Title:
Data Mining And Modelling

Word Count:
366

Summary:
The important processes that have to be clearly delineated for Data Mining,
Analysis and Modelling are:

Data model: what data will be available and how will it flow?
Data gathering: how will data be gathered both in physical and technological
terms?
Data gathered: what data will be gathered?
Data types: what types of data will be gathered?
Data formatting: how will data be held?
Data warehousing: where will data be held?
Data mining: how will we retrieve data from th...

Keywords:
data, mining, analysis, modelling, marketing, business, management

Article Body:
The important processes that have to be clearly delineated for Data Mining,
Analysis and Modelling are:

Data model: what data will be available and how will it flow?
Data gathering: how will data be gathered both in physical and technological
terms?
Data gathered: what data will be gathered?
Data types: what types of data will be gathered?
Data formatting: how will data be held?
Data warehousing: where will data be held?
Data mining: how will we retrieve data from the warehouse?
Information modelling: how will we create models and what of?
Information access: how will we access the data models and reports?
Presentation & reporting: on what will we report?

Most companies want to know essential information about customers at every point
of contact, for example:

Lifetime value

X sell and upgrade potential
Acquisition cost
Channel preferences
Loyalty/retention
Purchase behaviour patterns

Much of the data that they have will have different frequencies of change, refreshment or occurrence. It will be kept for different periods. In some cases, aggregated data may be kept rather than source data. All of these factors effect the data modelling exercise and the eventual modelling software requirements.

Turning the data into useful information requires:

Identifying the issue(s)
Assembling the data set(s)
Building models
Verify models
Interpretation of the results
Automation of the delivery

Thereafter, modelling tools and techniques have to be used. These can be divided into two groups: theory driven and data driven.

Theory driven modelling (hypothesis testing) attempts to substantiate or disprove preconceived ideas. Theory driven modelling tools require the user to specify most of the model based on prior knowledge and then tests to see if the model is valid.

Data driven modelling tools automatically create the model based on patterns they find in the data. This also needs to be tested before it can be accepted as valid.

Modelling is an iterative process with the final model usually being a combination of prior knowledge and newly discovered information. The engine(s) tools and techniques include:

Statistical techniques
Data driven tools
Correlation
Cluster analysis
t-tests
Factor analysis
Analysis of Variance

CHAID (Chi-square Automatic Interaction Detector) decision trees
Linear regression
Visualisation tools
Logistic regression
Neural networks
Discriminant analysis