

Документ подписан простой электронной подписью.  
Информация о владельце:

ФИО: Кандрашина Елена Александровна

Должность: Врио ректора ФГАОУ ВО «Самарский государственный экономический университет»

Дата подписания: 13.07.2023 09:39:56

Уникальный программный ключ:

b2fd765521f4c570b8c6e8e502a10b4f1de8ae0d

**Министерство науки и высшего образования Российской Федерации**  
**Федеральное государственное автономное образовательное учреждение**  
**высшего образования**  
**«Самарский государственный экономический университет»**

**Институт**      Институт экономики предприятий

**Кафедра**      Прикладной информатики

**УТВЕРЖДЕНО**

Ученым советом Университета

(протокол № 11 от 30 мая 2023 г.)

**РАБОЧАЯ ПРОГРАММА**

**Наименование дисциплины**      Б1.В.04 Машинное обучение на больших данных

**Основная профессиональная образовательная программа**      09.04.03 Прикладная информатика программа  
Искусственный интеллект и большие данные

Квалификация (степень) выпускника магистр

Самара 2023

## Содержание (рабочая программа)

Стр.

- 1 Место дисциплины в структуре ОП
- 2 Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе
- 3 Объем и виды учебной работы
- 4 Содержание дисциплины
- 5 Материально-техническое и учебно-методическое обеспечение дисциплины
- 6 Фонд оценочных средств по дисциплине

Целью изучения дисциплины является формирование результатов обучения, обеспечивающих достижение планируемых результатов освоения образовательной программы.

## 1. Место дисциплины в структуре ОП

Дисциплина Машинное обучение на больших данных входит в часть, формируемая участниками образовательных отношений блока Б1. Дисциплины (модули)

Предшествующие дисциплины по связям компетенций: Системы обработки и анализа больших массивов данных, Инструменты анализа данных и машинного обучения, Процесс, стадии и методологии разработки решений на основе искусственного интеллекта, Алгоритмы и технологии анализа сложных сетей, Агентно-ориентированное моделирование, Методы машинного обучения, Методы оптимизации, Интеллектуальное планирование, Качество данных, подходы и инструменты, Уровни предоставления данных

Последующие дисциплины по связям компетенций: Тестирование искусственного интеллекта, Современные методы проектирования систем искусственного интеллекта

## 2. Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе

Изучение дисциплины Машинное обучение на больших данных в образовательной программе направлено на формирование у обучающихся следующих компетенций:

### Универсальные компетенции (УК):

УК-3 - Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
УК-3	УК-3.1: Знать:	УК-3.2: Уметь:	УК-3.3: Владеть (иметь навыки):
	особенности управления и руководства работой команды, выработки командной стратегии для достижения поставленной цели	организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	навыками управления и руководства работой команды, выработки командной стратегии для достижения поставленной цели

### Профессиональные компетенции (ПК):

ПК-2 - Способен разрабатывать правила использования репозитория проекта

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
ПК-2	ПК-2.1: Знать:	ПК-2.2: Уметь:	ПК-2.3: Владеть (иметь навыки):
	особенности разработки правил использования репозитория проекта	разрабатывать правила использования репозитория проекта	навыками разработки правил использования репозитория проекта

ПК-6 - Способен инициировать запросы на изменение (в том числе корректирующие действия, предупреждающие действия, запросы на исправление несоответствий)

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
ПК-6	ПК-6.1: Знать:	ПК-6.2: Уметь:	ПК-6.3: Владеть (иметь навыки):

	особенности инициации запросов на изменение	инициировать запросы на изменение (в том числе корректирующие действия, предупреждающие действия, запросы на исправление несоответствий)	навыками инициирования запросов на изменение (в том числе корректирующие действия, предупреждающие действия, запросы на исправление несоответствий)
--	---	--	---

### 3. Объем и виды учебной работы

Учебным планом предусматриваются следующие виды учебной работы по дисциплине:

#### Очная форма обучения

Виды учебной работы	Всего час/ з.е.
	Сем 3
Контактная работа, в том числе:	22.3/0.62
Занятия лекционного типа	8/0.22
Занятия семинарского типа	12/0.33
Индивидуальная контактная работа (ИКР)	0.3/0.01
Групповая контактная работа (ГКР)	2/0.06
Самостоятельная работа:	123.7/3.44
Промежуточная аттестация	34/0.94
Вид промежуточной аттестации:	
Экзамен	Экз
Общая трудоемкость (объем части образовательной программы): Часы	180
Зачетные единицы	5

