

# CREDIT RISK MANAGEMENT PROJECT

## Pengantar Model Matematika Kelompok 8

**Dosen Pengampu: Dr. Nanang Susyanto, S.Si., M.Sc.**

M. Arifin Ilham (455502), Desy Safitri R. (456575), Banu Suharyadi (455494),  
Amira Latifa K. (459334), & Jerrico Nugroho (459342)

### Abstract

Risiko kredit merupakan risiko yang timbul atas kegagalan dari suatu nasabah dalam membayarkan kewajibannya atas pinjaman yang dilakukan kepada perusahaan. Data yang diberikan merupakan data pinjaman dari nasabah di salah satu perusahaan. Data tersebut terdiri dari 26 kolom yang berisikan seputar profil nasabah dan transaksi pinjaman yang dilakukan nasabah. Tujuan penulisan ini adalah untuk menganalisis serta membuat model klasifikasi untuk mengetahui apakah suatu pinjaman yang dilakukan oleh seorang nasabah adalah baik atau buruk.

Hasil analisis akan digunakan oleh perusahaan untuk melakukan *screening* dalam menerima atau menolak seorang nasabah ketika mengajukan pinjaman. Terdapat empat model yang digunakan untuk pengklasifikasian, yaitu: *logistic regression*, LDA (*linear discriminant analysis*), *decision tree*, dan *random forest*. Metode evaluasi model yang digunakan adalah *confusion matrix*. Hasil simulasi menunjukkan bahwa *decision tree* merupakan model yang paling baik digunakan untuk mengetahui suatu pinjaman oleh nasabah baik atau buruk sehingga perusahaan mampu menerima atau menolak seorang nasabah mengajukan pinjaman.

**Kata kunci:** Klasifikasi, *Logistic Regression*, LDA, *Decision Tree*, *Random Forest*

### 1 Pendahuluan

Pengklasifikasian merupakan salah satu metode untuk mengelompokkan, menyeparasi, atau mengklasifikasi suatu data yang disusun secara sistematis sesuai target kelompok yang diinginkan. Masalah klasifikasi dapat diterapkan dan dijumpai pada berbagai bidang kehidupan baik itu pengklasifikasian data pada bidang akademik, sosial, pemerintah maupun pada bidang lainnya. Klasifikasi ini diperlukan ketika terdapat sejumlah data dengan berbagai kategori, yang tidak dapat diidentifikasi langsung secara kasat mata sehingga harus diklasifikasikan atau dikelompokkan menggunakan suatu ukuran agar didapat interpretasi yang mudah dipahami.

### 2 Latar Belakang

Risiko kredit (*credit risk*) merupakan risiko yang disebabkan oleh nilai jaminan yang tidak memenuhi ketentuan atau

karena tidak tepat dalam memperhitungkan potensi gagal bayar atau gagal serah kepada lawan transaksi. Pengambilan keputusan debitur dalam menentukan diterima atau tidaknya pinjaman kredit dari kreditur didasarkan pada beberapa aspek yang diperhitungkan oleh debitur. Dalam pengambilan keputusan tersebut dapat dilakukan dengan cara klasifikasi. Terdapat beberapa model klasifikasi yang dapat digunakan antara lain *logistic regression*, *linear discriminant analysis* (LDA), *decision tree*, *random forest*, dan lain sebagainya.

Dalam kasus risiko kredit, klasifikasinya bertipe *binary*, yaitu *class* 1 dan 0. *Class* 0 mendefinisikan pinjaman kredit ditolak dan *class* 1 mendefinisikan pinjaman kredit diterima. Penentuan kreditur berada dalam kelas-kelas tersebut didasarkan pada parameter-parameter mengenai kreditur tersebut.

Pemberian keputusan pinjaman kredit yang salah akan mengakibatkan kerugian pada debitur maka dibutuhkan model klasifikasi dengan akurasi yang tinggi atau model yang menghasilkan nilai *false positive* sekecil mungkin. Dalam *project* ini, model yang digunakan untuk mengklasifikasikan kreditur pada awalnya adalah *logistic regression*. Namun, penulis lebih mengandalkan *decision tree* dalam model klasifikasi karena tiap hubungan fitur mudah ditangkap oleh model dan mudah dipahami. Selain itu, terdapat model lain seperti LDA dan *random forest* yang digunakan sebagai pembanding. Model-model ini yang di-*training* dan di-*testing* dengan data yang telah diberikan.

### 3 Identifikasi Paramater

Perusahaan menggunakan sejumlah data baik data internal maupun data nasabah yang dapat digunakan sebagai bahan pertimbangan dalam mengelola risiko kredit. Data yang tercantum diantaranya adalah sebagai berikut.

