

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Институт информационных технологий

Кафедра математики и информатики

УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ
«ВВЕДЕНИЕ В ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»

Направление подготовки (специальность):

01.03.02 Прикладная математика и информатика

Образовательная программа:

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Очная форма обучения

Составители:

Лягинова О.Ю., зав. кафедрой МиИ
канд.пед.наук, доцент

Лягинов Н.М., старший
преподаватель кафедры МиИ

Реутов Илья Сергеевич, старший
преподаватель кафедры математики
и информатики, ведущий инженер-
программист ООО «Малленом
Системс»

г. Череповец - 2022

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

Основная литература:

1. Сурова, Н. Ю. Искусственный интеллект / Н. Ю. Сурова, М. Е. Косов. – Москва : Юнити-Дана, 2021. – 408 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=690578> (дата обращения: 11.06.2022). – Библиогр.: с. 383-385. – ISBN 978-5-238-03513-0. – Текст : электронный.
2. Плас Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение. — (Серия «Бестселлеры O'Reilly»). - Санкт-Петербург : Питер, 2021. - 576 с. - ISBN 978-5-4461-0914-2. - URL: <https://ibooks.ru/bookshelf/376830/reading> (дата обращения: 11.06.2022). - Текст: электронный.

Дополнительная литература по дисциплине:

1. Залоило М. В. Искусственный интеллект в праве: научно-практическое пособие / М.В. Залоило. - Москва : Инфотропик Медиа, 2021. - 132 с. - ISBN 978-5-9998-0365-8. - URL: <https://ibooks.ru/bookshelf/382542/reading> (дата обращения: 11.06.2022). - Текст: электронный.
2. Бахтеев Д. В. Искусственный интеллект: этико-правовые основы : монография / Д.В. Бахтеев. - Москва : Проспект, 2021. - 176 с. - ISBN 978-5-392-33413-1. - URL: <https://ibooks.ru/bookshelf/380167/reading> (дата обращения: 11.06.2022). - Текст: электронный.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля), включая перечень информационных справочных систем (при необходимости)

1. Электронная библиотека «Университетская библиотека online». URL: <http://biblioclub.ru/>
2. Информационная система «Единое окно доступа к образовательным ресурсам». URL: <http://window.edu.ru/>
3. Образовательный портал Череповецкого государственного университета. URL: <https://edu.chsu.ru/>
4. Программирование на Python. URL: <https://stepik.org/course/67/syllabus> (03.08.2019)

Учебно-методические указания и рекомендации к изучению тем лекционных и практических занятий, самостоятельной работе студентов

Лекции

№ п/п	Тема лекции	Количество часов
1	Развитие искусственного интеллекта.	8
2	Сбор и подготовка данных для систем искусственного интеллекта.	8
3	Правовые и этические нормы в области искусственного интеллекта.	6
Итого		22

Лабораторные работы

№ п/п	Тема лекции	Количество часов
1	Развитие искусственного интеллекта.	8
2	Сбор и подготовка данных для систем искусственного интеллекта.	32
3	Правовые и этические нормы в области искусственного интеллекта.	8
Итого		48

Раздел 1. Развитие искусственного интеллекта

Содержание:

Информатизация общества, информационно-коммуникационные технологии в мире, перспективы их развития. Развитие искусственного интеллекта, эволюция и главные тренды систем искусственного интеллекта. Основные параметры идентификации задач искусственного интеллекта: назначение, сфера применения, виды используемых знаний, временные аспекты решения задач. Классификация систем искусственного интеллекта, методы и инструментальные средства решения задач с использованием систем искусственного интеллекта в зависимости от особенностей проблемной области. Функциональность программного обеспечения, методы оценки экономической эффективности применяемого программного и аппаратного обеспечения.

Самостоятельная работа по разделу:

По итогам самостоятельной работы студент готовит отчет, включающий в себя ответы на вопросы и решение заданий, предполагавшихся к выполнению в ходе самостоятельной работы. Отчет сдается преподавателю в электронной форме.

1. Напишите эссе, раскрыв в нем вопрос влияния информатизации на изменение качества жизни в РФ.
2. Разработайте интеллект-карту, отражающую этапы развития искусственного интеллекта и главные тренды современных систем искусственного интеллекта.

3. Приведите примеры систем искусственного интеллекта, которые могут быть использованы в черной металлургии.
4. Приведите примеры систем искусственного интеллекта, которые могут быть в процессе производства минеральных удобрений.
5. Разработайте интеллект-карту, отражающую методы оценки экономической эффективности применяемого аппаратного обеспечения (в контексте использования систем искусственного интеллекта).
6. Разработайте интеллект-карту, отражающую методы оценки экономической эффективности применяемого программного обеспечения (в контексте использования систем искусственного интеллекта).

