LAVAZZA

TORINO, ITALIA, 1895

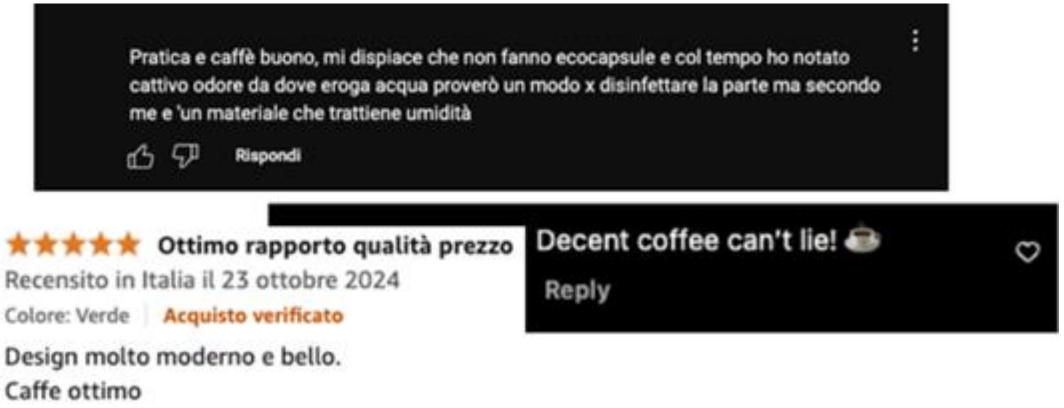
Leveraging Large Language Models for Marketing Analytics

Table of Contents

Final presentation

Project Value Proposition

For **marketers** seeking deeper **insights** on product launches, our **analytics software** transforms raw online comments and reviews into **actionable metrics**.



Project Goal

Politecnico di Torino

Developing **AI software** to accurately assess the **real impact of a product launch** by analyzing **consumer sentiment** across social media **comments** and third-party **reviews**.

Sustainable Development Goal:

12 RESPONSIBLE CONSUMPTION AND PRODUCTION

Responsible

Consumption

and Production

LAVAZZA

TORINO, ITALIA, 1895

Hypothesis

Politecnico di Torino

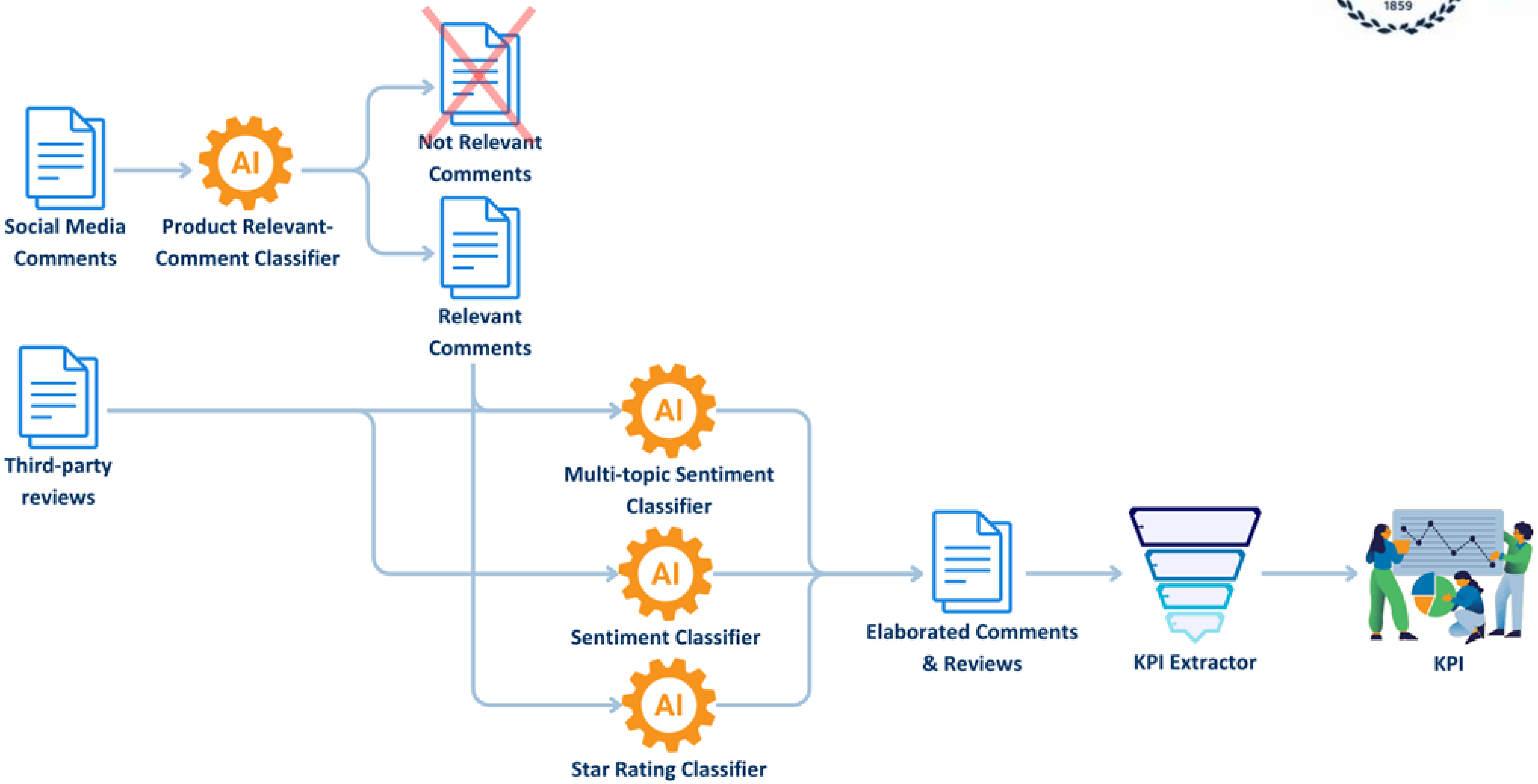
Large Language Models can be effectively used to extract valuable insights from consumer comments and reviews to enhance marketing strategies.

