

EVALUASI INTERNAL PROGRAM STUDI MENGUNAKAN METODE NAÏVE BAYES CLASSIFIER (NBC)

Yoseph Pius Kurniawan Kelen¹

¹Universitas Timor-Kefamenanu

¹e-mail: yosepkelen@gmail.com

ABSTRACT

The Development of department and study program can not be separated by its internal factors which is influenced number of candidate of students it self direct or indirectly. In this case we need a method to know the condition in each internal variables that will be evaluated in the department itself.

NBC method is kind of simple Bayesian Network that comes with assume statement that all features are similar one with another. Nbc shows the great work completely in accuracy and failed classification level. NBC could predict some assume for prediction needs.

This research generally made to use for evaluating department and study program internarly by clasifying some variables which can influence to know about the condition of the deprtment and the study program itself. This is also one 8f best strategy to developpe number of new cndidte of students and this method will shows about the study program to the new candidate of students.

Keywords : NBC,Naïve Bayes Clasification, Study Program Evaluating.

1. Pendahuluan

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). Metode NBC merupakan bentuk sederhana dari sebuah *Bayesian network* yang merupakan salah satu metode data mining. NBC mempunyai keunggulan akurasi dan minimnya tingkat kesalahan klasifikasi (Koc,2012). NBC mempunyai kemampuan inferensial dan keputusan yang baik sehingga memberikan fleksibilitas dalam penerapan untuk menganalisis faktor yang mempengaruhi sistem (Li, 2013).

Keputusan perlu dilakukan oleh pengambil keputusan dengan preferensi tertentu berdasarkan informasi dan pengetahuan. Kinerja NBC kompetitif dalam proses klasifikasi walaupun menggunakan asumsi atribut yang independen (tidak ada kaitan antar atribut). Metode NBC yang digunakan pada penelitian ini untuk mengevaluasi program studi sehingga dapat membantu meningkatkan kinerja fakultas khususnya Program Studi.

2. Kerangka Teori

a. Naïve Bayes Classifier (NBC)

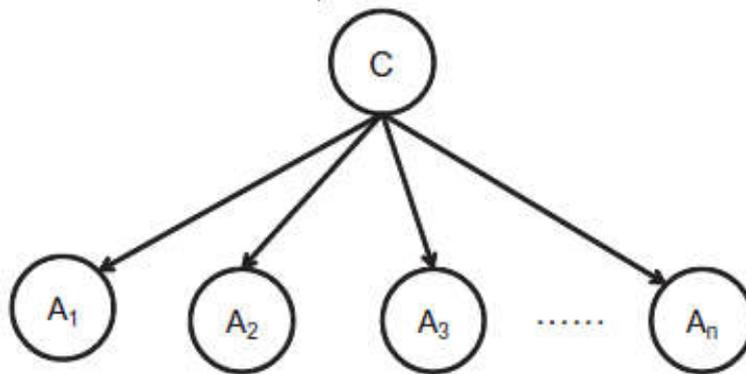
Naïve Bayesian Classifier mengansumsikan bahwa keberadaan sebuah atribut (*variabel*) tidak ada kaitannya dengan beradaan atribut yang lain (Abidin., 2012).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \dots\dots\dots(2.1)$$

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

Dalam NBC *H* adalah posterior dan *X* adalah prior. Prior adalah pengetahuan tentang karakteristik suatu parameter, sedangkan posterior adalah karakteristik yang diduga pada kejadian yang akan datang (Li, 2012). NBC menyediakan mekanisme merevisi atau memperbarui pengetahuan sebelumnya dengan data baru untuk menghasilkan laporan probabilitas posterior tentang parameter yang tidak diketahui atau berdasarkan hipotesis (Chu, 2011).



Gambar 2.1. Struktur *Naïve Bayes Classifier*

Sebuah NBC adalah *classifier* probabilistik sederhana berdasarkan pada penerapan teorema Bayes dengan asumsi bebas. Variabel acak yang melambangkan

kelas turunan dan X menjadi vektor variabel acak yang menunjukkan nilai-nilai atribut yang diamati (Soria, 2011).

Formulasi NBC untuk klasifikasi adalah :

$$P(Y|X) = \frac{P(Y)\prod_{i=1}^q P(X_i|Y)}{P(X)} \dots\dots\dots (2.2)$$

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 :

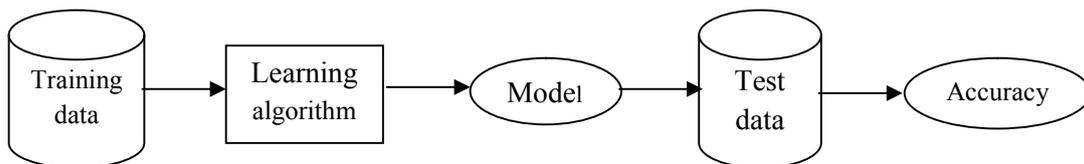
$$P(X|Y = y) = \prod_{i=1}^q P(X_i|Y = y) \dots\dots\dots (2.3)$$

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 :

$$P(X_i = x | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}\right) \dots\dots\dots (2.4)$$

b. Klasifikasi

Proses penting yang dilakukan saat melakukan klasifikasi yang pertama adalah *learning* yaitu proses pembelajaran menggunakan *training set*. dan yang kedua adalah proses *testing* yaitu menguji model menggunakan data testing.



