

SISTEM PENDUKUNG KEPUTUSAN UNTUK MENGEVALUASI INTERNAL PROGRAM STUDI

Indhitya R. Padiku^{a*}, Kusworo Adi^b, Achmad Widodo^c

^a Jurusan Sistem Informasi Universitas Negeri Gorontalo

^b Fakultas Sains dan Matematika Universitas Diponegoro

^c Jurusan Teknik Mesin Universitas Diponegoro

Abstract

The development of major and study program cannot be separated by some internal factors weather directly influenced number of new registration students or indirectly. It needs a method to both know and to analyze internal evaluation variables in major or study program.

Naive Bayes Classifier (NBC) method is the simple form of *Bayesian network* that assume all features are independent each other. NBC shows us a great performance entirely in accuracy and error level classification. NBC is able to differentiate irrelevance attribute and also classified some attributes in prediction needs.

This research hopefully can be useful for major internal evaluating and study program in order to increase the number of new registration students. The classification by influenced of variables to evaluate the condition of both major and study program for the new registration students.

Keyword : NBC, study program evaluation

1. Pendahuluan

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 program studi harus mampu meningkatkan kualitas secara berkelanjutan yang nantinya akan diimplementasikan kepada mahasiswa. Beberapa faktor yang mempengaruhi perkembangan tiap program studi 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, 2012).

Pada setiap program studi 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 program studi. Pimpinan fakultas dapat melakukan evaluasi awal pada setiap program studi, sehingga pimpinan dapat secara proporsional menentukan kuota jumlah mahasiswa pada penerimaan mahasiswa baru. Metode *naive 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, 2012).

Metode NBC adalah memanipulasi atribut saling bergantung, dan mewakili distribusi probabilitas gabungan lebih dari satu set diskrit, yaitu variabel stokastik (Ouali dkk., 2006). NBC dapat digunakan untuk memprediksikan objek kelas yang labelnya tidak diketahui atau dapat

memprediksikan data yang akan muncul di masa depan (Ginting, 2008).

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 dkk., 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 program studi sehingga dapat membantu untuk pengambilan keputusan sehingga dapat meningkatkan kinerja fakultas pada umumnya dan program studi. Hasil evaluasi tersebut dapat menjadi media untuk mempromosikan program studi, sehingga keminatan setiap calon mahasiswa baru meningkat, dan menjadi dasar pimpinan dalam penentuan secara proporsional jumlah mahasiswa yang akan di terima tiap tahunnya.

2. Kerangka Teori

2.1. *Naive Bayes Classifier* (NBC)

NBC adalah metode classifier yang berdasarkan probabilitas dari Teorema Bayes dengan asumsi bahwa setiap variabel X bersifat bebas (*independent*). Dengan kata lain, *Naive Bayesian Classifier* mengansumsikan bahwa keberadaan sebuah atribut (*variabel*) tidak ada kaitannya dengan beradaan atribut yang lain (Abidin., 2012).

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Keterangan :

X : data sampel dengan klas (label) yang tidak diketahui

H : hipotesa bahwa X adalah data dengan klas (label)

$P(H)$: peluang dari hipotesa H (*likelihood*)

*Alamat e-mail : indhitya@gmail.com

$P(X)$: peluang data sampel yang diamati (*evidence*)
 $P(X|H)$: peluang data sampel X , bila diasumsikan bahwa hipotesa H benar (*prior*)

Formulasi NBC untuk klasifikasi adalah (Prasetyo, 2012):

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \quad (2)$$

Keterangan :

- $P(Y|X)$: Peluang data dengan vector X pada kelas Y
- $P(Y)$: Peluang awal kelas Y
- $\prod_{i=1}^q P(X_i|Y)$: Peluang independen kelas Y dari semua fitur dalam vector X
- $P(X)$: Nilai konstan atau nilai tetap

Berdasarkan nilai $P(X)$ yang memang telah bernilai tetap maka nantinya hanya menghitung pada bagian $P(Y) \prod_{i=1}^q P(X_i|Y)$. Untuk peluang independen $\prod_{i=1}^q P(X_i|Y)$ adalah pengaruh semua atribut dari data terhadap kelas Y yang dinitasikan seperti dibawah ini :

$$P(X|Y = y) = \prod_{i=1}^q P(X_i|Y = y) \quad (3)$$

Atribut yang masuk dalam atribut kategorikal dapat langsung di hitung, tetapi untuk atribut yang kontinyu sebelum di implementasikan dengan formula *naive bayes classifier* harus melalui proses seperti berikut :

