**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ**

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ**

**«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет компьютерных наук

Образовательная программа бакалавриата «Прикладная математика и информатика»

|  |  |  |
| --- | --- | --- |
| СОГЛАСОВАНО  Руководитель проекта,  Место работы, должность    \_\_\_\_\_\_\_\_\_\_\_ / Инициалы, фамилия /  «27»  декабря  2018 г. |  | УТВЕРЖДАЮ  Академический руководитель образовательной программы «Прикладная математика и информатика»  доцент, канд. физ.-мат. наук    \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А.С. Конушин  «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_ 201\_ г. |

**Автоматизированная идентификация аномалий методами Process Mining**

**Техническое задание**

**ЛИСТ УТВЕРЖДЕНИЯ**

**RU.17701729.04.01-01 ТЗ 01-1-ЛУ**

                   Исполнители:

студент группы БПМИ 163

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ /Агузумцян С.А. /

студент группы БПМИ 164

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ /Шафиров И.Л. /

студент группы БПМИ 162

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ / Цыбакин А.Ю./

«\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 201\_ г.

**Москва 2018**

**Содержание**

[1. Введение 4](#_Toc385162099)

[**1.1. Наименование программы 4**](#_Toc385162100)

[**1.2. Общее описание проекта 4**](#_Toc385162101)

[2. Основания для разработки 5](#_Toc385162102)

[**2.1. Документы, на основании которых ведется разработка 5**](#_Toc385162103)

[**2.2. Наименование темы разработки 5**](#_Toc385162104)

[3. Назначение разработки 6](#_Toc385162105)

[**3.1. Функциональное назначение 6**](#_Toc385162106)

[**3.2. Эксплуатационное назначение 6**](#_Toc385162107)

[4. Требования к программе 7](#_Toc385162108)

[**4.1. Требования к функциональным характеристикам 7**](#_Toc385162109)

[**4.1.1. Требования к организации входных данных 7**](#_Toc385162112)

[**4.1.2. Требования к организации выходных данных 7**](#_Toc385162113)

[**4.2. Требования к интерфейсу 7**](#_Toc385162115)

[**4.3. Требования к надежности 8**](#_Toc385162116)

[**4.3.1. Требования к обеспечению надежного (устойчивого) функционирования программы 5**](#_Toc385162117)

[**4.3.2. Время восстановления после отказа 8**](#_Toc385162118)

[**4.3.3. Отказы из-за некорректных действий оператора 8**](#_Toc385162119)

[**4.4. Условия эксплуатации 8**](#_Toc385162120)

[**4.4.1. Климатические условия эксплуатации 8**](#_Toc385162121)

[**4.4.2. Требования к видам обслуживания 8**](#_Toc385162122)

[**4.4.3. Требования к численности и квалификации персонала 8**](#_Toc385162123)

[**4.5. Требования к составу и параметрам технических средств 8**](#_Toc385162124)

[**4.6. Требования к информационной и программной совместимости 8**](#_Toc385162125)

[**4.6.1. Требования к исходным кодам и языкам программирования 8**](#_Toc385162128)

[**4.7. Требования к маркировке и упаковке 8**](#_Toc385162130)

[**4.8. Требования к транспортировке и хранению 8**](#_Toc385162131)

[5. Требования к программной документации 9](#_Toc385162135)

[**5.1. Предварительный состав программной документации 9**](#_Toc385162136)

[**5.2. Специальные требования к программной документации 9**](#_Toc385162137)

[6. Технико-экономические показатели 10](#_Toc385162138)

[**6.1. Ориентировочная экономическая эффективность 10**](#_Toc385162139)

[**6.2. Предполагаемая потребность 10**](#_Toc385162140)

[7. Стадии и этапы разработки 11](#_Toc385162142)

[8. Порядок контроля и приемки 14](#_Toc385162143)

[**8.1. Виды испытаний 14**](#_Toc385162144)

[**8.2. Общие требования к приемке работы 14**](#_Toc385162145)

[Приложение 1. Список терминов 15](#_Toc385162146)

[Приложение 2. Распределение обязанностей на проекте 16](#_Toc385162148)

**1.Введение**

**1.1 Наименование программы**

Наименование программы - «Автоматизированная идентификация аномалий методами Process Mining».

