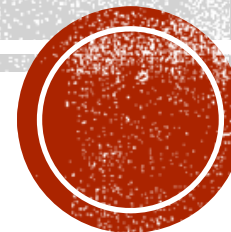


# ИССЛЕДОВАНИЕ ВТОРИЧНОГО РЫНКА ЖИЛОЙ НЕДВИЖИМОСТИ ГОРОДА ВЛАДИВОСТОК



# ВВЕДЕНИЕ

■ *Целью данной работы* является исследование ценообразования на вторичном рынке жилой недвижимости в городе Владивосток, а также кластеризация данного рынка.

■ *Задачи работы:*

1. сбор данных о цене объектов недвижимости и параметрах, характеризующих их;
2. исследование ценообразования на рынке недвижимости;
3. кластеризация объектов;
4. интерпретация полученных результатов.



# ПРИЗНАКОВОЕ ПРОСТРАНСТВО

1558

объектов всего

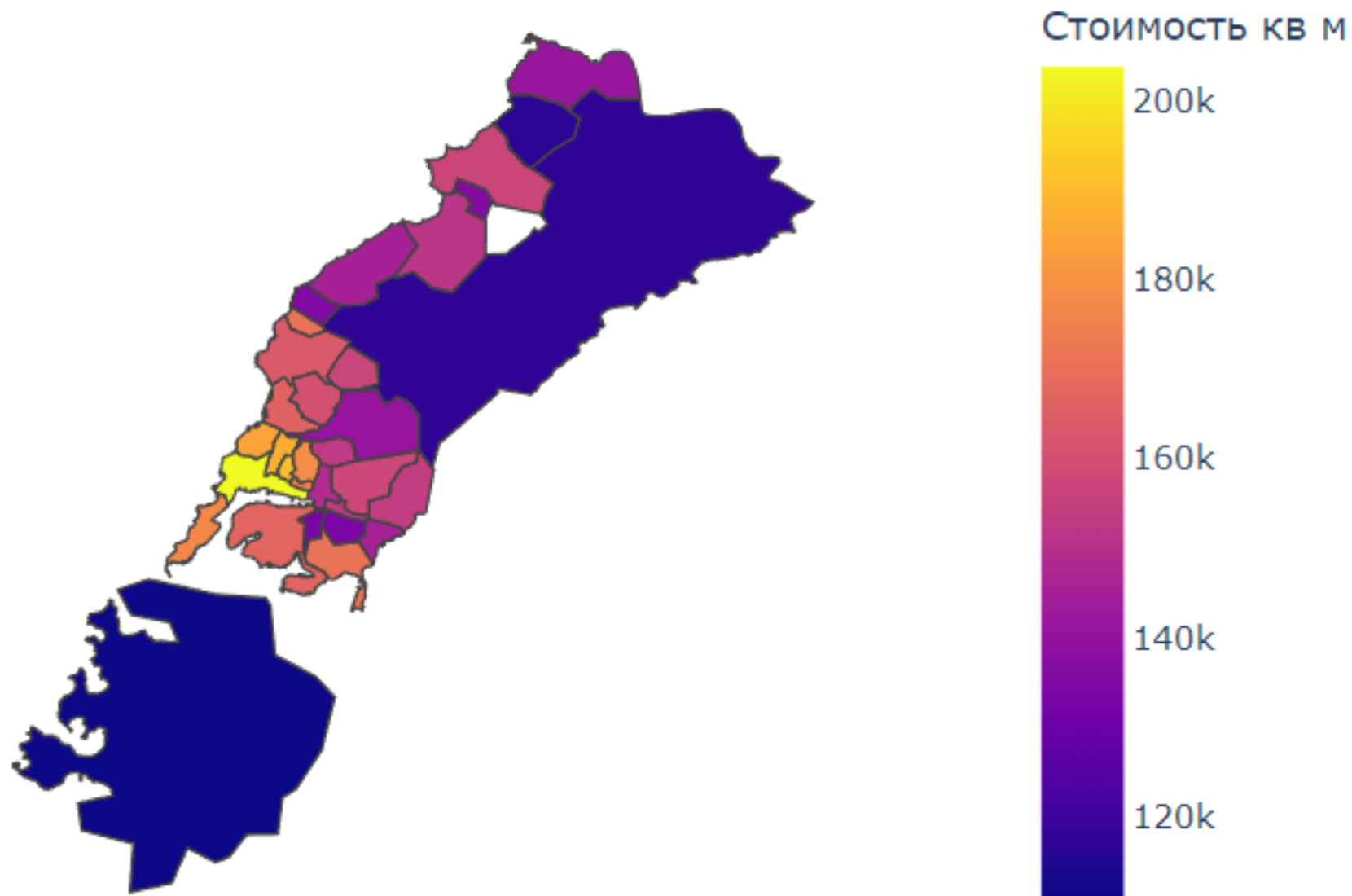
35

признаков

ФРАГМЕНТ ТАБЛИЦЫ С ДАННЫМИ ОБ ОБЪЕКТАХ НЕДВИЖИМОСТИ

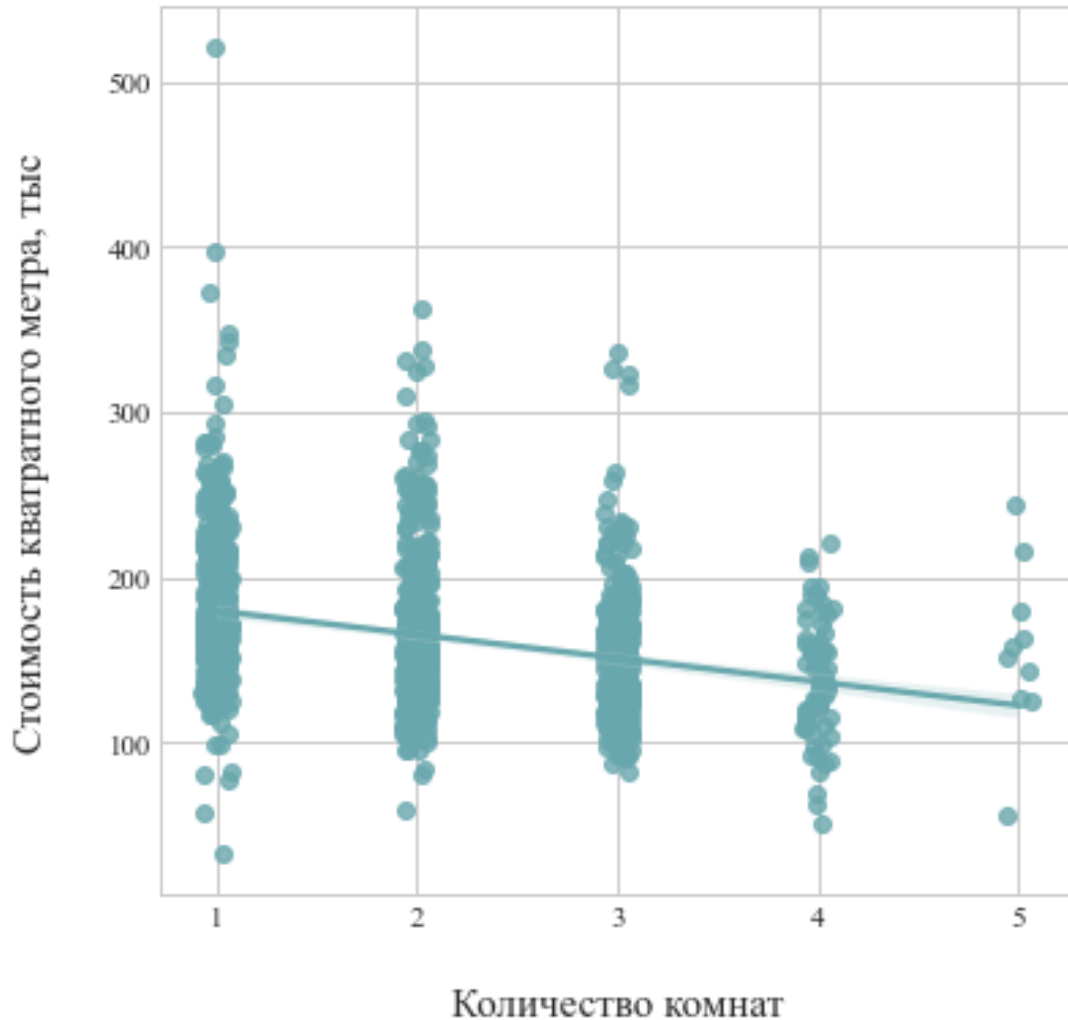
price	rooms	area	floor	region	floors_in_house	keep	building_year	wall_id
8 000 000	2	44,4	6	Фрунзенский	9	design	1967	12
3 300 000	1	33,4	1	Фрунзенский	3	cosmetic	1988	12
6 000 000	1	41	17	Ленинский	25	black	2021	64
4 900 000	1	23,1	2	Ленинский	9	good	1973	12
7 000 000	2	48,5	8	Советский	9	cosmetic	1965	12
5 300 000	1	30,8	3	Первореченский	5	cosmetic	1964	11
4 600 000	1	35,1	6	Советский	27	black	2020	64
4 000 000	1	17,4	4	Советский	5	good	1971	11
8 300 000	1	38,7	2	Ленинский	9	cosmetic	1996	11
13 950 000	3	89,6	16	Ленинский	26	good	2018	64

