Н.Ю. Прокопенко, Г.С. Артюх

ПРИМЕНЕНИЕ ТЕХНОЛОГИИ PROCESS MINING ДЛЯ АНАЛИЗА ДАННЫХ И ПРОЦЕССОВ

(на платформе Loginom Community)

Учебное пособие

Нижний Новгород
2025

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Нижегородский государственный архитектурно-строительный университет»

Н. Ю. Прокопенко, Г.С. Артюх

ПРИМЕНЕНИЕ ТЕХНОЛОГИИ PROCESS MINING ДЛЯ АНАЛИЗА ДАННЫХ И ПРОЦЕССОВ
(на платформе Loginom Community)

Утверждено редакционно-издательским советом университета
в качестве учебного пособия

Нижний Новгород
ННГАСУ
2025

УДК 004.89(075.8)
П 78
А 86
ББК 32.813я73

Рецензенты:

Н.Б. Паклин – канд. тех. наук, руководитель электронного учебного центра ООО «Аналитические технологии»

И.Н. Цветкова – канд. физ.-мат. наук, доцент кафедры информатики и информационных технологий ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации»

Прокопенко, Н.Ю. Применение технологии Process Mining для анализа данных и процессов (на платформе Loginom Community) : учебное пособие / Н.Ю. Прокопенко, Г.С. Артюх ; Министерство образования и науки Российской Федерации, Нижегородский государственный архитектурно-строительный университет. – Нижний Новгород : ННГАСУ, 2025. – 159 с. – ISBN 978-5-528-00617-8. – 1 CD ROM. – Текст : электронный.

В учебном пособии приводятся теоретические основы интеллектуального анализа процессов на основе данных из журналов событий, доступных в информационных системах, а также инструкции по выполнению практических работ с использованием аналитической low-code платформы Loginom. В заданиях практикума представлены кейсы из области ипотечного кредитования, медицины, электронного обучения, описана последовательность выполнения основных этапов методологии Loginom Process Mining.

Пособие предназначено для подготовки студентов бакалавриата и магистратуры, обучающихся по направлениям подготовки бакалавриата 09.03.03 Прикладная информатика (профиль: прикладная информатика в экономике) и магистратуры 09.04.03 Прикладная информатика (профиль: искусственный интеллект в бизнес-аналитике) при изучении дисциплин «Системы поддержки принятия решений», «Анализ и обработка данных», «Методы бизнес-аналитики» и «Бизнес-аналитика в практике предприятий», пособие также может быть использовано при написании выпускных квалификационных работ. Пособие направлено на развитие аналитических навыков, возможность аналитику ориентироваться в задачах и быстро двигаться к выявлению существенных проблем при анализе бизнес-процессов.

ISBN 978-5-528-00617-8

Оглавление

Введение

Повсеместное развитие цифровых технологий, роботизация и автоматизация, интернет вещей, глобальные сети и массовая цифровизация дает основу для качественно нового этапа исследования и оптимизации процессов, реализации возможностей искусственного интеллекта и машинного обучения.

В программных средствах, автоматизирующих бизнес-процессы, вся информация о выполненных экземплярах процесса записывается в протоколы работы информационных систем. За все время работы программных средств протоколы работы накапливают большой объем информации о реальных процессах, выполняемых в компании. Безусловно, данная информация является ценной, а ее анализ позволяет извлечь новые знания о бизнес-процессах. Для этого к регистрационным журналам применяются адаптированные методы Data Mining. Применение методов Data Mining для анализа информации о реальных процессах, выполняемых системами, автоматизирующими бизнес-процессы, получило в литературе название Process Mining.

Создателем концепции Process Mining является Вил ван дер Аалст (Wil van der Aalst) – голландский ученый, профессор математики и информатики технического университета Эйндховена. Именно он является автором фундаментального труда «Process Mining. Discovery, Conformance and Enhancement of Business Processes», который определил ключевые возможности, сферы применения и перспективы анализа бизнес-процессов. [1]

Сам термин «mining» применительно к моделям процессов и журналам событий впервые был введен Agrawal и др. [2] Согласно манифесту Process Mining [3], лозунгом дисциплины является обнаружение (discover), отслеживание (monitor) и улучшение (improve) реальных процессов в компаниях, у которых хорошо развита ИТ-инфраструктура и где основные процессы заложены в ИТ-систему, чтобы можно было отследить цепочку действий и восстановить картину происходящего.

Process Mining – интеллектуальный анализ процессов, мост между Data Mining и Process Management, это подход к извлечению, анализу и оптимизации процессов на основе данных из журналов событий (event logs), доступных в информационных системах.

Суть данного подхода в том, что в современном бизнесе, активно проходящем цифровую трансформацию, большинство действий клиентов и сотрудников оставляют цифровой след, и подход Process Mining позволяет анализировать эффективность процессов только на основе этой объективной информации. Process Mining включает в себя автоматическое обнаружение процессов (извлечение информации из логов – журналов событий и восстановление моделей процессов), поиск узких мест в бизнес-процессах, проверку соответствия (мониторинг отклонений путем сравнения моделей бизнес-процессов и журналов событий), поиск коротких путей выполнения бизнес-процесса, поведенческий анализ, анализ соблюдения требований, бенчмакинг (сравнение с лучшими практиками), рекомендации по улучшению процессов, анализ «что-если», выявление «узких мест», выявление лишних шагов, поиск нетипичного поведения, моделирование и стресс-тестирование, а также построение имитационных моделей для целей прогнозирования проблем в бизнес-процессах.

Process Mining обеспечивает наглядность и понимание реальных бизнес-операций и процессов, применяя набор алгоритмов к событиям, что приводит к адаптируемым, легко обслуживаемым и проверенным моделям процессов. Помимо выявления неэффективности процесса Process Mining дает представление о том, где улучшить операции, например, в рамках инициативы по цифровизации, и как достичь целевых результатов в бизнесе. Таким образом, поддерживая эффективность и результативность процессов, инструменты интеллектуального анализа процессов являются ключевыми инструментами инициатив по улучшению процессов и связанных с ними дисциплин.

Раздел 1. Описание методологии Process Mining

1.1. Глубинный анализ процессов [Process Mining]

Process Mining (PM) – технологии восстановления реального протекания бизнес-процессов на основании протоколов (логов) информационных систем, а также практики, методы и инструменты извлечения, мониторинга и улучшения существующих бизнес-процессов.

Process Mining – исследовательская дисциплина, объединяющая набор практик, методов и подходов для автоматического построения моделей процессов, проверки соответствия заданной модели некоторому процессу и улучшению процессов на основе информации, хранящейся в журналах событий.

Process Mining находит широкое применение во многих областях бизнеса и управления, так как, в отличие от обычного моделирования, получаемые с помощью алгоритмов Process Mining модели строятся на основе записей о реально происходивших событиях, поэтому более адекватно отображают действительность и несут информацию о том, что происходило в реальности, а не о том, что было запланировано. Кроме того, в некоторых областях моделирование может быть затруднено или очень трудоемко, тогда как Process Mining позволяет автоматизировать процесс.

Process Mining применяется в различных отраслях экономики:

- Финансы
- Дистрибуция
- Ритейл
- Промышленность
- Телекоммуникации
- Медицина

Описание методологии Process Mining представлено на рис. 1.

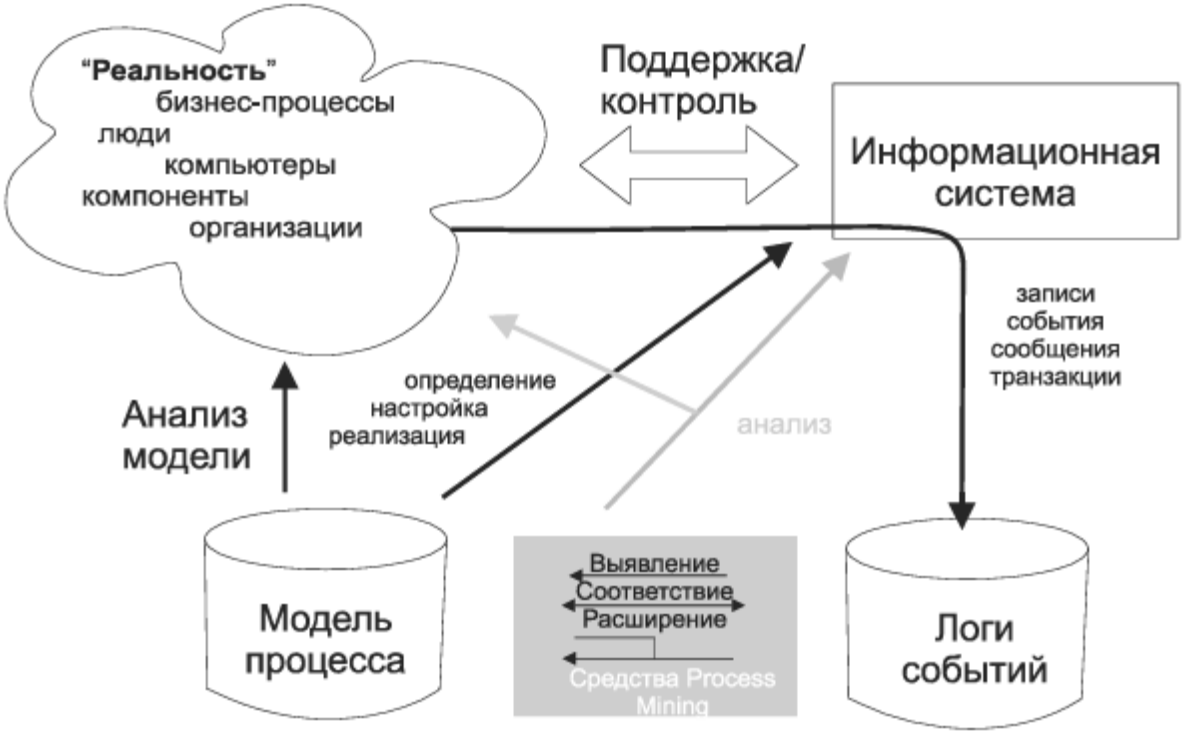


Рис. 1. Описание методологии Process Mining [4]

Process Mining решает следующие задачи:

1. Аудит соответствия – определение, насколько реальный процесс соответствует запроектированному, какие есть отклонения и типовые ошибки.

2. Операционная эффективность – оценка производительности, затрат и стоимости процесса, этапа или конкретных исполнителей.

3. Управление изменениями – сравнение пути процесса «до» и «после» внесенных изменений по метрикам операционной эффективности.

4. Повышение качества обслуживания клиентов – анализ путей взаимодействия клиента с системами компании для улучшения клиентского опыта.

Существует четыре основных подхода к проведению Process Mining.

Подход, направленный на *control-flow* процесса, основывается на анализе последовательности событий. Цель данного подхода – описать все возможные пути процесса. Результат, как правило, описывается с помощью сетей Петри или других нотаций (например, EPC, BPMN или UML).

Организационный подход направлен на анализ данных, скрытых в журнале событий. Например, какие пользователи, роли, департаменты задей-

ствованы в процессе и как они связаны между собой. Цель – структурировать пользователей благодаря их классификации по ролям или организационным подразделениям, или же предоставление сети их взаимоотношений.

Также существует подход, основанный на отдельном внимании к *уникальным цепочкам событий* и их характеристикам. Цель – определить уникальный путь, пользователей, задействованных в процессе его выполнения.

Подход, сконцентрированный на *времени*, уделяет особое внимание длительности и частоте событий. Когда события содержат метки времени, можно обнаружить узкие места, измерить уровни обслуживания, отследить использование ресурсов и прогнозировать оставшееся время обработки текущей цепочки.

Максимальный эффект от внедрения PM достигается, если речь идет:

- о простых массовых процессах, которые исполняются несколько тысяч раз ежедневно;
- трудных и долгих индивидуальных процессах, в которых задействовано много филиалов компании;
- длинных процессах из нескольких этапов, в которые вовлечено много участников.

Преимущества Process Mining:

- Автоматизированное восстановление модели бизнес-процессов на базе лог-файлов.
- Оптимизация расходов. Технологии Process Mining позволяют компаниям существенно сократить затраты путём автоматизации процессов и устранения лишних этапов.
- Возможность анализа бизнес-процесса до уровня отдельного экземпляра (негативный путь).
- Повышение качества сервиса. Инструменты Process Mining помогают выявить «узкие горлышки» бизнес-процесса – те операции, для выполнения которых не хватает ресурсов, и определить направления для оптимиза-

ции разных процессов, что приводит к сокращению общего времени производственного цикла.

- Позволяет провести анализ отклонений: выделение экземпляров процесса, отличающихся от эталонных и проанализировать те экземпляры процесса, выполнение которых длится намного дольше остальных.
- Позволяет искать наиболее успешные экземпляры процесса по различным параметрам, например, по времени от поступления заявки клиента до момента поступления денег на счет.
- Позволяет сверить регламентные и фактические сроки исполнения процесса даже после тестовой выгрузки данных.
- Возможность фактами доказать руководству неэффективность существующих бизнес-процессов на основе фактов.
- Возможность регулярного мониторинга процессов. Преимущества в соблюдении требований. Программы Process Mining способны в режиме реального времени выявлять процессы, не соответствующие требованиям, и оперативно уведомлять об этом руководство компании.

1.2. Этапы Process Mining

Методология Process Mining-а предполагает выполнение определенных этапов и операций.

1. Discovery (обнаружение): фиксация операций в логах, сбор и консолидация данных.

2. Conformance checking (проверка соответствия): расчет характеристик и метрик.

3. Enhancement (улучшение): воссоздание реального процесса, анализ, формирование выводов.

4. Monitoring (отслеживание): перепроектирование и моделирование, мониторинг процессов.

5. Process Mining и Machine Learning

Этап 1. Discovery – обнаружение реального процесса. [ETL & Data Preparation]. Логирование, сбор данных, входной аудит и подготовка данных.

Цель начального этапа Process Mining – собрать воедино все данные, фиксирующие фактическое течение процесса, оценить их полноту и качество, рассчитать базовые метрики и показатели, характеризующие текущий процесс.

Сбор и консолидация данных включает следующие этапы:

- подготовка данных (ETL);
- объединение из разных источников, подключение справочников;
- консолидация данных по всем логам;
- обогащение данных дополнительными атрибутами;
- извлечение сущностей;
- входной аудит;
- валидация данных.

На выходе – данные, соответствующие формализованным требованиям, гарантирующим корректную аналитику.

В основе Process Mining лежит извлечение понимания хода бизнес-процессов на основании журналов событий (цифровых следов, логов), кото-

рые массово доступны в современных информационных системах. Данные могут поступать из любых информационных систем: 1С, MySQL, Excel, CSV, файлы логов доступа к данным веб-сайтов, подойдут любые системы, которые собирают логи событий.

Оптимальный временной диапазон, с которыми следует работать в анализе (период выгрузки логов) можно вычислить как α*среднюю длительность экземпляра процесса ($\alpha \geq 100$). Такой диапазон позволяет иметь достаточную статистическую базу для расчёта метрик процессов и снижает долю неполных процессов в выгрузке (процессов без первых шагов, которые начались до момента начала периода выгрузки и без последних шагов для процессов, которые закончились после момента выгрузки).

Например, если вы эмпирически оцениваете среднюю длительность экземпляра процесса в один час, то желательный период выгрузки логов будет 100–150 часов. При этом оценочная доля неполных процессов может составлять 2%–1,5%.

Это позволит избежать лишних расходов на обслуживание системы бизнес-аналитики, а полученный результат можно спроецировать на остальной процесс.

Process Mining – это технология анализа процессов путем восстановления модели процесса через анализ набора цепочек событий исполнения этого процесса. Методология Process Mining включает в себя использование данных из различных процессно- и не процессно-ориентированных систем для построения и анализа реальной картины хода процесса. Помимо данных о названии и дате выполнения определенного события цепочки Process Mining также предполагают использование дополнительной информации для получения более полного и разностороннего анализа.

Цель Process Mining – автоматическое формирование точного представления о реальном ходе процесса. Журнал событий (event log) и модель процесса – это два ключевых артефакта, используемых в Process Mining. Журнал событий хранит информацию о датах выполнения событий процесса,

записанных информационной системой. Модель процесса используется для визуализации.

Журнал событий. Методы Process Mining предполагают, что данные исполнения процесса хранятся в журнале событий. В журнале событий информация хранится как последовательность событий.

Модель процесса. Модели процесса используются для описания поведения процесса организации, для достижения ряда различных целей: установления взаимосвязи между стейкхолдерами, усовершенствования процесса, управления процессом, автоматизации процесса, поддержки исполнения процесса. Чаще всего модель процесса используется для сравнения as-is процесса с to-be процессом, проверки соответствия процесса нормативным требованиям (как, например, ISO 9001), анализа проблем, связанных с производительностью, узких мест (bottlenecks) и других неэффективностей.

Журнал событий – это перечень упорядоченных во времени событий (действий, статусов), выполняемых в рамках реализации некоего бизнес-процесса.

Примеры источников журналов событий:

- информационные системы;
- сетевое оборудование;
- трекинг посетителей сайтов;
- устройства интернета вещей;
- мониторинг действий сотрудников на своих АРМ.

В качестве примеров бизнес-процессов можно назвать:

- процесс обработки кредитной заявки (услуги);
- процесс лечения пациента в отделении неотложной помощи (медицинское обслуживание);
- процесс изготовления автомобиля (производство).

В рамках одного процесса может осуществляться множество различных событий и изменений. При этом по одному процессу могут параллельно осуществляться несколько цепочек – экземпляров бизнес-процессов.

Обычно, выполнение каждого экземпляра бизнес-процесса сопровождается выполнением идущих друг за другом действий или функций (activities) и возникновением событий (events). Каждое событие, как правило, несет информацию о:

- дате исполнения (timestamp);
- имени или идентификаторе исполненного действия (activity name).

Например, в рамках одного экземпляра бизнес-процесса могут встречаться следующие события:

- утверждение заявки на кредит;
- проверка кредитного рейтинга;
- установление очерёдности медицинской помощи для пациента;
- заказ недостающей части для производства товара.

В рамках выявления процесса необходимо:

- отфильтровать шум (редкие и ошибочные события) от регулярных событий;
- выявить корректные поведенческие связи между событиями несмотря на неполное наблюдение за процессом.

Методология Process Mining-а предполагает, что можно последовательно записывать события так, чтобы каждое из них в дальнейшем поставить в соответствие с четко определенным шагом в некотором бизнес-процессе. Помимо событий, из журналов необходимо по возможности извлекать дополнительную информацию о том, кто какое действие выполнил и в какое время, а также связанные с этим действием данные, например наименование клиента или сумму платежа.

Необходимые требования для журнала логов:

- все события, записанные в протоколе, должны быть идентифицированы с экземплярами процессов;
- все события должны быть упорядочены по времени их выполнения;
- разнотипные события должны различаться.

Бизнес-операции фиксируются в журналах логов событий информационных систем. Данные в таких журналах включают, как правило, следующие атрибуты:

– идентификатор экземпляра процесса (Process ID, case ID). Характер ID зависит от специфики процесса, например это может быть:

- номер заявки, привязанный к обрабатываемому заказу;
- идентификатор сделки;
- идентификатор пациента для отслеживания лечения в логах медицинского учреждения, и т. д.

Это позволяет алгоритмам расставить действия по порядку и отнести их к конкретным экземплярам процессов, отследить прохождение каждого экземпляра процесса от начала и до конца даже если журнал не отсортирован или содержит ошибки.

– идентификатор события (event ID). Уникальный номер для каждого действия, обычно нумеруется последовательно по мере запуска действия;

– наименование операции – наименование события (Event);

– временная метка (Timestamp). Фиксируется точное время и дата, когда было совершенно действие;

– действие или мероприятие (Activity). Краткое описание действия;

– исполнитель «Name». (Performed by/Performer). Обычно фиксируется ответственное лицо непосредственно инициирующее действие.

Если есть дополнительные атрибуты, например исполнители, регионы или типы заявок, то на этапе разведочного анализа можно предложить больше гипотез.

В табл. 1 приведен фрагмент протокола работы информационной системы с базовыми полями.

Таблица 1. Фрагмент журнала событий с базовыми полями

Process ID	**Event**	**Timestamp**
1	Приёмка Ж/Д терминал	03.06.2024 13:59:51
1	Оформление приходных документов	03.06.2024 14:05:51
1	Оформление документов контрагентов	03.06.2024 14:26:42

1	Ввод данных контрагента	03.06.2024 14:35:09
1	Оформление приходных документов	04.06.2024 9:47:40
1	Заменить	04.06.2024 14:05:15
1	Оформление приходных документов	04.06.2024 14:05:28
1	Заменить	05.06.2024 8:28:17
1	Определение локации	05.06.2024 8:28:35
1	Вызов погрузчика	05.06.2024 8:29:41
1	Размещение груза	05.06.2024 13:40:53
1	Груз размещён	05.06.2024 13:41:01
2	Приёмка Ж/Д терминал	04.06.2024 2:28:58
2	Оформление приходных документов	04.06.2024 3:54:28
2	Груз размещён	04.06.2024 5:09:07
3	Приёмка	04.06.2024 3:02:31
3	Ввод данных контрагента	04.06.2024 4:45:41

Логи могут содержать ошибки (шум): пропуск некоторых шагов, изменение порядка действий, отсутствие начала или конца цепочки. Некорректные цепочки можно просто отбросить. Но всегда может оказаться, что то, что считают шумом в данных, на самом деле – редкие процессы. Возможно, что процесс на самом деле исполнялся с ошибками.

Исходя из таких соображений, лучше фильтровать события по частоте и отдельно анализировать редкие цепочки, а не отбрасывать их как шум.

Другая сложность состоит в том, что логи могут быть неполными, то есть не включать сведения обо всех возможных экземплярах процессов, их вариантов протекания. При параллельном выполнении действий количество вариантов растет со скоростью факториала. То есть процесс, состоящий из 10 действий, которые выполняются параллельно, имеет 10! = 3628800 вариантов выполнения. Конечно маловероятно, что каждый экземпляр такой цепочки встретится в отдельно взятом журнале событий, и, чтобы обобщить множество таких процессов в один, нужны эвристические методы.

Существует несколько критериев для оценки качества данных о событиях. События должны быть *достоверными*, т. е. надо быть уверенным в том, что зарегистрированные события действительно произошли и что характеристики событий корректны. Журналы событий должны быть *полными*, т. е., с учетом конкретного контекста, ни одно событие не должно быть пропущено.

Каждое зарегистрированное событие должно иметь четко определенную семантику.

Таблица 2 описывает пять уровней зрелости журналов событий, варьирующихся от великолепного качества (Уровень 5) до низкого качества (Уровень 1). [3]

Таблица 2 – Уровни качества журнала логов

Уровень	Характеристика
Уровень 5	Наивысший уровень: журнал событий великолепного качества (достоверный и полный), и события четко определены. События регистрируются в автоматическом, систематическом, надежном и безопасном режиме. Надлежащим образом обеспечены конфиденциальность и безопасность. Регистрируемые события (и все их атрибуты) обладают понятной семантикой, что подразумевает наличие одной или нескольких онтологий, а события и атрибуты указывают на эту онтологию.
Уровень 4	События регистрируются автоматически и в систематическом и надежном режиме, т. е. логи достоверны и полны. В отличие от систем, оперирующих на уровне 3, явным образом поддерживаются такие сущности, как экземпляр процесса (кейс) и действие.
Уровень 3	События регистрируются автоматически, но нет системного подхода к регистрации событий. Однако, в отличие от журналов уровня 2, есть некоторый уровень гарантии того, что регистрируемые события соотносятся с реальностью (журнал событий достоверен, но не обязательно полон). Рассмотрим, например, события, регистрируемые в системе ERP. Несмотря на то, что приходится извлекать события из многочисленных таблиц, информация в них, предположительно, корректна (например, наверняка можно предположить, что платеж, зарегистрированный в ERP, действительно существует, и наоборот).
Уровень 2	События регистрируются автоматически в режиме «побочного продукта» какой-либо информационной системы. Степень покрытия различна, нет системного подхода для принятия решений о том, какие события следует регистрировать. Кроме того, есть возможность обходить информационную систему. Следовательно, события могут отсутствовать, либо они могут быть неправильно зарегистрированы.
Уровень 1	Низший уровень: журналы событий низкого качества. Зарегистрированные события могут не соответствовать реальности, часть событий может отсутствовать.

	Обычно такими характеристиками обладают журналы событий, в которых события регистрируются вручную.

Методы Process Mining применимы к журналам логов уровня 5, 4 и 3. Принципиально методы Process Mining применимы и с журналами уровня 1 или 2, однако, анализ таких логов обычно проблематичен, а результаты недостоверны.

Этап 2. Conformance checking – проверка соответствия.

Автоматическое создание карты процессов. На основе собираемых данных строится карта рабочих процессов. Построенная карта позволяет увидеть эффективность операций внутри процесса – отклонения от регламентов и соответствия стандартам.

Цель данного этапа – определить, в какой мере фактический процесс совпадает с эталонным, выявить критические отклонения, мешающие запланированному течению процедур, провести своеобразное сравнение «ожидание/реальность».

Основные задачи проверки соответствия:

- Увидеть «реальное» состояние бизнес-процесса в формате графической или аналитической модели;
- Сравнить «реальное» состояние и целевое состояние, зафиксированное в регламентах;
- Найти «узкие места» в бизнес-процессах;
- Выделить для последующего выборочного анализа «экстремальные» экземпляры процесса;
- Найти инсайты по процессам (длительные или дорогие процессы; наиболее значимые процессы, типовые шаблоны (паттерны) процессов);
- Проанализировать процессные KPI;

- Проверить соответствие фактических параметров процесса Service Level Agreement (SLA) стандартам, заложенным в проекте (в первую очередь проверка длительности процессов).

На данном этапе реализуется проверка соответствия реального процесса эталонному, регламентированному в той или иной компании. Первоначально происходит воссоздание реального процесса в соответствии со следующим алгоритмом:

- выявление фактической, а не «экспертно-идеальной» последовательности действий;
- формирование «слова процесса»;
- обнаружение повторяющихся и стандартных операций;
- выявление процента соответствия процессов эталонному пути;
- обнаружение «счастливых путей» – последовательностей событий, чаще других приводящих к желаемому результату;
- поиск шаблонов поведения: циклы, «пинг-понг» и пр.

Расчет характеристик и метрик:

- метрики времени – определение чистого времени работы или трудозатрат – фактической длительности событий, простоя, переработки, сверхнормативных переработок и пр.;
- значения по графу – выявление зацикливаний (rework), «пинг-понгов» между этапами, лишних звеньев процесса;
- показатели производительности – вычисление нагрузки на каждого сотрудника, среднего значения производительности, среднего времени очереди каждого задания;
- метрики стоимости – расчет стоимости каждой операции процесса;
- расчет потерь – определение временных потерь по каждой операции;
- метрики конкуренции – бенчмаркинг – сравнение производительности сотрудников, филиалов и пр.;

- показатели эффективности – определение выгоды (business-value) от оптимизации процессов.

Данные о длительности и особенностях исполнения процесса можно наложить на визуальную схему (граф) исполняемого бизнес-процесса. Это позволяет выявить задержки в действиях отдельных исполнителей и целых подразделений, неэффективные взаимосвязи между пользователями, зацикленности и скрытые недостатки в процессе.

Анализ путей процесса может провести дата-сайентист, изучая граф процесса. Для того, чтобы построить граф по логам процесса, в распоряжении аналитиков в настоящее время достаточное многообразие профессиональных разработок, таких как Proceset, PM4PY (Process Mining for Python), Sber Process Mining, ProM и т.д.

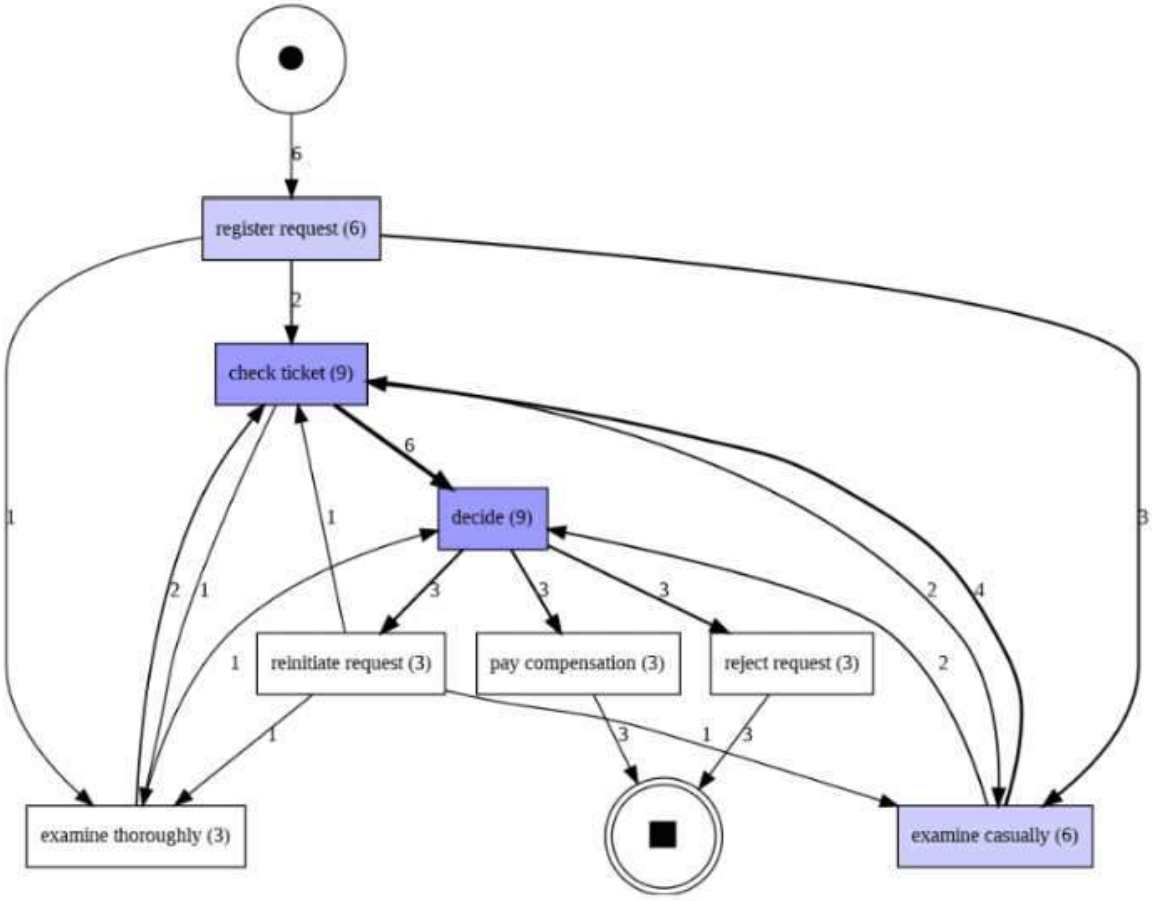


Рис. 2. Пример графа процесса с библиотекой PM4Py

В данном примере в 6 случаях запрос на регистрацию является первым наблюдаемым действием (представленным дугой, помеченной значением 6). В журнале событий действие проверки выполняется непосредственно после действия запроса регистрации. Тщательно изучить активность следует после регистрации один раз, внимательно изучить следует 3 раза. Всего за регистрационной активностью следует 6 различных событий, т. е. в журнале событий работающего примера есть 6 трассировок.

Граф хорошо читаем при малом количестве визуализируемых процессов или при визуализации основных путей процессов.

При большом количестве путей граф превращается в изображение в стиле «Spaghetti Style» и визуальная аналитика практически невозможна.

Для объёмного журнала событий анализ путей процесса рациональней проводить полностью автоматически, а специалист должен получать результаты в виде отчёта с метриками и примерами выбросов.

Граф должен использоваться как инструмент иллюстрации найденных отклонений. Он не может являться средством обнаружения проблемных зон. Основная причина – «Spaghetti Style» отображения графа. Конечно, при наложении различных фильтров на граф можно увидеть искомые нарушения последовательностей, но есть риск, что аналитик не переберет все фильтры и не составит полную картину нарушений. Работа с фильтрами требует значительного времени ручного труда и поэтому является достаточно дорогой процедурой. Таким образом, анализ отклонений и анализ соответствия должен быть выполнен полностью автоматически.

Анализ и интерпретация выявленных несоответствий зависит от типа модели:

- Для дескриптивной (описательной) модели выявленные расхождения между моделью и журналом событий указывают на необходимость улучшения самой модели.

- Для нормативной модели такие отклонения могут указывать на необходимость лучшего контроля процесса, а также не необходимость проверки неэффективных исполнителей.

Технологии проверки соответствия могут применяться при аудиторских проверках, оптимизации процессов и контроле исполнения нормативных актов.

Понимание реальной картины позволяет выявить цепочки операций, на которых надо сосредоточить внимание, чтобы снизить стоимость, увеличить скорость, сократить избыточный персонал.

Этап 3. Enhancement – улучшение

Цель этого шага – изменение и оптимизация согласно выводам, полученным на этапе Conformance checking. По результатам проведенного анализа происходит изменение процессов (устраняются «узкие места», лишние операции, тупиковые процессы) и предварительное апробирование внесенных улучшений.

Формирование выводов:

- визуализация результатов, построение дашбордов;
- формирование экспертных рекомендаций;
- прогнозирование эффекта от улучшения;
- оптимизация бизнес-процессов.

Перепроектирование и моделирование:

- предложение улучшений;
- тестирование улучшений с использованием математических моделей.

Если аналитические данные позволяют оценить стоимость выполнения этапов (например, известна стоимость человеко-часа или задержки в бизнес-процессе), то можно проводить стоимостный анализ. Например, определить, какой путь самый дорогой как для конкретного экземпляра процесса, так и по их ансамблю. Например, часто наблюдается, что наиболее редкие маршруты являются самыми дорогостоящими. Если его устранить, можно получить за-

метный эффект. Помимо стоимостных можно сравнивать KPI нескольких подмножеств одного процесса. Например, сравнить одинаковые процессы, реализованные в различных подразделениях, или процессы до и после реинжиниринга.

Этап 4. Monitoring – отслеживание

Цель этого этапа – наблюдение за правильностью течения обновленного процесса, контроллинг того, насколько задуманное соответствует полученному. Отслеживание состояния процесса «до» и «после» внесения изменений.

Фильтруя временные периоды при анализе восстановленной модели процесса, можно увидеть поведение процесса в динамике, в том числе происходящие изменения после того, как будут внедрены те или иные организационные мероприятия по его совершенствованию.

На завершающей ступени настраивается регулярный мониторинг процессов с предоставлением обратной связи о корректности реализации процедур заинтересованным пользователям:

- отслеживание ключевых показателей эффективности (KPI) по процессу с заданной частотой;
- проверка на соответствие процесса внутреннему регламенту;
- контроль и оповещение о некорректных переходах: ошибки, брак процесса, мошенничества;
- выявление ресурсоемких операций;
- оповещение владельцев процесса об отклонениях с помощью различных каналов коммуникаций.

Этап 5. (дополнительный) Process Mining и Machine Learning

Process Mining не только выявляет нежелательные изменения и отклонения, он также позволяет ответить на вопрос «почему?» – в чём причина возникших проблем и отклонений? Технологии машинного обучения дают возможность выявить глубинные причинно-следственные связи (root causes) и определить, почему процесс протекает именно таким образом.

Использование Machine Learning позволяет не просто увидеть факт, выявить «слово процесса», но понять проблему на глубинном уровне. Для Process Mining наиболее часто применяются следующие методы машинного обучения:

1. Поиск ассоциативных правил – автоматическое выявление основных и специфичных путей процессов.
2. Робастные методы – автоматическое выявление отклонений по времени, стоимости, частоте. Позволяют обнаружить резкие изменения, нивелирующиеся при агрегировании данных за длительный период.
3. Анализ временных рядов – предсказание времени выполнения процесса и допустимого разброса. Позволяет оценить границы вариативности процесса и необходимость реагирования.
4. Кластеризация позволяет выявлять сегменты событий, которые могут быть сгруппированы затем использованы для анализа и предсказания отдельных характеристик в конкретных кластерах;

1.3. Loginom Process Mining

Loginom Process Mining – модульное высокоуровневое решение автоматизированного интеллектуального анализа и оптимизации бизнес-процессов на основе «цифровых следов». [5]

Аналитическая платформа Loginom Process Mining применяется для оценки многоэтапных процессов со сложной иерархией принятия решений, с большим количеством типичных, повторяющихся операций, которые логируются информационной системой. Это позволяет восстановить фактическую, реальную модель массового бизнес-процесса, а не «экспертно-идеальную», регламентированную, игнорирующую многие варианты реализации событий.

Функционал Loginom Process Mining:

1) Discovery

- Разведочный анализ

- Расчёт процессных и кастомизированных метрик для процесса as-is
- Анализ очередей, нагрузки на процессы, загруженности сотрудников, производительности, зацикленности, длительностей и пр.
- Инструментарий для выявления «узких мест»
- Факторный анализ, использование алгоритмов машинного обучения, анализ шаблонов поведения – паттернов.

2) Conformance checking

- Проверка соответствия реального процесса проекту
- Выявление счастливых путей (happy path) в разрезе указанных аналитик (каналов коммуникации, типов клиентов и пр.)
- Определение эталонного пути, нахождение отклонений от эталона и от happy path
- Проверка бизнес-процессов на соответствие стандартам SLA (Service Level Agreement – соглашение об уровне сервиса)
- Проверка путей процесса на критические отклонения в последовательности шагов

3) Enhancement

- Определение бизнес-правил для процесса «to-be», создание виртуального журнала событий с бизнес-правилами
- Моделирование процессов «to-be», оптимизация на основе процессных метрик, what-if анализ
- Создание оптимальной математической модели «to-be», определение статистически допустимых отклонений от «идеального» пути процесса.

4) Monitoring

- Наблюдение за бизнес-процессом
- Настройка ETL процесса, регулярный импорт журнала событий
- Отслеживание KPI по процессу с заданной частотой
- Оповещение владельцев процесса об отклонениях с помощью различных каналов коммуникаций (мессенджеры, e-mail и пр.)

Используя готовые компоненты библиотеки Loginom Process Mining: метрики времени, слова процессов, характеристики событий и процессов, определение счастливого пути, анализ отклонений и другие на рисунке 3 показан сценарий «Типовой кейс воссоздания реального процесса».

В сценарий можно добавить фильтрацию удаления незавершённых событий и событий, начало которых в датасете отсутствует, рассчитать аналитику по событиям, по подразделениям, по пользователям и по любой их комбинации.

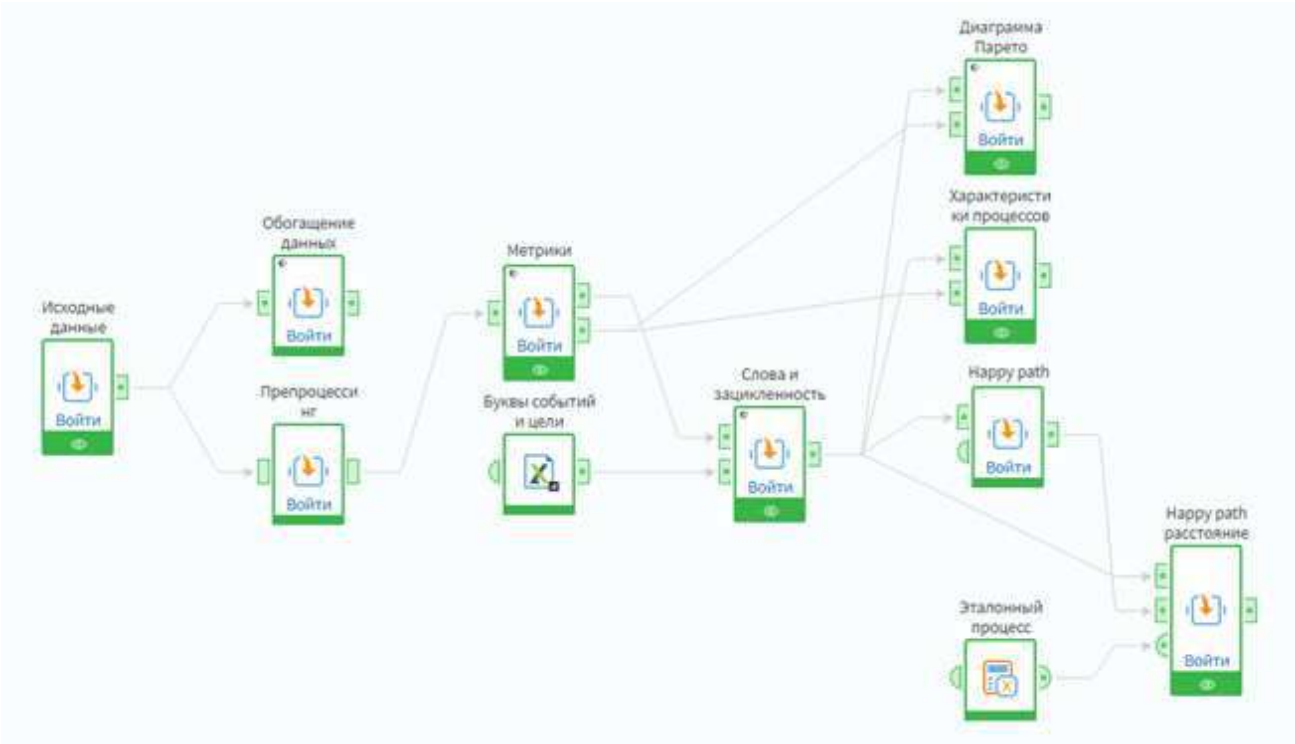


Рис. 3. Пример сценария Process Mining в Loginom «Типовой кейс воссоздания реального процесса»

Вычисленные метрики, слова процесса, зацикленность, обогащённые данные можно объединить в один датасет и использовать его для того, чтобы оценить какое влияние каждый из факторов оказывает на результат (рис 4.).

