Аннотация дисциплины Б.1.2.15 Дисциплина. Обработка больших данных

Дисциплина "Обработка больших данных" изучается обучающимися по основной профессиональной образовательной программе "Интеллектуальные технологии в приборостроении" направления подготовки "12.03.01 Приборостроение".
Дисциплина изучается в 7 семестре. Общая трудоемкость дисциплины составляет 108/3 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 зачет.
Целью изучения дисциплины является формирование следующих компетенций:

1. ПК-2 Способность к компьютерному моделированию с использованием методов системного подхода для определения принципа функционирования и параметров разрабатываемых оптических и оптико-электронных приборов, комплексов и их составных частей
2. ПК-4 Способность к созданию интегрированных оптических и оптико-электронных приборов и комплексов для решения сложных задач с учетом требований технического задания и возможностей организации-изготовителя

В ходе изучения дисциплины последовательно рассматриваются темы:

1. Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных.
2. Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных.
3. Технологии KDD и Data Mining. Подготовка данных к анализу. Методика извлечения знаний. Data Mining. Мультидисциплинарный характер Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining как средств обработки больших объемов информации.
4. Программное обеспечение в области анализа данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования.
5. Ассоциативные правила. Аффинитивный анализ, предметный набор. Поддержка и достоверность ассоциативного правила. Значимость ассоциативных правил, лифт и левередж. Поиск ассоциативных правил. Частые предметные наборы и их обнаружение. Алгоритм генерации ассоциативных правил. Иерархические ассоциативные правила. Методы поиска иерархических ассоциативных правил.
6. Определение кластеризации. Постановка задачи кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Меры расстояний. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации.
7. Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразие подходов.

Основными стратегическими образовательными технологиями являются: дискуссионные, лекционные занятия, практические занятия, процедуры самообучения.
В рамках указанных технологий применяются тактические образовательные технологии: задания, классическая лекция, лекция с элементами мозгового штурма.