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)

Управление корпоративными информационными процессами
(направленность (профиль))

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

на тему «Модели и алгоритмы аналитической обработки событий системы контроля и управления доступом»

Обучающийся Р.К. Бушмелев
(Инициалы Фамилия) (личная подпись)

Научный руководитель к.т.н., доцент, О.В. Аникина
(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Тольятти 2024

Оглавление

Введение

Обеспечение контроля и управления доступом на территорию вуза считается очень важной задачей, так как значительно ухудшилась криминальная обстановка в стране и возросла угроза террористических атак.

Целью контроля доступа является предоставление лицам, предъявившим учетные данные, подтверждения того, что им разрешено войти в здание.

Этими учетными данными могут быть ключ-карта, ПИН-код, пропуск на мобильный телефон или даже отпечаток пальца или сканирование глаз.

Контроль доступа можно использовать на входе и выходе из здания, на парковках, в серверных помещениях, на складах или в любых других помещениях, которые необходимо защитить от потенциального вторжения и кражи. Установка контроля доступа обеспечивает безопасную среду для сотрудников, посетителей и клиентов организации [19].

Для осуществления вышеуказанных целей существует система специального контроля и управления доступом – СКУД.

«СКУД – это объединение технических, информационных и программных средств контроля для их совместного действия с целью получения качественного управления доступом.

Внедрение СКУД позволит оградить вуз от несанкционированных посещений, контролировать рабочее время преподавателей, следить за посещаемостью студентов, а также применять универсальный идентификатор для доступа в различные помещения вуза» [14].

Следует отметить, что одной из основных функций современной СКУД является обеспечение анализа зарегистрированных событий, ключевой стадией которого является аналитическая обработка событийной информации, собранной за определенный период [39].

На основе результатов анализа событий менеджмент вуза может принять управленческие решения по обеспечению соблюдения трудового порядка

сотрудниками вуза. Эффективность принятых решений зависит от качества полученных результатов анализа, которое должно обеспечиваться на стадии аналитической обработки событий СКУД.

Для решения данной задачи необходимо использовать в процессе аналитической обработки событий СКУД эффективные модели и алгоритмы.

Таким образом, актуальность темы исследования обусловлена необходимостью исследования и разработки эффективных моделей и алгоритмов аналитической обработки событий СКУД.

Объектом настоящего исследования является СКУД.

Предметом исследования является аналитическая обработка событий СКУД.

Целью работы является исследование и разработка моделей алгоритмов аналитической обработки событий СКУД, обеспечивающих высокую эффективность принятых управленческих решений.

Для достижения поставленной цели необходимо решать следующие задачи:

- проанализировать современное состояние проблемы исследования;
- проанализировать и выбрать методы и технологии аналитической обработки событий СКУД;
- разработать модели и алгоритмы аналитической обработки событий СКУД, обеспечивающие высокую эффективность принятых управленческих решений;
- выполнить апробацию проектных решений и оценить их эффективность.

Гипотеза исследования: применение предлагаемых в диссертационном исследовании моделей и алгоритмов аналитической обработки событий СКУД позволит повысить эффективность принятых управленческих решений.

Теоретической основой диссертационного исследования являются научные труды российских и зарубежных ученых, занимающихся проблемами аналитической обработки событий СКУД.

Методы исследования. В процессе исследования будут использованы следующие положения и методы: системный анализ, методы и технологии аналитической обработки событий СКУД, методы и технологии проектирования информационных систем, методы машинного обучения.

Новизна исследования заключается в разработке моделей и алгоритмов аналитической обработки событий СКУД, обеспечивающих высокую эффективность принятых управленческих решений.

Практическая значимость исследования заключается в возможности применения предлагаемых моделей и алгоритмов при проектировании систем аналитической обработки событий СКУД, обеспечивающих высокую эффективность принятых управленческих решений.

Основные этапы исследования: исследование проводилось с 2022 по 2024 год в несколько этапов.

На первом (констатирующем) этапе формулировалась тема исследования, выполнялся сбор информации по теме исследования из различных источников, проводилась формулировка гипотезы, определялись постановка цели, задач, предмета исследования, объекта исследования и выполнялось определение проблематики данного исследования.

Второй этап – аналитический. В ходе проведения данного этапа осуществлялся анализ методов и технологий аналитической обработки событий СКУД, опубликована статья по теме исследования в научном сборнике.

На третьем этапе осуществлялась апробация предлагаемых проектных решений, произведена оценка их эффективности, сформулированы выводы о полученных результатах по проведенному исследованию.

На защиту выносятся:

- модели и алгоритмы аналитической обработки событий СКУД, обеспечивающие повышение эффективности принятых управленческих решений;
- результаты апробации и оценки эффективности предлагаемых

проектных решений.

По теме исследования опубликована 1 статья:

Бушмелев Р.К. Модель подсистемы аналитической обработки событий в СКУД // Вестник научных конференций. 2023. №12-1 (100). С. 21-22.

Диссертация состоит из введения, четырех глав, заключения и списка литературы.

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

В первой главе дан анализ современного состояния исследований в области проектирования систем аналитической обработки событий СКУД.

Во второй главе дан анализ методов и технологий аналитической обработки событий СКУД.

Третья глава посвящена разработке моделей и алгоритмов аналитической обработки событий СКУД, обеспечивающих повышение эффективности принятых управленческих решений

В четвертой главе выполнены апробация предлагаемых проектных решений и оценка их эффективности.

В заключении приводятся результаты исследования.

Работа изложена на 71 странице и включает 40 рисунков, 6 таблиц и 40 источников.

Глава 1 Анализ современного состояния исследований в области построения систем аналитической обработки событий СКУД

1.1 Обзор и анализ литературы и источников по теме исследования

В статье [14] представлена схема интеграции СКУД с подсистемами управления деятельностью вуза (рисунок 1).

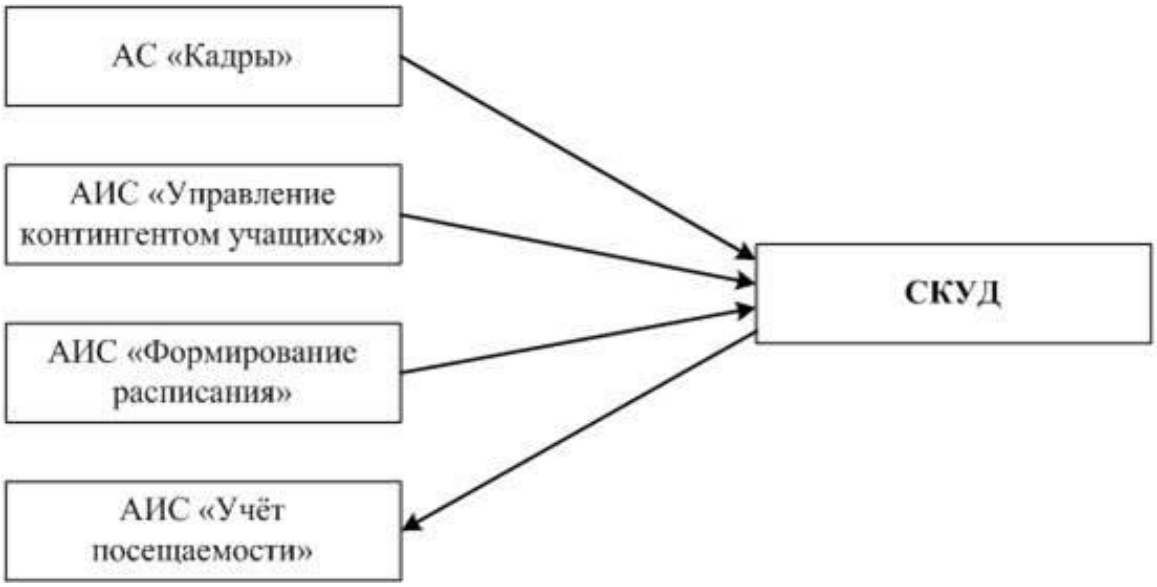

Рисунок 1 – Схема интеграции СКУД с подсистемами управления деятельностью вуза

В данном решении аналитическая обработка событий производится средствами каждой подсистемы с помощью механизмов обмена данными со СКУД.

По мнению авторов исследования, «имеющиеся на российском рынке СКУД, не следует использовать в вузах без соответствующей доработки».

В исследовании [30] представлена архитектура типовой системы обработки событий средств контроля доступа (рисунок 2).

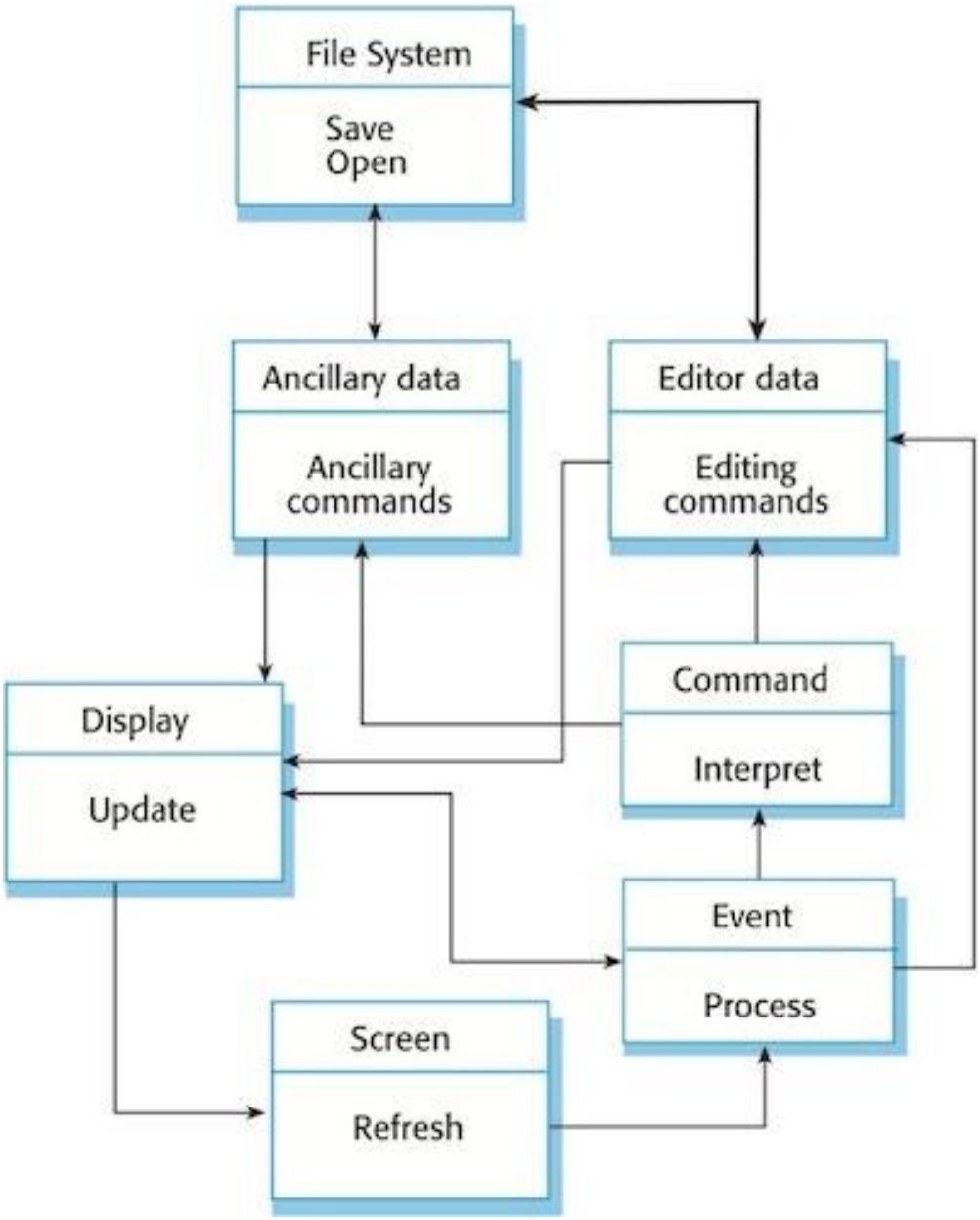

Рисунок 2 – Архитектура типовой системы обработки событий средств контроля доступа

Необходимо отметить, что для использования представленной архитектура для СКУД требуется предварительный анализ архитектурных и функциональных особенностей последней.

В работе [3] представлен пример решения задач, связанных с контролем прохода и регистрацией событий СКУД.

Необходимая функция СКУД – протоколирование событий входа и выхода, а также действий персонала охраны проходных по ручному

управлению точками прохода. «В совокупности с информацией систем наблюдения это позволяет обнаружить, например, несанкционированный выезд транспорта через КПП (что обычно также связано с хищениями). Кроме того, протокол событий позволяет контролировать перемещение сотрудников по территории объекта (рисунок 3)» [3].

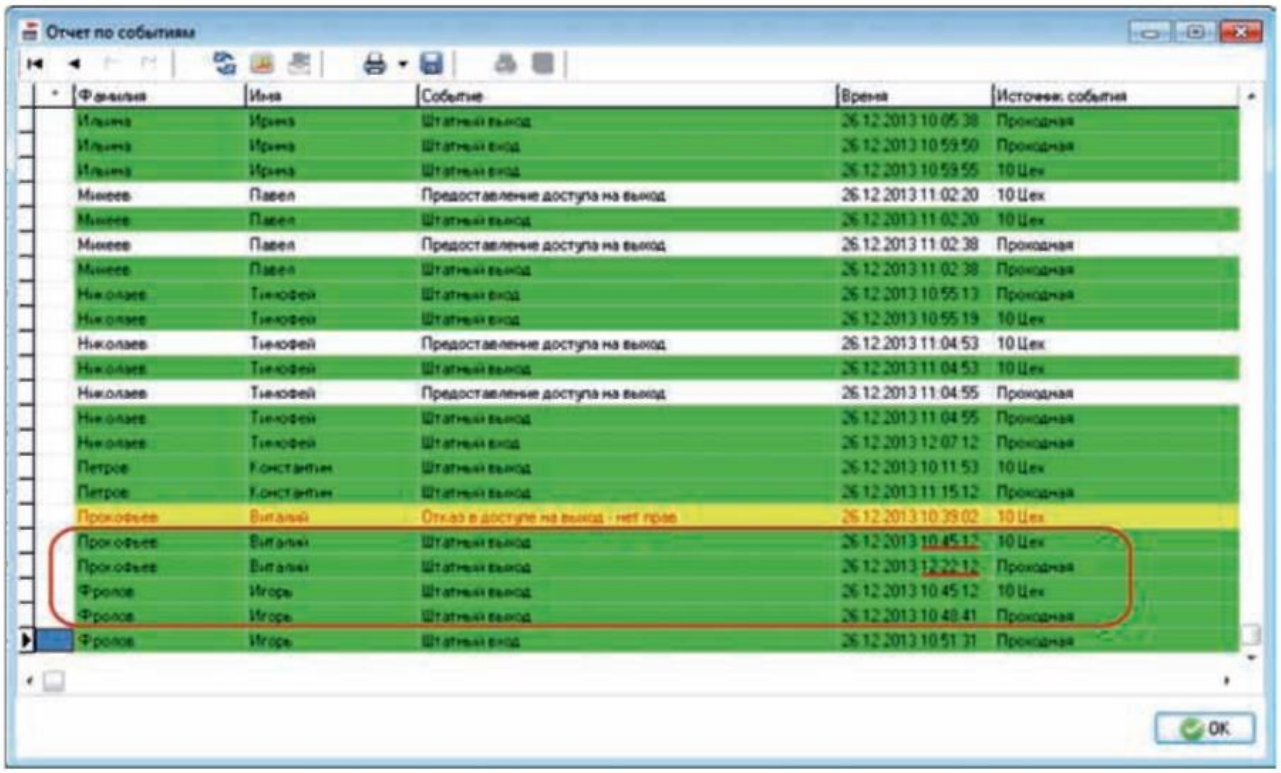

Рисунок 3 – Окно отчета по событиям

Следует отметить, что в статье отсутствует описание моделей и алгоритмов, используемых в аналитической системе.

В статье [21] отмечено, что компьютерные системы диспетчерского управления и сбора данных (SCADA) за последние четыре десятилетия превратились из автономных, разделенных операций в сетевые архитектуры, которые обмениваются данными на больших расстояниях.

Наметилась тенденция объединения SCADA и обычных ИТ-подразделений в сторону объединения некоторых дублирующих друг друга видов деятельности. Эта тенденция обусловлена снижением затрат за счет

объединения разрозненных платформ, сетей, программного обеспечения и средств обслуживания. «Из соображений эффективности, технического обслуживания, экономичности, сбора данных платформы управления мигрировали из изолированных внутризаводских сетей, использующих проприетарное аппаратное и программное обеспечение, в системы на базе ПК, использующие стандартное программное обеспечение, сетевые протоколы и Интернет. Авторы представляют новый подход к веб-системам SCADA, которые адаптируются к поведению целевого приложения» [21].

Кроме того, учтены ограничения реального времени, налагаемые характером проблемы (рисунок 4).

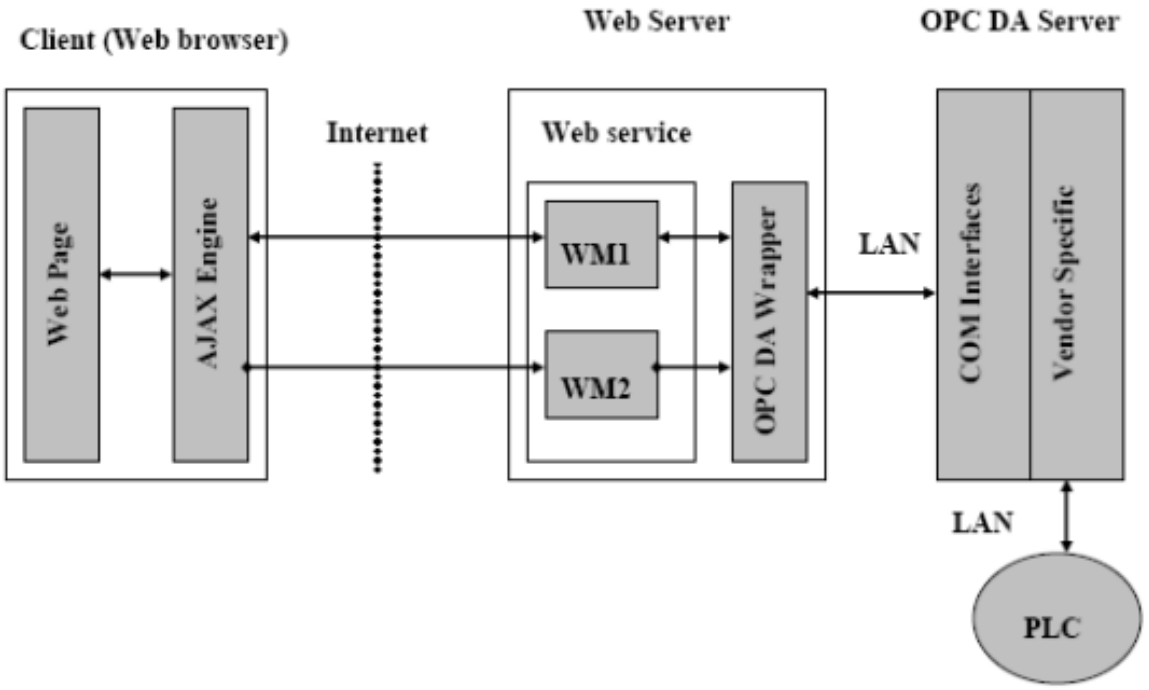

Рисунок 4 – Архитектура системы мониторинга событий СКУД

В работе [32] описывается веб-система, которая позволяет пользователю управлять доступом к образовательным помещениям и осуществлять мониторинг событий в режиме реального времени, используя две сети передачи данных.

Первая сеть основана на CAN-шине для сбора данных и задач управления, к которой подключены управляющие устройства, такие как считыватели магнитных карт, дверные замки и датчики (CAN, Controller Area Network — стандарт промышленной сети, ориентированный, прежде всего, на объединение в единую сеть различных исполнительных устройств и датчиков).

Второй представляет собой инфраструктуру Интернета/Интранета и использует стандартные веб-технологии, такие как PHP и Java, для обеспечения эффективного контроля и мониторинга в режиме реального времени (рисунок 5).

