Robot Adviser

G5 Adrian, Inna, Khushboo, Ralitza

From noisy feedback to confident buying decisions

Robot Adviser

Pick a category and sentiment to see a representative product and a concise summary.

Problem:

- Consumers drown in product reviews; teams waste hours turning noisy feedback into insights.

Our Solution:

- **Robot Adviser** is an **NLP-powered system** that turns **customer reviews** into buyer-ready **insights**. Robot Adviser is an **AI product-advisor** that uses real customer reviews, classifies sentiment, clusters products into clear categories, and writes buyer-friendly summaries of the top buying options.

Design and Methodology

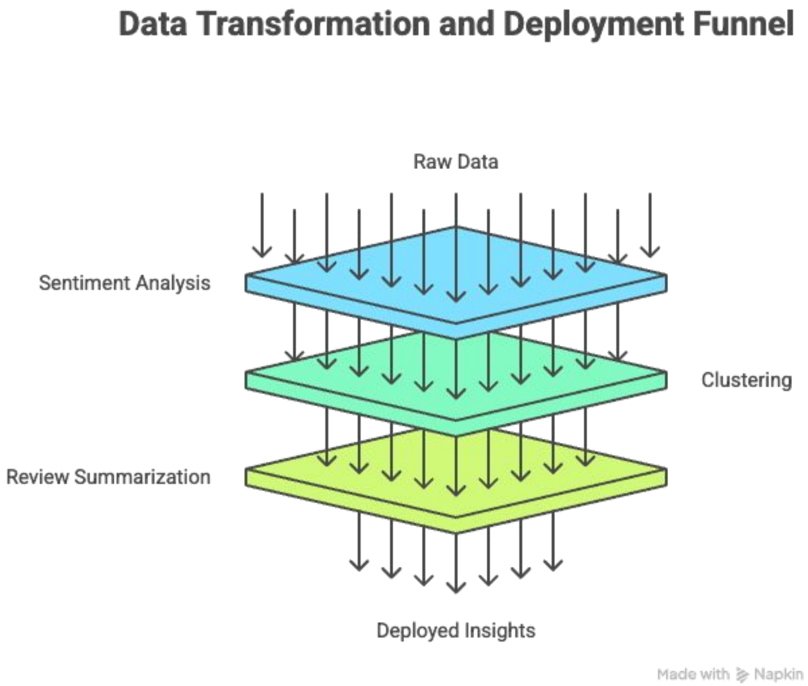
The System Design

- Data Layer
- Model 1: Sentiment Analysis (sentiment_m1)
- Model 2: Clustering (meta_category)
- Model 3: Review Summarizer (summaries)
- Deployment

The Methodology

- We approached the problem incrementally,

  then we iterated.



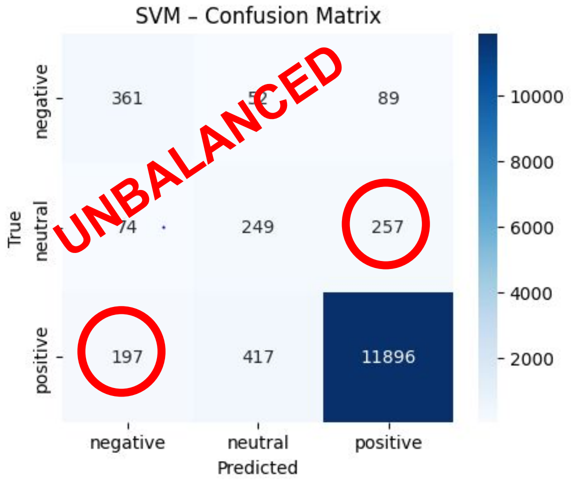
Methods

- We individually experimented with different approaches for
    - Data preprocessing and analysis the diferente CSVs
    - **Model 1:** Handling class imbalance and removing noise to improve macro F1
    - **Optimizing Models 1-3:** with various techniques**.**
    - Evaluating and visualizing models performance
    - Put all this information together during the meetings
- We also analyzed the impact of preprocessing on overall metrics.

1st Model 1

🔷 Overall Model Comparison:

	Accuracy	Macro F1	Weighted F1
Logistic Regression	0.9263	0.6504	0.9267
SVM	0.9201	0.6605	0.9246
XGBoost	0.9398	0.5906	0.9230