**1.2 Общее описание проекта**

Process Mining позволяет находить отклонения в бизнес-процессах, используя анализ данных, полученных из информационных систем, а также позволяет лучше понять схему, по которой работает компания. Благодаря этой информации представляется возможным улучшение работы компании на основе полученных данных (оптимизация, поиск более быстрых путей выполнения бизнес-процессов и т. д.), однако, полученных данных бывает недостаточно, чтобы выявить более глубокие и неочевидные проблемы (аномалии) в работе компании. Для нахождения таких проблем будут использоваться методы machine learning. Также необходимо вычислять нужные метрики качества и визуализировать полученные данные.

Цель данного проекта создать инструмент, который сможет находить аномалии в бизнес-процессах, кроме того, данный инструмент должен будет вычислять нужные метрики качества (KPI), а также визуализировать полученные данные. Также инструмент должен будет определять нужную степень визуализации графа бизнес-процессов. Для выявления аномалий и подсчёта метрик качества (KPI) будут анализироваться как данные из информационных систем компании, так и граф бизнес-процессов полученный путём Process Mining, кроме того, будут использованы методы machine learning для более детального анализа. Результатом работы инструмента будет являться подробный отчёт о всех аномалиях, выявленных в процессе работы программы, а также информация о метриках качества и визуализация этих данных. Полученные данные могут быть весьма полезны в целях улучшения работы компании.

**2. Основания для разработки**

**2.1 Документы на основании которых ведётся разработка**

*Приказ декана факультета компьютерных наук И.В. Аржанцева от 12.12.2018 № 2.3-02/1212-02 "Об утверждении тем, руководителей и консультантов курсовых работ студентовобразовательной программы Прикладная математика и информатика факультета компьютерных наук"*

**2.2 Наименование темы разработки**

«Автоматизированная идентификация аномалий методами Process Mining».

**3. Назначение разработки**

**3.1 Функциональное назначение**

Программа генерирует отчет об аномалиях и метриках качества используя данные из информационных систем и графа бизнес-процессов полученный путём Process Mining. Для составления отчета будут использованы методы машинного обучения и data science. На основании данного отчета в дальнейшем будет возможно исправление недостатков в анализируемых бизнес-процессах. Отчёт программы будет содержать визуализацию полученных данных, KPI которые были вычислены, а также информацию о найденных аномалия.

**3.2 Эксплуатационное назначение**

Программа используется для работы с данным компании и позволяет облегчить обработку данных об аномалиях и метриках качества в бизнес-процессах, визуализация данных также будет способствовать более глубокому пониманию бизнес-процессов. Компания, использующая программу, может использовать сгенерированный отчет для улучшения эффективности работы.

**4. Требования к программе**

**4.1 требования к функциональным характеристикам.**

На входе такой инструмент получает набор табличных данных в виде журнала событий из информационных систем. Впоследствии инструмент формирует отчёт по профилированию процесса, который отражает информацию о наиболее значимых KPI (ключевых показателей эффективности) процесса, его узких местах и аномалиях.

Всего в программе должно быть три основных модуля для корректной работы, а именно:

* 1. Модуль создания графа бизнес-процессов и её визуализации.
  2. Модуль для подбора и расчёта нужных KPI.
  3. Модуль обнаружения аномалий.
     1. **Требования к организации входных данных**

На вход программе подаётся журнал событий (EventLog), который формируется на основе данных из информационных систем

* + 1. **Требования к организации выходных данных**

Программа должна выводить информацию о найденных аномалиях и вычисленных KPI, затем весь этот процесс визуализируется.

**Требования к интерфейсу**

Программа должна иметь удобный Windows-интерфейс. А также программа должна отрисовывать выводимую информацию

* 1. **требования к надежности**

**4.2.1 Требования к обеспечению надежного (устойчивого) функционирования программы**

Для устойчивой работы программы необходимо соблюдать ряд организационно-технических мер:

1. обеспечить бесперебойное питание технических устройств;
2. обеспечить высокую защиту технических устройств для работы программы от воздействия шпионских программ, троянских программ, программ-шуток и других видов вредоносного программного обеспечения;
3. обеспечить регулярную проверку оборудования и программного обеспечения на наличие сбоев и неполадок;
4. обеспечить использование лицензионного программного обеспечения.

* + 1. **Время** **восстановления после** **отказа**

Если отказ был вызван какими-либо внешними факторами, например, сбоем электропитания, и при этом не произошел непоправимый сбой операционной системы, то время восстановления не должно превышать времени, требующегося на перезагрузку операционной системы и запуск программы.