Specifically:

- **LLMs** can effectively identify **product-relevant comments**.
- **LLMs** can extract **key product aspects** from reviews sand comments and assign **appropriate sentiments** to each.
- **LLMs** can successfully assess **sentiment** in comments and assign **rating stars** to reviews.

LAVAZZA TORINO, ITALIA, 1895

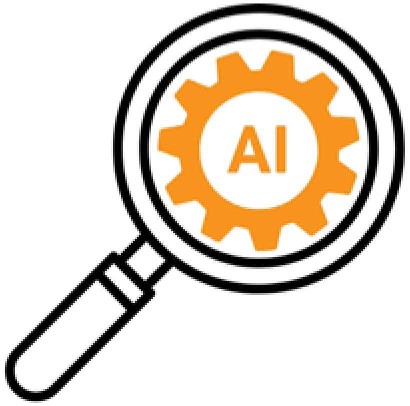
Method - Pipeline



Politecnico di Torino

LAVAZZA TORINO, ITALIA, 1895

Method - Classification Block and Models

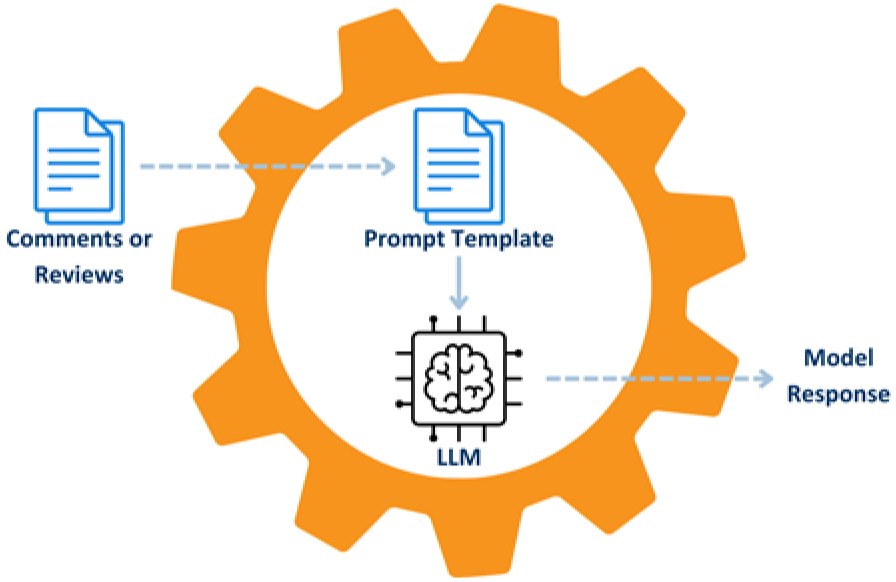


The **core** of the pipeline is the **classification block**, responsible for all classification tasks. Its structure adapts to the **model type** used:

- **General-Purpose Models (Chat-like LLMs):** Flexible and versatile for various tasks.
- **Specialized Task-Based Models:** Fine-tuned for specific tasks, offering higher precision in narrow domains.

Method - General Purpose Classification Block

The general purpose classification block comprises **two main components**:



Prompt Template:

- Clearly defines the **classification task**.
- Includes **examples** (few-shot prompting only).
- Specifies the **structured output format** for consistent model responses.

