

# Ивент-анализ: применение в социальном прогнозировании

Т. В. Коренькова<sup>1</sup>, А. А. Артамонов<sup>2</sup>, М. С. Улизко<sup>3</sup>

Национальный исследовательский ядерный университет «МИФИ», Москва, Россия

<sup>1</sup> ORCID: 0009-0000-6257-8500, [korenkova.tanya@mail.ru](mailto:korenkova.tanya@mail.ru)

<sup>2</sup> ORCID: 0000-0002-9140-5526, [aartamonov@kaf65.ru](mailto:aartamonov@kaf65.ru)

<sup>3</sup> ORCID: 0000-0003-2608-8330, [mulizko@kaf65.ru](mailto:mulizko@kaf65.ru)

## Аннотация

Мониторинг взаимосвязей между социальными событиями и явлениями и прогнозирование динамики их изменений необходимы в условиях нестабильности современного мира. Существует много отдельных методов анализа для социального прогнозирования, однако для работы выбрана методика ивент-анализа, которая в научной литературе в рамках данной задачи рассмотрена недостаточно, но имеет высокий потенциал. Целью статьи является адаптация методики ивент-анализа для ее использования в качестве инструмента социального прогнозирования. Основные данные для исследования собраны с российского информационно-новостного ресурса в период 2020-2023 года. Опираясь на классическую методику ивент-анализа, в работе определены классификаторы, представленные в работе в виде социальных сфер. В рамках этапа аналитического сравнения проведен графовый анализ (построены графы связей между категориями, выделены центральные узлы-категории); осуществлен анализ временных рядов (сегментация временных рядов алгоритмом PELT, кластеризация временных рядов алгоритмом k-средних); определены ключевые термины для событий-инфоповодов. Итоговым продуктом является аналитическая панель с фильтрами, статистикой и интерактивными графиками. Аналитическая панель позволяет сравнивать данные в статическом и динамическом состоянии, делать выводы о прошлых и будущих состояниях объектов социального прогнозирования. Главным результатом работы является разработанная автором методика проведения ивент-анализа, которую можно использовать для комплексного анализа новостных потоков, адаптировать под нужные категории, представляющие определенную сущность или сферу, и применять в различных социальных организациях или мониторинговых службах.

**Ключевые слова:** социальное прогнозирование, ивент-анализ, анализ новостей, анализ графов, вычисление центральности, анализ временных рядов, сегментация временных рядов, кластеризация временных рядов, PELT.

## 1. Введение

В условиях быстро меняющегося мира, эффективное прогнозирование социальных процессов становится одной из ключевых задач для исследователей и практиков [1]. Ивент-анализ, как метод, предлагающий системный подход к изучению значимых событий и их последствий, представляет собой инструмент для выявления закономерностей и тенденций в социальном развитии.

А. В. Мальцевой [1] предложен алгоритм применения ивент-анализа на примере изучения процессов на рынке труда, состоящий из этапов сбора данных, выделения классификаторов, подсчета результатов, сравнения полученных величин и верификации результатов. При этом выявлено, что не существует однозначной методологии

проведения данного вида анализа, поэтому становится возможным модифицировать некоторые этапы. М. В. Новоселов [2] описал некоторые дополнительные статистические, математические и графические методы, которые можно применять для проведения расширенного ивент-анализа, в частности, анализ временных рядов, кластерный анализ.

Количество как русскоязычных, так и зарубежных научных работ на стыке социального прогнозирования и ивент-анализа несущественно. А. А. Азаров [3] исследовал термин социального компьютинга — междисциплинарной области исследований, включающей изучение социального поведения средствами вычислительных систем. Одной из важных задач социального компьютинга является моделирование, анализ и прогнозирование социального поведения акторов с помощью таких методов, как интент-анализ, ивент-анализ, психографический анализ. В. И. Кудрявцевой [4] описаны основные подходы к проведению социального прогнозирования: общенаучные, интуитивные, формализованные и комплексные методы прогнозирования, и среди их числа упоминается применение ивент-анализа.

Гораздо больше исследований посвящено исследованию динамики развития социальных сущностей в новостях и их кластеризации, что связано с конкретными инструментами ивент-анализа. Дж. Л. Ортега [5] предложил и реализовал подход к построению сетевых графов на основе частоты, с которой СМИ совместно упоминают исследовательские работы. И. Блох и В. Александров [6] построили временные ряды, которые отображают распределение популярности кластеров, представляющих определенные социальные явления, с течением времени.

Таким образом, в отдельности тематики социального прогнозирования и ивент-анализа изучаются в научных публикациях довольно обширно, но работ, описывающих их совместное применение, ограниченное число. На основе обзора литературы определено, что анализ графов и временных рядов событий могут быть хорошими инструментами для изучения связей между социальными сущностями и отслеживания динамики их изменений. Стоит отметить, что эти инструменты редко описываются другими авторами на этапах ивент-анализа. Это свидетельствует о новизне и актуальности работы и необходимости изучать ивент-анализ и социальное прогнозирование в связке.

## **2. Постановка задачи**

Цель работы заключается в адаптации методики ивент-анализа для использования в качестве инструмента социального прогнозирования. В ходе исследования рассмотрены ключевые аспекты социального прогнозирования, выявлены его основные сферы применения и описаны этапы классической методики ивент-анализа. Особое внимание уделено разработке модификаций, которые позволяют усовершенствовать данную методику, а также её практическому применению на основе реальных данных.

В процессе работы исследованы как теоретические основы ивент-анализа, так и его практическая значимость в различных социальных контекстах. Результаты исследования направлены на создание эффективного инструмента, который не только повысит точность анализа, но и позволит более надежно предсказывать социальные изменения.

## **3. Теория**

Социальное прогнозирование - процесс изучения перспектив развития социальных объектов с целью повышения эффективности управления ими, исходя из работы с множеством альтернатив, сочетания разнообразных методов и абстрактного характера возможных решений. Основными особенностями социального прогнозирования являются отсутствие четких целей и направлений прогнозирования, сложность формализации социальной сферы и необходимость сочетания качественных и количественных методов [7]. Социальное прогнозирование, в отличие от обществоведческого — это прогнозирование только в области социологии. Однако социология с каждым годом все

более активно вторгается в другие отношения со стороны их социальной организации: экономические, национальные, морально-этические и т.д. На основе этих тенденций, некоторые авторы выделяют перечень областей и сфер, в которых возможно проводить социальное прогнозирование, см. Табл. 1 [7].

Таблица 1. Сферы применения социального прогнозирования

<b>Сфера</b>	<b>Области применения социального прогнозирования</b>
Наука	Перспективы развития научных кадров, исследовательских учреждений, финансирования научных открытий
Технологии	Перспективы развития новых технологий, информатизации общества, сохранения конфиденциальности
Экономика	Перспективы развития социальной организации труда, борьбы с безработицей и инфляцией
Политика	Перспективы развития государственных и международных отношений, мониторинг отношения народа к власти
Право	Перспективы внедрения новых законов в социальной сфере, мер поддержки населения, сохранения прав человека
Народонаселение	Перспективы изменения структуры общества, миграционных процессов
Образование	Перспективы развития различных учреждений образования, повышения квалификации и стимулирования кадров
Здравоохранение	Перспективы развития медицинских учреждений, открытий в области медицины, здорового образа жизни
Культура	Перспективы развития культурных благ, туризма, влияния СМИ, сохранения культурного наследия
Экология	Перспективы освоения Земли и космоса, сохранения окружающей среды, развития городов и транспорта
Социальная структура	Перспективы развития социально-производственной, профессиональной, образовательной и половозрастной структур
Общественная жизнь	Перспективы развития общественного порядка, социальных потребностей, борьбы с антиобщественными явлениями, неравенством и бедностью

