

Amazon Review Sentiment Analysis using NLP and Logistic Regression

Objective

To classify Amazon product reviews as **Positive**, **Neutral**, or **Negative** by analyzing review text using Natural Language Processing (NLP) and machine learning.

Dataset Overview

- **Total Reviews:** 25,000
 - **Columns:** reviews, sentiment (ratings 1–5)
 - **Label Mapping:**
 - 1–2 → Negative
 - 3 → Neutral
 - 4–5 → Positive
-

Preprocessing Steps

- Lowercased all text
 - Removed HTML tags, punctuation, numbers
 - Removed stopwords (NLTK)
 - Applied lemmatization
 - Created new cleaned column: cleaned_reviews
-

WordCloud Insights

Word clouds were used to visualize high-frequency terms per sentiment class:

- **Positive:** great, love, perfect, amazing, awesome
- **Negative:** waste, junk, disappointed, refund, poor
- **Neutral:** okay, decent, fine, overall, bit

These helped validate the emotional tone and polarity of each sentiment.

Model Used

- **TF-IDF Vectorizer** (max_features=5000)
 - **Logistic Regression** classifier (interpretable and baseline-friendly)
-

Model Performance

Metric	Value
Accuracy	69.58%
Macro F1	61%
Weighted F1	67%

Classification Report:

Sentiment Precision Recall F1-score Support

Negative	0.71	0.82	0.76	2000
Neutral	0.45	0.22	0.29	1000
Positive	0.73	0.81	0.77	2000

 **Note:** Neutral sentiment is harder to detect due to vague language and fewer examples.

Most Influential Words

Positive	Negative	Neutral
great, love, perfect, excellent, awesome	waste, refund, poor, junk, disappointed	okay, decent, fine, overall, bit

These were identified from logistic regression coefficients after TF-IDF transformation.

Business Insight

Although product-level analysis wasn't possible due to lack of product ID or name, this technique demonstrates how companies can scan thousands of reviews and extract actionable sentiment — helping in product improvement, customer satisfaction, and brand management.
