

Implementasi Algoritma *Naïve Bayes* Dalam Memprediksi Kenaikan Golongan Karyawan

Friska Pratiwi¹, Dedy Hartama², Irfan Sudahri Damanik³, Eka Irawan⁴, Ilham
Syahputra Saragih⁵

Sistem Infomasi, STIKOM Tunas Bangsa Pematangsiantar

¹Friskapратиwi123@gmail.com, ²dedyhartama@amiktunasbangsa.ac.id,

³irfansudahri@gmail.com, ⁴eka.irawan@amiktunasbangsa.ac.id,

⁵ilhamsyahputrasaragih@gmail.com

Abstrak

Kenaikan golongan merupakan suatu penghargaan yang diberikan oleh perusahaan atas prestasi kerja. Semakin berkembangnya suatu perusahaan semakin dibutuhkan tenaga kerja yang profesional, namun frekuensi tatap muka antara pihak manager dengan karyawan yang sangat minim sehingga menyulitkan dalam kegiatan penilaian kinerja karyawan. Data mining merupakan sebuah analisa dari observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui sebelumnya, data yang diolah dengan metode data mining akan menghasilkan suatu pengetahuan baru yang bersumber dari data lama, hasil dari pengolahan dapat digunakan untuk menentukan keputusan di masa depan. Dengan menggunakan algoritma *naïve bayes* akan memprediksi peluang dalam menentukan kelayakan kenaikan golongan pada karyawan PT. Perkebunan Nusantara IV. Sumber data penelitian ini dikumpulkan berdasarkan kuesioner (angket) yang di isi oleh karyawan pimpinan. Data akan diolah dengan menghitung kelas/label, menghitung jumlah kasus dengan kelas yang sama, menghitung jumlah data probabilitas lalu membandingkan hasil perkelas. Sehingga diperoleh hasil akhir yang diharapkan menjadi masukan kepada PT. Perkebunan Nusantara IV Unit Kebun Bah Jambi dalam menentukan kelayakan kenaikan golongan karyawan.

Kata Kunci : Kenaikan Golongan, Data Mining, *Naïve Bayes*

1. Pendahuluan

Sebuah perusahaan membutuhkan karyawan sebagai aset terpenting terhadap kesuksesan sebuah perusahaan. Kegiatan penilaian kinerja karyawan sangat sulit dilakukan, karena frekuensi tatap muka antara pihak manager dengan karyawan yang sangat minim juga dengan cara penilaian yang masih dilakukan hanya melihat kepribadian karyawan tanpa mengetahui aspek-aspek penting yang membuat karyawan layak dikatakan naik golongan atau tidak layak naik golongan. Oleh karena itu, dalam perusahaan ini dibutuhkan sistem yang dapat membantu dalam memprediksi kenaikan golongan karyawan, Didalam ilmu komputer banyak cabang ilmu dan metode yang terdapat didalamnya seperti Sistem Pendukung Keputusan (SPK), Sistem Pakar, *Data Mining* dan lain-lain. *Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan [1]. Data yang digunakan dalam penelitian ini berdasarkan kuesioner yang diisi oleh karyawan pimpinan PT. Perkebunan Nusantara IV Unit Kebun Bah Jambi. Dalam hal ini peneliti mengangkat topik kenaikan golongan karyawan dimana proses metode yang dipilih adalah prediksi menggunakan *Naïve Bayes*. Hasil dari penelitian ini diharapkan dapat menjadi masukan kepada pihak perkebunan dalam menentukan kelayakan kenaikan golongan karyawan kedepannya.

2. Tinjauan Pustaka

2.1. Data Mining

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data[2]. *Data mining* merupakan penyelesaian masalah dengan menganalisa data yang ada pada database, dimana data tersimpan secara elektronik dan pencariannya dilakukan otomatis seperti pada komputer [3].

2.2. Naïve Bayes

Naïve Bayes yang dikemukakan oleh ilmuwan inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes*[4]. Keuntungan penggunaan *Naïve Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan *estimasi* parameter yang diperlukan dalam proses pengklasifikasian[5].

