

Наука о данных (Data Science)

Синонимы: Даталогия, Datology

Наука о данных — междисциплинарная область, которая включает академическую дисциплину и сферу практической деятельности, связанные с изучением, проектированием и реализацией процессов извлечения знаний из массивов данных и их применением для поддержки принятия решений в управленческой и исследовательской деятельности.

Наука о данных включает такие дисциплины, как информатика, математика, логика, статистика, искусственный интеллект и машинное обучение.



Впервые термин «наука о данных» был введен датским пионером в области информатики Питером Науром в 1974 году. Он, определил ее как датологию — науку о процессах обработки цифровых данных и их жизненном цикле. Впоследствии, интегрировав в себя анализ данных, понятие получило более широкую трактовку и трансформировалось в **Data Science**. Хотя сам термин получил широкое распространение только в 1990-х годах, а всеобщее признание — в начале 2000-х.

Анализ данных обычно предполагает работу с относительно небольшими наборами структурированной информации для решения конкретных исследовательских и управленческих задач. В то время как наука ориентирована на решение всего комплекса

проблем, связанных с использованием данных, в том числе и неструктурированных, в управлении и принятии решений, начиная от их сбора и консолидации и заканчивая синтезом решений.

Однако наибольший всплеск интереса к науке о данных связан с появлением концепции Big data, в основу которой и легли новые технологии анализа и обработки огромных объемов информации, имеющих разнообразную природу и структуру, в том числе, за счет применения методов, разработанных в 2000-е годы в рамках Data Science.

Поэтому как в настоящее время, так и в обозримой перспективе наука о данных будет важнейшим направлением информационных технологий, а специалисты в этой области (data scientists) — наиболее востребованными в сфере IT.

Вместе с тем набирает популярность идея, что Data Science должна стать инструментом не только для специалистов в области IT, бизнес-аналитики и искусственного интеллекта, но и для широкого круга работников, у которых нет специального образования и навыков в этих сферах, но при этом необходимо решать проблемы, связанные с применением данных для принятия решений.

Эта идея воплотилась в корпоративную стратегию Гражданский дата-сайнс (Citizen Data Science), в рамках которой к решению проблем управления на основе данных привлекаются не только профильные специалисты и сотрудники IT-отделов, но и все работники организации, деятельность которых может стать более эффективной за счет применения технологий Data Science.