

Аннотация дисциплины М.1.2.1 Дисциплина. Анализ больших данных

Дисциплина "Анализ больших данных" изучается обучающимися по основной профессиональной образовательной программе "Интеллектуальные системы" направления подготовки "09.04.01 Информатика и вычислительная техника".

Дисциплина изучается в 3 триместре. Общая трудоемкость дисциплины составляет 144/4 часов/з.ед. Самостоятельная работа заключается в выполнении работ, указанных в разделе 4.

В ходе изучения дисциплины осуществляется текущий контроль в форме технологии рейтингового контроля в соответствии с технологической карты дисциплины, размещенной на электронном курсе, а также промежуточный контроль в форме балльно-рейтинговый контроль.

Целью изучения дисциплины является формирование следующих компетенций:

1. ПК-4 Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач искусственного интеллекта
2. ПК-5 Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях

В ходе изучения дисциплины последовательно рассматриваются темы:

1. Предмет курса. Определение Больших Данных. Области применения и примеры использования.
Сосредоточенные и распределенные базы данных.
CAP-теорема и ее следствия для бизнеса.
Серверные кластеры. Репликация и шардинг данных. Программные продукты для анализа Больших Данных.
2. Модели данных. NoSQL, базы данных. Сравнение NoSQL и реляционных баз данных.
3. Понятие хранилища данных. Технологии и архитектура систем хранения данных. Системы DAS, SAN и NAS. RAID-массивы. Многоуровневое хранение данных. Распределенные хранилища данных. Облачные хранилища.
4. Источники данных – Web-сайты, социальные сети, мобильные устройства, данные с измерительных устройств, видео- и аудио-оборудование, бизнес-информация и корпоративные БД. Методы сбора данных. Подготовка данных к анализу (алгоритмы выборки, очистки, сортировки);
Введение в распределенную файловую систему Hadoop. Технологии распределенной файловой системы HDFS. Концепции и терминология базы данных Oracle NoSQL. Введение в Oracle NoSQL - архитектура и компоненты.
5. Модель вычисления Map Reduce. Этапы вычислений.
6. Поиск закономерностей в данных. Визуализация данных. Статистический анализ данных. Кластерный анализ. Модели временных рядов.
7. Алгоритмы машинного обучения для решения задач регрессии, классификации и кластеризации. Алгоритмы работы рекомендательных систем и прогнозирования. Методы бустинга в задачах распознавания. Методы видеоаналитики.
Нейронные сети в задачах классификации. Задача классификации изображений. Оценка качества классификации.
8. Алгоритмы Data mining — глубинный анализ данных.
Классификация. Ассоциация. Регрессионный анализ.

Основными стратегическими образовательными технологиями являются: лекционные занятия, практика.

В рамках указанных технологий применяются тактические образовательные технологии: классическая лекция.