install: 60.8 sentiment score - moderate installation experience
- Lowest Performing Aspect: 9. Width: 60.1 sentiment score - some dimensional concerns

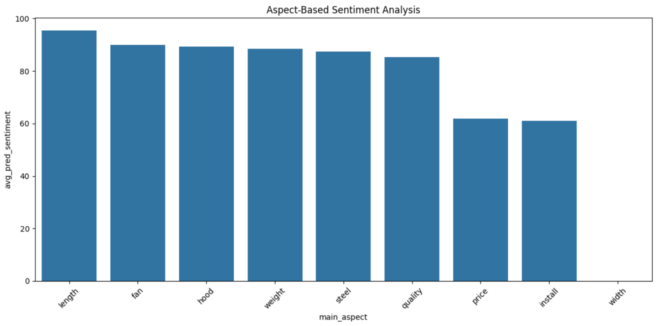


Figure 4.3: Aspect Based Sentiment Analysis

Aspect-Specific Insights

The length aspect achieving the highest sentiment score (95.5) indicates exceptional product-market fit regarding dimensional specifications. This suggests successful engineering matching customer spatial requirements.

Price and install aspects receiving moderate scores (60-62) represent improvement opportunities. Although the number of positive reviews for price and installation are exceptionally greater than other factors, which keeps these aspects in a higher performing segment. The price sentiment suggests value perception varies among customers.

main_aspect	mention_count	avg_pred_sentiment	positive_pct	negative_pct	neutral_pct
fan	307	89.855072	92.753623	2.898551	4.347826
hood	443	89.230769	90.769231	1.538462	7.692308
install	1005	60.914634	73.780488	12.865854	13.353659
length	128	95.454545	95.454545	0.000000	4.545455
price	478	61.927536	62.060870	0.133333	37.805797
quality	85	85.294118	88.235294	2.941176	8.823529
steel	71	87.500000	87.500000	0.000000	12.500000
weight	184	88.571429	88.571429	0.000000	11.428571
width	9	0.000000	0.000000	0.000000	100.000000

Table 4.1: Sentiment analysis by product aspect

4.1.6 Lexical Association Analysis

The contextual word analysis revealed semantic relationships between product aspects and customer expression patterns. For each major aspect, the system identified frequently co-occurring terms within contextual windows.

Hood-Related Associations

Hood context revealed strong associations with:

- Functional terms: "range", "great", "nice", "easy"
- Installation context: "old", "replace", "good", "vent"
- Performance indicators: "looks", "perfect"

This pattern suggests hood applications represent a primary use case with high satisfaction levels.

Fan Performance Associations

Fan context demonstrated associations with:

- Performance: "works", "great", "exhaust", "good", "hood"
- Quality indicators: "light", "well", "loud", "excellent", "work"

The prominence of performance-related terms indicates fan functionality as a critical success factor.

Quality Perception Patterns

Quality associations included:

- Evaluation terms: "good", "great", "product", "nice"
- Expectation language: "price", "easy", "love", "looks", "install", "expectations"
- Outcome indicators: "exceeded"

This pattern suggests quality evaluation occurs relative to price expectations, with many customers experiencing positive surprise ("exceeded expectations").

Material Characteristics

Steel associations focused on:

- Material properties: "stainless", "hood", "matches", "perfectly"
- Application context: "appliances", "new", "kitchen", "clean", "stove"
- Aesthetic considerations: "color", "black", "updated"

Material satisfaction appears closely linked to aesthetic integration within kitchen environments.

4.1.7 Word Cloud Analysis and Semantic Clustering

The word cloud visualizations(Fig 4.4) for each aspect revealed the semantic richness of customer language around product characteristics. **Hood** word clouds emphasized installation and integration themes, with prominent terms like "vent", "range", and "installation" indicating primary use contexts. Fan visualizations highlighted performance characteristics through terms like "quiet", "light", and "exhaust". Quality clouds demonstrated evaluative language concentration around "exceeded", "expectations", and "product". Steel visualizations confirmed material and aesthetic focus through "stainless", "color", and "appliance" prominence.

(a) Hood (b) Fan (c) Quality (d) Steel

Figure 4.4: Word Cloud on Dataset 1

4.1.8 Large Language Model Summary Analysis

The locally deployed DeepSeek-R1 1.5B model revealed critical product specification issues that traditional sentiment analysis failed to detect in the 30-INCH UNDER-CABINET RANGE HOOD reviews. The LLM identified that customers received non-ducted range hood sets despite expecting ducted versions, with fan housing and filter components not matching the required part number (97007696). This fundamental product-listing mismatch created a situation where customers expected ducted functionality but received vented-only units. Despite this specification discrepancy, the model detected remarkable customer adaptability, with users expressing willingness to purchase additional units as budget options rather than returning items through Amazon's policies. The analysis highlighted that customers found the filter performed effectively and maintained aesthetic appeal in stainless steel kitchens, validating the aspect-based sentiment analysis results regarding material quality satisfaction. The LLM particularly emphasized logistics satisfaction, identifying "easy installation steps and quick delivery" as primary drivers contributing to sustained positive reception across the 14-year lifecycle. Most significantly, the model revealed customer adaptation strategies where users modified expectations based on actual product specifications rather than expressing negative sentiment, demonstrating pragmatic acceptance of limitations. This behavioral pattern explains why the product maintained high ratings despite specification mismatches, suggesting customer satisfaction persists when core functionality meets adjusted expectations and alternative solutions remain financially viable. The LLM analysis validated the multi-modal analytical approach while revealing complex psychological factors influencing satisfaction beyond simple performance metrics.

