# Introduction to Business Intelligence:

Business Intelligence (BI) refers to the process of gathering, storing, analyzing, and visualizing data to help organizations make informed decisions. It involves a combination of technologies, applications, and practices aimed at transforming raw data into meaningful insights for strategic planning, operational improvements, and competitive advantage.

The key components of business intelligence include data collection from various sources, data storage in data warehouses or data lakes, data processing using analytical tools and algorithms, and data visualization through reports, dashboards, and interactive tools.

Business intelligence enables organizations to understand their internal operations, customer behaviors, market trends, and other critical factors affecting their business performance. By harnessing BI, companies can identify opportunities for growth, optimize processes, mitigate risks, and gain a competitive edge in their respective industries.

## What is business intelligence architecture (BI architecture)?

A business intelligence architecture is a framework for the various technologies an organization deploys to run business intelligence and analytics applications. It includes the IT systems and software tools that are used to collect, integrate, store and analyze BI data and then present information on business operations and trends to corporate executives and other business users.

The underlying BI architecture is a key element in the execution of a successful business intelligence program that uses data analysis and reporting to help an organization track business performance, optimize business processes, identify new revenue opportunities, improve strategic planning and make more informed business decisions.

## Benefits of BI architecture

In the absence of a BI architecture, businesses and enterprises are at risk of making costly errors while striving to optimize their data utilization.

A well-articulated BI framework can offer organizations the following key benefits:

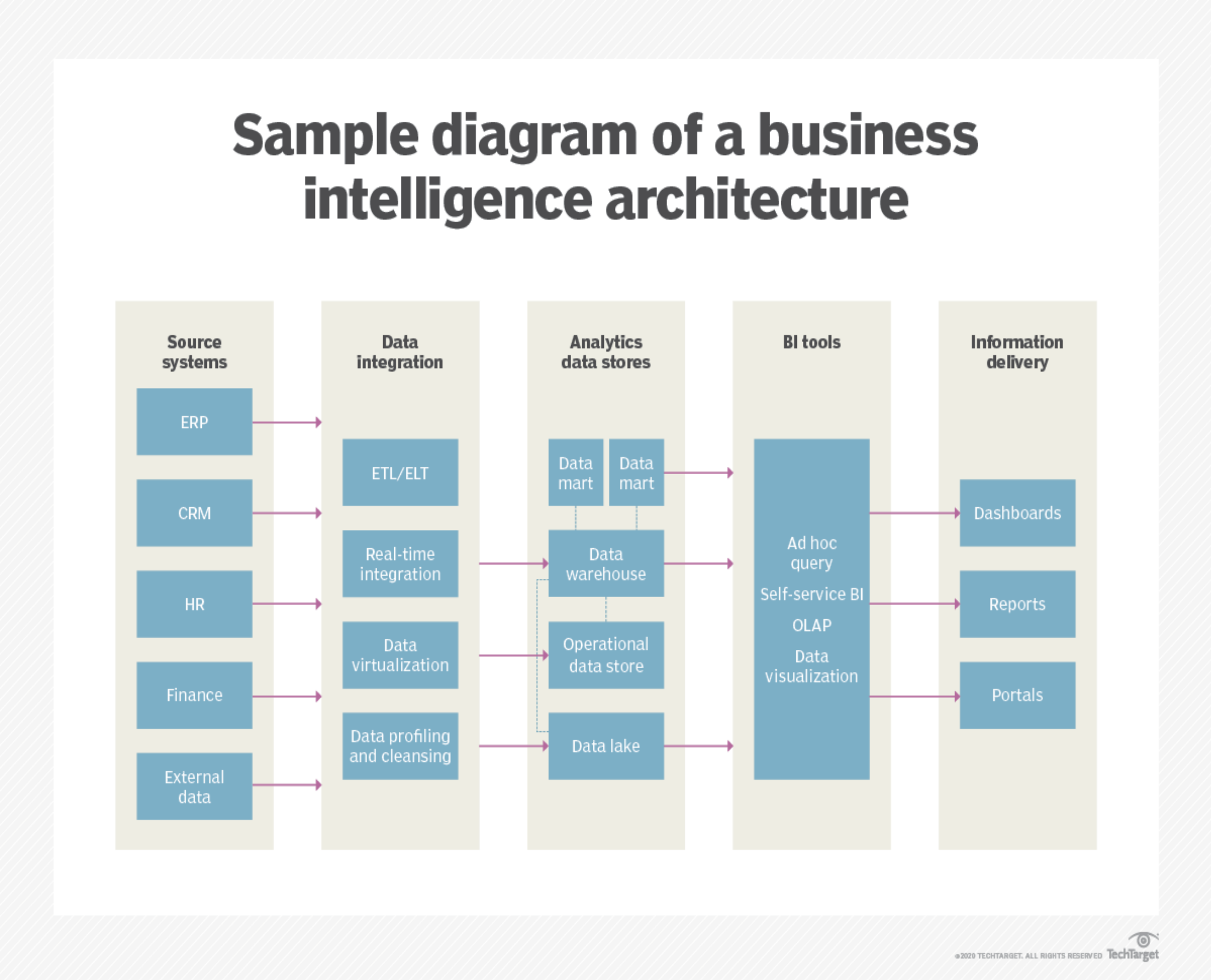
- **Technology benchmarks.** A BI architecture articulates the technology standards and data management and business analytics practices that support an organization's BI efforts, as well as the specific platforms and tools deployed.

