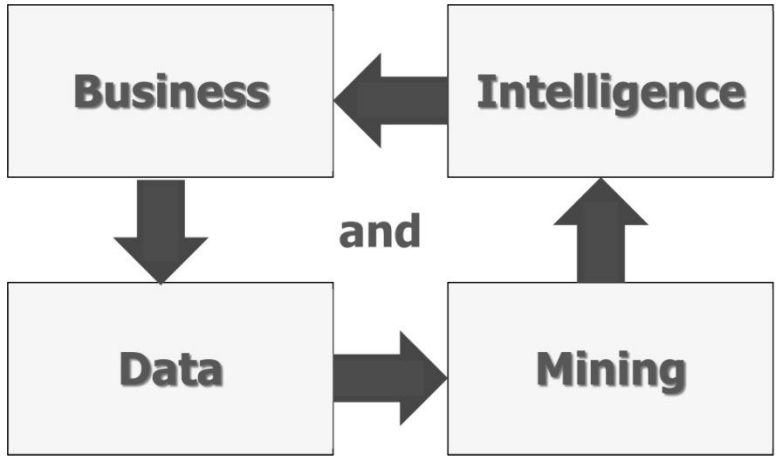
Business Intelligence
Concepts and Applications

Business intelligence (BI) is an umbrella term that includes a variety of IT applications that are used to analyze an organization's data and commu-nicate the information to relevant users. Its major components are data warehousing, data mining, querying, and reporting

The nature of life and businesses is to grow. Information is the life-blood of business. Businesses use many techniques for understanding their environment and predicting the future for their own benefit and growth. Decisions are made from facts and feelings. Data-based decisions are more effective than those based on feelings alone. Actions based on accurate data, information, knowledge, experimentation, and testing, using fresh insights, can more likely succeed and lead to sustained growth.



Business intelligence and data mining cycle

One's own data can be the most effective teacher. Therefore, organizations should gather data, sift through it, analyze and mine it, find insights, and then embed those insights into their operating procedures.

There is a new sense of importance and urgency around data as it is being viewed as a new natural resource. It can be mined for value, insights, and competitive advantage. In a hyperconnected world, where everything is potentially connected to everything else, with potentially infinite correlations, data represents the impulses of nature in the form of certain events and attributes. A skilled business person is motivated to use this cache of data to harness nature, and to find new niches of unserved opportunities that could become profitable ventures.

BI Applications

BI tools are required in almost all industries and functions. The nature of the information and the speed of action may be different across busi-nesses, but every manager today needs access to BI tools to have up-to-date metrics about business performance. Businesses need to embed new insights into their operating processes to ensure that their activities continue to evolve with more efficient practices. The following are some areas of applications of BI and data mining.

Customer Relationship Management

A business exists to serve a customer. A happy customer becomes a repeat customer. A business should understand the needs and sentiments of the customer, sell more of its offerings to the existing customers, and also, expand the

pool of customers it serves. BI applications can impact many aspects of marketing.

1. *Maximize the return on marketing campaigns:* Understanding the customer's pain points from data-based analysis can ensure that the marketing messages are fine-tuned to better resonate with customers.
2. *Improve customer retention (churn analysis):* It is more difficult and expensive to win new customers than it is to retain existing custom-ers. Scoring each customer on their likelihood to quit can help the business design effective interventions, such as discounts or free ser-vices, to retain profitable customers in a cost-effective manner.
3. *Maximize customer value (cross-selling, upselling):* Every contact with the customer should be seen as an opportunity to gauge their cur-rent needs. Offering a customer new products and solutions based on those imputed needs can help increase revenue per customer. Even a customer complaint can be seen as an opportunity to wow the cus-tomer. Using the knowledge of the customer's history and value, the business can choose to sell a premium service to the customer.
4. *Identify and delight highly valued customers:* By segmenting the cus-tomers, the best customers can be identified. They can be proactively contacted, and delighted, with greater attention and better service. Loyalty programs can be managed more effectively.
5. *Manage brand image:* A business can create a listening post to listen to social media chatter about itself. It can then do sentiment analysis of the text to understand the

nature of comments and respond ap-propriately to the prospects and customers.

Health Care and Wellness

Health care is one of the biggest sectors in advanced economies. Evi-dence-based medicine is the newest trend in data-based health care man-agement. BI applications can help apply the most effective diagnoses and prescriptions for various ailments. They can also help manage public health issues, and reduce waste and fraud.

