Statistics CA1

David Regan

9 April 2019

# Can we improve default prediction in the P2P banking sector?

# A test case using the Prosper Loan Dataset

# 1. Introduction

Definition: P2P Banking

Peer-to-peer banking is an online system which allows registered members to complete one to one or individual to individual loan transactions. Borrowers and lenders interact directly with one another, bypassing traditional financial institutions. P2P banking most often involves an auction process (Wang, Greiner & Aronson, 2009). Members can offer or request loans for a specific amount and at a specific rate. Block chain technologies are used to store loan contract listings and profiles. Borrowers can be assessed based on various profile statistics such as Credit Score or previous P2P performance. Lenders assess investments based on multiple factors for each individual. Borrowers are attracted by interest rates which are often significantly lower than retail bank offerings. There are many different P2P banking models on the market and hundreds of different P2P banking companies (Malekipirbazar & Aksakalli, 2015).

Credit Default Prediction

Credit default prediction is most often carried out by large financial institutions. Various methods are applied including credit scoring and grading, both for individuals and institutions. Statistical and Machine Learning Models such as Logistic Regression, SVM and Neural Networks can be also used. In P2P banking lenders can see a range of variables for each loan and individual. These factors can be used to construct effective credit-default classifier models. Factors include credit scores and grades which are calculated by the P2P ban and additional factors such as reason for loan, borrower state or district and borrower income and debt levels (Serrano-Cinca, Gutiérrez-Nieto & López-Palacios, 2015). In this report data from Prosper Loans an American P2P institution was used to create a credit default classifier.The data is described in the next section. The model constucted is a logistic regression classifier. Prosper Loans Dataset for The Prosper Loans Dataset contains approximately 120,000 instances and 40 features related to default prediction. A study was carried out using a dataset form Prosper Loans a leading American P2P institution. A Logisitic Regression Model was built to carry out default classification i.e. will a particular loan listing leading to default, yes or no. As such a target variable with a binary classification was needed.

# 2. Presentation of Dataset

Column Names There are 81 columns in the dataset and 113937 rows/instances.

Column Names

## Columns/Featues Overview

## Total number of columns 81

## Total number of rows 113937

## [1] "ListingKey"   
## [2] "ListingNumber"   
## [3] "ListingCreationDate"   
## [4] "CreditGrade"   
## [5] "Term"   
## [6] "LoanStatus"   
## [7] "ClosedDate"   
## [8] "BorrowerAPR"   
## [9] "BorrowerRate"   
## [10] "LenderYield"   
## [11] "EstimatedEffectiveYield"   
## [12] "EstimatedLoss"   
## [13] "EstimatedReturn"   
## [14] "ProsperRating..numeric."   
## [15] "ProsperRating..Alpha."   
## [16] "ProsperScore"   
## [17] "ListingCategory..numeric."   
## [18] "BorrowerState"   
## [19] "Occupation"   
## [20] "EmploymentStatus"   
## [21] "EmploymentStatusDuration"   
## [22] "IsBorrowerHomeowner"   
## [23] "CurrentlyInGroup"   
## [24] "GroupKey"   
## [25] "DateCreditPulled"   
## [26] "CreditScoreRangeLower"   
## [27] "CreditScoreRangeUpper"   
## [28] "FirstRecordedCreditLine"   
## [29] "CurrentCreditLines"   
## [30] "OpenCreditLines"   
## [31] "TotalCreditLinespast7years"   
## [32] "OpenRevolvingAccounts"   
## [33] "OpenRevolvingMonthlyPayment"   
## [34] "InquiriesLast6Months"   
## [35] "TotalInquiries"   
## [36] "CurrentDelinquencies"   
## [37] "AmountDelinquent"   
## [38] "DelinquenciesLast7Years"   
## [39] "PublicRecordsLast10Years"   
## [40] "PublicRecordsLast12Months"   
## [41] "RevolvingCreditBalance"   
## [42] "BankcardUtilization"   
## [43] "AvailableBankcardCredit"   
## [44] "TotalTrades"   
## [45] "TradesNeverDelinquent..percentage."   
## [46] "TradesOpenedLast6Months"   
## [47] "DebtToIncomeRatio"   
## [48] "IncomeRange"   
## [49] "IncomeVerifiable"   
## [50] "StatedMonthlyIncome"   
## [51] "LoanKey"   
## [52] "TotalProsperLoans"   
## [53] "TotalProsperPaymentsBilled"   
## [54] "OnTimeProsperPayments"   
## [55] "ProsperPaymentsLessThanOneMonthLate"  
## [56] "ProsperPaymentsOneMonthPlusLate"   
## [57] "ProsperPrincipalBorrowed"   
## [58] "ProsperPrincipalOutstanding"   
## [59] "ScorexChangeAtTimeOfListing"   
## [60] "LoanCurrentDaysDelinquent"   
## [61] "LoanFirstDefaultedCycleNumber"   
## [62] "LoanMonthsSinceOrigination"   
## [63] "LoanNumber"   
## [64] "LoanOriginalAmount"   
## [65] "LoanOriginationDate"   
## [66] "LoanOriginationQuarter"   
## [67] "MemberKey"   
## [68] "MonthlyLoanPayment"   
## [69] "LP\_CustomerPayments"   
## [70] "LP\_CustomerPrincipalPayments"   
## [71] "LP\_InterestandFees"   
## [72] "LP\_ServiceFees"   
## [73] "LP\_CollectionFees"   
## [74] "LP\_GrossPrincipalLoss"   
## [75] "LP\_NetPrincipalLoss"   
## [76] "LP\_NonPrincipalRecoverypayments"   
## [77] "PercentFunded"   
## [78] "Recommendations"   
## [79] "InvestmentFromFriendsCount"   
## [80] "InvestmentFromFriendsAmount"   
## [81] "Investors"

