

BAB II

TINJAUAN PUSTAKA

2.1 Regresi Logistik Biner

Regresi logistik biner merupakan suatu metode analisis regresi yang menggambarkan pola hubungan antara satu variabel respon (dependen) dengan satu atau lebih variabel prediktor (independen). Variabel respon terdiri dari dua kategori yaitu “sukses” dan “gagal” yang dinotasikan dengan $y = 1$ (sukses) dan $y = 0$ (gagal). Sehingga variabel akan mengikuti distribusi Bernoulli untuk setiap observasi tunggal. Fungsi probabilitas untuk setiap observasi adalah

$$f(y) = \pi^y(1 - \pi)^{1-y} \text{ dengan } y = 0,1 \quad (2.1)$$

Analisis regresi logistik merupakan suatu teknik untuk menganalisis data yang peubah responnya memiliki dua atau lebih kategori dengan satu atau lebih peubah penjelas yang berskala kategorik atau kontinu. Hosmer dan Lemeshow (2000) menjelaskan bahwa model regresi logistik dibentuk dengan nilai $P(Y = 1|x)$ sebagai $\pi(x)$, yang dinotasikan sebagai berikut:

$$\pi(x) = \frac{\exp(g(x))}{1 + \exp(g(x))} \quad (2.2)$$

Suatu fungsi dari $\pi(x)$ dicari dengan menggunakan transformasi logit, yaitu $g(x)$ yang dapat dinyatakan sebagai berikut:

$$g(x) = \ln \frac{\pi(x)}{1 - \pi(x)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_p x_p \quad (2.3)$$

2.2 Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasi probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua variabel prediktor independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kategori variabel respon (Patil dan Shereker, 2013). Definisi lain mengatakan Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya (Bustami, 2013).

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai variabel prediktor secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Ridwan, Suyono dan Sarosa, 2013). Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata kompleks dari pada yang diharapkan (Patterkari dan Parveen, 2012).

1. Teorema Bayes

Naive Bayes adalah algoritma yang digunakan dalam statistika untuk menghitung peluang dari suatu hipotesis, *Naive Bayes* menghitung peluang

suatu variabel respon berdasarkan pada variabel prediktor yang dimiliki dan menentukan variabel respon yang memiliki peluang paling tinggi. *Naive Bayes* mengklasifikasikan variabel respon berdasarkan pada probabilitas sederhana dengan mengasumsikan bahwa setiap variabel prediktor dalam data tersebut bersifat saling terpisah. *Naive Bayes* merupakan salah satu metode yang banyak digunakan berdasarkan beberapa sifatnya yang sederhana. *Naive Bayes* mengklasifikasikan data berdasarkan peluang P variabel prediktor X dari setiap variabel respon Y data. Pada model peluang setiap variabel respon dan jumlah variabel prediktor yang dapat dituliskan seperti persamaan berikut.

$$P(C_i|X_1, X_2, \dots, X_n) \quad (2.4)$$

Penghitungan *Naive Bayes* dapat dijelaskan dengan C_i adalah hipotesis data X yang merupakan suatu variabel respon $P(C_i|X)$ adalah peluang hipotesis variabel respon berdasarkan kategori X (*posteriori probability*). $P(C_i)$ adalah peluang hipotesis variabel respon (*prior probability*). $P(X|C_i)$ adalah peluang data X berdasarkan kategori pada hipotesis variabel respon. $P(X)$ adalah jumlah probabilitas data X yang nilainya adalah 1. Sehingga didapatkan rumus penghitungan *Naive Bayes* dituliskan pada persamaan.

$$P(C_i|X) = \frac{P(X|C_i) P(C_i)}{P(X)} \quad (2.5)$$

2. Naive Bayes Untuk Klasifikasi

Klasifikasi adalah proses pencarian sekumpulan model atau fungsi yang menggambarkan dan membedakan kategori variabel respon data dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kategori variabel respon dari suatu objek yang belum diketahui kategori variabel responnya. Selain itu, klasifikasi Naive Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam basis data dengan jumlah yang besar (Han dan Kamber, 2006). Salah satu penerapan theorem Bayes dalam klasifikasi adalah *Naive Bayes*.

