To ask or not to ask, that is the question

Hsin-Vonn Seow and Lyn C. Thomas

School of Management, University of Southampton, Highfield, SO17 1BJ, UK.

Corresponding e-mail:hvs@soton.ac.uk, L.Thomas@soton.ac.uk

Abstract

Applicants for credit have to provide information for the risk assessment process. In the current conditions of a saturated consumer lending market, and hence falling take rates, can such information be used to assess the probability of a customer accepting the offer?

Lenders do not want to make the application process too complicated, and with the growth in adaptive marketing channels like the Internet and the telephone, they can make the questions they ask depend on the previous answers. We investigate how one could develop such "adaptive" application forms; which would assess acceptance probabilities as well as risk of default.

Keywords: Data mining, classification trees, question selection

1.0 Introduction

The personal financial market is now a buyer's market rather than a seller's market. This is resultant from the many choices a customer has with regards to a single financial products. The competition among the different banking and financial institutions show that 'take' is now a major problem in this banking scene. In a sense, this would mean that the survival and success of a banking or financial product will depend on whether a customer accepts or not an offer made to him or her.

One way of maximizing the 'take' of offers is through the customization of offers. There have been several ways of where customization can commence as indicated by Murthi and Sarkar (2003). All of such research focuses on the following question: What offer is to be made to which segment of population/person such that they will accept it?

The telephone and the Internet has vastly influenced the way offer are and can be made to customers of banking and financial institutions. Real time offers can be made in confidence to each customer. But before any offers can be made, information about the customer must be collected (Raghu *et al.*, 2001). This can be done by asking them specific questions. Since this method of asking is to be done in real time, it is important that the number of questions posed is kept to a minimum so that the 'take' probability remains high.

The questions asked will depend of the response to an earlier question hence likening this exercise to an 'adaptive questionnaire'. Bearing in mind that only a limited number of questions are to be asked, it is important to identify which questions will result in the most relevant data to be collected.

So, which questions do we ask? What variant of the product do we offer in light of the responses to the questions asked? In this paper, we propose using classification and regression trees (CART) to answer these two questions.

2.0 Classification and Regression Trees (CART)

Classification trees remain a powerful classification tool in decision making. It is one of the more popular methods used where it is successfully used in many areas. In health, CART analysis is being used to decide on treatment that is both cost conscious and appropriate in the treatment of diabetic retinopathy (Harper *et al.*, 2003). It is again chosen in the development of a model which helped doctors detect possible complications for pregnancies so as to decide on a possible course of treatment at birth (Harper and Winslett, 2004).

Classification trees allow for the analysis of the interaction between the variables. This will prove useful when building a tree to classify the data using applicant characteristic(s) to identify an offer with the highest 'take'.

Classification trees identify significant variables in a data set that influences the target variable, which is the acceptance of offers. Classification trees work to find the best splits for a significant variable in a data set. Hence, it will find the best split even if a variable is forced in. This element is very useful given that we want the tree to be categorised first by the applicant characteristics which will lead to an offer characteristic.

This will help decide which two or three questions to ask the customer. Using the responses from the questions, the data collected will subsequently lead to the decision of the offer extended. Hence, classification trees used here are utilised to help decide what offer within a possibly limited set of offers that we are willing to make that has the highest 'take' rate and can be offered next to a customer, with a constraint of only asking two to three questions.

Unlike the work referenced above, the results of the classification trees cannot be taken directly. We need to order the variables that we split on so that the splits at the top of the tree are only applicant characteristics and those lower down the tree are only offer characteristics. This makes the classification non-optimal but is needed to devise a workable offering strategy. The classification tree that is proposed here consists of classifying via applicant characteristics to help decide on an offer with high 'take' probability. We refer to such a tree as APOFT (APplicant and OFfer Tree).

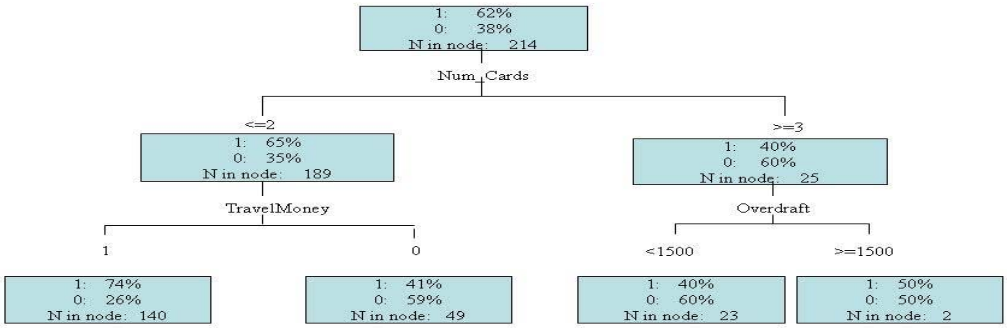


Figure 2.1: An example of a 1-Applicant characteristic-1-Offer Classification Tree

The construction of the APOFT begins with classification of the data using the most significant applicant characteristic. In a sense, its main aim is to successfully assign individuals from the data set to clusters, then find the offer that has the best 'take' probability for each of the clusters.