General-Purpose Model (LLM):

- **Processes** the **task-specific prompt** (combining prompt template and input data).
- Generates **task-specific responses**.

Method - General Purpose Models

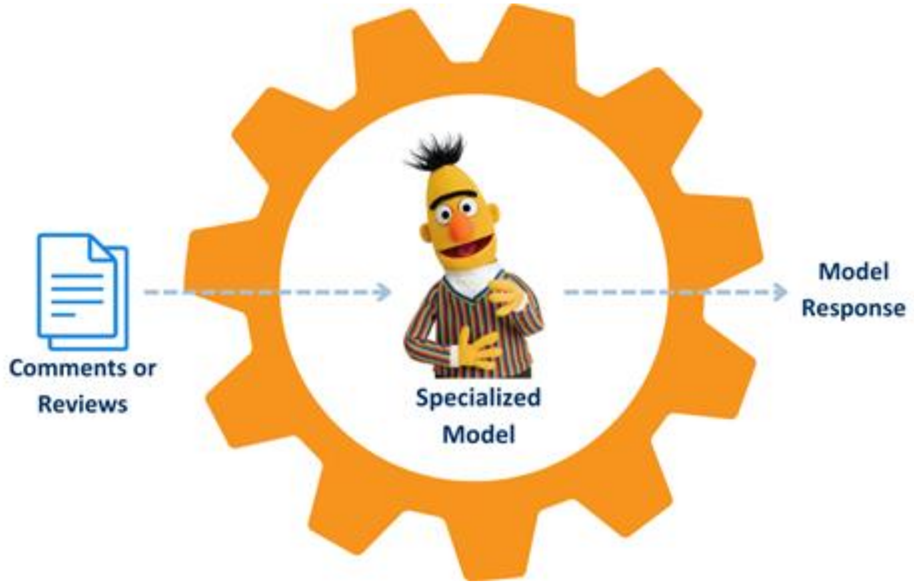
Gemma2: large language model from Google, different versions: 2B, 9B and 27B

Llama3: large language model from Meta, 8B, 70B and 405B

Mistral: large language model from Mistral AI

Method - Specialized Classification Block

The specialized model classification block comprises **two main components**:



Specialized BERT-base Model: Takes reviews or comments as input and generates a response based on the specific task.

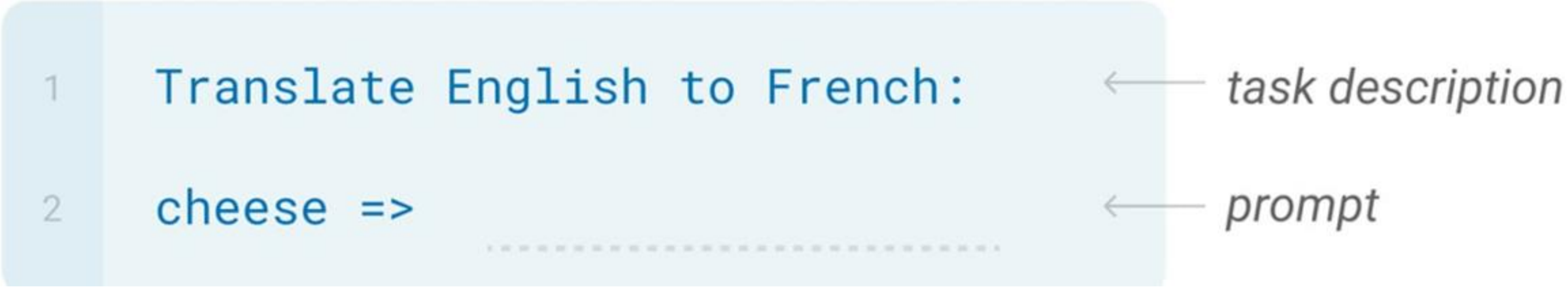
Method - Specialized Models

- **Sentiment Classification**
  - *distilbert-base-multilingual-cased-sentiments*
  - *twitter-xlm-roberta-base-sentiment-finetuned*
  - *twitter-roberta-base-sentiment-latest*
- **Star Rating Prediction**
  - *bert-base-multilingual-uncased-sentiment*
  - *multilingual-sentiment-analysis*

*from Hugging Face

Method - Prompting Approach

Zero-shot:
The model predicts the answer given only the description of the task in natural language.



