Review Analyzer

NLP Business Case - Automated Customer Reviews Analysis

Group 4
Felipe Doria & Krzysztof Giwojno

IronHack AI Engineering Bootcamp, February 2026

Dashboard

Project 3: NLP Business Case — Group 4, IronHack AI Engineering Bootcamp

TOTAL REVIEWS	PRODUCTS	CATEGORIES	AVERAGE RATING
28,332	65	6	4.51 / 5

Sentiment Distribution

● Positive 89.9%   ● Neutral 4.3%   ● Negative 5.8%

Rating Distribution

★★★★★	19,897
★★★★☆	5,648
★★★☆☆	1,206
★★☆☆☆	616
★☆☆☆☆	965

Reviews by Category

Fire Tablets	14,396
Batteries & Household	12,086
E-Readers	1,049
Smart Speakers	632
Accessories	138
Media & Home	31

Average Rating by Category

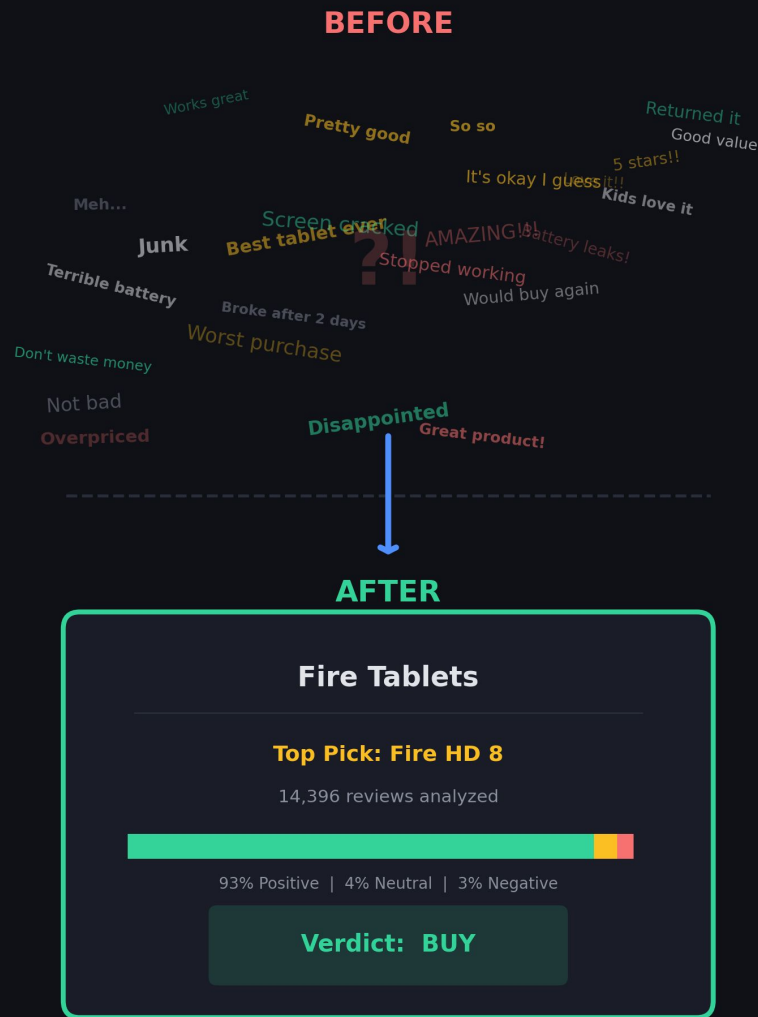|	
Fire Tablets	4.56/5
Batteries & Household	4.45/5
E-Readers	4.66/5
Smart Speakers	4.54/5
Accessories	4.31/5
Media & Home	4.39/5

We Read Them All!

