



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Магнитогорский государственный технический университет им. Г.И.  
Носова»



УТВЕРЖДАЮ  
Директор ИЭиАС  
В.Р. Храмшин

03.02.2026 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

***ТЕХНОЛОГИИ DATA MINING И BIG DATA***

Направление подготовки (специальность)  
09.03.01 Информатика и вычислительная техника

Направленность (профиль/специализация) программы  
Программное обеспечение средств вычислительной техники и автоматизированных систем

Уровень высшего образования - бакалавриат

Форма обучения  
очная

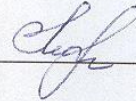
Институт/ факультет	Институт энергетики и автоматизированных систем
Кафедра	Вычислительной техники и программирования
Курс	4
Семестр	8

Магнитогорск  
2026 год

Рабочая программа составлена на основе ФГОС ВО - бакалавриат по направлению подготовки 09.03.01 Информатика и вычислительная техника (приказ Минобрнауки России от 19.09.2017 г. № 929)

Рабочая программа рассмотрена и одобрена на заседании кафедры  
Вычислительной техники и программирования  
29.01.2026, протокол № 7

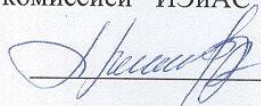
Зав. кафедрой



О.С. Логунова

Рабочая программа одобрена методической комиссией ИЭиАС  
03.02.2026 г. протокол № 5

Председатель



В.Р. Храмшин

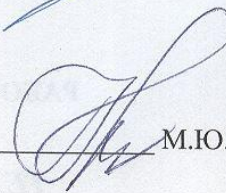
Рабочая программа составлена:  
ст. преподаватель кафедры кафедры ВТиП,



М.В.Зарецкий

Рецензент:

Директор НИИ "Промбезопасность", д-р техн. наук



М.Ю.Наркевич

## Лист актуализации рабочей программы

---

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2027 - 2028 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от \_\_\_\_\_ 20\_\_ г. № \_\_\_\_  
Зав. кафедрой \_\_\_\_\_ О.С. Логунова

---

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2028 - 2029 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от \_\_\_\_\_ 20\_\_ г. № \_\_\_\_  
Зав. кафедрой \_\_\_\_\_ О.С. Логунова

---

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2029 - 2030 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от \_\_\_\_\_ 20\_\_ г. № \_\_\_\_  
Зав. кафедрой \_\_\_\_\_ О.С. Логунова

---

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2030 - 2031 учебном году на заседании кафедры Вычислительной техники и программирования

Протокол от \_\_\_\_\_ 20\_\_ г. № \_\_\_\_  
Зав. кафедрой \_\_\_\_\_ О.С. Логунова

### 1 Цели освоения дисциплины (модуля)

Цель освоения дисциплины "Технологии Data Mining и Big Data":

- формирование у студентов представления о типах задач, возникающих в области интеллектуального анализа данных (Технологии Data Mining и Big Data);
- освоение основных подходов, применяемых при решении задач Data Mining и Big Data;
- освоение современных программных средств, применяемых при решении задач Data Mining и Big Data;
- получение навыков применения парадигм Data Mining и Big Data при решении задач в различных предметных областях.

### 2 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина Технологии Data Mining и Big Data входит в часть учебного плана формируемую участниками образовательных отношений образовательной программы.

Для изучения дисциплины необходимы знания (умения, владения), сформированные в результате изучения дисциплин/ практик:

Методы управления знаниями

Обработка экспериментальных данных на ЭВМ

Базы данных OLTP-систем

Базы и хранилища данных

Программные решения для бизнеса

Моделирование

Функциональное программирование

Объектно-ориентированное программирование

Философия

Прикладная математика

Программирование

Численные методы

Элементы линейной алгебры

Знания (умения, владения), полученные при изучении данной дисциплины будут необходимы для изучения дисциплин/практик:

Выполнение и защита выпускной квалификационной работы

### 3 Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля) и планируемые результаты обучения

В результате освоения дисциплины (модуля) «Технологии Data Mining и Big Data» обучающийся должен обладать следующими компетенциями:

