

PREDIKSI KELAYAKAN PEMBERIAN FASILITAS KARTU KREDIT KEPADA NASABAH DENGAN METODE KLASIFIKASI DATA MINING (Studi Kasus : Bank XYZ)

Ferdinand Stevanus Alvian¹, Selawati Carbini², Komarudin³
STMIK LIKMI BANDUNG^{1,2}, STMIK Mardira Indonesia³
fsalvian@gmail.com¹, lstjarbini@gmail.com², komarudin@stmik-mi.ac.id³

Abstract

As one of the largest banks in Indonesia and in the midst of competition between existing banking institutions, Bank XYZ participates in providing credit to customers in the form of credit cards. The provision of credit card facilities for customers at Bank XYZ must go through a customer screening process and check data in depth so that banks can avoid non-current customer credit. One way that banks can use is by using the data mining method. In this study, two models were compared in the classification method, namely Discriminant Analysis and Naïve Bayes.

Keywords: Data Mining, CRISP-DM, Discriminant Analysis.

Abstrak

Sebagai salah satu bank terbesar di Indonesia dan ditengah persaingan antar instansi perbankan yang ada, Bank XYZ ikut serta memberikan kredit kepada nasabah dengan bentuk kartu kredit. Pemberian fasilitas kartu kredit bagi nasabah pada bank XYZ harus melalui proses screening nasabah dan pengecekan data secara mendalam agar pihak perbankan dapat menghindari kredit nasabah tidak lancar. Salah satu cara yang dapat digunakan oleh pihak perbankan adalah dengan menggunakan metode data mining. Dalam penelitian ini dibandingkan dua model dalam metode klasifikasi yaitu *Discriminant Analysis* dan *Naïve Bayes*.

Kata Kunci: Data Mining, CRISP-DM, Discriminant Analysis.

PENDAHULUAN

Dewasa ini, pemberian kartu kredit semakin meningkat dan diikuti pula dengan meningkatnya pembayaran tagihan bermasalah atau kredit yang macet. Dampak yang timbul adalah tidak terbayarkannya uang dari kredit yang diberikan baik sebagian ataupun seluruhnya yang memungkinkan juga berdampak pada keuangan bank yang memberikan fasilitas kartu kredit. Profil nasabah sangat menentukan dalam keputusan pemberian kredit.

Bank XYZ merupakan salah satu perusahaan penyedia jasa keuangan milik swasta terbesar di Indonesia. Bank XYZ sendiri juga menyandang gelar sebagai the world best bank's menurut forbes dan mendapatkan peringkat 10 besar untuk di negara Indonesia.

Rank	Name	Headquarters City	Headquarters Country	Employees
1	Bank Central Asia	Jakarta	Indonesia	26,479
2	PT Bank Tabungan Pensiunan Nasional (BTPN)	Jakarta	Indonesia	19,770
3	HSBC	London	United Kingdom	229,500
4	Bank Negara Indonesia	Jakarta	Indonesia	21,209
5	Bank Mandiri	Jakarta	Indonesia	28,307
6	Bank Rakyat Indonesia (BRI)	Jakarta	Indonesia	92,888
7	Perco Bank	Jakarta	Indonesia	8,778
8	Bank OCBC NISP	Jakarta	Indonesia	6,477
9	CitiBank	New York	United States	164,337
10	DBS	Singapore	Singapore	34,776

Gambar 1 Data Peringkat Bank Terbaik di Indonesia

(Sumber : www.forbes.com)

Proses pemberian fasilitas kartu kredit pada Bank XYZ harus melalui pengajuan dan analisis terhadap profil nasabah yang mengajukan kartu kredit. Proses ini merupakan proses yang penting karena dapat menentukan kelayakan nasabah untuk menerima fasilitas kartu kredit dari Bank XYZ itu sendiri. Analisis profil nasabah dapat ditentukan dari banyak faktor seperti jumlah tabungan yang dimiliki, mutasi atau lalu lintas dana dari rekening nasabah, dan juga penghasilan dari nasabah yang mengajukan kredit.

