Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Волгоградский государственный технический университет»

Факультет электроники и вычислительной техники

Кафедра электронных вычислительных машин и систем

|  |  |
| --- | --- |
| Согласовано  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (должность гл. специалиста предприятия) | Утверждаю  заведующий кафедрой ЭВМ и систем |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (подпись) (инициалы, фамилия) | \_\_\_\_\_\_\_\_ А.Е. Андреев  (подпись) (инициалы, фамилия) |
| «\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2021 г | «\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_2021 г |

**ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

к магистерской диссертации на тему

(наименование вида работы)

Геоинформационная система для анализа инвестиционной привлекательности городских территорий

Автор \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Скринкович Виталий Сергеевич

(подпись и дата подписания) (фамилия, имя, отчество)

Обозначение МД – 40 461 806 – 10.44 – 07–21ПЗ

(код документа)

Группа \_ЭВМ-2.1н

(шифр группы)

Направление \_\_\_\_\_\_09.04.01 Информатика и вычислительная техника

(код по ОКСО, наименование направления, программы)

Программа «Высокопроизводительные, высоконагруженные и облачные вычислительные системы»

Руководитель работы \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ М. А. Кузнецов

(подпись и дата подписания) (инициалы и фамилия)

Консультанты по разделам:

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

(краткое наименование раздела) (подпись и дата подписания) (инициалы и фамилия)

Нормоконтролер \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Г.А. Кашина

(подпись, дата подписания) (инициалы и фамилия)

Волгоград 2021 г.

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Волгоградский государственный технический университет»

Кафедра электронных вычислительных машин и систем

|  |  |
| --- | --- |
|  | Утверждаю  зав. кафедрой |
|  | \_\_\_\_\_\_\_\_\_\_ А.Е. Андреев  (подпись) (инициалы, фамилия) |
|  | «\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_20\_\_ г. |

Задание на магистерскую диссертацию

(наименование вида работы)

Студент Скринкович Виталий Сергеевич

(фамилия, имя, отчество)

Код кафедры 10.44 Группа ЭВМ-2.1Н

Тема Геоинформационная система для анализа инвестиционной привлекательности городских территорий

Утверждена приказом по университету от «01» ноября 2019 г. № 1344-СТ

Срок представления готовой работы \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(дата, подпись студента)

Исходные данные для выполнения работы

Документация Django, OpenStreetMap, Wikimapia, GeoPy, 2Gis

Содержание основной части пояснительной записки

ВВЕДЕНИЕ

1 Обзор задач принятия решений на основе анализа привлекательности территорий \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2 Разработка модели принятия решения\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3 Разработка системы поддержки принятия решения на основе открытых ГИС 4 Использование разработанной системы\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ЗАКЛЮЧЕНИЕ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Перечень графического материала

1. Изображение программы “ ГИС «ИНВЕСТОРА» г. ИРКУТСКА
2. ГИС для выявления городских районов с более высоким потенциалом для социальных инвестиций
3. Тепловые карты концентрации компаний в г.Волгограде

Руководитель работы \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Кузнецов М. А.

(подпись и дата подписания) (инициалы и фамилия)

АННОТАЦИЯ

В ходе данной работы была создана система поддержки принятия решений на основе данных, собранных из открытых картографических и юридических источников. В системе использовалась модель попарного сравнения альтернатив ELECTRE второго типа, что позволило выполнять многокритериальную оценку.

Работа состоит из 85 страниц, содержит 15 рисунок, 35 источников.

ABSTRACT

During this work, a decision support system has been created. It is based on collected from open cartographic and legal sources. The system uses a pairwise comparison model of the alternatives ELECTRE of the second type, which performs a multi-criteria assessment of the typ.

The work consists of 85 pages, contains 15 figures, 35 bibliographic sources

Оглавление

[1 Обзор задач принятия решений на основе анализа привлекательности территорий 10](#_Toc74702581)

[1.1 Обзор прикладных областей, в которых необходимо использование систем поддержки принятия решений на основе ГИС 10](#_Toc74702582)

[1.2 Основные принципы определения привлекательности территории 12](#_Toc74702583)

[1.2.1 Оценка влияния ландшафта и структуры городской застройки 13](#_Toc74702584)

[1.2.2 Оценка влияния территории на психологоповеденческие и социальные факторы 14](#_Toc74702585)

[1.2.3 Оценка влияния городской инфраструктуры 16](#_Toc74702586)

[1.3 Способы получения информации для анализа 17](#_Toc74702587)

[1.3.1 Поисковые системы 20](#_Toc74702588)

[1.3.2 Предварительная обработка данных 21](#_Toc74702589)

[1.3.3 Специализированные поисковые системы 22](#_Toc74702590)

[1.4 Исследование технологий и аналитических ГИС 23](#_Toc74702591)

[1.5 Модели принятия решений 27](#_Toc74702592)

[1.6 Постановка задачи на разработку специализированной системы поддержки принятия решений на основе открытых ГИС 29](#_Toc74702593)

[2 Разработка модели принятия решения 30](#_Toc74702594)

[2.1 Модели представления информации 30](#_Toc74702595)

[2.1.1 Численные данные из открытых систем 33](#_Toc74702596)

[2.1.2 Статистические данные 34](#_Toc74702597)

[2.1.3 Нечеткие множества и нечеткие оценки 35](#_Toc74702598)

[2.1.4 Нормализация и стандартизация 37](#_Toc74702599)

[2.2 Математические модели для вычисления критериев оценки привлекательности территории на основе открытых данных ГИС 39](#_Toc74702600)

[2.2.1 Экономическая активность территории 39](#_Toc74702601)

[2.2.2 Населенность территории 41](#_Toc74702602)

[2.2.3 Критерий оценки посещаемости территории 42](#_Toc74702603)

[2.2.4 Насыщенность территории объектами культуры 43](#_Toc74702604)

[2.2.5 Учет фактора конкуренции 44](#_Toc74702605)

[2.3 Математические модели для многокритериального принятия решения 45](#_Toc74702606)

[2.3.1 Тепловые карты и когнитивный анализ карты экспертами 45](#_Toc74702607)

[2.3.2 Кривые безразличия 46](#_Toc74702608)

[2.1.3 Механизмы сверток 47](#_Toc74702609)

[2.1.4 Методики попарного сравнения 50](#_Toc74702610)

[3 РАЗРАБОТКА СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЯ НА ОСНОВЕ ОТКРЫТЫХ ГИС 54](#_Toc74702611)

[3.1 Разработка подсистемы извлечения данных из открытых ГИС 54](#_Toc74702612)

[3.2 Алгоритмы вычисления критериев 55](#_Toc74702613)

[3.3 Разработка базы данных для хранения промежуточных данных и представления результатов расчета 56](#_Toc74702614)

[3.4 Разработка подсистемы многокритериального ранжирования 58](#_Toc74702615)

[3.5. Разработка пользовательского интерфейса 59](#_Toc74702616)

[3.6 Программная архитектура системы поддержки принятия решения на основе ГИС 60](#_Toc74702617)

[3.7 Описание и причины использования выбранных технологий 64](#_Toc74702618)

[3.8 Оценка свойств разработанного программного обеспечения 69](#_Toc74702619)

[3.8.1 Оценка надежности программного обеспечения 69](#_Toc74702620)

[3.8.2 Оценка масштабируемости программного обеспечения 70](#_Toc74702621)

[4 ИСПОЛЬЗОВАНИЕ РАЗРАБОТАННОЙ СИСТЕМЫ 71](#_Toc74702622)

[4.1 Выбор территории для застройки 71](#_Toc74702623)

[4.2 Выбор торгового павильона для аренды 73](#_Toc74702624)

[Заключение 79](#_Toc74702625)

[Список литературы 81](#_Toc74702626)

Введение

Когда инвестор планирует открыть точку розничной продажи продукции, перед ним встает очевидный вопрос – какое место будет наиболее эффективно доставлять продукцию до клиента? Для бизнеса важно, чтобы его целевая аудитория находилась в непосредственной близости дислокации торговой точки, что может быть обеспечено расположением магазина в области активного движения людей. Такими зонами можно считать территорию в промежутке между остановкой и многоэтажной застройкой или парковая зона. Однако это не единственный фактор, влияющий на выбор местоположения, о чем будет подробно рассмотрено позже.

Ещё одной регулирующей выбор местоположения можно рассматривать насыщенность территории конкурентами. На первый взгляд можно решить, что конкуренция вредна для бизнеса, однако это мнение обманчиво – дело в том, что при отсутствии конкурентов возможна ситуация как монополии на территории, так и полного отсутствия экономической жизни, что тоже является негативным фактором. Ко всему прочему есть шанс, что отсутствие в локальной области торговых точек лишь свидетельствует о том, что множество крупных магазинов уже боролась за эту территорию и на данный момент установилось известное равновесие Нэша. О таком может свидетельствовать группа крупных магазинов, сгруппированных в некой области не далеко от рассматриваемой территории. Возможна также ситуация, когда успешным решением будет вариант составления конкуренции таким магазинам. Подробнее эта ситуация будет также более подробно рассмотрена

Не исключается и вероятность ситуации физической близости магазина, однако его реальной недоступности по причинам, связанным с особенностями искусственного или естественного ландшафта территории.

Поиск территории вручную может быть весьма трудо и время-затратным, в связи с этим, целью данной работы являлась оценка общей привлекательности территорий на основе многих факторов.

Для достижения данной цели необходимо решить следующие задачи:

1. Обзор и анализ предметной области, а также существующих методов и средств, используемых и уже использованных в решении подобных задач;
2. Разработка программного обеспечения, осуществляющего сбор необходимых данных из открытых источников;
3. Разработка математической модели, позволяющей выполнять требуемы оценки;
4. Оценка результатов, полученных в ходе экспериментов;

Для достижения поставленных задач были выполнены следующие действия:

1. Разработано программное обеспечение, осуществляющее сбор данных о местности и организациях, расположенных в требуемом радиусе, а также остановок общественного транспорта;
2. Разработано программное обеспечение, включающее математическую модель, использующую экономические, географические данные и иные оценки для выполнения необходимых расчетов;
3. Создана программа, осуществляющая визуализацию данных для удобного представления пользователю картографической информации;
4. Разработан программный интерфейс, посредством которого система может выдавать пользователю результат в удобной форме;

# 1 Обзор задач принятия решений на основе анализа привлекательности территорий

# 1.1 Обзор прикладных областей, в которых необходимо использование систем поддержки принятия решений на основе ГИС

Когда инвестор планирует открыть пространство, ему соответственно требуется сделать вложения капитала, что является рискованным действием. Для минимизации риска проводятся изучение рынка: изучается опыт коллег, чтение новостей, проводятся опросы, регистрируется посещаемость выбранной улицы с помощью специального прибора. Существуют и иные методы сбора необходимых для принятия решения о дальнейших действиях данных, однако ключевой вопрос остаётся актуален: достаточно ли мы собрали информации и достоверно ли она отражает реальность? Классические методы анализа рынка особенно самые простые, доступные любому человеку на самом деле отличаются низкой эффективностью и в условиях жесткой конкуренции и наличии крупных торговых сетей любого характера, располагающих огромными ресурсами они просто не могут показывать результаты большой надежности.

Для решения описанной проблемы требуется проводить намного более серьезные варианты анализа. Аналитика должна быть не просто глобальной, охватывающей многие аспекты но и на уровень более достоверной, чем доступные методики.

Вариантом решения этой задачи предлагается проведение сбора географических данных и построение разных видов моделей на основе полученных данных. Такой подход позволяет даже при самых неудачных методиках работы получать весьма неожиданные результаты. Причиной тому является в первую очередь факт того, что собранные географические данные имеют в своём составе параметры, которые не могут быть доступны при обычных наблюдениях. Эту ситуацию можно сравнить с поиском слабых мест в утеплении строения: можно сутками ходить по зданию с наружи и внутри в поисках тепловых мостов, но наличие даже плохого тепловизора сделает работу в десяток раз качественнее и быстрее.

Примерами случаев, когда географическая система может помочь поменять ход принятия решения можно считать планирование местоположения отдельного заведения. Предположим, вы хотите открыть школу для проведения курсов и обучения определенным навыкам некоторую группу людей. Проведение классической аналитики потребует от вас сперва поиск обычными поисковыми системами конкурентов с вероятностью упустить кого-либо. После чего поиск на страницах или рекламе каждого адреса их точек, расположение их на карте и только после этого предположения о благоприятности определенной зоны. При этом есть вероятность переоценить значимость одних заведений, недооценить значимость других, а факторов как посещаемости, благосостояния населения или наличия в соседстве областей, насыщенных пользой для людей, где они вынуждены появляться часто для удовлетворения обязательных потребностей (магазин или аптека) вы можете и не учесть вовсе.

Ещё одним примером можно считать ситуацию, когда вы хотите открыть кофейню по типу ”coffee to go”. При планировании размещения подобного типа заведения вы точно проведёте замеры посещаемости улицы на предполагаемом месте размещения кофейни, однако посещаемость эта будет необходимого ли качества? Вполне возможно, что посещаемость обеспечивается бабушками на промежутке между поликлиникой и аптеками. Бабушка с малой вероятностью зайдет в “крафтовую” кофейню, с ценой кофе в половину пенсии, хотя организация по этой точке даже второй, конкурирующей аптеки может оказаться весьма неплохим решением. Подобные ошибки, а главное задачи поиска наиболее оптимальных точек для создания отдельных типов бизнеса и услуг и есть ключевая задача, рассматриваемая в нашем анализе.

# 1.2 Основные принципы определения привлекательности территории

Как уже было ранее указано, существует множество параметров оценки территории. В каждом конкретном случае с точки зрения открывающейся локации предприятия каждый из них может как быть полезным, так оказаться вредным. Перечислим в общих чертах доступные теоретически критерии для оценки методом географического моделирования территории:

Наличие и расстояние до ближайшей остановки транспорта общественного типа. Использование этого критерия в расчетах позволяет опосредованно оценивать посещаемость локальной области.

Наличие и расстояние до ближайших организаций бюджетного характера, государственных предприятий и иных не ориентированных на предоставление услуг областей, привлекающих людей. Наличие подобных в целом благоприятно может сказаться на возможности оказания услуг частным лицам.

Наличие и расстояние до ближайших организаций, основанных с целью извлечения прибыли путем оказания разного типа услуг. Критерий намного более сложным образом, влияющий на принятие решения, поскольку подобного рода компании могут как быть конкурентами, так и благоприятными соседями.

Наличие, концентрация жилых домов. Важный параметр, характеризующий общее наличие потенциальных клиентов. Близость парковых и прогулочных зон увеличивает вероятность посещения заведений развлекательного характера.

Средняя цена квадратного метра жилой площади может характеризовать материальное состояние проживающего контингента.

Этажность застройки характеризует плотность заселенности.

Экологические параметры могут оказывать не очевидное влияние

Общая прибыльность организаций в области может характеризовать экономический климат зоны.

Перечисленные критерии вовсе не все, которые можно использовать при оценке территории, однако важно заметить, что каждый из них может оказывать влияние и быть значимым в разной степени. Если при сравнении двух точек ценность каждой из них ещё можно предположить эмпирически, то вопрос оценки встает более остро при сравнении например 14 областей. При начальном изучении рынка возможность автоматизировано получить оценку привлекательности для нескольких зон позволит сэкономить ресурсы за счет исключения подробного изучения заведомо неблагоприятных вариантов.

# 1.2.1 Оценка влияния ландшафта и структуры городской застройки

Наличие ландшафта природного и неприродного характера может влиять на ценность области. Расположение магазина в низине может как быть полезным при наличии рекреационных точек привлечения потенциальных клиентов, так и быть вредным, поскольку без лишней необходимости разумный потребитель не заинтересован в совершении действий.

Если оценка природного ландшафта предсказуема, то намного больший эффект намного менее очевидным образом оказывает ландшафт рукотворный. В формат такого можно отнести любого вида заграждения: заборы, длинные дома, дороги без пешеходных переходов или, как неожиданно пустые пространства, не имеющие источников тени, лавочек или иных следствий благоустройства территории. В случае последних оценку географическими системами можно вести только опосредованно : наблюдая отсутствие некоторых факторов, например низко количество располагающихся компаний.

Крайне важно оценивать параметр ландшафта, поскольку при построении модели может быть дана оценка близкого расположения благоприятного фактора, однако на деле по причине наличия границ реальное расстояние до влияющих объектов может быть намного больше. Соответственно их значимость сильно снижается.

Любой, кто подробно исследует поверхность Земли, будет поражен тем, как условия имеют тенденцию сохраняться локально и разделять поверхность на области, которые демонстрируют существенное внутреннее сходство. Например, пустынные регионы характеризуются отсутствием осадков, тропические леса - обильными дождями и густой растительностью, а полярные регионы - экстремальными холодами. Условия в близлежащих точках не выбираются независимо и случайным образом из глобальных распределений, а вместо этого демонстрируют значительные уровни взаимозависимости. Конечно, есть исключения, когда условия меняются очень быстро на небольших расстояниях, например, между равнинами Индии и высокими Гималаями или между прибрежными равнинами и прилегающим океаном.

Общий термин для этого явления - пространственная зависимость. Его распространенность была точно отражена Тоблером в том, что стало известно как его Первый закон географии: «Все вещи взаимосвязаны, но близлежащие предметы более взаимосвязаны, чем далекие». Величину эффекта можно измерить с помощью ряда статистических данных пространственной автокорреляции. Он также лежит в основе дисциплины, известной как геостатистика, которая описывает пространственные вариации в терминах функции (известной как коррелограмма), которая показывает, как пространственная автокорреляция уменьшается с увеличением расстояния.

# 1.2.2 Оценка влияния территории на психологоповеденческие и социальные факторы

Покупательское поведение представляет собой сочетание экономических, технологических, политических, культурных.

Геомаркетинг позволяет ориентироваться на потенциальных клиентов с большим потенциалом в данной области торговли. Исходя из принципа, что место проживания частично отвечает за покупательское поведение, можно, например, составить архетипический профиль лучших клиентов, наложив продажи по площади и данные о жителях, проживающих в этих зонах. Этот профиль может позволить ориентироваться на аналогичные группы населения, не являющиеся клиентами, и реагировать на гипотезу о том, что люди, живущие в одном районе или на одной территории, обладают относительно схожими социально-экономическими и культурными характеристиками. Демографические явления слишком часто рассматриваются как объекты постоянства и непрерывности, которыми они не всегда обладают. Возраст данных также является слабым местом этого типа подхода. Отчеты о количестве предприятий на 1000 жителей или торговых площадей на 1000 жителей, рассчитанные либо в глобальном масштабе, либо путем отделения продаж гипермаркетов и супермаркетов от остальной торговли, часто дают общую картину ситуации путем выделения продуктов питания, личная одежда, бытовая техника и культура или досуг. Недостаток этого метода заключается в том, что он не опускается до уровня розничной торговли, а остается на уровне большой категории при подсчете количества квадратных метров и количества магазинов в рамках деятельности. Кроме того, понятие насыщения рынка, определяемое достаточным количеством магазинов для удовлетворения потребностей населения, должно учитывать внешнюю привлекательность рассматриваемого района, что может быть чрезвычайно важным. Кроме того, расчет продаж на квадратный метр, когда он объединяет гипермаркеты и супермаркеты, не является очень сильным показателем, поскольку он не определяет различия между продуктами питания и непродовольственными товарами для этих форм продаж. Наконец, «теоретический» рынок кажется связанным с жителями этого района, поскольку поездки по работе, здоровью, отдыху и т. д.

