Этап 2

Разметка по ключевым словам

Class 0

В вопросе клиента:

"когда", "вернуть", "сдать",

"возврат", "поставк", "акция" и тд.

В ответе селлера:

wildberries", "ситуация",

"обратитесь"

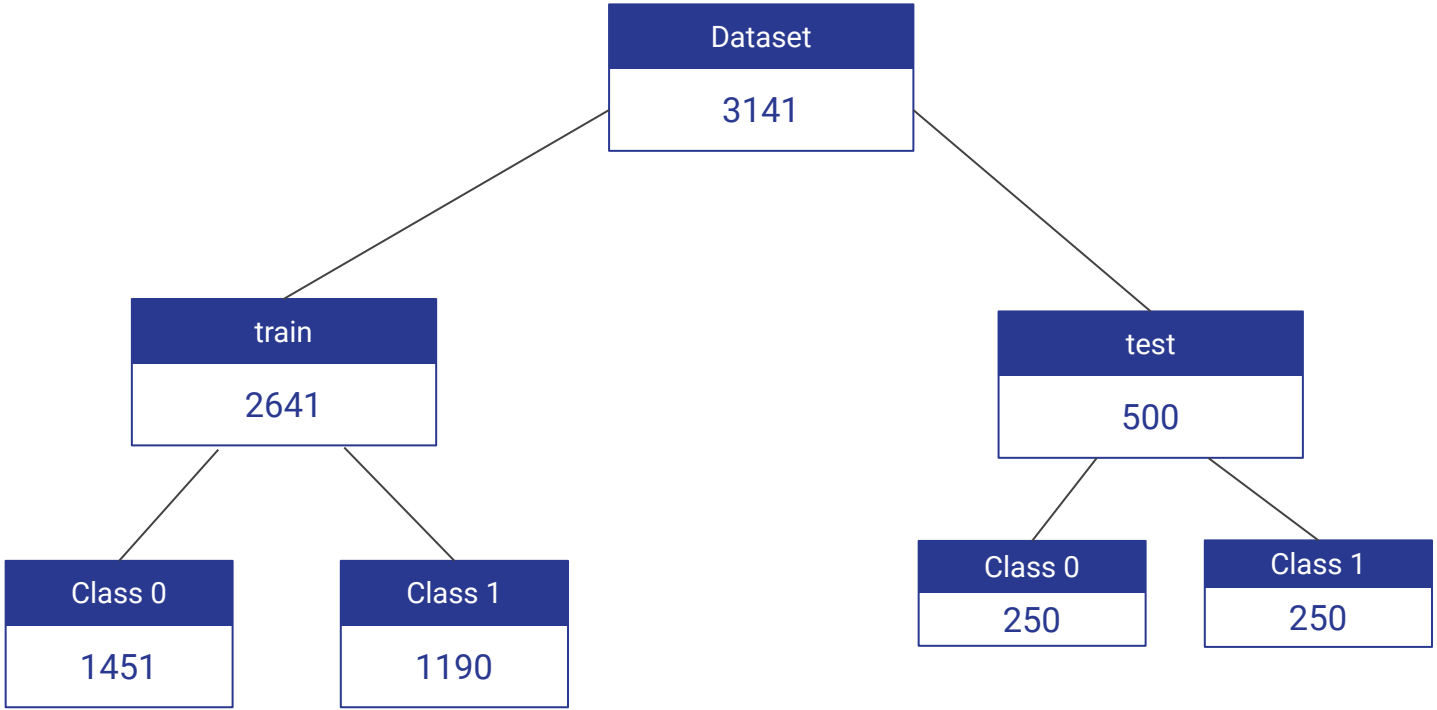
Class 1

В ответе селлера:

"карточка", "размер","комплект",

"фото", "да", "нет"

Итоговый датасет после разбиения по классам и удаления дубликатов вопросов



Обучающиеся данные

Извлечение описания из строки:

```
data['desc'] = data.apply(lambda x: x.Description.split('}')[-1].strip(), axis=1)
```

Итоговый столбец для обучения представляет собой конкатенацию названия товара и извлеченного описания.

cointegrated/LaBSE-en-ru

Precision: 0.9189

Recall: 0.952

```
BATCH_SIZE = 16
MAX_LENGTH = 512

train_data_loader = create_data_loader(train, tokenizer, MAX_LENGTH, BATCH_SIZE)
val_data_loader = create_data_loader(val, tokenizer, MAX_LENGTH, BATCH_SIZE)

training_args = TrainingArguments(
        output_dir='./results',
        num_train_epochs=4,
        per_device_train_batch_size=BATCH_SIZE,
        per_device_eval_batch_size=BATCH_SIZE,
        warmup_steps=500,
        weight_decay=0.01,
        logging_dir='./logs',
        logging_steps=10,
        eval_strategy='steps'
    )
```

intfloat/multilingual-e5-base

Precision: 0.926
Recall: 0.956

```
BATCH_SIZE = 16
MAX_LENGTH = 512

train_data_loader = create_data_loader(train, tokenizer, MAX_LENGTH, BATCH_SIZE)
val_data_loader = create_data_loader(val, tokenizer, MAX_LENGTH, BATCH_SIZE)

training_args = TrainingArguments(
        output_dir='./results',
        num_train_epochs=4,
        per_device_train_batch_size=BATCH_SIZE,
        per_device_eval_batch_size=BATCH_SIZE,
        warmup_steps=500,
        weight_decay=0.01,
        logging_dir='./logs',
        logging_steps=10,
        eval_strategy='steps'
    )
```

unsloth/Meta-Llama-3.1-8B-bnb-4bit

Precision: 0.876
Recall: 0.908

```
instruction = """
You are a moderator of questions on the marketplace.
You receive a customer's question and its description as input.
Your task is to evaluate whether the question is related to the characteristics
of the product,its capabilities, configuration, etc. or not.
If the input data contains a question about the delivery date,
delivery delay,
product defect, promotions, discounts, product prices, refunds,
the ability to buy, etc.,
or does not contain a question about the product at all, then
you need to output 0 .
If the question concerns its characteristics, size chart, material,
 capabilities and configuration of the product, etc.,
then you need to output 1.
The answer must contain one number - 0 or 1.
"""
```