# ТЕПЛОВАЯ КАРТА СТОИМОСТИ КВ М ПО МИКРОРАЙОНАМ ГОРОДА ВЛАДИВОСТОК

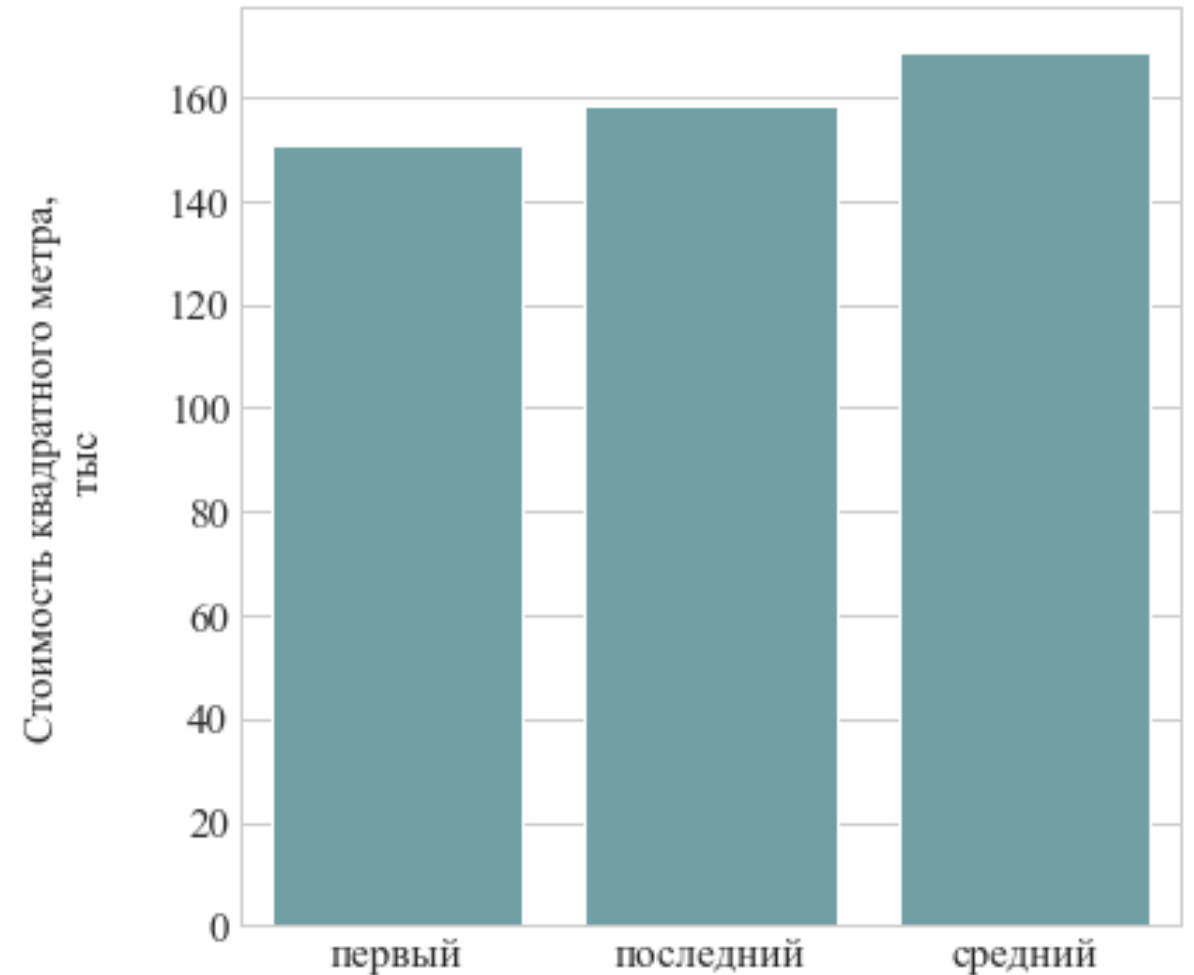


# ЗАВИСИМОСТЬ СТОИМОСТИ КВ М ОТ

## КОЛИЧЕСТВА КОМНАТ

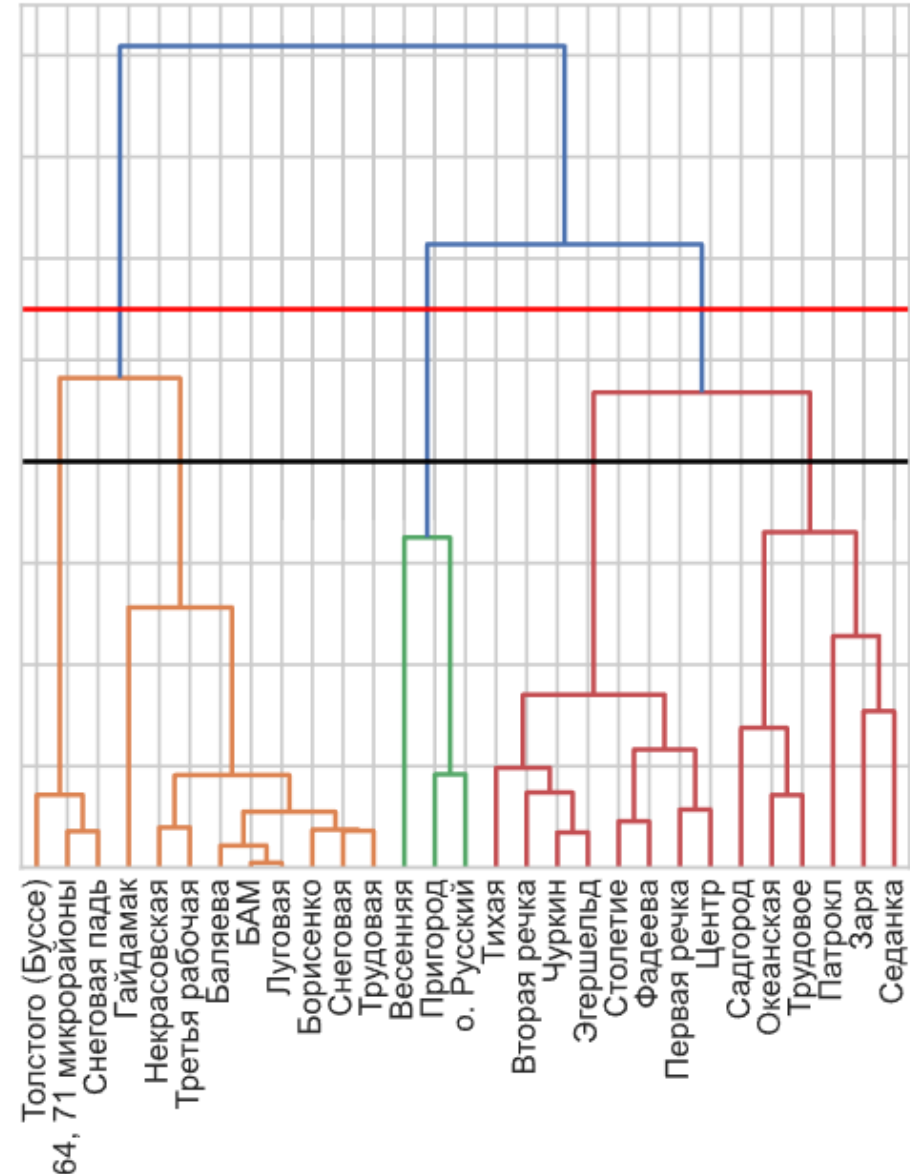


## ЭТАЖА



# КЛАСТЕРИЗАЦИЯ МИКРОРАЙОНОВ

## ИЕРАРХИЧЕСКИЙ МЕТОД

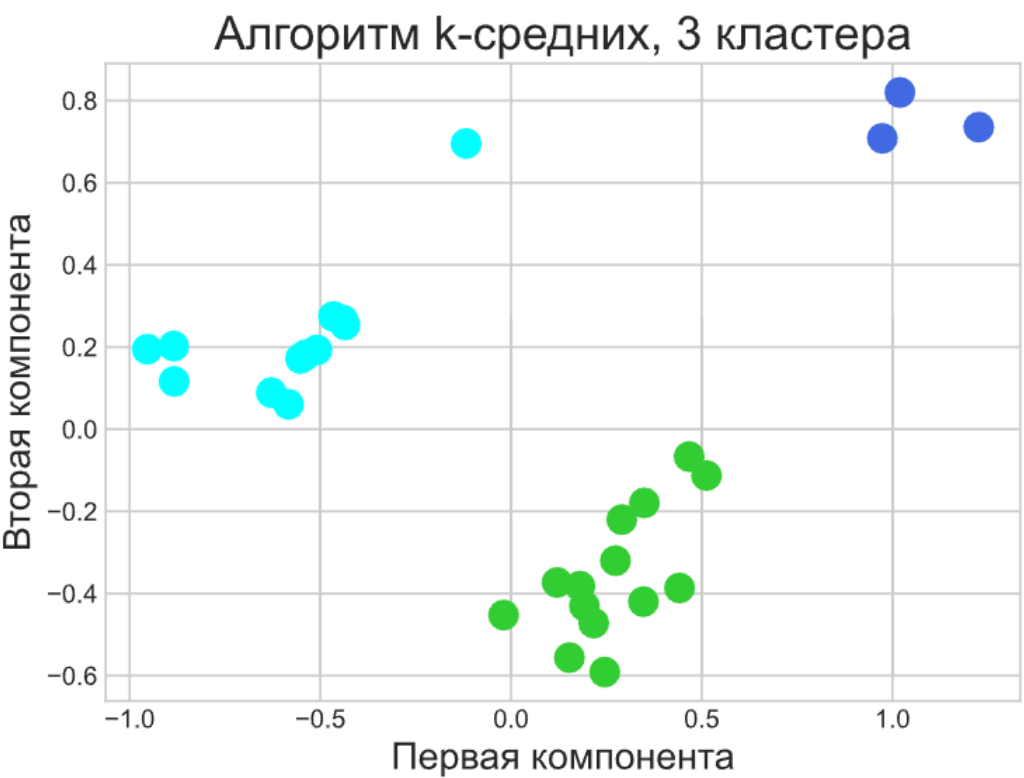


Первый кластер (оранжевый)		Второй (зеленый )	Третий кластер (красный)	
Первый	Второй	Третий	Четвертый	Пятый
Толстого, 64,71, Снеговая падь	Гайдамак, Некрасовская, Третья рабочая, Баляева, БАМ, Луговая, Борисенко, Снеговая, Трудовая	Весенняя, Пригород, о. Русский	Тихая, Вторая речка, Чуркин, Эгершельд, Столетие, Фадеева, Первая речка, Центр	Садгород, Океанская, Трудовое, Патрокл, Заря, Седанка



# КЛАСТЕРИЗАЦИЯ МИКРОРАЙОНОВ

## АЛГОРИТМ К-СРЕДНИХ



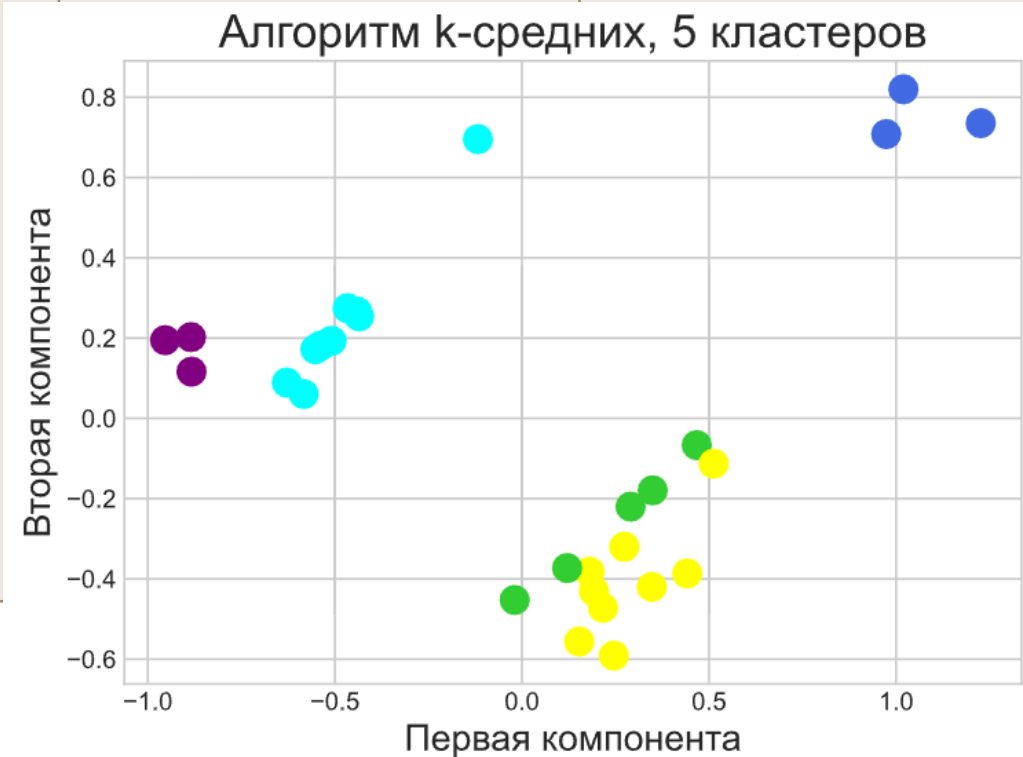
Первый кластер (зеленый кластер)	Второй кластер (голубой кластер)	Третий кластер (синий кластер)
Вторая речка, Заря, Океанская, Патрокл, Первая речка, Садгород, Седанка, Столетие, Тихая, Тудовое, Фадеева, Центр, Чуркин, Эгершельд	64, 71, БАМ, Баляева, Борисенко, Гайдамак, Луговая, Некрасовская, Снеговая, Снеговая падь, Толстого, Третья рабочая, Трудовая	Весенняя, Пригород, о. Русский



# КЛАСТЕРИЗАЦИЯ МИКРОРАЙОНОВ

## АЛГОРИТМ К-СРЕДНИХ

Первый кластер	Второй кластер	Третий кластер	Четвертый кластер	Пятый
Вторая речка, Первая речка, Столетие, Тихая, Трудовое, Фадеева, Центр, Чуркин, Эгершельд	Заря, Океанская, Патрокл, Садгород, Седанка	Весенняя, Пригород, о. Русский	64,71, Снеговая падь, Толстого	БАМ, Баляева, Борисенко, Гайдамак, Луговая, Некрасовская, Снеговая, Третья рабочая, Трудовая