28,332 reviews. 65 products. One question:
Which Amazon product should I actually buy?
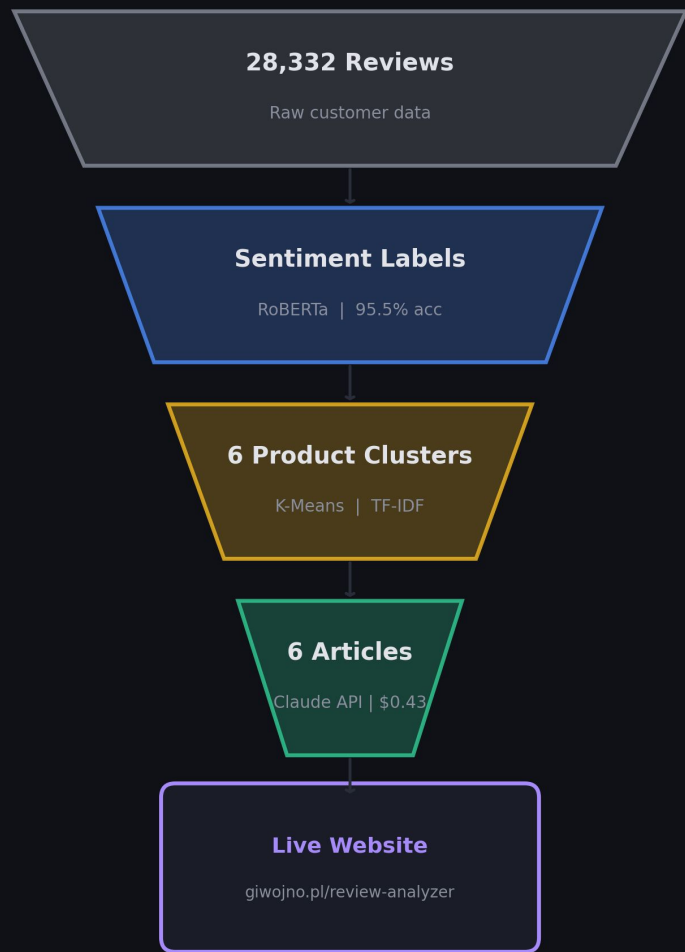
We built a system that:

- Reads every review and understands the sentiment
- Groups 65 products into 6 meaningful categories
- Writes recommendation articles - like a tech blogger, but cheaper ($0.43)

The result? A live website where you can browse categories, compare products, and read AI-generated buying guides.

No backend. No database. Just smart pre-computation.

Works great
Pretty good
So so
Returned it
Good value
5 stars!!
It's okay I guess it!!
Kids love it
Meh...
Screen cracked
Junk
Best tablet ever
AMAZING!!
Battery leaks!
Terrible battery
Stopped working
Would buy again
Broke after 2 days
Worst purchase
Don't waste money
Not bad
Disappointed
Great product!
Overpriced

AFTER

Fire Tablets

Top Pick: Fire HD 8

14,396 reviews analyzed

93% Positive | 4% Neutral | 3% Negative

Verdict: BUY

Pipeline

28,332 Reviews

Raw customer data

Sentiment Labels

RoBERTa | 95.5% acc

6 Product Clusters

K-Means | TF-IDF

6 Articles

Claude API | $0.43

Live Website

giwojno.pl/review-analyzer

Introduction

The Real Problem

Companies get thousands of reviews but lack the tools to extract actionable insights at scale.

Our Solution: 3 Models, 1 Pipeline

1. **Classification** feeds sentiment labels to...
2. **Clustering** which groups products, then...
3. **Summarization** writes articles per cluster using the classified reviews as evidence

Each model's output is the next model's input. Classification quality directly affects summarization - if we misclassify a negative review as positive, the AI writes misleading recommendations.

Methodology: Fine-tuned transformers + unsupervised clustering + LLM-powered generation, deployed as a static web app.

