

МИНОБРНАУКИ РОССИИ  
федеральное государственное бюджетное образовательное  
учреждение высшего образования  
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ  
ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ»  
(СПбГЭУ)

Факультет экономики и финансов  
Кафедра прикладной математики и экономико-математических методов

# Выпускная квалификационная работа Моделирование оценки объекта недвижимости

Обучающегося очной формы обучения группы Э-2010:

**Дмитриева Александра Ростиславовича**

Руководитель ВКР: к.э.н., доцент Заграновская Анна Васильевна





# Актуальность, цель, задачи



**Цель ВКР** – моделирование **стоимости квадратного метра** коммерческого торгового объекта путем анализа и оценки актуальных алгоритмов согласно качеству прогноза.

**Объект исследования** – рынок коммерческой торговой недвижимости города Санкт-Петербурга.

**Предмет исследования** – методы моделирования оценки стоимости коммерческой торговой недвижимости в Санкт-Петербурге.

**Использованные инструменты:** Python, QGIS, GitHub.

## **Задачи ВКР:**

- изучение теоретических основ рынка коммерческой недвижимости, его анализа и моделирования;
- сбор и обработка данных, образующих рынок;
- анализ собранных данных, включая геоспатические и корреляционные особенности объектов недвижимости;
- разработка моделей оценки, для выявления оптимального алгоритма, способного точно описывать стоимость квадратного метра коммерческого торгового объекта;
- сравнение и оценка результатов модели, ее интерпретация.



# СБОР ДАННЫХ



- **Средство сбора данных** – Парсер на Python
- **Источник данных** – Циан

Время сбора

Объем данных

Кол-во признаков

**14+**  
часов

**2500+**  
наблюдений

**16**  
столбцов

**Парсер** — это программа, сервис или скрипт, который собирает данные с указанных веб-ресурсов, обрабатывает их и выдает в нужном формате.

Пример наблюдения

The screenshot shows a real estate listing on a Russian website. The main title is "Продается Торговая площадь, 81,1 м² в ЖК «Солнечный город»". The location is "Санкт-Петербург, р-н Красносельский, Сосновая поляна, просп. Ветеранов, 185к3". The price is "25 500 000 Р.". The listing includes a large photo of the building, a smaller photo of the entrance, and a gallery of 6 photos. The agent is "Вера Щепанова" from "House Estate". The listing is marked as "Востребованная локация" (Popular location).



# ОБРАБОТКА ДАННЫХ



- Количественные признаки приведены к числовому формату;
- Исключены выбросы и аномалии;
- Адреса объектов были расшифрованы в географические координаты;
- Добавлены новые признаки на основе расположения объектов.

Для обработки адресов и перевода их в систему координат было решено обратиться к **геокодеру** от «Яндекс».

**Геокодер** — это программа, которая переводит адрес текстового формата в географические координаты согласно открытой базе данных того или иного источника.

## Формула расчета расстояния между двумя точками.

$$L = q * \arccos(\sin\varphi_1 * \sin\varphi_2 + \cos\varphi_1 * \cos\varphi_2 * \cos|\lambda_1 - \lambda_2|)$$

Где:

L — искомое расстояние между пунктами 1 и 2;

q — расстояние в одном градусе, константа ~ 111.2 км;

$\varphi_1, \varphi_2$  — географические широты пунктов 1 и 2;

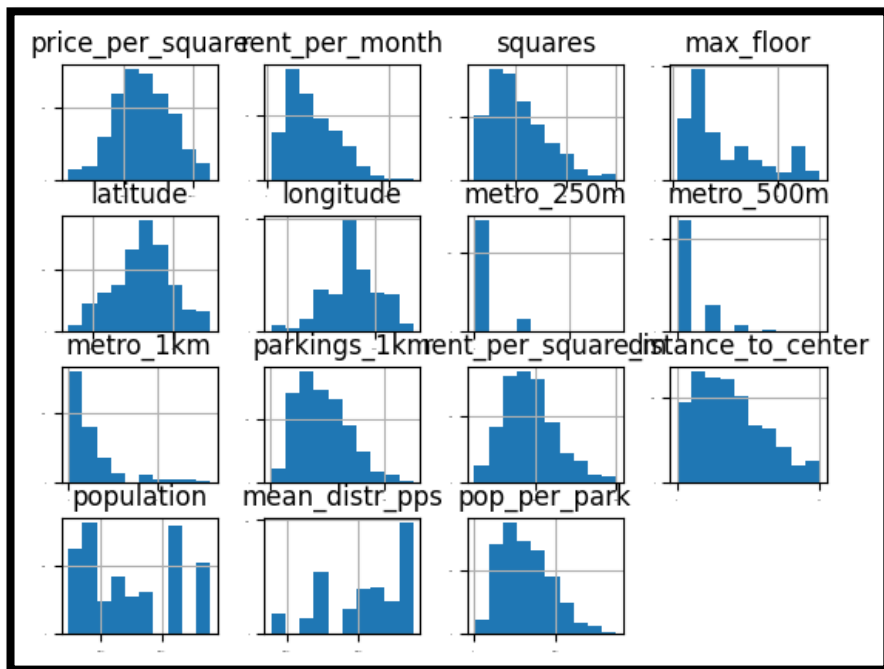
$\lambda_1, \lambda_2$  — географические долготы пунктов 1 и 2.



