Final Project

1. Tổng quan

346 khách hàng, 10 thuộc tính

1. Các thuộc tính

* Loan\_status: liệu một khoản vay đã được trả trong bộ sưu tập

+ PAIDOFF : đã trả (260)

+ COLLECTION: chưa trả (86)

* Principal ( tiền vốn): tiền vay gốc cơ bản
* Terms : điều khoản khởi đầu có thể là hàng tuần ( 7 ngày) , 2 tuần một lần và lịch thanh toán hàng tháng.
* Effective\_date : khi khoản vay có nguồn gốc và có hiệu lực.

VD: 9/8/2016

* Due\_date: kể từ khi nó lên lịch thanh toán một lần, mỗi khoản vay có một ngày đến hạn

VD: 10/7/2016

* Age : tuổi của người vay
* Education ( Trình độ học vấn): trình độ học vấn của người vay

+ High School or Below

+ Bechalor

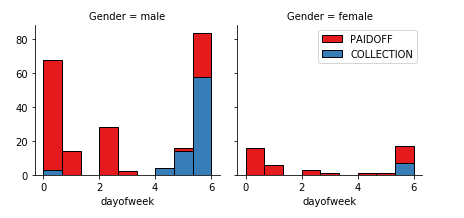
+ college

* Gender ( Giới tính): giới tính của người vay

+ Male

+ Female

1. Tiền xử lí
2. Effective\_date : nhìn vào ngày trong tuần mọi người nhận được khoản vay.



* Những người vay tiền vào cuối tuần không trả hết 🡪 sử dụng feature nhị phân để đặt giá trị ngưỡng ít hơn vào ngày 4 ( < 4 🡪 giá trị weekend = 0, >= 4 giá trị weekend = 1)
* Thêm 1 cột dayofweek, 1 cột weekend

1. Giới tính

Chuyển từ Categorical sang giá trị số

Male 🡪 0

Female 🡪 1

1. Education : one hot encoding 🡪 sử dụng kĩ thuật one hot encoding để chuẩn hóa các biến phân loại thành các biến nhị phân và nối chúng vào Data Frame

* Bechalor
* High School or Below
* Master or Above

COLLECTION : 0.5

PAIDOFF: 0.5

* Không có ý nghĩa gì cả, bỏ thuộc tính này đi.
* College

1. Bỏ 2 cột unname đầu tiên đi
2. Feauture selection

* Chọn ra 8 thuộc tính có ý nghĩa

+ Principal

+ Terms

+ Age

+ gender

+ weekend

+ bechalor

+ high school or below

+ collage

* Nhãn : loan\_status

1. Normalize Data ( Chuẩn hóa dữ liệu)

* Chuẩn hóa về dạng 0,1