Method - Prompting Approach

Few-shot: The model sees a few examples of the task.

```
Translate English to French:
sea otter => loutre de mer
peppermint => menthe poivrée
plush girafe => girafe peluche
cheese => ...............................
```

task description

examples

prompt

Experiments - Data

Politecnico di Torino

Data Sources:

Wonderflow | Digiming | YouTube | Instagram | Generative Models

Datasets:

Reviews | Comments | Synthetic Data

LAVAZZA
TORINO, ITALIA, 1895

Experiments - Datasets

Politecnico di Torino

Reviews:

- Records: **2930**
- Sampled records: **500**
- **Star Rating** label ration: **100%**
- **Multi-Topic Sentiment** label ratio: **95%**
- **Tiny Eco** Reviews**: 79**

Comments:

- Records: **577**
- **Labelled** Comments: **0%**
- **Tiny Eco** Comments**: 77**

Synthetic Data:

- **Comments** Records: **236**
- **Reviews** Records: **500**
- **Labelled** Comments and Reviews: **100%**

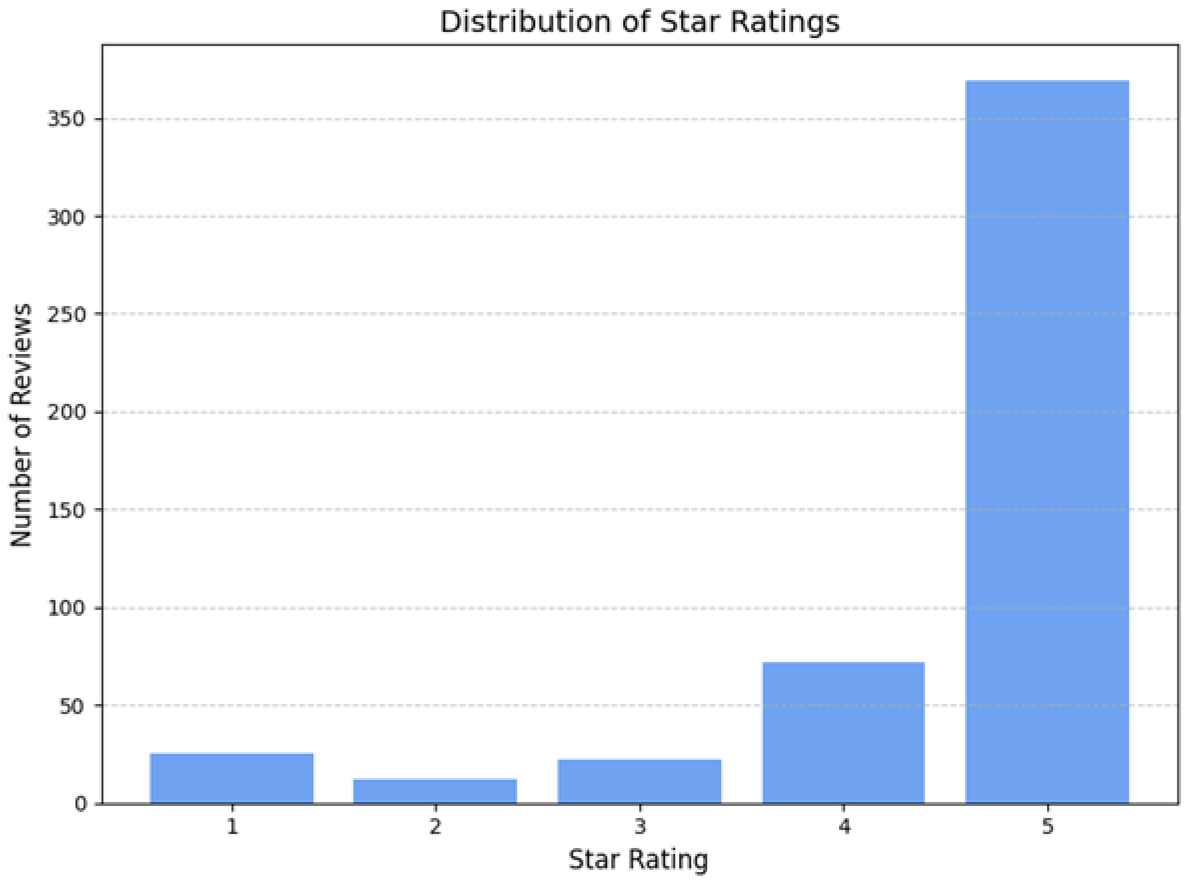
Some reviews may have incomplete labels, e.g.:

"Excellent machine. Fair price and guaranteed quality." **- Positive Aspects**: ["Price & worth"]

Where is "Coffee Quality"?

