

Article history

Received Okt 22, 2020

Accepted Dec 26, 2020

PERBANDINGAN METODE NAÏVE BAYES DAN EQUIVALENCE CLASSES DALAM MENENTUKAN STRUKTUR BAYESIAN NETWORK

Yussyafri¹, Maftahatul Hakimah², Farida³

^{1,2,3} Teknik Informatika, Institut Teknologi Adhi Tama Surabaya

email: cr.yussyafri@gmail.com, hakimahmafta@itats.ac.id, farida@itats.ac.id

Abstract

Bayesian Network is a model included in Bayes classification. This method assumes that independent variable value has a dependency to the value of other variables. Bayesian network has advantage, namely it is able to be relationship model between variables using graphs or a kind of depiction of the relationship flow between variables. There were several methods used to decide Bayesian network structure. Forming methods of Bayesian network structure in this study were naïve bayes dan equivalence classes method. Both method formed the structure were applied to classification of the feasibility of borrowing funds for Micro, Small and Medium Enterprises (MSMEs). In naïve bayes method structure, dependent variable became a center of independent variable and equivalence classes method structure, each variable has relationship between other variables. The average test results of naïve bayes method and equivalence classes in the form of Bayesian network structure were equivalence classes method of 79.53% and naïve Bayes of 80.93%.

Keywords: Bayesian Network, MSMEs, Classification.

Abstrak

Bayesian Network merupakan model yang termasuk dalam klasifikasi bayes, dimana metode ini mengasumsikan bahwa nilai variabel independen memiliki ketergantungan dengan nilai variabel lain. Bayesian Network memiliki keunggulan yaitu dapat memodelkan hubungan antar variabel dengan menggunakan graf atau semacam penggambaran alur hubungan antar variabel. Terdapat beberapa metode yang digunakan untuk menentukan struktur bayesian network. Metode pembentukan struktur jaringan Bayesian network pada penelitian ini adalah metode naïve bayes dan equivalence classes. Kedua metode pembentukan struktur ini diterapkan untuk klasifikasi kelayakan peminjaman dana Usaha Kecil Mikro Menengah (UMKM). Pada struktur metode naïve bayes variabel dependen menjadi pusat dari variabel independen sedangkan pada struktur metode equivalence classes setiap variabel memiliki hubungan antar variabel lain. Hasil pengujian dari metode naïve bayes dan equivalence classes dalam pembentukan struktur Bayesian network secara rata-rata adalah metode equivalence classes 79,53% dan naïve bayes 80,93%.

Kata Kunci: bayesian network, umkm, klasifikasi.

1. PENDAHULUAN

Perkembangan ekonomi di Indonesia semakin berkembang dan meningkat dari tahun ke tahun, salah satu sektor yang ikut menyumbang atau berkontribusi dalam perkembangan ekonomi di Indonesia adalah sektor Usaha Kecil Mikro Menengah (UMKM). Namun seiring berjalannya waktu persaingan usaha semakin meningkat, banyak UMKM yang tidak mampu bertahan

karena faktor internal maupun eksternal. Salah satu masalah internal bagi pelaku usaha UMKM adalah dana (Kemenperin, 2019). Modal dan dana operasional sangat penting untuk menunjang kebutuhan selama UMKM berjalan. Secara bertahap namun pasti penyaluran dana berupa kredit atau peminjaman dana ke sektor UMKM mengalami pertumbuhan. Proses penentuan kelayakan mendapatkan pinjaman umumnya dapat diselesaikan salah satunya dengan cara

pembobotan nilai kredit. Tujuannya untuk menentukan apakah calon peminjam termasuk layak atau tidak layak untuk mendapatkan pinjaman. Dalam prakteknya pembobotan nilai kredit menggunakan berbagai aspek di dalam laporan kredit dan sumber informasi pendukung lainnya.

