G-Tech : Jurnal Teknologi Terapan
Vol. 8, No. 4, Oktober 2024, hlmn. 2895-2902
E-ISSN: 2623-064X | P-ISSN: 2580-8737

Optimalisasi Akurasi Algoritma C4.5 dengan Metode *Adaptive Boosting* Memprediksi Siswa dalam Menerima Dana Pendidikan

Wahyu Aji Tri Riswandhana[1]✉, Alva Hendi Muhammad[2]

[1, 2] Teknik Informatika, PJJ MTI, Universitas Amikom Yogyakarta, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 02-10-2024
Direvisi : 17-10-2024
Diterima : 21-10-2024

ABSTRAK

Pentingnya peningkatan akurasi bagi institusi pendidikan dalam memprediksi pemberian bantuan dana pendidikan. Untuk membuat keputusan tentang siapa yang layak mendapatkan dana pendidikan. Pengolahan data penerima bantuan dapat diolah menjadi informasi. Studi ini bertujuan untuk meningkatkan akurasi algoritma C4.5 dengan menggunakan adaboost untuk menentukan apakah siswa layak mendapatkan dana bantuan pendidikan atau tidak dengan melakukan perbandingan hasil sebelum dan sesudah penerapan adaboost. Prediksi kelayakan siswa dalam memperoleh dana bantuan pendidikan dengan pohon keputusan. Dataset dikumpulkan dari siswa SMK Muhammadiyah 1 Ngoro dengan jumlah 414 siswa, digunakan untuk penelitian ini. Hasil penelitan menunjukkan peningkatan akurasi 2.69% dengan penerapan algoritma C4.5 memiliki akurasi 77.31% sedangkan akurasi dengan penerapan adaboost mencapai 80%.

Kata Kunci:

Adaboost; Algoritma C4.5; Mesin *Learning*;

Keywords **:**

Adaboost, Algoritym C4.5;Machine Learning

ABSTRACT

The importance of increasing accuracy for educational institutions in predicting the provision of educational financial assistance. To make decisions about who deserves education funding. Data processing on aid recipients can be processed into information. This study aims to improve the accuracy of the C4.5 algorithm by using adaboost to determine whether students deserve to receive educational assistance funds or not by comparing the results before and after implementing adaboost. Predicting students' eligibility for obtaining educational assistance funds using decision trees. The dataset was collected from 414 students of SMK Muhammadiyah 1 Ngoro, used for this research. The research results show an increase in accuracy of 2.69% with the application of the C4.5 algorithm which has an accuracy of 77.31%, while the accuracy with the application of Adaboost reaches 80%.

Corresponding Author **:**
Wahyu Aji Tri Riswandhana
Teknik Informatika, PJJ MTI, Universitas Amikom Yogyakarta, Indonesia
Jl. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, DIY
Email: wahyu.aji.tri.riswandhana@gmail.com

PENDAHULUAN

Bantuan Siswa Miskin (BSM) berdasarkan Undang-Undang nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional yang mengamanatkan, bahwa setiap peserta didik berhak mendapatkan biaya pendidikan bagi mereka yang orang tuanya tidak mampu. Melihat situasi adanya beberapa siswa yang masih kesulitan dalam membayar biaya sekolah, SMK Muhammadiyah 1 Ngoro Jombang menganggarkan dana untuk membantu siswa yang masih

This is an open access article under the CC BY license

kesulitan membayar biaya sekolah. Maka dilakukan sistem seleksi untuk melihat siapakah diantara para siswa yang layak mendapatkan bantuan dana pendidikan. Untuk menangani itu dibuatlah sebuah perhitungan yang bersifat prediktif.

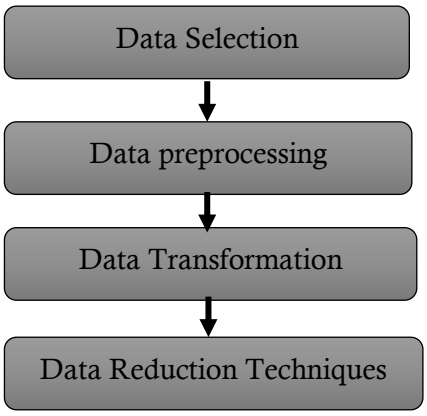
Sebagai upaya nyata untuk mengatasi masalah tersebut, maka diusulkan penerapan sistem seleksi calon penerima bantuan pendidikan. Metode pengambilan keputusan yang digunakan adalah metode *adaptive boosting*. Metode ini digunakan untuk menentukan nilai pembobotan setiap kriteria, yang kemudian dilakukan proses peringkat untuk menentukan alternatif terbaik dari sejumlah alternatif (Syaifuddin et al., 2022). Kriteria yang digunakan dalam sistem ini yaitu; jumlah penghasilan orang tua, kartu KPS,KIP dan yatim

