

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра математического обеспечения и применения ЭВМ

ОТЧЕТ
по научно-исследовательской работе
Тема: Модель поведения пользователей на основе коммерческих данных

Студентка гр. 6304

Вероха В.Н.

Преподаватель

Бекенева Я.А.

Санкт-Петербург

2021

ЗАДАНИЕ НА НАУЧНО-ИССЛЕДОВАТЕЛЬСКУЮ РАБОТУ

Студентка Вероха В.Н.

Группа 6304

Тема НИР: Модель поведения пользователей на основе коммерческих данных

Задание на НИР:

Провести содержательный обзор инструментов Process Mining, интегрировать данные в выбранный инструмент, а также построить модель на “сырых” данных на основе данных e-commerce сайта.

Сроки выполнения НИР: 01.09.2021-27.12.2021

Дата сдачи отчета: 27.12.2021

Дата защиты отчета: 28.12.2021

Студентка

Вероха В.Н.

Руководитель

Бекенева Я.А.

АННОТАЦИЯ

В результате развития информационных технологий возникла необходимость идти в ногу со временем во всех сферах, включая бизнес-процессы. Особенно актуален этот вопрос для e-commerce сайтов.

В данной работе представлен анализ коммерческих данных с помощью инструмента метод Process Mining.

SUMMARY

As a result of the development of information technology, it became necessary to keep up with the times in all areas, including business processes. This issue is especially relevant for e-commerce sites.

This paper presents the analysis of commercial data using the Process Mining method.

СОДЕРЖАНИЕ

АННОТАЦИЯ	3
SUMMARY	3
СОДЕРЖАНИЕ	4
ВВЕДЕНИЕ	5
ОСНОВНАЯ ТЕРМИНОЛОГИЯ	6
1. ПОСТАНОВКА ЗАДАЧИ	7
2. РЕЗУЛЬТАТЫ РАБОТЫ В ОСЕННЕМ СЕМЕСТРЕ	8
2.1. Содержательный обзор инструментов РМ	8
2.1.1. Aprimore	8
2.1.2. ProM	12
2.1.3. Disco	13
2.1.4. Результаты сравнения инструментов	15
2.2. Интеграция данных в Disco и построение модели на сырых данных	16
3. ОПИСАНИЕ ПРЕДПОЛАГАЕМОГО МЕТОДА РЕШЕНИЯ	19
4. ПЛАН РАБОТЫ НА ВЕСЕННИЙ СЕМЕСТР	20
5. ПЛАН РАБОТЫ НА ОСЕННИЙ СЕМЕСТР	21
6. ОТЗЫВ РУКОВОДИТЕЛЯ	22
ЗАКЛЮЧЕНИЕ	23
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	24
ПРИЛОЖЕНИЕ А. Скрипт подключения по API к JustControl	25
ПРИЛОЖЕНИЕ Б. Скрипт для анализа исходных данных	26

ВВЕДЕНИЕ

В результате развития информационных технологий [1] возникла необходимость идти в ногу со временем во всех сферах, включая бизнес-процессы.

Существуют такие бизнес-процессы, которые неподвластны для решения одним человеком или требуют большое количество времени и т. д. Например: отчетность за длинные периоды, закупки, графики посещения мероприятий, переводы финансов.

Существующий разрыв между анализом моделей бизнес-процессов и данных затрудняет поиск решений множества интересных и сложных задач современного мира, где значение данных давно сравнивается со значением нефти [2]. Process Mining (PM) призван ликвидировать данный разрыв, поднимая анализ бизнес-процессов на новый уровень.

Process Mining решает такие следующие проблемы:

- обнаружение реальных бизнес-процессов;
- поиск узких мест в бизнес-процессах;
- выявление отклонений в бизнес-процессах;
- поиск быстрых/ коротких путей выполнения бизнес-процессов;
- прогнозирование проблем в бизнес-процессах.

Данная работа концентрируется на внедрение интеллектуального анализа процессов в современное общество, а также на предобработке данных для модели и их влиянии на модель.

ОСНОВНАЯ ТЕРМИНОЛОГИЯ

Интеллектуальный анализ процессов, ИАП (en. Process Mining) — общее название ряда методов и подходов, предназначенных для анализа и усовершенствования процессов в информационных системах или бизнес-процессов на основании изучения системных данных о выполненных операциях.

E-commerce — это электронная покупка или продажа онлайн.

API (en. Application programming interface) — описание способов (набор классов, процедур, функций, структур или констант), которыми одна компьютерная программа может взаимодействовать с другой программой.

JustControl — это сервис, позволяющий собирать статистику с продвигаемых веб-сайтов.

BPMN (англ. Business Process Model and Notation, нотация и модель бизнес-процессов) — это язык моделирования бизнес-процессов, который является промежуточным звеном между формализацией/визуализацией и воплощением бизнес-процесса.

1. ПОСТАНОВКА ЗАДАЧИ

В настоящее время существует множество сайтов интернет-магазинов, то есть направление e-commerce, основной целью является продажа чего-либо и получение максимальной прибыли. Для этого необходимо понимать поведение пользователя на сайте.

Проблемой обнаружения зависимостей для их дальнейшей оптимизации является то, что человек не может корректно и быстро обработать большое количество данных.

Целью на данном этапе (осенний семестр) является изучение существующих инструментов для РМ, а также построение первой модели.

Задачи на осенний семестр:

- составление содержательного обзора популярных инструментов;
- интеграция данных в модель;
- построение первой модели на “сырых” данных;
- оформление отчета по проделанной работе.

2. РЕЗУЛЬТАТЫ РАБОТЫ В ОСЕННЕМ СЕМЕСТРЕ

В ходе осеннего семестра были решены следующие задачи:

- проведен содержательный обзор инструментов;
- интеграция данных в инструмент PM Disco и построение модели на сырых данных.

2.1. СОДЕРЖАТЕЛЬНЫЙ ОБЗОР ИНСТРУМЕНТОВ PM

На текущий момент инструментов для проведения интеллектуального анализа процессов не так много. Для данного обзора были выбраны наиболее популярные [4]:

- Apromore;
- ProM;
- Disco.

