

ANALISA DATA MINING DEFAULT KARTU KREDIT MENGGUNAKAN ALGORITMA NEURAL NETWORK

Siti Aisyah¹, Sulastris²

^{1,2}Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Stikubank
¹aissiyahs@gmail.com, ²sulastris@edu.unisbank.ac.id

ABSTRAK

Pemakaian teknologi informasi saat ini sudah sangat banyak dikalangan bisnis, sama halnya dengan pemakaian kartu kredit. Dengan kartu kredit bisa mendapatkan kemudahan dalam bertransaksi. Tapi kita harus bijak dalam penggunaan kartu kredit jika tidak maka akan menjadikan default kartu kredit.

Metode yang digunakan untuk menangani permasalahan yang ada bertujuan untuk memperlihatkan bagaimana sebuah model data mining dapat digunakan untuk membantu mengetahui pola nasabah dalam mengambil kredit di salah satu Negara yaitu Taiwan pada bulan April hingga September 2005 berdasarkan atribut-atribut dari data nasabah yang mempunyai kredit lancar ataupun macet. Data ini diambil dari UCI Machine Learning Repository. ANN dengan nilai keluaran kontinu biasanya menggunakan fungsi aktivasi sigmoid. Nilai sinyal keluaran y dihitung dengan fungsi kurva sigmoid dengan interval nilai keluaran mulai dari 0 sampai 1.

Hasil penelitian yang dilakukan dengan menggunakan Algoritman Neural Network dari 30.000 data dari data default Kartu Kredit Bank Taiwan dengan menggunakan RStudio untuk mendeteksi kredit macet pada Bank Taiwan dapat dijelaskan bahwa visualisasi model Neural Network tersebut memiliki 3 hidden layer. Sehingga dari proses traning diatas variable yang paling berpengaruh ada 3 yaitu X14(jumlah pernyataan tagihan pada bulan Juli 2005), X23(jumlah yang dibayarkan pada bulan April 2005) dan X20(jumlah yang dibayarkan pada bulan Juli 2005). Hasil akurasi menggunakan confusion matrix nilai RMSE prediksi 2 sebesar 0.1938333333. Dari prediksi 2 memiliki nilai RMSE yang sangat kecil, sehingga keakurasian prediksi sangat bagus.

Kata kunci: *Data Mining, Neural Network, Kartu Kredit.*

ABSTRACT

The use of information technology today is very much among businesses, as well as the use of credit cards. With a credit card you can get convenience in transactions. But we must be wise in using a credit card if not it will make a credit card default.

The method used to deal with existing problems aims to show how a data mining model can be used to help determine the pattern of customers in taking credit in one country, namely Taiwan in April to September 2005 based on the attributes of data customers who have current credit or This data is taken from the UCI Machine Learning Repository. ANN with a continuous output value usually uses the sigmoid activation function. The value of the output signal y is calculated by the sigmoid curve function with intervals of output values ranging from 0 to 1.

The results of research conducted using the Neural Network Algoritman of 30,000 data from the default data of the Taiwan Bank Credit Card by using RStudio to detect bad loans at the Bank of Taiwan can be explained that the visualization of the Neural Network model has 3 hidden layers. So from the transfer process above the most influential variables there are 3 namely X14 (number of billing statements in July 2005), X23 (amount paid in April 2005) and X20 (amount paid in July 2005). The accuracy of using the confusion matrix RMSE prediction value of 2 is 0.1938333333. Prediction 2 has a very small RMSE value, so the accuracy of the prediction is very good.

Keywords: *Data Mining, Neural Network, Credit Card.*

1. PENDAHULUAN

Penerapan sistem informasi dalam dunia bisnis sudah banyak digunakan untuk mempercepat proses bisnis. “Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data” [1].

Pemakaian teknologi informasi saat ini sudah sangat banyak dikalangan bisnis, sama halnya dengan pemakaian kartu kredit. Dengan kartu kredit bisa mendapatkan kemudahan dalam bertransaksi. Tapi kita harus bijak dalam penggunaan kartu kredit jika tidak maka akan menjadikan *default* kartu kredit.

