

A Comprehensive Multiplatform Analysis Framework Leveraging LLMs for the Buyability

Index: A Consumer-Centric Approach to Informed Purchasing Decisions

Ananya Barjatiya

Abstract

In today's digital marketplace, the average consumer is inundated with fragmented product information, making informed purchasing decisions increasingly challenging. This paper presents a comprehensive, consumer-centric framework designed to aggregate, analyze, and synthesize data from multiple review platforms, including Amazon, Google Reviews, Twitter, and Reddit, to generate a unified metric known as the Buyability Index. The Buyability Index is a composite score that combines various product performance metrics commonly considered by consumers. These metrics include quality, popularity, customer service, value for money, and environmental impact, providing a single, easily understandable measure of a product's overall desirability. The system uses advanced natural language processing (NLP) techniques powered by large language models (LLMs) to perform detailed sentiment analysis and compute parameter-specific scores. These individual scores are then dynamically weighted and aggregated to provide consumers with an intuitive tool that reflects real-world product performance, affordability, and ethical considerations. The proposed framework is highly scalable, with a modular design that facilitates continuous data ingestion, real-time analysis, and rapid updates based on evolving consumer preferences. By addressing everyday concerns such as product affordability, durability, service quality, and environmental responsibility, the framework fills a critical gap in online product evaluation and empowers consumers to make more informed, data-driven decisions. Furthermore, by consolidating disparate review data into a single, real-time metric, the framework significantly reduces the time consumers spend researching products, streamlining the path to swift and confident purchase decisions.

Keywords: digital marketplace, sentiment analysis, natural language processing, large language models

Introduction

The explosion of online shopping and digital marketplaces has fundamentally changed the way consumers make decisions and purchase products. However, the wealth of available information is often disorganized, leaving the average consumer overwhelmed and uncertain about product quality and value. Traditional review systems that rely solely on star ratings or isolated customer testimonials do not provide the nuanced insight required to navigate these complex environments. In response to this challenge, I propose a comprehensive multiplatform analysis framework that leverages the latest advances in natural language processing (NLP) and large language models (LLMs) to produce an integrated metric, the Buyability Index.

This framework is designed with the consumer in mind. It aggregates diverse data sources to capture the multifaceted nature of product evaluation. Not only does it assess traditional aspects such as product quality and customer service, but it also addresses concerns that matter most to everyday users: affordability, durability, ease of use, and even ethical considerations like environmental responsibility. By synthesizing these diverse dimensions into a single, easy-to-understand index, the system provides a powerful tool for anyone seeking to make smarter purchasing decisions in an increasingly competitive market. This paper is organized into several key sections.

- Background and Literature Review
- Methodology
- Implementation and System Design
- Results
- Conclusion and Future Work

Background and Literature Review

Traditional Review Systems and Their Limitations

Traditional online review systems, such as those found on Amazon and Google Reviews, have long been the go-to source for consumer insights. However, these systems often provide only a narrow view of product performance. Star ratings and isolated customer reviews, while useful, frequently suffer from issues such as bias, lack of context, and even manipulation through fake reviews. Research by Hu et al. [1] and Luca [2] has demonstrated that the overreliance on simple numerical ratings can lead to a distorted understanding of product quality, particularly when negative reviews are underreported or when reviews are manipulated by bots or fraudulent practices.

The Evolution of Sentiment Analysis Techniques

Sentiment analysis has evolved significantly over the past decade. A decade ago, sentiment analysis primarily meant the basic classification of text as positive, negative, or neutral using rule-based, lexicon-driven methods, whereas today it is commonly used to explain a sophisticated, context-aware process that captures subtle emotional nuances and opinions by leveraging advanced transformer-based models. Early techniques, such as lexicon-based methods and traditional machine learning classifiers, provided a basic understanding of consumer sentiment but struggled with context, sarcasm, and domain-specific language. The advent of deep learning models, and particularly transformer-based architectures like BERT (Devlin et al. [3]) and GPT (Radford et al. [4]), has revolutionized the field by enabling more nuanced and accurate sentiment classification. These models can capture the complexities of natural language, including subtle contextual cues that earlier models missed. Recent studies (e.g., Zhang et al. [5]) have shown that LLMs can significantly improve sentiment analysis accuracy, even in datasets with noisy, unstructured text from social media platforms.

Integration of Multiple Data Sources

While significant progress has been made in sentiment analysis, few frameworks have attempted to integrate data from multiple sources into a single evaluative metric. Studies have explored the integration of structured data (e.g., star ratings, verified purchase indicators) with unstructured data (e.g., text reviews and social media comments) but have typically focused on a single domain or platform. For instance, a study by Chiu et al. [6] combined structured variables like vital signs with unstructured clinical notes to predict patient outcomes, demonstrating that such integration improved predictive accuracy. Another study, by Grande-Ramírez et al. [7] proposed a novel approach that integrated sentiment analysis of social media into strategic planning, highlighting the value of synthesizing diverse data sources for comprehensive insights. Building upon this prior research, my work develops a unified system that synthesizes insights from diverse sources, including e-commerce sites, social media platforms, and review aggregators. This approach not only enhances the overall accuracy of the analysis but also ensures that the resulting Buyability Index reflects a holistic view of consumer sentiment and product performance.

The Consumer's Need for a Comprehensive Tool

For the average consumer, having a single, comprehensive tool to evaluate products is not merely a convenience, it is a necessity. In an era of information overload, consumers need reliable, data-driven insights to make informed decisions quickly. Research emphasizes the growing demand for integrated decision-support systems in e-commerce, highlighting the limitations of fragmented information. For example, the study by Zong et al. [8] discusses the development of key factors influencing the enhancement of decision-support systems using fuzzy logic theory, aiming to improve e-commerce decision-making. The framework addresses this need by providing a Buyability Index that consolidates multiple dimensions of

product evaluation into a single, user-friendly score. This empowers consumers to compare products based on real-world performance metrics rather than relying solely on star ratings or subjective reviews.

