Stemming Results

# Original Text:

Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 1 CONTOH KASUS NAÏVE BAYES CLASSIFIER MACHINE LEARNING Disusun Oleh: Dr. Jasman Pardede, S.Si., M.T. TEKNIK INFORMATIKA FAKULTAS TEKNIK INDUSTRI INSTITUT TEKNOLOGI NASIONAL BANDUNG   
  
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 2 Algoritma Naive Bayes merupakan sebuah metoda klasifikasi menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma Naïve Bayes adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk pembelajaran mesin dan data mining. Algoritma Naïve Bayes merupakan algoritma yang populer dalam aplikasi pembelajaran mesin karena kesederhanaan algoritmanya. Algoritma Naïve Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga dikenal sebagai Teorema Bayes. Algoritma Naïve Bayes menggunakan asumsi bahwa setiap atribut memiliki hubungan yang saling bebas yaitu tidak ada ketergantungan antara satu atribut dengan atribut lainnya. Sebagai contoh, buah dapat dianggap sebagai buah appel, jika buah tersebut memiliki atribut warna merah, bentuknya bulat, dan memiliki diameter sekitar 8 cm. Pada Naïve Bayes menggunakan asumsi bahwa tidak ada hubungan antara warna, bentuk, dan diameter dalam menentukan buah appel. Walaupun pada realitanya bahwa asumsi “hubungan yang saling bebas” atau independensi tersebut sangat jarang terjadi. Ciri utama dari klasifikasi Naïve Bayes ini adalah asumsi yang sangat kuat (naïf) akan independensi dari masing-masing atribut atau kondisi/kejadian. Setiap atribut pada algoritma Naïve Bayes berkontribusi terhadap keputusan akhir secara setara dan independen atau bebas dari atribut lainnya. Sehingga algoritma Naïve Bayes efisiensi secara komputasi dan cocok untuk berbagai domain. Walaupun asumsi independensi tersebut dilanggar, tetapi kinerja klasifikasi dengan Naïve Bayes cukup tinggi. Hal ini telah dibuktikan pada berbagai hasil penelitian empiris. Pada penelitian Xhemali, dkk. (2009) menyatakan bahwa klasifikasi Naïve Bayes memiliki kinerja akurasi yang lebih baik dibandingkan dengan algoritma klasifikasi lainnya. Algoritma Naïve Bayes memiliki kelebihan, diantaranya: mudah dipahami, mudah diimplementasikan pada suatu bahasa pemrograman tertentu, dapat digunakan untuk data kuantitatif maupun kualitatif, perhitungannya cepat dan efisien, tidak memerlukan jumlah data yang banyak, tidak memerlukan data training (pelatihan) yang banyak, dapat digunakan untuk klasifikasi biner atau multi-class, dapat mengabaikan data yang hilang dalam perhitungan, dan lain-lain.   
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 3 Sedangkan kekurangan algoritma Naïve Bayes diantaranya: memerlukan pengetahuan awal atau pengetahuan masa lalu dalam membuat suatu keputusan, tidak cocok digunakan untuk kasus yang memiliki korelasi antara satu atribut dengan atribut lainnya, probabilitas prediksi akan bernilai nol jika probabilitas kondisional bernilai nol, dan lain-lain. Rumus umum teorema Bayes adalah sebagai berikut: 𝑃(𝒄|𝒙) = 𝑃(𝒙|𝒄) 𝑃(𝒄)𝑃(𝒙) Dimana: x : data dengan class yang belum diketahui c : hipotesis data merupakan suatu class yang spesifik P(x|c) : probabilistik hipotesis x berdasarkan kondisi pada hipotesis c P(c) : probabilistik hipotesis c (prior probabilistik) P(c|x) : probabilistik hipotesis c berdasarkan kondisi pada hipotesis x (posteriori probabilistik) Contoh kasus 1: Diberikan data seseorang berolah raga seperti yang dinyatakan pada Tabel 1. Tentukanlah apakah seseorang akan berolah raga jika diketahui rain, mild, high, dan weak? Tabel 1. Data berolah raga sesorang Day Outlook Temperature Humidity Wind Play Tennis D1 Sunny Hot High Weak No D2 Sunny Hot High Strong No D3 Overcast Hot High Weak Yes D4 Rain Mild High Weak Yes D5 Rain Cool Normal Weak Yes D6 Rain Cool Normal Strong No D7 Overcast Cool Normal Strong Yes D8 Sunny Mild High Weak No D9 Sunny Cool Normal Weak Yes D10 Rain Mild Normal Weak Yes D11 Sunny Mild Normal Strong Yes D12 Overcast Mild High Strong Yes D13 Overcast Hot Normal Weak Yes   
