

Курс: PracticalMachine Learning with Python

Длительность: 32ак. часов

О курсе

В ходе курса слушатели не только познакомятся с основными концепциями и лучшими практиками в машинном обучении, но и научатся решать следующие задачи:

1. Кластеризация пользователей с целью создания уникального предложения для каждой группы пользователей.
2. Прогноз оттока пользователей.
3. Скоринг клиентов на возврат займа.

Особенности курса:

1. Упор идет не на теорию, а на практику и применимость в реальном бизнесе.
2. Вся теория подкреплена реальными данными из бизнеса.
3. В конце каждой темы задания на проверку материала. Каждый слушатель получает индивидуальный фидбек от преподавателя.

Аудитория

Продуктовые и бизнес аналитики, руководители, инженеры данных.

Предварительная подготовка

-

Программа

Неделя 1. Python3 для анализа данных.

Что такое Jupyter Anaconda.

Основные структуры данных в Python.

Операторы и управляющие конструкции.

Функции и классы.

Библиотека pandasдля анализа данных.

Визуализация данных в Python.

Домашнее задание: exploratorydataanalysisна реальном датасете.

Неделя 2. Основные концепции в машинном обучении.

Какие бывают задачи машинного обучения.

Как оценивать качество алгоритма: разбор метрик качества в задачах классификации и регрессии, их связь и решением бизнес-задачи.

Что такое валидация, какие бывают типы валидации.

Что такое регуляризация.

Работа с категориальными и пропущенными признаками.

Градиентный спуск – краеугольный камень всего машинного обучения.

Тест на понимание концепций.

Неделя 3. Основные алгоритмы.

Линейная регрессия, L1/L2 регуляризация.

Обобщения линейной регрессии: квантильная регрессия, полиномиальная регрессия, логистическая регрессия.

Метод k-ближайших соседей

Деревья принятия решений, регуляризация деревьев, модификации деревьев: bagging, randomforests, boosting.

Реализуем каждый из алгоритмов на реальном датасете, сравниваем их предсказательную силу.

Домашнее задание: самостоятельно произвести сравнение всех пройденных алгоритмов на других данных.

Неделя 4. Обучение без учителя, продвинутые подходы к работе с признаками, финальный проект.

Алгоритмы кластеризации, их оценка и область применения в реальном бизнесе.

Обзор подходов к featureengineering/featureselection. Использование генетических алгоритмов для создания новых признаков.

Продвинутые подходы к meantargetencoding.

Создание композиций алгоритмов: blending, stacking.

Объединяем полученные знания для построения production-ready модели скоринга.

Домашнее задание: реализация полного цикла data science проекта: EDA; формализация задачи; выбор метода валидации и метрики качества; генерация, оценка и отбор признаков; оценка и выбор модели, эксперименты с композициями моделей.