Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Макаренко Елена Николаевна
Должность: Ректор
Дата подписания: 30.08.2026 10:59:52
Уникальный программный ключ:
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ
Начальник
учебно-методического управления
Т.К. Платонова
«25» мая 2026 г.

Рабочая программа дисциплины
Введение в Data Mining

Направление подготовки
01.03.05 Статистика

Направленность (профиль) программы бакалавриата
01.03.05.01 Анализ больших данных

Для набора 2026 года

Квалификация
Бакалавр

КАФЕДРА **Статистики, эконометрики и оценки рисков**

Распределение часов дисциплины по семестрам / курсам

Семестр (<Курс>.<Семестр на курсе>)	**2 (1.2)**		Итого	
Недель	16 5/6			
Вид занятий	УП	РП	УП	РП
Лекции	8	8	8	8
Лабораторные	10	10	10	10
Итого ауд.	18	18	18	18
Контактная работа	18	18	18	18
Сам. работа	117	117	117	117
Часы на контроль	9	9	9	9
Итого	144	144	144	144

ОСНОВАНИЕ

Учебный план утвержден учёным советом Университета (протокол № 9 от 03.03.2026 г.).

Программу составил(и): к.э.н., доц., Федосова О.Н.;д.э.н., проф., Ниворожкина Л.И.

Зав. кафедрой: д.э.н., профессор Л.И. Ниворожкина

Методический совет: к.э.н., доцент О.В. Андреева

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	Выработка у студентов компетенций и навыков по работе с большими массивами данных, а также овладение методами обработки этих массивов данных и получение агрегированных показателей

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ПК-4. Способен формировать источники больших данных

В результате освоения дисциплины обучающийся должен:

Знать:

особенности наборов больших данных в части паспортизации, машиночитаемости, стандартизированности и практики публикации, включая нормативные ограничения (соотнесено с индикатором ПК-4.1)

Уметь:

выбирать исходные источники данных в соответствии требованиям, предъявляемым к наборам больших данных (соотнесено с индикатором ПК-4.2)

Владеть:

приемами исследования открытых наборов больших данных на соответствие условиям порядка распространения данных (соотнесено с индикатором ПК-4.3)

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Раздел 1. Возможности, практика и средства Data Mining

№	Наименование темы, краткое содержание	Вид занятия / работы / форма ПА	Семестр / Курс	Количество часов	Компетенции
1.1	Тема 1.1."Введение в Data Mining. Возможности и практика Data Mining в конкретных областях". Понятие Data Mining. Сходство и различия в инструментах статистической обработки данных и инструментах Data Mining. Типы задач Data Mining. Возможности и практика Data Mining в торговле, в банковском деле, в сфере телекоммуникаций, в маркетинге, в страховании, в автомобильной промышленности, в деятельности авиакомпаний, в торговле, в медицине, молекулярной генетике и генной инженерии, прикладной химии.	Лекционные занятия	2	2	ПК-4
1.2	Тема 1.1. "Введение в Data Mining. Возможности и практика Data Mining в конкретных областях". Понятие Data Mining. Сходство и различия в инструментах статистической обработки данных и инструментах Data Mining. Типы задач Data Mining. Возможности и практика Data Mining в торговле, в банковском деле, в сфере телекоммуникаций, в маркетинге, в страховании, в автомобильной промышленности, в деятельности авиакомпаний, в торговле, в медицине, молекулярной генетике и генной инженерии, прикладной химии. Работа с базами данных.	Лабораторные занятия	2	2	ПК-4
1.3	Тема 1.2. "Средства Data Minning". Средства Data Minning: предметно-ориентированные аналитические системы, статистические пакеты, нейросети, системы рассуждений на основе аналогичных случаев (case based reasoning), деревья решений (decision trees), эволюционное программирование, генетические алгоритмы, алгоритмы ограниченного перебора . Эксперименты с YandexGPT, GiGACHAT (сферы, примеры).	Лекционные занятия	2	2	ПК-4
1.4	Тема 1.2 "Средства Data Mining". Работа со средствами Data Mining. Работа с YandexGPT, GiGACHAT.	Лабораторные занятия	2	2	ПК-4
1.5	Тема 1.1. "Введение в Data Mining. Возможности и практика Data Mining в конкретных областях". Понятие Data Mining. Сходство и различия в инструментах статистической обработки данных и инструментах Data Mining. Типы задач Data Mining. Возможности и практика Data Mining в торговле, в банковском деле, в сфере телекоммуникаций, в маркетинге, в страховании, в автомобильной промышленности, в деятельности авиакомпаний, в торговле, в медицине, молекулярной генетике и генной инженерии, прикладной химии. Работа с базами данных. Работа по теме курсовой работы.	Самостоятельная работа	2	20	ПК-4
1.6	Тема 1.2. "Средства Data Minning". Средства Data Minning: предметно-ориентированные аналитические системы, статистические пакеты, нейросети, системы рассуждений на основе аналогичных случаев (case based reasoning), деревья решений (decision trees), эволюционное программирование, генетические алгоритмы,	Самостоятельная работа	2	27	ПК-4

	алгоритмы ограниченного перебора . Эксперименты с YandexGPT, GiGACHAT. Работа по теме курсовой работы.				

Раздел 2. Работа с данными в Data Mining

№	Наименование темы, краткое содержание	Вид занятия / работы / форма ПА	Семестр / Курс	Количество часов	Компетенции
2.1	Тема 2.1. " Источники данных. Открытые данные." Статистические данные. Российские источники статистических данных. Зарубежные источники статистических данных. Открытые данные. Стандарты для открытых данных. Государственные открытые данные. Имитация открытых данных, включая примеры.	Лекционные занятия	2	2	ПК-4
2.2	Тема 2.1. " Источники данных. Открытые данные." Статистические данные. Российские источники статистических данных. Зарубежные источники статистических данных. Открытые данные. Стандарты для открытых данных. Государственные открытые данные. Имитация открытых данных, включая примеры. Работа с данными.	Лабораторные занятия	2	2	ПК-4
2.3	Тема 2.2. "Организация работы с данными". Нормативная база по работе с данными. Признаки и характеристики общедоступной информации. Лицензирование баз данных. Creative Commons Zero. Паспортизация наборов данных. Выбор лицензии данных. Порядок распространения данных. Анонимизация данных – нормативная база и методы. Технические аспекты публикации данных. Особенности машиночитаемости, стандартизованности, используемой схемы и практики публикации наборов больших данных в форматах XLS, XLSX, CSV, JSON, XML, SQL, BSON, JSON lines, ZIP, GZ, 7z, RAR. Типичные ошибки при публикации данных. Парсинг данных.	Лекционные занятия	2	2	ПК-4
2.4	Тема 2.2. "Организация работы с данными". Нормативная база по работе с данными. Признаки и характеристики общедоступной информации. Лицензирование баз данных. Creative Commons Zero. Паспортизация наборов данных. Выбор лицензии данных. Порядок распространения данных. Анонимизация данных – нормативная база и методы. Технические аспекты публикации данных. Особенности машиночитаемости, стандартизованности, используемой схемы и практики публикации наборов больших данных в форматах XLS, XLSX, CSV, JSON, XML, SQL, BSON, JSON lines, ZIP, GZ, 7z, RAR. Типичные ошибки при публикации данных. Парсинг данных. Работа с данными.	Лабораторные занятия	2	4	ПК-4
2.5	Тема 2.1. " Источники данных. Открытые данные." Статистические данные. Российские источники статистических данных. Зарубежные источники статистических данных. Открытые данные. Стандарты для открытых данных. Государственные открытые данные. Имитация открытых данных, включая примеры. Работа с данными с рамках курсовой работы.	Самостоятельная работа	2	30	ПК-4
2.6	Тема 2.2. "Организация работы с данными". Нормативная база по работе с данными. Признаки и характеристики общедоступной информации. Лицензирование баз данных. Creative Commons Zero. Паспортизация наборов данных. Выбор лицензии данных. Порядок распространения данных. Анонимизация данных – нормативная база и методы. Технические аспекты публикации данных. Особенности машиночитаемости, стандартизованности, используемой схемы и практики публикации наборов больших данных в форматах XLS, XLSX, CSV, JSON, XML, SQL, BSON, JSON lines, ZIP, GZ, 7z, RAR. Типичные ошибки при публикации данных. Парсинг данных. Работа с данными в рамках курсовой работы. Выполнение курсовой работы.	Самостоятельная работа	2	40	ПК-4
2.7	Подготовка к промежуточной аттестации	Экзамен	2	9	ПК-4

4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущего контроля и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Учебные, научные и методические издания

	Авторы, составители	Заглавие	Издательство, год	Библиотека / Количество
1		Журнал "Вопросы статистики"		Читальный зал РГЭУ (РИНХ)

	Авторы, составители	Заглавие	Издательство, год	Библиотека / Количество
2	Чубукова И. А.	Data Mining: учебное пособие	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ)| Бином. Лаборатория знаний, 2008	ЭБС «Университетская библиотека онлайн»
3	Федин, Ф. О., Федин, Ф. Ф.	Анализ данных. Часть 2. Инструменты Data Mining: учебное пособие	Москва: Московский городской педагогический университет, 2012	ЭБС «IPR SMART»
4	Воронов, В. И., Воронова, Л. И., Усачев, В. А.	Data Mining - технологии обработки больших данных: учебное пособие	Москва: Московский технический университет связи и информатики, 2018	ЭБС «IPR SMART»
5	Чубукова, И. А.	Data Mining: учебное пособие	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2024	ЭБС «IPR SMART»

5.2. Профессиональные базы данных и информационные справочные системы

ИСС "Гарант"http://www.internet.garant.ru/
ИСС "КонсультантПлюс"
Базы данных Росстата https://rosstat.gov.ru/
База данных Центрального банка РФ http://cbr.ru/hd_base/

5.3. Перечень программного обеспечения

Операционная система РЕД ОС
LibreOffice

5.4. Учебно-методические материалы для обучающихся с ограниченными возможностями здоровья

При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Помещения для всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения:
- столы, стулья;
- персональный компьютер / ноутбук (переносной);
- проектор;
- экран / интерактивная доска.

