Project Report – NLP for Automated Customer Reviews

Project Title

Automated Customer Review Analysis using NLP

Project Goal

The aim of this project is to automate the analysis of customer product reviews using NLP techniques. The system processes large volumes of customer feedback across platforms, classifies sentiments, clusters products into meta-categories, and generates review-based summaries to support purchasing decisions or business insights.

Problem Statement

Manually analyzing thousands of customer reviews is impractical and time-consuming. This project introduces an automated NLP-driven approach to:

- Classify customer sentiments.
- Cluster products into logical categories.
- Summarize reviews into meaningful recommendation articles.

Dataset Overview

- **Primary Dataset:** Amazon Product Reviews
- **Source:** Kaggle – "Consumer Reviews of Amazon Products"
- **Size:** ~5,000 product reviews
- **Fields Used:** Review Text, Rating, Product Name

---

1. Sentiment Classification

Objective

Categorize customer reviews as:

- **Positive**
- **Neutral**

- **Negative**

Label Mapping (Star Rating to Sentiment)

Star Rating	Sentiment Class
1 – 2	Negative
3	Neutral
4 – 5	Positive

Preprocessing Steps

- Lowercasing
- Punctuation removal
- Stopword removal
- Tokenization & Lemmatization
- TF-IDF for traditional models
- Transformers: Tokenization via Hugging Face DistilBertTokenizer

---

Model Used

- **Transformer:** distilbert-base-uncased (from Hugging Face)
    - Lightweight yet powerful
    - Fine-tuned for text classification
- Training handled via Trainer API (transformers library)
- 3-class classification head

Model Evaluation

Accuracy:

95.54%

Classification Report

	precision	recall	f1-score	support

Negative	0.87	0.79	0.83	325
Neutral	0.60	0.54	0.57	254
Positive	0.98	0.99	0.98	5088
accuracy			0.96	5667
macro avg	0.82	0.77	0.79	5667
weighted avg	0.95	0.96	0.95	5667

Confusion Matrix

[[256 41 28]

[20 138 96]

[18 50 5020]]

---

2. Product Category Clustering

Objective

Group products into 4–6 broader meta-categories.

Approach

- Extracted keywords from product names and descriptions
- Used TF-IDF + KMeans clustering
- Evaluated optimal number of clusters using the elbow method

Resulting Categories

1. **Ebook Readers**
2. **Batteries & Chargers**
3. **Computer Accessories**
4. **Household & Non-electronics**
5. **Coffee & Kitchen Pods**

---

3. Review Summarization (Generative AI)

Objective

Automatically generate readable recommendation articles per product category.

Model Used

- facebook/bart-large-cnn from Hugging Face
- Summarization fine-tuned on grouped reviews by cluster
- Output structured like a blog article

Each Article Includes

- **Top 3 Recommended Products**
- **Common Complaints & Praise**
- **Worst Performing Product & Reason**