Methodology

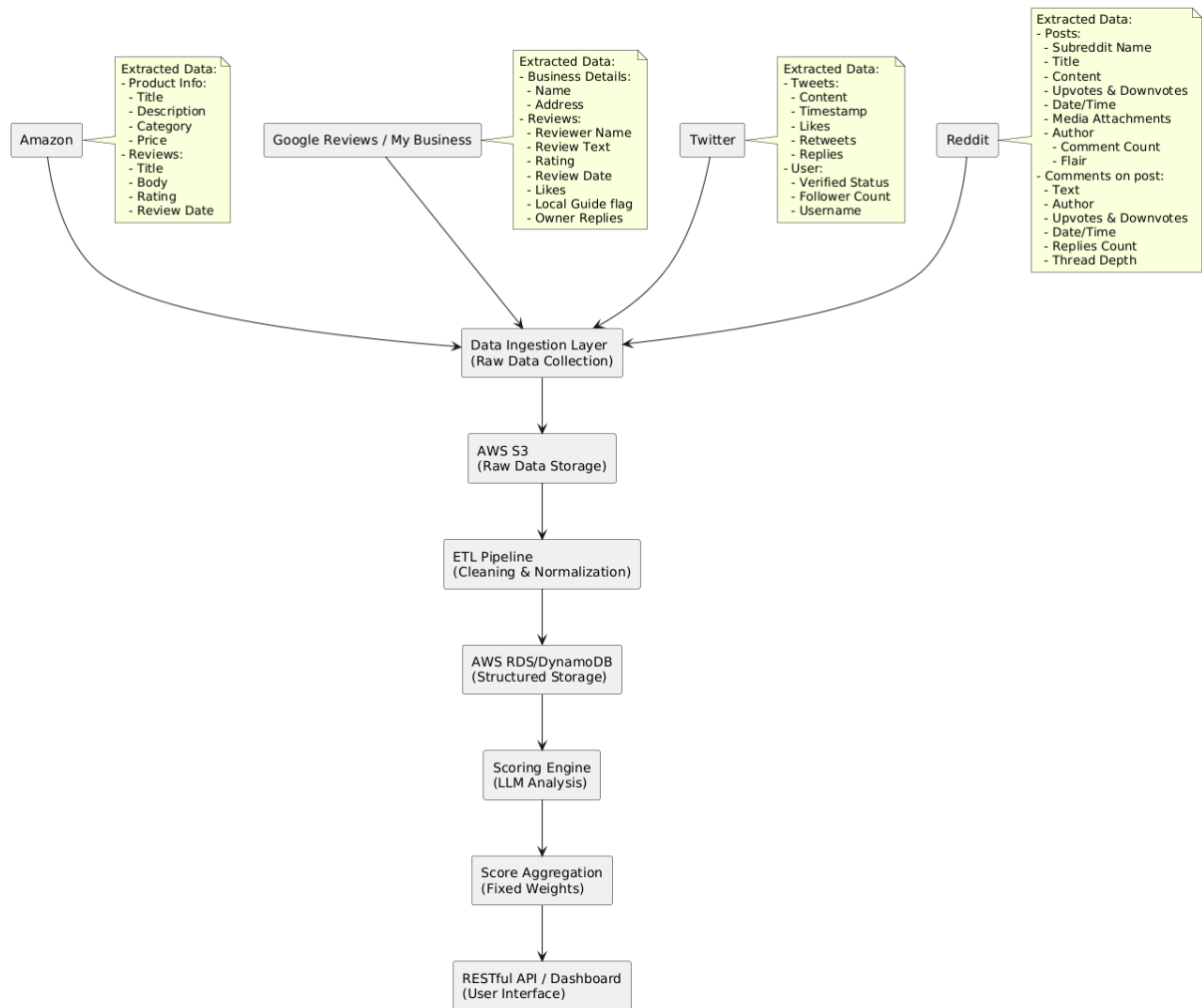
This section describes the conceptual design of the framework in detail. This system is organized into several interdependent stages, each of which plays a critical role in transforming raw, heterogeneous data into a coherent, consumer-centric Buyability Index.

Data Sources and Collection

I begin by identifying and integrating multiple data sources relevant to product evaluation. The primary data sources include:

- Amazon: Product descriptions, customer reviews, and ratings obtained via a combination of web scraping and the Amazon Product Advertising API.
- Google Reviews and Google My Business: Structured business details and review data accessed through the Google Places API and web scraping techniques.
- Twitter: Real-time public sentiment and engagement data, retrieved using Twitter's Search and Streaming APIs.
- Reddit: Community discussions, posts, and comments extracted using the Reddit API.

To ensure comprehensive data collection, I developed custom scripts and API integrations that adhere to each platform's rate limits and data policies.



Data Storage and Preprocessing

The data is collected, and stored in a cloud-based data lake (AWS S3) to ensure scalability and durability. A relational database (AWS RDS/DynamoDB) is then used to store preprocessed data for efficient querying and analysis.

Preprocessing steps:

1. Data Cleaning:

Raw data is cleaned using a hybrid approach that combines custom Python scripts with AI-powered techniques. This ensures the removal of duplicates, spam, and irrelevant content from various platforms such as Amazon, Google Reviews, and Reddit.

- Script-Based Cleaning:

- I developed Python scripts that leverage libraries like BeautifulSoup for HTML tag stripping and regular expressions (regex) to remove URLs.
- AI-Powered Cleaning for Low-Quality or Bot-Generated Reviews:

In addition to script-based cleaning, it uses a supervised machine learning classifier, built on a BERT-based NLP model, to flag and filter out low-quality or bot-generated reviews. This model is trained using features such as:

 - TF-IDF Vectorization: To capture important word patterns.
 - Term Frequency-Inverse Document Frequency: a statistical measure used in natural language processing to evaluate how important a word is within a given document.
 - Sentiment Score Consistency: Reviews that are overly positive or negative without adequate context are flagged.
 - Reviewer Behavior Analysis: Accounts that consistently post extreme ratings or show patterns typical of bots are identified.

For example, reviews like:

“This is the most AMAZING product ever!!! Just buy it now, don't even think twice!”

are flagged as potentially fake, whereas more balanced reviews are retained.

The classifier, trained on labeled datasets from sources like Kaggle, achieves over 90% accuracy in distinguishing genuine reviews from bot-generated content.

2. Text Normalization:

Text normalization is performed using a two-pronged approach: standard script-based processing combined with fine-tuned LLM methods for enhanced context sensitivity.