Если отказ был вызван неисправностью технических средств или непоправимым сбоем операционной системы, то время восстановления не должно превышать времени, необходимого для устранения неисправностей технических и программных средств.

* + 1. **Отказы из-за некорректных действий оператора**

Отказ программы возможен также вследствие некорректных действий пользователя при пользовании операционной системой. Для предотвращения случаев отказа программы по причине сбоев при пользовании операционной системой следует провести предварительный инструктаж конечного пользователя и обеспечить работу конечного пользователя без предоставления ему прав администратора.

* 1. **Условия эксплуатации**

**4.3.1 Климатические условия эксплуатации**

Климатические условия эксплуатации, при которых должны обеспечиваться заданные характеристики, должны удовлетворять требованиям, предъявляемым к персональным компьютерам и компакт-дискам (CD) в части условий их эксплуатации.

* + 1. **Требования к видам обслуживания**

На персональном компьютере, где производится эксплуатация программы необходимо обеспечить регулярные проверки оборудования и программного обеспечения на наличие сбоев и неполадок. Обеспечить защиту персонального компьютера от воздействия шпионских программ, программ-шуток, троянских программ и других видов вирусов.

**4.3.3 Требования к численности и квалификации персонала**

Минимальное количество персонала, требуемого для работы программы, должно составлять не менее 1 штатной единицы:

1. конечный пользователь – оператор ЭВМ.

Оператор ЭВМ должен:

1. иметь образование не ниже среднего (полного) общего;
2. обладать практическими навыками работы с пользовательским интерфейсом операционной системы.

**4.5 Требования к составу и параметрам технических средств**

Для бесперебойной работы программы требуется персональный компьютер, оснащенный 64-разрядным (x64) процессором и как минимум 4 Гб оперативной памяти

**4.6 Требования к информационной и программной совместимости**

**4.6.1** **Требования к** **исходным** **кодам и** **языкам** **программирования**

Исходные коды программы должны быть написаны на языке python.

**4.7 Требования к маркировке и упаковке.**

Требования к маркировке и упаковке не предъявляются.

**4.8 Требования к транспортированию и хранению.**

Специальные требования к транспортировке и хранению не предъявляются.

**5. Требования к программной документации**

**5.1 Предварительный состав документации и требования.**

1. «Автоматизированная идентификация аномалий методами Process Mining» Техническое задание (ГОСТ 19.201-78);
2. «Автоматизированная идентификация аномалий методами Process Mining». Программа и методика испытаний (ГОСТ 19.301-78);
3. «Автоматизированная идентификация аномалий методами Process Mining». Текст программы (ГОСТ 19.401-78);
4. «Автоматизированная идентификация аномалий методами Process Mining». Пояснительная записка (ГОСТ 19.404-79);
5. «Автоматизированная идентификация аномалий методами Process Mining». Руководство оператора (ГОСТ 19.505-79);
6. «Автоматизированная идентификация аномалий методами Process Mining». Руководство программиста (ГОСТ 19.504-79).

**5.2 Специальные требования к документации.**

1. Все документы к программе должны быть выполнены в соответствии с ГОСТ 19.106-78 [6] и ГОСТ к этому виду документа (см. п. 5.1.).
2. Пояснительная записка должна быть загружена в систему Антиплагиат через ЛМС НИУ ВШЭ. Лист, подтверждающий загрузку пояснительной записки, сдается в учебный офис вместе со всеми материалами не позже, чем за день до защиты курсовой работы.
3. Вся документация сдается в печатном виде, при этом она должна быть обязательно подписана академическим руководителем образовательной программы «Прикладная информатика и математика», руководителем разработки и исполнителем перед сдачей курсовой работы в учебный офис не позже одного дня до защиты.
4. Вся документация и программа также сдается в электронном виде в формате .pdf или .docx. в архиве формата .rar или .zip.
5. Все документы перед защитой курсовой работы должны быть загружены в информационно-образовательную среду НИУ ВШЭ LMS (Learning management system) в личном кабинете, дисциплина - «Курсовая работа», одним архивом (см. п.3)

**6. Технико-экономические показатели**

**6.1 Ориентировочная экономическая эффективность.**

В настоящее время в связи с автоматизацией бизнес-процессов, ведения электронного учета в информационных системах происходит накопление огромного объема данных. Разрабатываемый инструмент предназначается для:

1. Автоматизированного анализа бизнес-процесса.
2. Автоматизированной идентификации аномалий в бизнес-процессах.

