AI-Powered Automatic Replies in Customer Support

Prashant Anand

mercari

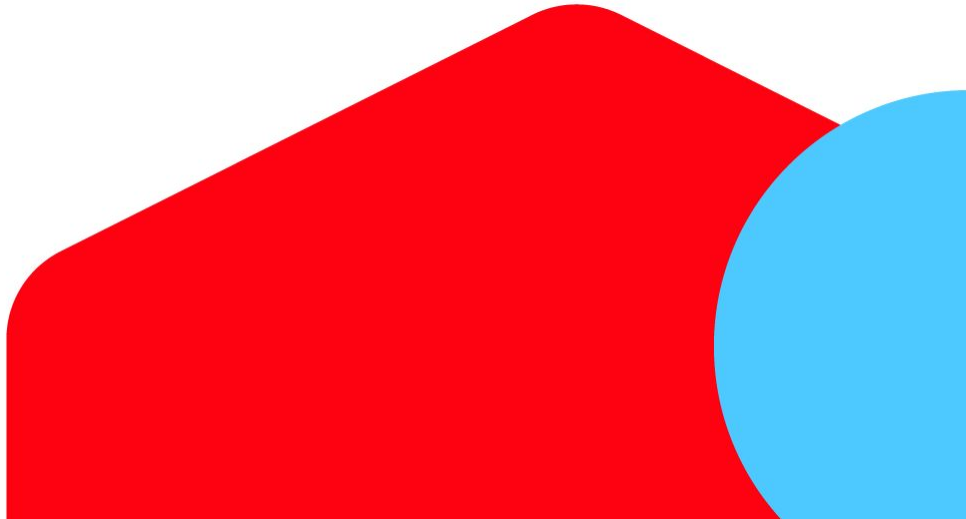


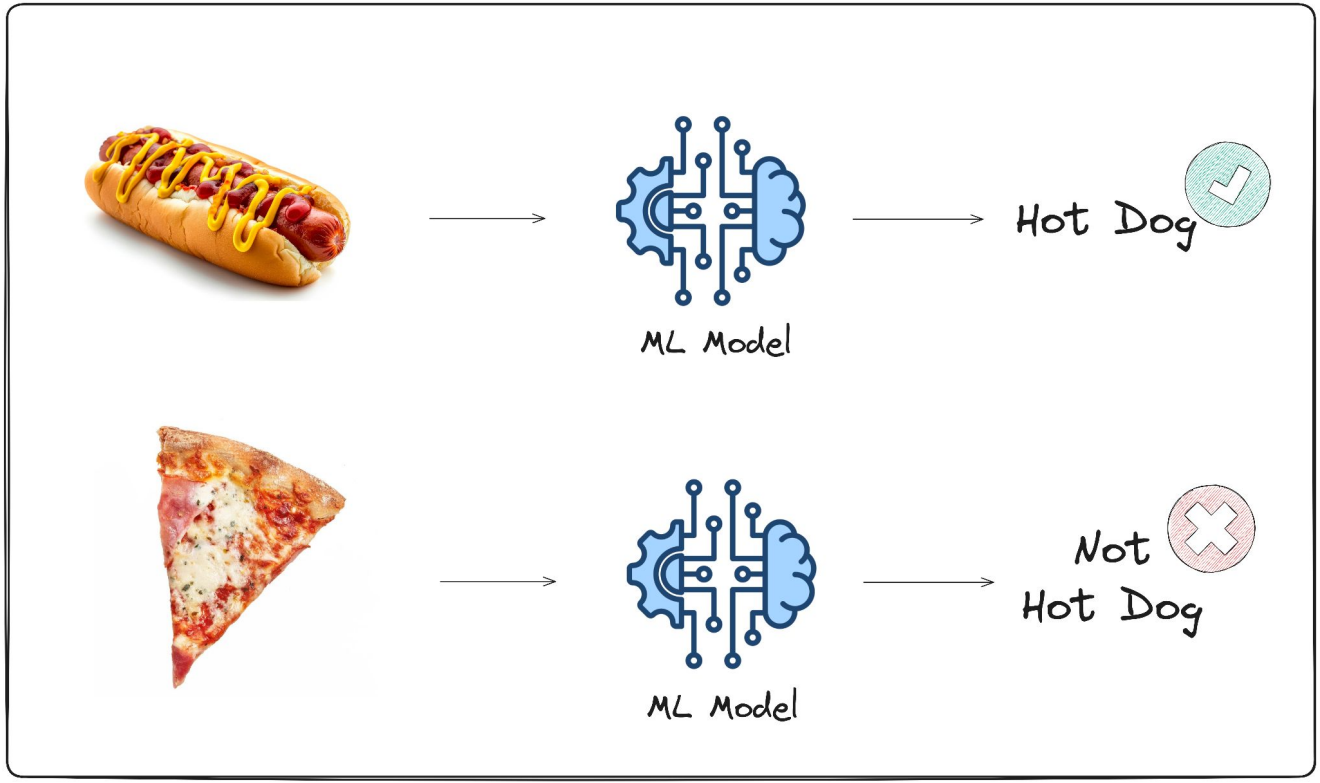
Table of Contents

mercari

01

Threshold Tuning

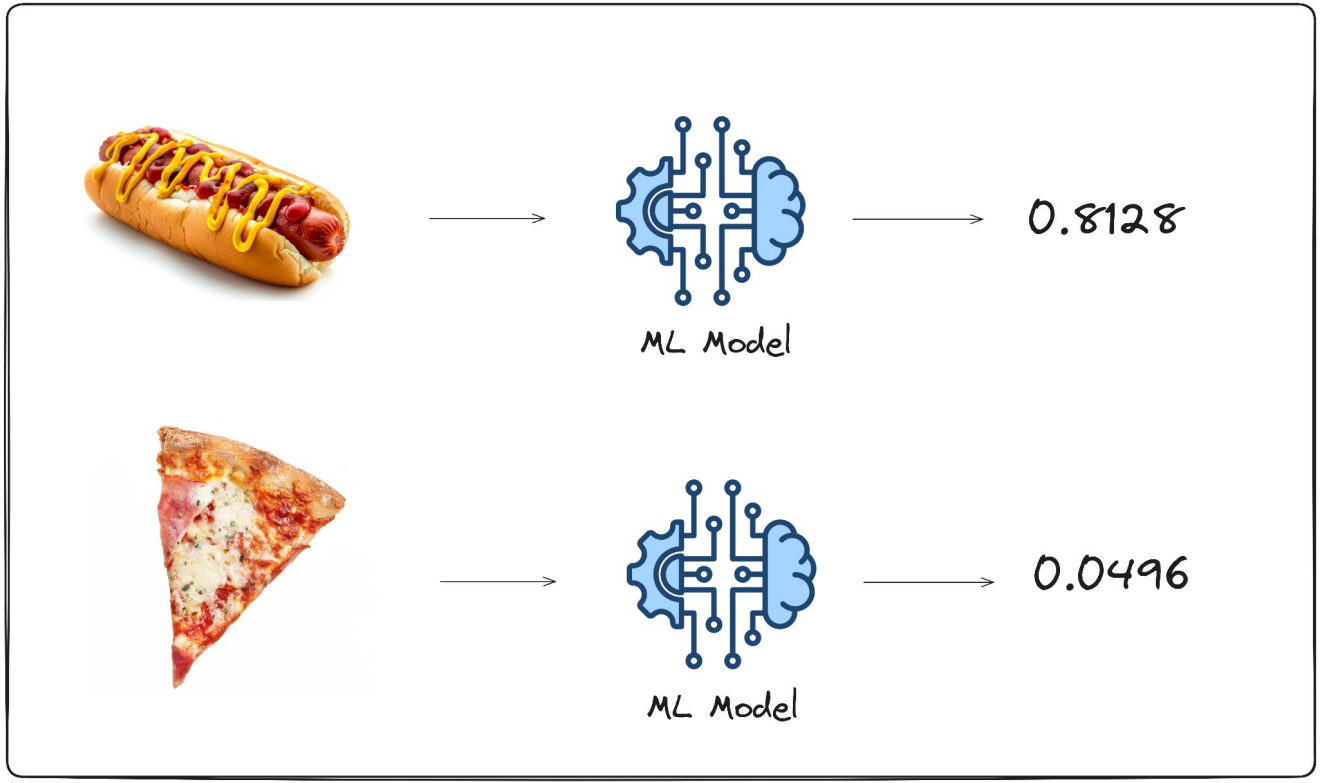
Binary Classifier



mercari

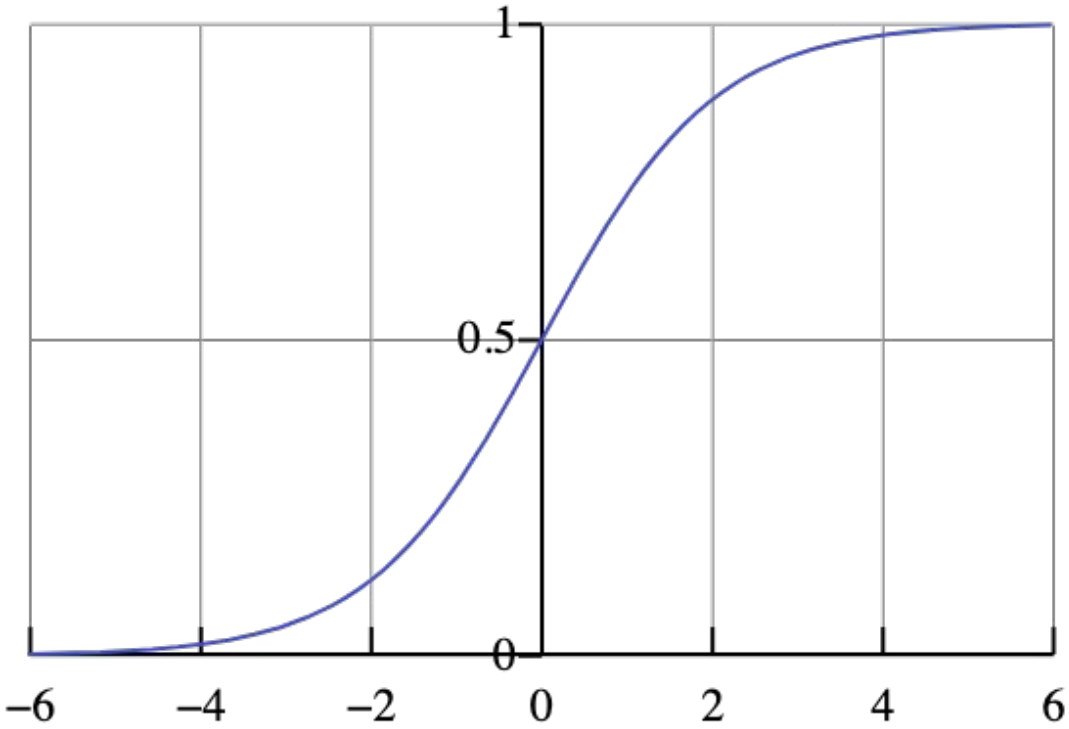