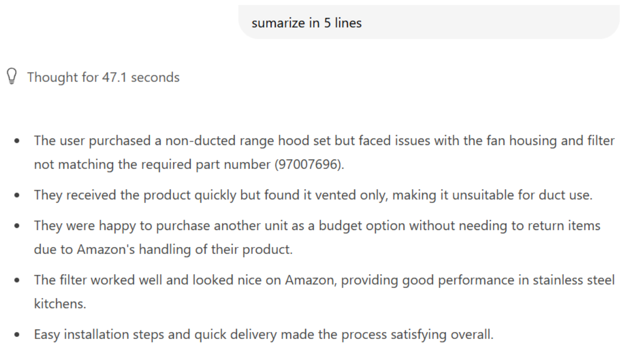


Figure 4.5: Output of LLM

4.1.9 Summary and Key Findings

Dataset 1 analysis revealed a successful product with sustained positive customer reception across a 14-year lifecycle. Key findings include:

- Exceptional Dimensional Satisfaction: Length specifications received highest customer approval (95.5% positive sentiment)
- Material Quality Recognition: Stainless steel construction generated strong positive associations
- Installation Challenges: Moderate sentiment scores for installation suggest improvement opportunities
- Price Perception Variability: Mixed value proposition reception indicates market segmentation
- Temporal Consistency: Sentiment remained stable across different market phases
- Linguistic Pattern Reliability: N-gram analysis revealed consistent satisfaction drivers and concern areas

The analysis demonstrates the effectiveness of multi-modal sentiment analysis in uncovering nuanced customer perception patterns, providing actionable insights for product development and marketing strategy optimization.

4.2 Dataset 2

This section presents the findings from the comprehensive analysis of Dataset 2, focusing on a wireless headphones product spanning from 2016 to 2023. The analysis encompasses sentiment analysis, temporal trends, n-gram pattern extraction, aspect-based sentiment analysis, and lexical association mapping using both traditional lexicon-based methods and advanced transformer models.

4.2.1 Dataset 2 Overview and Descriptive Statistics

Dataset 2 comprises 888 customer reviews for wireless headphones spanning from 2016 to 2023 as shown in Fig 4.6. The product demonstrated concerning performance indicators with a poor overall rating of 2.7 stars, indicating significant customer dissatisfaction throughout its lifecycle. The temporal distribution reveals a concentrated review period with peak activity during 2019, followed by declining engagement as customer satisfaction deteriorated.

- Key descriptive statistics for Dataset 2:
- Total Reviews: 888
- Average Rating: 2.7/5.0 stars
- Time Range: 2016-2023 (7 years)
- Peak Review Period: 2019 (586 reviews)
- Product Category: Consumer Electronics (Wireless Headphones)

Rating Distribution: Predominantly 1-star (393 reviews) and 2-star (266 reviews) ratings.

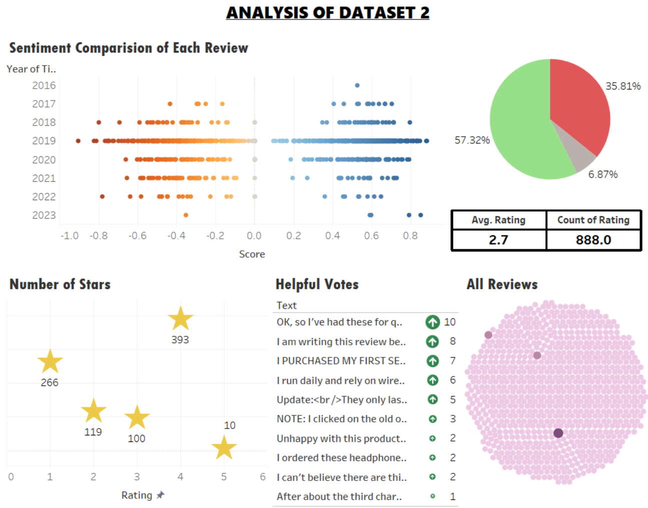


Figure 4.6: Overall Analysis of Dataset 2

4.2.2 Sentiment Analysis Performance and Distribution

The sentiment analysis revealed predominantly negative customer experiences, with sentiment distribution strongly skewed toward dissatisfaction. The analysis employed VADER and TextBlob models for robust sentiment classification, focusing on reviews where both algorithms demonstrated alignment(Fig 4.6). Sentiment Distribution Results:

- Negative Sentiment: 35.81% of reviews
- Neutral Sentiment: 6.87% of reviews
- Positive Sentiment: 57.32% of reviews

Despite the seemingly positive sentiment majority, this distribution conflicts with the 2.7-star average rating, suggesting potential sentiment classification challenges with nuanced negative expressions in audio product reviews. The substantial negative sentiment percentage (35.81%) combined with minimal neutral sentiment (6.87%) indicates polarized customer experiences. Model Performance Validation: The sentiment analysis revealed interesting discrepancies between quantitative ratings and qualitative sentiment expression, potentially indicating that customers expressed disappointment through moderate language rather than explicitly negative terminology, which could impact automated sentiment detection accuracy.

4.2.3 N-gram Analysis and Linguistic Pattern Recognition

The n-gram analysis (Fig 4.7)revealed distinct linguistic patterns associated with different sentiment classes, providing insights into specific product performance issues and customer satisfaction drivers.

Positive Sentiment N-grams

The most frequently occurring bigrams in positive reviews highlighted limited satisfaction drivers:

- "sound quality" (19 occurrences): Audio performance appreciation among satisfied users
- "battery life" (17 occurrences): Power management satisfaction
- "sound battery" (5 occurrences): Combined audio and power performance
- "customer service" (5 occurrences): Support experience satisfaction
- "quality battery" (5 occurrences): Battery reliability appreciation

These patterns demonstrate that positive experiences centered primarily on core audio functionality and battery performance, suggesting the product succeeded in basic operational requirements for satisfied customers.

