

Документ подписан простой электронной подписью Информация о владельце: ФИО: Таскаев Сергей Валерьевич Должность: Ректор	МИНИСТЕРСТВО НАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)	
Дата подписания: 08.04.2026 16:25:40 Уникальный идентификатор: 04c19ed8bfb98f3b6cb77a486b9a8788b8322323	Рабочая программа дисциплины "Методы машинного обучения" по направлению подготовки (специальности) 01.03.02 Прикладная математика и информатика" направленности (профилю) Прикладная математика и искусственный интеллект ФГБОУ ВО «ЧелГУ»	стр. 1

Рабочая программа дисциплины (модуля)*

Методы машинного обучения

Направление подготовки (специальность)

01.03.02 Прикладная математика и информатика

Направленность (профиль)

Прикладная математика и искусственный интеллект

Присваиваемая квалификация (степень)

бакалавр

Форма обучения

очная

Год набора 2026

*Рабочая программа дисциплины (модуля) адаптирована для инклюзивного обучения инвалидов и лиц с ограниченными возможностями здоровья

Челябинск 2026 г.



Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре ОПОП
3. Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля)
4. Объем дисциплины (модуля)
5. Структура и содержание дисциплины (модуля)
6. Фонд оценочных средств
 - 6.1. Перечень видов оценочных средств
 - 6.2. Типовые контрольные задания и иные материалы для текущей аттестации
 - 6.3. Типовые контрольные вопросы и задания для промежуточной аттестации
 - 6.4. Критерии оценивания
7. Учебно-методическое и информационное обеспечение дисциплины (модуля)
 - 7.1. Рекомендуемая литература
 - 7.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"
 - 7.3. Перечень информационных технологий
8. Материально-техническое обеспечение дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Специальные условия освоения дисциплины обучающимися с инвалидностью и ограниченными возможностями здоровья



1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

формирование теоретических знаний по основам машинного обучения, типам и классам задач машинного обучения; выработка умения по практическому применению методов машинного обучения для построения формальных математических моделей и интерпретации результатов моделирования; формирование умения использования различных программных инструментов анализа баз данных и систем машинного обучения при разработке алгоритмов для решения задач машинного обучения. Изучение дисциплины направлено на развитие следующих индикаторов:

ПК-4.1. Проводит анализ требований и определяет необходимые классы задач искусственного интеллекта;

Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи;

ПК-4.2. Определяет метрики оценки результатов моделирования и критерии качества построенных моделей;

Разрабатывает модели искусственного интеллекта для решения задач;

ПК-4.3. Принимает участие в оценке и выборе методов искусственного интеллекта;

Создаёт, поддерживает и использует системы искусственного интеллекта, включающие разработанные модели и методы с применением выбранных инструментов.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Цикл (раздел) ОПОП:

Б1.В.1.05

2.1 Требования к предварительной подготовке обучающегося:

Знает: [ПК-5.1. 3-1.] возможности современных инструментальных средств и систем программирования для решения задач анализа данных и машинного обучения Умеет: [ПК-2.2. У-1.] разрабатывать программные приложения систем искусственного интеллекта, с использованием современных языков программирования, библиотек и программных платформ функционального, логического, объектно-ориентированного программирования (Python, R, C++, C#)

Имеет практический опыт: участия в разработке программных приложений систем искусственного интеллекта, с использованием современных языков программирования

Технология программирования на языке C++

Учебная практика (практика по программированию)

Современные технологии поиска и обработки информации

Теория кодов

2.2 Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:

Выполнение и защита выпускной квалификационной работы

Методы анализа данных (научный семинар)

Теория нечетких множеств в системах искусственного интеллекта

Разработка приложений для интеллектуальных систем

Основы компьютерного зрения

Математическая статистика

Искусственные нейронные сети (научный семинар)

Технология баз данных

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

ПК-4: Способен разрабатывать и применять методы искусственного интеллекта для решения задач, в том числе используя инструментальные средства

Знать:

Для достижения ПК-4.1. студенту необходимо знать: как проводить анализ требований и определять необходимые классы задач в области машинного обучения.