📘 Key Studies on Label Noise and Sarcasm in Sentiment Analysis

Study	Findings (Noise & Effects)	Proposed Solutions / Methods	Results / Improvements	Sarcasm / Irony Insights
Wang et al. (2019) – ACL "Learning from Noisy Labels for Sentiment Classification"	Found ~25% of samples mislabeled (text polarity ≠ rating). Noise caused overfitting and poor minority recall.	Proposed **noise-robust training** using label smoothing and confidence reweighting to discount uncertain examples.	Improved **Macro F1 by 8–10%**, maintained accuracy while reducing false positives.	Mentions **figurative language and irony** as one of the hardest noise sources; recommends reweighting uncertain examples to handle linguistic ambiguity.
Ghosh & Veale (2021) – arXiv "Are Ratings Enough? Exploring Text–Rating Mismatch in Amazon Reviews"	Identified **systematic bias**: ~25 % mismatched samples where users give high ratings but text is sarcastic or humorous.	Developed a **BERT-based relabeling model** to correct mismatched samples and proposed hybrid labeling (rating + text).	After relabeling, **Macro F1 improved 7–9 %**, with better balance and interpretability.	Explicitly highlight **sarcasm and humor** as systematic causes of label mismatch; sarcastic text often scored falsely as positive.
Sun et al. (2022) – Expert Systems with Applications "Noisy Labels in Sentiment Analysis"	Quantified inconsistency (18–27 %) and showed that noise lowers minority recall and Macro F1.	Combined **BERT text classification + rule-based heuristics** to auto-correct noisy labels.	Achieved **Macro F1 +12 %** after correction; consistent across datasets.	Notes that **sarcasm creates ambiguous polarity** for both humans and BERT; recommends using sarcasm-aware fine-tuning or removing highly ambiguous

Preprocessing STEP to clean the noise:

"Mismatch Between Review Text and Star Rating: **Due to Error and Sarcasm**"

star_rating	review_text	review_title	sentiment
5.0	is product so far has not disappointed. My children love to use it and I like the ability to monitor control what content they see with ease.	K	
5.0	great for beginner or experienced person. Bought as a gift and she loves it	v	
5.0	I f c		

Classical models struggle to identify sarcasm and irony

2. 1-Star Rating → Actually Positive (Ironic/Playful)

Stars: ⭐

Text:

"Terrible... now my kids won't leave this tablet for a second because they love it too much. Thanks a lot!"

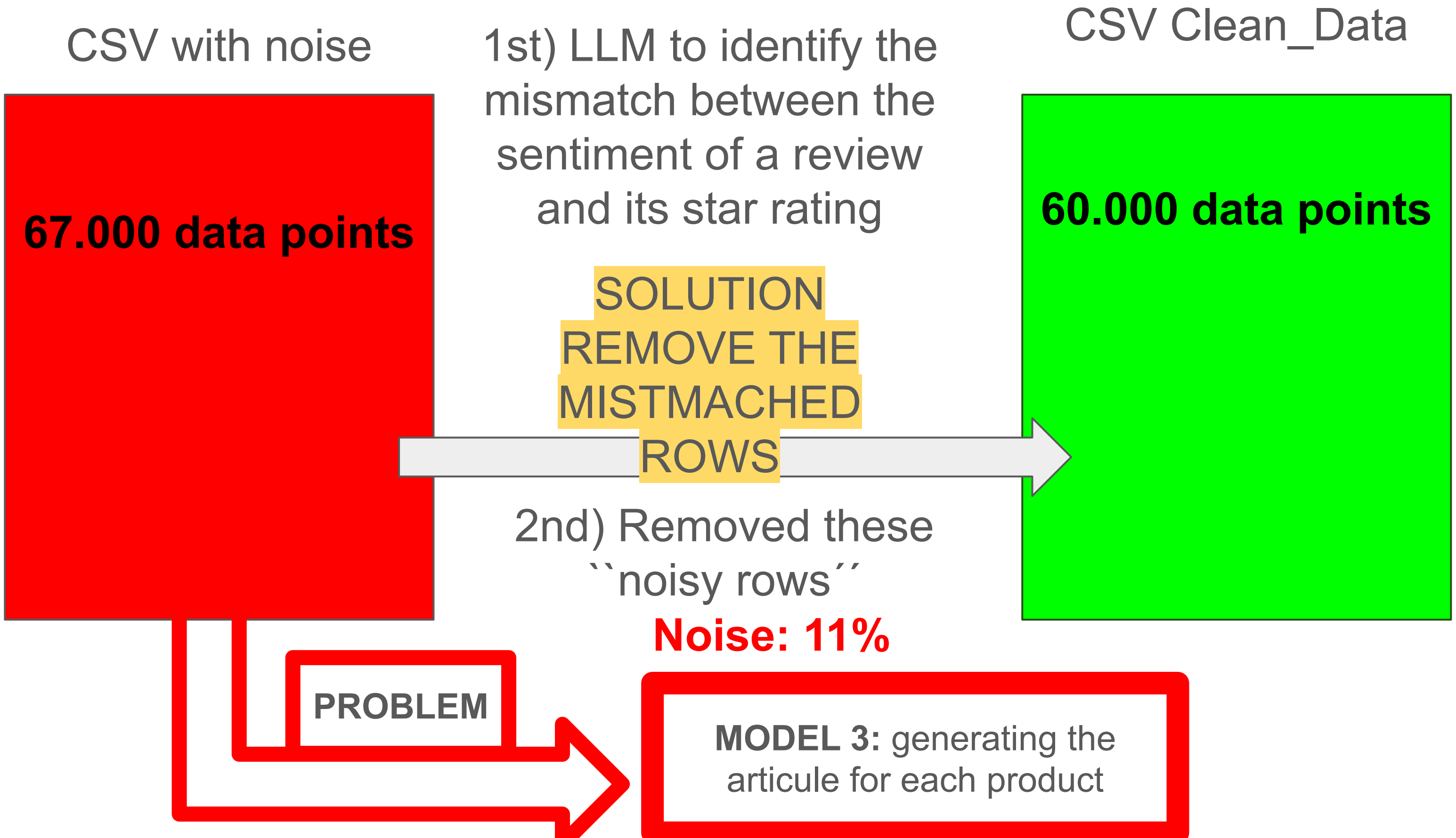
Meaning: Positive

Why? The complaint is sarcastic—the product works too well.

