

Penerapan Metode *Naive Bayes Classifier (Nbc)* Untuk Klasifikasi Kondisi Internal Program Studi

¹Indhitya R. Padiku

¹Program Studi Sistem Informasi, Fakultas Teknik, Universitas Negeri Gorontalo
email: indypadiku@ung.ac.id

Abstrak

Rendahnya minat calon mahasiswa pada program studi di Perguruan Tinggi dapat dipengaruhi oleh faktor internal program studi. Faktor-faktor tersebut menjadi variabel utama dalam menilai kondisi program studi. Untuk itu perlu dilakukan klasifikasi kondisi internal program studi mendapatkan hasil klasifikasi akhir dari penilaian program studi tersebut diperlukan metode yang baik dalam hal akurasi dan minim tingkat kesalahan klasifikasi. Tujuan penelitian ini adalah untuk melakukan klasifikasi kondisi internal program studi. Klasifikasi kondisi internal program studi dilakukan dengan metode *Naive Bayes Classifier (NBC)* yang merupakan bentuk sederhana dari Bayesian Network dengan asumsi bahwa semua fitur adalah independen satu sama lain. Metode NBC menunjukkan kinerja yang unggul secara keseluruhan dalam hal akurasi dan tingkat kesalahan klasifikasi. Metode NBC dapat digunakan untuk mengetahui kondisi internal program studi, sehingga dapat diketahui faktor-faktor yang perlu dibenahi untuk meningkatkan minat calon mahasiswa mendaftar di program studi.

Kata kunci: NBC; klasifikasi; kondisi internal; program studi; minat; calon mahasiswa.

Abstract

The low interest of prospective students in study programs at universities could be influenced by internal factors within the study program. These factors become the main variables in assessing the condition of the study program. For this reason, it is necessary to classify the internal conditions of the study program. A good method is needed in terms of accuracy and minimal misclassification to obtain the final classification results of the assessment. The purpose of this research is to classify the internal conditions of the study program. Classification of the internal conditions of the study program was carried out using the Naive Bayes Classifier (NBC) method which is a simple form of Bayesian Network with the assumption that all features are independent of each other. The NBC method shows an overall superior performance in terms of accuracy and misclassification rate. The NBC method can be used to determine the internal conditions of the study program, which could help identify factors that need to be addressed to increase the interest of prospective students enrolling in the study program.

Keywords: NBC; classification; internal conditions; study program; interest; prospective students.

Diterima: 12 Oktober 2020
Disetujui: 24 Juni 2021
Dipublikasi: 30 Juni 2021

©2021 Indhitya R. Padiku
Under the license CC BY-SA 4.0

Pendahuluan

Saat ini institusi Perguruan Tinggi berada dalam lingkungan yang sangat kompetitif. Setiap Perguruan Tinggi berusaha untuk terus memperbaiki menejemennya untuk meningkatkan mutu pendidikan dan meningkatkan akreditasi (BANPT, 2007). Jurusan terdiri dari beberapa program studi (prodi) yang melaksanakan fungsi pendidikan, penelitian serta pengabdian kepada masyarakat dan mengelola ilmu pengetahuan yang sesuai dengan bidang studi. Setiap prodi harus mampu meningkatkan kualitas secara berkelanjutan yang nantinya akan diimplementasikan kepada mahasiswa. Beberapa

faktor yang mempengaruhi perkembangan tiap prodi baik secara langsung maupun tak langsung dapat mempengaruhi jumlah mahasiswa baru, faktor-faktor utama tersebut diantaranya jumlah dosen, jumlah mahasiswa serta ketersediaan sarana prasarana (Vallejos, Alvarado & Puente, 2012). Pada setiap prodi mempunyai permasalahan yang dapat dianalisis berdasarkan data yang dikumpulkan, teknik pengumpulan data memerlukan metode tertentu, penggunaan metode yang tepat dapat menjadi informasi dan pengetahuan sebagai bahan pengamatan dan pertimbangan dalam memprediksi perkembangan prodi. Pimpinan fakultas dapat melakukan evaluasi awal pada prodi, sehingga pimpinan dapat secara proporsional menentukan kuota jumlah mahasiswa pada penerimaan mahasiswa baru.