Peningkatan akurasi dari 81% menjadi 85% dalam penelitian Mawaddah dan Pranoto (2023) menunjukkan bahwa penggunaan metode Adaboost berhasil dalam meningkatkan kinerja algoritma C4.5 pada kasus klasifikasi nilai mahasiswa. Adaboost sebagai metode boosting mampu memperbaiki hasil klasifikasi dengan memberikan bobot lebih pada data yang sulit diklasifikasikan oleh model dasar. Proses iteratif Adaboost ini cocok digunakan dalam konteks klasifikasi nilai mahasiswa untuk meningkatkan ketepatan prediksi.

Optimasi algoritma tersebut memungkinkan seleksi penerima bantuan menjadi lebih efisien, berpotensi mengurangi biaya operasi dan sumber daya manusia (Iskandar, 2022). Proses automatisasi yang lebih cepat dan tepat mengurangi kebutuhan evaluasi manual yang cenderung memakan waktu dan sumber daya (Rismawati et al., 2023). Secara keseluruhan, algoritma C4.5 dapat dioptimalkan untuk meningkatkan akurasi, efisiensi, dan keadilan dalam menetukan siswa yang patut menerima bantuan, memungkinkan penyandang dana memastikan bahwa bantuan yang mereka berikan tepat pada mereka yang layak berdasarkan data yang ada (Mosshananza, 2021).

METODE PENELITIAN

Dalam penelitian ini, algoritma C4.5 akan digunakan untuk memprediksi sejauh mana siswa berpotensi memperoleh dana Pendidikan (Kholil et al., 2022). Data yang diperoleh dari penelitian ini seperti pada tabel 1. Oleh karena itu, pemilihan metode perhitungan yang lebih dapat diandalkan adalah dengan menggunakan program algoritma C4.5 (Saputri et al., 2022). Langkah-langkah dari penelitian adalah sebagai berikut:



Gambar 1. Diagram Alur penelitian

1. Data *Selection*, pada gambar 1 data yang paling penting yang akan membantu dalam penentuan siswa mana yang dapat memperoleh dana pendidikan akan dipilih. Dalam data ini, informasi kasus akan dipilih pada titik ini karena akan menjadi bagian dari proses pengolahan data operasional. Dalam data ini, atribut keputusan yang dipilih adalah kartu BSM, pendapatan bulanan dari orang tua atau wali murid, Kartu Indonesia Pintar dan Yatim atau Yatim Piatu.
2. Data *preprocessing* / Data *cleaning*, metode ini menunjukkan semua isi yang memiliki atribut yang hilang atau kosong tidak digunakan.

3. Data *Transformation* pada tahap ini, data tersebut kemudian diubah modelnya menjadi tipe yang sesuai untuk proses pemrosesan.
4. Data *Reduction Techniques* merupakan berbagai metode dan teknik yang digunakan dalam rangka mengurangi jumlah atau kompleksitas data, seperti: teknik pemilihan fitur, pengelompokan, atau analisis faktor. Dengan menggunakan tahap pengurangan data, fitur-fitur yang tidak diperlukan akan dihilangkan dari skala data (Sunarto et al., 2023).

HASIL DAN PEMBAHASAN

Dataset

Data dalam penelitian ini dikumpulkan di lokasi analisis langsung, yaitu SMK Muhammadiyah 1 Ngoro Jombang. Kemudian penulis menemui staf bagian operator data sekolah, setelah itu wawancarai operator sekolah untuk mendapatkan data dan informasi tentang jumlah siswa di sekolah sebagai bahan penelitian seperti pada tabel 1 . Algoritma C4.5 menyatakan siapa yang layak mendapat dana pendidikan. Langkah pertama yang penulis lakukan ketika berhasil mendapat dataset yaitu mencari kriteria-kriteria apa saja yang kemudian nanti akan diseleksi alagoritma C4.5.

