

DATA MINING UNTUK ANALISA PENGAJUAN KREDIT DENGAN MENGGUNAKAN METODE LOGISTIK REGRESI

Data Mining To Analisizes Loan Proposal Using Regression Logistic Method

Jusia Amanda Ginting, jginting@bundamulia.ac.id¹⁾

¹⁾ Teknik Informatika, Fakultas Teknologi dan Desain, Universitas Bunda Mulia

ABSTRACT

The application of data mining as a medium for analyzing credit applications by customers is very important, due to the ability of data mining in data processing. The data used in this study is customer data from a banking industry taken from the Machine Learning Repository of ten thousand data and the results of data processing can be used as a reference by the bank to approve whether the customer's submission can be approved or not. The process of analyzing data on rapidminer through several processes, namely data retriive data limiter data, set roles, split data

Keywords: *Data Mining, Rapidminer, Classification, Logistic Regression, Banking Credit*

ABSTRAK

Penerapan data mining sebagai media untuk menganalisa pengajuan kredit oleh nasabah sangat penting, dikarenakan kemampuan data mining dalam pengolahan data. Data yang digunakan dalam penelitian ini merupakan data nasabah dari sebuah industri perbankan yang diambil dari Machine Learning Repository sebanyak sepuluh ribu data dan hasil dari pengolahan data dapat digunakan sebagai salah satu acuan oleh pihak bank untuk menyetujui apakah pengajuan nasabah tersebut dapat disetujui atau tidaknya. Proses analisa data pada rapidminer melalau beberapa proses yaitu data retriive data limiter dat, set role, spli data

Kata Kunci: *Data Mining, Rapidminer, klasifikasi, Logistik Regresi, Kredit Perbankan, Analisa Data*

PENDAHULUAN

Perkembangan teknologi saat ini, persaingan perusahaan-perusahaan dalam berbagai bidang bersaing ketat untuk menjadi yang terbaik oleh karena itu perusahaan harus mampu meningkatkan produktifitas bukan hanya untuk mempertahankan perusahaan, tetapi agar tujuan utama dari perusahaan dapat tercapai. Penerapan *Business Intelligence* pada industri per-bankan merupakan kunci sukses dalam mengidentifikasi dan mengefektifkan kegiatan bisnis utama dengan kemampuan dalam mendapatkan, mengelola dan menganalisa data nasabah, produk, layanan, kegitan operasi dalam

jumlah yang sangat besar. Salah satu contoh penerapan *Bussiness Intelegent* pada dunia perbankan adalah *customer credit analysis*. [1]. Penerapan *Customer Credit Analysis* adalah penerapan model penilaian kredit kepada nasabah [2]. Penilaian kredit nasabah merupakan kegiatan yang sangat penting untuk mengevaluasi aplikasi pinjaman atau aplikasi pengajuan kredit oleh nasabah, sistem penilaian kredit digunakan untuk memodelkan potensi resiko dan aplikasi pinjaman. Keuntungan dari sistem ini adalah sistem dapat menangani aplikasi pinjaman dan pengajuan kredit dalam jumlah yang besar dengan cepat tanpa membutuhkan sumber daya yang besar sehingga dapat

menurunkan biaya operasional dan efektif dalam penalaran dalam pengambilan keputusan.

Data mining merupakan bagian dari konsep teknologi informasi yang berkaitan dengan data dan informasi dan *data mining* juga dapat digunakan sebagai metode atau teknik pemodelan untuk menganalisis data dengan variasi data yang banyak untuk mendapatkan pola dan hubungan antar variasi tersebut.

Riset ini memiliki kontribusi dalam hal memahami pentingnya data mining dalam menganalisa penerimaan pengajuan kredit dengan menggunakan metode *Logistic Regresi*.

IDENTIFIKASI MASALAH

Dalam penelitian ini *data mining* digunakan sebagai *tools* untuk menganalisa data konsumen disebuah bank dengan menggunakan *data set Machine Learning Repository* dengan jumlah data konsumen sebanyak sepuluh ribu data yang diolah dengan menggunakan metode Logistik Regresi dan menggunakan aplikasi *rapid miner* sebagai alat bantu dalam pemrosesan data. Adapun batasan masalah dalam penelitian ini yaitu hanya menggunakan lima *variable* utama yaitu pekerjaan, status pernikahan nasabah, pendidikan, status kepemilikan rumah atau tempat tinggal dan status pinjaman atau kredit. Kelima *variable* tersebut menjadi acuan utama sebagai dasar pengambilan keputusan untuk memberikan kredit atau pengajuan kredit ke nasabah di bank tersebut sehingga dapat dibuat rumusan masalah yaitu: penerapan *Data Mining* dalam menganalisa pengajuan kredit dengan menggunakan metode logistik regresi.