## The dataset has lots of missing values.

# Pre-processing to reshape dataframe

This is necessary before proceeding to visualization and modelling. This includes reshaping the data and filling null cells with mean values. It was also necessary that each feature be assessed and judged as to whether its inclusion is appropriate. Some features were questionable as there meaning was difficult to ascertain. This is discussed further in the next section.

# Delete Useless and Invalid Columns

For many features such as “Member” or “LoanKey” it was obvious they should be deleted. However, it was difficult to find a perfect explanation for all the features in the dataset. The following webpage acted as the main guide <https://www.prosper.com/Downloads/Services/Documentation/ProsperDataExport_Details.html> Columns which were not deemed appropriate as predictors of default were deleted. To be more precise only columns which definitely belong to the original borrower loan listing were included. However, the guide used is old and many deleted features such as “Investors” are likely to also be valid, nontheless i deleted them as i could not be certain.

## The following columns/features were deleted:

##   
## IncomeVerifiable  
## OpenRevolvingAccounts  
## OpenRevolvingMonthlyPayment  
## InquiriesLast6Months  
## RevolvingCreditBalance  
## LoanKey  
## TotalProsperLoans  
## TotalProsperPaymentsBilled  
## OnTimeProsperPayments  
## ProsperPaymentsLessThanOneMonthLate  
## BorrowerState  
## Occupation  
## Investors  
## ScorexChangeAtTimeOfListing  
## Recommendations  
## InvestmentFromFriendsCount  
## TotalInquiries  
## InvestmentFromFriendsAmount  
## InvestmentFromFriendsCount  
## LoanOriginationQuarter  
## GroupKey  
## LoanOriginationDate  
## LP\_GrossPrincipalLoss  
## LP\_NetPrincipalLoss  
## ListingKey  
## ListingNumber  
## ListingCreationDate  
## ClosedDate  
## ListingCategory (numeric)  
## CurrentlyInGroup  
## GroupKey  
## DateCreditPulled  
## OpenCreditLines  
## EmploymentStatus  
## CurrentCreditLines  
## ProsperPaymentsOneMonthPlusLate  
## ProsperPrincipalBorrowed  
## ProsperPrincipalOutstanding  
## LoanFirstDefaultedCycleNumber  
## LoanMonthsSinceOrigination  
## LoanNumber  
## MemberKey  
## LP\_CustomerPayments  
## LP\_CustomerPrincipalPayments  
## LP\_InterestandFees  
## LP\_ServiceFees  
## PercentFunded  
## LP\_NonPrincipalRecoverypayments  
## LP\_CollectionFees  
## LoanCurrentDaysDelinquent  
## BorrowerRate  
## LenderYield  
## EstimatedEffectiveYield  
## EstimatedLoss  
## EstimatedReturn  
## FirstRecordedCreditLine