Код индикатора	Индикатор достижения компетенции
ПК-6	Способность к формализации и алгоритмизации поставленных задач, к написанию программного кода с использованием языков программирования, определения и манипулирования данными и оформлению программного кода в соответствии установленными требованиями
ПК-6.1	Оценивает качество математической модели при формализации задачи предметной области
ПК-6.2	Оценивает качество разработанных алгоритмов для последующего кодирования
ПК-6.3	Оценивает выбор программных средств для программирования и манипулирования данными в соответствии установленными требованиями

#### 4. Структура, объём и содержание дисциплины (модуля)

Общая трудоемкость дисциплины составляет 3 зачетных единиц 108 академических часов, в том числе:

- контактная работа – 57,3 академических часов;
- аудиторная – 56 академических часов;
- внеаудиторная – 1,3 академических часов;
- самостоятельная работа – 50,7 академических часов;
- в форме практической подготовки – 0 академических часов;

Форма аттестации - зачет с оценкой

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в академических часах)			Самостоятельная работа студента	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код компетенции
		Лек.	лаб. зан.	практ. зан.				
1. Концептуальные основы. Программный инструментарий.								
1.1 Данные, информация, знания.	8	2	2		2	Самостоятельное изучение учебной и научной литературы.	Беседа – обсуждение. Устный опрос	ПК-6.1, ПК-6.2
1.2 Основы языка R. Среда RStudio (RStudio Cloud). Хранилище CRAN и работа с ним.		4	4		6	Самостоятельное изучение учебной и научной литературы. Подготовка к лабораторному занятию. Выполнение лабораторной работы.	Беседа – обсуждение. Анализ программного кода. Устный опрос.	ПК-6.1, ПК-6.2
Итого по разделу		6	6		8			
2. Предварительная обработка данных. Проверка гипотез. Кластеризация.								
2.1 Предварительная обработка данных. Преобразование Raw Data в Tidy Data. Анализ выбросов.	8	4	6		10	Самостоятельное изучение учебной и научной литературы. Подготовка к лабораторному занятию. Выполнение лабораторной работы.	Беседа – обсуждение. Анализ программного кода. Устный опрос.	ПК-6.1
2.2 Проверка статистической гипотезы		4	6		10	Самостоятельное изучение	Беседа – обсуждение.	ПК-6.1

о параметрах генеральной совокупности. Проверка статистической гипотезы о законе распределения. Кластеризация.						учебной и научной литературы. Подготовка к лабораторному занятию. Выполнение лабораторной работы.	Анализ программного кода. Устный опрос.	
Итого по разделу		8	12		20			
3. Построение статистических зависимостей. Анализ и прогнозирование временных рядов. Обработка текстовой информации.								
3.1 Построение статистических зависимостей. Анализ временных рядов. Нахождение тренда.	8	4	8		10	Самостоятельное изучение учебной и научной литературы. Подготовка к лабораторному занятию. Выполнение лабораторной работы.	Беседа – обсуждение. Анализ программного кода. Устный опрос	ПК-6.3
3.2 Обработка "сырого" текста. Разметка по частям речи. Лемматизация и стеммирование. Построение корпусов текстов. Выявление именованных сущностей.		6	6		12,7	Самостоятельное изучение учебной и научной литературы. Подготовка к лабораторному занятию. Выполнение лабораторной работы.	Беседа – обсуждение. Анализ программного кода. Устный опрос.	ПК-6.3
Итого по разделу		10	14		22,7			
4. Закрепление изученного материала								
4.1 Закрепление изученного материала.	8					Изучение современных программных реализаций.	Критическое рассмотрение применения методов Data Mining и Big Data в реальных задачах.	ПК-6.1, ПК-6.2, ПК-6.3
Итого по разделу								
Итого за семестр		24	32		50,7		зао	
Итого по дисциплине		24	32		50,7		зачет с оценкой	

## **5 Образовательные технологии**

1. Традиционные образовательные технологии ориентируются на организацию образовательного процесса, предполагающую прямую трансляцию знаний от преподавателя к студенту (преимущественно на основе объяснительно-иллюстративных методов обучения). Учебная деятельность студента носит в таких условиях, как правило, репродуктивный характер.

Формы учебных занятий с использованием традиционных технологий:

Информационная лекция – последовательное изложение материала в дисциплинарной логике, осуществляемое преимущественно вербальными средствами (монолог преподавателя).