1. *Diagnose disease in patients:* Diagnosing the cause of a medical condi-tion is the critical first step in a medical engagement. Accurately diag-nosing cases of cancer or diabetes can be a matter of life and death for the patient. In addition to the patient's own current situation, many

other factors can be considered, including the patient's health history, medication history, family's history, and other environmental factors. This makes diagnosis as much of an art form as it is science. Systems, such as IBM Watson, absorb all the medical research to date and make probabilistic diagnoses in the form of a decision tree, along with a full explanation for their recommendations. These systems take away most of the guess work done by doctors in diagnosing ailments.

2. *Treatment effectiveness:* The prescription of medication and treatment is also a difficult choice out of so many possibilities. For example, there are more than 100 medications for hypertension (high blood pressure) alone. There are also interactions in terms of which drugs work well with others and which drugs do not. Decision trees can help doctors learn about and prescribe more effective treatments. Thus, the patients could recover their health faster with a lower risk of complications and cost.
3. *Wellness management:* This includes keeping track of patient health records, analyzing customer health trends, and proactively advising them to take any needed precautions.
4. *Manage fraud and abuse:* Some medical practitioners have unfortu-nately been found to conduct unnecessary tests and/or overbill the gov-ernment and health insurance companies. Exception-reporting systems can identify such providers, and action can be taken against them.

5. *Public health management:* The management of public health is one of the important responsibilities of any government. By using effec-tive forecasting tools and techniques, governments can better predict the onset of disease in certain areas in real time. They can thus be better prepared to fight the diseases. Google has been known to pre-dict the movement of certain diseases by tracking the search terms (like flu, vaccine) used in different parts of the world.

Education

As higher education becomes more expensive and competitive, it is a great user of data-based decision-making. There is a strong need for efficiency, increasing revenue, and improving the quality of student experience at all levels of education.

1. *Student enrolment (recruitment and retention):* Marketing to new po-tential students requires schools to develop profiles of the students that are most likely to attend. Schools can develop models of what kinds of students are attracted to the school, and then reach out to those students. The students at risk of not returning can be flagged, and corrective measures can be taken in time.
2. *Course offerings:* Schools can use the class enrolment data to develop models of which new courses are likely to be more popular with students. This can help increase class size, reduce costs, and improve student satisfaction.
3. *Alumni pledges:* Schools can develop predictive models of which alumni are most likely to pledge financial support to the school. Schools can create a profile for alumni more likely to pledge dona-tions to the school. This could lead to a reduction in the cost of mailings and other forms of outreach to alumni.

Retail

Retail organizations grow by meeting customer needs with quality prod-ucts, in a convenient, timely, and cost-effective manner. Understanding emerging customer shopping patterns can help retailers organize their products, inventory, store layout, and web presence in order to delight their customers, which in turn would help increase revenue and profits. Retailers generate a lot of transaction and logistics data that can be used to solve problems.

1. *Optimize inventory levels at different locations:* Retailers need to man-age their inventories carefully. Carrying too much inventory imposes carrying costs, while carrying too little inventory can cause stock-outs and lost sales opportunities. Predicting sales trends dynamically can help retailers move inventory to where it is most in demand. Retail organizations can provide their suppliers with real-time in-formation about sales of their items so that the suppliers can deliver their product to the right locations and minimize stock-outs.
2. *Improve store layout and sales promotions:* A market basket analy-sis can develop predictive models of which products sell together

often. This knowledge of affinities between products can help re-tailers co-locate those products. Alternatively, those affinity prod-ucts could be located farther apart to make the customer walk the length and breadth of the store, and thus be exposed to other products. Promotional discounted product bundles can be created to push a nonselling item along with a set of products that sell well together.

3. *Optimize logistics for seasonal effects:* Seasonal products offer tremen-dously profitable short-term sales opportunities, yet they also offer the risk of unsold inventories at the end of the season. Understand-ing which products are in season in which market can help retailers dynamically manage prices to ensure their inventory is sold during the season. If it is raining in a certain area, then the inventory of umbrella and ponchos could be rapidly moved there from nonrainy areas to help increase sales.
4. *Minimize losses due to limited shelf life:* Perishable goods offer chal-lenges in terms of disposing off the inventory in time. By tracking sales trends, the perishable products at risk of not selling before the sell-by date can be suitably discounted and promoted.