Образцы заданий для лабораторных работ:

По итогам выполнения лабораторной работы студент демонстрирует результаты работы программы преподавателю, предварительно разработав тестовые случаи, а также сдает в электронном виде отчет, содержащий порядок выполнения работы.

Лабораторная работа «Классификация систем искусственного интеллекта»

Задание: изучите основания классификации систем искусственного интеллекта и представленные классы. Приведите примеры реальных систем, относящихся к выделенным классам. Дайте заключение о возможности изменения классификации, обоснуйте заключение.

Искусственный интеллект как область знаний охватывает все области человеческой деятельности, включая: информатику, математику, философию, психологию, термодинамику, лингвистику, здравоохранение, инженерию, экономику, когнитивные науки и др.

Эти знания используются в таких приложениях, как: системы управления, системы принятия решений, многоагентные системы, системы обработки естественного языка, распознавание образов, распознавание речи, обработка знаний, интеллектуальный анализ данных, логистика и другие приложения.

Классификация должна отражать существенные (значимые) характеристики системы искусственного интеллекта (СИИ), включая особенности контура управления, в рамках которого используется СИИ, и технологии построения и использования знаний. В ГОСТ Р 59277-2020 «Системы искусственного интеллекта. КЛАССИФИКАЦИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА» рассмотрены следующие основания для классификации: 1) по степени автономности; 2) по степени автоматизации; 3) по архитектурному принципу; 4) по структуре и процессам обработки знаний: а) по модели знаний; б) по управлению знаниями; в) по методу обучения; 5) по специализации систем ИИ: а) специализированные (используют единый домен знаний); б) комплексные (используют множество доменов знаний); 6) по методам обработки информации; 7) по функциям в контуре управления; 8) по методам достижения интеграции и интероперабельности СИИ; 9) по опасности последствий; 10) по конфиденциальности; 11) по видам деятельности; 12) по взаимодействию с человеком-оператором. Возможно расширение видов классификации систем ИИ.

Таблица 1 - Схема классификации систем искусственного интеллекта

Основания для классификации	Классы
1 По степени автономности	1.1 Автономные системы 1.2 Встроенные системы 1.3 Гибридные системы
2 По степени автоматизации	2.1 Автоматизированные системы 2.2 Автоматические системы
3 По архитектурному принципу	3.1 Централизованные системы 3.2 Распределенные системы
4 По видам деятельности	4.1 Государственное управление 4.2 Безопасность 4.3 Общеотраслевое регулирование 4.4 Промышленность 4.5 Здравоохранение 4.6 Торговля 4.7 Финансы и банки 4.8 Транспорт и логистика 4.9 Сельское хозяйство 4.10 "Умный город" 4.11 Экология 4.12 Образование и наука 4.13 Нефть и газ 4.14 Прочее
5 По функциям контура управления	5.1 Системы с обратной связью 5.2 Системы реального времени 5.3 Адаптивные системы 5.4 Системы формирования цели (Системы целеполагания) 5.5 Системы формирования контура управления и обучения 5.6 Системы обработки измерений 5.7 Системы идентификации и диагностики 5.8 Системы когнитивного моделирования 5.9 Системы логического вывода 5.10 Системы принятия (поддержки) решений 5.11 Экспертно-аналитические системы 5.12 Системы оценки достижения цели 5.13 Ситуационные центры 5.14 Системы прогнозирования 5.15 Прочее
6 По специализации систем	6.1 Экспертные системы (управление знаниями) 6.2 Игровые системы 6.3 Системы естественного языка 6.4 Системы компьютерного зрения 6.5 Промышленные роботы 6.6 Беспилотные аппараты 6.7 Прочее
7 По комплексности и сложности систем	7.1 Многоагентные системы 7.2 Системы "Большие данные" 7.3 Промышленный интернет вещей 7.4 Киберфизические системы

