

VOL. 18 NO. 4 DESEMBER 2017

ISSN : 1411-3201

Jurnal Ilmiah

DASI

DATA MANAJEMEN DAN TEKNOLOGI INFORMASI



UNIVERSITAS
AMKOM
YOGYAKARTA

VOL. 18 NO. 4 DESEMBER 2017

ISSN:1411-3201

JURNAL
ILMIAH
DASI

**DATA MANAJEMEN DAN
TEKNOLOGI INFORMASI**



UNIVERSITAS
AMIKOM
YOGYAKARTA

VOL. 18 NO. 4 DESEMBER 2017
JURNAL ILMIAH
Data Manajemen Dan Teknologi Informasi

Terbit empat kali setahun pada bulan Maret, Juni, September dan Desember berisi artikel hasil penelitian dan kajian analitis kritis di dalam bidang manajemen informatika dan teknologi informatika. ISSN 1411-3201, diterbitkan pertama kali pada tahun 2000.

KETUA PENYUNTING

Abidarin Rosidi

WAKIL KETUA PENYUNTING

Heri Sismoro

PENYUNTING PELAKSANA

Emha Taufiq Luthfi

Hanif Al Fatta

Hartatik

Hastari Utama

STAF AHLI (MITRA BESTARI)

Jazi Eko Istiyanto (FMIPA UGM)

H. Wasito (PAU-UGM)

Supriyoko (Universitas Sarjana Wiyata)

Ema Utami (AMIKOM)

Kusrini (AMIKOM)

Amir Fatah Sofyan (AMIKOM)

Ferry Wahyu Wibowo (AMIKOM)

Rum Andri KR (AMIKOM)

Arief Setyanto (AMIKOM)

Krisnawati (AMIKOM)

ARTISTIK

Robert Marco

TATA USAHA

Nila Feby Puspitasari

PENANGGUNG JAWAB :

Rektor UNIVERSITAS AMIKOM YOGYAKARTA, Prof. Dr. M. Suyanto, M.M.

ALAMAT PENYUNTING & TATA USAHA

UNIVERSITAS AMIKOM YOGYAKARTA, Jl. Ring Road Utara Condong Catur Yogyakarta, Telp. (0274) 884201 Fax. (0274) 884208, Email : jurnal@amikom.ac.id

BERLANGGANAN

Langganan dapat dilakukan dengan pemesanan untuk minimal 4 edisi (1 tahun) pulau jawa Rp. 50.000 x 4 = Rp. 200.000,00 untuk luar jawa ditambah ongkos kirim.

VOL. 18 NO. 4 DESEMBER 2017

ISSN : 1411- 3201

JURNAL ILMIAH

DASI

DATA MANAJEMEN DAN TEKNOLOGI INFORMASI

UNIVERSITAS AMIKOM YOGYAKARTA

JURNAL ILMIAH

DASI

KATA PENGANTAR

Puji syukur kehadirat Tuhan Yang Maha Kuasa atas anugerahnya sehingga jurnal edisi kali ini berhasil disusun dan terbit. Beberapa tulisan yang telah melalui koreksi materi dari mitra bestari dan revisi redaksional dari penulis, pada edisi ini diterbitkan. Adapun jenis tulisan pada jurnal ini adalah hasil dari penelitian dan pemikiran konseptual. Redaksi mencoba selalu mengadakan pembenahan kualitas dari jurnal dalam beberapa aspek.

Beberapa pakar di bidangnya juga telah diajak untuk berkolaborasi mengawal penerbitan jurnal ini. Materi tulisan pada jurnal berasal dari dosen tetap dan tidak tetap Universitas AMIKOM Yogyakarta serta dari luar Universitas AMIKOM Yogyakarta.

Tak ada gading yang tak retak begitu pula kata pepatah yang selalu di kutip redaksi, kritik dan saran mohon di alamatkan ke kami baik melalui email, faksimile maupun disampaikan langsung ke redaksi. Atas kritik dan saran membangun yang pembaca berikan kami menghaturkan banyak terimakasih.

