Paper 2242-2014

Researching individual credit rating models

F.P.N. (Colin) Nugteren MSc., Direct Pay Services B.V.

Summary:

In this research paper we try to find an optimal model for predicting credit risk in a consumer-factoring environment. This is done for DirectPay and uses their data and knowledge. Both the possibilities of finding more and better data and using more intelligent models are explored. In the end the effect of better and more relevant data has the largest impact on the forecasting performance, but using more intelligent models can create an extra advancement.

Introduction:

DirectPay is a part of the Credit Exchange Group, this group provides several credit management solutions for numerous companies. Webcasso is the collection part of the group and has a much longer history in credit management than DirectPay. Credit management consists of billing debtors of clients and ensuring that they receive all money that their debtors owe them. This can be done in several forms, as a factoring provider, a collection agency, a payment solution provider and by giving credit advice. For most of the clients DirectPay is a factoring provider. Factoring is a financial transaction whereby a company buys accounts receivable (invoices) from a client. After this transaction the buying company will try to do whatever is legally possible to get the money from the debtors. DirectPay is quite different from other factoring agencies; this is partly because of the Dutch legal system and partly because of their strategy. As a factoring agency DirectPay does not have to buy every invoice, it is allowed to choose which invoices are bought and which are not bought. It is also possible to advise a client not to sell someone a subscription or request a guarantee paid upfront. This second form is especially important for energy and telecom companies. A schematic view on factoring and the credit check that helps DirectPay to choose these invoices is shown in Diagram 1.

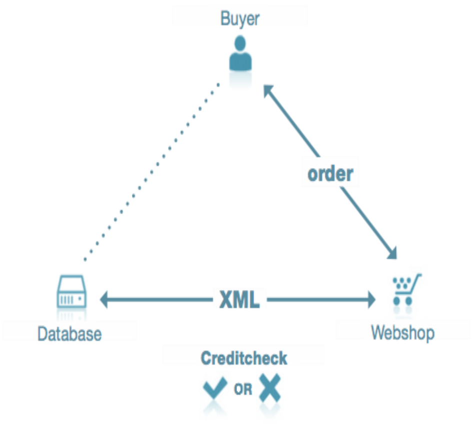

Diagram 1: A schematic view on factoring and the credit check in a webshop

At this moment DirectPay already uses a simple econometric model to estimate risk of non-payment. Based on recent research in the field of credit management and the growth of DirectPay in recent years it must be possible to find improvements for this model. The model DirectPay uses is described in the chapter "Point of origin" and provides an interesting benchmark for this paper. According to other researchers it might be possible to reach the following targets:

- Increase the maximum credit limit from 250 to 1000 euro
- Increase the acceptance level from 70% to 85%
- Keep equal risk or achieve lower risk

This research will describe the search for the best possible model to perform a customer credibility check in an online environment. This is also academically interesting, because most research on credit rating models is focused on bank loans and credit cards. Since this research will be used for a model on consumer behavior after buying something for a relatively small amount, there is a real difference. Both for credit cards and for a bank loan, the provider has much more freedom to ask for information on a future debtor. In a factoring environment only contact information is available in combination with statistical information on the household. Thus historic payment behavior is very interesting, since the most recent contact moment is often just a month ago.

The Dutch market is much harder to predict than foreign markets, because of the tight privacy laws and the lack of a single nationwide database. This provides credit managers with an important question: What criteria should be used to link one's credit history to future payment behavior? There are several levels at which we can identify a possible link. For example one can buy on the same account in a shop, which gives a relatively highly credible link. It is also possible to find a link based on an address, name and birth date, which is still quite credible, but links with lower credibility like address, email or phone number are also possible and might contribute to a stronger prediction.

Most credit information in the Netherlands is based on negative reports from collection agencies. In this research we will investigate whether such reports have added value. A negative report is at least 60 days old. Positive reports that are collected by DirectPay through registering paid invoices are not only more numerous, but also more recent. These negative reports are based on information from different credit management companies that use different strategies for collecting money. Using the own company database one is able to find all available information at the exact moment of the check, which provides a unique possibility to find the optimal combination of inputs. This is also an opportunity to look at the value of ad-hoc information that is delivered to the webshop compared to the value of static and often older information from collection agencies and data vendors.