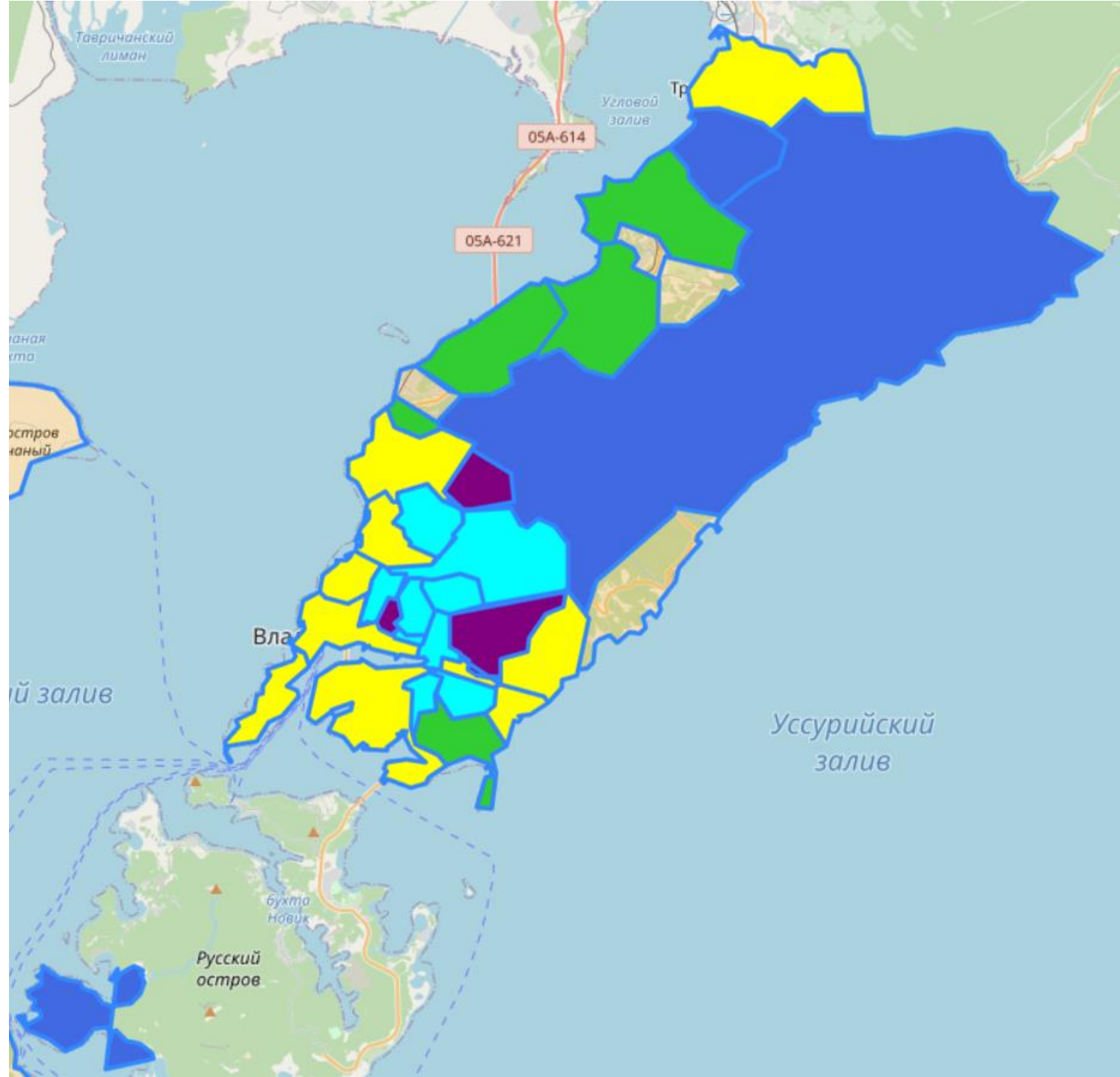


# КЛАСТЕРИЗАЦИЯ МИКРОРАЙОНОВ

## АЛГОРИТМ К-СРЕДНИХ

Цвет кластера	Фиолетовый кластер	Голубой кластер	Синий кластер	Зеленый кластер	Желтый кластер	Число объектов
Синий кластер			Весенняя, Пригород, о. Русский			3
Зеленый кластер				Заря, Океанская, Патрокл, Садгород, Седанка	Вторая речка, Первая речка, Столетие, Тихая, Трудовое, Фадеева, Центр, Чуркин, Эгершельд	14
Голубой кластер	64,71, Снеговая падь, Толстого	БАМ, Баляева, Борисенко, Гайдамак, Луговая, Некрасовская, Снеговая, Третья рабочая, Трудовая				12
Число объектов	3	9	3	5	9	29

# ВИЗУАЛИЗАЦИЯ КЛАСЕТРОВ

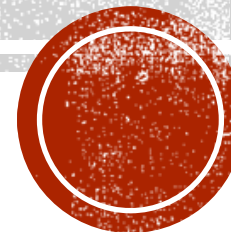


# ВЫВОД

- Самый дорогой квадратный метр у квартир маленьких по площади, в основном однокомнатных, находящихся не на первом и последнем этажах, в Ленинском и Фрунзенских районах. Самый дешевый квадратный метр у больших квартир, квартир на первом этаже и удаленных от центра города, то есть в Советском, Первомайском и Первореченских районах.
- Весь рынок вторичной жилой недвижимости города Владивосток можно разделить на три кластера.
- Город Владивосток можно разделить на 5 географических кластеров.



# МОДЕЛИ ЦЕНООБРАЗОВАНИЯ ЖИЛОЙ НЕДВИЖИМОСТИ



# ВВЕДЕНИЕ

- *Целью данной работы* является поиск модели, которая будет лучше всего предсказывать цену квартире на основе ее характеристик.

- *Задачи работы:*

1. выбор независимых переменных, характеризующих объект вторичной жилой недвижимости;
2. изучение регрессионных моделей машинного обучения;
3. поиск оптимальной модели;
4. выбор переменных оказывающих наибольшее влияние на прогностическую способность моделей.



# ОЦЕНКА РЕГРЕССИОННЫХ МОДЕЛЕЙ

## Коэффициент детерминации

$$R^2 = 1 - \frac{\sum_{i=1}^n (a(x_i) - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

