

33.9: Data Mining and Warehousing

Learning Objectives

- Explain the difference between data mining and data warehousing

Billions and billions of bits of data flood into an organization's information system, but how does that data get utilized effectively? The challenge lies not so much with the collection or storage of the data: today, it is possible to collect and even store vast amounts of information relatively cheaply. The main difficulty is figuring out the best and most efficient way to extract and manage the relevant data. In this section you will learn how organizations not only warehouse but then mine the data they collect.

Did you ever think about how much data you yourself generate? Just remember what you went through to start college. First, you had to fill out application forms asking you about test scores, high school grades, extracurricular activities, and finances, plus demographic data about you and your family. Once you'd picked a college, you had to supply data on your housing preferences, the curriculum you wanted to follow, and the party who'd be responsible for paying your tuition. When you registered for classes, you gave more data to the registrar's office. When you arrived on campus, you gave out still more data to have your ID picture taken, to get your computer and phone hooked up, to open a bookstore account, and to buy an on-campus food-charge card. Once you started classes, data generation continued on a daily basis: your food card and bookstore account, for example, tracked your various purchases, and your ID tracked your coming and going all over campus. And you generated grades.

And all these data apply to just one aspect of your life. You also generated data every time you used your credit card and your cell phone. Who uses all these data? How are they collected, stored, analyzed, and distributed in organizations that have various reasons for keeping track of you?

Warehousing and Mining Data

How do businesses organize all of this data so that they can transform it into useful information? For most businesses this is where **data warehousing** comes into play. A data warehouse collects data from multiple sources (both internal and external) and stores the data to later be used in an analysis. The primary purpose of a data warehouse is to store the data in a way that it can later be retrieved for use by the business. Despite the name, **Data Mining** is not the process of getting specific pieces of data out of the data warehouse, but rather the goal of data mining is the identification of patterns and knowledge from large amounts of data. Large retailers such as WalMart and Target track sales on a minute-by-minute basis and data mining allows these large retailers to recognize changes in purchasing behavior in an extremely short amount of time. They can quickly make adjustments to inventory levels based on the information gathered from thousands of individual transactions as a result of data mining. Clearly understanding consumer behavior is a primary goal of data mining. The following video explains just how businesses use data mining to understand and predict consumer behavior.



You can [view the transcript for "DATA MINING | The Checkout | ABC1"](#) (opens in new window).

? Practice Question

<https://assessments.lumenlearning.co...essments/11098>

Today businesses are treating the Internet as a massive data warehouse and are using data mining techniques to gather data about not just existing customers, but potential customers. Data mining tools such as Scrapy, Nutch and Splash allow businesses to learn more about customers, competitors, compare prices and even find new customers and sales targets. As the quantity of data businesses can collect continues to grow, having an effective data warehousing system that can be easily mined has become increasingly critical to business success.

Contributors and Attributions

CC licensed content, Original

- Data Mining and Warehousing. **Authored by:** Linda Williams. **Provided by:** Lumen Learning. **License:** [CC BY: Attribution](#)

All rights reserved content

- Data Mining. **Authored by:** The Checkout. **Located at:** <https://youtu.be/f2Kji24833Y>. **License:** *All Rights Reserved*. **License Terms:** Standard YouTube license

33.9: Data Mining and Warehousing is shared under a [not declared](#) license and was authored, remixed, and/or curated by LibreTexts.