Overall there is little information available about credit scoring in a consumer factoring environment, probably because of competitive decisions. Since the underlying data for most research is quite the same as the data of DirectPay, we expect that the outcomes and important models are alike too. Using all available information and portfolio knowledge will result in a specially designed model for DirectPay. This model will provide credit ratings for future and current debtors. The main question will be:

What is the best possible model to forecast future payment behavior in a consumer factoring environment?

The second research question is about the battle between the data and the models. Many research has focused on using other or more complicated models, since many researchers do not have the possibility of finding better data inside a database or increasing data quality at the client level. The data is considered a given thing by the company, the university or the IT department. The hypothesis is that putting interest and effort in the data, the type of data that is stored and the way the data is stored can provide much better results and that this is more influential than advances in modeling.

Credit data is often based on a negative 'mark' at a registered agency. The companies that build models on this information try to create a more or less 'static' score. In our case the target is to give scores to a person in a situation at a certain moment and that does not imply that this score is valid a day later in a different situation. The idea is that complete and recent information can provide better insights and thus that a single negative mark is less useful than knowing that a person owes DirectPay exactly 193 euro and 25 cents based on an invoice that was send 17 days ago from a well-known shop.

Point of Origin

When DirectPay was founded in 2006 it had only a small amount of data available based on earlier collection activities in Webcasso. So all decisions and credit checks were based on data bought from credit rating agencies combined with contractual agreements with clients like the amount of credit that was available for each individual. All data suppliers provided credit advices in different forms and the costs were quite high, but the advices were not tailored for DirectPay's specific needs.

In 2010 DirectPay developed a simple credit model based on logistic regression with the outcomes from the first few hundred thousand invoices. At this moment the dataset has grown with information about millions of invoices. This first model had been successfully active for over a year when we started this research.

One part of DirectPay's first model was based on statistical data about households and the second part on some information about the purchase. Using this model DirectPay was able to equal the scores from the bought-in credit checks. The finally used model included data about the house value, house code, income range, age, amount, shop id and the hour of purchase. This model resulted in a false negative rate of 0.537 and a false positive rate of 0.145, which resulted in a misclassification rate of 0.214.

Even though this misclassification rate is quite high this is even better than what was experienced by buying scores. Especially fraud was hard to detect given the tight privacy laws. In the live environment these scores proved soon to strongly reduce fraud, reduce non-payment and slightly increase overall acceptance.

Data Collection

The DirectPay database is in sheer size not really Big Data compared to huge international companies. Especially in the two main dimensions persons and transactions the numbers are not extremely large. There is information available for approximately one million individuals and four to five million transactions for the period from 2008 until 2012. This data is combined with an external database with statistical information about all 8 million Dutch households. Within this database a broad arrange of data is available about the persons that DirectPay has done business with. The data may not be very wide or high but it is very deep in all other dimensions and it is possible to look at the data from many points of view.

The hard part is to identify the individuals that are willing and able to pay within the regions where this is not the expected behaviour. In the map shown in diagram 2 all postal codes have a colour ranging from red, bad payment behaviour, to yellow, good payment behaviour. The cities Rotterdam, Amsterdam and The Hague are clearly marked in red. This map is drawn based on the same data set as used for the final models.

Based on personal and ad hoc information, like having an email address from an expensive Internet provider or buying something for a low amount, the final decision to accept the risk is made. It is unacceptable for a company like DirectPay to reject everyone that lives in a bad-scoring area. That would result in big losses in turnover and profit for the clients. That is why DirectPay gives individuals even from the worst areas a chance to prove themselves to us as a credit worthy person. When there is no payment information available the scores are just based on statistics at a street level, so every bit of information that is gathered is of the highest importance.