Metode Bayesian Network adalah model yang termasuk dalam klasifikasi bayes, dimana metode ini mengasumsikan bahwa nilai variabel independen memiliki ketergantungan dengan nilai variabel lain (Jiawei Han, 2012). Metode ini mampu memprediksi kemungkinan atau probabilitas dari suatu peristiwa kemudian menggambarannya dalam bentuk sebuah graf. Salah satu kelebihan dari metode ini adalah metode ini mampu menunjukkan hubungan sebab akibat dari sejumlah variabel yang terkait, sehingga setiap variabel mempunyai tingkat independensi masing-masing terhadap variabel lain. Setiap variabel pengamatan pada Bayesian Network saling berkaitan satu sama lain. Metode ini termasuk dalam kategori supervised learning dimana terdapat data training dan testing dalam pemrosesan klasifikasi dengan metode ini. Metode ini mempelajari pola yang didapat dari data training dan di terapkan pada data testing untuk pengujian data yang baru. Pada penelitian sebelumnya adalah membandingkan metode naïve bayes, maximum spanning tree dan equivalence classes sebagai struktur pembentukan jaringan Bayesian network (Purwadi, 2009) (Frendy, 2014). Metode yang digunakan pada penelitian kali ini adalah metode equivalence classes dan naïve bayes sebagai struktur pembentukan jaringan dari metode Bayesian network. Dimana pada metode equivalence classes antar variabel memiliki hubungan satu sama lain sedangkan pada metode naïve bayes variabel dependen menjadi pusat dari variabel independen. Tujuan dari penelitian ini adalah untuk mengetahui struktur pembentukan jaringan mana yang memiliki akurasi lebih besar.

2. METODE PENELITIAN

A. Bayesian Network

Bayesian Network merupakan model yang termasuk dalam klasifikasi bayes, dimana metode ini mengasumsikan bahwa nilai variabel independen memiliki ketergantungan dengan nilai variabel lain (Jiawei Han, 2012). Bayesian Network mempunyai komponen yang terdiri dari

kualitatif dan kuantitatif, pada komponen kualitatif bentuk model grafiknya berupa directed acyclic graph (DAG) dimana suatu edges menjelaskan hubungan keterikatan antar variabel yang terkait, selain itu untuk membangun Bayesian Network dengan komponen ini maka proses yang dikerjakan adalah mencari hubungan antar variabel. Sedangkan untuk komponen kuantitatif representasinya berupa prior probability, joint probability (probabilitas bersama), posterior probability, dan conditional probability table (CPT). Bayesian Network merupakan sebuah model grafik yang digunakan untuk merepresentasikan/menggambarkan interaksi antar variabel yang berkaitan. Selain itu model ini digunakan untuk mengetahui informasi hubungan probabilitas dari seluruh variabel dalam sebuah data. Struktur Bayesian Network direpresentasikan oleh sebuah graf yang terdiri dari sebuah simpul awal (node) dan busur (arc) atau biasa disebut juga dengan edges, busur/arah panah ini digunakan untuk mengetahui hubungan antar variabel terkait dengan sebab dan akibat. Jika terdapat sebuah variabel yang dihubungkan oleh busur maka artinya bahwa variabel tersebut saling berkaitan dan memiliki pengaruh satu sama lain, pengaruh tersebut nantinya dinyatakan dalam sebuah peluang bersyarat. Bayesian Network merupakan salah satu dari classifier bayes sehingga formula yang digunakan adalah aturan bayes.

$$P(H|E) = \frac{P(EVH)}{P(E)} P(H)$$

H = hipotesis (kelas yang akan di prediksi)

E = evidence (sekumpulan observasi pada data)

Untuk membuat struktur jaringan pada Bayesian Network ada beberapa metode yang dapat digunakan seperti, Naive Bayes, Chow-liu Tree, Tree-Augmented Naive Bayes (TAN), Maximum Spanning Tree, Equivalence Classes dan lain-lain. Pada penelitian ini metode yang digunakan adalah Equivalence Classes dan Naive Bayes.

a. Metode Equivalence Classes

Menurut (Jouffe, 2008) metode Equivalence Classes digunakan untuk menemukan hubungan antar variabel pada data. Uji kebebasan chi-square menjadi hal yang utama yang harus dikerjakan oleh metode ini. Pengujian chi-square digunakan untuk menyeleksi variabel-variabel yang saling bebas antara satu sama lain. Nilai dari hasil uji chi-square ini yang akan dijadikan acuan

dalam membangun struktur jaringan Bayesian Network.

b. Metode Naïve Bayes

Metode Naive Bayes merupakan salah satu metode yang digunakan untuk membentuk struktur Bayesian Network yang paling sederhana. Metode ini membangun struktur dengan cara menghubungkan semua variabel independen ke variabel dependen. Arah edges struktur pada metode ini semuanya mengarah ke variabel dependen. Metode ini menggambarkan variabel independen satu dengan yang lain tidak memiliki hubungan (saling bebas).

