

Implementasi Algoritma *Naive Bayes* Pada Data Set Kualitatif Prediksi Kebangkrutan

Fakhriza Firdaus, Ali Mukhlis

Jurusan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Darwan Ali, Indonesia Email ¹takiya.firdaus@gmail.com, ²alimukhlis19@gmail.com

Submitted 20-12-2019; Accepted 15-01-2020; Published 15-02-2020

Abstrak

Sejumlah penelitan tentang prediksi kebangkrutan telah secara luas menerapkan teknik Data Mining untuk menemukan pengetahuan yang berguna secara otomatis berdasarkan penilaian terhadap penilaian manajemen resiko-resiko yang ada pada suatu perusahaan. Dalam proses penilaian resiko sebenarnya pengetahuan dari para ahli masih dianggap sebagai tugas penting karena prediksi para ahli tergantung pada sujektivitas mereka. Penelitian ini bertujuan untuk menggali informasi dari data set kualitatif prediksi kebangkrutan agar dapat digunakan sebagai sumber pembelajaran yang bermanfaat untuk memperbaiki manajemen suatu perusahaan. Teknik yang digunakan dalam penelitian ini yaitu Klasifikasi dengan menggunakan algoritma Naive Bayes. Naive Bayes menggunakan prediksi probabilistik untuk mengklasifikan data

Kata Kunci: Data Mining, Klasifikasi, Naive Bayes.

Abstract

A number of studies about bankruptcy prediction have widely applied the Data Mining technique to find useful knowledge automatically based on an assessment of the management's assessment of the risks that exist in a company. In the process of risk assessment the actual knowledge of experts is still considered an important task because the predictions of experts depend on their effectiveness. This study aims to extract information from qualitative bankruptcy data sets so that they can be used as a useful learning resource for improving the management of a company. The technique used in this study is classification using the Naive Bayes algorithm. Naive Bayes uses probabilistic predictions to classify data.

Keywords: Data Mining, Classification, Naive Bayes.

1. PENDAHULUAN

Salah satu bentuk kemajuan dari teknologi informasi adalah berkembangnya metode-metode untuk melakukan penggalian dan pengumpulan informasi. *Data mining* menggali informasi yang dikumpulkan dari pola-pola tersembunyi yang ada pada data, hubungan antar elemen-elemen data, ataupun pembuatan model untuk keperluan peramalan data.

Penulis akan melakukan penelitian klasifikasi menggunakan algoritma *Naive Bayes* pada Data Set prediksi Kebangkrutan Kualitatif dari website *UCI Machine Learning Repository*. Berdasarkan data yang diperoleh dari *UCI Machine Learning Repository* terdapat 250 prediksi kebangkrutan perusahaan berdasarkan tingkatan pengelolaan resiko yang ada. Dimana hasil prediksi didapatkan dari hasil analisa terhadap atribut tingkat resiko industri, resiko manajemen, fleksibilitas finansial, kredibilitas, kompetitif, resiko operasional.

Tujuan dari penelitian yang dilakukan adalah:

- a. Melakukan klasifikasi pada data set kualitatif prediksi kebangkrutan .
- b. Mengukur nilai akurasi pada data set kualitatif prediksi kebangkrutan dengan metode *data mining* klasifikasi menggunakan algoritma *Naive bayes*.

2. METODE PENELITIAN

2.1 Kebangkrutan

Kebangkrutan adalah situasi dimana perusahaan mengalami penurunan dana sehingga tidak mampu untuk melanjutkan usahanya. Menurut Toto (2011:332) Kebangkrutan (bankcruptcy) merupakan kondisi dimana perusahaan tidak mampu lagi untuk melunasi kewajibannya. Kondisi ini biasanya tidak muncul begitu saja di perusahaan, ada indikasi awal dari perusahaan tersebut yang biasanya dapat dikenali lebih dini kalau laporan keuangan dianalisis secara lebih cermat dengan suatu cara tertentu. Rasio keuangan dapat digunakan sebagai indikasi adanya kebangkrutan di perusahaan.

Menurut Undang-undang N0.4 tahun 1998 tentang kepailitan, menyatakan bahwa kebangkrutan sebagai suatu situasi yang dinyatakan pailit oleh keputusan pengadilan.^[1]

2.2 Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dari pola-pola tersembunyi pada data, hubungan antar elemenelemen data, ataupun pembuatan model untuk keperluan peramalan data.^[2]

2.3 Algoritma Klasifikasi Data Mining





Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Klasifikasi data terdiri dari dua langkah proses, yaitu :

- 1. Proses *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi.
- Proses klasifikasi, dimana data test digunakan untuk memperkirakan akurasi dari rule klasifikasi.
 Proses klasifikasi didasarkan pada empat komponen :
- a. Kelas

Variabel dependen yang berupa kategorikal yang merepresentasikan "label" yang terdapat pada objek.