TINJAUAN PUSTAKA

Penelitian yang berjudul “Perancangan Data Mining untuk Analisis Kriteria Nasabah Kredit yang Potensial dan

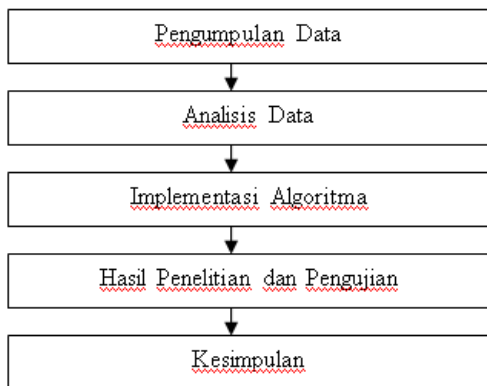
Manfaatnya untuk *Customer Relationship Management* Perbankan” merupakan penelitian mengenai penggunaan Data Mining untuk menganalisa kriteria nasabah kredit dengan tujuan meningkatkan loyalitas pelanggan ke bank meningkatkan profitabilitas dan mengurangi biaya akuisisi pelanggan.[4]

Penelitian yang berjudul “Penerapan *Business Intelligence* pada Industri Perbankan, Retail dan Pendidikan” merupakan penelitian yang membandingkan penerapan *Bussiness Intelligence* pada berbagai bidang, seperti contoh penggunaan clustering pada nasabah untuk menganalisa data pelanggan untuk menghasilkan keuntungan yang maksimal dan dengan menggunakan konsep pareto dengan mendesain produk dan layanan kepada 20% nasabah dapat memberikan hasil sebesar 80% terhadap keuntungan, dengan menganalisa 20% nasabah, manajemen percaya merupakan suatu langkah yang efektif dalam meningkatkan keuntungan dan menurunkan biaya operasional.[5]

Penelitian yang berjudul “*Using Data Mining For Bank Direct Marketing: An Application Of The Crisp-Dm Methodology*” merupakan penelitian yang mengimplementasikan data mining menggunakan metode *Cross Industry Standart Process* untuk mendapatkan sebuah model yang dapat digunakan untuk meningkatkan efisiensi persusahaan dengan mengidentifikasi karakteristik utama yang mempengaruhi kemajuan perusahaan dan manajemen sumber daya yang ada. Penelitian yang berjudul “Regresi Logistik” merupakan penelitian yang menggunakan metode regresi logistik untuk mengetahui pengaruh kualitas pelayanan public terhadap kepuasan pengguna dan menjelaskan penggunaan metode Regresi Logistik untuk mengetahui pengaruh satu *variable* independen atau lebih terhadap satu *variable* dependen.[6]

METODE PENELITIAN

Penelitian ini dilakukan dengan cara menganalisa data konsumen dalam industry perbankan yang bertujuan untuk mengetahui apakah konsumen tersebut bisa mendapatkan pinjaman atau pengajuan kredit di sebuah bank. Adapun tahapan metode penelitian seperti gambar 1 yang terdiri dari 5 tahap yaitu (1)Pengumpulan data, (2) Analisis Data, (3) Implementasi Algoritma, (4) Hasil Penelitian dan Pengujian, (5)Kesimpulan.



