

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Макаренко Елена Николаевна  
Должность: Ректор  
Дата подписания: 28.04.2023 17:29:06  
Уникальный программный ключ:  
с098bc0d111111111111111111111111

## АННОТАЦИЯ рабочей программы дисциплины

### Программирование аналитических приложений на базе хранилищ данных

#### 1. Общая трудоёмкость

Трудоёмкость дисциплины составляет 7 зачётных единиц (252 часа), из них 18 часов лекционных занятий, 36 часов практических занятий.

#### 2. Место дисциплины в структуре образовательной программы

Дисциплина относится к модулю профессиональных дисциплин, формируемому участниками образовательных отношений, части образовательной программы, формируемой участниками образовательных отношений.

Для изучения данной дисциплины необходимы знания, умения и навыки, формируемые предшествующими элементами образовательной программы: Программное и аппаратное обеспечение информационных систем; Экспертные системы и базы знаний.

Результаты обучения, формируемые данной дисциплиной, потребуются при освоении следующих элементов образовательной программы: производственная практика, преддипломная практика; выполнение и защита выпускной квалификационной работы..

#### 3. Цель изучения дисциплины

углубление знаний в вопросах применения методов аналитической обработки сверхбольших объемов информации, накапливаемой в современных хранилищах данных, совершенствование умений практического использования методов математической статистики и машинного обучения для решения задач обработки данных, а также совершенствование навыков использования аналитического программного обеспечения, предназначенного для интеллектуального анализа данных, направленного на формирование целостного представления об анализе и интерпретации данных, как о процессе поиска, так и применения скрытых в них закономерностей для достижения поставленных целей.

#### 4. Содержание дисциплины

Модуль 1.

Тема 1. Введение в анализ данных

Принципы анализа данных. Структурированные данные. Подготовка данных к анализу. Технологии KDD и Data Mining. Аналитические платформы»

Тема 2. Консолидация данных

Введение в хранилища данных. Основные концепции хранилищ данных. Многомерные хранилища данных. Реляционные хранилища данных. Гибридные хранилища данных. Виртуальные хранилища данных

Тема 3. Введение в ETL

Извлечение данных в ETL. Очистка данных в ETL. Преобразование данных в ETL. Загрузка данных в хранилище. Загрузка данных из локальных источников. Обогащение данных»

Тема 4. Введение в трансформацию данных

Трансформация упорядоченных данных. Группировка данных. Слияние данных. Квантование. Транспонирование. Нормализация и кодирование данных»

Модуль 2.

Тема 5. Введение в визуализацию

Визуализаторы общего назначения. OLAP-анализ. Визуализаторы для оценки качества моделей. Визуализаторы для интерпретации результатов анализа»

Тема 6. Оценка качества данных

Технологии и методы оценки качества данных. Очистка и предобработка данных. Фильтрация данных. Обработка дубликатов и противоречий. Выявление аномальных значений

Тема 7. Восстановление пропущенных значений

Введение в сокращение размерности. Сокращение числа признаков. Сокращение значений и записей. Сэмплинг

Тема 8. Введение в ассоциацию

Ассоциативные правила. Алгоритм Apriori. Иерархические ассоциативные правила. Введение в кластеризацию. Алгоритм кластеризации k-means. Сети Кохонена. Карты Кохонена. Проблемы алгоритмов кластеризации

### **5. Дополнительная полезная информация**

Дисциплина предназначена для формирования элементов следующих компетенций образовательной программы:

ПК-1. Способен адаптировать и применять методы и алгоритмы машинного обучения для решения прикладных задач в различных предметных областях

Форма промежуточной аттестации: экзамен.

Наименование оценочного собеседование по результатам выполнения практических работ; контрольная работа №1; контрольная работа №2; экзаменационные вопросы и билеты.