Credit Card Default Prediction

Xinyu Zhang

The George Washington University

CONTENT

- Introduction
- Data Description
- Exploratory Data Analysis
- Modeling
- Result
- Future work

Introduction

Background

- Credit card has become the preferred payment method for Americans because of its convenience, safety, and better rewards.
- With the growing number of credit card clients, banks have been facing default risk.

Goal

- Build different machine learning models to predict credit card default
- Determine the best one
- Reduce credit card default rate
- Deliver insights for banks about credit card strategies

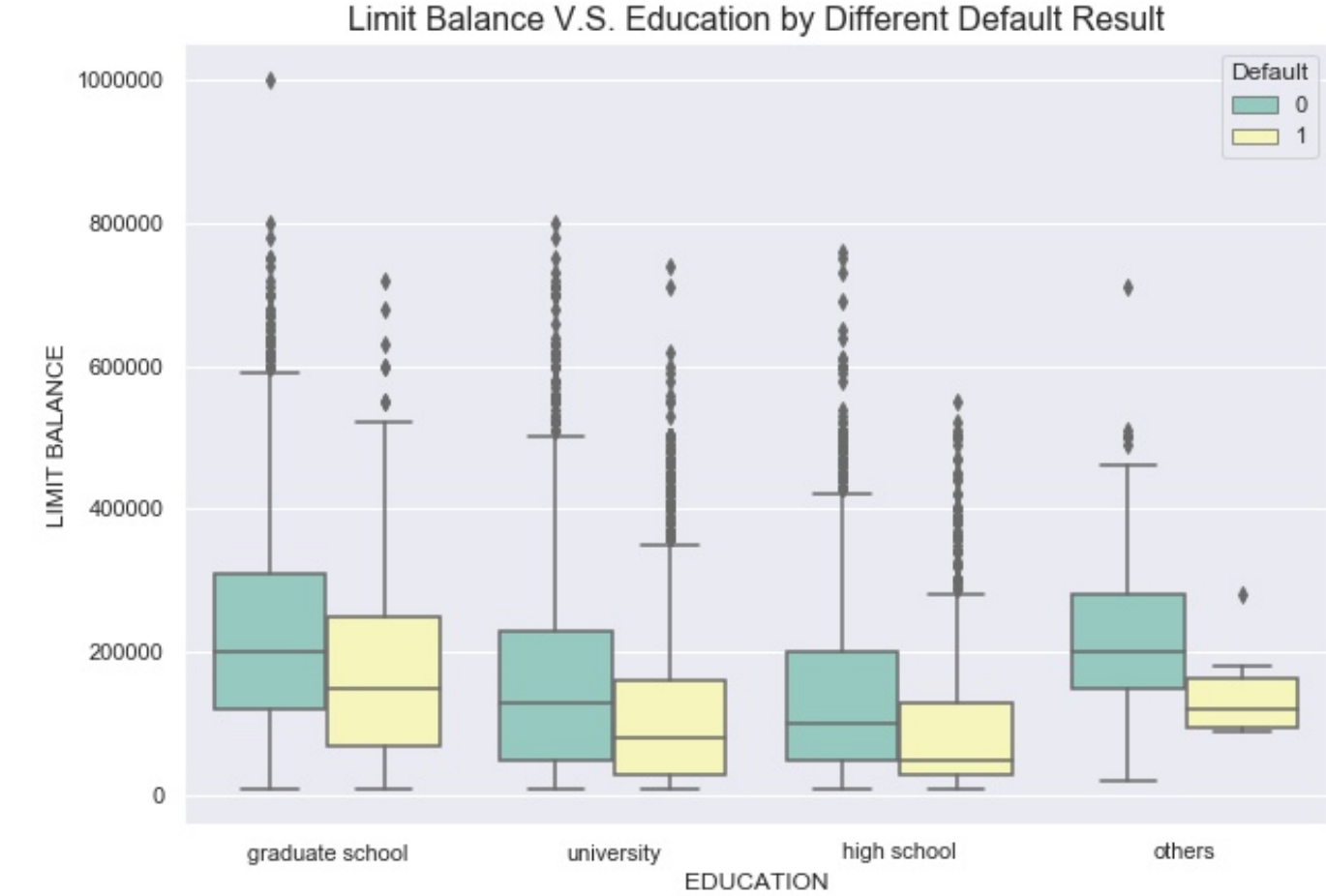
Data Description

- Taiwan-based Credit Card Default Data https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients

- 2005 April – 2005 September
- 30000 rows
- 1 binary label variable [default payment in Oct (1 = Yes, 0 = No)]
- 23 exploratory variables
  - Limit Balance - Amount of given credit
  - Gender
  - Education
  - Marital status
  - Repayment status
  - Bill statement amount
  - Previous payment amount

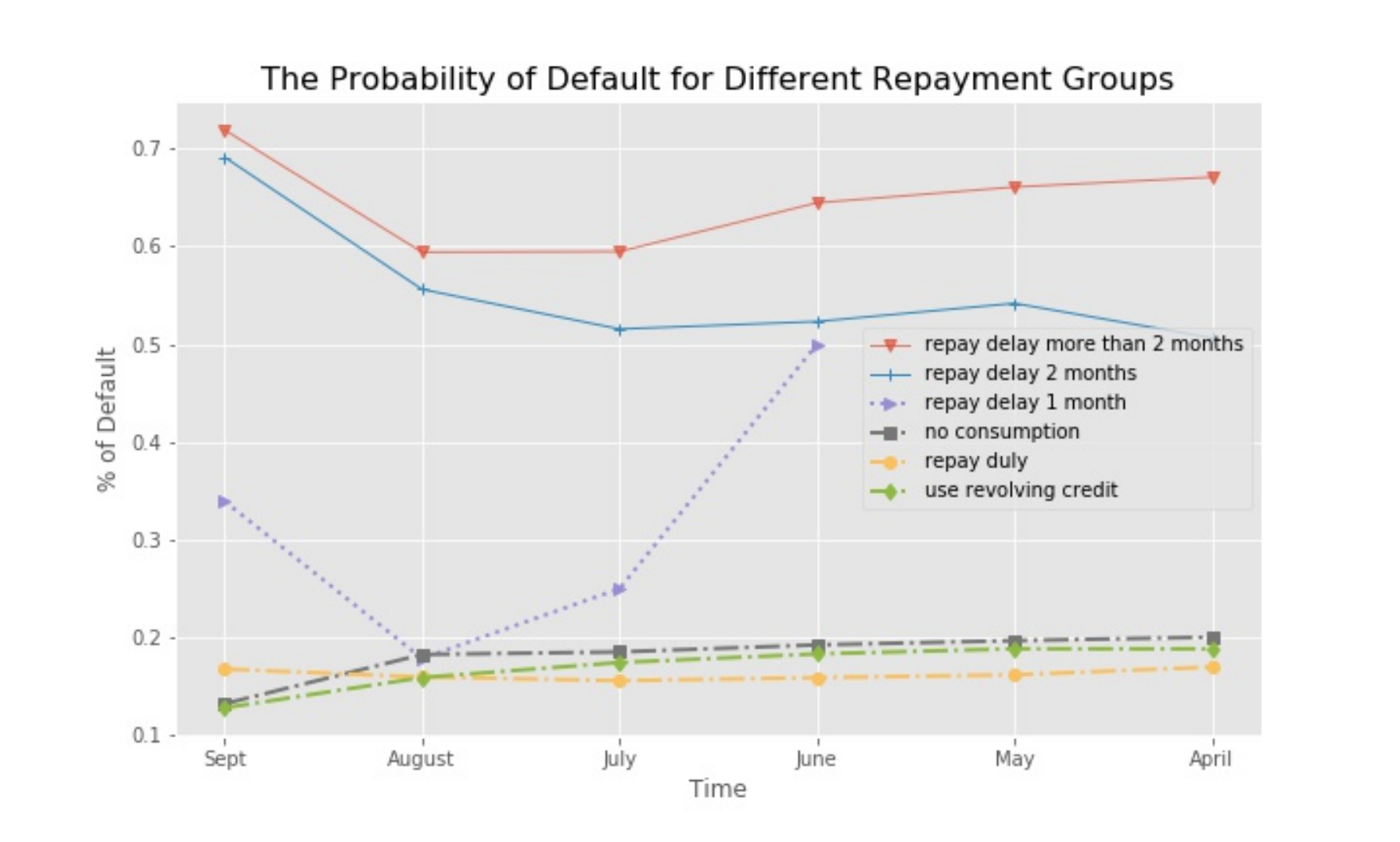
Exploratory Data Analysis

Credit line is positive related to education level



Exploratory Data Analysis

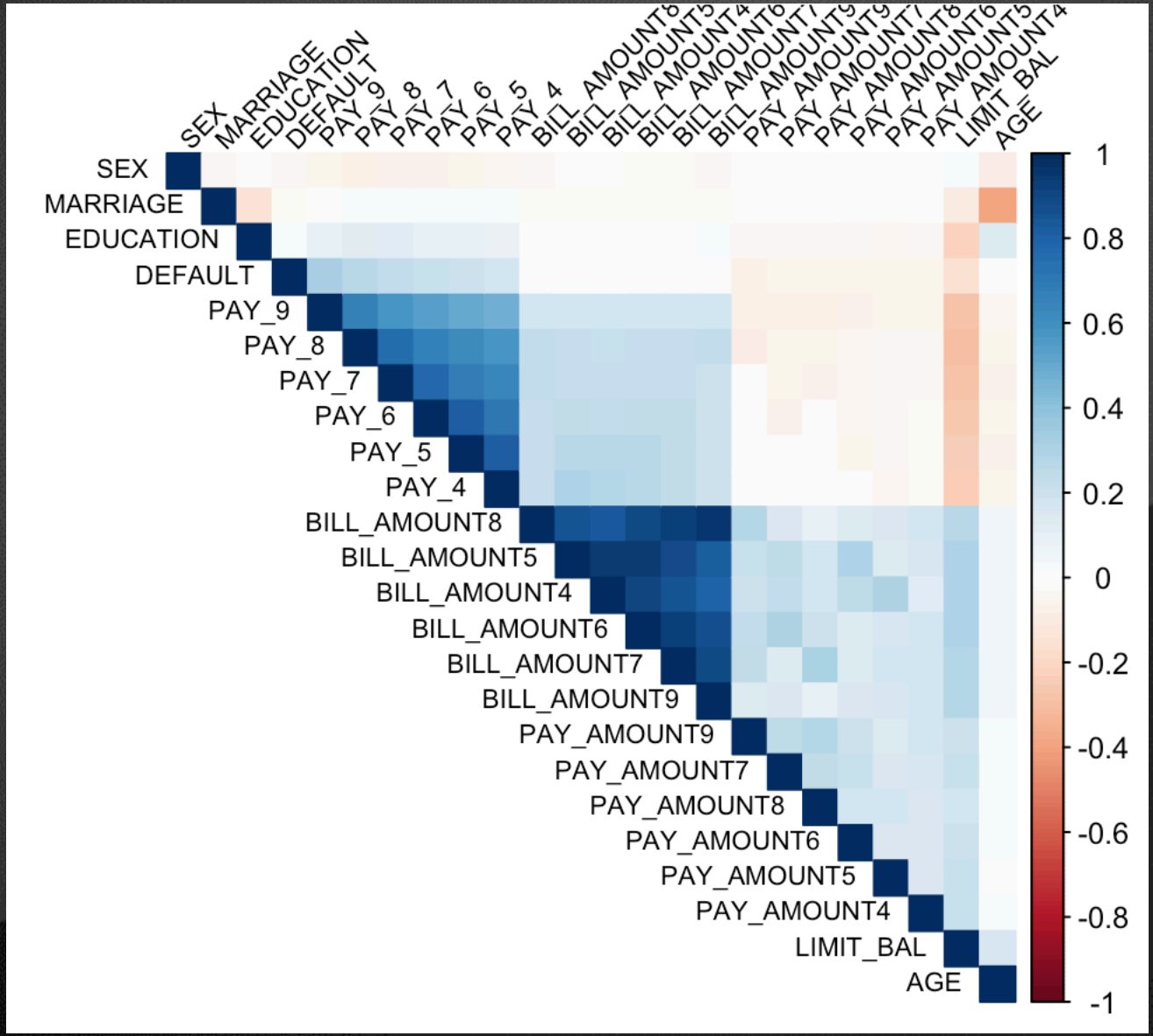
Customers who delayed repayment for at least 2 months have much higher probability to default