Лабораторные занятия проводятся в компьютерных классах, рабочие места в которых оборудованы необходимыми лицензионными и/или свободно распространяемыми программными средствами и выходом в Интернет, и/или в специализированных лабораториях, предусмотренных образовательной программой.
Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа к электронной информационно-образовательной среде.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины.

Приложение 1

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

1 Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

1.1 Показатели и критерии оценивания компетенций:

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
ПК-4: Способен формировать источники больших данных			
Знать: особенности наборов больших данных в части паспортизации, машиночитаемости, стандартизированности и практики публикации, включая нормативные ограничения	Описывает стандарты открытых данных, требования к машиночитаемости, практику публикации данных, указывает на нормативные ограничения	Полнота и содержательность ответа; умение приводить примеры во время устного опроса. Правильность и полнота выполнения лабораторных заданий. Соответствие темы и содержания курсовой работы современным направлениям развития data mining	Экзаменационные билеты (1-17), вопросы для устного опроса (1-33), лабораторные задания (1-4), курсовая работа (темы 1-22)
Уметь: выбирать исходные источники данных в соответствии требованиям, предъявляемым к наборам больших данных	выбирает исходные источники данных в соответствие требованиям, предъявляемым к наборам больших данных, в соответствии с требованиями к открытом данным.	Правильность и полнота выполнения лабораторных заданий. Соответствие содержания курсовой работы в части выбора исходных источников данных на соответствие требованиям, предъявляемым к наборам больших данных и/или открытым данным.	Экзаменационные билеты (1-17), вопросы для устного опроса (1-33), лабораторные задания (1-4), курсовая работа (темы 1-22)
Владеть: приемами исследования открытых наборов больших данных на соответствие условиям порядка распространения данных	Анализирует открытые наборы больших данных на соответствие условиям порядка распространения данных	Правильность и полнота выполнения лабораторных заданий. Соответствие содержания курсовой работы в части анализа отрытых наборов больших данных на соответствие условиям порядка распространения данных	Экзаменационные билеты (1-17), вопросы для устного опроса (1-33), лабораторные задания (1-4), курсовая работа (темы 1-22)

1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале:

84-100 баллов (оценка «отлично»)

67-83 баллов (оценка «хорошо»)

50-66 баллов (оценка «удовлетворительно»)

0-49 баллов (оценка «неудовлетворительно»)

2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Экзаменационные билеты

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 1

1. Понятие Data Mining.
2. Лицензирование баз данных.

Практическое задание

Найдите не менее трех источников открытых данных, характеризующих деятельность кредитных организаций. Охарактеризуйте найденные данные (дата публикации, дата последних изменений, тип файла, владелец данных, число объектов и признаков и т.д.)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 2

1. Сходство и различия в инструментах статистической обработки данных и инструментах Data Mining.
2. Порядок распространения данных.

Практическое задание

Найдите не менее трех источников открытых данных, характеризующих деятельность строительных организаций. Охарактеризуйте найденные данные (дата публикации, дата последних изменений, тип файла, владелец данных, число объектов и признаков и т.д.)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 3

1. Открытые данные. Стандарты для открытых данных.
2. Анализ корреляций в Python.

Практическое задание

Найдите не менее трех источников открытых данных, характеризующих деятельность страховых организаций. Охарактеризуйте найденные данные (дата публикации, дата последних изменений, тип файла, владелец данных, число объектов и признаков ит.д.)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 4

1. Признаки и характеристики общедоступной информации.
2. Технические аспекты публикации данных.

Практическое задание

Найдите не менее трех источников открытых данных, характеризующих деятельность организаций культуры. Охарактеризуйте найденные данные (дата публикации, дата последних изменений, тип файла, владелец данных, число объектов и признаков ит.д.)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 5

1. Возможности и практика Data Mining в торговле.
2. Типичные ошибки при публикации данных.

Практическое задание

Для решения задачи продвижения онлайн-школы по ораторскому искусству необходимо спарсить базу пользователей ВК. Опишите алгоритм парсинга данных. Какие фильтры необходимо установить и почему?

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 6

1. Возможности и практика Data Mining в банковском деле.
2. Выбор лицензии данных.

Практическое задание

Для решения задачи продвижения услуг стоматолога необходимо спарсить базу пользователей ВК. Опишите алгоритм парсинга данных. Какие фильтры необходимо установить и почему?

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 7

1. Эксперименты с ChatGPT (сферы, примеры).
2. Статистические данные.

Практическое задание

Для решения задачи продвижения фотостудии необходимо спарсить базу пользователей ВК. Опишите алгоритм парсинга данных. Какие фильтры необходимо установить и почему?

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 8

1. Средства Data Mining.
2. Первичная обработка данных в Python.

Практическое задание

Найдите не менее трех источников открытых данных, характеризующих деятельность кредитных организаций. Охарактеризуйте найденные данные (дата публикации, дата последних изменений, тип файла, владелец данных, число объектов и признаков и т.д.)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 9

1. Возможности и практика Data Mining в медицине, молекулярной генетике и генной инженерии, прикладной химии.
2. Парсинг данных.

Практическое задание

Найдите не менее трех источников открытых данных, характеризующих деятельность строительных организаций. Охарактеризуйте найденные данные (дата публикации, дата последних изменений, тип файла, владелец данных, число объектов и признаков и т.д.)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 10

1. Нормативная база по работе с данными.
2. Возможности и практика Data Mining в сфере телекоммуникаций.

Практическое задание

Найдите не менее трех источников открытых данных, характеризующих деятельность страховых организаций. Охарактеризуйте найденные данные (дата публикации, дата последних изменений, тип файла, владелец данных, число объектов и признаков ит.д.)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 11

1. Возможности и практика Data Mining в торговле.
2. Российские источники статистических данных.

Практическое задание

Найдите не менее трех источников открытых данных, характеризующих деятельность организаций культуры. Охарактеризуйте найденные данные (дата публикации, дата последних изменений, тип файла, владелец данных, число объектов и признаков ит.д.)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 12

1. Возможности и практика Data Mining в деятельности авиакомпаний.
2. Зарубежные источники статистических данных.

Практическое задание

Для решения задачи продвижения онлайн-школы по ораторскому искусству необходимо спарсить базу пользователей ВК. Опишите алгоритм парсинга данных. Какие фильтры необходимо установить и почему?

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 13

1. Возможности и практика Data Mining в автомобильной промышленности.
2. Государственные открытые данные.

Практическое задание

Для решения задачи продвижения фотостудии необходимо спарсить базу пользователей ВК. Опишите алгоритм парсинга данных. Какие фильтры необходимо установить и почему?

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 14

1. Возможности и практика Data Mining в страховании.
2. Creative Commons Zero.

Практическое задание

Для решения задачи продвижения онлайн-школы по ораторскому искусству необходимо спарсить базу пользователей ВК. Опишите алгоритм парсинга данных. Какие фильтры необходимо установить и почему?

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 15

1. Особенности машиночитаемости, стандартизованности, используемой схемы и практики публикации наборов больших данных в форматах XLS, XLSX, CSV, JSON, XML, SQL, BSON, JSON lines, ZIP, GZ, 7z, RAR.
2. Возможности и практика Data Mining в маркетинге.

Практическое задание

Найдите не менее трех источников открытых данных, характеризующих деятельность кредитных организаций. Охарактеризуйте найденные данные (дата публикации, дата последних изменений, тип файла, владелец данных, число объектов и признаков и т.д.)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 16

1. Анонимизация данных – нормативная база и методы.
2. Типы задач Data Mining.