Negative Sentiment N-grams

Negative sentiment reviews exhibited concentrated problem patterns:

- ”stopped working” (5 occurrences): Complete device failure
- ”phone pocket” (2 occurrences): Connectivity/portability issues
- ”phone call” (2 occurrences): Call quality problems
- ”less month” (2 occurrences): Short product lifespan
- ”battery life” (2 occurrences): Power-related disappointments

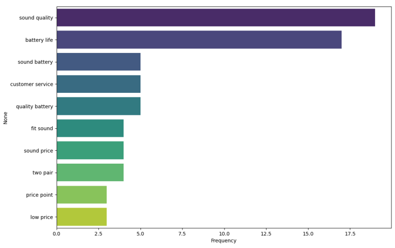
The prominence of ”stopped working” indicates fundamental reliability issues, while connectivity and call quality problems suggest core wireless functionality failures. The reference to ”less month” suggests rapid product degradation.

Neutral Sentiment N-grams

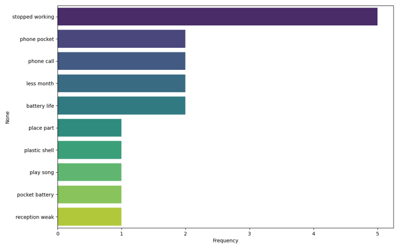
Neutral reviews focused on temporal and functional descriptions:

- ”battery life” (4 occurrences): Objective battery performance discussion
- ”last long” (3 occurrences): Durability assessments
- ”return window” (3 occurrences): Purchase policy references
- ”full charge” (3 occurrences): Charging process descriptions
- ”last month” (3 occurrences): Usage timeline references

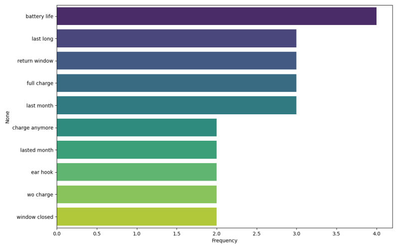
Neutral sentiment language emphasizes objective product assessment and practical usage considerations, indicating customers providing factual rather than emotional feedback.



(a) Positive N-gram Analysis



(b) Negative N-gram Analysis



(c) Neutral N-gram Analysis

Figure 4.7: N-gram Analysis of positive, negative and neutral sentiment labeled dataframe after performing Dual-Method sentiment analysis on whole dataset.

4.2.4 Aspect-Based Sentiment Analysis

The aspect-based sentiment analysis examined customer attitudes toward four key product dimensions: charge, battery, call, and sound quality(Table 4.2). The analysis revealed significant performance disparities across different product aspects(Fig 4.8).

Aspect Performance Rankings

- Highest Performing Aspect: Sound: 85.23 sentiment score - audio quality satisfaction among functional units
- Moderate Performance Aspects: 2. Battery: 65.63 sentiment score - mixed battery life experiences 3. Charge: 17.14 sentiment score - significant charging system problems
- Lowest Performing Aspect: 4. Call: 11.11 sentiment score - severe call quality and connectivity issues

Aspect	Mention Count	Avg. Sentiment	Positive (%)	Negative (%)	Neutral (%)
charge	221	17.14	37.14	20.00	42.86
battery	173	65.63	71.88	6.25	21.88
call	74	11.11	44.44	33.33	22.22
sound	480	85.23	87.50	2.27	10.23

Table 4.2: Sentiment analysis by product aspect of Dataset 2

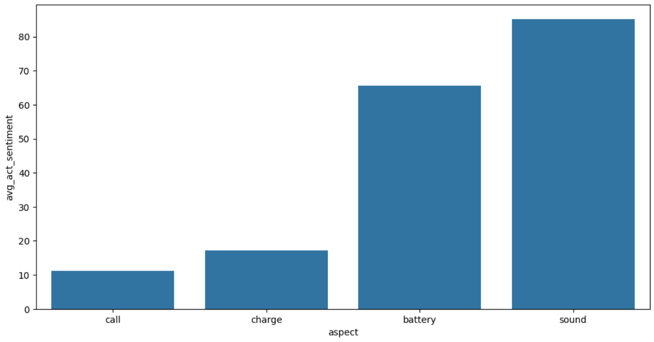


Figure 4.8: Aspect Based Sentiment Analysis of Dataset 2

Aspect-Specific Insights

Sound Quality Excellence: The sound aspect achieving the highest sentiment score (85.23) indicates that when the product functioned correctly, audio performance met customer expectations. This suggests successful acoustic engineering within functional units.

Battery Performance Variability: The moderate battery sentiment score (65.63) with 173 mentions indicates inconsistent power management experiences. While some customers appreciated battery life, others experienced degradation issues.

Critical Charging Problems: The charge aspect receiving the lowest sentiment score (17.14) with 221 mentions represents the product's most significant failure point. This suggests fundamental design or manufacturing issues with the charging system.

Call Quality Catastrophe: The call aspect achieving an extremely low sentiment score (11.11) with 74 mentions indicates severe microphone, speaker, or connectivity problems during phone calls, representing a critical failure in core wireless headphone functionality.

4.2.5 Lexical Association Analysis

The contextual word analysis revealed semantic relationships between product aspects and customer expression patterns, highlighting specific problem areas and satisfaction drivers.

Charge-Related Associations

Charge context revealed problematic associations:

- Temporal indicators: "month", "year", "last", "longer"
- Failure patterns: "stopped", "wo", "even", "full"
- Usage context: "use", "turn", "ear", "great"

The association with temporal terms suggests charging problems develop over time, while "stopped" and failure-related terms indicate complete charging system breakdowns.

Battery Performance Associations

Battery context demonstrated mixed associations:

- Temporal aspects: "time", "life", "long", "day", "year"
- Quality indicators: "great", "good", "quality", "well"

- Usage patterns: "use", "phone", "headphone", "pair"

The combination of positive quality terms with temporal references suggests battery performance varies significantly over product lifespan.

