Федеральное государственное образовательное учреждение

высшего образования

Санкт-Петербургский государственный университет

Институт «Высшая школа менеджмента»

**«Анализ рынка загородной недвижимости в Ленинградской области (на основе данных компании Циан) с помощью методов машинного обучения»**

Студента 4 курса программы бакалавриата по направлению «информационный менеджмент»

**Игнатенка Павла Дмитриевича**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Научный руководитель:

**д.т.н. Гаврилова Татьяна Альбертовна**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Санкт-Петербург

2022

ЗАЯВЛЕНИЕ О САМОСТОЯТЕЛЬНОМ ВЫПОЛНЕНИИ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Я, Игнатенок Павел Дмитриевич, студент 4 курса направления «Менеджмент» (профиль подготовки – Информационный менеджмент) заявляю, что в моей выпускной квалификационной работе на тему «Анализ рынка загородной недвижимости в Ленинградской области (на основе данных компании Циан) с помощью методов машинного обучения», представленной в службу обеспечения программ бакалавриата для публичной защиты, не содержится элементов плагиата. Все прямые заимствования из печатных и электронных источников, а также из защищённых ранее курсовых и выпускных квалификационных работ, кандидатских и докторских диссертаций имеют соответствующие ссылки.

Мне известно содержание п. 9.7.1 Правил обучения по основным образовательным программам высшего и среднего профессионального образования в СПбГУ о том, что «ВКР выполняется индивидуально каждым студентом под руководством назначенного ему научного руководителя», и п. 51 Устава федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Санкт-Петербургский государственный университет» о том, что «студент подлежит отчислению из Санкт-Петербургского университета за представление курсовой или выпускной квалификационной работы, выполненной другим лицом (лицами)».

Оглавление

[Введение 4](#_Toc105010537)

[Глава 1. Рынок загородной недвижимости Ленинградской области 6](#_Toc105010538)

[1.1. Ленинградская область 6](#_Toc105010539)

[1.2. Рынок загородной недвижимости Ленинградской области 6](#_Toc105010540)

[1.3. Компания ЦИАН 7](#_Toc105010541)

[1.4. Признаки объектов недвижимости 9](#_Toc105010542)

[1.5. Гипотезы и вопросы 17](#_Toc105010543)

[Выводы 18](#_Toc105010544)

[Глава 2. Методы машинного обучения для анализа рынка недвижимости 19](#_Toc105010545)

[2.1. Метод сбора данных 19](#_Toc105010546)

[2.2. Алгоритмы выявления выбросов 19](#_Toc105010547)

[2.3. Методы анализа данных 20](#_Toc105010548)

[2.4. Примеры использования методов машинного обучения для анализа рынка недвижимости 21](#_Toc105010549)

[2.5. Методология 24](#_Toc105010550)

[Выводы 25](#_Toc105010551)

[Глава 3. Исследование рынка с помощью методов машинного обучения 26](#_Toc105010552)

[3.1. Описание наборов данных 26](#_Toc105010553)

[3.2. Исследование набора данных об объектах, выставленных на продажу 26](#_Toc105010554)

[3.3. Исследование набора данных об объектах, сдающихся посуточно 45](#_Toc105010555)

[3.4. Исследование набора данных об объектах, сдающихся помесячно 55](#_Toc105010556)

[Выводы 63](#_Toc105010557)

[Заключение 65](#_Toc105010558)

[Список литературы 67](#_Toc105010559)

[Приложения 70](#_Toc105010560)

# Введение

Поскольку методы машинного обучения становятся все более популярными для анализа данных, с их помощью решается все большее количество задач. Так, рынок загородной недвижимости активно развивается после первых волн пандемии Covid-19. Следовательно, методы машинного обучения можно применить и к изучению этого рынка в том числе.

Таким образом, выпускная квалификационная работа посвящена анализу рынка загородной недвижимости Ленинградской области с помощью методов машинного обучения. В рамках нее будет представлена информация о ситуации на рынке загородной недвижимости региона, об условиях, в которых он формируется, собраны данные о рынке загородной недвижимости региона, проведен анализ с помощью методов описательной статистики, построены модели машинного обучения. Таким образом, будут изучены факторы, влияющие на стоимость объектов недвижимости.

Это исследование может быть использовано для получения информации о том, как оценивать объекты загородной недвижимости в данном регионе в текущей ситуации. Это может быть полезно как для собственников домов и коттеджей, которые бы хотели их продать, так и для агентств недвижимости, для аналитиков рынка, которые сталкиваются со сложностью оценки стоимости домов и коттеджей, стоимости аренды этих домов и коттеджей.

Форматом работы была выбрана исследовательская работа.

Целью работы является построение моделей машинного обучения для облегчения процесса оценки объектов недвижимости. Для достижения цели должны быть решены следующие задачи:

* Сбор данных о рынке загородной недвижимости Ленинградской области.
* Проведение анализа признаков, влияющих на цену объектов загородной недвижимости в этом регионе.
* Построение моделей, с помощью которых цена недвижимости, а также цена ее аренды, может быть определена.

Объектом исследования является рынок загородной недвижимости Ленинградской области, предметом исследования – признаки, влияющие на цену недвижимости в регионе.

Работа состоит из следующих частей:

* Введение. Описание управленческой проблемы, постановка целей, формулирование задач, определение формата работы.
* Глава 1. Рынок загородной недвижимости Ленинградской области. В главе приведено описание ситуации на рынке недвижимости, компании, которые на рынке функционируют. Также проводится анализ признаков объектов недвижимости, описываются их преимущества и недостатки, каким образом те или иные признаки могут влиять на стоимость дома или коттеджа.
* Глава 2. Методы машинного обучения для анализа рынка недвижимости. В главе описываются методы машинного обучения, которые могут быть использованы при анализе рынка недвижимости, рассказывается о предыдущем опыте похожих исследований, формируется методология исследования для этой работы.
* Глава 3. Исследования рынка с помощью методов машинного обучения. В главе исследуются данные об объектах недвижимости, которые выставлены на продажу или сдаются в аренду (как посуточную, так и помесячную). Используются методы описательной статистики и машинного обучения.
* Заключение. Описание результатов исследования, формулирование выводов, полученных во время исследования.
* Список литературы. Перечень использованных источников как для изучения рынка, так и формирование методологии исследования.
* Приложения. Примеры частей моделей, используемых для прогнозирования.

# Глава 1. Рынок загородной недвижимости Ленинградской области

## 1.1. Ленинградская область

Географически исследование в данной работе ограничено Ленинградской областью. Ленинградская область является одним из регионов Российской Федерации. В следующей таблице (табл. 1.1) приведена основная информация об этом субъекте РФ.

**Таблица 1.1.** Основная информация о Ленинградской области [13].

|  |  |
| --- | --- |
| **Параметр** | **Значение** |
| Площадь, квадратных километров | 83 908 (39-е место в РФ) |
| Население, человек | 1 911 586 (27-е место в РФ) |
| ВРП на душу населения, рос. рубли | 603 200 (19-е место в РФ) |
| Часовой пояс | Московское время |

Ленинградская область - один из самых экономически развитых регионов России [8]. Этому способствует развитая транспортная инфраструктура (в частности, морские и речные порты), близость к Санкт-Петербургу, граница с двумя странами Европейского союза - Эстонией и Финляндией, многоотраслевой характер промышленности.

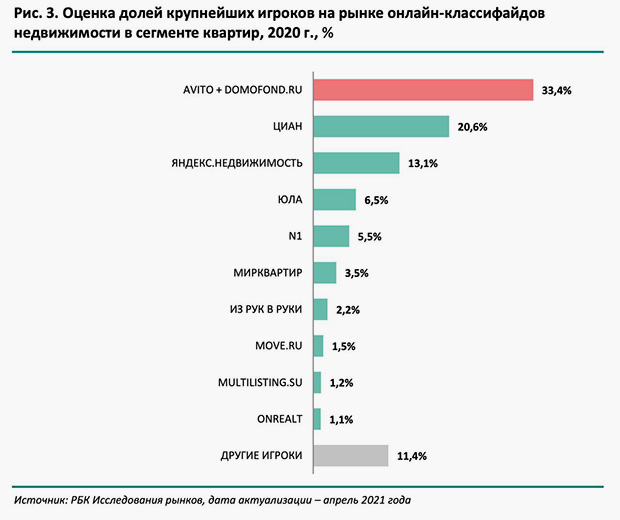
## 1.2. Рынок загородной недвижимости Ленинградской области

Определенные выше факторы также предопределили развитие рынка загородной недвижимости Ленинградской области. Так, по данным компании ЦИАН, рост цен на этом рынке в 2021 году составил 39 процентов, что сравнимо с ростом в Москве (41%) и Московской области (также 39%) [20]. Эти показатели заметно превышают показатели в других регионах страны. Также 77 процентов спроса приходится на готовые дома и дачи. При этом эксперты считают, что этот «бум» связан с прохождением пандемии Covid-19.

В апреле 2022 года на сайте cian.ru было опубликовано 7 862 объявления о продаже объектов загородной недвижимости в Ленинградской области [21]. На сайте avito.ru можно было найти 10 417 объявлений о продаже домов, дач и коттеджей [1]. В это же время на сайте cian.ru было опубликовано 677 объявлений о посуточной аренде домов и коттеджей, 515 объявлений о помесячной аренде [21].

## 1.3. Компания ЦИАН

ЦИАН - компания, которая предоставляет базу о продаже и аренде различных типов недвижимости (жилой, коммерческой и пр.). Иначе говоря, компания владеет онлайн-классифайдов в сфере недвижимости. Согласно исследованию «РБК Исследования рынков» в 2021 году компании принадлежало 20,6 процентов российского рынка онлайн-классифайдов в сегменте квартир, она уступала только Авито [10]. Весь российский рынок онлайн-классифайдов оценивался в 8,9 миллиардов рублей. На следующем рисунке приведена информация о том, какими долями российского рынка онлайн-классифайдов владеют упомянутые компании (рис. 1.1).



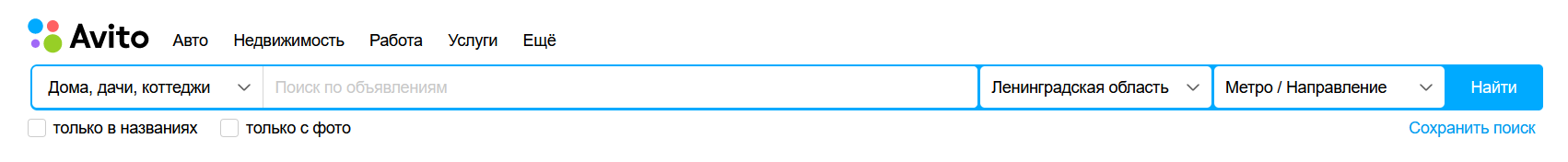
**Рис. 1.1.** Доли российского рынка онлайн-классифайдов [10]

Сайт компании ЦИАН был выбран, так как предоставляет возможность собрать информацию о большем количестве признаков, чем сайт Авито. Так, если ЦИАН позволяет пользователям фильтровать дома и коттеджи по размеру жилой площади и прилегающему участку, типу материала и отопления (рис. 1.1), в то время как лидирующий по доле рынка онлайн-классифайдов Авито (рис. 1.2, рис. 1.3) имеет меньший инструментарий для поиска и фильтра домов.

Изображение выглядит как текст

Автоматически созданное описание

**Рис. 1.2.** Фильтры в поиске сайта cian.ru.



**Рис. 1.3.** Фильтры в поиске сайта avito.ru.

Изображение выглядит как текст

Автоматически созданное описание

**Рис. 1.4.** Фильтры в поиске сайта avito.ru.

Таким образом, сайт Циана позволяет собрать информацию о типе материала (типе дома), типе отопления, наличии коммуникаций, площади и т. д., в то время как сайт Авито дает возможность фильтровать только по цене и направлению.

## 1.4. Признаки объектов недвижимости

У объектов недвижимости (домов, коттеджей) есть признаки, которые могут как повышать их стоимость, так и понижать. Далее будут рассмотрены некоторые из них, их преимущества и недостатки.

### 1.4.1. Типы отопления

При выборе загородной недвижимости многие покупатели обращают внимание на тип отопления (систему обогрева), которая установлена в доме. Например, есть три типа: электрическое, твердотопливное и газовое [18]. В следующей таблице (табл. 1.2) представлены преимущества и недостатки этих типов.