Target Metric for Production Use

Binary Classifier



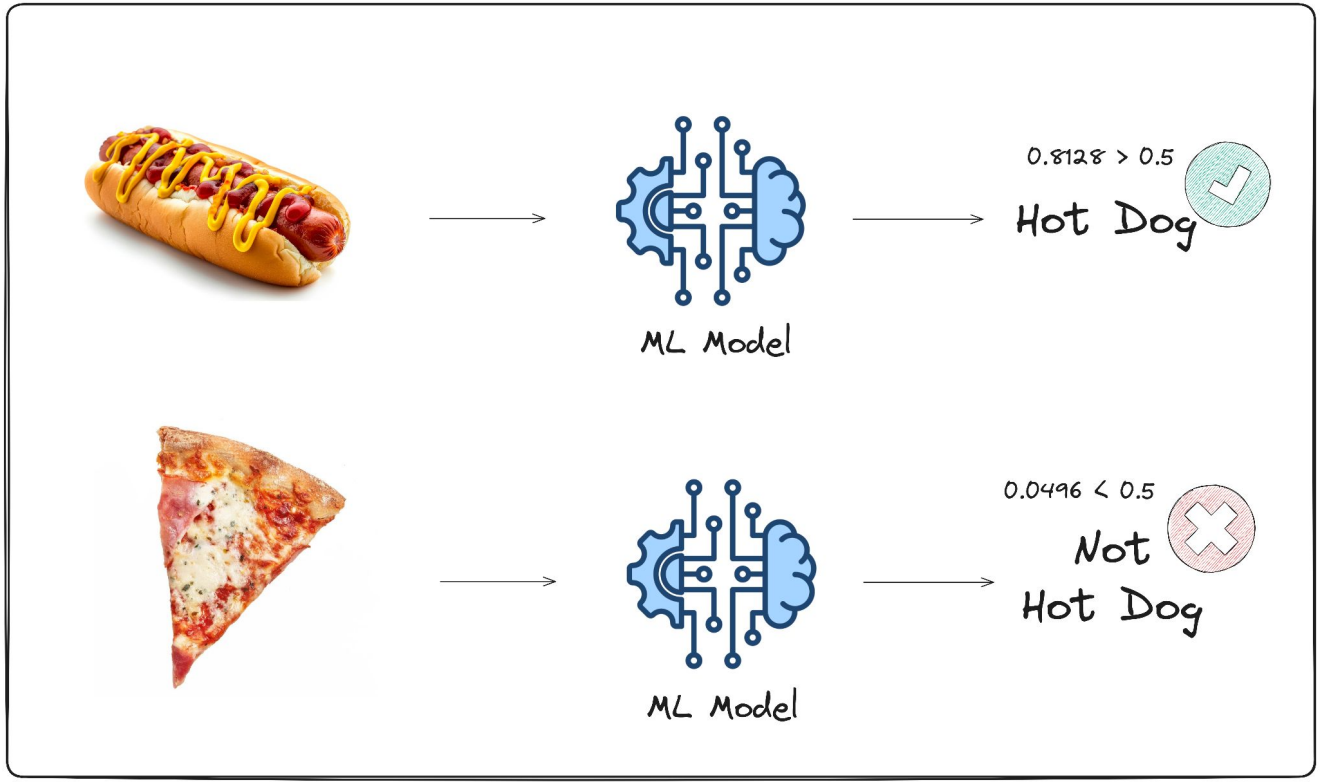
mercari

Sigmoid



mercari

Binary Classifier



Prediction

	prediction_score	prediction	ground_truth
img 1	0.8993	1	1
img 2	0.0791	0	0
img 3	0.4993	0	1
img 4	0.5341	1	0
img 5	0.4182	0	0
	...		