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 4 Day Outlook Temperature Humidity Wind Play Tennis D14 Rain Mild High Strong No Penyelesaian: Berdasarkan Tabel 1 diperoleh bahwa banyaknya sunny adalah 5, yaitu: D1, D2, D8, D9, dan D11. Jumlah orang yang berolah raga dengan syarat sunny adalah 2, yaitu: D9 dan D11. Jumlah orang yang tidak berolah raga ketika sunny adalah 3, yaitu: D1, D2, dan D8. Dengan cara yang sama diperoleh tabel prekuensi orang berolah raga atau tidak seperti yang dinyatakan pada Tabel 2. Tabel 2. Frekuensi seseorang bermain tennis Tabel Frekuensi Play Tennis Yes No Outlook Sunny 2 3 Overcast 4 0 Rain 3 2 Sehingga tabel likelihood dari tabel frekuensi seseorang bermain tennis adalah seperti yang dinyatakan pada Tabel 3. Tabel 3. Likelihood outlook seseorang bermain tennis Tabel Likelihood Play Tennis Persentasi Yes No Outlook Sunny 2/9 3/5 5/14 Overcast 4/9 0/5 4/14 Rain 3/9 2/5 5/14 9/14 5/14 Tabel 4. Likelihood temperature seseorang bermain tennis Tabel Likelihood Play Tennis Persentasi Yes No Temperature Hot 2/9 2/5 4/14 Mild 4/9 2/5 6/14 Cool 2/9 2/5 4/14 8/14 6/14 Tabel 5. Likelihood humidity seseorang bermain tennis Tabel Likelihood Play Tennis Persentasi Yes No Humidity High 3/9 4/5 7/14 Normal 6/9 1/5 7/14   
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 5 Tabel Likelihood Play Tennis Persentasi Yes No 9/14 5/14 Tabel 6. Likelihood wind seseorang bermain tennis Tabel Likelihood Play Tennis Persentasi Yes No Wind Weak 6/9 2/5 8/14 Strong 2/9 4/5 6/14 8/14 6/14 Berdasarkan Tabel 3 sampai dengan Tabel 6 diperoleh bahwa: a. peluang seseorang bermain tennis adalah P(yes) = 9/14 b. peluang seseorang tidak bermain tennis adalah P(no) = 5/14 c. peluang sunny adalah P (sunny) = 5/14 d. peluang overcast adalah P(overcast) = 4/14 e. peluang rain adalah P(rain) = 5/14 f. peluang seseorang bermain tennis ketika sunny adalah P(sunny|yes) = 3/9 g. peluang seseorang tidak bermain tennis ketika sunny adalah P(sunny|no) = 2/5 h. peluang seseorang bermain tennis ketika overcast adalah P(overcast|yes) = 4/9 i. peluang seseorang tidak bermain tennis ketika overcast adalah P(overcast|no) = 0/5 j. peluang seseorang bermain tennis ketika rain adalah P(rain|yes) = 2/9 k. peluang seseorang tidak bermain tennis ketika rain adalah P(rain|no) = 3/5 l. peluang seseorang tidak bermain tennis ketika rain adalah P(mild|yes) = 4/9 m. peluang seseorang tidak bermain tennis ketika rain adalah P(mild|no) = 2/5 n. peluang seseorang tidak bermain tennis ketika rain adalah P(high|yes) = 3/9 o. peluang seseorang tidak bermain tennis ketika rain adalah P(high|no) = 4/5 p. peluang seseorang tidak bermain tennis ketika rain adalah P(weak|yes) = 6/9 q. peluang seseorang tidak bermain tennis ketika rain adalah P(weak|no) = 2/5 P(bermain tennis = yes | X) = P(bermain\_tennis=yes) \* P(outlook = rain | bermain\_tennis= yes) \*P (temperatur = mild | bermain\_tennis= yes) \* P (humidity = high | bermain\_tennis= yes) \* P (wind = weak | bermain\_tennis= yes) = (9/14) \* (3/9) \* (4/9) \* (3/9) \* (6/9) = 0.02116402   
