

Документ подписан простой электронной подписью.
Информация о владельце:

ФИО: Кандрашина Елена Александровна

Должность: И.о. ректора ФГАОУ ВО «Самарский государственный экономический университет»

Дата подписания: 10.07.2025 14:16:30

Уникальный программный ключ:

2db64eb9605ce27edd3b8e8fdd32c70e0674ddd2

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

Институт Институт менеджмента
Кафедра Прикладной информатики

УТВЕРЖДЕНО

Ученым советом Университета
(протокол № 10 от 22 мая 2025 г.)

РАБОЧАЯ ПРОГРАММА

Наименование дисциплины	Б1.В.07 Машинное обучение на больших данных
Основная профессиональная образовательная программа	38.04.02 Менеджмент программа Международная логистика и управление цифровыми цепями поставок

Квалификация (степень) выпускника магистр

Самара 2025

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

Институт Институт менеджмента
Кафедра Прикладной информатики

АННОТАЦИЯ

Наименование дисциплины	Б1.В.07 Машинное обучение на больших данных
Основная профессиональная образовательная программа	38.04.02 Менеджмент программа Международная логистика и управление цифровыми цепями поставок

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

Институт Институт менеджмента
Кафедра Прикладной информатики

УТВЕРЖДЕНО

Ученым советом Университета
(протокол № 10 от 22 мая 2025 г.)

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ**

Наименование дисциплины	Б1.В.07 Машинное обучение на больших данных
Основная профессиональная образовательная программа	38.04.02 Менеджмент программа Международная логистика и управление цифровыми цепями поставок

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

Содержание (ФОС)

Стр.

- 6.1 Контрольные мероприятия по дисциплине
- 6.2 Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе
- 6.3 Паспорт оценочных материалов
- 6.4 Оценочные материалы для текущего контроля
- 6.5 Оценочные материалы для промежуточной аттестации
- 6.6 Шкалы и критерии оценивания по формам текущего контроля и промежуточной аттестации

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: Интеллектуальное планирование, Анализ данных и бизнес-аналитика в цепях поставок, Контроллинг в международной логистике, Функциональное логистическое управление в цепях поставок, 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:

Профессиональные компетенции (ПК):

ПК-2 - Способен к организации исследований и разработке перспективных методов, моделей и механизмов планирования и организации сетей поставок

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
	ПК-2.1: Знать:	ПК-2.2: Уметь:	ПК-2.3: Владеть (иметь навыки):
ПК-2	методы, модели и механизмы планирования и организации сетей поставок	проводить исследования с применением перспективных методик планирования и организации цепей поставок	методологией проведения исследований в цепях поставок с целью разработки и внедрения инновационных технологий планирования и организации логистических систем

ПК-3 - Способен к руководству проектами разработки и внедрения средств информационной поддержки управления сетями поставок

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
	ПК-3.1: Знать:	ПК-3.2: Уметь:	ПК-3.3: Владеть (иметь навыки):
ПК-3	основы проектной деятельности по внедрению информационных систем управления движением товарно-материальных ценностей в цепях поставок	принимать организационно-управленческие решения по разработке и внедрению инновационных технологий по информационной поддержке участников цепей поставок	навыками разработки проектов инновационных информационных систем и внедрения технологий информационной поддержки управленческих решений в сетях поставок

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 4
Контактная работа, в том числе:	18.3/0.51
Занятия семинарского типа	16/0.44
Индивидуальная контактная работа (ИКР)	0.3/0.01
Групповая контактная работа (ГКР)	2/0.06
Самостоятельная работа:	127.7/3.55
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 соотношении с результатами обучения по образовательной программе
		Занятия семинарского типа	ИКР	ГКР		
		Практич. занятия				
1.	Программный инструментарий машинного обучения и обработки данных	8	0.15	1	60	ПК-2.1, ПК-2.2, ПК -2.3, ПК-3.1, ПК- 3.2, ПК-3.3
2.	Работа с данными.Алгоритмы. Сети	8	0.15	1	67.7	ПК-2.1, ПК-2.2, ПК -2.3, ПК-3.1, ПК- 3.2, ПК-3.3
	Контроль	34				
	Итого	16	0.3	2	127.7	