Causes:

1. **Mistake**
2. **Sarcasm**

Preprocessing STEP to clean the noise: relabeling?
MAKE IT SIMPLE (remove the noise)



Model 1

Classifies customer reviews into **positive, negative, or neutral** categories to help the company improve its products and services.

SVM Sentiment Classifier

F1-Score (Macro):   0.7116

F1-Score (Weighted): 0.9692

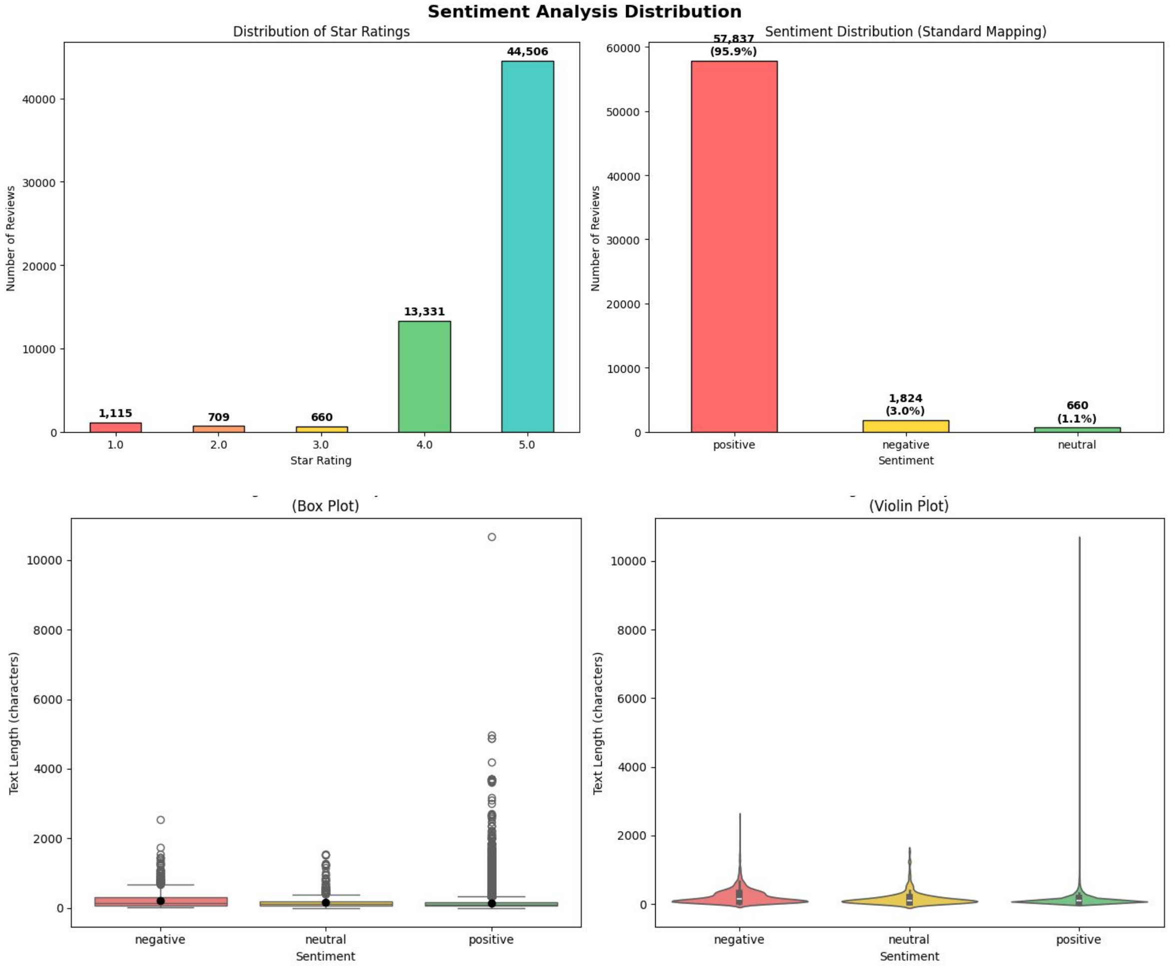
Data Imbalanced

- Sentiment Distribution Data Analysis
- Text Length Analysis by Sentiment

Optimized Pipeline for Imbalanced Data

- CURRENT CLASS DISTRIBUTION and Imbalance Ratio
- Preparing Features and Target
- Train-Test Split with Stratification
- Optimized TF-IDF Vectorization