	7.5 Системы жизненного цикла 7.6 Системы сетевой экспертизы 7.7 Распределенные системы управления 7.8 Система распределенных ситуационных центров 7.9 Прочее
8 По методам обработки информации	8.1 Нейросети 8.2 Обучение на примере 8.3 Эволюционные и генетические алгоритмы 8.4 Муравьиные алгоритмы 8.5 Иммунные вычисления 8.6 Глубокое обучение 8.7 Роевые вычисления 8.8 Метод Байеса 8.9 Уменьшение размерности 8.10 Природные вычисления 8.11 Мягкие вычисления 8.12 Кластеризация 8.13 Дерево решений 8.14 Регуляризация 8.15 Аналоговая обработка данных 8.16 Обработка фурье-образов 8.17 Регрессия 8.18 Решение обратных задач 8.19 Система правил 8.20 Прочее
9 По управлению знаниями, моделям и методам обучения	9.1 Процедурные 9.2 Декларативные 9.3 Онтологические 9.4 Семантические 9.5 Продукционные 9.6 Фреймовые 9.7 Нейросетевая 9.8 Генетическая 9.9 Логическая 9.10 Статистическая 9.11 Нечеткие знания 9.12 Классификации 9.13 Многомерное представление (3Д, 4Д) 9.14 Функциональные 9.15 Технологические 9.16 Методологические 9.17 Комбинированное обучение 9.18 Непрерывное обучение 9.19 Единовременное обучение 9.20 Прочее
10 По методам достижения интеграции и интероперабельности	10.1 Системы с интеграцией на базе онтологий 10.2 Системы на базе профилирования 10.3 Системы, использующие классификаторы 10.4 Прочее
11 По опасности последствий*	11.1 Социальная 11.2 Политическая

	11.3 Экономическая 11.4 Технологическая 11.5 Техногенная 11.6 Экологическая 11.7 Безопасность государства
12 По конфиденциальности**	12.1 Уровень конфиденциальности (0-3)
<p>* Классификация в соответствии с категорированием объектов критической информационной инфраструктуры: (1) Социальной значимости (здоровье и жизнь людей); Политической значимости (причинение ущерба государству); Экономической значимости (ущерб субъектам и/или бюджетам); Экологической значимости (воздействие на окружающую среду); Значимость для обороны/безопасности, Правопорядка.</p> <p>** Классификация соответствует следующим классам конфиденциальности: (0) Открытая информация; (1) Внутренняя информация; (2) Конфиденциальная информация; (3) Секретная информация.</p>	

Возможно дополнение классификации СИИ как по новым основаниям, так и путем детализации классов по специализированным классификациям.

Лабораторная работа «Методы и инструментальные средства решения задач искусственного интеллекта»

Задание: Подготовьте интеллект-карту, включающую в себя методы искусственного интеллекта, языки программирования интеллектуальных систем, фреймворки машинного обучения, облачные сервисы, используемые для решения задач искусственного интеллекта, а также аппаратные решения, используемые в системах искусственного интеллекта. Для каждого объекта, включенного в интеллект-карту, приведите краткую характеристику.

Литература:

1. Сурова, Н. Ю. Искусственный интеллект / Н. Ю. Сурова, М. Е. Косов. – Москва : Юнити-Дана, 2021. – 408 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=690578> (дата обращения: 11.06.2022). – Библиогр.: с. 383-385. – ISBN 978-5-238-03513-0. – Текст : электронный..

Раздел 2. Сбор и подготовка данных для систем искусственного интеллекта

Содержание:

Основные инструменты, библиотеки и технологии Data Science. Формирование требований к системе искусственного интеллекта. Виды представления данных. Методы поиска, сбора и обобщения информации о проблемной области путем опроса экспертов, исходных данных о функционировании проблемной и предметной областей, документированных источников знаний. Достоверные источники данных. Открытые источники данных. Целостность и непротиворечивость данных. Уровни представления данных (ODS, DDL, семантический слой, модель данных). Предварительная обработка и разметка данных.

Самостоятельная работа по разделу:

По итогам самостоятельной работы студент готовит отчет, включающий в себя ответы на вопросы и решение заданий, предполагавшихся к выполнению в ходе самостоятельной работы. Отчет сдается преподавателю в электронной форме.

Задания для самостоятельной работы:

1. Перечислите основные источники открытых данных.
2. Назовите основные этапы выполнения проекта в области искусственного интеллекта.
3. Перечислите программные платформы, используемые для реализации проектов в области искусственного интеллекта.
4. Перечислите основные библиотеки Python, используемые для работы с данными.
5. Как в Python организована работа со списками? Как распечатать элемент списка, как изменить элемент списка, как просуммировать элементы списка, как вычислить длину списка? Приведите пример.
6. Как из списка получить диапазон? Как добавить элемент в конец списка? Как отсортировать список по столбцу? Приведите пример.
7. Как в Python организовать циклическую обработку данных? Приведите пример.
8. Как в Python создаются функции пользователя? Приведите пример.
9. Изучите документацию разработчика по основным возможностям библиотеки Pandas.
10. Как вызвать библиотеку Pandas из программы на языке Python? Приведите пример.
11. Как прочитать файл формата CSV? Приведите пример.
12. Как создать таблицу, используя конструктор DataFrame()? Приведите пример.
13. Как просмотреть таблицу? Приведите пример.
14. Как вызвать n первых строк таблицы? Как вызвать n последних строк таблицы? Приведите пример.
15. Как реализуется логическая индексация для получения элементов по определенному условию? Приведите пример.
16. Как реализуется индексация в Series? Приведите пример.
17. Как изменить название столбца таблицы? Приведите пример.
18. Как найти пропущенные значения в таблице? Как удалить пропущенные значения в таблице? Приведите пример.
19. Как заполнить пропущенные значения? Чем обычно заполняются пропущенные значения? Приведите пример.
20. Как найти дубликаты в таблице? Как удалить дубликаты? Приведите пример.
21. Как просмотреть уникальные значения в столбце? Приведите пример.
22. Как заменить значения в таблице или столбце? Приведите пример.
23. Как сгруппировать данные в столбце? Приведите пример.
24. Как определить максимальное значение в столбце? Как определить минимальное значение в столбце? Приведите пример.
25. Как рассчитать среднее арифметическое значение в столбце? Как рассчитать медиану? Приведите пример.
26. Как отсортировать значения в столбце по возрастанию/убыванию? Приведите пример.

