Project on

Automated Customer Review Analysis with NLP

Rishi

The Pitch

Imagine you're shopping on Amazon for a new tablet...

You see a product with **10,000 reviews**. Who has time to read them all?

Enter Automated Review analysis! 🤖

We built an AI system that reads **60,000 Amazon reviews** in minutes
figures out what customers really think,
writes professional recommendations—all automatically.

From 4 weeks of manual work → 10 minutes of automated magic! ✨

Introduction: The Real Problem

The Problem

- 59,630 reviews to analyze
- Manual reading takes weeks
- Businesses lose valuable customer feedback

The Solution

3-Stage NLP Pipeline:
1. Sentiment Classification (ML)
2. Product Clustering (TF-IDFKMeans)
3. AI Summarization (GPT-3.5)
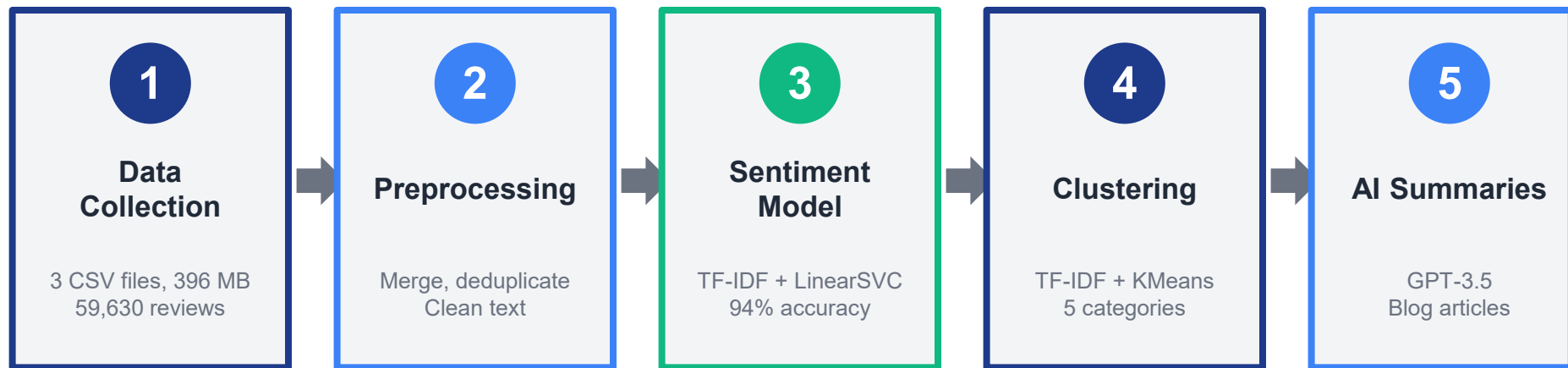
How the Models Work Together:

Model 1 (Sentiment) predicts if each review is positive/negative/neutral
Model 2 (Clustering) which groups products by similarity
Model 3 (AI Summaries) which generates blog-style recommendations

Methodology: Machine Learning + NLP + Generative AI

Methods: Pipeline Overview

1 **Data Collection**

3 CSV files, 396 MB
59,630 reviews

2 **Preprocessing**

Merge, deduplicate
Clean text

3 **Sentiment Model**

TF-IDF + LinearSVC
94% accuracy

4 **Clustering**

TF-IDF + KMeans
5 categories

5 **AI Summaries**

GPT-3.5
Blog articles

Dataset: Amazon Product Reviews | 59,630 reviews | 94 products | 5 categories | 8,358 duplicates removed

Model 1: Sentiment Classification

Problem: Classify 60,000 reviews as Positive, Neutral, or Negative

Approach:

Star Rating Mapping:
- 1-2 stars → Negative
- 3 stars → Neutral
- 4-5 stars → Positive

Model Architecture:

- TF-IDF Vectorization (1-2 word ngrams)
- LinearSVC Classifier
- Balanced class weighting
- 80-10-10 train-val-test split

Why This Approach?

✓ **Industry Standard**
LinearSVC is the go-to model for text classification. Fast, accurate, interpretable.

✓ **Handles Imbalance**
92% of reviews are positive (realistic!). Balanced weighting ensures model learns all classes.

✓ **Captures Context**
Ngrams capture phrases like 'not good' vs 'good'. Critical for sentiment.

Model 1: Evaluation Results
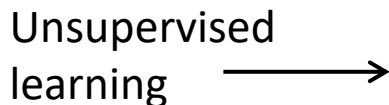
Overall Accuracy

93.85%

F1 Scores by Class

Positive: 0.97 (Excellent!) 🎯
Negative: 0.67 (Good) 👍
Neutral: 0.34 (Limited data) ⚠️

Also tried unsupervised learning
Weaklabels - lexcon dicitonary

Unsupervised learning

combined (review title+ review text) | weak_label

X Train | Y Train
Positive
positive
negative
.
.
.

