**Model Answers**

**CASE STUDY 1 solution**

Wireless Communication Company

1. Identify the business problem or opportunity

Identify customers who are like to respond to a new product offering, by using customer profiling.

There is a suggestion that the company needs to find 500 responders. Existing method of randomly mailing customers would require 16000 – 25000 prospects which is a 2-3 % response rate.

By using data mining methods they hope to they will need fewer prospects and therefore save money on their recruitment campaign.

1. Mine the data to transform it into actionable information

Their data mining process scores each prospect from 1 to 100, 1 being that customer is likely to purchase the product, 100 means the customer is unlikely to purchase the product. (NB this order can be reversed in other applications it is important that you know the context). This is an estimation problem. An alternative formulation of the problem could be a binary classification or responder as in the second tutorial.

1. Data was compiled from three sources: Customer information file; call detail database; demographic database.
2. Getting to know data, identifying inputs, minutes of use etc.
3. Transforming data to bring information to the surface, derived inputs such as Sphere of influence and minutes of use
4. No mention is made of building a model set, training, validation and test set.
5. Build a response model, detail about this step is absent, they probably used a linear regression, or neural network model
6. They did assess the performance of the data model by using a lift chart.
7. They did a field trial by using a controlled study of unspecified number and comparing the performance of the data mining model against selection prospects at random. The model provide to be 5 times more effective than random selection.
8. Act on the information

The applied the data mining model to a targeted group of prospects for the new product offering.

1. Measure the results

The measurement of results is not discussed but they run the campaign for the first several months of the new service. Did a round of iterative improvements to improve the model and rolled the model out to five different markets.

1. Repeat

**Case Study 2 SUV sales example**

1. Identify the business problem or opportunity

There is a suggestion that the expert system technology that they use was not very good at identifying responders to new products. They wanted to replace existing system of using expert interviews with data mining techniques.

To improve the level of response to a direct mail campaign for a new car model.

1. Mine the data to transform it into actionable information

Identify customers who would purchase a specific vehicle model (SUV).

The problem was divided into two steps, first step was to identify customers who would purchase a car from the company, binary classification problem. The second step was to identify which customers would be likely to purchase a specific advertised vehicle model, also a binary classification problem.

1. Data was compiled from five sources: mail file; Zip codes from PRIZM demographic and psychographic charaterisations of neighbourhoods; response card and telephone response files; Sales file of customers who purchased cars three months after the original mailing. (Note if you look at figure 2.5 only a fraction of purchases come from responses to cards or telephones. They are only modelling part of the overlap between mass mailing database and Sales).
2. Getting to know data, identifying inputs, minutes of use etc.
3. Transforming data to bring information to the surface, defined a target variable as

Received a mailing and bought a car, combination of two variables.

1. The mention is made of building a model set, training, validation and test set, they have two training sets one which is oversampled and one which they describe as sparse.
2. They generate a hybrid model, that is they combine two modelling techniques, the neural network works on the sparse data training set, to identify who was likely to purchase any car from the company. A second modelling technique a decision tree was then applied to predict who would purchase a chosen advertised model. The combination of both techniques was more effective than any one standalone method.
3. No information is provided on the assessment of their modelling techniques this information is missing.
4. No field trials or assessment of the model is discussed
5. Act on the information

There is an implication that they applied the model to a group of prospects and they made their offering more attractive by changing the offering from a pair of sunglasses to a pair of “nice” leather boots !

The new approach was more effective than the first lets assume that it was nothing to do with leather boots being more attractive than a pair of sunglasses !

1. Measure the results

The measurement of the success of the campaign is not discussed but there is a throw away comment that the new approach was more effective than the first !

1. Repeat

There is an indication that they would reiterate the process again but this is not discussed in any detail.

**Comments**

The first study has more details on the assessment part of the virtuous data mining cycle, which is missing from the SUV case study. However the SUV case study has more detail on building a data set and description of modelling techniques through the use of a hybrid model. These two case studies are designed to increase your exposure to a data mining methodology identified in week 2. Even though it is not explicitly expressed in the text hopefully you are now aware that there is a process behind it. The trick is to recognise and formulate the business problem, express it as a data mining problem which will typically be one of three classes of prediction: estimation; binary classification or segmentation. We will discuss segmentation, clustering later on in the course.

Refer to the slides virtuous data mining cycle discussed in the lecture in week 2.