##   
## Number of columns reduced to 28

# Delete Useless Rows

There are lots of useless instance/rows. Rows were the loan term is not finished. These will not help predict loan default. These rows will be deleted. Only rows were LoanStatus is either “Completed”,“Defaulted” or “Chargedoff”(a version of default) will be retained. Therefore, we will have only instances of loan repayment completion or default in the dataset to be modelled.

## Initially Column LoanStatus contained 12 different values:

##   
## Completed  
## Current  
## Past Due (1-15 days)  
## Defaulted  
## Chargedoff  
## Past Due (16-30 days)  
## Cancelled  
## Past Due (61-90 days)  
## Past Due (31-60 days)  
## Past Due (91-120 days)  
## FinalPaymentInProgress  
## Past Due (>120 days)

## Original number of rows = 113937

## Column LoanStatus now contains 12 different values:

## Completed   
## Current   
## Past Due (1-15 days)   
## Defaulted   
## Chargedoff   
## Past Due (16-30 days)   
## Cancelled   
## Past Due (61-90 days)   
## Past Due (31-60 days)   
## Past Due (91-120 days)   
## FinalPaymentInProgress   
## Past Due (>120 days)

## Number of rows reduced to 113937

# Create Target(y) Variable Defaulted

This involves a mapping of values from the “LoanStatus” feature. If “LoanStatus” contains the values “Defaulted” or “Chargedoff” the new target variable will be set as 1. If “LoanStatus” contains the value “Completed” the new target variable will be set to 0. Once this is done the loanStatus column will be deleted.

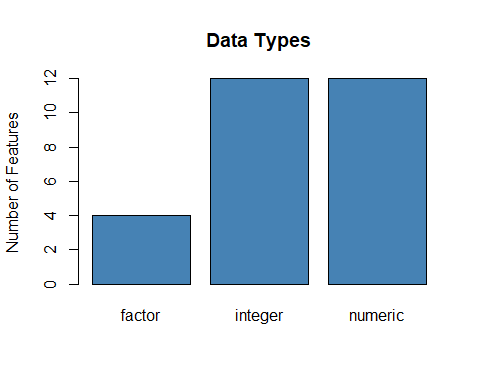
## New target variable Defaulted created

## Target variable Defaulted is binary. Values 0 or 1

## 0 1

## Column LoanStatus deleted.  
## Number of columns = 28

# Column Datatypes

 Most datatypes are numeric with only 4 categorical features remaining.

# Numerics

Missing values in the numeric columns were replaced with the column average.

# Categoricals: the levels are wrong

Some of the categorical variables stored as factors had incorrect levels.

## [1] "Incorrect Levels: CreditGrade, ProsperRating..Alpha and IncomeRange"

## [1] "" "A" "AA" "B" "C" "D" "E" "HR" "NC"

## [1] "" "A" "AA" "B" "C" "D" "E" "HR"

## [1] "$0" "$1-24,999" "$100,000+" "$25,000-49,999"  
## [5] "$50,000-74,999" "$75,000-99,999" "Not displayed" "Not employed"

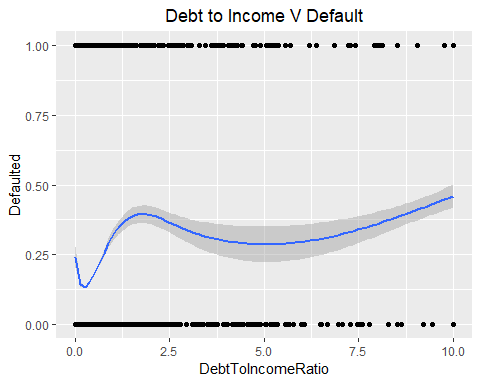
These were changed to the correct levels.