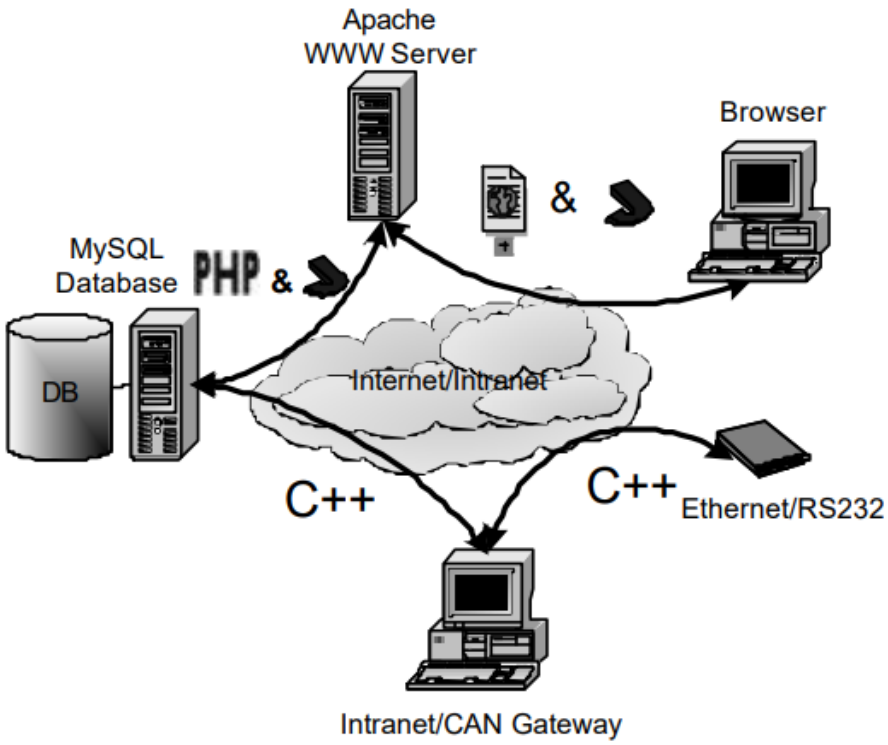

Рисунок 5 – Программная архитектура веб-системы мониторинга событий СКУД

Эта система реализована в операционной системе Linux с использованием HTTP-сервера Apache и использует стандартные технологии, используемые в Интернете, чтобы создать эффективную систему безопасности в масштабах всего кампуса. Чтобы сократить разрыв между

этими двумя сетями, использован шлюз CAN/Intranet.

Обзор и анализ источников по теме исследования подтвердили интерес ученых и специалистов к проблеме анализа и мониторинга событий СКУД.

Вместе с тем следует констатировать недостаточность работ, посвященных разработке моделей и алгоритмом аналитической обработки событий СКУД, что подтверждает актуальность настоящего исследования.

1.2 Обзор и анализ ИТ-решений для аналитической обработки событий СКУД

Как показал анализ, на рынке ИТ-решений для аналитической обработки событий СКУД широко представлены программные продукты, разработанные на платформе «1С: Предприятие 8» [15].

На рисунке 6 представлена типовая архитектура такого программного продукта.

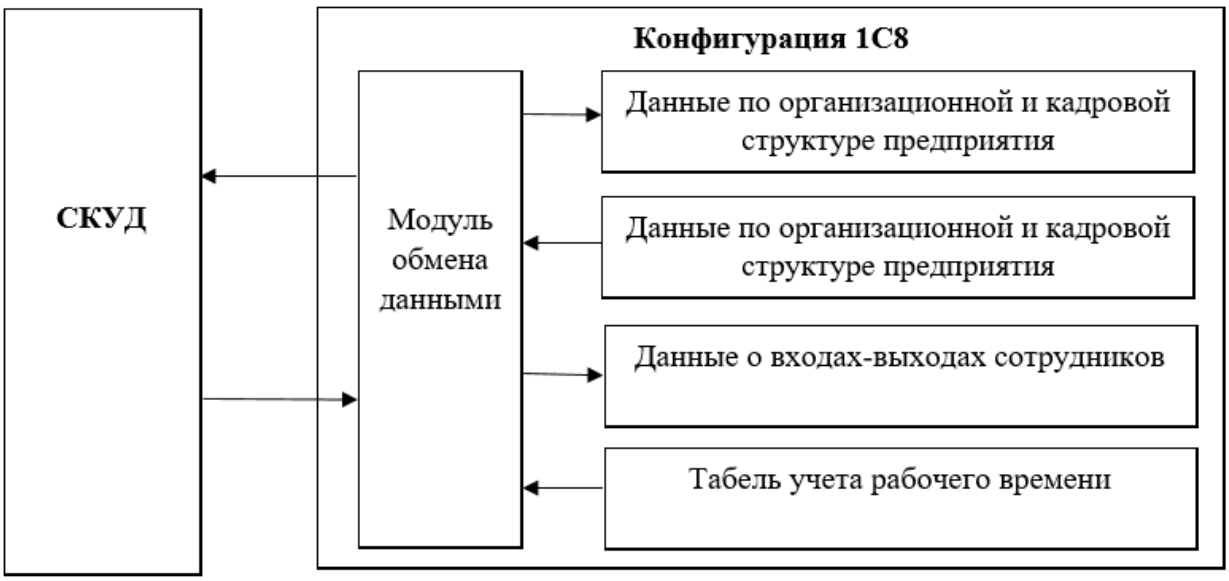

Рисунок 6 – Типовая архитектура ИТ-решения на платформе «1С: Предприятие 8» для аналитической обработки событий СКУД

Н основе такой архитектуры построен программный продукт (ПП)

«Болид: СКУД и УРВ для 1С: Предприятие 8», архитектура которого показан на рисунке 7 [12].

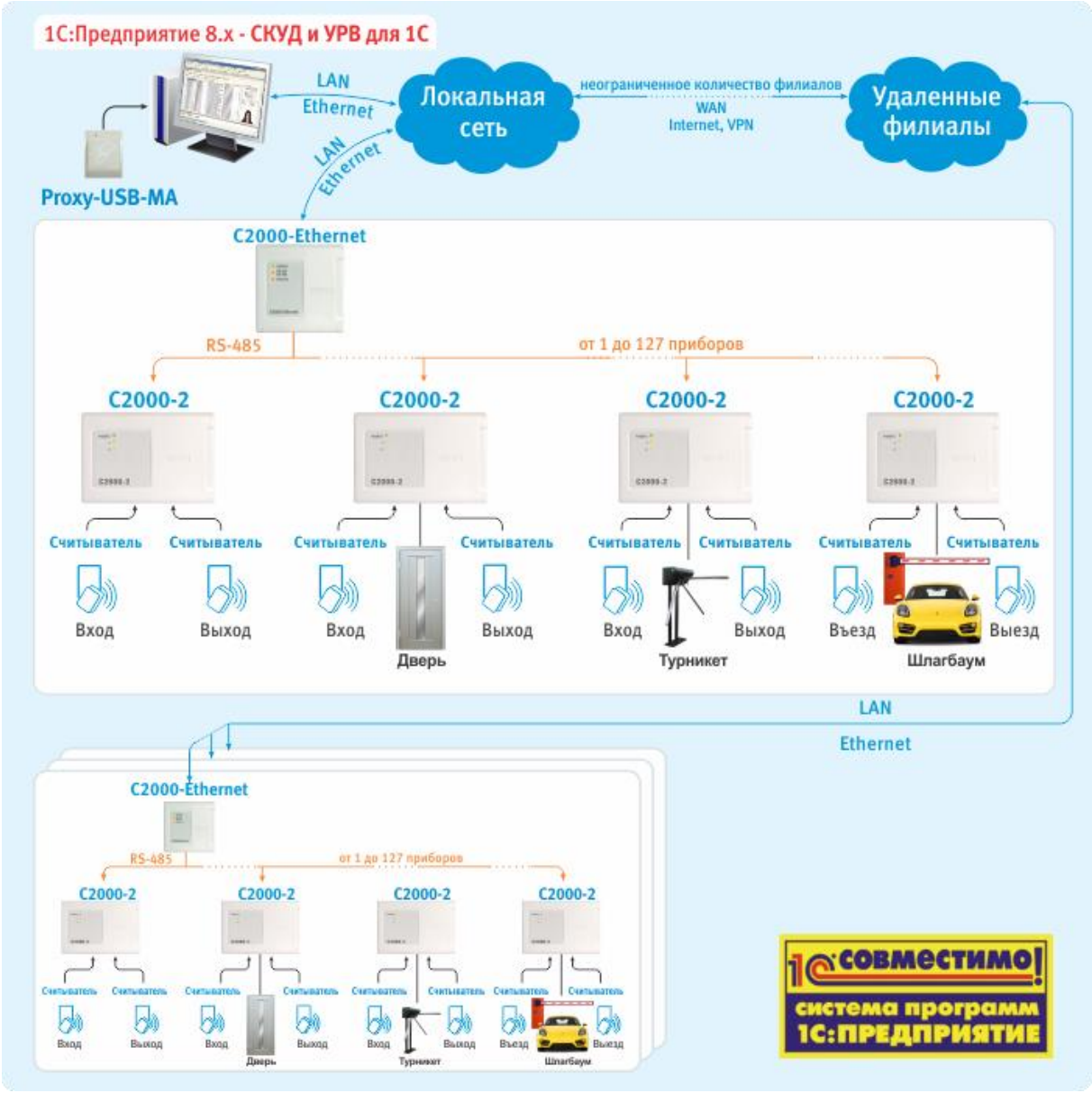

Рисунок 7 – Архитектура ПП «Болид: СКУД и УРВ для 1С: Предприятие 8»

В ИТ-решении «Аналитика по сотрудникам в подсистеме PROSTO: СКУД» Аналитика позволяет специалистам отдела кадров понимать, как работают сотрудники предприятия, а линейным руководителям – как работают сотрудники их отдела.

Форма для выполнения анализа показана на рисунке 8.

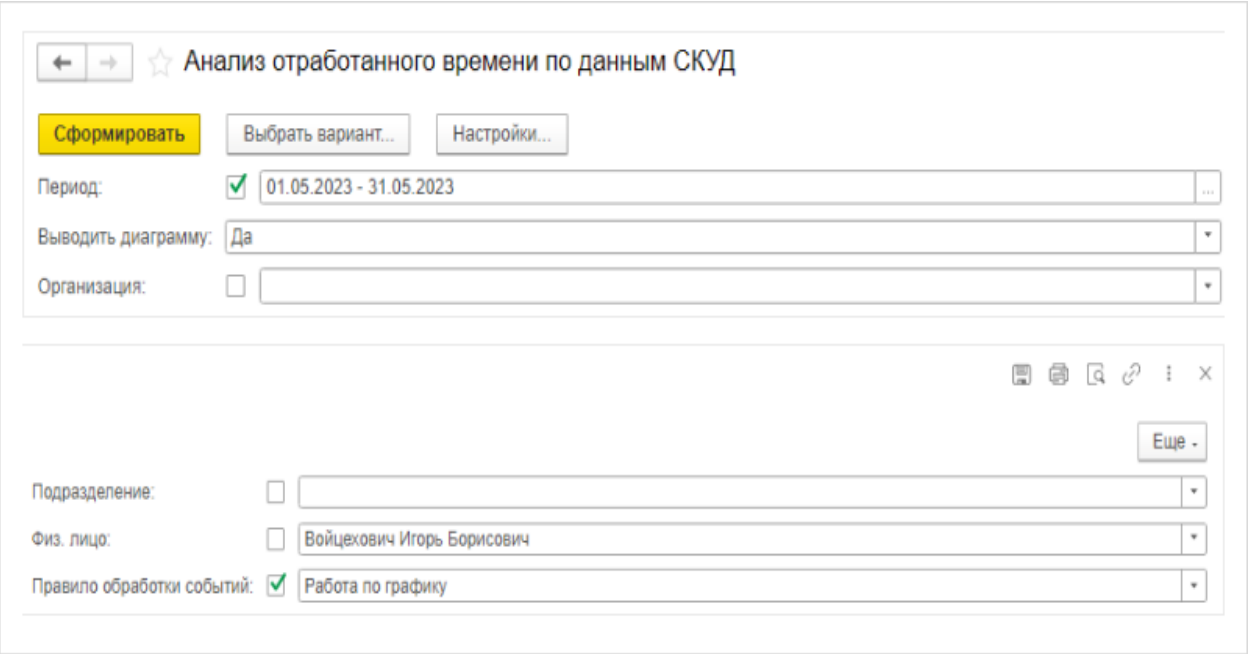

Рисунок 8 – Форма для выполнения анализа

Данная отчётность поможет оперативно назначать корректные графики и контролировать трудовую дисциплину.

«В отчёт включены более 15 типовых отборов, с помощью которых можно отслеживать все важные параметры работы предприятия и эффективности сотрудников.

Например, можно увидеть:

- детали работы по графику;
- отсутствия на рабочем месте;
- прогулы;
- работу во время выходного дня;
- приход позже начала рабочего дня;
- уход до окончания рабочего дня и другие данные.

Форма для создания отчета показана на рисунке 9» [2].

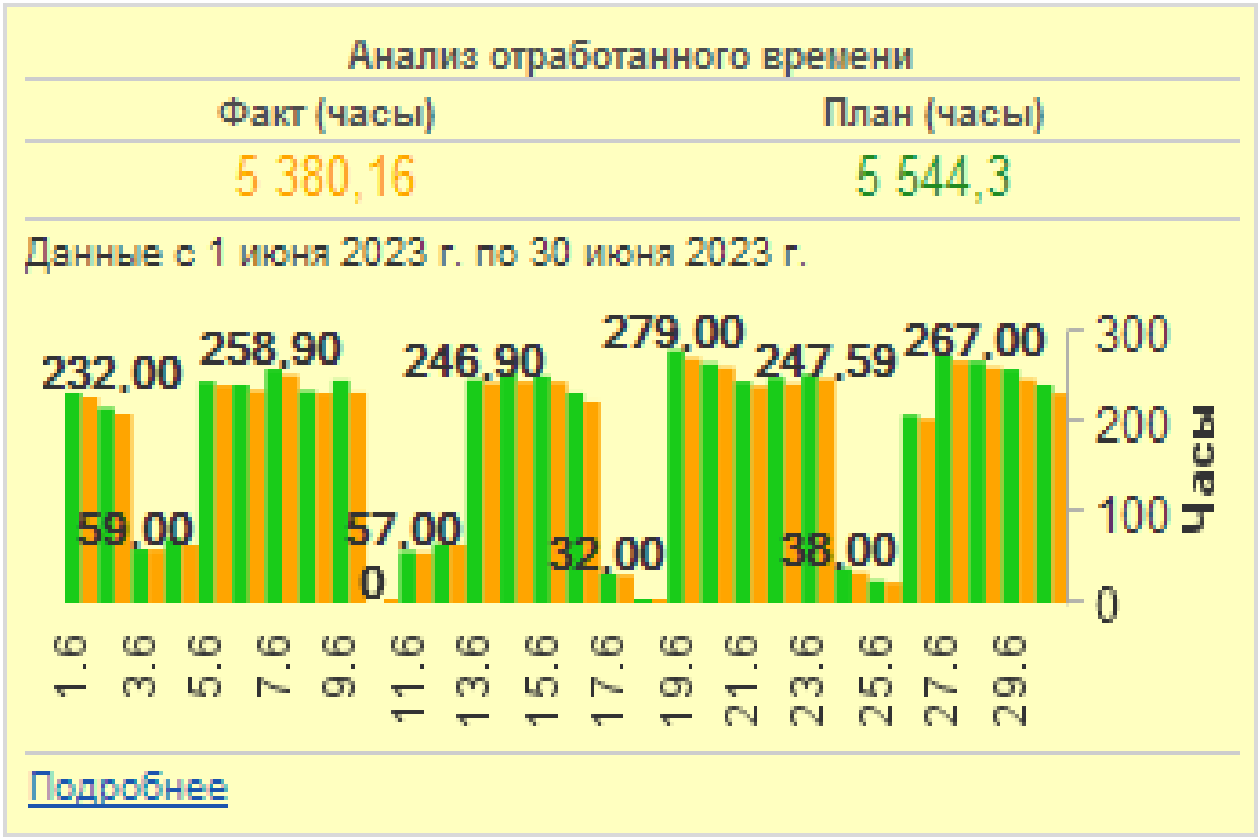

Рисунок 9 – Отчет «Анализ отработанного времени»

В ПП [5] реализован функционал анализа журнала событий СКУД (рисунок 10).

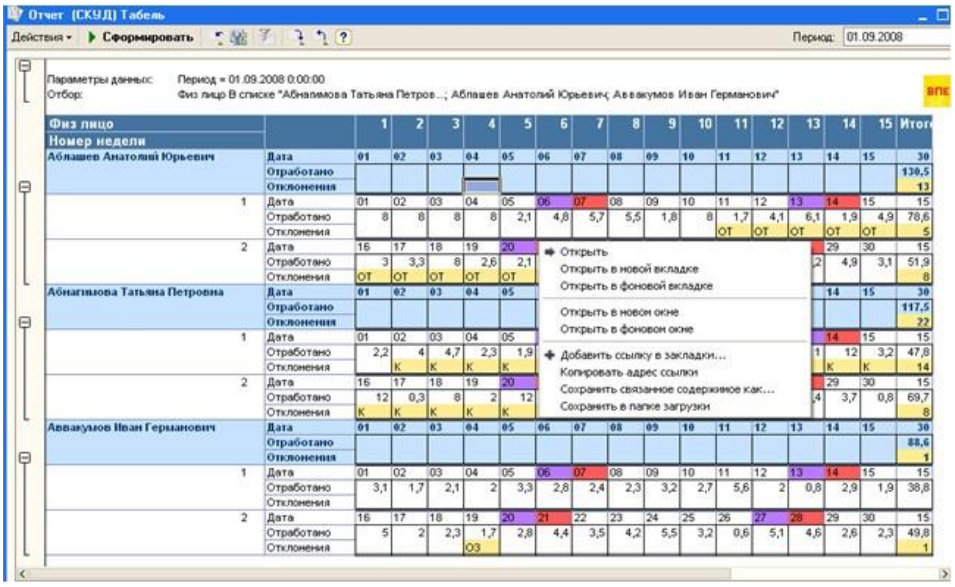

Рисунок 10 – Окно формирования аналитического отчета

Данный функционал обеспечивает:

- получение сводной информации об отработанном времени, опозданиях, перерывах, переработках, прогулах и т.д.;
- вывод информации о посещениях в диаграмму Ганта;
- ввод корректировок журнала событий по оправдательным документам и др.

ParsecNET – это профессиональная СКУД, которая по мнению ее вендора вобрала в себя перспективные наработки и уникальный опыт компании [8].

На рисунке 11 показано окно обозревателя монитора событий.

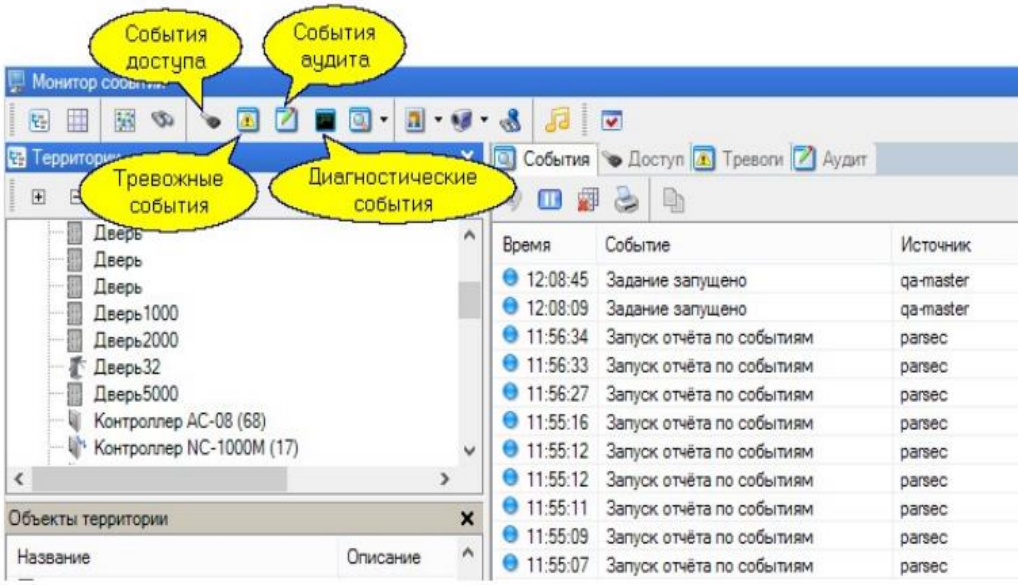

Рисунок 11 – Окно обозревателя монитора событий ParsecNET

Среди ИТ-решений на других платформах можно выделить модуль событий ParsecNET позволяет проводить ретроспективный анализ событий в системе с гибким назначением интервалов времени, типов событий.

