

Документ подписан простой электронной подписью Информация о владельце: ФИО: Гаскаев Сергей Валерьевич Должность: Ректор Дата подписания: 18.03.2025 14:53:16 Уникальный программный ключ: 04c19ed88fb98f3b6cb77a486b9a8788b8722727	МИНОБРАЗОВАНИЯ И НАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)	Рабочая программа дисциплины "Алгоритмы 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)*

Алгоритмы 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 2024

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

Челябинск 2024 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.

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

Программирование для анализа данных

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

Анализ требований и проектирование систем искусственного интеллекта

Высокопроизводительные параллельные вычисления на кластерных системах

Современные архитектуры глубоких искусственных нейронных сетей

Выполнение и защита выпускной квалификационной работы

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

ПК-5: Способен разрабатывать и применять методы машинного обучения для решения задач

Знать:

[ПК-4.1. 3-1.] принципы и методы машинного обучения, типы и классы задач машинного обучения, методологию ML Ops;
[ПК-4.1. 3-2.] статистические методы анализа данных;
[ПК-4.3. 3-1.] классические методы и алгоритмы машинного обучения: предиктивные - обучение с учителем, дескриптивные - обучение без учителя

Уметь:

[ПК-4.1. У-1.] сопоставить задачам предметной области классы задач машинного обучения;
[ПК-4.1. У-2.] использовать статистические методы анализа данных при решении задач машинного обучения;
[ПК-4.3. У1.] проводить сравнительный анализ и осуществлять выбор, настройку при необходимости, разработку методов и алгоритмов для решения задач машинного обучения

Владеть:

Имеет практический опыт: участия в разработке алгоритмов для решения задач машинного обучения

ПК-8: Способен использовать инструментальные средства для решения задач машинного обучения

Уметь:

[ПК-5.1. У-1.] проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения

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 : 64 самостоятельная работа : 35,5 : контактная работа: 72,5 ИКР: 8,5	Виды контроля в семестрах: экзамены 5

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/	5	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/	5	2	Л1.1 Л1.2Л2.1
1.3	Построение модели и сведение обучения к задаче оптимизации. Оценка качества работы алгоритма машинного обучения /Лек/	5	2	Л1.1 Л1.2Л2.1
1.4	Лабораторная работа 1. Визуализация данных /Лаб/	5	4	Л1.1 Л1.2Л2.1
1.5	Подготовка к лабораторной работе /Ср/	5	5	Л1.1 Л1.2Л2.1
	Раздел 2. Классические методы и алгоритмы машинного обучения. Деревья принятия решений. Регрессия. Кластеризация			
2.1	Ассоциативные правила. Алгоритм APRIORI. Алгоритм FP- Growth /Лек/	5	2	Л1.1 Л1.2Л2.1
2.2	Кластеризация. Метод К-средних. DBSCAN. Нечеткая кластеризация. /Лек/	5	2	Л1.1 Л1.2Л2.1
2.3	Линейные методы. Линейная регрессия. Линейная классификация. /Лек/	5	2	Л1.1 Л1.2Л2.1
2.4	Метод Парзенковского окна. Деревья принятия решений /Лек/	5	2	Л1.1 Л1.2Л2.1
2.5	Вероятностные модели. Байесовский классификатор. Наивный байесовский классификатор. /Лек/	5	2	Л1.1 Л1.2Л2.1
2.6	Ансамблевые методы. Бэггинг. Бустинг. Стэкинг /Лек/	5	2	Л1.1 Л1.2Л2.1
2.7	Лабораторная работа 2. Применение алгоритма APRIORI /Лаб/	5	4	Л1.1 Л1.2Л2.1



2.8	Лабораторная работа 3. Применение алгоритма линейной регрессии. Теоретический срез. /Лаб/	5	4	Л1.1 Л1.2Л2.1
2.9	Лабораторная работа 4. Ансамблевые методы машинного обучения /Лаб/	5	4	Л1.1 Л1.2Л2.1
2.10	Подготовка к лабораторным работам /Ср/	5	7	Л1.1 Л1.2Л2.1
Раздел 3. Обучение моделей машинного обучения				
3.1	Методы понижения размерности данных - метод главных компонент (PCA) и метод t-SNE. /Лек/	5	2	Л1.1 Л1.2Л2.1
3.2	Сжатие и визуализация данных. /Лек/	5	2	Л1.1 Л1.2Л2.1
3.3	Метрики в задачах машинного обучения. /Лек/	5	2	Л1.1 Л1.2Л2.1
3.4	Работа с несбалансированными классами. Обзор методов отбора признаков. /Лек/	5	2	Л1.1 Л1.2Л2.1
3.5	Применение машинного обучения к прикладным задачам /Лек/	5	2	Л1.1 Л1.2Л2.1
3.6	Лабораторная работа 5. Обнаружение мошеннических операций с кредитными картами. /Лаб/	5	6	Л1.1 Л1.2Л2.1
3.7	Лабораторная работа 6. Применение анализа главных компонент (PCA) /Лаб/	5	6	Л1.1 Л1.2Л2.1
3.8	Подготовка к лабораторным работам /Ср/	5	6	Л1.1 Л1.2Л2.1
Раздел 4. MLOps				
4.1	MLOps: самые известные модели /Лек/	5	4	Л1.1 Л1.2Л2.1
4.2	Лабораторная работа 7. Применение MLOps для командной разработки программного решения /Лаб/	5	4	Л1.1 Л1.2Л2.1
4.3	Подготовка к лабораторной работе /Ср/	5	5	Л1.1 Л1.2Л2.1
4.4	Подготовка к экзамену /Ср/	5	12,5	Л1.1 Л1.2Л2.1
Раздел 5. Иная контактная работа				
5.1	Индивидуальные консультации, текущий контроль /ИКР /ИКР/	5	8,5	

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

6.1. Перечень видов оценочных средств

Лабораторная работа №1
Лабораторная работа №2
Лабораторная работа №3
Лабораторная работа №4
Лабораторная работа №5
Лабораторная работа №6
Лабораторная работа №7
Теоретический срез

6.2. Типовые контрольные задания и иные материалы для текущей аттестации

Теоретический срез

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

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

Рейтинг обучающегося по дисциплине формируется по результатам текущего контроля. Контрольное мероприятие экзамена проводится в очной форме и не является обязательным, однако студент может прийти на экзамен и повысить свой рейтинг. Студенту на экзамене выдаётся билет. Дается 90 минут для подготовки к ответу. Проводится собеседование по выданным вопросам.