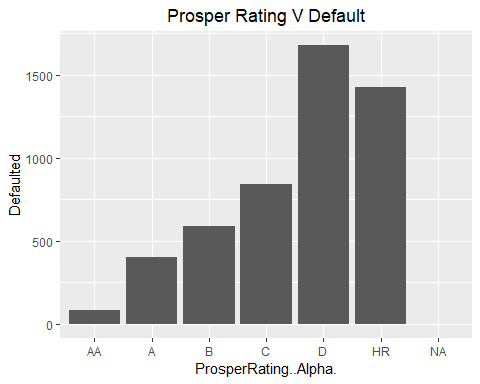
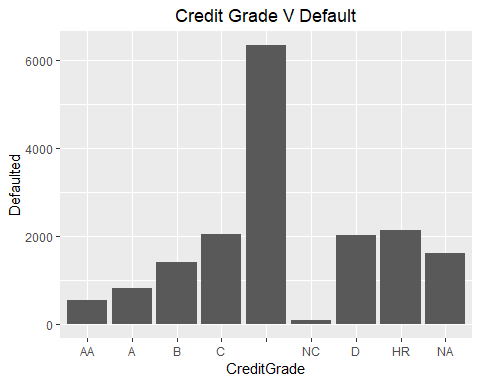
## [1] "AA" "A" "B" "C" "" "NC" "D" "HR"

## [1] "AA" "A" "B" "C" "" "D" "HR"

## [1] "Not employed" "Not displayed" "$0" "$1-24,999"   
## [5] "$100,000+" "$25,000-49,999" "$50,000-74,999" "$75,000-99,999"

Debt to income ratio is calculated as an individuals debt level proportional to their stated income. The figure below shows a steep incline as the ratio goes up initially. It then flattens out. However the majority of instances lie in the first or left section of the figure, with debt to income levels below 4. The right-side reduces the impact of the chart. Adjusting the scales would help.

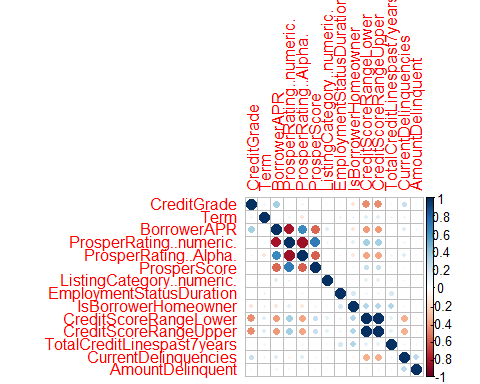


Prosper rating alpha is a categorical classifier created by Propser Loans. It is based on previous proper activity and some of the more obvious profiling features such as credit score and the very similar credit grade feature seen in the next chart.  

There are still some missing values in the categorical features. These will be converted to numeric variables using the factor levels. The missing values can then be assigned the column mean. The predictive variables are now ready to be used in a Logisitic Regression Model.

