

Приложение № 1

к договору №

от

Техническое задание

ИСПОЛНИТЕЛЬ

Самозанятый  
Владимировна

Кравцова

Кристина

ЗАКАЗЧИК

ООО "Застройщик в Казани"

Директор

\_\_\_\_\_/Кравцова К.В./

\_\_\_\_\_/\_\_\_\_\_/

## Оглавление

1. Общие положения .....	3
1.1 Наименование НИР .....	3
1.2 Формулировка проблемы.....	3
1.3 Нормативная документация.....	3
1.4 Источники разработки .....	4
1.5 Консультанты со стороны Заказчика .....	<b>Ошибка! Закладка не определена.</b>
1.6. Требования к патентной чистоте и лицензированию .....	4
2. Список терминов и определений .....	5
3. Список сокращений .....	5
4. Цели и задачи работы .....	5
5. Позиционирование .....	5
6. Входные и выходные данные .....	6
7. Функциональные требования .....	7
8. Ограничения .....	8
9. Допущения и зависимости .....	8
10. Системные требования и производительность.....	8
11. Атрибуты качества .....	9
12. Требования к защищенности .....	9
13. Требования к развертыванию .....	9
14. Документирование.....	10
15. Порядок приемки работ .....	10
16. Календарный план работ.....	<b>Ошибка! Закладка не определена.</b>
17. Приложения .....	<b>Ошибка! Закладка не определена.</b>

# 1. Общие положения

## 1.1 Наименование НИР

Калькулятор цены жилой недвижимости в Казани.

## 1.2 Формулировка проблемы

С ростом спроса на жилую недвижимость в Казани возникает необходимость в удобном и быстром способе расчета цены жилья для потенциальных покупателей.

Целевая аудитория - потенциальные покупатели и арендаторы жилья в Казани:

1. Возраст: 25-45 лет — возрастная группа, которая чаще всего занимается поиском жилья для себя или своей семьи.
2. Социальное положение: работающие люди среднего и выше среднего класса.
3. Цель использования калькулятора: быстрая оценка стоимости жилья перед принятием решения о покупке, инвестициях или планировании бюджета на будущую покупку недвижимости.
4. Географическое местоположение: жители или потенциальные покупатели недвижимости в Казани и прилегающих районах.

Существующие ресурсы для расчета цены жилья могут быть неудобными или не совсем точными. Недостаточно информации о рыночной стоимости жилья может привести к неправильным решениям при покупке. Данные, используемые в существующих калькуляторах, могут быть неактуальными. Необходимо создать удобный и достоверный калькулятор цены жилой недвижимости, который учтет все основные параметры и особенности рынка.

Основная бизнес-цель заказчика – увеличение комиссий агентств по продаже недвижимости.

Бизнес-требования и бизнес-правила: при определении стоимости жилой недвижимости первостепенными данными являются:

- Количество комнат
- Район города
- Год постройки дома

Помимо этого, на стоимость влияет доступность объектов социальной инфраструктуры, материал дома и качество его оборудования.

## 1.3 Нормативная документация

Федеральный закон "О государственной кадастровой оценке" от 03.07.2016 N 237-ФЗ (последняя редакция)

Федеральный закон "Об оценочной деятельности в Российской Федерации" от 29.07.1998 N 135-ФЗ (последняя редакция)

Законодательные акты и нормативные документы, регулирующие рынок недвижимости в Казани: полный перечень нормативно-правовых актов размещен на сайте ФИЛИАЛ ППК "РОСКАДАСТР" ПО РЕСПУБЛИКЕ ТАТАРСТАН ([https://roskadastr.tatarstan.ru/normativnye\\_aktvy\\_all.html](https://roskadastr.tatarstan.ru/normativnye_aktvy_all.html)) дата обращения 14.03.24

## 1.4 Источники разработки

В настоящее время крупные компании по продаже недвижимости предоставляют услуги по оценке квартир. Например, Циан.оценка использует следующие данные:

- адрес
- количество комнат
- площадь

И следующие подходы для получения оценки:

- Подбор и сравнение похожих объявлений в домах поблизости
- Доработке оценки с учётом десятков параметров: этаж, ремонт и другие
- Уточнение оценки на основе актуального тренда цен на рынке

Получение предсказания на цену недвижимости реализовано на сайте [irn.ru](http://irn.ru) (индикаторы рынка недвижимости), однако только для Москвы и Московской области. Калькулятор принимает такие параметры, как:

- Расположение
- Тип дома
- Этажность дома
- Этаж квартиры
- Количество комнат
- Общая площадь
- Площадь кухни
- Состояние квартиры

Результатом запроса является стоимость квартиры с такими параметрами и график его вероятного изменения за 3 месяца.

