Stemming Results

# 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, bentuknya bulat, memiliki diameter 8 cm. naïve bayes asumsi hubung warna, bentuk, diameter menentukan buah appel. realitanya asumsi “hubungan bebas” independensi jarang terjadi. ciri utama klasifikasi naïve bayes asumsi kuat (naïf) independensi atribut kondisi/kejadian. atribut algoritma naïve bayes berkontribusi keputusan tara independen bebas atribut lainnya. algoritma naïve bayes efisiensi komputasi cocok domain. asumsi independensi 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, perhitungannya 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 banyaknya 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 bersayaratnya 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 tanggungannya 10 keluarga. nilai 8 pengguna listrik tanggungannya 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 nya 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