#### заочная форма

Виды учебной работы	Всего час/ з.е.
	Сем 4
Контактная работа, в том числе:	22.3/0.62
Занятия лекционного типа	8/0.22
Занятия семинарского типа	12/0.33
Индивидуальная контактная работа (ИКР)	0.3/0.01
Групповая контактная работа (ГКР)	2/0.06
Самостоятельная работа:	123.7/3.44
Промежуточная аттестация	34/0.94
Вид промежуточной аттестации:	
Экзамен	Экз
Общая трудоемкость (объем части образовательной программы): Часы	180
Зачетные единицы	5

### 4. Содержание дисциплины

#### 4.1. Разделы, темы дисциплины и виды занятий:

Тематический план дисциплины Машинное обучение на больших данных представлен в таблице.

#### Разделы, темы дисциплины и виды занятий

##### Очная форма обучения

№ п/п	Наименование темы (раздела) дисциплины	Контактная работа				Самостоятельная работа	Планируемые результаты обучения в
		Лекции	Занятия семинарского типа	ИКР	ГКР		

			Практич. занятия				соотношении с результатами обучения по образовательной программе	
1.	Программный инструментарий машинного обучения и обработки данных	4	6			60	УК-3.1, УК-3.2, УК-3.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК- 6.1, ПК-6.2, ПК-6.3	
2.	Работа с данными. Алгоритмы. Сети	4	6			63,7	УК-3.1, УК-3.2, УК-3.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК- 6.1, ПК-6.2, ПК-6.3	
	Контроль	34						
	<b>Итого</b>	<b>8</b>	<b>12</b>	<b>0.3</b>	<b>2</b>	<b>123.7</b>		

#### заочная форма

№ п/п	Наименование темы (раздела) дисциплины	Контактная работа				Самостоятельная работа	Планируемые результаты обучения в соотношении с результатами обучения по образовательной программе	
		Лекции	Занятия семинарского типа		ИКР			ГКР
			Практич. занятия					
1.	Программный инструментарий машинного обучения и обработки данных	4	6			65	УК-3.1, УК-3.2, УК-3.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК- 6.1, ПК-6.2, ПК-6.3	
2.	Работа с данными. Алгоритмы. Сети	4	6			66,7	УК-3.1, УК-3.2, УК-3.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК- 6.1, ПК-6.2, ПК-6.3	
	Контроль	34						
	<b>Итого</b>	<b>8</b>	<b>12</b>	<b>0.3</b>	<b>2</b>	<b>131.7</b>		

## 4.2 Содержание разделов и тем

### 4.2.1 Контактная работа

#### Тематика занятий лекционного типа

№п/п	Наименование темы (раздела) дисциплины	Вид занятия лекционного типа*	Тематика занятия лекционного типа
1.	Программный инструментарий машинного обучения и обработки данных	лекция	Программный инструментарий Data Mining и Machine Learning
		лекция	Анализ и сжатие данных. Очистка. Регуляризация
2.	Работа с данными. Алгоритмы. Сети	лекция	Работа с данными. Алгоритмы.
		лекция	Байесовы и нейронные сети

\*лекции и иные учебные занятия, предусматривающие преимущественную передачу учебной информации педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях, обучающимся

#### Тематика занятий семинарского типа

№п/п	Наименование темы (раздела) дисциплины	Вид занятия семинарского типа**	Тематика занятия семинарского типа
1.	Программный	практическое занятие	Программный инструментарий Data

	инструментарий машинного обучения и обработки данных		Mining и Machine Learning
		практическое занятие	Анализ и сжатие данных.
		практическое занятие	Очистка.
2.	Работа с данными. Алгоритмы. Сети	практическое занятие	Работа с данными.
		практическое занятие	Алгоритмы.
		практическое занятие	Байесовы и нейронные сети

\*\* семинары, практические занятия, практикумы, лабораторные работы, коллоквиумы и иные аналогичные занятия

### Иная контактная работа

При проведении учебных занятий СГЭУ обеспечивает развитие у обучающихся навыков командной работы, межличностной коммуникации, принятия решений, лидерских качеств (включая при необходимости проведение интерактивных лекций, групповых дискуссий, ролевых игр, тренингов, анализ ситуаций и имитационных моделей, преподавание дисциплин (модулей) в форме курсов, составленных на основе результатов научных исследований, проводимых организацией, в том числе с учетом региональных особенностей профессиональной деятельности выпускников и потребностей работодателей).

Формы и методы проведения иной контактной работы приведены в Методических указаниях по основной профессиональной образовательной программе.