Методы оценки квартир аналитического центра IRN.RU основаны на использовании многомерной матрицы индексов стоимости жилья.

## 1.6. Требования к патентной чистоте и лицензированию

1. Проверка наличия патентов или других прав на интеллектуальную собственность, которые могут быть применимы к разрабатываемой модели калькулятора цены жилья.
2. Лицензирование необходимых технологий или программного обеспечения, которые будут использоваться в разработке калькулятора.
3. Соблюдение законодательства о защите данных и конфиденциальности.
4. Обеспечение правильной документации и оформления всех необходимых соглашений и договоров с разработчиками, партнерами и поставщиками услуг для обеспечения патентной чистоты и лицензирования проекта.

Получение данных с сайта [cian.ru](http://cian.ru) организовано при помощи библиотеки [cianparser](#), работающей под лицензией MIT.

## 2. Список терминов и определений

Датасет — обработанный и структурированный массив данных.

## 3. Список сокращений

## 4. Цели и задачи работы

Цель: Разработка модели оценки жилой недвижимости с отклонением цены в пределах 5 процентов

Задачи:

1. Анализ требований пользователей к калькулятору цены жилья
2. Сбор датасета на основе сайта cian.ru
3. Разработка алгоритма расчета цены жилья, учитывающего различные параметры (площадь, район, состояние квартиры и т.д.).
3. Создание удобного интерфейса калькулятора, который будет интуитивно понятен и легко использоваться пользователями.
4. Создание базы данных для хранения и актуализации информации о домах
5. Тестирование калькулятора на работоспособность, точность расчетов и удобство использования (модульное тестирование, интеграционное тестирование)
6. Оформление документации по проекту и подготовка к запуску веб-сервера.

## 5. Позиционирование

Модель калькулятора будет интегрирована в веб-сайт. Пользователи смогут заполнить необходимые параметры (площадь, район, состояние и прочее) и получить быстрый расчет ориентировочной стоимости жилья. Временные задержки по данным будут минимальными, так как информация о ценах на жилье будет обновляться регулярно. Работа конвейера будет организована таким образом, чтобы пользователи получали точные и актуальные данные о цене жилья в режиме реального времени.

Пользователь будет указывать желаемые критерии объекта недвижимости, заполняя поля веб-формы на сайте калькулятора.

На основании введенных данных будет формироваться строка датасета для предсказания стоимости.

Модель, решающая задачу регрессии, будет возвращать вещественную величину, обозначающую потенциальную стоимость квартиры с указанными параметрами. Стоимость будет отображаться на веб-странице калькулятора.

### 2. Условия эксплуатации:

Калькулятор должен быть доступен пользователям через веб-интерфейс, развернутый в Яндекс.Облаке и доступный по адресу, защищенному протоколом https .

### 3. Взаимодействие с другими системами и формат обмена данными:

Калькулятор может взаимодействовать с базами данных недвижимости, API сервисами для получения актуальной информации о ценах на жилье. Формат обмена данными — JSON.

## 6. Входные и выходные данные

### 1. Источники данных, используемых для обучения модели:

- Данные о ценах на жилье в различных районах Казани: открытые источники (avito.ru, cian.ru)
- Информация о площади, количестве комнат, состоянии жилья и других параметрах, влияющих на стоимость (ООО "Жилфонд Казань")

### 2. Датасеты для обучения будут формироваться при использовании библиотеки cianparser

- Признаки, получаемые в ходе сбора данных с предложений по продаже недвижимости
  - price - стоимость недвижимости
  - residential\_complex - название жилого комплекса
  - object\_type - тип жилья (вторичка/новостройка)
  - finish\_type - отделка
  - district - район
  - underground - метро
  - street - улица
  - house\_number - номер дома
  - floor - этаж
  - floors\_count - общее количество этажей
  - total\_meters - общая площадь
  - living\_meters - жилая площадь
  - kitchen\_meters - площадь кухни
  - rooms\_count - количество комнат
  - year\_construction - год постройки здания
  - house\_material\_type - тип дома (кирпичный/монолитный/панельный итд)
  - heating\_type - тип отопления
  - author - автор объявления
  - author\_type - тип автора
  - phone - номер телефона в объявлении
  - url - ссылка на объявление
- Также датасет будет содержать столбцы, указывающие на наличие социальной инфраструктуры:
  - Has\_school — школы
  - Has\_hospital — больницы
  - Has\_kindergarden — детские сады
  - Has\_shops — магазины
  - Has\_parks — парки
  - Has\_stops — остановки

