

Perbandingan Algoritma Naive Bayes Dan Decision Tree Dalam Pengujian Data Anemia Menggunakan K-Fold Cross Validation

Diterima:
10 Juni 2024
Revisi:
10 Juli 2024
Terbit:
1 Agustus 2024

¹Aurel Fransisca Kusuma Wardhani, ²Nila Kamilatutsaniya, ³M Alfianto Alamsyah, ⁴Erna Daniati, ⁵Aidina Ristyawan
¹⁻⁵Universitas Nusantara PGRI Kediri Fakultas Teknik dan Ilmu Komputer

¹aurelfransisca03@gmail.com, ²nila.kamila23@gmail.com,
³alfiantoalamsyah@gmail.com, ⁴ernadaniati@unpkediri.ac.id,
⁵aidinaristi@unpkediri.ac.id

Abstrak— *Anemia* mempengaruhi sekitar 1/3 dari populasi dunia, *anemia* adalah masalah kesehatan global yang perlu dipertimbangkan. Karena meningkatnya prevalensi *anemia* pada wanita, gadis remaja lebih rentan terhadap penyakit ini daripada remaja laki-laki. *Anemia* adalah istilah untuk konsentrasi *hemoglobin* rendah dalam *eritrosit*. Salah satu cara untuk menentukan anemia seseorang adalah dengan memeriksa kadar *hemoglobin* mereka. Kadar *hemoglobin* pada wanita di atas 15 lebih dari 12,0 g / dl (7,5 mmol). Penelitian ini bertujuan untuk menilai seberapa baik kinerja dua algoritma klasifikasi menggunakan dataset *Kaggle* untuk memprediksi anemia. *Decision Tree* dan *Naive Bayes* adalah algoritma yang diuji. Pada penelitian ini, menggunakan metode percobaan (eksperimen). Langkah-langkah pemrosesan yang digunakan termasuk pengambilan data, pemilihan, persiapan, perubahan, dan pembagian data untuk memisahkan data pelatihan dan pengujian. Nilai yang dievaluasi dinilai melalui validasi lintas batas, akurasi, presisi, recall, dan skor F1. Berdasarkan hasil, algoritma pohon keputusan ditemukan memiliki tingkat akurasi 99% setelah penyeimbangan data dan validasi silang. Setelah melakukan validasi silang kami menemukan bahwa proses penyeimbangan meningkatkan akurasi. Kesimpulannya, penggunaan algoritma *Decision Tree* dan *Naive Bayes* untuk memprediksi anemia memberikan wawasan penting dalam pengelolaan dan *diagnosis anemia*, terutama pada remaja putri.

Kata Kunci—*Anemia; Data Mining; Algoritma Klasifikasi; Cross Validation*

Abstract— *Anemia* affects about 1/3 of the world's population, *anemia* is a global health problem that needs to be considered. Due to the increasing prevalence of *anemia* in women, teenage girls are more susceptible to the disease than adolescent boys. *Anemia* is the term for low *hemoglobin* concentrations in *erythrocytes*. One way to determine a person's anemia is to check their *hemoglobin* levels. *Hemoglobin* levels in women above 15 are more than 12.0 g/dl (7.5 mmol). This study aims to assess how well the two classification algorithms perform using the *Kaggle* dataset to predict anemia. *Decision Tree* and *Naive Bayes* are the algorithms tested. An experimental methodology was employed in this investigation. The processing steps used include data capture, selection, preparation, change, and data sharing to separate training and testing data. The evaluated scores are assessed through cross-border validation, accuracy, precision, recall, and F1 scores. Based on the results, the decision tree algorithm was found to have an accuracy rate of 99% after data balancing and cross-validation. After cross-validation, we found that the balancing process improved accuracy.

Keywords—*Anemia; Data Mining; Classification Algorithm; Cross Validation*

This is an open access article under the CC BY-SA License.