### 4.2.2 Самостоятельная работа

№п/п	Наименование темы (раздела) дисциплины	Вид самостоятельной работы ***
1.	Программный инструментарий машинного обучения и обработки данных	- тестирование
2.	Работа с данными. Алгоритмы. Сети	- тестирование

\*\*\* самостоятельная работа в семестре, написание курсовых работ, докладов, выполнение контрольных работ

## 5. Материально-техническое и учебно-методическое обеспечение дисциплины

### 5.1 Литература:

#### Основная литература

1. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2023. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511121>

#### Дополнительная литература

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2023. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511020>

### 5.2. Перечень лицензионного программного обеспечения

1. Microsoft Windows 10 Education / Microsoft Windows 7 / Windows Vista Business  
 2. Office 365 ProPlus, Microsoft Office 2019, Microsoft Office 2016 Professional Plus (Word, Excel, Access, PowerPoint, Outlook, OneNote, Publisher) / Microsoft Office 2007 (Word, Excel, Access, PowerPoint)

### 5.3 Современные профессиональные базы данных, к которым обеспечивается доступ

## обучающихся

1. Профессиональная база данных «Информационные системы Министерства экономического развития Российской Федерации в сети Интернет» (Портал «Официальная Россия» - <http://www.gov.ru/>)
2. Профессиональная база данных «Финансово-экономические показатели Российской Федерации» (Официальный сайт Министерства финансов РФ - <https://www.minfin.ru/ru/>)
3. Профессиональная база данных «Официальная статистика» (Официальный сайт Федеральной службы государственной статистики - <http://www.gks.ru/>)

## 5.4. Информационно-справочные системы, к которым обеспечивается доступ обучающихся

1. 1. справочно-правовая система «Консультант Плюс»
2. справочно-правовая система «ГАРАНТ-Максимум»

## 5.5. Специальные помещения

Учебные аудитории для проведения занятий лекционного типа	Комплекты ученической мебели Мультимедийный проектор Доска Экран
Учебные аудитории для проведения практических занятий (занятий семинарского типа)	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Учебные аудитории для групповых и индивидуальных консультаций	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Учебные аудитории для текущего контроля и промежуточной аттестации	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Помещения для самостоятельной работы	Комплекты ученической мебели Мультимедийный проектор Доска Экран Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ
Помещения для хранения и профилактического обслуживания оборудования	Комплекты специализированной мебели для хранения оборудования

## 5.6 Лаборатории и лабораторное оборудование

Лаборатория информационных технологий в профессиональной деятельности	Комплекты ученической мебели Мультимедийный проектор Доска Экран
---	---

	Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ Лабораторное оборудование
--	--

## 6. Фонд оценочных средств по дисциплине **Машинное обучение на больших данных**:

### 6.1. Контрольные мероприятия по дисциплине

Вид контроля	Форма контроля	Отметить нужное знаком « + »
Текущий контроль	Оценка докладов	-
	Устный/письменный опрос	-
	Тестирование	+
	Практические задачи	-
	Оценка контрольных работ (для заочной формы обучения)	+
Промежуточный контроль	Экзамен	+

Порядок проведения мероприятий текущего и промежуточного контроля определяется Методическими указаниями по основной профессиональной образовательной программе высшего образования; Положением о балльно-рейтинговой системе оценки успеваемости обучающихся по основным образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры в федеральном государственном автономном образовательном учреждении высшего образования «Самарский государственный экономический университет».

### 6.2. Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе

#### Универсальные компетенции (УК):

УК-3 - Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
	УК-3.1: Знать:	УК-3.2: Уметь:	УК-3.3: Владеть (иметь навыки):
	особенности управления и руководства работой команды, выработки командной стратегии для достижения поставленной цели	организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	навыками управления и руководства работой команды, выработки командной стратегии для достижения поставленной цели
Пороговый	основы искусственного интеллекта	применять основы искусственного интеллекта	применения искусственного интеллекта
Стандартный (в дополнение к пороговому)	значение и сущность машинного обучения на больших данных	применять машинное обучение на больших данных	применения машинного обучения на больших данных
Повышенный (в дополнение к пороговому, стандартному)	особенности решения задач машинного обучения на больших данных в команде	решать задачи машинного обучения на больших данных в команде	решения задач машинного обучения на больших данных в команде