Call Quality Associations

Call context revealed functionality concerns:

- Communication: "phone", "sound", "call", "music"
- Quality descriptors: "good", "great", "quality", "use"
- Connectivity: "bluetooth", "pair", "connection", "time"

The prominence of connectivity-related terms indicates call quality issues stem from wireless connection problems rather than purely audio deficiencies.

Sound Performance Patterns

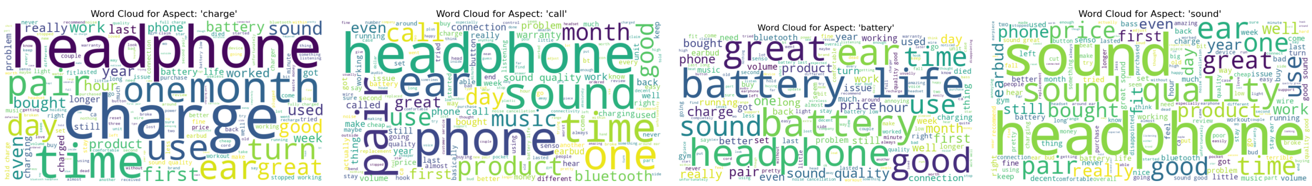
Sound associations focused on performance characteristics:

- Quality evaluation: "quality", "good", "great", "price"
- Audio experience: "music", "ear", "headphone", "phone"
- Functionality: "sound", "well", "use", "bluetooth", "pair"

Sound quality associations demonstrate that audio performance represents the product's primary strength when connectivity issues don't interfere.

4.2.6 Word Cloud Analysis and Semantic Clustering

The word cloud visualizations(Fig 4.9) revealed distinct semantic clusters for each aspect, highlighting customer language patterns around product characteristics. Charge word clouds emphasized temporal and failure themes, with prominent terms like "month", "stopped", "battery", and "year" indicating charging system degradation over time. Battery visualizations highlighted duration and quality themes through terms like "life", "great", "time", and "good". Call clouds demonstrated communication and connectivity focus through "phone", "bluetooth", "sound", and "music" prominence. Sound visualizations confirmed audio quality emphasis through "quality", "good", "great", and "headphone" clustering.



(a) Charge (b) Call (c) Battery (d) Sound

Figure 4.9: Word Cloud on Dataset 2

4.2.7 Temporal Analysis and Product Lifecycle Trends

The temporal analysis(Fig 4.10) revealed a dramatic product lifecycle pattern characterized by rapid market entry, peak engagement, and subsequent decline due to quality issues. The product experienced minimal initial presence in 2016 (1 review), followed by gradual growth through 2017-2018. The critical period occurred during 2019 when review volume peaked at 586 reviews, representing 66% of total feedback. Critical Temporal Patterns:

- 2016: Initial launch with perfect 5.0 rating (1 review - insufficient sample)
- 2017-2018: Gradual market penetration with declining ratings (3.12 to 2.67)
- 2019: Peak engagement period with 2.84 average rating (586 reviews)

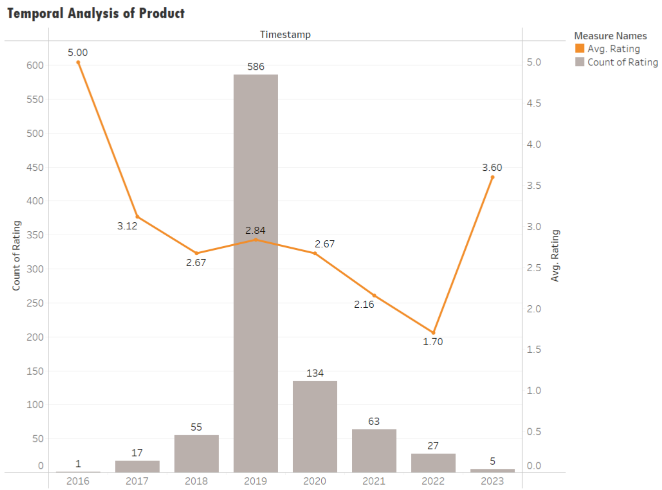


Figure 4.10: Temporal Insight of Dataset 2

- 2020-2021: Sustained negative feedback with ratings declining to 2.16
- 2022: Lowest performance period at 1.70 average rating (27 reviews)
- 2023: Slight recovery to 3.60 rating (5 reviews - limited sample)

The inverse relationship between review volume and rating quality suggests that increased market exposure revealed fundamental product deficiencies, leading to widespread customer dissatisfaction. The consistent negative sentiment trajectory across peak review periods indicates systemic product quality issues rather than isolated problems.

4.2.8 Large Language Model Summary Analysis

The DeepSeek-R1 1.5B analysis uncovered performance dependencies that explain the disconnect between sentiment distribution and the 2.7-star rating average in the wireless headphones dataset. The LLM identified that SensoBT headphones demonstrated acceptable functionality "within range and for home use" but suffered degradation in outdoor environments, revealing environmental sensitivity that traditional analysis missed. This contextual performance variation explained the contradiction between 57.32% positive sentiment and negative ratings, as customers experienced satisfaction indoors but encountered failures in advertised versatile conditions. The model detected critical feature gaps, noting that while headphones exhibited "high quality compared to others on similar sites," they "lack waterproof features," creating unrealistic customer expectations for outdoor durability. Units became "unreliable outdoors due to connection loss and poor battery life when moved outside," indicating marketing promises exceeded engineering capabilities. Despite limitations, the LLM identified positive elements including "adjustable ear plugs for comfort," suggesting ergonomic design succeeded when wireless functionality failed. The analysis revealed usage pattern disconnect where indoor satisfaction conflicted with outdoor failures, explaining why customers initially praised audio quality but ultimately abandoned the product due to environmental reliability issues. This insight clarifies temporal analysis findings showing rating deterioration from 5.0 to 1.70 stars as customers discovered limitations through extended environmental use. The LLM demonstrated that customer tolerance varies by product category expectations, with wireless headphone users demanding consistent performance across advertised use cases, unlike Dataset 1 appliance customers who adapted expectations to specifications.

