

Optimasi Akurasi Klasifikasi Pada Prediksi Smoke Detection dengan Menggunakan Algoritma Adaboost

Amin Nur Rais^{1,*}, Warjiyono²

¹Fakultas Teknik dan Informatika, Teknologi Komputer, Universitas Bina Sarana Informatika, Jakarta, Indonesia

²Fakultas Teknik dan Informatika, Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika, Tegal, Indonesia

Email: ^{1,*}amin.arv@bsi.ac.id, ²warjiyono.wrj@bsi.ac.id

Email Penulis Korespondensi: amin.arv@bsi.ac.id

Submitted: 21/11/2022; Accepted: 24/12/2022; Published: 31/12/2022

Abstrak—Permasalahan kebakaran menjadi salah satu ancaman bagi alam dan lingkungan. Untuk menanggulangi kejadian kebakaran maka dibuat dan dikembangkan alat pendeteksi asap yang dikombinasikan dengan perangkat IoT sehingga data kejadian dapat terekam dengan baik dimana data hasil rekaman akan digunakan sebagai acuan peningkatan akurasi deteksi dini.. Peningkatan akurasi pada alat pendeteksi asap sehingga dapat dikombinasikan dengan teknologi kecerdasan buatan. Pada penelitian ini mengusulkan optimasi prediksi dengan menggunakan algoritma adaboost yang dikombinasikan dengan algoritma klasifikasi naïve bayes dengan matrix pengukuran berdasarkan pada akurasi, recall, dan presisi. Hasil penelitian menunjukkan bahwa dengan menggunakan algoritma adaboost dapat meningkatkan nilai akurasi yang dihasilkan dengan nilai 0,987. Jika melihat evaluasi dari sisi presisi juga menunjukkan penggunaan algoritma adaboost dapat meningkatkan nilai presisi dengan nilai 0,971. Tetapi pada evaluasi recall menunjukkan tanpa adaboost mendapatkan nilai lebih baik dengan nilai 0,995

Kata Kunci: Adaboost; Naïve Bayes; Data Mining; Smoke Detection

Abstract—The problem of fire is a threat to nature and the environment. To deal with fire incidents, a smoke detector was created and developed in combination with an IoT device so that incident data can be recorded properly where the recorded data will be used as a reference for increasing the accuracy of early detection. Increasing the accuracy of smoke detectors so that they can be combined with artificial intelligence technology. This research proposes prediction optimization using the adaboost algorithm combined with the naïve Bayes classification algorithm with a measurement matrix based on accuracy, recall, and precision. The results showed that using the adaboost algorithm could increase the resulting accuracy value with a value of 0.987. If you look at the evaluation from the precision side, it also shows that the use of the adaboost algorithm can increase the precision value with a value of 0.971. But the recall evaluation showed that without boost it got a better score with a value of 0.995

Keywords: Adaboost; Naïve Bayes; Data Mining; Smoke Detection

1. PENDAHULUAN

Permasalahan pada insiden kebakaran menimbulkan ancaman bagi alam dan lingkungan disekitarnya. Untuk mencegah terjadinya kebakaran, teknik penanggulangan pendeteksian kebakaran mendapat perhatian yang luas dari berbagai pihak. Alat pendeteksian kebakaran terdapat 2 jenis, alat berbasis sensor dan berbasis penglihatan [1]. Wang mengungkapkan bahwa sebagian besar sensor asap menggunakan intensitas cahaya sebagai indikator asap untuk memicu alarm kebakaran, tetapi terkadang alarm kebakaran dipicu oleh aerosol non-api karena karakteristik konsentrasinya yang serupa [2]. Untuk menanggulangi hal tersebut, dikembangkan detektor asap yang digabungkan dengan perangkat Internet of Things (IoT) digunakan untuk mengumpulkan data kejadian kebakaran yang akan dijadikan dasar dalam pengembangan detektor kebakaran berbasis IoT dan Artificial Intelligence (AI) [3].

Perangkat pendeteksi asap berdasarkan fusi sensor AI untuk menentukan alarm kebakaran atau tidak dan menampung data kejadian [1]. Data yang terkumpul melalui pendeteksi asap kemudian diolah untuk memaksimalkan pendeteksi asap berbasis IoT dan AI. Proses awal dari pengolahan data dilakukan dengan proses preprocessing. Pada tahapan pengolahan data, proses pembersihan data dilakukan untuk menghilangkan nilai yang hilang dan mengurangi atribut yang tidak digunakan pada proses pemodelan [4].