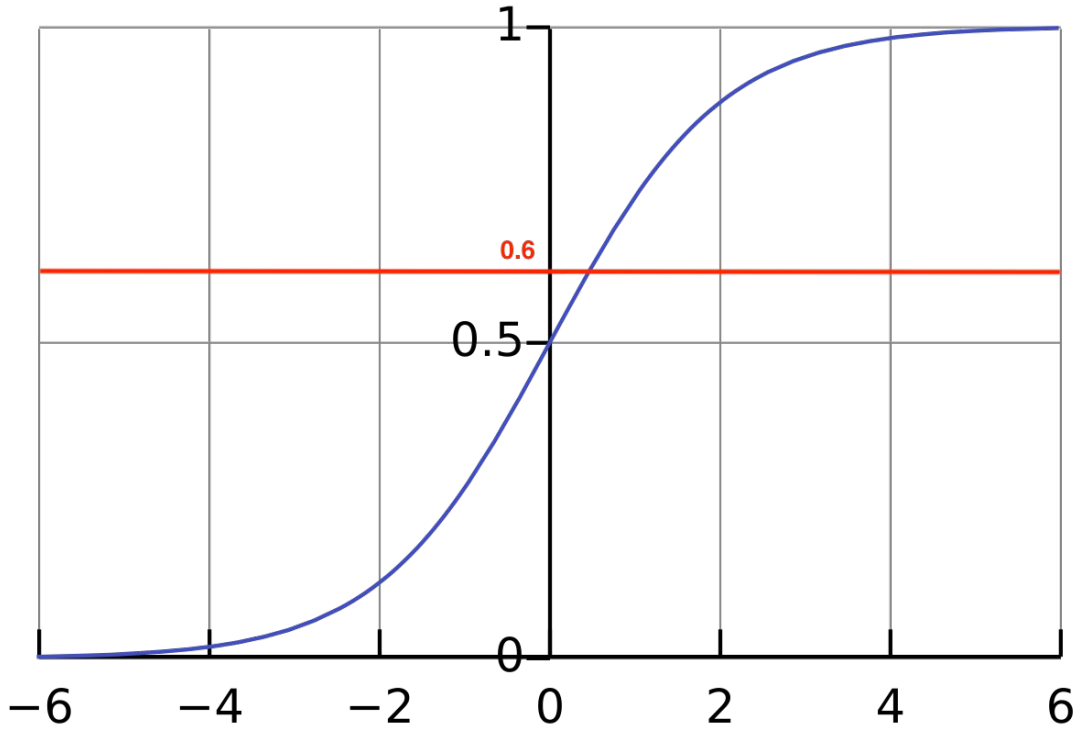
mercari

Classification Metrics

		Prediction: Hot Dog	Prediction: Not Hot Dog
Ground Truth	Hot Dog	540	213
	Not Hot Dog	312	632

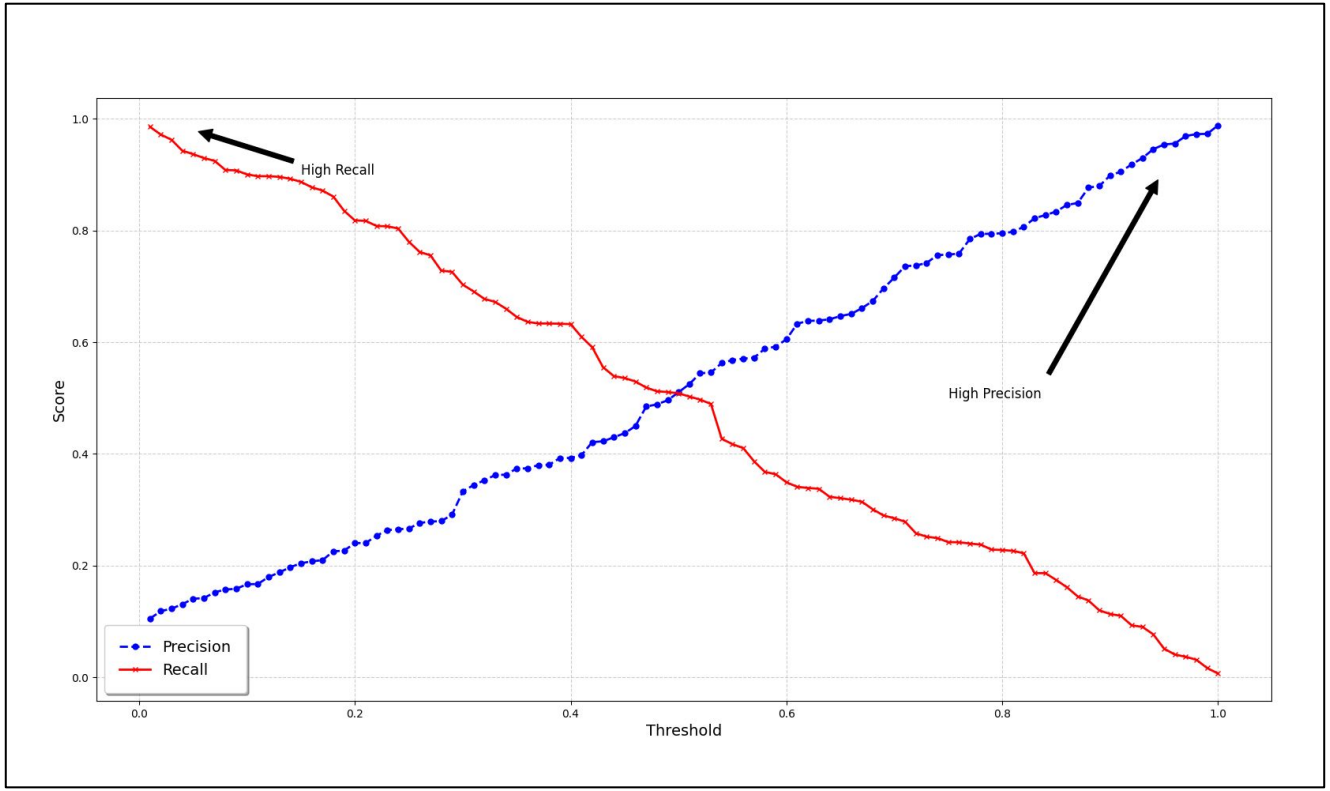
Precision = 540 / (540+312) = 63.38%

Sigmoid



mercari

Precision and Recall vs Threshold



mercari

Precision and Recall vs Threshold

```python
def calculate_precision_recall(y_pred: np.ndarray, y_true: np.ndarray) -> Dict[str, List[float]]:
    thresholds = np.arange(0.01, 1.01, 0.01).tolist()
    precision = []
    recall = []

    for threshold in thresholds:
        y_pred_binarized = (y_pred >= threshold).astype(int)
        tp = np.sum((y_pred_binarized == 1) & (y_true == 1))
        fp = np.sum((y_pred_binarized == 1) & (y_true == 0))
        fn = np.sum((y_pred_binarized == 0) & (y_true == 1))

        if tp + fp > 0:
            precision.append(tp / (tp + fp))
        else:
            precision.append(0.0)

        if tp + fn > 0:
            recall.append(tp / (tp + fn))
        else:
            recall.append(0.0)

    return {"thresholds": thresholds, "precision": precision, "recall": recall}
```