Training Multiple Models with Imbalance Handling



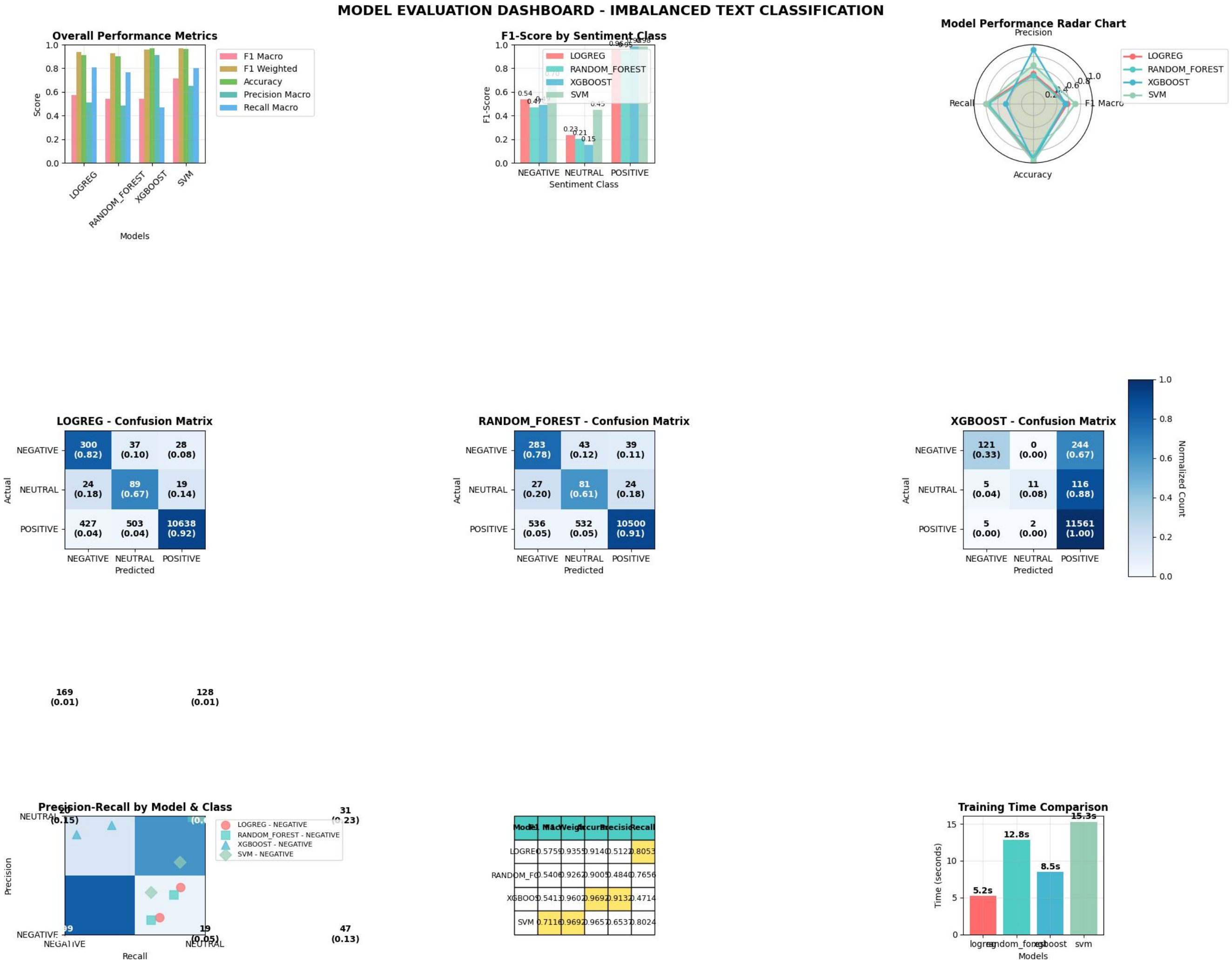
Model 1

Training Multiple Models with Imbalance Handling

1. Cost-sensitive Logistic Regression
2. Balanced Random Forest
3. XGBoost with Imbalance Handling
4. **SVM with Class Weights**
5. Model Evaluation and Comparison

