

Implementasi Teknik Ensemble Stacking pada Klasifikasi Penyakit Anemia

Nur Muklisah¹, Ria Indira², Anita Desiani^{3*}, Novi Rustiana Dewi⁴, Yuli Andriani⁵

^{1,2,3,4,5} Program Studi Matematika Universitas Sriwijaya

¹08011382227099@student.unsri.ac.id

²08011282227045@student.unsri.ac.id

^{3*}anita_desiani@unsri.ac.id

⁴novirustiana@unsri.ac.id

⁵yuliandriani@unsri.ac.id

Abstrak— Anemia adalah penyakit yang disebabkan oleh kondisi seseorang yang memiliki kadar *hemoglobin* (Hb) darah dibawah normal. Deteksi penyakit dapat menggunakan bantuan data mining untuk mengklasifikasikan penyakit. Algoritma yang digunakan dalam penelitian ini adalah *Gaussian Naive Bayes*, *K-Nearest Neighbor* dan *Support Vector Machine* yang kemudian diterapkan pada teknik *ensemble stacking*. Penerapan *Ensemble* bertujuan untuk mendapatkan nilai keakurasian yang lebih baik dari klasifikasi individu. Pengujian algoritma ini menggunakan dua teknik pengujian yaitu *percentage split* dan *k-fold cross validation*. Untuk *percentage split* menggunakan ukuran split sebesar 80% data training dan 20% data uji dan pada *k-fold cross validation* dipilih nilai $k=10$. Hasil klasifikasi dari algoritma-algoritma tersebut memperoleh bahwa *percentage split* mendapatkan hasil akurasi yang lebih baik dibandingkan *k-fold cross validation*. Algoritma *Support Vector Machine* (SVM), *Gaussian Naive Bayes* dan *k-Nearest Neighbor* (kNN) dengan teknik pengujian *percentage split* memperoleh hasil akurasi secara berturut-turut sebesar 90,16%, 94,61% dan 96,49%. *K-Nearest Neighbor* (kNN) menghasilkan nilai akurasi tertinggi dari ketiga algoritma tersebut, namun dengan penerapan teknik ensemble memberikan kenaikan akurasi sebesar 1.05% dari hasil *k-Nearest Neighbor* (kNN). *Ensemble* dengan model *stacking* memperoleh hasil akurasi sebesar 97,19%. Berdasarkan hasil yang diperoleh dapat disimpulkan bahwa ensemble dengan model *stacking* dengan teknik pengujian *percentage split* memperoleh kinerja yang terbaik dari algoritma lainnya pada klasifikasi penyakit anemia.

Kata Kunci— *Ensemble Learning*, *Support Vector Machine*, *Gaussian Naive Bayes*, *K-Nearest Neighbor*, Anemia

I. PENDAHULUAN

Anemia adalah suatu kondisi dimana seseorang memiliki kadar *hemoglobin* (Hb) dibawah kadar *hemoglobin* normal pada umumnya. Anemia juga menjadi salah satu masalah kesehatan masyarakat terbesar di dunia. Populasi penderita anemia semakin meningkat, sehingga memerlukan penanganan yang cepat dan perlunya dilakukan deteksi dini. Deteksi dini dapat dilakukan dengan menggunakan klasifikasi data mining. Anemia perlu diklasifikasikan untuk mendapatkan hasil yang lebih akurat, salah satunya yaitu dengan menggunakan metode algoritma *Gaussian Naive Bayes*. Algoritma *Gaussian Naive Bayes* telah diterapkan pada berbagai klasifikasi penyakit, seperti penyakit migrain, dengan akurasi sebesar 89,09% [1]. Pada kasus penyakit gagal ginjal kronis, algoritma ini juga digunakan dan menghasilkan akurasi sebesar 89,93% [2].