Penulis Korespondensi: Erna Daniati

Erna Daniati,
Sistem Informasi,
Universitas Nusantara PGRI Kediri,

I. PENDAHULUAN

Anemia terjadi ketika aliran darah mengandung jumlah *hemoglobin (Hb)* dan tingkat sel darah merah yang sangat rendah pada beberapa jenis kelamin dan kelompok umur. Hemoglobin adalah Sel darah merah memiliki tugas guna membawa oksigen dan nutrisi lainnya ke otak dan berbagai jaringan tubuh, ditemukan dalam butir darah merah orang yang sehat. Ada perbedaan dalam kadar Hb normal laki-laki dan perempuan. Kurang dari 13,5 g/dl adalah tingkat Hb pada pria dengan *anemia*, dan kurang dari 12 g/dl adalah tingkat Hb pada wanita.1, 2 [1]. Rona merah darah dikenal sebagai *hemoglobin*, yang dibuat oleh *protein kaya zat besi* yang mengangkut *karbon dioksida* dari tubuh ke paru-paru untuk ekskresi dan oksigen dari paru-paru ke seluruh tubuh.. Gadis remaja mengalami *anemia* defisiensi besi lebih sering daripada pria remaja. (Hafsah Us & Mey Elisa Safitri. 2023:2-3)[2].

Di Indonesia hampir 23% remaja putri menderita kurang darah atau dikenal dengan anemia. Setidaknya 4,8 juta dari perkiraan 21 juta gadis remaja mengalami defisit sel darah merah. Sel ini bertanggung jawab atas perjalanan oksigen dari jantung ke seluruh tubuh (The Conversation, 2019). Remaja perempuan memiliki anemia lebih besar daripada remaja laki-laki. Imunitas remaja, fokus, prestasi akademik, kebugaran, dan produktivitas semuanya terkena dampak negatif oleh *anemia* (Kemenkes, 2016)[3].

Kebutuhan oksigen penderita dan derajat dan kecepatan *anemia* memengaruhi gejala dan tanda *anemia*. Pada *anemia* yang *progresif*, gejala akan lebih ringan karena *mekanisme homeostatik* dapat menyesuaikan diri dengan berkurangnya kapasitas darah untuk membawa oksigen. Gejala *anemia* dapat disebabkan oleh berkurangnya *suplai* oksigen ke jaringan dan *hipovolemia*, yang merupakan faktor kompensasi volume *sekuncup* pada pasien dengan perdarahan hebat. Denyut jantung dan curah jantung pada nilai Hb 5 g% (Ht 15%) [4].

Perkembangan terbaru di bidang ilmu komputer, terutama dalam kecerdasan buatan, telah mendorong kemajuan dalam bidang ini. Penggunaan *AI* dan *ML* dalam pengaturan perawatan kesehatan perempuan telah berkembang pesat dalam beberapa tahun terakhir [5]. Data mining adalah proses mengumpulkan informasi dan pengetahuan dari suatu database dengan menggunakan metode seperti *machine learning*, statistik matematika, dan kecerdasan buatan untuk menemukan pola.[6].

Gejala *Anemia* saat ini dapat diprediksi menggunakan bantuan teknologi, teknologi ini bekerja dengan mengumpulkan data – data terdahulu lalu mengolahnya hingga menjadi sebuah

informasi. Namun dalam proses pengelolahannya sistem atau teknologi memerlukan sebuah metode dalam pengelolaannya. Klasifikasi penambangan data adalah pendekatan yang digunakan untuk mencari definisi fitur serupa dalam kelompok atau *class*. [7].

Pohon keputusan dan metode klasifikasi *Naïve Bayes* akan digunakan dalam penyelidikan. Salah satu teknik klasifikasi yang dapat digunakan untuk memperkirakan berapa banyak orang yang mungkin berada di kelas adalah algoritma *Naïve Bayes*. [8]. Pohon keputusan sangat membantu dalam menguraikan korelasi tersembunyi antara variabel tujuan dan beberapa variabel yang mungkin, atau data *ekspolarisasi*. [9]. Kelebihan algoritma ini adalah dapat digambarkan dalam bentuk pohon keputusan, yang dapat dipahami dengan mudah, lentur, dan mengagumkan [10]. *Diagnostik anemia* yang dipakai pada penelitian ini diunduh dari *Kaggle.com*. Dataset ini memiliki beberapa atribut mengenai nama-nama sel darah mulai dari *WBC*, *LYMp*, *NEUTp*, *LYMn*, *NEUTn*, *RBC*, *HGB*, *HCT*, *MCV*, *MCH*, *MCHC*, *PLT*, *PDW*, dan *PCT*.