По оценкам отечественных и зарубежных ученых, сегодня насчитывается 150–200 различных методов научного прогнозирования. Однако число методов, которые можно назвать базисными и наиболее распространенными в практике социальной прогностики, значительно меньше и достигает 15–20. Многие из этих методов относятся скорее к приемам и методикам прогнозирования, учитывающим нюансы динамики развития объектов. Как правило, применяются в отдельности либо экспертные, либо фактографические методы социального прогнозирования. Под критерии комплексности, гибкости, универсальности и изученности в научных работах, подходят не так много методов: в частности, экспликативные [3]. Главное преимущество этих методов в том, что они изначально рассчитаны на применение в социальной сфере, а не заимствованы из более точных наук. Некоторые из этих методов по большей части содержат в себе описательный анализ, однако на финальном этапе предполагается выработка рекомендаций и социальных прогнозов экспертами. Данные методы учитывают ограничения и сложность формализации данных из социальной сферы, напрямую связаны с обнаружением взаимосвязей и закономерностей в данных, а также предполагают использование фактографических методов в сочетании с экспертными. На практике это встречается, когда экспертам заранее представляют фактографическую информацию об объекте или знакомят их со сделанными ранее фактографическими прогнозами, либо, наоборот, в процессе экстраполяционного моделирования тенденций развития объекта наряду с фактическими данными учитываются данные экспертных оценок. Таким об-

разом, на основе представленных выше заключений и анализа научных работ выдвинута гипотеза, что экспликативный метод ивент-анализа обладает достаточной гибкостью и набором инструментов для его применения в задаче социального прогнозирования.

Ивент-анализ как метод политической науки зародился в 1960-х годах в научных трудах Чарльза Макклелланда [8]. Ивент-анализ является количественным методом изучения политической реальности, фокусирующимся на систематическом анализе сообщений о событиях. Его «родственник» - контент-анализ, оба метода осуществляют количественный анализ текстов, но разными способами. Объектом ивент-анализа являются не сами события, а сообщения о событиях, преимущественно из СМИ. Эти события систематизируются, анализируются, классифицируются и обрабатываются с использованием программно-математических методов [1].

В настоящее время ивент-анализ находит применение в конфликтологии, социологии, политологии и естественных науках. Широкая сфера его применения объясняется возможностью сравнения различных событий, их анализа по количеству участников, продолжительности и масштабам взаимодействия. Это позволяет не только сравнивать события, но и строить многовариантные сценарии, в связи с чем повышается точность не только тактического, но и стратегического прогнозирования. Таким образом, ивент-анализ обеспечивает более детальное представление об изменениях в политической и социальной обстановке по сравнению с традиционными исследовательскими методами.

Методика ивент-анализа направлена на мониторинг хода событий и их интенсивности с целью выявления основных тенденций эволюции обстановки как на национальном, так и на международном уровне. Изначально процесс анализа чаще всего включал в себя два подхода: первый, основанный на анализе данных «снизу», и второй, где исследователь формулировал нормативные модели для последующего заполнения фактами, - подход «сверху». Первый подход означает, что исследователь не предопределяет заранее важные аспекты изучаемого процесса, за исключением основного объекта наблюдения. Во втором случае исследование строится на структурированном сборе информации, где определенные элементы процесса заранее выделяются как наиболее значимые. Обычно в исследованиях оба подхода используются совместно, обогащая взгляд на анализируемые ситуации [1].

Практическую реализацию ивент-анализа можно разделить на две основные фазы. Первая фаза включает формализованное представление сообщений о событиях с использованием определенной кодировочной схемы, что создает «данные о событиях» (Event Data). Вторая фаза включает использование данных для формулирования содержательных гипотез и выводов относительно изучаемых политических процессов, а также для построения и тестирования моделей. В современной политической науке на этой стадии используется широкий набор статистических методов и математических подходов, таких как факторный, дискриминантный, корреляционный, кластерный анализ и т.д. На заключительном этапе происходит валидация результатов с последующим составлением прогнозов и экспертных оценок [1]. Модификации в виде внедрения графовых алгоритмов, сегментации и кластеризации временных рядов, выделения ключевых слов внесены на этапе аналитического исследования, а все остальные этапы выполнены в соответствии с установленной методологией. Описание каждого этапа ивент-анализа, реализованного в этой работе, приведено далее.

### **3.1. Составление информационного массива**

Первый этап работы с ивент-анализом - создание информационного массива или банка данных. Источники данных могут быть различными: официальные документы, отчеты, новостные статьи, статистика происшествий и т.д. В работе в качестве источника сбора данных выбран информационно-новостной ресурс, поскольку новости отражают практически все аспекты социальных взаимодействий и событий. Сбор данных

проведен автоматизированно методом веб-скрапинга на языке программирования Python с помощью библиотек BeautifulSoup и requests. Для этой цели использован только один российский информационно-новостной ресурс, чтобы избежать проблем дублирования и агрегирования данных из разных источников. Новости собраны за полных четыре года в период с 2020 по 2023 года вместе с такими метаданными, как дата, рубрика, заголовок исходной новости, а также заголовки новостей, упомянутых в контексте [9]. При этом в данной работе принимается допущение, что каждую из новостей в выбранном информационно-новостном ресурсе можно отнести к одной из социальных категорий, хотя среди них могут встречаться и те новости, которые трудно оценить с точки зрения вклада в социальную сферу – например, некоторые новости из научной тематики.

### **3.2. Построение системы классификаторов**

Второй этап – разработка системы классификации новостных сообщений о социальных событиях для формализации событий и явлений и анализа их взаимодействий. Результаты наблюдения могут быть зарегистрированы с использованием кодировки. Для учета событий может использоваться более сложная система – кодировочный бланк, который включает различные детали изучаемого явления, такие как данные об инициаторах событий, социальном контексте происходящего, типе события, объектах, на которые направлены действия акторов и т.д. На сегодняшний день существует множество баз данных ивент-анализа, которые постоянно пополняются новыми исследованиями. Все базы данных можно разделить на две основные группы. Первая группа – субъектно-ориентированные базы данных, сфокусированные на участниках международных политических процессов, включающие информацию о взаимодействии определенного набора участников за определенный период времени. Вторая группа – проблемно-ориентированные базы данных, сфокусированные на конкретных исторических событиях, таких как крупные конфликты [1].

В работе создана новая система классификация, основанная на выделенных ранее социальных сферах, которые могут требовать разработки социальных прогнозов, см. Табл. 1. При необходимости, некоторые из них можно расширять или объединять, что было сделано при помощи модели Word2Vec. Математическая модель Word2Vec, созданная Google, представляет собой нейронную сеть, которая обрабатывает текстовые данные, и включает в себя две модели обучения: «Continuous Bag of Words» (CBOW) и Skip-gram. CBOW – «непрерывный мешок со словами», архитектура, которая предсказывает текущее слово, исходя из окружающего его контекста. Архитектура типа Skip-gram использует текущее слово, чтобы предугадывать окружающие его слова. На вход в обучающую модель Word2Vec подается текстовый массив данных, а на выходе генерируются векторы слов. Далее вычисляются косинусные расстояния между всеми словами из входной выборки. Это значит, что для каждого слова из поданного текста можно найти список самых близких к нему слов, то есть таких, которые чаще всего упоминаются в одном контексте, на основе схожести их векторов [10].

Таким образом, модель Word2Vec была обучена на всем массиве новостных данных, и в результате получилась таблица соотнесения каждой сферы из Таблицы 1 со списком самых контексто-близких к ней слов. После небольшой ручной фильтрации был составлен окончательный список всех сфер и тех слов-категорий, которые в нее входят. При этом важно учитывать, чтобы каждая новая категория представляла собой социальный процесс или явление, чтобы к ней можно было отнести не одно, а ряд сообщений об уже конкретных социальных событиях.