SVM Sentiment Classifier

- F1-Score (Macro): 0.7116
- F1-Score (Weighted): 0.9692

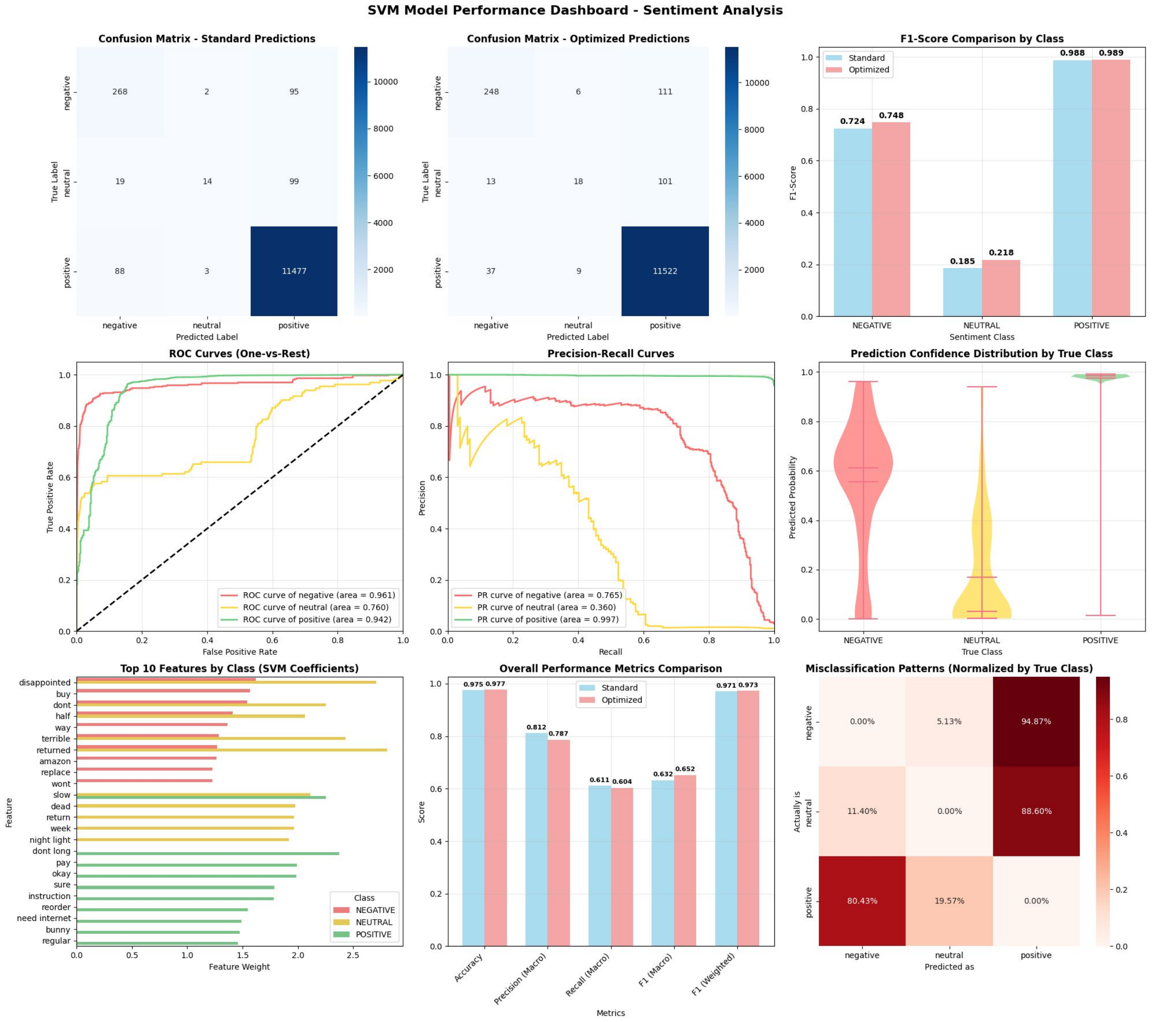


Mode	F1 Mac	Weigh	Accura	Precisio	Recall
LOGRE	0.5759	0.9355	0.9140	0.5122	0.8053
RANDOM_FO	0.5406	0.9262	0.9005	0.4840	0.7656
XGBOOS	0.5413	0.9602	0.9692	0.9132	0.4714
SVM	0.7116	0.9692	0.9657	0.6537	0.8024

Model 1

SVM Sentiment Classifier Optimized

- Enhanced Data Preparation
- Optimized Feature Engineering for SVM
- Advanced Class Weight Optimization
- Probability Calibration
- Advanced Threshold Optimization