4.2 Содержание разделов и тем

4.2.1 Контактная работа

Тематика занятий семинарского типа

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

		практическое занятие	
--	--	----------------------	--

** 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 инструментальный машинного обучения и обработки данных	- подготовка доклада - подготовка электронной презентации - тестирование
2.	Работа с данными. Алгоритмы. Сети	- подготовка доклада - подготовка электронной презентации - 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. Базовые методы анализа данных : учебник и практикум для вузов / Б. Г. Миркин. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 297 с. — (Высшее образование). — ISBN 978-5-534-19709-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/560414>.

Дополнительная литература

1. Анализ данных : учебник для вузов / под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2025. — 448 с. — (Высшее образование). — ISBN 978-5-534-19964-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/560311>.

Литература для самостоятельного изучения

1. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 89 с. — (Высшее образование). — ISBN 978-5-534-20732-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/558662>.

5.2. Перечень лицензионного программного обеспечения

1. Astra Linux Special Edition «Смоленск», «Орел»; РедОС.
2. МойОфис Стандартный 2, МойОфис Образование, Р7-Офис Профессиональный, МойОфис Стандартный 3, МойОфис Профессиональный 3

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. Государственная система правовой информации «Официальный интернет-портал правовой информации» (<http://pravo.gov.ru/>)

3. Профессиональная база данных «Финансово-экономические показатели Российской Федерации» (Официальный сайт Министерства финансов РФ - <https://www.minfin.ru/ru/>)

4. 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	+
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

Профессиональные компетенции (ПК):

ПК-2 - Способен к организации исследований и разработке перспективных методов, моделей и механизмов планирования и организации сетей поставок

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
	ПК-2.1: Знать:	ПК-2.2: Уметь:	ПК-2.3: Владеть (иметь навыки):
	методы, модели и механизмы планирования и организации сетей поставок	проводить исследования с применением перспективных методик планирования и организации цепей поставок	методологией проведения исследований в цепях поставок с целью разработки и внедрения инновационных технологий планирования и организации логистических систем
Пороговый	основные базисные условия поставки	использовать типовые методы и способы выполнения профессиональных задач в области организации цепей поставок, оценивать их эффективность и качество	навыками выбора необходимых характеристик проектируемой цепи поставок с учетом фактических параметров внешней среды
Стандартный (в дополнение к пороговому)	отечественный и зарубежный опыт рациональной организации сетей поставок продукции	отслеживать и осуществлять координацию материальных потоков в сети поставок	навыками оценки эффективности организации снабжения в сети поставок
Повышенный (в дополнение к пороговому, стандартному)	организацию управления и мониторинга процессов в сетях поставок	формировать базы данных и разрабатывать организационно-управленческую документацию с использованием современных технологий электронного документооборота	навыками разработки основных положений стратегии развития сетей поставок, обоснование стратегических решений по совершенствованию процессов

			стратегического и тактического планирования и организации цепей поставок
--	--	--	--

ПК-3 - Способен к руководству проектами разработки и внедрения средств информационной поддержки управления сетями поставок

