

**Analisis Deskriptif dan Pemilihan Model Pohon
Keputusan untuk menentukan Kualitas Nasabah
Peminjam Dana Menggunakan RStudio**



Alha Ahmada Belasae

16.8993

Kelas 3 KS 1

Tahun 2018

Latar Belakang

Perekonomian barter telah menimbulkan berbagai kesulitan sehingga tidak mengherankan apabila manusia primitif mulai mencari cara-cara yang lebih mudah dalam melaksanakan tukar-menukar barang, misalnya dengan menggunakan barang-barang berharga yang banyak disukai masyarakat luas. Akan tetapi, untuk membawa barang-barang berharga ke sana sini adalah berbahaya. Resiko dalam perjalanan terlalu besar. Dalam suasana itu, kredit dalam bentuknya yang sangat terbatas mulai dikenal, yakni sebagai pengurang resiko tersebut apabila harus membawa uang secara fisik dan dalam jumlah besar.

Semenjak uang mulai dikenalkan di Indonesia dengan nama Oeang Republik Indonesia atau ORI pada 1946 lalu, masyarakat sudah terlalu kreatif untuk meningkatkan efisiensi dan efektifitas kerja dari uang. Bahkan, sistem peminjaman baik bersifat formal maupun non-formal sudah tak asing lagi terdengar di masyarakat. Sistem tersebut mulai bekerja di Bank-Bank Indonesia sejak Tahun 1998 memberikan keleluasaan bagi rakyat untuk dapat mencukupi kebutuhannya dengan “ngutang” atau meminjam dengan konsekuensi harus mengembalikan dengan nominal sesuai dengan kesepakatan.

Bermunculan setelah itu perkembangan syarat-syarat perlu dan syarat-syarat cukup yang harus dipenuhi oleh seseorang ketika ingin meminjam uang di suatu bank. Yang terlewat yaitu syarat 5C, yaitu *capital*, *capacity*, *collateral*, *character*, dan *condition*. Tak segan-segan si pemohon biasanya menggunakan barang-barang berharga mereka sebagai jaminan, seperti mobil, motor, surat tanah, bahkan rumah.

Kebutuhan manusia yang tidak ada habisnya membuat manusia mudah berhasrat memiliki uang untuk membeli sesuatu. Biasanya, peminjaman dilakukan oleh seseorang untuk membeli barang-barang tertentu seperti mobil, biaya pendidikan, rumah, dll. Tulisan ini ingin menunjukkan bagaimana permodelan *decision tree* yang dapat dilakukan untuk menentukan kelas atau kualitas nasabah peminjam berdasarkan tujuan tertentu.

Tujuan

Tujuan dari tulisan ini adalah

1. Untuk melengkapi tugas mata kuliah Data Mining and Knowledge Management
2. Analisis deskriptif dari data kredit 1000 orang
3. Menganalisis permodelan pohon keputusan berdasarkan tujuan penggunaannya.

Deskripsi Dataset

Data yang akan digunakan dalam penelitian ini merupakan data mentah yang diambil dari <https://www.openml.org/search?type=data&from=100>. Data tersebut memiliki 1000 tuple yang akan peneliti kategorikan menjadi 10 kategori menurut tujuan penggunaan peminjaman uang, yaitu,

- a. Bisnis
- b. *New car*
- c. *Used car*
- d. Pendidikan
- e. Furnitur
- f. Perbaikan
- g. Elektronik
- h. *Domestics*
- i. *Retraining*
- j. *Others*

Selain itu, data pada penelitian kali ini mempunyai 21 atribut yaitu,

- a. *Duration* : Durasi pengembalian yang ditentukan masing-masing nasabah
- b. *Credit_history* : Riwayat peminjaman
- c. *Purpose* : Tujuan peminjaman
- d. *Credit_amount* : Jumlah dana yang dipinjam
- e. *Saving_status* : Status tabungan menurut periode bulanan
- f. *Employment* : Status kerja (dalam tahun)
- g. *Installment_commitment* :
- h. *Personal_status* : JK, Status pernikahan
- i. *Other_parties* : properti yang dijadikan jaminan oleh nasabah kepada bank
- j. *Residence_since* :