Redaksi

DAFTAR ISI

HALAMAN JUDUL.....	i
KATA PENGANTAR	ii
DAFTAR ISI.....	iii
Estimasi Data Missing Pada Dataset Penderita Dbd Menggunakan Metode Trend Moment.....	1-5
Ryan Putranda Kristianto (Informatika Universitas AMIKOM Yogyakarta)	
Rancang Sistem Direktori Berkas Arsip Siswa Sekolah Dasar.....	6-12
Moch Farid Fauzi (Sistem Informasi Universitas AMIKOM Yogyakarta)	
Perancangan Sistem Informasi Penjadwalan Program Kerja Penjaminan Mutu Universitas Amikom Dengan Metode Multilevel Feedback Queue.....	13-18
Ike Verawati ¹⁾ , Mulia Sulistiyono ²⁾ (¹⁾²⁾ Informatika Universitas AMIKOM Yogyakarta)	
Rancang Bangun Sistem Penjualan Untuk Minimarket.....	19-24
Ni Kadek Sukerti (Sistem Informasi STMIK STIKOM Bali)	
Implementasi Data Mining Untuk Menemukan Pola Asosiatif Data Tracer Study.....	25-33
Ferian Fauzi Abdulloh ¹⁾ , Kusnawi ²⁾ (¹⁾²⁾ Informatika Universitas AMIKOM Yogyakarta)	
Penentuan Kelayakan Penerima Bantuan Renovasi Rumah Warga Miskin Menggunakan Naïve Bayes.....	34-38
Bety Wulan Sari ¹⁾ , Donni Prabowo ²⁾ (^{1) 2)} Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta)	
Inovasi Sistem Ulangan Harian Online Bersama Antar Sekolah Berbasis Web.....	39-44
Ainul Yaqin ¹⁾ , Benadhed ²⁾ (¹⁾²⁾ Informatika Universitas AMIKOM Yogyakarta)	
Identifikasi Ikan Kering Berformalin Berbasis Histogram Dan Jaringan Syaraf Tiruan Backpropagation	45-50
Erni Seniwati ¹⁾ , Kumara Ari Yuana ²⁾ (¹⁾ Sistem Informasi Universitas AMIKOM Yogyakarta, ²⁾ Informatika Universitas AMIKOM Yogyakarta)	
Sistem Pendukung Keputusan Penentuan Hotel Menggunakan Metode Topsis.....	51-56
Ikmah ¹⁾ (Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta)	

Pemanfaatan Game Edukasi Berbasis Android Untuk Pembelajaran Anak Usia Dini Paud Sidoasih.....	57-66
Muhammad Tofa Nurcholis (Informatika Universitas AMIKOM Yogyakarta)	
Perbandingan Klasifikasi Pecemaran Air Sungai Dengan Metode Backpropagation Dan Naïve Bayes.....	67-71
Hartatik ¹⁾ , Andri Syafrianto ²⁾ , Wiwi Widayani ³⁾ (¹⁾ Manajemen Informatika Universitas AMIKOM Yogyakarta, ²⁾ Teknik Informatika STMIK EL-RAHMA, ³⁾ Sistem Informasi Universitas AMIKOM Yogyakarta)	

PERBANDINGAN KLASIFIKASI PECEMARAN AIR SUNGAI DENGAN METODE BACKPROPAGATION DAN NAÏVE BAYES

Hartatik¹⁾, Andri Syafrianto²⁾, Wiwi Widayani³⁾

¹⁾ Manajemen Informatika Universitas AMIKOM Yogyakarta

²⁾ Teknik Informatika STMIK EL-RAHMA

³⁾ Sistem Informasi Universitas AMIKOM Yogyakarta

email : hartatik@amikom.ac.id¹⁾, andrisyafrianto@gmail.com²⁾, wiji.w@amikom.ac.id³⁾

