

Документ подписан простой электронной подписью
 Информация о владельце:
 ФИО: Петренко Вячеслав Иванович
 Должность: и.о. директора Института цифрового развития
 Дата подписания: 01.06.2023 16:41:50
 Уникальный программный ключ:
 b47f96e6a03a76e175194ab7fb0b10000e7ad4d

Аннотация дисциплины

Наименование дисциплины	Обработка и анализ больших данных
Содержание	Статистические основы анализа данных: Роль анализа данных в современном мире. Научные исследования. Построение системы анализа данных. Источники данных. Сбор данных. Подготовка данных. Генеральная совокупность и выборка. Jupyter Notebook: интерактивные вычисления и среда. Типы переменных. Количественные и номинативные переменные. Меры центральной тенденции. Меры изменчивости. Квартили распределения и график box-plot. Нормальное распределение. Центральная предельная теорема. Основы NumPy: массивы и векторные вычисления: Доверительные интервалы для среднего. Идея статистического вывода, р-уровень значимости. Работа в интерактивной оболочке IPython. HTML-блокнот в IPython (Jupyter Notebook). Пользовательский интерфейс нового поколения JupyterLab. Назначение пакета NumPy. Сравнение NumPy с MATLAB. Основы работы с Pandas: Массивы NumPy. Векторизация вычислений в NumPy. Введение в структуры данных Pandas. Объект Series. Объект DataFrame. Индексные объекты Базовая функциональность Pandas. Переиндексация DataFrame. Удаление элементов из оси. Доступ по индексу, выборка и фильтрация. Переформатирование данных в Pandas: Арифметические операции и выравнивание данных. Применение функций и отображение. Сортировка и ранжирование. Индексы по осям с повторяющимися значениями. Обработка отсутствующих данных. Фильтрация отсутствующих данных. Замена отсутствующих данных. Иерархическое индексирование. Уровни переупорядочения и сортировки. Чтение и запись данных, форматы файлов, поддерживаемые Pandas: Сводная статистика по уровню. Работа со столбцами DataFrame. Комбинирование и слияние наборов данных. Слияние объектов DataFrame как в базах данных. Слияние по индексу. Конкатенация вдоль оси. Комбинирование перекрывающихся данных. Устранение дубликатов. Преобразование данных с помощью функции или отображения. Замена значений. Переименование индексов осей. Построение графиков и визуализация: Дискретизация и раскладывание. Обнаружение и фильтрация выбросов. Перестановки и случайная выборка. Вычисление индикаторных переменных. Чтение и запись данных в текстовом формате. Данные в формате JSON. Данные в формате XML. Двоичные форматы данных. Форматы данных BigData. Данные в формате Avro. Агрегирование данных и групповые операции Pandas: Данные в формате OCR. Данные в формате Parquet. Визуализация данных с помощью matplotlib. Встроенные средства визуализации pandas. Высокоуровневые средства визуализации seaborn. Средства интерактивной визуализации. Визуализация данных с помощью пакета Bokeh. Визуализация данных с помощью пакета Plotly. Визуализация данных с помощью пакета Altair. Работа с временными рядами в Pandas: Механизм GroupBy. Выборка столбца или подмножества столбцов. Группировка с помощью словарей

	и объектов Series. Группировка с помощью функций. Группировка по уровням индекса. Агрегирование данных. Применение функций, зависящих от столбца, и нескольких функций. Возврат агрегированных данных в "неиндексированном" виде.
Результаты освоения дисциплины (модуля)	Использовать программные и аппаратные средства персонального компьютера для анализа данных; ориентироваться в современной системе источников информации; квалифицировать задачи Data Mining, применять методы интеллектуального анализа данных; использовать современные информационные технологии в своей профессиональной деятельности; видеть и формулировать проблему; видеть конкретную ситуацию; прогнозировать и предвидеть; ставить цели и задачи. Представлять основные проблемы, возникающие при анализе данных, и пути их решения; отличия Data Mining от классических статистических методов анализа и OLAP-систем; типы закономерностей и сферы применения Data Mining; виды и способы организации хранилищ данных; классификацию аналитических систем; состав классов программных продуктов, образующих набор Business Intelligence. Понимать тенденции технологий интеллектуального анализа данных, стандарты и инструменты.
Трудоемкость, з.е.	3 з.е.
Формы отчетности	зачет с оценкой, 7 семестр
Учебно-методическое и информационное обеспечение дисциплины	
Основная литература	1. Крутиков, В. И; Анализ данных Электронный ресурс / Крутиков В. И., Мешечкин В. В. - Кемерово : КемГУ, 2018. - 138 с. - ISBN 978-5-8353-1770-7, экземпляров неограничено 2. Коэльо Луис Педро, Построение систем машинного обучения на языке Python: учебное пособие / Луис Педро Коэльо, Вилли Ричарт - Москва: ДМК Пресс, 2016. - 302 с. - ISBN 978-5-97060-330-7 - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : https://www.studentlibrary.ru/book/ISBN9785970603307.html (дата обращения: 12.03.2020). - Режим доступа : по подписке. 3. Data Mining – технологии обработки больших данных Электронный ресурс : Учебное пособие / В. И. Воронов, Л. И. Воронова, В. А. Усачев. - Data Mining - технологии обработки больших данных, 2024-02-26. - Москва : Московский технический университет связи и информатики, 2018. - 47 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 2227-8397, экземпляров неограниченно.
Дополнительная литература	1. Сузи, Р.А. Язык программирования Python Электронный ресурс : учебное пособие / Р.А. Сузи. - Язык программирования Python, 2020-07-28. - Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), 2016. - 350 с. - Книга находится в базовой версии ЭБС IPRbooks. - ISBN 5-9556-0058-2, экземпляров неограниченно. 2. Копырин, А. С. Программирование на Python Электронный ресурс / Копырин А. С., Салова Т. Л. : учебное пособие для студентов специальности 09.03.03 «прикладная информатика (в экономике)». - Сочи : СГУ, 2018. - 48 с., экземпляров неограниченно 3. Рутковская Д., Нейронные сети, генетические алгоритмы и нечеткие системы: учебное пособие / Д. Рутковская, М. Пилиньский, Л. Рутковский; пер. с польского И.Д. Рудинского. - 2-е изд., стер. - Москва: Горячая линия - Телеком, 2012. - 384 с. - ISBN 978-5-9912-0320-3 - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : https://www.studentlibrary.ru/book/ISBN9785991203203.html (дата

	обращения: 12.04.2021). - Режим доступа : по подписке. 4. Крутиков, В. Н. Анализ данных : учебное пособие / В.Н. Крутиков, В.В. Мешечкин ; Министерство образования и науки Российской Федерации ; Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Кемеровский государственный университет». - Кемерово: Кемеровский государственный университет, 2014. - 138 с. : ил. - http://biblioclub.ru/. - Библиогр. в кн. - ISBN 978-5-8353-1770-7, экземпляров неограничено
--	--