Уметь:

Для достижения ПК-4.2. студенту необходимо уметь: определять метрики оценки результатов моделирования и критерии качества построенных моделей;



уметь разрабатывать модели машинного обучения для решения задач;

Владеть:

Для достижения ПК-4.3. студенту необходимо владеть оптом: участия в оценке и выборе методов машинного обучения для решения конкретных задач; создания, поддержки и использования методов машинного обучения, включающие разработанные модели и методы с применением выбранных инструментов.

В результате освоения дисциплины обучающийся должен

3.1 Знать:	
3.1.1	принципы и методы машинного обучения, типы и классы задач машинного обучения, методологию ML Ops;
3.1.2	статистические методы анализа данных;
3.1.3	классические методы и алгоритмы машинного обучения: предиктивные - обучение с учителем, дескриптивные - обучение без учителя
3.2 Уметь:	
3.2.1	сопоставить задачам предметной области классы задач машинного обучения;
3.2.2	использовать статистические методы анализа данных при решении задач машинного обучения;
3.2.3	проводить сравнительный анализ и осуществлять выбор, настройку при необходимости, разработку методов и алгоритмов для решения задач машинного обучения
3.2.4	проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения
3.3 Владеть:	
3.3.1	Имеет практический опыт: участия в разработке алгоритмов для решения задач машинного обучения

4. ОБЪЕМ ДИСЦИПЛИНЫ (МОДУЛЯ)

Общая трудоемкость	3 ЗЕТ
Часов по учебному плану : 108 в том числе : аудиторные занятия : 50 самостоятельная работа : 18,7 часов на контроль : 36 контактная работа: 53,3 ИКР: 3,3	Виды контроля в семестрах: экзамены 4

5. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Литература
	Раздел 1. Введение в машинное обучение			
1.1	Введение в машинное обучение. Основные понятия. Определение предмета машинного обучения. Примеры задач и областей приложения. Образы и признаки. /Лек/	4	2	Л1.1 Л1.2Л2.1
1.2	Типы и классы задач машинного обучения. Задача регрессии. Задача классификации. Задача кластеризации. Задача уменьшения размерности. Задача выявления аномалий. /Лек/	4	2	Л1.1 Л1.2
1.3	Построение модели и сведение обучения к задаче оптимизации. Оценка качества работы алгоритма машинного обучения /Лек/	4	2	Л1.1 Л1.2
1.4	Лабораторная работа 1. Визуализация данных /Лаб/	4	2	Л1.1 Л1.2
1.5	Подготовка к лабораторной работе /Ср/	4	1,5	Л1.1 Л1.2
	Раздел 2. Классические методы и алгоритмы машинного обучения. Деревья принятия решений. Регрессия. Кластеризация			



