

Resume Pertemuan Ke-5 Machine Learning

Pengertian Naïve Bayesian

Naïve Bayes Classifier merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari Naïve Bayes Classifier ini adalah asumsi yang sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian.

Pengertian Bayesian Belief Network

Bayesian Belief Network yang disebut juga bayesian network atau belief network merupakan sebuah model peluang grafis dengan busur berarah yang merepresentasikan suatu gugus peubah dan peluang bebasnya. Metode Bayesian Belief Network terbagi atas dua kunci yang menjadi representasinya yaitu sebuah directed acyclic graph dan conditional probability table. Dua kunci ini menjadikan BBN memiliki kelebihan disbanding metodologi yang lain, karena sifatnya yang mengakomodasi perhitungan kualitatif dan kuantitatif.

Klasifikasi berdasarkan Naïve Bayesian & Bayesian Network

Klasifikasi dalam hal ini menggunakan *Naïve Bayesian* yang dikembangkan menjadi beberapa hirarki yang disebut dengan *Bayesian network* yang hirarkinya nanti akan membentuk sebuah graf dan dari grafnya itu yang menentukan nilai peluang terhadap target atau nilai peluang yang ingin kita cari. *Naive Bayesian* merupakan jumlah dari satu langkah probabilitas bersyarat. Sedangkan *Bayesian Network* merupakan beberapa langkah atau beberapa jalur dari peluang atribut atau nilai yang ingin kita capai. Perbedaan lain dari *Naïve Bayesian* dengan *Bayesian Network* adalah pada *Naïve Bayes* mengabaikan korelasi antar variabel, sedangkan pada *Bayesian Network* merupakan variabel input yang bisa saling dependen (berhubungan).

Classification & Prediction

Classification :

- Memprediksi label class (diskrit atau kontinu)
- Mengklasifikasi data (membangun model) didasarkan pada data training dan nilai label class dalam mengklasifikasikan atribut dan digunakannya pada saat mengklasifikasi data baru

Prediction :

- Memodelkan fungsi bernilai kontinu (memprediksi nilai yang tidak diketahui)
- Bentuk implementasi : pinjaman kredit, diagnosa medis, pendeteksi kegagalan, dll.

Data Preparation (Pre-processing)

Pre-processing merupakan proses yang paling banyak membutuhkan waktu karena menyangkut dengan solusi daripada masalah yang kita hadapi. Semakin bervariasi datanya (semi-terstruktur), semakin banyak waktu yang dibutuhkan. Adapun langkah-langkah dalam pre-processing yaitu:

- *Data Cleaning*, Memproses awal data untuk mengurangi noise dan mengatasi data-data yang hilang
- *Analisa Relevansi (seleksi fitur)*, Menghilangkan atribut-atribut yang tidak relevan dengan masalah atau atribut yang redundan. Contoh metodenya analisis korelasi, Q-Value, dll.
- *Transformasi Data*, Membangun normalisasi data atau menyamaratakan nilai-nilai setiap fitur supaya kedudukannya sama

Evaluating Classification Methods

Evaluating Classification Methods dilakukan untuk melihat ukuran kinerja yang mana yang menjadi patokan untuk melihat model klasifikasi mana yang akan digunakan. Adapun ukuran-ukuran yang diperhatikan yaitu sbb :

- *Akurasi*
Keakuratan klasifikasi : memperkirakan label class
Keakuratan prediksi : nilai yang ditebak dari atribut yang diprediksi
- *Kecepatan*
Waktu untuk membangun model (training time)
Waktu dalam menggunakan model (classification/prediction time)
- *Kehandalan* : mengatasi noise dan missing values

Bayesian Classification

Bayesian Classification awalnya berasal dari metode statistika. Klasifikasi berdasarkan statistik biasanya menggunakan teori peluang/probabilitas. Bayes Classification awalnya didasarkan pada Teorema Bayes dan performancinya sederhana.

