



# Предсказание пола пользователя соцсетей

**Авторы:** Малышева Анастасия, Петрухин Владислав,  
Киселев Виктор, Сидоренко Николай

# Проблема

Качественная реклама на специализированных площадках требует хорошего понимания аудитории. Важно, чтобы реклама была эффективной: демонстрировалась пользователям соответствующего гендера.

**Цель проекта:** разработать модель машинного обучения, которая будет предсказывать пол пользователя соцсетей.

# Данные



## Гендер пользователей

Тестовый и тренировочные датасеты с информацией о поле субъекта.



## Гео-информация

Тестовый и тренировочные датасеты с обезличенной гео-информацией.



## Ссылки

Информация о ссылках, которые были показаны пользователю.

# Обработка данных



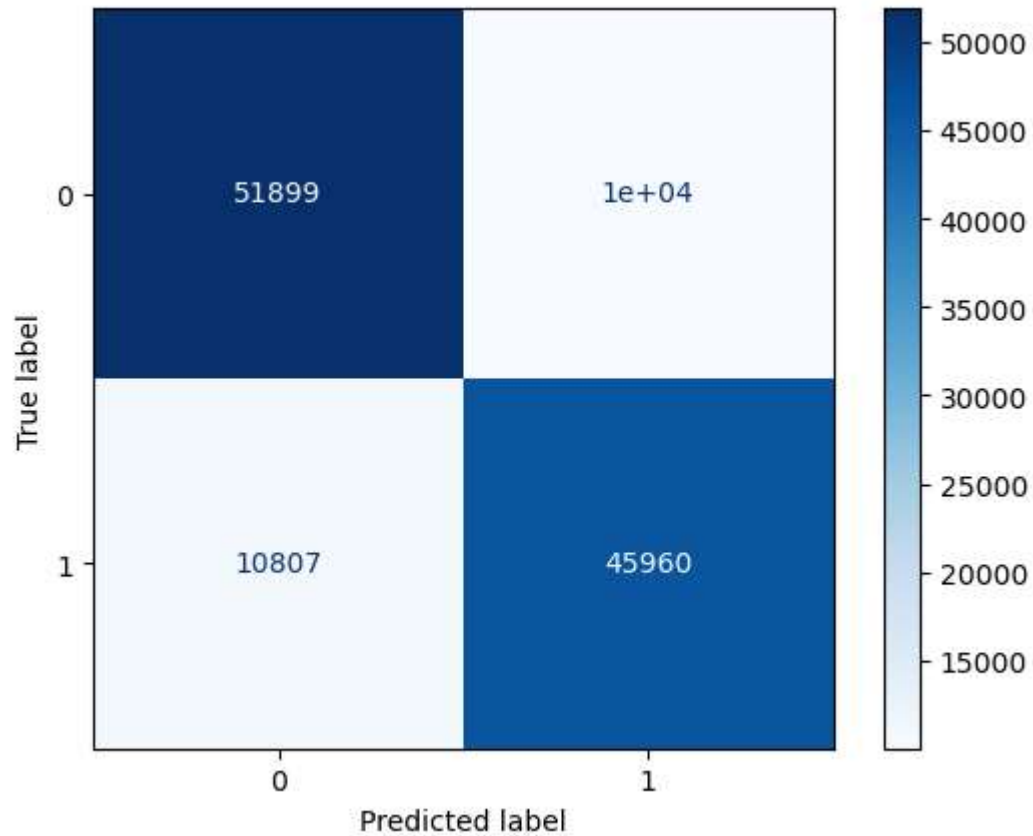
# Random forest

ACC

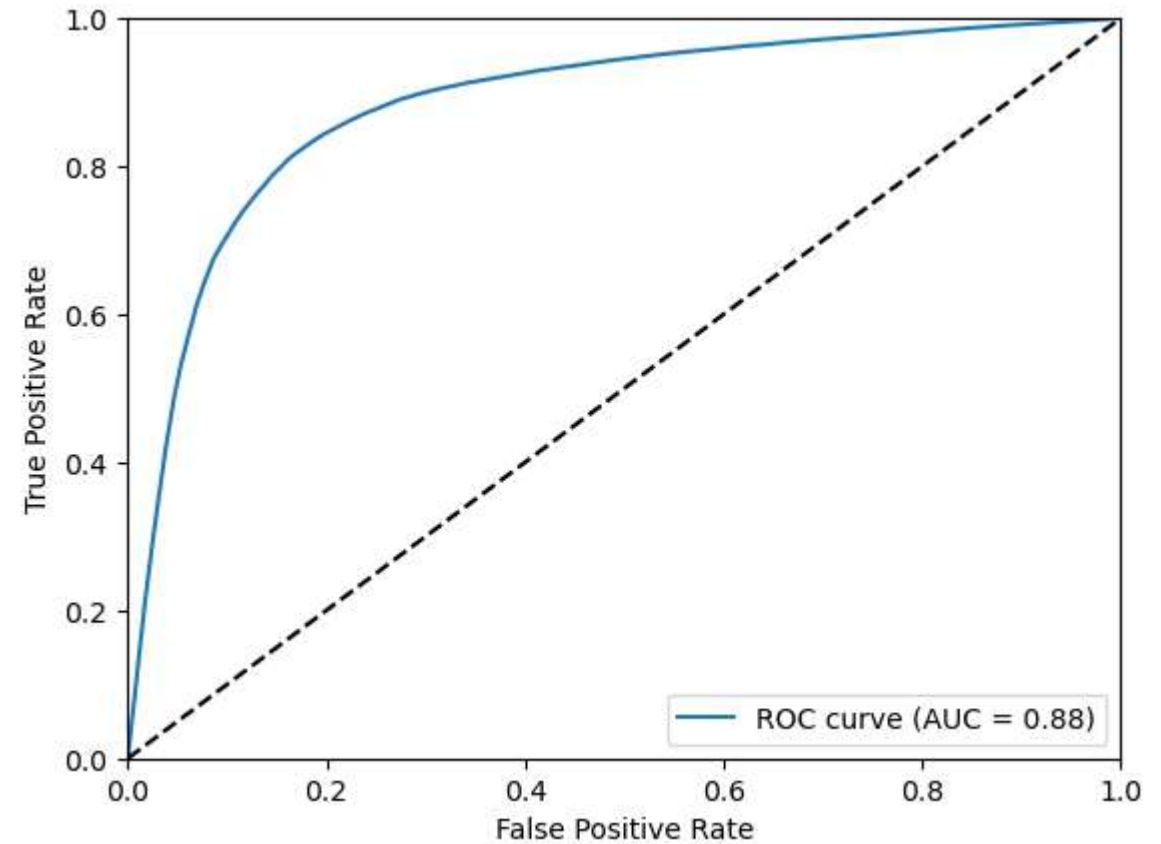
0.82

F1

0.81



Матрица ошибок



Кривая ROC-AUC

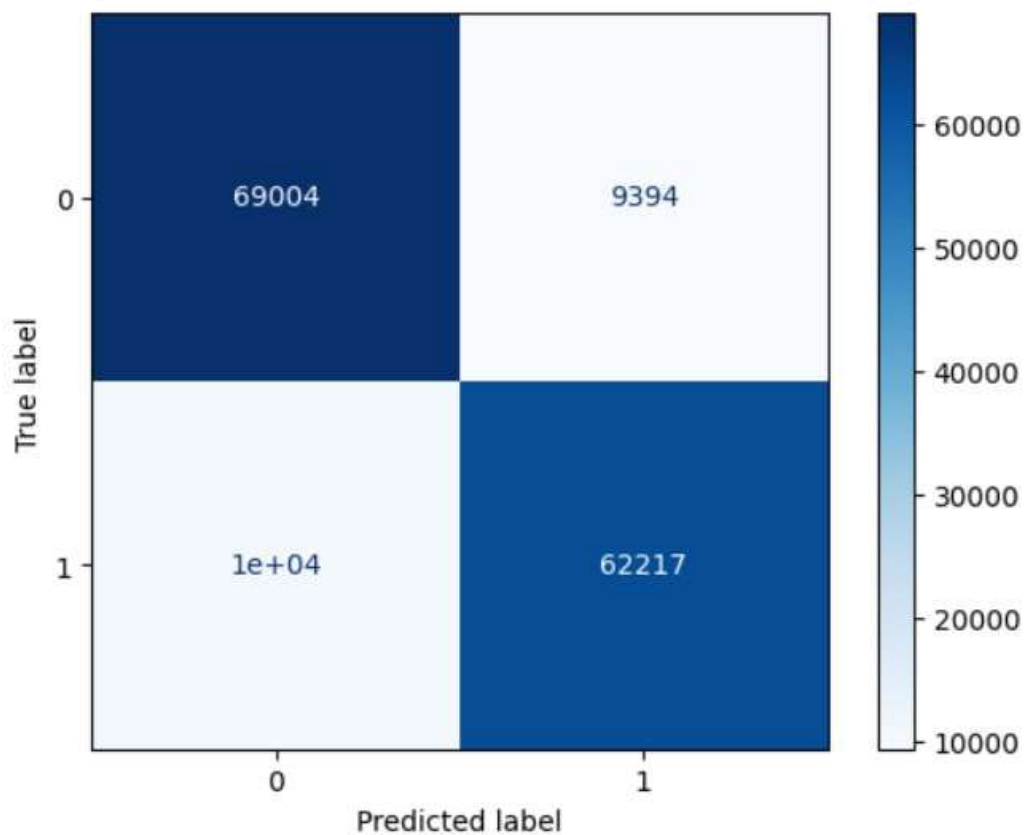