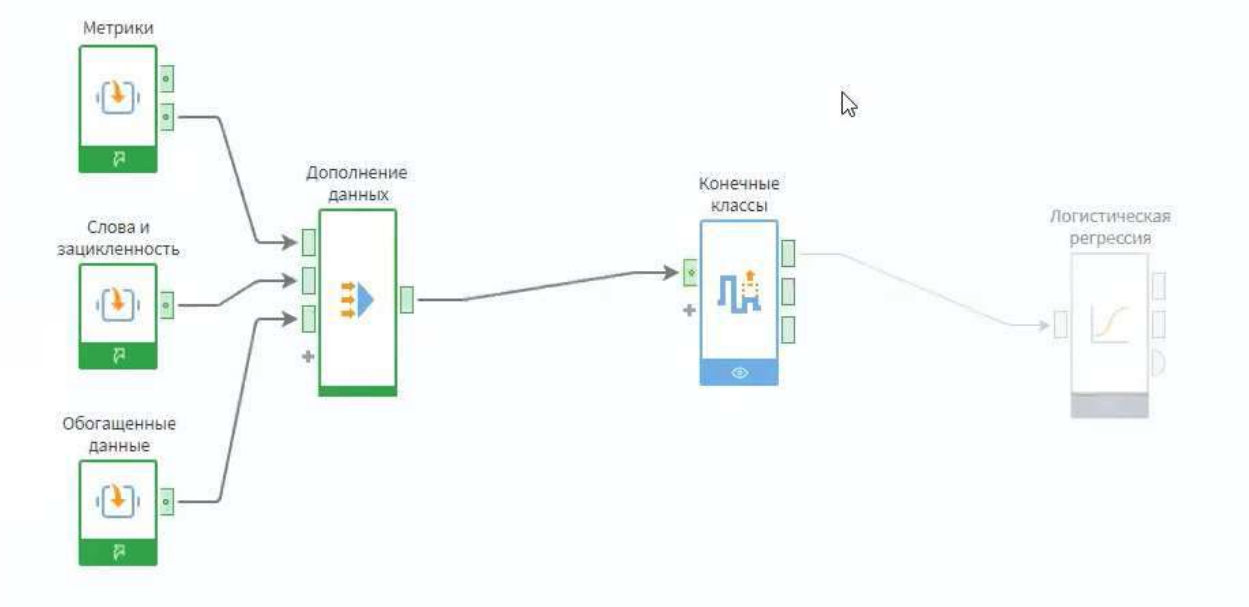


Рис. 4. Process Mining и Machine Learning в АП Loginom

Для мониторинга процесса построения и работы различных аналитических моделей, проверки гипотез и других целей, связанных с проведением Process Mining-а могут использоваться различные визуализаторы.

Например, на рис. 5 представлен график «диаграмма Парето» (определение длительности процессов), который показывает сколько процессов завершилось к тому или иному периоду. Все процессы были разделены на три типа (общие процессы, процессы, которые достигли положительного исхода и отрицательные процессы, которые не достигли положительного исхода.

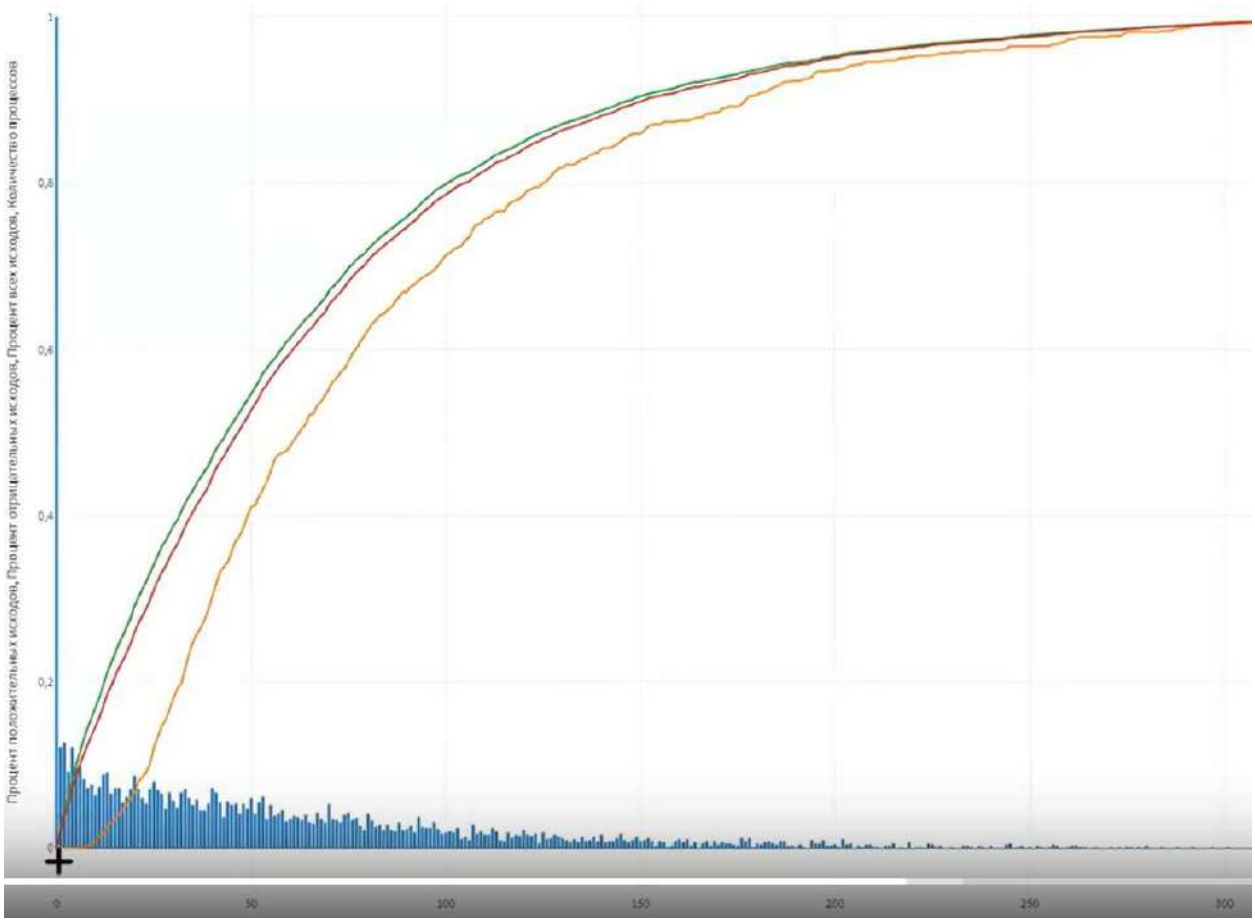


Рис. 5. Диаграмма Парето – определение длительности процессов

(по оси Ox – время в секундах)

Используя технологии OLAP-куба, можно сразу увидеть отличия между удачными и неудачными процессами (рис. 6).

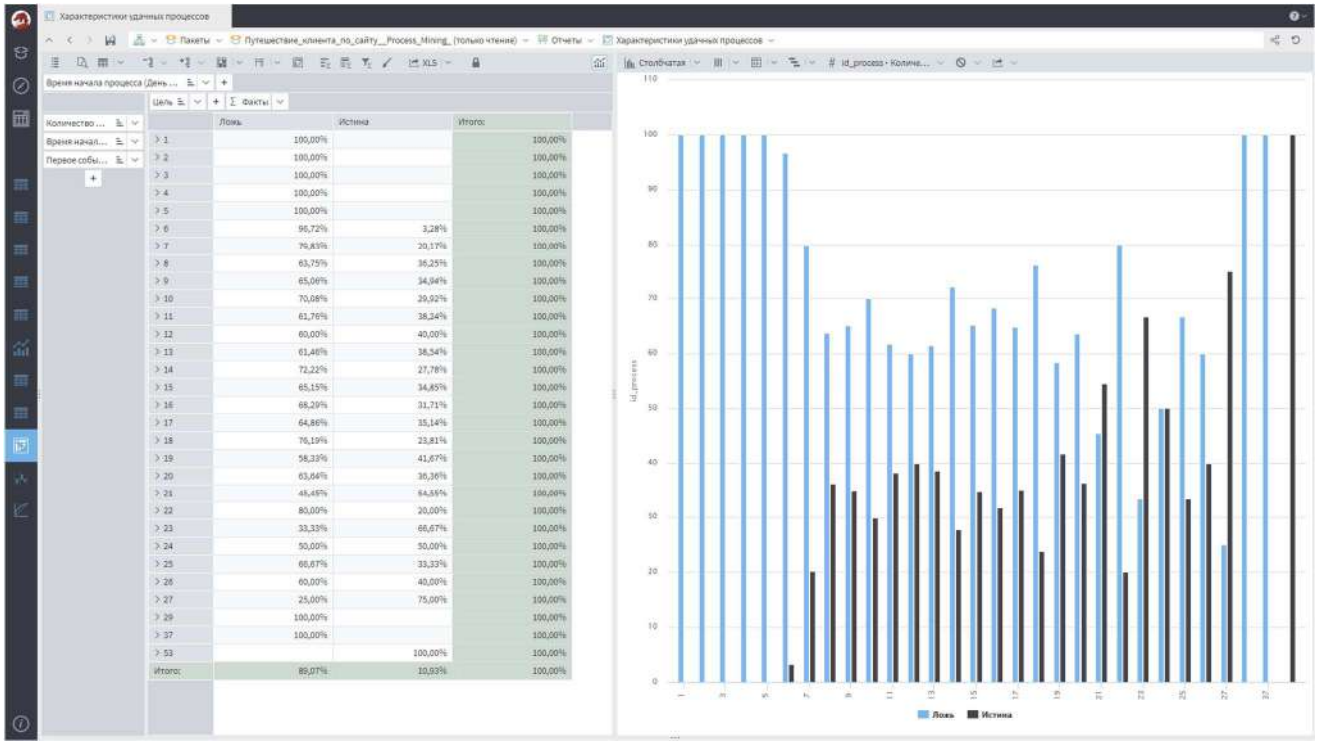


Рис. 6. OLAP-куб – удачные и неудачные процессы

Для воссоздания графа процесса и построения дашбордов можно интегрироваться с российской BI-системой Visiology [6].

Основными преимуществами модуля Process Mining системы Loginom вендор называет:

- low-code подход;
- возможность предобработки данных для подстройки под определенную задачу;
- использование готовых компонент библиотеки Process Mining_library;
- моделирование и прогнозирование;
- развёртывание и интеграция с различными источниками.

Раздел 2. Практикум по методологии Process Mining на платформе Loginom

2.1. Применение технологии Process Mining в ипотечном кредитовании

В банках Process Mining может использоваться для улучшения качества обслуживания клиентов, повышения эффективности работы, сокращения затрат и снижения рисков.

В контексте ипотечного кредитования эта технология позволяет выявить «узкие места», оптимизировать процессы, повысить скорость обслуживания клиентов и снизить риски. Process Mining используется в ипотечном кредитовании для решения следующих задач:

- **Анализ и визуализация процессов:**

Process Mining помогает визуализировать весь цикл ипотечного кредитования, от первоначального обращения до выдачи кредита, выявляя этапы, где возникают задержки или ошибки.

- **Выявление проблемных зон:**

С помощью Process Mining можно быстро определить «узкие места» в процессе, такие как задержки в обработке документов, неэффективные коммуникации между отделами или повторные ошибки в заполнении заявок.

- **Оптимизация процессов:**

Анализируя данные о реальном выполнении процессов, можно выявить возможности для их оптимизации, такие как автоматизация рутинных задач, сокращение времени ожидания или изменение порядка выполнения этапов.

- **Сокращение рисков:**

Process Mining помогает выявить отклонения от регламента и потенциальные риски, связанные с невыполнением процедур или неверными действиями сотрудников, что позволяет снизить вероятность ошибок и минимизировать финансовые потери.

- **Улучшение клиентского опыта:**

Оптимизация процессов и сокращение времени ожидания приводит к улучшению клиентского опыта и повышает лояльность клиентов.

- **Принятие обоснованных решений:**

Process Mining предоставляет объективные данные о реальном выполнении процессов, что позволяет принимать обоснованные решения о необходимости изменений и оптимизаций.

- **Увеличение эффективности:**

Оптимизация процессов и сокращение времени ожидания приводит к увеличению эффективности работы сотрудников и снижению затрат на обслуживание клиентов.

- **Автоматизация процессов:**

Process Mining помогает выявить процессы, которые можно автоматизировать, что приводит к сокращению ручного труда и повышению производительности.

- **Контроль качества:**

Process Mining позволяет контролировать качество выполнения процессов и выявлять отклонения от регламента, что помогает улучшить соблюдение стандартов и снизить риски.

- **Оценка эффективности мер оптимизации:**

Process Mining позволяет оценивать эффективность реализованных мер по оптимизации процессов и выявлять области для дальнейшего улучшения.

Примеры применения:

- **Анализ процесса одобрения заявки на кредит:**

Process Mining позволяет выявить, какие этапы в процессе одобрения заявки занимают наибольшее время и выявить причины задержек.

- **Анализ процесса подготовки документов:**

Process Mining помогает выявить, какие типы документов наиболее часто вызывают ошибки и выявить причины этих ошибок.

- **Анализ процесса передачи документов между отделами:**

Process Mining позволяет выявить, какие процессы передачи документов занимают наибольшее время и выявить причины задержек.

В итоге использование Process Mining в ипотечном кредитовании позволяет:

- Повысить скорость обслуживания клиентов.
- Сократить время ожидания.
- Снизить риски и ошибки.
- Улучшить клиентский опыт.
- Увеличить эффективность работы сотрудников.
- Повысить прибыль компании

Рассмотрим процесс ипотечного кредитования, представленный в виде блок-схемы на рис. 7, который может иметь несколько вариантов исходов:

- Негативный исход – заявки, завершающиеся событиями «Отказ Банка», «Отказ клиента до принятия решения», «Отказ клиента после принятия решения».
- Позитивный исход – заявки, завершающиеся событиями «Кредит готов к рефинансированию», «Кредитный договор закрыт», «Сделка подписана».

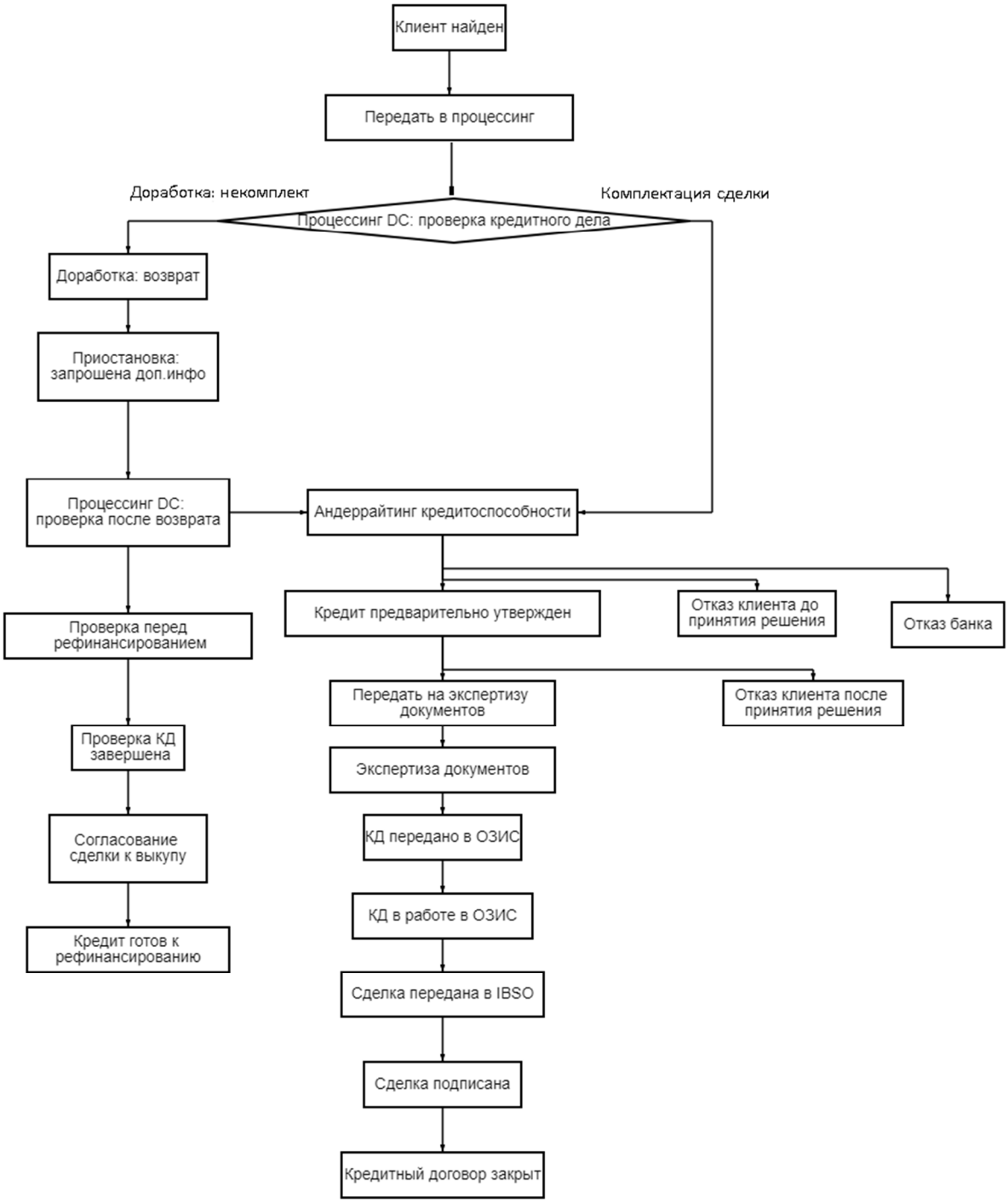


Рис. 7. Процесс ипотечного кредитования

Предусмотреть все возможные исходы прохождения заявок на выдачу кредита заранее просто невозможно. Технология Process Mining позволяет оценить эффективность любого текущего процесса по выдаче ипотеки.

Первым этапом применения методологии Process Mining в ипотечном кредитовании на платформе Loginom является постановка задачи и подготовка данных с входным контролем в соответствии с требованиями анализа данных:

объединение из разных источников, обогащение данных дополнительными атрибутами, входной аудит и валидация данных.

Для формирования журнала логов необходимо выполнить следующие шаги:

1. Определение типа процесса, который мы будем анализировать, и формирование правил, определяющих его идентификатор в системе;

2. Определение правил формирования идентификатора экземпляра процесса;

3. Определение правил обозначения этапа и роли исполнителя процесса;

4. Определение временных метрик, которые мы будем регистрировать в журнале, и способ их расчета;

5. Определение дополнительных метаданных, которые нам потребуются для более глубокой аналитики.

На рисунке 8 представлен фрагмент датасета по ипотечному кредитованию (объем данных составлял 1048547 записей).

ab Process…	ab Event	31 Timestamp	31 Data_End	ab Кредитный …	ab MANAGER	ab CHANEL	ab Структур
0000B522-…	Клиент найден	05.10.2020 20:42:32	05.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57
0000B522-…	Передать в процессинг	05.10.2020 20:51:14	06.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57
0000B522-…	Процессинг DC: пров…	06.10.2020 10:48:27	06.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57
0000B522-…	Андеррайтинг креди…	06.10.2020 10:59:52	06.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57
0000B522-…	Доработка: некомплект	06.10.2020 14:21:05	06.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57
0000B522-…	Процессинг DC: пров…	06.10.2020 16:37:09	06.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57
0000B522-…	Андеррайтинг креди…	06.10.2020 16:41:13	06.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57
0000B522-…	Кредит предварител…	06.10.2020 17:14:57	06.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57
0000B522-…	Передать на эксперт…	06.10.2020 22:29:38	07.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57
0000B522-…	Андеррайтинг креди…	07.10.2020 14:03:25	07.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57
0000B522-…	Кредит предварител…	07.10.2020 15:27:53	08.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57
0000B522-…	Передать на эксперт…	08.10.2020 06:13:41	08.10.2020 …	Готовое жилье …	Manager 185	Физ. Лицо	Партнёр 57

Рис. 8. Фрагмент датасета по ипотечному кредитованию

В исходном журнале логов представлена следующая информация:

- Process_ID – определяет экземпляр процесса.
- Event – упорядоченный перечень событий, выполняемых в рамках экземпляров процесса.
- Timestamp – момент начала совершения события.
- Data_End – момент завершения процесса.
- Кредитный продукт – наименования кредитного продукта.

- Chanel – физические и юридические лица.
- Manager – сотрудник банка.
- Структурное подразделение – наименование подразделения банка.

Экземпляр процесса представляет упорядоченный перечень операций для описания одного завершенного действия в рамках бизнес-процесса. Экземпляр процесса маркируется уникальным Process_ID.

На рисунке 9 представлен экземпляр процесса с Process_ID 0000B522–3207–EB11–8476–005056BF4460.

#	Process_ID	Event	Timestamp
1	0000B522-3207-EB11-8476-005056BF4460	Клиент найден	05.11.2020 23:39:39
2	0000B522-3207-EB11-8476-005056BF4460	Передать в процессинг	05.11.2020 23:48:21
3	0000B522-3207-EB11-8476-005056BF4460	Процессинг DC: проверка кредитного дела	06.11.2020 13:45:34
4	0000B522-3207-EB11-8476-005056BF4460	Андеррайтинг кредитоспособности	06.11.2020 13:56:59
5	0000B522-3207-EB11-8476-005056BF4460	Доработка: некомплект	06.11.2020 17:18:12
6	0000B522-3207-EB11-8476-005056BF4460	Процессинг DC: проверка после возврата	06.11.2020 19:34:16
7	0000B522-3207-EB11-8476-005056BF4460	Андеррайтинг кредитоспособности	06.11.2020 19:38:20
8	0000B522-3207-EB11-8476-005056BF4460	Кредит предварительно утвержден	06.11.2020 20:12:04
9	0000B522-3207-EB11-8476-005056BF4460	Передать на экспертизу документов	07.11.2020 1:26:45
10	0000B522-3207-EB11-8476-005056BF4460	Процессинг перед повторным андеррайтингом	07.11.2020 17:00:32
11	0000B522-3207-EB11-8476-005056BF4460	Кредит предварительно утвержден	07.11.2020 18:25:00
12	0000B522-3207-EB11-8476-005056BF4460	Передать на экспертизу документов	08.11.2020 9:10:48
13	0000B522-3207-EB11-8476-005056BF4460	Процессинг перед повторным андеррайтингом	08.11.2020 14:22:18
14	0000B522-3207-EB11-8476-005056BF4460	Кредит предварительно утвержден	08.11.2020 16:36:34
15	0000B522-3207-EB11-8476-005056BF4460	Передать на экспертизу документов	08.11.2020 20:15:56
16	0000B522-3207-EB11-8476-005056BF4460	КД передано в ОЗИС	23.11.2020 17:27:49
17	0000B522-3207-EB11-8476-005056BF4460	Сделка передана в IBSO	26.11.2020 12:06:56
18	0000B522-3207-EB11-8476-005056BF4460	Сделка подписана	28.11.2020 9:59:00

Рис. 9. Экземпляр процесса ипотечного кредитования

События упорядочены по времени (рис. 10). В различных экземплярах процессов могут встречаться одни и те же события, но они будут отличны между собой по Timestamp, что будет делать события уникальными.

ab Event	31 Timestamp
Клиент найден	05.11.2020 23:39:39
Передать в процессинг	05.11.2020 23:48:21
Процессинг DC: проверка кредитного дела	06.11.2020 13:45:34
Андеррайтинг кредитоспособности	06.11.2020 13:56:59
Доработка: некомплект	06.11.2020 17:18:12
Процессинг DC: проверка после возврата	06.11.2020 19:34:16
Андеррайтинг кредитоспособности	06.11.2020 19:38:20
Кредит предварительно утвержден	06.11.2020 20:12:04
Передать на экспертизу документов	07.11.2020 1:26:45
Процессинг перед повторным андеррайтингом	07.11.2020 17:00:32
Кредит предварительно утвержден	07.11.2020 18:25:00
Передать на экспертизу документов	08.11.2020 9:10:48
Процессинг перед повторным андеррайтингом	08.11.2020 14:22:18
Кредит предварительно утвержден	08.11.2020 16:36:34
Передать на экспертизу документов	08.11.2020 20:15:56
КД передано в ОЗИС	23.11.2020 17:27:49
Сделка передана в IBSO	26.11.2020 12:06:56
Сделка подписана	28.11.2020 9:59:00

Рис. 10. Упорядоченность событий по времени

Выделяют следующие необходимые требования для журнала логов:

- все события, записанные в протоколе, должны быть идентифицированы с экземплярами процессов;
- все события должны быть упорядочены по времени их выполнения;
- разнотипные события должны различаться.

После загрузки данных в Loginom, первичный аудит журнала логов можно выполнить помощью визуализатора «Качество данных» (рис. 11, 13).

Метка	Вид	Проблемы	Виды проблем
Data_End		16,04%	Пропуски - 16,04% (168 232)
Кредитный продукт		2,61%	Экстремальные - 0,40% (4 165) Выбросы - 2,21% (23 182)
Структурное подр…		1,69%	Экстремальные - 0,00% (9) Выбросы - 1,69% (17 751)
MANAGER		0,85%	Экстремальные - 0,01% (85) Выбросы - 0,84% (8 829)
Event		0,66%	Экстремальные - 0,12% (1 254) Выбросы - 0,54% (5 659)
Process_ID			
Timestamp			
CHANEL			

Рис. 11. Визуализатор «Качество данных»

Рис. 12. Анализ поля Process_ID

Есть пропуски в поле Data_End – момент завершения процесса. Следующий отчет показывает список событий, у которых отсутствует дата завершения:

	ab Event	12 Data_End|Количество
1	Сделка подписана	23 638
2	Отказ клиента до принятия решения	53 568
3	Приостановка: запрошена доп.инфо	2 038
4	Отказ клиента после принятия решения	37 725
5	Отказ Банка	20 513
6	Комплектация сделки	997
7	Передать на экспертизу документов	2 915
8	Кредитный договор закрыт	1 199
9	Кредит предварительно утвержден	5 380
10	Доработка: возврат	1 236
11	Кредит готов к рефинансированию	568
12	Сделка передана в IBSO	201
13	Доработка: некомплект	1 356
14	Процессинг DC: проверка кредитного дела	76
15	Андеррайтинг кредитоспособности	272
16	Проверка КД завершена	13
17	Рефинансирование DC: доработка	11
18	КД передано в ОЗИС	102
19	Клиент найден	47
20	Передать в процессинг	31
21	Процессинг DC: проверка после возврата	5
22	КД в работе в ОЗИС	3

Рис. 13. Отчет «Список событий, у которых пропуски в поле дата окончания»

Можно добавить фильтрацию удаления незавершённых событий и событий, начало которых в датасете отсутствует.

Восстановим пропущенные значения в поле Data_End (рис. 14).

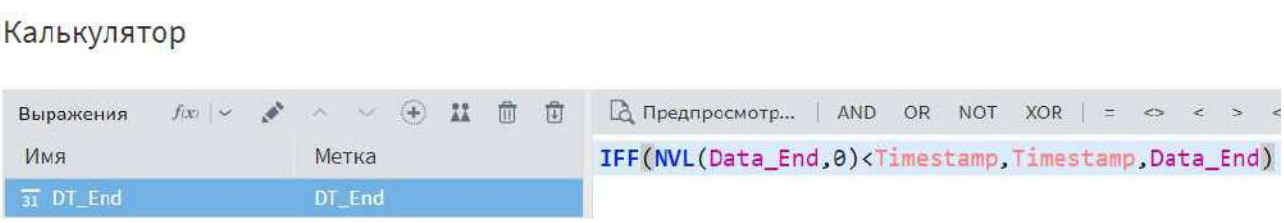
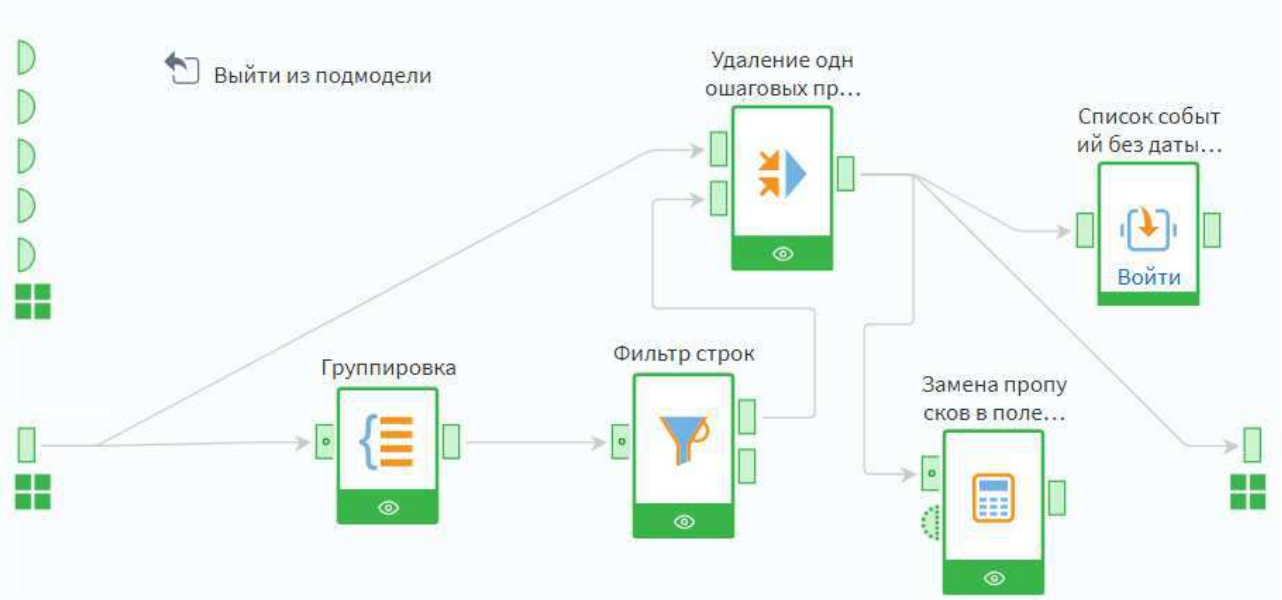
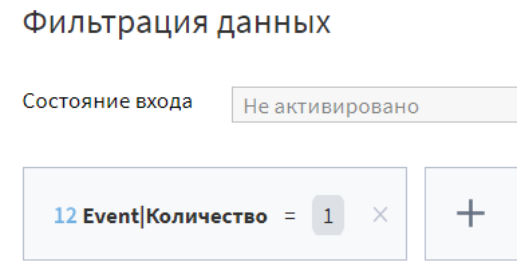


Рис. 14. Замена пропущенных значений в поле Data_End

Удалим одношаговые процессы с помощью компонент: группировка, фильтрация и разность (рис. 15).





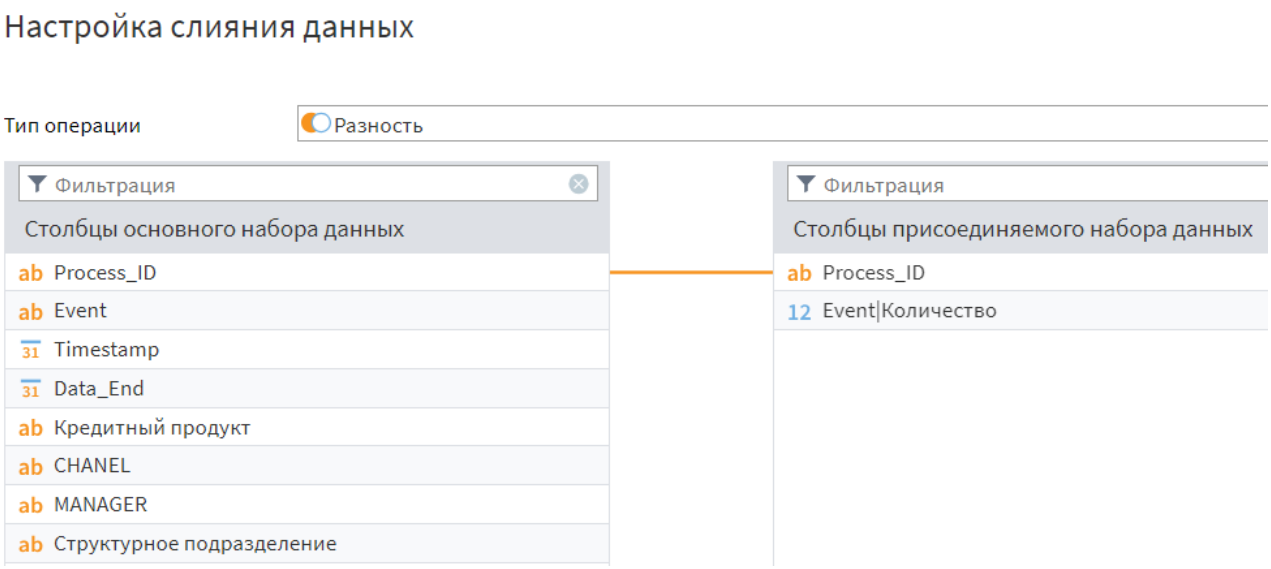


Рис. 15. Подмодель «Удаление одношаговых процессов и аудит»

Анализируем статистические показатели и рассчитываем основные метрики по событиям, по подразделениям, по сотрудникам и по любой их комбинации (рис. 16).

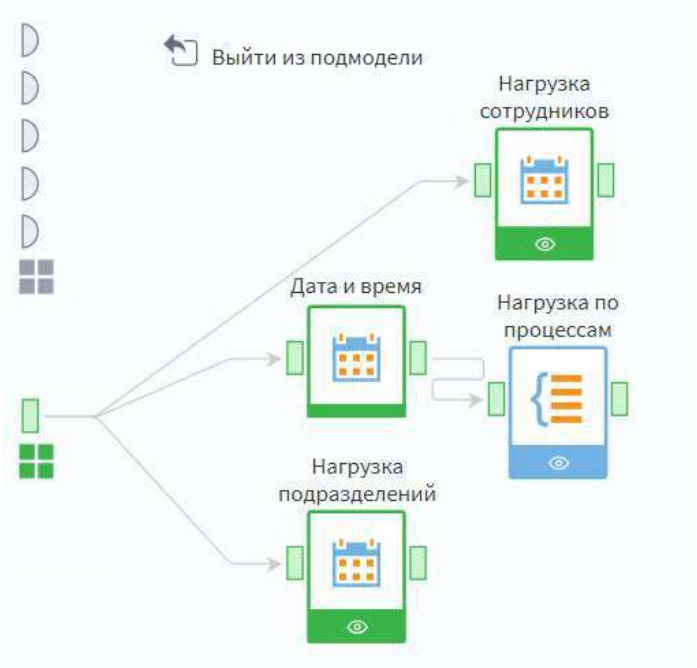


Рис. 16. Подмодель «Аналитика по процессам, по подразделениям, по сотрудникам»

Ключевые метрики показывают количество процессов (количество кредитных продуктов) в день, количество событий в день, средний срок прохождения и рост заявок по месяцам. Например, визуализатор «Нагрузка про процессам» (рис. 17) показывает, какое количество процессов запущено одновременно.

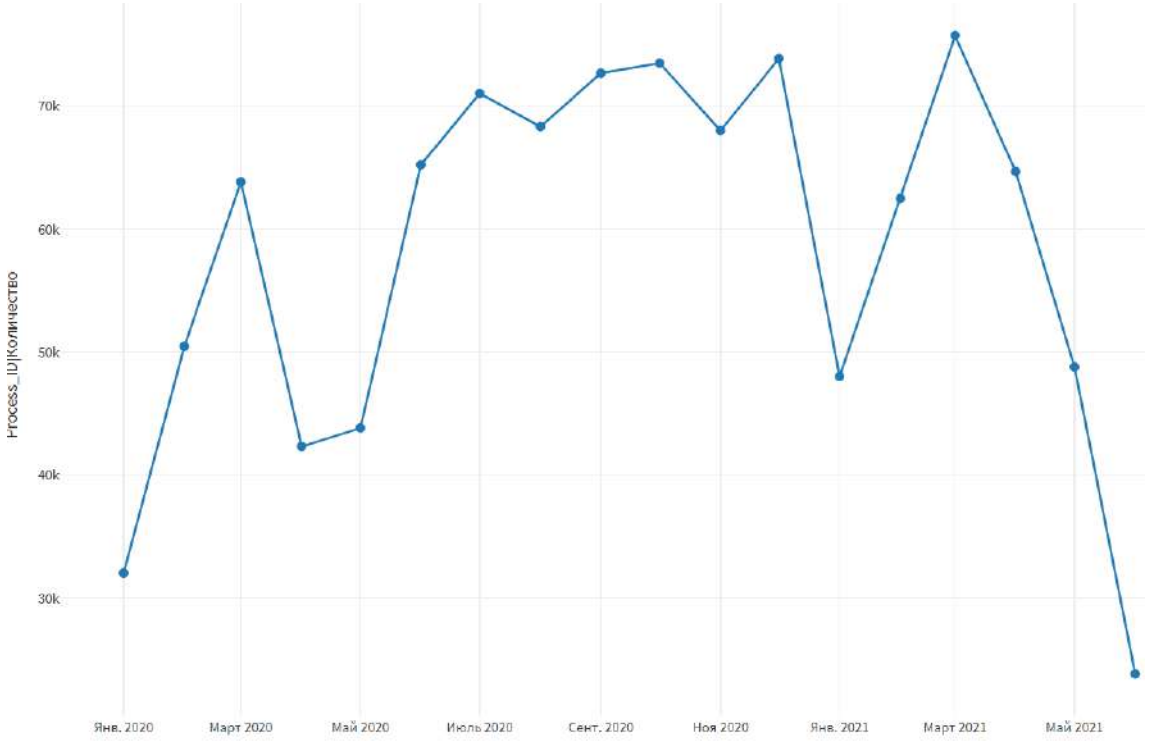


Рис. 17. График динамики количества процессов

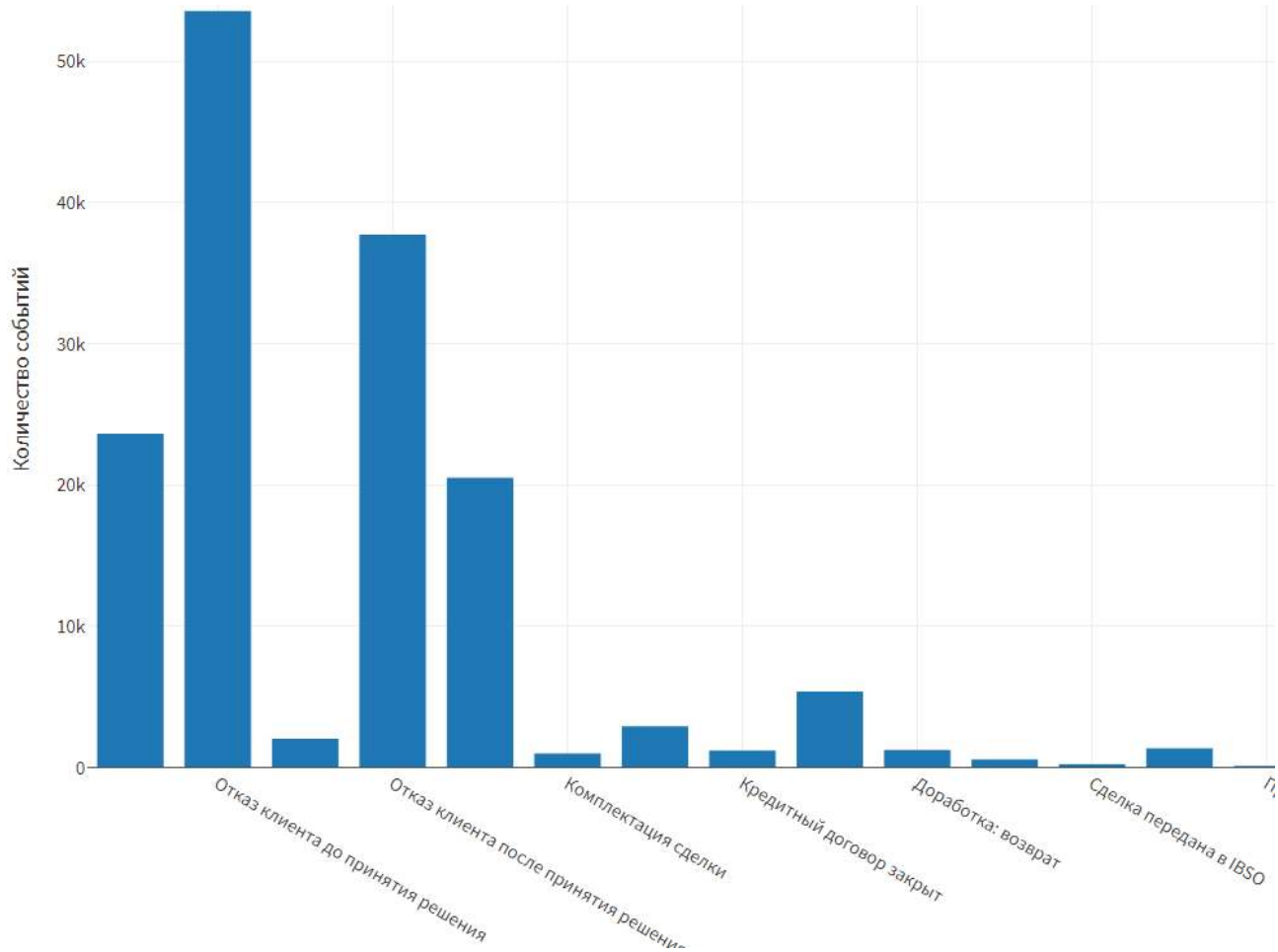


Рис. 18. Диаграмма «Распределение событий»

OLAP-куб «Нагрузка сотрудников» (рис. 19) показывает, сколько задач сотрудники выполняют каждый день.

MANAGER | Факты

Timestamp (Г…	Manager 10	Manager 100	Manager 101	Manager 102	Manager 103	Manager 104	Manager 105
24.01.2020			7		4		12
25.01.2020							
26.01.2020							
27.01.2020			11		9		11
28.01.2020	8		25		6		6
29.01.2020	2		19		8		4
30.01.2020	1		18		5		4
31.01.2020	6		11		8		5
01.02.2020					1		
02.02.2020			1				
03.02.2020	7		17		15		2
04.02.2020	2		3		6		5
05.02.2020	11		4		5		15
06.02.2020	4		1		19		4
07.02.2020	1		6		7		7

Рис. 19. OLAP-куб «Нагрузка сотрудников»

По подразделениям можно посмотреть какие работают активно (рис. 20).

Структурное подразделение | Факты

Timestamp (Г…	Волгоград	Воронеж	Екатери…	Ижевск
27.01.2020	2	11	29	
28.01.2020		19	32	
29.01.2020	9	29	43	3
30.01.2020	4	10	41	1
31.01.2020	7	17	23	3
01.02.2020				
02.02.2020				
03.02.2020	2	29	38	1
04.02.2020	2	16	36	3
05.02.2020	9	18	27	1
06.02.2020	8	19	18	3
07.02.2020	9	5	23	3
08.02.2020			1	
09.02.2020				
10.02.2020	5	21	34	12
11.02.2020	3	13	24	22
12.02.2020	10	24	30	7
13.02.2020	11	12	44	9

Рис. 20. OLAP-куб «Нагрузка подразделений»

Для реализации второго этапа Process Mining «Проверка соответствия» нужно каждое событие заменить своей уникальной буквой, тогда каждый – последовательность его событий (путь) можно будет представить в виде «слова процесса».