Teorema Bayes

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

Keterangan :

- \mathbf{X} adalah data sampel (*evidence*) : label kelas tidak diketahui
- H adalah dugaan (*hypothesis*) bahwa \mathbf{X} adalah anggota C
- Klasifikasi ditentukan $P(H|\mathbf{X})$, (*posteriori probability*), probabilitas bahwa dugaan terhadap data sampel \mathbf{X}
- $P(H)$ (*prior probability*), initial probability
- $P(\mathbf{X})$: probabilitas dari sampel data yang diamati
- $P(\mathbf{X}|H)$ (*likelihood*), probabilitas dari sampel \mathbf{X} , dengan memperhatikan dugaan
- Data training data \mathbf{X} , *posteriori probability* dari *hypothesis* H atau class, $P(H|\mathbf{X})$. teorema Bayes ini dapat ditulis dengan

Prediksi berdasarkan Teorema Bayes

- Dengan data training \mathbf{X} , probabilitas posteriori dari hipoteses H , $P(H|\mathbf{X})$ mengikuti teorema Bayes
- Prediksi \mathbf{X} adalah milik C_i jika probabilitas $P(C_i|\mathbf{X})$ adalah yang tertinggi di antara semua $P(C_k|\mathbf{X})$ untuk semua kelas k

Klasifikasi untuk memperoleh Posteriori Maksimum

- Perhatikan D adalah record training dan ditetapkan label-label kelasnya dan masing-masing record dinyatakan n atribut (n field) $\mathbf{X} = (x_1, x_2, \dots, x_n)$
- Misalkan terdapat m kelas C_1, C_2, \dots, C_m
- Klasifikasi adalah diperoleh maksimum posteriori yaitu maksimum $P(C_i|\mathbf{X})$
- Yang diperoleh dari teorema Bayes

$$P(C_i|\mathbf{X}) = \frac{P(\mathbf{X}|C_i)P(C_i)}{P(\mathbf{X})}$$

- Karena $P(\mathbf{X})$ adalah konstan untuk semua kelas, hanya $P(C_i|\mathbf{X}) = P(\mathbf{X}|C_i)P(C_i)$ perlu dimaksimumkan

Klasifikasi Naïve Bayes Classifier

- Diasumsikan atribut dalam kondisi saling bebas (independen) yaitu tidak ada ketergantungan antara atribut-atribut :

$$P(\mathbf{X}|C_i) = \prod_{k=1}^n P(x_k|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times \dots \times P(x_n|C_i)$$

- A_k adalah kategorikal, $P(x_k|C_i)$ adalah jumlah record dalam kelas C_i yang memiliki nilai x_k sama dengan A_k dibagi dengan $|C_i|$ jumlah record dalam C_i dalam D
- Jika A_k bernilai kontinu, $P(x_k|C_i)$ biasanya dihitung berdasarkan pada distribusi Gaussian dengan mean μ dan standar deviasi σ

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

dan $P(x_k|C_i)$ adalah

$$P(\mathbf{X}|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i})$$

Menghindari Masalah Probabilitas 0

Prediksi Naïve Bayesian membutuhkan masing-masing probabilitas tidak nol. Dengan kata lain, probabilitas yang dihitung tidak menjadi nol

$$P(\mathbf{X}|C_i) = \prod_{k=1}^n (x_k|C_i)$$

Implementasi Naive Bayes

Contoh perhitungan metode naive bayes untuk sistem pakar penentuan kerusakan pada laptop, Kerusakan laptop yang dibahas disini adalah tentang kerusakan dibagian hardware didalam laptop. Berikut adalah data yang disajikan.

Data kerusakan laptop :

K1 = IC Charger Rusak

K2 = IC Power Rusak

K3 = Resistor Rusak

K4 = Kapasitor Rusak

K5 = Mofset Rusak

K6 = Embeded Controllor Rusak

Data gejala yang timbul :

G1 = Indikator pengisian baterai nyala tapi laptop tidak bisa dinyalakan.

G2 = Indikaor pengisian baterai mati, laptop tidak bisa dinyalakan.

G3 = Indikaor pengisian baterai nyala, bisa dinyalakan tapi tidak tampil pada layar.

G4 = Input seperti USB tidak berfungsi

Keterangan :

K = Kerusakan

G = Gejala

Dari data gejala dan kerusakan diatas kita menentukan tabel keputusan antara kerusakan dan gejala yang timbul. Tabel keputusan berfungsi untuk menentukan laptop tersebut mengalami kerusakan apa, berdasarkan gejala yang timbul. Berikut adalah tabel keputusan yang sudah ditentukan.

