

Санкт-Петербургский государственный университет

*на правах рукописи*

Фурсов Дмитрий Викторович

**Интеллектуальная система  
поддержки принятия управленческих решений  
в задаче распространения информации**

Научная специальность: 2.3.1. Системный анализ, управление и обработка информации, статистика

Диссертация на соискание ученой степени  
кандидата технических наук

Научный руководитель: доктор  
физ.-мат. наук, доцент  
Крылатов Александр Юрьевич

Санкт-Петербург  
2024

## Оглавление

<b>Введение</b> . . . . .	3
<b>Глава 1. Моделирование наборов площадок в задаче распространения информации на основе методов оптимизации</b> . . . . .	18
1.1. Постановка и описание задачи . . . . .	18
1.2. Оптимизационная модель . . . . .	22
1.2.1. Анализ источников релевантных статистических данных	22
1.2.2. Алгоритм предобработки статистических данных . . . . .	26
1.2.3. Особенности реализации оптимизационной модели . . . . .	37
1.3. Численное моделирование и анализ результатов . . . . .	41
1.4. Основные результаты и выводы по первой главе . . . . .	48
<b>Глава 2. Моделирование наборов площадок в задаче распространения информации на основе методов машинного обучения</b> .	50
2.1. Постановка и описание задачи . . . . .	50
2.2. Комплексная модель с применением методов кластеризации . . . . .	55
2.2.1. Методы отбора и выделения признаков . . . . .	56
2.2.2. Методы кластерного анализа и метрики качества разбиений	64
2.2.3. Описание и реализация комплексной модели . . . . .	69
2.3. Численное моделирование и сравнительный анализ результатов .	72
2.4. Основные результаты и выводы по второй главе . . . . .	86
<b>Глава 3. Интеллектуальная система сценарного моделирования</b>	87
3.1. Проектирование интеллектуальной системы . . . . .	87
3.2. Реализация и особенности применения интеллектуальной системы	91
3.3. Сравнительный анализ результатов моделирования . . . . .	98
3.4. Основные результаты и выводы по третьей главе . . . . .	105
<b>Заключение</b> . . . . .	106
<b>Список литературы</b> . . . . .	109

## Введение

### Актуальность темы исследования

В современном мире передача информации играет важную роль и охватывает все сферы жизнедеятельности человека: от личной переписки до обеспечения безопасности целых регионов. Сегодня имеется возможность распространять информацию с использованием разнообразных каналов и средств, таких как: сеть Интернет, телевидение, радио, печатные издания и другие. Благодаря развитию технологий, время передачи информации уменьшилось, а технические средства обмена данными стали доступными, вследствие чего вырос объем передаваемой информации. Согласно анализу отчета «Global Digital 2023» [5] за 2023 год число пользователей сети Интернет превысило 64% относительно всего населения Земли в сравнении с 2010 годом – 29%, а количество активных пользователей социальных сетей составило более 59%. Такие глобальные изменения в области информационных технологий существенно повлияли на процессы, протекающие в различных предметных областях. Компании, организации и государственные институты вынуждены адаптироваться к новой цифровой реальности и для этого им необходимы современные инструменты анализа больших данных и моделирования процессов, связанных с распространением информации<sup>1</sup>. Исходя из особенностей определенной области науки, специалисты применяют соответствующие математические методы и виды моделирования для осуществления эффективного распределения ресурсов, автоматизации процессов управления и поддержки принятия решений.

В различных исследованиях освещается проблема резкого роста количества пользователей сети Интернет по всему миру и кратно возросшего объема генерируемых ими данных, начиная с 1990 года по настоящее время. Для решения этого вопроса применяются различные подходы и методы из таких разделов математики как: теория графов, имитационное моделирование, исследование операций, теория вероятностей, математическая статистика, машинное обучение,

---

<sup>1</sup> В соответствие со ст. 2 Федерального закона РФ от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации».

нейронные сети и другие. В данном диссертационном исследовании рассматриваются оптимизация и методы машинного обучения в решении проблемы определения площадок распространения информации в средствах массовой коммуникации (СМК). Поскольку под СМК в данной работе понимается социальная сеть<sup>2</sup>, а площадками распространения информации являются сообщества социальной сети, то результаты и выводы, полученные для задач, сформулированных в диссертационном исследовании, могут быть использованы не только в заданной предметной области — информационно-коммуникационных технологий. Это говорит о том, что предлагаемый инструмент поддержки принятия управленческих решений обладает свойством масштабируемости.

Особенностью применения предлагаемого в диссертации подхода является возможность проведения анализа больших данных и пространства признаков, а также формирования различных сценариев распространения информации и рекомендаций для лиц принимающих решения (ЛПР). Применение оптимизационного подхода позволяет сформировать набор площадок распространения информации в СМК при заданных ограничениях и предпочтениях, а методы кластерного анализа позволяют во-первых, понизить размерность в задаче оптимизации при большом количестве объектов, во-вторых получить хорошо отделимое разбиение с компактными кластерами. Преимущество предлагаемого подхода заключается в том, что у ЛПР есть возможность анализировать сценарии распространения информации, полученные путем поиска оптимального решения и построения разбиения с учетом особенностей признакового описания объектов.

Актуальность исследования проблем, связанных с распространением информации в цифровой среде, обусловлена прикладной востребованностью, которая позволяет успешно применять оптимизационный подход и методы машинного обучения для проведения анализа спроса и предложения на рынке товаров и услуг, реализации агитационных мероприятий в СМК для политических целей, анализа поведенческой активности пользователей по территориальной принад-

---

<sup>2</sup> В соответствие с ч. 1 п. 1 ст. 10.6 Федерального закона РФ от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации».

лежности и других задачах, где использование современных коммуникационных технологий имеет определяющее значение для участников информационного обмена. Таким образом, разработка интеллектуальных инструментов поддержки принятия управленческих решений в задаче распространения информации в СМК с применением средств для анализа больших данных является одной из ключевых задач в современном технологическом обществе.

### **Степень разработанности проблемы в литературе**

Анализ научной литературы по применению математических методов для решения задачи распространения информации в СМК показывает, что существуют работы рассматривающие различные подходы и методы [11, 27, 35, 58, 92, 106, 107]. Ярким примером интегрирования систем поддержки принятия управленческих решений являются инструменты с применением алгоритмов машинного обучения и искусственного интеллекта такие как, «Albert» – маркетинговая платформа искусственного интеллекта [40] и «МТС Маркетолог» – рекламная платформа на базе «Big Data МТС» [56] и другие. Однако многие из них обладают существенным недостатком – большие затраты на внедрение, которые могут себе позволить только крупные участники рынка. Кроме того, отметим, что большая часть работ, где рассматривается вопрос распространения информации, имеют прикладное значение либо для маркетинга, либо моделируют информационные противоборства [3, 4, 22, 38, 39, 42, 43, 47, 57].

Изучая научные работы, освещающие различные подходы к моделированию процесса распространения информации, встречаются такие как - [6, 7, 8, 9, 37, 60, 61, 90], в которых предлагаются модели оптимизации распределения ресурсов, в том числе финансовых, при проведении рекламных кампаний. Разработка подобных инструментов для поддержки принятия решений является хорошим средством анализа текущей ситуации спроса-предложения и тенденций на рынке товаров и услуг для ЛПР, однако результатом работы таких моделей не является ответ на вопрос о том, где лучше разместить рекламное объявление при заданных бюджетных ограничениях. Следует отметить, что большой процент работ в данном направлении посвящен построению моделей, которые не удовле-

творяют требованиям масштабируемости и не представляют возможности интегрирования с другими системами для удобства и упрощения автоматизации процессов управления в области коммуникационных технологий [14, 15, 41].

В работе [17] рассматривается задача математического моделирования рекламной компании: проводится анализ современных методов оценки эффективности рекламной деятельности и применяется математическая модель оптимального распределения рекламного бюджета. Отметим, что в этом случае задача распространения информации формулируется так, что применять ее на практике становится возможным только в контексте проведения узконаправленных маркетинговых исследований, а модель оптимального распределения бюджета автор применяет для выявления рекомендаций по оптимизации рекламной политики. Кроме того, отличительной особенностью данной работы является то, что при построении оптимизационной модели, в качестве переменной величины принимается вид рекламы, а целевым показателем служит эффективность. Недостаток предлагаемого подхода к решению задачи заключается в том, что не проводится анализ больших данных и не применяются современные средства и пакеты программ для осуществления процесса моделирования.

Важно отметить, что существуют такие работы как [28, 29, 30, 69, 89, 110, 111], где поднимается вопрос выбора площадок для размещения информации. Это говорит о том, что данное направление исследования является достаточно популярным среди специалистов различных отраслей. При этом анализ научно-предметных публикации показал, что оптимизационные модели с применением статистического анализа данных [103], численные методы и дифференциальные уравнения [18, 19, 25, 31, 44, 68, 86, 97] стали основным инструментом для выявления закономерностей в сложных социально-экономических системах [23]. Однако в силу непрерывного усовершенствования технических и математических прикладных решений существует возможность постановки новых типов задач в данной предметной области.

Некоторые исследователи выбирают методы оптимизации [95] и инструменты численного моделирования [94] для определения периода размещения ре-

кламных записей, эффективных характеристик [53, 54] и метрик оценивания качества проводимых маркетинговых мероприятий [21], формирования оценки влияния информационного воздействия на потребителей [51], а также для поиска закономерностей в сфере медиапространства [93]. Отметим, что диссертационное исследование является продолжением исследования, проводимым автором в бакалавриате и магистратуре [72, 73, 84]. Ранее рассматривалась задача информационного воздействия в СМК с применением методов имитационного моделирования, дескриптивной статистики и построением базы знаний для создания прототипа экспертной системы с использованием актуальных данных. В связи с чем можно заключить, что задача моделирования распространения информации с применением современных технических средств и прикладного математического аппарата хоть и является хорошо изученной, она все же сохраняет актуальность в исследовании различных моделей.

## **Цель и задачи исследования**

Целью данной работы является *построение интеллектуальной системы с применением методов системного анализа сложных прикладных объектов и современных методов обработки информации для оптимизации процесса поддержки принятия управленческих решений в задаче распространения информации в СМК путем разработки новых и совершенствования существующих методов и средств анализа обработки информации и управления сложными системами*. В качестве объекта исследования рассматривается распространение информации в СМК, а предметом исследования является совокупность методов теоретического и экспериментального анализа указанного процесса, а также инструментов моделирования результатов продвижения информации с применением методов оптимизации и машинного обучения.

Для достижения поставленной цели исследования необходимо выполнить следующие задачи:

1. *Сформулировать постановки задач определения набора площадок распространения информации*. Для этого требуется проанализировать источни-

ки научной литературы на предмет существующих постановок задач в данной предметной области.

2. *Провести анализ сервисов, предоставляющих актуальные данные СМК и реализовать импорт данных.* Для этого необходимо сформулировать критерии для выбора наиболее подходящего сервиса.
3. *Провести анализ существующих подходов к построению интеллектуальной системы поддержки принятия решений.* Необходимо проанализировать источники научной литературы, посвященных теоретическим и прикладным исследованиям для решения научных и технических проблем, связанных с распространением информации.
4. *Разработать программный компонент, реализующий алгоритм предобработки статистических данных о пользовательской активности в СМК.* Необходимо произвести анализ структуры файлов содержащих статистические данные, выделить соответствующие типы данных и реализовать циклический алгоритм обработки выделенных типов данных.
5. *Разработать архитектуру и реализовать программный комплекс для формирования сценариев распространения информации для ЛПР с применением современных средств программирования.* Выделить ряд методов машинного обучения без учителя и оптимизации, а также методы сжатия пространства признаков и реализовать интеллектуальную систему в виде программного комплекса, состоящего из программных компонент для решения соответствующих типов задач.

## **Научная новизна**

В диссертационном исследовании предложены архитектура и схема интеллектуальной системы поддержки принятия управленческих решения для ЛПР в задаче распространения информации в СМК. Разработка и внедрение таких систем является актуальной задачей во многих сферах жизнедеятельности человека в том числе, в сфере коммуникационных технологий. Построены и про-

граммно реализованы оптимизационная модель и комплексная модель кластерного анализа для формирования различных сценариев продвижения информации с рекомендациями по размещению записи для выбранных площадок.

В моделях исследовано влияния сезонности, бюджета, предпочтений клиента и типов товаров и услуг к динамике поведенческой активности аудиторий площадок. На основе анализа полученных результатов численного моделирования установлено, что существует сезонная составляющая активности участников площадок, а также детерминирована чувствительность критериев при изменении предпочтений. Под критериями в данном исследовании понимается количественная характеристика оценивания вовлеченности аудитории площадки к размещаемой в ней информации. Применение на практике разработанной системы позволит корректировать предпочтения ЛПР в соответствии с заданными номенклатурой товаров, бюджетом и временным интервалом, а также выделенными ключевыми количественными характеристиками.

Проанализировано пространство признаков и выделены наиболее значимые для достижения целей моделирования. Изучено влияние наборов значимых признаков на получение лучшего с точки зрения значений количественных характеристик сценария в зависимости от временного интервала, бюджета, а также номенклатуры товаров и услуг. Сформулированы новые постановки задач для моделирования процесса распространения информации в СМК. Кроме того, успешно применены методы кластерного анализа в оптимизационной задаче для понижения размерности, что позволило сократить время получения различных сценариев распространения информации. Продемонстрирована зависимость значений количества кластеров от бюджета для рассматриваемых входных параметрах системы в программных блоках с применением методов машинного обучения без учителя. Показано, что при применении методов кластеризации для понижения размерности в задаче оптимизации количество кластеров меняется на некоторую малую величину, вне зависимости от используемых методов кластерного анализа и входных параметров модели.

Все основные результаты, представленные в работе, получены автором лично и являются новыми.

## Теоретическая и практическая значимость работы

Данная научная работа имеет теоретическое значение, так как она способствует развитию ключевого направления в анализе цифрового пространства. Исследование в диссертационной работе процесса распространения информации, а также анализ научной литературы позволили сформулировать новые постановки задач в контексте автоматизации вопросов управления в области коммуникационных технологий. Несмотря на то, что в работе исследуются широко известные модели анализа больших данных и оптимизации, хотелось бы отметить, что полученные результаты носят универсальный прикладной характер — в смысле моделирования процессов, встречающихся за пределами маркетинга. Действительно, с учетом специфики заданной предметной области, результаты исследования могут быть соответствующим образом перенесены на такие процессы как, например, оптимизация ресурсов предприятия (экономика и управления), анализ мнений и настроений по регионам (социология), воздействие на общественные массы для формирования образа политического деятеля (политология). Таким образом, важность данного исследования заключается в развитии теории, изучающей процессы, протекающие в СМК, а также в возможности получения различных сценариев размещения информации на определенных площадках с рекомендациями.

Разработанный автором программный комплекс представляет собой практическую значимость для ЛПР в вопросе эффективного распределения финансовых ресурсов и может быть использован для поддержки принятия управленческих решений в задаче распространения информации. Полученные в ходе проведения исследования результаты могут быть учтены в будущем при развитии и реализации новых функциональных возможностей предлагаемой интеллектуальной системы. Кроме того, хотелось бы отметить, что архитектура рассматриваемого программного обеспечения построена таким образом, что удовлетворяет свойствам масштабируемости и интегрирования с другими информационно-коммуникационными системами.

На основе вышеизложенного можно сделать вывод о том, что была предпри-

нята успешная попытка предложить универсальный подход к моделированию процесса распространения информации и разработать новый прикладной инструмент для поддержки принятия управленческих решений, который может быть использован в задачах менеджмента для эффективного распределения ограниченных ресурсов организации при проведении соответствующих мероприятий в цифровой среде, а также в экономическом анализе текущей ситуации на рынке товаров и услуг, задачах социологии и политологии.

### **Краткое описание структуры диссертации**

Структура диссертационного исследования включает в себя введение, три главы, представленные с разбиением на разделы и подразделы, описание основных результатов и выводов — в каждой главе, заключение и список литературы, содержащий 111 источников. Общий объем работы составляет 121 страницу машинописного текста, содержит 3 таблицы и 59 рисунков.

**Во введении** обосновывается актуальность темы диссертационного исследования, приведен обзор литературных источников и обозначена степень разработанности проблемы в литературе, изложены цели и задачи, указаны методология, сформулированы основные результаты исследования. Представлена научная новизна, теоретическая и практическая значимость работы.

**В первой главе** сформулированы постановки задач для моделирования процесса распространения информации в СМК с применением методов оптимизации (раздел 1.1). Проводится анализ сервисов предоставляющих статистические данные СМК, исходя из структуры файлов разработан и имплементирован алгоритм обработки статистических данных, формирующий матрицу объектов-признаков для площадок распространения информации. Используя сформированную матрицу, определены характеристики, позволяющие оценивать вовлеченность и обратную связь аудитории площадок. Реализован блок формирования рекомендаций по размещению платных публикаций на площадках СМК. Предложена схема и программно реализована оптимизационная модель определения оптимального набора площадок распространения информации (раздел 1.2). Проведено численное моделирование при заданных входных параметрах

модели, а также анализ чувствительности критериев в задаче многокритериальной оптимизации, что позволяет корректировать предпочтения пользователей разработанного автором ИТ-продукта. На основе анализа полученных результатов, сформулирован ряд выводов данного исследования (раздел 1.3).

**Во второй главе** сформулированы постановки задач для моделирования процесса распространения информации в СМК с применением методов машинного обучения и оптимизации с предварительной кластеризацией для понижения размерности и сокращения времени формирования сценариев распространения информации при увеличении количества объектов-площадок (раздел 2.1). Разработана архитектура и программно реализована комплексная модель с применением методов оптимизации и машинного обучения без учителя и предложено ее описание. Рассмотрены методы отбора и выделения признаков для задач обучения без учителя, приведены результаты применения указанных методов и сформулированы наблюдения. Рассмотрены методы кластерного анализа и заданы соответствующие гиперпараметры, а также приведены метрики оценивания качества получаемых разбиений. Описана программная реализация блоков, реализующих предлагаемую комплексную модель (раздел 2.2). Проведены обучение моделей машинного обучения и сравнительный анализ результатов моделирования, сформулирован ряд выводов данного исследования (раздел 2.3).

**В третьей главе** представлена в виде схемы разработанная архитектура интеллектуальной системы поддержки принятия управленческих решений в задаче распространения информации в СМК (раздел 3.1). Рассмотрены особенности реализации и применения системы, приведена схема структуры хранения данных (раздел 3.2), а также сформулирован ряд выводов исследования на основе сравнительного анализа результатов численного моделирования (раздел 3.3).

**В заключении** приведено краткое обсуждение полученных результатов и возможных направлений дальнейших исследований.

## **Методология и методы исследования**

Задействованные в работе инструменты являются общепризнанными правилами и подходами к исследовательской деятельности в области прикладной

математики: математическое программирование (теория и методы решения задачи оптимизации), машинное обучение (методы кластерного анализа и сжатия пространства признаков), математическое моделирование, сравнительный анализ, численное моделирование в кроссплатформенной интегрированной среде разработки для языка программирования Python — PyCharm.

### **Степень достоверности и апробация результатов**

Основные результаты, полученные в ходе выполнения исследования, обсуждались и были оформлены в виде докладов на следующих научных мероприятиях: Международная онлайн конференция «The 6th Computational Methods in Systems and Software 2022 (CoMeSySo2022)», секция: «Data Science and Algorithms in Systems», г. Прага, Чешская Республика [101]; Всероссийская конференция по естественным и гуманитарным наукам с международным участием «Наука СПбГУ 2023», секция: «Математика, механика, информатика», г. Санкт-Петербург, Российская Федерация; «XIII Конгресс молодых ученых ИТ-МО 2024», секции: «Большие данные и машинное обучение» [75], «Искусственный интеллект и поведенческая экономика» [71], г. Санкт-Петербург, Российская Федерация; VI Всероссийская с международным участием научно-практическая конференция студентов, аспирантов и работников образования и промышленности — «Системы управления, информационные технологии и математическое моделирование — 2024» в рамках I Международного Форума «ИТ. Наука. Креатив» (iFORUM), секция: «Прикладная математика и информатика в гуманитарных и социально-экономических науках», г. Омск, Российская Федерация [83]; Научный семинар кафедры математической теории экономических решений Санкт-Петербургского государственного университета, г. Санкт-Петербург.

Достоверность и обоснованность результатов данного диссертационного исследования обеспечивается корректностью постановок задач, аргументов и выводов, а также получением положительных рецензий от членов редакционных коллегий периодических научных изданий, в которых были опубликованы основные результаты работы.

## Публикации

Результаты проведенного исследования опубликованы в 6 научных изданиях [71, 74, 75, 76, 77, 83, 101], в том числе, основные результаты диссертационной работы опубликованы в трех научных журналах [74, 76, 77], включенных в перечень рецензируемых научных изданий, рекомендованных ВАК РФ и входящих в РИНЦ. Получено 5 свидетельств о регистрации программы для ЭВМ [78, 79, 80, 81, 82] в Федеральном институте промышленной собственности (ФИПС).

**Личный вклад автора** состоит в самостоятельном определении цели, задач и плана исследований диссертационной работы. Автором самостоятельно проведены численные эксперименты и разработаны программные компоненты и комплекс. Интерпретация, статистическая обработка и окончательная оценка полученных результатов, анализ научной литературы, а также написание текста диссертации осуществлены соискателем самостоятельно.

## Основные научные результаты

1. Сформулированы постановки задач для моделирования процесса распространения информации в СМК с применением методов оптимизации. Описанные результаты получены в первой главе исследования и опубликованы в работе [76].
2. Сформулированы постановки задач для моделирования процесса распространения информации в СМК с применением методов машинного обучения. Описанные результаты получены во второй главе исследования и опубликованы в работе [77].
3. Разработан программный компонент, реализующий на языке программирования Python в кроссплатформенной интегрированной среде разработки PyCharm, циклический алгоритм предобработки статистических данных о пользовательской активности информационных площадок в задаче распространения информации в СМК. Описанные результаты получены в первой главе исследования и опубликованы в работе [76, 79].

4. Разработан программный компонент с рекомендательным блоком для формирования сценариев распространения информации в СМК и решения оптимизационных задач с возможностью трансформации и визуализации информации на языке программирования Python в кроссплатформенной интегрированной среде разработки PyCharm. Описанные результаты получены в первой главе исследования и опубликованы в работе [76, 81].
5. Разработаны программные компоненты с применением методов машинного обучения и отбора признаков без учителя для решения задачи кластеризации информационных площадок и задачи снижения размерности в оптимизационной задаче с возможностью трансформации и визуализации информации на языке программирования Python в кроссплатформенной интегрированной среде разработки PyCharm. Описанные результаты получены во второй главе исследования и опубликованы в работе [77, 80, 82].
6. Разработана архитектура и программно реализована интеллектуальная система поддержки принятия управленческих решений в задаче распространения информации в СМК, а также предложена схема хранения данных и результатов моделирования с возможностью трансформации и визуализации информации на языке программирования Python в кроссплатформенной интегрированной среде разработки PyCharm. Проведено сравнение сценариев распространения информации и продемонстрирована целесообразность формирования нескольких наборов информационных площадок. Описанные результаты получены в третьей главе исследования и опубликованы в работе [74, 78].
7. Разработан инструмент проведения численного моделирования исследуемой системы, позволяющий осуществлять анализ чувствительности критериев, а также анализировать процесс формирования уникальных сценариев распространения информации в результате изменения предпочтений в задаче многокритериальной оптимизации на примере рынка товаров-услуг в цифровой среде с учетом номенклатуры товаров, бюджете

та и временного промежутка. Применение разработанных программных компонент позволяет корректировать предпочтения пользователя, посредством настройки гиперпараметров методов машинного обучения, а также сократить время формирования сценариев распространения информации. Проведение анализа важности признаков позволило определить базовый набор значимых характеристик объектов выбранными методами сжатия пространства признаков. Описанные результаты получены в первой и второй главе исследования и опубликованы в работах [76, 77].

### **Основные положения, выносимые на защиту**

1. Формализация и постановка задач обработки информации и моделирования процесса распространения информации в СМК с применением методов оптимизации и машинного обучения.
2. Специальное алгоритмическое и математическое обеспечение интеллектуальной системы поддержки принятия управленческих решений в области распространения информации в СМК с использованием современных методов обработки и анализа информации, оптимизации и машинного обучения.
3. Программный комплекс для реализации проблемно-ориентированной интеллектуальной системы поддержки принятия управленческих решений в задаче распространения информации в СМК.
4. Функции визуализации, трансформации и анализа информации с применением актуальных компьютерных методов обработки информации в ВІ-инструменте для задачи распространения информации в СМК.
5. Методология проведения прикладного исследования, направленного на выявление, измерение и анализ складывающейся конъюнктуры рынка товаров-услуг в цифровой среде, а также моделирование сценариев распространения информации в СМК.

## Благодарности

Автор выражает благодарность за оказанную всестороннюю поддержку, отзывчивость и помощь в подготовке и оформлении диссертации своему научному руководителю, доктору физико-математических наук, профессору кафедры математической теории экономических решений Санкт-Петербургского государственного университета Александру Юрьевичу Крылатову. Также хотелось бы выразить слова благодарности своему первому научному руководителю и наставнику, кандидату физико-математических наук, доценту Санкт-Петербургского государственного университета Михаилу Владимировичу Свиркину, за поддержку и консультации во время проведения диссертационного исследования, ценные замечания и рекомендации. Кроме того, невозможно не отметить своего коллегу и не выразить признательность кандидату физико-математических наук Кочевадову Виталию Алексеевичу, за оценку трудов и рекомендации в подготовке и оформлении некоторых разделов диссертации. В заключении, хочется поблагодарить своего школьного учителя математики Нелли Искандеровну Амирханову, профессионала своего дела, человека, который демонстрировал каким интересным и увлекательным может быть мир математики.

## Глава 1.

# Моделирование наборов площадок в задаче распространения информации на основе методов оптимизации

### 1.1. Постановка и описание задачи

Применение математических моделей для моделирования процессов в различных прикладных областях играет ощутимую роль при принятии управленческих решений. С их помощью специалисты могут предсказывать или оценивать поведение системы, моделируя различные сценарии при изменении входных данных или при воздействии на систему внешних факторов, прогнозировать результаты каких-либо процессов, тем самым минимизировать потери, как в денежном, так и в репутационном отношении, максимизировать прибыль предприятия и многое другое. Оптимизационные модели являются неотъемлемой частью математического инструментария, используемого как различными государственными институтами, так и бизнесом, для помощи лицам, принимающим решение в сложных условиях с целью проведения полного и объективного анализа предметной деятельности. В данной главе будут рассмотрены некоторые возможные постановки оптимизационных задач выбора площадок распространения информации в СМК.

Поскольку рассматриваемые понятия охватывают весьма широкий спектр средств, инструментов и способов предоставления информации неопределенному кругу лиц, а также исходя из специфики экспериментальной части данного исследования предлагается под распространением информации понимать ее продвижение путем публикации рекламной записи<sup>1</sup>, а под СМК - социальные сети<sup>2</sup>. Для проведения численного моделирования и удобства интерпретации результатов была взята социальная сеть «ВКонтакте» [64]. В данной работе про-

---

<sup>1</sup> В соответствии с Федеральным законом РФ от 13.03.2006 N 38-ФЗ (ред. от 11.03.2024) "О рекламе"

<sup>2</sup> В соответствии со статьей 10.6 Федерального закона РФ от 27.07.2006 N 149-ФЗ (ред. от 12.12.2023) "Об информации, информационных технологиях и о защите информации"

движение информации будет осуществляться в части коммуникативного воздействия [26] на ее потребителя. Для демонстрации работоспособности предлагаемого подхода к моделированию рассмотрим ряд задач из области маркетинга, например, увеличение продаж, позиционирование бренда [32, 59] или создание корпоративного имиджа [20]. В следствие того, что продвижение информации будет происходить в рамках социальных сетей, под площадкой понимается сообщество [46] социальной сети. В сообществе администраторами опубликовываются записи как на бесплатной основе в рамках темы группы, так и в коммерческих целях. Такие записи называются - рекламными. Стоимость размещения подобной записи зависит от многих факторов: активности аудитории, тематики, времени года и так далее. Соответственно, лицо, которое обращается к администраторам сообщества с целью публикации рекламной записи является клиентом.

Для того, чтобы оценивать возможные результаты проведения рекламных кампаний, требуются данные, отражающие активность в сообществе. В данной работе оценивание активности аудитории в сообществе происходит по различным показателям и метрикам записи за выбранный временной период в силу ограниченной возможности получения данных содержащих иную дополняющую информацию.

Итак, рассмотрим возможные постановки задачи формирования набора сообществ для размещения в них рекламных записей.

### **Задача целочисленного линейного программирования**

*Содержательная постановка задачи:* клиенту требуется увеличить объем реализуемой им продукции в рамках определенного бюджета. Клиент хочет провести рекламную кампанию таким образом, чтобы как можно больше пользователей сети узнали о его продукте. Требуется при заданных тематике, временном периоде, целевом параметре и бюджете максимизировать суммарное количество просмотров опубликованных рекламных записей. Отметим, что одна и та же рекламная запись может быть опубликована в нескольких сообществах, а также рекламная запись может быть отредактирована отдельно для каждого выбран-

ного сообщества с учетом особенностей ее аудитории.

*Математическая постановка задачи:* пусть определены универсальное множество сообществ  $X$ , тематики сообществ, желаемый месяц для размещения информации  $t$  ( $t = \overline{1, 12}$ ), бюджет рекламной кампании ( $P > 0$ ). Упорядочим множество  $X$ , то есть установим взаимно однозначное соответствие между множествами  $X$  и  $M = \{1, \dots, n\} \subset N$ , что позволяет конкретизировать стоимость размещения рекламной записи в  $i$ -м сообществе как  $b_i > 0$ , значение выбранного целевого показателя в  $i$ -ом сообществе для среднестатистической записи в заданном месяце года  $t - c_i(t) > 0$ ,  $i \in M$ . Решение задачи представляется определением набора сообществ  $x = (x_1, \dots, x_m) \subseteq X$ ,  $1 \leq m \leq n$ , который удовлетворяет следующим требованиям:

$$f(x) = \sum_{j=1}^m c_j \cdot x_j \rightarrow \max,$$

$$\sum_{j=1}^m b_j \cdot x_j \leq P,$$

$$x_j \in \{0; 1\}, j = \overline{1, m}.$$

### **Задача многокритериальной оптимизации**

*Содержательная постановка задачи:* клиенту требуется провести позиционирование бренда в рамках определенного бюджета. Однако, клиент отдает предпочтение сообществам, в которых люди оставляют обратную связь в комментариях под записями, а также активно делятся записями сообщества у себя на личной странице. Требуется при заданных тематике, временном периоде и бюджете максимизировать количество комментариев и отметок «поделиться» опубликованной рекламной записи. Таким образом, для клиента равновесны два показателя, значения которых и будут максимизироваться. Отметим, что одна и та же рекламная запись может быть опубликована в нескольких сообществах, а также рекламная запись может быть отредактирована отдельно для каждого выбранного сообщества с учетом особенностей ее аудитории.

Под критериями в данном исследовании понимаются отметки «Мне нравится», «Поделиться», «Комментарии», «Просмотры». В общем случае, имея большее разнообразие данных, критериями могут быть, например, охват<sup>3</sup>, количество переходов по ссылкам и так далее.

*Математическая постановка задачи:* к обозначениям введенным на странице 21 добавим 4 критерия, заданных с помощью функций:  $f_1(x)$ ,  $f_2(x)$ ,  $f_3(x)$ ,  $f_4(x)$ ; переопределим стоимость размещения рекламной записи в  $i$ -ом сообществе —  $g_i > 0$ , а также зададим значения соответствующих критериев в  $i$ -ом сообществе для среднестатистической записи в заданном месяце года  $t$  —  $a_i(t) > 0$ ,  $b_i(t) > 0$ ,  $c_i(t) > 0$ ,  $d_i(t) > 0$ ,  $i \in M$ . Решение задачи представляется определением набора сообществ  $x = (x_1, \dots, x_m) \subseteq X$ ,  $1 \leq m \leq n$ , который удовлетворяет следующим требованиям:

$$\begin{cases} f_1(x) = \sum_{j=1}^m a_j \cdot x_j \rightarrow \max, f_2(x) = \sum_{j=1}^m b_j \cdot x_j \rightarrow \max, \\ f_3(x) = \sum_{j=1}^m c_j \cdot x_j \rightarrow \max, f_4(x) = \sum_{j=1}^m d_j \cdot x_j \rightarrow \max, \\ \sum_{j=1}^m g_j \cdot x_j \leq P, \\ x_j \in \{0; 1\}, j = \overline{1, m}. \end{cases} \quad (1.1.1)$$

Применив метод свертки критериев, сведем систему (1.1.1) к системе (1.1.2) и будем ее решать известными методами математического программирования.

$$\begin{cases} f(x) = \sum_{j=1}^m w_j \cdot x_j \rightarrow \max, \\ \sum_{j=1}^m g_j \cdot x_j \leq P, \\ x_j \in \{0; 1\}, j = \overline{1, m}. \end{cases} \quad (1.1.2)$$

где  $w_j = \alpha_1 \cdot a_j + \alpha_2 \cdot b_j + \alpha_3 \cdot c_j + \alpha_4 \cdot d_j$  - это показатель суммарной активности в  $j$ -ом сообществе для «среднестатистической» записи в заданном месяце года  $t$ ,  $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$  - весовые коэффициенты критериев или предпочтения клиента.

---

<sup>3</sup> Уникальные просмотры записи пользователями социальной сети.

## 1.2. Оптимизационная модель

Архитектура оптимизационной модели представлена в виде блок-схемы на рисунке 1.1. Рассматриваемый программный компонент реализован с помощью четырех основных функциональных блоков: 1) предобработка данных; 2) обработка данных; 3) формирование рекомендаций для публикации рекламных записей; 4) оптимизация.



Рис. 1.1: Архитектура программного компонента «Оптимизационная модель»

Однако прежде чем переходить к рассмотрению структур и особенностям реализации этих функций, необходимо провести анализ источников, которые предоставляют релевантные статистические данные сообществ социальной сети «ВКонтакте». Отметим, что разработка собственного программного комплекса, решающего проблему парсинга данных<sup>4</sup> для различных интернет-платформ является отдельной трудоемкой задачей, поэтому в данном исследовании автором было принято решение воспользоваться существующими инструментами по извлечению статистических данных по сообществам социальной сети.

### 1.2.1. Анализ источников релевантных статистических данных

Для проведения мероприятий по распространению информации в СМК необходимо анализировать данные, доступ к которым предоставляют специальные

<sup>4</sup> Парсинг данных — это извлечение структурированной информации из неструктурированных или полуструктурированных данных.

сервисы. В данном разделе будет приведен ряд таких ресурсов и выбран такой, который отвечает следующим критериям:

1. Универсальность использования — сервис предоставляет статистику по разным социальным медиа ресурсам.
2. Возможность делать выгрузки по выбранному количеству сообществ.
3. Возможность указывать период выгрузки.
4. Достаточное количество характеристик сообществ, предоставляемых сервисом (более 10).
5. Формат выгрузки: табличный по характеристикам. Желательный формат файлов: `xlsx`, `txt`, `csv`, `json`, `xml`.
6. Стоимость использования сервиса.

Сформулированные критерии позволят исследователю найти релевантные статистические данные за короткий временной промежуток.

Рассмотрим несколько наиболее популярных ресурсов среди специалистов отрасли интернет-маркетинга:

- «LiveDune» — сервис осуществляет проверку аккаунтов на накрутки и предоставляет другую различную статистику, необходимую для принятия решения о размещении рекламы [63]. Также сервис предоставляет ежечасную статистику по аккаунтам кампании и конкурентов. Соответствие сформулированным критериям:

1. Медиа ресурсы: «ВКонтакте», «Одноклассники», «Telegram», «TikTok», «YouTube».
2. Есть, но сильно ограничена. Чтобы получить данные по 3000 сообществам, требуется заплатить более 9990 рублей.
3. Есть.