- **Improved decision-making.** Enterprises benefit from an effective BI architecture by using the insights generated by business intelligence tools to make data-driven decisions that help increase revenue and profits.
- **Technology blueprint.** A BI framework serves as a technology blueprint for collecting, organizing and managing BI data and then making the data available for analysis, data visualization and reporting. A strong BI architecture automates reporting and incorporates policies to govern the use of the technology components.
- **Enhanced coordination.** Putting such a framework in place enables a BI team to work in a coordinated and disciplined way to build an enterprise BI program that meets the organization's data analytics needs. The BI architecture also helps BI and data managers create an efficient process for handling and managing the business data that's pulled into the environment.
- **Time savings.** By automating the process of collecting and analyzing data, BI helps organizations save time on manual and repetitive tasks, freeing up their teams to focus on more high-value projects.
- **Scalability.** An effective BI infrastructure is easily scalable, enabling businesses to change and expand as necessary.
- **Improved customer service.** Business intelligence enhances customer understanding and service delivery by helping track customer satisfaction and facilitate timely improvements. For example, an e-commerce store can use BI to track order delivery times and optimize shipping for better customer satisfaction.

To ensure their needs are met, C-level executives, business managers and other users who rely on data analysis to formulate strategies and guide their decision-making should have a stake in creating the architecture.

### Business intelligence architecture components and diagram

A BI architecture can be deployed in an on-premises data center or in the cloud. In either case, it contains a set of core components that collectively support the different stages of the BI process from data collection, integration, data storage and analysis to data visualization, information delivery and the use of BI data in business decision-making.




The core components of a BI architecture include the following:

**Source systems.** These are all of the systems that capture and hold the transactional and operational data identified as essential for the enterprise BI program. For example, this can include enterprise resource planning, customer relationship management, flat files, application programming interfaces, finance, manufacturing and supply chain management systems as well as secondary sources, such as market data and customer databases from outside information providers. As a result, both internal and external data sources are often incorporated into a BI architecture. Customer relationship management (CRM) is a technology for managing all your company's relationships and interactions with customers and potential customers. ERP stands for enterprise resource planning. It's a software system that includes all the tools and processes required to run a successful company, including HR, manufacturing, supply chain, finance, accounting, and more.

Important criteria in the data source selection process include data relevancy, data currency, data quality and the level of detail in the available data sets. In addition, a combination of structured, semi-structured and unstructured data types might be required to meet the data analysis and decision-making needs of executives and other end users.

**Data integration and cleansing tools.** To effectively analyze the collected data for a BI program, an organization must integrate and capture and streaming integration to support real-time analytics applications, and data virtualization, which combines data from different source systems virtually. consolidate different data sets to create unified views of them. The most widely used data integration technology for BI applications is Extract, Transform and Load (ETL) software, which pulls data from source systems in batch processes. A variant of ETL is extract, load and transform, a technology in which data is extracted and loaded as-is and transformed later for specific BI uses. Other methods include real-time data integration, such as change data

A BI architecture typically also includes data profiling and data cleansing tools that are used to identify and fix data quality issues. They help BI and data management teams provide clean, consistent data that's suitable for BI uses.