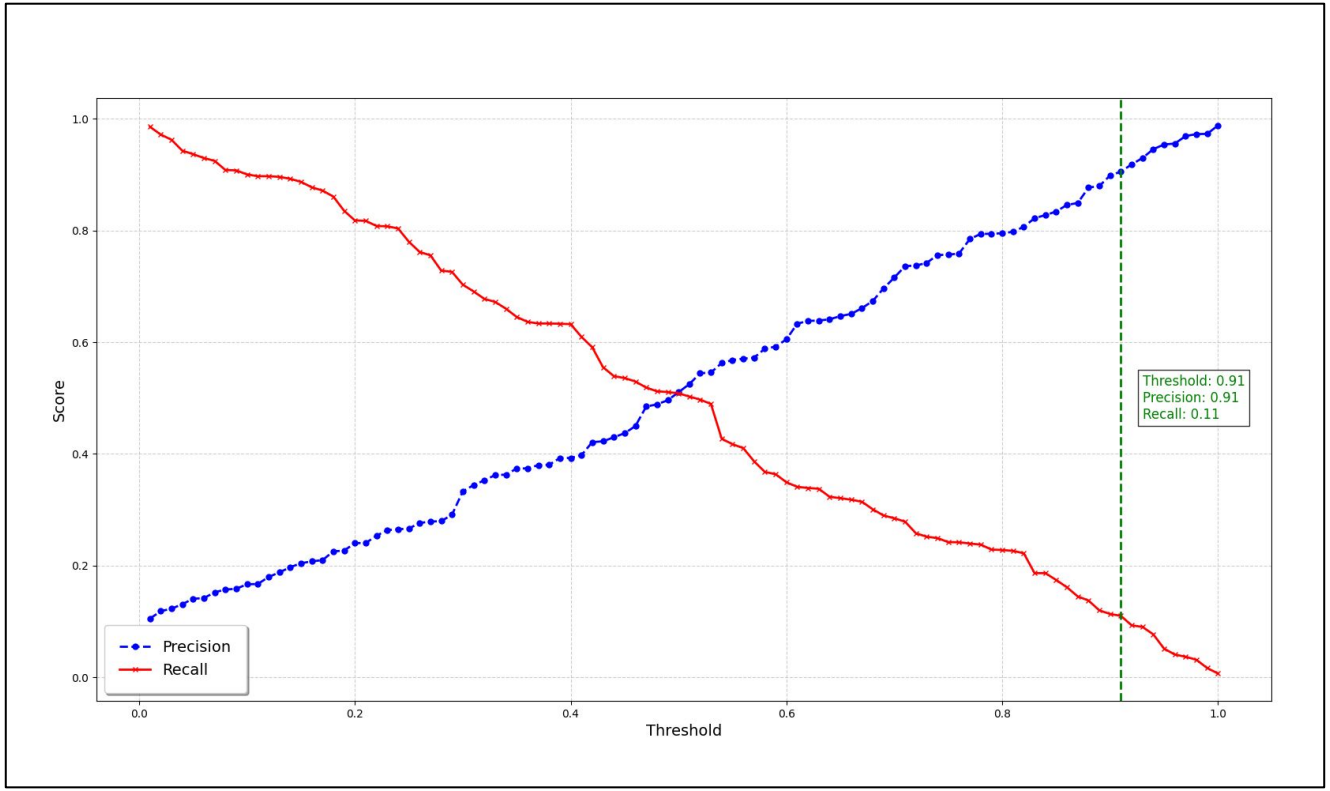
Precision and Recall vs Threshold

```python
def plot_precision_recall(
    thresholds: List[float], precision: List[float], recall: List[float]
) -> None:
    plt.figure(figsize=(12, 8), dpi=100)

    plt.plot(thresholds, precision, label="Precision", color="blue", marker="o", markersize=5)
    plt.plot(thresholds, recall, label="Recall", color="red", marker="x", markersize=5)

    plt.xlabel("Threshold", fontsize=14)
    plt.ylabel("Score", fontsize=14)
    plt.title("Threshold Analysis: Precision and Recall vs. Threshold", fontsize=16)
    plt.legend(
        loc="best",
        fontsize=14,
        frameon=True,
        fancybox=True,
        framealpha=1,
        shadow=True,
        borderpad=1,
    )
    plt.grid(True, linestyle="--", alpha=0.6)
    plt.show()
```

Threshold Tuning



mercari

AI-Powered Automatic Replies in Customer Support

Precision-Focused Approach at Mercari

mercari

Prashant Anand

ML Engineer at Mercari

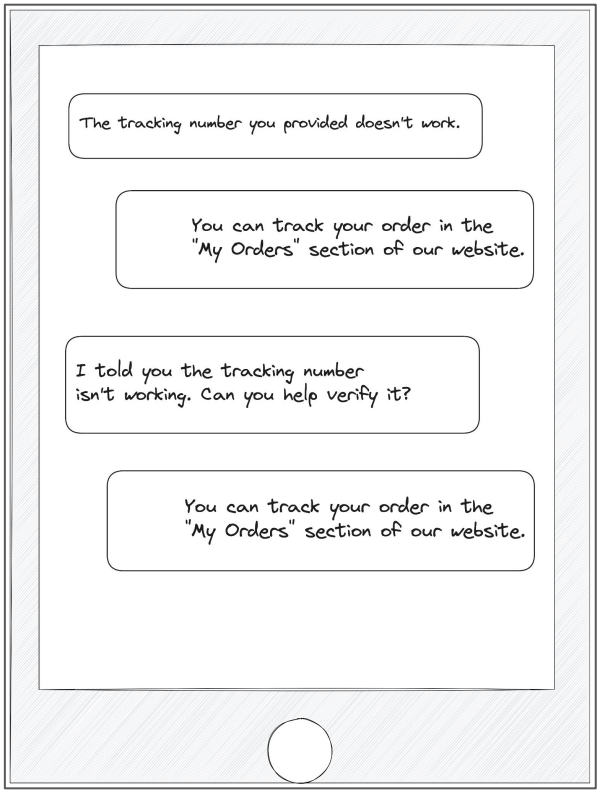
@primaprashant

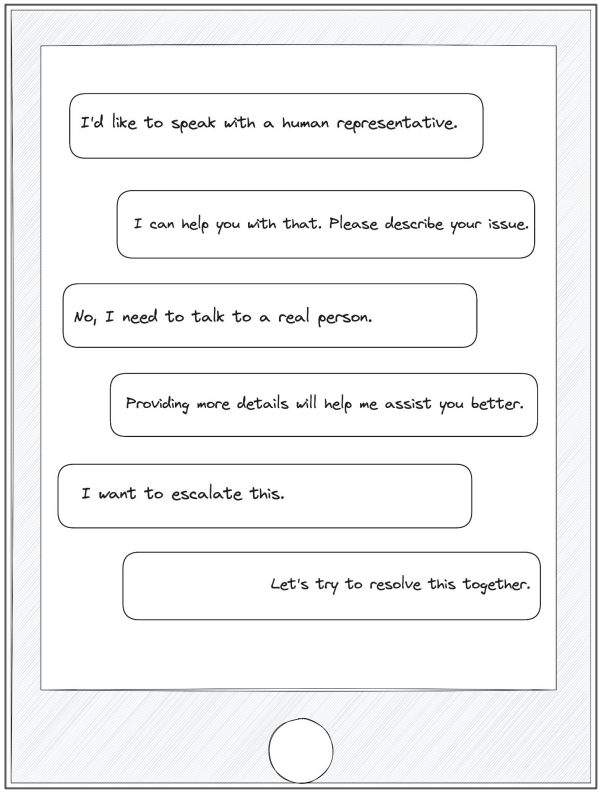
mercari

02

Why use AI?

AI Chatbots


The tracking number you provided doesn't work.
You can track your order in the "My Orders" section of our website.
I told you the tracking number isn't working. Can you help verify it?
You can track your order in the "My Orders" section of our website.

