NLP Group Project

Improving Online Shopping Decisions by Ranking Product Reviews.

Prepared by:

| Masa Jalamneh | Aseel Dar Assi | Tasneem Hamed | Ahlam Abuqare |

15 June, 2025

Introduction

Our project aims to help users find the best and most helpful reviews for Amazon products

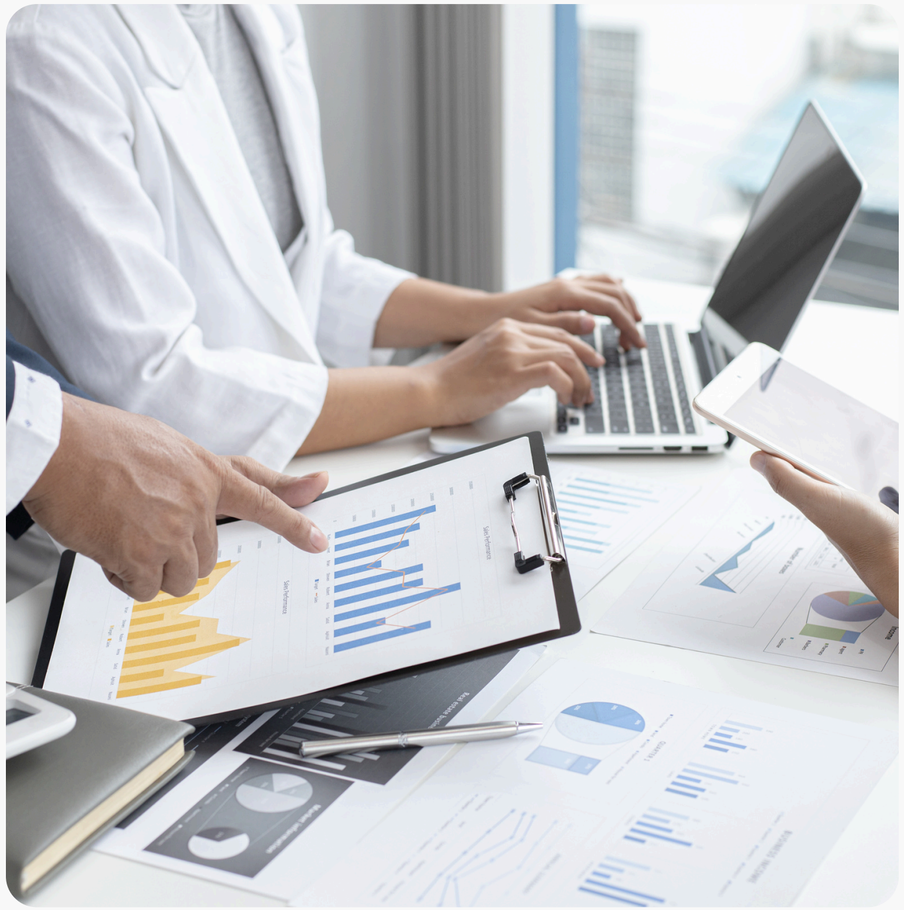
Project Idea

Our project focuses on enhancing the way we analyze and rank product reviews on e-commerce platforms like Amazon.

Normally, users are shown reviews in a random or chronological order, which doesn't always highlight the most helpful or reliable ones.

Our goal is to give customers better insights, and make the shopping experience smarter and more trustworthy.

Dataset Selection &Columns



Dataset Overview

- Dataset Source: Amazon Beauty Product ReviewsBeauty category products
- File Format: CSV
- Number of Reviews Used: 1,000
- Category: Beauty Products

Strategic Goals

25% sales increase through

Dataset Columns:

- marketplace, product_category, product_title, star_rating, review_headline, review_body, helpful_votes, verified_purchase

Why This Dataset?

- NLP-ready | Sentiment-rich | Verified trust | Accurate | Growth-driven

Work Plan

Data Collection & Cleaning:

Preprocessed Amazon review dataset for consistency and quality.