Banking

Banks make loans and offer credit cards to millions of customers. They are most interested in improving the quality of loans and reducing bad debts. They also want to retain more good customers and sell more services to them.

1. *Automate the loan application process:* Decision models can be gen-erated from past data that predict the likelihood of a loan proving successful. These can be inserted in business processes to automate the financial loan application process.
2. *Detect fraudulent transactions:* Billions of financial transactions hap-pen around the world every day. Exception-seeking models can iden-tify patterns of fraudulent transactions. For example, if money is being transferred to an unrelated account for the first time, it could be a fraudulent transaction.

3. *Maximize customer value (cross-selling, upselling):* Selling more prod-ucts and services to existing customers is often the easiest way to increase revenue. A checking account customer in good standing could be offered home, auto, or educational loans on more favorable terms than other customers, and thus, the value generated from that customer could be increased.
4. *Optimize cash reserves with forecasting:* Banks have to maintain cer-tain liquidity to meet the needs of depositors who may like to with-draw money. Using past data and trend analysis, banks can forecast how much to keep, and invest the rest to earn interest.

Financial Services

Stock brokerages are an intensive user of BI systems. Fortunes can be made or lost based on access to accurate and timely information.

1. *Predict changes in bond and stock prices:* Forecasting the price of stocks and bonds is a favorite pastime of financial experts as well as lay people. Stock transaction data from the past, along with other variables, can be used to predict future price patterns. This can help traders develop long-term trading strategies.
2. *Assess the effect of events on market movements:* Decision models using decision trees can be created to assess the impact of events on changes in market volume and prices. Monetary policy changes (such as Fed Reserve interest rate change) or geopolitical changes (such as war in a part of the world) can be factored into the predictive model to help take action with greater confidence and less risk.

3. *Identify and prevent fraudulent activities in trading:* There have un-fortunately been many cases of insider trading, leading to many prominent financial industry stalwarts going to jail. Fraud detection models can identify and flag fraudulent activity patterns.

Insurance

This industry is a prolific user of prediction models in pricing insurance proposals and managing losses from claims against insured assets.

1. *Forecast claim costs for better business planning:* When natural disas-ters, such as hurricanes and earthquakes, strike, loss of life and prop-erty occurs. By using the best available data to model the likelihood (or risk) of such events happening, the insurer can plan for losses and manage resources and profits effectively.
2. *Determine optimal rate plans:* Pricing an insurance rate plan requires covering the potential losses and making a profit. Insurers use actu-ary tables to project life spans and disease tables to project mortality rates, and thus price themselves competitively yet profitably.
3. *Optimize marketing to specific customers:* By microsegmenting potential customers, a data-savvy insurer can cherry-pick the best customers and leave the less profitable customers to its competitors. Progressive Insur-ance is a U.S.-based company that is known to actively use data mining to cherry-pick customers and increase its profitability.

4. *Identify and prevent fraudulent claim activities:* Patterns can be iden-tified as to where and what kinds of fraud are more likely to occur. Decision-tree-based models can be used to identify and flag fraudu-lent claims.

Manufacturing

Manufacturing operations are complex systems with interrelated subsys-tems. From machines working right, to workers having the right skills, to the right components arriving with the right quality at the right time, to money to source the components, many things have to go right. Toyota's famous lean manufacturing company works on just-in-time inventory systems to optimize investments in inventory and to improve flexibility in their product mix.

1. *Discover novel patterns to improve product quality:* Quality of a prod-uct can also be tracked, and this data can be used to create a predic-tive model of product quality deteriorating. Many companies, such as automobile companies, have to recall their products if they have found defects that have a public safety implication. Data mining can help with root cause analysis that can be used to identify sources of errors and help improve product quality in the future.

2. *Predict/prevent machinery failures:* Statistically, all equipment is likely to break down at some point in time. Predicting which machine is likely to shut down is a complex process. Decision models to forecast machinery failures could be constructed using past data. Preventive maintenance can be planned, and manufacturing capacity can be adjusted, to account for such maintenance activities.

Telecom

BI in telecom can help with churn management, marketing/customer profiling, network failure, and fraud detection.

