# E-commerce Customer Review Analysis: Predicting Product Success through Sentiment Intelligence

# Final Project Report

# Dataset Description and Source:

# This project utilizes the Amazon Sales Dataset (Karkavelraja, 2023), which contains over 1,000 customer product reviews scraped from Amazon and made publicly available through Kaggle. The dataset includes 1,465 entries and 16 columns, each providing relevant attributes about the product, its pricing and discount structure, customer ratings, and review content. Key features include ‘product\_name’, ‘category’, ‘discounted\_price’, ‘actual\_price’, ‘rating’, ‘review\_title’, ‘review\_content’, and ‘rating\_count’. The dataset provides a valuable mixture of numerical, categorical, and text data, ideal for exploring customer sentiment, extracting product insights, and evaluating relationships between product attributes and customer satisfaction. It spans multiple categories, especially electronics and accessories, making it a representative sample for e-commerce applications.

**Business Problem or Question**

In the current competitive e-commerce environment, customer reviews provide valid insights. Still, most businesses fail to see the bigger picture in the written reviews and base their decisions primarily on numeric scales. This project will address whether the sentiment associated with customer reviews can provide valuable patterns that can be used to predict the success of a product and how to enhance it. We would like to learn more about customer satisfaction and how sentiment analysis may have been used as a reliable indicator by researching the relationship between sentiment and product category and ratings. This information is used to inform decisions regarding pricing, product updates, or marketing focus, instead of relying solely on the surface-level metrics.

**Methodology**

The development was done in Python under Google Colab. The preprocessing was as follows:

- Cleaning of column prices and rating by eliminating currency expressions and turning strings into floats or integers.

Median imputation of missing values in the column rating\_count.

Then, combining and cleaning text data in 'review\_title’ and ‘review\_content’ was conducted based on the Natural Language Toolkit (NLTK) and regular expressions. Preprocessing included:

- Lowercasing all characters

- Removing stopwords, special characters, and numbers

- Tokenizing and lemmatizing the words

**Two primary sentiment analysis tools were used:**

1. TextBlob, which provided a polarity score between -1 and 1 and a subjectivity score between 0 and 1.

2. VADER (Valence Aware Dictionary and sEntiment Reasoner), which provided compound sentiment scores, more suitable for short, informal reviews.

A bag-of-words approach using TF-IDF vectorization was used to identify top keywords associated with positive and negative sentiment. Word clouds were generated for visual inspection of term distributions.

Latent Dirichlet Allocation (LDA) was applied to model topics to discover latent themes in customer feedback. This helped classify reviews based on common phrases and topics, such as "charging speed", "durability", or "compatibility".

Finally, linear regression was applied to determine the predictive power of sentiment scores and discount percentage on the numeric rating.

**Results and Evaluation**

Preliminary exploratory analysis revealed that the most common product categories were USB cables, adapters, and TV accessories. The average rating across the dataset was 4.1, indicating generally positive customer perception. However, significant discrepancies existed between numeric ratings and review sentiment in about 15% of cases. For example, some reviews scored low in sentiment yet had ratings of 4 or above, suggesting sarcastic or ambiguous language.

Sentiment scores from VADER aligned with numeric ratings about 78% of the time. Positive reviews generally had compound scores above 0.5 and ratings of 4.0 or higher. Negative sentiment (compound score below -0.2) corresponded to ratings below 3.5. TextBlob polarity scores showed a slightly weaker correlation with ratings due to their lower contextual sensitivity.

TF-IDF analysis showed that high-rated reviews frequently contained terms such as “value for money”, “fast charging”, “good quality”, and “durable”. Conversely, low-rated reviews included terms like “not working”, “cheap material”, “broke after use”, and “waste of money”. These patterns suggest that review language can strongly indicate performance and product build quality.

**Topic modeling revealed recurring themes**

**Topic 1**: "Charging, Cable, Fast, Type-C" — centered on performance and compatibility

**Topic 2:** "Product, Return, Refund, Support" — focused on post-purchase issues

**Topic 3:** "Build, Material, Quality, Price" — related to product construction and cost-effectiveness

Regression analysis showed that sentiment compound scores and discount percentage together explained 61% of the variance in product ratings, suggesting that these features strongly predict customer satisfaction.

**Discussion of Challenges, Limitations, and Potential Improvements:**

One major challenge encountered was noise in the review content, such as emojis, links, and non-standard abbreviations, which affected tokenization and sentiment scoring. Also, some reviews contained sarcasm or humor, which TextBlob and VADER struggle to interpret correctly.

Another limitation was the imbalance in the data: most reviews were positive, with relatively few low-rated products. This skew made it difficult to train models distinguishing between neutral and negative feedback.

Additionally, while TextBlob and VADER are useful for baseline sentiment analysis, their accuracy is limited for domain-specific or complex expressions. To improve, the project could incorporate transformer-based models (e.g., DistilBERT or BERT) that understand context better and perform classification at a more sophisticated level.

The dataset also lacked time-related features (e.g., review dates), which would be useful for tracking changes in sentiment over time or during promotions. Including temporal data could enrich the analysis with trend and seasonality insights.

**Future improvements could include:**

- Sentiment classification using fine-tuned transformer models

- Multiclass classification of sentiment (positive, neutral, negative) instead of scalar regression

- Incorporating user-level features (e.g., verified purchase, helpful votes) to weigh review impact

Link to GitHub Repository:

[https://github.com/yourusername/amazon-sentiment-analysis](https://github.com/yourusername/amazon-sentiment-analysis)

**Reference**

Karkavelraja, J. (2023). Amazon Sales Dataset. Kaggle. <https://www.kaggle.com/datasets/karkavelrajaa/amazon-sales-dataset>