Удобное расположение организации - один из ключевых факторов успеха торговой организации. Предприниматели при выборе места для открытия бизнеса ориентируются на степень насыщенности тем или иным товаром на конкретной территории и планируют открытие с точки зрения приемлемой рентабельности. Посещаемость и прогнозируемый товарооборот. Эти два показателя связаны между собой: чем больше посещаемость, тем в большинстве случае выше предполагаемый оборот товаров.

Важно учесть и то, на что ориентируются потенциальные покупатели. Если в районе живут богатые люди, их не заинтересуют сэкондхэнды и одежда, произведенная неизвестными брендами. Такой класс общества предпочитает приобретать вещи в дорогих бутиках. Необходимо учесть специфику магазина с точки зрения эксклюзивности товара (услуги) и частоты потребления продукции покупателями. Чем эксклюзивнее продукция, тем больше времени покупатели готовы тратить на дорогу, чем чаще они потребляют этот продукт – тем ближе его желают купить. Поэтому, прежде чем открыть магазин, необходимо четко понимать, к какому виду продукции относятся товары, представленные в нем. От этого напрямую будет зависеть его местоположение. Большое значение имеет и подходящая целевая аудитория потока. Например, возле ВУЗов трафик практически всегда очень высок, при этом молодежь не делает значимых покупок — их средний чек несущественен. Сам трафик в дневное время может достигать 1000 человек в час.

# 1.2.3 Оценка влияния городской инфраструктуры

Наличие инфраструктуры как таковой повышает ценность территории в любом случае. Не секрет, что наличие транспортного узла повышает цену близко расположенных квартир. Можно с уверенностью сказать, что любой параметр связанный с инфраструктурой сильно влияет на свойства территории и на другие параметры. При повышении благоустройства территории, увеличении пассажиропотока в области и появлении транспортных остановок меняется интенсивность других факторов. Вместе они увеличивают характеристики ценности для всех видов бизнеса. Наличие достопримечательностей так-же может служить регулирующим фактором при оценке территории

Вопрос конкуренции в целом стоит отметить не очевидным. На первый взгляд обыватель скажет “конкуренция вредит прибыли” и будет прав с формальной точки зрения, однако в торговых центрах кафе часто располагаются в фудкортах, что напрямую противоречит тезису о необходимости избегания конкуренции. Причиной этому может быть несколько факторов: например клиенты предпочитают пройти в точку, где у них будет выбор в ассортименте, чем потратить ресурсы на передвижение в область с кафе с выбором куда меньшим.

Ещё одним вариантом можно считать классическое понимание равновесия Нэша. При его учете следует понимать, что хотя для покупателей будет куда более удобным равномерное распределение точек предоставляющих услуги, сами магазины заинтересованы в максимизации прибыли. Сдвиг одного из магазинов в центр области населения людей среди расположенных равномерно повлечет значительно сильное изменение распределения покупателей, поскольку указанный магазин будет охватывать большую область. Соответственно равновесным состоянием будет считаться нахождение всех магазинов в центре региона.

Последним очевидным параметром можно считать экономический: Строительство отдельного сооружения сопряжено с множеством тормозящих процесс действий. Экономически более выгодно строительство одного большого здания и единоразовые расходы на его проектирование и планирование, чем строительство каждого здания по отдельности с прохождением всех этапов разработки каждым экономическим субъектом.

# 1.3 Способы получения информации для анализа

На сегодняшний день ни один научный проект не обходится без получения данных для исследования. В случае, если мы проводим анализ показателей физических явлений, нам требуется провести сбор данных с датчиков, оцифровать их и только после этого мы сможем провести точные и удобные исследования. При производстве анализа социальных и экономических показателей распространен метод опросов. Результаты каждого должны быть сохранены для дальнейшей аналитики.

Оба указанных случая являются частными случаями сбора научных данных. В случае, когда мы осуществляем сбор экономических показателей и должны проводить расчеты нам требуется крайне большой объем данных для работы, сбор которых ручными методами будет крайне время затратным. В таких случаях как наш разумно прибегнуть к разработке программного обеспечения, выполняющую функцию сбора требуемых данных. Такой подход позволяет несмотря на труд во время процесса разработки ПО получить актуальные данные для анализа и обновить их в случае необходимости.

Для автоматического извлечения данных с веб-сайтов необходимо написать компьютерную программу с учетом требований проекта. Эта компьютерная программа может быть написана с нуля на языке программирования или может представлять собой набор инструкций, вводимых в специализированное программное обеспечение для очистки веб-страниц. Затем компьютерная программа отправляет на веб-сайты запросы на определенные страницы. На этом этапе могут иметь место различные типы аутентификации, и веб-сайт может отвечать или не отвечать запрошенным контентом. Предполагая, что контент возвращен, программа затем анализирует ответ, чтобы извлечь данные в структурированный формат для дальнейшей обработки. Фактическое извлечение данных с веб-сайтов обычно является лишь первым шагом в проекте веб-скрейпинга. Обычно необходимо предпринять дополнительные шаги для очистки, преобразования и агрегирования данных, прежде чем они могут быть доставлены конечному пользователю или приложению. Например, имена людей могут быть разделены на имя и фамилию, аббревиатуры штатов могут быть сопоставлены с их полным именем, а цены на продукты могут быть преобразованы в числовые типы данных путем удаления знаков доллара и запятых. Наконец, данные могут быть обобщены на более высоком уровне детализации, например, для отображения средних цен по категории. Для выполнения этих задач пост-очистки можно использовать множество типов программного обеспечения и языков программирования. Общие варианты включают Excel, Python и R. Excel предлагает самую низкую кривую обучения, но ограничен в своих возможностях и масштабируемости. R и Python - это языки программирования с открытым исходным кодом, которые требуют навыков программирования, но практически безграничны в возможностях манипулирования данными. Для крупномасштабных приложений могут потребоваться более продвинутые методы, использующие такие языки, как Scala, и более продвинутые аппаратные архитектуры.

Этапы работы при подготовке к парсингу следующие: настроить среду Python на своем компьютере. Наладить IDE. Установить библиотеки Beautiful Soup и requests, которые необходимы для сбора данных с веб-страниц.

BeautifulSoup - одна из самых плодовитых существующих библиотек Python, в какой-то мере сформировавшая сеть в том виде, в каком мы ее знаем. BeautifulSoup - это легкий, простой в освоении и высокоэффективный способ программной изоляции информации на одной веб-странице за раз. Обычно BeautifulSoupin используется в сочетании с библиотекой запросов, где запросы извлекают страницу, а BeautifulSoup извлекает полученные данные.

Фреймворк Scrapy: программный пакет созданный специально для максимальной эффективности сбора данных. Является инструментосм для создания поисковых роботов и технологии, использующие scrapy это как правило роботы, выполняющие задачи беспрепятственного перехода по ссылкам и максимальный захват данных там, где данные существуют. Поскольку Scrapy служит для массового собирания данных с сайтов, он содержит в своём составе инструментарий для логирования, хранения пакетов куки и поддерживает работу с сервисами VPN.

Selenium - это не только инструмент для парсинга, но и инструмент автоматизации, который можно использовать для парсинга сайтов. Selenium - это вариант для программной навигации по сайтам в случаях, когда сайт использует разную методику борьбы с автоматическими программами и роботами, его следует рассматривать как инструмент в случае острой необходимости, поскольку скорость работы указанного намного медленнее вышеперечисленных.

regular expression – Инструмент используемый не для сбора данных, а для их очистки и выделения требуемых. формальный язык поиска и осуществления манипуляций с подстроками в тексте, основанный на использовании метасимволов. Для поиска используется строка-образец (англ. pattern, по-русски её часто называют «шаблоном», «маской»), состоящая из символов и метасимволов и задающая правило поиска. Для манипуляций с текстом дополнительно задаётся строка замены, которая также может содержать в себе специальные символы.

# 1.3.1 Поисковые системы

Многая информация может быть получена с использованием разного рода поисковыми системами. Инструменты вроде Google и Yandex являются частными и не единственными случаями, поскольку они могут быть использованы во многом только ручным методом, что не соответствует нашим задачам массового сбора данных. Существуют альтернативные методы сбора данных, некоторые из которых не доступны обычным пользователям, однако позволяют добиваться результатов весьма интересных благодаря своим инструментам .

OSINT (англ. Open source intelligence) или разведка на основе открытых источников включает в себя поиск, сбор и анализ информации, полученной из общедоступных источников. Главной целью ставится поиск информации, которая представляет ценность для конкурента либо злоумышленника. К такому типу инструментов можно отнести например .

TheHarvester - способный осуществлять сбор определенного типа данных использую открытые источники данных и доступные поисковые системы.

SpiderFoot – инструмент, позволяющий собирать с сайтов данные о имена пользователей, телефонные адреса или технические данные. Например домены и субдомены.

Указанные здесь методы не обязательно должны использоваться даже в случае поставленной научной задачи, поскольку работа подобных систем хотя и не вредит напрямую создателям сайта или владельцам контента, работают в масштабах не предполагавшихся создателями и владельцами информационных источников.

# 1.3.2 Предварительная обработка данных

При сборе информации возможна ситуация, когда информация собрана, однако формат полученных данных не соответствует желаемому и данные в таком виде не возможно будет обрабатывать методами математики. Очевидный пример это запись числовых данных в формате строки. При передаче такого типа данных в функцию расчетов мы не получим результатов, а надежность программы соответственно будет низкой. Наша задача обеспечить надежную работу системы в первую очередь.

Самый простой метод решения этой задачи это использование встроенных в язык методов преобразования, как int() или float() и перехват ошибки в случае невозможности произвести такой. Вполне возможно, что вместо необходимых цифровых данных нам удалось получить кусочек строки или вместо точки в цифре координат будет стоять запятая, что соответственно требует отдельных методов обработки. Во многом собранные данные можно редактировать, используя технологии bigdata, поскольку они изначально были сделаны для задач исправления, фильтрации и изменений больших объемов табличных данных. Нельзя исключать вариант хранения данных в формате строки как есть, поскольку требования к ПО меняются с течением времени и может оказаться, что отфильтрованная часть данных будет нужна для одной из новых моделей. Вероятность такого события тем не менее мала.

# 1.3.3 Специализированные поисковые системы

При сборе данных можно пользоваться многими методами. Каждый имеет свои преимущества и недостатки. Классический метод сбора информации, самый доступный, но не теряющий своей надежности – анкетирование и опросы. Владелец бизнеса может проводить сбор предложений и рекомендаций от клиентов, либо случайных прохожих и проводить различные методы анализа на этих данных. Эта технология не новая и даже закреплена законом путем обязательного наличия доступной книги жалоб и предложений.

Вторым по популярности можно считать сбор количественной информации – количество чеков, ассортимент, наиболее популярные пункты меню и цена среднего чека. Полученная информация тоже имеет ценность и может помочь в изучении целевой аудитории, свойства которой в свою очередь будут очень зависеть от местоположения организованного бизнеса.

Инвестору ещё до реализации своей задумки желательно максимально тщательно изучить среду в которой он планирует открыть свою точку предоставления услуг, поскольку целевая аудитория является основным потребителем любого заведения и будет весьма неприятным сюрпризом узнать, что область в которой было открыто заведение на самом деле имеет малую долю целевой аудитории в своём составе.

Для предотвращения подобной ситуации следует проводить аналитику более глобальными инструментами, как например сервисы, предоставляющие картографические данные. Известный 2GIS активно развивает указанный подход и занимает позицию среди лидеров анализа географических данных.

# 1.4 Исследование технологий и аналитических ГИС

При разработке любого программного обеспечения, крайне важно ориентироваться на предыдущий опыт разработки альтернативных систем. Одной из подобных программ можно считать разработанную Г.М. Ружниковым, А.Е. Хмельновым, А.С. Гаченко проекта “ ГИС «ИНВЕСТОРА» г. ИРКУТСКА” [28]. В их проекте реализовано облегчение доступа и навигации по существующим адресным планам города с учетом обновляемых экономических данных.

“В качестве базовой топоосновы используется цифровой адресный план г. Иркутска, который регулярно автоматически обновляется по дежурному цифровому топографическому плану города. Информация для ГИС «Инвестора» автоматически формируется из тематических слоев муниципальной геоинформационной системы г. Иркутска (МГИС). Кроме того, у системы имеется собственная БД, которая предназначена для хранения служебной информации, а также быстрого добавления табличных данных, получаемых из внешних источников.”



Рисунок 1 - Внешний вид программы ГИС «Инвестора»

В статье “Возможность использования ГИС-технологий для динамического планирования инвестиционного процесса в городах” (possibilities of using GIS technology for dynamic planning of investment process in cities) изданная в журнале Procedia computer science за авторством Cesar Stepniak и Tomas Turek исследует вопрос возможности и оправданности использования ГИС технологий для планирования инвестиций.

Ещё одной системой является “Пространственный анализ для выявления городских районов с более высоким потенциалом для социальных инвестиций” (Spatial Analysis to Identify Urban Areas with Higher Potential for SocialInvestment) за авторством Marcos Roberto Martines, Ricardo Vicente, Ferreira Rogйrio, Hartung Toppa, Alexandre Cavagis. [8] В разработанной системе осуществляется анализ территории для выявления наиболее недофинансированных территорий с точки зрения качества жизни проживающих людей с целью устранения неравенства повышения уровня общего благосостояния.

Этот метод позволяет анализировать данные в различных пространственных масштабах из базы ГИС, позволяя систематизировать области с более высоким потенциалом для инвестиций в социальное оборудование, услуги и объекты. Кроме того, он дает возможность моделирования нескольких переменных, что позволяет адаптировать его к различным географическим областям.

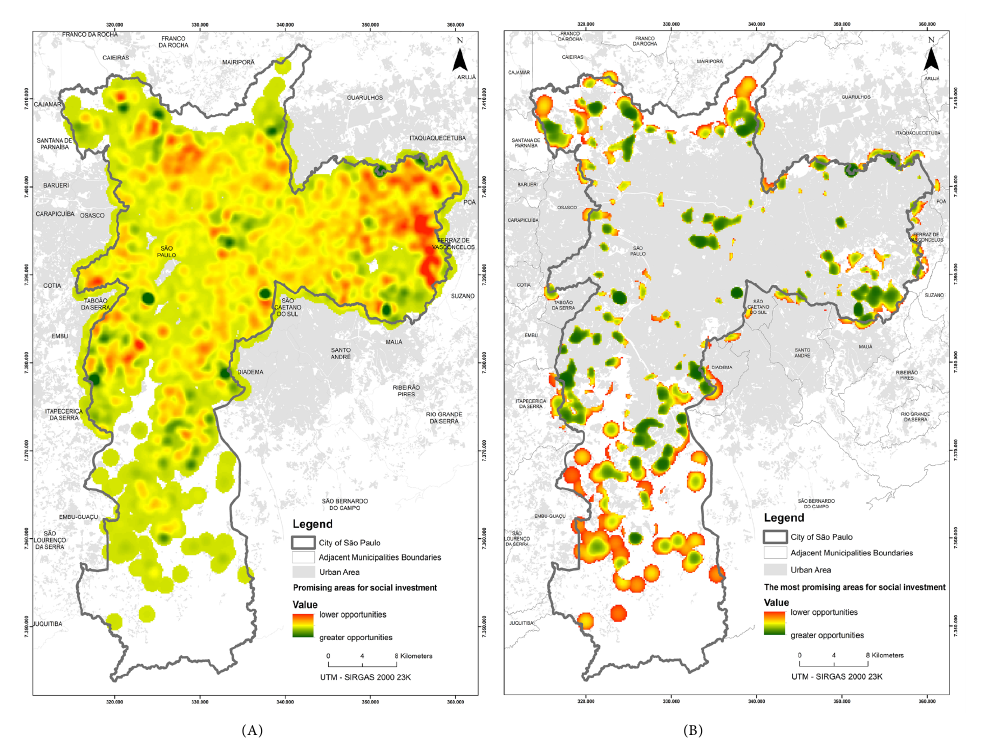


Рисунок 2 – Тепловая карта представляющая распределение областей, обладающих потенциалом для социальных инвестиций

В статье “Разработка системы поддержки принятия решений на основе данных” за авторством А. С. Быканова, В. В. Соболев [19] Рассуждается на тему разработки Системы поддержки принятия решения и авторы приходят к выводу, что подобная система хотя и возможна, может быть весьма трудоёмкой в разработке и совсем не универсальной .

“Системы поддержки принятия решения на основе данных достаточно удобны для сбора, хранения и обработки больших массивов данных за счет их упорядоченности и использования различных сложных запросов, позволяющих оперировать с ними. Но далеко не всегда они просты в реализации и их применимость отличается в зависимости от выбранной предметной области: необходимо предусмотреть вид данных, необходимость их обновления и наилучшую модель их хранения.” [19]

В статье “Оценка недвижимости г. Браслава с использованием гис-технологий” за авторством Е.Т.Борковский и Л.В.Гурьянова [26] исследуются моделирующие возможности ГИС для моделирования и оценки недвижимости в г. Браслава. Авторы проводят анализ данных и формируют статистику.

В статье “ Многокритериальный анализ принятия решений на основе ГИС” (GIS-Based Multiple-Criteria Decision Analysis) за авторством Рэндал Грин, Родольф Девиллерс, Джоан Э. Лютер и Брайан Г. Эдди (Randal Greene, Rodolphe Devillers, Joan E. Luther and Brian G. Eddy) [4]

В статье исследуются возможности проведения анализа решений по множеству критериев (MCDA). “поддерживая конечную цель - сделать его более доступным для лиц, принимающих решения. Описывается ряд факторов для описания проблем MCDA и выбора методов, которые затем упрощаются в виде дерева решений, которое организует введение основных методов. Подходы варьируются от математического программирования и эвристических алгоритмов для одновременной оптимизации нескольких целей до более распространенных одноцелевых методов, основанных на взвешенном сложении значений критериев, достижении пороговых значений критериев или превосходстве альтернатив.”