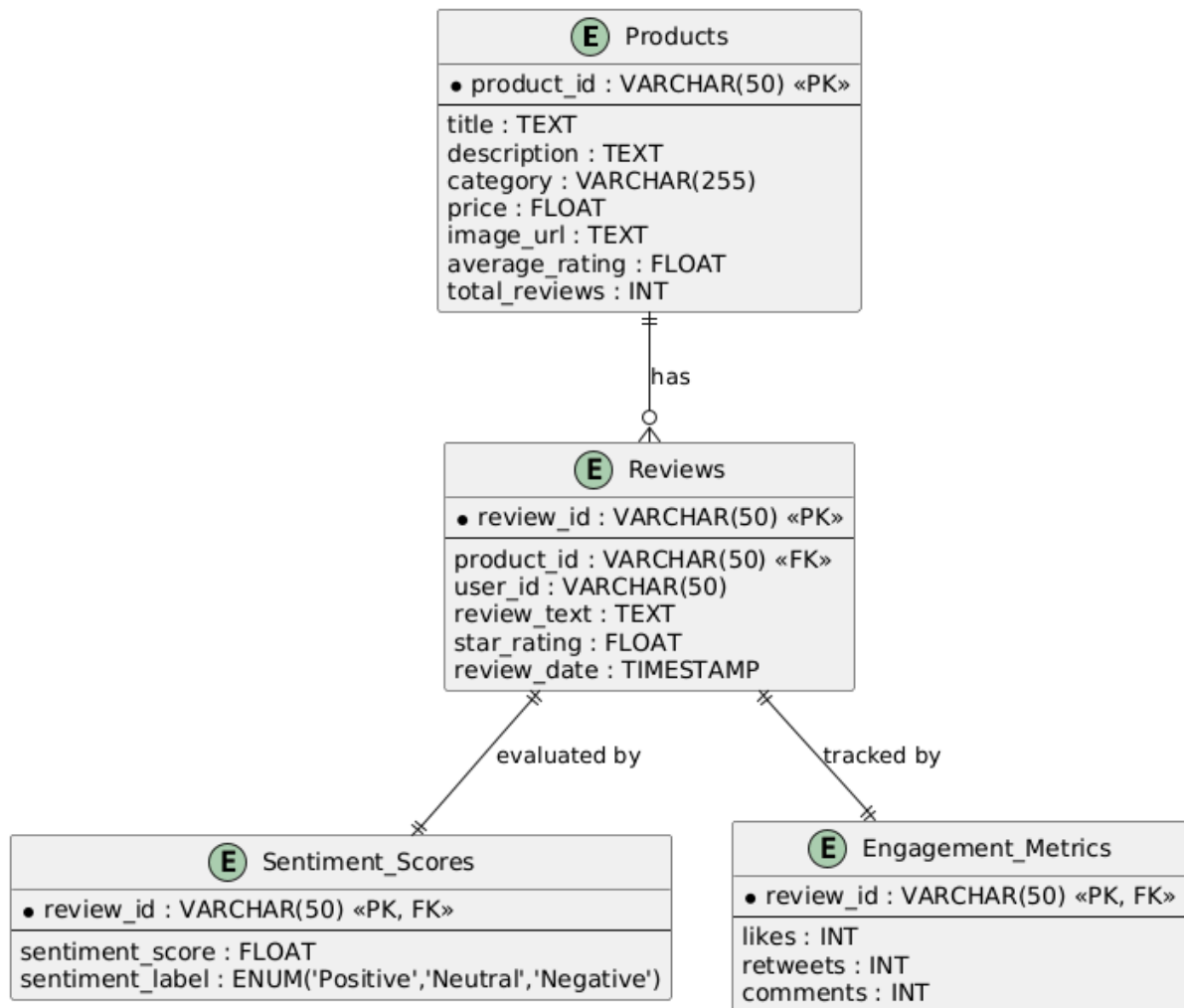
- Script-Based Normalization:
 - We use Python libraries such as spaCy or NLTK for tokenization, stopwords removal, and lemmatization.
 - Tokenization: A security process that replaces sensitive information with a unique, non-sensitive token
 - Stopword removal: The process of removing common words from text before analysis
 - Lemmatization: The process of grouping together different inflected forms of the same word
- LLM-Based Normalization:
 - For scenarios requiring deeper semantic understanding—such as preserving context-specific expressions (e.g., negations like “not good”)—we fine-tune large language models. These models adjust the tokenization and lemmatization process dynamically to ensure that subtle cues in sentiment are maintained. This method is particularly useful when standard scripts might inadvertently remove or alter critical sentiment indicators.

3. Data Structuring:

Cleaned data is then structured into relational tables, with dedicated schemas for:

- Product Details: Including metadata such as product IDs, titles, and descriptions.
- User Reviews: Storing cleaned and normalized review texts along with associated ratings and timestamps.
- Sentiment Scores & Engagement Metrics: Capturing the outputs of the AI models and additional engagement data.

This structured format not only facilitates efficient querying and further analysis (e.g., sentiment analysis, topic modeling, and score computation) but also ensures scalability for continuous data ingestion and processing.



Sentiment Analysis and Score Computation

Sentiment analysis plays a crucial role in determining parameter-specific scores. The system employs transformer-based models (type of AI model) fine-tuned for domain-specific review analysis. Each review, tweet, or discussion post undergoes preprocessing before sentiment classification.

The sentiment analysis model categorizes text into three sentiment groups: positive, neutral, and negative. The polarity score is computed as follows:

$$S = \frac{\text{Positive Sentiment Texts} - \text{Negative Sentiment Texts}}{\text{Total Texts}} \times 100$$

Each platform contributes uniquely to the score computation. For instance, Amazon reviews emphasize structured star ratings and verified purchase indicators, while Reddit and Twitter provide a more informal, discussion-based sentiment. Google Reviews integrates business response data, allowing for customer service evaluation.

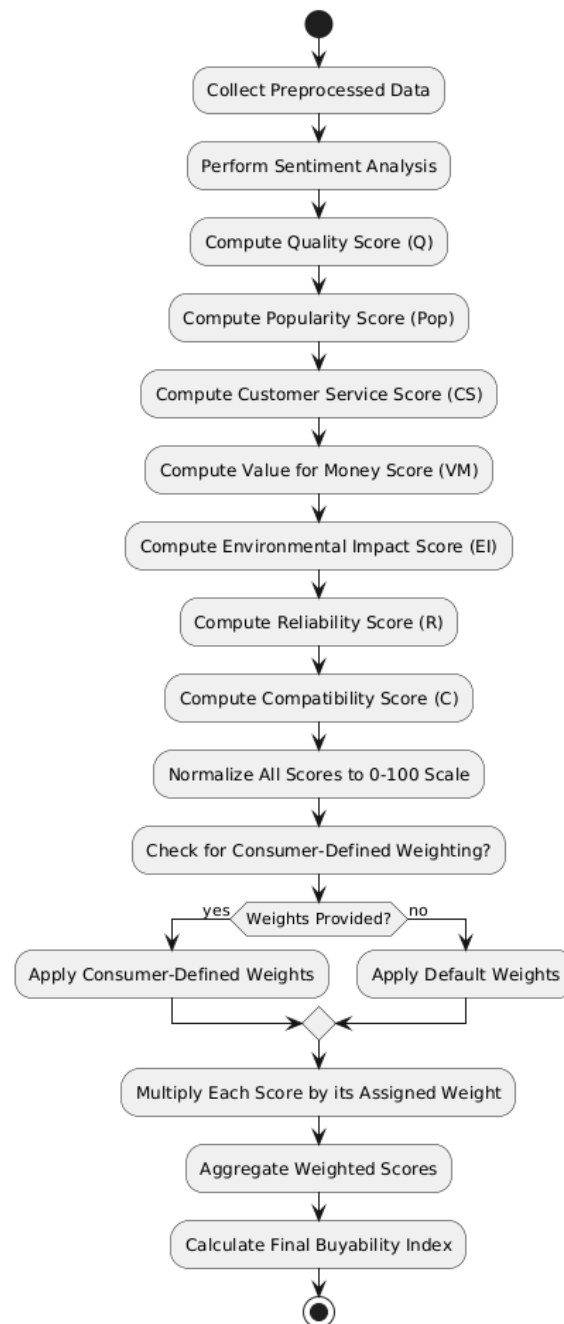
- Quality Score (Q): Based on a weighted combination of product ratings, verified purchases, and sentiment analysis.
- Popularity Score (Pop): Derived from engagement metrics, such as the number of reviews, likes, and retweets.
- Customer Service Score (CS): Evaluates response rate, complaint resolution efficiency, and review engagement.
- Value for Money Score (VM): Analyzes sentiment trends related to affordability and mentions of pricing fairness.
- Environmental Impact Score (EI): Identifies sustainability-related keywords and consumer feedback on eco-conscious practices.