Gambar 2.2. Tahapan proses klasifikasi (Abidin, 2012)

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 2.2. *confusion matrix*

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

Akurasi dapat didefinisikan sebagai berikut

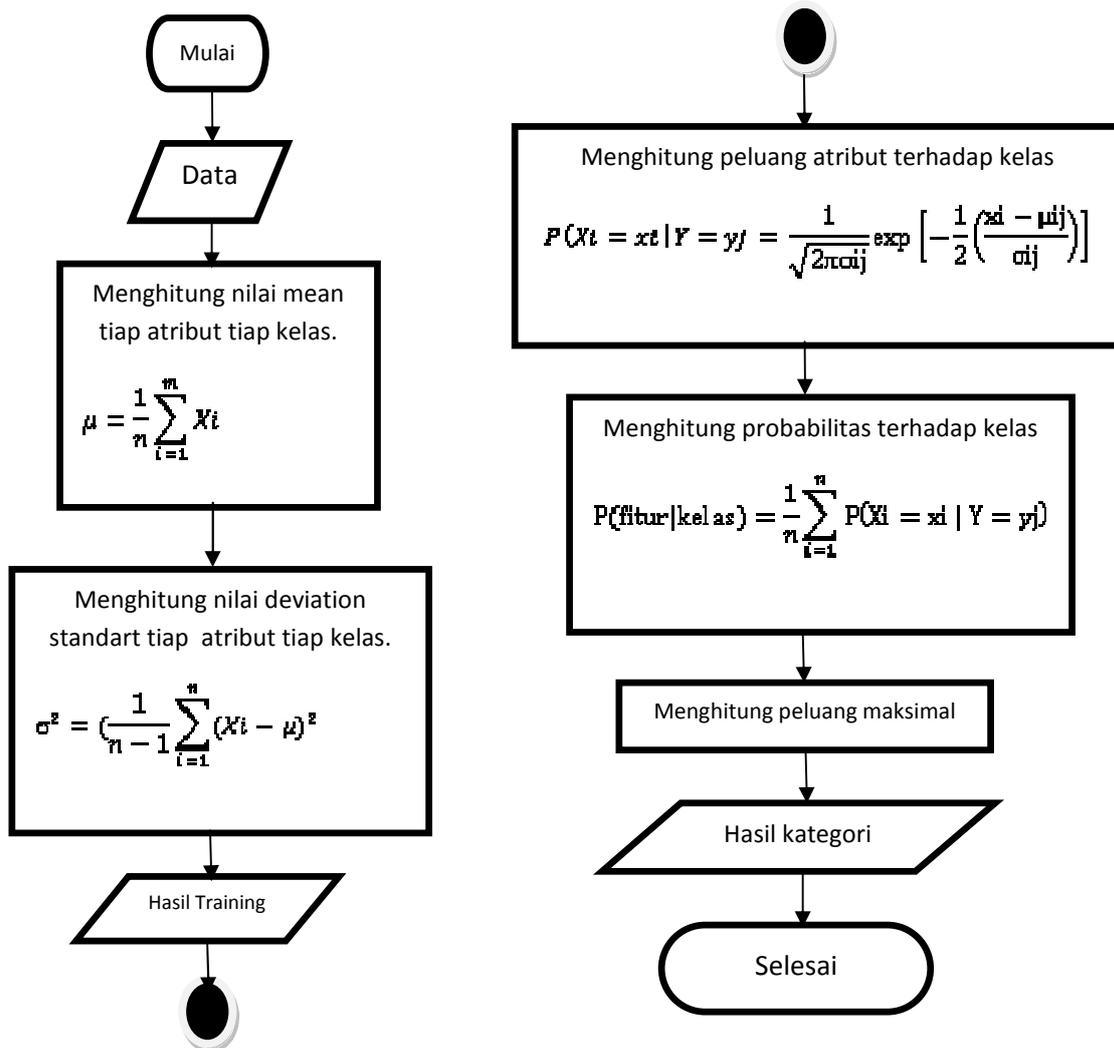
$$\text{Akurasi} = \frac{\text{jumlah prediksi benar}}{\text{jumlah total prediksi}} \dots\dots\dots (2.5)$$

3. Metodologi

Pengumpulan data; di tahap ini adalah mengumpulkan semua data historis 2011-2016 di setiap prodi yaitu data penelitian, data dosen, data pegawai, data mahasiswa, data alumni, data sarana prasarana (kelas), data sarana prasarana (laboratorium), dan data beasiswa. Data yang menjadi variabel utama yang berpengaruh dan ketentuan penilaiannya adalah :

