

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования

**АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
(ФГБОУ ВО «АмГУ»)**

Факультет математики и информатики
Кафедра информационных и управляющих систем
Направление подготовки 09.03.02 – Информационные системы и технологии
Направленность (профиль) образовательной программы Безопасность информационных систем

ДОПУСТИТЬ К ЗАЩИТЕ

И.о. зав. кафедрой

 А.В. Бушманов


« 26 » 06 2020 г.

БАКАЛАВРСКАЯ РАБОТА

на тему: Разработка информационной системы «прогнозирования рейтинга фильма», используя технологию «большие данные»

Исполнитель

студент группы 655-об


 20.06.2020

(подпись, дата)

А.Ф. Филин

Руководитель

доцент, физ.-мат. наук

 19.06.2020

(подпись, дата)

В.В. Еремина

Консультант

по безопасности и экологичности

доцент, канд. техн. наук

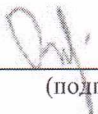
 19.06.2020

(подпись, дата)

А.Б. Булгаков

Нормоконтроль

доцент, канд. техн. наук

 25.06.2020

(подпись, дата)

О.В. Жилиндина

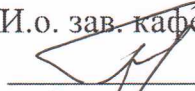
Благовещенск 2020

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
(ФГБОУ ВО «АмГУ»)

Факультет математики и информатики
Кафедра информационных и управляющих систем

УТВЕРЖДАЮ

И.о. зав. кафедрой

 А.В. Бушманов
« 20 » 02. 2020

ЗАДАНИЕ

К выпускной квалификационной работе студента Филина Александра Федоровича

1. Тема дипломной работы: Разработка информационной системы «прогнозирования рейтинга фильма» используя технологию «большие данные» (утверждена приказом от 30.04.2020 №810-уч)

2. Срок сдачи студентом законченной работы: 26.06.2020 г.


3. Исходные данные к выпускной квалификационной работе: отчет о прохождении преддипломной практики, нормативная документация, специальная литература.

4. Содержание выпускной квалификационной работы (перечень подлежащих разработке вопросов): обоснование необходимости разработки и определение требований, проектирование программного продукта, оценка надежности и качества функционирования объекта проектирования, руководство пользователя, описание способов защиты информации для программы, обоснование безопасности и экологичности продукта.

6. Консультанты по выпускной квалификационной работе:
по безопасности и экологичности – Булгаков А.Б., доцент, кандидат технических наук.

7. Дата выдачи задания: 20.02.2020 г.

Руководитель выпускной квалификационной работы: Еремина В.В. доцент, кандидат физико-математических наук.

Задание принял к исполнению: Филин Александр Федорович 

РЕФЕРАТ

Выпускная квалификационная работа содержит 59 с., 19 рисунков, 9 таблиц, 1 приложение, 15 источников.

РАЗРАБОТКА, АНАЛИЗ СУЩЕСТВУЮЩИХ РЕШЕНИЙ, ПРОЕКТИРОВАНИЕ, ПРОГНОЗИРОВАНИЕ, ИНФОРМАЦИОННАЯ СИСТЕМА.

В работе произведено проектирование и разработка информационной системы «прогнозирование рейтинга фильма», используя технологию «большие данные».

Цель работы: проектирование и разработка информационной системы «прогнозирование рейтинга фильма». Для достижения этой цели необходимо выполнить следующий перечень задач: анализ предметной области и разработка информационной системы, включающей следующие функции:

- загрузка данных;
- предобработка данных;
- настройка методов машинного обучения;
- отображение результатов прогнозирования;
- сохранение результатов прогнозирования.

На основе теории было разработано программное обеспечение, соответствующие поставленной задаче. Таким образом, результатом выполнения выпускной квалификационной работы является информационная система «прогнозирование рейтинга фильма» на основе технологии Большие данные.

СОДЕРЖАНИЕ

Введение	6
1 Анализ предметной области	7
1.1 Кинематограф и технологии	7
1.2 Большие данные	8
1.2.1 Понятие Больших данных	9
1.2.2 Процесс работы с Большими данными	9
1.3 Машинное обучение	11
1.3.1 Линейная регрессия	13
1.3.2 Метод опорных векторов	13
1.3.3 Градиентный бустинг (XGBoost)	14
1.4 Сравнительный анализ существующих решений	14
2 Проектирование приложения	16
2.1 Цели и назначение программы	17
2.2 Характеристика функциональных модулей программы	18
2.3 Требования к программе	21
2.3.1 Общие требования	21
2.3.2 Требования к лингвистическому обеспечению	22
2.3.3 Требования к информационному обеспечению	22
2.3.4 Требования к техническому обеспечению	22
3 Описание информационной системы	24
3.1 Выбор методологии разработки программного обеспечения	24
3.2 Выбор средств разработки	24
3.3 Взаимодействие модулей	26
3.4 Описание базы данных	26
3.5 Описание экранных форм	29
4 Угрозы информационной безопасности приложения	34
4.1 Модели угроз	34
4.2 Модель нарушителя	35
4.3 Меры по предотвращению действий злоумышленника	39
5 Безопасность и экологичность	40
5.1 Безопасность	40
5.1.1 Определение безопасных и вредных факторов на рабочем столе пользователя ПЭВМ.	41

5.1.2 Освещение	42
5.1.3 Механические колебания	43
5.1.4 Электромагнитное и ионизирующие излучение	44
5.1.5 Микроклимат	45
5.1.6 Анализ помещения с установленными ПЭВМ	46
5.2 Экологичность	47
5.3 Комплекс физических упражнений при работе на ПЭВМ	49
5.3.1 Упражнения для глаз	49
5.3.2 Упражнения для снятия локального утомления	50
5.3.3 Упражнения для рук.	50
5.3.4 Упражнения для позвоночника	51
5.3.5 Упражнения для ног	52
5.3.6 Упражнения для улучшения кровообращения в мозговой области	52
5.4 Защита от чрезвычайных ситуаций	52
Заключение	56
Библиографический список	57
Приложение А	59

ВВЕДЕНИЕ

Благодаря компьютерным технологиям появилась возможность упростить многие процессы в современном мире, а часть из них и вовсе автоматизировать. Во многом это стало возможно благодаря стремительному развитию искусственного интеллекта в последние года, в частности машинного обучения.

Среди задач классификации, кластеризации, уменьшения размерности и других чаще всего исследователи в области машинного обучения и больших данных работают с задачей регрессии или в общем — прогнозирования. Это объясняется колоссальным ростом количества информации в последние годы, что позволяет выявлять закономерности с помощью информационных технологий.

Кинематограф также использует современные технологии для улучшения качества картин и уменьшения затрат. Стриминговый сервис Netflix один из первых начал активно использовать машинное обучение, а в частности глубокое, для создание рекомендательной системы, которая позволяла бы пользователям за короткий промежуток времени подбирать себе фильм или сериал для просмотра, на основе интересов и уже просмотренных картин. Помимо этого уже сейчас именно искусственный интеллект помогает продюсерам Netflix делать правильные решения еще на стадии препродакшена картины, чтобы угодить пользовательским вкусам.

Объектом разработки является система, позволяющая прогнозировать, благодаря собранным данным, содержащей пользовательские оценки и информацию о фильмах. В данной работе описываются основные этапы работы над программным продуктом и методы, которые были выбраны в качестве основных при обучение модели на основе больших массивов информации.

1 АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

1.1. Кинематограф и технологии

Кинематограф — отрасль человеческой деятельности, основа которой создание движущихся изображений. Кинематограф был изобретен в XIX веке и стал крайне популярен в XX веке. Одна из составляющих частей кинематографа это киноискусство и киноиндустрия. Киноискусство — это вид современного изобразительного искусства, произведения которого создаются при помощи движущихся изображений, киноиндустрия — это отрасль экономики, производящая кинофильмы, спецэффекты для кинофильмов, мультипликацию и, демонстрирующая эти произведения для зрителей.

Кинематограф с момента своего основания напрямую зависел от технологий и технологического прогресса. Исторически создание кинематографа определяется как решения задачи по закреплению на материальном носителе изображения непрерывного движения объектов и проекции этого движения на экран, с целью выполнить замысел режиссера и запечатлить метафору или какой-либо образ на киноплёнке. В дальнейшем, связь кино и современных технологий только усиливалась: разработка магнитной записи и записи звука будущи важными научными открытиями, стали революционными для кино.

Начиная с 70-х годов огромное влияние на кинематограф стали оказывать информационные технологии. Компьютеры из громоздких аналитических машин постепенно превращались в портативные и многофункциональные устройства. Информационные технологии нашли применение в различных сферах жизнедеятельности и, очевидно, стали основой модернизации кинооборудования, процессов съемок и монтажа. Существует общераспространенное представление, что компьютерные технологии очень быстро распространились и изменили образ жизни человека, однако в кинематографе цифровая революция скорее носит характер поступательного эволюционного прогресса. В настоящее время благодаря использованию небольших цифровых камер стало возможно

создание малобюджетного кино, а крупнобюджетные картины используют компьютерную графику для изображения самых причудливых фантазий режиссера.

Информационные технологии способствовали образованию новых жанров и направлений в кино, но при этом были задействованы и в менее творческих сферах, таких как: кинопроизводство и кинобизнес. Использование современных информационных технологий таких как машинное обучение, большие данные, нейронные сети стало стандартом при производстве и продвижении фильмов.

На основе интересов пользователей, выявленных по запросам в сети Интернет и предпочтений в социальных сетях, проанализированных с помощью методов машинного обучения, создаются рекомендательные системы, используемые сейчас во многих онлайн кинотеатрах. Эта технология позволяет на основе некоторого прогноза, составленного индивидуально для каждого человека, подсказать пользователю фильм, который он еще не видел, но с высокой долей вероятности захочет посмотреть.

Компания ScriptBook предлагает технологию, основанную на больших данных и машинном обучении. Разработанная система, проанализировав только сценарий фильма, выдает детальный отчет, где указывает возрастной рейтинг фильма, указывает антагонистов и протагонистов, предсказывает аудиторию проекта и даже потенциальные кассовые сборы.

Информационные технологии уже стали неотделимой частью индустрии развлечений в целом, в том числе и в кино. На данный момент практически каждую новую технологию пытаются использовать как для облегчения труда кинематографистов с целью повысить зрелищность продукта, так и для возможной минимизации риска при производстве кинокартин, связанного с экономической составляющей по окупаемости затрат на производство при выходе продукта в прокат. Одной из таких технологий являются Большие данные.

1.2. Большие данные

1.2.1. Понятие Больших данных

Исходя из названия можно подумать, что термин «большие данные» определяет просто управление и анализ большого объема информации. Но все же данный термин подразумевает нечто большее, чем просто анализ массивных объемов данных. Основная проблема не в том, что организации и предприятия создают огромные объемы данных, а в том, что большая их часть представлена в плохо структурированном формате, «большие данные» (далее БД) — это веб-журналы, видеозаписи, фотографии, программный код или, например, геопространственные данные. Вся эта информация хранится во множестве различных хранилищ, иногда даже за пределами предприятий. В результате организации могут иметь доступ к огромному объему своих данных и не иметь необходимых инструментов, чтобы установить зависимости между этими данными и сделать на их основе значимые выводы. Также стоит отметить, что на сегодняшний день информация обновляется все чаще и чаще, и может сложиться ситуация, когда традиционные методы анализа информации не могут угнаться за огромными объемами постоянно обновляемых данных, что в итоге и открывает дорогу технологиям БД. В сущности, понятие больших данных подразумевает работу с информацией огромного объема и разнообразного состава, весьма часто обновляемой и находящейся в разных источниках в целях увеличения эффективности работы, создания новых продуктов и повышения конкурентоспособности. Консалтинговая компания Forrester дает краткую формулировку: «Большие данные объединяют техники и технологии, которые извлекают смысл из данных на экстремальном пределе практичности».

1.2.2. Процесс работы с Большими данными

Процесс работы с Большими данными обычно состоит из шести шагов:

- Назначение цели исследования;
- Сбор данных;
- Подготовка данных;
- Исследование данных;

- Моделирование данных;
- Отображение и автоматизация.

Обычно работа с большими данными производится в пределах организации или в конкретной сфере. Перед работой с самими данными необходимо описать, что именно необходимо исследовать, какую пользу это принесет компании или сфере, какие данные и ресурсы потребуются в ходе проведения работ.

Далее происходит сбор данных. Необходимо проверять интересующие данные на доступность и возможность использования в исследованиях. Помимо этого, необходимо учесть количество полученной информации и ее представление, например, формат или качество.

При подготовке данных необходимо повысить качество данных и подготовить для дальнейшего исследования. Обычно данная фаза разделяется на три подфазы:

- очистка;
- интеграция;
- преобразование данных.