# ОБРАБОТКА ДАННЫХ



Распределение всех признаков в  
выборке:



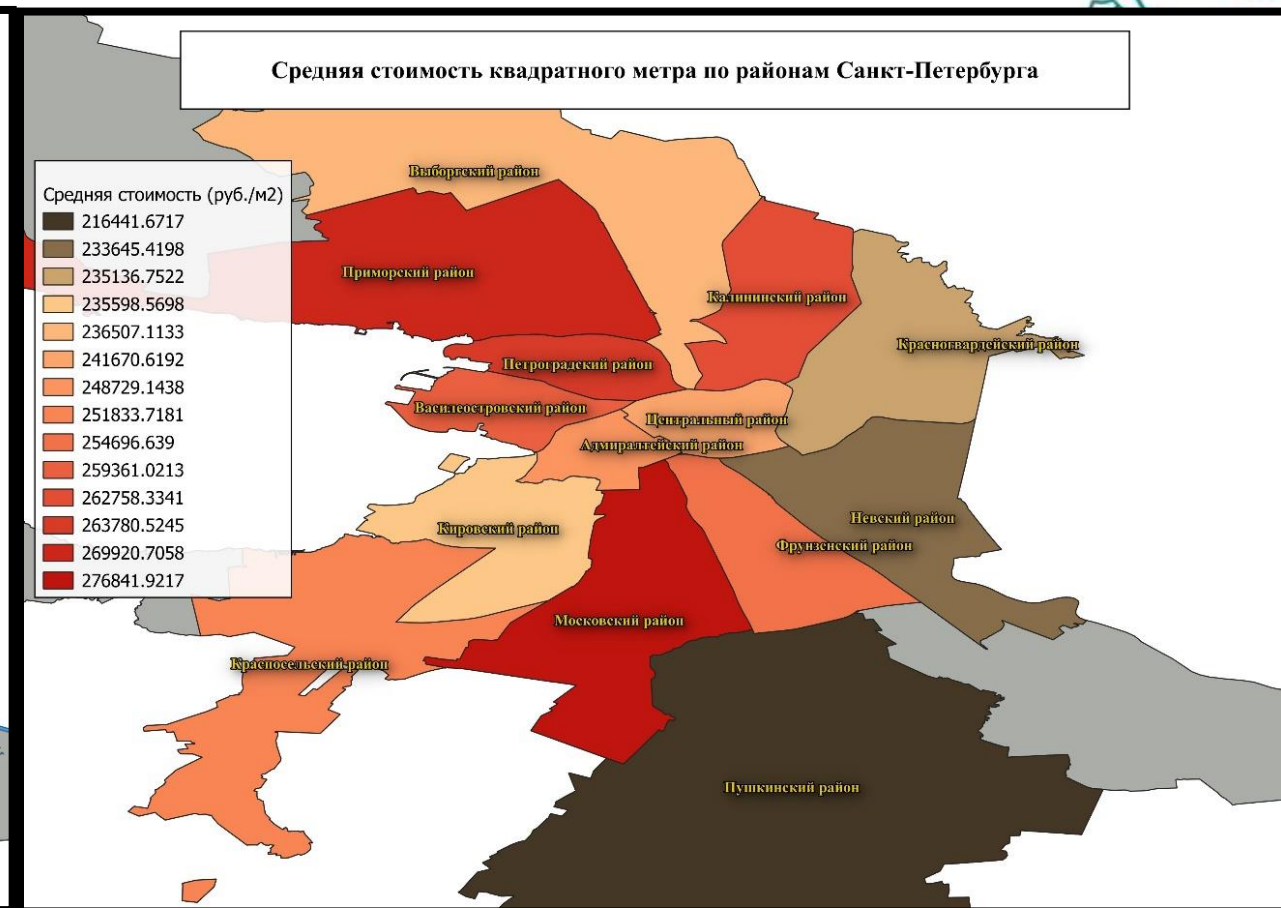
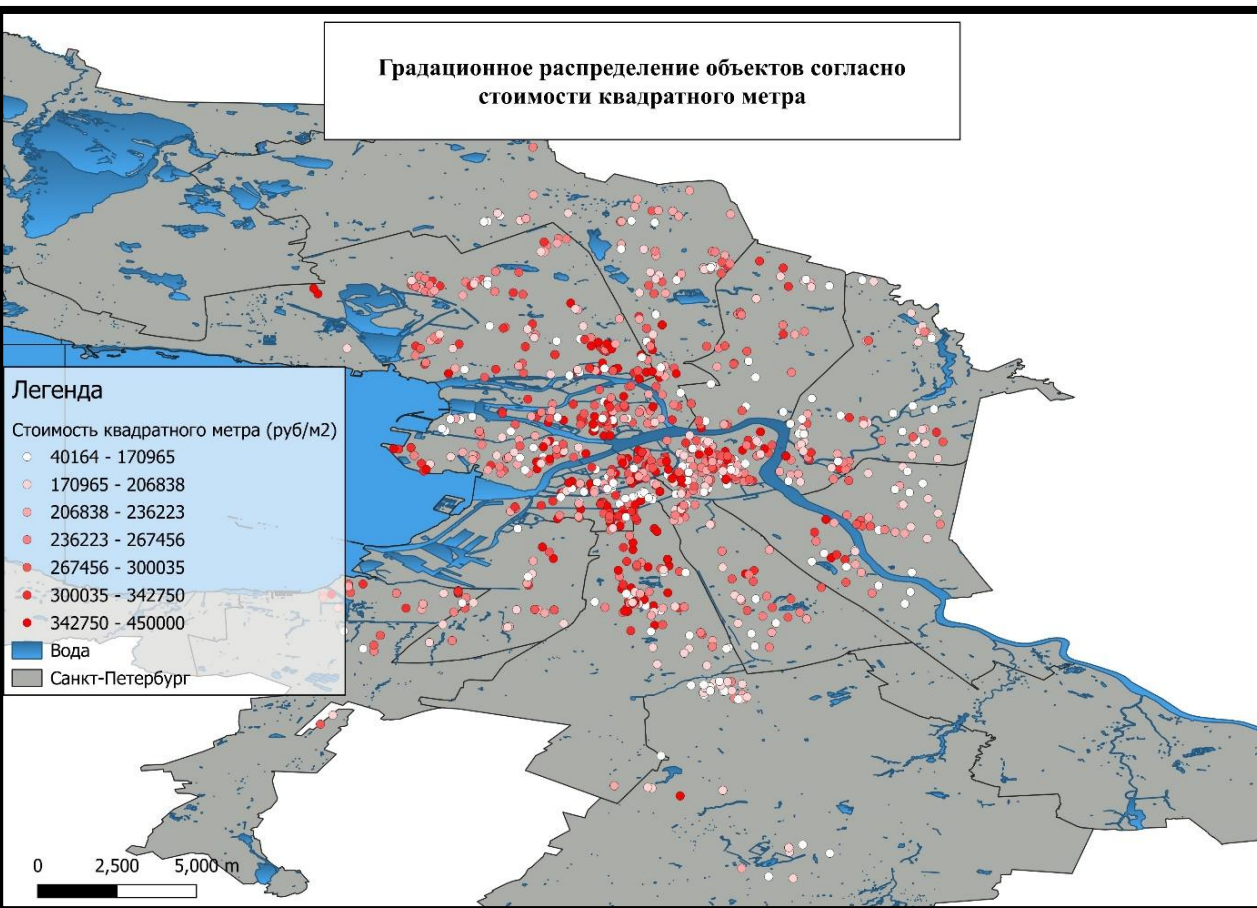
Столбцы после обработки данных и расчета новых признаков:

Технические факторы объекта	Географические факторы объекты
Стоимость кв. метра (целевой признак)	Координаты (долгота и широта)
Площадь	Расстояние до центра города
Максимальный этаж здания объекта	Количество парковок в радиусе 1 км
Стоимость аренды кв. метра аналогичного объекта в месяц.	Количество станций метро в радиусе (250м, 500м и 1км)
	Средняя стоимость кв. метра по району
	Население по району и в радиусе 1 км



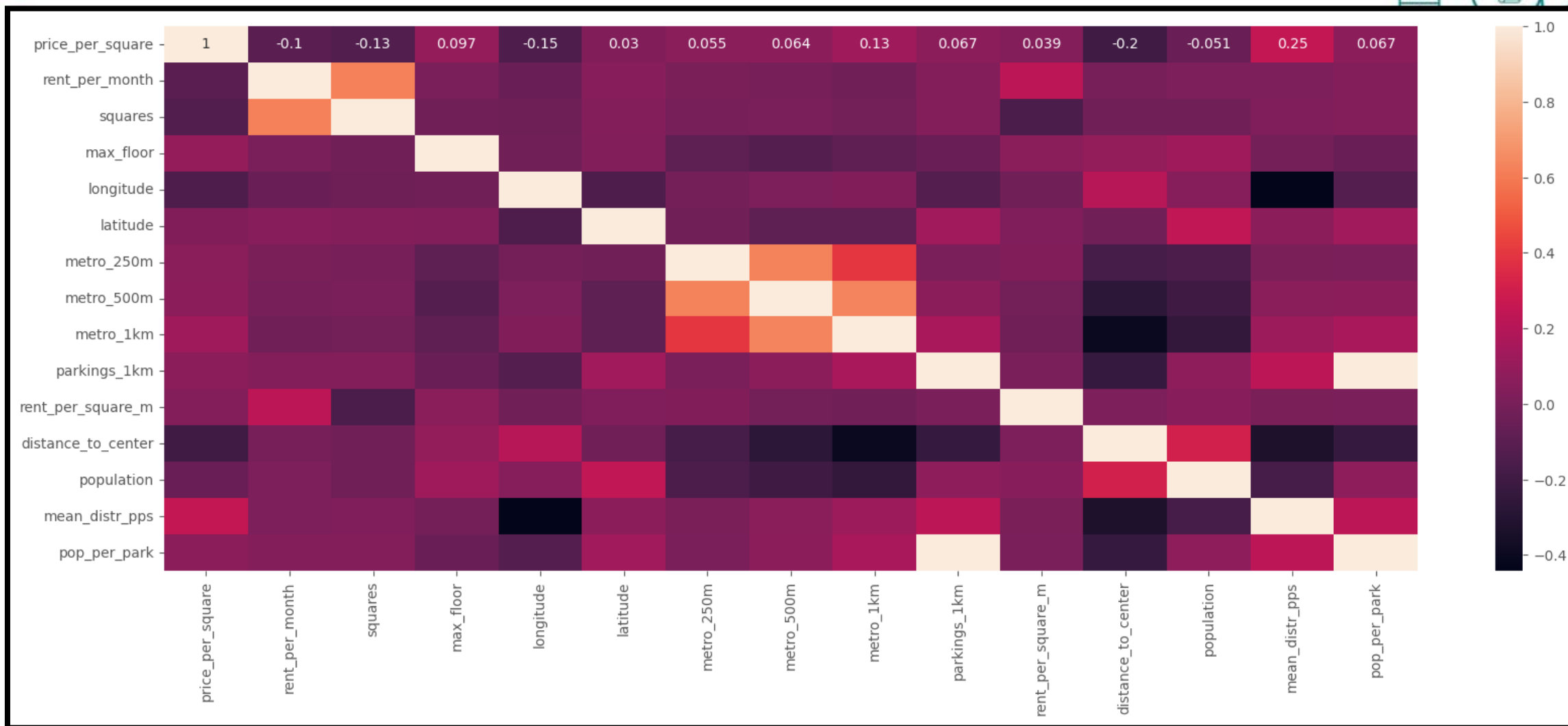
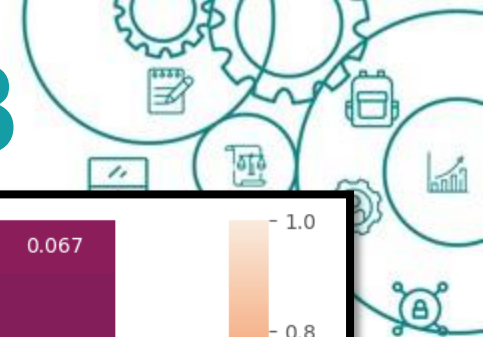