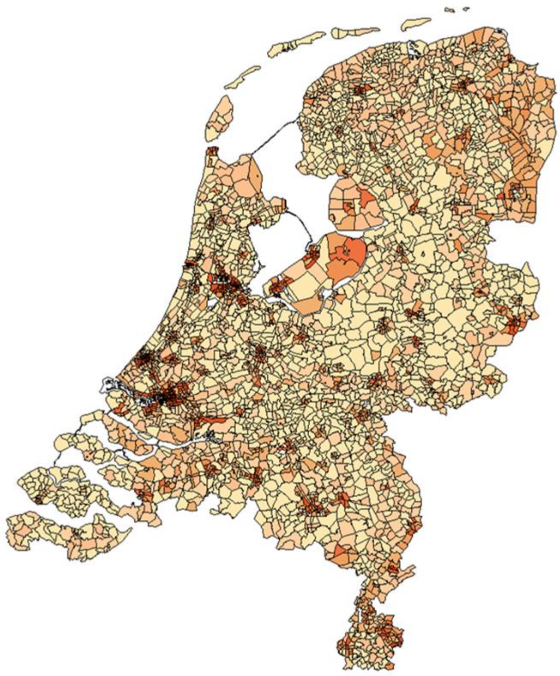

Diagram 2: Payment behaviour in the Netherlands

From all of the payment history data it is known whether the debtor was scored or not and whether he was accepted or not. To gather the data for this research we started to randomly accept applicants without scoring. This decision was made to provide this research the best possible dataset. Each application had the same chance to be marked as a 'test case' during a year. In total almost 75.000 test cases were selected by this process. Next to this project we searched for clients that did not want scoring on their clients because of marketing concerns and high margin products. Combining the results from both projects resulted in a unique dataset with almost 150.000 payment experiences.

The Dataset

The available data consist of contact information like postal code, a house number with a possible suffix, phone (mobile of fixed), e-mail (free or paid), date of birth and gender. Combining the postal code, house number and suffix provides statistical information based on a particular household. This consists of information on the house value, type of house (apartment, house with garden, houseboat etc.), age of the house, age group of the house owner, small office at home, children, education, social status, income and car. Third we have our own information on historical payment results of a particular household like the number of invoices, amount paid, payment time, invoice amount, debt amount and the transaction type (direct debit or account receivable). Transaction types are very important, because the risk differs strongly between these types. This is due to the fact that the debt will be automatically collected with direct debit, while to debtor has to wire transfer the money self with an account receivable.

An important part of this research is to find the optimal "level of knowledge". Thus based on which criteria does one's payment history provide the strongest predictions? One can imagine that the quality of the data that we receive is not optimal. Each client asks the same information differently and end users make mistakes and some even use false information on purpose. For example some o clients require birth dates and some do not and some combine the street name and the house number in one field. Especially married women are hard to match since they are free to use their maiden name and their husbands name and many have other initials than the first letter of their name.

There are three methods available for matching payment behavior. The first only uses data from the same account in the same shop the second uses all data available at an address and the third is in the middle of those two. The matches are found based on textual comparison using the Jaro-Winkler distance. The first level provides more comparable data while the other provides more accurate data. Adding this information multiplies the number of variables since the payment history can be considered at several "levels of knowledge". Because of data constraints we will consider the following three levels:

- same account
- ID: name, address and birth date or chamber of commerce number
- Address

The most important part of the data is the result and thus target variable. This variable is defined as not paid within three months. This decision is based on the Dutch legal system that requires a specific amount of effort done by a collecting agency before allowing a bailiff to use force to retrieve the money. For DirectPay it is important to retrieve as much money as possible before going to court to reduce costs. All data is shown in diagram 3 in the appendix.

All available data is randomly divided in three groups with 40% of the data in the train dataset, 30% for the validation dataset and 30% for the test dataset. This division helps to avoid overparameterisation, since all discriminatory power that scores well must be present in all three datasets. This method will prevent finding false correlations that might perform well in this dataset, but will appear to be wrong in real-time scoring.

Models

For all models y = 1 is defined as a paid invoice and y = 0 is defined as an unpaid invoice. In general the other variables in the equations are transformed into ranges or groups according to similarity in the proportion of good and bad results. It is debatable whether to group for example age in 3, 4 or 7 groups or use the age as a number. This decision can be made based on the log odds ratio (Crook. et al 2002). This process is known as 'coarse classification' and in large datasets like ours this results in dummy variables for the different groups. It also allows for treating missing values as informational values. For example not filling in a date of birth in a webshop might say something about the person's intentions. Applying this is essential in credit scoring for both categorical characteristics and continuous characteristics. For the first type to ensure that each attribute is available in the sample to make the analysis robust and for the second type to capture nonlinear effects. Based on this method 432 new grouped explanatory variables where created from the dataset. All models are computed in SAS® Enterprise Miner™, the Enterprise Miner™ Diagram can be found in Diagram 4 in the appendix. After the computation the final outcomes were compared using Excel and SAS® Visual Analytics.

In the comparison multiple models will be evaluated and compared to the original model:

- Logit Model
- Decision Tree
- Combining Multiple Logit Models
- Combining a Decision Tree and a Logit Model
- Partial Least Squares
- Gradient Boosting
- Neural Network

Comparing models
For comparing these models we need strong measures of discriminatory power in combination with the actual gains and losses that DirectPay would incur. The model must also be capable of answering slightly different questions with mostly the same set of data. This is important for telecom and energy clients. Finally we need a model that can compute scores within milliseconds on a server, because the decision in an online environment must be made in a split second. This does not include estimation, which can be done separately.