Algoritma *Gaussian Naive Bayes* digunakan untuk menyelesaikan masalah klasifikasi dengan memanfaatkan data pelatihan yang relatif kecil untuk menentukan parameter yang dibutuhkan dalam proses klasifikasi [3]. Namun, keterbatasannya menunjukkan bahwa diperlukan algoritma alternatif yang lebih sederhana dan efisien, seperti kNN.

Diagnosa penyakit anemia dengan klasifikasi algoritma *k-Nearest Neighbor* (kNN) merupakan metode pengklasifikasian objek berdasarkan data training tetangga terdekat. Prinsip kerja algoritma kNN adalah menentukan dan mencari jarak terdekat terhadap k nilai terdekat pada data training dan data uji. Metode kNN memiliki beberapa keunggulan seperti pelatihan sederhana, cepat, mudah dipahami, dan efektif bila data pelatihan berukuran besar. Namun Algoritma kNN juga mempunyai kelemahan yaitu nilai k yang cenderung tidak seimbang[4]. Penelitian lain yang menggunakan metode kNN pada klasifikasi penyakit kanker payudara mendapatkan hasil akurasi yang tergolong baik yaitu 98% [5]. Pada klasifikasi penyakit ginjal juga menggunakan algoritma kNN dengan akurasi mencapai 92,59% [2]. Namun metode kNN kurang baik untuk dataset yang besar dan kompleks, sehingga perlu menggunakan algoritma untuk dataset yang besar dan lebih kompleks seperti metode SVM.

Selain kNN dan GNB, klasifikasi penyakit anemia juga dapat menggunakan metode lain seperti algoritma SVM. *Support Vector Machine* (SVM) merupakan metode yang dapat meminimalkan error pada training set dan memilih hyperline dengan margin terbesar. Algoritma ini juga digunakan dalam klasifikasi penyakit diabetes melitus dengan tingkat akurasi yang baik yaitu sebesar 96,27% [3]. Pada klasifikasi penyakit jantung, algoritma ini juga diterapkan dan menghasilkan akurasi sebesar 83% [6]. Namun algoritma SVM dirancang untuk menyelesaikan masalah klasifikasi dua kelas, meskipun beberapa waktu belakangan metode SVM dipakai untuk menyelesaikan tiga atau lebih kelas, namun masih dalam pengembangan. Dengan demikian diperlukan metode lain untuk menyelesaikan klasifikasi lebih dari dua kelas, yakni metode ensemble.

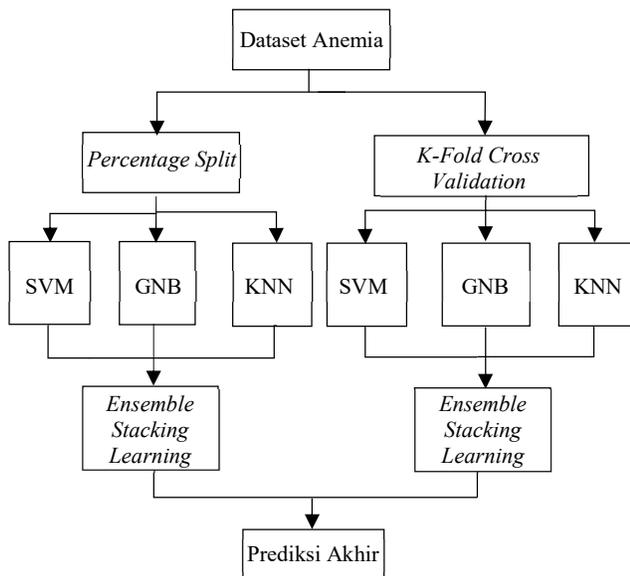
Ensemble Learning dikenal dapat meningkatkan hasil nilai akurasi dan meminimalkan kesalahan klasifikasi yang diterapkan dengan menggunakan tiga base classifiers [7]. Beberapa tipe yang dimiliki metode *Ensemble Learning* diantaranya voting, bagging, AdaBoost, stacking dan masih banyak lainnya [8]. Menurut Annisa & Ismasari [9] pada

penelitian perbandingan metode Ensemble dalam meningkatkan akurasi Algoritma Machine Learning dalam memprediksi penyakit Breast Cancer (Kanker Payudara) mendapatkan hasil nilai akurasi sebesar 83,76%. Indri & Ari [10] juga menuliskan pada penelitiannya, Metode Ensemble Learning dapat memberikan hasil nilai akurasi sebesar 95% dengan nilai k optimal yaitu k=10. Namun penelitian tersebut hanya menggunakan single classifier, sehingga dapat mempengaruhi akurasi secara keseluruhan.