В научных трудах при подготовке систем классификации предполагалось составление таких категорий, которым возможно дать конкретную эмоциональную окраску – либо позитивную, либо негативную: например, в сфере изучения конфликтов. В работе также перенята данная особенность, и к каждой из вновь образованных категорий добавлена эмоциональная составляющая. Некоторые из категорий изначально предпола-

гают ту или иную эмоциональную окраску – например, бедность и преступление – это негативные явления, а сотрудничество и импортозамещение – позитивные. Некоторые категории представлены более обобщенно путем добавления приставок «развитие», «достижения», «проблемы» – например, «развитие образования». Итоговый список категорий представлен в результатах.

Следующим важным шагом является классификация новостного массива по выбранным категориям. Для этой цели выбран подход тематического моделирования с помощью Zero-Shot классификации. Тематическое моделирование – одно из современных приложений машинного обучения к анализу текстов, которое позволяет определить, к каким тематикам относится каждый документ и какие слова образуют каждую из них [11]. При этом одна из главных трудностей – формирование обучающих данных под каждую категорию, что в рамках исследования очень проблематично, так как сделать качественную разметку данных под каждую категорию не представляется возможным в связи с несбалансированностью всей выборки. В связи с этим выбран подход Zero-Shot классификации для тематического моделирования, который позволяет обойти все эти ограничения.

Zero-Shot классификация текста – задача классификации, в которой модели могут классифицировать текст, не обучаясь при этом на наборе данных, созданном для этой задачи классификации [12]. Модель способна предсказать, к какому из предложенных классов вероятнее всего относится текст на основе анализа ключевых слов и контекста. Для классификации новостей по новым категориям выбрана готовая мультязычная модель тематического моделирования с помощью подхода Zero-Shot классификации [13]. Данный этап выполнен на языке программирования Python с помощью библиотеки transformers.

### **3.3. Подсчет результатов заполнения матричного классификатора**

Третьим этапом в классической методике ивент-анализа является подсчет результатов заполнения матричного классификатора. Количественные данные для аналитического сравнения качественных характеристик ситуации выражают через определение их относительных значений, а также через построение индексов. Определение относительных значений целесообразно, если требуется статистическая обработка данных, особенно при использовании событийной информации. Построение индекса применяется для объединения различных количественных данных в единый комплексный показатель с целью последующего отслеживания ситуации [1].

Поскольку основной этап ивент-анализа состоит из совокупности методов анализа, то для каждого из них требуется свой подход к подсчету итоговых результатов. В большинстве случаев предполагается подсчет абсолютного количества новостных сообщений по каждой тематике за определенный период времени.

### **3.4. Проведение аналитического исследования**

Самый трудоемкий этап исследования – проведение аналитических сравнений полученных значений показателей, описывающих типы событий или их аспекты на различных временных этапах. Весь анализируемый период поделен на интервалы, а события, наблюдаемые в каждом из них, сравнены по различным критериям в рамках данных периодов.

Методы анализа выбраны таким образом, чтобы после выполнения данного этапа можно было оценить параметры, предложенные Ч. Макклелландом, в соответствии с которыми осуществляется обработка данных, см. Табл. 2 [8].

Таблица 2. Параметры ивент-анализа

Параметр	Вопрос, на который отвечает параметр
Оценка сюжета	Что происходит?
Оценка субъекта-инициатора	Кто за этим стоит?
Оценка объекта	По отношению к кому?
Оценка времени события	Когда?

Этот этап разделен на три основные части: анализ графов, анализ временных рядов, выделение ключевых слов. Выбор этих подходов обоснован параметрами, выделенными выше. Графовый анализ используется для оценки объекта («по отношению к кому/чему?») и частично субъектов-инициаторов («кто/что за этим стоит?»), анализ временных рядов – для оценки времени события («когда?»), выделение ключевых слов – для оценки сюжета («что происходит?»).

Графовый анализ в рамках ивент-анализа описывается очень редко, однако он наилучшим образом позволяет увидеть силу взаимосвязей между различными социальными категориями [14]. Граф представляет собой пару  $G = (V, E)$ , где  $V$  – множество, элементы которого называются вершинами, а  $E$  – набор неупорядоченных пар вершин, элементы которых называются ребрами [15]. Графы взаимосвязей построены на основе появления категорий в контексте одной новости. Узлы – сами категории, а ребра показывают наличие связи между ними в рамках одного новостного текста. Ребра также имеют веса, и чем больше новостей связывает две категории, тем больше значения этих весов. Чтобы проследить динамику взаимосвязей между социальными событиями и явлениями, исследователю представляется возможность построить графы за годовой интервал и сравнивать значения в разрезе четырех лет.

Для алгоритма ивент-анализа вычисление центральности также является важным шагом, поскольку позволяет выявить самые межотраслевые и связующие категории. Центральность – это один из важнейших показателей, который используется для того, чтобы показать релевантность или структурную важность узла в сети. Для каждой категории в разрезе лет возможно рассчитать показатель центральности по степени: согласно этому подходу, большее значение центральности получают узлы с большим количеством связей, а сам показатель считается как отношение количества узлов, с которым у рассматриваемого узла есть связи, к общему количеству узлов [15]. Графовый анализ проведен на языке программирования Python с использованием библиотек `networkx` и `plotly` для визуализации результатов.

Для оценки времени события выбран метод анализа временных рядов. Временные ряды построены отдельно для каждой категории, где на оси  $x$  – дата, на оси  $y$  – количество новостей по данной категории за эту дату. В научных работах по ивент-анализу временные ряды, как правило, используются для построения частотного распределения классификаторов по заданным периодам [6]. Однако такой подход может быть не очень информативным, и здесь главная цель – определить ключевые события, в результате которых появились экстремальные точки или выбивающиеся из общего ряда периоды, а для решения этого вопроса ручная обработка может быть трудоемкой и неэффективной. Из-за этого было решено добавить этап сегментации временных рядов, в результате которого такие события будут определяться автоматически и показывать изменения более наглядно.

Для этой цели использован алгоритм `Pruned Exact Linear Time (PELT)`. Этот алгоритм ищет набор точек «перегиба» для заданного временного ряда таким образом, чтобы их количество и местоположение минимизировали заданную «стоимость» сегментации. Основные шаги алгоритма заключаются в определении функции «стоимости» для сегмента, затем итерации по всем возможным начальным и конечным точкам сегмента и проверке того, уменьшает ли разделение на новые сегменты значение функции стоимости по сравнению с сегментом без разделения. Одним из часто исполь-

зуемых подходов к идентификации нескольких точек изменений является сведение к минимуму суммы, представленной в формуле ниже [16].

$$\sum_{i=1}^{m+1} \left[ C_{(y_{(\tau_{i-1}+1):\tau_i})} + \beta f(m) \right]$$

Здесь  $C$  – функция «стоимости» сегмента,  $\tau$  – точка «перегиба»,  $m$  – общее количество точек «перегиба»,  $\beta f(m)$  – регуляризатор для предотвращения переобучения.

Выявлено, что некоторые точки и периоды сниженной новостной активности на временных рядах совпадают с праздничными днями и выходными. Поэтому данные по этим дням удалены из выборки, чтобы временные ряды получились более сглаженными и не находилось ложных корреляций. Данный этап реализован на языке программирования Python с использованием библиотеки `ruptures` для сегментации и библиотеки `plotly` для визуализации результатов.

Еще один метод, позаимствованный из научных работ – это кластеризация временных рядов, которая позволяет обнаруживать корреляции между временными рядами, представляющими различные категории. Для проведения кластеризации построены временные ряды не по дням, а по месяцам, так как это позволяет сгладить временные ряды и избавиться от шума. Для данной задачи взят метод  $k$ -средних, по которому считается евклидово расстояние между векторами несмещенных временных рядов, для них ищутся центроиды и наконец определяются кластеры в результате перемещения центроид по количеству итераций [17]. Сам алгоритм  $k$ -средних кластеризует временные ряды, построенные по месяцам за интересующий год, так как это помогает определить корреляции между ними лучше, чем по датам.