Практическое задание

Найдите не менее трех источников открытых данных, характеризующих деятельность строительных организаций. Охарактеризуйте найденные данные (дата публикации, дата последних изменений, тип файла, владелец данных, число объектов и признаков и т.д.)

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ 17

1. Имитация открытых данных, включая примеры.
2. Паспортизация наборов данных.

Практическое задание

Для решения задачи продвижения фотостудии необходимо спарсить базу пользователей ВК. Опишите алгоритм парсинга данных. Какие фильтры необходимо установить и почему?

Критерии оценивания:

84-100 баллов- оценка «отлично»
67-83 баллов -оценка «хорошо»
50-66 баллов -оценка «удовлетворительно»
0-49 баллов -оценка «неудовлетворительно»

Максимальное количество баллов – 100.

Каждый вопрос оценивается отдельно, максимально в 25 баллов. Максимальный балл за вопросы – 50 баллов. Критерии оценивания отдельного вопроса:

- 21,5-25 баллов. Ответ на вопрос верный; продемонстрировано наличие глубоких исчерпывающих / твердых и достаточно полных знаний, грамотное и логически стройное изложение материала при ответе.
- 17,5-21,49 балла. Ответ на вопрос верный, но с отдельными погрешностями и ошибками, уверенно исправленными после дополнительных вопросов; продемонстрировано наличие глубоких исчерпывающих / твердых и достаточно полных знаний, грамотное и логически стройное изложение материала при ответе.
- 12,5-17,49 балла. Ответ на вопрос частично верен, продемонстрирована некоторая неточность ответов на дополнительные и наводящие вопросы.
- 0-12,49 балла. Ответ на вопрос не верен, продемонстрирована неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Максимальный балл за решение практической задачи – 50 баллов.
Критерии оценивания задачи:

- 42-50 баллов. Задача решена в полном объеме, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов.

- 32-41,9 балла. Задача решена в полном объеме с небольшими погрешностями, выбраны верные инструментальные методы и приемы решения, проведены верные расчеты, сделан полный, содержательный вывод по результатам проведенных расчетов, в расчетах и выводах содержатся незначительные ошибки.
- 25-31,9 балла. Задача решена частично, частично выбраны верные инструментальные методы и приемы решения, проведены частичные расчеты, сделан вывод по результатам проведенных расчетов с отдельными, незначительными погрешностями.
- 0-24,9 балла. Задача не решена или решена частично, частично выбраны необходимые инструментальные методы и приемы решения, расчеты не проведены или проведены частично, вывод по результатам проведенных расчетов не сделан или ошибочен.

Вопросы для устного опроса

1. Раскройте понятие Data Mining.
2. Выделите сходство и различия в инструментах статистической обработки данных и инструментах Data Mining.
3. Опишите типы задач Data Mining.
4. Охарактеризуйте возможности и практику Data Mining в торговле.
5. Охарактеризуйте возможности и практику Data Mining в банковском деле.
6. Охарактеризуйте возможности и практику Data Mining в сфере телекоммуникаций.
7. Охарактеризуйте возможности и практику Data Mining в маркетинге.
8. Охарактеризуйте возможности и практику Data Mining в страховании.
9. Охарактеризуйте возможности и практику Data Mining в автомобильной промышленности.
10. Охарактеризуйте возможности и практику Data Mining в деятельности авиакомпаний.
11. Охарактеризуйте возможности и практику Data Mining в торговле.
12. Охарактеризуйте возможности и практика Data Mining в медицине, молекулярной генетике и генной инженерии, прикладной химии.
13. Перечислите и опишите средства Data Mining.
14. Опишите порядок работы ChatGPT, приведите примеры.
15. Раскройте понятие статистические данные.
16. Охарактеризуйте российские источники статистических данных.
17. Охарактеризуйте зарубежные источники статистических данных.
18. Раскройте понятие открытых данных. Опишите стандарты для открытых данных.
19. Укажите источники государственных открытых данных, приведите конкретные примеры.
20. Выделите и опишите требования нормативной базы по работе с данными, включая меры ответственности за нарушение требований.
21. Раскройте практику имитации открытых данных, приведите примеры.
22. Выделите признаки и характеристики общедоступной информации.
23. Опишите нормативные требования и практику лицензирования баз данных.
24. Охарактеризуйте Creative Commons Zero.
25. Опишите процесс паспортизация наборов данных. Выделите обязательные характеристики, которыми должен обладать набор данных, прошедший паспортизацию.
26. Опишите принципы выбора лицензии данных.
27. Охарактеризуйте порядок распространения данных.
28. Опишите процедуру анонимизация данных, опираясь на нормативную базу.
29. Опишите методы анонимизации данных.
30. Раскройте технические аспекты публикации данных.
31. Охарактеризуйте особенности машиночитаемости, стандартизованности, используемой схемы и практики публикации наборов больших данных в форматах XLS, XLSX, CSV, JSON, XML, SQL, BSON, JSON lines, ZIP, GZ, 7z, RAR.
32. Выделите типичные ошибки при публикации данных, приведите конкретные примеры.
33. Опишите процесс парсинга данных.

Критерии оценивания:

Студенту в течение семестра задается 10 вопросов. Максимальное количество баллов – 20 баллов. Критерии приведены относительно ответа студентом на один вопрос, заданный в ходе устного опроса.

1,5-2,0 балл выставляется студенту, если изложенный материал фактически верен, продемонстрированы глубокие исчерпывающие знания в объеме пройденной программы в соответствии с поставленными программой курса целями и задачами обучения, изложение материала при ответе - грамотное и логически стройное;

1-1,49 балла выставляется студенту, если продемонстрированы твердые и достаточно полные знания в объеме пройденной программы дисциплины в соответствии с целями обучения; материал изложен достаточно полно с отдельными логическими и стилистическими погрешностями;

0,5-0,99 балла выставляется студенту, если продемонстрированы твердые знания в объеме пройденного курса в соответствие с целями обучения, ответ содержит отдельные ошибки, уверенно исправленные после дополнительных вопросов;

0-0,49 балла выставляется студенту, если ответы не связаны с вопросами, допущены грубые ошибки в ответе, продемонстрированы непонимание сущности излагаемого вопроса, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Лабораторные задания

Задание к лабораторной работе № 1

Используя данные сети Internet найти фактически реализованные примеры статистической обработки данных и применения инструментов Data Mining (в торговле, в банковском деле, в сфере телекоммуникаций, в маркетинге, в страховании, в автомобильной промышленности, в деятельности авиакомпаний, в торговле, в медицине, молекулярной генетике и генной инженерии, прикладной химии), идентифицировать какие именно типы задач были решены при помощи инструментов Data Mining, а какие являются результатом статистической обработки. Привести примеры баз данных в конкретной области. Оформить отчет в LibreOffice. В отчете приложить принт-скрины примеров, ссылки на базы данных, сделать пояснения и выводы.

Задание к лабораторной работе № 2

Используя YandexGPT и GiGACHAT описать средства Data Minin: предметно-ориентированные аналитические системы, статистические пакеты, нейросети, системы рассуждений на основе аналогичных случаев (case based reasoning), деревья решений (decision trees), эволюционное программирование, генетические алгоритмы, алгоритмы ограниченного перебора. Сравнить и проанализировать полученные описания. Оформить отчет в LibreOffice.

Задание к лабораторной работе № 3

1. Подготовить сравнительный анализ двух открытых стандартов данных. Указать на сходства и различия стандартов.
2. Найти и привести не менее 5 примеров имитации открытых данных.
3. Найти не менее двух примеров. В которых соблюдены стандарты открытых данных.

Оформить отчет в LibreOffice, приложить принтскрины и ссылки. Сделать выводы.

Задание к лабораторной работе № 4

1. На основе Mind Card составить карту данных по конкретной отрасли экономики (по согласованию с преподавателем), в которую включить открытые данные и базы данных, имеющие паспортизацию и не имеющих паспортизацию (вывести в отдельные ветки). В описании к карте данных указать
   А) факт машиночитаемости, формат, тип лицензии, наличие паспортизации (полная, частичная, отсутствует),
   Б) наличие типичных ошибок при публикации конкретных данных.

2. Провести парсинг данных (по согласованной с преподавателем тематике) на основании данных социальной сети VK (не 50 наблюдений).

Критерии оценивания:

Задание к лабораторной работе №1
Максимальное количество баллов - 10

9-10 баллов выставляется, если обучающийся: выполнил работу в полном объеме; провел анализ различий реализованной статистической обработки данных и применения инструментов Data Mining, нашел требуемые базы данных в конкретной предметной области, грамотно оформил представленный отчет;

7–8 баллов выставляется, если обучающийся: выполнил работу в полном объеме с соблюдением необходимой последовательности; грамотно оформил представленный отчет; провел анализ различий реализованной статистической обработки данных и применения инструментов Data Mining, нашел требуемые базы данных в конкретной предметной области. материал изложен четко; допускаются отдельные логические и стилистические погрешности, уверенно исправленные после дополнительных вопросов;

5-6 баллов выставляется, если обучающийся: выполнил работу в полном объеме; но провел неполный анализ различий реализованной статистической обработки данных и применения инструментов Data Mining, нашел требуемые базы данных в конкретной предметной области, допускаются отдельные логические и

стилистические погрешности; обучающийся может испытывать некоторые затруднения в формулировке суждений;
0-4 балла выставляется, если работа не выполнена или выполнена не в полном объеме; обучающийся практически не владеет теоретическим материалом, допуская в отчете грубые ошибки, испытывает затруднения в формулировке собственных суждений, неспособен ответить на дополнительные вопросы.