- k. *Property_magnitude* : properti yang dijadikan jaminan oleh nasabah kepada bank
- l. *Age* : Umur nasabah
- m. *Other_payment_plants* : Rencana pengembalian lainnya (boleh tidak ada)
- n. *Housing* : Status kepemilikan rumah nasabah
- o. *Existing_credits* : Kredit/peminjaman lain yang dimiliki nasabah yang bersangkutan
- p. *Job* : Level pekerjaan nasabah
- q. *Num_dependents* : Jumlah orang yang dibiayai nasabah
- r. *Own_telephone* : Status kepemilikan telepon
- s. *Foreign_worker*
- t. *Class* : Klasifikasi nasabah (*good/bad*)

Metode

Penelitian kali ini memiliki alur sebagai berikut:

1. Menentukan Dataset

Jujur saja, peneliti mencari dataset terlebih dahulu ketika ingin melakukan penelitian. Peneliti sudah mencoba meneliti 5 buah dataset namun kebingungan apa yang akan diteliti dari dataset tersebut. Akhirnya di pencarian ke-6, peneliti sudah menentukan apa yang akan diteliti dari data kredit bank ini.

2. Studi Literatur

Studi literatur digunakan untuk mempelajari data, deskripsi data, teori mengenai penelitian yang akan digunakan, dan juga menentukan urutan penelitian dengan memilih software yang terbaik menurut peneliti.

3. Menentukan tujuan penelitian

Tujuan penelitian yang disusun sebagaimana di Bab 1 poin 3 ditentukan oleh peneliti setelah melakukan studi literatur dan tentu saja ketika dataset sudah ditemukan pada langkah-langkah sebelumnya.

4. Analisis Deskriptif data gabungan

Analisis deskriptif digunakan untuk menentukan karakteristik data yang akan diteliti. Analisis yang akan digunakan antara lain dengan mencari rata-rata, varians, dsb. Yang dirasa dapat menunjang langkah penelitian selanjutnya.

5. Pemodelan data yang sudah dikategorikan menurut tujuan penggunaan kredit dengan metode *Decision Tree*.

6. Menyusun Kesimpulan

Diakhir penelitian, peneliti akan menyampaikan kesimpulan dari penelitian yang telah digunakan yang diharapkan selanjutnya dapat menjadi acuan penelitian selanjutnya yang serupa atau dapat digunakan dalam rujukan penelitian yang lainnya.

Pembahasan dan Hasil

1. Analisis Deskriptif data

Setelah diinput kedalam program Rstudio dengan *syntax* sebagai berikut

```
crtfull <- read.csv("C:/Users/X441S/Downloads/dataset_31_credit-full.csv")
```

Dan juga dilihat karakter dari masing-masing atribut, diperoleh ringkasan dari beberapa atribut yang penting

Dari *syntax* >summary(crtfull) sebagai berikut.

a. Atribut *duration*

Rata-rata peminjam meminjam dana ke bank dengan angsuran sebanyak 18 kali atau 20,9 bulan. Dengan jumlah angsuran paling sedikit adalah 4 bulan dan jumlah angsuran maksimal adalah 72 bulan (6 Tahun). Hal itu tentunya dilihat dari atribut employment dan saving_status dari masing-masing nasabah dengan juga memperhatikan kondisi dari nasabah tersebut.

b. Atribut *credit_amount*

Dari 1000 nasabah, rata-rata mereka meminjam sejumlah 3271 USD.

Apabila di rupiahkan maka sekitar 49 juta (1 USD = Rp. 15.000).

c. Atribut *purpose*

Para nasabah ternyata paling banyak menggunakan kredit untuk membeli barang-barang elektronik seperti TV, radio, dll. Tercatat sebanyak 28% nasabah menggunakan pinjaman bank untuk hal tersebut. Disusul dengan pembelian mobil baru dan furnitur yang masing-masing digunakan oleh 23,4% dan 18,1% dari total 1000 nasabah.

Dengan rata-rata pinjaman sejumlah 3271 USD dan rata-rata jumlah angsuran sebanyak 21 kali, maka nasabah rata rata dalam sebulan mengembalikan sebanyak 155-156 USD dengan bunga setiap bulan adalah 5%, yang menjadi 163,55 USD perbulan. Keuntungan ini tentunya menjadi daya tarik tersendiri bagi bank-bank baik negeri maupun swasta untuk menawarkan pinjaman kepada nasabahnya. Bisnis ini dirasa sangat menguntungkan apalagi melihat jumlah nasabah yang ingin meminjam uang di bank yang tidak sedikit, sehingga akan memberikan keuntungan tambahan yang sangat besar bagi bank yang bersangkutan.