Так как оптимальное количество кластеров заранее не известно, необходимо определять это значение по методу «локтя» и по метрике силуэта [17]. Метод локтя показывает оптимальное количество кластеров по следующему принципу: если после визуального «локтя» на графике идет резкое убывание общей ошибки, то такое количество считается оптимальным, но если кластеров много, то ошибка будет минимизироваться, но не будет смысла в кластеризации в принципе. Считается сумма квадратов расстояний от объектов до центра кластера – иначе говоря, ошибок. По методу силуэта оптимальное количество кластеров – пиковое значение на графике, после которого идет резкий спад. Метрика считает для каждого объекта среднее расстояние между ним и объектами внутри кластера (а) и между ним и объектами в ближайшем кластере (b). Чем больше нормализованное  $b-a$ , тем лучше. Для реализации этой задачи использовался алгоритм  $k$ -средних, реализованный на языке программирования Python с использованием библиотек `sklearn` для нормализации временных рядов, `tslearn` для кластеризации и `plotly` для визуализации результатов.

Для оценки сущности событий, участвующих в них субъектов и определения более содержательной корреляции между категориями, добавлен еще один метод – выделение ключевых терминов из новостных заголовков для экстремальных значений на временных рядах. Под такими значениями подразумеваются даты, совпадающие с максимальной частотной активностью по конкретным категориям, а также точки «перегиба», которые обозначают появление нового сегмента на временном ряде: некоторых переломных событий, в результате которых новостная активность в определенные периоды стала резко отличаться.

Для реализации данного шага использовался подход для определения ключевых терминов в тексте, основанный как на лингвистических инструментах (определение части речи, токенизация), так и на относительной частотности [18]. Для каждой экстремальной даты в рамках одной категории и временного периода выводится список наиболее частотных ключевых терминов, которые позволяют делать более информативные выводы о ключевых событиях, их акторах и эмоциональной составляющей, которые повлияли на изменение новостной активности. Для этой задачи взята модель из библиотеки `TermExtractor`, код реализован на языке программирования Python.



### 3.5. Верификация результатов и формулирование выводов

Поскольку в данном варианте методики не предполагается прогнозирование программно-математическими методами, то этот этап является полностью ручным, а главный результат реализуемой методики – проанализированные данные, оформленные в виде интерактивных графиков на итоговой аналитической панели, реализованной на языке программирования Python с использованием библиотеки dash [19]. Выбирая различные фильтры, такие как интересующий временной период и категория, можно сравнивать результаты анализа по годам, определять изменение поведения анализируемых объектов в динамике и на основе этого строить социальные прогнозы.

Выполнение финального этапа методики ивент-анализа – валидация результатов и социальное прогнозирование – проводится на основании итоговой аналитической панели, которая включает в себя:

- совокупность интерактивных графиков, построенных по перечисленным ранее методам: графовый анализ, сегментация и кластеризация временных рядов, распределение ключевых терминов;
- выбор фильтров – «Категория» и «Год»;
- статистику по абсолютному и относительному количеству новостей по «Категории» и «Году», а также центральные категории за взятый период.

## 4. Результаты экспериментов

Фрагмент собранного массива данных с российского информационно-новостного ресурса представлен на рисунке (рис. 1).

дата	рубрика	заголовок	новости, упомянутые в контексте
2022/06/28	Россия	Свердловский губернатор вмешался в конфликт коммунистов с Ельцин Центром	['Никита Михалков потребовал признать Ельцин Центр иноагентом', 'Российский депутат предложил Шойгу закрыть Ельцин Центр']
2020/10/04	Экономика	Названы самые популярные валюты для сбережений у россиян	['Россияне решили массово забрать валюту из банков']
2023/10/10	Мир	ХАМАС начало обстрел аэропорта Тель-Авива	['В Израиле опровергли проникновение боевиков на север страны']
2023/08/18	Моя страна	В Северной Осетии появились пляжные спортивные площадки	['Во Владикавказе появятся скульптуры двух героев романа «12 стульев»']
2021/06/15	Бывший СССР	Зеленский оценил позицию НАТО насчет членства Украины	['В НАТО подтвердили вступление в альянс Украины и Грузии']

Рисунок 1. Фрагмент массива новостей

Результаты по второму и третьему этапу методики объединены на одном рисунке – продемонстрирован список сформированных социальных категорий, а также их абсолютное количество в тысячах в разрезе четырех лет (рис. 2).

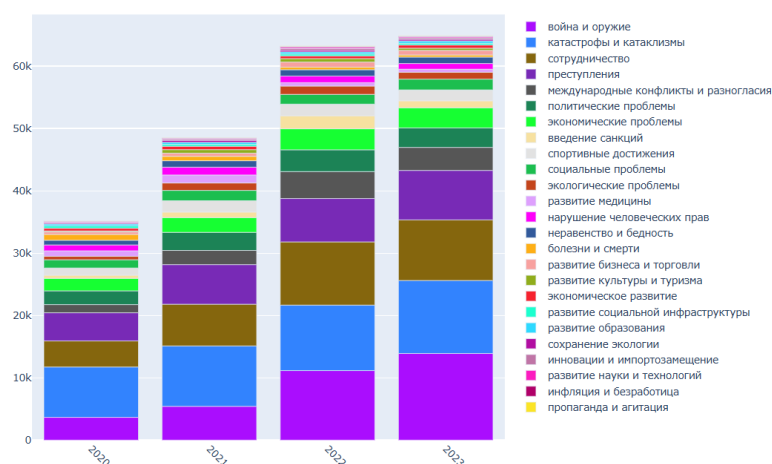


Рисунок 2. Распределение количества новостей по категориям и годам

На рисунке представлен граф, построенный для 2023 года: при наведении на узлы на аналитической панели можно увидеть название соответствующих категорий (рис. 3). Размер узлов-категорий пропорционален количеству новостей по данной категории за взятый период. Толщина ребер пропорциональна количеству новостей, в которых упоминались пары новостей, принадлежащие соответствующим категориям.

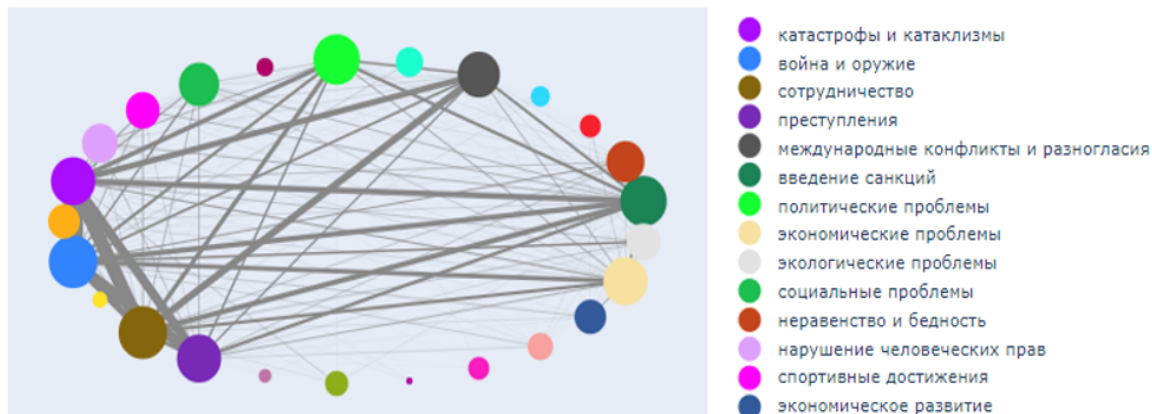


Рисунок 3. Граф категорий за 2023 год

Результат определения топ-5 центральных узлов-категорий для 2023 года представлен в таблице, см. Табл. 3.