Methods

Dataset: Datafiniti Amazon Consumer Reviews (Kaggle)

- 28,332 reviews, 65 products, 24 columns
- Reviews from 2009–2019
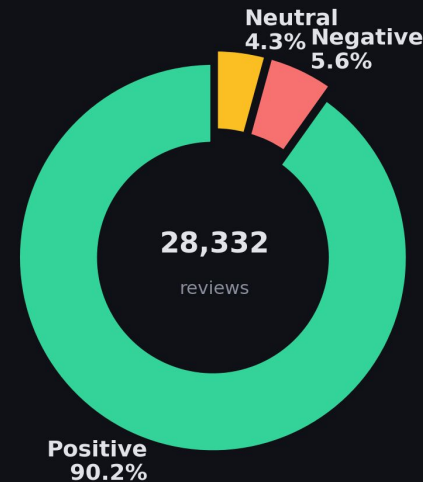- All Amazon-branded: Fire Tablets, Kindle, Batteries, Echo

Preprocessing:

- Dropped 10 columns (>99% null or metadata)
- Kept 35.87% duplicate texts (legitimate short reviews)
- Mapped ratings: 1–2★ → Negative, 3★ → Neutral, 4–5★ → Positive
- No text cleaning for transformers (they handle raw text)

The Challenge: 90% Positive, 6% Negative, 4% Neutral
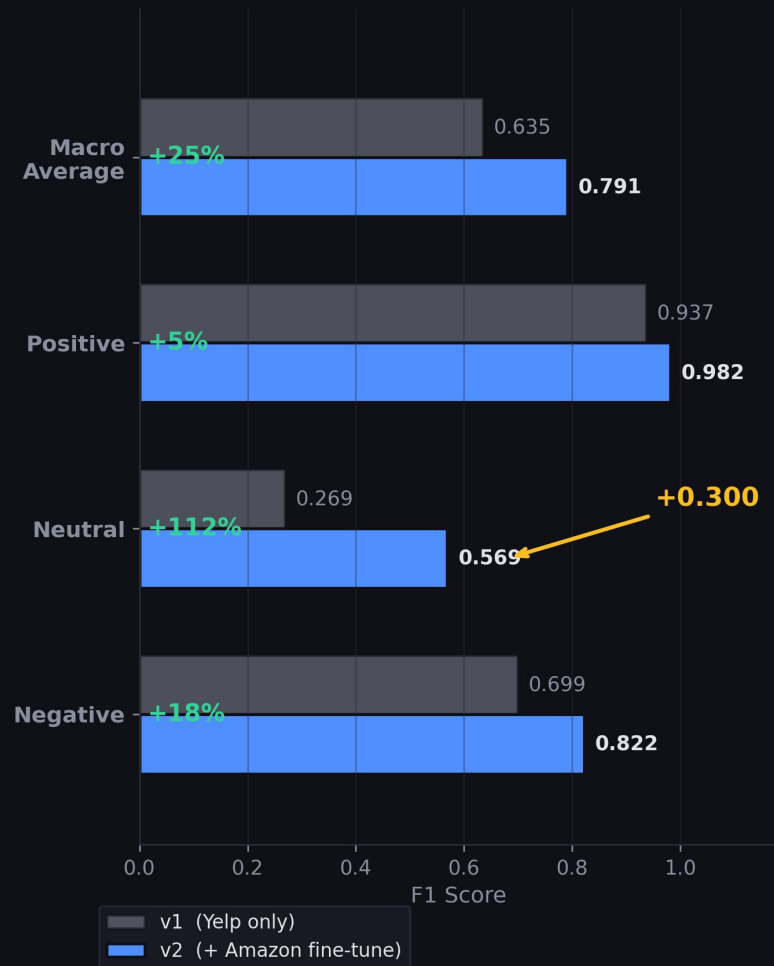Training a model on this will predict "Positive" for everything and score 90% accuracy - but that's useless.

Sentiment Distribution

Neutral 4.3%
Negative 5.6%
Positive 90.2%

28,332 reviews

Rating Distribution

Star Rating	Reviews
1 ★	908
2 ★	673
3 ★	1,206
4 ★	3,965
5 ★	21,580

Classification F1: v1 vs v2

- Macro Average: **+25%** — v1 0.635, v2 **0.791**
- Positive: **+5%** — v1 0.937, v2 **0.982**
- Neutral: **+112%** — v1 0.269, v2 **0.569** (+0.300)
- Negative: **+18%** — v1 0.699, v2 **0.822**