Представление пути как слова процесса позволяет проводить анализ процесса с использованием алгоритмов: выявлять статистику некорректных переходов, определять возвраты и циклы. Длина пути процесса (количество событий в процессе) равно длине слова процесса.

Опишем шаги формирования букв и слов процесса. В сценарии под буквой процесса подразумевается одно действие, словом обозначена последовательность действий.

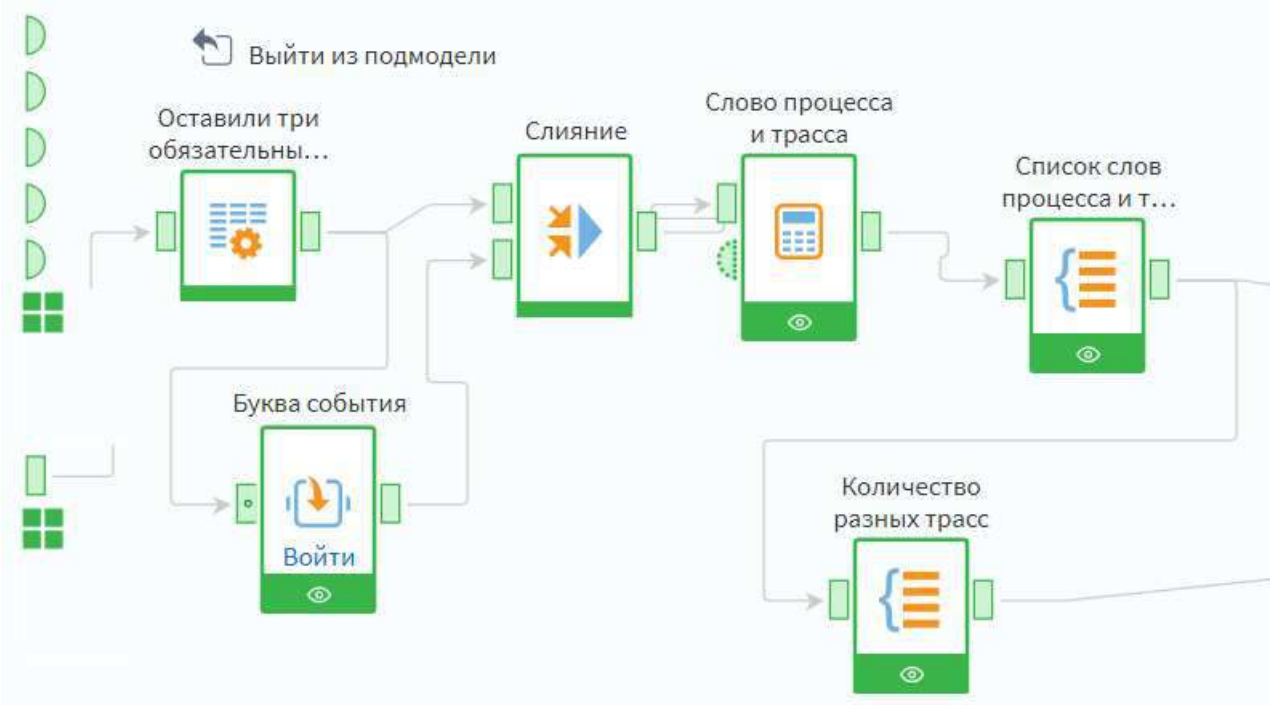


Рис. 21. Подмодель «Формирование слов процесса и трасс»

Все события в бизнес-процессе ипотечного кредитования должны быть обозначены буквами с помощью компонента Замена (рис. 22). Например, клиент найден – буква «А», заявка передана в процессинг – «В» и т.д.

Замена

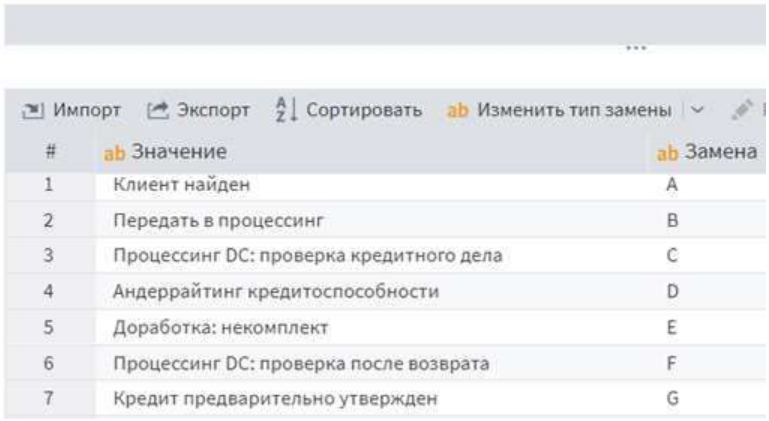


Рис. 22. Компонент «Замена» сопоставление каждому событию – уникальной буквы

Все буквенные обозначения событий приведены в таблице 3.

Таблица 3. Буквенные обозначения событий

ab Event	ab Буква события
Клиент найден	A
Передать в процессинг	B
Процессинг DC: проверка кредитного дела	C
Андеррайтинг кредитоспособности	D
Доработка: некомплект	E
Процессинг DC: проверка после возврата	F
Кредит предварительно утвержден	G
Передать на экспертизу документов	H
КД передано в ОЗИС	I
Сделка передана в IBSO	J
Сделка подписана	K
Отказ клиента до принятия решения	L
Приостановка: запрошена доп.инфо	M
Отказ клиента после принятия решения	N
Отказ Банка	O
Комплектация сделки	P
Кредитный договор закрыт	Q
Доработка: возврат	R
КД в работе в ОЗИС	S
Проверка перед рефинансированием	T
Проверка КД завершена	U
Согласование сделки к выкупу	V
Кредит готов к рефинансированию	W
Рефинансирование DC: доработка	X
Экспертиза документов	Z

Находим слова экземпляров процесса с помощью калькулятора (рис. 23):

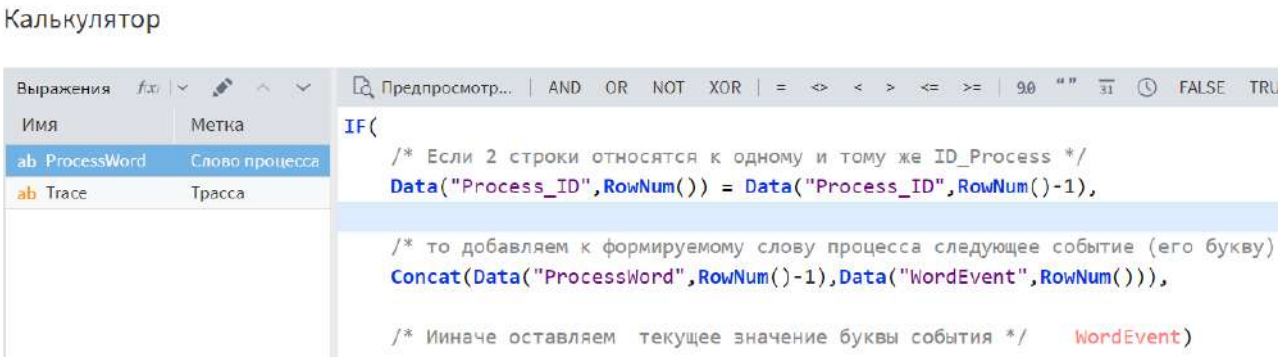


Рис. 23. Компонент «Калькулятор». Формируем слова экземпляров процесса

Для одного экземпляра процесса результат построения слова и трассы показан на рис 24.

Process_ID 1	Слово процесса	Трасса
0000B522-3207-EB11-8476-005056BF4460	A	A
0000B522-3207-EB11-8476-005056BF4460	AB	AB
0000B522-3207-EB11-8476-005056BF4460	ABC	ABC
0000B522-3207-EB11-8476-005056BF4460	ABCD	ABCD
0000B522-3207-EB11-8476-005056BF4460	ABCDE	ABCDE
0000B522-3207-EB11-8476-005056BF4460	ABCDEF	ABCDEF
0000B522-3207-EB11-8476-005056BF4460	ABCDEFD	ABCDEF
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDG	ABCDEFG
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDGH	ABCDEFGH
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDGHD	ABCDEFGH
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDGHDG	ABCDEFGH
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDGHDGH	ABCDEFGH
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDGHDGHD	ABCDEFGH
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDGHDGHDG	ABCDEFGH
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDGHDGHDGH	ABCDEFGH
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDGHDGHDGHI	ABCDEFGHI
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDGHDGHDGHIJ	ABCDEFGHIJ
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDGHDGHDGHIJK	ABCDEFGHIJK

Рис. 24. Слово и трасса для одного экземпляра процесса

Например ABCDEFDGHDGHDGHIJK– это слово процесса – уникальный путь процесса, где каждой букве соответствует определенная операция или этап. Например, этап обращения клиента обозначается буквой «А», передача заявки на рассмотрение – буквой «В» и т. д. (см. таблицу 3). В процессе рассмотрения заявок на ипотеку каждый клиент достигает определенных этапов.

Например, у кого-то слово процесса может состоять из одной буквы A – это значит, что клиент отказался от услуг уже на этапе обращения в банк.

Повторяющиеся в одном слове буквы означают, что процесс зациклен.

В зависимости от слова процесса можно проанализировать разные характеристики, например различие в длительности (рис. 25).

ab Слово процесса	12 Длительность процесса (дней)
ABCGHIJKQ	512
BCDGHDML	511
ABCDGDMCDGHDDGHDMCDGHDDGHDDGHDDEFDGDDGHN	505
BCGHIJPDGHIKQ	505
ABCDEFDGDGHDGHIJKQ	500
ABCDGDMCGDGDDGDMCGDGDDGDMCDGDMFGHDMFGH	496
BCMCDGHIJKQ	494
BCMCDEFDRFDGHDDGDDGHDDGHDDGHDDEFML	493
ABCDGHIJKQ	491
BCMCMCDGHDDGHIJKKQ	490
ABCMLBCML	486
BCDODDODDGHIJKQ	486

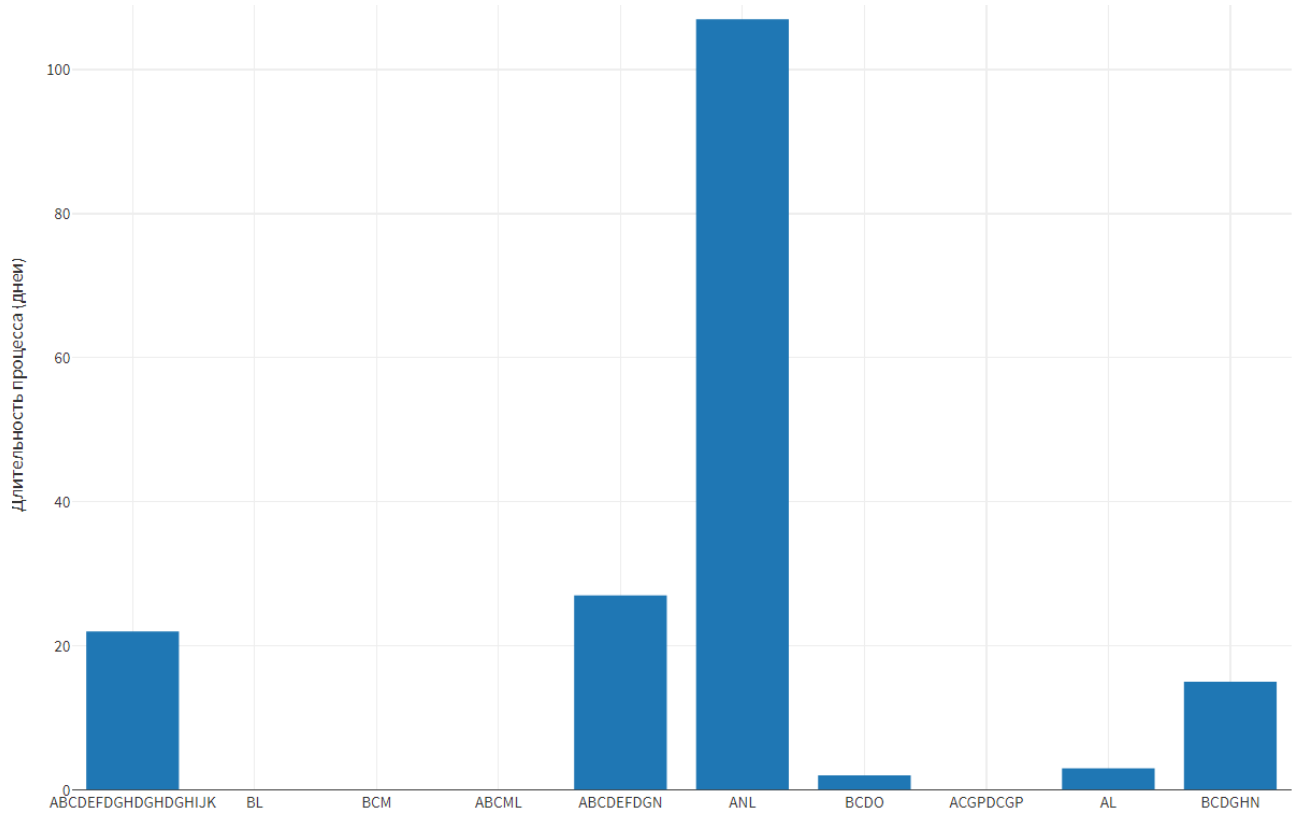


Рис. 25. Длительность в зависимости от слова

Некоторые события начинаются в одно и тоже время и идут параллельно. Можно оценить какое время затрачивается на параллельность, отфильтровать параллельные события и т. д.

В подмодели «Анализ длительности», также можно рассчитать следующие характеристики (рис. 26):

Отказ в %

Метка	Значение
Отказ клиента до принятия решения (%)	64,73
Отказ клиента после принятия решения (%)	75,16

Рис. 26. Процент отказа клиентов до и после принятия решения банком

Дополнительно, используя слова процесса, также можно:

1) анализировать длительность процесса каждого отдельного этапа, время задержки между ними, определять вероятность тех или иных ветвлений в процессах, количество циклов;

2) давать статистические оценки или определять различные аномалии в реализации процессов: обнаруживать экземпляры процессов или этапов, которые исполняются аномально долго, обнаруживать зависшие процессы и ситуации, когда процесс зацикливается между отдельными его этапами (ситуация пинг-понга) и пр.

3) анализировать нарушение регламентов. Например, как часто происходят пропуски обязательных этапов процесса, начало процесса не с первого этапа, нарушение последовательности этапов, выявлять параметры тех процессов, которые идут с нарушением регламентов и пр.

Для того, чтобы найти внутри групп экземпляров процесса пути, последовательность смены событий в которых некорректна или процессы, не достигшие результата (не валидное финальное событие), найдем трассы – последовательность событий, в которых каждое событие появляется только один раз (рис. 27).

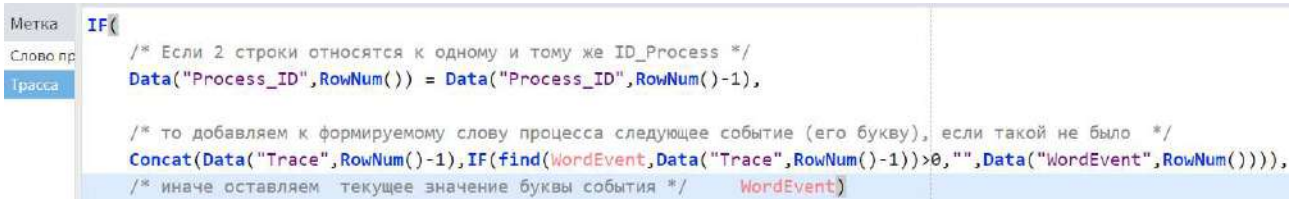


Рис. 27. Компонент «Калькулятор». Находим трассы

Применяя компонент «Группировка» для всех экземпляров процессов находим слова и трассы. На рис. 28 представлены примеры полученных последовательностей событий – слова процессов и трассы.

ab Process_ID	ab Слово процесса	ab Трасса
0000B522-3207-EB11-8476-005056BF4460	ABCDEFDGHDGHDGHIJK	ABCDEFGHIJK
0000C950-1386-EA11-8470-005056BF4460	BL	BL
0000DA01-5D8D-EB11-8485-005056BF5AD0	BCM	BCM
0000FE27-F3A8-EB11-848C-005056BF5AD0	ABCML	ABCML
00010D1E-8A1B-EB11-8482-005056BF5AD0	ABCDEFDGN	ABCDEFGN
00013BAB-1B6D-EA11-846F-005056BF4460	ANL	ANL
00013F42-853B-EA11-846E-005056BF4460	BCDO	BCDO
00014737-7637-EA11-846E-005056BF5AD0	ACGPDCGP	ACGPD
0001A436-A15C-EB11-8476-005056BF4460	AL	AL
0001E9DB-93EB-EA11-8476-005056BF4460	BCDGHN	BCDGHN
0001FADC-358D-EB11-8477-005056BF4460	BCGN	BCGN
00024331-8B0F-EB11-8476-005056BF4460	AL	AL
00025F08-9AA6-EB11-8478-005056BF4460	BCMCODDEFDGDGH	BCMODEFGH
00026C29-D981-EB11-8485-005056BF5AD0	ACODCDGPIJK	ACODGPIJK

Рис. 28. Слова и трассы для экземпляров процессов

ab Трасса 1	12 Количество
AB	29
ABC	18
ABCD	22
ABCDE	93
ABCDEF	9
ABCDEFG	12
ABCDEFGH	37
ABCDEFGHI	3
ABCDEFGHIJ	2
ABCDEFGHIJK	266
ABCDEFGHIJKQ	17
ABCDEFGHIJN	1

Рис. 29. Трассы и их количество в журнале логов

Используя найденные трассы были определены следующие группы процессов, имеющие несколько вариантов исходов событий (рис. 30):

Негативный исход – заявки, завершающиеся событиями:

- «Отказ Банка» – 680 процессов
- «Отказ клиента до принятия решения» – 459 процессов
- «Отказ клиента после принятия решения» – 1740 процессов

Позитивный исход – заявки, завершающиеся событиями:

- «Кредит готов к рефинансированию» – 78 процессов
- «Кредитный договор закрыт» – 270 процессов
- «Сделка подписана» – 1244 процесса
- «Незавершенные процессы» – 1401 процесс.

ab События	12 Количество	9.0 Доля
Кредитный договор закрыт	270	0,05
Отказ клиента до принятия решения	459	0,08
Отказ банка	680	0,12
Отказ клиента после принятия решения	1 740	0,30
Сделка подписана	1 244	0,21
Кредит готов к рефенансированию	78	0,01
Незавершенные процессы	1 401	0,24

Рис. 30. Результат подмодели «Анализ трасс»

«Отказ банка» – отказ банка в выдаче кредита по определенном причинам, например, испорченная кредитная история, судебные тяжбы с банковскими организациями, предоставление недостоверных сведений о себе, а также поддельных документов. Сюда также относится некорректно заполненная документация, неоплаченные задолженности по штрафам или налогам, наличие открытых кредиток, действующих потребительских займов и др. Приходится примерно 12% процессов.

«Отказ клиента до принятия решения» – те процессы, когда клиент вероятно мог найти более подходящие условия для до того, чтобы взять ипотеку. Приходится примерно 8% процессов.

«Отказ клиента после принятия решения» – те процессы, когда клиент вероятно мог найти более подходящие условия для до того, чтобы взять ипотеку, либо вовсе передумал. Приходится примерно 30% процессов

«Кредит готов к рефинансированию» – смена одной кредитной организации на другую: должник берет деньги у другого банка, как правило, на более выгодных условиях, чтобы погасить имеющийся кредит. Приходится примерно 1% процессов.

«Кредитный договор закрыт» – окончание срока кредитного договора и полного погашения задолженности по займу. Приходится примерно 5% процессов.

«Сделка подписана» – принято положительное решение о выдаче кредита, между заемщиком и кредитной организацией. Приходится примерно 20% процессов.

«Незавершенные процессы» – незаконченные или прерванные процессы. Приходится примерно 24% процессов.

Важным этапом сценария PM являются определение эталонного и счастливого пути для процесса выдачи ипотечного кредитования.

Эталонный путь (Reference path) – это запроектированный путь процесса, который может быть использован как эталон (перечень событий среди экземпляров процесса, приводящий к желаемому результату). Эталонный путь должен приводить к требуемому результату и не иметь зацикливаний. Часто эталонный процесс задаётся экспертом.

Возможна ситуация, когда эталонный путь отсутствует в исследуемом журнале событий, тогда допустимо принимать в качестве эталонного пути существующий счастливый путь – путь, по которому проходит наибольшее количество процессов или это наиболее часто встречающаяся последовательность событий среди экземпляров процесса, где нет ошибочных ветвлений или циклов, начинаются и заканчивается на валидные начальные / конечные события, по которому прошло максимальное количество экземпляров процесса.

Эталонный путь в данном примере: ABCDGHIJKQ: обращение клиента (A), передача заявки на рассмотрение (B), проверка кредитного дела (C), андеррайтинг кредитоспособности (D), кредит предварительно утвержден (G), документы переданы на экспертизу (H), кредитное дело передано в ОЗИС (I), сделка передана в IBSO (J), сделка подписана (K), кредитный договор закрыт (Q).

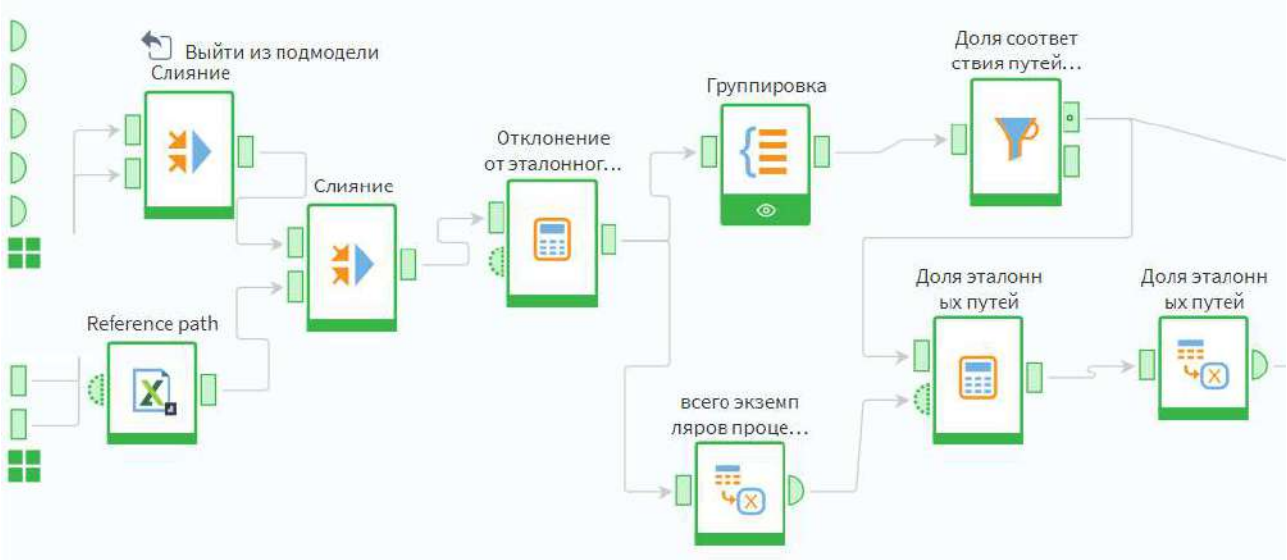
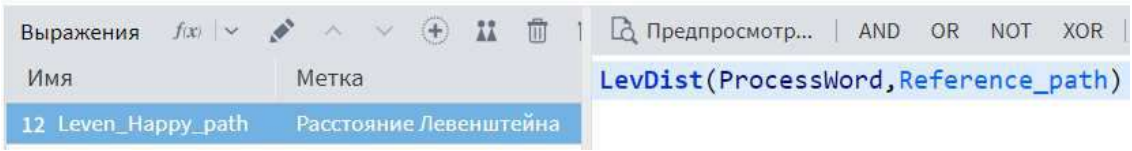


Рис. 31. Подмодель «Поиск эталонных путей в журнале логов»

Мерой отличия отдельного пути процесса от эталонного пути может быть расстояние Левенштейна. Метрика, измеряющая по модулю разность между двумя последовательностями символов. Она определяется как минимальное количество одно символьных операций, необходимых для превращения одной последовательности символов в другую.

С помощью встроенного в платформу Loginom «Калькулятора» отклонение от эталонного пути рассчитывается с помощью функции DemLevDist(рис. 32).

Калькулятор



ab Process_ID	ab Слово процесса	12 Расстояние Левенштейна
0003328D-847D-EA11-8470-005056BF4460	ABCML	7
00034AA9-DDC9-EB11-848C-005056BF5AD0	AL	9
00034CAA-063D-EA11-846E-005056BF5AD0	ABCDGHIJK	1
0003753D-395A-EB11-8483-005056BF5AD0	AL	9
00039053-E8C5-EB11-848C-005056BF5AD0	AL	9
000473F3-48FE-EA11-8482-005056BF5AD0	BCODDO	8
00056CB7-E39A-EB11-8489-005056BF5AD0	ABCDO	6

Рис. 32. Расстояние слов процесса от эталонного пути

Доля эталонных путей оказалась очень маленькой 0,08%.

Счастливый (основной) путь – это путь, в котором нет ошибочных ветвлений или циклов, начинающийся и заканчивающийся на валидные начальные / конечные события, по которому прошло максимальное количество экземпляров процесса.

На рисунке 33 изображена подмодель нахождения счастливых путей.

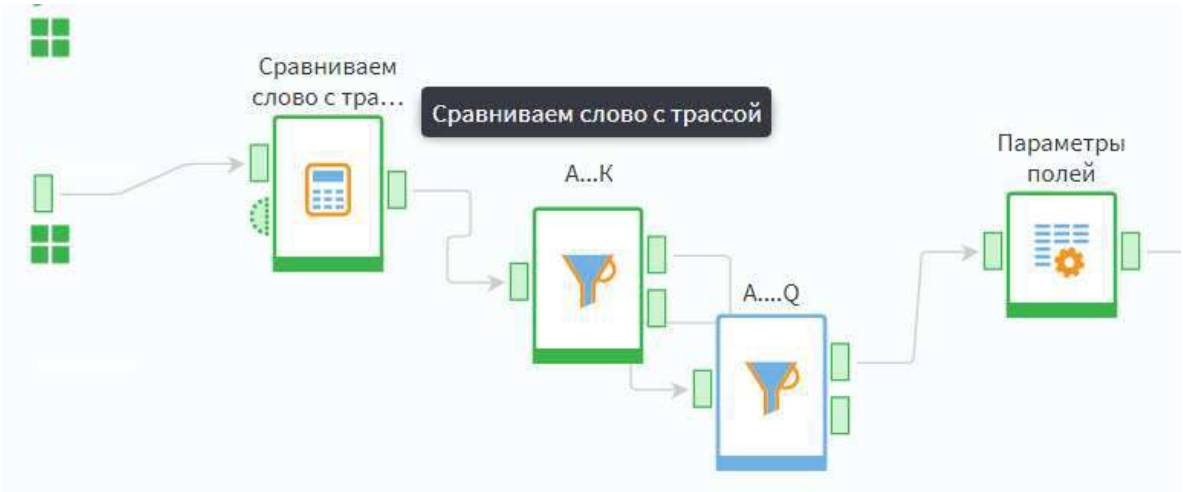


Рис. 33. Подмодель «Счастливые пути HappyPath»

ab Process_ID	ab Счастливый путь
00034CAA-063D-EA11-846E-005056BF5AD0	ABCDGHIJK
0006065A-76E7-EA11-8481-005056BF5AD0	ABCGHIJK
0008A90D-1765-EA11-846E-005056BF4460	ABCGHIJK
00098CF8-5C55-EA11-846E-005056BF4460	ABCGHIJKQ
000C9DF7-81B0-EA11-8475-005056BF4460	ABCGHIJK
00455420-644E-EA11-846E-005056BF4460	ABCGHIJK
005A274E-4452-EA11-846E-005056BF4460	ABCDGHIJK
0065FCF5-29A7-EB11-8478-005056BF4460	ACDGPIJK
006AFAA6-C09E-EA11-847D-005056BF5AD0	ABCDGHIJK
0076550A-5CD1-EA11-8481-005056BF5AD0	ABCGHIJK
0078591F-BE72-EA11-8474-005056BF5AD0	ABCGHIJK
007F5C68-DC11-EB11-8476-005056BF4460	ABCGHIJK
008B8B19-8243-EB11-8476-005056BF4460	ABCGHIJK

Рис. 34. Примеры «Счастливых путей HappyPath»

В данном кейсе счастливый путь – это путь, который начинается событием обращение клиента (A), а заканчивается событием сделка подписана (K) или кредитный договор закрыт (Q).

В результате получили 2883 «счастливых путей», это примерно 1,9% от общего количества экземпляров процессов – 2883/151896=0,019.

На основе данных, представленных на рис. 32 (отклонение слов от эталонного), можно делать сводные таблицы, рассчитывая средние и медианные отклонения путей от эталона и любые прочие метрики, которые могут быть полезны аналитику.

Добавим подмодель для вычисления KPI. Рассчитаем показатели эффективности, описывающие процент разнообразия путей, вариацию отклонения от эталона и долю несоответствия путей эталонному пути.

Разнообразие путей = (количество путей) / количество уникальных экземпляров процесса. Для используемых данных процент разнообразия путей=11,95%.

9.0 Количество уникальных ID	12 Количество Слово процесса	9.0 Процент разнообразия путей
151 896,00	18 158	11,95

Рис. 35. Процент разнообразия путей

Вариация отклонений от эталона рассчитывается как стандартный коэффициент вариации на основе данных по отличию путей.

Коэффициент вариации = стандартное отклонение / математическое ожидание.

Для используемых данных = 5.16 / 6.39 = 0.802. Метрика вариации сама по себе является мало что говорящим значением.

По степени вариации можно судить об однородности совокупности, устойчивости значений признака, типичности средней, о взаимосвязи между какими-либо признаками. Вариацию следует рассчитать для различных аналитик, сравнивая их можно видеть разницу: на сколько отличаются реальные пути процесса от запроектированного эталона.

Следующий этап PM – Enhancement –улучшение включает определение приоритетов в работе над выявленными проблемами и продолжение анализа метрик для оценки эффективности внедрённых инициатив.

Далее проводится Monitoring – отслеживание, финальный этап, цель которого заключается в наблюдении за правильностью течения обновленного процесса.

Применяя технологии Process Mining в бизнес-процессах, можно сэкономить до 50% времени. Process Mining чётко показывает места для приложения усилий по оптимизации бизнес-процессов. Анализ данных журнала логов с помощью Process Mining позволил выявить и в дальнейшем сократить операции, которые были дополнительной нагрузкой для клиентов, а также уменьшить время ответа от банка и в конечном счете сохранить клиентов.

2.2. Использование технологии Process Mining в медицине

Process Mining можно применить не только в коммерческих компаниях, но и в организациях государственного сектора, в учебных и даже медицинских заведениях. Положительный опыт внедрения Process Mining в учреждениях, оказывающих медицинские услуги в мире, уже имеется [7].

Технология Process Mining в медицине может применяться для анализа и оптимизации различных медицинских процессов, таких как диагностика, лечение, госпитализация и выписка пациентов. Цель применения Process Mining в медицине – улучшение эффективности и качества медицинских услуг, а также снижение затрат на их предоставление.

Process Mining может быть использован для анализа следующих процессов в медицине:

- **Диагностика:**

Анализ данных о маршрутизации пациентов, обследованиях, диагностических процедурах и телах, чтобы выявить узкие места и оптимизировать процесс.

- **Лечение:**

Анализ данных о лечении пациентов, включая рецепты, анализы, лекарственные препараты и процедуры, чтобы выявить наиболее эффективные методы лечения и оптимизировать процесс.

- **Госпитализация:**

Анализ данных о госпитализации пациентов, включая время ожидания и поступления, чтобы выявить узкие места и оптимизировать процесс.

- **Выписка:**

Анализ данных о выписке пациентов, включая время ожидания и освобождение мест, чтобы выявить узкие места и оптимизировать процесс.

- **Медицинская документация:**

Анализ данных о медицинской документации, включая записи в электронных картах здоровья, чтобы выявить узкие места и оптимизировать процесс.

В процессе применения Process Mining в медицине можно использовать различные инструменты и методы, такие как:

- **Анализ данных:**

Анализ данных о медицинских процессах, чтобы выявить узкие места и отклонения от стандартного процесса.

- **Визуализация процессов:**

Визуализация процессов в виде диаграмм, чтобы лучше понять их структуру и выявить узкие места.

- **Автоматизированная оптимизация:**

Автоматизированная оптимизация процессов с использованием машинного обучения и искусственного интеллекта.

Применение Process Mining в медицине может привести к следующим результатам:

- **Улучшение качества медицинских услуг:**

Анализ процессов может помочь выявить области, где качество услуг может быть улучшено.

- **Снижение затрат на медицинские услуги:**

Оптимизация процессов может помочь снизить затраты на предоставление медицинских услуг.

- **Повышение эффективности медицинских учреждений:**

Анализ процессов может помочь выявить области, где эффективность работы может быть повышена.

- **Улучшение координации между различными медицинскими учреждениями:**

Process Mining может помочь выявить области, где улучшение координации между медицинскими учреждениями может быть достигнуто.

В системах здравоохранения ряда стран разработана и активно применяется методология инцидентного анализа, где под инцидентом понимается событие или обстоятельство (триггер), которое могло привести или привело к

причинению непреднамеренного вреда пациенту при оказании медицинской помощи.

Важнейшим этапом инцидентного анализа является определение мер по устранению или контролю выявленных системных причин. Цель применения Process Mining – понять и улучшить поток оказания услуги в медицинских учреждениях, проанализировав записи, разбросанные по сотням таблиц базы данных. Process Mining может помочь ответить на следующие вопросы:

– Почему больные так долго ждут операции;

– Следуют ли врачи регламенту;

– Можем ли мы предсказать время ожидания;

– Сколько людей нужно и где;

– Как снизить затраты (времени, денег);

– Как на самом деле используются медицинские приборы;

– Почему они ломаются;

– Какие компоненты нужно заменить;

– Можем ли мы предсказать поломку;

– Какие части нужно улучшить;

– Понять и улучшить работу и характеристики приборов, используя данные, полученные от сети сенсоров.

Для реализации технологий Process mining в медицине также можно использовать аналитическую платформу Loginom (см. рис. 3).

Российские нормативные требования к организации внутреннего контроля качества и безопасности медицинской деятельности обязывают медицинскую организацию проводить учет и анализ нежелательных событий при осуществлении медицинской деятельности.

Подавляющее большинство нежелательных событий в лечебных учреждениях и в больницах, в частности, связано с развитием осложнений оказания медицинской помощи. На рубеже XXI века в Институте совершенствования здравоохранения США была разработана методика «Глобальной оценки триггеров» позволяющая выявлять осложнения лечения, ретроспективно

анализируя случайную выборку данных законченных случаев лечения в стационаре. На рис. 36 представлена блок-схема процесса обнаружения и предотвращения внутри больничных осложнений, связанных с лечением пневмонии.

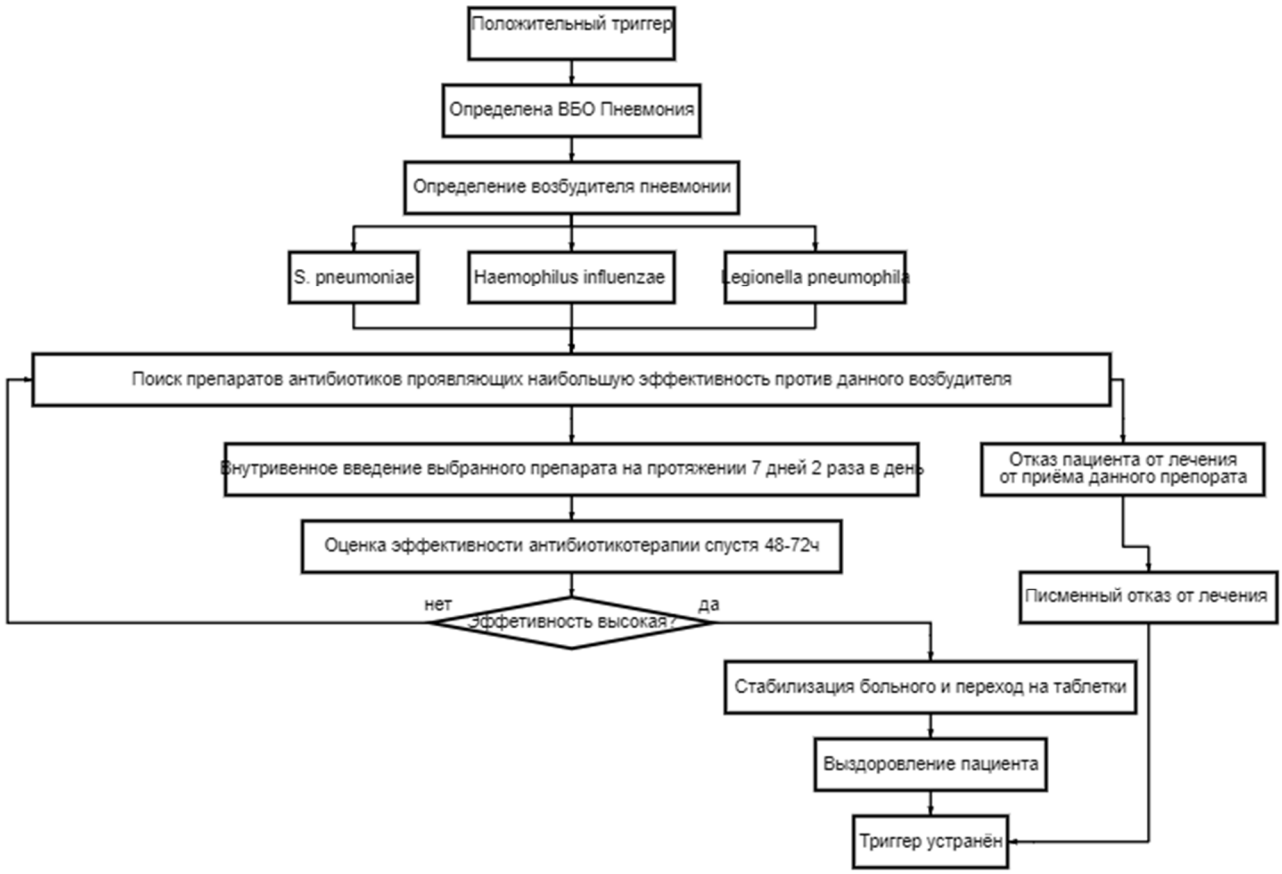


Рис. 36. Блок-схема процесса обнаружения и предотвращения внутри больничных осложнений, связанных с пневмонией

Используя результаты исследования методов лечения внутрибольничной пневмонии, был сгенерирован журнал событий, содержащий записи, показывающие действия в информационной системе больницы за некоторый период времени. Объём выборки составил 20897 записей.

#	12 Process_ID_Больного	ab Event_Действие	TimeStamp	ab Вид_Больного	ab Лечащий_Врач	ab Вид_Стационар
1	5 078 917	Получен положительный триггер	07.01.2010, 05:14	Бесплатный	Петров И.И.	Полный
2	5 078 917	Определение типа вещества	07.01.2010, 05:18	Бесплатный	Петров И.И.	Полный
3	5 078 917	Изучение дефектов вещества	07.01.2010, 05:21	Бесплатный	Петров И.И.	Полный
4	5 078 917	Развитие у пациента Нежелательной лекарственной реакции	07.01.2010, 16:37	Бесплатный	Петров И.И.	Полный
5	5 078 917	Определение вида Нежелательной лекарственной реакции	07.06.2010, 16:40	Бесплатный	Петров И.И.	Полный
6	5 078 917	Предовтратимость Нежелательной лекарственной реакции	07.06.2010, 16:42	Бесплатный	Петров И.И.	Полный
7	5 078 917	Изменение вида применяемого вещества	07.06.2010, 16:54	Бесплатный	Петров И.И.	Полный
8	5 078 917	Предовращение использование данного вещества для больного	07.06.2010, 17:54	Бесплатный	Петров И.И.	Полный
9	5 078 917	Триггер устранён	07.06.2010, 17:59	Бесплатный	Петров И.И.	Полный
10	5 078 921	Получен положительный триггер	07.01.2010, 01:09	Платный	Басуров В.В.	Полный
11	5 078 921	Определение типа вещества	07.01.2010, 01:10	Платный	Басуров В.В.	Полный
12	5 078 921	Изучение дефектов вещества	07.01.2010, 01:10	Платный	Басуров В.В.	Полный
13	5 078 921	Нежелательной лекарственной реакции не найденно	07.01.2010, 01:30	Платный	Басуров В.В.	Полный
14	5 078 921	Изменение вида применяемого вещества	07.01.2010, 01:51	Платный	Басуров В.В.	Полный
15	5 078 921	Предовращение использование данного вещества для больного	07.01.2010, 02:12	Платный	Басуров В.В.	Полный
16	5 078 921	Триггер устранён	07.01.2010, 03:10	Платный	Басуров В.В.	Полный
17	5 078 978	Получен положительный триггер	07.01.2010, 01:19	Платный	Вилков И.С.	Дневной
18	5 078 978	Определение падения	07.01.2010, 01:40	Платный	Вилков И.С.	Дневной
19	5 078 978	Определение причины падения	07.01.2010, 02:00	Платный	Вилков И.С.	Дневной
20	5 078 978	Изучение внутренних факторов падения	07.01.2010, 02:21	Платный	Вилков И.С.	Дневной
21	5 078 978	Больной не принимал лекарственных препаратов	07.01.2010, 03:21	Платный	Вилков И.С.	Дневной

Рис.37. Датасет по медицинскому стационару

Исходный файл данных содержит следующую информацию:

1) Process_ID_Больного

2) Event_Действие

3) TimeStamp (Начало времени действия)

4) Data_End (Конец времени действия)

5) Вип_Больного (Платный/Бесплатный)

6) Вид_Стационара (Полный/Дневной)

На этапе аудита данных удалим записи, которые не являются процессами, то есть их Event_Действие равно 1. Неполные данные – это те процессы (9% записей), у которых не закончены действия или действия вырваны из идеальных процессов.

Путь описывает выполнение одного конкретного экземпляра или случая зарегистрированного процесса. Пример пути можно увидеть на рис. 38.

