

4.8: Data Mining

Businesses go from lacking of data to analyze to gain better insights about their company. Databases solved that problem. Now, business has the challenge of too much data to review and analyze which leads to data overload. This becomes an issue because the user needs to evaluate which information is useful and which is not. Data mining helps solve this issue.

Definition: Data Mining

Data mining is the process of discovering patterns and knowledge from large amounts of data using automated methods. It applies statistics, machine learning, and database techniques to parse data and identify useful information to extract actionable insights efficiently from both structured and unstructured data to inform data-driven decision making.

Data mining is the process of sorting through big data (measured in terabytes). Many businesses do mining to get detailed insight on their customers, products and to optimize business decisions. The analysis is executed with sophisticated programs. The programs can combine multiple databases. The end effect is so complex that companies must find a way to store the data. Data warehouses are needed. The data warehouse is where the information is stored and processed from the data mining. The price for a simple warehouse could start at \$10 million.

Companies like Google, Netflix, Amazon, and Facebook are big users of data mining. They seek to find out who their consumer is and how best to keep them and sell them more products. They also review their products. The means used are reviewing data and finding trends, patterns, and associations to make decisions. Generally, data mining is accomplished through automated means against extensive data sets, such as a data warehouse.

Examples of data mining include:

- An analysis of sales from a large grocery chain might determine that milk is purchased more frequently the day after it rains in cities with a population of less than 50,000.
- A bank may find that loan applicants whose bank accounts show particular deposit and withdrawal patterns are not good credit risks.
- A baseball team may find those collegiate baseball players with specific statistics in hitting, pitching, and fielding for more successful major league players.

In some cases, a data-mining project is begun with a hypothetical result in mind. For example, a grocery chain may already have some idea that the buying patterns change after it rains and want to get a deeper understanding of exactly what is happening. In other cases, there are no presuppositions, and a data-mining program is run against large data sets to find patterns and associations.

This page titled [4.8: Data Mining](#) 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](#)) .

- [4.8: Data Mining](#) by Ly-Huong T. Pham, Tejal Desai-Naik, Laurie Hammond, & Wael Abdeljabbar is licensed [CC BY 3.0](#).