Gambar 1 Metode Penelitian

Tahap pengumpulan data merupakan tahap mengumpulkan data, data nasabah pada industry perbankan diambil pada *Machine Learning Repository* dengan jumlah data sebanyak sepuluh ribu data dan terdapat 5 variabel utama yaitu pekerjaan, status pernikahan, pendidikan, kepemilikan rumah, dan status pinjaman. Tahap Analisi data menggunakan rapidminer sebagai *tools* dalam pengolahan data mining. Tahap implementasi algoritma merupakan tahap dimana algoritma diimplementasikan pada data yang sudah diperoleh sebelumnya. Jika diketahui ada p *variable* independen maka bisa ditunjukkan dengan vector pada rumus 1.

$$X' = (X_1, X_2, \dots, X_p) \quad \dots\dots\dots(1)$$

Hal ini diasumsikan bahwa beberapa *variable* bersekala interval sehingga

probabilitasnya dapat dituliskan dengan rumus 2

$$P(Y=1|x) = \pi(x) \quad \dots\dots\dots(2)$$

Persamaan *logic* dengan model regresi logistik seperti pada rumus 3 merupakan salah satu jenis regresi yang menghubungkan antara satu dan beberapa variabel independen dengan variabel dependen yang berupa kategori 0 dan 1

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad \dots\dots(3)$$

Persamaan untuk menentukan nilai peluang dari persamaan logistik:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad \dots\dots\dots(4)$$

Keterangan:

- $\pi(x)$ adalah peluang
- $g(x)$ adalah peluang estimasi logit
- $\beta_0, \beta_1, \beta_2$ dan β_3 adalah nilai koefisien untuk *variable-variabel* konstan

Statisk W digunakan untuk uji signifikansi parameter Regresi Logistik:

$$W_i = \frac{\beta_i}{SE_i} \quad \dots\dots\dots(5)$$

Wilayah Kristis:

$$W_i > \chi_{k,\alpha}^2 \quad \dots\dots\dots(6)$$

Keterangan rumus 6:

- β_i adalah nilai koefisien regresi logistic untuk *variable* ke- i
- SE_i adalah nilai *standar error* untuk *variable* ke- i
- k adalah banyaknya *variabel* bebas yang dapat digunakan
- α adalah taraf nyata.

Odds ratio digunakan untuk mengukur hubungan antara variabel A dan variabel B dalam sebuah populasi. Secara khusus *Odds ratio* digunakan untuk menentukan efek dari variabel A terhadap variabel B seperti pada rumus 7.

Rumus untuk menentukan *odds ratio*:

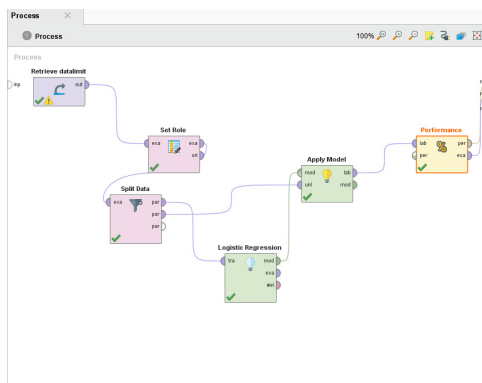
$$\frac{p_1/(1-p_1)}{p_2/(1-p_2)} \dots\dots\dots(7)$$

Keterangan :

- p1 adalah peluang kejadian pada kelompok pertama
- p2 adalah peluang kejadian pada kelompok ke-dua.

HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan data nasabah pada sebuah bank yang diambil dari *Machine Learning Repository* dan dibagi berdasarkan *variable*-variabelnya. *Variable* yang dimaksud terdiri dari lima variabel yaitu pekerjaan, status pernikahan, pendidikan, status kepemilikan rumah dan pinjaman. Dari kelima *variable* tersebut menjadi penentu apakah seorang nasabah dapat mengajukan kredit dan *variable* tersebut dapat dijadikan sebagai acuan oleh pihak perbankan untuk memutuskan apakah permohonan kredit dari seorang nasabah dapat disetujui atau tidak.



Gambar 2. Alur Proses Rapidminer

Pada Gambar 2 bagian *data retrieve data limiter data* yang digunakan berjumlah sepuluh ribu data dikarenakan keterbatasan pada *rapidminer* dalam mengolah data yang besar, proses selanjutnya adalah proses *set role*, pada proses ini data yang telah diimputkan akan di set sesuai dengan fungsi dan tujuan dari data tersebut seperti contoh data dengan *variabel* pekerjaan berpengaruh terhadap pinjaman kredit ke nasabah atau tidak. Proses selanjutnya adalah *split data*, pada tahapan ini data akan dipecah menjadi dua bagian yaitu 70% data akan digunakan untuk melakukan *training data* yang dilakukan menggunakan *algoritma Logistic Regression* dan 30% data digunakan sebagai data yang digunakan untuk mengambil keputusan.