Data yang telah didapatkan dari proses perekaman dengan menggunakan perangkat pendeteksi asap kemudian diolah menggunakan metode data mining. Data mining adalah proses penambangan untuk menemukan informasi baru dengan cara mencari pola atau aturan tertentu dari kumpulan data yang sangat besar [5]. Data mining juga digunakan untuk menggani nilai tambah berupa pengetahuan dan informasi yang selama ini belum diketahui jika dilakukan pemrosesan secara manual [6]. Dalam pengolahan dataset pendeteksian asap, digunakan teknik klasifikasi data mining. Teknik klasifikasi pada data mining adalah proses untuk menemukan fungsi dan model yang dapat membedakan dan atau menjelaskan konsep pada kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek [7].

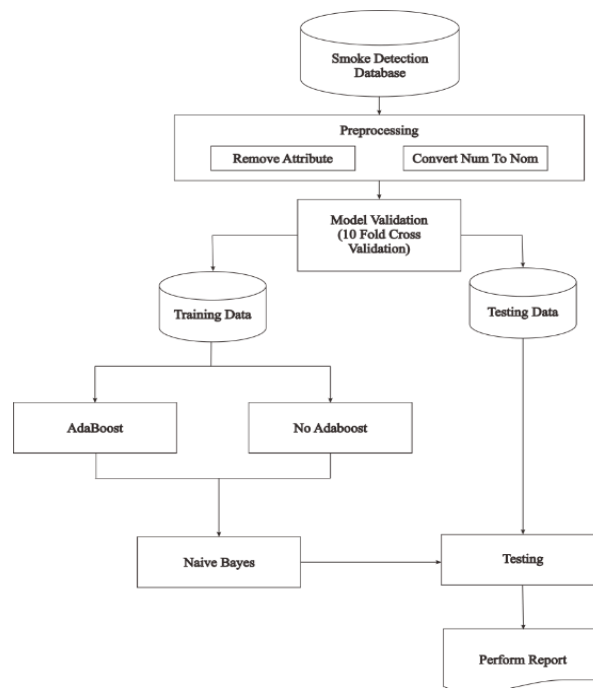
Pada penelitian yang dilakukan oleh elvira dan muhammad tentang smoke and fire detection base on convolutional neural network menghasilkan bahwa diketahui bahwa CNN memberikan kinerja yang baik dalam deteksi kebakaran dan asap [8]. Sedangkan Dileep dkk pada penelitiannya menyatakan bahwa dengan fitur yang diberikan kepada SVM untuk mendiskriminasi data secara lebih menyeluruh dan memberikan deteksi asap yang lebih akurat dengan eksperimen dilakukan dengan kumpulan data tolok ukur sehingga menunjukkan bahwa pendekatan yang telah dilakukan dapat bekerja secara efektif [6].

Pada penelitian ini akan dilakukan proses klasifikasi berdasarkan data yang telah terekam oleh perangkat pendeteksi asap. Hasil luaran pada penelitian ini akan menginformasikan model pengklasifikasian dengan algoritma naïve bayes dengan algoritma adaboost sehingga menghasilkan nilai akurasi, recall, dan presisi yang terbaik.

2. METODOLOGI PENELITIAN

2.2 Tahapan Penelitian

Tahapan penelitian secara garis besar dibagi menjadi 4 bagian : pengumpulan data, preprocessing, pengujian model, dan performa report [9]. Pada penelitian ini sebagaimana digambarkan pada gambar 1 dimulai dengan proses pengumpulan data menggunakan data dengan menggunakan data sekunder yang besar dari kaggle repositori. Dari kaggle repositori menggunakan dataset klasifikasi smoke detection. Dataset smoke detection terdiri dari 15 atribut dan 2 kelas (0/1). Jumlah data yang ada pada dataset smoke detection terdiri dari 62.629. Data yang telah diperoleh kemudian dilakukan proses *preprocessing* dengan cara menghilangkan atribut yang tidak diperlukan dan mengkonversi tipe data *numtonom*.