Lower down the tree when we are in the 'offer layer', we need to chose the most significant offer characteristic. Taking into account these requirements, there was a need for a statistical software package that had the ability to force splits on subsets of the variables. SAS 9.1.3.has Enterprise Miner 4.3 that allows the user to experiment with forced variables in the classification trees and is very user friendly and presents results well in the forms of charts and graphs. It also calculates statistical test like chi-squared test and entropy to show how one variable is more significant than another. It is suitable for the analysis that needs to be done here.

An alternative to using classification trees would be logistic regression. Logistic performs the same classification functions as a classification tree but indicates the significance of a variable by the value of its corresponding coefficients. However it cannot be used to split on applicant characteristics and then identify the best offer characteristic for each segment.

Its strength though is that it can give a prediction of the acceptance of any variant of the product offered by any applicant. We will exploit this in later work by building a tree using these hypothetical acceptances, which gives a much larger data set that that of the actual acceptances, This is a form of 'bootstrapping' which may make APOFT more robust.

2.1 'Take' Classification Trees

The aim of this piece of research is to be able to find a way to decide on offer with a high 'take' probability to extend to customers. Before any prediction can be made, certain information regarding the customers' characteristics will have to be collected. To do so, questions will have to be asked, be it in the shape of questionnaires or in real time via the internet or on the telephone. When asking for personal information, one has to be careful so as to not ask more questions than is needed for fear that the prospective customer will be uncomfortable thus reducing the probability of accepting an offer. This condition holds regardless whether this line of questioning is done face to face or through a medium like the internet.

In such a situation, it is crucial to remember to learn as much as possible by asking as few questions as one can. The difficulty of such a situation is choosing the 'right' questions to ask. The 'right' questions are defined as a minimum number of questions with responses that will generate enough data about the potential customer so that an offer made to him, or her, will have a high probability of acceptance.

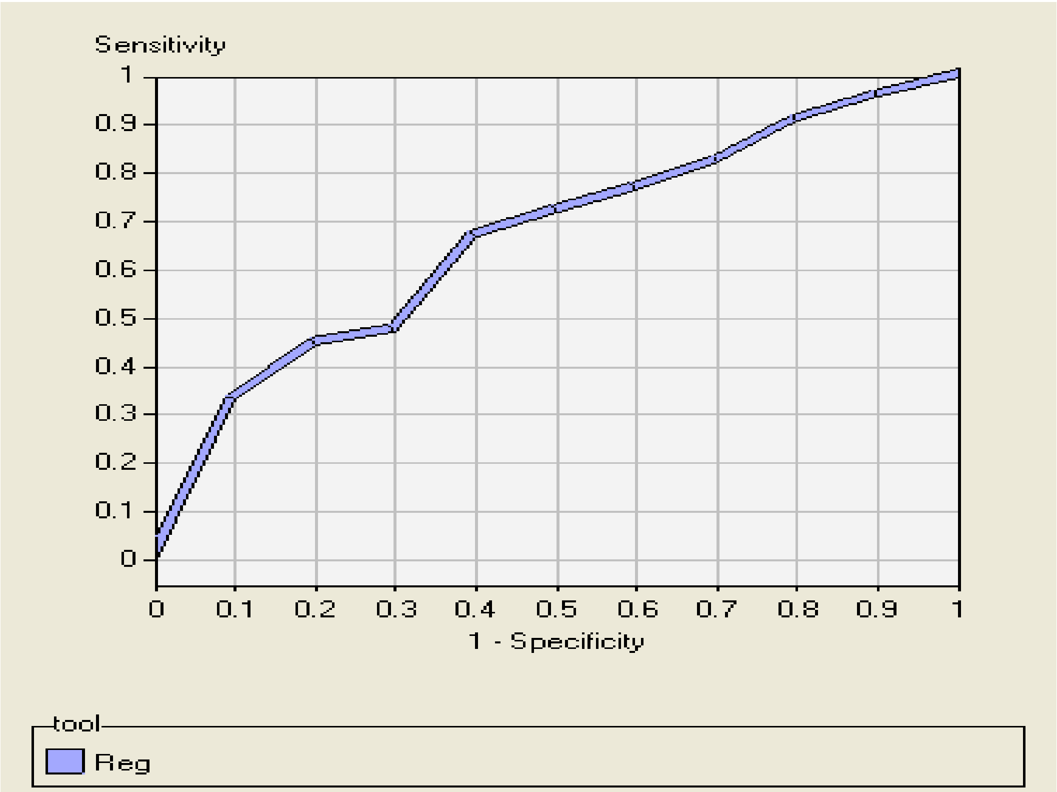
There is no publicly available data on the offer strategies of the personal financial sector. So we utilise the data gathered from a Fantasy Student account which was started in 2003. This is a web based application form of a fantasy student bank account which first year university students were encouraged to fill in. There were 17 applicant characteristics that can be obtained from this information. The software then made the students an offer of a specific account, where the overdraft limit, interest when in credit, commission-free travel money, discount on insurance and whether a credit card was also offered varied from student to student. The offer characteristics were chosen randomly in 25% of the cases and followed a decision tree which was used to ensure a variety of offers was being made in the other 75% of the cases. Further details can be found in Jung and Thomas (2004).