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 6 P(bermain tennis = no | X) = P(bermain\_tennis=no) \* P(outlook = rain | bermain\_tennis= no) \*P (temperatur = mild | bermain\_tennis= no) \* P (humidity = high | bermain\_tennis= no) \* P (wind = weak | bermain\_tennis= no) = (5/14) \* (3/5) \* (2/5) \* (4/5) \* (2/5) = 0.02743 Jadi, berdasarkan hasil perhitungan peluang bermain tennis di atas diperoleh bahwa hasil peluang (ya) bermain tennis = 0.01411 < peluang (tidak) bermain tennis = 0.02743 yaitu 1.41% < 2.74%, maka dapat disimpulkan bahwa seseorang bermain tennis dengan kondisi outlook rain, temperature mild, humidity high, dan wind weak TIDAK AKAN BERMAIN TENNIS. Contoh kasus 2: Tabel 7. Data sesorang melanggar rambu lalu lintas Warna Tipe Asal Tercuri (X1) (X2) (X3) (X4) Merah Sport Domestik Ya Merah Sport Domestik Tidak Merah Sport Domestik Ya Kuning SUV Domestik Tidak Kuning Sport Import Ya Kuning SUV Import Tidak Kuning SUV Import Ya kuning SUV Domestik Tidak Merah SUV Import Tidak Merah Sport Import Ya Dari Tabel 7 di atas, data mobil yang melanggar rambu lalu lintas bisa dilihat dari atribut warna, tipe, dan asal. Misalkan kita ingin mengelompokkan mobil warna merah, tipe SUV, dan asal domestik. Tentukan probabilitas pelanggaran lalu lintas dan probabilitas tidak melanggar rambu lalu lintas, dan kemudian tentukan berapa persen mobil yang melanggar dan berapa persen mobil yang tidak melanggar, serta tentukan mobil dengan warna merah, tipe SUV, dan asal domestik tersebut melanggar lalu lintas atau tidak? Penyelesaian: Tabel 8. Frekuensi warna Tercuri Ya Tidak Warna Merah 3 2 Kuning 2 3   
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 7 Tabel 9. Likelihood warna Likelihood Tercuri Persentasi Ya Tidak Warna Merah 3/5 2/5 5/10 Kuning 2/5 3/5 5/10 5/10 5/10 Tabel 10. Frekuensi tipe Melanggar Ya Tidak Tipe Sport 4 1 SUV 1 4 Tabel 11. Likelihood tipe Likelihood Melanggar Persentasi Ya Tidak Tipe Sport 4/5 1/5 5/10 SUV 1/5 4/5 5/10 5/10 5/10 Tabel 12. Frekuensi Asal Melanggar Ya Tidak Asal Domestik 2 3 Import 3 2 Tabel 13. Likelihood Asal Likelihood Melanggar Persentasi Ya Tidak Asal Domestik 2/5 3/5 5/10 Import 3/5 2/5 5/10 5/10 5/10 Sehingga: a. Peluang melanggar lalu lintas, P(ya) = 5/10 = 0.5 b. Peluang tidak melanggar lalu lintas, P(tidak) = 5/10 = 0.5 c. Peluang merah melanggar lalu lintas, P(merah| ya) = 3/5 = 0.6 d. Peluang SUV melanggar lalu lintas, P(SUV|ya) = 1/5 = 0.2 e. Peluang domestik melanggar lalu lintas, P(domestik|ya) = 2/5 = 0.4 f. Peluang merah tidak melanggar lalu lintas, P(merah| tidak) = 2/5 = 0.4 g. Peluang SUV tidak melanggar lalu lintas, P(SUV|tidak) = 4/5 = 0.8 h. Peluang domestik tidak melanggar lalu lintas, P(domestik|tidak) = 3/5 = 0.6 i. Peluang melanggar lalu lintas, P(melanggar=ya|X) = P(ya) \* P(merah|ya) \* P(SUV|ya) \* P(domestik|ya) = 0.5 \* 0.6 \* 0.2 \* 0.4 = 0.024 = 2.4%   
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 8 j. Peluang tidak melanggar lalu lintas, P(melanggar =tidak|X) = P(tidak) \* P(merah|tidak) \* P(SUV| tidak) \* P(domestik| tidak) = 0.5 \* 0.6 \* 0.8 \* 0.6 = 0.144 = 14.4%. Jadi, berdasarkan hasil perhitungan pelanggaran lalu lintas di atas dengan hasil pelanggaran (tidak) > pelanggaran (ya) yaitu 14.4% > 2.4% maka dapat disimpulkan mobil dengan warna merah, tipe SUV, dan asal domestik TIDAK MELANGGAR rambu lalu lintas. Contoh kasus 3: Dari hasil pengamatan sebelumnya, diperoleh bahwa penggunaan Listrik dipengaruhi oleh jumlah tanggungan keluarga, luas rumah, pendapatan per bulan, daya listrik yang digunakan, dan perlengkapan yang dimiliki. Adapun data hasil pengamatan yang dilakukan adalah sebagai berikut: Tabel 14. Data penggunaan listrik No Jumlah Tanggungan Keluarga Luas Rumah Pendapatan/ Bulan Daya Listrik Perlengkapan Yang Dimiliki Penggunaan Listrik 1 banyak besar besar tinggi sedang tinggi 2 sedang kecil sedang tinggi tinggi tinggi 3 sedang standar besar rendah tinggi sedang 4 sedikit standar kecil tinggi tinggi tinggi 5 sedang besar besar tinggi tinggi