Таблица 3. Топ-5 самых центральных узлов-категорий за 2023 год

Категория	Центральность
Катастрофы и катаклизмы	0.958
Сотрудничество	0.958
Введение санкций	0.917
Политические проблемы	0.917
Преступления	0.875

Для демонстрации работы методики в качестве примера взята категория «развитие науки и технологий», по которой в дальнейшем будет составлен социальный прогноз. На рисунке представлен пример графика для сегментации временного ряда по категории «развитие науки и технологий» за 2023 год (рис. 4), где на оси x указана дата, а на оси y — количество новостей. В данном случае выбрано объединение в сегменты размером не меньше недели. Параметр регуляризации подбирается экспериментально, чтобы предотвратить переобучение алгоритма. Один из популярных подходов, который использован в работе — брать регуляризацию как два логарифма от длины исходного ряда. Чем меньше значение параметра регуляризации, то есть меньше «штраф», тем больше сегментов выделяется.

Каждый сегмент представляет собой отдельный период особой новостной активности: есть сегменты с маленькой амплитудой разброса частотной активности, а есть с высокой. Можно предположить, что в первом случае не было сильных инфоповодов. Вертикальной красной полосой отмечены те точки перегиба, то есть даты, которые являются границами сегмента и обозначают изменение поведения временного ряда на определенном промежутке. Помимо этого, красным цветом выделены топ-5 точек с максимальными значениями новостной активности за весь взятый период. Поскольку график интерактивный, нет необходимости выводить даты, которые совпадают с точками максимума, и их можно увидеть, наведя курсор на соответствующую точку. По данному графику можно определить те даты, в которые произошли самые интересные с точки зрения анализа события, которые привели к появлению инфоповода. Более информативное описание этих событий в указанные даты приведено на этапе анализа ключевых слов.

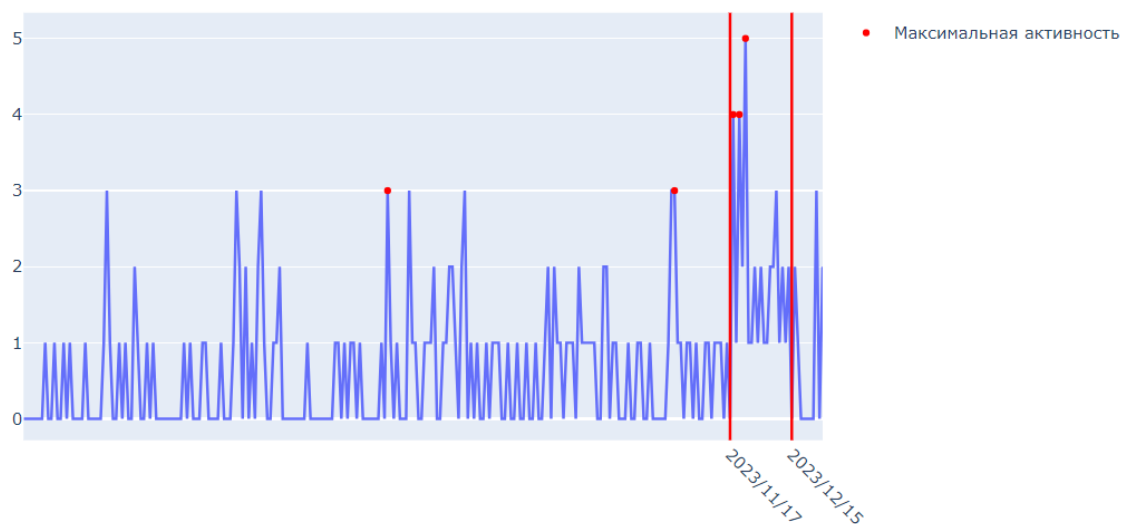


Рисунок 4. График с результатами сегментации временного ряда для категории «развитие науки и технологий» за 2023 год

В результате работы алгоритма кластеризации на графике выводится тот кластер временных рядов, который содержит выбранную в фильтре аналитической панели категорию. На рисунке представлен пример кластера для категории «развитие науки и технологий» за 2023 год), где на оси x указана дата, а на оси y – количество новостей (рис. 5). Благодаря построению данного графика возможно проследить в разрезе лет то, как меняется корреляция временных рядов различных категорий, определить причины, по которым поведение одних категорий в определенный период схоже с выбранной, а в другой период разительно отличается.

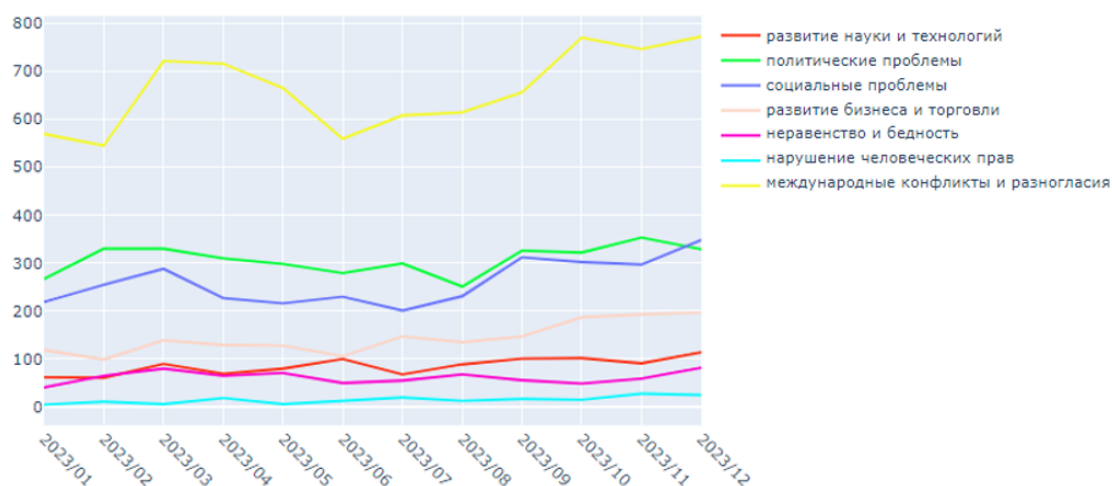


Рисунок 5. График с результатами кластеризации временного ряда для категории «развитие науки и технологий» за 2023 год

Метод, применяемый для более информативного описания событий, произошедших в даты, совпадающий с точками экстремума или перегиба на временном ряде – выделение ключевых терминов. На рисунке представлен результат работы алгоритма для категории «развитие науки и технологий» за 2023 год, где на оси x указана дата, на оси y – суммарное количество ключевых терминов (рис. 6).

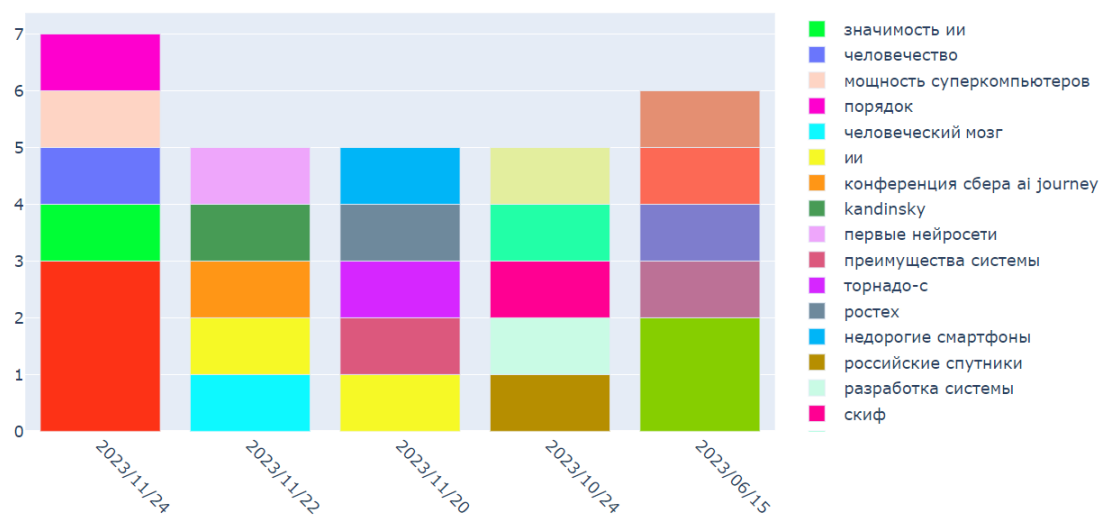


Рисунок 6. График с результатами определения ключевых терминов для категории «развитие науки и технологий» за 2023 год

Аналитическая панель реализована на языке программирования Python с использованием библиотеки dash, результаты выводятся на локальном сервере. Пример аналитической панели для категории «развитие науки и технологий» за 2023 год представлен на рисунке (рис. 7).