Рабочая программа дисциплины "Методы машинного обучения" по направлению подготовки (специальности) 01.03.02 "Прикладная математика и информатика" направленности (профилю) Прикладная математика и искусственный интеллект ФГБОУ ВО «ЧелГУ»				стр. 5
2.1	Ассоциативные правила. Алгоритм APRIORI. Алгоритм FP- Growth /Лек/	4	2	Л1.1 Л1.2
2.2	Кластеризация. Метод К-средних. DBSCAN. Нечеткая кластеризация. /Лек/	4	2	Л1.1 Л1.2
2.3	Линейные методы. Линейная регрессия. Линейная классификация. /Лек/	4	2	Л1.1 Л1.2
2.4	Метод Парзенковского окна. Деревья принятия решений /Лек/	4	2	Л1.1 Л1.2
2.5	Вероятностные модели. Байесовский классификатор. Наивный байесовский классификатор. /Лек/	4	2	Л1.1 Л1.2
2.6	Ансамблевые методы. Бэггинг. Бустинг. Стэкинг /Лек/	4	2	Л1.1 Л1.2
2.7	Лабораторная работа 2. Применение алгоритма APRIORI /Лаб/	4	2	Л1.1 Л1.2
2.8	Лабораторная работа 3. Применение алгоритма линейной регрессии. Теоретический срез. /Лаб/	4	2	Л1.1 Л1.2
2.9	Лабораторная работа 4. Ансамблевые методы машинного обучения /Лаб/	4	2	Л1.1 Л1.2
2.10	Подготовка к лабораторным работам /Ср/	4	3,5	Л1.1 Л1.2
Раздел 3. Обучение моделей машинного обучения				
3.1	Методы понижения размерности данных - метод главных компонент (PCA) и метод t-SNE. /Лек/	4	2	Л1.1 Л1.2Л2.1
3.2	Сжатие и визуализация данных. /Лек/	4	2	Л1.1 Л1.2
3.3	Метрики в задачах машинного обучения. /Лек/	4	2	Л1.1 Л1.2
3.4	Работа с несбалансированными классами. Обзор методов отбора признаков. /Лек/	4	2	Л1.1 Л1.2
3.5	Применение машинного обучения к прикладным задачам /Лек/	4	4	Л1.1 Л1.2
3.6	Лабораторная работа 5. Обнаружение мошеннических операций с кредитными картами. /Лаб/	4	3	Л1.1 Л1.2
3.7	Лабораторная работа 6. Применение анализа главных компонент (PCA) /Лаб/	4	3	Л1.1 Л1.2
3.8	Подготовка к лабораторным работам /Ср/	4	3,5	Л1.1 Л1.2Л2.1
Раздел 4. MLOps				
4.1	MLOps: самые известные модели /Лек/	4	4	Л1.1 Л1.2
4.2	Лабораторная работа 7. Применение MLOps для командной разработки программного решения /Лаб/	4	2	Л1.1 Л1.2
4.3	Подготовка к лабораторной работе /Ср/	4	3,5	Л1.1 Л1.2
4.4	Подготовка к экзамену /Ср/	4	6,7	Л1.1 Л1.2Л2.1
Раздел 5. Иная контактная работа				
5.1	Индивидуальные консультации, текущий контроль /ИКР /ИКР/	4	3,3	

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

6.1. Перечень видов оценочных средств

Лабораторные работы №1-7

Теоретический срез

Активная познавательная деятельность

6.2. Типовые контрольные задания и иные материалы для текущей аттестации

Теоретический срез

1) Какой код лучше использовать при загрузке данных из csv-файла?

а) dataset = read_csv("data.csv")

б) dataset = import("data.csv")

в) dataset = read.csv("data.csv")

г) dataset = import.csv("data.csv")

д) dataset = read_xls("data.csv")



- 2) Опишите принципы работы алгоритма APRIORI
- 3) Перечислите и опишите виды регрессионного анализа

Пример лабораторной работы:

Визуализация данных

1. Откройте дата сет об Ирисах. Рассмотрите основные признаки, представленный в наборе. Загрузите набор данных с использованием pandas и выведите признаки набора данных.
2. Для детального изучения признаков воспользуйтесь методом info() класса DataFrame
3. Для представления распределения простого количественного признака постройте обычная гистограмма, содержащаяся во всех библиотеках.
4. Визуализируйте соотношения количественных признаков является диаграмма по нескольким признакам.
5. Определите корреляции признаков обучающей выборки. В пакете Pandas имеется встроенный инструмент для этого – метод corr() класса DataFrame.

6.3. Типовые контрольные вопросы и задания для промежуточной аттестации

Примерные вопросы на экзамене.