В статье “ Агентно-ориентированная модель динамики города с использованием гис-технологий ”за авторством С. П. Семенов, В. В. Славский, А. О. Ташкин. [17]

В работе проводится моделирование динамики города с использованием комбинированных технологий: агентного подхода, методов системной динамики и ГИС- технологий. Разработанная Компьютерная модель имеет возможность быть примененной в качестве инструмента поддержки принятия решения при управлении системами города. Представлена концептуальная модель динамики города и технология ее реализация в среде ПО AnyLogic, а также проведено исследование влияния изменения уровня комфорта районов на динамику населения города.

# 1.5 Модели принятия решений

При разработке системы поддержки принятия решения, помимо задачи сбора данных требуется разработка модели анализа данных.

Модели для системы поддержки принятия решения могут быть разработанны абсолютно разными подходами, однако первое, что стоит отметить, так это то, что лучшие модели это модели, разработанные под конкретную ситуацию с специфическими для неё данными и проблемами. Остальные подходы и варианты решения, созданные как универсальные методы не смогут конкурировать со специализированными моделями.

Какие могут быть созданы модели для поддержки принятия решений? В следствие анализа уже существующих решений в этой сфере, было установлено, что на данный момент не существует определенной тенденции или моды по разработке моделей систем принятия решений определенным образом. Каждый отдельный случай и задача рассматривается в индивидуальном порядке и инструменты используются абсолютно разные.

В первую очередь в нашей ситуации характерно использование моделей искусственного обучения: Нейросети, деревья принятия решений, логистические регрессии и многие другие технологии, используемые в datascience очень хорошо себя показывают в нашей сфере. У них так-же есть ещё одно неоспоримое преимущество – возможность самообновления.

Маленькие изменения, характерные для нашей реальности не так сильно влияют на модели, созданные с технологией обучения, поскольку такие модели могут автоматически производить до обучение на новых объемах данных и менять параметры весов автоматически и регулярно даже без вмешательства разработчика. Изменения самой модели в этом случае потребуются только при серьезных изменениях работы организации и почти при полном изменении обучающего параметра. Помимо указанного, данные системы не так сложно обслуживать, поскольку специалистов в области анализа данных на сегодняшний день больше, чем математиков.

Ещё один подход заключается в использовании математических моделей. Математические модели позволяют получить больший контроль надо происходящим, в отличие от систем машинного обучения. Результаты их работы могут быть легко интерпретированы и расчеты без особых сложностей выполняются на машинах с низкой вычислительной мощностью.

В целом у математических моделей множество преимуществ перед системами искусственного интеллекта, однако есть и недостатки:

1. Отсутствие механизмов самостоятельно проводить изменения
2. Дороговизна изменений
3. Затрудненная возможность в найме стороннего специалиста, поскольку математические модели могут быть достаточно сложны, чтобы их корректировка была проведена правильно в сжатые сроки

Ещё одним подходом можно считать вариант предоставления достаточного объема данных для принятия решений, однако без создания непосредственно моделей. В этом случае пользователю как правило предоставляют все возможные инструменты для проведения сравнений, накладывании и манипулировании имеющимися данными, однако решение он формирует единолично.

Такой подход можно считать оправданным в некоторых случаях, однако его использование сопряжено с тем, что обязательно требуется участие человека при принятии решения, что влечет за собой человеческий фактор.

# 1.6 Постановка задачи на разработку специализированной системы поддержки принятия решений на основе открытых ГИС

При описанных условиях необходимо разработать систему, осуществляющую сбор, систематизацию картографических и экономических данных с последующим проведением анализа всей доступной информации.

Разрабатываемая система должна иметь достаточный объем функционала для возможности пользователем вводить и получать данные после обработки. Важным является и фактор точности её работы, поскольку именно он обеспечивает ценность разработанной модели.

Нами было принято решение разработки программного обеспечения, содержащего в себе систему сбора информации , модели расчетов и модуль для взаимодействия с пользователем.

Метод расчетов выбран комбинированный математический. Основная часть расчетов должна проводиться методом индексного попарного сравнения альтернатив (РИПСА), однако до обработки данных таковым предполагается предобработка данных построением тепловых карт и дополнением данных экономической информацией, получаемой из открытых источников.

# 2 Разработка модели принятия решения

# 2.1 Модели представления информации

При проектировании модели, обрабатывающей данные нам в первую очередь, нужно понимать, что мы работаем с данными, чем с моделями. Именно данные диктуют условия и возможности разрабатываемых моделей. Мы можем с уверенностью заявить, что данные являются ключевым в нашей и подобных работах .

Данные могут быть прежде всего разделены по параметру четкости данных. К четким относят те данные, которые имеют натуральные значения, как количество домов или остановок, поскольку их можно измерить в штуках.

Части данных, которые мы используем для построения модели весьма желательно быть четкими, поскольку это позволяет иметь больший контроль над моделью и предсказуемость. Такие данные как правило не характеризуют реальный показатель, однако их использование не требует дополнительной обработки и проведения вычислительных операций.

К нечетким данным относят данные, которые невозможно напрямую интерпретировать. Таковыми можно считать данные, которые были получены благодаря построению моделей или полученные нормализацией изначально четких данных до числовых показателей в промежутке от 0 до 1. Оценка пассажиропотока не может иметь точных показателей, поскольку даже при учете количества проданных билетов мы не можем быть уверены в том, что не был пропущен ни один пассажир. Результаты полученные после построения модели оценки пассажиропотока косвенно по оценке плотности остановок могут быть только косвенными.

Данные можно оценивать по тому, что они характеризуют – это может быть распределение географической величины по территории – например влажность может быть распределена достаточно линейно, если построить линию перпендикулярно водоему. Возможно, данные будут характеризовать принадлежность объекта какому-либо явлению. Эти данные можно назвать категориальными. Таковым можно считать разбиение имеющейся непрерывной величины на определенные категории. В нашей работе данные о компаниях имеют показатель категории принадлежности их к одному из типов предприятия, как продуктовые магазины розничной торговли или реализации автомобильной продукции.

В целом любые данные могут быть полезны, поскольку в худшем случае они будут оказывать низкое влияние на принимаемое моделью решение. Чем больше данных, тем больше возможность сделать качественный результат.

Вполне возможно, что для построения модели придется проводить фильтрацию, формировать дополнительные синтетические показатели с использованием сочетания нескольких параметров и даже менять модель. В отдельных случаях данные могут быть полезны только для корректировки уже имеющихся результатов.

Одна из наиболее полезных концепций пространственного анализа - это плотность. Плотность людей в многолюдном городе, или плотность следов на участке пустыни, или плотность розничных магазинов в торговом центре. Плотность обеспечивает эффективную связь между концептуализацией дискретного объекта и непрерывного поля, поскольку плотность выражает количество дискретных объектов на единицу площади, и сама является непрерывным полем. Математически плотность какого-либо объекта рассчитывается путем подсчета количества таких объектов в области и деления на размер области. Это легко сделать при вычислении плотности населения по данным переписи, например, путем деления населения каждого переписного участка или округа на его физическую площадь. Но это приводит ко множеству значений плотности в зависимости от выбранных объектов и определения областей, используемых для расчета.

Наборы пространственных данных позволяют создавать операционные модели реального мира на основе концепций полей и объектов:

- Поля, и использования координатной геометрии для представления классов объектов.

- Объекты. К ним относятся: дискретные наборы точечных местоположений; упорядоченные наборы точек (открытые множества, образующие составные линии или полилинии, и замкнутые непересекающиеся множества, образующие простые многоугольники); и множество изображений непрерывно меняющихся явлений, иногда описываемых как поверхности или поля. Последние часто представляют собой непрерывную сетку квадратных ячеек, каждая из которых содержит значение, указывающее (оценочную) среднюю высоту или напряженность поля в этой ячейке. В большей части литературы и в программных пакетах модель точек / линий / областей описывается как векторные данные, тогда как сеточная модель описывается как растровые (или изображения) данные.

Векторные данные могут быть преобразованы в растровое представление и наоборот. Преобразование в большинстве случаев приведет к потере информации (например, разрешения, топологической структуры), и, следовательно, такие преобразования могут быть необратимыми. Например, предположим, что у нас есть растровая карта, содержащая ряд отдельных зон (групп смежных ячеек), представляющих тип почвы. Чтобы преобразовать эту карту в векторную форму, вам нужно будет указать целевую векторную форму (многоугольник в этом примере), а затем применить операцию преобразования, которая найдет границы зон и заменит их сложным набором с зубцами многоугольников, следующих по контуру формы сетки. Эти многоугольники затем могут быть автоматически или выборочно сглажены, чтобы обеспечить упрощенное и более приемлемое векторное представление данных. Если обратить этот процесс вспять, взяв сглаженную векторную карту и сгенерировав выходной растровый файл, обычно получается несколько иной выходной файл, чем тот, с которого мы начали, по разным причинам, включая:

- Степень обнаружения границ и упрощения, предпринятого во время векторизации;

- Точный характер алгоритмов обнаружения и преобразования границ, применяемых как при векторизации, так и при растрировании;

- Способ обработки особых случаев, например края карты, «открытые зоны», изолированные ячейки или ячейки с пропущенными значениями;

# 2.1.1 Численные данные из открытых систем

Некоторые открытые системы абсолютно бесплатно могут предоставить картографические данные. Таковыми можно считать Openstrrtnap и Wikimapia.

Эти данные как правило относятся к категории четких. Таковыми могут быть, например остановки или дома. Впоследствии для задач расчетов данные могут быть преобразованы в достаточно полезные характеристики – количество домов в изучаемой зоне, количество зданий определенной категории, количество остановок в обрабатываемой области и.т.д.

Четкие данные хороши надежностью показателей, мы можем быть уверены в этих данных, и они являются человекочитаемой характеристикой, удобной для представления и обработке.

Благодаря ним могут быть построены многие модели, как модель характеризующая плотность населения, проживающего на некоторой области или интенсивность движения, если мы будет строить модель базируясь на данных о остановках.

Допускается возможность получения данных не только из систем, предоставляющих картографические данные. Вполне возможно использование альтернативных методов сбора информации, однако в этом случае сразу стоит понимать, что в итоге их нужно будет привязать к местоположению.

# 2.1.2 Статистические данные

Важно использование наиболее качественных и достоверных данных. Хорошая математическая модель может быть построена только благодаря качественным данным.

Статистические данные являются примером таковых. При возможности использовать данные из официальной статистики лучше воспользоваться ей и добавить таковые в свою модель, однако стоит подумать о том, какие это будут данные и могут ли они быть полезны в нашем конкретном случае.

Статистические данные, собираемы государственными органами или организациями, имеющими высокий уровень качества данных как правило не доступны в сыром виде или может быть получен за достаточно крупную плату. В наше время получить данные во многом можно, однако процесс этот достаточно дорогой с финансовой точки зрения.

Вторым важным пунктом следует отметить, что данные статистики могут быть недостаточно точными. Во многом благодаря смещению их результатов в сторону, желаемых заказчиком. Подобная проблема может быть решена, однако это стоит учитывать.

Третьим недостатком статистических данных можно считать их обобщение. Для государства апеллирование масштабами сотен метров не оправдано и как правило данные таких масштабов имеют тенденцию к группированию многих случаев и событий в единые блоки, внутри которых возможно отсутствие уточнений.

Использование статистических данных может быть оправдано для увеличения точности, однако не следует рассматривать этот источник информации как основной на этапах разработки системы.

Следует отличать статистические данные от пространственной статистики:

Термин пространственная статистика относится к применению статистических концепций и методов к данным, имеющим явную пространственную структуру, которая важна для понимания этих данных. Если данные доступны со связанной пространственной информацией, и последняя ничего не добавляет к пониманию данных, тогда анализ может продолжаться, как если бы данные не имели пространственного контекста, то есть как если бы данные были пространственными.

Пространственная статистика - это самостоятельная дисциплина, тесно связанная с традиционными статистическими методами и более современной вычислительной статистикой, но с дополнительной сложностью пространственной зависимости - обычно выборки пространственных данных не являются независимыми, как это обычно требуется в традиционных статистических методах.

# 2.1.3 Нечеткие множества и нечеткие оценки

Использование нечеткой логики может быть весьма хорошим решением в условиях нашей задачи. Поскольку мы вынуждены определять параметры комплексным и иногда опосредованным образом связанные с результатом, нам требуется использовать новаторские методы расчетов. Одним из таковых можно считать расчеты с использованием нечеткой логики.

Нечеткая логика подразумевает в первую очередь работу с относительно сырыми данными: любые данные, которые получает модель для расчетов должны быть преобразованы в формат от 1 до 0, поскольку математические (логические) операции могут быть произведены только с ними. В теории датасаенс такие действия принято называть нормализацией и в нечеткой логике важным моментом является использование степени истинности – вещественные числа с значением от 0 до 1, что и формирует понятие нечёткости.

После этой операции уже можно строить систему обработки данных с использованием стандартных для классической логики операций: конъюнкции (&&), дизъюнкции (||) и отрицания (!). Построенная таким образом модель в отличие от систем использующих нейросети будет работать предсказуемым образом, однако получит подобные преимущества и будет способна выдавать требуемый результат.

В нашем случае стоит задача формирования логической структуры вокруг каждого отдельного заведения или территории, которую ввел пользователь и выдачи результатов средних между предсказанием сравнительной эффективности этих областей. Что обобщая ведет к пониманию разности между положительными свойствами и свойствами отрицательными – разными в каждом конкретном случае. Под свойствами следует понимать совокупность параметров, формируемых за счет количественной оценки ценностных свойств а также оценки значимости влияния каждого свойства и его интенсивности.

Модель расчетов в этом случае:

for i in 1..n

for j in 1..n

E[i,j] = A[i] == B[j]

end

end

A и B - это сравниваемы области. A[i] и B[j] - это их свойства (характеристики), оператор '= =' вычисляет их нечеткое равенство.

Важно отметить, что обе области имеют общее количество свойств - n. А в случае количественного различия этих свойств имеет необходимость выделения количественной различности в отдельный параметр для сравнения и приведение этих свойств к количественной идентичности. Однако это особый случай, который встречать пока не удавалось.

После этого нам следует свести данные в матрицу:

for i in 1..n

M[i] = E.max\_in\_row(i)

end

Метод max\_in\_row вычисляет максимальное значение в строке матрицы. После этого превращаем вектор совпадений в одно единственное значение вычисляя отношение суммы элементов к длине вектора:

e = M.sum / M.length

Итоговое значение отражает требуемый запрос.

# 2.1.4 Нормализация и стандартизация

- Средние. Значение атрибута в каждой зоне (например, количество детей) корректируется путем деления, например, на количество домашних хозяйств в каждой зоне, чтобы получить среднее количество детей на домашнее хозяйство, которое затем может быть сохраненным в новом поле и / или отображенным.

- Проценты. В случае процентов или преобразования в диапазон [0,1] выбранное значение атрибута делится на максимальное значение или (в случае данных подсчета) на общее количество в наборе, из которого оно было взято. Например, количество людей в зоне, зарегистрированной как безработные, деленное на общую численность населения трудоспособного возраста.

- Плотности. Делитель — это площадь зоны (которая может храниться в некоторых программных пакетах как внутренний атрибут, а не как поле), в которой находится указанный атрибут. Результатом является набор значений, представляющих население на единицу площади, например, жилищ на квадратный километр, зрелые деревья на гектар, тонны пшеницы на акр и т. д. - все это можно охарактеризовать как меры плотности.

Термин стандартизация тесно связан с термином нормализация и также широко используется в анализе. Он имеет тенденцию появляться в двух разных контекстах. Первый - как средство обеспечения сопоставимости данных из разных источников по рассматриваемой проблеме. Например, предположим, что изучается показатель успешности лечения по районам здравоохранения. Для каждого медицинского округа процент успеха каждого типа операции можно было проанализировать, нанести на карту и составить отчет. Однако различия в демографическом составе медицинских округов не учитывались в уравнении. Если в некоторых районах очень высока доля пожилых людей, показатели успеха для некоторых видов лечения могут быть ниже, чем в других районах. Стандартизация - это процесс, при котором ставки корректируются с учетом таких вариаций в округах, например, по возрасту, полу, социальной депривации и т.д. Пропорции и корректировка заявленных коэффициентов с использованием этих значений для создания коэффициента, взвешенного по возрасту. Косвенная стандартизация включает вычисление ожидаемых значений для каждого района на основе региональных или национальных показателей с разбивкой по возрасту / полу / депривации и сравнение этих ожидаемых значений с фактическими показателями. Последний подход менее чувствителен к небольшим размерам выборки в сочетании стандартизирующих переменных (то есть, когда все они объединены и объединены с конкретными методами лечения в определенных районах).

Второе использование термина «стандартизация» включает корректировку величины конкретного набора данных для облегчения сравнения или комбинации с другими наборами данных. Обычно применяются две основные формы такой стандартизации.

Первый — это стандартизация z-показателя. Это включает в себя вычисление среднего и стандартного отклонения для каждого набора данных и изменение всех значений на z-оценку, в которой среднее значение вычитается из значения, а результат делится на стандартное отклонение (которое должно быть отличным от нуля). Это генерирует значения, которые имеют нулевое среднее значение и стандартное отклонение 1. Эта процедура широко используется при статистическом анализе наборов данных.

Второй подход предполагает стандартизацию на основе диапазонов. В этом методе вычисляются минимальное и максимальное значения данных (таким образом, диапазон = max ‑ min), и из каждого значения вычитается минимальное значение, которое затем делится на диапазон. Чтобы уменьшить влияние крайних значений, стандартизация диапазона может опускать количество или процентное соотношение наименьших и наибольших значений (например, наименьшее и наибольшее 10%), давая форму усеченного диапазона. Обрезанный стандарт на основе диапазона

# 2.2 Математические модели для вычисления критериев оценки привлекательности территории на основе открытых данных ГИС

Задача принятия решения не стояла бы перед нами, если она была на столько простой, как оценка единственного критерия. Однако, реальность заставляет нас сталкиваться с многокритериальными системами принятия решения и соответственно появляется необходимость создания комплексных методов оценки и ранжирования.

# 2.2.1 Экономическая активность территории

Экономическая активность территории является одним из важнейших показателей, которые мы можем передать в модель для последующей оценки и принятия решения. Сразу стоит сказать, что экономическая активность — это показатель нечеткой логики и не может быть охарактеризована натуральным параметром. В нашем случае экономическая активность это параметр, характеризующийся сразу несколькими параметрами, каждый из которых в той или иной степени свидетельствует о наличии экономической активности. К таковым параметрам можно отнести:

1. Построенная модель тепловой карты, в которой показательным параметров являются обороты или прибыль компаний
2. Параметр насыщенности территории компаниями любого рода.
3. Параметр плотности офисов или офисных зданий.
4. Наличие и концентрация остановок.

Каждый из этих параметров вносит вклад в общую оценку экономической интенсивности в зоне. Следует отметить, что экономически успешный район не гарантирует непосредственно высокую вероятность успеха, а характеризуют зону, как более благоприятную, нежели соседнюю.

