



Классическое Машинное Обучение

Прогнозирование сердечного приступа и анализ данных

Выполнили:
Пелых Александр
Мельников Максим

Описание полей данных

Age – возраст пациента

Sex - пол пациента

Ca – кол-во круглых сосудов

Cr – тип боли в груди

Exang – стенокардия вызванная физической нагрузкой

Trtbps – артериальное давление в состоянии покоя

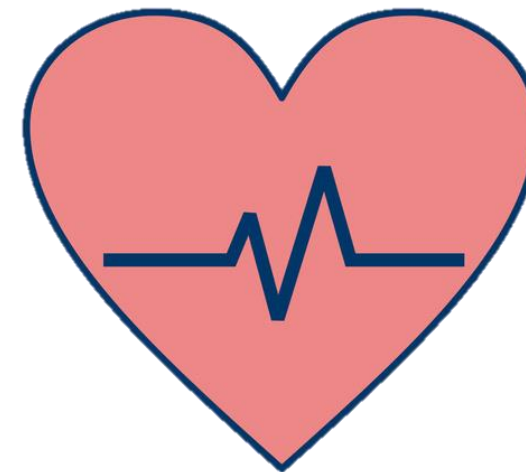
Chol – холестерин, определяемый с помощью ИМТ датчика

Fbs – уровень сахара в крови натощак

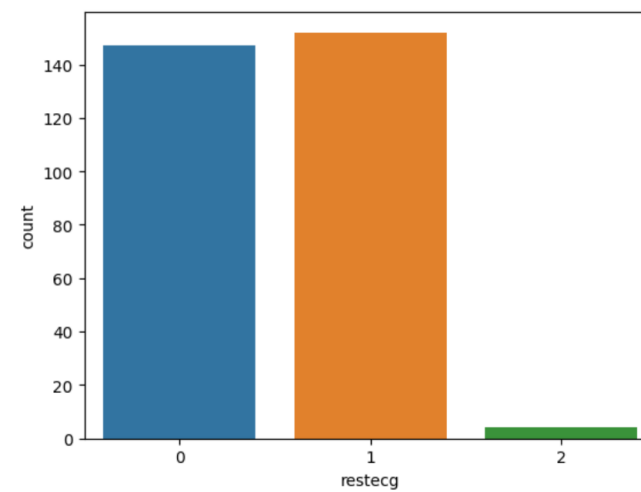
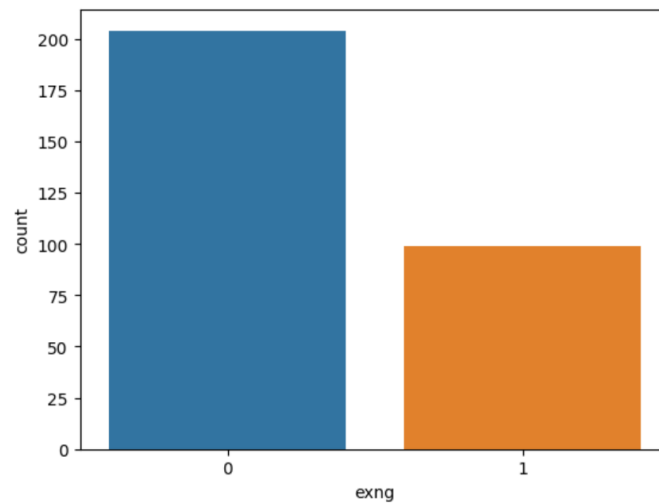
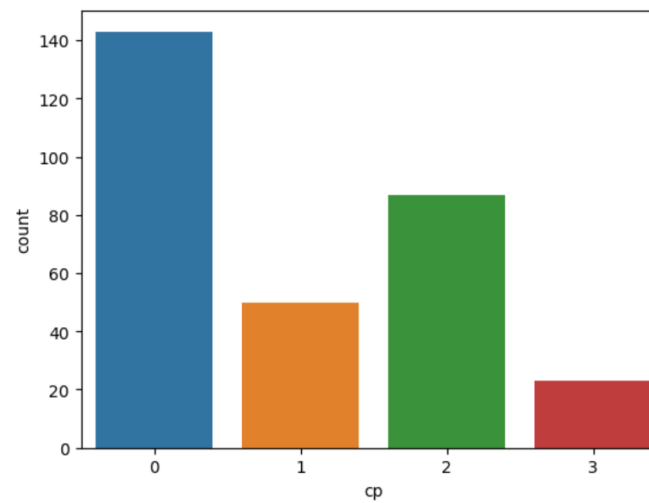
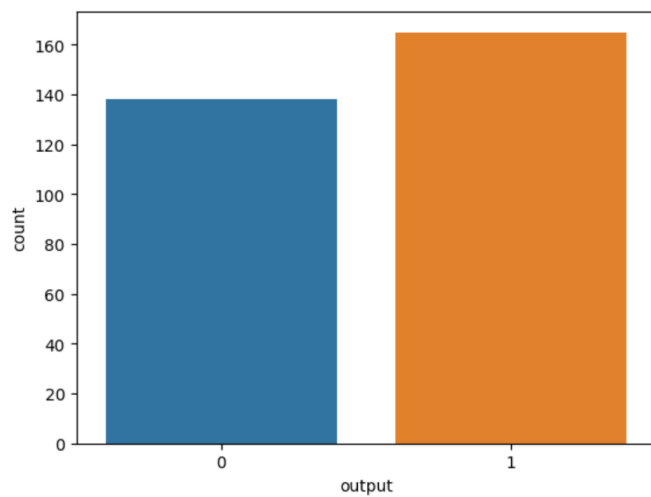
Rest ecg – результаты электрокардиографии в состоянии покоя