- ID (*Identity*), yaitu identitas (*unique number*) dari transaksi pinjaman dari seorang nasabah
- Member ID*, yaitu identitas (*unique number*) dari seorang nasabah
- Loan amount*, yaitu jumlah pinjaman dari nasabah.
- Funded amount*, yaitu jumlah pinjaman yang disetujui oleh perusahaan.
- Term*, yaitu lama waktu (*tenor*) dari pinjaman yang diberikan oleh perusahaan kepada nasabah.

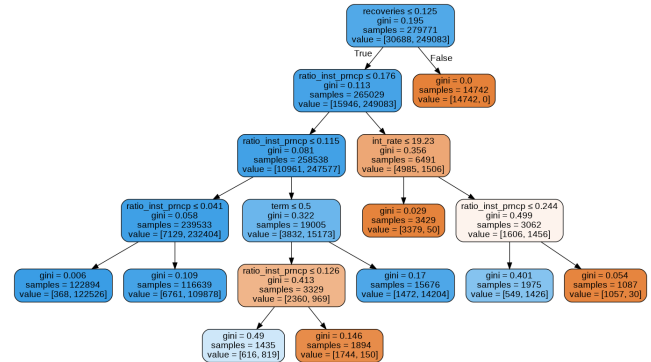
- f. *Interest rate*, yaitu bunga yang diberikan perusahaan kepada nasabah.
- g. *Installment*, yaitu cicilan yang dibayarkan oleh nasabah.
- h. *Grade*, yaitu klasifikasi penilaian internal (*rating*) dari pinjaman seorang nasabah.
- i. *Subgrade*, yaitu detail klasifikasi penilaian internal (*rating*) dari pinjaman seorang nasabah.
- j. *Employment title* merupakan level pekerjaan dari seorang nasabah.
- k. *Employment length* merupakan lama pengalaman bekerja dari seorang nasabah.
- l. *Home ownership*, yaitu status kepemilikan rumah dari seorang nasabah.
- m. *Annual income* merupakan pendapatan tahunan dari seorang nasabah.
- n. *Verification status*, yaitu status verifikasi pinjaman nasabah oleh perusahaan.
- o. *Purpose* merupakan tujuan dari pinjaman yang dilakukan nasabah.
- p. *Title* merupakan detail tujuan dari pinjaman yang dilakukan nasabah.
- q. *Zip code* merupakan kode pos dari alamat nasabah.
- r. *Address state* merupakan kode alamat nasabah.
- s. *DTI* (Debt to income) merupakan perbandingan antara pinjaman dan pendapatan dari nasabah.
- t. *Total accrual* merupakan jumlah transaksi pinjaman yang dimiliki oleh seorang nasabah.
- u. *Total payment* merupakan total cicilan yang sudah dibayarkan oleh nasabah.
- v. *Total recoveries principal* merupakan biaya yang diterima perusahaan berkaitan dengan kelalaian pembayaran biaya pokok dari pinjaman yang dilakukan nasabah.
- w. *Recoveries* merupakan biaya yang diterima kembali oleh perusahaan ketika suatu nasabah sudah gagal membayar pinjaman yang dilakukan.
- x. *Last payment amount* merupakan cicilan terakhir yang dibayarkan oleh nasabah kepada perusahaan.
- y. *Application type* merupakan tipe pinjaman yang dilakukan nasabah.
- z. *Loan status class* merupakan status klasifikasi pinjaman suatu nasabah. Bernilai nol (0) jika status pinjaman nasabah tersebut buruk dan bernilai satu (1) jika status pinjaman nasabah tersebut baik.

Dalam meminimalkan risiko kredit, penulis menggunakan beberapa parameter yang memiliki korelasi dengan kemampuan para nasabah untuk mengembalikan pinjamannya.

## 4 Analisis

Untuk mendapatkan pemahaman atas hubungan fitur dengan kelas target dan untuk mengetahui fitur apa saja yang penting dalam klasifikasi, penulis menggunakan metode *backward selection* dengan model *decision tree* yang menggunakan hampir seluruh fitur yang kemudian diseleksi dan dioptimisasi sehingga didapat fitur-fitur yang penting.

Dari model *decision tree* yang dibuat sudah cukup optimal, diperoleh diagram sebagai berikut.