Data yang mendukung dalam proses analisis profil nasabah dapat diraih oleh pihak bank

menggunakan metode data mining. Data mining dapat membantu pihak bank untuk membantu proses pengambilan keputusan dalam pemberian fasilitas kartu kredit. Dengan metode dan penilaian dari nasabah yang tepat maka dapat meminimalisir kemungkinan adanya pembayaran tagihan kartu kredit nasabah yang bermasalah.

Sementara, Bank XYZ memiliki *dataset* penerimaan pengajuan kartu kredit tahun-tahun sebelumnya yang dapat dimanfaatkan untuk mengkaji kelayakan pemberian fasilitas kartu kredit nasabah. Berangkat dari kondisi tersebut, maka dilakukanlah penelitian ini, yaitu mengkaji *dataset* penerimaan kartu kredit nasabah tahun-tahun sebelumnya dengan proses *data mining* menggunakan metode klasifikasi. Dalam penelitian ini dilakukan perbandingan metode *Discriminant Analysis*, dan *Naïve Bayes*.

RUMUSAN MASALAH

Permasalahan yang menjadi fokus dalam penelitian ini yaitu:

1. Bagaimana memprediksi kelayakan pemberian kartu kredit kepada nasabah dengan metode *Naïve Bayes* ?
2. Bagaimana mengimplementasikan model *Naïve Bayes* untuk menganalisa profil nasabah yang memiliki peluang paling besar untuk mendapatkan fasilitas kartu kredit ?
3. Faktor apa yang paling mempengaruhi profil nasabah untuk mendapatkan fasilitas kredit?

Penelitian ini menggunakan bantuan program aplikasi Rapid Miner

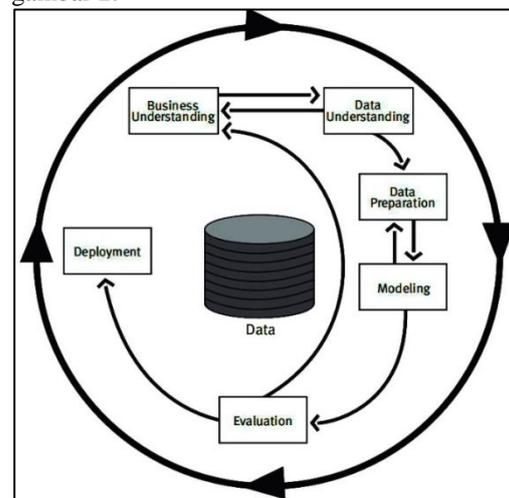
KAJIAN PUSTAKA

Data mining didefinisikan sebagai proses menemukan pola dan pengetahuan menarik dari sejumlah besar data. Sumber data dapat mencakup *database*, gudang data, Web, repositori informasi lain, atau data yang dialirkan ke sistem secara dinamis. (Han, 2012:8)

Sementara pengertian lain menyebutkan bahwa *data mining* adalah proses transformasi perampangan kumpulan informasi menjadi pengetahuan yang bermakna. Hal ini merupakan proses untuk membantu dalam mengidentifikasi peluang baru dengan menemukan kebenaran mendasar pada data yang terlihat acak. Pola yang terungkap dapat menjelaskan masalah aplikasi dan membantu pembuatan keputusan yang lebih bermanfaat dan proaktif. (Sumathi, 2006:9)

Langkah-langkah Data Mining

Dalam penelitian ini penulis menggunakan siklus CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang sudah dikembangkan sejak tahun 1996 oleh tim dari NCR, SPSS, dan Daimler Chrysler. Siklus ini dipilih karena fleksibilitasnya terhadap unit bidang bisnis apapun, baik untuk pemecahan permasalahan maupun untuk unit penelitian. Terdapat enam (6) fase dalam CRISP-DM yang saling berkaitan antara fase satu dengan fase lainnya, seperti gambar 2.