Algoritma klasifikasi yang cukup populer adalah decision tree, naïve bayes, dan neural network (Gorunescu, 2011). Metode *naïve bayes classifier* (NBC) merupakan bentuk sederhana dari sebuah *Bayesian network* yang merupakan salah satu metode data mining, kesederhanaan metode ini bergantung pada asumsi bahwa semua fitur adalah independen satu sama lain. NBC mempunyai keunggulan akurasi dan minimnya tingkat kesalahan klasifikasi (Koc, et.al, 2012). NBC adalah sebuah algoritma *data mining* yang cocok untuk menemukan data yang berguna dalam representasi yang berbeda. Hasil dari *data mining* dievaluasi, jika ada pengetahuan yang ditemukan tidak sesuai yang dibutuhkan maka diulangi kembali dengan langkah-langkah proses *data mining*. Klasifikasi dengan Naïve Bayes bekerja berdasarkan teori probabilitas yang memandang semua fitur dari data sebagai bukti dalam probabilitas (Prasetyo, 2012). Metode NBC adalah memanipulasi atribut saling bergantung, dan mewakili distribusi probabilitas gabungan lebih dari satu set diskrit, yaitu variabel stokastik (Ouali, et.al, 2006). NBC dapat digunakan untuk memprediksikan objek kelas yang labelnya tidak diketahui atau dapat memprediksikan data yang akan muncul di masa depan (Ginting & Trinanda, 2014).

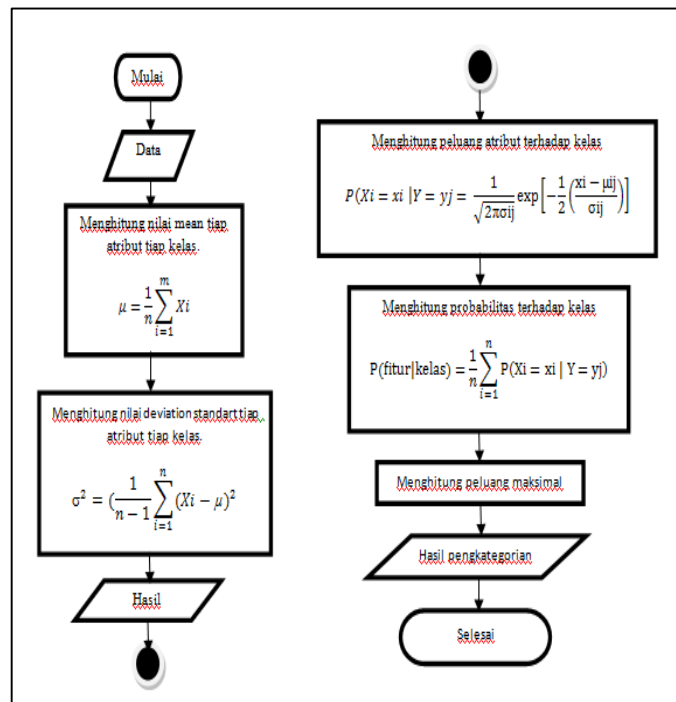
Keputusan perlu dilakukan oleh pengambil keputusan dengan preferensi tertentu berdasarkan informasi dan pengetahuan. Metode NBC dikembangkan dengan menggunakan rasio kemungkinan sebagai validasi metric model penilaian, dengan memprediksikan segala kemungkinan maka resiko dapat diminimalisir (Jiang & Sankaran, 2007). Kinerja NBC kompetitif dalam proses klasifikasi walaupun menggunakan asumsi atribut yang independen (tidak ada kaitan antar atribut).