#### Профессиональные компетенции (ПК):



ПК-2 - Способен разрабатывать правила использования репозитория проекта

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
	ПК-2.1: Знать:	ПК-2.2: Уметь:	ПК-2.3: Владеть (иметь навыки):
	особенности разработки правил использования репозитория проекта	разрабатывать правила использования репозитория проекта	навыками разработки правил использования репозитория проекта
Пороговый	наиболее распространенные методы машинного обучения	производить поиск и отбор публикаций по машинному обучению в различных источниках	навыками построения и проверки качества формальных математических моделей
Стандартный (в дополнение к пороговому)	наиболее значимые отечественные и зарубежные журналы в области машинного обучения; электронные ресурсы, связанные с машинным обучением, анализом данных,	анализировать, обобщать и формировать сравнительные обзоры функциональных возможностей и технологических характеристик программных инструментов машинного обучения	навыками интерпретации формальных математических моделей в терминах прикладной области с целью получения новых нетривиальных знаний и выводов
Повышенный (в дополнение к пороговому, стандартному)	основные положения теории обучения по прецедентам; методы предварительной обработки данных	планировать исследование, основывающееся на анализе прецедентов и направленное на предсказательное моделирование	навыками использования высокоуровневых программных средств для предварительной обработки исходных данных

ПК-6 - Способен инициировать запросы на изменение (в том числе корректирующие действия, предупреждающие действия, запросы на исправление несоответствий)

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
	ПК-6.1: Знать:	ПК-6.2: Уметь:	ПК-6.3: Владеть (иметь навыки):
	особенности инициации запросов на изменение	инициировать запросы на изменение (в том числе корректирующие действия, предупреждающие действия, запросы на исправление несоответствий)	навыками инициирования запросов на изменение (в том числе корректирующие действия, предупреждающие действия, запросы на исправление несоответствий)
Пороговый	основы интеллектуального анализа данных	основные способы визуализации данных	методы понижения размерности данных
Стандартный (в дополнение к пороговому)	решать нестандартные профессиональные задачи, с применением элементов машинного обучения	проводить анализ данных, проводить предобработку и очистку данных, работать с пропущенными значениями	выбирать наиболее подходящие алгоритмы решения задач машинного обучения и оценивать качество построенных моделей
Повышенный (в дополнение к пороговому, стандартному)	инструментами анализа данных и решения задач	навыками применения новых научных	Навыками составления композиции моделей,

пороговому, стандартному)	машинного обучения	принципов и методов исследования для решения профессиональных задач	проведения отбора признаков
---------------------------	--------------------	---	-----------------------------

### 6.3. Паспорт оценочных материалов

№ п/п	Наименование темы (раздела) дисциплины	Контролируемые планируемые результаты обучения в соотношении с результатами обучения по программе	Вид контроля/используемые оценочные средства	
			Текущий	Промежуточный
1.	Программный инструментарий машинного обучения и обработки данных	УК-3.1, УК-3.2, УК- 3.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК-6.1, ПК- 6.2, ПК-6.3	Оценка контрольных работ (для заочной формы обучения) Тестирование	Экзамен
2.	Работа с данными. Алгоритмы. Сети	УК-3.1, УК-3.2, УК- 3.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК-6.1, ПК- 6.2, ПК-6.3	Оценка контрольных работ (для заочной формы обучения) Тестирование	Экзамен

### 6.4. Оценочные материалы для текущего контроля

**Задания для тестирования по дисциплине для оценки сформированности компетенций (min 20, max 50 + ссылку на ЭИОС с тестами) <https://lms2.sseu.ru/course/index.php?categoryid=1863>**

Отметьте верные высказывания о принципах Hadoop:

Вертикальное масштабирование

Инкапсуляция сложности реализации распределенной системы Горизонтальное масштабирование

Отправка кода к данным Отказоустойчивость Перемещение данных к коду

Какие функции выполняют вендоры дистрибутивов?

Разрабатывают дополнительные фичи в стандартных компонентах Hadoop Обеспечивают совместимость разных компонентов Hadoop

Предоставляют дистрибутивы в различных форматах (rpm, tar.gz, образ виртуальных машин)

Исправляют ошибки в компонентах Hadoop

Как осуществляется доступ к данным на чтение в HDFS? Произвольный доступ к нужной части файла Последовательное чтение всего файла с данными

Как можно менять данные в файле HDFS? Удалять данные в конце файла

Удалять произвольную часть файла Дописывать данные в конец файла Удалять все данные (удалять файл)

Дописывать данные в произвольное место файла (начало, середина, конец)

Для какого шаблона взаимодействия с данными был разработан HDFS? Много раз поменяли данные / один раз прочитали

Один раз записали / один раз прочитали Один раз записали / много раз прочитали

Много раз поменяли данные / много раз прочитали

Чем больше размер блока HDFS, тем...