Задание к лабораторной работе № 2
Максимальное количество баллов - 10
9-10 баллов выставляется, если обучающийся: выполнил работу в полном объеме; использовал YandexGPT и GiGACHAT, сравнил и проанализировал полученные описания Data Mining, грамотно оформил представленный отчет;
7–8 баллов выставляется, если обучающийся: выполнил работу в полном объеме, использовал YandexGPT и GiGACHAT, сравнил и проанализировал полученные описания Data Mining, в отчете материал изложен четко; допускаются отдельные логические и стилистические погрешности, уверенно исправленные после дополнительных вопросов;
5-6 баллов выставляется, если обучающийся: выполнил работу в полном объеме; использовал YandexGPT и GiGACHAT, но провел анализ не всех требуемых средств Data Mining, сравнил и проанализировал полученные описания Data Mining, в отчете материал изложен четко; допускаются отдельные логические и стилистические погрешности, уверенно исправленные после дополнительных вопросов;
0-4 балла выставляется, если работа не выполнена или выполнена не в полном объеме; обучающийся практически не владеет теоретическим материалом, допуская в отчете грубые ошибки, испытывает затруднения в формулировке собственных суждений, неспособен ответить на дополнительные вопросы.

Задание к лабораторной работе №3
Максимальное количество баллов - 20
17-20 баллов выставляется, если обучающийся: подготовил сравнительный анализ двух открытых стандартов данных, указал на сходства и различия стандартов; нашел и привел не менее 5 примеров имитации открытых данных; привел не менее двух примеров открытых данных, которых соблюдены стандарты открытых данных; оформил отчет в LibreOffice, приложил принтскрины и ссылки; сделал выводы.
13-16 баллов выставляется, если обучающийся: подготовил сравнительный анализ двух открытых стандартов данных, указал на сходства и различия стандартов; нашел и привел не менее 5 примеров имитации открытых данных; привел не менее двух примеров открытых данных, которых соблюдены стандарты открытых данных; оформил отчет в LibreOffice, приложил принтскрины и ссылки; сделал выводы, в которых содержатся незначительные погрешности.
8-12 баллов выставляется, если обучающийся подготовил сравнительный анализ двух открытых стандартов данных, указал на сходства и различия стандартов; нашел и привел не менее 5 примеров имитации открытых данных; привел не менее двух примеров открытых данных, которых соблюдены стандарты открытых данных; оформил отчет в LibreOffice, приложил принтскрины и ссылки; сделал выводы, в которых содержатся незначительные погрешности. В приводимых примерах допускаются незначительные ошибки, уверенно исправленные после замечаний преподавателя.
0-7 баллов выставляется, если работа не выполнена или выполнена не в полном объеме; обучающийся практически не владеет теоретическим материалом, допуская грубые ошибки, испытывает затруднения в формулировке собственных суждений, неспособен ответить на дополнительные вопросы.

Задание к лабораторной работе №4
Максимальное количество баллов - 40
34-40 баллов выставляется, если обучающийся выполнил задание в полном объеме: составил карту данных по конкретной отрасли экономики (по согласованию с преподавателем), в которую включил открытые данные и базы данных, имеющие паспортизацию и не имеющих паспортизацию (вывел их в отдельные ветки), в описании к карте данных указал факт машиночитаемости, формат, тип лицензии, наличие паспортизации (полная, частичная, отсутствует), наличие типичных ошибок при публикации конкретных данных; провел парсинг данных (по согласованной с преподавателем тематике) на основании данных социальной сети VK , получив не менее 50 наблюдений.
26-33 баллов выставляется, если обучающийся выполнил задание в полном объеме: составил карту данных по конкретной отрасли экономики (по согласованию с преподавателем), в которую включил открытые данные и базы данных, имеющие паспортизацию и не имеющих паспортизацию, но не вывел их в отдельные ветки), в описании к карте данных указал факт машиночитаемости, формат, тип лицензии, наличие паспортизации (полная, частичная, отсутствует), наличие типичных ошибок при публикации конкретных данных; провел парсинг данных (по согласованной с преподавателем тематике) на основании данных социальной сети VK , получив менее 50 наблюдений, при этом в работе содержатся незначительные погрешности.
16-25 баллов выставляется, если обучающийся выполнил задание в полном объеме: составил карту данных по конкретной отрасли экономики (по согласованию с преподавателем), в которую включил открытые

данные и базы данных, имеющие паспортизацию и не имеющих паспортизацию, но не вывел их в отдельные ветки, в описании к карте данных указал факт машиночитаемости, формат, тип лицензии, наличие паспортизации (полная, частичная, отсутствует), наличие типичных ошибок при публикации конкретных данных; провел парсинг данных (по согласованной с преподавателем тематике) на основании данных социальной сети VK , получив менее 50 наблюдений. В карте данных и ее описании допускаются незначительные ошибки, уверенно исправленные после замечаний преподавателя.
0-15 баллов выставляется, если работа не выполнена или выполнена не в полном объеме; обучающийся практически не владеет теоретическим материалом, допуская грубые ошибки, испытывает затруднения в формулировке собственных суждений, неспособен ответить на дополнительные вопросы.

Курсовая работа

Методические рекомендации по подготовке и защите курсовой работы содержатся в Приложении 2.

Примерный перечень тем курсовых работ

1. Возможности Data Mining и практика открытых данных в маркетинге.
2. Возможности Data Mining и практика открытых данных в медицине.
3. Возможности Data Mining и практика открытых данных в экологии.
4. Возможности Data Mining и практика открытых данных в сельском хозяйстве.
5. Возможности Data Mining и практика открытых данных в налогообложении.
6. Возможности и практика Data Mining в торговле.
7. Возможности и практика Data Mining в банковском деле.
8. Возможности и практика Data Mining в сфере телекоммуникаций.
9. Возможности и практика Data Mining в маркетинге.
10. Возможности и практика Data Mining в страховании.
11. Возможности и практика Data Mining в автомобильной промышленности.
12. Возможности и практика Data Mining в деятельности авиакомпаний.
13. Возможности и практика Data Mining в медицине.
14. Возможности и практика Data Mining в экологии.
15. Возможности и практика Data Mining в сельском хозяйстве.
16. Возможности и практика Data Mining в налогообложении.
17. Анализ практики предоставления государственных открытых данных в РФ.
18. Анализ практики имитации открытых данных.
19. Анализ практики нарушения паспортизации данных.
20. Особенности порядка распространения данных: российская практика.
21. Особенности порядка распространения данных: зарубежная практика.
22. Анализ типичных ошибок при публикации данных.

Критерии оценивания:

- 84-100 баллов (оценка «отлично») выставляется за работу, которая носит аналитический характер, содержит грамотно изложенную теоретическую и нормативную базу, подробный, аргументированный анализ и наличие достаточного количества практических примеров данных, наборов или баз данных, имитации данных в соответствии с выбранной темой, характеризуется логичным, последовательным изложением материала с соответствующими выводами и обоснованными предложениями, имеет положительный отзыв руководителя, при защите обучающийся показывает глубокое знание вопросов темы, свободно оперирует данными исследования, вносит обоснованные предложения по решению проблемы, во время доклада использует наглядные средства, легко отвечает на поставленные вопросы;
- 67-83 баллов (оценка «хорошо») выставляется за работу, которая носит аналитический характер, содержит грамотно изложенную теоретическую базу и нормативную базу, аргументированный анализ и наличие достаточного количества практических примеров данных, наборов или баз данных, имитации данных в соответствии с выбранной темой, критический анализ фактического материала, характеризуется логичным, последовательным изложением материала с соответствующими выводами, однако с не вполне обоснованными предложениями, имеет положительный отзыв руководителя, при защите обучающийся показывает знание вопросов темы, оперирует данными исследования, вносит предложения по решению проблемы, во время доклада использует наглядные средства, легко отвечает на поставленные вопросы;
- 50-66 баллов (оценка «удовлетворительно») выставляется за работу, которая, наряду с вышеуказанными положительными качествами, содержит частичную нормативную базу, частичный анализ и наличие недостаточного количества практических примеров данных, наборов или баз данных, имитации данных в соответствии с выбранной темой, в ней просматривается непоследовательность изложения материала, представлены необоснованные предложения, в отзыве руководителя имеются замечания по содержанию работы и методике анализа, при защите обучающийся проявляет неуверенность, показывает неглубокое знание вопросов темы, не дает полного, аргументированного ответа на заданные вопросы;

- 0-49 баллов (оценка «неудовлетворительно») выставляется за работу, которая не соответствует заявленной теме, объекту, предмету исследования, в которой не реализованы поставленные цели и не решены указанные задачи, не отвечает требованиям, предъявляемым к курсовым работам, в отзыве руководителя имеются критические замечания, при защите обучающийся затрудняется отвечать на поставленные вопросы по теме, не знает теории вопроса, при ответе допускает существенные ошибки.