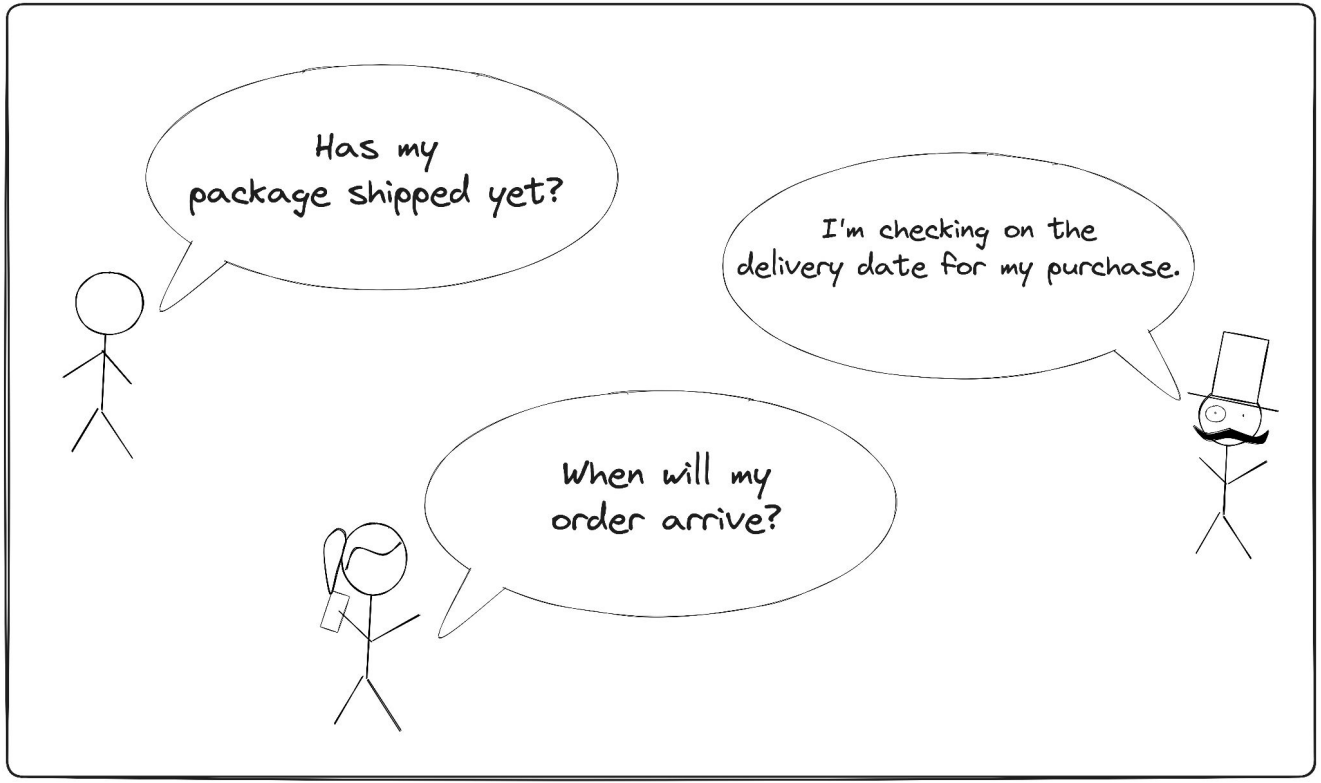

I'd like to speak with a human representative.
I can help you with that. Please describe your issue.
No, I need to talk to a real person.
Providing more details will help me assist you better.
I want to escalate this.
Let's try to resolve this together.

Benefits to Businesses

- Lower operational costs
- Resource reallocation
- Improved scalability

Sample Inquiries

Sample Inquiries



Benefits to Users

- Quicker resolution
- 24/7 availability

Considerations for Our AI Response

- Highly precise
- Easy escalation to human agents

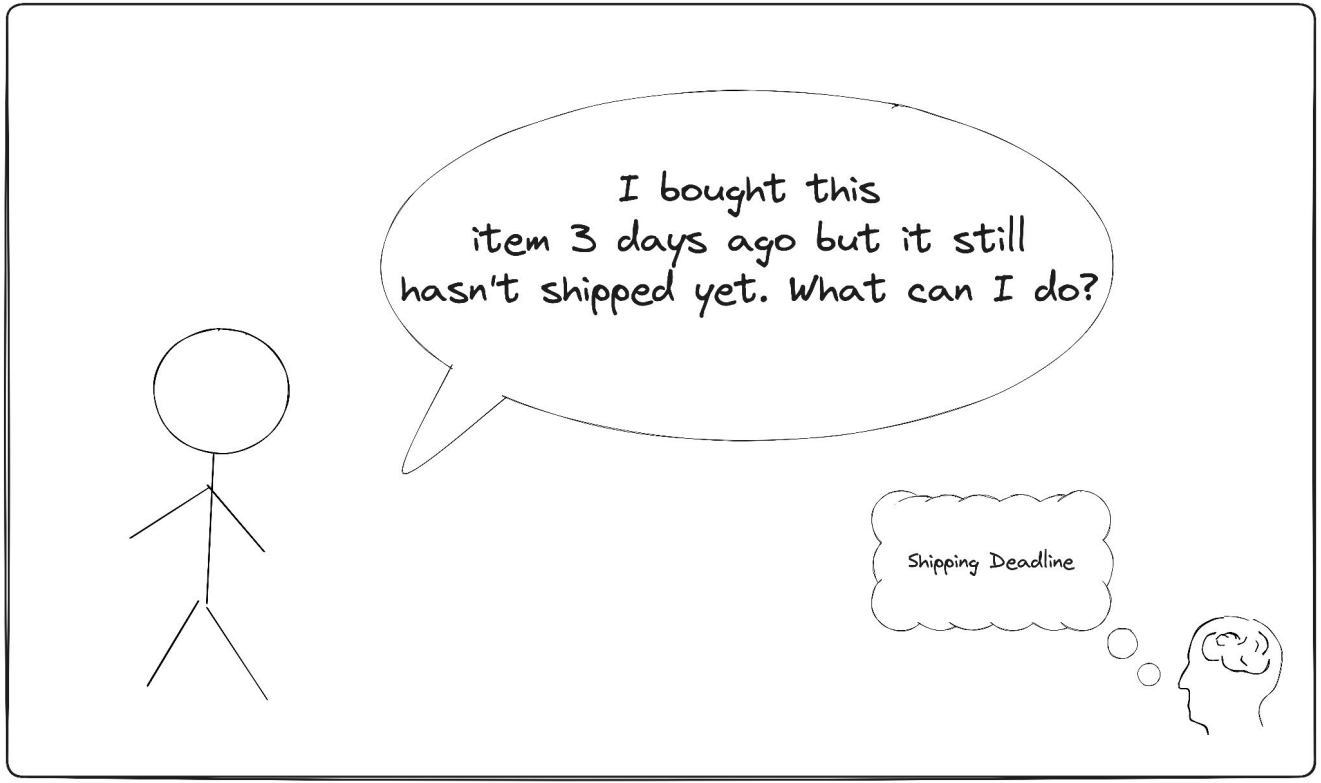
03

Designing a Precise System

Identify Patterns

- Analyze raw data
- Look at different segments
- Most used response templates
- Talk with domain experts