Во время очистки удаляются некорректные значения из источника данных и устраняются расхождения.

При интеграции происходит расширение информации за счет объединения данных из различных источников.

Наконец преобразование данных гарантирует, что данные находятся в подходящем формате для использования в будущих моделях машинного обучения.

Исследование данных направлено на более глубокое понимание данных. Необходимо понять, как переменные, признаки взаимодействуют друг с другом, оценить распределение данных и определить наличие выбросов в данных. Для достижения данных целей часто удобно использовать описательные статистики, визуальные методы и простое моделирование.

Следующий этап — это моделирование знания предметной области и информации о данных, полученная на предыдущих этапах, используется для получения ответа на вопрос исследования. В ходе моделирования используются методы из статистики, машинного обучения и т. д. Построение модели является итеративным процессом, в ходе которого выбираются переменные для модели, применяется модель и проводится диагностика модели.

Результаты исследования должны быть представлены пользователям и быть понятны другим людям. Эти результаты могут существовать в разных формах, от красочных презентаций до сухих отчетов, а также в виде графиков.

Несмотря на то, что процесс работы с большими данными представляется, как последовательный алгоритм, в действительности он носит больше итеративный характер. Каждая фаза данного процесса может повторяться в зависимости от поступающих новых данных или, например, при обнаружении выбросов во время фазы исследования данных, в связи с неверным импортированием данных. А именно качество входных данных напрямую влияет на работу модели, а, следовательно, и результат, полученный с помощью методов машинного обучения.

1.3. Машинное обучение

Машинное обучение — это область математики, которая занимается изучением закономерностей в данных.

Данные в математике — это некоторая совокупность объектов. Что такое объект в каждой предметной области обозначается индивидуально. Например, объектами могут быть пациенты в больнице, изображение, текст, банковский клиент, фильм и т. д. С точки зрения математики единственное требование, предъявляемое к объекту, заключается в том, чтобы объект мог быть описан некоторой совокупностью переменных, которые собственно этот объект и характеризуют. Так вот машинное обучение заключается в том, что переменные, описывающие объект, можно явно разделить на наблюдаемые и скрытые.

Наблюдаемые переменные — это такие переменные, которым мы явно

можем придать значение для произвольного объекта. Скрытые (или латентные) переменные мы можем соотнести лишь для узкого набора объектов, просто потому, что обычно их измерение зависит от финансовых затрат, либо от человеческих, либо от временных, либо в принципе невозможно измерить, например, эти переменные устанавливают свойства объекта в будущем. И, конечно, ожидается, что между наблюдаемыми и скрытыми переменными есть какая-нибудь зависимость друг от друга.

Задача машинного обучения состоит в том, чтобы найти эту связь. Если же существуют строгие математические модели, связывающие наблюдаемые и скрытые компоненты, машинное обучение не нужно.

Например, в моделях из классической физики. Если нам даны наблюдаемые переменные масса и ускорение, то чтобы найти латентную переменную - силу, нам не нужно применять машинное обучение, а достаточно использовать второй закон Ньютона. Но все-таки в большинстве задач, которые возникают в реальной жизни, таких строгих математических моделей не существует и эта связь не так очевидна. Но зато можно собрать достаточно большую обучающую выборку - такую совокупность объектов, для которых известны и наблюдаемые, и скрытые компоненты.

Классическим примером машинного обучения является задача банковского кредитного скоринга - система оценки кредитоспособности лица, основанная на численных статистических методах. Часто используется в потребительском экспресс-кредитовании на небольшие суммы. Пусть есть банк и есть клиенты, которые хотят получить кредит. Встает вопрос - кому можно выдавать кредит, а кому нет. В качестве наблюдаемых переменных есть характеристики клиентов с заполняемых ими анкет: пол, возраст, образование, уровень доходов, состав семьи и т.д.. Эти переменные мы можем легко «измерить». В качестве скрытой компоненты в простейшем случае выступает бинарная величина - вернет клиент деньги или нет. В более сложных случаях предполагается, что оценивается риск возврата кредита как некоторая вероятность того, что кредит будет не возвра-

щен и тогда возникают более сложные экономические модели принятия решений, когда же кредит выдавать. В качестве обучающей выборки выступает совокупность клиентов из прошлого. Предположим, что у нас было 400 клиентов, из которых 350 кредит вернули. Т.е. это та выборка клиентов, у которых скрытая или латентная переменная принимала одно значение, а 50 клиентов — та выборка, у которой скрытая переменная принимала другое значение. Этого достаточно, чтобы ставить задачу машинного обучения и пытаться математически находить закономерности между наблюдаемыми и скрытыми компонентами в надежде, что они помогут нам для новых клиентов определить, стоит выдавать им кредит или нет.

В данной работе будут приведены наиболее популярные классические методы машинного обучения для прогнозирования переменных.

1.3.1. Линейная регрессия

Линейная регрессия – регрессионная модель, где зависимость критериальной переменной от регрессоров носит линейный характер:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_mX_m$$

$b_0 \dots b_m$ – коэффициенты(параметры) регрессии.

1.3.2. Метод опорных векторов

Основной идеей метода опорных векторов является перевод исходных векторов (данных) в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве.

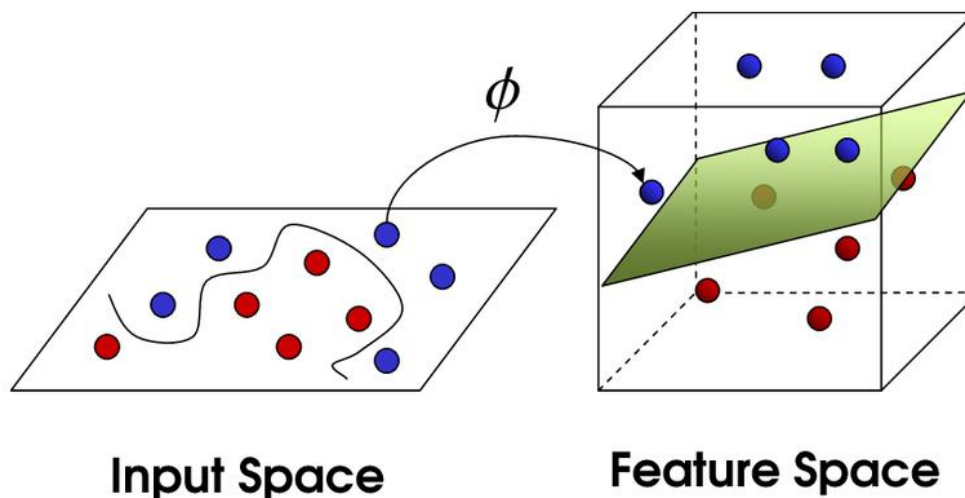


Рисунок 1 – Пример перевода векторов в пространство более высокой размерности

1.3.3. Градиентный бустинг (XGBoost)

Идея градиентного бустинга состоит в построении ансамбля последовательно уточняющих друг друга элементарных моделей. n -ная элементарная модель обучается на «ошибках» ансамбля из $n-1$ моделей, ответы моделей взвешенно суммируются. «Ошибки» здесь в кавычках, поскольку на самом деле каждая последующая модель приближает антиградиент функции потерь, который не обязательно равен разности фактических и предсказанных значений (т.е. ошибке в буквальном смысле). В общем случае в качестве функции ошибки используется среднеквадратичная функция потерь (RMSE).

1.4. Сравнительный анализ существующих решений

Задача прогнозирования пользовательского рейтинга фильмов с использованием современных информационных технологий является одной из основных для киностудий. В случае достаточно точного прогноза еще до снятия фильма можно оценить его успех или провал.

Одну из самых успешных попыток предприняла студия Inventale, которая попыталась на основе данных IMDb выявить факторы успеха будущих проектов киностудий. На основе 418 334 фильмов и сериалов с помощью алгоритма «Решающих лесов» были выявлены сильные зависимости между жанром и попу-

лярностью у зрителей, годом выхода и жанром, а также влиянием актерского состава фильма на восприятие зрителей. Благодаря данному исследованию можно проследить некоторые зависимости, но никак не использовать его для прикладных задач прогнозирования.

Также существуют аналоги у крупных компаний, таких как Netflix или HBO, но их программное обеспечение используется лишь внутри компании и не является открытым, а тем более портативным. В связи с чем отсутствует возможность подробного сравнения их решений и анализа недостатков и достоинств.

Из этого можно сделать вывод, что на данный момент нельзя выявить прямых аналогов разрабатываемому приложению. Представленные аналоги либо не являются открытыми, либо представляют научные исследования. Разрабатываемое же приложение должно находиться в открытом доступе и быть портативным.

2 ПРОЕКТИРОВАНИЕ ПРИЛОЖЕНИЯ

Проектируемая программа состоит из трех частей: модель, представление и контроллер. Модель предоставляет данные пользователю, а также реагирует на команды контроллера, меняя свое состояние. Представление отвечает за отображение данных и информации пользователю, реагируя на изменения модели. Контроллер интерпретирует действия пользователя, оповещая модель о необходимости изменений.



Рисунок 2 – Структурная схема работы приложения

Для работы с информационной системой пользователю необходимо установить ее на компьютер и открыть. В главном окне пользователю доступна загрузка данных, а также возможность из базовой обработки. При загрузке необходимых данных пользователь может настроить необходимый ему метод машинного обучения и набор данных. Далее программа предложит пользователю

ознакомиться с дополнительной информацией по результатам обучения модели, а также с информацией о результатах прошедших сессий.

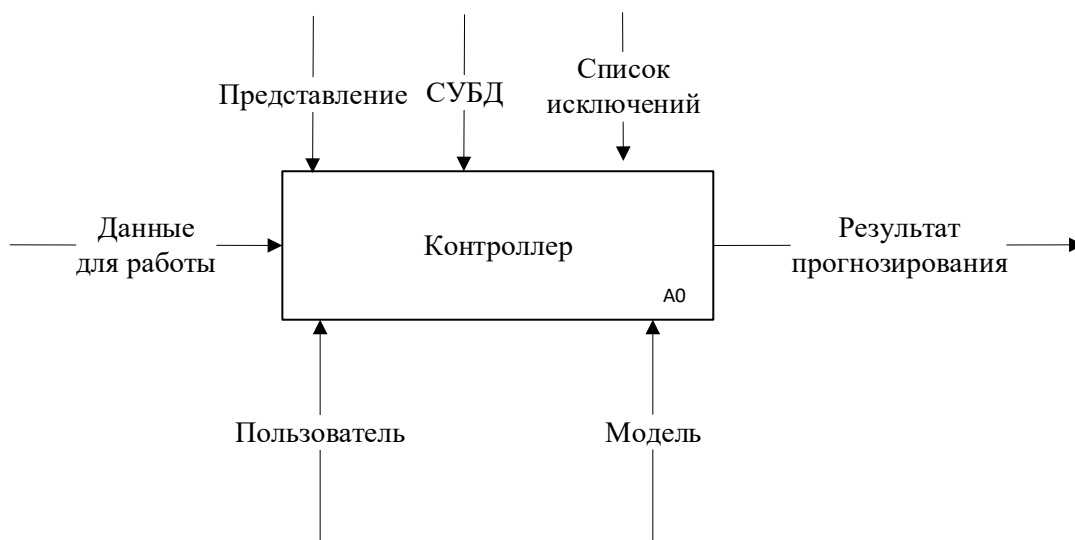


Рисунок 3 - Функциональная модель программы в нотации IDEF0

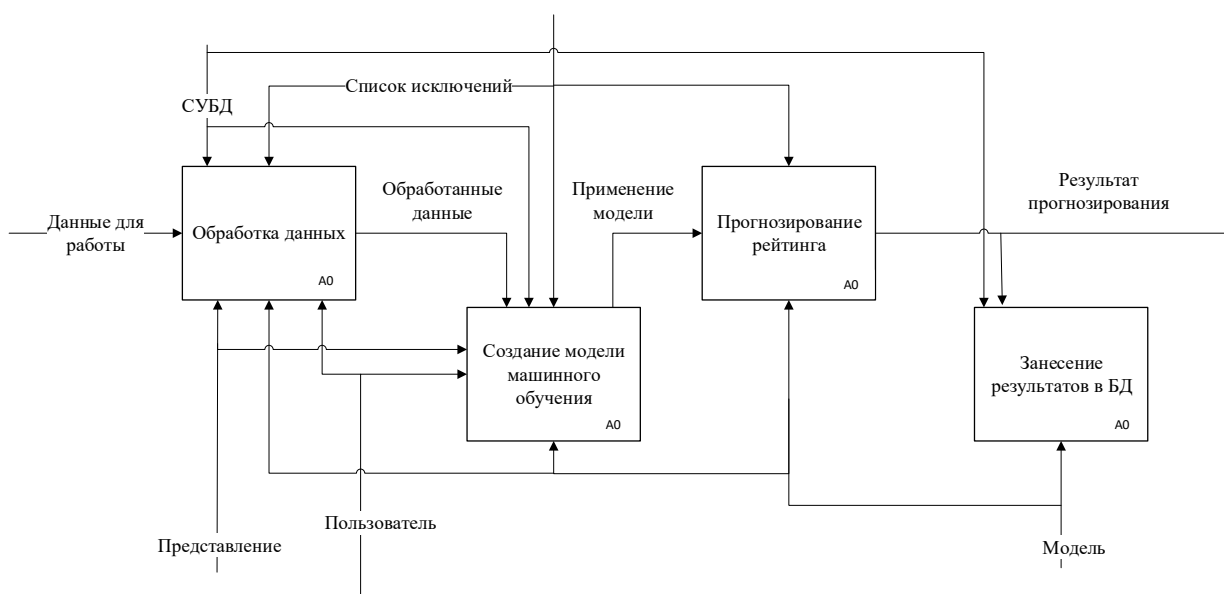


Рисунок 4 - Декомпозиция функциональной модели приложения

2.1. Цели и назначение программы

Программа предназначена для обработки данных табличного вида формата .csv. Перед работой с программой необходимые данные должны быть собраны и представлены в нескольких файлах с конкретными названиями.

