

# How to make Money with Lending Club (and Machine Learning)

Younes Iferd, Marco Zagermann

July 2, 2020

# How can one make money with machine learning?

## Possibility 1:

Get paid for solving **somebody else's** problems with machine learning.

# How can one make money with machine learning?

## Possibility 1:

Get paid for solving **somebody else's** problems with machine learning.

## Possibility 2:

Use machine learning to **your own** advantage and try to **outsmart the system!**

# How can one make money with machine learning?

## Possibility 1:

Get paid for solving **somebody else's** problems with machine learning.

## Possibility 2:

Use machine learning to **your own** advantage and try to **outsmart the system!**



**Attempt in this talk:**

**Improve your investment decisions at “Lending Club” with ML**

# Outline for the rest of the talk

- 1. What is Lending Club and why ML?**
- 2. Some interesting observations**
- 3. Results of machine learning models**

# **1. What is Lending Club and why ML?**

# The concept of Lending Club

**Lending Club** : An internet platform that brings together **borrowers** and **investors**



# The concept of Lending Club

**Lending Club** : An internet platform that brings together **borrowers** and **investors**



## Borrower:

- applies for a credit
- provides information



# The concept of Lending Club

**Lending Club** : An internet platform that brings together **borrowers** and **investors**



## Lending Club:

- screens borrowers
- facilitates transactions
- services the loan

## Borrower:

- applies for a credit
- provides information

# The concept of Lending Club

**Lending Club** : An internet platform that brings together **borrowers** and **investors**



## Investor/Lender:

- Select loans they want to invest in
- make lending offer

## Lending Club:

- screens borrowers
- facilitates transactions
- services the loan

## Borrower:

- applies for a credit
- provides information

# The concept of Lending Club

**Lending Club** : An internet platform that brings together **borrowers** and **investors**

Maybe one  
could use  
machine  
learning?



## Investor/Lender:

- Select loans they want to invest in
- make lending offer

## Lending Club:

- screens borrowers
- facilitates transactions
- services the loan

## Borrower:

- applies for a credit
- provides information

# Challenges for Investors

1. Pick the **most reliable borrowers** (to avoid credit defaults)

# Challenges for Investors

1. Pick the **most reliable borrowers** (to avoid credit defaults)
2. **Maximize interest rate** (usually is in tension with reliability)

# Challenges for Investors

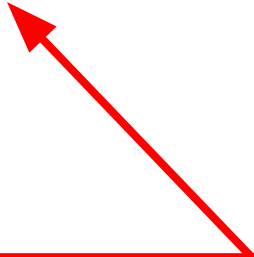
1. Pick the **most reliable borrowers** (to avoid credit defaults)
2. **Maximize interest rate** (usually is in tension with reliability)
3. **Diversify** the portfolio ( i.e. distribute investment over **many** borrowers)

# Challenges for Investors

1. Pick the **most reliable borrowers** (to avoid credit defaults)
2. **Maximize interest rate** (usually is in tension with reliability)
3. **Diversify** the portfolio ( i.e. distribute investment over **many** borrowers)
4. Be **fast!**

# Challenges for Investors

1. Pick the **most reliable borrowers** (to avoid credit defaults)
2. **Maximize interest rate** (usually is in tension with reliability)
3. **Diversify** the portfolio ( i.e. distribute investment over **many** borrowers)
4. Be **fast!**



Four postings every day: **6am, 10am, 2pm, and 6pm Pacific Time**

Many of the loans are fully invested **within minutes** and **disappear** from the list!



# What the investors get:

Browse Loans

Summary | Portfolio Builder | **Browse Loans** | Alert | Transfer | Trading Account | Automated Investing