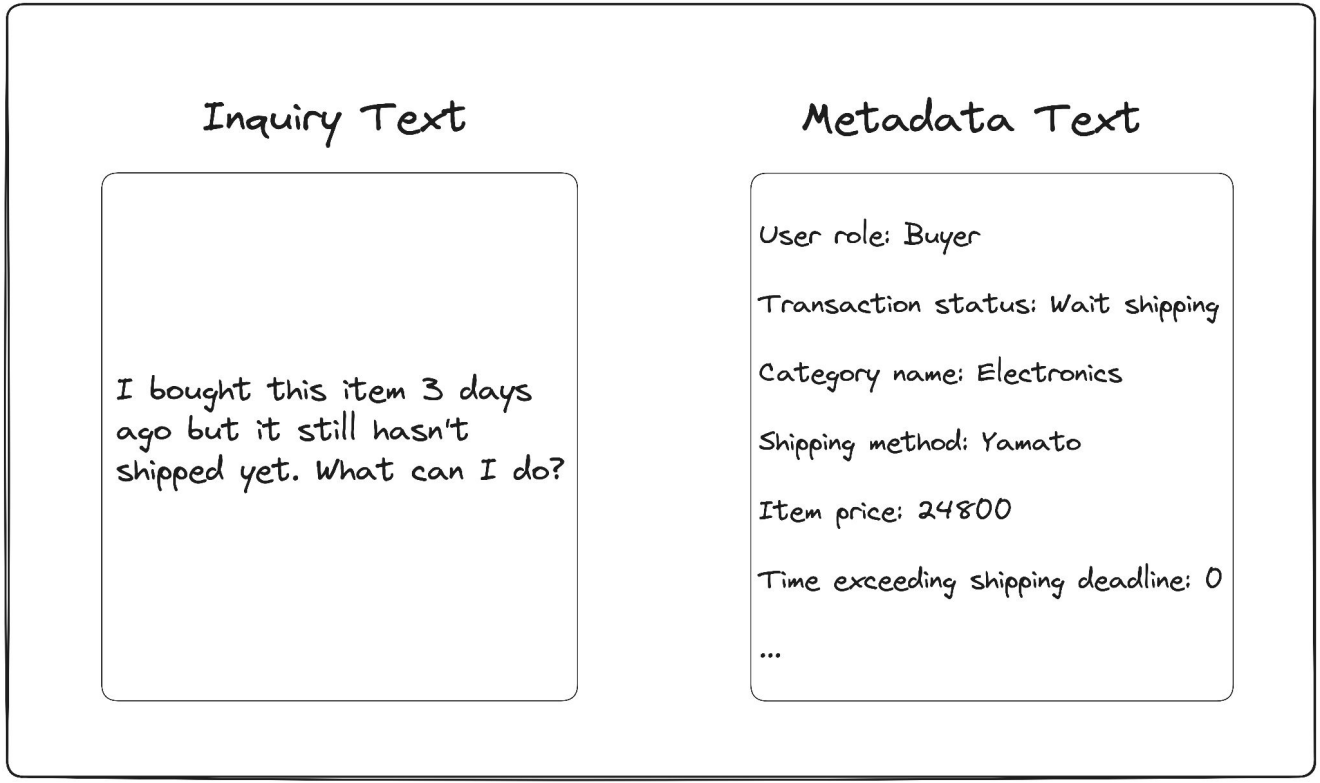
Utilize Metadata



Metadata

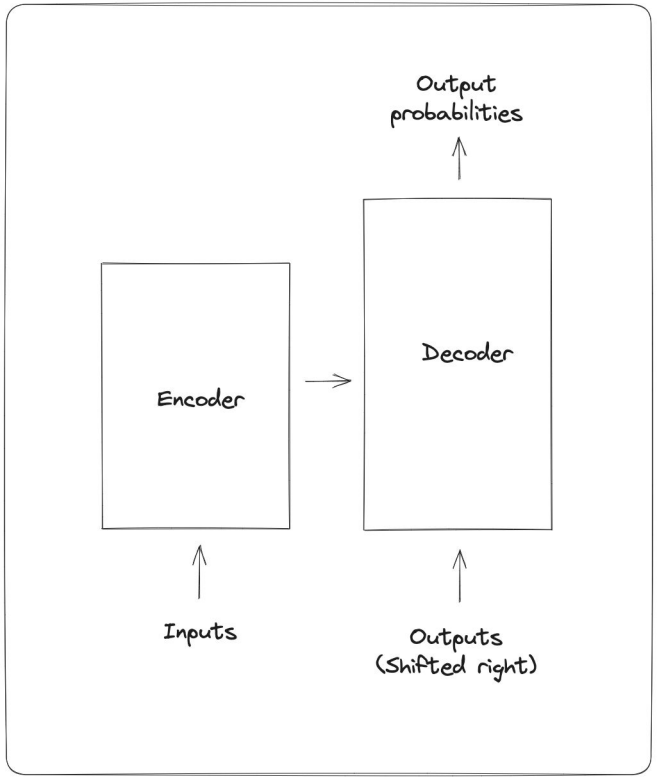
- User role (buyer or seller)
- Transaction status
- Category of item
- Shipping method
- Item price
- Time exceeding shipping deadline
- High value item or not
- Large item or not
- Restrictions on seller
- eKYC audit state

Dataset

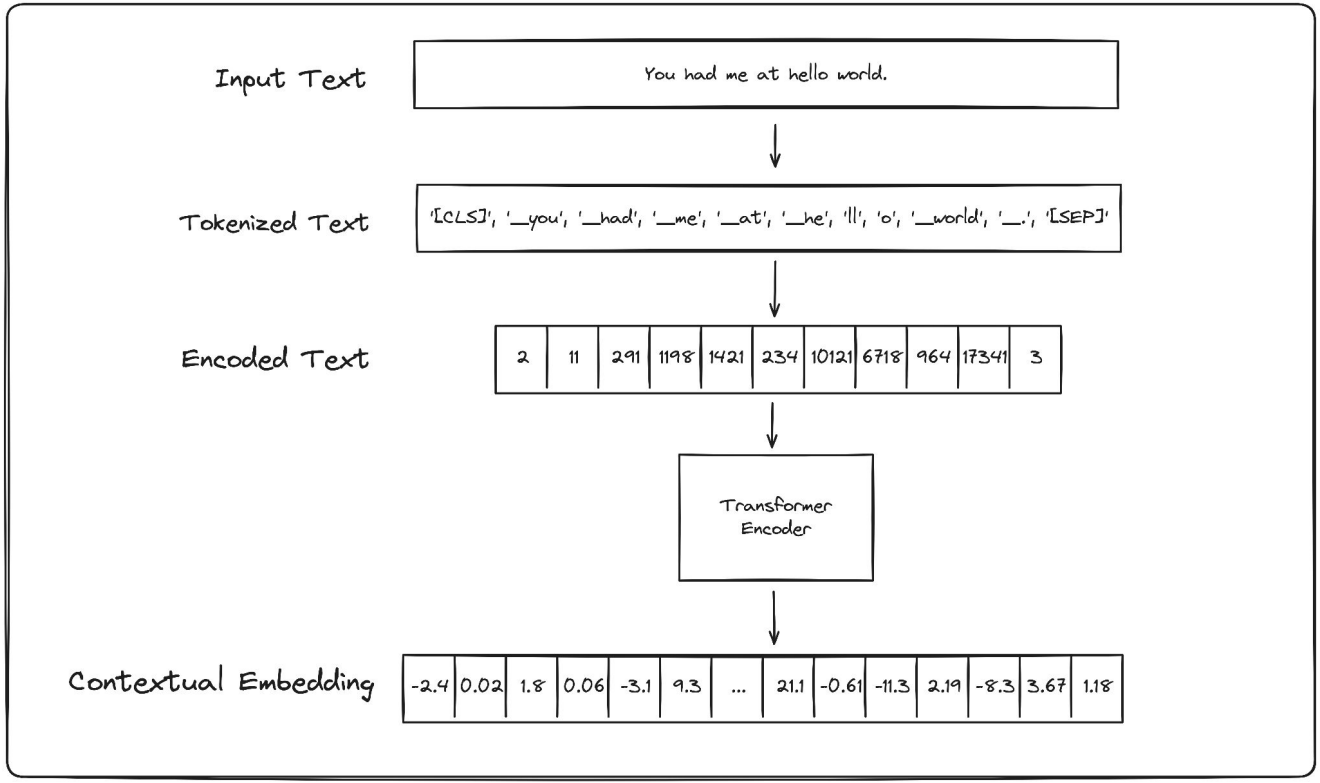

Inquiry Text
I bought this item 3 days ago but it still hasn't shipped yet. What can I do?
Metadata Text
User role: Buyer
Transaction status: Wait shipping
Category name: Electronics
Shipping method: Yamato
Item price: 24800
Time exceeding shipping deadline: 0
...