$$P(X|C_i) = \prod_{k=1}^n P(X_k | C_i) \quad (2.6)$$

Untuk menentukan nilai variabel respon pada variabel prediktor tertentu digunakan persamaan:

$$\operatorname{argmax}_{c \in C} = \frac{P(X|C_i) P(C_i)}{P(X)} \quad (2.7)$$

2.3 Klasifikasi

Klasifikasi merupakan proses untuk menemukan sebuah model atau fungsi untuk menjelaskan dan membedakan kategori data atau konsep yang bertujuan untuk memprediksi atau memperkirakan kategori dari suatu objek dimana kategori variabel responnya belum diketahui (Han et al., 2012). Metode klasifikasi terdiri dari *Naive Bayes Classifier*, *Decision tree*, *k-Nearest Neighbors* (k-NN), *Neural Network*, Jaringan Syarat Tiruan (JST), Analisis statistik yaitu Regresi Logistik, Algoritma Genetik, *Rough Sets*, *Metode Rule Based* dan *Support Vector Machines* (SVM) (Gorunescu, 2011).

Ada tiga tahap dalam proses pembuatan model klasifikasi (Vercellis, 2009):

1. Tahap Pembelajaran

Tahap pembelajaran adalah tahap dimana algoritma klasifikasi diterapkan ke dalam data contoh untuk mendapatkan relasi data dalam setiap kategori variabel responnya. Tahap ini akan membentuk model yang berisikan aturan-aturan variabel prediktor dalam menentukan kategori variabel respon data.

2. Tahap Pengujian

Tahap pengujian adalah tahap penerapan aturan-aturan yang sudah terbentuk pada tahap pembelajaran ke dalam contoh data yang tidak termasuk dalam data pembelajaran. Dalam tahap ini, aturan model yang dimiliki akan diterapkan pada setiap variabel prediktor dalam data pengujian dan dilihat kecocokan antara kategori variabel respon yang diprediksi dengan kategori variabel respon pada data sebenarnya.

3. Tahap Prediksi

Tahap prediksi adalah tahap dimana model yang dihasilkan benar-benar diterapkan pada data yang belum diketahui variabel responnya. Penilaian algoritma klasifikasi biasanya dilihat dari akurasi model. Akurasi model merupakan ketepatan model dalam memprediksi kategori variabel respon data.

Adapun proses klasifikasi didasarkan pada empat komponen (Gorunescu, 2011):

1. Variabel Respon

Data yang berupa kategorikal yang merepresentasikan “variabel respon” yang terdapat pada objek. Contoh: resiko penyakit jantung, resiko kredit, customer loyalty, jenis gempa.

2. Prediktor (*predictor*)

variabel prediktor dari data yang diklasifikasikan berdasarkan klasifikasi yang dibuat. Contoh: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.

3. Data Pelatihan (*training dataset*)

Merupakan dataset yang berisi dua komponen nilai yang digunakan untuk pelatihan untuk mengenali model yang sesuai dengan kategori variabel responnya, berdasarkan prediktor yang ada.

4. Data Pengujian (*testing dataset*)

Merupakan dataset baru yang akan diklasifikasikan oleh model yang dibangun sehingga dapat dievaluasi hasil akurasi klasifikasi tersebut. Akurasi merupakan salah satu indikator yang dipakai dalam penelitian ini untuk mengukur performa dari metode yang diusulkan. Nilai akurasi dihitung dengan mengambil persentase prediksi yang benar dari keseluruhan data. Prediksi yang tepat berarti hasil dimana kategori variabel respon hasil prediksi adalah sama dengan kategori variabel respon dari data. Perhitungan akurasi dapat dilihat ditabel 2.1