Available: \$36.06

Add Funds

Add to Order

Showing Loans 1 - 15 of 463

<< < 1 2 3 4 5 > >> 15 ▾

<input type="checkbox"/> Investment	Rate	Term	FICO®	Amount	Purpose	% Funded	Amount / Time Left
<input type="checkbox"/> \$0	<b>B 3</b> 9.99%	60	685-689	\$12,000	Loan Refinancing & Consolidation	<div><div></div></div> 81%	\$2,175 10 days
<input type="checkbox"/> \$0	<b>B 3</b> 9.99%	60	725-729	\$24,000	Loan Refinancing & Consolidation	<div><div></div></div> 83%	\$3,875 10 days
<input type="checkbox"/> \$0	<b>B 3</b> 9.99%	60	720-724	\$16,600	Loan Refinancing & Consolidation	<div><div></div></div> 93%	\$1,100 11 days
<input type="checkbox"/> \$0	<b>B 2</b> 9.17%	60	680-684	\$15,000	Other	<div><div></div></div> 62%	\$5,600 10 days
<input type="checkbox"/> \$0	<b>A 5</b> 7.89%	60	790-794	\$25,600	Major Purchase	<div><div></div></div> 68%	\$8,075 10 days
<input type="checkbox"/> \$0	<b>B 5</b> 11.53%	36	675-679	\$11,000	Credit Card Payoff	<div><div></div></div> 94%	\$650 12 days

**Build a Portfolio**

Per Loan: \$25

**Filter Loans** Save | Open

**Loan Term** ▾

☒ 36-month

☒ 60-month

**Public Records** ▾

☐ Exclude Loans with Public Records

**Location State** ▾

**Earliest CREDIT line** ▾

**Funding Progress** ▾

Borrower 1

Borrower 2

Borrower 3

•  
•  
•

Grade and interest rate (+ a lot more info)<sup>17</sup>

# How should the investor (quickly) pick their loans?

A helpful criterion:

The **grade** provided by Lending Club:

A1, A2, ..., A5, B1, B2,..., B5, ..... , G1, G2, ..., G5



# How should the investor (quickly) pick their loans?

A helpful criterion:

The **grade** provided by Lending Club:

A1, A2, ..., A5, B1, B2,..., B5, ..... , G1, G2, ..., G5

← Reliability

→ Interest rate

# Poor man's quick and dirty strategy:

Choose (for example):

- Grade **A1, A2, A3, A4, A5**
- Annual income **at least 50 000 USD**

# Poor man's quick and dirty strategy:

Choose (for example):

- Grade **A1, A2, A3, A4, A5**
- Annual income **at least 50 000 USD**

**Result:**

- **Pay-back probability = 95 %**
- But: **Detection rate** of pay-back borrowers: **Only 20 % !**

# Poor man's quick and dirty strategy:

Choose (for example):

- Grade **A1, A2, A3, A4, A5**
- Annual income **at least 50 000 USD**

**Result:**

- **Pay-back probability = 95 %**
- But: **Detection rate** of pay-back borrowers: **Only 20 % !**

I.e. **80%** of the good investment opportunities would be **missed!**  
→ **Diversification** of portfolio more **difficult**  
→ **Low** average **interest rate**

# How much can this be improved with ML?

Use a **Lending Club data set** to train **machine learning** algorithms:

- Data from **42 500** borrowers
- Number of features: **52** (Annual income, purpose, loan amount ...)
- Average interest rate: **12%**
- Average **pay back probability**: **q = 85.5 %** ( → **class imbalance!** )

## **2. Some interesting observations**



# What influences the pay-back probability?

Three **less surprising** features:

- Grade
- Interest rate
- Annual income

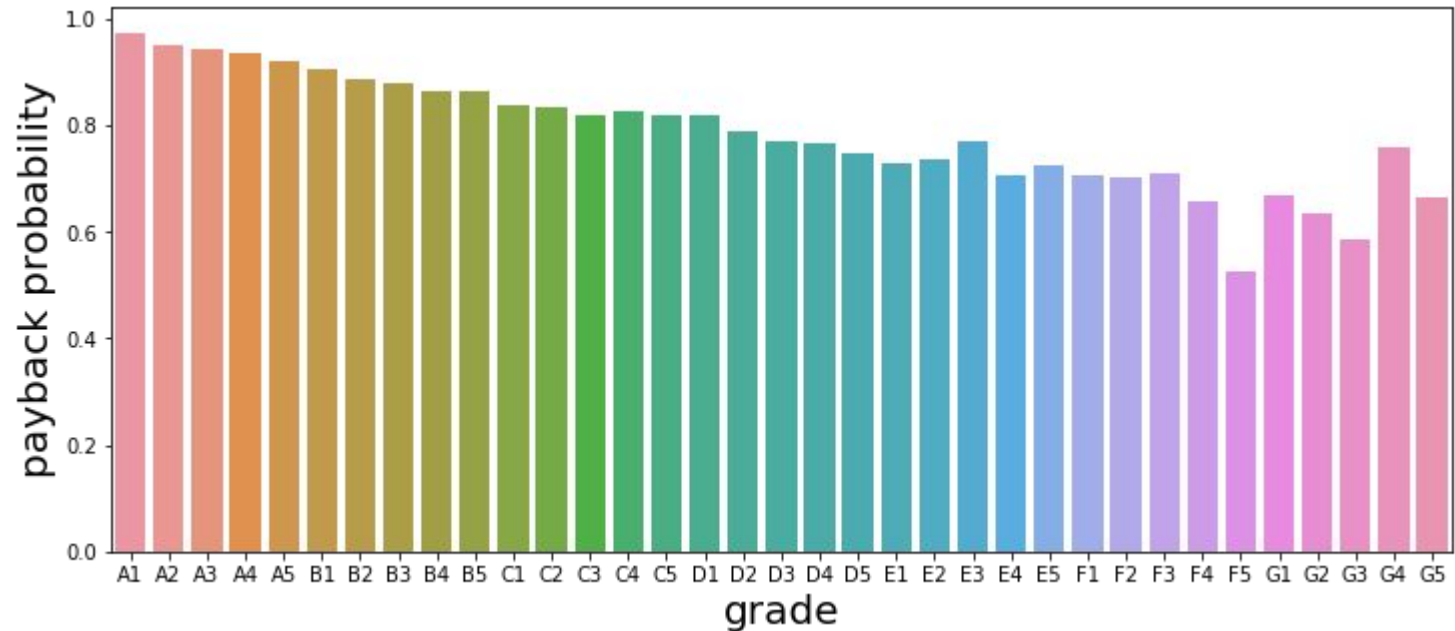
Three **more surprising** features:

- Purpose of the loan
- Employment title
- State

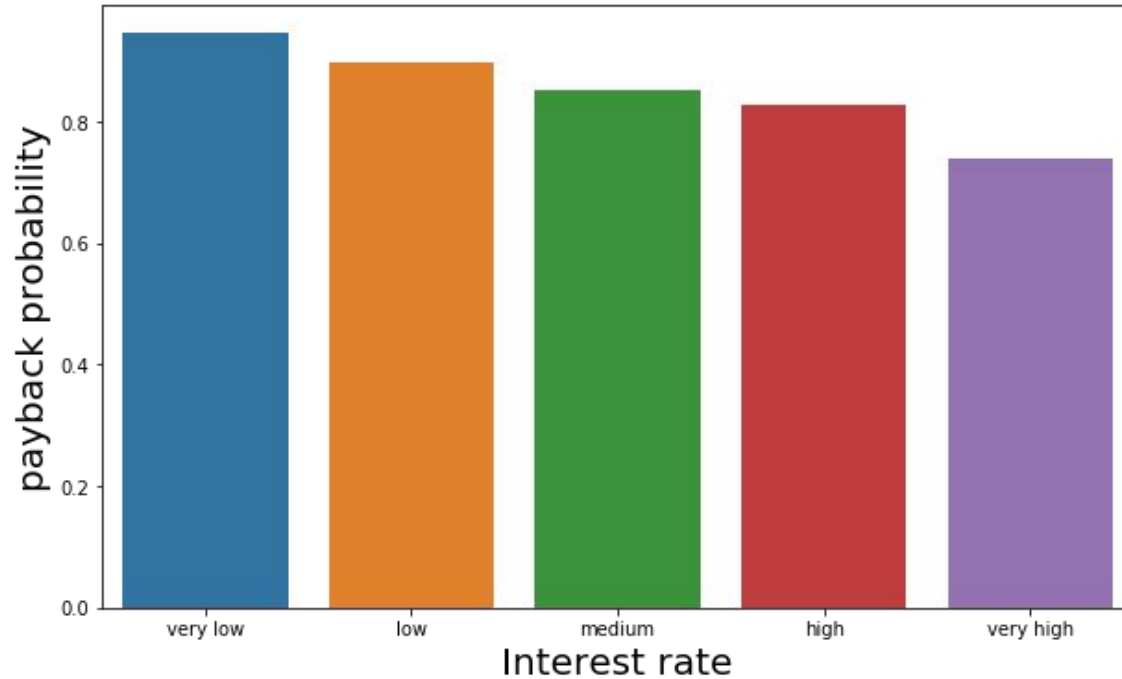
+ many more

# Grade

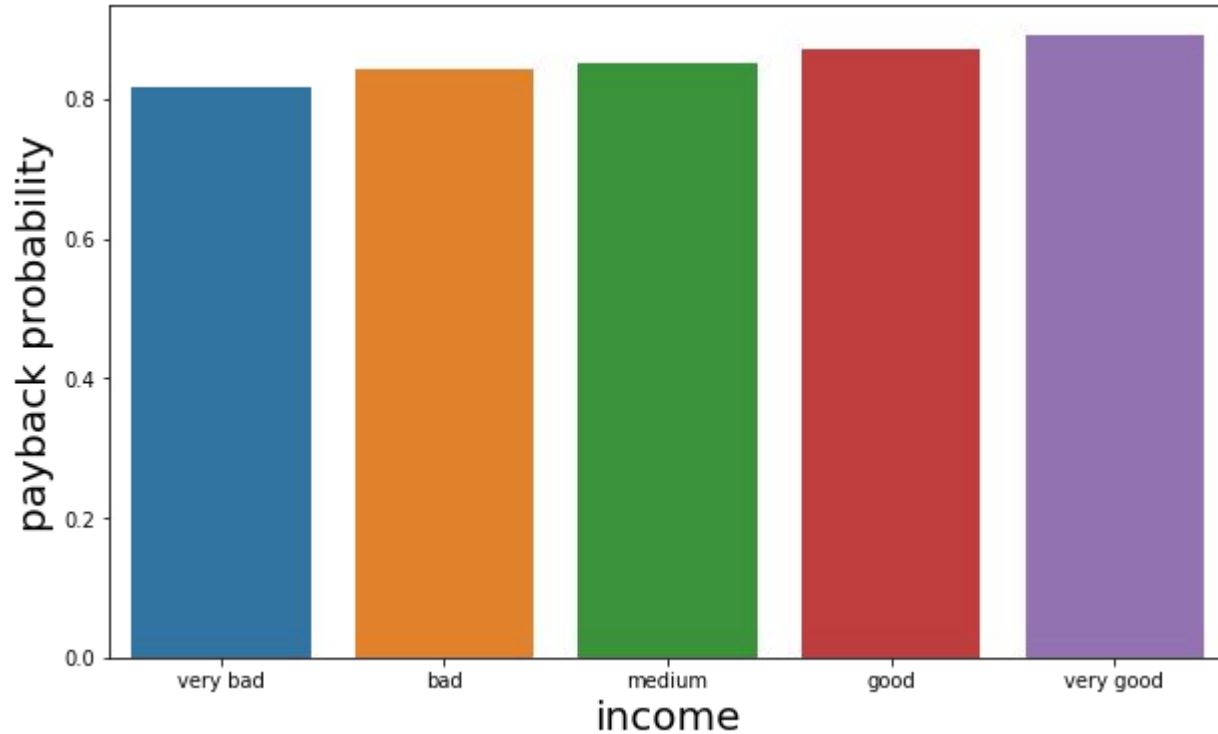
With deterioration of the grade, the payback probability decreases