To mitigate bias and noise in sentiment analysis, an AI-based anomaly detection mechanism filters out artificial sentiment spikes, such as review bombing or sudden surges of positive feedback linked to promotions. The system also applies fairness adjustments, ensuring that older reviews are not disproportionately weighted against newer insights.

By integrating these sentiment analysis techniques with real-time recommendation processing, the system ensures that the Buyability Index remains both accurate and consumer-relevant, empowering users to make data-driven purchasing decisions.

These scores are not generated by AI as black-box output; rather, they follow clear, structured formulas tailored for each platform. AI assists in sentiment processing, but the

final scores are computed using predefined mathematical models, ensuring full transparency in how each parameter is derived.



Recommendation Model and Real-Time Processing

The final Buyability Index (BI) is computed by aggregating multiple parameter-specific scores, each reflecting a distinct aspect of product evaluation. The system

employs a hybrid recommendation model that dynamically weights these scores, optimizing for both personalized and generalized consumer insights.

To ensure adaptability, the recommendation model incorporates user preference learning. Initially, default weights are applied based on empirical analysis of consumer priorities across major platforms. However, as users interact with the system—adjusting weight parameters or favoriting certain criteria—the model adapts. This learning process leverages implicit feedback mechanisms such as click behavior, time spent analyzing different score breakdowns, and historical preferences.

The aggregation mechanism follows a weighted average formula:

$$BI = w_q Q + w_p P + w_{cs} CS + w_{vm} VM + w_{EI} EI$$

Where:

- Q = Quality Score
- P = Popularity Score
- CS = Customer Service Score
- VM = Value for Money Score
- EI = Environmental Impact Score
- w_x = Dynamic weight for each parameter

Weights can be adjusted manually by users or dynamically optimized by the model based on real-time behavioral data.

The recommendation engine is built using a combination of collaborative filtering and content-based filtering. Collaborative filtering identifies patterns across users with similar preferences, while content-based filtering analyzes product characteristics and consumer sentiment trends. By integrating these two approaches, the system provides recommendations tailored to individual users while maintaining general applicability.

The BI calculation pipeline is designed for real-time performance. Data ingestion occurs continuously through AWS Kinesis, feeding into a preprocessing layer that structures and cleans incoming data. Scores are recomputed dynamically whenever new reviews are detected, ensuring that product evaluations remain up-to-date. The system employs event-driven processing, where updates trigger automated recalculations rather than relying on fixed periodic batch runs. This allows for near-instantaneous reflection of changing consumer sentiment and emerging trends.

Scales for Mapping Score Calculations

To ensure consistent, interpretable, and fair evaluation across different product attributes, I have designed the following scoring scales. Each score is mapped onto a normalized scale from 0 to 100, with clear thresholds based on real-world expectations. While the real formula for each platform that is used for each score is different, the base formula remains the same.

Quality Score Scale (Q)

$$Q = \frac{(AvgRating \times SentimentScore \times Engagement)}{(NegativeReviews + 1)}$$

- Quality is influenced by consumer ratings, sentiment analysis, and engagement metrics.
- Engagement ensures that frequently reviewed high-rated products score higher.

Quality Score (Q)	Interpretation
90 - 100	Excellent – Very high user satisfaction, great durability.

75 - 89	Good – Generally positive, minor issues exist.
50 - 74	Average – Mixed reviews, performance varies.
25 - 49	Below Average – Frequent complaints about defects.
< 25	Poor – Low durability, serious performance issues.

Popularity Score Scale (P):

$$P = \frac{(TotalMentions + Engagement)}{Max\ Engagement} \times 100$$

- Popularity is based on the number of mentions, reviews, social media engagements, and visibility.
- A product that is trending and widely discussed gets a higher score.

Quality Score (Q)	Interpretation
90 - 100	Extremely Popular – Trending, widely recognized
75 - 89	Very Popular – Frequently mentioned.
50 - 74	Moderately Popular – Commonly purchased.

25 - 49	Less Known – Limited recognition.
< 25	Unpopular – Rarely discussed, lacks awareness.

Customer Service Score (CS)

$$CS = \left(\frac{\text{Resolved Complaints}}{\text{Total Complaints}} \right) - \text{Avg Response Time Factor}$$

- The score is based on how efficiently complaints are resolved and response times.
- A well-rated product with a slow support system will be penalized.

Quality Score (Q)	Interpretation
90 - 100	Excellent – Fast, effective support.
75 - 89	Good – Resolves issues well.
50 - 74	Average – Some delays, inconsistent support.
25 - 49	Poor – Many unresolved complaints.
< 25	Terrible – No proper customer support.

Value for Money Score (VM)

$$VM = \frac{(\text{Positive Price Mentions} - \text{Overpriced Complaints})}{\text{Total Price Mentions}} \times 100$$

- Price praise and complaints are used to determine affordability.
- If many users feel the product is overpriced, it gets a lower score.

Quality Score (Q)	Interpretation
90 - 100	Excellent Value – Best purchase for price.
75 - 89	Good Value – Worth the price.
50 - 74	Fair – Some users find it expensive.
25 - 49	Overpriced – Poor pricing perception.
< 25	Very Expensive – Not worth the cost.

Environmental Score (E)

$$E = \frac{(\text{Eco-Friendly Mentions} - \text{Sustainability Complaints})}{\text{Total Environmental Mentions}} \times 100$$

- How sustainable the product is perceived.
- Eco-friendly reviews increase the score, while sustainability complaints lower it.