Он может использовать шаблоны типовых отчетов, созданные пользователем (оператором).

В модуле имеется функция формирования различных отчетов.

На рисунке 12 представлен пример отчета по событиям.

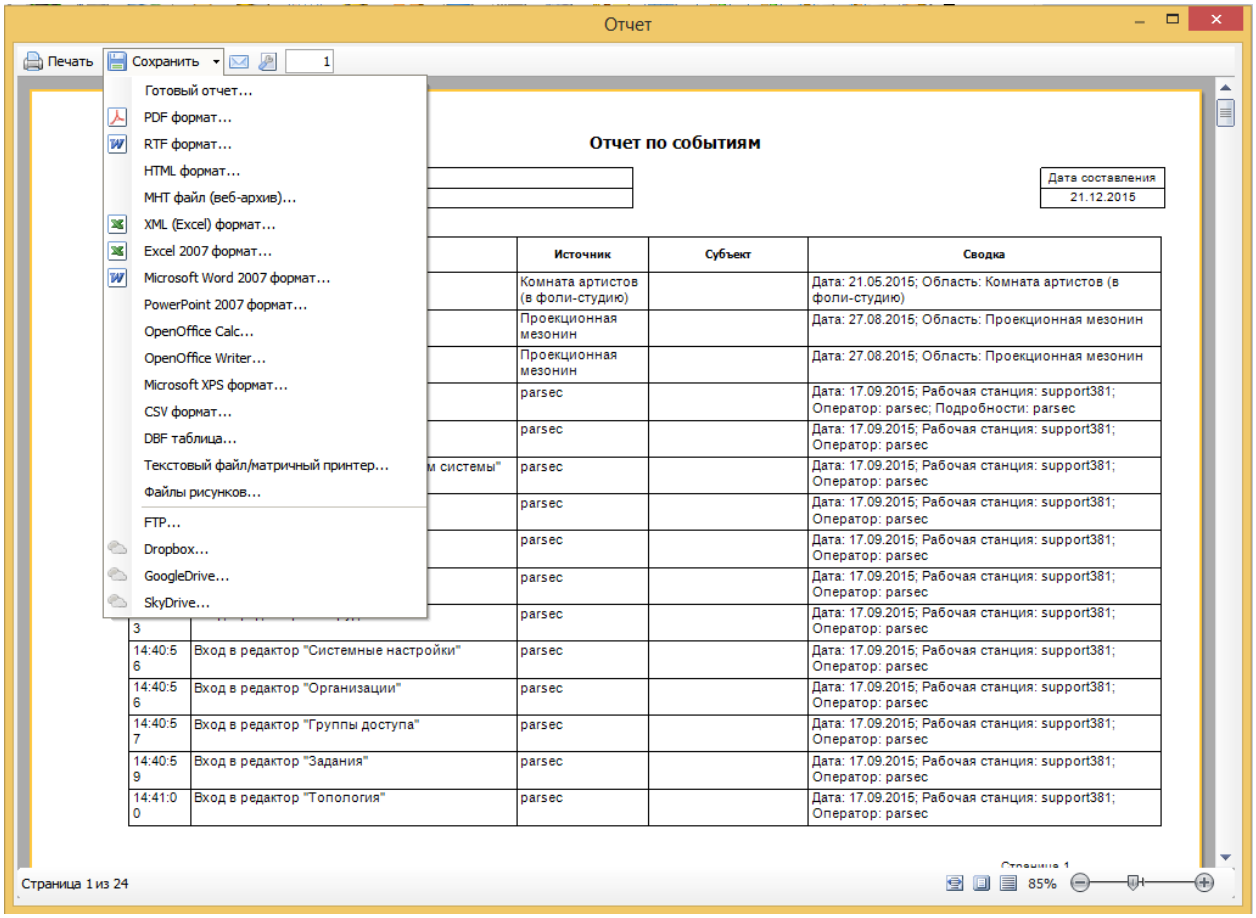

Рисунок 12 – Отчет по событиям ParsecNET

Web-приложение для анализа посещаемости сотрудников по данным СКУД предоставляет широкий функционал различным категориям его пользователей.

«Приложение в табличном виде позволяет посмотреть общее время нахождения в офисе каждого сотрудника за конкретный месяц.

Итоговое значение можно развернуть кликом по нему и получить расшифровку по дням (рисунок 13).

Данные по всем сотрудникам за конкретный месяц можно выгружать в файл Excel» [1].

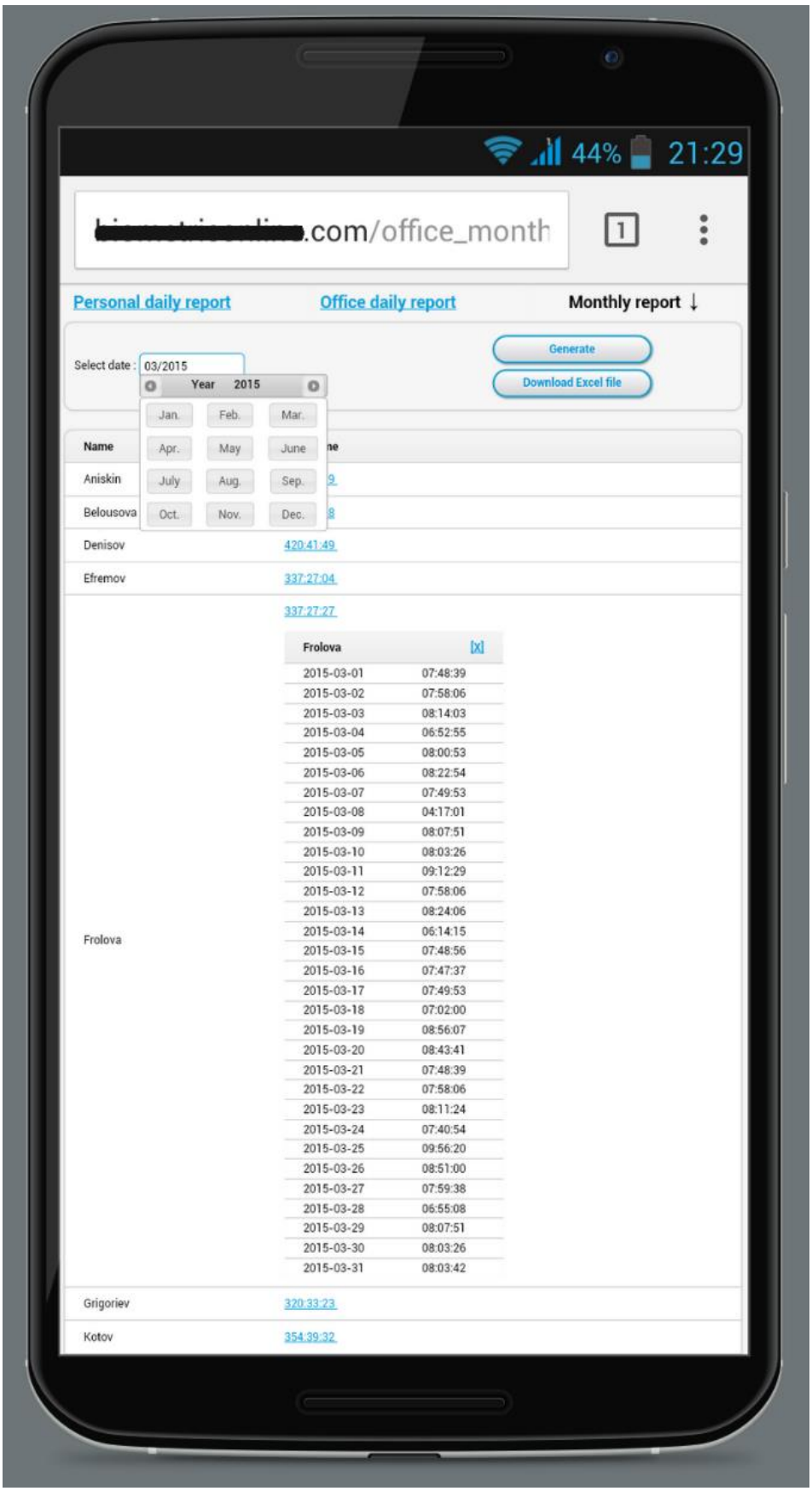

Рисунок 13 – Окно просмотра событий по сотрудникам конкретный месяц

В [10] описан модуль обработки событий ИСБ КОДОС.

«Все события, происходящие в ИСБ КОДОС отображаются на экране оператора и сохраняются в базе данных. Для удобства восприятия, в зависимости от типа события отображаются разным цветом. Представлена возможность отложить обновления списка для более удобного просмотра определенных моментов» [10].

Реализована функция обработки тревожных событий.

В решении имеется возможность проведения и формирования отчета статистического анализа за определенный период, который можно экспортировать в файлы MS Office (рисунок 14).

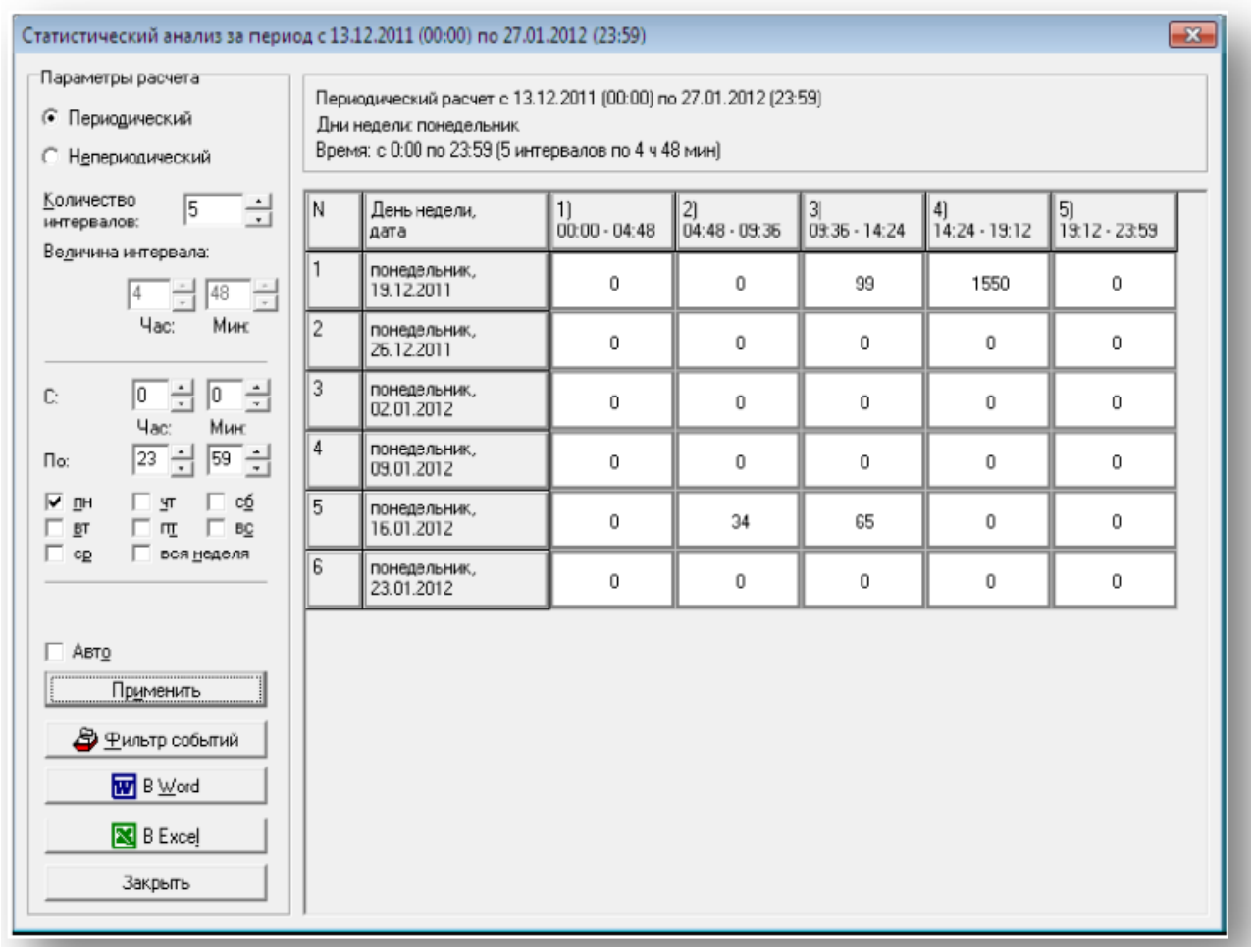

Рисунок 14 – Окно формирования отчета статистического анализа событий ИСБ КОДОС

Как показали обзор и анализ готовых ИТ-решений для аналитической обработки событий СКУД, последние главным образом ориентированы на решение задач контроля рабочего времени сотрудников организаций.

Следует также отметить, что представленные ИТ-решения разработаны под конкретные модели СКУД, что ограничивает их функциональные возможности.

Кроме того, в описаниях к рассмотренным ИТ-решениям отсутствуют сведения о моделях и алгоритмах, положенных в их основу.

Выводы по главе 1

В результате проделанной работы были сделаны следующие выводы:

- обзор и анализ источников по теме исследования подтвердили интерес ученых и специалистов к проблеме анализа и мониторинга событий СКУД;
- как показали обзор и анализ готовых ИТ-решений для аналитической обработки событий СКУД, последние главным образом ориентированы на решение задач контроля рабочего времени сотрудников организаций;
- анализ показал, что представленные ИТ-решения разработаны под конкретные модели СКУД, что ограничивает их функциональные возможности.

Необходимо отметить, что в описаниях к рассмотренным ИТ-решениям отсутствуют сведения о моделях и алгоритмах, положенных в их основу.

Глава 2 Анализ методов и технологий аналитической обработки событий СКУД

2.1 СКУД как объект Интернета вещей

В зарубежных источниках системы контроля и управления доступом СКУД рассматриваются как объекты Интернета вещей (IoT).

Под IoT понимается сеть устройств, соединенных между собой посредством цифровой сети (государственной или частной), которая позволяет взаимодействовать и обмениваться потоками данных между ними.

Это могут быть бытовые приборы, промышленное оборудование, роботизированные средства или любое другое устройство, способное взаимодействовать без вмешательства человека, то есть M2M (машина-к-машине) [18].

Системы обработки событий реагируют на события в среде системы или пользовательском интерфейсе.

Ключевой характеристикой систем обработки событий является то, что время событий непредсказуемо, и система должна быть в состоянии справиться с этими событиями, когда они происходят.

Системы контроля доступа (ACS) через IoT – это системы, устройства которых подключены к сети, что экспоненциально увеличивает количество их применений.

Когда в компании используется система контроля доступа, целью является не только обеспечение безопасности, но и облегчение организационного управления с помощью ряда параметров.

Преимущество этих усовершенствованных систем заключается в том, что они могут содержать решения для всей компании в целом, предоставляя расширенные функциональные возможности на основе систем контроля пользователей и идентификации, такие как контроль рабочего времени, контроль доступа посетителей и зон ограниченного доступа в режиме

реального времени и даже планы эвакуации и действий в чрезвычайных ситуациях.

Действительно интересная особенность использования технологии IoT для контроля доступа заключается в том, что она позволяет осуществлять мониторинг и управление в режиме реального времени через централизованную платформу управления.

Более того, ее эффективность возрастает в сочетании с технологией больших данных.

При использовании технологий искусственного интеллекта (ИИ), СКУД становится передовой технологией, которая открывает бесконечные возможности для организационной среды.

Интеллектуальные системы контроля доступа, взаимосвязанные через IoT, предлагают ряд преимуществ и функций, которые заставляют организации все чаще делать выбор в пользу их внедрения.

Современные системы контроля доступа обеспечивают высокую безопасность. Тот факт, что они используют технологию IoT, позволяет им подключаться и обмениваться данными с централизованной платформой, которая обрабатывает информацию за миллисекунды и предоставляет разрешения на доступ.

Кроме того, мы можем пойти еще на один шаг дальше.

Эта система применима к использованию определенных активов внутри организации, где пользователь имеет уникальный идентификатор с кодировкой разрешений на использование компьютеров и других “активных” элементов для своего профиля.

Рассмотрим пример СКУД, построенной на основе событийной платформы.

Эта платформа в основном представлена в модуле обработки событий, в котором политика используется для контроля и управления потоком обработчика событий для вычисления данных.

Поскольку эти ресурсы ограничены, очень важно предотвратить истощение ресурсов во время процесса вычислений. Управление доступом состоит из контроллера на основе событий контроля доступа и модуля политики, которые применяются в процессе обработки событий в модуле обработки, как показано на рисунке 15 [34].

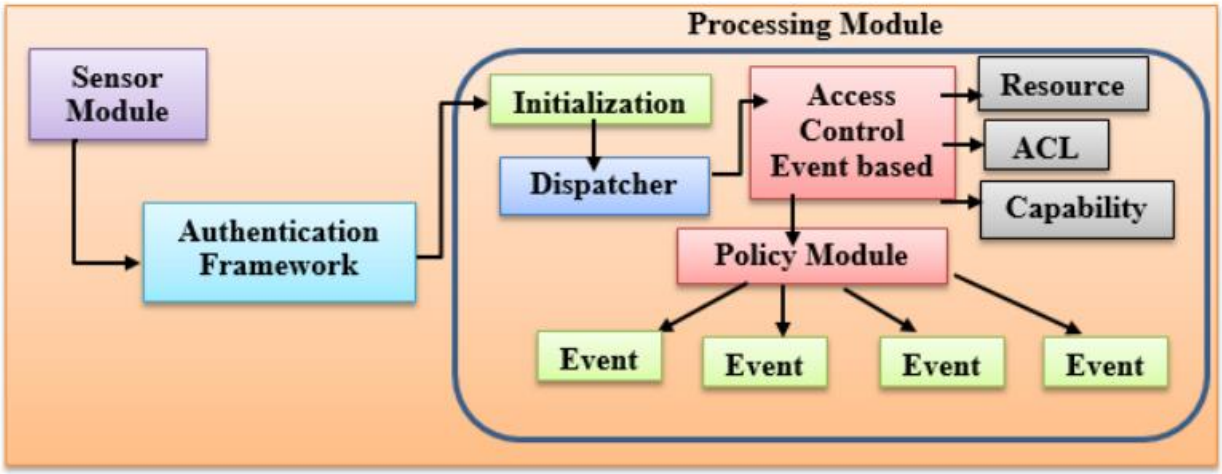

Рисунок 15 – Архитектура СКУД на основе событийной платформы

Предположим, что процесс планирования не применял упреждающее действие и должен выполняться до завершения. Если процесс необходимо прервать, он должен запросить опрос процесса. Опрос — это сигнал для выполнения процесса, связанный с обработчиком прерываний.

Состояние гонки с планировщиком ядра переднего плана произойдет, если будет какая-либо попытка отправить контекст прерывания.

Как показано на рисунке 13, событие инициируется либо ядром, либо процессом после проверки get в процессе аутентификации. Каждый процессор имеет структуру данных очереди событий, в которой хранятся необработанные события.

После подтверждения аутентификации процессор начинает инициализацию функции main(), ожидая отправки данных.

Следует отметить, что из-за огромного масштаба политик и количества объектов контроля доступа в открытых распределенных информационных системах, таких как большие данные, Интернет вещей и облачные вычисления, существующие методы анализа событий в СКУД являются неэффективными.

Чтобы преодолеть вышеупомянутую проблему, предлагается эффективная подсистема анализа событий СКУД, основанная искусственном интеллекте (ИИ) и машинном обучении (МО).

