



ФКН НИУ ВШЭ

7 мая 2025 г.

# Игра с генерацией заданий и оценкой ответов методами машинного обучения

Научный руководитель: Макарова Мария

Спиридонов Денис  
студент МОВС, 2 курс



# Введение

Игра с генерацией заданий и оценкой ответов методами машинного обучения

## Цель проекта

- Разработка игры, позволяющей пользователю в развлекательной форме закреплять материал по изучаемому предмету.
- В качестве учебной области выбрана статистика, модель может быть дообучена для других дисциплин на соответствующих данных.
- Механика игры поддерживается двумя нейросетями: энкодер-декодером для генерации вопросов и кросс-энкодером для оценки ответов игрока.



# Демо

## Игра с генерацией заданий и оценкой ответов методами машинного обучения

Шурик: 3

Questions: 1 / 5

Time: 00:30

Балл

Шурик (You)

Score: 3

1 of 5 questions

Score progress: 3 / 15

Last Answer:  
Yes, it can be applied to both regression and classification problems.

Балбес

Score: 1

1 of 5 questions

Score progress: 1 / 15

Last Answer:  
Some believe it guarantees perfect model performance, but it only estimates generalization.

Topic: p-values

What are some limitations of using p-values?

Click for hint

00:30

Type your answer...

Submit

Skip



# Анализ данных

Игра с генерацией заданий и оценкой ответов методами машинного обучения

input	question	answer_1	answer_2
Missing 1299 Unique values	Missing 1299 Unique values	Missing 1299 Unique values	Missing 1299 Unique values
0 generate a effect quest	What happens to the mea	The mean may increase o	It adjusts to reflect t
1 generate a definition q	What is the mean in sta	The mean is the average	It represents a central
2 generate a suitability	Can the mean be used wi	The mean is best suited	It is not meaningful fo
3 generate a application	How is the mean used in	People use the mean to	Businesses use it to an
4 generate a importance q	Why is the mean importa	It helps summarize larg	It is often used to com
5 generate a comparison q	How does the mean compa	The mean calculates the	The mean can be influen
6 generate a limitation q	What are some limitatio	The mean can be skewed	It might not represent
7 generate a preference q	Why might someone choos	The mean uses all data	It is straightforward a
8 generate a explanation	How does the mean help	It provides a simple wa	Researchers use it to s
9 generate a caution ques	What is a situation whe	When the data has outli	If the data is heavily

- Промпты для генерации вопроса отличаются по типу (explanation, definition etc.) и по топику (mean, variance etc.)
- В процессе игры произвольно комбинируются для инференса



# Анализ данных

Игра с генерацией заданий и оценкой ответов методами машинного обучения

topic	context
Missing 130 Unique values	Missing 130 Unique values
0 mean	The mean is a way to find the average value of
1 median	The median is a way to find the middle value in
2 mode	The mode is a concept used to describe the most
3 variance	Variance is a measure of how spread out or disp
4 standard deviation	Standard deviation is a way to measure how spre
5 normal distribution	The normal distribution is a way to describe ho
6 z-score	A z-score is a way to measure how far a particu
7 probability	Probability is a way to describe how likely it
8 sampling	Sampling is the process of selecting a small gr
9 confidence interval	A confidence interval is a range of values that

- Для улучшения обобщающей способности опционально добавляем контекст к промπτу генератора, которой представляет собой описание топика



# Анализ данных

Игра с генерацией заданий и оценкой ответов методами машинного обучения

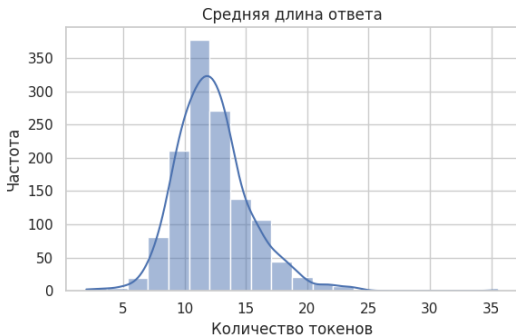
question	answer_1	answer_2	student_answer	score
Missing	Missing	Missing	Missing	Missing
How is correlation diff...	0.2% The mean may increas...	0.1% It adjusts to reflect the...	0.1%	10362
What happens to the ...	0.1% They are listed as a so...	0.1% You might see them W...	0.1%	Unique values
0 What happens to the mea	The mean may increase	o It adjusts to reflect t	The mean changes depend	3
1 What happens to the mea	The mean may increase	o It adjusts to reflect t	Adding a new value can	3
2 What happens to the mea	The mean may increase	o It adjusts to reflect t	The mean will go up if	2
3 What happens to the mea	The mean may increase	o It adjusts to reflect t	The mean changes when a	2
4 What happens to the mea	The mean may increase	o It adjusts to reflect t	The mean is affected by	1
5 What happens to the mea	The mean may increase	o It adjusts to reflect t	The mean might be diffe	1
6 What happens to the mea	The mean may increase	o It adjusts to reflect t	The mean always stays t	0
7 What happens to the mea	The mean may increase	o It adjusts to reflect t	The mean is not related	0
8 What is the mean in sta	The mean is the average	It represents a central	The mean is the sun of	3
9 What is the mean in sta	The mean is the average	It represents a central	The mean is the average	3