**Таблица 1.2.** Преимущества и недостатки различных типов отопления.

|  |  |  |
| --- | --- | --- |
| **Тип отопления** | **Преимущества** | **Недостатки** |
| Электрическое | 1. Экологичность - угарный газ не выделяется при использовании электроприборов. 2. Самый большой коэффициент полезного действия - до 98 процентов 3. может перейти из электроэнергии в тепло. 4. Зонирование обогрева - в различных комнатах дома можно установить собственный температурный режим. | 1. Высокая стоимость электрического обогрева. |

**Таблица 1.2.** Преимущества и недостатки различных типов отопления (продолжение).

|  |  |  |
| --- | --- | --- |
| **Тип отопления** | **Преимущества** | **Недостатки** |
| Твердотопливное | 1. Автономность - дом не зависит от газовых и электрических сетей. 2. Мощность - даже достаточно просторные дома могут быть отапливаемыми данной системой обогрева. | 1. Риск выделения опасного угарного газа при поломке системы. 2. Высокие затраты на непрерывную систему подачи топлива. 3. Содержание склада для топлива. |
| Газовое | 1. Экономичность - наименее затратная система обогрева. 2. Простота эксплуатации - не требует частых наладок оборудования. 3. Мощность. | 1. Бюрократические сложности - необходимость согласовывать установку газового оборудования с местной администрацией согласно текущему законодательству. |

### 1.4.2. Тип участка

Также при покупке недвижимости за городом большую роль тип участка. Среди многих типов участка можно выделить три: ИЖС (индивидуальное жилищное строительство), СНТ (садовое некоммерческое товарищество), ДНП (дачное некоммерческое партнерство) [23]. В следующей таблице (табл. 1.3) представлены преимущества и недостатки этих типов.

**Таблица 1.3.** Преимущества и недостатки различных типов участка.

|  |  |  |
| --- | --- | --- |
| **Тип участка** | **Преимущества** | **Недостатки** |
| ИЖС | 1. Покупателям легче будет получить прописку, а также официальный адрес. 2. Зачастую к таким участкам расположена социальная инфраструктура | 1. Обязательное согласование дома с точки зрения круглогодичного проживания (соблюдение ГОСТов, СНиПов). |
| ДНП | 1. Покупателям легче будет получить «прописку» в новом доме. 2. Дом уже признан жилым. 3. Покупатель становится членом партнерства, тем самым, получает возможность влияет на решения партнерства. 4. Более низкая стоимость. | 1. Так как этот тип участков предназначен для земледелия, участки этого типа должны быть оформлены в собственность. 2. Зачастую такие участки отдалены от социальной инфраструктуры. |
| СНТ | 1. Сниженная стоимость участков. 2. Участки можно использовать без строительства домов. | 1. Зачастую такие участки менее связаны инженерными коммуникациями. 2. Покупателям сложнее оформить «прописку». |

### 1.4.3. Тип материала

При выборе загородного дома также необходимо определиться с материалом, из которого он построен [19]. Среди многих типов можно выделить несколько типов материала: кирпич, газобетонные блоки, пенобетонные блоки, дерево, каркасное жилье.

В следующей таблице (табл. 1.4) представлены основные преимущества и недостатки выделенных типов материала.

**Таблица 1.4.** Преимущества и недостатки различных типов материала.

|  |  |  |
| --- | --- | --- |
| **Тип материала** | **Преимущества** | **Недостатки** |
| Кирпич | 1. Сохранение тепла. 2. Защита от шума. 3. Долговечность. 4. Подходит под любую реализацию. 5. Защита от коррозии и грибка. | 1. Трудность в укладке. 2. Необходим мастер для строительства. 3. Необходим основательный фундамент и теплоизолятор. 4. Длительность строительства. |
| Газобетонный блок | 1. Быстрое строительство. 2. Теплопроводность. 3. Хорошо сжимается в холод. 4. Не горит. 5. Не может быть испорчен вредителями. 6. Безопасен для экологии. 7. Низкая стоимость. | 1. Хорошо поглощает воду. 2. Плохо выдерживает разгибания. 3. Для крепежа можно использовать дюбели лишь определенных видов. |
| Пенобетонный блок |
| Дерево | 1. Быстрое строительство. 2. Экологичность. | 1. Подвержено гниению. 2. Подвержено возгоранию. 3. Недолговечность. 4. Страдает от вредителей. |

**Таблица 1.4.** Преимущества и недостатки различных типов материала (продолжение).

|  |  |  |
| --- | --- | --- |
| **Тип материала** | **Преимущества** | **Недостатки** |
| Каркасное жилье | 1. Низкая стоимость. 2. Быстрое строительство. 3. Легко перепланировать дом. | 1. Непрочность стен. 2. Недолговечность. 3. Сложность вентиляции. 4. Необходимость ежегодной обработки стен. |

### 1.4.4. Районы Ленинградской области

В состав Ленинградской области входит 17 муниципальных районов и один городской округ (рис. 1.5).

Изображение выглядит как карта

Автоматически созданное описание

**Рис. 1.5.** Карта Ленинградской области [13].

Эти муниципальные образование отличаются между собой динамикой их экономического развития [17]. Так, по подсчетам издания «Деловой Петербург» за 2020 год, рейтинг экономического развития районов выглядел следующим образом (табл. 1.5). При подсчетах использовались четыре фактора: промышленность, инвестиции, сельское хозяйство и доходы жителей.

**Таблица 1.5.** Рейтинг экономического развития муниципальных районов Ленинградской области.

|  |  |
| --- | --- |
| **№** | **Муниципальный район** |
| 1 | Кингисеппский |
| 2 | Выборгский |
| 3 | Приозерский |
| 4 | Тихвинский |
| 5 | Тосненский |
| 6 | Ломоносовский |
| 7 | Лужский |
| 8 | Кировский |
| 9 | Сосновоборский (городской округ) |
| 10 | Бокситогорский |
| 11 | Гатчинский |
| 12 | Сланцевский |
| 13 | Волховский |
| 14 | Киришский |
| 15 | Волосовский |
| 16 | Всеволожский |
| 17 | Лодейнопольский |

**Таблица 1.5.** Рейтинг экономического развития муниципальных районов Ленинградской области (продолжение).

|  |  |
| --- | --- |
| **№** | **Муниципальный район** |
| 18 | Подпорожский |

### 1.4.5. Ближайшие трассы

При выборе объекта недвижимости покупатели обращают внимание на ближайшую трассу, которая проходит близ поселения, в котором находится дом [7]. Аналитики даже замеряют спрос на загородную недвижимости в разрезе «направлений», то есть, ближайших трасс, исходящих из центра области. В этом случае речь идет о шоссе, исходящих из Санкт-Петербурга.

В следующей таблице (табл. 1.6) приведена информация об основных трассах, исходящих из Санкт-Петербурга.

**Таблица 1.6.** Трассы, исходящие из Санкт-Петербурга [24].

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Название** | **Конечная точка** | **Расположение** |
| 1 | Приморское | город Выборг, Ленинградская область |  |
| 2 | Скандинавия | МАПП «Торфяновка», государственная граница РФ и Финляндии |  |

**Таблица 1.6.** Трассы, исходящие из Санкт-Петербурга [24] (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Название** | **Конечная точка** | **Расположение** |
| 3 | Выборгское | посёлок Красносельское |  |
| 4 | Новоприозерское (часть трассы «Сортавала») | посёлок Лосево |  |
| 5 | Дорога Жизни | посёлок при станции Ладожское озеро |  |
| 6 | Мурманское (часть трассы «Кола») | административная граница Ленинградской области с Республикой Карелия |  |
| 7 | Московское (М-10 «Россия») | город Москва |  |

**Таблица 1.6.** Трассы, исходящие из Санкт-Петербурга [24] (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Название** | **Конечная точка** | **Расположение** |
| 8 | Киевское (часть трассы Р-23 «Псков») | государственная граница РФ и Республики Беларусь |  |
| 9 | Таллиннское (часть трассы «Нарва») | город Ивангород, государственная граница РФ и Эстонии |  |
| 10 | Санкт-Петербург - Большая Ижора | деревня Ручьи |  |

## 1.5. Гипотезы и вопросы

Предварительно изучив рынок загородной недвижимости Ленинградской области, условия, в которых он развивается и факторы, которые на него влияют, можно выделить несколько гипотез на основе экспертных оценок, которые были приведены в первой главе:

* СНТ и ДНП являются более дешевыми типами участков, чем другие, например, ИЖС [16];
* Поскольку электрический тип отопления является более дорогим к установке [18], дома с таким типом отопления тоже стоят дороже;
* Поскольку центральный газовый тип отопления является более дешевым к установке [18], дома с таким типом отопления тоже стоят дешевле;
* Участки с домами, построенными из газо- и пенобетонных блоков стоят дешевле [12];
* Участки с каркасными домами стоят дешевле [12];

Соответственно, в рамках исследования должны быть поставлены следующие вопросы:

* Участки каких типов являются в среднем более дорогими?
* Участки с домами из каких материалов стоят дороже, чем другие?
* Участки с домами, имеющими какой тип отопления оцениваются выше, чем остальные?
* Как расстояние от административной границы с Санкт-Петербургом влияет на цену недвижимости?
* Как наличие коммуникаций влияет на стоимость объекта недвижимости?
* Как ближайшая трасса влияет на стоимость объекта недвижимости?
* Какие муниципальные районы Ленинградской области являются более «дорогими» в плане загородной недвижимости, чем другие?

## Выводы

В главе была приведена информация о рынке загородной недвижимости Ленинградской области, об условиях, в которых он формируется, компаниях, которые в нем функционируют. Были описаны факторы (тип отопления, тип материала и проч.), которые могут иметь влияние на стоимость объектов загородной недвижимости. Также были сформулированы гипотезы и вопросы, которые касаются описанных факторов: например, как тип отопления влияет на стоимость дома, правда ли, что каркасное жилье стоит меньше, чем остальные типы.

# Глава 2. Методы машинного обучения для анализа рынка недвижимости

## 2.1. Метод сбора данных

Далее будет рассмотрен парсинг как метод сбора данных, подходящий заявленному исследованию.

Парсинг – это автоматизированный синтаксический и лексический анализ текстовых документов [9]. Особая потребность в парсинге и в программах-парсерах возникла с появлением интернета как основного источника информации.

Компаниям же парсинг веб-сайтов дает следующие возможности [11]:

* Автоматизировать процессы сбора данных в нужном масштабе;
* Получить доступ к источникам данных в интернете, которые могут повысить эффективность бизнеса;
* Принимать решения с опорой на данные.

Процесс парсинга интернет-сайтов состоит из следующих шагов [11]:

* Идентификация целевых URL-адресов;
* Если есть необходимость обойти инструменты противодействия парсингу, то должен быть выбран прокси-сервер, с помощью которого будет отправляться запрос;
* Отправка запросов на идентифицированные URL-адреса с целью получения HTML-кода;
* Использование указателей для выделения данных в полученном коде;
* Аналитический разбор строки данных, которая содержит необходимую информацию;
* Преобразование собранных данных в нужный формат;
* Передача данных в хранилище данных.

## 2.2. Алгоритмы выявления выбросов

Далее будет представлено несколько алгоритмов выявления выбросов из наборов данных.

### 2.2.1. Стандартизированная оценка

Также этот метод известен как z-оценка. Известно, что согласно свойствам нормального распределения, 99,7 процентов значений должны отстоять не более, чем на 3δ. Таким образом, для i-го наблюдения значение ( – значение наблюдения, – среднее, – среднеквадратическое отклонение) не должно превышать 3 по модулю [4].

Преимуществом данного метода является расчет оценки для каждого значения. Однако этот алгоритм можно использовать исключительно для признаков, имеющих нормальное распределение [4].

### 2.2.2. Метод Тьюки

Для данного метода требуется рассчитать межквартильный размах – разность первого и третьего квартиля [15]. Если значение отстает на три (или полтора, в случае выбора «мягких границ») межквартильных размаха от указанных квартилей, оно должно быть удалено. Этот метод работает для величин, распределение которых близко к нормальному.

## 2.3. Методы анализа данных

### 2.3.1. Описательная статистика

В описательную статистику входит вычисление следующих мер [25]:

* Меры частоты – пропорция, частота – описывают то, как часто то или иное значение встречается в наборе;
* Меры центральной тенденции – среднее, мода, медиана – указывают на распределение величины;
* Меры дисперсии или вариации – размах, вариация, стандартное отклонение – указывают на размах значений;
* Меры позиции – перцентили, квартили – описывают то, где значения находятся по отношению друг к другу.

### 2.3.2. Корреляционные связи

Коэффициент корреляции является отношением момента корреляции двух случайных величин к произведению их среднеквадратических отклонений [3]. Для выборок с объектами, имеющими два признака и более может быть построена корреляционная матрица для оценки тесноты связи между признаками.

### 2.3.3. Линейная регрессия

Линейная регрессия – модель зависимости переменной x от одной или нескольких других переменных (т. н. факторов, регрессоров, независимых переменных) с линейной функцией зависимости [14].

Цель линейной регрессии – поиск линии, которая наилучшим образом соответствует точкам на корреляционном поле. Также линейная регрессия может быть множественной, то есть, иметь несколько объясняющих переменную x факторов.

Одной из характеристик этой модели, которая дает оценку качеству модели – коэффициент детерминации, R2 [5]. Он является отношением объясненной суммы квадратов отклонений к общей сумме. Чем ближе R2 к единице, тем качество модели выше.

### 2.3.4. Случайный лес

Случайный лес – это множество деревьев решений [6]. В задаче регрессии их ответы усредняются, а в задаче классификации решение принимается «большинством голосов». Все деревья строятся независимо следующим образом:

* Создается подвыборка обучающей выборки размера n (возможно, с повторениями);
* Для построения каждого расщепления в дереве просматриваем k случайных признаков (для каждого нового расщепления — свои случайные признаки).
* Выбирается наилучшие признак и расщепление по нему (по заранее заданному критерию). Дерево строится, как правило, до исчерпания выборки (пока в листьях не останутся представители только одного класса, в случае задачи классификации).