ExampleSet(10000 examples, 0 special attributes, 6 regular attributes)

Row No.	job	marital	education	housing	loan	label
1	management	married	tertiary	yes	no	no
2	technician	single	secondary	yes	no	no
3	entrepreneur	married	secondary	yes	yes	no
4	blue-collar	married	unknown	yes	no	no
5	unknown	single	unknown	no	no	no
6	management	married	tertiary	yes	no	no
7	management	single	tertiary	yes	yes	no
8	entrepreneur	divorced	tertiary	yes	no	no
9	retired	married	primary	yes	no	no
10	technician	single	secondary	yes	no	no
11	admin.	divorced	secondary	yes	no	no
12	admin.	single	secondary	yes	no	no
13	technician	married	secondary	yes	no	no
14	technician	married	unknown	yes	no	no
15	services	married	secondary	yes	no	no
16	retired	married	primary	yes	no	no
17	admin.	single	unknown	yes	no	no
18	blue-collar	married	primary	yes	no	no
19	retired	married	primary	yes	no	no

Gambar 3. Data Nasabah

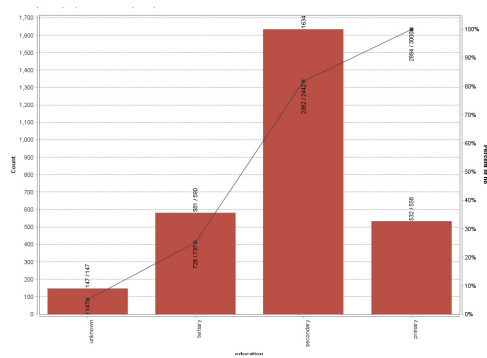
Pada Gambar 3 menunjukkan data nasabah yang terdapat pada sebuah bank yang berjumlah sepuluh ribu data dan terdiri dari lima variabel utama yaitu pekerjaan, status pernikahan, pendidikan, kepemilikan rumah dan pinjaman yang saling keterkaitan sebagai penentu apakah

permintaan untuk pengajuan kredit dapat disetujui atau tidak oleh pihak bank.

precision: 4.72% (positive class: yes)			
	true no	true yes	class precision
pred no	2794	100	96.54%
pred yes	101	5	4.72%
class recall	96.51%	4.76%	

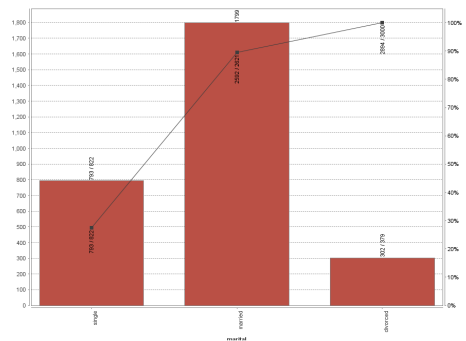
Gambar 4. Tabel Precision

Gambar 4 menjelaskan hasil dari analisa pengajuan kredit oleh nasabah bank. *Precision no* bernilai 96.54% yang berarti penolakan terhadap pengajuan kredit ditolak oleh bank dan *precision yes* bernilai 4.72 yang berarti pengajuan kredit oleh nasabah telah disetujui oleh pihak bank.



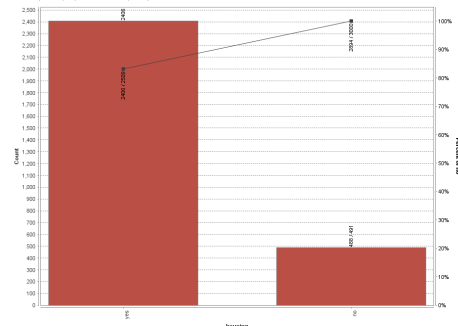
Gambar 5. Dampak Pendidikan terhadap pengajuan Kredit

Gambar 5 menunjukkan hubungan pendidikan terhadap pengajuan kredit. Pendidikan nasabah yang tinggi dapat menjadi pengaruh yang besar terhadap bank untuk mempertimbangkan pengajuan kredit dari seorang nasabah