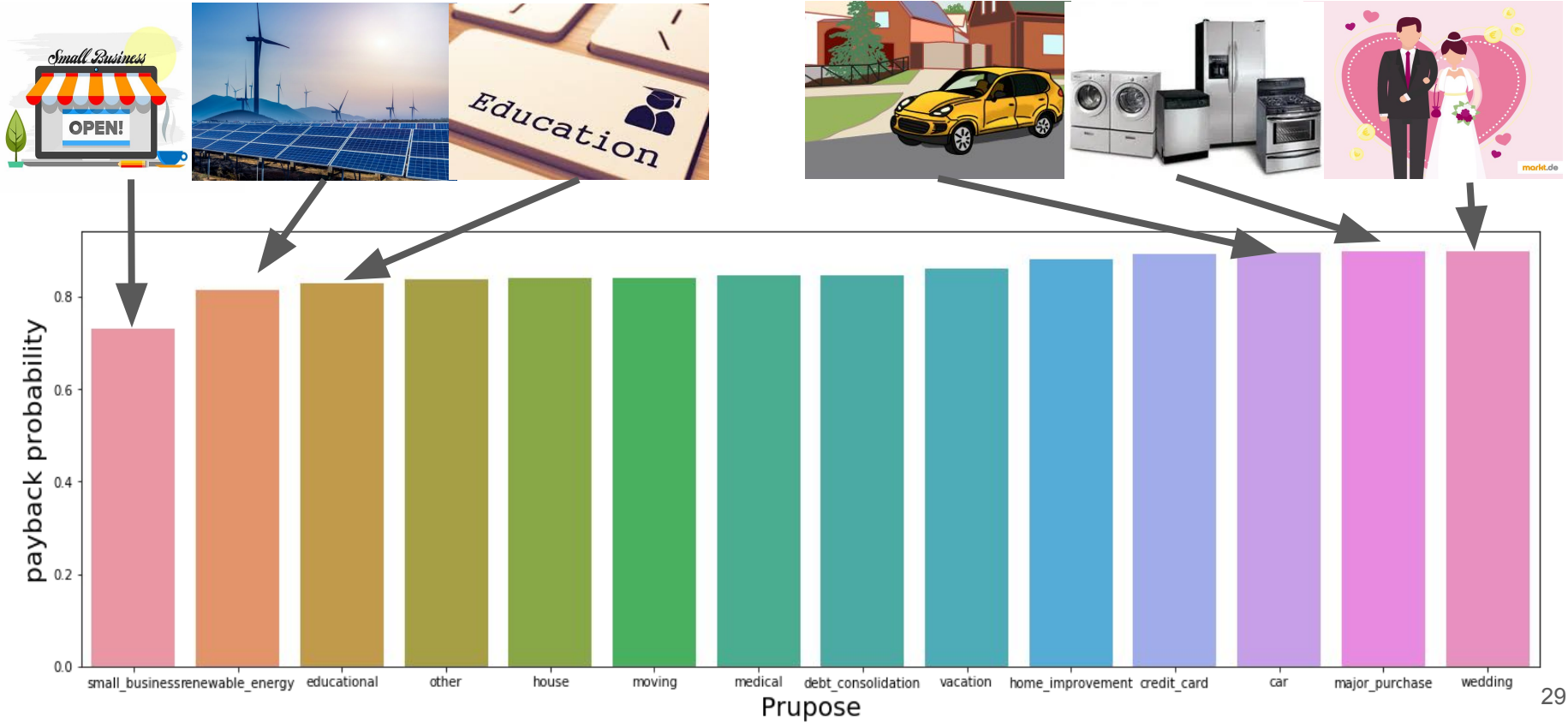
# Interest rate



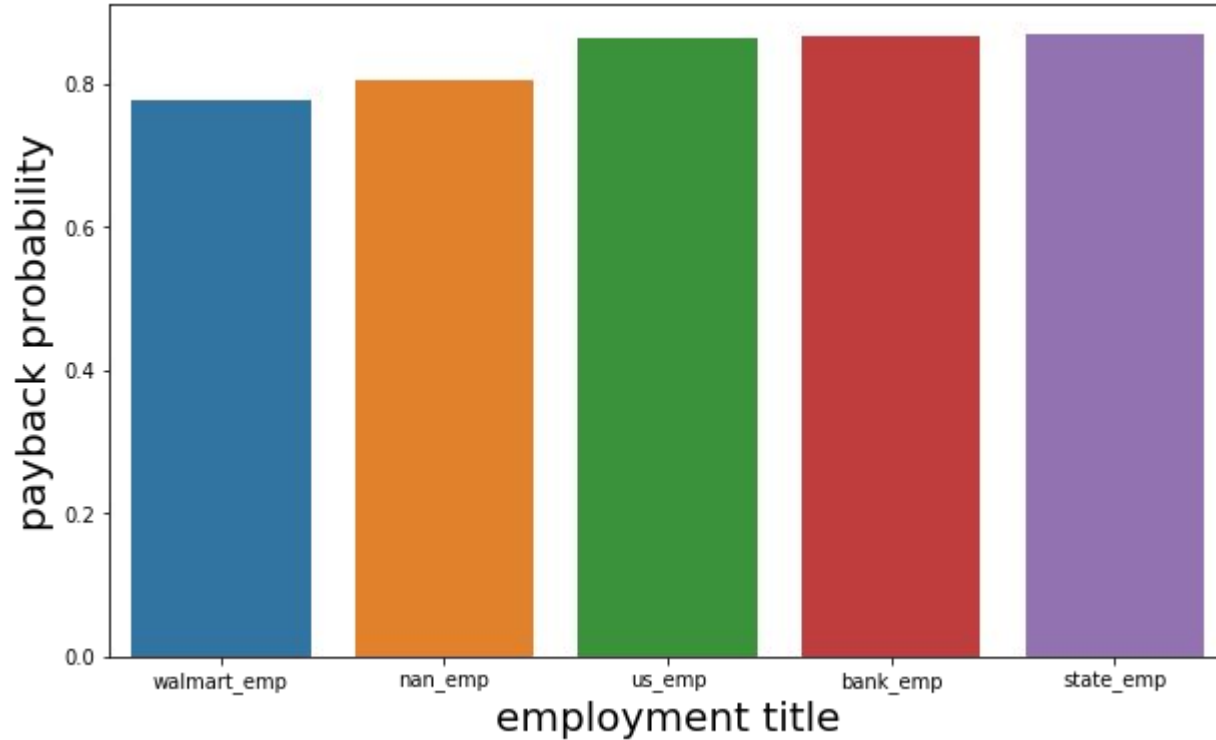
# Annual income



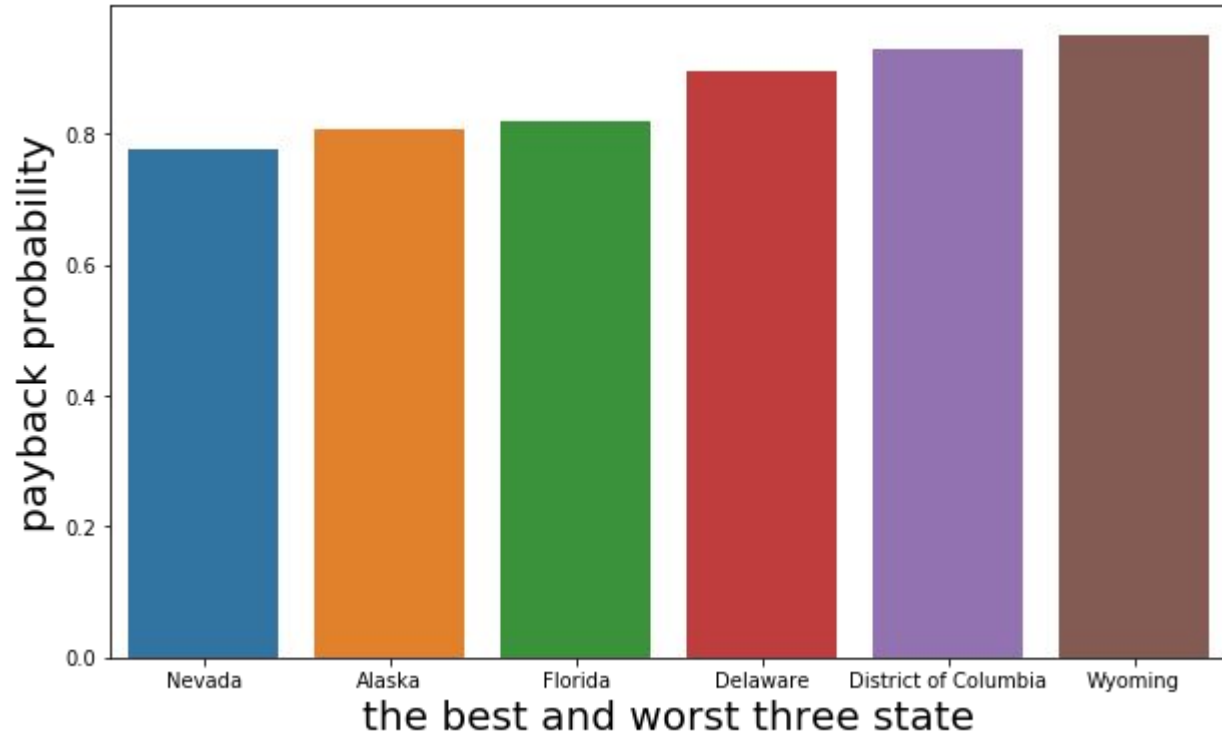
# Purpose



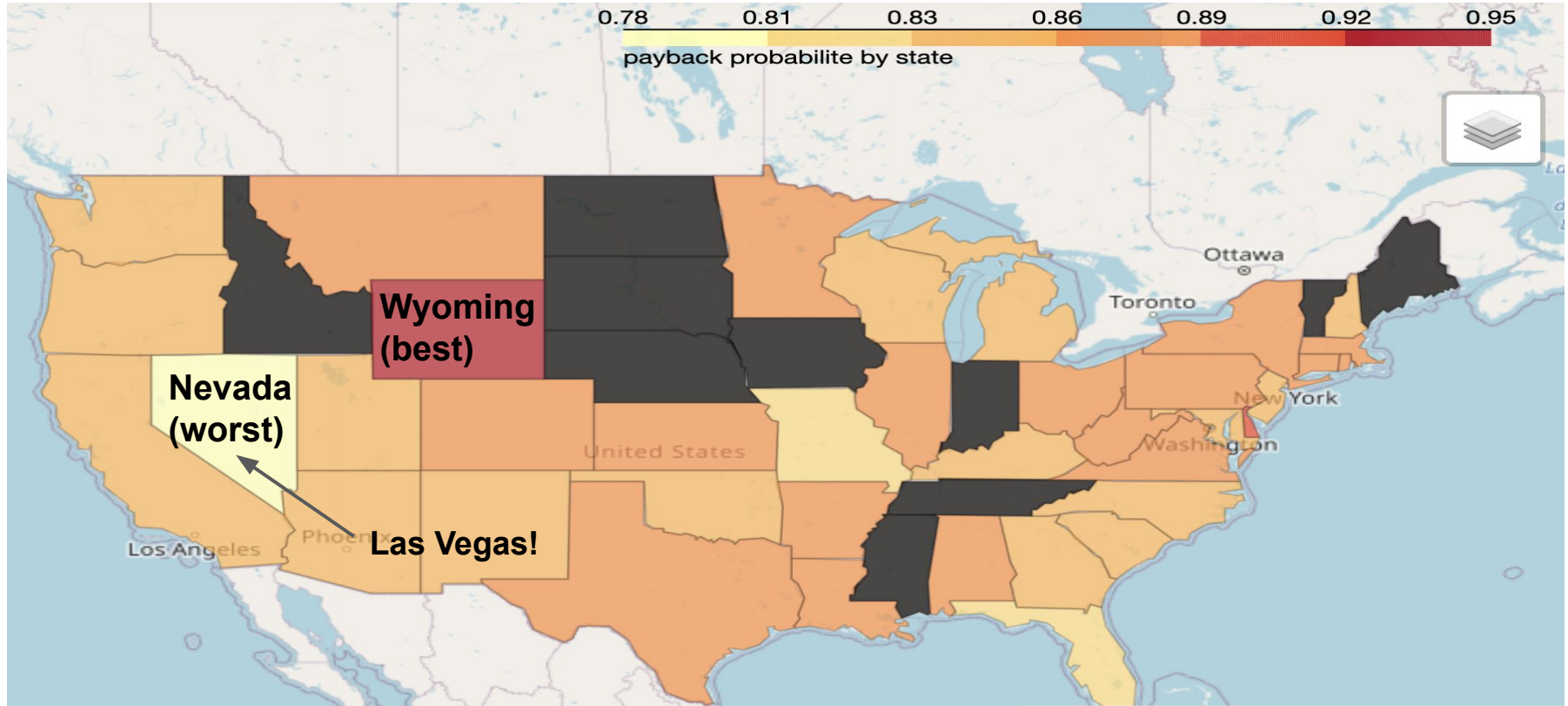
# Employment title



# State



# State





### **3. Results of machine learning models**

# Systematic analysis with machine learning

**Benchmark:** Poor man's quick and dirty strategy (grade A5 or better, income >50k)

Pay-back probability:  $q = 95 \%$

Detection rate:  $20 \%$