rendah 6 sedikit besar besar rendah sedang tinggi 7 sedang besar sedang sedang tinggi tinggi 8 banyak besar besar tinggi tinggi sedang 9 sedang standar besar sedang tinggi tinggi 10 sedikit standar sedang sedang sedang tinggi 11 sedikit besar kecil tinggi sedang sedang 12 sedang kecil kecil tinggi tinggi sedang 13 banyak besar besar tinggi sedang tinggi 14 banyak besar besar sedang tinggi tinggi 15 sedang besar besar sedang tinggi tinggi 16 sedang standar besar tinggi tinggi tinggi 17 banyak standar sedang tinggi tinggi sedang 18 sedang besar besar sedang tinggi tinggi 19 banyak besar sedang tinggi tinggi tinggi   
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 9 No Jumlah Tanggungan Keluarga Luas Rumah Pendapatan/ Bulan Daya Listrik Perlengkapan Yang Dimiliki Penggunaan Listrik 20 sedikit besar besar sedang rendah sedang 21 sedang standar besar sedang tinggi tinggi 22 banyak standar besar tinggi tinggi sedang 23 banyak besar kecil tinggi tinggi tinggi 24 banyak besar sedang tinggi tinggi tinggi 25 sedang besar sedang sedang sedang sedang 26 banyak kecil kecil rendah tinggi rendah 27 sedang standar sedang tinggi sedang sedang 28 banyak besar sedang sedang sedang tinggi 29 sedang standar besar tinggi tinggi sedang 30 banyak kecil sedang tinggi sedang tinggi 31 banyak besar besar sedang rendah rendah 32 sedikit besar sedang rendah tinggi tinggi 33 sedang besar besar tinggi rendah rendah 34 banyak kecil besar sedang tinggi sedang 35 banyak besar besar tinggi rendah sedang 36 sedang standar kecil tinggi tinggi sedang 37 banyak standar besar tinggi rendah sedang 38 banyak besar sedang rendah sedang rendah 39 sedikit besar kecil tinggi sedang rendah 40 banyak besar besar tinggi sedang tinggi 41 banyak standar besar tinggi tinggi sedang 42 sedikit besar besar tinggi tinggi sedang 43 banyak besar besar tinggi tinggi tinggi 44 sedikit besar sedang sedang tinggi tinggi 45 banyak kecil kecil tinggi tinggi tinggi 46 banyak standar sedang sedang tinggi rendah 47 banyak kecil sedang tinggi tinggi tinggi 48 sedang besar sedang sedang tinggi sedang 49 banyak besar sedang tinggi tinggi rendah 50 sedang besar besar tinggi tinggi rendah 51 banyak kecil kecil tinggi tinggi tinggi 52 sedang standar kecil tinggi sedang tinggi 53 sedang kecil sedang rendah tinggi sedang   
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 10 No Jumlah Tanggungan Keluarga Luas Rumah Pendapatan/ Bulan Daya Listrik Perlengkapan Yang Dimiliki Penggunaan Listrik 54 banyak besar sedang tinggi tinggi sedang 55 banyak standar besar sedang tinggi tinggi 56 banyak kecil sedang tinggi tinggi sedang 57 sedang besar besar rendah tinggi tinggi 58 banyak besar besar tinggi tinggi sedang Tentukanlah Correctly dan incorrectly classified instance dari data penggunaan listrik yang diberikan. Penyelesaikan: 1. Menghitung Probabilitas class penggunaan listrik: Jumlah kelas penggunaan listrik adalah 3, yaitu: tinggi, sedang, dan rendah. Dengan jumlah masing-masing adalah: 28, 21, dan 9. Seperti yang dinyatakan pada Tabel 15. Probabilitas (Tinggi) = jumlah\_tinggi/total\_data =∑\*+,-\_/0123045678869:6;<, =>?> = 0.4828 Tabel 15. Probabilitas Penggunaan Listrik 2. Menghitung Probabilitas bersyaratnya: Pada kasus ini, atribut bersayaratnya ada 5 (lima), yaitu: a. Jumlah tanggungan keluarga b. Luas tanah c. Pendapatan per bulan d. Daya listrik’ e. Perlengkapan yang dimiliki. Perhitungan numerik terhadap peluang bersyarat Penggunaan Listrik dengan syarat Jumlah tanggungan adalah sebegai berikut: Pada atribut Jumlah tanggungan memiliki 3 kelas, yaitu: banyak, sedang, dan sedikit. Adapun jumlah pengguna listrik dengan syarat jumlah tanggungan adalah seperti yang dinyatakan pada Tabel 16.   