1. Untuk setiap atribut kontinyu diganti ke nilai interval diskret yaitu dengan mentransformasi atribut kontinyu ke dalam ordinal.
2. Mengasumsikan bentuk tertentu dari distribusi peluang untuk atribut kontinyu dan memperkirakan parameter distribusi dengan data pelatihan. Distribusi Gaussian biasanya dipilih untuk merepresentasikan peluang bersyarat dari atribut kontinyu pada sebuah kelas $P(X_i|Y)$, sedangkan distribusi Gaussian dikarakteristikan dengan dua parameter yaitu : mean (μ) dan varian (σ^2). Untuk setiap kelas y_j , peluang bersyarat kelas y_j , untuk atribut X_i adalah :

$$P(X_i = x | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp \left(-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right) \quad (4)$$

2.2. Data Mining

Data mining atau *Knowledge Discovery in Database* (KDD) adalah proses menentukan informasi yang berguna serta pola-pola yang ada dalam data. Informasi ini terkandung dalam basis data yang berukuran besar yang sebelumnya tidak diketahui dan potensial bermanfaat (Han & Kamber, 2006). Data Mining merupakan salah satu langkah dari serangkaian proses *iterative* KDD.

Fungsi-fungsi dalam data mining mengacu pada Larose (2005) yang membaginya ke dalam enam fungsi yaitu (Susanto & Suryadi, 2010) :

1. Fungsi Deskripsi (*description*)
2. Fungsi Estimasi (*estimation*)
3. Fungsi Prediksi (*prediction*)
4. Fungsi Klasifikasi (*classification*)
5. Fungsi Klasterisasi (*clustering*)
6. Fungsi Asosiasi (*association*)

Algoritma yang dapat digunakan diantaranya :

- a) *Generalizes Association Rules*
- b) *Quantitative Association Rules*
- c) *Assynchronous Parallel Mining (APM)*

2.3. Akurasi klasifikasi

Akurasi menunjukkan kedekatan nilai hasil pengukuran dengan nilai sebenarnya. Untuk menentukan tingkat akurasi perlu diketahui nilai sebenarnya dari parameter yang diukur dan kemudian dapat diketahui seberapa besar tingkat akurasinya. Akurasi pada sebuah klasifikasi berpengaruh pada performa dari suatu klasifikasi. Untuk menganalisa performa tersebut dapat menggunakan sebuah matrik dengan membandingkan data kelas yang asli dengan prediksi dari data yang di input atau disebut dengan *confusion matrix*. Untuk setiap kolom dari dari matriks berhubungan dengan *output* klasifikasi dan yang berhubungan dengan input adalah pada setiap baris (neila, 2012).

Confusion matrix terdapat pada weka *classifier* yang memberikan informasi dalam bentuk angka sehingga mempermudah untuk menghitung rasio keberhasilan dari suatu klasifikasi dalam menganalisis performa algoritma. *Confusion matrix* adalah matrik yang berbentuk 2x2 untuk mendapatkan hasil klasifikasi dengan tepat, tiap kelas yang diprediksi mempunyai kemungkinan keluaran yang berbeda-beda seperti diantaranya *true positives* (TP) yaitu menunjukkan ketepatan klasifikasi dan *false positive* (FP) adalah jika nilai aslinya dan nilai yang diprediksi berbeda. Tabel *confusion matrix* seperti berikut ini :

Tabel 1. *confusion matrix*

		Predicted Class	
		C1	C2
Actual Class	C1	True Positive	False negative
	C2	False positive	True negatives

Dalam *confusion matrix* beberapa hal yang dilakukan dengan menggunakan data hasil klasifikasi diantaranya adalah :

- a. menghitung persentase kelas positif (*true positive*) yang diperoleh dalam klasifikasi.
- b. *Precision* berfungsi menghitung persentase *false positive*
- c. Menghitung nilai rata-rata keberhasilan klasifikasi dengan cara membagi jumlah data yang terklasifikasi dengan benar dengan seluruh data yang diklasifikasi