Penerapan algoritma classification diatas menggunakan *tools jupyter Notebook versi 7.0.8*. Studi ini membandingkan algoritma dengan akurasi tertinggi antara *Naïve bayes* dan pohon keputusan. Dari penelitian ini penulis berharap dapat memberikan wawasan baru.

II. METODE

2.1 Klasifikasi

Penggunaan teknik klasifikasi merupakan aspek penting dalam penambangan data. Klasifikasi melibatkan pengkategorian atau penamaan data/objek baru berdasarkan atribut tertentu [11].

2.2 Algoritma Naïve Bayes

Naïve Bayes Classifier berdasarkan *Teorema Bayes*, meramalkan peluang masa depan berdasarkan pengalaman sebelumnya dengan menggunakan *teknik statistik* dan *probabilitas*. Fitur utama dari Pengklasifikasi *Naïve Bayes* ini adalah keyakinannya yang ekstrem (naif) bahwa setiap keadaan atau kejadian adalah unik. [12]. Algoritma ini sensitif terhadap panjang rekaman yang dibuat oleh kumpulan data yang digunakan (Sidiq dkk., 2020) [7].

2.3 Algoritma Decission Tree

Teknik yang populer dan efektif untuk *kategorisasi* dan prediksi saat *ekspolarisasi* data adalah penggunaan pohon keputusan, yang mengungkapkan korelasi tersembunyi antara berbagai *variabel prospektif* dan *variabel target*. Banyak data diubah menjadi keputusan yang mencerminkan aturan menggunakan teknik pohon keputusan. Tabel dan catatan atribut

merupakan cara standar untuk menyajikan data dalam pohon keputusan. Atribut menentukan *parameter* yang dibuat sebagai kriteria saat membangun pohon. [9].

Metode eksperimen digunakan dalam penelitian ini untuk menguji keakuratan setiap algoritma klasifikasi dalam memprediksi *anemia*. Setiap model algoritma klasifikasi diuji dengan data pelatihan dan digunakan untuk melakukan prediksi pada data pengujian setelah data dikumpulkan dan dilakukan pre-prosesing. Untuk mengevaluasi performa model, metrik seperti akurasi, presisi, recall, dan melihat terjadinya *kolinearitas* pada setiap tabel atribut. Hasil akhir akan ditampilkan dalam bentuk table yang berisi nilai – nilai dari hasil pengujian.

2.4 Pengumpulan Data

Kumpulan data penelitian terdiri dari *kompilasi diagnosis anemia* yang dikumpulkan dari Kaggle.com. Pada langkah pertama, data tersebut dianalisis dan diidentifikasi untuk pemodelan [13].

2.5 Data Selection

Untuk menyelidikan lebih lanjut, proses *Knowledge Discovery Database (KDD)* akan diterapkan pada data. Pemilihan data dari sekumpulan data *operasional* adalah langkah pertama dalam proses *KDD* [14].

2.5 Pre-processing

Melakukan pengecekan jika ada *field* dengan nilai yang terduplikat dan *field* dengan nilai data yang nihil. Melakukan normalisasi untuk mengubah nilai *numerik* dalam himpunan data dalam rentang nilai dan kehilangan informasi

2.6 Pembagian Data

Setelah menyelesaikan langkah *pra-pemrosesan*, data akan dibagi menjadi set pelatihan dan tes. Sebagian besar dataset akan digunakan untuk pelatihan model, dan sebagian kecil lagi akan digunakan untuk pengujian model.