2.1.1. APROMORE

Apromore — это программное обеспечение с открытым исходным кодом, которое доступно с 2009 года. Основное внимание уделяется интеллектуальному анализу и улучшению процессов с помощью искусственного интеллекта. Программа бесплатна для образовательных целей. Ключевые особенности инструмента PM описаны ниже.

Общая рабочая область

Эта функция позволяет пользователю делиться моделями процессов и журналами событий в репозитории с другими людьми в бизнесе.

Отображение процессов и модели BPMN

Данный функционал позволяет пользователю создать карту процесса из модели BPMN по данным журнала событий, а также динамически переключаться между двумя представлениями. Интерфейс функционала представлен на рис. 2.1.1.1.

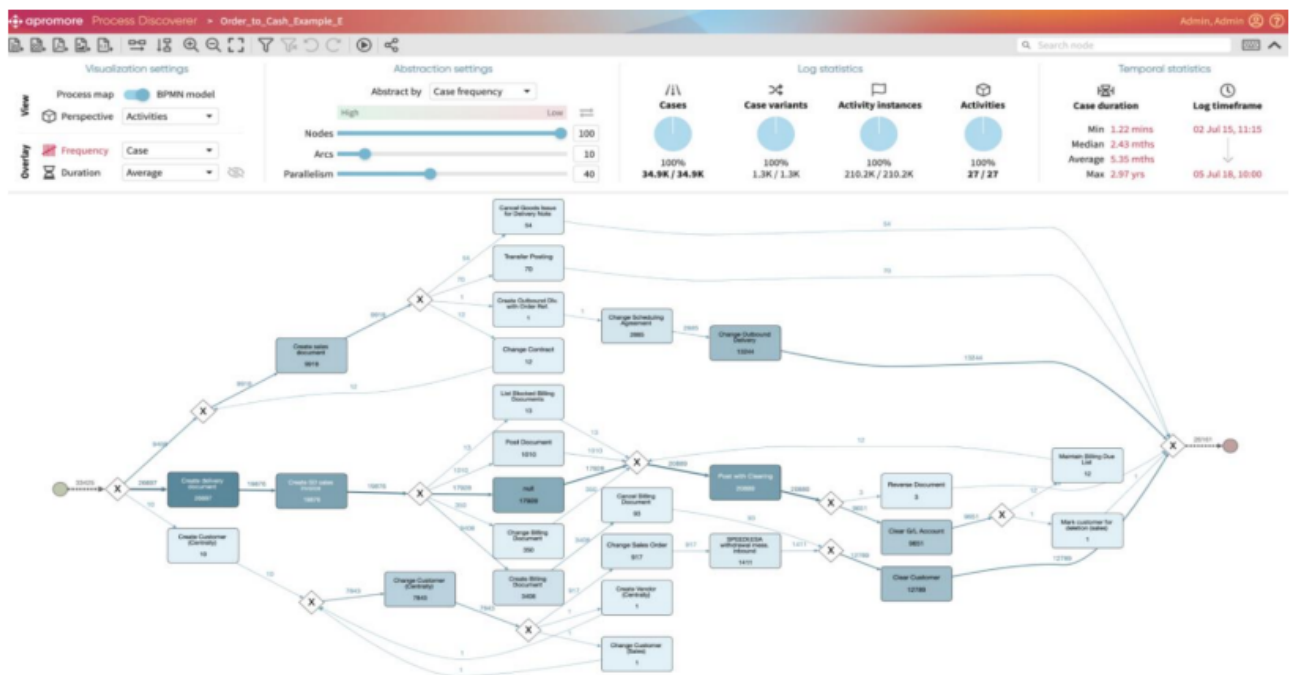


Рисунок 2.1.1.1 — Отображение процессов и модели BPMN

Наложения производительности

С помощью данного функционала можно проводить анализ различных частот и продолжительности процесса. На рис. 2.1.1.2 представлена функция наложения производительности, где время транзакции и время обработки различных действий визуально показаны для каждого узла и соединены стрелками.