3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

Текущий контроль успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

Промежуточная аттестация проводится в форме экзамена и защиты курсовой работы.

Экзамен проводится по расписанию промежуточной аттестации в письменном виде. В задании – 2 теоретических вопроса и 1 практическое задание. Проверка ответов и объявление результатов производится в день экзамена. Результаты аттестации заносятся в ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику, должны ликвидировать задолженность в установленном порядке.

Приложение 2

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены следующие виды занятий:

- лекции;
- лабораторные.

В ходе лекционных занятий рассматриваются понятие Data Mining, сходство и различия в инструментах статистической обработки данных и инструментах Data Mining, типы задач Data Mining, возможности и практика Data Mining в конкретных отраслях, средства Data Mining, источники данных., открытые данные, стандарты для открытых данных, имитация открытых данных, нормативная база по работе с данными, признаки и характеристики общедоступной информации, лицензирование баз данных, паспортизация наборов данных, выбор лицензии данных, порядок распространения данных, анонимизация данных – нормативная база и методы, технические аспекты публикации данных, особенности машиночитаемости, стандартизованности, используемой схемы и практики публикации наборов больших данных.

В ходе лабораторных занятий углубляются и закрепляются знания студентов по ряду рассмотренных на лекциях вопросов, развиваются и закрепляются навыки анализа сходства и различий в инструментах статистической обработки данных и инструментах Data Mining, типах задач Data Mining, возможностях и практики Data Mining в конкретных отраслях, средств Data Mining, источников данных., открытых данных, стандартов для открытых данных, имитации открытых данных, нормативной базы по работе с данными, признаков и характеристики общедоступной информации, лицензирования баз данных, паспортизации наборов данных, выбора лицензии данных, порядка распространения данных, анонимизации данных – нормативной базы и методов, технических аспектов публикации данных, особенностей машиночитаемости, стандартизованности, используемой схемы и практики публикации наборов больших данных.

При подготовке к лабораторным занятиям каждый студент должен:

– изучить рекомендованную учебную литературу;

– изучить конспекты лекций;

– подготовить ответы на все вопросы по изучаемой теме;

В процессе подготовки к лабораторным занятиям студенты могут воспользоваться консультациями преподавателя.

Вопросы, не рассмотренные на лекциях и лабораторных занятиях, должны быть изучены студентами в ходе самостоятельной работы. В ходе самостоятельной работы каждый студент обязан прочитать основную и по возможности дополнительную литературу по изучаемой теме, дополнить конспекты лекций недостающим материалом, выписками из рекомендованных первоисточников. Выделить непонятные термины, найти их значение в энциклопедических словарях. Студент должен готовиться к предстоящему лабораторному занятию по всем, обозначенным в рабочей программе дисциплины вопросам.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации студенты могут воспользоваться электронно-библиотечными системами. Также обучающиеся могут взять на дом необходимую литературу на абонементе университетской библиотеки или воспользоваться читальными залами.

Методические рекомендации по подготовке и защите курсовой работы

Введение

Методические рекомендации написаны в соответствии с «Положением о текущем контроле успеваемости и промежуточной аттестации обучающихся по программам высшего образования», принятым Ученым советом ФГБОУ ВО «РГЭУ(РИНХ)», протокол №9 от 12.05.2020.

Курсовая работа является оценочным средством промежуточной аттестации. Это форма контроля, предназначенная для демонстрации обучающимся умений работать с объектами изучения, критическими источниками, справочной и энциклопедической литературой, логично и грамотно излагать собственные умозаключения и выводы, обосновывать и строить априорную модель изучаемого объекта или процесса, создавать содержательную презентацию выполненной работы. Применяется для оценки знаний, умений и навыков обучающихся.

Курсовая работа, представляет собой самостоятельное законченное исследование на заданную тему, свидетельствующее об умении обучающегося работать с научной и нормативной литературой, обобщать и анализировать фактический материал, используя теоретические знания, умения и практические навыки, полученные в процессе освоения основной профессиональной образовательной программы.

Целью курсовой работы по дисциплине ««Введение в Data Mining» является систематизация и конкретизация теоретических знаний, полученных при изучении курса, приобретение навыков ведения самостоятельной исследовательской работы, включая поиск источников данных, а также их анализ на соответствие требованиям, предъявляемым к наборам больших данных и условиям порядка распространения данных.

1. Общие требования к курсовой работе, выбор темы, порядок и контроль работы, защита.

ОБЩИЕ ТРЕБОВАНИЯ. К курсовой работе предъявляются следующие общие требования:

- соблюдение рекомендуемой структуры работы, логичность и структурированность изложения материала;
- соблюдение рекомендаций по оформлению текста, ссылок, библиографического списка, приложений.
- достижение сформулированной цели курсовой работы через решение поставленных задач, обязательное наличие выводов;
- использование учебной и научной литературы с обязательными ссылками на цитируемые фрагменты;
- поиск и анализ конкретных примеров наборов данных, примеров имитации данных, использования персональных данных в рамках выбранной темы курсовой работы с указанием обязательных, полных, действующих интернет-ссылок.

ВЫБОР ТЕМЫ. Темы курсовой работы (примерный перечень) разрабатываются кафедрой статистики, эконометрики и оценки рисков кафедрой. Обучающийся выбирает тему из Примерного перечня тем курсовых работ (Приложение А) **и обязательно согласовывает ее с научным руководителем.** Обучающийся имеет право самостоятельно предложить тему курсовой работы при условии обоснования ее актуальности, уточнить или изменить тему по **согласованию с научным руководителем.** Без согласования с научным руководителем темы не могут быть уточнены или изменены. В рамках одной академической группы повтор тем курсовых работ не допускается.

ПОРЯДОК РАБОТЫ И ЗАЩИТА КУРСОВОЙ РАБОТЫ. В рамках подготовки курсовой работы, обучающемуся предлагается придерживаться следующего порядка действий:

1. Выбрать тему курсовой работы (Приложение А) и согласовать ее научным руководителем.
2. Написать заявление на тему курсовой работы (стандартная форма приведена в Приложении Б), подписать его лично, подписать заявление у научного руководителя.
3. Составить план курсовой работы и согласовать его с научным руководителем, учитывая рекомендации научного руководителя скорректировать план при необходимости.
4. Подобрать учебную и научную литературу, нормативные акты в соответствии с темой курсовой работы, с учетом источников, рекомендуемых научным руководителем.
5. Самостоятельно выполнить курсовую работу, регулярно отчитываясь перед научным руководителем о ходе работы. В случае нарушения графика работы научный руководитель доводит данный факт до сведения заведующего кафедрой в виде служебной записки.
6. Оформить курсовую работу в строгом соответствии с требованиями, изложенными в данных методических рекомендациях, факт нарушения требований будет учтен в процессе итогового оценивания работы.
7. Сдать в установленные сроки готовую курсовую работу научному руководителю (за 14 дней до защиты для дневной формы обучения, за 5 дней до защиты для заочной формы обучения) в печатной форме.
8. Разместить готовую курсовую работу в Moodle, для проверки гиперссылок (обязательны) и карты данных (при наличии) (за 14 дней до защиты для дневной формы обучения, за 5 дней до защиты для заочной формы обучения).
9. Ознакомиться с замечаниями к курсовой работе при их наличии, внести дополнения при необходимости.
10. Явиться на защиту и защитить курсовую работу в соответствии с утвержденным расписанием промежуточной аттестации. Защита проходит в форме собеседования, предполагает обоснование обучающимся правильности выполнения работы, полноты достижения цели и решения поставленных задач. В процессе собеседования студент имеет право подготовить и использовать итоговую презентацию по курсовой работе.

В рамках защиты курсовой работы научный руководитель выявляет качество знаний, умений и навыков обучающегося, а также самостоятельность выполнения работы. В оценке курсовой работы учитывается качество и аргументированность защиты, содержание работы и ее оформление, достижение целей и реализация задач работы, глубина анализа и наличие достаточного количества практических примеров данных, наборов или баз данных, имитации данных в соответствии с выбранной темой. Студент, не представивший в установленный срок работу, или не защитивший ее, считается имеющим академическую задолженность.

2. Рекомендации по выполнению курсовой работы

2.1. Общие требования к структуре, содержанию и техническому оформлению курсовой работы

Структура и содержание курсовой работы определяются ее темой, содержанием курса «Введение в Data Mining» и направлением подготовки.

