Бауманский учебный центр «Специалист»

Программа курса Основы работы с большими данными (Data Science)

Код курса: **ДС0**

Тема	Ак. часов
Модуль 1. Область применения больших данных. Типовые задачи • Цели курса • Определение основных понятий • История науки о данных • Выгоды от работы с большими данными • Типовые задачи: прогноз продаж, производства, спроса. Анализ поведения. Распознавание образов. Экспертные системы.	**1**
Модуль 2. Сбор и подготовка исходных данных. Методика CRISP-DM • С чего начать. Межотраслевая стандартная методика работы с данными CRISP-DM. • Описательное и ассоциативное исследование исходных данных. • Сегментирование и очистка данных (slice and dice). • «Озера данных» (Data lakes). • Отличия озер данных от структурированных хранилищ • Обзор инструментария. • **Практическая работа.** Сегментировать и очистить тестовый набор данных.	**2**
Модуль 3. Задачи численного прогнозирования. Понятие машинного обучения. Корреляция. Регрессионный анализ. • Виды распределений. • Постановка задачи оценки взаимосвязи между различными факторами и построение прогноза. • Корреляция. Коэффициент Пирсона. • Критерий Стьюдента (Т-анализ). • Основы машинного обучения. • Регрессионный анализ. • **Практическая работа.** Определить наличие корреляции и регрессионную зависимость между двумя выборками данных. • **Практическая работа.** Определить статистические характеристики выборки данных. • Обзор прикладных средств работы с данными.	**2**
Модуль 4. Задачи классификации и распознавания образов, видео, речи, текста. Понятие нейронных сетей. Примеры применения • Задача сегментации дискретных данных на примере задач распознавания (графика, речь, текст). • Нейронные сети как инструмент решения задач классификации. • Разбиение на выборки (обучающую, тестовую, проверочную). • Анализ ошибок обучения. Базис и отклонения. Ручная корректировка. • Демонстрация источников библиотек и no-code. • **Опциональная практика.** Работа с nocode приложением.	**3**

Модуль 5. GPT нейросети. глубокое машинное обучение, искусственный интеллект. Помощники и агенты • Понятие Deep Machine Learning. • Генеративные (GPT) сети. Конкурентное обучение. • Понятия больших, малых и микро-моделей (LLM, SLM, MLM). • ИИ помощники – общего назначения и специализированные. • ИИ помощники (ассистенты) и ИИ агенты. • Фокусированный промптинг. Диалог с ИИ помощником (ассистентом). • **Практическая работа.** Взаимодействие с ИИ помощником.	**3**
Модуль 6. Задачи построения маршрутов (логистика, соцсети, телеком). Задача прогнозирования поведения пользователя. Понятие графов • Задачи классификации данных в социальных сетях и поиска оптимального решения (маршрута). • Графы как инструмент решения задач на социальных, логистических, телекоммуникационных графах и прогнозирования поведения. • Дерево решений. • **Практическая работа:** провести классификацию набора данных и его разбиение на сегменты.	**3**
Модуль 7. Обработка естественного языка. Нечеткие множества • Понятия естественного языка и нечетких логик. • Многофакторный бизнес-анализ на примере нечетких логик.	**1**
Модуль 8. Профориентация по специальностям и в Data Science • Роли специалистов по DS: аналитик данных, ученый по данным, программист, цифровой директор. • Требования к компетенциям. • Состав и требования к проектной команде для DS.	**1**
Аудиторная нагрузка в классе с преподавателем	**16 ак.ч. + 16 ак.ч. самостоятельно**

По окончании обучения на курсе проводится итоговая аттестация. Аттестация проводится в виде теста на последнем занятии или на основании оценок практических работ, выполняемых во время обучения на курсе.

Контакты центра

Для частных лиц
+7 (495) 232-32-16

Для организаций
+7 (495) 780-48-44

E-mail
info@specialist.ru