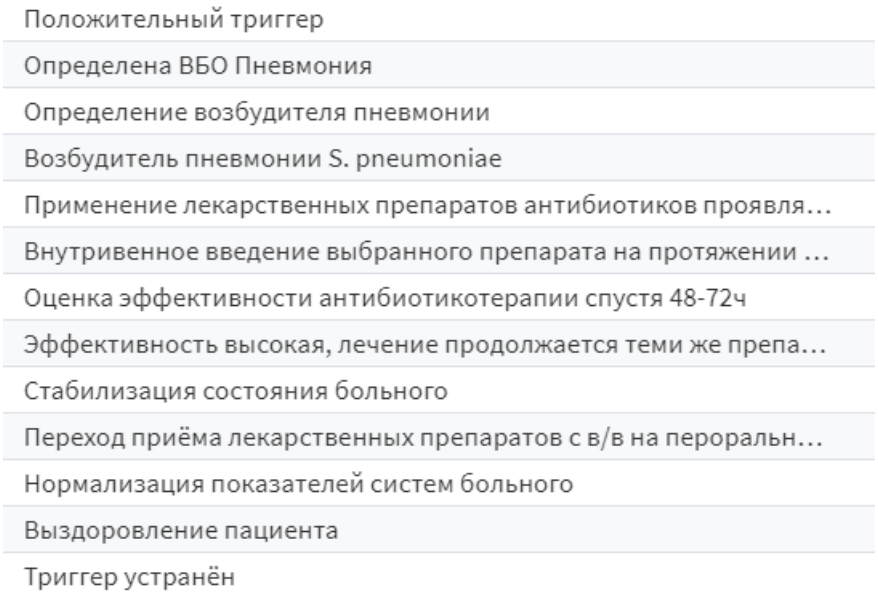


Рис. 38. Пример пути

Следующим шагом является расчёт основных метрик Process Mining: количество шагов процесса; первый и последний шаг процесса; длительность события; нарастающая длительность события; следующее по очереди событие.

Далее переходим к формированию букв и слов процесса. Буквенное обозначение событий позволяет делать процессный лог более компактным и позволяет сильно упростить технологию анализа процессов. Соответствие представлено в таблице 4.

Таблица 4 – Соответствие между событиями и буквами событий

Положительный триггер	A
Определена ВБО Пневмония	B
Определение возбудителя пневмонии	C
Возбудитель пневмонии S. pneumoniae	D
Применение лекарственных препаратов антибиотиков, проявляющих наибольшую эффективность против данного возбудителя	E
Внутривенное введение выбранного препарата на протяжении 7 дней 2 раза в день	F
Оценка эффективности антибиотикотерапии спустя 48-72ч	G
Эффективность высокая, лечение продолжается теми же препаратами	H
Стабилизация состояния больного	I
Переход приёма лекарственных препаратов с в/в на пероральный приём.	J
Нормализация показателей систем больного	K
Выздоровление пациента	L

Триггер устранён	M
Возбудитель пневмонии Haemophilus influenzae	N
Возбудитель пневмонии Legionella pneumophila	O
Эффективность низкая, смена препарата	P
Состояние больного не нормализовалась, применение другого препарата	Q
Пациент отказался от приёма данного препарата	R
Посменный отказ от лечения	S
Внутривенное введение выбранного препарата на протяжении 5 дней 2 раза в день	T

Все последовательности событий начинаются с шага A и заканчиваются шагом M (кроме последовательностей с неполными данными). Примеры возможных последовательностей событий: ABCNEFGHQJKLM, ABCDERSM, ABCOERSM, ABCNER, ABCDETGHIJKLM.

Подмодель маркировки путей представлена на рисунке 39.

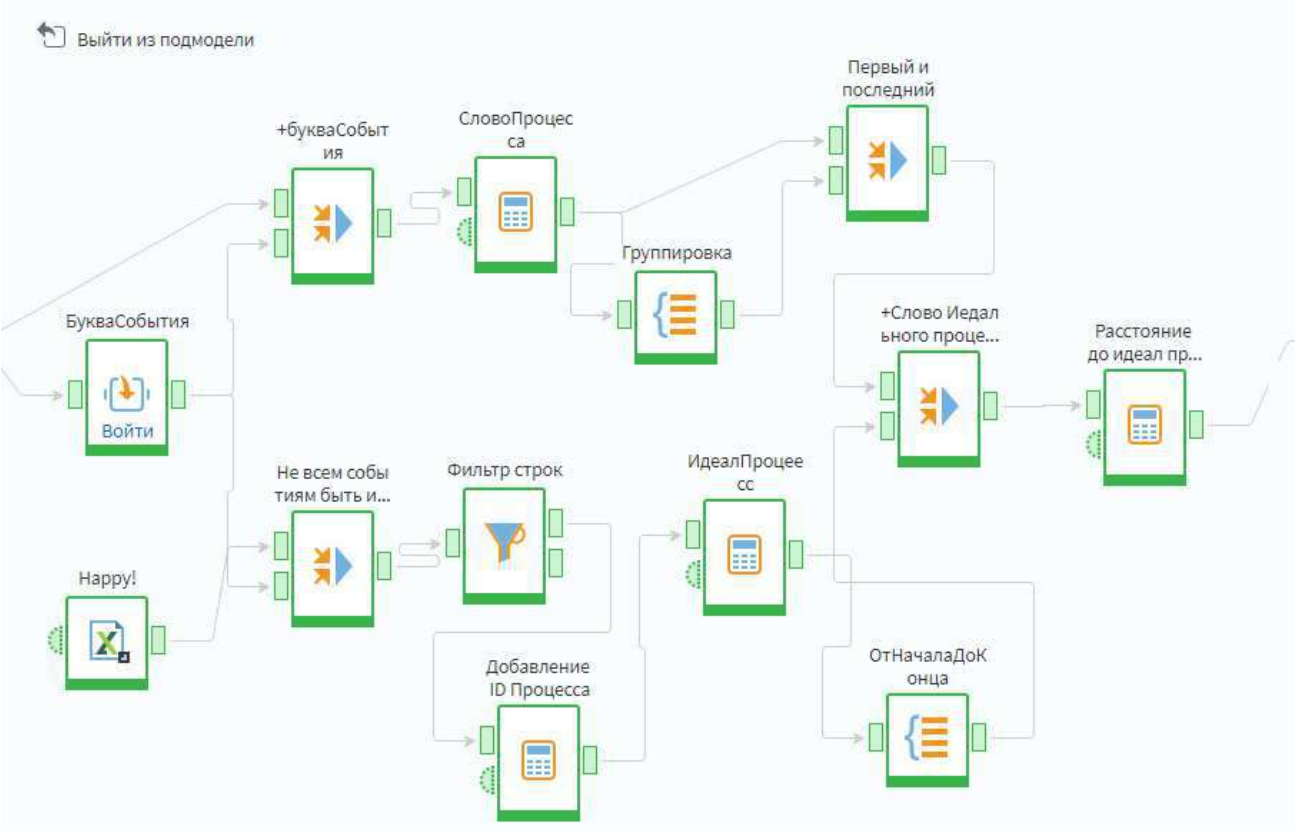


Рис. 39. Подмодель маркировки путей и их анализа

Анализ последовательностей событий показал, что в выборке уникальных путей оказалось 22 (рис. 40).

#	ab УникальноеСлово|Последний	12 Process_ID_Больного|Количество
1	ABCDEFGHIJKLM	322
2	ABCNEFGHIJKLM	184
3	ABCOEFGHIJKLM	216
4	ABCDEFGPHIJKLM	70
5	ABCOEFGPHIJKLM	74
6	ABCDEFGHQJKLM	130
7	ABCOEFGHQJKLM	104
8	ABCNEFGHQJKLM	44
9	ABCDERSM	54
10	ABCOERSM	18
11	ABCNERSM	36
12	ABCDETGHIJKLM	78
13	ABCOETGHIJKLM	78
14	ABCNETGHIJKLM	52
15	DEFGPHIJKLMABC	56
16	OEFGPHIJKLMABC	32
17	OEFGPHIJKLM	8
18	ABCNEFGH	8
19	QJKLMABC	8
20	DERSMABC	24
21	OERSMABC	8
22	NERSMABC	8

Рис. 40. Уникальные пути

Каждый уникальный путь имеет свою последовательность событий, в зависимости от того, как они отличаются от эталонного пути – оптимального порядка действий, приводящих к выздоровлению пациента, были определены следующие группы (рис. 41):

ab Слово_процесса (последнее событие)	9.0 Количество	9.0 Процент
Верное лечение	1000	62
Лечение требовало дополнительных анализов	144	9
Больной отказался от лечения	108	7
Отклонение от стандартов лечения	208	13
Неполные данные	152	9

Рис. 41. Статистика по группам процессов

- «Верное лечение», 1000 процессов.
- «Лечение требовало дополнительных анализов», 144 процесса.
- «Больной отказался от лечения», 108 процессов.

- «Отклонение от стандартов лечения», 208 процессов.
- «Неполные данные», 152 процесса.

Верное лечение – это те процессы (их выявлено 62%), которые выполнялись согласно правильному лечению и в итоге больной выздоравливал.

Лечение требовало дополнительных анализов – это те процессы (9% записей), в которых больным давали не те антибиотики, что увеличивало время их лечения.

Больной отказался от лечения – это те процессы (7% записей), когда больной остался невылеченным, но покинул больницу.

Отклонение от стандартов лечения – это те процессы (13% записей), когда лечение было выполнено неверно. В качестве примера можно увидеть, что некоторые антибиотики принимались 5 дней место 7.

В статистических данных было рассчитано:

1) Количество уникальных процессов от первого до последнего действия;

2) Число действий, выполненных на одну конкретную дату;

3) Количество уникальных экземпляров процессов и их доля.

За эталонный путь возьмем путь, который принадлежит группе «Верное лечение» и приводит к желаемому результату. Данный путь: ABCDEFGHIJKLM загрузим в сценарий АП Loginom через подключение новой таблицы Excel (рис. 42).

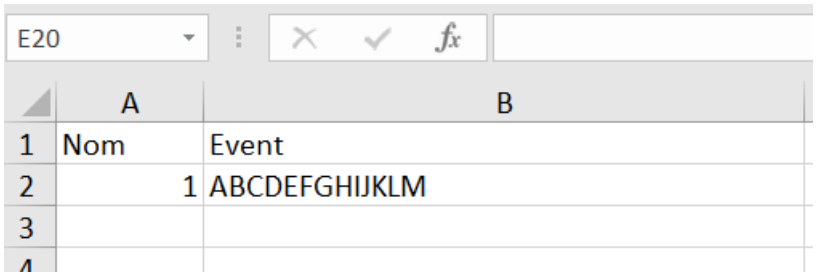


Рис. 42. Таблица Excel: Reference path

Определим отклонения процессов в журнале лога от эталонного пути с помощью расстояния Левенштейна (рис.43.)

ab Process_ID_Больного	ab СловоПроцесса	9.0 Расстояние Левенштейна
5142840	NERSMABCNERSM	5,00
5365062	NERSMABCNERSM	5,00
5698395	NERSMABCNERSM	5,00
6253950	NERSMABCNERSM	5,00
7031728	NERSMABCNERSM	5,00
7253950	NERSMABCNERSM	5,00
7587283	NERSMABCNERSM	5,00
8142838	NERSMABCNERSM	5,00
5142776	ABCOEFGHIJKLMABC	3,00
5364998	ABCOEFGHIJKLMABC	3,00
5698331	ABCOEFGHIJKLMABC	3,00
6253886	ABCOEFGHIJKLMABC	3,00

Рис. 43. Отклонение путей от эталонного

Было определено, что 84% путей в датасете совпадают с эталонным.

Для вычисления счастливых путей добавим в сценарий новую подмодель «Счастливый путь» (рис. 44).

ab Счастливый путь	9.0 Их количество в датасете
ABCDEFGHIJKLM	322
ABCOEFGHIJKLM	216
ABCNEFGHIJKLM	184
ABCDEFGHQJKLM	130
ABCDEFGPHIJKLM	126
ABCOEFGPHIJKLM	106
ABCOEFGHQJKLM	104
ABCDERSM	78
ABCDETGHIJKLM	78
ABCOETGHIJKLM	78
ABCNETGHIJKLM	52
ABCNEFGHQJKLM	44
ABCNERSM	44
ABCOERSM	26

Рис. 44. Счастливые пути

В ходе проведенного процессного анализа было выявлено, что некоторые действия сильнее влияют на выздоровление больного и эффект от них гораздо выше, к примеру, использование нужных антибиотиков при различ-

ном виде пневмонии. В ходе исследования было обнаружено, что только 62% всех больных лечатся верно. Были найдены отклонения от лечения 13%. Зафиксированы случаи, когда больные лечились антибиотиками не подходящего вида, тем самым увеличивая время лечения и затраты больницы на дополнительные лекарства, а также случаи, когда больные самостоятельно отказывались от лечения.

Благодаря технологии Process Mining можно оптимизировать следующие процессы:

- повышение качества оказания услуг больным;
- снижение внутри больничных осложнений в целом;
- перераспределение ресурсов больницы;
- уменьшение промежутков времени между событиями;
- автоматизацию отдельных событий, например, внесение данных в специальную базу;
- сокращение трудозатрат на выполнение задач.

В зависимости от целей и ситуации можно вносить изменения в разработанный сценарий для усовершенствования системы.

2.3. Использование технологии Process Mining в образовании

Принятие решений в сфере образования – сложный, многоаспектный процесс, в который вовлечен большой круг заинтересованных лиц. В течение длительного периода в информационных системах учебных заведений накапливается информация о различных аспектах образовательного процесса: об учащихся и их успеваемости, о преподавателях и их научно-образовательной работе.

В связи с ростом применения информационных технологий в образовании возник интерес к автоматизированному выявлению новых взаимосвязей в данных и их интерпретации. Цель кейса – показать возможности применения методологии Process Mining для анализа образовательных данных.

Educational Process Mining (EPM) – технология извлечения, мониторинга и усовершенствования реальных процессов с помощью извлечения новых знаний из журналов событий, которые доступны в современных информационных системах [8].

Образовательные данные содержат информацию, состоящую из организационных и дидактических мероприятий образовательного процесса. Эти данные предоставляют информацию об успеваемости на основе государственных стандартов образования.

Например, данные об образовании могут включать в себя следующую информацию:

- доступ учащихся к учебным материалам в сверхурочное время в системе управления обучением (например, темы, ресурсы);
- взаимодействие учащихся с образовательной платформой в сверхурочное время (например, вход в систему, выход из системы, видео, тест);
- процесс решения учеником задачи (например, использование калькулятора, исключение ответов, сброс вопроса).

На рисунке 45 представлена диаграмма потоков данных процесса получения образовательных данных.

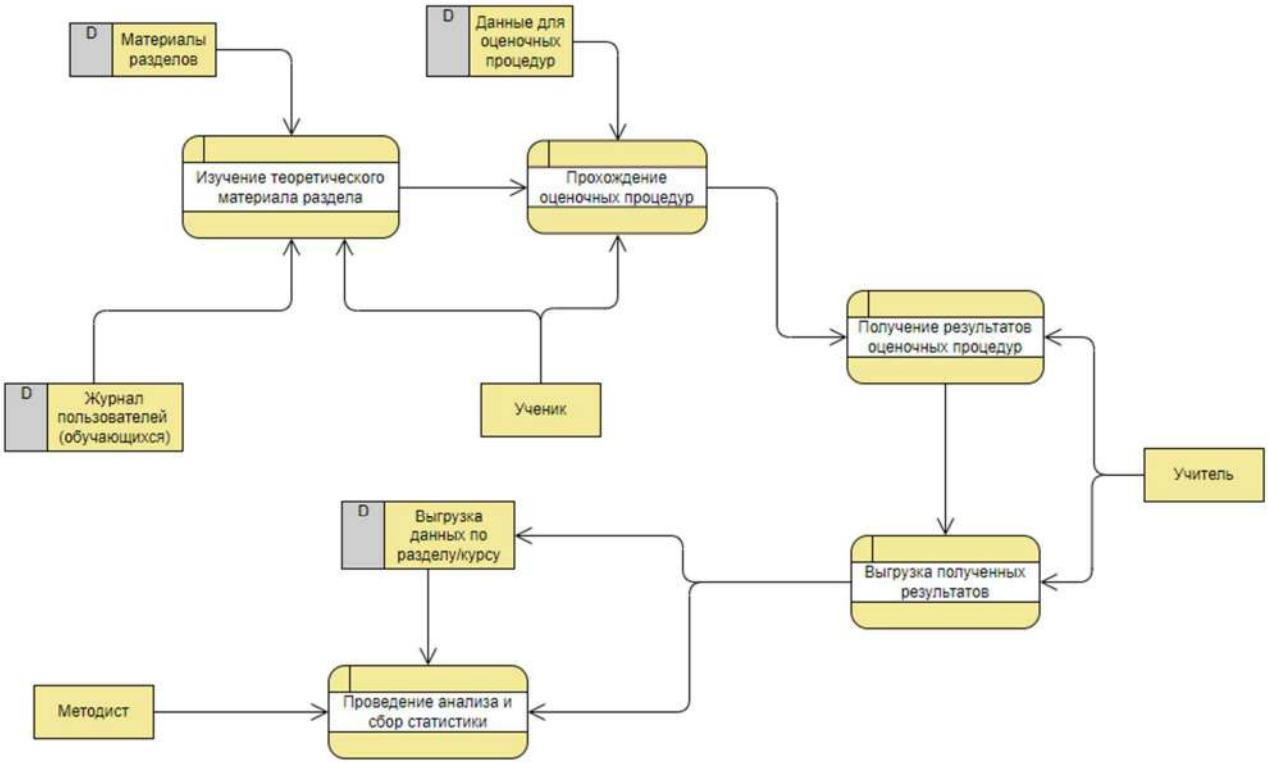


Рис. 45. DFD диаграмма, описывающая этапы работы с онлайн курсом

Процесс выглядит следующим образом. В самом начале обучающийся, входит в раздел курса, изучает теоретическим материал, который в этом разделе размещён (это могут быть презентации, картинки, видео, книги и статьи и т.д.) После изучения теоретических материалов необходимо провести проверку остаточных знаний с помощью прохождения оценочных процедур (тесты, эссе, вопросы с развёрнутым ответом и т. д.) Далее после того, как обучающийся прошёл раздел курса, преподаватель получает данные о прохождении раздела. Платформа выдаёт некоторый набор данных, которые можно выгрузить и отправить для анализа администратору курса.

При внедрении технологии Process Mining, модель потоков данных изменится следующим образом.

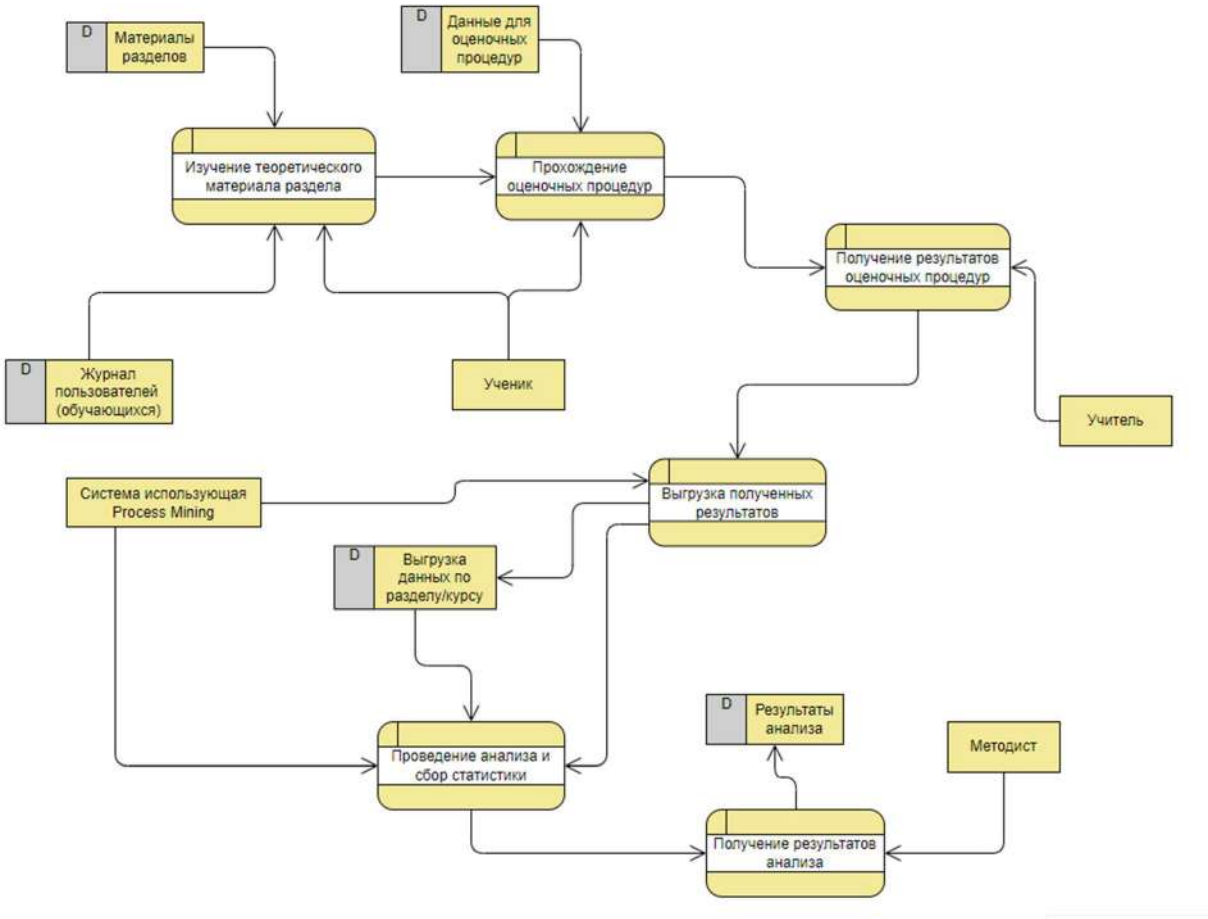


Рис. 46. Модель, описывающая этапы работы с онлайн курсом после внедрения Process Mining

Внедрение технологии Process Mining в образование позволит получить очевидные преимущества в виде понимания и улучшения:

- индивидуальных привычек обучающихся;
- ключевых показателей эффективности (KPI);
- основных компетенций учащихся;
- предоставления консультаций для обучающихся и преподавателей;
- оптимизации образовательных материалов и рабочих программ.

Постановка цели и задач для реализации технологии Process Mining.

Одной из актуальных проблем EPM является задача получения журнала логов событий. Такой журнал можно будет получить с помощью выгрузки данных из онлайн платформы обучения. В журнале должны быть указаны

данные об ученике, раздел, который он проходил и информацию по срокам прохождения раздела.

Далее нужно определить эталонный путь для рассматриваемого процесса. Такой путь представляет из себя полное прохождение всех разделов курса на оценку выше «2». То есть в контексте прохождения онлайн курса, эталонный процесс – последовательное успешное закрытие оценочных процедур по каждому разделу курса.

В итоге необходимо создать сценарий на платформе Loginom, который позволит, подавая на вход данные из выгрузки с онлайн платформы обучения, возможно с небольшими корректировками, получать выгрузку разных статистических данных, а также некоторых выводов по курсу в целом, например, есть ли слишком лёгкие разделы, которые необходимо усложнить или напротив, есть потенциально сложные разделы, где нужно добавить теоретический материал или изменить оценочную процедуру.

Одним из главных требований к сценарию является универсальность. То есть, если входные данные изменятся, без изменения структуры, то сценарий должен работать без дополнительных настроек.

Также должна присутствовать возможность изменения и дополнения сценария, для развития процесса анализа или в случае изменения структуры выходных данных.

Для обеспечения удобства пользователей создаваемой системы на основе сценария, выгруженные данные должны храниться в отдельной директории, как и данные, загружаемые в сценарий.

Создаваемая система должна обладать проверками соответствия структуры данных и не пропускать до анализа «грязные» данные. Для обеспечения этого можно воспользоваться простым механизмом фильтрации, поскольку данные получаются при выгрузке из онлайн платформы и как правило не имеют ошибок, связанных с человеческим фактором. К тому же есть возможность повторной выгрузки и перезаписи данных, в случае возникновения системных ошибок.

Все вышеперечисленные требования к системе на основе сценария в Loginom позволят выполнить главную задачу – автоматизацию процесса проведения анализа образовательных данных.

User(Proces...	Event	Timestamp	Timestamp Fin
2000101	Табличный процессор. Основные сведения	06.09.2022	09.09.2022
2000101	Табличный процессор. Основные сведения	10.09.2022	11.09.2022
2000101	Редактирование и форматирование в табличном процессоре	14.09.2022	14.09.2022
2000101	Входная контрольная работа	21.09.2022	22.09.2022
2000101	Встроенные функции и их использование	29.09.2022	30.09.2022
2000101	Логические функции Инструменты анализа данных	05.10.2022	06.10.2022
2000101	Обобщение и систематизация изученного материала по теме «Обработка информации...	13.10.2022	15.10.2022
2000101	Основные сведения об алгоритмах	20.10.2022	22.10.2022
2000101	Алгоритмические структуры	25.10.2022	06.11.2022
2000101	Запись алгоритмов на языке программирования Паскаль	08.11.2022	13.11.2022

Рис. 47. Фрагмент журнала логов

- «User» – поле, куда записывается сначала ФИО ученика, а потом заменяется на id;
- «Process_id» –уникальный идентификатор пользователя, который дублирует информацию из поля User;
- «Event» – наименование задания;
- «Timestamp» – дата, когда было выдано задание;
- «Timestamp Fin» – дата, когда ученик выполнить задание.

Из-за того, что оценки могут быть пересмотрены учителем, то есть должна быть возможность вносить изменения именно в оценки, было принято решение записывать их в отдельную таблицу, а потом объединять с журналом событий. Для этого был использован компонент «Слияние» в Loginom.

Получение оценок • Быстрый просмотр

Выходной набор данных

#	Process_id	Event	Timestamp Fin	Timestamp	User	Score
1	2 000 101	Табличный процессор. Основные сведения	09.09.2022, 00:00	06.09.2022, 00:00	2000101	2
2	2 000 101	Табличный процессор. Основные сведения	11.09.2022, 00:00	10.09.2022, 00:00	2000101	5
3	2 000 101	Редактирование и форматирование в табличном процессоре	14.09.2022, 00:00	14.09.2022, 00:00	2000101	3
4	2 000 101	Входная контрольная работа	22.09.2022, 00:00	21.09.2022, 00:00	2000101	5
5	2 000 101	Встроенные функции и их использование	30.09.2022, 00:00	29.09.2022, 00:00	2000101	4
6	2 000 101	Логические функции Инструменты анализа данных	06.10.2022, 00:00	05.10.2022, 00:00	2000101	3
7	2 000 101	Обобщение и систематизация изученного материала по теме «О…	15.10.2022, 00:00	13.10.2022, 00:00	2000101	5
8	2 000 101	Основные сведения об алгоритмах	22.10.2022, 00:00	20.10.2022, 00:00	2000101	5
9	2 000 101	Алгоритмические структуры	06.11.2022, 00:00	25.10.2022, 00:00	2000101	4
10	2 000 101	Запись алгоритмов на языке программирования Паскаль	13.11.2022, 00:00	08.11.2022, 00:00	2000101	2
11	2 000 101	Запись алгоритмов на языке программирования Паскаль	14.11.2022, 00:00	14.11.2022, 00:00	2000101	2
12	2 000 101	Запись алгоритмов на языке программирования Паскаль	17.11.2022, 00:00	15.11.2022, 00:00	2000101	3
13	2 000 101	Анализ программ с помощью трассировочных таблиц	19.11.2022, 00:00	19.11.2022, 00:00	2000101	4
14	2 000 101	Функциональный подход к анализу программ	27.11.2022, 00:00	26.11.2022, 00:00	2000101	5
15	2 000 101	Структурированные типы данных. Массивы	04.12.2022, 00:00	04.12.2022, 00:00	2000101	5
16	2 000 101	Структурное программирование	10.12.2022, 00:00	09.12.2022, 00:00	2000101	4
17	2 000 101	Рекурсивные алгоритмы	17.12.2022, 00:00	17.12.2022, 00:00	2000101	5
18	2 000 101	Обобщение и систематизация изученного материала по теме «А…	25.12.2022, 00:00	24.12.2022, 00:00	2000101	4
19	2 000 101	Модели и моделирование	06.01.2023, 00:00	04.01.2023, 00:00	2000101	4
2 799	2 000 101	Моделирование на графах	16.01.2023, 00:00	14.01.2023, 00:00	2000101	5

Рис. 48. Добавили поле (Оценка)

Также, поскольку курсы для трёх разных классов отличаются, соответственно отличаются данные полей (Material) и ученики (Process_id), а также информация по курсам хранится в разных таблицах, то их нужно объединить в одну таблицу с помощью компонента «Объединение». Эти данные будут нужны для аналитической работы, поэтому была предусмотрена выгрузка в единый файл Excel со всеми курсами за 9–11 класс.

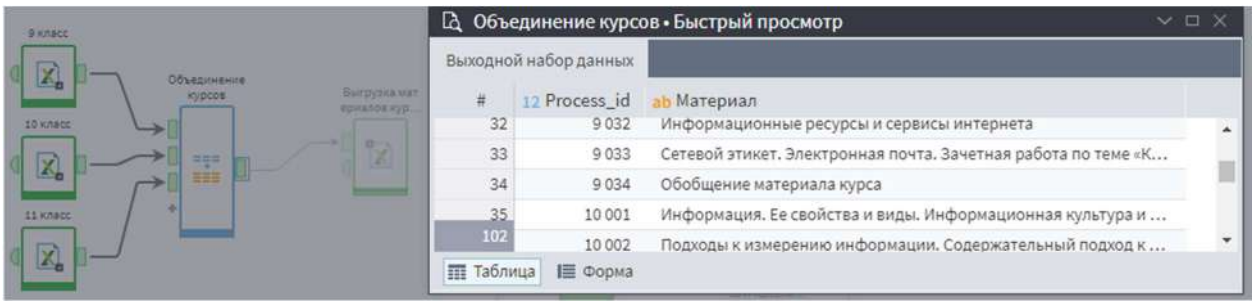


Рис. 49. Получение файла со всеми курсами

Далее были предусмотрены выгрузки данных, которые не подходят для исследования или содержат ошибку. Таким образом в проекте появилась директория «errors» в которой содержатся выгрузки:

- мало шаговых процессов, то есть тех процессов, которые содержат меньше одного события;

- строк с отрицательной длительностью события, что может возникнуть при ошибочном заполнение данных.

Также были выполнены все основные этапы разведочного анализа (очистка данных, фильтрация, кодирование, форматирование) (рис. 50).

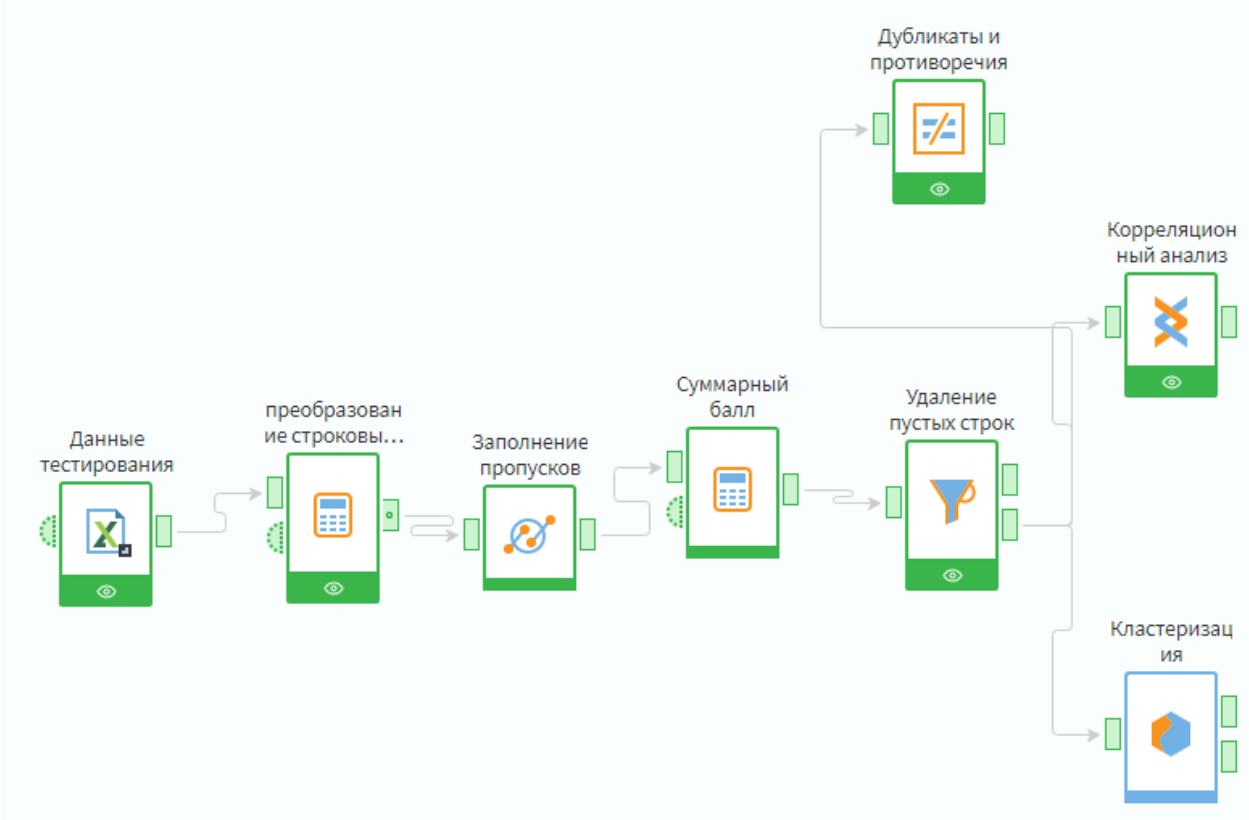


Рис. 50. Загрузка данных, аудит и разведочный анализ

Тип	Метка	Вид	Проблемы	Виды проблем	
ab	Учреждение (орга...		17,99%	Пропуски - 16,40% (62)	Экстремальные - 1,59% (6)
ab	Тест:ИТОГОВЫЙ Т...		4,50%	Выбросы - 4,50% (17)	
ab	Тест:2.4. Подпрост...		3,17%	Выбросы - 3,17% (12)	
ab	Тест:2.1. Введение...		2,91%	Выбросы - 2,91% (11)	
ab	Тест:2.2. Векторно...		2,12%	Выбросы - 2,12% (8)	
ab	Тест:2.3. Евклидов...		2,12%	Выбросы - 2,12% (8)	
ab	Тест:Входное тест...		1,85%	Выбросы - 1,85% (7)	

Рис. 51. Визуализатор «Качество данных»

Данные строковые, так как содержат "-" , нужно их переконвертировать в числа и пропущенные значения заменить нулями.

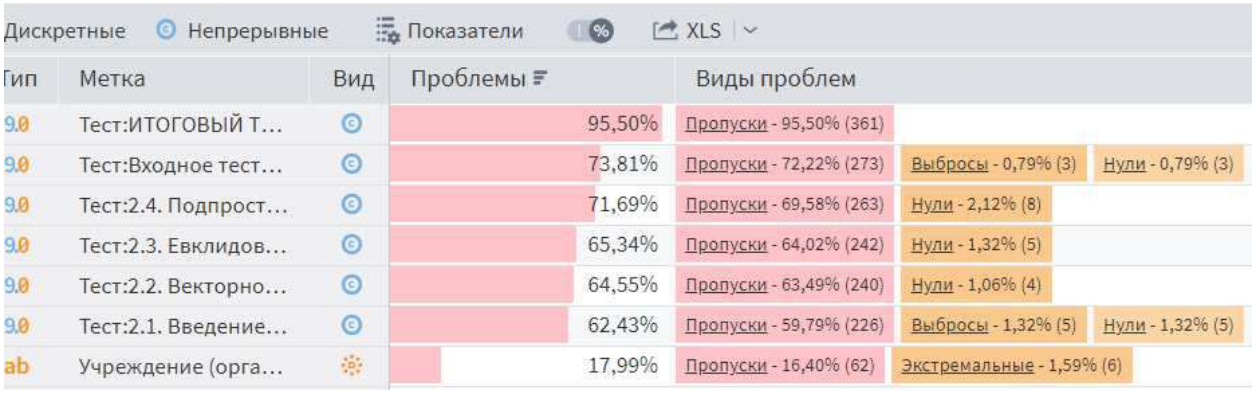


Рис. 51. Визуализатор «Качество данных». Пропуски

Чтобы мы смогли использовать эти данные для анализа и построения моделей:

- Надо заполнить пропуски нулями;

- убрать пустые/нулевые строки;

- убрать пустые/нулевые столбцы (если есть);

- добавить столбец с суммой всех баллов (обогащение).

Для определения нулевых строк сценарии были дополнительно выполнены действия (замена «Null» на 0, вычисление суммарный балла, удаление строк, где суммарный балл=0).

Анализ статистики числовых признаков (анализ гистограмм)

Диапазон	Доля	Кол-во	%
[0,33: 3,22625)		12	7,6
[3,22625: 6,1225)		12	7,6
[6,1225: 9,01875)		16	10,2
[9,01875: 11,915)		13	8,3
[11,915: 14,81125)		15	9,6
[14,81125: 17,7075)		50	31,8
[17,7075: 20,60375)		35	22,3
[20,60375: 23,5]		4	2,5

Рис. 52. Анализ статистики числовых признаков (анализ гистограмм)

Анализ гистограммы по суммарному баллу: 4 человека получили в сумме максимальные баллы 20–24, 50 человек получили 14–18 баллов.

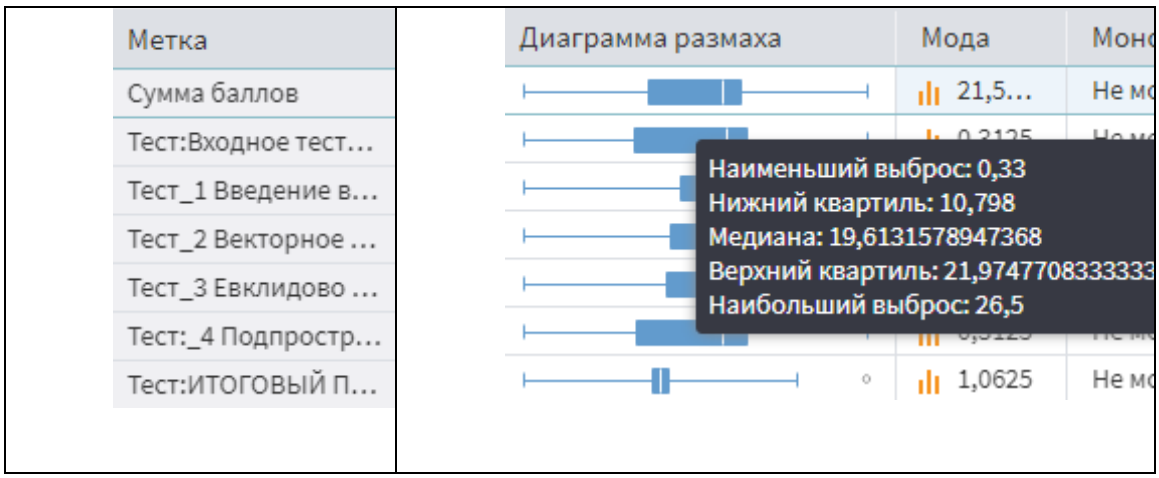


Рис. 53. Анализ статистики числовых признаков (Анализ boxplota)

Анализ boxplota для поля "sum": нижний квартиль примерно 10,7, верхний квартиль примерно 22, медиана 19,6, наблюдаемый максимум 26,5. Самый простой тест T_1, самый сложный T_4.

Для выполнения поставленных задач реализации технологии Process Mining на платформе Loginom требуется создать таблицу с кодификаторами наименований разделов курса (рис. 54).

Поскольку рассматриваются сразу несколько курсов необходимо указать к какому классу относится данный раздел (Class), также нужно указать раздел курса, чтобы связывать данную таблицу с другими (Material) и непосредственно сам код раздела курса в виде буквенного символа (Letter).

Касаемо поля Letter, в технологии Process Mining принято использовать символы из английского алфавита, но поскольку точно известно, что курсов сильно больше, чем символов латинского алфавита, было принято решение использовать кириллические символы.

ab Event	ab Буква процесса
Табличный процессор. Основные сведения	А
Редактирование и форматирование в табличном процессоре	Б
Входная контрольная работа	В
Встроенные функции и их использование	Г
Логические функции Инструменты анализа данных	Д
Обобщение и систематизация изученного материала по теме «Обработка информации…	Е
Основные сведения об алгоритмах	Ё
Алгоритмические структуры	Ж
Запись алгоритмов на языке программирования Паскаль	З
Анализ программ с помощью трассировочных таблиц	И
Функциональный подход к анализу программ	Й
Структурированные типы данных. Массивы	К
Структурное программирование	Л
Рекурсивные алгоритмы	М
Обобщение и систематизация изученного материала по теме «Алгоритмы и элементы п…	Н
Модели и моделирование	О
Моделирование на графах	П
Знакомство с теорией игр	Р
База данных как модель предметной области	С
Реляционные базы данных	Т
Системы управления базами данных	У
Проектирование и разработка базы данных	Ф
Обобщение и систематизация изученного материала по теме «Информационное модел…	Х
Основы построения компьютерных сетей	Ц
Как устроен Интернет	Ч
Службы Интернета	Ш
Интернет как глобальная информационная система	Щ

Рис. 54. Лист с кодами разделов курсов

Из систем с онлайн курсами также необходимо получить журнал с пользователями. Данный журнал нужен для идентификации id обучающегося и его ФИО, а также класса и группы. Поскольку тестовые данные обезличены, в них указывается два поля id пользователя (User) и класс, к которому ученик относится (Group).

User	Group
2000552	9a-1
2000650	9a-1
2000386	9a-1
2000658	9a-1
2000607	9a-1
2000190	9a-1
2000131	9a-1
2000619	9a-1
2000424	9a-1
2000586	9a-1
2000498	9a-1
2000520	9a-1
2000540	9a-1
2000527	9a-2
2000728	9a-2
2000425	9a-2
2000208	9a-2
2000503	9a-2
2000682	9a-2
2000710	9a-2
2000294	9a-2
2000513	9a-2
2000535	9a-2
2000117	9a-2
2000871	9a-2
2000375	10a-1
2000291	10a-1
2000410	10a-1
2000883	10a-1
2000784	10a-1
2000638	10a-1
2000129	10a-1
2000616	10a-1
2000250	10a-1
2000488	10a-1

Рис. 55. Лист пользователей

Осталось заполнить текстовый файл, где поочерёдно будут перечислены кодификаторы разделов курсов, тем самым создавая счастливый путь.

