Predictive Model for Prosper.com

BIDM Group 4
Natasha Sood
Nikhil Goenka
Salil Das
Sourabh Kukreja
Vikas Shah

What is prosper.com

- Prosper is the world's largest peer-to-peer lending marketplace, with more than 1,020,000 members and over $213 million in funded loans."
- Borrowers list loan requests between $2,000 and $25,000 and individual lenders invest as little as $25 in each loan listing they select.

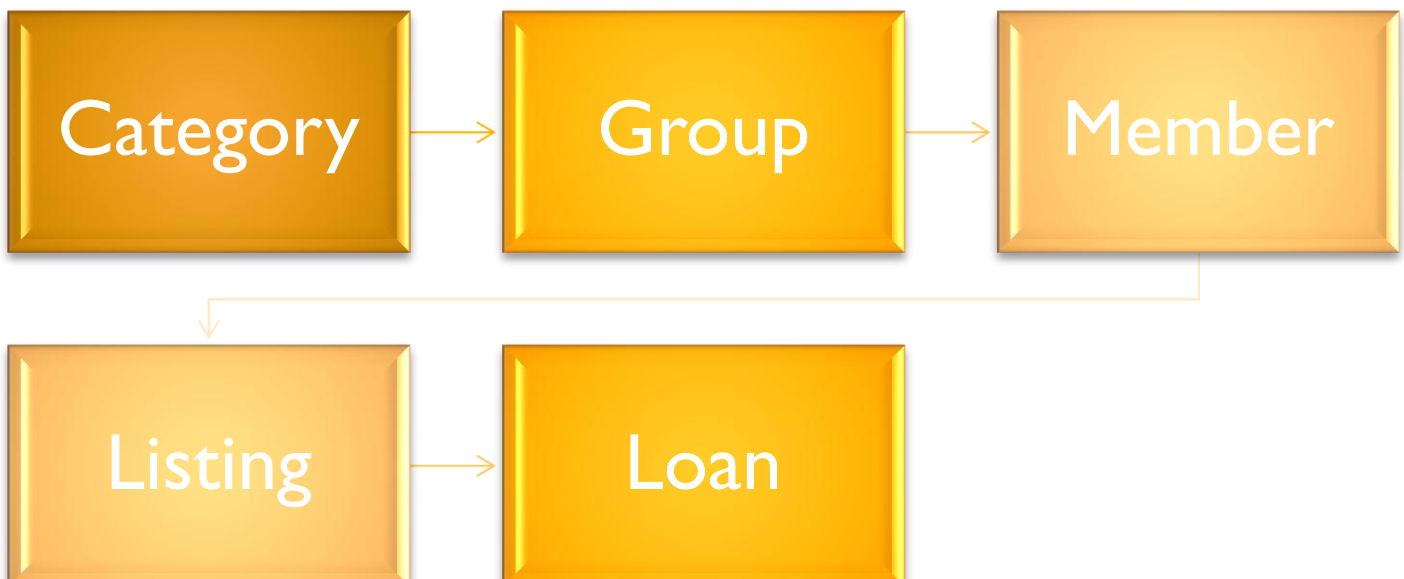
Objectives

- Build a predictive model for investors to be able to classify "Success" loans vs "Probable Default" Loans