Pada penelitian ini digunakan metode NBC untuk mengevaluasi prodi sehingga dapat membantu meningkatkan kinerja fakultas pada umumnya dan prodi. Hasil evaluasi tersebut dapat menjadi media untuk mempromosikan prodi, sehingga keminatan setiap

calon mahasiswa baru meningkat, dan menjadi dasar pimpinan dalam penentuan secara proporsional jumlah mahasiswa yang akan di terima tiap tahunnya.

Metode

Pada penelitian ini diimplementasikan dengan menggunakan perhitungan *Naïve Bayes Classifier*. Pada *naïve bayes classifier* terdapat dua proses penting yaitu *training* dan *testing*. *Training* digunakan untuk melakukan proses pembelajaran berdasarkan data yang telah memiliki keputusan atau *class label*. Sedangkan tahap atau fase *testing* adalah proses penentuan keputusan itu sendiri berdasarkan hasil training yang telah dilakukan. Alur proses NBC yaitu pada gambar 1.



Gambar 1. Alur proses NBC

Hasil dan Pembahasan

Hasil

Metode NBC digunakan untuk menghitung hingga mendapatkan probabilitas atau peluang yang paling besar nilainya sehingga bisa di dapatkan *class* dari data yang menjadi data *test* tersebut. Untuk nilai *class* yang paling mendekati angka 1 merupakan hasil dari status prodi pada tahun 2020, terdapat 43 data perprodi dan tahun 2020 tersebut digunakan sebagai data *testing*. Untuk hasil inputan data *test* dapat dilihat pada tabel 1.

$$P(Cukup) = \frac{\text{Likelihood Cukup}}{\text{Likelihood Cukup} + \text{likelihood Kurang}}$$

$$= \frac{0,000010}{0,000010 + 0,000005} = 0,6$$

$$P(Kurang) = \frac{\text{Likelihood Kurang}}{\text{Likelihood Cukup} + \text{likelihood Kurang}}$$

$$= \frac{0,000005}{0,000010 + 0,000005} = 0,3$$

Jadi inputan pada data test hasilnya = cukup

Tabel 1. Hasil *input data testing*

Prodi	Tahun	Riset	Dosen	Pegawai	Jmlh Mhs	Alumni	Kelas	Lab	Beasiswa	Status Prodi
Teknik Elektro	2020	5	S2	PNS	44	17	2	terawat	5	<u>Cukup</u>

Seperti yang telah digambarkan sebelumnya alur untuk penggunaan formula *naïve bayes classifier*, langkah pertama adalah menentukan *mean* dan varian untuk setiap atribut yang bersifat kontinyu, kemudian menghitung peluang tiap atribut, hingga mendapatkan hasil likelihood masing-masing *class* untuk mendapatkan status prodi. Penggunaan formula NBC pada sistem evaluasi internal prodi menghasilkan nilai-nilai yang dianalisis sehingga menghasilkan klasifikasi untuk status prodi.

Penentuan *mean* dan varian pada masing-masing atribut yang bernilai kontinyu terlihat pada tabel 2.

Tabel 2 Data, *mean* dan varian

Data	Mean	Varian
Penelitian Cukup	7	2.16
Mahasiswa Cukup	48.25	9.215
Alumni Cukup	17.25	8.77
Beasiswa Cukup	4	1.41
Penelitian Kurang	6.5	0.707
Mahasiswa Kurang	37	16.97
Alumni Kurang	17.5	4.949
Beasiswa Kurang	3.5	0.707

Probabilitas setiap kelas dengan diberikan masukan yang menjadi data test. Perhitungan untuk kelas cukup adalah seagai berikut:

1. Penelitian

$$\mu = \frac{10+7+6+5}{4} = 7$$

$$\sigma^2 = \frac{(10-7)^2 + (7-7)^2 + (6-7)^2 + (5-7)^2}{4-1} = 4,66$$

$$\sigma = \sqrt{4,66} \approx 2,16$$

2. Mahasiswa

$$\mu = \frac{47+61+39+46}{4} = 48,25$$

$$\sigma^2 = \frac{(47-48,25)^2 + (61-48,25)^2 + (39-48,25)^2 + (46-48,25)^2}{4-1} = 84,917$$

$$\sigma = \sqrt{84,917} \approx 9,215$$

3. Alumni

$$\mu = \frac{11+12+30+16}{4} = 17,25$$

$$\sigma^2 = \frac{(11-17,25)^2 + (12-17,25)^2 + (30-17,25)^2 + (16-17,25)^2}{4-1} = 76,9166$$

$$\sigma = \sqrt{76,9166} \approx 8,77$$

4. Beasiswa

$$\mu = \frac{6+4+3+3}{4} = 4$$

$$\sigma^2 = \frac{(6-4)^2 + (4-4)^2 + (3-4)^2 + (3-4)^2}{4-1} = 2$$

$$\sigma = \sqrt{2} \approx 1,41$$

Perhitungan untuk kelas kurang adalah sebagai berikut.

1. Penelitian

$$\begin{aligned}\mu &= \frac{7 + 6}{2} = 6,5 \\ \sigma^2 &= \frac{(7 - 6,5)^2 + (6 - 6,5)^2}{2 - 1} = 0,5 \\ \sigma &= \sqrt{0,5} \approx 0,707\end{aligned}$$

2. Mahasiswa

$$\begin{aligned}\mu &= \frac{25 + 49}{2} = 37 \\ \sigma^2 &= \frac{(25 - 37)^2 + (49 - 37)^2}{2 - 1} = 288 \\ \sigma &= \sqrt{288} \approx 16,97\end{aligned}$$

3. Alumni

$$\begin{aligned}\mu &= \frac{14 + 21}{2} = 17,5 \\ \sigma^2 &= \frac{(14 - 17,5)^2 + (21 - 17,5)^2}{2 - 1} = 24,5 \\ \sigma &= \sqrt{24,5} \approx 4,949\end{aligned}$$

4. Beasiswa

$$\begin{aligned}\mu &= \frac{4 + 3}{2} = 3,5 \\ \sigma^2 &= \frac{(4 - 3,5)^2 + (3 - 3,5)^2}{2 - 1} = 0,5 \\ \sigma &= \sqrt{0,5} \approx 0,707\end{aligned}$$

Setelah nilai masing-masing kelas diperoleh, maka kemudian dicari masing-masing *likelihood*-nya, dengan hasil sebagai berikut:

$$\begin{aligned}\text{Likelihood Cukup} &= 0,073 \times 1 \times 1 \times 0,016 \times 0,044 \times 1 \times 1 \times 0,199 &= 0,000010 \\ \text{Likelihood Kurang} &= 0,115 \times 1 \times 1 \times 0,005 \times 0,079 \times 0,115 &= 0,000005\end{aligned}$$

Tahapan selanjutnya adalah menghitung probabilitas dengan hasilnya yang mendapatkan nilai paling mendekati 1 adalah kelas cukup, dengan hasil sebagai berikut:

$$\begin{aligned}
 P(Cukup) &= \frac{\text{Likelihood Cukup}}{\text{Likelihood Cukup} + \text{likelihood Kurang}} \\
 &= \frac{0,000010}{0,000010 + 0,000005} = 0,6 \\
 P(Kurang) &= \frac{\text{Likelihood Kurang}}{\text{Likelihood Cukup} + \text{likelihood Kurang}} \\
 &= \frac{0,000005}{0,000010 + 0,000005} = 0,3
 \end{aligned}$$

Pada penelitian ini menghitung akurasinya menggunakan *confusion matrix* pada weka *classifier* yang menggunakan Teknik *Split Validation* adalah teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai data training dan sebagian lainnya sebagai data testing. Dengan menggunakan *Split Validation* akan dilakukan percobaan *training* berdasarkan *split ratio* yang telah ditentukan sebelumnya, untuk kemudian sisa dari *split ratio data training* akan dianggap sebagai data testing. Data *training* adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan data testing adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakurasian hasil pembelajaran (Witten, et.al, 2011). Dari hasil *confusion matrix* 25 *record* di prediksi tepat pada kelas a atau cukup sedangkan 1 *record* diprediksikan tidak tepat untuk kelas cukup, karena terdapat pada kelas b atau kelas kurang. Pada class b diprediksikan tepat yaitu 16 *record*, sedangkan 1 *record* di prediksikan tidak tepat karena diprediksikan ada pada kelas cukup. Dari hasil tersebut, dapat dihitung nilai :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{25 + 16}{25 + 16 + 1 + 1} \times 100\% = 95,3 \%$$

