Nova Information Management School

Advanced Analytics Master's degree

Predictive models

Project 1: Delicatessen Company, Classification

Group 4

Carolina Bellani - M20170098

Gonçalo Passão - M20170450

Sofia Jerónimo - M20170070

Table of contents

Introduction

The goal of this project is to support the marketing department of the organization Delicatessen and produce the highest profit for the next campaign. With this goal in mind, we created a predictive model in order to generate the highest theoretical profit using the information at our disposable.

Analytical process

1. Feature transformations

a. Extracting new features

In order to have an interval variable, we created in excel a new variable called 'Relative0630' which is the date of the customer ('Dt_Customer') minus the last date of 'Dt_Customer'+1 (06/30/2014). After exploring the variable 'Daysactive', we realized that this variable had negative values, so we summed 94 to have only positive values.

- **Relative0630** – The time in days since the customer first bought in the company (Dt_Customer-(06/30/2014)+94)

In SAS code node, we created the following new features:

- **Daysactive –** The time difference in days since the customer first bought and last bought from Delicatessen (Relative0630-Recency)
- **Age –** Age of the customer (2018-Year_Birth)
- **AcceptedCmpTotal** – Total accepted campaigns (AcceptedCmp1+AcceptedCmp2+AcceptedCmp3+AcceptedCmp4+AcceptedCmp5)
- **MntTotal** - Total money the customer spent in the organization, some of the products may overlap but it should give a rough estimate and we don’t have enough information to reverse the overlaps. (MntFishProducts+MntFruits+MntGoldProds+MntMeatProducts+MntSweetProducts+ MntWines)
- **DependentHome** – Number of dependents the customer has (Kidhome+Teenhome)
- **PercSpent** – Percentage of the customers income spent in Delicatessen (((MntTotal/2)/Income)*100)
- **PercSpentIncomex2** – Percentage of the customers income spent in our company standardized ((MntTotal/(Income*2))*100)
- **Npurchcatwebstore** – Number of total purchases NumCatalogPurchases+NumStorePurchases+NumWebPurchases

- **Discountpurchasespercentage** – Percentage of purchases in discount products ((NumDealsPurchases/Npurchcatwebstore)*100
- **MntGoldPerc –** Percentage of monetary spent only in gold products ((MntGoldProds/MntTotal) * 100)
- **MntTotWthoutGold** – Total money spent in Delicatessen except gold products (MntFishProducts + MntFruits + MntMeatProducts + MntSweetProducts + MntWines)
- **PercSpentwithoutgold** – Percentage of income spent without accounting for gold products (((MntTotWthoutGold / 2) / Income) * 100)
- **PercSpentIncomex2withoutGold** - Percentage of income spent without accounting for gold products standardized ((MntTotWthoutGold / (Income*2))*100)

The following interactions:

- **Tpurch_perspent**=Npurchcatwebstore*PercSpentIncomex2
- **TMnt_Compl**=MntTotal*Complain
- **Tacc_compl**=AcceptedCmpTotal*Complain
- **Tdep_ed**=DependentHome*Education (obs. before Education was transformed from nominal to values and the formula is in the excel file)
- **Tmaritstatus_depend**=DependentHome*Marital_Status (obs. before Marital_Status was transformed from nominal to values and the formula is in the excel file)

And finally:

- **ClusterID**: indicated which customers belong to cluster 1 or 2. This two clusters were obtained by 'High Performance Clustering' using 'Align Box Criterion', standardize and PCA initialization. The appropriate number of clusters was accessed by using 'Global Peak' as criterion. We used as input variables 'daysactive', 'AcceptedCmpTotal', 'MntTotal' and 'Recency' because every selection study (see Feature selection) defined these variables as relevant. To have this variable, we needed to merge the results of the clustering with the dataset. (obs. We did not add this variable and its formula in the excel file to not create problems in the document analysis)

b. Binning, correcting non-normality and standardize the variables

We standardized and maximum normal the variables before performing any models. We did the 'best' transformation method which tries several transformations and selects the one that has the R-square value with the target. In addition to that, we performed optimal binning transformations to maximize the discriminant ability before performing decision tree models and the random forest.

2. Feature selection

We used all original and transformed variables as an input and performed various methods to select the main candidate features for our analysis by using:

- Principal Component Analysis of a correlation matrix: selecting the nine top principal components, they together explain 53.80% of the variance.
- 'HP Variable Selection': this is an automatic way of selecting nine variables by estimating the R-square.
- Clustering
- Worth, Scale Mean Deviation and Chi-square from the 'StatExplore' node
- The variable importance from the tree decision models
- Gini reduction from the random forest model
- Absolute coefficient effect using different regression selections (Forward, Backward and Stepwise) estimated by AIC.
- Hinton diagrams showing the weights of neural networks using misclassification selection

After selecting the top-ranking candidate features, we explored the correlation matrix (Pearson and Spearman's coefficients) and cluster analysis results to have the final decision of the nine sets of variables to further test.

