Data Science

REVIEWED BY JAKE FRANKENFIELD | Updated Mar 2, 2018

What is Data Science

Data science is a field of Big Data geared toward providing meaningful information based on large amounts of complex data. Data science, or data-driven science, combines different fields of work in statistics and computation in order to interpret data for the purpose of decision making.

BREAKING DOWN Data Science

Data is drawn from different sectors and platforms including cell phones, social media, e-commerce sites, healthcare surveys, internet searches, etc. The increase in the amount of data available opened the door to a new field of study called Big Data — or the extremely large data sets that can help produce better operational tools in all sectors. The continually increasing sets of and easy access to data are made possible by a collaboration of companies known as fintech, which use technology to innovate and enhance traditional financial products and services. The data produced creates even more data which is easily shared across entities thanks to emergent fintech products like cloud computing and storage. However, the interpretation of vast amounts of unstructured data for effective decision making may prove too complex and time consuming for companies, hence the emergence of data science.

History of Data Science

The term data science has been floating around for the better part of the last 30 years, and was originally used as a substitute for "computer science" in 1960. It wasn't until about 15 years later that the term was used to define the survey of data processing methods that are used in different applications. In 2001, data science was introduced as an independent discipline. The Harvard Business Review published an article in 2012 calling the data scientist the “sexiest job of the 21st century.”

How Data Science Works

Data science incorporates tools from multi disciplines to gather a data set, process and derive insights from the data set, extract meaningful data from the set, and interpret it for decision-making purposes. The disciplinary areas that make up the data science field include mining, statistics, machine learning, analytics, and some programming. Data mining applies algorithms in the complex data set to reveal patterns that are then used to extract usable and relevant data from the set. Statistical measures like predictive analytics utilize this extracted data to gauge events that are likely to happen in the future based on what the data shows happened in the past. Machine learning is an artificial intelligence tool that processes mass quantities of data that a human would be unable to process in a lifetime. Machine learning perfects the decision model presented under predictive analytics by matching the likelihood of an event happening to what actually happened at the predicted time.

Under analytics, the data analyst collects and processes the structured data from the machine learning stage using algorithms. S/he interprets, converts, and summarizes the data to a cohesive language that the decision-making team can understand. These areas mentioned are by no means a complete list of what data science involves. As the role of a data scientist is better understood, more skill sets will be added to the field that encompass sectors like data architecture, data engineering, and data administrator.

Data Scientist Defined

A data scientist collects, analyzes and interprets large volumes of data to help a company improve its operations. These professionals develop statistical models in order to analyze