1. Этапы решения задач машинного обучения
2. Основные определения в машинном обучении: объект, целевая функция, признак, модель, обучающая выборка, функционал качества, обучение, переобучение.
3. Задачи машинного обучения - обучение с учителем, без учителя.
4. Задачи регрессии и классификации. Задачи снижения размерности и кластеризации.
5. Типы признаков в машинном обучении. Приведите примеры различных признаков.
6. Определение ROC-кривой.
7. Метод k ближайших соседей в задаче классификации.
8. Методы отбора признаков. Жадный метод.
9. Определение отступа в метрических алгоритмах классификации. Алгоритм Condensed Nearest Neighbor.
10. Метод k ближайших соседей в задаче регрессии.
11. Обобщение метода k ближайших соседей через взвешенный учет объектов. Ядерная оценка плотности.
12. Проклятие размерности. Зависимость метода ближайших соседей от масштабирования признаков. Способы стандартизации признаков.
13. Вероятностная постановка задачи классификации. Функция правдоподобия и априорная вероятность.
14. Функционал среднего риска. Общая формула байесовского классификатора.
15. Наивный байесовский классификатор.
16. Восстановление плотности распределения по выборке.
17. Аддитивное сглаживание для байесова классификатора.
18. Модель МакКаллока-Питтса
19. Обобщённая модель линейного классификатора. Определение отступа. Минимизация эмпирического риска.
20. Метод градиентного спуска. Выбор величины шага.
21. Метод стохастического градиента. Недостатки метода SG и как с ними бороться.
22. Использование сингулярного разложения для решения задачи наименьших квадратов.
23. Нелинейная регрессия. Метод Ньютона-Гаусса
24. Задача уменьшения размерности. Метод главных компонент.

6.4. Критерии оценивания

7 лабораторных работ оцениваются максимально - 56 баллов.
Теоретический срез оценивается максимально - 10 баллов.
Активная познавательная деятельность оценивается максимально - 14 баллов.

На экзамене студент может получить максимально 30 баллов.

Если за семестр студент набирает:

1. от 61 до 78 баллов - оценка удовлетворительно,
2. от 79 до 88 баллов - оценка хорошо,
3. более 88 баллов - оценка отлично,
4. менее 61 балла - оценка неудовлетворительно.



7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1. Рекомендуемая литература

7.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Ресурс
Л1.1	Вохминцев А. В., Ботов Д. С., Шицелов А. В., Петриченко Ю. В.	Машинное обучение. Нейронные сети: практикум (https://library.csu.ru/rbooks2/view2?code=local/007890/007890)	Челябинск : Издательство Челябинского государственног о университета, 2022	ЭБС
Л1.2	Кугаевских А. В., Муромцев Д. И., Кирсанова О. В.	Классические методы машинного обучения: учебное пособие (https://e.lanbook.com/book/283928)	Санкт- Петербург : НИУ ИТМО, 2022	ЭБС

7.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Ресурс
Л2.1	Рашка С.	Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения (https://e.lanbook.com/book/100905)	Москва : ДМК Пресс, 2017	ЭБС

7.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

Э1	Лань [Электронный ресурс] : электронно-библиотечная система (ЭБС) / издательство Лань. URL: http://e.lanbook.com/ .
Э2	Университетская библиотека онлайн [Электронный ресурс] : электронно-библиотечная система (ЭБС) / ООО Директмедиа Паблишинг. URL: http://biblioclub.ru/
Э3	Znanium.com [Электронный ресурс] : электронно-библиотечная система (ЭБС) / Научно-издательский центр ИНФРА-М. URL: http://znanium.com/
Э4	Юрайт [Электронный ресурс] : электронно-библиотечная система (ЭБС) / издательство Юрайт. URL: https://urait.ru

7.3 Перечень информационных технологий

7.3.1 Программное обеспечение

Python

LMS Moodle

Visual Studio

LibreOffice

OpenOffice

PostgreSQL

7.3.2 Профессиональные базы данных и информационно-справочные системы

1. Электронный каталог научной библиотеки ЧелГУ [Электронный ресурс] : база данных / Челяб. гос. ун-т. – Челябинск, 1992 .
2. eLIBRARY.RU : научная электронная библиотека : сайт. – Москва, 2000 – . – URL: https://elibrary.ru (дата обращения: 28.02.2025). – Режим доступа: для зарегистрир. пользователей. – Текст : электронный.
3. Mathematical Reviews (MR) : реферативная база данных / American Mathematical Society. – URL: http://www.ams.org/mathscinet/ (дата обращения: 28.02.2025). – Яз. рус., англ. – Режим доступа: для зарегистрир. пользователей ЧелГУ. – Текст : электронный.