After pre-processing the initial data, 305 entries were used to build the classification trees. Out of this, 214 (70%) was used to train the tree, while the remaining 91 entries were used to validate the trees.

pkUserNumber	Sex	Status	num_Children	num_Cards	Wage	Loan
1	Male	single	0	1	FALSE	TRUE
2	Male	divorced	0	3	TRUE	TRUE
6	Male	married	0	1	TRUE	FALSE
7	Female	single	0	6	TRUE	TRUE
8	Male	single	0	1	TRUE	TRUE
9	Female	other	0	0	TRUE	TRUE
10	Female	single	0	1	FALSE	TRUE
11	Female	single	0	0	FALSE	TRUE
12	Female	single	0	0	FALSE	TRUE

Table 2: A sample of the data from the Fantasy Student account

The data is difficult to classify well which, as one might accept when one is trying to predict a target like acceptance, especially in the case of an acceptance of a hypothetical product. We can see the results of the logistic regression classification we built on the data (using both offer and applicant characteristics). The ROC curve for this is shown in Graph 4.



Graph 2: Receiver Operating Characteristic Curve (ROC) for logistic regression performed on data from the Fantasy Student account

The Gini Coefficient for this data set is 0.50729. This data set yields a. Kolmogorov-Smirnoff statistic and area under the ROC is 0.75364. Hence, the scorecard generated using the logistic regression is satisfactory.

2.2.1 Applicant characteristics

The applicant characteristics will be used to form clusters to which each of the customers will be assigned to, matching each customer applicant characteristic to the corresponding classification cluster. Then, the APOFT will find an offer to match each cluster where the 'take' probability is optimal for each different cluster.

Consider first a tree where only applicant characteristics are used to split the accepts from the rejects. The best such tree is given in Figure 2.2.

The tree denoted that 62% of the 214 people in the data set have accepted offers made to them with 38% rejecting them. From the 214 people, 198 of them enjoy do-it-yourself (DIY) activities while the remaining 16 do not enjoy DIY activities. Out of the 198 that do not enjoy DIY, 60% of them have accepted offers, while 34% of those who enjoy DIY have accepted the offers.

It goes on that the people that enjoy DIY can be classified in to a further two groups; people with one or less number of credit cards and people who have 2 or more credit cards. All of the people from the former group have accepted offers made while 60% of people in the former group have accepted. For those who do not enjoy DIY, 63% of people with two or less cards accepted the offer made while 33% of people with three or more cards have done so. This is a classification exercise and nothing more. The absence of offer characteristics in the classification tree prevents the matching of an offer to a cluster of customers.