Data Properties

- 1.5 GB of Data
- Over 2 MM rows
- 5 Tables of user data

Data Schema

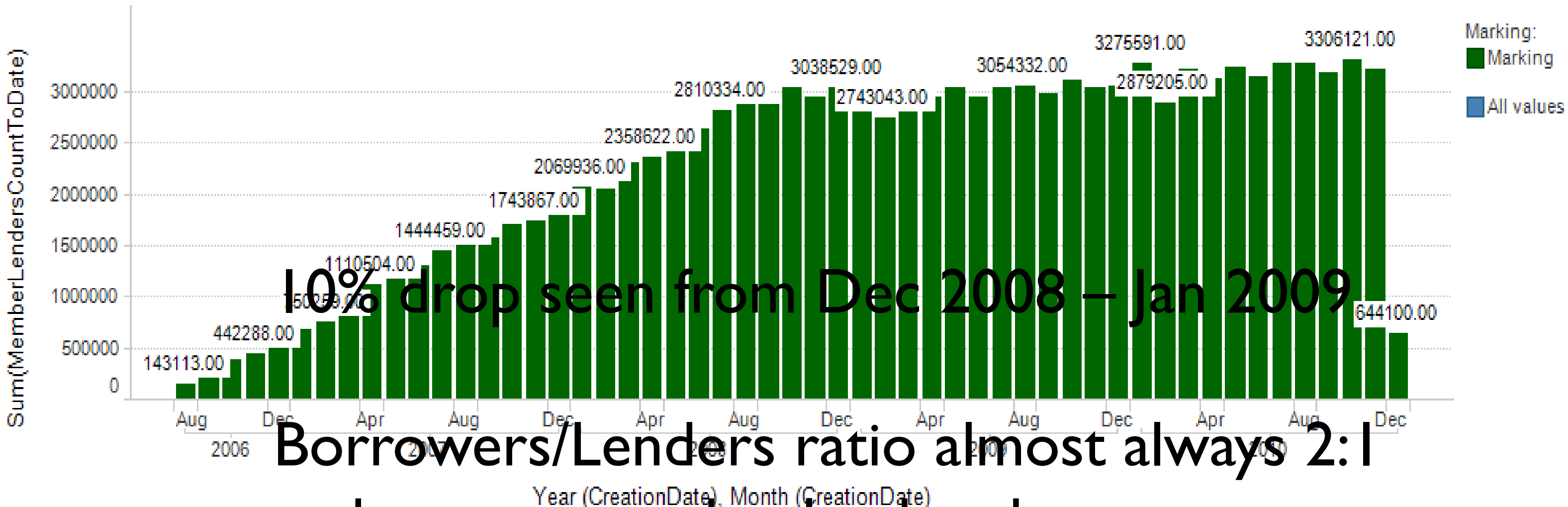


Data Processing

- Merging of data; Account Foreign Key Constraints
- Inner Join / Left Outer / Full Outer
- Random Sample from Data Set
- Random No generator
- Pick ~50k Rows (limitation with XLMiner)
- Working Data Sets
  - Listings – Members - Loans
  - Listings – Members – Groups
- Missing Data Values
- Data Binning
- Merging Columns (Credit Grade , Prosper Rating)