X Test | Y Test

Compare with ratings

Similar accracy
Lower f1 scores
For negative
And neutral

Model 2: Product Clustering

Problem: Group 114 products into 5 meaningful categories

Approach:

Feature Engineering:

• TF-IDF on reviews,titles, etc
• High-dimensional text features

Dimensionality Reduction:
• PCA (Principal Component Analysis)
• Reduced to 2D for visualization

Clustering Algorithm:
• KMeans (k=5 clusters)
• Elbow method + silhouette analysis

Why This Approach?

✓ **Unsupervised Learning**
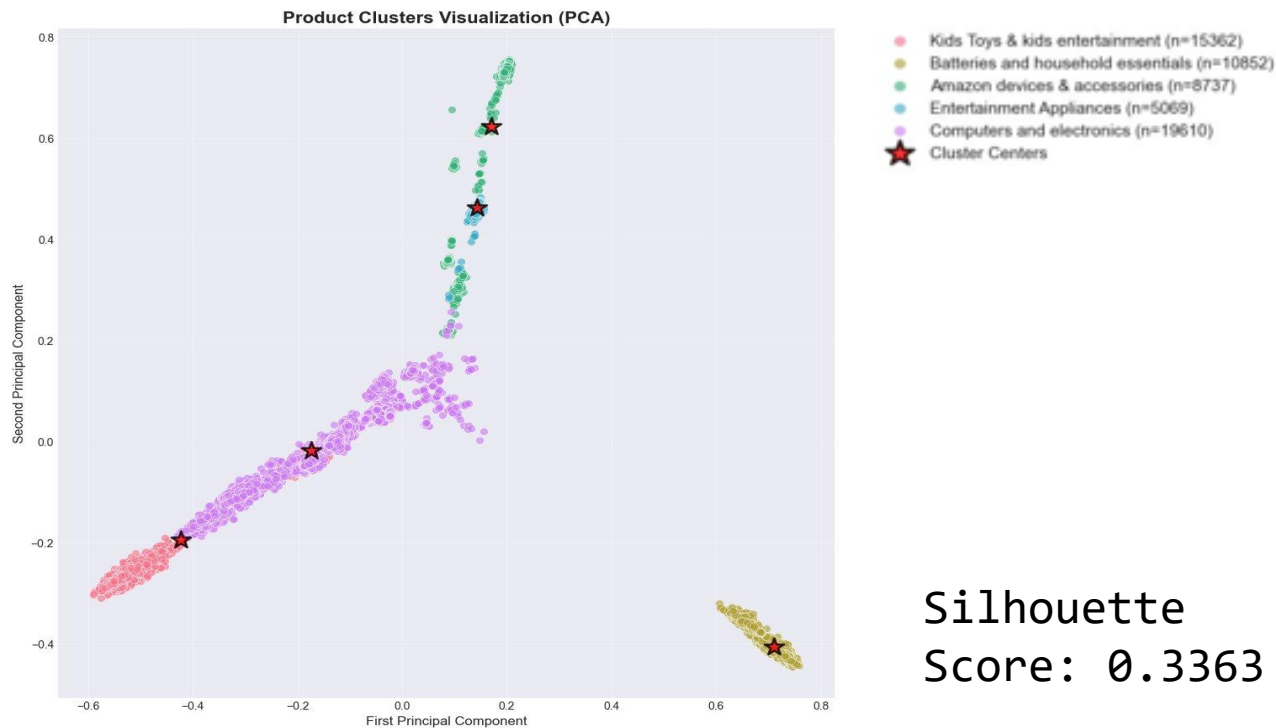No pre-labeled categories needed. Algorithm discovers natural groupings in data.