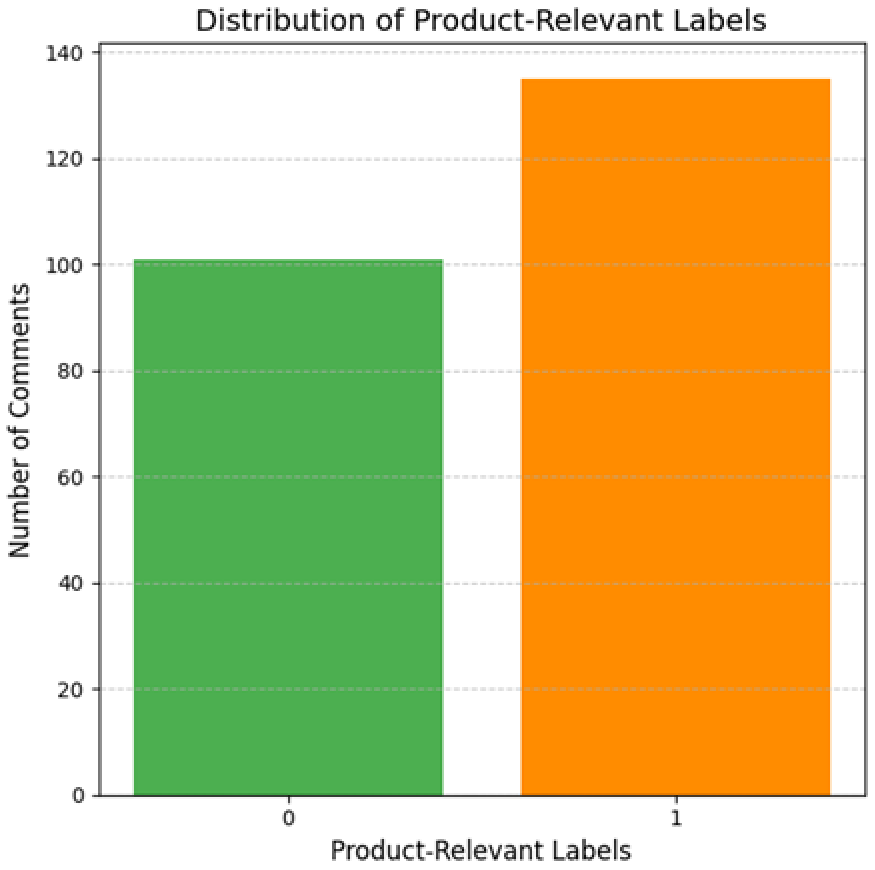
LAVAZZA TORINO, ITALIA, 1895

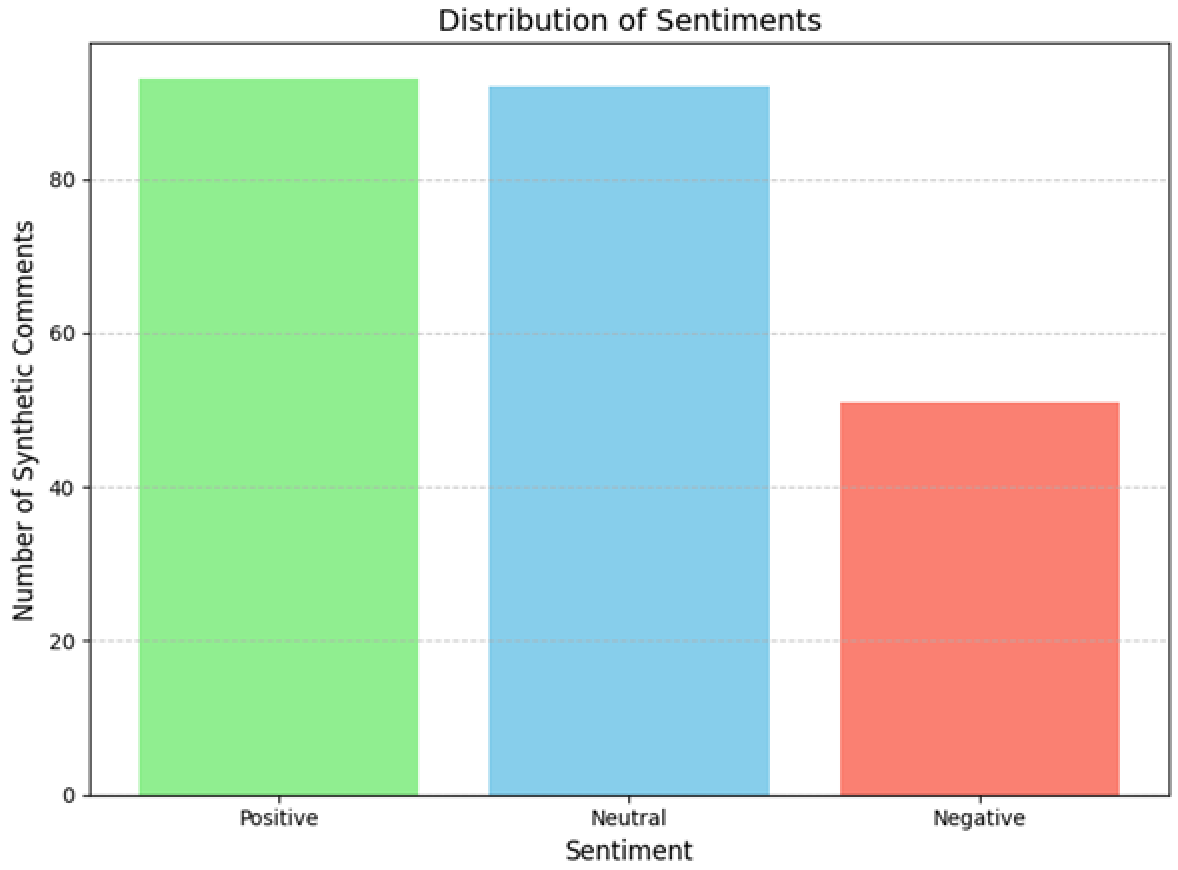
Experiments - Labels Distribution



Heavily skewed toward **higher ratings**!