Klasifikasi algoritma SVM, kNN, dan GNB memiliki kekurangan dan kelebihan masing-masing, dengan menerapkan metode ensemble maka tingkat akurasi akan semakin tinggi. Dengan menerapkan metode stacking ensemble learning, kelemahan dari masing-masing algoritma dapat diminimalisir dengan menggabungkan kekuatan dari beberapa model. Stacking ensemble learning bekerja dengan cara menggabungkan prediksi dari beberapa model dasar (base learners) dan menggabungkannya dengan menggunakan model meta-learner, dengan harapan dapat meningkatkan performa secara keseluruhan. Penelitian ini akan menggunakan dua metode validasi, yaitu *percentage split* dan *k-fold cross-validation*, untuk memastikan hasil klasifikasi yang konsisten dan dapat digeneralisasi. Dengan evaluasi berdasarkan akurasi, presisi, dan recall, penelitian memiliki tujuan untuk mendapatkan model yang lebih akurat dalam pengklasifikasian penyakit anemia.

II. METODOLOGI PENELITIAN

Metodologi penelitian memiliki peran penting dalam sebuah penelitian agar penelitian menjadi lebih sistematis dan sesuai dengan alur yang telah ditetapkan. Dengan adanya metodologi yang tepat, diharapkan penelitian mendapatkan hasil yang baik. Tahap dari penelitian ini disajikan pada Gambar 1.



Gbr. 1 Alur Penelitian

Pada Gambar 1. dataset anemia akan diuji menggunakan dua metode, yaitu metode *percentage split* dan metode *k-fold cross validation*. Masing-masing teknik pengujian tersebut akan menerapkan metode klasifikasi dengan tiga algoritma yaitu algoritma SVM, algoritma GNB, dan algoritma kNN. Hasil dari setiap metode klasifikasi kemudian akan digabungkan dengan menggunakan teknik ensemble stacking learning untuk setiap teknik pengujian. Prediksi akhir untuk klasifikasi penyakit anemia akan didasarkan pada hasil akurasi tertinggi dari proses ensemble stacking learning.

A. Dataset

Data yang digunakan merupakan data yang berkaitan dengan penyakit anemia yang terdiri atas 1.421 data yang diambil dari situs kaggle [11]. Dataset anemia ini memiliki 5 atribut dan 1 label yaitu Result dimana label 0 berarti tidak terdeteksi penyakit anemia dan label 1 terdeteksi penyakit anemia. Ada sebanyak 801 data yang tidak terdeteksi penyakit anemia dan ada 620 data terdeteksi penyakit anemia. Berikut merupakan atribut yang digunakan dalam dataset anemia yang disajikan pada Tabel 1.

TABEL 1
ATRIBUT PADA DATASET ANEMIA

Atribut	Keterangan	Tipe data	Range
Gender	Kategori Gender	Kategori	0=Laki-laki 1=Perempuan
Hemoglobin	Parameter dari Hemoglobin	Numerik	6.6 – 16.9
MCH	Parameter dari Mean Corpuscular Hemoglobin (MCH)	Numerik	16 – 30
MCHC	Parameter dari Mean Corpuscular Hemoglobin Concentration (MCHC)	Numerik	27.8 – 32.5
MCV	Parameter dari Mean Corpuscular Volume (MCV)	Numerik	69.4 – 101.6
Result	Kategori terdeteksi anemia atau tidak terdeteksi anemia	Kategori	0=Tidak anemia 1=Anemia

B. Metode Training

1) Percentage Split

Metode training dengan memisahkan data menjadi 2 set yaitu data *training* dan data uji. Dimana persentase yang digunakan pada data *training* dapat sebesar 80% dan persentase pada data uji sebesar 20%. Masing-

masing algoritma akan diuji menggunakan *percentage split* [12].