$$Precision = \frac{TP}{TP + FP} = \frac{25}{25 + 1} \times 100\% = 96,1 \%$$

$$Recall = \frac{TP}{TP + FN} = \frac{25}{25 + 1} \times 100\% = 96,1 \%$$

Dengan didapatkannya presentase mencapai 95,3 % untuk tingkat akurasi , 96,1% nilai *precision* dan 96,1% nilai *recall* dari perhitungan *confusion matrix*, maka data prodi yang menggunakan metode *naïve bayes classifier* dinyatakan akurat.

Pembahasan

Berdasarkan hasil klasifikasi pengolahan data hingga mendapatkan data yang akan diimplementasikan dengan menggunakan Naïve Bayes Classifier memerlukan

beberapa proses, yaitu menentukan data yang diperlukan, memeriksa data yang tidak konsisten, mengintegrasikan data, seleksi data yang relevan hingga tahap data mining. Pada penelitian ini data yang digunakan adalah dataset Fakultas Teknik Universitas Negeri Gorontalo dari tahun 2014 sampai 2019 dalam format Ms.Excel yang terdiri atas 8 tabel (data), yaitu :

Prodi	Tahun	Riset	Dosen	Pegawai	Jml Mhs	Alumni	Kelas	Lab.	Beasiswa	Status Prodi
Elektronika	2019	11	S2	PNS	47	1	2	Memadai	6	Cukup
Elektronika	2018	7	S2	PNS	61	12	2	Memadai	4	Cukup
Elektronika	2017	6	S2	PNS	39	30	2	Memadai	3	Cukup
Elektronika	2016	0	S2	PNS	46	6	2	Kurang Memadai	3	Kurang
Elektronika	2015	7	S2	PNS	25	14	1	Kurang Memadai	4	Kurang
Elektronika	2014	3	S2	Honoror	49	21	1	Kurang Memadai	3	Kurang
Sipil	2019	1	S3	PNS	91	15	2	Memadai	8	Cukup
Sipil	2018	6	S3	PNS	71	21	2	Memadai	5	Cukup
Sipil	2017	3	S2	PNS	57	24	2	Memadai	6	Cukup
Sipil	2016	1	S2	PNS	67	35	2	Memadai	2	Cukup
Sipil	2015	1	S2	PNS	42	40	1	Kurang Memadai	3	Kurang
Sipil	2014	1	S2	Honoror	64	32	1	Kurang Memadai	2	Kurang
Sistem Informasi	2019	0	S3	PNS	127	11	3	Memadai	7	Cukup
Sistem Informasi	2018	1	S3	PNS	111	19	3	Memadai	5	Cukup
Sistem Informasi	2017	4	S2	PNS	121	0	3	Memadai	5	Cukup
Sistem Informasi	2016	0	S2	PNS	109	0	2	Memadai	4	Cukup
Sistem Informasi	2015	0	S2	PNS	75	0	2	Kurang Memadai	1	Cukup
Sistem Informasi	2014	1	S2	Honoror	21	0	2	Kurang Memadai	0	Sangat Kurang
Industri	2019	8	S2	PNS	37	3	2	Memadai	4	Cukup
Industri	2018	5	S2	PNS	4	17	2	Memadai	3	Cukup
Industri	2017	9	S2	Honoror	1	24	1	Kurang Memadai	3	Kurang
Industri	2016	3	S2	PNS	10	29	1	Kurang Memadai	4	Kurang
Industri	2015	5	S2	PNS	4	8	1	Kurang Memadai	4	Kurang
Industri	2014	4	S2	PNS	2	2	1	Kurang Memadai	3	Kurang
Arsitektur	2019	5	S2	PNS	33	4	3	Memadai	7	Cukup
Arsitektur	2018	2	S2	PNS	42	18	2	Memadai	5	Cukup
Arsitektur	2017	1	S2	PNS	32	22	2	Memadai	4	Cukup
Arsitektur	2016	1	S2	PNS	27	20	1	Kurang Memadai	4	Kurang
Arsitektur	2015	0	S2	Honoror	9	15	1	Kurang Memadai	4	Sangat Kurang
Arsitektur	2014	4	S2	Honoror	8	17	1	Kurang Memadai	4	Kurang
Kriya	2019	8	S2	PNS	14	32	3	Memadai	5	Cukup
Kriya	2018	3	S2	PNS	18	14	2	Memadai	4	Kurang
Kriya	2017	2	S2	Honoror	12	27	2	Memadai	4	Kurang
Kriya	2016	2	S2	PNS	7	14	2	Kurang Memadai	3	Kurang
Kriya	2015	10	S2	PNS	14	20	2	Kurang Memadai	3	Kurang
Kriya	2014	1	S2	Honoror	1	2	2	Kurang Memadai	3	Kurang