**Analytics data stores.** This encompasses the various repositories where BI data is stored and managed. The primary repository is a data warehouse, which usually stores structured data in a relational, columnar or multidimensional database and makes it available for querying and analysis. An enterprise data warehouse can also be tied to smaller data marts set up for individual departments and business units with data that's specific to their BI needs.

In addition, BI architectures often include an Operational Data Store (ODS) that's an interim repository for data before it goes into a data warehouse. An ODS can also be used to run analytical queries against recent transaction data. Depending on the size of a BI environment, a data warehouse, data mart and an ODS can be deployed on a single database server or separate business intelligence systems.

A data lake running on a Hadoop cluster or other big data platform can also be incorporated into a BI architecture as a repository for raw data of various types. The data can be analyzed in the data lake itself or filtered and loaded into a data warehouse for analysis. A wellplanned architecture should specify which of the different data stores is best suited for particular BI uses.

**BI and data visualization tools.** The tools used to analyze data and present information to business users include a suite of technologies that can be built into a BI architecture -- for example, ad hoc query, data mining and online analytical processing software. In addition, the growing adoption of self-service BI tools enables business analysts and managers to run queries themselves instead of relying on the members of the BI team to do that for them.

BI software also includes data visualization tools that can be used to create graphical representations of data in the form of charts, graphs and other types of visualizations designed to illustrate trends, patterns and outlier elements in data sets.

**Dashboards, portals and reports.** These information delivery tools give users visibility into the results of BI and analytics applications with built-in data visualizations and, often, self-service capabilities to do additional data analysis. For example, BI dashboards and online portals can be designed to provide real-time data access with configurable views and give users the ability to drill down into data. Reports tend to present data in a more static format.

Other components that increasingly are part of a business architecture include data preparation software used to structure and organize data for analysis and a metadata repository, a business glossary and a data catalog, which can help users find relevant data and understand its lineage and meaning.

## BI architecture tools

BI architecture tools facilitate the centralization of data collection as well as data analysis and visualization. These tools play an integral role in empowering businesses to make informed decisions and extract insights from extensive data sets.

Some examples of BI tools on the market include the following:

1. **Datapine.** Datapine lets users access, view, analyze and share their company data on a single analytics platform. Users can perform data analysis, create interactive business dashboards and obtain new business insights through a simple drag-and-drop interface.

2. **Domo.** The Domo cloud-based platform unifies data, systems and people for seamless business operations. It provides enterprise tools for data aggregation, analytics, dashboards and reporting for organizations looking to maximize data value.
3. **Dundas BI.** This enterprise-level BI tool lets users create and customize interactive dashboards and reports. The software can either act as a central data hub or integrate into existing websites for customized BI capabilities.
4. **GoodData.** As part of the GoodData platform, this tool offers an enterprise-level option for data analytics and business intelligence. It helps users analyze data coming from multiple sources and create reports.
5. **Infor Birst.** Infor Birst is a cloud-based platform that uses a networked approach and modern enterprise-class architecture with a focus on multi-tenancy. Birst ensures that a company's data remains connected by centralizing both decentralized and centralized data.
6. **Microsoft Power BI.** Users can run analytics either in the cloud or in a reporting server. The tool comes with built-in artificial intelligence features and offers end-to-end encryption features.
7. **Oracle Business Intelligence.** This integrated set of tools lets users gather, store, analyze and report data for smart decision-making. In addition, it includes a scalable BI server, dashboards, a content library, web-based reporting and analytics tools.
8. **SAS Business Intelligence.** This collection of tools lets corporate users conduct self-service analytics. Its two components -- Enterprise Business Intelligence and Business Visualization -- provide interactive visualizations and analytics to aid with data analysis and decision-making.
9. **Tableau.** In addition to data visualization features, this tool offers live visual analytics and supports most databases and numerous data sources.
10. **Zoho Analytics.** This self-service BI and data analytics software lets users analyze data, generate data visualizations and uncover insights quickly and easily. This tool is accessible to both small and large-sized organizations.

