

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Кандрашина Елена Александровна

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

Дата подписания: 22.11.2022 09:38:51

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 № 9 от 31 мая 2022 г.)

РАБОЧАЯ ПРОГРАММА

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

Квалификация (степень) выпускника Бакалавр

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

Профессиональные компетенции (ПК):

ПК-1 - Способность 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: особенности подготовки коммерческого	ПК-1.2: Уметь: готовить коммерческое предложение заказчику	ПК-1.3: Владеть (иметь навыки): навыками подготовки коммерческого предложения

	предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС	по созданию (модификации) и вводу в эксплуатацию ИС	заказчику по созданию (модификации) и вводу в эксплуатацию ИС
--	---	---	---

ПК-4 - Способность 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:	ПК-4.3: Владеть (иметь навыки):
ПК-4	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		
			Лаборат. работы				
1.	Машинное обучения при анализе больших массивов данных	9	18	0,075		17,925	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.	Современные технологии анализа больших массивов данных	9	18	0,075		17,925	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					
	Итого	18	36	0.15		35.85	

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		
			Лаборат. работы				
1.	Машинное обучения при анализе больших массивов данных	1	1	0,075		42,925	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.	Современные технологии анализа больших массивов данных	1	1	0,075		42,925	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					
	Итого	2	2	0.15		85.85	

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.	Машинное обучения при анализе больших массивов данных	лекция	Основы машинного обучения.
		лекция	Обучение без учителя
		лекция	Обучение с учителем
		лекция	Обучение по прецедентам

2.	Современные технологии анализа больших массивов данных	лекция	Визуализация на основе нейронных сетей при анализе больших массивов данных
		лекция	Кластеризация на основе нейронных сетей при анализе больших массивов данных
		лекция	Регрессионный анализ данных при анализе больших массивов данных
		лекция	Обнаружение аномалий в данных при анализе больших массивов данных

*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.	Машинное обучения при анализе больших массивов данных	лабораторные работы	Кластеризация методом к-средних
		лабораторные работы	Метод главных компонент
		лабораторные работы	Ассоциативные правила
		лабораторные работы	Нейронные сети, обучение с подкреплением
2.	Современные технологии анализа больших массивов данных	лабораторные работы	Регрессионный анализ
		лабораторные работы	Метод опорных векторов
		лабораторные работы	Дерево решений
		лабораторные работы	Случайный лес

** 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.	Машинное обучения при анализе больших массивов данных	- подготовка доклада - подготовка электронной презентации - тестирование
2.	Современные технологии анализа больших массивов данных	- 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. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2022. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/489100>

Дополнительная литература

1. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2022. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/469306>

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, 2022. — 246 с. — (Высшее образование). — ISBN 978-5-9916-8262-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/498879>

5.2. Перечень лицензионного программного обеспечения

Обязательное программное обеспечение

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

Программное обеспечение по выбору

1. 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	Комплекты ученической мебели

и промежуточной аттестации	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, 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-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 задач с учетом правовых	УК-2.2: Уметь: определять круг задач в рамках поставленной цели и выбирать

	норм, имеющих ресурсы и ограничений	оптимальные способы их решения, исходя из действующих правовых норм, имеющих ресурсы и ограничений	оптимальных способов их решения
Пороговый	оптимальные способы решения поставленных задач с учетом правовых норм, имеющих ресурсы и ограничений	определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющих ресурсы и ограничений	навыками определения круга задач в рамках поставленной цели, выбора оптимальных способов их решения
Стандартный (в дополнение к пороговому)	оптимальные способы решения поставленных задач с учетом правовых норм, имеющих ресурсы и ограничений в инструментальных средствах	определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющих ресурсы и ограничений в инструментальных средствах	навыками определения круга задач в рамках поставленной цели, выбора оптимальных способов их решения в инструментальных средствах
Повышенный (в дополнение к пороговому, стандартному)	оптимальные способы решения поставленных задач с учетом правовых норм, имеющих ресурсы и ограничений используя типовые методики	определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющих ресурсы и ограничений используя типовые методики	навыками определения круга задач в рамках поставленной цели, выбора оптимальных способов их решения используя современные методики

Профессиональные компетенции (ПК):

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

Планируемые результаты обучения по программе	Планируемые результаты обучения по дисциплине		
	ПК-1.1: 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	особенности подготовки коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС с элементами искусственного	готовить коммерческое предложение заказчику по созданию (модификации) и вводу в эксплуатацию ИС с элементами искусственного интеллекта	навыками подготовки коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС с элементами искусственного интеллекта