  
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 11 Tabel 16. Probabilitas Jumlah Tanggungan Nilai 10 menyatakan bahwa jumlah pengguna listrik sedang dan jumlah tanggungannya banyak adalah 10 keluarga. Nilai 8 menyatakan bahwa jumlah pengguna listrik sedang dan jumlah tanggungannya sedang adalah 8 keluarga. Sedangkan untuk pengguna listrik sedang dengan jumlah tanggungan sedikit ada sebanyak 3 keluarga. Sehingga jumlah total pengguna listrik sedang adalah (10+8+3) = 21. Probabilitas pengguna listrik sedang dengan jumlah tanggungan banyak adalah 0.4762. Dengan cara yang sama dilakukan untuk probabilitas setiap kelas penggunaan listrik dengan syarat kelas jumlah tanggungan. Dengan cara yang sama, untuk probabilitas bersyarat lainnya diperoleh seperti yang dinyatakan tabel berikut:   
  
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 12 3. Pengujian Metode Naïve Bayes terhadap penggunaan Listrik Hitung probabilitas penggunaan listrik dengan syarat dari setiap kategori kelas yang diberikan. Pada kasus pertama, jumlah tanggungan keluarga (x1) = banyak, luas rumah (X2) = besar, pendapatan per bulan (X3) = besar, daya listrik yang digunakan (X4) = tinggi, dan perlengkapan yang dimiliki (X5) = sedang. Sehingga P(penggunaan=rendah | X) = (P(jumlah tanggungan = banyak | penggunaan=rendah) \* P(luas tanah =besar | penggunaan=rendah) \* P(pendapatan = besar |penggunaan = rendah) \* P(daya listrik = tinggi | penggunaan = rendah) \* P(perlengkapan = sedang | penggunaan = rendah)) \* P(penggunaan = rendah) = 0,5556 \*0,7778 \*0,4444\*0,5556\*0,2222\*0,15517241 =0,003679 Dengan cara yang sama untuk - P(penggunaan=sedang|X) = 0,003950 - P(penggunaan=tinggi|X) = 0,009960 Dari probabilitas ketiga kelas peluang terbesar adalah 0.009960, sehingga prediksi sistem adalah pengguna tinggi. Berdasarkan hasil prediksi yang dilakukan diperoleh data jumlah yang benar diprediksi sebanyak 36, sedangkan yang tidak tepat sebanyak 22. Sehingga nilai Correctly classified-nya adalah 62.07%, sedangkan incorrectly classified nya adalah 37.93%.   