Образцы заданий для лабораторных работ:

По итогам выполнения лабораторной работы студент демонстрирует результаты работы программы преподавателю, предварительно разработав тестовые случаи, а также сдает в электронном виде отчет, содержащий порядок выполнения работы.

Лабораторная работа «Сбор данных, отображение данных на экране, простейшие расчеты»

Задание: Проведите исследование частоты использования эмодзи на разных российских платформах.

1. Соберите данные о Топ-10 эмодзи на каждой из платформ (название, рус.; название, англ.; значок; название сети 1, млн.; название сети 2, млн.).
2. Организуйте хранение данных средствами Python.
3. Выведите на экран таблицу с данными.
4. Посчитайте сколько всего топовых эмодзи отправляется в социальной сети. Для всех топовых эмодзи посчитайте, сколько в среднем (в миллионах) сообщений с каждым из них отправляется в сети.
5. Для каждого эмодзи посчитайте соотношение его количества в каждой из социальных сетей.
6. Выведите на экран отсортированные результаты по использованию эмодзи в каждой из социальных сетей по убыванию.
7. Создайте срезы так, чтобы в каждом случае выводилась не вся таблица, а только первые пять строк таблицы.
8. Оцените, какую долю среди всех эмодзи, используемых на платформе, составляют выбранные вами.
9. Измените исходную таблицу, добавив в нее столбцы с долями использования по каждому из эмодзи.
10. Рассчитайте индекс использования эмодзи для каждой из социальных сетей. Для этого:
 - для каждого столбца оцените его масштаб (как среднее значение);
 - для каждого значения в столбце выполните нормирование (разделите на масштаб);
 - нормированные значения сложите.
11. Сделайте выводы по данным.

Лабораторная работа «Использование функций для работы с данными»

Задание: Проведите исследование изменения предпочтений российской академии кинематографических искусств, начиная с 1987 года по настоящее время.

1. Соберите данные о фильмах, получивших премию «Ника» с 1987 года (название фильма, год вручения премии, рейтинг, длина (в мин.), жанр, бюджет (в руб.), сборы (в руб.)).
2. Организуйте хранение данных в Python.
3. Определите лидеров по разным параметрам: например, по рейтингу и бюджету, используя сортировку.

4. Напишите функцию перевода продолжительности фильма из минут в часы. Измените данные, используя написанную функцию.
5. Напишите функцию, выводящую на экран первые пять строк данных. Данные в таблице предварительно должны сортироваться по столбцу, указанному пользователем, и в порядке по возрастанию/убыванию значений в указанном столбце.
6. Напишите функцию, которая вычисляет и выводит на экран прибыль, полученную от фильма. Прибыль вычисляется как разность сборов и бюджета.
7. Напишите функцию, которая считает и выводит на экран среднее значение по заданному столбцу.
8. Отфильтруйте таблицу данных, оставив в ней фильмы от 2010 года и новее. Затем отсортируйте её по сборам по убыванию и напечатайте на экране топ-5 результатов.
9. Отфильтруйте таблицу по жанру. Для этой цели напишите функцию, имеющую два аргумента: таблицу с данными и название жанра; и возвращающую новую таблицу с фильмами переданного жанра.
10. Отберите самые популярные жанры.
11. Выберите из таблицы данных фильмы ТОП-5 самых популярных жанров. Выведите следующую информацию по фильмам: название фильма; год вручения премии; прибыль, полученная от фильма; стоимость минуты фильма. Отсортируйте данные по годам вручения премии.
12. Сформулируйте выводы по проделанной работе.

Лабораторная работа «Получение и предобработка данных»

Задание: получите данные из файла формата .csv, выполните предобработку полученных данных.