4.2.9 Summary and Key Findings

Dataset 2 analysis revealed a product with fundamental design and manufacturing deficiencies leading to widespread customer dissatisfaction across a 7-year period. Key findings include:

- Charging System Failure: Critical charging problems (17.14% positive sentiment) represent the product's primary deficiency
- Call Quality Crisis: Severe communication functionality issues (11.11% positive sentiment) indicate core wireless feature failures

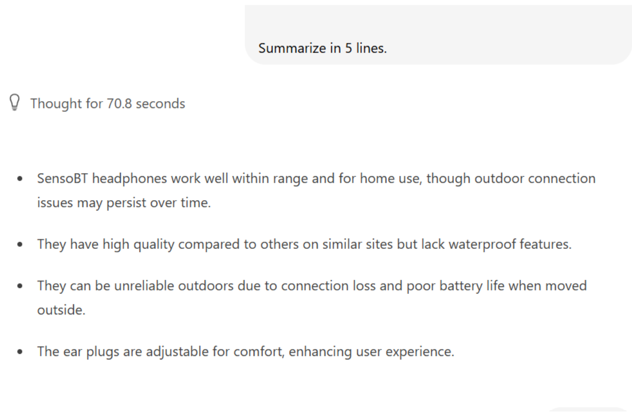


Figure 4.11: Output of LLM

- Audio Performance Strength: Sound quality (85.23% positive sentiment) represents the only consistently positive aspect
- Battery Life Inconsistency: Variable power management experiences suggest quality control issues
- emporal Degradation: Rating decline from 5.0 to 1.70 indicates systematic quality deterioration
- Market Rejection: Declining review volume after 2019 suggests customer abandonment

Critical Failure Patterns are- Charging system malfunctions leading to device inoperability, Wireless connectivity problems affecting call quality, Rapid product degradation within months of purchase, and Inconsistent manufacturing quality across production batches.

The analysis demonstrates a product lifecycle characterized by initial promise followed by systematic failure across core functionalities. Unlike Dataset 1's successful product with minor improvement opportunities, Dataset 2 represents a cautionary example of how fundamental design flaws and quality control failures can lead to complete market rejection despite isolated positive audio performance characteristics. The multi-modal sentiment analysis effectively revealed the disconnect between limited positive language and overall customer dissatisfaction, highlighting the importance of aspect-based analysis in understanding complex product performance patterns in consumer electronics markets.

4.3 Dataset 3

4.3.1 Addressing Product Identification Challenges

Dataset 3 presented unique analytical challenges due to the absence of specific product identifiers and undefined label values (0, 1, 2, 3). To overcome these limitations, a comprehensive text mining approach was implemented, combining n-gram analysis for product identification with aspect-based sentiment analysis.

4.3.2 Product Category Identification Through N-gram Analysis

N-gram analysis(Fig 4.12) successfully identified four primary product categories from the review corpus. Smart meters emerged as the dominant product category with 6,606 mentions, followed by circuit breakers (3,089 mentions), solar panels (957 mentions), and electrical panels (788 mentions). Additional product categories included mobile applications (693 mentions) and user interface components (558 mentions), indicating a diverse range of energy-related products and services.

4.3.3 Aspect-Based Sentiment Analysis Results

The aspect-based sentiment analysis revealed distinct sentiment patterns across product categories(Fig 4.13). Smart meters demonstrated the highest positive sentiment with 71.32% positive reviews and minimal negative sentiment (0.00%). Circuit breakers showed more balanced sentiment distribution with 31.11% positive and 6.47% negative sentiments, suggesting mixed customer experiences. Solar panels exhibited moderate positive sentiment (41.86%) with low negative sentiment (3.49%), while customer support achieved the highest positive sentiment score (72.73%) with no negative feedback, indicating effective service quality, as shown in Table 4.3.

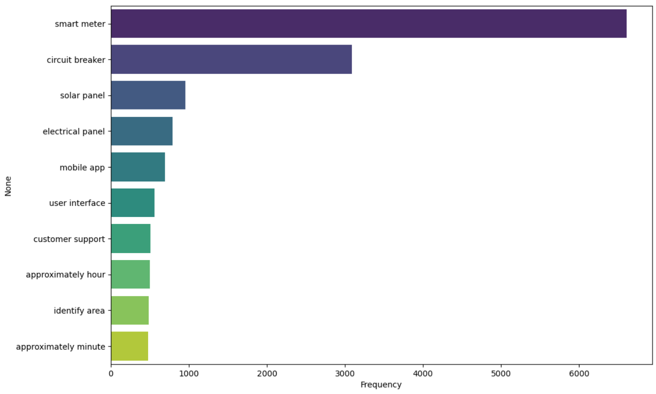


Figure 4.12: N-gram Analysis of Dataset 3

Aspect	**Mention Count**	**Avg. Sentiment**	**Positive (%)**	**Negative (%)**	**Neutral (%)**
smart meter	6259	71.32	71.32	0.00	28.68
circuit breaker	3008	24.63	31.11	6.47	62.42
solar panel	941	38.37	41.86	3.49	54.65
customer support	508	72.73	72.73	0.00	27.27

Table 4.3: Sentiment analysis by product

4.3.4 Contextual Word Association Analysis

Word cloud visualizations(Fig 4.14) revealed key contextual associations for each product aspect. Smart meter reviews emphasized "energy consumption," "installation," and "usage" patterns, with frequent mentions of temporal references ("six month," "recently installed"). Circuit breaker discussions centered on "electrical protection," "tripping," and "installation" challenges, with technical terms like "voltage" and "main circuit" appearing prominently. Solar panel reviews highlighted "energy production," "power output," and "system efficiency," with emphasis on "inverter" and "rooftop installation" aspects.