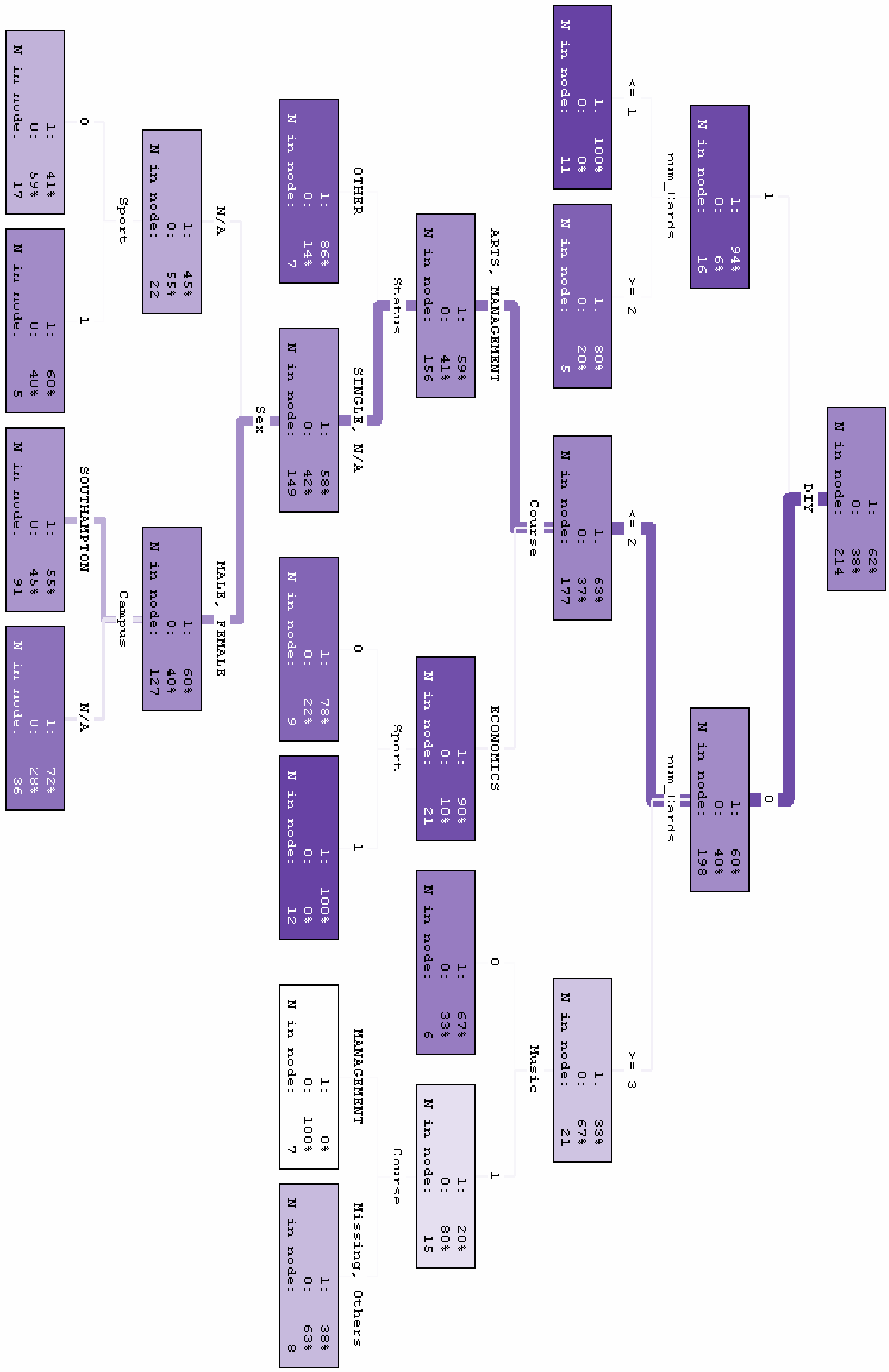


Figure 2.2: Applicant characteristics only classification tree

2.2.2 Offer characteristics

The offer characteristics will guide in making the decision of which offer will generate a high acceptance given that the customer has been classified and put into a cluster. This is effectively matching the offer to the cluster. But a classification tree with only offer characteristics cannot perform such a task. As seen in Figure 2.2, the Offer characteristics only classification tree only gives one even more information about the types of offers being extended and the 'take' probability of these offers in the data set. This tree says more about what offer characteristics were more popular and nothing about who should be targeted with what offer.

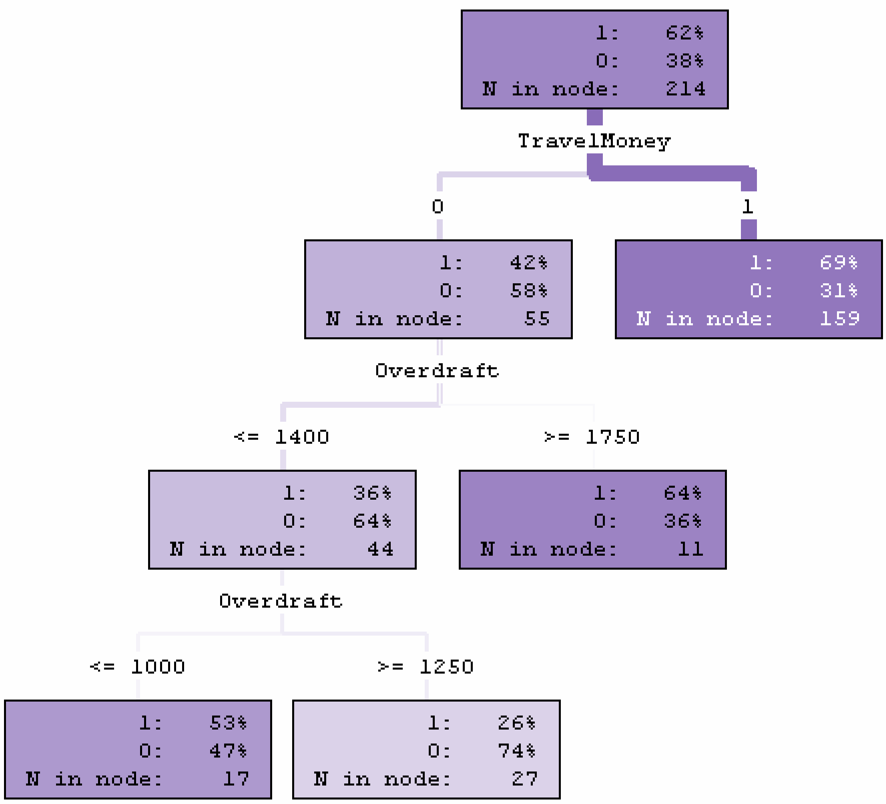


Figure 2.3: Offer characteristics only classification tree

Notice again that this is a classification exercise showing that of 62% of the total 214 people have accepted offers. 69% of the 159 people that had travel money offered to them accepted while 42% of 55 people who were not offered travel money accepted. But from this 55, 44 were offer an overdraft of £1400 and less with 36% accepting while the others were offered an overdraft of £1750 and above, 64% of those accepted. Out of the 44 offered an overdraft of £1400 and less, 53% accepted an offer with an overdraft of £1000 and less while 26% from the 27 offered £1250 and more accepted the offer.

So, the appropriate step is to build a classification tree which classifies the individuals in the data and forms clusters via the applicant characteristic. Then, using the tree, pick the most significant offer (based on 'take' probability). The number of questions asked will have to depend on how the individuals in the data set are segmented further. For example, if the specific cluster is formed from the classification using two applicant characteristics, then two questions can be asked of the customer to get information based on those two applicant characteristics.

2.2.3 Both applicant and offer characteristics

The following is the tree built by Enterprise Miner which is the best tree using all 24 applicant and offer characteristic variables. However, this tree does not help decide which offer to make to which applicant segment as this is an offer characteristic above an applicant one. We have to make that offer to all, not just those with that applicant characteristic. No offer can be made for a sense of logic as the variables are randomly scattered among the nodes of the tree. See Figure 2.4.

For example, from Figure 2.4, it can be said that there are 5 people who do go to the cinema and have an overdraft of £1500 and have been offered travel money before. There is no help identifying offers with a high acceptance rate. Hence one needs to have applicant characteristics first to classify the data set before looking at the offer characteristic.