1. *Churn management:* Telecom customers have shown a tendency to switch their providers in search for better deals. Telecom companies tend to respond with many incentives and discounts to hold on to customers. However, they need to determine which customers are at a real risk of switching and which others are just negotiating for a better deal. The level of risk should to be factored into the kind of deals and discounts that should be given. Millions of such customer calls happen every month. The telecom companies need to provide a consistent and data-based way to predict the risk of the customer switching, and then make an operational decision in real time while the customer call is taking place. A decision-tree- or a neural network-based system can be used to guide the customer-service call operator

to make the right decisions for the company, in a consistent manner.

2. *Marketing and product creation:* In addition to customer data, tele-com companies also store call detail records (CDRs), which pre-cisely describe the calling behavior of each customer. This unique data can be used to profile customers and then can be used for creating new products/services bundles for marketing purposes. An American telecom company, MCI, created a program called Friends & Family that allowed calls with one's friends and family on that network to be totally free and thus, effectively locked many people into their network.

Data Warehousing

A data warehouse (DW) is an organized collection of integrated, subject-oriented databases designed to support decision support functions. DW is organized at the right level of granularity to provide clean enterprise-wide data in a standardized format for reports, queries, and analysis. DW is physically and functionally separate from an operational and transac-tional database. Creating a DW for analysis and queries represents signifi-cant investment in time and effort. It has to be constantly kept up-to-date for it to be useful. DW offers many business and technical benefits.

DW supports business reporting and data mining activities. It can facilitate distributed access to up-to-date business knowledge for depart-ments and functions, thus improving business efficiency and customer service. DW

can present a competitive advantage by facilitating decision making and helping reform business processes.

DW enables a consolidated view of corporate data, all cleaned and or-ganized. Thus, the entire organization can see an integrated view of itself. DW thus provides better and timely information. It simplifies data access and allows end users to perform extensive analysis. It enhances overall IT performance by not burdening the operational databases used by Enter-prise Resource Planning (ERP) and other systems.

Design Considerations for DW

The objective of DW is to provide business knowledge to support deci-sion making. For DW to serve its objective, it should be aligned around those decisions. It should be comprehensive, easy to access, and up-to-date. Here are some requirements for a good DW:

1. *Subject-oriented:* To be effective, DW should be designed around a subject domain, that is, to help solve a certain category of problems.
2. *Integrated:* DW should include data from many functions that can shed light on a particular subject area. Thus, the organization can benefit from a comprehensive view of the subject area.
3. *Time-variant (time series):* The data in DW should grow at daily or other chosen intervals. That allows latest comparisons over time.
4. *Nonvolatile:* DW should be persistent, that is, it should not be cre-ated on the fly from the operations databases.

Thus, DW is consis-tently available for analysis, across the organization and over time.

5. *Summarized:* DW contains rolled-up data at the right level for queries and analysis. The rolling up helps create consistent granularity for effec-tive comparisons. It helps reduces the number of variables or dimensions of the data to make them more meaningful for the decision makers.
6. *Not normalized:* DW often uses a star schema, which is a rectangular central table, surrounded by some lookup tables. The single-table view significantly enhances speed of queries.
7. *Metadata:* Many of the variables in the database are computed from other variables in the operational database. For example, total daily sales may be a computed field. The method of its calculation for each variable should be effectively documented. Every element in DW should be sufficiently well-defined.
8. *Near real-time and/or right-time (active):* DWs should be updated in near real-time in many high-transaction volume industries, such as air-lines. The cost of implementing and updating DW in real time could discourage others. Another downside of real-time DW is the possibili-ties of inconsistencies in reports drawn just a few minutes apart.

DW Development Approaches

There are two fundamentally different approaches to developing DW: top down and bottom up. The top-down approach is to make a comprehensive DW that covers all the

reporting needs of the enterprise. The bottom-up approach is to produce small data marts, for the reporting needs of different departments or functions, as needed. The smaller data marts will eventually align to deliver comprehensive EDW capabilities. The top-down approach provides consistency but takes time and resources. The bottom-up approach leads to healthy local ownership and maintainability of data.