Business Intelligence: What and Hows of PAST and Present Data?

Analytics: WHYs of Present and Whats and Hows of Future Data?

- **Designing a business intelligence (BI) application** : It involves several key steps to ensure it effectively meets the needs of users and provides actionable insights. Here's a high-level outline of the process:

1. Define Objectives and Requirements: Identify the business goals and objectives the BI application aims to support. Gather requirements from stakeholders across different departments to understand their needs and pain points.Define key performance indicators (KPIs) and metrics that will be tracked to measure success.
2. Data Collection and Integration: Identify all relevant data sources within the organization, including databases, spreadsheets, cloud applications, etc. Implement processes for extracting, transforming, and loading (ETL) data into a centralized data warehouse or data lake. Ensure data quality by cleansing and standardizing data to maintain accuracy and consistency.
3. Data Modeling and Architecture: Design a data model that reflects the relationships between different data entities and supports the required analytical queries. Choose an appropriate data architecture (e.g., star schema, snowflake schema) based on the nature of the data and analytical needs. Optimize the data architecture for performance and scalability.
4. Visualization and Reporting: Select visualization tools and techniques that best represent the data and insights for end users. Design intuitive dashboards and reports that provide a clear overview of KPIs and trends. Ensure interactivity and drill-down capabilities for users to explore data at different levels of detail.
5. Advanced Analytics and Predictive Modeling: Implement advanced analytics techniques such as predictive modeling, machine learning, and data mining to uncover hidden patterns and trends. Develop predictive models to forecast future outcomes and support decision-making. Integrate these advanced analytics capabilities into the BI application to provide additional value to users.
6. Security and Governance: Implement robust security measures to protect sensitive data and ensure compliance with regulatory requirements. Define user roles and permissions to control access to data and features based on users' roles and responsibilities. Establish data governance policies and procedures to maintain data integrity and enforce data quality standards.

7. Deployment and Maintenance: Deploy the BI application in a production environment, ensuring scalability and reliability. Provide training and support for end users to maximize adoption and usage of the BI application. Establish a process for ongoing maintenance and updates to keep the BI application aligned with evolving business needs.
8. Continuous Improvement: Monitor usage metrics and user feedback to identify areas for improvement. Iterate on the BI application based on user insights and changing business requirements. Stay informed about emerging technologies and best practices in BI to keep the application competitive and relevant.

By following these steps, you can design a business intelligence application that empowers users with actionable insights and drives informed decision-making within the organization.

- **Requirements Gathering:**

1. Stakeholder Interviews: Conduct interviews with key stakeholders from various departments (e.g., finance, sales, marketing) to understand their specific reporting and analytics needs. Identify the types of data they require, the frequency of reporting, and the level of detail needed for analysis.
2. User Personas and Use Cases: Develop user personas representing different roles within the organization (e.g., executives, analysts, operational staff). Define use cases that illustrate how each persona will interact with the BI application to achieve their goals.
3. Data Requirements: Determine the types of data needed for analysis, including structured and unstructured data from internal and external sources. Identify any data transformations or aggregations required to meet reporting requirements.
4. Performance and Scalability Requirements: Determine the expected volume of data and concurrent users to establish performance benchmarks. Define scalability requirements to accommodate future growth in data volume and user base.
5. Integration Requirements: Identify existing systems and applications that will need to integrate with the BI application. Determine the data integration methods and protocols required to extract data from these systems.

