

04 марта 2020

Какие алгоритмы и структуры данных нужно освоить начинающему специалисту по Data Science

Data Science – обширная сфера, поэтому начинающим специалистам может быть сложно понять, что им нужно изучать. На какие алгоритмы и структуры данных нужно обратить внимание в первую очередь, рассказывает Александр Зенькович, руководитель проектов департамента аналитических решений ГК «КОРУС Консалтинг».

Начинающему data scientist'у необходимо определиться с ближайшей целью своего развития:

1. Применение своих навыков в прикладных задачах.
2. Экспериментальные и в большей степени научные задачи, участие в конкурсах на платформе Kaggle, анализ идей цифровых гигантов.

В первом случае нужно обладать знаниями моделей регрессии (линейная, градиентный спуск, полиномиальная, логистическая, регуляризованные регрессии, метод опорных векторов, деревья принятия решений), методах понижения размерности, ансамблевых методах (бэггинг, бустинг), моделей классификации и кластеризации.



Для этого стоит изучить библиотеки NumPy, Pandas, SciPy, Sklearn, Matplotlib, XGBoost, LightGBM, после чего специалисту открыты двери для решения задач в ритейл-, телеком- и банковской сферах.

Если в приоритете экспериментальные задачи, помимо всего вышесказанного нужно разобраться в библиотеках NLP, Keras, TensorFlow, Theano, проанализировать множество видео-, аудиоданных, собрать различные датасеты, например, с изображениями котиков. Это откроет потенциальную возможность разработать нечто крутое, что потом купит Apple или Facebook.

Вне зависимости от того, какая цель будет выбрана, начинающий специалист должен владеть основами аналитики, статистики и математики. А в фундаментальных знаниях – Python и R, базовые навыки работы в Linux, Excel и, конечно же, SQL, потому что иногда проще написать две строчки select-а, чем ломать голову, как минимизировать количество циклов над датасетом.

Источник: Tproger