F1 Score

Legend:
- v1 (Yelp only)
- v2 (+ Amazon fine-tune)

Class weights doubled Neutral F1

Model 1 — Classification

Task: Classify each review as Positive, Negative, or Neutral

What we tried:

- **RoBERTa-base** (125M params) — transformer pre trained on large text corpora

Two-stage training:

1. Pre-trained on 650K Yelp reviews (3-class) → learns general sentiment
2. Fine-tuned on Amazon data with **class weights** → adapts to our domain

Key technique: Class Weights

- Neutral gets 7.8x more importance during training
- Negative gets 6.0x
- Positive gets 0.4x (downweighted)

This forces the model to pay attention to rare classes instead of just predicting "Positive" every time.

Model 1 - Evaluation

95.48% Accuracy | 0.791 Macro F1

Positive: Precision 0.985 | Recall 0.979 | F1 0.982
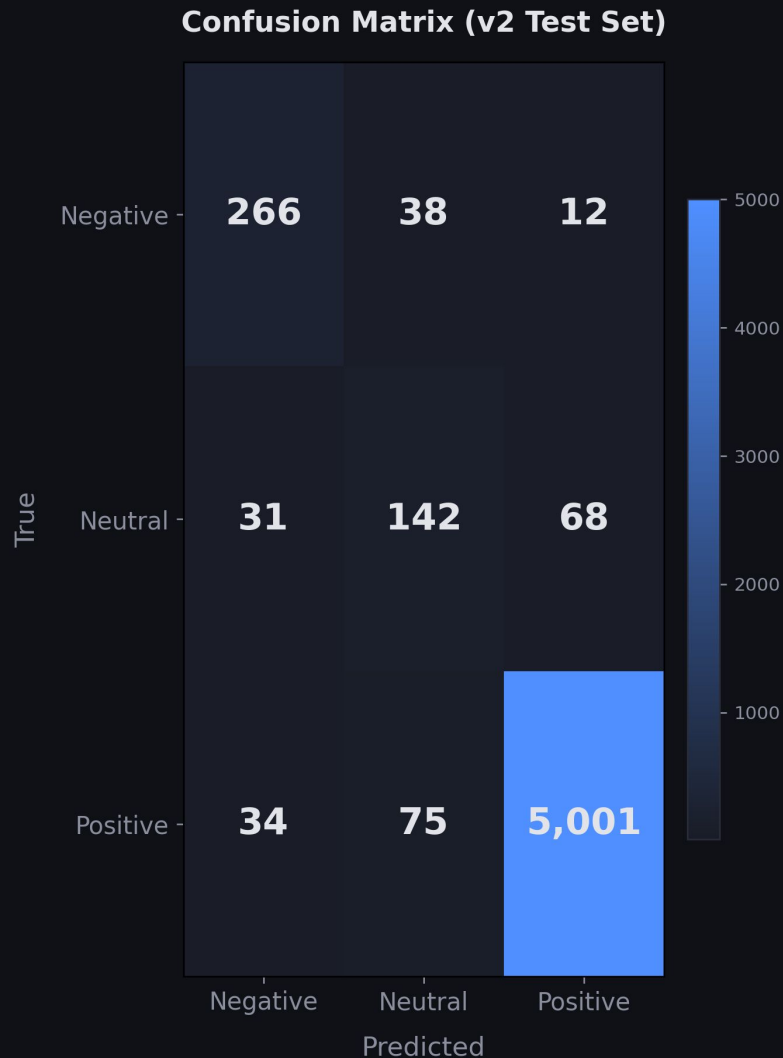**Neutral**:   Precision 0.550 | Recall 0.589 | F1 0.569
Negative: Precision 0.804 | Recall 0.842 | F1 0.822
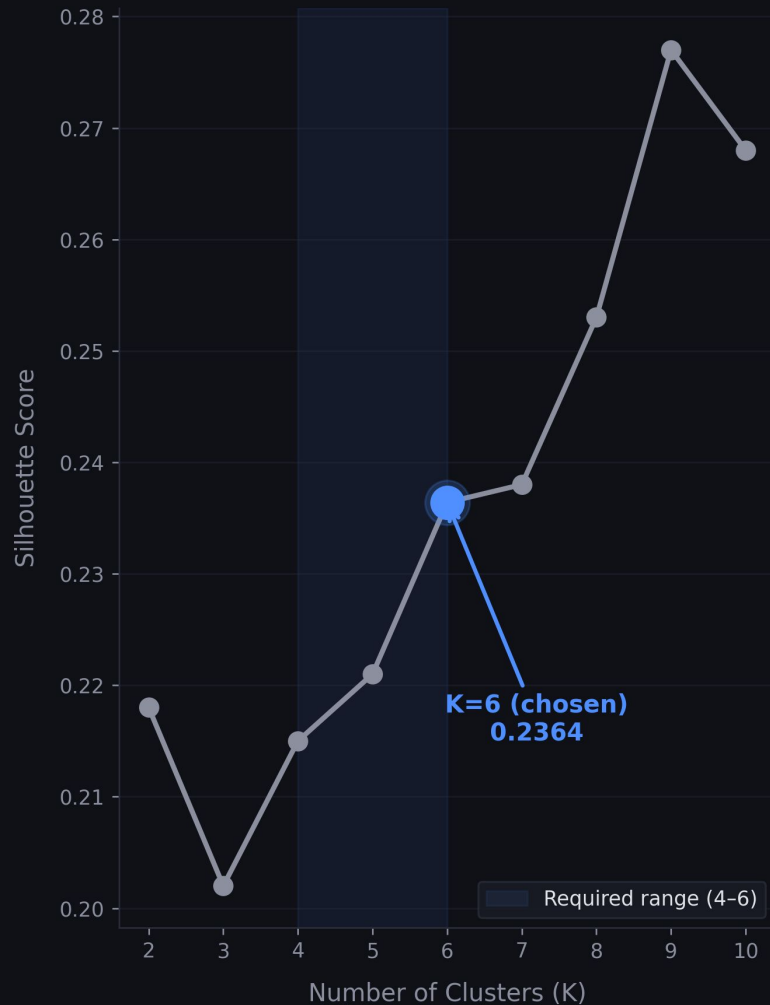
Only 256 errors out of 5,667 test reviews (4.5%)