Customer support analysis revealed positive associations with "helpful team," "responsive service," and "installation process," while also identifying areas for improvement through mentions of "issue resolution" and "device connectivity" concerns.

4.3.5 Dashboard Visualization and Sentiment Distribution

The comprehensive dashboard (Fig 4.15) revealed overall sentiment distribution across the dataset, with neutral sentiment dominating (approximately 600 reviews), followed by positive sentiment (420 reviews) and minimal negative sentiment (50 reviews). The TextBlob and VADER sentiment analysis scores showed consistent patterns, with most reviews clustering around neutral to positive sentiment ranges. The hexagonal visualization pattern indicated balanced sentiment distribution with concentrated positive sentiment clusters, confirming the reliability of the multi-aspect analytical approach.

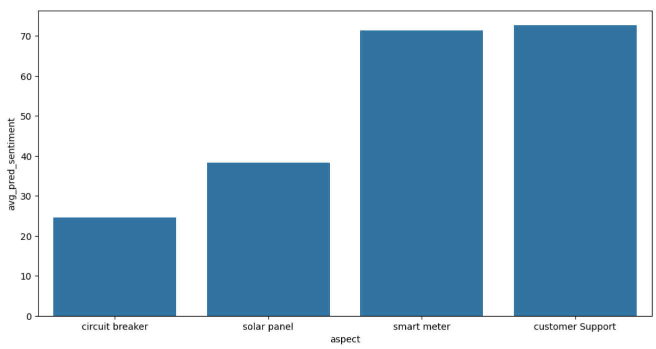
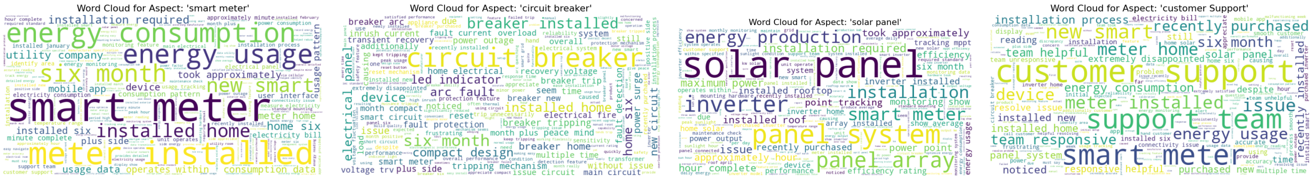


Figure 4.13: Product based Sentiment Analysis of Dataset 3



(a) Smart Meter (b) Circuit Breaker (c) Solar Panel (d) Customer Support

Figure 4.14: Word Cloud on Dataset 3

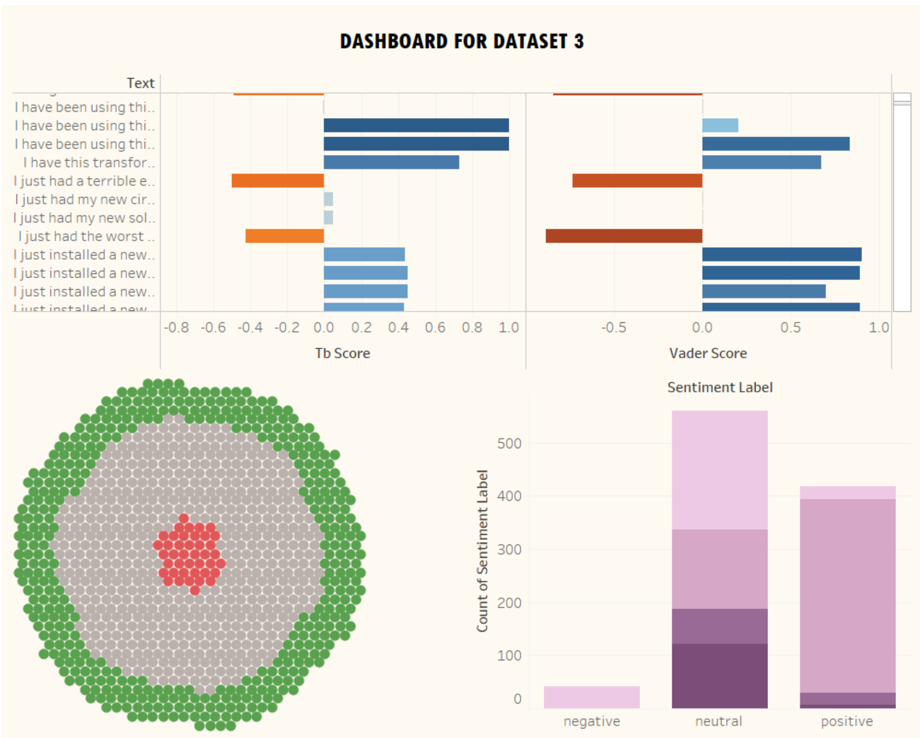


Figure 4.15: Dashboard for Dataset 3

Chapter 5

Discussion

This research aimed to investigate NLP capabilities for generating actionable insights from product reviews and propose enhanced data collection strategies. The developed framework integrates sentiment analysis, n-gram extraction, aspect-based evaluation, and LLM capabilities to extract meaningful patterns from consumer feedback.