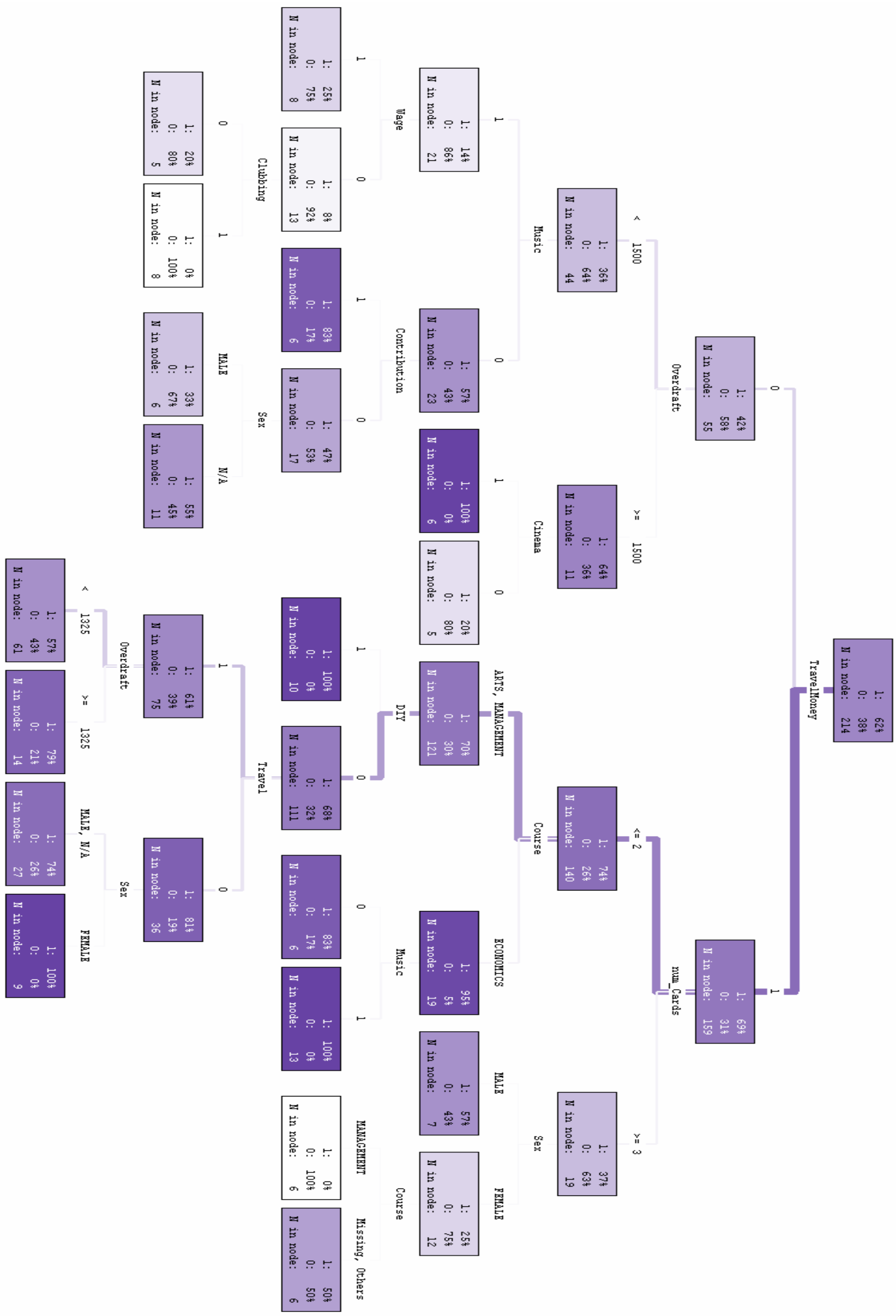


Figure 2.4: **Applicant and offer characteristics classification tree (with no 'initial classification using applicant characteristic' rule)**

3.0 Results

We outline the procedure for finding the best offers to make, if we restrict ourselves to only asking two questions. Clearly we can extend this to allow more questions to be asked. Initially we look at the offer only tree and recognise that the most significant offer characteristic is whether or not to give commission-free currency exchange (travel money).

In Figure 3.1, we build a tree with two level of applicant characteristics splits (corresponding to the two questions we are allowed to ask) but with the first level split being that on travel money which is the most significant in the whole population.

The tree suggests that the first question should be on how many credit cards does the applicant have. Hence, the responses split the population into those with 0, 1 and 2 cards and those with 3 or more credit cards. The second question for the first group is that subject they study and for the second group, whether they receive a wage or not (i.e. do they have a part-time job). Notice the questions asked are adaptive in that they depend on answers to the previous question.

The problem with this tree is that we have forced the offer decision to be the same for all groups – to whom do we give travel money? The tree suggests we give travel money to those with 2 or less credit cards. We assume that if an offer feature is not significant in the tree, we will leave it at some pre-assigned level.

The obvious question is whether it is better to vary the offer on other characteristics to the few applicant groups we have identified. So Figure 3.2 is the result of allowing this third level splits to be on any offer characteristic. In this case, we see that it is both travel money and overdraft which most affect the acceptance rate. In this case, it is best to offer commission-free currency exchange to those with 2 or less credit cards and to offer overdrafts of more than £1125 to those with more than 2 credit cards but have a wage, but overdraft of less than £1125 for those who have more than 2 credit cards but no wage.