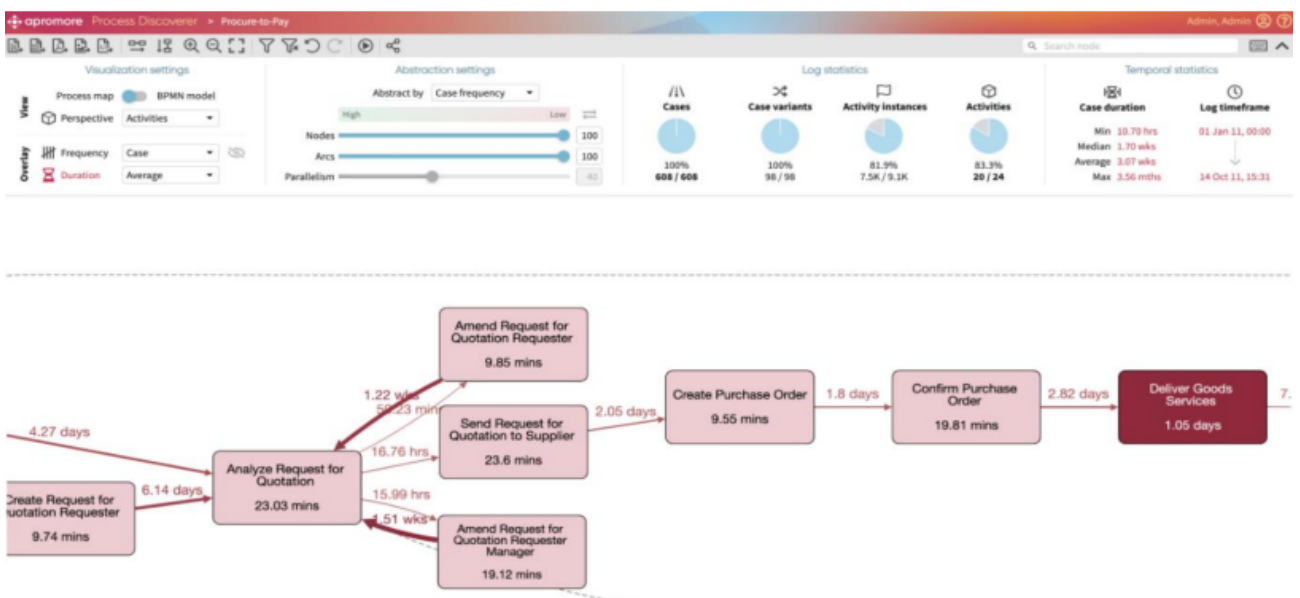


Рисунок 2.1.1.2 — Пример наложения производительности

Визуальная фильтрация

На рис. 2.1.1.3 представлена визуальная фильтрация данных журнала событий в ArgoMore. Функционал помогает сэкономить время при различных анализах с помощью простого интерфейса. Фильтры могут применяться по различным доступным показателям производительности, временным рамкам журнала процессов, вариантам событий, путям выполнения и т. д.



Рисунок 2.1.1.3 — Пример визуальной фильтрации

Визуализация данных в дашборд

Панель управления полностью настраивается в зависимости от предпочтений пользователя. Она позволяет пользователю создавать, визуализировать, проверять и настраивать информационную панель. Это может помочь в анализе производительности с более функциональной визуализацией различных параметров. На рис. 2.1.1.4 представлена панель показателей производительности.

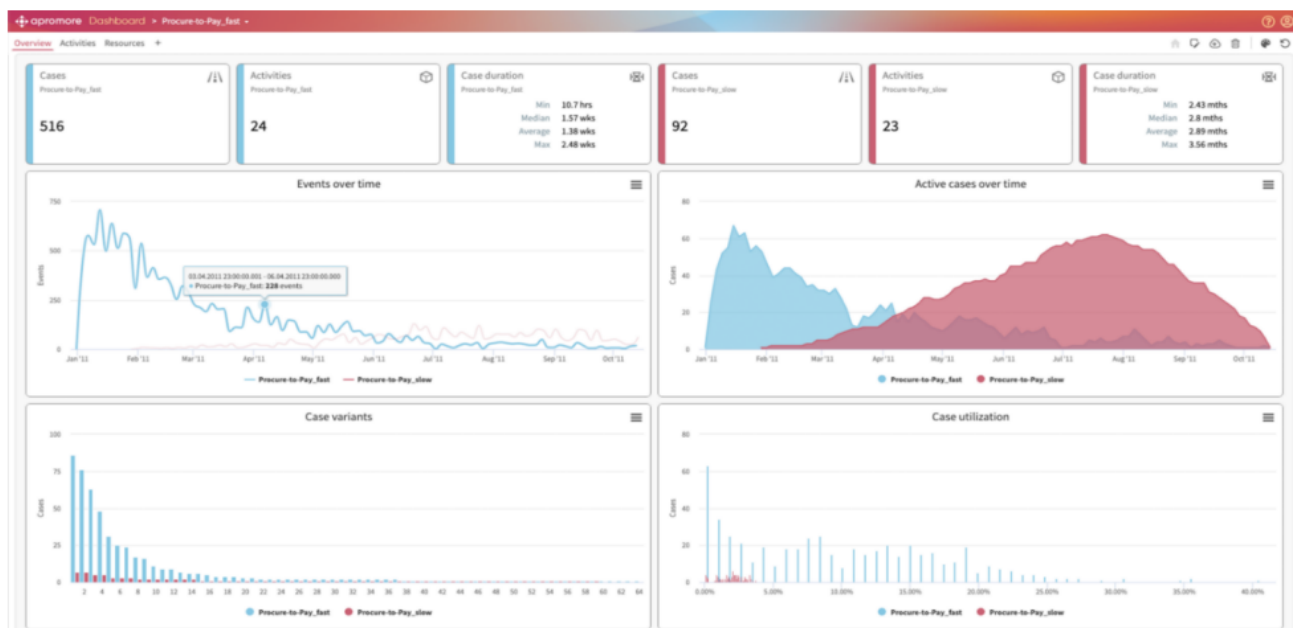


Рисунок 2.1.1.4 — Пример дашборда

Сравнение потоков и анимация журнала процессов

Различные варианты процесса можно легко сравнить друг с другом с помощью функции анимации потока процесса в реальном времени. Функцию анимации также можно использовать для выявления узких мест в процессе, которые в дальнейшем можно использовать для оптимизации. На рис. 2.1.1.5 представлена анимация журнала процесса, которая позволяет пользователю оценить рабочую нагрузку каждого ресурса в процессе.