Exploratory Data Analysis

Bill amounts of previous six months are strongly correlated.

Repayment status between adjacent months are strongly related.



Modeling

1

DATA PRE-PROCESSING

- Check null data
- Normalization
- Dummy variables
- Split data into training and test sets

2

MODEL TRAINING

- Build SVM
- Feature selection
- Hyper-parameters tuning
- Train model

3

MODEL EVALUATION

- Precision & Recall
- Confusion matrix
- Top 10 positive & Top 10 negative features

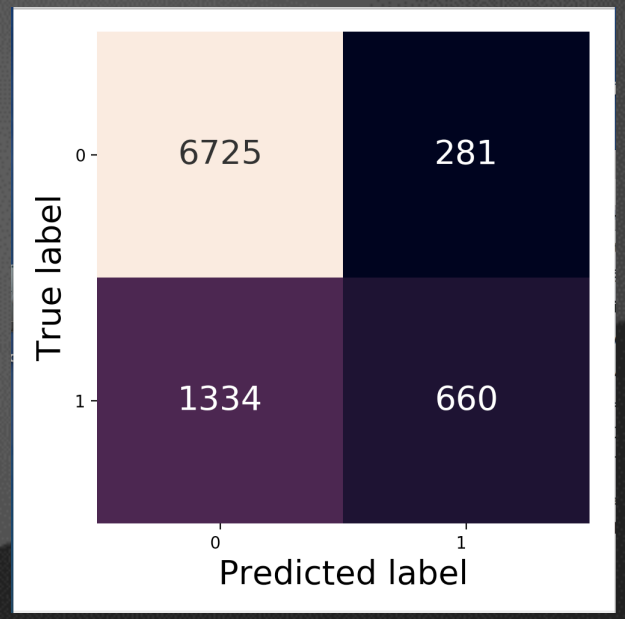
Result

- 1994 default in 9000 testing observations, model detects 941 default.

Classification Report

Class	Precision	Recall	F1-score	Support
0 (non-default)	0.83	0.96	0.89	7006
1 (default)	0.70	0.33	0.45	1994
Avg / Total	0.80	0.82	0.79	9000