Структура, содержание и техническое оформление работы должны отвечать следующим **обязательным требованиям:**

1. Обязательные элементы курсовой работы:
   Титульный лист (установленного образца, Приложение В).
   Содержание (образец оформления в Приложении Г).
   Введение
   Основная часть (2 главы по 2-4 параграфа в каждой).
   Заключение
   Библиографический список
   Приложения.
2. Количественные характеристики работы:
   - Минимальное количество глав в работе – 2.
   - Количество параграфов в отдельной главе – 2-4.
   - Объем отдельного параграфа - минимально 3 полных листа текста, максимально 7-8. Параграфы **не должны быть значительно асимметричны** по своему объему, например параграф 1 – объемом 3 листа, параграф 2 – объемом 10 листов.
   - Объем курсовой работы- 25 - 40 листов, без учета приложений.
3. В работе должны присутствовать:
   - обоснование актуальности темы и проблемы, исследуемые автором в работе;
   - теоретические положения и анализ нормативных актов в выбранной предметной области, изложенные на основании изучения литературы по теме;
   - примеры существующих данных, наборов данных, баз данных, анализ соответствия рассматриваемых примеров требованиям, предъявляемым к базам данных, к отрытым данным;
   - аргументированные выводы автора, сформулированные на основании проведенного исследования.
4. В работе должна выдерживаться логичность изложения, связность глав и параграфов для достижения цели работы, полнота изложения материала в рамках параграфа в соответствии с его названием. Содержание каждого параграфа должно полностью соответствовать его названию. Содержание параграфов одной главы должно раскрывать главу в целом. Главы в целом, в совокупности, должны раскрывать тему курсовой работы.
5. Использование материалов научных работ, учебников и прочих материалов без оформления соответствующих ссылок не допускается и расценивается как ПЛАГИАТ, что выявляется в рамках проверки работы в системе «Антиплагиат». Цитирование в рамках работы оформляется в виде ссылок.
6. В работе соблюдается научный стиль изложения (см. пункт 2.2 Научный стиль изложения в курсовой работе).
7. Технические требования к тексту и печати:
   - работа печатается на стандартном листе бумаги формата А4,
   - положение основного текста – вертикальное, горизонтальное положение допускается для приведения примеров (например, скрины материалов по имитации открытых данных), организации карты данных, приложений (при необходимости);
   - абзацный отступ - 1,25 см.
   - поля - левое поле – 30 мм, правое –15 мм, верхнее и нижнее – 20 мм;
   - межстрочный интервал − основной текст и библиографический список 1,5, таблицы и рисунки – 1;
   - шрифт - Liberation Serif;
   - размер кегля − основной текст и библиографический список – 14 пт, таблицы, рисунки, формулы – 12 пт, в случае таблицы большого объема может быть применим кегль 10 пт.
   - цвет шрифта - черный, только рисунки или графики могут быть выполнены и в цветной гамме;
   - сплошная заливка запрещена;
   - выделение **полужирным,** *курсивом,* подчеркивания **НЕ ДОПУСКАЮТСЯ** в тексте и библиографическом списке;
   - выравнивание − основной текст и библиографический список - по ширине Заголовок – по центру Каждая новая глава начинается с новой страницы; это же правило относится к другим основным структурным частям работы (введению, заключению, библиографическому списку литературы, приложениям и т.д.);
   - нумерация страниц – сквозная, по всему тексту работы начиная с титульного печатается только со второго листа, оглавления (в центре или справа, в нижней части листа).

2.2 Научный стиль изложения в курсовой работе

Курсовая работа должна быть выдержана в научном стиле, как в изложении, так и в оформлении работы.

Научный стиль предполагает использование конструкций, исключающих употребление местоимений первого лица единственного и множественного числа, местоимений второго лица единственного числа. В данном случае предполагается использовать неопределенно-личные предложения, формы изложения от третьего лица. Разговорно-просторечная лексика недопустима.

Неправильно	Правильно
Я считаю…	Автор полагает…
Я сформулировал основные признаки…	В работе сформулированы основные признаки…

В научной работе крайне важно соблюдать целость, связность и логичность изложения. В этих целях рекомендуется использовать следующие слова и словосочетания:

Неправильно	Правильно
Отображение последовательности мыслей и рассуждений	
во 1-х, во 2-х, да и…, не то…, раз…, один, два	Во-первых, во-вторых, значит, итак, прежде всего…
Переход от одной мысли к другой	
ранее я выделил…, а теперь…	прежде чем перейти к…, обратимся к…, рассмотрим…, необходимо остановиться на…
Выделение противоречивости отношений к отдельным частям материала	
но почему – то, этот…, а тот…	однако, в то время как, тем не менее
Установление и отображение причинно-следственных связей	
во 1-х, во 2-х, один, два	следовательно, поэтому, благодаря этому, вследствие этого, кроме того, к тому же
Степень уверенности и источник сообщения	
может быть, говорят, я уверен…	конечно, разумеется, действительно, видимо, надо полагать, возможно, вероятно, по сообщению, по сведениям, по мнению, по данным
Выражение логической последовательности	
да и…, не то…	благодаря тому что, между тем как, так как, вместо того чтобы, ввиду того что, оттого что, вследствие того что, после того как, в то время как, в течение, в соответствии с, в результате, в отличие от, наряду с, в связи с, вследствие
Логические связи	
я думаю, размышляя о…, возможно	приведем результаты; как показал анализ; на основании полученных данных
Средства связи	
что-то, кое-что, какой-нибудь	данные, этот, такой, названные, указанные, перечисленные
Сравнение и превосходство	
повыше, побыстрее, поскорее, побольше	наиболее, наименее, предпочтительнее
Обобщение результатов	
я делаю вывод, я резюмирую, я понимаю…	итак, таким образом, значит, в заключение отметим, сказанное позволяет сделать вывод, подводя итог, резюмируя сказанное, отсюда следует, что

Подготовлено авторами с использованием следующей научной работы:
Пособие по русскому языку. Научный стиль речи. Оформление научной работы [Текст] : учеб. пособие / А. К. Демидова. - Москва: Русский язык, 1991.

2.3. Структура и оформление курсовой работы

Титульный лист

Образец титульного листа курсовой работы приведен в приложении В. При оформлении титульного листа подчеркивания в местах указания Ф.И.О. автора курсовой работы, Ф.И.О. руководителя, его ученой степени, должности заменяются текстом. Сами примечания (Ф.И.О., должность и т.д.) удаляются. На титульном листе не проставляется номер листа, но лист оглавления нумеруется как 2.

Содержание

Следует после титульного листе. Рекомендуется оформить в виде таблицы, сделав ее границы невидимыми. Шрифт - Liberation Serif, интервал - 1,5, Названия глав печатаются заглавными буквами, названия параграфов – строчными, начиная с заглавной.
Образец оформления содержания приведен в приложении Б.

Введение

Название параграфа **Введение**, также как и параграфов **Заключение**, **Содержание**, **Библиография**, пишутся строчными буквами, начиная с заглавной буквы и выравниваются по центру. После них также ставится абзацный отступ.

Введение должно содержать следующие компоненты, логически связанные друг с другом в нижеприведенной последовательности (пункты не нумеровать и не выделять жирным).

Актуальность. Обоснование актуальности темы предполагает несколько предложений об актуальности исследуемой проблемы. Например, если тема касается возможностей и практики Data Mining в медицине, необходимо сказать о значимости самой отрасли, медицины, затем перейти (кратко) к актуальности теоретических возможностей Data Mining, а затем – к важности и актуальности практического применения Data Mining в медицине. Обоснование актуальности 4-8 предложений.

Обзор публикаций по теме. Нужно указать авторов, исследователей, опубликованные работы которых связаны с тематикой курсовой работы, сгруппировав их по направленности. Например,
Проблема формирования культуры открытых данных отражена в работах А.В.Морозова, В.И. Петровой, К.В. Размышляева и др. Анализ практики противодействия формирования открытых данных представлены в исследованиях К.Л. Величко, Г.А. Ветвицкой, Д.У. Устинова, З.Ф.Алексеенко и др.

Объект исследования – это конкретно существующий объект, к которому относится курсовая работа, то, что исследуется и рассматривается в курсовой работе. Например, открытые данные.

Предмет исследования – отражает особенности явления или процесса, характеристик или признаков, аспектов исследования. Например, противодействие формированию открытых данных в РФ.

Цель курсовой работы – искомый результат работы в соответствии с темой.

Задачи исследования – порядок, этапы, выполнив которые можно достичь цели курсовой работы.

Характеристика методологического инструментария – какие методы (методы каких дисциплин, наук) использовались.

Характеристика эмпирической базы – источники данных, краткая характеристика источника (если нужно) и самой базы. Например, открытые данные российских сайтов.

Характеристика использованных программных средств – работа выполняется в LibreOffice.