Abstraksi

Algoritma Backpropagation dan Naïve Bayes bisa digunakan untuk membangun model klasifikasi dari sekumpulan data input. Masing-masing algoritma memiliki pendekatan yang berbeda-beda dalam melakukan klasifikasi. Penelitian ini akan mencoba untuk membandingkan 2 algoritma klasifikasi yaitu Backpropagation dan Naive Bayes. Kedua algoritma akan diperbandingkan nilai keakuratannya dalam melakukan klasifikasi pencemaran air sungai. Arsitektur yang dibangun menggunakan dataset sebanyak 150 data dengan neuron input sebanyak 22 dan neuron output sebanyak 3. Pengujian dilakukan menggunakan *validation matrix* dengan 4 kondisi fold dan split yang berbeda. Hasilnya didapatkan algoritma Naive Bayes memiliki nilai rata-rata akurasi yang lebih baik yaitu 72,79% dibandingkan dengan algoritma Backpropagation yang hanya memiliki nilai akurasi 64,75%.

Kata Kunci :

Klasifikasi, Naive Bayes, Backpropagation

Abstract

Backpropagation and Naïve Bayes algorithms can be used to construct a classification model from a set of input data. Each algorithm has a different approach in classification. This study will try to compare two classification algorithms, namely Backpropagation and Naive Bayes. Both algorithms will be compared the value of accuracy in the classification of river water pollution. The architecture is built using a dataset of 150 data with 22 input neurons and 3 output neurons. The test is done using validation matrix with 4 different fold and split conditions. The results obtained Naive Bayes algorithm has a better average accuracy of 72.79% compared with Backpropagation algorithm which only has an accuracy of 64.75%.

Keywords :

Classification, Naive Bayes, Backpropagation

Pendahuluan

Penelitian mengenai klasifikasi air sungai telah dilakukan sebelumnya oleh penulis menggunakan Algoritma *Learning Vector Quantization* (LVQ) dan Algoritma Perceptron [1][2]. Permasalahan yang muncul, ternyata kedua algoritma belum memiliki nilai akurasi yang baik. Dengan jumlah data latih sebanyak 100, data uji sebanyak 50, learning rate sebesar 0,01, pengurangan learning rate sebesar 0,001 dan maksimal iterasi 20.000, Algoritma LVQ hanya menghasilkan nilai akurasi sebesar 76% [1]. Untuk penelitian yang kedua menggunakan algoritma Perceptron ternyata memiliki nilai akurasi yang tidak lebih baik daripada algoritma LVQ. Dengan jumlah data latih dan data uji yang sama yaitu 100 dan 50, nilai learning sebesar 0,05, epoch sebanyak 50000 dan threshold bernilai 0, Algoritma Perceptron memiliki nilai akurasi sebesar 75% [2].

Perbandingan kinerja dari klasifikasi dengan menggunakan 3 algoritma ANN yang lain yaitu *Neuro Fuzzy Network*, Backpropagation, dan SVM

(*Support Vector Machine*) juga telah dilakukan oleh Kim dkk [3]. Percobaan pertama menggunakan dataset iris milik Fisher-Anderson dengan 5 parameter neuron input. Hasilnya didapatkan nilai akurasi rata-rata algoritma SVM sebesar 94,45% ; Algoritma *Neuro Fuzzy Network* sebesar 97,335% dan Algoritma Backpropagation sebesar 95,33%. Percobaan yang kedua menggunakan dataset kanker payudara dari Wisconsin yang terdiri dari 9 neuron input. Hasilnya, didapatkan nilai akurasi rata-rata algoritma SVM sebesar 95,79% ; Algoritma *Neuro Fuzzy Network* sebesar 96,185% dan Algoritma Backpropagation sebesar 96,28%. Percobaan yang ketiga menggunakan dataset wine dengan 13 parameter neuron input. Hasilnya didapatkan nilai akurasi rata-rata algoritma SVM sebesar 97,183% ; Algoritma *Neuro Fuzzy Network* sebesar 98,876% dan Algoritma Backpropagation sebesar 90,44%. Dari ketiga percobaan tersebut didapatkan kesimpulan bahwa algoritma Backpropagation memiliki nilai akurasi yang lebih baik.