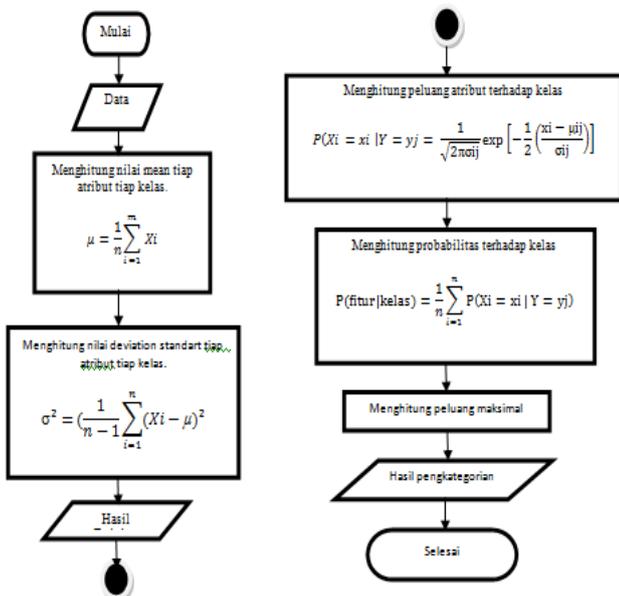
Dengan algoritma klasifikasi dapat dilakukan dengan menghitung jumlah dari *record* yang di prediksi secara benar (akurasi) untuk mengevaluasi performa sebuah model yang dibangun. Akurasi dapat didefinisikan sebagai berikut :

$$\text{Akurasi} = \frac{\text{jumlah prediksi benar}}{\text{jumlah total prediksi}} \quad (5)$$

3. Metodologi

Pada penelitian ini di implementasikan dengan menggunakan perhitungan *Naive Bayes Classifier*. Pada *naive 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. Berikut ini alur proses NBC yaitu pada Gambar 1 :



Gambar 1. alur proses NBC

4. Hasil Dan Pembahasan

4.1 Hasil Penelitian

Dalam penelitian ini dihasilkan sebuah aplikasi yang didalamnya terdiri dari *front end system* dan *back end system*, isi dari *front end* adalah info yang berupa grafik perkembangan setiap program studi dari tahun 2008-2013, serta login dan isi dari *back end system* adalah input data, data *training*, data *testing*, hasil pengolahan data serta menampilkan laporan pada setiap program studi yang hanya bisa di akses oleh admin fakultas dan dekan fakultas teknik. Dengan aplikasi tersebut dekan selaku pimpinan fakultas bisa mengevaluasi kekurangan dan kebutuhan setiap program studi pada fakultas teknik.

Penggunaan metode NBC ini 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 2014, yang tahun 2014 tersebut digunakan sebagai data *testing*. Untuk hasil inputan data *test* dapat di lihat pada Gambar 2 berikut :

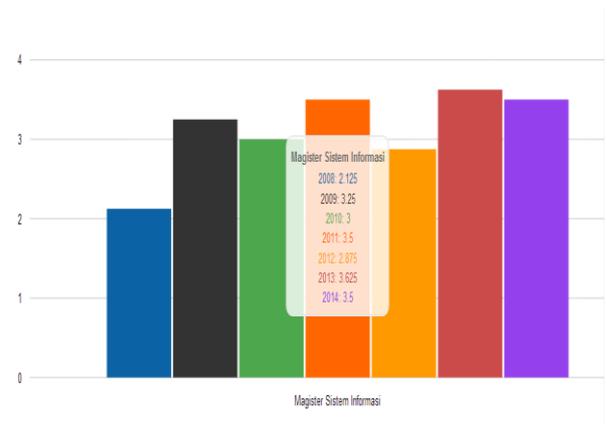
```
[Baik] => 1
[Cukup] => 0
)
Baik
```

Hasil Analisis Menggunakan Naive Bayes Classifier:

Tahun	publikasi	Dosen	Pegawai	Mahasiswa	Alumni	Kelas	Laboratorium	seminar	Status Prodi
2014	11	S3	PNIS	55	45	2	Terawat	8	Baik