2.7 Kolinearitas

Proses ini juga termasuk ke dalam kegiatan atau langkah *pre-processing*, disini data akan dilihat seberapa kuat hubungannya dengan data dari atribut tabel lain jika semakin tinggi nilai kolinearitasnya maka semakin tinggi pula terjadi double data atau data duplikat, hal ini bisa diatasi dengan cara menghapus dan menggabungkan tabelnya

2.8 Data Mining

Data merupakan suatu proses yang dipakai guna mengekstraksi informasi atau *knowledge* yang berharga dari kumpulan data yang sangat besar dan kompleks. Data mining bertujuan untuk mendapatkan wawasan yang lebih dalam dan bernilai dengan mengidentifikasi pola, hubungan, atau informasi yang mungkin tidak terlihat secara langsung dalam data [15].

2.9 Evaluasi Model

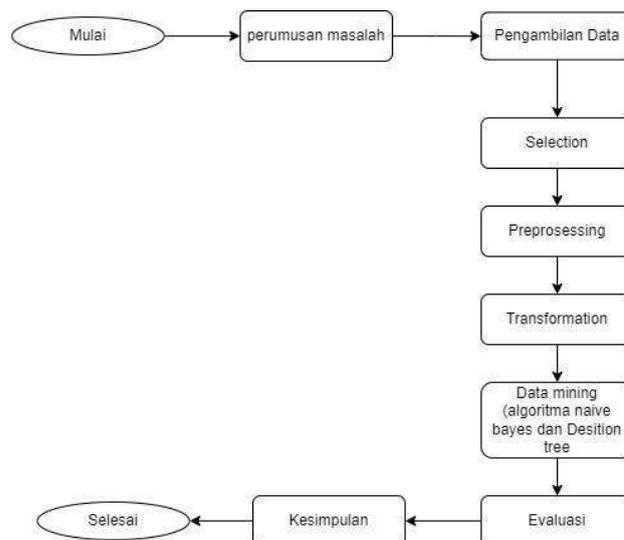
Setelah pelatihan, model yang dihasilkan dari evaluasi diuji dengan data pengujian yang berbeda. Menilai *performa* model masing-masing algoritma klasifikasi menggunakan *metrik* seperti akurasi, presisi, recall, dan skor f1. Kemudian, hasil *performa* akan ditampilkan dalam bentuk tabel.

2.9 Cross Validation

Ketika data dibagi menjadi set tes dan pelatihan, teknik penambangan data yang disebut validasi silang digunakan untuk mendapatkan nilai *akurasi optimal*. Teknik ini membagi data sampel secara acak dan kemudian menggabungkannya menggunakan nilai K sebanyak yang diperlukan untuk menilai kinerja proses suatu algoritma. [16]. Kinerja model pembelajaran mesin (akurasi) dapat dievaluasi dan diuji menggunakan teknik *validasi* silang. Semua subset selain dari satu yang akan digunakan untuk menguji kemampuan *prediksi* model akan digunakan untuk melatih model *Machine Learning*.

2.10 Tahapan Penelitian

Tahap pada penelitian ini bisa dilihat pada Gambar 1 berikut ini:



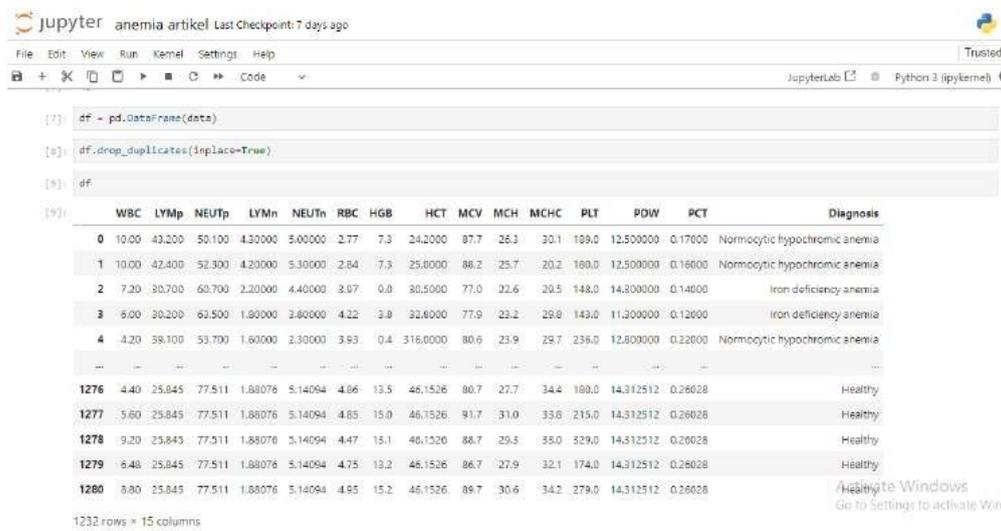
Gambar 1. Tahapan Penelitian

Tahap penelitian yang akan dilakukan digambarkan pada Gambar 1. Membentuk masalah, pengumpulan data yang diperlukan, dan setelah itu masuk ke tahap *KDD* — yang memerlukan *seleksi*, *pra-pemrosesan*, *transformasi*, pemrosesan data, dan *evaluasi* — adalah langkah pertama dalam proses. Pada akhirnya, ini mengarah pada kesimpulan penelitian setelah selesai.

III. HASIL DAN PEMBAHASAN

3.1 Klasifikasi dengan Algoritma Naïve Bayes

Algoritma Decision Tree dan *Naive Bayes* digunakan untuk kategorisasi dalam karya ini. *Algoritma NB* dan *Decision Tree* digunakan dalam penyelidikan ini. Bahasa pemrograman *Python* digunakan di *Jupyter* untuk melakukan pengujian *K-Fold Cross Validation*. Metode memasukkan data ke *Jupyter* digambarkan dalam gambar berikut.



Gambar 2. Proses input data

Dapat dilihat Gambar 2, menunjukkan proses penggunaan data dan isi data *Anemia*, yang berjumlah 1232 serta terdapat 49 data *duplikasi* dan tidak ada data yang hilang. Kemudian dilakukan pembagian data pengujian untuk *normalisasi* data di *Jupyter* seperti pada gambar 3 berikut ini:

Sebelum aktivitas kolinearitas:

	feature	VIF
0	WBC	6.992793
1	LYMp	12.212816
2	NEUTp	1.258841
3	LYMn	4.244656
4	NEUTn	4.227171
5	RBC	3.048324
6	HGB	35.507562
7	HCT	1.889000
8	MCV	29.043023
9	MCH	1.646901
10	MCHC	27.953705
11	PLT	7.047739
12	PDW	4.945497
13	PCT	1.129533

Gambar 5. Hasil VIF (Kolinearitas) Sebelum Dihapus

Pada Gambar 4 dan Gambar 5, setelah dicek *kolinearitas* datanya ternyata atribut data yang paling tinggi *kolinearitasnya* yaitu *HGB (Hemoglobin)* dan jika terjadi kolinearitas maka data harus dihapus atau digabungkan seperti pada Gambar 6.

Setelah perubahan kolinearitas (gabungan fitur HGB, MCV, MCHC):

	feature	VIF
0	WBC	8.182030
1	LYMp	19.192178
2	NEUTp	1.274645
3	LYMn	4.933245
4	NEUTn	4.962024
5	RBC	4.828926
6	HGB	inf
7	HCT	1.923972
8	MCV	inf
9	MCH	1.723267
10	MCHC	inf
11	PLT	7.688253
12	PDW	21.887589
13	PCT	1.145830
14	new	inf