2) *K-fold Cross Validation*

Metode training yang tidak memperhatikan bagaimana data dibagi. Pada data uji, setiap titik data muncul tepat satu kali, namun pada data latih, titik data muncul sebanyak k-1 kali[13]. Pada metode ini akan digunakan k=10, dengan membagi data menjadi 10 bagian, dimana untuk data latih dilakukan pada satu bagian dan untuk data uji dilakukan pada data lainnya. Proses tersebut akan dilakukan secara bergantian.

C. *Algoritma Support Vector Machine (SVM)*

Algoritma yang digunakan sebagai klasifikasi dengan mencari *Hyperplane marjinal maksimum (MMH)*, atau *hyperplane optimal* dengan margin tertinggi, dengan cara membagi data menjadi dua kelas[14]. Data learning dengan titik data x_i ($i = 1, 2, \dots, m$) dimana untuk m adalah jumlah data yang mempunyai dua kelas yaitu $y_i = \pm 1$ (dimana +1 merupakan kelas positif dan -1 merupakan kelas negatif) [15]. Pada metode SVM, menggunakan kernel membantu menyelesaikan masalah data non-linier dengan menempatkan data tersebut ke dalam ruang dimensi yang lebih besar. Kernel sigmoid merupakan jenis kernel pada model SVM yang dapat mengubah data input yang awalnya tidak dapat dipisahkan secara linier, dapat dipisahkan dengan baik[16]. Tahap yang digunakan untuk mencari *hyperplane* pada SVM yaitu sebagai berikut:

1. Data $\vec{x}_i \in (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n)$ dimana x_i merupakan suatu data yang terdiri dari n atribut dengan dua kelas yaitu $y_i \in \{-1, +1\}$.
2. Anggaplah data merupakan linier dan kelas berada antara +1 dan -1, dimana data tersebut dapat dipisahkan dengan *hyperplane*. Sehingga dapat didefinisikan sebagai:

$$\vec{w} \cdot \vec{x} + b = 0 \tag{1}$$

Dan diperoleh persamaan berupa:

$$\vec{w} \cdot \vec{x} + b \geq 1, \text{ untuk kelas } +1 \tag{2}$$

$$\vec{w} \cdot \vec{x} + b \geq -1, \text{ untuk kelas } -1 \tag{3}$$

Dimana:

- \vec{w} = Nilai
- \vec{x} = Data input
- b = Posisi bidang relatif

3. Untuk mendapatkan suatu *hyperplane* yang bernilai optimal, dapat dilakukan dengan mencari suatu *hyperplane* yang berguna sebagai pemisah yang dapat memaksimalkan jarak antara dua kelas tersebut. Pencarian titik minimal dapat dilakukan dengan persamaan sebagai berikut:

$$\min_w \frac{1}{2} (|\vec{w}|)^2 \tag{4}$$

Kendala:

$$y_i(\vec{x}_i, \vec{w} + b) - 1 \geq 0 \tag{5}$$

Dimana:

- y_i = Kelas target ke- i
- \vec{x}_i = Input ke- i
- \vec{w} = Bobot
- b = Posisi bidang relatif

4. Data non-linier dapat diatasi dengan menggunakan kernel yang dapat mengubah bentuk dari yang semula *input space* menjadi bentuk *feature space*.