## CreditGrade Term BorrowerAPR ProsperRating..numeric.  
## Min. :1.000 Min. :12.00 Min. :0.00653 Min. :1.000   
## 1st Qu.:5.000 1st Qu.:36.00 1st Qu.:0.15629 1st Qu.:3.000   
## Median :5.000 Median :36.00 Median :0.20979 Median :4.072   
## Mean :4.842 Mean :40.83 Mean :0.21883 Mean :4.072   
## 3rd Qu.:5.000 3rd Qu.:36.00 3rd Qu.:0.28370 3rd Qu.:5.000   
## Max. :8.000 Max. :60.00 Max. :0.51229 Max. :7.000   
## ProsperRating..Alpha. ProsperScore ListingCategory..numeric.  
## Min. :1.000 Min. : 1.00 Min. : 0.000   
## 1st Qu.:3.000 1st Qu.: 5.00 1st Qu.: 1.000   
## Median :4.169 Median : 5.95 Median : 1.000   
## Mean :4.169 Mean : 5.95 Mean : 2.774   
## 3rd Qu.:5.000 3rd Qu.: 7.00 3rd Qu.: 3.000   
## Max. :7.000 Max. :11.00 Max. :20.000   
## EmploymentStatusDuration IsBorrowerHomeowner CreditScoreRangeLower  
## Min. : 0.00 Min. :1.000 Min. : 0.0   
## 1st Qu.: 28.00 1st Qu.:1.000 1st Qu.:660.0   
## Median : 74.00 Median :2.000 Median :680.0   
## Mean : 96.07 Mean :1.504 Mean :685.6   
## 3rd Qu.:130.00 3rd Qu.:2.000 3rd Qu.:720.0   
## Max. :755.00 Max. :2.000 Max. :880.0   
## CreditScoreRangeUpper TotalCreditLinespast7years CurrentDelinquencies  
## Min. : 19.0 Min. : 2.00 Min. : 0.000   
## 1st Qu.:679.0 1st Qu.: 17.00 1st Qu.: 0.000   
## Median :699.0 Median : 25.00 Median : 0.000   
## Mean :704.6 Mean : 26.75 Mean : 0.592   
## 3rd Qu.:739.0 3rd Qu.: 35.00 3rd Qu.: 0.000   
## Max. :899.0 Max. :136.00 Max. :83.000   
## AmountDelinquent DelinquenciesLast7Years PublicRecordsLast10Years  
## Min. : 0.0 Min. : 0.000 Min. : 0.0000   
## 1st Qu.: 0.0 1st Qu.: 0.000 1st Qu.: 0.0000   
## Median : 0.0 Median : 0.000 Median : 0.0000   
## Mean : 984.5 Mean : 4.155 Mean : 0.3126   
## 3rd Qu.: 0.0 3rd Qu.: 3.000 3rd Qu.: 0.0000   
## Max. :463881.0 Max. :99.000 Max. :38.0000   
## PublicRecordsLast12Months BankcardUtilization AvailableBankcardCredit  
## Min. : 0.00000 Min. :0.0000 Min. : 0   
## 1st Qu.: 0.00000 1st Qu.:0.3300 1st Qu.: 1017   
## Median : 0.00000 Median :0.5613 Median : 4870   
## Mean : 0.01509 Mean :0.5613 Mean : 11210   
## 3rd Qu.: 0.00000 3rd Qu.:0.8200 3rd Qu.: 12078   
## Max. :20.00000 Max. :5.9500 Max. :646285   
## TotalTrades TradesNeverDelinquent..percentage.  
## Min. : 0.00 Min. :0.0000   
## 1st Qu.: 15.00 1st Qu.:0.8300   
## Median : 23.00 Median :0.9300   
## Mean : 23.23 Mean :0.8859   
## 3rd Qu.: 29.00 3rd Qu.:1.0000   
## Max. :126.00 Max. :1.0000   
## TradesOpenedLast6Months DebtToIncomeRatio IncomeRange   
## Min. : 0.0000 Min. : 0.0000 Min. :1.000   
## 1st Qu.: 0.0000 1st Qu.: 0.1500 1st Qu.:5.000   
## Median : 0.8023 Median : 0.2300 Median :6.000   
## Mean : 0.8023 Mean : 0.2759 Mean :5.966   
## 3rd Qu.: 1.0000 3rd Qu.: 0.3100 3rd Qu.:7.000   
## Max. :20.0000 Max. :10.0100 Max. :8.000   
## StatedMonthlyIncome LoanOriginalAmount MonthlyLoanPayment  
## Min. : 0 Min. : 1000 Min. : 0.0   
## 1st Qu.: 3200 1st Qu.: 4000 1st Qu.: 131.6   
## Median : 4667 Median : 6500 Median : 217.7   
## Mean : 5608 Mean : 8337 Mean : 272.5   
## 3rd Qu.: 6825 3rd Qu.:12000 3rd Qu.: 371.6   
## Max. :1750003 Max. :35000 Max. :2251.5   
## Defaulted   
## Min. :0.0000   
## 1st Qu.:0.0000   
## Median :0.0000   
## Mean :0.1493   
## 3rd Qu.:0.0000   
## Max. :1.0000

# Vizualisations of Correlation



The chart above is a correlation matrix for 14 of the predictor variables. There are a couple of clusters of highly correlated features. Some of these can be dropped from the Logistic Regression model, this is discussed further below. Their co-linearity makes them redundant. Prosper Rating Numeric, Prosper Rating Alpha, Prosper Score and Borrower APR are clearly heavily correlated.