2. Analisis Pemodelan Data dengan *Decision Tree*

Pemodelan akan dilakukan dari data yang mempunyai kuantitas terbesar yaitu data kredit dengan tujuan untuk membeli barang elektronik yang dilakukan oleh 28% dari 1000 nasabah yang terdapat dalam dataset penelitian kali ini.

Pemodelan akan dilakukan dalam metode *decision tree*, dengan variabel yang digunakan adalah sebagai berikut:

- i. Credit_amount
- ii. Class
- iii. Job
- iv. Housing

Syntax tree

```
> library(tree)
> select_rows<-sample(1:nrow(crtel), round(0.2:nrow(crtel)), replace =
F)
> select_rows
integer(0)
> testt<-crtel[select_rows,]
> trainn<-crtel[-(select_rows),]
> trainn
[1] checking_status      duration
[3] credit_history        purpose
[5] credit_amount         savings_status
[7] employment            installment_commitment
[9] personal_status       other_parties
[11] residence_since        property_magnitude
[13] age                   other_payment_plans
[15] housing               existing_credits
[17] job                   num_dependents
[19] own_telephone         foreign_worker
[21] class
<0 rows> (or 0-length row.names)
> RegTree<-tree(class~job+credit_amount+housing, data = crtel)
```

```

> plot(RegTree)
> text(RegTree, pretty = 5, cex = 1)
> pred <- predict(RegTree, newData = testt)
> head(pred, n = 5)

```

	bad	good
1	0.0000000	1.0000000
2	0.0000000	1.0000000
3	0.0000000	1.0000000
4	0.2142857	0.7857143
5	0.0000000	1.0000000

```

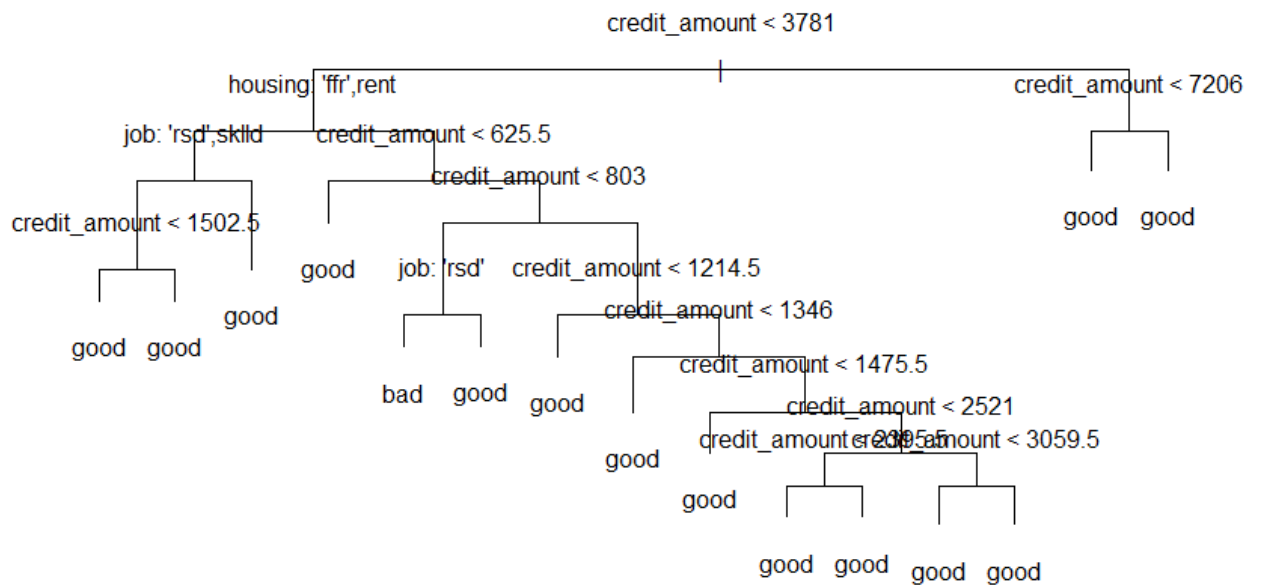
> pred