Такая схема построения алгоритма является примером ансамблирования: построения алгоритма машинного обучения на основе нескольких.

## 2.4. Примеры использования методов машинного обучения для анализа рынка недвижимости

Далее будет представлен обзор нескольких статей, авторы которых использовали методы машинного обучения, в частности, множественную линейную регрессию и случайный лес для оценки объектов недвижимости.

### 2.4.1. Shetty D., Prakash Rao B., Prakash C., S. Vaibhava - Multiple regression analysis to predict the value of a residential building and to compare with the conventional method values

Авторы статьи для анализа рынка недвижимости выбрали множественный регрессионный анализ. Сначала были выбраны независимые признаки:

* площадь застройки (*Built Up Area*), в квадратных футах;
* количество этажей *(No. of Floors*);
* «возраст» дома (*Age of Building*), в годах;
* количество комнат (*No. of Rooms*);
* количество парковочных мест (*No. of Parking Spaces*);
* форма участка (*Plot Shape*) - три степени «конвенциональности» форм;
* локация (*Location*) - три степени качества ближайшей инфраструктуры;
* доступ к дороге (*Access to Road*) - три степени качества ближайшей дороги;
* близость к удобствам (*Nearness to amenities*) - три степени близости к удобствам (от нуля до трех километров, от трех до шести километров, от шести до десяти километров).

Отметим, что последние четыре признака являются качественными, которым были присвоены числовые значения.

В качестве зависимой переменной была выбрана рыночная стоимость жилья (в индийских рупиях).

Для построения регрессии был использован набор данных из 20 объектов недвижимости. В результате была построена модель (множественная линейная регрессия) со скорректированным коэффициентом детерминации (R-squared) в 99,6 процентов.

Далее авторы статьи сравнивают результат оценки, полученный с помощью регрессионной модели с другими, конвенциональными подходами:

* Land and Building Method (метод земли и строения) - стоимость земли и стоимость строения оцениваются отдельно и стоимость объекта недвижимости равна их сумме;
* Rental Income Approach (подход дохода от аренды) - стоимость объекта недвижимости равна доходу от рыночной оценки аренды, капитализированный по ставке, равной норме прибыли;
* Composite Rate Method (метод составной ставки) - составная ставка получается сравнением с ближайшими похожими зданиями;
* Valuation by Detailed Estimation (подетальная оценка) - стоимость объекта недвижимости состоит из большего количества стоимостей, чем стоимость земли и строения.

После сравнения результатов различными методами оценки выяснилось, что наиболее схожими по точности оценками оказались оценки, полученные с помощью метода составной ставки и подетальной оценки (отличаются от оценки с помощью регрессионной модели на 4 и 8 процентов соответственно).

### 2.4.2. Крылова Д. А., Аббакумов Р. Г., Моргунова О. Н. - Использование методов регрессионного анализа при оценке стоимости недвижимости

Авторы статьи применили регрессионный анализ для оценки стоимости недвижимости. Была поставлена цель - сделать оценку конкретному объекту недвижимости («объект оценки»). Ими был построен набор данных из 11 объектов-аналогов (аналогов по отношению к объекту оценки) со следующими признаками:

* адрес;
* цена за квадратный метр, в тысячах рублей;
* расстояние до центра города, в километрах;
* расстояние до остановки общественного транспорта, в километрах;
* этаж;
* наличие отдельного входа - есть/нет;
* класс помещения - три градации класса;
* парковочные места - наличие/отсутствие платной/бесплатной парковки.

Как и в предыдущей статье, последние три признака на самом деле являются качественными, которым были присвоены числовые значения.

Далее была построена модель (множественная линейная регрессия), скорректированный коэффициент детерминации которой равен 99,1 процентов.

### 2.4.3. Abigail Bola Adetunji, Oluwatobi Noah Akande, Funmilola Alaba Ajala, Ololade Oyewo, Yetunde Faith Akande, Gbenle Oluwadara – House Price Prediction using Random Forest Machine Learning Technique

Авторы статьи анализировали рынок недвижимости субурбий Бостона (США, штат Массачусетс). В набор данных вошло порядка пятисот объектов, обладающих четырнадцатью признаками. Такими признаками были, например:

* уровень преступности в городе;
* процент чернокожего населения;
* находится ли объект у реки Чарльз.

Далее была построен случайный лес. На рисунке (рис. 2.1) изображен сравнение настоящих значений цены дома и прогнозных. Также дана абсолютная разница между этими значениями.

Изображение выглядит как стол

Автоматически созданное описание

**Рис. 2.1.** Сравнение настоящих и прогнозных значений.

Итак, можно наблюдать следующее: разница между значениями составляет несколько процентов, значения близки друг другу. Ошибка аппроксимации (относительная ошибка) для данного кейса составляет 8 процентов.

## 2.5. Методология

Так, основываясь на изученной литературе [10, 26, 27] для решения поставленных ранее задач был разработан перечень следующих шагов:

* Сбор данных с сайта cian.ru. Сайт Циана более удобен для парсинга. В итоге должен получиться набор данных об объектах недвижимости и их признаках. Для сбора данных об объектах недвижимости с сайта cian.ru должен быть использован набор библиотек Selenium для Python [22]. Selenium позволяет автоматизировать управление интернет-браузерами и собирать данные с веб-страниц, которые таким образом будут загружены. Иначе говоря, будет использован парсинг веб-страниц [9, 11]. Для выделения необходимых данных будут использованы регулярные выражения. Для построения наборов данных, совмещающих информацию из различных источников (например, базы данных Петростата [2]) возможно использовать библиотеку Pandas для Python и Microsoft Excel, в частности, инструмент Power Query.
* Очищение полученного набора от проблемных значений (пустые значения, ошибки), а также выбросов, с помощью z-оценки [4].
* Анализ переменных, включенных в набор данных, которые могут быть, как и признаками объектов [10, 26, 27], так и характеризовать местность, в котором объект недвижимости располагается [26], например, муниципалитета. Также некоторые признаки будут качественными [27].
* Применение методов описательной статистики для выявления потенциально важных факторов при построении моделей оценки стоимости объектов загородной недвижимости. В частности, будут построены корреляционные матрицы для количественных признаков, и средние показатели цен в различных категориях, определяемых качественными признаками.
* Изучение характера распределения признака, который будет выбран как зависимая переменная, будет применено правило «трёх сигма» [4].
* Построение моделей машинного обучения для оценки стоимости объектов загородной недвижимости [10, 26, 27].
* Для применения инструментов описательной статистики и машинного обучения будут использованы такие программные продукты, как Google Colab, Deductor, RapidMiner Studio и Stata.

## Выводы

В главе приведено описание методов машинного обучения, которые могут быть использованы в этом исследовании. Также был сделан обзор статей, авторы которых используют эти методы для анализа рынка недвижимости, для прогнозирования стоимости объектов недвижимости. На этой основе была разработана методология для следующего исследования рынка загородной недвижимости Ленинградской области.

# Глава 3. Исследование рынка с помощью методов машинного обучения

## 3.1. Описание наборов данных

Для исследования было собрано три набора данных, которые содержат информацию об объектах загородной недвижимости:

* набор данных 1 - данные об объектах недвижимости, которые были выставлены на продажу (порядка 1500 объектов до очистки, 855 – после);
* набор данных 2 - данные об объектах недвижимости, которые сдаются в аренду посуточно (около 300 объектов до очистки, 275 – после);
* набор данных 3 - данные об объектах недвижимости, которые сдаются в аренду помесячно (порядка 150 объектов до очистки, 111 – после).

## 3.2. Исследование набора данных об объектах, выставленных на продажу

Далее будет представлено исследования набора данных 1. Оно имело целью построение моделей, с помощью которых можно будет спрогнозировать стоимость определенного объекта недвижимости.

### 3.2.1. Описание данных

В следующей таблице (табл. 3.1) представлена информация о количественных признаках объектов в наборе данных 1.

**Таблица 3.1.** Количественные признаки, набор данных 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Наименование (на английском языке)** | **Объяснение** | **Размерность** | **Промежуток** |
| 1 | *Meters* | Общая площадь жилой зоны (дома) | квадратный метр | [0; +∞) |
| 2 | *Land* | Общая площадь участка при жилой зоне (доме) | ар («сотка») | [0; +∞) |
| 3 | *Price* | Цена объекта недвижимости | российский рубль | [0; +∞) |

**Таблица 3.1.** Количественные признаки, набор данных 1 (продолжение).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Наименование (на английском языке)** | **Объяснение** | **Размерность** | **Промежуток** |
| 4 | *price\_per\_meter* | Цена объекта недвижимости, разделенная на площадь жилой зоны (*prices/meters*) | российский рубль | [0; +∞) |
| 5 | *Electricity* | Фиктивная переменная: наличие электричества на участке | - | {0, 1} |
| 6 | *Gas* | Фиктивная переменная: наличие газа на участке | - | {0, 1} |
| 7 | *Water* | Фиктивная переменная: наличие водоснабжения на участке | - | {0, 1} |
| 8 | *Drainage* | Фиктивная переменная: наличие канализации на участке | - | {0, 1} |
| 9 | *Commodities* | Сумма переменных *electricity, gas, water, drainage* | - | {0, 1, 2, 3, 4} |
| 10 | *Floors* | Количество этажей в доме | этажи | {1, 2, 3} |
| 11 | *Distance* | Дистанция от объекта до КАД (кольцевой автодороги) | километр | [0; +∞) |

В наборе данных 1 используются следующие качественные признаки (табл. 3.2).

**Таблица 3.2.** Качественные признаки, набор данных 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Наименование (на английском языке)** | **Объяснение** | **Возможные значения** |
| 1 | *Type* | Тип жилого строения | Дом, Коттедж |

**Таблица 3.2.** Качественные признаки, набор данных 1 (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Наименование (на английском языке)** | **Объяснение** | **Возможные значения** |
| 2 | *Form* | Вид участка | ИЖС, ДНП, Садоводство |
| 3 | *Road* | Ближайшая трасса | Кола, Нарва, Сортавала |
| 4 | *District* | Район Ленинградской области (один из семнадцати) | Выборгский, Всеволожский, Ломоносовский |
| 5 | *Settlement* | Муниципалитет в составе района области | Аннинское городское поселение, Лесколовское сельское поселение |
| 6 | *Locality* | Населенный пункт в муниципалитете | Рощино (посёлок городского типа), Мендсары (деревня), Новая Ропша (садоводческий массив) |
| 7 | *Description* | Описание, данное объекту недвижимости при загрузке объявления |  |
| 8 | *Actor* | Тип лица, которое опубликовало объявление | Собственник, риэлтор, агентство недвижимости |
| 9 | *Agents* | Компания, лицо, которые отправили объявление |  |

**Таблица 3.2.** Качественные признаки, набор данных 1 (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Наименование (на английском языке)** | **Объяснение** | **Возможные значения** |
| 10 | *Material* | Материал, из которого изготовлено жилое строение | Кирпичный, Монолитный |
| 11 | *Heating* | Тип отопления | Угольное, Автономное газовое |

К этому набору (и двум другим) добавлена информация о муниципалитетах (*settlement*), которую предоставляет Петростат [2]. В следующей таблице (табл. 3.3) представлена информация об этих количественных признаках. Объектам недвижимости дается значение этих признаков в зависимости от муниципалитетов, в которых они находятся.

**Таблица 3.3.** Количественные признаки муниципалитетов.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Наименование (на английском языке)** | **Объяснение** | **Размерность** | **Промежуток** |
| 1 | *new\_homes* | Введено в действие жилых домов на территории муниципального образования, квадратный метр общей площади,  значение показателя за 2020 год | Квадратный метр | [0; +∞) |
| 2 | *profitable\_firms* | Удельный вес прибыльных организаций в общем числе организаций по данным бухгалтерской отчетности, процент,  значение показателя за 2020 год | Процент | [0; 100] |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Наименование (на английском языке)** | **Объяснение** | **Размерность** | **Промежуток** |
| 3 | *culture\_workers* | Количество работников досуговых и культурных учреждений, 2016 год | Человек | [0; +∞) |

### 3.2.2. Проверка на нормальное распределение

В первом наборе данных есть пять признаков, имеющих непрерывное распределение: *meters, land, price, price\_per\_meter, distance*. В этом разделе будет рассмотрено распределение этих признаков, является ли оно нормальным.

В следующей таблице (табл. 3.4) представлены данные о том, соответствует ли распределение нормальному согласно правилу трех сигм [4]. В трех колонках указана доля объектов набора данных, попадающих в промежуток [μ-n\*σ, μ+n\*σ] (μ – среднее, n – количество отклонений, σ – среднеквадратическое отклонение).