In the end scores are divided into groups by using a cut-off point. This way a confusion matrix can be made for all possible outcomes of the prediction $\hat{y}_i$ and the result y_i. Predicting only Good payments will result in a large loss for DirectPay, while only predicting Bad payments will result in loss of all their clients. The debtors that result in errors are the hardest to predict and a model that predicts less errors performs better in essence. When model predicts the debtors in both error types more distinct, this will results in a higher overall score.

Overall we lose approximately 40% of the transaction value in case of a type 1 error; in contrast the profits from good paying customers are approximately 10% of this value. Type 2 errors cause a minor loss, because of a loss of profit, dissatisfaction of clients and their customers and the call centre that handles information requests. The model with the highest value is the one that scores the best given this profit model. Based on different gains for correct predictions and different losses for type 1 and type 2 errors it is important to find the optimal model based on profit. The transaction value must be multiplied with the following values:

Prediction/ Outcome	Bad	Good
Bad	0	- 0.10
Good	- 0.4	0.10

Table 2: Profit and loss ratio's

These scores are difficult to compare against each other between different models on the same dataset, because the chosen cut-off level, has quite a lot of influence on the optimal outcome. Given this expected loss the optimal cut off level ρ can be computed that minimizes expected loss. The expected loss is minimized when a cut off level is chosen that gives both Bad predictions and Good predictions an equal chance compared to their respective profits:

$$-0.4 \cdot (1-\rho) + 0.1\rho = -0.1\rho + 0 \cdot (1-\rho)$$
$$\rho = \frac{2}{3}$$

Minimizing expected loss results in a cut off level of 2/3 so the results from these models can be compared easily by computing the total loss and profit for DirectPay.

Results
The parameter estimates for the standard Logistic model are presented in table 4 in the appendix. Based on these results, we learn a lot about the influential variables. In the end only 40 variables are used. For almost all variables that were available at all three levels the model chooses to use the Address level. In retrospect this could be expected because of the high correlation between those variables and the high correlation in payment behaviour within streets and areas within a city or village. Having more information about an address is always better than having less information. Even using information that is not necessarily about the same person will help in credit risk management.

Even though some of our findings are expected, this still proves that actual and clear findings about someone's payment history are of greater use than information from unclear sources like standard rating agencies. The useful results like the amount of open invoices at the time of the order or that our database holds nothing about a person provide a clear understanding of payment behaviour. Having more than 1 open invoice is a very bad sign where 1 open invoice can be a coincidence especially combined with the open amount and the total amount of all invoices at the address level. The impact of the variable that counts the number of debtors at an address and the variable that counts the number of shops that is visited from an address is negative. We believe that this is subject to constant change based on the companies that DirectPay works for, but in general we expect that people who buy a lot on invoice are more risky than people who pay up front.

Model name	Error rate test dataset	Profit
Original Model (OM)	0.214	€ 7,847.22
Grouped Logistic Regression (GR)	0.135	€ 26,515.87
Gradient Boosting (GB)	0.137	€ 22,814.47
Logistic Regression (LR)	0.138	€ 24,555.06
Decision Tree (DT)	0.138	€ 18,416.02
Partial Least Squares (PLS)	0.142	€ 24,775.35
Neural Network (NN)	0.156	€ 9,306.90
Decision Tree Regression (DTR)	0.166	€ - 2,761.99

Table 1: Error rates and profit for each model

Comparing the original model with the regression model in table 1 shows the impact of the better data that is available. A gain of 35 per cent is quite impressive in a credit-scoring environment. This result confirms that the standard logistic regression performs strong in many different settings. Within the newly tried models the grouped regression scores best in both profit and error rate. Apparently the difference in information between the unknown, positively known and negatively known customers is extremely relevant. Based on profit the results are also quite impressive, the amount of profit for DirectPay increases threefold. The profits that are computed based on the different models differ much more than expected. Apparently there is a large difference between models that score the higher valued invoices better. The grouped regression also performs best at this point. The results from the Decision tree regression are quite weak. We expect that this method can be optimized to perform better; this could be an interesting point for further research. The Partial least squares model performs better profit wise than expected, especially compared to the decision tree. The DTR performs extremely bad on final profit compared to the other models even though the error rate is not that much worse.