Transformers Overview

- Introduced in 2017
- State-of-the-art architecture for NLP tasks
- Pre-trained on massive amount of data
- Contextual understanding



Transformer



Pre-Trained Models

Encoder models

General purpose

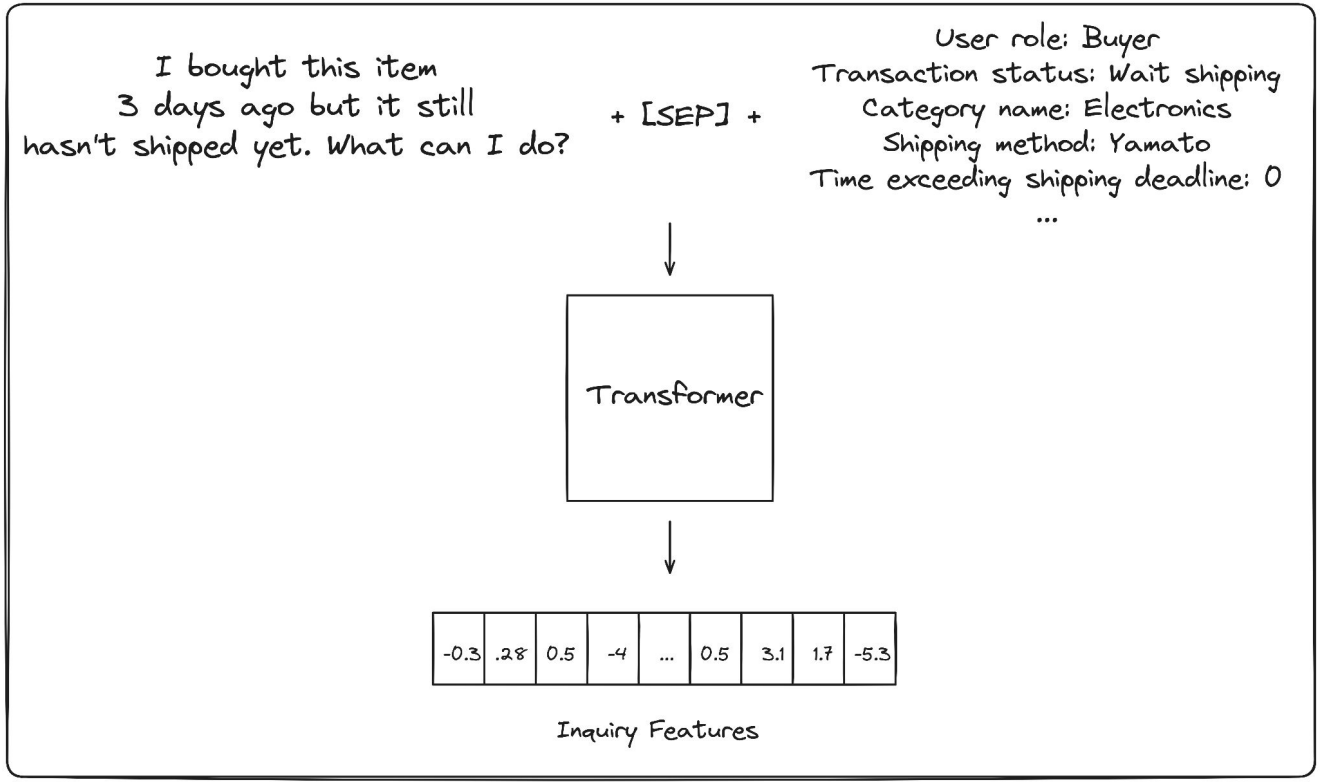
	Architecture	Training Data	Developer	License	HuggingFace? [9]
KyotoUniBERT	BERT (base, large)	Japanese Wikipedia (18M articles)	Kyoto University Language Media Processing Lab	Apache 2.0	△
TohokuUniversityBERT	BERT (base, large)	base (v1): Japanese Wikipedia (17M articles / 2.6GB) base (v2) & large: Japanese Wikipedia 4.0GB base (v3) & large (v2): Japanese Wikipedia (4.9GB), Japanese CC-100 (74.3GB)	Tohoku University NLP Group	base (v1, v2) & large: CC BY-SA 3.0 base (v3) & large (v2): Apache 2.0	○ (base (v1), base (v1, char-level), base (v2), base (v2, char-level), large, large (char-level), base (v3), base (v3, char-level), large (v2), large (v2, char-level))
NICT BERT	BERT (base)	Japanese Wikipedia	NICT	CC BY 4.0	△
Laboro BERT	BERT (base, large)	Japanese Web Corpus (News and blogs, etc) (12GB)	Laboro.AI	CC BY-NC 4.0	×
colorfulscoop BERT	BERT (base)	Japanese Wikipedia	Colorful Scoop	CC BY-SA 3.0	○
UniversityOfTokyoBERT	BERT (small)	Japanese Wikipedia (2.9GB)	University of Tokyo Izumi Lab	CC BY-SA 4.0	○

llm-jp/awesome-japanese-llm

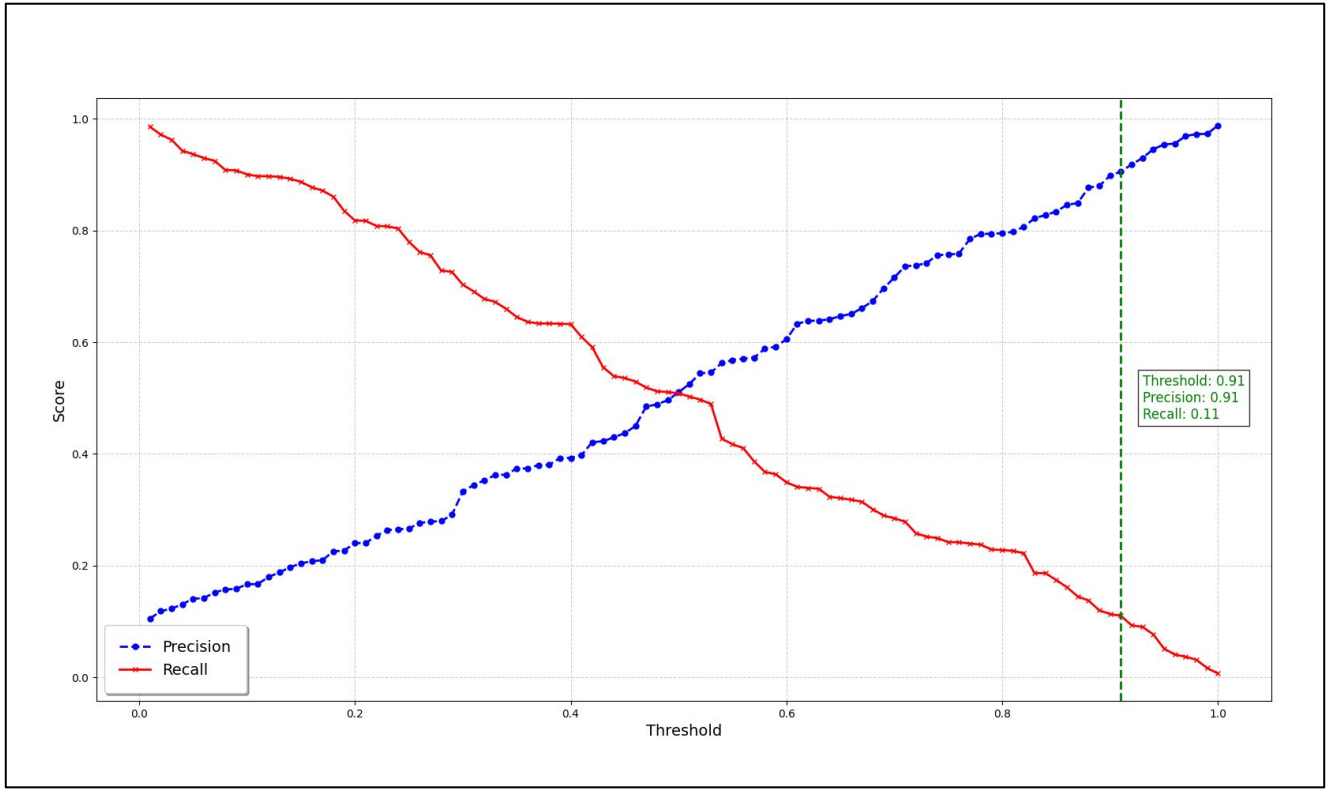
Pre-Trained Models

Model Name	Release Year	Link
Tohoku BERT base Japanese V3	2023	https://huggingface.co/cl-tohoku/bert-base-japanese-v3
LINE DistilBERT Japanese	2023	https://github.com/line/LINE-DistilBERT-Japanese
Rinna Japanese RoBERTa Base	2021	https://huggingface.co/rinna/japanese-roberta-base
Bandai Namco DistilBERT-base-jp	2020	https://github.com/BandaiNamcoResearchInc/DistilBERT-base-jp
FacebookAI/xlm-roberta-base	2020	https://huggingface.co/FacebookAI/xlm-roberta-base