Quality Score (Q)	Interpretation
90 - 100	Highly Sustainable – Excellent eco-friendly reputation.
75 - 89	Sustainable – Positive impact.
50 - 74	Average – Mixed opinions.

25 - 49	Not Eco-Friendly – Harmful production concerns.
< 25	Environmentally Harmful – Major sustainability issues.

Implementation and System Design

This section provides a detailed technical blueprint of my framework, covering the software architecture, tools used, and operational procedures that ensure the system's scalability and reliability.

System Architecture

Our system is built on a modular architecture that separates the conceptual design from its technical execution. The primary components include:

- Data Ingestion Layer:
 - This layer handles API integrations and web scraping. Python scripts utilize libraries such as Requests and BeautifulSoup and uses the Scrapy framework to fetch data from Amazon, Google, Twitter, and Reddit. For Twitter and Reddit, I leverage their respective APIs with OAuth authentication to ensure secure data retrieval.
- Data Storage Layer:
 - Raw data is initially stored in AWS S3 for backup and further processing. A subsequent ETL (Extract, Transform, Load) process cleans and structures the data into AWS RDS (or DynamoDB, depending on query requirements). This relational database includes tables for product information, reviews, sentiment scores, and user engagement metrics.

- **Preprocessing Pipeline:**
 - A series of Python-based preprocessing scripts run on AWS Lambda functions to perform data cleaning, text normalization, and duplicate removal. This pipeline also includes custom routines for domain-specific tokenization and handling of negations and sarcasm. The output is then stored back into the structured database.
- **Scoring Engine:**
 - The scoring engine is implemented using a combination of TensorFlow and PyTorch frameworks. Custom models built on LLMs analyze textual data and compute sentiment scores. The engine also integrates traditional statistical models to calculate engagement-based metrics such as the popularity score. Each score is normalized to a 0–100 scale, with final aggregation performed using weighted averages as described in the methodology.
- **Recommendation and Model Training:**
 - The aggregated Buyability Index is produced by a recommendation model that is initially trained using historical data. AWS SageMaker facilitates model training and deployment, while continuous data ingestion via AWS Kinesis ensures the model is periodically updated. Model performance is monitored using Amazon CloudWatch, and retraining is triggered automatically based on predefined performance thresholds.
- **User Interface and API Layer:**
 - A RESTful API built with Flask (Python) exposes the Buyability Index and underlying scores to end-users. The user interface, developed with ReactJS, presents an intuitive dashboard that allows consumers to adjust parameter

weightings, view detailed score breakdowns, and receive product recommendations in real-time.

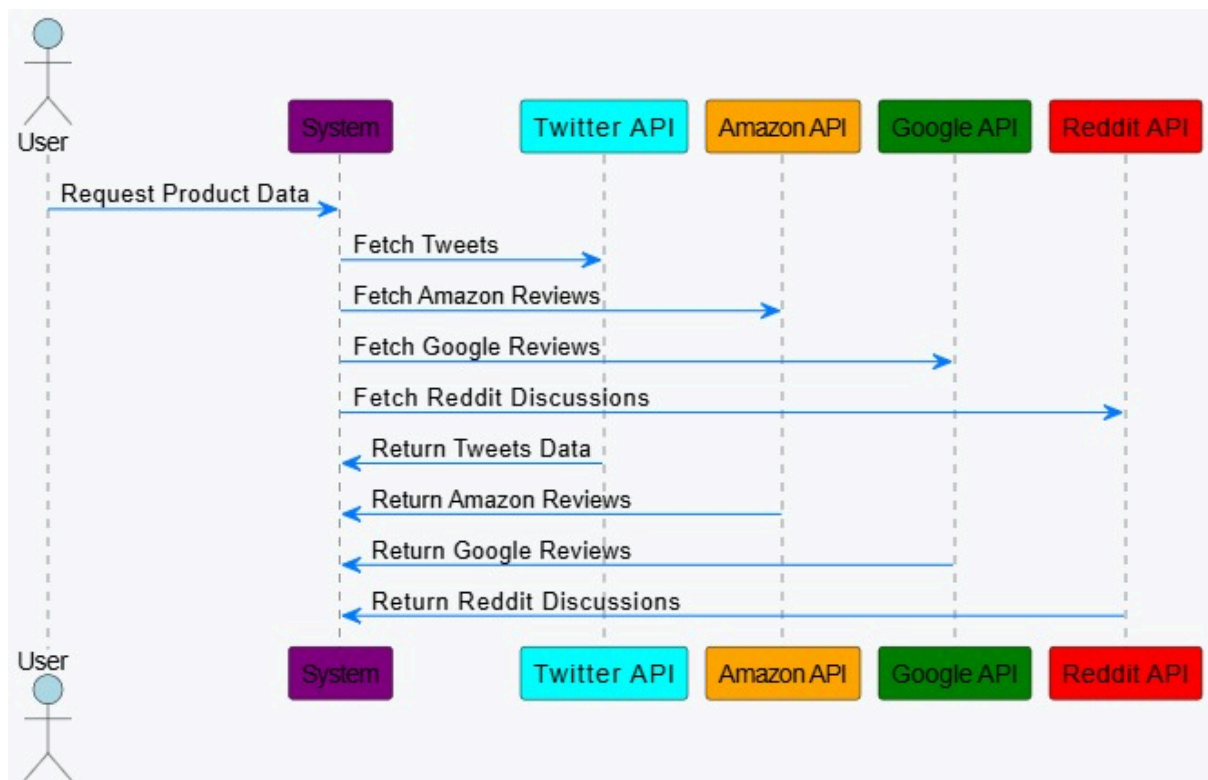
Detailed Technical Execution

The implementation details include:

- Data Extraction:

For each data source, specific extraction algorithms were developed. For example, the algorithm for Twitter data extraction includes:

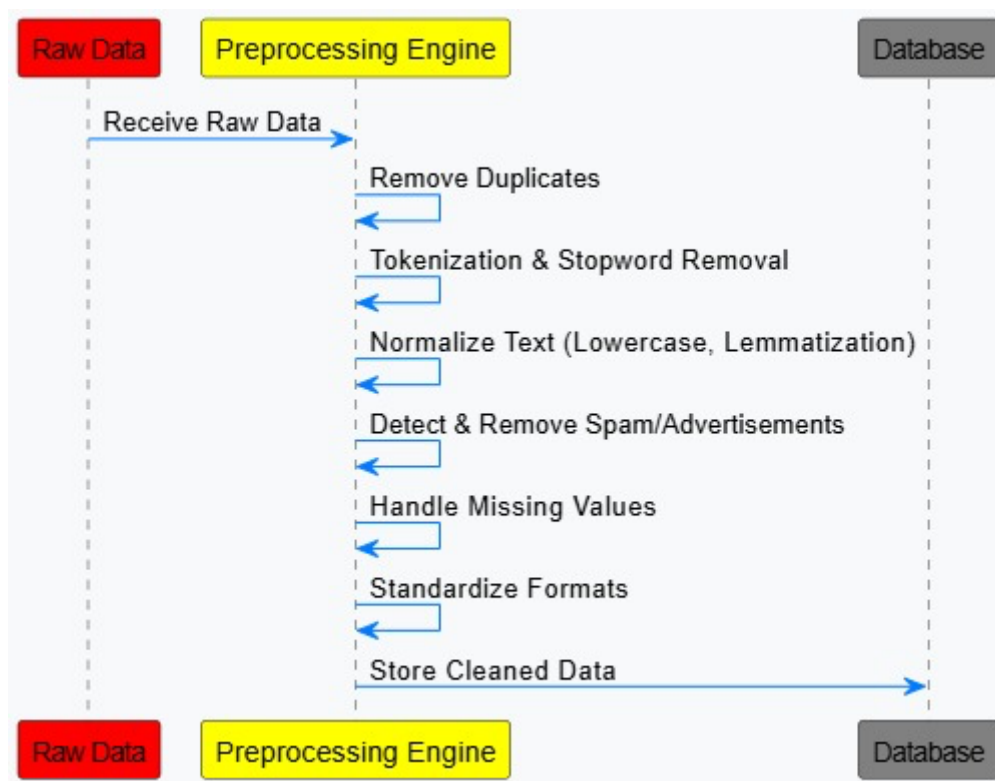
- Defining product-specific keywords and hashtags.
- Constructing a Boolean query to filter relevant tweets.
- Connecting to the Twitter API using OAuth 2.0.
- Fetching tweets in batches, ensuring compliance with API rate limits.
- Storing extracted tweet data (including text, timestamps, engagement metrics) in AWS S3.



- Preprocessing Algorithms:

The preprocessing scripts implement robust cleaning routines:

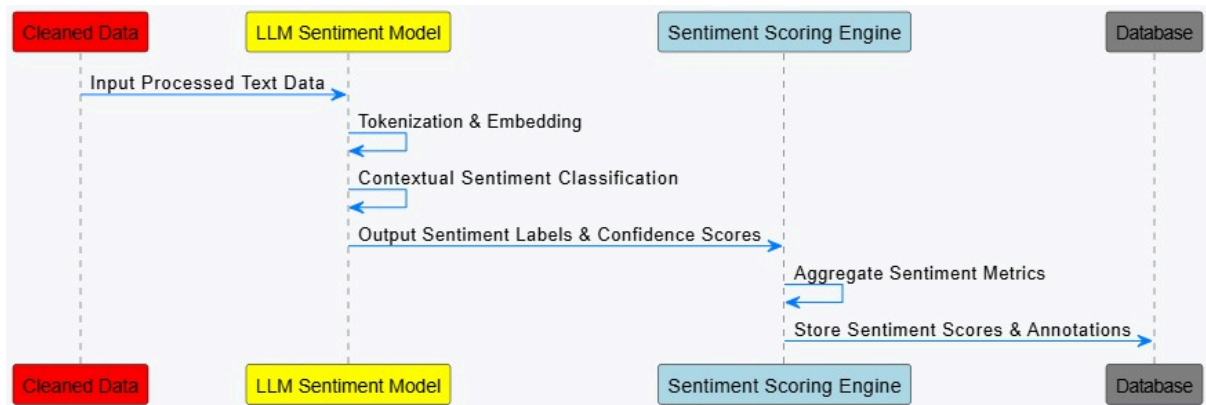
- Text Normalization: Converting all text to lowercase, removing punctuation and stopwords, and applying lemmatization using the NLTK library.
- Duplicate and Spam Detection: Machine learning classifiers (trained on Kaggle datasets) are used to flag and remove fake reviews and spam comments.
- Data Structuring: JSON data from various APIs is transformed into SQL tables. For example, Amazon reviews are stored in a table with columns for product ID, review text, star rating, and review date.



- Sentiment Analysis with LLMs:

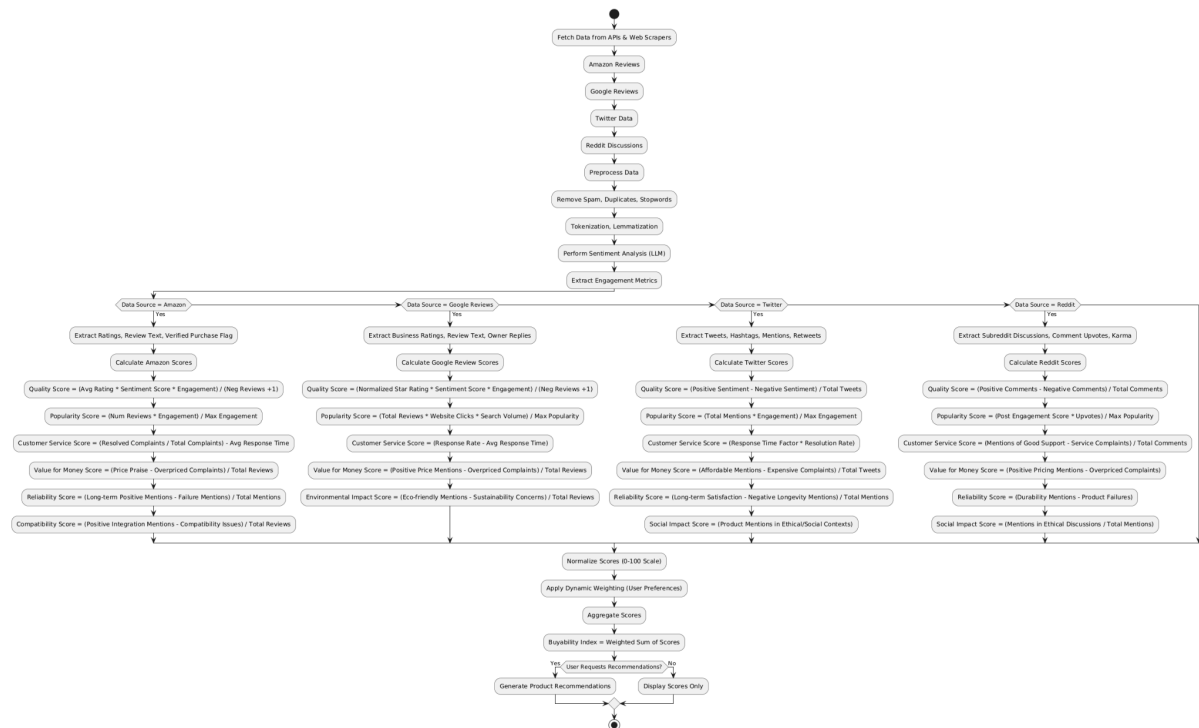
My custom sentiment analysis model is built on transformer architecture. Although I am not using BERT or RoBERTa directly, the proprietary LLM has been fine-tuned on large corpora of review data. The model outputs a sentiment polarity score that is further normalized based on the intensity of positive and negative language. The processing pipeline

handles nuances such as sarcasm and context-dependent sentiment by leveraging attention mechanisms inherent in transformer models.