Семинар – беседа преподавателя и студентов, обсуждение заранее подготовленных сообщений по каждому вопросу плана занятия с единым для всех перечнем рекомендуемой обязательной и дополнительной литературы.

Практическое занятие, посвященное освоению конкретных умений и навыков по предложенному алгоритму.

Лабораторная работа – организация учебной работы с реальными материальными и информационными объектами, экспериментальная работа с аналоговыми моделями реальных объектов.

2. Технологии проблемного обучения – организация образовательного процесса, которая предполагает постановку проблемных вопросов, создание учебных проблемных ситуаций для стимулирования активной познавательной деятельности студентов.

3. Интерактивные технологии – организация образовательного процесса, которая предполагает активное и нелинейное взаимодействие всех участников, достижение на этой основе лично значимого для них образовательного результата. Наряду со специализированными технологиями такого рода принцип интерактивности прослеживается в большинстве современных образовательных технологий. Интерактивность подразумевает субъект - субъектные отношения в ходе образовательного процесса и, как следствие, формирование саморазвивающейся информационно-ресурсной среды.

Формы учебных занятий с использованием специализированных интерактивных технологий:

Лекция «обратной связи» – лекция–провокация (изложение материала с заранее запланированными ошибками), лекция-беседа, лекция-дискуссия, лекция–пресс-конференция.

4. Информационно-коммуникационные образовательные технологии – организация образовательного процесса, основанная на применении специализированных программных сред и технических средств работы с информацией.

## **6 Учебно-методическое обеспечение самостоятельной работы обучающихся**

Представлено в приложении 1.

## **7 Оценочные средства для проведения промежуточной аттестации**

Представлены в приложении 2.

## **8 Учебно-методическое и информационное обеспечение дисциплины**

### **а) Основная литература:**

1. Митяков, Е. С. Искусственный интеллект и машинное обучение : учебное пособие для вузов / Е. С. Митяков, А. Г. Шмелева, А. И. Ладынин. — 2-е изд., стер. —

Санкт-Петербург : Лань, 2026. — 252 с. — ISBN 978-5-507-51198-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/507451> (дата обращения: 08.03.2026). — Режим доступа: для авториз. пользователей.

2. Мэтлофф, Н. Искусство программирования на R. Погружение в большие данные : практическое руководство / Н. Мэтлофф. - Санкт-Петербург : Питер, 2019. - 416 с. - (Серия «Библиотека программиста»). - ISBN 978-5-4461-1101-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1733504> (дата обращения: 03.05.2023). – Режим доступа: по подписке.

#### **б) Дополнительная литература:**

1.

Дейтел, П. Python: Искусственный интеллект, большие данные и облачные вычисления : практическое руководство / П. Дейтел, Х. Дейтел. - Санкт-Петербург : Питер, 2020. - 864 с. - (Серия «Для профессионалов»). - ISBN 978-5-4461-1432-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1733685> (дата обращения: 03.05.2023). – Режим доступа: по подписке.

2. Зарова, Е. В. Методы Datamining в обработке и анализе статистических данных (решения в R) : монография / Е.В. Зарова. — Москва : ИНФРА-М, 2021. — 232 с. : ил. - ISBN 978-5-16-016814-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1240276> (дата обращения: 03.05.2023). – Режим доступа: по подписке.

#### **в) Методические указания:**

1. Станкевич, Л. А. Интеллектуальные системы и технологии : учебник и практикум для вузов / Л. А. Станкевич. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 495 с. — (Высшее образование). — ISBN 978-5-534-16238-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/530657> (дата обращения: 03.05.2023).