Neural Network adalah jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan susunan syaraf manusia. Algoritma *Neural Network* yang paling populer adalah *Backpropagation*, Algoritma *Backpropagation* melakukan pembelajaran pada jaringan saraf *multi layer*

feed forward yang terdiri dari tiga lapisan/layer, yaitu: lapisan input, lapisan tersembunyi, dan lapisan keluaran [2]. Penelitian yang digunakan untuk mencari pola/trend peminjaman yang dilakukan oleh UKM. Dengan mengetahui trend peminjaman dari ukm tersebut maka BPR_CTR akan dapat memprediksi besarnya alokasi dana yang harus mereka siapkan setiap bulannya [3].

Telah dilakukan penelitian dengan tujuan membandingkan dari kedua metode tersebut, metode mana yang menghasilkan akurasi yang lebih baik [4]. Penelitian yang digunakan untuk mengetahui bagaimana menentukan nasabah yang akan mengajukan kredit akan masuk kedalam kelompok kredit macet atau lancar. Dengan menggunakan metode Neural Network algoritma yang digunakan Algoritma Backpropagation, dalam penelitian ini menggunakan tools data mining rapid miner [5].

2. METODE PENELITIAN

Metode yang digunakan untuk menangani permasalahan yang ada bertujuan untuk memperlihatkan bagaimana sebuah model data mining dapat digunakan untuk membantu mengetahui pola nasabah dalam mengambil kredit di salah satu Negara yaitu Taiwan pada bulan April hingga September 2005 berdasarkan atribut-atribut dari data nasabah yang mempunyai kredit lancar ataupun macet. Data ini diambil dari *UCI Machine Learning Repository*.

ANN dengan nilai keluaran kontinu biasanya menggunakan fungsi aktivasi sigmoid. Nilai sinyal keluaran y dihitung dengan fungsi kurva sigmoid dengan interval nilai keluaran mulai dari 0 sampai 1. Nilai y diformulasikan dengan:

$$y = \text{sign}(v) = \frac{1}{1 + e^{-av}} \quad (1)$$

Parameter a adalah parameter kemiringan (slope) pada kurva sigmoid yang dihasilkan semakin besar nilai a , semakin tegak kurva yang diberikan, dan semakin kecil nilainya, semakin landai kurva yang diberikan. Untungnya nilai a yang digunakan adalah 1 sehingga formula yang umum digunakan menjadi

$$y = \frac{1}{1 + e^{-av}} \quad (2)$$

Setelah mendapat informasi dari semua atribut yang dihitung, atribut yang paling mempengaruhi akan menjadi faktor paling penting dalam pemberian kartu kredit terhadap nasabah. Berikut adalah atribut data nya :

Tabel 2. Variabel Data

ATR	Keterangan	
X1	LIMIT_BAL	Jumlah kredit yang diberikan dalam dolar NT (termasuk kredit individu dan keluarga / pelengkap).
X2	SEX	Jenis kelamin (1 = laki-laki, 2 = perempuan)
X3	PENDIDIKAN	(1 = sekolah pascasarjana, 2 = universitas, 3 = sekolah tinggi, 4 = lainnya, 5 = tidak diketahui, 6 = tidak diketahui)
X4	MARRIAGE	Status perkawinan (1 = menikah, 2 =
X5	UMUR	Usia di tahun
X6	PAY_0	Status pembayaran pada bulan September 2005 (-1 = bayar sepatutnya, 1 = penundaan pembayaran untuk satu bulan, 2 = penundaan pembayaran selama dua bulan, ... 8 = penundaan pembayaran selama delapan bulan, 9 = penundaan pembayaran selama sembilan bulan dan atas)
X7	PAY_2	Status pelunasan pada bulan Agustus 2005 (skala sama seperti di atas)
X8	PAY_3	Status pelunasan pada bulan Juli 2005 (skala sama seperti di atas)
X9	PAY_4	Status pelunasan pada bulan Juni 2005 (skala sama seperti di atas)
X10	PAY_5	Status pembayaran kembali pada bulan Mei 2005 (skala sama seperti di atas)
X11	PAY_6	Status pelunasan pada bulan April 2005 (skala sama seperti di atas)
X12	BILL_AMT1	Jumlah pernyataan tagihan pada bulan September 2005 (dolar NT)