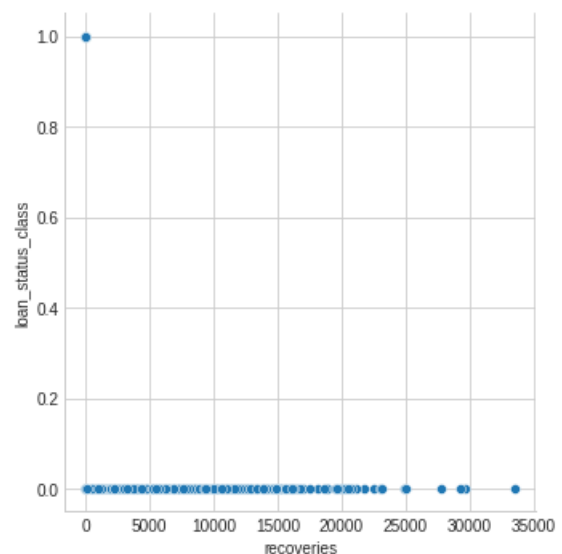


Dengan catatan, "*ratio inst prncp*" merupakan rasio dari *installment* + *c* dan *total recoveries principal* + *c* dengan *c* bilangan tak nol yang dibuat sekecil mungkin untuk mengatasi masalah di mana data *total recoveries principal* bisa bernilai nol. Peneliti mengambil nilai *c* = 0.0001 pada penelitian ini.

Dari diagram di atas, penulis berhasil mengidentifikasi empat klaster populasi kelas nol (kelompok populasi di mana kebanyakan anggota kelompok tersebut merupakan kelas nol) yang ditandai dengan *node* berwarna jingga.

Keempat klaster kelas nol yang berhasil diidentifikasi tersebut adalah sebagai berikut.

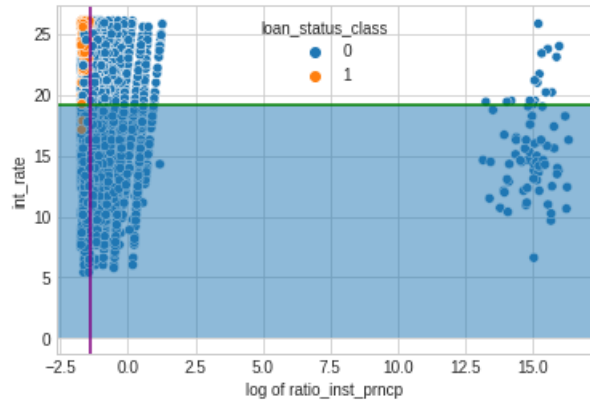
- a. Kelompok dengan nilai *recoveries* tidak sama dengan nol. Hal ini dibuktikan dengan grafik dari data keseluruhan berikut.



Karena *recoveries* adalah biaya yang diterima perusahaan

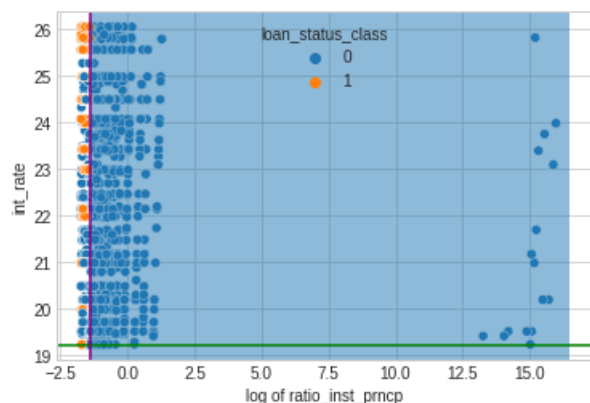
ketika nasabah sudah gagal bayar maka nasabah yang memiliki data *recoveries* bernilai tak nol bisa dipastikan sebagai nasabah gagal bayar.

- b. Kelompok dengan nilai *recoveries* sama dengan nol, nilai *ratio inst prncp* > 0.176 dan *interest rate* ≤ 19.23. Hal ini dibuktikan dengan grafik dari data dengan nilai *ratioinstprncp* > 0.176 berikut.



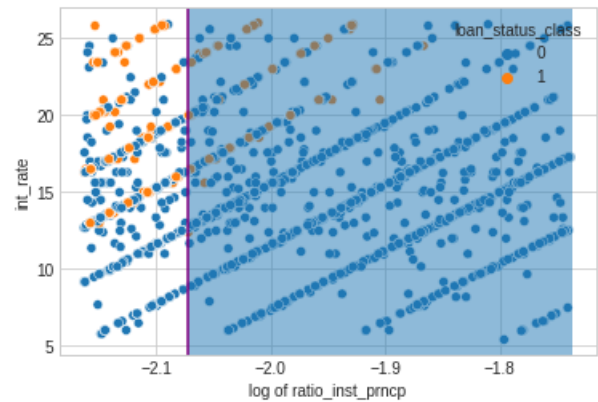
Garis ungu merupakan garis *ratio inst prncp* = 0.244 yang disesuaikan dengan *log ratio inst prncp*, sedangkan garis hijau merupakan garis *interest rate* = 19.23. Kedua garis ini merupakan garis pemisah yang ditentukan oleh *decision tree*.

