



## Data Mining

- let the data tell its story

'If thou art able, O stranger, to find out all these things and gather them together in your mind, giving all the relations, thou shalt depart crowned with glory and knowing that thou hast been adjudged perfect in this species of wisdom.'

- In Ivor Thomas "Greek Mathematics" in J. R. Newman (ed.) *The World of Mathematics*

**Data mining is the art and science of extracting hidden patterns from the accumulated data for decision-making.**

It has emerged as a valuable decision support tool with the recognition that:

1. Data Mining and advanced statistical techniques provide insights into data that mere **slicing and dicing** does not.
2. The human mind's ability to handle complexity is limited.
3. The advances in computing have made the cost of storing and processing data very affordable.

**T**he three essential requisites of good data mining initiatives are:

1. Domain expertise in the area of business
2. Extensive knowledge of data mining tools, advanced statistics and **modeling expertise**
3. A data mining vision that includes willingness to commit time and other resources.

As time progresses, the returns on investment in data mining grow **exponentially**. There are **two mistakes** one can make along the road to data mining: not going all the way, and not starting. Data mining has to be driven

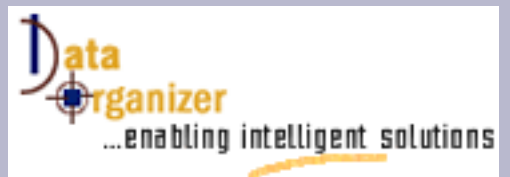
## Modeling Lab

Dedicated to data collection, analysis and modeling, the Modeling Lab



at DecisionCraft is headed by [Dr. Jitendra C Parikh](#). Dr. Parikh, a theoretical physicist (PhD-Physics, University of Chicago) with 40 years' experience of research and teaching, has expertise in modeling **complex real systems**. He is supported by a team of MBAs, engineers, and statisticians.

The team at DecisionCraft has expertise in state-of-the-art modeling techniques such as artificial neural networks, stochastic processes, chaos theory, statistical methods, simulation, data mining algorithms, etc.



**dataOrganizer™**

The first step in data mining is

iteratively to generate a **virtuous cycle of returns**. Insights gathered initially are used to collect more relevant data, leading to more actionable insights.

Data mining projects require substantial initial effort in data preparation. Research indicates that 75% of total project time goes in data preparation.

Many companies embark on data warehousing initiatives without first developing a data mining vision. *Developing a **data warehousing solution** in the context and background of a data mining vision increases the value of the data warehousing initiative manifolds.*

Data mining is used mostly in applications where a **large amount of data** is generated/available. The typical sectors are Finance, Insurance, Banking, Retail, Telecommunications, Airlines, Public Utilities, etc.

## Data Mining Techniques:

### 1. Artificial Neural Networks:

Non-linear predictive models that learn through training, and resemble biological neural networks in structure.

### 2. Decision Trees:

Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID)

### 3. Genetic Algorithms:

Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.

### 4. Nearest Neighbor Method:

A technique that classifies each record in a dataset based on a combination of the classes of k record(s) most similar to it in a historical data set.

### 5. Rule Induction:

The extraction of useful if-then rules from data based on statistical significance.

Next issue: [Simulation](#)

Previous issue: [Irrationality & Decision Sciences](#)

getting together clean usable data onto one database.

**dataOrganizer™** is a web-enabled application capable of browsing, cleaning and integrating data from diverse sources onto one destination.

## qcCharts™

qcCharts™ features interactive data visualization with a range of charts capable of exploring patterns in data. In combination with dataOrganizer™, it can provide enterprise-wide visibility to data, charts and analysis.

## More Resources on Data Mining

[-Data Mining](#)

[-Overview of Data Mining](#)

[-Executive's Guide to Data Mining](#)

[-Analysis of Data Mining](#)

[Algorithms](#)

[-Glossary of Data Mining Terms](#)

## Some Data Mining Tools

SPSS Clementine

SAS Enterprise Miner

SYSTAT

StatSoft Statistica

RuleQuest C 5.0

Salford Systems CART

Megaputer PolyAnalyst

## DecisionCraft Products

qcCharts™

dataOrganizer™

Supply Chain Simulator

Travel Route Optimizer  
Logistics Planner