Tabel 2.1 Perhitungan Akurasi

Label	Kenyataan Benar	Kenyataan Salah
Prediksi Benar	TP (True Positive)	FP (False Positive)
Prediksi Salah	FN (False Negative)	TN (True Negative)

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (2.8)$$

$$\text{Sensitifitas} = \frac{TP}{TP+FN} \times 100\% \quad (2.9)$$

$$\text{Spesifisitas} = \frac{TN}{FP+TN} \times 100\% \quad (2.10)$$

2.4 Kualitas Kredit

2.4.1 Kredit

Beberapa orang menganggap bahwa istilah kredit dalam konteks perbankan memiliki arti yang sama dengan hutang piutang pada umumnya, namun apabila dilihat lebih lanjut keduanya memiliki pengertian yang jauh berbeda (Wijaya, 2011). Dalam bahasa Latin, kredit disebut “creder” yang artinya percaya, yaitu kepercayaan dari kreditur bahwa debiturnya akan mengembalikan pinjaman beserta bunganya sesuai dengan perjanjian kedua belah pihak (Kasmir, 2005). Sedangkan menurut H. V. Rivai & Veithzal, 2006 kredit adalah penyerahan barang, jasa, atau uang dari satu pihak (kreditur/atau pemberi pinjaman) atas dasar kepercayaan kepada pihak lain (debitur atau pengutang/*borrower*) dengan janji membayar dari penerima kredit kepada pemberi kredit pada tanggal yang telah disepakati kedua belah pihak. Selain pengertian kredit di atas, berikut adalah beberapa pengertian kredit menurut:

Pengertian kredit menurut Undang-undang Perbankan No. 10 Tahun 1998 adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan jumlah bunga imbalan atau pembagian hasil keuntungan (Indonesia, 1998).

Ketentuan Pasal 1 angka 3 Peraturan Bank Indonesia 5 Oktober 2006, Nomor 8/19/PBI/2006 tentang Kualitas Aktiva Produktif dan Pembentukan Penyisihan Penghapusan Aktiva Produktif Bank Perkreditan Rakyat, kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara bank dengan pihak peminjam yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga.

Setelah membahas pengertian kredit maka akan dilanjutkan dengan unsur-unsur kredit. Unsur-unsur yang terkandung dalam pemberian suatu fasilitas kredit adalah sebagai berikut (Suyatno, dkk, 2003):

Kepercayaan yaitu keyakinan dari si pemberi kredit bahwa prestasi yang diberikannya, baik dalam bentuk uang, barang, atau jasa, akan benar-benar diterimanya kembali dalam jangka waktu tertentu di masa yang akan datang.

1. Tenggang Waktu

Yaitu suatu masa yang memisahkan antara pemberian prestasi dengan kontra prestasi yang akan diterima pada masa yang akan datang. Dalam unsur waktu ini, terkandung pengertian nilai dari uang, yaitu uang yang ada sekarang lebih tinggi nilainya dari uang yang akan diterima pada masa yang akan datang.

2. Degree of Risk

Yaitu tingkat resiko yang akan dihadapi sebagai akibat dari adanya jangka waktu yang memisahkan antara pemberian prestasi dengan kontra prestasi yang akan diterima kemudian hari. Semakin lama kredit yang diberikan, semakin tinggi pula tingkat resikonya karena sejauh jauh kemampuan manusia untuk menerobos hari depan itu, maka masih selalu terdapat unsur ketidaktentuan yang tidak dapat diperhitungkan. Inilah yang menyebabkan timbulnya unsur resiko. Dengan adanya unsur resiko inilah maka timbullah jaminan dalam pemberian kredit (Kasmir, 2011).

3. Prestasi

Prestasi atau objek kredit itu tidak saja diberikan dalam bentuk uang, tetapi juga dapat berbentuk barang atau jasa. Namun, karena kehidupan ekonomi modern sekarang ini didasarkan pada uang, maka transaksi-transaksi kredit yang menyangkut uanglah yang sering kita jumpai dalam praktek perkreditan (Suyatno et al., 2003).