Практическая значимость – где могут быть применены полученные результаты курсовой работы. Например, практическая значимость работы состоит в возможности использования материалов исследования в практике формирования открытых данных хозяйствующими субъектами, органами власти и управления.

Структура работы – из чего состоит, сколько страниц и т.д. Например, работа состоит из введения, двух глав, заключения, библиографического списка и трех приложений, общим объемом 32 страницы. Основной текст включает в себя 3 таблицы и 6 рисунков. Библиографический список состоит состоит из 24 источников.

Содержательная часть работы

Все темы курсовых работ предназначены для работы с открытыми данными.

В первой главе, как правило, излагаются основные теоретические положения по выбранной теме, рассматриваются основные категории изучаемого явления или процесса, приводятся определения, в том числе и нормативные, дается анализ исследований разных авторов, формулируются основные гипотезы исследования.

В зависимости от темы курсовой работы в первом параграфе рекомендуется рассмотреть следующие понятия: агрегированные данные, база данных, данные, машиночитаемые форматы данных, метаданные, набор данных, неструктурированная информация, обладатель данных (в контексте открытых данных), обезличенные данные, открытые данные, персональные данные, повторное использование, свободная (открытая) лицензия, сырые данные, человекочитаемые форматы данных, чувствительные данные.

Второй параграф может быть посвящен анализу нормативных ограничений и здесь могут рассматриваться ограничения в соответствии с законодательством, правил и периодичности публикации открытых данных, типов лицензий, работе с персональными данными.

Третий параграф может быть посвящен анализу технических ограничений на публикацию данных: требованиям к паспортизации, выбору лицензии, анонимизации данных, метаданными, форматам файлов/публикаций.

Во второй главе, проводится анализ сложившейся практики, выявляются особенности и проблемы в соответствии со сформулированной ранее гипотезой. В первом параграфе рекомендуется проанализировать основные источники открытых данных, в рамках тематики курсовой работы, составив карту данных. Необходимо разбирать назначение и кратко анализировать содержание каждого приведенного источника данных. Во втором параграфе следуют остановиться на проблемах, выявленных в исследовании, в

соответствии с тематикой работы. Глава завершается конкретными рекомендациями и изложением потенциальных направлений применения на практике полученных автором работы результатов и выводов.

Название глав и параграфов выделяется заголовками, напечатанными строчными буквами, начиная с заглавной, шрифтом Liberation Serif 14 с выделением полужирным, с выравнивание производится по центру. Название главы отделяется от названия параграфа абзацным отступом, название параграфа от текста – также. ***Например:***

2 Практика открытых данных в России

2.1 Карта данных: примеры публикации открытых данных

Использование ссылок. При оформлении работы необходимо обязательно ссылаться на первоисточник данных. Прямое заимствование текста, таблиц, графиков является плагиатом и недопустимо. Ссылки бывают внутритекстовые и подстрочные. Внутритекстовая ссылка – ссылка, помещаемая внутри основного текста издания. Внутритекстовая библиографическая ссылка содержит сведения об объекте ссылки, не включенные в текст документа, оформляется в квадратных скобках с указанием порядкового номера источника (согласно библиографии) и страницы, с которой взята цитата или другой материал, например:

Метаданные – данные, описывающие набор данных или документ. Например, даты его создания и обновления; автора; публикатора; контактную и лицензионную информацию [8, с. 15].

Подстрочная ссылка –ссылка, помещаемая в сноске (сноска – элемент аппарата издания, содержащий вспомогательный текст пояснительного или справочного характера). Помещается внизу страницы, после основного текста и связывается с текстом знаком сноски – соответствующим цифровым номером. Например: Цели и основные задачи развития искусственного интеллекта - повышение доступности и качества данных, необходимых для развития технологий искусственного интеллекта. [1]

Оформление иллюстративного материала. Графики, диаграммы, схемы (рамках карты данных принтскрины данных), рисунки, называются «рисунок» и нумеруются сквозным образом в пределах одной главы. Например, вне зависимости от параграфа все рисунки главы 2 будут пронумерованы Рисунок 2.1, Рисунок 2.2. В тесте следует описывать рисунок. Например: основные источники открытых данных в медицине изображены на рисунке 2.1.

Номер и название рисунка пишутся в одну строку, с выравниванием от центра. Обязательно отделение названия рисунка от последующего текста абзацным отступом. После рисунка обязательна трактовка его содержания. Нельзя размещать два и более рисунков подряд без комментария их содержимого.

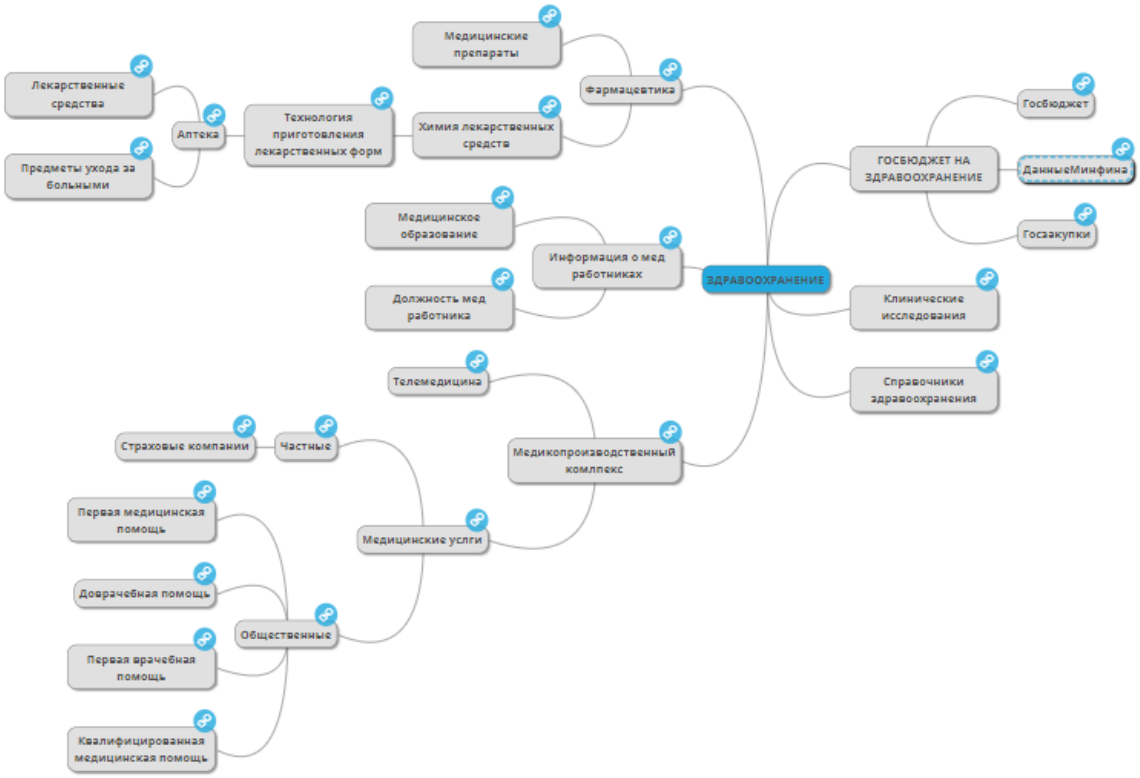


Рисунок 2.1 – Карта данных отрасли Здравоохранение.

[1] Российская Национальная стратегия развития искусственного интеллекта на период до 2030 года.

Оформление таблиц. В рамках работы приветствуется использование таблиц, позволяющих компактно проводить сравнения и излагать цифровой материал. Каждая таблица должна иметь подробное название (заглавие), где отражено ее основное содержание и нумероваться сквозным образом в пределах одной главы. Например, вне зависимости от параграфа все таблицы главы 2 будут пронумерованы Таблица 2.1, Таблица 2.2. Надпись Таблица 2.1 и заглавие располагают над таблицей с выравниванием по центру. Они отделяются друга тире.

Таблицу с большим количеством строк возможно перенести на другую страницу. В этом случае над продолжением с выравниванием по левому краю *курсивом* пишут «*Продолжение таблицы 2.1*»

Например:

Продолжение таблицы 2.1

Обязательно указывать источник данных приведенных в таблице, под ней, 10 шрифтом или в квадратных скобках, рядом с заглавием (в примере оба варианта, нужно выбрать один.

Объемные таблицы могут быть оформлены не 14, а 12 шрифтом, через одинарный интервал.

Таблица 1.1 – Принципы Хартии Открытых данных [12].