Gejala	Kerusakan					
	K1	K2	K3	K4	K5	K6
G1	1	0	1	0	0	0
G2	0	1	0	0	0	1
G3	0	0	1	0	0	0
G4	0	1	0	1	1	1

Keterangan :

1 = Gejala muncul

0 = Tidak ada gejala yang muncul

Contoh Kasus :

Misalnya gejala yang tampak pada laptop ada dua gejala yaitu :

G1 : Indikator pengisian baterai nyala tapi laptop tidak bisa dinyalakan, dan

G3 : Indikator pengisian baterai nyala, bisa dinyalakan tapi tidak tampil pada layar.

Berdasarkan gejala yang muncul tersebut maka langkah perhitungannya adalah sebagai berikut :

Langkah 1 : menentukan penyakit yang muncul berdasarkan tabel keputusan

Berdasarkan gejala yang muncul G1 dan G3 , maka bisa dilihat dari tabel keputusan indikasi kerusakan yang akan di prediksi yaitu K1 dan K3. karena pada K1 terdapat G1 dan G3 yang bernilai 1 dan pada K3 terdapat G3 yang bernilai 1.

Maka untuk tahap selanjutnya yang di hitung menggunakan algoritma naive bayes adalah menghitung nilai probabilitas gejala dari K1 dan K3.

Langkah 2 : menghitung nilai probabilitas kerusakan dan gejala.

Pada langkah 1 sudah di dapatkan indikasi penyakit yang di prediksi berdasarkan gejala yang timbul, sesuai tabel keputusan. Langkah selanjutnya yaitu menghitung nilai probabilitas dari masing-masing kerusakan dan gejala yang timbul.

Perhitungan Probabilitas K1 (IC Charger Rusak)

Rumus menghitung probabilitas nilai K1

$$\text{Rumus Probabilitas K1} = \frac{\text{jumlah kemungkinan kerusakan yang muncul}}{\text{jumlah semua kerusakan}} = \frac{1}{6} = 0.16$$

Keterangan :

- Angka 1 di dapatkan dari prediksi minimal kerusakan yang muncul
- Angka 6 di dapatkan dari jumlah semua kerusakan yang ada pada tabel keputusan

Rumus menghitung probabilitas gejala yang muncul

G1 : Indikator pengisian baterai nyala tapi laptop tidak bisa dinyalakan.

$$G1 = \frac{\text{jumlah kemungkinan}}{\text{jumlah kemungkinan kerusakan akibat gejala}} = \frac{1}{2} = 0.5$$

G3 : Indikator pengisian baterai nyala, bisa dinyalakan tapi tidak tampil pada layar.

$$G3 = \frac{\text{jumlah kemungkinan}}{\text{jumlah kemungkinan kerusakan akibat gejala}} = \frac{0}{2} = 0$$

Keterangan :

jumlah kemungkinan = jumlah gejala G1/G3 yang muncul pada K1 di tabel keputusan

jumlah kemungkinan kerusakan akibat gejala = kerusakan yang muncul yang di akibatkan gejala dalam perhitungan kali ini didapatkan 2 kerusakan yang muncul yaitu K1 dan K3

Perhitungan Probabilitas K3 (Resistor Rusak)

Rumus menghitung probailitas nilai K3

$$\text{Rumus Probabilitas K3} = \frac{\text{jumlah kemungkinan kerusakan yang muncul}}{\text{jumlah semua kerusakan}} = \frac{1}{6} = 0.16$$

Keterangan :

- Angka 1 di dapatkan dari prediksi minimal kerusakan yang muncul
- Angka 6 di dapatkan dari jumlah semua kerusakan yang ada pada tabel keputusan

Rumus menghitung probabilitas gejala yang muncul

G1 : Indikator pengisian baterai nyala tapi laptop tidak bisa dinyalakan.

$$G1 = \frac{\text{jumlah kemungkinan}}{\text{jumlah kemungkinan kerusakan akibat gejala}} = \frac{1}{2} = 0.5$$

