Промежуточный отчет

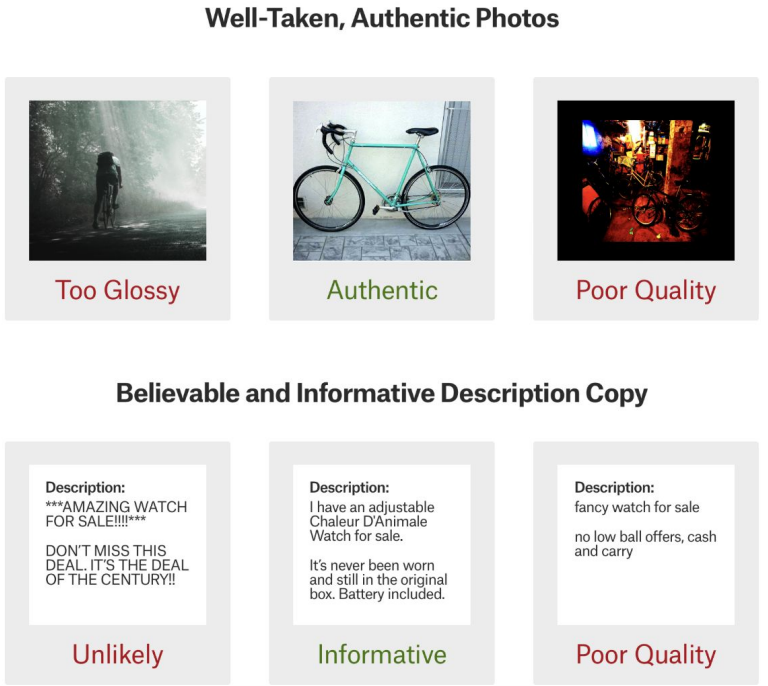
Предмет: Мультимодальные модели: Архитектуры, Обучение и Применение

Коряковская Дарья Олеговна

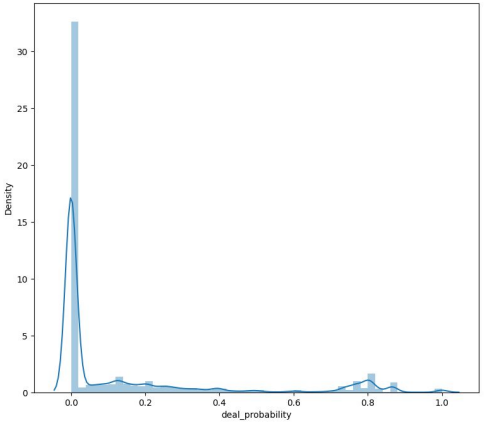
Соревнование

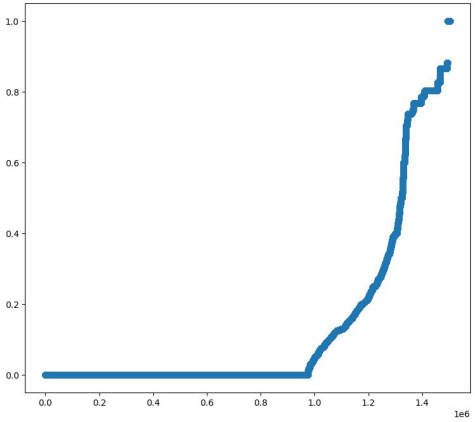
Цель - предсказание вероятности продажи товара на платформе Avito на основе данных предложения (заголовок, описание, изображение, цена, город и тд).