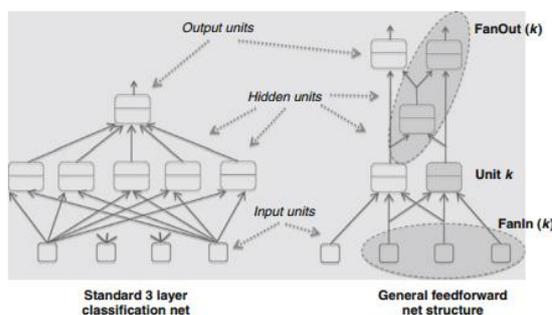
Perbandingan algoritma klasifikasi lainnya juga dilakukan oleh Chaurasia dkk[4]. Jurnal yang berjudul "Data Mining Approach to Detect Heart Dises" membandingkan 3 algoritma klasifikasi yaitu Naive Bayes, J48 Decision Tree dan Bagging Algorithm. Attribute yang dipakai untuk melakukan klasifikasi penyakit hati berjumlah 11 atribut dengan dataset yang digunakan sebagai data training berjumlah 294 data. Dengan pengujian menggunakan metode fold validation disimpulkan algoritma Naive Bayes memiliki nilai akurasi paling rendah yaitu 82,31%, kemudian disusul algoritma J48 dengan nilai akurasi 84,35% dan terakhir Algoritma Bagging dengan akurasi 94,50%.

Dari beberapa literatur, penulis mencoba untuk membandingkan 2 algoritma yaitu Naive Bayes dan Backpropagation. Data yang digunakan pada penelitian ini menggunakan 22 parameter kandungan fisika, kimia dan biologis.

Tinjauan Pustaka

Backpropagation atau disingkat BP adalah metode untuk melatih jaringan syaraf umpan balik ke depan. Implementasi spesifik BP adalah prosedur iteratif yang menyesuaikan parameter bobot jaringan sesuai dengan gradien pengukuran kesalahan [5]. Prosedur diimplementasikan dengan menghitung nilai kesalahan untuk setiap unit output, dan dengan menukar kembali nilai kesalahan melalui jaringan [5].

Jaringan syaraf umpan balik adalah fungsi matematis yang terdiri dari unsur penyusun "semilinear". Fungsi dibatasi oleh arsitektur jaringan feed-forward, dimana fungsi penyusunnya sesuai dengan node (sering disebut unit atau neuron buatan) dalam grafik, seperti pada Gambar 1. [5]



Gambar 1. Jaringan Backpropagation

Arsitektur jaringan feed-forward memiliki Sebuah struktur konektivitas yang merupakan grafik asiklik; Artinya, tidak ada loop tertutup. Dalam kebanyakan kasus, fungsi unit memiliki rentang yang terbatas seperti [0, 1]. Dengan demikian, peta jaringan RN ke [0, 1] M, dimana N adalah jumlah nilai input dan M adalah jumlah unit output. Misalkan FanIn (k) mengacu pada himpunan unit yang memberikan masukan ke unit k, dan biarkan FanOut (k)

menunjukkan himpunan Dari unit yang menerima input dari unit k.

Untuk prosedur BP, aktivitas masing-masing unit dihitung dalam dua tahap [5]:

1. Langkah linier: neuron input masing-masing dikalikan dengan parameter "bobot" independen, yang mana parameter "bias" ditambahkan; Setiap unit komputasi memiliki parameter bias tunggal, tidak bergantung pada unit lainnya. Misalkan jumlah ini dinotasikan x_k untuk unit k.
2. Langkah nonlinear: Aktivitas ak unit k adalah suatu fungsi nonlinier terdiferensiasi dari x_k . Favorit Fungsi adalah logika $\alpha = 1/(1+\exp(-x))$, karena memetakan kisaran $(-\infty, +\infty)$ sampai $[0,1]$ dan turunannya memiliki sifat yang kondusif bagi Implementasi BP.

$$a_k = f_k(x_k); \text{ where } x_k = b_k + \sum_{j \in \text{FanIn}(k)} w_{kj} s_j \quad (1)$$

Algoritma Naive bayes adalah suatu algoritma klasifikasi probalistik yang digunakan untuk menghitung frekuensi dan kombinasi nilai di kumpulan data yang diberikan [6].

