

План курса

Основная часть: упор делается на статистику.

1. Введение в R. Установка R+RStudio. Основы языка:
 - основные контейнеры хранения данных: переменные, массивы, матрицы
 - циклы for и while
 - оператор if
 - функции, глобальные и локальные переменные
2. Основы RMarkdown для лаконичных отчетов. Введение в теорию вероятностей и математическую статистику (случайная величина, распределения, математическое ожидание).
3. Продолжение 2 лекции: t-test и доверительные интервалы. Разведочный анализ данных. Среднее, медиана, дисперсия, среднеквадратическое отклонение, процентиля, корреляция и так далее.
4. Введение в data frame(cross-section). Работа с данными (dplyr, tidyr).
5. Визуализация данных (ggplot2).
6. Линейная регрессия. Оценки коэффициентов и их интерпретация.
7. Применение линейной регрессии на различных данных. Оценка качества моделей.

Дополнительная часть: если успеваем по времени.

- Логистическая регрессия.
- Графы в R (пакет igraph). Простой парсер для скачивания данных из интернета. Скачиваем данные из VK с помощью VkAPI. Строим граф друзей. Основы кластеризации.
- Что такое машинное обучение? Основные задачи машинного обучения. Метод ближайших соседей (kNN).