МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Кафедра «Прикладная математика и информатика»
(наименование)

09.04.03 Прикладная информатика
(код и наименование направления подготовки / специальности)

Технология бизнес-анализа
(направленность (профиль) / специализация)

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

на тему «Моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных»

Обучающийся А.О. Пилипчук
(Инициалы Фамилия) (личная подпись)

Научный руководитель доцент, канд. пед. наук, О.М. Гущина
(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Тольятти 2025

Содержание

Введение

Актуальность работы. «В условиях развития цифровой экономики и увеличения объема информации, а также цифрового следа, который оставляют клиенты, задача принятия решений в организациях, становится сложной задачей и требует использование новых методов и инструментов анализа этого большого количества данных для принятия верных и оперативных решений» [33].

Современные методы анализа и прогнозирования временных рядов часто недостаточно точно учитывают сложные нелинейные взаимосвязи и шум в статистических данных, что приводит к снижению достоверности прогноза. Возникает необходимость разработки интеллектуальных подходов к обработке статистических данных, способных выявлять скрытые тренды, учитывать многомерные зависимости и адаптироваться к изменениям во временных рядах, с целью повышения точности и надежности прогнозных моделей.

«Требуется математическое, информационно-алгоритмическое и техническое обеспечение, позволяющее вести сбор, хранение, прогнозирование и интеллектуальную обработку статистических данных, а также возможность построения различных трендов больших объемов разных данных, генерируемых как внутри организации, так и получение из внешних источников. Наличие полной информации о клиенте организации позволит установить явные и скрытые взаимосвязи между персональными и групповыми профилями, обнаружения причинно-следственных зависимостей в информационном поле, т. е. позволит спрогнозировать поведение клиента организации для минимизации оттока клиентов, определение профиля проблемных клиентов, а также выявлять новые взаимосвязи между собираемыми данным» [33].

«Одним из перспективных методов является интеграция новейших информационных технологий, связанных со сбором и анализом больших

неструктурированных данных, работой с данными с помощью интеллектуальной обработки данный (анализ временных рядов), а также внедрение отдельных прогностических моделей в обработку данных в работающую CRM-систему организации» [34].

Объект исследования – статистические данные, получаемые из различных источников.

Предмет исследования – интеллектуальный анализ статистических данных, применяемых для моделирования трендов и прогнозирования.

Цель исследования – моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных, для повышения точности и надежности прогнозов.

Задачи исследования

- Проанализировать современное состояние и тенденции развития моделирования трендов и прогнозов.
- Определить понятия интеллектуализации данных для выбранной предметной области.
- Разработать алгоритмы для моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных.
- Экспериментально подтвердить эффективность предлагаемых решений.
- Разработать практические рекомендации по внедрению результатов исследования.

Значительный вклад в развитие моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных внесли работы следующий российских и зарубежных ученых Arunalatha, G. A [50], Kaplan R.S. [21], Norton D.P, [22] Remya, S. [32], Астраханцева, И. А. [1], Богатырев, С. Ю.[2], Богданов А. Б.[3], Борисова И. А. [3], Бром, А. Е.[4], Витязев, В.В.[5], Власов, М. П.[6], Гончаренко, А. Н.[7], Дайзенрот Марк Питер [8], Чен Сунь Он [8], Альдо Фейзал А. [8], Дорофеюк Ю. А. [10], Драганчук, Л. С.О [11], Жердев, А. А. [14], Дубров А.М. [12], Мхитрян В.С. [12], Трошин Л.И. [12]

В процессе исследования использованы следующие подходы и методы:

- «методы математического и статистического анализ данных, а также методы интеллектуализации и поиска зависимости данных» [34];
- «определены ключевые факторы, влияющие интеграция новейших информационных технологий, связанных со сбором и анализом больших неструктурированных данных, работой с данными с помощью интеллектуальной обработки данный (анализ временных рядов), а также внедрение отдельных прогностических моделей в обработку данных в работающую CRM-систему организации» [34].

Основная гипотеза исследования заключается в том, интеграция современных методов машинного обучения и анализа временных рядов с традиционными подходами к обработке статистических данных позволит значительно повысить точность и надежность прогнозирования трендов в различных областях.

В работе предлагается новый подход к моделированию тренда и построению прогноза, основанный на интеллектуализации обработки статистических данных, который включает в себя интеграцию адаптивных алгоритмов и автоматизированного выявления скрытых закономерностей. В отличие от традиционных подходов, разработанная модель обладает устойчивостью к шумам и выбросам, а также адаптацией к изменяющейся структуре данных. Это обеспечивает повышение точности и надежности прогнозов в условиях неопределенности и многомерности информации.

Практическая значимость исследования заключается в возможности практического применения реализованной программы для работы с клиентами финансовой организации.

«Теоретической основой диссертационного исследования служат научные труды российских и зарубежных исследователей» [3]. В ходе анализа учитываются современные подходы к проектированию, внедрению и оптимизации таких систем, а также их влияние на повышение эффективности управления финансовыми потоками, автоматизацию бизнес-процессов и

обеспечение информационной безопасности. Кроме того, исследование опирается на междисциплинарные концепции, включающие теорию систем, методы математического и имитационного моделирования, а также новейшие достижения в области финансовых технологий и цифровой трансформации.

На защиту выносятся:

- модели и методы для моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных;
- результаты апробации и оценки эффективности предлагаемых проектных решений.

Диссертация состоит из введения, трех глав, заключения, списка используемой литературы и используемых источников.

Во введении обоснована актуальность темы исследования, представлены объект, предмет, цели, задачи и положения, выносимые на защиту диссертации.

В первой главе дан анализ состояния исследований в области моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных. Во второй главе дан анализ методов и моделей для моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных. Третья глава посвящена разработке приложения с использованием предложенных методов и моделей для моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных, выполнена апробация предлагаемых проектных решений и оценка их эффективности.

В заключении приводятся результаты исследования.

Работа изложена на 71 странице и включает 53 рисунка, 5 таблиц и 50 источников.

1 Обзор исследований по теме моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных

1.1 Моделирования тренда на основе интеллектуализации обработки статистических данных

Основным видом деятельности организации, на базе которой выполнялась выпускная квалификационная работа, является 62.02 - Деятельность консультативная и работы в области компьютерных технологий.

Дополнительные виды деятельности по ОКВЭД:

- 62.0 разработка компьютерного программного обеспечения, консультационные услуги в данной области и другие сопутствующие услуги,
- 62.09 деятельность, связанная с использованием вычислительной техники и информационных технологий, прочая,
- 63.1 деятельность по обработке данных, предоставление услуг по размещению информации, деятельность порталов в информационно-коммуникационной сети интернет,
- 63.11.1 деятельность по созданию и использованию баз данных и информационных ресурсов,
- 95.1 ремонт компьютеров и коммуникационного оборудования.

«Структура организации представляет собой целостную, упорядоченную систему взаимосвязанных элементов» [8], определяющую порядок их взаимодействия, распределение полномочий, ответственности и ресурсов.

Линейно-функциональная модель является одной из наиболее распространенных в корпоративной среде, так как она позволяет достигать высокой степени управляемости, оперативного принятия решений и эффективного использования ресурсов. (рисунок 1).

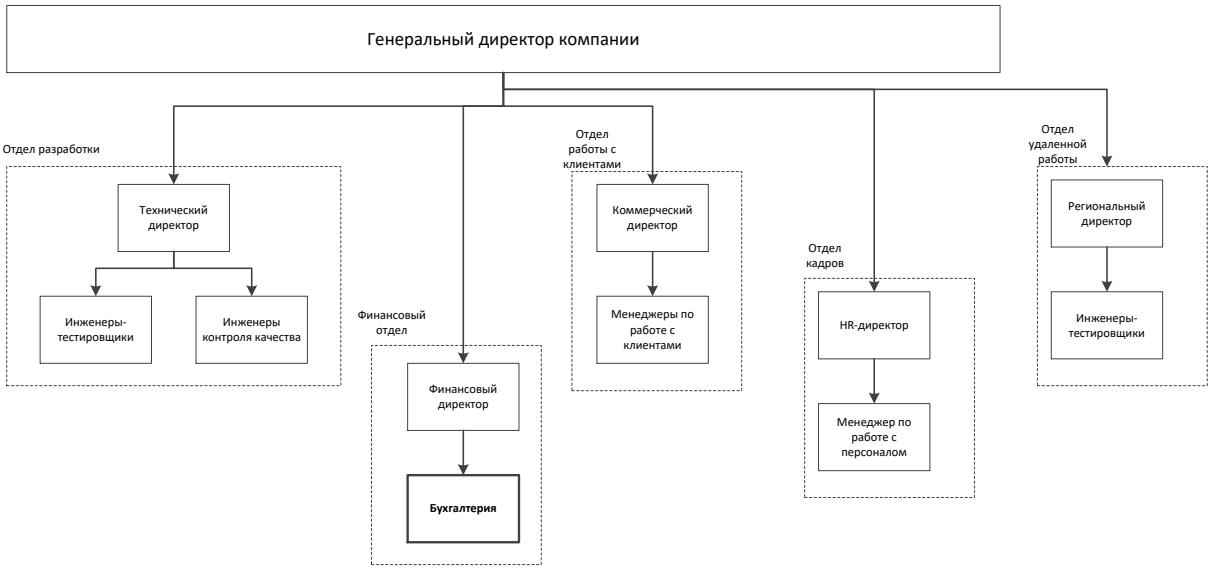

Рисунок 1 - Обобщенная организационная структура компании

«Администрация компании напрямую работает с разработчиками программного обеспечения, вендорами, производителями аппаратных средств защиты и устройств и т. д., далее поставляя данный товар (программное обеспечение и аппаратные средства) по подразделениям и заказчикам, а также осуществляя его тестирование» [21].

«Финансовый отдел включает в себя финансового директора и бухгалтерию. Бухгалтерия отвечает за ведение бухгалтерского, финансового, налогового и управленческого учета» [4].

«Отдел кадров состоит из HR-директора и менеджеров по работе с персоналом.

Отдел продаж состоит из коммерческого директора и менеджеров по продажам. Отдел продаж занимается поиском клиентов, обзвоном клиентов, ведет с ними переписку и т. д.» [5].

«Отдел функционального и нагрузочного тестирования - туда входят технический директор и команда тестировщиков и инженеров по качеству» [6].

В научной литературе очень хорошо освещены вопросы математической обработки статистических данных. В научных работах Arunalatha, G. A [50],

Kaplan R.S. [42], Norton D.P, [41] Remya, S. [22], Астраханцева, И. А. [1], Богатырев, С. Ю.[2], Богданов А. Б.[3], Борисова И. А. [23], Бром, А. Е.[4] рассматриваются использование метода главных компонент для построения прогнозов и трендов. Рассмотрим краткое содержание данных работ.

Метод главных компонент (principal component analysis, PCA) – это линейный метод по сокращению количества признаков в больших наборах данных. Это достигается путем преобразования большого набора признаков в более маленький, при этом сохраняя большую часть информации [50].

Конечно, часть информации при выполнении данного действия теряется, однако особенность данного алгоритма – потерять как можно меньше ценных данных при этом максимально сократив и упростив датасет.

Метод главных компонент – лишь один из алгоритмов, решающих задачу понижения размерности данных. Существуют и другие, как линейные, так и нелинейные алгоритмы, в частности, SVD, PCA с ядрами, t–SNE и другие.

С геометрической точки зрения, метод главных компонент заключается в поэтапном построении ортогональных по отношению друг к другу векторов (рисунок 2) [49]:

- Выбор центра всей выборки.
- Поиск «лучшей» прямой, которая проходит через центр: прямая, у которой сумма расстояний от точек выборки до прямой минимальна, либо, что эквивалентно, у которой сумма расстояний от центра до проекций точек выборки максимальна.
- Построение следующей «лучшей» прямой, которая была бы ортогональна к предыдущей.
- Повторяем процесс до тех пор, пока не закончатся измерения (признаки).

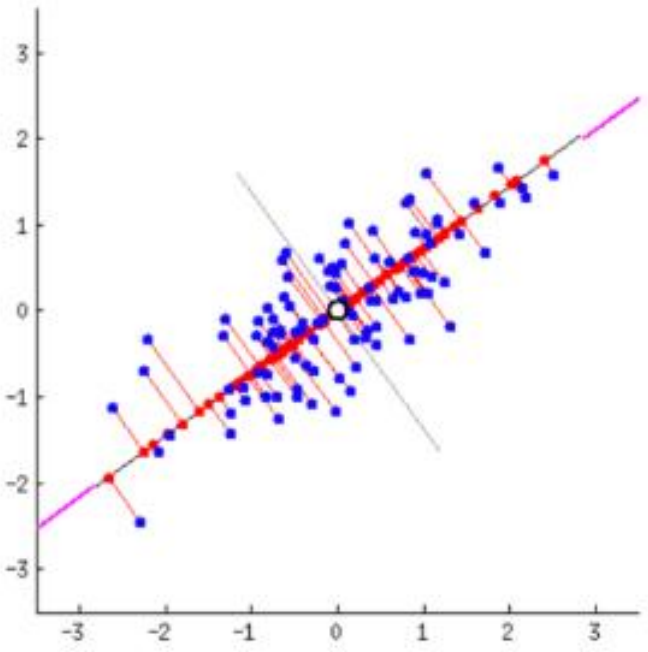

Рисунок 2 – Поиск первой главной компоненты

С алгебраической точки зрения, главные компоненты определяются с помощью собственных значений и собственных векторов матрицы ковариации наших признаков. Так, собственные векторы задают направление прямой таким образом, чтобы захватить максимум информации (дисперсии), а собственные значения – это коэффициенты у этих векторов, показывающие объем информации, который несет в себе каждый компонент [29].

Таким образом, если мы упорядочим главные компоненты по их собственным значениям мы получим список главных компонент в порядке убывания информативности [22].

Рассмотрим ключевые этапы применения PCA [48]:

– Стандартизация и нормализация данных - на данном шаге выполняется подготовка датасета к применению основного алгоритма. Так, мы стремимся привести все признаки к одной шкале (например, чтобы признаки, измеряющиеся тысячами единиц, не оказывали большего влияния, чем признаки с маленькими значениями) [24]. Помимо этого, мы стремимся привести данные к

нормальному распределению.

- Вычисление матрицы ковариаций или корреляций - на этом этапе наша цель выявить зависимости между признаками и отбросить явно линейно зависимые признаки. В реализации мы будем использовать корреляционную матрицу, предоставляемую библиотекой pandas и тепловую карту для удобной визуалиации.
- Применение PCA - в реализации мы будем использовать метод главных компонент из библиотеки scikit–learn. На этом этапе и происходит преобразование изначального датасета с данными с помощью метода главных компонент [27].
- Визуализация и анализ результатов.

В работе Дайзенрот Марк Питер [8], Чен Сунь Он [8], Альдо Фейзал А. [8], Дорофеюк Ю. А. [10], Драганчук, Л. С.О [11], «рассмотрена задача поиска корректных эмпирических аналогий в выборочных данных, имеющих вид булевых функций. Указанные в» [34], [2] «три направления использования эмпирических аналогий: для синтеза продукций как элементов наполнения баз знаний, для восстановления пропусков информации и построения алгоритмов автоматической классификации применим и для задач прогноза» [10].

«Использование аналогии (рисунок 3) предполагает наличие эталонного процесса (например, представленного временным рядом) для которого синтезируются модели прогноза с необходимым горизонтом и качеством прогнозирования. Для изучаемого процесса существует или установлена информация о подобии эталонному процессу. Это позволяет получить модели прогноза даже при неполноте данных изучаемых процессов [28].

Важным этапом в задачах синтеза знание ориентированных моделей является задача сжатого описания исходных данных (индексы, интегральные показатели, рейтинги и т. п.) [47].

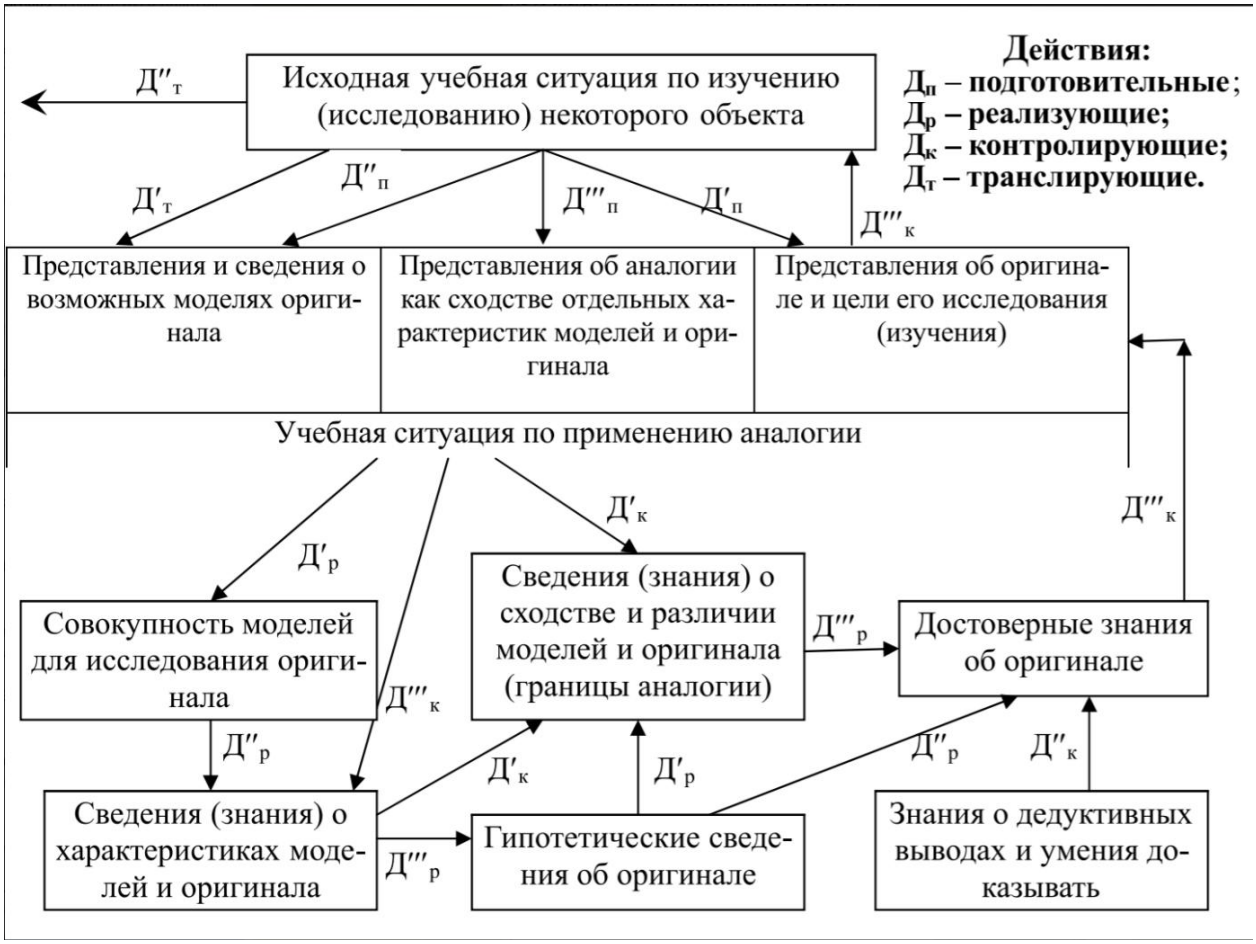

Рисунок 3 – Использование аналогии

В работе в основном используются метод главных компонент, корреляционно–регрессионный анализ, но с успехом может использоваться подход, применяемый в работе [3], в которой структуризация параметров проводится методами экстремальной группировки: выбирается число групп и тип групп. Результатом экстремальной группировки являются группы параметров (факторов) – интегральные показатели групп (центры), т. е. линейные комбинации показателей соответствующей группы на некотором уровне иерархии изучаемой системы. На базе группировки выбираются информативные показатели (легко интерпретируемые интегральные показатели, либо исходные показатели, ближайшие к этим факторам).