In the total prediction error the positive category scores best as shown in table 2, while the negative category scores worst. The unknown category scores in the middle even though that was not expected. We expected that unknown customers would be harder to predict, since there was less information available on those customers. In the end the negatively known customers provide no profit for DirectPay, it might be interesting to find a way to improve scoring for these three individual models.

Grouped Logistic Regression	Unknown	Negatively Known	Positively Known
Observations in test dataset	6294	2777	8792
False Negatives	0.399	0.310	0.516
False Positive	0.103	0.131	0.073
Error rate test dataset	0.152	0.212	0.099
Profit	€ 8990.44	€ - 409.83	€ 17935.26

Table 2: Results for the individual models of the grouped logistic regression model

For DirectPay the targets of this research were quite clear:

- Increase the maximum credit limit from 250 to 1000 euro
- Increase the acceptance level from 70% to 85%
- Keep equal risk or achieve lower risk

Al these points were combined during our research into the expected profit, but it is interesting for DirectPay to know whether these individual goals were reached too. In the dataset the maximum credit limit for an individual debtor is just below 1000 euro, but the average credit limit is 75 euro. In table 3 the acceptance rates for each order amount in the test dataset is shown.

Amount in €	**Denied**	**Approved**	**Acceptance Rate**
0-50	1133	10069	0.899
50-100	1004	3463	0.775
100-150	355	913	0.720
150-250	254	474	0.651
250-1000	69	129	0.652
Total	**2815**	**15048**	**0.842**

Table 3: The acceptance rate for each group of order amounts

The lowest levels show very high acceptance rates between 78% and 90%, but above 100 euro these levels fall down to 65% for the groups 150-250 euro and 250-1000 euro. This seems to prove that higher credit limits are possible for DirectPay, but the cases in these groups are quite low compared to the total number of observations. We suggest some extra tests and research for the levels above 250 euro. The average acceptance rate for the test dataset is 84.2% and this shows that the goal for the acceptance level was reached. The goal for equal or lower risk is also reached by this research, since both the scores for false negatives and false positives are much lower than in the original model. In the end the profit is much higher with this new model than before and we believe that these results are very useful for DirectPay.

Conclusion

At the beginning of this assay there are two questions formulated about credit rating models in a consumer factoring environment. The first question considers the search for the best possible forecasting model and the second question compares the influence of better data compared to the influence of better models. Based on a dataset with more than 100.000 observations and 80 variables 8 models were tested to find the optimal forecast. These models were also compared to a baseline model that is based on a Logistic Regression with much less available explanatory variables. This baseline model provides insight in the extra forecasting performance that the newly found variables provided to the models.

During this research it became clear that various models perform very well. This is obviously good for business, but from an econometric perspective this could be quite problematic. In the end four models perform comparable in terms of profit and misclassification rates. Given the model we started to work with and the improvements we have made during this research it appears that the influence of more advanced models is modest compared to the influence of better and more advanced data. The knowledge about a customer or the lack thereof is highly important. This is not a surprise but it does provide a lot of interesting things for future research, since it is not very profitable to keep buying invoices on people that are not expected to pay the invoices. For DirectPay this research provides a method to improve both acceptance rates and profit.