Across all the possible models tested a few variables seem to remain constant as good predictors for the target variable. ‘AcceptedCmpTot’ is a remarkable predictor, which makes sense taking into account we are attempting to predict the result of a new campaign and this variable shows us the result of previous campaigns. Two other variables that showed to be good predictors were ‘MntTotal’, ‘MntGoldPerc’, ‘NumCatalogPurchases’ and ‘PercSpentIncome’ which are variables related to the amount of money our clients have already spent in our company. The more money they have invested and trusted into our business, higher the chances that they will buy from us again. For last but not least, the time commitment into our business seems to be a decent predictor. These time variables include ‘Recency’, time since last purchase, ‘RelativeDate’, time since they are our customers and ‘daysactive’, a variable transformation mixing the information of both previous variables.

The previously mentioned variables, although apparently good predictors, may not be used across all the models tested so we can have a wider span of possibilities and diversity in our models.

For the remaining used variables, we combined all the information we arranged from our analysis, in particular, the lack of correlation with the other chosen variables to get the most diversified set as possible. Since there is not an easy way to combine the huge amount of information, to select the perfect combination of variables we tried different possibilities with several pre-processing steps in our quest for the champion model and understand which variables interact better.

To be sure the elimination of outliers did not influence the model we tried all the mentioned models without removing the outliers and for the majority of the indexes, they did not improve our models. After imputation of tree surrogate values on the outliers, the models did not improve as well.

3. List of algorithms tested

After the selection of the sets of nine variables, we tested the decision-trees, random forest, neural networks, logit and probit regression algorithms.

For the neural network nodes, we selected always standardized variables as inputs and in particular:

- Default activation function, momentum 0.1, weight decay 0.1, Rprop to have a dynamic learning rate
- Activation function logistic, momentum 0.1, weight decay 0.1, backpropagation as learning technique
- Activation function logistic, momentum 0.1, weight decay 0.1, default as learning technique
- Default activation function, momentum 0.1, weight decay 0.1, backpropagation as learning technique
- No preliminary phase, all default
- Default but with preliminary phase
- Auto Neural
- 20 hidden units, misclassification
- 6 hidden units, weight decay 0.2
- Backpropagation learning technique, momentum 0.1
- Tanh as activation function, weight decay 0.1, learning technique backpropagation, learning rate 0.1, momentum 0.1. (Obs. In this case, to have more accuracy we could change the domain of the target from (0,1) to (–1,1).)

For the regression nodes we select probit and logistic regressions with:

- Selection criterion AIC with intercept
- Backward selection
- Forward selection
- Stepwise selection

We noticed that not all the assumptions of the regression models are verified and in particular there is not multicollinearity among the values of the input variables (this could be solved taking in consideration Partial Least Squares regression for example).

Decision trees (using Bonferroni adjustment):

- Gini, Largest
- Gini, Largest, Depth3
- Gini, Misclassification
- CHi-squared Automatic Interaction Detection (CHAID)
- Chisq, largest subtree with at most 5 leaves
- Chisq, largest
- With several stopping criteria customized: split size 30, leaf size 25, minimum categorical size 15, number of surrogate rules 3

HP forest:

Loss Reduction with the k number of attributes randomly sampled at each node chosen minimizing the misclassification rate.

Ensemble:

When we use 'Ensemble' node to combine different model, we defined the average of the predicted values for the interval target and voting for the posterior probability of the class target.

4. Set of variables tested

The nine set of variables obtain by using the PCA did not perform well when testing the different algorithms (maximum ROC result was 0.882 obtained by using the NN default functions, mom 0.1, decay 0.1, Rprop method and the profit was 799.5).

a. Set 1: AcceptedCmpTotal, NumWebVisitsMonth, ClusterID, Daysactive, Discountpurchasesperc, NumCatalogPurchases, MntGoldPerc, Recency, tmstatus

We can notice there is no ´MntTotal´ or any combinations of it in this set. As justification, we can notice that in the clustering (Table 1) ´MntTotal´ has high difference between the two clusters in mean and in standard deviation; this gap, together with the information given by the clustering, makes ´ClusterID´ a possible candidate substitute of ´MntTotal´.

Table 1 - Descriptive statistics for cluster 1 and 2.