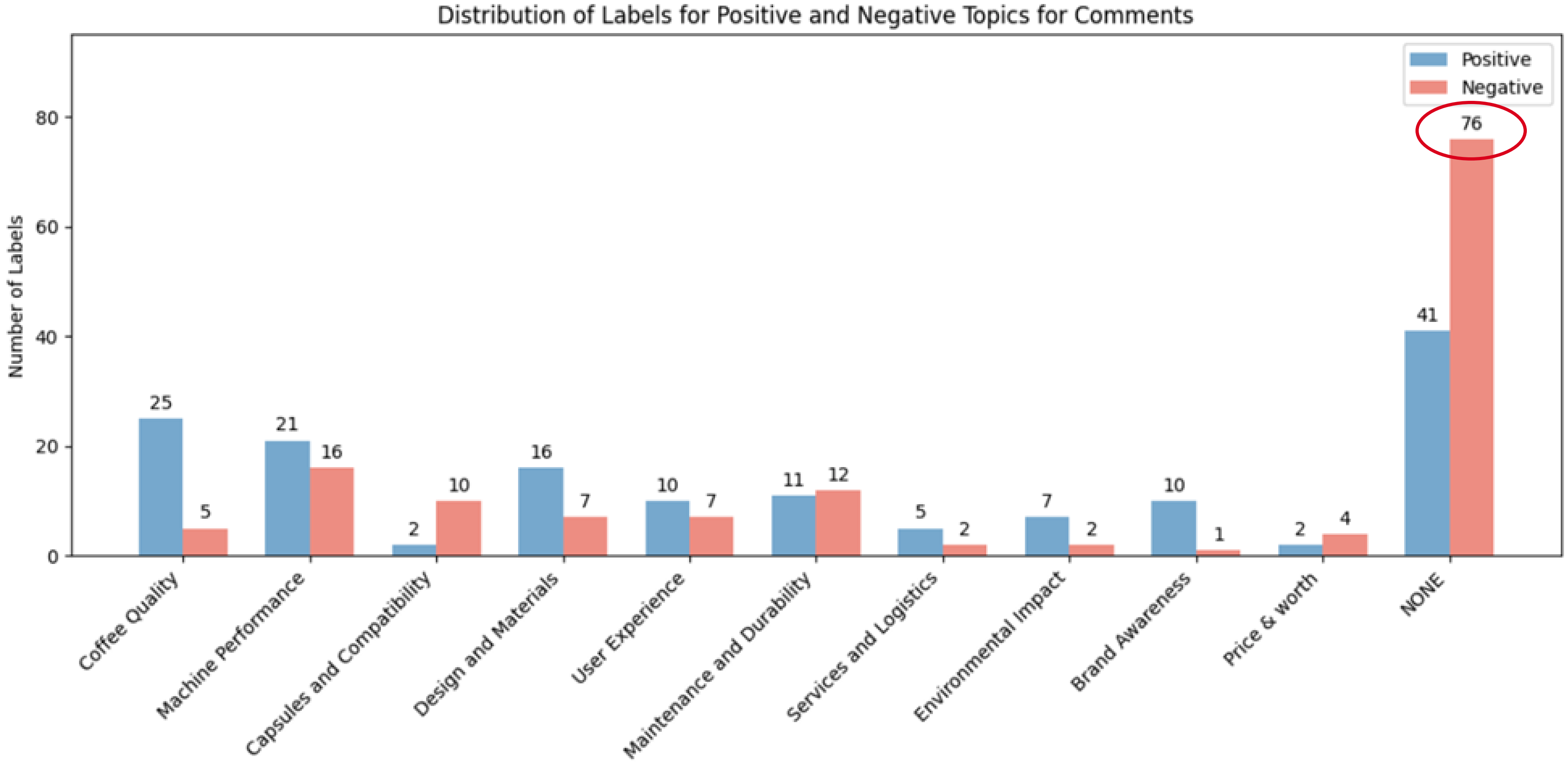
Experiments - Labels Distribution