Gambar 2. hasil input data *testing*

Hasil dari keseluruhan data yang telah di analisis di semua program studi beserta hasil status prodi yang didapatkan. Untuk hasil analisis dalam bentuk grafik terlihat pada Gambar 4.13 berikut :



Gambar 3. Tampilan grafik hasil analisis

Grafik pada gambar diatas adalah grafik yang menggambarkan perkembangan pada program studi Magister Sistem Informasi Universitas Diponegoro dari tahun 2008 sampai 2014. Sehingga dapat menjadi bahan pertimbangan pimpinan untuk pengambilan keputusan. Untuk melihat lebih detail lagi perkembangan untuk masing-masing atribut pada masing program studi

4.2 Pembahasan

Seperti yang telah di gambarkan pada bab sebelumnya alur untuk penggunaan formula *naive bayes classifier* adalah pertama 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.

Dari penggunaan formula NBC pada sistem evaluasi internal program studi menghasilkan nilai-nilai yang di analisis sehingga mengasilkan klasifikasi untuk status prodi. Berikut ini hasil-hasil dari setiap nilai yang di *input* pada masing-masing atribut. Sesuai dengan langkah-langkah untuk NBC :

- Menentukan mean dan varian pada masing-masing atribut yang bernilai kontinyu terlihat pada Gambar 4:

Data	Mean	Varian
Publikasi Baik	12,25	5,56
Maahasiswa Baik	47,75	10,37
Alumni Baik	26,5	12,767
Kelas Baik	1,75	0,5
Seminar Baik	12,5	5,196
Publikasi Cukup	1,5	2,121
Mahasiswa Cukup	40,5	40,30
Alumni Cukup	39	43,84
Kelas Cukup	1,5	0,707
Seminar Cukup	1	1,41

- Probabilitas setiap kelas dengan diberikan masukan yang menjadi data test.

a. Untuk kelas Baik terlihat pada Gambar 5 berikut :

```
[publikasi] => 0.0623595583185
[dosen] => 1
[pegawai] => 1
[mahasiswa] => 0.00305301763381
[alumni] => 4.71890227121E-8
[kelas] => 0.75
[laboratorium] => 1
[seminar] => 0.0109418479246
```

Gambar 5. Probabilitas kelas Baik

b. Untuk kelas Cukup terlihat pada Gambar 6 berikut :

```
[publikasi] => 1.08653989691E-10
[dosen] => 1
[pegawai] => 1
[mahasiswa] => 0.000729322702129
[alumni] => 0.00603714587471
[kelas] => 0.5
[laboratorium] => 1
[seminar] => 8.44767383233E-9
```

Gambar 6 Probabilitas kelas Cukup

3. Setelah di dapatkan nilai masing-masing kelas, maka kemudian mencari masing-masing likelihoodnya. Hasilnya terlihat pada Gambar 7:

```
[Baik] => 7.37267795957E-14
[Cukup] => 2.02071107453E-24
```

Gambar 7. Likelihood masing-masing kelas

4. Menghitung probabilitas dengan hasilnya yang mendapatkan nilai paling mendekati 1 adalah kelas cukup. Hasilnya terlihat pada Gambar 8 berikut :

```
[Baik] => 1
[Cukup] => 0
```

Gambar 8. hasil kelas yang paling mendekati angka 1

Pada penelitian ini menghitung akurasinya menggunakan *confusion matrix* pada *weka classifier*. Dari hasil *confusion matrix* 25 record di prediksi tepat pada class 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 di hitung nilai akurasinya :

$$\begin{aligned} \text{Presentase akurasi} &= \frac{\text{Banyaknya prediksi yang benar}}{\text{Total banyaknya prediksi}} \\ &= \frac{2 + 5}{2 + 1 + 1 + 5} \times 100\% = 100\% \end{aligned}$$

Dengan didapatkannya presentase mencapai 100% untuk tingkat akurasi dari perhitungan *confusion matrix*, maka data program studi yang menggunakan metode *naïve bayes classifier* dinyatakan akurat.