Расчет экономической интенсивности производился оценкой всех параметров, а также значимости каждого из них. В случае создания синтетического параметра из нескольких других важно учесть сложность их взаимодействия. Как правило такая операция проводится простым делением, однако в нашем случае учитываются четыре параметра и в следствие этого возможны специфические проблемы. Одной из таковых можно считать преобладание значимости маленьких компаний за счет числа таковых над значимостью одной, однако крупной компании с большим объемом оборотов. Этот недостаток решается тем, что мы вводим дополнительный параметр прибыли, учет которого способен исправить эту погрешность.

В этом случае у нас появляется новая проблема – оборот крупной компании распространяется на все её филиалы, а не зациклен только на одном конкретном. В этой ситуации наиболее простым решением будет установление значения количества филиалов у компании и разделения прибыли или оборота на количество филиалов. Данное решение хотя не решает задачу полностью, способно сгладить проблему достаточным для дальнейших расчетов образом.

Приведем формулу для расчета указанного синтетического параметра для определенной выбранной зоны:

|  |  |  |
| --- | --- | --- |
|  |  | (1). |

Где M – среднемесячная выручка предприятия, количество филиалов компании, O – офис, S – остановка

Приведенная формула в целом описывает ценность выбранной зоны, однако требуется для верности провести нормализацию этого параметра между всему анализируемыми случаями, для этого рассчитывается стандартным образом нормализация по формуле:

|  |  |  |
| --- | --- | --- |
|  |  | (2). |

Где E – значение, подлежащее нормализации. , – интервал значений E.

# 2.2.2 Населенность территории

Совершенно очевидно, что ни один разумный инвестор не будет открывать точку рознично торговли посреди пустырей, однако, когда речь идет о сравнении густоты заселённости территории, сложность той-же задачи становится намного выше.

Поскольку мы разрабатываем географическую систему поддержки принятия решений, нам, конечно, следует использовать максимально возможное количество параметров и населенность можно считать одним из таковых. Самый простой статистический учет позволит это сделать, однако получение государственных данных — это задача сложная и не надежная. Мы не можем быть уверенны, что данные репрезентативны, актуальны, характеризуют области с достаточной точностью и границы этих областей достаточного размера. Наша модель требует минимально характеристик заселенности в несколько квадратных километров, а государственные данные скорее обобщают районы. Подытоживая, можно заключить, что данные Росстата нас скорее не устроят, чем устроят.

Альтернативным вариантом можно считать возможность расчета предположительной заселенности путем оценки этажности домов и среднего количества комнат. В этом случае нам потребуется получить более подробные данные о высоте домов и его свойствах: жилая ли застройка или офисная. После этого мы сможем несложно рассчитать условное количество проживающих.

Есть альтернативный вариант в проведении оценки путем получения данных прописках граждан, однако эта опция доступна только при использовании разработанной системы государством, поскольку подразумевает использование конфиденциальных данных.

Использованная формула расчетов населенности:

|  |  |  |
| --- | --- | --- |
|  |  | (3). |

Где P – населенность, – площадь анализируемой области (м2), N – количество зданий жилого типа

# 2.2.3 Критерий оценки посещаемости территории

Оценка посещаемости на таких больших территориях с достаточной точностью на первый взгляд достаточно сложно производится. Нам потребуется ходить со специальными аппаратами, расставлять их и замерять количество людей в потоке, что по времени может занять столько-же ресурсов, сколько построение модели.

На самом деле вовсе не обязательно собирать требуемые параметры напрямую. Мы можем, как и в предыдущем случае, воспользоваться распространённой статистикой и на её основании построить некоторую несложную модель. В случае посещаемости речь идет о динамике движения потенциальных клиентах любого качества. Когда речь идет о движении сразу на ум приходит транспорт и остановки, обеспечивающие жителей города возможностями перемещения или же просто мобильностью.

Показатель количества транспортных узлов любого рода: остановки, количество маршрутов транспорта или даже ширина дорог может характеризовать проходимость в общих чертах. Хотя, когда речь идет о точке розничной торговли важным оказывается движение людей, а не людей посредством транспорта – проезжающий поток машин в большом объеме вовсе не характеризует область как привлекательную напрямую, ведь транспорт без необходимости и тем более возможности остановиться не добавляет количества потенциальных покупателей в реальности. Следовательно, более важными оказываются остановки.

Остановки тоже можно характеризовать по условной пропускной способности, но такая точность излишняя для нас, поэтому можно обойтись количеством остановок в некотором заданном радиусе.

Есть ещё одна важная деталь, не учитываемая таким методом расчетов. Речь идет о том факте, что остановки могут располагаться в местах, населенных людьми, но область населения будет находиться только с одной стороны от остановки. Наша модель не может учитывать такой параметр как нахождение точки в промежутке между областями посещения людей, кроме как косвенно благодаря оценкам самих этих областей или точек притяжений.

Использованная формула расчетов посещаемости:

|  |  |  |
| --- | --- | --- |
|  |  | (4). |

Где A – посещаемость. – площадь анализируемой области (м2), T – количество маршрутов транспорта, G – количество остановок

# 2.2.4 Насыщенность территории объектами культуры

Оценка насыщения некоторой территории предметами, обеспечивающими времяпрепровождение людей на открытом воздухе тоже важный параметр. К таковым можно отнести Парки, кинотеатры, музеи и базовые предметы обустройства территории как наличие лавочек, качелей, спортивных снарядов и фонтанов.

# 2.2.5 Учет фактора конкуренции

Фактор конкуренции оказывает сильное влияние на оценку пригодности территории к организации эффективной торговли. Конкуренты могут сильно менять экономический ландшафт и решение при выборе определенной зоны.

Поскольку конкуренты как правило серьезные оппоненты при работе бизнеса, следует их оценивать как вредный фактор в первую очередь. Наличие конкурентов оценивается путем учета параметра типа организации и сравнения его с предполагаемым типом организации введенной пользователем.

Указанный метод на первый взгляд достаточно прост, и конкуренты вредят бизнесу, однако внимательный аналитик заметит существование таких мест, как “фудкорты”, где концентрируются в большом количестве компании предоставляющие услуги одного типа: продажи пищи в частности. Такое обстоятельство не может быть случайным и ранее уже описывалась её природа – следствие закона равновесия Нэша.

Для учета подобной ситуации модель можно улучшить, если область поисков будет достаточно большой, чтобы можно быть заявить о том, что поставленная точка находится в некотором локальном пике проходимости. В узкой зоне, где людей много и их практически нет вокруг неё. Задача сложная, но её можно решить при необходимости комбинацией моделирования тепловых карт и разработкой нейросети, занимающейся определением всего двух параметров - нахождения точки внутри области активности и изолированность этой зоны от других.

# 2.3 Математические модели для многокритериального принятия решения

С целями аналитики реальных объектов, ученым требуется создавать математические модели, поскольку они намного лучше функционируют в задачах прогнозирования и планирования. Модели как правило выполняют функции реальных объектов, но в отличие от таковых их создание проще, а эксплуатация может повторяться множество раз без серьезных затрат, что отличает их от создания прототипов. Благодаря моделям мы можем наблюдать предположительные изменения в объекте исследования при внесении изменений.

Важно заметить, что модели, используемые в естественных науках, отличны от моделей экономических, поскольку в последних моделируется общее поведение людей на базе имеющейся статистики и такие модели как правило актуальны непродолжительное время, поскольку люди это живые организмы.

# 2.3.1 Тепловые карты и когнитивный анализ карты экспертами

Пользователь оценивает сервис намного более высоко, если сервис может предоставить информацию несколькими способами. Одним из дополнительных вариантов можно считать предоставление информации пользователю методом формирования тепловых карт (heatmap).

Тепловая карта — это метод графического представления данных, который показывает величину явления в виде цвета в двух измерениях. Изменение цвета может быть связано с оттенком или интенсивностью, что дает читателю очевидные визуальные подсказки о том, как явление группируется или изменяется в пространстве. Есть две принципиально разные категории тепловых карт: тепловая карта кластера и пространственная тепловая карта. На тепловой карте кластера величины представлены в виде матрицы с фиксированным размером ячеек, строки и столбцы которой являются дискретными явлениями и категориями, а сортировка строк и столбцов является преднамеренной и в некоторой степени произвольной с целью предложить кластеры или изобразить их как обнаружено посредством статистического анализа. Размер ячейки произвольный, но достаточно большой, чтобы ее было хорошо видно. Напротив, положение магнитуды на пространственной тепловой карте обусловлено расположением магнитуды в этом пространстве, и здесь нет понятия ячеек. считается, что явление постоянно меняется.

В нашей задаче важно понимать, что тепловые карты могут отображать разные критерии. Например, существуют универсальные оценки, как плотность компаний или цена жилья, косвенно характеризующие уровень средств населения проживающего в указанной области., так и специфических: как плотность организаций определенной категорий, характеризующий уже меру конкуренции.

# 2.3.2 Кривые безразличия

Если мы придём в магазин с условными тысячей рублей и необходимостью закупить продуктов, то мы столкнемся с тем, что мы ограничены с набором продуктов, укладывающихся суммарной ценой в одну тысячу рублей. Тем не менее выбор у нас есть и предположим, что мы можем покупать в магазине только творог и яйца. Если мы потратим все деньги на покупку творога, то получим 8 кг. Если мы потратим все деньги на покупку яиц, то получим 4 кг яиц. Очевидно, что могут существовать и варианты комбинаций этих параметров: 6,5кг творога и 0,4 кг яиц или 3 кг творога и 1 кг яиц. Любой вариант комбинации нас в целом устраивает, поскольку мы получаем продуктов ровно на тысячу и именно такие комбинации параметров, в которых значения параметров различны, однако для нас сами комбинации равноценны являются безразличными параметрами. Если мы построим на двумерном графике все возможные комбинации, то они будут выстроены в так называемую кривую безразличия.

Рисунок 3 – Графическое представление кривой безразличия

Все точки на этой кривой для нас считаются удовлетворительными. Точки, находящиеся с лева или под кривой считаются неудовлетворительными и не устраивают нас, а область точек с права или с верху это отличный случай и нас такое устроит.

В случае построения линий кривых безразличия на географической плоскости, можно говорить о том, что линии будут превращены в двумерные плоскости - зоны, которые являются в конкретном анализируемом случае одинаковыми несмотря на различные параметры, каждый из которых в отдельности дает зонам отдельную характеристику.

Использование теории кривой безразличия позволяет разрабатывать систему, упрощающую понимание сложных данных в многомерном пространстве за счет проведения многих операций и выведения в качестве результата непосредственно требуемых данных

# 2.1.3 Механизмы сверток

Направлением решения задачи многокритериального анализа является отказ от множества критериев путем сведения их к одному. Простейший подход, когда один критерий считают главным и упорядочивают лишь по нему, а остальные используют, только если у двух альтернатив значения главного критерия одинаковы (если одинаковы значения и главного, и второго по важности критерия, используют третий и т.д.), оказывается удовлетворительным лишь в редких случаях. Обычно среди критериев невозможно выделить важнейший. Лучше работают методы, учитывающие все значения вектора критериев. Такие составные критерии принято именовать свертками.

Свертка - мощный оператор, который широко используется в глубоких нейронных сетях в самых разных вариантах. Это позволяет выражать в компактной форме операции над структурированным набором экземпляров данных аналогичной формы.

Произведение критериев является мультипликативной сверткой. В этом случае, подобно введению весов в аддитивной свертке, можно перед перемножением критериев возвести их в степень тем большую, чем больше важность, придаваемая критерию. Очевидно, что мультипликативная свертка оправданна, если критерии неотрицательны–иначе правило «минус на минус дает плюс» сыграет с нами плохую шутку, сделав «хорошее» значение свертки из двух заведомо плохих критериев. Впрочем, если только один из критериев принимает отрицательные значения, подобного рода парадоксы не возникают, и мы можем пользоваться мультипликативной сверткой. Также нужно учитывать, что если один из критериев равен нулю, то и мультипликативная свертка равна нулю, для аддитивной же свертки такое правило не выполняется. Вообще, в мультипликативной свертке по сравнению с аддитивной большее влияние оказывают те критерии, которые для данного объекта имеют низкие значения.

Аддитивная свертка наиболее приемлема для критериев, представляющих собой однородные по смыслу, и близкие по масштабу значений величины, каковыми в нашей классификации являются прогнозные критерии. Например, комбинируя «математическое ожидание прибыли по логнормальному распределению» и «математическое ожидание прибыли по эмпирическому распределению», естественно взять в качестве критерия их сумму. С другой стороны, для свертывания таких классов критериев, как «математическое ожидание прибыли» и «вероятность прибыли» (по любому из распределений), лучше применять мультипликативную свертку. В этом случае мы используем полезное свойство произведения – если прогнозируемая вероятность прибыли близка к нулю, то и сводный критерий также будет стремиться нулю. Впрочем, в применении произведения есть дополнительная тонкость – если матожидание прибыли отрицательно, то, умножая его на меньшую вероятность, получаем величину более близкую к нулю и, следовательно, большую. Однако это не создает трудностей, если комбинации с отрицательным матожиданием прибыли просто не принимаются к рассмотрению.

Кроме аддитивной и мультипликативной, существует также селективная свертка, когда для каждого элемента исходного множества принимается в качестве значения свертки наименьшее (или наибольшее) значение из всего набора критериев.

При расчете свертки не стоит забывать о том, что критерии могут измеряться в разных единицах и иметь различный масштаб величин. Существует несколько способов их приведения к единой мере. Так, можно вычесть из значений критериев их средние значения и разделить на стандартные отклонения (метод нормализации) или же вычесть минимальные (минимальные по данной выборке или минимальные принципиально достижимые) значения, разделив затем на разность между максимальным и минимальным значением (в этом случае значения критерия будут лежать в интервале от нуля до единицы). Первый из предложенных способов более пригоден для построения аддитивной, второй–для мультипликативной свертки. [3]

# 2.1.4 Методики попарного сравнения

Ключевой теорией для понимания использованной в диссертации методологии расчета является методология Парето-доминирования, которая несмотря на хорошее математическое обоснование имеет значимый недостаток – результат расчета в реальных условиях может быть трудно интерпретирован в случаях, когда метод предоставил большое количество множеств и когда мы получаем после расчетов пустое множество. Ненадежность такого подхода вынудила профессор Франции Бернарда Роя и его коллег в консалтинговой компании “SEMA” разработать альтернативную систему расчетов, которой впоследствии стала ELECTRE I.

Аббревиатура “ ELECTRE ” расшифровывается как “ELimination Et Choix Traduisant la REalité”, что в свою очередь переводится, как “Исключение и выбор в условиях реальности”, что достаточно неплохо описывает подход, который использовала новая методика расчетов. Важно заметить разработанное нововведение – использование дополнительного, регулируемого со стороны параметра рисков.

Метод предназначен для решения задач, в которых из имеющегося множества альтернатив требуется выбрать заданное количество лучших альтернатив с учетом их оценок по нескольким критериям, а также важности этих критериев.

Принцип работы метода следующий. Для каждой пары альтернатив (Aj и Ak) выдвигается предположение (гипотеза) о том, что альтернатива Aj лучше, чем Ak. Затем для каждой пары альтернатив находятся два индекса: индекс согласия (величина, подтверждающая предположение о превосходстве Aj над Ak) и индекс несогласия (величина, опровергающая это предположение). На основе анализа этих индексов выбирается одна или несколько лучших альтернатив ("ядро" альтернатив).

Расчет в нашей работе велся методом ELECTRE II. Однако прежде, чем мы перейдем к нему нам следует рассмотреть ELECTRE I

Метод ELECTRE I был разработан первым в семействе методов, принадлежащих к подходу разработки индексов попарного сравнения альтернатив (РИПСА). В нем используются четкие бинарные отношения между альтернативами.

Алгоритм действий для расчетов методом ELECTRE I строится следующим образом:

Выделяются альтернативные варианты и критерии их оценки , которые влияют на эффективность принимаемого решения.

Оба выдвинутых варианта сравниваются покритериально и в результате расчетов мы получаем по два индекса на каждую пару сравнений. Индексы распределяются в промежутке от 0 до 1 и характеризуют согласие и несогласие с гипотезами превосходства одной альтернативы над другой.

Далее для сравнения полученных промежуточных результатов нам нужно задать две пороговые величины – уровни согласия и не согласия. С этими уровнями сравниваются полученные на предыдущем этапе индексы согласия и не согласия. Утверждаемая гипотеза о том, что выбранная альтернатива лучше другой принимается за верную, если индекс согласия с этим утверждением будет выше порогового значения, а индекс несогласия ниже порогового значения несогласия. Все остальные варианты считаются несравнимыми. Уровни позволяют регулировать риски в задаче принятия решения. Чем выше уровень согласия и ниже уровень несогласия, тем менее рискованным является решение. При уровне согласия = 1, на уровне несогласия = 0 выделяется множество Парето.

Множество альтернатив фильтруется, оставляя наиболее приемлемые варианты.

Дальнейший выбор строится благодаря тому, что мы повышаем альтернативам сложность прохождения наши уровни путем изменения уровней согласия (уменьшая его) и несогласия (повышая его). Смысл заключается в уменьшении количества несравнимых альтернатив за счет повышения рисков принятия неправильного решения.

Указанный выше пункт выполняется многократно до тех пор, пока не останется единственная альтернатива, которую в свою очередь и можно считать верной. Последовательность ядер определяет упорядоченность альтернатив по качеству решения.

Указанная система действий верна для первой версии ELECTRE. [25] Для примера рассмотрим две альтернативы с гипотезой превосходства одной из них: АI над альтернативой АJ Множество I, состоящее из N критериев, разбивается на три подмножества:

I+ — подмножество критериев, по которым АI предпочти­тельнее Аj;

I= - подмножество критериев, по которым АI равноценно Аj;

I- — подмножество критериев, по которым АJ предпочтительнее АI;

Далее формулируется индекс согласия с гипотезой о превосходстве АI над Аj. Индекс согласия подсчитывается на осно­ве весов критериев. В методе ELECTRE I этот индекс определяется как отношение суммы весов критериев подмножеств I+ и I= к общей сумме весов:

индексы «согласия» для метода ELECTRE I по следующей формуле (3):

|  |  |  |
| --- | --- | --- |
|  |  | (5). |

Где АI, АJ альтернативные гипотезы.Wi – вес критерия

В случае расчетов методом ELECTRE II индекс согласия определяется как отношение суммы весов критериев подмножеств I+ к сумме весов критериев подмножеств I-.

Формула индекса «согласия» метода ELECTRE II (4):

|  |  |  |
| --- | --- | --- |
|  |  | (6). |

Где АI, АJ альтернативные гипотезы.Wi – вес критерия.

Расчёт индексов несогласия ведется согласно утверждению, что мы должны предоставить наиболее невыгодные условия для нашей гипотезы и поэтому критерием берется наиболее различающийся критерий по которой альтернатива АJ наиболее превосходит альтернативу АI . Расчет производится по формуле ELECTRE I, однако для учета возможной разницы длин шкал критериев, разность оценок относят к длине наибольшей шкалы.

