

4.12: What is data science?

Definition: Data Science

Data science takes structured and unstructured data and uses scientific methods, processes, algorithms, and systems to extract knowledge and insight.

It begins by procuring data from many sources such as web servers, logs, databases, APIs (application program interface), and online repositories. Once the acquisition has happened, the data must be cleaned and pipeline data. This is done by sorting and organizing relevant and usable data; this is the transformation process. Data Modeling is next; the goal is to create the best modeling that suits the company's needs when using the data. This can be done using metrics, algorithms, and analytics. The goal is to progress to AI and deep learning or machine learning. Data science problem solves company issues using data.

- **Structured Data** - Is data that is found in a fixed field within a record or file. It includes data contained in relational databases and spreadsheets. Such as:
 - Date
 - Time
 - Census Data
 - Facebook “Likes”
- **Unstructured Data** - Is information that is not organized and does not have a pre-defined model. Such as:
 - Body of emails
 - Tweets
 - Facebook Status
 - Video Transcripts

4.12.1: What is data analytics?

Data Analytics takes raw data gathered from data mining and analyzes the information to uncover relationships and patterns to find insight into the data when using it. Companies use these analytics to optimize problem-solving and assist in decision-making. The information is helpful to understand who your consumer is as well as marketing your company or product. This is all helpful to create efficiency and streamline operations. Data continuously being collected can then be adjusted as new criteria happen. Today's data analytics are deeper, larger in abundance, and retrieved quicker than yesteryear. The information is more accurate and detailed, which accelerates successful problem-solving.



Figure 4.12.1: Analytic information. Image by xresch from Pixabay is licensed CC BY-SA 2.0

4.12.2: Business Intelligence and Business Analytics

This trend remains a competitive activity for companies. With tools such as data warehousing and data mining at their disposal, businesses learn how to use the information to their advantage. The term business intelligence is used to describe how organizations use to take data they are collecting and analyze it to obtain a competitive advantage. Besides using data from their internal databases, firms often purchase information from data brokers to understand their industries' big-picture understanding. Business analytics is the term used to describe internal company data to improve business processes and practices.

4.12.3: Data-Driving Business Leaders

Data analytics, used correctly, help management make data-driven and timely decisions. Companies like Amazon, Netflix, and Walmart have shown how harnessing rich customer and operational data can create competitive advantages.

The techniques of databases, data warehousing, and data mining are fundamental to building business intelligence. When raw data and information are transformed into valuable assets that provide a deep understanding of customer needs and wants for today and tomorrow, it becomes an essential tool for any company. Integrating business intelligence in day-to-day operations is what sets one company apart from another.

As more data is continually generated and massive volume of data can be analyzed using these techniques, forward-thinking leaders recognize that data mastery opens doors to new ways of understanding customers, streamlining processes, adapting to market changes, and designing winning business models. The story of every major corporate success today inevitably includes how they successfully acquired the right data, organized it in databases built for analytics, and mined it for actionable insights.

This page titled [4.12: What is data science?](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Ly-Huong T. Pham and Tejal Desai-Naik \(Evergreen Valley College\)](#).

- [4.12: Sidebar- What is data science?](#) by Ly-Huong T. Pham, Tejal Desai-Naik, Laurie Hammond, & Wael Abdeljabbar is licensed [CC BY 3.0](#).