Использование структурного прогнозирования развития сложных систем позволяет прогнозировать не точные значения параметров, описывающих состояние объекта, а только класса объектов в рамках структуры объектов изучаемой системы» [8].

1.2 Моделирования прогноза

Вопросами интеллектуального анализа данных занимаются следующие авторы: Альдо Фейзал А. [8], Дорофеюк Ю. А. [10], Драганчук, Л. С.О [11], Жердев, А. А. [14], Дубров А.М. [12], Мхитрян В.С. [12], Трошин Л.И. [12].

«Интеллектуальный анализ данных связан с поиском шаблонов, трендов, ассоциаций, аномалий, важных атрибутов, структур и зависимостей в данных. Он является мультидисциплинарной областью, которая использует и усовершенствует идеи из различных областей знаний, таких как обработка сигналов и изображений, машинное обучение, распознавание образов, экспертные системы, оптимизация, статистика и другие. ИАД анализирует большие объемы сложных необработанных данных, помогая принимать на их основе решения» [7].

«Интеллектуальный анализ данных (ИАД) является итеративным процессом, который можно разбить на следующие этапы» [8][9]:

- «хранение и управление данными» [30],
- «предварительная обработка и подготовка данных к анализу,
- процессы непосредственного анализа данных и извлечения знаний,
- оценка и интерпретация полученных результатов.

Последний из перечисленных этапов в основном касается нетехнической работы, такой как документация и оценка результатов аналитиком» [31].

«Одним из направлений ИАД, перспективных с точки зрения исследований, является анализ временных рядов, Временной ряд – это последовательность собранных в разные моменты времени значений каких–либо параметров исследуемого процесса или объекта» [13].

«Методы интеллектуального анализа призваны исследовать временные ряды с целью нахождения в них скрытых шаблонов. Несмотря на то, что все эти методы используются различные математические подходы, все они имеют одну общую особенность: для их применения скорее необходимо некоторое высокоуровневое представление данных, чем исходные сырые данные» [12].

«Такое высокоуровневое представление необходимо как для выделения характеристик временного ряда, так и для его эффективного хранения, передачи и обработки.

Вейвлет–преобразование по своей природе призвано предоставить такое высокоуровневое представление временных рядов (рисунок 4).

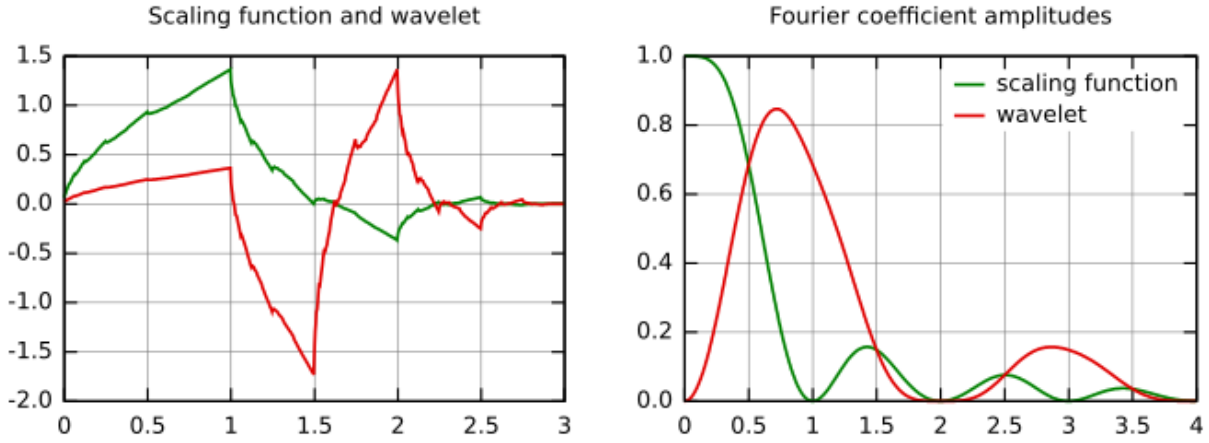

Рисунок 4 – Вейвлет–преобразование

Вейвлет–преобразование может быть эффективно применено в следующих направлениях» [32].

– «Поиск подобия во временных рядах предполагает отыскание полного или частичного совпадения с заданным рядом (шаблоном). При этом вводится параметр допустимого расстояния между рядами.

Решение данной задачи состоит из двух этапов: индексирования рядов и выполнения запросов. Индексирование» [46] – «это процесс создания указателей для ускорения доступа к данным, предполагающий выделение признаков

временного ряда и его сжатие. ДВП в данной задаче может быть применено для выполнения индексирования данных и создания метрики расстояния между рядами» [35].

– «Классификация временных рядов (рисунок 5) состоит в назначении ряду одной из заранее известных меток класса. ДВП может быть интегрировано в классификацию временных рядов двумя путями: применение методов классификации к результатам вейвлет–преобразования и применение многомасштабного представления данных» [36].

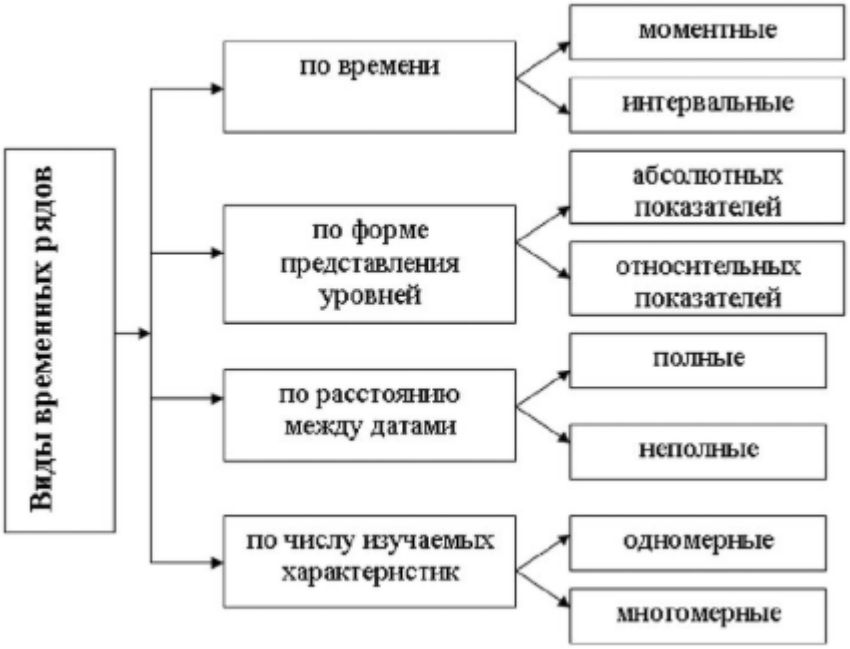

Рисунок 5 – Классификация временных рядов

– «Кластеризация временных рядов состоит в их разбиение на группы по подобию. Кластеризация позволяет идентифицировать шаблоны и тренды для каждой из групп.
– Выявление аномалий в поведении временных рядов (рисунок 6) включает: определение внезапных изменений в рядах, выявление аномалий путем сравнения расстояний между несколькими рядами и

выявление нерегулярных шаблонов» [38].

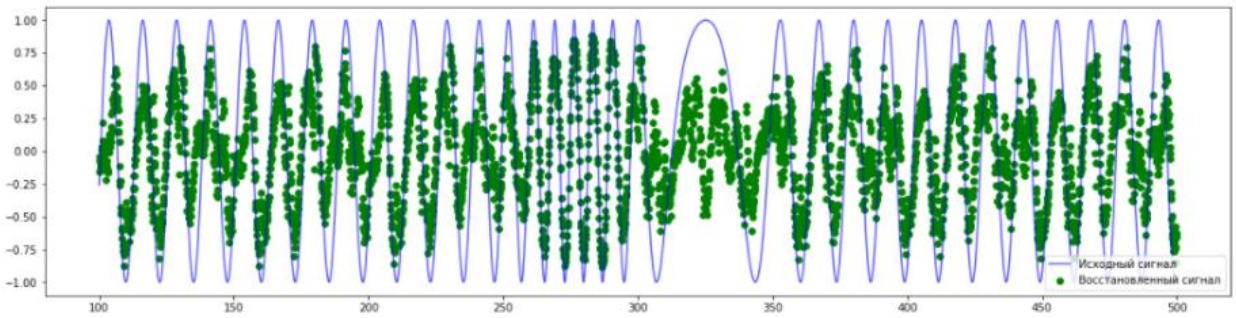

Рисунок 6 – Выявление аномалий в поведении временных рядов

– «Прогнозирование значений временного ряда в будущем на основании исторических данных.

Многие теоретические вопросы геометрической кластеризации и классификации слабо–структурированной информации остаются нерешенными до настоящего времени» [10]. «Слабоструктурированные данные (semi–structured data) – данные, для которых точная структура заранее неизвестна и может меняться в потоке. К подобным данным обычно относят тексты и снимки, которые могут как порознь, так и одновременно присутствовать в документах, представленных в различных форматах.

Для анализа таких данных необходимы математические модели представления информации и алгоритмы преобразования, выделения признаков и обработки. Возникающие при этом трудности преодолеваются наборами различных методов» [37].

«Можно выделить, на наш взгляд, четыре наиболее значимые системы представления исходных данных:

– «табличный способ, который широко применяется при решении задач классификации алгебраическим методом» [18], нейронными сетями, деревьями решений и при построении опорного множества кластеров;

– «способ представления данных мультимножествами» [12], эффективность которого показана на экономических задачах и исследуется применительно к задачам распознавания графических образов, например, печатных букв;
– «метод фазовых траекторий (рисунок 7), предложенный для обработки мультимодальной информации» [44], «и пока недостаточно изученный на практике. Задача формирования траекторий в фазовом пространстве и интеллектуального управления ими на основе правил рассмотрена в работе» [15];

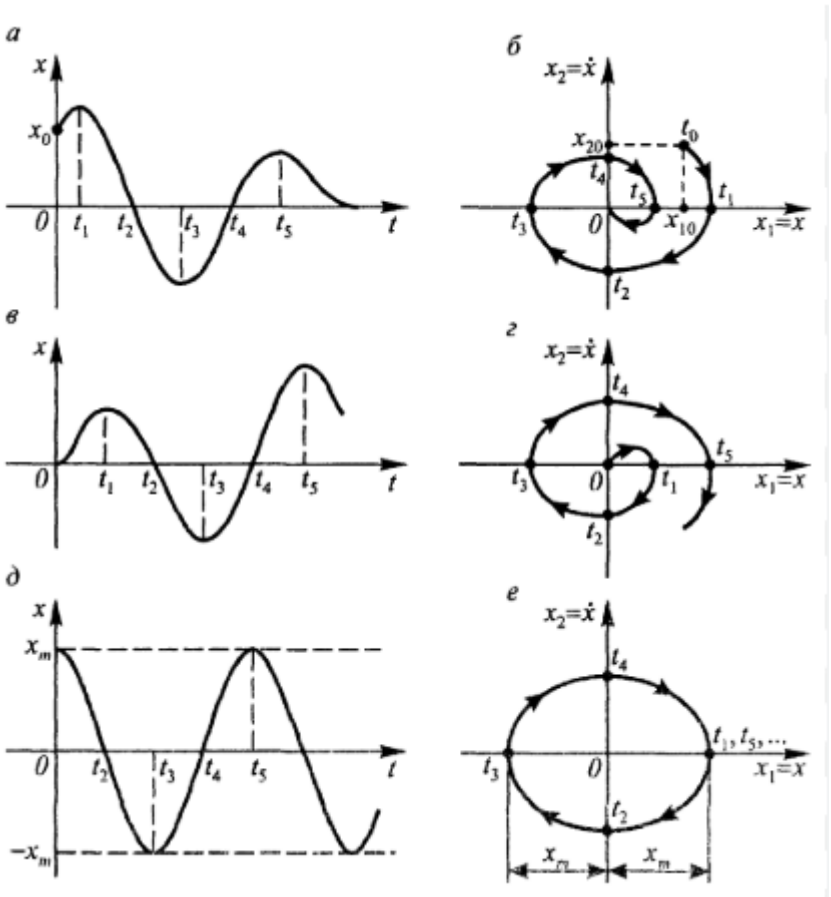

Рисунок 7 – Метод фазовых траекторий

– «процедурное представление, когда знания о конкретной проблемной области задаются в виде набора правил, и декларативное

представление знаний, когда информация хранится в виде базы данных и/или базы знаний» [16].

«Некоторые модели представления данных, которые могут быть применены в задачах кластерного анализа» [17].

«Один из способов снижения размерности данных заключается в применении анализа главных компонент (PCA). Популярность PCA (рисунок 8) определяется рядом свойств, важнейшими из которых является его оптимальность при сжатии множества векторов высокой размерности в множество векторов более низкой размерности, а затем их восстановления» [11].

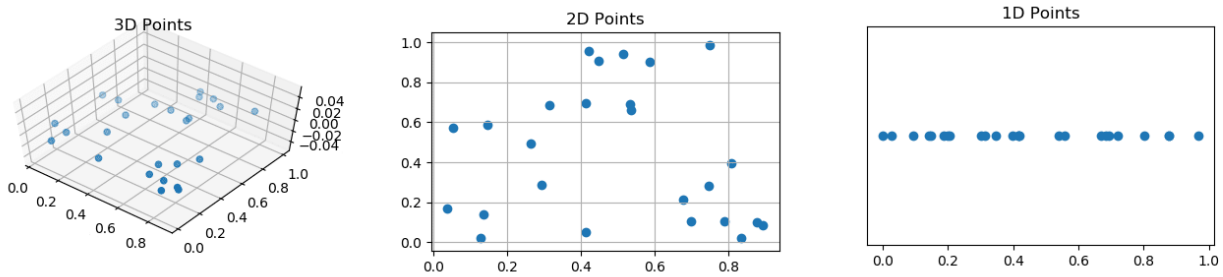

Рисунок 8 – Метод главных компонент (PCA)

«Использовать PCA в задаче обучаемой классификации данных можно двояко. Во–первых, безотносительно к сложной структуре результатов наблюдений, подразумевающей наличие классов данных. В этом случае данные без уточнения их статистической модели сжимаются, а затем подвергаются анализу» [13].

В работах Кривенко М. П. приведено описаний методики формирования интегральных показателей оценки тренда и прогноза по статистическим данным, рассмотрим подробно ее работу.

«Проблемным является этап формирования целей, их показателей достижимости и целевых значений. Многокритериальность существенно усложняет задачу. Целевые значения показателей не равнозначны, но их параметры (минимальные и максимальные) должны обеспечивать стабильность

исследуемой системы, а некоторый набор показателей должен отражать приоритетное развитие. Соответствующие организационные мероприятия должны обеспечить рациональность и эффективность всего процесса развития системы» [39].

«Древовидная иерархическая структура формирования интегральных показателей образует нейронную сеть с точно известными процедурами логического вывода (а не «черный ящик»). Задаваясь набором целевых интегральных показателей, можно находить весовые коэффициенты в свертках (например, методом обратного распространения ошибки) и тем самым выявлять куда необходимо вкладывать инвестиции (усилия, ресурсы и т. п.) для повышения значения показателя, которому соответствует больший вес. Задачи такого вида находятся в стадии исследования.

Веса сворачиваемых показателей при настройке также выбираются экспертно, а затем адаптируются в зависимости от поведения интегральных показателей и выбранных ограничений, что соответствует принципу адаптивной оценки» [40].

«Так как интегральные показатели (индексы) должны отражать эффективность мероприятий по всем видам деятельности, то это требование накладывает условие на выбор исходных показателей. Представляет интерес показатели, наиболее точно отражающие динамику процесса (рисунок 9). Показатели, которые не меняются на протяжении длительного периода, но могут реагировать на резкие изменения (событийный характер воздействия), формируют отдельный блок (блок «безопасности»), предупреждающий об изменяющихся условиях событий (особенно актуально в период пандемии)» [41] [42].

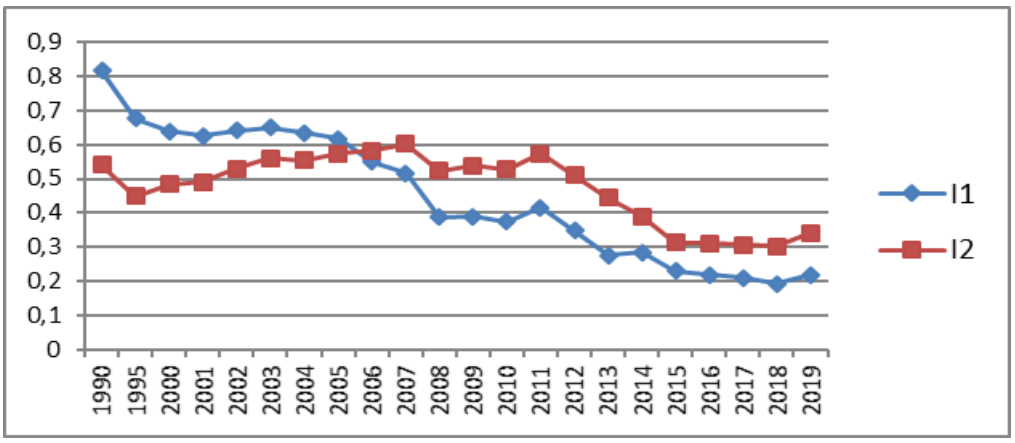

Рисунок 9 –Интегральные показатели по блокам 1,2

«В качестве примера для расчета интегральных показателей можно взять статистические данные по региону, которые разделены на блоки (рисунок 10)» [43].