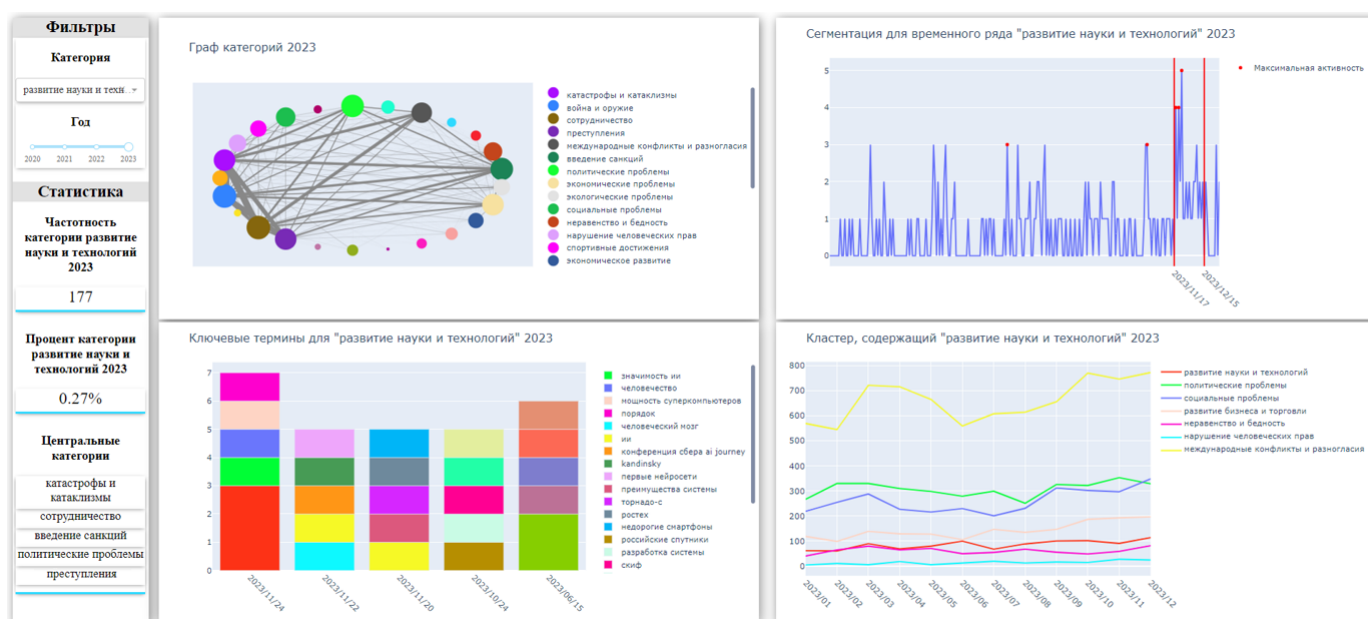


Рисунок 7. Результат построения итоговой аналитической панели с фильтрами «Категория» – «развитие науки и технологий», «Год» – 2023.

В качестве метода построения социального прогноза выбрано написание аналитической записки по тенденциям развития сферы «развитие науки и технологий» в России на ближайшие годы на основе анализа ретроспективных данных из аналитической панели. Для удобства аналитическая записка представлена в виде таблицы ниже с анализом тенденций, выдвижением наиболее вероятных сценариев и рекомендаций как для государства, так и для бизнеса, см. Табл. 4.

Таблица 4. Социальный прогноз по категории «Развитие науки и технологий» в России

Тенденции по годам			Прогнозы и рекомендации		
	Тенденции	Источники	Прогноз	Государство	Бизнес
2020	<p><u>Медицина:</u> развитие вакцинационных технологий.</p> <p><u>Экология:</u> технологии для экологически чистого строительства, инновации в материаловедении, освоение Арктической зоны.</p>	<p><u>Ключевые тер-мины и сегментация времен-ных рядов (ВР):</u> вакцины, зондирование ионосферы, морозостойкий бетон.</p> <p><u>Графы и кла-стеризация ВР:</u> экологическое развитие, раз-витие медици-ны, сотрудни-чество.</p>	Развитие эко-логически чистых тех-нологий, а также сферы здравоохра-нения.	Обеспечение до-статочного фи-нансирования для научных ис-следований в сферах экологии и медицины.	Инвестирование в исследования по усовершен-ствованию ин-фраструктуры и подходов в сфе-ре здравоохра-нения. Активное участие в эколо-гических ини-циативах и про-ектах по устой-чивому разви-тию.
2021	<p><u>Образование:</u> цифровизация образовательных процессов, разви-тие технопарков и инновацион-ных кластеров.</p> <p><u>Медицина:</u> Рас-ширение приме-нения ИИ для диагностики и лечения, про-должение работы по усовершен-ствованию и рас-пространению российской вак-цины.</p>	<p><u>Ключевые тер-мины и сегментация ВР:</u> вак-цины, техно-парки, искус-ственный ин-теллект.</p> <p><u>Графы и кла-стеризация ВР:</u> развитие соци-альной инфра-структуры, раз-витие медици-ны, развитие образования.</p>	Ускоренное развитие цифровых технологий, включая ис-кусственный интеллект, в ответ на по-требности в цифровой трансформа-ции экономи-ки и обще-ства.	Формирование национальных стратегий циф-ровизации, включая разра-ботку законода-тельства и ин-фраструктуры для поддержки цифровых техно-логий. Поддерж-ка исследований в области искус-ственного интел-лекта и его при-ложений.	Инвестирование в разработку цифровых платформ, раз-витие иннова-ционных про-дуктов на основе искусственного интеллекта. Ак-тивное исполь-зование цифро-вых технологий для улучшения эффективности бизнес-процессов.

2022	<p><u>Промышленность:</u> развитие высокотехнологичных отраслей, включая ВПК, разработка новых технологий для повышения производительности.</p> <p><u>Инновации и сотрудничество:</u> укрепление технологического суверенитета через развитие отечественных технологий.</p>	<p><u>Ключевые тер-мины и сегментация ВР:</u> российские беспилотники и радиолокационные спутники, технологический суверенитет России, искусственный интеллект.</p> <p><u>Графы и кластеризация ВР:</u> экономическое развитие, развитие бизнеса и торговли, инновации и импортозамещение.</p>	Активное развитие международных научных и технологических инициатив, особенно с участием стран Восточной Азии.	Поддержка международных научных проектов и программ обмена, стимулирование технологического трансфера и совместных исследовательских проектов.	Активное участие в международных научных и технологических партнерствах, создание совместных исследовательских и инновационных лабораторий с зарубежными партнерами.
2023	<p><u>Безопасность:</u> развитие технологий в области безопасности и обороны, включая кибербезопасность и радиолокационные технологии.</p> <p><u>ИТ:</u> инвестиции в разработку искусственного интеллекта, особенно в областях кибербезопасности и автоматизации процессов.</p>	<p><u>Ключевые тер-мины и сегментация ВР:</u> искусственный интеллект, суперкомпьютеры, российские радиолокационные спутники.</p> <p><u>Графы и кластеризация ВР:</u> политические проблемы, введение санкций, международные конфликты и разногласия.</p>	Активное развитие отечественных технологий, направленных на укрепление национальной безопасности и суверенитета.	Повышение государственного финансирования в области технологий безопасности, киберзащиты и разработки высокотехнологичных продуктов. Внедрение мер поддержки национальных производителей и исследовательских центров.	Инвестирование в разработку и инновации, направленные на укрепление технологической базы страны. Сотрудничество с государственными заказчиками и научными учреждениями.