2.3. Prediksi

Prediksi atau dalam hal ini juga disebut peramalan (*forecast*) pada dasarnya merupakan dugaan atau peramalan mengenai peristiwa di waktu yang akan datang berdasarkan informasi masa lalu dan sekarang yang dimiliki agar kesalahannya dapat diperkecil[6].

2.4. Kenaikan Golongan

Kenaikan golongan merupakan suatu penghargaan yang diberikan pihak perusahaan atas prestasi kerja yang didapatkan oleh karyawan. Besar tidaknya tingkat pendapatan bagi seorang karyawan ditetapkan pada tingkat golongan yang telah ditentukan untuk setiap karyawan. Pentingnya suatu golongan bagi karyawan membuat seorang pengambil keputusan harus teliti dalam menentukan kenaikan golongan.

2.5. Teknik Pengumpulan Data

Pengumpulan data dapat dilakukan dalam berbagai aturan, berbagai sumber dan berbagai cara. Bila dilihat dari sumber datanya, maka pengumpulan data dapat menggunakan sumber primer dan sekunder. Sumber primer adalah sumber data yang langsung memberikan data kepada pengumpul data, dan data sekunder merupakan sumber tidak langsung memberikan data pada pengumpul data, misalnya lewat orang lain atau lewat dokumen. Selanjutnya jika dilihat dari segi cara atau teknik pengumpulan data, maka teknik pengumpulan data dapat dilakukan dengan interview, kuesioner (angket), observasi. Untuk memperoleh data dalam penelitian ini, penulis menyebarkan kuesioner kepada karyawan pimpinan PT. Perkebunan Nusantara IV Unit Kebun Bah Jambi sebanyak 200 sampel karyawan.

2.6. Teknik Pengolahan Data

Data yang digunakan pada penelitian ini adalah data primer, dimana data tersebut diperoleh dari hasil kuesioner yang disebarkan kepada karyawan pimpinan PT. Perkebunan Nusantara IV Unit Kebun Bah Jambi. Sampel yang digunakan adalah sebanyak 200 karyawan. Setelah data yang diperlukan diperoleh, kemudian ditentukan variabel-variabel yang akan digunakan dalam penelitian ini. Variabel independen yang digunakan antara lain Usia, Pendidikan Terakhir, Lama Bekerja, Kinerja Karyawan, Sikap, Utusan Karyawan Pimpinan, Keterampilan Komunikasi, dan Motivasi Kerja. Sedangkan variabel responnya adalah kelayakan, yaitu

karyawan yang layak naik dan tidak layak naik. Kemudian data akan diuji menggunakan *software RapidMiner* untuk membuktikan hasil perhitungan manual adalah sama dengan system.

3. Hasil dan Pembahasan

3.1. Proses Perhitungan

Proses perhitungan untuk memprediksi kenaikan golongan karyawan PT. Perkebunan Nusantara IV Unit Kebun Bah Jambi menggunakan algoritma *naïve bayes* adalah sebagai berikut :

1. Menghitung Jumlah *Class/Label*

Jumlah data yang digunakan sebanyak 100 data, data naik sebanyak 80 data dan tidak naik sebanyak 20 data. Sehingga untuk menghitung nilai probabilitasnya adalah sebagai berikut :

$$P(\text{Naik}) = 80/100$$

$$P(\text{Naik}) = 0,8$$

$$P(\text{Tidak Naik}) = 20/100$$

$$P(\text{Tidak Naik}) = 0,2$$

Tabel 1. Data Training

ALT	A	B	C	D	E	F	G	H	KET.
A1	41>55	SLTA	>15	SS	S	S	KS	S	NAIK
A2	41>55	SLTA	>15	S	S	S	S	S	NAIK
A3	41>55	S1	>15	SS	S	SS	S	SS	NAIK
A4	41>55	SLTA	>15	S	S	S	S	S	NAIK
A5	41>55	SLTA	>15	S	SS	CS	SS	S	NAIK
A6	41>55	SLTA	>15	S	S	S	S	S	NAIK
A7	41>55	S1	<15	S	SS	S	SS	S	NAIK
A8	41>55	SLTA	>15	SS	SS	S	SS	S	NAIK
A9	25>40	SLTA	<15	S	CS	KS	CS	KS	TIDAK NAIK
A10	25>40	S1	<15	CS	S	CS	CS	KS	TIDAK NAIK
...
A100	41>55	SLTP	>15	TS	CS	S	KS	KS	TIDAK NAIK