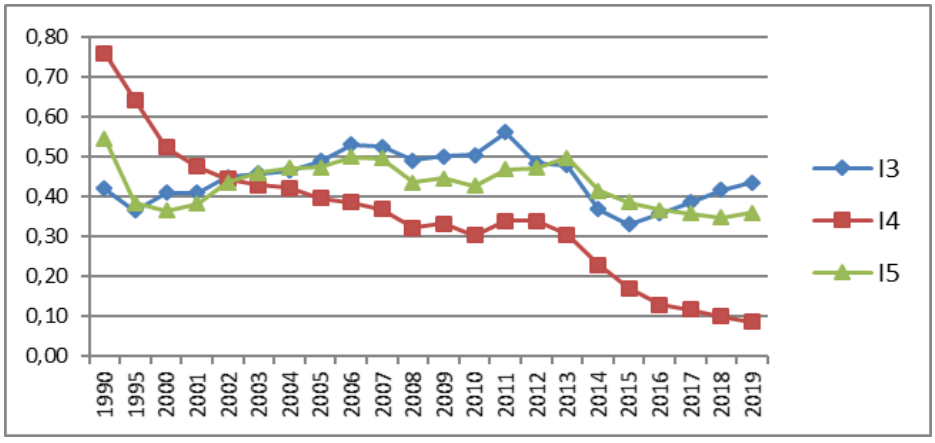

Рисунок 10 – График трех интегральных показателей I^3, I^4, I^5

«Аналогично строятся интегральные показатели по всем блокам данных на различных уровнях иерархии с различной степенью детализации» [44].

«Сильное запаздывание или незначительный рост может свидетельствовать о нерациональном управлении процессом работы системы. Построенные интегральные показатели позволяют прогнозировать дальнейшие тенденции. Заметим, что весовые коэффициенты, построенные по методу

главных компонент, хорошо отражают показатели с большой дисперсией, соответствуют коэффициентам информативности, что позволяет существенно уменьшать размерность данных (использовать только значимые компоненты)» [45]. К таким же результатам приводит и экспертный выбор коэффициентов (случай профессионального подхода и четким целями на плановые стратегические показатели).

Можно выделить недостатки существующих решений. К ним относится недостаточная адаптивность к изменяющимся данным - многие традиционные и даже интеллектуальные модели плохо справляются с изменением структуры данных (смена тренда, сезонности, внезапные аномалии). Ограниченное выявление скрытых закономерностей - классические статистические методы (например, линейная регрессия, ARIMA) не всегда способны уловить сложные нелинейные взаимосвязи в данных. Низкая устойчивость к шумам и выбросам - модели плохо работают с «грязными» или неполными данными, требуя предварительной ручной очистки. Слабая масштабируемость и универсальность решений - ограниченная гибкость при обработке больших объемов данных.

Вывод по первой главе

В первой главе магистерской работы рассмотрены основные публикации российских и зарубежных авторов.

В современных публикациях авторов был сделан упор на существующие математические и статистические методы анализ данных.

2 Анализ методов и моделей моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных

2.1 Метод моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных

«В рамках выполнения магистерской работы «схемы бизнес-процессов, используемые для моделирования трендов и прогнозов на основе интеллектуальной обработки статистических данных, будут описаны с использованием методологии структурного подхода» [34].

«Входной информацией для бизнес-процессов, связанных с моделированием трендов и прогнозов, служит набор первичных данных, собираемых организацией в ходе своей деятельности (рисунок 11).

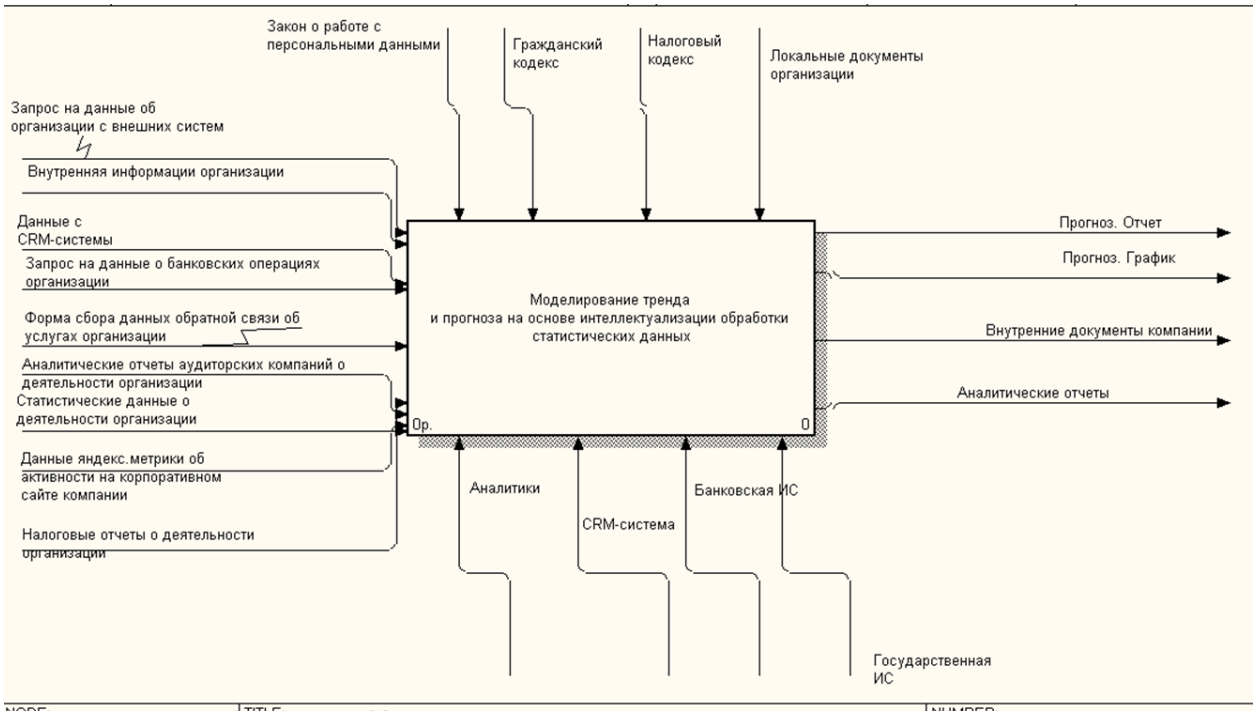

Рисунок 11 – Контекстная диаграмма «Моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных. Как есть»

К таким данным относятся» [34]:

– «государственные и банковские информационные системы;
– CRM-системы;
– внешние запросы на данные об организации;
– аналитика Яндекс.Метрики о посещаемости сайта;
– формы обратной связи;
– отчеты аудиторских компаний;
– налоговая и внутренняя отчетность;
– запросы на банковские операции;
– статистические данные о деятельности организации» [34].

Эти данные формируют основу для построения моделей, которые обеспечивают прогнозирование и анализ динамики процессов.

«Выходная информация бизнес-процесса моделирования трендов и прогнозов на основе интеллектуальной обработки статистических данных» [1] представлена различными типами отчетов, включая:

– «прогнозы в текстовой форме (документация, отчеты);
– прогнозы в виде графиков и трендов;
– аналитические отчеты;
– отчетные документы для регулирующих органов» [25].

Рассмотрим составные части процесс «Моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных» подробно (рисунок 12).

«Основными операциями верхнего уровня являются:

– сбор информации,
– подготовка данных для анализа,
– построение прогноза,
– моделирование тренда» [19].

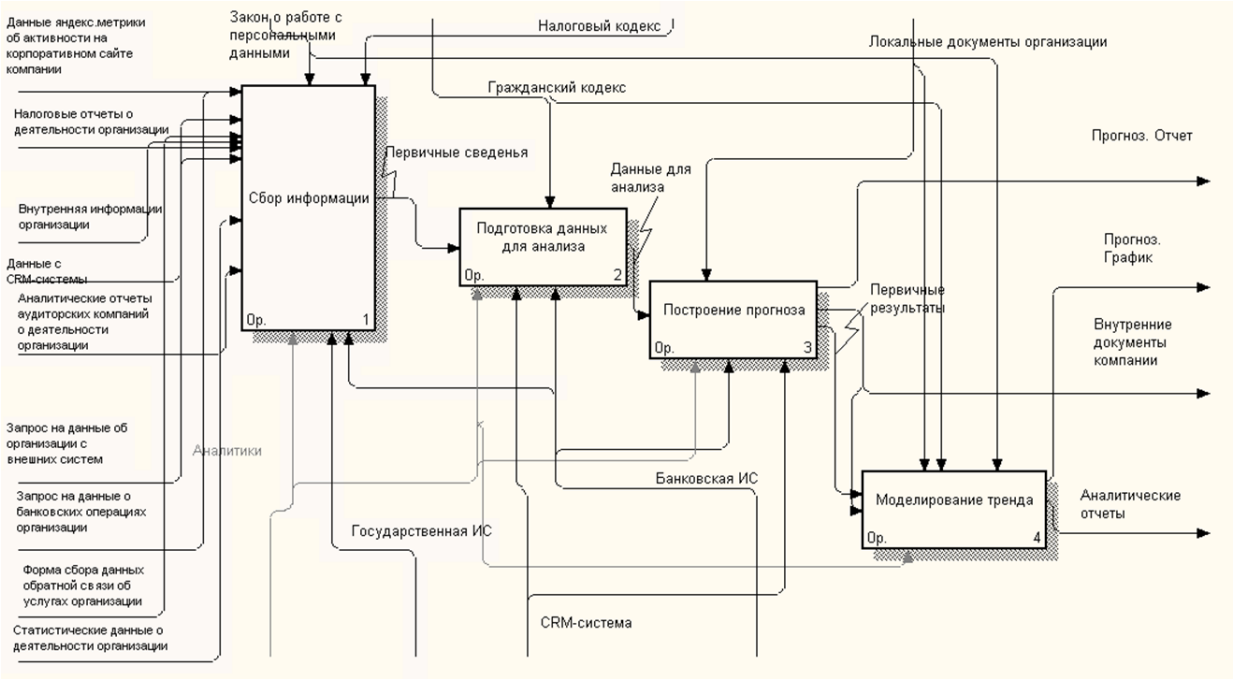

Рисунок 12 – Моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных. Как есть

Бизнес-процесс, описывающий построение прогноза и моделирования тренда, показан на рисунке 13.

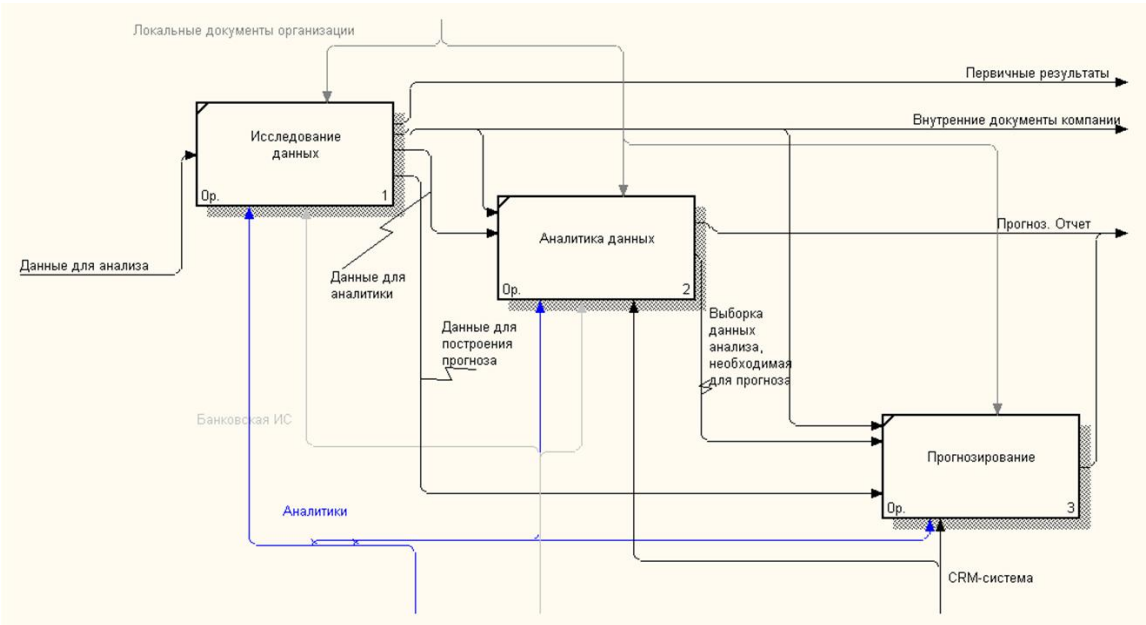

Рисунок 13 – Построение прогноза. Как есть

Описание процесса «Моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных представлено в таблице 1.

Таблица 1 – Процесс «Моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных. Как есть»

Имя блока	Управление блока	Механизм блока	Вход блока	Выход блока
Сбор информации	Закон защите персональных данных, Налоговый и гражданский кодекс	Аналитики, Государственная информационная система, Банковская информационная система	Государственные и банковские системы; CRM-системы; Внешние запросы на данные об организации; Аналитика Яндекс.Метрики о посещаемости сайта; Отчеты аудиторских компаний; Налоговая и внутренняя отчетность; Запросы на банковские операции; Статистические данные о деятельности организации.	Первичные сведенья
Подготовка данных для анализа	Налоговый и гражданский кодекс	CRM-система, Банковская ИС, Аналитики	Первичные сведенья	Данные для анализа
Построение прогноза	Внутренние регламенты компании	Банковская ИС, CRM-система, Аналитики	Данные для анализа	Отчет по среднесрочному прогнозированию Отчетные документы для регулирующих органов Первичные результаты

Продолжение таблицы 1

Имя блока	Управление блока	Механизм блока	Вход блока	Выход блока
Моделирование тренда	Внутренние регламенты компании Налоговый и гражданский кодекс Закон защите персональных данных	Аналитики,	Первичные результаты Отчетные документы для регулирующих органов	Аналитические отчеты Графическое представление прогнозных данных

«Для более детального рассмотрения процесса «Построение прогноза. Как есть» составлена таблица 2, в которой показаны входная и выходная информация, а также управление и механизмы каждого блока» [2].

Таблица 2 – Процесс «Построение прогноза. Как есть»

Имя блока	Управление блока	Механизм блока	Вход блока	Выход блока
Исследование данных	Внутренние регламенты компании	Аналитики Банковская ИС	Данные для анализа	Данные для аналитики Первичные результаты Данные для построения прогноза Отчетные документы для регулирующих органов
Аналитика данных	Внутренние регламенты компании	Аналитики Банковская ИС CRM-система	Данные для аналитики Отчетные документы для регулирующих органов	Выборка данных анализа, необходимая для прогноза Отчет по среднесрочному прогнозированию
Прогнозирование	Внутренние регламенты компании	Аналитики CRM-система	Выборка данных анализа для прогноза Данные для построения прогноза Отчетные документы для регулирующих органов	Отчет по среднесрочному прогнозированию

Современные компании, стремящиеся к повышению эффективности своей деятельности, активно используют передовые информационные технологии, позволяющие автоматизировать процессы анализа и прогнозирования. Одним из наиболее перспективных направлений является интеллектуальной обработки данных для работы с большими объемами информации, включая как структурированные, так и неструктурированные данные.

Одним из ключевых элементов аналитического процесса является анализ временных рядов, который позволяет выявлять закономерности и тренды на основе исторических данных. Этот метод широко используется в прогнозировании продаж, оценке сезонных колебаний спроса, анализе потребительского поведения и оптимизации цепочек поставок.

Для максимального эффекта прогнозные модели должны быть интегрированы в существующую CRM-систему компании. Такая интеграция обеспечивает:

- Автоматическое обновление прогнозов в режиме реального времени.
- Использование исторических данных о клиентах, сделках и заказах для повышения точности предсказаний.
- Создание персонализированных предложений на основе поведенческого анализа.
- Автоматизацию процессов принятия решений, что позволяет оперативно реагировать на изменения в рыночной среде.

Внедрение интеллектуального анализа данных позволяет компаниям не только прогнозировать будущие продажи, но и разрабатывать стратегию роста на основе объективных данных, минимизируя влияние субъективных факторов.

С учетом предложенных решений была разработана модифицированная модель прогнозирования, представленная на рисунке 14.

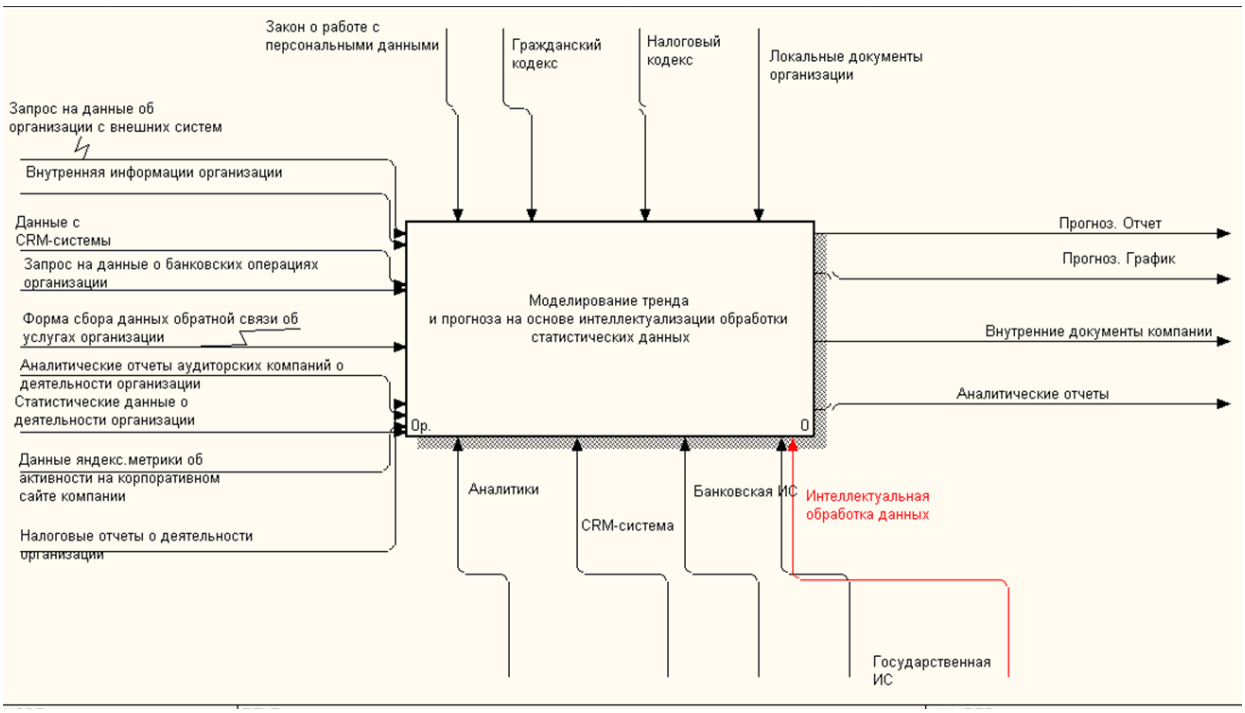

Рисунок 14 – Контекстная диаграмма «Моделирование тренда и прогноза на основе обработки статистических данных. Как должно быть»

Данная модель учитывает следующие нюансы:

- Расширенные источники данных (включая внешние факторы, такие как экономические индикаторы и конкурентная среда).
- Автоматизированную систему сбора, очистки и обработки информации.
- Улучшенные алгоритмы машинного обучения для более точного прогнозирования.
- Ключевые преимущества внедрения интеллектуальной аналитики
- Повышение точности прогнозов, что позволяет минимизировать риски и эффективно управлять ресурсами.
- Оптимизация маркетинговых стратегий, за счет глубокого анализа поведения клиентов и персонализированного подхода.
- Автоматизация принятия решений, что снижает влияние человеческого фактора и ускоряет реакцию на изменения рынка.