2.4.2 Kualitas Kredit

Jumlah kredit yang disalurkan sangat berpengaruh terhadap hidup matinya lembaga keuangan. Banyaknya jumlah kredit yang disalurkan juga harus memperhatikan kolektibilitas kredit tersebut. Berdasarkan Peraturan Bank Indonesia No. 8/19/ 2006 Indonesia tentang Kualitas Aktiva Produktif dan Pembentukan Penyisihan Penghapusan Aktiva Produktif Bank Perkreditan Rakyat, membagi tingkat kolektibilitas kredit menjadi 4 (empat) golongan yaitu sebagai berikut:

1. Lancar

Kredit Lancar yaitu kualitas kredit apabila tidak terdapat tunggakan pembayaran pokok dan/atau bunga.

2. Kurang Lancar

Kredit Kurang Lancar adalah *sub-standard* yaitu kualitas kredit yang tingkat pengembaliannya mencerminkan keadaan yang kurang baik karena terdapat tunggakan pembayaran dalam jangka waktu tertentu yang ditetapkan.

3. Diragukan

Kredit Diragukan adalah *doubtfull loan* yaitu kredit yang digolongkan diragukan karena kredit yang bersangkutan tidak memenuhi kriteria lancar dan kredit kurang lancar serta akan mempengaruhi kualitas kolektibilitas kredit, tetapi berdasarkan penilaian, dapat disimpulkan bahwa kredit masih dapat diselamatkan dan agunan (nya) senilai sekurang-kurangnya 75% dari utang peminjam, termasuk bunga tersebut atau kredit tidak dapat

diselamatkan, tetapi agunannya masih bernilai sekurang-kurangnya 100% dari utang peminjam.

4. Macet

Kredit Macet adalah *bad debt* yaitu kredit yang tidak memenuhi kriteria lancar, kredit kurang lancar dan kredit diragukan serta akan berpengaruh terhadap kualitas kolektibilitas kredit, dan/atau, memenuhi kriteria diragukan, tetapi dalam jangka waktu 21 bulan sejak digolongkan diragukan belum ada pelunasan atau usaha penyelamatan kredit atau, penyelesaiannya telah diserahkan kepada Pengadilan Negeri atau Badan Urusan Piutang Negara (BUPN) atau telah diajukan penggantian ganti rugi kepada perusahaan asuransi kredit. Lalu, membagi kualitas kredit menjadi dua golongan yakni kredit tidak bermasalah dan kredit bermasalah. Kredit bermasalah adalah kredit dengan kualitas kurang lancar, diragukan dan macet (Riyadi, 2006).

2.4.3 Non Performing Loan (NPL)

Non Performing Loan (NPL) sering disebut kredit bermasalah, dapat diartikan sebagai rasio yang menunjukkan kemungkinan terjadinya risiko tidak tertagihnya piutang terhadap sejumlah pinjaman yang telah diberikan (V. Rivai, Veithzal, & Idroes, 2007). Kredit bermasalah adalah kredit dengan kualitas kurang lancar, diragukan dan macet (Riyadi, 2006). Menurut Dendawijaya, kemacetan fasilitas kredit disebabkan oleh 2 faktor yaitu (Dendawijaya, 2005):

1. Dari pihak perbankan

Dalam hal ini pihak analisis kredit kurang teliti baik dalam mengecek kebenaran dan keaslian dokumen maupun salah dalam menghitung rasio-rasio yang ada. Akibatnya, apa yang seharusnya terjadi, tidak diprediksi sebelumnya.

2. Dari pihak Debitur

Kemacetan kredit yang disebabkan debitur diakibatkan oleh adanya unsur kesengajaan ataupun tidak sengaja. Kriteria penilaian tingkat kesehatan rasio NPL dapat dilihat pada tabel dibawah ini (B. Indonesia, 2011):

Tabel 2.2 Tingkat Kesehatan Rasio NPL

Rasio	Predikat
$NPL \leq 5\%$	Sehat
$NPL \geq 5\%$	Tidak Sehat