Gambar 2 Fase dari CRISP-DM
(Sumber: Chapman, 2000:10)

- a. *Business Understanding*, merupakan fase pertama yang fokus pada pemahaman tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menerjemakan pengetahuan ini kedalam pendefinisian masalah dalam data mining. Selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut.
- b. *Data Understanding*, fase ini dimulai dengan pengumpulan data yang kemudian akan dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi masalah kualitas data, atau untuk mendeteksi adanya bagian yang menarik dari data yang dapat digunakan untuk hipotesa untuk informasi yang tersembunyi.
- c. *Data Preparation*, fase ini meliputi semua kegiatan untuk membangun

dataset akhir (data yang akan diproses pada tahap modeling) dari data mentah. Fase ini dapat diulang hingga beberapa kali. Pada fase ini ini juga mencakup pemilihan tabel, record, dan atribut-atribut data, termasuk proses pembersihan dan transformasi data untuk kemudian dijadikan masukan dalam tahap pemodelan (modeling).

- d. *Modeling*. Dalam fase ini akan dilakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Secara khusus, ada beberapa teknik berbeda yang dapat diterapkan untuk masalah data mining yang sama. Pada fase ini masih memungkinkan kembali ke tahap sebelumnya.
- e. *Evaluation*. Pada fase ini, model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisa data. Pada fase ini akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal (Business Understanding). Kunci dari fase ini adalah menentukan apakah ada masalah bisnis yang belum dipertimbangkan. Pada akhir dari tahap ini harus ditentukan penggunaan hasil proses data mining.
- f. *Deployment*. Pada fase ini, pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna. Fase deployment dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses data mining yang berulang dalam sebuah unit bisnis.

Discriminant Analysis

Singgih Santoso mendefinisikan analisis diskriminan sebagai teknik multivariat yang termasuk dependece method, yakni adanya variabel dependen dan independen. Dengan

demikian ada variabel yang hasilnya tergantung dari data variabel independen. Ciri khusus data variabel dependen harus berupa data kategori, sedangkan data variabel independen berupa data non-kategori (Santoso, 2017). Adapun tujuan dari analisis diskriminan yaitu:

- a. Ingin mengetahui apakah ada perbedaan yang jelas antar grup pada variabel dependen? Dengan kata lain, apakah ada perbedaan antara anggota grup satu dengan anggota grup lainnya?
- b. Jika ada perbedaan, variabel independen manakah pada fungsi diskriminan yang membuat perbedaan tersebut?
- c. Membuat fungsi atau model diskriminan
- d. Melakukan klasifikasi terhadap objek (dalam terminologi SPSS disebut baris), apakah suatu objek termasuk ke dalam grup satu, grup dua, atau lainnya.

Untuk memodelkan analisis diskriminan digunakan persamaan berikut:

$$Y_1 = X_1 + X_2 + X_3 + \dots + X_n \quad (1)$$

Keterangan:

X1 : Variabel independen (X1, X2, dan seterusnya) yang merupakan data metrik, yakni data berjenis interval atau rasio. Misal: Usia seseorang, tinggi badan, dan sebagainya.

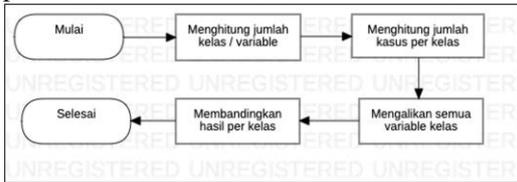
Y1 : Variabel dependen yang merupakan data kategorikal atau nominal. Misal: kategori seseorang yang banyak minum (kode 1), sedangkan kategori seseorang yang sedikit minum (kode 0), dan sebagainya.

Naïve Bayes

Naïve Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas (Patil and Sherekar : 2013).

