

Аннотация дисциплины Б.1.2.13 Дисциплина. Обработка больших данных

Дисциплина "Обработка больших данных" изучается обучающимися по основной профессиональной образовательной программе "Искусственный интеллект в бизнес-аналитике" направления подготовки "09.03.03 Прикладная информатика".

Дисциплина изучается в 6 семестре. Общая трудоемкость дисциплины составляет 180/5 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 курсовая работа, экзамен.

Целью изучения дисциплины является формирование следующих компетенций:

1. BD-1 Способен осуществлять поиск, сбор, очистку и предварительный анализ данных
2. BD-2 Способен определять требования к наборам данных для решения задач машинного обучения, проводить разметку и анализ наборов данных, оценивать качество данных, обеспечивать непрерывную интеграцию данных
3. BD-4 Способен применять различные модели и (или) технологии хранения данных

В ходе изучения дисциплины последовательно рассматриваются темы:

1. Лекция 1. Предмет курса. Понятие Big Data: 3V (Volume, Velocity, Variety) и расширенные V (Veracity, Value). Жизненный цикл данных (Data Lifecycle).
2. Лекция 2. Обзор типичных задач обработки больших данных: очистка данных, разведочный анализ, визуализация, машинное обучение. Знакомство с библиотеками Python для работы с большими данными: NumPy, Pandas, Matplotlib и надстройка Seaborn, Plotly.
3. Лекция 3. Структурированные и неструктурированные данные. Алгоритмы сбора и предварительного анализа данных: парсинг веб-страниц, работа с API, сбор логов.
4. Лекция 4. Алгоритмы предобработки больших данных: процессы очистки данных, работа с пропусками, дубликатами, выбросами. Разметка данных. Инструменты разметки данных.
5. Лекция 5. Разметка и анализ качества данных, обеспечение непрерывной интеграции данных.
6. Лекция 6. Параллельная обработка больших данных. Технология MapReduce для работы с BigData.
7. Лекция 7. Знакомство с методами: метод Монте-Карло, метод Монте-Карло на основе Марковской цепи, методы блужданий.

Основными стратегическими образовательными технологиями являются: имитационное моделирование, исследовательские, лекционные занятия, практические и лабораторные занятия, процедуры самообучения, лекция с обсуждением новых научных достижений в сфере ии по результатам конференции и публикаций, открытые образовательные ресурсы.

В рамках указанных технологий применяются тактические образовательные технологии: задания, информационные, классическая лекция, лекция с элементами мозгового штурма, мини-проекты, case-study, проблемная лекция, семинары с ит-практиками, конкурсы, хакатоны, кейс-чемпионаты, мастер-классы, командные проектные задания.