#### **г) Программное обеспечение и Интернет-ресурсы:**

##### **Программное обеспечение**

Наименование ПО	№ договора	Срок действия лицензии
Deductor Studio Academic	Соглашение о сотрудничестве №06-2901\08 от 29.01.2008	бессрочно
Anaconda Python	свободно распространяемое ПО	бессрочно
Scilab Computation Engine	свободно распространяемое ПО	бессрочно
MathWorks MathLab v.2014 Classroom License	К-89-14 от 08.12.2014	бессрочно
NotePad++	свободно распространяемое ПО	бессрочно

##### **Профессиональные базы данных и информационные справочные системы**

Название курса	Ссылка
Национальная информационно-аналитическая система – Российский индекс научного цитирования (РИНЦ)	URL: <a href="https://elibrary.ru/project_risc.asp">https://elibrary.ru/project_risc.asp</a>

## **9 Материально-техническое обеспечение дисциплины (модуля)**

Материально-техническое обеспечение дисциплины включает:

Лекционная аудитория ауд. 282 – Мультимедийные средства хранения, передачи и представления информации;

Компьютерные классы Центра информационных технологий ФГБОУ ВПО «МГТУ им. Г.И. Носова» – Персональные компьютеры, объединенные в локальные сети с выходом в Internet, оснащенные современными программно-методическими комплексами для решения задач в области информатики и вычислительной техники;

Аудитории для самостоятельной работы: компьютерные классы; читальные залы библиотеки – ауд. 282 и классы УИТ и АСУ;

Помещения для самостоятельной работы обучающихся, оснащенных компьютерной техникой с возможностью подключения к сети «Интернет» и наличием доступа в электронную информационно-образовательную среду организации – классы УИТ и АСУ;

Помещения для хранения и профилактического обслуживания учебного оборудования – Центр информационных технологий – ауд. 372

### ***Приложение 1***

#### ***Учебно-методическое обеспечение самостоятельной работы обучающихся***

В течение семестра каждый студент выполняет лабораторные работы.

Лабораторная работа №1. Данные информация, знания.

Задание 1 (пороговый уровень). Элементарная работа с данными.

1.1. Возьмите в открытом доступе текстовый файл в формате docx. Текст должен содержать диакритические знаки. Преобразуйте файл в представление pdf. Оцените изменение объема файла.

1.2. Сгенерируйте с помощью свободно доступной нейросети (например, GigaChat) растровое цветное изображение животного или растения (при наличии художественных способностей можно нарисовать самостоятельно). Разместите его в файле (jpeg, png). Оцените объем полученного файла. Вставьте файл с изображением в текстовый документ.

Задание 2 (пороговый уровень). Элементарная обработка данных.

2.1 Подсчитайте количество слов в текстовом файле.

2.2 Подсчитайте количество стоп-слов в текстовом файле.

Задание 3(пороговый уровень) Извлечение информации.

3.1. Выделите в цветном изображении составляющие, соответствующие основным цветам.

3.2. Выведите эти изображения.

Задание 4 (средний уровень)

4.1. С помощью программ пакета cv2 преобразуйте изображение из цветного в монохромное представление

4.2. С помощью программ из пакета cv2 выделите контуры на монохромном изображении.

Задание 5 (высокий уровень)

5.1. Запрограммируйте поиск наиболее часто встречающихся в текстовом файле слов.

5.2. Запрограммируйте те же действия, исключив из рассмотрения стоп-слова.

5.3. Сгенерируйте аналогичный программный код с помощью нейросетей GigaChat и Deepseek

5.4. Сравните полученный каждым из способов программный код по критериям: правильность, удобочитаемость, приспособленность к доработке.

Лабораторная работа №2.

Задание 1 (пороговый уровень) Основы языка программирования R. Перед выполнением работы студент должен создать учетную запись в среде POSIT Cloud и получить для работы свободно распространяемый файл в формате csv.

- 1.1. Выполните средствами языка R арифметические действия над числовыми переменными (сложение, вычитание, умножение, деление).
- 1.2. Выполните средствами языка R действия над строковыми переменными (конкатенация, извлечение подстроки).

Задание 2 (пороговый уровень)

- 2.1. Составьте из однотипных данных (числовых, строковых) не менее трех массивов.
- 2.2. Выполните над массивами операции, допустимые типом данных, из которых составлены массивы.

Задание 3 (средний уровень)

- 3.1. Составьте из разнотипных данных не менее трех списков.
- 3.2. Выполните над списками действия, допустимые для их структуры.
- 3.3. Выясните, допустимы ли поэлементные операции для массивов и списков.

