

Understanding Customer Behaviour through the Extraction, Storage, Modeling and Reporting of Customer Data

Andy Twaits, Amadeus Software Limited



ABSTRACT

This paper will examine how an integrated approach to data management – the extraction, storage, analysis, reporting and modeling of customer data – can add significant value to an organisation throughout the customer lifecycle. This includes delivery of significant business benefits in the areas of prospect targeting and marketing, customer acquisition and take-on, customer management (cross-sell, retention etc.), and collections and recoveries. The subject is approached from a business rather than a technical point of view.

INTRODUCTION

This paper focuses on the use of statistical models within consumer facing businesses and is particularly relevant to the financial services, telecommunications, utilities, mail order and insurance sectors. It provides an overview of the basic methodology used in building models. It then looks at the various sources of data that are used, how that data can be accessed and the significant benefits to be gained from holding consolidated data for modeling and reporting in a structured data mart environment.

STATISTICAL MODELING

Statistical models, often referred to as scorecards or propensity models, aim to predict a particular aspect of the future behaviour of individual customers. For example, an application scorecard will predict the likelihood of a new applicant defaulting on their loan, or an attrition model will predict the likelihood of an existing customer switching to an alternative supplier. This then enables organisations to take actions based on that likely future behaviour, either to mitigate negative behaviours or capitalize on positive behaviours. There are numerous different statistical techniques that can be used to build models, the most popular being linear regression. The aim here is not to examine or explain these differing techniques¹ but to consider the common underlying methodologies and therefore the requirements in terms of data needed to build models and monitor their performance.

Underlying Methodology

The basic methodology for the large majority of all scorecards and propensity models is the same, and based on one single assumption – the future will be like the past. That is to say, if a group of customers with particular characteristics behaved in a particular way in the past, then customers with those same characteristics will behave in the same way in the future. To illustrate this, the process for building an Application Scorecard to assess risk of default at application is outlined below.

Application Scorecard

The purpose of an Application Scorecard is to determine the likelihood or probability of a customer failing to pay for the service for which they are applying, for example a credit card, personal loan or electricity supply. The results of the scorecard can then be used to make decisions about the applicant, such as to accept or decline the loan application, to set the credit limit on a credit card or determine if a security deposit is required before supplying electricity.

The process to develop an Application Scorecard is illustrated in Figure 1. A sample is taken of customers that applied over a given period in the past (usually 18 to 24 months). A large amount of data relating to the characteristics of the customer at the point of application is required within the sample (e.g. age, income, previous credit history etc). The performance of these accounts between the date of application and the present is then examined to determine the outcome status. Simplistically each account is defined as being “good” or “bad” during the outcome period. The definition of good or bad will vary, but for example might be “ever more than 3 months in arrears” for bad accounts and “never more than 1 month in arrears” for good accounts.

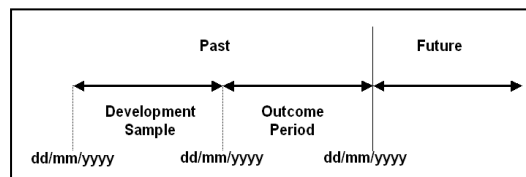


Figure 1. Application Scorecard development

The next stage uses statistical techniques such as linear regression to determine which application characteristics were associated with customers that were subsequently defined as either good or bad, and the strength of that association. Once this is complete, a weighting or score is assigned to each of the characteristics that are found to be predictive. See Figure 2.

Characteristic	Value	Score
Age	18 – 25	22
	26 – 40	35
	41+	48
Time @ Bank	0 – 2 years	20
	3 – 7 years	28
	8+ years	40
Residential Status	Owner	38
	Occupier	27
	Other	22

Figure 2. Application Score Characteristics

The final application score, which is used for the decision-making, is created by summing the individual scores from each characteristic. As shown in Figure 3, there is a direct relationship between the application score and the good/bad odds i.e. the higher the score, the greater the probability the subsequent performance of the account will be “good” rather than “bad”.

Application Score	Good/Bad Odds	Decision
<120	2 / 1	Decline
121 – 140	3 / 1	Decline
141 – 160	4 / 1	Decline
161 – 180	7 / 1	Refer
181 – 200	8 / 1	Refer
201 – 220	10 / 1	Refer
221 – 240	14 / 1	Accept
241 – 260	18 / 1	Accept
261+	25 / 1	Accept

Figure 3. Simple Application Scorecard

The example here shows the score being used to determine whether the application is accepted, declined or referred for further investigation. The score could as easily be used to determine the credit limit on a credit card or the rate of interest to be charged on a personal loan.

