

LAMPIRAN

CONTOH PERHITUNGAN NAÏVE BAYES

1. Menghitung Probabilitas prior (awal) setiap kelas

Tahapan klasifikasi metode *naive* Bayes yang pertama yaitu menghitung probabilitas *prior* pada kedua kelompok menggunakan data *training*. Adapun nilai probabilitas *prior* setiap kelompok sebagai berikut:

a. Kelas Pertama Alumni dengan lama mendapat pekerjaan < 12 bulan.

Pada 320 data training yang digunakan. Terdapat 151 data dengan lama mendapat pekerjaan < 12 bulan , sehingga nilai probabilitas prior kelas pertama adalah sebagai berikut.

$$\begin{aligned} P(Y_1) &= \frac{151}{320} \\ &= 0,47 \end{aligned}$$

b. Kelas ke dua Alumni dengan lama mendapat pekerjaan >= 12 bulan.

Pada 320 data training yang digunakan. Terdapat 169 data dengan lama mendapat pekerjaan >= 12 bulan , sehingga nilai probabilitas prior kelas pertama adalah sebagai berikut.

$$\begin{aligned} P(Y_1) &= \frac{169}{320} \\ &= 0,53 \end{aligned}$$

2. Menghitung setiap probabilitas variabel bebas pada setiap kelas.

Perhitungan nilai variabel bebas pada kedua kelompok kelas dilakukan pada data *testing* berdasarkan data *training*. Pada data testing diketahui alumni memiliki atribut variabel (X_1) Jurusan , (X_2) Jenis Kelamin , (X_3) Ekstrakurikuler , (X_4) Rata- rata nilai . untuk menghitung probabilitas setiap variabel bebas pada kedua kelas menggunakan rumus 2.3 maka perhitungannya adalah sebagai berikut :

a. Jurusan (X_1).

Adapun perhitungan nilai probabilitas variabel bebas jurusan pada data *testing* di kedua kelas adalah sebagai berikut :

Probabilitas variabel jurusan TKJ terhadap kelas < 12 bulan

$$P(X_1 = \text{TKJ} \mid < 12) = 76 / 151 = 0.5033$$

Probabilitas Varabel jurusan AP terhadap kelas < 12 bulan

$$P(X_1 = \text{AP} \mid < 12) = 75 / 151 = 0.4966$$

Probabilitas variabel jurusan TKJ terhadap kelas ≥ 12 bulan

$$P(X_1 = \text{TKJ} \mid \geq 12) = 78 / 169 = 0.461$$

Probabilitas Varabel jurusan AP terhadap kelas ≥ 12 bulan

$$P(X_1 = \text{AP} \mid \geq 12) = 91 / 169 = 0.538$$

Berikut hasil perhitungan variabel bebas jurusan terhadap setiap kelas dapat dilihat pada tabel 4.1.

Tabel Probabilitas varabel jurusan terhadap kedua kelas

P(Jurusan)	< 12	≥ 12
TKJ	0.503311	0.461538
AP	0.496689	0.538462

b. Jenis Kelamin (X_2)

Adapun perhitungan nilai probabilitas variabel bebas Jenis Kelamin pada data *testing* di kedua kelas adalah sebagai berikut :

Probabilitas variabel jenis kelamin Laki-laki terhadap kelas < 12 bulan

$$P(X_2 = \text{L} \mid < 12) = 55 / 151 = 0.364$$

Probabilitas Varabel jenis kelamin Perempuan terhadap kelas < 12 bulan

$$P(X_2 = \text{P} \mid < 12) = 96 / 151 = 0.636$$

Probabilitas variabel jenis kelamin Laki-laki terhadap kelas ≥ 12 bulan

$$P(X_2 = \text{L} \mid \geq 12) = 68 / 169 = 0.402$$

Probabilitas Varabel jenis kelamin Perempuan terhadap kelas ≥ 12 bulan

$$P(X_2 = \text{P} \mid \geq 12) = 101 / 169 = 0.597$$

Berikut hasil perhitungan variabel bebas Jenis Kelamin terhadap setiap kelas dapat dilihat pada tabel 4.2.