Dari penelitian ini dihasilkan sebuah data test yang bisa melihat kondisi prodi pada Fakultas Teknik ditahun selanjutnya, dari kondisi data tersebut dekan fakultas secara umum dan ketua prodi secara khusus dapat mengevaluasi kekurangan dan kebutuhan prodi sehingga bisa meminimalisir kesalahan pada saat kegiatan akreditasi prodi.

Kesimpulan

Berdasarkan penelitian yang dilakukan, Penggunaan metode NBC untuk sistem evaluasi internal prodi dapat mengevaluasi setiap atribut yang berpengaruh pada setiap prodi, sehingga dapat diketahui keadaan atau status masing-masing prodi. Sistem evaluasi ini juga dapat menjadi bahan pertimbangan untuk pengambilan keputusan pimpinan dalam merekomendasikan kebutuhan untuk setiap prodi sehingga dapat menciptakan fakultas menjadi lebih baik lagi. Penggunaan NBC sangat efektif digunakan pada data prodi karena mempunyai nilai akurasi mencapai 95,3 %.

Daftar Pustaka

- BANPT. (2007). *Buku 1 naskah akademik akreditasi instistusi perguruan tinggi*. Jakarta: BAN-PT.
- Ginting, S.L.Br., & Trinanda, R.P. (2014). Teknik data mining menggunakan metode Bayes Classifier untuk optimalisasi, pencarian pada perpustakaan (Studi Kasus: Perpustakaan Universitas Pasundan-Bandung), 1(6), 1-14.
- Gorunescu, F. (2011). *Data mining concept, models and techniques*. Verlag Berlin Heidelberg: Springer.
- Jiang, X., & Sankaran, M. (2007). Bayesian risk-based decision method for model validation under uncertainty. *Journal of Reliability Engineering and System Safety*, 92, 707-718
- Koc, L., Mazzuchi, Thomas, A., & Sarkani, S. (2012). A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier. *Journal of Expert Systems with Applications*, 39, 13492-13500.
- Ouali, A., Cherif, A., Ramdane. K., & Marie-Odile. (2006). Data mining based Bayesian networks for best classification. *Journal of Computational Statistics & Data Analysis*, 51, 1278-1292.
- Prasetyo, E. (2012). *Data mining: Konsep dan aplikasi menggunakan Matlab*. Yogyakarta: Andi Offset.
- Vallejos, M., Alvarado, J., & Puente, A.I. (2012). *College performance prediction test*. *Journal of Procedia - Social and Behavioral Sciences* 31, 846-851.

Witten, I.H. , Frank, E., & Hall, M.A. (2011). *Data mining practical machine learning tools and technique*. Burlington: Morgan Kaufmann Publisher.