2.2 Методы аналитической обработки данных на основе технологии машинного обучения

Система управления информацией о безопасности и событиями — это отраслевой термин в области безопасности, обозначающий сбор данных, обычно файлов журналов или журналов событий, из различных источников в центральное хранилище для анализа.

Журналы событий генерируются различными сетевыми устройствами, операционными системами и серверами приложений и СКУД.

Журналы событий предоставляют необработанные данные обо всей активности, происходящей в ИТ-инфраструктуре любой организации. Эти необработанные данные действуют как входные данные для системы анализа событий СКУД, которая предоставляет предупреждения безопасности и отчеты в качестве выходных данных.

Обработка всех необработанных данных осуществляется с использованием технологии интеллектуального анализа данных (Data mining) [40].

Интеллектуальный анализ данных, синоним «обнаружения знаний в базах данных», представляет собой процесс анализа данных с разных точек зрения и их обобщения в полезную информацию. Это процесс, который

позволяет пользователям понять суть взаимосвязей между данными (рисунок 16) [29].

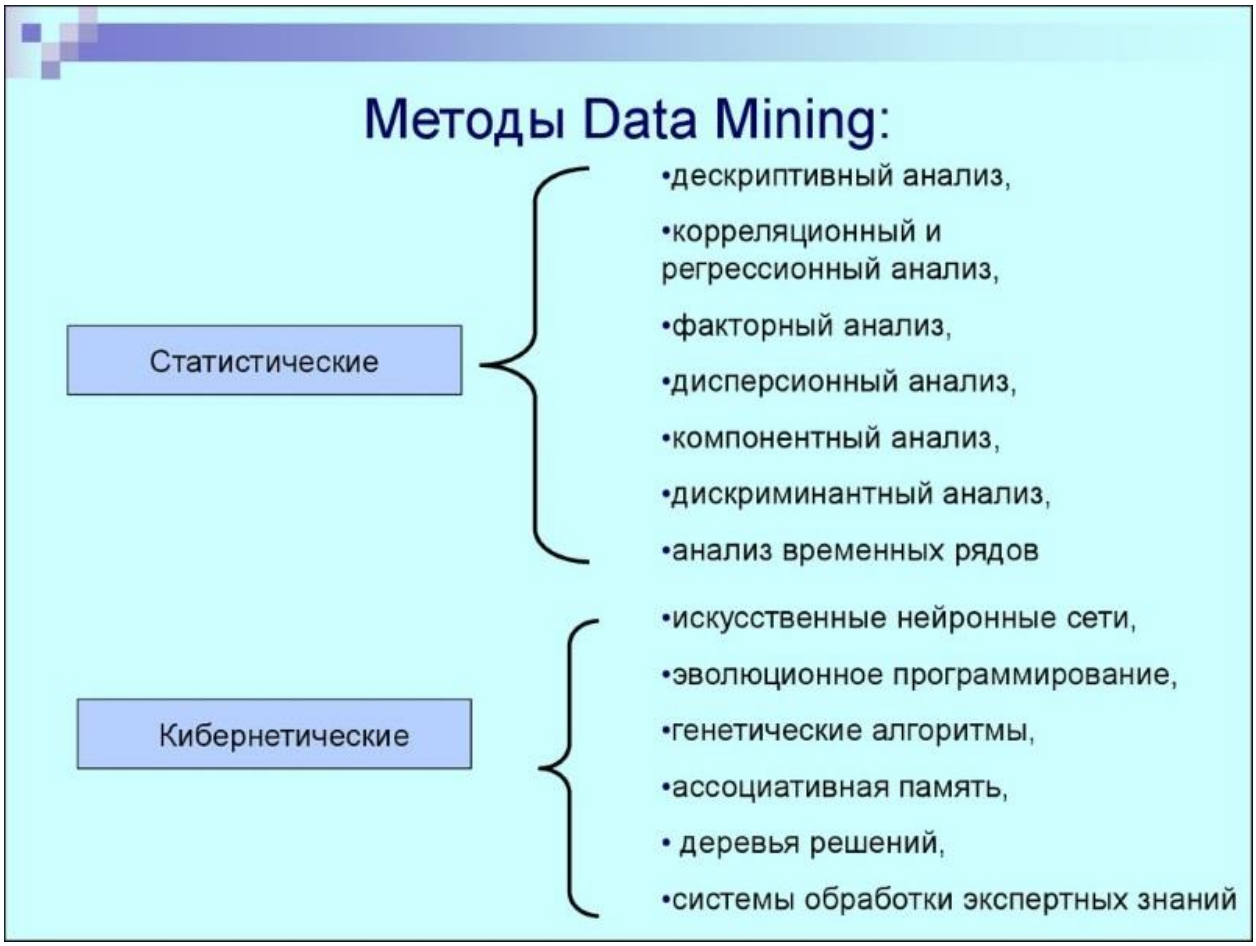

Рисунок 16 – Методы интеллектуального анализа данных

«Data mining выявляет закономерности и тенденции, скрытые среди данных. Его часто рассматривают как процесс извлечения достоверной, ранее неизвестной, нетривиальной и полезной информации из больших баз данных» [40].

Интеллектуальный анализ данных становится все более распространенным как в частном, так и в государственном секторах. Если интеллектуальный анализ данных применяется ко всем журналам событий, генерируемым различными сетевыми устройствами, системами и серверами

приложений. тогда эффективность корпоративной безопасности может быть значительно повышена.

Преимущество использования интеллектуального анализа данных заключается в его способности анализировать огромный набор данных.

Интеллектуальный анализ данных состоит из следующих основных этапов [27]:

– изучение данных для раскрытия тем и тенденций. Этот этап может включать довольно сложный анализ с использованием широкого спектра статистических методов.
– создание модели для объяснения данных и выявления закономерностей с помощью анализа. На этом этапе рассматриваются несколько моделей.
– применение моделей к новым данным для прогнозирования.

Развернутая схема процесса интеллектуального анализа данных показана на рисунке 17 [17].

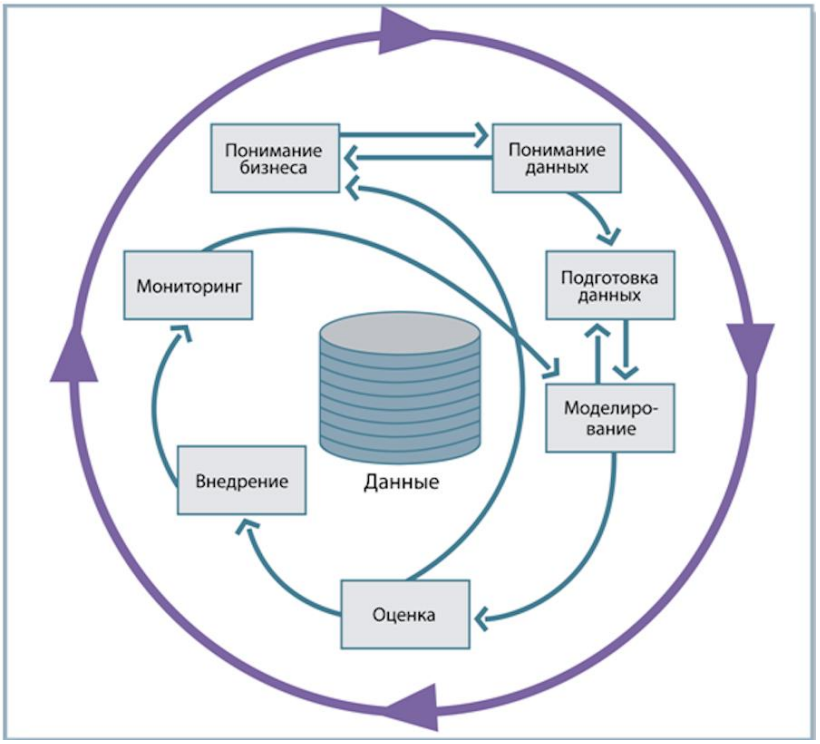

Рисунок 17 – Развернутая схема процесса интеллектуального анализа данных

Предварительная обработка — это важная работа по подготовке данных для интеллектуального анализа данных.

Подготовка данных имеет решающее значение в любом проекте машинного обучения, поскольку она напрямую влияет на производительность и точность модели обучения.

Подготовка данных для интеллектуального анализа данных — это процесс очистки, преобразования и организации необработанных данных в формат, понятный алгоритмам машинного обучения.

Процесс подготовки данных состоит из следующих этапов [28]:

- сбор данных из различных источников, таких как базы данных, электронные таблицы или API СКУД;
- очистка данных, которая заключается в удалении или исправлении пропущенных значений, выбросов или несоответствий;
- преобразуете данные с помощью таких процессов, как нормализация и кодирование, чтобы сделать их совместимыми с алгоритмами машинного обучения;

Наконец, уменьшение сложности данных, не теряя при этом информации, которую они могут предоставить модели машинного обучения, часто используя такие методы, как уменьшение размерности.

Следует учесть, что подготовка данных — это непрерывный процесс, а не разовая задача. По мере развития вашей модели или получения новых данных необходимо пересматривать и уточнять этапы подготовки данных.

Использование интеллектуального анализа данных в качестве инструмента анализа, применяемого к базам данных событий СКУД, может оказать огромное положительное влияние на деятельность организации и ее сотрудников.

Преимущества проведения тщательного анализа баз данных событий СКУД включают лучшее понимание показателей безопасности, лучшее понимание того, как сосредоточить усилия на уменьшении событий, а также

лучшее понимание того, как эти события влияют на деятельность организации и ее сотрудников.

Технология интеллектуального анализа данных полезна для поиска закономерностей в большом наборе данных с использованием различных междисциплинарных методов, таких как машинное обучение (МО), статистика и искусственный интеллект.

Машинное обучение — это наука о разработке алгоритмов, которые самостоятельно обучаются на основе данных и адаптируются без вмешательства человека.

Жизненный цикл машинного обучения показан на рисунке 18.

Рисунок 18 – Жизненный цикл машинного обучения

Когда мы передаем данные в эти алгоритмы, они выстраивают свою собственную логику и, в результате, создают решения, актуальные для таких

разнообразных аспектов нашего мира, как обнаружение мошенничества, веб-поиск, классификация болезней и прогнозирование цен.

В глубоком обучении, подвиде машинного обучения, программы открывают сложные концепции, создавая их из более простых. Эти алгоритмы работают, подвергая многослойные (следовательно, «глубокие») нейронные сети воздействию огромных объемов данных.

Приложения для машинного обучения, такие как обработка естественного языка, значительно повышают производительность за счет использования глубокого обучения [25].

Вот некоторые сценарии, в которых машинное обучение может помочь в решении проблем интеллектуального анализа данных:

- качество результатов инструментов интеллектуального анализа данных зависит от качества данных. Иногда это может даже не решить проблемы качества данных. Это приводит к неправильным результатам, поскольку инструмент анализирует ошибочные данные. Поэтому важно очистить данные перед их обработкой. В таких ситуациях рекомендуется использовать алгоритмы машинного обучения, поскольку их можно объединить с инструментами интеллектуального анализа данных для автоматизации процесса ввода данных и получения качественных данных. Эта комбинация позволяет легко выявить любые повторяющиеся данные и устранить их. После этого для классификации данных можно использовать алгоритм случайного леса;
- инструменты интеллектуального анализа данных можно использовать для выявления проблем, связанных с процессами, но они не могут найти основную причину проблем. Алгоритмы машинного обучения, наоборот, могут помочь в решении проблемы. Мы также можем представить программное обеспечение с инструментами анализа первопричин и интеллектуального анализа данных, которые могут решать подобные проблемы;

– данные реального времени могут быть структурированными и неструктурированными. Некоторые традиционные инструменты интеллектуального анализа данных могут обрабатывать только структурированные данные и, следовательно, неприменимы к неструктурированным данным. Эту проблему можно решить, используя два алгоритма машинного обучения — оптическое распознавание символов (OCR) и язык естественной обработки (NLP). Методы машинного обучения помогают преобразовать неструктурированные данные в машиночитаемый формат, чтобы инструмент интеллектуального анализа данных мог лучше анализировать и принимать решения. Обратите внимание, что разработчикам следует проявлять осторожность при преобразовании неструктурированных данных в машиночитаемый формат, поскольку это может привести к несовершенству данных и возникновению ошибок;
– иногда инструменты интеллектуального анализа данных обеспечивают меньшую ясность при обработке большого количества переменных. Добавление данных увеличивает сложность результатов интеллектуального анализа данных, которые людям трудно понять. Преодолеть эту проблему помогают инструменты интеллектуального анализа данных, интегрированные с алгоритмами машинного обучения и компьютерным зрением. Следовательно, обработанные данные могут быть собраны и получены соответствующие выходные данные;
– инструменты интеллектуального анализа данных анализируют прошлую эффективность процесса, а не анализируют текущий процесс. Они не могут гарантировать прогнозирование производительности в будущем. Использование приложений машинного обучения с интеллектуальным анализом данных позволяет предсказать окончательные результаты и будущие события.

Они также отправляют пользователям предупреждающее сообщение, если есть какие-либо недостатки и требуются ли какие-либо улучшения.

Интеллектуальный анализ данных в основном заключается в обнаружении скрытых и непредсказуемых взаимосвязей между данными путем обнаружения закономерностей данных, извлечения знаний и раскрытия неизвестной информации.

Понимание этих стратегий интеллектуального анализа данных может быть использовано для оценки вероятности будущих событий, которые можно использовать в различных областях маркетинга, научных открытий, мошенничества, обнаружения вторжений и т. д.

Использование методов интеллектуального анализа данных на основе технологий МО для анализа событий СКУД приобретает важность в настоящее время для прогнозного анализа подозрительных попыток проникновения в помещения организации [22],[23].

Выводы по главе 2

Интеллектуальный анализ данных заключается в обнаружении скрытых и непредсказуемых взаимосвязей между данными путем обнаружения закономерностей данных, извлечения знаний и раскрытия неизвестной информации.

Использование методов интеллектуального анализа данных на основе технологий МО для анализа событий СКУД приобретает важность в настоящее время для прогнозного анализа подозрительных попыток проникновения в помещения организации.

Глава 3 Разработка моделей и алгоритмов аналитической обработки событий СКУД

3.1 Разработка моделей аналитической обработки событий СКУД

На рисунке 19 показан процесс интеллектуального анализа событий СКУД на основе технологий машинного обучения.

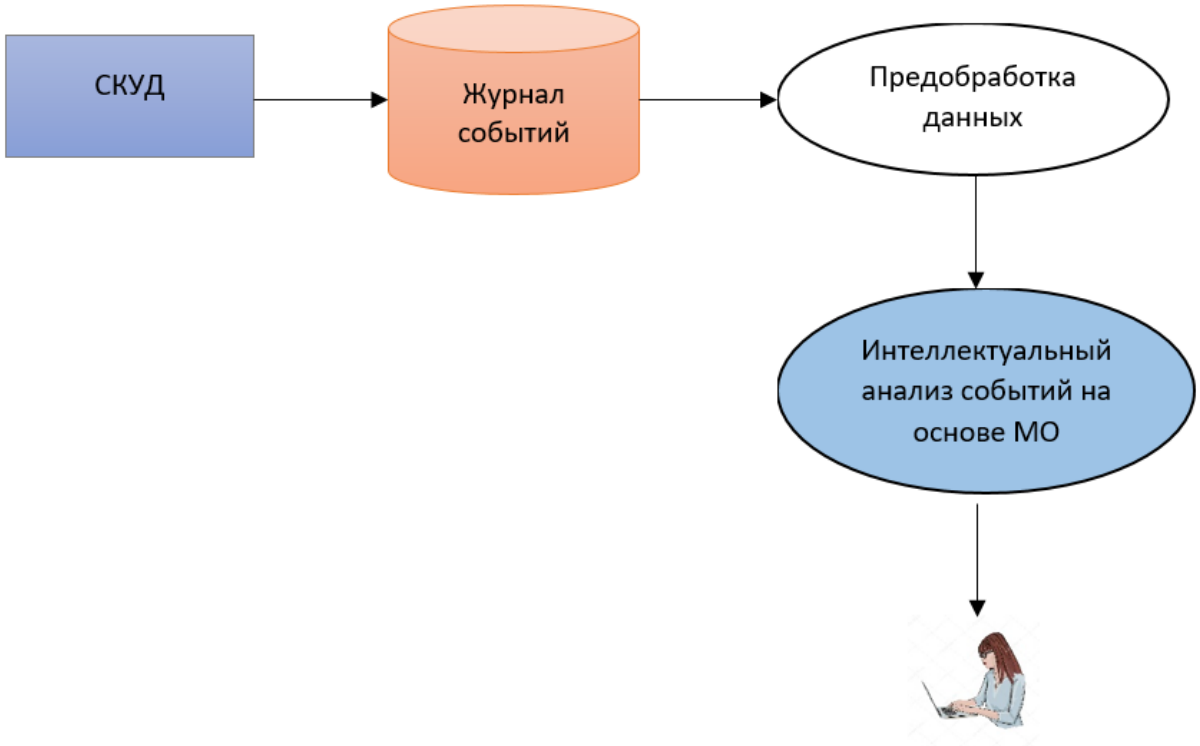

Рисунок 19 – Схема процесса интеллектуального анализа событий СКУД на основе технологий машинного обучения

Разработан модель подсистемы анализа событий СКУД на основе машинного обучения, показанная на рисунке 20.

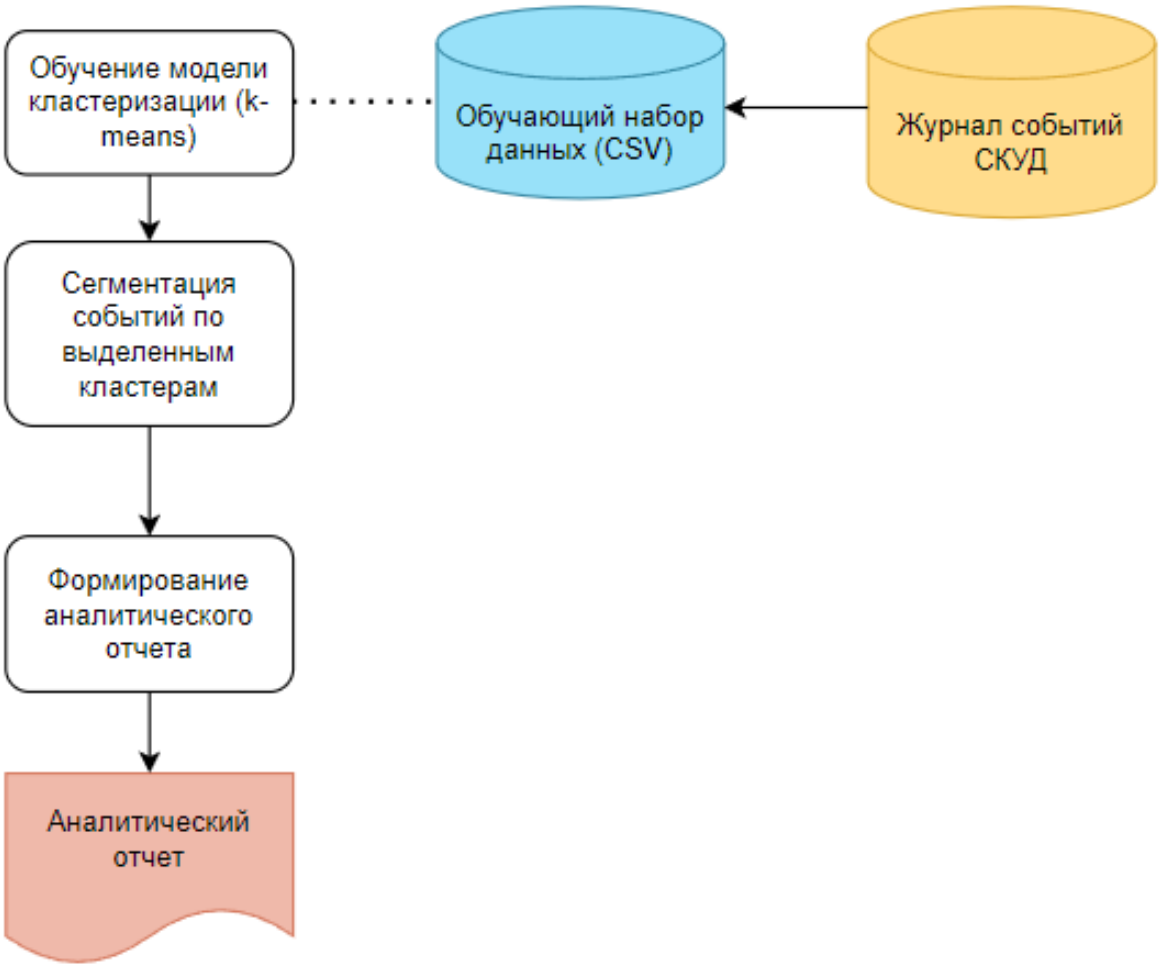

Рисунок 20 – Модель подсистемы анализа событий СКУД

Для отражения основных аспектов системы аналитической обработки событий СКУД выполнено ее логическое проектирование.