X13	BILL_AMT2	Jumlah pernyataan tagihan pada bulan Agustus 2005 (dolar NT)
X14	BILL_AMT3	Jumlah pernyataan tagihan pada bulan Juli 2005 (dolar NT)
X15	BILL_AMT4	Jumlah pernyataan tagihan pada bulan Juni 2005 (dolar NT)
X16	BILL_AMT5	Jumlah pernyataan tagihan pada bulan Mei 2005 (dolar NT)
X17	BILL_AMT6	Jumlah pernyataan tagihan pada bulan April 2005 (dolar NT)
X18	PAY_AMT1	Jumlah pembayaran sebelumnya pada bulan September 2005 (dolar NT)
X19	PAY_AMT2	Jumlah pembayaran sebelumnya pada bulan Agustus 2005 (dolar NT)
X20	PAY_AMT3	Jumlah pembayaran sebelumnya pada bulan Juli 2005 (dolar NT)
X21	PAY_AMT4	Jumlah pembayaran sebelumnya pada bulan Juni 2005 (dolar NT)
X22	PAY_AMT5	Jumlah pembayaran sebelumnya pada bulan Mei 2005 (dolar NT)
X23	PAY_AMT6	Jumlah pembayaran sebelumnya pada bulan April 2005 (dolar NT)
Y	Default.payment.	Pembayaran default (1 = ya, 0 = tidak)

Sebuah proses data mining ini melalui tahap dalam analisisnya menurut *knowledge discovery in databases*(KDD) sebagai berikut :

2.1. Persiapan Data

Tahapan awal yang dilakukan dalam penelitian sebelum diproses ke dalam data mining:

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	Y
1	20000	2	2	1	24	2	2	-1	-1	-2	-2	3913	3102	689	0	0	0	0	689	0	0	0	0	1
2	120000	2	2	2	26	-1	2	0	0	0	2	2682	1725	2682	3272	3455	3261	0	1000	1000	1000	0	2000	1
3	90000	2	2	2	34	0	0	0	0	0	0	29239	14027	13559	14331	14948	15549	1518	1500	1000	1000	1000	5000	0
4	50000	2	2	1	37	0	0	0	0	0	0	46990	48233	49291	28314	28959	29547	2000	2019	1200	1100	1069	1000	0
5	50000	1	2	1	57	-1	0	-1	0	0	0	8617	5670	35835	20940	19146	19131	2000	36681	10000	9000	689	679	0
6	50000	1	1	2	37	0	0	0	0	0	0	64400	57069	57608	19394	19619	20024	2500	1815	657	1000	1000	800	0
7	500000	1	1	2	29	0	0	0	0	0	0	367965	412023	445007	542653	483003	473944	55000	40000	38000	20239	13750	13770	0
8	100000	2	2	2	23	0	-1	-1	0	0	-1	11876	380	601	221	-159	567	380	601	0	581	1687	1542	0
9	140000	2	3	1	28	0	0	2	0	0	0	11285	14096	12108	12211	11793	3719	3329	0	432	1000	1000	1000	0
10	20000	1	3	2	35	-2	-2	-2	-2	-1	-1	0	0	0	0	13007	13912	0	0	0	13007	1122	0	0
11	200000	2	3	2	34	0	0	2	0	0	-1	11073	9787	5535	2513	1828	3731	2306	12	50	300	3738	66	0
12	260000	2	1	2	51	-1	-1	-1	-1	-1	2	12261	21670	9966	8517	22287	13668	21818	9966	8583	22301	0	3640	0
13	630000	2	2	2	41	-1	0	-1	-1	-1	-1	12137	6500	6500	6500	6500	2870	1000	6500	6500	6500	2870	0	0
14	70000	1	2	2	30	1	2	2	0	0	2	65802	67369	65701	66782	36137	36894	3200	0	3000	3000	1500	0	1
15	250000	1	1	2	29	0	0	0	0	0	0	70887	67060	63561	59696	56875	55512	3000	3000	3000	3000	3000	3000	0
16	50000	2	3	3	25	1	2	0	0	0	0	50614	29173	28116	28771	29531	30211	0	1500	1100	1200	1300	1100	0
17	20000	1	1	2	24	0	0	2	2	2	2	15376	18010	17428	18338	17905	19104	3200	0	1900	0	1650	0	1
18	320000	1	1	1	49	0	0	0	-1	-1	-1	253286	246536	194663	70074	5856	195599	10358	10000	75940	20000	195599	50000	0
19	360000	2	1	1	49	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0	0
20	180000	2	1	2	29	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0	0