**Таблица 3.4.** Применение правила трех сигм для набора данных 1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№** | **Признак** | **1σ** | **2σ** | **3σ** | **Количество соответствий** |
| 1 | *Meters* | 90,25% | 96,09% | 98,42% | 2 из 3 |
| 2 | *Land* | 99,03% | 99,18% | 99,18% | 2 из 3 |
| 3 | *Price* | 94,92% | 98,01% | 98,63% | 2 из 3 |
| 4 | *price\_per\_meter* | 87,23% | 96,77% | 97,87% | 2 из 3 |
| 5 | *Distance* | 85,65% | 95,05% | 97,94% | 1 из 3 |
|  | При нормальном распределении | ≥68,26% | ≥95,44% | ≥99,72% |  |

Таким образом, строгого соответствия правилу трех сигм не достигает ни один из признаков, однако от необходимых показателей полученным долям не хватает не более 2 процентов. Поэтому можно выбрать метод Тьюки для удаления выбросов.

В следующей таблице (табл. 3.5) представлены гистограммы, построенные для этих признаков.

**Таблица 3.5.** Гистограммы пяти признаков набора данных 1.

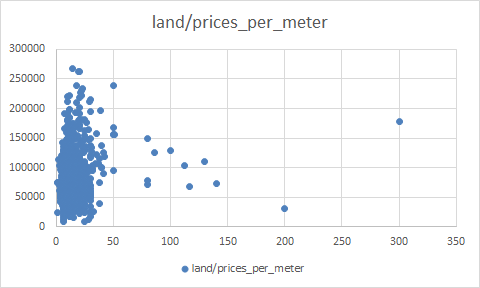
|  |  |  |
| --- | --- | --- |
| **№** | **Признак** | **Гистограмма** |
| 1 | *Price* |  |
| 2 | *price\_per\_meter* |  |
| 3 | *meters* |  |

**Таблица 3.5.** Гистограммы пяти признаков набора данных 1 (продолжение).

|  |  |  |
| --- | --- | --- |
| **№** | **Признак** | **Гистограмма** |
| 4 | *Land* |  |
| 5 | *distance* |  |

### 3.2.3. Проблема выбросов

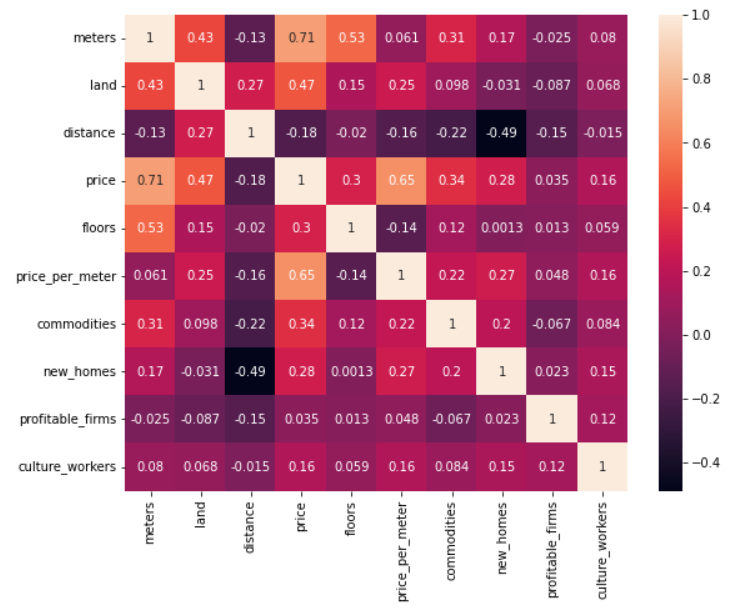
В первую очередь, для очищения набора данных 1 от выбросов с помощью метода Тьюки. Однако были выявлены (рис. 3.1) сильно отличающиеся значения по осям *price\_per\_meter* и *land*. После чего было принято убрать из набора объекты со значением *land* в 50 и более соток.



**Рис. 3.1.** Корреляционное поле для набора данных 1, построенное по признакам land и price\_per\_meter.

### 3.2.4. Корреляционные связи

На следующем рисунке представлена корреляционная матрица для набора данных 1.



**Рис. 3.2.** Корреляционная матрица для набора данных 1.

Оценим силу корреляционных связей по шкале Чеддока [4]. Корреляционные связи между переменной *price\_per\_meter* и:

* *land* - 0,25, слабая связь;
* *floors* - -0,14, слабая связь;
* *commodities* - 0,22, слабая связь;
* *new\_homes* – 0,27, слабая связь;
* *culture\_workers* – 0,16, слабая связь.
* *distance* - -0,16, слабая связь.

### 3.2.5. Вопрос о типах участка

В следующей таблице (табл. 3.6) представлены данные о средних ценах за квадратный метр в зависимости от типа участка (*form*).

**Таблица 3.6.** Средние цены за квадратный метр по типам участка.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип участка** | **Количество объектов** | **Средняя цена за квадратный метр, рубли** |
| 1 | ДНП | 151 | 91262 |
| 2 | ИЖС | 468 | 83788 |
| 3 | Фермерское хозяйство | 3 | 72519 |
| 4 | Садоводство | 233 | 71037 |
|  | Итого | 855 | 81594 |

Наиболее дорогими являются объекты, построенные на участках типа ДНП (дачное некоммерческое партнерство), наименее - в садоводствах. Наиболее «популярным» типом участка является ИЖС (индивидуальное жилищное строительство).

### 3.2.6. Вопрос о ближайшей трассе

В следующей таблице (табл. 3.7) представлены данные о средних ценах за квадратный метр в зависимости от ближайшей трассы (*road*).

**Таблица 3.7.** Средние цены за квадратный метр по ближайшей трассе у объекта.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Ближайшая трасса** | **Количество объектов** | **Средняя цена за квадратный метр, рубли** |
| 1 | Скандинавское шоссе | 15 | 123262 |
| 2 | ЗСД | 11 | 116987 |
| 3 | Зеленогорское | 5 | 109669 |
| 4 | Приморское | 86 | 100501 |
| 5 | Выборгское | 51 | 97988 |
| 6 | Сортавала | 166 | 94746 |
| 7 | Санкт-Петербург – Большая Ижора | 11 | 89806 |
| 8 | Дорога Жизни | 35 | 86331 |
| 9 | Пулковское | 9 | 82105 |
| 10 | Московское, М-10 | 108 | 74220 |
| 11 | Колтушское | 2 | 73515 |
| 12 | Кола | 121 | 70855 |
| 13 | Киевское | 36 | 69541 |
| 14 | Нарва | 199 | 64125 |
|  | Итого | 855 | 81594 |

Наиболее дорогим направлением является Скандинавское шоссе, которое пролегает по территории Выборгского района области. Далее будет изучен вопрос о том, как нахождение объекта недвижимости в том или ином районе Ленинградской области.

### 3.2.7. Вопрос о районах области

В следующей таблице (табл. 3.8) представлены данные о средних ценах за квадратный метр в зависимости от района, в котором находится объект недвижимости (*district*).

**Таблица 3.8.** Средние цены за квадратный метр по районам области.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Район области** | **Количество объектов** | **Средняя цена за квадратный метр, рубли** |
| 1 | Выборгский | 110 | 103744 |
| 2 | Приозерский | 51 | 98874 |
| 3 | Всеволожский | 281 | 91608 |
| 4 | Тосненский | 48 | 76229 |
| 5 | Кингисеппский | 11 | 73786 |
| 6 | Гатчинский | 132 | 70282 |
| 7 | Ломоносовский | 135 | 69495 |
| 8 | Лодейнопольский | 2 | 61833 |
| 9 | Кировский | 47 | 56957 |
| 10 | Волосовский | 12 | 53899 |
| 11 | Лужский | 11 | 40205 |
| 12 | Волховский | 11 | 39394 |
| 13 | Сланцевский | 3 | 29208 |
| 14 | Киришский | 1 | 26000 |
|  | Итого | 855 | 81594 |

При рассмотрении вопроса о ближайшей трассе было выяснено, что Скандинавское шоссе, которое пролегает по территории Выборгского района, является самым дорогим направлением. Здесь было снова подтверждено, что недвижимость в Выборгском районе является самой дорогой. Наиболее «населенным» районом является Всеволожский.

### 3.2.8. Вопрос о лицах, публикующих объявления

В следующей таблице (табл. 3.9) дана информация о средних ценах за квадратный метр в зависимости от лица, которое опубликовало объявление (*actor*).

**Таблица 3.9.** Средние цены по лицам, опубликовавшим объявления.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Лицо** | **Количество объектов** | **Средняя цена за квадратный метр, рубли** |
| 1 | Агенство недвижимости | 438 | 82073 |
| 2 | Риелтор | 207 | 81166 |
| 3 | Собственник | 210 | 81016 |
|  | Итого | 850 | 81954 |

Исходя из предыдущей таблицы (табл. 3.9) разница в средней цене за квадратный метр между этими категориями несущественна. Однако заметно следующее: чаще всего объявления публикуются агентствами недвижимости.

### 3.2.9. Вопрос о типах отопления

В следующей таблице (табл. 3.10) дана информация о средних ценах за квадратный метр в зависимости от типа отопления (*heating*), который используется в доме или коттедже.

**Таблица 3.10.** Средние цены по типам отопления объектов недвижимости.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип отопления** | **Количество объектов** | **Средняя цена за квадратный метр, рубли** |
| 1 | Угольное | 6 | 106487 |
| 2 | Дизельное | 24 | 105371 |
| 3 | Центральное газовое | 159 | 97280 |
| 4 | Автономное газовое | 132 | 87141 |

**Таблица 3.10.** Средние цены по типам отопления объектов недвижимости (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип отопления** | **Количество объектов** | **Средняя цена за квадратный метр, рубли** |
| 5 | Камин | 32 | 80300 |
| 6 | Электрическое | 336 | 79838 |
| 7 | Твердотопливный котел | 23 | 68498 |
| 8 | Без отопления | 16 | 66030 |
| 9 | Печь | 127 | 59823 |
|  | Итого | 855 | 81954 |

Самым «популярным» типом отопления является электрическое. Однако самым «дорогим», не считая угольного, у которого всего лишь шесть объектов, является дизельное. Самым «дешевым» - печное или объекты без отопления.

### 3.2.10. Вопрос о типах материала

В следующей таблице (табл. 3.11) дана информация о средних ценах за квадратный метр в зависимости от типов материала (*material*), из которого построен дом или коттедж.

**Таблица 3.11.** Средние цены по типам материала объектов недвижимости.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип материала** | **Количество объектов** | **Средняя цена за квадратный метр, рубли** |
| 1 | Газосиликатный блок | 2 | 150431 |
| 2 | Монолитный | 34 | 97206 |
| 3 | Кирпичный | 113 | 90550 |
| 4 | Газобетонный блок | 167 | 90521 |
| 5 | Каркасный | 211 | 82069 |

**Таблица 3.11.** Средние цены по типам материала объектов недвижимости (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип материала** | **Количество объектов** | **Средняя цена за квадратный метр, рубли** |
| 6 | Деревянный | 217 | 79391 |
| 7 | Пенобетонный блок | 167 | 59980 |
| 8 | Щитовой | 25 | 51650 |
|  | Итого | 855 | 81954 |

Наиболее дорогими являются монолитные и кирпичные дома. Наиболее дешевые – щитовые. Наиболее «популярные» - каркасные и деревянные.

### 3.2.11. Вопрос о типах объекта

В следующей таблице (табл. 3.12) дана информация о средних ценах за квадратный метр в зависимости от типа объекта (*type*).

**Таблица 3.12.** Средние цены по лицам, опубликовавшим объявления.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип объекта** | **Количество объектов** | **Средняя цена за квадратный метр, рубли** |
| 1 | Коттедж | 618 | 93479 |
| 2 | Дом | 237 | 77036 |
|  | Итого | 855 | 81954 |

Объекты, обозначенные как коттеджи в среднем стоят дороже, чем дома. Однако домов в наборе больше, чем коттеджей.

В следующей таблице (табл. 3.13) представлено распределение объектов различного типа (*type*) по районам области (*district*).

**Таблица 3.13.** Количество объектов разного типа по районам области.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Район области** | **Количество домов** | **Количество коттеджей** |
| 1 | Волосовский | 10 | 2 |
| 2 | Волховский | 9 | 2 |
| 3 | Всеволожский | 176 | 105 |

**Таблица 3.13.** Количество объектов разного типа по районам области (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Район области** | **Количество домов** | **Количество коттеджей** |
| 4 | Выборгский | 78 | 32 |
| 5 | Гатчинский | 105 | 27 |
| 6 | Кингисеппский | 11 | 0 |
| 7 | Киришский | 1 | 0 |
| 8 | Кировский | 39 | 8 |
| 9 | Лодейнопольский | 2 | 0 |
| 10 | Ломоносовский | 102 | 33 |
| 11 | Лужский | 10 | 1 |
| 12 | Приозерский | 33 | 18 |
| 13 | Сланцевский | 3 | 0 |
| 14 | Тосненский | 39 | 9 |
|  | Итого | 618 | 237 |

Больше всего и домов, и коттеджей находятся во Всеволожском районе области.

### 3.2.12. Вопрос о муниципалитетах

В следующей таблице (табл. 3.14) представлены десять муниципалитетов, в которых находится наибольшее количество объектов недвижимости.