В случае прохождения онлайн курса, где каждый раздел открывается постепенно и есть чёткая последовательность этих разделов, эталонный путь в случае курса из 34 разделов счастливый путь выглядит так:

«АБВГДЕЁЖЗИЙКЛМНОПРСТУФХЦЧШЩЬЫЪЭЮЯ1»

Далее можно преступить к расчёту метрик и проведению проверок.

Для этого создаётся подмодель «Метрики и проверки» (рис. 56).

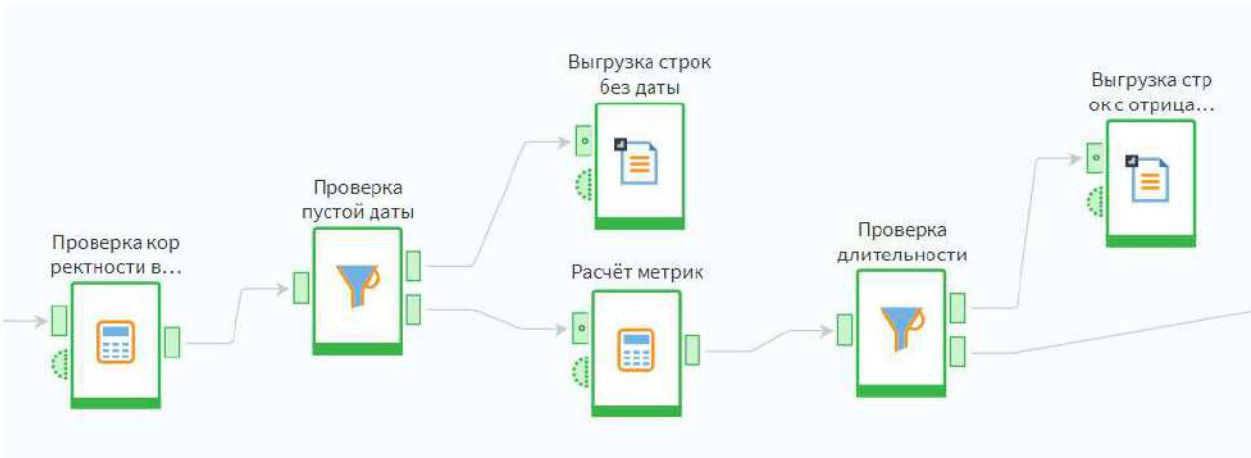


Рис. 56. Подмодель с расчётом метрик и проведению проверки корректности параметра времени

Для этого сначала нужно провести проверку корректности введённого временного параметра, то есть определить:

- время начала изучения раздела курса точно меньше времени окончания изучения раздела;
- заполнено ли время начала изучения раздела;
- заполнено ли время окончания изучения раздела;
- если что-то из вышеперечисленного не выполняется, то необходимо логически заполнить данные с помощью следующих вычислений.

Листинг 1. Проверка корректности временных параметров

```
IFF(Timestamp_Fin<Timestamp and IsNull(Timestamp_Fin)=False,Timestamp_Fin,Timestamp)
IFF(NVL(Timestamp_Fin,0)<Timestamp,Timestamp_Fin,Timestamp_Fin)
```

При выполнении расчётов будет получено 10 метрик:

- Счёт шагов – для расчёта используется функция, возвращающая накапливающуюся сумму, где «Размер шага» - поле, по которому необходимо получить накапливающуюся сумму, а полем, по которому происходит группировка данных, станет «Process_id» то есть сам ученик, точнее его процесс прохождения всего курса;
- Ключ – строка, объединяющая «Process_id» и название события;

- Последний шаг процесса – на последней строке, связанной с одним и тем же учеником будет фиксироваться количество разделов, которые он прошёл;
- Длительность события – количество затраченных дней (часов) на работу;
- Фактическая длительность события – разница во времени до текущего события до следующего события;
- Нарастающая длительность события – накапливающаяся сумма по предыдущей метрике;
- Следующее события – название следующего по очереди события;
- Количество событий процесса – сколько одинаковых событий было внутри процесса;
- Количество повторяющихся элементов – логическое выражение, при котором, если предыдущее событие будет равно следующему, то значение 1, иначе 0;
- Цикл – логическое выражение, при котором, если дата события совпадает с датой предыдущего события, то значение 1, иначе 0;
- В итоге производиться проверка ненулевой длительности события.

В результате корректные данные пойдут на дальнейшую обработку, а данные с пустой датой и с нулевой длительностью пойдут в отдельную выгрузку, причём, поскольку это ошибка в данных, то они будут храниться в отдельной директории – «errors».

Далее рассчитаем статистические показатели, полученные при реализации технологии Process Mining.

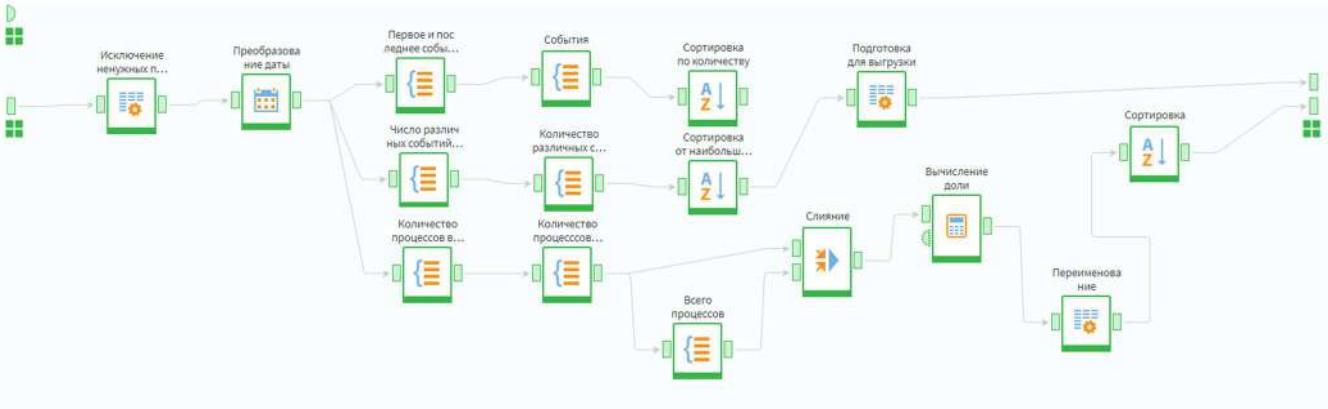


Рис. 57. Подмодель «Statis Process Mining»

Данная подмодель позволяет получить две статистические таблицы:

- Пройденные разделы в разрезе дней по количеству учеников;
- Проценты двоек за разделы курсов.

Statis Process Mining • Быстрый просмотр

Таблица 1 | Таблица 2

#	12 Количество ученико	ab Проходили событие	31 В этот день
1	25	Обобщение и систематизация изученного материала по теме «О...	16.05.2023, 00:00
2	25	Информационные ресурсы и сервисы интернета	16.05.2023, 00:00
3	25	Объекты компьютерной графики. Компьютерная графика и ее в...	16.05.2023, 00:00
4	14	Итоговая контрольная работа	18.05.2023, 00:00
5	12	Встроенные функции и их использование	27.09.2022, 00:00
6	12	Обработка информации. Задачи обработки информации. Коди...	05.10.2022, 00:00
7	12	Как устроен интернет. IP- адрес компьютера	19.04.2023, 00:00
8	11	Компьютерные презентации. Виды компьютерных презентаций...	18.05.2023, 00:00
9	11	Алгебра логики. Логические высказывания и переменные. Логи...	23.02.2023, 00:00
10	10	Логические функции Инструменты анализа данных	05.10.2022, 00:00
11	10	Как устроен Интернет	14.03.2023, 00:00
12	10	Одномерные массивы целых чисел. Описание массива. Использ...	09.12.2022, 00:00
13	10	Сетевой этикет. Электронная почта. Зачетная работа по теме «К...	18.05.2023, 00:00
14	10	Преобразование логических выражений. Логические функции. ...	11.04.2023, 00:00
15	10	Моделирование как метод познания.	21.09.2022, 00:00
16	10	Информация. Ее свойства и виды. Информационная культура и ...	07.09.2022, 00:00
17	10	Логические задачи и способы их решения. Метод рассуждений. ...	26.04.2023, 00:00
18	9	Знакомство с теорией игр	18.01.2023, 00:00
19	9	Обобщение и систематизация изученного материала по теме «С...	13.04.2023, 00:00
20	9	Итоговое повторение. Основные идеи и понятия курса	23.05.2023, 00:00
21	9	Обобщение и систематизация изученного материала по теме «О...	11.10.2022, 00:00
22	9	Информационное право	27.04.2023, 00:00
23	9	Интернет как глобальная информационная система	05.04.2023, 00:00
24	9	Интерфейс электронных таблиц. Данные в ячейках таблицы	01.03.2023, 00:00
25	9	Всемирная паутина. Файловые архивы	25.04.2023, 00:00
26	9	Сетевой этикет. Электронная почта. Зачетная работа по теме «К...	19.05.2023, 00:00
27	9	Подходы к измерению информации. Содержательный подход к ...	13.09.2022, 00:00
28	9	Основополагающие принципы устройства ЭВМ. Принципы Ней...	09.11.2022, 00:00
741	9	Контрольная работа "Компьютер и его программное обеспечен...	21.12.2022, 00:00

Рис. 58. Пройденные разделы в разрезе дней по количеству учеников

Первая выгрузка позволяет отследить, сколько учеников и в какой день проходили тот или иной раздел курса. Зная количество учеников, можно понять, сколько обучающихся начали раздел с задержкой.

Statis Process Mining • Быстрый просмотр

Таблица 1 | Таблица 2

#	9.0 Процент от общего количества учеников	12 Количество учеников	12 Получено двоек за раздел
1	4,00	3	8
2	9,33	7	7
3	4,00	3	6
4	13,33	10	5
5	13,33	10	4
6	17,33	13	3
7	12,00	9	2
8	14,67	11	1
9	12,00	9	0

Рис. 59. Проценты двоек за разделы курсов

Вторая выгрузка позволяет понять долю учеников, получивших определённое количество двоек за раздел, то есть понять долю обучающихся, которым понадобилось повторное прохождение того или иного раздела.

Следующий шаг – определение узких мест. Узкое место – это событие (или паттерн), к которому возникает наиболее длительная очередь, которая может быть рассчитана только на основании двух полей Timestamp.

В случае с онлайн курсами, узким местом является тот раздел курса, который имеет или самую большую длительность.

В данном процессе присутствует подмодель по счёту шагов по времени начала процесса. Для этого используется функция, возвращающая накапливающуюся сумму. Накопление считаются в пределах идентификатора процесса, который и является идентификатором ученика.

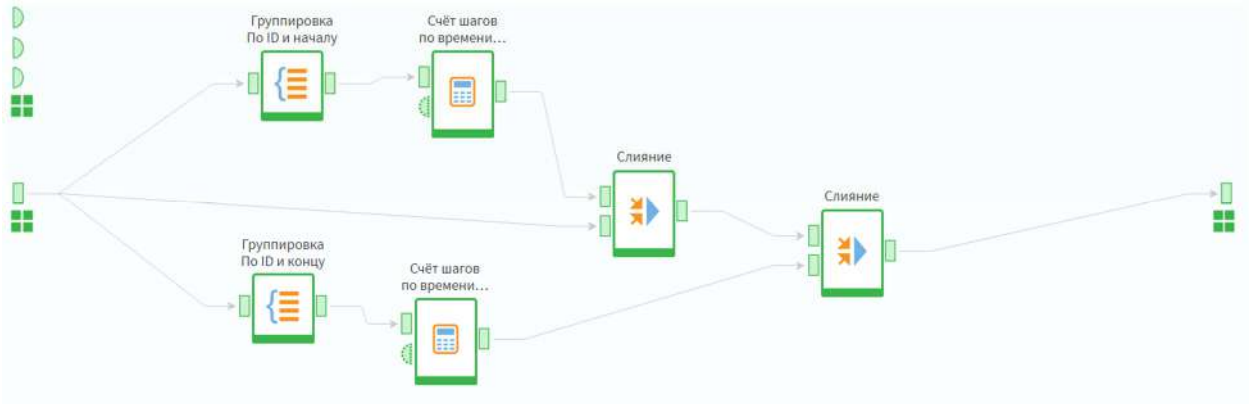


Рис. 60. Подмодель «Счёт шагов по времени начала процесса»

Также присутствует подмодель, в которой вычисляется нагрузка на процесс.

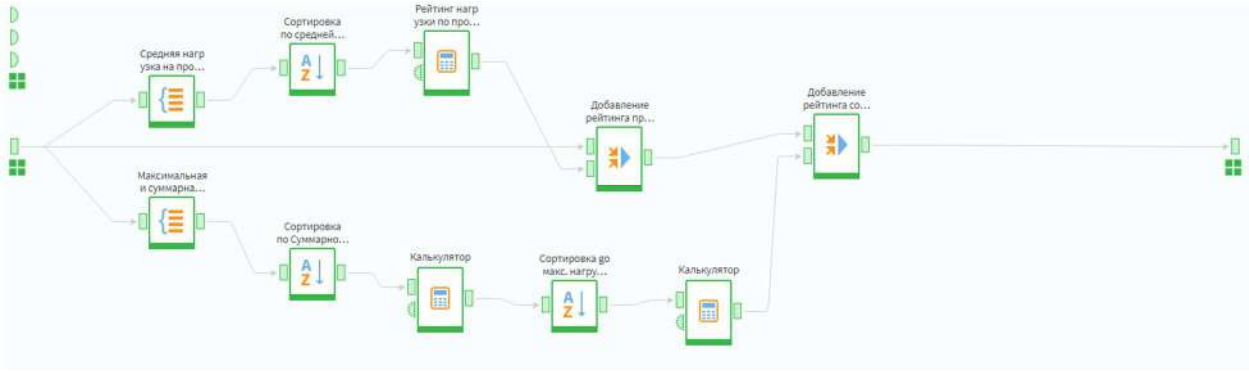


Рис. 61. Подмодель «Нагрузка на процесс»

Нагрузка на процесс – это количество экземпляров процесса, выполняющихся одновременно в какой-то момент времени. В данном случае это количество проходимых разделов курса за один день.

Поясним, что значит метрика «нагрузка на процесс». Например, в календаре на утро было 3 срочные задачи (3 курса), к которым надо приступить одновременно, и со вчерашнего дня осталась одна недоделанная задача. Итого – в параллели есть 4 задачи. К обеду 3 задачи сделаны, но «нападало» ещё две, одну из которых вы взяли в работу. Итого, ваша нагрузка в середине дня: 4–3+1 = 2, и одна задача в очереди. Если к концу дня удастся всё сделать, то нагрузка будет равна 0.

Есть ещё метрика производительности – это количество задач, которые исполняются в единицу времени. Для данного примера за рабочий день про-

изводительность = 5,5 = 3 (на начало дня) + 2 («упавших») +0,5 (остаток вчерашнего дня).

Аналогично можно рассчитать нагрузку на каждое событие – то есть количество событий, исполняемых параллельно.

Нагрузка, в общем случае, это разница между входным потоком (запуском процесса или события) и выходным потоком (завершением процесса или события).

После определения нагрузки проводится корреляционный анализ – находится корреляция между нагрузкой и длительностью.

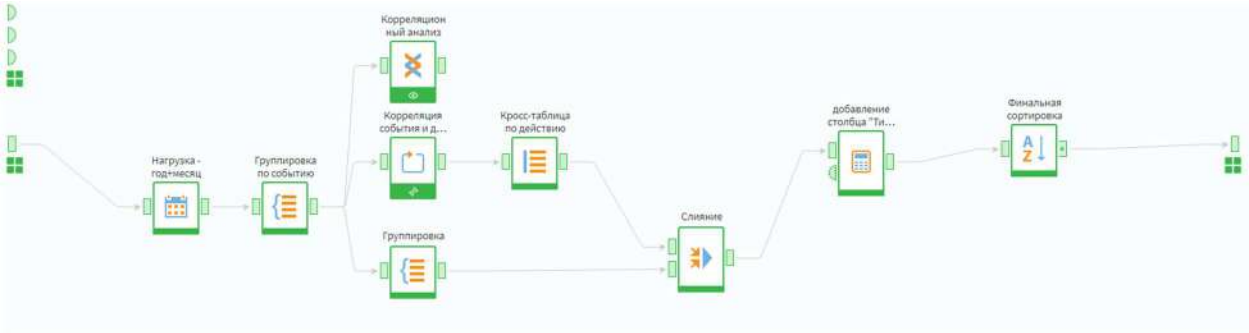


Рис. 62. Подмодель с расчётом корреляции

В результате проведённого корреляционного анализа оказалась, что чем больше задач вы берёте в параллельное выполнение, тем дольше длительность исполнения каждой задачи: каждый раздел влияет на длительность его изучения, то есть корреляция может быть заметно положительной, скажем, больше +0,7. В таком случае это событие – первый кандидат в «бутылочные горлышки» (или узкие места процесса).

Рис. 63. Подмодель «Опредение узких мест»

Важный вопрос – как рассчитать усреднённую нагрузку, скажем, за неделю, при том, что характерная длительность события – около часа или менее. Здесь есть пространство для творчества: можно, например, разбить неделю на, скажем часовые промежутки. Для каждого часа выбирать максимальную нагрузку, а затем усреднить все часовые максимумы за неделю.

Но увеличение длительности события при увеличении нагрузки наблюдается не всегда. Если мы имеем дело с системой, имеющей «запас прочности», то увеличение нагрузки может приводить к незначительным увеличениям корреляции или иметь флуктуации корреляции (скажем, от +0,5 до –0,5). В таком случае система толерантна к нагрузке и не является бутылочным горлышком.

Бывают случаи, что при увеличении нагрузки на событие мы видим выраженные отрицательные корреляции. Это явление, как правило, возникает в случае, если в процесс были добавлены ресурсы или процесс был перестроен. То есть, если в результате вашей работы по улучшению процесса корреляция нагрузки и длительности снизилась, то это можно смело приписать себе в заслугу.

Далее начинается работа с буквами процесса и эталонным путём.

Так как нам известен идеальный маршрут (эталонный путь) прохождения курса учеником, можно подать его в качестве входных данных. А так как события для всех учеников одинаковые, то и буквы событий изначально нам уже даны и будут поданы на вход.

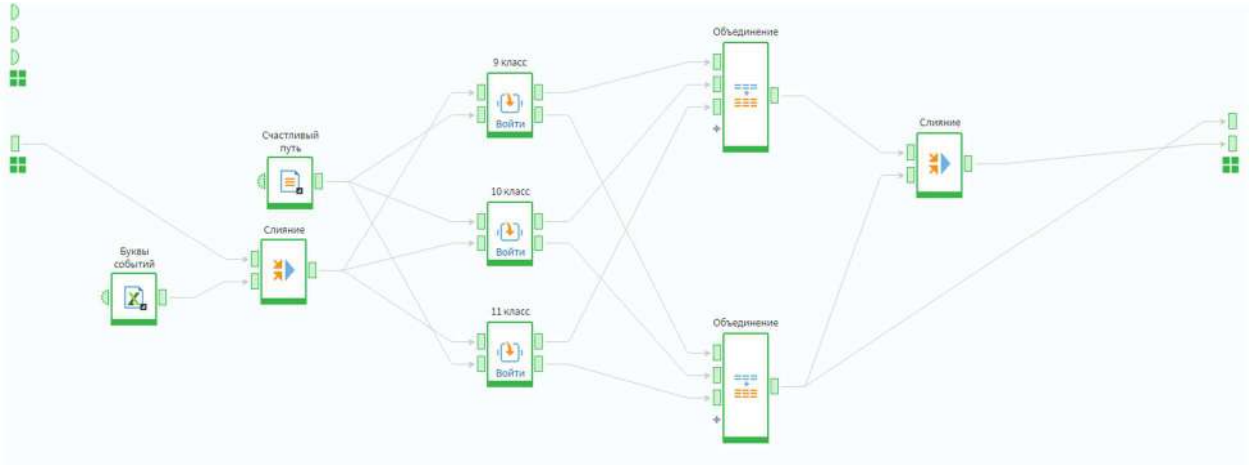


Рис. 64. Подмодель «Слово процесса и эталонный процесс»

В подмодели «Слово процесса и идеальный процесс» реализовано разделение по классам, это сделано для того, чтобы можно было доработать анализ для каждого отдельно взятого класса или даже для группы учеников.

В каждой отдельно взятой подмодели с классом реализован идентичный механизм определения «расстояния до идеального процесса», где используется расстояния Дамерау-Левенштейна (это расстояние также называют расстоянием редактирования с учетом перестановок, которое является мерой похожести двух строк). Результатом является значение минимального количества операций удалений, вставки, замены и перестановок символов, которое нужно произвести, чтобы преобразовать одну строку в другую.

В итоге получается набор данных, который послужит основой для оценки эффективности, а также для выявления трасс.

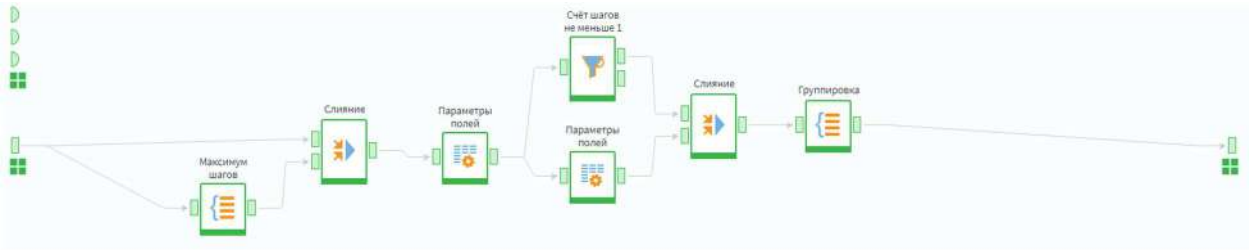


Рис. 65. Подмодель «Выявление трасс»

Подмодель «Трассы» позволяет определить все существующие трассы на основе слов процесса. В данном случае трасса представляет уникальную комбинацию событий, то есть разделов курсов, пройденных учеником. Получается набор разнообразных маршрутов выполнения курсов.

Подмодель «Кластеризация» служит для кластеризации учеников с последующей интерпретацией. Каждый кластер является группой учеников относительно средней успеваемости за курс.

Стоит отметить, что обычно для определения числа кластеров используют различные методы, например метод «локтя», но для данного сценария нет необходимости искать оптимальное количество кластеров, ведь нужно определить степень освоения программы, поэтому кластеров будет три.

Таким образом выделено три кластера учеников по успеваемости со следующей интерпретацией:

- Cluster 0 – Ученики, нормально освоившие программу
- Cluster 1 – Ученики, отлично освоившие программу
- Cluster 2 – Ученики, плохо освоившие программу

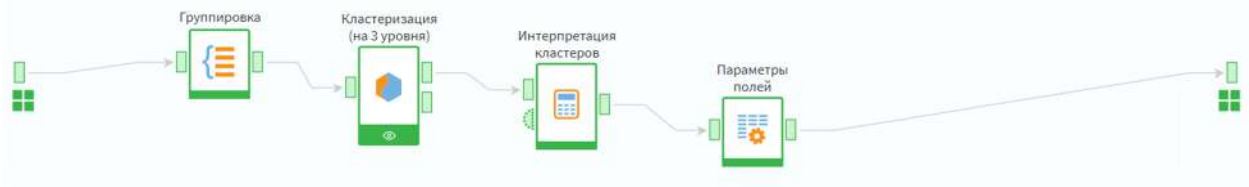


Рис. 66. Подмодель «Кластеризация»

Результат работы данной подмодели потребуется для следующей подмодели в качестве дополнительных данных для выгрузки результатов каждого ученика.

Последним шагом в создании сценария обработки данных на платформе Loginom будет реализация поставленной задачи по анализу курсов в целом, то есть необходимо создать такую подмодель, которая будет выступать в роли системы поддержки принятия решения по модернизации курсов на основе результатов прохождения их учениками. Для этого создаётся подмодель «Решение задачи по организации курсов».

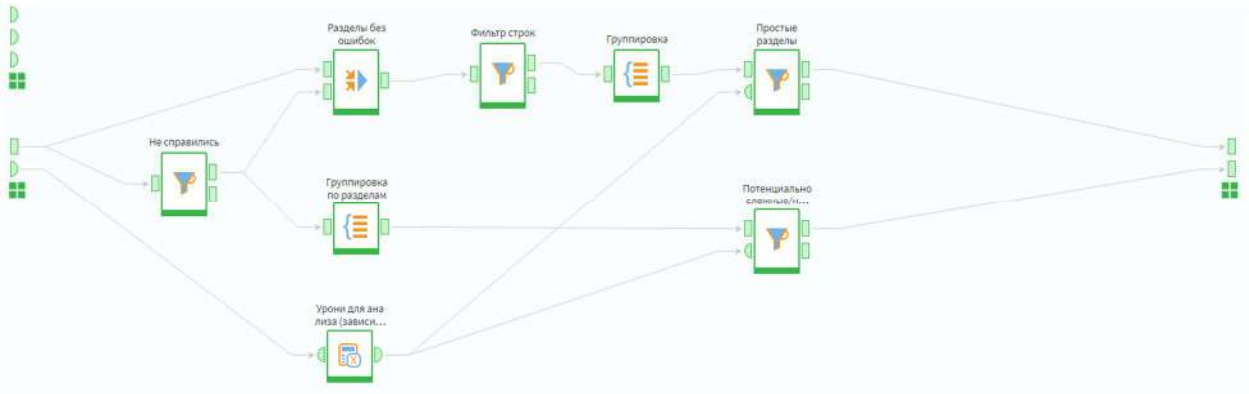


Рис. 67. Подмодель «Решение задачи по организации курсов»

В результате выполнения данной подмодели администратор получает две таблицы. В одной из них указываются те разделы курса, с которыми все ученики справились безошибочно, то есть потенциально простые разделы, которые нуждаются в доработке. Во вторую попадают разделы, с которыми

не справились хотя бы один раз не менее 30% учеников, то есть потенциально сложные или некачественно организованные курсы.

Данная информация может «подсветить» проблемные места курсов и подскажет администратору, какие разделы нужно дополнить методическими материалами, а где-то пересмотреть процедуру оценивания.

При определении ключевых показателей эффективности применения Process Mining (KPI) учитывался дистанционный формат обучения. По итогу такими показателями стали:

- время, которое пользователь провел на платформе за изучением курса;
- средний балл;
- количество неудовлетворительных оценок, полученных учеником;
- количество пройденных учеником тестовых заданий и заданий с открытым ответом и результаты их выполнения;
- уровень активности прохождения итоговых тестов;
- уровень активности работы с теоретическими материалами;
- количество просмотренных пользователем образовательных медиа материалов;
- уровень активности выполнения заданий, проверяемых вручную;
- уровень активности на форумах;
- последовательность выполнения заданий (по траекториям).

Некоторые перечисленные показатели были определены в рассматриваемом сценарии (рис.68):

- средняя оценка учеников на курсе составила 4 балла из 5.
- неудовлетворительных оценок, полученных учениками – 248 (9%).
- отличных оценок – 672 (24%).

№	Метка	Доля	Кол-во	%
1	2		248	9
2	3		976	35
3	4		903	32
4	5		672	24

Рис. 68. Статистика оценок учеников на курсе

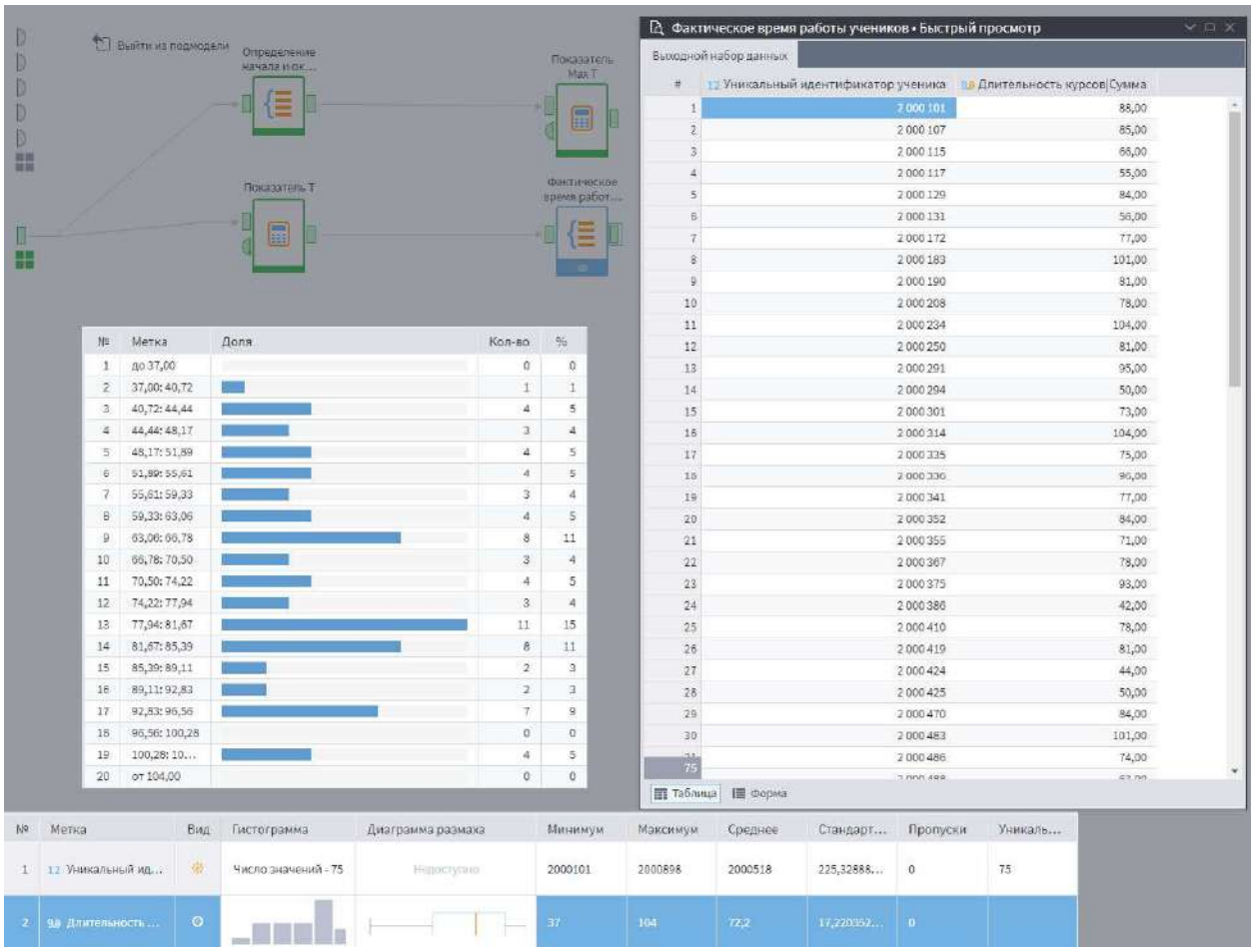


Рис. 69. Статистика времени учеников на курсе

Время, которое ученики провели на платформе за изучением курсов, составило – 265 день, однако были те, кто закончили курс раньше. Открытие курсов произошло 06.09.2022, а последний день сдачи заданий – 29.05.2023.

Фактическое время работы учеников было получено в Loginom с помощью группировки и составило в среднем 72 дня.

Количество пройденных учеником тестовых заданий и заданий с открытым ответом и результаты их выполнения равен количеству выполненных заданий на курсе, поскольку курс состоит из лекционной части и тестирования по материалу. Количество таких заданий составило 34.

Что касается прохождений контрольных заданий, то можно отметить, что ученики в основном писали эти работы на оценку 4 из 5 и преимущественно за 1–2 дня. (см. рис. 70.)

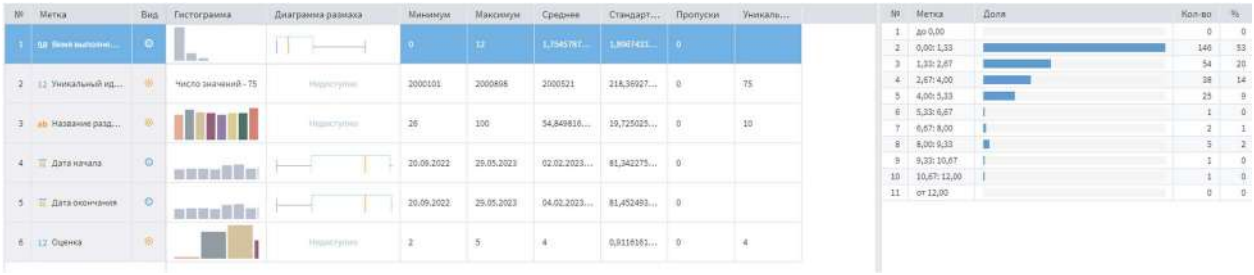


Рис. 70. Статистика по контрольным работам

Поскольку курс был выстроен по принципу – «Прохождение онлайн материала после очного урока», то все задания выполнялись последовательно.

Идеальный маршрут прохождения курса – пройти на оценку выше 3 каждое из 34 заданий. Курсы предполагали пересдачу материалов, при получении оценки 2. Учеников, не допустивших не одной ошибки – 9. То есть, идеальной траектории достигли лишь 12% учеников.

Подробная статистика представлена на рисунке 71.

№	Метка	Доля	Кол-во	%
1	1-2 неудовлетворительные оценки		20	27
2	3-5 неудовлетворительных оценок		33	44
3	Неудовлетворительных оценок больше 5		13	17
4	Неудовлетворительных оценок нет		9	12

Рис. 71. Статистика по контрольным работам

Преимущества использования Educational Process Mining:

– Улучшение платформ онлайн-обучения. Анализ процессов позволяет обнаруживать процессы навигации по пользовательскому интерфейсу на обучающих платформах.

– Поиск эффективных процессов обучения. Интеллектуальный анализ процессов можно применять для понимания процессов обучения учащихся. Например, разделяя учащихся по их успеваемости, можно выявить процессы обучения, которые приносят более высокие результаты, и рекомендовать их другим учащимся.

– Улучшение навыков решения задач. Основываясь на результатах анализа взаимодействия ученика и компьютера, можно побудить следовать более выгодным стратегиям.

2.4. Вопросы и задания для самоконтроля

1. Что такое «процессный подход»?
2. Что такое «майнинг процессов» (Process Mining)?
3. В каких организациях используется процессный майнинг. Приведите примеры.
4. Как представляются данные о процессах в различных информационных системах.
5. Требования к журналам событий (Event Logs).
6. Опишите возможные способы описания бизнес-процессов по данным из журналов Event Logs.
7. Опишите особенности форматов представления данных о процессах.
8. Какие атрибуты являются обязательными для проведения анализа Process Mining?
9. Что такое состояние, событие, переход, действие. Как они записываются в корпоративных информационных системах.
10. Как выбранный для выгрузки данных временной интервал должен соотноситься со средней длительностью одного экземпляра процесса.
11. Какие формы применения Process Mining существуют?
12. Из скольких этапов состоит стандартное исследование Process Mining?
13. Что такое экземпляр и путь процесса?
14. Что такое поиск узких мест (bottlenecks) в бизнес-процессах. Как он реализуется?
15. Как выявляют отклонения в бизнес-процессах в Loginom?
16. Как реализуется поиск счастливых путей выполнения бизнес-процессов.
17. Опишите метрики процесса в Process Mining.
18. Как можно реализовать оптимизацию бизнес-процессов, представленных в результате Process Mining?
19. Как используются искусственный интеллект и машинное обучение в Process Mining?

20. Разработайте концепцию применения Process Mining в интернет-торговле.

21. Разработайте концепцию применения Process Mining в разработке программного обеспечения.

22. Разработайте концепцию использования Process Mining в логистике.

23. Разработайте концепцию использования Process Mining для транспортной компании.

24. Разработайте концепцию использования Process Mining для аэропорта.

25. Разработайте концепцию использования Process Mining в киберспорте.

Раздел 3. Глоссарий

1. Датасет [Dataset] (Синонимы «Лог», «Цифровые следы»)

Датасет **-** набор данных, включающий в себя множество экземпляров процесса в формате PM Ready. Датасет – это собранные в единый источник данных логи. Датасет может быть представлен как файл или витрина данных и пр. Пример в приложении Приложение 2. Пример датасета.

2. Валидация датасета [Dataset validation]

Валидация – процесс проверки датасета на соответствие требованиям PM Ready. Формат описан в Приложение 1. PM Ready формат датасета.

Валидация выполняется сразу после загрузки и сборки данных из источников и является составной частью ETL процесса.

Основными требованиями является отсутствие пропусков данных в ключевых полях (ID процесса, Event, DataTime_Start).

Результатом валидации является датасет, подготовленный к дальнейшему исследованию.

3. Формат Process Mining Ready [PM Ready]

Format PM Ready – формат датасета, пригодного для загрузки на платформы Process Mining.

Формат подразумевает наличие обязательных полей и неограниченного количества дополнительных. Состав полей процессного датасета является открытым: реальный процессный датасет может иметь множество полей, но в нем должно быть 3 обязательных поля.

Обязательные поля:

- **код экземпляра процесса [ID_Process] (синонимы «Case_ID», «ID»)** – ID экземпляра процесса, позволяющая отнести последовательность некоторых событий к выполнению этого экземпляра процесса. В рамках Case_ID выполняются все события (активности), относящие к данному экземпляру процесса. Экземпляром процесса может быть обработка заявки, единичный эпизод обслуживания клиента, подготовка договора и пр.

- **событие процесса [Event] (Синонимы «Activity», «Status»)** – человеко-понимаемое содержательное наименование действия, либо статуса (результата действия). Событие – неделимая в рамках одного лога сущность. Допустимо использование логов с указанием статусов [Status] вместо событий. Статусы обозначают завершенное действие и определяются моментом завершения. Например, для события – «Приёмка Ж/Д терминал» статусом была бы формулировка – «Товар в Ж/Д терминале принят».
- **Момент времени начала совершения события [Timestamp_Start] (Синонимы «DateTime», «StartTime»)]** – Момент начала совершения события, или момент смены статуса процесса.

В таблице 1 приведен примерный перечень полей процессного датасета.

В поле «Обязательность» указано – обязательное ли это поле, желательное или дополнительное.

- Без обязательных полей Process Mining анализ процесса невозможен практически на всех PM платформах
- Желательное поле открывает большие дополнительные аналитические возможности для анализа и позволяет реализовать дополнительный функционал PM платформ
- Дополнительные поля могут нести функции аналитик, либо под дополнительные поля можно разработать кастомизированный анализ процесса.

Таблица 1. Перечень обязательных, и некоторых дополнительных полей процессного датасета

№	Наименование поля (с синонимами)	Русское наименование	Обязательность	Иерархия	Описание – логическое предназначение	Комментарий
1.	**Process_ID, Case_ID**	Идентификатор экземпляра процесса	Обязательное	нет	ID экземпляра процесса	**Идентификатор экземпляра процесса**. Обязательное поле.

2.	**Event, Activity, Status**	Событие, активность, статус	Обязательное	нет	Текстовое наименование статуса (шага) процесса	**Наименование события**. Обязательное поле. Поле несёт смысловую нагрузку о ходе бизнес-процесса
3.	**DateTime, TimeStamp_Start**	Момент начала (события или экз. процесса)	Обязательное	нет	Дата / время начала выполнения события процесса	**Временная метка начала события**. Обязательное поле.
4.	**Datetime_End, Timestamp_End**	Момент завершения (события или экз. процесса)	Желательное	нет	Дата / время окончания времени выполнения события процесса	**Временная метка окончания события**. Желательное, но не обязательное поле. Формат аналогичен полю «DateTime». Наличие поля позволяет существенно увеличить глубину анализа: анализировать чистое время, очереди, занятость и пр.
5.	**Measure_Name, Measure_Code,**	Аналитика, измерение	Желательное	да	Наименование измерения, аналитический разрез	**Наименования измерений.** Желательное, не обязательное поле. Полей измерений может быть множество, их наименования специально не регламентируются. Наличие измерений позволяет проводить более глубокий анализ процесса. ***Пример***: тип товара, канал сбыта Измерения могут быть иерархичными. ***Пример иерархичности:*** область, город, терр.представительство

6.	**User_ID, User**	Исполнитель	Желательное	нет	ID сотрудника, выполняющего шаг процесса и чьи трудозатраты ложаться на процесс	**Идентификатор пользователя** Желательное, но не обязательное поле. USER – менеджер, сотружник банка, пользователь IT систем Банка. (Не клиент)
7.	**User_Role**	Роль исполнителя	Желательное	нет	Роль пользователя	**Идентификатор роли пользователя в системе** Желательное, но не обязательное поле.
8.	**Client_ID**	Идентификатор клиента	Дополнительное	нет	ID клиента	**Идентификатор клиента** Не обязательное поле
9.	**Unit_Name, Office**	Юнит, подразделение	Желательное	да	Объект управления: внутреннее подразделение, филиал, офис.	**Обособленное территориальное представительство** Желательное поле Выделяет Объект (Unit), в котором происходят бизнес – процессы, имеющие минимальные связи с другими объектами. **Пример**: территориальные продающие подразделения, точки продаж, магазины. Для любых объектов можно выделить бизнес-процессы, которые выполняются внутри этого объекта. Однако есть процессы, например процессы централизованного снабжения, которые выполняются в объекте иерархическим уровнем выше.
10.	**Minute_Cost**	Стоимость минуты	Желательное	нет	Стоимость минуты чистого процессного времени	**Стоимость единицы времени user** при выполнении статуса. Желательное, но

						не обязательное поле.
11.	**Text, Comment**	Комментарий	Желательное	нет	Комментарии, описания, фрагменты чатов, прочая текстовая информация	**Текстовые комментарии, описания, переписка.** Желательное, но не обязательное поле. Полей может быть несколько. Анализ текстовой информации позволяет глубже идентифицировать проблемные зоны процесса.
12.	**IT_System**	ИТ система	Дополнительное	нет	Наименование информационной системы процесса.	**Наименование или ID ИС источника данных.** Не обязательное поле. Желательно, если данные для анализа поступают из нескольких ИС. Используется для анализа интеграционных потоков между системами, нагрузки на системы и пр.
13.	**Region_latitude**	Географическая широта	Желательное	да	Широта и долгота для отображения на картах.	**Географические координаты.** Не обязательное поле. Используется для отображения на картах, если требуется указать место положение объекта, отличающегося от области, города. (напр. Положение ВСП). **Для информации**: справочник координат городов РФ Коды регионов в соответствии с Классификатором адресов Российской Федерации (КЛАДР).
14.	**Region_longitude**	Географическая долгота				
15.	**City**	Город	Желательное	нет	Город, населенный пункт	**Город**. Не обязательное поле. Используется для

						обозначения места нахождения объекта управления (Unit). Может быть использовано для отрисовки географических карт.