Лабораторная работа №1 - №7: 3 балла: Студент отвечает на все теоретические вопросы по теме лабораторной работы. Программа работает правильно и корректно.
2 балла: Студент отвечает с затруднениями на теоретические вопросы по теме лабораторной работы. Программа работает правильно и корректно.
1 балл: Алгоритм составлен верно, но программа не работает.
0 баллов: Алгоритм составлен неверно, программа не работает.

Теоретический срез: 3 балла: верно решены все три задания теоретического среза.
2 балла: верно решены два задания теоретического среза.
1 балл: верно решено одно задание теоретического среза.
0 баллов: неверно решены все три задания теоретического среза.

Экзамен: 4 балла получает студент, обнаруживший всестороннее, систематическое и глубокое знание учебно-программного материала, умение свободно выполнять задания, предусмотренные экзаменационным билетом и свободно отвечающий на дополнительные вопросы
3 балла заслуживает студент, обнаруживший полное знание учебно-программного материала, успешно выполняющий предусмотренные в билете задания, но отвечающий на дополнительные вопросы с затруднениями
2 балла получает студент, допустивший погрешности в ответе на экзамене и при выполнении заданий, но обладающим необходимыми знаниями для их устранения под руководством преподавателя;
1 балл ставится студенту, обнаружившему пробелы в знаниях основного учебно-программного материала, допустившему принципиальные ошибки в выполнении предусмотренных билетом заданий.
0 баллов ставится студенту, который не смог выполнить ни одно задание в билете.

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

7.1. Рекомендуемая литература

7.1.1. Основная литература

Авторы, составители	Заглавие	Издательство, год	Ресурс
---------------------	----------	-------------------	--------



	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 (http://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	Ростовцев В. С.	Искусственные нейронные сети: учебник для вузов (https://e.lanbook.com/book/364517)	Санкт- Петербург : Лань, 2024	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

7.3.2 Профессиональные базы данных и информационно-справочные системы

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)

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 голо информационных технологий, предоставляемых Ресурсным учебно-методическим центром по обучению инвалидов и лиц с ограниченными возможностями здоровья ЧелГУ по запросу обучающегося.



1. Мобильные специальные технические средства для лиц с нарушениями зрения: портативный компьютер с вводом/выводом шрифтом Брайля с синтезатором речи «EiBraille-W14J G2»; ноутбуки с программной экранного доступа NVDA; электронные увеличители для удаленного просмотра; видеоувеличители портативные; тифлоплеер; цифровые диктофоны.

2. Мобильные специальные технические средства для лиц с нарушениями слуха: система свободного звукового поля со встроенной совместимостью с FM-устройствами; радиоклассы «Сонет-PCM» с передатчиком, заушным индуктором и индукционной петлей; система информационная для слабослышащих переносная «Исток» А2 со встроенным плеером – звуковым информатором; документ-камера; программируемые слуховые аппараты индивидуального пользования.

3. Ассистивные информационные технологии: программное обеспечение экранного доступа с синтезом речи NVDA; программы экранного увеличения; программы речевого синтеза для компьютеров и ноутбуков; программы речевого синтеза для мобильных устройств; экранная клавиатура; экранная лупа.

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

В учебные аудитории обеспечивается беспрепятственный доступ для обучающихся инвалидов и обучающихся 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 к электронным источникам, представленным в форме электронного документа в фонде научной библиотеки ЧелГУ или электронно-библиотечных системах, с помощью специальных технических и программных средств (рабочее место для незрячего пользователя с программным обеспечением экранного доступа с синтезом речи NVDA, рабочее место с компьютерным роллером и клавиатурой Clevu с большими кнопками и с разделяющей клавиши накладкой).

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

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,
- в печатной форме шрифтом Брайля.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,
- в форме аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

Для инвалидов и лиц с ограниченными возможностями здоровья освоение дисциплины может быть частично или полностью осуществлено с использованием дистанционных образовательных технологий (Moodle, Adobe Connect Pro и пр.).

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья используется 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 направлены на индивидуализацию обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или обучающимся с ограниченными возможностями здоровья.

При проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине обеспечивается выполнение следующих дополнительных требований в зависимости от индивидуальных особенностей, обучающихся:

- а) инструкция по порядку проведения процедуры оценивания предоставляется в доступной форме (устно, в письменной форме, в письменной форме шрифтом Брайля, устно с использованием услуг сурдопереводчика);
- б) доступная форма предоставления заданий оценочных средств (в печатной форме, в печатной форме увеличенным шрифтом, в печатной форме шрифтом Брайля, в форме электронного документа, задания зачитываются ассистентом, задания предоставляются с использованием сурдоперевода);
- в) доступная форма предоставления 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 технических средств, необходимых им в связи с их индивидуальными особенностями. Эти средства могут быть предоставлены ЧелГУ или могут использоваться собственные технические средства. 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.

Проведение процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья допускается с использованием дистанционных образовательных технологий.

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. Попробуйте 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.