SVM Model Performance - Sentiment Analysis

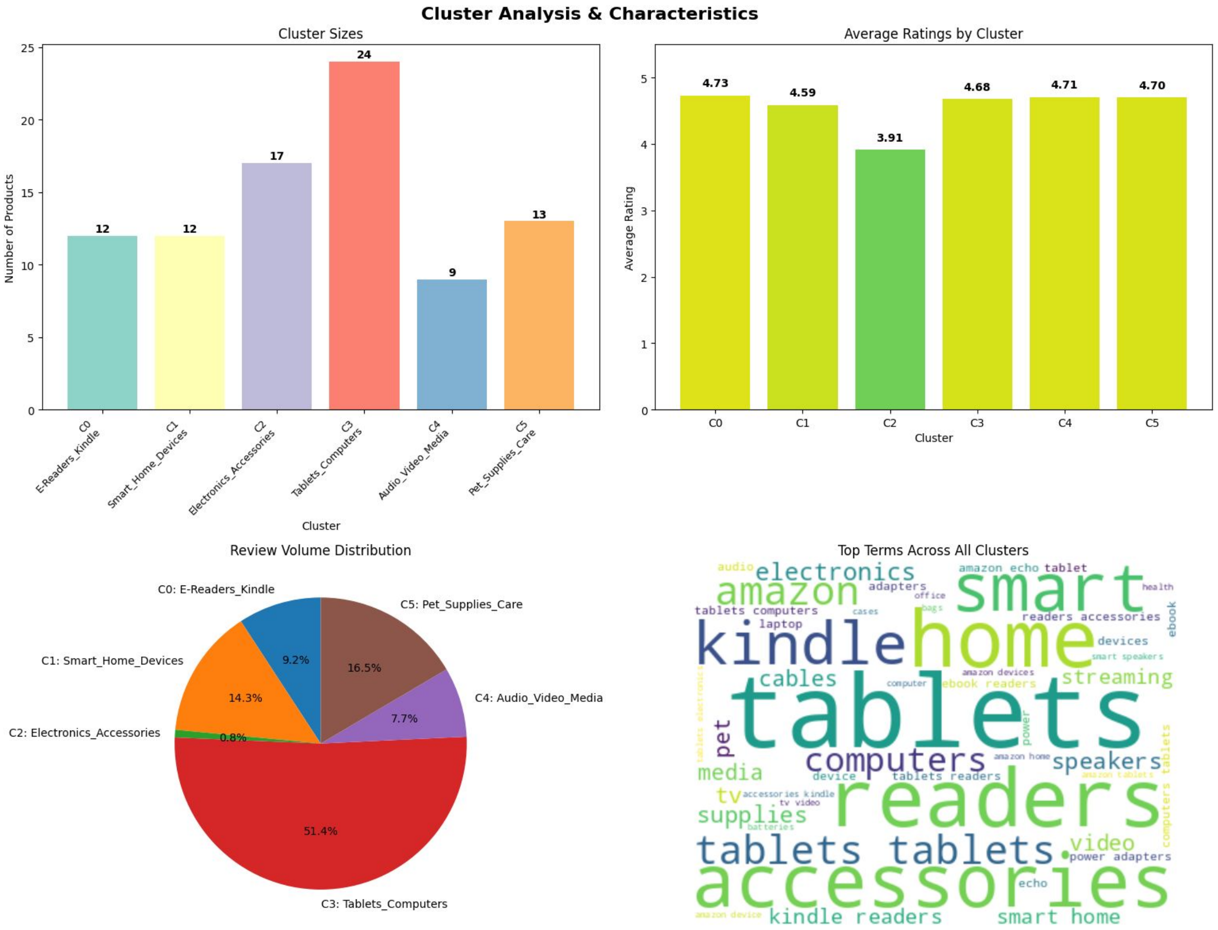
```
=========================================
🎯 MODEL EVALUATION COMPLETE - KEY INSIGHTS
==========================================

📊 OVERALL PERFORMANCE SUMMARY:
   F1 Macro (Optimized):      0.6517
   F1 Weighted:               0.9732
   Best Performing Class:     POSITIVE
   Worst Performing Class:    NEUTRAL

🚀 OPTIMIZATION IMPACT:
   Overall F1 Macro Improvement: +3.04%
```

Model 2

K-means Product Clustering

- It was challenging to find the correct no clusters
- We used optimal k Means

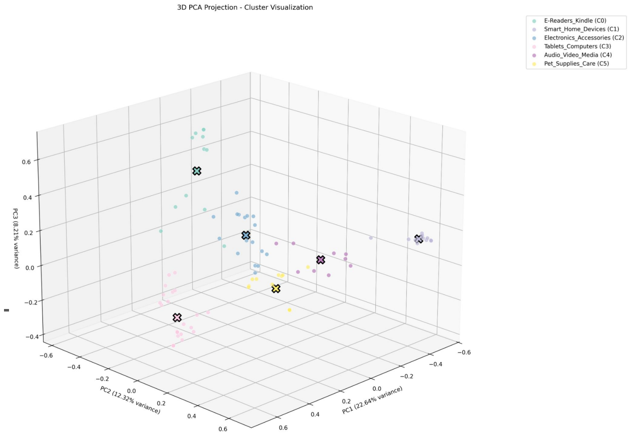
- **Data Exploration and Understanding**
- Data Analysis
- Data Cleaning & Text Preprocessing
  Extract useful columns & preprocess text
- Feature Extraction with TF-IDF Vectorizer
- Clustering with optimal k selection based on Elbow method & Silhouette score
- Cluster analysis and labeling
- Cluster validation