## 1. Входные и выходные данные:

- Пользователь заполняет форму на сайте, результаты POST-запроса извлекаются на стороне сервера. Форма содержит следующие поля
  - Количество комнат (возможно несколько вариантов)
  - Общая площадь
  - Площадь кухни
  - Жилая площадь
  - Состояние квартиры (новостройка/вторичка/неважно)
  - Район
  - Диапазон стоимостей
  - Расстояние до метро
  - Этаж
  - Количество этажей в доме
  - Вид ремонта (косметический/евроремонт/дизайнерский/неважно)
  - Год постройки
  - Балкон (балкон/лоджия/нет/неважно)
  - Тип дома(кирпичный/деревянный/монолитный/панельный/блочный/кирпично-монолитный/сталинский/неважно)
  - Лифт (любой/грузовой/неважно)
  - Кухонная плита(газовая/электрическая/неважно)
  - Высота потолков
  - Парковка(наземная/многоуровневая/подземная/неважно)
  - Тип продавца(собственник/агент/застройщик/неважно)
  - Валюта результата

На стороне сервера в базе данных PostgreSQL хранятся картографические данные для определения наличия объектов социальной инфраструктуры в районе, в котором ищется недвижимость.

## 2. Требования к хранилищу данных и СУБД:

- Использование СУБД PostgreSQL для хранения данных о недвижимости, доступ к СУБД осуществляется при помощи JDBC-драйвера и библиотеки Spring-data.

## 3. Структура и формат результатов:

- Пользователь получает результат в виде числового значения в рублях. Результат отображается на веб-странице калькулятора под формой ввода параметров.

## 4. Дополнительные требования:

Датасет для обучения модели должен быть репрезентативным и отображать актуальную ситуацию на рынке недвижимости в Казани.

## 7. Функциональные требования

На стороне front-end:

1. Осуществлять прием входных данных от пользователя через форму на сайте;
2. Отправлять на сервер ввод пользователя в формате json;
3. Принимать ответ от сервера в формате json, распаковывать из него значение стоимости недвижимости и отображать на сайте.

На стороне back-end:

1. Принимать входные данные от клиента в формате json;
2. Запрашивать из базы данных информацию об инфраструктуре в указанном районе;
3. Извлекать факты наличия школ, детских садов, автобусных остановок, больниц, магазинов в указанном районе;
4. Формировать на основании них входные данные для модели в формате pandas DataFrame;
5. Запуск экземпляра модели машинного обучения;
6. Получать предсказание в виде вещественного числа;
7. Отправлять результат на клиент в формате json.

Архитектура решения:

Серверная часть будет состоять из двух микросервисов: микросервиса на python, который содержит модель МО, и микросервиса на java, который принимает web-запросы от клиента.

Необходимо решать задачу предсказания вещественной величины на основе набора параметров. Для этого будет использована реализация алгоритма градиентного бустинга XGBoost.

## 8. Ограничения

Аппаратные ограничения: не более 0.5 ГБ оперативной памяти на запущенный экземпляр модели МО.

Языки программирования и хранилища данных: Java, Python, СУБД PostgreSQL.

