OTUS

Big Data для менеджеров

Погружение бизнес-специалистов без технического бэкграунда в экосистему Data Science-проектов. Раскрываем принципы Data-driven трансформации бизнеса, изучаем технологии анализа данных. Разбираем подводные камни, сложности внедрения и реализации проектов.

Длительность курса: 102 академических часа

1 Введение в дисциплину управления данными

1	**Введение в управление данными организации**	Познакомим с актуальностью проблемы, основными терминами дисциплины управления данными, историей развития и основными историческими событиями и основами дисциплины: - проблематика и обзор современных тенденций; - Большие Данные, основы концепции, терминология и область применимости; - закон Мура; - понятие Gartner Hype Cycle. - Скорость и стоимость вычислений, следствие закона Мура - Цифровизация - Big Data maturity index; - индекс индустриальной цифровизации; - Искуственный интеллект - Основные проблемы внедрения AI в бизнес компаний - Роль CDO в процессе data-трансформации
2	**Концепция "Data-Driven организация"**	Раскрываем основные аспекты концепции "data-driven организация", в чем ценность и выгоды данного подхода: - Данные — определения, термины, процесс преобразования данных, характеристики данных, качество данных; - откуда берутся большие данные, источники, машинные данные и способы их получения, рынок данных; - основные принципы data-driven управления и роль data-driven подхода в бизнес-трансформации; - основные виды аналитики данных; - Data-аудит организации.

2 Роль данных в трансформации бизнеса (Data-Driven Business)

1	**Введение в стратегический менеджмент организации**	- Введение в стратегический менеджмент организации, виды бизнес-процессов - Основные тренды развития экономики, бизнес-среды и теории управления бизнесом - Теория экспоненциальных организаций, основные области автоматизации бизнеса для достижения конкурентных преимуществ
2	**Использование данных для оптимизации управления организацией**	- Поддержка принятия решений - Мониторинг информационного пространства - Автоматизация принятия решений и основные преимущества такого подхода
3	**Использование данных для оптимизации функционирования организации**	- Оптимизация бизнес-процессов компаний - Примеры реальный кейсов в областях: маркетинг, финансы, логистика, производство, обслуживание клиентов, профилактика и диагностика оборудования, продажи

3 Основы аналитики данных

1	**Введение в анализ данных**	- Виды анализа данных - Обзор основных задач Машинного обучения
2	**Базовые элементы теории вероятности и статистики**	Базовые понятия теории вероятности: - математическое ожидание; - теорема Байеса; - Центральная Предельная Теорема; - Закон Больших Чисел; - основные распределения. Базовые элементы статистики: - построение гипотез; - проверка гипотез с помощью тестов.
3	**Базовые элементы Python**	Базовые элементы языка Python: - основы синтаксиса; - работа с библиотеками Numpy, Scipy; - работа с Pandas; - работа с Sklearn;
4	**Базовые алгоритмы кластеризации**	Базовые алгоритмы кластеризации: - Kmeans; - иерархическая кластеризация; - Dbscan; - метрики качества для задачи кластеризации.

5 **Основные алгоритмы машинного обучения и метрики качества**

Базовые алгоритмы машинного обучения:
- логистическая регрессия;
- деревья решений;
- метод ближайших соседей.

Базовые метрики качества для задач:
- классификации;
- регрессия.

---

6 **Работа с ансамблями**

Работа с ансамблями:
- Random Forest;
- Gradient Boosting.

Домашние задания

1 Анализ поведения человека по данным смартфона

В практической работе будет рассмотрен рабочий пример анализа поведения человека по данным смартфона. Будет проведено сравнение моделей классификации и продемонстрированы признаки переобучения. Будет рассмотрен процесс подбора параметров алгоритма RandomForest и проведена интерпретация результатов.

---

7 **Базовые элементы нейронных сетей**

Базовые понятия для работы с нейронными сетями:
- сигмойда и другие функции активации;
- метод обратного распространения ошибки;
- глубокие нейронные сети;
- автокодировщики;
- рекуррентные нейронные сети;
- сверточные нейронные сети.

Домашние задания

1 использование нейронной сети для оценки эмоциональной окраски текста

Цель: В практической работе будет продемонстрировано, каким образом построить и натренировать нейронную сеть для решения задачи оценки эмоциональной окраски текстов.

---

8 **Рекомендательные системы**

Основные подходы в построении рекомендательных систем:
- коллаборативная фильтрация;
- ассоциативные правила;
- применение нейронных сетей для построения рекомендаций.

4 Основы технологий обработки больших данных (Data Engineering)

1 **Основы теории управления данными. Ограничения и трудности классического подхода**

Изучение теоретических основ:
- управление данными;
- ограничения и трудности классических подходов хранения и обработки данных;
- вопросы масштабирования систем обработки данных;
- виды и методы масштабирования систем хранения и обработки данных.

Домашние задания

1 Основы теории управления данными. Quiz.

Цель: Закрепляем пройденный в лекции материал

Для выполнения домашнего задания вам необходимо:
- Пройти по ссылке https://forms.gle/tz4HbwrKMvN1vkX49
- Ответить на вопросы
- В личном кабинете OTUS написать преподавателю, что тест пройден.

2 **Распределенные файловые системы. Введение в концепцию Map-Reduce**

Знакомство с:
- распределенными файловыми системами;
- объектными хранилищами данных;
- отличиями распределенных файловых систем от объектных хранилищ;
- представителями распределенных файловых систем и объектных хранилищ.