```

	bad	good
1	0.00000000	1.00000000
2	0.00000000	1.00000000
3	0.00000000	1.00000000
4	0.21428571	0.7857143
5	0.00000000	1.00000000
6	0.21428571	0.7857143
7	0.00000000	1.00000000
8	0.00000000	1.00000000
9	0.00000000	1.00000000
10	0.00000000	1.00000000
11	0.21428571	0.7857143
12	0.00000000	1.00000000
13	0.00000000	1.00000000
14	0.00000000	1.00000000
15	0.66666667	0.3333333
16	0.25000000	0.7500000
17	0.25000000	0.7500000
18	0.25000000	0.7500000
19	0.21428571	0.7857143
20	0.66666667	0.3333333
21	0.25000000	0.7500000
22	0.66666667	0.3333333
23	0.25000000	0.7500000
24	0.66666667	0.3333333
25	0.21428571	0.7857143
26	0.66666667	0.3333333
27	0.25000000	0.7500000
28	0.66666667	0.3333333
29	0.66666667	0.3333333
30	0.25000000	0.7500000
31	0.66666667	0.3333333
32	0.66666667	0.3333333
33	0.25000000	0.7500000
34	0.03571429	0.9642857
35	0.21428571	0.7857143
36	0.03571429	0.9642857
37	0.03571429	0.9642857
38	0.03571429	0.9642857
39	0.03571429	0.9642857
40	0.03571429	0.9642857
41	0.03571429	0.9642857
42	0.03571429	0.9642857
43	0.03571429	0.9642857
44	0.03571429	0.9642857
45	0.03571429	0.9642857
46	0.03571429	0.9642857
47	0.03571429	0.9642857

48	0.03571429	0.9642857
49	0.03571429	0.9642857
50	0.03571429	0.9642857
51	0.00000000	1.00000000
52	0.03571429	0.9642857
53	0.21428571	0.7857143
54	0.03571429	0.9642857
55	0.03571429	0.9642857
56	0.03571429	0.9642857
57	0.03571429	0.9642857
58	0.03571429	0.9642857
59	0.03571429	0.9642857
60	0.03571429	0.9642857
61	0.03571429	0.9642857
62	0.03571429	0.9642857
63	0.03571429	0.9642857
64	0.03571429	0.9642857
65	0.40000000	0.60000000
66	0.00000000	1.00000000
67	0.40000000	0.60000000
68	0.40000000	0.60000000
69	0.40000000	0.60000000
70	0.40000000	0.60000000
71	0.40000000	0.60000000
72	0.21428571	0.7857143
73	0.40000000	0.60000000
74	0.40000000	0.60000000
75	0.40000000	0.60000000
76	0.21428571	0.7857143
77	0.40000000	0.60000000
78	0.40000000	0.60000000
79	0.40000000	0.60000000
80	0.40000000	0.60000000
81	0.40000000	0.60000000
82	0.40000000	0.60000000
83	0.21428571	0.7857143
84	0.00000000	1.00000000
85	0.00000000	1.00000000
86	0.00000000	1.00000000
87	0.21428571	0.7857143
88	0.00000000	1.00000000
89	0.21428571	0.7857143
90	0.00000000	1.00000000
91	0.21428571	0.7857143
92	0.00000000	1.00000000
93	0.00000000	1.00000000
94	0.00000000	1.00000000
95	0.00000000	1.00000000
96	0.00000000	1.00000000
97	0.00000000	1.00000000
98	0.00000000	1.00000000
99	0.21428571	0.7857143
100	0.00000000	1.00000000

Didapatkan pohon keputusan sebagai berikut :



KESIMPULAN

Berdasarkan analisis yang telah dilakukan, pohon keputusan mampu untuk menentukan kualitas baik atau buruknya nasabah dengan beberapa variabel yang menurut peneliti mampu menjelaskan secara keseluruhan. Analisis dari tulisan ini pun masih jauh dari kata sempurna, maka dari itu, penulis mengharapkan adanya beberapa analisis atau bahkan penelitian baru dari data ini untuk dapat menentukan model yang tepat mengenai penetapan kualitas nasabah peminjam dana di suatu bank.

DAFTAR PUSTAKA

- Pramana, Setia.dkk.2017. Dasar-Dasar Statistika dengan Software R Konsep dan Aplikasi .Edisi Kedua.Jakarta:In Media
- Hoffman, Hans. “German Credit Data” 1994. Diakses Pada 6 Oktober 2018.
<https://www.openml.org/search?type=data&from=100>