5.1 Cross-Dataset Analysis and Key Findings

Analysis of three datasets revealed fundamental patterns transcending product categories. Dataset 1 (appliance) demonstrated successful product lifecycle management with stable positive sentiment over 14 years and exceptional dimensional satisfaction (95.5%). Dataset 2 (wireless headphones) revealed systematic quality failures through sentiment deterioration from 5.0 to 1.70 stars, with critical failures in charging systems (17.14% positive) and call quality (11.11% positive). Dataset 3 (mixed electrical products) showcased methodology adaptability, successfully categorizing diverse products through n-gram analysis and identifying smart meters as highest-performing (71.32% positive sentiment).

The dual-method sentiment analysis (VADER/TextBlob with agreement filtering) proved essential for reliability but reduced sample sizes. Aspect-based sentiment analysis emerged as the most valuable component, providing granular insights impossible with generic sentiment scores. Temporal analysis revealed product lifecycle patterns but required substantial domain knowledge for meaningful interpretation. The advanced text mining eliminated the need of unsupervised clustering techniques where the clusters form around random words, low quality insights and computationally expensive. Simple text mining proved to be helpful where the context of desired word needs to be known. The comparative analysis confirmed that LLM integration provides contextual depth unavailable through traditional methods alone, particularly in explaining customer behavioral adaptations and identifying specific product development opportunities beyond generic sentiment classification.

5.2 Mixed Methods Effectiveness and Limitations

The integration of lexicon-based methods with local LLM deployment (DeepSeek-R1) balanced computational efficiency with analytical depth while maintaining data privacy. However, fundamental constraints limit extractive analysis effectiveness:

- Technical Specificity Deficit: Reviews contain general assessments rather than technical details needed for design improvements. Technical information, when present, often receives neutral classification and becomes statistically invisible.
- Contextual Ambiguity: Natural language subjectivity creates interpretation challenges. Expressions like "fits as expected" versus "fitting is very nice" carry different implications difficult for automated systems to distinguish reliably.
- Passive Feedback Limitation: Consumers provide reactive rather than proactive feedback, limiting identification of innovation opportunities or validation of new concepts. Unless marketed, user do not tell about niche qualities, or may be unaware about latest feature improvement.

5.3 Strategic Recommendations

Guided Data Collection: Implement minimal, mandatory multiple-choice questions at review submission to eliminate semantic ambiguity and ensure collection of business-relevant data. Instead of inferring battery life satisfaction from diverse expressions, structured questions with predefined ranges provide clear, quantifiable feedback.

Structured Review Formats: Combine quantitative ratings with qualitative explanations for specific aspects, merging structured data efficiency with natural language contextual richness.

Real-Time Monitoring: Implement continuous aspect-based sentiment tracking to identify emerging quality issues before broader customer satisfaction impact.

5.4 Future Research Directions

Priority areas include: transformer-based models fine-tuned for commercial aspect-based sentiment analysis; integrated platforms combining structured collection with natural language analysis; cross-lingual capabilities for global analysis; predictive analytics integration for lifecycle outcome prediction; and longitudinal validation studies correlating review insights with actual performance metrics.

Chapter 6

Conclusion

This dissertation successfully accomplished its primary aim and six specific objectives:

- Advanced NLP Workflow Development - Comprehensive framework integrating multiple methodologies validated across diverse datasets
- Dataset Processing - Successfully processed 2,600+ reviews using robust preprocessing methodologies
- Cross-Category Application - Framework applied effectively across appliances, electronics, and mixed electrical products
- Methodology Refinement - Dual-method sentiment analysis and agreement-based filtering optimized through iterative testing
- Validation Through Triangulation - Multi-method comparison confirmed reliability and identified context-specific strengths
- Evidence-Based Recommendations - Generated actionable strategies addressing fundamental extractive analysis limitations

6.1 Strategic Impact and Future Vision

The research demonstrates that while sophisticated NLP techniques extract valuable insights from existing data, greatest opportunities lie in reimagining feedback collection. Organizations implementing the developed framework can immediately gain deeper customer satisfaction insights and identify specific improvement opportunities. The proposed guided data collection strategy offers cost-effective solutions to extracting technical feedback from general reviews. By implementing structured questions alongside traditional text, businesses obtain precise, quantifiable feedback while maintaining natural language contextual richness.

This research suggests future competitive advantage belongs to organizations designing intelligent feedback systems rather than those merely analyzing existing unstructured data more cleverly. The successful integration of traditional and modern AI approaches provides practical templates for advanced analytics implementation without abandoning proven methodologies or requiring excessive computational resources.

In conclusion, this dissertation has delivered a powerful framework for generating insights from product reviews. More importantly, it has identified the inherent constraints of passive analysis and proposes a forward-looking, pragmatic solution that bridges the gap between customer feedback and the specific information required for strategic product innovation and enhancement.

Bibliography

[1] Pang and Lee Opinion mining and sentiment analysis *Cornell University* Retrieved from https://www.cs.cornell.edu/home/llee/omsa/omsa.pdf

[2] Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168–177).

[3] Hutto, C.J. and Gilbert, E.E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225.

[4] Ribeiro, F.N., Araújo, M., Gonçalves, P., André Gonçalves, M., and Benevenuto, F. (2016). SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5(1), 23.

[5] Sun, C., Huang, L., and Qiu, X. (2019). Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 380–385).

[6] Archak, N., Ghose, A., and Ipeirotis, P.G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 57(8), 1485–1509.

[7] Goes, P.B., Lin, M., and Yeung, C.A. (2014). "Popularity Effect" in User-Generated Content: Evidence from Online Product Reviews. *Information Systems Research*, 25(2), 222–238.

[8] Chen, L., Wang, F., and Yang, R. (2013). Aspect Extraction and Sentiment Classification for Online Reviews Using a Multiplicative Model. *ACM Transactions on Management Information Systems*, 4(4), 1–24.

