

Klasifikasi Pemahaman Santri Dalam Pembelajaran Kitab Kuning Menggunakan Algoritma Naive Bayes Berbasis Forward Selection

Adzhal Arwani Mahfudh¹, Hery Mustofa²

^{1,2} Universitas Islam Negeri Walisongo Semarang Indonesia

adzhal@walisongo.ac.id, herymustofa@walisongo.com

Abstract

The yellow book is a traditional book containing Islamic tradition taught in Islamic boarding schools, ranging from Arabic language structures (nahwu and shorof science), 'ulumul qur'an, hadith, aqidah, tasawuf / akhlaq, commentary, fiqh to social and social sciences (nahwu and shorof), 'ulumul qur'an, hadith, aqidah, tasawuf / akhlaq, interpretation, fiqh to social and social sciences (social science and society) mu'amalah). Also called the bare book because it does not have harakat (fathah, kasroh, dhammah, breadfruit) to be able to read and understand thoroughly takes a relatively long time. This study aims to obtain a classification model of the yellow book learning data in Islamic boarding schools. The method used in this research is forward selection as a preprocess in reducing data dimensions, eliminating irrelevant data and naive bayes that are useful for classifying data. The results of the yellow book learning data classification use attributes that have been classified based on its features and iterated on cross validation so as to produce precise accuracy. Based on the results of testing with two methods, testing with the Naive Bayes algorithm alone produces an accuracy of 96.02%, for the Naive Bayes algorithm based on forward selection produces an accuracy of 97.38%. There is an increase in accuracy with the addition of the selection feature.

Keywords: Data Mining, classification, forward selection, naive bayes, yellow book

Abstrak

Kitab kuning merupakan kitab tradisional yang mengandung dirasah islamiyah yang diajarkan pada pondok pesantren, mulai dari struktur bahasa arab (ilmu nahwu dan shorof), 'ulumul qur'an, hadits, aqidah, tasawuf/akhlaq, tafsir, fiqh sampai ilmu sosial dan kemasyarakatan (mu'amalah). Disebut juga dengan kitab gundul karena tidak memiliki harakat (fathah, kasroh, dhammah, sukun) untuk bisa membaca dan memahami secara menyeluruh dibutuhkan waktu yang relatif lama. Penelitian ini bertujuan untuk mendapatkan model klasifikasi dari data pembelajaran kitab kuning di pondok pesantren. Metode yang digunakan dalam penelitian ini adalah *forward selection* sebagai praproses dalam mengurangi dimensi data, menghilangkan data yang tidak relevan dan naive bayes yang berguna untuk mengklasifikasi data. Hasil dari klasifikasi data pembelajaran kitab kuning menggunakan atribut yang telah diklasifikasi berdasarkan fitur-fiturnya dan dilakukan iterasi pada *cross validation* sehingga menghasilkan akurasi yang tepat. Berdasarkan hasil pengujian dengan dua metode, pengujian dengan algoritma Naive bayes saja menghasilkan akurasi 96,02%, untuk algoritma Naive bayes berbasis forward selection menghasilkan akurasi 97,38%. Terdapat peningkatan akurasi dengan penambahan fitur seleksi.

Kata kunci: Data Mining, klasifikasi, forward selection, naive bayes, kitab kuning

I. PENDAHULUAN

Dalam proses klasifikasi data *training* atau data latih dibutuhkan untuk menentukan pola suatu data, data pembelajaran kitab kuning merupakan data pengalaman yang bisa digunakan pada algoritma klasifikasi. Sehingga algoritma klasifikasi dapat menggunakan data tersebut untuk mendapatkan pengetahuan dan menggolongkan pemahaman santri dalam pembelajaran kitab kuning berdasarkan parameter-parameter yang telah ditentukan. Sebagai salah satu sarana dalam menentukan seorang santri dapat dikatakan layak atau tidak, pondok pesantren telah menentukan parameter yang harus dicapai oleh seorang calon santri sebelum memulai pembelajaran kitab kuning, antara lain parameter yang digunakan adalah sebagai berikut, kemampuan Jurumiyah, Tasyrifiyah, kelancaran dalam membaca alqur'an, imlaq (penulisan alqur'an), penguasaan bahasa Arab serta label yang akan digunakan adalah kemampuan santri dalam pembelajaran kitab kuning. Penelitian ini memberikan kontribusi yang positif terhadap perkembangan pondok pesantren dalam menentukan santri yang memenuhi kriteria, karena ditemukan banyak santri yang telah mengikuti proses pembelajaran selama 2 tahun akan tetapi belum dapat mengikuti proses pembelajaran kitab kuning dikarenakan kemampuan santri serta seleksi yang kurang cermat untuk mengikuti pembelajaran kitab kuning.

