Data Scientist II Technical Challenge

Task 4: Real-world Scenario

Cristina Sánchez Maíz | csmaiz@gmail.com | LinkedIn

Consider the following business problem:

Your company wants to improve customer satisfaction by understanding the main topics and sentiments expressed in customer reviews. Your task is to:

- Use topic modeling to identify the main topics in the customer reviews.
- Summarize the findings and suggest actionable insights for business improvements.

Deliverables:

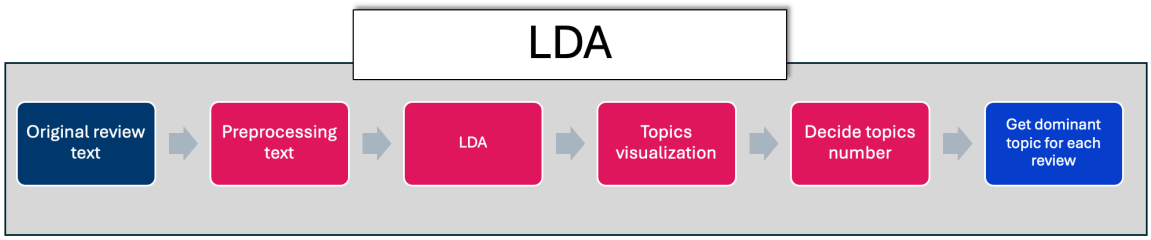
- - Explanation of the topic modeling process.
- - Summary of the main topics identified.
- - Actionable insights based on the analysis.

Explanation of the topic modeling process

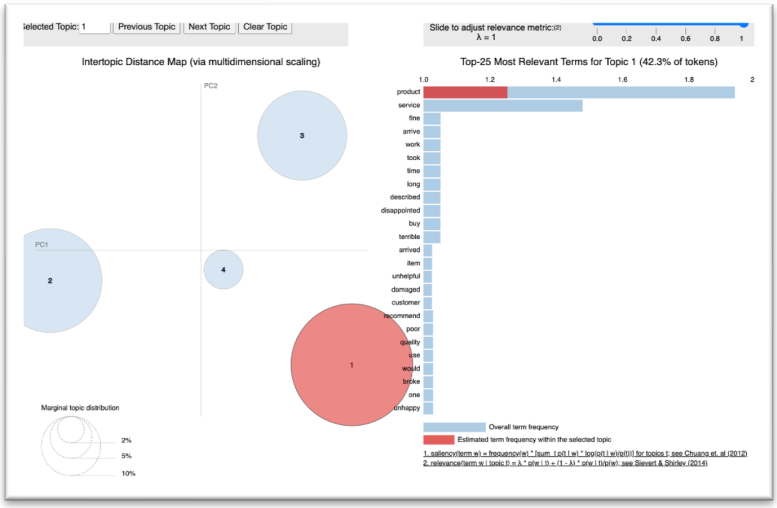
Topic modeling is applied to find the main topics within the reviews. By combining topic modeling and sentiment analysis (Task 3), we can gain valuable insights into customer feedback and drive business improvements.

I applied Latent Dirichlet Allocation (LDA), one of the most popular topic modeling methods, and a Large Language Model (LLM). I explaing both processes and then I compare the results.

1. LDA



- Preprocessing text: refer to Task 3 for more details.
- LDA: The aim of LDA is to find topics a review belongs to, based on the words in it.
- Topics visualization: I used pyLDAvis to visualize the topics. Choosing 4 as the number of topics, we can see that they do not overlap.



The reviews are distributed as follows:

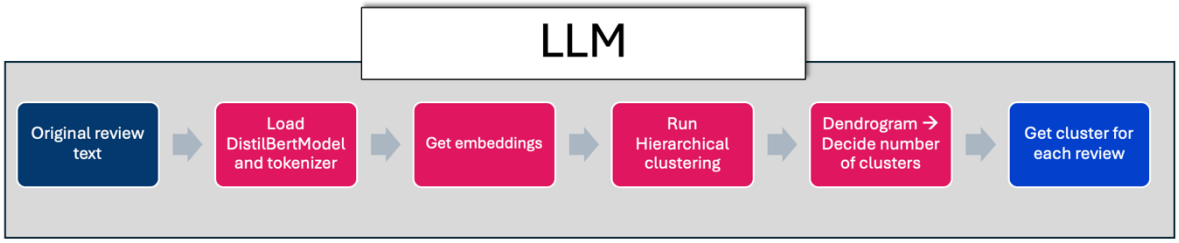
Topic 1:
- Very disappointed with the product, not as described.
- Poor quality, would not recommend.