Метрика $$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

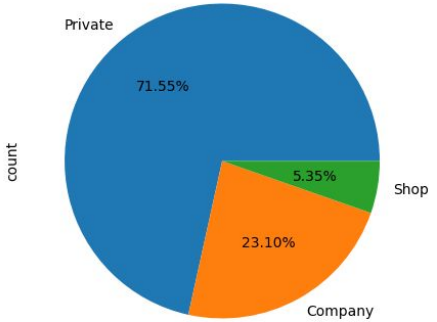


EDA



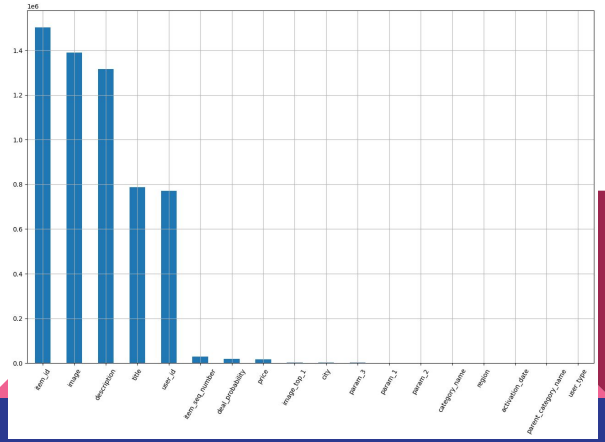


Распределение целевой переменной



Тип пользователей

Количество уникальных значений



Полный EDA в репозитории

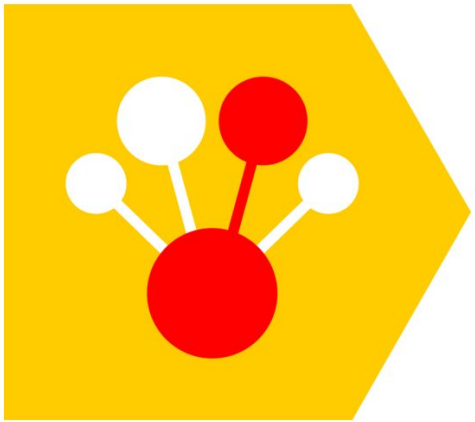
Baseline - Catboost

В качестве baseline я выбрала catboost.

Использованы табличные данные (в том числе категориальные) и текстовые.

Произведен перебор параметров.

Результат: RMSE = 0.33028



В качестве первой части работы я использовала 2 модальности: текстовые и табличные данные.

В качестве экстрактора признаков для текста я использовала jina-embeddings-v3. Эта модель одна из самых используемых на HF + показала наилучшие результаты в моих исследованиях.

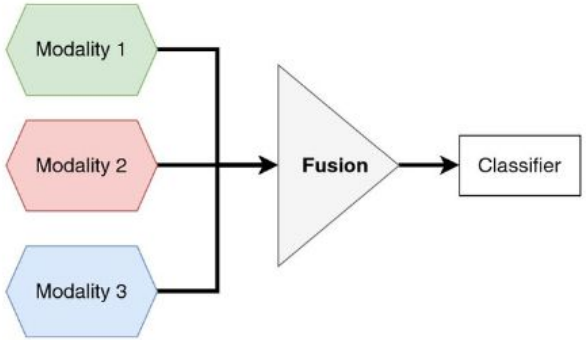
Для работы с несколькими модальностями я использовала разные подходы:

- Early Fusion
- Late Fusion
- Hybrid Fusion

Early Fusion (feature-level fusion)

Используемые модели и полученные лучшие результаты:

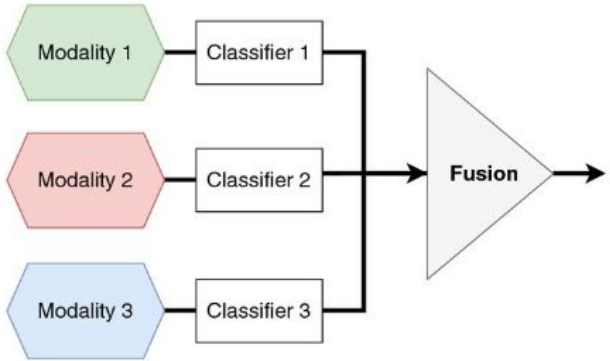
- Transformer: RMSE = 0.30322
- Mamba: RMSE = 0.45980
- LSTM: RMSE = **0.27722**



Late Fusion (decision-level fusion)

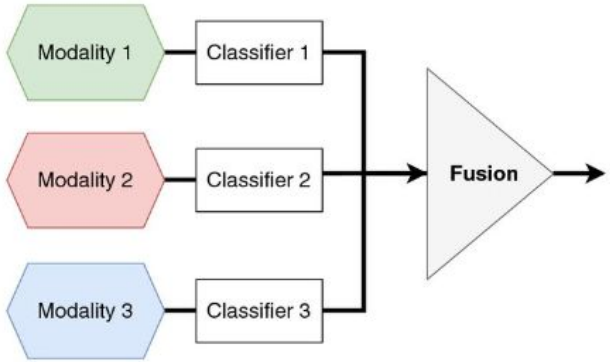
Для начала я протестировала модели на каждой из модальностей отдельно.