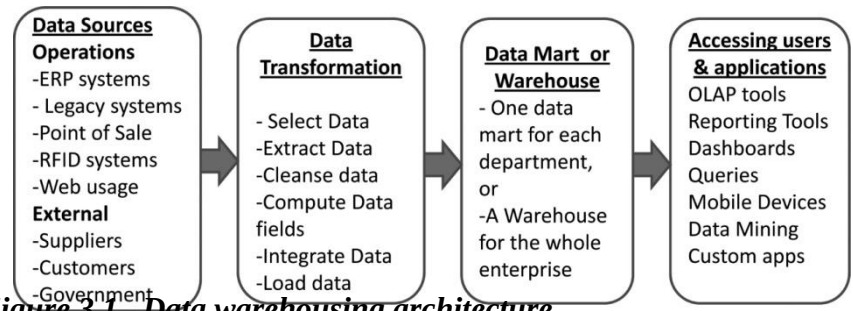


Figure 3.1 Data warehousing architecture

DW Architecture

DW has four key elements (Figure 3.1). The first element is the data sources that provide the raw data. The second element is the process of transforming that data to meet the decision needs. The third element is the methods of regularly and accurately loading of that data into EDW or data marts. The fourth element is the data access and analysis part, where devices and applications use the data from DW to deliver insights and other benefits to users.

Data Sources

DWs are created from structured data sources. Unstructured data, such as text data, would need to be structured before inserted into DW.

1. Operations data include data from all business applications, includ-ing from ERPs systems that form the backbone of an organization's IT systems. The data to be extracted will depend upon the subject matter of DW. For example, for a sales/marketing DW, only the data about customers, orders, customer service, and so on would be extracted.
2. Other applications, such as point-of-sale (POS) terminals and -e-commerce applications, provide customer-facing data. Supplier data could come from supply chain management systems. Planning and budget data should also be added as needed for making com-parisons against targets.
3. External syndicated data, such as weather or economic activity data, could also be added to DW, as needed, to provide good contextual information to decision makers.

Data Transformation Processes

The heart of a useful DW is the processes to populate the DW with good quality data. This is called the extract-transform-load (ETL) cycle.

1. Data should be extracted from many operational (transactional) da-tabase sources on a regular basis.
2. Extracted data should be aligned together by key fields. It should be cleansed of any irregularities or missing values. It should be rolled up together to the same level of granularity. Desired fields, such as daily sales totals, should be computed. The entire data should then be brought to the same format as the central table of DW.
3. The transformed data should then be uploaded into DW.

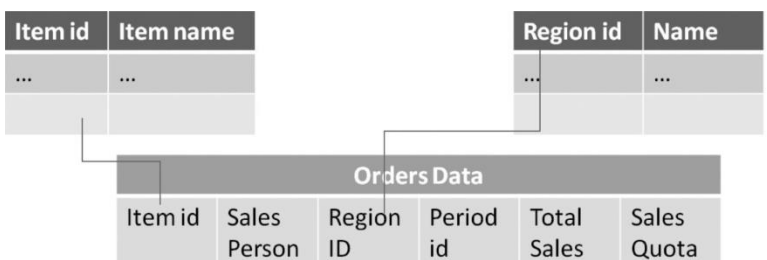
This ETL process should be run at a regular frequency. Daily trans-action data can be extracted from ERPs, transformed, and uploaded to the database the same night. Thus, DW is up-to-date next morning. If DW is needed for near-real-time information access, then the ETL pro-cesses would need to be executed more frequently. ETL work is usually automated using programing scripts that are written, tested, and then deployed for periodic updating DW.

DW Design

Star schema is the preferred data architecture for most DWs. There is a central fact table that provides most of the information of interest. There are lookup tables that provide detailed values for codes used in the central table. For example, the central table may use digits to represent a sales person. The lookup table will help provide the name for that sales person code. Here is an example of a star schema for a data mart for monitoring sales performance (Figure 3.2).

Other schemas include the snowflake architecture. The difference be-tween a star and snowflake is that in the latter, the lookup tables can have their own further lookup tables.

There are many technology choices for developing DW. This includes selecting the right database management system and the right set of data management tools. There are a few big and reliable providers of DW sys-tems. The provider of the operational DBMS may be chosen for DW also.



Alternatively, a best-of-breed DW vendor could be used. There are also a variety of tools out there for data migration, data upload, data retrieval, and data analysis.

DW Access

Data from DW could be accessed for many purposes, through many devices.

1. A primary use of DW is to produce routine management and moni-toring reports. For example, a sales performance report would show sales by many dimensions, and compared with plan. A dashboard-ing system will use data from the warehouse and present analysis to users. The data from DW can be used to populate customized performance dashboards for executives. The dashboard could in-clude drill-down capabilities to analyze the performance data for root cause analysis.
2. The data from the warehouse could be used for ad hoc queries and any other applications that make use of the internal data.

