

Аннотация дисциплины М.1.1.7 Дисциплина. Анализ больших данных

Дисциплина "Анализ больших данных" изучается обучающимися по основной профессиональной образовательной программе "Искусственный интеллект в строительной отрасли" направления подготовки "08.04.01 Строительство".

Дисциплина изучается в 3 семестре. Общая трудоемкость дисциплины составляет 48/6 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, курсовая работа.

Целью изучения дисциплины является формирование следующих компетенций:

1. ПК-3 Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика
2. УК-6 Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки

В ходе изучения дисциплины последовательно рассматриваются темы:

1. Предмет курса. Основные понятия, области применения и примеры использования Больших Данных. Архитектура хранилищ Больших Данных. Распределенные базы данных (БД). CAP-теорема и ее следствия для бизнеса. Введение в экосистему Hadoop.
2. Модели данных. NoSQL, базы данных. Операции CRUD в реляционных и NoSQL БД. Технология MapReduce для работы с BigData.
3. Структурированные и неструктурированные данные. Подготовка данных к анализу. Алгоритмы выборки, очистки, сортировки;
4. Поиск закономерностей в данных. Визуализация данных. Статистический анализ данных. Кластерный анализ. Модели временных рядов.
5. Основы языка Python. Библиотеки для анализа данных (NumPy, Matplotlib, scikit-learn)
6. Смешение и интеграция данных - набор техник интегрировать разнородных данных. Имитационное моделирование – построение моделей имитирующих реальные процессы, разновидность экспериментальных испытаний. Пространственный анализ - класс методов, использующих топологическую, геометрическую и географическую информацию в данных.
7. Алгоритмы машинного обучения для решения задач регрессии, классификации и кластеризации. Алгоритмы работы рекомендательных систем и прогнозирования. Методы бустинга в задачах распознавания и видеоаналитики. Нейронные сети в задачах классификации. Задача классификации изображений. Оценка качества классификации.
8. Алгоритмы Data mining — глубинный анализ данных. Классификация. Ассоциация. Регрессионный анализ. Задача и алгоритмы видеоаналитики.

Основными стратегическими образовательными технологиями являются: лекционные занятия, практические занятия.

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