Where it fails:

- **Positive ↔ Neutral** boundary (most errors)
- **Sarcasm and mixed-sentiment** reviews
- "It's okay I guess" — is that Neutral or Positive?

The worst error (**Negative → Positive**) happens only 12 times

Confusion Matrix (v2 Test Set)

	Negative	Neutral	Positive
Negative	266	38	12
Neutral	31	142	68
Positive	34	75	5,001

Silhouette Score by K

Model 2 — Clustering

Task: Group 65 products into 4–6 meaningful categories

Why not use existing categories?

- `primaryCategories` is useless — 92% of products are just "Electronics" or "Health & Beauty"
- We needed categories a shopper would recognize

Method:

- Combined product names + cleaned category labels
- TF-IDF vectorization (unigrams + bigrams, 200 features)
- K-Means clustering
- Tested K=2 to K=10

Chose K=6 - best silhouette within required 4-6 range (0.2364)
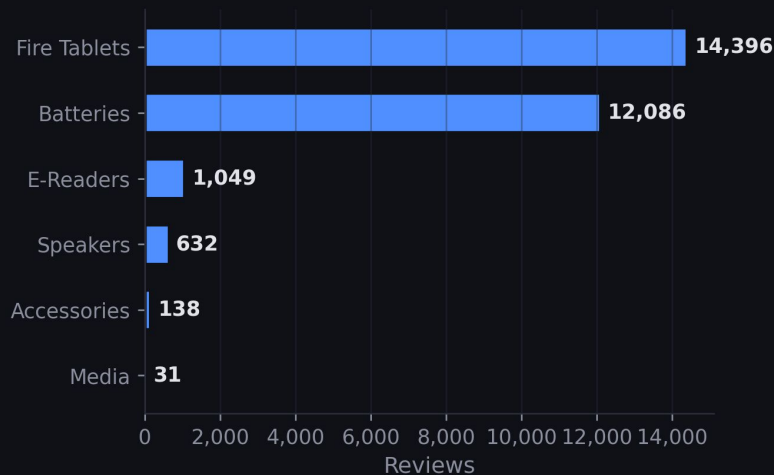
Model 2 - Evaluation