Задание 4 (высокий уровень)

- 4.1. Загрузите выданный вам файл в формате csv.
- 4.2. Внесите данные из этого файла в data.frame.
- 4.3. Внесите данные из этого файла в tibble.
- 4.4. Сформулируйте сходства и различия data.frame и tibble.

Лабораторная работа №3. Предварительная обработка данных. Преобразование Raw Data в Tidy Data. Анализ выбросов.

Задание 1 (пороговый уровень)

- 1.1. Загрузите представленный набор данных в data.frame или tibble.
- 1.2. Загрузите представленный набор данных в data.frame или tibble.
- 1.3. Проанализируйте строки таблицы с отсутствующими данными.
- 1.4. Определите способ работы с ними.
- 1.5.
- 1.6.

Задание 2 (пороговый уровень)

- 2.1. Получите средствами R и Python основные характеристики для числовых и категориальных данных.
- 2.2. Получите робастные оценки центрального положения и разброса
- 2.2. Создайте из нескольких атомов с помощью рекуррентного применения функции **cons** многоэлементный список.

Задание 3 (пороговый уровень)

- 3.1. Выполните средствами графических библиотек Python и R графический анализ данных.
- 3.2. Внимательно проанализируйте построенные «ящики с усами» (boxplots). Особое внимание уделите данным, не попавшим в «ящики»

Задание 4 (средний уровень)

- 4.1. Для пар числовых данных постройте диаграммы рассеяния. Визуально оцените наличие зависимости (это не обоснование, а только предположение!).
- 4.2. Оцените наличие корреляции между данными, представленными в столбце. Определите, какой из коэффициентов корреляции целесообразно использовать.
- 4.3. Проверьте гипотезу о незначимости (значимости) вычисленного коэффициента корреляции.

Задание 5 (высокий уровень)

- 5.1. Составьте на языке R программу для выполнения разведочного анализа данных.
- 5.2. Составьте на языке Python программу для выполнения разведочного анализа данных.
- 5.2. Сгенерируйте аналогичные программы на языках Python и R с помощью нейросетей GigaChat и Deepseek.
- 5.3. Сравните полученный каждым из способов программный код по критериям: правильность, удобочитаемость, приспособленность к доработке.

Лабораторная работа №4. Проверка статистической гипотезы о параметрах генеральной совокупности. Проверка статистической гипотезы о законе распределения. Кластеризация. Проверку статистической гипотезы следует выполнять при уровне значимости 0.95

*Задание 1(пороговый уровень)*

- 1.1. Проверьте статистическую гипотезу о равенстве математического ожидания заданной величине.
- 1.2. Проверьте статистическую гипотезу о законе распределения для выборки по критерию Колмогорова-Смирнова.

*Задание 2(пороговый уровень)*

- 2.1. Выполните кластеризацию данных по методу К средних (K\_Means).
- 2.2. Выполните иерархическую кластеризацию тех же данных.

*Задание 3 (средний уровень)*

- 3.1. Проверьте гипотезу о наличии различий между двумя выборками методами однофакторного дисперсионного анализа (ANOVA)
- 3.2. Проверьте гипотезу о наличии различий между двумя выборками методами двухфакторного дисперсионного анализа (ANOVA).

*Задание 5 (высокий уровень)*

- 5.1. Разработайте программу на языке R, предназначенную для решения вышеперечисленных задач
- 5.2. Разработайте аналогичную программу на языке Python.
- 5.3. Сгенерируйте аналогичные программы с помощью нейросетей GigaChat и Deepseek.
- 5.4. Сравните полученный каждым из способов программный код по критериям: правильность, удобочитаемость, приспособленность к доработке.

Лабораторная работа №5. Построение статистических зависимостей. LASSO и Ridge-регрессия.

*Задание 1(пороговый уровень)*

- 1.1. Для двух числовых переменных постройте линейную зависимость. Оцените результат.
- 1.2. Для двух числовых переменных постройте полиномиальную зависимость. Оцените результат.

*Задание 2(пороговый уровень)*

- 2.1. Постройте линейную зависимость одной числовой переменной от нескольких числовых переменных. Оцените результат.
- 2.2. Постройте полиномиальную зависимость одной числовой переменной от нескольких числовых переменных. Оцените результат.