Variable	Cluster	Mean	Standard Deviation
daysactive	1	0.3052	1.9314
	2	-0.1774	1.9387
AcceptedCmpTotal	1	0.5007	3.2634
	2	-0.2910	1.4323
MntTotal	1	1.0989	2.4887
	2	-0.6387	1.4998
Recency	1	0.0406	1.9291
	2	-0.0236	1.9660
MntGoldPerc	1	-0.5611	1.2170
	2	0.3261	1.9584

Moreover, in the figure 1 we can notice that ´MntTotal´ vs. all the other variables, has high discrimination between ´ClusterID´ 1(blue) and 2 (red). So, considering ´ClusterID´ we hoped to have all the importance we have with ´MntTotal´ in addition to other information given by the clustering.

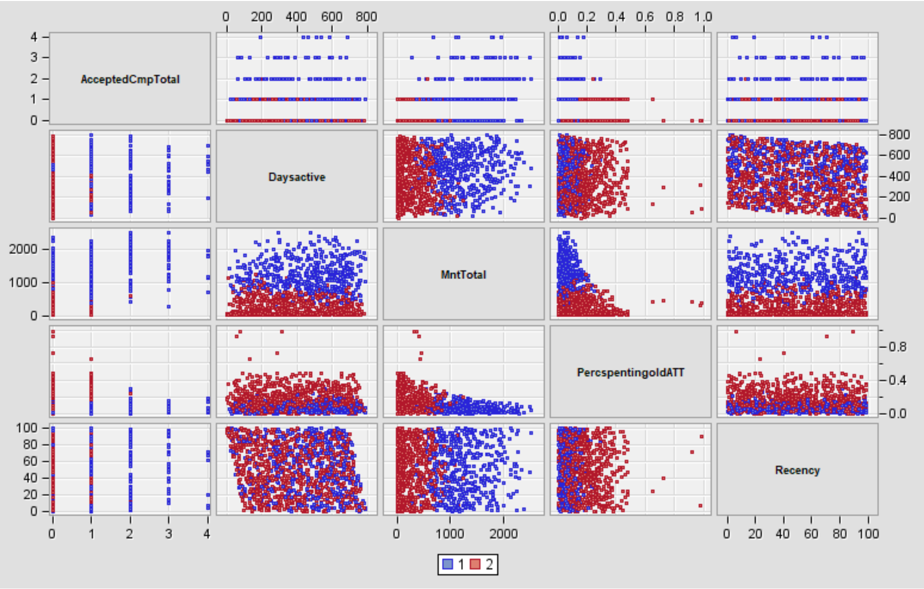

***Figure 1** - Plot of all the five variables using for the HP Clustering. Blue represents cluster 1, red represents cluster 2*

In this selection, we took to the extremes the importance to not have high correlation coefficients (neither Pearson and Spearman) trying to find the best trade off given by the relevance and the redundancy. (Figure 2)

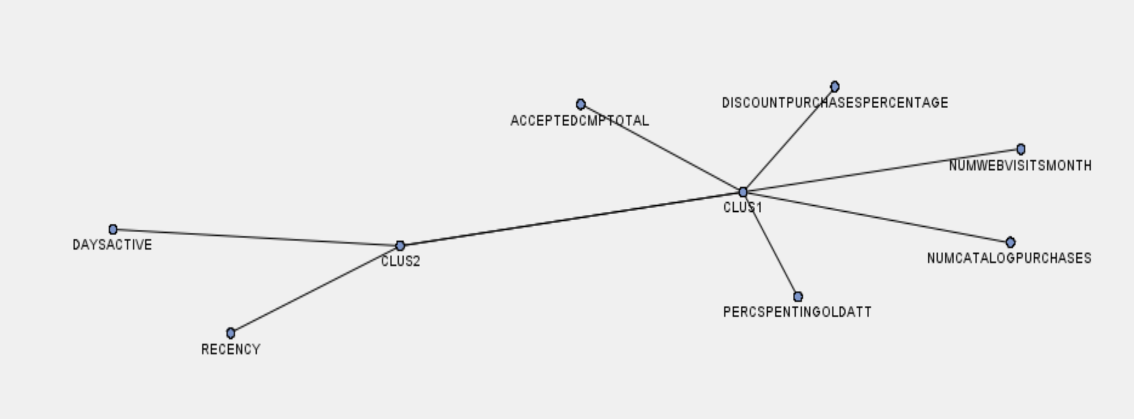

Figure 2 *- Variable clustering of the selected interval variables.*

We tried possible candidate transformations as optimal binning and maximum normal. To find the best ones, we compared for the same variables the results of Gini Reduction in HP Forest and the worth variable.

We considered the following mix set: AcceptedCmpTotal, SqrtDaysactive, LogNumCatPurchases, ClusterID, Recency, Discountpurchases, LogMntGoldPerc, tmstatus, NumWebVisitsMonth.