6 Clusters Found

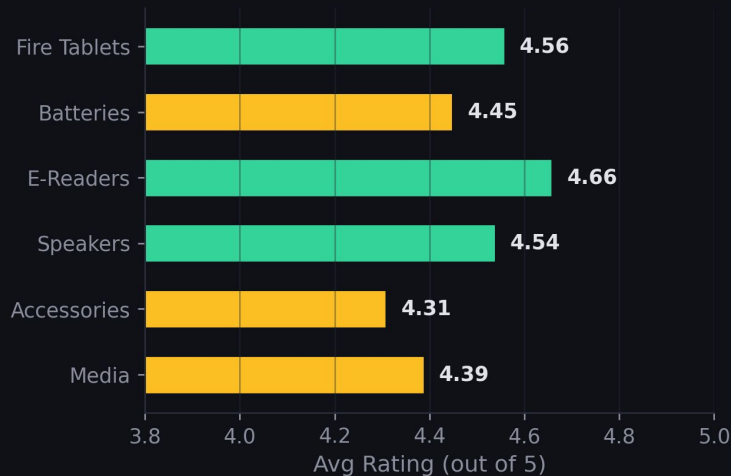|			
Fire Tablets	20 products	14,396 reviews	⭐ 4.56
Batteries	7 products	12,086 reviews	⭐ 4.45
E-Readers	10 products	1,049 reviews	⭐ 4.66
Smart Speakers	9 products	632 reviews	⭐ 4.54
Accessories	11 products	138 reviews	⭐ 4.31
Media & Home	8 products	31 reviews	⭐ 4.39

Do they make sense? Yes:

- **Every Kindle → E-Readers ✓**
- **Every Fire tablet variant → Fire Tablets ✓**
- **Batteries separated from electronics ✓**
- **Named clusters manually after inspecting members**

Reviews by Cluster

Cluster	Reviews
Fire Tablets	14,396
Batteries	12,086
E-Readers	1,049
Speakers	632
Accessories	138
Media	31

Average Rating by Cluster

Cluster	Avg Rating (out of 5)
Fire Tablets	4.56
Batteries	4.45
E-Readers	4.66
Speakers	4.54
Accessories	4.31
Media	4.39

Product Clusters (TF-IDF + PCA)

- Accessories (11)
- Batteries & Household (7)
- E-Readers (10)
- Fire Tablets (20)
- Media & Home (8)
- Smart Speakers (9)

Kindle Voyage E-reader, 6 Hi...

Amazon 9W PowerFast Official...

AmazonBasics AAA Performance...
AmazonBasics External Hard D...

Amazon Tap Smart Assistant A...

Fire HD 8 Tablet with Alexa,...

PCA Component 1

PCA Component 2

API Output (Claude Sonnet)

The Fire HD 8 with Alexa stands out as
the category leader with 2,445 reviews
and a 4.6 rating. Users consistently
praise the value for money, screen
quality, and Alexa integration.

However, app limitations and occasional
slowness are recurring complaints.
For budget-conscious buyers seeking a
reliable entertainment tablet, the Fire
HD 8 remains the top recommendation.

$0.43

vs

Local Output (Flan-T5-base)

Fire HD 8 Tablet with Alexa has
rating 4.6 from 2445 reviews.

$0.00

Same data. Same task.