	интеллекта		
Стандартный (в дополнение к пороговому)	особенности подготовки коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС с элементами искусственного интеллекта в виде искусственных нейронных сетей	готовить коммерческое предложение заказчику по созданию (модификации) и вводу в эксплуатацию ИС с элементами искусственного интеллекта в виде искусственных нейронных сетей	навыками подготовки коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС с элементами искусственного интеллекта в виде искусственных нейронных сетей
Повышенный (в дополнение к пороговому, стандартному)	особенности подготовки коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС с элементами искусственного интеллекта в виде искусственных нейронных сетей для анализа больших данных	готовить коммерческое предложение заказчику по созданию (модификации) и вводу в эксплуатацию ИС с элементами искусственного интеллекта в виде искусственных нейронных сетей для анализа больших данных	навыками подготовки коммерческого предложения заказчику по созданию (модификации) и вводу в эксплуатацию ИС с элементами искусственного интеллекта в виде искусственных нейронных сетей для анализа больших данных

ПК-4 - Способность 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:	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
Пороговый	особенности верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС с элементами искусственного интеллекта	верифицировать структуру программного кода ИС относительно архитектуры ИС и требований заказчика к ИС с элементами искусственного интеллекта	навыками верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС с элементами искусственного интеллекта
Стандартный (в дополнение к пороговому)	особенности верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС с элементами искусственного	верифицировать структуру программного кода ИС относительно архитектуры ИС и требований заказчика к ИС в виде искусственных нейронных сетей	навыками верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС в виде искусственных нейронных сетей

	интеллекта в виде искусственных нейронных сетей		
Повышенный (в дополнение к пороговому, стандартному)	особенности верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС с элементами искусственного интеллекта в виде искусственных нейронных сетей для анализа больших данных	верифицировать структуру программного кода ИС относительно архитектуры ИС и требований заказчика к ИС с элементами искусственного интеллекта в виде искусственных нейронных сетей для анализа больших данных	навыками верификации структуры программного кода ИС относительно архитектуры ИС и требований заказчика к ИС с элементами искусственного интеллекта в виде искусственных нейронных сетей для анализа больших 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.	Машинное обучения при анализе больших массивов 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	Оценка практических работ Тестирование Оценка курсовых проектов	Экзамен
2.	Современные технологии анализа больших массивов 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	Оценка практических работ Тестирование Оценка курсовых проектов	Экзамен

6.4.Оценочные материалы для текущего контроля

Примерная тематика докладов

Раздел дисциплины	Темы
Машинное обучения при анализе больших массивов данных	Объекты, признаки, ответы, функционал качества. Вероятностная постановка задачи обучения. Переобучение, обобщающая способность. Задачи классификации, восстановления регрессии, ранжирования, кластеризации, поиска ассоциаций.
Современные технологии анализа больших массивов данных	Качество кластеризации. Статистические, эвристические методы кластеризации. Иерархическая кластеризация. Сети Кохонена. t-SNE. PCA.

Вопросы для устного/письменного опроса

Раздел дисциплины	Вопросы
Машинное обучения при анализе больших массивов данных	Обобщённый метрический классификатор. Метод ближайшего соседа. KNN. Метод

массивов данных	потенциальных функций. Метод парзеновского окна. Понятие отступа. Понятие эталонного объекта.
Современные технологии анализа больших массивов данных	Модель нейрона. Многослойный перцептрон. Метод обратного распространения ошибки. Свёрточные нейронные сети. Рекуррентные нейронные сети.

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>

1.

Для распределенного глубокого машинного обучения (Deep Learning) больше подходит фреймворк

PyTorch
TensorFlow
Scikit-learn
Flask

2.

Формат Parquet считается

колоночным (столбцовым)
строковым
неструктурированным
полуструктурированным

3.

Автоматизировать запуск пакетных задач в рамках конвейера обработки больших данных по расписанию можно с помощью

Apache Hive
Apache AirFlow
Apache Kafka
Apache Hadoop

4.

Выберите технологию потоковой обработки событий в режиме реального времени

Apache Kafka
Spark Streaming
Apache Hadoop
MapReduce

5.

Анализировать данные, хранящиеся в Apache Hadoop, с помощью стандартного инструментария SQL-запросов

Нельзя
Можно

6.

Повысить производительность Apache Kafka можно с помощью:

Увеличения размера сообщений
Замены HDD-дисков на SSD
Увеличения плотности разделов на каждом брокере
Повышения коэффициента репликации

7.

Для реализации микросервисной архитектуры и интеграции разрозненных систем подходит

Apache Hadoop
Apache Kafka
Apache AirFlow
Apache Spark

8.

Apache NiFi используется для

оптимизации SQL-запросов к DWH
маршрутизации потоков Big Data и построения ETL-конвейеров
визуализации результатов аналитики
эффективного хранения больших данных

9.

Для машинного обучения подходят данные

Предварительно подготовленные, очищенные от ошибок, пропусков и выбросов, а также нормализованные и представленные в виде числовых векторов

Любых форматов в цифровом виде

Бинарные

Числовые типа int

10.

Для полнотекстового интеллектуального поиска и аналитики по полуструктурированным данным в формате JSON отлично подходит СУБД

Elasticsearch

Hive

Cassandra

HBase

11

Что, из ниже перечисленного, относится к обучающей выборке?