– Рост конкурентоспособности, благодаря оперативному анализу трендов и быстрому внедрению инновационных решений.

Далее более подробно рассмотрим на какие блоки повлияет внедрение данных механизмов. Моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных будет выглядеть следующим образом (рисунок 15).

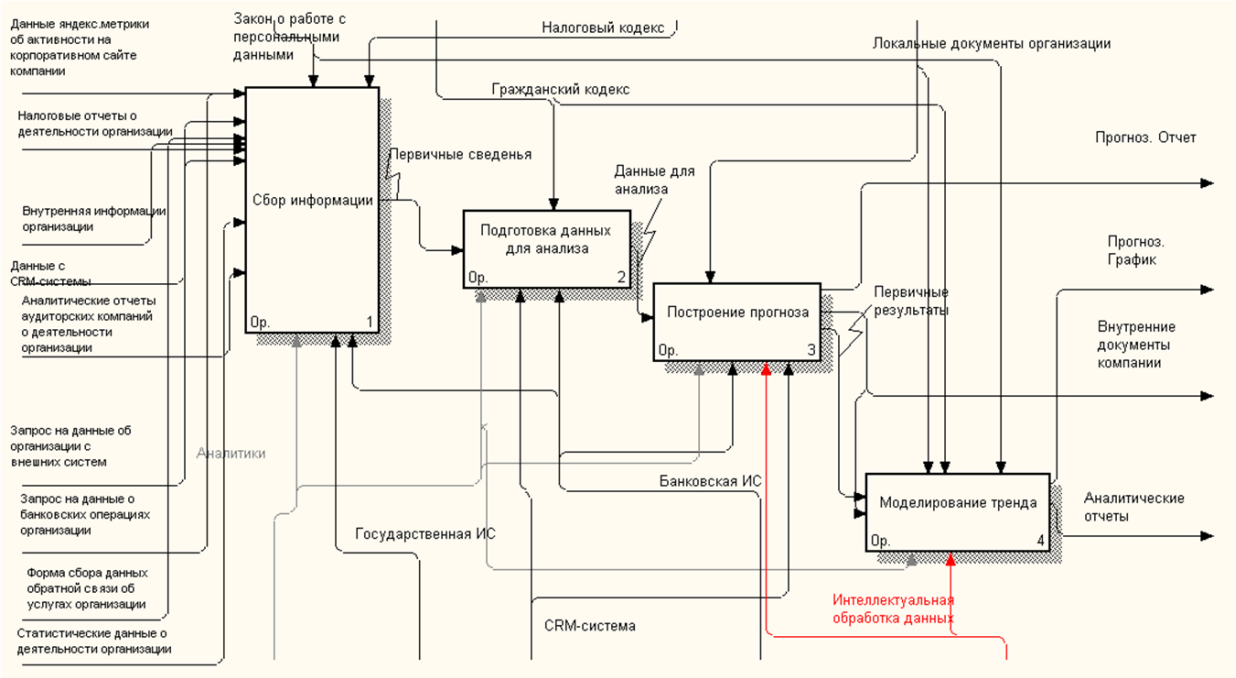

Рисунок 15 – Моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных. Как должно быть

Внедрение современных информационных технологий в процессы анализа и прогнозирования является стратегически важным шагом для компании, стремящейся к устойчивому развитию. Использование интеллектуальных инструментов обработки данных позволяет повысить эффективность аналитики, улучшить стратегическое планирование и сформировать конкурентные преимущества.

«Для более детального рассмотрения процесса «Моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных»

после внедрения предложенных механизмов составлена таблица 3, в которой показаны входная и выходная информация, а также управление и механизмы каждого блока» [20][23].

Таблица 3 – Процесс «Моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных. Как должно быть»

Имя блока	Управление блока	Механизм блока	Вход блока	Выход блока
Сбор информации	Закон о работе с персональными данными, Налоговый кодекс	Аналитики, Государственная ИС, Банковская ИС	Государственные и банковские системы; CRM-системы; Внешние запросы на данные об организации; Аналитика Яндекс.Метрики о посещаемости сайта; Отчеты аудиторских компаний; Налоговая и внутренняя отчетность; Запросы на банковские операции; Статистические данные о деятельности организации.	Первичные сведенья
Подготовка данных для анализа	Гражданский кодекс	CRM-система, Банковская ИС, Аналитики	Первичные сведенья	Данные для анализа
Построение прогноза	Внутренние регламенты компании	Банковская ИС, CRM-система, Аналитики Модели и алгоритмы обработки информации	Данные для анализа	Отчет по среднесрочному прогнозированию Отчетные документы для регулирующих органов Первичные результаты

Продолжение таблицы 3

Имя блока	Управление блока	Механизм блока	Вход блока	Выход блока
Моделирование тренда	Внутренние регламенты компании Гражданский кодекс Закон о работе с персональными данными	Аналитики, Модели и алгоритмы обработки информации	Первичные результаты Отчетные документы для регулирующих органов	Аналитические отчеты Графическое представление прогнозных данных

Новая модифицированная модель, основанная на применении машинного обучения и анализа больших данных, обеспечивает компании возможность гибкого реагирования на рыночные изменения, точного прогнозирования и эффективного управления ресурсами, что в итоге приводит к увеличению прибыли и укреплению позиций на рынке.

Бизнес-процесс «Построение прогноза. Как должно быть» представлен на рисунке 16.

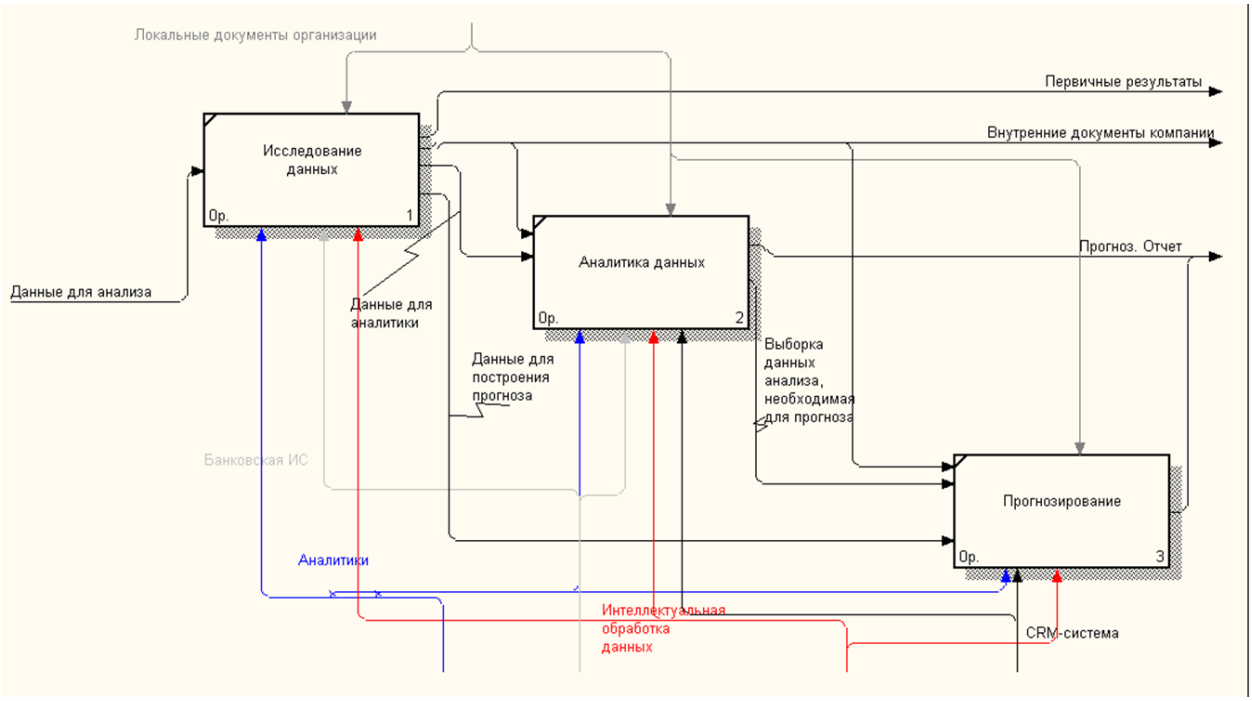

Рисунок 16 – Построение прогноза. Как должно быть

Для более детального рассмотрения изменённого процесса «Представление результатов. Как должно быть» составлена таблица 4.

Таблица 4 – Процесс «Представление результатов. Как должно быть»

Имя блока	Управление блока	Механизм блока	Вход блока	Выход блока
Исследование данных	Внутренние регламенты компании	Аналитики Банковская ИС Модели и алгоритмы обработки информации	Данные для анализа	Данные для аналитики Первичные результаты Данные для построения прогноза Отчетные документы для регулирующих органов
Аналитика данных	Внутренние регламенты компании	Аналитики Банковская ИС CRM-система Модели и алгоритмы обработки информации	Данные для аналитики Отчетные документы для регулирующих органов	Выборка данных анализа, необходимая для прогноза Отчет по среднесрочному прогнозированию
Прогнозирование	Внутренние регламенты компании	Аналитики CRM-систем Модели и алгоритмы обработки информации	Выборка данных анализа, необходимая для прогноза Данные для построения прогноза Отчетные документы для регулирующих органов	Отчет по среднесрочному прогнозированию

«Результатом является проектирования бизнес-процессов, которые отражают текущее состояние и состояние после внедрения новых методов моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных. Внедрение новых методов моделирования и

прогнозирования данных позволит взаимодействовать с различными источниками входных данных, а также позволит получать различные прогнозы (долгосрочные, среднесрочные и краткосрочные)»[21][26].

2.2. Модель моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных

«Основные этапы процесса моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных, построенные с помощью методологии BPMN, представлены на рисунке 17» [33].

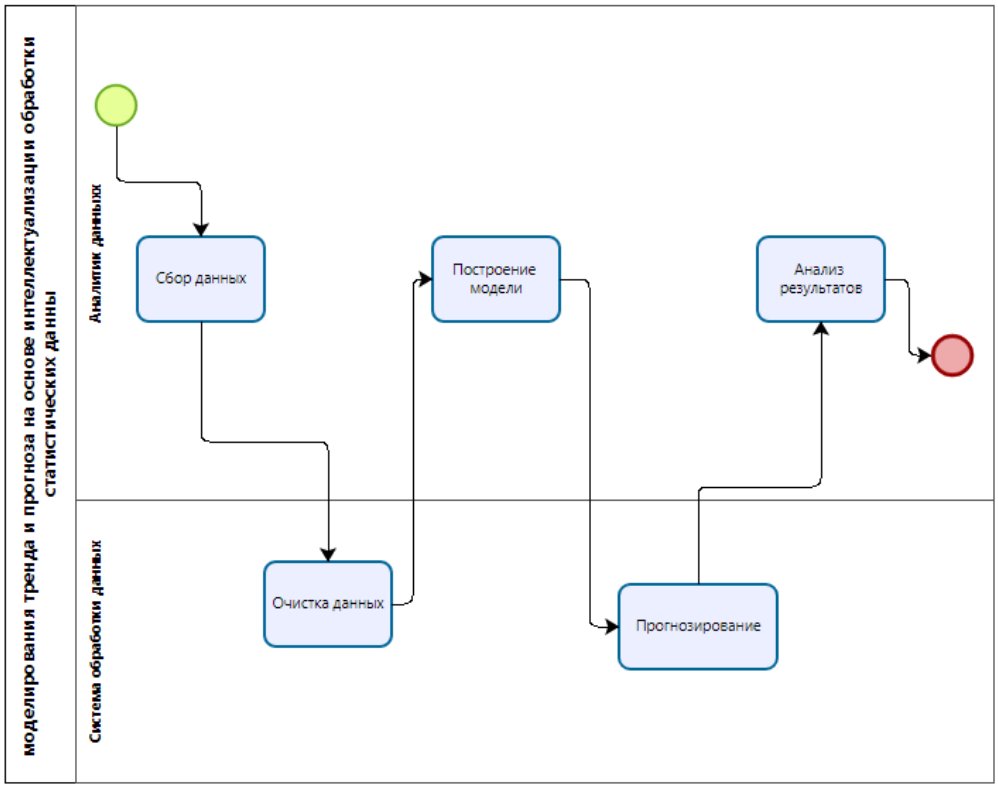

Рисунок 17 – Процесс моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных. Как есть

– «Инициация процесса – на этом этапе необходимо определить цели и задачи анализа. Далее согласовать входные данных и параметры

прогноза» [34].

- «Сбор и предобработка данных – на этом этапе получаем исходные статистические данных» [34] и проводим очистку данных (удаление пропусков, аномалий, нормализация). Также если необходимо проводим агрегацию данных (группировка по временным или другим параметрам).
- Анализ данных - на этом этапе происходит исследовательский анализ данных (EDA): визуализация, расчет статистических характеристик, также планируется выявление трендов и закономерностей в данных.
- Модели и алгоритмы обработки информации - на этом этапе происходит построение модели тренда с использованием алгоритмов машинного обучения (например, линейная регрессия, ARIMA, нейронные сети) и происходит оценка качества модели (метрики, кросс-валидация).
- Прогнозирование – на этом этапе для построения прогноза необходимо применить модели для предсказания будущих значений. Провести визуализацию результатов прогноза (графики, таблицы).
- Контроль и корректировка – на этом этапе необходимо провести анализ качества прогноза. И провести корректировку модели при необходимости (оптимизация параметров, пересмотр гипотез).
- Документирование и принятие решений – «это завершающий этапе процесса моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных на котором происходит генерация отчетов по результатам анализа, а также формируются рекомендации для принятия решений» [33].

«Процесс прогнозирования включает несколько этапов, которые объединяют подготовку данных, построение моделей и интерпретацию результатов. Рассмотрим подробное описание каждого шага, процесса прогнозирования, который представлен на рисунке 18» [34].

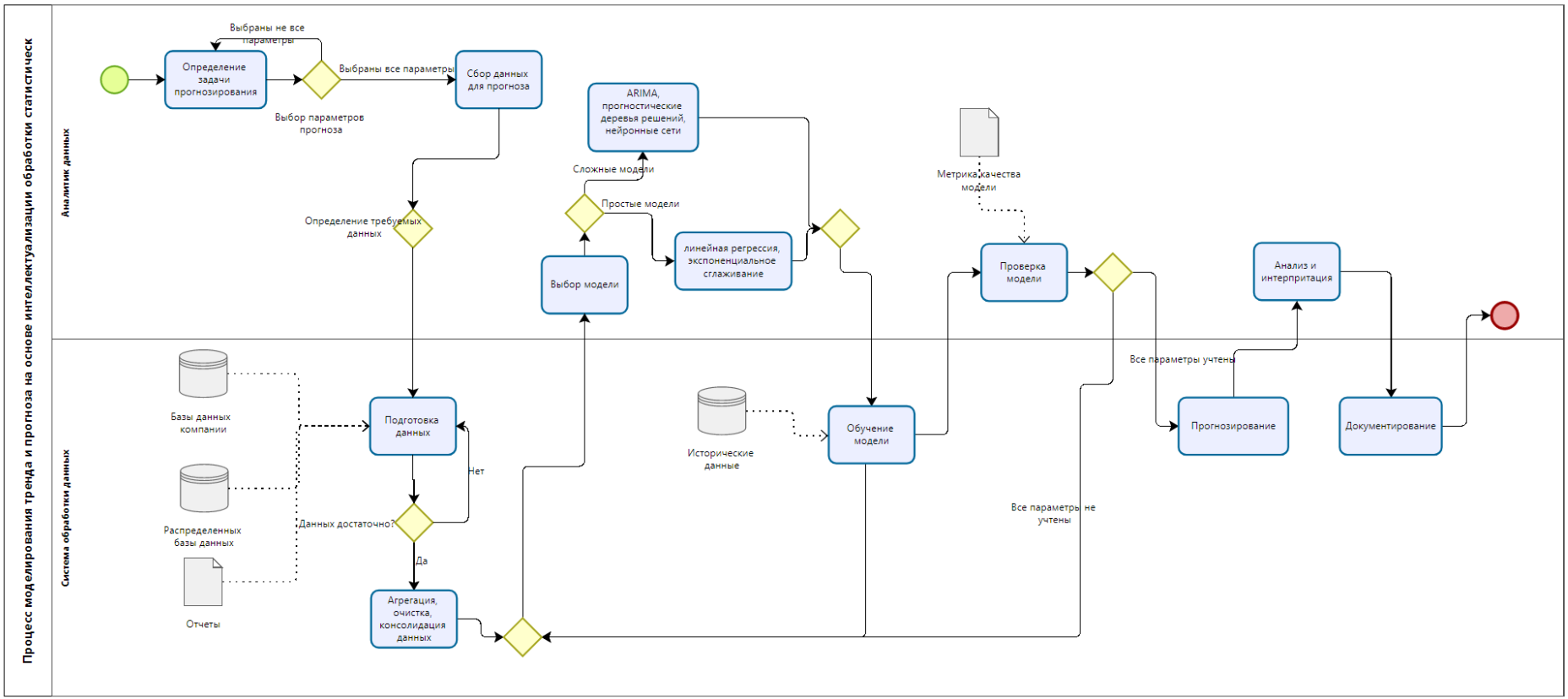

Рисунок 18 – Процесс моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных. Как должно быть

«Определение задачи прогнозирования - цель данного этапа: четкое понимание того, что нужно прогнозировать (например, продажи, спрос, уровень доверия к клиенту).

Тип данных, необходимый на этом этапе: временные ряды, кросс-секционные данные или панельные данные.

Горизонт прогнозирования: как далеко в будущее вы хотите сделать прогноз (день, неделя, месяц, год)» [34].

«Сбор данных - источники данных: системы учета, базы данных, API, отчеты. Уточнение доступных данных (включая временные метки, категориальные переменные, числовые значения и др.). Важность достоверности данных, чтобы прогноз базировался на надежных показателях» [34].

«Подготовка данных - на этом этапе необходимо провести очистку данных и агрегацию, а также трансформацию и нормализацию и стандартизацию. Этап очистки, проводился и до прогнозирования, но на этом этапе необходимо вернутся и провести удаление или обработка пропущенных значений. Исправление ошибок в данных (дубликаты, неверные форматы). Этап агрегации: группировка данных по времени (дни, недели, месяцы) или категориям» [34].

«Этап трансформации данных включает нормализацию или стандартизацию, преобразование временных рядов (например, логарифмирование для стабилизации дисперсии). Этап разделения данных включает разделение на обучающую и тестовую выборки (например, 80% для обучения, 20% для тестирования)» [34].

«Выбор и построение модели - на этом этапе предлагается использовать простые модели, такие как линейная регрессия, экспоненциальное сглаживание. А также при необходимости использовать сложные модели, такие как ARIMA (Авто-регрессионная интегрированная скользящая средняя), прогностические деревья решений (например, Gradient Boosting, Random Forest) и нейронные сети (LSTM, GRU для временных рядов)» [34].