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Для реализации дисциплины используются учебные аудитории для проведения занятий лекционного типа, лабораторных занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы.



Учебные аудитории укомплектованы специализированной мебелью и техническими средствами обучения (компьютерная техника с подключением к сети "Интернет", предустановленным программным обеспечением MS Office, Python для лабораторных занятий).

Для проведения занятий лекционного типа предлагаются наборы демонстрационного оборудования и учебно-наглядных пособий, такие как презентации.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с подключением к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета.

9. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

В результате изучения теоретических основ дисциплины и реализации в процессе обучения требований по прикладной направленности дисциплины, а также самостоятельной работы студент должен знать, уметь и владеть составляющими компетенций, определенных в программе.

Основными видами занятий являются: лекции, практические занятия, самостоятельная работа и зачет.

Аудиторные лекции студента имеют своей целью формирование целостной системы знаний по изучаемому предмету.

Студент может воспользоваться основной и дополнительной литературой.

Самостоятельная работа студента начинается с внимательного ознакомления с программой данной дисциплины.

Требуется творческое отношение к самой Программе учебного курса. Вопросы, составляющие ее содержание, обладают разной степенью важности. Есть вопросы, выполняющие функцию логической связки содержания темы и всего курса, имеются вопросы описательного или разъяснительного характера. Эти вопросы не составляют сути, понятийного, концептуального содержания темы, но необходимы для целостного восприятия изучаемых проблем. Успешно освоив теоретический материал, студент будет готов к проведению практических заданий, которые рассматриваются как дальнейшее углубление и расширение знаний по предмету.

В освоении дисциплины (модуля) инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная работа. Под индивидуальной работой подразумевается две формы взаимодействия с преподавателем: индивидуальная учебная работа (консультации), т.е. дополнительное разъяснение учебного материала и углубленное изучение материала с теми обучающимися, которые в этом заинтересованы, и индивидуальная воспитательная работа.

Индивидуальные консультации по предмету является важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или обучающимся с ограниченными возможностями здоровья.

В случае применения при обучении дисциплины электронного обучения, дистанционных образовательных технологий общение обучающихся и преподавателя осуществляется в режиме реального времени (чаты, видео-конференции и др.) или отложенного времени (система дистанционного обучения Moodle, форумы, электронная почта).

Большую часть времени обучающиеся самостоятельно работают с учебно-методическими материалами. Студенты имеют возможность консультироваться с преподавателем по всем вопросам, возникающим в ходе самостоятельной работы посредством электронной почты, социальных сетей, Moodle.

Доступ обучающегося к учебным ресурсам в режиме отложенного времени, самостоятельной работы осуществляется через сеть Интернет в удобном для него месте, времени и темпе.

При обучении лиц с ограниченными возможностями здоровья электронное обучение, дистанционные образовательные технологии предусматривают возможность приема-передачи информации в доступных для них формах.

Реализация дисциплины с применением электронного обучения, дистанционных образовательных технологий (далее – ЭО, ДОТ) осуществляется на основании «Положения о реализации основных и дополнительных образовательных программ с применением электронного обучения и дистанционных образовательных технологий в федеральном государственном бюджетном образовательном учреждении высшего образования «Челябинский государственный университет», «Положения о порядке зачета обучающимися по основным профессиональным образовательным программам высшего образования в ФГБОУ ВО «ЧелГУ» результатов освоения в организациях, осуществляющих образовательную деятельность, учебных предметов, курсов, дисциплин (модулей), практик, дополнительных образовательных программ» посредством электронной информационно-образовательной среды ФГБОУ ВО «ЧелГУ». В исключительных случаях (форс-мажор и т.п.) при реализации образовательной деятельности с применением ЭО, ДОТ могут применять компоненты, не входящие в перечень электронной информационно-образовательной среды.