Moreover, we considered also the optimal binning variables and the originals variables subsets. For each subset, we studied the presence of outliers and we applied the algorithms. Comparing the results, the best in terms of Roc index and profit was obtained with no transformed variables.

In conclusion, we did the cross validation with this set for five seeds and we obtained (Table 2)

Table 2 - Cross validation with five different partition seeds.

Model	**Depth**	**Profit Average**	**Profit STD**	**Roc Average**	**Roc STD**
EnsembleNN	**15**	960.45	76.6535	0.8686	0.04495
	20	970.4	151.241		
HP Forest	**15**	839.45	104.216	0.837	0.05759
	20	893.4	130.636		

Unfortunately, with such attention about the correlation we did not expect to have these relative low ROC and profit results; so in the following set we did not consider no correlations as priority but importance.

b. Set 2: AcceptedCmpTot, MntTotal (Mntgold included), NumCatalogPurchases, PercSpentIncomex2, Recency, Relative0630, T1_AcceptedCmpTota_Complain[1], T3_DependentHome_Marital_Status[1], Discountpurchasespercentage.

The choice of this set came out for the analysis of the importance of naïve decision trees, HP forest and neural networks. We then, transformed these candidate variables with 4 optimal bins. After, we performed the research for the potential outliers and we decided to remove 6 outliers, all of them have 'OPT_PercSpentIncomex2' with too much high values comparing to the general tendency. We noticed that these 6 observations have depvar=0. We compared the model with and without outliers (identified after the binning transformation) and in terms of profit the second one maximizes it more.

We applied all the mentioned algorithms above and we ensembled (voting, average) some NNs (default funs with Rprop, default funs with Backprop, default with preliminary phase, HU=20 misclassification) with Probit Stepwise regression model (the differences in ROC between the regression models are not remarkable; 0.904 for Probit models and 0.902 for Logit) and a second ensemble for the decision tree algorithms.

We reported the results in the cross-validation tables (Table 3 and 7). Comparing with the other obtained results, this selection from the naïve algorithms have correlated variables and with comparing the validation set with training set and trying different partition seeds, this model do not have so competitive profits.

Table 3. Measures for the two best models for the set 2 of variables with five seeds.

Models	Depth	Profit Average	Profit STD	ROC Average	ROC STD
EnsembleNNReg	15	1002.25	101.296	0.878	0.0219
	20	1016.6	107.912		
HP Forest	15	942.85	205.744	0.8798	0.0292
	20	981.4	195.3669		

[1] In some sets, there will be just one of the dummy variables of the interaction, because during the first exploratory we studied the importance of them separately.

c. Set 3: AcceptedCmpTotal, Recency, Relative0630, MntFishProducts, MntMeatProducts, Education, MntGoldProds, Income, NumWebPurchases

For this set, we standardized the interval variables. Considering only the original variables, the six with higher variable importance obtained by the 'HP Selection' node were 'AcceptedCmpTotal', 'Recency', 'Relative0630', 'MntFishProducts', 'MntMeatProducts', 'daysactive', 'MntGoldProds', 'Income', 'NumWebPurchases'. We verified that 'Relative0630' and 'daysactive' were highly correlated, so we kept 'Relative0630' and add the next higher important variable, which was 'Education'. For these variables, we tried optimal binning of 4 and 5, and compared these models with models without these variable transformations.

We tried also models with and without potential outliers for 'MntGold' and 'MntMeat'. With and without outliers and optimal binning of 4, the result was similar, being the best result for NeuralHU=6, weight decay=0.2, with a ROC of 0.909 and profit of 1197 at depth 20.

With binning of 5, the results improved, having the best profit the neural networks using logistic function, momentum 0.1, decay of 0.1, default learning with 1428 at depth 20 and ROC equals to 0.911. The ensemble had a profit of 1189.25 at depth 15 and the best ROC of 0.929.

In the following table, we show the cross validation of these two models, the models seemed good at the starting, but after changing partition seeds the variance was noteworthy.

Table 4 - Measures for the two best models for the set 3 of variables with five seeds.

