

Документ подписан простой электронной подписью.  
Информация о владельце:

ФИО: Кандрашина Елена Александровна

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

Дата подписания: 07.07.2023 15:32:02

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**

<b>Наименование дисциплины</b>	Б1.В.ДЭ.05.02 Машинное обучение на больших данных
<b>Основная профессиональная образовательная программа</b>	09.03.03 Прикладная информатика программа Цифровые технологии в экономике

Квалификация (степень) выпускника Бакалавр

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 (дисциплина по выбору) блока Б1.Дисциплины (модули)

Предшествующие дисциплины по связям компетенций: Интеллектуальные информационные системы, Основы проектной деятельности, Инженерия знаний, Методы оптимизации и теория игр, Хранение, обработка и анализ данных, Системы искусственного интеллекта, Разработка интерфейсов и адаптивный Веб-дизайн, Технологии работы в социальных сетях, Информационно-коммуникационные технологии в профессиональной деятельности, Основы алгоритмизации и программирования, Современные технологии и языки программирования, Встроенные языки программирования, Организация вычислительных процессов, Предпринимательское дело, Основы права, Гражданское право, Облачные технологии и услуги

Последующие дисциплины по связям компетенций: Проектирование информационных систем, Проектный практикум, Управление ИТ-проектами, Управление качеством разработки приложений, Цифровые технологии управления предприятием, Современные цифровые платформы, Разработка профессиональных приложений, Разработка мобильных приложений, Интернет-предпринимательство

## 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-2 - Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
УК-2	УК-2.1: Знать:	УК-2.2: Уметь:	УК-2.3: Владеть (иметь навыки):
	оптимальные способы решения поставленных задач с учетом правовых норм, имеющихся ресурсов и ограничений	определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	навыками определения круга задач в рамках поставленной цели, выбора оптимальных способов их решения

### Профессиональные компетенции (ПК):

ПК-1 - Способен к подготовке коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
ПК-1	ПК-1.1: Знать:	ПК-1.2: Уметь:	ПК-1.3: Владеть (иметь навыки):

	особенности подготовки коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС	готовить коммерческое предложение заказчику по созданию (модификации) и вводу в эксплуатацию ИС	навыками подготовки коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС
--	--	---	---

ПК-4 - Способен к верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС

Планируемые результаты обучения по программе	<b>Планируемые результаты обучения по дисциплине</b>		
ПК-4	ПК-4.1: Знать:	ПК-4.2: Уметь:	ПК-4.3: Владеть (иметь навыки):
	особенности верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС	верифицировать структуру программного кода ИС относительно архитектуры ИС и требований заказчика к ИС	навыками верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС

### 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 6
Контактная работа, в том числе:	54.15/1.5
Занятия лекционного типа	18/0.5
Лабораторные работы (лабораторный практикум)	36/1
Индивидуальная контактная работа (ИКР)	0.15/0
Самостоятельная работа:	35.85/1
Промежуточная аттестация	18/0.5
Вид промежуточной аттестации: Зачет	Зач
Общая трудоемкость (объем части образовательной программы): Часы	108
Зачетные единицы	3

#### заочная форма

Виды учебной работы	Всего час/ з.е.
	Сем 7
Контактная работа, в том числе:	4.15/0.12
Занятия лекционного типа	2/0.06
Лабораторные работы (лабораторный практикум)	2/0.06
Индивидуальная контактная работа (ИКР)	0.15/0
Самостоятельная работа:	85.85/2.38
Промежуточная аттестация	18/0.5
Вид промежуточной аттестации: Зачет	Зач
Общая трудоемкость (объем части образовательной программы): Часы	108
Зачетные единицы	3