Tabel 1. Dataset Siswa SMK Muhammadiyah 1 Ngoro

Nama	Penerima KPS	Rata-Rata Penghasila...	Penerima KIP	Yatim/ Piatu/...	Rekomendasi
ADAM ILHAM FIRDAUS	Tidak	Rp3,996.000	Tidak	Tidak	Tidak
AGUSTIAN NUGRAHA	Tidak	Rp815.000	Tidak	Tidak	Tidak
ALIEFIAN NANDA WIJAYA	Tidak	Rp784.000	Tidak	Tidak	Ya
ANDRI TRI WIJAYA	Tidak	Rp1,560.000	Ya	PIATU	Ya
ARDIAN DESCA DAFFAREL	Tidak	Rp750.000	Tidak	Tidak	Ya
AURIEL BRAHMANTYA SEPTIANDIKA	Tidak	Rp1,940.000	Tidak	Tidak	Ya

Pemodelan

Pada penelitian ini menggunakan Algoritma C4.5 dan *Adaptive Boosting* sebagai optimalisasi dari pemodelan

1. Algoritma C4.5

   Algoritma C4.5 adalah algoritma yang digunakan untuk membangun pohon keputusan (*decision tree*) dalam data mining (Mortara et al., 2023). Algoritma ini merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser 3*), yang digunakan untuk klasifikasi (Mawaddah & Pranoto, 2023).

   C4.5 bekerja dengan membagi dataset berdasarkan atribut yang paling membedakan kelas target menggunakan perhitungan Gain Ratio. Algoritma ini digunakan dalam pengambilan keputusan berdasarkan data, seperti klasifikasi dan pengenalan pola (Sunarto et al., 2023).

   Entropy (H(S)): Entropi digunakan untuk mengukur tingkat ketidakpastian dalam satu set data S. Entropi dihitung dengan rumus:

$$H(S) = -\sum_{i=1}^{n} p_1 \, log_2(p_i) \quad (1)$$

   Dimana p_i adalah probabilitas kemunculan kelas i dan n adalah jumlah kelas

   Information Gain digunakan untuk memilih atribut terbaik untuk memisahkan data (Rijal et al., 2023). Information Gain dihitung sebagai perbedaan antara entropi sebelum dan sesudah dataset dibagi menurut atribut A:

$$Gain\ (S,A) = H(S) - \sum_{v \epsilon V(A)} \frac{|Sv|}{|S|} H(Sv) \quad (2)$$

Dimana :

- S adalah dataset sebelum pemisahan
- Sv adalah subset dari S untuk setiap nilai v dari atribut A
- $V(A)$ adalah himpunan nilai yang mungkin dari atribut A

Split Information digunakan untuk menghitung seberapa bagus atribut tersebut memisahkan data. *Split Information* dihitung dengan rumus:

$$Gain\ (S,A) = H(S) - \sum_{v \epsilon V(A)} \frac{|Sv|}{|S|} log_2 \frac{|Sv|}{|S|} \quad (3)$$

Gain Ratio adalah perbandingan antara *Information Gain* dan *Split Information*. Rumusnya adalah:

$$Gain\ Ratio\ (A) = \frac{Gain\ (S,A)}{SplitInfo\ (A)} \quad (4)$$

Gain Ratio digunakan untuk menghindari atribut dengan banyak nilai yang cenderung memiliki Information Gain tinggi tetapi tidak memberikan pemisahan yang baik.

2. Adaptive Boosting (adaboost)

Adaboost ini adalah metode untuk meningkatkan aturan yang benar dengan menggabungkan beberapa aturan yang kurang akurat dan lemah (Quang et al., 2023). Pembelajaran ensemble yang dikenal dengan Adaboost atau peningkatan adaptif Adaboost adalah pembelajaran ensemble yang digunakan dalam perbaikan formula (Madyatmadja et al., 2023). AdaBoost tetap menjadi salah satu optimasi paling sering dipakai dan diterapkan serta aturan pencegah awal untuk meningkatkan kinerja dan aturan klasifikasi penyempurnaan aturan yang sangat bervariasi. Apabila jenisnya benar-benar berlainan, model akan membantu (Saifullah et al., 2023).