References

- Baesens B., Egmont-Petersen M., Castelo R. and Vanthienen J., Learning Bayesian network classifiers for credit scoring using Markov Chain Monte Carlo search, Proceedings of the 16th International Conference on Pattern Recognition 3, pp 49-52 (2002)
- Baesens B., Gestel T. van, Viaene S., Stepanova, M., Suykens J. and Vanthienen J., Benchmarking state-of-the-art classification algorithms for credit scoring, Journal of Operational Research Society 54-6, pp. 627-635 (2003)
- Bishop, C.: Neural Networks for Pattern Recognition, Oxford University Press, Oxford (1995)
- Breiman, L.: Arcing Classifiers, The annals of statistics 26-3, pp. 801-849 (1998)
- Breiman, L.: Random Forests, Machine Learning 45-1, pp. 5-32 (2001)
- Breiman, L., Friedman, J., Stone, C. and Olshen, R.: Classification and Regression Trees, Wadsworth, Belmont (1984), since 1993 this book has been published by Chapman & Hall, New York
- Crook J., Edelman D. and Thomas L., Recent developments in consumer credit risk assessment, European Journal of Operational Research 183-3, pp. 1447-1465 (2007)
- Friedman, J. H., Stochastic Gradient Boosting, Computational Statistics & Data Analysis 38-4, pp. 367-378 (2002)
- Hastie T., Tibshirani R. and Friedman J.: The elements of statistical learning (2nd ed.), Springer, New York (2009)
- Jacobsen T. and Roszbach K., Bank lending policy, credit scoring and Value at Risk, Journal of Banking and Finance 27-4, pp. 615-533 (2003)
- Martin, D.: Early warning of bank failure: A logit regression approach, Journal of Banking & Finance 1-3, pp. 249–276 (1977)
- Ripley, B.: Pattern Recognition and Neural Networks, Cambridge University Press, Cambridge (1996)
- Ryser M. and Denzler S., Selecting credit rating models: a cross-validation-based comparison of discriminatory power, Financial Markets and Portfolio Management 23-2, pp. 187-203 (2009)
- Thomas L., Edelman D. and Crook J., Credit Scoring and Its Applications, SIAM, Philadelphia (2002)
- Tin Kam Ho, Random Decision Forests, Proceedings of the 3rd International Conference on Document Analysis and Recognition 1, pp. 278-282 (1995)
- West D., Neural network credit scoring models, Computers & Operations Research 27 11-12, pp. 1131-1152 (2000)
- Zhang J. and Härdle W., The Bayesian Additive Classification Tree applied to credit risk modelling, Computational Statistics and Data Analysis 54-5, pp. 1197-1205 (2010)

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.