D. *Algoritma Gaussian Naive Bayes (GNB)*

Sebuah teori yang digunakan untuk menghitung nilai dari probabilitas bersyarat, dimana peluang bersyarat dari A ketika B dapat dinotasikan sebagai $P(A | B)$. Metode ini merupakan metode yang paling umum dan metode yang penting dalam menghitung probabilitas dan menghitung statistik[17]. Adapaun persamaan umum dari algoritma Gaussian Naive Bayes adalah sebagai berikut:

$$P(X_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)} \tag{6}$$

Keterangan:

$$P(X_i|y) = \text{Probabilitas}$$

$$x_i = \text{Nilai variabel acak } X \text{ pada } i$$

$$\mu = \text{Rata-rata}$$

$$\sigma = \text{Varians}$$

Untuk mendapatkan μ dan σ maka menggunakan persamaan sebagai berikut:

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \tag{7}$$

Dimana, n adalah nilai sampel ke n .

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1} \tag{8}$$

E. *Algoritma k-Nearest Neighbor (kNN)*

Algoritma ini bekerja berdasarkan dengan jarak terdekat dari data uji ke data latih, Kemudian diambil data yang menjadi data terbanyak dari kNN yang akan dijadikan sebagai prediksi dari data uji[18]. Tahap yang digunakan untuk menghitung metode kNN adalah yaitu:

1. Mempersiapkan data.
2. Menentukan besar nilai K.
3. Menghitung besar jarak antara data latih dan data uji dengan persamaan berikut.

$$\text{euclidean} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \tag{9}$$

Dimana untuk p_i adalah data *training*, q_i adalah data *testing*, i adalah variabel data, dan untuk n adalah dimensi dari data.

4. Mengurutkan data berdasarkan jarak yang telah dihitung.
5. Menentukan tetangga terdekat (nearest neighbor) sampai dengan urutan yang ke K.
6. Menentukan kelas yang sesuai.
7. Mengklasifikasi data uji dan mencari jumlah kelas dari tetangga terdekat dan menetapkan kelas tersebut sebagai kelas yang akan dievaluasi.

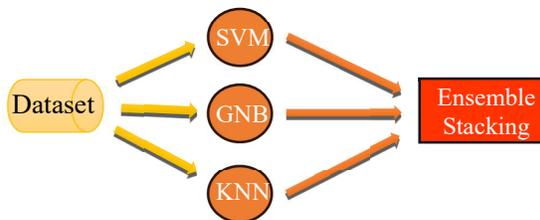
F. *Stacking Ensemble Learning*

Ensemble merupakan proses menggabungkan beberapa algoritma atau model yang berbeda untuk memberikan hasil dengan akurasi yang lebih tinggi. Salah satu model metode *ensemble learning* yaitu *stacking*. *Stacking* merupakan *training*

meta models untuk memprediksi berdasarkan output beberapa model dasar[19]. Adapun langkah-langkah kerja *stacking* yaitu:

1. Data latih dibagi menjadi dua subset atau lebih.
2. Beberapa model dasar dilatih pada satu subset data.
3. Setelah model selesai dilatih, setiap model terlatih akan digunakan untuk menentukan prediksi pada subkumpulan data yang lainnya.
4. Hasil dari prediksi setiap model dasar di kombinasikan dengan kumpulan data baru.
5. Melakukan pelatihan *meta models* pada dataset baru yang menggunakan nilai target asli sebagai label dan *meta models* untuk memprediksi pada data uji.

Secara metode *ensemble learning* (termasuk model *stacking*), bekerja seperti pada gambar 2. Metode yang akan digunakan diantaranya yaitu Algoritma *Support Vector Machine* (SVM), Algoritma *Gaussian Naive Bayes* (GNB), dan Algoritma *k-Nearest Neighbor* (kNN).



Gbr. 2 Skema Metode *Ensemble Learning* pada SVM, GNB, dan kNN.