**Таблица 3.14.** Количество объектов недвижимости по муниципалитетам области.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Муниципалитет** | **Район области** | **Количество объектов** |
| 1 | Колтушское сельское поселение | Всеволожский | 51 |
| 2 | Куйвозовское сельское поселение | Всеволожский | 40 |
| 3 | Агалатовское сельское поселение | Всеволожский | 36 |

**Таблица 3.14.** Количество объектов недвижимости по муниципалитетам области (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Муниципалитет** | **Район области** | **Количество объектов** |
| 4 | Первомайское сельское поселение | Выборгский | 29 |
| 5 | Ропшинское сельское поселение | Ломоносовский | 28 |
| 6 | Аннинское городское поселение | Ломоносовский | 28 |
| 7 | Полянское сельское поселение | Выборгский | 27 |
| 8 | Рощинское городское поселение | Выборгский | 26 |
| 9 | Лесколовское сельское поселение | Всеволожский | 25 |
| 10 | Бугровское сельское поселение | Всеволожский | 24 |
|  | Итого | | 855 |

### 3.2.13. Линейная регрессия

Для построения линейной регрессии была выбрана зависимая переменная (*price\_per\_meter*) и несколько независимых: *meters*, *land*, *distance*, *floors*, *commodities, form, heating, material, wc\_site*.

В следующей таблице (табл. 3.15) представлены результаты построения этой модели.

**Таблица 3.15.** Результаты построения линейной регрессии.

|  |  |  |  |
| --- | --- | --- | --- |
| **Коэффициент детерминации** | **Скорректированный коэффициент детерминации** | **Стандартное отклонение** | **Размер выборки** |
| 0,3163 | 0,2927 | 36043 | 812 |

В следующей таблице (табл. 3.16) представлены коэффициенты при переменных регрессии.

**Таблица 3.16.** Коэффициенты регрессии.

|  |  |  |
| --- | --- | --- |
| **№** | **Переменная** | **Коэффициент** |
| 1 | Константа | 65422,0653 |
| 2 | *Meters* | -102,4828 |
| 3 | *Land* | 1774,3465 |
| 4 | *Form = ИЖС* | 837,1126 |
| *Form = Фермерское хозяйство* | -17270,6421 |
| *Form = Садоводство* | -4549,6283 |
| 5 | *Distance* | -154,5458 |
| 6 | *Heating = Без отопления* | -30849,9140 |
| *Heating = Дизельное* | 17272,0501 |
| *Heating = Камин* | -11700,5645 |
| *Heating = Печь* | -27867,7332 |
| *Heating = Твердотопливный котел* | -18075,0740 |
| *Heating = Угольное* | 10423,0518 |
| *Heating = Центральное газовое* | -320,6662 |
| *Heating = Электрическое* | -11058,7295 |
| 7 | *Material = Газосиликатный блок* | 58366,0943 |
| *Material = Деревянный* | -1563,4656 |
| *Material = Каркасный* | -5343,0600 |
| *Material = Кирпичный* | 5170,8201 |
| *Material = Монолитный* | -2043,1073 |
| *Material = Пенобетонный блок* | -29383,1119 |

**Таблица 3.16.** Коэффициенты регрессии (продолжение).

|  |  |  |
| --- | --- | --- |
| **№** | **Переменная** | **Коэффициент** |
| 7 | *Material = Щитовой* | -17038,5777 |
| 8 | *Wc\_site = На улице* | -3276,5806 |
| 9 | *Floors* | -10444,9195 |
| 10 | *Commodities* | 4936,2772 |
| 11 | *New\_homes* | 0,2083 |
| 12 | *Profitable\_firms* | 300,6356 |
| 13 | *Culture\_workers* | 370,9986 |
|  | *Form = ДНП* | 0 |
|  | *Heating = Автономное газовое* | 0 |
|  | *Material = Газобетонный блок* | 0 |
|  | *Wc\_site = В доме* | 0 |

Ввиду того, что коэффициент детерминации у модели составляет всего лишь 25,59 процентов, ее нельзя использовать при оценке объектов недвижимости.

### 3.2.14. Случайный лес

Также был построен случайный лес. Зависимой переменной была выбрана *price\_per\_meter*, независимыми - *commodities*, *distance*, *floors*, *form*, *heating*, *land*, *material*, *meters, new\_homes, culture\_workers, profitable\_firms*.

В следующей таблице (табл. 3.17) представлены характеристики построенной модели.

**Таблица 3.17.** Характеристики модели (случайный лес).

|  |  |  |
| --- | --- | --- |
| **№** | **Характеристика** | **Значение** |
| 1 | Относительная ошибка (relative error) | 26,30% |
| 2 | Множественный коэффициент корреляции | 0,542 |

Множественный коэффициент корреляции превышает 50 процентов, что по шкале Чеддока сигнализирует о «заметной» силе связи.

Приведем несколько примеров использования модели для оценки (табл. 3.18).

**Таблица 3.18.** Примеры использования модели (случайный лес).

|  |  |  |  |
| --- | --- | --- | --- |
| **Описание объекта** | **Цена за кв. м. в объявлении, рубли** | **Прогноз модели, рубли** | **Разница** |
| [Дом в деревне Силино, Красноозерное с/пос, Приозерский район](https://spb.cian.ru/sale/suburban/273776289/):   * 1 этаж * 89 кв. м. * 12 соток * Деревянный * Электрическое отопление * Все коммуникации * ИЖС * 73 км от КАД * Санузел в доме | 98876 | 91563 | 7,40% |
| [Коттедж во Всеволожске](https://spb.cian.ru/sale/suburban/261527869/):   * 3 этажа * 350 кв. м. * 25 соток * Кирпичный * Центральное газовое отопление * Все коммуникации * ИЖС * 6 км от КАД * Санузел в доме | 107143 | 124092 | 15,82% |

**Таблица 3.18.** Примеры использования модели (случайный лес) (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **Описание объекта** | **Цена за кв. м. в объявлении, рубли** | **Прогноз модели, рубли** | **Разница** |
| [Дом во Всеволожске](https://spb.cian.ru/sale/suburban/271900419/):   * 2 этажа * 150 кв. м. * 10 соток * Пенобетонный блок * Автономное газовое отопление * Все коммуникации * ИЖС * 11 км от КАД * Санузел в доме | 76667 | 83056 | 8,33% |

## 3.3. Исследование набора данных об объектах, сдающихся посуточно

Далее будет представлено исследования набора данных 2. Оно имело целью построение моделей, с помощью которых можно будет спрогнозировать стоимость суточной аренды определенного объекта недвижимости.

### 3.3.1. Описание данных

В набор данных 2 входят объекты загородной недвижимости, которые сдаются посуточно. В следующей таблице (табл. 3.19) представлены количественные признаки этих объектов.

**Таблица 3.19.** Количественные признаки объектов из набора данных 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Наименование (на английском языке)** | **Объяснение** | **Размерность** | **Промежуток** |
| 1 | *Meters* | Общая площадь жилой зоны (дома) | квадратный метр | [0; +∞) |
| 2 | *Land* | Общая площадь участка при жилой зоне (доме) | ар («сотка») | [0; +∞) |

**Таблица 3.19.** Количественные признаки объектов из набора данных 2 (продолжение).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Наименование (на английском языке)** | **Объяснение** | **Размерность** | **Промежуток** |
| 3 | *Price* | Стоимость суточной аренды объекта недвижимости | российский рубль | [0; +∞) |
| 4 | *Age* | «Возраст» дома | года | {1, 2, … 21} |
| 5 | *Bedrooms* | Количество спален в доме | спальни | {1, 2, 3…} |
| 6 | *Floors* | Количество этажей в доме | этажи | {1, 2, 3} |
| 7 | *Distance* | Дистанция от объекта до КАД (кольцевой автодороги) | километр | [0; +∞) |

В следующей таблице (табл. 3.20) представлены качественные признаки объектов из набора данных 2.

**Таблица 3.20.** Качественные признаки, набор данных 2.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Наименование (на английском языке)** | **Объяснение** | **Возможные значения** |
| 1 | *Road* | Ближайшая трасса | Кола, Нарва, Сортавала |
| 2 | *District* | Район Ленинградской области (один из семнадцати) | Выборгский, Всеволожский, Ломоносовский |
| 3 | *Settlement* | Муниципалитет в составе района области | Селезневское сельское поселение, Виллозское городское поселение, Ропшинское сельское поселение |
| 4 | *Locality* | Населенный пункт в муниципалитете | Деревня Яльгелево, город Всеволожск |

**Таблица 3.20.** Качественные признаки, набор данных 2 (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Наименование (на английском языке)** | **Объяснение** | **Возможные значения** |
| 5 | *Material* | Материал, из которого изготовлено жилое строение | Кирпичный, Монолитный |
| 6 | *Heating* | Тип отопления | Угольное, Автономное газовое |
| 7 | *wc\_site* | Расположение санузла | В доме, на улице |
| 8 | *Repair* | Тип ремонта | Косметический, Евроремонт, Дизайнерский, без ремонта |

Также в набор данных 2 входят признаки *new\_homes*, *culture\_workers* и *profitable\_firms*, описанные в п. 3.2.1.

### 3.2.2. Проверка на нормальное распределение

Во втором наборе данных есть четыре признака, имеющих непрерывное распределение: *meters, land, price, distance*. В этом разделе будет рассмотрено распределение этих признаков, является ли оно нормальным.

В следующей таблице (табл. 3.21) представлены данные о том, соответствует ли распределение нормальному согласно правилу трех сигм [4]. В трех колонках указана доля объектов набора данных, попадающих в промежуток [μ-n\*σ, μ+n\*σ] (μ – среднее, n – количество отклонений, σ – среднеквадратическое отклонение).

**Таблица 3.21.** Применение правила трех сигм для набора данных 2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№** | **Признак** | **1σ** | **2σ** | **3σ** | **Соответствий** |
| 1 | *Meters* | 68,84% | 97,63% | 98,81% | 2 из 3 |
| 2 | *Land* | 97,33% | 98,81% | 98,81% | 2 из 3 |
| 3 | *Price* | 90,21% | 96,16% | 98,22% | 2 из 3 |
| 4 | *Distance* | 69,73% | 96,74% | 97,92% | 2 из 3 |
|  | При нормальном распределении | ≥68,26% | ≥95,44% | ≥99,72% |  |

Таким образом, строгого соответствия правилу трех сигм не достигает ни один из признаков, однако от необходимых показателей полученным долям не хватает не более 2 процентов. Поэтому можно выбрать метод Тьюки для удаления выбросов.

В следующей таблице (табл. 3.22) представлены гистограммы, построенные для этих признаков.

**Таблица 3.22.** Гистограммы пяти признаков набора данных 2.

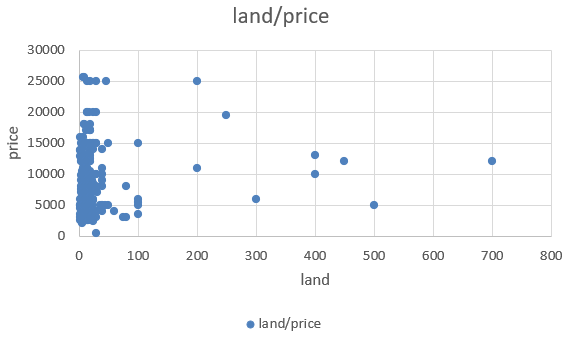
|  |  |  |
| --- | --- | --- |
| **№** | **Признак** | **Гистограмма** |
| 1 | *Price* |  |
| 2 | *meters* |  |
| 3 | *land* |  |

**Таблица 3.22.** Гистограммы пяти признаков набора данных 2 (продолжение).

|  |  |  |
| --- | --- | --- |
| **№** | **Признак** | **Гистограмма** |
| 4 | *distance* |  |

### 3.3.3. Проблема выбросов

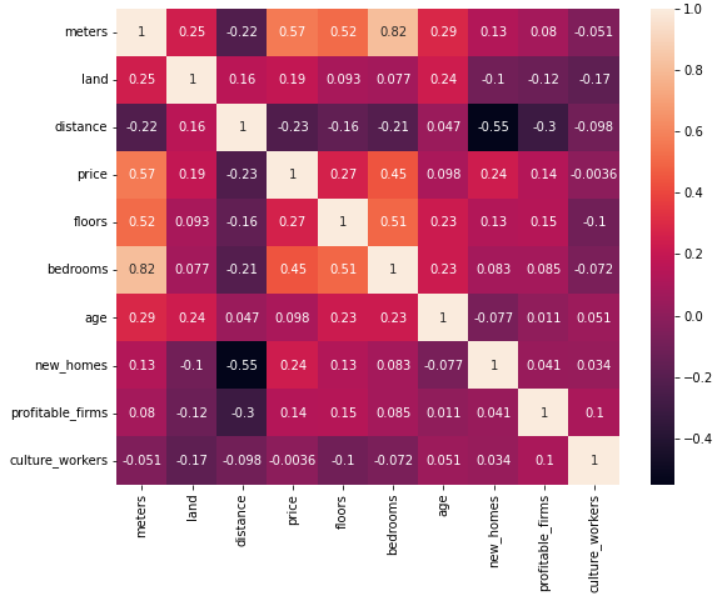
Предварительно были удалены выбросы с помощью метода Тьюки, а также ошибочные значения. Как и в предыдущем наборе данных, были удалены объекты со значением параметра *land*, превышающие 50. На следующем рисунке (рис. 3.3) изображено корреляционное поле для объектов, построенное по признакам *land* и *price*.