B. Confusion Matrix

Confusion matrix merupakan metode yang digunakan untuk mengukur kinerja dari klasifikasi yang dilakukan dengan cara membandingkan antara hasil prediksi yang didapatkan dari data testing dengan data training. Kuantitas matriks konfusi diringkas menjadi 3 nilai yaitu akurasi, presisi dan recall. Setelah melalui tahap pengukuran kinerja klasifikasi menggunakan konfusi matriks, maka dapat disimpulkan seberapa besar akurasi, presisi dan recall data testing terhadap data training dan apakah perlu tahapan lebih lanjut jika akurasi, presisi dan recall belum sesuai harapan.

Umumnya confusion matrix dilakukan setelah tahapan data cleaning (preprocessing) dan klasifikasi pada data, setelah proses tersebut sudah dilakukan maka confusion matrix dapat diterapkan. Selain itu terdapat beberapa cara untuk mengukur kinerja tergantung kebutuhan dan jenis metode yang digunakan. Rumus akurasi, presisi dan recall pada confusion matrix adalah sebagai berikut :

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

Dimana,

TP = True Positif (jumlah data positif yang terklasifikasi dengan benar)

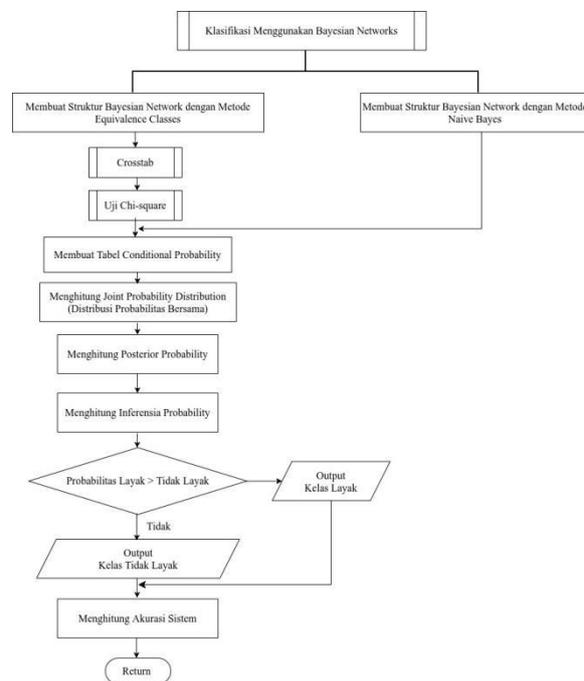
TN = True Negatif (jumlah data negatif yang terklasifikasi dengan benar)

FP = False Positif (jumlah data positif yang salah terklasifikasi)

FN = False Negatif (jumlah data negatif yang salah terklasifikasi)

Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Nilai Recall menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem.

Metode penelitian yang dilakukan pada penelitian ini yaitu studi literatur, perancangan dan implementasi sistem, analisa, dan pengujian. Berikut adalah gambar blok diagram sistem seperti pada Gambar 1.



Gambar 1. Flowchart Klasifikasi Menggunakan Bayesian Network.

Pada penelitian ini dataset yang digunakan adalah data yang diambil dari <http://umkm.kendalkab.go.id>. Berikut tahapan yang dilakukan pada penelitian :

Langkah 1. Preprocessing data :

Data melalui tahapan preprocessing dimana pada proses ini terdapat beberapa proses yang harus dilakukan guna mendapatkan data yang

berkualitas baik seperti valid, akurat, dapat diulang, konsisten dan seragam. Preprocessing juga melakukan penanganan pada missing value, standarisasi format data dan pengkodean data.