G. Model Evaluasi

Pengujian untuk tingkat akurasi dapat dilakukan dengan cara menerapkan *confusion matrix*. *Confusion matrix* dapat digunakan untuk menghitung nilai dari akurasi, presisi, dan *recall* pada data mining anemia. Nilai akurasi merupakan persentase data yang telah dilakukan proses klasifikasi secara akurat setelah dilakukannya tahap pengujian pada data. Presisi mengacu pada kondisi yang positif dan yang benar-benar menghasilkan temuan positif sejati pada data yang aktual. *Recall* atau yang disebut juga sebagai sensitivitas merupakan kejadian positif yang nyata kemudian diprediksi secara tepat sebagai data yang positif[20]. Bentuk dari *confusion matrix* disajikan dalam Tabel 2.

TABEL 2
CONFUSION MATRIX

Kelas		Aktual	
		Negative	Positive
Prediksi	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

Berdasarkan Tabel 2. *True Positive* (TP) merupakan jumlah data yang bernilai positif dan diprediksikan sebagai data positif. *False positive* (FP) merupakan jumlah data yang negatif kemudian diprediksikan sebagai data yang bernilai positif. *True*

Negative (TN) merupakan jumlah data negatif yang diprediksikan sebagai negatif. *False Negative* (FN) merupakan jumlah data yang bernilai negatif dan diprediksikan sebagai data yang positif.

Untuk mengitung akurasi, presisi, dan *recall* pada tabel *confusion matrix* dapat dilakukan dengan menggunakan persamaan sebagai berikut.

$$Akurasi = \frac{(TP+TN)}{(TP+TN+FP+F)} \tag{10}$$

$$Presisi = \frac{(TP)}{(TP+FP)} \tag{11}$$

$$Recall = \frac{(TP)}{(TP+F)} \tag{12}$$

III. HASIL DAN PEMBAHASAN

Pada penelitian ini, klasifikasi data mining dilakukan dengan menggunakan tiga metode, baik dilakukan secara individu maupun secara bersamaan dengan menggunakan metode *stacking*. Tiga metode yang digunakan yaitu *Support Vector Machine* (SVM), *Gaussian Naive Bayes* (GNB), dan *k-Nearest Neighbor* (kNN). Metode training yang digunakan yaitu *percentage split* dan *k-fold cross validation*.

Perbandingan nilai akurasi, presisi, dan *recall* pada metode *stacking* dengan 3 algoritma lain dengan metode training yaitu *percentage split* dapat dilihat pada Tabel 3.

TABEL 3
HASIL PERBANDINGAN METODE TRAINING PERCENTAGE SPLIT

Algoritma	Akurasi	Presisi	Recall
SVM	90.53%	90.52%	90.53%
GNB	95.09%	94.53%	94.53%
kNN	96.14%	96.06%	95.31%
Stacking	97.19%	97.21%	97.19%

Berdasarkan Tabel 3. hasil perbandingan beberapa algoritma berdasarkan metode *Training Percentage Split*, algoritma *stacking* menunjukkan kinerja terbaik dengan akurasi sebesar 97,19%, presisi 97,21%, dan *recall* 97,19%. Algoritma kNN menempati posisi kedua dengan akurasi 96,14%, presisi 96,06%, dan *recall* 95,31%. Algoritma GNB memiliki akurasi 95,09%, presisi 94,53%, dan *recall* 94,53%, sementara algoritma SVM memiliki kinerja terendah dengan akurasi dan *recall* masing-masing sebesar 90,53% serta presisi sebesar 90,52%. Hal ini menunjukkan bahwa metode *Stacking* lebih unggul dalam memprediksi hasil dibandingkan algoritma lainnya.

Perbandingan nilai akurasi, presisi, dan *recall* pada metode *stacking* dengan 3 algoritma lain dengan metode training yaitu *k-fold cross validation* dapat dilihat pada Tabel 4.

TABEL 4
HASIL PERBANDINGAN DENGAN METODE TRAINING K-FOLD CROSS VALIDATION

Algoritma	Akurasi	Presisi	Recall
SVM	89.30%	89.29%	89.30%
GNB	93.81%	91.05%	95.16%
kNN	95.35%	93.92%	95.47%
Stacking	95.77%	95.79%	95.77%

Tabel 4. menyajikan hasil perbandingan performa algoritma menggunakan metode training *k-fold cross-validation*.