  
Naïve Bayes – IF 358 Machine Learning Oleh Jasman Pardede 13 https://www.datacamp.com/community/tutorials/categorical-data https://pbpython.com/categorical-encoding.html

# Stemmed Text:

naïve bayes – if 358 machine learning jasman pardede 1 contoh naïve bayes classifier machine learning susun oleh: dr. jasman pardede, s.si., m.t. teknik informatika fakultas teknik industri institut teknologi nasional bandung naïve bayes – if 358 machine learning jasman pardede 2 algoritma naive bayes metoda klasifikasi metode probabilitas statistik dikemukakan ilmuwan inggris thomas bayes. algoritma naïve bayes salah algoritma pembelajaran induktif efektif efisien pembelajaran mesin data mining. algoritma naïve bayes algoritma populer aplikasi pembelajaran mesin kesederhanaan algoritmanya. algoritma naïve bayes memprediksi luang berdasarkan pengalaman sebelumnya, kenal teorema bayes. algoritma naïve bayes asumsi atribut memiliki hubung bebas ketergantungan atribut atribut lainnya. contoh, buah anggap buah appel, buah memiliki atribut warna merah, bentuk bulat, memiliki diameter 8 cm. naïve bayes asumsi hubung warna, bentuk, diameter menentukan buah appel. realita asumsi “hubungan bebas” independen jarang terjadi. ciri utama klasifikasi naïve bayes asumsi kuat (naïf) independen atribut kondisi/kejadian. atribut algoritma naïve bayes berkontribusi keputusan tara independen bebas atribut lainnya. algoritma naïve bayes efisiensi komputasi cocok domain. asumsi independen dilanggar, kinerja klasifikasi naïve bayes tinggi. dibuktikan hasil penelitian empiris. penelitian xhemali, dkk. (2009) klasifikasi naïve bayes memiliki kinerja akurasi dibandingkan algoritma klasifikasi lainnya. algoritma naïve bayes memiliki kelebihan, diantaranya: mudah dipahami, mudah diimplementasikan bahasa pemrograman tertentu, data kuantitatif kualitatif, perhitungan cepat efisien, data banyak, data training (pelatihan) banyak, klasifikasi biner multi-class, mengabaikan data hilang perhitungan, lain-lain. naïve bayes – if 358 machine learning jasman pardede 3 kekurangan algoritma naïve bayes diantaranya: pengetahuan pengetahuan keputusan, cocok memiliki korelasi atribut atribut lainnya, probabilitas prediksi bernilai nol probabilitas kondisional bernilai nol, lain-lain. rumus teorema bayes berikut: 𝑃(𝒄|𝒙) = 𝑃(𝒙|𝒄) 𝑃(𝒄)𝑃(𝒙) dimana: x : data class c : hipotesis data class spesifik p(x|c) : probabilistik hipotesis x berdasarkan kondisi hipotesis c p(c) : probabilistik hipotesis c (prior probabilistik) p(c|x) : probabilistik hipotesis c berdasarkan kondisi hipotesis x (posteriori probabilistik) contoh 1: data o raga dinyatakan tabel 1. tentu o raga rain, mild, high, weak? tabel 1. data o raga sorang day outlook temperature humidity wind play tennis d1 sunny hot high weak no d2 sunny hot high strong no d3 overcast hot high weak yes d4 rain mild high weak yes d5 rain cool normal weak yes d6 rain cool normal strong no d7 overcast cool normal strong yes d8 sunny mild high weak no d9 sunny cool normal weak yes d10 rain mild normal weak yes d11 sunny mild normal strong yes d12 overcast mild high strong yes d13 overcast hot normal weak yes naïve bayes – if 358 machine learning jasman pardede 4 day outlook temperature humidity wind play tennis d14 rain mild high strong no penyelesaian: berdasarkan tabel 1 diperoleh banyak sunny 5, yaitu: d1, d2, d8, d9, d11. orang o raga syarat sunny 2, yaitu: d9 d11. orang o raga sunny 3, yaitu: d1, d2, d8. diperoleh tabel prekuensi orang o raga dinyatakan tabel 2. tabel 2. frekuensi bermain tennis tabel frekuensi play tennis yes no outlook sunny 2 3 overcast 4 0 rain 3 2 tabel likelihood tabel frekuensi bermain tennis dinyatakan tabel 3. tabel 3. likelihood outlook bermain tennis tabel likelihood play tennis persentasi yes no outlook sunny 2/9 3/5 5/14 overcast 4/9 0/5 4/14 rain 3/9 2/5 5/14 9/14 5/14 tabel 4. likelihood temperature bermain