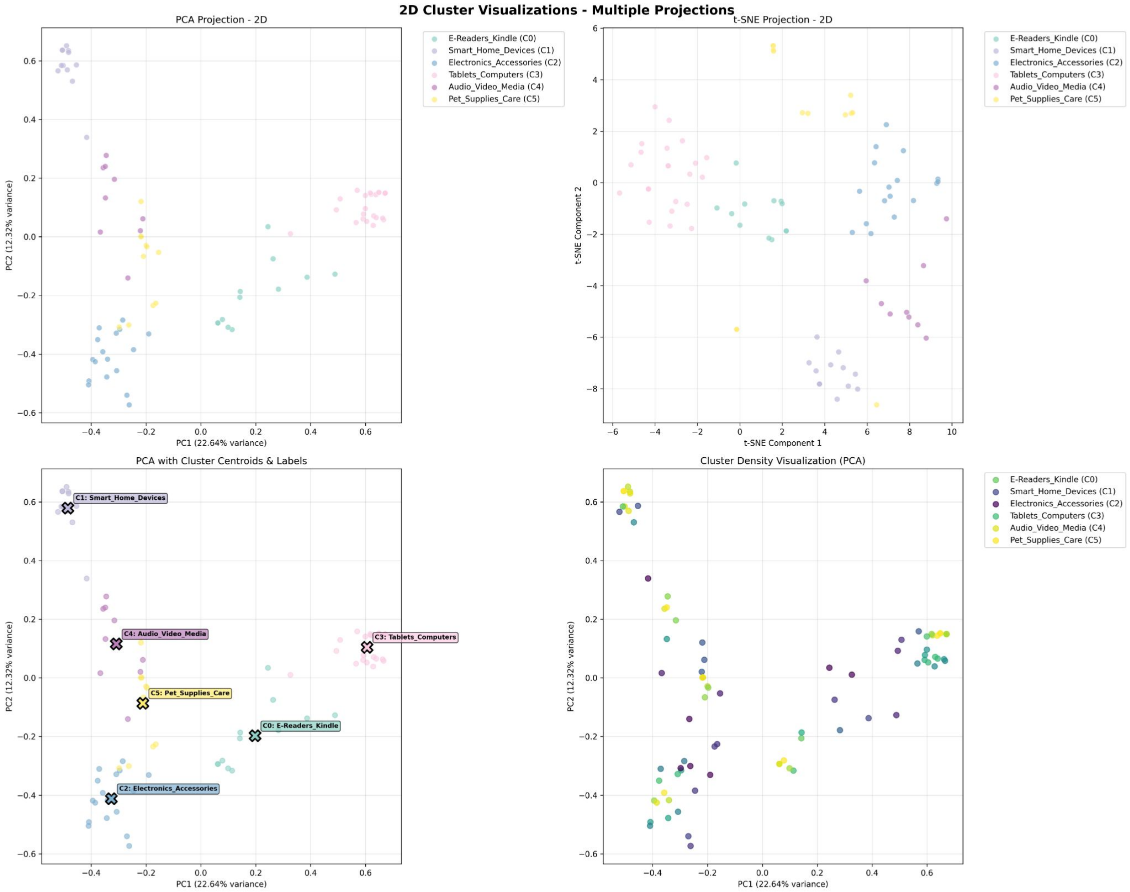


Model 2

K-means Product Clustering

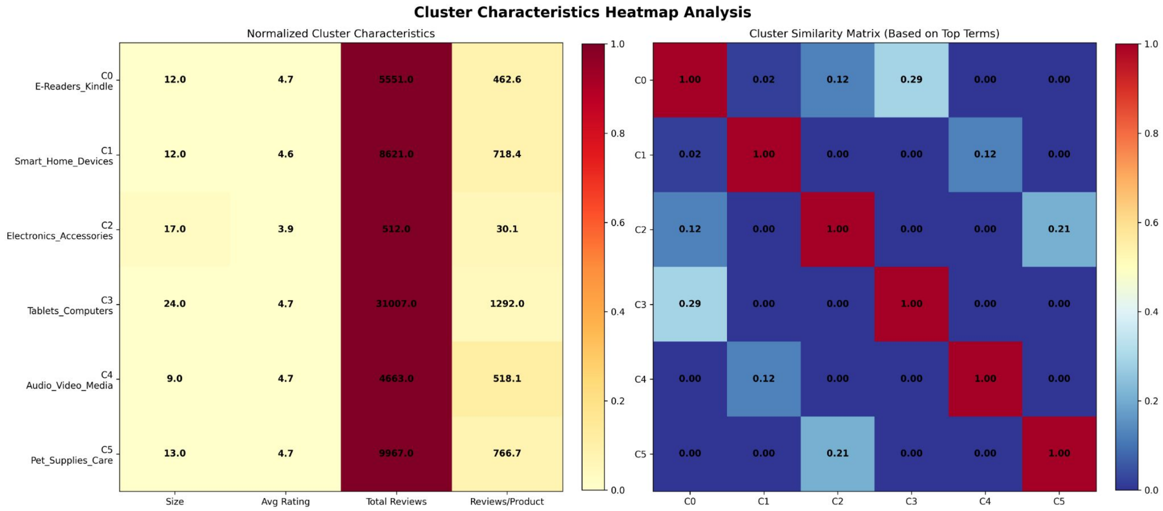
6 Meta - Cluster Visualization





Model 2 - Evaluation

```
===================================
META-CLUSTERING QUALITY EVALUATION
===================================
1. Evaluating Meta-Clustering
Quality...

   A. INTERNAL VALIDATION METRICS:
   • Silhouette Score: 0.283
     ⚠️  Fair: Weak cluster structure
   • Calinski-Harabasz Index: 16.35
     ❌ Poor: Weak cluster separation
   • Davies-Bouldin Index: 1.714
     ❌ Poor: Poor cluster separation

   B. CLUSTER STABILITY ANALYSIS:
   • Cluster Size Statistics:
     - Sizes: [12, 12, 17, 24, 9, 13]
     - Mean: 14.5, Std: 4.9
     - Min: 9, Max: 24
     - Size Ratio (max/min): 2.7x
     ✅ Good: Balanced cluster sizes
```



Model 3

An automated solution that summarizes reviews and highlights top products can save time and guide purchasing decisions.