Berdasarkan tabel tersebut, algoritma *stacking* mencapai hasil tertinggi dengan akurasi 95,77%, presisi 95,79%, dan recall 95,77%. Algoritma kNN berada di posisi kedua dengan akurasi 95,35%, presisi 93,92%, dan recall 95,47%. Selanjutnya, algoritma GNB mencatatkan akurasi 93,81%, presisi 91,05%, dan recall 95,16%. Sementara itu, algoritma SVM menunjukkan performa terendah dengan akurasi, presisi, dan recall sebesar 89,30%. Hasil ini menunjukkan keunggulan algoritma Stacking dibandingkan dengan algoritma lainnya pada metode pelatihan ini.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat diambil kesimpulan bahwa untuk kasus pengklasifikasian pada dataset penyakit anemia dengan menggunakan algoritma Gaussian Naive Bayes (GNB), Algoritma k-Nearest Neighbors (kNN), Algoritma Support Vector Machine (SVM) dan *stacking ensemble learning* tergolong baik. Masing-masing dari model tersebut mempunyai kelebihan dan kekurangan dalam menghadapi dataset tersebut. Pada metode training dengan *percentage split* untuk algoritma SVM menghasilkan nilai akurasi sebesar 90.53%. Untuk Algoritma GNB menghasilkan menghasilkan akurasi sebesar 95.09%. Sedangkan untuk Algoritma kNN menghasilkan nilai akurasi sebesar 96.14%. Dan untuk *stacking ensemble learning* memberikan akurasi yang paling tinggi yaitu sebesar 97.19%. Metode training dengan *k-fold cross validation* untuk algoritma SVM menghasilkan nilai akurasi sebesar 89.30%. Untuk Algoritma GNB menghasilkan menghasilkan akurasi sebesar 93.81%. Sedangkan untuk Algoritma kNN menghasilkan nilai akurasi sebesar 95.35%. Dan untuk *stacking ensemble learning* memberikan akurasi yang paling tinggi yaitu sebesar 95.77%. Penggunaan metode *stacking ensemble learning* memberikan keunggulan dalam menangani masalah kelas dataset dengan menggabungkan kelebihan metode dan meminimalkan kekurangan serta meningkatkan kinerja model secara keseluruhan dengan mengurangi *overfitting* dan bias. Sehingga diperoleh nilai akurasi terbaik menggunakan ensemble model *stacking* dengan teknik pengujian *percentage split* pada penyakit anemia.