Цели создания информационной системы – это создание портативного приложения для прогнозирования рейтинга фильмов, которое может обрабаты-

вать большие массивы информации в табличном виде с помощью методов машинного обучения.

Основные задачи программы:

- загрузка пользовательских данных;
- предобработка пользовательских данных;
- прогнозирование пользовательского рейтинга фильмов;
- возможность настройки методов машинного обучения для прогнозирования;
- предпросмотр необходимых табличных данных;
- хранение информации о моделях машинного обучения, данных и результатах обучения.

2.2. Характеристика функциональных модулей программы

Описание основных функций программы:

- функция «открыть таблицу» позволяет загрузить необходимый файл в программу и отобразить его в главном окне;
- функция «подготовка данных» необходима для обработки большого массива данных в форму, подходящую для машинного обучения. Время работы данной функции может занимать от пары минут до нескольких часов в зависимости от количества данных;
- функция «сохранить шаблоны» позволяет пользователю скачать заранее подготовленные шаблоны таблиц для заполнения данными в понятном для программы порядке;
- функция «сохранить тестовые данные» необходима, чтобы пользователь мог опробовать возможности и работоспособность программы, если у него нет возможности собрать данные самостоятельно;
- функция «применить машинное обучение» необходима для подтверждения пользовательских настроек на этапе выбора типа машинного обучения и типа данных. В зависимости от выбранных параметров и количества данных функция может выполняться от пары секунд до десятков минут;

- функция «результаты XGB» позволяет просмотреть предыдущие итерации обучения модели на основе градиентного бустинга и данные, которые были использованы;

- функция «результаты SVM» позволяет просмотреть предыдущие итерации обучения модели на основе метода опорных векторов и данные, которые были использованы;

- функция «результаты LR» позволяет просмотреть предыдущие итерации обучения модели на основе линейной регрессии и данные, которые были использованы;

На основе функционала программы можно выделить несколько модулей:

- модуль загрузки данных - является частью представления, обеспечивает загрузки файлов с данными в программу, а также их проверку на повторы, с помощью базы данных;

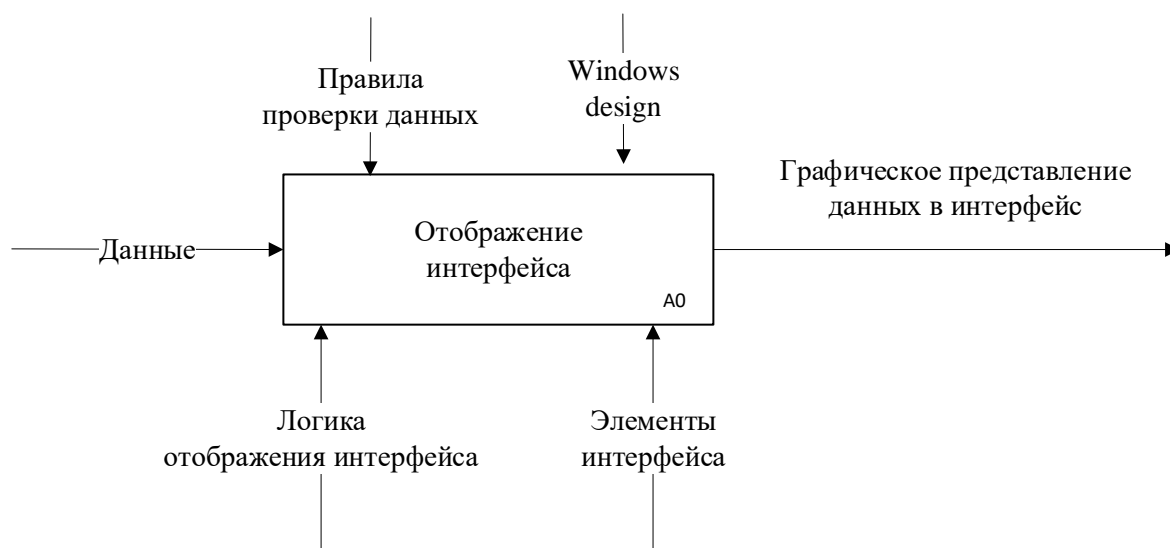


Рисунок 5 - Функциональная модель модуля интерфейса в нотации IDEF0

- модуль интерфейса пользователя – является представления, то есть обеспечивает взаимодействие пользователя с программой через графический интерфейс;

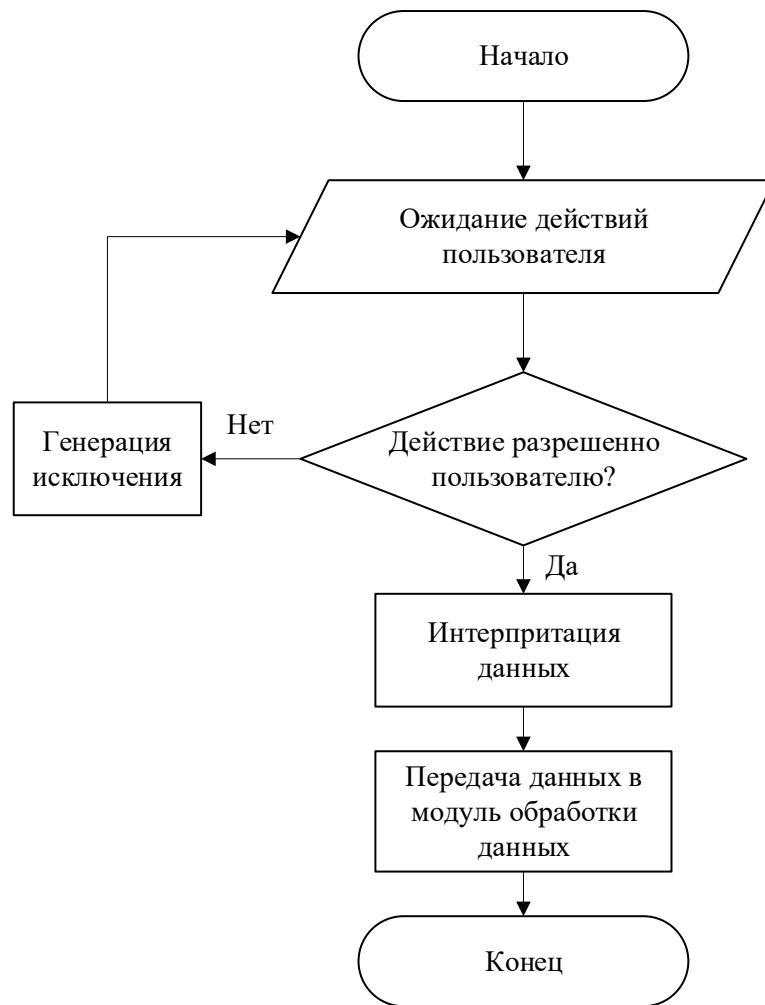


Рисунок 6 – Диаграмма работы модуля передачи данных

- модуль передачи данных – является контроллером, передает информацию о действиях пользователя в модуль обработки данных;

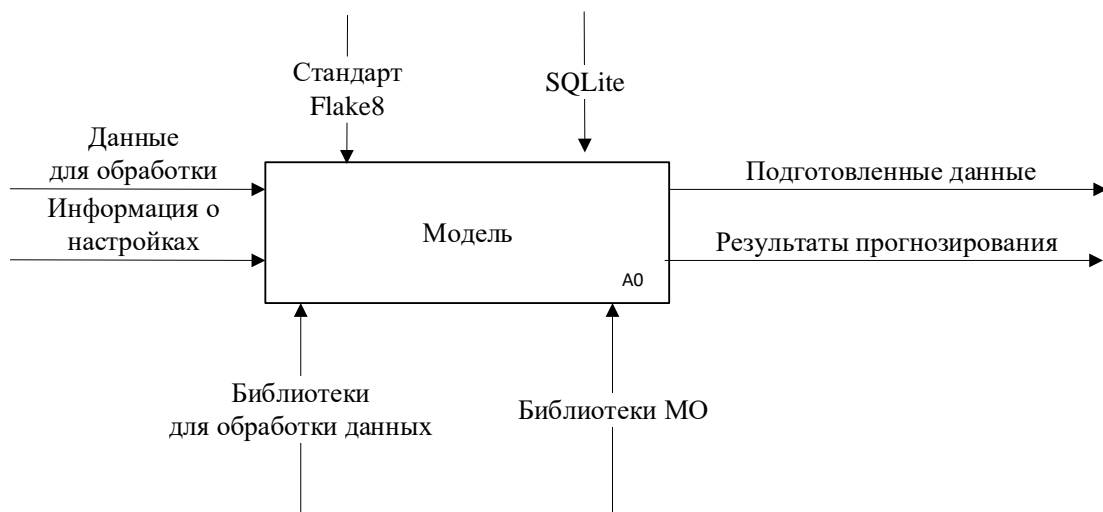


Рисунок 7 - Функциональная модель модуля обработки данных в нотации IDEF0

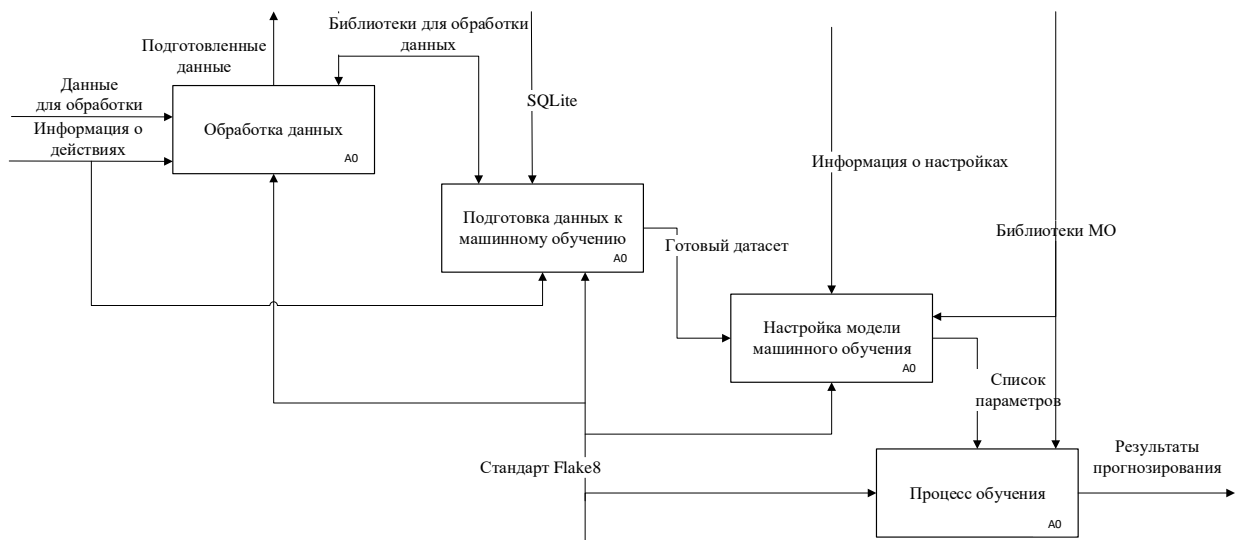


Рисунок 8 – Декомпозиция модуля обработки данных

- модуль обработки данных – является моделью, выполняет все вычислительные задачи и обеспечивает хранение информации о данных. В зависимости от входных данных, модуль либо занимается обработкой больших объемов данных для их дальнейшего использования при машинном обучении, либо сам машинным обучением.