Инструмент при работе в этих направлениях основывается на работе с методами машинного обучения и методов Process Mining. Одной из ключевых возможностей инструмента является обнаружения часто встречающихся отклонений от стандартного процесса и выявление потенциальных причин его появления, что позволит оптимизировать сам бизнес-процесс и уйти от лишних издержек. Также немаловажным приложением инструмента будет обнаружение нетипичных цепочек событий, что может потенциально снизить риск мошенничества.

**6.2 Предполагаемая потребность.**

Предполагаемая потребность обуславливается тем фактом, что на данный момент не существует инструмента, позволяющего помимо обнаружения аномалий в бизнес-процессах выявлять причины их появления, который существенно поможет бизнес-аналитикам в исследовании цепочек событий.

**7. Стадии и этапы разработки**

**7.1** **Основные модули программы.**

Программа будет иметь 3 основных модуля:

1. Модуль создания графа бизнес-процессов и её визуализации. Этот модуль программы также включает в себя алгоритм, который должен будет определить необходимый процент точности визуализации

2. Модуль для подбора нужных KPI метрик. После того, как нужные метрики будут выбраны их надо будет вычислить с помощью подпрограмм. Всего есть два типа метрик

1. Общие метрики. Основные метрики качества, которые будут отражать эффективность деятельности предприятия (например, в качестве основной метрики может быть взято время выполнения того или иного Process Path в графе бизнес-процессов)

2. P2P (procure to pay) метрики или специальные метрики они специфичны для модели покупок (Procurement).

3. Модуль обнаружения аномалий. Будет включать в себя подпрограммы, которые в соответствии с выбранными критериями будут искать изъяны в бизнес-процессах предприятия. Будем выделять три критерия для выявления аномалий.

1. Длительность выполнения того или иного процесса. Если время выполнения бизнес-процесса превышает средние показатели значит это будет определено как аномалия.

2. Неэффективные процессы, в которых присутствуют циклы или присутствуют слишком часто повторяющиеся циклы также будут определены как аномалии.

3. SOD (segregation of duties). Этот критерий подразумевает под собой разделение обязанностей внутри предприятия., соответственно если разделение обязанностей не соблюдается тогда эти случаи будут идентифицированы как аномалия. Также эта часть модуля должна выполнять функцию prediction, которая будет предсказывать вероятные повторения такого рода аномалий на предприятии.

**7.2 Анализ требований к проекту.**

Цель данного проекта — это разработка программы для выявления аномалий в данных полученных путём Process Mining. Аномалии, найденные в процессе анализа, должны быть исправлены для оптимизации работоспособности фирмы. Также должны быть подобраны нужные KPI которые впоследствии будут вычисляться, кроме того, надо сформулировать требования к стилю визуализации бизнес-процессов.

**7.3 Проектирование.**

Программа должна быть реализована на языке python. Во время проектирования также должен быть проведён анализ задачи и моделирование. Далее должен быть выбран алгоритм наиболее подходящий под требования задачи. Также можно выбрать один из существующих алгоритмов и усовершенствовать под решаемые задачи. После этого требуется сделать проектирование общей структуры инструмента на этом этапе должна быть определена структура функциональных модулей программы, а также модель спецификаций и уровней разработки.

**7.4 Реализация.**

В процессе реализации должен быть написан чёткий и понятный код на языке python. Кроме этого, требуется создать основные функциональные модули программы. Если в процессе тестирования были обнаружены баги, ошибки в алгоритме или же изъяны, тогда код программы должен быть переписан и повторно отправлен на тестирование.

**7.5 Тестирование продукта.**

В процессе тестирования программа должна быть проверена на соответствие требованиям. Кроме того, должна быть проделана проверка корректности работы алгоритма и проверка на наличие багов. После этого программа должна пройти тесты на производительность и точность работы выдав корректный вывод на тестовых данных. Результатом тестирования является устранение всех недостатков системы и заключение о ее качестве.