REFERENSI

- [1] M. Yuichi and Y. A. Susetyo, "Klasifikasi Penyakit Migrain dengan Metode Naïve Bayes pada Dataset Kaggle," *J. Indones. Manaj. Inform. dan Komun.*, no. 1, pp. 139–151, 2025.
- [2] N. Fatimah Indrianti, A. Kania Ningsih, and R. Ilyas, "Implementasi Data Mining Untuk Klasifikasi Penyakit Gagal Ginjal Kronis Menggunakan Metode K-Nearest Neighbor," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 2255–2260, 2024, doi: 10.36040/jati.v8i2.9464.
- [3] H. Apriyani and K. Kurniati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.issuue3.year2020.page133-143.
- [4] A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 1109–1119, 2022, doi: 10.35957/jatisi.v9i2.2048.
- [5] W. Ramdhani, D. Bona, R. B. Musyaffa, and C. Rozikin, "Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 12, pp. 445–452, 2022.
- [6] M. D. F. Tino, Herliyani Hasanah, and Tri Djoko Santosa, "Perbandingan Algoritma Support Vector Machines (Svm) Dan Neural Network Untuk Klasifikasi Penyakit Jantung," *INFOTECH J.*, vol. 9, no. 1, pp. 232–235, 2023, doi: 10.31949/infotech.v9i1.5432.
- [7] A. M. Siregar, "Klasifikasi Untuk Prediksi Cuaca Menggunakan Esemble Learning," *Petir*, vol. 13, no. 2, pp. 138–147, 2020, doi: 10.33322/petir.v13i2.998.
- [8] V. H. Alves Ribeiro, S. Moritz, F. Rehbach, and G. Reynoso-Meza, "A novel dynamic multi-criteria ensemble selection mechanism applied to drinking water quality anomaly detection.," *Sci. Total Environ.*, vol. 749, p. 142368, Dec. 2020, doi: 10.1016/j.scitotenv.2020.142368.
- [9] A. M. Majid and I. Nawangsih, "Perbandingan Metode Ensemble Untuk Meningkatkan Akurasi Algoritma Machine Learning Dalam Memprediksi Penyakit Breast Cancer (Kanker Payudara)," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 23, no. 1, p. 97, 2024, doi: 10.53513/jis.v23i1.9563.
- [10] L. Ati and A. Kusyanti, "Metode Ensemble Classifier untuk Mendeteksi Jenis Attention Deficit Hyperactivity Disorder (SDHD) pada Anak Usia Dini," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 3, pp. 301–308, 2019, doi: 10.25126/jiitk.2019631313.
- [11] E. Cicekyurt, "Anemia Classification with EDA (100% Acc)," Kaggle. [Online]. Available: <https://www.kaggle.com/code/emreikyurt/anemia-classification-with-eda-100-acc/input>
- [12] Adi Nugroho, Agustinus Bimo Gumelar, Adri Gabriel Sooi, Dyana Sarvasti, and Paul L Tahalele, "Perbandingan Performansi Kinerja Algoritma Pengklasifikasian Terpandu Untuk Kasus Penyakit Kardiovaskular," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 998–1006, 2020, doi: 10.29207/resti.v4i5.2316.
- [13] S. Prusty, S. Patnaik, and S. K. Dash, "SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer," *Front. Nanotechnol.*, vol. 4, no. August, pp. 1–12, 2022, doi: 10.3389/fnano.2022.972421.
- [14] K. Khadijah and R. Kusumaningrum, "Ensemble Classifier untuk Klasifikasi Kanker Payudara," *It J. Res. Dev.*, vol. 4, no. 1, pp. 61–71, 2019, doi: 10.25299/itjrd.2019.vol4(1).3540.
- [15] A. Desiani, M. Akbar, I. Irmeilyana, and A. Amran, "Implementasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular," *J. Tek. Elektro dan Komputasi*, vol. 4, no. 2, pp. 207–214, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/ELKOM/article/view/7691>
- [16] A. Zalvadila, "Klasifikasi Penyakit Tanaman Bawang Merah Menggunakan Metode SVM dan CNN," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 255–260, 2023, doi: 10.30591/jpit.v8i3.5341.
- [17] Nurul A'ayunnisa, Y. Salim, and H. Azis, "Analisis Performa Metode Gaussian Naïve Bayes untuk Klasifikasi Citra Tulisan Tangan Karakter Arab," *Indones. J. Data Sci.*, vol. 3, no. 3, pp. 115–121, 2022, doi: 10.56705/ijodas.v3i3.54.
- [18] D. Cahyanti, A. Rahmayani, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
- [19] B. Sunarko *et al.*, "Penerapan Stacking Ensemble Learning untuk Klasifikasi Efek Kesehatan Akibat Pencemaran Udara," *Edu Komputika J.*, vol. 10, no. 1, pp. 55–63, 2023, doi: 10.15294/edukomputika.v10i1.72080.
- [20] S. Wulan Dari and J. Triloka, "Kajian Algoritme Mask Region-Based Convolutional Neural Network (Mask R-CNN) dan You Look Only Once (YOLO) Untuk Deteksi Penyakit Kulit Akibat Infeksi Jamur," *Semin. Nas. Has. Penelit. dan Pengabd. Masy.*, pp. 132–138, 2022.