4. Более 50 различных метрик.
  5. Формат выгрузки – pdf, xlsx.
  6. Стоимость – более 9990 рублей в месяц.
- «Pur Ninja» — это сервис для отложенного постинга и аналитики в социальных сетях [1]. Умеет публиковать видео, добавлять водяные знаки на медиа, отображать, как точно будет выглядеть пост после публикации в социальные сети. Соответствие сформулированным критериям:
    1. Медиа ресурсы: «Telegram», «ВКонтакте», «Одноклассники».
    2. Есть, но сильно ограничена. Чтобы получить данные по 3000 сообществам, требуется оплатить 30-кратный тарифный план «Бизнес L».
    3. Есть.
    4. Достаточное количество метрик.
    5. Формат выгрузки - xlsx, csv.
    6. Стоимость – от 200 тысяч рублей в месяц.
  - «Popsters» — это сервис аналитики контента, статистики и сравнения сообществ в 12 социальных сетях [62]. Он позволяет: оценить популярность различных постов, учитывая содержание, формат, объем текста и др.; позволяет произвести быстро в автоматическом режиме расчет показателей эффективности постов, имеющих разные хештеги и прикрепления; позволяет получить статистику, проанализировать и сравнить результативность различных кампаний в разных сообществах. Соответствие сформулированным критериям:
    1. Медиа ресурсы: «ВКонтакте», «Одноклассники», «Telegram», «TikTok», «Pinterest» и «YouTube».
    2. Есть, без ограничений.
    3. Есть.
    4. Более 20 метрик.

5. Формат выгрузки - позволяет сформировать и выгрузить в удобном формате (xlsx, jpg, pdf, pptx, png, csv) отчеты и графики на основе статистики.
  6. Стоимость – от 499 рублей в месяц.
- «JagaJam» — это компания, которая создает сервисы для работы с данными из социальных сетей [65]. Статистика аккаунтов в социальных сетях: лучшие посты, динамика подписчиков, вовлеченности, отметок «Просмотры», «Мне нравится», «Комментарии» и «Поделиться» за любой период. Компания предоставляет: рейтинг сообществ, позволяя выбрать лучшие для продвижения; данные в удобном для сравнения виде (одновременно до 10 сообществ) и многое другое. Соответствие сформулированным критериям:
    1. Медиа ресурсы: «ВКонтакте», «Одноклассники», «Telegram», «TikTok», «YouTube» и другие.
    2. Есть, без ограничений.
    3. Есть.
    4. Достаточное количество.
    5. Стоимость – от 9890 рублей в месяц.
    6. Формат выгрузки - xlsx.
  - «AllSocial» — это сервис с большим количеством полезной аналитической информации сообществ социальной сети «ВКонтакте» [66]. Сервис предоставляет информацию о стоимости размещения рекламной записи в сообществах, которая соответствует биржевой цене «Sociate». Соответствие сформулированным критериям:
    1. Медиа ресурсы: «ВКонтакте».
    2. Есть, без ограничений.
    3. Есть.

4. Достаточное количество.
5. Стоимость – бесплатно.
6. Формат выгрузки - xlsx.

Из перечисленных сервисов были выбраны — «Popsters» и «AllSocial». Первый сервис удовлетворяет большинству критериев и имеет достаточно удобную структуру файлов выгрузки статистических данных для дальнейшего их преобразования и использования. Второй сервис содержит дополнительные метрики, статистические данные, а также одну из ключевых характеристик для данного исследования — стоимость размещения рекламной записи в сообществе социальной сети, которая была получена вторым сервисом из источника<sup>5</sup> под названием «Sociate», являющимся биржей рекламы [13].

Следует отметить, что все собранные данные являются общедоступными, не являются конфиденциальными и не имеют персональных данных, они могут быть получены любым пользователем сети Интернет и принадлежат его открытому сегменту.

Разработка алгоритма предобработки статистических данных пользовательской активности в социальных сетях является ключевым этапом для обеспечения качественного анализа пользовательского поведения. Перейдем к описанию данного алгоритма.

### 1.2.2. Алгоритм предобработки статистических данных

#### **Описание структуры файлов со статистикой и пространства признаков**

В результате проведенного анализа сервисов, предоставляющих аналитическую информацию, была произведена выгрузка данных из сервисов «Popsters» и «AllSocial» за период с 01.06.2021-31.06.2022. В первом сервисе количество сообществ составило более чем 3600, а во втором сформирована выгрузка из 10000 сообществ. Каждая из выгрузок имеет свою структуру и свое признаковое описание сообщества. Разберем подробнее особенности каждой из них.

---

<sup>5</sup> Все представленные источники, в том числе их работоспособность, актуальны на 01.06.2022.



Для применения оптимизационного подхода или методологии машинного обучения необходимо, чтобы данные были представлены в виде матрицы объектов-признаков. Кроме того, статистические данные должны отражать активность пользователей социальной сети, как обратную реакцию на размещаемый в сообществе контент. Данные из первого набора преобразовывать не надо, а для данных из второго набора требуется разработать алгоритм для формирования матрицы объектов-признаков.

Признаки, представленные в выгрузке из сервиса «AllSocial», являются дополнительными, а основными выбраны характеристики из сервиса «Popsters». Для более подробного ознакомления с признаками из сервиса «Popsters» необходимо перейти по ссылке [87]. Используя данные в наборе 2, были введены новые признаки, такие как:

- Коэффициент вовлеченности (Engagement Rate) аудитории в день в течение месяца на количество подписчиков сообщества:

$$ER_{day\ month_j} = \frac{\sum_{i=1}^{n_j} (L_i + R_i + C_i)}{Subscribers_j \times n_j} \times 100,$$

где  $L_i, R_i, C_i$  - сумма «Мне нравится»/«Поделиться»/«Комментариев» со всех вышедших публикаций в  $i$ -ый день в  $j$ -ом месяце;  $i = \overline{1, n_j}$ ,  
 $n_j$  - количество дней в  $j$ -ом месяце,  $Subscribers_j$  - количество подписчиков сообщества в  $j$ -ом месяце.

- Коэффициент вовлеченности аудитории на публикацию в течение месяца на количество подписчиков сообщества:

$$ER_{post\ month_j} = \frac{\sum_{i=1}^{n_j} (L_i + R_i + C_i)}{Subscribers_j \times \sum_{i=1}^{n_j} Publication_i} \times 100,$$

где  $L_i, R_i, C_i$  - сумма «Мне нравится»/«Поделиться»/«Комментариев» со всех вышедших публикаций в  $i$ -ый день в  $j$ -ом месяце,  $i = \overline{1, n}$ ;  
 $n_j$  - количество дней в  $j$ -ом месяце,  $Subscribers_j$  - количество подписчиков сообщества в  $j$ -ом месяце;  $Publication_i$  - количество публикаций в  $i$ -ый

день.

- Коэффициент видимости (Visibility Rate) аудитории в день в течение месяца на количество подписчиков сообщества:

$$VR_{day\ month_j} = \frac{\sum_{i=1}^{n_j} V_i}{Subscribers_j \times n_j} \times 100,$$

где  $V_i$  - сумма «Просмотров» в  $i$ -ый день со всех вышедших публикаций в  $j$ -ом месяце;  $i = \overline{1, n_j}$ ,  $n_j$  - количество дней в  $j$ -ом месяце;  $Subscribers_j$  - количество подписчиков сообщества в  $j$ -ом месяце.

- Коэффициент видимости аудитории на публикацию в течение месяца на количество подписчиков сообщества:

$$VR_{post\ month_j} = \frac{\sum_{i=1}^{n_j} V_i}{Subscribers_j \times \sum_{i=1}^{n_j} Publication_i} \times 100,$$

где  $V_i$  - сумма «Просмотров» в  $i$ -ый день со всех вышедших публикаций в  $j$ -ом месяце,  $Subscribers_j$  - количество подписчиков сообщества в  $j$ -ом месяце;  $Publication_i$  - количество публикаций в  $i$ -ый день;  $i = \overline{1, n_j}$ ,  $n_j$  - количество дней в  $j$ -ом месяце.

- Коэффициент привлекательности (Love Rate) аудитории в день в течение месяца на количество подписчиков сообщества:

$$LR_{day\ month_j} = \frac{\sum_{i=1}^{n_j} L_i}{Subscribers_j \times n_j} \times 100,$$

где  $L_i$  - сумма «Мне нравится» в  $i$ -ый день со всех вышедших публикаций в  $j$ -ом месяце;  $i = \overline{1, n_j}$ ,  $n_j$  - количество дней в  $j$ -ом месяце;  $Subscribers_j$  - количество подписчиков сообщества в  $j$ -ом месяце.

- Коэффициент привлекательности аудитории на публикацию в течение ме-

сяца на количество подписчиков сообщества:

$$LR_{post\ month_j} = \frac{\sum_{i=1}^{n_j} L_i}{Subscribers_j \times \sum_{i=1}^{n_j} Publication_i} \times 100,$$

где  $L_i$  - сумма «Мне нравится» в  $i$ -ый день со всех вышедших публикаций в  $j$ -ом месяце,  $Subscribers_j$  - количество подписчиков сообщества в  $j$ -ом месяце;  $Publication_i$  - количество публикаций в  $i$ -ый день;  $i = \overline{1, n_j}$ ,  $n_j$  - количество дней в  $j$ -ом месяце.

- Коэффициент общительности (Talk Rate) аудитории в день в течение месяца на количество подписчиков сообщества:

$$TR_{day\ month_j} = \frac{\sum_{i=1}^{n_j} C_i}{Subscribers_j \times n_j} \times 100,$$

где  $C_i$  - сумма «Комментариев» в  $i$ -ый день со всех вышедших публикаций в  $j$ -ом месяце;  $i = \overline{1, n_j}$ ,  $n_j$  - количество дней в  $j$ -ом месяце;  $Subscribers_j$  - количество подписчиков сообщества в  $j$ -ом месяце.

- Коэффициент общительности аудитории на публикацию в течение месяца на количество подписчиков сообщества:

$$TR_{post\ month_j} = \frac{\sum_{i=1}^{n_j} C_i}{Subscribers_j \times \sum_{i=1}^{n_j} Publication_i} \times 100,$$

где  $C_i$  - сумма «Комментариев» в  $i$ -ый день со всех вышедших публикаций в  $j$ -ом месяце,  $Subscribers_j$  - количество подписчиков сообщества в  $j$ -ом месяце;  $Publication_i$  - количество публикаций в  $i$ -ый день;  $i = \overline{1, n_j}$ ,  $n_j$  - количество дней в  $j$ -ом месяце.

- Коэффициент распространения (Amplification Rate) среди аудитории в день в течение месяца на количество подписчиков сообщества:

$$AR_{day\ month_j} = \frac{\sum_{i=1}^{n_j} R_i}{Subscribers_j \times n_j} \times 100,$$

где  $R_i$  - сумма «Поделиться» в  $i$ -ый день со всех вышедших публикаций в  $j$ -ом месяце;  $i = \overline{1, n_j}$ ,  $n_j$  - количество дней в  $j$ -ом месяце;  $Subscribers_j$  - количество подписчиков сообщества в  $j$ -ом месяце.

- Коэффициент распространения среди аудитории на публикацию в течение месяца на количество подписчиков сообщества:

$$AR_{post\ month_j} = \frac{\sum_{i=1}^{n_j} R_i}{Subscribers_j \times \sum_{i=1}^{n_j} Publication_i} \times 100,$$

где  $R_i$  - сумма «Поделиться» в  $i$ -ый день со всех вышедших публикаций в  $j$ -ом месяце,  $Subscribers_j$  - количество подписчиков сообщества в  $j$ -ом месяце;  $Publication_i$  - количество публикаций в  $i$ -ый день;  $i = \overline{1, n_j}$ ,  $n_j$  - количество дней в  $j$ -ом месяце.

- Всего публикаций в течение месяца:

$$Publication_j = \sum_{i=1}^{n_j} Publication_i,$$

где  $Publication_j$  - количество публикаций в  $j$ -ом месяце,  $j = \overline{1, m}$ ;  $m$  - количество месяцев,  $i = \overline{1, n_j}$ ,  $n_j$  - количество дней в  $j$ -ом месяце.

- Интенсивность публикационной активности в день в течение месяца (далее «ИПА в день (за год)»):

$$IPA_{month_j} = \frac{\sum_{i=1}^{n_j} Publication_i}{n_j},$$

где  $\sum_{i=1}^{n_j} Publication_i$  - количество публикаций в  $j$ -ом месяце,  $j = \overline{1, m}$ ;  $m$  - количество месяцев,  $i = \overline{1, n_j}$ ,  $n_j$  - количество дней в  $j$ -ом месяце.

- Среднее значение отметок («Мне нравится», «Поделиться», «Комментариев», «Просмотров») публикации в течение месяца:

$$L_{post\ month_j} = \frac{\sum_{i=1}^{n_j} L_i}{\sum_{i=1}^{n_j} Publication\_day_i},$$

$$R_{post\ month_j} = \frac{\sum_{i=1}^{n_j} R_i}{\sum_{i=1}^{n_j} Publication\_day_i},$$

$$C_{post\ month_j} = \frac{\sum_{i=1}^{n_j} C_i}{\sum_{i=1}^{n_j} Publication\_day_i},$$

$$V_{post\ month_j} = \frac{\sum_{i=1}^{n_j} V_i}{\sum_{i=1}^{n_j} Publication\_day_i},$$

где  $\sum_{i=1}^{n_j} L_i, \sum_{i=1}^{n_j} R_i, \sum_{i=1}^{n_j} C_i, \sum_{i=1}^{n_j} V_i$  - сумма отметок «Мне нравится», «Поделиться», «Комментариев», «Просмотров» под всеми публикациями за  $j$ -ый месяц;  $\sum_{i=1}^{n_j} Publication\_day_i$  - количество публикаций в  $i$ -ый день,  $j = \overline{1, m}$ ;  $m$  - количество месяцев,  $i = \overline{1, n_j}$ ,  $n_j$  - количество дней в  $j$ -ом месяце.

- Среднее значение отметок («Мне нравится», «Поделиться», «Комментариев», «Просмотров») публикации в течение года:

$$L_{post\ year} = \frac{\sum_{j=1}^m L_j}{\sum_{j=1}^m Publication\_month_j},$$

$$R_{post\ year} = \frac{\sum_{j=1}^m R_j}{\sum_{j=1}^m Publication\_month_j},$$

$$C_{post\ year} = \frac{\sum_{j=1}^m C_j}{\sum_{j=1}^m Publication\_month_j},$$

$$V_{post\ year} = \frac{\sum_{j=1}^m V_j}{\sum_{j=1}^m Publication\_month_j},$$

где  $\sum_{j=1}^m L_j, \sum_{j=1}^m R_j, \sum_{j=1}^m C_j, \sum_{j=1}^m V_j$  - сумма отметок «Мне нравится», «Поделиться», «Комментариев», «Просмотров» под всеми публикациями за  $j$ -ый месяц;  $Publication\_month_j$  - количество публикаций в месяц  $j$ ,  $j = \overline{1, m}$ ,  $m$  - количество месяцев.

Используя данные в наборе 1, были введены новые признаки, такие как:

- Средний возраст целевой аудитории сообщества (далее «Средний возраст

ЦА»):

$$Age_{mean} = \frac{\sum_{i=1}^n Age\_interval\_mean_i \times Percent\_value\_interval_i}{100},$$

где  $Age\_interval\_mean_i$  - средний возраст в  $i$ -ом интервале, например, для «% 27-30 лет» средний возраст будет составлять 28,5 лет (крайние значения: «% до 18 лет» и «% от 45 лет», были взяты следующие: 16,5 и 50 лет соответственно);  $Percent\_value\_interval_i$  - какой процент от общего количества аудитории принадлежит заданному возрастному интервалу.

- Соотношение мужчин и женщин в сообществе (далее «Пол (М/Ж)»):

$$Sex (male/female) = \frac{Percent\_male}{Percent\_female},$$

где  $Percent\_male / Percent\_female$  - доля аудитории мужчин и женщин соответственно.

Часть из введенных выше признаков будет использоваться для реализации алгоритмов кластеризации, часть для формирования рекомендаций по размещению контента и для комплексного отображения активности внутри сообщества.

Кроме того, был введен признак - «Тематика», который заполнялся вручную, так как в выбранных сервисах такая характеристика отсутствует.

В итоге, из сервиса «AllSocial» имеем следующие характеристики сообществ: «Сообщества», «id», «URL», «Прирост за 1 день», «Прирост за 7 дней», «Прирост за 30 дней», «% Offline более месяца», «ER (100 последних постов)», «Цена Sociate», «CPM Sociate», «Посетители (среднее за 7 дней)», «% мобильных», «% компьютеров», «% Муж.», «% Жен.», «% до 18 лет», «% 18-21 лет», «% 21-24 лет», «% 24-27 лет», «% 27-30 лет», «% 30-35 лет», «% 35-45 лет», «% от 45 лет», « $Age_{mean}$ », « $Sex (male/female)$ », «% Россия», «% Белоруссия», «% Другие страны».

Из сервиса «Popsters» получаем следующие характеристики: «Дни недели», «Время суток», «Дни недели/объем текста», «Время суток/объем текста»,

«Относительная активность дни недели/объем текста», «Относительная активность время суток/объем текста», «Мне нравится», «Поделиться», «Комментарии», «Просмотры», «Подписчики», «Количество публикаций», «Количество по длине текста», «ER по длине текста», «Количество по типу контента», «ER по типу контента», «Относительная активность по типу контента», « $ER_{day\ month_j}$ », « $ER_{post\ month_j}$ », « $VR_{day\ month_j}$ », « $VR_{post\ month_j}$ », « $LR_{day\ month_j}$ », « $LR_{post\ month_j}$ », « $TR_{day\ month_j}$ », « $TR_{post\ month_j}$ », « $AR_{day\ month_j}$ », « $AR_{post\ month_j}$ », « $Publication_j$ », « $IPA_{month_j}$ », « $L_{post\ month_j}$ », « $R_{post\ month_j}$ », « $C_{post\ month_j}$ », « $V_{post\ month_j}$ », « $L_{post\ year}$ », « $R_{post\ year}$ », « $C_{post\ year}$ », « $V_{post\ year}$ ».

Для проведения моделирования и реализации методов машинного обучения требуется построение матрицы объектов-признаков. В соответствие с данным требованием был разработан алгоритм предобработки данных. Рассмотрим его более подробно.

### Алгоритм формирования матрицы объектов-признаков

Необходимо набор данных 2 преобразовать в матрицу объектов-признаков, где объектом является сообщество, а признаками его характеристики. Для реализации требуемого преобразования был разработан алгоритм (см. рис. 1.4), состоящий из следующих частей:

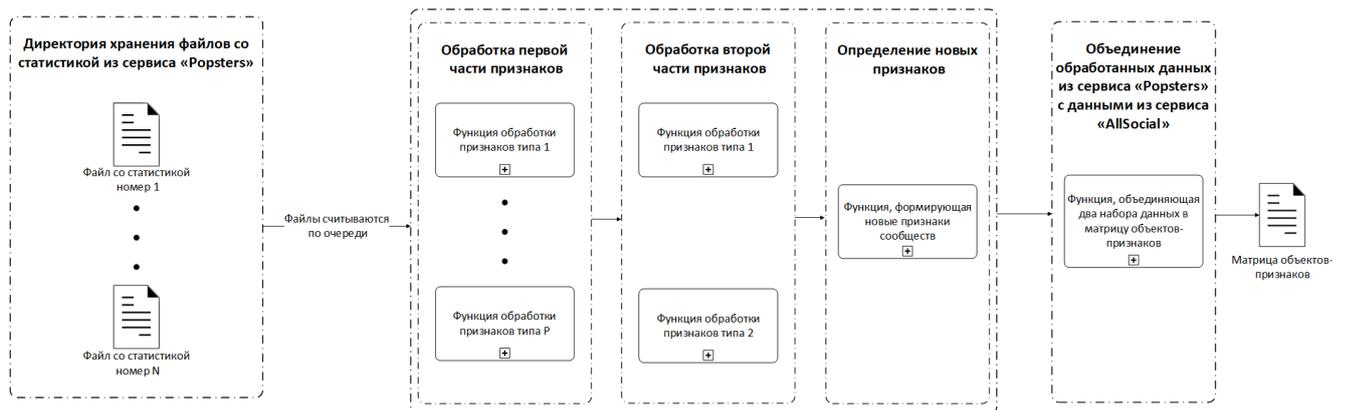


Рис. 1.4: Схема алгоритма предобработки данных

1. Реализован цикл, в котором считываются по очереди файлы и начинается их предобработка по принципу – один файл разбивается на таблицы, количество которых равно количеству признаков. Следует отметить, что признаки объединены по типам, для которых функции обработки бу-

дут одинаковыми.

1.1 Вложенный цикл, удаляющий пустые строки и ставящий в соответствие названиям признаков таблицы;

1.2 Далее реализованы функции обработки всех типов признаков и функции проверки одинаковых названий сообществ в одном файле, а также в тех случаях, когда одно и то же сообщество есть в разных файлах:

- Если такие существуют, то в название сообществ добавляется их порядковый номер строки, например: «Моя дача», «Моя дача\_6», «Моя дача\_7». Это необходимо для корректной работы метода `merge` и `concat` библиотеки `pandas`;
- Если одно и то же сообщество есть в разных файлах, то для такого случая реализована функция проверки и удаления таких повторов, причем удаляются все кроме первого.

1.3 Таким образом, сформированы матрицы по количеству признаков, в каждой из которых хранятся данные по всем сообществам.

2. Реализована функция, которая определяет и вычисляет новые признаки по формулам, описанным ранее, из уже имеющихся данных.
3. Реализованы функции для обработки двух типов таблиц, содержащих признаки объектов, и формирования матрицы объектов-признаков.
4. Далее происходит объединение преобразованных данных с набором данных 1 в единую матрицу объектов-признаков.

Как итог работы алгоритма, создается матрица объектов-признаков, в которой 3604 сообщества, более 20 целевых признаков и более 30 характеристик для формирования рекомендаций по размещению записей в каждом сообществе (см. рис. 1.5). Программный компонент, реализующий данный алгоритм программно реализован и зарегистрирован в ФИПС [79].

1	0	1	2	3	4	5	6	7	8	9	10	11	12	13		
2	0	Подписчиков	Всего публикаций	Мне нравится	Поделился	Комментариев	Просмотров	ER Day_year	ER Post_year	ER View_year	LR Post_year	TR Post_year	Период	VR Post_year		
3	1	MDK	11822715	14585	37259544	27224948	1565814	4800829108	1.526	0.038	1.296	0.022	0.001	01.06.2021 - 01.06.2022	2,784	
4	2	4ch	5071206	2253	1336557	3486182	305833	763120438	0.923	0.150	2.277	0.117	0.003	01.06.2021 - 01.06.2022	6,679	
5	3	NR.Music	5291361	8814	9844436	4496675	655541	1569389740	0.774	0.032	0.869	0.021	0.001	01.06.2021 - 01.06.2022	3,365	
6	4	БОРЦ	7057083	9608	22405135	15947605	919675	2185478037	1.520	0.058	1.768	0.033	0.001	01.06.2021 - 01.06.2022	3,223	
7	5	Академия Порядочных Парней	5393647	5216	4622969	2922419	611221	1200581400	0.413	0.029	0.689	0.016	0.002	01.06.2021 - 01.06.2022	4,267	
8	6	Science   Наука	5230844	5557	8935943	6435616	233322	2221480677	0.925	0.061	0.737	0.031	0.008	01.06.2021 - 01.06.2022	7,642	
9	7	ПОЗОР	4557846	13979	43874125	4751586	88569	2396305273	3.686	0.076	2.100	0.069	0.000	01.06.2021 - 01.06.2022	3,761	
10	8	Лепрозорий	3481261	1775	4228431	4031156	1047800	882556824	0.730	0.151	1.118	0.068	0.017	01.06.2021 - 01.06.2022	14,283	
11	9	MARVEL/DC	3703678	6352	40593633	4880416	919600	1566532984	3.423	0.197	2.801	0.173	0.004	01.06.2021 - 01.06.2022	6,659	
12	10	Идеи дизайна интерьера	5831257	3389	2630301	1324220	166410	618186253	0.193	0.021	0.644	0.013	0.001	01.06.2021 - 01.06.2022	3,128	
13	11	Книги	2159267	8980	5804354	1773459	316706	435153324	0.605	0.099	0.041	1.341	0.030	0.002	01.06.2021 - 01.06.2022	2,75
14	12	Цитаты и статусы	3585503	4400	666863	230101	34196	162473930	0.071	0.006	0.492	0.004	0.000	01.06.2021 - 01.06.2022	1,03	
15	13	В приколе (18+)	5258065	12095	3414599	2540972	353613	669237223	0.328	0.010	0.976	0.005	0.001	01.06.2021 - 01.06.2022	1,052	
16	14	• Неприличные Анекдоты	3688681	2677	1471915	911649	35834	186808899	0.179	0.025	1.154	0.015	0.000	01.06.2021 - 01.06.2022	1,892	
17	15	Экспериментатор   Наука	1641143	8471	2638863	726776	270150	435153324	0.605	0.026	0.883	0.019	0.002	01.06.2021 - 01.06.2022	3,13	
18	16	Кулинарное искусство	3115565	1829	1511781	779355	25462	136355640	0.203	0.041	2.019	0.027	0.000	01.06.2021 - 01.06.2022	2,393	
19	17	Простые рецепты	3251767	2385	1454300	1255197	19068	193581958	0.229	0.035	1.235	0.019	0.000	01.06.2021 - 01.06.2022	2,496	
20	18	КАЕФ	1599307	9554	2309271	1166501	32319	233995670	0.599	0.023	1.420	0.015	0.000	01.06.2021 - 01.06.2022	1,531	
21	19	Боги Смеха	1507737	1502	3595186	254900	50695	66028494	0.707	0.172	5.860	0.159	0.002	01.06.2021 - 01.06.2022	2,916	
22	20	PINK PARADISE	445362	3282	1143357	1036158	11314	72786601	1.344	0.150	3.005	0.078	0.001	01.06.2021 - 01.06.2022	4,98	
23	21	Дерева РУКОДЕЛКИНО рукоделие, вязание, дизайн	588480	3440	352518	239732	5768	52215744	0.278	0.030	1.020	0.017	0.000	01.06.2021 - 01.06.2022	2,579	
24	22	Сама Себе Тренер	473304	4076	685022	589163	9599	62371222	0.741	0.067	1.971	0.036	0.000	01.06.2021 - 01.06.2022	3,233	
25	23	Пилер Live	303433	5417	722949	161466	40833	61124528	0.833	0.056	1.400	0.044	0.002	01.06.2021 - 01.06.2022	3,719	
26	24	Знать и познавать	421557	996	821587	329895	24884	42479634	0.762	0.280	2.471	0.196	0.006	01.06.2021 - 01.06.2022	10,117	
27	25	Queen	644965	6173	2376196	557917	9868	115287131	1.247	0.074	2.083	0.060	0.000	01.06.2021 - 01.06.2022	2,896	
28	26	Новинки Музыки 2022   Свежая Музыка	345278	657	71387	27570	13755	8604158	0.089	0.050	1.226	0.031	0.006	01.06.2021 - 01.06.2022	3,793	

Рис. 1.5: Матрица объектов-признаков

В данном пункте были рассмотрены наиболее популярные сервисы по предоставлению статистических данных из социальной сети «ВКонтакте», определены характеристики сообществ, которые будут использоваться в задачах оптимизации и кластеризации, разработан алгоритм, формирующий матрицу объектов-признаков.

### Алгоритм обработки статистических данных

Алгоритм обработки данных, представляет собой ряд преобразований с матрицей объектов-признаков, таких как:

- Удаление объектов у которых пусто в столбце «Цена Sociate». Это объясняется необходимостью использования для моделирования объектов, имеющих известную стоимость размещения рекламной записи;
- Удаление объектов у которых в столбце «Всего публикаций» значение равно 0. Это объясняется необходимостью использования для моделирования множества объектов с ненулевой активностью в течение временного периода: 01.06.2021 - 31.05.2022;
- Преобразование категориальных признаков со строковыми значениями, например, «Тематика», «Название сообщества», «URL», в числовой тип данных.

Кроме того, в этом блоке используются данные из технического задания клиента, такие как: «Тематика», «Целевой показатель», «Месяц». Исходя из этого

формируется: подмножество объектов, заданных тематик, с ненулевыми значениями целевых показателей в заданный клиентом месяц. Если месяцев несколько, то для каждого месяца требуется проводить моделирование в отдельности, поскольку подмножества объектов с ненулевыми значениями целевого показателя, в общем случае, могут отличаться от месяца к месяцу. Данные преобразования объясняются необходимостью выделения определенного подмножества объектов под конкретное техническое задание клиента с целью проведения численного моделирования и формирования решения.

На выходе получается два набора данных, один предназначен для применения методов оптимизации и машинного обучения, другой для формирования рекомендаций по размещению рекламной записи для каждого объекта. В первом случае набор включает признаки: «Подписчиков», «ER Post\_year», «ER View\_year», «LR Post\_year», «TR Post\_year», «VR Post\_year», «AR Post\_year», «Лайки поста среднее (за год)», «Репосты поста среднее (за год)», «Комментарии поста среднее (за год)», «Просмотры поста среднее (за год)», «Посетители (среднее за 7 дней)», «ИПА в день (за год)», «% Offline более месяца», «Цена Sociate», «CPM Sociate», «Пол (М/Ж)», «Тематика\_id», «ER (100 последних постов)», «Средний возраст ЦА», «Прирост за 30 дней». Отметим, что под «Лайками» и «Репостами» понимаются соответственно отметки «Мне нравится» и «Поделиться».

### 1.2.3. Особенности реализации оптимизационной модели

#### Реализация рекомендательного блока

Поддержка принятия управленческих решений для ЛПР в задаче распространения информации в СМК осуществляется путем проведения анализа поведенческой активности аудитории площадок. Следовательно, помимо характеристик введенных в разделе 1.2.2., необходимо разработать функции для составления рекомендаций по размещению записей в сообществах социальной сети.

Для реализации блока формирования рекомендаций написано более 10 функций (см. рис. 1.6).

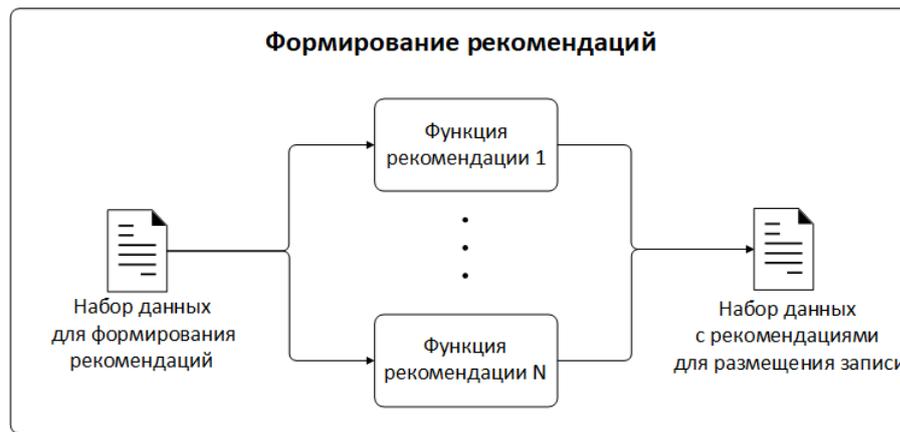


Рис. 1.6: Схема реализации разработанной библиотеки для формирования рекомендаций по размещению публикации

Благодаря формированию рекомендаций имеется возможность определять наилучший временной период для публикации записи, объем текста в ней, необходимость прикрепления к ней фото/ауди/видео материалов, а также ссылок, таким образом, чтобы отклик аудитории на размещаемую запись был максимальным в данном сообществе, то есть сделать так, чтобы опубликованная запись набрала больше отметок «Мне нравится», «Поделиться», «Комментариев», «Просмотров» (см. рис. 1.7).

```
def best_dayweek_func(data: pd.DataFrame) -> pd.DataFrame:
    """..."""
    data_copied = data.copy()
    list = ['Дни недели Пн', 'Дни недели Вт', 'Дни недели Ср', 'Дни недели Чт', 'Дни недели Пт', 'Дни недели Сб',
            'Дни недели Вс']
    data_copied = data_copied[list]
    best_res = pd.DataFrame(
        [data_copied.columns[i].tolist() for i in (data_copied.values == data_copied.max(axis=1)[:. None])])
    best_res.rename(columns={0: 'День недели для публикации', 1: 'Значение (День недели)'}, inplace=True)
    best_res['Значение (День недели)'] = data_copied.max(axis=1)
    best_res = best_res.replace(list,
                                ['Понедельник', 'Вторник', 'Среда', 'Четверг', 'Пятница', 'Суббота', 'Воскресенье'])
    return best_res
```

Рис. 1.7: Пример рекомендательной функций

Кроме основных рекомендаций, клиенту предоставляются все имеющиеся метрики, отражающие активность в сообществе, например, выводится не «Средний возраст ЦА», а процентное соотношение количества подписчиков по возрастным интервалам.

## Реализация оптимизационной модели

Для программной реализации оптимизационной модели необходимо сформулировать математическую постановку задачи, сформировать матрицу объектов-признаков и функции рекомендаций для каждого объекта-сообщества по размещению записи. Решаться поставленные задачи будут с помощью известных методов линейного программирования, методов многокритериальной оптимизации, а программно реализованы алгоритмы решения задач будут на языке программирования «Python» с соответствующими библиотеками в кроссплатформенной интегрированной среде разработки PyCharm. В силу оптимизации времени работы алгоритмов, решающих поставленные задачи, было принято решение реализовать одну функцию, в которой решается как задача многокритериальной оптимизации, так и задача целочисленного линейного программирования.

Перейдем к описанию архитектуры функции, реализовывающей предлагаемый оптимизационный подход к решению задачи распространения информации в СМК:

1. Задаются: списки тематик, месяц года, бюджет клиента -  $P$ , список критериев: «Мне нравится», «Поделиться», «Комментариев», «Просмотров» -  $f_1, f_2, f_3, f_4$ , матрица весовых коэффициентов в виде списка со вложенными списками -  $A$ , таблица для фиксации результатов моделирования. Следует отметить, что бюджет клиента не превосходит максимальный бюджет для заданных тематик.
2. Задаются 3 вложенных цикла: первый по списку тематик, второй по бюджету клиента, третий по срокам матрицы весовых коэффициентов.
3. Вызывается функция обработки данных, которая отбирает объекты из заданного списка тематик, с ненулевыми значениями критериев в заданный месяц года. Таким образом, формируется подмножество объектов на котором будет производиться моделирование.
4. Вызывается функция, реализующая методы оптимизации, где формиру-

ются рекомендации для объектов заданного подмножества, отбираются объекты, стоимость размещения рекламной записи в которых меньше либо равна  $P$ .