Модель для табличных данных	Модель для текстовых данных	RMSE
Catboost	-	0.33298
RNN	-	**0.27903**
-	Transformer	0.28390
-	LSTM	**0.27610**



Late Fusion (decision-level fusion)

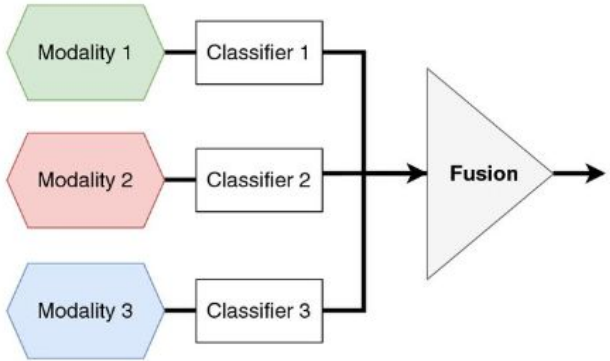
Модель для табличных данных	Модель для текстовых данных	RMSE
Catboost	Transformer	0.26810
RNN	Transformer	**0.25335**
Catboost	LSTM	0.26472
RNN	LSTM	0.27541



Late Fusion (decision-level fusion)

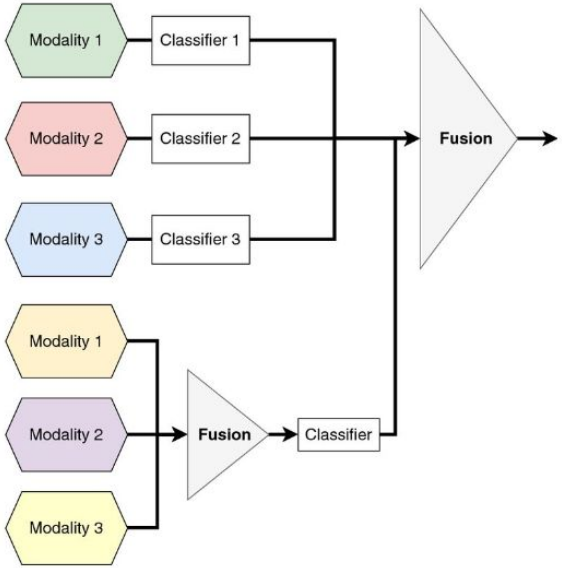
Взвешивание результатов (представлены лучшие комбинации)

Модель для табличных данных	Модель для текстовых данных	RMSE
Catboost	Transformer (x2)	0.26360
RNN (x2)	Transformer	**0.25573**
Catboost	LSTM (x2)	0.25867
RNN	LSTM (x2)	0.27516



Hybrid Fusion

Модель для Early Fusion	Модель для табличных данных	Модель для текстовых данных	RMSE
Transformer	Catboost	Transformer	**0.24671**
Transformer	Catboost	LSTM	0.26203
Transformer	RNN	Transformer	0.26640
Transformer	RNN	LSTM	0.28659



Hybrid Fusion

Модель для Early Fusion	Модель для табличных данных	Модель для текстовых данных	RMSE
Mamba	Catboost	Transformer	0.33429
Mamba	Catboost	LSTM	0.33412
Mamba	RNN	Transformer	0.31043
Mamba	RNN	LSTM	0.31562
LSTM	Catboost	Transformer	0.26219
LSTM	Catboost	LSTM	0.26934
LSTM	RNN	Transformer	**0.25185**
LSTM	RNN	LSTM	0.26570

Планы

Добавить новую модальность: изображение. (При этом уже произведена подготовка к этому этапу - например, извлечение признаков из изображения).

Протестировать Cross-Attention.