# XGBoost

ACC

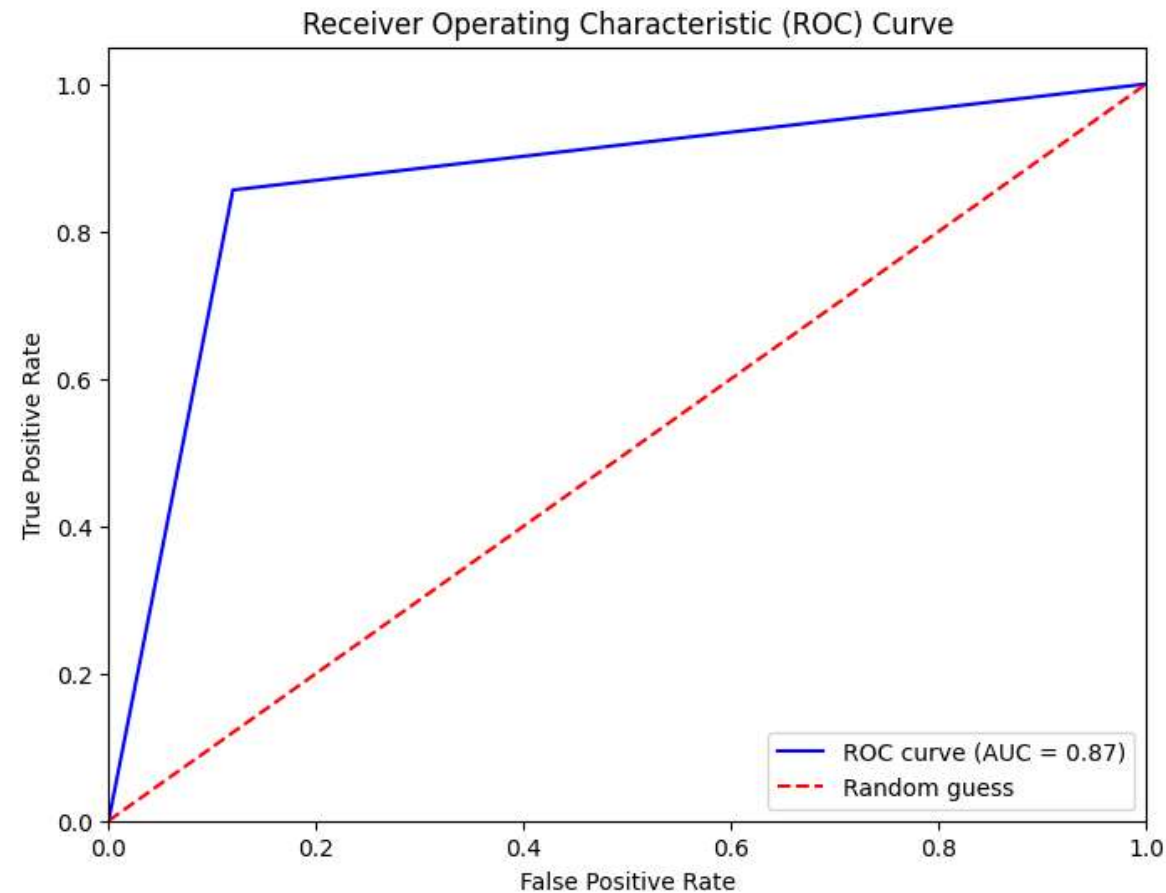
0.87

F1

0.86



Матрица ошибок



Кривая ROC-AUC

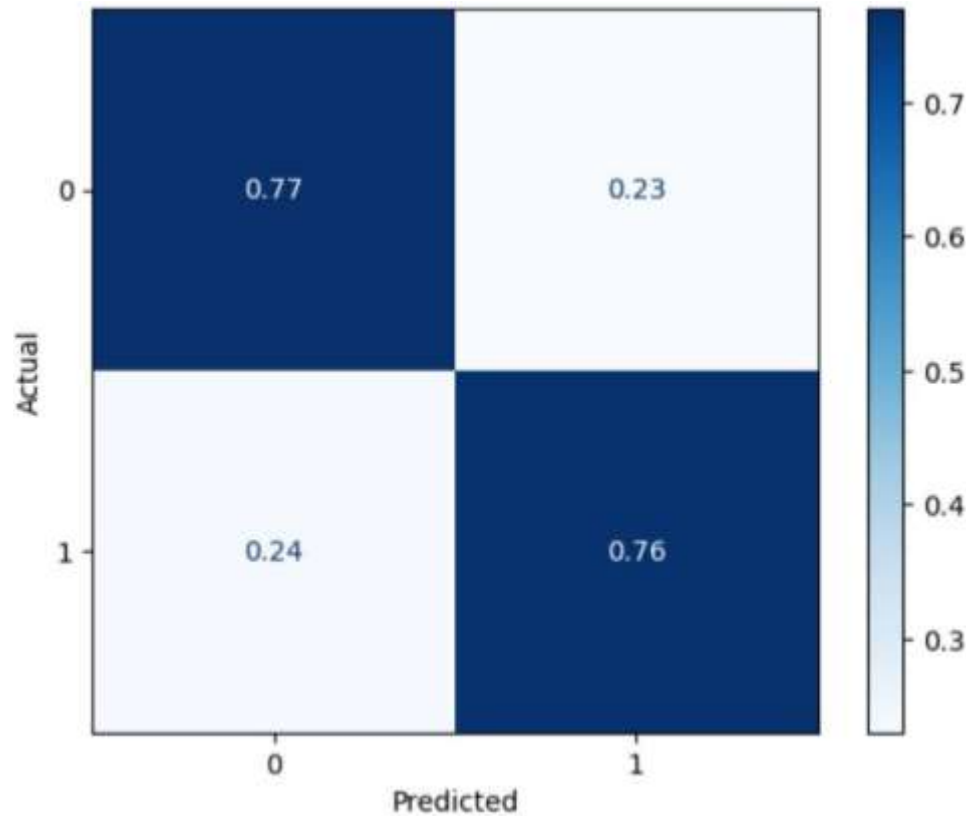
# LightGBM

ACC

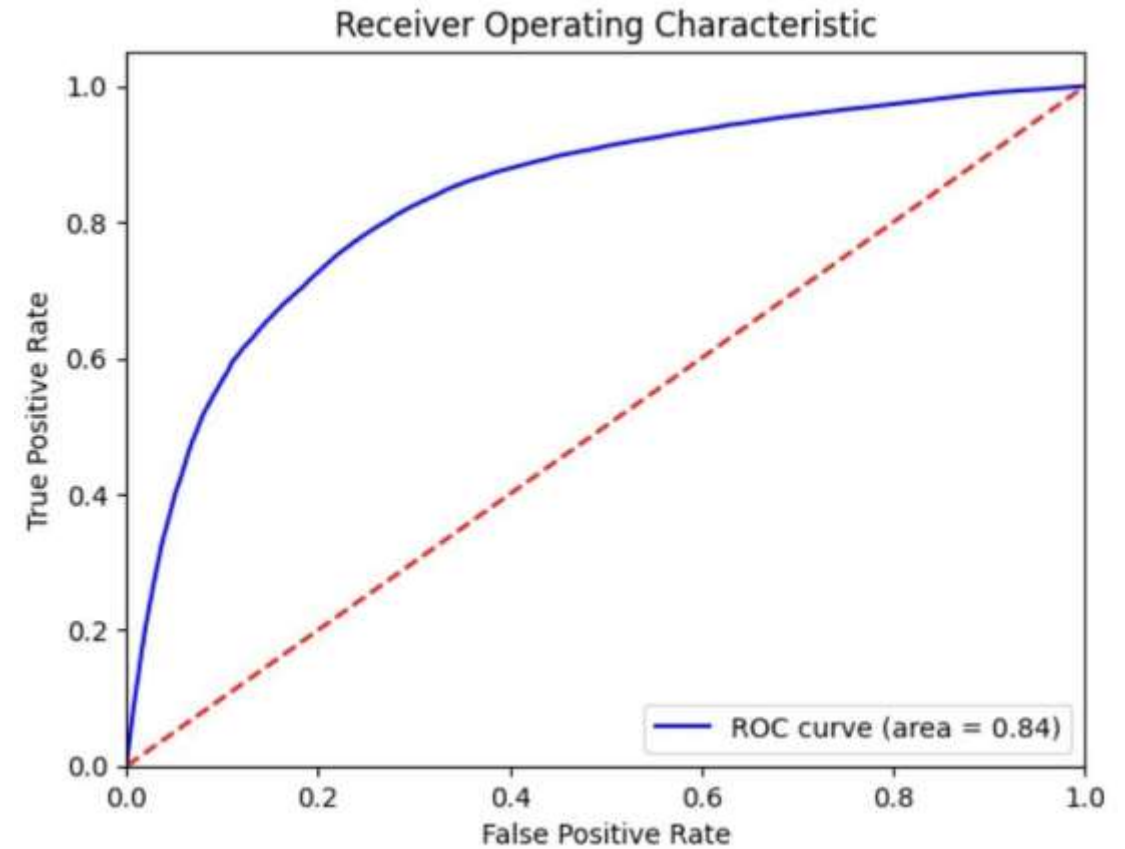
0.77

F1

0.76



Матрица ошибок



Кривая ROC-AUC

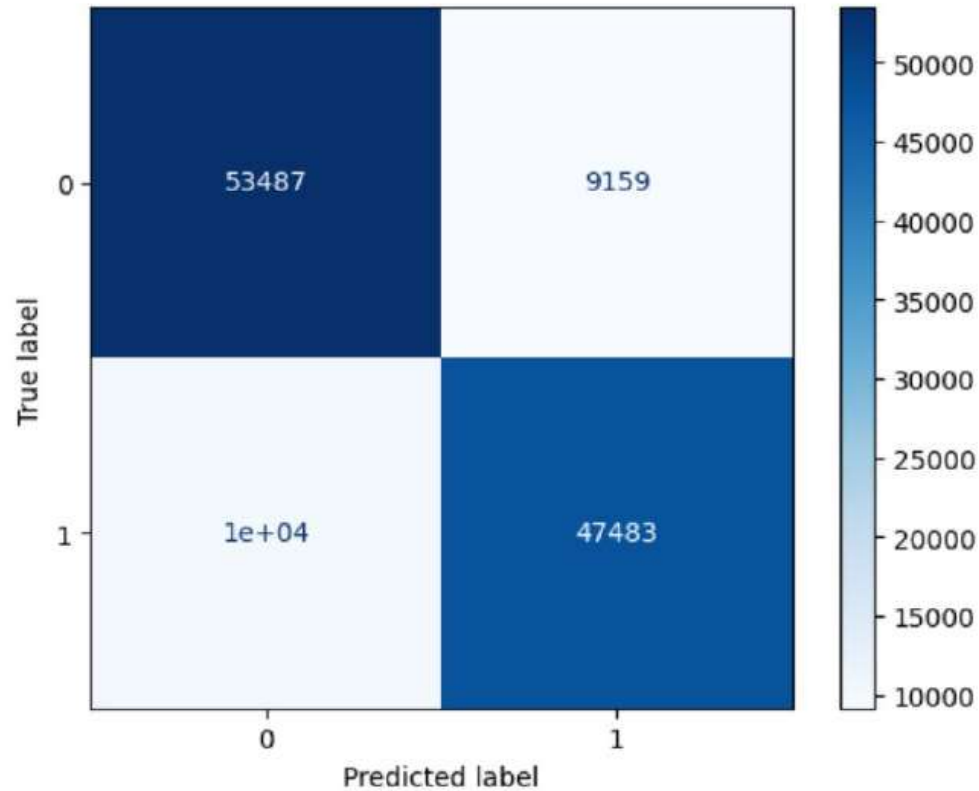