Thalach – максимальная частота сердечных сокращений

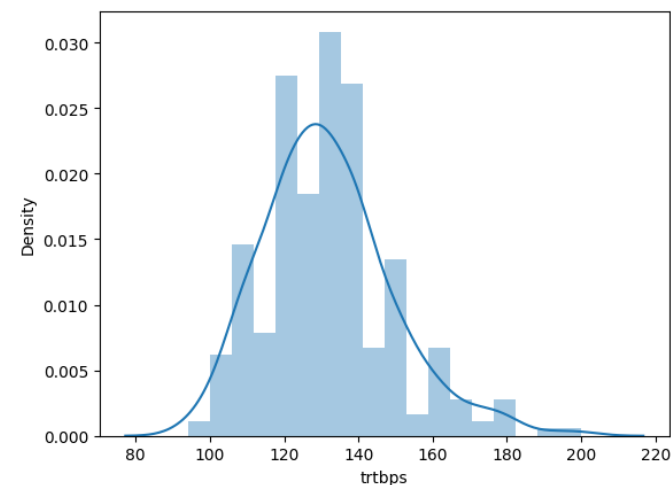
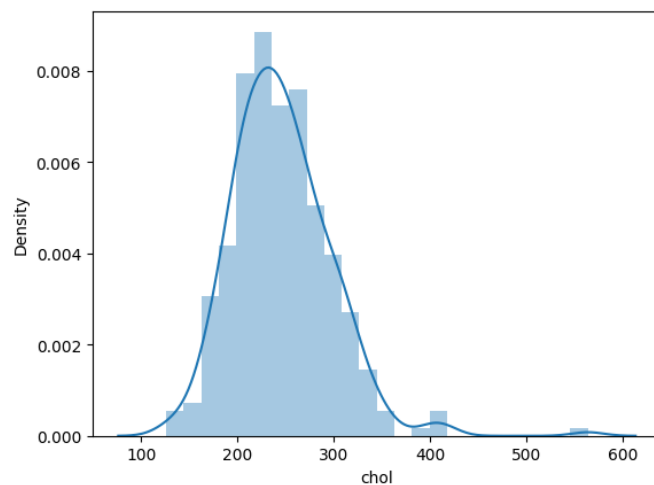
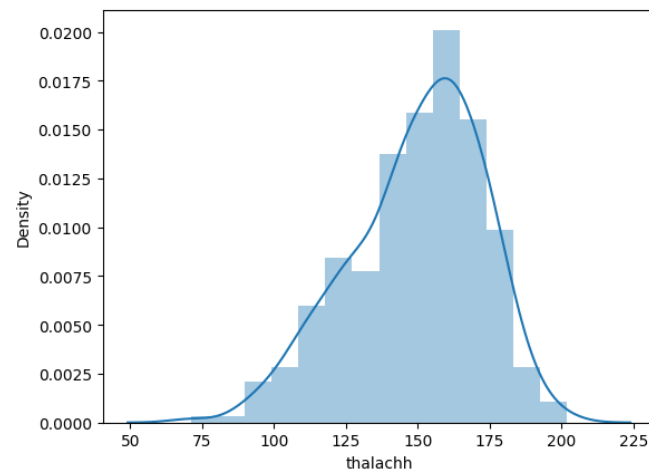
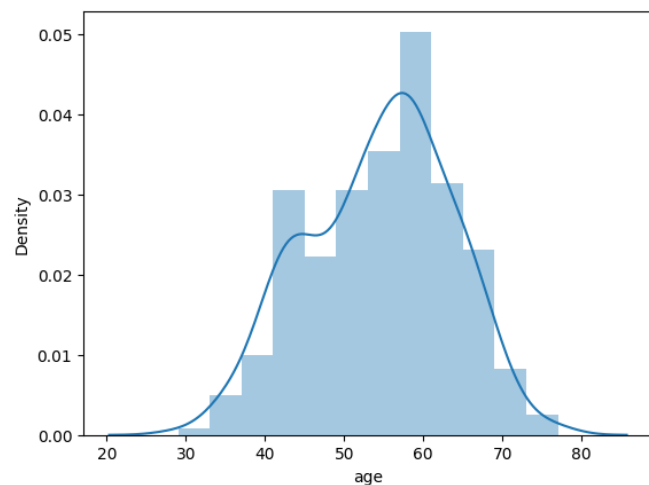
Target – бинарный таргет по предсказанию сердечного приступа