## 5. Обсуждение результатов

Ивент-анализ является адаптируемой технологией для использования в качестве инструмента социального прогнозирования, поскольку выбор системы классификации данных и методов на этапе аналитического исследования зависит от поставленных задач. Основным результатом работы стала авторская модификация алгоритма ивент-анализа, в частности, для его использования в качестве инструмента социального прогнозирования. Для выдвижения конкретных прогнозов необходима экспертная оценка, однако она должна опираться на проанализированные с различных сторон данные, что подтверждает актуальность и значимость проведенного исследования.

По сравнению с классической методикой ивент-анализа, где главным методом на этапе аналитического сравнения, как правило, выступает вычисление частотного распределения категорий в определенные периоды, добавлены дополнительные подходы, которые позволяют рассмотреть взаимосвязи и тенденции более подробно. Модифицированные этапы сделали возможным более обоснованно отвечать на главные вопросы, поставленные в классической методике: об оценке сюжета, субъектов-инициаторов, объектов, времени события.

Выводы, сделанные на основе данных из интерактивных графиков, позволили определить тенденции поведения сферы «развитие науки и технологий» по соответствующим годам. Достоверность полученных результатов оценена посредством ручной верификации выявленных тенденций путем сравнения с ранее составленными экспертами выводами. Это доказывает, что адаптированная методика ивент-анализа обладает достаточной степенью достоверности и полноты, что позволяет использовать ее возможности как одного из этапов в проведении масштабных социальных исследований и прогнозов. Разработанные на ее основе наиболее вероятные сценарии развития объекта социального прогнозирования и рекомендации позволят государственным органам вовремя предпринимать меры по поддержке актуальных исследований частных организаций, а бизнесу – искать новые и прибыльные ниши для развития. Это приведет к синхронизации действий государства и бизнеса, их взаимовыгодному сотрудничеству и ускоренному развитию различных сфер жизнедеятельности.

Основное преимущество разработанной методики в том, что ее можно адаптировать под некоторые задачи социального прогнозирования, где требуется классификация данных по различным категориям и их разносторонний анализ как в статическом, так и в динамическом состоянии. В качестве дальнейшей работы планируется увеличение количества источников сбора данных, улучшение качества классификации данных по категориям и усовершенствование методов, используемых на этапе аналитического сравнения.

## **6. Выводы и заключение**

В ходе проведенного исследования разработана и протестирована адаптированная методика ивент-анализа, которая продемонстрировала свою эффективность в контексте социального прогнозирования. Основные выводы, сделанные в результате работы, можно обобщить следующим образом. Адаптированная методика ивент-анализа показала высокую степень универсальности, позволяя применять ее в различных областях социального исследования. Она может быть адаптирована под специфические задачи, что делает ее подходящей для анализа как краткосрочных, так и долгосрочных социальных явлений.

Внедрение новых аналитических инструментов, таких как анализ графов и временных рядов, значительно повысило уровень достоверности получаемых данных. Ручная верификация выявленных тенденций подтвердила, что результаты анализа соответствуют реальным событиям и их контексту.

Создание интерактивных графиков с возможностью фильтрации данных улучшило восприятие информации и обеспечило возможность динамического сравнения различных сценариев. Это позволяет не только визуализировать результаты, но и углубить анализ, выявляя скрытые взаимосвязи и закономерности.

Результаты исследования могут быть использованы как для государственных органов, так и для бизнеса. Сценарии развития событий, сформулированные на основе полученных данных, могут служить основой для принятия стратегических решений. Это способствует синхронизации действий различных субъектов и обеспечению взаимовыгодного сотрудничества.

В дальнейшем планируется расширение источников данных, включая новые информационные потоки и социальные медиа. Это позволит улучшить классификацию событий и усовершенствовать аналитические методы. Также стоит рассмотреть возможность интеграции машинного обучения для автоматизации прогнозов и повышения его точности.

Таким образом, адаптированная методика ивент-анализа представляет собой мощный инструмент для социального прогнозирования, который сочетает в себе традиционные подходы с современными аналитическими технологиями. Ее использование может значительно улучшить качество прогнозов и способствовать более эффективно-



му взаимодействию между государственными структурами и бизнесом в области научных исследований и технологий.

## Список литературы

1. Мальцева, А. В. Использование методики event-анализа для изучения процессов на рынке труда / А. В. Мальцева и др. // Вестник евразийской науки. – 2012. – № 3 (12). – С. 12.
2. Новоселов, М. В. Горизонты социологического применения EVENT-анализа / М. В. Новоселов // Социальные и гуманитарные науки: теория и практика. – 2018. – № 1 (2). – С. 497-506.
3. Азаров, А. А. Predictor Mining: применение методов интеллектуального анализа данных в задачах социального компьютинга / А. А. Азаров // Информатика и автоматизация. – 2013. – № 26. – С. 136-161.
4. Кудрявцева, В. И. Особенности социального прогнозирования в международных отношениях / В. И. Кудрявцева // Вестник БГУ. – 2004. – № 2. – С. 67-71.
5. Ortega, J. L. «How do media mention research papers? Structural analysis of blogs and news networks using citation coupling» / J. L. Ortega // Journal of Informetrics. – 2021. – № 15 (3). – P. 101175. – doi.org/10.1016/j.joi.2021.101175.
6. Blokh, I. «News clustering based on similarity analysis» / I. Blokh, V. Alexandrov // Procedia Computer Science. – 2017. – № 122. – P. 715-719. – doi.org/10.1016/j.procs.2017.11.428.
7. Нехамкин, А. Н. Социальное прогнозирование: достижения, недостатки, пути совершенствования / Нехамкин А. Н., Нехамкин В. А. // Вестник Московского государственного областного университета. Серия: Философские науки. – 2020. – №. 2. – С. 57-68.
8. McClelland, C. A. Let the user beware / C. A. McClelland // International Studies Quarterly. – 1983. – № 27(2). – P. 169-177. – doi: 10.2307/2600544.
9. Лента.ру: сайт / учредитель ООО «МИНС». – Москва. – Обновляется в течение суток. – URL: <https://lenta.ru/> (дата обращения: 20.08.2024).
10. Johnson, S. J. A detailed review on word embedding techniques with emphasis on word2vec / S. J. Johnson, M. R. Murty, I. Navakanth // Multimedia Tools and Applications. – 2024. – № 13 (83). – С. 37979-38007.
11. A, Daud. Knowledge discovery through directed probabilistic topic models: a survey / Daud A., Li J., Zhou L., Muhammad F. // Frontiers of Computer Science in China. – 2010. – № 2 (4). – P. 280-301. – doi:10.1007/s11704-009-0062-y.
12. Z, Ji. Zero-shot classification with unseen prototype learning / Ji Z., Cui B., Yu Y., Pang Y., Zhang Z. // Neural computing and applications. – 2023. – P. 1-11. – doi: 10.1007/s00521-021-05746-9
13. Hugging Face: сайт / MoritzLaurer/mDeBERTa-v3-base-mnli-xnli. – Обновляется в течение суток. – URL: <https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-mnli-xnli> (дата обращения: 20.08.2024).
14. Ulizko, M. S. Visual analytics of twitter and social media dataflows: a casestudy of covid-19 rumors / M. S. Ulizko, E. V. Antonov, M. A. Grigorieva, E. S. Tretyakov, R. R. Tukumbetova, A. A. Artamonov // Scientific Visualization. – 2021. – № 4 (13). – P. 144-163.
15. Camacho, D. The four dimensions of social network analysis: An overview of research methods, applications, and software tools / D. Camacho, Á. Panizo-Lledot, G. Bello-Orgaz, A. Gonzalez-Pardo, E. Cambria // Information Fusion. – 2020. – № 63. – P. 88-120. – doi: 10.1016/j.inffus.2020.05.009.
16. Killick, R. Optimal detection of changepoints with a linear computational cost / R. Killick, P. Fearnhead, I. A. Eckley // Journal of the American Statistical Association. – 2012. – № 107(500). – P. 1590-1598. – doi: 10.1080/01621459.2012.737745.