Индексы «несогласия» рассчитываются по формуле (5):

|  |  |  |
| --- | --- | --- |
|  |  | (7). |

Где и – оценки альтернатив АJ, АI по критериям i.

- длина шкалы i-го критерия

# 3 РАЗРАБОТКА СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЯ НА ОСНОВЕ ОТКРЫТЫХ ГИС

В этой главе описываются преимущественно технологии, использованные при разработке системы и технические детали, обнаруженные во время работ.

# 3.1 Разработка подсистемы извлечения данных из открытых ГИС

2GIS имеет широкую базу данных, весьма ценную для проведения исследований подобных нашему. Однако API 2GIS предоставляет разные возможности демонстрации и настройки карт, но не предоставляет возможность получения данных с их серверов. Тем не менее данные можно собирать почти без препятствий и компания на сегодняшний день не блокирует доступ к ним годами даже на базовом уровне, что свидетельствует о крайне интересной политике их действий и о том, что сбор данных нам придется производить методом парсинга сайта сервиса.

Сбор данных производился с помощью программы, написанной на языке Python с библиотеками, позволяющими выполнять веб-запросы (requests) и проводить парсинг json строк, которые мы получали в ответ на запросы. Для эффективности мы также выполнили распараллеливание операций с использованием модуля multiprocessing.

Полученные данные заносились в таблицу excel и затем после обработки ручным методом и удостоверения в наглядном уровне качества полученных данных переводились в формат таблицы базы данных с той целью, что база данных работает заметно быстрее, чем таблицы формата xls.

# 3.2 Алгоритмы вычисления критериев

Критерии, используемые для проведения оценок разные и для каждого, должна использоваться, своя система расчетов. Базово алгоритм выполнения расчетов можно описать в виде порядка выполнения следующих операций:

Выдача параметра веса для каждого критерия согласно задаваемым параметрам значимости пользователем или автоматически, используя метод попарного сравнения .

Фильтрация путем удаления случаев, когда параметр имеет нерепрезентативную характеристику (“Null”, ”None”, ”Undefined”, 0.0).

Удаление случаев дублирования

Нормализация наших критериев до значений от 0 до 1 стандартным методом нормализации или выполняем хеширование значений, если его тип данных является категорией.

В отдельных случаях возможно наличие выбросов (значений, сильно отличающихся от средних) тогда допускается удаление таковых. Для этого мы будем пользоваться методом среднеквадратических отклонений. Среднеквадратическое отклонение вычисляется в значениях атрибута, что не допустимо поскольку в каждом атрибуте необходимое для нас число будет своим, поэтому этот этап производится обязательно после этапа нормализации.

Расчет показателей, используемых в качестве характеристики плотности организаций, проводился следующим образом:

1. Добавление к нашим данным нового параметра - количество соседей, который в начале был заполнен нулевыми значениям.
2. У каждой компании мы берем её координаты и из базы данных выбираем все компании, находящиеся не дальше, чем 200 метров в каждую сторону.
3. Считаем количество компаний в получившемся наборе данных
4. Присваиваем полученное значение к выбранной компании
5. Повторяем предыдущие операции до окончания списка компаний.
6. В результате мы получили критерий, оценивающий плотность компании и косвенным образом оценивающий экономическую интенсивность из предположения, что предприятия предпочитают открывать филиалы и точки реализации продукции в экономически активных зонах, при этом многие компании, которые открылись не в экономически благоприятных зонах имеют больше шанса столкнуться с рынком слабо насыщенным покупателями.

# 3.3 Разработка базы данных для хранения промежуточных данных и представления результатов расчета

Вопрос выбора базы данных для нового проекта является весьма опасным, поскольку ошибка на ранних этапах может повлечь за собой множество проблем в будущем, когда проект уже выложен в открытый доступ и его данные постоянно изменяются пользователями. Перенос информации из одной базы данных в другую также осложняется и наличием связей между таблицами, которые надо будет обязательно соблюсти.

Мы не будем погружаться во все возможные проблемы в случае ошибок, поскольку нашу систему они по самой архитектуре затрагивают достаточно слабо. Причина для этой уверенности достаточно проста – Геоинформационная система для поддержки принятия решений является обрабатывающим запросы пользователя, однако не предполагающее обилия функционала, позволяющего хранить большие объемы данных пользователей, а возможно и вовсе не должна иметь возможности загрузки сторонних данных кем либо, кроме разработчиков.

Указанная ситуация ставит наш проект в достаточно выгодное положение и возможно делает этот проект одним из немногих, которые могут позволить себе вольности в вариантах использования базы данных.

Важно отметить, что база данных тем не менее должна быть достаточно надежной и её функционал обязан покрывать хотя-бы минимальные классические задачи, выполняемые SQLзапросами.

В нашем случае была использована база данных SQLite благодаря наличию требуемого функционала, надежности и простоте использования.

Данные хранятся в формате единой таблицы. Такое решение обусловлено тем, что модификация данных в этом случае производится намного проще, чем при наличии многих таблиц и связей между ними. Для каждого города может использоваться отдельная таблица, что позволит разделить хранимые данные и увеличит быстродействие всей программы при необходимости.

Поскольку мы использовали базу данных SQLite, которая является однопоточной и требует обязательного закрытия после совершения действий, нами было принято решение о разработке дополнительной программы, которая в случае отсутствия доступа к базе данных (что случается например, когда мы пытаемся сделать запрос к базе данных из соседнего потока ) перехватывала исключение и ожидала половину секунды до тех пор, пока операция не совершалась. В контексте классификаций паттернов проектирования эту систему можно отнести к шаблону “адаптер”

Хранение промежуточных результатов допускается с помощью таблиц типа .xls, поскольку их этот формат достаточно неплохо справляется с таблицами больших объемов, способен хранить данные различного типа легко избегая конфликтов типов данных, поскольку нам не требуется указывать тип данных при заполнении ячеек.

# 3.4 Разработка подсистемы многокритериального ранжирования

Система многокритериального ранжирования является ключевой в нашей работе и осуществляет выполнение основной задачи разработанной программы.

Расчеты выполнялись на языке Python с использованием модуля numpy. Система расчетов выполнена путем непосредственного выполнения требуемых операций с помощью циклов, функций и базового, встроенного функционала в язык программирования. Каждый отдельный этап выполнен в отдельном модуле (функции), которая принимает требуемые параметры и возвращает результаты расчета.

Модуль расчетов на вход получает следующие параметры:

1. Массив данных, представляющий список, где каждый элемент является данными по каждой отдельной точке.
2. Значения весов для каждого параметра
3. Вещественные значения для учета их в качестве пороговых величин при вычислении групп привлекательности.
4. Пороговые значения согласия и несогласия в формате двух списков.

Расчеты производятся путем вызова необходимых функций, передачи данных и получения готовых или промежуточных результатов.

В нашем проекте использовался метод ELECTRE II и итоговый расчет производился в следующие этапы:

Базово алгоритм выполнения расчетов можно описать в виде выполнения следующих операций:

1. Вычисление веса каждого критерия
2. Классификация комбинаций (подмножества) на слабые (f), средние (m) и сильные (F) в соответствии с заданными порогами c0, c1, c2.
3. Функция, возвращающая несоответствие из таблицы критериев, веса каждого критерия и порядок ранжирования.
4. Функция классифицирующая расхождения на слабые (f), средние (m) и сильные (F) в соответствии с пороговыми значениями d1 и d2.
5. Проверка утверждения гипотезы вычислением устойчивости утверждения по отношению к выбранным весам, изменяя их один за другим

В результате мы получаем три списка точек, которые разгруппированы в соответствии с заданными значениями порогов. .

# 3.5. Разработка пользовательского интерфейса

Пользовательский интерфейс весьма важная часть любой программы и сервиса, поскольку пользователь, хотя и хочет получить требующийся ему результат, ко всему прочему не хочет столкнуться с проблемами при работе, сложностями и непониманием. Интерфейс должен одновременно предоставлять доступ к функционалу, который может быть иногда сложным и при этом не обременять пользователя настройками, если многие из них не обязательны.

В общих чертах можно описать хороший интерфейс следующим образом:

1. Пользователь видит только обязательные параметры.
2. Необязательные настройки скрыты в отдельное меню и имеют значения по умолчанию.
3. Сложные вещи желательно пояснять в подсказках.
4. Ключевой функционал, как кнопка проведения расчетов или отправки формы должны быть на видном доступном месте.

Разработанная программа имеет интерфейс реализованный с помощью Фреймворка Django и классических языков разметки текста и стилей: html, CSS. Использовались стандартные методы форматирования.

изменения на странице и её обновление производится благодаря использованию шаблонизатора Jinja. С помощью него мы добились выведения результатов, требуемых данных для пользователя.

Разработанный программный интерфейс удовлетворяет базовым нуждам пользователей и не имеет ненужного функционала



Рисунок 4 - Разработанный пользовательский интерфейс.

# 3.6 Программная архитектура системы поддержки принятия решения на основе ГИС

Программный комплекс может быть рассмотрен с разных точек зрения. Для пользователя работа с ней может быть представлена следующими этапами:

1. Введение координат для сравнения
2. Введение уточняющих показателей
3. Получение результата

В свою очередь программное обеспечение работает не совсем так линейно и просто. Во многом работа уже совершена благодаря алгоритмам сбора данных, которые будут обработаны при запросе пользователя. С программной точки зрения после запроса пользователя проводятся следующие этапы:

1. Расчет точек координат, расположенных на кругах в определенных радиусах вокруг указанных пользователем точек.
2. Определение реального (пешеходного) расстояния до каждой точки путем прокладывания маршрута и присваивание к круговым точкам этих расстояний.
3. Сбор информации о адресах домов вокруг каждой точки методом запросов к апи геоинформационных систем. Присваивание к круговым точкам списков домов и удаление дубликатов.
4. Поиск по базе данных юридических лиц, находящихся в доме. Дополнение собранных вокруг каждой точки окружности данных информацией о находящихся в области юридических лицах.
5. Проверка наличия данных о оборотах компании и в случае их отсутствия сбор таковых из соответствующих открытых источников и дополнение базы данных.
6. Передача всей собранной информации в модуль системы расчетов с указанием, выбранной пользователем методологии.
7. Выведение результатов расчетов

При этом стоит отметить, что работа самой программы не имеет явного вида конвейера. В ней участвует управляющий модуль, с которым по средством модуля пользовательского интерфейса взаимодействует с пользователем, получает запросы и контролирует выполнение каждого из этапов операций. По завершении выполнения операций он получает результаты с каждого этапа и направляет данные в следующий модуль.

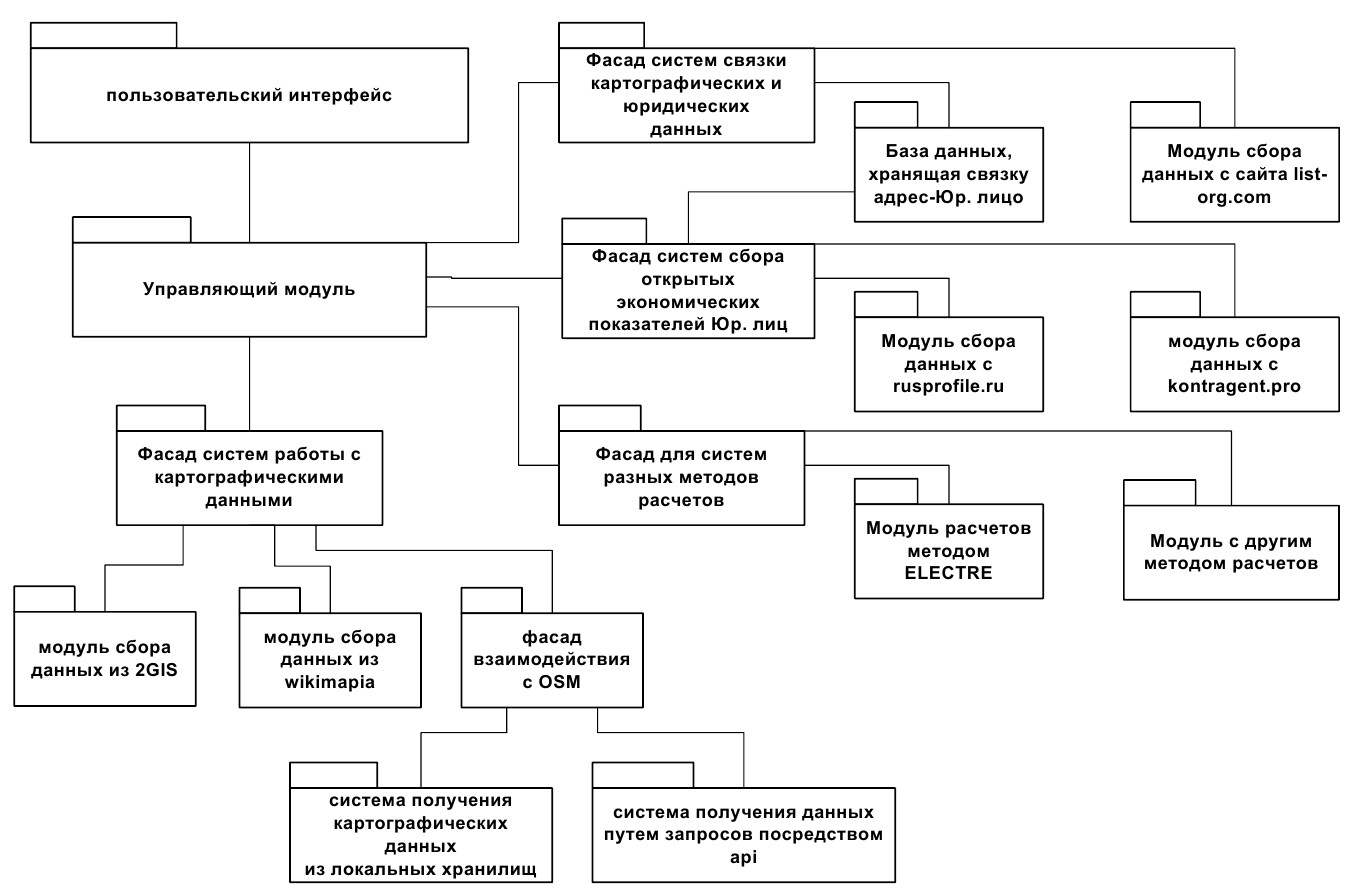


Рисунок 5 - UML-диаграмма системы

Расчеты, проводимые сервисом, заключаются в следующих этапах:

Проведение запросов в базы данных и получение компаний, находящихся в области вокруг каждой точки. Система подсчетов следующая – получая определенные координаты, функция считает вокруг неё квадрат размером 100 футов (это обусловлено особенностями апи сервиса wikimapia) с точкой в центре. После этого в зависимости от требований мы можем получить координаты сетки квадратов вокруг указанной точки. Каждый квадрат в сетке имеет свои координаты и уровень близости к изначальной точке.

К каждому квадрату из базы данных добавляются компании соответствующие их местоположению. На этом этапе мы имеем круги с квадратами , которые имеют свойство близости к изначальной точке и список компаний в качестве свойства каждого квадрата . У компаний есть параметры ширины и долготы, благодаря которым мы вычисляем их принадлежность к определенным зонам.

Знание местоположения также позволяет аналитику определять расстояние и направление между объектами. Расстояние между точками легко рассчитывается с использованием формул для расстояния по прямой на плоскости или на искривленной поверхности Земли, и, приложив немного больше усилий, можно определить фактическое расстояние, которое будет пройдено по дороге или уличной сети, и даже предсказать время, которое потребуется, чтобы совершить путешествие. Расстояние и направление между линиями или областями часто рассчитываются с использованием репрезентативных точек. Их также можно вычислить с использованием ближайшей пары точек в двух объектах или как среднее расстояние и направление между всеми парами точек.

Многие типы пространственного анализа требуют вычисления таблицы или матрицы, выражающей относительную близость разных мест. Близость может быть одним из факторов в объяснении различий во множестве явлений, включая потоки мигрантов, интенсивность социального взаимодействия или скорость распространения эпидемии.

Делается проверка наличия экономических показателей. В случае отсутствия таковых делается вызов дополнительного функционала, собирающего характеристики прибыльности, каждого предприятия, осуществляя запрос в виде ИНН в сервисы открытых источников, как sbis.ru.

Полученные данные отправляются в модуль расчетов, где для них применяется система расчетов методом ELECTRE 2. Полученные результаты выводятся пользователю

# 3.7 Описание и причины использования выбранных технологий

При разработке продукта использовались подходы модульной архитектуры ПО

Архитектура программного обеспечения - это шаблон как для системы, так и для проекта. Он определяет рабочие задания, которые должны выполняться командами разработчиков и разработчиков. Архитектура является основным носителем таких качеств системы, как масштабируемость, производительность, модифицируемость, безопасность и снижение затрат, ни одно из которых не может быть достигнуто без единого архитектурного видения. Архитектура - это артефакт для раннего анализа, чтобы убедиться, что подход к проектированию приведет к приемлемой системе. Создавая эффективную архитектуру, вы можете определять риски проектирования и снижать их на ранних этапах процесса разработки. Архитектура программного обеспечения диктует технические стандарты, включая стандарты кодирования программного обеспечения, инструменты и платформы. Он дает правильные технические решения для обеспечения вашего успеха.

Модульное программирование - это процесс, который включает в себя разделение компьютерной программы на отдельные подпрограммы. Модуль - это отдельный компонент, который может использоваться с рядом функций и приложений в сочетании с другими компонентами в системе. Идентичные функции организованы вместе с одним и тем же программным кодом, а отдельные функции создаются как отдельные блоки кода, что делает их доступными для повторного использования с другими приложениями.

К преимуществам Модульной архитектуры можно отнести следующие пункты:

Разработка может быть разделена между разработчиками: Модульное программирование позволяет разделить разработку путем разделения программы на более мелкие программы для выполнения различных задач. Это позволяет разработчикам работать одновременно и сводит к минимуму время, необходимое для разработки.

Повышение читаемости кода: Модульное программирование помогает разрабатывать программы, которые намного легче читать, поскольку они могут быть включены как определяемые пользователем функции. За программой, несущей несколько функций, легче следить, тогда как за программой, не имеющей функции, следовать гораздо труднее.

Ошибки программирования легко обнаружить: Модульное программирование сводит к минимуму риски, связанные с ошибками программирования, а также упрощает обнаружение ошибок, если таковые имеются. Это связано с тем, что ошибки могут быть сужены до конкретной функции или подпрограммы.

Удобство повторного использования кода: Программный модуль можно повторно использовать в программе, что сводит к минимуму разработку избыточных кодов. Также удобнее повторно использовать модуль, чем писать программу с самого начала. Это также требует написания очень небольшого количества кода.