Тем меньше файлов в HDFS можно хранить Тем больше файлов в HDFS можно хранить

Как осуществляется доступ к HDFS? Только прямой доступ (Direct Access) Прямой доступ или

через проху-сервер Только через проху-сервер

Какие предложения характеризуют HDFS?

Данные не теряются, если выходят из строя диски/сервера Является приложением пользовательского уровня ОС

Высокая пропускная способность вместо быстрого доступа к данным Несколько процессов могут писать данные в один файл HDFS Работает на кластере из суперкомпьютеров

Лучше хранить миллиард небольших файлов, чем миллион больших

Какие варианты соответствуют URI схеме и могут быть использованы в командах shell-клиента HDFS? /tmp/output.txt

file:///home/user/ hdfs:data.txt

hdfs://localhost:9000//data/log.txt hdfs://hdfs/ file:///

Как лучше скопировать файл file.txt размером 10Тб из папки /data/ в папку /tmp/

```
$hdfs distcp /data/file.txt /tmp/
```

```
$hdfs dfs -get /data/file.txt . ; hdfs dfs -put file.txt /tmp/
```

```
$hdfs dfs -cp /data/file.txt /tmp/
```

Как лучше скопировать файл file.txt размером 10Мб из папки /data/ в папку /tmp/

```
$hdfs distcp /data/file.txt /tmp/
```

```
$hdfs dfs -cp /data/file.txt /tmp/
```

```
$hdfs dfs -get /data/file.txt . ; hdfs dfs -put file.txt /tmp/
```

Что означает, если файл имеет такие права доступа: -rw-r-----

Владелец и пользователи из той же группы могут читать и писать в файл, другие могут только читать из файла

Все пользователи могут читать и писать в файл

Владелец может читать и писать в файл, пользователи из той же группы могут читать файл, другие ничего не могут делать с файлом

Только владелец может читать и писать в файл.

Остальные пользователи не могут ни читать, ни писать в файл

Какие реализации абстрактного класса FileSystem существуют в HDFS API? FTPFileSystem

LocalFileSystem HftpFileSystem DistributedFileSystem

В каком виде осуществляется передача данных между Map и Reduce? Массив байт Ключ / Значение

В виде xml

Можно самому определять формат Текстовые сообщения

Как определяется число mapper-процессов в MapReduce задаче? Задается пользователем при описании MapReduce задачи Обычно равно число сплитов во входных данных

Что такое data locality?

Запуск mapper на той же ноде, где хранятся нужные данные Процесс определения, на какой ноде расположены нужные данные Процесс копирования нужных данных на ноду с запущенным mapper

Как хранятся промежуточные данные между фазами Map и Reduce? На локальном диске

В HDFS

Что происходит, если выходит из строя диск с промежуточными данными между Map и Reduce?

Данные восстанавливаются из других реплик

Перезапускается mapper, который создал эти данные

Задача завершается с ошибкой, т.к. потеряны промежуточные данные

Что такое shuffle?

Определение номера редьюсера для каждой пары Ключ-Значение в промежуточных данных

Процесс перемешивания промежуточных данных для равномерного распределения по reducer'ам

Сортировка промежуточных данных между Map и Reduce

Процесс передачи промежуточных данных из mapper'ов в reducer

Как определяется число reducer-процессов в MapReduce задаче?

Определяется системой, исходя из размера промежуточных данных между Map и Reduce. Задается пользователем при описании MapReduce задачи.

### Тематика контрольных работ

Раздел дисциплины	Темы
Программный инструментарий машинного обучения и обработки данных	Тема 1: Источники данных. Цель: Получить представление об открытых источниках данных, используемых в научно-исследовательской деятельности. Задание: 1. Изучить различные источники геоданных. 2. Оценить количественные характеристики баз данных. Используя возможности информационных технологий реализовать решение поставленной научной задачи. Тема 2: Основные понятия Big Data. Цель: Ознакомиться с предметной областью, относящейся к машинному обучению и большим данным. Задание: 1. Дать определение основным терминам и понятиям, относящимся к предметной области. 2. Определить взаимосвязи прикладных и теоретических дисциплин, являющихся базовыми к машинному обучению. Тема 3: Технологии работы с Big Data. Цель: Ознакомиться с основными технологиями работы с Big Data и машинным обучением. Задание: Ознакомиться с применением языка программирования Python в качестве системы обработки данных и машинного обучения.
Работа с данными. Алгоритмы. Сети	Тема 4: Аналитика данных. Цель: Ознакомиться с основными технологиями работы с Big Data и машинным обучением на примере программных комплексов Weka и Rubi. Задание: Оценить возможности программных комплексов Weka и Rubi на примере решения задачи машинного обучения, приведенного в качестве базового примера разработчиками. Тема 5: Управление проектами Big Data. Цель: Изучить типовой пример развития и управления проектом машинного обучения. Задание: 1. Разработать структуру проекта для реализации научной задачи. 2. Используя возможности информационных технологий реализовать решение поставленной научной задачи.