Growth in the member base over the years

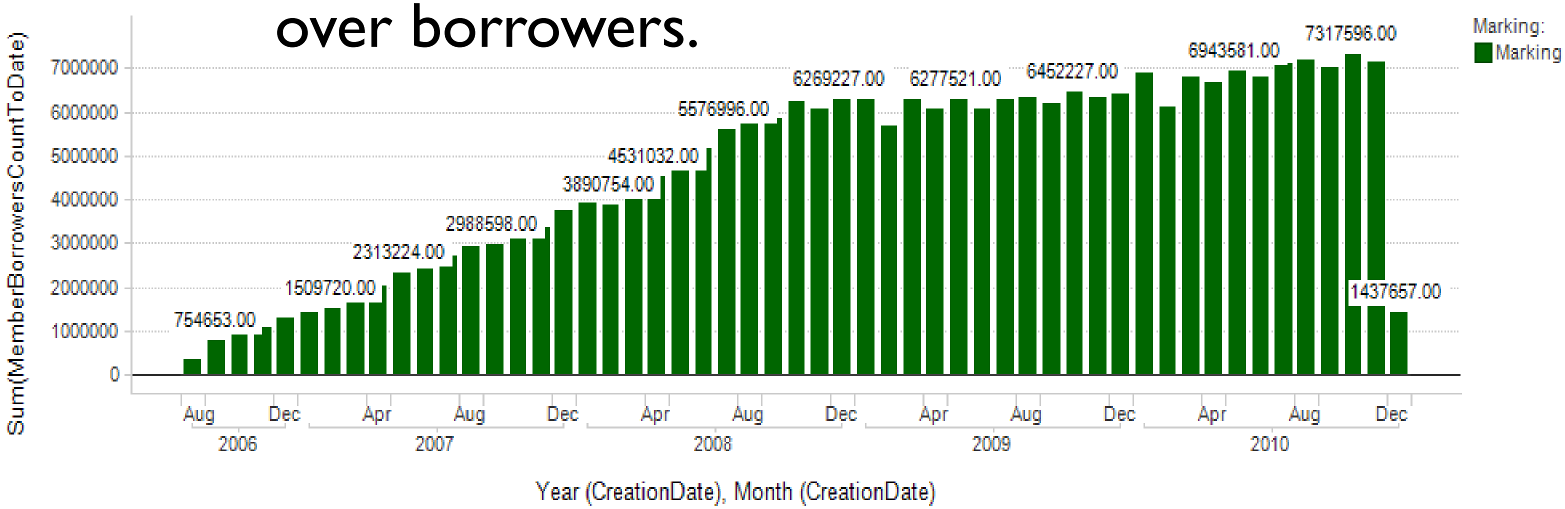
Lenders



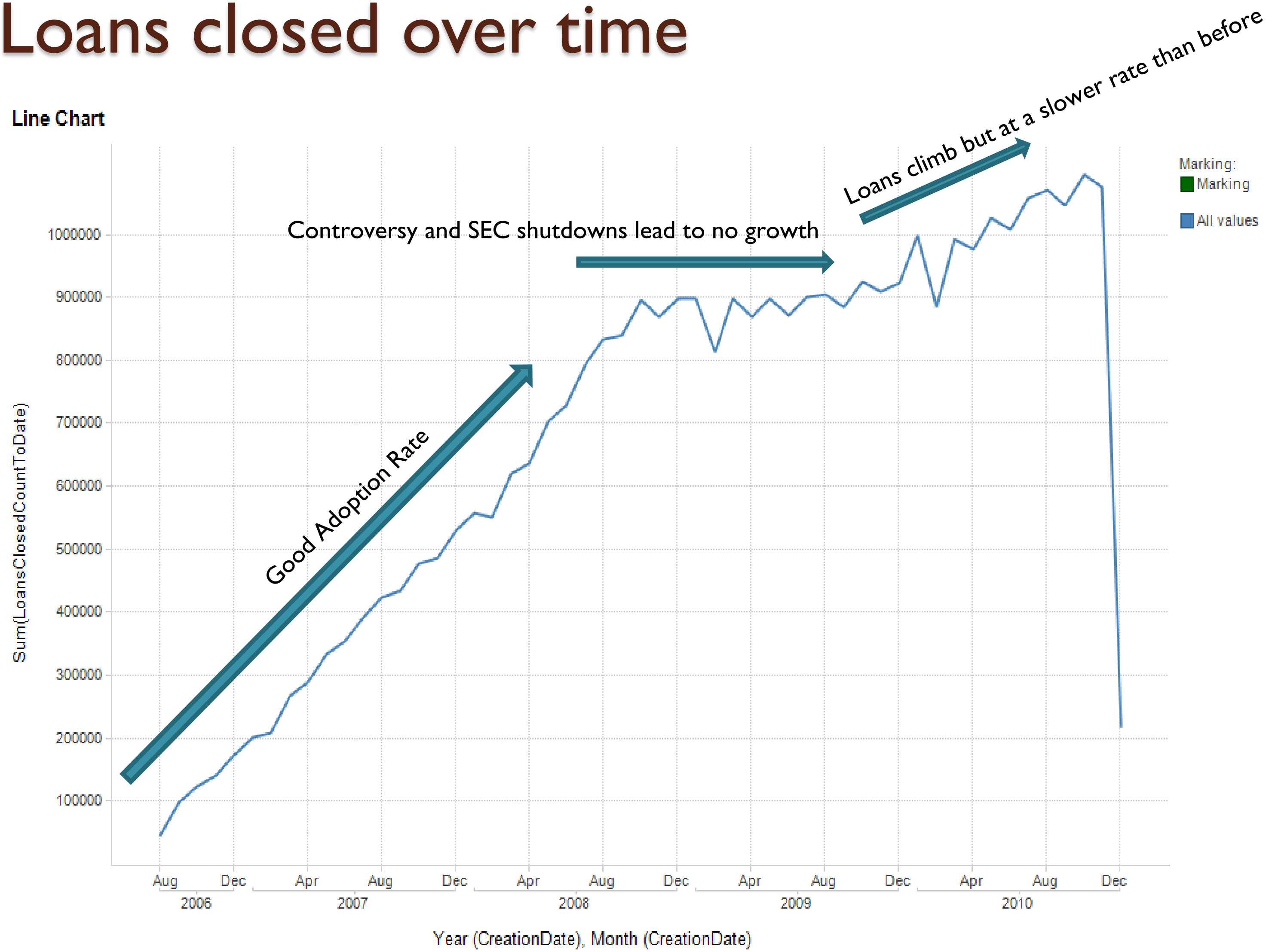
10% drop seen from Dec 2008 – Jan 2009

Borrowers/Lenders ratio almost always 2:1 and we can see that lenders have power over borrowers.