- a. Publikasi : Sangat kurang = 0
Kurang = 1 – 3
Cukup = 4 – 7
Baik = 8 – 10
Sangat baik = 11-15
- b. Dosen : S2 = baik
S3 = Sangat Baik
- c. Pegawai: Honorer = Kurang
PNS = Baik
- d. Mahasiswa : Sangat Kurang = 0 – 5
Kurang = 6 – 10
Cukup = 11 – 20
Baik = 21 – 40
Sangat Baik = 41 – 70
- e. Alumni : Sangat Kurang = 0 – 5
Kurang = 6 – 10
Cukup = 11 – 20
Baik = 21 – 40
Sangat Baik = 41 – 70
- f. Kelas : Cukup = 1
Baik = 2
- g. Laboratorium : Cukup = tidak terawat
Baik = terawat
- h. Seminar : Sangat kurang = 0
Kurang = 1 – 5
Cukup = 6 – 10
Baik = 11 – 15
Sangat Baik = 16 - 20

Berikut ini alur proses NBC yaitu pada Gambar 3.1 :



Gambar 3.1 Alur proses NBC

4. Hasil dan Pembahasan

4.1.1. Pengujian Metode Naïve Bayes Classifier

Sesuai dengan data terdapat pada program studi Ilmu Komputer Universitas Nusa Cendana, untuk data *testing* sesuai input sebelumnya menggunakan tahun 2016. Untuk data *training* dari tahun 2011 sampai 2015 sesuai data pada sistem terdapat pada tabel berikut :

Tabel 4.1. Data *training* prodi Ilmu Komputer

Prodi	Tahun	Publikasi	Dosen	Pegawai	Mahasiswa	Alumni	Kelas	Lab.	Seminar	Status Prodi
Ilkom	2011	0	S3	PNS	12	8	1	Terawat	2	Cukup
Ilkom	2012	8	S3	PNS	36	21	1	Terawat	18	Baik
Ilkom	2013	7	S3	PNS	45	16	2	Terawat	6	Baik
Ilkom	2014	16	S3	PNS	61	24	2	Terawat	15	Baik
Ilkom	2015	3	S3	PNS	69	70	2	Terawat	0	Cukup
Ilkom	2016	18	S3	PNS	49	45	2	Terawat	11	Baik

Tabel 4.2 hasil data *testing*

Prodi	Tahun	Publikasi	Dosen	Pegawai	Mahasiswa	Alumni	Kelas	Lab	Seminar	Status Prodi
Ilkom	2016	11	S3	PNS	55	45	2	Terawat	8	Baik

5. Penutup

Pada penelitian evaluasi internal program studi menggunakan *naïve bayes classifier* ini dapat disimpulkan :

1. Penggunaan metode NBC untuk evaluasi internal program studi dapat mengevaluasi setiap atribut yang berpengaruh pada setiap program studi, sehingga dapat diketahui keadaan atau status masing-masing program studi.
2. Evaluasi internal program studi dengan menggunakan metode NBC ini dapat menjadi media untuk calon mahasiswa baru mengetahui keadaan program studi yang akan dipilih nantinya.
3. Evaluasi ini dapat menghasilkan suatu klasifikasi yang tepat untuk menjadi bahan pertimbangan untuk pengambilan keputusan pimpinan dalam merekomendasikan kebutuhan untuk setiap program studi sehingga dapat menciptakan fakultas menjadi lebih baik lagi.
4. Penggunaan NBC cukup efektif digunakan pada data program studi karena mempunyai nilai akurasi mendekati 100 %.

5.2. Saran

Saran yang dapat diberikan pada penelitian selanjutnya agar penelitian ini dapat diperbaiki atau dikembangkan adalah :

1. Pada penelitian ini menggunakan delapan data (atribut) yang dianggap paling berpengaruh, pada penelitian selanjutnya dapat meningkatkan jumlah pemakaian data.
2. Penelitian ini menggunakan metode *naïve bayes classifier*, untuk penelitian yang berhubungan dengan klasifikasi dapat menggunakan metode lainnya dan dapat dibandingkan nilai akurasinya dengan penggunaan NBC.

DAFTAR PUSTAKA

- Chu P., S., dan Zhao X., 2011, Bayesian analysis for extreme climatic events: A review, Hawaii. *Journal of Atmospheric Research* 102. 243–262
- Koc L., Mazzuchi T., dan Sarkani S., 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 D., Yang H., Z., dan Liang X., F., 2013, Prediction analysis of a wastewater treatment system using a Bayesian network, China. *Journal of Environmental Modelling & Software* 40. 140-150
- Li G., dan Shi J., 2012, Applications of Bayesian methods in wind energy conversion systems, USA. *Journal of Renewable Energy* 43. 1-8
- Soria D., Garibaldi Jonathan M., Ambrogi F., 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 M., Alvarado Jesus M., dan Puente A., 2012, College performance prediction test, Spain. *Journal of Procedia - Social and Behavioral Sciences* 31. 846 – 851.