Model	Depth	Profit Average	Profit STD	ROC Average	ROC STD
Neural Network	**15**	1066.05	206.4252	0.884	0.031837
	20	1093.6	233.0962		
Ensemble	**15**	1085.85	199.2913	0.8966	0.020304
	20	1010	123.2295		

d. **Set 4: Age, AcceptedCmpTotal (with optimal binning), Days active (with optimal binning), Recency (with optimal binning), Education, NumDealsPurchases, Marital Status, NumWebVisitsMonth (with optimal binning), MntGoldProds (with optimal binning)**

We selected nine standardized variables using the 'HP selection node' when considering all variables. I used this set of variables for the decision tree models and with and without optimal binning for the remaining models. Comparing the models, the best three results were obtained using the optimal binning in the following models: probit regression using Forward and Backward selection, neural network no preliminary weights. The ROC for the training set was 0.904 and the profit was 1177.56$ at depth 20. At depth 15, the profit was 1062.106$ for neural network no preliminary weights, 977.9$ for probit regression using forward selection. The ensemble of the three best models improves the ROC index of the valid set, being 0.909 and maintain the profit of 1177.56$.

e. **Set 5: PercSpentIncomex2 (with optimal binning), NumCatalogPurchases (with optimal binning), Recency (with optimal binning), Income (with optimal binning), AcceptedCmpTotal (with optimal binning), MntTotal (with optimal binning), MntGoldPerc (with optimal binning), daysactive (with optimal binning and square root transformation), discountpurchasespercentage (with optimal binning).**

These variables were selected by a mix of variable worth from different models, good results across a wide range of models and different information each variable provides. In the beginning, taking into account the outliers, six observations were removed from our dataset. The first one came from an outlier in the variable ‘PercSpentIncomex2’ that corresponded to a customer that spent 17.5% of his total income in the business (Figure 3). The other five observations removed were due to customers with higher values in income compared to the general tendency (Figure 4). These removed outliers correspond to 0.004% of our dataset.

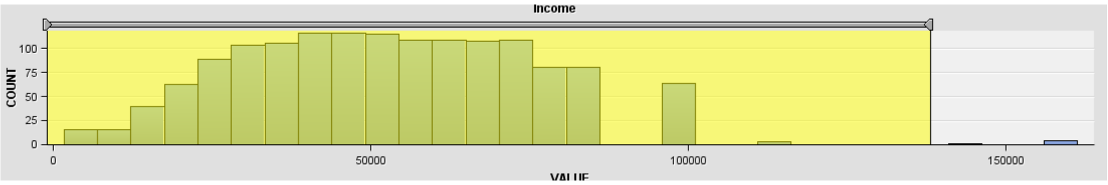

***Figure 3** – Outliers on income*

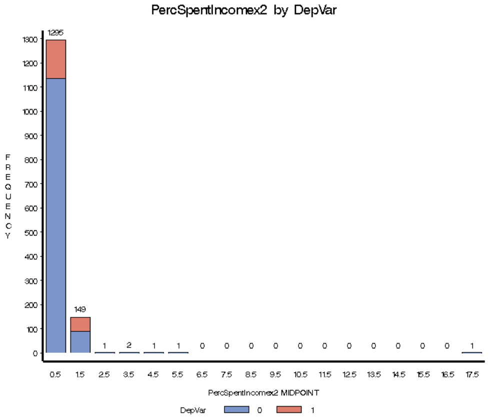

***Figure 3** – Outlier on PercSpendIncomex2*

Every variable was transformed using best method transformations except 'daysactive' that was square root transformed. Regarding these back to back transformation in this variable, it proves to be able to provide a higher level of discrimination through a previous transformation to the binning.

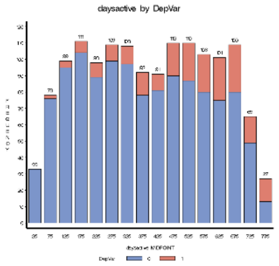

Figure 5 – daysactive

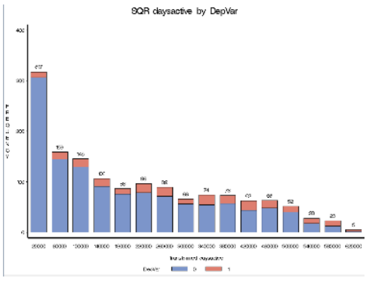

Figure 6 – SQR_daysactive

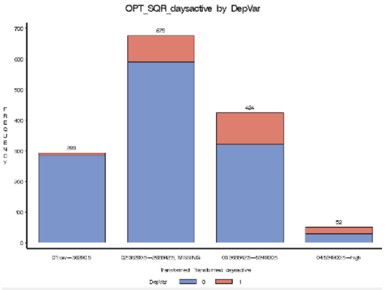

Figure 7 – OPT_SQR_daysactive

Following this process, the 4 optimal binnings generated the best results when compared to the other possibilities. The best result was a neural network using misclassification as a parameter, with 0 weight decay and 999 maximum iterations for optimization and using 50 iterations for preliminary training, of depth 20 with a total profit of 1306$ across the 5 seeds used for cross validation, with a standard deviation of 174$. Considering the average and standard deviation profit, it has a quite high deviation, which could be argued that it could be better to use the neural network or the ensemble, both of depth 15. Since they have a smaller standard deviation even though a slightly smaller profit, meaning it would be more robust in an unseen dataset.