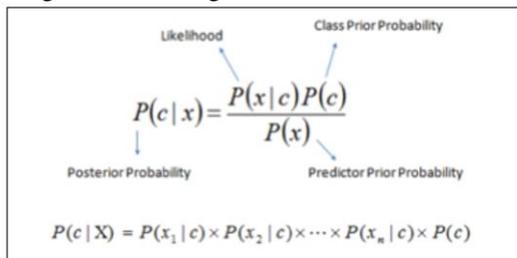
Konsep penggunaan metode naïve bayes adalah dengan mencari nilai probabilitas label atau kelas dan nilai probabilitas setiap atribut sebagai data training, yang kemudian nilai probabilitas tersebut digunakan untuk memprediksi data baru yang belum memiliki label dengan cara mengalikan nilai probabilitas atribut pada data

training sesuai atribut pada data yang di uji terhadap masing-masing nilai probabilitas label atau kelas, hasil perkalian pada masing-masing label kemudian dibandingkan dan nilai probabilitas terbesar digunakan sebagai hasil prediksi.



Gambar 3 Langkah-langkah Algoritma Naive Bayes
 (Sumber: Saleh, 2016)

Untuk menghitung probabilitas setiap kasus. Perhitungan dilakukan dengan menghitung jumlah kasus yang terjadi di masing-masing variabel, sesuai yang bersangkutan dengan data tambahan, dengan masing-masing kelas kejadian dengan rumus sebagai berikut :



Gambar 1 Rumus probabilitas
 (Sumber: Vembandasamy, 2015)

Keterangan :

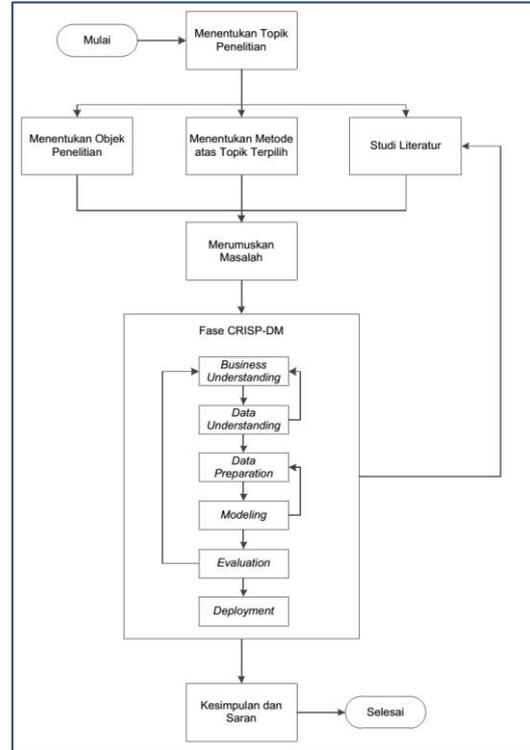
- P (c | x) adalah probabilitas posterior kelas (target) yang diberikan prediktor (atribut).
- P (c) adalah probabilitas kelas sebelumnya.
- P (x | c) adalah kemungkinan yang merupakan probabilitas kelas yang diberikan prediktor.
- P (x) adalah probabilitas prediktor sebelumnya

3.1. Akurasi Klasifikasi
 Selanjutnya mengacu pada (Gorunescu, 2011) akurasi klasifikasi yang dihasilkan dapat dibandingkan sesuai dengan tabel 1

Tabel 1 Akurasi Klasifikasi

Presentase Akurasi	Klasifikasi
50 – 59	Failure
60 – 69	Poor
70 – 79	Fair
80 – 89	Good
90 – 100	Excellent