**7.6 Сроки и даты.**

**Университетские сроки сдачи.**

**Контрольная точка 1:**

До 28.12.2018 нужно

Сдать проект Технического задания, согласованный с ментором (руководителем проекта)

Сдать в учебный офис:

Титульный лист Технического задания с подписями студента и руководителя с 10-балльной оценкой руководителя

Загрузить в ЛМС:

Техническое задание

**Контрольная точка 2:**

До 15.04.2019 нужно

Сдать проект Пояснительной записки, согласованный с ментором (руководителем проекта)

Сдать в учебный офис:

Титульный лист проекта Пояснительной записки с подписями студента и руководителя с 10-балльной оценкой от руководителя

Загрузить в ЛМС:

Проект Пояснительной записки

Выложить на Github:

Предоставить доступ к репозиторию проекта на github Центру практик и проектной работы

До 30.05.2019 нужно

Сдать в учебный офис:

Техническое задание с подписями

Пояснительную записку с подписями

Отзыв руководителя с 10-балльной оценкой и подписью

Отчет о проверке Пояснительной записки на плагиат

Загрузить в ЛМС:

Техническое задание

Пояснительную записку

Скан отзыва руководителя с 10-балльной оценкой и подписью

Отчет о проверке Пояснительной записки на плагиат

Выложить на Github:

Исходный код

**Контрольная точка 3:**

До 01.06.2019-30.06.2019 нужно

Защитить проект перед комиссией

Комиссия оценивает защиту с учетом отзыва руководителя, выступления студента, ответов на вопросы.

**8. Порядок контроля и приёмки**

* 1. **Виды испытаний.**

Должна быть произведена проверка корректного выполнения программой заложенных в неё функций, то есть осуществляется функциональное тестирование программы.

Общие требования к приёмке программы:

* 1. **Общие требования к приемке работы**

Приём программы будет утверждён при корректной работе программы при условии, что программа выдаст правильный и корректный вывод на тестовых данных и пройдёт все необходимые тесты.

**Приложение 1**

**Словарь терминов**

Информационные системы – это ERP, CRM, BPM и другие системы в которых содержится информация о деятельности компании (бухгалтерские и финансовые отчёты, информация о закупках и продажах и т. д.)

Process Mining— это подход к анализу бизнес-процессов, базирующийся на автоматизированной реконструкции фактической последовательности событий процесса на основании данных из информационных систем. Он является одним из наиболее перспективных направлений развития анализа бизнес-процессов. После завершения процесса майнинга создаётся блок схема бизнес-процессов.

Граф бизнес-процессов – графическое представление бизнес-процессов компании, полученное после применения Process Mining.

Аномалия в бизнес-процессе – это некое отклонение в бизнес-процессе, которое мешает ходу его нормального выполнения или некоторая характеристика бизнес-процесса (например, время выполнения), которая отклонена от нормального значения.

Модульное программирование — это организация программы как совокупности небольших независимых блоков, называемых модулями, структура и поведение которых подчиняются определённым правилам.

Модуль - функционально законченный фрагмент программы.

Подпрограмма – часть модуля, выполняющая какую-то определённую функцию.

KPI метрики - показатели деятельности подразделения (предприятия), которые помогают организации в достижении стратегических и тактических (операционных) целей. В каком-то смысле это критерии на основании которых измеряется качество работы компании.

**Приложение 2**

**Распределение обязанностей на проекте**  
Всего три человека работают над проектом Агузумцян Смбат (Менеджер проекта, разработчик, специалист по Machine Learning), Шафиров Илья (разработчик, тестировщик, специалист по Machine Learning), Цыбакин Александр (специалист по Machine Learning, ответственный за очистку данных)

Агузумцян Смбат:  
1) KPI модуль программы. Ответственен за разработку специальных P2P метрик и их реализации в данном модуле программы.

2) Модуль анализа и поиска аномалий. Должен разработать подпрограмму для анализа SOD (segregation of duties) аномалий. Также должна быть разработана подпрограмма для предсказания (prediction) возникновения данных аномалий

Шафиров Илья:  
1) Ответственен за разработку подпрограммы для определения нужной точности визуализации данных для которых были применены алгоритмы Process Mining

2) Модуль анализа и поиска аномалий. Должен разработать подпрограмму для аномалий, связанных с длительностью исполнения в Process Path.

3) Окончательное тестирование программы.

Цыбакин Александр:

1) KPI модуль программы. Ответственен за разработку общих метрик качества и их реализации в данном модуле программы.

2) Модуль анализа и поиска аномалий. Ответственен за разработку методов для поиска неэффективных процессов (слишком часто повторяющихся циклов и тому подобное) и реализации этих методов в данном модуле программы.