Berdasarkan penelitian sebelumnya, khususnya dalam bidang pendidikan, para peneliti telah menggunakan berbagai macam algoritma data mining untuk menyelesaikan masalah klasifikasi, diantaranya Naive Bayes, Decision tree, C4.5 dan lain sebagainya. Berdasarkan penelitian yang telah dilakukan Naive Bayes (NB) merupakan algoritma klasifikasi yang sederhana, efisiensi, komputasi tinggi, dan akurasi klasifikasi yang baik terutama untuk data berdimensi tinggi [4]. NB merupakan algoritma klasifikasi berbasis probabilistik sederhana yang berdasarkan pada penerapan teorema Bayes dengan asumsi independensi yang kuat pada fitur, artinya bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama, Namun NB memiliki kekurangan yaitu atribut atau fitur independen sering salah dan hasil estimasi probabilitas tidak dapat berjalan optimal [5].

Bondhan Arya [8] dalam penelitiannya pada tahun 2015 melakukan klasifikasi kelulusan mahasiswa dengan menerapkan fitur seleksi Forward selection menggunakan algoritma Naive Bayes untuk menentukan atribut yang berpengaruh, penelitian ini menunjukkan bahwa algoritma Naive Bayes terbukti meningkat setelah dilakukan fitur seleksi terhadap data set yang digunakan. Dari penelitian tersebut metode forward selection dapat mengurangi dimensi data yang besar dan dapat membantu meningkatkan hasil akurasi

klasifikasi algoritma naive bayes sebesar 3,34%. Selain itu Wiwit supriyanti dkk [9] pada tahun 2016 melakukan penelitian mengenai perbandingan kinerja Algoritma C.45 dengan Naive Bayes untuk klasifikasi ketepatan pemilihan konsentrasi mahasiswa. Pada uji coba yang dilakukan menggunakan data dari universitas muhamadiyah surakarta. Penggabungan seleksi fitur forward selection pada algoritma C4.5 dan naive bayes menunjukkan peningkatan pada algoritma naive bayes sebesar 3,54% sedangkan pada algoritma C4.5 hanya meningkat sebesar 0,55%.

Pada penelitian yang dilakukan Yusuf S. Nugroho dan kawan-kawan [10], dimana dari penelitian itu membandingkan algoritma Naive bayes dengan algoritma C4.5 dan K-means dalam pemilihan jurusan pada sekolah menengah atas. Dari penelitian tersebut diperoleh hasil Precision sebesar 77,51%. Berdasarkan nilai yang diperoleh, metode naive bayes mendapatkan hasil yang terbaik dibandingkan dengan metode yang lain. Pada tahun 2016 Adhi Indra Irawan dan kawan kawan [11] melakukan penelitian dengan metode Naive Bayes dan AdaBoost dalam mengklasifikasikan penyakit ginjal kronik atau *chronic kidney disease* (CKD). Dari hasil perhitungan dengan *confusion matrix* didapatkan akurasi sebesar 0,95 untuk metode Naive Bayes.