Keterangan :

A : Usia

B : Pendidikan Terakhir

C : Lama Bekerja

D : Kinerja Karyawan

E : Sikap

F : Utusan Karyawan Pimpinan

G : Keterampilan Komunikasi

H : Motivasi Kerja

2. Menghitung Jumlah Kasus Yang Sama Dengan *Class* Yang Sama

Setelah mencari nilai probabilitas setiap kriteria, yang didapatkan dari data *training* maka nilai probabilitas untuk masing-masing kriteria dapat digunakan seperti tabel 2.

Tabel 2. Tabel Nilai Probabilitas Usia

Usia	Tanggapan		Probabilitas	
	Naik	Tidak Naik	Naik	Tidak Naik
41>55	69	16	1	0,80

25>40	11	4	0,14	0,20
Jumlah	80	20	1	1

Tabel 3. Tabel Nilai Probabilitas Pendidikan Terakhir

Pendidikan Terakhir	Tanggapan		Probabilitas	
	Naik	Tidak Naik	Naik	Tidak Naik
SD	12	1	0,15	0,05
SLTP	5	4	0,06	0,20
SLTA	54	14	0,68	0,70
Diploma	4	0	0,05	0,00
S1	5	1	0,06	0,05
Jumlah	80	20	1	1

Tabel 4. Tabel Nilai Probabilitas Lama Bekerja

Lama Bekerja	Tanggapan		Probabilitas	
	Naik	Tidak Naik	Naik	Tidak Naik
>15	12	5	0,15	0,25
<15	68	15	0,85	0,75
Jumlah	80	20	1	1

Tabel 5. Tabel Nilai Probabilitas Kinerja Karyawan

Kinerja Karyawan	Tanggapan		Probabilitas	
	Naik	Tidak Naik	Naik	Tidak Naik
Sangat Setuju	29	0	0,36	0,00
Setuju	43	3	0,54	0,15
Cukup Setuju	1	4	0,01	0,20
Kurang Setuju	7	0	0,09	0,00
Tidak Setuju	0	13	0,00	0,65
Jumlah	80	20	1	1

Tabel 6. Tabel Nilai Probabilitas Sikap Karyawan

Sikap	Tanggapan		Probabilitas	
	Naik	Tidak Naik	Naik	Tidak Naik
Sangat Setuju	14	0	0,18	0,00
Setuju	54	2	0,68	0,10
Cukup Setuju	4	15	0,05	0,75
Kurang Setuju	8	2	0,10	0,10
Tidak Setuju	0	1	0,00	0,05
Jumlah	80	20	1	1

Tabel 7. Tabel Nilai Probabilitas Utusan Karyawan Pimpinan

Utusan Karpim	Tanggapan		Probabilitas	
	Naik	Tidak Naik	Naik	Tidak Naik

Sangat Setuju	11	0	0,14	0,00
Setuju	67	12	0,84	0,60
Cukup Setuju	0	4	0,00	0,20
Kurang Setuju	2	4	0,03	0,20
Tidak Setuju	0	0	0,00	0,00
Jumlah	80	20	1	1

Tabel 8. Tabel Nilai Probabilitas Keterampilan Komunikasi

Komunikasi	Tanggapan		Probabilitas	
	Naik	Tidak Naik	Naik	Tidak Naik
Sangat Setuju	15	0	0,19	0,00
Setuju	45	4	0,56	0,20
Cukup Setuju	9	7	0,11	0,35
Kurang Setuju	11	9	0,14	0,45
Tidak Setuju	0	0	0,00	0,00
Jumlah	80	20	1	1

Tabel 9. Tabel Nilai Probabilitas Motivasi Kerja

Motivasi	Tanggapan		Probabilitas	
	Naik	Tidak Naik	Naik	Tidak Naik
Sangat Setuju	5	0	0,06	0,00
Setuju	75	0	0,94	0,00
Cukup Setuju	0	2	0,00	0,10
Kurang Setuju	0	18	0,00	0,90
Tidak Setuju	0	0	0,00	0,00
Jumlah	80	20	1	1

3. Kalikan Semua Hasil Variabel Naik Dan Tidak Naik
Langkah selanjutnya adalah mengalikan semua hasil variabel naik dan tidak naik dengan menggunakan data *testing*.