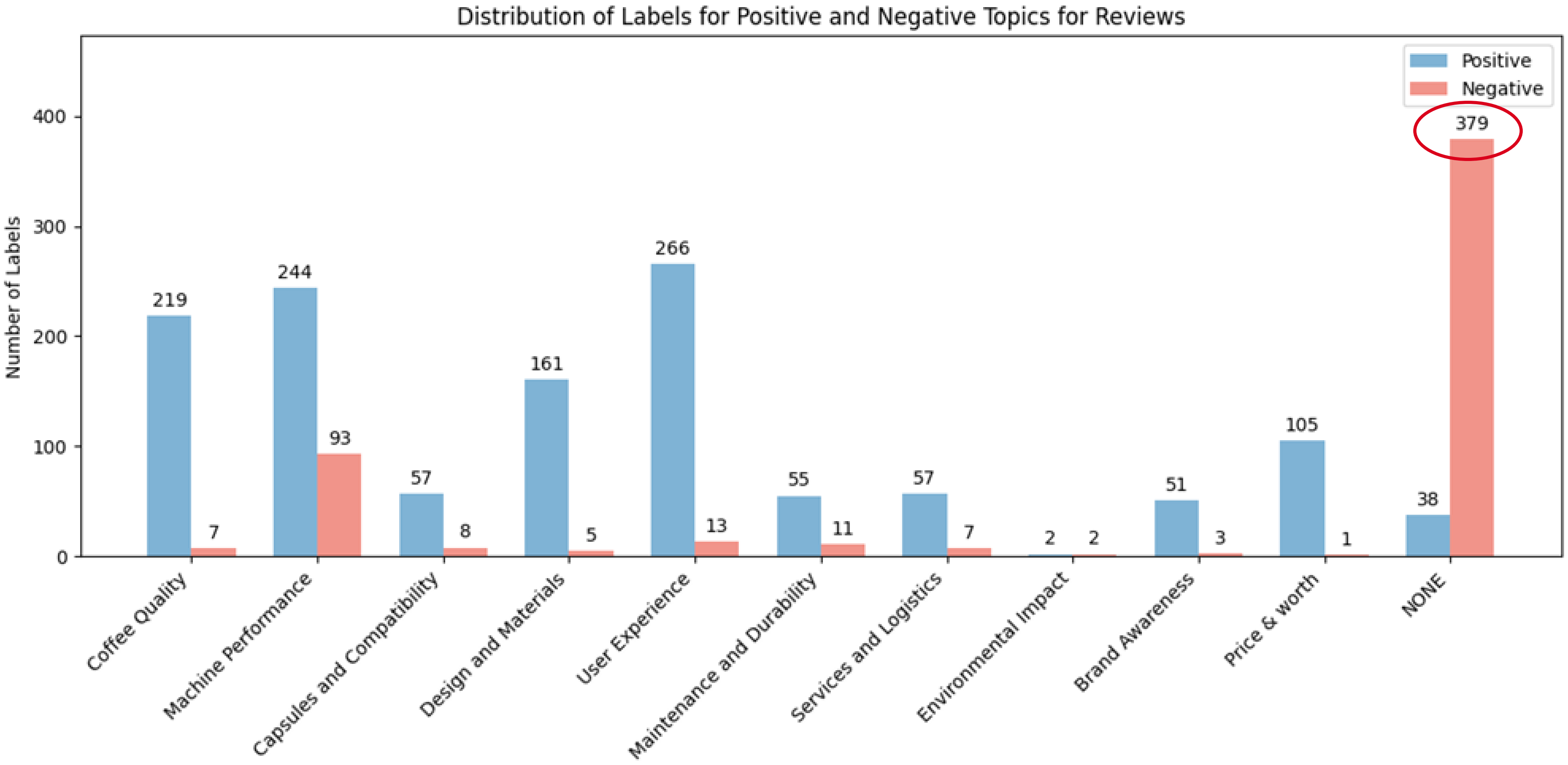


Much more balanced!