## Средняя абсолютная ошибка

$$\text{MAPE} = 100\% * \frac{1}{n} \sum_{i=1}^n \frac{|a(x_i) - y_i|}{y_i}$$

## Медианная абсолютная ошибка

$$\text{MedAPE} = 100\% * \text{median}\left(\frac{|a(x_i) - y_i|}{y_i}\right)$$



# ЛИНЕЙНАЯ РЕГРЕССИЯ

$$R^2 = 0,806$$

$$\text{MAPE} = 14,94\%$$

$$\text{MedAPE} = 11,61\%$$

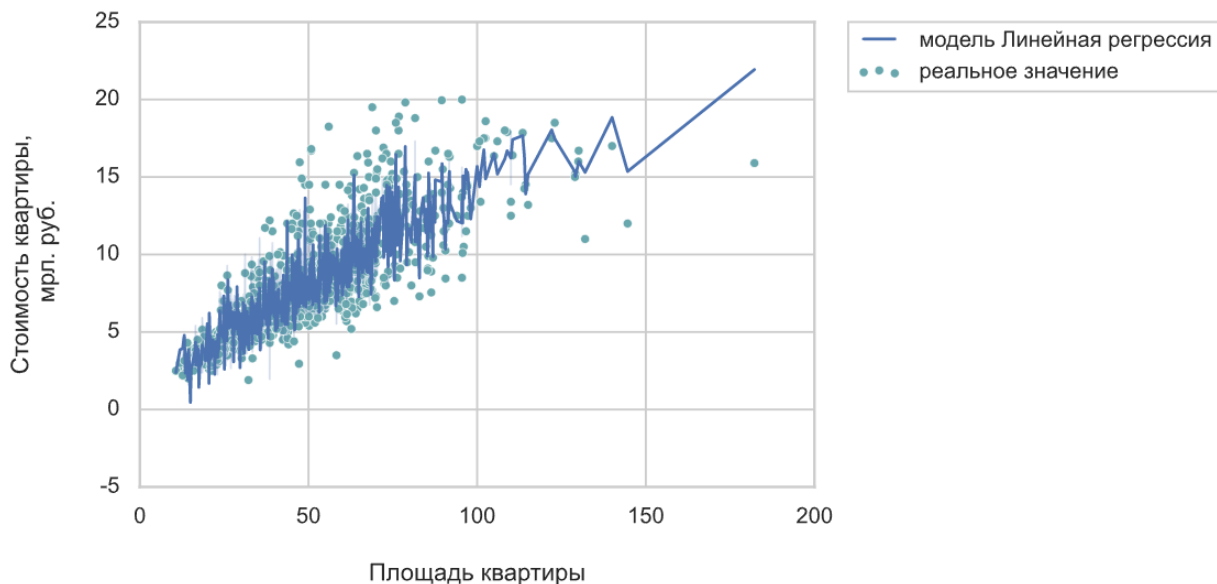


Рисунок 3 – Модель линейная регрессия

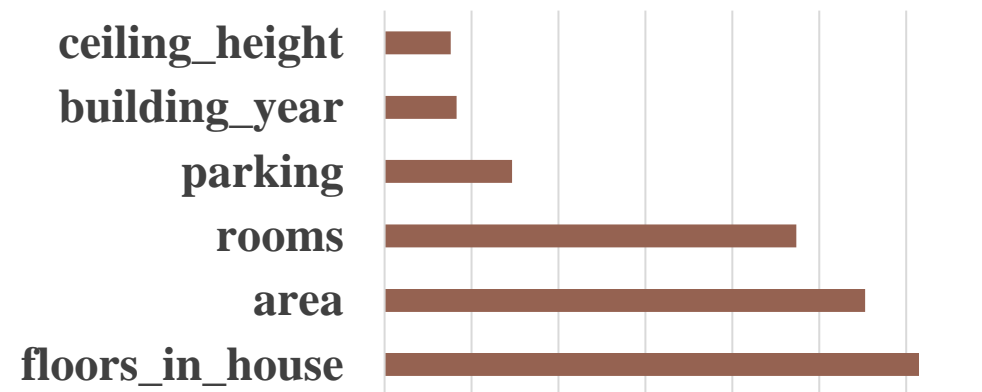


Рисунок 1 - Коэффициенты линейной регрессии

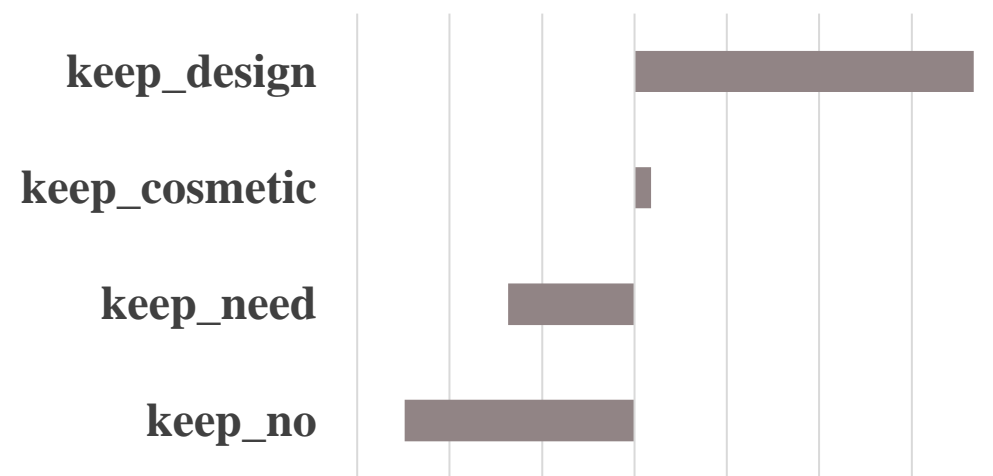


Рисунок 2 - Коэффициенты линейной регрессии, указывающие на тип ремонта



# ДЕРЕВО РЕШЕНИЙ

$$R^2 = 0,76$$

$$\text{MAPE} = 15,24\%$$

$$\text{MedAPE} = 11,98\%$$

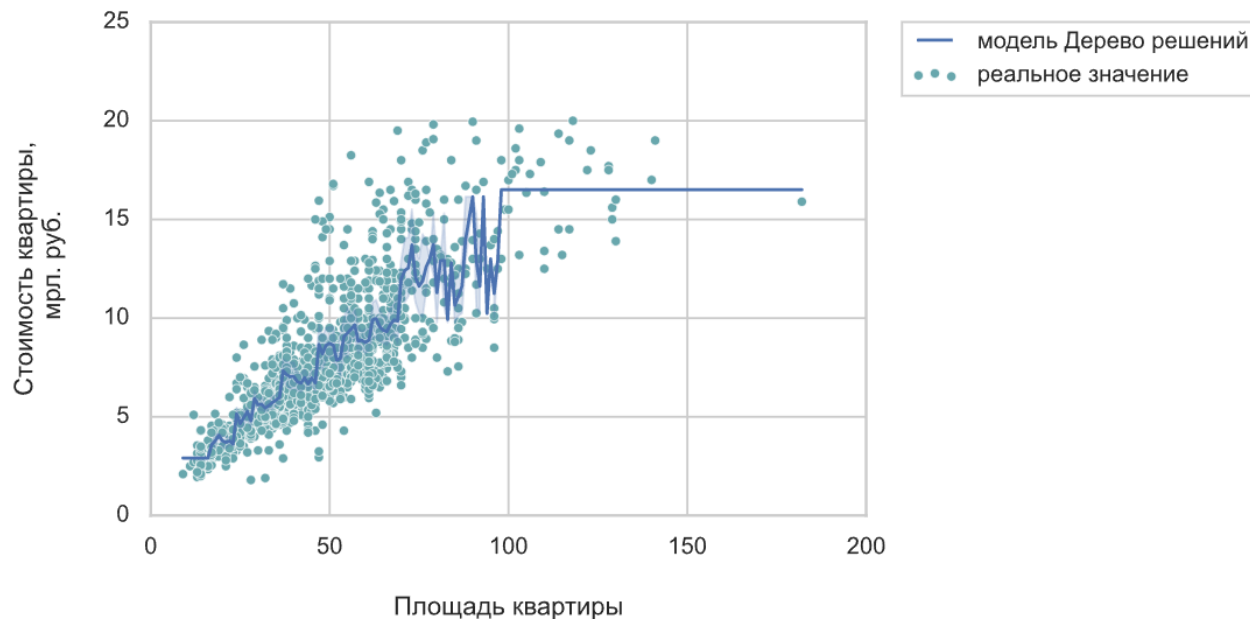