Улучшение управляемости: Разделение программы на более мелкие подпрограммы упрощает управление. Отдельные модули легче тестировать, внедрять или проектировать. Эти отдельные модули затем можно использовать для разработки всей программы.

Реализация видения: Анализ архитектуры - это эффективный способ увидеть общее состояние ИТ и разработать видение того, в чем организация нуждается или хочет развиваться со своей ИТ-структурой. Архитектор программного обеспечения видит общую картину. Чтобы довести архитектуру до успешного завершения, важно, чтобы кто-то владел общей картиной и продавал видение на протяжении всего жизненного цикла разработки программного обеспечения, при необходимости развивая его на протяжении всего проекта и принимая на себя ответственность за его успешное выполнение.

Определяет области для потенциальной экономии затрат: Архитектура помогает организации анализировать текущую ИТ-инфраструктуру и определять области, в которых изменения могут привести к экономии затрат. Например, архитектура может показывать, что несколько систем баз данных могут быть изменены таким образом, что используется только один продукт, что снижает затраты на программное обеспечение и поддержку. Предоставляет основу для повторного использования. Процесс проектирования может поддерживать как использование, так и создание повторно используемых ресурсов. Ресурсы многократного использования полезны для организации, так как они могут снизить общую стоимость системы, а также улучшить ее качество, учитывая, что повторно используемый актив уже был доказан.

Разработанная программа использует множество дополнительных инструментов. Поскольку в задачи входит использование картографических данных, были использованы технологии, предоставляемые сервисами Graphhopper, OpenStreetMap, Wikimapia, GeoPY.

Работа с базой данных осуществлялась с помощью SQLite и openpyxl.

Сбор данных осуществлялся методом Веб-скрининга с помощью технологий BeautifulSoup4, Selenium, Regular expression.

Интерфейс программы написан с помощью языков HTML, Javascript и шаблонизатора Jinja.

Визуализация данных осуществлялась с помощью пакетов Pandas, numpy и matplotlib

Сбор экономических показателей осуществляется с помощью парсинга сервисов sbis.ru, list-org.com, zachestnyibiznes.ru, rusprofile.ru

GraphHopper - это быстрый и эффективный механизм маршрутизации Java, выпущенный под лицензией Apache License 2.0. По умолчанию он использует данные OpenStreetMap и GTFS, но может импортировать другие источники данных.

OpenStreetMap – сервис, предоставляющий картографические данные для сайтов, мобильных приложений и устройств. OpenStreetMap создан сообществом картографов, которые добавляют и поддерживают данные о дорогах, тропах, кафе, вокзалах и о многих других объектах по всему миру.

Wikimapia - Международный бесплатный сервис, географическая онлайн-энциклопедия, цель которой заключается в том, чтобы отметить и описать все географические объекты на Земле. Викимапия совмещает в себе интерактивную карту с принципом свободного редактирования вики. Сейчас в Викимапии добавлено на карту более 30 млн объектов. Все данные Викимапии доступны для общего пользования под лицензией Creative Сommons Attribution-ShareAlike 3.0.

GeoPY - это клиент Python для нескольких популярных веб-сервисов геокодирования. GeoPY позволяет разработчикам Python легко определять координаты адресов, городов, стран и достопримечательностей по всему миру с помощью сторонних геокодеров и других источников данных.

SQLite - это библиотека на языке C, которая реализует небольшой, быстрый, автономный, высоконадежный, полнофункциональный механизм базы данных SQL. SQLite - это наиболее часто используемый механизм баз данных в мире. SQLite встроен во все мобильные телефоны и большинство компьютеров и входит в состав множества других приложений, которые люди используют каждый день.

Формат файла SQLite является стабильным, кроссплатформенным и обратно совместимым, и разработчики обещают сохранить его в таком виде до 2050 года. Файлы базы данных SQLite обычно используются в качестве контейнеров для передачи богатого контента между системами

openpyxl - это библиотека Python для чтения / записи файлов Excel

Beautiful Soup - это библиотека, которая упрощает очистку информации с веб-страниц. Он располагается поверх анализатора HTML или XML, предоставляя идиомы Pythonic для итерации, поиска и изменения дерева синтаксического анализа.

Selenium WebDriver – это программная библиотека для управления браузерами. Употребляется также более короткое название WebDriver. Это основной продукт, разрабатываемый в рамках проекта Selenium.

WebDriver представляет собой семейство драйверов для различных браузеров плюс набор клиентских библиотек для этих драйверов на разных языках программирования:

Регулярные выражения —язык поиска и осуществления манипуляций с подстроками в строках и тексте

HTML (от англ. HyperText Markup Language — «язык гипертекстовой разметки») — стандартизированный язык разметки документов в веб-страницах. Язык HTML интерпретируется браузерами. полученный в результате интерпретации форматированный текст отображается на экране используемого устройства

JavaScript – мультиплатформенный язык программирования. Поддерживает объектно-ориентированную, императивную и функциональную парадигмы. Является реализацией спецификации.

Язык обычно используется как встраиваемый для программного доступа к объектам приложений. Наиболее широкое применение находит в браузерах как язык сценариев для придания интерактивности веб-страницам

Jinja — это шаблонизатор для языка программирования Python и предоставляет Python-подобные выражения, обеспечивая исполнение шаблонов в песочнице. Поскольку Jinja текстовый шаблонизатор, он может использоваться для создания любого вида разметки, а также исходного кода. Лицензирован под BSD лицензией.

pandas - быстрый, мощный, гибкий и простой в использовании инструмент для анализа и обработки данных с открытым исходным кодом, созданный на основе языка программирования Python.

NumPy - это фундаментальный пакет для научных вычислений на Python. Это библиотека Python, которая предоставляет объект многомерного массива, различные производные объекты (такие как маскированные массивы и матрицы) и набор процедур для быстрых операций с массивами, включая математические, логические, манипуляции с формами, сортировку, выбор, ввод-вывод. , дискретные преобразования Фурье, базовая линейная алгебра, основные статистические операции, случайное моделирование и многое другое.

Matplotlib - это комплексная библиотека для создания статических, анимированных и интерактивных визуализаций в Python.

Производится попарный расчет с помощью метода Electre второго типа, поскольку он более качественно позволяет оценить критерии и в отличие от первой версии можно работает эффективнее, поскольку можно ещё до сравнений убрать из выборки не подходящие вариант только по наличию критериев оценки.

list-org.com – каталог организаций, содержащий регистрационные сведения, контакты, реквизиты, учредители. статистических показателей - финансовых, бухгалтерских, производственных результатов.

# 3.8 Оценка свойств разработанного программного обеспечения

# 3.8.1 Оценка надежности программного обеспечения

Надежность программного обеспечения - это вероятность того, что программное обеспечение будет правильно работать в определенной среде и в течение определенного периода времени.

В случае с разработанным программным продуктом можно говорить о достаточной безотказности, поскольку для работы используется протестированная технология Django последней версии.

Если говорить про отдельные элементы системы, то их надежность соответствует необходимой и в случае отказа одного из них предполагается возможность использования дублирующей. Так, например при невозможность осуществить построения маршрута с помощью апи сервиса graphhopper, можно осуществить измерение расстояния функцией geodesic из библиотеки geopy. Так-же работает и возможность сбора данных. Программа обращается к локальному хранилищу и при необходимости может осуществлять дополнительные запросы к необходимым сервисам для получения недостающих данных.

Устойчивость к нагрузкам тоже может быть увеличена. В первую очередь это касается распределения решаемых управляющим модулем задач на несколько отдельных. Так получение и обработка данных от пользователя может быть отдана одному модулю, контроль этапности задач другому, а представление данных третьему. Тем не менее для текущих задач подобной системы во многом достаточно. Нагрузка так-же вполне выдерживается, во-первых, благодаря используемой многопоточности. Каждый выполняемый этап задачи расчет выделен в отдельный поток. Задачи, выдаваемые программе пользователями, также могут быть выделены в отдельные процессы, что тоже увеличит быстродействие системы. Отдельно имеет смысл поговорить про оптимизацию работы модуля, собирающего данные. Выполняемые им действия не затрагивают расчеты и могут быть выполнены асинхронно с запросами пользователя, что влечёт за собой возможность использования архитектуры микросервисов, функционирующих отдельно друг от друга. Более того мы можем контролировать загрузку вычислительных мощностей настроив сбор данных в то время, когда запросов пользователей достаточно мало, и ограничивать при загрузке основного сервиса.

# 3.8.2 Оценка масштабируемости программного обеспечения

Программное обеспечение благодаря модульной архитектуре при необходимости может быть масштабировано, расширена и дополнена без особых проблем, поскольку при разработке архитектуры программы учитывалась возможность для улучшения. Можно отметить, что не смотря на наличие сложностей вычислительной масштабируемости, поскольку у нас во все данные проходят через управляющий модуль мы можем без особых сложностей изменять, настраивать и добавлять дополнительные этапы работы программы. Подобный подход позволяет добавить как дополнительные предобработку или постобработку для улучшения представления данных, так и параллельные этапы расчета, которые могут выполняться отдельно от представленных систем решения задачи, что позволит расширить возможности программы или предоставить дополнительные результаты, которые возможно обнаружить только одним из нескольких методик решения.

# 4 ИСПОЛЬЗОВАНИЕ РАЗРАБОТАННОЙ СИСТЕМЫ

# 4.1 Выбор территории для застройки

Расчет возможности подключения дома новой застройки к существующим ТЭЦ. Подобная задача на первый взгляд не схожа с поставленной задачей, однако следует отметить, что разработанная система расчетов предполагает возможность выдачи результатов в бинарном формате (Да/Нет). Так, например при необходимости пользователь может ввести данные, формирующие дополнительный параметр для оценки и указать метод учета нового показателя для своей задачи.

Одним из таких случаев можно считать задачу расчета возможности получения достаточного объема тепла за счет распределения нагрузки на тепловые камеры, установленные в области вокруг расчетной точки.

Допускаются изменения с целью проведения расчетов другого характера. В качестве примера представлены следующие изображения программы, выполняющей указанный расчет.

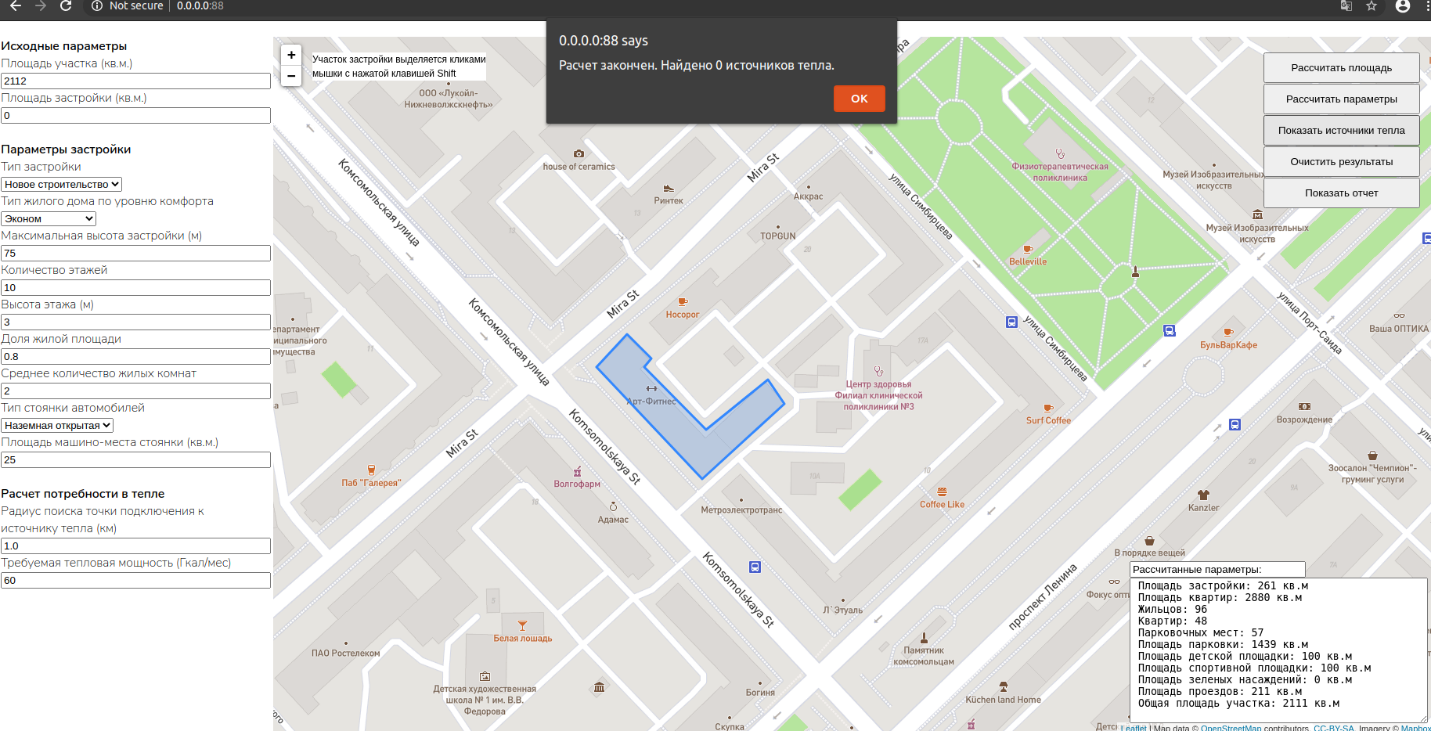


Рисунок 6 - Пример заполнения данных для расчета территории

В ряде случаев возможна потребность выведения более подробных результатов расчетов и дополнительных характеристик. В этом случае допускается использование подхода, в котором может быть открыто отдельное окно с выведением результатов

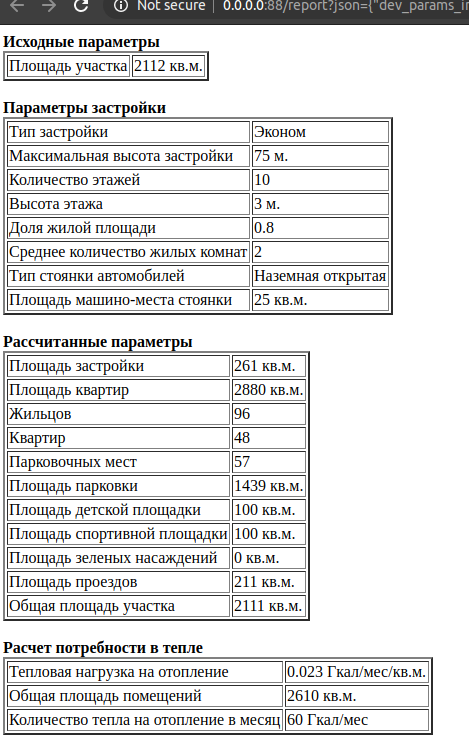


Рисунок 7 – Подробный отчет о проделанных расчетах

# 4.2 Выбор торгового павильона для аренды

Благодаря проделанной работе мы имеем программное обеспечение, способное совершать сбор географических и экономических данных, производить расчеты, инвестор получает возможность получить автоматически созданное экспертное заключение с параметрами причинности выбора определенности зоны.

Разработанная модель учитывает факторы, которые не возможно оценить эмпирическим путем, что позволяет заметно облегчить работу принимающего решение субъекта и улучшить эффективность экономики как следствие.

Важно заметить, что разработанная система не является наиболее эффективным методом решения поставленной задачи и скорее свидетельствует о возможности разработки проведения указанных операций

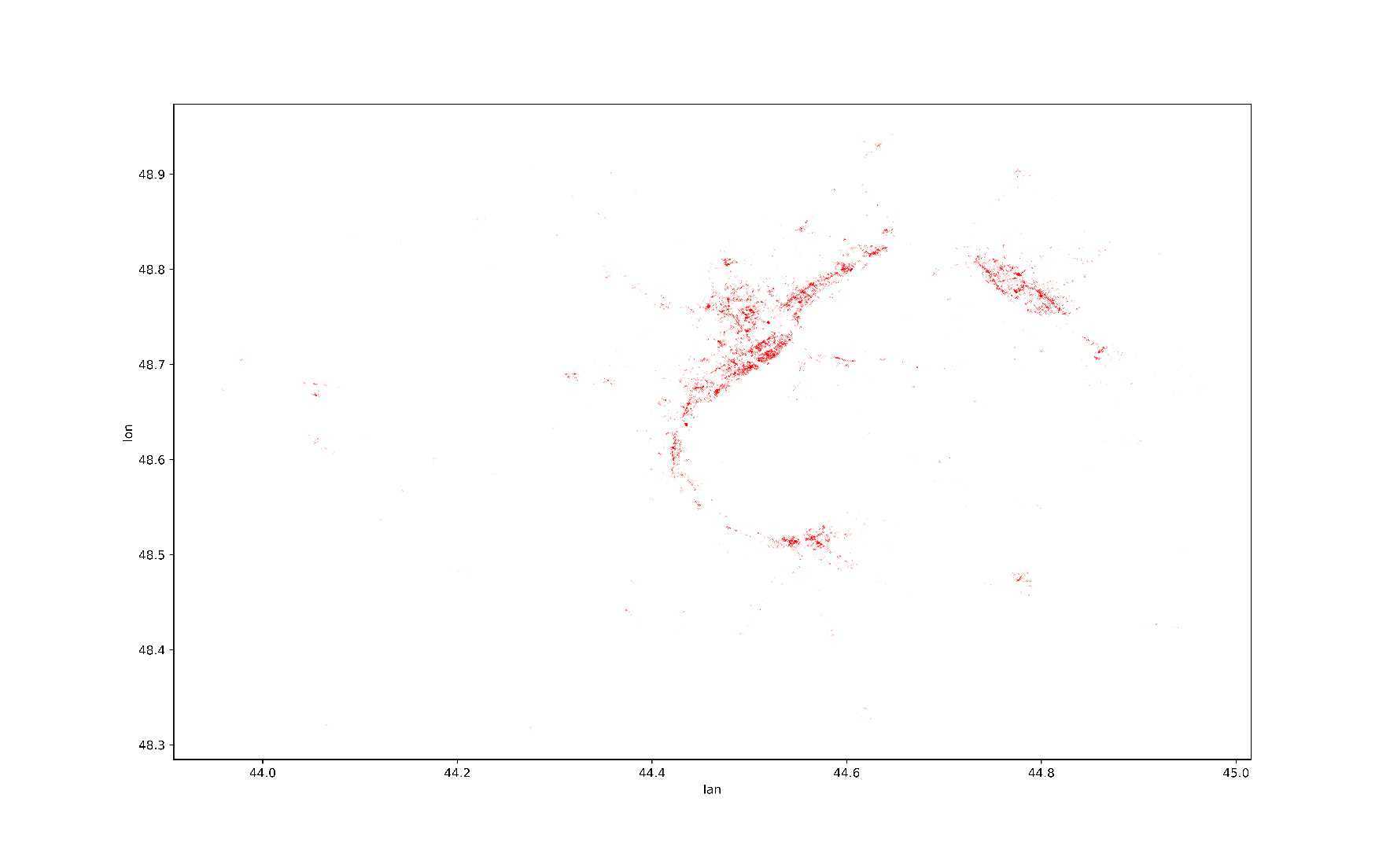


Рисунок 8 - График расположения компаний по всей территории Волгограда и Волгоградской области