Langkah 2. Menyusun Struktur Bayesian Network :

Pembentukan struktur jaringan Bayesian network, Setelah data melalui tahapan proses preprocessing maka langkah selanjutnya adalah proses pembentukan struktur Bayesian network menggunakan metode equivalence classes dan naïve bayes.

Langkah 2a. Struktur Bayesian Network dengan Equivalence Classes :

Pada pembentukan struktur equivalence classes terdapat beberapa proses seperti crosstab, uji chi-square. Pada metode equivalence classes proses yang harus dilakukan adalah proses crosstab terlebih dahulu, yaitu melakukan penggambaran data dengan cara mengeksplorasi data untuk mengetahui karakteristik pada data tersebut. Kemudian setelah proses crosstab telah dilakukan proses selanjutnya adalah uji chi-square, pengujian dilakukan untuk menguji tingkat independensi dan dependensi variabel pengamatan pada data. Hasil dari pengujian chi-square ini digunakan untuk membuat struktur bayesian network menggunakan metode equivalence classes . Cara untuk membuat struktur bayesian network adalah dengan mengidentifikasi variabel yang berpengaruh pada variabel lain berdasarkan tingkat independensinya, dengan cara mencari urutan suatu kejadian sehingga dapat membentuk klasifikasi kelas yang diprediksi. Kemudian membuat tabel conditional probability, menghitung nilai joint probability distribution, posterior probability, inference probability.

Langkah 2b. Struktur Bayesian Network dengan Naïve Bayes :

Pada metode naïve bayes struktur telah ditentukan sebelumnya yaitu kelas merupakan pusat dari semua variabel, karena struktur telah ditentukan maka proses berikutnya sama dengan metode equivalence classes yaitu membuat tabel conditional probability, menghitung nilai joint probability distribution, posterior probability, inference probability.

Langkah 3. Mencetak Kelas :

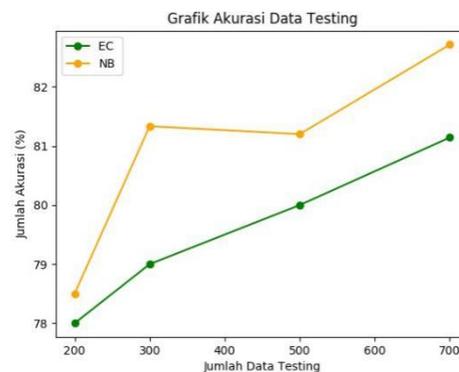
Pada proses ini setelah mendapatkan nilai akhir dari metode Bayesian network dengan struktur equivalence classes dan naïve bayes maka mencetak nilai kelas yang memiliki probabilitas lebih besar.

Langkah 4. Menghitung Akurasi :

Pada proses ini perhitungan akurasi sistem dilakukan menggunakan confusion matrix pada kedua metode pembentukan struktur jaringan Bayesian network.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan pengujian menggunakan beberapa jumlah data *testing* sebagai parameter jumlah data *testing*. Parameter jumlah data *testing* ini adalah 200, 300,500 dan 700 data *testing*. Berikut adalah gambar akurasi dari pengujian pada parameter jumlah data *testing* pada sistem seperti pada Gambar 2.



Gambar 2 Hasil Pengujian Akurasi

Gambar 2 menunjukkan perbandingan jumlah akurasi pada metode *equivalence classes* dengan *naïve bayes* terhadap parameter jumlah data *testing*. Berdasarkan Gambar 2 akurasi pada metode *naïve bayes* lebih besar dari *equivalence classes* dan terdapat kenaikan jumlah akurasi data pada kedua metode dari pengujian data *testing* sebelumnya namun pada *naïve bayes* terjadi penurunan akurasi dari 300 data *testing* ke 500 data *testing*. Meskipun *naïve bayes* mengalami penurunan akurasi namun rata-rata *naïve bayes* lebih tinggi dari *equivalence classes*. Hasil pengujian akurasi dari keseluruhan data *testing* dapat disimpulkan dalam Tabel 1 berikut ini.