Table 5 - Measures for the two best models for the set 5 of variables with five seeds.

Models	Depth	Profit Average	Profit STD	ROC Average	ROC STD
Ensemble	**15**	1275.42	132.17	0.9284	0.014622
	20	1264.57	179.54		
Neural Networks	**15**	1258.58	118.666	0.925	0.01713
	20	1306.67	174.83		

f. **Set 6: Relative0630 (with log10 transformation), MntGoldPerc (with optimal binning), Income (with optimal binning), AcceptedCmpTotal (with optimal binning), Recency (with optimal binning), PercSpentIncomex2 (with optimal binning), discountpurchasespercentage (with optimal binning), MntTotal (with optimal binning), NumCatalogPurchases (with optimal binning).**

After the initial transformations 6 observations were excluded due to having income and PercSpentIncome too high compared to the general sample. After, 8 out of the 9 variables were transformed using binning with the exception of the time variable Relative_0630 that was transformed using log10 manually due with the goal of using binning after, although the different model testing proved that would be more useful without binning.

For the modeling we used Neural Networks with a different set of parameters, a stepwise regression and ensemble using both average and maximum for the predicted values. The best results were of the both ensembles when combined with the regressions and the default neural network, with the winning model being the ensemble average with a profit of 1272.6$ and a depth of 20. Although the ensemble maximum is really close in terms of profit, the standard deviation indicates that the ensemble average would be more robust with less variance. In terms of the ROC, the results show no difference.

Table 6 - Measures for the two best models for the set 6 of variables with five seeds.

Model	Depth	Profit Average	Profit STD	ROC Average	ROC STD
Ensemble Maximum Neural and Regression	**15**	1268.5	177.5401	0.92475	0.019805
	20	1235.2	163.0451		
Ensemble Average Neural and Regression	**15**	1226.7	100.9366	0.92225	0.01031
	20	1272.6	158.949		

Table 7 - Cross Validation with 5 seeds: 937162211, 12345, 654321, 1249821, 10270119. The best model corresponds to the set 5. Set 1 was omitted in the table due to a lack of good results.

Set	Model	Depth	Profit Average	Profit STD	ROC Average	ROC STD
2	**EnsembleNNReg**	15	1002.25	101.296	0.8780	0.0219
		20	1016.6	107.912		
	HP Forest	15	942.85	205.744	0.8798	0.0292
		20	981.4	195.367		
3	**Neural Network with weight decay**	15	1066.05	206.4252	0.884	0.031837
		20	1093.6.67	233.0962		
	Ensemble	15	1085.85	199.2913	0.8966	0.020304
		20	1010	123.2295		
4	**Probit Forward Regression**	20	1012.4	163.833	0.8980	0.0123
		15	935.3	155.649		
	Probit Back Regression	20	1021.2	158.825	0.8976	0.0133
		15	941.9	176.468		
	Neural Network Backpropagation	20	966.2	116.848	0.8916	0.0130
		15	957.3	213.358		
	Ensemble	20	1030	163.806	0.9014	0.0143
		15	1018.9	213.473		
5	**EnsembleNNReg Average**	15	1275.42	132.170	0.9284	0.0146
		20	1264.57	179.540		
	Neural Network Misclassification	15	1258.58	118.666	0.9250	0.0171
		20	1306.67	174.830		
6	**EnsembleNNReg Average**	15	1268.5	177.540	0.9248	0.0198
		20	1235.2	163.045		
	EnsembleNNReg Maximum	15	1226.7	100.937	0.9223	0.0103

5. Cross-validation

For the data partition, we decided to follow the 'rule of thumb'; having a sample size between 1000 and 10000 (our case is 1450 customers), it is recommended that the % training set should be 70%, the % validation set should be 30% and % test set should be 0%. We did a stratified partition to maintain the proportion of the target variable in the partition.

For a set of best 13 models, we tried five different random seeds in the data partition to check the potential idiosyncrasies of our samples. We calculated the average and standard deviation of the ROC index and the average and standard deviation of the profit to better compare and select the best final model. These results are presented in table 1. The *priori* probability of the target be 1 (response rate) is ~15%, for this reason, the best models should have a depth close to 15 to avoid having high number of false positives.

6. Best model

Most of the models improved when using optimal binning independently of the algorithm used, this decreases the levels of the interval variables and 'simplifies' our predictive model without losing the power of prediction.

The ROC curve of the validation set was used as an indicator to predict the most robust model. We are aware that is an optimistic measure comparatively when using a test set, although the small sample size did not justify the use of a test set.