Appendix

Diagram 3: All available data grouped by source and type

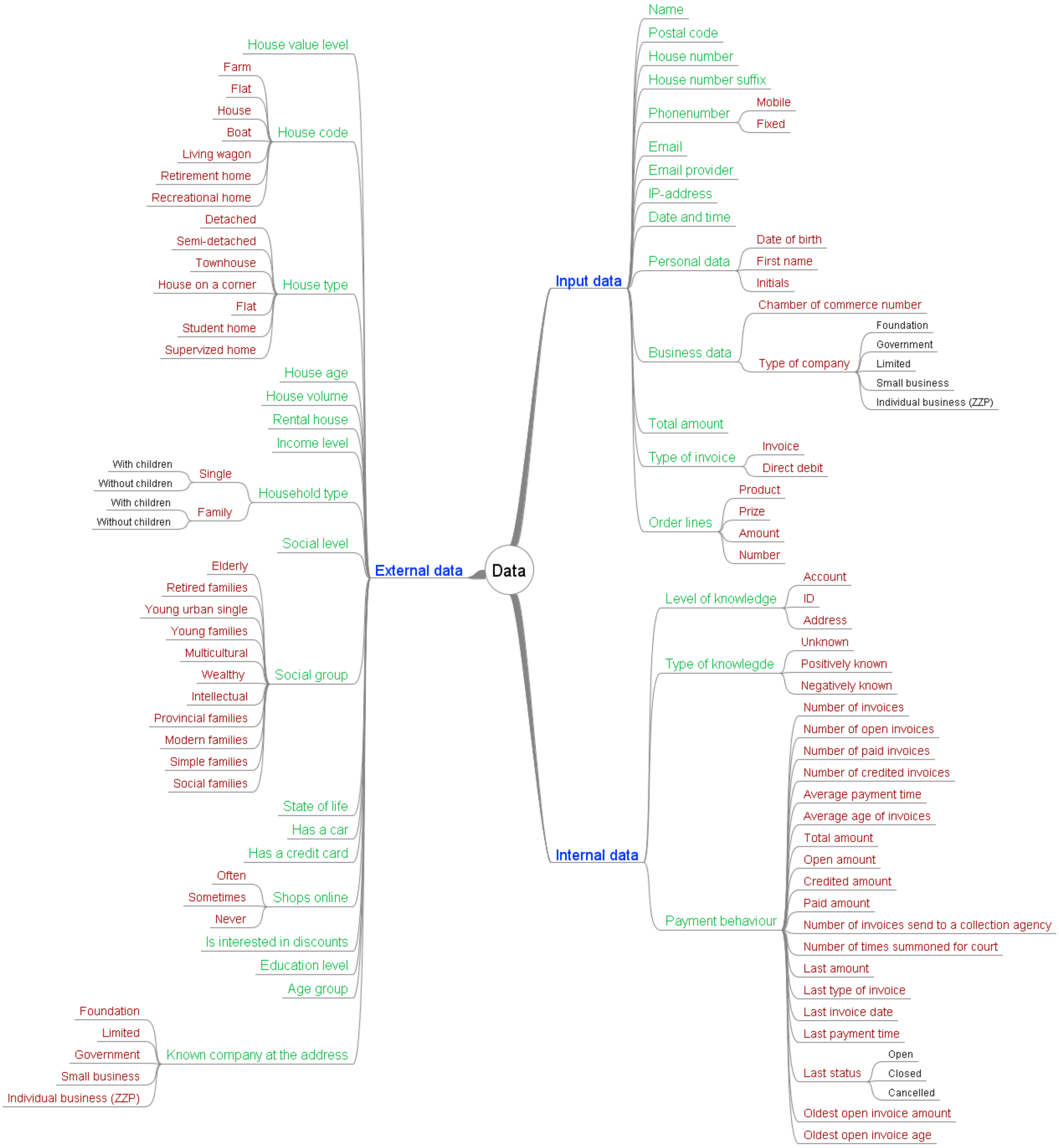

Diagram 4: The SAS® Enterprise Miner™ diagram

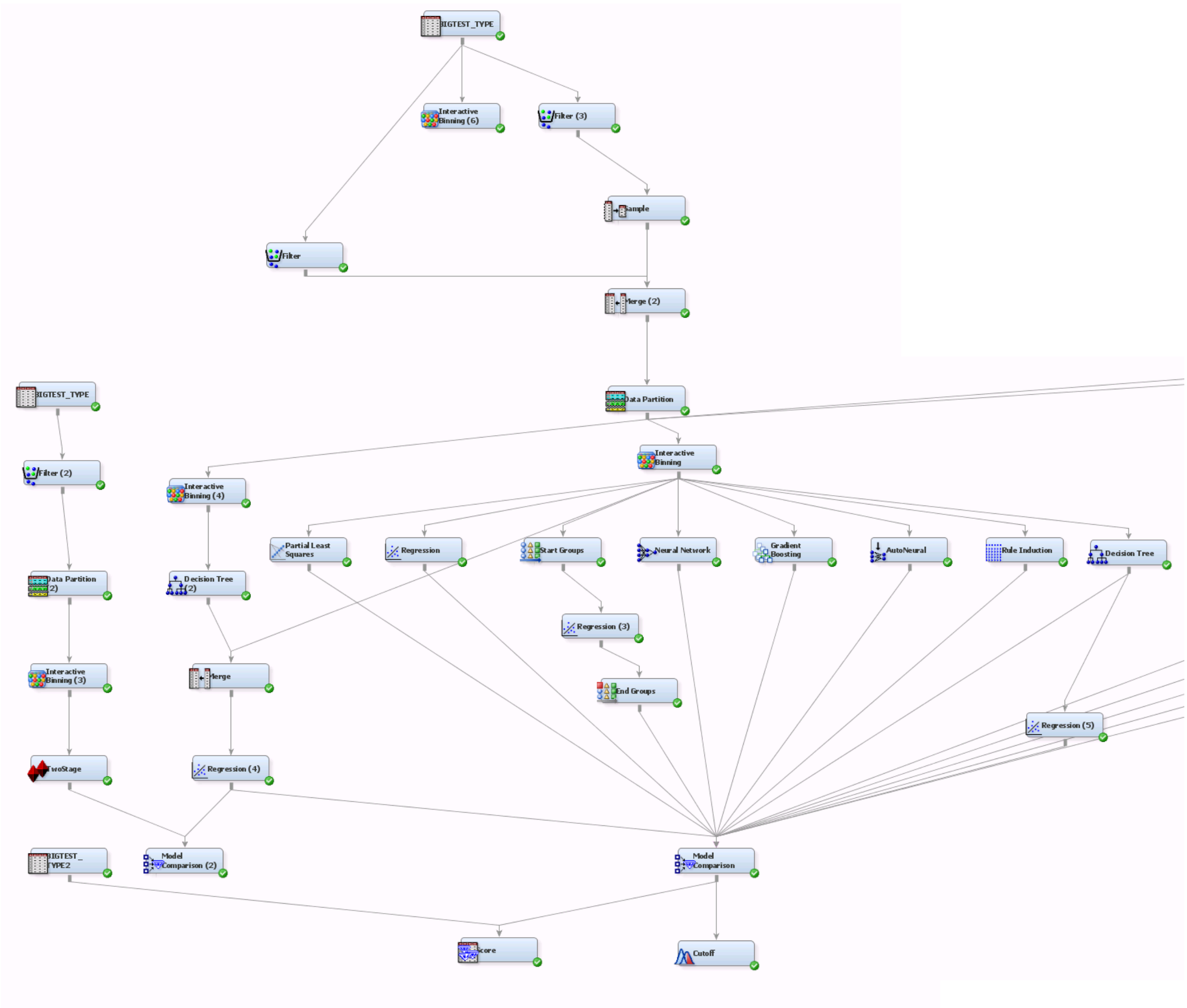

Table 4: Results Standard Logit Regression Model

Name	Coefficient	T-Value	P-Value	Description
GRP_bedragopen2	0,744	3,086	0,002	0<= bedragopen< 1
GRP_bedragopen3	0,405	3,149	0,002	1<= bedragopen< 50
GRP_bedragopen4	0,535	4,619	0,000	50<= bedragopen< 150
GRP_bedragtotaal3	-0,477	-4,306	0,000	0<= bedragtotaal< 35.9
GRP_bedragtotaal4	-0,393	-4,762	0,000	35.9<= bedragtotaal< 157.85
GRP_betaalsnelheid2	0,833	9,286	0,000	betaalsnelheid< 20
GRP_cnt_debiteur3	0,677	7,259	0,000	1<= cnt_debiteur< 2
GRP_cnt_debiteur4	0,500	7,262	0,000	2<= cnt_debiteur< 3
GRP_cnt_entiteit2	1,417	13,147	0,000	1
GRP_cnt_entiteit3	0,299	3,550	0,000	2
GRP_cnt_entiteit4	-0,206	-2,046	0,041	10, 11, 12, 18, 4, 5, 6, 7, 8, 9
GRP_Email1	-0,620	-6,770	0,000	Missing
GRP_Email2	-0,477	-8,687	0,000	GMAIL.COM, HOTMAIL.COM
GRP_Email3	-0,926	-11,701	0,000	LIVE.NL
GRP_entiteit_id2	0,850	9,744	0,000	1084, 1096, 1121, 50012, 50019, 50020, 50036
GRP_entiteit_id3	0,307	3,623	0,000	1104, 50000, 50001, 50006