1. Библиотека Pandas.
 - 1). Получите доступ к библиотеке Pandas, используйте имя переменной `pd`.
 - 2). Создайте список `music` с 5 парами «имя вашего любимого исполнителя - название его песни».
 - 3). Создайте список `entries` с названиями для двух столбцов — *artist* и *track*.
 - 4). Используя конструктор `DataFrame()`, создайте таблицу из списка ваших любимых исполнителей *music* и списка столбцов *entries*. Сохраните таблицу в переменной *playlist* и выведите эту сборную таблицу на экран.
2. Получение данных
 - 1). Прочитайте файл `music_log.csv` и сохраните его в переменной `df`. Сохраните первые 5 строк с данными из `music_log.csv` в переменной `music_head` и выведите значение переменной на экран.
 - 2). Прочитайте файл `music_log.csv` и сохраните его в переменной `df`. Сохраните последние 10 строк с данными из `music_log.csv` в переменной `music_tail` и выведите значение переменной на экран.
3. Объект `DataFrame`
 - 1). Прочитайте файл `music_log.csv` и сохраните его в переменной `df`. Создайте переменную `shape_table` и сохраните в ней размеры таблицы `music_log.csv`. Напечатайте на экране размер таблицы в таком виде: Размер таблицы: ...
 - 2). Сколько наблюдений в наборе данных? В переменной `shape_table` хранится кортеж. Его первый элемент — количество наблюдений, который надо сохранить

в переменной *observations_table* (не забывайте, что индексация элементов идёт с 0). Напечатайте на экране ответ в таком виде: Количество наблюдений: ...

- 3). Сравните полученные результаты в переменных *observations_info_table* и *observations_table*. Если значения переменных совпадают, то выведите количество наблюдений и сообщение: «Решение верно, количество наблюдений равно», *observations_table*. Если значения переменных не совпадают, то выведите сообщение: «Решение неверно, проверьте еще раз!»

4. Индексация в DataFrame

- 1). Получите таблицу, состоящую из столбцов *genre* и *Artist*. Сохраните её в переменной *genre_fight*. Посчитайте число прослушанных треков в жанре поп. Сохраните результат в переменной *genre_pop*. Напечатайте ответ на экране в виде: Число прослушанных треков в жанре поп равно ...
- 2). Посчитайте число прослушанных треков в жанре рок. Допишите в код подсчёт, похожий на предыдущий, только с логическим условием `df['genre'] == 'rock'`. Сохраните результат в переменной *genre_rock*. Напечатайте ответ на экране в виде: Число прослушанных треков в жанре поп равно ... Число прослушанных треков в жанре рок равно ...
- 3). Напишите условную конструкцию, которая сравнивает полученные значения и выводит информацию о победителе в этом бою! Если победил жанр рок, то выведите сообщение "Рок победил!", а если победил жанр поп - сообщение "Попса forever!"

5. Объект Series

- 1). Получите таблицу только с жанром *rock* и сохраните её в переменной *rock*.
- 2). Выделите время прослушивания роковых композиций в особую структуру данных. Сохраните столбец 'total play' таблицы *rock* в переменной *rock_time*.
- 3). Обратитесь к новой Series с именем *rock_time* и посчитайте количество треков жанра рок, пропущенных в течение 5 секунд. Логическим условием укажите `rock_time <= 5`. Результат сохраните в переменной *rock_haters* и напечатайте на экране с пояснением: Количество пропущенных треков жанра рок равно ...
- 4). Выберите из исходной таблицы только строки с жанром 'pop' и сохраните эту новую таблицу в переменной *pop*.
- 5). По аналогии с роком создайте Series, где хранятся только данные о времени воспроизведения композиций в жанре поп. Назовите его *pop_time* и сохраните в нём данные столбца 'total play' из таблицы *pop*.
- 6). По аналогии с роком обратитесь к Series *pop_time*, чтобы посчитать количество пропущенных в течение 5 секунд треков жанра поп. Используйте условие `pop_time <= 5`. Результат сохраните в переменной *pop_haters* и напечатайте на экране в таком виде: Количество пропущенных треков жанра поп равно ...
- 7). Для обоих жанров посчитайте долю быстро пропущенных пользователями композиций в процентах. Разделите количество треков, которые пользователи пропустили — соответственно *rock_haters* и *pop_haters* — на общее количество треков жанра рок и жанра поп.
Общее количество треков жанра равно количеству наблюдений в таблицах *rock* и *pop*, т.е. значению атрибута `shape[0]` этих таблиц.

Результаты сохраните в переменных `rock_skip` и `pop_skip`. Выведите значения новых переменных в процентах с точностью до одного знака после запятой в форме:

Доля пропущенных композиций жанра рок равна: ...

Доля пропущенных композиций жанра поп равна: ...