Анализ исходных данных

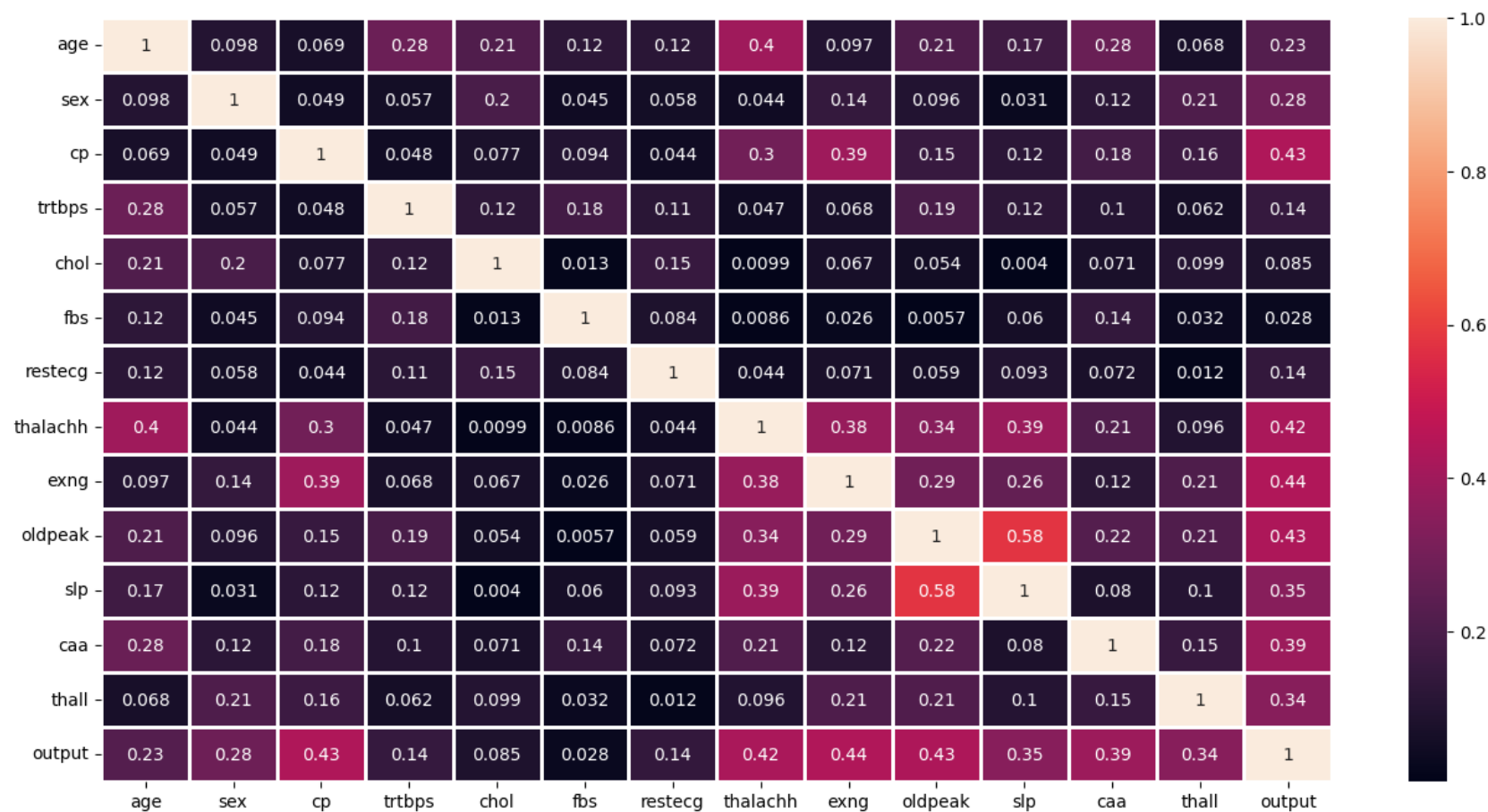


Анализ исходных данных



Анализ исходных данных

Корреляция признаков и их влияние на выходные данные



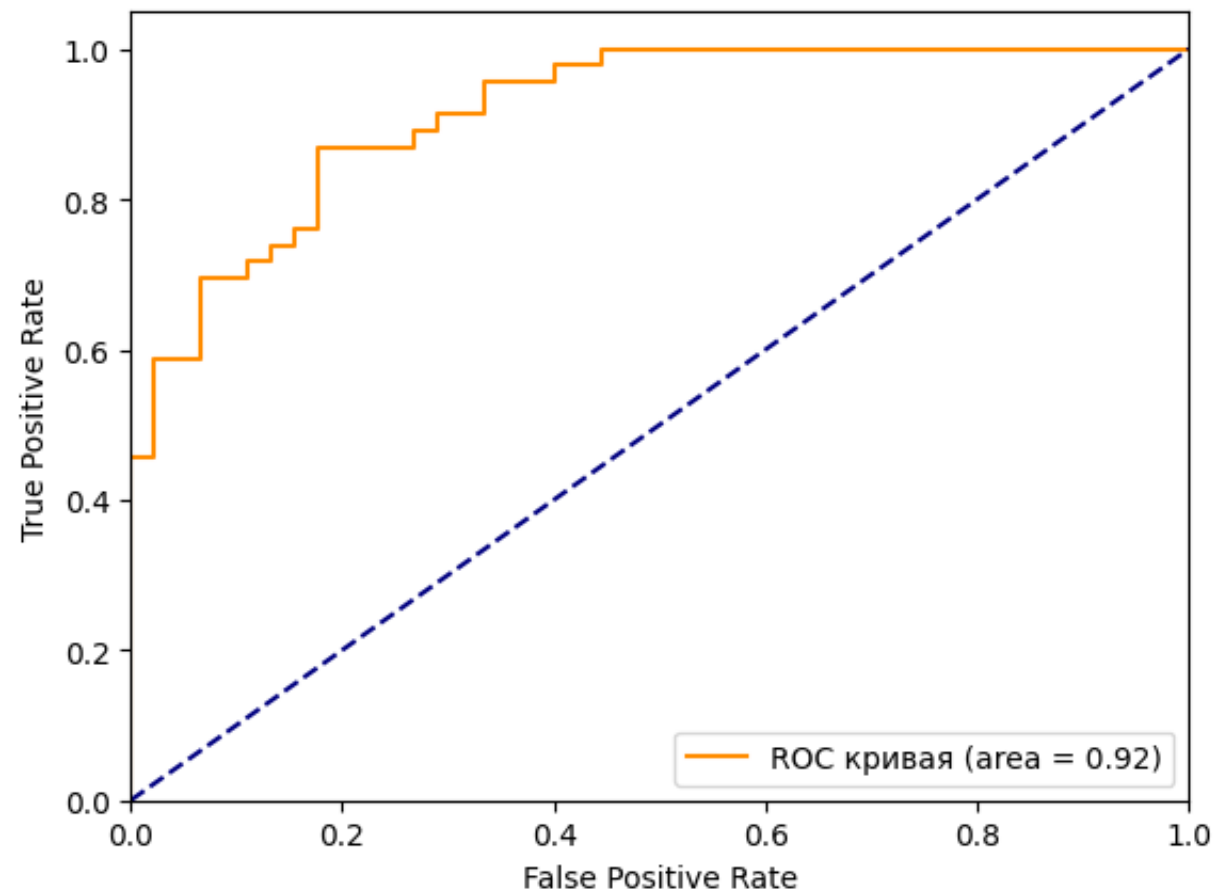
Логистическая регрессия

Модель бинарной классификации, использующая функцию сигмоиды для расчета вероятности на основе коэффициентов функции независимой переменной, рассчитанных в процессе обучения