Данные модули необходимы для корректной работы информационной системы.

2.3. Требования к программе

2.3.1. Общие требования

Информационная система должна содержать следующие компоненты:

- модуль загрузки данных;
- модуль интерфейса пользователя;
- модуль передачи данных;
- модуль обработки данных.

Данный продукт может подвергаться модификациям и улучшениям со стороны на основе желаний или интересов пользователей, так как является открытым продуктом.

Пользователями системы должны быть люди знакомые с основными понятиями машинного обучения.

Программный продукт можно развертывать на множестве персональных компьютеров в силу его портативности.

Требования к надежности оборудования и ПО:

- интерфейс должен быть простым и понятным с низкой сложностью освоения;
- при наличии опыта работы со схожими интерфейсами пользователь должен работать быстрее;
- цветовая гамма интерфейса не должна мешать работе пользователя;
- возможность ввода и вывода данных с клавиатуры и с мыши;
- возможность выбора пути в файловой системе для сохранения данных программы.

2.3.2. Требования к лингвистическому обеспечению

Информационная система должна быть полностью на русском языке. Ввод данных должен быть реализован исключительно цифрами. При попытке ввода данных символами или буквами должна быть выведена ошибка.

Для написания кода был выбран высокоуровневый язык программирования общего назначения - Python.

2.3.3. Требования к информационному обеспечению

При работе с программным продуктом, подразумевается работа как с малыми, так и с большими объемами данных. Из этого следует, что приложение должно использовать накопители данных большого размера. В угоду портативности приложения и его легковесности используется база данных SQLite лишь для хранения результатов прогнозирования и информации о расположении данных. Стандартных ресурсов компьютера считать недостаточным.

2.3.4. Требования к техническому обеспечению

Основная нагрузка по обработке данных, обучению моделей и хранению лежит на ПК, на котором установлена информационная система. В связи с этим программный продукт требует значительных вычислительных мощностей со стороны ПК пользователя. Среди операционных систем поддерживается

лишь Windows 10, так как часть библиотек для языка Python, используемых в процессе прогнозирования, написаны исключительно под эту операционную систему. Для процесса обучения моделей машинного обучения необходим процессор с минимум 4 ядрами. Для комфортной работы с большим объемом данных потребуется минимум 4 Гб оперативной памяти.

Таким образом минимальные требования к устройству пользователя следующие:

- ОС: Windows 10;
- Процессор: AMD FX-6350, Intel Core i5 6600K и новее;
- Объем оперативной памяти 4 Гб;
- Место на накопителе 500 Гб;
- Доступ в интернет.

3 ОПИСАНИЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ

3.1. Выбор методологии разработки программного обеспечения

На начальном этапе проектирования необходимо определить методологию разработки программного продукта. Исходя из требований к информационной системе, которые были определены на предыдущем этапе, оптимальными являются методологии, которые могут обеспечить качество и скорость разработки программы, а также позволят вносить изменение в ПО при смене требований.

Одной из подходящих методологий является «Экстремальное программирование» (XP). Она включает в себя следующие особенности, которые позволят полноценно реализовать требования к программному продукту:

а) Короткий цикл обратной связи включает в себя разработку через тестирование и парное программирование, что позволяет добиться более высокого качества приложения.

б) Непрерывный процесс включает в себя непрерывную интеграцию, рефакторинг и небольшие релизы. Непрерывная интеграция и небольшие релизы позволяют еще на ранних этапах разработки оценить перспективность приложения. А рефакторинг кода обеспечит качественную кодовую базу.

3.2. Выбор средств разработки

Исходя из выше установленных требований к информационной системе и совокупности функций, которые она должна выполнять, необходимо выбрать средство разработки с поддержкой СУБД SQLite, с библиотеками для работы с большим объемом информации, с возможностью реализации методов машинного обучения, а также наивысшей скоростью обработки данных. Помимо этого, инструментарий в полной мере должен обеспечивать разработку через тестирование.

Кандидатами стали:

- Python;

- C++.

Для наглядности и удобства сравнения сформируем таблицу, с помощью которой оценим обоих кандидатов:

Таблица 1 – Сравнение инструментов разработки

Наименование	Вариант 1	Вариант 2
Язык программирования	C++	Python
Поддержка SQLite	Да	Да
Библиотеки для работы с большими объемами данных	Xtensor	Pandas, Numpy
Библиотеки для работы с методами машинного обучения	MLPack, Shogun	Scikit-learn, Xgboost
Выделение памяти под переменные	Статическое	Динамическое; частично статическое
IDE	VS 2019, VS code	PyCharm, VS code

При сравнении библиотек для работы с большими данными стоит учитывать, что Xtensor разработана для решения разнообразных задач, что усложняет работу с ней. Pandas же вместе с Numpy созданы для использования в более прикладных задачах, а их функционал напоминает расширенный функционал программы Microsoft Excel.

Библиотеки для работы с методами машинного обучения также отличаются во многом лишь сложностью работы. MLPack и Shogun рассчитаны исключительно для научных задач, проверки сложных гипотез и построения моделей практически с нуля. Scikit-learn, Xgboost же наоборот позволяют из частей собрать необходимый метод машинного обучения и использовать его в прикладных задачах.

К сожалению, Python в основе своей имеет динамическое выделение памяти под переменные, что замедляет его работу. При обычных операциях это можно и не заметить, но при работе с большим количеством данных этот фактор становится критическим. С++ же имеет статическую типизацию, что делает работу с данными на нем максимально быстрой. К счастью, библиотеки Numpy, Pandas в основе своей используют реализацию языка Python – CPython. Благодаря этому объекты, которые обрабатываются функциями из этих библиотек работают с такой же скоростью, как они бы выполнялись на языке С++.

Таким образом наиболее подходящим, исходя из требований к информационной системе, является язык Python.

3.3. Взаимодействие модулей

В качестве архитектурного паттерна всего приложения был выбран MVC (модуль-представление-контроллер), который был описан выше.

Итоговая схема взаимодействия модулей представлена на рисунке 9.

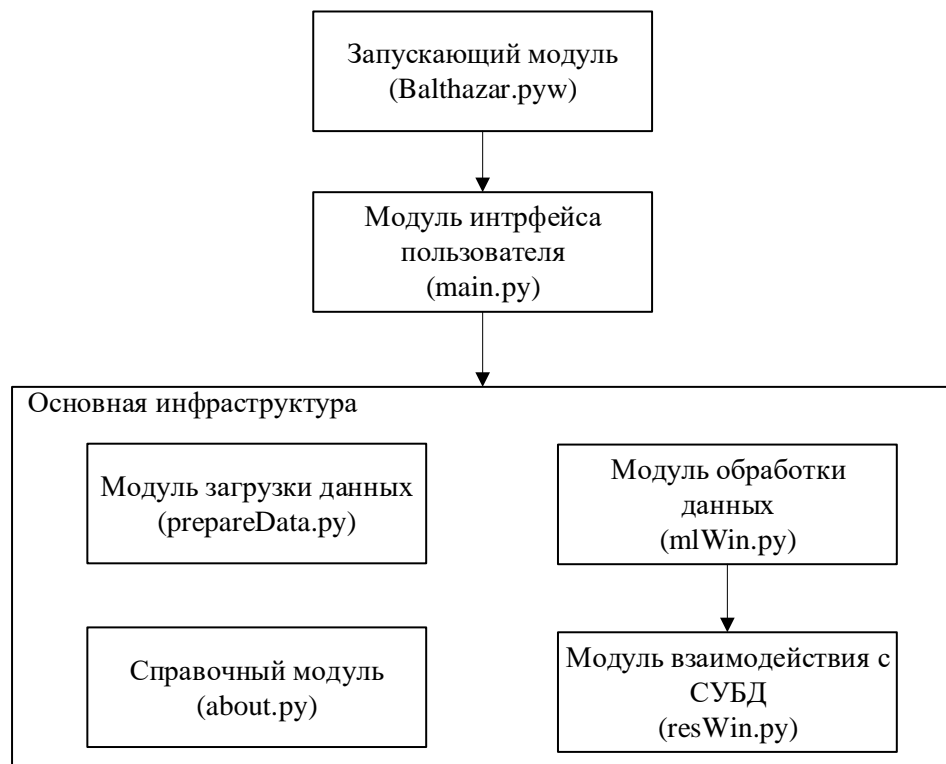


Рисунок 9 – Взаимодействие модулей приложения

3.4. Описание базы данных

Исходя из требований база данных не должна хранить данные, с которыми будет работать информационная система. Ее цель хранить лишь пути к этим данным и результаты предыдущих сессий прогнозирования.

В качестве СУБД была выбрана легковесная SQLite. Эта СУБД позволяет разворачивать БД, не затрачивая большое количество ресурсов компьютера и быстро обрабатывать входящие запросы.

Было разработано четыре таблицы, каждая, из которых хранит определенную часть информации.

Таблица «FILES» хранит пути к файлам, на основе которых обучаются модели. При загрузке каждого нового файла СУБД автоматически применяет функцию автоинкремент и создает новую запись в таблице.

Таблицы «LR», «SVM» и «XGB» хранят информацию о параметрах модели, а также величину ошибки, которую эта модель показала на данных. Информация о данных хранится в последнем столбце в качестве внешнего ключа на таблицу «FILES».

Также связи и сущности в БД предоставлены в нотации Чена в ПРИЛОЖЕНИИ А. Таким образом связь между таблицей «FILES» и таблицами «LR», «SVM» и «XGB» является один-ко-многим.

Таблица 2 – Спецификация атрибутов сущности «FILES»

Название атрибута	Описание атрибута	Тип данных	Диапазон значений	Пример атрибута
idFile	Номер файла	Числовой	-	1
path	Путь к файлу	Текст	>0	C:\path

Таблица 3 – Спецификация атрибутов сущности «LR»

Название атрибута	Описание атрибута	Тип данных	Диапазон значений	Пример атрибута
idLR	Номер модели	Числовой	-	1
Fit intercept	Параметр пересечения с осью	Булевый	-	0

Продолжение таблицы 3

1	2	3	4	5
Normalize	Параметр нормализации	Булевый	-	1
RMSE train	Ошибка тренировочной выборки	Текст	>20	0.5
RMSE valid	Ошибка проверочной выборки	Текст	>20	0.5

Таблица 4 – Спецификация атрибутов сущности «SVM»

Название атрибута	Описание атрибута	Тип данных	Диапазон значений	Пример атрибута
idSVM	Номер модели	Числовой	-	1
Regularization parameter	Параметр регуляризации	Текст	>10	1.0
Kernel	Выбор ядра модели	Текст	>20	poly
Shrinking	Параметр сужения	Булевый	-	0
Max iterations	Параметр максимального количества итераций	Текст	>20	100
RMSE train	Ошибка тренировочной выборки	Текст	>20	0.5
RMSE valid	Ошибка проверочной выборки	Текст	>20	0.5

Таблица 5 – Спецификация атрибутов сущности «XGB»

Название атрибута	Описание атрибута	Тип данных	Диапазон значений	Пример атрибута
idXGB	Номер модели	Числовой	-	1

Продолжение таблицы 5

1	2	3	4	5
Max depth	Параметр глубины дерева	Текст	>10	1000
Learning rate	Параметр точности обучения	Текст	>10	0.3
Booster	Выбор основной функции в модели	Текст	>20	dart
Gamma	Параметр гаммы	Текст	>10	0
Max delta step	Максимальный шаг	Текст	>10	100
RMSE train	Ошибка тренировочной выборки	Текст	>20	0.5
RMSE valid	Ошибка проверочной выборки	Текст	>20	0.5

3.5. Описание экранных форм

Интерфейс реализован с помощью современной и надежной библиотеки PyQt5.

Главная экранная форма содержит одну таблицу, для вывода данных, один выпадающий список и шесть кнопок. Кнопки «Открыть» и «Удалить» позволяет провести выбор файла для загрузки в программу или наоборот удаления из нее. Для удобной работы с загруженными файлами, все их названия помещаются в интерактивный выпадающий список, активный элемент которого отображается в таблице. Кнопка «Справка» открывает справочную форму, в которой можно узнать подробности о программе, инструктаж, а также тестовые файлы и шаблоны для заполнения данными. Кнопка «Подготовить данные» может быть активирована только при загрузки семи ключевых файлов: “credits.csv”, “keywords.csv”, “movies_metadata.csv”, “oscar.csv”, “oscar_director_editor_screenplay.csv”, “oscar_movies.csv”, “ratings.csv”. Шаблоны для заполнения этих файлов и их тестовые версии находятся в окне «Справка», как было сказано выше. Результатом «Подготовки данных» является файл

“movies_metadata_final.csv”, который сохраняется в корневую папку программы.