Gambar 1. Model Penelitian

Setelah dilakukan tahap *preprocessing*, kemudian dilanjutkan dengan mensplit data menggunakan model 10 fold validation untuk training data dan testing data. Pada training data, dilakukan percobaan menggunakan algoritma Naïve Bayes, kemudian dilakukan juga sebelum algoritma Naïve Bayes, diberi algoritma Adaboost sebagai optimasi hasil evaluasinya.

Hasil pengujian dihitung evaluasi sebagai laporan hasil percobaan dan digunakan sebagai alat ukur optimasi dari adaboost dengan melihat sisi akurasi, presisi, dan recall.

2.2 Naïve Bayes

Algoritma Naive Bayes adalah pengklasifikasi probabilitas sederhana yang menghitung sekumpulan probabilitas dengan menghitung frekuensi dan kombinasi nilai dalam kumpulan data yang diberikan [10]. Algoritma ini menggunakan teorema Bayes dan mengasumsikan bahwa semua variabel adalah independen dengan mempertimbangkan nilai dari variabel kelas [11]. Asumsi independensi bersyarat ini jarang berlaku dalam aplikasi dunia nyata, tetapi algoritma cenderung belajar dengan cepat dalam berbagai masalah klasifikasi terkontrol. Teorema bayes dapat dijelaskan sebagai berikut :

$$P(C | x) = \frac{P(C)P(x|C)}{P(x)} \tag{1}$$

Dimana:

C : Kelas

$P(x|C)$: Kemungkinan posterior x pada kondisi kelas C

$P(C)$: Kemungkinan kelas C tanpa memandang bukti apapun

$P(x)$: Kemungkinan posterior x tanpa memandang kelas/bukti lain

2.3 Adaboost

Adaboost merupakan ensemble learning yang sering digunakan pada algoritma boosting yang bisa dikombinasikan dengan classifier algoritma yang lain untuk meningkatkan performa klasifikasi [12]. Jadi AdaBoost juga suatu meta-algorithm yang dapat digunakan bersamaan dengan banyak algoritma pembelajaran lain untuk meningkatkan kinerjanya. AdaBoost bersifat adaptif, dimana classifiers berikutnya dibangun untuk mendukung data-data yang mengalami kesalahan klasifikasi oleh classifier sebelumnya. AdaBoost sensitif terhadap data yang noisy dan outliers. Dalam beberapa hal, AdaBoost menjadi kurang rentan terhadap masalah overfitting, jika dibandingkan dengan algoritma pembelajaran pada umumnya. AdaBoost dapat dideskripsikan melalui persamaan sebagai berikut :

$$F(x) = \sum_{t=1}^T \alpha_t h_t(x) \tag{2}$$

Dimana:

α_t : Tingkat pembelajaran (learning rate)

h_t : Pengklasifikasian dasar/ lemah

$F(x)$: Pengklasifikasian kuat

Pembentukan classifier pada Adaboost didasari input dataset training (D_i) adalah $(X_1, Y_1), \dots, (X_m, Y_m)$ dimana setiap X_i dimiliki beberapa domain atau ruang instance X dan setiap label Y_i dalam beberapa set label Y . Diasumsikan $Y = \{-1, +1\}$. AdaBoost disebut algoritma pembelajaran yang lemah dalam serangkaian putaran $t = 1 \dots T$. Salah satu ide utama dari algoritma ini adalah untuk memelihara distribusi bobot lebih dari training set. Bobot dari distribusi ini pada contoh training i dalam putaran t adalah dilambangkan $D_t(i)$. Awalnya, semua bobot ditetapkan sama, tetapi dalam setiap putaran, bobot dari contoh yang salah diklasifikasikan meningkat sehingga pembelajaran dasar dipaksa untuk fokus pada contoh yang sulit pada set training.

2.4 Evaluasi

Hasil percobaan akan di metode cross fold validation untuk menganalisa dan mengevaluasi model yang diusulkan [10]. Cross fold validation digunakan untuk memilih model dalam mempraktekkan masalah pembelajaran dalam k-iterasi [13]. Pada percobaan ini, kami menggunakan K-10 pada metode cross fold validation yang dilakukan. Cross fold validation meminimalkan generalisasi kesalahan [14]. Evaluasi dilakukan untuk melihat hasil uji berupa akurasi, presisi, dan recall sehingga dapat diketahui performa yang dihasilkan dari percobaan tersebut [15]. Perhitungan evaluasi menggunakan confusion matrix sebagai metode untuk menentukan kinerja yang didasarkan pada benar atau salah pada sebuah klasifikasi [16] [17].