METODE PENELITIAN



Gambar 5 Metode Penelitian

HASIL DAN PEMBAHASAN

Model Discriminant Analysis

Model Discriminant Analysis (Analisis Diskriminan) menghasilkan fungsi yang mampu memetakan objek yang diprediksi sehingga dapat diklasifikasikan menurut grup status penerumaan kartu kredit (disetujui atau tidak disetujui) dengan akurasi 61,8% (Poor) dan dengan error sebesar 38.20 yang artinya model ini kurang sesuai untuk menentukan penerimaan kartu kredit dengan kemungkinan a-priori disetujui sebesar 0.545 dan tidak disetujui sebesar 0.455

accuracy: 61.80%

	true Disetujui	true Tidak Disetujui	class precision
pred. Disetujui	384	221	63.47%
pred. Tidak Disetujui	161	234	59.24%
class recall	70.46%	51.43%	

Gambar 6 Hasil Pengujian Dengan Model Discriminant Analysis

Model Naive Bayes

Setelah langkah-langkah algoritma model Naive Bayes dilakukan, Maka diperoleh tingkat akurasi prediksi model ini dalam melakukan fungsi klasifikasi sebesar 90.61%, seperti terlihat pada gambar 4.5. Mengacu pada subab 2.2. tabel 2.4 mengenai tingkat akurasi klasifikasi, maka model naive bayes menempati tingkat akurasi

excellent, yang artinya model naïve bayes layak dan dapat digunakan untuk melakukan prediksi kelayakan pemberian fasilitas kartu kredit kepada nasabah bank XYZ.

classification_error: 9.39%			
accuracy: 90.61%			
	true Disetujui	true Tidak Disetujui	class precision
pred. Disetujui	411	75	84.57%
pred. Tidak Disetujui	0	313	100.00%
class recall	100.00%	80.67%	

Gambar 7 Hasil Pengujian Dengan Model Naïve Bayes

Dapat dilihat bahwa model ini memiliki error sebesar 9.39 % dengan true disetujui sebesar 411 dimana kasus nasabah yang diprediksi (Positif) disetujui pengajuan kartu kreditnya, benar (True) disetujui. False disetujui sebesar 75 data dimana nasabah yang diprediksi disetujui ternyata tidak disetujui yang berarti prediksi salah (false).

True tidak disetujui sebesar 313 data dimana kasus nasabah yang diprediksi (Positif) tidak disetujui pengajuan kartu kreditnya, benar (True) tidak disetujui. Sedangkan untuk false tidak disetujui sebesar 0 yang berarti tidak terjadi kesalahan prediksi dalam data tidak disetujui.

Class Recall dari pemodelan ini sebesar 100% dan 80.67% dimana rasio prediksi positif dibandingkan dengan kekeliruan data benar yang positif adalah 100% untuk kelas disetujui dan 80.67% untuk kelas tidak disetujui. Hal ini berbanding terbalik dengan class precision yang menunjukkan angka 84.57% dan 100 %. Class precision merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang positif sehingga dapat disimpulkan bahwa persentase nasabah benar disetujui kartu kreditnya dari keseluruhan nasabah yang diprediksi disetujui sebesar 84.57%, persentase nasabah yang diprediksi disetujui dibandingkan keseluruhan nasabah yang sebenarnya disetujui adalah sebesar 100% dan persentase nasabah benar tidak disetujui kartu kreditnya dari keseluruhan nasabah yang diprediksi tidak disetujui sebesar 100%, sedangkan persentase nasabah yang diprediksi tidak disetujui dibandingkan keseluruhan nasabah yang sebenarnya tidak disetujui sebesar 80.67%.

Selanjutnya berdasarkan pada gambar 10 dapat dianalisa bahwa atribut yang paling berpengaruh dalam penentuan kelayakan pemberian fasilitas kartu kredit kepada nasabah adalah penghasilan / tahun dengan bobot 0.204, lalu status rumah dengan bobot 0.082, jenis kelamin sebesar 0.057,

data pekerjaan 0.045, dan yang terakhir status perkawinan sebesar 0.014.

