

## PREDICTING FUNDING TIME FOR MICRO-LOANS

James Salisbury, Nader Esmael, Sid Carter



#### Problem Statement

#### PREDICTING FUNDING TIME FOR MICRO-LOANS

Microlending is a form of peer-to-peer financing that has seen continued expansion over the past couple of decades. Microloans are most commonly granted to impoverished borrowers who typically lack collateral.

Kiva.org is an internet-based microlending platform that was founded in 2005. Since Kiva's inception, they have processed almost 2 million loans via their platform in over 75 countries. Similar to other microcredit organizations, Kiva promotes a 95.8% repayment rate and over 1.9 million lenders.

While most of the loans processed through Kiva's site are \$500 or less, there are many factors which may affect the time to funding besides just the loan amount. Kiva has reached out to us to help them build a predictive model which will ultimately be implemented as a tool for borrowers. We have been tasked to build a predictor that will help borrowers estimate the total time required to fund a loan.

				•				
kiva	Lend 🔻				Borrow	About 🔻	Sign in	
			įþu	43 days Total loc Powerso Tim	funded lieft n: \$9,375 lby 194 lenders my Secta, Uganda /	\$3.; / Water Distribi	175.00 to go ution	
	, cust	loan of \$9,3 tomers who co	75 helps to bu ould not previo water	iy a truck to ously afford 7.	e reach mor I clean drin	re king		
Timmy	's story	,		< Loan	details			
With the help customers or community. His volumes I the rise. The demand	o of the previou ad deliver wate have grown tr current delive	us Kiva loans, Jean has l er to more than 60% of f emendously and the der ry tools that he has are r	been able to recruit households in his nand for water is also on hot sufficient to meet this	(3) Loan 26 I Repays	<u>length</u> : nonths entschedule: Mon	thly		

Kiva.org This screenshot shows what a typical borrower page looks like. Borrowers fill out a profile including items such as some personal information, country, activity the loan will be used for, loan information, etc.



\$1.5 billion In Ioans

1.9 million

95.8% Repayment rate



### Data Discovery and Cleaning



### Data Discovery

- Data "snapshot" from Kiva
- 1.96 M loans from 2006 -August 2020
- 1.85 M funded loans from 99 countries



### Data Cleaning

#### Trimmed Columns

First, we trimmed the unnecessary columns to reduce our dataset size; We also dropped all loans with no 'raised time'.

#### **Funding Status**

Next, we dropped loan statuses of 'expired', 'refunded', and 'fundraising', only keeping 'funded' loans.

#### **Date Calculation**

We then calculated the total time to raise the loan and manipulated the date time information to calculate 'hours to fund'.

#### Images / Videos

We coded all image and video IDs to just represent whether an image or video was included.

#### Mapping Genders

Genders were mapped to a single character; individual females and males were mapped to 'F' and 'M', respectively; multiple borrowers were mapped to 'G'.

#### **Dropping Columns**

Any remaining columns that were no longer needed or used were drop to limit the dataset file size.



### 2. Exploratory Data Analysis

#### Average and Maximum Loan Amounts, by Year





#### Average and Maximum Hours To Fund, by Year







### Loans by Country



#### Lowest Average Loan Values by Country

Top 10 Lowest Loan Amount averages by country



#### Borrower Genders and Repayment Interval Distributions





#### Hours to Fund and Loan Amount Distributions



Loan Amount in USD

Distribution of Loan Amounts (outliers removed)

Loans <= 1008 hours to fund</p>

► Loans < \$2000</p>

14

## Correlations Between NumericalFeatures





### 3. Modeling and Tuning



#### Parametric Models

- Models
  - OLS Linear Regression
  - GLM Poisson
- Fast to run; easy to understand and interpret
  Works better with highly

correlated, numeric data



### Non-parametric Models

- ⊷ Models
  - → Decision Tree
  - Bagging Decision Tree
  - Random Forest
- Can result in better models for prediction
- Slower to train

Model M	letrics Cor	nparis	nparison		
		RMSE	R2		
	OLS Linear Regression	298	.16		
	GLM Poisson	277	N/		
	Decision Tree	280	.34		
	Bagging Decision Tree	282	.34		
	Random Forest	283	.35		

	RMSE	R2
OLS Linear Regression	298	.163
GLM Poisson	277	N/A
Decision Tree	280	.344
Bagging Decision Tree	282	.343
Random Forest	283	.353





### Conclusion



#### Conclusion

 Non-parametric models performed better underperforming for a production model Outside factors may lead to unpredictability in time to fund for loans



#### Next Steps

- Development of a non-parametric neural network model.
- Learn about their lending process.

# Thanks!

#### ANY QUESTIONS?