# Systematic analysis with machine learning

**Benchmark:** Poor man's quick and dirty strategy (grade A5 or better, income >50k)

Pay-back probability:  $q = 95\%$

Detection rate:  $20\%$

**Trained and tuned ten machine learning algorithms:**

Logistic regression

KNN

Gaussian NB

SVM with four kernels

Decision tree

Random Forest

Decision tree with Adaboost

+ Grid search

# Systematic analysis with machine learning

**Benchmark:** Poor man's quick and dirty strategy (grade A5 or better, income >50k)

Pay-back probability:  $q = 95\%$

Detection rate:  $20\%$

**Trained and tuned ten machine learning algorithms:**

**Results:** Pay-back probability:  $q = 95\%$  (Logistic regression tuned to  $q = 95\%$ )

Detection rate:  $35\%$

# Systematic analysis with machine learning

**Benchmark:** Poor man's quick and dirty strategy (grade A5 or better, income >50k)

Pay-back probability:  $q = 95 \%$

Detection rate:  $20 \%$

**Trained and tuned ten machine learning algorithms:**

**Results:** Pay-back probability:  $q = 95 \%$  (Logistic regression tuned to  $q = 95 \%$ )

Detection rate:  $35 \%$

( $q > 95 \%$  can be easily achieved at the expense of lower detection rate)

# Conclusion:

**Machine learning can improve your investment decisions!**

# Future work

- Make better use of job titles and perhaps ZIP-codes
- Improve KNN and Naive Bayes by taking into account class imbalance
- Use more ensemble methods
- Use dataset to predict early repayments
- Register at Lending Club and test the model

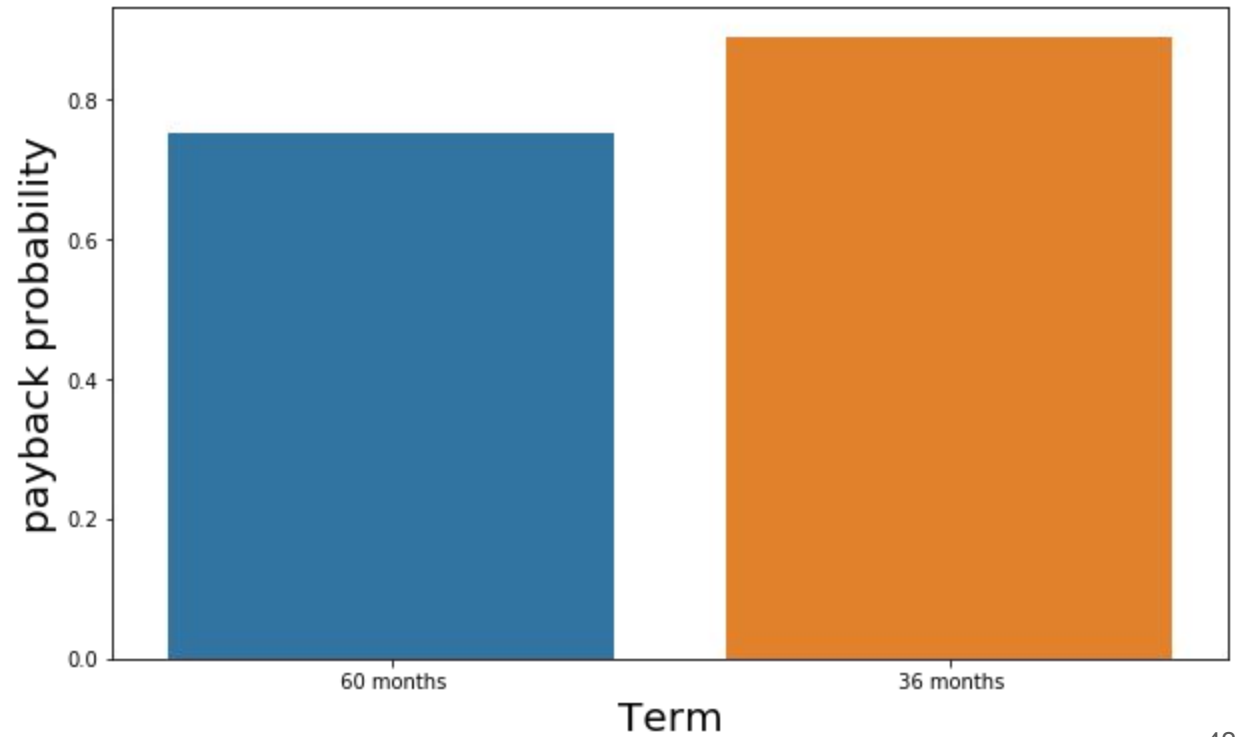
Thank you !







# Term



# Revolving utilization Rate

Used credit/Available credit