Из приведенного изображения видно, что местоположение компаний характеризует области интенсивной экономики и области низкой значимости. Для удобства сравнения представим схематическую карту Волгограда и области:

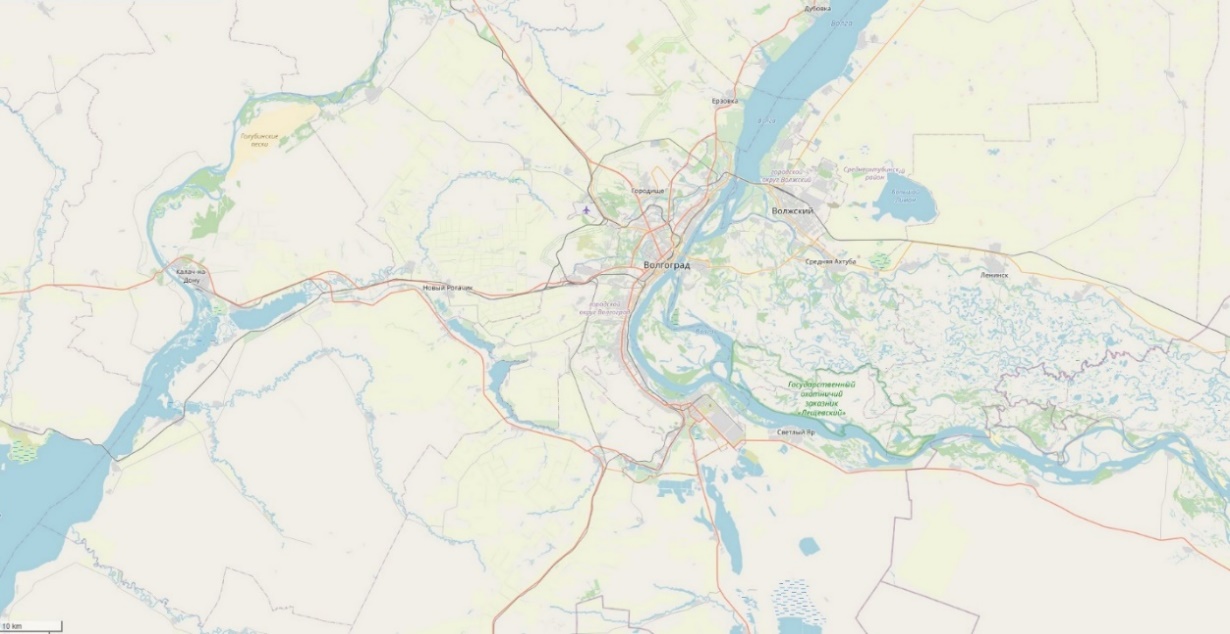


Рисунок 9 - Карта территории Волгограда и области

При приближении мы можем наблюдать плотность расположения компаний и оценивать ценность территорий.

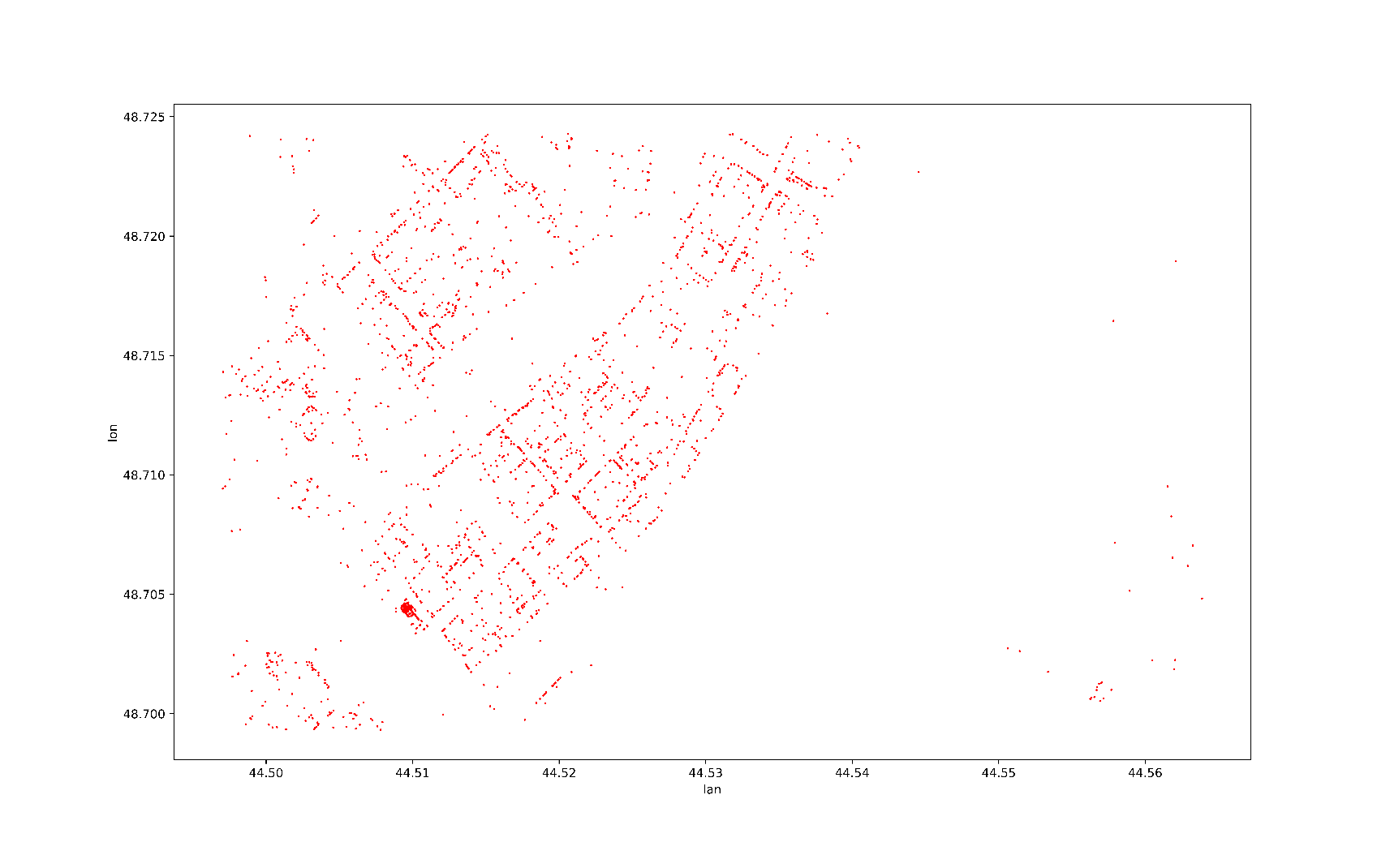


Рисунок 10 - График расположения компаний по территории Центрального района г. Волгограда

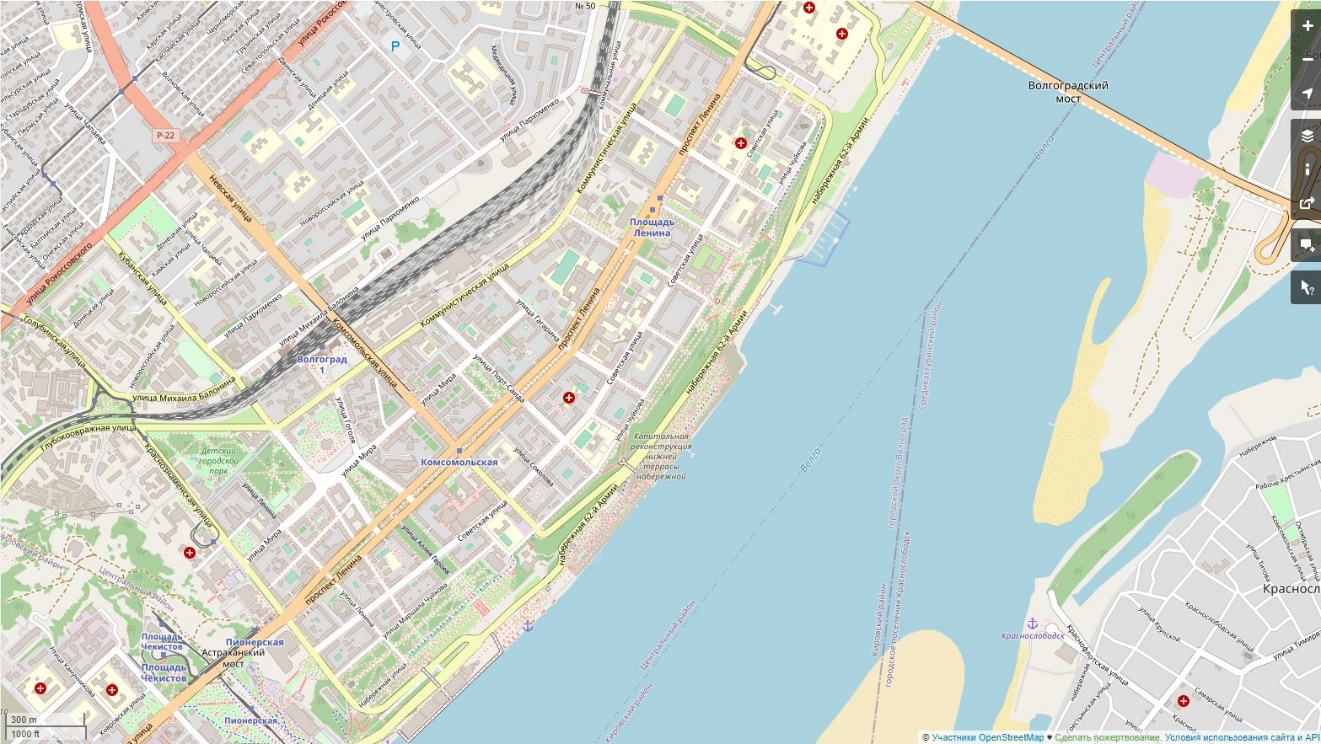


Рисунок 11 - Карта территории Центрального района г. Волгограда

Однако простое расположение точек на карте нам дает мало преимуществ и не позволяет построить реальный график распределения плотности.

Стоит заметить, что тепловые карты – трехмерный график, на котором третья мерность обозначается цветом точки или цветом ограниченной зоны. Поскольку у нас нет готового показателя для указания его в качестве третьей мерности и построения тепловой карты, мы можем таковой создать путем добавлений нового параметра, присваивая некоторое значение к каждой точке. Параметром этим в нашем случае будет являться количество компаний вокруг выбранной. Такой подход исходит из той логики что рыночный отношения уже установили некоторые зоны экономической интенсивности, которые нам следует лишь только выявить. Параметр концентрации компаний является как раз таковым, и он описывает в общих чертах зонирование потребительской активности.

По итогу мы получили новую шкалу, позволяющую построить график тепловой карты. Таковые представлены на следующих изображениях:

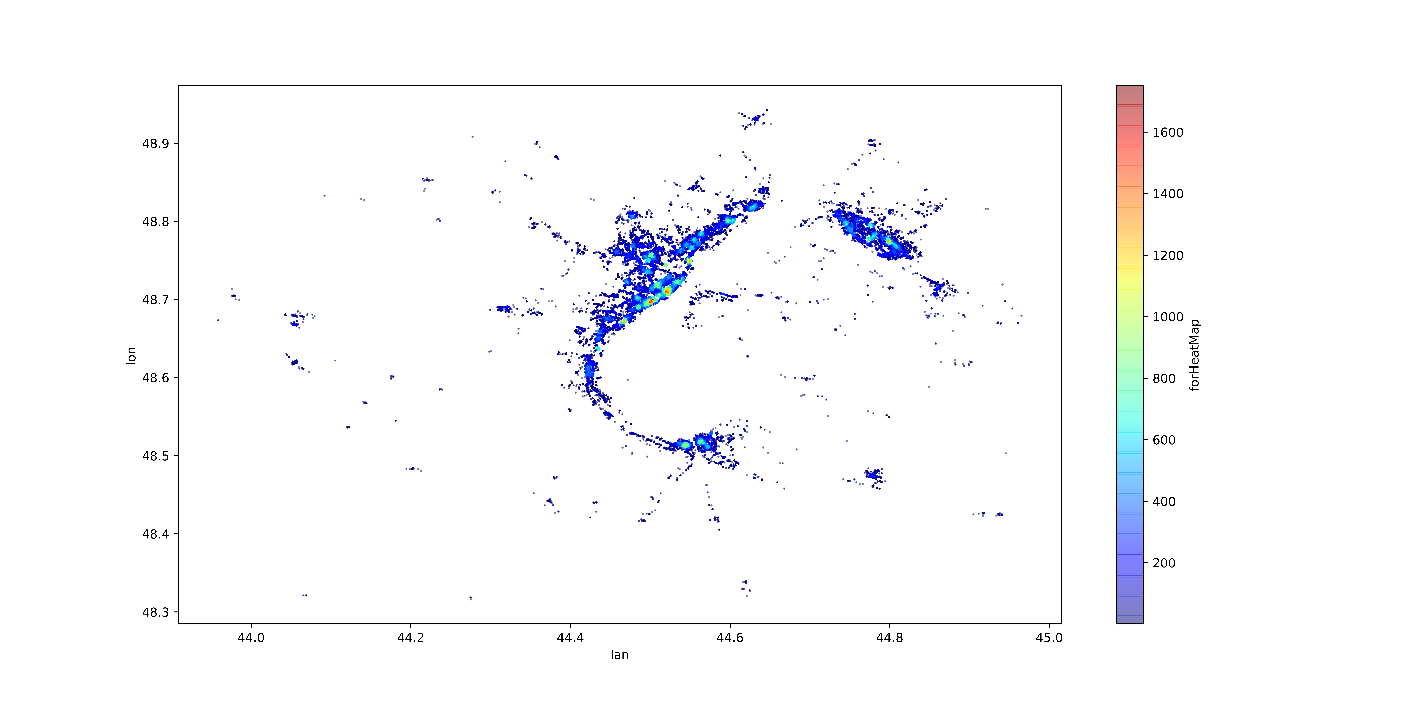


Рисунок 12 - Тепловая карта территории г. Волгограда и области

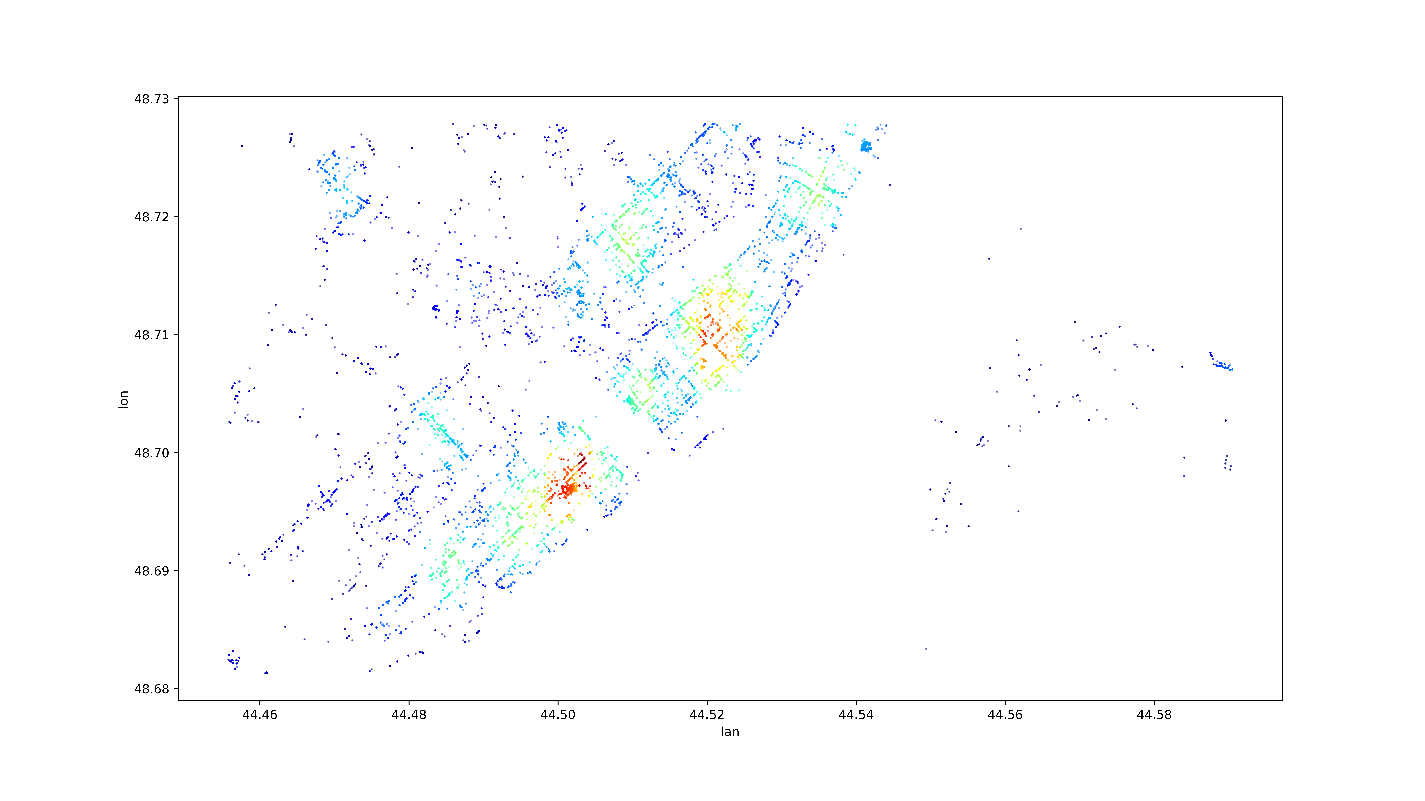


Рисунок 13 - Тепловая карта территории Центрального и ворошиловского района г. Волгограда