Tabel 2 hasil bobot dari setiap atribut

Attribute	Weight
Penghasilan / Tahun	0.204
Status Rumah	0.082
Jenis Kelamin	0.057
Pekerjaan	0.045
Status Kawin	0.014

Perbandingan Hasil Pemodelan

Pada bagian ini akan diuraikan mengenai perbandingan hasil pemodelan prediksi kelulusan mahasiswa dari ketiga model di atas: Discriminant Analysis dan Naïve Bayes berdasarkan hasil pengamatan penulis selama pemrosesan berlangsung, dapat dilihat pada tabel 3.

Kriteria	Model Prediksi	
	<i>Discriminant Analysis</i>	<i>Naïve Bayes</i>
Waktu pemrosesan	15 detik	12 detik
Kemudahan pemrosesan	Mudah	Mudah
Akurasi prediksi	61,80%	90,61%
Kemudahan pemahaman model/ fungsi / rule	Sedang	Mudah

Tabel 3 Perbandingan Hasil Model Prediksi

KESIMPULAN

Setelah melakukan proses analisis dan pembahasan hasil model prediksi penerimaan kartu kredit nasabah dalam penelitian yang telah dilakukan penulis di Bank XYZ hingga penyelesaiannya, maka dapat disimpulkan bahwa:

Pertanyaan-pertanyaan penulis dalam rumusan masalah dalam penelitian ini sebagian telah terjawab, yaitu:

- Pada model *Discriminant Analysis* dihasilkan dua buah fungsi untuk menentukan status predikat kelulusan seorang mahasiswa (Diterima atau Tidak Diterima) berdasarkan perhitungan dengan model *Discriminant Analysis* dihasilkan nilai akurasi sebesar 61.80 (poor).
- Pada model Naïve Bayes dihasilkan nilai akurasi sebesar 90,61 dengan faktor yang paling mempengaruhi profil nasabah

berdasarkan prediksi bobot setiap atribut adalah penghasilan / tahun dengan bobot 0.204, lalu status rumah dengan bobot 0.082, jenis kelamin sebesar 0.057.

Dengan membandingkan kedua model, maka dapat disimpulkan bahwa untuk mengetahui penerimaan kartu kredit nasabah bank XYZ dapat menggunakan model Naïve Bayes dengan nilai akurasi sebesar 90,61.

Technology, Vol. 2, Issue 9. ISSN : 2348-7968.

REFERENSI

- [1] Chapman, Peter. dkk. 2000. *CRISP-DM 1.0 Step-by-Step Data Mining Guide*.
- [2] Gorunescu, F. 2011. *Data Mining Concept Model and Techniques*. ISBN 978-3-642-19720-8. Berlin: Springer.
- [3] Gorunescu, F. 2011. *Data Mining Concept Model and Techniques*. ISBN 978-3-642- 19720-8. Berlin: Springer.
- [4] Han, J. dkk. 2012. *Data Mining: Concepts and Techniques 3rd edition*. Massachusetts (US): Morgan Kaufmann.
- [5] Larose, Daniel T dan Chantal D. Larose. 2015. *Data Mining and Predictive Analytics*. New Jersey: John Willey & Sons, Inc.
- [6] Patil, T. R., Sherekar, M. S., 2013. Performance Analysis of Naïve Bayes and J48 Classification Algorithm for Data Classification, *International Journal of Computer Science and Applications*, Vol. 6, No. 2, Hal 256-261.
- [7] Saleh, A. 2015. Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*, Vol. 2, No. 3, ISSN: 2354-5771
- [8] Santoso, Singgih. 2017. *Statistik Multivariat dengan SPSS*. Jakarta: PT Elex Media Komputindo.
- [9] Sumathi, S. dan Sivanadam. 2006. *Introduction to Data Mining and its Applications*. Verlag Berlin Heidelberg: Springer.
- [10] Susanto, Sani dan Suryadi, D. *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. CV. ANDI OFFSET, 2010.
- [11] Vembandasamy, K. dkk. 2015. Heart Diseases Detection Using Naïve Bayes Algorithm. *International Journal of Innovative Science, Engineering &*