Dari beberapa penelitian diatas dapat disimpulkan bahwa algoritma naive bayes dengan penambahan fitur

seleksi mampu melakukan klasifikasi terhadap berbagai macam data set. Dari hasil penelitian yang telah dilakukan oleh para peneliti sebelumnya klasifikasi pembelajaran kitab kuning sangat mungkin untuk menggunakan beberapa atribut seperti yang dijelaskan diatas. Penelitian ini dengan penelitian yang dilakukan Muhammad Al Fajri dkk sama sama menggunakan algoritma fitur seleksi Forward selection dan Naive Bayes. Namun dalam penelitian ini terdapat perbedaan yaitu pada dataset serta penggunaan atribut yang akan digunakan.

II. METODOLOGI

Metode yang diusulkan untuk melakukan klasifikasi pemahaman santri dalam belajar kitab kuning di pondok pesantren yaitu penerapan Forward selection pada algoritma Naive Bayes. Dataset akan dilakukan seleksi dengan metode Forward Selection untuk menyeleksi fitur terhadap atribut atribut yang kurang berpengaruh atau tidak relevan pada dataset dengan tujuan dapat meningkatkan akurasi. Selanjutnya dilakukan pengklasifikasian dengan algoritma naive bayes sebagai algoritma yang melakukan klasifikasi santri dalam belajar kitab kuning di pondok pesantren. Sedangkan untuk validasi menggunakan *10-fold crossvalidation*. Tahapan dan metode digambarkan seperti gambar dibawah:

1. Dalam pengolahan data awal, data melewati preprocessing yaitu data dipilah dan dibersihkan. Proses memilah data yaitu dengan

memilih atribut yang akan dipakai dalam menentukan klasifikasi santri dalam belajar kitab di pondok pesantren. Atribut yang tidak berpengaruh tidak dipakai dalam penelitian. Sedangkan pembersihan data dilakukan untuk menghilangkan data data yang inkonsistensi seperti missing value atau data yang tidak lengkap. Pada data yang missing value dilakukan penghapusan pada record. Setelah dibersihkan dan dipilah, data menjadi sebuah dataset baru untuk dijadikan data training dan testing dari atribut yang telah ditentukan

2. Kemudian pada tahapan selanjutnya, data diuji dengan dua metode yaitu data diuji hanya menggunakan algoritma Naive bayes, data diuji menggunakan naive bayes berbasis forward Selection.
3. Hasil dari perbandingan kedua metode tersebut masing masing akan menghasilkan sebuah model dan akurasi dari data yang telah diuji coba.
4. Model yang dihasilkan selanjutnya akan di evaluasi menggunakan confusion matrix. Dari hasil diagram Confusion Matrix tersebut akan diketahui hasil prediksi positive yang benar dan salah, serta hasil prediksi negative yang benar dan salah. Kemudian dilakukan proses validasi menggunakan metode *k-fold cross validation* untuk mengetahui rata-rata keberhasilan dari suatu metode, caranya yaitu melakukan perulangan dengan mengacak

atribut masukan sehingga metode tersebut teruji untuk beberapa data input yang di acak.

III. KERANGKA TEORI

Dalam pendidikan agama islam merujuk pada kitab-kitab tradisional yang berisi pelajaran agama islam yang diajarkan pada pondok – pondok pesantren mulai dari fiqih, aqidah, akhlak/tasawuf, tata bahasa arab, hadist ,tafsir,‘ullumul quran hingga pada ilmu sosial dan kemasyarakatan. Dikenal juga dengan kitab gundul karena memang tidak memiliki harakat, tidak seperti kitab Al quran. Oleh sebab itu, untuk bisa membaca kitab kuning tentang harfiah kalimat per kalimat agar bisa dipahami secara menyeluruh dibutuhkan waktu yang relaif lama. Untuk pembelajaran kitab kuning dibutuhkan penguasaan ilmu nahwu (*syntax*), ilmu shorof (*morfologi*), menerjemahkan kata per kata (*translate*) dan menerapkan teori susunan kalimat pada kitab kuning.

Data mining adalah proses menggunakan teknik statistik, perhitungan, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar [14]. Keluaran dari data mining dapat dipakai untuk memperbaiki pengambilan keputusan dimasa depan [15].