# ГЕОСПАТИЧЕСКИЙ АНАЛИЗ





# КОРРЕЛЯЦИОННЫЙ АНАЛИЗ



Тепловая карта корреляции признаков по Кендаллу



# ПОСТРОЕНИЕ МОДЕЛЕЙ



**Основной принцип** определения наилучшей модели – **минимизация САПО** (среднее абсолютное процентное отклонение).

## Используемые модели:

- Линейная регрессия;
- Лассо;
- Эластичная сеть;
- Метод ближайших соседей;
- Регрессор деревьев решений;
- Метод опорных векторов.

## Используемые ансамблевые модели:

- АдаБустинг;
- Градиентный бустинг;
- Случайный лес;
- Ансамбль дополнительных деревьев.

Отношение тестовой выборки к тренировочной: 2 к 8.

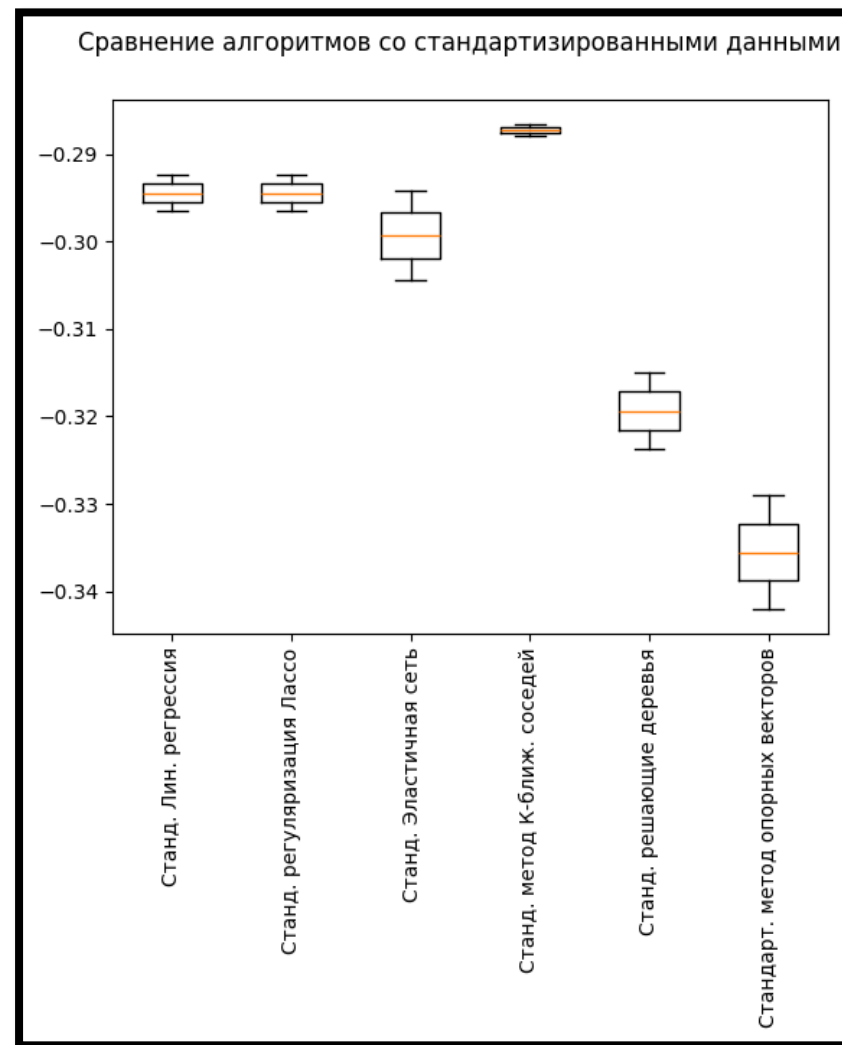
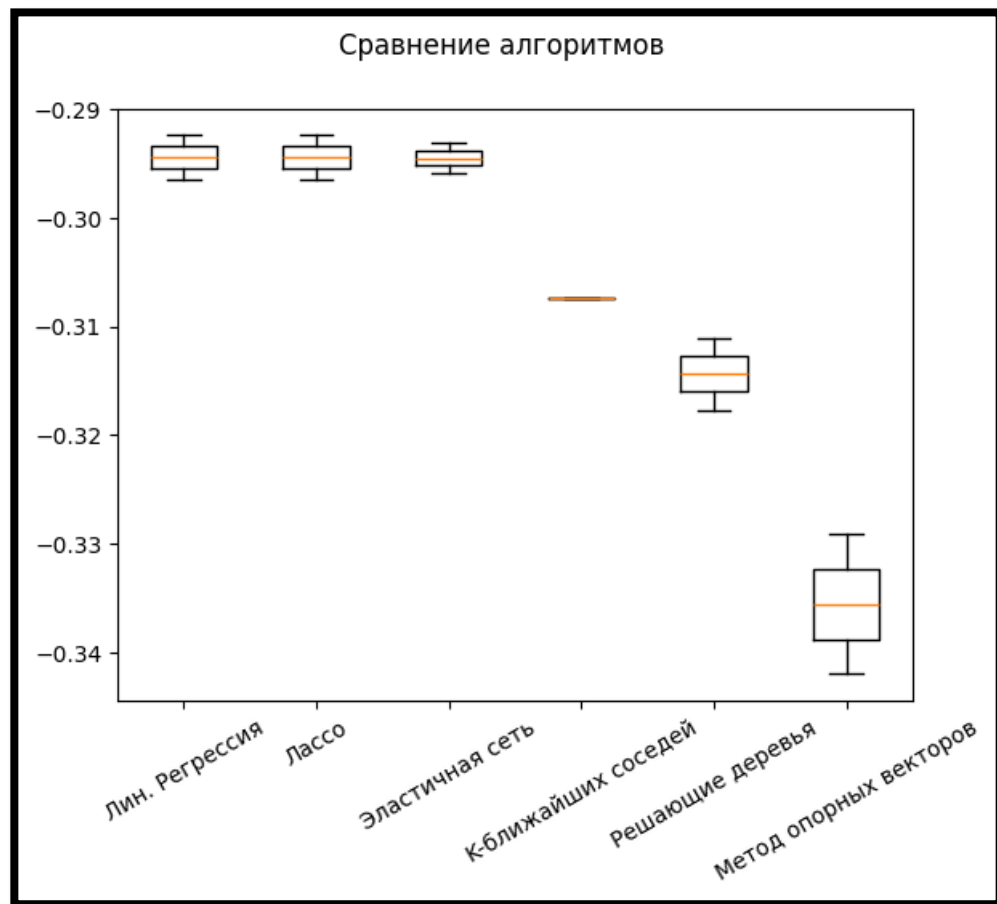
## Столбцы, включенные в процесс построения модели:

Ценообразующие факторы объекта
Координаты (широта и долгота) (°)
Расстояние до центра города (км)
Количество парковок в радиусе 1 км (шт.)
Количество станций метро в радиусе 1км (шт.)
Площадь объекта (м2)
Максимальный этаж здания
Средняя стоимость квадратного метра в районе, в котором находится объект (руб./м2)





# СРАВНЕНИЕ МОДЕЛЕЙ

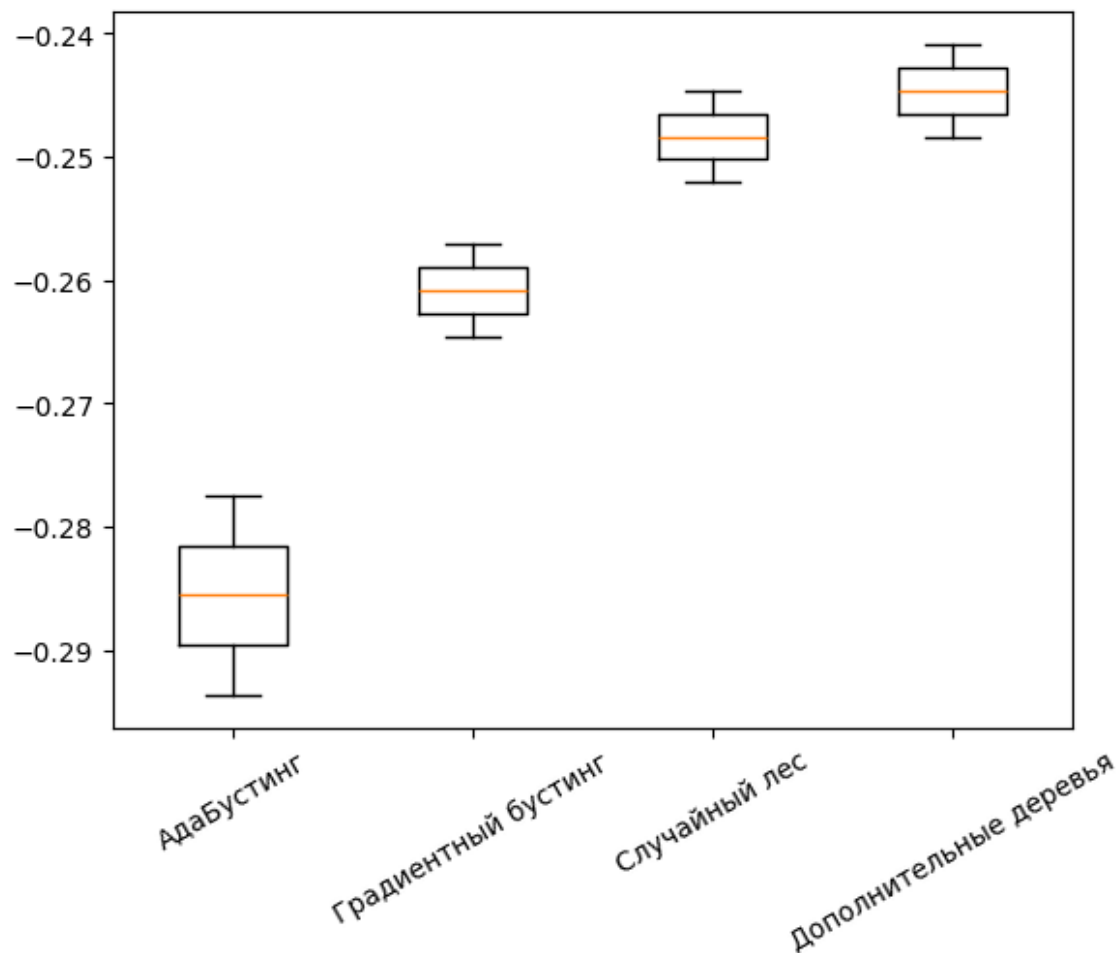




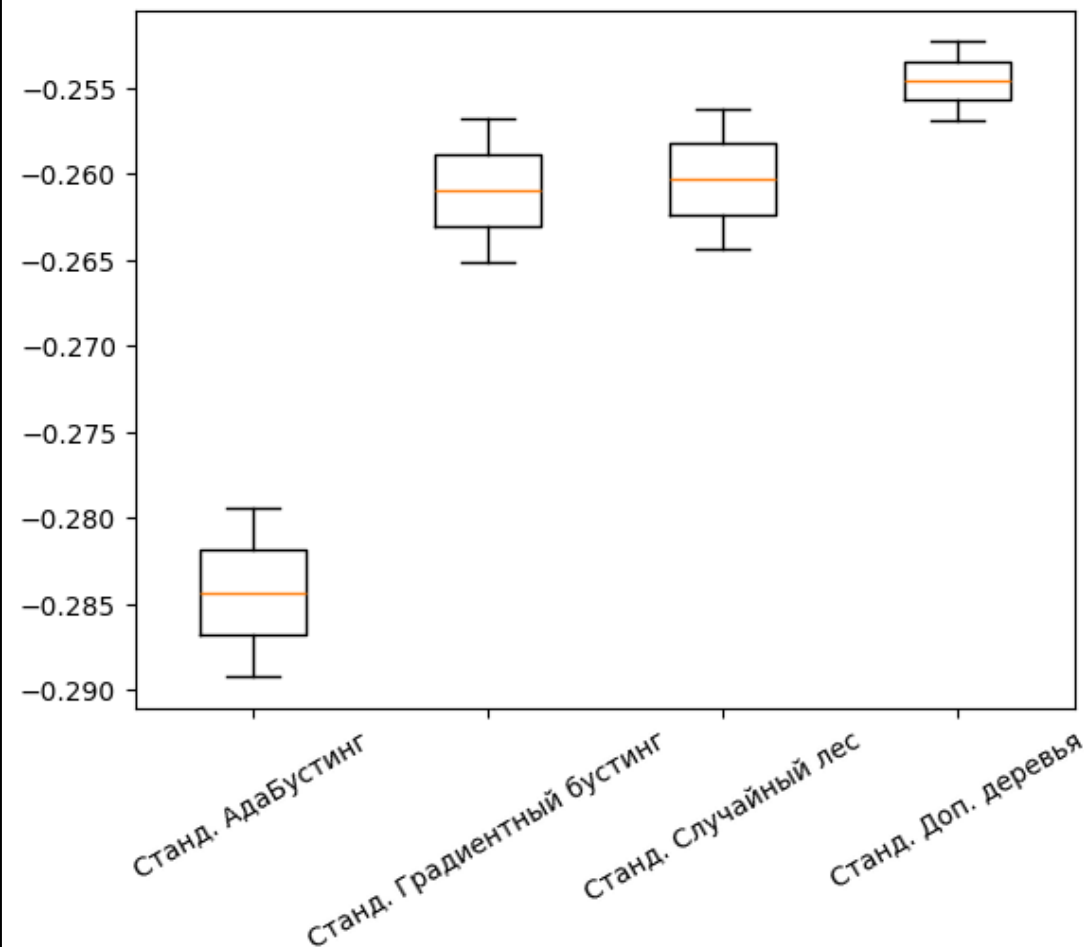
# СРАВНЕНИЕ МОДЕЛЕЙ



Сравнение ансамблевых алгоритмов



Сравнение станд. ансамблевых алгоритмов





# СРАВНЕНИЕ МОДЕЛЕЙ



**Статистические показатели** моделей после подбора их параметров для улучшения точности на тестовой выборке.