5. Задается: массив цен на размещение рекламной записи -  $g_j$ , где  $j = \overline{1, m}$ ; массивы значений для критериев  $f_1, f_2, f_3, f_4$ .
6. Методом свертки критериев задается один критерий:

$$\begin{aligned} \hat{f}(x) &= \alpha_1 \cdot f_1(x) + \alpha_2 \cdot f_2(x) + \alpha_3 \cdot f_3(x) + \alpha_4 \cdot f_4(x) = \\ &= \alpha_1 \cdot \sum_{j=1}^m a_j \cdot x_j + \alpha_2 \cdot \sum_{j=1}^m b_j \cdot x_j + \alpha_3 \cdot \sum_{j=1}^m c_j \cdot x_j + \alpha_4 \cdot \sum_{j=1}^m d_j \cdot x_j = \\ &= \sum_{j=1}^m (\alpha_1 \cdot a_j + \alpha_2 \cdot b_j + \alpha_3 \cdot c_j + \alpha_4 \cdot d_j) \cdot x_j = \sum_{j=1}^m w_j \cdot x_j \end{aligned}$$

7. Далее с помощью библиотеки «scipy.optimize» [12] решается задача оптимизации и определяется результирующий вектор  $x$ .
8. Вычисляются значения соответствующих критериев для оптимальных наборов сообществ при заданных коэффициентах значимости.
9. Фиксируются результаты моделирования в сравнительные таблицы (пример см. рис. 1.8).

	alpha_0	alpha_1	alpha_2	alpha_3	Всего Лайки поста	Всего Репосты поста	Всего Комментарии поста	Всего Просмотры поста	F_value	Obj_fun	Бюджет клиента	Цена стоимости	Количество сообществ	Выбранное тематике	Время работы	Итого
1	1.00	0.00	0.00	0.00	4346	469	140	134267	139222	-4346.00	1000	995	8	Образование	0 days 00:00:00.003001	Иль
2	0.00	1.00	0.00	0.00	2332	1455	45	144614	148446	-1455.00	1000	974	6	Образование	0 days 00:00:00.003001	Иль
3	0.00	0.00	1.00	0.00	3782	453	168	131835	136238	-168.00	1000	995	7	Образование	0 days 00:00:00.003001	Иль
4	0.00	0.00	0.00	1.00	3486	664	110	168863	173123	-168863.00	1000	989	8	Образование	0 days 00:00:00.015004	Иль
5	0.25	0.25	0.25	0.25	3486	664	110	168863	173123	-43208.75	1000	989	8	Образование	0 days 00:00:00.011002	Иль
6	0.57	0.01	0.01	0.01	4249	556	116	148254	153275	-5611.79	1000	995	8	Образование	0 days 00:00:00.007002	Иль
7	0.01	0.97	0.01	0.01	2763	1320	62	162099	166244	-2929.64	1000	988	7	Образование	0 days 00:00:00.001999	Иль
8	0.01	0.01	0.97	0.01	3523	618	121	167930	172192	-1838.08	1000	997	9	Образование	0 days 00:00:00.012002	Иль
9	0.01	0.01	0.01	0.97	3486	664	110	168863	173123	-163839.71	1000	989	8	Образование	0 days 00:00:00.013003	Иль
10	0.94	0.02	0.02	0.02	3960	646	112	167236	171954	-7882.28	1000	974	9	Образование	0 days 00:00:00.005002	Иль
11	0.02	0.94	0.02	0.02	2763	1320	62	162099	166244	-4539.28	1000	988	7	Образование	0 days 00:00:00.002001	Иль
12	0.02	0.02	0.94	0.02	3486	664	110	168863	173123	-3563.66	1000	989	8	Образование	0 days 00:00:00.011506	Иль
13	0.02	0.02	0.02	0.94	3486	664	110	168863	173123	-158826.42	1000	989	8	Образование	0 days 00:00:00.015004	Иль
14	0.91	0.03	0.03	0.03	3960	646	112	167236	171954	-8643.42	1000	974	9	Образование	0 days 00:00:00.000999	Иль
15	0.03	0.91	0.03	0.03	2763	1320	62	162099	166244	-6148.92	1000	988	7	Образование	0 days 00:00:00.003001	Иль
16	0.03	0.03	0.91	0.03	3486	664	110	168863	173123	-5208.49	1000	989	8	Образование	0 days 00:00:00.014003	Иль
17	0.03	0.03	0.03	0.91	3486	664	110	168863	173123	-153793.13	1000	989	8	Образование	0 days 00:00:00.014003	Иль
18	0.04	0.04	0.04	0.04	3960	646	112	167236	171954	-18204.56	1000	974	9	Образование	0 days 00:00:00.000999	Иль
19	0.04	0.04	0.04	0.04	2763	1320	62	162099	166244	-7758.56	1000	988	7	Образование	0 days 00:00:00.007001	Иль
20	0.04	0.04	0.04	0.04	3486	664	110	168863	173123	-7817.32	1000	989	8	Образование	0 days 00:00:00.012002	Иль

Рис. 1.8: Пример сравнительной таблицы результатов моделирования

### 1.3. Численное моделирование и анализ результатов

Для демонстрации работы предлагаемого подхода и возможности проведения сравнительного анализа полученных результатов зададим следующие входные параметры:

1. Тематики - «Автомобили, Автовладельцы», «Кулинария, рецепты», «Образование». Количество объектов в каждой тематике: 124, 126, 103;
2. Временные интервалы - ежемесячно, с января по декабрь;
3. Бюджет - от 1000 рублей до максимально возможного бюджета в рамках заданных тематик и месяцев года с шагом в 5000 рублей;
4. Предпочтения клиента заданы в векторной форме  $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$  и представлены в таблице 1.1:

Таблица 1.1: Весовые коэффициенты для проведения анализа на чувствительность при изменении предпочтений клиента

Часть 1				Часть 2				Часть 3			
$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$
1.0	0.0	0.0	0.0	0.05	0.85	0.05	0.05	0.1	0.1	0.1	0.7
0.0	1.0	0.0	0.0	0.05	0.05	0.85	0.05	0.67	0.11	0.11	0.11
0.0	0.0	1.0	0.0	0.05	0.05	0.05	0.85	0.11	0.67	0.11	0.11
0.0	0.0	0.0	1.0	0.82	0.06	0.06	0.06	0.11	0.11	0.67	0.11
0.25	0.25	0.25	0.25	0.06	0.82	0.06	0.06	0.11	0.11	0.11	0.67
0.97	0.01	0.01	0.01	0.06	0.06	0.82	0.06	0.64	0.12	0.12	0.12
0.01	0.97	0.01	0.01	0.06	0.82	0.06	0.82	0.12	0.64	0.12	0.12
0.01	0.01	0.97	0.01	0.79	0.07	0.07	0.07	0.12	0.12	0.64	0.12
0.01	0.01	0.01	0.97	0.07	0.79	0.07	0.07	0.12	0.12	0.12	0.64
0.94	0.02	0.02	0.02	0.07	0.07	0.79	0.07	0.61	0.13	0.13	0.13
0.02	0.94	0.02	0.02	0.07	0.07	0.07	0.79	0.13	0.61	0.13	0.13

Продолжение таблицы 1.1

0.02	0.02	0.94	0.02	0.76	0.08	0.08	0.08	0.13	0.13	0.61	0.13
0.02	0.02	0.02	0.94	0.08	0.76	0.08	0.08	0.13	0.13	0.13	0.61
0.91	0.03	0.03	0.03	0.08	0.08	0.76	0.08	0.58	0.14	0.14	0.14
0.03	0.91	0.03	0.03	0.08	0.08	0.08	0.76	0.14	0.58	0.14	0.14
0.03	0.03	0.91	0.03	0.73	0.09	0.09	0.09	0.14	0.14	0.58	0.14
0.03	0.03	0.03	0.91	0.09	0.73	0.09	0.09	0.14	0.14	0.14	0.58
0.88	0.04	0.04	0.04	0.09	0.09	0.73	0.09	0.55	0.15	0.15	0.15
0.04	0.88	0.04	0.04	0.09	0.09	0.09	0.73	0.15	0.55	0.15	0.15
0.04	0.04	0.88	0.04	0.7	0.1	0.1	0.1	0.15	0.15	0.55	0.15
0.04	0.04	0.04	0.88	0.1	0.7	0.1	0.1	0.15	0.15	0.15	0.55
0.85	0.05	0.05	0.05	0.1	0.1	0.7	0.1	–	–	–	–

Такое количество вариантов векторов весовых коэффициентов обусловлено тем, что будет проведен анализ чувствительности заданной системы с целью формирования рекомендаций для лиц принимающих управленческие решения в зависимости от бюджета, сезонности и номенклатуры товаров или услуг. Заметим, что для оценки влияния изменения значений весов критериев на конечный результат используется принцип выделения главного критерия и дальнейшего равномерного распределения остатка между другими. Перейдем к анализу результатов моделирования.

**Наблюдение 1.1.** *Количество уникальных решений не превышает 30 процентов вне зависимости от изменений значений весовых коэффициентов и бюджета по указанным тематикам и месяцам года (см. рис. 1.9, 1.10, 1.11, 1.12).* В таблице на рисунке 1.9 представлены максимальные значения процента уникальных решений среди всех возможных бюджетов.

Тематики	Январь	Февраль	Март	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
Кулинария, рецепты	25	26,7	28,3	26,7	21,7	25	26,7	28,3	30	26,7	26,7	28,3
Автомобили, Автовладельцы	16,7	16,7	20	25	18,3	21,7	21,7	15	15	16,7	16,7	16,7
Образование	23,3	23,3	23,3	23,3	26,7	18,3	25	23,3	30	25	23,3	23,3

Рис. 1.9: Максимальный процент уникальных решений среди всех бюджетов

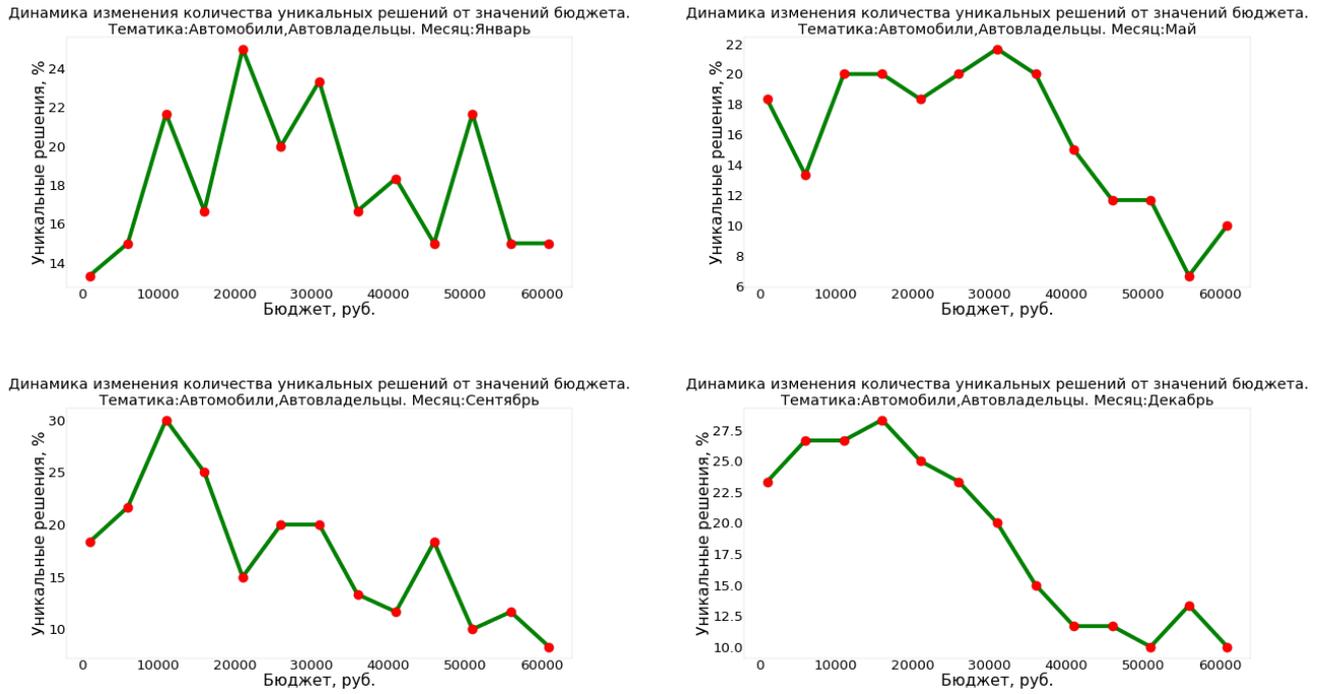


Рис. 1.10: Динамика изменения процента уникальных решений в зависимости от бюджета для тематики: «Автомобили, Автовладельцы»

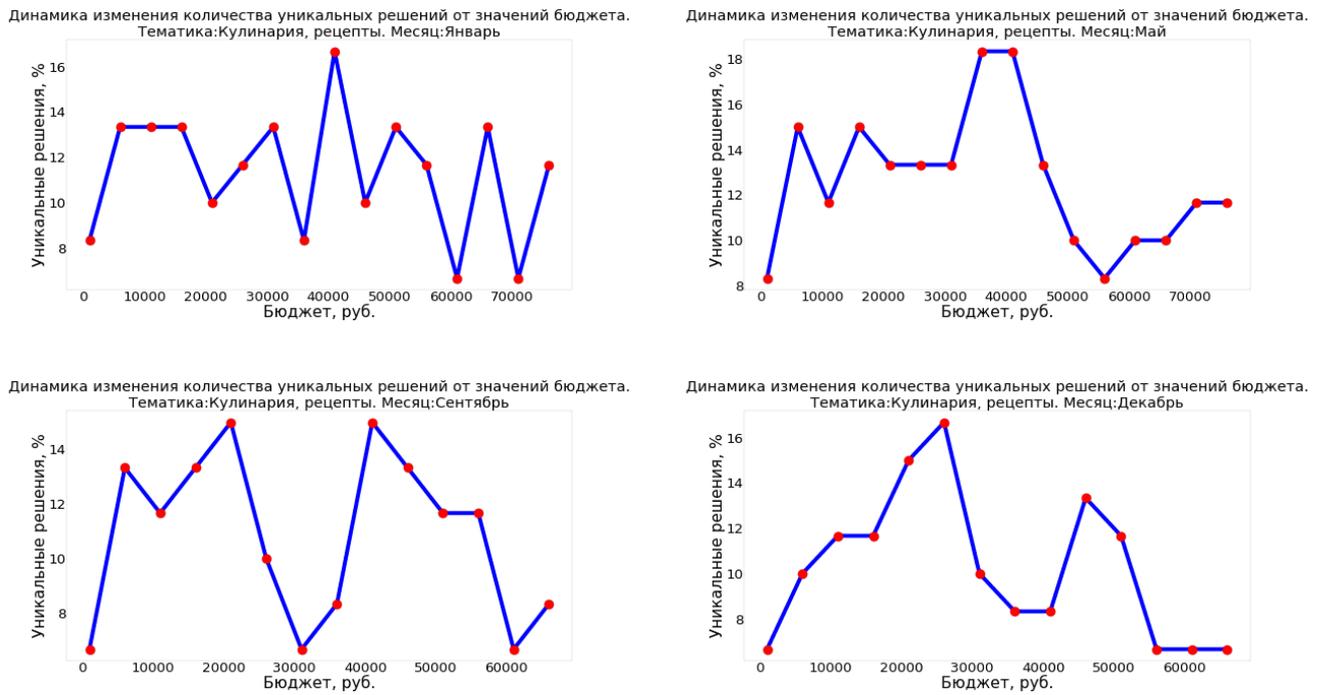


Рис. 1.11: Динамика изменения процента уникальных решений в зависимости от бюджета для тематики: «Кулинария, рецепты»

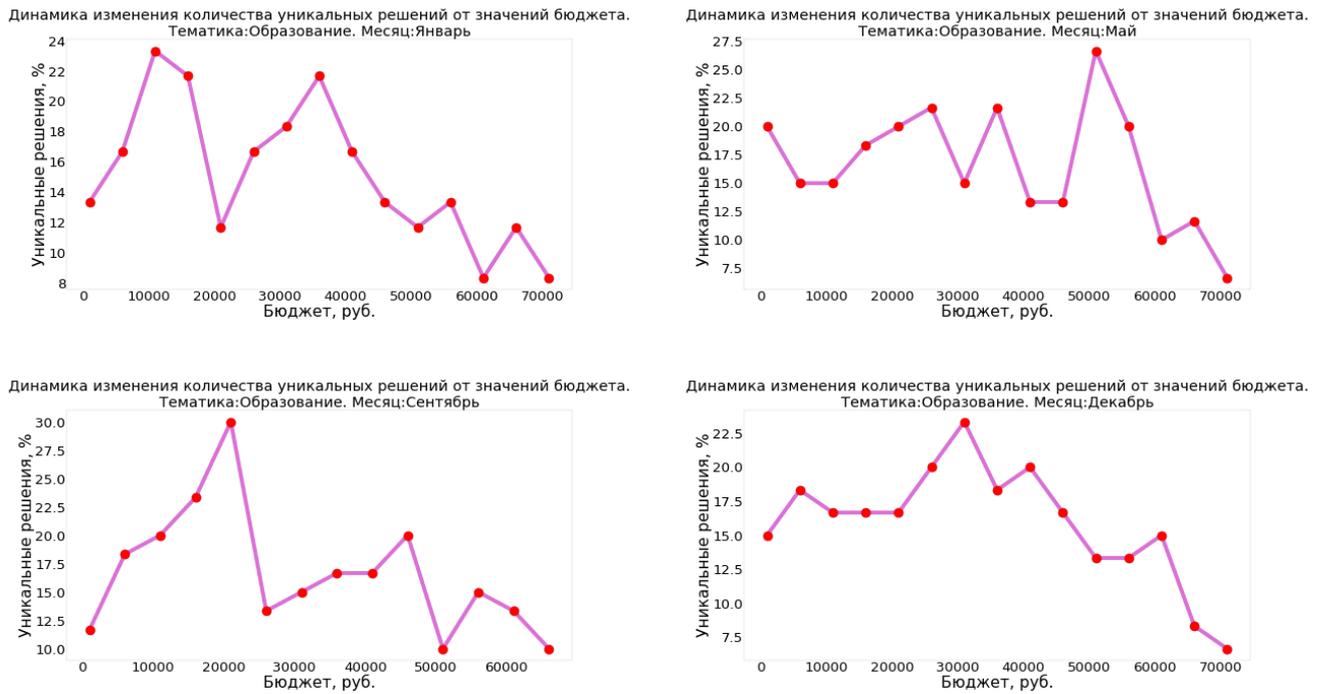


Рис. 1.12: Динамика изменения процента уникальных решений в зависимости от бюджета для тематики: «Образование»

**Вывод 1.1.** Специфика поведенческой активности пользователей социальной сети такова, что изменение предпочтений на треть влияет на процесс формирования уникального решения.

**Замечание 1.1.** Динамика изменения процента уникальных решений от изменений значений весовых коэффициентов по тематикам и месяцам года показывает насколько активно и отлично друг от друга ведет себя аудитория в различных группах одной и той же тематики, вне зависимости от выбранных критериев.

Действительно, на представленном графике (см. рис. 1.13) можно видеть, что в различные временные интервалы в результате повышения интереса пользователей сети к тем или иным сферам их жизнедеятельности, в силу сезонности спроса на товары (или услуги) и проявлении других внешних факторов, увеличивается вероятность получения уникального набора площадок. Данное замечание позволяет сделать следующий вывод для обоснования рекомендаций по формированию предпочтений.

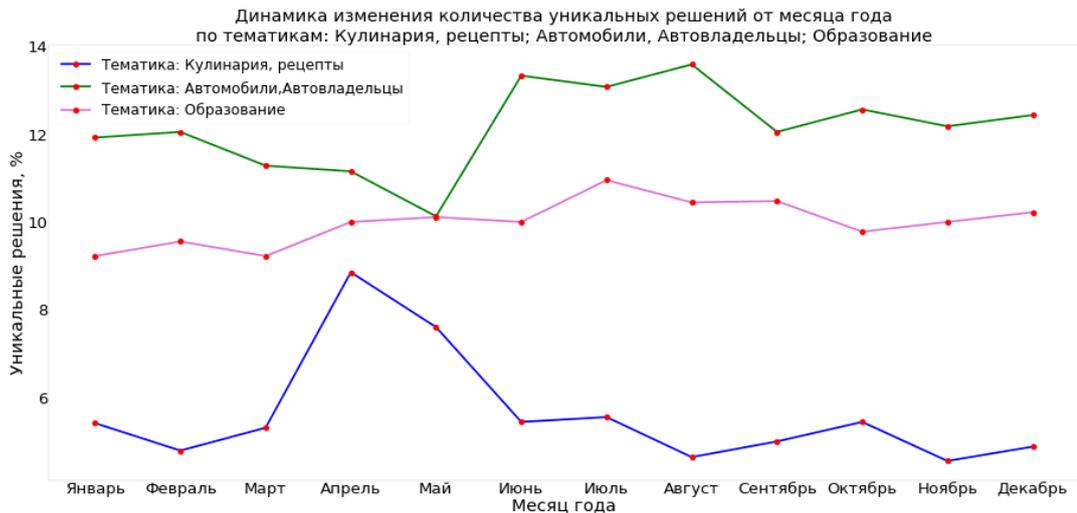


Рис. 1.13: Динамика изменения количества уникальных решений от месяца года по тематикам

**Вывод 1.2.** Чем разнообразнее проявляет себя аудитория сообществ, тем выше процент получения уникальных результатов при небольших изменениях предпочтений.

**Замечание 1.2.** На основе данных по 12 месяцам, 3 тематикам и установленным значениям бюджета заметим, что прослеживается наличие чувствительности критериев.

Данное замечание указывает на то, что имеется возможность корректировать предпочтения клиента в зависимости от задаваемых им входных параметров, а также анализировать степень активности аудитории площадок.

**Вывод 1.3.** Текущие тенденции на рынке таковы, что различные критерии имеют отличную друг от друга чувствительность к изменениям предпочтений (значений весовых коэффициентов) вне зависимости от тематик, бюджета и месяца года (см. рис. 1.14, 1.15, 1.16, 1.17).

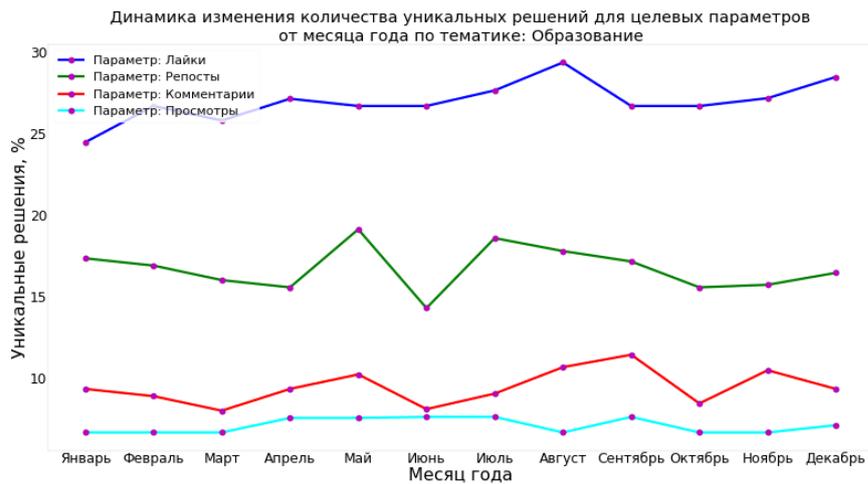
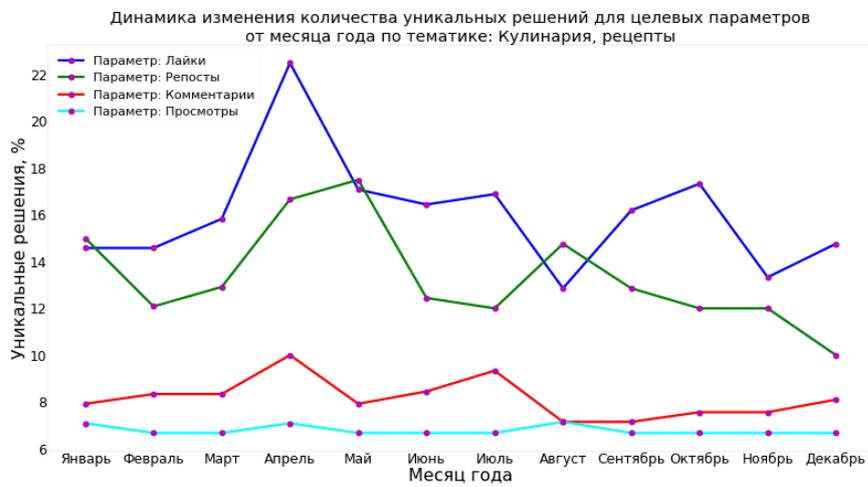
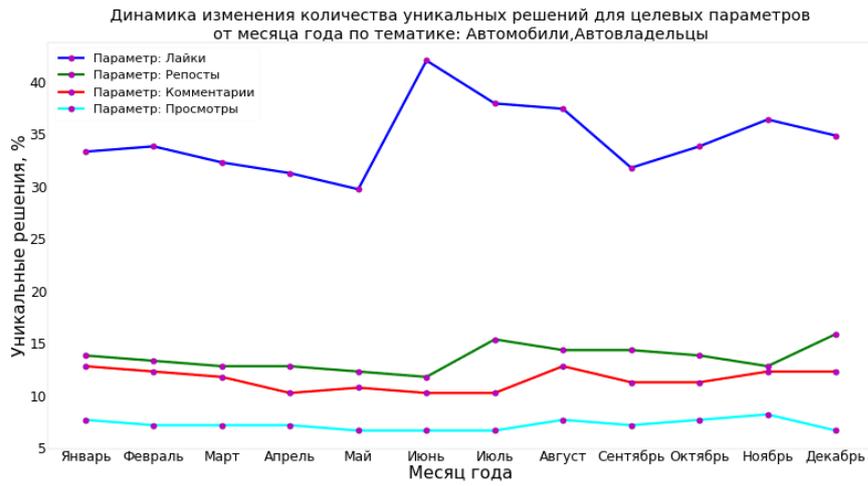


Рис. 1.14: Динамика процента уникальных решений по тематикам и месяцам года в зависимости от критерия

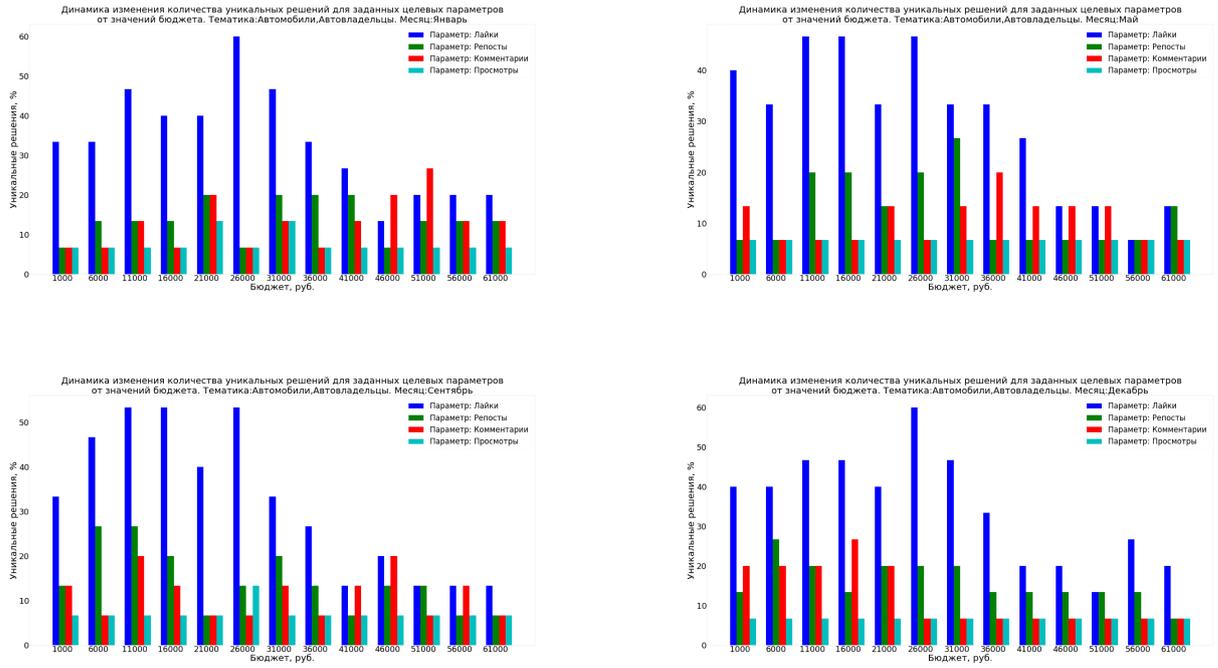


Рис. 1.15: Динамика изменения количества уникальных решений по критериям в зависимости от бюджета. Тематика: «Автомобили, Автовладельцы»

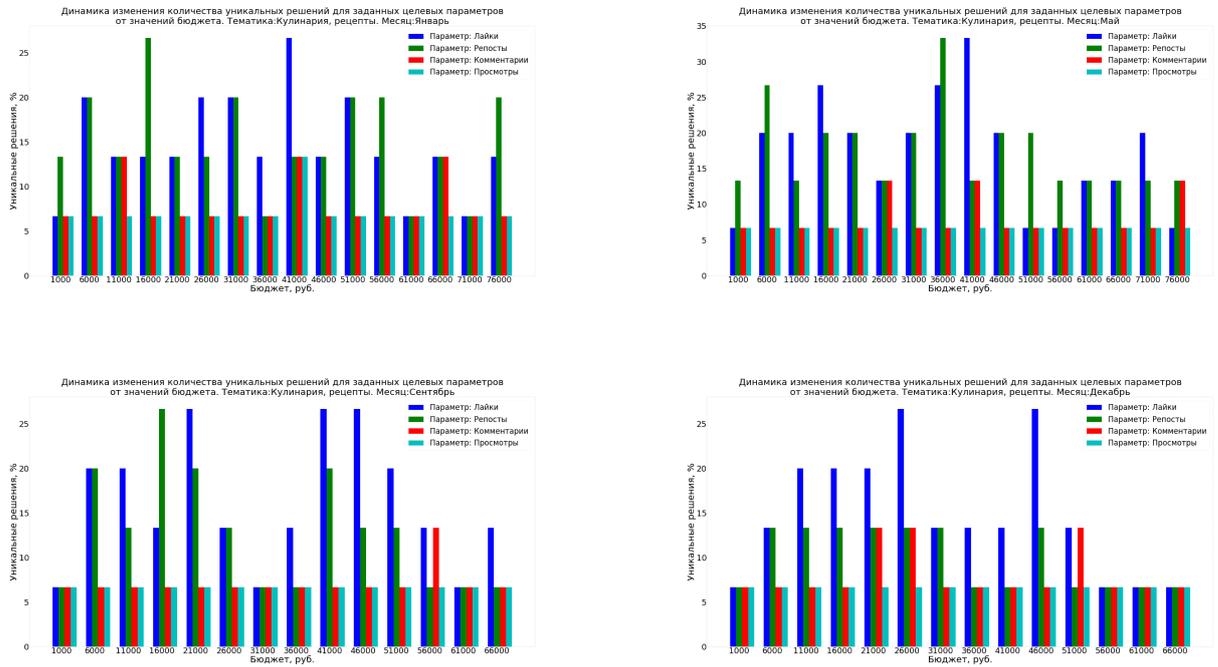


Рис. 1.16: Динамика изменения количества уникальных решений по критериям в зависимости от бюджета. Тематика: «Кулинария, рецепты»

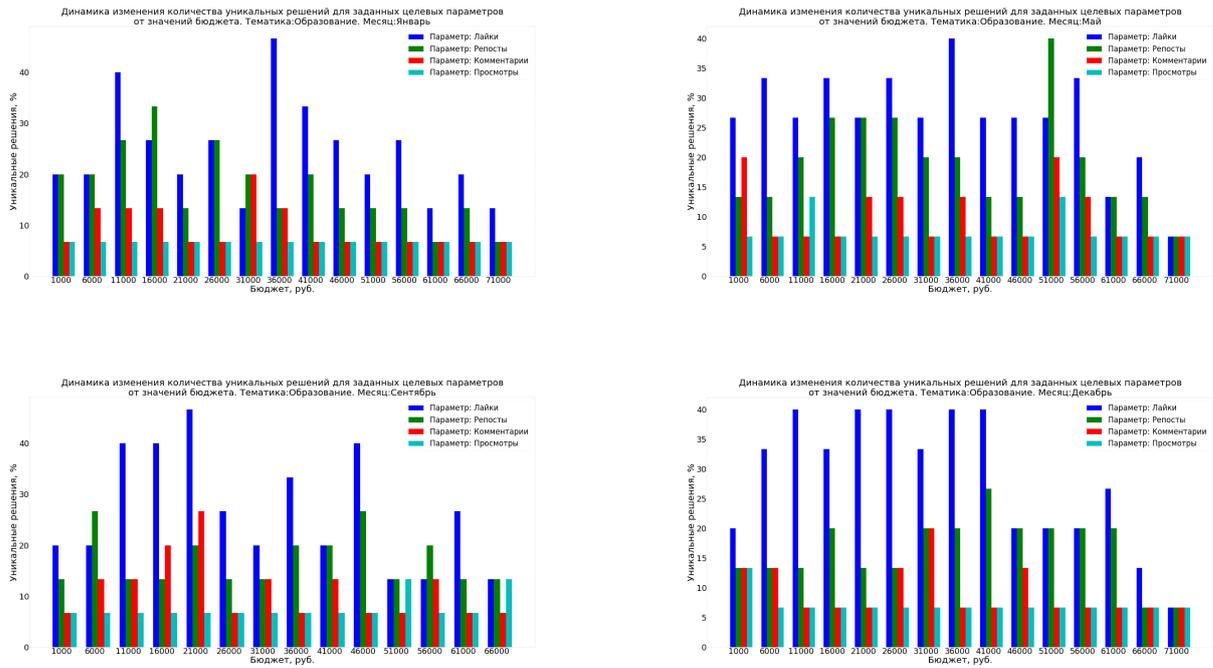


Рис. 1.17: Динамика изменения количества уникальных решений по критериям в зависимости от бюджета. Тематика: «Образование»

#### 1.4. Основные результаты и выводы по первой главе

Данная глава посвящена разработке оптимизационной модели, являющейся инструментом для осуществления консультаций по коррекции изначальных предпочтений заказчика с учетом текущей динамики изменений ключевых характеристик выбранной номенклатуры товара на рынке.

В главе сформулированы постановки задач для моделирования процесса распространения информации в СМК с применением методов оптимизации. Проведен анализ сервисов, предоставляющих данные социальных сетей. Разработаны и имплементированы алгоритмы обработки статистических данных о пользовательской активности информационных площадок в задаче распространения информации в СМК. Предложена архитектура и программно реализована оптимизационная модель с визуализацией, позволяющая формировать набор сообществ социальной сети с рекомендациями по размещению в них информации. Проведены численное моделирование и анализ чувствительности критериев в задаче многокритериальной оптимизации, который показал, что суще-

ствует чувствительность критериев вне зависимости от номенклатуры товаров, бюджета и временного промежутка на примере рынка товаров-услуг в цифровой среде. Показано, что специфика поведенческой активности пользователей социальной сети такова, что изменение предпочтений в задаче многокритериальной оптимизации на треть влияет на процесс формирования уникального сценария распространения информации.

## Глава 2.

# Моделирование наборов площадок в задаче распространения информации на основе методов машинного обучения

### 2.1. Постановка и описание задачи

Методы кластерного анализа относят к разделу машинного обучения без учителя, поскольку испытываемая система обучается выполнять поставленную задачу без вмешательства со стороны экспериментатора, что является одним из видов кибернетического эксперимента. Именно такой подход позволяет решать задачи, когда известны описания множества объектов и требуется обнаружить внутренние взаимосвязи или закономерности между объектами (см. рис. 2.1).