tennis tabel likelihood play tennis persentasi yes no temperature hot 2/9 2/5 4/14 mild 4/9 2/5 6/14 cool 2/9 2/5 4/14 8/14 6/14 tabel 5. likelihood humidity bermain tennis tabel likelihood play tennis persentasi yes no humidity high 3/9 4/5 7/14 normal 6/9 1/5 7/14 naïve bayes – if 358 machine learning jasman pardede 5 tabel likelihood play tennis persentasi yes no 9/14 5/14 tabel 6. likelihood wind bermain tennis tabel likelihood play tennis persentasi yes no wind weak 6/9 2/5 8/14 strong 2/9 4/5 6/14 8/14 6/14 berdasarkan tabel 3 tabel 6 diperoleh bahwa: a. luang bermain tennis p(yes) = 9/14 b. luang bermain tennis p(no) = 5/14 c. luang sunny p (sunny) = 5/14 d. luang overcast p(overcast) = 4/14 e. luang rain p(rain) = 5/14 f. luang bermain tennis sunny p(sunny|yes) = 3/9 g. luang bermain tennis sunny p(sunny|no) = 2/5 h. luang bermain tennis overcast p(overcast|yes) = 4/9 i. luang bermain tennis overcast p(overcast|no) = 0/5 j. luang bermain tennis rain p(rain|yes) = 2/9 k. luang bermain tennis rain p(rain|no) = 3/5 l. luang bermain tennis rain p(mild|yes) = 4/9 m. luang bermain tennis rain p(mild|no) = 2/5 n. luang bermain tennis rain p(high|yes) = 3/9 o. luang bermain tennis rain p(high|no) = 4/5 p. luang bermain tennis rain p(weak|yes) = 6/9 q. luang bermain tennis rain p(weak|no) = 2/5 p(bermain tennis = yes | x) = p(bermain\_tennis=yes) \* p(outlook = rain | bermain\_tennis= yes) \*p (temperatur = mild | bermain\_tennis= yes) \* p (humidity = high | bermain\_tennis= yes) \* p (wind = weak | bermain\_tennis= yes) = (9/14) \* (3/9) \* (4/9) \* (3/9) \* (6/9) = 0.02116402 naïve bayes – if 358 machine learning jasman pardede 6 p(bermain tennis = no | x) = p(bermain\_tennis=no) \* p(outlook = rain | bermain\_tennis= no) \*p (temperatur = mild | bermain\_tennis= no) \* p (humidity = high | bermain\_tennis= no) \* p (wind = weak | bermain\_tennis= no) = (5/14) \* (3/5) \* (2/5) \* (4/5) \* (2/5) = 0.02743 jadi, berdasarkan hasil perhitungan luang bermain tennis diperoleh hasil luang (ya) bermain tennis = 0.01411 < luang (tidak) bermain tennis = 0.02743 1.41% < 2.74%, disimpulkan bermain tennis kondisi outlook rain, temperature mild, humidity high, wind weak bermain tennis. contoh 2: tabel 7. data sorang langgar rambu lintas warna tipe curi (x1) (x2) (x3) (x4) merah sport domestik ya merah sport domestik merah sport domestik ya kuning suv domestik kuning sport import ya kuning suv import kuning suv import ya kuning suv domestik merah suv import merah sport import ya tabel 7 atas, data mobil langgar rambu lintas atribut warna, tipe, asal. mengelompokkan mobil warna merah, tipe suv, domestik. tentu probabilitas pelanggaran lintas probabilitas langgar rambu lintas, tentu persen mobil langgar persen mobil melanggar, tentu mobil warna merah, tipe suv, domestik langgar lintas tidak? penyelesaian: tabel 8. frekuensi warna curi ya warna merah 3 2 kuning 2 3 naïve bayes – if 358 machine learning jasman pardede 7 tabel 9. likelihood warna likelihood curi persentasi ya warna merah 3/5 2/5 5/10 kuning 2/5 3/5 5/10 5/10 5/10 tabel 10. frekuensi tipe langgar ya tipe sport 4 1 suv 1 4 tabel 11. likelihood tipe likelihood langgar persentasi ya tipe sport 4/5 1/5 5/10 suv 1/5 4/5 5/10 5/10 5/10 tabel 12. frekuensi langgar ya domestik 2 3 import 3 2 tabel 13. likelihood likelihood langgar persentasi ya domestik 2/5 3/5 5/10 import 3/5 2/5 5/10 5/10 5/10 sehingga: a. luang langgar lintas, p(ya) = 5/10 = 0.5 b. luang langgar lintas, p(tidak) = 5/10 = 0.5 c. luang merah langgar lintas, p(merah| ya) = 3/5 = 0.6 d. luang suv langgar lintas, p(suv|ya) = 1/5 = 0.2 e. luang domestik langgar lintas, p(domestik|ya) = 2/5 = 0.4 f. luang merah langgar lintas, p(merah| tidak) = 2/5 = 0.4 g. luang suv langgar lintas, p(suv|tidak) = 4/5 = 0.8 h. luang domestik langgar lintas, p(domestik|tidak) = 3/5 = 0.6 i. luang langgar