# LOGISTIC REGRESSION

First, the data must be split into a training and test set. An 80/20 split was used. 80% for training purposes and 20% to test the models accuracy on unseen data. The goal is to create a model or models which generalize to unseen data.

The dataset is unbalanced approximately 69-70% of the instances are completed, the rest being defaults. Therefore, the training data must be adjusted to 50/50 between default and completed instances in order to create a useful Logistic regression model.

##   
## Call:  
## glm(formula = Defaulted ~ . - Defaulted, family = binomial(link = "logit"),   
## data = train\_set)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.4497 -0.8682 -0.0048 0.9052 4.3117   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 1.199e+00 3.879e-01 3.091 0.001994  
## CreditGrade -2.735e-01 1.491e-02 -18.344 < 2e-16  
## Term -1.116e-02 2.354e-03 -4.741 2.12e-06  
## BorrowerAPR 8.255e+00 3.777e-01 21.854 < 2e-16  
## ProsperRating..numeric. 4.314e-01 3.005e-02 14.357 < 2e-16  
## ProsperRating..Alpha. 5.558e-01 1.906e-02 29.164 < 2e-16  
## ProsperScore 2.161e-01 1.160e-02 18.629 < 2e-16  
## ListingCategory..numeric. -2.175e-02 4.099e-03 -5.306 1.12e-07  
## EmploymentStatusDuration -7.392e-04 1.732e-04 -4.269 1.96e-05  
## IsBorrowerHomeowner 4.158e-01 3.297e-02 12.610 < 2e-16  
## CreditScoreRangeLower -8.452e-03 3.795e-04 -22.274 < 2e-16  
## CreditScoreRangeUpper NA NA NA NA  
## TotalCreditLinespast7years 4.167e-03 2.411e-03 1.728 0.084000  
## CurrentDelinquencies 1.135e-01 1.009e-02 11.252 < 2e-16  
## AmountDelinquent -5.611e-06 2.308e-06 -2.431 0.015041  
## DelinquenciesLast7Years -1.251e-02 1.649e-03 -7.585 3.31e-14  
## PublicRecordsLast10Years -8.731e-03 1.999e-02 -0.437 0.662317  
## PublicRecordsLast12Months 2.786e-01 9.999e-02 2.787 0.005326  
## BankcardUtilization -5.120e-01 5.344e-02 -9.580 < 2e-16  
## AvailableBankcardCredit -7.875e-07 9.795e-07 -0.804 0.421405  
## TotalTrades -1.038e-02 2.954e-03 -3.513 0.000443  
## TradesNeverDelinquent..percentage. -6.210e-01 1.184e-01 -5.245 1.56e-07  
## TradesOpenedLast6Months 2.137e-01 1.347e-02 15.866 < 2e-16  
## DebtToIncomeRatio 6.498e-02 2.587e-02 2.511 0.012023  
## IncomeRange -1.416e-01 9.195e-03 -15.403 < 2e-16  
## StatedMonthlyIncome -4.893e-05 4.715e-06 -10.376 < 2e-16  
## LoanOriginalAmount -8.275e-05 1.221e-05 -6.778 1.22e-11  
## MonthlyLoanPayment 2.948e-03 3.455e-04 8.531 < 2e-16  
##   
## (Intercept) \*\*   
## CreditGrade \*\*\*  
## Term \*\*\*  
## BorrowerAPR \*\*\*  
## ProsperRating..numeric. \*\*\*  
## ProsperRating..Alpha. \*\*\*  
## ProsperScore \*\*\*  
## ListingCategory..numeric. \*\*\*  
## EmploymentStatusDuration \*\*\*  
## IsBorrowerHomeowner \*\*\*  
## CreditScoreRangeLower \*\*\*  
## CreditScoreRangeUpper   
## TotalCreditLinespast7years .   
## CurrentDelinquencies \*\*\*  
## AmountDelinquent \*   
## DelinquenciesLast7Years \*\*\*  
## PublicRecordsLast10Years   
## PublicRecordsLast12Months \*\*   
## BankcardUtilization \*\*\*  
## AvailableBankcardCredit   
## TotalTrades \*\*\*  
## TradesNeverDelinquent..percentage. \*\*\*  
## TradesOpenedLast6Months \*\*\*  
## DebtToIncomeRatio \*   
## IncomeRange \*\*\*  
## StatedMonthlyIncome \*\*\*  
## LoanOriginalAmount \*\*\*  
## MonthlyLoanPayment \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 37729 on 27215 degrees of freedom  
## Residual deviance: 29359 on 27189 degrees of freedom  
## AIC: 29413  
##   
## Number of Fisher Scoring iterations: 5