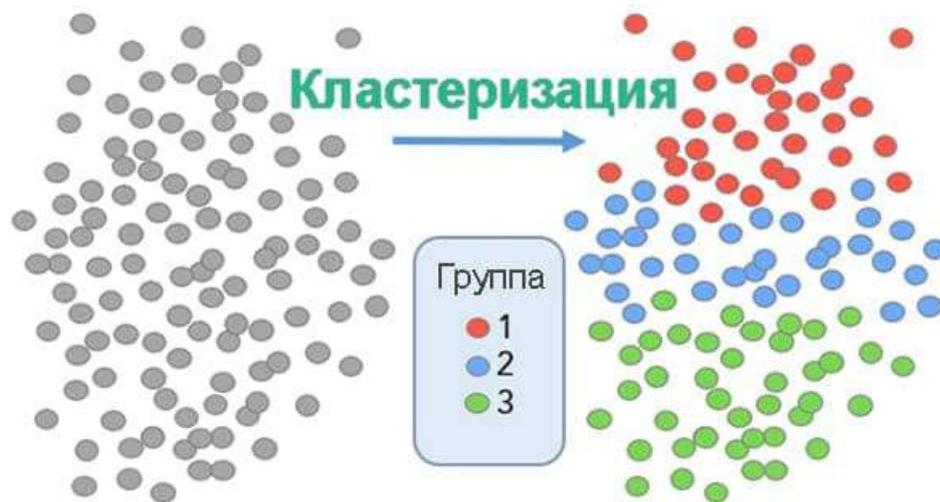


Рис. 2.1: Пример работы методов кластерного анализа

В современном мире методы машинного обучения и искусственного интеллекта становятся все более распространенными и востребованными. Практическая значимость применения указанных методов заключается в возможности создания более эффективных и инновационных решений в различных областях. В данной главе рассмотрена задача распространения информации в СМК с применением методов кластерного анализа и оптимизации. Для демонстрации рабо-

ты предлагаемого подхода к моделированию процесса распространения информации взята одна из таких областей — маркетинг, а именно задача разбиения множества всех клиентов на кластеры для выявления типичных предпочтений. В данном исследовании задача будет сформулирована иначе, однако её суть останется неизменной [70].

В главе предложена комплексная модель распространения информации в СМК, формирующая набор сообществ социальной сети, исходя из заданных входных параметров. Применение такого подхода позволит автоматизировать процесс принятия управленческих решений в рассматриваемой задаче, а также сократить временные и финансовые издержки при проведении соответствующих мероприятий в сети Интернет. Предлагаемая модель учитывает особенности поведенческой активности аудитории каждого сообщества, что способствует максимизации ключевых показателей обратной связи при размещении платных объявлений. Использование методов математической статистики позволяет ЛПР комплексно оценивать смоделированные сценарии продвижения информации и эффективно принимать оперативные и стратегические решения по распределению ограниченного количества ресурсов.

Представленное в работе решение позволяет моделировать набор сообществ при небольших бюджетах, учитывая важность сформулированных заказчиком критериев и особенности поведенческой активности аудитории. Актуальность использования предлагаемого подхода для автоматизации процессов, протекающих в сложных организационных системах, определяется потребностью оптимизировать процесс принятия управленческих решений в области коммуникационных технологий в условиях возрастающего объема распространяемой информации. Следует отметить, что разработка таких прикладных инструментов позволит моделировать процессы в области экономики и управления, социологии и политологии.

В данной главе будут рассмотрены эвристические методы для решения задачи по определению набора площадок распространения информации в СМК, а также для снижения размерности в задаче оптимизации.

## Задача кластерного анализа

*Содержательная постановка задачи:* клиенту требуется сформировать имидж своего предприятия в социальных медиа в рамках определенного бюджета. Клиент хочет провести рекламные мероприятия таким образом, чтобы как можно больше пользователей сети узнали о его компании. Требуется при заданной тематике, временном периоде и бюджете, максимизировать количество просмотров опубликованной рекламной записи. Отметим, что одна и та же рекламная запись может быть опубликована в нескольких сообществах, а также рекламная запись может быть отредактирована отдельно для каждого выбранного сообщества с учетом особенностей ее аудитории.

*Формальная постановка задачи:* используем обозначения введенные на странице 21. Требуется определить такой кластер (набор сообществ)  $U_j$  из разбиения  $\{U_\alpha\}_{\alpha \in \Theta}$  множества  $X$ , где  $\Theta \subseteq M$  — некоторое множество индексов кластеров, используя заданные методы кластеризации и их гиперпараметры, чтобы он максимизировал заданный целевой показатель  $j$ -го кластера в заданном месяце года  $t$  —  $C_j(t) = \sum_{k=1}^{r_j} c_k$ ,  $c_k(t) > 0$  — значение целевого показателя в  $k$ -ом сообществе, при ограничении по бюджету  $B_j = \sum_{k=1}^{r_j} b_k \leq P$ ,  $b_k > 0$  — стоимость размещения записи в  $k$ -ом сообществе,  $j \in \Theta$ ,  $r_j$  — количество сообществ в  $j$  кластере.

## Задача оптимизации с предварительной кластеризацией

*Содержательная постановка задачи целочисленного линейного программирования:* клиенту требуется увеличить объем реализуемой им продукции в рамках определенного бюджета. Клиент хочет провести рекламную кампанию таким образом, чтобы как можно больше пользователей сети узнали о его продукте. Требуется при заданных тематике, временном периоде, целевом параметре и бюджете максимизировать суммарное количество просмотров опубликованных рекламных записей. Отметим, что одна и та же рекламная запись может быть опубликована в нескольких сообществах, а также рекламная запись может быть отредактирована отдельно для каждого выбранного сообщества с учетом особенностей ее аудитории.

*Математическая постановка задачи целочисленного линейного программирования:* используем обозначения введенные на страницах 21, 53. Требуется определить такую комбинацию кластеров (наборов сообществ)  $\{U_j\}_{j=1}^s$ ,  $1 \leq s \leq |\Theta|$ , из разбиения  $\{U_\alpha\}_{\alpha \in \Theta}$  множества  $X$ , где  $\Theta \subseteq M$  — некоторое множество индексов кластеров, используя заданные методы кластеризации и их гиперпараметры, при которой будет достигаться максимальное значение заданного целевого показателя  $\sum_{j=1}^s C_j$  для данной комбинации кластеров, при ограничении  $\sum_{j=1}^s B_j \leq P$ ,  $j \in \Theta$ . Решение задачи представляется определением набора кластеров  $U = (U_1, \dots, U_{|\Theta|})$ , который удовлетворяет следующим требованиям:

$$\begin{aligned} f(U_\alpha) &= \sum_{\alpha \in \Theta} C_\alpha \cdot U_\alpha \rightarrow \max, \\ \sum_{\alpha \in \Theta} B_\alpha \cdot U_\alpha &\leq P, \\ U_\alpha &\in \{0; 1\}, \alpha \in \Theta. \end{aligned}$$

*Содержательная постановка задачи многокритериальной оптимизации:* клиенту требуется провести позиционирование бренда в рамках определенного бюджета. Клиент хочет провести рекламные мероприятия таким образом, чтобы как можно больше пользователей сети узнали о его продукте. Однако, клиент отдает предпочтение сообществам, в которых люди оставляют обратную связь в комментариях под записями, а также активно делятся записями сообщества у себя на личной странице. Следовательно, требуется при заданной тематике, временном периоде и бюджете, возможно наличии портрета целевой аудитории, максимизировать количество комментариев и отметок «поделиться» опубликованной рекламной записи. Таким образом, для клиента равновесны два показателя, значения которых мы и будем максимизировать. Отметим, что одна и та же рекламная запись может быть опубликована в нескольких сообществах, а также рекламная запись может быть отредактирована отдельно для каждого выбранного сообщества с учетом особенностей ее аудитории.

*Математическая постановка задачи многокритериальной оптимизации:* к обозначениям введенным на страницах 21, 53, 54 добавим 4 критерия, заданных с помощью функций:  $f_1(U_\alpha)$ ,  $f_2(U_\alpha)$ ,  $f_3(U_\alpha)$ ,  $f_4(U_\alpha)$  и зададим значения соответствующих критериев в  $j$ -ом кластере для среднестатистической записи в заданном месяце года  $t$  —  $A_j(t)$ ,  $B_j(t)$ ,  $C_j(t)$ ,  $D_j(t)$ , где  $A_j(t) = \sum_{k=1}^{r_j} a_k$ ,  $B_j(t) = \sum_{k=1}^{r_j} b_k$ ,  $C_j(t) = \sum_{k=1}^{r_j} c_k$ ,  $D_j(t) = \sum_{k=1}^{r_j} d_k$  и  $a_k(t) > 0$ ,  $b_k(t) > 0$ ,  $c_k(t) > 0$ ,  $d_k(t) > 0$ ; переопределим стоимость размещения рекламной записи в  $j$ -ом кластере (наборе сообществ) —  $G_j = \sum_{k=1}^{r_j} g_k$ , где  $g_k > 0$  — стоимость рекламной записи в  $k$ -ом сообществе. Решение задачи представляется определением набора кластеров  $U = (U_1, \dots, U_{|\Theta|})$ , который удовлетворяет следующим требованиям:

$$\left\{ \begin{array}{l} f_1(U_\alpha) = \sum_{\alpha \in \Theta} A_\alpha \cdot U_\alpha \rightarrow \max, \\ f_2(U_\alpha) = \sum_{\alpha \in \Theta} B_\alpha \cdot U_\alpha \rightarrow \max, \\ f_3(U_\alpha) = \sum_{\alpha \in \Theta} C_\alpha \cdot U_\alpha \rightarrow \max, \\ f_4(U_\alpha) = \sum_{\alpha \in \Theta} D_\alpha \cdot U_\alpha \rightarrow \max, \\ \sum_{\alpha \in \Theta} G_\alpha \cdot U_\alpha \leq P, \\ U_\alpha \in \{0; 1\}, \alpha \in \Theta. \end{array} \right. \quad (2.1.1)$$

Используя метод свертки критериев, преобразовать систему (2.1.1):

$$\begin{aligned} \hat{f}(U_\alpha) &= \beta_1 \cdot f_1(U_\alpha) + \beta_2 \cdot f_2(U_\alpha) + \beta_3 \cdot f_3(U_\alpha) + \beta_4 \cdot f_4(U_\alpha) = \\ &= \beta_1 \cdot \sum_{\alpha \in \Theta} A_\alpha \cdot U_\alpha + \beta_2 \cdot \sum_{\alpha \in \Theta} B_\alpha \cdot U_\alpha + \beta_3 \cdot \sum_{\alpha \in \Theta} C_\alpha \cdot U_\alpha + \\ &+ \beta_4 \cdot \sum_{\alpha \in \Theta} D_\alpha \cdot U_\alpha = \sum_{\alpha \in \Theta} (\beta_1 \cdot A_\alpha + \beta_2 \cdot B_\alpha + \beta_3 \cdot C_\alpha + \beta_4 \cdot D_\alpha) \cdot U_\alpha = \\ &= \sum_{\alpha \in \Theta} W_\alpha \cdot U_\alpha. \end{aligned}$$

Применяя представленные выше линейные преобразования к (2.1.1), получим систему вида:

$$f(U_\alpha) = \sum_{\alpha \in \Theta} W_\alpha \cdot U_\alpha \rightarrow \max,$$

$$\sum_{\alpha \in \Theta} G_\alpha \cdot U_\alpha \leq P,$$

$$U_\alpha \in \{0; 1\}, \alpha \in \Theta.$$

где  $W_\alpha = \beta_1 \cdot A_\alpha + \beta_2 \cdot B_\alpha + \beta_3 \cdot C_\alpha + \beta_4 \cdot D_\alpha$  — это показатель суммарной активности публикации за выбранный временной период;  $\beta = \{\beta_1, \beta_2, \beta_3, \beta_4\}$  — весовые коэффициенты критериев или предпочтения клиента.

## 2.2. Комплексная модель с применением методов кластеризации

Архитектура модели представлена в виде блок-схемы на рисунке 2.2 и реализована с помощью пяти основных функциональных блоков:

1. Предобработка данных;
2. Обработка данных;
3. Формирование рекомендаций для публикации рекламных записей;
4. Построение разбиения (кластеризация объектов);
5. Оптимизация.

Сформулированная задача кластеризации в параграфе 2.1 отличается от задачи понижения размерности наличием или отсутствием блока с оптимизацией. В этом можно убедиться изучив схему, представленную на рисунке 2.2.



Рис. 2.2: Архитектура модели с применением методов кластерного анализа и оптимизации

Рассмотрим более подробно каждый из этапов. Отметим, что «Процесс предобработки данных (циклический алгоритм обработки выделенных типов данных)», «Процесс обработки данных», «Процесс формирования рекомендаций» были описаны в главе 1.

### 2.2.1. Методы отбора и выделения признаков

В данной работе для осуществления процедуры сжатия пространства признаков применяются как методы отбора, так и выделения признаков.

Отбор признаков — это процедура для оценивания значимости тех или иных признаков с помощью статистических методов и, в том числе, алгоритмов машинного обучения с целью формирования пространства признаков меньшей размерности. Отбор признаков используется по четырем причинам:

- для «упрощения» модели и повышения интерпретируемости результатов;
- для сокращения времени обучения;

- для избежания «проклятия» размерности;
- для улучшения обобщающей способности модели и борьбы с переобучением.

Целью применения техники отбора признаков является удаление без существенной потери информации излишних данных или незначимых признаков. «Незначимость» и «избыточность» признаков — отличные друг от друга понятия, это объясняется тем, что один значимый признак может быть излишним при наличии другого значимого признака, сильно коррелирующего с ним. Выбор подходящего метода отбора признаков зависит от специфики задачи и имеющихся данных.

Выделение признаков, в отличие от процедуры отбора, направлено на формирование новых характеристик, как функции от оригинальных признаков, в то время как результатом процедуры отбора является подмножество признаков.

Существует достаточное количество алгоритмов, подходящих для решения различных прикладных задач и анализа данных. В данном исследовании будут использоваться методы отбора и выделения признаков для задачи обучения без учителя. В разработанной модели методы отбора признаков представлены в виде функций (см. рис. 2.3). Следует также отметить, что некоторые алгоритмы возвращают один вектор признаков, а другие один и более. Многокластерные методы отбора признаков в задачах машинного обучения без учителя представляют собой эффективный подход к снижению размерности данных и выделению наиболее информативных признаков. Эти методы основаны на идее группировки признаков в несколько кластеров, что позволяет выявить структуру данных и выделить наиболее значимые признаки для каждого кластера.

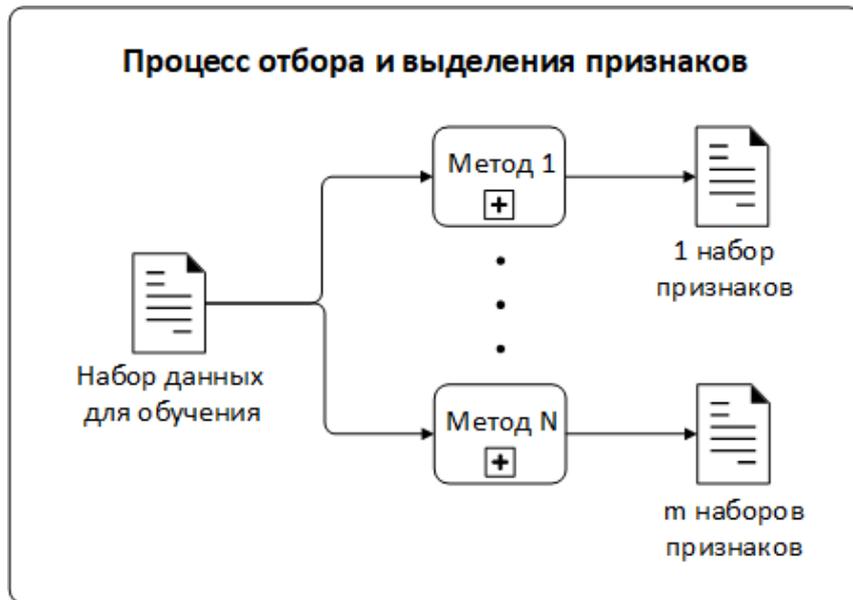


Рис. 2.3: Схема реализации отбора и выделения признаков

Рассмотрим более подробно используемые методы:

1. Дисперсия: было показано, что оценка дисперсии признака может быть эффективным способом отбора признаков. Как правило признаки с почти нулевой дисперсией не являются значимыми, и их можно удалить [50].
2. Средняя абсолютная разница: вычисляется средняя абсолютная разность между значениями признака и его средним значением. Более высокие значения, как правило, имеют более высокую предсказательную силу [85].

$$MAD_i = \frac{1}{n} \times \sum_{j=1}^n |X_{ij} - \bar{X}_i|$$

3. Соотношение дисперсий: среднее арифметическое, деленное на среднее геометрическое. Более высокая дисперсия соответствует более релевантным признакам [85].

$$AM_i = \bar{X} = \frac{1}{n} \times \sum_{j=1}^n X_{ij}$$

$$GM_i = \prod_{j=1}^n X_{ij}$$

Так как  $AM_i \geq GM_i$  выполняется тогда и только тогда, когда  $X_{i1} = X_{i2} = \dots = X_{in}$ , следовательно:

$$R_i = \frac{AM_i}{GM_i} \in [1, +\infty)$$

4. Оценка Лапласа: в основе лежит наблюдение, что данные из одного класса часто расположены ближе друг к другу, поэтому можно оценить важность признака по его способности отражать эту близость. Метод состоит из встраивания данных в граф ближайших соседей с помощью измерения произвольного расстояния с последующим вычислением матрицы весов. Затем для каждого признака вычисляем критерий Лапласа и получаем такое свойство, что наименьшие значения соответствуют самым важным размерностям [102].
5. Многокластерный отбор признаков: выбираются такие признаки, чтобы лучше всего сохранить многокластерную структуру данных. Количество кластеров задано от 2 до максимального количества объектов с шагом в 5 процентов от общего числа объектов. Из всех получившихся вариантов наборов признаков были выбраны три уникальные [98].
6. Метод главных компонент: представляет собой алгебраический метод преобразования набора наблюдений, возможно коррелированных переменных, в набор значений линейных некоррелированных переменных [55].

Таким образом, методы отбора признаков в задачах машинного обучения без учителя играют важную роль в обработке данных, позволяя сфокусироваться на наиболее значимых аспектах и улучшить качество последующего анализа без необходимости разметки данных экспертами.

Представим результаты применения первых пяти методов отбора признаков (см. рис. 2.4, 2.5, 2.6 , 2.7 , 2.8).

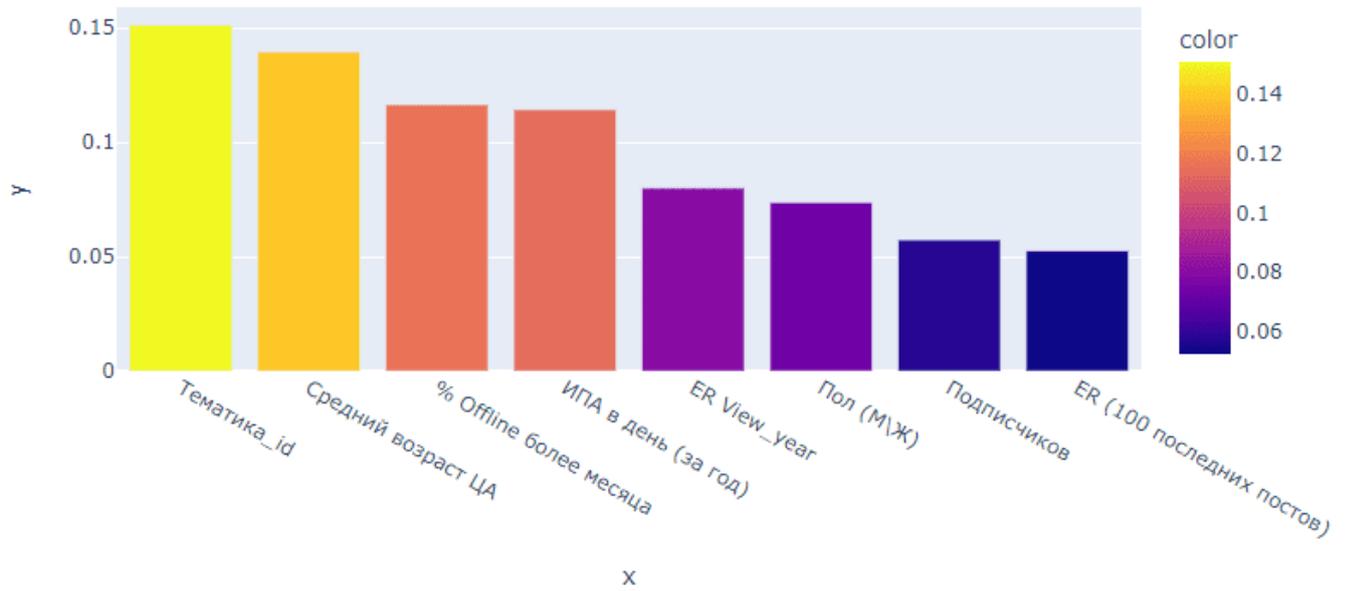


Рис. 2.4: Результат применения метод: «Дисперсия»

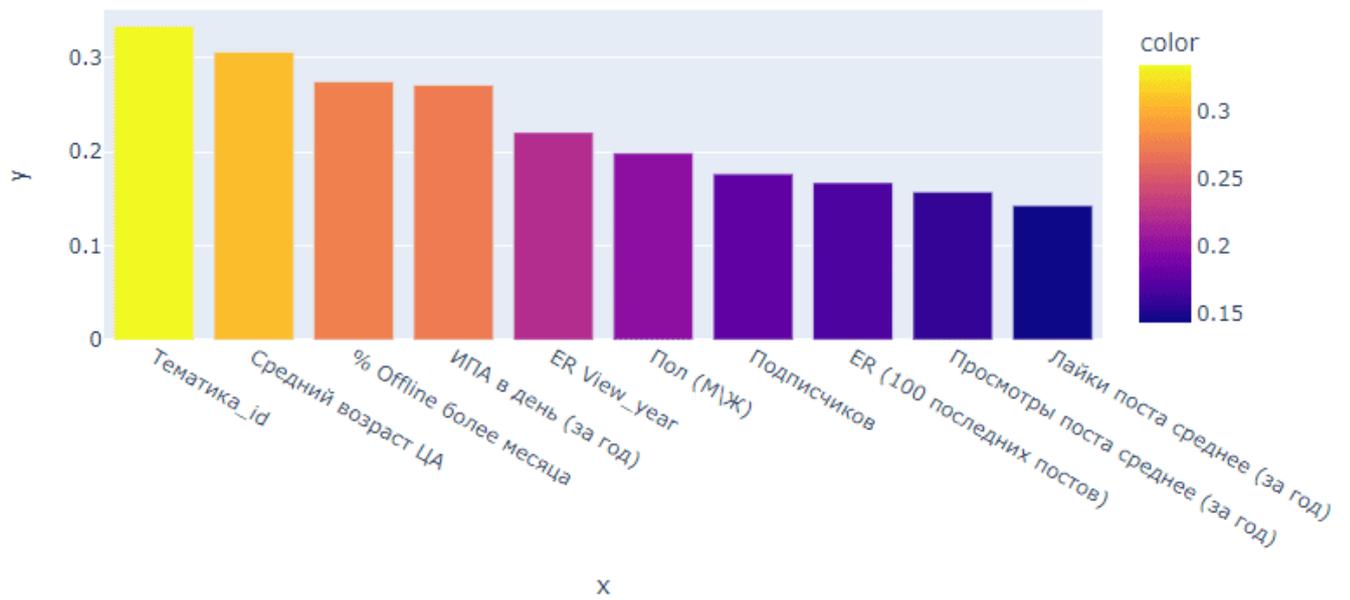


Рис. 2.5: Результат применения метода: «Средняя абсолютная разность»

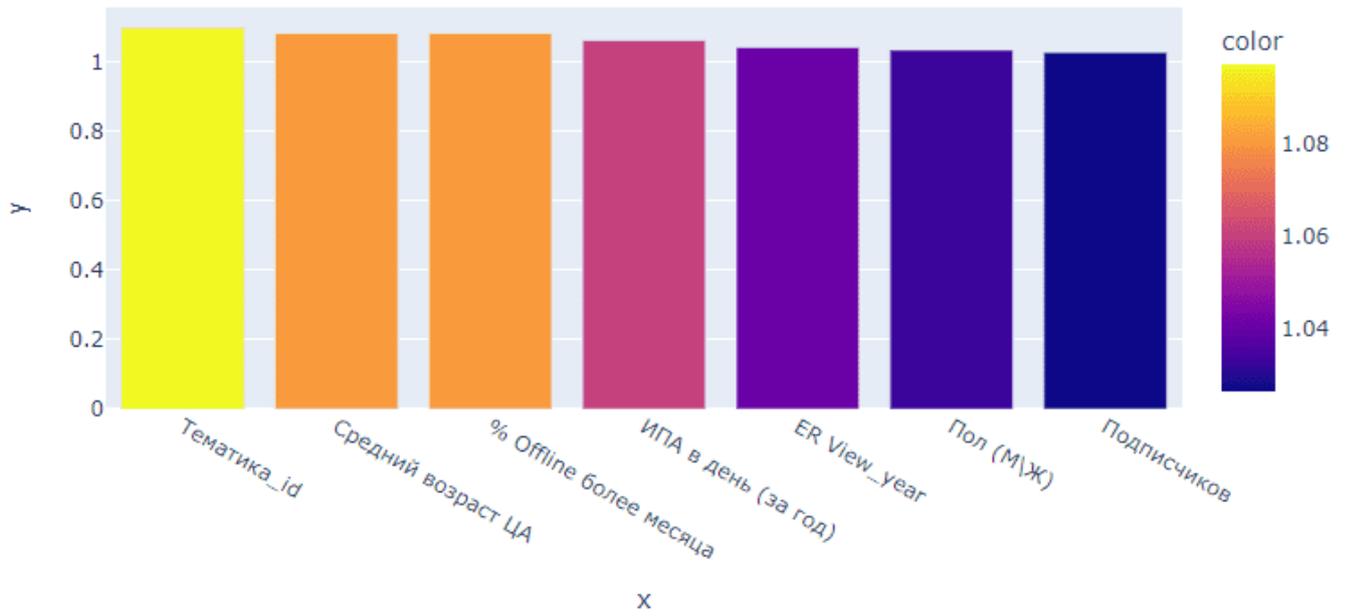


Рис. 2.6: Результат применения метода: «Соотношение дисперсий»

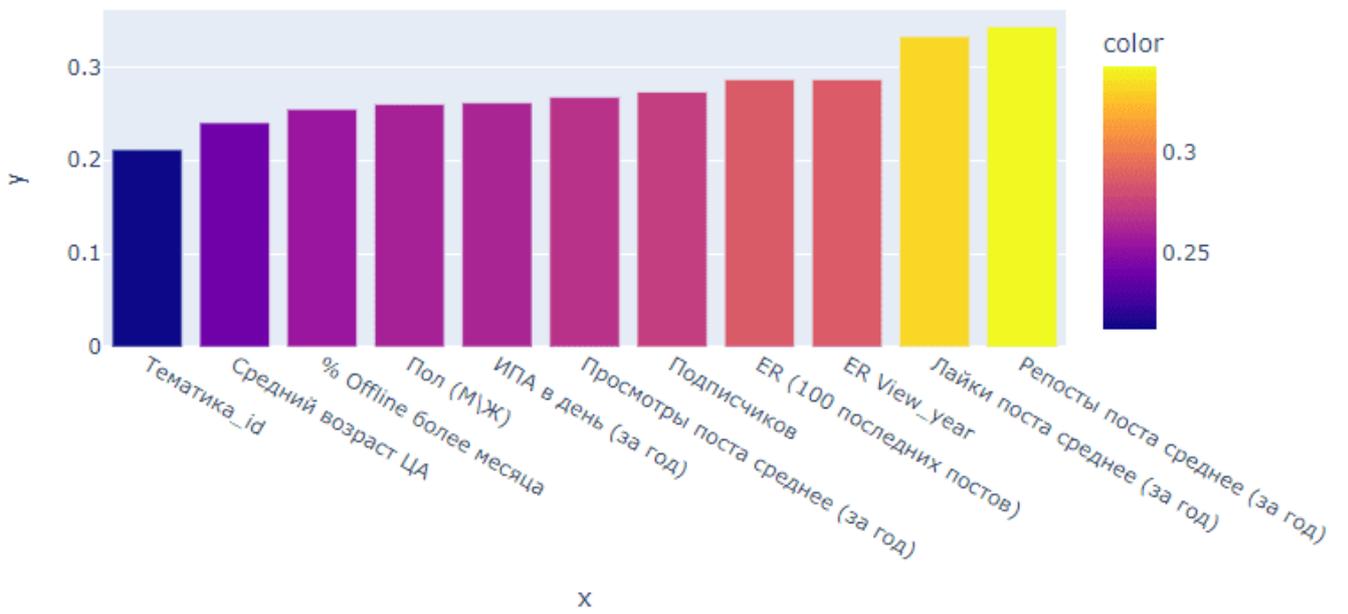


Рис. 2.7: Результат применения метода: «Оценка Лапласа»

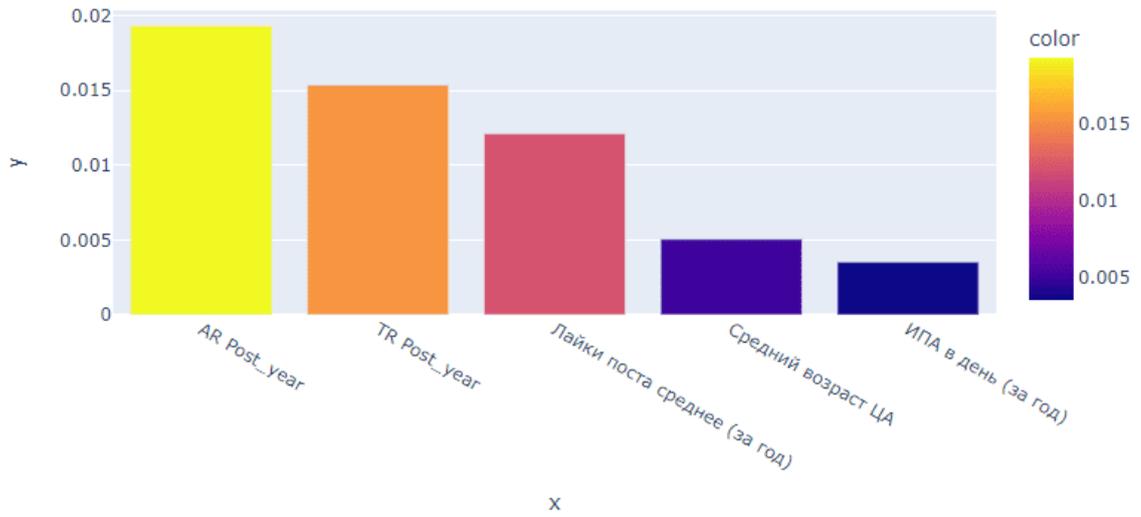
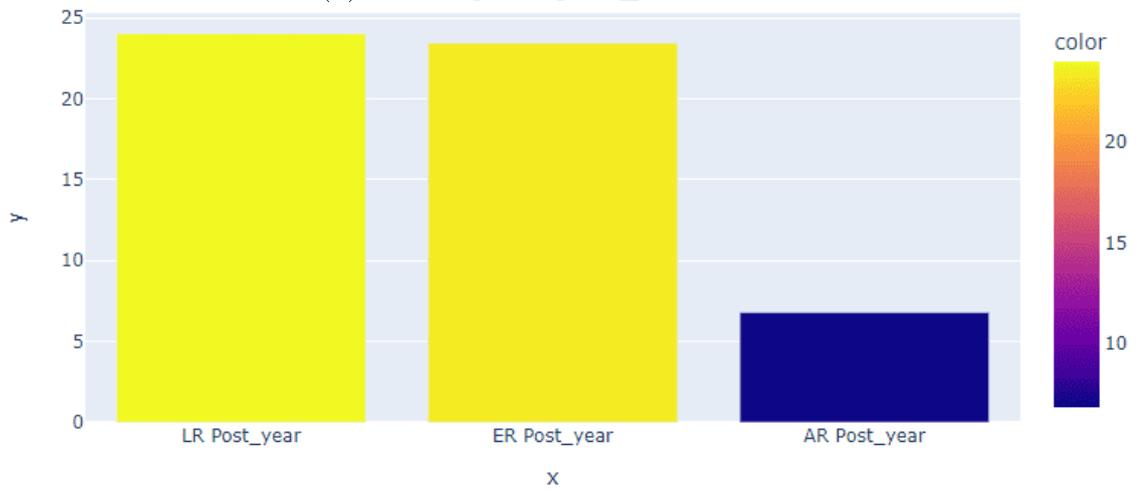
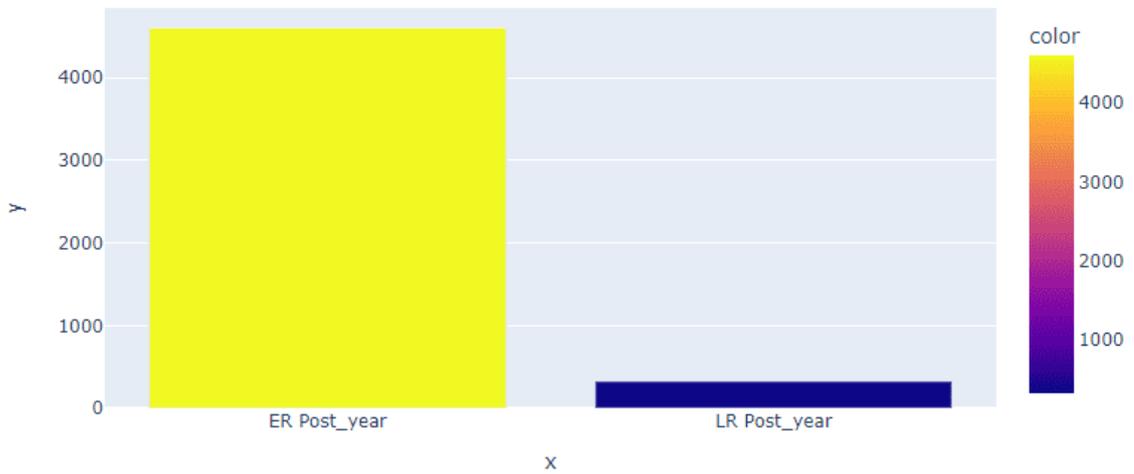
(a) Для параметра  $n\_clusters = 2$ .(b) Для параметра  $n\_clusters = 171$ .(c) Для параметра  $n\_clusters = 678$ .

Рис. 2.8: Результат применения метода: «Многокластерный отбор признаков»

Особенность реализации первых пяти методов заключается в том, что были вычислены контрольные значения для каждого признака, а затем вычислялось среднее и выбирались признаки, которые принимали контрольные значения выше среднего. Таким образом, мы получили 7 наборов наиболее значимых признаков для методов 1-5. Далее эти наборы подаются в модель кластеризации, происходит обучение, подбор лучших гиперпараметров моделей кластеризации по внешнему критерию качества получаемых разбиений.

Следует отметить, что базовым набором значимых характеристик являются: «Тематика\_id», «Средний возраст ЦА», «% Offline более месяца», «ИПА в день (за год)», «ER View\_year», «Пол (М/Ж)», «Подписчиков». Удаление избыточных признаков позволяет лучше понять данные, а также сократить время настройки модели, улучшить её точность и облегчить интерпретируемость результатов. Иногда эта задача и вовсе может быть самой значимой, например, нахождение оптимального набора признаков может помочь расшифровать механизмы, лежащие в основе исследуемой проблемы.

Метод главных компонент — один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации. С математической точки зрения метод главных компонент представляет собой ортогональное линейное преобразование, которое отображает данные из исходного пространства признаков в новое пространство меньшей размерности. Задача метода главных компонент заключается в том, чтобы построить новое пространство признаков меньшей размерности, дисперсия между осями которой будет перераспределена так, чтобы максимизировать дисперсию по каждой из них. Для этого выполняется последовательность следующих действий:

- Вычисляется общая дисперсия исходного пространства признаков. Это нельзя сделать простым суммированием дисперсий по каждой переменной, поскольку они, в большинстве случаев, не являются независимыми. Поэтому суммировать нужно взаимные дисперсии переменных, которые определяются из ковариационной матрицы.
- Вычисляются собственные векторы и собственные значения ковариацион-

ной матрицы, определяющие направления главных компонент и величину связанной с ними дисперсии.