#### 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 соотношении с результатами обучения по образовательной программе
		Лекции	Занятия семинарского типа	ИКР	ГКР		
			Лабораг. работы				
1.	Основы машинного обучения на больших данных.	9	18	0.1		18	УК-2.1, УК-2.2, УК-2.3, ПК-1.1, ПК-1.2, ПК-1.3, ПК-4.1, ПК-4.2, ПК-4.3
2.	Глубокое машинное обучение на больших данных	9	18	0.05		17.85	УК-2.1, УК-2.2, УК-2.3, ПК-1.1, ПК-1.2, ПК-1.3, ПК-4.1, ПК-4.2, ПК-4.3
	Контроль	18					
	<b>Итого</b>	<b>18</b>	<b>36</b>	<b>0.15</b>		<b>35.85</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		
			Лабораг. работы				
1.	Основы машинного обучения на больших данных.	1	1	0.1		45	УК-2.1, УК-2.2, УК-2.3, ПК-1.1, ПК-1.2, ПК-1.3, ПК-4.1, ПК-4.2, ПК-4.3
2.	Глубокое машинное обучение на больших данных	1	1	0.05		40.85	УК-2.1, УК-2.2, УК-2.3, ПК-1.1, ПК-1.2, ПК-1.3, ПК-4.1, ПК-4.2, ПК-4.3
	Контроль	18					
	<b>Итого</b>	<b>2</b>	<b>2</b>	<b>0.15</b>		<b>85.85</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.		лекция	Программный инструментарий Data Mining и Machine Learning

	Основы машинного обучения на больших данных.	лекция	Анализ и сжатие данных. Очистка. Регуляризация
		лекция	Программный инструментарий машинного обучения и обработки данных
		лекция	
2.	Глубокое машинное обучение на больших данных	лекция	Работа с данными. Алгоритмы.
		лекция	Байесовы и нейронные сети
		лекция	Работа с данными. Алгоритмы. Сети
		лекция	
		лекция	

\*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.	Основы машинного обучения на больших данных.	лабораторные работы	Программный инструментарий Data Mining и Machine Learning
		лабораторные работы	Анализ и сжатие данных
		лабораторные работы	Анализ и сжатие данных и очистка данных
2.	Глубокое машинное обучение на больших данных	лабораторные работы	Работа с данными.
		лабораторные работы	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.	Основы машинного обучения на больших данных.	- тестирование
2.	Глубокое машинное обучение на больших данных	- 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. 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>

## Дополнительная 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>

2. Инженерная и компьютерная графика : учебник и практикум для вузов / Р. Р. Анамова [и др.] ; под общей редакцией Р. Р. Анамовой, С. А. Леоновой, Н. В. Пшеничной. — Москва : Издательство Юрайт, 2023. — 246 с. — (Высшее образование). — ISBN 978-5-9916-8262-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/512176>

### 5.2. Перечень лицензионного программного обеспечения

#### Обязательное программное обеспечение

Microsoft Windows 10 Education / Microsoft Windows 7 / Windows Vista Business

#### Программное обеспечение по выбору

1. 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. 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

## 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	Зачет	+

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-2 - Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
		УК-2.1: Знать: оптимальные способы решения поставленных	УК-2.2: Уметь: определять круг задач в рамках поставленной



	задач с учетом правовых норм, имеющихся ресурсов и ограничений	цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	поставленной цели, выбора оптимальных способов их решения
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 принципов и методов исследования для решения профессиональных задач	Навыками составления композиции моделей, проведения отбора признаков

#### Профессиональные компетенции (ПК):

ПК-1 - Способен к подготовке коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС

Планируемые результаты обучения по программе	<b>Планируемые результаты обучения по дисциплине</b>		
	ПК-1.1: Знать:	ПК-1.2: Уметь:	ПК-1.3: Владеть (иметь навыки):
	особенности подготовки коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС	готовить коммерческое предложение заказчику по созданию (модификации) и вводу в эксплуатацию ИС	навыками подготовки коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС
Пороговый	основы искусственного интеллекта	применять основы искусственного	основы искусственного интеллекта применять основы искусственного интеллекта применения искусственного
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

ПК-4 - Способен к верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
		ПК-4.1: Знать:	ПК-4.2: Уметь:
	особенности верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС	верифицировать структуру программного кода ИС относительно архитектуры ИС и требований заказчика к ИС	навыками верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС
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