«Выбор модели, используемой в прогнозе, зависит от:

– Объема данных.
– Сложности трендов и сезонности.
– Требуемого уровня точности» [34] (рисунок 19).

Рисунок 19 – Параметры, от которых зависит выбор модели

«Обучение модели на этом этапе происходит усовершенствование модели и появляется возможность обучить модель на исторических данных. Для этого предлагается использовать параметры, минимизирующих ошибку (например, среднеквадратическая ошибка, MAE) или настроить гиперпараметры с помощью кросс-валидации» [34].

Проверка модели - метрики качества прогноза предлагаются использовать следующие (рисунок 20).

«Рассмотрим метрики более подробно:

– MAE (Mean Absolute Error): Средняя абсолютная ошибка.
– MAPE (Mean Absolute Percentage Error): Средняя абсолютная процентная ошибка.
– Визуальная проверка: сопоставление фактических данных с прогнозом.
– Тестирование модели на тестовом наборе данных» [34].

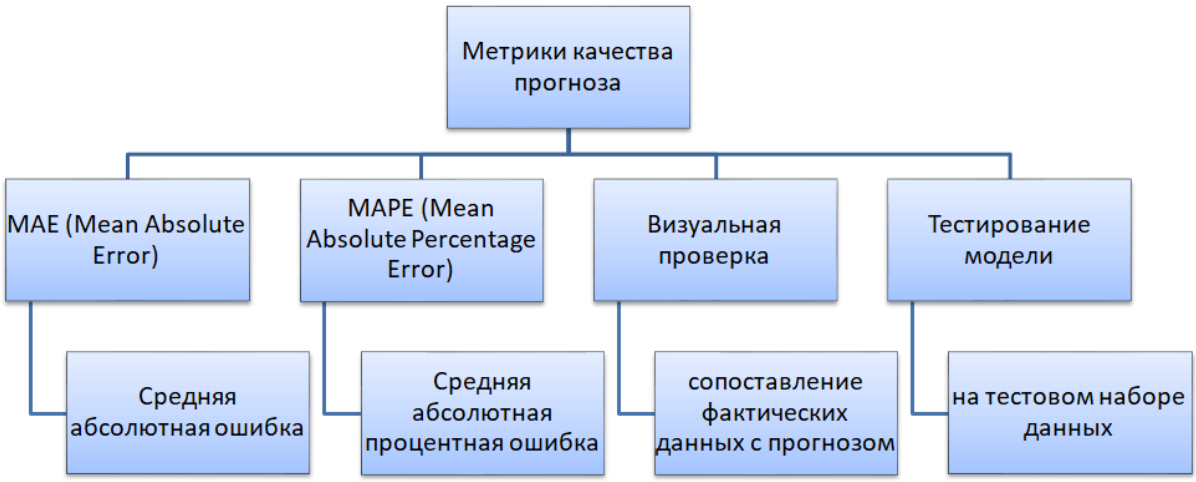

Рисунок 20 – Метрики качества прогноза

«Прогнозирование - на этом этапе происходит следующие:

– Применяется обученная модель к новым данным.
– Генерация прогнозов на заданный горизонт (например, следующие 3 месяца).
– Обработка неопределенности (например, доверительные интервалы)» [34].

«Анализ и интерпретация:

– Оценка значимости прогнозируемых показателей.
– Визуализация прогнозов (графики, таблицы, дашборды).

Определение рисков и сценариев (например, оптимистический, пессимистический)» [34].

«Документирование результатов:

– Подготовка отчетов с результатами прогнозирования.
– Выводы и рекомендации на основе прогноза.
– Передача данных заинтересованным сторонам (руководству, маркетинговым или операционным отделам)» [34].

«Внедрение и пересмотр:

– Использование прогнозов для принятия решений.
– Постоянный мониторинг реальных данных.

– Адаптация модели на основе новых данных (переподготовка)» [34].

«Процесс моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных является итеративным процессом, т.е может проходить некоторые этапы многократно для улучшения прогнозов» [34].

Вывод по второй главе

Во второй главе проведено проектирования бизнес-процессов, которые отражают текущее состояние и состояние после внедрения новых методов моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных.

Процесс прогнозирования включает несколько этапов, которые объединяют подготовку данных, построение моделей и интерпретацию результатов.

Внедрение новых методов моделирования и прогнозирования данных позволит взаимодействовать с различными источниками входных данных, а также позволит получать различные прогнозы (долгосрочные, среднесрочные и краткосрочные).

Процесс моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных является итеративным процессом.

3 Разработка приложения с использованием предложенных методов и моделей моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных

3.1 Разработка приложения моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных

Для практической части магистерской работы предлагается разработать модули прогнозирования и анализа трендов на Python с последующей интеграцией в CRM-систему организации.

Работа с данными (рисунок 21), на этом этапе используются две библиотеки pandas – в качестве обработки временных и numpy – в качестве операции с массивами и проведение математических вычислений.

```
import pandas as pd
import numpy as np

# Генерация временного ряда
dates = pd.date_range('2023-01-01', periods=12, freq='M')
values = np.random.rand(12) * 100
data = pd.DataFrame({'Date': dates, 'Values': values})
```

Рисунок 21 – Работа с данными

Анализ трендов на этом этапе используются следующие методы методы скользящих средних, декомпозицию временных рядов и линейная регрессия (рисунок 22).

```
from statsmodels.tsa.seasonal import seasonal_decompose

result = seasonal_decompose(data['Values'], model='additive', period=12)
result.plot()
```

Рисунок 22 – Скрипт для анализа трендов

Построение прогнозов на этом этапе используются ARIMA / SARIM и Auto-ARIMA (рисунок 23).

```
from statsmodels.tsa.arima.model import ARIMA

# Создание и обучение модели
model = ARIMA(data['Values'], order=(1, 1, 1))
model_fit = model.fit()

# Прогноз
forecast = model_fit.forecast(steps=5)
```

Рисунок 23 – Скрипт для построения прогнозов

Регрессия (линейная, полиномиальная, регрессия с ядром): подходит для построения прогнозов на основе набора признаков (рисунок 24).

```python
from sklearn.linear_model import LinearRegression

# Обучение модели
X = np.arange(len(data)).reshape(-1, 1)  # Временной индекс
y = data['Values']
model = LinearRegression()
model.fit(X, y)

# Прогноз
future_index = np.arange(len(data), len(data) + 5).reshape(-1, 1)
forecast = model.predict(future_index)
```

Рисунок 24 – Скрипт для реализации регрессии

Визуализация данных и прогнозов использует две библиотеки matplotlib и seaborn (рисунок 25).

```python
import matplotlib.pyplot as plt

plt.plot(data['Date'], data['Values'], label='Исходные данные')
plt.plot(future_index, forecast, label='Прогноз', linestyle='--')
plt.legend()
plt.show()
```

Рисунок 25 – Скрипт для визуализации данных

Преимущества Python для прогнозирования:

- Универсальность: подходит как для простых моделей (линейная регрессия), так и для сложных (глубокое обучение).
- Богатый экосистемный набор библиотек: для анализа данных: pandas, numpy, scipy. Для моделей: statsmodels, scikit-learn, xgboost, tensorflow, prophet. Для визуализации: matplotlib, seaborn, plotly.
- Доступность готовых решений: Большинство библиотек хорошо документированы, а в интернете много примеров.

– Интеграция с другими инструментами: Python легко соединяется с базами данных, веб-интерфейсами и облачными сервисами.

Рассмотрим модули, написанные на Python, которые используют метрики, разработанные ранее. Для построения прогнозов с использованием библиотеки scikit-learn будем строить прогноз с использованием линейной регрессии (рисунок 26).

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# 1. Подготовка данных
# Пример данных: Зависимость продаж от расходов на рекламу
data = {
    'Ad_Spend': [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000],
    'Sales': [4, 7, 10, 15, 21, 27, 34, 44, 50, 60]
}

df = pd.DataFrame(data)

# 2. Разделение данных на признаки (X) и целевую переменную (y)
X = df[['Ad_Spend']]
y = df['Sales']

# 3. Разделение данных на обучающую и тестовую выборки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 4. Построение модели
model = LinearRegression()
model.fit(X_train, y_train)

# 5. Прогнозирование
y_pred = model.predict(X_test)

# 6. Оценка модели
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("Mean Squared Error (MSE):", mse)
print("R-squared (R2):", r2)

# 7. Визуализация
plt.scatter(X, y, color='blue', label='Исходные данные')
plt.plot(X, model.predict(X), color='red', label='Линия регрессии')
plt.xlabel('Расходы на рекламу')
plt.ylabel('Продажи')
plt.title('Прогноз продаж')
plt.legend()
plt.show()
```

Рисунок 26 – Скрипт для реализации линейной регрессии

Рассмотрим, что выполняется в коде, первое – загружаются данные, а именно создаются данные о расходах на рекламу и продажах. Далее, происходит разделение данных. Данные делятся на обучающую и тестовую выборки. Строится модель линейной регрессии с использованием библиотеки scikit-learn. Выполняется прогнозирование значений на тестовой выборке. Рассчитываются метрики качества модели, такие как MSE и R^2. Визуализация: рисуется график, показывающий исходные данные и линию регрессии.

Используются те же данные, что и в примере с линейной регрессией. Используется модель дерева решений для регрессии. Параметр max_depth=4 задаёт максимальную глубину дерева, чтобы избежать переобучения. Метрики качества (MSE и R^2) показывают, насколько хорошо дерево решений справляется с задачей. Линия (или ступени) прогнозируемых значений показывают, как дерево решений разделяет данные.

Далее для построения прогноза с использованием деревьев решений на Python с библиотекой scikit-learn будем использовать модель регрессии DecisionTreeRegressor (рисунок 27).

Для более сложных задач можно экспериментировать с параметрами дерева, такими как max_depth, min_samples_split и min_samples_leaf, чтобы улучшить производительность.

Деревья решений хорошо справляются с нелинейными зависимостями и небольшими наборами данных, но могут переобучаться на сложных выборках.

Далее для использования деревьев решений для задачи классификации с библиотекой scikit-learn будем использовать модель DecisionTreeClassifie (рисунок 28).

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error, r2_score

# 1. Подготовка данных
# Пример данных: Зависимость продаж от расходов на рекламу
data = {
    'Ad_Spend': [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000],
    'Sales': [4, 7, 10, 15, 21, 27, 34, 44, 50, 60]
}

df = pd.DataFrame(data)

# 2. Разделение данных на признаки (X) и целевую переменную (y)
X = df[['Ad_Spend']]
y = df['Sales']

# 3. Разделение данных на обучающую и тестовую выборки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 4. Построение модели
model = DecisionTreeRegressor(random_state=42, max_depth=4)
model.fit(X_train, y_train)

# 5. Прогнозирование
y_pred = model.predict(X_test)

# 6. Оценка модели
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("Mean Squared Error (MSE):", mse)
print("R-squared (R2):", r2)

# 7. Визуализация
# Построим прогнозную кривую
X_range = np.arange(min(X['Ad_Spend']), max(X['Ad_Spend']), 1).reshape(-1, 1)
y_range_pred = model.predict(X_range)

plt.scatter(X, y, color='blue', label='Исходные данные')
plt.plot(X_range, y_range_pred, color='red', label='Дерево решений (прогноз)')
plt.xlabel('Расходы на рекламу')
plt.ylabel('Продажи')
plt.title('Прогноз продаж (Дерево решений)')
plt.legend()
plt.show()
```

Рисунок 27 – Реализация дерева решений

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.metrics import classification_report, accuracy_score

# 1. Подготовка данных
# Пример данных: Зависимость успешности рекламной кампании (1 - успех, 0 - не успех)
data = {
    'Ad_Spend': [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000],
    'Sales': [4, 7, 10, 15, 21, 27, 34, 44, 50, 60],
    'Success': [0, 0, 0, 1, 1, 1, 1, 1, 1, 1]  # Метка классов
}

df = pd.DataFrame(data)

# 2. Разделение данных на признаки (X) и целевую переменную (y)
X = df[['Ad_Spend', 'Sales']]
y = df['Success']

# 3. Разделение данных на обучающую и тестовую выборки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 4. Построение модели
model = DecisionTreeClassifier(random_state=42, max_depth=3)
model.fit(X_train, y_train)

# 5. Прогнозирование
y_pred = model.predict(X_test)

# 6. Оценка модели
print("Classification Report:")
print(classification_report(y_test, y_pred))
print("Accuracy:", accuracy_score(y_test, y_pred))

# 7. Визуализация дерева
plt.figure(figsize=(10, 6))
plot_tree(model, feature_names=['Ad_Spend', 'Sales'], class_names=['Not Success', 'Success
plt.title("Дерево решений")
plt.show()
```

Рисунок 28 – Реализация дерева решения для задачи классификации

В примере используется набор данных, где Ad_Spend (расходы на рекламу) и Sales (продажи) используются как признаки. Целевая переменная Success (успешность) имеет два класса: 1 (успех) и 0 (неуспех). Модель дерева

решений для классификации. Параметр max_depth=3 ограничивает глубину дерева, чтобы избежать переобучения. Оценка: classification_report показывает метрики, такие как точность (precision), полнота (recall) и F1-score для каждого класс и accuracy_score вычисляет общую точность модели.

Используется plot_tree, чтобы отобразить структуру дерева решений, включая разделения по признакам, значения и вероятности классов.

Этот пример показывает, как классифицировать данные с использованием деревьев решений.

Если данные содержат больше признаков или классов, то можно адаптировать код, добавляя дополнительные данные и изменяя параметры модели, такие как max_depth или criterion (например, gini или entropy).

Далее для анализа и прогнозирования временных рядов используем библиотеку statsmodels, prophet или scikit-learn. Для построения прогноза временного ряда с использованием библиотеки statsmodels и модели ARIMA (рисунок 29).

Данные генерируются временные ряды с линейным трендом и шумом. Можно заменить их своими данными, загрузив их из файла, например, с помощью pd.read_csv(). ARIMA — это модель, основанная на авторегрессии (AR), интегрировании (I) и скользящем среднем (MA).

Параметры:

– p: Порядок авторегрессии.
– d: Разности для стационарности.
– q: Порядок скользящего среднего.

Для оценки модели используется Mean Squared Error (MSE), который показывает, насколько предсказания близки к реальным значениям. График показывает исходные данные, тестовую выборку и прогноз.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from sklearn.metrics import mean_squared_error

# 1. Генерация данных временного ряда
# Пример: Ежемесячные продажи
np.random.seed(42)
dates = pd.date_range(start="2020-01-01", periods=36, freq="M")
sales = 50 + 2 * np.arange(36) + np.random.normal(scale=5, size=36)  # Линейный тренд + шу

df = pd.DataFrame({"Date": dates, "Sales": sales})
df.set_index("Date", inplace=True)

# 2. Визуализация временного ряда
plt.figure(figsize=(10, 6))
plt.plot(df.index, df['Sales'], label='Продажи')
plt.title('Временной ряд: Продажи')
plt.xlabel('Дата')
plt.ylabel('Продажи')
plt.legend()
plt.show()

# 3. Построение модели ARIMA
# Задание параметров ARIMA (p, d, q)
p, d, q = 2, 1, 2

# Разделение на обучающую и тестовую выборки
train_size = int(len(df) * 0.8)
train, test = df[:train_size], df[train_size:]

# Обучение модели
model = ARIMA(train['Sales'], order=(p, d, q))
model_fit = model.fit()

# 4. Прогнозирование
forecast = model_fit.forecast(steps=len(test))

# 5. Оценка модели
mse = mean_squared_error(test['Sales'], forecast)
print(f"Mean Squared Error (MSE): {mse}")

# 6. Визуализация прогноза
plt.figure(figsize=(10, 6))
plt.plot(train.index, train['Sales'], label='Обучающая выборка')
plt.plot(test.index, test['Sales'], label='Тестовая выборка')
plt.plot(test.index, forecast, label='Прогноз', linestyle='dashed', color='red')
plt.title('Прогнозирование временного ряда (ARIMA)')
plt.xlabel('Дата')
plt.ylabel('Продажи')
plt.legend()
plt.show()
```

Рисунок 29 – Реализация построения прогноза модели ARIMA

Для работы с более сложными временными рядами используются библиотеки показанные на рисунке 30.

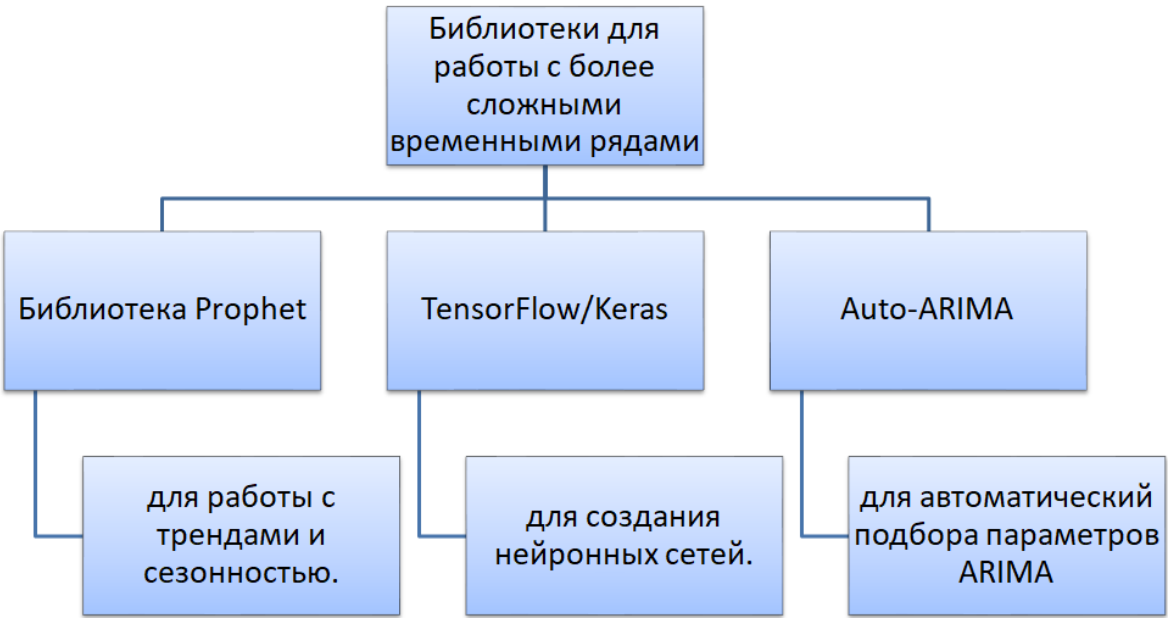

Рисунок 30 – Библиотеки для сложного прогнозирования

Рассмотрим содержание библиотек подробнее:

– Библиотека Prophet - используется для работы с трендами и сезонностью.
– TensorFlow/Keras – используется для создания нейронных сетей (например, LSTM) для прогнозирования временных рядов.
– Auto-ARIMA – используется для автоматический подбора параметров ARIMA (библиотека pmdarima).

Далее рассмотрим, как шаги проектирования отразились в разработке приложения.