Рисунок 4 – Модель дерево решений

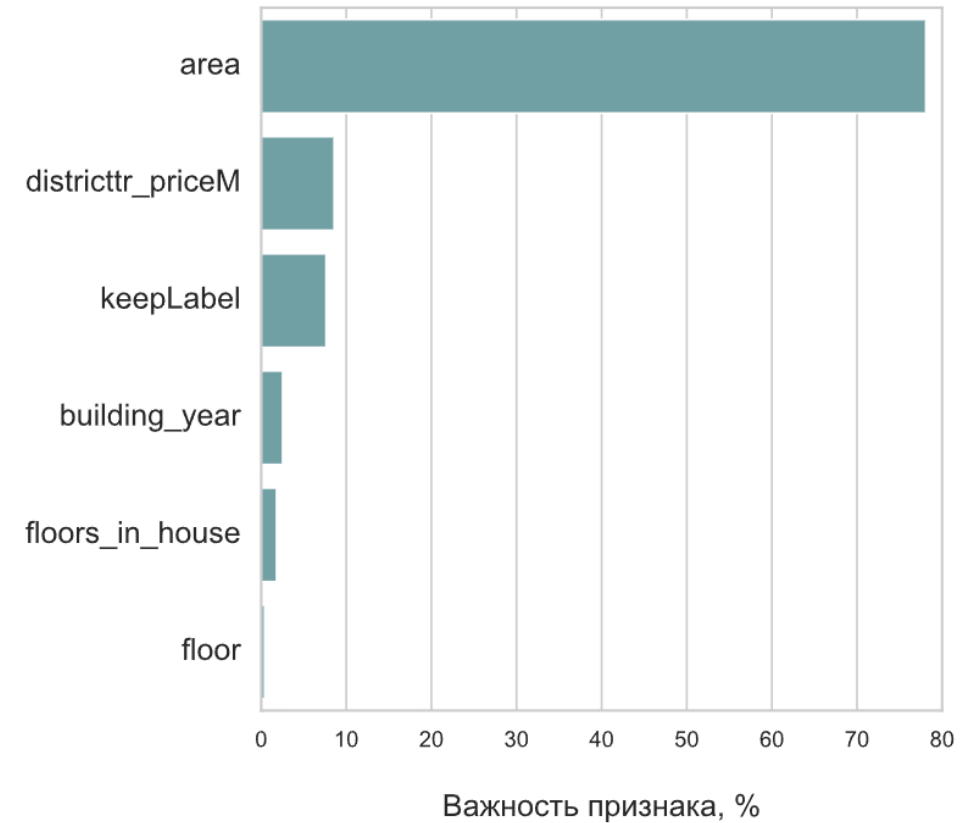


Рисунок 5 - Важность признаков в модели дерево решений





# СЛУЧАЙНЫЙ ЛЕС

$$R^2 = 0,839$$

$$\text{MAPE} = 11,87\%$$

$$\text{MedAPE} = 8,76\%$$

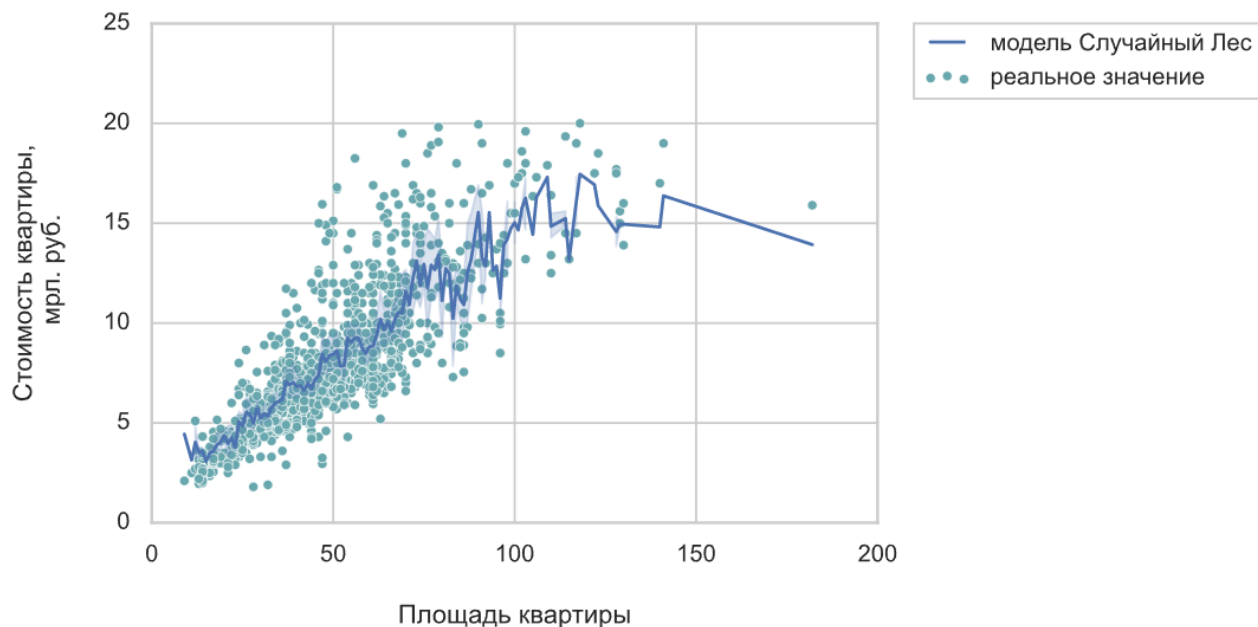


Рисунок 6 – Модель случайный лес

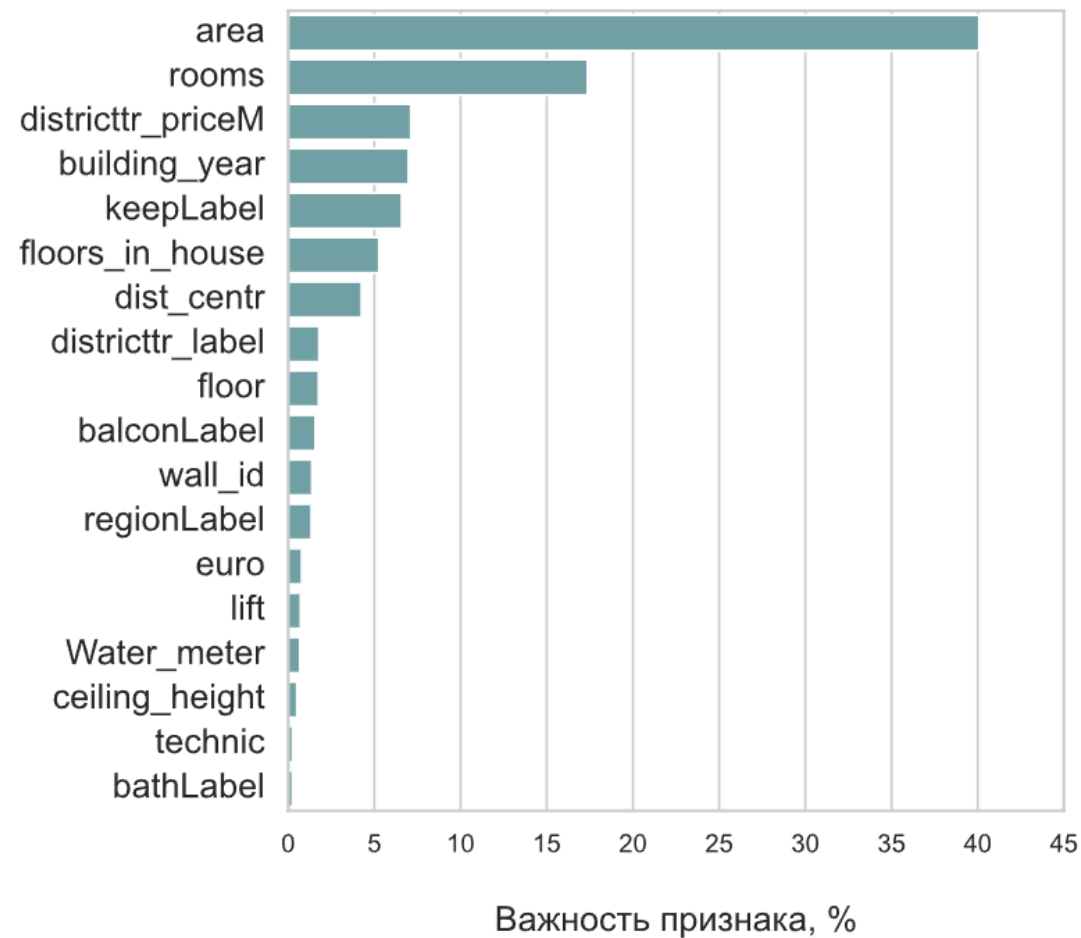


Рисунок 7 - Важность признаков в модели случайный лес



# ГРАДИЕНТНЫЙ БУСТИНГ

$R^2 = 0,855$

MAPE = 12,27%

MedAPE = 9,68%

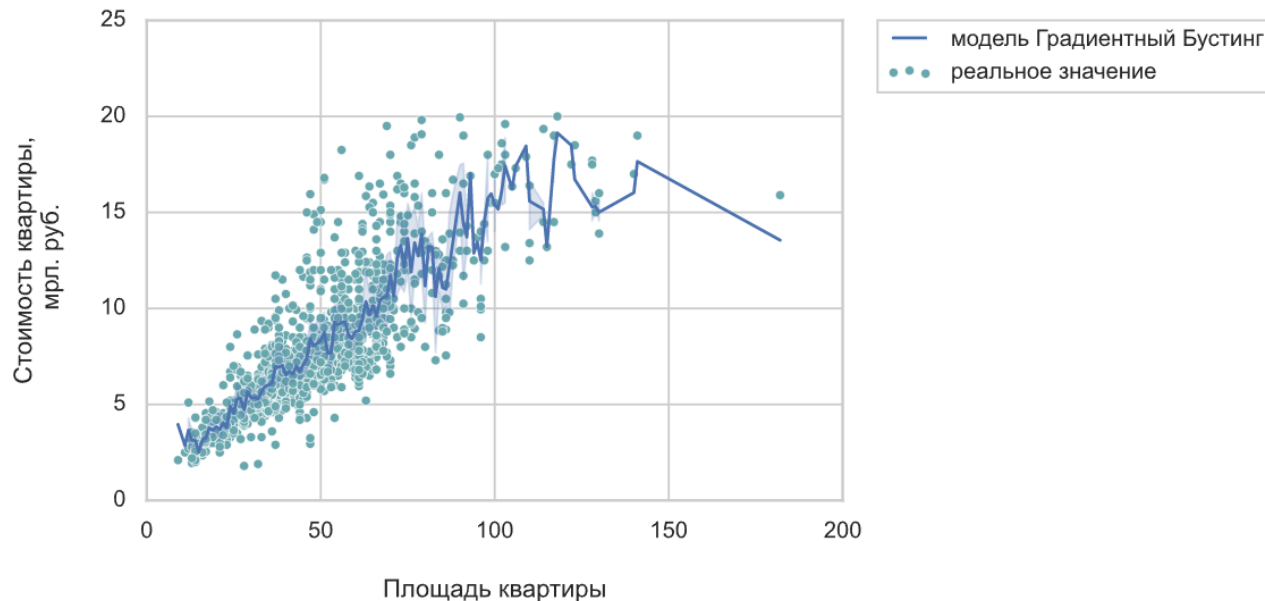


Рисунок 8 – Модель градиентный бустинг

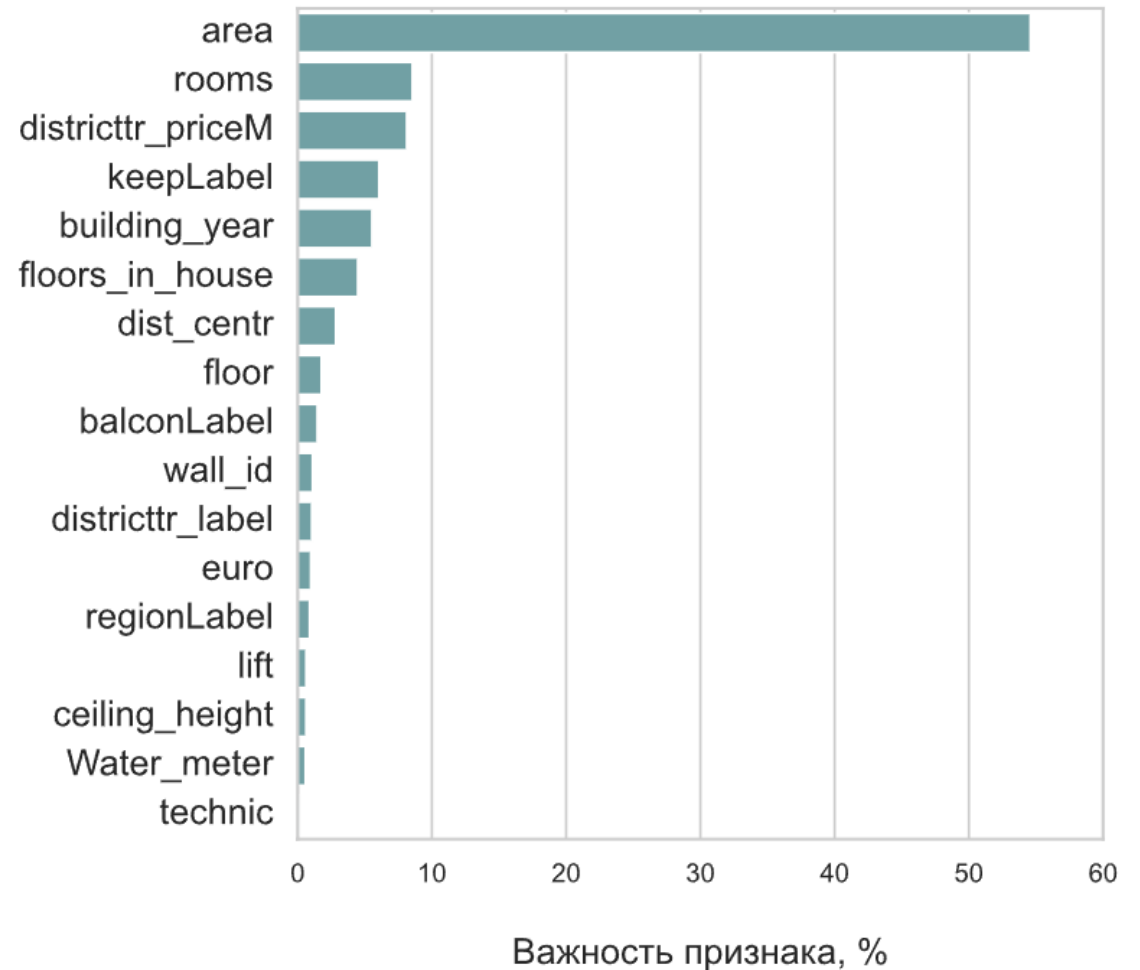


Рисунок 9 - Важность признаков в модели градиентный бустинг



# CATBOOST

$$R^2 = 0,864$$

$$\text{MAPE} = 11,33\%$$

$$\text{MedAPE} = 8,67\%$$