- c. Kelompok dengan nilai *recoveries* sama dengan nol, nilai *ratio inst prncp* > 0.244 dan *interest rate* > 19.23. Hal ini dibuktikan dengan grafik dari data dengan nilai *ratioinstprncp* > 0.176 dan *interest rate* > 19.23 sebagai berikut.



Garis ungu merupakan garis *ratio inst prncp* = 0.244 yang disesuaikan dengan *log ratio inst prncp*, sedangkan garis hijau merupakan garis *interest rate* = 19.23. Kedua garis ini merupakan garis pemisah yang ditentukan oleh *decision tree*, sama seperti pada kluster sebelumnya.

- d. Kelompok dengan nilai *recoveries* sama dengan nol, nilai *ratio inst prncp* ≤ 0.176, tetapi juga *ratio inst prncp* > 0.126 dan *term* bernilai 36 bulan. Hal ini dibuktikan dengan grafik dari data dengan nilai *ratio inst prncp* berada di antara 0.115 dan 0.176 dan *term* 36 bulan sebagai berikut.

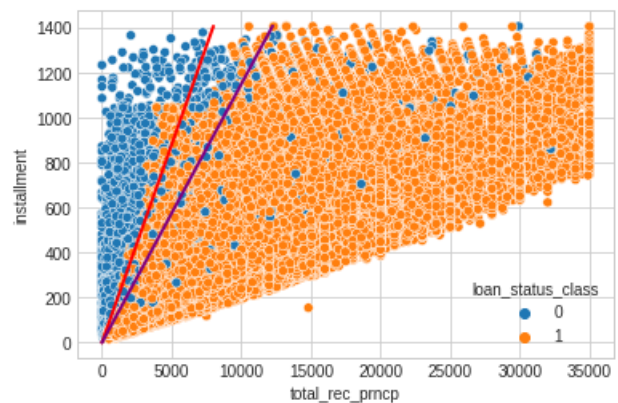


Garis ungu merupakan garis *ratio inst prncp* = 0.126 yang disesuaikan dengan *log ratio inst prncp*. Garis ini merupakan garis pemisah yang ditentukan oleh *decision tree*.

Pada kluster ini, garis pemisah tersebut tidak sebaik pada kluster-kluster sebelumnya dan masih bisa diperbaiki pada penelitian selanjutnya.

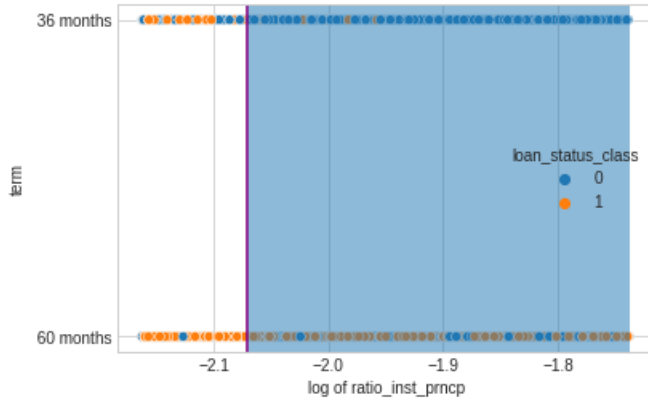
Dengan meninjau keempat kluster tersebut, penulis mengidentifikasi hubungan antara beberapa fitur dengan kelas target sebagai berikut.

- a. Terdapat hubungan antara fitur *installment* dan *total recoveries principal* dengan kelas target, sebagaimana bisa dilihat pada grafik dari data dengan *recoveries* sama dengan nol sebagai berikut.



Garis merah merupakan garis yang melewati titik *origin* dengan gradien bernilai 0.176, sedangkan garis ungu merupakan garis yang juga melewati titik *origin* dengan gradien bernilai 0.115. Jika data berada di sebelah kiri garis merah, probabilitas data tersebut kelas nol cukup besar. Lebih lanjut, jika data berada di antara garis merah dan ungu, data tersebut bisa dilakukan segmentasi lebih lanjut sehingga proses klasifikasi kelas target lebih akurat. Letak data terhadap garis merah dan ungu bisa ditentukan melalui rasio dari *installment* dengan *total recoveries principal* yang disesuaikan (*ratio inst prncp*). Jika suatu data memiliki rasio yang lebih tinggi dari 0.176, data tersebut berada di sebelah kiri garis merah. Suatu data juga akan terletak di antara garis merah dan ungu jika nilai rasio berada di antara 0.115 dan 0.176. Hal ini juga merupakan justifikasi lain dari kluster kedua dan ketiga.