### 6.5. Оценочные материалы для промежуточной аттестации

#### Фонд вопросов для проведения промежуточного контроля в форме экзамена

Раздел дисциплины	Вопросы
Программный инструментарий	1. Введение в машинное обучение и обработку данных. Постановка основных классов задач в машинном обучении. 2. Обучение с учителем; регрессия и классификация; обучение без

<p>машинного обучения и обработки данных</p>	<p>учителя; кластеризация, снижение размерности; рекомендательные системы; обработка текстов; обработка изображений; обучение с подкреплением.</p> <p>3. Программный инструментарий Data Mining и Machine Learning.</p> <p>4. Основной аппарат комбинаторики и мат. статистики.</p> <p>5. Основные понятия математической статистики: статистические оценки, их свойства, проверка гипотез. Регрессионный анализ и сжатие данных.</p> <p>6. Задача регрессии.</p> <p>7. Минимизация квадрата отклонения.</p> <p>8. Регрессионная функция: условное мат. ожидание.</p> <p>10. Линейная регрессия и метод k ближайших соседей.</p> <p>11. Переобучение и недообучение.</p> <p>12. Разложение ошибки на шум, смещение и разброс.</p>
<p>Работа с данными. Алгоритмы. Сети</p>	<ol style="list-style-type: none"> <li>1. Детектирование выбросов и аномалий.</li> <li>2. Что такое выбросы, типы выбросов.</li> <li>3. Методы обнаружения выбросов.</li> <li>4. Поиск аномалий.</li> <li>5. Цензурирование выборки.</li> <li>6. Отсев объектов-выбросов, удаление выбросов.</li> <li>7. Очистка данных и технологии регуляризации.</li> <li>8. Основные виды регуляризации.</li> <li>9. Метод редукции размерности.</li> <li>10. Методы отбора признаков.</li> <li>11. Технологии кластеризации и классификации.</li> <li>12. K-means.</li> <li>13. EM-алгоритм.</li> <li>14. Другие методы кластеризации.</li> <li>15. Задачи классификации.</li> <li>16. Байесовский классификатор.</li> <li>17. Линейные методы для классификации.</li> <li>18. Логистическая регрессия, максимизация правдоподобия.</li> </ol> <p>Нейронные сети: общая архитектура.</p> <ol style="list-style-type: none"> <li>19. Многослойные сети.</li> <li>20. Обратное распространение ошибки.</li> <li>21. Стохастический градиентный спуск.</li> <li>22. Генетические алгоритмы.</li> <li>23. Извлечение признаков / выделение особенностей (Feature detection).</li> <li>24. Преобразования признаков.</li> <li>25. Нормализация данных.</li> <li>26. Методы нормализации данных.</li> <li>27. Нормализация по методу минимакса.</li> <li>28. Нормализация по Z-показателю.</li> <li>29. Десятичное масштабирование.</li> <li>30. Нечеткие множества.</li> <li>31. Байесовы сети.</li> <li>32. Задачи байесовского вывода.</li> <li>33. Методика построения нечеткой байесовой сети</li> </ol>

## 6.6. Шкалы и критерии оценивания по формам текущего контроля и промежуточной аттестации

### Шкала и критерии оценивания

Оценка	Критерии оценивания для мероприятий контроля с применением
--------	--

	<b>4-х балльной системы</b>
<b>«отлично»</b>	Повышенный УК-3.1, УК-3.2, УК-3.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК-6.1, ПК-6.2, ПК-6.3
<b>«хорошо»</b>	Стандартный УК-3.1, УК-3.2, УК-3.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК-6.1, ПК-6.2, ПК-6.3
<b>«удовлетворительно»</b>	Пороговый УК-3.1, УК-3.2, УК-3.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК-6.1, ПК-6.2, ПК-6.3
<b>«неудовлетворительно»</b>	Результаты обучения не сформированы на пороговом уровне