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
	ПК-3.1: Знать:	ПК-3.2: Уметь:	ПК-3.3: Владеть (иметь навыки):
	основы проектной деятельности по внедрению информационных систем управления движением товарно-материальных ценностей в цепях поставок	принимать организационно-управленческие решения по разработке и внедрению инновационных технологий по информационной поддержке участников цепей поставок	навыками разработки проектов инновационных информационных систем и внедрения технологий информационной поддержки управленческих решений в сетях поставок
Пороговый	основы проектной деятельности, информационные системы управления, товарно-материальные ценности	применять знания по основам проектной деятельности, информационных систем управления, товарно-материальных ценностей	навыками разработки проектов информационных систем управления
Стандартный (в дополнение к пороговому)	типовые программные продукты для проектирования и выполнения процессов в цепях поставок, планирования и оперативного управления материальными потоками	применять типовые программные продукты для проектирования и выполнения процессов в цепях поставок, планирования и оперативного управления материальными потоками	навыками использования типовых программных продуктов для проектирования и выполнения процессов в цепях поставок, планирования и оперативного управления материальными потоками
Повышенный (в дополнение к пороговому, стандартному)	методы оптимизации процессов поставки товара и систем управления движением товарно-материальных ценностей в цепях поставок	применять методы оптимизации процессов поставки товара и систем управления движением товарно-материальных ценностей в цепях поставок	навыками использования методов оптимизации процессов поставки товара и систем управления движением товарно-материальных ценностей в цепях поставок

6.3. Паспорт оценочных материалов

№ п/п	Наименование темы (раздела) дисциплины	Контролируемые планируемые результаты обучения в соотношении с	Вид контроля/используемые оценочные средства	
			Текущий	Промежуточный

		результатами обучения по программе		
1.	Программный инструментарий машинного обучения и обработки данных	ПК-2.1, ПК-2.2, ПК- 2.3, ПК-3.1, ПК-3.2, ПК-3.3	Тестирование	Экзамен
2.	Работа с данными. Алгоритмы. Сети	ПК-2.1, ПК-2.2, ПК- 2.3, ПК-3.1, ПК-3.2, ПК-3.3	Тестирование	Экзамен

6.4.Оценочные материалы для текущего контроля

Примерная тематика докладов

Раздел дисциплины	Темы
Программный инструментарий машинного обучения и обработки данных	1. Методы машинного обучения: методы регрессии. 2. Методы машинного обучения: методы классификации. 3. Методы машинного обучения: методы кластеризации. 4. Многослойные нейронные сети. 5. Приложения сверточных нейронных сетей. 6. Автоматическая обработка естественного языка с использованием методов машинного обучения. 7. Метод Reinforcement Learning. 8. Методы машинного обучения в автомобильной промышленности. 9. Методы машинного обучения в робототехнике. 10. Персональные умные ассистенты.
Работа с данными. Алгоритмы. Сети	1. Введение в работу с данными: основные понятия и инструменты 2. Методы сбора и очистки данных 3. Базы данных и системы управления данными 4. Анализ данных с помощью SQL 5. Инструменты для визуализации данных 6. Машинное обучение и анализ данных 7. Этика и безопасность данных 8. Большие данные (Big Data) 9. Практическое применение анализа данных в бизнесе 10. Будущее работы с данными 11. Введение в алгоритмы: основные понятия и классификация 12. Эффективные алгоритмы сортировки: сравнение и анализ 13. Алгоритмы поиска: от линейного до бинарного 14. Граф-алгоритмы: пути, потоки и связность 15. Динамическое программирование: принципы и примеры 16. Рекурсивные алгоритмы: преимущества и недостатки 17. Алгоритмы шифрования и криптографии 18. Параллельные и распределенные алгоритмы 19. Эвристические алгоритмы и методы оптимизации 20. Анализ сложности алгоритмов: Big O и другие нотации 21. Модель взаимодействия открытых систем (OSI) ISO 22. Семейство протоколов IEEE 802.11 (WiFi) 23. Стек протоколов TCP/IP v4 24. Протокол IPv6 25. Технология Network Address Translation 26. Динамическая маршрутизация 27. Автономные системы и маршрутизация в Internet 28. Протокол HTTP 29. Протокол BitTorrent 30. TOR (The Onion Router)