The residual deviance value of 29359 which is less than the null deviance of 37729 indicates that the model fit is good and should have additional predictive ability.We can see that most of the p-values are small, below 0.05, indicating that the correlation coefficient has significance. Some features such as “AvailableBankcardCredit” and “PublicRecordsLast10Years” have large p-values and can be deleted from the model.

TRAINING Below are the results for the training data, including confusion matrix and accuracy scores.

##   
## Training Results:

##   
## FALSE TRUE  
## 0 9849 3759  
## 1 3674 9934

##   
## Accuracy = (TP + TN)/(TP + TN + FP + FN)

## Accuracy : 0.7268886

##   
##   
## Precision = (TP)/(TP + FP)

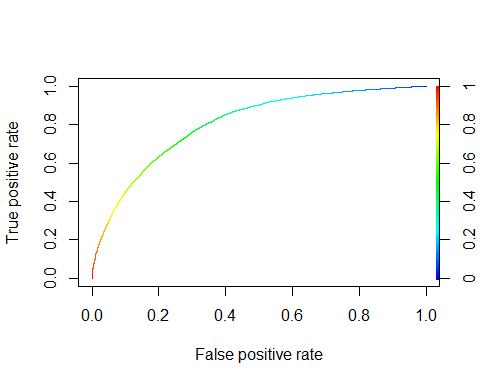
## Precision : 0.7254802

##   
##   
## Specificity = (TN)/(TN + FP)

## Specificity: 0.7237654

##   
##   
## Recall = (TP)/(TP + FN)

## Recall : 0.7300118

ROC Curve: training data 

TEST DATA

Below are the results for the testing data, including confusion matrix and accuracy scores.

##   
## Test Results:

##   
## FALSE TRUE  
## 0 60658 22661  
## 1 935 2467

##   
## Accuracy = (TP + TN)/(TP + TN + FP + FN)

## Accuracy : 0.727909

##   
##   
## Precision = (TP)/(TP + FP)

## Precision : 0.09817733

##   
##   
## Specificity = (TN)/(TN + FP)

## Specificity: 0.7280212

##   
##   
## Recall = (TP)/(TP + FN)

## Recall : 0.7251617

# Conclusions & Recommendations

The Logistic Regression Model created was of limited success in predicting default classification. The baseline dataset was 69-70% default. The Logistic Regression Classifier has approximately 72-73% accuracy in its test predictions. This is only 3% above the baseline. A detailed iterative feature selection process could improve the model.

I would recommend further research using this dataset. If the validity of more predictor variables could be shown the Logistic Regression Model would likely increase with their addition to the model.

It is also advisable that other Machine Learning Algorithms be used. A Random Forest Classifier or Neural Network should see better results.

The above test case used the relatively small Prosper Loan Dataset. Future work will be carried out on similar but much larger datasets from Lending Club P2P. Additionally, regional macro-economic factors will be sourced and tested. For example, with regards the Prosper Loan Dataset the relevant macro-economic data could be referenced using the feature “State” aka. US State. This feature is in the original dataset but was not used in this study and report.

# References

1. References: 1. Mach, T., Carter, C. and Slattery, C. (2014). Peer-to-Peer Lending to Small Businesses.SSRN Electronic Journal.
2. Malekipirbazari, M. and Aksakalli, V. (2015). Risk assessment in social lending via random forests.Expert Systems with Applications, 42(10), pp.4621-4631.
3. Serrano-Cinca, C., Gutiérrez-Nieto, B. and López-Palacios, L. (2015). Determinants of Default in P2P Lending.PLOS ONE, 10(10), p.e0139427.
4. Wang, H., Greiner, M. and Aronson, J. (2009). People-to-People Lending: The Emerging E-Commerce Transformation of a Financial Market.