Введение в концепцию Map-Reduce:
- знакомство с историей и предпосылками;
- теоретические основы Map-Reduce;
- практическое применение парадигмы Map-Reduce.

Домашние задания

1 Распределенные файловые системы и Map-Reduce. Quiz.

Цель: Закрепляем пройденный в лекции материал

Для выполнения домашнего задания вам необходимо:
- Пройти по ссылке https://forms.gle/UsFF7kgrsvexiqaZ9
- Ответить на вопросы
- В личном кабинете OTUS написать преподавателю, что тест пройден.

3 **Введение в Hadoop. Экосистема Hadoop**

Введение в Hadoop:
- история Hadoop и критерии его применимости;
- Hadoop и его составные части;
- распределенная файловая система hdfs;
- Yarn и управление ресурсами;
- Yarn и map-reduce;
- дистрибутивы Hadoop;
- сайзинг.

Экосистема Hadoop:
- обзор экосистемы Hadoop;
- обзор hive, spark, impala, presto, pig;
- обзор oozie, airflow.

Домашние задания

1 Экосистема Hadoop. Quiz.

Цель: Закрепляем пройденный в лекции материал

Для выполнения домашнего задания вам необходимо:
- Пройти по ссылке https://forms.gle/Frujiq6uPRkf34mT6
- Ответить на вопросы
- В личном кабинете OTUS написать преподавателю, что тест пройден.

4 **Платформы хранения данных класса NoSQL. Платформы обработки данных реального времени**

Платформы хранения данных класса NoSQL:
- предпосылки;
- обзор экосистемы;
- SQL;
- NoSQL (key-value, document, wide-column, graph);
- NewSQL;
- In-Memory DataGrids.

Платформы обработки данных реального времени:
- предпосылки;
- обзор экосистемы;
- обработка данных;
- доставка данных;
- Spark Streaming, Flink, Samza, Storm, Heron, и др.;
- Kafka, Pulsar и др.

Домашние задания

1 NoSQL & Stream Processing. Quiz.

Цель: Закрепляем пройденный в лекции материал

Для выполнения домашнего задания вам необходимо:
- Пройти по ссылке https://forms.gle/8oMMEjy95m5Cjuja6
- Ответить на вопросы
- В личном кабинете OTUS написать преподавателю, что тест пройден.

5	**Интеграция. Визуализация. Управление**	- Интеграция данных - Средства визуализации данных - Управление ресурсами и инфраструктурой обработки данных Домашние задания 1 Интеграция и визуализация данных. Quiz. Цель: Закрепляем пройденный в лекции материал Для выполнения домашнего задания вам необходимо: - Пройти по ссылке https://forms.gle/k82vMenqAqLmxF366 - Ответить на вопросы - В личном кабинете OTUS написать преподавателю, что тест пройден.

6 **Комплексные архитектуры хранения и обработки данных**

Комплексные архитектуры хранения и обработки данных:
- корпоративное хранилище данных (DWH);
- озеро данных (Data Lake);
- отличия и сходства хранилищ данных и озер данных;
- лямбда-архитектура;
- каппа-архитектура;
- дзета-архитектура.

Домашние задания

1 Комплексные архитектуры хранения и обработки данных. Quiz.

Цель: Закрепляем пройденный в лекции материал

Для выполнения домашнего задания вам необходимо:
- Пройти по ссылке https://forms.gle/QyW2LkGhAx9Kgo3P6
- Ответить на вопросы
- В личном кабинете OTUS написать преподавателю, что тест пройден.

5 Основы реализации проектов по аналитике данных (Data Science Management)

1	**Основные этапы проекта по анализу данных**	- Обзор имеющихся методологий - Методология CRISP-DM и ее особенности - Основные этапы проекта по анализу данных, постановка задачи, оценка результата, коммуникация результата - Оценка экономической эффективности
2	**Особенности управления проектами, связанными с аналитикой и большими данными**	- Построение проектной команды, различные организационные структуры - Управление персоналом, вопросы найма и развития компетенций - Взаимодействие с бизнес-заказчиком - Вопросы выбора инструментария - Практические рекомендации и подводные камни

6 Основы управления данными организации (Data Governance)

1	**Основы дисциплины Data Governance, часть 1**	- Основные цели и задачи - Развитие компетенций, развитие организационной структуры и культуры организации - Вопросы управления качеством данных и мета-данными, политиками ввода данных, интеграции данных - Вопросы аутсорсинга, краудсорсинга, инсорсинга - Качество данных
2	**Основы дисциплины Data Governance, часть 2**	- Вопросы развития инфраструктуры (облачные модели, собственная инфраструктура) - Особенности законодательства и регуляторных требований по сбору и обработки данных (модели США, Европейского союза, Китая, обзор текущей ситуации Российского рынка, введение в GDPR) - Вопросы этики, приватности, владения данными, безопасности данных - Концепция Human in the loop
3	**Дорожная карта бизнес-трансформации**	- Индекс цифровой зрелости организации - Стратегия монетизации данных, новые бизнес модели в эпоху больших данных, Индустрия 4.0 - Дорожная карта бизнес-трансформации, карта бенефициаров

7 Проектная работа

1	**Консультация по проектной работе**	Слушатели курса смогут определиться с темой проекта и получить понимание, какие ресурсы им необходимо использовать для работы. Домашние задания 1 Проектная работа
2	**Консультация по проектной работе**	Слушатели курса получат комментарии относительно прогресса проектной работы, ответы на вопросы, рекомендации по реализации.
3	**Защита проектных работ**	По окончании занятия слушатели курса получат разбор проектов, комментарии и оценку своей работы.