

Программа курса

«VIP: Визуализация данных на языке Python»

О курсе: Данный курс является введением в тематику визуализации данных. В курсе будут описаны основные подходы к визуализации данных, позволяющие наглядно подчеркнуть имеющиеся в них закономерности и особенности. Вы познакомитесь с возможностями библиотек визуализации на языке Python (matplotlib, seaborn, plotly, pandas, bokeh) для решения таких задач, как построение графиков функций распределения, отображение статистических закономерностей, визуализация изменяющихся во времени данных, демонстрация взаимосвязи отдельных компонент данных и многих других. Также вы узнаете, как можно использовать полученные визуальные представления данных для их дальнейшей подготовки и обработки. В курсе делается уклон в сторону развития практических навыков по визуализации данных и даются только самые необходимые теоретические сведения. Самостоятельный итоговый проект, который будет затрагивать все пройденные в рамках курса тематики, продемонстрирует приобретенные навыки и послужит хорошим сборником рецептов для вашей дальнейшей работы.

Аудитория: Специалисты по работе с большими данными, разработчики, руководители, желающие понять подходы к визуализации данных как при построении стратегий развития бизнеса с использованием больших данных, так и решения бизнес задач на основе алгоритмов машинного обучения, а также получить практический навык в данной области.

Уровень подготовки: опыт программирования на языке Python, знание основ математического анализа и математической статистики

Продолжительность курса: 32 академических часа, 8 дней по 4 ак. часов дистанционно

Содержание программы

1. Основные статистические характеристики данных. Графики функций распределения и визуализация зависимости компонент данных

Цель: познакомить участников с основными статистическими характеристиками данных, их визуализацией, анализом и подходами к обработке данных, использующими данные характеристики

Теоретическая часть:

- обзор библиотек языка Python для обработки данных (pandas) и визуализации (matplotlib, seaborn, plotly, pandas, bokeh)
- краткий обзор основных статистических приемов работы с данными:
 - первичный анализ данных
 - получение описательных статистик
 - изменение типа данных
 - построение сводных таблиц
- визуализация статистических характеристик данных
 - гистограммы
 - графики плотностей распределений
 - «ящики с усами»
 - «виолончели»
- визуализация зависимости компонент данных, отбор значимых компонент и понижение размерности данных
 - классические и пузырьковые диаграммы рассеяния
 - тепловые карты
 - таблица со встроенными диаграммами

Практическая часть: решение практических задач по построению графиков одномерных и

многомерных функций распределения, точечных диаграмм, тепловых карт.

2. Сравнительный анализ компонент данных

Цель: познакомить участников с подходами к анализу компонент данных относительно временных изменений и сравнению компонент данных с использованием группировки и агрегирования

Теоретическая часть:

- сравнительные графики вдоль временной оси
 - линейные диаграммы
 - вертикальные столбиковые диаграммы
 - круговые диаграммы площади
- сравнительные графики с использованием группировки и агрегирования
 - группировка и агрегирование данных в библиотеке pandas
 - вертикальные столбиковые диаграммы переменной ширины
 - таблица со встроенными диаграммами

Практическая часть: решение задач анализа изменяющихся во времени данных, а также подходам к анализу данных с использованием группировки и агрегирования.

3. Композиционные методы визуализации данных

Цель: познакомить участников с основными подходами к композиционным представлениям данных при работе со статичными и изменяющимися во времени данными

Теоретическая часть:

- композиционные графики изменяющихся во времени компонент данных
 - столбиковые диаграммы со стыковкой компонент данных
 - площадные диаграммы со стыковкой компонент данных
- композиционные графики статичных данных
 - круговая диаграмма
 - древовидная карта
 - столбиковые диаграммы со стыковкой компонент данных

Практическая часть: анализ данных с использованием композиционных представлений компонент данных.

4. Проектная работа

Цель: закрепить полученные слушателями курса знаний по визуализации данных

Теоретическая часть: краткий обзор пройденного материала с отсылками на рабочие блокноты, в которых решалась та или иная задача визуализации данных

Практическая часть: самостоятельное решение задачи анализа данных с использованием всех изученных подходов визуализации данных на собственной базе данных или же на лабораторном наборе, предоставляемом организаторами курса. Итоговый разбор работ слушателей курса.