- b. Predictor
 - Variabel independen yang direpresentasikan oleh karakteristik (atribut) data.
- c. Training dataset

Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.

d. Testing dataset

Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

2.4 Algoritma Naive bayes

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediski probabilitas keanggotaan suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decesion tree dan neural network. Bayesian classification terbukti memiliki akurasai dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. (Kusrini,2009).^[3] Teorema Bayes memiliki bentuk umum sebagai berikut:

$$P(H \mid X) = \frac{P(X \mid H)P(H)}{P(X)} \tag{1}$$

Keterangan:

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

2.5 Weka

Weka adalah sebuah perangkat lunak yang memiliki banyak algoritma machine learning untuk keperluan data mining. Weka juga memiliki banyak tools untuk pengolahan data, mulai dari pre-processing, classification, regression, clustering, association rules, dan visualization. Weka adalah perangkat lunak open source berbasis Java dan kita dapat menggunakannya secara langsung atau melalui program Java kita. Weka juga bisa diimplementasikan ke program python. [4]

3. ANALISA DAN PEMBAHASAN

3.1 Persiapan Data Set

Data set kualitatif prediksi kebangkrutan di dapatkan dari halaman web UCI Machine Learning Repository. Link untuk mendapatkan data set adalah : https://archive.ics.uci.edu/ml/datasets/Qualitative Bankruptcy .



Gambar 1. Halaman data set kualitatif prediksi kebangkrutan.



3.2 Mendeskripsikan data

Dataset kualitatif prediksi kebangkrutan terdiri dari 6 atribut yang antara lain :

1. IR = Industry Risk

2. MR = Management Risk

3. FF = Financial Flexibility

4. CR = Credibility
5. CO = Competitiveness
6. OP = Operating Risk

Dan nilai masing-masing atribut antara lain:

1. P = Positive

2. A = Average

3. N = Negative

Dan diklasifikasikan menjadi 2 yaitu:

1. B = Bankruptcy

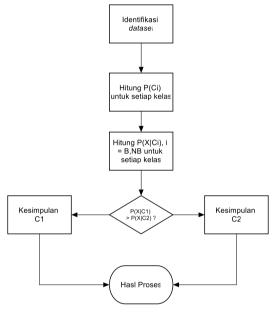
2. NB = Non-Bankruptcy

3.3 Pemodelan data

Pemodelan adalah fase yang secara langsung melibatkan teknik *data mining* yaitu dengan melakukan pemilihan teknik *data mining* dan menentukan algoritma yang akan digunakan.

a. Pemilihan teknik pemodelan

Tool yang digunakan adalah WEKA versi 3.8.3 :



Gambar 2. Flowchar WEKA

Dari data kelas yang ada, kemudian hitung P(Ci) untuk setiap atribut, dalam kasus dataset pada penelitian ini yaitu atribut *class* yang terdiri dari dua kelas yaitu *Bankruptcy* dinyatakan dengan "B" dan *Non-Bankruptcy* dinyatakan dengan "NB".

Kemudian hitung P(X|Ci), i=1,2 untuk setiap kelas atau atribut. Setelah itu bandingkan, jika P(X|C1) > P(X|C2) maka kesimpulannya C1 yaitu Bankruptcy = "B". Jika jika P(X|C1) < P(X|C2) maka kesimpulanya C2 yaitu Non-Bankruptcy = "NB".

b. Perhitungan Data Mining

Berikut perhitungan manual naive bayes dengan menggunakan dataset kualitatif prediksi kebangkrutan jika data raining dan data testing dipilih secara acak :

Tabel 1. Penjelasan data training dan data testing

IR	MR	FF	CR	CO	OP	Class
A	A	P	Α	P	A	NB
P	N	Α	Α	Α	P	NB
N	P	Α	Α	A	N	NB
P	N	Α	Α	Α	N	NB



P	N	P	P	P	P	NB	
N	P	P	P	P	N	NB	
P	N	P	P	P	N	NB	
N	N	A	P	P	P	NB	
P	N	P	Α	A	P	NB	
N	P	A	P	A	P	NB	
N	P	A	Α	P	N	NB	
A	N	N	N	N	A	В	
P	N	N	N	N	N	В	
N	P	N	N	N	N	В	
A	P	N	Α	N	N	В	
A N	N	N	N	N	N	В	
N	N	N	Α	N	Α	В	
N	N	N	N	N	P	В	
N	N	N	N	N	A	В	
N	N	N	A	N	P	В	

Keterangan:

Data Training
 Data Testing

Menghitung jumlah class berdasarkan klasifikasi yang terbentuk (Prior Probability):

- 1. C1 (Class = "B") = jumlah "B" pada kolom class = 9/20 = 0.45
- 2. C2 (Class = "NB") = jumlah "NB" pada kolom class = 11/20 = 0.55

c. Menghitung jumlah kasus yang sama pada setiap atribut dari Class (B/NB) berdasarkan data testing

Tabel 2. Mencari P(IR = "N" | Class = "B")

IR	Class
Α	NB
P	NB
N	NB
P	NB
P	NB
N	NB
P	NB
N	NB
P	NB
N	NB
N	NB
Α	В
P	В
N	В
Α	В
N	В
N	В
N	В
N	В
N	В

- 1. P(IR = "N" | Class = "B") = 6/9 = 0,666667
- 2. P (IR = "N" | Class = "NB") = 5/11 = 0,454545