Лабораторная работа «Раздельный сбор информационного мусора»

Задание: получите данные из файла формата `.csv`, удалите информационный мусор.

1. Получение данных

- 1). Прочитайте файл `music_log.csv` и сохраните его в переменной `df`.
- 2). Просмотрите информацию о наборе данных, воспользовавшись методом `info()`.

2. Переименование столбцов

- 1). Выведите список столбцов.
- 2). Подготовьте список `new_names` с новыми именами для столбцов:
 - `user_id` → `user_id`
 - `total play` → `total_play_seconds`
 - `Artist` → `artist_name`
 - `genre` → `genre_name`
 - `track` → `track_name`
- 3). Переименуйте столбцы таблицы, которая хранится в переменной `df`.
- 4). Проверьте, что получилось, запросив для структуры данных `df` атрибут `columns`.

3. Обработка пропущенных значений

- 1). Посчитайте количество пропущенных значений в наборе данных и выведите его на экран.
- 2). Заполните отсутствующие значения столбца `'track_name'` строкой `'unknown'`.
- 3). Заполните отсутствующие значения столбца `'artist_name'` строкой `'unknown'`.
- 4). Удалите пропущенные значения из столбца `'genre_name'`.
- 5). Проверьте полученный результат. Просмотрите информацию о наборе данных: воспользуйтесь методом `info()`.

4. Обработка дубликатов

- 1). Сохраните текущий размер таблицы в переменной `shape_table`.
- 2). Посчитайте и выведите на экран суммарное количество дубликатов в таблице.
- 3). Удалите дубликаты. Используйте метод `reset_index()` для сохранения порядка индексов.
- 4). Сохраните в переменную `shape_table_update` размер таблицы после удаления дубликатов.
- 5). Сравните переменные `shape_table` и `shape_table_update`. Если они равны, выведите сообщение 'Размер таблицы не изменился, текущий размер: ' и значение переменной `shape_table_update`. В ином случае сообщение должно быть таким: 'Таблица уменьшилась, текущий размер: ' и значение переменной `shape_table_update`.
- 6). Получите уникальные значения столбца `'genre_name'`, используйте метод `unique()`. Просмотрите результат и найдите название жанра, которое выпадает из общего ряда.
- 7). Оцените изменения: пересчитайте количество значений 'электроника' в столбце `'genre_name'`. Если удалось всё заменить, результат должен быть равен 0.

Сохраните этот результат в переменной `genre_final_count`, выведите на экран. Примените к отобранным по логическому условию `df['genre_name'] == 'электроника'` значениям столбца `'genre_name'` метод `count()` для подсчёта. Результат сохраните в переменной `genre_final_count`, значение которой напечатайте на экране.

5. Сделайте выводы по проделанной работе.

Лабораторная работа «Анализ данных и оформление результатов»

Задание: получите данные из файла формата `.csv`, выполните расчеты, подготовьте презентацию по результатам работы.

1. Знакомство с набором данных

- 1). Прочитайте данные из файла `music_log_2.csv` и выведите первые 10 строк (`music_log_2.csv` — обновлённый файл с данными, которые прошли предобработку).
- 2). Получите список названий столбцов, запросив атрибут `columns`. Результат выведите на экран.
- 3). Посчитайте количество пустых значений в наборе данных, сохраните результат в переменной `na_number`. Выведите её значение на экран.
- 4). Посчитайте количество дубликатов в наборе данных, сохраните результат в переменной `duplicated_number`. Выведите её значение на экран.

2. Группировка данных

- 1). Узнайте `user_id` меломанов. Для этого сгруппируйте данные по каждому пользователю, чтобы собрать жанры прослушанных им композиций. Сгруппируйте `DataFrame` по столбцу `user_id`, сохраните полученный результат в переменной `genre_grouping`. Посчитайте количество жанров, которые выбрали пользователи, методом `count()`, указав, что выбираем один столбец `genre_name`. Сохраните результат в переменной `genre_counting` и выведите первые 30 строк этой таблицы.
- 2). Быть может, те, кто за день слушает больше 50 песен, имеют более широкие предпочтения. Чтобы найти такого, напишите функцию `user_genres`, которая принимает некую группировку как свой аргумент `group`. Функция должна перебирать группы, входящие в эту группировку. В каждой группе два элемента — имя группы с индексом 0 и список значений с индексом 1. Обнаружив такую группу, в которой список (элемент с индексом 1) содержит более 50 значений, функция возвращает имя группы (значение элемента с индексом 0).
- 3). Вызовите функцию `user_genres`, как аргумент передайте ей `genre_grouping`. Результат — `user_id` неведомого нам любителя музыки — сохраните в переменной `search_id` и выведите значение на экран.