- Производится снижение размерности. Диагональные элементы ковариационной матрицы показывают дисперсию по исходной системе координат, а ее собственные значения — по новой. Тогда разделив дисперсию, связанную с каждой главной компонентой на сумму дисперсий по всем компонентам, получаем долю дисперсии, связанную с каждой компонентой. После этого отбрасывается столько главных компонент, чтобы доля оставшихся составляла 80-90%.

Существенным недостатком применения данного метода для осуществления процедуры выделения признаков является невозможность смысловой интерпретации компонент, поскольку они «вбирают» в себя дисперсию от нескольких исходных переменных. Кроме того, необходимо проверить все или большинство главных компонент, то есть, если в предыдущих методах мы получали набор признаков, который подавали в модель кластеризации затем по внешнему критерию определяли лучшие гиперпараметры и некоторый набор сообществ, то здесь, необходимо проанализировать результаты кластеризации для  $2, 3, 4, 5, \dots, n$  главных компонент, где  $n$  - общее количество признаков, и выбрать лучшее значение для получения оптимального результата по внешнему критерию. В следствие чего, данный метод требует отдельной от предыдущих архитектуры программной реализации.

### **2.2.2. Методы кластерного анализа и метрики качества разбиений**

Кластеризация данных является одним из ключевых инструментов машинного обучения основная цель которого заключается в группировке объектов на основе их сходства, таким образом, что объекты в одной группе (кластере) более похожи друг на друга, чем на объекты из других групп. Этот метод находит широкое применение в различных областях, таких как биоинформатика, финансовая аналитика, обработка изображений и многие другие. Существует несколько основных методов кластеризации данных, включая иерархические

методы (агломеративная и дивизионная кластеризация), метод  $k$ -средних, метод DBSCAN, методы спектральной кластеризации и многие другие. Каждый из этих методов имеет свои преимущества и недостатки, и выбор конкретного метода зависит от характеристик исходных данных, поставленных задач и требуемой интерпретируемости результатов. Результаты кластерного анализа могут быть использованы для выявления групп схожих объектов, создания сегментации аудитории, определения паттернов поведения и многих других прикладных задач. Для проведения данного исследования было рассмотрено несколько методов кластеризации (см. рис. 2.9), проведен анализ их применимости для решения сформулированных задач при использовании различных методов отбора признаков.

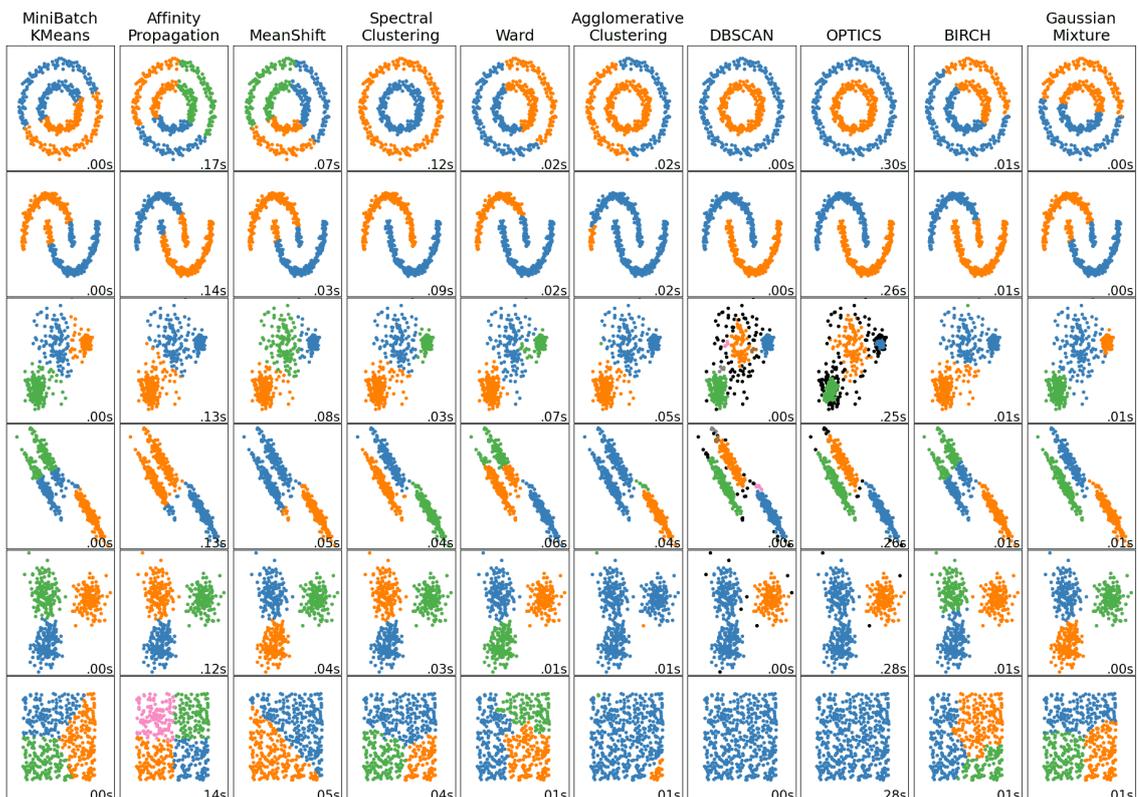


Рис. 2.9: Методы кластерного анализа

В данном разделе методы кластерного анализа будут представлены в виде таблицы 2.1 с описанием заданных гиперпараметров и интервалов их значений. Отметим, что такое количество выбранных методов кластеризации обосновывается проведением сравнительного анализа в рассматриваемой прикладной задаче. Перейдем к рассмотрению методов кластерного анализа [48].

Таблица 2.1: Методы кластерного анализа

№	Метод кластеризации	Гиперпараметры	Значения гиперпараметров
1	«Agglomerative clustering»	«n_clusters»	от 2 до $(n - 1)$ с шагом $(0.1 \times n)$ , где $n$ - это количество объектов в заданном множестве
		«linkage»	[«complete»; «average», «single»]
2	«Ward»	«n_clusters»	от 2 до $(n - 1)$ с шагом $(0.1 \times n)$ , где $n$ - это количество объектов в заданном множестве
3	«K-Means»	«n_clusters»	от 2 до $(n - 1)$ с шагом $(0.1 \times n)$ , где $n$ - это количество объектов в заданном множестве
4	«MiniBatch K-Means»	«n_clusters»	от 2 до $(n - 1)$ с шагом $(0.1 \times n)$ , где $n$ - это количество объектов в заданном множестве
5	«Affinity propagation»	«damping»	от 0.5 до 0.95 с шагом 0.05
6	«Mean-shift»	-	-
7	«Spectral clustering»	«n_clusters»	от 2 до $(n - 1)$ с шагом $(0.1 \times n)$ , где $n$ - это количество объектов в заданном множестве
		«gamma»	$[10^{-2}; 10^{-1}; 1; 10^1; 10^2]$

Продолжение таблицы 2.1

8	«DBSCAN»	«eps»	от 0.1 до 0.9 с шагом 0.2
		«min_samples»	от 5 до $(n - 1)$ с шагом $(0.1 \times n)$ , где $n$ - это количество объектов в заданном множестве
9	«OPTICS»	«min_samples»	от 2 до $(n - 1)$ с шагом $(0.1 \times n)$ , где $n$ - это количество объектов в заданном множестве
10	«Gaussian mixtures»	«n_components»	от 2 до $(n - 1)$ с шагом $(0.1 \times n)$ , где $n$ - это количество объектов в заданном множестве
		«covariance_type»	[«full»; «tied»; «diag»; «spherical»]
11	«BIRCH»	«n_clusters»	от 2 до $(n - 1)$ с шагом $(0.1 \times n)$ , где $n$ - это количество объектов в заданном множестве
		«threshold»	от 0.1 до 0.9 с шагом 0.2

Далее приведены меры оценивания качества получаемых разбиений, опирающиеся только непосредственно на структуру кластеров, не используя внешней информации. Проблема оценки качества в задаче кластеризации трудноразрешима, как минимум, по двум причинам: 1) Теорема невозможности Клейнберга [105] — не существует оптимального алгоритма кластеризации; 2) Многие алгоритмы кластеризации не способны определить настоящее количество кластеров в данных, чаще всего количество кластеров подается на вход алгоритма и под-

бирается несколькими запусками алгоритма. Поэтому в данной работе рассматривается несколько метрик качества получаемых разбиений [49] в результате работы алгоритмов кластеризации для проведения сравнительного анализа результатов численного моделирования:

- Индекс «Силуэт» (англ. «Silhouette»).

Значение силуэта показывает, насколько объект похож на свой кластер по сравнению с другими кластерами [108]. Оценка для всей кластерной структуры:

$$Sil(C) = \frac{1}{N} \cdot \sum_{c_k \in C} \sum_{x_i \in c_k} \frac{b(x_i, c_k) - a(x_i, c_k)}{\max\{a(x_i, c_k); b(x_i, c_k)\}},$$

где:  $a(x_i, c_k) = \frac{1}{|c_k|} \cdot \sum_{x_j \in c_k} \|x_i - x_j\|$  - среднее расстояние от  $x_i \in c_k$  до других объектов из кластера  $c_k$  (компактность);

$b(x_i, c_k) = \min_{c_l \in C \setminus c_k} \left\{ \frac{1}{|c_l|} \cdot \sum_{x_j \in c_l} \|x_i - x_j\| \right\}$  - среднее расстояние от  $x_i \in c_k$  до объектов из другого кластера  $c_l : k \neq l$  (отделимость).

Чем ближе данная оценка к 1, тем лучше.

- Индекс «Калински-Харабаша» (англ. «Calinski-Harabasz»).

Компактность основана на расстоянии от точек кластера до их центроидов, а разделимость - на расстоянии от центроидов кластеров до глобального центроида [99]. Представляет собой отношение среднего значения дисперсии между кластерами и дисперсии внутри кластера. Для набора данных  $E$  размера  $n_E$ , который был размечен на  $k$  кластеров:

$$CH = \frac{tr(B_k)}{tr(W_k)} \times \frac{n_E - k}{k - 1},$$

где:  $W_k = \sum_{q=1}^k \sum_{x \in C_q} (x - c_q) \cdot (x - c_q)^T$  - матрица дисперсии между кластерами;  $B_k = \sum_{q=1}^k n_q \cdot (c_q - c_E) \cdot (c_q - c_E)^T$  - матрица внутрикластерной дисперсии;  $C_q$  - набор объектов кластера  $q$ ;  $c_q$  - центр кластера  $q$ ;  $c_E$  - центр  $E$ ;  $n_q$  - количество объектов в кластере  $q$ .

Чем выше значение, тем лучше произведена кластеризация.

- Индекс «Дэвиса-Боулдина» (англ. «Davies–Bouldin Index»).

Это, возможно, одна из самых используемых мер оценки качества кластеризации [100]. Она вычисляет компактность как расстояние от объектов кластера до их центроидов, а отделимость - как расстояние между центроидами. То есть показывает среднее «сходство» кластеров: расстояние между ними сравнивается с их размером.

$$DB = \frac{1}{k} \cdot \sum_{i=1}^k \max_{i \neq j} \{R_{ij}\},$$

где:  $R_{ij} = \frac{s_i + s_j}{d_{ij}}$  - мера сходства кластеров  $i$  и  $j$ ,  $s_i$  - среднее расстояние между каждой точкой кластера  $i$  и центроида этого кластера, то есть диаметр  $i$ -ого кластера;  $d_{ij}$  - расстояние между центроидами кластеров  $i$  и  $j$ .

Чем меньше значение, тем лучше произведено разделение на кластеры.

### 2.2.3. Описание и реализация комплексной модели

Комплексные математические модели широко применяются в науке, инженерии, экономике и других областях для анализа, прогнозирования и оптимизации различных систем. Они позволяют исследователям и специалистам получить глубокое понимание поведения системы в различных условиях, проводить численные эксперименты и моделировать сценарии. Использование комплексных математических моделей позволяет ЛПР принимать обоснованные решения, оптимизировать процессы и улучшать работу системы в целом. Используя комплексную модель с методами машинного обучения, управленцы могут анализировать большие объемы данных, выявлять скрытые зависимости и оптимизировать стратегии принятия решений. Такие модели способны учитывать множество переменных и факторов, что позволяет принимать более точные и обоснованные управленческие решения.

Разработанная комплексная модель реализована следующим образом: на вход функции, реализующей метод кластеризации, подаются данные: для обучения, для рекомендаций, тематики, целевой показатель, бюджет клиента, месяц года. Далее выполняются следующие пункты:

1. Формирование подмножества объектов указанных тематик.
2. Для него определяется задается стартовый бюджет, шаг и максимально возможный бюджет. Поскольку данные ограничены, необходимо определить наибольший возможный бюджет клиента для каждого набора тематик.
3. Создаются таблицы для записи результатов моделирования.
4. Для заданного бюджета, целевого показателя, тематик, месяца года и одного из нормализованных наборов данных для обучения выполняются вложенные циклы.
5. В теле цикла вызывается функция, реализующая методы кластерного анализа. В ней происходит вызов функций из блока «Формирования рекомендаций».
6. Задаются векторы значений гиперпараметров метода кластеризации и запускаются вложенные циклы.
7. Выполняется метод кластеризации и получается некоторое разбиение  $\{U_\alpha\}_{\alpha \in \Theta}$  множества  $X$ .
8. Вычисляются метрики качества разбиения.
9. Для каждого кластера вычисляется суммарная стоимость размещения записи и отбираются те кластеры, стоимость которых не превышает заданный клиентом бюджет:  $\{G_\alpha \mid G_\alpha \leq P, \alpha \in \Theta\}$ .
10. Далее в зависимости от подхода выполняется:

- если это задача определения лучшего кластера среди полученных разбиений, то для каждого кластера вычисляется суммарное значение целевого показателя  $C_j(t)$ , вычисляется максимум по  $C_j(t)$  и таким образом определяется лучший кластер;
- если это много задача оптимизации с предварительной кластеризацией, то как и в главе 1 для оптимизации времени работы алгоритмов, решающих поставленные задачи, была принято решение реализовать одну функцию, в которой решается как задача многокритериальной оптимизации, так и задача целочисленного линейного программирования.
- далее с помощью библиотеки «`scipy.optimize`» [12] решается задача оптимизации и определяется результирующий вектор  $U$ .
- вычисляются значения соответствующих критериев для оптимальных комбинаций кластеров при заданной матрице весовых коэффициентов.

#### 11. Фиксируются результаты моделирования в сравнительные таблицы.

Таким образом, предлагаемая комплексная модель, основанная на методах машинного обучения и оптимизации, представляет собой интегрированный подход к анализу данных и моделированию сценариев распространения информации в СМК для принятия управленческих решений. Эта модель объединяет в себе алгоритмы машинного обучения с комплексным представлением системы или процесса, что позволяет получить глубокое понимание взаимосвязей и паттернов в данных. Применение комплексных моделей с методами машинного обучения и оптимизации в управленческой деятельности позволяет повысить эффективность принятия решений, минимизировать риски и улучшить стратегическое планирование на основе данных и аналитики.

### 2.3. Численное моделирование и сравнительный анализ результатов

Для демонстрации работы предлагаемого подхода и возможности проведения сравнительного анализа полученных результатов зададим следующие входные параметры:

1. Тематики - «Автомобили, Автовладельцы», «Кулинария, рецепты», «Образование». Количество объектов в каждой тематике: 124, 126, 103;
2. Временные интервалы - ежемесячно, Январь, Май, Сентябрь, Декабрь;
3. Бюджет - от 1000 рублей до максимально возможного бюджета в рамках заданных тематик и месяцев года с шагом в 5000 рублей;
4. Предпочтения клиента заданы в векторной форме  $\beta = \{\beta_1, \beta_2, \beta_3, \beta_4\}$  и представлены в таблице 2.2:

Таблица 2.2: Весовые коэффициенты для моделирования

$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
1.0	0.0	0.0	0.0
0.0	1.0	0.0	0.0
0.0	0.0	1.0	0.0
0.0	0.0	0.0	1.0
0.25	0.25	0.25	0.25

Отметим, что параметры указанные в пункте 4 будут использоваться только при решении задачи оптимизации с предварительной разметкой объектов, а такое количество вариантов векторов весовых коэффициентов объясняется необходимостью демонстрации возможности получения различных решений при изменении предпочтений. Для программной реализации использовался язык программирования Python, среда разработки PyCharm, а также соответствующие библиотеки: Pandas, Numpy, Datetime, Scikit-learn, Math, SciPy, Matplotlib, Glob. Важно отметить, что компоненты комплексной модели были программно реализованы и зарегистрированы соответствующие

программы для ЭВМ в ФИПС [80, 82]. Следует также указать, что данная комплексная модель была спроектирована в виде отдельных блоков и функций таким образом, что позволяет исследователю при появлении внешних сбоев во время моделирования беспрепятственно продолжить его с момента, где оно было прервано. Всего таких блоков четыре:

- Блок 1: кластеризация с применением интерпретируемых методов отбора признаков.
- Блок 2: кластеризация с применением метода главных компонент для сжатия пространства признаков.
- Блок 3: предварительная кластеризация с использованием интерпретируемых методов отбора признаков в задаче оптимизации.
- Блок 4: предварительная кластеризация с использованием метода главных компонент для сжатия пространства признаков в задаче оптимизации.

Вследствие чего, формулирование выводов, исходя из результатов моделирования, будет происходить по типу наблюдений. Перейдем к их рассмотрению и анализу.

**Вывод 2.1.** *В полученных решениях прослеживается уменьшение количества кластеров с увеличением бюджета при применении методов кластерного анализа вне зависимости от выбираемых тематик, целевых параметров и месяцев года (см. рис. 2.10, 2.11).*

Данный вывод позволяет заключить, что при увеличении бюджета необходимо уменьшать значения гиперпараметра «n\_clusters» с целью сокращения времени обучения в соответствующих методах кластеризации. Это объясняется процедурой выбора лучшего кластера среди получившегося разбиения при заданных гиперпараметрах метода, описанной в параграфе 2.2.



**Вывод 2.2.** При применении методов оптимизации с предварительной кластеризацией изменение количества кластеров с увеличением бюджета если и происходит, то на некоторую малую величину вне зависимости от выбираемых тематик, целевых параметров, бюджета и месяцев года (см. рис. 2.12, 2.13).

Следует отметить, что при выполнении процедуры предварительной разметки объектов одни методы кластерного анализа работают так, что понижение размерности в задаче линейного программирования происходит до 25 – 30%, а другие уменьшают количество переменных на 70% и более. Это можно объяснить различными принципами построения разбиения множества в рассматриваемых методах кластеризации. Кроме того, данный вывод позволяет заключить, что имеется возможность сокращения времени обучения за счет задания ограничений для значений гиперпараметра «n\_clusters» в соответствующих алгоритмах.

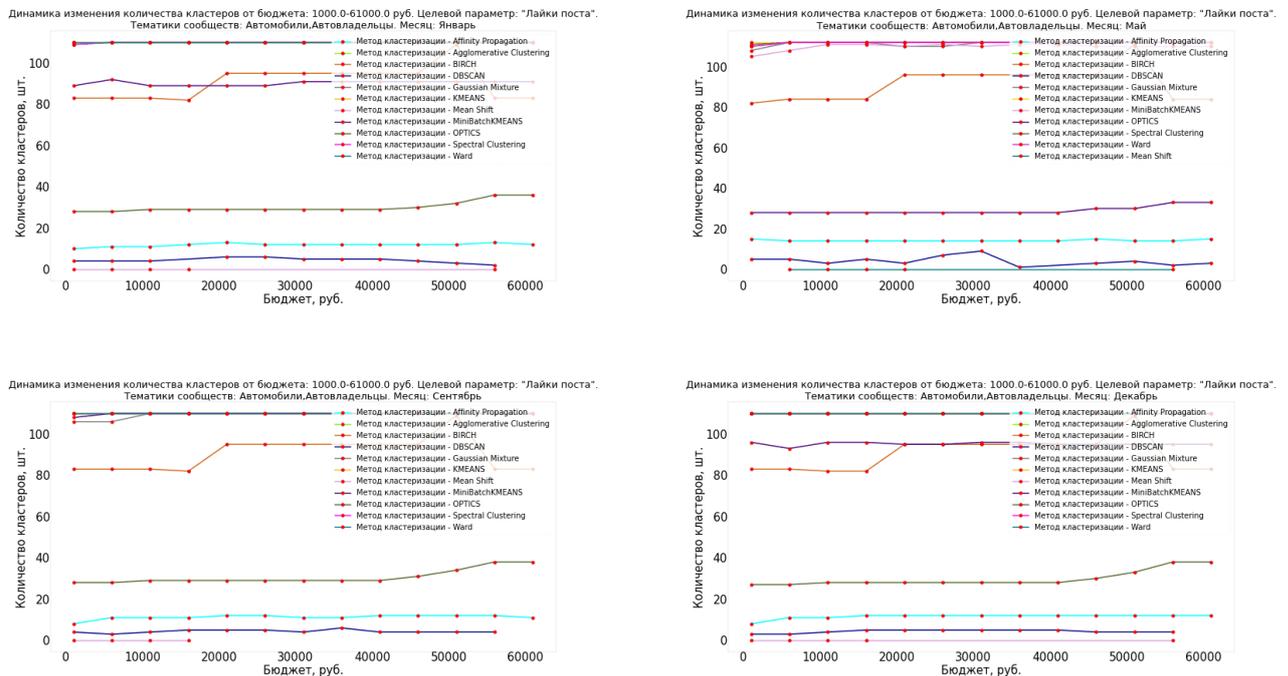


Рис. 2.12: Динамика изменения количества кластеров в зависимости от бюджета. Блок 3

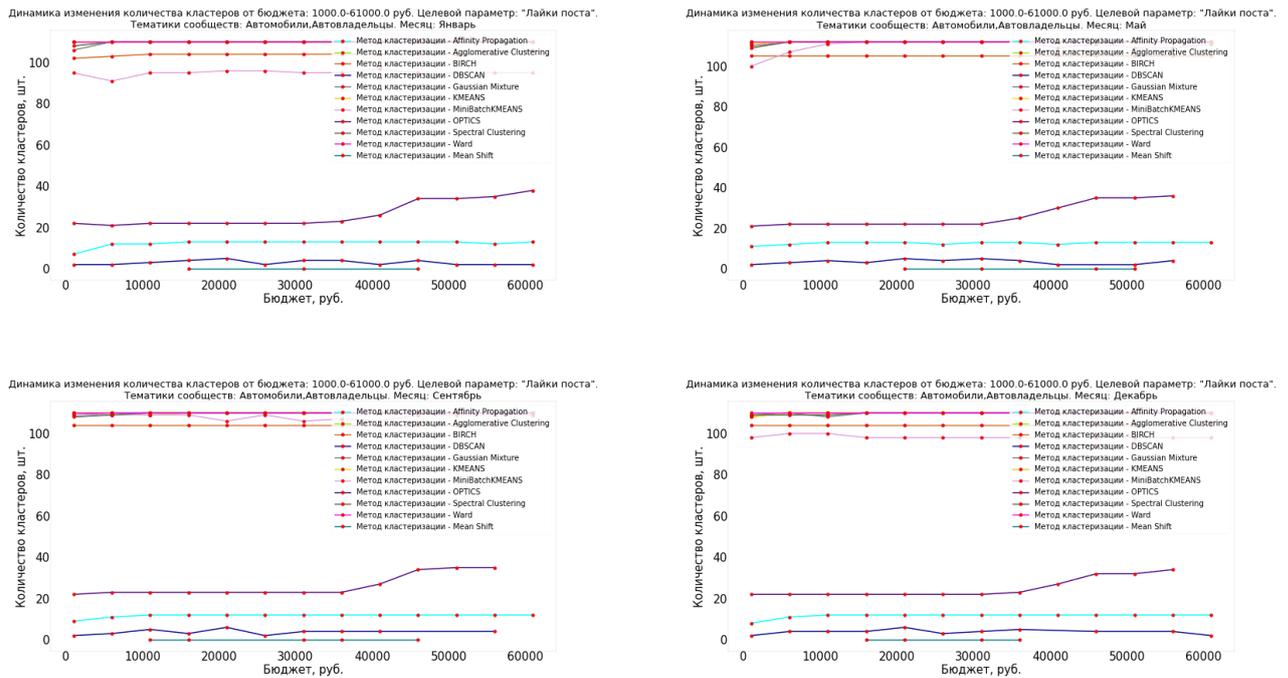


Рис. 2.13: Динамика изменения количества кластеров в зависимости от бюджета. Блок 4

**Вывод 2.3.** Динамика изменения количества объектов в получаемых решениях такова, что, при прочих равных, с увеличением бюджета она является возрастающей вне зависимости от применяемого математического инструментария, выбираемых тематик, целевых параметров и месяцев года (см. рис. 2.14, 2.15, 2.16, 2.17).

Данный вывод является закономерным, поскольку с увеличением бюджета имеется возможность разместить публикацию в большем количестве площадок социальной сети. Однако при применении методов кластерного анализа и наличии в заданной тематике сообществ, у которых соотношение единицы вложенных денежных средств к количественным показателям обратной связи аудитории выше чем у других, мы можем получить, что при увеличении бюджета, с некоторых его значений, число объектов в получаемых решениях будет являться невозрастающей величиной. Такое встречается редко при использовании оптимизационного подхода и также является следствием процедуры выбора лучших кластеров среди получившегося разбиения.

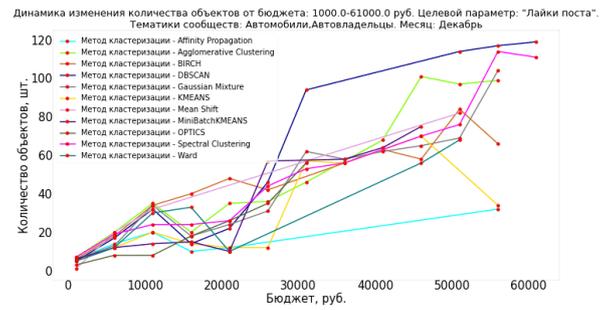
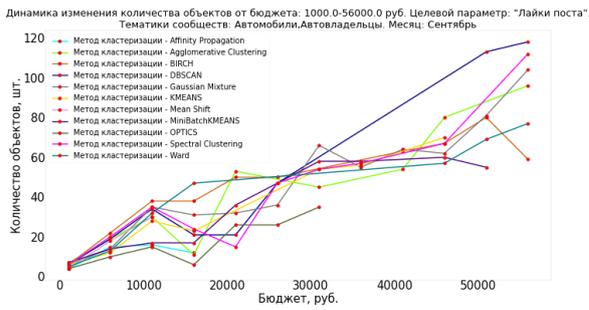
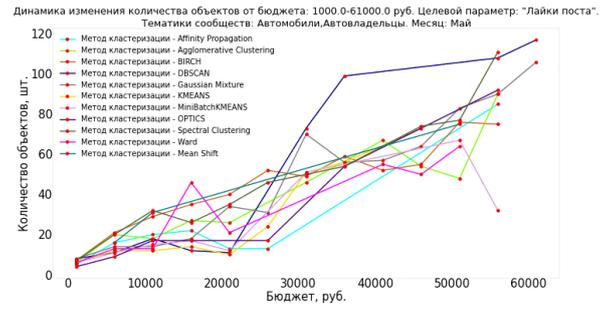
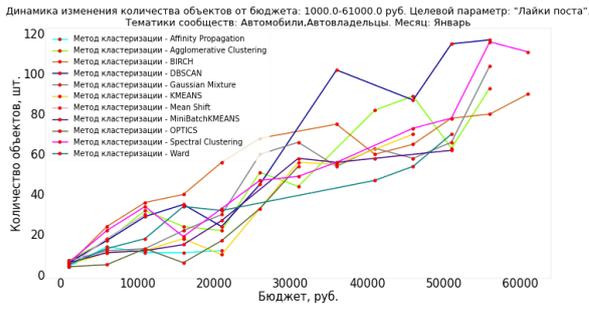


Рис. 2.14: Динамика изменения количества объектов в зависимости от бюджета. Блок 1

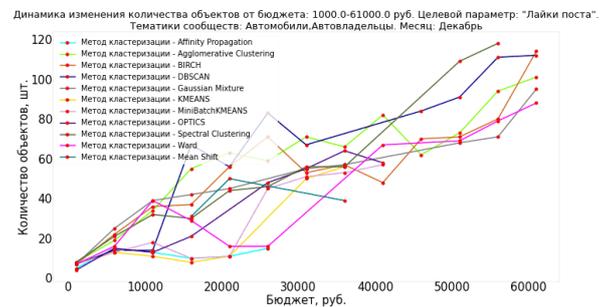
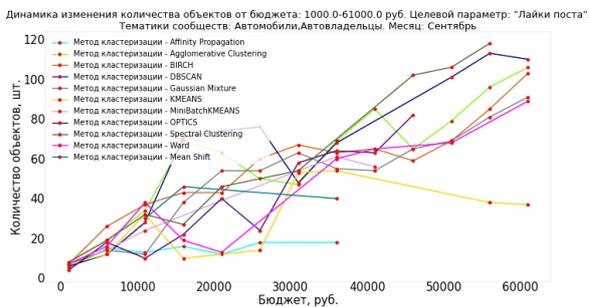
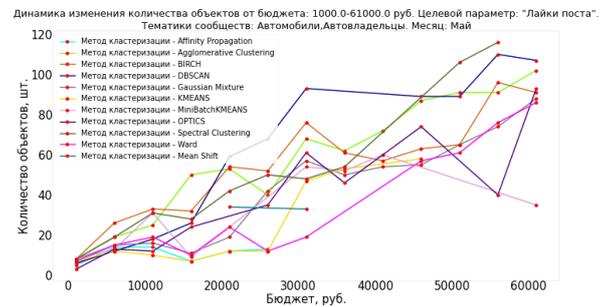
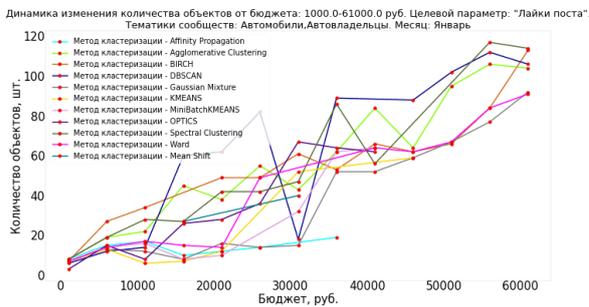


Рис. 2.15: Динамика изменения количества объектов в зависимости от бюджета. Блок 2

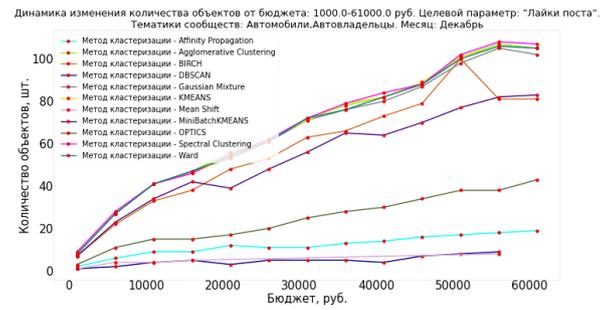
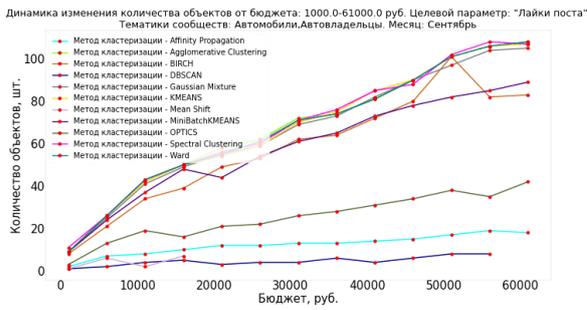
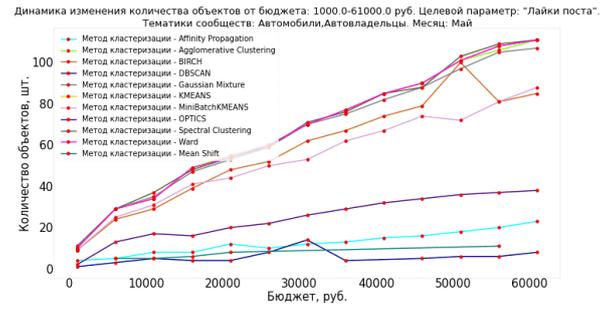
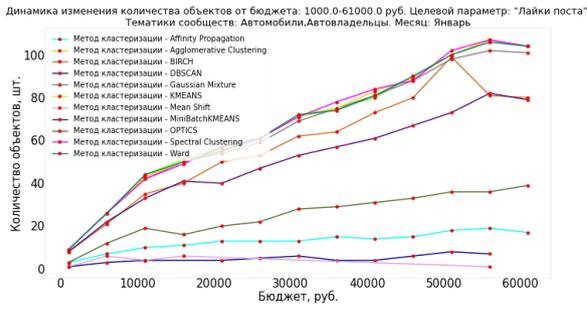


Рис. 2.16: Динамика изменения количества объектов в зависимости от бюджета.  
Блок 3

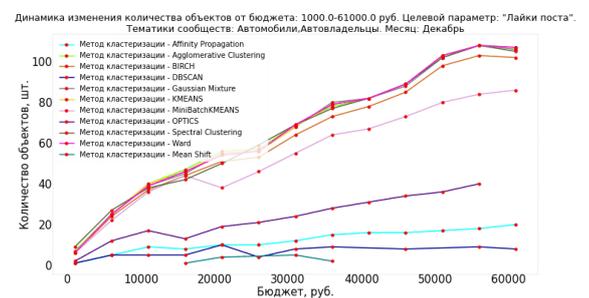
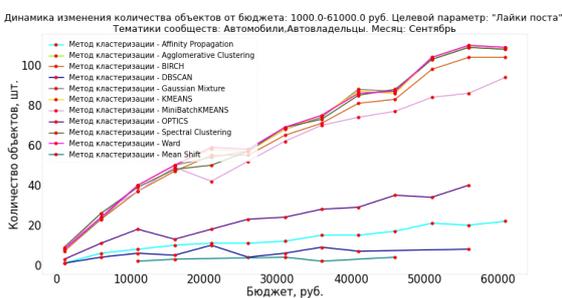
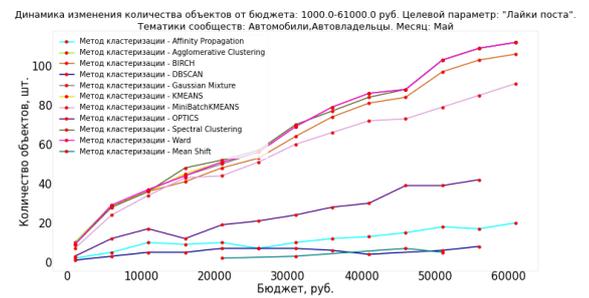
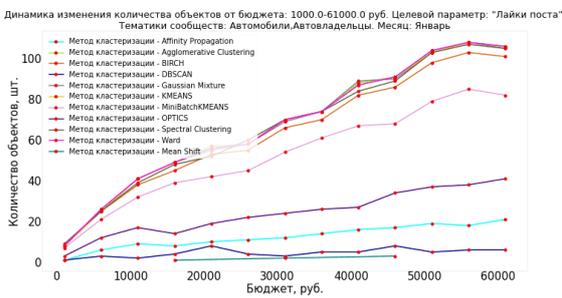


Рис. 2.17: Динамика изменения количества объектов в зависимости от бюджета.  
Блок 4

Кроме того, разработанный инструмент позволяет предоставлять клиенту право выбора сценариев с различным количеством сообществ для публикации рекламной записи, а исследователям определять причинно-следственные связи между получаемыми результатами и применением соответствующих методов кластерного анализа.