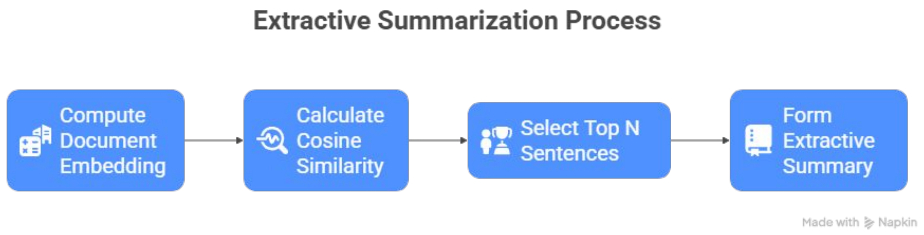
we leveraged **Generative AI** combined with **NLP** techniques to create concise, meaningful summaries of product reviews. Our approach was two-fold:

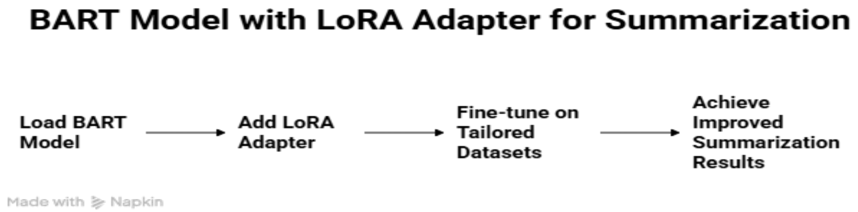
1 Step: Extract Summarization

model:distilbert-base-nli-mean-tokens

2 step: Abstractive Summarization

model:facebook/bart-large-cnn with LoRA optimization





Model 3 - Evaluation

We have Used ROUGE and BERTScore for Evaluation :

ROUGE: {'rouge1': np.float64(0.3398268398268398), 'rouge2': np.float64(0.09664351851851852), 'rougeL': np.float64(0.22835497835497834), 'rougeLsum': np.float64(0.22835497835497834)} BERTScore: {'precision': [0.8687559366226196, 0.8922217488288879], 'recall': [0.8964695334434509, 0.9034714698791504], 'f1': [0.8823952078819275, 0.8978114128112793], 'hashcode': 'roberta-large_L17_no-idf_version=0.3.12(hug_trans=4.57.1)'}

This the summary in 2 steps:

Extractive Summary:

LOVE the Echo, it's everything expected and moreEasy setup.Easily connected to homes Harmony setup. The various reminders you can set are really cool! Just getting use to the many applications it can be use for.Love the Echo. am now looking forward to get the lights connected to this as well.. Love it so far. love it and use it all the time.This device is awesome if you already have a prime subscription. Love having music available on command. It does what say and more.. very happy with the purchaseThis product is amazing. would definitely recommend.Easy to set up and lots of great functions. As well as the information predownloaded is awesome. 'm glad took the opportunity to buy the echo especially when it was on ! love Echo; the technology is user friendly helps with everyday items. It's pretty addictive too.Just as you would expect it sounds great and works.Item works as advertised. bought one and are having all kinds of interesting fun. Has turned into a great every day companion.Bought it to try some features. love that the Echo responds well to anything say. Very useful and fun item.The Echo is everything that's advertised: smart, intuitive, and has a plethora of information and features at her metaphorical fingertips. This is an awesome device.Coupled with Wink, this becomes an invaluable tool. Needs lot of improvement on the API's.This product is so easy to set up and use. This exceeded expectations and is so nice to tell Alexa to play a certain song or playlist. Happy that got it on !Echo was easy to install and start using.

Abstractive Summary:

The Echo is everything that's advertised: smart, intuitive, and has a plethora of information and features at her metaphorical fingertips. This exceeded expectations and is so nice to tell Alexa to play a certain song or playlist.

Deployment

Robot Adviser

Pick a category and sentiment to see a representative product and a concise summary.

Category

Smart Speakers & Portable Audio (631 items)

Sentiment

All

Matches: **631** rows from **clustered_reviews_kmeans_tfidf (3).csv**

Summary

Smart Speakers & Portable Audio: The top three are Certified Refurbished Amazon Echo (ASIN B01GAGYVU2), Amazon Tap Smart Assistant Alexaenabled (black) Brand New and Amazon Echo (2nd Generation) Smart Assistant Oak Finish Priority Shipping. Rankings weigh both average rating and review volume to reduce noise from small samples. Avoid Amazon Echo (1st Generationcertified) White Free Shipping (ASIN B00Y3QOH5G); it trails the category with a 4.31★ average across 13 reviews. Overall, choose models that pair strong ratings with substantial review counts for dependable performance.

Takeaways

- It was difficult to divide the work so we worked on the project individually and together at the same time
- We selected the most promising models based on evaluation and comparison
- We optimized reviews using different fine-tuning techniques
- We tried deployment with streamlit
- Finally put an extra effort in the whole process from data preprocessing to developing models, optimizing them, trying to connect them, to deploy them and to improve our end result - the summaries

Demo

- Demo Time!