✓ **Interpretability**
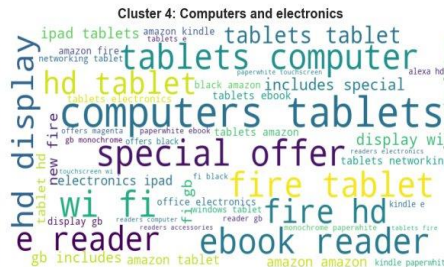PCA allows visual inspection. Can validate clusters make semantic sense.
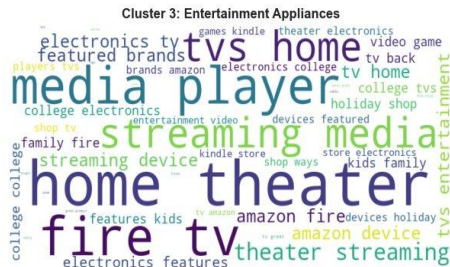
✓ **Optimal k=5**
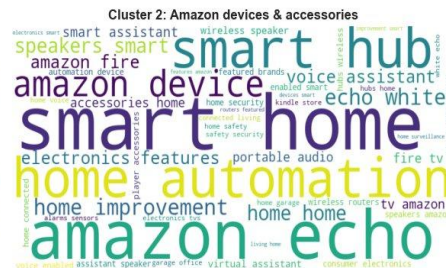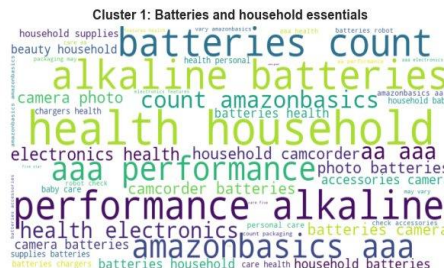Tested k=3 to k=8. Five clusters balanced granularity with manageability.

Model 2: Cluster Evaluation

Product Clusters Visualization (PCA)

Silhouette
Score: 0.3363

Cluster Inspection: Word Clouds

Cluster 0: Kids Toys & kids entertainment

Cluster 1: Batteries and household essentials

Cluster 2: Amazon devices & accessories

Cluster 3: Entertainment Appliances

Cluster 4: Computers and electronics
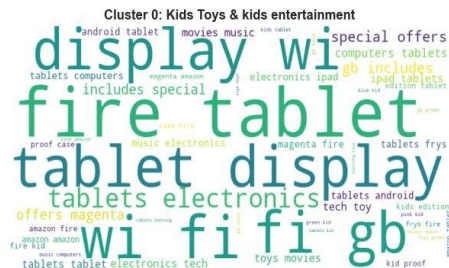
Final Clusters: Categories Identified

Each cluster shows clear semantic meaning and product similarity

Kids Toys & Kids Entertainment
Key terms: tablet, display, fire, wi-fi, electronics

n=15,362 reviews

Batteries and Household Essentials
Key terms: batteries, alkaline, performance, health

n=10,852 reviews

Amazon Devices & Accessories
Key terms: smart home, echo, alexa, voice, assistant

n=8,737 reviews

Entertainment Appliances
Key terms: tv, media player, streaming, home theater

n=5,069 reviews

Computers and Electronics
Key terms: computers, tablets, ebook reader, fire

n=19,610 reviews

Model 3: AI-Powered Summarization

Problem: Generate blog articles recommending top products per category

Approach:

Data Preparation:

• Merged sentiment + cluster data
• Quality Score = 60% sentiment + 40% rating
• Ranked products within each category

AI Integration:
• OpenAI GPT-3.5 Turbo API
• Structured prompts (600-800 word target)
• Includes: top 3, comparisons, complaints

Why This Approach?

✓ **Data-Driven Content**
Not generic descriptions—summaries based on actual review sentiment and ratings.

✓ **Structured Output**
Consistent format: intro, top products, comparisons, complaints, recommendations.

✓ **Business-Ready**
Professional tone suitable for e-commerce sites, blogs, or marketing materials.

Model 3: Example Output

Sample: Amazon Devices & Accessories Category

Amazon Devices & Accessories: Your Smart Home Starts Here

🏆 Top Pick: Amazon Echo Dot (4th Gen)

Quality Score: 0.965 | Avg Rating: 4.7 ★ | 3,242 reviews

Customers love: *"Great sound quality," "Alexa is responsive," "Perfect for small spaces"*

Common concerns: *"Setup can be tricky for non-tech users"*

[Article continues with 2 more products, detailed comparisons, and final recommendations...]

CATEGORY 1: COMPUTERS AND ELECTRONICS

Top 3 Products:

Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Black
- **Average Rating**: 4.67/5.0
- **Total Reviews**: 15
- **Quality Score**: 1.05/5.0

What Customers Love:
- Customers appreciate the responsiveness and ease of use of the touch screen.
- Reading on the Fire Tablet is comfortable and doesn't strain the eyes like some other devices.
- International shipping options from Amazon ensure timely delivery without customs hassles.

Key Features and Strengths:
- Receptive touch screen for easy navigation.
- Comfortable reading experience without eye strain.
- International shipping options for hassle-free delivery.

Common Complaints:
- Limited storage capacity for those with larger digital libraries.
- Some users experienced issues with app compatibility.

Who Should Buy This Product:
The Fire Tablet is ideal for casual readers and users looking for a budget-friendly device for reading and basic entertainment.

Key Takeaways

59,630

Reviews Analyzed

93.85%

Classification Accuracy

5

Product Categories

10 min

Total Processing Time

Questions?

Thank you!