Дополнительные сведения находятся в приложениях:

1. Учебное пособие "Нейронные сети"
2. Методические указания к СРС



10. СПЕЦИАЛЬНЫЕ УСЛОВИЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ ОБУЧАЮЩИМИСЯ С ИНВАЛИДНОСТЬЮ И ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

Освоение дисциплины инвалидами и лицами с ограниченными возможностями здоровья осуществляется с использованием специальных технических средств и информационных технологий, предоставляемых Ресурсным учебно-методическим центром по обучению инвалидов и лиц с ограниченными возможностями здоровья ЧелГУ по запросу обучающегося (мобильные специальные технические средства для лиц с нарушениями зрения и с нарушением слуха, ассистивные информационные технологии).

При необходимости для обучающихся с нарушениями зрения на рабочих местах для проведения практических или лабораторных занятий устанавливается специальное программное обеспечение (программа речевой навигации, речевые синтезаторы, экранные лупы).

В учебные аудитории обеспечивается беспрепятственный доступ для обучающихся с инвалидностью и с ограниченными возможностями здоровья. В каждой аудитории, где обучаются инвалиды и лица с ограниченными возможностями здоровья, предусматривается соответствующее количество мест для обучающихся с учетом нарушений их здоровья.

Для освоения дисциплины инвалидам и лицам с ограниченными возможностями здоровья предоставляется доступ к печатным источникам, имеющимся в научной библиотеке ЧелГУ, с помощью специальных технических средств; доступ с помощью специальных технических и программных средств к электронным источникам, представленным в форме электронного документа в фонде научной библиотеки ЧелГУ или электронно-библиотечных системах.

Учебно-методические материалы для обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья предоставляются в формах, адаптированных к ограничениям их здоровья и особенностям восприятия информации.

Для инвалидов и лиц с ограниченными возможностями здоровья освоение дисциплины может быть частично или полностью осуществлено с использованием дистанционных образовательных технологий.

При проведении промежуточной аттестации по дисциплине обучающимся с инвалидностью и с ограниченными возможностями здоровья обеспечивается по их заявлению предоставление в доступной форме в зависимости от их индивидуальных особенностей инструкции о порядке проведения промежуточной аттестации, оценочных средств и возможности ответов на задания (письменно на бумаге, набор ответов на компьютере, письменно шрифтом Брайля, с использованием услуг ассистента, устно).

При проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование предоставленных ЧелГУ или собственных технических средств, необходимых им в связи с их индивидуальными особенностями. При необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на задания, процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Контрольное мероприятие 1

Лабораторная работа 1

Визуализация данных

1. Откройте дата сет об Ирисах. Рассмотрите основные признаки, представленный в наборе. Загрузите набор данных с использованием `pandas` и выведите признаки набора данных.
2. Для детального изучения признаков воспользуйтесь методом `info()` класса `DataFrame`
3. Для представления распределения простого количественного признака постройте обычная гистограмма, содержащаяся во всех библиотеках.
4. Визуализируйте соотношения количественных признаков является диаграмма по нескольким признакам.
5. Определите корреляции признаков обучающей выборки. В пакете `Pandas` имеется встроенный инструмент для этого – метод `corr()` класса `DataFrame`.

Контрольное мероприятие 2

Лабораторная работа 2

Применение алгоритма APRIORI

С помощью алгоритма Apriori, можно определить товары, купленные вместе – то есть установить ассоциативные правила. Это необходимо для повышения покупательской активности.