Showing 1 to 21 of 30,000 entries

Gambar 1. Persiapan Data

2.2. Transformasi Data

Menormalisasikan data dan membersihkan data yang tidak digunakan. Berikut adalah gambar hasil transformasi :

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21
1	0.01010101	1	0.3333333	0.3333333	0.05172414	0.4	0.4	0.1	0.1	0.0	0.0	0.1499817	0.06916432	0.08672289	0.1601378	0.08064809	0.2609787	0.0000000000	4.090820e-04	0.000000e+00	0.000000e+00
2	0.11111111	1	0.3333333	0.6666667	0.08620690	0.1	0.4	0.2	0.2	0.2	0.4	0.1488924	0.06785751	0.08781713	0.1632199	0.08407395	0.2634847	0.0000000000	5.937329e-04	1.116022e-03	1.610306e-03
3	0.08080808	1	0.3333333	0.6666667	0.22413793	0.2	0.2	0.2	0.2	0.2	0.2	0.1723923	0.07953247	0.09378907	0.1736374	0.09547003	0.2729278	0.0017377328	8.905994e-04	1.116022e-03	1.610306e-03
4	0.04040404	1	0.3333333	0.3333333	0.27586207	0.2	0.2	0.2	0.2	0.2	0.2	0.1800999	0.11199497	0.11340745	0.1868092	0.10936287	0.2836851	0.0022895031	1.198747e-03	1.339226e-03	1.771337e-03
5	0.04040404	0	0.3333333	0.3333333	0.62068966	0.1	0.2	0.1	0.2	0.2	0.2	0.1541442	0.07160143	0.10601954	0.1798630	0.09963262	0.2756805	0.0022895031	2.177872e-02	1.116022e-02	1.449275e-02
6	0.04040404	0	0.6666667	0.6666667	0.27586207	0.2	0.2	0.2	0.2	0.2	0.2	0.2035057	0.12038060	0.11797384	0.1784066	0.10010164	0.2763668	0.0028618789	1.077625e-03	7.332262e-04	1.610306e-03
7	0.49494949	0	0.6666667	0.6666667	0.13793103	0.2	0.2	0.2	0.2	0.2	0.2	0.4721257	0.45714242	0.33067231	0.6713097	0.59557779	0.6251961	0.0629613349	2.374932e-02	4.240882e-02	3.259098e-02
8	0.09090909	1	0.3333333	0.6666667	0.03448276	0.2	0.1	0.1	0.2	0.2	0.1	0.1570281	0.06658106	0.08667458	0.1603459	0.08049043	0.2614145	0.0004350056	3.563335e-04	0.000000e+00	9.355878e-04
9	0.13131313	1	0.5000000	0.3333333	0.12068966	0.2	0.2	0.4	0.2	0.2	0.2	0.1565051	0.07959795	0.09289241	0.1716404	0.09234163	0.2638367	0.0038108779	0.000000e+00	4.821213e-04	1.610306e-03
10	0.01010101	0	0.5000000	0.6666667	0.24137931	0.0	0.0	0.0	0.0	0.1	0.1	0.1465192	0.06622043	0.08634460	0.1601378	0.09354540	0.2716698	0.0000000000	0.000000e+00	0.000000e+00	2.094525e-02
11	0.19191919	1	0.5000000	0.6666667	0.22413793	0.2	0.2	0.4	0.2	0.2	0.1	0.1563175	0.07550858	0.08938355	0.1625050	0.08246067	0.2638459	0.0026397971	7.124795e-06	5.580108e-05	4.830918e-04
12	0.25252525	1	0.6666667	0.6666667	0.51724138	0.1	0.1	0.1	0.1	0.1	0.4	0.1573687	0.08678590	0.09181636	0.1681607	0.10274714	0.2714823	0.0149761892	5.917141e-03	9.578813e-03	3.591143e-02
13	0.62626263	1	0.3333333	0.6666667	0.34482759	0.1	0.2	0.1	0.1	0.1	0.1	0.1572590	0.07238912	0.08991338	0.1662607	0.08709327	0.2631843	0.0011447515	3.89264e-03	7.254140e-03	1.046699e-02
14	0.06060606	0	0.3333333	0.6666667	0.15517241	0.3	0.4	0.4	0.2	0.2	0.4	0.2047463	0.13015560	0.12241724	0.2230455	0.11648033	0.2893311	0.0036632049	0.000000e+00	3.348065e-03	4.830918e-03
15	0.24242424	0	0.6666667	0.6666667	0.13793103	0.2	0.2	0.2	0.2	0.2	0.2	0.2092460	0.12986235	0.12124229	0.2163706	0.13704345	0.3036387	0.0034342546	1.781199e-03	3.348065e-03	4.830918e-03
16	0.04040404	1	0.5000000	1.0000000	0.03448276	0.3	0.4	0.2	0.2	0.2	0.2	0.1913067	0.09390647	0.10178148	0.1872397	0.10993004	0.2841953	0.0000000000	8.905994e-04	1.227624e-03	1.832367e-03
17	0.01010101	0	0.6666667	0.6666667	0.05172414	0.2	0.2	0.4	0.4	0.4	0.4	0.1601252	0.08331245	0.09591331	0.1774119	0.09840209	0.2756598	0.0036632049	0.000000e+00	1.674032e-03	0.000000e+00
18	0.31313131	0	0.6666667	0.3333333	0.48275862	0.2	0.2	0.2	0.1	0.1	0.1	0.3706480	0.30019038	0.19322284	0.2261465	0.08645470	0.4112930	0.00118573365	5.937329e-03	8.475068e-02	3.220612e-02