*Задание 3 (средний уровень)*

- 3.1. Постройте линейную зависимость одной числовой переменной от нескольких числовых переменных с помощью перцептрона. Оцените результат.
- 3.2. Постройте полиномиальную зависимость одной числовой переменной от нескольких числовых переменных с помощью перцептрона. Оцените результат.

*Задание 4(средний уровень)*

- 4.1. Постройте полиномиальную зависимость одной числовой переменной от нескольких числовых переменных методами LASSO-регрессии.
- 4.2. Постройте полиномиальную зависимость одной числовой переменной от нескольких числовых переменных методами Ridge-регрессии.

*Задание 5 (высокий уровень)*

- 5.1. Разработайте программу на языке R, предназначенную для решения вышеперечисленных задач.
- 5.2. Разработайте аналогичную программу на языке Python.
- 5.3. Сгенерируйте аналогичные программы с помощью нейросетей GigaChat и Deepseek.
- 5.4. Сравните полученный каждым из способов программный код по критериям: правильность, удобочитаемость, приспособленность к доработке.

Лабораторная работа №6. Обработка «сырого» текста. Разметка по частям речи. Лемматизация и стемминирование. Выявление именованных сущностей.

Задание 1(пороговый уровень)

- 1.1. В одном из текстов, входящих в состав корпуса gutenber, вычислить частоты появления заданных слов.
- 1.2. Ту же работу провести с сырым текстом.

Задание 2(пороговый уровень)

- 2.1. Для заданных слов определите близость по метрике Левенштейна
- 2.2. Для заданных слов определите семантическую близость

Задание3(средний уровень)

- 3.1. Выполните лемматизацию заданного текста.
- 3.2. Выполните стеммирование заданного текста.

Задание4(средний уровень)

- 4.1. Выявите именованные сущности в тексте на русском языке.
- 4.2. Выявите именованные сущности в тексте на английском языке

Задание 5 (высокий уровень)

- 5.1. Разработайте программу на языке Python, предназначенную для решения вышеперечисленных задач.
- 5.2. Сгенерируйте аналогичные программы с помощью нейросетей GigaChat и Deepseek.
- 5.3. Сравните полученный каждым из способов программный код по критериям: правильность, удобочитаемость, приспособленность к доработке.

## Приложение 2. Оценочные средства для проведения промежуточной аттестации

а) Планируемые результаты обучения и оценочные средства для проведения промежуточной аттестации:

Код индикатора	Индикатор достижения компетенции	Оценочные средства
ПК-6: Способность к формализации и алгоритмизации поставленных задач, к написанию программного кода с использованием языков программирования, определения и манипулирования данными и оформлению программного кода в соответствии установленными требованиями		
ПК-6.1	Оценивает качество математической модели при формализации задачи предметной области	<p><i>Перечень теоретических вопросов</i></p> <ol style="list-style-type: none"> <li>1. Данные и информация. Концепция Data Mining.</li> <li>2. Специфика больших данных.</li> <li>3. Концепция “Data Driven Organization</li> <li>4. Применение концепций функционального программирования при обработке больших данных.</li> <li>5. Концепция качества данных.</li> <li>6. Числовые и категориальные данные.</li> <li>7. Графические данные.</li> <li>8. Видеоданные.</li> <li>9. Текстовые данные.</li> <li>10. Принципы разведочного анализа данных.</li> <li>11. Робастные оценки центрального положения данных.</li> <li>12. Робастные оценки вариабельности данных.</li> </ol>

