

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Кандрашина Елена Александровна

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

Дата подписания: 29.06.2022 10:08:10

Уникальный программный ключ:

2db64eb9605ce27edd3b8e8fdd32c70e0674ddd2

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 № 9 от 31 мая 2022 г.)

РАБОЧАЯ ПРОГРАММА

Наименование дисциплины	Б1.В.ДЭ.04.01 Инструменты анализа данных и машинного обучения
Основная профессиональная образовательная программа	09.04.03 Прикладная информатика программа Искусственный интеллект и большие данные

Квалификация (степень) выпускника магистр

Самара 2022

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 Инструменты анализа данных и машинного обучения входит в часть, формируемая участниками образовательных отношений (дисциплина по выбору) блока Б1. Дисциплины (модули)

Предшествующие дисциплины по связям компетенций: Современные проблемы и тренды прикладной информатики, Системы обработки и анализа больших массивов данных, Интеллектуальное планирование, Методы оптимизации

Последующие дисциплины по связям компетенций: Управление проектной деятельностью в профессиональной сфере, Тестирование искусственного интеллекта, Интеллектуальные информационные системы, Анализ прикладных систем, Машинное обучение на больших данных, Управление проектами создания интеллектуальных информационных систем

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

Изучение дисциплины Инструменты анализа данных и машинного обучения в образовательной программе направлено на формирование у обучающихся следующих компетенций:

Универсальные компетенции (УК):

УК-1 - Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
УК-1	УК-1.1: Знать:	УК-1.2: Уметь:	УК-1.3: Владеть (иметь навыки):
	особенности осуществления критического анализа проблемных ситуаций на основе системного подхода, выработки стратегии действий	осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	навыками осуществления критического анализа проблемных ситуаций на основе системного подхода, выработки стратегии действий

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-4 - Способен сравнивать фактическое исполнение проекта с планами работ по проекту

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
ПК-4	ПК-4.1: Знать:	ПК-4.2: Уметь:	ПК-4.3: Владеть (иметь

			навыки):
	особенности исполнения ИТ-проекта, построение планов работ ИТ-проекта	сравнивать фактическое исполнение проекта с планами работ по проекту	навыками исполнения ИТ-проекта, построения планов работ ИТ-проекта

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 2
Контактная работа, в том числе:	12.15/0.34
Занятия лекционного типа	4/0.11
Занятия семинарского типа	8/0.22
Индивидуальная контактная работа (ИКР)	0.15/0
Самостоятельная работа:	113.85/3.16
Промежуточная аттестация	18/0.5
Вид промежуточной аттестации: Зачет	Зач
Общая трудоемкость (объем части образовательной программы): Часы	144
Зачетные единицы	4

заочная форма

Виды учебной работы	Всего час/ з.е.
	Сем 3
Контактная работа, в том числе:	12.15/0.34
Занятия лекционного типа	4/0.11
Занятия семинарского типа	8/0.22
Индивидуальная контактная работа (ИКР)	0.15/0
Самостоятельная работа:	113.85/3.16
Промежуточная аттестация	18/0.5
Вид промежуточной аттестации: Зачет	Зач
Общая трудоемкость (объем части образовательной программы): Часы	144
Зачетные единицы	4

4. Содержание дисциплины

4.1. Разделы, темы дисциплины и виды занятий:

Тематический план дисциплины Инструменты анализа данных и машинного обучения 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		
			Практич. занятия			

1.	Большие данные и машинное обучение. Метрические методы классификации	2	4			50	УК-1.1, УК-1.2, УК-1.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК-4.1, ПК-4.2, ПК-4.3
2.	Логические и линейные методы классификации	2	4			63,85	УК-1.1, УК-1.2, УК-1.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК-4.1, ПК-4.2, ПК-4.3
Контроль		18					
Итого		4	8	0.15		113.85	

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.	Большие данные и машинное обучение. Метрические методы классификации	2	4			50	УК-1.1, УК-1.2, УК-1.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК-4.1, ПК-4.2, ПК-4.3
2.	Логические и линейные методы классификации	2	4			63,85	УК-1.1, УК-1.2, УК-1.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК-4.1, ПК-4.2, ПК-4.3
Контроль		18					
Итого		4	8	0.15		113.85	

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.	Большие данные и машинное обучение. Метрические методы классификации	лекция	Большие данные и машинное обучение. Метрические методы классификации
2.	Логические и линейные методы классификации	лекция	Логические и линейные методы классификации