4. Журнал событий [Event Log]

Синонимы: лог событий, процессный лог, журнал цифровых следов

Журнал событий – это перечень упорядоченных во времени событий (действий, статусов), выполняемых в рамках реализации некоего бизнес-процесса. По журналам событий можно построить граф процесса и провести аналитику по процессу, сравнить процесс AS IS с проектом этого процесса.

Размер журнала событий (лога) определяется количеством содержащихся в нем строк. Также размер лога определяется его объёмом (Мбайт) или количеством экземпляров процесса.

Примеры источников журналов событий:

- информационные системы;
- сетевое оборудование;
- трекинг посетителей сайтов;
- мониторинг действий сотрудников на своих АРМ (при наличии специальных агентов логирования);
- устройства интернета вещей.

Данные из журналов событий являются источником информации для анализа по технологии Process Mining. События могут храниться в таблицах баз данных, журналах регистрации сообщений, архивах электронных писем, журналах транзакций. Типичными форматами для хранения логов событий являются формат CSV или Excel (для небольших логов), а также XES (eXtensible Event Stream) и MXML (Mining eXtensible Markup Language).

Для Process Mining очень важное значение имеет качество логов событий, так как качество результатов Process Mining напрямую зависит от входных данных.

Существует несколько критериев оценки качества данных о событиях:

- события должны быть достоверными, то есть нужно быть уверенными, что записанные события действительно случились и что атрибуты событий являются верными;
- лог событий должен быть завершенным, то есть предоставлять полную картину действий;
- у любых записываемых событий должна быть четко определенная семантика (разнотипные события должны различаться).

Пример фрагмента журнала событий представлен в таблице 2. Подобные журналы событий можно рассматривать как совокупности экземпляров процессов. Экземпляры процессов – последовательность событий в рамках выполнения одного экземпляра процесса.

Обязательные поля журнала событий:

1. Process ID – определяет экземпляр процесса, т.е. перечень событий, объединённых одной меткой ID экземпляра процесса.

2. Event – упорядоченный перечень событий, выполняемых в рамках экземпляров процесса. В рамках лога «Event» является неделимой сущностью. События также часто называют шагом процесса. Есть логи, в которых события заменены статусом, то есть результатом выполнения события. (например: товар принят, документы оформлены и пр.)

3. Timestamp – момент начала совершения события.

Таблица 2. Фрагмент журнала событий с базовыми полями

Process ID	**Event**	**Timestamp**
1	Приёмка Ж/Д терминал	03.06.2024 13:59:51
1	Оформление приходных документов	03.06.2024 14:05:51
1	Оформление документов контрагентов	03.06.2024 14:26:42
1	Ввод данных контрагента	03.06.2024 14:35:09
1	Оформление приходных документов	04.06.2024 9:47:40
1	Заменить	04.06.2024 14:05:15
1	Оформление приходных документов	04.06.2024 14:05:28
1	Заменить	05.06.2024 8:28:17
1	Определение локации	05.06.2024 8:28:35
1	Вызов погрузчика	05.06.2024 8:29:41
1	Размещение груза	05.06.2024 13:40:53
1	Груз размещён	05.06.2024 13:41:01

2	Приёмка Ж/Д терминал	04.06.2024 2:28:58
2	Оформление приходных документов	04.06.2024 3:54:28
2	Груз размещён	04.06.2024 5:09:07
3	Приёмка	04.06.2024 3:02:31
3	Ввод данных контрагента	04.06.2024 4:45:41

Лог также должен соответствовать критериям качества. Некоторые из них:

- Лог должен быть отсортирован. Как правило, лог сортируется сперва по Process_ID, затем по Timestamp.
- Не должно быть пропусков указанных ключевых полей. Иначе требуется или перевыгрузка, или удаление всего Process_ID с пропусками.
- Если один Process_ID содержит только одно событие – это «желтый» сигнал: необходимо перепроверить качество логирования.

Каждое событие имеет один или несколько атрибутов. Расширенный лог может содержать формально не ограниченный перечень дополнительных полей – атрибутов.

Из значимых полей-атрибутов в логе могут быть:

- Timestamp Fin – момент завершения события
- Пользователь (User)
- Аналитики (продукт, канал коммуникации, город и пр.)

Таблица 2 отражает стандартный подход к логированию, но в последнее время получает распространение новый стандарт состава полей. Добавляются поля:

Event_request – или событие запрашивающее. Например, в любых IT системах есть некие (запрашивающие) функции, которые могут вызвать некоторые другие функции или сервисы. А те, в свою очередь, могут вызвать последующие функции.

Для фиксации иерархии вызовов вводится поле «**hierarchy_of_nesting**» – иерархия вложенности. Она показывает уровень иерархии вложенности последовательности взаимных вызовов событий или функций.

В таблице 3 приведен абстрактный пример такого лога с дополнительным полем запрашивающего события.

Таблица 3 Перечень обязательных, и некоторых дополнительных полей процессного датасета

hierarchy_of_nesting	**Event_request**	**Event**	**TimeStamp_Start**	**TimeStamp_End**
1	create	update_info	1	2
1	create	update_status	1	3
1	create	Import_Data	1	2,5
2	Import_Data	credit-potential	3	5
2	Import_Data	Credit_start	3	4
3	Credit_start	NBKI_start	4	6
3	Credit_start	NH_RA	4	7
3	Credit_start	NBKI_end	4	12

Новые системы процесс майнинговых исследований имеет смысл строить с учётом такого лога, так как всё больше данных будет появляться именно в такой логике. В таком логе мы отчётливо видим параллелизм вызовов и выполнения различных событий. По полям начала и завершения события можно легко вычислить чистую длительность событий.

Из такого лога можно вычленить дополнительную информацию по очереди и пути.

Очередь – это состояние, когда предшествующее событие завершилось, а последующее ещё не началось.

В примере (табл. 3) «в очереди» стояло событие Import_Data. Предшествующее ему событие «create» завершилось на 2,5 временной единице, а «Import_Data» стартовал с 3-й временной единицы, то есть 0,5 времени потеряно на очередь.

Для модели лога с параллельно текущими событиями, представленной в таблице 3, расчёт пути становится не столь тривиальной задачей как в логе с одним Event и одним полем дат. Путь в контексте параллельности должен показывать дерево последовательности вызовов событий (или функций) в иерархии вложенности. Причём, в рамках одной иерархии вложенности нет

разницы во временной последовательности старта вызываемых событий (event).

Пример с магазином для наглядного описания параллельных иерархических событий в процессе.

Чтобы показать на примере, рассмотрим процесс покупок в магазине мама с дочерью. Скажем, условно, что «**Event_request»** в таком случае будет – «Покупатель зашёл в торговый зал». Они параллельно могут производить «**event**» – наполнять продуктами свою корзину и что они положат в неё в какой последовательности: например, кефир или молоко, или «растишку» – значения не имеет. Данные в полях **TimeStamp** в логе будут, но для сортировки последовательности и построения пути они не используются. Если последующая семья (новый ID экземпляра процесса) положит в корзину сперва «растишку», потом молоко и потом кефир – то у этих двух семей будет одинаковый путь. А если приобретение, скажем, кефира напомнит им о том, что надо бы купить муку для блинов, то в логах это отразится уже как новый уровень иерархии, и «**Event_request**»-ом станет покупка кефира, а «**event**» будет – покупка муки. И так далее по уровням вложенности.

Но если покупка кефира не является причиной последующей покупки муки, то данные события не связаны, и будут находиться на одном иерархическом уровне.

В случае разбора логов IT систем причинно-следственная связь в таких логах чётко отслеживается. Явно понятно какая функция какую последующую вызывает – то есть чётко отслеживается причинность последовательности и иерархия событий.

Трансформация логов. Между логами в стандарте таблицы 2 и таблицы 3 есть возможность взаимного пересчёта. Предшествующее событие таблицы 2 является «**Event_request**» для лога таблицы 3. Аналогично для полей времени начала и конца – момент начала последующего события – это момент завершения предыдущего.

Конечно, такой пересчёт является некоей «натяжкой» и может использоваться в Loginom для приведения логов к единому стандарту состава полей.

5. Событие [Event] (Синонимы «Шаг процесса», «Этап», «Действие»)

Событие (Event) – это базовая единица информации, используемая в PM.

Каждое событие описывает действие (Event и Activity синонимы с точки зрения Process mining) четко определенным этапом некоторого процесса) и является составной частью экземпляра процесса.

Примерами событий являются – запись персональной информации о клиенте в базу данных, проверка валидности QR кода, осуществление снимка с помощью рентгеновского аппарата и пр.

Таблица 4. Фрагмент лога событий

ID экземпляра процесса	ID события	Время	Действие	Ресурс	Стоимость
1	35653323	30-12-2024:11.02	Регистрировать запрос	Петр	50
1	35653324	31-12-2023:10.06	Проверить (тщательно)	Светлана	400
1	35653325	05-01-2024:15.12	Проверить билет	Михаил	100
1	35653326	06-01-2024:11.18	Принять решение	Светлана	200
1	35653327	07-01-2024:14.24	Отклонить запрос	Петр	200
2	35654483	30-12-2023:11.32	Регистрировать запрос	Михаил	50
2	35654485	30-12-2023:12.12	Проверить билет	Михаил	100
2	35654487	30-12-2023:14.16	Проверить (базовые проверки)	Петр	400
2	35654488	05-01-2024:11.22	Принять решение	Светлана	200
2	35654489	08-01-2024:12.05	Выплатить компенсацию	Елена	200

В таблице 4 представлены события, которые относятся к процессу обработки запросов для получения компенсационных выплат: «Регистрировать запрос», «Проверить билет», «Отклонить запрос», «Принять решение» и другие.

События могут иметь уникальные идентификаторы – ID события. Использование идентификаторов вместо наименований событий обосновано, когда в написании наименований возникает вариативность, или когда надо закодировать, например, для внешних аналитиков смысловое содержание процесса.

Следует различать лог событийный и лог статусный. Например, в таблицах 5 и 6 представлены два типа логов.

Таблица 5. Событийный лог

Событие	**Момент начала**	**Момент завершения**
Регистрация запроса	30-12-2023 11:02	30-12-2023 11:04
Проверка билета	05-01-2024 15:12	05-01-2024 15:20
Принятие решения	06-01-2024 11:18	06-01-2024 17:55
Отклонение запрос	07-01-2024 14:24	07-01-2024 14:30

Таблица 6. Статусный лог

Статус	**Момент совершения**
Запрос поступил	30-12-2023 11:02
Запрос зарегистрирован	30-12-2023 11:04
Билет проверен	05-01-2024 15:12
Решение принято	06-01-2024 11:18
Запрос отклонен	07-01-2024 14:24

Событие обозначает продолжительное во времени действие, то есть событие имеет длительность. Статус – это смена состояния процесса. Статус имеет только момент его совершения.

В классическом событийном логе с двумя столбцами дат можно рассчитывать чистое время операции как разницу между моментами конца и начала операции. Также событийный лог даёт возможность рассчитать длительность простоя или очереди.

Статусный лог позволяет рассчитать только фактическую длительность между моментами смены статусов (моментами начала обработки событий) и фактическую длительность экземпляра процесса.

Событийный лог гораздо предпочтительнее для анализа, в некотором смысле он даёт 3D картину процесса, вместо плоской последовательности статусного лога.

6. Атрибуты события [Event Attributes]

Атрибуты события являются его характеристиками и описывают его более подробно. Каждое событие может имеет один или несколько атрибутов, при этом значения атрибутов могут варьироваться.

У событий могут быть атрибуты, которые характеризуют некоторое действие в рамках определенного бизнес-процесса. Наличие атрибутов является необязательным, но на практике каждое событие имеет один или несколько атрибутов. Технически атрибуты события являются измерениями (Dimension), на основе которых строится BI-аналитика по процессу.

Примерами атрибутов являются ресурсы [resource] (сотрудник User, клиент, исполнитель работы и пр.), состояние [Status] – результат события, аналитики [Dimension] – продукты, каналы, коммуникации, местонахождение и пр. Также примерами атрибутов могут быть стоимость единицы времени, подразделение исполнитель и пр.

Эта информация может оказаться полезной при анализе процессов (например, можно определить время ожидания между двумя действиями).

7. Валидные начальные и конечные события [Valid start and end events]

Валидные начальные и конечные события – это два множества событий: множество событий, на которые штатно может начинаться процесс, и множество событий, на которые он может штатно завершаться.

Процессы, начинающиеся и завершающиеся на валидные события, являются полными.

Перечень вводится для того, чтобы автоматически различать фрагменты незавершенных процессов (без валидных начальных/конечных событий), и полностью завершившиеся процессы.

8. Буква события [Event letter]

Буква события – каждому событию сопоставляется одна буква (символ), которая и является «буквой события».

Буквенное обозначение событий позволяет делать процессный лог более компактным и позволяет сильно упростить технологию анализа процессов.

Кодировка процесса буквами интуитивно понятна и технологически транспарентна.

Событие и буква события образуют соответствие 1:1.

«Буквой» события должен являться один символ.

Если использовать кодировку «Юникод», то в ней достаточно символов для кодировки практически любого широкого набора уникальных событий процесса.

Таблица 7. Фрагмент журнала событий, где показано соответствие между событиями и буквами событий

Process ID	**Event**	**Буква события**	**Timestamp**
1	Приёмка Ж/Д терминал	A	03.06.2024 13:59:51
1	Оформление приходных документов	B	03.06.2024 14:05:51
1	Оформление документов контрагентов	C	03.06.2024 14:26:42
1	Ввод данных контрагента	D	03.06.2024 14:35:09
1	Оформление приходных документов	E	04.06.2024 9:47:40
1	Заменить	F	04.06.2024 14:05:15
1	Оформление приходных документов	E	04.06.2024 14:05:28
1	Заменить	F	05.06.2024 8:28:17
1	Определение локации	G	05.06.2024 8:28:35
1	Вызов погрузчика	H	05.06.2024 8:29:41
1	Размещение груза	I	05.06.2024 13:40:53
1	Груз размещён	J	05.06.2024 13:41:01
2	Приёмка Ж/Д терминал	A	04.06.2024 2:28:58
2	Оформление приходных документов	B	04.06.2024 3:54:28
2	Груз размещён	J	04.06.2024 5:09:07

9. Экземпляр процесса [Process Instance] (Синоним кейс [Case])

Экземпляр процесса (Process Instance) – лог реализации одного конкретного случая процесса для получения единичного результата процесса в Process Mining.

Экземпляр процесса представляет упорядоченный перечень операций для описания одного завершенного действия в рамках бизнес-процесса. Например, размещение товара можно описать последовательностью: Разгрузка товара, оформление накладных, вызов погрузчика, размещение товара на складе, занесение записи о месте хранения в реестр.

Экземпляр процесса маркируется уникальным Process ID. Например, ID может быть номером одной конкретной заявки на кредит, выплаты одного страхового возмещения, приёма одного клиентского звонка, лога одного контакта клиента с провайдером услуг и пр. В таблице 8 указаны два экземпляра процесса с Process ID – 1 и 2.

Таблица 8. Фрагмент журнала событий с базовыми полями

Process ID	Event	Timestamp
1	Приёмка Ж/Д терминал	03.06.2024 13:59
1	Оформление приходных документов	03.06.2024 14:05
1	Оформление документов контрагентов	03.06.2024 14:26
1	Ввод данных контрагента	03.06.2024 14:35
1	Оформление приходных документов	04.06.2024 9:47
1	Заменить	04.06.2024 14:05
1	Оформление приходных документов	04.06.2024 14:05
1	Определение локации	05.06.2024 8:28
1	Вызов погрузчика	05.06.2024 8:29
1	Размещение груза	05.06.2024 13:40
1	Груз размещён	05.06.2024 13:41

2	Приёмка Ж/Д терминал	04.06.2024 2:28
2	Оформление приходных документов	04.06.2024 3:54
2	Груз размещён	04.06.2024 5:09
3	Приёмка	04.06.2024 3:02
3	Ввод данных контрагента	04.06.2024 4:45

Множество логов экземпляров процесса образуют журнал событий. С содержательной точки зрения, лог одного экземпляра процесса отображает конкретную реализацию процесса, его путь.

Экземпляры процесса могут быть завершенными или незавершенными.

Завершенным является экземпляр процесса, при котором получен результат процесса, или, иными словами, когда последним событием процесса является событие из перечня валидных конечных событий. Например: «Груз размещен».

Незавершенным является экземпляр процесса, ещё не приведший к результату. Иначе: процесс завершается на промежуточном событии. Например: «Оформление приходных документов».

Ситуация неполноты экземпляров процессов, например, возникает при выгрузке лога, когда часть недавно начавшихся процессов ещё не успела завершиться. Для отрисовки графа процесса такие экземпляры желательно удалить из процессного лога, так как они вносят искажения в пути процесса. Но решение об отрисовке незавершенных экземпляров процессов на графе должно быть за аналитиком. Для расчёта процессных метрик должен использоваться весь процессный лог без изъятий.

Чтобы получать выгоды от Process Mining, организациям следует стремиться к максимально высокому качеству журналов событий.

Таблица 9. Пример процессного лога, соответствующего требованиям PM Ready

Case_ ID	**Event**	**TimeStamp Start**	**TimeStamp End**	**User_ Role**	**USER**	**City**	**PRODUCT_NAME**	**Channel**
ID5	Приёмка	26.07.2024 11:02:31	26.07.2023 11:41:20	Комплектовщик	1181	Барнаул	Ответственное хранение	Авто

ID5	Данные контрагента введены	26.07.2024 12:45:41	26.07.2023 12:47:11	Оператор приёмщик	534	Барнаул	Ответственное хранение	Авто
ID5	Определение локации работником склада	26.07.2024 12:51:46	26.07.2023 13:01:05	Оператор склада	2412	Барнаул	Ответственное хранение	Авто
ID5	Согласование локации начальником смены	26.07.2024 13:02:01	26.07.2023 13:06:58	Начальник смены	1555	Барнаул	Ответственное хранение	Авто
ID5	Корректировка входных документов	26.07.2024 13:23:45	26.07.2023 13:28:26	Оператор приёмщик	560	Барнаул	Ответственное хранение	Авто
ID5	Определение локации работником склада	26.07.2024 13:40:54	26.07.2023 13:41:12	Оператор склада	2412	Барнаул	Ответственное хранение	Авто
ID5	Согласование локации начальником смены	26.07.2024 13:41:37	26.07.2023 13:44:12	Начальник смены	1555	Барнаул	Ответственное хранение	Авто
ID5	Определение локации	28.07.2024 18:44:29	28.07.2023 18:46:04	Диспетчер склада	15	Барнаул	Ответственное хранение	Авто
ID5	Груз размещён	12.08.2024 19:48:30	12.08.2023 19:48:35	Контролёр	561	Барнаул	Ответственное хранение	Авто
ID5	Определение локации	13.08.2024 19:07:28	13.08.2023 19:13:21	Диспетчер склада	15	Барнаул	Ответственное хранение	Авто
ID5	Груз размещён	13.08.2024 21:04:49	13.08.2023 21:06:18	Контролёр	561	Барнаул	Ответственное хранение	Авто
ID5	Определение локации	16.08.2024 14:35:54	16.08.2023 14:49:02	Диспетчер склада	15	Барнаул	Ответственное хранение	Авто
ID5	Определение локации работником склада	16.08.2024 15:37:23	16.08.2023 15:37:40	Оператор склада	2412	Барнаул	Ответственное хранение	Авто
ID5	Согласование локации начальником смены	16.08.2024 18:04:01	16.08.2023 18:07:02	Начальник смены	1555	Барнаул	Ответственное хранение	Авто
ID5	Замена сопроводительных документов	17.08.2024 13:38:06	17.08.2023 13:39:23	Оператор приёмщик	542	Барнаул	Ответственное хранение	Авто
ID5	Определение локации работником склада	17.08.2024 22:57:16	17.08.2023 22:57:43	Оператор склада	2412	Барнаул	Ответственное хранение	Авто
ID5	Согласование локации начальником смены	18.08.2024 11:30:17	18.08.2023 12:15:51	Начальник смены	1555	Барнаул	Ответственное хранение	Авто
ID5	Возврат документов контрагенту	18.08.2024 16:18:35	18.08.2023 16:22:22	Начальник смены	1389	Барнаул	Ответственное хранение	Авто
ID5	Замена сопроводительных документов	18.08.2024 16:47:16	18.08.2023 16:53:36	Начальник смены	1541	Барнаул	Ответственное хранение	Авто
ID5	Определение локации	19.08.2024 14:04:21	19.08.2023 14:04:46	Диспетчер склада	15	Барнаул	Ответственное хранение	Авто
ID5	Вызов погрузчика	19.08.2024 14:05:49	19.08.2023 14:05:49	Диспетчер склада	1499	Барнаул	Ответственное хранение	Авто
ID5	Размещение груза	19.08.2024 18:22:27	19.08.2023 18:22:28	Оператор склада	2416	Барнаул	Ответственное хранение	Авто
ID5	Груз размещён	19.08.2024 18:22:36	19.08.2023 18:22:32	Контролёр	561	Барнаул	Ответственное хранение	Авто
ID26	Приёмка Ж/Д терминал	26.07.2024 12:29:46	26.07.2023 12:29:46	Комплектовщик	397	Новосибирск	Транзитное хранение	Ж/Д
ID26	Оформление приходных документов	26.07.2024 16:58:59	26.07.2023 17:00:01	Оператор приёмщик	1183	Новосибирск	Транзитное хранение	Ж/Д
ID26	Определение локации работником склада	26.07.2024 17:01:43	26.07.2023 17:01:52	Оператор склада	691	Новосибирск	Транзитное хранение	Ж/Д
ID26	Согласование локации начальником смены	26.07.2024 22:36:05	26.07.2023 22:41:30	Начальник смены	1721	Новосибирск	Транзитное хранение	Ж/Д
ID26	Корректировка вх. документов	27.07.2024 10:16:59	27.07.2023 10:19:21	Оператор приёмщик	1176	Новосибирск	Транзитное хранение	Ж/Д
ID26	Определение локации работником склада	27.07.2024 13:44:56	27.07.2023 13:45:01	Оператор склада	691	Новосибирск	Транзитное хранение	Ж/Д
ID26	Согласование локации начальником смены	27.07.202413 :45:38	27.07.2023 13:47:09	Начальник смены	1721	Новосибирск	Транзитное хранение	Ж/Д
ID26	Согласование локации начальником смены	27.07.2024 13:47:26	27.07.2023 13:49:10	Начальник смены	1721	Новосибирск	Транзитное хранение	Ж/Д
ID26	Определение локации	28.07.2024 12:29:05	28.07.2023 12:29:52	Диспетчер склада	1997	Новосибирск	Транзитное хранение	Ж/Д
ID26	Определение локации работником склада	28.07.202416 :23:13	28.07.2023 16:24:15	Оператор склада	691	Новосибирск	Транзитное хранение	Ж/Д
ID26	Согласование локации начальником смены	28.07.2024 16:44:39	28.07.2023 17:08:00	Начальник смены	1721	Новосибирск	Транзитное хранение	Ж/Д
ID26	Определение локации	28.07.2024 17:50:26	28.07.2023 17:55:05	Диспетчер склада	1997	Новосибирск	Транзитное хранение	Ж/Д
ID26	Определение локации работником склада	28.07.2024 18:01:47	28.07.2023 18:03:16	Оператор склада	691	Новосибирск	Транзитное хранение	Ж/Д
ID26	Согласование локации начальником смены	28.07.2024 18:20:08	28.07.2023 18:24:29	Начальник смены	1721	Новосибирск	Транзитное хранение	Ж/Д
ID26	Груз размещён	29.07.2024 13:35:31	29.07.2023 13:35:27	Контролёр	561	Новосибирск	Транзитное хранение	Ж/Д

10. Слово процесса [Process word]

Слово процесса – это последовательность букв событий, отражающая порядок наступления событий в экземпляре процесса.

Например, последовательность событий для экземпляров процесса из таблицы 10:

- Process ID 1: слово «ABCDEFEFGHIJ»
- Process ID 2: слово «ABJ»

Таблица 10. Фрагмент журнала событий, где показано соответствие между событиями и буквами событий

Process ID	**Event**	**Буква события**	**Timestamp**
1	Приёмка Ж/Д терминал	A	03.06.2024 13:59:51
1	Оформление приходных документов	B	03.06.2024 14:05:51
1	Оформление документов контрагентов	C	03.06.2024 14:26:42
1	Ввод данных контрагента	D	03.06.2024 14:35:09
1	Оформление приходных документов	E	04.06.2024 9:47:40
1	Заменить	F	04.06.2024 14:05:15
1	Оформление приходных документов	E	04.06.2024 14:05:28
1	Заменить	F	05.06.2024 8:28:17
1	Определение локации	G	05.06.2024 8:28:35
1	Вызов погрузчика	H	05.06.2024 8:29:41
1	Размещение груза	I	05.06.2024 13:40:53
1	Груз размещён	J	05.06.2024 13:41:01
2	Приёмка Ж/Д терминал	A	04.06.2024 2:28:58
2	Оформление приходных документов	B	04.06.2024 3:54:28
2	Груз размещён	J	04.06.2024 5:09:07

Обозначение пути как слова процесса позволяет использовать обширный алгоритмический функционал обработки строковых выражений, в том числе делает возможным применение инструмента регулярных выражений.

Использование графов для визуализации процесса остаётся самым наглядным способом. Слова процесса более приспособлены для анализа и автоматизированной обработки данных и анализа.

11. Путь – [Path]

Путь процесса – объединяет в себе все экземпляры процесса с полностью одинаковой последовательностью операций. Сколько разных вариантов протекания продемонстрировал процесс – столько и путей.

Путь – последовательность событий, которое хоть раз встречалась в ходе процесса (из множества экземпляров процесса исследуемого датасета).

Путь описывает выполнение конкретного экземпляра или случая зарегистрированного процесса. Например, одно конкретное страховое возмещение, одно обследование с использованием рентгеновского аппарата, одно посещение веб-сайта одним конкретным пользователем.

Путь может быть обозначен в виде графа, либо в виде «Слова процесса».

12. Дина пути [Path length]

Длина пути процесса (Path length) – количество событий в процессе.

Если процесс представляется в виде слова процесса, то длина пути равна длине слова процесса.

Таблица 11. Длина пути

Trace	Path	Длина пути
ABCFNED	ABCFNED	7
ABCFNED	ABCFNECED	9
ABCFNED	ABCECFNED	9
ABCFNED	ABCECFNECED	11
ABCFNED	ABCFNFNED	9
ABCFNED	ABCFNFCED	9
ABCFNED	ABCFNEDCED	10

13. Счастливый путь [Happy Path (happy flow)]

(синонимы: Счастливый поток (Happy flow), Золотой путь (Golden path)

Счастливый (Happy path) – это наиболее часто встречающаяся последовательность событий среди экземпляров процесса.

Счастливый (основной) путь – это путь, в котором нет ошибочных ветвлений или циклов, начинающийся и заканчивающийся на валидные начальные / конечные события, по которому прошло максимальное количество экземпляров процесса.

Счастливый путь не может быть задан, он, может быть, только выявлен в процессе анализа логов событий.

В таблице 12, в поле «Путь» выделены наиболее часто встречающиеся последовательности событий процесса оформления поступления товара на склад в разрезе различных аналитик.

Таблица 12. Фрагмент таблицы различных путей процессов

Аналитика	Путь (слово)	Кол. Экз. процессов (id)
Аренда площадей	BKN	**826**
	AKN	469
	BEKN	420
Хранение транзитных грузов для таможенного оформления	BGKPSTN	**435**
	BGEGKPSTN	413
	BGKN	390
	BEGKPSTN	314
	BKN	306
	AGGKPSTZ	264
Ответственное хранение	BCKN	**557**
	BCGKPSTN	323
	ACKN	304
	BCN	265
Транзитное хранение	BGKN	**642**
	BKN	441
	AGKN	397

В таблице 12 наиболее часто встречающиеся пути процесса в разрезе каждой аналитики: BKN, BGKPSTN, BCKN, BGKN – это счастливые (основные) пути. Они начинаются валидным (предусмотренным) событием «B» и завершаются валидным событием «N», то есть приводят процесс к ожидаемому полезному результату.

Другие пути, не являющиеся «Happy Path», являются альтернативными путями процесса. Среди альтернатив можно выделить антипод «Happy Path» – «unhappy path», переводят по-разному: «Несчастный путь», «Путь исключения», «Бракованный путь». «Unhappy path» позволяют выделять брак в ходе осуществления процесса. Если счастливый путь один для каждой аналитики, то альтернативных путей может быть множество.

«Бракованный путь» не достигает требуемого результата – это главный критерий его определения. В примере выше – это путь «AGGKPSTZ» так как

начальное событие «A» и конечное событие «Z» не являются валидными начальными / конечными событиями.

14. Эталонный путь [Reference Path]

Эталонный путь (Reference path) – это спроектированный упорядоченный перечень событий среди экземпляров процесса, приводящий к желаемому результату.

Проектирование эталонного пути производится на начальном этапе проектирования бизнес-процесса, либо на этапе усовершенствования модели бизнес-процесса. Как правило, эталонный путь проектируется и фиксируется во внутренней нормативной документации.

Как и счастливый путь, эталонный путь должен быть различен для различных аналитик. Например, эталонные пути для получения клиентом наличности в банкомате или в отделении банка различны, хотя это один процесс – снятие наличности.

Не исключается случай, когда в реальном процессе или совокупности путей процесса может не встретиться эталонного пути процесса, или эталонный путь не будет соответствовать счастливому пути.

Не для всех бизнес-процессов должен быть запроектирован эталонный путь, так как не все, особенно мелкие процессы, проектируются. Поэтому допустимо принимать в качестве эталонного пути существующий счастливый путь.

Мерой отличия отдельного пути процесса от эталонного пути может быть расстояние Левенштейна.

Мерой соответствия полного графа процесса эталонному пути может быть стандартное отклонение (дисперсия) и вариация всей совокупности расстояний редактирования.

15. Трасса [Trace]

Трасса – перечень событий, присутствующих в экземпляре процесса или это конечная последовательность событий, в которых каждое событие появляется только один раз.

На практике часто встречаются случаи, когда в различных экземплярах процесса встречаются одни и те же события, но они могут повторяться, иметь различную последовательность и пр. Таким образом, **трасса** может являться объединяющим признаком для некоего перечня путей процесса (см. таблицу 13). В разрезе трассы можно рассчитывать различные процессные метрики.

Можно использовать трассу как идентификатор группы процессов (группы путей), как некую возможность кластеризации путей процесса.

С использованием трассы иерархия экземпляров процессов объединяется следующим образом: Путь экземпляра процесса -> Путь процесса -> Трасса.

Практическая применимость трассы для исследования процессов состоит в том, чтобы выявлять внутри небольших групп экземпляров процесса пути, последовательность смены событий в которых некорректна или процессы, не достигшие результата (не валидное финальное событие). Такие пути в таблице 13 выделены красным.

Для анализа трасса снижает размерность количества путей, что облегчает анализ.

Анализ трасс не является основным, или обязательным элементом анализа бизнес-процессов, но в некоторых случаях может иметь место.

Таблица 13. Пример трассы, объединяющей 139 экземпляров процесса и 28 путей процесса

Trace	Path	Кол. Экз. процессов (id)	Зацикленность %	Длина процесса (шагов)
ABCFNED	**ВСЕГО по трассе**	139	22,57%	11,61
ABCFNED	ABCFNED	69	0,00%	7
ABCFNED	ABCFNECED	17	18,18%	9
ABCFNED	ABCECFNED	11	18,18%	9
ABCFNED	ABCECFNECED	8	35,29%	11
ABCFNED	ABCFNFNED	6	18,18%	9
ABCFNED	ABCFNFCED	2	18,18%	9
ABCFNED	ABCFNEDCED	2	23,08%	10
ABCFNED	ABCFNECECED	2	35,29%	11
ABCFNED	ABCECECFNED	2	35,29%	11
ABCFNED	ABCECECFNECED	2	48,00%	13

ABCFNED	BDABCFNED	1	18,18%	9
ABCFNED	BDABCFNECE	1	23,08%	10
ABCFNED	BDABCECFNECE	1	36,84%	12
ABCFNED	BCECFNFNFNEDABCED	1	45,16%	17
ABCFNED	ADBCFNFNE	1	18,18%	9
ABCFNED	ADBCFNED	1	11,11%	8
ABCFNED	ADADABCECFNF	1	33,33%	12
ABCFNED	ADABCFNECECED	1	38,10%	13
ABCFNED	ABCFNFNEDABCF	1	35,00%	13
ABCFNED	ABCFNECEDAD	1	26,67%	11
ABCFNED	ABCEDBCFNED	1	26,67%	11
ABCFNED	ABCEDABCFNF	1	26,67%	11
ABCFNED	ABCECFNFNFNED	1	38,10%	13
ABCFNED	ABCECFNFNFCECED	1	46,43%	15
ABCFNED	ABCECFNFNED	1	26,67%	11
ABCFNED	ABCECFNECECECECED	1	63,83%	17

Рассмотрим в качестве примера логи, которые содержат несколько трасс.

Пример 1. Поиск полностью идентичных событий в логах

К полностью идентичным события относятся такие события, название и порядок которых в последовательностях событий логов полностью совпадают.

В качестве примера даны два лога, каждый из которых содержит следующие последовательности событий.

Лог1: ABCD, ACBD, AED.

Лог2: ABCBD, AED.

Каждая буква в логе событий является отдельным событием. Сочетание нескольких подряд идущих букв (например, AED) является последовательностью событий.

В рассмотренных двух логах события A, E, D последовательности событий AED являются полностью идентичными. Результатом нахождения полностью идентичных событий является последовательность AED.

Лог1: ABCD, ACBD, AED.

Лог2: ABCBD, AED.

Пример 2. Поиск различий в логах на один элемент в одном логе относительно другого лога (на одно событие)

К таким различиям относятся различия между последовательностями событий на один элемент (событие). Событие, может быть, как пропущено в одной из последовательностей событий относительно другой, как и добавлено лишнее.

Рассмотрим для примера два лога событий. Лог1 содержит уникальные последовательности событий ABCD, AED, в то время как Лог2 содержит события ABCED, AED. Каждая буква является отдельным событием.

Лог1: ABCD, AED.

Лог2: ABCED, AED.

В рассматриваемых логах событие E последовательности ABCED второго лога является так называемым лишним событием по сравнению с последовательностью ABCD первого лога. В то же время, для последовательности ABCD первого лога событие E является пропущенным (удаленным) событием.

Последовательности AED для первого и второго логов являются полностью идентичными последовательностями, как было рассмотрено в примере 1.

Лог1: ABCD, AED.

Лог2: ABCED, AED.

Пример 3. Поиск различий в логах на один элемент в обоих логах

К таким различиям относятся все различия в последовательностях событий, которые отличаются на одно событие, то есть в обоих логах присутствует событие, которое не происходило в другом логе.

Рассмотрим два небольших лога в качестве примера.

Лог1: AED, ABCD.

Лог2: ABFD, AED.

В данных логах последовательности событий ABCD и ABFD являются последовательностями, которые отличаются друг от друга ровно на одно со-

бытие. В последовательности ABCD первого лога событие C является событием, на которое можно заменить событие F в последовательности ABFD, чтобы две последовательности стали одинаковыми. Такие события будем называть измененные события в рамках данной работы.

Лог1: AED, ABCD.

Лог2: ABFD, AED.

Пример 4. Поиск различий в логах событий на несколько элементов в обоих логах

Ограничением данной работы является то, что будут рассмотрены все возможные различия в логах событий, описанные выше (удаленные, лишние, измененные события) с точностью до одного события. В качестве примера рассмотрим два лога, каждый из которых содержит по одной последовательности событий.

Лог1: ABCD

Лог2: ABCEFD

В рассматриваемом случае события E и F второй последовательности уже не будут отмечены как «лишние» события, так как это различие между последовательностями уже затрагивает два события.

16. Паттерны [Patterns]

Паттерны – это фрагменты пути процесса, состоящие из более чем одного события.

Паттерн имеет длину от 2-х событий до «длина пути – 1»

Например, для пути «ABJKD» паттерны это: AB, BJ, JK, KD, ABJ, BJK, JKD, ABJK, BJKD.

Часто паттерн может повторяться внутри одного пути процесса. Например, в слове процесса «ABJKDABJKS» паттерн BJK повторяется дважды, следовательно, один раз зацикливается. Этот же паттерн может повторяться и в других процессных путях.

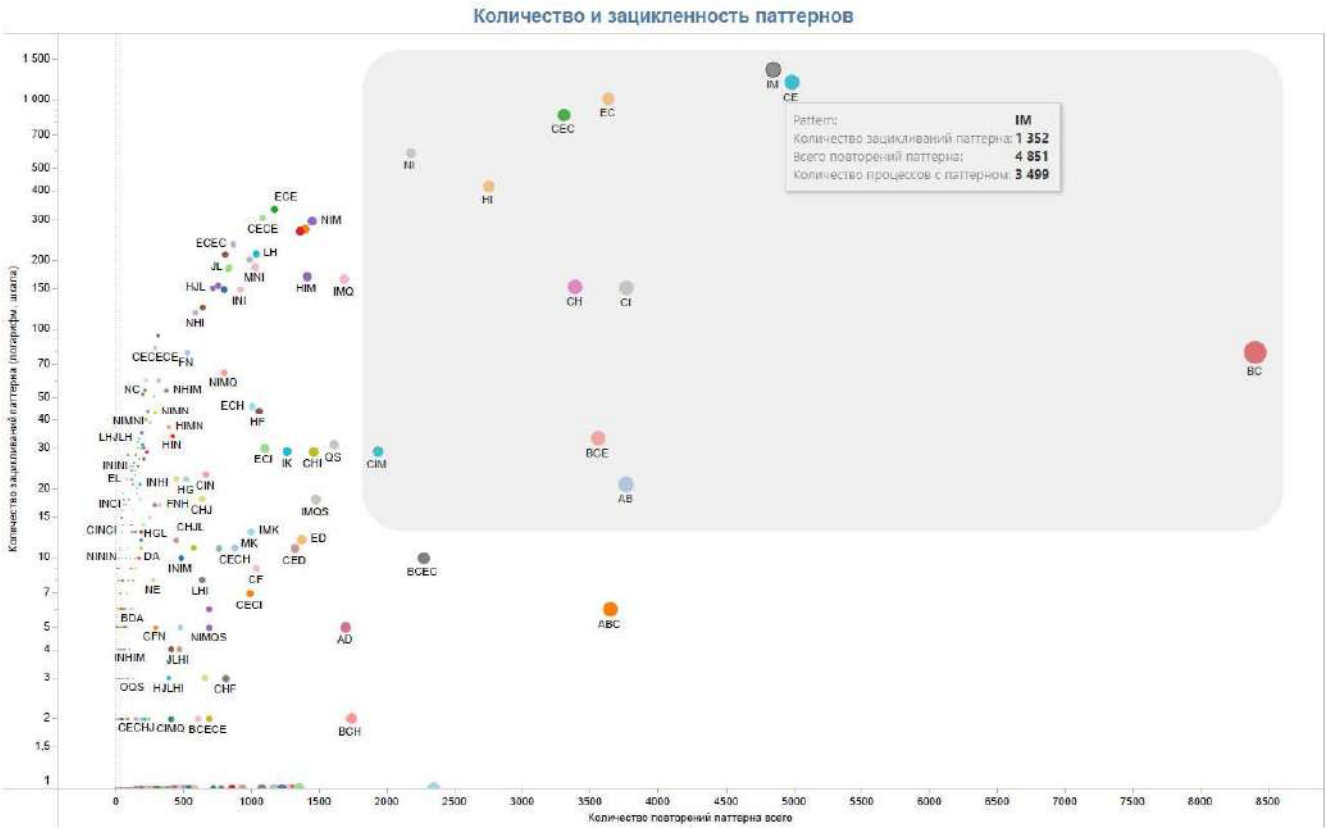


Рис. 1. Диаграмма рассеяния в осях: количество зацикливаний / всего повторений

На практике анализ паттернов может подсказать точки оптимизации процессов.

Примеры неэффективностей:

- Повторные паттерны, или зацикленные паттерны. Для анализа удобно построить диаграмму рассеяния как на рис.1. В правый верхний квадрант попадают паттерны, требующие оптимизации: необходимо уменьшать зацикленность. Оптимизация заключается в «спрямлении» процесса, удалении повторов.

- Если паттерн часто воспроизводится в различных путях, но мало зациклен (лежит или слабо отклоняется от оси «X»), то следует сосредоточиться на оптимизации событий, входящих в паттерн и оптимизации скорости реализации паттерна. Возможно, качественно поменять процесс, например, автоматизировав его с внедрением RPA.

Анализ паттернов не является основным средством анализа процессов, но в некоторых случаях он может дать ценную информацию о точках оптимизации бизнес-процессов.

Таблица 14. Статистика паттернов (фрагмент)

Pattern	Количество зацикливаний паттерна	Количество процессов с паттерном	Всего повторений паттерна	Доля зацикленных паттернов	Длина паттерна
IM	**1 352**	3 499	4 851	27,87%	2
CE	**1 188**	3 797	4 985	23,83%	2
EC	**1 011**	2 622	3 633	27,83%	2
CEC	**856**	2 452	3 308	25,88%	3
NI	**585**	1 593	2 178	26,86%	2
HI	**421**	2 328	2 749	15,31%	2
ECE	**332**	842	1 174	28,28%	3
CECE	**305**	785	1 090	27,98%	4
NIM	**297**	1 156	1 453	20,44%	3
IN	**272**	1 131	1 403	19,39%	2
MN	**268**	1 095	1 363	19,66%	2
IMN	**268**	1 093	1 361	19,69%	3
ECEC	**234**	638	872	26,83%	4
LH	**213**	823	1 036	20,56%	2
CECEC	**212**	597	809	26,21%	5
HJ	**202**	785	987	20,47%	2
MNI	**187**	847	1 034	18,09%	3
IMNI	**187**	846	1 033	18,10%	4
JL	**185**	658	843	21,95%	2

17. Концептуальный дрифт [Concept Drift]

Дрифт – это изменение с течением времени наиболее часто встречающихся путей процесса.

Наличие дрифта свидетельствует о происходящих изменениях в процессе.

Процесс под мониторингом может постепенно или скачкообразно измениться. Причиной изменения могут быть целенаправленные действия по улучшению процессов, а также эволюционные причины: изменение основных каналов продаж или коммуникации с клиентами, сезонных изменений и пр.

Концептуальный дрифт можно обнаружить, сравнивая Happy Path для одних и тех же измерений в различные длительные промежутки времени.

Например, какой наиболее частый путь встречается для процесса снятия наличности в отделениях банка за прошлый год или полгода, и за текущий период такой же длительности.

Дрифт не обязательно фиксируется при анализе Happy Path. Надо исследовать первые 1–5 путей процесса, по которым проходит примерно 60% всех экземпляров процесса.