Tabel 1. Confusion Matrix

	Actual Positive	Actual Negative
Prediction Positive	True Positive (TP)	False Positive (FP)
Prediction Negative	False Negative (FN)	True Negative (TN)

Pada tabel 1 menunjukkan matriks untuk merekap hasil pengujian. Hasil pengujian kemudian dimasukkan kedalam tabel confusion matrix seperti pada tabel 1 yang akan digunakan sebagai landasan perhitungan akurasi, recall, dan precision. Perhitungan akurasi, recall, dan precision menggunakan rumus sebagai berikut :

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Dataset yang digunakan dalam pengolahan ini menggunakan dataset sekunder yang bersasal dari kaggle repositori. Dari kaggle repository diambil dataset smoke detection (<https://www.kaggle.com/datasets/deepcontractor/smoke-detection-dataset>) yang terdiri dari 15 atribut dan 2 kelas (1 dan 0). Jumlah data yang digunakan sebanyak 62.629 data. Daftar atribut yang terdapat pada dataset dapat dilihat pada tabel 2.

Tabel 2. Daftar Atribut

No	Atribut	Keterangan
1	No	No Urut
2	UTC	Timestamp UTC Seconds
3	Temperature[C]	Air Temperature

No	Attribut	Keterangan
4	Humidity[%]	Air Humidity
5	TVOC[ppb]	Total Volatile Organic Compounds
6	eCO2[ppm]	Co2 Equivalent Concentration
7	Raw H2	Raw Molecular Hydrogen
8	Raw Ethanol	Raw Ethanol Gas
9	Pressure[hPa]	Air Pressure
10	PM1.0	Particulate Matter Size < 1.0 µm
11	PM2.5	Particulate Matter Size < 2.5 µm
12	NC0.5	Number Concentration Of Particulate Matter By The Particle Size: < 0.5 µm
13	NC1.0	Number Concentration Of Particulate Matter By The Particle Size: < 1.0 µm
14	NC2.5	Number Concentration Of Particulate Matter By The Particle Size: < 2.5 µm
15	CNT	Sample Counter
16	Fire Alarm	Result

3.2 Preprocessing

Dari data yang telah didapatkan, sebelum masuk ke proses pemodelan, dilakukan proses preprocessing terlebih dahulu agar data dapat diolah. Proses preprocessing dilakukan dengan 2 tahapan : menghilangkan atribut yang membuat bias dan mengkonversi tipe data. Pada tahapan pertama (menghilangkan atribut), ada 2 atribut yang dihilangkan yaitu atribut no dan atribut UTC. Atribut no dan UTC dihilangkan karena isi dari atribut tersebut sangat unik dan setiap data memiliki nilai no dan UTC yang berbeda sehingga menurunkan tingkat akurasi yang dihasilkan. Kemudian, setelah kedua atribut tersebut dihilangkan, data yang ada tipe datanya dirubah dari numerik ke nominal (numtom). Proses ini dilakukan karena pada saat proses modeling tipe data sebelum dirubah tidak bisa dilakukan proses modeling, sehingga perlu dilakukan konversi tipe data sehingga dapat dilanjutkan ke proses pemodelan.

3.3 Modeling

Dari data yang telah dilakukan proses preprocessing, kemudian data dilakukan proses pemodelan. Tahap awal yang dilakukan pada proses pemodelan dengan menentukan model validasi dengan menggunakan cross fold validation dengan k-10. Dengan model validasi tersebut data akan secara otomatis dibagi menjadi data training dan data testing. Pada tahap pemodelan klasifikasi, dilakukan 2 kali proses percobaan. Yang pertama dengan menggunakan algoritma naïve bayes, kemudian menggunakan algoritma naïve bayes dengan penambahan algoritma adaboost.

Tabel 3. Hasil percobaan pada algoritma naïve bayes

	Actual Positive	Actual Negative
Prediction Positive	17.781	3.302
Prediction Negative	92	41.455

Hasil percobaan dari algoritma naïve bayes dapat dilihat pada tabel 3 dapat diketahui bahwa nilai TP sebesar 17.781, FP sebesar 3.302, FN sebesar 92, dan TN sebesar 41.455.