Experiments - Labels Distribution



Experiments - Labels Distribution



Experiments - Tasks

1. Product Relevant-Comment Classification

1. Multi-topic Sentiment Classification
- Positive Topics
- Negative Topics

3. Sentiment Classification
- Positive
- Negative
- Neutral

3. Star Rating Prediction:

⭐⭐⭐⭐⭐

Experiments - Product Relevant-Comment Classification

- **Data:** 236 synthetic comments
- **Batch-size:** 15 comments

Model	Acc.	Pre.	Rec.	F1	Time (s)
Gemma 0-shot	0.758	0.738	0.896	0.809	3.40
Gemma Few-shot	0.814	0.783	0.933	0.851	3.63
Llama 0-shot	0.763	0.737	0.911	0.815	**2.77**
Llama Few-shot	**0.826**	**0.794**	**0.941**	**0.861**	2.93
Mistral 0-shot	0.758	0.747	0.874	0.805	3.25
Mistral Few-shot	0.797	0.764	0.933	0.840	3.53

Table 1: Metrics and Execution Times for Models for Product Relevant-Comment Classification

Experiments - Multi-topic Sentiment Classification

- **Data:** 500 labeled reviews + 236 synthetic comments
- **Batch-size:** 5 comment/review

Model	F1 (Neg.)	F1 (Pos.)	Prec. (Neg.)	Prec. (Pos.)	Rec. (Neg.)	Rec. (Pos.)	Time (s)
Gemma 0-shot	0.895	0.732	0.901	0.834	0.900	0.705	12.10
Gemma Few-shot	**0.899**	0.732	**0.905**	**0.844**	**0.904**	0.694	12.82
Llama 0-shot	0.880	**0.745**	0.901	0.795	0.879	0.718	**8.27**
Llama Few-shot	0.878	0.730	0.895	0.811	0.874	0.694	9.55
Mistral 0-shot	0.873	0.738	0.885	0.761	0.873	**0.735**	11.12
Mistral Few-shot	0.880	0.715	**0.905**	0.789	0.873	0.679	20.96

Table 2: Weighted Metrics and Execution Times for Models for Multi-topic Sentiment Classification

Experiments - Sentiment Classification

- **Data:** 236 synthetic comments
- **Batch-size:** 20 (general purpose model) and 50 (specialized model)

Model	Acc.	Prec.	Rec.	F1	Time (s)
Gemma 0-shot	0.869	0.878	0.869	0.866	3.75
Gemma Few-shot	0.852	0.856	0.852	0.849	3.95
Llama 0-shot	0.818	0.818	0.818	0.815	2.76
Llama Few-shot	0.814	0.825	0.814	0.812	2.96
Mistral 0-shot	**0.877**	**0.880**	**0.877**	**0.876**	4.75
Mistral Few-shot	0.860	0.860	0.860	0.859	3.71
DB-sentimet	0.496	0.436	0.496	0.388	$\mathbf{6.52 \times 10^{-3}}$
TR-sentimet	0.682	0.751	0.682	0.680	6.77×10^{-3}
TRL-sentimet	0.831	0.837	0.831	0.829	8.41×10^{-3}

Table 3: Weighted Metrics and Execution Times for Models for Sentiment Classification

Experiments - Star Rating Prediction

- **Data:** 500 labeled reviews
- **Batch-size:** 20 (general purpose model) and 50 (specialized model)

Model	Acc.	Prec.	Rec.	F1	Off-1 Acc.	Time (s)
Gemma 0-shot	0.596	0.763	0.596	0.650	0.908	8.07
Gemma Few-shot	0.592	0.765	0.592	0.647	0.908	8.17
Llama 0-shot	0.682	0.764	0.682	0.713	0.948	5.73
Llama Few-shot	0.682	**0.777**	0.682	0.715	0.942	5.92
Mistral 0-shot	0.498	0.741	0.498	0.567	0.912	6.12
Mistral Few-shot	0.434	0.732	0.434	0.510	0.902	6.45
BBM-rating	**0.692**	0.757	**0.692**	**0.718**	**0.955**	0.158
MLA-rating	0.552	0.706	0.552	0.606	0.845	$\mathbf{10.45 \times 10^{-3}}$

Table 4: Weighted Metrics, Accuracy, Off-by-One Accuracy, and Execution Times for Models for Star Rating Classification

Conclusion

We can conclude by saying that:

- **LLMs** are able to identify **relevant product comments**, with an accuracy of up to **82.6%.**

- **LLMs** are able to extract **product aspects** from reviews and comments and **assign sentiment** to them, with an F1 score of up to **89.9%** for **negative** aspects and **74.5%** for **positive** aspects**.**

- **LLMs** are able to **assign sentiment** to comments and **rating stars** to reviews, with an accuracy of up to **87.7%** and **69.2% (95,5 off-by-one) respectively.**

Conclusion

Politecnico di Torino

In conclusion we chose these models:

- **Gemma 2 few-shot:** for **Product Relevant-Comment Classification** and **Multi-topic Sentiment Classification**.

- **Twitter RoBERTa latest:** for **Sentiment Classification**.

- **BERT base multilingual:** for **Star Rating Prediction**.

LAVAZZA TORINO, ITALIA, 1895

TORINO, ITALIA, 1895

Thank you

Alessio Gioè - Catalano Vincenzo - Tommaso Mazzarini