	1	2	3	4
1	userId	movieId	rating	timestamp
2	1	110	1.0	1425941529
3	1	147	4.5	1425942435
4	1	858	5.0	1425941523
5	1	1221	5.0	1425941546
6	1	1246	5.0	1425941556
7	1	1968	4.0	1425942148
8	1	2762	4.5	1425941300

Рисунок 10 – Главная экранная форма

Сохранить шаблоны

По нажатию кнопки "Сохранить шаблоны" вы скачаете архив с шаблонами .csv для заполнения их данными. Шаблоны представляют собой ряд .csv-файлов со строгой разметкой по признакам и их порядку. При заполнении данных минимальное количество записей не должно быть ниже 200 элементов.

Сохранить тестовые данные

По нажатию кнопки "Сохранить шаблоны" вы скачаете архив с предоставленными данными для тестирования программы и ее возможностей. Архив включает в себя семь неподготовленных .csv таблиц полностью подходящих для обработки в программе "Бальтазар".

Рисунок 11 – Экранная форма справки

При загрузке в программу ключевого файла “movies_metadata_final.csv” и нажатия кнопки «Машинное обучение» пользователь перейдет на экран настройки машинного обучения. Иначе же пользователя ожидает ошибка, о которой его уведомит диалоговое окно.

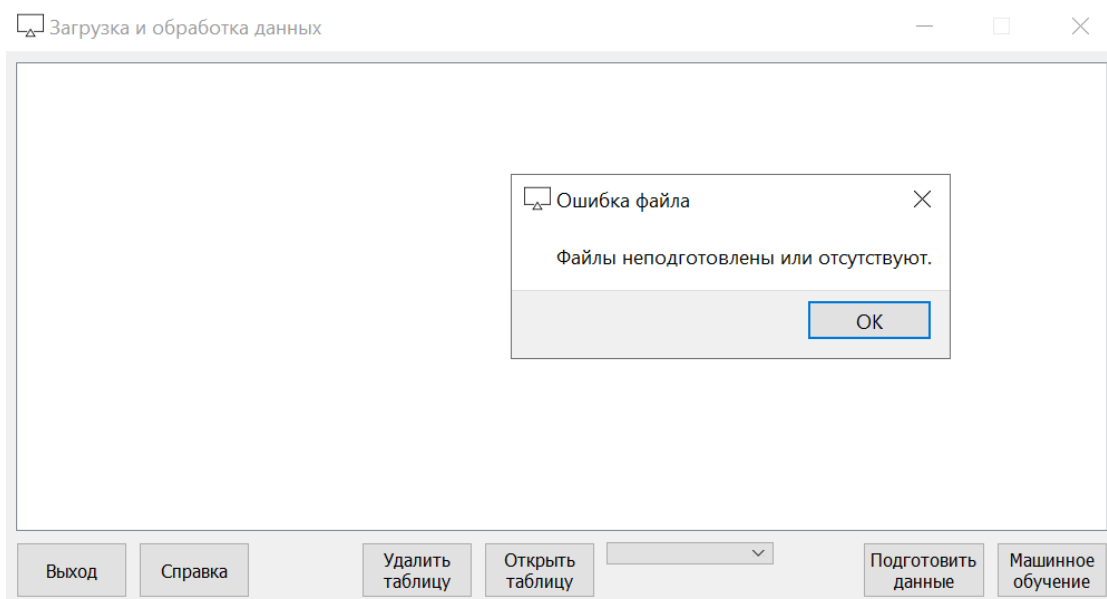


Рисунок 12 – Ошибка при неправильной загрузке файлов

На экране настройки методов машинного обучения пользователю доступны три вида данных и три метода машинного обучения. Каждый из методов машинного обучения можно настроить вручную, а можно использовать значения по умолчанию, если работает неопытный пользователь. Чтобы дополнительно обезопасить пользователя, при выборе одного метода машинного обучения, параметры остальных блокируются. Также при вводе некорректных значений в качестве параметров метода, система автоматически меняет их на значения по умолчанию.

При нажатии кнопки «Продолжить» информационная система применяет выбранные пользователем параметры и переходит к обучению модели на данных, что может занять большое количество времени.

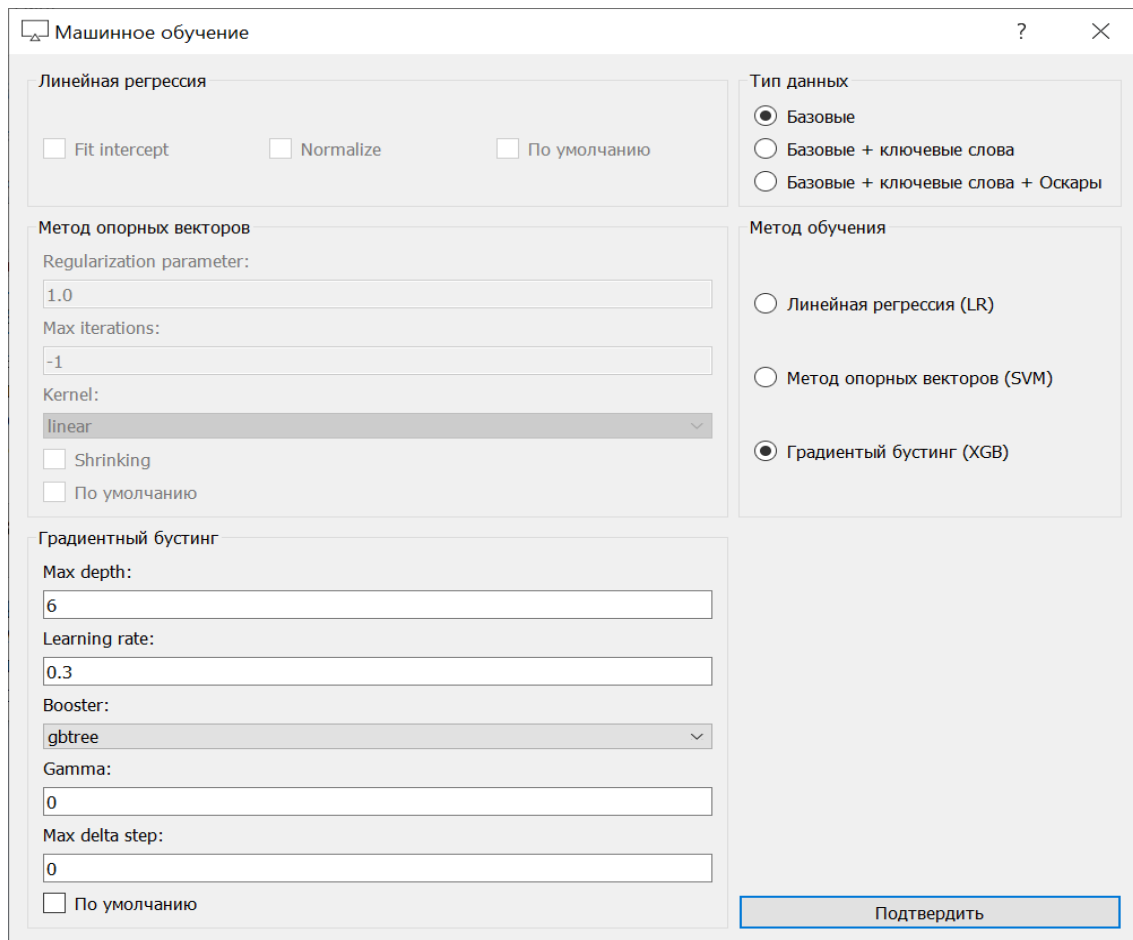


Рисунок 13 – Экран настройки методов машинного обучения

После обучения модели информационная система выведет экран результатов.

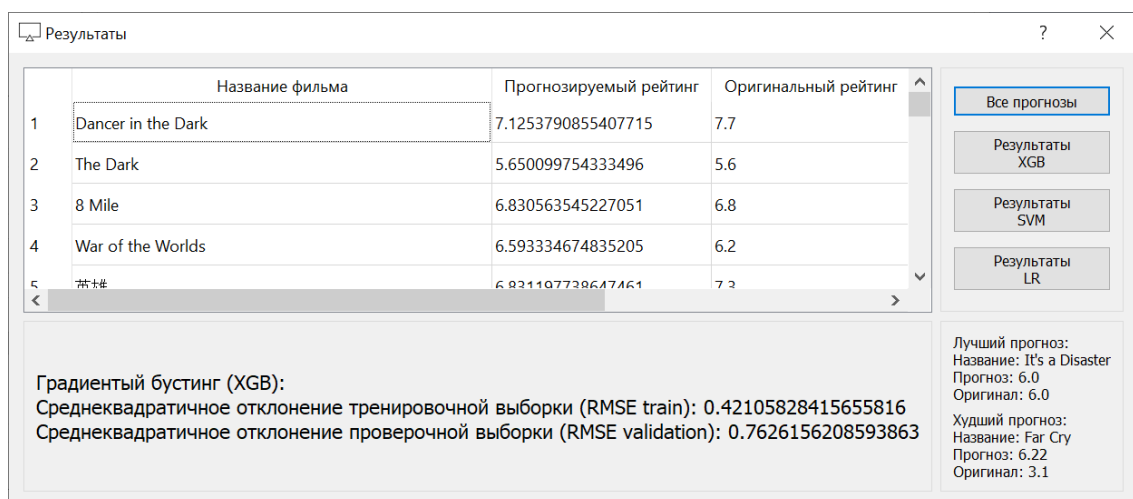


Рисунок 14 – Экранная форма отображения результатов

В рамках данной формы можно узнать величину ошибки, лучший и худший прогноз системы, все прогнозы в табличном виде, а также обратиться к базе данных, чтобы узнать результаты иных обученных моделей.

The screenshot shows a window titled "Результаты" (Results) with a table of model parameters and a summary section. The table has columns for id, Max depth, Learning rate, Booster, Gamma, and Max depth. The summary section includes information about Gradient Boosting (XGB) and the best/worst predictions.

	id	Max depth	Learning rate	Booster	Gamma	Max depth
1	1	6	0.3	gbtree	0	0
2	2	6	0.3	gbtree	0	0
3						
4						
5						

Градиентный бустинг (XGB):
Среднеквадратичное отклонение тренировочной выборки (RMSE train): 0.42105828415655816
Среднеквадратичное отклонение проверочной выборки (RMSE validation): 0.7626156208593863

Лучший прогноз:
Название: It's a Disaster
Прогноз: 6.0
Оригинал: 6.0
Худший прогноз:
Название: Far Cry
Прогноз: 6.22
Оригинал: 3.1

Рисунок 15 – Информация о иных моделях

4 УГРОЗЫ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ПРИЛОЖЕНИЯ

4.1. Модели угроз

Информация является важным ресурсом любой информационной системы, поэтому обеспечение сохранности и доступности данных представляет собой одну из важных задач при разработке и функционировании ИС.

Безопасность информационной системы характеризуется целостностью секретностью информации, чтобы обеспечить её нормальное функционирование и сохранность конфиденциальных данных. Для реализации конфиденциальности, достоверности, доступности и целостности информации необходимо внедрить механизмы предотвращения несанкционированного доступа с целью кражи, модификации и уничтожения информации.

Стоит отметить, что под угрозой понимается потенциально возможное случайное или преднамеренное событие, процесс или действие, приводящее к нарушению конфиденциальности, целостности, доступности информации и поддерживающей её инфраструктуры, нанося ущерб владельцу или пользователю информации.

При решении вопроса защищённости информационной системы прибегают к моделям угроз, модель возможного нарушителя и мерам по предотвращению действий злоумышленника:

- Модель угрозы – это свойства или характеристика угроз информационной безопасности, которая может быть описана физически, математически или визуально;
- Модель возможного нарушителя – это человек, который случайно или намеренно нарушил информационную безопасность в следствии своих действий;
- Меры предотвращения действия злоумышленника – это комплекс организационных, правовых, технических мер по обеспечению информационной безопасности.

Рассматривая разработанное приложение можно выделить следующие модели угроз:

- Атака «отказ в обслуживании», цель которой вывести ИС из рабочего состояния путём длительной нагрузки на вычислительные мощности системы из-за чего пользователи не могут получить доступ к системе;
- Незащищённость важных данных – получение информации через уязвимости в ПО информации о пользователях;
- Использование компонентов или пакетов с известными уязвимостями, которые есть в публичном доступе в интернете и хорошо документированы;
- Перехват информации.

4.2. Модель нарушителя

Говоря о модели нарушителя, вводится понятие нарушитель – это лицо, умышленно или неумышленно предпринявшее попытку реализации угрозы информационной безопасности. Для атаки принимаются любые действия нарушителем, которые приводят к реализации угрозы путём использования уязвимостей программного или аппаратного обеспечения, применяющиеся для функционирования ИС.