При внесении целевых изменений в процесс необходимо мониторить процесс и отслеживать дрифт. Дрифт показывает на сколько быстро и качественно вносятся изменения в бизнес-процесс.

Термин *концептуальный дрифт* описывает ситуацию, при которой процесс видоизменяется по ходу анализа. Например, в начале журнала событий два действия могут осуществляться параллельно, в то время как ближе к концу журнала эти действия становятся последовательными. Процессы могут видоизменяться в силу периодических/сезонных колебаний (например, «в декабре спрос возрастает» или «по пятницам после обеда на рабочих местах меньше сотрудников») или по причине изменившихся условий (например, «рынок становится более конкурентным»). Подобные изменения влияют на процессы, и их выявление и анализ принципиально важны. Концептуальный дрифт процесса может быть обнаружен посредством разделения журнала событий на журналы меньшего размера и анализа «следов» в этих меньшего размера журналах. Такой анализ «второго порядка» требует существенно большего объема данных о событиях. Однако, стабильных во времени процессов совсем немного, и понимание концептуального дрифта является важнейшей задачей в управлении процессами. Поэтому для адекватного анализа концептуального дрифта нужны дополнительные исследования и разработка дополнительного функционала программного обеспечения.

18. Цикл [Cycle]

Циклы – это повторение событий или цепочек событий. Цикл – это повторный переход к событию «X» через промежуточные N событий.

Различают:

- «пинг – понг» [Ping pong] – это последовательность событий: A→B→A→B.
- Петля [Loop] – это последовательность событий: A→B→C→…→A.
- Короткие циклы [Short loops] – повторяющиеся вхождения одного и того же события подряд.
- Повторы [Iteration] – последовательно выполнение одинаковых событий: A→A→…→A.

19. Извлечение процесса [Process Discovery]

Process Discovery: один из трех основных видов PM. Process Discovery – это технология извлечение моделей процессов работы из протоколов работы в контексте программной инженерии На основе журнала событий формируется модель процесса.

Используемые нотации для описания моделей процесса:

- Конечные автоматы (FA)
- Системы переходов (Transition Systems)
- Сети Петри (Petri Nets)
- Сети потоков работ (Workflow nets)
- BPMN (Business Process Model and Notation)
- BPEL (Business Process Execution Language)
- YAWL (Yet Another Workflow Language)
- EPC (Event-driven Process Chains)
- Деревья процессов (Process Trees)
- Fuzzy and Heuristic Models
- Социальные графы (Social Graphs)

20. Алгоритмы извлечения процессов [Process extraction algorithms]

Алгоритмы извлечения процессов – это алгоритмы построения графа процесса, основываясь на последовательности событий в логе.

Обычно выделяют три основные, задающие целые семейства подходов

1) Alpha-алгоритм и Alpha+ алгоритм;

2) эвристические алгоритмы;

3) индуктивные алгоритмы

Алгоритмы обнаружения процессов:

- Alpha miner
- Alpha+, Alpha++, Alpha#
- FSM miner
- Fuzzy miner
- Heuristic miner
- Multi phase miner
- Genetic process mining
- Single/duplicate tasks
- Distributed GM
- Region- based process mining
- State-based regions
- Language based regions
- Classical approaches not dealing with concurrency
- Inductive inference (Mark Gold, Dana Angluin et al.)
- Sequence mining

Процедуры распознавания процессов на основе этих алгоритмов состоят из трех фаз: фаза извлечения (abstration) фаза, основная (induction) фаза и фаза конструирования (construction). Во время первой фазы просматривается каждое событие из журнала событий, и между событиями устанавливается отношение порядка (например, «событие 1 следует за событием 2», «событие 3 никогда не может следовать за событием 4»). Во время основной фазы выводятся более сложные отношения из базовых и могут быть созданы новые виды событий (невидимые и дублирующиеся). В качестве сложных отношений, например, можно выделить, зависимы ли события, или параллельны. На фазе построения (construction) конечная модель составляется на основе зави-

симостей выявленных на предыдущем этапе при использовании дополнительных эвристических правил.

21. Майнеры [Miner]

Майнер – алгоритм, определяющий последовательность следования нод (задач) относительно друг друга и связи между нодами. Результатом работы майнера является создание визуального графа процесса.

1. SimpleMiner – отображает все ребра, которые имеются в логе, фильтрация отсутствует. Такой вариант выглядит как куча связей, его анализировать сложно.
2. CausalMiner – отображает связь между узлами X и Y, если X→Y, в таком случае отфильтровываются связи, которые идут из Y→X.
3. HeuMiner – эвристический майнер. Удаляет неважные ребра, вес которых ниже заданного порога. Чем порог больше, тем меньше ребер изображено.
4. AlphaMiner – принимает во внимание причинно-следственные, параллельные и независимые отношения и создает сеть Петри, содержащую информацию о параллельном и XOR-подобном выполнении действий.
5. AlphaPlusMiner – модификация AlphaMiner, которая может обрабатывать случаи с циклом (логи, где присутствует связь: X→X).

22. Сети Петри [Petri nets, PN]

Сети Петри – основная форма представления модели бизнес-процесса в PM. Сети Петри – это высокоуровневое представление моделей процессов, которые позволяют компактно представить последовательное, параллельное, выборочное и циклическое выполнение между различными действиями в процессах.

Сеть Петри [4] (СП) – это ориентированный двудольный граф СП = (P, T, F, V, R), где P – конечное множество мест, T – конечное непустое множество переходов, $F \subseteq (P \times T) \cup (T \times P)$– множество направленных дуг, называемое

отношением маршрутизации, V – множество строковых имен, , а R – функция именования переходов.

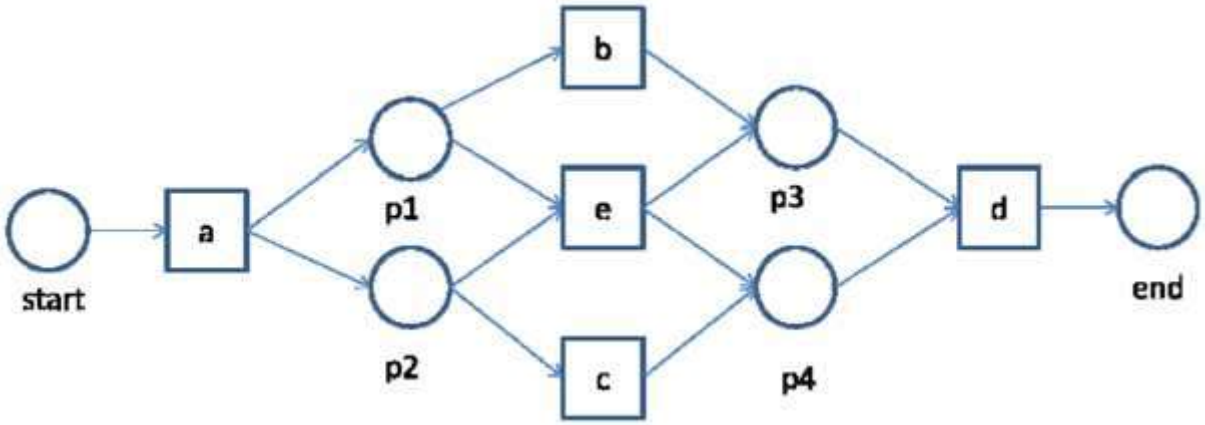


Рис. 2. Пример сети Петри (a, b, c, d, e – переходы (события), p1, p2, p3, p4 – позиции (условия наступления события)

Важные свойства сети Петри:

1. Асинхронная природа (поддерживается параллельное выполнение событий).

2. Выполнение сети Петри – это последовательность дискретных событий. Порядок появления событий является одним из возможных, допускаемых основной структурой. Это приводит к недетерминированности.

23. Сети потока работ Workflow nets (WF)

Workflow nets – это сеть, построенная на основе сетей Петри, позволяет представлять и анализировать рабочие процессы. Отличительными чертами WF являются:

- четкое начало – уникальное действие, которое только запускает все цепочки действий;
- четкий конец – уникальное действие, которое только заканчивает все цепочки действий.

24. DFG граф [Directly Follows Graph]

Directly Follows Graph – граф процесса, где вершины – это события журнала событий, а направленные ребра соединяют пары событий, происшедшие одно после другого хотя бы один раз. Плюс такого представления –

в подробном отображении всех возможных переходов. Минусом является чрезмерная загроможденность картинки соединяющими линиями, число которых резко возрастает при увеличении количества действий в журнале и многовариантности фактических шагов, случившихся в наблюдаемом процессе. DFG можно построить в разрезе частоты, т. е. сколько переходов было от одной вершины к другой. А можно построить в разрезе эффективности по времени, выбрав в качестве показателя среднее значение времени между событиями. Также в качестве показателя можно выбрать минимальное значение, максимальное или медиану.

25. Проверка соответствия [Conformance Checking]

Проверка соответствия (Conformance Checking) – это анализ соответствия фактической реализации бизнес-процесса, восстановленного на основе журнала событий и модели этого бизнес-процесса, выполненного, например, в BPM-системе. Проверка соответствия – одна из трех основных форм PM.

Основные задачи проверки соответствия:

- Увидеть «реальное» состояние бизнес-процесса в формате графической или аналитической модели;
- Сравнить «реальное» состояние и целевое состояние, зафиксированное в регламентах;
- Найти «узкие места» в бизнес-процессах;
- Выделить для последующего выборочного анализа «экстремальные» экземпляры процесса;
- Проанализировать процессные KPI;
- Проверить соответствие фактических параметров процесса Service Level Agreement (SLA) стандартам, заложенным в проекте (в первую очередь проверка длительности процессов).

Выявлять расхождения между смоделированным и наблюдаемым поведением помогают алгоритмы проверки соответствия. Они выдают показатели степени соответствия и диагностические сведения, объясняющие

наблюдаемые различия. С их помощью можно детально анализировать прецеденты, не соответствующие построенной модели.

Соответствие модели определяется только для конкретного журнала событий. Для проведения проверки соответствий требуется, помимо лога событий, некоторая заранее определенная модель. Это модель может быть построена вручную или получена с помощью методов извлечения. Для расчета соответствия журнал событий последовательно исполняется моделью.

Методы проверки соответствия принимают в качестве входных данных модель процесса и журнал событий и возвращают набор различий между поведением, зафиксированным в модели процесса, и поведением, зафиксированным в журнале событий. Эти различия могут быть представлены визуально (например, перегружены поверх модели процесса) или текстуально в виде списков операторов естественного языка (например, действие x выполняется многократно в журнале, но это не допускается в соответствии с моделью). Некоторые методы могут также производить нормализованные измерения (между 0 и 1), указывающие, в какой степени модель процесса и журнал событий совпадают друг с другом.

Трасса журнала событий считается соответствующей модели, если может быть получена в результате исполнения этой модели. Журнал событий идеально соответствует модели, если все его трассы соответствуют модели. На практике редко встречаются пары журнал – модель с идеальным соответствием. Поэтому обычно журнал считают соответствующим модели, если некоторая доля трасс в журнале соответствует модели. Значение уровня отсечки задается экспертом в каждом конкретном случае.

Интерпретация несоответствия зависит от цели модели:

- Если модель должна быть дескриптивной, разрыв между моделью и журналом указывает на то, что модель должна быть улучшена для лучшего отражения реальности.
- Если модель нормативна, то такие отклонения могут быть интерпретированы двумя способами: они могут обнаруживать нежелательные от-

клонения, то есть сигналы проверки соответствия, указывающие на необходимость лучшего контроля процесса, или могут обнаруживать нежелательные отклонения, то есть работники могут отклоняться, чтобы лучше обслуживать клиентов или справляться с обстоятельствами, не предусмотренными моделью процесса.

Механизмы проверки соответствия помогают аналитикам оценивать качество распознавания моделей процессов, но важнее, что эти механизмы применяются как вспомогательная технология при аудиторских проверках, оптимизации процессов и контроле исполнения нормативных актов.

26. Методы анализа процессов [Methods of process analysis]

1. Воспроизведение лога. Данный метод заключается в воспроизведении каждой последовательности лога по исходной модели сети Петри, с помощью чего можно обнаружить пропущенные фишки (токены), необходимые для достижения очередного перехода в модели. Все данные о проведенном исследовании могут быть собраны по завершении анализа и использованы в дальнейшем.
2. Анализ состояний. Метод заключается в обходе графа достижимости модели процесса, представленной в виде сети Петри, до тех пор, пока петли в графе не будут встречены хотя бы дважды.
3. Структурный анализ. Метод представляет собой анализ структуры модели процесса.

27. Гипотеза анализа [Analysis Hypothesis]

Ограничить анализ поиском гипотезы для конкретной проблемы можно путем фильтрации – в интуитивно понятном интерфейсе пользователь задает условия, по которым программа собирает только те кейсы, которые соответствуют задачам анализа.

28. Анализ причин [Root Cause Analysis, (RCA)]

Анализ причин **–** представляет собой метод решения задач PM, используется для выявления основных причин несоответствий или проблем. Цель

RCA – определить конкретные основные причины отклонений от желаемого процесса.

29. Предвзятость представления [Representational Bias]

Предвзятость представления – это выбор нотации для визуализации результатов извлечения процесса.

Некоторые нотации класса workflow **[workflow notations]**: IDEF0; BPMN; EPC (event-driven process chain).

30. Метрики [Metrics]

Процессная метрика – величина, рассчитываемая из исходных данных, содержащихся в процессном логе.

Метрика в организационном управлении – численный показатель качества, мера эффективности процесса.

Синонимами термина «метрика» являются: «показатель» [indicator], или «мера» [measure].

Метрики можно классифицировать следующим образом:

- **Процессные метрики** – описывают процесс и рассчитываются на данных, имеющихся в процессном логе. Перечень основных процессных метрик приведен в таблице 15 «Приложение 3. Таблица. Основные процессные метрики».

Перечень процессных метрик является ограниченным. В различных проектах рассчитываются дополнительные «пользовательские» процессные метрики. Process Mining в основном имеет дело с процессными метриками, но в некоторых случаях может дополняться бизнес-метриками.

- **Бизнес-метрики** – описывают состояние бизнеса. Как правило, бизнес-метрики рассчитываться с использованием дополнительных источников данных, ассоциативно связанных с процессным логом. *Примеры*: Воронка продаж, количество клиентов и прочее.

Перечень бизнес-метрик является открытым, он практически бесконечен и постоянно дополняем.

Замечание: в реальной практике Process mining, возникают вопросы:

- Какой ценой достигаются бизнес-KPI.
- Нет ли перенапряжения сотрудников в процессе.
- Следует ли ожидать текучки кадров.
- Совместный расчёт процессных и бизнес-метрик позволит обоснованно (Digital Driving) ответить на эти и сходные вопросы.

- **Ключевые показатели эффективности, KPI [*Key Performance Indicators*, KPI]** – метрики, которым придан статус «важных», на которые ориентируются менеджеры при принятии решений.

31. Основные процессные метрики [Key process Indicators, KPI]

В табл.15 указаны важнейшие процессные метрики (процессные KPI). Детальный перечень представлен в разделе «Приложение 3. Таблица. Основные процессные метрики».

Приложение 3. Таблица. Основные процессные метрики

Группа	№	Наименование метрики	Расчёт
Статистика	1.	Входной поток	Количество экземпляров процесса, начавшихся в течение наблюдаемого периода (час, сутки, пр.)
	2.	Количество процессов в работе	Количество экземпляров процесса, находящихся в работе в течение наблюдаемого периода (час, сутки, пр.)
	3.	Количество экземпляров процесса	Количество уникальных Case_ID
	4.	Количество сотрудников (User)	Количество уникальных User
	5.	Количество разновидностей путей процесса	Количество разновидностей путей процесса = количество разновидностей «слов» процесса
	6.	Количество уникальных событий	Количество уникальных событий “Event”
	7.	Количество событий (в разрезе событий)	Количество событий в журнале событий, в разрезе каждого события
Длительность	8.	Трудозатраты (Длительность чистая)	Разница между TimeStapm_End и TimeStamp_Start в рамках одного Case_ID. Рассчитывается только для журналов событий, имеющих два поля ДатаВремя (TimeStapm)
	9.	Длительность очереди (длительность бездействия)	Разница между TimeStamp_Start последующего события и TimeStapm_End текущего события в рамках одного Case_ID. Рассчитывается только для журналов событий, имеющих два поля ДатаВремя (TimeStapm)
	10.	Длительность фактическая	Разница между TimeStamp_Start по-

			следующего события и TimeStamp_Start текущего события в рамках одного Case_ID
	11.	Длительность чистая или фактическая рабочего времени	Длительности, из которых вычтено не рабочее время (напр., после конца рабочего дня, выходные, отпуска сотрудников и пр.)
	12.	Длительность на стороне клиента	Состоит из длительности событий, которые в экземпляре процесса выполнял клиент. Например, если клиент подготавливает документы для получения кредита, то этот период времени включается в фактическую длительность процесса, но может быть выделен как отдельное время, требующееся клиенту для деятельности на своей стороне
	13.	Длительность на стороне компании	Состоит из длительности событий, которые в экземпляре процесса выполнялись на стороне Компании. Например, если клиент подал документы для получения кредита, то время рассмотрения документов и решения по кредиту совокупно показывает длительность клиентского ожидания.
	14.	Длительность от начала экземпляра процесса (Running sum)	Сумма длительностей событий (трудозатрат или фактического времени), от начала процесса до каждого из событий в рамках одного Case_ID
Стоимость		Стоимость процесса	Сумма прямых затрат на осуществление процессной деятельности.
Эффективность		Зацикленность событий %	Рассчитывается как процент повторных, «лишних» событий, от всего количества событий в журнале
		Зацикленность по длительности % = (Потери времени на повторное исполнение элементов процесса)	Рассчитывается как процент длительности повторных, «лишних» событий, от совокупной длительности всех событий в журнале
Нагрузка (параллельное выполнение)		Нагрузка на процесс	Количество одновременно выполняющихся экземпляров процесса на момент времени
		Нагрузка на событие	Количество одновременно выполняющихся событий в параллельно выполняющихся, экземплярах процесса на момент времени
		Нагрузка на сотрудника	Количество одновременно выполняющихся событий на момент времени, исполнителем которых является User
		Производительность • процессная и • событийная	Количество экземпляров процесса или событий, выполнившихся (имеющих завершение) в течение наблюдаемого периода (час, сутки, пр.)

		Производительность сотрудников	Количество событий, завершенным сотрудником в течение наблюдаемого периода (час, сутки, пр.)
		Занятость сотрудников логируемая %	Доля рабочего времени сотрудника по данным журнала событий, которую он занят выполнением различных событий процесса
Очередь		Очередь событий	Количество событий на некий момент времени, для которых предшествующее событие завершилось, а текущее событие ещё не началось. Рассчитывается только для журналов событий, имеющих два поля ДатаВремя (TimeStapm)
		Очередь к ресурсу	Если событие выполняется с использованием ресурса, например, человеком), то очередь событий в разрезе ресурса позволяет оценить узкие места. Рассчитывается только для журналов событий, имеющих два поля ДатаВремя (TimeStapm
		Производительность	Количество событий в единицу времени. Максимальная производительность – количество обрабатываемых ресурсом событий до образования к нему очереди. Так как очередь рассчитывать не всегда возможно, то за производительность можно принимать максимальное за период количество событий, (или 90-й перцентиль количества), исполненных ресурсом.
Граф		% Переходов между событиями	Отношение количества переходов от события «X» к событию «Y» к сумме всех переходов от события «X» к другим событиям. Рассчитывается в разрезе каждого из событий
		Брак экземпляров процесса	Последовательность шагов в экземпляре процесса, которых не должно быть в штатной последовательности процесса. Цепочки «брак» определяются экспертом.
		Схожесть путей процесса	Рассчитывается как расстояние Левенштейна между рассматриваемыми путями процесса. В прикладном смысле, при мониторинге результатов мероприятий по улучшению процесса, можно рассматривать схожесть путей в различные промежутки времени и наблюдать уровень изменения процессов

Прежде чем прокомментировать расчёт некоторых метрик, приведём наглядный пример для понимания того, как рассчитываются метрики и как они взаимосвязаны.

Например, рассмотрим процесс обслуживания в супермаркете с одним, (условно), входом для покупателей, просторным торговым залом и кассовой зоной из «n» касс. Через вход свободно проходят люди, но если вдруг входной поток увеличится, то вход может оказаться узким, возникнет очередь и время доступа в магазин увеличится. Интерпретация: если, входной поток возрастает до некоторой критической производительности то после этого момента возможно возникновение очереди.

Для ресурса, работающего в «зелёной» зоне своей производительности, нет зависимости длительности события (здесь – входа в магазин) и количества событий (входящих людей). Но стоит входному потоку повысится, и может возникнуть очередь и время доступа резко увеличится и вход станет «бутылочным горлышком».

Математически можно рассчитать бутылочное горлышко как повышенную корреляцию нагрузки и длительности процессов под этой нагрузкой.

Далее, когда покупатели попадают в торговый зал, то они параллельно (каждый за себя) наполняют товаром свои продуктовые тележки. Если зал просторный и людей не так много, то они не ждут друг друга для подхода к полкам магазина. Интерпретация: В данном случае, нагрузка на торговый зал – это количество одновременно производимых обслуживаний = количеству человек в торговом зале. Если количество людей в торговом зале сильно вырастет, они будут мешать друг другу при доступе к полкам и длительность набора продуктовой корзины будет расти. Как раз проявится зависимость длительности процесса от нагрузки.

Длительность нахождения покупателя в торговом зале зависит от него: сколько товара он берёт, какой путь обхода полок выбирает, выбирает ли покупатель товар (положил в корзину – затем выставил обратно на полку)

Интерпретация: длительность обслуживания в данном случае, является ответственностью покупателя и если мы рассчитываем «узкие места» в магазине, то длительность процесса на стороне клиента можно исключить из расчёта длительность обслуживания клиента. Общая длительность процесса, конечно же, учитывает полный цикл нахождения покупателя в магазине от длительности входа до выхода.

Далее покупатель подходит к кассовой зоне. Если кассы свободны – он быстро оплачивает товар и уходит. А если кассы не справляются с нагрузкой, то на них стоит очередь. Очередь, это когда предыдущее событие завершилось, а следующее ещё не началось. С точки зрения покупателя – он завершил сборку продуктовой корзины, а к оплате не приступил. Наличие очереди – это чистая потеря его времени.

С точки зрения магазина – наличие очереди обеспечивает полную загрузку мощностей, а её отсутствие – означает простой кассиров. Поэтому магазинам часто выгодно создавать очереди, но небольшие, чтобы покупатели не ушли к конкурентам.

Но в целом, в процессах от очередей стремятся избавляться.

Если за мелкое событие процесса принимать добавление отдельного продукта в корзину, и, возможно, выкладку его обратно из корзины на полку, то мы можем рассчитать зацикленность таких событий. Таких событий, скорее всего, будет не одно.

Есть события, для которых зацикленность –обычное явление. В расчёт общей зацикленности по процессу они входить не должны.

Итого, для корректного расчёта времени обслуживания и зацикленности мы должны иметь два справочника:

1. Справочник событий на стороне клиента, который позволяет рассчитывать длительность на стороне клиента и длительность на стороне компании. Для расчёта общей длительности процесса такой справочник не нужен.

2. Справочник событий, не входящих в расчёт зацикленности.

Математически, для расчёта как нагрузки, так и очереди, требуется датасет с двумя полями DataTime: начала и завершения события, иначе нельзя будет понять, когда покупатель приступил к сборке продуктовой корзины, когда завершил сборку и когда начал «пробивать» покупки на кассе, и какой период времени (очереди) прошёл между этими событиями.

Каждая касса имеет свою производительность и пропускную способность. Это два взаимно обратных показателя. Пропускная способность – это количество покупателей в единицу времени, скажем, в час. Производительность – это время на обслуживания одного клиента.

Если входной поток увеличится выше пропускной способности, то начнут образовываться очереди.

Время обслуживания на кассе – это период от начала передачи покупателем товара кассиру, до проведения платежа. Это время на стороне компании, и его требуется минимизировать, как и очереди к кассе.

Нагрузкой для кассовой зоны из нескольких касс будет являться количество касс, которые параллельно обслуживают покупателей.

Если, например, при «пробивке» товара оказывается, что штрих-код не читается, или приходится заменять товар, то это следует отнести к зацикливанию и непроизводительным потерям времени.

Приведенный пример демонстрирует каким образом можно исследовать процесс на наличие «бутылочных горлышек», что очень часто ожидают от Process Mining. Более строго процесс поиска бутылочного горлышка описан в п. 31.4.

31.1. Длительность экземпляра процесса, события [Duration of process instance]

Длительность экземпляра процесса – это разница Timestamp от начала первого события в экземпляре процесса и максимального Timestamp в рамках одного Process ID.

Если в датасете указаны моменты начала и моменты завершения события, то из максимума Timestamp завершения событий вычитается минимум Timestamp начала событий.

Длительность принято измерять в целых секундах, хотя возможно использовать минуты или сутки или др. единицы времени.

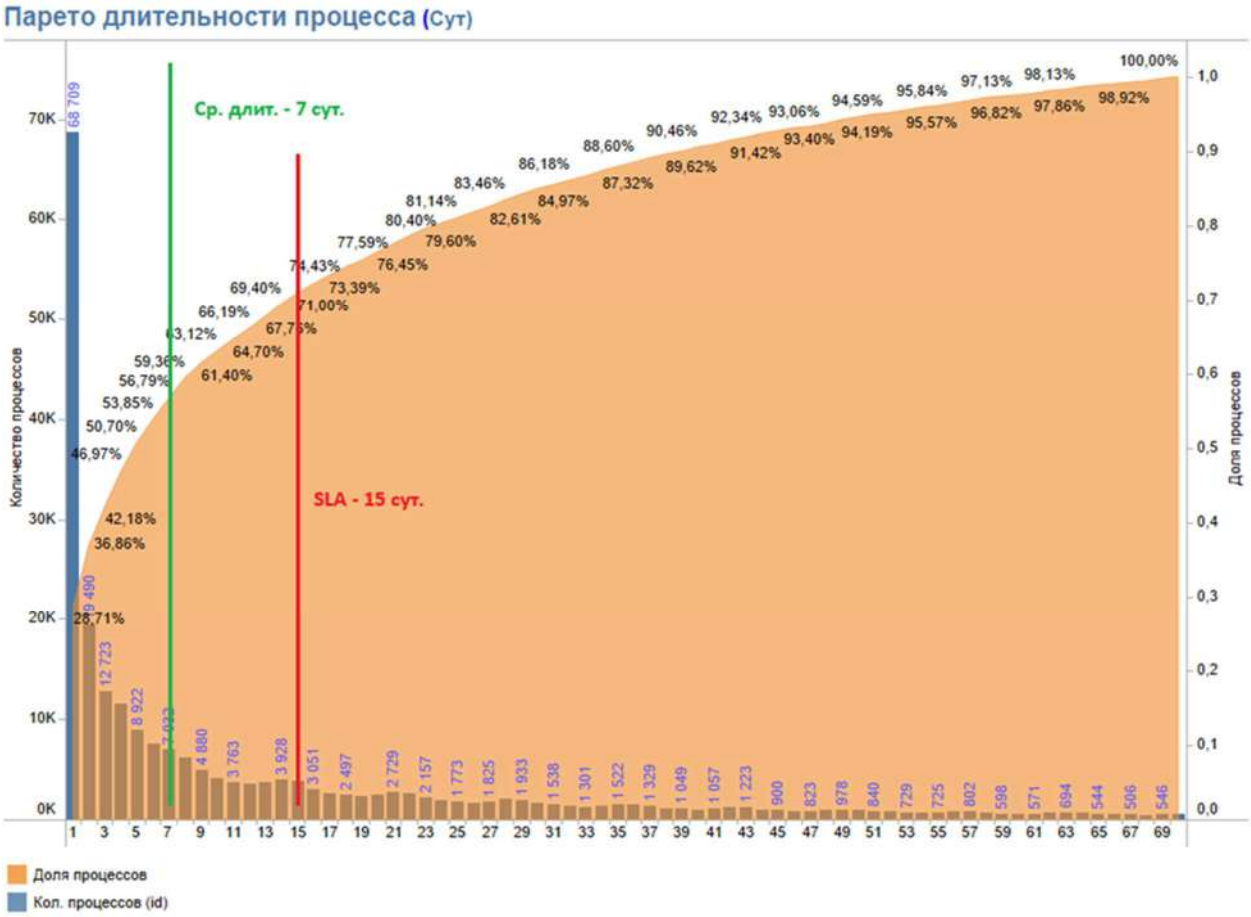


Рис. 3. Диаграмма Парето длительности процесса

Виды длительностей:

- **Трудозатраты** – или «чистая длительность» – рассчитываются как разница между моментами завершения и начала события (End Timestamp – Start Timestamp). Может быть рассчитана только в датасетах с двумя полями Timestamp. В случае наличия одного поля рассчитываются как «Фактическая длительность».
- **Длительность очереди** (длительность ожидания). Физический смысл – это период времени, когда поток процесса ожидает дальнейшей обработки, например ввиду занятости обрабатывающего ресурса.

Длительность ожидания – это время между моментом завершения предыдущего события и началом следующего, рассчитывается как разница между моментом начала текущего события и моментом завершением предыдущего события. Может быть рассчитана только в датасетах с двумя полями Timestamp. В случае наличия одного поля – не рассчитывается и приравнивается нулю.

В таблице 16 приведен пример расчёта длительностей с учётом параллелизма протекания событий (с двумя столбцами Timestamp).

Параллелизм возникает из-за того, что событие «Оформление приходных документов» завершается самым последним и имеет максимальное значение Timestamp в экземпляре процесса.

- **Фактическая длительность события** – разница между моментом начала и моментом окончания события.

Таблица 16. Пример расчёта длительностей

Process ID	Event	Start Timestamp	End Timestamp	Трудозатраты	Длит. Очереди	Фактическая длит. события
3	Приёмка Ж/Д терминал	03.06.2023 13:59:51	03.06.2023 14:05:21	330	NaN	360
3	Оформление приходных документов	03.06.2023 14:05:51	03.06.2023 14:23:55	1 084	30	1 251
3	Оформление документов контрагентов	03.06.2023 14:26:42	03.06.2023 14:34:24	462	167	507
3	Ввод данных контрагента	03.06.2023 14:35:09	03.06.2023 14:54:59	1 190	45	69 151
3	Оформление приходных документов	04.06.2023 9:47:40	04.06.2023 10:22:32	2 092	67 961	15 455
3	Заменить	04.06.2023 14:05:15	04.06.2023 14:05:19	4	13 363	13
3	Оформление приходных документов	04.06.2023 14:05:28	05.06.2023 14:21:20	87 352	9	66 169
3	Заменить	05.06.2023 8:28:17	05.06.2023 8:28:28	11	66 178	18
3	Определение локации	05.06.2023 8:28:35	05.06.2023 8:28:59	24	7	66
3	Вызов погрузчика	05.06.2023 8:29:41	05.06.2023 8:30:45	64	42	18 672
3	Размещение груза	05.06.2023 13:40:53	05.06.2023 13:57:53	1 020	18 608	1 040
3	Груз размещён	05.06.2023 13:58:13	05.06.2023 13:58:13	0	20	NaN
			Всего	**93 633**	**166 430**	**172 702**

Расчёт длительность процесса через min max длительность			
03.06.2023 13:59:51	05.06.2023 14:21:20	**Длительность процесса:**	**174 089**

31.2. Зацикленность [Rework]

Зацикленность – показывает долю повторного прохождения по одним и тем же событиям как внутри экземпляра процесса, в разрезе измерений, и в датасете в целом.

Зацикленность рассчитывается как отношение количества повторных событий в экземпляре процесса к количеству событий всего.

Например:

- В процессе ABCD – нет повторных событий и зацикленность = 0/4=0.
- В процессе ABBC – 1 повтор события «B» из 4-х шагов. Зацикленность = ¼ или 25%.
- В процессах ABBC и ABCDEDFC – в сумме 3 повтора из 12 шагов. Зацикленность для этих 2-х экземпляров процесса составляет 3/12 = 25%.

31.3. Нагрузка [Capacity]

Нагрузка – количество событий, одновременно исполняемых одним ресурсом.

Целесообразно рассматривать нагрузки:

- Нагрузка на событие – количество одновременно исполняемых событий.
- Нагрузка на User – количество событий (возможно, различных), которые одновременно находятся в работе у одного сотрудника.
- Нагрузка на процесс – количество одновременно исполняемых экземпляров процесса.

Нагрузка может быть рассчитана в разрезе любых ресурсов.

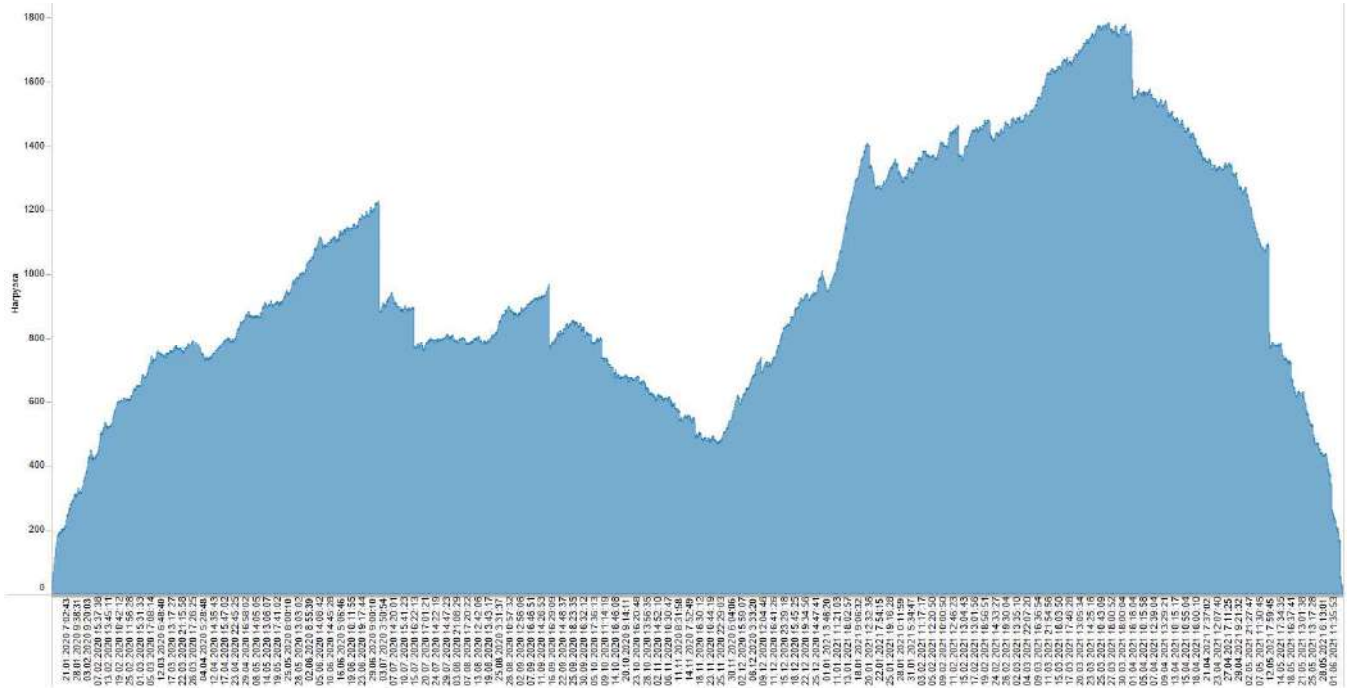


Рис. 4. Пример графика динамики нагрузки в посуточном разрезе

Нагрузка является одной из метрик, которые используются в расчёте показателя «**Ошибка! Источник ссылки не найден.**.

31.4. Бутылочное горлышко (узкое место) [Bottleneck]

Bottleneck – это ограничение системы, при котором теряется доля производительности или пропускной возможности системы. Подобные ограничения мешают системе достигать своих целей максимально эффективно.

«Бутылочное горлышко» – это слабое звено в цепочке действий и событий, из-за которого затрудняется и удлиняется весь процесс. Причиной может стать человеческий фактор – некомпетентность сотрудника, высокая нагрузка на людей, которые не справляются с одной из операций.

Иногда возникают проблемы в работе всего отдела: например, длительный цикл согласований от одного руководителя может приводить к невыполнению KPI всего департамента.

Узкое место часто возникает из-за ограничения ресурса оборудования или технических проблем. Например, станок способен произвести только определенное количество деталей или компьютеры, сервера или другие компоненты ИТ-инфраструктуры не справляются с нагрузкой.

Узкие места могут определяться по-разному, в зависимости от задачи: по самым длительным события или по самым зацикленным.

Наиболее обоснованным подходом является определение узких мест на основании расчёта очереди и нагрузки, в сочетании с анализом длительности и зацикленности.

Узкое место – это событие (или паттерн), к которому возникает наиболее длительная очередь. Очередь может быть рассчитана только на основании двух полей Timestamp.

Альтернативно можно определить узкое место как наиболее нагруженное событие, имеющее большую длительность и зацикленность.

Для определения узких мест следует проводить кластеризацию по набору метрик.

Примеры:

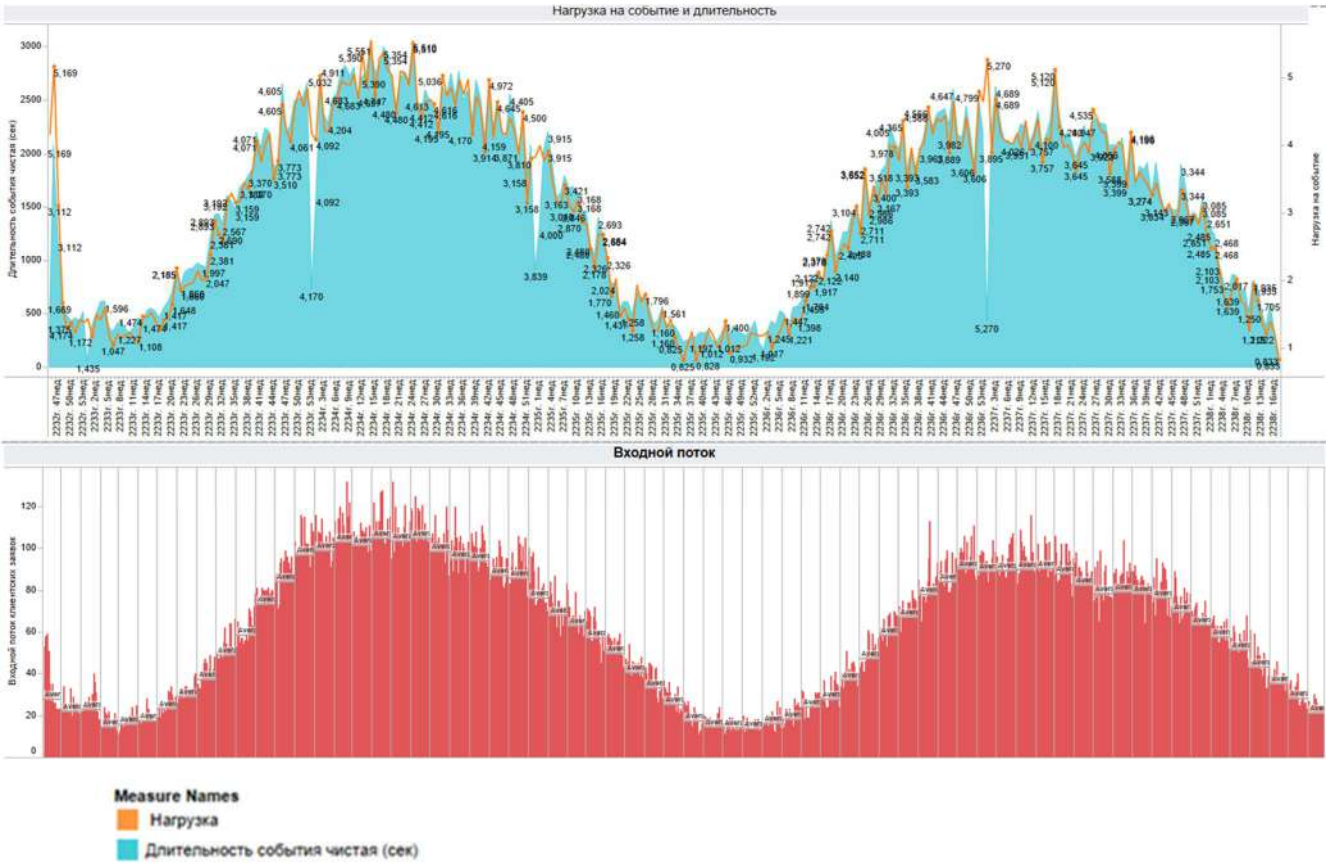


Рис 5. Зависимость нагрузки и длительности исполнения события от интенсивности входного потока

На рисунке 5 приведен кейс наблюдения за процессом в течение двух суток. Красная область внизу – это входной поток запросов, поступающих от

клиентов в информационную систему компании. Явно видны сезонные (тут суточные) колебания интенсивности, и «зашумление» входного потока.

В верхней части рисунка – наложение графиков нагрузки и длительности исполнения некоего функционала информационно-аналитической системы.

В терминологии Process Mining это зависимость нагрузки и длительности некоего события от динамики входного потока.