Tabel Probabilitas varabel Jenis Kelamin terhadap kedua kelas

P(JK)	< 12	≥ 12
L	0.364238	0.402367
P	0.635762	0.597633

c. Ekstrakurikuler (X_3)

Adapun perhitungan nilai probabilitas variabel bebas Ekstrakurikuler pada data *testing* di kedua kelas adalah sebagai berikut :

Probabilitas variabel Ekstrakurikuler ikut terhadap kelas < 12 bulan

$$P(X_3 = \text{IKUT} \mid < 12) = 138 / 151 = 0,9139$$

Probabilitas Variabel Ekstrakurikuler tidak ikut terhadap kelas < 12 bulan

$$P(X_3 = \text{TIDAK IKUT} \mid < 12) = 13 / 151 = 0,086$$

Probabilitas variabel Ekstrakurikuler ikut terhadap kelas ≥ 12 bulan

$$P(X_3 = \text{IKUT} \mid \geq 12) = 154 / 169 = 0,9112$$

Probabilitas Variabel Ekstrakurikuler tidak ikut terhadap kelas ≥ 12 bulan

$$P(X_3 = \text{TIDAK IKUT} \mid \geq 12) = 19 / 169 = 0,0887$$

Berikut hasil perhitungan variabel bebas Ekstrakurikuler terhadap setiap kelas dapat dilihat pada tabel 4.3.

Probabilitas variabel Ekstrakurikuler terhadap kedua kelas

P(EKSKUL)	2	1
IKUT	0.913907	0.911243
TIDAK IKUT	0.086093	0.088757

d. Rata –rata nilai (X_4)

Adapun perhitungan nilai probabilitas variabel bebas Rata –rata nilai pada data *testing* di kedua kelas adalah sebagai berikut :

$$P(X_4 = \leq 70 \mid < 12) = 1 / 151 = 0,006623$$

Probabilitas Variabel Nilai tidak ikut terhadap kelas < 12 bulan

$$P(X_4 = > 70 \text{ dan } < 85 \mid < 12) = 144 / 151 = 0,947$$

Probabilitas variabel Nilai ikut terhadap kelas ≥ 12 bulan

$$P(X_4 = \geq 85 \mid < 12) = 6 / 151 = 0,04635$$

Probabilitas variabel Nilai ikut terhadap kelas ≥ 12 bulan

$$P(X_4 = \leq 70 \mid \geq 12) = 0 / 169 = 0$$

Probabilitas Variabel Nilai tidak ikut terhadap kelas ≥ 12 bulan

$$P(X_4 = >70 \text{ dan } <85 | \geq 164) = 19 / 169 = 0.94$$

Probabilitas variabel Nilai ikut terhadap kelas ≥ 12 bulan

$$P(X_2 = \geq 85 | \geq 12) = 5 / 151 = 0.059$$

Berikut hasil perhitungan variabel bebas Nilai terhadap setiap kelas dapat dilihat pada tabel 4.4.

Probabilitas variabel Ekstrakurikuler terhadap kedua kelas

P(NILAI)	2	1
1	0.006623	0
2	0.94702	0.940828
3	0.046358	0.059172

3. Menghitung perkalian probabilitas prior dan probabilitas setiap variabel bebas pada setiap kelompok (posterior).

Tahap ketiga adalah menghitung perkalian probabilitas *prior* dan probabilitas

posterior pada data *testing*. Pada data *testing* pertama diketahui alumni memiliki Jurusan (X_1) TKJ, Jenis Kelamin (X_2) Laki-laki, Ekstrakurikuler (X_3) Ikut, Rata-rata nilai (X_4) 81 atau simbol 2. Adapun perhitungan perkalian probabilitas *prior* dan probabilitas *posterior* sebagai berikut:

a. Kelas Pertama < 12 bulan

$$\begin{aligned} P(Y_1 / X_1, X_2, X_3, X_4) &= P(Y_1) * P(Y_1 / X_1) * P(Y_1 / X_2) * P(Y_1 / X_3) * P(Y_1 / X_4) \\ &= P(<12) * P(X_1 = TKJ / <12) * P(X_2 = L / <12) * P(X_3 = IKUT / <12) * P(X_4 = 2 / <12) \\ &= 0,47 * 0,503 * 0,3642 * 0,913 * 0,94702 \\ &= 0,07487 \end{aligned}$$

b. Kelas kedua ≥ 12 bulan

$$\begin{aligned} P(Y_2 / X_1, X_2, X_3, X_4) &= P(Y_2) * P(Y_2 / X_1) * P(Y_2 / X_2) * P(Y_2 / X_3) * P(Y_2 / X_4) \\ &= P(\geq 12) * P(X_1 = TKJ / \geq 12) * P(X_2 = L / \geq 12) * P(X_3 = IKUT / \geq 12) * P(X_4 = 2 / \geq 12) \\ &= 0,53 * 0,461 * 0,402 * 0,911 * 0,9408 \\ &= 0,084084 \end{aligned}$$

Setelah melakukan perhitungan probabilitas posterior pada kedua kelas, dicari nilai maksimum pada kedua kelas untuk menentukan hasil klasifikasi obyek. Berdasarkan hasil perhitungan dapat diketahui bahwa kelompok yang memiliki posterior terbesar adalah ≥ 12 bulan, yaitu sebesar 0,084084 dibandingkan dengan posterior < 12 bulan, yaitu sebesar 0,07487, sehingga dapat disimpulkan data testing pertama diprediksi masuk ke dalam kelas kedua yaitu ≥ 12 Bulan. Untuk hasil selengkapnya dapat dilihat di lampiran.

4. Mengevaluasi Hasil Prediksi

Untuk mengevaluasi hasil klasifikasi pada metode naive Bayes dilakukan pengukuran tingkat akurasi dengan menghitung peluang kesalahan prediksi menggunakan *Mean Square Error* (MSE). Pada proses prediksi, semakin kecil kesalahan menunjukkan bahwa semakin baik hasil prediksi yang didapatkan. Perhitungan MSE dengan menggunakan rumus 3.5

Diketahui Nilai N / Banyaknya data *testing* adalah 80 dan hasil Penjumlahan dari $(\text{data aktual} - \text{hasil prediksi})^2 = 27$ Maka hasil $MSE = \frac{27}{80} = 0.3375$

CONTOH PERHITUNGAN K-NN

1. Menentukan nilai Parameter K.

Prediksi Waktu tunggu alumni dalam mendapatkan pekerjaan menggunakan metode K-NN dilakukan dengan menentukan nilai parameter K terlebih dahulu. Adapun nilai K yang digunakan yaitu K=1, K=3, K=5, K=7 dan K=9. Pada kali ini akan diberikan contoh perhitungan menggunakan parameter K=1

2. Menghitung Jarak Euclid

Pada tahap selanjutnya dilakukan perhitungan jarak Euclid pada objek antara

data testing dan data training berdasarkan Persamaan (3.4). Perhitungan jarak

pada penelitian ini menggunakan angka yang berasal dari kodingan kategori pada variabel bebas. Adapun perhitungan jarak Euclid antara data testing pertama dan data training pertama hingga data training ke-320 sebagai berikut:

$$\begin{aligned}d(X_1, X_1) &= \sqrt{(x_{11} - x_{11})^2 + (x_{12} - x_{12})^2 + (x_{13} - x_{13})^2 + (x_{14} - x_{14})^2} \\ &= \sqrt{(1 - 1)^2 + (2 - 1)^2 + (1 - 1)^2 + (2 - 1)^2} \\ &= 1\end{aligned}$$

$$\begin{aligned}d(X_2, X_1) &= \sqrt{(x_{21} - x_{11})^2 + (x_{22} - x_{12})^2 + (x_{23} - x_{13})^2 + (x_{24} - x_{14})^2} \\ &= \sqrt{(2 - 1)^2 + (2 - 1)^2 + (1 - 1)^2 + (2 - 1)^2} \\ &= 1.414214\end{aligned}$$