В процессе логического моделирования построены диаграммы отражающие функциональный, структурный и динамический аспекты подсистемы анализа событий СКУД.

Для построения функциональной модели аналитической обработки событий СКУД использована диаграмма вариантов использования UML.

При разработке диаграммы вариантов использования выделены следующие акторы: Аналитик и Подсистема анализа событий.

Варианты использования представлены в таблицах 1-3.

Таблица 1 – Описание прецедента: Подготовить обучающую выборку

«Элемент диаграммы	Описание
Прецедент	Подготовить обучающую выборку
ID	1
Краткое описание	Подготовить обучающую выборку
Главный актор	Аналитик
Второстепенный актор	Подсистема анализа событий
Предусловие	Импорт журнала событий СКУД
Основной поток	Аналитик выполняет подготовку обучающей выборки средствами подсистемы анализа событий
Постусловие	Нет
Альтернативные потоки	Нет» [7]

Таблица 2 – Описание прецедента: Сегментация событий

«Элемент диаграммы	Описание
Прецедент	Сегментация событий
ID	2
Краткое описание	Сегментация событий СКУД
Главный актор	Аналитик
Второстепенный актор	Подсистема анализа событий
Предусловие:	Использование модели МО
Основной поток	Аналитик производит сегментацию событий, используя метод кластеризации
Постусловие	Нет
Альтернативные потоки	Нет» [7]

Таблица 3 – Описание прецедента: Формирование аналитического отчета

«Элемент диаграммы	Описание
Прецедент	Формирование аналитического отчета
ID	3
Краткое описание	Формирование аналитического отчета
Главный актор	Аналитик
Второстепенный актор	Подсистема анализа событий
Предусловие	
Основной поток	Аналитик формирует аналитический отчет
Постусловие	Нет
Альтернативные потоки	Нет» [7]

Для разработки диаграмм UML использовано CASE-средство Rational Rose [36].

Диаграмма вариантов использования аналитической обработки событий СКУД показана на рисунке 21.

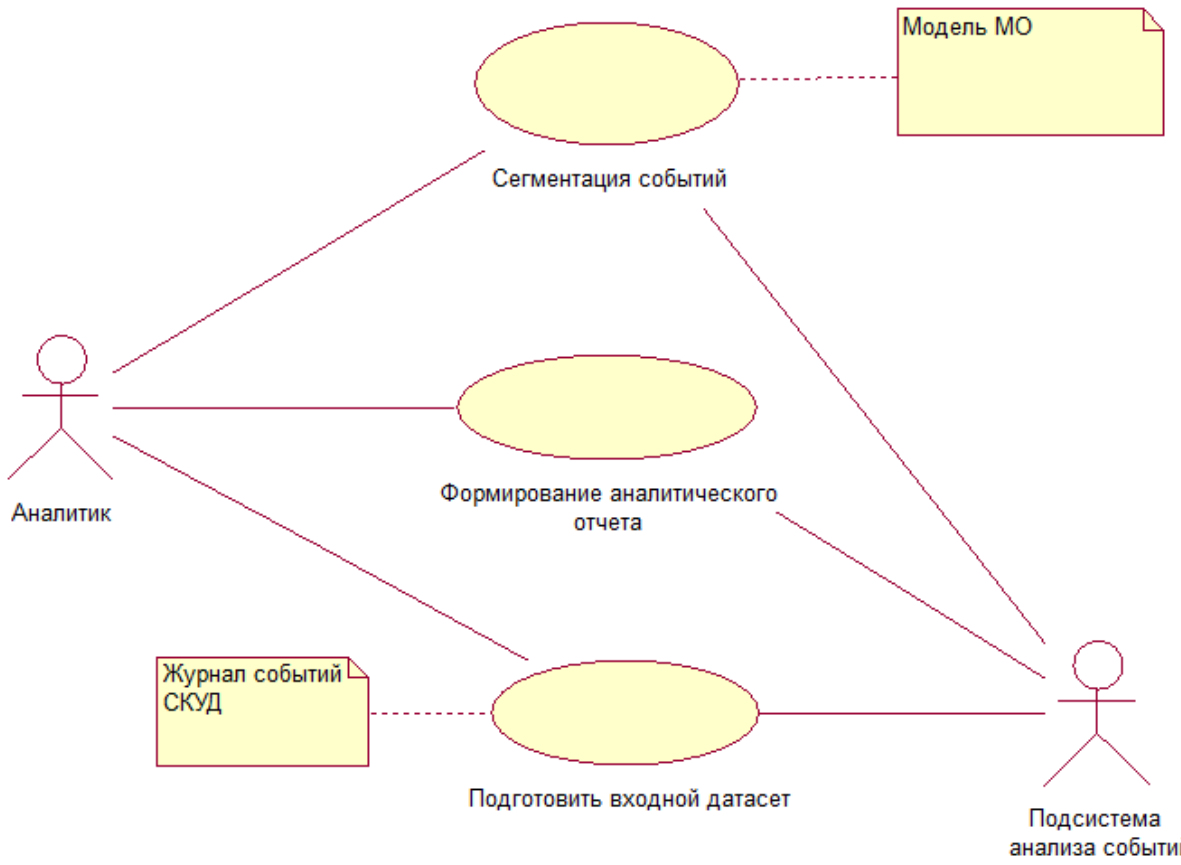

Рисунок 21 – Диаграмма вариантов использования аналитической обработки событий СКУД

«Диаграмма классов — это диаграмма, которая показывает структуру системы, состоящей из классов, их атрибутов, методов и связей между ними.

Диаграмма классов используется для моделирования объектно-ориентированных систем с разных точек зрения: концептуальной, спецификации и реализации» [38].

Диаграмма классов является одним из типов структурных диаграмм в языке моделирования UML.

На рисунке 22 показана диаграмма классов подсистемы анализа событий СКУД.

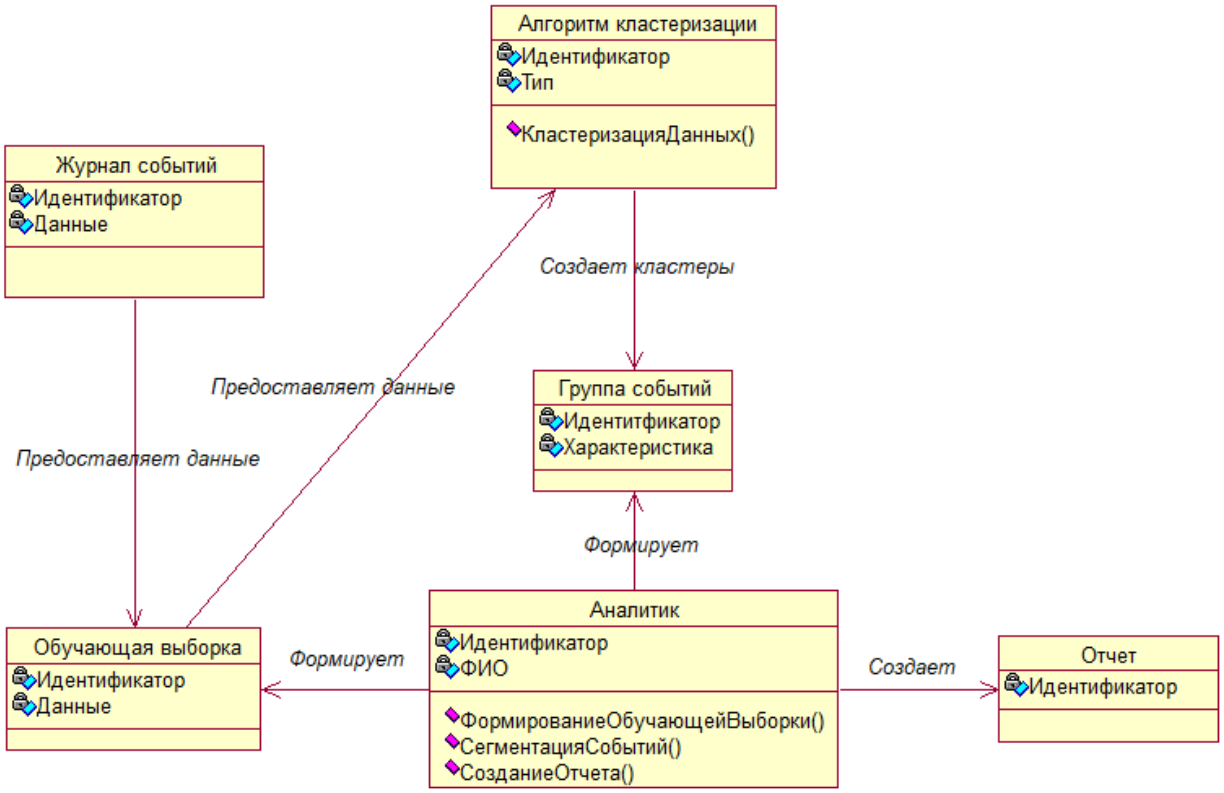

Рисунок 22 – Диаграмма классов подсистемы анализа событий СКУД

В таблице 4 представлена спецификация диаграммы классов подсистемы анализа событий СКУД.

Таблица 4 – Спецификация диаграммы классов подсистемы анализа событий СКУД

Класс	Описание
«Аналитик	Класс объектов, моделирующих на логическом уровне пользователей подсистемы
Журнал событий	Класс объектов, моделирующих на логическом уровне источник данных для анализа» [38]

Продолжение таблицы 4

Класс	Описание
«Обучающая выборка	Класс объектов, моделирующих на логическом уровне датасет для обучения модели МО
Алгоритм кластеризации	Класс объектов, моделирующих на логическом уровне алгоритмы кластеризации обучающей выборки
Группа событий	Класс объектов, моделирующих на логическом уровне группы событий СКУД
Отчет	Класс объектов, моделирующих на логическом уровне аналитические отчеты» [38]

Для описания процесса анализа событий СКУД используем диаграмму деятельности UML.

Диаграммы деятельности — это визуальное представление рабочего процесса системы, показывающее последовательность действий, действий и решений. Они широко используются в различных областях, включая моделирование бизнес-процессов и системный анализ. Понимание преимуществ и недостатков диаграмм деятельности может помочь профессионалам принимать обоснованные решения при использовании этого мощного инструмента [20].

Диаграммы деятельности обладают рядом преимуществ, что делает их ценным инструментом системного моделирования и анализа:

- визуальное представление. Диаграммы действий обеспечивают четкое и интуитивно понятное визуальное представление сложных систем, что упрощает понимание и передачу рабочего процесса;
- легко понять. В диаграммах действий используются простые графические обозначения, такие как действия, узлы принятия решений и переходы, что упрощает их понимание заинтересованными сторонами с ограниченными техническими знаниями;

- стандартизация процессов. Документируя последовательность действий и точки принятия решений в системе, диаграммы действий помогают стандартизировать процессы и обеспечить последовательное выполнение;
- выявление узких мест. Диаграммы действий подчеркивают критические пути и потенциальные узкие места в системе, что позволяет улучшить оптимизацию и повысить производительность;
- эффективное планирование и анализ Диаграммы действий помогают планировать и анализировать поведение системы, определять необходимые ресурсы и зависимости, а также улучшать общее управление проектом.

Диаграмма деятельности процесса аналитической обработки событий СКУД показана на рисунке 23.

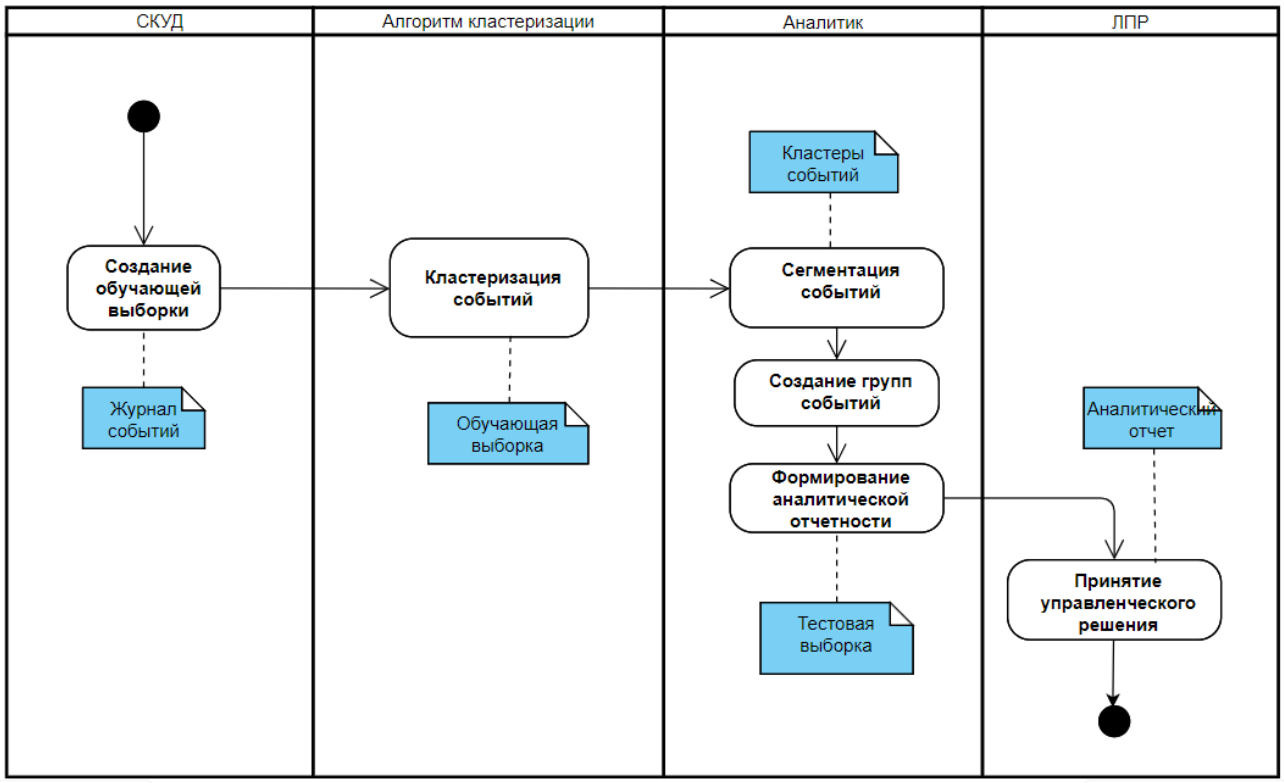

Рисунок 23 - Диаграмма деятельности процесса аналитической обработки событий СКУД

В качества лица, принимающего решения (ЛПР) может выступать руководители службы безопасности и других подразделений организации, имеющие необходимые полномочия.

Диаграмма деятельности UML – это вид диаграммы, которая показывает, какие действия выполняются в системе или бизнес-процессе. Она состоит из разных элементов, таких как действия, узлы управления, потоки управления и объектов, которые соединяются стрелками. Диаграмма деятельности UML может использоваться для моделирования различных аспектов поведения системы, таких как последовательность, параллелизм, ветвление и синхронизация. Диаграмма деятельности UML также может включать разделы, которые группируют действия по разным действующим лицам или ролям.

Процесс аналитической обработки событий СКУД организован следующим образом:

- на основе данных журнала событий СКУД создается обучающая выборка. Для повышения эффективности обучения и точности прогнозирования целесообразно использовать большой объем достоверных данных;
- выполняется кластеризацию данных обучающей выборки с помощью выбранного алгоритма кластеризации;
- выполняются сегментация и создание групп событий на основе результатов кластеризации;
- формируется аналитический отчет для поддержки принятия управленческих решений лицом, принимающим решения (ЛПР).

Рассмотрим процесс выбора алгоритмов аналитической обработки событий СКУД.

3.2 Алгоритмы аналитической обработки событий СКУД

«В аналитике данных мы часто имеем очень большие данные (много наблюдений — «строки в плоском файле»), которые, однако, похожи друг на друга, поэтому мы можем захотеть организовать их в несколько кластеров со схожими наблюдениями внутри каждого кластера. Например, в случае с данными о клиентах, хотя у нас могут быть данные от миллионов клиентов, эти клиенты могут принадлежать только к нескольким сегментам: клиенты похожи внутри каждого сегмента, но различны в разных сегментах. Часто нам может потребоваться проанализировать каждый сегмент отдельно, поскольку они могут вести себя по-разному (например, разные сегменты рынка могут иметь разные предпочтения в отношении продуктов и модели поведения).

В таких ситуациях для идентификации сегментов данных можно использовать статистические методы, широко называемые методами кластеризации.

Основываясь на том, как мы определяем «сходства» и «различия» между наблюдаемыми данными (например, клиентами или активами), которые также можно определить математически с использованием показателей расстояния, можно найти различные решения для сегментации.

Ключевым компонентом кластеризации и сегментации является именно определение этих показателей расстояния (между наблюдениями), которые необходимо определять творчески, основываясь на контекстуальных знаниях, а не только с использованием математических уравнений и методов «черного ящика».

Методы кластеризации используются для группировки данных/наблюдений в несколько сегментов, чтобы данные внутри любого сегмента были одинаковыми, а данные в разных сегментах различались.

Определение того, что мы имеем в виду, когда говорим про похожие или разные наблюдения, является ключевой частью кластерного анализа, который часто требует большого количества контекстуальных знаний и творческого

подхода, выходящего за рамки того, что могут обеспечить статистические инструменты» [31].

Для выбора алгоритмов МО для описанной модели рассмотрим и сравним алгоритмы кластеризации.

Сравнение алгоритмов кластеризации включает в себя анализ различных методов группировки данных на основе их сходства.

Рассмотрим некоторые из наиболее распространенных методов и алгоритмов кластеризации:

- к-means: этот алгоритм разделяет данные на K кластеров, минимизируя сумму квадратов расстояний от точек до центра их кластера. Он хорошо работает с большими наборами данных и является одним из самых популярных методов кластеризации;
- иерархическая кластеризация: этот метод создает дерево вложенных кластеров, позволяя анализировать данные на разных уровнях детализации. Он может быть полезен для интерпретации структуры данных;
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise): алгоритм, который группирует точки на основе плотности их распределения, позволяя обнаруживать кластеры произвольной формы и выделять выбросы;
- алгоритмы на основе плотности: Эти методы определяют кластеры как области высокой плотности, окруженные областями низкой плотности. Они могут быть эффективны для данных с шумами и выбросами;
- спектральная кластеризация: использует собственные значения матрицы сходства для уменьшения размерности перед кластеризацией. Этот метод подходит для сложных структур.

Для сравнения алгоритмов кластеризации используем таблицу 5.

Таблица 5 – Сравнение характеристик алгоритмов кластеризации

Алгоритм	Преимущества	Недостатки
«k-means	Простота использования; быстрота использования; понятность и прозрачность алгоритма	Алгоритм слишком чувствителен к выбросам, которые могут искажать среднее; медленная работа на больших базах данных; необходимо задавать количество кластеров
Иерархическая кластеризация	Выполняет кластеризацию на высоком уровне даже при наличии выбросов, выделяет кластеры сложной формы и различных размеров, обладает линейно зависимыми требованиями к месту хранения данных и временную сложность для данных высокой размерности.	Есть необходимость в задании пороговых значений и количества кластеров» [9]

На основании результатов анализа характеристик алгоритмов выбираем алгоритм k-means, как наиболее простой в реализации.

«Алгоритм k-means – это итеративный алгоритм, который пытается разделить набор данных на заранее определенные *k*-отдельных неперекрывающихся групп-кластеров, где каждая точка данных принадлежит только одной группе.

Алгоритм k-means пытается сделать точки данных внутри кластера как можно более похожими, но при этом сохраняя кластеры как можно более разными.