## 9. Допущения и зависимости

В проекте будут использованы библиотеки:

- Для языка python 3:
  - cianparser 1.0.4 для формирования датасета для обучения модели
  - pandas 2.2.2 для обработки датасета
  - xgboost 1.2.3 для формирования предсказания цены недвижимости
- Для языка java будет использован фреймворк Spring boot и в составе него библиотеки:
  - Spring-boot-starter
  - Spring-boot-data
  - Spring-boot-web

## 10. Системные требования и производительность

1. Доступность ресурсов: необходимо обеспечить надежный доступ к данным и вычислительным мощностям как на локальном ноутбуке, так и в облачной среде Яндекс.Облака.
2. Время реакции: приложение должно быстро отвечать на запросы пользователей, обеспечивая плавную работу калькулятора.
3. Пропускная способность: система должна быть способна обрабатывать большое количество запросов без потери производительности (1000 пользователей одновременно).
4. Предсказуемость систем реального времени: приложение должно предсказуемо работать в режиме реального времени, обеспечивая точные результаты.
5. Время доступа: минимизация времени доступа к данным и вычислениям для повышения эффективности работы калькулятора. (Не более 0.XX секунд),



## 11. Атрибуты качества

### 11.1 Атрибуты качества модели

На тестовой выборке модель должна удовлетворять следующим метрикам качества:

- MSE (среднеквадратичная ошибка):  $< 0,1$
- RMSE (корень из среднеквадратичной ошибки):  $< 0,3$
- SMAPE (симметричная средняя процентная ошибка):  $< 10\%$

### 11.2 Атрибуты качества обработки входных и выходных данных:

- Содержимое датасета для обучения модели должно быть релевантным и адекватным
- Датасет, подающийся на вход модели, не должен содержать неопределенных значений (NaN).

### 11.2 Надежность и доступность

- Среднее время наработки на отказ (MTTF) должно равняться 99,9% процентам времени (или 8760 часов в год).
- Среднее время восстановления после отказа (MTTR) должно составлять не более 1 часа.

### 11.3 Устойчивость:

- Веб-сервис должен гарантировать возможность одновременного доступа 100 пользователей и обрабатывать 100 запросов в минуту.

### 11.4 Скорость оперирования данными:

- Обработка пользовательского запроса должна занимать не более 2 секунд
- Время первичной загрузки страницы должно быть не более 5 секунд.

### 11.5 Отображение веб-интерфейса

- Веб-сервис должен корректно отображаться в основных веб-браузерах (Chrome, Firefox, Safari):
  - масштаб и расположение компонентов на экране должно позволять пользователю беспрепятственно установить значения фильтров;
  - результат, отображенный на странице веб-сервиса, должен быть явно различим и понятен пользователю;
- Выпадающие списки фильтров отображаются корректно.
- У всех форм текстового ввода присутствует валидация.

## 12. Требования к защищенности

Веб-сервис не будет обрабатывать или собирать биометрические данные пользователей или их персональные данные.

## 13. Требования к развертыванию

Приложение будет развернуто при использовании Docker-контейнера.

Для развертывания необходимо, чтобы на целевой сервер был установлен Docker, была создана учетная запись на Docker Hub и был доступен репозиторий Docker Hub, содержащему образы приложения.

## 14. Документирование

По окончании работ предъявляются следующие документы:

1. Отчет о НИР;
2. Руководство пользователя;
3. Инструкция по развертыванию приложения.

## 15. Порядок приемки работ

### 15.1. Тестирование

Для оценки качества необходимо провести следующие тесты:

Тест 1. Характеристики запроса:

- Выполнить запросы на предсказание цены с использованием различных комбинаций фильтров и параметров (например, площадь, количество комнат, район).
- Ожидаемый результат: цены на квартиры должны изменяться соответствующим образом при изменении фильтров.

Тест 2. Доверительный интервал:

- Выполнить запросы на предсказание цены для определенного района.
- Ожидаемый результат: предсказанная цена недвижимости должна попадать в доверительный интервал ожидаемой цены для данного района.

Тест 3. Отображение в различных браузерах

- Проверить работоспособность и корректность отображения контента веб-страницы в различных браузерах (Chrome, Firefox, Safari).
- Ожидаемый результат: веб-страница должна корректно отображаться и быть удобной в использовании во всех протестированных браузерах.

Тест 4. Проверка релевантности и корректности содержимого датасетов.

- Выполнить запросы на предсказание цены с использованием различных комбинаций фильтров и параметров (например, площадь, количество комнат, район).
- Ожидаемый результат: для всех комбинаций вычисленная стоимость не является аномальной (слишком низкой или высокой) и не дублируется для различных входных данных.

### 15.2. Условия приемки

Работа считается принятой, если:

- Результаты тестов, перечисленных в Разд. 15.1, удовлетворительны.
- Полученный результат соответствует критериям, описанным в разделе Атрибуты качества настоящего Технического задания.