Под уязвимостью понимается любая характеристика или свойство сетевого, или системного компонента ИС, использование которого нарушителем может привести к реализации угрозы.

Для приложения подразделяются следующие категории нарушителей:

- пользователи ИС – пользователи ИС, осуществляющие доступ к ресурсам ИС;
- программисты – программисты-разработчики ПО, которые осуществляют поддержку и администрирование работоспособности ИС;
- злоумышленник – лицо (группа лиц), не являющееся зарегистрированным пользователем, не имеющим полномочий для доступа к ресурсам ИС, но пытающийся получить доступ к информации в системе.

Источниками угроз ИБ можно разделить на внутренние и внешние. К внутренним источникам угроз относят:

- аппаратное обеспечение, применяющиеся в ИС;
- программное обеспечение;
- сетевое оборудование;
- системы жизнеобеспечения.

К внешним источникам угроз информационной безопасности информационной системы относят:

- нарушители информационной безопасности ИС;
- форс-мажорные обстоятельства.

По мотивации атаки на информационную систему можно разделить на две категории:

- преднамеренные угрозы связаны с корыстными целями нарушителей.
- случайные угрозы вызваны ошибками в работе программного обеспечения информационной системы.

Из исходя вышесказанного, для разработанного приложения выделяются следующие модели угроз информационной безопасности:

- угрозы ввиду возникновения ошибок в системном и функциональном программном обеспечении компонентов системы;
- угрозы из-за преднамеренных или непреднамеренных действий человека;
- угрозы, которые возникают вследствие физических повреждений, отказов;
- неисправностей технических средств системы, её отдельных компонентов и вспомогательных коммуникаций;
- угрозы, возникающие вследствие форс-мажорных обстоятельств.

В таблицах 6, 7, 8 и 9 приведены модели угрозы.

Таблица 6 – Модели угроз, возникающие в системном и функциональном программном обеспечении компонентов системы

Вид угрозы	Описание угрозы	Источник
Отказ системного ПО	Отказ системного ПО из-за наличия ошибок, влекущих возникновение уязвимостей и сбоев в работе ПО	Программные средства
Отказ прикладного ПО	Отказ прикладного ПО вследствие наличия ошибок, влекущих возникновение уязвимостей и сбоев в работе системы	Программные средства

Таблица 7 – Модели угроз из-за преднамеренных или случайных действий человека

Вид угрозы	Описание угрозы	Источник
Отказ в обслуживании	Выполнение намеренных действий, направленных на возникновение отказа в обслуживании в системах, приложениях, базах и сетях передачи данных	Злоумышленники, программисты, пользователи
Внедрение вредоносного программного обеспечения	Внедрение вредоносного ПО: «тройных коней», «червей», «логические бомбы» приводящее к сбою или нарушению в работе компонентов ИС, а также получению полного контроля над уязвимой системой	Злоумышленники, пользователи, программисты
Ошибка пользователя	Совершение ошибок пользователями при работе с приложениями	Пользователи, временные пользователи

Таблица 8 – Модели угроз у технических средств

Вид угроз	Описание угрозы	Источник
Отказы и сбои дисковых массивов	Отказ в работе дисковых массивов в результате влияния различных факторов	Технические средства

Таблица 9 – Модели угроз в форс-мажорных обстоятельствах

Вид угроз	Описание угрозы	Источник
Отказ системы энерго-снабжения	Отказ подачи электропитания в помещение	Технические средства
Отказ системы кондиционирования	Отказ системы воздушного кондиционирования, приводящий к приостановке работы вследствие выхода рабочих температур за предельнодопустимые рамки	Технические средства
Пожар	Повреждение огнём физических средств, составляющих систему, включая документацию и магнитные носители данных	Форс-мажорные обстоятельства
Затопления	Повреждение водой физических средств, составляющих систему, включая документацию и магнитные носители данных	Форс-мажорные обстоятельства
Стихийные бедствия	Землетрясения, сели, ураганы, наводнения и прочие катастрофические природные явления	Форс-мажорные обстоятельства

4.3. Меры по предотвращению действий злоумышленника

Для защиты от вышеперечисленных моделей угроз, предлагается придерживаться следующих мер:

Для моделей угроз, возникающих при работе программного обеспечения:

- использование последних версий компонентов и пакетов, полученных с надёжных источников;
- быстрое устранение ошибок программистами;
- наличие нескольких запущенных экземпляров приложения на различных серверах, каждый из которых может использоваться в качестве резервного.

Для моделей угроз, возникающих по вине человека:

- ограничение прав пользователей;
- использование специализированной системы защиты конфиденциальной информации на серверах и рабочих станциях от несанкционированного доступа.

Для моделей угроз технических средств и в результате форс-мажора:

- наличие персонала, который будет отслеживать работоспособность оборудования;
- создание резервных копий;
- установка приложения на нескольких машинах.

5 БЕЗОПАСНОСТЬ И ЭКОЛОГИЧНОСТЬ

Информационной безопасностью называется процесс обеспечения конфиденциальности, целостности и доступности информации. При прогнозировании рейтинга фильмов на основе больших данных используются рабочие места для сотрудников, выполняющих прогноз и соответственно необходимо задействовать помещения, в которых они будут находиться. Рабочие места создаются в соответствии с нормативными документами и стандартами (СанПин) в которых предусмотрены вопросы о сохранении здоровья сотрудников при работе на ЭВМ. СанПиНы и другие нормативно – правовые акты, принятые в нашей, направлены на обеспечение условия труда, отвечающих требованиям сохранения жизни и здоровья работников в процессе трудовой деятельности. Они содержат положения, обеспечивающие работникам гарантии прав на охрану труда. Обеспечение здоровых и безопасных условий труда работникам является главной задачей работодателя.

Безопасность жизнедеятельности – наука о комфортном и безопасном взаимодействии человека со средой обитания.

Цель этого раздела в ВКР изучить вопросы техники безопасности и экологии труда для работников, которые занимаются прогнозом на основе больших данных.

При создании рабочего места таким работникам также необходимо разработать рекомендации и комплексы упражнений с целью охраны здоровья сотрудников. В работе будут исследоваться 4 основных положения:

- безопасность;
- экологичность;
- комплекс физических упражнений при работе на ПЭВМ;
- защита от ЧС.

5.1. Безопасность

5.1.1. Определение безопасных и вредных факторов на рабочем столе пользователя ПЭВМ.

Стандартом СанПин 2.2.2/2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы» установлены требования к ПЭВМ, которые необходимо соблюдать.

В соответствии с ГОСТом 12.03.003-2015 выделяются при работе с ПЭВМ следующие опасные факторы, наносящие вред здоровью пользователя:

- электростатические поля;
- электромагнитное излучение;
- опасность поражения электрическим током;
- повышенная или пониженная температура воздуха на рабочем месте;
- повышенный уровень шума на рабочем месте;
- монотонность рабочего процесса;
- утомляемость глаз;
- отсутствие или недостаток в помещении естественного света;
- недостаточная искусственная освещенность рабочего места;
- нервно-эмоциональный перегрузки пользователя.

Согласно СанПиНу 2.2.2/2.2.1340-03, к рабочему месту пользователя, занимающегося прогнозированием на основе больших данных, предъявляются следующие требования:

- рабочая поверхность стола для взрослых пользователей по высоте должна регулироваться в пределах 680-800 мм, при отсутствии такой возможности высота должна составлять 725мм;
- пространство для ног рабочего стола должно составлять не менее 600мм, шириной не менее 500 мм, глубиной на уровне колен не менее 450 мм и на уровне вытянутых ног не менее 650 мм;

- поверхность сидения должна иметь 400 мм, передний край должен быть закругленным и регулироваться в пределах 400-500 мм, углы наклона вперед до 15градусов и назад до 5 градусов;
- подлокотники могут быть стационарными или съемными, их длина не менее 250 мм и ширина 50-70 мм, они регулируются над сиденьями в пределах 230 мм, внутреннее расстояние между подлокотниками в пределах 350 – 500 мм;
- клавиатура располагается на поверхности стола на расстоянии 100-300мм от края, обращенного к пользователю или на специальной рабочей поверхности, которая регулируется по высоте и отделена от стола;
- рабочее место оборудуется подставкой для ног, имеющей ширину не менее 300 мм, глубину не менее 400 мм и которое регулируется по высоте в пределах 150 мм и по углу наклона подставки до 20 градусов.

В соответствии с СанПиНом 2.2.2/2.2.1340-03 необходимо учитывать площадь на одно рабочее место. В случае отсутствия периферийного оборудования и использовании жидкокристаллического монитора на одно рабочее место требуется 4,5 кв.м. площади. Если используются периферийные устройства площадь увеличивается до 6 кв.м.

На рисунке № отобрано рекомендуемое размещение пользователя на рабочем месте.

5.1.2. Освещение

Для оптимальной световой среды применяют различные системы освещения.

Под освещением понимают свет от какого – либо источника, который создает освещенность поверхности, предметов и обеспечивает зрительное восприятие этих предметов.

Выделяют следующие виды освещения:

– естественное освещение – это освещение светом, проникающим через световые проемы в наружных ограждающих конструкциях, в зависимости от расположения оно может быть боковым, верхним или комбинированным;

– искусственное освещение – это освещение помещений и других мест иными источниками, в случае недостаточного естественного освещения. Искусственное освещение подразделяется на рабочие, аварийное, охранное, дежурное, общее, местное и комбинированное. Искусственное освещение должно обеспечивать равномерное освещение всего рабочего стола.

При работе на ПЭВМ используется совмещенное освещение. Оно применяется при недостатке естественного освещения.

Аварийное освещение используется при чрезвычайных ситуациях, в случаи отключения искусственного освещения.

Требования к освещению на рабочих местах четко изложены в СанПиНе 2.2.2/2.1340-03 «Гигиенические требования к персональным электронно – вычислительным машинам и организации работы». В соответствии с этими требованиями коэффициент естественной освещенности не должен быть менее 1,2% в зонах с неустойчивым снежным покровом и не ниже 1,5 % на остальной территории. На рабочем месте должно быть одностороннее естественное освещение. Если естественного освещения недостаточно необходимо применять искусственное освещение. Для этих целей применяются люминесцентные лампы, которые имеют высокую световую отдачу. Освещенность рабочей поверхности стола должна быть в пределах 300 – 500 лк.

5.1.3. Механические колебания

К механическим колебаниям относятся вибрация и шум.

Шум – это беспорядочное сочетание различных по частоте и уровню звуков. Это одна из причин быстрого утомления работающих, которая может вызвать головокружение и даже привести к несчастному случаю.

В соответствии с СанПиНом 2.2.2/2.4.1340-03 шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки

имеет четкие ограничения. В помещениях с созданным рабочим местом пользователя ПЭВМ уровень шума не должен превышать 50дБ. В таблице № 1 представлены значения уровней звукового давления в октавных полосах частот и уровней звука, который создается ПЭВМ.

Уровни звукового давления в октавных полосах со среднегеометрическими частотами									Уровень звука в ДВА
31,5	63	125	250	500	1000	2000	4000	8000	50
Гц	Гц	Гц	Гц	Гц	Гц	Гц	Гц	Гц	
86	71	61	54	49	45	42	40	38	
дБ	дБ	дБ	дБ	дБ	дБ	дБ	дБ	дБ	

Рисунок 16 - Допустимые значения уровней звукового давления в октавных полосах частот и уровня звука, создаваемого ПЭВМ

Вибрация – механическое колебание упругих тел при низких частотах (3-100Гц) с большими амплитудами (0,5 – 0,003 мм).

В случаи если на рабочем месте на работника длительный период воздействует такой фактор как вибрация может наступить потеря трудоспособности, вызванная вибрационной болезнью (невритом). Эта проявляется в виде головных болей, белях в суставах, судорог пальцев, спазмов сосудов и нарушении питания тканей тела. В тяжелых случаях это приводит к полной потери трудоспособности и инвалидности.

5.1.4. Электромагнитное и ионизирующие излучение

При включенном ПЭВМ на рабочем месте создается электромагнитное и ионизирующие излучение. При работе монитор ПЭВМ излучает мягкое рентгеновское излучение. В течении работы на корпусе накапливается статическое электричество. При включенном ПЭВМ запрещается:

- прикасаться к монитору;
- вытирать пыль;

– производить любые манипуляции с отключением отдельных частей ПЭВМ.

Экран монитора должен находиться на расстоянии 60-70 см. от глаз пользователя.

Допустимые уровни электромагнитного излучения регламентируются СанПиНом 2.2.2/2.4.1340-03 «Гигиенические требования к ПЭВМ и организация работы».