Other Statistical Modeling Applications

The same basic techniques as used in Application Scoring can be applied in to build models for use throughout the life of a customer relationship. These models can be used for the targeting of prospective new customers, assessing the risk or value of new applicants, managing existing customers through retention or cross-sell activities and in the collections and recoveries processes where a customer has failed to keep up with their required payments. Specific examples include:

- **Response Scoring** to predict the likelihood of a customer responding to a specific marketing message

- Determination of the outcome status for individual accounts

Data will be obtained either from internal systems or external sources, including:

Application Processing System

Internal. Data from the application form (name, address, date of birth, income etc) required for the development sample.

Account Management Systems

Internal. Data for determination of the outcome status (good / bad, response / no response etc.). In many organisations there will be more than one account management system due to historical mergers and acquisitions.

Bureau Data

External. Data from Credit Reference Agencies about individuals, providing information such as payment performance with other suppliers / lenders, voters roll information etc.

Demographic Data

External. Data about groups of individuals, usually at postcode level, providing information such as lifestyle, wealth, spending preferences etc.

Fulfilment Data

Internal / External. Data indicating which customers have been targeted with which communications, and their responses to that communication. This data may come from internal sources such as a call centre or external, for example a direct mailing supplier.

Data Extraction and Storage

Data from external sources will in most cases be well documented and defined. The same however cannot be said for data obtained from many organisations' internal systems, with data definitions and data dictionaries in many cases being very poor or non-existent.

Further, data will also be provided in numerous different formats from a variety of different sources, including databases such as Oracle, SQL Server or DB2, flat files in a variety of different formats (e.g. CSV) or other sources such as SAS data sets.

As a result, the task of obtaining the required data is often the most time-consuming and difficult part of the model building process. Each time a new model is built or an existing one updated the process of extracting, formatting and matching data from a variety of sources has to be repeated, with the whole process taking anything from a couple of days to several weeks.

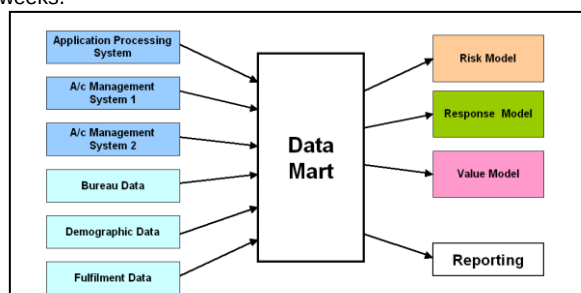


Figure 6. Credit Risk and Marketing Data Mart

Significant benefits can be gained the deployment of a single data mart to hold the data required for model building, see Figure 6. The process of pulling the data together can be highly automated and once constructed the data mart requires minimal maintenance. The data in the mart is structured in such a way as to make it extremely easy for sample data to be selected and extracted for model building, with a process that may have taken several weeks being completed with a few clicks of a mouse.

REPORTING

An additional benefit of the data mart that is worth touching on here is that the data mart can also be used to store data in a format that will support flexible report generation using standard tools such as Business Objects, Cognos or Futrix or through a bespoke reporting application.

CONCLUSION

The building and deployment of statistical models within consumer facing businesses can provide significant business benefits. These include reductions in bad debt through application scorecards and behavioural scoring, and increased revenue through cross-sell, up-sell and attrition models.

The construction of models for differing purposes for both credit risk and marketing departments is based on the same underpinning methodology, with all models requiring the extraction and manipulation of very similar data from a wide variety of sources. Significant benefits can be gained from the creation of a structured data mart to support model building. The time taken to obtain sample data can be reduced from many days or weeks to a matter of minutes.

Additional benefits include the ability to utilise the data mart to provide an extremely flexible and responsive reporting solution across both the credit risk and marketing functions within an organisation.

NOTES

¹ If interested in the various statistical techniques available for model building, a good summary is provided in the chapter "Optimal Use of Statistical Techniques in Model Building" by John M. L. Gruenstein, contained in "Credit Risk Modeling Design and Application", Elizabeth Mays, Glenlake Publishing, 1998.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author: Andy Twaits

Amadeus Software Limited

Address: Orchard Farm, Witney Lane, Leafield,
Oxon OX29 9PG

Phone: 01993 878287

Fax: 01993 878042

Email: andy.twaits@amadeus.co.uk

Web: www.amadeus.co.uk/customermanagement

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.