Gambar 7. Hasil VIF (Kolinearitas) Setelah Dihapus

Pada Gambar 7, menunjukkan tabel yang memiliki tingkat kemungkinan kolinearitas tinggi dihapus agar tidak terjadi duplikasi data. Setelah pengecekan kolinearitas sudah dilakukan maka data bisa langsung diproses untuk melihat hasil *accuration*, *precision*, serta *recall* seperti di Gambar 8 berikut ini.



```
Jupyter anemia artikel Last Checkpoint: 7 days ago
File Edit View Run Kernel Settings Help
nb_precision = precision_score(y_test, y_pred_nb, average='weighted', zero_division=1)
nb_recall = recall_score(y_test, y_pred_nb, average='weighted', zero_division=1)

[29]: # Train & Evaluate Algorithm Decision Tree
dt_model.fit(X_train, y_train)
y_pred_dt = dt_model.predict(X_test)
dt_accuracy = accuracy_score(y_test, y_pred_dt)
dt_precision = precision_score(y_test, y_pred_dt, average='weighted', zero_division=1)
dt_recall = recall_score(y_test, y_pred_dt, average='weighted', zero_division=1)

[30]: # Hasil
print("Naive Bayes:")
print(f"Accuracy: {nb_accuracy:.2f}")
print(f"Precision: {nb_precision:.2f}")
print(f"Recall: {nb_recall:.2f}")

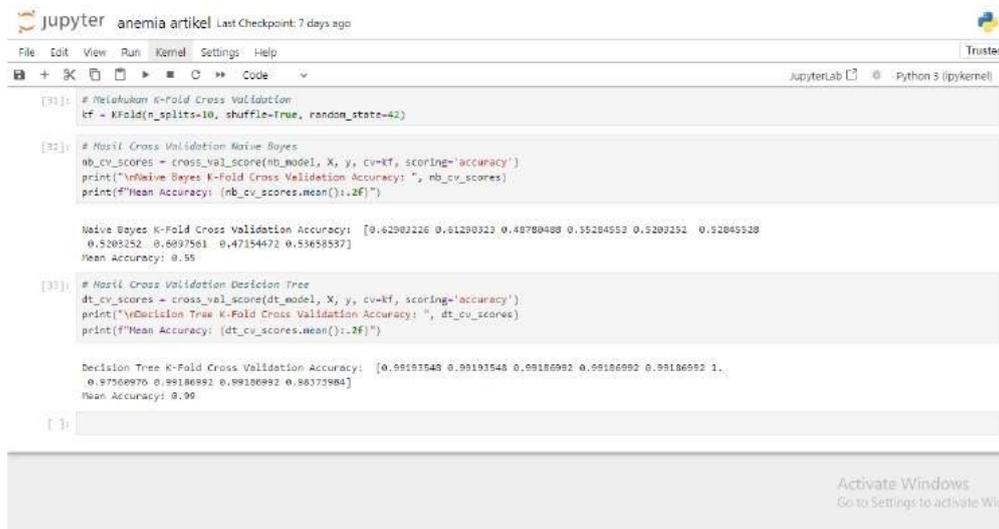
print("\nDecision Tree:")
print(f"Accuracy: {dt_accuracy:.2f}")
print(f"Precision: {dt_precision:.2f}")
print(f"Recall: {dt_recall:.2f}")

Naive Bayes:
Accuracy: 0.66
Precision: 0.69
Recall: 0.66

Decision Tree:
Accuracy: 0.95
Precision: 0.96
Recall: 0.95
```

Gambar 8. Hasil *Accuration*, *Precision*, dan *Recall*

Pada Gambar 8, menunjukkan hasil akurasi, precision dan recallnya bahwa ternyata algoritma *Naive Bayes* memiliki hasil akurasi 66%, hasil precision 69%, serta hasil recall 66%. Sedangkan algoritma *Decision Tree* memiliki hasil akurasi 95%, hasil precision 96%, dan hasil recall 95%. Dari hasil tersebut perlu dilakukan pengujian menggunakan *Cross Validation* untuk hasil yang lebih akurat seperti yang terlihat di gambar 9.



```
Jupyter anemia artikel Last Checkpoint: 7 days ago
File Edit View Run Kernel Settings Help
kf = KFold(n_splits=10, shuffle=True, random_state=42)

[31]: # Melakukan K-Fold Cross Validation

[32]: # Hasil Cross Validation Naive Bayes
nb_cv_scores = cross_val_score(nb_model, X, y, cv=kf, scoring='accuracy')
print("\nNaive Bayes K-Fold Cross Validation Accuracy: ", nb_cv_scores)
print(f"Mean Accuracy: {nb_cv_scores.mean():.2f}")

Naive Bayes K-Fold Cross Validation Accuracy: [0.62902226 0.61290323 0.48780488 0.55284553 0.5202252 0.52865238
0.5205252 0.8897561 0.47154472 0.53658537]
Mean Accuracy: 0.55

[33]: # Hasil Cross Validation Decision Tree
dt_cv_scores = cross_val_score(dt_model, X, y, cv=kf, scoring='accuracy')
print("\nDecision Tree K-Fold Cross Validation Accuracy: ", dt_cv_scores)
print(f"Mean Accuracy: {dt_cv_scores.mean():.2f}")

Decision Tree K-Fold Cross Validation Accuracy: [0.99193548 0.99193548 0.99186992 0.99186992 0.99186992 1.
0.97569970 0.99186992 0.99186992 0.98373984]
Mean Accuracy: 0.99
```