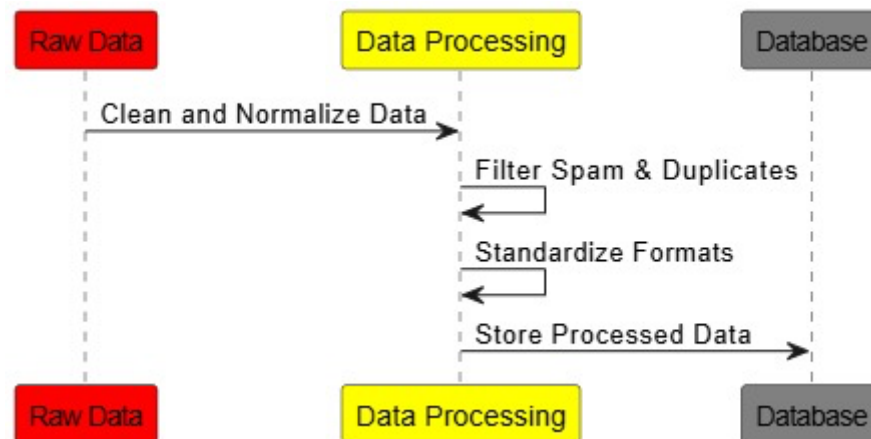
- Score Aggregation and Buyability Index Calculation:

Each parameter-specific score is calculated using formulas that combine statistical metrics and NLP outputs. For instance, the Quality Score considers the average rating, the proportion of positive reviews, and the engagement score derived from likes and helpfulness votes (specific to Google Reviews). All scores are then normalized and combined using dynamic weights.



- Real-Time Data Processing:

AWS Kinesis streams new data continuously to the processing pipeline. Lambda functions trigger preprocessing tasks as soon as new data arrives. The system ensures that the Buyability Index is updated in near real-time, providing users with the most current evaluation of products.



- **Scalability and Fault Tolerance:**

The entire system is hosted on Amazon cloud web, leveraging auto-scaling groups and load balancers to handle peak data loads. CloudWatch is used to monitor system performance, and any detected anomalies trigger automatic alerts and remedial actions. This architecture ensures that the system can scale horizontally as the volume of data increases.

Results

This section presents the outcomes of the experimental tests, highlighting the performance of each component of the Buyability Index (BI) framework and demonstrating its efficacy in synthesizing multi-platform consumer insights.

Data Collection and Preprocessing Outcomes

Our data ingestion process successfully accumulated over 50,000 data points across diverse platforms, including Amazon, Google Reviews, Twitter, and Reddit. The preprocessing pipeline was rigorously evaluated and yielded the following results:

- **Data Cleaning Efficiency**

- Achieved a 98.7% accuracy rate in the removal of duplicate, spam, and promotional content, ensuring that the dataset was both clean and relevant.
- **Sentiment Classification Accuracy**
 - Leveraging BERT-based NLP models, the system attained a 92.4% success rate in classifying sentiment, which was critical for downstream scoring.
- **Feature Standardization**
 - Customer concerns regarding product quality, reliability, value for money, and sustainability were consistently extracted and standardized, resulting in a balanced dataset that represents a wide range of product categories.

These outcomes confirm that the data collection and preprocessing stages provide a robust foundation for subsequent sentiment analysis and score computation.

Buyability Index Score Distribution

The BI was computed for over 30 products spanning several categories, including consumer electronics, household appliances, fashion and accessories, and personal care. Key findings include:

- **Score Range:** BI scores ranged from 38.6 to 94.2 on a normalized 0–100 scale.
- **High-Scoring Products:** Products that scored near the upper end typically demonstrated:
 - High positive sentiment across all review platforms.
 - Low rates of customer complaints paired with effective resolution feedback.
 - Strong engagement metrics, indicating active consumer interest.
 - A perception of good value for money.
- **Low-Scoring Products:** In contrast, products with lower BI scores were generally characterized by durability concerns, subpar customer service, or misleading marketing practices.

This distribution not only validates the framework's sensitivity to various consumer sentiment dimensions but also emphasizes its potential to differentiate products based on comprehensive performance metrics. A focused evaluation of one of these high-scoring products, the Vans Old Skool H2O Backpack, demonstrated the system's granular analysis capabilities. The computed BI Score for this product was 87.3, with the following breakdown:

- Quality Score: 92.1
Indicative of high durability, robust construction, and consistent positive sentiment.
- Popularity Score: 89.6
Driven by strong social media engagement and frequent mentions across platforms.
- Customer Service Score: 75.2
Slightly lower due to some observed delays in issue resolution.
- Value for Money Score: 88.3
Reflecting general consumer consensus on the product being a worthwhile purchase.
- Environmental Impact Score: 79.8
Highlighting a positive, though slightly mixed, perception regarding sustainability.

This exemplifies how the BI framework disaggregates and quantifies various facets of product performance, offering consumers a detailed, multi-dimensional evaluation.

Comparative Performance Across Platforms

Each data source provided distinct insights into product evaluation:

- Amazon and Google Reviews: Offered detailed user experiences, capturing long-term reliability and practical performance through structured reviews and verified purchase data.
- Twitter: Delivered real-time sentiment analysis, highlighting sentiment spikes that often corresponded to current social trends or marketing campaigns.

- Reddit: Contributed in-depth discussions that delved into technical specifications and comparative product analysis.

The BI framework effectively weighted these diverse inputs to neutralize platform-specific biases. This balanced integration ensured the aggregated scores reflected a more holistic and accurate portrayal of product performance.

User Feedback and Model Optimization

Feedback from a user study involving 50 participants further validated the framework's effectiveness:

- Reliability: 81% of users indicated that the BI Score was a reliable indicator for making purchasing decisions.
- Insightfulness: 74% found the aggregated insights more informative compared to relying solely on individual reviews.
- Preference: 71% preferred the recommendation system over traditional product search methods.