Gambar 2. Hasil Proses Transformasi

2.3. Pembagian Data

Proses data mining ini dilakukan untuk membagi data training dan data testing. Pembagian data ini adalah 60% untuk data train dan 40% data test. Sehingga menghasilkan 18.000 data train dan 12.000 data test dengan 24 variabel.

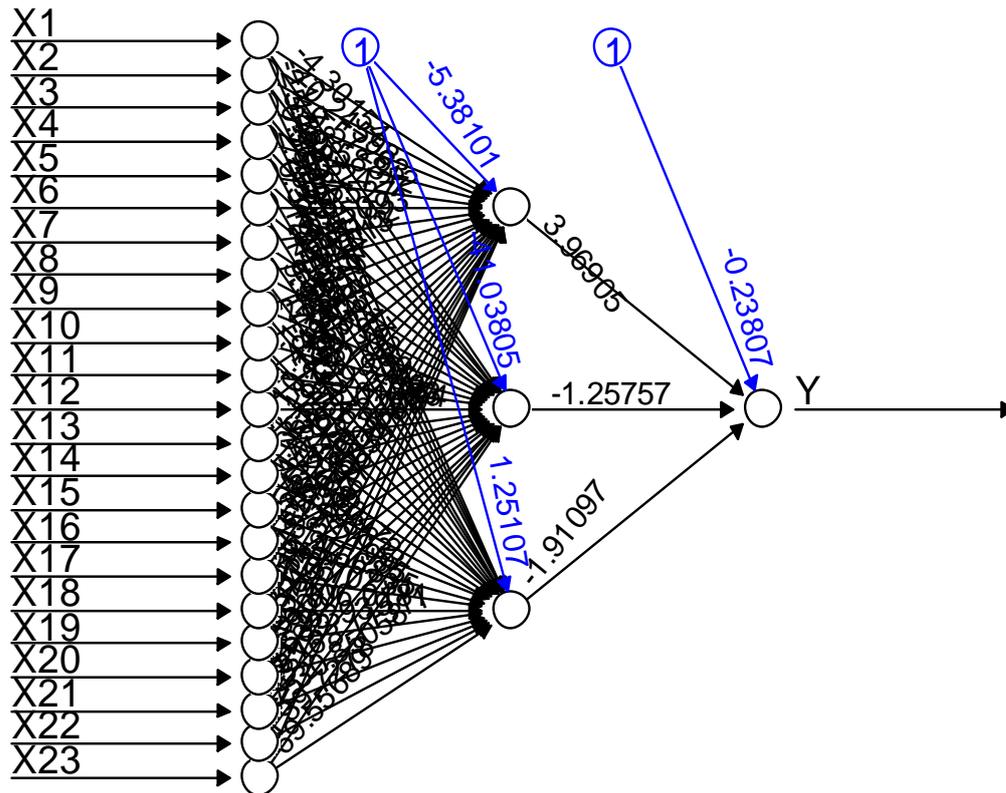
3. HASIL DAN PEMBAHASAN

3.1. Pemodelan Data

Proses data mining ini untuk memvisualisasikan *Neural Network* dengan bobot masing-masing variable menggunakan package *neuralnet* dan menampilkan plot NN.

```

NN = neuralnet(Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 + X12 + X13 + X14
+ X15 + X16 + X17 + X18 + X19 + X20 + X21 + X22 + X23, trainNN, hidden = 3 , linear.output =
FALSE )
plot(NN)
    
```



Gambar 3. Hasil visualisasi Neural Network

Pemodelan ini menghasilkan prediksi dan akurasi dengan menggunakan *confusion matrix*:

```
pred2    0    1
         0 8543 1469
         1  857 1131
> 1-sum(diag(tab2))/sum(tab2)
[1] 0.1938333333
```

Gambar 4. Prediksi dengan Confusion Matrix

5. KESIMPULAN

Hasil penelitian yang dilakukan dengan menggunakan *Algoritman Neural Network* dari 30.000 data dari data *default* Kartu Kredit Bank Taiwan dengan menggunakan RStudio untuk mendeteksi kredit macet