(Hitung Probabilitas dari seluruh atribut IR, MF, FF, CR, CO, OP seperti contoh gambar 3.3).**

d. Kalikan semua hasil variabel

- a. Untuk semua atribut class = "B" $P(X|Class = "B") = 0,666667 \times 0,777778 \times 1 \times 0,333333 \times 1 \times 0,222222 \\ = 0,038409$
- b. Untuk semua atribut class = "NB" P(X|Class = "NB") = 0,454545 x 0,545455 x 0 x 0,545455 x 0 x 0,454545 = 0
- c. Perkalian prior probability dengan semua atribut class "B" $P(Ci) \mid Class = "B") \times P(X \mid Class = "B") = 0,45 \times 0,38409 = 0,1728405$





d. Perkalian prior probability dengan semua atribut class "NB"
 P(Ci) | Class = "NB") x P(X|Class = "NB") = 0,55 x 0
 = 0

e. Bandingkan hasil kelas

1. P(Ci) | Class = "B") x P(X|Class = "B") > P(Ci) | Class = "NB") x P(X|Class = "NB")

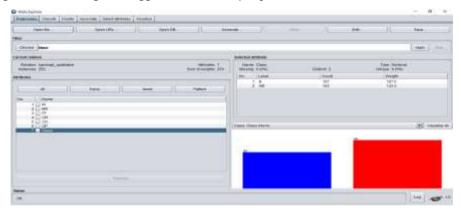
Kesimpulan:

Class = "B"

(Perbandingan antara hasil perkalian class "B" dengan class "NB" menunjukan bahwa nilai class = "B" lebih besar dibandikan class = "NB" sehingga didapatkan keputusan bahwa untuk record IR = "N", MR= "N", FF= "N", CR= "A", CO= "N", OP= "P" masuk ke class "B".

f. Implementasi dengan WEKA

Berikut adalah pengolahan data dengan menggunakan naive bayes pada WEKA:



Gambar 3. Preprocessing data pada WEKA



Gambar 4. Pemodelan naive bayes pada WEKA

- 1. Data Preparation atau bisa disebut juga dengan data preprocessing adalah suatu proses/langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas(input yang baik untuk data mining tools).^[5]
- 2. Pemodelan adalah tahapan dalam membuat model dari suatu sistem nyata
- 3. WEKA merupakan software terintegrasi yang berisi implementasi dari metode-metode data mining.
- 4. 10 Fold-Cross validation untuk membantu menghasilkan tingat keakurasian berdasarkan dataset kualitatif prediksi kebangkrutan yang telah dilakukan klasifikasi.





g. Hasil Pengujian dan Percobaan

Gambar 5. Classifier Output

Pada percobaan dengan algoritma naive bayes dengan menggunakan tools WEKA diperoleh waktu komputasi adalah 0 second. Dimana dapat diartikan bahwa komputasi menggunakan naive bayes berjalan cukup cepat. Hal ini sesuai dengan kelebihan naive bayes dibandingkan beberapa algoritma lain seperti neural network yang membutuhkan waktu yang cukup lama untuk melakukan komputasi data. Dan pada gambar 3.5 juga menunjukan tingkat akurasinya adalah 99.2% artinya model klasifikasi data kualitatif prediksi kebangkrutan menggunakan naive bayes terbukti baik. Hal ini dapat dilihat dari tingkat akurasinya yang mencapai 99.2% akan tetapi hl ini perlu di tinjau ulang dari sudut pandang kompleksitas dan jumlah datasetnya.

4. KESIMPULAN

Dari hasil observasi terhadap penerapan data mining klasifikasi menggunakan algoritma naive bayes pada dataset kualitatif prediksi kebangkrutan didapatkan sebuah hasil bahwa nilai akurasi terhadap klasifikasi prediksi kebangkrutan adalah sebesar 99,2%. Dimana hasil tersebut bisa juga disebabkan oleh kurangya kompleksitas data yang mengakibatkan model dapat memprediksi dengan akurat.

Peneliti telah membahas penggunaan serta perhitungan akurasi menggunakan 10 fold-cross validation dengan aplikasi WEKA pada metode naive bayes pada dataset kualitatif prediksi kebangkrutan, diharapkan pada penelitian selanjutnya dapat dibandingkan dengan memanfaatkan metode klasifikasi lainnya seperti metode C4.5, metode nearest neighbor guna menentukan kelas berdasarkan atribut-atribut yang telah ditentukan sehingga dengan menggunakan banyak metode dapat mengetahui kelebihan masing-masing metode.

REFERENCES

- Dr.M.Handi Shubhan., SH., MH., M.Si, Hukum Kepailitan. Kencana, 2015.
- L. Muflikhah, D. E. Ratnawati, dan R. R. MP, Data Mining. Universitas Brawijaya Press, 2018. 12.1
- [3] E. luthfi dan U. Amikom, Algoritma Data Mining. Penerbit Andi.
- S. Adinugroho dan Y. A. Sari, Implementasi Data Mining Menggunakan Weka. Universitas Brawijaya Press, 2018.
- S. García, J. Luengo, dan F. Herrera, Data Preprocessing in Data Mining. Springer International Publishing, 2014.