Tabel 10. Data Testing

ALT	A	B	C	D	E	F	G	H	KETERANGAN
A101	25>40	SLTA	<15	CS	S	S	CS	S	?
A102	41>55	SLTA	>15	SS	KS	S	S	S	?
A103	25>40	SLTP	<15	TS	KS	S	CS	KS	?
A104	25>40	SLTA	<15	SS	KS	S	S	S	?
A105	41>55	SLTA	>15	S	CS	S	S	S	?
A106	41>55	SLTP	<15	TS	KS	S	CS	KS	?
A107	41>55	SD	>15	SS	KS	S	S	S	?
A108	41>55	SLTA	>15	SS	KS	S	S	S	?
A109	25>40	SLTA	<15	S	CS	S	S	S	?
A110	41>55	SLTA	>15	SS	KS	S	S	S	?
...
A200	41>55	SLTA	>15	S	SS	S	SS	S	?

Keterangan :

- A : Usia
- B : Pendidikan Terakhir
- C : Lama Bekerja
- D : Kinerja Karyawan
- E : Sikap
- F : Utusan Karyawan Pimpinan
- G : Keterampilan Komunikasi
- H : Motivasi Kerja

Hasil belum diketahui klasifikasinya karena variabel tersebut merupakan hasil prediksi dari data yang dihitung dari setiap variabel. Berikut adalah proses perhitungannya :

$$P(25\backslash\text{Naik}) \times P(\text{SLTA}\backslash\text{Naik}) \times P(<15\backslash\text{Naik}) \times P(\text{CS}\backslash\text{Naik}) \times P(\text{S}\backslash\text{Naik}) \times P(\text{S}\backslash\text{Naik}) \times P(\text{CS}\backslash\text{Naik}) \times P(\text{S}\backslash\text{Naik}) \times P(\text{Naik})$$

$$= \frac{11}{80} \times \frac{54}{80} \times \frac{68}{80} \times \frac{1}{80} \times \frac{54}{80} \times \frac{67}{80} \times \frac{9}{80} \times \frac{75}{80} \times \frac{80}{100}$$

$$= 0,1375 \times 0,6750 \times 0,8500 \times 0,0125 \times 0,6750 \times 0,8375 \times 0,1125 \times 0,9375 \times 0,8000$$

$$= 0,0000470368017$$

$$P(25\backslash\text{Tidak Naik}) \times P(\text{SLTA}\backslash\text{Tidak Naik}) \times P(<15\backslash\text{Tidak Naik}) \times P(\text{CS}\backslash\text{Tidak Naik}) \times P(\text{S}\backslash\text{Tidak Naik}) \times P(\text{S}\backslash\text{Tidak Naik}) \times P(\text{CS}\backslash\text{Tidak Naik}) \times P(\text{S}\backslash\text{Tidak Naik}) \times P(\text{Tidak Naik})$$

$$= \frac{4}{20} \times \frac{14}{20} \times \frac{15}{20} \times \frac{4}{20} \times \frac{2}{20} \times \frac{12}{20} \times \frac{7}{20} \times \frac{0}{20} \times \frac{20}{100}$$

$$= 0,4500 \times 0,4500 \times 0,6000 \times 0,1500 \times 0,5000 \times 0,7500 \times 0,4000 \times 0,1000 \times 0,2000$$