3. Сортировка данных

- 1). Выполняя предыдущее задание, вы обнаружили меломана с уникальными данными. Он за день послушал больше 50 композиций. Получите таблицу с прослушанными им треками. Для этого запросите из структуры данных `df` строки, отвечающие сразу двум условиям: 1) значение в столбце `'user_id'` должно быть равно значению переменной `search_id`; 2) время прослушивания, т.е.

- значение в столбце *'total_play_seconds'*, не должно равняться 0. Сохраните результат в переменной *music_user*.
- 2). Узнайте, сколько времени он слушал музыку каждого жанра. Сгруппируйте данные таблицы *music_user* по столбцу *'genre_name'* и получите сумму значений столбца *'total_play_seconds'*. Сохраните результат в переменной *sum_music_user* и выведите её значение на экран.
 - 3). Важно знать, сколько треков каждого жанра он включил. Сгруппируйте данные по столбцу *genre_name* и посчитайте, сколько значений в столбце *genre_name*. Сохраните результат в переменной *count_music_user* и выведите её значение на экран.
 - 4). Чтобы предпочтения были видны сразу, нужно крупнейшие значения расположить наверху. Отсортируйте данные в группировке *sum_music_user* по убыванию. Внимание: когда применяете метод *sort_values()* к *Series* с единственным столбцом, аргумент *by* указывать не нужно, только порядок сортировки. Сохраните результат в переменной *final_sum* и выведите её значение на экран.
 - 5). Теперь то же самое надо сделать с числом прослушанных меломаном композиций. Отсортируйте данные группировки *count_music_user* по убыванию. Сохраните результат в переменной *final_count*, значение которой выведите на экран.
4. Описательная статистика
- 1). Получите таблицу с композициями самого популярного жанра — поп, исключив пропущенные треки. Сохраните результат в переменной *pop_music*.
 - 2). Найдите максимальное время прослушивания песни в жанре поп. Сохраните результат в переменной *pop_music_max_total_play* и выведите её значение на экран.
 - 3). Получите строку таблицы *pop_music* с информацией о самой длинной по времени прослушивания песне жанра 'поп' и сохраните её в переменной *pop_music_max_info*. Выведите эту строку на экран.
 - 4). Найдите минимальное ненулевое время прослушивания композиции в жанре поп. Сохраните его в переменной *pop_music_min_total_play*, значение выведите на экран.
 - 5). Выведите на экран информацию о композиции жанра поп, которую запустили, но быстрее всех остальных выключили. Результат сохраните в переменную *pop_music_min_info* и выведите на экран.
 - 6). Рассчитайте медиану времени прослушивания произведений жанра поп. Сохраните результат в переменной *pop_music_median* и выведите на экран.
 - 7). Рассчитайте среднее арифметическое времени прослушивания произведений жанра поп. Сохраните результат в переменной *pop_music_mean* и выведите на экран.
5. Решение задачи и оформление результатов
- 1). Рассчитайте метрику *happiness* для всего набора данных. Сохраните полученный результат в переменной *current_happiness* и выведите на экран.

Метрика *happiness* рассчитывается так: считаем, как долго каждый пользователь слушал музыку. Для этого сгруппируем *DataFrame* по пользователю. Посчитаем

общее время прослушивания музыки. Находим медианное значение для суммы прослушиваний по пользователю.

- 2). Рассчитайте разность двух значений метрики happiness до (57.456 секунд) и после эксперимента. Сделайте выводы об изменении удовлетворенности пользователей сервисом.

Литература:

1. Плас Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение. — (Серия «Бестселлеры O'Reilly»). - Санкт-Петербург : Питер, 2021. - 576 с. - ISBN 978-5-4461-0914-2. - URL: <https://ibooks.ru/bookshelf/376830/reading> (дата обращения: 11.06.2022). - Текст: электронный.

Раздел 3. Правовые и этические нормы в области искусственного интеллекта

Содержание:

Информационная безопасность, правовая база информационного законодательства, правовые и этические нормы в области искусственного интеллекта, международные и национальные стандарты в области искусственного интеллекта и смежных областях, защищенности информационных систем и систем искусственного интеллекта.

Самостоятельная работа по разделу:

По итогам самостоятельной работы студент готовит отчет, включающий в себя ответы на вопросы и решение заданий, предполагавшихся к выполнению в ходе самостоятельной работы. Отчет сдается преподавателю в электронной форме.

Задания для самостоятельной работы:

1. Разработайте интеллект-карту, отражающую правовую базу информационного законодательства.
2. Какие правовые и этические нормы в области искусственного интеллекта используются в настоящее время (приведите ссылки на законодательство).
3. Перечислите основные международные и национальные стандарты в области искусственного интеллекта и смежных областях.
4. Что подразумевает под собой «защищенность информационных систем и систем искусственного интеллекта».