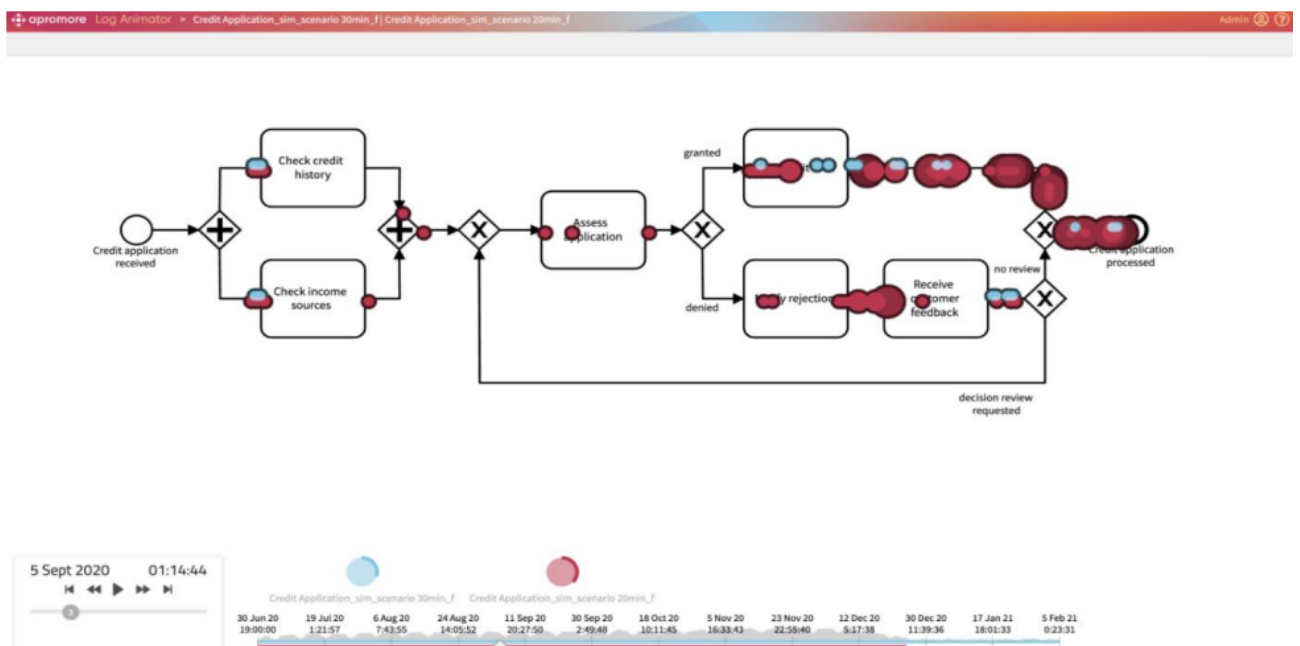


Рисунок 2.1.1.5 — Анимация журнала процесса

Соответствие процессов на основе правил и моделей

Можно установить правила для проверки фактического потока процесса против ожидаемого потока процессов. Правила выбираются на основе запланированных потоков процессов. Это позволяет пользователю идентифицировать несоответствия процесса и основные причины таких различий и оценивать соответствующие воздействия на фактический процесс.

Прогнозирование РМ

Еще одна функция Arprome позволяет пользователю обучать модели машинного обучения, которые можно использовать для прогнозирования и отслеживания различных характеристик процессов в режиме реального времени. Модели можно обучить предсказывать исход дела, время для завершения действия, определять последующее действие в процессе и т. д.

2.1.2. ProM

ProM — это инструмент РМ с доступными надстройками и открытым исходным кодом, реализованный на Java, которым можно воспользоваться без какой-либо оплаты. Различные версии ProM совместимы с такими операционными системами, как Windows, Mac и Linux.

Фреймворк ProM основан на следующих концепциях:

- Data Objects;
- Plug-ins;
- Visualizers.

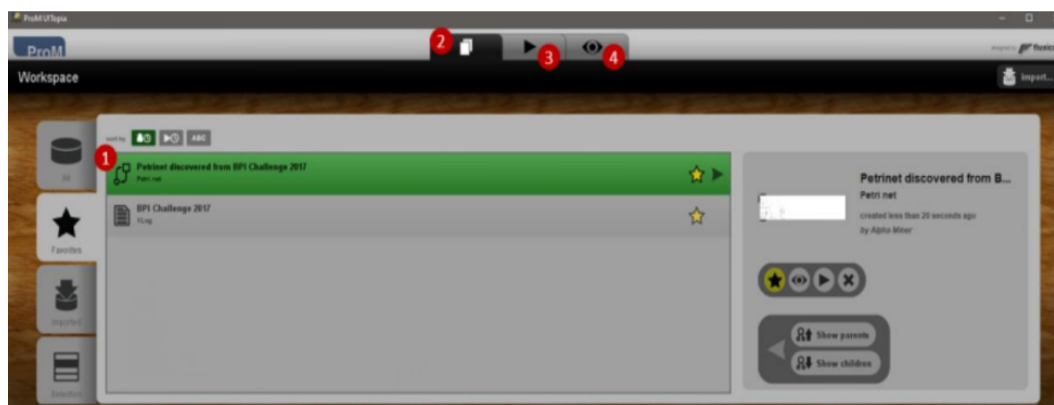


Рисунок 2.1.2.1 — Пользовательский интерфейс ProM : (1) Data Objects, (2) Plug-ins, (3) Visualizers

На рис. 2.1.2.1 представлен пользовательский интерфейс инструмента.

Data Objects

Все модели данных и процессов считаются объектами данных, а наиболее важными из них являются журналы событий и модели процессов. ProM требует обработки данных журнала событий в формате XES, если данные журнала событий в формате CSV, то их можно преобразовать внутри инструмента. ProM может создавать различные типы моделей процессов, наиболее важными из которых являются Petri Net и диаграммы BPMN. Модели Petri Net могут использоваться для поведенческого анализа, а также для проверки соответствия.

Plug-ins

Существуют следующие плагины в ProM:

- Event log filtering;
- Process Discovery;
- Conformance Checking;
- Model enhancement.

Каждый из вышеперечисленных плагинов может быть использован в зависимости от цели исследования.

Visualizers

ProM предлагает визуализаторы для большинства объектов, а наиболее часто используемые визуализаторы — журналы событий, Petri Net и диаграммы BPMN. Визуализатор фильтра событий можно использовать для разных целей, например: фильтрация, предварительная обработка и повторный анализ журналов событий.

2.1.3. Disco

Disco — это платный инструмент для анализа и оптимизации фактических процессов. Данный инструмент разработан с использованием алгоритмов быстрого анализа процессов Fluxicon BV. Disco доступно для операционных систем Mac 10.10 или новее, Windows 7 SP1 (x64) или новее. Программа поддерживает такие форматы файлов, как CSV, MS Excel (XLS и XLSX), MXML и MXML.GZ (ProM 5), XES и XES.GZ (ProM 6), файлы журнала

FXL Disco и файлы проекта DSC Disco. Ниже перечислены ключевые особенности рассматриваемого инструмента.

Автоматическое обнаружение процессов

Disco может создавать модели процессов непосредственно из необработанных данных журнала событий с минимальным вмешательством пользователя. Программное обеспечение может генерировать до шести визуализаций метрик процесса. Шесть показателей процесса включают абсолютную частоту, частоту наблюдений, максимальное количество повторений, общую, среднюю и максимальную продолжительность.