$$= 0,00000000000000$$

4. Bandingkan Hasil *Class* Naik Dan Tidak Naik

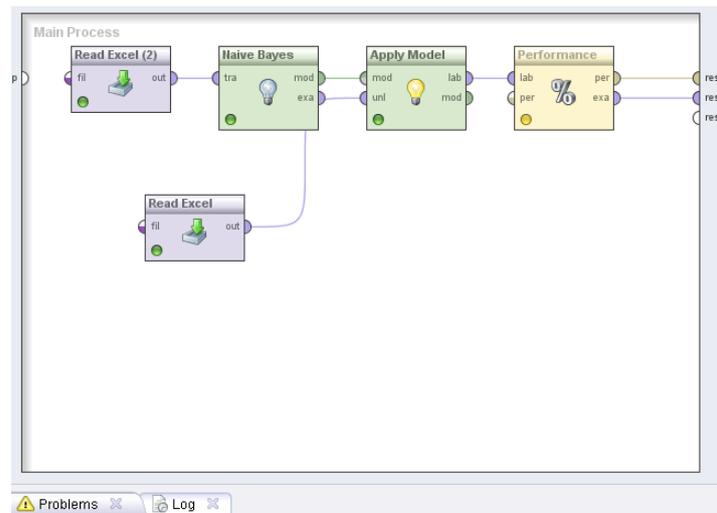
Dari hasil perhitungan tersebut, bandingkan kilai probabilitas antara kelas (P|Naik) dan kelas (P|Tidak Naik) nilai probabilitas tertinggi ada pada kelas (P|Naik) sehingga dapat disimpulkan bahwa tanggapan karyawan pimpinan tersebut masuk dalam klasifikasi “Naik”

Row No.	Alt	Keterangan	confidence	confidence	prediction(K...	Usia	Pendidikan	Lama Beke...	Kinerja Kary...	Sikap	Utusan Kar...	Ket. Komi
1	A101	?	0.998	0.002	Naik	25>40	SLTA	<15	CS	S	S	CS
2	A102	?	1.000	0.000	Naik	41>55	SLTA	>15	SS	KS	S	S
3	A103	?	0	1	Tidak Naik	25>40	SLTP	<15	TS	KS	S	CS
4	A104	?	1.000	0.000	Naik	25>40	SLTA	<15	SS	KS	S	S
5	A105	?	1.000	0.000	Naik	41>55	SLTA	>15	S	CS	S	S
6	A106	?	0	1	Tidak Naik	41>55	SLTP	<15	TS	KS	S	CS
7	A107	?	1.000	0.000	Naik	41>55	SD	>15	SS	KS	S	S
8	A108	?	1.000	0.000	Naik	41>55	SLTA	>15	SS	KS	S	S
9	A109	?	1.000	0.000	Naik	25>40	SLTA	<15	S	CS	S	S
10	A110	?	1.000	0.000	Naik	41>55	SLTA	>15	SS	KS	S	S
11	A111	?	1.000	0.000	Naik	41>55	SLTA	>15	SS	KS	S	S
12	A112	?	1.000	0.000	Naik	41>55	SD	>15	SS	S	S	KS
13	A113	?	1.000	0.000	Naik	41>55	SD	>15	SS	S	S	KS
14	A114	?	0	1	Tidak Naik	41>55	SLTA	>15	TS	CS	S	KS
15	A115	?	1.000	0.000	Naik	25>40	SD	<15	SS	S	S	KS
16	A116	?	1.000	0.000	Naik	41>55	SLTP	>15	S	S	S	CS
17	A117	?	0	1	Tidak Naik	25>40	SLTP	<15	TS	CS	S	KS
18	A118	?	1.000	0.000	Naik	41>55	SD	>15	SS	S	S	KS
19	A119	?	1.000	0.000	Naik	25>40	SLTP	<15	S	S	S	S
20	A120	?	1.000	0.000	Naik	41>55	SD	>15	S	S	S	S
21	A121	?	1.000	0.000	Naik	41>55	SD	>15	S	S	S	S

Gambar 1. Pembuktian Menggunakan *Tools RapidMiner*

Pada gambar 1 dapat dilihat bahwa hasil dari perhitungan manual adalah sama dengan hasil perhitungan menggunakan tools rapidminer. Dalam melakukan validasi data terdapat hal-hal yang harus diperhatikan diantaranya : perhitungan manual algoritma harus telah menampilkan hasil akhir berupa pohon keputusan, serta data yang digunakan haruslah data yang valid dan sama dengan yang dipakai pada tools.