Rumus naive bayes classifier didasarkan pada teorema bayes dan teorema probabilitas. Probabilitas dari suatu h dengan vector $e = \langle e_1, \dots, e_n \rangle$ untuk hipotesa h adalah [15] :

$$P(h_i | e_1 \dots e_n) = \frac{P(e_1 | h_i) \dots P(e_n | h_i) P(h_i)}{\sum_{k=1}^n P(e_1 | h_k) \dots P(e_n | h_k) P(h_k)} \quad (2)$$

Dimana, $P(h_i | e_i)$ adalah posterior probability (probabilitas akhir), sedangkan $P(h_i)$ adalah prior probability (probabilitas awal) yang berasosiasi dengan hipotesa h_i [6].

Pada naive bayes semua atribut akan memiliki kontribusi dalam pengambilan suatu keputusan, dimana setiap bobot dari atribut sama penting dan saling bebas. Apabila diberikan k atribut yang saling bebas, maka nilai probabilitas bisa dihitung dengan [7]:

$$P(e_1, \dots, e_k | h) = P(e_1 | h) \dots P(e_k | h) \quad (3)$$

Namun jika atribut ke-i memiliki bentuk diskret, maka $P(e_i | h)$ diestimasi sebagai frekuensi relatif dari sampel yang memiliki nilai e_i sebagai atribut kelas ke-i dalam suatu kelas h. Namun, jika atribut ke-i kontinue, maka nilai $P(e_i | h)$ diestimasi dengan fungsi densitas gauss [7].

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4)$$

prosentase split data latih dan data uji. Metode cross validation fold adalah suatu pengujian yang akan menguji secara bergantian data latih dan data uji. Di percobaan yang pertama sebagian data akan menjadi data latih dan sebagian lainnya menjadi data uji. Di percobaan yang kedua, data latih di percobaan pertama akan dipakai menjadi data uji dan sebaliknya data uji di percobaan yang pertama akan menjadi data latih di percobaan yang kedua. Pengujian Cross Validation fold dilakukan dengan pemberian 2 skema penilaian yaitu 10 dan 40, sedangkan pada pengujian split data dilakukan dengan melakukan pembagian data sebesar 20% (artinya 80% data menjadi data set dan 20% sisanya menjadi data uji) dan 50% (artinya 50% data menjadi data set dan 50% sisanya menjadi data uji). Nilai akurasi di pengujian cross validation fold maupun persentase split Algoritma Naive Bayes ditunjukkan pada gambar 4.

Data ke-	Cross Validation fold	Percentage Split (%)	Correctly Classified Instances (%)	Incorrectly Classified Instances (10%)
1	10	-	88.6667	11.3333
	40	-	90.6667	9.3333
	-	20	80.8333	19.1667
	-	50	84	16
2	10	-	82	18
	40	-	81.3333	18.6667
	-	20	70	30
	-	50	82.6667	17.3333
3	10	-	72.6667	27.3333
	40	-	73.3333	26.6667
	-	20	61.6667	38.3333
	-	50	76	24
4	10	-	40.6667	59.3333
	40	-	41.3333	58.6667
	-	20	64.1667	35.8333
	-	50	74.6667	25.3333