Borrowers

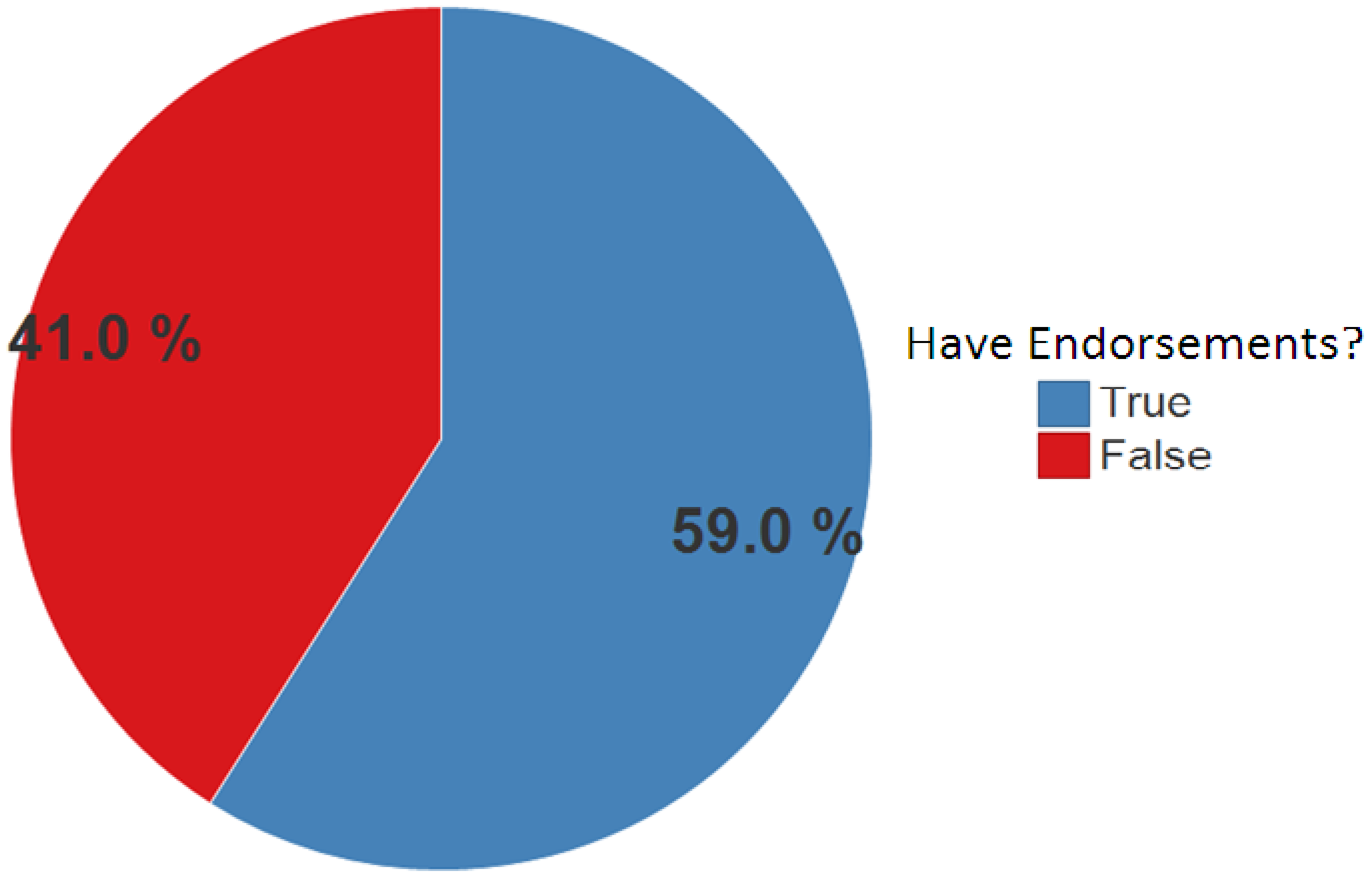


Loans closed over time



Loans closed rate grew significantly until Nov 2008 which coincides with the class action lawsuit

Having an endorsement increases your chance of being funded by 18%



Logistic Regression

- Output : Status (Default , Late , Paid)
- Predictors
  - Bid Count
  - Borrower rate
  - Lender rate
  - Age in months
  - Amount Borrowed
  - Is Home Owner
  - Debt To Income Ratio
- Co-relation analysis
- Significant predictors
- Goodness of Fit : Performance & Error Rate

Logistic Regression

Input variables	Coefficient	Std. Error	p-value	Odds