Hasil pengujian model *Algoritma Naïve Bayes Classifier* ditunjukkan pada gambar 2 berikut :



Gambar 2. Desain Klasifikasi *Training* dan *Testing*

Penjelasan gambar 2 :

1. Operator *apply model* digunakan untuk memprediksi dan menerapkan model dari data *training*
2. *Performance* digunakan untuk mengetes hasil dari data *training*, *performance* memberikan nilai *accuracy*

Running dari desain adalah :

1. Accuracy

Akurasi adalah ukuran dari seberapa baik model mengkorelasikan antara hasil dengan atribut dalam data yang telah disediakan.

accuracy: 100.00%			
	true Naik	true Tidak Naik	class precision
pred. Naik	80	0	100.00%
pred. Tidak Naik	0	20	100.00%
class recall	100.00%	100.00%	

Gambar 3. Nilai *Accuracy Performance*

Table / Plot View
 Text View
 Annotations

PerformanceVector

```

PerformanceVector:
accuracy: 100.00%
ConfusionMatrix:
True:   Naik   Tidak Naik
Naik :  80     0
Tidak Naik:  0     20
    
```

Gambar 4. Detail *Performance Vector*

Berdasarkan gambar yang telah dijelaskan di atas, maka dapat diketahui bahwa pengujian data yang dilakukan dengan menggunakan apply model dan %Performance diperoleh nilai akurasi 100% dapat dikategorikan model algoritma *Naïve Bayes* ini adalah *excellent*.

4. Kesimpulan

Dapat disimpulkan bahwa Penerapan *Datamining* dengan menggunakan Algoritma *Naïve Bayes* pada kelayakan kenaikan golongan karyawan PT. Perkebunan Nusantara IV Unit Kebun Bah Jambi dapat diterapkan. Sumber data yang digunakan pada penelitian ini adalah data yang diperoleh langsung dari karyawan pimpinan. Jumlah data yang data uji sebanyak 100 karyawan dengan menggunakan dua kelas. Dari hasil perhitungan Algoritma *Naïve Bayes* diperoleh klasifikasi dengan kelas Naik sebanyak 80 karyawan dan kelas Tidak Naik sebanyak 20 karyawan.

5. Ucapan Terima Kasih

Terima kasih kepada program studi sistem informasi, STIKOM Tunas Bangsa Pematangsiantar atas dukungannya dalam terlaksananya penelitian ini serta bapak pembimbing Dr. Dedy Hartama, S.T, M.Kom dan Bapak Irfan Sudahri Damanik, M.Kom.

Daftar Pustaka

- [1] I. Technology and C. Science, "No Title," vol. 1, no. 1, pp. 72–77, 2018.
- [2] D. Larose, *An Introduction to Data Mining*. 2016.
- [3] R. R. Putra and C. Wadisman, "IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA K-MEANS IMPLEMENTATION," vol. 1, no. 1, pp. 72–77, 2018.
- [4] M. Syarli, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus : Data Mahasiswa Baru Perguruan Tinggi)," vol. 2, no. 1, pp. 22–26, 2016.
- [5] A. Saleh, "PENERAPAN DATA MINING DENGAN METODE KLASIFIKASI NAÏVE BAYES UNTUK MEMPREDIKSI KELULUSAN MAHASISWA DALAM MENGIKUTI ENGLISH PROFICIENCY TEST (Studi Kasus : Universitas Potensi Utama) PENERAPAN DATA MINING DENGAN METODE KLASIFIKASI NAÏVE MENGIKUTI ENGLISH ," no. February 2015, 2016.
- [6] E. Manalu, F. A. Sianturi, and M. R. Manalu, "PENERAPAN ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI JUMLAH PRODUKSI BARANG BERDASARKAN DATA PERSEDIAAN DAN JUMLAH PEMESANAN PADA CV . PAPADAN MAMA PASTRIES," vol. 1, no. 2, 2017.