·
..
·

$$d(X_{320}, X_1)$$

$$\begin{aligned}
&= \sqrt{(x_{3201} - x_{11})^2 + (x_{3202} - x_{12})^2 + (x_{3203} - x_{13})^2 + (x_{3204} - x_{14})^2} \\
&= \sqrt{(2 - 1)^2 + (1 - 2)^2 + (1 - 1)^2 + (2 - 2)^2} \\
&= 1
\end{aligned}$$

Adapun jarak perhitungan euclid selengkapnya dapat dilihat pada lampiran.

3. Menentukan perhitungan K dari hasil perhitungan jarak.

Setelah jarak euclid didapatkan kemudian dilakukan perbandingan jarak dengan tetangga atau K yang ditentukan misal k=1 , dimana jarak yang diambil adalah jarak yang paling kecil dari jumlah tetanga tersebut. Berikut hasil perhitungan jarak antara data testing pertama dengan data training :

Hasil perhitungan jarak dan K=1

data training	data testing	jarak (d)	k =1
1	1	1	-
2	1	1.414214	-
3	1	1	-
4	1	1.732051	-
5	1	1.732051	-
6	1	1	-
7	1	0	1
8	1	1	-
9	1	1	-
10	1	1.414214	-
11	1	1	-
12	1	1	-
13	1	1	-
.	.	.	.
..
.	.	.	.
320	1	1	-

Pada tabel 3.5 diatas dilakukan perbandingan jarak terhadap semua data training dimana jumlah K yang ditentukan adalah 1 , pada tahap ini penulis menggunakan bantuan ms Excel untuk menghitung proses KNN , dimana setiap jarak dibandingkan dengan data *range jarak* 1 – 320 dimana jika data jarak pada data training tersebut lebih kecil dari data range tersebut dilakukan pelabelan waktu berdasarkan waktu real data baris tersebut, misal pada data training ke 7 dimana jarak terhadap data testing pertama adalah 0 maka 0 akan dibandingkan dengan data jarak 1 sampai 320 dan jika 0 merupakan terkecil maka akan dilakukan pelabelan waktu sesuai dengan data waktu di data training ke 7 yaitu 1 atau ≥ 12 bulan , untuk perhitungan nilai K selengkapnya dapat dilihat pada lampiran.

Dari hasil data testing pertama maka akan dilakukan pencarian terhadap label data yang paling sering muncul dimana untuk data testing pertama jumlah label data waktu 1 (≥ 12 bulan) Muncul sebanyak 28 kali sedangkan Waktu dengan Label 2 (< 12 bulan) Muncul sebanyak 21 Kali dengan hasil ini pada data testing pertama maka diprediksi lama masa tunggu alumni untuk mendapatkan pekerjaan yaitu ≥ 12 bulan.

4. Evaluasi Hasil Prediksi

Untuk mengevaluasi hasil klasifikasi pada metode KNN dilakukan pengukuran tingkat akurasi dengan menghitung peluang kesalahan prediksi menggunakan *Mean Square Error* (MSE). Pada proses prediksi, semakin kecil kesalahan menunjukkan bahwa semakin baik hasil prediksi yang didapatkan. Perhitungan MSE dengan menggunakan rumus 3.5

Diketahui Nilai N / Banyaknya data *testing* adalah 80 dan hasil Penjumlahan dari (data aktual - hasil prediksi)² = 34 Maka hasil $MSE = \frac{34}{80} = 0.425$