BidCount	0.00045379	0.00028042	0.10560962	1.000454
BorrowerRate	24.77111626	4.57036781	**0.00000006**	5.727E+10
DebtToIncomeRatio	0.03414476	0.02391896	0.15343051	1.0347344
IsBorrowerHomeowner	0.17194989	0.05088718	**0.00072739**	1.1876184
LenderRate	-15.33367443	4.61400652	**0.00088963**	2.2E-07
AgeInMonths	0.05707799	0.0022076	**0**	1.0587384
AmountBorrowed	0.00003511	0.0000068	**0.00000024**	1.0000352
Term	-0.141791	0.00396481	**0**	0.8678026

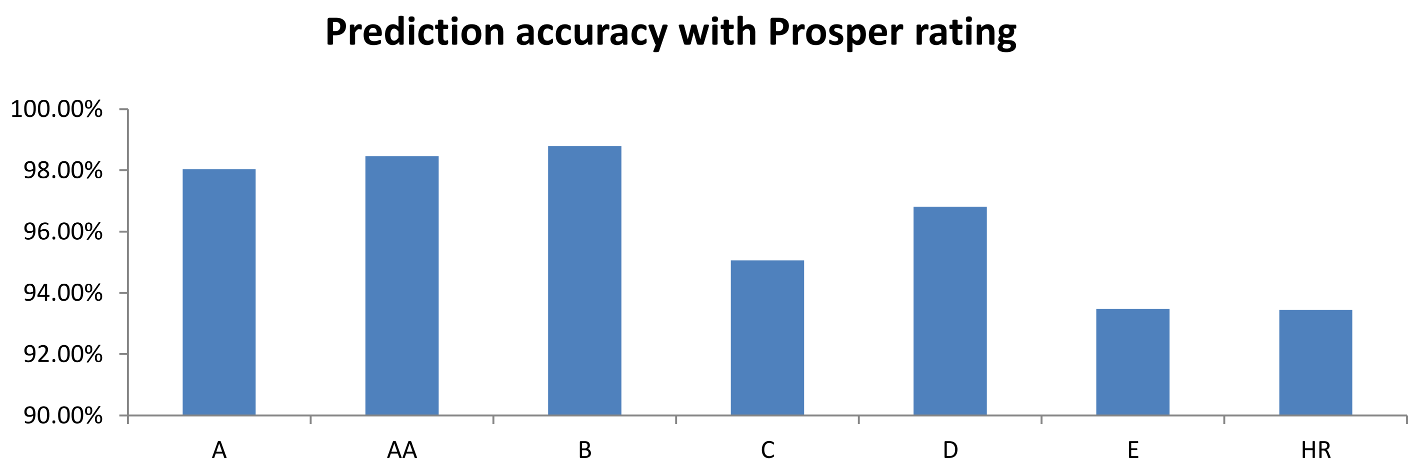
Training data			
Class	**# Cases**	**# Errors**	**% Error**
0	2986	1951	65.3
1	7013	688	9.8
Overall	9999	2639	26.4

