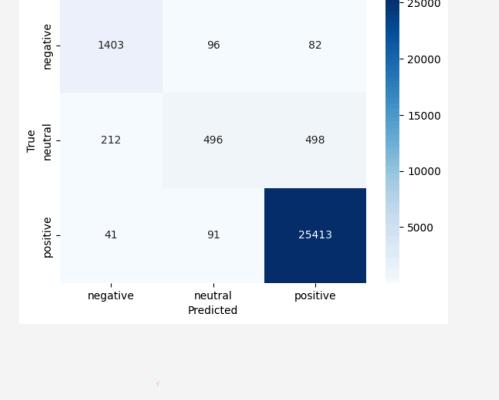


RoboReviews

- a. Classify customer reviews into positive, negative or neutral
- b. Cluster product categories
- c. Summarize reviews using GenAI

CLASSIFICATION

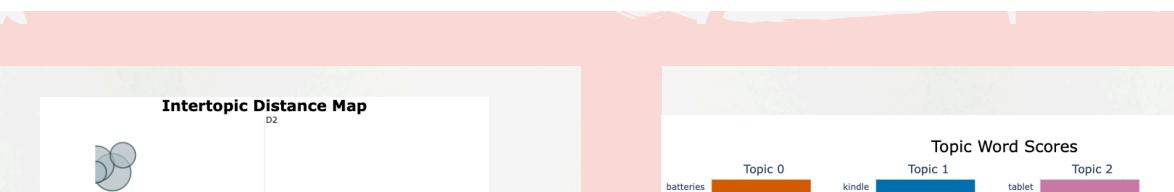
- Used Amazon reviews dataset to make the classification or the sentiment analysis (positive, negative or neutral)
- used pre trained 'bert-base-uncased' model to check the result, the initial results weren't satisfactory therefore fine-tuned this model.



CLUSTERING

- Used BERTopic – a topic modeling technique for clustering
- used the input from the classification i.e the 'positive' sentiments data to form clustering.

Hierarchical Clustering



Intertopic Distance Map



Topics

Most topics overlaps



Bar chart

to visualize the top 7 topics and the appearance of words into them

SUMMARIZATION

I used 2 models to summarize the reviews – facebook/bart-large-cnn & gpt-4o and compared output between them.

Observation

- BART-Large-CNN Strengths:
 - Concise and consistent summaries.
 - Less prone to redundancy.
 - Effective for high-level insights.
- GPT-4o Strengths:
 - Rich detail and specificity.
 - Highlights nuanced pros and cons of products.
 - Better context for different audiences.
- BART-Large-CNN Weaknesses:
 - Limited depth and detail.
 - Misses nuanced context.
- GPT-4o Weaknesses:
 - Verbose and occasionally repetitive.
 - Summaries can become overly detailed for broader overviews.

Recommendation:

- Use BART-Large-CNN for quick, high-level overviews.
- Use GPT-4o when detailed, audience-specific insights are required.

References:

<https://maartengr.github.io/BERTopic/index.html> – on Bertopic