Table 8 - Best model results of median and standard deviation profit and ROC for the five seed partitions (12345, 654321, 937162211, 1249821, 10270119). The chosen model for customer extraction was the one of the seed 937162211 highlighted in bold.

Model	**12345**	**654321**	**937162211**	**1249821**	**10270119**	**Median**	**Standard Deviation**
Profit	1087	1230	**1230**	1516	1472	1230	140.3093
ROC	0.909	0.915	**0.914**	0.941	0.946	0.915	0.01532

Considering the predicted profit and, once again, considering that it will be optimistically measured, we are expected to have a smaller profit that the one calculated theoretically due to the false positives. Although to reduce the number of false positives and be sure we do the best decision for the model taking in consideration the limitation of the size of the data size and the potential bias of the sample, we calculated the median and standard deviation of both ROC and profit for different five data partitions seeds (Table 8).

Five seeds are a good trade-off between time and accuracy to decide the best model. Considering these measures, the model we chose was neural network because it was the one that had the best compromise between the ROC curve and profit. This model seems to have been the best mostly due to the data preprocessing steps, taking into account that most of the models generated with these procedures obtained good results. Starting with variable transformation and taking advantage of the most information gain in the transformation, to the filter of 6 of the observations that could skew the model in the training set. But most importantly would be the variable selection that chooses a good equilibrium between the monetary aspects to the customer commitment to the organization, proving to be the best predictor model.

In relation to the depth of choice, we opted for 20 because it will contact more customers and therefore will have more people accepted the offer. A higher depth than 15 have a bigger risk of having a higher number of false positives, although this risk is not problematic in this case since the revenue is 11 and the cost is 3 by client. For this, we expect to have a positive and maximum net profit.

One thing to take into consideration, looking at the standard deviations of the profit, is that the best model chosen has some bias between different partitions, although it is not problematic and could not be reduced below 100. This could be a limitation of the data set itself and number of variables used in the prediction during the training set.

In this particular case we choose a depth of 20, although it could be riskier. Observing the pattern of the average and the standard deviation, we gain approximately 50$ if we increase the depth even though we risk to increase the standard deviation by 50$ as well, which actually would result in negating the risk and very likely increasing the profit (Figure 8).