**Замечание 2.1.** *Анализируя динамику изменения максимальных значений целевого параметра, можно выделить ряд методов кластеризации, дающих лучшее решение вне зависимости от применяемого математического инструментария, выбираемых тематик, целевых параметров, бюджета и месяцев года (см. рис. 2.18, 2.19, 2.20, 2.21).*

Данное замечание говорит о том, что имеется возможность определить какие методы кластерного анализа следует применять для получения максимальных значений целевого параметра в различных сценариях продвижения информации, сократить время их формирования, а также формулировать рекомендации для клиента по целесообразности задания определенного значения бюджета.

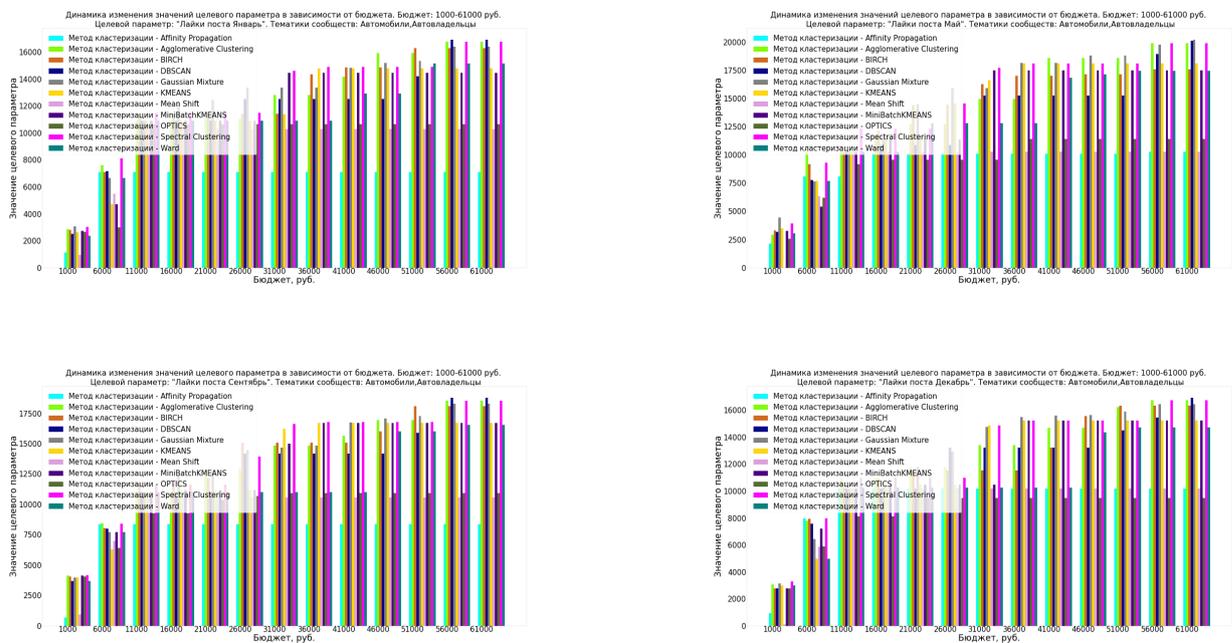


Рис. 2.18: Динамика изменения значений целевого параметра в зависимости от бюджета. Блок 1

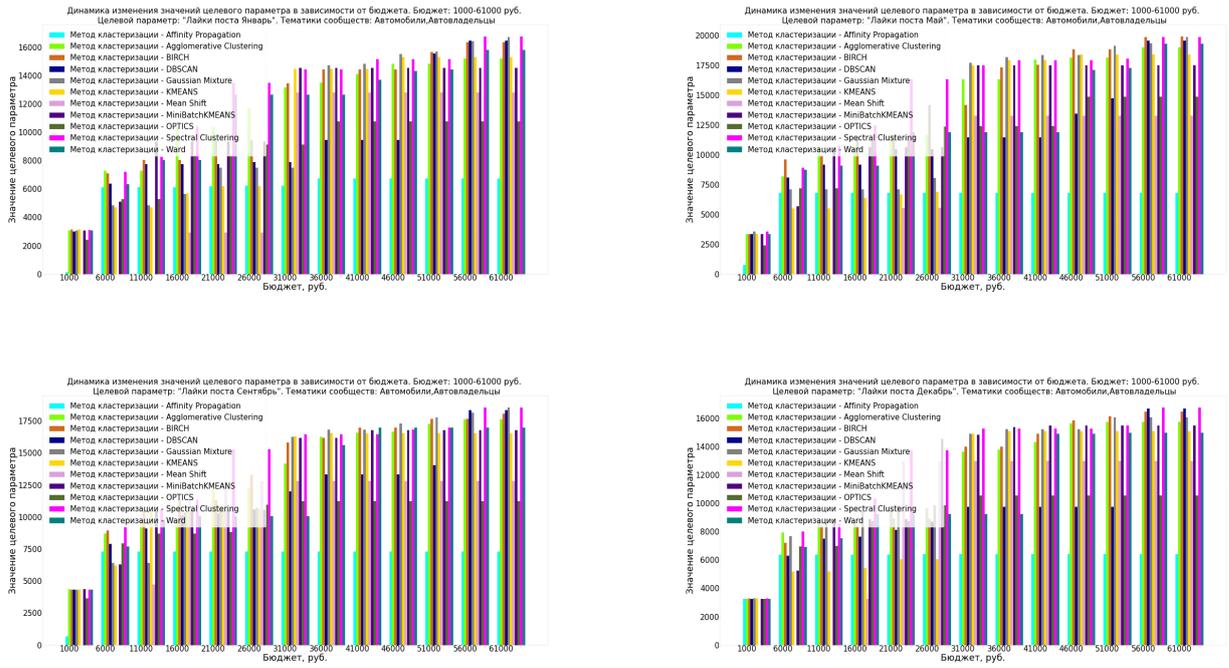


Рис. 2.19: Динамика изменения значений целевого параметра в зависимости от бюджета. Блок 2

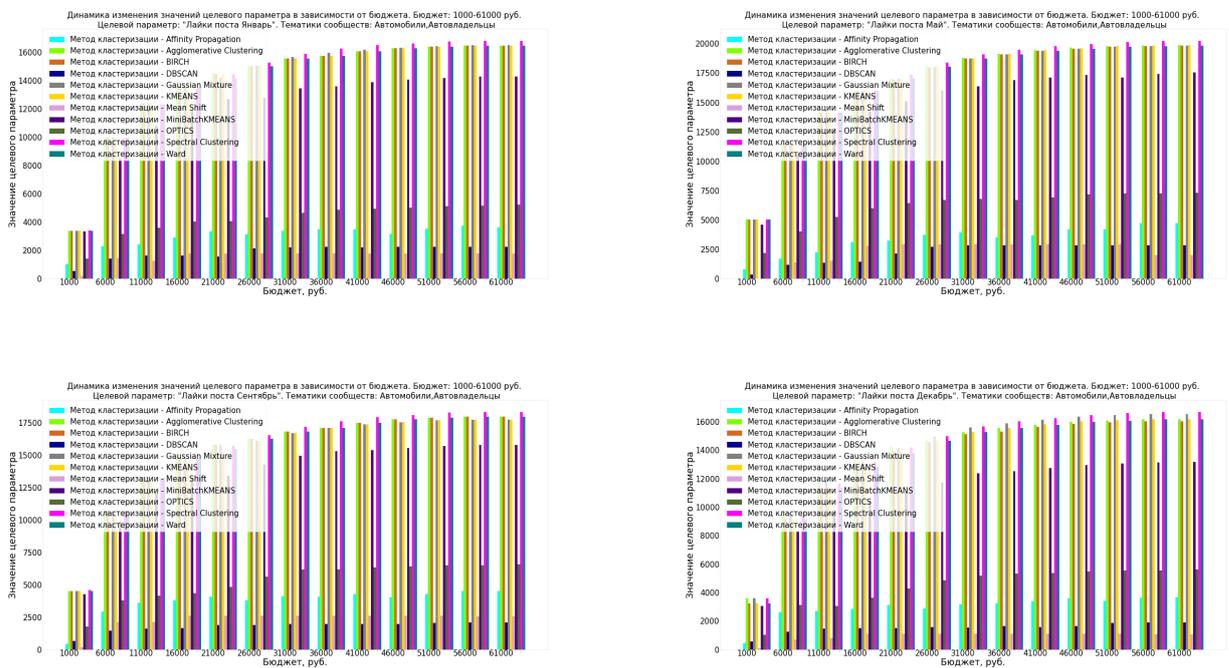


Рис. 2.20: Динамика изменения значений целевого параметра в зависимости от бюджета. Блок 3

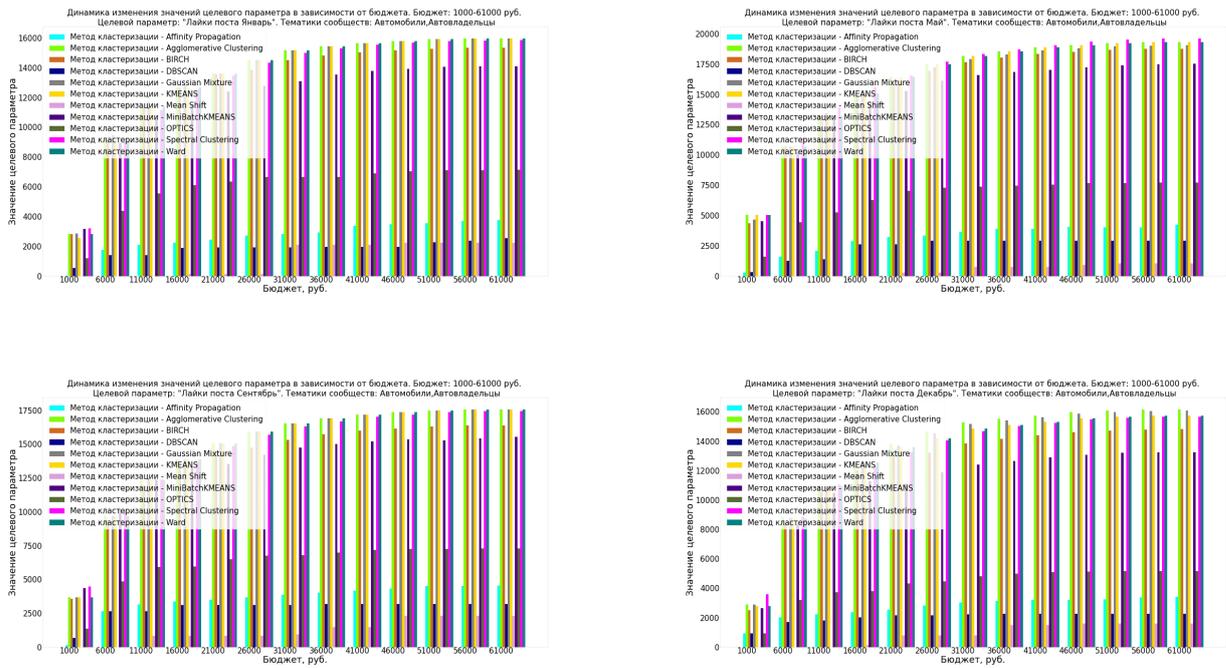
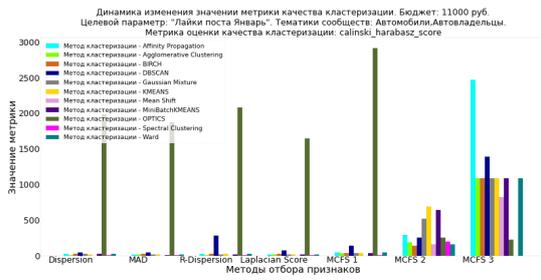


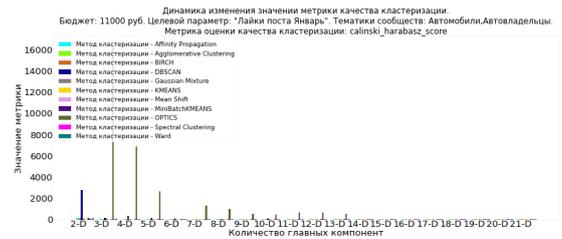
Рис. 2.21: Динамика изменения значений целевого параметра в зависимости от бюджета. Блок 4

**Замечание 2.2.** Динамика изменения значений метрик качества показывает, что можно выбрать лучшие методы уменьшения размерности признакового пространства для соответствующих методов кластеризации и входных параметров (см. рис. 2.22, 2.23, 2.24).

Поскольку существует несколько способов оценивания качества разбиения, в данном исследовании были рассмотрены основные из них для раздела обучения без учителя. Отметим, что предлагаемый подход для решения поставленных задач, позволит формулировать выводы о разделимости и плотности получаемых разбиений. Кроме того, появится возможность определять в первом случае, лучший интерпретируемый метод отбора признаков, во втором, необходимое количество главных компонент для используемых методов кластерного анализа, что может положительно отразиться на сокращении времени получения различных сценариев продвижения информации в средствах массовой коммуникации.



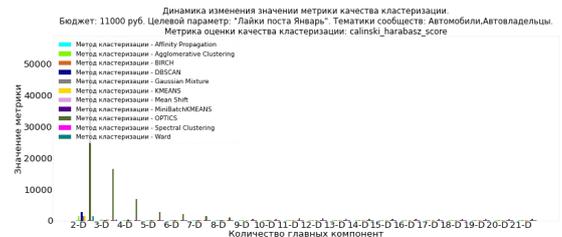
(a) Блок 1



(b) Блок 2

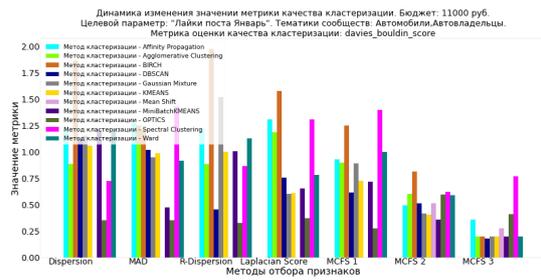


(c) Блок 3



(d) Блок 4

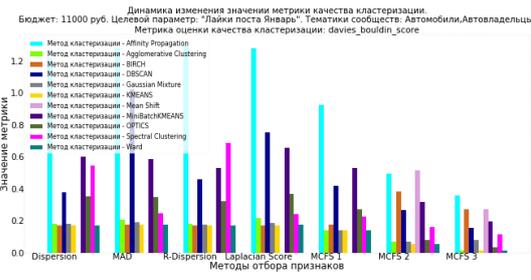
Рис. 2.22: Динамика изменения значений индекса «Калински-Харабаша»



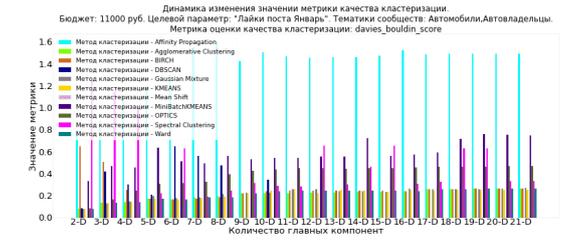
(a) Блок 1



(b) Блок 2

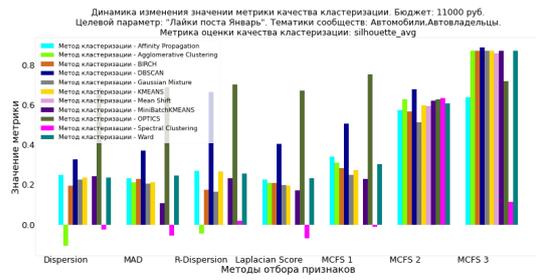


(c) Блок 3

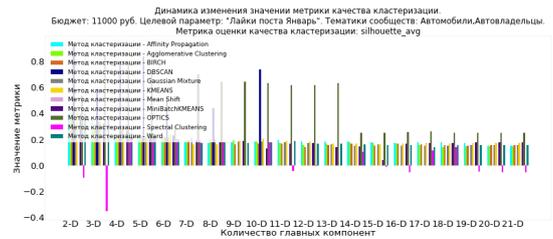


(d) Блок 4

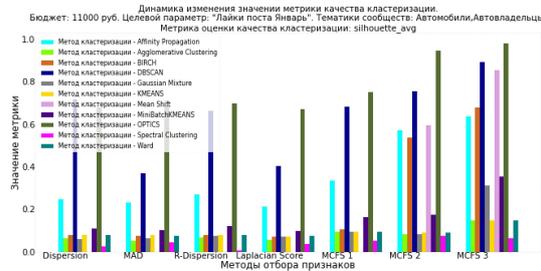
Рис. 2.23: Динамика изменения значений индекса «Дэвиса-Боулдина»



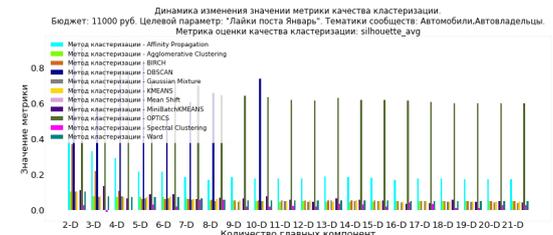
(a) Блок 1



(b) Блок 2



(c) Блок 3



(d) Блок 4

Рис. 2.24: Динамика изменения значений индекса «Силуэт»

**Замечание 2.3.** Динамика изменения значений целевого параметра в зависимости от бюджета показывает, что можно определять лучшие методы кластеризации для соответствующих способов отбора признаков и входных параметров (см. рис. 2.25, 2.26).

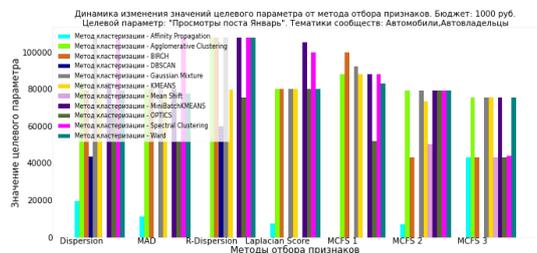
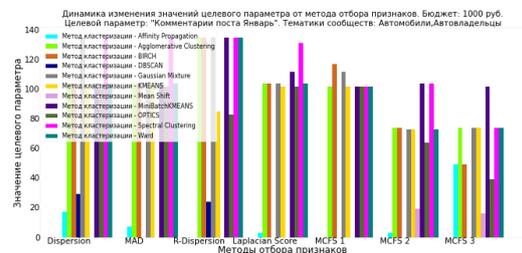
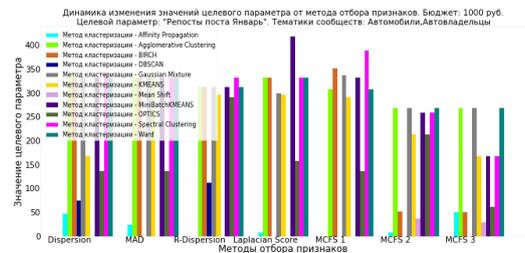
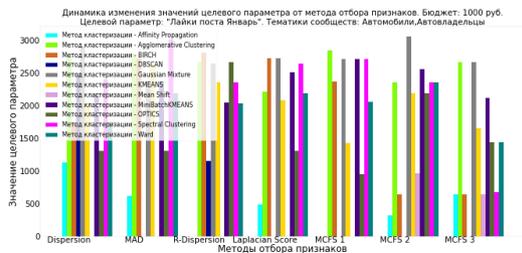


Рис. 2.25: Динамика изменения значений целевого параметра в зависимости от бюджета. Блок 1

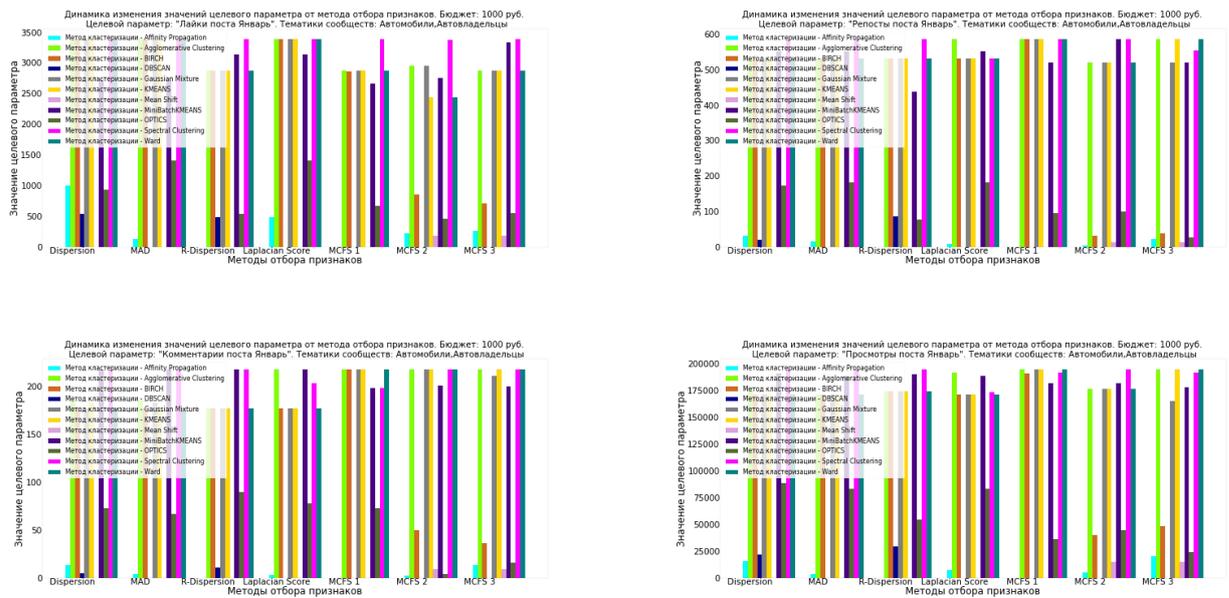


Рис. 2.26: Динамика изменения значений целевого параметра в зависимости от бюджета. Блок 3

Отметим, что данное замечание позволит сократить время для обучения моделей за счет выборки определенных методов кластерного анализа при различных условиях задачи.

**Замечание 2.4.** Динамика изменения количества главных компонент в зависимости от бюджета такова, что позволяет ограничить массив значений для сжатия признаковового пространства при соответствующих входных параметрах (см. рис. 2.27, 2.28).

Каждый столбец в приведенных диаграммах это то количество главных компонент, которое соответствует лучшему, в количественном отношении, численному результату для заданных параметров и используемых методов кластерного анализа. Таким образом, анализ данных визуальных представлений результатов моделирования позволит выдвигать гипотезы об ограничении значений главных компонент для формирования решения с некоторой точностью. Кроме того, это позволит определить лучшие значения для гиперпараметров методов кластеризации, а также существенно сократить время моделирования.

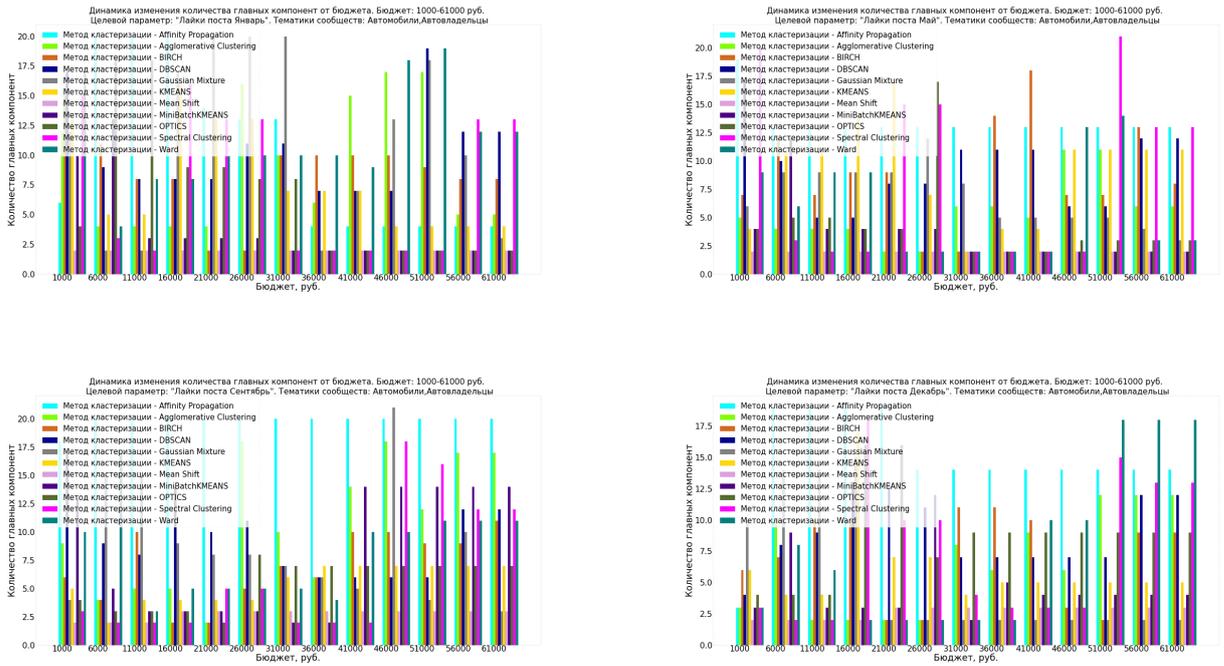


Рис. 2.27: Динамика изменения количества главных компонент от бюджета. Блок 2

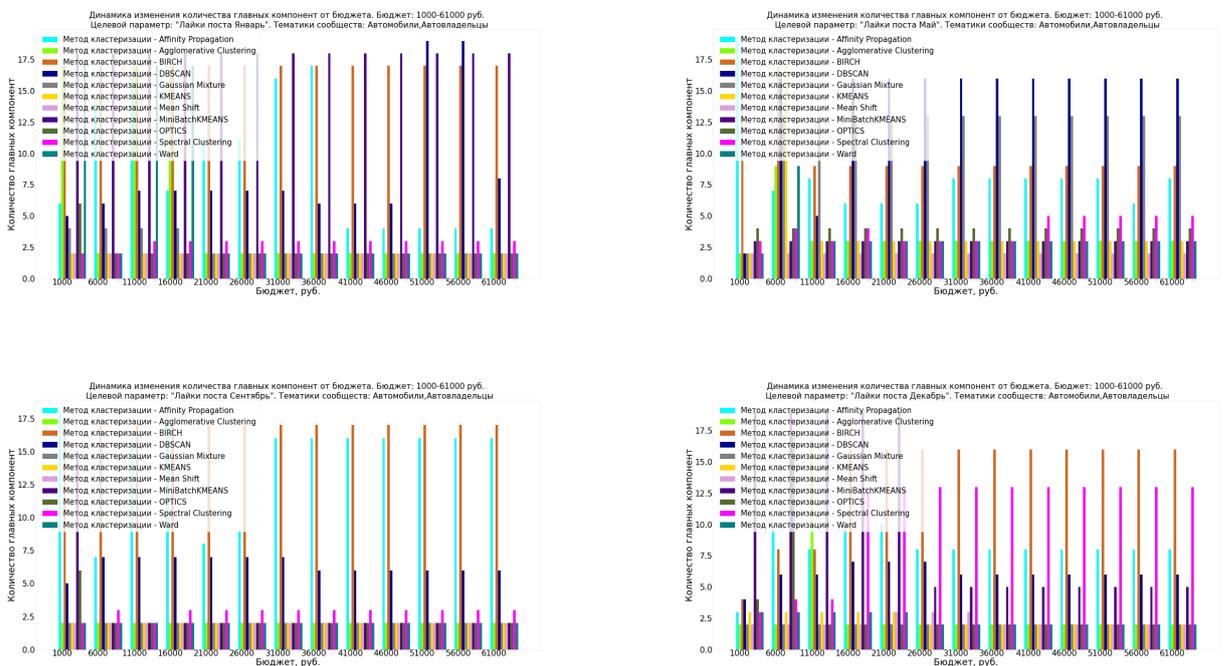


Рис. 2.28: Динамика изменения количества главных компонент от бюджета. Блок 4

## 2.4. Основные результаты и выводы по второй главе

В главе сформулированы постановки задач для моделирования процесса распространения информации в СМК с применением методов машинного обучения и оптимизации с предварительной кластеризацией для решения задачи снижения размерности и сокращения времени формирования сценариев распространения информации. Предложена архитектура и программно реализована комплексная модель с визуализацией, позволяющая формировать набор сообществ социальной сети с рекомендациями по размещению в них информации. Рассмотрены методы отбора и выделения признаков для задач обучения без учителя, приведены результаты применения указанных методов. Рассмотрены методы кластерного анализа и заданы их гиперпараметры, а также приведены метрики оценивания качества получаемых разбиений. Описана программная реализация блоков комплексной модели. Проведены обучение моделей кластеризации и сравнительный анализ результатов моделирования, который показал, что применение методов кластерного анализа позволяет решать задачу снижения размерности в оптимизационной задаче, что позволило сократить время формирования сценариев распространения информации. Применение разработанных программных компонент позволяет корректировать предпочтения пользователя, посредством настройки гиперпараметров методов машинного обучения. Проведение анализа важности признаков позволило определить базовый набор значимых характеристик объектов выбранными методами сжатия пространства признаков.

Таким образом, разработанная комплексная модель с методами машинного обучения представляет собой инструмент для сценарного моделирования, который позволяет формировать различные сценарии распространения информации. Использование данной комплексной модели с применением методов машинного обучения поможет управленческим структурам быть более гибкими и адаптивными при принятии решений в цифровой среде на основе смоделированных сценариев.

## Глава 3.

### Интеллектуальная система сценарного моделирования

#### 3.1. Проектирование интеллектуальной системы

Методы машинного обучения и искусственного интеллекта повсеместно интегрируют и адаптируют для применения в различных прикладных областях. Проектирование системы поддержки принятия управленческих решений [2, 36, 67, 91] как инструмента сценарного моделирования является ключевым этапом в создании эффективного и надежного инструмента для управления бизнес-процессами [16, 96, 104, 109]. Сценарное моделирование позволяет оценить возможные варианты развития событий, результаты принимаемых решений и анализировать их влияние на бизнес-процессы.

При проектировании интеллектуальной системы поддержки принятия управленческих решений с использованием сценарного моделирования важно также учитывать специфику бизнеса и особенности отрасли [10, 88]. Также необходимо определить методы и инструменты сценарного моделирования, которые будут использоваться в рамках системы. Это могут быть математические модели, аналитические инструменты, статистические методы и другие технологии, позволяющие проводить анализ данных и численное моделирование. При разработке данной системы были сформулированы постановки задач для определения набора площадок распространения информации (разделы 1.1 и 2.1). Отметим, что для анализа больших данных при формировании различных сценариев продвижения информации использовались методы оптимизации, а также методы машинного обучения без учителя.

Важным аспектом проектирования интеллектуальной системы является также обеспечение ее надежности, безопасности и масштабируемости. Система должна быть гибкой и адаптивной к изменяющимся условиям рынка и бизнес-среды, чтобы обеспечивать высокую эффективность принимаемых управленческих решений. Как отмечалось ранее, архитектура предлагаемого решения такова, что удовлетворяет перечисленным аспектам. Ниже на рисунке 3.1 пред-

ставлена схема интегрирования разработанной интеллектуальной системы для моделирования процесса распространения информации на примере социальных сетей.

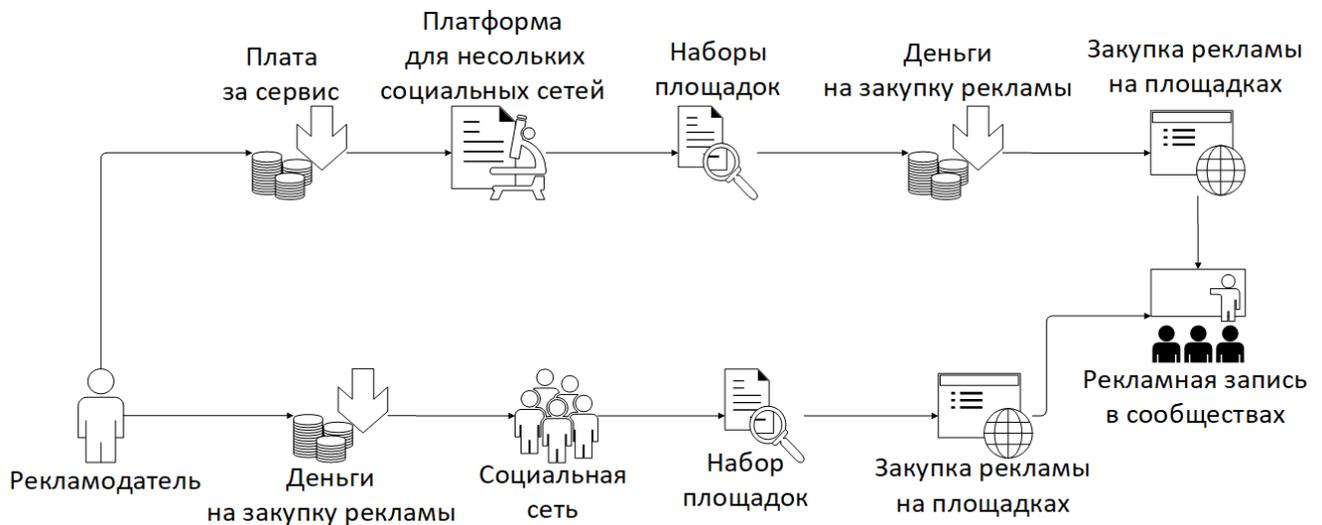


Рис. 3.1: Схема интегрирования интеллектуальной системы

В данном случае, с точки зрения целесообразности сотрудничества и возможной выгоды для каждой из сторон, предлагаемый подход является взаимовыгодным. Поскольку качество предоставляемой услуги растет, срок проведения рекламных кампаний сокращается, за счет осуществления анализа больших данных и построения рекомендаций для размещения записей в соответствующих сообществах. Отметим, что рекламодатели получают услугу надлежащего качества с привлечением меньшего количества ресурсов, а социальные сети дополнительный приток оборотных денежных средств. Внедрение такого инструмента возможно как на кросс-платформенной основе, так и для конкретной социальной сети, это будет выгодно как в первом, так и во втором случае. Отличие между этими вариантами реализации интеллектуальной системы заключается в функциональных возможностях и в том, куда будет обращаться рекламодатель напрямую к инструментам выбранной социальной сети или к платформе предоставляющей подобную услугу для нескольких интернет-площадок. Кроме того, на текущий момент времени в рассматриваемой социальной сети нет выбора различных сценариев продвижения информации, за вложенные деньги предлагается один набор площадок для публикации рекламной записи.

Проблема выбора сценариев для ЛПР заключается в необходимости оценить и анализировать различные варианты развития событий, чтобы принять обоснованное и эффективное решение. Также ЛПР сталкиваются с неопределенностью, сложностью и разнообразием факторов, которые могут повлиять на исход принимаемого решения. Выбор оптимального сценария требует глубокого понимания бизнес-процессов, анализа данных, прогнозирования результатов и учета рисков. Ошибочный выбор сценария может привести к нежелательным последствиям, потере времени, ресурсов и потере конкурентоспособности компании.

Для успешного принятия управленческих решений необходимо использовать инструменты сценарного моделирования, проводить анализ данных и учитывать возможные изменения во внешней среде. Важно иметь такой инструмент поддержки принятия решений, которая поможет ЛПР в выборе оптимального сценария и минимизации рисков. Следует также отметить, что ЛПР смогут оценивать различные сценарии продвижения информации и влияние их на ключевые показатели эффективности проведения рекламной кампании в сети Интернет до момента закупки рекламы и публикации записи в сообществах социальных сетей, что позволит более точно планировать и распределять бюджет при проведении рекламных кампаний. Разработанная интеллектуальная система как инструмент сценарного моделирования благодаря своей архитектуре позволяет решать поставленные задачи, может быть применена и адаптирована в различных предметных областях, масштабируема и, благодаря своей блочной структуре, легко изменяема в соответствие с выдвигаемыми требованиями по развитию ее функциональных возможностей. С архитектурой системы можно ознакомиться изучив рис. 3.2.