Он назначает точки данных кластеру таким образом, чтобы сумма квадратов расстояния между точками данных и центроидом кластера (среднем арифметическим всех точек данных, принадлежащих этому кластеру) была минимальной.

Чем меньше вариаций внутри кластеров, тем более однородные

(похожие) точки данных находятся в одном кластере.

Подход, который использует k-means для решения проблемы, называется EM (Expectation-Maximization)-алгоритмом.

EM-алгоритм состоит из итерационного повторения двух шагов. На E-шаге вычисляется ожидаемое значение (expectation) вектора скрытых переменных G по текущему приближению вектора параметров Θ. На M-шаге решается задача максимизации правдоподобия (maximization) и находится следующее приближение вектора Θ по текущим значениям векторов G и Θ» [6].

Блок-схема алгоритма k-means показана на рисунке 24.

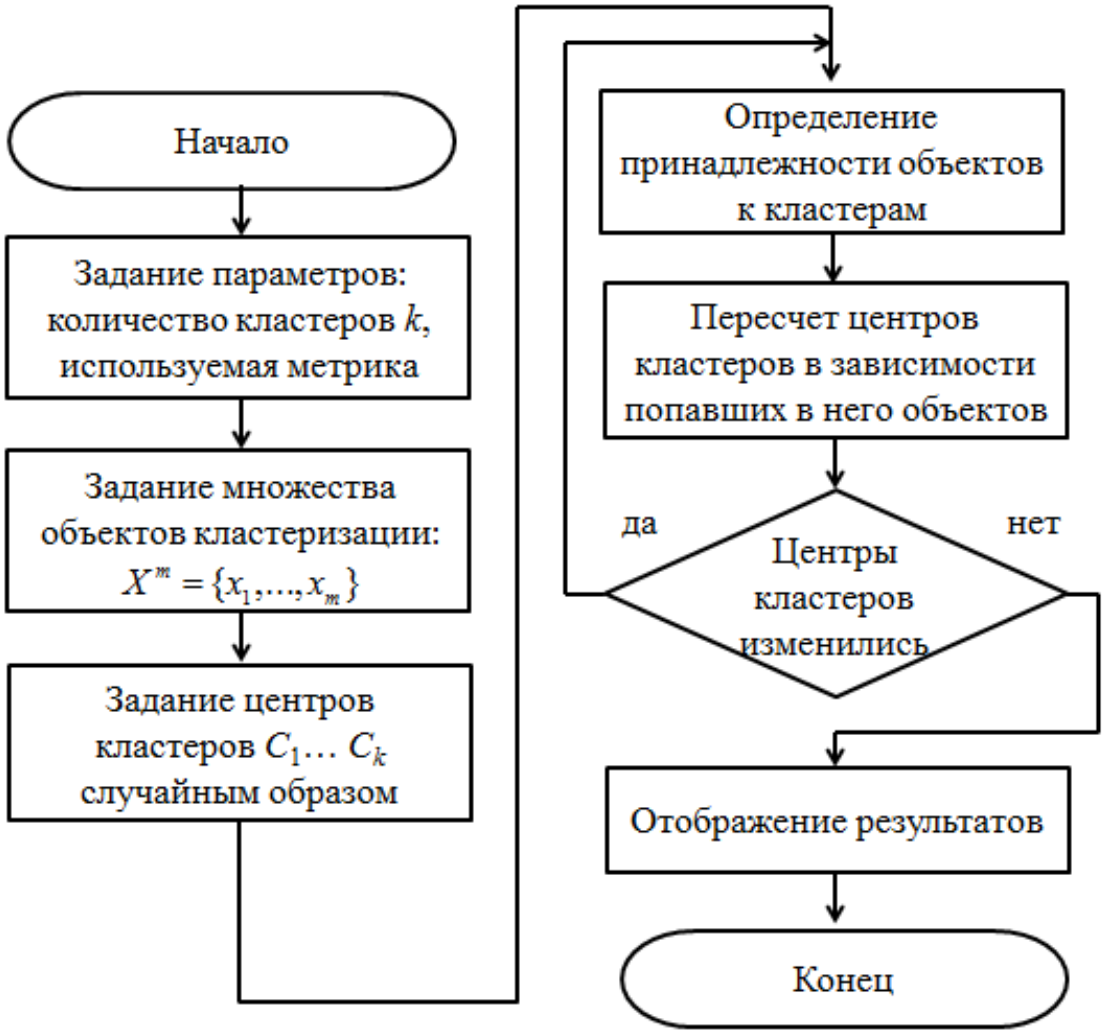

Рисунок 24 – Блок-схема алгоритма k-means

«Алгоритм k-means состоит из следующих шагов:

Шаг 1. Случайным образом выбирается *k* объектов обучающей выборки, которые будут служить начальными центрами кластеров.

Шаг 2. Для каждого объектов обучающей выборки определяется ближайший к ней центр кластера. Для этого вычисляется расстояние между объектами и центрами кластеров. Считается, что объект принадлежит тому кластеру, к которому он ближе. В качестве формулы для оценки близости объектов в многомерном пространстве признаков используется одна из известных метрик.

Шаг 3. Как только состав кластеров на данной итерации известен, производится расчёт новых центров кластеров. Это делается путем расчета средних значений для каждого числового признака по всем объектам рассматриваемого кластера. Например, в двухмерном пространстве координаты центр кластера на основе вошедших в него *t* объектов рассчитывается следующим образом (1)» [6]:

$$(P_{2ц}, P_{1ц}) = \left(\frac{\sum_{1}^{t} P_1(t)}{t}, \frac{\sum_{1}^{t} P_2(t)}{t} \right) \quad (1)$$

«Шаг 4. Шаги 2 и 3 повторяются до тех пор, пока не выполниться один из двух критериев остановки:

- границы кластеров и расположения центров кластеров не перестанет изменяться от итерации к итерации, т.е. на каждой итерации в каждом кластере будет оставаться один и тот же набор записей. На практике алгоритм k-means обычно находит набор стабильных кластеров за несколько десятков итераций;
- достигнут критерий сходимости. Чаще всего используется критерий суммы квадратов ошибок между центром кластера и всеми

вошедшими в него объектами (2):

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} (p - m_i)^2 \quad (2)$$

где $p \in C_i$ - произвольная точка данных, принадлежащая кластеру C_i, m_i – центр данного кластера. Иными словами, алгоритм остановится тогда, когда ошибка E достигнет достаточно малого значения.

Преимуществом алгоритма k-means является высокая производительность, поскольку все, что мы на самом деле делаем, это вычисляем расстояния между точками и центроидами групп.

Алгоритм имеет линейную сложность $O(n)$.

С другой стороны, у k-means есть недостатки:

- необходимо выбрать количество групп/классов. Это не всегда тривиально, и в идеале с алгоритмом кластеризации мы хотели бы, чтобы он выяснил это за нас, потому что его цель - получить некоторое представление о данных;
- алгоритм начинается со случайного выбора центров кластеров и, следовательно, может давать разные результаты кластеризации при разных прогонах алгоритма.

Таким образом, результаты могут быть неповторимыми и противоречивыми.

Поэтому большая область исследований в области кластеризации была сосредоточена на улучшении процесса кластеризации в том числе с помощью адаптивных алгоритмов k-means.

Идея адаптивного алгоритма состоит в том, чтобы оптимизировать набор выборочных данных с помощью правила локтя для выявления и устранения выбросов» [6].

На рисунке 25 представлена блок-схема адаптивного алгоритма k-means.

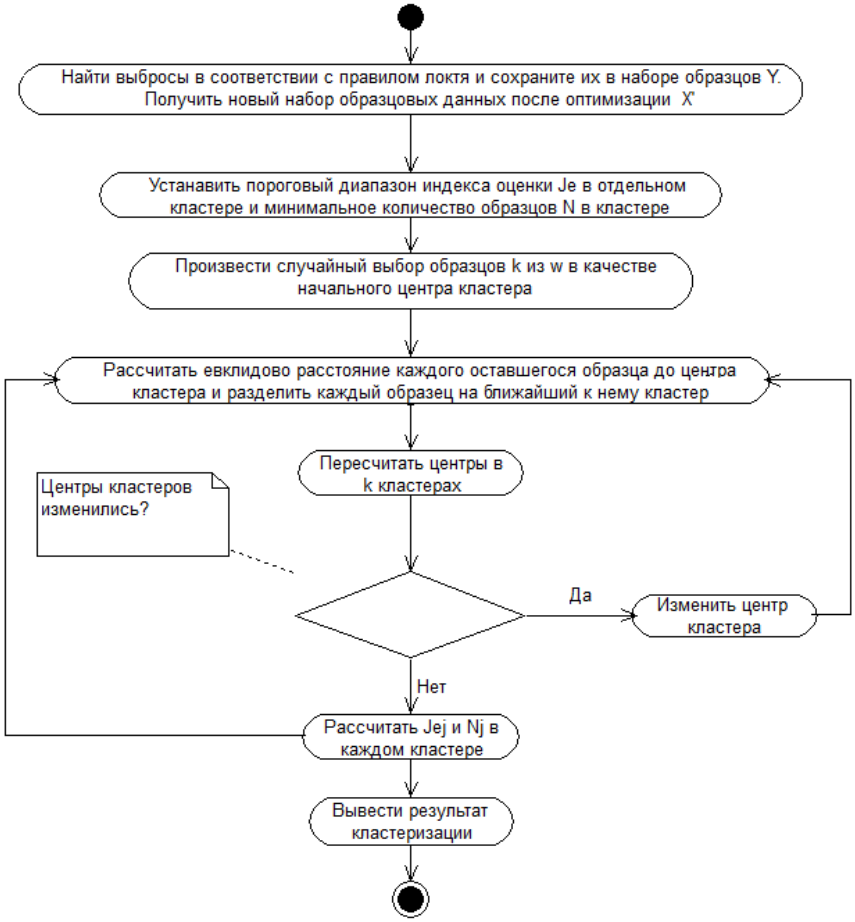

Рисунок 25 – Блок-схема адаптивного алгоритма k-means

«Метод локтя (Elbow Rule) – один из самых известных методов, с помощью которого вы можете выбрать правильное значение k и повысить производительность модели. Этот эмпирический метод вычисляет сумму квадратов расстояний между точками и вычисляет среднее значение.

Когда алгоритм реализован, используется выборочный набор данных, который устраняет выбросы, а после завершения алгоритма окончательный выброс определяется в соответствии с сходством между выбросами и каждым кластером.

На основе адаптивной идеи после завершения каждой итерации значение автоматически корректируется в соответствии с ошибкой индекса оценки кластера для каждого кластера до тех пор, пока не будет достигнут диапазон ошибок.

Определение подобия классического алгоритма k-means основано на евклидовом расстоянии. Выбросы будут влиять на оценку значения, тем самым увеличивая временную сложность алгоритма.

Используем метод локтя для эффективного обнаружения выбросов в наборе данных для оптимизации алгоритма» [26].

Выводы по главе 3

Результаты проделанной работы позволили сделать следующие выводы:

- для отражения основных аспектов системы аналитической обработки событий СКУД выполнено ее логическое проектирование. В процессе логического моделирования построены диаграммы отражающие функциональный, структурный и динамический аспекты подсистемы анализа событий СКУД;
- на основании результатов анализа характеристик алгоритмов для сегментации событий СКУД выбраны метод кластеризации и алгоритм k-means как наиболее простой для реализации и визуализации.

Представленные модели и алгоритмы являются основой для построения подсистемы аналитической обработки событий СКУД.

Глава 4 Апробация проектных решений и оценка их эффективности

4.1 Разработка прототипа и апробация проектного решения

Для представления архитектуры подсистемы анализа событий СКУД разработана диаграмма компонентов системы.

Диаграмма компонентов подсистемы анализа событий СКУД показана на рисунке 26.

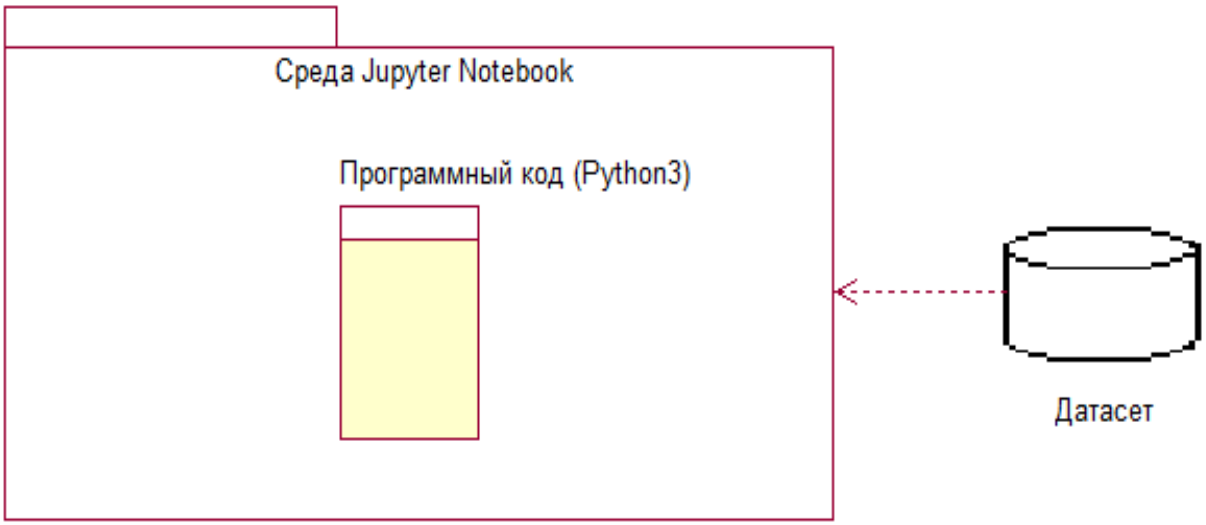

Рисунок 26 – Диаграмма компонентов прототипа подсистемы анализа событий СКУД

Преимущества диаграмм компонентов [16]:

- они просты, стандартизированы и понятны;
- помогают команде разработчиков визуализировать физическую структуру системы и понять взаимосвязь между различными компонентами;
- полезны для представления реализации системы;
- оказывают помощь при проектировании системы, содержащей интерфейс ввода-вывода;

– применение компонентов многократного использования может помочь снизить общую стоимость разработки;
– дают возможность представить влияние результатов на сервис;
– помогают визуализировать существующие процессы и разработать стратегию будущих.

Для разработки прототипа подсистемы анализа событий СКУД использованы язык Python, среда Jupyter Notebook и библиотека scikit-learn [35], [37].

В качестве источника данных для обучения модели машинного обучения используется файл CSV, сформированный путем конвертации журнала событий СКУД, который импортируется в виде файла рабочей книги Excel.

В журнале событий СКУД зарегистрированы события типов, представленных на рисунке 27.

Номер	Событие
1	Зарегистрирован проход
2	Зарегистрирован проход, санкционированный с кнопки
3	Доступ запрещен. Обработка предыдущего объекта не завершена
4	Зарегистрирован проход при открытой двери
5	Доступ запрещен. Неизвестный номер пропуска

Рисунок 27 – Типы событий СКУД

На рисунке 28 показан программный код загрузки и структура входного датасета.

```
1  import pandas as pd
2  import numpy as np
3  import seaborn as sns
4  import matplotlib.pyplot as plt
5  import pathlib
6  p =  pathlib.Path('scuddataset1.csv')
7  ds = pd.read_csv(p)
8  ds.shape
9  ds
```

	Date	Time	Direction	Event	Object
0	26122017	100518	1	1	101
1	20171226	104403	2	1	101
2	20171226	104758	1	1	101
3	20171226	113842	1	1	102
4	20171226	114136	2	1	102
...	...	...	...	...	...
152	20180110	183909	1	1	102
153	20180110	190524	2	1	102
154	20180110	202907	2	4	101
155	20180110	202933	1	1	101
156	20180110	203022	2	1	101

157 rows × 5 columns

Рисунок 28 – Программный код загрузки и структура входного датасета

Далее произведены очистка, трансформация и предподготовка датасета. Результатом является датафрейм для обучения модели МО, код формирования и статистика которого показаны на рисунке 29.

```python
#Переименование столбцов
ds = ds.rename(columns={"Date":"Дата", "Time": "Время",
                        "Direction": "Направление",
                        "Event":"Событие","Object":"ТабНом"})
ds.head(5)
#замена значений None на значение 1
ds = ds.fillna(1)
#убрать дубликаты
#ds.drop_duplicates()
ds.describe()
```

	Дата	Время	Направление	Событие	ТабНом
count	1.570000e+02	157.000000	157.000000	157.000000	157.000000
mean	2.021156e+07	143271.815287	1.611465	1.656051	154.573248
std	4.747459e+05	33522.509508	0.488977	1.010924	88.168342
min	2.017123e+07	91353.000000	1.000000	1.000000	101.000000
25%	2.017123e+07	114136.000000	1.000000	1.000000	101.000000
50%	2.017123e+07	142830.000000	2.000000	1.000000	102.000000
75%	2.018011e+07	170002.000000	2.000000	2.000000	300.000000
max	2.612202e+07	224606.000000	2.000000	5.000000	300.000000

Рисунок 29 – Программный код формирования и статистика датафрейма для обучения модели МО

Разработаны аналитические отчеты для формирования рекомендаций для поддержки принятия управленческих решений.

Разработана тепловая карта датасета.

На рисунке 30 показаны программный код и представление тепловой карты признаков набора данных датафрейма.

```
#Тепловая карта
plt.figure(figsize=(10,5))
c= ds.corr()
sns.heatmap(c,cmap="YlGnBu",annot=True)
plt.title('Тепловая карта признаков набора данных')
c
```

	Дата	Время	Направление	Событие	ТабНом
Дата	1.000000	-0.101963	-0.101041	-0.053237	-0.050033
Время	-0.101963	1.000000	0.170434	-0.010057	-0.081161
Направление	-0.101041	0.170434	1.000000	-0.129431	0.452005
Событие	-0.053237	-0.010057	-0.129431	1.000000	0.248549
ТабНом	-0.050033	-0.081161	0.452005	0.248549	1.000000

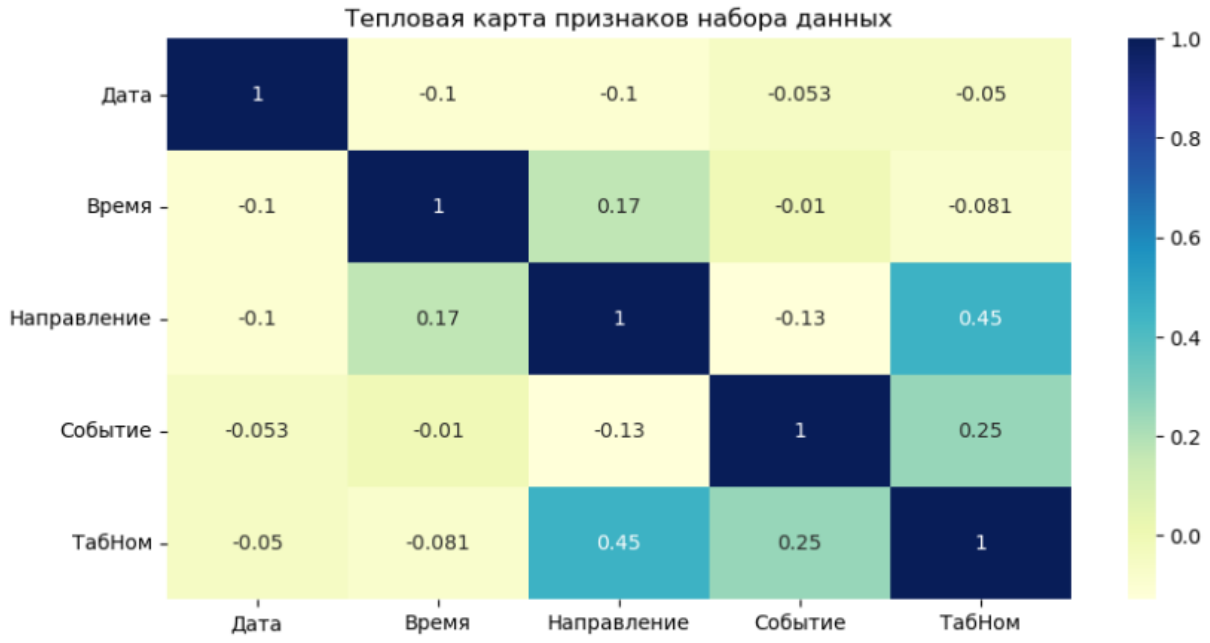

Рисунок 30 – Программный код создания и представление тепловой карты признаков набора данных

В тепловой карте обычно все строки относятся к одной категории (метки отображаются слева или справа), а все столбцы относятся к другой категории (метки отображаются вверху или внизу).