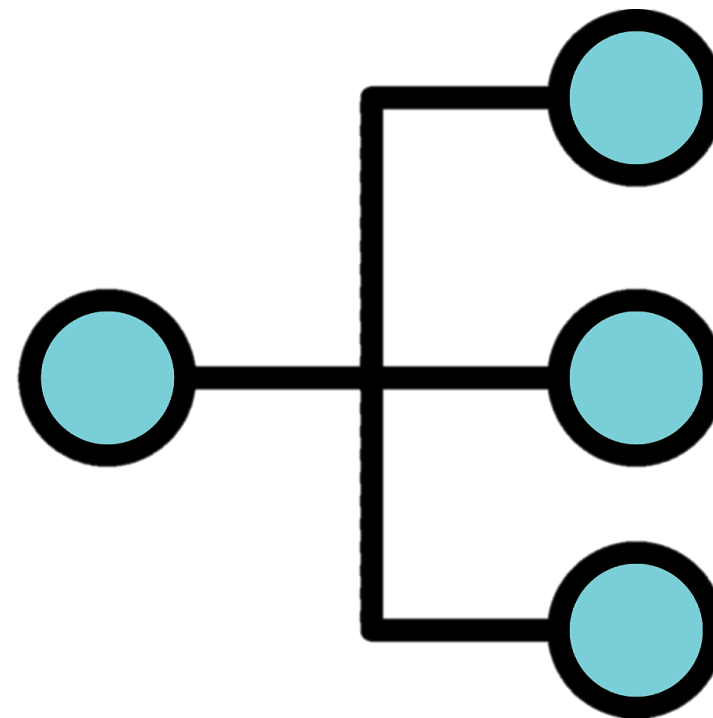
Логистическая регрессия

ROC - кривая



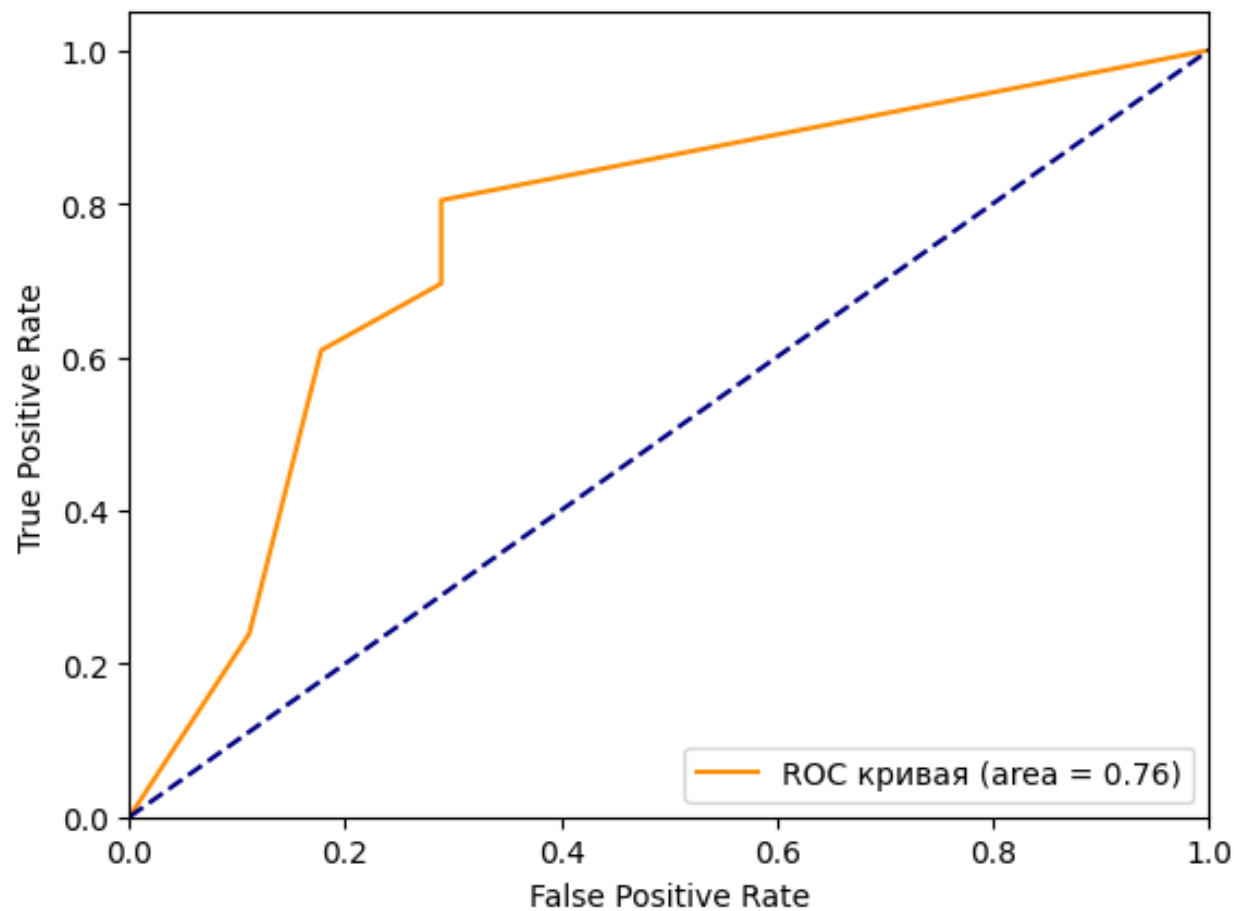
Решающее дерево

Модель нейронной сети, основанной на разделении выборок объектов по их критериям до тех пор, пока не будет установлен класс объекта, или не будет получено значение для регрессии