3.2 Апробация полученных результатов

В ходе работы над магистерской диссертацией и ее практической реализацией, выбрана система, использующаяся на предприятии для сбора

данных - ELMA, которая позволяет добавлять для анализа программные модули, разработанные ранее на языке Python. Добавление модулей происходит с помощью встроенного инструмента создания приложений.

ELMA предоставляет мощные и гибкие инструменты для работы с аналитикой, что позволяет компаниям не только отслеживать текущие показатели, но и прогнозировать будущее, выявлять тренды и оптимизировать бизнес-процессы. Интеграция с BI-системами и другими платформами делает ELMA универсальным решением для стратегического управления, помогая компаниям принимать обоснованные решения на основе данных и повышать свою конкурентоспособность на рынке.

Последовательность работы с приложением и добавление в него блоков показана на рисунке 31.

Рисунок 31 – Последовательность работы с приложением

Весь процесс работы в CRM системе ELMA основан на LowCode. Работа с индивидуальными настройками и кодом начинается в панели Администрирования, которая показана на рисунке 32.

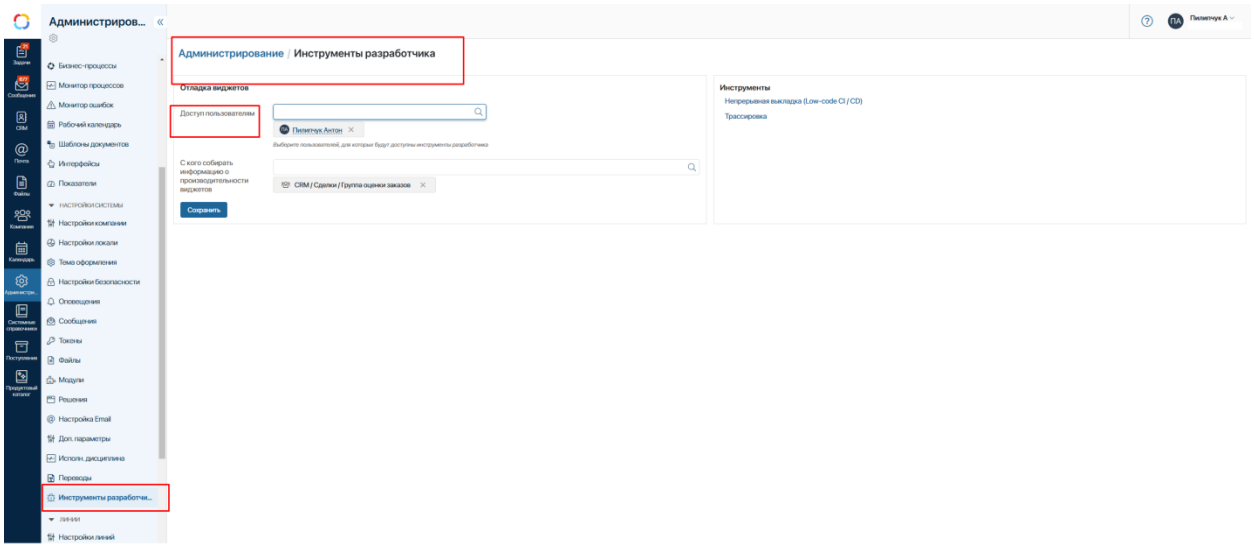

Рисунок 32 – Панель администрирования. Добавление модулей

На первом этапе выбирается процесс, которого будем вести мониторинг и проводить анализ (рисунок 33).

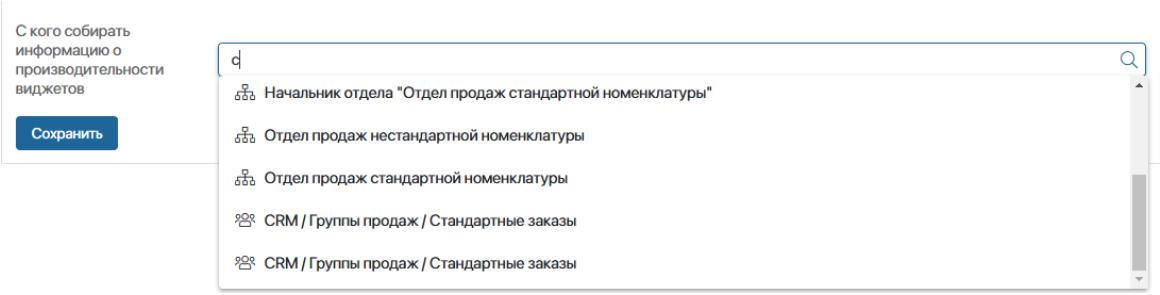

Рисунок 33 – Сбор данных о процессе

В системе предусмотрена гибкая настройка прав доступа согласно бизнес-процессам. Видимость данных и разрешенные действия определяются в зависимости от ролей и обязанностей сотрудника. Для корректного

функционирования бизнес-процесса необходимо определить:

- Уровень доступа – какие модули и данные доступны конкретному пользователю.
- Роли – привязка пользователя к определенной роли (например, аналитик, менеджер, администратор).
- Ограничения и фильтрация данных – защита конфиденциальной информации, ограничение доступа по регионам, подразделениям или иным критериям.

После настройки доступа можно переходить к анализу данных. В качестве объекта анализа выбраны продажи компании. Основные цели исследования:

- Оценка текущей динамики продаж – выявление трендов, сезонности и факторов, влияющих на объем продаж.
- Прогнозирование будущих продаж – построение предсказательной модели для планирования производства, логистики и стратегического развития.

Процессы должны быть формализованы в нотации BPMN. Каждая операция в рамках анализа и прогнозирования должна быть описана с учетом логики работы бизнес-процессов. Готовая модель загружается на сервер для дальнейшего выполнения и интеграции с системой управления бизнес-процессами.

Применение BPMN позволяет четко структурировать логику процесса, автоматизировать выполнение задач и обеспечить гибкость в изменении бизнес-логики. После загрузки модели система автоматически выполняет шаги анализа и прогнозирования, позволяя оперативно получать обновленные прогнозы и анализировать изменения в продажах. (рисунок 34).

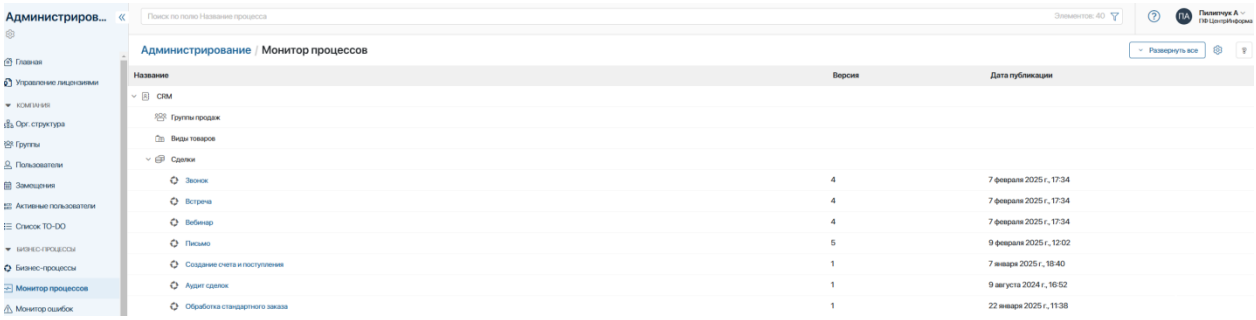

Рисунок 34 – Выбор модели

Изучаемые модели процесса продаж компании, показаны на рисунках ниже (рисунок 35 -36).

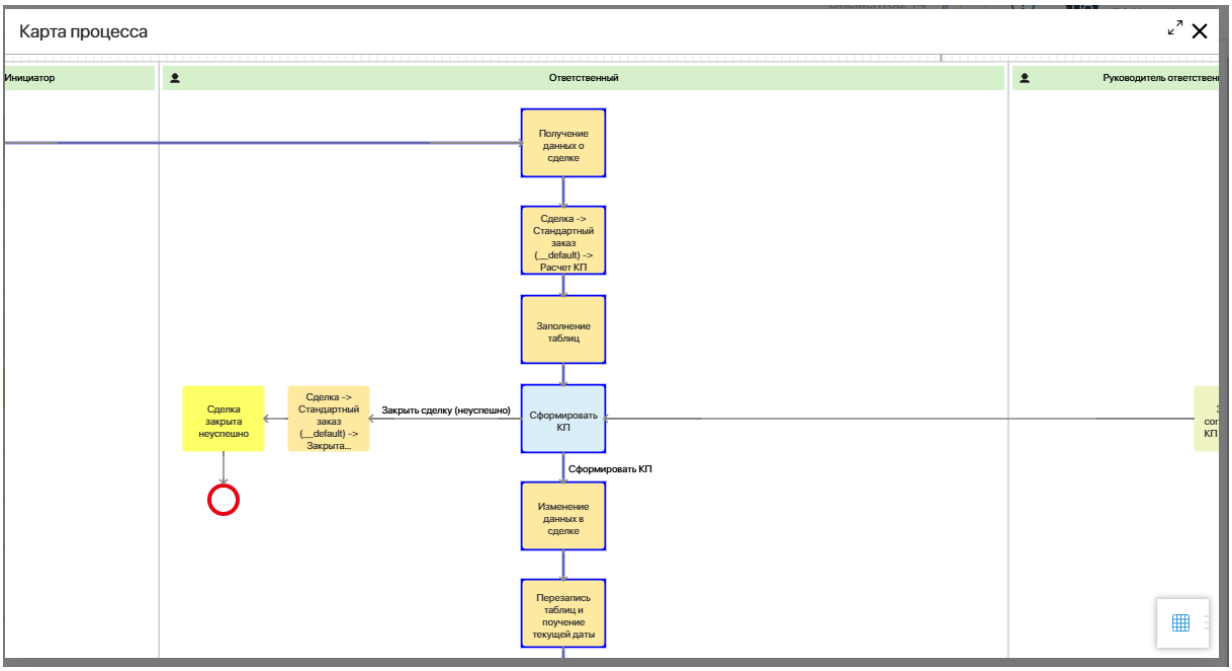

Рисунок 35 –Модель процесса стандартной сделки

В модели кроме стандартов блоков, могут быть использованы процессы, ранее загруженные на сервер, например при анализе счетов и поступлений денежных средств, есть вложенный процесс стандартной сделки (рисунок 37).

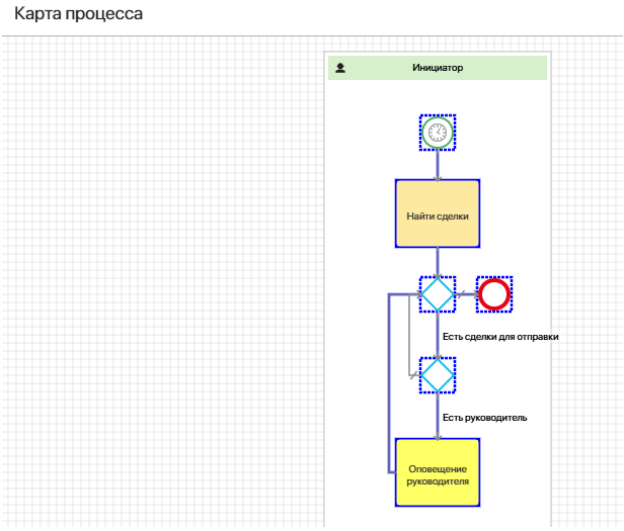

Рисунок 36 –Модель процесса аудита сделки

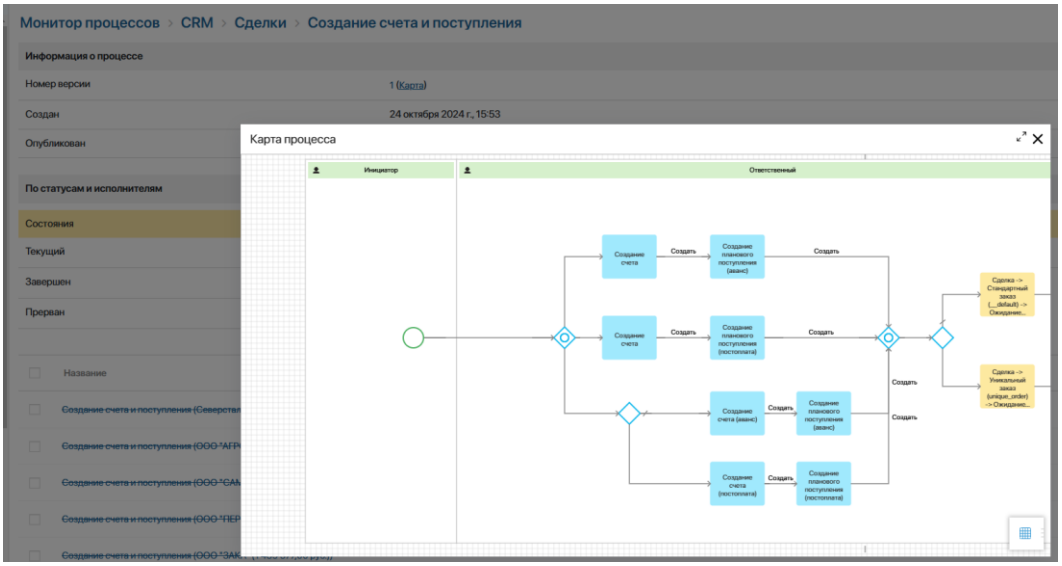

Рисунок 37 –Модель сложного процесса

Далее согласно алгоритму, показанному на рисунке 38 необходимо провести обучение модели и ее проверку. Для этого будем собирать показатели каждого процесса, чтобы в дальнейшем проводить необходимые корректировки и анализировать количественные показатели. При создании имени показателя код показателя автоматически переименовывается транслитом.

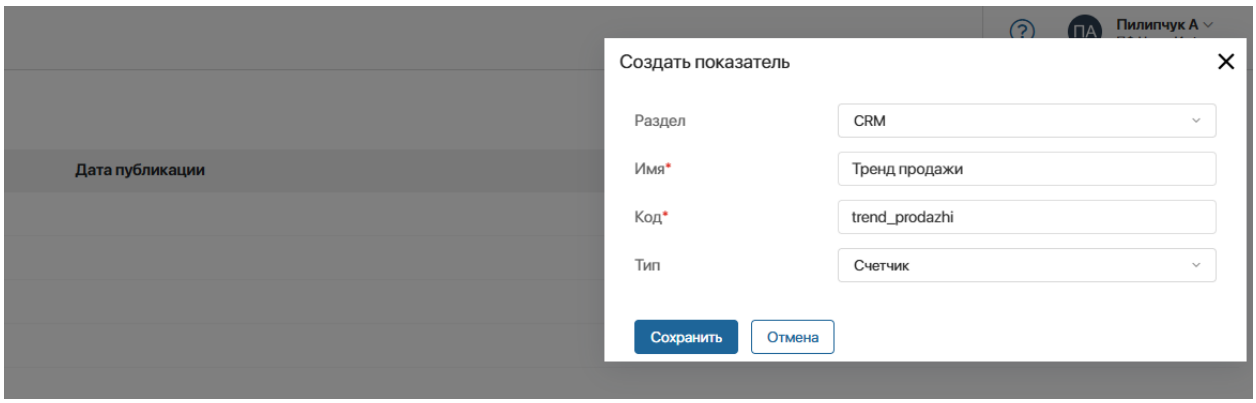

Рисунок 38 –Сбор показателей каждого процесса

В дальнейшем для анализа и построение тренда использованы следующие два значения показателей – счетчик (отражение количества сделок, количество проданного товара, количества клиентов) и временной интервал (временной интервал на сделку, временной интервал ответа клиента, частота временного повторного и последующего обращения клиента) (рисунок 39).

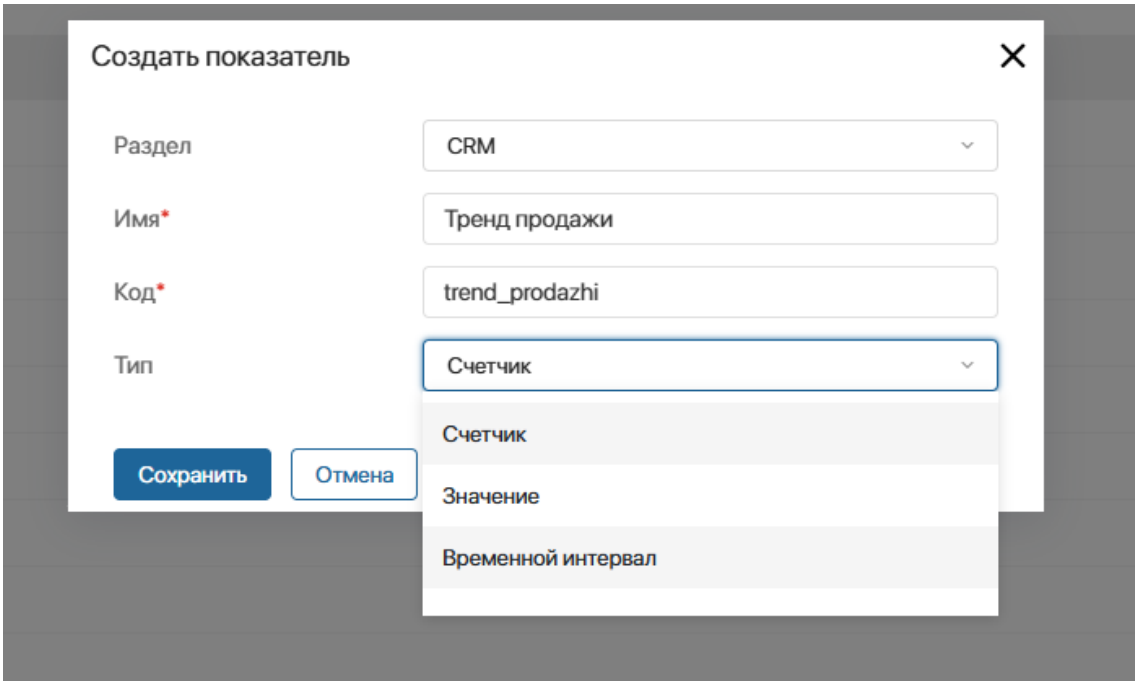

Рисунок 39 – Тип показателя процесса

Настройки отображения и загрузки показателей и их обработка находится в панели настроек и в ходе работы над выпускной работой проводились различные конфигурации настроек, для удобства отображения и учета всех способов проведения анализа, рассмотренных выше (рисунок 40 - 41).