Feature Engineering:

Extracted key attributes such as rating scores and vote-based scores.

Ranking Conditions:

Applied custom rules for review ranking (rating bias, helpfulness score).

Interface Development:

Built an interactive system to visualize and explore reviews.

Testing & Evaluation:

Verified ranking logic and evaluated sample cases.

Presentation & Finalization:

Prepared results, insights, and demo for final delivery.

Review Ranking Algorithm

Features & Their Weights :

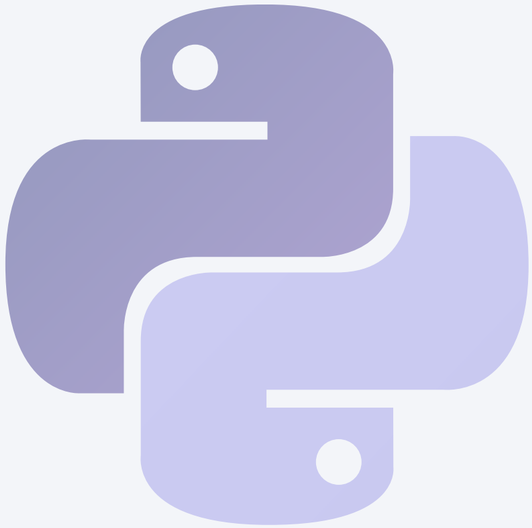
1. Sentiment Subjectivity – 30%
2. Sentiment Polarity – 15%
3. Keyword Relevance – 15%
4. User Trust & Activity Score – 10%
5. Review Length (Normalized) – 10%
6. Helpfulness Votes (Wilson Score) – 10%
7. Usage Classification (Used / Returned) – 5%
8. Star Rating (Normalized) – 5%

Final Score Formula:

Weighted sum of all the above features

Core Python Libraries Used:

- pandas – for data handling and preprocessing
- spaCy – to clean and lemmatize text
- TextBlob – for analyzing sentiment
- YAKE – to extract top keywords from reviews
- Sentence-Transformers – for semantic similarity using BERT and cosine similarity
- SciPy – to compute Wilson Score for helpfulness votes
- NumPy – for score normalization and scaling



User Profiling

Generated a profile per user based on their reviews

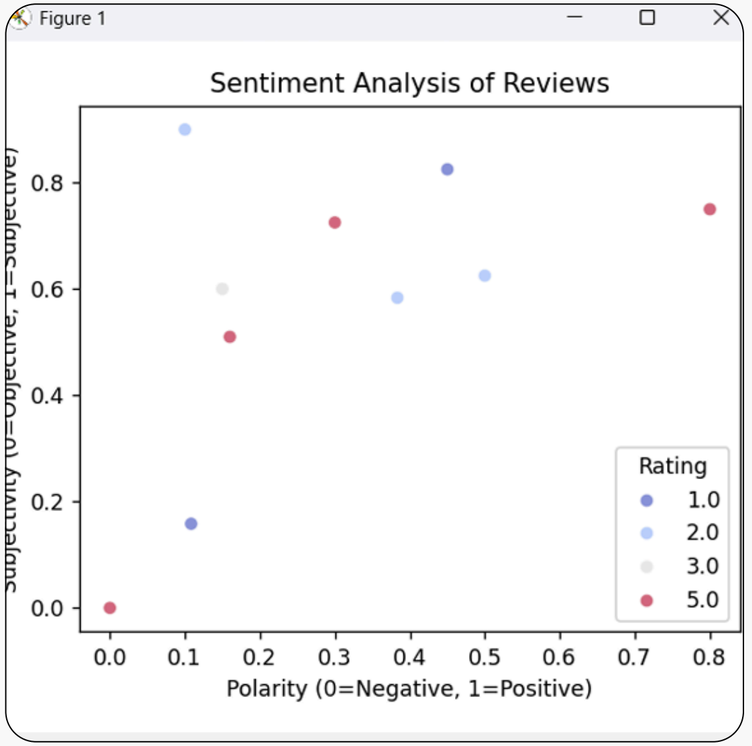
Includes:

- Average rating & sentiment
- Helpfulness & review length
- Purchase verification ratio
- Top used keywords

Helps detect behavior patterns and credibility

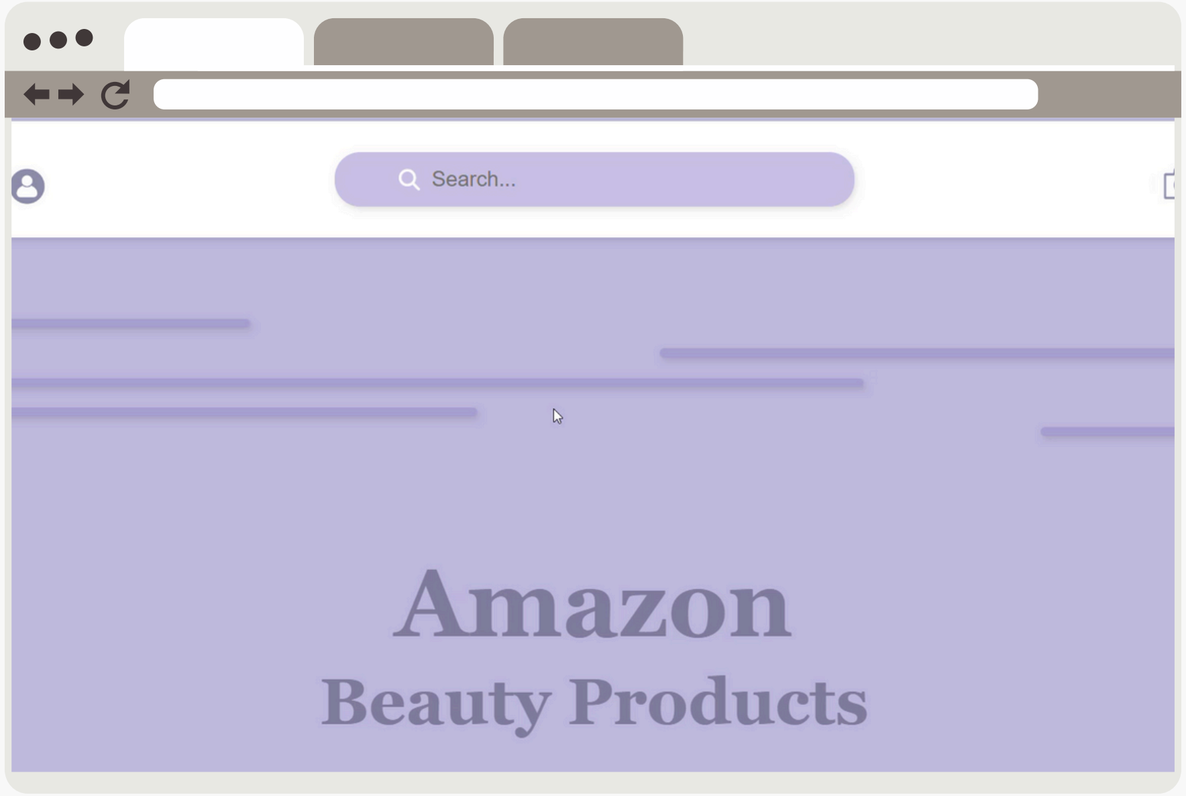
Testing & results with UI

- The system was tested using real Amazon product reviews
- Implemented smart ranking conditions
- Each review received a score reflecting its value and usefulness
- The reviews were reordered based on their computed scores



Testing & results with UI

The following interface shows the output of our ranking system



Evaluation & Conclusion

Evaluation

✓ Successfully ranked real Amazon reviews

✓ Highlighted the most relevant and helpful feedback

✓ Detected usage patterns and sentiment effectively

Conclusion

This system makes it easier for people to find helpful product reviews. It supports smarter shopping decisions by showing the most useful feedback.

Thank You
Any Questions