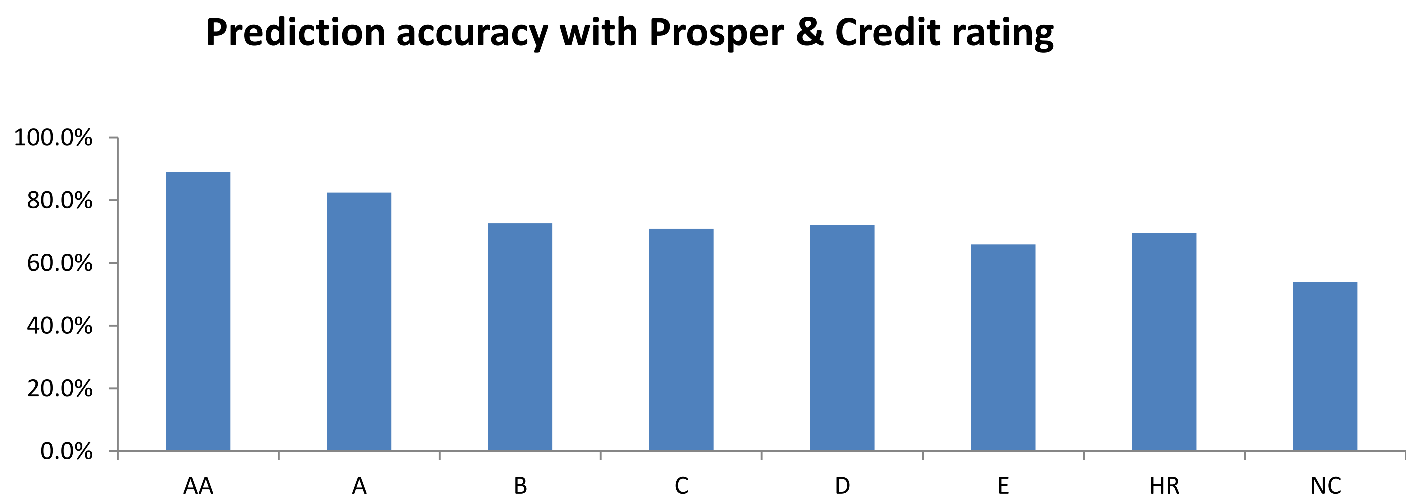
Test data			
Class	**# Cases**	**# Errors**	**% Error**
0	3118	2031	65.1
1	7378	739	10.0
Overall	10496	2770	26.39

Classification Tree

- Output : Status (Default , Late , Paid)
- Predictors
  - Bid Count
  - Borrower rate
  - Lender rate
  - Age in months
  - Amount Borrowed
  - Is Home Owner
  - Debt To Income Ratio
- Parameters of the tree
  - Best Pruned Tree
  - 100 in the terminal node
- Classification tree top 3 predictors
  - Age in Months
  - Borrower Rate
  - Bid Count

Prediction accuracy





What We tried …… but Failed

- Role of Social Network
  - Relation between Member Friends in network and probability of getting a loan
- Some categories better than others
  - Religious / Ethnic ?
- Does time of the year has role to play in role approcal
  - Thanks Giving / Christmas time !!!

Conclusion

- Real use of our model
  - Usage along with Prosper Rating for identifying "Lemons"
  - 50+% opportunities without Prosper Rating can also be lucrative.

Questions