A. Bobot awal untuk setiap data yang mana setiap data awalnya diberikan bobot yang sama.

$$w_i^{(1)} = \frac{1}{N}, \quad \text{untuk } i = 1, 2, \ldots, N \quad (5)$$

Dimana N adalah jumlah total data

B. Menghitung error dari *weak learner* dengan I merupakan fungsi indikator yang bernilai 1 jika prediksi salah, dan) jika benar.

$$\epsilon_t = \frac{\sum_{i=1}^{N} w_i^{(t)} \cdot I(y_i \neq h_t(x_i))}{\sum_{i=1}^{N} w_i^{(t)}} \quad (6)$$

C. Hitung bobot model ini menentukan seberapa besar pengaruh weak learner $h_t(x)$ dalam model akhir

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) \quad (7)$$

D. Update bobot data untuk data yang diklasifikasikan dengan benar, kurangi bobotnya, dan untuk data yang salah, tingkatkan bobotnya:

$$w_i^{(t+1)} = w_i^{(t)} \cdot \exp\left(-\alpha_t \cdot y_i \cdot h_t(x_i)\right) \quad (8)$$

Setelah itu normalisasi bobot:

$$w_i^{(t+1)} = \frac{w_i^{(t+1)}}{\sum_{i=1}^{N} w_i^{(t+1)}} \quad (9)$$

E. Weak learners di kombinasi menjadi strong learner Prediksi akhir dari model adalah kombinasi linier dari semua weak learners dengan bobot α_t dimana $H(x)$ merupakan prediksi akhir dari model dan T yaitu jumlah iterasi

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t \cdot h_t(x)\right) \quad (10)$$

Evaluasi

Setelah model C4.5 diterapkan bersama AdaBoost, evaluasi kinerjanya dapat dilakukan dengan berbagai metrik evaluasi. Model di evaluasi dengan menggunakan data uji dengan metrik sebagai berikut :

Tabel 2. Definisi Metrik Evaluasi Model Machine Learning

Metrik Evaluasi	Keterangan
Akurasi	Persentase prediksi yang benar dari keseluruhan prediksi
Presisi	Proporsi prediksi positif yang benar dari total prediksi positif
Recall	Proporsi kejadian positif yang benar-benar teridentifikasi dari total kejadian positif

Tabel 3. Akurasi Algoritma C4.5 Sebelum Optimalisasi

accuracy: 77.36% +/- 4.53% (micro average: 77.31%)

	true Tidak	true Ya	class precision
pred. Tidak	1	6	14.29%
pred. Ya	21	91	81.25%
class recall	4.55%	93.81%	

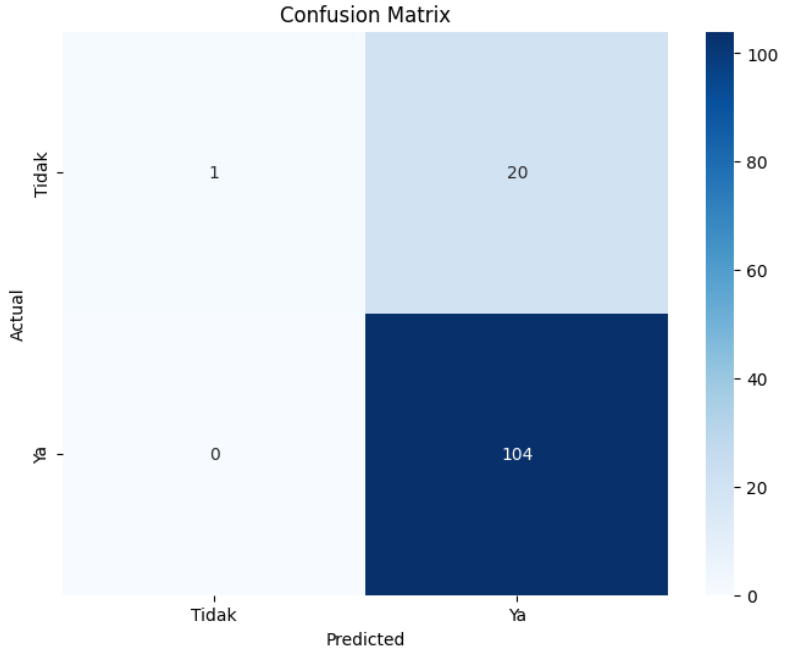
Dengan mempertimbangkan kinerja algoritma C4.5 pada tabel 3, *class precision* tampak tinggi pada beberapa dataset yang terdefinisi dengan baik dan bersih. Namun, kinerjanya mulai menurun ketika dihadapkan pada data yang bising dan kompleks (Okfalisa et al., 2022). Sebagai contoh, untuk dataset dengan banyak tumpang tindih antar kelas, C4.5 mungkin mengalami kesulitan dalam generalisasi, yang mengakibatkan peningkatan tingkat kesalahan. Ini secara efektif dibatasi oleh kenyataan bahwa algoritma ini cenderung *overfit* pada data latih