G3 : Indikator pengisian baterai nyala, bisa dinyalakan tapi tidak tampil pada layar.

$$G3 = \frac{\text{jumlah kemungkinan}}{\text{jumlah kemungkinan kerusakan akibat gejala}} = \frac{1}{2} = 0.5$$

Keterangan :

jumlah kemungkinan = jumlah gejala G1/G3 yang muncul pada K3 di tabel keputusan

jumlah kemungkinan kerusakan akibat gejala = kerusakan yang muncul yang di akibatkan gejala dalam perhitungan kali ini didapatkan 2 kerusakan yang muncul yaitu K1 dan K3

Langkah 3 : Menghitung nilai bayes berdasarkan probabilitas kerusakan dan gejala yang timbul

Dari nilai probabilitas diatas selanjutnya tahap perhitungan nilai bayes dengan rumus sebagai berikut

Menghitung Nilai Bayes K1

$$K(K1|G1) = \frac{[K(G1|K1) \times K(K1)]}{[K(G1|K1) \times K(K1) + K(G1|K3) \times K(K3)]}$$

$$= \frac{0.5 \times 0.16}{0.5 \times 0.16 + 0.5 \times 0.16} = \frac{0.08}{0.16} = 0.5$$

$$K(K1|G3) = \frac{[K(G3|K1) \times K(K1)]}{[K(G3|K1) \times K(K1) + K(G3|K3) \times K(K3)]}$$

$$= \frac{0 \times 0.16}{0 \times 0.16 + 0.5 \times 0.16} = \frac{0}{0.08} = 0$$

Total nilai bayes dari K1 yaitu :

$$Total K1 = K(K1 | G1) + K(K1 | G3)$$

$$Total K1 = 0.5 + 0 = 0.5$$

Menghitung Nilai Bayes K3

$$K(K3|G1) = \frac{[K(G1|K3) \times K(K3)]}{[K(G1|K1) \times K(K1) + K(G1|K3) \times K(K3)]}$$

$$= \frac{0.5 \times 0.16}{0.5 \times 0.16 + 0.5 \times 0.16} = \frac{0.08}{0.16} = 0.5$$

$$K(K3|G3) = \frac{[K(G3|K3) \times K(K3)]}{[K(G3|K1) \times K(K1) + K(G3|K3) \times K(K3)]}$$

$$= \frac{0.5 \times 0.16}{0.5 \times 0.16 + 0.5 \times 0.16} = \frac{0.08}{0.16} = 0.5$$

Total nilai bayes dari K3 yaitu :

$$Total\ K3 = K(K3 | G1) + K(K3 | G3)$$

$$Total\ K1 = 0.5 + 0.5 = 1$$

Menjumlahkan hasil nilai bayes dari K1 dan K3

$$Hasil\ total = Total\ Bayes\ K1 + Total\ Bayes\ K2$$

$$= 0.5 + 1$$

$$= 1.5$$

Langkah 4 : Menghitung presentase nilai prediksi kerusakan

Dari perhitungan hasil total didapatkan nilai 1.5 . Angka tersebut nantinya di gunakan sebagai pembagi masing-masing nilai bayes dari K1 dan K3 untuk di ketehaui presentasenya. Berikut ini adalah hasil yang didapatkan dari perhitungan tersebut.

1. Kerusakan Pada IC charger (K1)

$$= \frac{Total\ Bayes\ K1}{Total\ Hasil} \times 100\% = \frac{0.5}{1.5} \times 100\% = 33.3\%$$

2. Kerusakan Pada Resistor (K3)

$$= \frac{Total\ Bayes\ K3}{Total\ Hasil} \times 100\% = \frac{1}{1.5} \times 100\% = 66.6\%$$

Dari hasil presentase diatas maka didapatkan nilai presentase tertinggi adalah hasil kerusakan yang didapatkan. Dengan demikian jika ada laptop yang mengalami gejala kerusakan G1 (Indikator pengisian baterai nyala tapi laptop tidak bisa dinyalakan.) dan G3 (Indikator pengisian baterai nyala, bisa dinyalakan tapi tidak tampil pada layar.). Maka laptop tersebut mengalami kerusakan K3 (Kerusakan Pada Resistor).