Наименование принципа	Краткое содержание принципа
Открыто по умолчанию	Правительственные данные должны быть открыты по умолчанию. Термин “правительственные данные” включает в себя данные национальных, региональных, местных и городских органов власти, международных правительственных органов и других типов учреждений в более широком государственном секторе, а также данные организаций предоставляющих услуги государству
Своевременность	Для получения данных необходимы как человеческий ресурс, так и время. Так же необходимо определить, какие данные следует приоритизировать для публикации. И ключевое эти данные должны быть всеобъемлющими и высококачественными.
Доступность	Публикации открытых данных должны быть легко обнаруживаемыми и доступными, а также предоставляться без бюрократических или административных барьеров, которые могут помешать людям получить доступ к данным.
Сопоставимые и интероперабельные	Данные обладают мультипликативным эффектом. Чем больше качественных наборов данных у вас есть и чем проще им общаться друг с другом, тем большую потенциальную ценность вы можете извлечь из них. Решающую роль в этом играют общепринятые стандарты данных.
Для улучшения управления и вовлечения граждан	Открытые данные позволяют гражданам (и другим представителям правительства) лучше понять, чем занимаются чиновники и политики. Эта прозрачность может улучшить государственные услуги и помочь привлечь правительства к ответственности.
За инклюзивное развитие и инновации	Открытые данные могут способствовать инклюзивному экономическому развитию. Например, более широкий доступ к данным может сделать сельское хозяйство более эффективным или его можно использовать для борьбы с изменением климата. Наконец, мы часто думаем об открытых данных просто как об улучшении работы правительства, но существует целая вселенная предпринимателей, зарабатывающих деньги на открытых данных.

Источник: Хартия открытых данных – URL:www.opendatacharter.net/principles

Материалы, изложенные в таблице, должны быть проанализированы. Пересказ недопустим.

Заключение

Содержит выводы по всей курсовой работе и практические рекомендации по избежанию ошибок в практике публикаций открытых данных, сделанные студентом по результатам проведенного анализа.

Библиографический список

В библиографический список следует включать источники, которые были использованы при подготовке текста работы и непосредственно в самом тексте работы (**не менее 15 источников помимо сайтов, включенных в карту данных**). При этом сайты, содержащие открытые данные и включенные в карту данных также включаются в список.

В состав библиографии включаются:

1) официальные документы (Конституция РФ, Кодексы, Федеральные законы РФ, Указы Президента РФ и так далее) даются в хронологическом порядке;

2) нормативные акты и инструктивный материал (Приказы, Положения, Указания, Методические рекомендации, Письма, Инструкции), материалы Росстата РФ даются в хронологическом порядке;

3) монографии, учебная литература, статьи из научных сборников и журналов, материалы периодической печати даются в алфавитном порядке;

4) адреса сети Internet.

Источники в библиографии располагаются в алфавитном порядке.

Общая схема библиографического описания отдельно изданного документа включает следующие обязательные элементы.

1. Сведения об авторе. Если у ресурса один, два или три автора, то в описании приводят первого автора. Сведения о всех остальных авторах помещают за косой чертой (/). Если авторов более трех, то документ описывают под названием, а фамилии авторов приводят за косой чертой. Для списков допускается сокращать количество приводимых сведений. В этом случае в сведениях об ответственности указывают имена одного, двух, трех или четырех авторов. При наличии информации о пяти и более авторов приводят имена первых трех и сокращение «[и др.]». Фамилия как правило, отделяется от имени (имен), имени и отчества, инициалов запятой (,).
2. Заглавие (название книги, указанное на титульном листе).
3. Сведения, относящиеся к заглавию (раскрывают тематику, вид, жанр, назначение документа и т. д.), пишутся после заглавие через двоеточие.
4. Сведения об ответственности (содержат информацию об авторах, составителях, редакторах, переводчиках и т. п.; об организациях, от имени которых опубликован документ). Сведения об ответственности приводят в описании в той форме, в какой они представлены в ресурсе, за косой чертой (/).
5. Сведения об издании (повторности, переработка, дополнения).
6. Место издания (название города, где издан документ). Города в списках не сокращают. Если на титульном листе указано два места издания и более, то они приводятся через точку с запятой или достаточно привести первое место издания. (например Москва ; Санкт-Петербург)
7. Издательство или издающая организация.
8. Год издания.
9. Объем (сведения о количестве страниц, листов).

При оформлении библиографического списка выдерживается ГОСТ Р 7.0.5 – 2018 Система стандартов по информации, библиотечному и издательскому делу. Библиографическая ссылка. Общие требования и правила составления.

Например:

1. Об информации, информационных технологиях и о защите информации: Федеральный закон от 27.07.2006 N 149-ФЗ// КонсультантПлюс:[сайт].– https://www.consultant.ru/document/cons_doc_LAW_61798/ (дата обращения: 03.04.2024).
2. Регионы России. Социально-экономические показатели. 2023: Стат. сб. / Росстат. – М., 2024.
3. Стефогло С.Н Концепция открытых данных, основные принципы открытых данных/ С.Н. Стефогло, Н.И. Томилова, А.Ж. Амиров // Инновационная наука. – 2015. – №4. –С. 129–137.
4. Чубукова И.А. Data Mining: учебное пособие – Москва; Саратов: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. - URL:http://www.iprbookshop.ru/89404.html (дата обращения 03.04.2024)
5. Правительство Российской Федерации: официальный сайт. – Москва. – URL: http://government.ru (дата обращения: 03.04.2024).

Приложения

Приложения следуют после библиографии как продолжение работы со сквозной нумерацией страниц и располагаются в порядке их упоминания в тексте. Приложения включают таблицы, графики, диаграммы, схемы и иные иллюстрации; вспомогательные и объемные материалы к основной текстовой части работы.

- Если в работе больше одного приложения, то после Библиографического списка вставляют лист с заголовком **Приложения**, строчными буквами, начиная с заглавной с выделением полужирным шрифтом Liberation Serif (размер шрифта выбирается ориентировочно в границах 24-27). Выравнивание производится по центру.
- Вне зависимости от размера приложения, каждое из них располагается на отдельном листе.
- Чтобы содержимое приложений не сместилось при форматировании курсовой работы, их рекомендуется разделять с помощью функции «Разрыв страницы».

- Если приложение большое и не помещается на одной странице, разрешается переместить его часть на следующий лист. В этом случае, на следующей странице вверху указывается «Продолжение Приложения» и проставляется его порядковая буква (А, Б, С и т.д.).
- Если приложений несколько, то их обозначение будет состоять из буквы, например Приложение А, Приложение Б.
- Название приложения располагается в следующей за обозначением строке и выравниванием по центру.
- Внутри текста работы ссылки на материалы приложения обязательны. Например, Карта данных отрасли медицина приведена в приложении А.

Приложение А

Форма заявления на выполнение курсовой работы

Заведующему кафедрой
Статистики, эконометрики и оценки рисков
д.э.н., проф. Ниворожкиной Л.И.
обучающегося группы _____
_____
_____
ФИО обучающегося полностью

ЗАЯВЛЕНИЕ

Прошу разрешить мне выполнение курсовой работы по дисциплине «Введение в Data Mining» на тему: «_____
_____
_____
_____».

«____» _____ 20__г. _____
подпись обучающегося

"СОГЛАСОВАНО"
Научный руководитель
курсовой работы,
д.э.н., профессор Л.И. Ниворожкина

Зав. кафедрой СЭиОР,
д.э.н., профессор Л.И. Ниворожкина

Приложение Б
Форма титульного листа курсовой работы

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

РОСТОВСКИЙ ГОСУДАРСТВЕННЫЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ (РИНХ)

Учетно-экономический факультет

Кафедра статистики, эконометрики и оценки рисков

ДОПУСТИТЬ К ЗАЩИТЕ
Зав. кафедрой _____
д.э.н., профессор Ниворожкина Л.И.
«_____» _____ 20__ г.

КУРСОВАЯ РАБОТА
по дисциплине «Введение в Data mining»
на тему:
«ТЕМА КУРСОВОЙ РАБОТЫ»

Выполнил
студент группы _____ _____ *И.О.Фамилия*
Направление 01.03.05 Статистика
Направленность 01.03.05.01 Анализ больших данных

Руководитель курсовой работы
ученая степень, звание, должность _____ *И.О.Фамилия*

Ростов-на-Дону, 20__

Приложение В
Образец оформления содержания курсовой работы

Содержание

ВВЕДЕНИЕ…… 3
1. РАБОТА С ОТКРЫТЫМИ ДАННЫМИ: ОСОБЕННОСТИ ПУБЛИКАЦИЙ И ПРАВОВЫЕ ОГРАНИЧЕНИЯ
1.1. Основные понятия и определения Data Mining для открытых данных…… 5
1.2. Нормативные ограничения по публикации открытых данных …….. 9
1.3. Технические аспекты публикации открытых данных …… 14
2. ПРАКТИКА ОТКРЫТЫХ ДАННЫХ В РОССИИ
2.1. Карта данных: примеры публикации открытых данных…… 18
2.2. Проблемы и препятствия в публикации открытых данных …… 22
ЗАКЛЮЧЕНИЕ…… 27
БИБЛИОГРАФИЧЕСКИЙ СПИСОК…… 29
ПРИЛОЖЕНИЯ …..……..... 31