## Ансамбль доп. деревьев

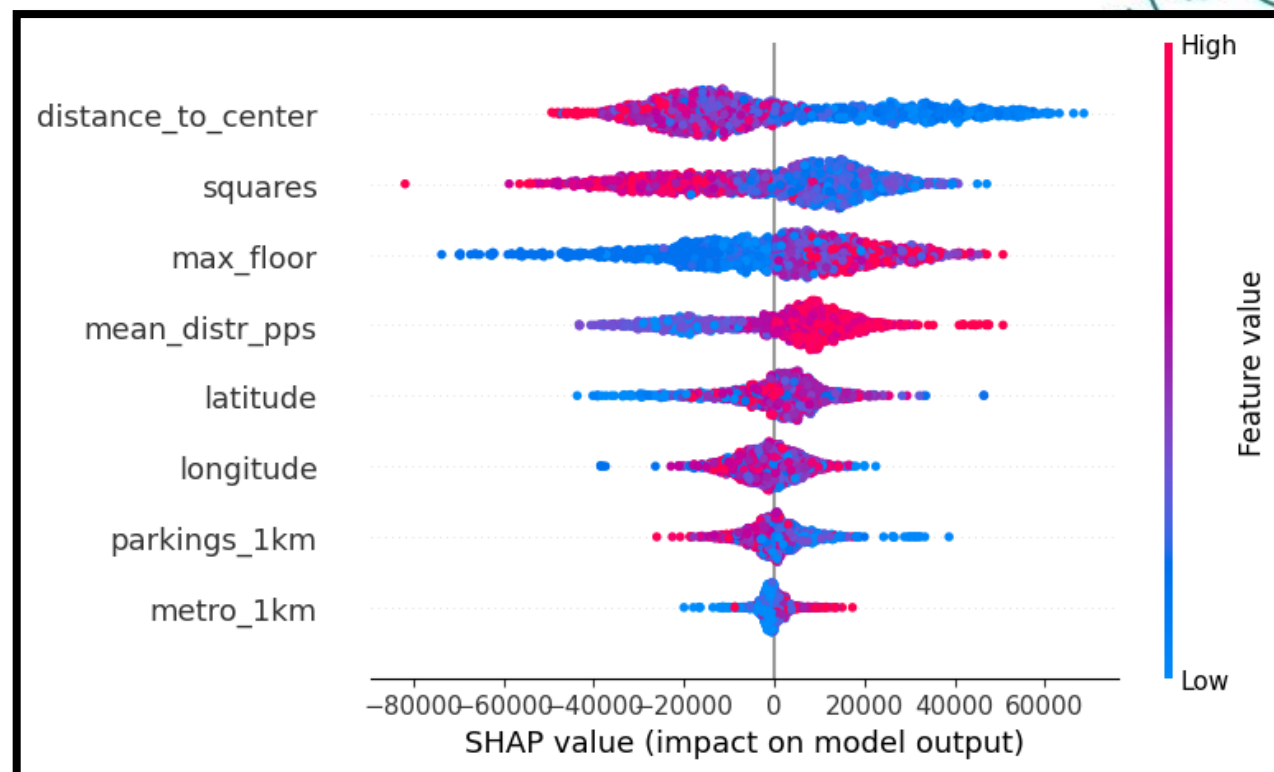
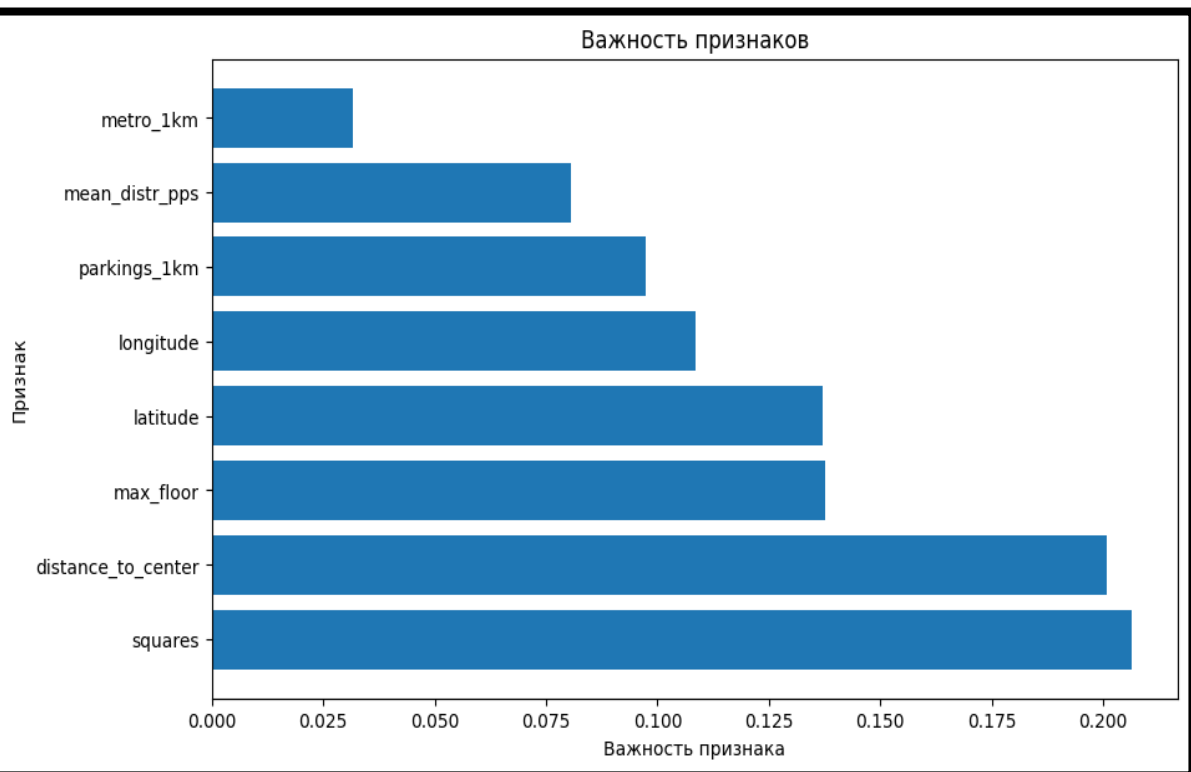
- Коэффициент детерминации = 40.99 (%)
- СКО = 59850.0712
- САО = 43456.8520 (руб.)
- САПО = 21.91 (%)

## Случайный лес

- Коэффициент детерминации = 44.97 (%)
- СКО = 57794.9583
- САО = 43021.9394 (руб.)
- САПО = 21.49 (%)