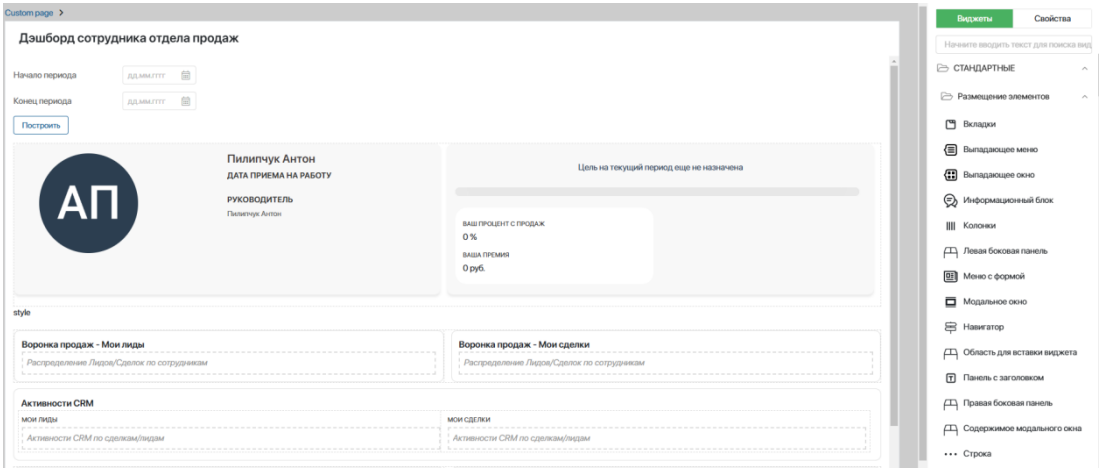

Рисунок 40 – Работа по настройке страниц, с использованием кода

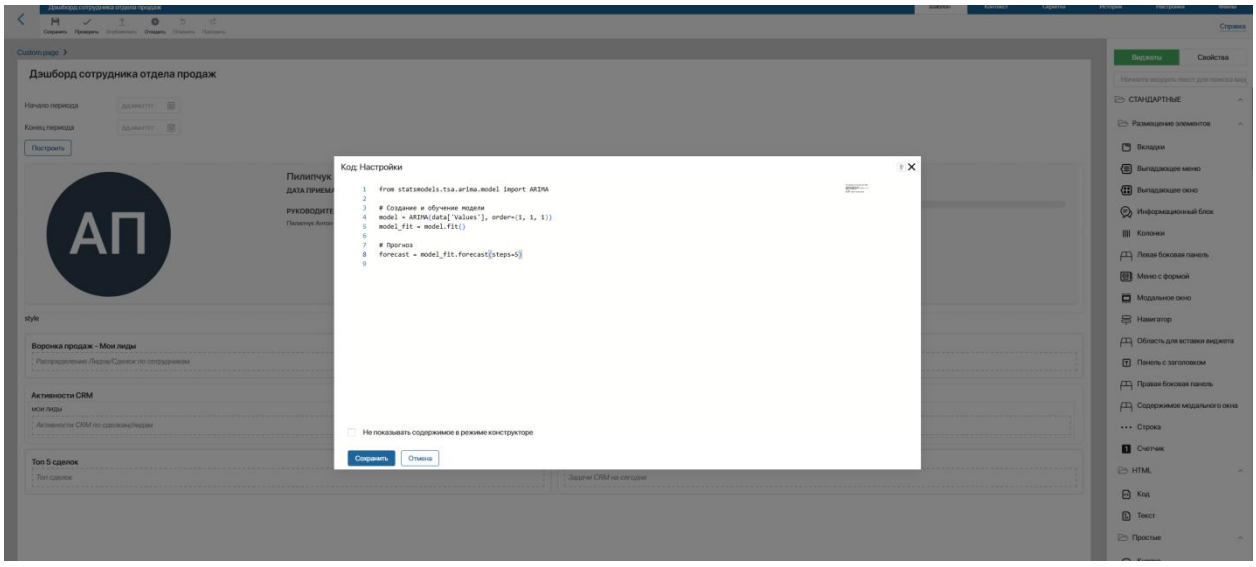

Рисунок 41 – Загрузка скриптов и настройка вывода результатов на страницу

В коде важно указать имя параметров, которое будет совпадать с ранее созданными и хранимыми показателями (рисунок 42).

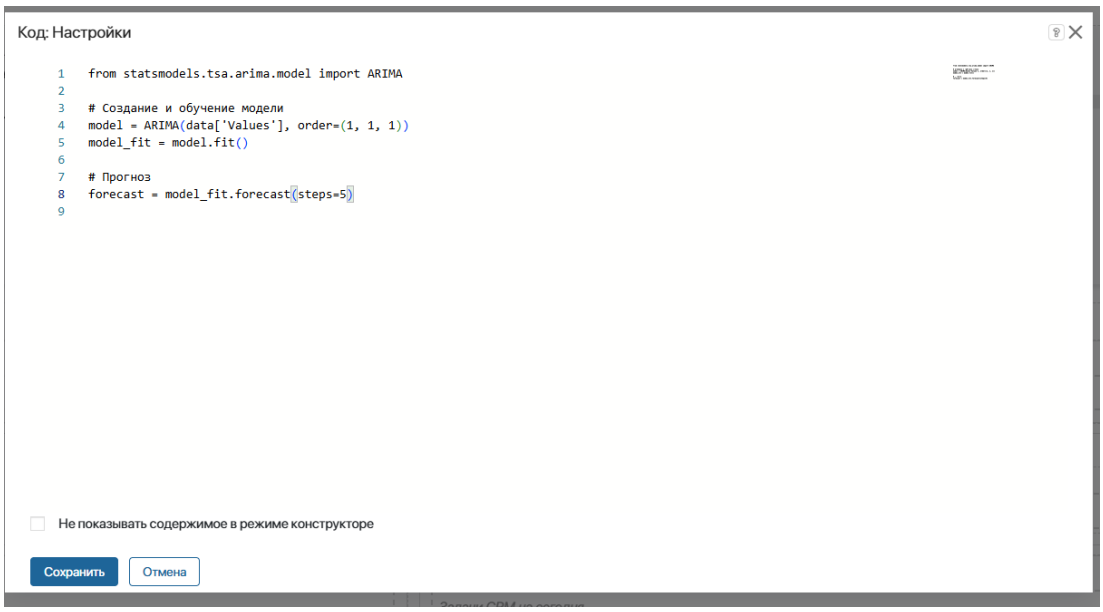

Рисунок 42 – Вставка кода в страницу

Все настройки графиков, тип графиков и показатели ранее созданные выводятся с помощью настройки страницы и показаны на рисунке ниже (рисунок 43).

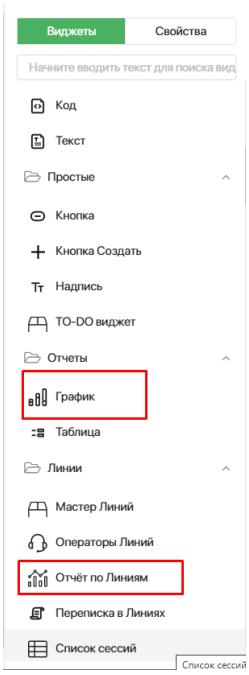

Рисунок 43 – Настройка и добавления графика на страницу

Выбор типа графика, показатели, цветовые схемы – это все стандартные процессы в работе CRM системы (рисунок 44-45).

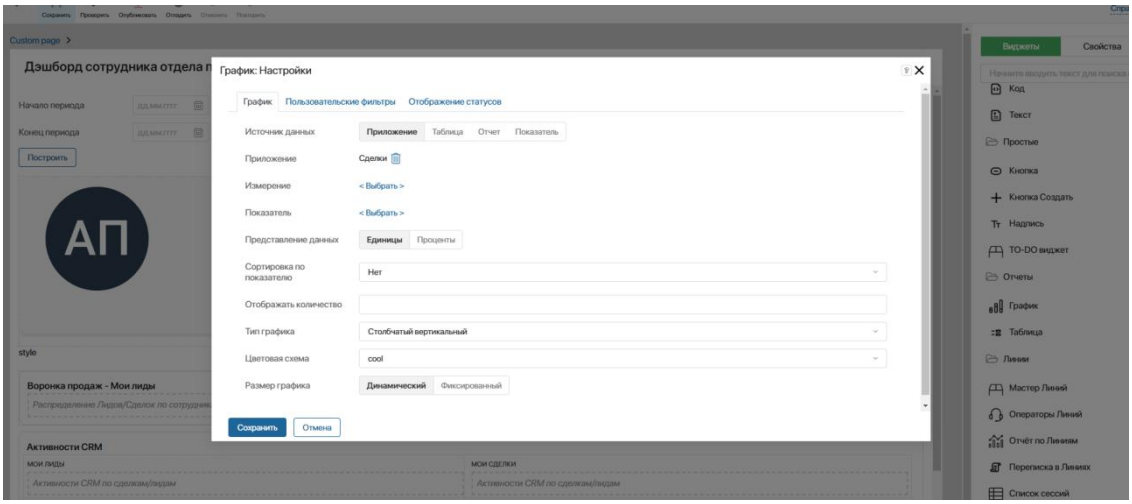

Рисунок 44 – Настройка графика и параметров вывода на страницу

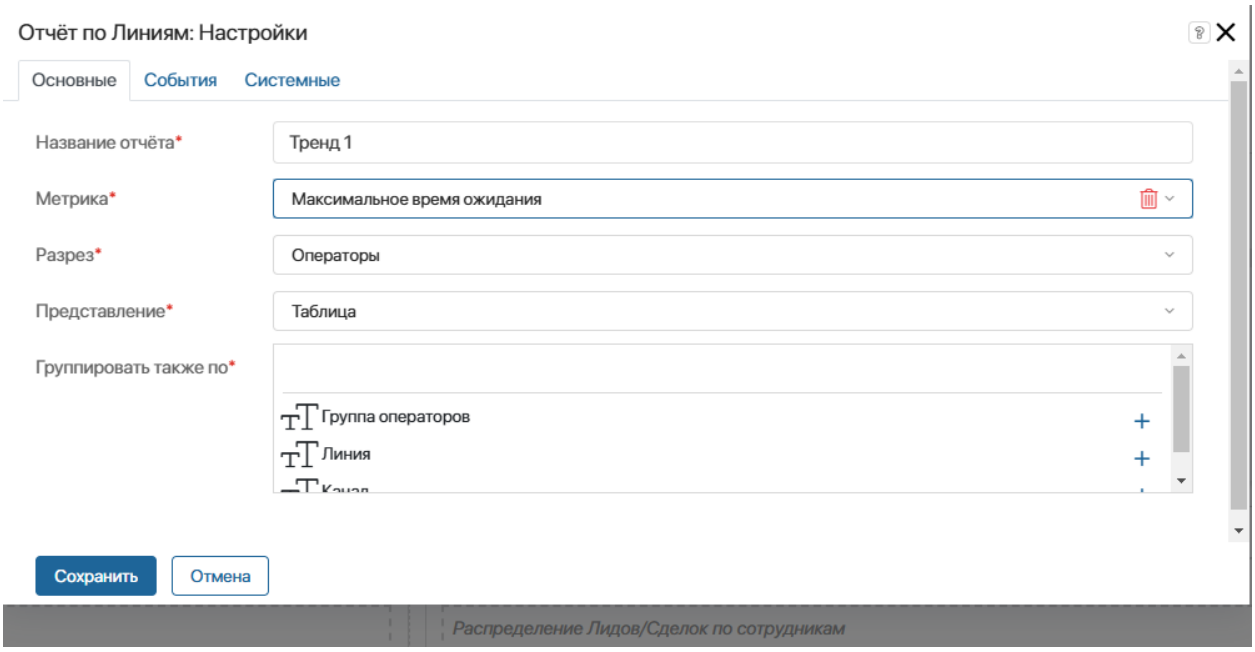

Рисунок 45 – Загрузка скриптов и настройка вывода результатов на страницу

В ходе выпускной работы, были проанализированы различные подходу к отображению данных и выбрана следующая стратегия, которая используется во множестве рекомендаций, а именно:

– линейный график используют, если анализируете тренды и временные ряды.
– столбчатая или круговая диаграмма используется, если сравнивают категории.
– гистограмму используют, если необходим анализ распределения
– круговую диаграмму используют, если исследуют структуру.

Результаты настроек по анализу продаж показаны с помощью различных типов графиков, с учетом рекомендаций показаны на рисунках 46-48.

Рисунок 46 – Настройка отображения анализа с помощью гистограмм

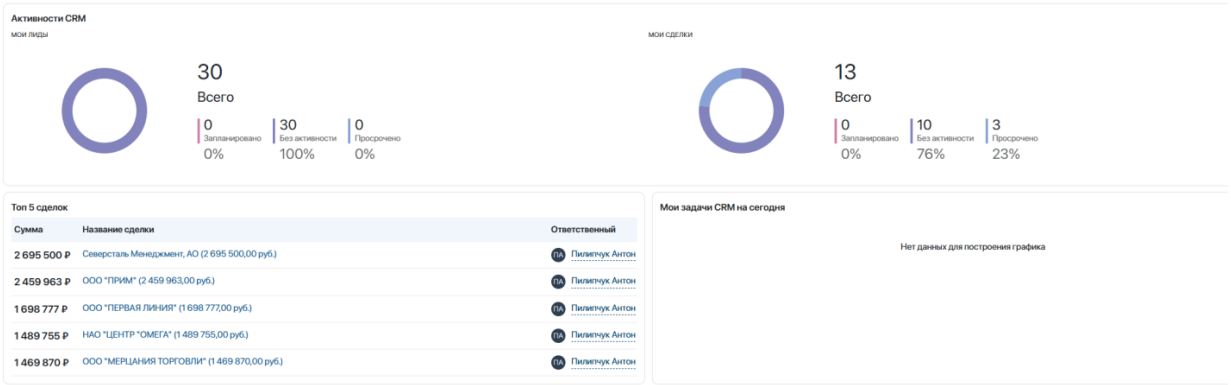

Рисунок 47 – Настройка отображения анализа с помощью круговых диаграмм

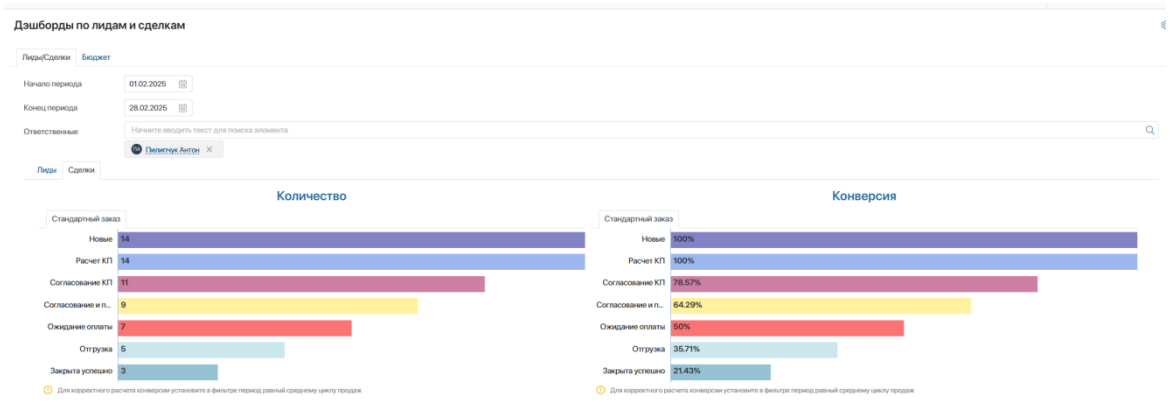

Рисунок 48 – Настройка отображения анализа с помощью диаграмм

В работе также кроме статического анализа используется работа с помощью динамического анализа моделей стандартной сделки и уникальных заказов, на рисунке 49 и 50 показан как раз результат анализа показателей собранных по каждому блоку моделей.