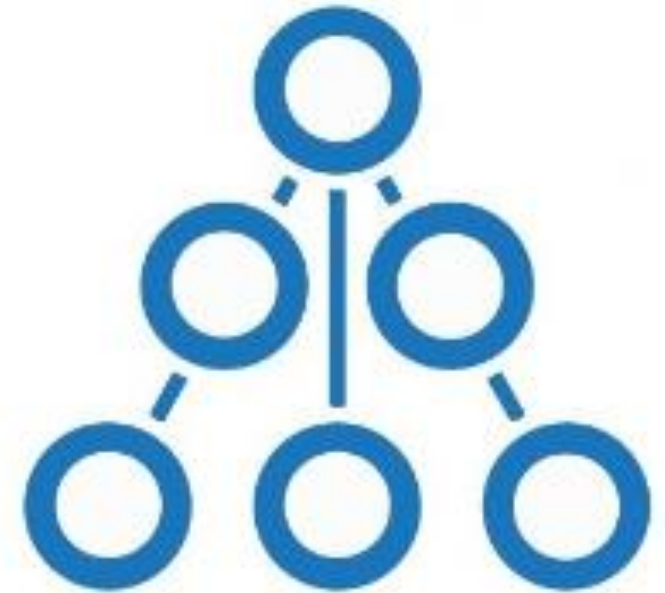
Решающее дерево

ROC - кривая



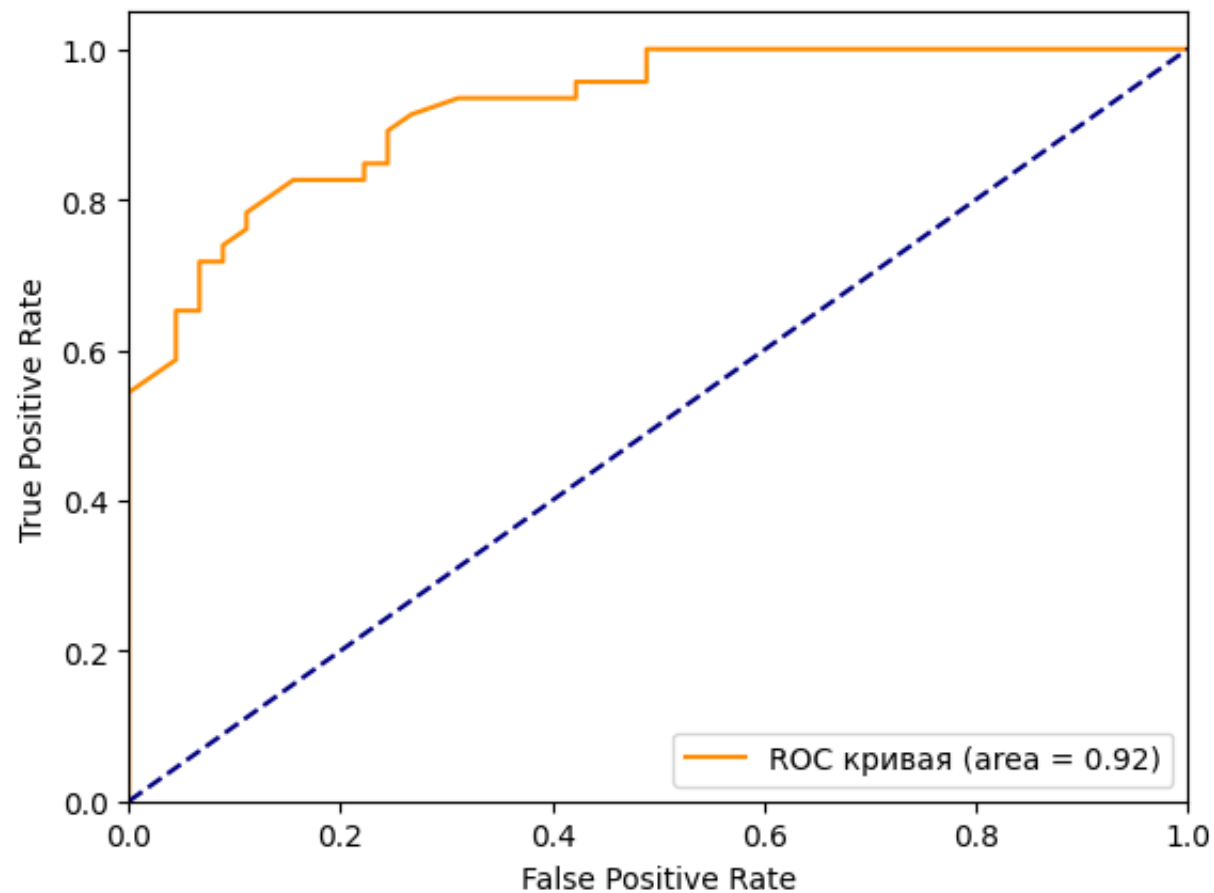
Случайный лес

Модель состоящая из множества отдельных решающих деревьев и основанная на беггинге и случайности признаков при построении каждого дерева



Случайный лес

ROC - кривая



Сравнение моделей

	Логистическая регрессия	Решающее дерево	Случайный лес
Точность	81.32%	72.53%	80.22%
Precision	0.774	0.723	0.792
Recall	0.891	0.739	0.826
F1-Score	0.828	0.731	0.809



Спасибо за
внимание