250M params vs 70B+ params

Model 3 - Summarization

Task: Generate recommendation articles per category

Two parallel approaches:

API - Claude Sonnet

- Sample 5 positive, 5 negative, 3 neutral reviews per product
- Send with stats to LLM via structured prompts
- Output: 600-word articles with top products, complaints, recommendations
- Cost: **$0.43 total** for 39 summaries

Local - Flan-T5-base (250M params)

- Build evidence briefs with product rankings + TF-IDF complaints
- Smart ranking: `avg_rating × log(review_count)`
- Run entirely on GPU, no API needed
- Cost: **$0.00**

Model 3 - Evaluation

How Do You Evaluate Generated Text?

```
                | API (Claude)  | Local (T5)
-----------------------------------------------------------------
Structure       | All sections  | No sections
Length          | ~600 words    | 1–2 sentences
Factual accuracy| Correct stats | Correct but minimal
Actionable advice| Buy/Consider | None
Complaints      | In narrative  | Raw TF-IDF terms
Cost            | $0.43         | $0.00
```

Both approaches deployed to the website side-by-side
Users can switch between API and Local tabs to compare
Demonstrates the quality-cost tradeoff transparently

https://giwojno.pl/review-analyzer/

Claude Sonnet (API) — $0.43

Output length	~600 words
Sections	5 of 5
Actionable advice	Buy / Consider / Avoid
Complaint detail	Woven into narrative
Quality	

Flan-T5-base (Local) — $0.00

Output length	1-2 sentences
Sections	0 of 5
Actionable advice	None
Complaint detail	Raw TF-IDF terms
Quality	

Model 3 - Evaluation

How Do You Evaluate Generated Text?

	API (Claude)	Local (T5)
Structure	All sections	No sections
Length	~600 words	1–2 sentences
Factual accuracy	Correct stats	Correct but minimal
Actionable advice	Buy/Consider	None
Complaints	In narrative	Raw TF-IDF terms
Cost	$0.43	$0.00

Both approaches deployed to the website side-by-side
Users can switch between API and Local tabs to compare
Demonstrates the quality-cost tradeoff transparently

https://giwojno.pl/review-analyzer/

Output Length (words)

~600 (API Claude) · ~15 (Local Flan-T5)

Structured Sections

5/5 (API Claude) · 0/5 (Local Flan-T5)

Total Cost (USD)

$0.43 (API Claude) · $0.00 (Local Flan-T5)

Categories — API Model Output

Generated by **Claude Sonnet** (claude-sonnet-4-20250514) via the Anthropic API. Each category summary is a full recommendation article (~600 word top products, common complaints, worst product, and buying advice. Total cost: $0.43 USD for 39 API calls. Temperature: 0.3 for consistent, factual outp

← Back to categories

Fire Tablets

20 products   14,396 reviews   ★ 4.56

> Generated by Claude Sonnet (claude-sonnet-4-20250514) via Anthropic API

Fire Tablets: A Comprehensive Buying Guide Based on Customer Reviews

Category Overview

Fire Tablets represent Amazon's budget-friendly entry into the tablet market, with 20 different models earning an impressive 4.56/5 averag rating across 14,396 customer reviews. The category shows strong customer satisfaction, with most models achieving over 90% positive reviews, making it a compelling choice for budget-conscious consumers seeking basic tablet functionality.

Top 3 Recommended Products

1. Fire HD 8 Tablet with Alexa (16GB, Tangerine)
With 2,443 reviews and a 4.6/5 rating, this model leads in both popularity and satisfaction. Customers praise its excellent value at around $89, calling it "fast at everything except gameplay" with "great stereo speakers." The 8" HD display receives consistent praise for movies and reading, though some note it struggles with direct sunlight.

2. All-New Fire HD 8 Tablet (16GB)
Nearly matching the top model with 2,370 reviews and 4.58/5 rating, customers highlight its superior screen quality compared to the 7" models and adequate performance for web browsing and streaming. Parents particularly appreciate the FreeTime parental controls availab across Fire models.

3. Fire Kids Edition Tablet (7" Display, Pink)
Leading the kids' category with 1,676 reviews and 4.53/5 rating, this tablet stands out for its exceptional durability. Parents rave about the "thick, heavy duty" protective case that withstands drops, bites, and rough handling. The 2-year no-questions-asked replacement warranty including cracked screens, receives universal praise.