классификация данных

объекты с известными ответами

алгоритм решающий функцию

12

Объекты состоят из признаков?

Да

Нет

13

Что называют данными в машинном обучении?

матрицы

объекты

признаки

алгоритм

функция

14

Выберите правильный ответ. Задача классификации - это:
множество объектов, разделенных на классы

исследование влияние одного или нескольких признаков на объект
определение порядка признака согласно рангу

15

Выберите правильный ответ. Задача регрессии - это:

множество объектов, разделенных на классы

исследование влияние одного или нескольких признаков на объект

определение порядка признака согласно рангу

16

Кроме рисования искусственный интеллект научился разбираться в музыке. Но насколько хорошо работает программа по определению музыкальных стилей?

Да, лучше, чем программа, написанная вручную

Да, но программа написанная вручную будет точнее

Нет, в режиме реального времени программа не справится

17

Способен ли ИИ сравниться с человеческим интеллектом в общей совокупности способностей?

Нет, пока этот уровень недостижим

Нет, но технология уже близка к уровню мозга человека

Да, и уже идёт работа над созданием искусственного сверхума

18

К какому времени относятся первые упоминания об искусственно созданных человекоподобных существах?

Период до н.э. (в источниках Древнего Египта)

Период Возрождения (в трудах Да Винчи)

XX век (в материалах, появившихся после возникновения обучающих алгоритмов)

19

Новая программа от Google научилась рисовать на основе эскизов, сделанных людьми. Что при этом учитывала программа?

Стиль и цветовую гамму, типичную для похожих изображений

Только конечный результат

Концепцию (идею) рисунка

20

Выберите правильный ответ. Задача ранжирования - это:

множество объектов, разделенных на классы

исследование влияние одного или нескольких признаков на объект

определение порядка признака согласно рангу

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

Раздел дисциплины	Задачи
Машинное обучения при анализе больших массивов данных	Обобщённый метрический классификатор. Метод ближайшего соседа. KNN. Метод потенциальных функций. Метод парзеновского окна. Понятие отступа. Понятие эталонного объекта.
Современные технологии анализа больших массивов данных	Модель нейрона. Многослоный перцептрон. Метод обратного распространения ошибки. Свёрточные нейронные сети. Рекуррентные нейронные сети.

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
<p>Основы машинного обучения. Алгоритмы классификации</p>	<ol style="list-style-type: none"> 1. Основные определения в машинном обучении: объект, целевая переменная, признак, модель, обучающая выборка, функционал ошибки, обучение, переобучение. Приведите примеры задач обучения с учителем и без учителя. 2. Линейная модель регрессии. Аналитическое решение для среднеквадратичной ошибки (с выводом). Градиентное обучение линейной регрессии. 3. Функционалы ошибки для регрессии: MSE, MAE, коэффициент детерминации. MSE как максимум правдоподобия. 4. Градиентные методы обучения. Свойство градиента о направлении наискорейшего убывания. Градиентный спуск. Методы оценивания градиента. 5. Сингулярное разложение матриц. Сингулярные числа. Сингулярное разложение в задаче регрессии со среднеквадратичной ошибкой. 6. L-2 регуляризация. Аналитический вид вектора весов в линейной регрессии со среднеквадратичной ошибкой и L-2 регуляризатором (с выводом). 7. L1-регуляризация. Почему использование L1-регуляризатора приводит к отбору признаков? 8. Линейная модель классификации. Отступ. Обучение линейных классификаторов через верхнюю оценку на долю ошибок. Примеры верхних оценок. 9. Функционалы ошибки для классификации: матрица ошибок, accuracy, precision, recall, F-мера. ROC-кривая и AUC-ROC. Precision-recall-кривая и площадь под ней.
<p>Глубокое машинное обучение</p>	<ol style="list-style-type: none"> 10. Логистическая регрессия. Оценивание вероятностей. 11. Метод опорных векторов. Вывод постановки задачи для разделимого и неразделимого случаев. 12. Использование ядер в методе опорных векторов. Теорема Мерсера. 13. Многоклассовая классификация: one-vs-all, all-vs-all. Многоклассовая логистическая регрессия. Микро и макро – усреднения. 14. Решающие деревья: определение и жадный алгоритм обучения. Функционал качества при выборе предиката. Общий вид критерия информативности (через функцию потерь) и конкретные примеры для регрессии (дисперсия) и классификации (критерий Джини и энтропийный критерий). 15. Композиции алгоритмов. Разложение ошибки на смещение и разброс (с выводом). 16. Бэггинг и случайные леса. 17. Градиентный бустинг. Обучение базовых алгоритмов для произвольной дифференцируемой функции потерь. Сокращение шага. 18. Нейронные сети, их обучение методом обратного распространения ошибки. 19. Задача кластеризации. Карты Кахоннена. 20. Нелинейные методы классификации: наивный байесовский классификатор, метод Kmeans.

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