### 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.	Основы машинного обучения на больших данных.	УК-2.1, УК-2.2, УК-2.3, ПК-1.1, ПК-1.2, ПК-1.3, ПК-4.1, ПК-4.2, ПК-4.3	Тестирование	Зачет
2.	Глубокое машинное обучение на больших данных	УК-2.1, УК-2.2, УК-2.3, ПК-1.1, ПК-1.2, ПК-1.3, ПК-4.1, ПК-4.2, ПК-4.3	Тестирование	Зачет

### 6.4. Оценочные материалы для текущего контроля

### Примерная тематика докладов

Раздел дисциплины	Темы

### Вопросы для устного/письменного опроса

Раздел дисциплины	Вопросы

**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=1819>**  
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

**Практические задачи (min 20, max 50 + ссылку на ЭИОС с электронным изданием, если имеется)**

Раздел дисциплины	Задачи
Основы машинного обучения на больших данных.	Тема 1: Источники данных. Цель: Получить представление об открытых источниках данных, используемых в научно-исследовательской деятельности.

	<p>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: Ознакомится 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: Ознакомится с основными технологиями работы с Big Data и машинным обучением. Задание: Ознакомится с применением языка программирования Python в качестве системы обработки данных и машинного обучения.</p>
Глубокое машинное обучение на больших данных	<p>Тема 4: Аналитика данных. Цель: Ознакомится 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.</p>

#### Тематика контрольных работ

Раздел дисциплины	Темы

#### 6.5. Оценочные материалы для промежуточной аттестации

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

Раздел дисциплины	Вопросы
Основы машинного обучения на больших данных.	<p>1. Введение в машинное обучение и обработку данных. Постановка основных классов задач в машинном обучении. Обучение с учителем; регрессия и классификация; обучение без учителя; кластеризация, снижение размерности; рекомендательные системы; обработка текстов; обработка изображений; обучение с подкреплением.</p>
Глубокое машинное обучение на больших данных	<p>3. Программный инструментарий Data Mining и Machine Learning. 4. Основной аппарат комбинаторики и мат. статистики.</p>

	<p>5. Основные понятия математической статистики: статистические оценки, их свойства, проверка гипотез.</p> <p>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>Разложение ошибки на шум, смещение и разброс.</p> <p>1. Детектирование выбросов и аномалий.</p> <p>2. Что такое выбросы, типы выбросов.</p> <p>3. Методы обнаружения выбросов.</p> <p>4. Поиск аномалий.</p> <p>5. Цензурирование выборки.</p> <p>6. Отсев объектов-выбросов, удаление выбросов.</p> <p>7. Очистка данных и технологии регуляризации.</p> <p>8. Основные виды регуляризации.</p> <p>9. Метод редукции размерности.</p> <p>10. Методы отбора признаков.</p> <p>11. Технологии кластеризации и классификации.</p> <p>12. K-means.</p> <p>13. EM-алгоритм.</p> <p>14. Другие методы кластеризации.</p> <p>15. Задачи классификации.</p> <p>16. Байесовский классификатор.</p> <p>17. Линейные методы для классификации.</p> <p>18. Логистическая регрессия, максимизация правдоподобия.</p> <p>Нейронные сети: общая архитектура.</p> <p>19. Многослойные сети.</p> <p>20. Обратное распространение ошибки.</p> <p>21. Стохастический градиентный спуск.</p> <p>22. Генетические алгоритмы.</p> <p>23. Извлечение признаков / выделение особенностей (Feature detection).</p> <p>24. Преобразования признаков.</p> <p>25. Нормализация данных.</p> <p>26. Методы нормализации данных.</p> <p>27. Нормализация по методу минимакса.</p> <p>28. Нормализация по Z-показателю.</p> <p>29. Десятичное масштабирование.</p> <p>30. Нечеткие множества.</p> <p>31. Байесовы сети.</p> <p>32. Задачи байесовского вывода.</p> <p>Методика построения нечеткой байесовой сети</p>
--	--

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

### Шкала и критерии оценивания

<b>Оценка</b>	<b>Критерии оценивания для мероприятий контроля с применением 2-х балльной системы</b>
«зачтено»	УК-2, ПК-1, ПК-4
«не зачтено»	Результаты обучения не сформированы на пороговом уровне