Credit Score Analysis

Ajay Shankar A

2023-12-28

Table of contents

## Introduction

#### Welcome to **Analyzing Credit Scores with tidymodels in R**

In this project we will explore the what differentiates consumer credit scores levels and demonstrate how dimensionality reduction can retain much of information in a dataset while reducing it’s size. We’ll use tidymodels and embed to build UMAP and decision tree models.

library(tidyverse)  
library(tidymodels)  
library(embed)

data\_url <- "https://assets.datacamp.com/production/repositories/6081/datasets/e02471e553bc28edddc1fe862666d36e04daed80/credit\_score.csv"  
  
credit\_df <- read\_csv(data\_url)

Rows: 18965 Columns: 23  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (6): month, occupation, credit\_mix, payment\_of\_min\_amount, payment\_beha...  
dbl (17): age, annual\_income, monthly\_inhand\_salary, num\_bank\_accounts, num\_...  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

glimpse(credit\_df)

Rows: 18,965  
Columns: 23  
$ month <chr> "January", "July", "April", "January", "Febru…  
$ age <dbl> 44, 19, 39, 43, 22, 52, 32, 45, 37, 22, 26, 2…  
$ occupation <chr> "Doctor", "Doctor", "Manager", "Developer", "…  
$ annual\_income <dbl> 32625.590, 31041.460, 25876.180, 15928.750, 1…  
$ monthly\_inhand\_salary <dbl> 2922.7992, 2501.7883, 2293.3483, 1301.3958, 1…  
$ num\_bank\_accounts <dbl> 0, 5, 4, 9, 7, 5, 7, 9, 1, 7, 5, 7, 9, 4, 4, …  
$ num\_credit\_card <dbl> 5, 2, 6, 9, 9, 5, 8, 8, 5, 7, 5, 5, 5, 5, 7, …  
$ interest\_rate <dbl> 8, 5, 7, 24, 18, 5, 26, 21, 10, 25, 3, 23, 32…  
$ num\_of\_loan <dbl> 4, 1, 3, 4, 2, 4, 9, 2, 2, 7, 1, 3, 6, 1, 2, …  
$ delay\_from\_due\_date <dbl> 5, 5, 5, 26, 40, 23, 62, 18, 13, 51, 10, 15, …  
$ num\_of\_delayed\_payment <dbl> 10, 11, 15, 22, 17, 1, 18, 18, 3, 16, 7, 21, …  
$ changed\_credit\_limit <dbl> 2.55, 5.13, 10.29, 1.26, 6.09, 0.75, 22.13, 8…  
$ num\_credit\_inquiries <dbl> 3, 0, 4, 12, 12, 1, 12, 13, 3, 8, 0, 9, 12, 1…  
$ credit\_mix <chr> "\_", "Good", "\_", "Bad", "Bad", "Good", "Bad"…  
$ outstanding\_debt <dbl> 177.90, 291.77, 71.54, 2240.56, 2063.45, 636.…  
$ credit\_utilization\_ratio <dbl> 39.41268, 25.26184, 25.24044, 31.18439, 23.13…  
$ payment\_of\_min\_amount <chr> "No", "NM", "NM", "Yes", "NM", "No", "NM", "Y…  
$ total\_emi\_per\_month <dbl> 73.125008, 21.021180, 37.374350, 31.846679, 1…  
$ amount\_invested\_monthly <dbl> 139.43374, 185.31083, 117.93599, 161.38822, 1…  
$ payment\_behaviour <chr> "Low\_spent\_Small\_value\_payments", "Low\_spent\_…  
$ monthly\_balance <dbl> 369.7212, 333.8468, 354.0245, 226.9047, 343.9…  
$ credit\_history\_months <dbl> 344, 398, 257, 145, 161, 199, 31, 121, 244, 1…  
$ credit\_score <chr> "Standard", "Good", "Standard", "Poor", "Poor…