Tabel 4. Akurasi C4.5 dengan Adaptive Boosting

accuracy: 80.00% +/- 5.43% (micro average: 80.00%)

	true Tidak	true Ya	class precision
pred. Tidak	0	0	0.00%
pred. Ya	24	96	80.00%
class recall	0.00%	100.00%	

Setelah diterapkan, AdaBoost dengan C4.5 Tabel 4 menunjukkan peningkatan dalam akurasi. AdaBoost meningkatkan proses pengambilan keputusan dengan mengidentifikasi dan memprioritaskan kasus-kasus yang sulit dijelaskan, sehingga mengurangi kesalahan secara keseluruhan (Qois Syafi, 2022). Penelitian komparatif dilakukan dan ditemukan bahwa penggunaan AdaBoost dengan C4.5 menunjukkan peningkatan akurasi yang sangat mencolok, terutama untuk dataset yang lebih sulit dan bising. Hal ini disebabkan oleh proses penyesuaian ulang bobot iteratif AdaBoost di mana pengklasifikasi berikutnya cenderung pada instance yang telah salah klasifikasi (Crismayella et al., 2023).



Gambar 2. Confusion Matrix Algoritma C4.5 dengan Adaboost

Pada gambar 1 menunjukan hasil confusion Matrik dari hasil evaluasi model algoritma C4.5 dengan menggunakan adaptive boosting

KESIMPULAN DAN SARAN

Kesimpulan

Menggunakan algoritma C4.5 bersama dengan AdaBoost dalam menentukan siswa yang memperoleh dana bantuan pendidikan dapat meningkatkan kinerja prediksi. Evaluasi kinerja dapat dilakukan dengan berbagai metrik seperti akurasi, precision dan recall. Dengan pendekatan ini, diharapkan bahwa hasil klasifikasi lebih akurat dan dapat membantu dalam pengambilan keputusan terkait pemberian bantuan pendidikan secara lebih adil dan tepat. Dalam klasifikasi siswa memperoleh dana bantuan pendidikan SMK Muhammadiyah 1 Ngoro algoritma adaboost mampu meningkatkan akurasi algoritma C4.5 dengan peningkatan sebesar 2.69% .

Saran

Menambahkan analisis perbandingan dengan metode klasifikasi lainnya untuk menunjukkan keunggulan pendekatan yang diusulkan serta melakukan validasi tambahan menggunakan dataset yang berbeda untuk menguji generalisasi dari model klasifikasi yang dikembangkan .

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada kepala serta Staf SMK Muhammadiyah 1 Ngoro yang telah memberikan izin dan dukungan yang sangat berharga

selama proses pembuatan jurnal ini. Ucapan terima kasih juga kepada keluarga yang terus mendoakan dan kepada semua orang yang telah membantu dan mendukung penulis dalam menyusun dan menyelesaikan jurnal ini.

REFERENSI

Crismayella, Y., Satyahadewi, N., & Perdana, H. (2023). Comparison of Adaboost Application to C4.5 and C5.0 Algorithms in Student Graduation Classification. *Pattimura International Journal of Mathematics (PIJMath)*, *2*(1), 07–16. https://doi.org/10.30598/pijmathvol2iss1pp07-16

Iskandar, A. (2022). Sistem Pendukung Keputusan Kelayakan Penerima Bantuan Dana KIP Kuliah Menggunakan Metode ROC-EDAS. *Building of Informatics, Technology and Science (BITS)*, *4*(2). https://doi.org/10.47065/bits.v4i2.2265

Kholil, M. M. arij, Alzami, F., & Soeleman, M. A. (2022). AdaBoost Based C4.5 Accuracy Improvement on Credit Customer Classification. *2022 International Seminar on Application for Technology of Information and Communication: Technology 4.0 for Smart Ecosystem: A New Way of Doing Digital Business, ISemantic 2022*. https://doi.org/10.1109/iSemantic55962.2022.9920463