Categories — Local Model Output

Generated by **Flan-T5-base** (google/flan-t5-base) — a 250M parameter seq2seq transformer running locally on GPU (NVIDIA RTX 4050). Unlike the API-based summaries, this model runs entirely offline with no API costs. The smaller model size results in shorter, more factual outputs compared to the Claude Sonnet API approach.

← Back to categories

Fire Tablets

20 products   14,396 reviews   ★ 4.56

> Generated by Flan-T5-base (250M params) running locally on GPU

Fire Tablets

The Fire HD 8 Tablet with Alexa, 8 HD Display, 16 GB, Tangerine - with Special Offers has rating 4.6 from 2445 reviews. The All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Black has rating 4.58 from 2370 reviews. The lowest rated is Fire Tablet, 7 Display, Wi-Fi, 16 GB - Includes Special Offers, Black with rating 4.51 from 1028 reviews.

Top Picks

Fire HD 8 Tablet with Alexa, 8 HD Display, 16 GB, Tangerine - with Special Offers
Rating: 4.6 • Reviews: 2445
Complaints: tablet; just; good; bought; amazon

All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Black
Rating: 4.58 • Reviews: 2370
Complaints: tablet; just; good; amazon; apps

Fire Kids Edition Tablet, 7 Display, Wi-Fi, 16 GB, Pink Kid-Proof Case
Rating: 4.53 • Reviews: 1702
Complaints: tablet; charge; apps; slow; good

Takeaway

What we built: A complete NLP pipeline from raw reviews → deployed web app

- **95.48% accuracy → 6 clusters → 39 summaries → Live website**
- **Both API ($0.43) + Local models deployed side-by-side**
- Live at https://giwojno.pl/review-analyzer

What We Learned

- **Class weights matter** - they doubled Neutral F1 from 0.27 to 0.57
- **Model size matters** - 250M params can't do what an API model does for $0.43
- **Keep it simple** - static site, no backend, pre-compute everything

Bonus point pitch: We deployed a fully functional public website on real hosting - not localhost, not a notebook, not Gradio. A static site architecture that pre-computes all results, needs zero backend, and works on the cheapest shared hosting available. You can open it right now on your phone.

giwojno.pl/review-analyzer

Takeaway

What we built: A complete NLP pipeline from raw reviews → deployed web app

- **95.48% classification accuracy (RoBERTa)**
- **6 meaningful product clusters (K-Means)**
- **39 AI-generated summaries for $0.43**
- **Both API + Local models deployed side-by-side**
- Live at https://giwojno.pl/review-analyzer

Bonus point pitch: We deployed a fully functional public website on real hosting - not localhost, not a notebook, not Gradio. A static site architecture that pre-computes all results, needs zero backend, and works on the cheapest shared hosting available. You can open it right now on your phone.

giwojno.pl/review-analyzer

Demo

giwojno.pl/review-analyzer/

ReviewAnalyzer

Dashboard | Categories (API) | Categories (Local) | Products | Reviews | About

Dashboard

Project 3: NLP Business Case — Group 4, IronHack AI Engineering Bootcamp
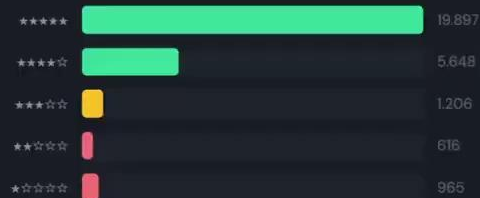
TOTAL REVIEWS	PRODUCTS	CATEGORIES	AVERAGE RATING
28.332	65	6	4.51 / 5

Sentiment Distribution

● Positive 89.9%   ● Neutral 4.3%   ● Negative 5.8%

Rating Distribution

★★★★★	19.897
★★★★☆	5.648
★★★☆☆	1.206
★★☆☆☆	616
★☆☆☆☆	965

Reviews by Category

Fire Tablets	14.396
Batteries & Household	12.086
E-Readers	1.049
Smart Speakers	632

Average Rating by Category

Fire Tablets	4.56/5
Batteries & Household	4.45/5
E-Readers	4.66/5
Smart Speakers	4.54/5