Beberapa ahli lain berpendapat, data mining merupakan serangkaian proses yang digunakan untuk menggali nilai tambah dari kumpulan

data yang berupa pengetahuan dan tidak diketahui secara manual. Istilah data mining dengan *Knowledge Discovery in Databases* atau bisa disingkat dengan KDD sering digunakan secara bergantian untuk menjelaskan proses penggalian informasi yang tersembunyi dalam suatu basis data yang besar. Kedua konsep tersebut memiliki konsep yang berbeda, akan tetapi memiliki keterkaitan yang sama [16].

Salah satu fungsi data mining adalah prediksi, prediksi banyak digunakan untuk menentukan keputusan sesuai pengetahuan baru yang didapat dari pengolahan data terdahulu menggunakan perhitungan suatu algoritma. Teknik klasifikasi dapat diterapkan dalam berbagai bidang misalnya bidang kesehatan [17], bidang pendidikan [18], bidang teknik bangunan [19], serta banyak digunakan dalam bidang lain. Atribut inilah yang akan dicari dari data baru dengan dasar atribut lain pada data terdahulu. Banyaknya atribut dapat mempengaruhi performa suatu algoritma [20]. Masalah klasifikasi pada dasarnya adalah sebagai berikut [21].

1. Masalah klasifikasi berangkat dari data training yang tersedia.

2. Data training akan diolah dengan menggunakan algoritma klasifikasi

Masalah klasifikasi berakhir dengan dihasilkannya sebuah pengetahuan yang direpresentasikan dalam bentuk diagram, aturan atau pengetahuan.

Naive Bayes atau NB adalah pengklasifikasian statistik yang didasarkan pada teorema Bayes yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. NB terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database besar. NB pertama kali dikemukakan oleh Revered Bayes sejak tahun 1702-1761. Naive Bayes disebut juga *Bayesian Classification* merupakan metode pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan dari satu class. Naive Bayes berdasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa decision tree dan neural network. Selain itu, Naive Bayes terbukti memiliki akurasi kecepatan yang tinggi saat diaplikasikan ke dalam database yang besar [27]. Berikut bentuk umum Algoritma *Naïve Bayes*:

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

Dimana :

X = data dengan kelas yang belum diketahui

H = Hipotesa data X merupakan suatu kelas spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X (posterior probability)

P(H) = probabilitas hipotesis H (prior Probability)

Peluang bersyarat atribut kategorikal dinyatakan dalam bentuk [28]:

$$P(A_i|C_j) = \frac{|A_{ij}|}{N_{Cj}}$$

Dimana $|A_{ij}|$ adalah jumlah contoh training dari kelas A_i yang menerima nilai C_j . Jika hasilnya adalah nol, maka menggunakan pendekatan yang dinyatakan bentuk[29]:

$$P(A_i|C_j) = \frac{n_c + n_{equiv} P}{n + n_{equiv}}$$

Dimana n adalah total dari jumlah record dari kelas C_j . N_c adalah jumlah contoh training dari kelas A_i yang menerima nilai C_j . N_{equiv} adalah nilai konstanta dari ukuran sampel yang ekuivalen. P adalah peluang estimasi prior, $P=1/k$ dimana k adalah jumlah kelas dalam variable target.

Peluang bersyarat atribut kontinu dinyatakan dalam bentuk[28]:

$$P(A_i|C_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp\left[-\frac{(A_i - \mu_{ij})^2}{2(\sigma_{ij})^2}\right] \tag{2.4}$$

Seleksi fitur adalah salah satu tahapan praproses yang berguna terutama dalam mengurangi dimensi data, menghilangkan data yang tidak relevan serta meningkatkan hasil akurasi [30]. Tujuan dari seleksi fitur adalah untuk mengurangi tingkat kompleksitas dari sebuah algoritma klasifikasi, meningkatkan akurasi dari algoritma klasifikasi dan mampu mengetahui fitur fituryang paling berpengaruh terhadap tingkat akurasi [9]. Seleksi fitur digunakan untuk memberikan karakteristik dari data. Seleksi fitur merupakan salah satu penelitian yang banyak

dilakukan di berbagai bidang. Salah satu metode fitur seleksi yaitu forward selection. Forward selection termasuk dalam teknik wrapper approach. Banyak yang menggunakan Teknik seleksi fitur wrapper approach karena kinerjanya lebih baik dibandingkan dengan *filter*, *embedded* dan *hybrid approach* dan dapat menghasilkan *subset fit* yang lebih sesuai pada klasifikasi sehingga hasil akurasi yang di dapatkan lebih tinggi [31] [32]. Tahapan tahapan di dalam metode forward selection dijelaskan seperti dibawah ini [9]:

1) Membuat model dengan meregresikan variabel respon Y dengan setiap predictor. Kemudian memilih model yang memiliki nilai R^2 tertinggi. Jika model tersebut adalah yang memuat predictor X_a , yaitu:

$$Y = b_0 + b_a X_a$$

2) Meregresikan variabel respon Y , dengan predictor X_a ditambah dengan setiap predictor selain X_a dan predictor lain. Kemudian dipilih model yang nilai R^2 nya tertinggi. Jika mengandung tambahan prediktor X_b , yaitu model

$$Y = b_0 + b_a X_a + b_b X_b$$

Prediktor terpilih X_b berarti mempunyai $F_{sequensial}$ tertinggi. Formula $F_{sequensial}$ untuk X_b adalah

$$F_{seq} = R(\beta_b / \beta_0 \beta_a) \dots MSE/db$$

Nilai $F_{sequensial}$ untuk X_b juga dapat diperoleh dengan cara mengkuadratkan nilai statistik uji T prediktor X_b .

3) Proses diulang sampai mendapatkan $F_{\text{sequensial}} > F_{\text{in}}$. nilai $F_{\text{in}} = F(1, \nu, \alpha_{\text{in}})$, sehingga model terbaik yang dipilih adalah model yang tidak mempunyai predictor dengan $F_{\text{sequensial}} < F_{\text{in}}$.

Confusion matrix adalah alat visualisasi yang biasa digunakan pada supervised learning. Tiap kolom pada matrix adalah contoh kelas prediksi, sedangkan tiap baris mewakili kejadian di kelas yang sebenarnya. Tabel 2.2 adalah contoh tabel confusion matrix yang menunjukkan klasifikasi dua kelas

Tabel 2.2 Evaluasi Confusion Matrix

Classification	Predicted Class		
		Class = Yes	Class = No
Observed Class	Class = Yes	A (true positive – tp)	B (false negative – m)
	Class = No	C (false positive – fp)	D (true negative – tn)

Keterangan :

True Positive (tp) = proporsi positif pada dataset yang diklasifikasikan positif

True Negative (tn) = proporsi negatif pada dataset yang diklasifikasikan

Negative False Positive (fp) = proporsi negative pada dataset yang diklasifikasikan positif

False Negative (fn) = proporsi negatif pada dataset yang diklasifikasikan sebagai negatif

Berikut adalah persamaan model Confusion Matrix:

a. Nilai Akurasi adalah jumlah prediksi yang benar. Dihitung dengan formulasi:

$$acc = \frac{tp+tn}{tp+tn+fp+fn}$$

b. *Sensitivity* sebagai pembanding tp terhadap tupel yang positif, dihitung dengan formulasi :

$$sensitivity = \frac{tp}{tp+fn}$$

c. *Specificity* sebagai pembanding tn terhadap tupel yang negatif, dihitung dengan formulasi:

$$specitivity = \frac{tn}{tn+fb}$$

d. *PPV(Positive Predictive Value)* adalah perbandingan kasus dengan hasil diagnosa positif, dihitung dengan formulasi:

$$ppv = \frac{tp}{tp+fp}$$

e. *NPV (Negative Prediktive Value)* adalah perbandingan kasus dengan hasil diagnosis negatif, yang dihitung dengan formulasi:

$$npv = \frac{tn}{tn+fn} \quad (2.12)$$

IV PEMBAHASAHAN

Tujuan dari penelitian ini yaitu untuk meningkatkan nilai akurasi yang dihasilkan oleh algoritma klasifikasi naive bayes dengan mengurangi beberapa atribut pada dataset menggunakan algoritma fitur seleksi berupa forward selection dan menyeimbangkan data menggunakan algoritma adaboost. Dataset diolah menggunakan perbandingan dari hasil yang didapatkan antara algoritma naive bayes, naive bayes

berbasis forward selection dan penerapan algoritma adaboost pada Nb+Fs. Perbandingan dari ketiga pengujian dapat dilihat pada tabel dibawah:

Tabel 4.11 Perbandingan Akurasi

Metode	Akurasi
Naive bayes	96,02%
Naive bayes + FS	97,38%

Dapat dilihat pada tabel diatas, perbandingan ketiga eksperimen yang telah dilakukan dengan data yang sama, terlihat dari hasil eksperimen menggunakan model adaboost pada algoritma Nb+Fs dengan 1030 data training menghasilkan akurasi yang baik. Dengan menggunakan metode evaluasi 10-fold cross validation yang digunakan untuk mengetahui rata-rata keberhasilan nilai akurasi yang didapat valid atau tidak, serta mengukur performa dari algoritma didapatkan hasil akurasi rata-rata

dari 10 kali iterasi pengujian sebesar 97,57%

V. KESIMPULAN

Kesimpulan yang didapatkan berdasarkan hasil penerapan algoritma naive bayes berbasis forward selection untuk klasifikasi pemahaman santri dalam pembelajaran kitab kuning sebagai berikut :

1. Hasil pengujian dari kinerja algoritma Naive bayes berbasis forward selection dengan 1030 data menghasilkan akurasi lebih baik sebesar 97,38 %.
2. Penerapan forward selection sebagai metode seleksi fitur terbukti bahwa metode tersebut dapat meningkatkan akurasi dari naive bayes, dimana atribut yang terlibat semula pada dataset berkurang dengan hanya memilih atribut yang berpengaruh saja dan memperbaiki klasifikasi yang salah.

DAFTAR PUSTAKA

- Kamaru, R. Jurnal Pembaharuan Pendidikan Islam (JPPI). 2014.
- Thoriqussu'ud, Muhammad. MODEL-MODEL PENGEMBANGAN, 225–239. Jurnal Ilmu tarbiyah “At-Tajdid”. 2012.
- Mukroji. METODE TAMYIZ (Sebuah Formulasi Teori Nahwu Shorof Quantum, II(1), 161–184. Jurnal Kependidikan. 2014.
- Wu, J., Pan, S., Zhu, X., Cai, Z., Zhang, P., & Zhang, C. (2015). Expert Systems with Applications Self-adaptive attribute weighting for Naive Bayes classification. *EXPERT SYSTEMS WITH APPLICATIONS*, 42(3), 1487–1502.
- Zaidi, N. A., Carman, M. J., & Webb, G. I. (2013). Alleviating Naive Bayes Attribute Independence Assumption by Attribute Weighting, 14, 1947–1988.
- Quan, Z., G. Lin-gang, W. Chong-jun, W.-j. and C. Shi-fu, "Using An Improved C4.5 for Imbalanced Dataset of Intrusion," National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093
- Yu, L. H, "Feature Selection for High Dimensional Data: A FastCorrelation-Based Filter Solution," 2003
- Bondhan, A.P. Penerapan Fitur Seleksi Forward Selection Menggunakan Algoritma Naive Bayes Untuk Menentukan Atribut yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa”. “Tugas Akhir Teknik Informatika Universitas Dian Nuswantoro Semarang, 2015.
- Supriyati, W., Kusriani, Amborowati, A., "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa," *INFORMA Politeknik Indonusa Surakarta*, vol. 1, pp. 2442- 7942, 2016.
- Nugroho, Y. S. Graha, I. "Data Mining Jasa Pengiriman Titipan Kilat di PT Citra Van dengan Metode Decision Tree," Naskah Publikasi Program Studi Teknik Informatika Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta, Juni 2014.
- Irawan, A.I., Saptono, R., Doewes, A., “Implementation of Naive Bayes Classifier Method And Adaboost Algorithm For Prediction Of Chronic Kidney Disease”.*Digilib.uns.ac.id*.2016.
- Wibowo, K., Sfenrianto, Nainggolan, K.,”Kalsifikasi Gangguan Motorik Anak Menggunakan Naive Bayes serta Optimasi dengan PSO dan Adaboost”. Vol. 1. No. 1 E-ISSN: 2527-4864, Agustus 2015.
- Hongbo S, Xiaoyong Lv. “The Naïve Bayesian Classifier Learning Algorithm based on Adaboost and Parameter Expectations” Third International Joint Conference on Computational Science and Optimization. 978-0-7695-4030-6/10 \$26.00 © 2010 IEEE DOI 10.1109/CSO.2010.161. 2010.
- Sutrisno. (2013). Penerapan Data Mining Pada Penjualan Menggunakan Metode Clustering Study Kasus Pt . Indomarco, 1–11.
- Y.Y.W, “Perbandingan Performa Algoritma Decision Tree C5.0, CART dan CHAID”, Seminar Nasional Aplikasi Teknologi Informasi 2007 (SNATI 2007), pp. 1907-5022, 16 Juni 2007.