Образцы заданий для лабораторных работ:

По итогам выполнения лабораторной работы студент демонстрирует результаты работы программы преподавателю, предварительно разработав тестовые случаи, а также сдает в электронном виде отчет, содержащий порядок выполнения работы.

Лабораторная работа «Правовые нормы в области искусственного интеллекта»

Тезис: система, которая учится на информации, полученной из внешнего мира, может действовать таким образом, который создатели не могли предсказать, а предсказуемость критична для современных правовых подходов.

Опишите примеры ситуаций, когда система искусственного интеллекта действует не так, как предсказывали создатели. Какую ответственность несет владелец системы искусственного интеллекта, если та совершила действия, нанеся вред третьей стороне. По какой статье гражданского кодекса будет наступать ответственность владельца на территории РФ? Может ли к владельцу системы искусственного интеллекта применяться правовые нормы об источниках повышенной опасности? Какие системы могут быть отнесены к источникам повышенной опасности? Может ли роботизированная система быть юридически ответственной за определенные действия?

Лабораторная работа «Этика и искусственный интеллект»

«Проблема вагонетки»: Представьте, что неуправляемая вагонетка несётся по рельсам, а на ее пути к рельсам привязаны пять человек. Вы стоите возле рычага, с помощью которого можно переключить стрелку, и вагонетка повернёт и поедет по другому пути, где к рельсам привязан один человек. Переключите ли Вы стрелку?

Опишите примеры ситуаций, связанных с использованием систем искусственного интеллекта, в которых «проблема вагонетки» должна быть решена. Представьте различные сценарии развития событий. Как нужно поступить в той или иной ситуации? Какие этические и правовые нормы используются должны быть использованы в приведенных ситуациях?

Некоторые примеры ситуаций: фильм «Я, робот» (чью жизнь нужно спасти?); авария при движении беспилотного транспорта (чья жизнь в приоритете – пассажиров или пешеходов?) и др.

Литература:

1. Сурова, Н. Ю. Искусственный интеллект / Н. Ю. Сурова, М. Е. Косов. – Москва : Юнити-Дана, 2021. – 408 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=690578> (дата обращения: 11.06.2022). – Библиогр.: с. 383-385. – ISBN 978-5-238-03513-0. – Текст : электронный.
2. Залоило М. В. Искусственный интеллект в праве: научно-практическое пособие / М.В. Залоило. - Москва : Инфотропик Медиа, 2021. - 132 с. - ISBN 978-5-9998-0365-8. - URL: <https://ibooks.ru/bookshelf/382542/reading> (дата обращения: 11.06.2022). - Текст: электронный.
3. Бахтеев Д. В. Искусственный интеллект: этико-правовые основы : монография / Д.В. Бахтеев. - Москва : Проспект, 2021. - 176 с. - ISBN 978-5-392-33413-1. - URL: <https://ibooks.ru/bookshelf/380167/reading> (дата обращения: 11.06.2022). - Текст: электронный.

Средства контроля качества обучения

Вопросы к экзамену:

1. Информатизация общества.
2. Информационно-коммуникационные технологии в мире, перспективы их развития.
3. Развитие искусственного интеллекта.
4. Главные тренды систем искусственного интеллекта.
5. Основные параметры идентификации задач искусственного интеллекта: назначение и сфера применения.

6. Основные параметры идентификации задач искусственного интеллекта: виды используемых знаний.
7. Основные параметры идентификации задач искусственного интеллекта: временные аспекты решения задач.
8. Классификация систем искусственного интеллекта.
9. Методы решения задач с использованием систем искусственного интеллекта в зависимости от особенностей проблемной области.
10. Инструментальные средства решения задач с использованием систем искусственного интеллекта в зависимости от особенностей проблемной области.
11. Функциональность программного обеспечения.
12. Методы оценки экономической эффективности применяемого программного и аппаратного обеспечения.
13. Основные инструменты, библиотеки и технологии Data Science.
14. Формирование требований к системе искусственного интеллекта.
15. Виды представления данных.
16. Методы поиска, сбора и обобщения информации о проблемной области. Опрос экспертов.
17. Методы сбора исходных данных о функционировании проблемной и предметной областей.
18. Документированные источники знаний.
19. Достоверные источники данных. Открытые источники данных.
20. Целостность и непротиворечивость данных.
21. Уровни представления данных (ODS, DDL, семантический слой, модель данных).
22. Предварительная обработка и разметка данных.
23. Информационная безопасность, правовая база информационного законодательства.
24. Правовые и этические нормы в области искусственного интеллекта.
25. Международные и национальные стандарты в области искусственного интеллекта и смежных областях.
26. Защищенности информационных систем и систем искусственного интеллекта.