5.1.5. Микроклимат

Микроклимат производственных помещений – это характеристика внутренней среды помещений в которых находятся рабочие места. Микроклимат определяется совместно действующими на организм человека температурой, относительной влажностью и скоростью движения воздуха окружающих поверхностей. Эти показатели определены в нормативно – правовом документе ГОСТ 12.1.002 «Общие санитарно-гигиенические требования к воздуху рабочей зоны».

Задача работодателя поддерживать микроклимат рабочего места в пределах гигиенических норм, с целью охраны труда работников.

В помещениях, оборудованных ПЭВМ, должна проводиться ежедневная влажная уборка и систематическое проветривание после каждого часа работы на ПЭВМ.

Оптимальная температура воздуха в помещениях согласно ГОСТу 12.1.005-76 «Воздух рабочей зоны. Общие санитарно-гигиенические требования» в холодные и переходные периоды должна находиться в пределах 20 – 30⁰С, в теплый период года 20-25⁰С. Относительная влажность воздуха 60-40%, скорость движения воздуха не более 0,2м/с во все периоды года.

При работе на ПЭВМ на рабочем месте необходимо поддерживать оптимальную температуру воздуха. ПЭВМ на рабочих местах являются источником существенных выделений, которые повышают температуру тела работника, что приводит к снижению работоспособности. ПЭВМ повышает

температуру всего помещения в целом. Исходя из выше изложенного поддержание температуры на требуемом уровне позволяет обеспечить безопасность и комфортность при работе на ПЭВМ.

В целях поддержания микроклимата в помещении используются системы вентиляции. Система вентиляции – это смена воздуха в помещении, предназначенная для поддержания оптимальных метеорологических параметров помещения и подача чистого воздуха с наружи. В рабочих помещениях используют систему естественной вентиляции, в случаи её недостаточности применяют систему кондиционирования для регулирования микроклиматических параметров в рабочих помещениях для создания комфортных условий труда.

В рабочих помещениях в холодный период года постоянную и равномерную температуру поддерживает система отопления. Виды систем отопления:

- водяные;
- паровые;
- воздушные;
- комбинированные.

Наиболее эффективными в санитарно-гигиеническом отношении признаны системы водяного отопления, их достоинство заключается в том, что они надежны и обеспечивают возможность регулировать температуру в помещении. В помещениях с установленными ПЭВМ в холодный период температура не должна превышать 22-24⁰С, в теплый период – 20-25⁰С. Относительная влажность должна составлять 40-60%, скорость движения воздуха не должна превышать 0,1м/с.

5.1.6. Анализ помещения с установленными ПЭВМ

Для прогнозирования пользовательского рейтинга фильма с использованием больших данных используется помещение, которое имеет площадь 18м². В этом помещении находится два рабочих места с ПЭВМ,

содержащих ЖК монитор, клавиатуру и мышь. Помещение соответствует требованиям СанПиН 2.2.2/2.4 1340-03. На 1 рабочее место приходится 9 м². Рабочая поверхность, сидения соответствуют всем требованиям нормативного документа. Оконные проемы закрыты вертикальными пластиковыми жалюзи. Помещение оснащено защитным заземлением. Температура в помещении круглый год поддерживается в диапазоне 22 – 25⁰С.

Схема размещения ПЭВМ представлена на рисунке № 1

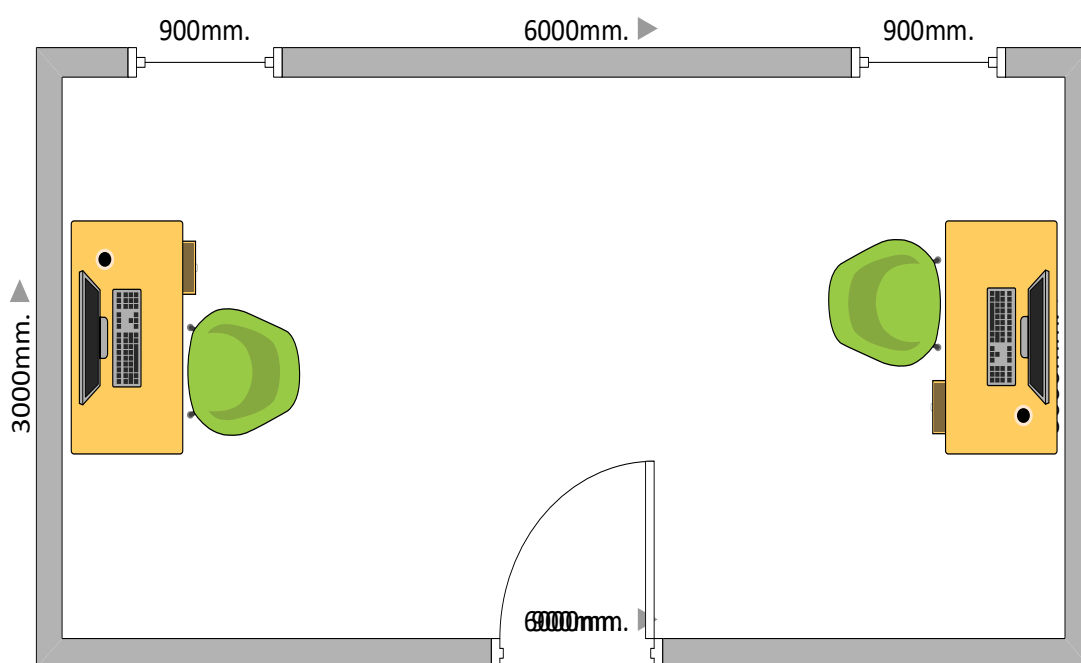


Рисунок 17 - Схема помещения с ПЭВМ

Исходя из анализа всех параметров делаем вывод что данное помещение соответствует всем требованиям предъявляемых СанПиН 2.2.2/2.4 1340.

5.2. Экологичность

Экологичность – качество чего – либо, отражающее его способность не наносить вред окружающей природе. Конструкция ПЭВМ состоит из многих компонентов. Эти компоненты содержат токсические вещества, которые вредны для окружающей среды.

К таким веществам относятся:

- ртуть, которая находится в подсветке ЖК-мониторов. Она может поражать мозг и нервную систему;
- щелочи, находятся в щелочных аккумуляторах. Они поражают слизистые оболочки и кожу;
- никель, цинк располагаются в материнской плате, в батареях питания для ноутбуков. Их воздействие на организм пользователя может вызвать дерматит;
- поливинилхлорид есть в составе кабелей. Воздействие разрушает нервную систему и вызывает раковые заболевания.

Принимая во внимание наличие вредных веществ в ПЭВМ для утилизации пришедших в негодность ПЭВМ используется специальный комплекс методов утилизации. Комплекс включает следующие последовательные действия:

- сортировка металлических и неметаллических частей;
- отправка металлических частей на переплавку для последующего использования в качестве вторичного сырья;
- неметаллические части утилизируются специальным способом.

Утилизируемое оборудование хранится в подсобном помещении, при хранении оно не выделяет вредных веществ. Транспортировка к месту утилизации производится без специальных средств и методов.

При утилизации отходов, содержащих свинец или ртуть, таких как ртутные лампы и аккумуляторы источников бесперебойного питания заключается договор со специальной организацией, которая имеет лицензию на действия, связанные с утилизацией этого класса отходов. Вывоз и непосредственно утилизация осуществляется этой организацией.

Таким образом можно сделать вывод, что в настоящее время создана и внедрена малоотходная технология при утилизации отходов при списании ПЭВМ. В государстве поставлена глобальная задача в перспективе решить

вопрос о переходе на безотходную технологию. Однако это требует комплекса сложных технологических, конструкторских и организационных задач.

5.3. Комплекс физических упражнений при работе на ПЭВМ

Комплекс физических упражнений при работе на ПЭВМ направлен на сохранение индивидуального здоровья и обеспечение полноценной профессиональной деятельности.

Составление рейтинга прогноза фильмов с использованием больших данных предполагает длительную и напряженную работу с ПЭВМ. При неправильной эксплуатации ПЭВМ, а также при нарушении рекомендаций по работе с ПЭВМ возникают проблемы со здоровьем пользователя. Основными проблемами являются:

- проблемы со зрением;
- проблемы с опорно-двигательным аппаратом.

С целью предотвратить возникновения этих проблем необходимо выполнять рекомендации по работе с ПЭВМ.

В качестве наиболее распространённых рекомендаций предлагается:

- чередование работ с использованием ПЭВМ и без него. Во время перерыва необходимо встать с рабочего места и переключить свою деятельность на иные виды деятельности;
- проведение упражнений для глаз с интервалом 20-25 минут работы на ПЭВМ;
- проведение физических упражнений (физкультминутки) в течении 1-2 минут с целью снятия локального утомления. Они выполняются индивидуально каждым пользователем при наступлении локального утомления.

5.3.1. Упражнения для глаз

Упражнения для глаз выполняются в любом положении, которое комфортно пользователю. Необходимо отвернуться от экрана монитора. Этот комплекс необходимо выполнять для уменьшения нагрузки на глаза и для

укрепления глазных мышц. Каждое упражнение выполняется 10-12 раз. В этой работе предлагаются следующие упражнения:

- закрыть рукой один глаз, затем посмотреть вдаль перед собой, время выполнения 2-3 секунды;
- выбрать предмет, находящийся на расстоянии 15-20 см от глаз, посмотреть на него 3-5 секунд и потом перевести взгляд вдаль;
- взять ручку и перемещать её от расстояния вытянутой руки к кончику носа и обратно, следя за её движением;
- в такт дыхания открытыми глазами плавно рисовать «восьмерку» в пространстве.

Сотрудники, страдающие заболеваниями органов зрения должны выполнять следующие правила:

- следовать рекомендациям офтальмолога и терапевта;
- учитывать состояние здоровья;
- помнить об ограничениях, связанных с состоянием органа зрения при выполнении некоторых видов упражнений;
- нельзя выполнять упражнения связанные с сотрясанием тела и требующие напряжения.

5.3.2. Упражнения для снятия локального утомления

Офисная работа предполагает работу в сидячем положении, в результате этого снижается активность в мышцах и мышцы слабеют. С течением времени их сокращение становится медленнее, препятствуя нормальной циркуляции крови. Чтобы сократить нагрузки на собственный организм необходимо регулярно выполнять гимнастику для опорно-двигательного аппарата.

5.3.3. Упражнения для рук.

При работе на ПЭВМ особенно страдают руки. Из-за одинакового положения руки могут затекать. Что бы предотвратить это необходимо выполнять действия, представленные на рисунке № 2.

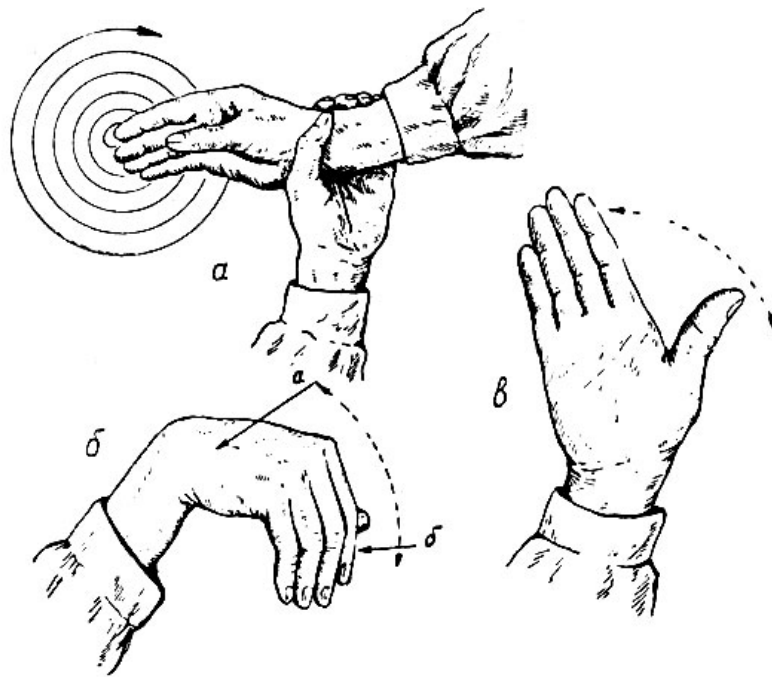


Рисунок 18 - Упражнения для рук

5.3.4. Упражнения для позвоночника

С целью убрать напряжение в спине, теле, суставах необходимо выполнять упражнения представленные на рисунке № 3. Упражнения выполняются 10-12 раз.