Анимация карты процесса

Функция анимации помогает визуализировать процесс по модели, созданной из импортированных данные журнала событий; а также определить узкие места в модели. Пример визуализации представлен на рис. 2.1.3.1.

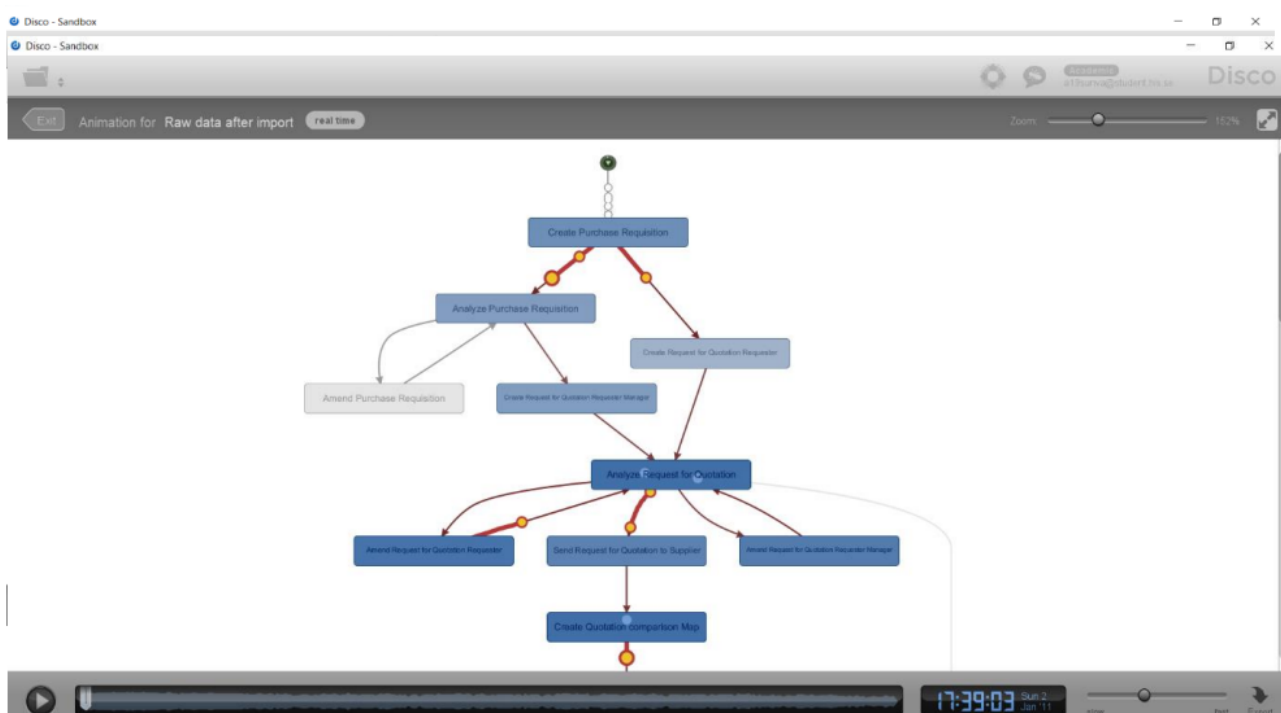


Рисунок 2.1.3.1 — Визуализация модели

Подробная статистика

Статистика Disco помогает определить действия, на выполнение которых требуется больше времени, их частотность и влияние. Эти статистические данные могут быть полезны для повышения производительности всего

процесса вместо того, чтобы сосредоточиться на улучшении одного вида деятельности. На рис. 2.1.3.2 показана панель статистики.

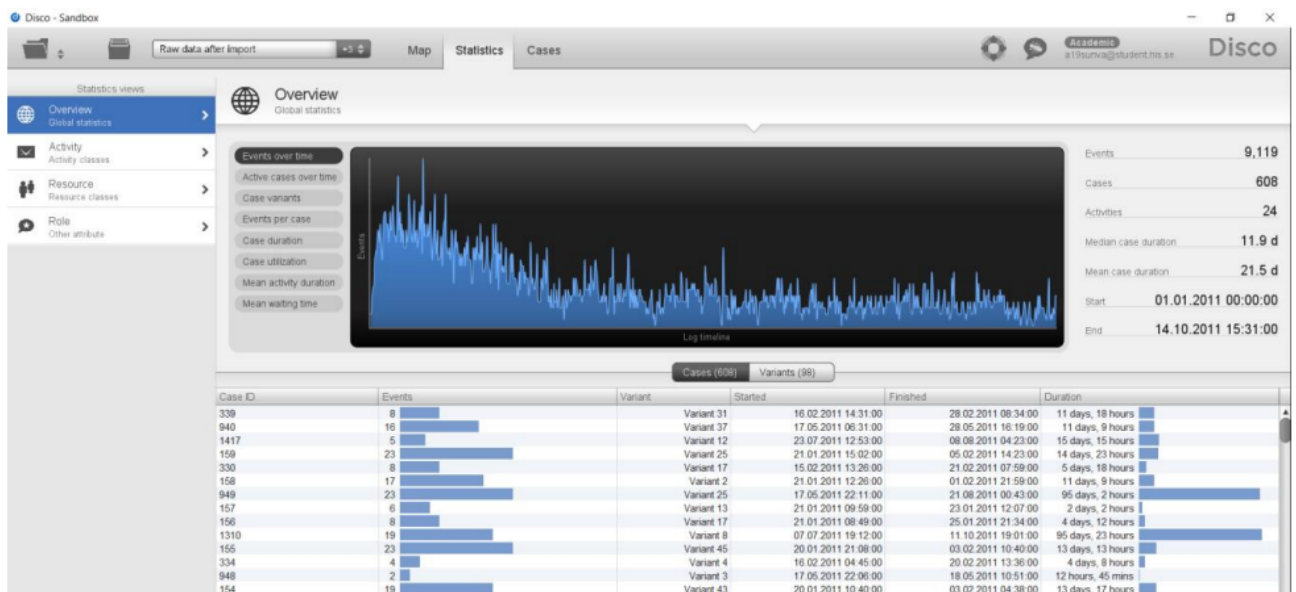


Рисунок 2.1.3.2 — Панель статистики

2.1.4. РЕЗУЛЬТАТЫ СРАВНЕНИЯ ИНСТРУМЕНТОВ

Сравнение инструмента выполнено по следующим признакам:

- поддержка моделирования — способность инструмента интеллектуального анализа процессов создавать валидную модель процессов из данных журнала событий;
- удобство — легкость и удобство для пользователя при создании карты процессов, анализа проведения и производительности, визуализации и использования различных статистических параметров и т. д;
- гибкость — возможность загрузки пользователем журнала событий в более распространенных форматах файлов, обрабатывать данные журнала событий, просматривать статистику, возможностью без каких-либо трудностей перемещаться между другими вкладками и проектами, а также выполнять изменения для получения различных представлений карт процесса;
- стабильность — способность инструментов не давать сбоев при продолжительном использовании, а также без сбоев читать и обрабатывать все записи данных;

- поддержка анимации — способность имитировать потоки продукта, соответствующие записанному времени в данных журнала событий на карте процесса, который обнаруживается с использованием тех же данных журнала событий;
- поддержка моделирования BPMN — возможность позволить компаниям понять вовлеченность в продукт с помощью графической нотации.

Результаты сравнения приведены в табл. 2.1.4.1.

Таблица 2.1.4.1 — Результаты сравнения

Сравнительный признак	Apromore	ProM	Disco
Поддержка моделирования	Высокий уровень	Средний уровень	Высокий уровень
Удобство	Высокий уровень	Средний уровень	Высокий уровень
Гибкость	Высокий уровень	Низкий уровень	Средний уровень
Стабильность	Высокий уровень	Низкий уровень	Высокий уровень
Поддержка анимации	Средний уровень	Низкий уровень	Средний уровень
Поддержка моделирования BPMN	Высокий уровень	Средний уровень	Нет такого функционала

Исходя из проделанного сравнительного анализа инструментов, наиболее подходящим по функционалу является Apromore.

2.2. Интеграция данных в Disco и построение модели на сырых данных

Так как инструменты Apromore и Disco [5] являются платными, у их представителей были запрошены академические лицензии. Представители Disco дали академическую лицензию, а лицензия Apromore находится на этапе согласования. Поэтому на данном этапе работ используется инструмент Disco.

Интеграция данных в модель представлена на рис. 2.2.1. В качестве данных case был выбран столбец “Client ID (Cust. Dim. 2 - S)”, activity —

“Категория событий”, timestamp — “Дата” (данный столбец также преобразован в корректный формат).

Disco - New project

Academic veronavi111@mail.ru Disco

Дата
column is used

Timestamp

Pattern... matches all rows.

	Client ID (Cust. Dim. 2 - S)	Категория событий	Дата	Пользователи
1	28457714.1617992843	Product List	20210409	2
2	1000002639.1619770024	flocktory-precheckout	20210430	1
3	1000002639.1619770024	Internal Marketing	20210430	1
4	1000002639.1619770024	Product List	20210430	1
5	1000013384.1617961179	Account	20210409	1
6	1000013384.1617961179	Cart	20210409	1
7	1000013384.1617961179	Checkout	20210409	1
8	1000013384.1617961179	flocktory-precheckout	20210409	1
9	1000013384.1617961179	Internal Marketing	20210409	1
10	1000013384.1617961179	Product List	20210409	1
11	1000037713.1619360436	Internal Marketing	20210425	1
12	1000037713.1619360436	Product List	20210425	1
13	1000042093.1618663768	flocktory-precheckout	20210417	1
14	1000042093.1618663768	Internal Marketing	20210417	1
15	1000042093.1618663768	Product List	20210417	1
16	1000050879.1619126149	flocktory-precheckout	20210423	1
17	1000050879.1619126149	Internal Marketing	20210423	1
18	1000050879.1619126149	Product List	20210423	1
19	1000062086.1618653704	Cart	20210417	1
20	1000062086.1618653704	Checkout	20210417	1

Рисунок 2.2.1 — Интеграция данных

Построенная модель на “сырых” данных без фильтров представлена на рис. 2.2.2.

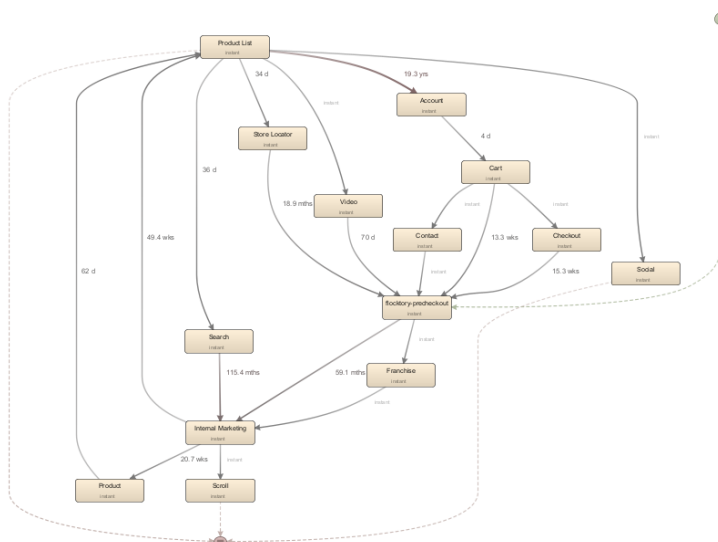


Рисунок 2.2.2 — Модель данных без фильтров

Также построена модель с фильтром по исключению следующих событий: Internal Marketing и flocktory-precheckout — так как они являются внутренними для маркетологов и разработчиков сайта соответственно. Результаты представлены на рис. 2.2.3.