- ❖ **Designing a Business Intelligence (BI) solution involves several key components, including designing dimensional models and physical databases.**

**Designing Dimensional Models:**

1. Identify Business Processes and Dimensions: Understand the business processes that the BI solution will support. Identify key business dimensions (e.g., time, product, customer) that will be used for analysis.
2. Define Dimensional Hierarchies: Define hierarchical relationships within each dimension (e.g., year > quarter > month > day for time dimension). Establish attribute hierarchies to organize dimension members at different levels of granularity.
3. Choose Fact Tables and Measures: Identify fact tables that represent the primary business processes and contain quantitative measures (e.g., sales revenue, quantity sold). Determine the granularity of each fact table, representing the level of detail at which data is captured.
4. Design Dimensional Schemas: Choose a dimensional modeling schema (e.g., star schema, snowflake schema) based on the complexity and relationships between dimensions and facts. Design star schema for simpler structures or snowflake schema for more normalized dimensions.
5. Handle Slowly Changing Dimensions (SCDs): Identify dimensions that have attributes that change over time. Implement appropriate slowly changing dimension techniques (e.g., Type 1, Type 2, Type 3) to manage historical changes in dimension data.

**Designing the Physical Databases:**

1. Select Database Platform: Choose a database platform that supports the requirements of the BI solution (e.g., relational database management system (RDBMS), columnar databases, NoSQL databases). Consider factors such as scalability, performance, and compatibility with existing infrastructure.
2. Design Tables and Indexes: Translate the dimensional model into physical tables and define appropriate indexes to optimize query performance. Normalize or denormalize tables based on the access patterns and reporting requirements.
3. Partitioning and Data Distribution: Implement partitioning strategies to manage large datasets efficiently (e.g., partitioning by date range). Distribute data across multiple physical nodes or servers to improve scalability and parallel processing.

4. Optimize Storage and Compression: Implement compression techniques to reduce storage requirements and improve query performance. Choose appropriate data storage formats (e.g., row-based, columnar) based on access patterns and query requirements.
5. Backup and Recovery: Establish backup and recovery procedures to ensure data integrity and availability. Define backup schedules and retention policies to meet recovery point objectives (RPOs) and recovery time objectives (RTOs).
6. Security and Access Control: Implement security measures such as encryption, access controls, and authentication mechanisms to protect sensitive data. Define user roles and permissions to restrict access to data based on users' roles and responsibilities.
7. Data Integration and ETL Processes: Design ETL (Extract, Transform, Load) processes to populate the dimensional model from source systems. Implement error handling and data validation routines to ensure data quality and integrity.

By designing dimensional models and physical databases effectively, you can create a robust foundation for your BI solution that supports efficient data analysis and reporting, while ensuring scalability, performance, and data integrity.

❖ **Predictive Analytics and Data Mining Concepts:**

1. **Definitions and Characteristics:**
    - **Predictive analytics** involves using data, statistical algorithms, and machine learning techniques to identify the likelihood of future outcomes based on historical data.
    - **Data mining** refers to the process of discovering patterns, trends, and insights from large datasets.
    - Characteristics include pattern recognition, classification, clustering, regression, and anomaly detection.
2. **Benefits:**
    - Predictive analytics and data mining enable organizations to make data-driven decisions, forecast future trends, optimize processes, improve customer satisfaction, and mitigate risks.

- They provide actionable insights for strategic planning, marketing campaigns, fraud detection, and personalized recommendations.

❖ **What Is Data Mining? How It Works, Benefits, Techniques, and Examples**

Data mining is the process of searching and analyzing a large batch of raw data in order to identify patterns and extract useful information.

Companies use data mining software to learn more about their customers. It can help them to develop more effective marketing strategies, increase sales, and decrease costs. Data mining relies on effective data collection, warehousing, and computer processing.

**Key Features:**

- Data mining is the process of analyzing a large batch of information to discern trends and patterns.
- Data mining can be used by corporations for everything from learning about what customers are interested in or want to buy to fraud detection and spam filtering.
- Data mining programs break down patterns and connections in data based on what information users request or provide.
- Social media companies use data mining techniques to commodify their users in order to generate profit.
- This use of data mining has come under criticism lately as users are often unaware of the data mining happening with their personal information, especially when it is used to influence preferences.

**How Data Mining Works**

Data mining involves exploring and analyzing large blocks of information to glean meaningful patterns and trends. It is used in credit risk management, fraud detection, and spam filtering. It also is a market research tool that helps reveal the sentiment or opinions of a given group of people.

The data mining process breaks down into four steps:

- Data is collected and loaded into data warehouses on-site or on a cloud service.
- Business analysts, management teams, and information technology professionals access the data and determine how they want to organize it.
- Custom application software sorts and organizes the data.
- The end user presents the data in an easy-to-share format, such as a graph or table.