5. Kesimpulan

Berdasarkan penelitian yang dilakukan, Penggunaan metode NBC dalam sistem pendukung keputusan untuk mengevaluasi setiap atribut yang berpengaruh pada setiap program studi, sehingga dapat diketahui keadaan atau status masing-masing program studi. Sistem evaluasi ini juga dapat menjadi bahan pertimbangan untuk pengambilan keputusan pimpinan dalam merekomendasikan kebutuhan untuk program studi sehingga dapat menciptakan fakultas menjadi lebih baik lagi. Penggunaan NBC sangat efektif digunakan pada data program studi karena mempunyai nilai akurasi mencapai 100%.

Ucapan Terima Kasih

Terima kasih diucapkan pada Magister Sistem Informasi Universitas Diponegoro yang telah memberikan data dalam penelitian ini.

Daftar Pustaka

- Abidin Taufik Fuadi, Subianto Muhammad, 2012, Praktikum Data Mining Naive Bayesian Classifier.
- Berger,C, 1990. *Statistical Inference*. Pasific Grove; New York.
- Bolstad, W.M. 2007. *Introduction to Bayesian Statistics Second Edition*. A John Wiley & Sons. Inc; America.
- Chu Pao-Shin, Zhao Xin, 2011, Bayesian analysis for extreme climatic events: A review, Hawaii. *Journal of Atmospheric Research* 102. 243–262
- Han Jiawei, Kamber Micheline, 2006, *Data Mining Concepts and Techniques*.
- Hsu Chung-Chian, Huang Yan-Ping, Chang Keng-Wei, 2008, Extended Naive Bayes classifier for mixed data, Taiwan. *Journal of Information Sciences* 163. 103–122
- Jiang Xiaomo, Mahadevan Sankaran, 2007, Bayesian risk-based decision method for model validation under uncertainty. *Journal of Reliability Engineering and System Safety* 92. 707–718
- Koc Levent, Mazzuchi Thomas A, Sarkani Shahram, 2012, A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier, USA. *Journal of Expert Systems with Applications* 39. 13492–13500
- Li Dan, Yang Hai Zhen, Liang Xiao Feng, 2013, Prediction analysis of a wastewater treatment system using a Bayesian network, China. *Journal of Environmental Modelling & Software* 40. 140-150
- Li Gong, Shi Jing, 2012, Applications of Bayesian methods in wind energy conversion systems, USA. *Journal of Renewable Energy* 43. 1-8
- Marco A. de Oliveira, Osmar Possamai, Luiz V.O. Dalla Valentina, dan Carlos A. Flesch, 2012, Applying Bayesian networks to performance forecast of innovation projects: A case study of transformational leadership influence in

- organizations oriented by projects. *Journal of Expert Systems with Applications* 39. 5061–5070
- Mukherjee Saurabh, Sharma Neelam., 2012, Intrusion Detection using Naïve Bayes Classifier with feature Reduction, Spain. *Journal of Procedia Technology* 4. 119 – 128
- Neila Ramdhania, Djamaludin Ancok, Yuliardi Swasono, Peno Suryanto., 2012., Teacher Quality Improvement Program: Empowering teachers to increasing a quality of Indonesian's education, Indonesia. *Journal of Procedia - Social and Behavioral Sciences* 69 (2012) 1836 – 1841
- Ouali Abdelaziz, Cherif Amar Ramdane, Krebs Marie-Odile, 2006, Data mining based Bayesian networks for best classification. *Journal of Computational Statistics & Data Analysis* 51. 1278 – 1292
- Ragg Thomas, Menzel Wolfram, Baum Walter, Wigbers Michael, 2002, Bayesian learning for sales rate prediction for thousands of retailers. *Journal of Neurocomputing* 43. 127–144
- Shadiq M. Ammar, 2012, Keoptimalan Naïve Bayes Dalam Klasifikasi.
- Soria Daniele, Garibaldi Jonathan M., Ambrogi Federico, Biganzoli Elia M., Ellis Ian O., 2011, A 'non-parametric' version of the naive Bayes classifier. *Journal of Knowledge-Based Systems* 24. 775–784
- Vallejos Maria, Alvarado Jesus M, Puente Anibal., 2012, College performance prediction test, Spain. *Journal of Procedia - Social and Behavioral Sciences* 31. 846 – 851
- Wibisono, D., 2011, *Manajemen Kinerja Korporasi dan Organisasi Panduan Penyusunan Indikator*, Penerbit Erlangga, Jakarta.