*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.	Большие данные и	практическое занятие	Ответы и типы задач машинного

	машинное обучение. Метрические методы классификации		обучения. Модель алгоритмов. Метод обучения. Этап обучения и этап применения. Функционалы качества. Сведение задачи обучения к задаче оптимизации. Переобучение и обобщение. Пример переобучения (Рунге). Эмпирические оценки обобщающей способности. Примеры задач машинного обучения: задачи классификации и регрессии; задачи ранжирования. Эксперименты в машинном обучении: эксперименты на реальных и синтетических данных.
		практическое занятие	Метод ближайшего соседа. Метод взвешенных ближайших соседей. Метод парзеновского окна. Метод потенциальных функций. Отбор эталонных объектов. Понятие отступа объекта. Типы объектов в зависимости от отступа. Отбор эталонов, алгоритм STOLP. Задача выбора метрики. Жадное добавление признаков.
2.	Логические и линейные методы классификации	практическое занятие	Определение бинарного решающего дерева. Жадный алгоритм построения дерева ID 3. Варианты критериев ветвления в ID 3. Обработка пропусков, алгоритм обработки пропусков на этапе обучения и этапе классификации. Алгоритм ID3: достоинства и недостатки. Стратегии редукции решающих деревьев. Небрежные решающие деревья. Бинаризация вещественного признака.
		практическое занятие	Принцип максимума правдоподобия. Оптимальная разделяющая гиперплоскость. Метод SVM. Нелинейное обобщение SVM.

** 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.	Большие данные и машинное обучение. Метрические методы классификации	- тестирование
2.	Логические и линейные методы классификации	- 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, 2020. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450262>

Дополнительная литература

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2021. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/469022>

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. 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 Инструменты анализа данных и машинного обучения:

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, утвержденными Ученым советом ФГАОУ ВО СГЭУ, протокол № 9 от 31.05.2022; 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-1 - Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
	УК-1.1: Знать:	УК-1.2: Уметь:	УК-1.3: Владеть (иметь навыки):
	особенности осуществления критического анализа проблемных ситуаций на основе системного подхода, выработки стратегии действий	осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий	навыками осуществления критического анализа проблемных ситуаций на основе системного подхода, выработки стратегии действий
Пороговый	основы интеллектуального анализа данных	основные способы визуализации данных (гистограммами, диаграммами плотности, диаграммами рассеяния, ящиками с усами и т.п.), реализованные в библиотеках matplotlib, seaborn	методы понижения размерности данных, реализованные в библиотеке sklearn
Стандартный (в дополнение к пороговому)	решать нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных, социально-экономических и профессиональных знаний	проводить разведочный анализ данных, проводить предобработку и очистку данных, работать с пропущенными значениями	обоснованно выбирать наиболее подходящие алгоритмы решения задач машинного обучения и оценивать качество построенных моделей
Повышенный (в дополнение к пороговому, стандартному)	базовыми инструментами анализа данных и решения задач машинного обучения, реализованными	навыками применения новых научных принципов и методов исследования для	Навыками составления композиции моделей (блендинг, стеккинг), проведения отбора

	в библиотеках pandas и sklearn.	решения профессиональных задач	признаков
--	---------------------------------	--------------------------------	-----------

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	возможности, условия применимости и свойства наиболее распространенных методов машинного обучения при построении, проверке качества и эксплуатации формальных математических моделей;	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, извлечением знаний из баз 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 (переформатирования, устранения выбросов, заполнения пропусков, шкалирования, агрегации);	планировать исследование, основывающееся на анализе прецедентов и направленное на предсказательное моделирование;	навыками использования высокоуровневых программных средств для предварительной обработки исходных данных ;

ПК-4 - Способен сравнивать фактическое исполнение проекта с планами работ по проекту

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
	ПК-4.1: Знать:	ПК-4.2: Уметь:	ПК-4.3: Владеть (иметь навыки):
	особенности исполнения ИТ-проекта, построение планов работ ИТ-проекта	сравнивать фактическое исполнение проекта с планами работ по проекту	навыками исполнения ИТ-проекта, построения планов работ ИТ-проекта
Пороговый	методы анализа многомерных данных; методы снижения размерности данных и отбора информативных признаков;	применять методы машинного обучения при решении задач построения формальных математических моделей в различных прикладных областях;	навыками использования высокоуровневых программных средств для решения типичных задач машинного обучения: кластеризации, классификации, регрессии
Стандартный (в дополнение к	методы кластеризации; методы классификации;	использовать различные программные системы для	Навыками применения математических моделей в