# Data Warehousing and Mining Software

Data mining programs analyze relationships and patterns in data based on user requests. It organizes information into classes.

For example, a restaurant may want to use data mining to determine which specials it should offer and on what days. The data can be organized into classes based on when customers visit and what they order.

In other cases, data miners find clusters of information based on logical relationships or look at associations and sequential patterns to draw conclusions about trends in consumer behavior.

Warehousing is an important aspect of data mining. Warehousing is the centralization of an organization's data into one database or program. It allows the organization to spin off segments of data for specific users to analyze and use depending on their needs.

# Data Mining Techniques

Data mining uses algorithms and various other techniques to convert large collections of data into useful output. The most popular types of data mining techniques include:

- **Association rules**, also referred to as market basket analysis, search for relationships between variables. This relationship in itself creates additional value within the data set as it strives to link pieces of data. For example, association rules would search a company's sales history to see which products are most commonly purchased together; with this information, stores can plan, promote, and forecast.
- **Classification** uses predefined classes to assign to objects. These classes describe the characteristics of items or represent what the data points have in common with each. This data mining technique allows the underlying data to be more neatly categorized and summarized across similar features or product lines.
- **Clustering** is similar to classification. However, clustering identifies similarities between objects, then groups those items based on what makes them different from other items. While classification may result in groups such as "shampoo," "conditioner," "soap," and "toothpaste," clustering may identify groups such as "hair care" and "dental health."
- **Decision trees** are used to classify or predict an outcome based on a set list of criteria or decisions. A decision tree is used to ask for the input of a series of cascading

questions that sort the dataset based on the responses given. Sometimes depicted as a tree-like visual, a decision tree allows for specific direction and user input when drilling deeper into the data.

- **K-Nearest neighbor (KNN)** is an algorithm that classifies data based on its proximity to other data. The basis for KNN is rooted in the assumption that data points that are close to each other are more similar to each other than other bits of data. This non-parametric, supervised technique is used to predict the features of a group based on individual data points.
- **Neural networks** process data through the use of nodes. These nodes are comprised of inputs, weights, and an output. Data is mapped through supervised learning, similar to the ways in which the human brain is interconnected. This model can be programmed to give threshold values to determine a model's accuracy.
- **Predictive analysis** strives to leverage historical information to build graphical or mathematical models to forecast future outcomes. Overlapping with regression analysis, this technique aims at supporting an unknown figure in the future based on current data on hand.

# The Data Mining Process

To be most effective, data analysts generally follow a certain flow of tasks along the data mining process. Without this structure, an analyst may encounter an issue in the middle of their analysis that could have easily been prevented had they prepared for it earlier. The data mining process is usually broken into the following steps.

**Step 1: Understand the Business**

Before any data is touched, extracted, cleaned, or analyzed, it is important to understand the underlying entity and the project at hand. What are the goals the company is trying to achieve by mining data? What is their current business situation? What are the findings of a SWOT analysis? Before looking at any data, the mining process starts by understanding what will define success at the end of the process.

**Step 2: Understand the Data**

Once the business problem has been clearly defined, it's time to start thinking about data. This includes what sources are available, how they will be secured and stored, how the information will be gathered, and what the final outcome or analysis may look like. This step also includes determining the limits of the data, storage, security, and collection and assesses how these constraints will affect the data mining process.

**Step 3: Prepare the Data**

Data is gathered, uploaded, extracted, or calculated. It is then cleaned, standardized, scrubbed for outliers, assessed for mistakes, and checked for reasonableness. During this stage of data mining, the data may also be checked for size as an oversized collection of information may unnecessarily slow computations and analysis.

**Step 4: Build the Model**
With our clean data set in hand, it's time to crunch the numbers. Data scientists use the types of data mining above to search for relationships, trends, associations, or sequential patterns. The data may also be fed into predictive models to assess how previous bits of information may translate into future outcomes.

**Step 5: Evaluate the Results**
The data-centered aspect of data mining concludes by assessing the findings of the data model or models. The outcomes from the analysis may be aggregated, interpreted, and presented to decision-makers that have largely been excluded from the data mining process to this point. In this step, organizations can choose to make decisions based on the findings.