pada Bank Taiwan dapat dijelaskan bahwa visualisasi model *Neural Network* tersebut memiliki 3 hidden layer. Sehingga dari proses training diatas variable yang paling berpengaruh ada 3 yaitu X14(jumlah pernyataan tagihan pada bulan Juli 2005), X23(jumlah yang dibayarkan pada bulan April 2005) dan X20(jumlah yang dibayarkan pada bulan Juli 2005). Hasil akurasi menggunakan *confusion matrix* nilai RMSE prediksi 2 sebesar 0.1938333333. Dari prediksi 2 memiliki nilai RMSE yang sangat kecil, sehingga keakurasian prediksi sangat bagus.

DAFTAR PUSTAKA

- [1] Efraim, T., Jay, E. A., Liang, T. P., & McCarthy, R. V. (2005). Decision support systems and intelligent systems. *Andi, Yogyakarta*.
- [2] Han, J., Kamber, M., & Pei, J. (2006). Data mining: concepts and techniques. 2001. *San Francisco: Morgan Kauffman*.
- [3] Novita, A., Atmodjo, D., & Rahayu, P. (2013). Study Awal Peminjaman Kredit Nasabah UKM Pada BPR Dengan Menggunakan Neural Network.

- [4] Diaprina, S. R., & Suhartono, S. (2014). Analisis Klasifikasi Kredit Menggunakan Regresi Logistik Biner Dan Radial Basis Function Network di Bank 'X' Cabang Kediri. *Jurnal Sains dan Seni ITS*, 3(2), D218-D223.
- [5] Sucipto, A. (2015). Credit Prediction with Neural Network Algorithm.