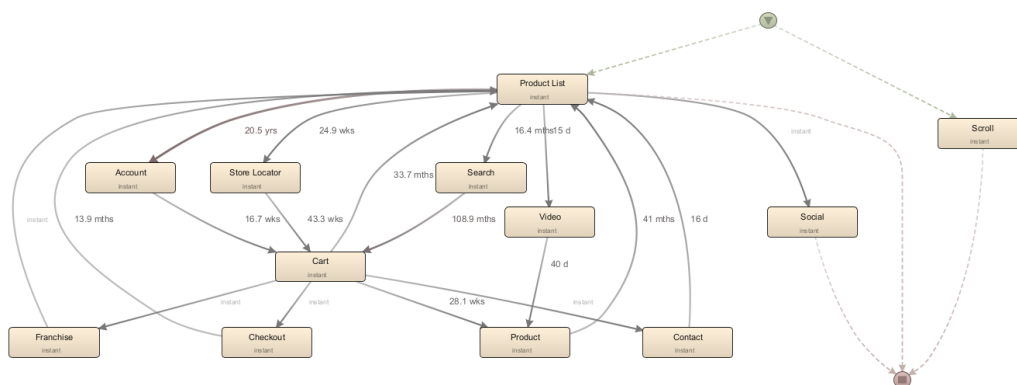


Рисунок 2.2.3 — Модель данных с фильтром исключения *Internal Marketing* и *flocktory-precheckout*

Запуск анимации построенной модели поведения пользователей на “сырых” данных представлен на рис. 2.2.4.

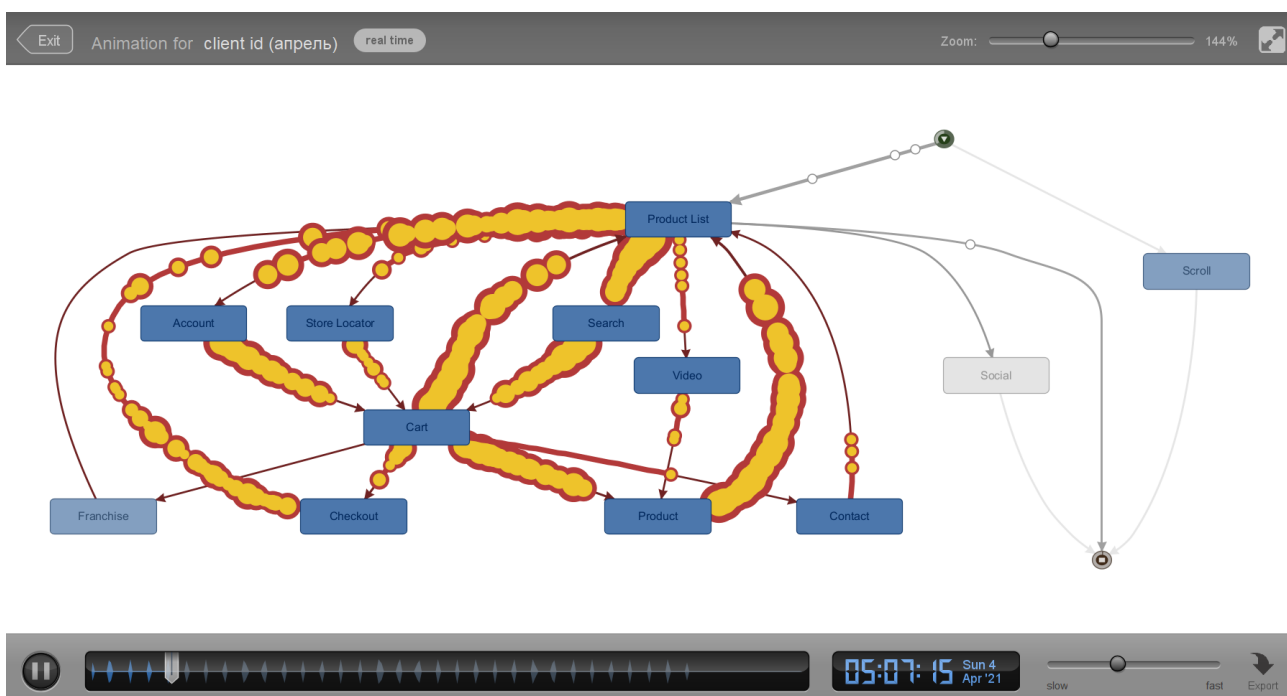


Рисунок 2.2.4 — Анимация построенной модели

Таким образом видно самую популярную цепочку действий: пользователь заходит на сайт, далее посещает страницу каталога, заходит в свой аккаунт, добавляет товар в корзину и оформляет заказ.

3. ОПИСАНИЕ ПРЕДПОЛАГАЕМОГО МЕТОДА РЕШЕНИЯ

В ходе осеннего семестра была построена примитивная модель. На осенний семестр предполагается более глубокое изучение данных с помощью инструмента РМ.

Предполагается изучить все пути перемещения пользователей по сайту. Для этого необходимо будет составить список гипотез по уже построенной модели поведения пользователей. Далее опровергнуть или подтвердить гипотезы и на их основе построить модель, наиболее корректную и “продающую”. В соответствии с этой моделью клиент сможет оптимизировать сайт для более эффективного и удобного пользования.

Также планируется взять данные за более длительный период времени, чтобы улучшить точность будущей модели, с академической лицензией это будет возможно, так как без нее были ограничения по количеству данных.

4. ПЛАН РАБОТЫ НА ВЕСЕННИЙ СЕМЕСТР

План работы на осенний семестр представлен в табл. 4.1.

Таблица 4.1 — План работы

Период	Задача
10.02.2022-05.03.2022	Составление гипотез
06.03.2022-16.03.2022	Выгрузка данных за более длительный промежуток времени
17.03.2022-16.04.2022	Опровержение или принятие гипотез
17.04.2022-10.05.2022	Подготовка модели на основе проведенного анализа
11.05.2022-31.05.2022	Подведение итогов и подготовка отчета ВКР

5. ПЛАН РАБОТЫ НА ОСЕННИЙ СЕМЕСТР

План работы на осенний семестр представлен в табл. 5.1.