Gambar 9. Hasil Pengujian Menggunakan Cross Validation

Pada Gambar 9, menunjukkan hasil akurasi, precision dan recall setelah dilakukan pengujian atau pengecekan ulang menggunakan *cross validation* ternyata algoritma *Naive Bayes* mengalami penurunan yaitu dengan hasil akurasi menjadi 55%. Sedangkan algoritma *Decision Tree* mengalami kenaikan dengan hasil akurasi 99%. Ini menunjukkan bahwa ketika metode Pohon Keputusan digunakan dibandingkan dengan algoritma *Naive Bayes*, ia menghasilkan nilai akurasi paling akurat yaitu 99%. Berikut hasil perbandingannya

Algoritma	Akurasi	Precision	Recall	Cross Validation
Naive Bayes	0.66 (66%)	0.69 (69%)	0.66 (66%)	0.55 (55%)
Decision Tree	0.95 (95%)	0.96 (96%)	0.95 (95%)	0.99 (99%)

Hasil *Confusion Matrix* (Decision Tree)

	Prediction Positive	Prediction Negative
Actual Positive	95	5
Actual Negative	2	98

Hasil *Confusion Matrix* (Naive Bayes)

	Prediction Positive	Prediction Negative
Actual Positive	66	34
Actual Negative	10	90

IV. KESIMPULAN

Dari hasil analisa yang dilakukan ditemukan algoritma terbaik yaitu *Decision Tree* dengan hasil akurasi 0.99 atau 99%, meskipun pada saat diuji sebelum menggunakan *cross validation* sudah terlihat bahwa algoritma *Decision Tree* memiliki hasil akurasi yang lebih akurat tapi tetap membutuhkan *cross validation* agar hasilnya dapat dibandingkan ternyata algoritma *Naive Bayes* mengalami penurunan yang sebelumnya memiliki tingkat akurasi 0.66 atau 66% setelah di uji menggunakan *cross validation* turun menjadi 0.55 atau 55%. Sedangkan algoritma *Decision Tree* terjadi peningkatan dimana sebelumnya hasil akurasinya adalah 0.95 atau 95% setelah diuji menggunakan *cross validation* hasil akurasinya meningkat menjadi 0.99 atau 99%. Dan sebelum dilakukan pengujian hasil akurasi, dilakukan juga pengecekan *kolinearitas* yang menunjukkan *HGB* atau *Hemoglobin* menunjukkan kemungkinan *kolinearitas* paling tinggi. Jadi, penulis dapat menyimpulkan bahwa pengaruh *diagnosa anemia* paling tinggi adalah kadar *hemoglobin* darah yang tinggi.

UCAPAN TERIMAKASIH

Ungkapan terima kasih penulis ucapkan kepada Bapak/Ibu Dosen Pembimbing Universitas Nusantara PGRI Kediri yang telah membantu penelitian dari proses awal sampai dengan dipublikasikan.