Topic 2:
- The product works fine, but took a long time to arrive.
- The product broke after one use, very unhappy.

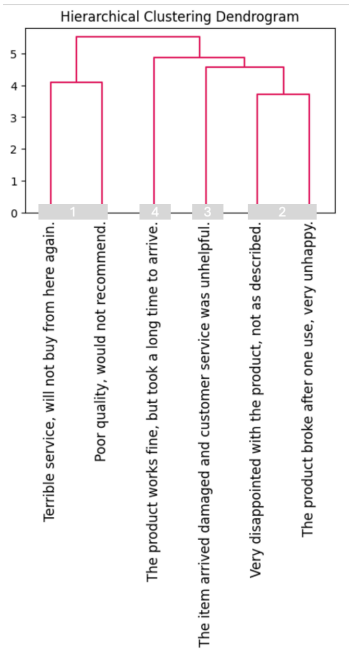
Topic 3:
- Terrible service, will not buy from here again.

Topic 4:
- The item arrived damaged and customer service was unhelpful.

2. Large Language Model



Use a pre-trained LLM like DistilBERT to generate dense vector representations (embeddings) for each review. Then, I run an aglomerative clustering algorithm to group reviews based on their similarity in the embedding space. The dendrogram allows us to determine the number of clusters.



The reviews are distributed as follows:

Topic 1:
- Terrible service, will not buy from here again.
- Poor quality, would not recommend.

Topic 2:
- Very disappointed with the product, not as described.
- The product broke after one use, very unhappy.

Topic 3:
- The item arrived damaged and customer service was unhelpful.

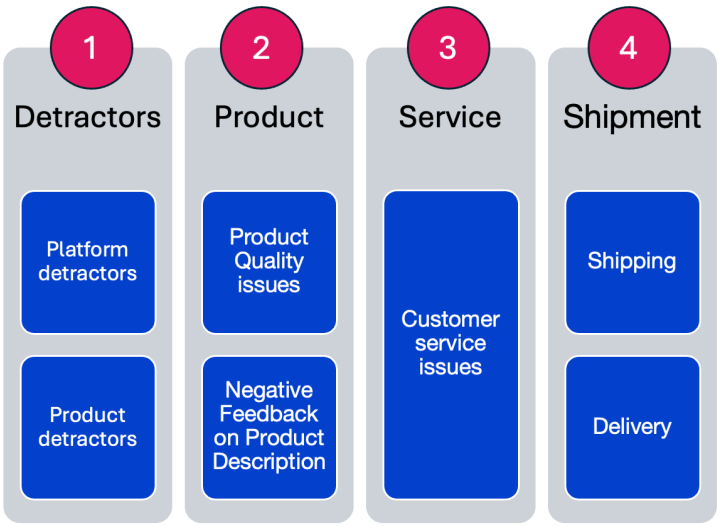
Topic 4:
- The product works fine, but took a long time to arrive.

Summary of the main topics identified.

topic_lda and topic_llm are the topics assigned by LDA and LLM, respectively. Note that the number of the cluster is just a label.

review_text	topic_lda	topic_llm
Terrible service, will not buy from here again.	2	1
Very disappointed with the product, not as described.	1	2
The item arrived damaged and customer service was unhelpful.	4	3
Poor quality, would not recommend.	2	1
The product works fine, but took a long time to arrive.	1	4
The product broke after one use, very unhappy.	4	2

The assignment made by LDA is a bit more confusing than with LLM. With this second method, it is easier to find the meaning of the topics. I take the topic assigment done by LLM.



Actionable insights based on the analysis.

Based on the identified topics and sentiments, here are some actionable insights for the business:

Improve Customer Service:
- Conduct training sessions for customer service representatives to improve their communication skills and responsiveness.
- Implement a better tracking system for customer inquiries to ensure timely responses.

Enhance Product Descriptions:
- Review and update product descriptions to ensure they accurately reflect the product features and quality.
- Include more detailed images and customer reviews to set realistic expectations.

Address Product Quality Issues:
- Investigate and address any recurring quality issues in the products.
- Consider quality checks and improvements in the production process.

Streamline Delivery and Shipping:
- Partner with reliable shipping companies to ensure timely and accurate deliveries.
- Implement a tracking system for customers to monitor their orders in real-time.

Additionally, compare topics over time for detecting trends and emerging issues.