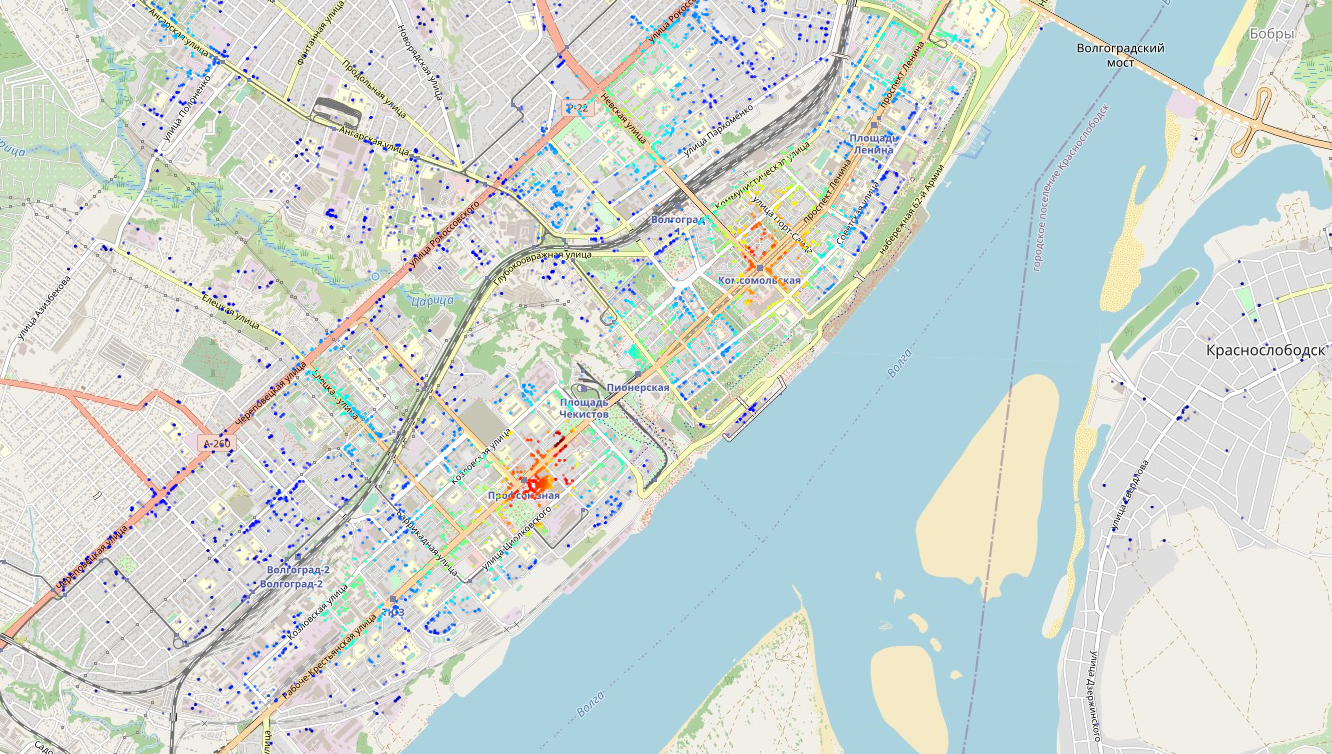


Рисунок 14 - Карта территории Центрального района г. Волгограда c учетом распределения

По окончанию расчетов мы получаем список зон, ранжированных в соответствии с их ценностью. Стоит отметить, что программа хотя и выполняет полезную функцию, всё же выполняет моделирование и установленная точка розничной торговли также будет показывать различия в экономической эффективности от методов управления и маркетинга. К тому-же оценка производится на основании имеющихся данных, которые могут сформировать искаженную определенным образом выборку. Ключевой задачей является автоматизация и увеличение эффективности отбора территорий для размещения торговых точек.

По результатам построения тепловых карт и генерации новых синтетических характеристик могут быть рассчитаны группы точек, среди введенных пользователем в соответствии с ценностью и особенностями

На рисунке 1 представлены результаты расчетов модели для магазина розничной торговли при использовании стандартных дополнительных параметров, которыми являются учет привилегии отдельных факторов (не были заданы особые оценки) и указании пороговых границ риска для отбора результатов в группы.

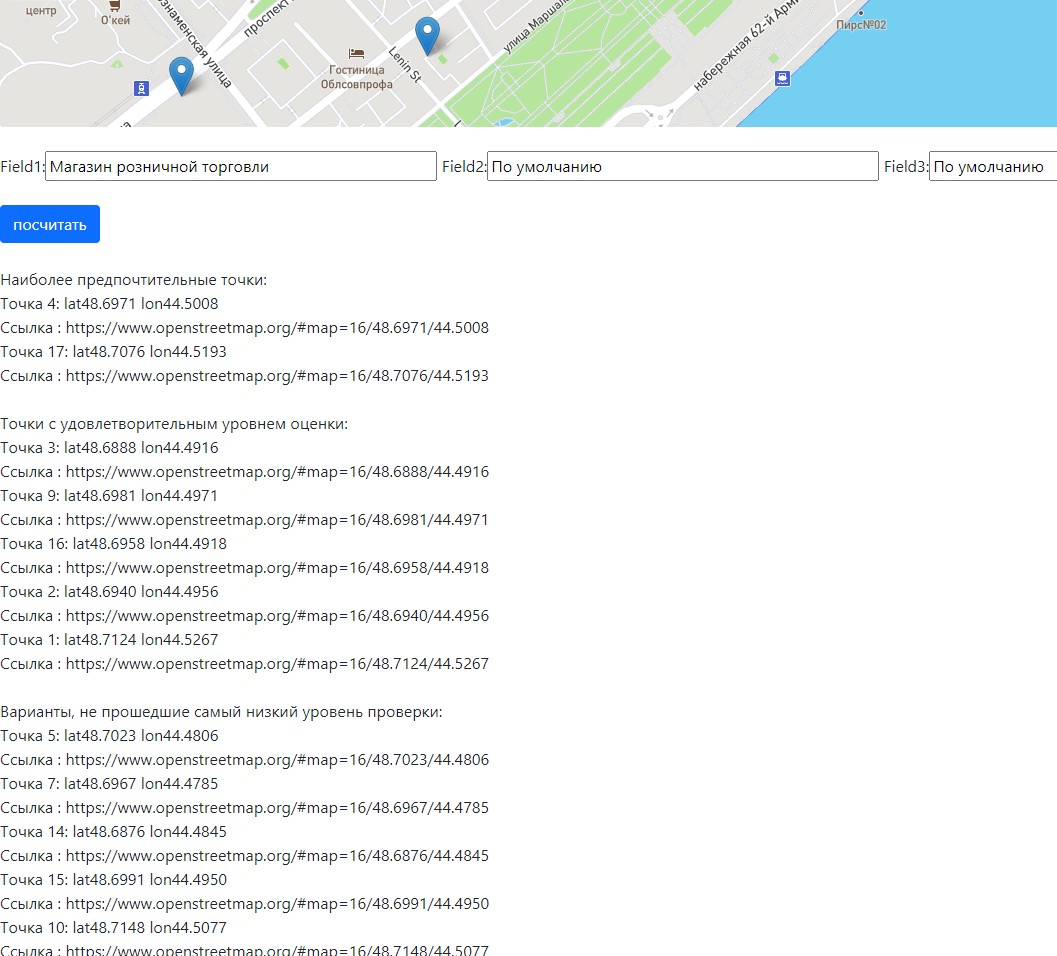


Рисунок 15 – Выведенный результат расчетов

# Заключение

Благодаря проделанной работе мы имеем программное обеспечение, способное предоставлять пользователю аналитику городских территорий с точки зрения их инвестиционной привлекательности путем обработки многофакторных экономических и географических показателей с учетом, дополнительных условий, указанных пользователем в качестве дополнительного параметра.

Архитектура программы позволяет добавлять и менять многие её части с целью её расширения и дополнения новыми возможностями. Для демонстрации этой возможности была продемонстрирована дополнительная система расчета.

Нами были собраны и обработаны данные о компаниях на территории города Волгограда, Волжского и Краснослободска. Построены тепловые карты, характеризующие экономическую активность территорий и разработана математическая модель.

Также допускается добавление вывода дополнительного отчета, содержащего более подробные сведения, картографические данные и графики, потенциально увеличивающие как наглядность принятых решений, так и помогающие дополнить результаты математического анализа экспертным.

К результатам подобных дополнений можно отнести различные тепловые карты натуральных данных, как плотности компаний или прибыльности компаний, но и созданных синтетических характеризующих общую экономическую интенсивность благодаря выполненной свертке нескольких имеющихся параметров.

Расчеты, производимые разработанной программой, могут быть полезны не только для поиска привлекательных территорий, но в теории позволят оценивать недооцененные зоны городов, а в случае дополнения базы данных дополнительными временными срезами появится возможность создавать прогнозные модели. Подобных расчетов в нашем исследовании не проводилось по той причине, что такие данные не представлены в общий доступ.

Помимо указанного был разработан графический интерфейс в формате веб-приложения для предоставления пользователю удобного инструмента доступа к производимым расчетам.

Помимо выполнения своей основной задачи, благодаря заранее продуманной архитектуре допускается дополнение части расчетов новыми методиками, целевая аудитория которых может сильно отличаться от планируемой изначально. Для демонстрации возможности были представлены изображения результатов расчета программы, задача которой заключалась в определении возможности подключения нового строения к тепловым камерам, используемым уже построенными зданиями.

Допускается использование методов нейросетей и иных алгоритмов машинного обучения, однако, поскольку обучение таковых требует наличия выборки данных, содержащей заранее установленные целевые значения, эти расчеты не проводились. В качестве целевых расчетов можно использовать данные полученные построенной нами моделью и обучать модель таким образом, что позволит получить относительно верные результаты, но уже методами машинного обучения.

.

# СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Chang, K. T. (1989). "A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model". International Journal of Geographical Information Science. 3 (4): 323–334.: https://www.tandfonline.com/doi/abs/10.1080/02693798908941519
2. David A Holdstock "Strategic GIS Planning and Management in Local Government". CRC Press. Retrieved 25 October 2017.: Taylor and Francis 280p: https://www.crcpress.com/Strategic-GIS-Planning-and-Management-in-Local-Government/Holdstock/p/book/9781466556508
3. Dr Bhumika Das, Dr Deepak Kumar (May 2015)." Recent Trends in GIS Applications ". Shukla University2015: 1–6.: https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2609707
4. GIS-Based Multiple-Criteria Decision Analysis / Randal Greene, Rodolphe Devillers, Joan E. Luther, Brian G. Eddy. - Текст : электронный // Сибирский психологический журнал. - 2011. - N 5. – P.412-432. - URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-8198.2011.00431.x (дата обращения: 17.10.2020). - Режим доступа: Научная электронная библиотека Wiley Online Library.
5. Greene, R.; Devillers, R.; Luther, J.E.; Eddy, B.G. (2011). "GIS-based multi-criteria analysis". Geography Compass. 5/6 (6): 412–432: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-8198.2011.00431.x
6. Hossein Aghajani. Identifying urban neighborhoods with higher potential for social investment using GIS-FIS approach / Hossein Aghajani & Ali Alizadeh-Zoeram // Applied Geomatics. – 2020. – № 13. – P.1-13.
7. How are maps made? - 2019. / Heather Smith. - URL: https://www.esri.com/about/newsroom/blog/how-are-maps-made/ (дата обращения: 04.04.2020)
8. Marcos M.R. Spatial Analysis to Identify Urban Areas with Higher Potential for Social Investment / Martines, Marcos & Toppa, Rogério & Ferreira, Ricardo & Cavagis, Alexandre & Kawakubo, Fernando & Morato, Rúbia. //Journal of Geographic Information System. – 2017. – № 09. – P.591-603.
9. Mennecke, Brian E.; Lawrence, A. West Jr. (October 2001). "Geographic Information Systems in Developing Countries: Issues in Data Collection, Implementation and Management". Journal of Global Information Management. 9 (4): 45–55.: https://lib.dr.iastate.edu/scm\_pubs/6/
10. Ming-Che Wu.The ELECTRE multicriteria analysis approach based on Atanassov’s intuitionistic fuzzy sets / Ming-Che Wu, Ting-Yu Chen // Expert Systems with Applications. – 2011. – Vol. 38. № 10. – P. 12318-12327.
11. Overpass API User's Manual Режим доступа: https://dev.overpass-api.de/overpass-doc/en/ (дата обращения 13.11.2020), свободный. – Заглавие с экрана. – Яз. анг.
12. Quantitative Methods and Socio-Economic Applications in GIS / Fahui Wang / CRC Press. – 2014. – № 2. – P.1-13. ISBN: 9781466584723
13. Quantitative Methods and Socio-Economic Applications in GIS / П.Л.Кириллов, М.С.Савоскул - Текст : электронный // socio-economic geography in russia - 2015. - N 1. - С. 37-42. - URL: https://elibrary.ru/item.asp?id=23890376&pff=1 (дата обращения: 13.03.2020). - Режим доступа: Научная электронная библиотека eLIBRARY.RU.
14. Stępniaka Cezary. Possibilities of Using GIS Technology for Dynamic Planning of Investment Processes in Cities / Cezary Stępniaka, Tomasz Turek.// Procedia Computer Science . – 2020. – Vol. 176 № 2. – P. 3225-3234.
15. Wang, F. (October 2005). "Job Access and Homicide Patterns in Chicago: An Analysis at Multiple Geographic Levels Based on Scale-Space Theory ". Journal of Quantitative Criminology. 21: 195–217.: https://link.springer.com/article/10.1007%2Fs10940-005-2492-5#Abs1
16. Вид со спутника: как геоаналитика помогает искать места для торговых точек. / Д.Р.Струков - 2016. - URL: https://retail.cnews.ru/articles/2016-10-24\_vid\_so\_sputnika\_kak\_geoanalitika\_pomogaet\_iskat\_mesta\_dlya (дата обращения: 05.10.2020)
17. Агентно-ориентированная модель динамики города с использованием гис-технологий / С. П. Семенов, В. В. Славский, А. О. Ташкин - Текст : электронный // Вестник югорского государственного университета - 2019. - N 1. - С. 66-77. - URL: https://e.lanbook.com/reader/journalArticle/535494/#2 (дата обращения: 21.04.2021). - Режим доступа: Научная электронная библиотека Лань.
18. Будикин А. Е. Современные тенденции, проблемы и перспективы развития географических информационных систем в России / А. Е. Будикин, Д. В. Андреев. // Московский экономический журнал. – 2018. – № 5-3 –С. 16-19.
19. Быканова А. С. Разработка системы поддержки принятия решений на основе данных / А. С. Быканова, В. В. Соболе.// Сибирский аэрокосмический журнал. – 2017. – № 21-2. –С. 311-112.
20. Геоинформатика : в 2 кн. Кн. 1 : учебник для студ. высш. учеб. заведений / [Е.Г. Капралов, А.В.Кошкарев, В.С.Тикунов и др.]; под ред. В. С.Тикунова. — 3 е изд., перераб. и доп. — М. Издательский центр «Академия», 2010. — 400 с.
21. Геоинформационные системы в менеджменте. Учебное пособие/ С. Г. Казаков, К. Г. Дочева, Г. Н. Сухорукова. — Москва: Изд-во ФГБОУ ВПО «Российский экономический университет им. Г. В. Плеханова», 2015.- 134.С.
22. Геоинформационные системы для бизнеса / Е.А.Колупаева. - Текст : электронный // Сервис в россии и за рубежом. - 2009. - N 1. - С. 96-99. - URL: https://elibrary.ru/item.asp?id=12000354 (дата обращения: 23.02.2021). - Режим доступа: Научная электронная библиотека eLIBRARY.RU.
23. Опционы: Разработка, оптимизация и тестирование торговых стратегий / Израйлевич С.И, Цудикман В.Б. // - 2017.. –С. 343-145.
24. Кому на Руси жить хорошо? Как мы искали самый “зеленый” город с помощью OpenStreetMap и Overpass API Режим доступа: https://habr.com/ru/post/520524/ (дата обращения 27.04.2021), свободный. – Заглавие с экрана. – Яз. рус.
25. Кузнецов М.А. Использование методов ELECTRE в задачах принятия решения / М.А. Кузнецов, Т.У.Н. Нгуен // Прикаспийский журнал: управление и высокие технологии. – 2010. – № 2 (10). – С. 40-46.
26. Ловцов Д. А., Черных А. М. Uеоинформационные системы. -учебное пособие // Москва, Издательство: Российская академия правосудия 2012, 192 с.
27. Оценка недвижимости г.Браслава с использованием гис-технологий / Е.Т.Борковский, Л.В. Гурьянова. - Текст : электронный // Вестник БГУ. - 2011. - N 2. - С. 91-93. - URL: https://e.lanbook.com/reader/journalArticle/206007/#1 (дата обращения: 24.05.2021). - Режим доступа: Научная электронная библиотека Лань.
28. Регулярные выражения (дата обращения: 02.06.2021). - Режим доступа: Общедоступная электронная библиотека https://ru.wikipedia.org/wiki/Регулярные выражения
29. Скворцов А.В Геоинформатика: Учебное пособие: Томск: Изд-во Том. ун-та. 2006. — 336 с.
30. Создание и развитие гис «инвестора» г. иркутска / Г.М. Ружников, А.Е. Хмельнов, А.С. Гаченко. - Текст : электронный // Вестник бурятского государственного университета . - 2013. - N 9. - С. 76-80. - URL: https://e.lanbook.com/reader/journalArticle/215472/#1 (дата обращения: 03.12.2020). - Режим доступа: Научная электронная библиотека Лань.
31. Технология создания и ведения муниципальной геоинформационной системы / А.С. Гаченко, Г.М. Ружников, А.Е. Хмельнов, Ю.А.Новицкий, Е.С.Фереферов - Текст : электронный // Информационные системы и технологии. - 2016. - N 32. - С. 32-46. - URL: https://cyberleninka.ru/article/n/tehnologiya-sozdaniya-i-vedeniya-munitsipalnoy-geoinformatsionnoy-sistemy/viewer (дата обращения: 04.06.2021). - Режим доступа: Научная электронная библиотека https://cyberleninka.ru/.
32. Тикунов В.С., Капралов Е.Г. Геоинформатика. Учебник для студентов ВУЗов: М Издательский центр "Академия",2008. — 386 с.
33. урлапов В. Е. Геоинформационные системы в экономике: Учебно-методическое пособие. — Нижний Новгород: НФ ГУ-ВШЭ, 2007. — 104 с.
34. Хортонен А. А. Географические информационные системы как класс систем поддержки принятия решений при управлении пространственной информацией (на примере банковской сферы) / А. А. Хортонен. // Вестн. Астрахан. гос. техн. ун-та. Сер. управление, вычисл. техн. информ. – 2011. – № 1 –С. 195-199.
35. Цветков В. Я. Системный анализ ГИС // Образовательные ресурсы и технологии. — 2015. — № 1(9). — с.97-103.
36. Шайтура С. В. Обзор технологий создания геоинформационной продукции // Информационные технологии. 2001. № 9. — С. 27-30.
37. Юрченко Т. В. Информационные технологии в экономике. Решение экономических задач средствами MS EXCEL 2007 : Учебное пособие / Т. В. Юрченко; Нижегород. гос. архит.-строит. ун-т — Н.Новгород: ННГАСУ, 2010. — 132 с.