**Рис. 3.3.** Корреляционное поле для набора данных 2, построенное по признакам land и price.

### 3.3.4. Корреляционные связи

На следующем рисунке (рис. 3.4) изображена корреляционная матрица для набора данных 2.



**Рис. 3.4.** Корреляционная матрица для набора данных 2.

Оценим силу корреляционных связей по шкале Чеддока [4]. Между переменной *price* и:

* *meters* – 0,57, заметная связь;
* *land* – 0,19, слабая связь;
* *distance* - -0,23, слабая связь;
* *floors* – 0,27, слабая связь;
* *bedrooms* – 0,45, умеренная связь;
* *age* – 0,1, слабая связь;
* *new\_homes* – 0,24, слабая связь;
* *profitable\_firms* – 0,14, слабая связь.

Наблюдается высокая корреляционная связь между переменными *meters* и *bedrooms*. Во избежание мультиколлинеарности, переменная *bedrooms* не будет использована при построении моделей для оценки стоимости посуточной аренды объектов загородной недвижимости.

### 3.3.5. Вопрос о районах

В следующей таблице (табл. 3.23) представлены данные о средних ценах за квадратный метр в зависимости от района, в котором находится объект недвижимости (*district*).

**Таблица 3.23.** Средние цены за квадратный метр по районам области.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Район области** | **Количество объектов** | **Средняя цена, рубли** |
| 1 | Кировский | 1 | 12000 |
| 2 | Выборгский | 60 | 9297 |
| 3 | Всеволожский | 81 | 9192 |
| 4 | Ломоносовский | 26 | 7796 |
| 5 | Приозерский | 61 | 7512 |
| 6 | Гатчинский | 17 | 6953 |
| 7 | Лодейнопольский | 4 | 6325 |
| 8 | Тосненский | 8 | 5000 |
| 9 | Лужский | 13 | 4462 |
| 10 | Волховский | 3 | 4000 |
| 11 | Волосовский | 1 | 3000 |
|  | Итого | 275 | 8116 |

Наиболее «дорогими» для посуточной аренды являются Всеволожский и Выборгский районы (не берем в расчет Кировский, в котором находится всего лишь один дом, предназначенный для аренды). Всеволожский, Приозерский и Выборгский имеют больше всего объектов для посуточной аренды.

### 3.3.6. Вопрос о типах отопления

В следующей таблице (табл. 3.24) дана информация о средних ценах за квадратный метр в зависимости от типа отопления (*heating*), который есть в доме или коттедже.

**Таблица 3.24.** Средние цены по типам отопления объектов недвижимости.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип отопления** | **Количество объектов** | **Средняя цена за сутки рубли** |
| 1 | Центральное газовое | 33 | 11470 |

**Таблица 3.24.** Средние цены по типам отопления объектов недвижимости (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип отопления** | **Количество объектов** | **Средняя цена за сутки рубли** |
| 2 | Автономное газовое | 33 | 10812 |
| 3 | Дизельное | 6 | 8883 |
| 4 | Электрическое | 160 | 7482 |
| 5 | Угольное | 9 | 6638 |
| 6 | Камин | 26 | 5742 |
| 7 | Твердотопливный котел | 5 | 4659 |
| 8 | Печь | 3 | 4556 |
|  | Итого | 275 | 8116 |

Самым «популярным» типом отопления является электрическое. Однако самыми «дорогими» (для посуточной аренды домов и коттеджей) являются газовые типы отопления: автономное газовое и центральное газовое.

### 3.3.7. Вопрос о типах материала

В следующей таблице (табл. 3.25) дана информация о средних ценах за сутки в зависимости от типа материала (*material*), из которого построен дом или коттедж.

**Таблица 3.25.** Средние цены по типам материала объектов недвижимости.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип материала** | **Количество объектов** | **Средняя цена за сутки, рубли** |
| 1 | Газобетонный блок | 27 | 10009 |
| 2 | Пенобетонный блок | 2 | 9750 |
| 3 | Кирпичный | 20 | 9600 |

**Таблица 3.25.** Средние цены по типам материала объектов недвижимости (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип материала** | **Количество объектов** | **Средняя цена за сутки, рубли** |
| 4 | Щитовой | 2 | 9000 |
| 5 | Деревянный | 127 | 8131 |
| 6 | Каркасный | 87 | 7298 |
| 7 | Монолитный | 10 | 6450 |
|  | Итого | 275 | 8116 |

Так, наиболее дорогими для аренды посуточно являются дома из бетонных блоков: газобетонных и пенобетонных.

### 3.3.8. Случайный лес

Далее была построена модель, случайный лес. Как зависимая была выбрана переменная price. Как независимые – *age*, *culture\_workers, distance*, *floors*, *heating*, *land*, *material*, *meters, new\_homes, profitable\_firms, repair, wc\_site*. Из-за низкой корреляции, почти равной нулю, была исключена переменная *culture\_workers*. Из-за того, что 98,12 процентов объектов имеют одно и то же значение признака *wc\_site* («в доме»), эта переменная тоже была исключена.

В следующей таблице (табл. 3.26) представлены метрики построенной модели.

**Таблица 3.26.** Характеристики модели (случайный лес).

|  |  |  |
| --- | --- | --- |
| **№** | **Характеристика** | **Значение** |
| 1 | Относительная ошибка (relative error) | 29,30% |
| 2 | Множественный коэффициент корреляции | 0,536 |

Множественный коэффициент корреляции превышает 50 процентов (по шкале Чеддока [21] силу связи можно оценить как высокую).

В следующей таблице (табл. 3.27) представлены примеры использования построенной модели для оценки стоимости посуточной аренды для некоторых объектов загородной недвижимости.

**Таблица 3.27.** Примеры использования модели для оценки стоимости посуточной аренды.

|  |  |  |  |
| --- | --- | --- | --- |
| **Описание объекта** | **Цена в объявлении, рубли за сутки** | **Прогноз модели, рубли за сутки** | **Разница** |
| [Дом в деревне Касимово, Агалатовское сельское поселение, Всеволожский район](https://spb.cian.ru/rent/suburban/273827294/):   * 1 этаж * 250 кв. м. * 6 соток * 2015 года постройки * Пенобетонный блок * Автономное газовое отопление * Дизайнерский ремонт * 13 км от КАД | 10000 | 11320 | 13,20% |
| [Дом в деревне Борисово, Раздольевское сельское поселение, Приозерский район](https://spb.cian.ru/rent/suburban/267913192/):   * 1 этаж * 127 кв. м. * 10 соток * 2021 год постройки * Монолитный * Электрическое отопление * Дизайнерский ремонт * 72 км от КАД | 7500 | 7273 | 3,03% |

**Таблица 3.27.** Примеры использования модели для оценки стоимости посуточной аренды (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **Описание объекта** | **Цена в объявлении, рубли за сутки** | **Прогноз модели, рубли за сутки** | **Разница** |
| [Дом в посёлке Волочаевка, Рощинское городское поселение, Выборгский район](https://spb.cian.ru/rent/suburban/271923975/):   * 1 этаж * 68 кв. м. * 6 соток * 2020 год постройки * Деревянный * Камин * Санузел в доме * 61 км от КАД * Косметический ремонт | 5000 | 5568 | 11,36% |

## 3.4. Исследование набора данных об объектах, сдающихся помесячно

Далее будет представлено исследования набора данных 3. Оно имело целью построение моделей, с помощью которых можно будет спрогнозировать стоимость месячной аренды определенного объекта недвижимости.

### 3.4.1. Описание данных

По своей структуре (набору признаков) набор данных 3 идентичен набору данных 2. В нем содержатся те же самые признаки: *meters*, *land*, *road*, *distance*, *district*, *settlement*, *locality*, *price*, *heating*, *material*, *wc\_site*, *repair*, *floors*, *bedrooms*, *age*. Однако есть дополнительный признак *status* – обозначающий то, насколько долго можно арендовать объект недвижимости. Имеет два значения: «от года» и «несколько месяцев».

Также в набор данных 3 входят признаки *new\_homes*, *culture\_workers* и *profitable\_firms*, описанные в п. 3.2.1.

### 3.4.2. Проверка на нормальное распределение

Во втором наборе данных есть четыре признака, имеющих непрерывное распределение: *meters, land, price, distance*. В этом разделе будет рассмотрено распределение этих признаков, является ли оно нормальным.

В следующей таблице (табл. 3.28) представлены данные о том, соответствует ли распределение нормальному согласно правилу трех сигм [4]. В трех колонках указана доля объектов набора данных, попадающих в промежуток [μ-n\*σ, μ+n\*σ] (μ – среднее, n – количество отклонений, σ – среднеквадратическое отклонение).

**Таблица 3.28.** Применение правила трех сигм для набора данных 3.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№** | **Признак** | **1σ** | **2σ** | **3σ** | **Соответствий** |
| 1 | *Meters* | 93,30% | 98,88% | 99,44% | 2 из 3 |
| 2 | *Land* | 97,77% | 98,32% | 98,88% | 2 из 3 |
| 3 | *Price* | 96,65% | 98,88% | 98,88% | 2 из 3 |
| 4 | *Distance* | 75,42% | 96,09% | 98,88% | 2 из 3 |
|  | При нормальном распределении | ≥68,26% | ≥95,44% | ≥99,72% |  |

Таким образом, строгого соответствия правилу трех сигм не достигает ни один из признаков, однако от необходимых показателей полученным долям не хватает не более процента. Поэтому можно выбрать метод Тьюки для удаления выбросов.

В следующей таблице (табл. 3.29) представлены гистограммы, построенные для этих признаков.

**Таблица 3.29.** Гистограммы пяти признаков набора данных 3.

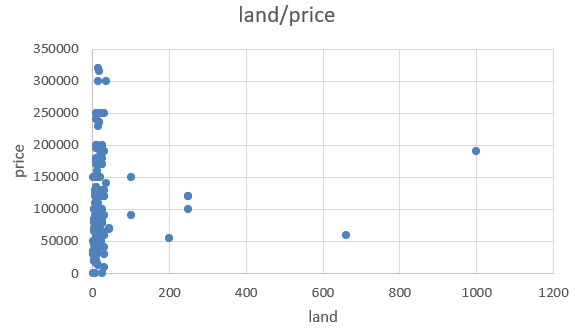
|  |  |  |
| --- | --- | --- |
| **№** | **Признак** | **Гистограмма** |
| 1 | *Price* |  |

**Таблица 3.29.** Гистограммы пяти признаков набора данных 3 (продолжение).

|  |  |  |
| --- | --- | --- |
| **№** | **Признак** | **Гистограмма** |
| 2 | *Meters* |  |
| 3 | *Land* |  |
| 4 | *Distance* |  |

### 3.4.3. Проблема выбросов

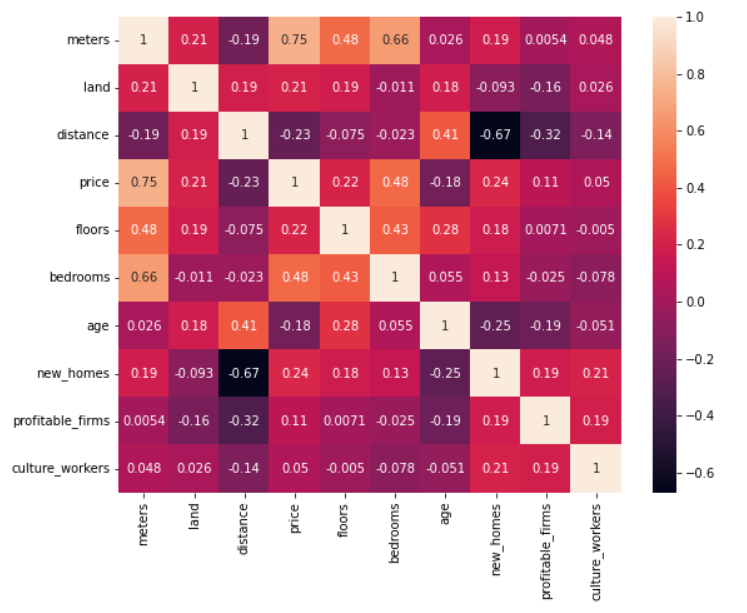
Как и в предыдущем наборе данных, был использован метод Тьюки, а также были удалены объекты со значением параметра *land*, превышающие 50. На следующем рисунке (рис. 3.5) изображено корреляционное поле для объектов, построенное по признакам *land* и *price*.



**Рис. 3.5.** Корреляционное поле для набора данных 3, построенное по признакам land и price.

### 3.4.4. Корреляционные связи

На следующем рисунке (рис. 3.6) изображена корреляционная матрица для набора данных 3.