Таблица 5.1 — План работы

Период	Задача
01.09.2021-30.09.2021	Знакомство с разными средами создания моделей process mining
01.10.2021-20.10.2021	Интеграция данных в модель
21.10.2021-20.11.2021	Построение модели process mining
21.11.2021-15.12.2021	Подведение итогов и подготовка отчета

6. ОТЗЫВ РУКОВОДИТЕЛЯ

Ниже представлен (рис. 6.1) скан отзыва научного руководителя — Бекеновой Яны Андреевны.

ОТЗЫВ

на научно-исследовательскую работу

студентки гр. 6304

Вероха В. Н.

В своей научно-исследовательской работе студентка Вероха В.Н. выполнила все поставленные перед ней задачи. В рамках проделанной работы был получен и предварительно обработан исходный набор данных, произведена его выгрузка, проведены эксперименты по анализу данных.

Полученные результаты описаны и проанализированы, на их основании сделаны выводы о дальнейшем плане работы.

Считаю, что студентка Вероха В.Н. заслуживает оценки «отлично».

Руководитель к.т.н. ассистент



Бекенева Я.А.

Рисунок 6.1 — Отзыв руководителя

ЗАКЛЮЧЕНИЕ

В результате весеннего семестра были выполнены следующие поставленные задачи:

- составление содержательного обзора популярных инструментов;
- интеграция данных в модель;
- построение первой модели на “сырых” данных.

После содержательного обзора популярных инструментов Process Mining: Apromore, ProM и Disco — было решено использовать инструмент Apromore. Однако из-за того, что академическая лицензия еще не предоставлен, построение модели на “сырых” данных проводилось с помощью инструмента Disco.

Сырые данные были интегрированы в модель, после чего она была построена.

Для более качественного построения были исключены такие события, как Internal Marketing и flocktory-precheckout, так как они являются внутренними для маркетологов и разработчиков сайта соответственно. В результате построения модели была выявлена самая популярная цепочка действий пользователя на сайте: пользователь заходит на сайт, далее посещает страницу каталога, заходит в свой аккаунт, добавляет товар в корзину и оформляет заказ.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. Knowledge management and data mining for marketing // Decision Support Systems. 2001. Vol 31(1), 127–137.
2. Ana blogs // Medium. URL: https://ana.blogs.com/maestros/2006/11/data_is_the_new.html (дата обращения: 10.03.2021).
3. JustControl // Medium. URL: https://docs.justcontrol.it/public_docs/docs/intro/about.html (дата обращения: 20.03.2021).
4. Amos H.C. // A COMPARATIVE STUDY OF PROCESS MINING SOFTWARE FOR SUPPORTING BOTTLENECK ANALYSIS OF PRODUCTION SYSTEMS. 2021.
5. Process Mining Book // <https://fluxicon.com/book/read/> (дата обращения: 13.12.2021).

ПРИЛОЖЕНИЕ А. Скрипт подключения по API к JustControl

```
import requests
import json

api_token = '****'

# Общий запрос для того, чтобы узнать метрики
url='https://****/v1/modules/5/stats/metrics'

headers = {
    "Content-Type": "text/csv",
    "Authorization": api_token
}

r = requests.get(url, headers=headers)
r.text

# Гугл-выгрузка
url='https://****/v1/modules/5/stats/csv'
params = {
    "dateFrom": "2021-04-01",
    "dateTo": "2021-04-30",
    "metrics": json.dumps(["user"]),
    "breakdown[]": ["date","client_id","category"]
}

headers = {
    "Content-Type": "text/csv",
    "Authorization": api_token
}

r = requests.get(url=url, params=params, headers=headers)

it = iter(list(r.text.replace('\n','').split(',')))
data_list=list(map(list, zip(it, it, it)))
import pandas as pd
data_frame=pd.DataFrame(data_list)

data_frame.to_csv('JC_google_id5_api.csv', index=False, encoding='utf-8', header=False)
```

ПРИЛОЖЕНИЕ Б. Скрипт для анализа исходных данных

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
import datetime as dt

data=pd.read_csv('C:/Users/user/Desktop/НИР/client id (апель).csv', dtype={'Дата':str, 'Client ID (Cust. Dim. 2 - S)':str})

# исходные данные
data.head()

# преобразование столбца с датами в формат даты
data["Дата"] = pd.to_datetime(data["Дата"]).dt.strftime('%Y-%m-%d')
data.head()

# вывод уникальных значений категорий событий
data["Категория событий"].unique()

# количество уникальных совпадений в data["Категория событий"]
data["Категория событий"].value_counts()

data_category=pd.DataFrame(data["Категория событий"].value_counts().reset_index().values,
columns=["category", "count"])
print(data_category)
plt.bar(data_category["category"], data_category["count"])
plt.xticks(rotation=45, ha='right')
plt.show()

# количество уникальных совпадений в data["Client ID (Cust. Dim. 2 - S)"]
data["Client ID (Cust. Dim. 2 - S)"].nunique()

# количество уникальных совпадений в data["Client ID (Cust. Dim. 2 - S)"]
```

```

# то есть данные пользователи совершали следующее количество событий
data["Client ID (Cust. Dim. 2 - S)"].value_counts()

data_client=pd.DataFrame(data["Client ID (Cust. Dim. 2 - S)"].value_counts().reset_index().values,
columns=["client", "count"])

# можно удалить тех клиентов, у кого количество совершенных событий менее 3, т к они слишком
мало совершенный действий (это 81015 Client ID)
print(data_client[data_client["count"]<3])

data['Дата']=data['Дата'].astype('datetime64[ns]')
day=pd.Grouper(key="Дата", freq="D")

# распределение пользователей по дням
data_day=data.groupby([day]).sum().reset_index()

# распределение пользователей по дням
plt.plot(data_day["Дата"], data_day["Пользователи"], label = 'Пользователи')
plt.xticks(rotation=45, ha='right')
plt.legend( loc=2)
plt.show()

data_day.sort_values(by='Пользователи', ascending = False)

```