Gambar 4. Nilai akurasi Algoritma Naive Bayes

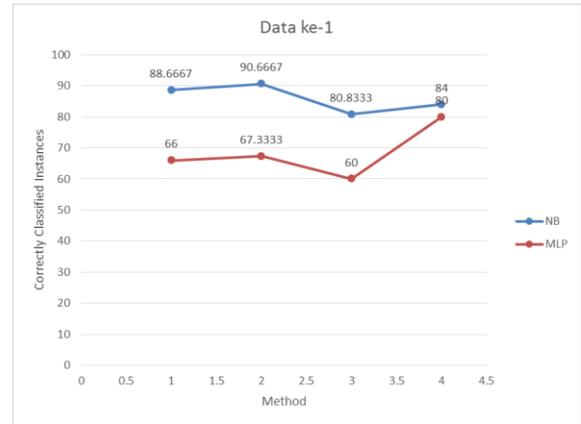
Sedangkan nilai akurasi hasil pengujian cross validation fold maupun persentase split untuk Algoritma Backpropagation ditunjukkan pada gambar 5.

Data ke-	Cross Validation fold	Percentage Split (%)	Correctly Classified Instances (%)	Incorrectly Classified Instances (10%)
1	10	-	66	34
	40	-	67.3333	32.6667
	-	20	60	40
	-	50	80	20
2	10	-	64	36
	40	-	56.6667	43.3333
	-	20	50.8333	49.1667
	-	50	68	32
3	10	-	58	42
	40	-	62	38
	-	20	52.5	47.5
	-	50	64	36
4	10	-	72	28
	40	-	72.6667	27.3333
	-	20	63.3333	36.6667
	-	50	78.6667	21.3333

Gambar 5. Nilai akurasi Algoritma Backpropagation

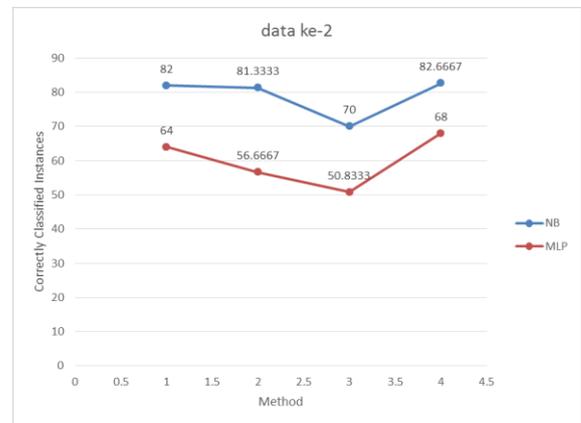
Perbandingan nilai akurasi dari klasifikasi pencemaran air sungai oleh Algoritma Naive Bayes dan Backpropagation pada keempat kelas ditunjukkan pada gambar 6 sampai dengan 9.

Pada percobaan yang pertama (ditunjukkan di gambar 5), algoritma naive bayes memiliki nilai akurasi yang lebih baik dengan nilai tertinggi yang bisa didapatkan yaitu 90,6667%.



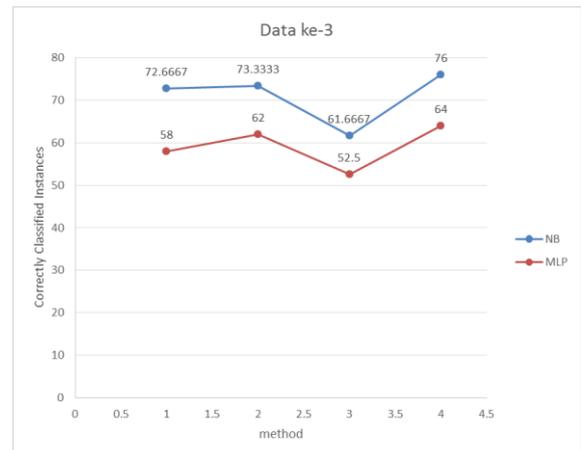
Gambar 5. Perbandingan akurasi Algoritma Backpropagation dan Naive Bayes di kelas data yang pertama

Begitu juga dengan percobaan untuk data yang kedua, algoritma naive bayes memiliki nilai akurasi yang baik dengan nilai tertinggi ada di angka 82,667%