Tabel 1. Hasil Pengujian Akurasi

Metode	Akurasi %				Rata-rata
	200 data	300 data	500 data	700 data	
Equivalence Classes	78	79	80	81,1	79,53
Naïve Bayes	78,5	81,33	81,2	82,7	80,93

Berdasarkan Tabel 1 dan Gambar 2 dapat disimpulkan bahwa akurasi kinerja sistem pada metode *equivalence classes* mengalami kenaikan secara bertahap dan pada metode *naïve bayes* akurasi mengalami penurunan pada 500 data testing sebesar 0,1% namun mengalami kenaikan kembali pada 700 data testing. Nilai akurasi yang terbesar dalam penelitian ini adalah 82,7% dari pengujian metode *naïve bayes* dengan 700 data testing dan akurasi terendah adalah 78% dari pengujian metode *equivalence classes* dengan menggunakan 200 data testing. Rata-rata parameter jumlah testing dari metode *equivalence classes* adalah 79,53% dan *naïve bayes* 80,93%.

Pembentukan struktur dari dataset UMKM Kendal menggunakan metode *equivalence classes* sangat berpengaruh pada proses *crosstab* dan uji *chi-square* karena pada kedua proses tersebut menentukan *node* yang saling terhubung. Jumlah *node* dan arah panah yang digunakan untuk proses metode *equivalence classes* dan *naïve bayes* sangat berpengaruh pada proses perhitungan *inference probability*.

4. PENUTUP

Kesimpulan

1. Nilai akurasi terbesar yaitu 82,7% pada *bayesian network* menggunakan metode *naïve bayes* terhadap 700 data testing dan nilai akurasi terendah adalah 78% dari pengujian metode *equivalence classes* terhadap 300 data testing. Rata-rata akurasi metode *equivalence classes* dalam membentuk struktur *bayesian network* sebesar 79,53%, sedangkan akurasi pada metode *naïve bayes* mempunyai rata-rata 80,93%.
2. Struktur pada *bayesian network* dengan metode *equivalence classes* dapat berubah jika ada pengurangan atau penambahan pada data *training*, terutama dalam hal hubungan antar variabel. Karena nilai hasil *crosstab* dan uji

chi-square juga akan berubah mengikuti pola data yang terbaru. Sedangkan pada struktur metode *naïve bayes* tidak berubah karena variabel dependen tetap menjadi pusat dari variabel independen.

Saran

Perlu dilakukan uji coba metode bayesian network dengan pembentukan struktur yang berbeda seperti metode maximum spanning tree, tree augmented naïve bayes (TAN) dan lain-lain.

5. REFERENSI

- Kemenperin. (2019, April 04). Diambil kembali dari <https://kemenperin.go.id/artikel/14200/Kontribusi-UMKM-Naik>
- Jiawei Han, M. K. (2012). Data Mining Concepts and Techniques Third Edition. Dalam M. K. Jiawei Han, Data Mining Concepts and Techniques Third Edition (hal. 394-397).
- Purwadi, I. (2009). Penerapan Bayesian Network Dalam Penetapan Daerah Tertinggal. Penerapan Bayesian Network Dalam Penetapan Daerah Tertinggal.
- Prasetyo, E. (2012). Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab. Yogyakarta: CV.ANDI.
- Jouffe. (2008). BayesiaLab Online Help. Diambil kembali dari BayesiaLab: http://www.bayesia.com/Result_Analysis/BayesiaLab-_satisfaction-_poll-_analysis_BAK.pdf
- <https://www.slideshare.net/hendrikarisma/bayesian-belief-network>
- Khoerudin, Asep, Aji Hamim Wigena, dan Anang Kurnia (2011). Analisis Tingkat Kesukaan Konsumen dengan Metode Bayesian Network (Studi Kasus Produk Biskuit).
- Frendy, I. (2014). Pembentukan Model Credit Scoring Dengan Menggunakan Metode Bayesian Network: Studi Kasus Permohonan Aplikasi Kredit Pemilikan Rumah (KPR).
- <https://achmatim.net/2017/03/19/mengukur-kinerja-algoritma-klasifikasi-dengan-confusion-matrix/>.
- <https://cs.ipb.ac.id/~yeni/files/ai/Kuliah%2012%20%20Bayesian%20Network%20print.pdf>.