3. Data from DW is used to provide data for mining purposes. Parts of the data would be extracted, and then combined with other relevant data, for data mining.

Data Mining

Gathering and Selecting Data

The total amount of data in the world is doubling every 18 months. There is an ever-growing avalanche of data coming with higher velocity, volume, and variety. One has to quickly use it or lose it. Smart data mining re-quires choosing where to play. One has to make judicious decisions about what to gather and what to ignore, based on the purpose of the data min-ing exercises. It is like deciding where to fish; not all streams of data will be equally rich in potential insights.

To learn from data, one needs to effectively gather quality data, clean and organize it, and then efficiently process it. One requires the skills and technologies for consolidation and integration of data elements from many sources. Most organizations develop an enterprise data model (EDM), which is a unified, high-level model of all the data stored in an organiza-tion's databases. The EDM will be inclusive of the data generated from all internal systems. The EDM provides the basic menu of data to create a data warehouse for a particular decision-making purpose. Data warehouses help organize all this data in a useful manner so that it can be selected and deployed for mining. The EDM can also help imagine what relevant exter-nal data should be gathered to develop good predictive relationships with the internal data. In the United States, the governments and their agencies make a vast variety and quantity of data available at data.gov.

Gathering and curating data takes time and effort, particularly when it is unstructured or semistructured. Unstructured data can come in many forms like databases, blogs, images, videos, and chats. There are streams of unstructured social media data from blogs, chats, and tweets. There are also streams of machine-generated data from connected machines, RFID tags, the internet of things, and so on. The data should be put in rectangular data shapes with clear columns and rows before submitting it to data mining.

Knowledge of the business domain helps select the right streams of data for pursuing new insights. Data that suits the nature of the problem being solved should be gathered. The data elements should be relevant, and suitably address the problem being solved. They could directly im-pact the problem, or they could be a suitable proxy for the effect being measured. Select data will also be gathered from the data warehouse.

Industries and functions will have their own requirements and con-straints. The health care industry will provide a different type of data with different data names. The HR function would provide different kinds of data. There would be different issues of quality and privacy for these data.

Data Cleansing and Preparation

The quality of data is critical to the success and value of the data mining project. Otherwise, the situation will be of the kind of garbage in and garbage out (GIGO). Duplicate data needs to be removed. The same data may be received from multiple sources. When merging the data sets, data must be de-duped.

1. Missing values need to be filled in, or those rows should be removed from analysis. Missing values can be filled in with average or modal or default values.
2. Data elements may need to be transformed from one unit to an-other. For example, total costs of health care and the total number of patients may need to be reduced to cost/patient to allow comparabil-ity of that value.
3. Continuous values may need to be binned into a few buckets to help with some analyses. For example, work experience could be binned as low, medium, and high.
4. Data elements may need to be adjusted to make them comparable over time. For example, currency values may need to be adjusted

for inflation; they would need to be converted to the same base year for comparability. They may need to be converted to a common currency.

6. Outlier data elements need to be removed after careful review, to avoid the skewing of results. For example, one big donor could skew the analysis of alumni donors in an educational setting.
7. Any biases in the selection of data should be corrected to ensure the data is representative of the phenomena under analysis. If the data in-cludes many more members of one gender than is typical of the popu-lation of interest, then adjustments need to be applied to the data.
8. Data should be brought to the same granularity to ensure compara-bility. Sales data may be available daily, but the sales person compen-sation data may only be available monthly. To relate these variables, the data must be brought to the lowest common denominator, in this case, monthly.
9. Data may need to be selected to increase information density. Some data may not show much variability, because it was not properly recorded or for any other reasons. This data may dull the effects of other differences in the data and should be removed to improve the information density of the data.

Outputs of Data Mining

Data mining techniques can serve different types of objectives. The out-puts of data mining will reflect the objective being served. There are many representations of the outputs of data mining.

One popular form of data mining output is a decision tree. It is a hierarchically branched structure that helps visually

follow the steps to make a model-based decision. The tree may have certain attributes, such as probabilities assigned to each branch. A related format is a set of busi-ness rules, which are if-then statements that show causality. A decision tree can be mapped to business rules. If the objective function is predic-tion, then a decision tree or business rules are the most appropriate mode of representing the output.