**Step 6: Implement Change and Monitor**
The data mining process concludes with management taking steps in response to the findings of the analysis. The company may decide the information was not strong enough or the findings were not relevant, or the company may strategically pivot based on findings. In either case, management reviews the ultimate impacts of the business and recreates future data mining loops by identifying new business problems or opportunities.

Different data mining processing models will have different steps, though the general process is usually pretty similar. For example, the Knowledge Discovery Databases model has nine steps, the CRISP-DM model has six steps, and the SEMMA process model has five steps.

## Applications of Data Mining

In today's age of information, almost any department, industry, sector, or company can make use of data mining.

**Sales**
Data mining encourages smarter, more efficient use of capital to drive revenue growth. Consider the point-of-sale register at your favorite local coffee shop. For every sale, that coffeehouse collects the time a purchase was made and what products were sold. Using this information, the shop can strategically craft its product line.

**Marketing**
Once the coffeehouse above knows its ideal line-up, it's time to implement the changes. However, to make its marketing efforts more effective, the store can use data mining to understand where its clients see ads, what demographics to target, where to place digital ads, and what marketing strategies most resonate with customers. This includes aligning marketing campaigns, promotional offers, cross-sell offers, and programs to the findings of data mining.

**Manufacturing**

For companies that produce their own goods, data mining plays an integral part in analyzing how much each raw material costs, what materials are being used most efficiently, how time is spent along the manufacturing process, and what bottlenecks negatively impact the process. Data mining helps ensure the flow of goods is uninterrupted.

**Fraud Detection**

The heart of data mining is finding patterns, trends, and correlations that link data points together. Therefore, a company can use data mining to identify outliers or correlations that should not exist. For example, a company may analyze its cash flow and find a reoccurring transaction to an unknown account. If this is unexpected, the company may wish to investigate whether funds are being mismanaged.

**Human Resources**

Human resources departments often have a wide range of data available for processing including data on retention, promotions, salary ranges, company benefits, use of those benefits, and employee satisfaction surveys. Data mining can correlate this data to get a better understanding of why employees leave and what entices new hires.

**Customer Service**

Customer satisfaction may be caused (or destroyed) for a variety of reasons. Imagine a company that ships goods. A customer may be dissatisfied with shipping times, shipping quality, or communications. The same customer may be frustrated with long telephone wait times or slow e-mail responses. Data mining gathers operational information about customer interactions and summarizes the findings to pinpoint weak points and highlight what the company is doing right.

## Advantages and Disadvantages of Data Mining

Pros of Data Mining

- It drives profitability and efficiency
- It can be applied to any type of data and business problem
- It can reveal hidden information and trends

Cons of Data Mining

- Complexity
- Results and benefits are not guaranteed
- It can be expensive

**Pros Explained**

- Data mining ensures a company is collecting and analyzing reliable data. It is often a more rigid, structured process that formally identifies a problem, gathers data related to the problem, and strives to formulate a solution. Therefore, data mining helps a business become more profitable, more efficient, or operationally stronger.

- Data mining can look very different across applications, but the overall process can be used with almost any new or legacy application. Essentially any type of data can be gathered and analyzed, and almost every business problem that relies on qualifiable evidence can be tackled using data mining.
- The end goal of data mining is to take raw bits of information and determine if there is cohesion or correlation among the data. This benefit of data mining allows a company to create value with the information they have on hand that would otherwise not be overly apparent. Though data models can be complex, they can also yield fascinating results, unearth hidden trends, and suggest unique strategies.

**Cons Explained**

- This complexity of data mining is one of its greatest disadvantages. Data analytics often requires technical skill sets and certain software tools. Smaller companies may find this to be a barrier of entry too difficult to overcome.
- Data mining doesn't always guarantee results. A company may perform statistical analysis, make conclusions based on strong data, implement changes, and not reap any benefits. Through inaccurate findings, market changes, model errors, or inappropriate data populations, data mining can only guide decisions and not ensure outcomes.
- There is also a cost component to data mining. Data tools may require costly subscriptions, and some bits of data may be expensive to obtain. Security and privacy concerns can be pacified, though additional IT infrastructure may be costly as well. Data mining may also be most effective when using huge data sets; however, these data sets must be stored and require heavy computational power to analyze.