- b. Terdapat hubungan antara fitur *term* dan *total recoveries principal* dengan kelas target, sebagaimana bisa dilihat pada grafik dengan nilai *ratio inst prncp* berada di antara 0.115 dan 0.176 (data yang diapit oleh garis merah dan ungu pada grafik sebelumnya) sebagai berikut.



Garis ungu merupakan garis *ratio inst prncp* = 0.126 yang disesuaikan dengan *log ratio inst prncp*. Dari grafik tersebut, terlihat bahwa data dengan *term* 36 bulan lebih mudah diklasifikasi dibandingkan data dengan *term* 60 bulan. Lebih lanjut, jika data memiliki *term* 36 bulan dan berada di sebelah kanan garis ungu, probabilitas data tersebut kelas nol cukup besar. Hal inilah yang merupakan salah satu justifikasi dari kluster keempat.

Diperoleh fitur *recoveries* dan *interest rate* serta fitur sintetis *ratio inst prncp* sebagai fitur yang signifikan dalam penentuan klasifikasi data pada model *decision tree*.

Selanjutnya, pada model klasifikasi *logistic regression* dan *linear discriminant analysis* (LDA), penulis mencoba menggunakan tujuh fitur, di antaranya: (1) *loan amount*, (2) *total payment*, (3) *total recoveries principal*, (4) *recoveries*, (5) *installment*, (6) *last payment amount*, dan (7) *grade*. Sementara itu, pada model klasifikasi *decision tree* dan *random forest*, penulis menggunakan seluruh fitur yang kemudian dioptimisasi dan diseleksi.

## 5 Interpretasi Model

Pada *project* ini model yang digunakan untuk mengklasifikasi kreditur terdapat empat (4) jenis model di antaranya:

### a. Logistic Regression

Dari model ini diperoleh *confusion matrix*-nya adalah

	0	1
0	4857	5229
1	574	82597

Table 1: *Confusion Matrix Logistic Regression.*

Selanjutnya, dapat diperoleh:  
*Accuracy* = 0.937774108109847  
*Precision* = 0.48155860

*Specificity* = 0.94046182  
*Sensitivity* = 0.89431044

### b. LDA ( Linear Discriminant Analysis )

Dari model ini diperoleh *confusion matrix*-nya adalah

	0	1
0	2824	7262
1	10	83161

Table 2: *Confusion Matrix LDA.*

Selanjutnya, dapat diperoleh:

*Accuracy* = 0.9220219393718434  
*Precision* = 0.27999207  
*Specificity* = 0.91968857  
*Sensitivity* = 0.99647142

### c. Decision Tree

Dari model ini diperoleh *confusion matrix*-nya adalah

	0	1
0	8530	1556
1	271	82900

Table 3: *Confusion Matrix Decision Tree.*

Selanjutnya, dapat diperoleh:

*Accuracy* = 0.9804089773421835  
*Precision* = 0.84572675  
*Specificity* = 0.98157621  
*Sensitivity* = 0.96920804

### d. Random Forest

Dari model ini diperoleh *confusion matrix*-nya adalah

	0	1
0	7694	2392
1	117	83054

Table 4: *Confusion Matrix Random Forest.*

Selanjutnya, dapat diperoleh:

*Accuracy* = 0.9730958533943833  
*Precision* = 0.76283958  
*Specificity* = 0.97200571  
*Sensitivity* = 0.98502112

Diperoleh kesimpulan, dengan menggunakan *confusion matrix* dapat mengetahui keakuratan dari model yang dibuat dengan *performance matrix* seperti: *accuracy*, *precision*, *specificity* dan *sensitivity*, serta dapat dilihat bahwa model yang memiliki tingkat *accuracy* yang cukup besar adalah *decision tree* sebesar 0.9804089773421835.

## 6 Video

Penjelasan lebih lanjut mengenai *credit risk management project* dapat diakses melalui tautan berikut:  
<https://youtu.be/5WxBYWihoig>

## References

Joseph M. Hilbe. 2015. Practical Guide to Logistic Regression. New York: CRC Press.

Sheldon M. Ross. 2010. Introduction to Probability Models Tenth Edition. Los Angeles, California: Elsevier.