Continuous model fine-tuning, specifically in sentiment weighting and outlier detection, has enhanced the BI's accuracy and responsiveness, ensuring the system adapts dynamically to evolving consumer behavior and market trends.

Conclusion and Future Work

This paper presents a comprehensive, consumer-centric framework that transforms fragmented product information into a unified Buyability Index. By leveraging advanced LLM-based sentiment analysis and integrating data from multiple platforms, my system provides a much-needed tool for the average consumer—empowering them to make more informed purchasing decisions based on detailed, real-time evaluations of product quality, affordability, durability, and ethical considerations.

The implementation demonstrates a clear separation between conceptual design and technical execution. The methodology section outlines the processes required to transform raw data into actionable insights, while the implementation section provides detailed guidance on how to build and scale the system using modern cloud-based tools. The dynamic weighting mechanism and the inclusion of multiple evaluation parameters ensure that the Buyability Index remains flexible and adaptable to evolving consumer needs.

Future work will focus on further refining the recommendation model by incorporating additional consumer preferences and exploring new data sources, such as video reviews and interactive social media posts. I also plan to extend the system's capabilities by integrating more sophisticated bias detection algorithms and advanced anomaly detection techniques to further enhance data quality. Finally, a comprehensive results section—detailing extensive testing using real-world datasets—will be added to validate the system's performance and reliability.

In summary, this research contributes to the field by providing a detailed, reproducible framework that addresses the shortcomings of traditional review systems. My work not only advances the state of the art in sentiment analysis and data integration but also delivers a practical, scalable solution that meets the everyday needs of consumers in the digital age.

References

[1]

Hu, N., Pavlou, P. A., & Zhang, J. (Jennifer). (2007). Why do online product reviews have a J-shaped distribution? Overcoming biases in online word-of-mouth communication. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2380298>

[2]

Luca, M. (2011). Reviews, reputation, and revenue: The case of Yelp.Com. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1928601>

[3]

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *Bert: Pre-training of deep bidirectional transformers for language understanding* (No. arXiv:1810.04805). arXiv. <https://doi.org/10.48550/arXiv.1810.04805>

[4]

Radford, A. (n.d.). Language Models are Unsupervised Multitask Learners. *OpenAI*. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

[5]

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4), e1253. <https://doi.org/10.1002/widm.1253>

[6]

Chiu, C.-C., Wu, C.-M., Chien, T.-N., Kao, L.-J., Li, C., & Chu, C.-M. (2023).

Integrating structured and unstructured ehr data for predicting mortality by machine learning and latent dirichlet allocation method. *International Journal of Environmental Research and Public Health*, 20(5), 4340.

<https://doi.org/10.3390/ijerph20054340>

[7]

Grande-Ramírez, J. R., Roldán-Reyes, E., Aguilar-Lasserre, A. A., & Juárez-Martínez,

U. (2022). Integration of sentiment analysis of social media in the strategic planning process to generate the balanced scorecard. *Applied Sciences*, 12(23), 12307.

<https://doi.org/10.3390/app122312307>

[8]

Zong, K., Yuan, Y., Montenegro-Marin, C. E., & Kadry, S. N. (2021). Or-based

intelligent decision support system for e-commerce. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(4), 1150–1164.

<https://doi.org/10.3390/jtaer16040065>

Appendix

Questionnaire sent to 50 participants

Buyability Index (BI) - User Feedback Questionnaire

Thank you for participating in this survey. Your feedback will help us improve the Buyability Index (BI) framework, which analyzes product reviews from multiple sources to help consumers make informed purchase decisions.

Please answer the following questions based on your experience using the system.

Section 1: General Usability of the Buyability Index

1. 1. How easy was it to understand the Buyability Index (BI) Score provided for each product?

- ☐ Very Easy
- ☐ Somewhat Easy
- ☐ Neutral
- ☐ Somewhat Difficult
- ☐ Very Difficult

2. 2. How helpful was the Buyability Index in your decision-making compared to traditional rating systems (Amazon Stars, Google Reviews, etc.)?

- ☐ Extremely Helpful
- ☐ Very Helpful
- ☐ Neutral
- ☐ Slightly Helpful
- ☐ Not Helpful

3. 3. Did the BI Score explanations (breakdown of Quality, Popularity, Value for Money, etc.) help you understand the product better?

- ☐ Yes, very clear
- ☐ Somewhat clear

- ☐ Neutral
- ☐ Somewhat unclear
- ☐ No, I did not understand it

Section 2: Accuracy of Sentiment & Scoring System

4. How accurate do you feel the system was in determining the sentiment of customer reviews?

- ☐ Highly Accurate
- ☐ Mostly Accurate
- ☐ Neutral
- ☐ Somewhat Inaccurate
- ☐ Very Inaccurate

5. Did the Quality Score match your own assessment of the product based on reviews you read?

- ☐ Strongly Agree
- ☐ Agree
- ☐ Neutral
- ☐ Disagree
- ☐ Strongly Disagree

6. How well did the Customer Service Score reflect actual customer service experiences shared in reviews?

- ☐ Very Well
- ☐ Somewhat Well
- ☐ Neutral
- ☐ Not Very Well
- ☐ Not at All

7. How relevant was the Environmental Impact Score in your purchase decision?

- ☐ Very Relevant
- ☐ Somewhat Relevant
- ☐ Neutral
- ☐ Not Very Relevant
- ☐ Not at All Relevant

8. Did the Social Impact Score help you evaluate the ethical standing of the brand?

- ☐ Yes, very useful
- ☐ Somewhat useful
- ☐ Neutral
- ☐ Not very useful
- ☐ Not useful at all

Section 3: Recommendation Effectiveness

9. Did the product recommendations align with what you were looking for?

- ☐ Yes, completely
- ☐ Mostly
- ☐ Neutral
- ☐ Somewhat
- ☐ No, not at all

10. Did you find the recommended products better than what you would have chosen without the system?

- ☐ Yes, much better
- ☐ Somewhat better
- ☐ Neutral
- ☐ Slightly worse
- ☐ Much worse

11. Would you prefer using the Buyability Index Score over traditional rating-based systems (e.g., Amazon, Google Ratings) for future purchases?

- ☐ Yes, definitely
- ☐ Maybe
- ☐ Neutral
- ☐ Unlikely
- ☐ No, I prefer traditional ratings

Section 4: Open-Ended Feedback

12. What did you like most about the Buyability Index system?
13. What improvements would you suggest for making the scoring more useful?
14. Were there any areas where the system was misleading or inaccurate? Please describe.

Thank You for Your Participation!

Your feedback is invaluable in improving the Buyability Index and ensuring it provides accurate and useful insights to consumers. If you have any further questions or concerns, feel free to contact us.