lintas, p(melanggar=ya|x) = p(ya) \* p(merah|ya) \* p(suv|ya) \* p(domestik|ya) = 0.5 \* 0.6 \* 0.2 \* 0.4 = 0.024 = 2.4% naïve bayes – if 358 machine learning jasman pardede 8 j. luang langgar lintas, p(melanggar =tidak|x) = p(tidak) \* p(merah|tidak) \* p(suv| tidak) \* p(domestik| tidak) = 0.5 \* 0.6 \* 0.8 \* 0.6 = 0.144 = 14.4%. jadi, berdasarkan hasil perhitungan pelanggaran lintas hasil pelanggaran (tidak) > pelanggaran (ya) 14.4% > 2.4% disimpulkan mobil warna merah, tipe suv, domestik langgar rambu lintas. contoh 3: hasil pengamatan sebelumnya, diperoleh penggunaan listrik dipengaruhi tanggung keluarga, luas rumah, pendapatan bulan, daya listrik digunakan, perlengkapan dimiliki. data hasil pengamatan berikut: tabel 14. data penggunaan listrik no tanggung keluarga luas rumah pendapatan/ daya listrik perlengkapan dimiliki penggunaan listrik 1 2 3 standar rendah 4 standar 5 rendah 6 rendah 7 8 9 standar 10 standar 11 12 13 14 15 16 standar 17 standar 18 19 naïve bayes – if 358 machine learning jasman pardede 9 no tanggung keluarga luas rumah pendapatan/ daya listrik perlengkapan dimiliki penggunaan listrik 20 rendah 21 standar 22 standar 23 24 25 26 rendah rendah 27 standar 28 29 standar 30 31 rendah rendah 32 rendah 33 rendah rendah 34 35 rendah 36 standar 37 standar rendah 38 rendah rendah 39 rendah 40 41 standar 42 43 44 45 46 standar rendah 47 48 49 rendah 50 rendah 51 52 standar 53 rendah naïve bayes – if 358 machine learning jasman pardede 10 no tanggung keluarga luas rumah pendapatan/ daya listrik perlengkapan dimiliki penggunaan listrik 54 55 standar 56 57 rendah 58 tentu correctly incorrectly classified instance data penggunaan listrik diberikan. penyelesaikan: 1. menghitung probabilitas class penggunaan listrik: kelas penggunaan listrik 3, yaitu: tinggi, sedang, rendah. adalah: 28, 21, 9. dinyatakan tabel 15. probabilitas (tinggi) = jumlah\_tinggi/total\_data =∑\*+,-\_/0123045678869:6;<, =>?> = 0.4828 tabel 15. probabilitas penggunaan listrik 2. menghitung probabilitas bersyaratnya: ini, atribut bersayarat 5 (lima), yaitu: a. tanggung keluarga b. luas tanah c. pendapatan d. daya listrik’ e. perlengkapan dimiliki. perhitungan numerik luang bersyarat penggunaan listrik syarat tanggung sebegai berikut: atribut tanggung memiliki 3 kelas, yaitu: banyak, sedang, sedikit. pengguna listrik syarat tanggung dinyatakan tabel 16. naïve bayes – if 358 machine learning jasman pardede 11 tabel 16. probabilitas tanggung nilai 10 pengguna listrik tanggung 10 keluarga. nilai 8 pengguna listrik tanggung 8 keluarga. pengguna listrik tanggung 3 keluarga. total pengguna listrik (10+8+3) = 21. probabilitas pengguna listrik tanggung 0.4762. probabilitas kelas penggunaan listrik syarat kelas tanggungan. sama, probabilitas bersyarat diperoleh dinyatakan tabel berikut: naïve bayes – if 358 machine learning jasman pardede 12 3. pengujian metode naïve bayes penggunaan listrik hitung probabilitas penggunaan listrik syarat kategori kelas diberikan. pertama, tanggung keluarga (x1) = banyak, luas rumah (x2) = besar, pendapatan (x3) = besar, daya listrik (x4) = tinggi, perlengkapan dimiliki (x5) = sedang. p(penggunaan=rendah | x) = (p(jumlah tanggung = | penggunaan=rendah) \* p(luas tanah =besar | penggunaan=rendah) \* p(pendapatan = |penggunaan = rendah) \* p(daya listrik = | penggunaan = rendah) \* p(perlengkapan = | penggunaan = rendah)) \* p(penggunaan = rendah) = 0,5556 \*0,7778 \*0,4444\*0,5556\*0,2222\*0,15517241 =0,003679 - p(penggunaan=sedang|x) = 0,003950 - p(penggunaan=tinggi|x) = 0,009960 probabilitas tiga kelas luang besar 0.009960, prediksi sistem pengguna tinggi. berdasarkan hasil prediksi diperoleh data prediksi 36, 22. nilai correctly classified-nya 62.07%, incorrectly classified 37.93%. naïve bayes – if 358 machine learning jasman pardede 13 https://www.datacamp.com/community/tutorials/categorical-data https://pbpython.com/categorical-encoding.html