Develop ML Model



Threshold Tuning



mercari

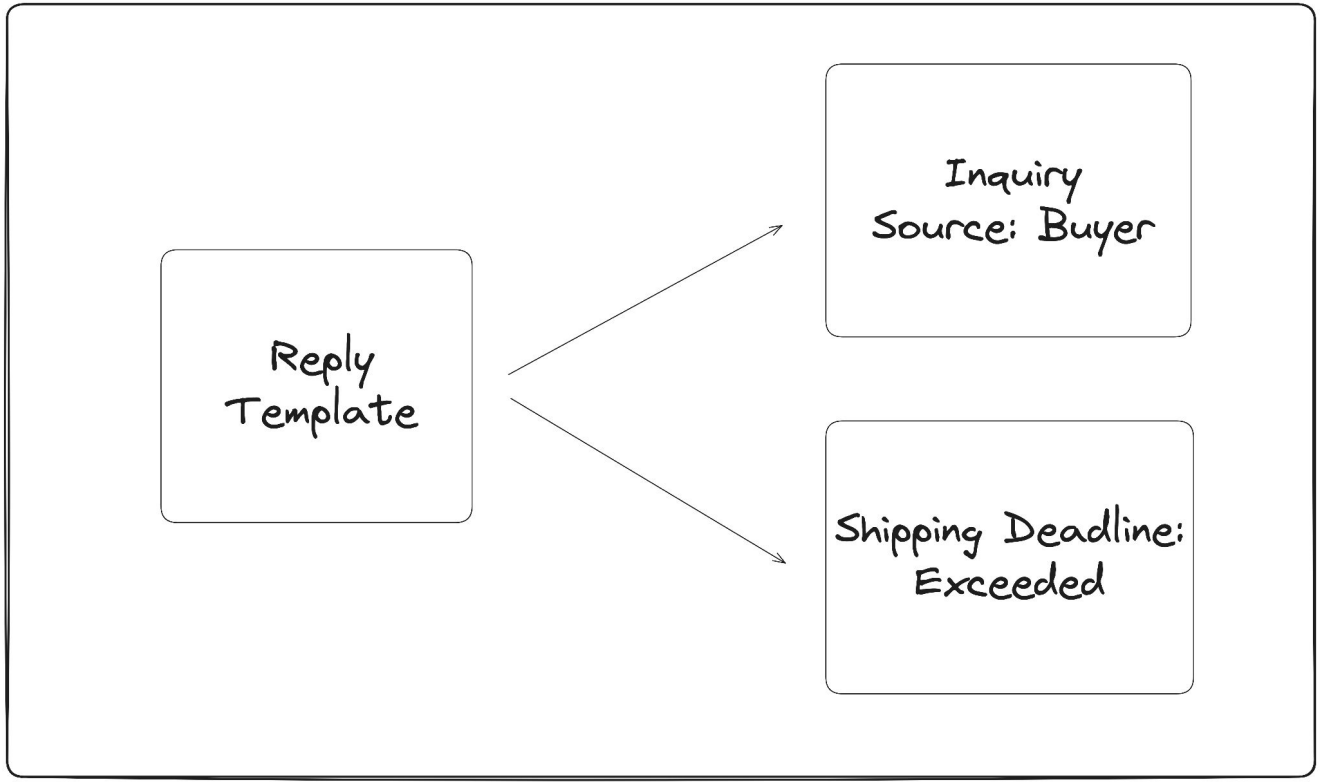
Sample Reply Template

Thank you for your inquiry. We have notified the seller to either ship the item within 24 hours or inform you of the shipping progress via transaction message.

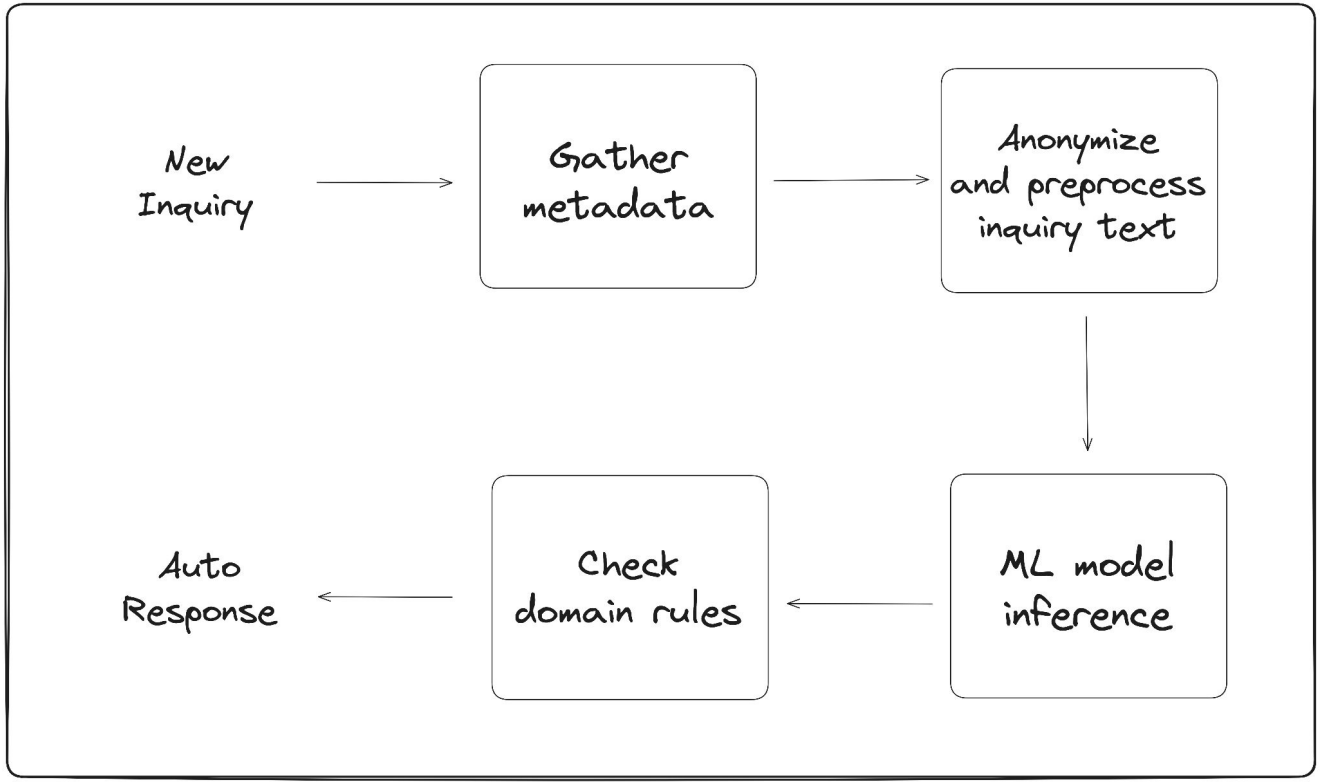
While we understand that you have been waiting for a long time, we kindly ask you to wait for the seller's response for 24 hours from this notification.

Additionally, if there is no response from the seller after the aforementioned time, you can proceed to cancel the transaction by following this guide: https://help.jp.mercari.com/guide/articles/281/

Rules Using Domain Knowledge



Inference Flow



04

Business Impact

Metrics

- Goal: Average number of manual replies per case
- Guardrails: Average resolution time per user, Customer satisfaction

Results

- Average number of manual replies per case reduced by 5.5% and 1.3%.

- No negative impact on customer satisfaction. Average resolution time per user improved.

Results

- Total time required for customer support operation reduced by 1000 hours and 800 hours per months.

- Automatic replies are about 5% of all the replies sent per month.

What We Learnt!

- Designing high precision system for text classification
- Importance of utilizing metadata
- Threshold tuning for production use of ML models
- Using rules based on domain knowledge along with the ML model