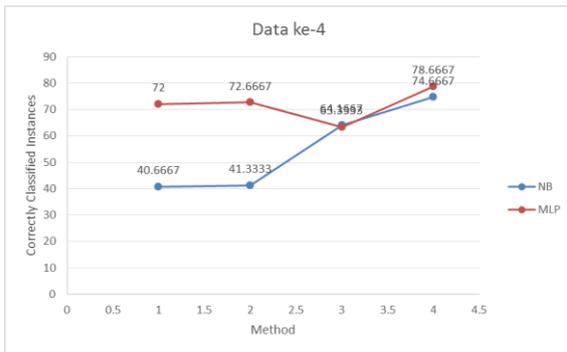
Gambar 6. Perbandingan akurasi Algoritma Backpropagation dan Naive Bayes di kelas data yang kedua

Pada data yang ketiga, naive bayes memiliki nilai akurasi tertinggi 76%.

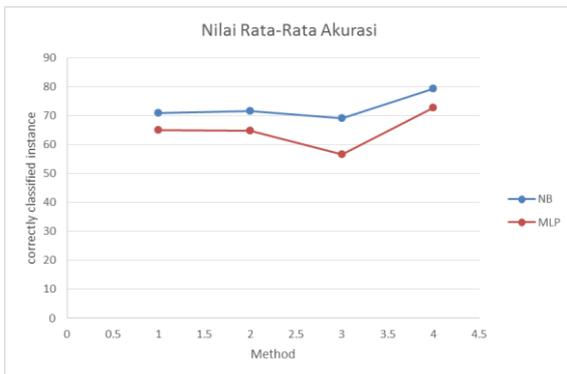


Gambar 7. Perbandingan akurasi Algoritma Backpropagation dan Naive Bayes di kelas data yang ketiga

Pada data yang keempat, nilai akurasi Backpropagation lebih baik dibandingkan algoritma naive bayes.



Gambar 8. Perbandingan akurasi Algoritma Backpropagation dan Naive Bayes di kelas data yang keempat



Gambar 9. Perbandingan akurasi Algoritma Backpropagation dan Naive Bayes di semua data

Kesimpulan dan Saran

Dari serangkaian pengujian menggunakan metode uji cross validation fold dan prosentase split data terhadap 4 kelas didapatkan suatu kesimpulan bahwa nilai akurasi rata-rata Naive Bayes memiliki nilai yang lebih besar yaitu :

Tabel 1 Nilai akurasi Naive Bayes dan Backpropagation

	Naive Bayes (%)	Backpropagation (%)
Folds : 10	71.00003	65
folds : 40	71.66665	64.66668
split : 20%	69.16668	56.66665
split : 50%	79.33335	72.66668

Daftar Pustaka

[1] M.R. Rahimi, Hartatik., Penerapan Algoritma Learning Vector Quantization Dalam Pengklasifikasian Tingkat Pencemaran Air Sungai, *Seminar Semnasteknomedia*, Vol. 4, No. 1, 2016.

[2] Hartatik, A. Fatkhurohman., Penentuan Kualitas Air Tanah Menggunakan Algoritma Perceptron, *Jurnal DASI*, Vol. 17, No. 4, 2016.

[3] Y.G.Kim, M.S. Jang, K.S. Cho, G.T. Park., Performance Comparison Between Backpropagation, Neuro-Fuzzy Network, and SVM., *International Computer Science Symposium in Russia CSR 2006: Computer Science – Theory and Applications*, 2016, pp 438-446

[4] V.Chaurasia, S.Pal., Data Mining Approach to Detect Heart Diseases., *International Journal of Advanced Computer Science and Information Technology (IJACSIT)* Vol. 2, No. 4, 2013, Page: 56-66

[5] P. Munro, *Backpropagation*, Encyclopedia of Machine Learning and Data Mining, pp 93-97, Springer, 14 April 2017

[6] T.R. Patil, S.S. Sherekar, Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification, *International Journal of Computer Science and Applications*, Vo. 6, No. 2, April 2013.

[7] S. Kusumadewi, Klasifikasi Status Gizi Menggunakan Naive Bayes Classification, *Jurnal CommIT*, Vol. 3, No. 1, pp. 6-11, Mei 2009