- Сгенерированный вопрос подаем в классификатор
- Опционально дополняем эталонными вопросами
- Добавляем ответ игрока оцениваемый от 0 до 3



# Анализ данных

Игра с генерацией заданий и оценкой ответов методами машинного обучения

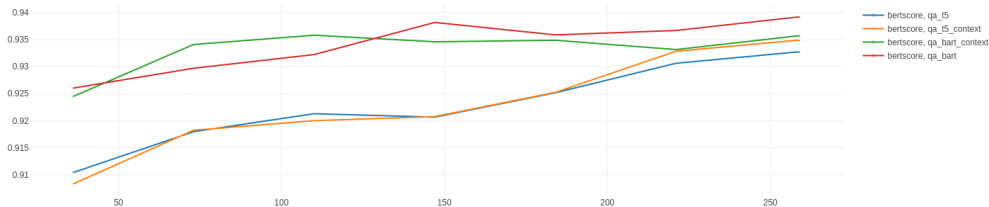


- Большинство ответов имеют длину 8-14 токенов, распределение почти нормальное
- Оценки сбалансированы, что исключает необходимость в взвешивании классов при обучении



# Сравнение моделей

Игра с генерацией заданий и оценкой ответов методами машинного обучения



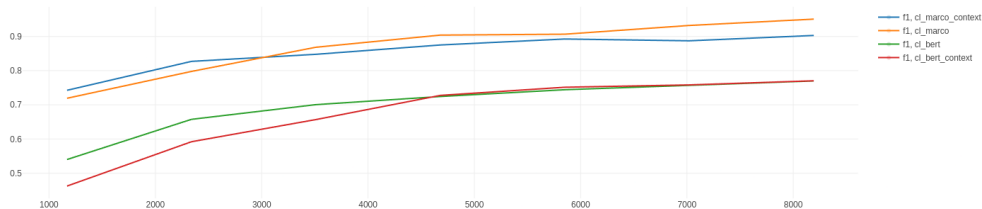
- Генерируемые вопросы должны быть семантически близки к исходным, для оценки использовался BERTScore
- Проведены эксперименты с двумя энкодер-декодерами с добавлением контекста к промπτу и без





# Сравнение моделей

Игра с генерацией заданий и оценкой ответов методами машинного обучения

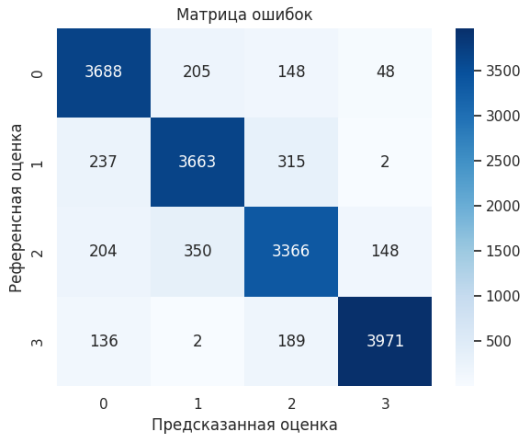


- Все оценки одинаково значимы, для оценки использовалась F1 сочетающая полноту и точность
- Проведены эксперименты с двумя кросс-энкодерами, оценивающими вопрос-ответ совместно на уровне всей последовательности



# Качество классификатора

Игра с генерацией заданий и оценкой ответов методами машинного обучения



- Классификатор точно оценивает большинство ответов. Очень мало сильных промахов
- Модель не склонна занижать или завышать оценки



## Клиент

Игра с генерацией заданий и оценкой ответов методами машинного обучения

- Фронтенд реализован на React+TypeScript+Vite. Архитектура построена по принципу компонентного UI с минимальными зависимостями
- Состояние игры управляется через MobX, что упростило взаимодействие с API и навигацию
- TypeScript обеспечивает строгую типизацию, повышает надежность кода и простоту отладки



## Заключение

Игра с генерацией заданий и оценкой ответов методами машинного обучения

### Выводы

- Добавление контекста полезно для слабых моделей, но не дает прироста сильным моделям
- При достаточно небольшом объеме обучающих данных классификатор достиг высоких F1 значений
- Объединение моделей для генерации вопросов и оценки ответов дает работоспособную архитектуру
- Разработка игры на React позволяет перенести основную логику игры на клиент, снижает нагрузку на сервер и упрощает масштабирование