- Maharani, W., "Klasifikasi Data Menggunakan JST Backpropagation," Seminar Nasional Informatika 2009 (semnasIF 2009), vol. 2009, no. semnasIF, pp. 25–31, 2009
- Christobel, Angeline, Sivaprakasam, D.r., "An Empirical Comparison of Data Mining Classification Methods" 3 (2): 24–28. 2011.
- Ragab, Abdul Hamid M., Amin Y. Noaman, Abdullah S. Al-Ghamdi, and Ayman I. Madbouly. "A Comparative Analysis of Classification Algorithms for Students College Enrollment Approval Using Data Mining." Proceedings of the 2014 Workshop on Interaction Design in Educational Environments - IDEE '14. New York, New York, USA: ACM Press, 106–13. doi:10.1145/2643604.2643631. 2014.
- Ashari, Ahmad, Iman Paryudi, and A Min Tjoa. "Performance Comparison between Naïve Bayes , Decision Tree and K-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool" 4 (11): 33–39. 2013.
- Prasetyo, Eko. "Data Mining Konsep Dan Aplikasi Menggunakan Matlab". Yogyakarta: Andi Offset. 2012.
- Susanto, Sani, Suryadi, Dedi. Pengantar Data Mining: Menggali Pengetahuan Dari Bongkahan Data. Yogyakarta: Andi Offset. 2010.
- Hermawati, F.A. "Data Mining Bab I Teknik Data Mining, Yogyakarta: Cv Andioffset, 2013.
- Kusrini, Emha, t.l. "Algoritma Data Mining", Yogyakarta, 2009.
- Sijabat, A., "Penerapan Data Mining Untuk Pengolahan Data Siswa dengan Metode Decision Tree(Studi Kasus: Yayasan Kristen Andreas)," informasi dan teknologi ilmiah (INTI). Vol V, no. 3, pp. 2339-210X, Pebruari 2015.
- Prastya, F.A., "Penerapan Algoritma C4.5 untuk Prediksi Jurusan Siswa SMAN 3 Rembang, "Skripsi Teknik Informatika Universitas Dian Nuswantoro Semarang, 2015.
- Wiradhaningsih, K. P., Ratnawati, D. E.. , "Penerapan ALgoritma Decision Tree C5.0 untuk Peramalan Forex," Proceeding Universitas Brawijaya Malang.
- Fitriani, I. R. (2007). Peningkatan Metode Naive Bayes Classification Untuk Penentuan Tingkat Keganasan Kanker Payudara.
- Gorunescu, F. (2011). Data Mining.
- Larose, D. T. (2006). METHODS AND.
- Yu, L. H, "Feature Selection for High Dimensional Data: A FastCorrelation-Based Filter Solution," 2003.