We notice that the waged, more than 2 credit card group has an acceptance rate of less than 30% and so one might decide not to make them any sort of offer.

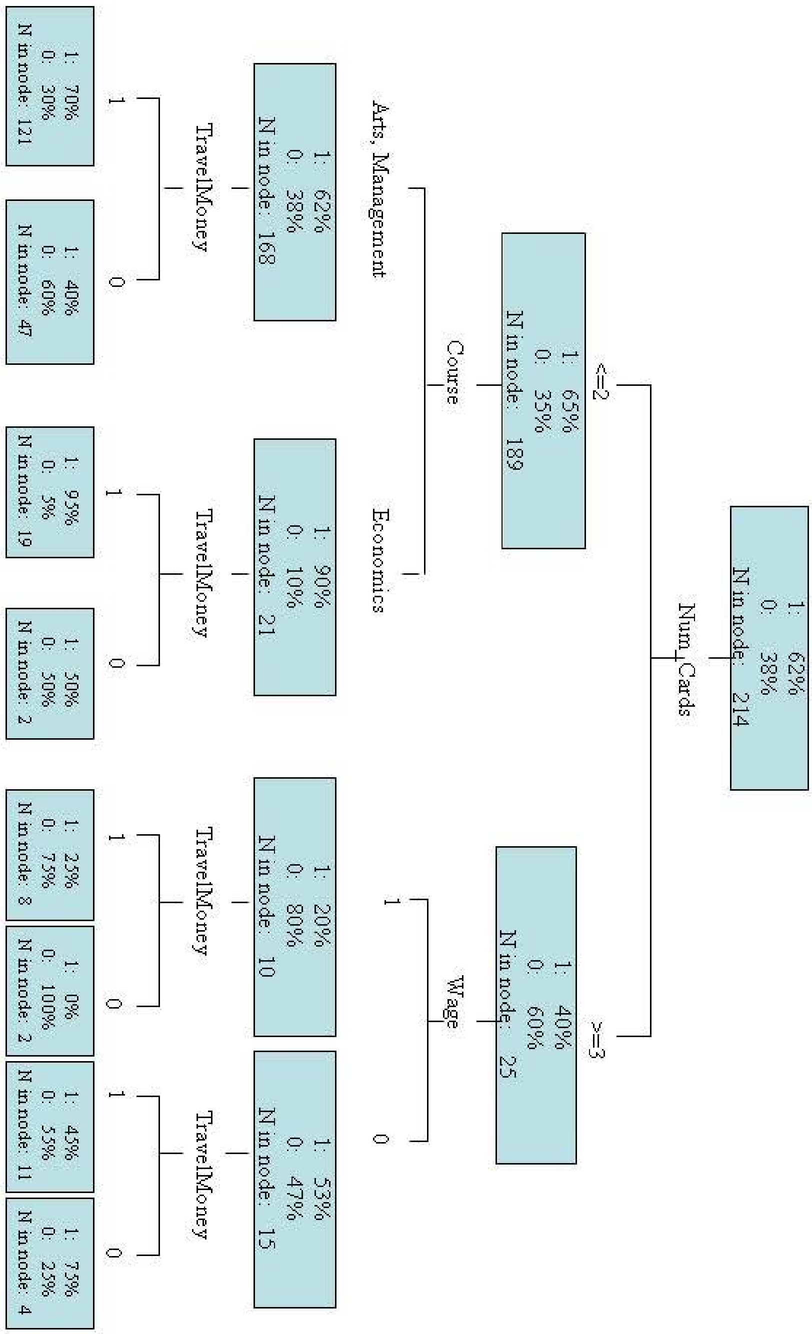


Figure 3.1: Applicant and offer characteristic tree- Travel money

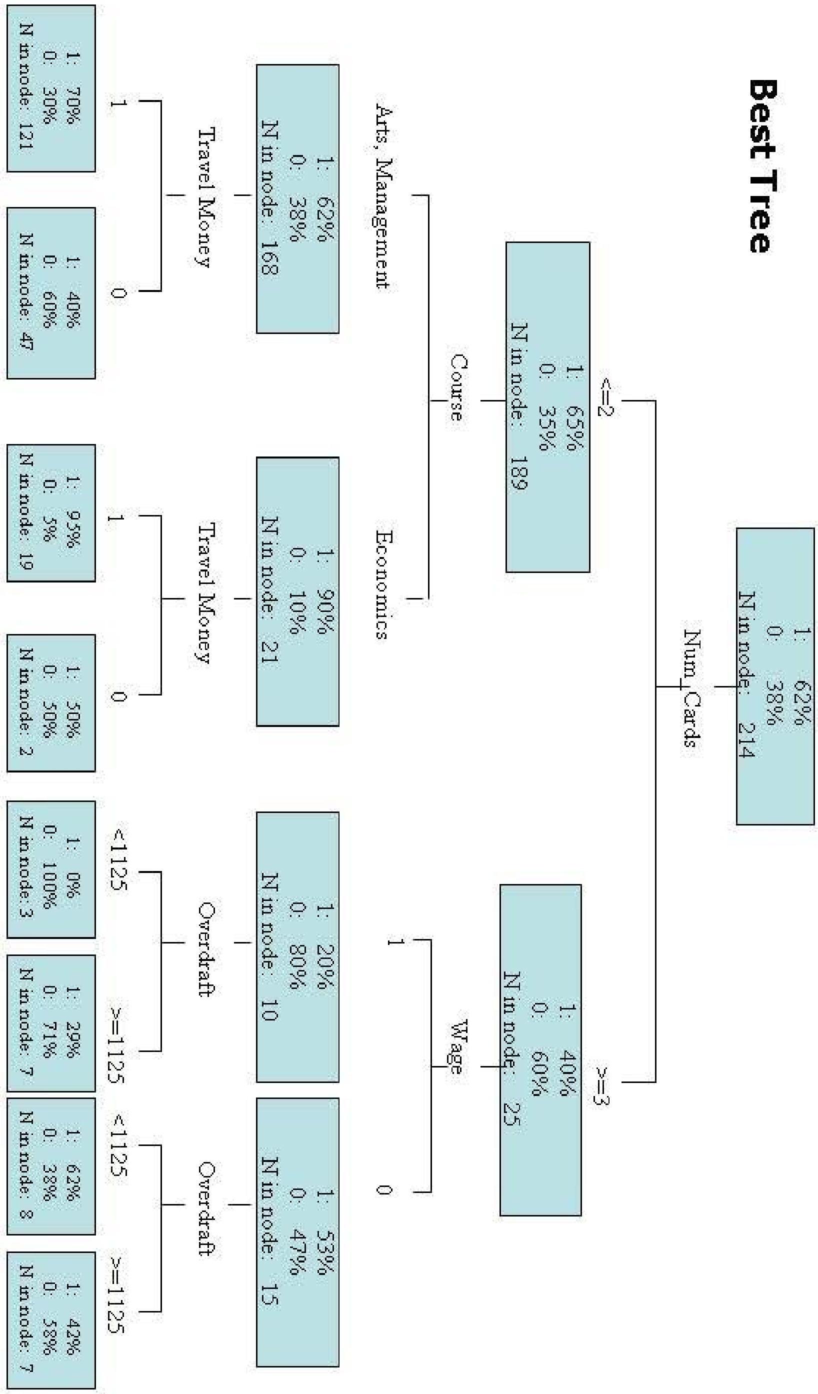


Figure 3.2: The best classification for the clusters in the Fantasy Student data set

What can be interpreted from the best tree is the following:

a) If a customer possesses two or less credit cards and is a student of Arts or Management, hence the best offer to make to this individual is travel money.
b) If a customer possesses two or less credit cards and is a student of Economics, hence the best offer to make to this individual is travel money.
c) If a customer possesses three or more credit cards and is earning a wage, the best offer to make to this individual is overdraft of £1125.
d) If a customer possessed three or more credit cards and is not earning, the APOFT predicts the best offer to make is an overdraft of less than £1125.

4.1 Alternate Best Tree

There is an alternate which is quite similar to the tree in Figure 3.3. In the first tree, we were only allowing offer or applicant characteristics to vary in a tree. Suppose we allowed the final level of applicant characteristic and the offer characteristic both to vary. We get this tree in Figure 3.3, which leads to better acceptance rates. In this tree, we check the set of those with 3 or more credit cards and if male, give them 0% interest on credit, and if female an overdraft of less than £1125. These seem odd results in that in both cases, one is making an apparently worse offer. This could be due to the data coming from a fantasy situation, but it is recognised that some students do not like having large overdraft limits, while others believe that interest on current accounts is compensated for by hidden charges and is no benefit of you are never in the black.

