

Документ подписан простой электронной подписью
 Информация о владельце:
 ФИО: Петренко Вячеслав Иванович
 Должность: и.о. директора Института цифрового развития
 Дата подписания: 15.04.2024 15:20:29
 Уникальный программный ключ:
 b47f96e6a03a7c611914ab7fb0b10000e7ad4d

Аннотация дисциплины

Наименование дисциплины	Анализ данных
Краткое содержание	Компоненты систем Data Mining. Введение в интеллектуальный анализ данных. Области использования. Компоненты систем Data Mining. Основные этапы анализа данных. Исторический обзор. Основные математические пакеты языка Python для анализа данных: NumPy, matplotlib, Pandas, SciPy. Предварительная обработка данных. Очистка данных. Интеграция и преобразование данных. Сокращение данных. Основы работы с Docker. Введение в Big Data и NoSQL СУБД ClickHouse. Архитектура Apache Hadoop. Основные компоненты Apache Hadoop: HDFS, MapReduce, Hive, HBase. Основные форматы данных Apache Hadoop: Avro, Parquet и ORC. Архитектура Apache Spark. Основные компоненты Spark Core. Основы работы с Apache Kafka. Обработка потоковых данных в Apache Spark. Управление данными ETL/ELT. Архитектура хранилищ Data Warehouse и Data Lake. Облачные решения для анализа данных. Планирование задач обработки данных в Apache Airflow. Работа с данными в Superset и Microsoft Power BI.
Результаты освоения дисциплины (модуля)	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. Использовать основные возможности пакета pandas для решения задач анализа данных. Применять основной инструментарий для работы с Big Data. 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. Использовать основной инструментарий Data Mining; средства многомерного анализа Data Mining. Понимать тенденции технологий интеллектуального анализа данных, стандарты и инструменты.
Трудоемкость, з.е.	8
Формы отчетности	зачет с оценкой, экзамен
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