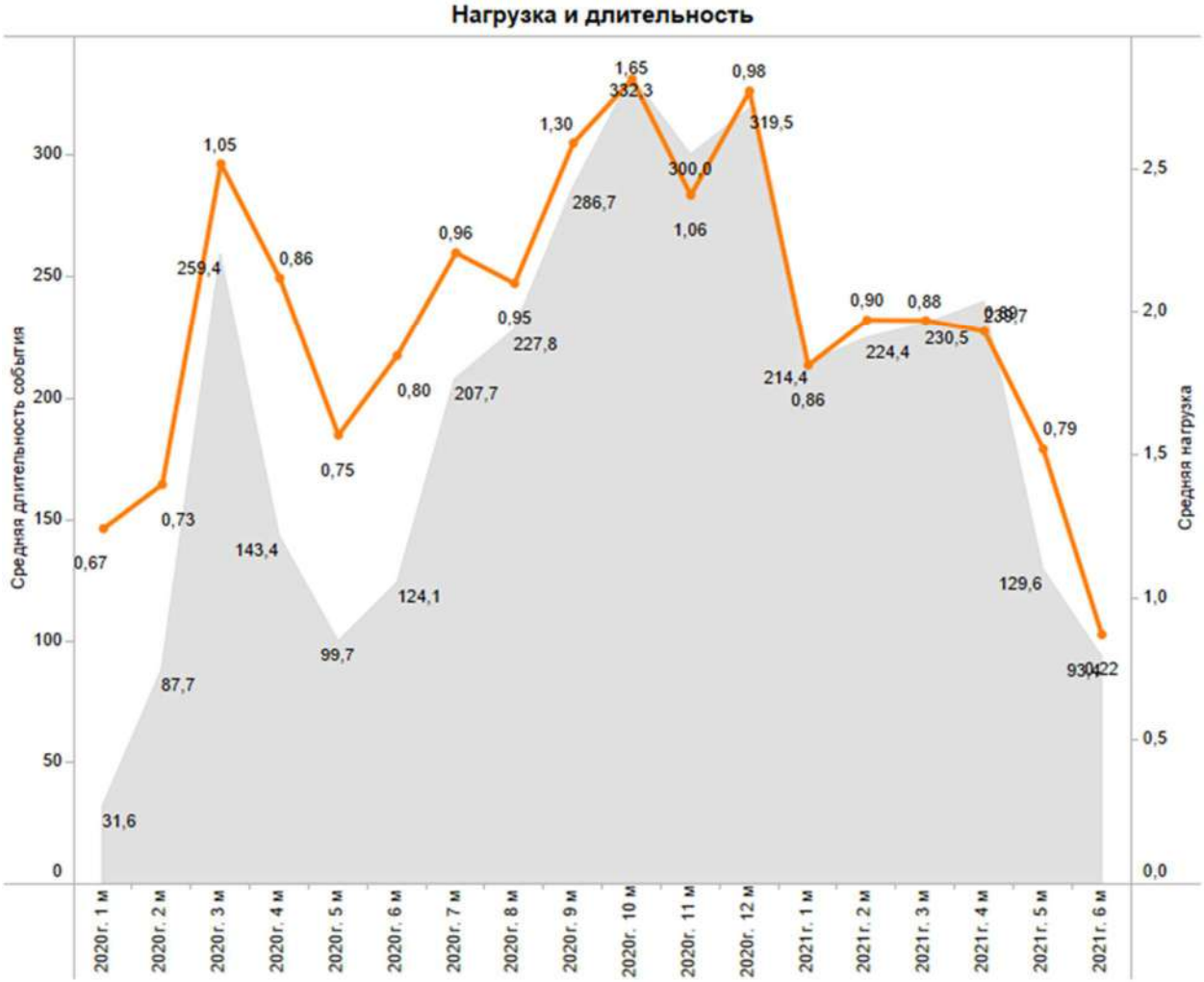


Рис. 6. Пример корреляции между нагрузкой на событие и его длительностью

На рисунке 6 выше построены графики средней нагрузки (оранжевая линия) и средней длительности события (серая область).

Если мы видим корреляцию: при увеличении нагрузки длительность исполнения события возрастает, то это событие – первый кандидат на статус «бутылочного горлышка».

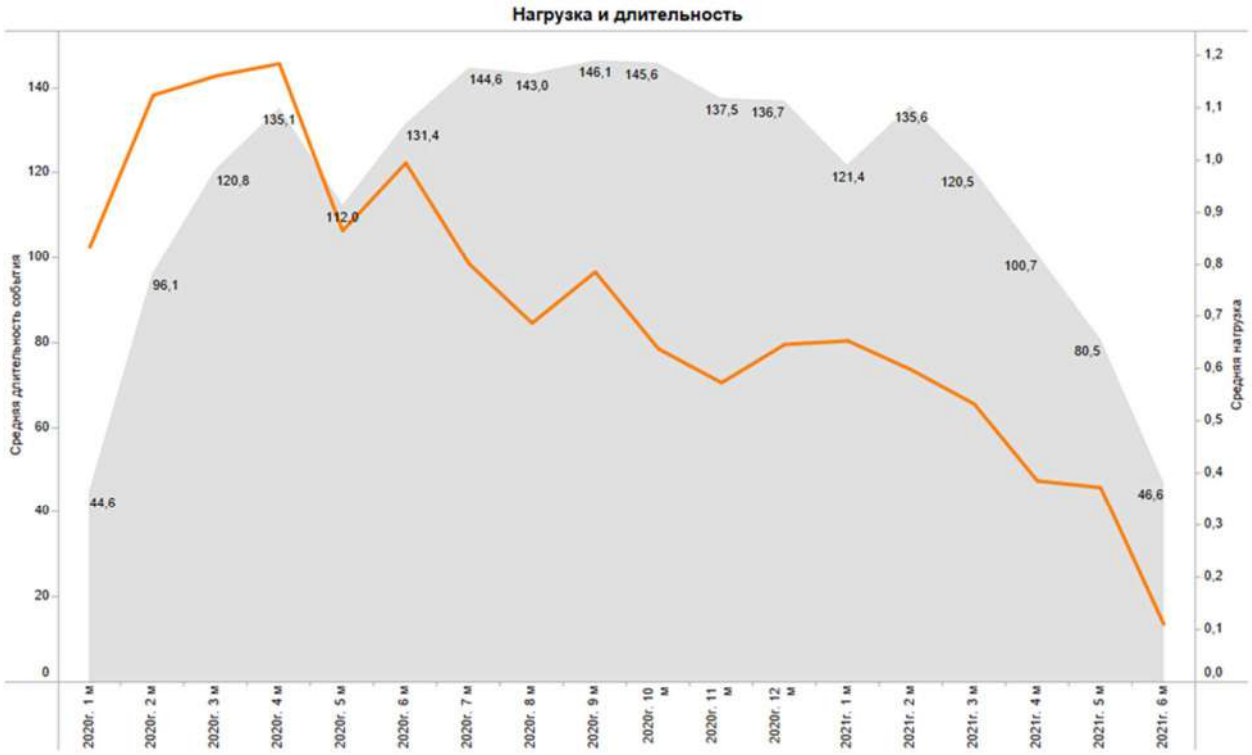


Рис. 7. Корреляция между нагрузкой и длительностью незначительна

На рисунке 7 приведён пример низкой зависимости длительности от нагрузки. В данном случае риска «бутылочного горлышка» нет.

Но надо понимать, что любой ресурс в процессе имеет некую пропускную способность, при превышении которой возникают узкие места, то есть для штатного входного потока процесса событие на рис. 7 более подвержено риску торможения процесса, чем другие события и ресурсы с ними связанные.

Для расчёта узких мест с использованием нагрузки и длительности не обязательно строить графики и просматривать их для всех событий. График – только иллюстративный материал.

Задача решается аналитически, через расчёт корреляций между нагрузкой и чистой длительностью события. Для корреляционного анализа следует подготовить данные, а именно усреднить значения нагрузки и длительности за некий временной интервал. Для данного анализа принималось усреднение за 1 месяц.

Так как нагрузку не всегда можно рассчитать, (для этого нужен датасет с полями начала и окончания события), то в качестве упрощения допустимо

рассчитывать корреляцию между входным потоком и длительностью. Но такой расчёт является более грубым приближением к реальности.

Для выявления узких мест, или подтверждения найденных при корреляционном анализе, можно использовать некоторые показатели в разрезе событий.

Таблица 17. Пример расчёта некоторых показателей в разрезе событий

Событие	Кол. событий	Доля % длительности событий от длительности экземпляра процесса	Зацикленность %	Ср. Нагрузка
Всего	**1 786 134**	**100,0%**	**23,55%**	**-**
Экспертиза документов на входе	118 959	29,3%	30,46%	4 617
Кредит предварительно утвержден	207 726	25,4%	38,34%	3 918
Клиент найден	153 157	13,0%	0,75%	2 022
Запрошена доп.инфо	121 596	7,6%	29,57%	1 127
Сделка подписана	46 232	6,3%	6,90%	1 025
Сделка передана в ДВУ	47 062	2,3%	5,96%	385
Возврат на доработку	14 537	1,6%	4,94%	267
Некомплект документов. Доработка	26 648	2,1%	21,73%	325
Андеррайтинг кредитоспособности	126 493	1,9%	34,24%	281
Отказ клиента после одобрения кредита	72 308	2,0%	4,33%	319
Передать в процессинг	198 682	1,5%	2,63%	237
Комплектация сделки	14 150	1,7%	20,26%	263
Передача договора в юр.отдел	50 942	1,4%	11,98%	229
Процессинг: проверка кредитного дела	300 470	1,1%	31,19%	182
Отказ клиента до принятия решения	99 480	1,2%	2,48%	185
Отказ Банка	50 620	0,8%	11,23%	122
Процессинг	96 101	0,5%	37,28%	79
Рефинансирование: доработка	1 025	0,1%	30,98%	18
Проверка документов после возврата	28 932	0,1%	25,46%	15
Проверка кредитного дела завершена	1 734	0,1%	42,64%	13
Кредит готов к рефинансированию	1 125	0,1%	8,76%	17
КД в работе юр.отдел	2 207	0,0%	12,80%	7
Проверка перед рефинансированием	2 130	0,0%	42,82%	1
Согласование сделки	1 669	0,0%	40,88%	1
На доработке	30	0,0%	3,23%	1
Экспертиза документов	18	0,0%	50,00%	1
Кредитный договор закрыт	2 079	0,0%	1,19%	5
Встреча проведена	22	0,0%	24,14%	1

В таблице 17 представлены показатели: усреднённые значения доли длительности события от длительности процесса, зацикленность и средняя

нагрузка (всё в разрезе событий). Таблица отсортирована по показателю доли длительности события в процессе.

В первой же строке видим событие «Экспертиза документов на входе». Оно занимает почти 30% времени процесса, более других событий зациклено и средняя нагрузка на него (в среднем 4617 дел рассматривается параллельно) тоже самая большая из всех. Количество таких событий достаточно большое, чтобы оказать заметное влияние на замедление процесса.

Надо сказать, что в этом примере «повезло» и простая сортировка позволяет выявить или подтвердить узкие места процесса.

Аналитически, в более общем случае, требуется выполнить кластеризацию по этим показателям. Кластер событий с максимальными значениями этих показателей является зоной внимания к поиску узких мест процесса.

31.5. Стоимость процесса

Стоимость процесса – это объём прямых затрат, в управленческом учёте относящихся на процесс.

В Process Mining к стоимости процесса относят прямые затраты. Это, в первую очередь, заработная плата (стоимость секунды / минуты / часа рабочего времени сотрудника). Или стоимость расходных материалов, электроэнергии и пр., ресурсов, потребляемых в процессе.

Как правило, в управленческом учёте такие котировки отсутствуют, и на практике используются драйверы расчётов. В качестве драйверов выступает чистая длительность (синоним трудозатраты), которая умножается на стоимость единицы времени. Более точный результат временных затрат даёт автоматический хронометраж.

32. Драйвер [Driver]

Драйвер **–** параметр, на основании которого могут быть рассчитаны другие метрики. Например, длительность может быть драйвером для расчёта стоимости процесса, стоимости аренды и пр.

33. Хронометраж [Timing]

Хронометраж – это определение чистого времени, которые пользователи затрачивают на работу в том или ином приложении на своей рабочей станции.

Цель хронометража Process Mining – мониторинг трудозатрат, определение порядка действий пользователя в рамках задачи и определение неэффективности различного рода и потерь времени на непроизводительные операции.

Данные по хронометражу, как правило, используются для:

- выработки нормативов трудозатрат,
- оценки совокупного рабочего времени на выполнение некоей задачи,
- оценка количества сотрудников для выполнения работы,
- контроль рабочей нагрузки на сотрудников, контроль напряженности ресурсов

Традиционно хронометраж выполняется контролирующим сотрудником, который с секундомером составляет карту рабочего времени исполнителей, участвующих в хронометраже. Таким способом можно охватить только малую часть всего объёма работ, слабо учитываются тренды изменения трудозатрат со временем, трудозатраты сотрудников различной квалификации.

Хронометраж в PM выполняется при помощи специального ПО, так называемых «агентов логирования», фиксирующих операции пользователя на своём рабочем месте при работе в различных приложениях. Автоматическая фиксация позволяет создать реальную картину хронометража, отслеживать динамику изменения трудозатрат, фиксирует дополнительные внеплановые, или маятниковые операций, выполняемые пользователем в рамках решения бизнес-задачи.

Хронометраж позволяет определить:

- Отработанное время
    - суммарное время работы сотрудников

- деление время на продуктивное время и время отвлечений

- Фотографию рабочего дня
  - перечень и длительности работы сотрудника с ПО, с которыми работал сотрудник
  - загруженность сотрудников: суточный профиль, общая переработка или недогруз сотрудников
- Тайм – трекинг
  - учет распределения времени в разрезе проектов, на задачи в разрезе проектов
  - оценку вклада сотрудников в проект
- Анализ бизнес-процессов
  - при «продвинутом» логировании возможно построение графа бизнес-процессов с использованием логов работы сотрудников на своих рабочих станциях. Для достижения такого результата требуется устанавливать соответствие между журналом событий информационной системы и логами, полученными «агентом логирования».
- Формирование табелей рабочего времени. Требуется интеграция с кадровыми системами
  - данные о рабочем графике
  - выходные, отпуска, отгулы и пр.
  - сверхурочные работы

34. Цифровой двойник [Digital twin]

Цифровой двойник – создание цифрового представления взаимосвязанных процессов. (Связь родительских и дочерних процессов). В идеале, во взаимосвязанную систему должны быть положены все основные массовые процессы организации. В реальности это трудно разрешимая задача.

При построении связи родительского и дочернего процессов, некоторые события родительского процесса, ранее являвшиеся неделимыми, могут быть представлены в виде дочерних процессов. Уровень вложенности теоре-

тически не лимитируется. Практически – отсутствует коммерческое ПО, которое могло бы системно анализировать взаимосвязанные бизнес-процессы, построенные на основе цифровых следов.

«Цифровой двойник» теоретически позволяет производить моделирование связанных бизнес-процессов, проектировать совершенствование бизнес-процессов с учётом их системного взаимного влияния.

35. Показатели качества восстановления процесса

- **Соответствие модели журналам событий (синоним «пригодность», «подгонка») [Fitness]** – метрика, которая определяет, насколько полученная модель допускает поведение, отраженное в журнале событий. Чем больше экземпляров бизнес-процесса из журнала описывается моделью, тем выше мера пригодности. Точная подгонка модели означает, что все трассы журнала возможно воспроизвести на модели от начала до конца.

- **Воспроизводимость [Replay Fitness] –** метрика, которая выражается так: если мы возьмем модель процесса и проведем ее симуляцию, получим ли мы такой же лог процессов, какой лег в основу графа процесса.

- **Точность [Precision]** – метрика, которая определяет, насколько полученная модель допускает поведение, значимо отличающееся от поведения, наблюдаемого в журнале событий. Модель с малой точностью – это модель с «недоподгонкой».

Можно построить универсальную простую модель, которая может воспроизвести любую конечную последовательность событий указанных типов. У такой модели будет высокая «пригодность», но низкая «точность», так как очень много вариантов выполнения процесса, которые допускаются моделью, но не присутствуют в журнале.

- **Простота [Simplicity]** – метрика, описывающая соответствие принципу «бритвы Оккама»: наиболее простая модель, с помощью которой можно описать поведение, наблюдаемое в журнале событий, является наилучшей

моделью. Метрика простоты модели количественно может быть измерена разными способами, например количеством нодов и дуг в модели.

Сложность модели бизнес-процесса приводит к ее бесполезности с точки зрения последующего анализа. Существует класс алгоритмов восстановления процессов, которые выдают хорошее значение метрики «простота», но за счет снижения показателей «пригодности» и «точности».

- **Обобщенность [Generalization]** – метрика, которая определяет, насколько полученная модель допускает поведение, которое ранее не наблюдалось.

Для получения высокого показателя метрики «обобщенность» поведение модели не должно ограничиваться только поведением, присутствующем в журнале. Модель с недостаточной «обобщенностью» является «переобученной».

Метрика, по смыслу являющаяся противоположностью «точности». Можно построить такую модель, в которой каждая отдельная ветка будет предназначена для отражения каждого отдельного экземпляра бизнес-процесса из журнала. Подобная модель характеризуется высокой мерой «пригодности», но низкой мерой «обобщенности», так как не будет позволять ни малейшей модификации экземпляров процесса в журнале.

Не имеет смысла анализировать вышеперечисленные метрики качества модели бизнес-процесса в отдельности, так как часто значительное улучшение одной метрики может привести к значительному ухудшению другой. Поэтому стоит принимать во внимание только средневзвешенную сумму метрик при минимально заданном пороге каждой отдельно взятой метрики.

36. Усовершенствование модели бизнес-процесса [Model Enhancement]

Улучшение бизнес-процесса через моделирование и создание проекта обновлённых процессов включает:

- Анализ модели процесса (алгоритма, отражающего системные связи)

- Определение возможностей (Process Discovery, Conformance Checking)
- Моделирование – поиск оптимальных и реализуемых параметров будущего процесса.
- Проектирование – практические рекомендации по изменению процесса для достижения модельных KPI.
- Изменения проектируются таким образом, чтобы снизить риск возникновения узких мест или сделать определенные последовательности процессов невозможными.
- Реализация изменений. Мониторинг.

37. Соглашение об уровне услуг [Service Level Agreement, SLA]

SLA в общем случае, содержит описание услуги, права и обязанности сторон и, **согласованный уровень качества предоставления** услуги.

Для Process mining часто применяется метрика длительности – как установленной допустимой границы события или процесса.

38. Мониторинг [Monitoring]

Мониторинг бизнес-процесса – постоянное наблюдение за состоянием процесса:

- Регулярный расчёт KPI за различные периоды,
- Отслеживание соответствия KPI проектным значениям и SLA
- Отслеживание порядка следования событий и обнаружение «запрещённых» переходов – или «брака» процесса.

Мониторинг является инструментом внедрения изменений. Инструментом мониторинга является система оповещений владельца процесса и ответственных лиц о выходе KPI за допустимые границы.

39. Моделирование бизнес-процесса, what – if анализ

Под моделированием, а точнее под анализом «что, если» понимается расчёт процессных метрик для состояния «как есть» и некоего желаемого состояния процесса или стрессовой ситуации.

«Как есть» – это перечень всех метрик процесса, рассчитанных на текущем периоде (текущем датасете).

Суть моделирования состоит в том, чтобы на базе исходного лога сгенерировать новый виртуальный процессный лог, но с тремя существенными изменениями:

1. Внести изменения в длительность событий.
2. Внести изменения в интенсивность входного потока процессов.
3. Использовать только допустимые пути процесса (пути без брака).

После генерации такого лога он загружается в систему Process Mining, которая рассчитает метрики процесса и покажет новые узкие места, производительность и прочие параметры.

Критичные метрики различных датасетов сравниваются и подбираются параметры процесса, к которым следует стремиться уже в организационном плане, меняя бизнес.

Основная математическая задача сводится к генерации нового датасета на основе изменённых параметров математического распределения текущего датасета.

У многих параметров есть распределение вероятности. Например, фактическое распределение длительности процессов на рисунке 8 соответствует распределению Парето. Если проанализировать длительность процессов, можно выявить, какие из них занимают больше всего времени, но приносят меньше всего пользы.

Распределение Парето описывает неравномерное распределение ресурсов. Для моделирования данных в областях, где нужно учесть широкий диапазон асимметрий и эксцессов, часто подходит **распределение Бёрра** (рис. 8).

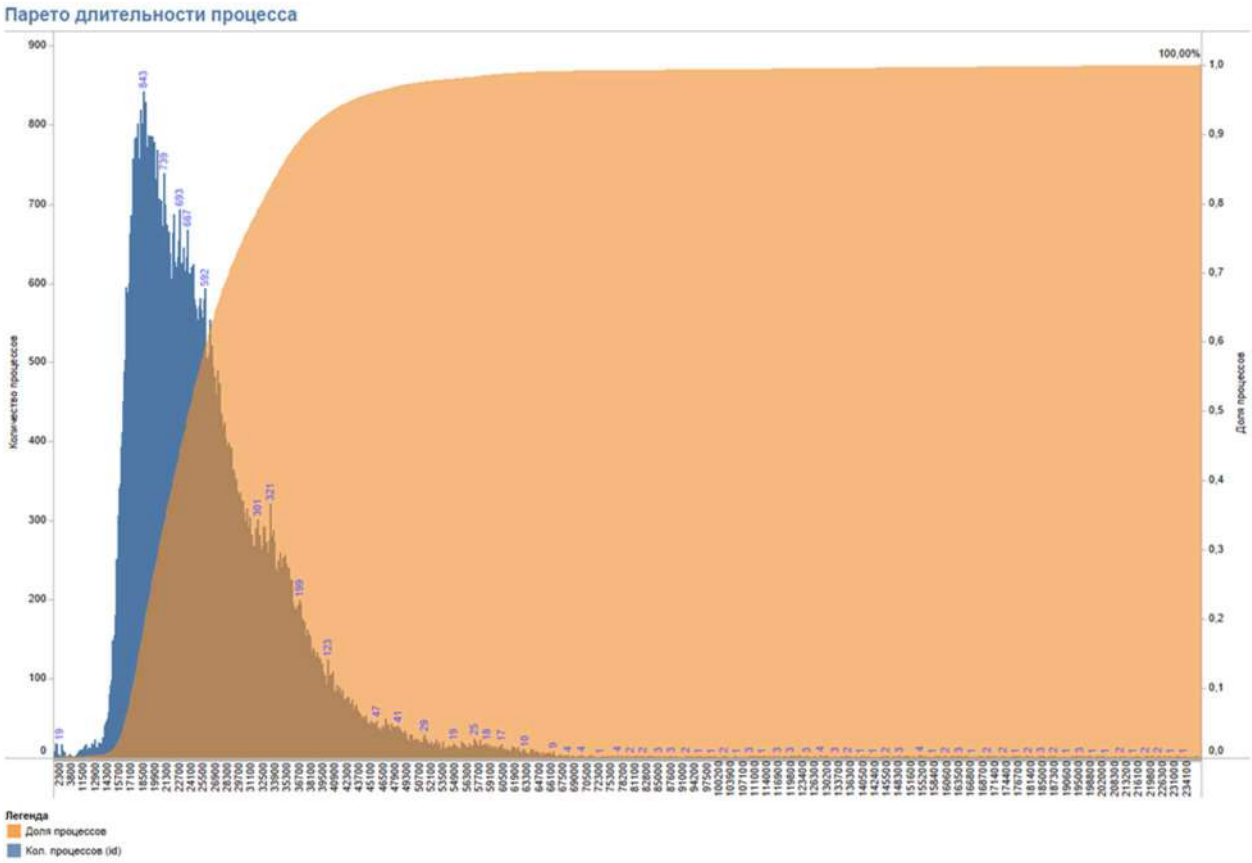


Рис. 8. Диаграмма Парето распределения длительности процессов

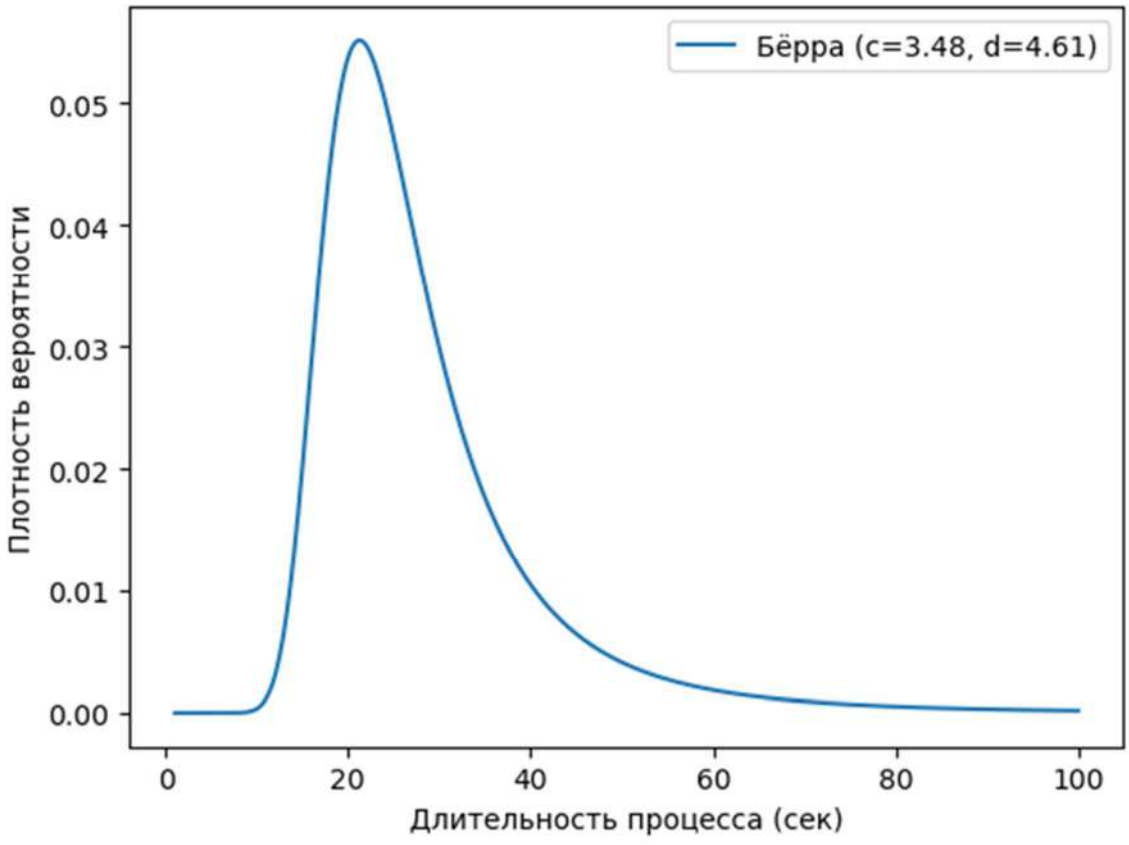


Рис. 9. Распределение Бёрра

Распределение Бёрра (Burr Distribution) – это гибкое распределение, которое может моделировать различные формы распределений, включая тяжелые хвосты.

Можно написать код на Python для оценки распределения длительностей процессов и каждого события.

Для генерации тестового журнала логов требуется решить обратную задачу:

- Изменить параметры распределения одного или всех как того требует задача моделирования: формируем целевые параметры распределения;
- Создать новый сгенерированный датасет, в котором длительности событий имеют уже целевое распределение.

Сгенерированный датасет может быть загружен в аналитическую систему Process Mining, где будут рассчитываться прочие метрики процесса.

Критичные метрики различных датасетов сравниваются и подбираются параметры процесса, к которым следует стремиться уже в организационном плане, меняя бизнес.

40. Workflow Mining

Workflow Mining – технология выявления часто встречающихся экземпляров процессов (шаблонов) из протоколов работы систем. Понятие по смыслу близкое Process Mining.

41. Этапы Process Mining

Подготовительный этап. Логирование, входной аудит и подготовка данных [**ETL & Data Preparation**]. Сбор и обработка данных.

1. Открытие, обнаружение реального процесса [**Discovery**] (модель AS IS, расчёт метрик). Автоматическое создание карты процессов. На основе собираемых данных строится карта рабочих процессов в режиме реального времени.

2. Проверка соответствия [Conformance checking] (Сравнение с проектом процесса или экспертной моделью)**.** Анализ всей цепочки событий

бизнес-процесса. Построенная карта позволяет увидеть эффективность операций внутри процесса – отклонения от регламентов и соответствия стандартам

3. Усовершенствование [Enhancement] (анализ, оптимизация, моделирование, имплементация). На основании полученной картины устраняются «узкие места», лишние операции, тупиковые процессы

4. Мониторинг [Monitoring] – инструмент управления изменениями. Суть мониторинга в отслеживании динамики выбранных процессных KPI и управления процессом для достижения целевых значений KPI. Отслеживание состояния процесса «до» и «после» внесения изменений.

Литература

1. Aalst W. M. P. van der Process Mining – Discovery, Conformance and Enhancement of Business Processes. – Springer, 2023. – С. I–XVI, 1–352

2. Agrawal R., Gunopulos D., Leymann F. Mining process models from workflow logs // Advances in Database Technology – EDBT'98 / под ред. H.-J. Schek [и др.]. – Berlin, Heidelberg : Springer Berlin Heidelberg, 1998. – С.467–483

3. Манифест Process Mining [Электронный ресурс], – URL: https://www.tf-pm.org/upload/1590128200840.pdf дата обращения: 20.01.2025).

4. Баргесян А.А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. – 3-е издание перераб. и доп. – СПб.: БХВ-Петербург, 2009.

5. Loginom Process Mining [Электронный ресурс] – URL : https://marketplace.loginom.ru/solutions/lpm/ (дата обращения: 20.01.2025).

6. Опыт интеграции ETL Loginom и BI Visiology в эпоху больших данных [Электронный ресурс], – URL : https://loginom.ru/sites/default/files/global-files/textpage/2023/vlasov_visiology.pdf (дата обращения: 20.01.2025). – Режим доступа: свободный. – Текст: электронный.

7. Использование технологии Process Mining (процесс майнинг) в медицине [Электронный ресурс] – URL : https://processmi.com/blog/ispolzovanie-tehnologii-process-mining-v-mediczine / (дата обращения: 20.01.2025). – Режим доступа: свободный. – Текст: электронный.

8. Nikola Trcˇka, Mykola Pechenizkiy, and Wil van der Aalst. Process Mining from Educational Data/ Taylor and Francis Group, LLC, 2011 – URL: https://www.win.tue.nl/~mpechen/publications/pubs/ch9edmbook.pdf (Дата обращения: 10.01.2025) – Режим доступа: свободный. – Текст: электронный

Приложение 1. PM Ready формат датасета

№	Наименование поля	Тип данных	Обязательность	Множественность	Иерархия	Описание – логическое предназначение	Комментарий
16.	**ID_Process**	String	Обязательное	1 поле	нет	ID экземпляра процесса	**Идентификатор экземпляра процесса**. Обязательное поле.
17.	**Event**	String	Обязательное	1 поле	нет	Текстовое наименование статуса (шага) процесса	**Наименование события**. Обязательное поле. Поле несёт смысловую нагрузку о ходе бизнес-процесса
18.	**DataTime_Start**	Datatime	Обязательное	1 поле	нет	Дата / время начала выполнения события процесса	**Временная метка начала события**. Обязательное поле.
19.	**DataTime_End**	Datatime	Желательное	1 поле	нет	Дата / время окончания времени выполнения события процесса	**Временная метка окончания события**. Желательное, но не обязательное поле. Формат аналогичен полю «**DateTime».** Наличие поля позволяет существенно увеличить глубину анализа: анализировать чистое время, очереди, занятость и пр.
20.	**Unit**	String	Желательное	1 поле	нет	Организационная часть компании, исполняющая процесс.	**Unit -** может быть территориальной организационной единицей, или подразделением в компании, исполняющим процесс. За рамки Unit-а процесс выходить не должен, но различные экземпляры процесса могут параллельно исполнять несколько Unit. Unit используется для расчёта метрик очереди и нагрузки.
21.	**City**	String	Желательное	1 поле	нет		
22.	**User_ID**	String	Желательное	1 поле	нет	ID сотрудника, выполняющего шаг процесса	**Идентификатор пользователя** Желательное, но не обязательное поле. USER – менеджер, сотружник банка, пользователь IT систем Банка. (Не клиент)
23.	**Measure_Name**	String	Желательное	Множество полей	да	Наименование измерения	**Наименования измерений.** Желательное, но не обязательное поле. Полей измерений может быть множество, их наименования специально не регламентируют-

							ся. Наличие измерений позволяет проводить более глубокий анализ процесса. **Например**, измерениями может быть структура подчинения. Либо наименование вида события, digital канала или наименования продуктов и пр. Измерения могут быть иерархичными. **Например,** 2 поля: продукт, субпродукт
24.	**Stage**	String	Не обязательное	1 поле	нет	Этап, к которому относится событие	**Наименование этапа к которому относится событие или группа событий.** Может подгружаться отдельно, справочником
25.	**ID_Status**	String	Не обязательное	1 поле	нет	Код статуса (события) процесса	ID события (если есть)
26.	**User_Role**	String	Желательное	1 поле	нет	Роль пользователя	**Идентификатор роли пользователя в системе** Желательное, но не обязательное поле.
27.	**Client_ID**	String	Не обязательное	1 поле	нет	ID клиента	**Идентификатор клиента Банка** Не обязательное поле
28.	**Minute_Cost**	Float	Желательное	1 поле	нет	Стоимость минуты чистого времени	**Стоимость единицы времени user** при выполнении статуса. Желательное, но не обязательное поле.
29.	**ID_Parent_Process**	String	Желательное	Множество полей	Возможно	ID родительского процесса	**Является ключевым полем для связи с родительским** процессом, породившим дочерний процесс, или связкой с событием в родительском процессе, являющееся отдельным бизнес-процессом в дочернем. Используется для построения Digital Twin.
30.	**Text**	String	Желательное	Множество полей	нет	Комментарии, текст переписки, прочая текстовая информация коммуникация в процессе исполнения шага	**Текстовые комментарии и переписка.** Желательное, но не обязательное поле. Полей может быть несколько. Наличие текстовой переписки позволяет глубже идентифицировать проблемные зоны процесса.
31.	**IT_System**	String	Не обязательное	1 поле	нет	Наименование информационной системы процесса.	**Наименование или ID ИС источника данных.** Не обязательное поле. Желательно, если дан-

							ные для анализа поступают из нескольких ИС. Используется для анализа интеграционных потоков между системами, нагрузки на системы и пр.
32.	**Region_latitude**	Geoteg	Не обязательное	Множество полей	да	Широта и долгота для отображения на картах.	**Географические координаты.** Не обязательное поле. Используется для отображения на картах, если требуется указать место положение объекта, отличающегося от области, города. (напр. Положение ВСП). ***Для информации***: справочник координат городов РФ Коды регионов в соответствии с Классификатором адресов Российской Федерации (КЛАДР).
33.	**Region_longitude**						

Приложение 2. Пример датасета

Бизнес – процесс «Размещение товаров на складе»

Process ID	Event	Start Timestamp	End Timestamp	City	Unit Name	User	Product	Channel	User Function
3	Приёмка Ж/Д терминал	03.06.2023 13:59:51	03.06.2023 14:05:21	Саранск	Базовый склад	895	Таможня	Ж/Д	Комплектовщик
3	Оформление приходных документов	03.06.2023 14:05:51	03.06.2023 14:23:55	Саранск	Базовый склад	93	Таможня	Ж/Д	Оператор приёмщик
3	Оформление документов контрагентов	03.06.2023 14:26:42	03.06.2023 14:34:24	Саранск	Базовый склад	922	Таможня	Ж/Д	Оператор приёмщик
3	Ввод данных контрагента	03.06.2023 14:35:09	03.06.2023 14:54:59	Саранск	Базовый склад	897	Таможня	Ж/Д	Оператор приёмщик
3	Оформление приходных документов	04.06.2023 9:47:40	04.06.2023 10:22:32	Саранск	Базовый склад	93	Таможня	Ж/Д	Оператор приёмщик
3	Заменить	04.06.2023 14:05:15	04.06.2023 14:05:19	Саранск	Базовый склад	2182	Таможня	Ж/Д	Контролёр
3	Оформление приходных документов	04.06.2023 14:05:28	04.06.2023 14:06:51	Саранск	Базовый склад	93	Таможня	Ж/Д	Оператор приёмщик
3	Заменить	05.06.2023 8:28:17	05.06.2023 8:28:28	Саранск	Базовый склад	2182	Таможня	Ж/Д	Контролёр
3	Определение локации	05.06.2023 8:28:35	05.06.2023 8:28:59	Саранск	Базовый склад	1127	Таможня	Ж/Д	Диспетчер склада
3	Вызов погрузчика	05.06.2023 8:29:41	05.06.2023 8:29:41	Саранск	Базовый склад	1126	Таможня	Ж/Д	Диспетчер склада
3	Размещение	05.06.2023	05.06.2023	Са-	Базовый склад	602	Таможня	Ж/Д	Оператор скла-

	груза	13:40:53	13:40:53	ранск					да
3	Груз размещён	05.06.2023 13:41:01	05.06.2023 13:40:57	Саранск	Базовый склад	2150	Таможня	Ж/Д	Контролёр
10	Приёмка Ж/Д терминал	04.06.2023 2:28:58	04.06.2023 2:42:23	Саранск	Территория хранения "Южная"	101	Транзит до 30 сут	Ж/Д	Комплектовщик
10	Оформление приходных документов	04.06.2023 3:54:28	04.06.2023 3:57:02	Саранск	Территория хранения "Южная"	105	Транзит до 30 сут	Ж/Д	Оператор приёмщик
10	Груз размещён	04.06.2023 5:09:07	04.06.2023 3:57:02	Саранск	Территория хранения "Южная"	102	Транзит до 30 сут	Ж/Д	Контролёр
5	Приёмка	04.06.2023 3:02:31	04.06.2023 3:41:20	Тула	Склад №1	1181	Транзит до 30 сут	Авто	Комплектовщик
5	Ввод данных контрагента	04.06.2023 4:45:41	04.06.2023 5:50:11	Тула	Склад №1	534	Транзит до 30 сут	Авто	Оператор приёмщик
5	Определение локации работником склада	04.06.2023 4:51:46	04.06.2023 5:01:05	Тула	Склад №1	2412	Транзит до 30 сут	Авто	Оператор склада
5	Согласование локации начальником смены	04.06.2023 5:02:01	04.06.2023 5:06:58	Тула	Склад №1	1555	Транзит до 30 сут	Авто	Начальник смены
5	Корректировка вх. документов	04.06.2023 5:23:45	04.06.2023 5:28:26	Тула	Склад №1	560	Транзит до 30 сут	Авто	Оператор приёмщик
5	Определение локации работником склада	04.06.2023 5:40:54	04.06.2023 10:51:12	Тула	Склад №1	2412	Транзит до 30 сут	Авто	Оператор склада
5	Согласование локации начальником смены	04.06.2023 5:41:37	04.06.2023 5:44:12	Тула	Склад №1	1555	Транзит до 30 сут	Авто	Начальник смены

5	Определение локации	06.06.2023 10:44:29	06.06.2023 10:46:04	Тула	Склад №1	15	Транзит до 30 сут	Авто	Диспетчер склада
5	Груз размещён	22.06.2023 13:04:49	22.06.2023 13:06:18	Тула	Склад №1	561	Транзит до 30 сут	Авто	Контролёр
5	Заменить	23.06.2023 13:06:18	23.06.2023 13:10:37	Тула	Склад №1	2402	Транзит до 30 сут	Авто	Оператор склада
5	Определение локации	25.06.2023 6:35:54	25.06.2023 6:49:02	Тула	Склад №1	15	Транзит до 30 сут	Авто	Диспетчер склада
5	Определение локации работником склада	25.06.2023 7:37:23	25.06.2023 7:37:40	Тула	Склад №1	2412	Транзит до 30 сут	Авто	Оператор склада
5	Согласование локации начальником смены	25.06.2023 10:04:01	25.06.2023 10:07:02	Тула	Склад №1	1555	Транзит до 30 сут	Авто	Начальник смены
5	Замена сопроводительных документов	26.06.2023 5:38:06	26.06.2023 5:39:23	Тула	Склад №1	542	Транзит до 30 сут	Авто	Оператор приёмщик
5	Определение локации работником склада	26.06.2023 14:57:16	26.06.2023 14:57:43	Тула	Склад №1	2412	Транзит до 30 сут	Авто	Оператор склада
5	Согласование локации начальником смены	27.06.2023 3:30:17	27.06.2023 4:15:51	Тула	Склад №1	1555	Транзит до 30 сут	Авто	Начальник смены
5	Возврат документов контрагенту	27.06.2023 8:18:35	27.06.2023 8:22:22	Тула	Склад №1	1389	Транзит до 30 сут	Авто	Начальник смены
5	Замена сопроводительных документов	27.06.2023 8:47:16	27.06.2023 8:53:36	Тула	Склад №1	1541	Транзит до 30 сут	Авто	Начальник смены

5	Заменить	27.06.2023 9:32:23	27.06.2023 9:32:52	Тула	Склад №1	565	Транзит до 30 сут	Авто	Контролёр
5	Определение локации	28.06.2023 6:04:21	28.06.2023 6:04:46	Тула	Склад №1	15	Транзит до 30 сут	Авто	Диспетчер склада
5	Вызов погрузчика	28.06.2023 6:05:49	28.06.2023 6:05:49	Тула	Склад №1	1499	Транзит до 30 сут	Авто	Диспетчер склада
5	Размещение груза	28.06.2023 10:22:27	28.06.2023 10:22:28	Тула	Склад №1	2416	Транзит до 30 сут	Авто	Оператор склада
5	Груз размещён	28.06.2023 10:22:36	28.06.2023 10:22:32	Тула	Склад №1	561	Транзит до 30 сут	Авто	Контролёр

Приложение 3. Таблица. Основные процессные метрики

ГРУППА	№	НАИМЕНОВАНИЕ МЕТРИКИ	Диаграмма
Стати-стика	15.	Входной поток (количество Process ID)	
	16.	Количество разновидностей путей процесса	
	17.	Разнообразие путей (%) – (*Количество разновидностей путей / Входной поток*)	
	18.	Количество уникальных событий	да
	19.	Количество процессов с зацикливанием	
	20.	Количество участников (User)	
	21.	Текучка (динамика) кадров во временном периоде	
	22.	Количество ролей	
Длитель-ность	23.	Длительность чистая (трудозатраты)	да
	24.	Длительность очереди	да
	25.	Длительность фактическая	да
	26.	Длительность фактическая рабочего времени	
	27.	Длительность на стороне клиента	
	28.	Длительность на стороне компании	
	29.	Длительность от начала процесса до ноды (Running sum)	да
Стоимость		Стоимость чистая	да
		Стоимость от начала процесса Running sum()	да
Эффек-тивность		Зацикленность % по нодам	да
		Зацикленность % по факт времени = (Потери времени на циклы)	да
		Зацикленность % по стоимости	
		Стоимость зацикливания	да
		Длительность зацикливания	
Нагрузка (парал-лельное выполне-ние)		Нагрузка на процесс	да
		Нагрузка на ноду	да
		Нагрузка на User	да
Очередь		Очередь	да
		Производительность $USER_n$ | обработанных нод во времени. (напр. 20 операций “X” в час)	да
		Занятость USER	да
Граф		% Переходов (доля входящих/исходящих по ноде)	
		Рейтинг отдалённости всех последующих нод	
		Рейтинг отдалённости всех предыдущих нод	
		Количество процессов в разрезе длины цепочки	
		Брак – это последовательности шагов, которых не должно быть в штатном течении процесса. Цепочки «брак» определяются экспертом	
		Схожесть цепочек (расстояние Ливенштейна)	

Прокопенко Наталья Юрьевна
Артюх Геннадий Сергеевич

ПРИМЕНЕНИЕ ТЕХНОЛОГИИ PROCESS MINING ДЛЯ АНАЛИЗА ДАННЫХ И ПРОЦЕССОВ
(на платформе Loginom Community)

Учебное пособие

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Нижегородский государственный архитектурно-строительный университет»
603000, Нижний Новгород, ул. Ильинская, 65.
http://www. nngasu.ru, rector@nngasu.ru