**Рис. 3.6.** Корреляционная матрица, построенная для набора данных 3.

Оценим силу корреляционных связей с помощью шкалы Чеддока [4]. Между переменной *price* и:

* *meters* – 0,75, высокая связь;
* *land* – 0,21, слабая связь;
* *distance* - -0,23, слабая связь;
* *floors* – 0,22, слабая связь;
* *bedrooms* – 0,5, заметная связь;
* *age* - -0,17, слабая связь;
* *new\_homes* – 0,24, слабая связь;
* *profitable\_firms* – 0,11, слабая связь.

Наблюдается высокая корреляционная связь между переменными *meters* и *bedrooms*. Во избежание мультиколлинеарности, переменная *bedrooms* не будет использована при построении моделей для оценки стоимости посуточной аренды объектов загородной недвижимости.

### 3.4.5. Вопрос о районах области

В следующей таблице (табл. 3.30) представлены данные о средних ценах за квадратный метр в зависимости от района, в котором находится объект недвижимости (*district*).

**Таблица 3.30.** Средние цены за квадратный метр по районам области.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Район области** | **Количество объектов** | **Средняя цена, рубли** |
| 1 | Ломоносовский | 3 | 148000 |
| 2 | Всеволожский | 30 | 132000 |
| 3 | Выборгский | 28 | 114925 |
| 4 | Приозерский | 26 | 106958 |
| 5 | Гатчинский | 12 | 92083 |
| 6 | Тосненский | 3 | 66667 |
| 7 | Кингисеппский | 4 | 60000 |
| 8 | Волховский | 1 | 57000 |
| 9 | Кировский | 2 | 49500 |
| 10 | Лужский | 2 | 40000 |
|  | Итого | 111 | 109764 |

Наиболее «дорогими» является Всеволожский район области, в нём же находится больше всего объектов, сдающихся под помесячную аренду. По количеству объектов и средней цене к нему близки Выборгский и Приозерский районы.

### 3.4.6. Вопрос о типах отопления

В следующей таблице (табл. 3.31) дана информация о средних ценах за квадратный метр в зависимости от типа отопления (*heating*), который установлен на объекте недвижимости.

**Таблица 3.31.** Средние цены по типам отопления объектов недвижимости.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип отопления** | **Количество объектов** | **Средняя цена за месяц, рубли** |
| 1 | Дизельное | 1 | 169000 |
| 2 | Центральное газовое | 17 | 148824 |
| 3 | Автономное газовое | 18 | 146556 |
| 4 | Твердотопливный котел | 1 | 100000 |
| 5 | Электрическое | 59 | 99369 |
| 6 | Камин | 5 | 82800 |
| 7 | Печь | 9 | 48889 |
| 8 | Без отопления | 1 | 30000 |
|  | Итого | 111 | 109764 |

Самым «популярным» в этом сегменте типом отопления также является электрическое. Самыми «дорогими» (для помесячной аренды домов и коттеджей) являются снова газовые типы отопления: автономное газовое и центральное газовое (в расчет не берется дизельное отопление, которое имеет лишь один объект из набора).

### 3.4.7. Вопрос о типах материала

В следующей таблице (табл. 3.32) дана информация о средних ценах за квадратный метр в зависимости от типа материала (*material*), из которого построен объект недвижимости.

**Таблица 3.32.** Средние цены по типам материала объектов недвижимости.

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Тип материала** | **Количество объектов** | **Средняя цена за месяц, рубли** |
| 1 | Кирпичный | 10 | 156000 |
| 2 | Монолитный | 1 | 150000 |
| 3 | Газобетонный блок | 17 | 130294 |
| 4 | Деревянный | 55 | 100780 |
| 5 | Каркасный | 26 | 100535 |
| 6 | Пенобетонный блок | 1 | 57000 |
| 7 | Щитовой | 1 | 45000 |
|  | Итого | 111 | 109754 |

Так, наиболее дорогими для аренды посуточно являются кирпичные дома. Однако чаще всего сдаются посуточно деревянные дома.

### 3.4.8. Случайный лес

Далее была построена модель, случайный лес. Как зависимая была выбрана переменная *price*. Как независимые – *age*, *distance*, *floors*, *heating*, *land*, *material*, *meters, wc\_site, status*.

В следующей таблице (табл. 3.33) представлены метрики построенной модели.

**Таблица 3.33.** Характеристики модели (случайный лес).

|  |  |  |
| --- | --- | --- |
| **№** | **Характеристика** | **Значение** |
| 1 | Относительная ошибка (relative error) | 34,84% |
| 2 | Множественный коэффициент корреляции | 0,592 |

Множественный коэффициент корреляции превышает 50 процентов. По шкале Чеддока [21] такая сила связи может быть оценена как «заметная».

В следующей таблице (табл. 3.34) представлены примеры использования построенной модели для оценки стоимости посуточной аренды для некоторых объектов загородной недвижимости.

**Таблица 3.34.** Примеры использования модели для оценки стоимости помесячной аренды.

|  |  |  |  |
| --- | --- | --- | --- |
| **Описание объекта** | **Цена в объявлении, рубли за месяц** | **Прогноз модели, рубли за месяц** | **Разница** |
| [Дом в Кингисеппе](https://spb.cian.ru/rent/suburban/200488617/):   * 2 этажа * 100 кв. м. * 12 соток * 2017 год постройки * Деревянный * Электрическое отопление * На несколько месяцев * Санузел в доме * Косметический ремонт * 118 км от КАД | 80000 | 85047 | 6,31% |
| [Коттедж в посёлке Лебедевка, Красносельское сельское поселение, Выборгский район](https://spb.cian.ru/rent/suburban/272522617/):   * 2 этажа * 250 кв. м. * 30 соток * 2012 год постройки * Кирпичный * Электрическое отопление * На несколько месяцев * Санузел в доме * Косметический ремонт * 79 км от КАД | 250000 | 176386 | 29,45% |

**Таблица 3.34.** Примеры использования модели для оценки стоимости помесячной аренды (продолжение).

|  |  |  |  |
| --- | --- | --- | --- |
| **Описание объекта** | **Цена в объявлении, рубли** | **Прогноз модели, рубли** | **Разница** |
| [Дом в Колтушском сельском поселении, Всеволожский район](https://spb.cian.ru/rent/suburban/263627018/):   * 2 этажа * 180 кв. м. * 10 соток * 2021 год постройки * Каркасный * Электрическое отопление * От года * Санузел в доме * Дизайнерский ремонт * 12 км от КАД | 155000 | 127350 | 17,84% |

## Выводы

Были проанализированы признаки объектов недвижимости и их влияние на цену объектов. Было описано построение моделей, позволяющих предсказывать стоимость домов и коттеджей, а также стоимость их аренды, были даны их характеристики.

На основании проведенного исследования можно сделать следующие выводы в соответствии со сформулированными гипотезами и поставленными вопросами:

* в среднем, более дорогие участки находятся в Выборгском и Приозерском районах Ленинградской области. Однако больше всего объектов находится во Всеволожском и Ломоносовском районах;
* объекты с участками типа ДНП стоят дороже чем те, которые располагаются в садоводствах и ИЖС;
* самыми дорогими являются монолитные дома, самыми дешевыми – щитовые;
* самыми дорогими являются дома с дизельным, самыми дешевыми – с печным и без отопления;
* положительно на цену квадратного метра влияют: общая площадь участка, количество типов коммуникации в доме, количество введенных квадратных метров жилья в муниципалитете;
* отрицательно на цену квадратного метра влияет количество этажей в доме, расстояние от КАДа до объета;
* положительно на стоимость аренды объектов загородной недвижимости (как посуточной, так и помесячной) влияют: общая жилая площадь, количество этажей, количество введенных квадратных метров жилья в муниципалитете, процент прибыльных предприятий в муниципалитете;
* отрицательно на стоимость аренды дома влияет расстояние от КАДа до дома.

# Заключение

Итак, работа посвящена анализу рынка загородной недвижимости Ленинградской области. Были использованы данные, собранные на сайте cian.ru.

В первой главе была приведена информация о рынке загородной недвижимости Ленинградской области, об условиях, в которых он формируется, компаниях, которые в нем функционируют. Были описаны факторы, которые могут иметь влияние на стоимость объектов загородной недвижимости.

Во второй главе приведено описание методов машинного обучения, которые могут быть использованы в этом исследовании. Была разработана методология для следующего анализа рынка загородной недвижимости Ленинградской области на основе предыдущего опыта подобного рода исследований.

В третьей главе был проведен анализ рынка загородной недвижимости Ленинградской области. В рамках этого анализа было сделано следующее:

* собраны данные об объектах недвижимости, которые выставлены на продажу, сдаются посуточно или помесячно;
* проанализированы факторы, влияющие на стоимость объектов загородной недвижимости;
* построена модель, определяющая стоимость объектов загородной недвижимости;
* построена модель, определяющая стоимость посуточной аренды объектов загородной недвижимости;
* построена модель, определяющая стоимость помесячной аренды объектов загородной недвижимости.

Таким образом, все задачи этой работы выполнены, и поставленная цель достигнута. Было выяснено следующее:

* в среднем, более дорогие участки находятся в Выборгском и Приозерском районах Ленинградской области. Однако больше всего объектов находится во Всеволожском и Ломоносовском районах;
* объекты с участками типа ДНП стоят дороже чем те, которые располагаются в садоводствах и ИЖС;
* самыми дорогими являются монолитные дома, самыми дешевыми – щитовые;
* самыми дорогими являются дома с дизельным, самыми дешевыми – с печным и без отопления;
* положительно на цену квадратного метра влияют: общая площадь участка, количество типов коммуникации в доме, количество введенных квадратных метров жилья в муниципалитете;
* отрицательно на цену квадратного метра влияет количество этажей в доме, расстояние от КАДа до объета;
* положительно на стоимость аренды объектов загородной недвижимости (как посуточной, так и помесячной) влияют: общая жилая площадь, количество этажей, количество введенных квадратных метров жилья в муниципалитете, процент прибыльных предприятий в муниципалитете;
* отрицательно на стоимость аренды дома влияет расстояние от КАДа до дома.

# Список литературы

1. Авито [Электронный ресурс]. – Режим доступа: <https://avito.ru>, свободный. – Загл. с экрана.
2. База данных показателей муниципальных образований Ленинградской области [Электронный ресурс]. – Режим доступа: <https://www.gks.ru/dbscripts/munst/munst41/DBInet.cgi>, свободный. – Загл. с экрана.
3. Березинец И. В. Практикум по теории вероятностей и математической статистике / И. В. Березинец. – СПб.: Издательство «Высшая школа менеджмента», 2013 - 162 с.
4. Бизнес-статистика: учебник и практикум для академического бакалавриата / под ред. И. И. Елисеевой – М., Издательство Юрайт. – 2018. – 445 с.
5. Домбровский В. В., Эконометрика / В. В. Домбровский. – М.: Новый учебник, 2004. – 342 с.
6. Дьяконов А., Случайный лес (Random Forest) [Электронный ресурс] / dyakonov.org, 2016. – Режим доступа: <https://dyakonov.org/2016/11/14/%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D1%8B%D0%B9-%D0%BB%D0%B5%D1%81-random-forest/>, свободный. – Загл. с экрана.
7. Загородный дом против летней дачи [Электронный ресурс] / Интерфакс Недвижимость, 2021. – Режим доступа: <https://realty.interfax.ru/ru/analytics/comments/130402/>, свободный. – Загл. с экрана.
8. Конева Л. Ю. Экономико-географическое положение Ленинградской области [Электронный ресурс]. – СПб.: Северо-западная академия государственной службы, 2010 -. – Режим доступа: <https://works.doklad.ru/view/UyRpW6mvE2Q.html>, свободный. – Загл. с экрана.
9. Королев Р. И. История развития технологии парсинга / Р. И. Королев, Е. З. Никонова // Современное программирование. III Международная научно-практическая конференция. – 2020. – С. 169–172.
10. Крылова Д. Д., Использование методов регрессионного анализа при оценке стоимости недвижимости / Д. Д. Крылова, Р. Г. Абакумов, О. Н. Моргунова // Инновационная экономика: перспективы развития и совершенствования. – 2018. - №3 (29). – С. 88–93.
11. Кульгин М., Руководство по парсингу веб-сайтов в 2021 году [Электронный ресурс] / vc.ru, 2021. – Режим доступа: <https://vc.ru/marketing/239915-rukovodstvo-po-parsingu-veb-saytov-v-2021-godu>, свободный. – Загл. с экрана.
12. Лабинцев Е. Метрики в задачах машинного обучения [Электронный ресурс] / Хабр, 2017. – Режим доступа: https://habr.com/ru/company/ods/blog/328372/, свободный. – Загл. с экрана.
13. Ленинградская область [Электронный ресурс] / Википедия. – Режим доступа: https://ru.wikipedia.org/wiki/Ленинградская\_область, свободный. – Загл. с экрана.
14. Линейная регрессия в машинном обучении [Электронный ресурс] / Neurohive, 2018. – Режим доступа: <https://neurohive.io/ru/osnovy-data-science/linejnaja-regressija/>, свободный. – Загл. с экрана.
15. Определение выбросов [Электронный ресурс] / baguzin.ru, 2016. - Режим доступа: <https://baguzin.ru/wp/opredelenie-vybrosov/>, свободный. – Загл. с экрана.
16. Разница между СНТ, ДНП и ИЖС [Электронный ресурс] / City Sip. – Режим доступа: https://citysip.ru/informacija/raznica-mezhdu-snt-dnp-izhs/, свободный. – Загл. с экрана.
17. Рейтинг экономического развития районов Ленобласти – 2020 [Электронный ресурс] / Деловой Петербург, 2020 – Режим доступа: https://www.dp.ru/a/2020/07/31/Kak\_mi\_schitali, свободный. – Загл. с экрана.
18. С каким типом отопления стоит покупать загородный дом? [Электронный ресурс] / Garda Estate, 2019. – Режим доступа: https://garda-estate.ru/sovety/pokupka/s-kakim-tipom-otopleniya-stoit-pokupat-zagorodnyy-dom/, свободный. – Загл. с экрана.
19. Семенов П. С. Из какого материала лучше всего построить дом: как выбрать самый долговечный для частного и загородного жилья [Электронный ресурс] / Авирон Девелопмент. – Режим доступа: https://avirondevelopment.ru/blog/iz-kakogo-materiala-luchshe-vsego-postroit-dom-kak-vybrat-samyj-dolgovechnyj-strojmaterial/, свободный. – Загл. с экрана.
20. Тирская М. На рынке недвижимости под Петербургом – новый тренд [Электронный ресурс] / РБК, 2021 – Режим доступа: https://www.rbc.ru/spb\_sz/12/09/2021/61361b479a79473126714d41, свободный. – Загл. с экрана.
21. ЦИАН [Электронный ресурс]. – Режим доступа: https://spb.cian.ru, свободный. – Загл. с экрана.
22. Что такое Selenium? [Электронный ресурс] / Хабр, 2012. – Режим доступа: https://habr.com/ru/post/152653/, свободный.
23. Шугаев Г. Avito и ЦИАН заняли больше половины рынка онлайн-классифайдов недвижимости [Электронный ресурс] / Лента, 2021 – Режим доступа: https://lenta.ru/news/2021/06/17/avito\_zian/, свободный. – Загл. с экрана.
24. Яндекс.Карты [Электронный ресурс]. – Режим доступа: https://yandex.ru/web-maps/, свободный. – Загл. с экрана.
25. Descriptive Statistics / Parampreet K. et al // International Journal of Academic Medicine. – 2018. - № 4 (1).
26. House Priсe Prediction using Random Forest Machine Learning Technique / Adetunji A. B. et al // Procedia Computer Science: The 8th International Conference on Information Technology and Quantitative Management. – 2021. – 199.
27. Multiple regression analysis to predict the value of a residential building and to compare with the conventional method values / Shetty D. et al // Journal of Physics: Conference Series. – 2020. – 1706 012118.