17. Ulizko, M. S. Clustering Thematic Information in Social Media / M. S. Ulizko, A. A. Ar-tamonov, J. E. Fomina, E. V. Antonov, R. R. Tukumbetova // Proceedings of the International Conference on Computer Graphics and Vision «Graphicon». – 2022. – № 32. – P. 403-413.
18. Мотовских, Л. В. Выделение ключевых слов для классификации текстов / Л. В. Мотовских // Вестник Московского государственного лингвистического университета. Гуманитарные науки. – 2020. – № 9 (838). – С. 235-242.
19. Dabbas, E. Interactive Dashboards and Data Apps with Plotly and Dash: Harness the power of a fully fledged frontend web framework in Python–no JavaScript required. / E. Dabbas. – Packt Publishing Ltd, 2021. – 336 p. – ISBN 978-1-80056-891-4.

# Event Analysis: Application in Social Forecasting

T. V. Korenkova<sup>1</sup>, A. A. Artamonov<sup>2</sup>, M. S. Ulizko<sup>3</sup>

National Research Nuclear University MEPhI, Moscow, Russia

<sup>1</sup> ORCID: 0009-0000-6257-8500, [korenkova.tanya@mail.ru](mailto:korenkova.tanya@mail.ru)

<sup>2</sup> ORCID: 0000-0002-9140-5526, [aartamonov@kaf65.ru](mailto:aartamonov@kaf65.ru)

<sup>3</sup> ORCID: 0000-0003-2608-8330, [mulizko@kaf65.ru](mailto:mulizko@kaf65.ru)

## **Abstract**

Monitoring the interrelationships between social events and phenomena and forecasting the dynamics of their changes are necessary in the conditions of instability of the modern world. There are many separate methods of analysis for social forecasting, however, for this research, the method of event analysis has been chosen, which is insufficiently considered in the scientific literature within the framework of this task, but has high potential. The purpose of the article is to adapt the event analysis methodology for its use as a social forecasting tool. The main data for the study was collected from the Russian information and news resource in the period 2020-2023. Based on the classical methodology of event analysis, the classifiers presented in this paper in the form of social spheres are defined in the research. As part of the analytical comparison stage, a graph analysis was carried out (graphs of relationships between categories were constructed, central nodes-categories were identified); time series analysis was performed (segmentation of time series by the PELT algorithm, clustering of time series by the k-means algorithm); key terms for press events were defined. The final product is an analytical dashboard with filters, statistics and interactive graphs. The analytical dashboard makes it possible to compare data in a static and dynamic state, to draw conclusions about the past and future states of objects of social forecasting. The main result of the research is the event analysis methodology developed by the author, which can be used for a comprehensive analysis of news streams, adapted to the necessary categories representing a certain entity or sphere, and applied in various social organizations or monitoring services.

**Keywords:** social forecasting, event analysis, news analysis, graph analysis, centrality calculation, time series analysis, time series segmentation, time series clustering, PELT.

## **References**

1. Maltseva, A. V. Using the event analysis technique to study processes in the labor market / A. V. Maltseva et al. // Bulletin of Eurasian Science. - 2012. - No. 3 (12). - P. 12.
2. Novoselov, M. V. Horizons of sociological application of EVENT analysis / M. V. Novoselov // Social and humanitarian sciences: theory and practice. - 2018. - No. 1 (2). - P. 497-506.
3. Azarov, A. A. Predictor Mining: application of data mining methods in social computing tasks / A. A. Azarov // Computer Science and Automation. - 2013. - No. 26. - P. 136-161.
4. Kudryavtseva, V. I. Features of social forecasting in international relations / V. I. Kudryavtseva // Bulletin of BSU. - 2004. - No. 2. - P. 67-71.
5. Ortega, J. L. «How do media mention research papers? Structural analysis of blogs and news networks using citation coupling» / J.L Ortega // Journal of Informetrics. - 2021. - No. 15 (3). - P. 101175. - doi.org/10.1016/j.joi.2021.101175.
6. Blokh, I. «News clustering based on similarity analysis» / I. Blokh, V. Alexandrov // Procedia Computer Science. - 2017. - No. 122. - P. 715-719. - doi.org/10.1016/j.procs.2017.11.428.

7. Nekhamkin, A. N. Social forecasting: achievements, shortcomings, ways of improvement / Nekhamkin A. N., Nekhamkin V. A. // Bulletin of the Moscow State Regional University. Series: Philosophical Sciences. - 2020. - No. 2. - P. 57-68.
8. McClelland, C. A. Let the user beware / C. A. McClelland // International Studies Quarterly. - 1983. - No. 27(2). - P. 169-177. - doi: 10.2307/2600544.
9. Lenta.ru: website / founder OOO MINS. - Moscow. - Updated within 24 hours. - URL: <https://lenta.ru/> (date accessed: 20.08.2024).
10. Johnson, S. J. A detailed review on word embedding techniques with emphasis on word2vec / S. J. Johnson, M. R. Murty, I. Navakanth // Multimedia Tools and Applications. - 2024. - No. 13 (83). - P. 37979-38007.
11. A, Daud. Knowledge discovery through directed probabilistic topic models: a survey / Daud A., Li J., Zhou L., Muhammad F. // Frontiers of Computer Science in China. - 2010. - No. 2 (4). - P. 280-301. - doi:10.1007/s11704-009-0062-y.
12. Z, Ji. Zero-shot classification with unseen prototype learning / Ji Z., Cui B., Yu Y., Pang Y., Zhang Z. // Neural computing and applications. - 2023. - P. 1-11. - doi: 10.1007/s00521-021-05746-9
13. Hugging Face: website / MoritzLaurer/mDeBERTa-v3-base-mnli-xnli. - Updated within 24 hours. - URL: <https://huggingface.co/MoritzLaurer/mDeBERTa-v3-base-mnli-xnli> (date of access: 20.08.2024).
14. Ulizko, M. S. Visual analytics of twitter and social media data flows: a case study of covid-19 rumors / M. S. Ulizko, E. V. Antonov, M. A. Grigorieva, E. S. Tretyakov, R. R. Tukumbetova, A. A. Artamonov // Scientific Visualization. - 2021. - No. 4 (13). - P. 144-163.
15. Camacho, D. The four dimensions of social network analysis: An overview of research methods, applications, and software tools / D. Camacho, Á. Panizo-LLedot, G. Bello-Organ, A. Gonzalez-Pardo, E. Cambria // Information Fusion. - 2020. - No. 63. - P. 88-120. - doi: 10.1016/j.inffus.2020.05.009.
16. Killick, R. Optimal detection of changepoints with a linear computational cost / R. Killick, P. Fearnhead, I. A. Eckley // Journal of the American Statistical Association. - 2012. - No. 107(500). - P. 1590-1598. - doi: 10.1080/01621459.2012.737745.
17. Ulizko, M. S. Clustering Thematic Information in Social Media / M. S. Ulizko, A. A. Artamonov, J. E. Fomina, E. V. Antonov, R. R. Tukumbetova // Proceedings of the International Conference on Computer Graphics and Vision «Graphicon». - 2022. - No. 32. - P. 403-413.
18. Motovskikh, L. V. Keyword extraction for text classification / L. V. Motovskikh // Bulletin of the Moscow State Linguistic University. Humanities. - 2020. - No. 9 (838). - P. 235-242.
19. Dabbas, E. Interactive Dashboards and Data Apps with Plotly and Dash: Harness the power of a fully fledged frontend web framework in Python—no JavaScript required. / E. Dabbas. - Packt Publishing Ltd, 2021. - 336 p. - ISBN 978-1-80056-891-4.