Рис. 3.2: Архитектура программного комплекса «Интеллектуальная система поддержки принятия управленческих решений» как инструмента сценарного моделирования

Кроме того заметим, что при разработке прототипа интеллектуальной системы для ЛПР используются методы машинного обучения, оптимизации и аналитики данных для обработки информации и предоставления пользователям-клиентам актуальной и достоверной информации, а также формируются рекомендации для принятия оптимальных стратегических и оперативных решений.

### 3.2. Реализация и особенности применения интеллектуальной системы

Реализация интеллектуальной системы поддержки принятия управленческих решений начинается с определения бизнес-целей и задач, которые необходимо решить с ее помощью. Затем происходит сбор и анализ данных, необходимых для функционирования системы. После этого происходит настройка и обучение алгоритмов машинного обучения для работы с конкретными данными и поставленными задачами, а также проводится анализ на чувствительность при определении предпочтений рекламодателя в задаче многокритериальной оптимизации. Подробнее компоненты из которых состоит система описаны в предыдущих главах и разделах.

Разработанная интеллектуальная система поддержки принятия управленческих решений (ИСППУР) представляет собой инновационный инструмент, который объединяет в себе возможности искусственного интеллекта, аналитики данных и бизнес-процессов для помощи руководителям и менеджерам в принятии обоснованных и эффективных решений. Особенности применения ИСППУР включают:

- *Автоматизация процесса принятия решений:* ИСППУР предоставляет возможность автоматизировать часть процесса принятия управленческих решений, основываясь на анализе данных, моделировании сценариев (прогнозировании результатов). Это позволяет сократить время на принятие решений и уменьшить вероятность появления ошибок.
- *Анализ больших объемов данных:* ИСППУР способен обрабатывать и анализировать большие объемы данных из различных источников, выявляя скрытые паттерны и зависимости. Это помогает руководителям принимать информированные решения на основе фактов и аналитики.
- *Персонализация рекомендаций:* ИСППУР способен создавать персонализированные рекомендации и предложения на основе индивидуальных потребностей и целей каждого руководителя. Это позволяет учитывать уни-

кальные особенности бизнеса и принимать решения, соответствующие конкретным обстоятельствам.

- *Мониторинг и оценка результатов:* ИСППУР предоставляет возможность мониторинга и оценки результатов принятых решений, а также анализа их эффективности. Это позволяет корректировать стратегии действий в реальном времени и оптимизировать бизнес-процессы.
- *Интеграция с другими системами:* ИСППУР может быть легко интегрирован с другими информационными системами организации, такими как CRM, ERP и BI-системы, что обеспечивает единый доступ к данным и повышает эффективность работы всей компании.
- *Масштабируемость:* система может быть легко масштабирована для работы с различными объемами данных и задачами, что делает ее универсальным инструментом для различных организаций и отраслей.

Таким образом, применение данной ИСППУР позволяет организациям повысить качество принимаемых решений, оптимизировать бизнес-процессы и достичь конкурентных преимуществ на рынке при проведении мероприятий по распространению информации в сети Интернет, а также позволяет организациям-рекламодателям повысить эффективность управления, минимизировать риски и ошибки, а также принимать обоснованные стратегические и оперативные решения на основе данных и аналитики.

Разработанный прототип интеллектуальной системы состоит из 6 программных компонент, каждый из которых может работать автономно и предоставлять сценарии для продвижения информации [34, 45, 52]. Исходя из потребностей клиента будет использовано необходимое количество программных компонент из всего программного комплекса для получения решения в короткий срок. Рассмотрим более подробно схему реализации и функционирования прототипа системы.

Отметим, что для обеспечения устойчивости работы системы при внешних сбоях было решено реализовать программный комплекс состоящий из 6 программных блоков это:

1. Программная реализация циклического алгоритма предобработки статистических данных о пользовательской активности информационных площадок в задаче распространения информации в СМК. Он реализован в виде одного файла с расширением «.irunb» для удобства внесения изменений при написании программного кода.
2. Программная реализация оптимизационной модели, включающая в себя решение задачи многокритериальной оптимизации. Модель был реализована так, что состоит из 6 файлов: 2 из которых с расширением «.ру», использующиеся как библиотеки для обработки данных и формирования рекомендаций; 1 с расширением «.irunb» и используется для проведения численного моделирования; оставшиеся 3 файла с расширением «.irunb» необходимы для визуализации, анализа и интерпретации результатов.
3. Программная реализация методов кластерного анализа с применением интерпретируемых методов отбора признаков. Блок состоит из 15 файлов: 11 с расширением «.ру», использующиеся как библиотеки и реализующие методы кластеризации, 2 с расширением «.ру», использующиеся как библиотеки для обработки данных и формирования рекомендаций, 1 с расширением «.irunb» для проведения численного моделирования, 1 с расширением «.irunb» для визуализации, анализа и интерпретации результатов.
4. Программная реализация методов кластерного анализа с применением метода главных компонент для сжатия пространства признаков. Структура аналогична п.3.
5. Программная реализация оптимизационной модели с проведением предварительной кластеризации и использованием интерпретируемых методов отбора признаков. Структура аналогична п.3.

6. Программная реализация оптимизационной модели с проведением предварительной кластеризации и использованием метода главных компонент для сжатия признакового пространства. Структура аналогична п.3.

Далее, рассмотрим разработанную структуру хранения результатов моделирования. Разработка подобных структур является важным этапом в процессе работы с данными и аналитикой. Эта система позволяет эффективно сохранять, управлять и обрабатывать результаты моделирования, что имеет ряд преимуществ:

- *Сохранение результатов:* предлагаемая структура хранения позволяет сохранять все результаты моделирования в структурированном и удобном для дальнейшего доступа виде. Это обеспечивает возможность повторного использования результатов, анализа и сравнения различных моделей.
- *Удобство доступа и обмена:* предлагаемая структура хранения позволяет легко получать доступ к результатам моделирования, делиться ими с коллегами и другими участниками проекта. Это способствует совместной работе и обмену знаниями.
- *Улучшение процесса принятия решений:* доступ к сохраненным результатам моделирования помогает принимать обоснованные решения на основе данных и аналитики. Анализ предыдущих результатов может помочь определить тренды и оптимальные стратегии действий.
- *Повышение эффективности работы:* предлагаемая структура хранения результатов моделирования упрощает процессы управления данными, сокращает время на поиск нужной информации и повышает общую эффективность работы команды.

Таким образом, разработка структуры хранения результатов моделирования играет ключевую роль в успешной работе с данными и аналитикой, обеспечивая безопасность, удобство доступа к информации и повышение эффективности процессов принятия управленческих решений.

Исходя из вышеизложенного, было принято решение сохранять данные по блокам в форматах «.csv» и «.xlsx». Так, например, в блоке 2 данные хранятся по соответствующим тематикам, в которых есть вложенные папки по месяцам года, где созданы еще 2 папки для исследователей и для клиентов. Последние отличаются набором столбцов. А если рассматривать блоки 3-6, то в этих случаях будет гораздо больше различной информации, которую придется хранить и обрабатывать.

Возьмем блок с номером 5, в нем создана папка с названием «Результаты моделирования», где хранятся выходные сводные табличные данные для соответствующих тематик, в каждой имеется следующая структура хранения как на рисунке 3.3.

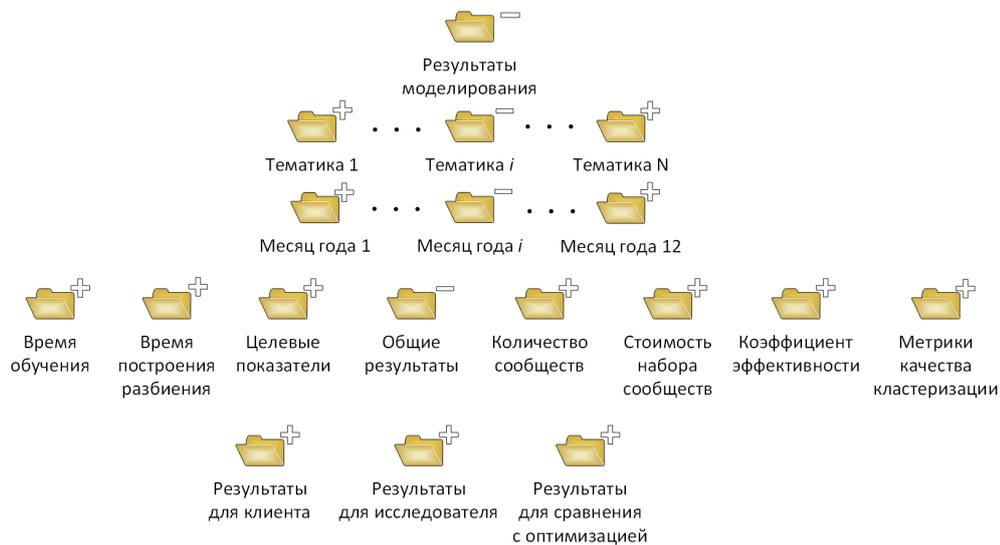


Рис. 3.3: Схема структуры хранения данных

Каждая папка лежащая находящаяся в «Месяц года  $i$ » имеет 11 файлов формата либо «.csv», либо «.xlsx», количество которых равно количеству применяемых методов кластерного анализа. Отметим, что в папке «Общие результаты» сделано разделение еще на 3 раздела, отличающихся друг от друга набором столбцов в файлах. Приведем примеры структуры файлов каждой папки они представлены на соответствующих рисунках 3.4, 3.5, 3.6.





### 3.3. Сравнительный анализ результатов моделирования

В рамках диссертационного исследования были рассмотрены результаты моделирования, проанализированы сценарии предложенные системой и сделаны соответствующие выводы в рамках поставленных задач. Однако следует отметить, что программный компонент «Модель кластерного анализа» может быть применен и для других постановок задач, где требуется не столько максимизировать значения целевого показателя при заданных ограничениях бюджета для некоторого перечня информационных площадок, сколько получить такой набор, в котором эти площадки будут объединены по определенным принципам, а выбор лучшего набора будет осуществляться по разработанной системе правил. В этом смысле, рассматриваемый программный комплекс обладает свойством масштабируемости, описанным в предыдущем разделе. Кроме того, методы кластерного анализа могут быть использованы не только как отдельный инструмент для формирования решения, но и как метод для понижения размерности в задаче оптимизации.

В данном параграфе будут рассматриваться результаты моделирования представленные в виде сводных таблиц. Сводные таблицы играют важную роль в анализе данных и представлении информации. Они позволяют суммировать, группировать и агрегировать данные из различных источников, что помогает быстро и эффективно выявлять закономерности, а также сравнивать различные параметры. Таким образом, использование сводных таблиц является необходимым инструментом для эффективного анализа данных и принятия обоснованных решений.

Для примера возьмем следующие входные параметры:

1. Тематики: «Автомобили, Автовладельцы»;
2. Месяц: Январь;
3. Целевые параметры: Лайки, Репосты, Комментарии, Просмотры.

Было проведено численно моделирования по заданным параметрам. Перейдем к проведению сравнительного анализа различных сценариев продвижения информации. Ниже представлены 4 таблицы, которые позволяют оценить численные результаты и скорость их формирования. Отметим, что обоснование анализа результатов моделирования является ключевым элементом в процессе принятия решений на основе данных, это помогает убедиться в правильности выводов, сделанных на основе моделирования, и обеспечивает основу для разработки стратегий и дальнейших действий на основе полученных результатов.

Номер программного блока	Лайки поста	Репосты поста	Комментарии поста	Просмотры поста	Бюджет клиента	Общая стоимость	Количество сообщений	Время работы оптимизации	Время работы кластеризации	Время обучения
2	3392	334	81	111452	1000	1000	10	0 days 00:00:00.004001		
3	3067	107	46	58642	1000	974	8		0 days 00:00:00.012003	0 days 00:00:03.708894
4	3130	94	44	55888	1000	974	8		0 days 00:00:00.003001	0 days 00:00:02.141678
5	3392	334	81	111452	1000	1000	10	0 days 00:00:00.003001	0 days 00:00:00.105024	0 days 00:00:14.424262
6	3201	117	69	81774	1000	986	11	0 days 00:00:00.002000	0 days 00:00:00.106024	0 days 00:00:14.410813
2	9972	1140	475	529371	6000	5998	26	0 days 00:00:00.012003		
3	8123	559	240	280543	6000	5954	31		0 days 00:00:00.101024	0 days 00:00:07.766416
4	7281	494	168	236201	6000	5451	25		0 days 00:00:00.001000	0 days 00:00:01.664153
5	9972	1140	475	529371	6000	5998	26	0 days 00:00:00.004095	0 days 00:00:00.001000	0 days 00:00:10.580808
6	9449	1118	467	522588	6000	5882	25	0 days 00:00:00.003000	0 days 00:00:00.015003	0 days 00:00:17.013147
2	12387	1804	688	783894	11000	10992	44	0 days 00:00:00.007002		
3	11481	1587	628	652937	11000	10217	37		0 days 00:00:00.079019	0 days 00:00:09.455923
4	9339	924	288	378373	11000	8543	37		0 days 00:00:00.038009	0 days 00:00:03.947927
5	12387	1804	688	783894	11000	10992	44	0 days 00:00:00.004000	0 days 00:00:00.000001	0 days 00:00:16.665632
6	11693	1800	718	789461	11000	10640	43	0 days 00:00:00.003001	0 days 00:00:00.015004	0 days 00:00:17.013322
2	13621	2150	799	965854	16000	15991	52	0 days 00:00:00.010002		
3	12134	2083	852	884869	16000	15481	45		0 days 00:00:00.004059	0 days 00:00:04.073123
4	10407	1641	564	817133	16000	15945	28		0 days 00:00:00.050012	0 days 00:00:09.520101
5	13621	2150	799	965854	16000	15991	52	0 days 00:00:00.004001	0 days 00:00:00.001000	0 days 00:00:14.874318
6	12831	2059	818	913892	16000	15365	58	0 days 00:00:00.010002	0 days 00:00:00.016003	0 days 00:00:20.773879

Рис. 3.8: Сводная таблица результатов. Тематики: «Автомобили, Автовладельцы». Месяц: Январь. Целевой параметр: Лайки

Номер программного блока	Лайки поста	Репосты поста	Комментарии поста	Просмотры поста	Бюджет клиента	Общая стоимость	Количество сообщений	Время работы оптимизации	Время работы кластеризации	Время обучения
2	1597	587	129	161250	1000	987	6	0 days 00:00:00.002000		
3	1043	419	112	105276	1000	699	4		0 days 00:00:00.045011	0 days 00:00:01.292892
4	1058	442	121	121012	1000	915	6		0 days 00:00:00.005002	0 days 00:00:03.508108
5	1597	587	129	161250	1000	987	6	0 days 00:00:00.001000	0 days 00:00:00.000001	0 days 00:00:05.016178
6	1597	587	129	161250	1000	987	6	0 days 00:00:00.002000	0 days 00:00:00.106024	0 days 00:00:13.750781
2	6698	1736	604	660727	6000	5991	25	0 days 00:00:00.002000		
3	5126	1271	482	467840	6000	5847	10		0 days 00:00:00.006002	0 days 00:00:00.584957
4	3739	1052	356	422726	6000	5424	5		0 days 00:00:00.107026	0 days 00:00:08.179697
5	6343	1727	595	652826	6000	6000	23	0 days 00:00:00.002000	0 days 00:00:00.109025	0 days 00:00:24.606837
6	5957	1728	594	653051	6000	5954	23	0 days 00:00:00.001000	0 days 00:00:00.102023	0 days 00:00:22.498651
2	8374	2277	706	909412	11000	10992	33	0 days 00:00:00.004001		
3	5337	1811	577	662371	11000	10074	16		0 days 00:00:00.100709	0 days 00:00:00.247215
4	4756	1651	615	675489	11000	10415	12		0 days 00:00:00.089021	0 days 00:00:08.092901
5	8441	2267	707	899505	11000	10980	32	0 days 00:00:00.003000	0 days 00:00:00.108025	0 days 00:00:33.718610
6	7214	2254	702	879700	11000	10995	31	0 days 00:00:00.004001	0 days 00:00:00.106024	0 days 00:00:28.992862
2	10359	2674	850	1087611	16000	15989	47	0 days 00:00:00.003001		
3	12134	2083	852	884869	16000	15481	45		0 days 00:00:00.004059	0 days 00:00:04.073123
4	5869	1953	614	816331	16000	14597	12		0 days 00:00:00.008002	0 days 00:00:00.590138
5	10771	2657	843	1090513	16000	15991	43	0 days 00:00:00.003001	0 days 00:00:00.106024	0 days 00:00:36.735366
6	9049	2638	814	1081554	16000	15991	35	0 days 00:00:00.003001	0 days 00:00:00.104024	0 days 00:00:32.382823

Рис. 3.9: Сводная таблица результатов. Тематики: «Автомобили, Автовладельцы». Месяц: Январь. Целевой параметр: Репосты

Номер программного блока	Лайки поста	Репосты поста	Комментарии поста	Просмотры поста	Бюджет клиента	Общая стоимость	Количество сообщений	Время работы оптимизации	Время работы кластеризации	Время обучения
2	875	365	219	128951	1000	1000	7	0 days 00:00:00.003001		
3	385	312	135	107877	1000	944	6		0 days 00:00:00.039010	0 days 00:00:08.175789
4	634	390	202	131086	1000	944	5		0 days 00:00:00.041010	0 days 00:00:03.814897
5	871	357	218	123571	1000	974	6	0 days 00:00:00.001999	0 days 00:00:00.108025	0 days 00:00:16.061855
6	871	357	218	123571	1000	974	6	0 days 00:00:00.002000	0 days 00:00:00.105024	0 days 00:00:13.909219
2	6554	1494	757	655108	6000	5991	27	0 days 00:00:00.002001		
3	2838	924	529	416453	6000	5855	11		0 days 00:00:00.177775	0 days 00:00:09.330109
4	3739	1052	356	422726	6000	5424	5		0 days 00:00:00.107026	0 days 00:00:08.179697
5	6910	1485	756	646598	6000	5994	26	0 days 00:00:00.002001	0 days 00:00:00.109025	0 days 00:00:24.606837
6	5766	1396	751	644514	6000	5964	25	0 days 00:00:00.001000	0 days 00:00:00.001000	0 days 00:00:10.641500
2	9105	1799	952	900895	11000	10971	47	0 days 00:00:00.002001		
3	11409	1589	649	659514	11000	10649	38		0 days 00:00:00.002000	0 days 00:00:02.626978
4	4756	1651	615	675489	11000	10415	12		0 days 00:00:00.089021	0 days 00:00:08.092901
5	8639	1770	944	883800	11000	10837	44	0 days 00:00:00.002001	0 days 00:00:00.001001	0 days 00:00:16.349813
6	8027	1671	929	846029	11000	10564	42	0 days 00:00:00.002001	0 days 00:00:00.001001	0 days 00:00:13.154460
2	10964	2118	1073	1120841	16000	15976	61	0 days 00:00:00.002000		
3	12134	2083	852	884869	16000	15481	45		0 days 00:00:00.004059	0 days 00:00:04.073123
4	5122	1798	664	767519	16000	12431	13		0 days 00:00:00.092022	0 days 00:00:08.104014
5	9607	2111	1051	1081789	16000	15481	51	0 days 00:00:00.000999	0 days 00:00:00.001001	0 days 00:00:17.544121
6	9302	2048	1044	1071474	16000	15575	48	0 days 00:00:00.001001	0 days 00:00:00.001001	0 days 00:00:14.685395

Рис. 3.10: Сводная таблица результатов. Тематики: «Автомобили, Автовладельцы». Месяц: Январь. Целевой параметр: Комментарии

Номер программного блока	Лайки поста	Репосты поста	Комментарии поста	Просмотры поста	Бюджет клиента	Общая стоимость	Количество сообщений	Время работы оптимизации	Время работы кластеризации	Время обучения
2	1382	537	148	194726	1000	992	10	0 days 00:00:00.002000		
3	385	312	135	107877	1000	944	6		0 days 00:00:00.039010	0 days 00:00:08.175789
4	634	390	202	131086	1000	944	5		0 days 00:00:00.041010	0 days 00:00:03.814897
5	1382	537	148	194726	1000	992	10	0 days 00:00:00.000999	0 days 00:00:00.107025	0 days 00:00:16.061855
6	1382	537	148	194726	1000	992	10	0 days 00:00:00.001000	0 days 00:00:00.106024	0 days 00:00:13.909219
2	5917	1539	662	746262	6000	5990	27	0 days 00:00:00.002001		
3	5126	1271	482	467840	6000	5847	10		0 days 00:00:00.006002	0 days 00:00:00.584957
4	3739	1052	356	422726	6000	5424	5		0 days 00:00:00.107026	0 days 00:00:08.179697
5	5814	1534	661	743160	6000	5983	28	0 days 00:00:00.001000	0 days 00:00:00.108026	0 days 00:00:22.318820
6	5917	1539	662	746262	6000	5990	27	0 days 00:00:00.001000	0 days 00:00:00.104024	0 days 00:00:24.078659
2	9650	2019	822	1010102	11000	11000	44	0 days 00:00:00.004001		
3	5337	1811	577	662371	11000	10074	16		0 days 00:00:00.100709	0 days 00:00:00.247215
4	6509	1394	499	678536	11000	10537	21		0 days 00:00:00.106025	0 days 00:00:09.774296
5	7007	1888	801	994974	11000	10978	40	0 days 00:00:00.003013	0 days 00:00:00.158036	0 days 00:00:27.603781
6	8509	1986	816	1001131	11000	11000	42	0 days 00:00:00.002003	0 days 00:00:00.001001	0 days 00:00:14.717530
2	10064	2480	922	1235878	16000	15990	46	0 days 00:00:00.002000		
3	12134	2083	852	884869	16000	15481	45		0 days 00:00:00.004059	0 days 00:00:04.073123
4	5437	1613	607	849227	16000	14259	13		0 days 00:00:00.112027	0 days 00:00:08.628373
5	8438	2488	904	1215932	16000	15961	48	0 days 00:00:00.001001	0 days 00:00:00.105185	0 days 00:00:33.014212
6	9056	2401	883	1186803	16000	15338	42	0 days 00:00:00.001001	0 days 00:00:00.104024	0 days 00:00:32.382823

Рис. 3.11: Сводная таблица результатов. Тематики: «Автомобили, Автовладельцы». Месяц: Январь. Целевой параметр: Просмотры

Структура таблиц такова, что позволяет сравнивать сценарии различных программных блоков. В таблицах представлена не вся информация доступная для лиц принимающих решения, однако достаточная для проведения анализа применения рассматриваемых моделей и методов, как с точки зрения временных затрат на построение решения, так и сравнения численных значений различных параметров. Используя данные из представленного примера, сформулированы соответствующие замечания и выводы.

**Вывод 3.1.** *В полученных сценариях продвижения информации по заданным бюджетам лучший результат демонстрируют программные блоки, где применяются методы оптимизации.*

Действительно, оптимизационный подход позволяет точно решить поставленную задачу. Однако, если обратить внимание на таблицу на рисунке 3.8, где бюджет равен 6000 рублей, и сравнить результаты блоков 2 и 3, то есть оптимизации и кластеризации с применением интерпретируемых методов отбора признаков, то значения целевых показателей этих двух сценариев будут отличаться на 18,54% в пользу блока 2. Но если сравнить количество сообществ в которых будет размещена запись, то окажется, что в блоке 3 таких на 5 штук больше, что может быть принципиально для лица принимающего решение.

Аналогично рассмотрим сценарии представленные на рисунке 3.11, а именно бюджет 11000 рублей, блоки 2 и 3. Здесь значения целевых показателей этих двух сценариев будет отличаться на 34,42% в пользу блока 2, однако количество сообществ больше в 2,75 раза. Это говорит о том, что данный набор площадок, который содержит 36,36% сообществ от оптимального сценария, дает 65,58% результата, при условии, что имеется остаток от использованного бюджета в 926 рублей, в отличии от блока 2, где бюджет освоен полностью. Это, в том числе, может указывать на существенную разницу показателей пользовательской активности в различных сценариях.

Таким образом, сценарное моделирование помогает принимать более обоснованные решения, основанные на моделировании и анализе данных, оценивать возможные риски и выгоды различных стратегий. Кроме того, исследователи могут проводить симуляции, чтобы оценить вероятные последствия различных действий. Такой подход помогает сформировать и учитывать различные варианты развития событий и выбирать наиболее эффективные решения.

Поэтому при проектировании данной интеллектуальной системы учитывалась возможность анализа получаемых решений по различным аспектам, в том числе по соотношению суммы значений соответствующих целевых показателей на единицу вложенных денежных средств (см. рис. 3.12).

Анализ такого рода данных помогает оценивать различные наборы сообщений и принимать решения о размещении публикации.

	Dispersion	MAD	R-Dispersion	Laplacian Score	MCFS 1	MCFS 2	MCFS 3	Метод кластеризации	Бюджет клиента	Выбранные тематики	Целевой параметр	Месяц
0	3,79527559	3,34257206	2,91707921	3,400513479	2,92481203	2,47794118	1,89166667	Spectral Clustering	1000	Автомобили,Автовладельцы	Лайки поста	Январь
1	0,35233161	0,42045455	0,42045455	0,420454545	0,42190889	0,51190476	2,31944444	Spectral Clustering	1000	Автомобили,Автовладельцы	Репосты поста	Январь
2	0,14300847	0,14300847	0,14300847	0,16375	0,10725552	0,20634921	0,08286674	Spectral Clustering	1000	Автомобили,Автовладельцы	Комментарии поста	Январь
3	114,276483	114,276483	114,276483	124,5825	101,855324	84,6517094	76,2534722	Spectral Clustering	1000	Автомобили,Автовладельцы	Просмотры поста	Январь
4	1,48184818	1,03250658	1,70389232	0,981002742	1,22746579	1,36429291	1,36327053	Spectral Clustering	6000	Автомобили,Автовладельцы	Лайки поста	Январь
5	0,18307463	0,24784483	0,22247254	0,30260599	0,18802041	0,21450713	0,23802203	Spectral Clustering	6000	Автомобили,Автовладельцы	Репосты поста	Январь
6	0,06166277	0,07093569	0,17430556	0,067037954	0,08350896	0,08472825	0,09035013	Spectral Clustering	6000	Автомобили,Автовладельцы	Комментарии поста	Январь
7	70,4142037	87,7269089	96,4498794	74,24381188	96,3227513	79,0028932	71,1277541	Spectral Clustering	6000	Автомобили,Автовладельцы	Просмотры поста	Январь
8	1,48184818	1,03250658	1,70389232	0,981002742	1,22746579	1,04909718	1,12371538	Spectral Clustering	11000	Автомобили,Автовладельцы	Лайки поста	Январь
9	0,18307463	0,10907424	0,22247254	0,110435758	0,12908096	0,18537524	0,15062916	Spectral Clustering	11000	Автомобили,Автовладельцы	Репосты поста	Январь
10	0,03554549	0,03858845	0,03554549	0,039274359	0,04425363	0,06137147	0,06146618	Spectral Clustering	11000	Автомобили,Автовладельцы	Комментарии поста	Январь
11	41,3368532	56,4137489	42,5061242	57,36525154	59,600585	70,517105	63,9069198	Spectral Clustering	11000	Автомобили,Автовладельцы	Просмотры поста	Январь
12	1,48184818	1,03250658	1,70389232	0,442118432	0,53541638	0,84613592	1,12371538	Spectral Clustering	16000	Автомобили,Автовладельцы	Лайки поста	Январь
13	0,18307463	0,11204867	0,0728773	0,113210075	0,1187941	0,15215751	0,15062916	Spectral Clustering	16000	Автомобили,Автовладельцы	Репосты поста	Январь
14	0,03554549	0,0390625	0,03554549	0,041791252	0,03601073	0,05219417	0,06146618	Spectral Clustering	16000	Автомобили,Автовладельцы	Комментарии поста	Январь
15	33,9881523	56,2868917	36,4579869	56,04642283	54,4480909	60,266898	63,9069198	Spectral Clustering	16000	Автомобили,Автовладельцы	Просмотры поста	Январь
16	1,48184818	0,29225077	1,70389232	0,442118432	0,60711438	0,84613592	0,65365132	Spectral Clustering	21000	Автомобили,Автовладельцы	Лайки поста	Январь
17	0,0522259	0,11204867	0,06731946	0,113210075	0,1187941	0,1381416	0,12578974	Spectral Clustering	21000	Автомобили,Автовладельцы	Репосты поста	Январь
18	0,03554549	0,03350878	0,02096083	0,041791252	0,03765716	0,04389654	0,0413797	Spectral Clustering	21000	Автомобили,Автовладельцы	Комментарии поста	Январь
19	31,8731688	47,4596904	36,9070787	56,04642283	55,1034652	58,2426407	52,1630144	Spectral Clustering	21000	Автомобили,Автовладельцы	Просмотры поста	Январь
20	0,44777373	0,29225077	1,70389232	0,442118432	0,60711438	0,84613592	0,65365132	Spectral Clustering	26000	Автомобили,Автовладельцы	Лайки поста	Январь
21	0,06300756	0,11204867	0,06731946	0,113210075	0,1187941	0,11275791	0,12578974	Spectral Clustering	26000	Автомобили,Автовладельцы	Репосты поста	Январь
22	0,01884226	0,03350878	0,02096083	0,041791252	0,03765716	0,03653482	0,0413797	Spectral Clustering	26000	Автомобили,Автовладельцы	Комментарии поста	Январь
23	36,0494459	47,4596904	36,9070787	56,04642283	55,1034652	51,3689522	52,1630144	Spectral Clustering	26000	Автомобили,Автовладельцы	Просмотры поста	Январь

Рис. 3.12: Таблица коэффициентов эффективности по целевым показателям

**Вывод 3.2.** В рассматриваемых примерах процедура предварительной кластеризации для понижения размерности в задаче оптимизации позволяет формировать сценарии продвижения информации не позднее, чем с использованием модели оптимизации и с минимальными отклонением в значениях целевых параметрах.

Применение такого подхода позволит, при больших размерностях задачи оптимизации, существенно уменьшить время работы алгоритма при условии, что в количественном отношении будет получено решение эквивалентное оптимальному или вовсе такое же, в силу применения соответствующего метода машинного обучения для построения разбиения. Наличие большого количества объектов в выбранной(ых) тематике(ах) является актуальной проблемой при проведении больших рекламных кампаний.

**Замечание 3.1.** В приведенных примерах при увеличении количества объектов продемонстрирована необходимость применения предварительной разметки данных в задаче оптимизации.

Отметим, что разница в значениях целевых показателей между блоками 2, 5, 6 если и есть, то находится в пределах 5-10% от оптимальных значений, при этом скорость выполнения процедуры поиска решения в задаче оптимизации значительно уменьшается. Возьмем для примера блоки 2 и 5 из таблиц, представленных на рисунках 3.9, 3.10, и сравним, в первом случае сценарии в бюджете 11000 рублей (см. рис. 3.13), во втором 16000 (см. рис. 3.14).

<b>Блок 2</b>		<b>Блок 5</b>	
	Название сообщества		Название сообщества
0	AUDI	0	AUDI
1	AUTO	1	AUTO
2	Deutsche Autos	2	Best Cars - Авто клуб
3	JDM	3	JDM
4	JDM РФ	4	JDM РФ
5	LOW BASS TEAM   Автозвук	5	LOW BASS TEAM   Автозвук
6	Men`s Academy	6	Men`s Academy
7	Nissan Club	7	Nissan Club
8	Vazzz i Basss™ (18+)   Бот	8	TURBO GARAGE
9	[В] Дороге   АВТО	9	Vazzz i Basss™ (18+)   Бот
10	АВТОБУГУРТ	10	АВТОБУГУРТ
11	АВТОРЫНОК ПЕРМЬ КУПИТЬ АВТО С ПРОБЕГОМ	11	Авто приколы
12	Авто приколы	12	Авторынок   Орел   Орловская область
13	Авторынок   Орел   Орловская область	13	Авторынок ДНР-ЛНР (Донецк, Макеевка, Харцызск)
14	Авторынок   Чебоксары   Чувашия	14	Авторынок Корчей
15	Авторынок ДНР-ЛНР (Донецк, Макеевка, Харцызск)	15	Авторынок Крым (Симферополь, Севастополь)
16	Авторынок Корчей	16	Боевой Корч
17	Боевой Корч	17	Видеопродажа АВТО СВЕРДЛОВСКАЯ ОБЛАСТЬ
18	Видеопродажа АВТО СВЕРДЛОВСКАЯ ОБЛАСТЬ	18	Главная Дорога
19	Главная Дорога	19	ДТП Кострома
20	ДТП Кострома	20	Дорожные войны!
21	Дорожные войны!	21	МИР ГРУЗОВИКОВ И ВНЕДОРОЖНИКОВ
22	МИР ГРУЗОВИКОВ И ВНЕДОРОЖНИКОВ	22	Музыка в машину
23	Музыка в машину	23	Немецкие хищники
24	Немецкие хищники	24	Продажа Авто   Авто Подбор
25	Продажа Авто   Авто Подбор	25	Тачка за сотку
26	Тачка за сотку	26	Тачка за сотку   Урал
27	Тачка за сотку   Урал	27	Тачки за копейки
28	Тачки за копейки	28	Тачки за копейки'77   Москва
29	Тачки за копейки'77   Москва	29	ЭТО ДРИФТ, ДЕТКА
30	ЭТО ДРИФТ, ДЕТКА	30	Японские пушки   JDM
31	Японские пушки   JDM	31	министерство оперов
32	министерство оперов		

Рис. 3.13: Сообщества из решений для бюджета 11000 рублей. Целевой параметр: Репосты

Блок 2			Блок 5				
Название сообщества	Название сообщества	Название сообщества	Название сообщества	Название сообщества			
0	+300 км/ч   Автожурнал	35	Видеопродажа АВТО СВЕРДЛОВСКАЯ ОБЛАСТЬ	0	Best Cars - Авто клуб	35	За рулём
1	AUDI	36	Внедорожники	1	JDM	36	Клуб Автомобилистов
2	AVTO ZONA 163 Самара Тольятти Авто Барахолка	37	ГАРАЖ   АВТО   GIF	2	JDM   ДРИФТ	37	Лада Веста & Lada Vesta
3	Best Cars - Авто клуб	38	Гараж	3	JDM РФ	38	Лучшие автомобили
4	JDM	39	Главная Дорога	4	Life 4x4: полный привод!	39	Любители Авто
5	JDM   ДРИФТ	40	Главная Дорога: Авто / ДТП	5	MQ: Men's Quality	40	МИР ГРУЗОВИКОВ И ВНЕДОРОЖНИКОВ
6	JDM РФ	41	ДТП Кострома	6	Nissan Club	41	Музыка в машину
7	Life 4x4: полный привод!	42	Дальнбойщики	7	Off-Road   Мир Внедорожников	42	Немецкие хищники
8	MQ: Men's Quality	43	Дорожные войны!	8	[ Клуб Внедорожников Уаз 4x4 ]™	43	Тачка за сотку
9	Nissan Club	44	За рулём	9	[В] Дороге   АВТО	44	Тачка за сотку   Урал
10	Off-Road   Мир Внедорожников	45	Лада Веста & Lada Vesta	10	АВТОБУГУРТ	45	Тачки за копейки
11	[ Клуб Внедорожников Уаз 4x4 ]™	46	Лучшие автомобили	11	АВТОМЕХАНИК ОТ БОГА	46	Тачки за копейки'77   Москва
12	[В] Дороге   АВТО	47	Любители Авто	12	АВТОТЕМА™   АВТО   МАШИНЫ   ТАЧКИ	47	ЭТО ДРИФТ, ДЕТКА
13	АВТО ДО 1000\$ БЕЛАРУСЬ [1К.АУТО]	48	МИР ГРУЗОВИКОВ И ВНЕДОРОЖНИКОВ	13	Авто + Маляр	48	ЯПОНСКИЕ БРИЧКИ
14	АВТО КЛУБ	49	Музыка в машину	14	Авто - Подслушано	49	Японские пушки   JDM
15	АВТОБУГУРТ	50	Немецкие хищники	15	Авто Алматы	50	министерство оперов
16	АВТОМЕХАНИК ОТ БОГА	51	Под тюнингом	16	Авто Пермь		
17	АВТОРЫНОК ПЕРМЬ КУПИТЬ АВТО С ПРОБЕГОМ	52	Тачка за 200	17	Авто после осмотра   Беларусь		
18	АВТОТЕМА™   АВТО   МАШИНЫ   ТАЧКИ	53	Тачка за сотку	18	Авто приколы		
19	Авто + Маляр	54	Тачка за сотку   Урал	19	Автоподбор		
20	Авто - Подслушано	55	Тачки за копейки	20	Авторынок   Орел   Орловская область		
21	Авто Алматы	56	Тачки за копейки'77   Москва	21	Авторынок   Чебоксары   Чувашия		
22	Авто Пермь	57	ЭТО ДРИФТ, ДЕТКА	22	Авторынок ДНР-ЛНР (Донецк, Макеевка, Харцызск)		
23	Авто после осмотра   Беларусь	58	ЯПОНСКИЕ БРИЧКИ	23	Авторынок Корчей		
24	Авто приколы	59	Японские пушки   JDM	24	Авторынок Краснодар		
25	Автоподбор	60	министерство оперов	25	Авторынок Крым (Симферополь, Севастополь)		
26	Авторынок   Орел   Орловская область			26	Авторынок Питер СПб		
27	Авторынок   Чебоксары   Чувашия			27	Видеопродажа АВТО СВЕРДЛОВСКАЯ ОБЛАСТЬ		
28	Авторынок Беларусь   До 5000\$			28	Внедорожники		
29	Авторынок ДНР-ЛНР (Донецк, Макеевка, Харцызск)			29	ГАРАЖ   АВТО   GIF		
30	Авторынок Корчей			30	Главная Дорога		
31	Авторынок Краснодар			31	Главная Дорога: Авто / ДТП		
32	Авторынок Крым (Симферополь, Севастополь)			32	ДТП Кострома		
33	Авторынок Питер СПб			33	Дальнбойщики		
34	Боевой Корч			34	Дорожные войны!		