The output can be in the form of a regression equation or mathe-matical function that represents the best fitting curve to represent the data. This equation may include linear and nonlinear terms. Regression

Evaluating Data Mining Results

There are two primary kinds of data mining processes: supervised learn-ing and unsupervised learning. In supervised learning, a decision model can be created using past data, and the model can then be used to predict the correct answer for future data instances. Classification is the main category of supervised learning activity. There are many techniques for classification, decision trees being the most popular one. Each of these techniques can be implemented with many algorithms. A common met-ric for all of classification techniques is predictive accuracy.

Predictive Accuracy 5 (Correct Predictions) / Total Predictions

Suppose a data mining project has been initiated to develop a predictive model for cancer patients using a decision tree. Using a relevant set of variables and data instances, a decision tree model has been created. The model is then used to predict other data instances. When a true positive data point is positive, that is a correct prediction, called a true positive (TP). Similarly, when a true negative data point is classified as negative, that is a true negative (TN). On the other hand, when a true-positive data

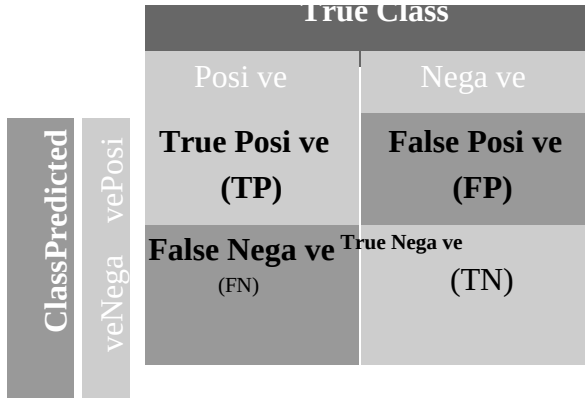


Figure Confusion matrix

point is classified by the model as negative, that is an incorrect prediction, called a false negative (FN). Similarly, when a true-negative data point is classified as positive, that is classified as a false positive (FP). This is called the confusion matrix (Figure 4.1).

Thus, the predictive accuracy can be specified by the following formula.

Predictive Accuracy = (TP + TN) / (TP + TN + FP + FN).

All classification techniques have a predictive accuracy associated with a predictive model. The highest value can be 100 percent. In practice, predictive models with more than 70 percent accuracy can be considered usable in business domains, depending upon the nature of the business.

There are no good objective measures to judge the accuracy of unsupervised learning techniques, such as cluster analysis. There is no single right answer for the results of these techniques. The value of the segmentation model depends upon the value the decision maker sees in those results.

Data Mining Techniques

Data may be mined to help make more efficient decisions in the future. Or it may be used to explore the data to find interesting associative pat-terns. The right technique depends upon the kind of problem being solved

Important Data Mining Techniques		
Supervised Learning: Classification	Machine Learning Techniques	Decision Trees
		Artificial Neural Networks
	Statistical Techniques	Regression
Unsupervised Learning: Exploration	Machine Learning Techniques	Cluster Analysis
		Association Rule Mining

Figure : Important data mining techniques

The most important class of problems solved using data mining are classification problems. These are problems where data from past deci-sions is mined to extract the few rules and patterns that would improve the accuracy of the decision-making process in the future. The data of past decisions is organized and mined for decision rules or equations, which are then codified to produce more accurate decisions. Classifica-tion techniques are called supervised learning as there is a way to super-vise whether the model's prediction is right or wrong.

A decision tree is a hierarchically organized branched, structured to help make decision in an easy and logical manner. *Decision trees* are the most popular data mining technique, for many reasons.

1. Decision trees are easy to understand and easy to use, by analysts as well as executives. They also show a high predictive accuracy.
2. They select the most relevant variables automatically out of all the available variables for decision-making.
3. Decision trees are tolerant of data quality issues and do not require much data preparation from the users.
4. Even nonlinear relationships can be handled well by decision trees.

Data Visualization

As data and insights grow in number, a new requirement is the ability of the executives and decision makers to absorb this information in real time. There is a limit to human comprehension and visualization capac-ity. That is a good reason to prioritize and manage with fewer but key variables that relate directly to the key result areas of a role.

Here are few considerations when presenting data:

1. Present the conclusions and not just report the data.
2. Choose wisely from a palette of graphs to suit the data.
3. Organize the results to make the central point stand out.
4. Ensure that the visuals accurately reflect the numbers. Inappropriate visuals can create misinterpretations and misunderstandings.
5. Make the presentation unique, imaginative, and memorable.