## Data Mining and Social Media

One of the most lucrative applications of data mining has been undertaken by social media companies. Platforms like Facebook, TikTok, Instagram, and X platform (formerly Twitter) gather reams of data about their users, based on their online activities.

That data can be used to make inferences about their preferences. Advertisers can target their messages to the people who appear to be most likely to respond positively.

Data mining on social media has become a big point of contention, with several investigative reports and exposes showing just how intrusive mining users' data can be. At the heart of the issue, users may agree to the terms and conditions of the sites not realizing how their personal information is being collected or to whom their information is being sold.

## Examples of Data Mining

Data mining can be used for good, or it can be used illicitly. Here is an example of both.

**eBay and e-Commerce**

eBay collects countless bits of information every day from sellers and buyers. The company uses data mining to attribute relationships between products, assess desired price ranges, analyze prior purchase patterns, and form product categories.

eBay outlines the recommendation process as:

1. Raw item metadata and user historical data are aggregated.
2. Scrips are run on a trained model to generate and predict the item and user.
3. A KNN search is performed.
4. The results are written to a database.
5. The real-time recommendation takes the user ID, calls the database results, and displays them to the user.3

   eBay. "Building a Deep Learning Based Retrieval System for Personalized Recommendations."

**Facebook-Cambridge Analytica Scandal**

Another cautionary example of data mining is the Facebook-Cambridge Analytica data scandal. During the 2010s, the British consulting firm Cambridge Analytica Ltd. collected personal data from millions of Facebook users. This information was later analyzed for use in the 2016 presidential campaigns of Ted Cruz and Donald Trump. It is suspected that Cambridge Analytica interfered with other notable events such as the Brexit referendum.

In light of this inappropriate data mining and misuse of user data, Facebook agreed to pay $100 million for misleading investors about its uses of consumer data. The Securities and Exchange Commission claimed Facebook discovered the misuse in 2015 but did not correct its disclosures for more than two years.5

## **What Are the Types of Data Mining**?

There are two main types of data mining: **predictive data mining and descriptive data mining**. Predictive data mining extracts data that may be helpful in determining an outcome. Description data mining informs users of a given outcome.

## How Is Data Mining Done?

Data mining relies on big data and advanced computing processes including machine learning and other forms of artificial intelligence (AI). The goal is to find patterns that can lead to inferences or predictions from large and unstructured data sets.

## What Is Another Term for Data Mining?

Data mining also goes by the less-used term "knowledge discovery in data," or KDD.

## Where Is Data Mining Used?

Data mining applications have been designed to take on just about any endeavor that relies on big data. Companies in the financial sector look for patterns in the markets. Governments try to identify potential security threats. Corporations, especially online and social media companies, use data mining to create profitable advertising and marketing campaigns that target specific sets of users.

## The Bottom Line

Modern businesses have the ability to gather information on their customers, products, manufacturing lines, employees, and storefronts. These random pieces of information may not tell a story, but the use of data mining techniques, applications, and tools helps piece together information.

The ultimate goal of the data mining process is to compile data, analyze the results, and execute operational strategies based on data mining results.

**Data Mining and Privacy Issues:**

- **Privacy Concerns:**
    - Data mining techniques may uncover sensitive information about individuals.
    - Privacy-preserving methods such as anonymization, encryption, and differential privacy can help mitigate privacy risks.

**Techniques for Predictive Modeling:**

1. **Artificial Neural Networks (ANN):**
    - Mimic the structure and function of the human brain to learn complex patterns from data.
    - Consist of interconnected nodes (neurons) organized in layers (input, hidden, output).
2. **Support Vector Machines (SVM):**
    - Supervised learning algorithm used for classification and regression tasks.
    - Finds the optimal hyperplane that separates different classes or predicts a continuous value.

These techniques, along with regression, classification, association rules, and clustering, form the foundation of predictive analytics and data mining. They enable organizations to extract valuable insights from data to drive business decisions and gain a competitive advantage.