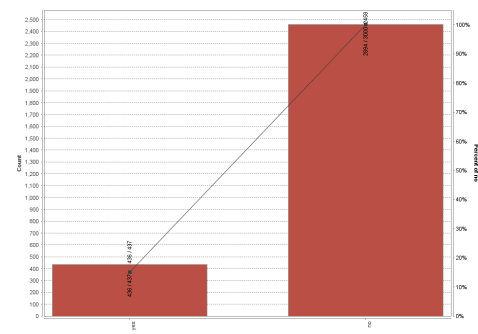
Gambar 6 Dampak Status Pernikahan Terhadap Pengajuan Kredit

Gambar 6 menunjukkan pengaruh status pernikahan nasabah terhadap pengajuan kredit. Status pernikahan “Menikah” menjadi pertimbangan yang cukup kuat oleh pihak bank untuk menyetujui pengajuan kredit oleh nasabah abk tersebut.



Gambar 7 .Dampak Kepemilikan Rumah terhadap Pengajuan Kredit

Gambar 7 menunjukkan pengaruh kepemilikan rumah terhadap pengajuan kredit. Pengajuan kredit oleh nasabah yang memiliki rumah cenderung disetujui oleh pihak bank dibandingkan dengan nasabah yang tidak memiliki rumah



Gambar 8 Pengaruh Hutang Terhadap pengajuan Kredit

Gambar 8 menunjukkan pengaruh hutang nasabah terhadap pengajuan kredit. Nasabah yang tidak memiliki hutang permohonan pengajuan kredit disetujui oleh pihak bank mencapai 100% dibandingkan dengan nasabah yang memiliki hutang persentasi bank untuk menyetujui pengajuan bank mencapai 18%.

KESIMPULAN

Penerapan *data mining* sebagai media untuk menganalisa pengajuan kredit oleh nasabah sangat penting, dikarenakan kemampuan data mining dalam pengolahan data. Data yang digunakan dalam penelitian ini merupakan data nasabah dari sebuah industri perbankan yang diambil dari *Machine Learning Repository* dan hasil dari pengolahan data dapat digunakan sebagai salah satu acuan bank untuk menyetujui apakah pengajuan nasabah tersebut dapat disetujui atau tidaknya. Dari penelitian yang sudah dilakukan dengan menggunakan data mining dan metode logistik regresi maka diperoleh kesimpulan nasabah yang memiliki pendidikan yang tinggi(100%), memiliki sebuah rumah(100%) dan sudah menikah (100%) serta tidak memiliki hutang (100%) berpotensi besar pengajuan kredit akan disetujui oleh pihak bank.

DAFTAR PUSTAKA

- [1] Hair, J. F. (2007). *Knowledge Creation in Marketing :The Role of Predictive Analysis*. *European Business Review* , 303-315.
- [2] Ince, H., & Aktan, B. (2009). *A Comparison of Data Mining Techniques for Credit Scoring in Banking: A Managerial Perspective*. *Journal of Business Economics and Management* , 233-240.
- [3] Mawoli, M. A., & Abdulsalam, D. (2012). *Effectiveness Market Segmentation and Viability of Islamic Banking in Nigeria*. *Australian Journal of Business and Management Research* , 1-9.
- [4] Kurniawan, Sukma (2015). *Perancangan Data Mining Untuk Analisis kriteria Nasabah Kredit Yang Potensial Dan Manfaatnya Untuk Customer Relation Management Perbankan*. Universitas Pendidikan Ganesha. Bali.
- [5] Sundjaja (2013). *Penerapan Business Intelegent Pada Industry Perbankan, Retail, Dan Pendidikan*. Binus University. Jakarta.
- [6] Rokhman,Ali. (2010) *Regresi Logistik*. Univeritas Jedral Sudirman.Purwokerto
- [7] Miranda, Eka (2008) *Pengembangan Business Intelligence Bagi Perkembangan Bisnis Perusahaan*.Universitas Bina Nusantara. Jakarta Barat
- [8] Dass, R.. (2008). *Data Mining in Banking and Finance: A Note for Bankers. Working Paper, Indian Institute of Management Ahmedabad* .
- [9] Kapiudin, M. 2008. *Data Mining untuk Klasifikasi Pelanggan dengan Ant Colony Optimization*. Jurusan Teknik Informatika, Fakultas Teknologi Industri – Universitas Kristen Petra.
- [10] Baye, M. R. 2010. *Managerial Economics and Business Strategy (7th Ed.)*. USA: McGraw-Hill.