Confusion Matrix

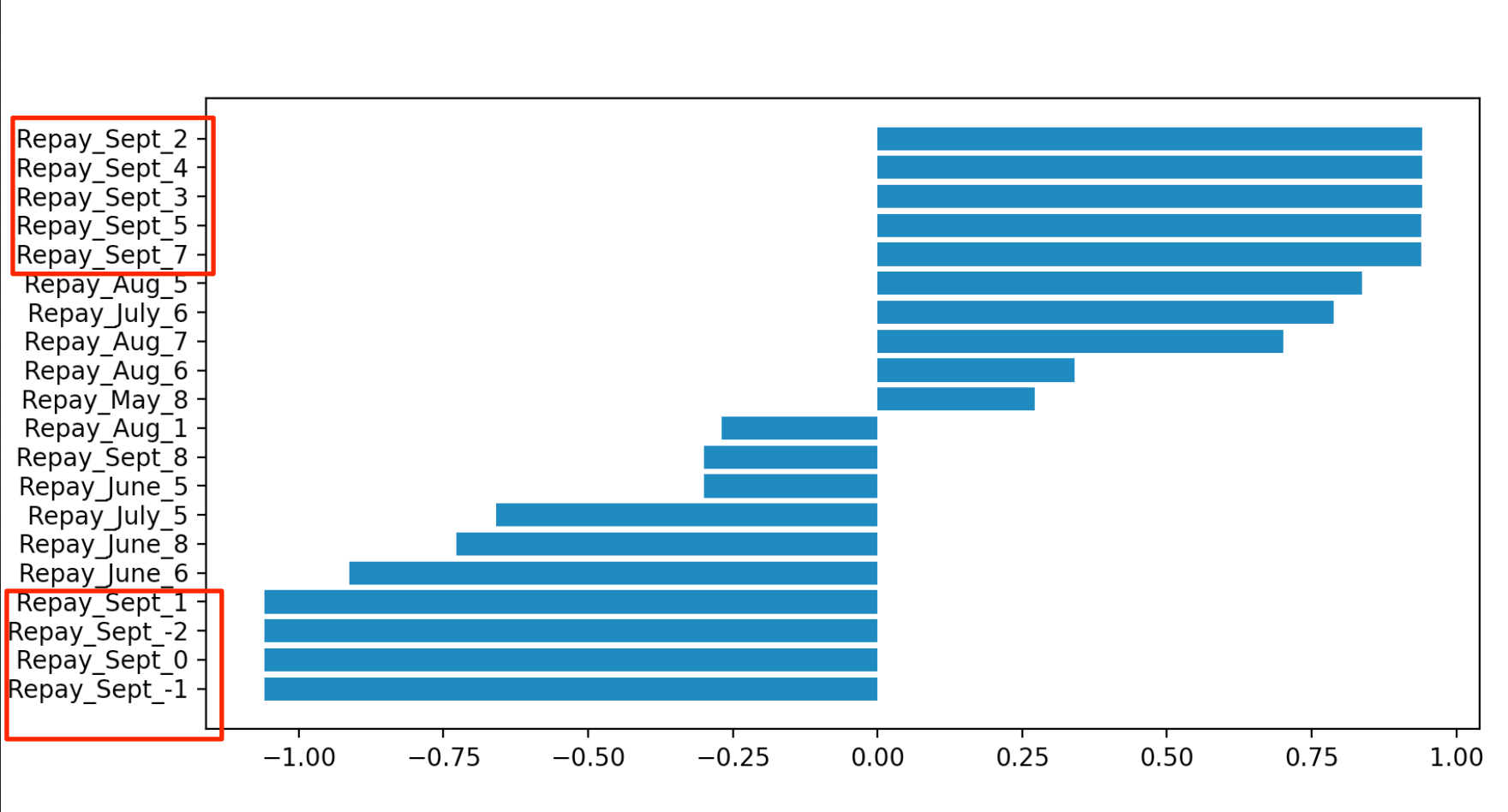


Result

Delay Repayment
more than 1 month in Sept.
is significantly positive features

Repay duly or delay 1 month
in Sept. is significantly
negative features

Top 20 features (top 10 positive and negative coefficient values)



Future Work

- Data Reconstruction
  - Default status of current month is strongly related to repayment status in the previous one month
  - Reconstruct a dataset with much more observations directly from original data

- Time Series
  - Recurrent Neural Network (RNN)
  - Long Short Term Memory (LSTM)

- Reinforcement Learning
  - Markov Decision

THANK YOU

Q & A