1. Откройте файл с набором данных о покупательской активности. Рассмотрите основные признаки, представленный в наборе. Загрузите набор данных с использованием pandas и выведите признаки набора данных.
2. Реализуйте алгоритм APRIORI на одном из языков программирования (ПИФАГОР, Python и др.)
3. Примените реализованный алгоритм для выявления товаров, которые покупают вместе чаще всего (комплиментарных товаров).

Контрольное мероприятие 3

Лабораторная работа 3

Применение алгоритма линейной регрессии

1. Откройте набор данных о стоимости автомобилей. Рассмотрите основные признаки, представленный в наборе. Загрузите набор данных с использованием `pandas` и выведите признаки набора данных.
2. Реализуйте алгоритм линейной регрессии на одном из языков программирования (ПИФАГОР, Python и др.)
3. Примените реализованный алгоритм для прогнозирования цены на автомобили.

Контрольное мероприятие 4

Лабораторная работа 4

Ансамблевые методы машинного обучения

1. Откройте набор данных о дефектах текстиля. Рассмотрите основные признаки, представленный в наборе. Загрузите набор данных с использованием `pandas` и выведите признаки набора данных.
2. При изготовлении текстильных изделий на фабриках могут возникать редкие аномалии, что приводит к ухудшению качества тканей. Чтобы избежать этого в некоторых случаях, крайне важно обнаружить дефект. Приведенный набор данных содержит:
 - Размер изображения: 32x32 или 64x64
 - Классы: ['good', 'color', 'cut', 'hole', 'thread', 'metal_contamination']
 - Ракурсы: 8 различных вращений в [0, 20, 40, 60, 80, 100, 120, 140]
 - При заданном размере изображения доступны обучающий и тестовый наборы данных со случайно сгенерированными исправлениями. Исходные изображения из набора и теста не перекрываются.
3. Выполните задачу классификации ракурсов с использованием только "хороших" изображений и протестируйте другие классы.

Контрольное мероприятие 5

Лабораторная работа 5

Обнаружение мошеннических операций с кредитными картами. Работа с несбалансированными классами

1. Откройте набор данных об операциях с кредитными картами. Рассмотрите основные признаки, представленный в наборе. Загрузите набор данных с использованием `pandas` и выведите признаки набора данных.
2. Важно, чтобы компании, выпускающие кредитные карты, могли распознавать мошеннические транзакции по кредитным картам, чтобы с клиентов не взималась плата за товары, которые они не покупали. Приведенный набор данных содержит:
 - Набор данных содержит транзакции, совершенные европейскими держателями карт с помощью кредитных карт в сентябре 2013 года.
 - В этом наборе данных представлены транзакции, которые произошли за два дня, где мы имеем 492 мошеннических действия из 284 807 транзакций. Набор данных сильно несбалансирован, на положительный класс (мошенничество) приходится 0,172% всех транзакций.
 - Он содержит только числовые входные переменные, которые являются результатом преобразования PCA. К сожалению, из-за проблем с конфиденциальностью мы не можем предоставить исходные функции и дополнительную справочную информацию о данных. Характеристики V_1, V_2, \dots, V_{28} являются основными компонентами, полученными с помощью PCA, единственными характеристиками, которые не были преобразованы с помощью PCA, являются "Время" и "Количество". Функция "Время" содержит секунды, прошедшие между каждой транзакцией и первой транзакцией в наборе данных. Функция "Сумма" – это сумма транзакции, эту функцию можно использовать, например, для обучения, зависящего от затрат. Функция 'Class' является переменной ответа, и она принимает значение 1 в случае мошенничества и 0 в противном случае.
 - Учитывая коэффициент дисбаланса в классе, рекомендуется измерять точность, используя площадь под кривой возврата точности (AUPRC). Точность матрицы путаницы не имеет значения для несбалансированной классификации.
3. Попытайтесь избавиться от несбалансированности классов.
4. Обучите модель определять мошеннические транзакции.
5. Оцените точность модели.