пороговому)	методы регрессионного анализа;	построения и эксплуатации моделей машинного обучения; использовать формальные математические модели для имитационного моделирования в режиме "что-если";	рамках специализированных программных продуктов
Повышенный (в дополнение к пороговому, стандартному)	иноязычную терминологию в области машинного обучения; международные стандарты в области машинного обучения;	анализировать многомерные данные и преодолевать вычислительные проблемы связанные с высокой размерностью данных; пользоваться иноязычной литературой и электронными ресурсами в области машинного обучения;	Навыками пользования иноязычной литературой и электронными ресурсами в области машинного обучения

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.	Большие данные и машинное обучение. Метрические методы классификации	УК-1.1, УК-1.2, УК-1.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК-4.1, ПК-4.2, ПК-4.3	Оценка контрольных работ (для заочной формы обучения) Тестирование	Зачет
2.	Логические и линейные методы классификации	УК-1.1, УК-1.2, УК-1.3, ПК-2.1, ПК-2.2, ПК-2.3, ПК-4.1, ПК-4.2, ПК-4.3	Оценка контрольных работ (для заочной формы обучения) Тестирование	Зачет

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 числу 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
Большие данные и машинное обучение. Метрические методы классификации	<ol style="list-style-type: none">1. Примеры использования методов машинного обучения для практических задач.2. Краткий обзор синтаксиса языка Python.3. Встроенные операции и функции, типы и структуры данных. Библиотеки NumPy и SciPy.4. Матрицы. Разреженные матрицы.5. Индексирование, срезы.6. Объединение массивов.7. Библиотека pandas.8. Запросы к таблицам: выборка строк/столбцов по заданным критериям. Модификация элементов таблицы.9. Добавление строк/столбцов.10. Группировка и агрегирование.11. Объединение таблиц (различные виды join).12. Многомерные данные: мультииндексы.13. Операции stack-unstack.14. Построение сводных таблиц (pivot tables).
Логические и линейные методы классификации	<ol style="list-style-type: none">1. Описательные статистики.2. Обзор библиотек matplotlib, seaborn, bokeh.3. Базовые типы визуализации данных.4. Знакомство с библиотекой scikit-learn (sklearn). Предобработка данных5. Метод главных компонент.6. Работа с пропущенными значениями.

6.5. Оценочные материалы для промежуточной аттестации

Фонд вопросов для проведения промежуточного контроля в форме зачета

Раздел дисциплины	Вопросы
	1. Специализированные библиотеки языка программирования Python

<p>Большие данные и машинное обучение. Метрические методы классификации</p>	<p>для научных расчетов и анализа данных. 2. NumPy, SciPy, pandas. 3. Методы предобработки данных, описательные статистики и основные способы визуализации данных. 4. Методы снижения размерности. 5. Метод главных компонент. 6. Важность нормировки данных. 7. Предобработка данных. 8. Работа с пропущенными значениями. 9. Основы машинного обучения и основные типы задач. 10. Классификация задач машинного обучения.</p>
<p>Логические и линейные методы классификации</p>	<ol style="list-style-type: none"> 1. 1. Обучение на неразмеченных данных. 2. 2. Кластеризация. 3. 3. Иерархическая кластеризация. 4. 4. Метод K-средних, DBSCAN и др. 5. 5. Обзор методов кластеризации, реализованных в библиотеке sklearn. 6. Задачи обучения с учителем. 7. Разделение данных на обучающие и тестовые. 8. Нормировка данных. 9. Определение переобученности модели. 10. Критерии оценки качества полученных моделей. 11. Постановка задачи регрессии. Линейный регрессионный анализ. 12. Отбор признаков, коллинеарность, влиятельные наблюдения, анализ остатков. 13. Непараметрическая регрессия (ядерное сглаживание). L1 и L2 регуляризация. Метрики качества. 14. Постановка задачи классификации, обзор основных методов ее решения. 15. Бинарная и многоклассовая классификация. 16. Логистическая регрессия. 17. Решающие деревья. 18. Метрики качества классификации (точность/специфичность, ROC-кривая, площадь под кривой). 19. Ансамбли алгоритмов машинного обучения. Агрегирование моделей. 20. Ансамбли решающих деревьев. Метод случайного леса. Градиентный бустинг.

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 2-х балльной системы
«зачтено»	УК-1, ПК-2, ПК-4
«не зачтено»	Результаты обучения не сформированы на пороговом уровне