Код индикатора	Индикатор достижения компетенции	Оценочные средства
		<p>13. Корреляция по Пирсону и ее применение в работе с большими данными.</p> <p>14. Корреляция по Спирмену и ее применение в работе с большими данными.</p> <p><i>Задания на решение задач из профессиональной области, комплексные задания</i></p> <p>Задание 1.</p> <p>1.1. Получите робастную оценку центрального положения заданного набора данных.</p> <p>1.2. Получите робастную оценку вариабельности для заданного набора данных.</p> <p>Задание 2.</p> <p>2.1. Разработайте программу для нахождения корреляции по Пирсону с использованием функций Python или R. Разработайте аналогичную программу с помощью ИИ-систем. Сопоставьте результаты.</p> <p>2.2. Разработайте программу для нахождения корреляции по Спирмену с использованием функций Python или R. Разработайте аналогичную программу с помощью ИИ-систем. Сопоставьте результаты.</p>
ПК-6.2	Оценивает качество разработанных алгоритмов для последующего кодирования	<p><i>Перечень теоретических вопросов</i></p> <ol style="list-style-type: none"> <li>1. Распределение данных и его анализ.</li> <li>2. Бутстреп в анализе данных.</li> <li>3. Визуализация данных числовых данных</li> <li>4. Визуализация категориальных данных.</li> <li>5. «Опрятные» данные.</li> <li>6. Методы превращения сырых данных в опрятные.</li> <li>7. Оценки качества моделей данных.</li> <li>8. Ресэмплинг при оценке качества модели.</li> <li>9. Классификация.</li> <li>10. Кластеризация.</li> <li>11. Работа с текстовыми данными.</li> <li>12. Корпусы текстов и их применение.</li> <li>13. Выявление именованных сущностей в тексте.</li> <li>14. Семантические метрики.</li> </ol> <p>Задание 1.</p> <p>1.1. Визуализируйте несколько набор данных.</p> <p>1.2. С помощью графического средства «ящик с усами» выявите медиану, 25% квартиль, 75% квартиль, подозрительные данные.</p> <p>Задание 2.</p> <p>2.1. Разработайте программу с использованием функций Python или R для построения «ящика с усами».</p>

Код индикатора	Индикатор достижения компетенции	Оценочные средства
		2.2. Разработайте программу с использованием функций Python для вычисления значения метрики Левенштейна..
ПК-6.3	Оценивает выбор программных средств для программирования и манипулирования данными в соответствии установленными требованиями	<p><i>Перечень теоретических вопросов</i></p> <ol style="list-style-type: none"> <li>1. Принципы организации программного средства PySpark.</li> <li>2. Организация таблицы данных (DataFrame) в PySpark.</li> <li>3. Методы работы с недостающими данными в PySpark.</li> <li>4. Методы группировки данных в PySpark.</li> <li>5. Методы построения эмпирических зависимостей в PySpark.</li> <li>6. Методы кластеризации в PySpark.</li> <li>7. Методы построения решений в PySpark.</li> <li>8. Выявление аномалий в PySpark.</li> <li>9. Графический анализ данных в PySpark</li> <li>10. Фильтрация данных в PySpark.</li> <li>11. Проверка статистических гипотез в PySpark.</li> </ol> <p>Задание 1.</p> <p>1.1. Выполните фильтрацию данных в PySpark.</p> <p>1.2. Выполните группировку данных в PySpark.</p> <p>Задание 2.</p> <p>2.1. Разработайте программу для работы с недостающими данными средствами PySpark..</p> <p>2.2. Разработайте программу для фильтрации данных методами PySpark..</p>

**б) Порядок проведения промежуточной аттестации, показатели и критерии оценивания:**

Промежуточная аттестация по дисциплине «Технологии Data Mining и Big Data» включает теоретические вопросы, позволяющие оценить уровень усвоения обучающимися знаний, и практические задания, выявляющие степень сформированности умений и владений, проводится в форме в форме зачета с оценкой.

Зачет с оценкой по дисциплине проводится по результатам отчетности за выполненные самостоятельные работы с опросом в устной форме по этапам выполнения в беседе-обсуждении на лекционных занятиях.

**Критерии оценки**

– на оценку **«отлично»** – полно раскрыто содержание материала; чётко и правильно даны определения и раскрыто содержание материала; ответ самостоятельный, при ответе использованы знания, приобретённые ранее;

– на оценку **«хорошо»** – раскрыто основное содержание материала в объёме; в основном правильно даны определения, понятия; материал изложен неполно, при ответе допущены неточности, нарушена последовательность изложения; допущены небольшие неточности при выводах и использовании терминов; практические навыки нетвёрдые;

– на оценку **«удовлетворительно»** – усвоено основное содержание материала, но изложено фрагментарно, не всегда последовательно; определения и понятия даны не чётко; практические навыки слабые;

– на оценку **«неудовлетворительно»** – основное содержание учебного материала не раскрыто; не даны ответы на дополнительные вопросы преподавателя