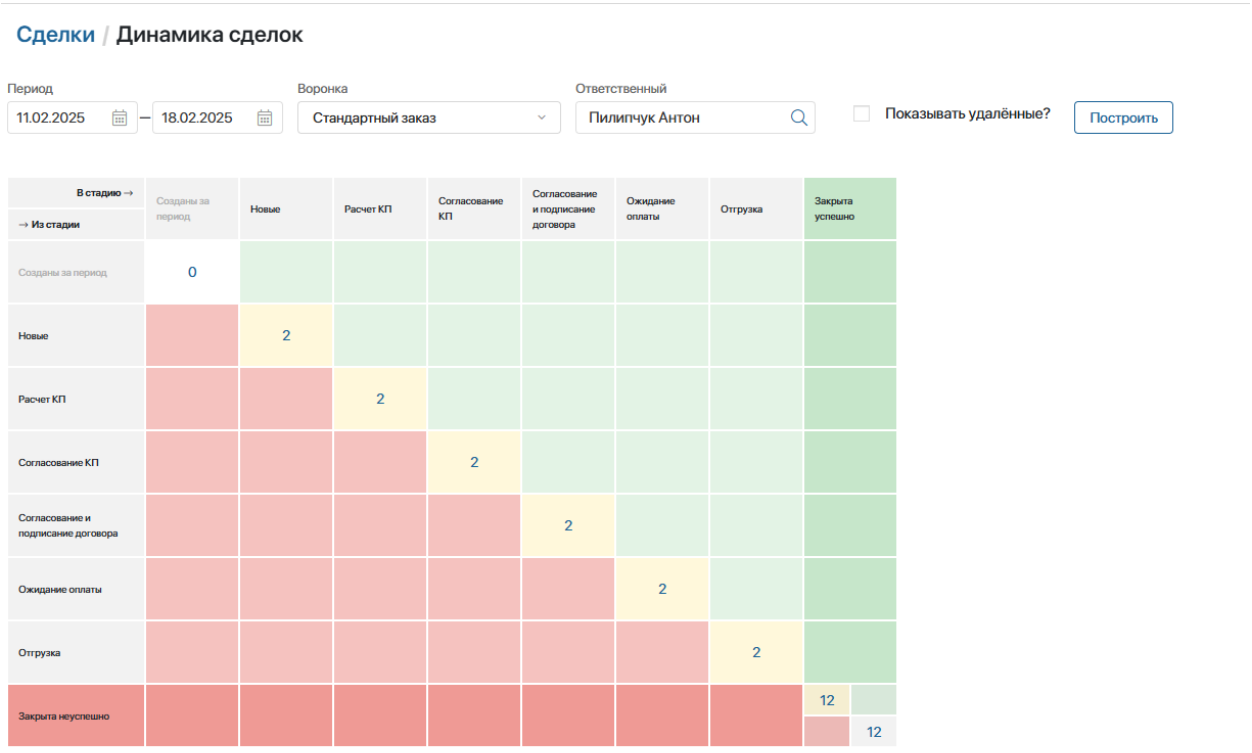

Рисунок 49 – Загрузка скриптов и настройка вывода результатов на страницу

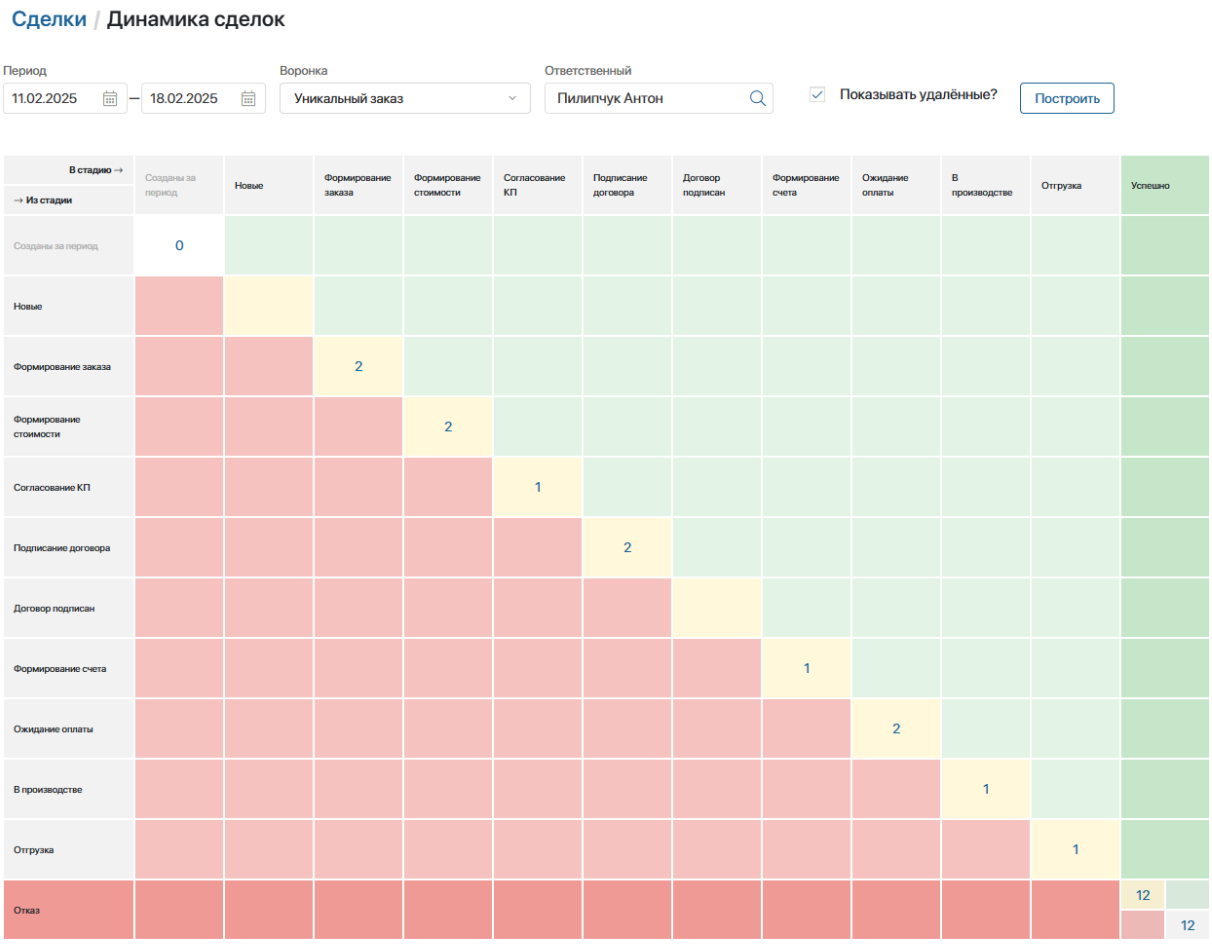

Рисунок 50 – Загрузка скриптов и настройка вывода результатов на страницу

ELMA предлагает гибкие инструменты для работы с графиками и трендами, что позволяет бизнесу отслеживать показатели, анализировать данные и прогнозировать будущее. Используя встроенные возможности визуализации и интеграции с BI-платформами, компании могут эффективно управлять аналитикой и принимать стратегические решения.

В результате проделанной работы, показано, что программа работает с данными системы и может строить тренда и использовать разработанные модели (рисунок 51).

Внедрение аналитических инструментов ELMA позволяет компаниям:

– Повышать прозрачность бизнес-процессов, за счет глубокого анализа данных и выявления узких мест.

– Сокращать время на подготовку отчетности, благодаря автоматизированному сбору и обработке данных.
– Повышать точность прогнозов, что дает возможность минимизировать риски и более эффективно управлять ресурсами.
– Оптимизировать стратегические решения, за счет комплексного анализа данных и построения моделей развития.

Рисунок 51 – Анализ трендов

В ходе работы изучена возможность работы с LowCode система, которые позволяют сосредоточиться на анализе полученных данных, предоставляя инструменты анализа и настройки графиков, а также с возможностью работать со сторонними языками программирования и библиотеками. Настройка анализа каждого показателя длительный процесс, в ходе работы проанализированы возможности сбора информации о каждом процессе, причем каждый процесс требовалась описать в виде модели, у каждого блока настроить сбор показателей и потом эти показатели отобразить.

Анализирую работу этого этапа по моделированию тренда, можно привести следующие результаты (таблица 5).

Таблица 5- Принятие решений по продажам

Прогноз с помощью моделирования тренда		
Критерий	ДО внедрения	ПОСЛЕ внедрения
Время на создание запроса для анализа, ч	3	0,5
Процент запросов, используемых для анализа, %	67%	75%
Процент запросов, которые необходимо переделать, %	85%	30%

Полученные данные представлены в виде графика, для наглядности результата и визуализация решений по разрабатываемым аналитиком запросам представлена на рисунке 52 и 53.

Рисунок 52 – Графическое представление времени на создания запроса

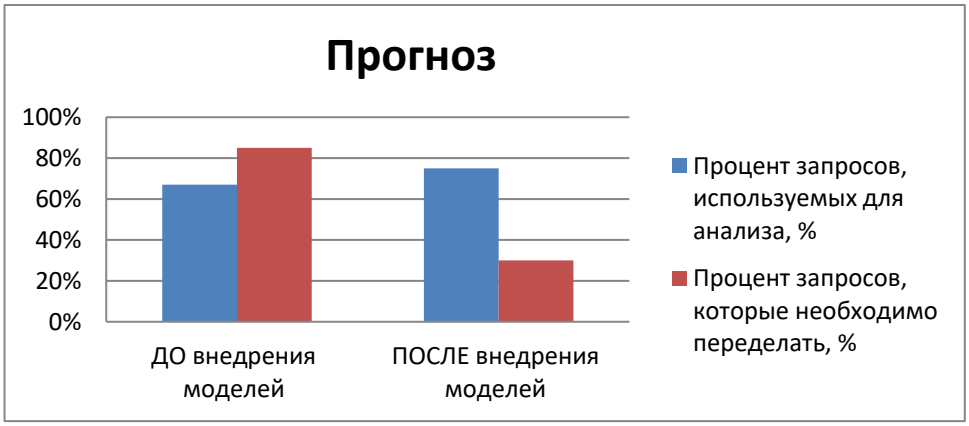

Рисунок 53 – Графическое представление прогнозных решения

Как видно из приведенного анализа после внедрения предложенных моделей, за счет наличия в системе различных фильтров, время создания запроса уменьшается, процент запросов, которые в дальнейшем используются при оперативном анализе выше и наличие тренда упрощает представление прогноза.

Вывод по третьей главе

Полученное программное решение позволило сократить время на анализ данных, тиражировать решение на каждого менеджера компании, а также позволило получить опыт работы с CRM-системой

При выполнении магистерской работы «Моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных» показана практическая реализация модуля системы, в которой используются ранее разработанные модели и строятся тренды на основе данных.

Метрики качества прогноза используются для количественной оценки точности модели прогнозирования, чтобы понять, насколько модель хорошо справляется с задачей.

Метрики и методы используются вместе, чтобы получить объективную и всестороннюю оценку качества модели

После внедрения предложенных моделей, за счет наличия в системе различных фильтров, время создания запроса уменьшается, процент запросов, которые в дальнейшем используются при оперативном анализе выше и наличие тренда упрощает представление прогноза

Заключение

В ходе выполнения магистерской диссертации по теме «Моделирование тренда и прогноза на основе интеллектуализации обработки статистических данных» получены выводы по всем задачам, которые были поставлены в рамках работы.

В первой главе «Обзор исследований по теме моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных» проведён анализ современного состояния исследований в области управления моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных. Этот анализ показал, что данная тема является актуальной и требует более глубокого изучения.

В этой главе были рассмотрены основополагающие вопросы моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных, а также показано практическое применение и характеристика каждой метрики процесса. Были определены основные принципы моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных, которые позволяют повысить эффективность работы компании.

Одним из перспективных методов является интеграция новейших информационных технологий, связанных со сбором и анализом больших неструктурированных данных, работой с данными с помощью интеллектуальной обработки данный (анализ временных рядов), а также внедрение отдельных прогностических моделей в обработку данных в работающую CRM-систему организации.

Во второй главе магистерской работы «Анализ методов и моделей моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных» рассмотрены методы и модели моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных.

В рамках исследования были определены источники данных для каждого бизнес-процесса модели, а также выявлены зависимости между каждым информационным потоком и метрикой, которая в дальнейшем используется для оценки эффективности процесса.

В третьей главе «Разработка приложения с использованием предложенных методов и моделей моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных» Выполнена апробация и оценка эффективности проектных решений по качественным и количественным характеристикам, которые показали увеличение эффективности работы аналитиков при моделировании тренда и прогноза на основе интеллектуализации обработки статистических данных.

Метрики качества прогноза используются для количественной оценки точности модели прогнозирования, чтобы понять, насколько модель хорошо справляется с задачей.

Эти метрики и методы используются вместе, чтобы получить объективную и всестороннюю оценку качества модели. Например, MAE показывает общую ошибку, RMSE подчеркивает крупные отклонения, MAPE полезен для сравнения, а визуальная проверка позволяет глубже понять ошибки модели.

Процесс моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных является итеративным процессом, т.е может проходить некоторые этапы многократно для улучшения прогнозов. Для реализации практической части магистерской работы предлагается разработать модули прогноза и построения тренда на языке Python и внедрить их в CRM системы организации.

В результате проделанной работы, показано, что программа работает с данными системы и может строить тренда и использовать разработанные модели.

Гипотеза исследования подтверждена.

Список используемых источников

1. Астраханцева, И. А. Моделирование систем : учебное пособие / И. А. Астраханцева, С. П. Бобков. — Москва : ИНФРА-М, 2023. — 216 с.
2. Богатырев, С. Ю. Информационные системы в корпоративных финансах [Электронный ресурс]: учеб. пособие / С. Ю. Богатырев. - Москва : РИОР; ИНФРА-М, 2017. - 173 с.
3. Богданов А. Б., Борисова И. А. и др. Интеллектуальный анализ спектральных данных // Автометрия, 2009, № 1. –С. 92–101.
4. Бром, А. Е. Теория и практика моделирования динамики экономических систем в промышленности : монография / А. Е. Бром, В. М. Картвелишвили, И. Н. Омельченко ; под ред. А. Е. Бром. - Москва : МГТУ им. Баумана, 2018. - 216 с.
5. Витязев, В.В. Вейвлет–анализ временных рядов / В.В.Витязев. – Спб.: Изд–во С.–Петерб. унт–та, 2001. – 58 с.
6. Власов, М. П. Моделирование экономических систем и процессов : учеб. пособие / М.П. Власов, П.Д. Шимко. — М.: ИНФРА-М, 2019. — 336 с.
7. Гончаренко, А. Н. Моделирование систем : лабораторный практикум / А. Н. Гончаренко. - Москва : Издательский Дом НИТУ «МИСиС», 2022. - 56 с.
8. Дайзенрот Марк Питер, Чен Сунь Он, Альдо Фейзал А. Математика в машинном обучении. – СПб.: Питер, 2024. – 512 с.: ил. – (Серия «Для профессионалов»)
9. Донской В. И. Алгоритмы поиска аналогий в булевых таблицах эмпирических данных // В. И. Донской, А. В. Ильченко. – Искусственный интеллект. – 2002. – №2. – С. 99–107.
10. Дорофеюк Ю. А. Структурно–классификационные методы анализа и прогнозирования в системах управления // Таврический вестник информатики и математики. – 2008. – №1. – С.166–170.

11. Драганчук, Л. С. Поведение потребителей : учебное пособие / Л.С. Драганчук. — Москва : ИНФРА-М, 2024. — 192 с.

12. Дубров А.М., Мхитрян В.С., Трошин Л.И. Многомерые статистические методы: Учебник. – М.: Финансы и статистика, 2020.

13. Ежов А.А., Шумский С.А. Нейрокомпьютинг и его применения в экономике и бизнесе. Оптимизация с помощью сети Кохонена.

14. Жердев, А. А. Корпоративные информационные системы : практикум / А. А. Жердев. - Москва : Изд. Дом НИТУ «МИСиС», 2021. - 64 с.

15. Интегрированные информационные системы управления объектами. Корпоративные информационные системы : учебное пособие / А. А. Григорьев, Е. А. Исаев, В. В. Корнилов [и др.] ; под ред. А. А. Григорьева. — Москва : ИНФРА-М, 2024. — 273 с.

16. Калянов, Г. Н. Консалтинг: от бизнес-стратегии к корпоративной информационно-управляющей системе: Учебник для вузов / Калянов Г.Н., - 2-е изд., дополн. - Москва :Гор. линия-Телеком, 2016. - 210 с.

17. Кваснов, А. В. Корпоративные информационные системы на промышленных предприятиях : учебное пособие / А. В. Кваснов. - Санкт-Петербург : ПОЛИТЕХ-ПРЕСС, 2019. - 90 с.

18. Кондауров, В. И. Процесс формирования научного знания (онтологический, гносеологический и логический аспекты) : монография / В.И. Кондауров. — Москва : ИНФРА-М, 2024. — 128 с.

19. Кривенко М. П. Обучаемая классификация данных с учётом анализа главных компонент //. Информатика и её применение. 2018. Т 12, №3. – 59–61.

20. Лебедев, В. А. Методологические основы подготовки специалистов-разработчиков технических объектов в условиях сетевой системы корпоративного обучения : учеб. пособие / В.А. Лебедев, Н.П. Игнатьев, О.А. Захарова. — Москва : ИНФРА-М, 2019. — 161 с.

21. Лыгина, Н. И. Моделирование : учебное пособие / Н. И. Лыгина, О. В. Лауферман. - Новосибирск : Изд-во НГТУ, 2020. - 87 с.

22. Люгер Д.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем. – М.: Издательский дом «Вильямс», 2003 – 864 с.

23. Малла, С. Вэйвлеты в обработке сигналов [Текст]/С. Малла. – М.: Мир, 2005. – 671 с.

24. Милорадова, Н. Г. Поведение людей в организации : учебное пособие / Н. Г. Милорадова ; М-во образования и науки Рос. Федерации, Моск. гос. строит. ун-т. - 2-е изд. (эл.). - Электрон. текстовые дан. (1 файл pdf : 169 с.). - Москва : Издательство МИСИ—МГСУ, 2017. — Систем. требования: Adobe Reader XI либо Adobe Digital Editions 4.5 ; экран 10".- ISBN 978-5-7264-1749-3. - Текст : электронный. - URL: https://znanium.ru/catalog/product/971767 (дата обращения: 27.02.2025).

25. Моделирование систем : практикум / сост. Р. В. Кузьменко, Н. А. Андреева, Е. В. Корчагина [и др.]. - Иваново : ПресСто, 2022. - 96 с.

26. Наумов, В. Н. Поведение потребителей : учебник / В.Н. Наумов. — 2-е изд., перераб. и доп. — Москва : ИНФРА-М, 2023. — 345 с.

27. Нивен П.Р. Сбалансированная система показателей: шаг за шагом. Максимальное повышение эффективности и закрепление полученных результатов. – М.: Баланс–Клуб, 2004. – 328 с.

28. Никитаева, А. Ю. Корпоративные информационные системы: Учебное пособие / Никитаева А.Ю. - Таганрог:Южный федеральный университет, 2017. - 149 с.

29. Ногин В.Д. Парето–оптимальные решения многокритериальных задач. – М.: ФИЗМАТЛИТ, 2007, – 256 с. ISBN 978–5–9221–0821–6.

30. Ногин В.Д. Сужение множества Парето: аксиоматический подход. – М.: ФИЗМАТЛИТ, 2016, – 272 с. ISBN 978–5–9221–1638–1.

31. Павлов, В. А. Инструменты и методы поточно-финансового моделирования деятельности предприятий : учебное пособие / В. А. Павлов. - Москва : Изд-во МГТУ им. Баумана, 2016. - 44 с.

32. Петровский А.Б. Пространства множеств и мультимножеств. – М.: УРСС, 2003. – 248 с. http: // www.raai.org/about/persons/petrovsky/pages/Petrovsky 2003.pdf

33. Пилипенко, А. М. Методы математического и компьютерного моделирования элементов и устройств инфокоммуникационных систем : учебное пособие / А. М. Пилипенко ; Южный федеральный университет. - Ростов-на-Дону : Издательство Южного федерального университета, 2023. - 130 с.

34. Пилипчук, А. О. Модели моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных / А. О. Пилипчук // Вестник науки. – 2024. – Т. 4, № 10(79). – С. 794-800. – EDN KFCWTC

35. Пилипчук, А. О. Процесс моделирования тренда и прогноза на основе интеллектуализации обработки статистических данных / А. О. Пилипчук // Вестник науки. – 2024. – Т. 5, № 12(81). – С. 345-350. – EDN KFCWTC

36. Поведение потребителей : учебник / под общ. ред. О. Н. Романенковой. - Москва : Вузовский учебник : ИНФРА-М, 2022. - 320 с.

37. Поворозник, И. Корпоративные закупки. Как построить эффективную систему закупок в компании : практическое руководство / И. Поворозник. - Москва : Альпина Пабл., 2023. - 345 с. –

38. Пойа Дж. Математика и правдоподобные рассуждения // Дж. Пойа. – Москва : Наука, 1975. – 464 с.

39. Пятецкий, В. Е. Моделирование и регламентация бизнес-процессов с использованием Business Studio 4 : практикум / В. Е. Пятецкий, Л. Н. Калошина, М. А. Поддубный. - Москва : Изд. Дом НИТУ «МИСиС», 2017. - 77 с.

40. Сбоева, И. А. Поведение потребителей : учебное пособие / И. А. Сбоева. - Йошкар-Ола : Поволжский государственный технологический университет, 2017. - 128 с.

41. Сигал, А. В. Моделирование экономики : учебное пособие / А.В. Сигал. — Москва : ИНФРА-М, 2023. — 283 с.

42. Шаньгин, В. Ф. Комплексная защита информации в корпоративных системах : учебное пособие / В.Ф. Шаньгин. — Москва : ФОРУМ : ИНФРА-М, 2022. — 592 с.

43. Юрчук, С. Ю. Основы математического моделирования : учебное пособие / С. Ю. Юрчук. - Москва : Изд. Дом МИСиС, 2014. - 108 с.

44. Якубов, С. Х. Моделирование и управление объектами с распределенными параметрами при неполной информации о векторе переменных состояния : монография / С.Х. Якубов, Р.К. Пирова. — Москва : ИНФРА-М, 2022. — 128 с.

45. Addison, P.S. Illustrated wavelet transform handbook. Introductory Theory and Applications in Science, Engineering, Medicine and Finance / Paul S. Addison. – Bristol: Institute of Physics Publishing, 2022. – 400 p.

46. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). Wiley.

47. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer , 2019. – 386 p.

48. Kuhn, M., & Johnson, K. Applied predictive modeling. Springer , 2013. – 600 p.

49. Siegel, E.. Predictive analytics: The power to predict who will click, buy, lie, or die (Revised and Updated Edition). Wiley, 2019. – 400 p.

50. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann , 2020. – 470 p.