# CatBoost

ACC

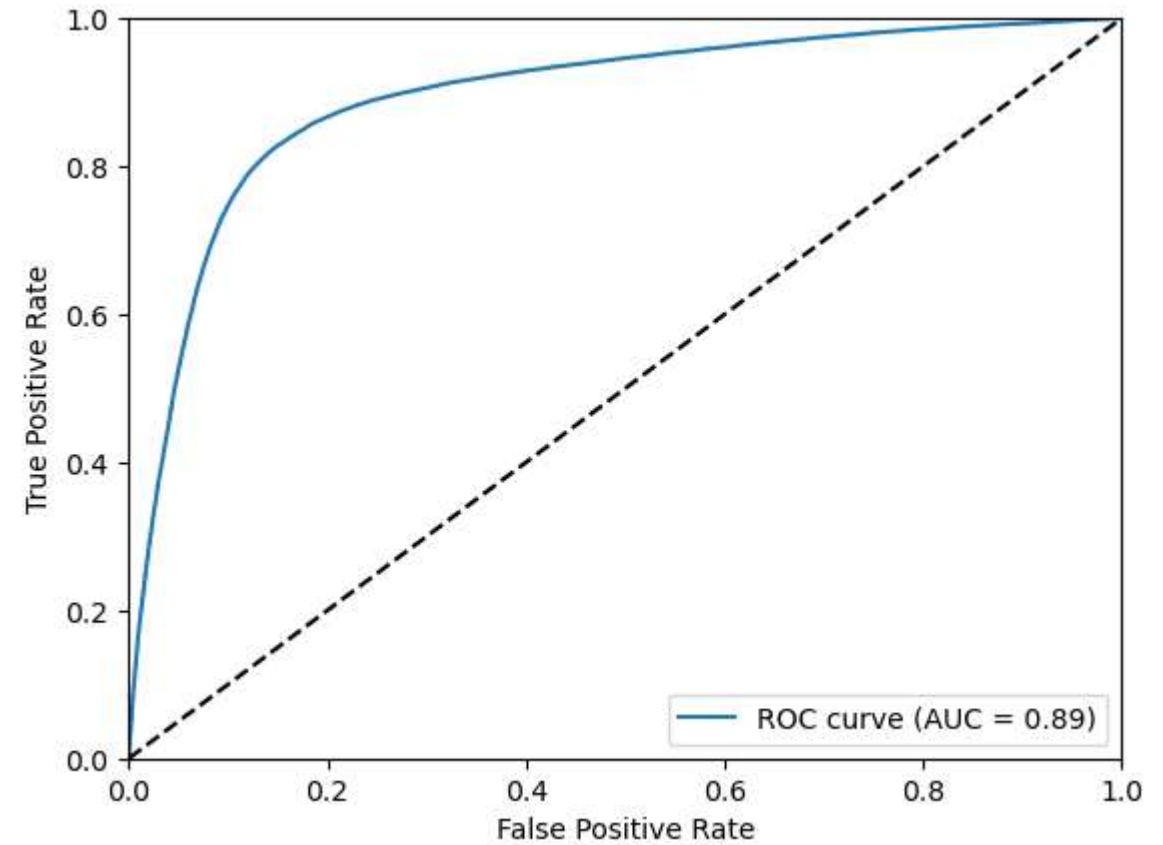
0.84

F1

0.83



Матрица ошибок



Кривая ROC-AUC



# Выводы

## Основные метрики, свод

		Accuracy	F1	ROC-AUC
1	Random forest	0.82	0.82	0.824
2	XGBoost	0.87	0.86	0.868
3	CatBoost	0.84	0.83	0.894
4	LightGBM	0.77	0.77	0.84

## Лучшие метрики:

ACC

0.87

2

F1

0.86

2

ROC-AUC

0.89

3

# Бизнес-ценность и планы

Обученная модель может помочь при формировании таргетированной рекламы.

## Планы для развития:

1. Расширение диапазона исходных данных.
2. Инженерия признаков.
3. Детальный анализ ошибок модели при обучении.



Спасибо за  
ВНИМАНИЕ