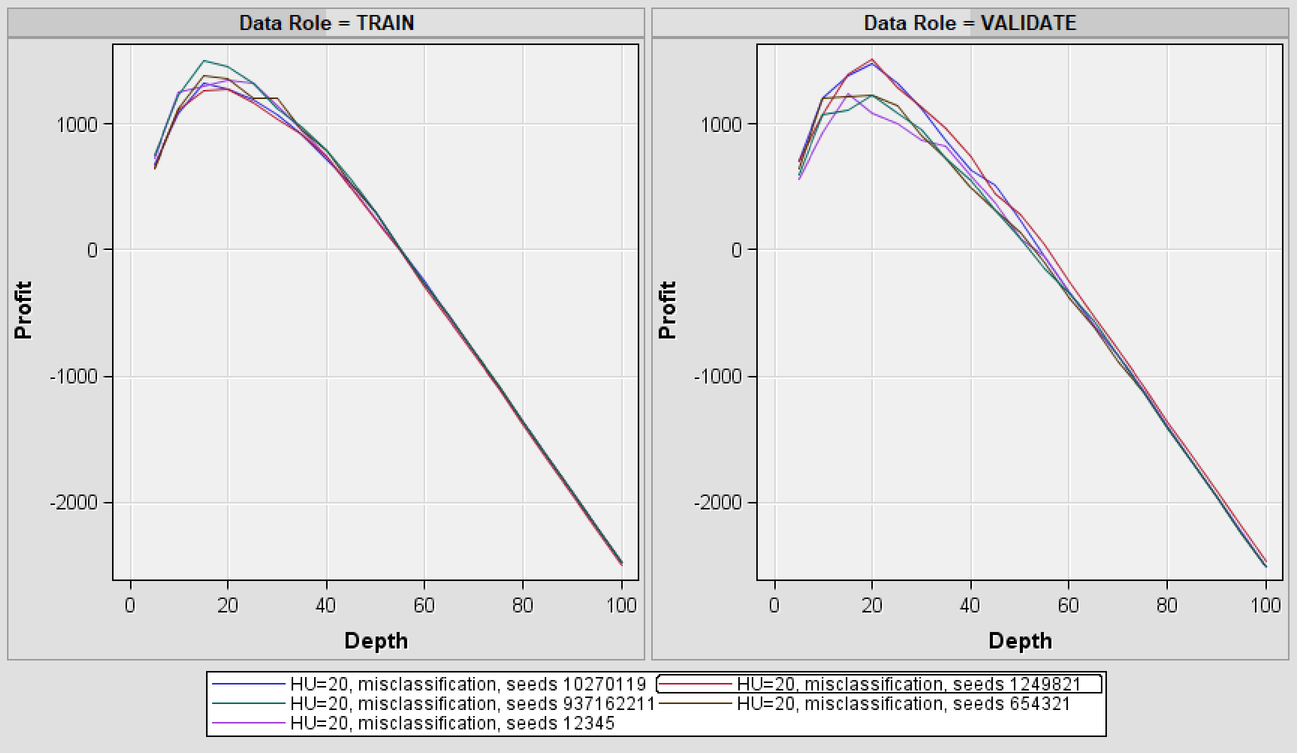

Figure 8 *– Best Model Profit for each partition seed considering different depths. The left graph is for the training set and the right graph is for the validate set.*

For the seed to be selected there were two choices, both with a profit of 1230$ and a ROC difference of 0,01. In the end the seed that was selected to extract the customer to be contacted in the next campaign was 937162211, taking into account our knowledge that across possible different partitions the best global depth would be 20, and this seed gives its best result at exactly that depth while the seed 654321 seems to present a greater deal of ambiguity and contradiction with our cross validated results (see Table 8). From the 1855 customers, we selected 371 customers and calculated the probability of accepting the next campaign. This probability ranged between 1 and 0.16.

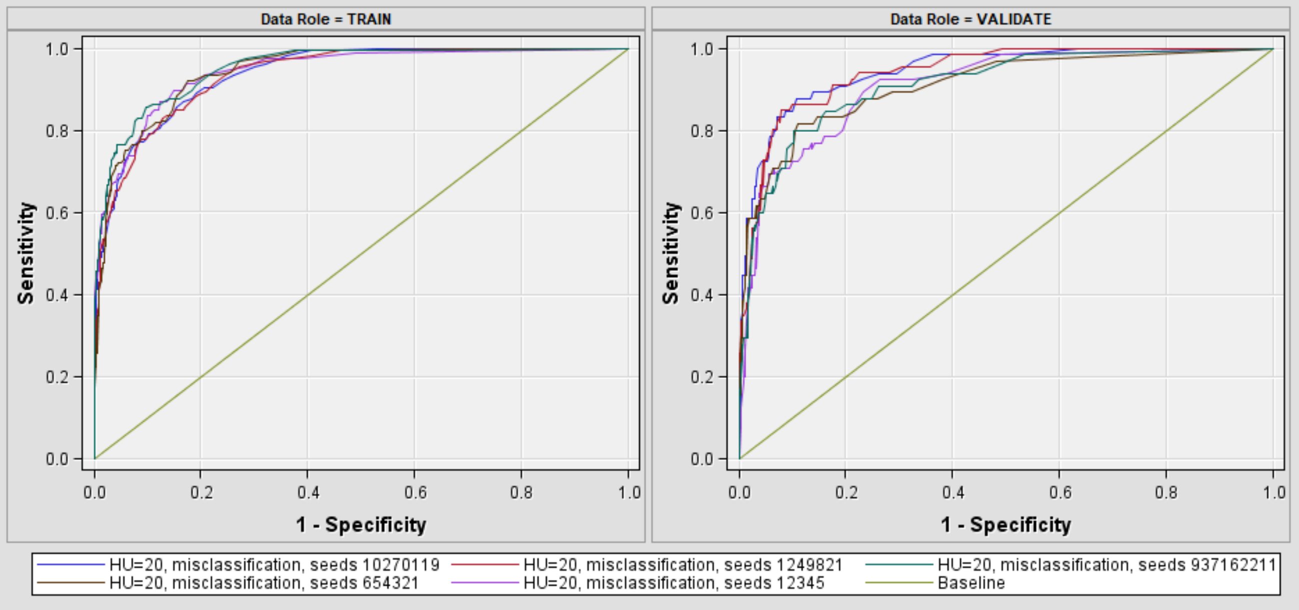

Figure 9 – *Best Model ROC Curve considering the five seeds for training set (left) and validate set (right).*

Business conclusions

Our goal was to predict the future marketing campaign based on past customer behavior. With this in mind, seems like the best predictors to understand if clients will buy the new gadget are related with the number of days since the last purchase and the money spent in the past in products. A good subcategory predictor would be the monetary spent on gold products, since they indicate top products specially designed for top clients. Besides that, the date measure seems like a relevant predictor as well, the date since they have been clients and date since the last purchase. In conclusion, we built a model based on these main features and a few specific ones that fit the problem, in a particular useful way, such as the number of catalog purchases or discount purchases percentage, in our quest to predict which clients will accept future campaigns.

Finally, Data science like so many other fields, falls into the same famous quote by Confucius 'Study the past, if you would divine the future'.