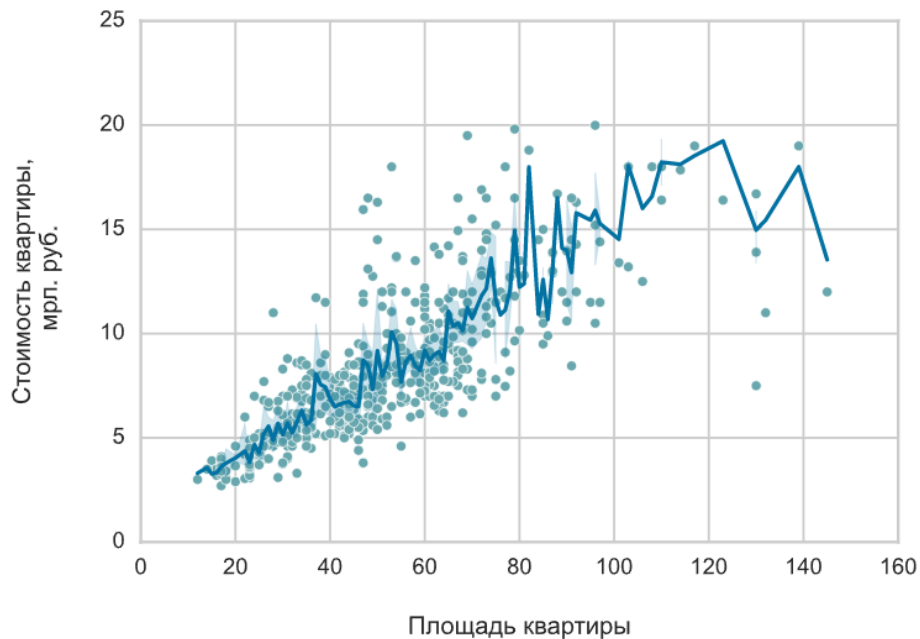


Рисунок 10 – Модель CatBoost

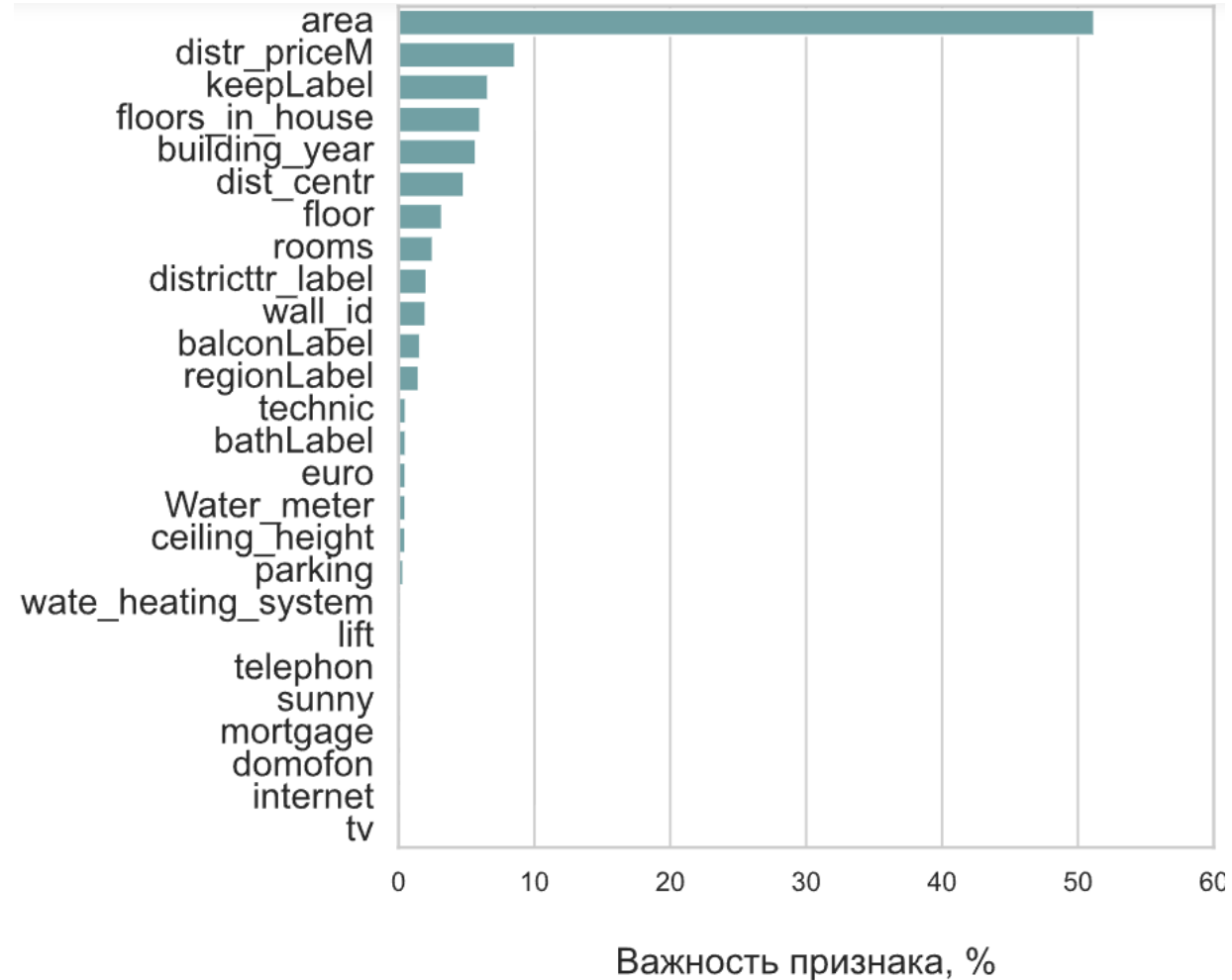


Рисунок 11 - Важность признаков в модели CatBoost



# СРАВНЕНИЕ МОДЕЛЕЙ

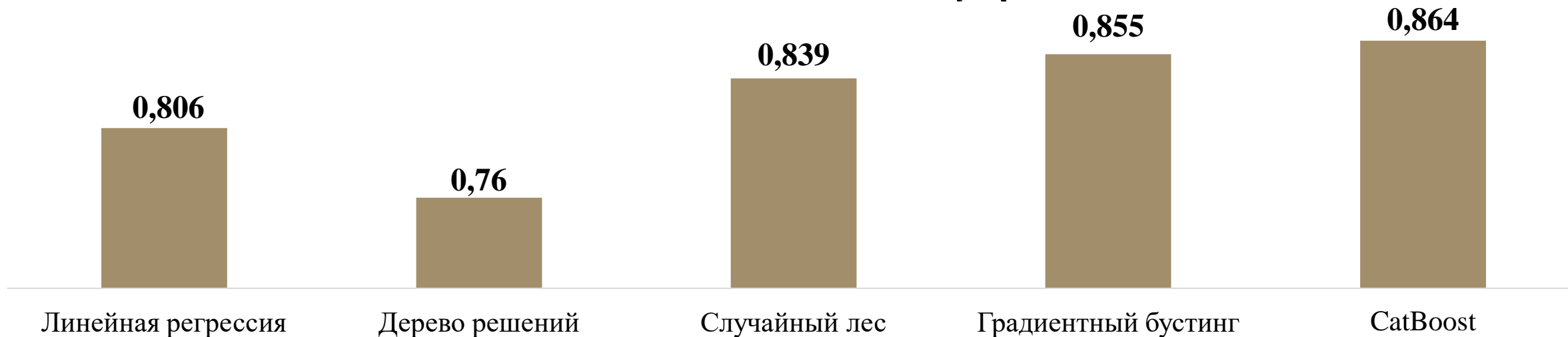


Рисунок 12 - Коэффициенты детерминации рассмотренных моделей

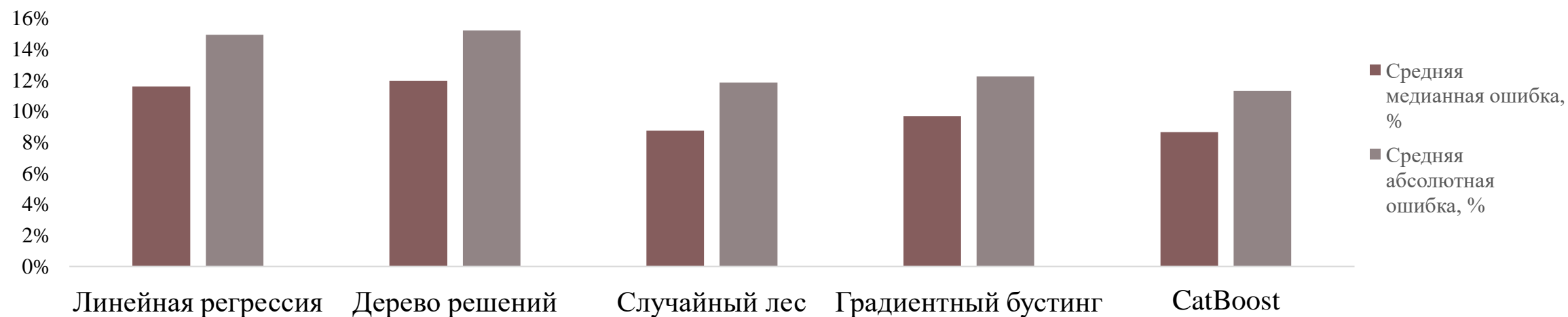


Рисунок 13 - Средние абсолютная и медианная ошибки рассмотренных моделей



# ВЫВОД

- В ходе данной работы были изучены модели машинного обучения и применены для предсказания цены объектов рынка жилой вторичной недвижимости города Владивосток. Всего были рассмотрены 5 моделей.
- Наилучшей обобщающей способностью обладает метод машинного обучений CatBoost. Он способен объяснить 86,4% изменчивости цены и в среднем он ошибается в 8,67% случаев.
- Худший прогноз показывает модель дерево решений, способная предсказать только 76% изменчивости цены объектов и в среднем ошибается в 12% случаев. Данная модель плохо справляется с объектами, площадь которых более 100 м<sup>2</sup>. Какая бы площадь ни была сверх 100 м<sup>2</sup> модель будет предсказывать одно и то же значение, среднее значение цены по таким квартирам на обучающей выборке.
- Также в ходе данного исследования на основе анализа значимости независимых переменных во всех рассмотренных моделях было выявлено, что наиболее влиятельными переменными являются площадь объекта, количество комнат, средняя цена по микрорайону, тип ремонта, год постройки и его этажность, а также расстояние до центра города.