DAFTAR PUSTAKA

- [1] A. Muhayati, D. Ratnawati, F. Ilmu Kesehatan, and U. Pembangunan Nasional Veteran Jakarta Jalan Limo Raya Kelurahan Limo Kecamatan Limo Kota, "Hubungan Antara Status Gizi dan Pola Makan dengan Kejadian Anemia Pada Remaja Putri," *J. Ilm. Ilmu Keperawatan Indones.*, vol. 9, no. 01, pp. 563–570, Apr. 2019, doi: 10.33221/JIIKI.V9I01.183.
- [2] S. U. Chasanah, P. P. Basuki, and I. M. Dewi, "Anemia penyebab, strategi pencegahan dan penanggulangannya bagi remaja," *Stikes Wira Husada Kemenristek Dikti*, 2019.
- [3] M. Komara Sari, R. Fadhillah, R. Novita Ratna, D. Program Studi Kebidanan Universitas Malahayati, and M. D. Prodi, "Penyuluhan Tentang Anemia Pada Remaja Sman 14 Bandar Lampung Kemiling Permai Tanjung Karang Barat Lampung Tahun 2020," *J. Perak Malahayati Pengabd. Kpd. Masy.*, vol. 2, no. 1, pp. 29–33, Jun. 2020, doi: 10.33024/JPM.V2I1.2697.
- [4] I. Putu Wibina Karsa Gumi *et al.*, "Perbandingan Algoritma Naïve Bayes dan Decision Tree Pada Sentimen Analisis," *Indones. J. Comput. Sci. Res.*, vol. 1, no. 2, Jul. 2022, doi: 10.59095/IJCSR.V1I2.11.
- [5] S. E. Dunkerton, Y. B. Jeve, N. Walkinshaw, E. Breslin, and T. Singhal, "Predicting Postpartum Hemorrhage (PPH) during Cesarean Delivery Using the Leicester PPH Predict Tool: A Retrospective Cohort Study," *Am. J. Perinatol.*, vol. 35, no. 2, pp. 163–169, Jan. 2018, doi: 10.1055/S-0037-1606332.
- [6] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 2, pp. 437–444, Apr. 2020, doi: 10.30865/MIB.V4I2.2080.
- [7] S. N. Chotimah and A. R. Rozzaqi, "Klasifikasi Diagnosis Penyakit Ginjal Kronis Dengan Menerapkan Konsep Algoritma Naïve Bayes," *JIPETIK Jurnal Ilm. Penelit. Teknol. Inf. Komput.*, vol. 4, no. 1, pp. 8–15, Jul. 2023, doi: 10.26877/JIPETIK.V4I1.16174.
- [8] E. Darnila, M. Maryana, and M. Azmi, "Aplikasi Klasifikasi Status Gizi Balita Menggunakan Metode Naïve Bayes Berbasis Android," *METHOMIKA J. Manaj. Inform. Komputerisasi Akunt.*, vol. 5, no. 2, pp. 135–141, Oct. 2021, doi: 10.46880/JMIKA.VOL5NO2.PP135-141.
- [9] A. Aziz and K. Karpen, "Diagnosa Penyakit Kulit WAJAH MENGGUNAKAN METODE DECESSION TREE DAN ALGORITMA C4.5," *J. Teknol. DAN OPEN SOURCE*, vol. 2, no. 1, pp. 74–86, Jun. 2019, doi: 10.36378/JTOS.V2I1.148.
- [10] U. M. Kudus, J. Ganesha, and P. Kudus, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, Oct. 2020, doi: 10.47970/SISKOM-KB.V4I1.173.
- [11] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 452–458, Mar. 2023, doi: 10.36040/JATI.V7I1.6303.
- [12] A. Felicia Watratan, A. B. Puspita, D. Moeis, S. Informasi, and S. Profesional Makassar,

- “Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia,” *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 7–14, Jul. 2020, doi: 10.52158/JACOST.VIII.9.
- [13] E. ABOELNAGA, “Anemia.” [Online]. Available: <https://www.kaggle.com/datasets/ehababoelnaga/anemia-types-classification>
- [14] N. Nurainun, E. Haerani, F. Syafria, and L. Oktavia, “Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation,” *J. Comput. Syst. Informatics*, vol. 4, no. 3, pp. 578–586, May 2023, doi: 10.47065/JOSYC.V4I3.3414.
- [15] P. Rahayu *et al.*, *Buku Ajar Data Mining*, vol. 1, no. January. 2018.
- [16] A. Hutapea, M. T. Furqon, and I. Indriati, “Penerapan Algoritme Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3957–3961, Feb. 2018, Accessed: Jun. 10, 2024. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2800>