Отдельные строки и столбцы разделены на подкатегории, которые совпадают друг с другом в матрице. Ячейки представляют собой пересечения строк и столбцов, которые могут содержать категориальные данные или числовые данные [33].

На рисунке 31 представлены программный код и представление диаграммы распределения событий по типу.

```
ds.Событие.value_counts().nlargest(100).plot(kind='bar')  #GH
plt.title("Распределение по типу событий")
plt.ylabel('Количество событий')
plt.xlabel('Типы событий');
```

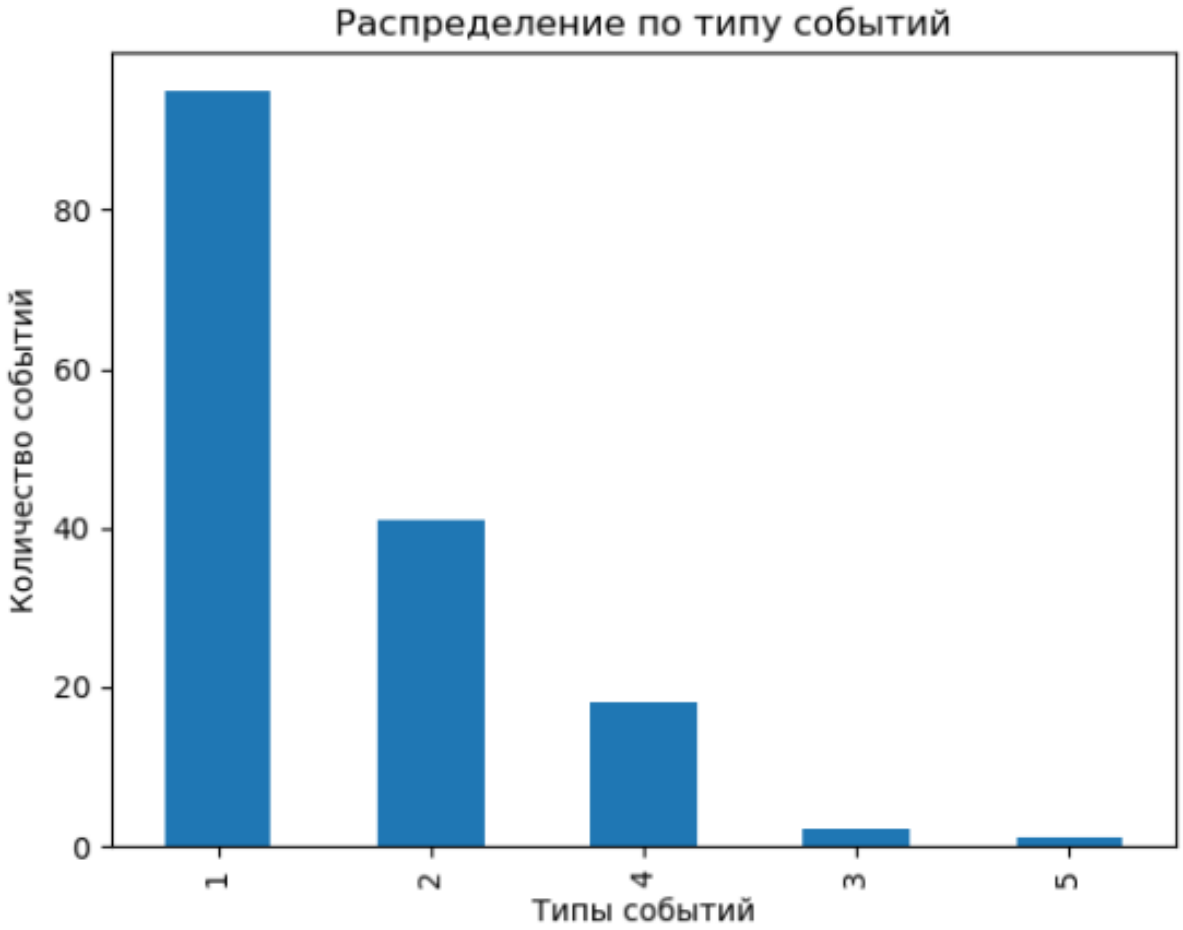

Рисунок 31 – Диаграмма распределения событий по типу

На рисунках 32 и 33 представлены программный код и представление диаграммы распределения событий по табельным номерам сотрудников в долях, соответственно.

```
#Распределение по долям
params = {'font.size' : 14}
color = 'Set2'
palette_color = sns.color_palette(color)
fig,ax = plt.subplots(figsize = (9,8))
plt.title("Распределение по таб.номерам (%)")
ax = sns.countplot(x = "ТабНом", data=ds, palette=color, order = ds.ТабНом.value_counts().index)
ax.set_ylabel('Количество событий')
ax.set_xlabel('Табельные номера')
patches = ax.patches
for j in range(len(patches)):
        percentage = list(ds.ТабНом.value_counts())[j]/ds.ТабНом.value_counts().sum()
        offset = ds.Событие.value_counts().max() * 0.01
        x = patches[j].get_x() + patches[j].get_width()/2
        y = patches[j].get_height() + offset
        ax.annotate('{:.1f}%'.format(percentage*100), (x, y), ha='center')
```

Рисунок 32 – Программный код диаграммы распределения событий по табельным номерам сотрудников (в долях)

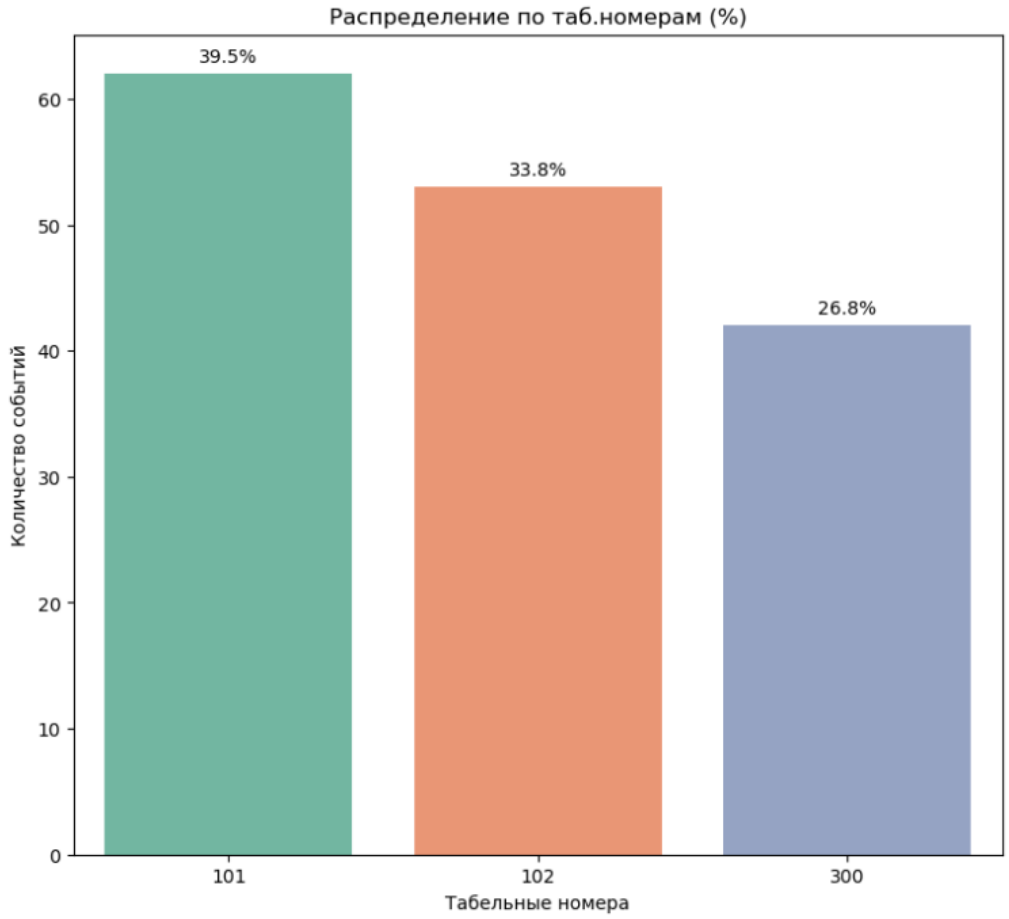

Рисунок 33 – Диаграмма распределения событий по табельным номерам сотрудников (в долях)

Диаграмма «ящик с усами» (или ящичковая диаграмма) — это удобный способ визуального отображения распределения данных через их квартили.

«Линии, идущие параллельно от прямоугольников, известны как «усы» и используются для обозначения изменчивости за пределами верхнего и нижнего квартилей.

Выбросы иногда изображаются в виде отдельных точек, расположенных на одной линии с усами. Ящичные диаграммы можно рисовать как вертикально, так и горизонтально» [24].

На рисунке 34 показаны код и изображение ящичковой диаграммы.

```
#Ящики с усами
import seaborn as sns
sns.boxplot(x=ds['Событие'],color=".8")
```

```
<Axes: xlabel='Событие'>
```

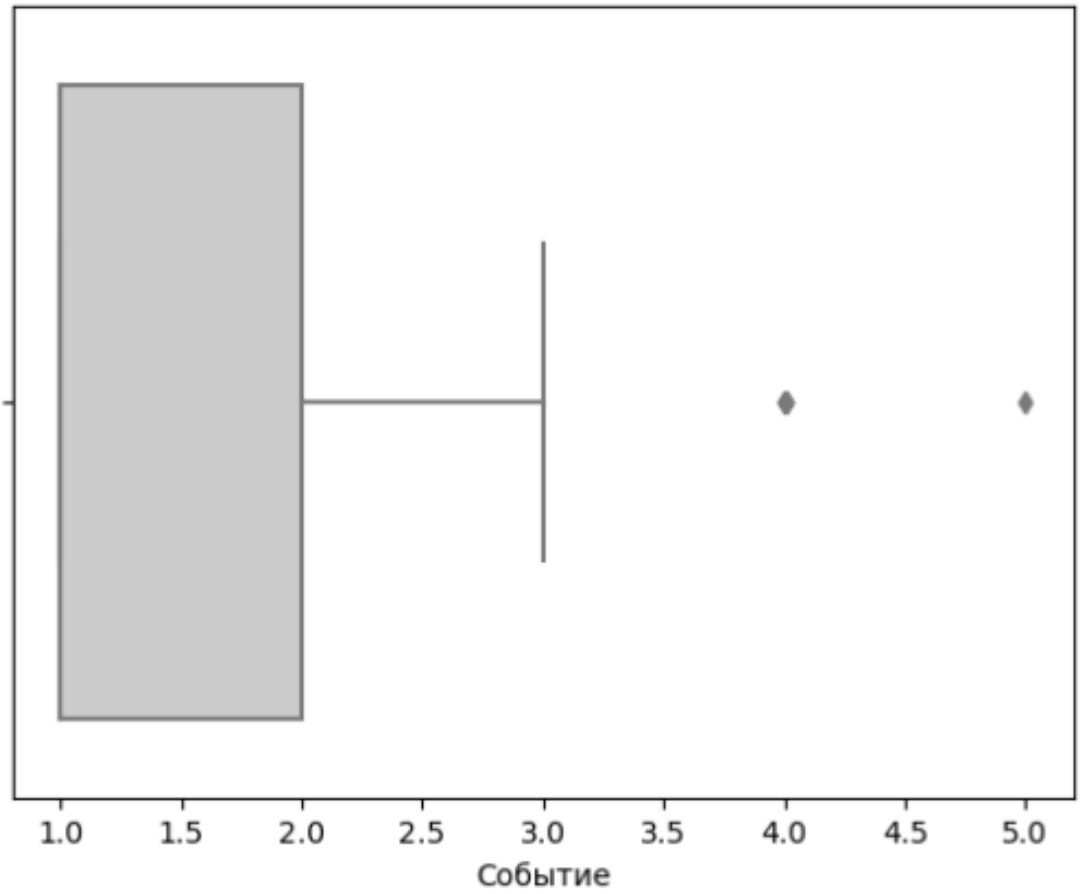

Рисунок 34 – Ящичковая диаграмма событий СКУД

На рисунках 35 и 36 показаны программный код алгоритма k-means и график результатов кластеризации датафрейма по признакам «ТабНом» и «Тип события», соответственно.

```
#KMeans
from sklearn.cluster import KMeans
indata = pd.read_csv(p)
kmeans = KMeans(5)
kmeans.fit(indata.values)
import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap
%matplotlib inline
customcmap = ListedColormap(["crimson", "mediumblue", "darkmagenta"])
fig, ax = plt.subplots(figsize=(10, 8))
plt.scatter(x=indata['Object'], y=indata['Event'], s=200,
            c=kmeans.labels_,
            cmap=customcmap)
ax.set_xlabel(r'ТабНом', fontsize=14)
ax.set_ylabel(r'Событие', fontsize=14)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```

Рисунок 35 – Программный код алгоритма k-means

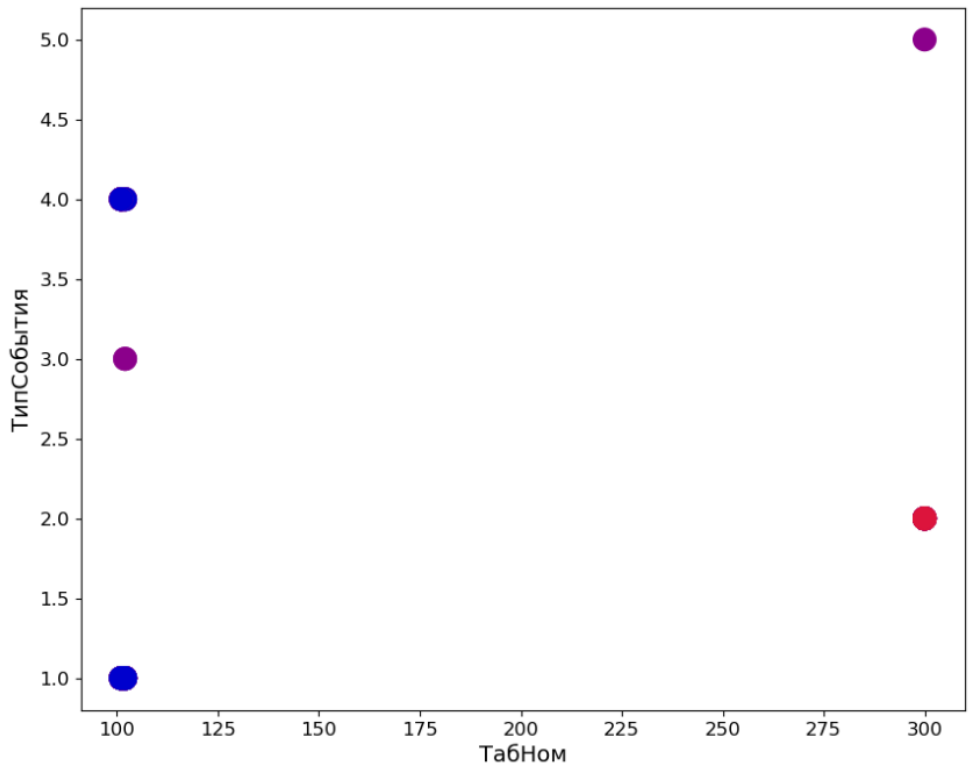

Рисунок 36 – График результатов кластеризации датафрейма по признакам «ТабНом» и «Тип события»

Одним из преимуществ алгоритма k-means является возможность прогнозирования вхождения конкретной записи в журнале в определенную группу событий.

На рисунке 37 показан пример прогнозирования вхождения события с заданными параметрами в конкретную группу событий.

```
#Прогнозирование
print(kmeans.labels_)
print(kmeans.cluster_centers_)
predicted_class = kmeans.predict([[20180108,195640,2,1,101]])
print('Группа событий:',predicted_class)
```

```
[1 3 3 3 3 3 3 3 3 3 0 0 0 0 0 0 0 0 0 4 4 4 4 4 4 4 4 2 2 2 2 3 3 3 3 3
 3 3 3 3 0 0 0 0 0 0 0 0 0 0 0 0 4 4 4 4 4 4 4 4 4 4 4 4 4 4 2 2 2 2 2 2
 2 2 2 2 3 3 3 3 3 3 3 3 3 0 0 0 0 0 0 0 0 0 4 4 4 4 2 2 2 3 3 3 3 3 3 3 0
 0 0 2 2 2 3 3 3 3 0 0 0 0 0 0 4 4 4 4 2 2 2 2 3 3 3 3 3 3 3 0 0 0 0 4 4
 4 4 4 2 2 2 2 2 2]
[[2.01732458e+07 1.33833727e+05 1.65909091e+00 1.47727273e+00
  1.69136364e+02]
 [2.61220170e+07 1.00518000e+05 1.00000000e+00 1.00000000e+00
  1.01000000e+02]
 [2.01747800e+07 1.94349333e+05 1.66666667e+00 1.40000000e+00
  1.01466667e+02]
 [2.01735445e+07 1.05355217e+05 1.50000000e+00 1.73913043e+00
  1.48869565e+02]
 [2.01734475e+07 1.61879250e+05 1.66666667e+00 2.00000000e+00
  1.89805556e+02]]
Группа событий: [2]
```

Рисунок 37 – График результатов кластеризации датафрейма по признакам «ТабНом» и «Тип события»

Следует отметить, что в прототипе подсистемы анализа событий СКУД также реализован метод классификации событий СКУД с помощью классификатора дерева решений.

Классификаторы дерева решений успешно используются во многих различных областях. Их наиболее важной особенностью является способность извлекать описательные знания для принятия решений из предоставленных

данных. Дерево решений может быть создано на основе обучающих наборов.

Для реализации классификатора дерева решений используется соответствующий модуль библиотеки scikit-learn (рисунок 38).