# ИНТЕРПРЕТАЦИЯ МОДЕЛИ



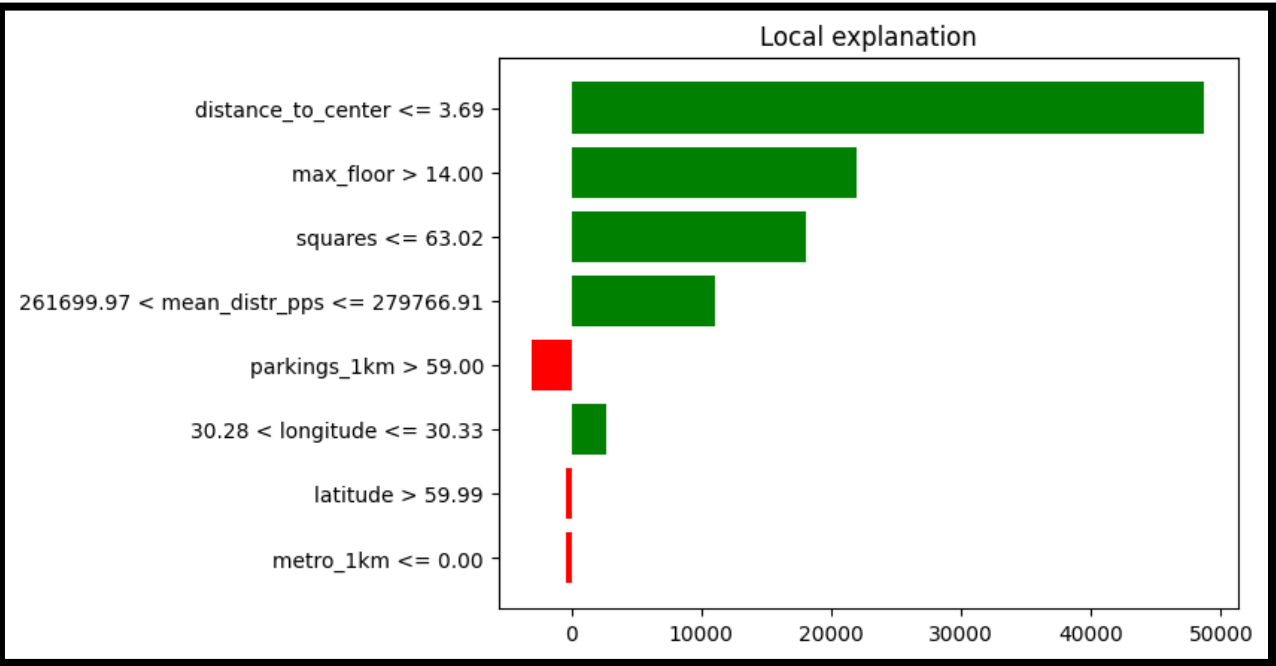
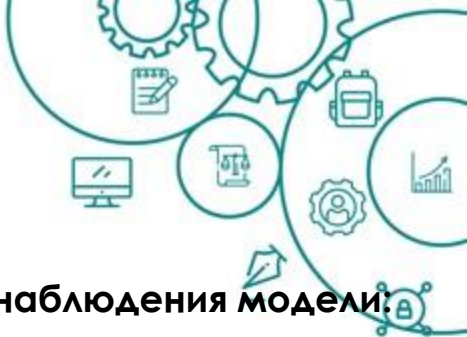
Среднее влияние каждого признака на стандартный прогноз

- модели:
- Площадь = -116.29201861829223
  - Максимальный этаж = 251.0200610573661
  - Количество станций метро в радиусе 1 км = -164.29105783762756
  - Широта = 931.5595059651902
  - Долгота = -1371.0091214121162
  - Количество парковок в радиусе 1 км = 531.3646769899101



# ИНТЕРПРЕТАЦИЯ МОДЕЛИ

## ЛОКАЛЬНАЯ



От факт. значения

Прогноз модели: 339291.04 **[-5708.16]** (руб./м2)

### Признаки интерпретируемого прогноза наблюдения модели:

- **Стоимость кв. метра** = 344999.2 (руб./м2)
- Площадь = 63 (м2)
- Максимальный этаж = 16
- Количество станций метро в радиусе 1 километра = 0
- Координаты широты = 59.987226°
- Координаты долготы = 30.310128°
- Количество парковок в радиусе 1 километра = 84
- Расстояние до центра = 2.54 (км)
- Средняя стоимость кв. метра в районе = 279766.910244 (руб./м2)

Площадь +17076.332813606758 руб.	Максимальный этаж +21076.97972767247 руб.	Долгота +3769.7083065712513 руб.	Кол-во парковок -3923.067180290473 руб.
Расстояние до центра +50359.64958314775 руб.	Ср. стоим. по району +9658.796439006705 руб.	Широта -184.37957627598684 руб.	Кол-во станций метро -1450.3270034932632 руб.





# РЕЗУЛЬТАТЫ РАБОТЫ



- Полнота решений поставленных задач для достижения цели исследования.
- Были собраны данные с основного «онлайн-рынка» недвижимости Санкт-Петербурга.
- Был проведен подробный геоспатический анализ и отбор признаков для прогнозирования стоимости квадратного метра коммерческой торговой недвижимости.
- Построена и выбрана модель с оптимальной точностью.
- Дальнейшая работа: модель следует подготовить к практическому применению.
- Экономическая эффективность определяется средней **ошибкой модели на 29,49%**.



**Спасибо за внимание!**