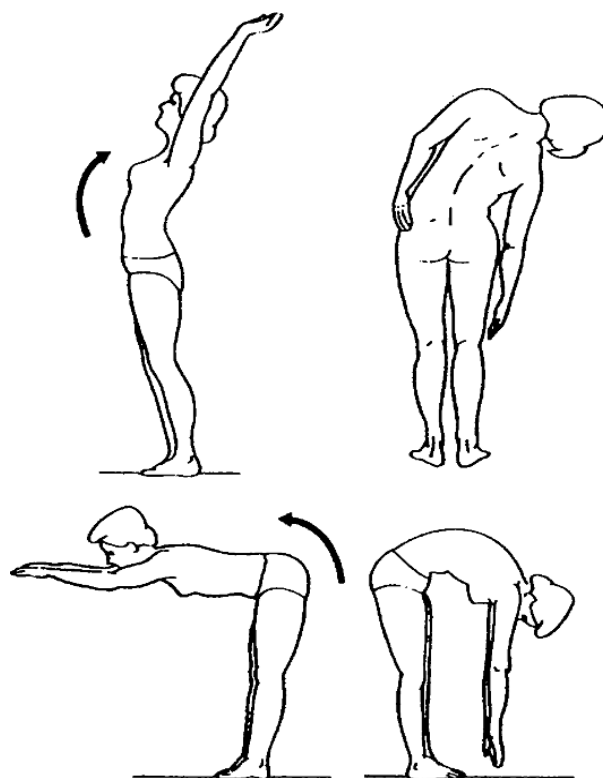


Рисунок 19 - Упражнения для спины

5.3.5. Упражнения для ног

Для восстановления кровообращения в нижней части тела, с целью охраны здоровья пользователя необходимо выполнять следующие упражнения:

- поочередно поднимать носки от пола, пятки при этом неподвижны;
- действия обратные первому упражнению, т.е. поочередное поднятие пяток, носки неподвижны;
- сжатие и расслабление ягодиц;
- поднятие ног под углом 15-30° и выполнение упражнения ножницы.

5.3.6. Упражнения для улучшения кровообращения в мозговой области

Для улучшения кровообращения в мозговой области необходимо занять положение на стуле, руки свесить, расслабиться. Небыстро наклонить голову назад, посчитать медленно до трех и затем принять исходное положение. Далее наклонить голову вперед, посчитать до трех и снова принять исходное положение.

5.4. Защита от чрезвычайных ситуаций

В помещении где находятся ПЭВМ при выполнении работы могут произойти различные чрезвычайные ситуации: пожар, взрыв в здании, разрушение здания от сейсмической активности, затопление, получение урона вследствие электрического разряда. В следствии этого возможны следующие аварийные ситуации:

- обрыв проводов;
- неисправность заземления;
- повреждение электрооборудования;
- повреждение инженерных коммуникаций.

При возникновении аварийной ситуации или при резком ухудшении самочувствия, а также при любых ситуациях, когда возникает угроза жизни и здоровью людей, необходимо выполнение комплекса действий:

- остановить работы;

- обеспечить при необходимости оказания первой помощи пострадавшим;
- отключить электроэнергию;
- открыть аварийные выходы для эвакуации работников;
- доложить о создавшейся ситуации административным лицам организации;
- доложить оперативному дежурному.

Наиболее вероятная ситуация в помещении – пожар.

В помещении где работают ПЭВМ существует электропроводка, которая обеспечивает питание все электроприборы, а также систему освещения.

Помещения относятся к категории В – пожароопасные, согласно СП 12.13130.2009. Поэтому проблема противопожарной безопасности в них является одной из основополагающих при рассмотрении аспектов БЖД.

Эксплуатация ПЭВМ подразумевает наличие большого количества электроприборов, токопроводящих кабелей и высоких нагрузок на электросеть. Их установка, эксплуатация, техническое обслуживание, проверка, замена, утилизация должны соответствовать принятым законодательным нормам и стандартам.

Эксплуатирую ПЭВМ необходимо проверять целостность токопроводящих кабеле, вилки, розетки на отсутствие их повреждений.

ПЭВМ и другая офисная техника должны иметь функцию самоотключения при повышении температуры в период эксплуатации этим приборов.

В помещениях, в которых располагаются ПЭВМ необходима установка средств пожаротушения. Это могут быть огнетушители различных конструкций.

Для более быстрого реагирования на чрезвычайную ситуацию в виде пожара необходимо оснастить помещения автоматической системой пожаротушения, дымовыми датчиками, пожарной сигнализацией, звуковыми извещателями. По всему учреждению должны быть размещены планы

эвакуации при чрезвычайной ситуации в обозначенными на них эвакуационными выходами.

При оборудовании рабочего места для работы ПЭВМ необходимо учесть следующие факторы, которые обеспечат предотвращение нанесения вреда здоровью сотрудникам при пожаре:

- обеспечение наличие проходов к путям эвакуации и эвакуационным выходам;
- установление ПЭВМ и монитора на надежную опору, не допускать его падения;
- запрещено устанавливать в нишах мебельных стенок, в тумбочках и т.п.;
- запрещено установление ближе 1 метра от нагревательных приборов, от горючих предметов (тюлей, декоративных украшений, новогодних елок и т.п.)
- запрещено установление на путях передвижения и эвакуации людей.

Для безаварийной работы перед началом эксплуатации ПЭВМ требуется провести следующие мероприятия:

- провести внешний осмотр места установки ПЭВМ и монитора, убедиться в исполнении требований безопасности, предъявляемых выше;
- провести внешний осмотр ПЭВМ, монитора, электровилки, электрошнура, убедиться в их исправности. Если есть сомнения в их исправности ПЭВМ эксплуатировать запрещается;
- осмотреть место установки на предмет горючих предметов. По возможности убрать их от места установки. При невозможности выбрать другое место установки;
- проверить вентиляционные отверстия в ПЭВМ, в случае их закрытия какими – либо предметами, открыть их или изменить условия установки;

– проверить наличие в помещении, где устанавливается ПЭВМ средства огнетушения, в случае их отсутствия обеспечить их наличие.

Исполнение всех перечисленных выше правил сократит риск возникновения чрезвычайной ситуации, а именно пожара.

ЗАКЛЮЧЕНИЕ

В настоящее время, формируется огромное количество информации о фильмах и зрительских впечатлениях о фильмах, в связи с чем возрастает актуальность приложений, которые могут анализировать эти данные, и на их основе делать выводы, которые помогут в продвижении будущих кинолент.

В ходе выполнения бакалаврской работы были исследованы существующие решения и выявлены их недостатки. Это и закрытость из-за политики крупных киностудий, и большое количество средств на поддержку приложения. На основании этих данных было принято решение создать портативное, открытое приложение для прогнозирования рейтинга фильмов. После чего было выполнено проектирование структуры приложения и базы данных.

По итогу проектирования стали определены требования к продукту, его структура и функционал программных модулей. Во время этапа разработки была выбрана методология разработки программного обеспечения, ей стала практика «Экстремального программирования», так как она обладает значительными преимуществами при разработке современных объектно-ориентированных приложений.

Разработанное приложение предлагает пользователю систему для обработки, моделирования и управления данными. В результате пользователь сможет получать прогнозы на интересующие его фильмы или модели фильмов, в зависимости от введенных в систему данных.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- 1 Чистый код: создание, анализ и рефакторинг. Библиотека программиста. / Пер. с англ. – СПб.: Питер, 2010. – 464 с.
- 2 НПБ 105-03. Определение категорий помещений, зданий и наружных уста-новок по взрывопожарной и пожарной опасности
- 3 Python для сложных задач: наука о данных и машинное обучение / П. Дж. Вандер – СПб.: Питер, 2018 – 572 с.
- 4 Основы Data Science и Big Data. Python и наука о данных / Дэви Силен, Арно Мейсман, Мохамед Али – СПб.: Питер, 2018 – 336 с.
- 5 Python 3 и PyQt5. Разработка приложений. – 2-е изд., перераб. и доп. / Н.А. Прохоренко, В.А. Дронов. – СПб.: БХВ-Петербург, 2019. – 832 с.: ил. – (Профессиональное программирование)
- 6 Вероятностное программирование на Python: байесовсовский вывод и алгоритмы. – СПб.: Питер, 2019. – 256 с.: ил. – (Серия «Библиотека программиста»)
- 7 История кино: Киносъемки, кинопромышленность, киноискусство / Игорь Беленький. – М.: Альпина Паблишер, 2019. – 405.
- 8 Habr: MVVM: полное понимание [Электронный ресурс]. URL: <https://habr.com/ru/post/338518/>
- 9 Pandas: User Guid [Электронный ресурс]. URL: https://pandas.pydata.org/docs/user_guide/index.html
- 10 NumPy: Quickstart tutorial [Электронный ресурс]. URL: <https://numpy.org/doc/stable/user/quickstart.html>
- 11 XGBoost: XGBoost Documentation [Электронный ресурс]. URL: <https://xgboost.readthedocs.io/en/latest/>
- 12 Scikit-learn: Support Vector Machines [Электронный ресурс]. URL: <https://scikit-learn.org/stable/modules/svm.html>

13 Scikit-learn: Linear Models [Электронный ресурс]. URL: https://scikit-learn.org/stable/modules/linear_model.html

14 ГОСТ 12.0.003-2015. Система стандартов по безопасности труда. Опасные и вредные производственные факторы. Классификация. – Введ. 2017-03-01. – М.: Стандартиформ, 2016. – 10 с.

15 СанПиН 2.2.2/2.4.1340-03. Гигиенические требования к персональным электронно-вычислительным машинам и организации работы : утв. постановлением гл. гос. санитар. врача Рос. Федерации от 30.05.2003 №118. – М.: Рид Групп, 2011. – 32 с.

ПРИЛОЖЕНИЕ А

Инфологическая модель в нотации Чена

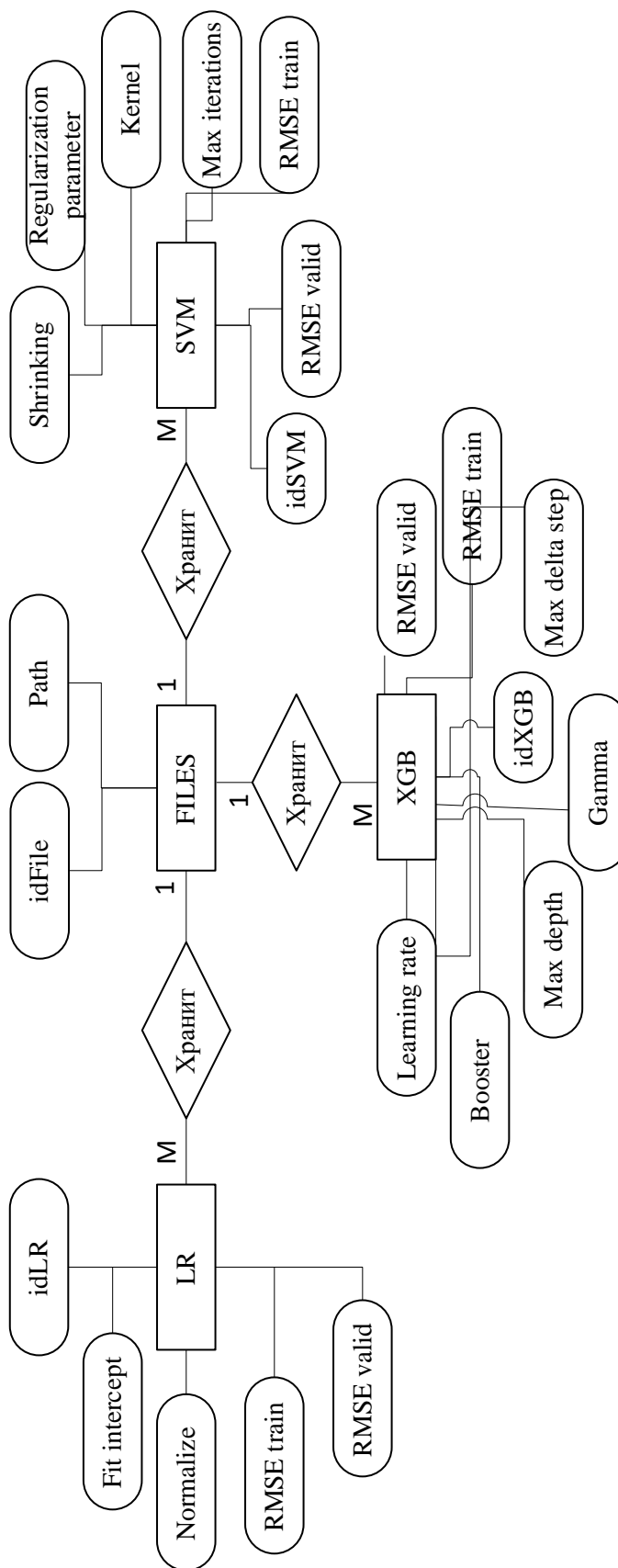


Рисунок А - Инфологическая модель в нотации Чена