Рис. 3.14: Сообщества из решений для бюджета 16000 рублей. Целевой параметр: Комментарии

В случае с целевым параметром «Репосты» отличия минимальны как в количестве сообществ, так и в суммарных значениях показателя для смоделированного набора площадок. Но если мы обратимся к целевому параметру «Комментарии», то заметим, что количество сообществ отличается на целых 10 штук, а суммарные значения выбранного показателя на 22 комментария или всего на 2,05% от оптимального. Анализ и сравнение подобных сценариев поможет ЛПР объективно оценивать необходимость публикации рекламной записи в предлагаемых наборах сообществ. Для этого в данной работе введены относительные метрики, которые определены во втором параграфе первой главы. Например, важную роль могут сыграть коэффициенты распространения, общительности, средний возраст целевой аудитории, а также интенсивность публикационной активности в заданный временной промежуток для соответствующих групп социальной сети. Все эти факторы могут привести к тому, что выбран будет вариант с меньшим значением целевого параметра, но более отвечающим требованиям рекламодателя по другим не менее важным критериям.

**Замечание 3.2.** *Исходя из приведенных результатов установлено, что при решении задачи кластеризации для сжатие пространства признаков следует применять интерпретируемые методы отбора признаков.*

Проведенный сравнительный анализ результатов моделирования является важным этапом диссертационного исследования, поскольку он позволяет оптимизировать процесс моделирования, заключить ряд выводов, замечаний и наблюдений относительно рассматриваемых математических моделей и методов для достижения научных и практических результатов при решении поставленных задач с учетом особенностей указанной предметной области.

### **3.4. Основные результаты и выводы по третьей главе**

Данная глава посвящена построению схемы, разработке и программной реализации интеллектуальной системы поддержки принятия управленческих решений в задаче распространения информации в СМК. Применение интеллектуальной системы сценарного моделирования позволит оптимизировать распределение ресурсов компании, сформировать различные сценарии распространения информации с учетом особенностей площадок, поведенческой активности их участников и моделирования обратной связи аудитории.

Были рассмотрены особенности реализации и применения разработанной интеллектуальной системы, приведены схемы ее интегрирования в промышленный контур на примере социальных сетей, а также представлена архитектура программного комплекса и схема хранения данных и результатов моделирования с возможностью трансформации и визуализации информации. Проведено численное моделирование и сравнительный анализ результатов, который продемонстрировал важность применения сценарного подхода, как инструмента стратегического анализа, а также интеллектуальных инструментов обработки больших массивов данных в задаче распространения информации в СМК для поддержки принятия управленческих решений. Отметим, что система программно реализована в кроссплатформенной интегрированной среде разработки для языка программирования Python — PyCharm.

## Заключение

Диссертационное исследование посвящено анализу процесса распространения информации в СМК и разработке методологии для проведения сценарного моделирования с применением методов оптимизации и машинного обучения без учителя. В ходе проведенного исследования были разработаны инструменты для интеллектуального анализа статистических данных информационных площадок и формирования сценариев распространения информации в СМК. Это позволило реализовать интеллектуальную систему поддержки принятия управленческих решений в области информационно-коммуникационных технологий для автоматизации процессов управления на примере маркетинговой задачи распространения информации в социальных сетях, где в качестве информационной площадок выступают сообщества этой сети.

В рамках выполненного диссертационного исследования были получены следующие основные результаты:

1. Сформулированы постановки задач для моделирования процесса распространения информации в СМК с применением методов оптимизации.
2. Сформулированы постановки задач для моделирования процесса распространения информации в СМК с применением методов машинного обучения.
3. Разработан программный компонент, реализующий на языке программирования Python в кроссплатформенной интегрированной среде разработки PyCharm, циклический алгоритм предобработки статистических данных о пользовательской активности информационных площадок в задаче распространения информации в СМК.
4. Разработан и реализован программный компонент с рекомендательным блоком для формирования сценариев распространения информации в СМК и решения оптимизационных задач с возможностью трансформации и визуализации информации на языке программирования Python в кроссплат-

форменной интегрированной среде разработки PyCharm.

5. Разработаны программные компоненты с применением методов машинного обучения и отбора признаков без учителя для решения задачи кластеризации информационных площадок и задачи снижения размерности в оптимизационной задаче с возможностью трансформации и визуализации информации на языке программирования Python в кроссплатформенной интегрированной среде разработки PyCharm.
6. Разработана архитектура и программно реализована интеллектуальная система поддержки принятия управленческих решений в задаче распространений информации в СМК, а также предложена схема хранения данных и результатов моделирования с возможностью трансформации и визуализации информации на языке программирования Python в кроссплатформенной интегрированной среде разработки PyCharm. Осуществлено сравнение сценариев распространения информации и продемонстрирована целесообразность формирования нескольких наборов информационных площадок, посредством проведения численного моделирования.
7. Разработан инструмент проведения численного моделирования исследуемой системы, позволяющий осуществлять анализ чувствительности критериев, а также анализировать процесс формирования уникальных сценариев распространения информации в результате изменения предпочтений в задаче многокритериальной оптимизации на примере рынка товаров-услуг в цифровой среде с учетом номенклатуры товаров, бюджета и временного промежутка. Применение разработанных программных компонент позволяет корректировать предпочтения пользователя, посредством настройки гиперпараметров методов машинного обучения, а также сократить время формирования сценариев распространения информации. Проведение анализа важности признаков позволило определить базовый набор значимых характеристик объектов выбранными методами сжатия пространства признаков.

В заключение, отметим, что все задачи, сформулированные в рамках исследования, выполнены, а поставленная цель достигнута в полном объеме.

**Дальнейшая работа** по данной тематике может быть направлена на развитие функциональных возможностей интеллектуальной системы в рамках решения задачи по моделированию сценариев распространения информации в СМК. Представим перспективные направления и задачи будущих исследований:

- Интеграция с другими IT-решениями для проведения датамайнинга различных дополнительных характеристик информационных площадок.
- Применение методов анализа временных рядов и текстов для прогнозирования значений ключевых характеристик и формирования рекомендаций по размещению информации в различных СМК.
- Определение по скрытым паттернам ботов и «фейковых» аккаунтов на информационных площадках.
- Применение алгоритмов глубокого машинного обучения и нейросетевых технологий для осуществления анализа и генерирования нового медиа контента.

Кроме того, актуальными задачами будут являться доработка представленных схем прикладных математических моделей и программного комплекса, посредством редактирования или добавления новых программных и функциональных модулей для достижения большей эффективности, в смысле времени формирования сценариев распространения информации и их количества при заданных входных параметрах системы, при решении задач в различных областях знаний (экономика и управления, социология, политология).

## Список литературы

1. Автопостинг в социальных сетях — «Pur Ninja» [Электронный ресурс]. — Режим доступа: <https://pur.ninja/> (дата обращения: 11.06.2022).
2. Адигамов А. Э. Система поддержки принятия решений при проектировании сложных технических систем / А. Э. Адигамов, П. В. Макаров, И. Н. Симачева, Н. Д. Симачев // История и педагогика естествознания. — 2013. — №2. — С. 22–24.
3. Ажмухамедов, И.М. Моделирование процесса распространения информации в социальных сетях / И. М. Ажмухамедов, Д. А. Мачуева, Д. А. Жолдобов // Фундаментальные исследования. — 2017. — № 5. — С. 9–14.
4. Ажмухамедов, И.М. Моделирование распространения информации в социальных сетях / И.М. Ажмухамедов, Н. А. Аншаков, В. М. Гудонис // Проблемы Науки. — 2017. — № 20(102) — С. 16–19.
5. Анализ отчета «Global Digital 2023» [Электронный ресурс]. — Режим доступа: <https://www.web-canape.ru/business/statistika-interneta-i-socsetej-na-2023-god-cifry-i-trendy-v-mire-i-v-rossii/> (дата обращения: 15.03.2024).
6. Астафьева, Е. В. Модель рекламной компании с эффектом «недоедания» рекламы / Е.В. Астафьева, А.Ф. Терпугов // Вестник Томского государственного университета. — 2004. — № 284. — С. 34–36.
7. Ахмедова, Д. Д. Математическая модель функционирования страховой компании с учетом расходов на рекламу / Д. Д. Ахмедова, А. Ф. Терпугов // Известия вузов. Физика. — 2001. — № 1. — С. 25–29.
8. Ахмедова, Д. Д. Оптимизация деятельности страховой компании с учетом расходов на рекламу / Д. Д. Ахмедова, О. А. Змеев, А. Ф. Терпугов // Вестник Томского государственного университета. — 2002. — № 275. — С. 181–184.
9. Ахмедова, Д. Д. Оптимизация расходов на рекламу при деятельности страховой компании / Д. Д. Ахмедова, О. А. Змеев // Известия вузов. Физика. — 2001. — № 6. — С. 3–7.

10. Балашова И. В. СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ / И. В. Балашова, Т. А. Терещенко // The Scientific Heritage. — 2021. — № 79(4). — С. 3—7.
11. Бадрызлов, В. А. Оценка эффективности распространения информации в социальных сетях с использованием имитационного моделирования / В. А. Бадрызлов, В. В. Сидельцев // Креативная экономика. — 2018. — № 9(12). — С. 1359—1372.
12. Библиотека «SciPy», алгоритмы оптимизации «scipy.optimize» [Электронный ресурс]. — Режим доступа: <https://sociate.ru/> (дата обращения: 21.09.2023).
13. Биржа рекламы — «Sociate» [Электронный ресурс]. — Режим доступа: <https://sociate.ru/> (дата обращения: 11.06.2022).
14. Бородина, И. П. Модель оптимального поведения фирмы с учетом влияния рекламы / И. П. Бородина, С. В. Жак // Экономическая наука современной России. — 2004. — №3. — С. 80—87.
15. Бородина, И. П. Системы управления рекламными коммуникациями фирмы / И. П. Бородина. — Ростов н/Д : изд-во РБПХЛ СП РФ, 2004. — 123 с.
16. Висаитова, Л. Р. Системы поддержки принятия решений: эффективность принятия решений / Л. Р. Висаитова // Научно-практический электронный журнал Аллея Науки. — 2020. — № 11(50). — С. 219—225.
17. Гоманюк, С. О. Математическое моделирование рекламной кампании образовательного центра [Электронный ресурс] / С.О. Гоманюк // ВКР. НИУ ВШЭ. Москва. — 2015. — 53 с. Режим доступа: <https://www.hse.ru/edu/vkr/153015895?ysclid=lwewjxodyp495766419>.
18. Горковенко, Д. К. Обзор моделей распространения информации в социальных сетях / Д.К. Горковенко // Молодой ученый. — 2017. — № 8 (142). — С. 23—28.
19. Горковенко, Д. К. Сравнительный анализ моделей эпидемии и клеточного автомата при моделировании распространения информации в социальных сетях / Д. К. Горковенко // Научно-технические ведомости СПбГПУ.

- Информатика. Телекоммуникации. Управление. — 2017. — № 3(10). — С. 103—113.
20. Горчакова, Р. Р. Особенности формирования корпоративного имиджа / Р. Р. Горчакова // Известия высших учебных заведений. Поволжский регион. Общественные науки. — 2012. — № 2(22). — С. 185—192.
21. Грибанова Е. Б. Алгоритмы моделирования распространения информации при маркетинговых мероприятиях в группах онлайн-социальной сети / Е. Б. Грибанова // Проблемы управления. — 2018. — № 1. — С. 66—73.
22. Губанов Д. А. Модели влияния в социальных сетях / Д. А. Губанов, Д. А. Новиков, А. Г. Чхартишвили // Управление большими сетями. — 2009. — № 27. — С. 205—281.
23. Губанов, Д. А. Социальные сети: модели информационного влияния, управления и противоборства [Текст] / Д. А. Губанов, Д. А. Новиков, А. Г. Чхартишвили; Российская академия наук, Институт проблем управления. — Изд. 3-е, перераб. и доп. — Москва : Изд-во МЦНМО, 2018. — 223 с.
24. Гусс, С. В. Разработка семейства программных систем в специфической предметной области / С. В. Гусс // Математические структуры и моделирование. — 2011. — № 22. — С. 55—68.
25. Делицын, Л. Л. Количественные модели распространения нововведений в сфере информационных и телекоммуникационных технологий [Текст] : монография / Л. Л. Делицын ; Московский гос. ун-т культуры и искусства. — Москва : Московский гос. ун-т культуры и искусства, 2009. — 106 с.
26. Дзялошинский, И. М. Коммуникативное воздействие: мишени, стратегии, технологии: монография / И. М. Дзялошинский. — М.: НИУ ВШЭ, — 2012. — 572 с.
27. Ершов, В. Е. Исследование отношения пользователей социальных сетей к активности по продвижению на их основе [Электронный ресурс] / В. Е. Ершов // Интернет-журнал «Науковедение». — 2017. — № 5(9). — С. 1—13. Режим доступа: <https://cyberleninka.ru/article/n/issledovanie-otnosheniya-polzovateley-sotsialnyh-setey-k-aktivnosti-po-prodvizheniyu-na-ih-osnove/viewer>.

28. Ехлаков, Ю. П. Многокритериальная задача поддержки принятия решений при выборе вариантов поставки на целевые рынки линейки программных продуктов / Ю. П. Ехлаков, Д. Н. Бараксанов // Доклады Томского государственного университета систем управления и радиоэлектроники. — 2015. — № 3(37). — С. 114—119.
29. Ехлаков, Ю. П. Основные положения по разработке программы продвижения программных продуктов в сети Интернет / Ю. П. Ехлаков, Д. Н. Бараксанов // Бизнес-информатика. — 2012. — № 4(22). — С. 26—32.
30. Ехлаков, Ю. П., Структура и содержание коммуникационного сообщения при организации Интернет-рекламы программных продуктов / Ю. П. Ехлаков, Д. Н. Бараксанов // Интернет-маркетинг. — 2016. — № 3(93). — С. 146—156.
31. Измоденова, К. В. Об оптимальном управлении процессом распространения информации / А. П. Михайлов, К. В. Измоденова // Математическое моделирование. — 2005. — № 5(17). — С. 67—76.
32. Карамова, В. В. Позиционирование бренда/продукции организации как инструмент повышения эффективности её деятельности / В. В. Карамова // Вестник УГНТУ. Наука, образование, экономика. Серия экономика. — 2018. — № 3(25). — С. 105—112.
33. Кознов, Д. В. Инструменты для управления вариантивностью: готовность к промышленному применению / Д. В. Кознов, И. А. Новицкий, М. Н. Смирнов // Труды СПИИРАН. — 2013. — № 3(26). — С. 297—331.
34. Коэльо, Л. П. Построение систем машинного обучения на языке Python / Л. П. Коэльо, В. Ричарт. — 2-е издание; пер. с англ. А.А. Слинкина. — М.: ДМК Пресс, 2016. — 302 с.
35. Кряжева, Ю. С. Использование Social Media Marketing как инструмента продвижения компании / Ю. С. Кряжева // Формула менеджмента. — 2017. — № 1(1). — С. 5—11.
36. Ларичев, О. И. Системы поддержки принятия решений. Современное состояние и перспективы их развития [Электронный ресурс] / О. И. Ларичев, А. Б. Петровский // Итоги науки и техники. Сер. Техническая кибернети-

- ка. — Москва : ВИНТИ, 1987. — Т. 21. — С. 131—164. — Режим доступа: [http://www.raai.org/library/papers/Larichev/Larichev\\_Petrovsky\\_1987.pdf](http://www.raai.org/library/papers/Larichev/Larichev_Petrovsky_1987.pdf) (дата обращения: 23.04.2019).
37. Лившин, Д. А. Математическое моделирование в маркетинге при построении рекламных кампаний / Д. А. Лившин, Л. И. Воронова // Современные наукоемкие технологии. — 2014. — № 5-2. — С. 207—209.
38. Маревцева, Н. А. Некоторые математические модели информационного нападения и информационного противоборства / Н. А. Маревцева // Социология. — 2011. — № 3. — С. 26—35.
39. Маревцева, Н. А. Простейшие математические модели информационного противоборства / Н. А. Маревцева // Математическое моделирование социальных процессов. — 2010. — № 11 — С. 59—72.
40. Маркетинговая платформа искусственного интеллекта «Albert» [Электронный ресурс]. — Режим доступа: <https://albert.ai/> (дата обращения: 15.03.2024).
41. Мирская, С. Ю. Оценка влияния рекламы на формирование потребительского спроса / С. Ю. Мирская, В. И. Сидельников, И. П. Бородина // Известия вузов. Северо-Кавказский регион. Серия: Естественные науки. — 2004. — № 6. — С. 13—18.
42. Михайлов, А. П. Модели информационной борьбы / А. П. Михайлов, Н. А. Маревцева // Математическое моделирование. — 2011. — № 10(23). — С. 19—32.
43. Михайлов, А. П. Модель информационного противоборства в социуме при периодическом дестабилизирующем воздействии / А. П. Михайлов, А. П. Петров, О. Г. Прончева, Н. А. Маревцева // Математическое моделирование. — 2017. — № 2(29). — С. 23—32.
44. Михайлов, А. П. Развитие модели распространения информации в социуме / А. П. Михайлов, А. П. Петров, Н. А. Маревцева, И. В. Третьякова // Математическое моделирование. — 2014. — № 3(26). — С. 65—74.
45. Мюллер А., Гвидо С. Введение в машинное обучение с помощью Python: руководство для специалистов по работе с данными / А. Мюллер, С. Гвидо.

- пер. с англ. — СПб.: ООО «Альфа-книга», 2017. — 480 с.
46. Нетология. Сообщество социальной сети [Электронный ресурс]. — Режим доступа: <https://netology.ru/glossariy/gruppa-v-socialnoy-seti> (дата обращения: 09.02.2024).
  47. Носова, М. В. Моделирование распространения информации в децентрализованных сетевых системах с нерегулярной структурой / М. В. Носова, Л. И. Сенникова // Новые информационные технологии в автоматизированных системах. — 2014. — № 17. — С. 329—335.
  48. Описание методов кластерного анализа. Библиотека Scikit-learn [Электронный ресурс]. — Режим доступа: <https://scikit-learn.org/stable/modules/clustering.html> (дата обращения: 22.03.2023).
  49. Описание методов оценки качества кластеризации. Библиотека Scikit-learn [Электронный ресурс]. — Режим доступа: <https://scikit-learn.org/stable/modules/clustering.html#clustering-performance-evaluation> (дата обращения: 22.03.2023).
  50. Описание метода порога отклонения. Библиотека Scikit-learn [Электронный ресурс]. — Режим доступа: [https://scikit-learn.org/stable/modules/feature\\_selection.html](https://scikit-learn.org/stable/modules/feature_selection.html) (дата обращения: 22.03.2023).
  51. Оптимизация медиаплана с помощью математического моделирования [Электронный ресурс]. — Режим доступа: [https://www.marketing.spb.ru/lib-special/case/math\\_model.htm](https://www.marketing.spb.ru/lib-special/case/math_model.htm) (дата обращения: 17.03.2024).
  52. Орельен, Ж. Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем / Ж. Орельен. — 2-е изд., пер. с англ. — СПб.: ООО «Диалектика», 2020. — 1040 с.
  53. Попов Е. В. Теория вычисления охвата СМИ / Е. В. Попов, Г. А. Шматов // Проблемы управления. — 2009. — № 5. — С. 22—27.

54. Попов Е. В. Вычисление охвата СМИ / Е. В. Попов, Г. А. Шматов // Проблемы управления. — 2010. — № 2. — С. 34—38.
55. Реализация и описание метода главных компонент. Библиотека Scikit-learn [Электронный ресурс]. — Режим доступа: <https://scikit-learn.org/stable/modules/decomposition.html#pca> (дата обращения: 22.03.2023)
56. Рекламная платформа на базе «Big Data МТС» — «МТС Маркетолог» [Электронный ресурс]. — Режим доступа: <https://marketolog.mts.ru/> (дата обращения: 15.03.2024).
57. Самарский, А. А. Математическое моделирование: Идеи. Методы. Примеры. / А. А. Самарский, А. П. Михайлов. — 2-е изд., испр. — М.: Физматлит, 2001. — 320 с.
58. Свиркин, М. В. Имитационное моделирование социальной сети с двумя лидерами / М. В. Свиркин, В. В. Карелин, В. М. Буре, А. В. Екимов // Вестник СПбГУ. Серия 10. Прикладная математика. Информатика. Процессы управления. — 2016. — № 1. — С. 92—98.
59. Семенова, О. В. Важность позиционирования бренда и актуальные методы продвижения бренда организации / О. В. Семенова // Вестник Самарского университета. Экономика и управление. — 2023. — № 2(14). — С. 155—163.
60. Семиглазов, А. М. Математическое моделирование рекламной кампании / А. М. Семиглазов, В. А. Семиглазов, К. И. Иванов // Управление, вычислительная техника и информатика, доклады ТУСУРа. — 2010. — № 2(22) — С. 342—349.
61. Семиглазов, В. А. Оптимизация расходов на рекламную кампанию / В. А. Семиглазов // Маркетинг. — 2007. — № 1. — С. 63—70.
62. Сервис аналитики контента социальных сетей — «Popsters» [Электронный ресурс]. — Режим доступа: <https://popsters.ru> (дата обращения: 11.06.2022).
63. Сервис управления социальными сетями — «LiveDune» [Электронный ресурс]. — Режим доступа: <https://livedune.com/ru> (дата обращения: 11.06.2022).

64. Социальная сеть «ВКонтакте» [Электронный ресурс]. — Режим доступа: <https://vk.com/> (дата обращения: 15.03.2024).
65. Статистика аккаунтов в социальных сетях — «JagaJam» [Электронный ресурс]. — Режим доступ: <https://jagajam.com> (дата обращения: 11.06.2022).
66. Статистика сообществ социальной сети «ВКонтакте» — «AllSocial» [Электронный ресурс]. — Режим доступ: <http://allsocial.ru> (дата обращения: 11.06.2022).
67. Терелянский, П. В. Системы поддержки принятия решений. Опыт проектирования : монография / П. В. Терелянский. — Волгоград : ВолгГТУ. 2009. — 127 с.
68. Тимофеев, С. В., Математическое моделирование информационного противоборства / С. В. Тимофеев, А. В. Баенхаева // Научно-технические ведомости СПбГПУ. Физико-математические науки. — 2021. — № 1(14). — С. 164—176.
69. Трофимов, В. М. Модель информационного воздействия в социальных сетях / В. М. Трофимов, Л. А. Видовский, Р. А. Дьяченко // Научный журнал КубГАУ. — 2015. — № 110. — С. 1788—1801.
70. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Текст] : пер. с англ. / П. Флах. — Москва : ДМК Пресс, 2015. — 399 с.
71. Фурсов, Д. В. Анализ применения методов машинного обучения и оптимизации в задаче формирования набора сообществ для продвижения информации в социальных медиа [Электронный ресурс] / Д. В. Фурсов (науч. рук. Крылатов А.Ю.) // Сборник тезисов докладов конгресса молодых ученых. Электронное издание. — СПб: Университет ИТМО. — 2024. Режим доступа: <https://kmu.itmo.ru/digests/article/13613>.
72. Фурсов, Д. В. Имитационное моделирование информационного воздействия с использованием средств массовой коммуникации / Д. В. Фурсов // Процессы управления и устойчивость. — 2018. — № 1(5). — С. 517—522.
73. Фурсов, Д. В. Имитационное моделирование продвижения информации в сложных организационных системах / Д. В. Фурсов // Процессы управле-

- ния и устойчивость. — 2019. — № 1(6). — С. 486—490.
74. Фурсов, Д. В. Интеллектуальная система поддержки принятия решений в задаче распространения информации в средствах массовой коммуникации / Д. В. Фурсов, А. Ю. Крылатов // Информационные системы и технологии. — 2024. — № 2(142). — С. 80—89.
75. Фурсов, Д. В. Интеллектуальная система поддержки принятия управленческих решений в задаче формирования набора площадок распространения информации в средствах массовой коммуникации [Электронный ресурс] / Д. В. Фурсов (науч. рук. Крылатов А.Ю.) // Сборник тезисов докладов конгресса молодых ученых. Электронное издание. — СПб: Университет ИТМО. — 2024. Режим доступа: <https://kmu.itmo.ru/digests/article/13603>.
76. Фурсов, Д. В. Применение методов оптимизации в задаче распространения информации в средствах массовой коммуникации / Д. В. Фурсов, А. Ю. Крылатов, М. В. Свиркин // Прикладная математика и вопросы управления. — 2024. — № 1. — С. 139—159.
77. Фурсов, Д. В. Применение методов машинного обучения в задаче распространения информации в средствах массовой коммуникации / Д. В. Фурсов // Автоматизация процессов управления. — 2024. — № 2(76) — С. 75—86.
78. Фурсов, Д. В. Программный комплекс «Интеллектуальная система поддержки принятия управленческих решений в задаче распространения информации в средствах массовой коммуникации» / Д. В. Фурсов // Программа для ЭВМ. Свидетельство о регистрации программы для ЭВМ RU2024614394, дата регистрации 22.02.2024, номер и дата поступления заявки – 2024613251, 19.02.2024, дата публикации 22.02.2024, Бюллетень ФИПС №3.
79. Фурсов, Д. В. Программный компонент «Обработка статистических данных акторов средств массовой коммуникации» / Д. В. Фурсов // Программа для ЭВМ. Свидетельство о регистрации программы для ЭВМ RU2024615049, дата регистрации 01.03.2024, номер и дата поступления

- заявки – 2024613270, 19.02.2024, дата публикации 01.03.2024, Бюллетень ФИПС №3.
80. Фурсов, Д. В. Программный компонент «Программная реализация методов кластерного анализа для формирования решающего набора акторов в задаче распространения информации» / Д. В. Фурсов // Программа для ЭВМ. Свидетельство о регистрации программы для ЭВМ RU2024614766, дата регистрации 28.02.2024, номер и дата поступления заявки – 2024613298, 19.02.2024, дата публикации 28.02.2024, Бюллетень ФИПС №3.
81. Фурсов, Д. В. Программный компонент «Программная реализация методов оптимизации в задаче распространения информации» / Д. В. Фурсов // Программа для ЭВМ. Свидетельство о регистрации программы для ЭВМ RU2024614893, дата регистрации 29.02.2024, номер и дата поступления заявки – 2024613262, 19.02.2024, дата публикации 01.03.2024, Бюллетень ФИПС №3.
82. Фурсов, Д. В. Программный компонент «Программная реализация методов оптимизации с предварительной кластеризацией в задаче распространения информации» / Д. В. Фурсов // Программа для ЭВМ. Свидетельство о регистрации программы для ЭВМ RU2024614590, дата регистрации 27.02.2024, номер и дата поступления заявки – 2024613238, 19.02.2024, дата публикации 27.02.2024, Бюллетень ФИПС №3.
83. Фурсов, Д. В. Разработка прототипа интеллектуальной системы поддержки принятия управленческих решений в задаче распространения информации / Д. В. Фурсов // ИТ. Наука. креатив : Материалы VI Всероссийской конференции, в рамках I Международного форума, Омск, 14–16 мая 2024 года. – Москва: Общество с ограниченной ответственностью "Издательско-книготорговый центр "Колос-с 2024. – С. 432-436.
84. Фурсов, Д. В. Статистическое имитационное моделирование продвижения информации в социальных сетях / Д. В. Фурсов, М. В. Свиркин // Перспективы науки. — 2020. — № 6(129). — С. 39–44.

85. Хабр. Статья: «Обзор методов отбора признаков» [Электронный ресурс]. — Режим доступа: <https://habr.com/ru/companies/jetinfosystems/articles/470622/> (дата обращения: 22.03.2023).
86. Храпов П. В., Математическое моделирование процесса распространения новостей в социальных сетях / П. В. Храпов, В. А. Столбова // Современные информационные технологии и ИТ-образование. — 2019. — № 1(15). — С. 225—231.
87. Часто задаваемые вопросы. «Popsters» [Электронный ресурс]. — Режим доступа: <https://popsters.ru/app/faq/> (дата обращения: 11.06.2022).
88. Черноруцкий, И. Г. Методы принятия решений. — СПб.: БХВ–Петербург, 2005. — 416 с.: ил.
89. Чесноков, В. О. Применение алгоритма выделения сообществ в информационном противоборстве в социальных сетях / В. О. Чесноков // Вопросы кибербезопасности. — 2017. — № 1(19). — С. 37—44.
90. Чибисова, А. В., Математическое моделирование рекламной кампании / А. В. Чибисова, Д. С. Шинаков // Математическое моделирование и численные методы. — 2022. — № 3. — С. 84—97.
91. Шабанов, Р. М. Интеллектуальная информационная система поддержки принятия решений / Р. М. Шабанов, Н. А. Микушин // Молодой исследователь Дона. — 2019. — № 4(19). — С. 91—97.
92. Шигина, Я. И. Маркетинг в социальных медиа: современные инструменты продвижения для малого бизнеса / Я. И. Шигина, К. А. Зорина // Вестник технологического университета. — 2015. — № 23(18). — С. 96—102.
93. Шматов, Г. А. К проблеме поиска закономерностей в сфере медиавоздействий / Г. А. Шматов // Вестник УГТУ–УПИ. Серия экономика и управление. — 2010. — № 6. — С. 102—111.
94. Шматов, Г. А. Математическая теория медиапланирования : монография / Г. А. Шматов ; Российская акад. наук, Уральское отд-ние, Ин-т экономики. — Екатеринбург : Ин-т экономики УрО РАН, 2009. — 329 с.

95. Шматов, Г. А. Оптимизация периода размещения рекламы / Г. А. Шматов // Реклама: теория и практика. — 2008. — № 1. — С. 8—15.
96. Alyoubi, B. A. Decision support system and knowledge-based strategic management / B. A. Alyoubi // ELSEVIER. — 2015. — № 65. — P. 278—284.
97. Bass, F. M. A new product growth for model consumer durables / F. M. Bass // Management Science. — 1969. — № 5(15). — P. 215—227.
98. Cai, D. Unsupervised feature selection for Multi-Cluster data / D. Cai, C. Zhang, X. He // Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — 2010. — P. 333—342.
99. Calinski, T. A dendrite method for cluster analysis / T. Calinski, J. Harabasz // Communications in Statistics. — 1974. — Vol. 3, iss. 1, — P. 1—27.
100. Davies, D. L. A Cluster Separation Measure / D. L. Davies, D. W. Bouldin // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1979. — Vol. PAMI-1, iss. 2. — P. 224—227.
101. Fursov, D. Problems of Data Processing in the Problem of Modeling Advertising Campaigns in Social Networks Using Python Libraries / D. Fursov, A. Krylatov, M. Svirkin, F. Prokhorenko // R. Silhavy, P. Silhavy, Z. Prokopova (eds) Data Science and Algorithms in Systems. CoMeSySo 2022. Lecture Notes in Networks and Systems. Springer, Cham. — 2023. — Vol 597, iss. 2. — P. 990—1001.
102. He, X. Laplacian Score for Feature Selection [Электронный ресурс] / X. He, D. Cai, P. Niyogi // Proceeding of Advances in Neural Information Processing Systems. — 2005. — Vol. 18. — Режим доступа: <https://api.semanticscholar.org/CorpusID:14270830>.
103. Jin, J. HPC Simulations of Information Propagation over Social Networks / J. Jin, S. J. Turner, B. S. Lee, J. Zhong, B. He // Procedia Computer Science. — 2012. — Vol. 9. — P. 292—301.
104. Khodashahri, N. G. Decision support system (DSS) / N. G. Khodashahri, M. H. Sarabi // Singaporean journal of business economics and management studies. — 2013. — №6. — P. 94—102.

105. Kleinberg, J. M. An Impossibility Theorem for Clustering [Электронный ресурс] / J. M. Kleinberg // Adv Neural Inform Process Syst (NIPS). — 2002 — Vol. 15. — Режим доступа: <https://api.semanticscholar.org/CorpusID:1487131>
106. Klepek, M. Marketing communication model for social networks / M. Klepek, H. Starzyczna // Journal of business economics and management. — 2018. — Vol. 19, iss. 3. — P. 500—520.
107. Pistol, L. Model of simulation for optimizing marketing mix through conjoint analysis case study: launching a product on a new market / L. Pistol, R. Bucea-Manea-Tonis // Economics World. — 2017. — Vol. 5, iss 4. — P. 311—315.
108. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis / P. J. Rousseeuw // Journal of Computational and Applied Mathematics. — 1987. — Vol. 20. — P. 53—65.
109. Tripathi, K. P. Decision support system is a tool for making better decisions in the organization / K. P. Tripathi // Indian Journal of Computer Science and Engineering. — 2017. — № 21. — P. 112—117.
110. Yekhlakov, Yu. P. Mathematical model and algorithm for selection of Internet sites and places for display of communication messages in planning advertising campaigns / Yu. P. Yekhlakov, D. N. Baraksanov // Business Informatics. — 2017. — № 1(39). — P. 55—60.
111. Yekhlakov, Yu. P. Mathematical model and algorithm of selecting software promotion options differentiated by functionality and business models / Yu. P. Yekhlakov, D. N. Baraksanov // Business Informatics. — 2015. — № 4(34). — P. 55—62.