



A Classification Analysis of Credit Payments

Caleb Elgut - August 2020



TOC

Introduction

Problems to solve

Visualizations

Findings

Future Research

Vision

Conclusions

Thanks



Introduction

- Target Audience: Bank Managers & Loan Officers
- Visualizations powered by classification models
 - Individual Decision Trees, Bagged Trees, Random Forest, Logistic Regression, K-Nearest Neighbors
- Data Source: UCI Database
 - Includes payment history between April & September 2005 in Taiwan
 - 30,000
 - Features: how far behind one is in payments, amount paid monthly, monthly balance, credit limit
 - Categorical variables: Sex, Education, Marriage Status, Age Group





Problems to solve

1

Given personal data and payment history, what is the basic probability that someone will default on their payment?

2

What features should we focus on when we want to predict whether or not someone will default on their next credit payment?

Note on Age Groups (Who are we dealing with?):

- Minimum: 21
- 1st Quartile: 28
- Median Age: 34
- 3rd Quartile: 41
- Maximum Age: 79



Visualization #1: Basic Probability of Default

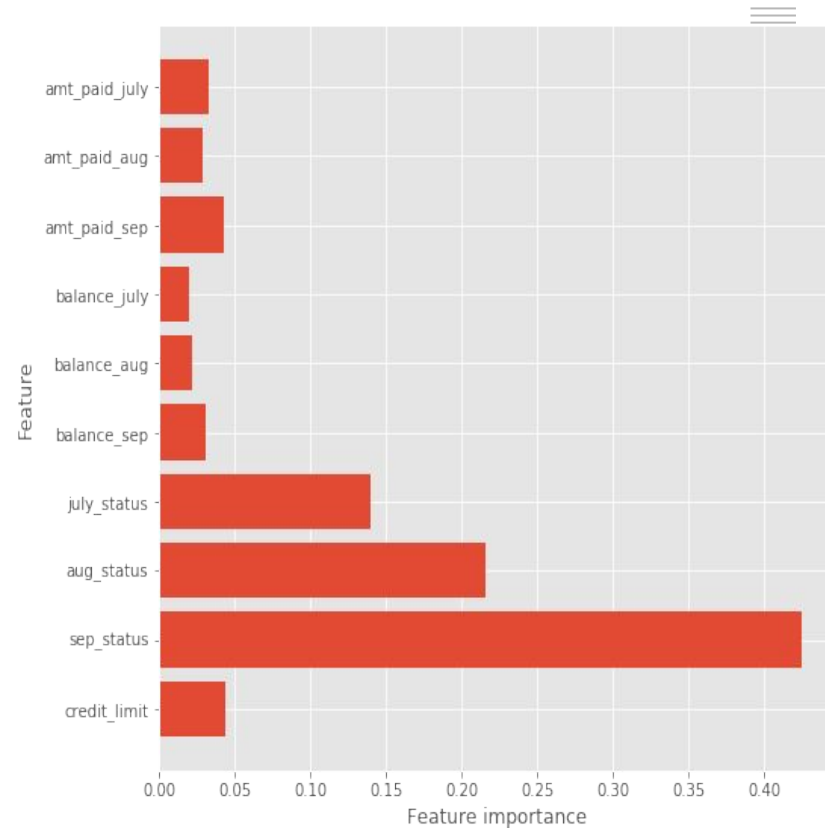
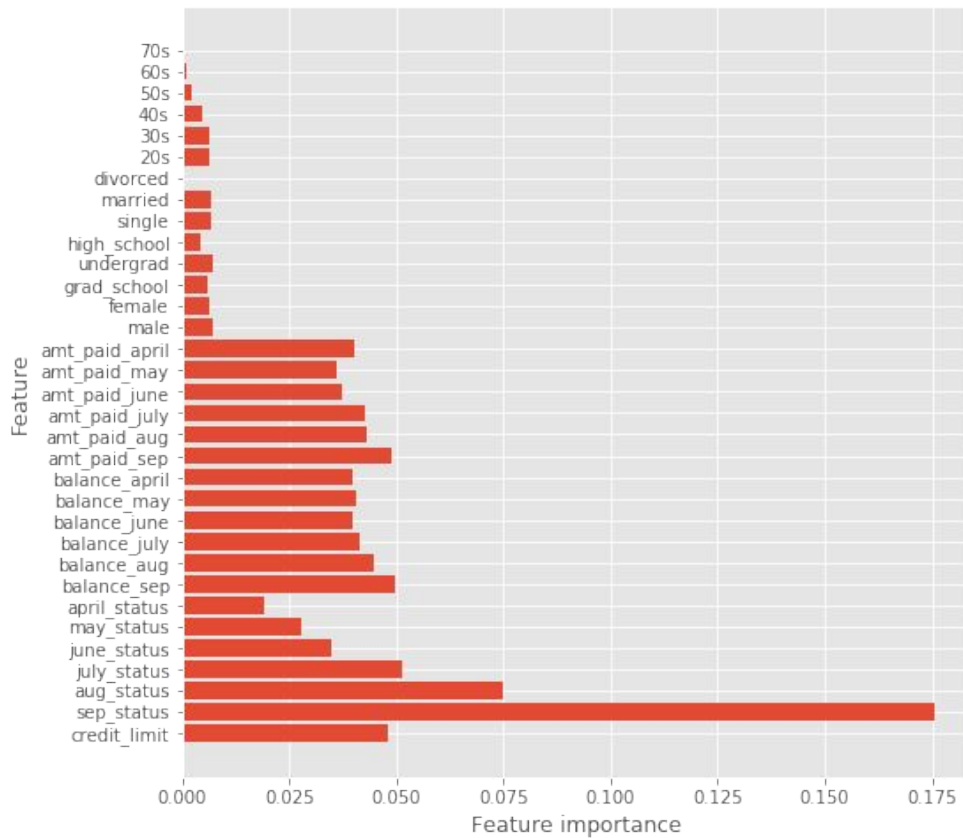
- Over 23,000 made their payments.
- Over 6,600 defaulted.
- Basic Probability of Default: 22%





Features of Import in the Next Slide

- **Status:** Paid up? User of Revolving Credit? # of Months Behind?
- **Balance:** Measured Monthly
- **Amount Paid:** Monthly



Visualizing the Results that Matter for You



Results Analysis in Summary

When analyzing whether or not to give out credit, look at the recent payment status history.



Future Research?

- Aid to those who default
- More recent data
- More categorical data
 - Employment
 - Asset Ownership
 - Number of Children
- Longer timeline
 - Would like to work with data that shows span of years.



Thank you.