Контрольное мероприятие 6

Лабораторная работа 6

Применение анализа главных компонент (РСА)

1. Откройте набор данных. Рассмотрите основные признаки, представленный в наборе. Загрузите набор данных с использованием pandas и выведите признаки набора данных.
2. После завершения текущих программ финансирования Международная гуманитарная неправительственная организация собрала около 10 миллионов долларов. Теперь генеральному директору НПО необходимо решить, как использовать эти деньги стратегически и эффективно. Важные вопросы, возникающие при принятии этого решения, в основном связаны со странами, остро нуждающимися в помощи. Наша работа как аналитика данных заключается в классификации стран с использованием социально-экономических факторов и факторов здравоохранения, которые определяют общее развитие наций. После этого анализа нам нужно предложить страны, на которых генеральному директору необходимо сосредоточиться и которым он должен уделить наивысший приоритет.
3. Проведите стандартизацию функций.
4. Проверьте взаимосвязь между функциями без РСА.
5. Примените анализ основных компонент.
6. Проверьте взаимосвязи между функциями после РСА.

Контрольное мероприятие 7

Лабораторная работа 7

Командный проект

1. Разделитесь на команды по 2–4 человека.
2. С помощью сети Интернет найдите задачу, которая решается с помощью машинного обучения. Это может быть задача, которая решается любым из изученных алгоритмов.
3. С помощью практик MLOps разделите проект на небольшие задачи, проведите «совещание», используйте продукты для внедрения MLOps.
4. Подготовьте мини-презентацию своего готового решения.

Контрольное мероприятие

Теоретическая работа

- 1) Какой код лучше использовать при загрузке данных из csv-файла?
 - а) `dataset = read_csv("data.csv")`
 - б) `dataset = import("data.csv")`
 - в) `dataset = read.csv("data.csv")`
 - г) `dataset = import.csv("data.csv")`
 - д) `dataset = read_xls("data.csv")`
- 2) Опишите принципы работы алгоритма APRIORI
- 3) Перечислите и опишите виды регрессионного анализа.

Контрольное мероприятие 9

Промежуточная аттестация

Примерные вопросы на экзамене.

1. Этапы решения задач машинного обучения
2. Основные определения в машинном обучении: объект, целевая функция, признак, модель, обучающая выборка, функционал качества, обучение, переобучение.
3. Задачи машинного обучения - обучение с учителем, без учителя.
4. Задачи регрессии и классификации. Задачи снижения размерности и кластеризации.
5. Типы признаков в машинном обучении. Приведите примеры различных признаков.
6. Определение ROC-кривой.
7. Метод k ближайших соседей в задаче классификации.
8. Методы отбора признаков. Жадный метод.
9. Определение отступа в метрических алгоритмах классификации. Алгоритм Condensed Nearest Neighbor.
10. Метод k ближайших соседей в задаче регрессии.
11. Обобщение метода k ближайших соседей через взвешенный учет объектов. Ядерная оценка плотности.
12. Проклятие размерности. Зависимость метода ближайших соседей от масштабирования признаков. Способы стандартизации признаков.
13. Вероятностная постановка задачи классификации. Функция правдоподобия и априорная вероятность.
14. Функционал среднего риска. Общая формула байесовского классификатора.
15. Наивный байесовский классификатор.
16. Восстановление плотности распределения по выборке.
17. Аддитивное сглаживание для байесова классификатора.
18. Модель МакКаллока-Питтса
19. Обобщённая модель линейного классификатора. Определение отступа. Минимизация эмпирического риска.
20. Метод градиентного спуска. Выбор величины шага.
21. Метод стохастического градиента. Недостатки метода SG и как с ними бороться.
22. Использование сингулярного разложения для решения задачи наименьших квадратов.
23. Нелинейная регрессия. Метод Ньютона-Гаусса
24. Задача уменьшения размерности. Метод главных компонент.