Tabel 4. Hasil percobaan pada algoritma naïve bayes dengan algoritma adaboost

	Actual Positive	Actual Negative
Prediction Positive	17.622	533
Prediction Negative	251	44.224

Hasil percobaan dari algoritma naïve bayes dengan algoritma adaboost dapat dilihat pada tabel 4 dapat diketahui bahwa nilai TP sebesar 17.622, FP sebesar 533, FN sebesar 251, dan TN sebesar 44.224.

3.4 Perform Report

Dari hasil perhitungan pemodelan, kemudian dilakukan proses penghitungan performa dari setiap percobaan. Hasil perhitungan performa dapat dilihat pada tabel 5.

Tabel 5. Hasil Perhitungan Performa Algoritma Naïve Bayes

	Naïve Bayes
Akurasi	0,946
Recall	0,995
Presisi	0,843

Pada tabel 5 menjelaskan tentang hasil perhitungan performa menggunakan algoritma naïve bayes diketahui bahwa hasil akurasi yang dihasilkan sebesar 0,946, recall 0,995, dan presisi 0,843. Sedang pada tabel 6

menjelaskan tentang hasil perhitungan performa menggunakan algoritma naïve bayes + adaboost diketahui bahwa hasil akurasi yang dihasilkan sebesar 0,987, recall 0,986, dan presisi 0,971.

Tabel 6. Hasil Perhitungan Performa Algoritma Naïve Bayes + Adaboost

	Naïve Bayes	Naïve Bayes + Adaboost
Akurasi	0,946	0,987
Recall	0,995	0,986
Presisi	0,843	0,971

Pada tabel 5 dan tabel 6 dapat diketahui bahwa pada performa akurasi diketahui dengan menambahkan algoritma adaboost akurasi dapat ditingkatkan dari 0.946 menjadi 0.987 yang berdampak pada semakin kecilnya tingkat kesalahan prediksi yang akan dilakukan. Pada pengujian hasil recall atau sensitifitas yang mengukur rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif diketahui bahwa dengan algoritma naïve bayes tanpa adaboost mendapatkan hasil recall atau sensitifitas lebih baik dengan nilai 0.995 dengan 0.986. Sedangkan pada presisi, pada pengujian algoritma naïve bayes yang ditambahkan algoritma adaboost mendapatkan hasil lebih baik dengan nilai 0.971, jika tidak menggunakan algoritma adaboost mendapatkan hasil nilai 0.843 yang menunjukkan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan terkait optimasi akurasi smode detection dengan algoritma adaboost yang dikombinasikan dengan algoritma naïve bayes menunjukkan bahwa dari data yang telah diambil dari repositori kaggle dataset menunjukkan bahwa jumlah data yang ada berjumlah 62.629 dengan 15 atribut dan 2 kelas. Data yang telah didapatkan tidak langsung dapat diolah, tetapi harus melewati proses preprocessing terlebih dahulu dengan menghilangkan 2 atribut, yaitu atribut no dan UTC. Setelah atribut no dan UTC dihilangkan kemudian tipe data dikonversi numtonom agar dapat diolah menggunakan algoritma klasifikasi. Pada tahap testing menggunakan metode 10 k-fold cross validation untuk membagi menjadi data training dan data testing. Data training kemudian dilakukan 2 kali percobaan dengan menggunakan algoritma naïve bayes dan algoritma naïve bayes yang dikombinasikan dengan algoritma adaboost. Hasil pengujian yang telah dicatat dan dihitung berdasarkan akurasi, recall, dan presisi menunjukkan bahwa nilai akurasi prediksi dengan menggunakan algoritma adaboost dapat meningkatkan nilai akurasi yang dihasilkan dengan nilai 0,987. Jika melihat evaluasi dari sisi presisi juga menunjukkan penggunaan algoritma adaboost dapat meningkatkan nilai presisi dengan nilai 0,971. Tetapi pada evaluasi recall menunjukkan tanpa adaboost mendapatkan nilai lebih baik dengan nilai 0,995