GRP_entiteit_id4	0,933	9,700	0,000	1040, 1088, 1112, 50003, 50011, 50023
GRP_Geboortejaar1	0,896	6,386	0,000	1899, 1900, 1901, Missing
GRP_Geboortejaar2	0,689	5,116	0,000	42 and older
GRP_Geboortejaar3	0,461	3,242	0,001	35-42 years old
GRP_Geboortejaar4	0,290	2,166	0,030	23-35 years old
GRP_IsBedrijf_NeeJa2	-1,128	-6,775	0,000	0
GRP_last_open2	0,499	5,240	0,000	0
GRP_last_paidtime2	-0,266	-2,259	0,024	last_paidtime< 1
GRP_last_paidtime3	0,442	4,642	0,000	1<= last_paidtime< 16
GRP_Last_same_shop1	-1,122	-6,672	0,000	Missing
GRP_Last_same_shop2	-0,576	-7,332	0,000	1
GRP_last_time2	0,305	3,418	0,001	last_time< 14
GRP_MXCTHT2	0,379	6,045	0,000	10, 11, 5, 7, 8, 9 (Goed betalende burgers)
GRP_MXkoophuur2	0,289	5,246	0,000	1 (Koop)
GRP_MXwoz2	-0,580	-8,488	0,000	1, 2 (Laag)
GRP_MXwoz3	-0,307	-5,824	0,000	3, 4, 5, 6 (Middel)
GRP_Totaal_in_btw2	0,894	12,812	0,000	Totaal_in_btw< 60
GRP_Totaal_in_btw3	0,299	4,100	0,000	60<= Totaal_in_btw< 125
GRP_xafgehandeld2	0,603	2,752	0,006	0
GRP_xafgehandeld3	1,103	4,324	0,000	1
GRP_xbetaald3	-0,619	-7,290	0,000	1<= xbetaald< 5
GRP_xfraude2	2,206	10,317	0,000	xfraude< 1
GRP_xnogopen2	0,762	3,087	0,002	xnogopen< 1
GRP_xnogopen3	0,690	6,702	0,000	1<= xnogopen< 2
Intercept	-3,585	-9,841	0,000	