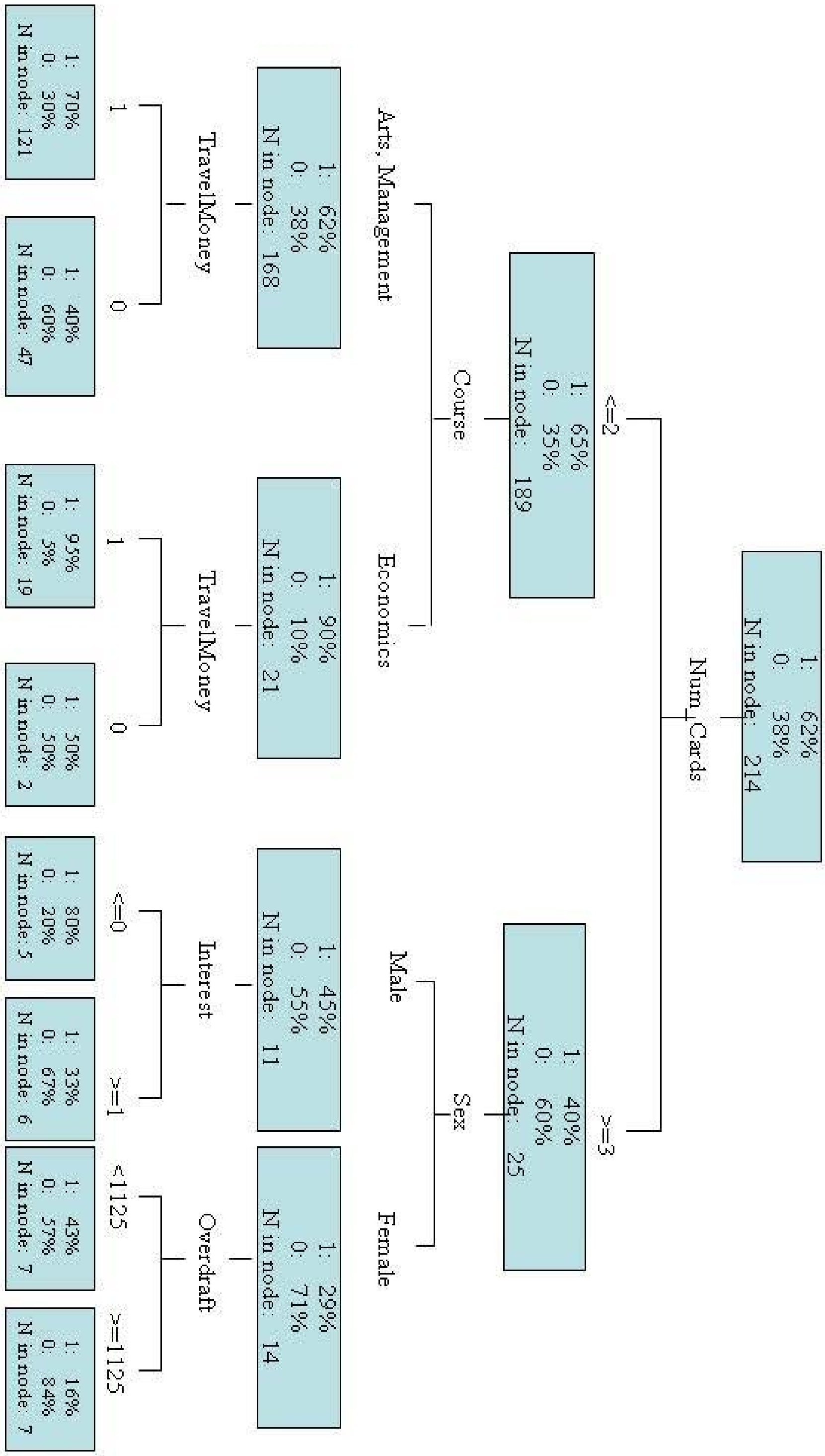


Figure 4.5: Alternate best applicant and offer characteristics tree

These two best trees show what information about the applicant should be collected as they will help in deciding on an offer with a high acceptance rate.

3.2 One question classification tree

The following trees show the acceptance rate if only one question is asked. This is simpler compared to trees for 2 questions.

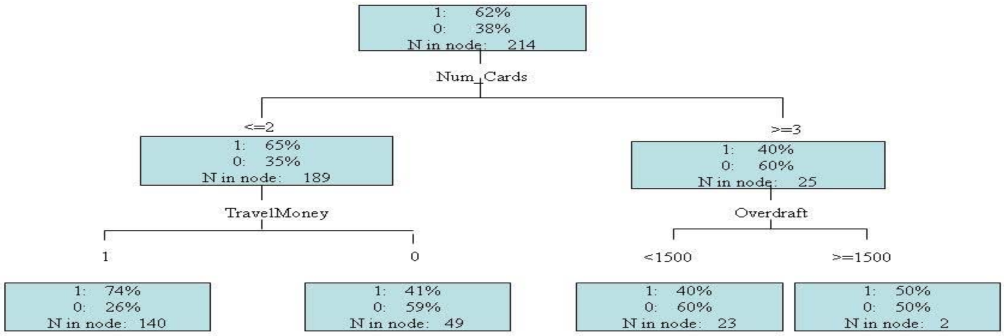


Figure 3.4: One question APOFT with predicted offer

Hence it is possible to match an offer to a cluster by means of only asking a question.

Conclusions

The APOFT classification trees can be successfully used to help decide which offer to extend to each cluster formed from a classification using applicant characteristics. As well as showing the feasibility of this approach, it illustrates the importance of acceptance scoring. This piece of work shows that it is possible to predict the acceptance of offers and the selection of such offers can be achieved via classification trees. Hence the APOFT is a successful in deciding on a high 'take' offer to best match a cluster by asking one to two questions.

To alleviate the problem of the small sample of data, we are investigating logistic regression as means to get extended data. In this paper, we build a logistic regression scorecard based on offer and applicant characteristics to estimate the likelihood of a specific applicant accepting a specific offer. We then use the hypothetical probabilities to build a scorecard. This means that for each offer split, the daughter sets can remain the same size as the original set that was being split.

References

[1] P. R. Harper., M.G. Sayyad, V. de Senna., A.K. Shahani, C.S. Yajnik and K.M. Shelgikar, A systems modelling approach for the prevention and treatment of diabetic retinopathy, European Journal of Operational Research. 150(2003) 81-91.

[2] P.R. Harper and D.J. Winslett, Classification trees: A possible method for maternity risk grouping, European Journal of Operational Research(2004)-to be published.

[3] K.M. Jung, and L. C. Thomas, A note on coarse classifying in accepted score cards, Discussion Papers in Management, M04-16 Southampton: University of Southampton (February 2004).

[4] B.P.S. Murthi and S. Sarkar, The Role of the Management Sciences in Research on Personalization, Management Science 49(10)(2003) 1344-1362.

[5] T.S. Raghu, P.K. Kannan, H.R. Rao, A.B. Whinston, Dynamic Profiling of Consumers for Customized Offerings over the Internet: A Model and Analysis, Decision Support Systems 32(2001) 117-134.