# Приложения

### Приложение 1. Пример дерева из случайного леса (прогноз цены объекта)

Так, в модель случайного леса входит 140 построенных деревьев решений. Далее приведен одно из них из модели для прогнозирования цены квадратного метра объекта недвижимости.

|  |
| --- |
| material = Газобетонный блок  | distance > 16.500  | | land > 14.500: 150000.000 {count=3}  | | land ≤ 14.500  | | | culture\_workers > 10: 135000.000 {count=2}  | | | culture\_workers ≤ 10  | | | | age > 7  | | | | | heating = Камин: 65000.000 {count=3}  | | | | | heating = Электрическое: 80000.000 {count=3}  | | | | age ≤ 7: 40000.000 {count=2}  | distance ≤ 16.500  | | status = На несколько месяцев: 250000.000 {count=3}  | | status = От года: 205000.000 {count=4}  material = Деревянный  | meters > 145  | | culture\_workers > 3.500  | | | new\_homes > 21610.500  | | | | repair = Дизайнерский: 180600.000 {count=5}  | | | | repair = Косметический: 190000.000 {count=3}  | | | new\_homes ≤ 21610.500  | | | | land > 20  | | | | | meters > 175: 189500.000 {count=2}  | | | | | meters ≤ 175: 140000.000 {count=3}  | | | | land ≤ 20: 108300.000 {count=3}  | | culture\_workers ≤ 3.500: 300000.000 {count=2}  | meters ≤ 145  | | age > 9.500  | | | land > 18: 62500.000 {count=2}  | | | land ≤ 18  | | | | meters > 49.600: 37500.000 {count=2}  | | | | meters ≤ 49.600: 50000.000 {count=7}  | | age ≤ 9.500  | | | distance > 40.500  | | | | age > 5  | | | | | heating = Камин: 99000.000 {count=2}  | | | | | heating = Электрическое: 115000.000 {count=2}  | | | | age ≤ 5  | | | | | new\_homes > 20171.500: 98000.000 {count=2}  | | | | | new\_homes ≤ 20171.500: 70000.000 {count=2}  | | | distance ≤ 40.500  | | | | meters > 115: 113333.333 {count=3}  | | | | meters ≤ 115  | | | | | distance > 25: 22000.000 {count=3}  | | | | | distance ≤ 25: 65200.000 {count=5}  material = Каркасный  | meters > 95  | | distance > 67.500  | | | meters > 105: 110000.000 {count=2}  | | | meters ≤ 105: 80000.000 {count=2}  | | distance ≤ 67.500  | | | status = На несколько месяцев: 192000.000 {count=5}  | | | status = От года  | | | | meters > 145: 150000.000 {count=3}  | | | | meters ≤ 145: 124000.000 {count=2}  | meters ≤ 95  | | meters > 55: 68000.000 {count=5}  | | meters ≤ 55  | | | land > 13.500: 40000.000 {count=2}  | | | land ≤ 13.500  | | | | distance > 54.500: 32300.000 {count=3}  | | | | distance ≤ 54.500: 29000.000 {count=2}  material = Кирпичный  | profitable\_firms > 91.850: 250000.000 {count=2}  | profitable\_firms ≤ 91.850  | | floors > 2.500: 118750.000 {count=4}  | | floors ≤ 2.500  | | | status = На несколько месяцев: 150000.000 {count=2}  | | | status = От года: 200000.000 {count=2}  material = Монолитный: 150000.000 {count=3}  material = Щитовой: 45000.000 {count=3} |

### Приложение 2. Пример дерева из случайного леса (прогноз цены суточной аренды)

Так, в модель случайного леса входит 140 построенных деревьев решений. Далее приведен одно из них из модели для прогнозирования цены суточной аренды объекта недвижимости.

|  |
| --- |
| material = Газобетонный блок  | distance > 16.500  | | land > 14.500: 150000.000 {count=3}  | | land ≤ 14.500  | | | culture\_workers > 10: 135000.000 {count=2}  | | | culture\_workers ≤ 10  | | | | age > 7  | | | | | heating = Камин: 65000.000 {count=3}  | | | | | heating = Электрическое: 80000.000 {count=3}  | | | | age ≤ 7: 40000.000 {count=2}  | distance ≤ 16.500  | | status = На несколько месяцев: 250000.000 {count=3}  | | status = От года: 205000.000 {count=4}  material = Деревянный  | meters > 145  | | culture\_workers > 3.500  | | | new\_homes > 21610.500  | | | | repair = Дизайнерский: 180600.000 {count=5}  | | | | repair = Косметический: 190000.000 {count=3}  | | | new\_homes ≤ 21610.500  | | | | land > 20  | | | | | meters > 175: 189500.000 {count=2}  | | | | | meters ≤ 175: 140000.000 {count=3}  | | | | land ≤ 20: 108300.000 {count=3}  | | culture\_workers ≤ 3.500: 300000.000 {count=2}  | meters ≤ 145  | | age > 9.500  | | | land > 18: 62500.000 {count=2}  | | | land ≤ 18  | | | | meters > 49.600: 37500.000 {count=2}  | | | | meters ≤ 49.600: 50000.000 {count=7}  | | age ≤ 9.500  | | | distance > 40.500  | | | | age > 5  | | | | | heating = Камин: 99000.000 {count=2}  | | | | | heating = Электрическое: 115000.000 {count=2}  | | | | age ≤ 5  | | | | | new\_homes > 20171.500: 98000.000 {count=2}  | | | | | new\_homes ≤ 20171.500: 70000.000 {count=2}  | | | distance ≤ 40.500  | | | | meters > 115: 113333.333 {count=3}  | | | | meters ≤ 115  | | | | | distance > 25: 22000.000 {count=3}  | | | | | distance ≤ 25: 65200.000 {count=5}  material = Каркасный  | meters > 95  | | distance > 67.500  | | | meters > 105: 110000.000 {count=2}  | | | meters ≤ 105: 80000.000 {count=2}  | | distance ≤ 67.500  | | | status = На несколько месяцев: 192000.000 {count=5}  | | | status = От года  | | | | meters > 145: 150000.000 {count=3}  | | | | meters ≤ 145: 124000.000 {count=2}  | meters ≤ 95  | | meters > 55: 68000.000 {count=5}  | | meters ≤ 55  | | | land > 13.500: 40000.000 {count=2}  | | | land ≤ 13.500  | | | | distance > 54.500: 32300.000 {count=3}  | | | | distance ≤ 54.500: 29000.000 {count=2}  material = Кирпичный  | profitable\_firms > 91.850: 250000.000 {count=2}  | profitable\_firms ≤ 91.850  | | floors > 2.500: 118750.000 {count=4}  | | floors ≤ 2.500  | | | status = На несколько месяцев: 150000.000 {count=2}  | | | status = От года: 200000.000 {count=2}  material = Монолитный: 150000.000 {count=3}  material = Щитовой: 45000.000 {count=3} |

### Приложение 3. Пример дерева из случайного леса (прогноз цены месячной аренды)

Так, в модель случайного леса входит 140 построенных деревьев решений. Далее приведен одно из них из модели для прогнозирования цены месячной аренды объекта недвижимости.

|  |
| --- |
| material = Газобетонный блок  | distance > 16.500  | | land > 14.500: 150000.000 {count=3}  | | land ≤ 14.500  | | | culture\_workers > 10: 135000.000 {count=2}  | | | culture\_workers ≤ 10  | | | | age > 7  | | | | | heating = Камин: 65000.000 {count=3}  | | | | | heating = Электрическое: 80000.000 {count=3}  | | | | age ≤ 7: 40000.000 {count=2}  | distance ≤ 16.500  | | status = На несколько месяцев: 250000.000 {count=3}  | | status = От года: 205000.000 {count=4}  material = Деревянный  | meters > 145  | | culture\_workers > 3.500  | | | new\_homes > 21610.500  | | | | repair = Дизайнерский: 180600.000 {count=5}  | | | | repair = Косметический: 190000.000 {count=3}  | | | new\_homes ≤ 21610.500  | | | | land > 20  | | | | | meters > 175: 189500.000 {count=2}  | | | | | meters ≤ 175: 140000.000 {count=3}  | | | | land ≤ 20: 108300.000 {count=3}  | | culture\_workers ≤ 3.500: 300000.000 {count=2}  | meters ≤ 145  | | age > 9.500  | | | land > 18: 62500.000 {count=2}  | | | land ≤ 18  | | | | meters > 49.600: 37500.000 {count=2}  | | | | meters ≤ 49.600: 50000.000 {count=7}  | | age ≤ 9.500  | | | distance > 40.500  | | | | age > 5  | | | | | heating = Камин: 99000.000 {count=2}  | | | | | heating = Электрическое: 115000.000 {count=2}  | | | | age ≤ 5  | | | | | new\_homes > 20171.500: 98000.000 {count=2}  | | | | | new\_homes ≤ 20171.500: 70000.000 {count=2}  | | | distance ≤ 40.500  | | | | meters > 115: 113333.333 {count=3}  | | | | meters ≤ 115  | | | | | distance > 25: 22000.000 {count=3}  | | | | | distance ≤ 25: 65200.000 {count=5}  material = Каркасный  | meters > 95  | | distance > 67.500  | | | meters > 105: 110000.000 {count=2}  | | | meters ≤ 105: 80000.000 {count=2}  | | distance ≤ 67.500  | | | status = На несколько месяцев: 192000.000 {count=5}  | | | status = От года  | | | | meters > 145: 150000.000 {count=3}  | | | | meters ≤ 145: 124000.000 {count=2}  | meters ≤ 95  | | meters > 55: 68000.000 {count=5}  | | meters ≤ 55  | | | land > 13.500: 40000.000 {count=2}  | | | land ≤ 13.500  | | | | distance > 54.500: 32300.000 {count=3}  | | | | distance ≤ 54.500: 29000.000 {count=2}  material = Кирпичный  | profitable\_firms > 91.850: 250000.000 {count=2}  | profitable\_firms ≤ 91.850  | | floors > 2.500: 118750.000 {count=4}  | | floors ≤ 2.500  | | | status = На несколько месяцев: 150000.000 {count=2}  | | | status = От года: 200000.000 {count=2}  material = Монолитный: 150000.000 {count=3}  material = Щитовой: 45000.000 {count=3} |