REFERENCES

- [1] R. Zafar, S. Zaib, and M. Asif, "False fire alarm detection using data mining techniques," *Int. J. Decis. Support Syst. Technol.*, vol. 12, no. 4, pp. 21–35, 2020, doi: 10.4018/IJDSST.2020100102.
- [2] S. Wang, X. Xiao, T. Deng, A. Chen, and M. Zhu, "A Sauter mean diameter sensor for fire smoke detection," *Sensors Actuators, B Chem.*, vol. 281, no. September 2018, pp. 920–932, 2019, doi: 10.1016/j.snb.2018.11.021.
- [3] E. V. V. D. D. P. C.K. Gomathy, "IRJET- The Fire Alarm System based on IoT," *Irjet*, vol. 8, no. 10, 2021.
- [4] M. T. Anwar, W. Hadikurniawati, E. Winarno, and W. Widiyatmoko, "Performance Comparison of Data Mining Techniques for Rain Prediction Models in Indonesia," *2020 3rd Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2020*, no. December, pp. 83–88, 2020, doi: 10.1109/ISRITI51436.2020.9315460.
- [5] F. Marisa, "Educational Data Mining (Konsep dan Penerapan)," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 91–93, 2013.
- [6] D. K. Appana, R. Islam, S. A. Khan, and J. M. Kim, "A video-based smoke detection using smoke flow pattern and spatial-temporal energy analyses for alarm systems," *Inf. Sci. (Ny.)*, vol. 418–419, pp. 91–101, 2017, doi: 10.1016/j.ins.2017.08.001.
- [7] R. Yendra, L. Marifni, and I. Suryani, "Klasifikasi Data Mining Untuk Seleksi Penerimaan Calon Pegawai Negeri Sipil Tahun 2017 Menggunakan Metode Naïve Bayes," *J. Sains Mat. dan Stat.*, vol. 6, no. 1, p. 65, 2020, doi: 10.24014/jsms.v6i1.9254.
- [8] E. S. WAHYUNI and M. HENDRI, "Smoke and Fire Detection Base on Convolutional Neural Network," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 7, no. 3, p. 455, 2019, doi: 10.26760/elkomika.v7i3.455.
- [9] A. N. Rais and A. Subekti, "Integrasi SMOTE Dan Ensemble AdaBoost Untuk Mengatasi Imbalance Class Pada Data Bank Direct Marketing," *J. Inform.*, vol. 6, no. 2, pp. 278–285, 2019, doi: 10.31311/ji.v6i2.6186.
- [10] F. Paquin, J. Rivnay, A. Salleo, N. Stingelin, and C. Silva, "Multi-phase semicrystalline microstructures drive exciton dissociation in neat plastic semiconductors," *J. Mater. Chem. C*, vol. 3, pp. 10715–10722, 2015, doi: 10.1039/b000000x.
- [11] Fadila, W. I. Rahayu, and M. H. K. Saputra, *Penerapan Metode Naïve Bayes dan Skala Likert Pada Aplikasi Prediksi Kelulusan Mahasiswa*, Rolly Maul. Bandung: Kreatif Industri Nusantara, 2020.
- [12] E. Listiana and M. A. Muslim, "Prosiding Seminar Nasional Teknologi dan Informatika, 2017 : Kudus, 25 Juli 2017," *Pros. SNATIF*, no. 2015, pp. 875–881, 2017.
- [13] X. Zhang and Q. Song, "A multi-label learning based kernel automatic recommendation method for support vector machine," *PLoS One*, vol. 10, no. 4, pp. 1–30, 2015, doi: 10.1371/journal.pone.0120455.
- [14] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, 2014, doi: 10.1016/j.compeleceng.2013.11.024.



- [15] M. D. Alkhusayid, “Implementasi Algoritma K-Nearest Neighbors Pada Penentuan Jurusan Siswa,” vol. 4, no. September, pp. 25–36, 2022, doi: 10.30865/json.v4i1.4759.
- [16] S. T. Rizaldi and M. Mustakim, “Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air pada Algoritma K- Nearest Neighbor dan Naïve Bayes Classifier,” *Semin. Nas. Teknol. Informasi, Komun. dan Ind.* 12, pp. 130–137, 2020.
- [17] L. Munkhdalai, T. Munkhdalai, K. H. Park, H. G. Lee, M. Li, and K. H. Ryu, “Mixture of Activation Functions with Extended Min-Max Normalization for Forex Market Prediction,” *IEEE Access*, vol. 7, pp. 183680–183691, 2019, doi: 10.1109/ACCESS.2019.2959789.