Вопросы для устного/письменного опроса

Раздел дисциплины	Вопросы
Программный инструментарий машинного обучения и обработки данных	1. Методы машинного обучения: методы регрессии. 2. Методы машинного обучения: методы классификации. 3. Методы машинного обучения: методы кластеризации. 4. Многослойные нейронные сети. 5. Приложения сверточных нейронных сетей. 6. Автоматическая обработка естественного языка с использованием методов машинного обучения. 7. Метод Reinforcement Learning. 8. Методы машинного обучения в автомобильной промышленности. 9. Методы машинного обучения в робототехнике. 10. Персональные умные ассистенты.
Работа с данными. Алгоритмы. Сети	1. Введение в работу с данными: основные понятия и инструменты 2. Методы сбора и очистки данных 3. Базы данных и системы управления данными 4. Анализ данных с помощью SQL 5. Инструменты для визуализации данных 6. Машинное обучение и анализ данных 7. Этика и безопасность данных 8. Большие данные (Big Data) 9. Практическое применение анализа данных в бизнесе 10. Будущее работы с данными 11. Введение в алгоритмы: основные понятия и классификация 12. Эффективные алгоритмы сортировки: сравнение и анализ 13. Алгоритмы поиска: от линейного до бинарного 14. Граф-алгоритмы: пути, потоки и связность 15. Динамическое программирование: принципы и примеры 16. Рекурсивные алгоритмы: преимущества и недостатки 17. Алгоритмы шифрования и криптографии 18. Параллельные и распределенные алгоритмы 19. Эвристические алгоритмы и методы оптимизации 20. Анализ сложности алгоритмов: Big O и другие нотации 21. Модель взаимодействия открытых систем (OSI) ISO 22. Семейство протоколов IEEE 802.11 (WiFi) 23. Стек протоколов TCP/IP v4 24. Протокол IPv6 25. Технология Network Address Translation 26. Динамическая маршрутизация 27. Автономные системы и маршрутизация в Internet 28. Протокол HTTP 29. Протокол BitTorrent 30. TOR (The Onion Router)

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=514>

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
Программный инструментарий машинного обучения и обработки данных	1. Разведочный анализ данных 2. Задача бинарной классификации сбалансированного набора данных 3. Задача многоклассовой классификации сбалансированного набора данных 4. Задача классификации несбалансированного набора данных 5. Задача восстановления регрессии 6. Задача кластеризации 7. Задача выявления аномалий 8. Задача прогнозирования временных рядов
Работа с данными. Алгоритмы. Сети	1. Проектирование базы данных 2. Установка соединения с сервером Microsoft SQL Server и принципы создания баз данных 3. Разработка таблиц и ограничений 4. Введение в язык SQL. Создание таблиц и ограничений на SQL 5. Создание запросов на выборку. Отбор строк по условию 6. Создание многотабличных запросов. Запросы на соединение 7. Создание запросов на группировку и сортировку данных. Запросы на изменение. Использование встроенных функций 8. Создание и управление представлениями 9. Основы программирования с помощью встроенного языка Transact-SQL в Microsoft SQL Server

	10. Создание, изменение, применение и удаление функций и хранимых процедур 11. Создание, программирование и управление триггерами 12. Создание и управление транзакциями 13. Создание, применение и управление курсорами 14. Система безопасности SQL Server 15. Администрирование сервера баз данных MS SQL Server
--	--

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 машинного обучения и обработки данных	1. Введение в машинное обучение и обработку данных. Постановка основных классов задач в машинном обучении. 2. Обучение с учителем; регрессия и классификация; обучение без учителя; кластеризация, снижение размерности; рекомендательные системы; обработка текстов; обработка изображений; обучение с подкреплением. 3. Программный инструментарий Data Mining и Machine Learning. 4. Основной аппарат комбинаторики и мат. статистики.

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

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 4-х балльной системы
«отлично»	Повышенный ПК-2.1, ПК-2.2, ПК-2.3, ПК-3.1, ПК-3.2, ПК-3.3
«хорошо»	Стандартный ПК-2.1, ПК-2.2, ПК-2.3, ПК-3.1, ПК-3.2, ПК-3.3
«удовлетворительно»	Пороговый ПК-2.1, ПК-2.2, ПК-2.3, ПК-3.1, ПК-3.2, ПК-3.3
«неудовлетворительно»	Результаты обучения не сформированы на пороговом уровне