```
#выполним классификацию
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier()
classifier.fit(X_train, y_train)
```

```
▾ DecisionTreeClassifier
DecisionTreeClassifier()
```

```
#построим дерево решений
from sklearn import tree
tree.plot_tree(classifier)
```

Рисунок 38 – Программный код классификатора дерева решений

На рисунке 39 показано полученное дерево решений.

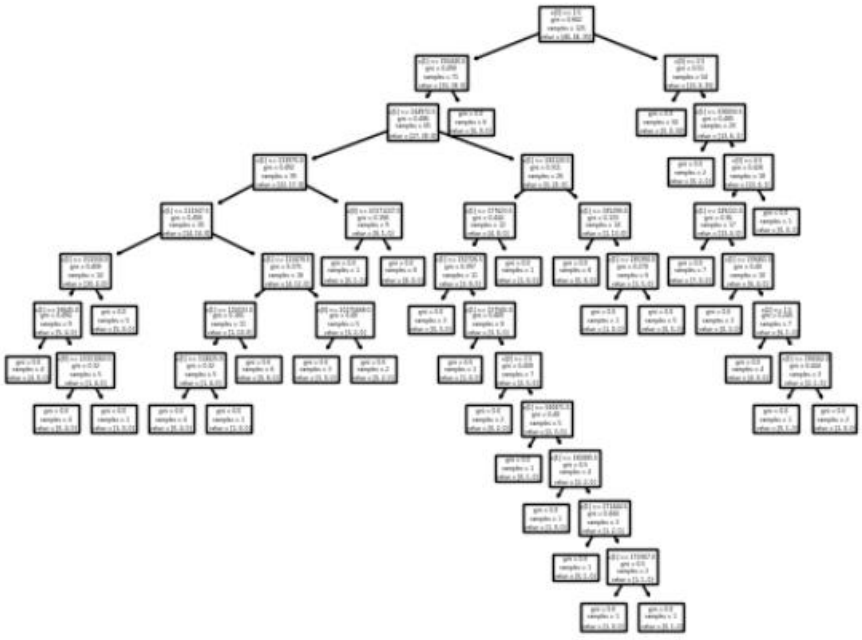

Рисунок 39 – Дерево решений событий СКУД

Таким образом, апробация подтвердила работоспособность прототипа подсистемы анализа событий СКУД.

4.2 Оценка эффективности проектного решения

Для оценки эффективности управления подсистемы анализа событий СКУД используем формулу (3):

$$K_{эу} = \frac{\sum_{i=1}^{n} P_{yi}}{n}, \quad (3)$$

где n – «количество функций управления, реализуемых подсистемой анализа событий СКУД;

P_{yi} – вероятность подсистемой анализа событий СКУД эффективного управляющего воздействия при реализации *i*-й функции управления» [4].

«Для решения задач управления в подсистеме анализа событий СКУД используются следующие функции:

- формирование аналитического отчета;
- принятие решения ЛПР.

Как показывает практика, на выполнение функции «Принятие решения ЛПР» может негативно повлиять человеческий фактор.

Пусть вероятность выработки эффективного управляющего воздействия для данной функции равна 0.5.

В этом случае значение показателя функциональной эффективности управления подсистемы анализа событий СКУД будет равно:

$$К_{эу} = 1.5/2 = 0,75$$

Таким образом, коэффициент эффективности управления предлагаемой подсистемы анализа событий СКУД Кэу> 0,5, что свидетельствует о высокой функциональной эффективности аналитической обработки событий СКУД» [4].

«Для оценки экономической эффективности проектных решений используем методику сравнения затрат на разработку подсистемы анализа событий СКУД внешним программистом по договору аутсорсинга (базовый вариант) и программистом образовательной организации (проектный вариант), соответственно.

В калькуляцию себестоимости заказной разработки подсистемы анализа событий СКУД включаются следующие статьи затрат:

- зарплата исполнителя проекта по трудовому договору ($ЗБ_1$);
- социальные страховые взносы ($ЗБ_2$);
- прочие прямые расходы ($ЗБ_3$);
- накладные расходы ($ЗБ_4$).

В заказной доработке задействован внешний программист» [11].

Средняя стоимость часа работы программиста Python по договору составляет 1500 руб [13].

Ориентировочное время разработки составляет 100 час.

«Итого затраты базового варианта $С_{баз}$ составят:

$$С_{баз} = ЗБ_1 + ЗБ_2 + ЗБ_3 + ЗБ_4$$

Таким образом:

$$Сбаз = 1500*100 +0,271*1500*100+0+0 =190650 \text{ руб}$$

В собственной разработке подсистемы анализа событий СКУД задействованы программист и аналитик образовательной организации» [11].

«В калькуляцию себестоимости собственной разработки подсистемы анализа событий СКУД включаются следующие статьи затрат:

- зарплата исполнителей проекта с учетом затраченного времени 100 час ($ЗП_1$);
- социальные страховые взносы ($ЗП_2$);
- прочие прямые расходы ($ЗП_3$);
- накладные расходы ($ЗП_4$)» [11].

«Итого затраты проектного варианта $С_{пр}$ составят (4):

$$С_{пр} = ЗП_1 + ЗП_2 + ЗП_3 + ЗП_4 \quad (4)$$

Таким образом:

$С_{пр}$ = (50000+40000) руб+0,3*(50000+40000)+0+0 = 117000 руб

Сформируем таблицу и график показателей экономической эффективности проекта разработки подсистемы анализа событий СКУД (таблица 6, рисунок 40)» [11].

Таблица 6 – Показатели эффективности проекта разработки подсистемы анализа событий СКУД

«Затраты		Абсолютное изменение затрат	Коэффициент относительного снижения затрат	Индекс снижения затрат
Базовый вариант	Проектный вариант			
$С_{баз}$ (руб.)	$С_{пр}$ (руб.)	$\Delta С = С_{баз} - С_{пр}$ (руб.)	$К_С = \Delta С / С_{баз} \times 100\%$	$У_С = С_{баз} / С_{пр}$
190650	117000	73650	39	1,6» [11]

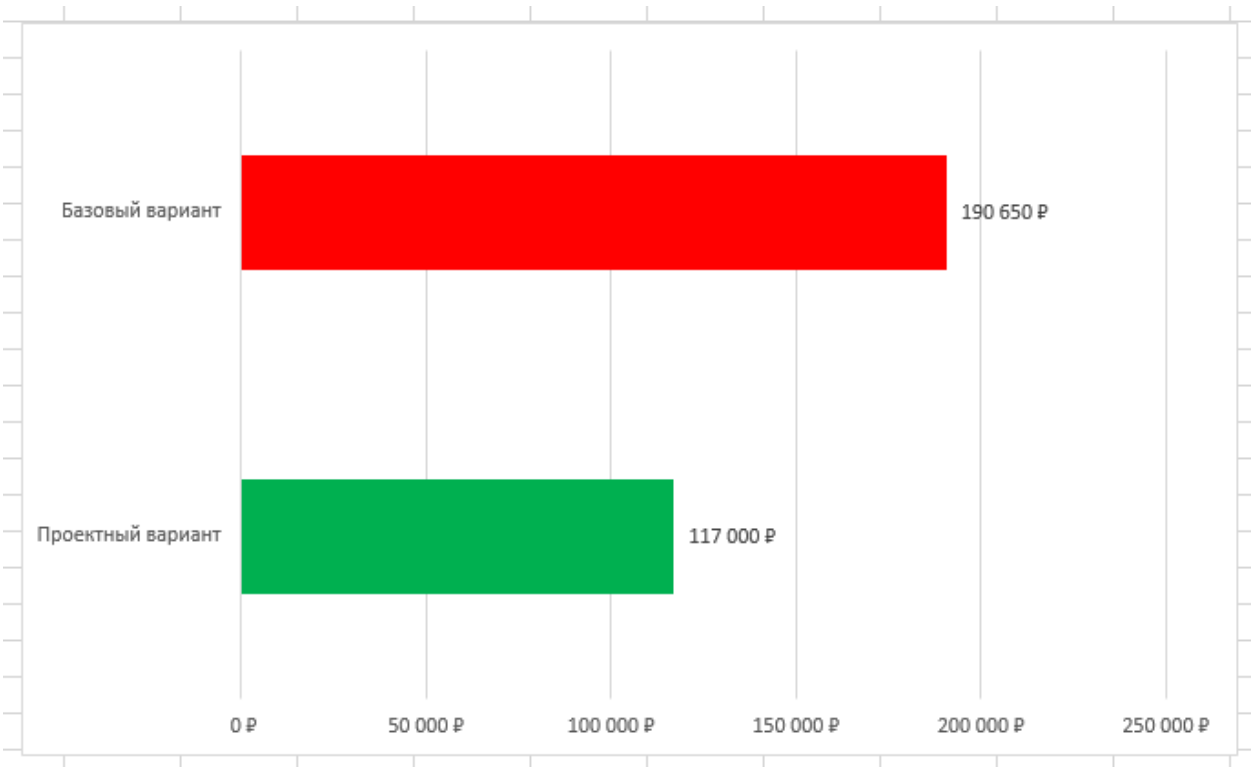

Рисунок 40 – Диаграмма сравнения затрат на разработку подсистемы анализа событий СКУД

Таким образом, затраты при проектном варианте разработки подсистемы анализа событий СКУД сократились в 1,6 раза.

«Срок окупаемости затрат на внедрение проектного решения ($Т_{ок}$) определяется по формуле (5):

$$Т_{ок} = К_{П} / \Delta С \text{ (мес.)}, \quad (5)$$

где $К_{П}$ – затраты на реализацию проектных решений (проектирование и внедрение подсистемы анализа событий СКУД).

Следовательно, срок окупаемости подсистемы анализа событий СКУД равен:

$$Т_{ок} = 117000/73650 \approx 2 \text{ мес.}$$

Представленные расчеты подтвердили существенное снижение затрат на проектирование и эффективность проектного решения» [11].

Выводы по главе 4

В результате проделанной работы были сделаны следующие выводы:

- для разработки прототипа подсистемы анализа событий СКУД использованы язык Python, среда Jupyter Notebook и библиотека scikit-learn;
- в качестве источника данных для обучения модели машинного обучения используется файл CSV, сформированный путем конвертации журнала событий СКУД, который импортируется в виде файла рабочей книги Excel;
- апробация подтвердила работоспособность прототипа подсистемы анализа событий СКУД.

Расчеты подтвердили высокую функциональную и экономическую эффективность подсистемы анализа событий СКУД, разработанной на основе предлагаемых моделей и алгоритмов.

Заключение

Одной из основных функций современной СКУД является обеспечение анализа зарегистрированных событий, ключевой стадией которого является аналитическая обработка событийной информации, собранной за определенный период.

На основе результатов анализа событий менеджмент вуза может принять управленческие решения по обеспечению соблюдения трудового порядка сотрудниками вуза. Эффективность принятых решений зависит от качества полученных результатов анализа, которое должно обеспечиваться на стадии аналитической обработки событий СКУД.

Для решения данной задачи необходимо использовать в процессе аналитической обработки событий СКУД эффективные модели и алгоритмы.

Магистерская диссертация посвящена актуальной проблеме исследования и разработки моделей и алгоритмов аналитической обработки событий СКУД.

В процессе выполнения магистерской диссертации были решены следующие задачи:

- проведен анализ современного состояния исследований в области построения систем аналитической обработки событий СКУД. Обзор и анализ источников по теме исследования подтвердили интерес ученых и специалистов к проблеме анализа и мониторинга событий СКУД. Следует констатировать недостаточность исследований, посвященных разработке моделей и алгоритмов аналитической обработки событий СКУД, что подтверждает актуальность темы исследования магистерской диссертации. Как показали обзор и анализ готовых ИТ-решений для аналитической обработки событий СКУД, последние главным образом ориентированы на решение задач контроля рабочего времени сотрудников организаций. Анализ показал, что представленные ИТ-решения разработаны под

конкретные модели СКУД, что ограничивает их функциональные возможности. Необходимо отметить, что в описаниях к рассмотренным ИТ-решениям отсутствуют сведения о моделях и алгоритмах, положенных в их основу;

- произведен анализ методов и технологий построения аналитической обработки событий СКУД. Как показал анализ, в зарубежных источниках системы контроля и управления доступом СКУД рассматриваются как объекты Интернета вещей (IoT), для анализа данных которых используются методы интеллектуального анализа данных (Data mining). Интеллектуальный анализ данных заключается в обнаружении скрытых и непредсказуемых взаимосвязей между данными путем обнаружения закономерностей данных, извлечения знаний и раскрытия неизвестной информации. Использование методов интеллектуального анализа данных на основе технологий МО для анализа событий СКУД приобретает важность в настоящее время для прогнозного анализа подозрительных попыток проникновения в помещения организации;
- разработаны модели и выбраны алгоритмы аналитической обработки событий СКУД. Для отражения основных аспектов системы аналитической обработки событий СКУД выполнено ее логическое проектирование. В процессе логического моделирования построены диаграммы отражающие функциональный, структурный и динамический аспекты подсистемы анализа событий СКУД. На основании результатов анализа характеристик алгоритмов для сегментации событий СКУД выбраны метод кластеризации и алгоритм k-means как наиболее простой для реализации и визуализации. Представленные модели и алгоритмы являются основой для построения подсистемы аналитической обработки событий СКУД;
- выполнены апробация и оценка эффективности проектных решений.

Для разработки прототипа подсистемы анализа событий СКУД использованы язык Python, среда Jupyter Notebook и библиотека scikit-learn. В качестве источника данных для обучения модели машинного обучения используется файл CSV, сформированный путем конвертации журнала событий СКУД, который импортируется в виде файла рабочей книги Excel. Апробация подтвердила работоспособность прототипа подсистемы анализа событий СКУД.

Расчеты подтвердили высокую функциональную и экономическую эффективность подсистемы анализа событий СКУД, разработанной на основе предлагаемых моделей и алгоритмов.

Таким образом, была решена проблема исследования и разработки моделей и алгоритмов аналитической обработки событий СКУД.

Гипотеза исследования подтверждена.

Работа может представлять интерес для бизнес-аналитиков и разработчиков систем аналитической обработки событий СКУД.

Список используемой литературы и используемых источников

1. Анализируем посещаемость сотрудников по данным СКУД [Электронный ресурс]. URL:https://habr.com/ru/articles/253493/ (дата обращения: 02.05.2024).

2. Аналитика по сотрудникам в подсистеме PROSTO: СКУД [Электронный ресурс]. URL: https://infostart.ru/1c/articles/1925738/ (дата обращения: 02.05.2024).

3. Батманов О., Куляс М., Суконщиков Ю. Какова роль СКУД в составе ИСБ? Взгляд разработчиков [Электронный ресурс]. URL: http://www.techportal.ru/189399 (дата обращения: 02.05.2024).

4. Вдовин В.М., Суркова Л.Е., Шурупов А.А. Предметно-ориентированные экономические информационные системы. М.: Дашков и К, 2016. 388 с.

5. Интеграция RusGuard и БИТ. Управление доступом (СКУД) 8 [Электронный ресурс]. URL: https://www.rgsec.ru/integration/corporate-systems/1c-bit (дата обращения: 02.05.2024).

6. Котов К., Красильников Н. Кластеризация данных [Электронный ресурс]. URL: https://logic.pdmi.ras.ru/~yura/internet/02ia-seminar-note.pdf (дата обращения: 02.05.2024).

7. Леоненков А. В. Объектно-ориентированный анализ и проектирование с использованием UML и IBM Rational Rose : учебное пособие. М. : ИНТУИТ, Ай Пи Ар Медиа, 2020. 317 с. [Электронный ресурс]. URL: https://www.iprbookshop.ru/97554.html (дата обращения: 25.02.2024).

8. Модуль формирования отчетности ParsecNET [Электронный ресурс]. URL: https://www.parsec.ru/support/kb/parsecnet3/reports/8821 (дата обращения: 02.05.2024).

9. Обзор алгоритмов кластеризации данных [Электронный ресурс]. URL: https://habr.com/ru/articles/101338/ (дата обращения: 02.05.2024).

10. Обработка событий в СКУД [Электронный ресурс]. URL:

https://kodos.ru/capabilities/obrabotka-sobytij/ (дата обращения: 02.05.2024).

11. Поршкевич Н.Ю., Огнева А.Ю., Чинчукова Е.П. Оценка экономической эффективности применения информационных систем // Экономика и социум. 2016. №7 (26). URL: https://cyberleninka.ru/article/n/otsenka-ekonomicheskoy-effektivnosti-primeneniya-informatsionnyh-sistem (дата обращения: 20.03.2024).

12. Программный продукт «Болид: СКУД и УРВ для 1С:Предприятие 8» [Электронный ресурс]. URL: https://bolid.ru/production/urv1c/urv1c8.html#download (дата обращения: 02.05.2024).

13. Сколько стоят услуги программистов? Цены студий и фрилансеров Источник: https://www.kadrof.ru/articles/46641 (дата обращения: 20.03.2024).

14. СКУД в высших учебных заведениях [Электронный ресурс]. URL: https://nppohrana.ru/stati/skud-v-vysshikh-uchebnykh-zavedeniiakh (дата обращения: 02.05.2024).

15. СКУД: принцип работы, виды и преимущества для компании [Электронный ресурс]. URL: https://sky-dynamics.ru/stati/skud-princip-raboty-vidy-i-preimushhestva-dlya-kompanii/ (дата обращения: 02.05.2024).

16. Создание схем компонентов UML [Электронный ресурс]. URL: https://support.microsoft.com/ru-ru/topic/%D1%81%D0%BE%D0%B7%D0%B4%D0%B0%D0%BD%D0%B8%D0%B5-%D1%81%D1%85%D0%B5%D0%BC-%D0%BA%D0%BE%D0%BC%D0%BF%D0%BE%D0%BD%D0%B5%D0%BD%D1%82%D0%BE%D0%B2-uml-aa924ecb-e4d2-4172-976e-a78fa157b074 (дата обращения: 02.05.2024).

17. Технология и методы Data Mining [Электронный ресурс]. UML: https://trends.rbc.ru/trends/industry/6634bec99a79472903630d92 (дата обращения: 02.05.2024).

18. Access control system through Internet of Things [Электронный ресурс]. URL: https://nexusintegra.io/access-control-system/ (дата обращения:

02.05.2024).

19. Access Control Systems: The Complete Guide [Электронный ресурс]. URL: https://getsafeandsound.com/access-control-system-buyers-guide/ (дата обращения: 02.05.2024).

20. Activity: Diagram Advantages & Disadvantages [Электронный ресурс]. URL: https://websitesium.com/activity-diagram-advantages-and-disadvantages/ (дата обращения: 02.05.2024).

21. Ahmed M. Mohamed, Hosny A. Abbas “Efficient Web-based Monitoring and Control System”, ICAS 2011 : The Seventh International Conference on Autonomic and Autonomous Systems. P. 18-23.

22. Anita Rajendra Zope et al. Data Mining Approach in Security Information and Event Management, International Journal of Future Computer and Communication, Vol. 2, No. 2, P. 80-84.

23. Aodi Liu, Xuehui Du, Na Wang, “Efficient Access Control Permission Decision Engine Based on Machine Learning”, Security and Communication Networks, vol. 2021, Article ID 3970485, 11 pages, 2021. https://doi.org/10.1155/2021/3970485 (дата обращения: 02.05.2024).

24. Box and Whisker Plot [Электронный ресурс]. URL: https://datavizcatalogue.com/methods/box_plot.html (дата обращения: 02.05.2024).

25. Brown S. Machine learning, explained. URL: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained (дата обращения: 02.05.2024).

26. Chen H. Adaptive K-means clustering algorithm [Электронный ресурс]. URL: https://www.spiedigitallibrary.org/profile/Chen.Hailin-86165 (дата обращения: 02.05.2024).

27. Data Mining: Simple Definition, Uses & Techniques [Электронный ресурс]. URL: https://www.statisticshowto.com/data-mining/ (дата обращения: 02.05.2024).

28. Data Preparation for Machine Learning: The Ultimate Guide to Doing It

Right [Электронный ресурс]. URL: https://www.pecan.ai/blog/data-preparation-for-machine-learning/ (дата обращения: 02.05.2024).

29. Essential Data Mining Techniques [Электронный ресурс]. URL: https://www.dataversity.net/15-essential-data-mining-techniques/ (дата обращения: 02.05.2024).

30. Event processing systems [Электронный ресурс]. URL: https://ifs.host.cs.st-andrews.ac.uk/Books/SE9/Web/Architecture/AppArch/EventProc.html (дата обращения: 02.05.2024).

31. Evgeniou T. Cluster Analysis and Segmentation. URL: https://inseaddataanalytics.github.io/INSEADAnalytics/CourseSessions/Sessions45/ClusterAnalysisReading.html (дата обращения: 02.05.2024).

32. F. Ribeiro, J. C. Metrôlho, E. R. Lopes “Web-CAN Interface for Access Control and Monitoring”, Departamento de Engenharia Informática, Escola Superior de Tecnología de Castelo Branco. URL: https://repositorio.ipcb.pt/bitstream/10400.11/579/1/2003-09%20SAAEI.pdf (дата обращения: 02.05.2024).

33. Heatmap (Matrix) [Электронный ресурс]. URL: https://datavizcatalogue.com/methods/heatmap.html (дата обращения: 02.05.2024).

34. Nurul Huda Nik Zulkipli and Gary B.Wills, “An Event-based Access Control for IoT”, Conference Paper · March 2017 DOI: 10.1145/3018896.3025170 (дата обращения: 02.05.2024).

35. Project Jupyter [Электронный ресурс]. URL: https://jupyter.org/ (дата обращения: 15.03.2024).

36. Rational Rose [Электронный ресурс]. UML: https://www.kpms.ru/Automatization/Rational_Rose.htm (дата обращения: 02.05.2024).

37. Scikit-learn [Электронный ресурс]. URL: https://blog.skillfactory.ru/glossary/scikit-learn/ (дата обращения: 15.03.2024).

38. UML Class Diagram Tutorial [Электронный ресурс]. URL: https://www.lucidchart.com/pages/uml-class-diagram (дата обращения: 02.05.2024).

39. What is access monitoring? [Электронный ресурс]. URL: https://www.imprivata.com/blog/what-is-access-monitoring (дата обращения: 02.05.2024).

40. What is Data Mining? [Электронный ресурс]. URL: https://economictimes.indiatimes.com/definition/data-mining (дата обращения: 02.05.2024).