Madyatmadja, E. D., Sianipar, C. P. M., Wijaya, C., & Sembiring, D. J. M. (2023). Classifying Crowdsourced Citizen Complaints through Data Mining: Accuracy Testing of k-Nearest Neighbors, Random Forest, Support Vector Machine, and AdaBoost. *Informatics*, *10*(4), 84. https://doi.org/10.3390/informatics10040084

Mawaddah, S., & Pranoto, W. J. (2023). Optimasi Algoritma C4.5 Menggunakan Metode Adaboost Classification Pada Klasifikasi Nilai Mahasiswa Studi Kasus: Universitas Muhammadiyah Kalimantan Timur Optimizing the C4.5 Algorithm Using Adaboost Classification Method for Student Grade Classification Case Study: Muhammadiyah University of East Kalimantan. In *Jurnal Sains Komputer dan Teknologi Informasi e-issn* (Vol. 6, Issue 1).

Mortara, A. A., Permatasari, M., Desiani, A., Andriani, Y., & Arhami, M. (2023). Perbandingan Algoritma C4.5 dan Adaptive Boosting dalam Klasifikasi Penyakit Alzheimer. *Jurnal Teknologi Dan Informasi*, *13*(2). https://doi.org/10.34010/jati.v13i2.10525

Mosshananza, H. (2021). IMPLEMENTATION OF SMART INDONESIA PROGRAMME (PIP) AT SMP NEGERI 11 BANDAR LAMPUNG. *DEMOKRASI*, *1*(2). https://doi.org/10.36269/dmkr.v1i2.483

Okfalisa, Budianita, E., Yuliani, R., Suanmali, L., Megawati, Rusnedy, H., & Saktioto. (2022). Forecasting company financial distress: C4.5 and adaboost adoption. *Engineering and Applied Science Research*, *49*(3), 300–307. https://doi.org/10.14456/easr.2022.31

Qois Syafi, M. (2022). Increasing Accuracy of Heart Disease Classification on C4.5 Algorithm Based on Information Gain Ratio and Particle Swarm Optimization Using Adaboost Ensemble. *Journal of Advances in Information Systems and Technology*, *4*(1). https://journal.unnes.ac.id/sju/index.php/jaist

Quang, V. D., Viet, H. H., Long, V. H., & Khang, T. D. (2023). An Improved AdaBoost Algorithm for Highly Imbalanced Datasets in the Co-Authorship Recommendation Problem. *IEEE Access*, *11*, 89107–89123. https://doi.org/10.1109/ACCESS.2023.3306783

Rijal, S., Cakranegara, P. A., Ciptaningsih, E. M. S. S., Pebriana, P. H., Andiyan, A., & Rahim, R. (2023). Integrating Information Gain methods for Feature Selection in Distance Education Sentiment Analysis during Covid-19. *TEM Journal*, *12*(1). https://doi.org/10.18421/TEM121-35

Rismawati, R., Hakim, L., & Nababan, R. (2023). The Effectiveness of The Smart Indonesia Program at The Education Level of Junior High School in Plered District, Purwakarta Regency (Case Study at Junior High School 2 Plered). *EduLine: Journal of Education and Learning Innovation*, *3*(2). https://doi.org/10.35877/454ri.eduline1802

Saifullah, H., Adiwijaya, N. O., & Dharmawan, T. (2023). Optimization of Machine Learning Algorithms with Bagging and AdaBoost Methods for Stroke Disease Prediction. In *Applied Medical Informatics Original Research* (Vol. 45, Issue 2).

Saputri, N. D., Khalid, K., & Rolliawati, D. (2022). Komparasi penerapan metode Bagging dan Adaboost pada Algoritma C4. 5 untuk prediksi Penyakit Stroke. *SISTEMASI: Jurnal Sistem Informasi*, *11*(3).

Sunarto, A., Kencana, P. N., Munadjat, B., Dewi, I. K., Abidin, A. Z., & Rahim, R. (2023). Application of Boosting Technique with C4.5 Algorithm to Reduce the Classification Error Rate in Online Shoppers Purchasing Intention. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *14*(2), 01–11. https://doi.org/10.58346/JOWUA.2023.I2.001

Syaifuddin, S., Solikhin, S., & Riyanto, E. (2022). Aplikasi untuk Mencari Kelayakan Siswa Penerima Bantuan Pendidikan dengan Metode Simple Additive Weighting (Studi Kasus : SMK NU Ma'arif Kudus). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *9*(1). https://doi.org/10.25126/jtiik.2021864023