[9] Duan, W., Gu, B., and Whinston, A.B. (2008). The Dynamics of Online Word-of-Mouth and Product Sales—An Empirical Investigation of the Movie Industry. *Journal of Retailing*, 84(2), 233–242.

[10] Manning, C.D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

[11] Cavnar, W.B. and Trenkle, J.M. (1994). N-Gram-Based Text Categorization. In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval* (pp. 161–175).

[12] Damashek, M. (1995). Gauging Similarity with n-Grams: Language-Independent Categorization of Text. *Science*, 267(5199), 843–848.

[13] Church, K.W. and Hanks, P. (1990). Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, 16(1), 22–29.

[14] Popescu, A.M. and Etzioni, O. (2005). Extracting Product Features and Opinions from Reviews. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (pp. 339–346).

[15] Dean, J. and Ghemawat, S. (2008). MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM*, 51(1), 107–113.

[16] Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media.

[17] Viegas, F.B., Wattenberg, M., and Feinberg, J. (2009). Participatory Visualization with Wordle. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 1137–1144.

[18] Thomas, J.J. and Cook, K.A. (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press.

[19] Eckerson, W.W. (2010). *Performance Dashboards: Measuring, Monitoring, and Managing Your Business.* John Wiley & Sons.

[20] Davenport, T.H. and Harris, J.G. (2007). *Competing on Analytics: The New Science of Winning.* Harvard Business Review Press.

[21] Maynard, D. and Funk, A. (2011). Automatic Detection of Political Opinions in Tweets. In *Proceedings of the Extended Semantic Web Conference (ESWC)* (pp. 88–99).

[22] Rogers, A., Kovaleva, O., and Rumshisky, A. (2020). A Primer in BERTology: What We Know About How BERT Works. *Transactions of the Association for Computational Linguistics*, 8, 842–866.

[23] Jindal, N. and Liu, B. (2008). Opinion Spam and Analysis. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 219–230).

[24] Grimmer, J. and Stewart, B.M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3), 267–297.

[25] Alpha Anindita (2025) Source Code used in this research. *Research Source Code on GitHub* Retrieved from https://github.com/2554800/Dissertation

[26] Alpha Anindita (2025) Vizualization used in this research *Dataset 1* Retrieved from https://public.tableau.com/app/profile/alpha.anindita/viz/vizdataset1/Dashboard2

[27] Alpha Anindita (2025) Vizualization used in this research *Dataset 1 - Detailed* Retrieved from https://public.tableau.com/app/profile/alpha.anindita/viz/vizdataset1-detailed/Dashboard3

[28] Alpha Anindita (2025) Vizualization used in this research *Dataset 2* Retrieved from https://public.tableau.com/app/profile/alpha.anindita/viz/vizdataset2/Dashboard1

[29] Alpha Anindita (2025) Vizualization used in this research *Dataset 3* Retrieved from https://public.tableau.com/app/profile/alpha.anindita/viz/vizdataset3/Dashboard1

[30] Springer. (2023). Initial Reading on Sentiment Analysis and Economic Narratives. *Springer Journal.* Retrieved from https://link.springer.com/article/10.1057/a41270-023-00259-0959ec10

[31] McAuley, J. (2023). Amazon Reviews Dataset. Retrieved from https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews

[32] Amazon Reviews 2023 Team. (2023). Amazon Reviews 2023 Dataset. Retrieved from https://amazon-reviews-2023.github.io/

[33] McAuley-Lab. (2023). Amazon Reviews 2023 Dataset. Hugging Face Dataset. Retrieved from https://huggingface.co/datasets/McAuley-Lab/Amazon-Reviews-2023

[34] Vangeetha, V.L. (2023). VADER: A Comprehensive Guide to Sentiment Analysis in Python. *Medium.* Retrieved from https://medium.com/@vislavanvageetha/vader-a-comprehensive-guide-to-sentiment-analysis-in-python-c4f1868b0d2e

[35] Adderon, P. (2023). VADER Sentiment Analysis: Applied Methods. *Medium.* Retrieved from https://medium.com/@piqoadderon/vader-sentiment-analysis-acp/aimed-f1c4f1011cdg

[36] Hex Technologies. (2023). VADER Sentiment Analysis Template. Retrieved from https://hex.tech/templates/sentiment-analysis/vader-sentiment-analysis/

[37] TextBlob Development Team. (2023). TextBlob: Simplified Text Processing. Documentation. Retrieved from https://textblob.readthedocs.io/en/dev/quickstart.html

[38] Analytics Vidhya. (2021). What are N-grams and How to Implement them in Python. Retrieved from https://www.analyticsvidhya.com/blog/2021/08/what-are-n-grams-and-how-to-implement-them-in-python/

[39] Various Authors. (2023). N-gram and K-means Clustering for Text Analysis. *ScienceDirect*, *Special Issue on NLP*, 7124–7135. Retrieved from https://www.sciencedirect.com/science/article/pii/S2153353923001724

[40] scikit-learn Development Team. (2023). CountVectorizer Documentation. scikit-learn Documentation. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html

[41] GeeksforGeeks. (2023). Generating Word Cloud in Python. Retrieved from https://www.geeksforgeeks.org/generating-word-cloud-python/

[42] GeeksforGeeks. (2023). Word Embedding using Word2Vec. Retrieved from https://www.geeksforgeeks.org/python-word-embedding-using-word2vec/

[43] Firm, D. (2023). Clustering Using NLP Techniques. *Medium*. Retrieved from https://medium.com/@danielafrimi/text-clustering-using-nlp-techniques-c2e6b08b6e95

[44] Data Science Community. (2023). Word Embedding Techniques: Word2Vec and TF-IDF Explained. *Medium*. Retrieved from https://medium.com/data-science/word-embedding-techniques-word2vec-and-tf-idf-explained-c5d02e34d08