

PREDIKSI DIAGNOSIS ANEMIA MELALUI METODE KLASIFIKASI BERBASIS MACHINE LEARNING

Riski Susanti, Lena Elfianti

Universitas Dehasen

Email : Riski Susanti2@gmail.com, lenaelfianti@gmail.com

ABSTRAK

Abstrak: Anemia adalah kondisi di mana jumlah sel darah merah dalam darah berada di bawah kadar normal, mengakibatkan terhambatnya distribusi oksigen ke organ tubuh. Hal ini dapat menghambat fungsi pernapasan sel dan mengurangi aktivitas tubuh secara optimal. Akibatnya, penderita anemia seringkali mengalami gejala seperti kebingungan, kelelahan, rasa dingin, dan kulit yang pucat. Sayangnya, banyak orang tidak menyadari tanda-tanda kecil ini, dan seringkali hanya mengunjungi rumah sakit ketika anemia telah menjadi penyakit kronis. Perlunya estimasi hasil dan mengevaluasi metode mana yang menghasilkan nilai akurasi yang lebih tepat. Estimasi menggunakan metode Klasifikasi dimana dalam metode ada 3 metode lagi meliputi (1) Decision Tree, (2) Naïve Bayes, (3) K-Nearst Neghtbors. Berdasarkan 3 metode klasifikasi tersebut akan dicari 1 metode terbaik dengan nilai akurasi paling terbaik yang akan digunakan dalam proses pengolahan deploy untuk prediksi penyakit anemia. Decision Tree keluar sebagai metode terbaik diantara metode klasifikasi lainnya. Dimana metode Decision Tree paling banyak menghasilkan nilai akurasi terbaik dalam 4 pengujian akurasi dengan rasio 90:10, 80:20, 70:30 dan 60:40. Hasilnya metode Decision Tree memiliki tingkat akurasi sebesar 99% pada rasio pengujian data 90:10 dan 96% pada rasio pengujian data 70:30. Pada rasio 90:10 . Dession Tree menghasilkan nilai estimasi terbaik yaitu Akurasi sebesar 1.0 dan pada Precision sebesar 1.0, serta pada F1-SCORE hasil estimasi sebesar 0.1. Hasil pengujian menunjukkan bahwa Decision Tree secara konsisten mencapai nilai akurasi, presisi, recall, dan F1-Score tertinggi dibandingkan dengan metode klasifikasi lainnya pada semua rasio pembagian data. Penemuan ini mengindikasikan bahwa Decision Tree efektif dalam memprediksi keberadaan anemia berdasarkan parameter yang diuji.

Kata Kunci: Anemia, Klasifikasi, Machine Learning

ABSTRACT

Abstract: Anemia is a condition where the number of red blood cells in the blood is below normal levels, resulting in obstruction of the distribution of oxygen to the body's organs. This can inhibit cell respiratory function and reduce optimal body activity. As a result, anemia sufferers often experience symptoms such as confusion, fatigue, feeling cold, and pale skin. Unfortunately, many people are unaware of these small signs, and often only visit hospital when anemia has become a chronic disease. There is a need to estimate results and evaluate which method produces more precise accuracy values. Estimation uses the Classification method where in the method there are 3 more methods including (1) Decision Tree, (2) Naïve Bayes, (3) K-Nearst Neghtbors. Based on these 3 classification methods, the 1 best method with the best accuracy value will be searched for. in the deployment processing process for predicting anemia. Decision Tree emerged as the best method among other classification methods. Where the Decision Tree method produces the best accuracy values in 4 accuracy tests with ratios of 90:10, 80:20, 70:30 and 60:40. As a result, the Decision Tree method has an accuracy rate of 99% at a data testing ratio of 90:10 and 96% at a data testing ratio of 70:30. At a ratio of 90:10 . The Dession Tree produces the best estimated value, namely Accuracy of 1.0 and Precision of 1.0, and the F1-SCORE estimation result is 0.1. Test results show that Decision Tree consistently achieves the highest accuracy, precision, recall and F1-Score values compared to other classification methods at all data sharing ratios. These findings indicate that the Decision Tree is effective in predicting the presence of anemia based on the parameters tested.

Keywords: Anemia, Classification, Machine Learning

PENDAHULUAN

Di era globalisasi saat ini, kemajuan teknologi komputer dalam bidang kesehatan telah mengalami perkembangan yang signifikan. Penggunaan sistem komputerisasi menjadi sangat penting, terutama dalam penanganan berbagai penyakit. Kehadiran sistem komputer dapat memberikan dukungan yang dibutuhkan, karena seringkali dokter dan perawat kesulitan dalam menentukan jenis penyakit yang dialami oleh pasien, baik itu orang dewasa, anak-anak, atau orang lanjut usia, tanpa adanya sistem komputerisasi yang efisien. Pada sistem ini akan membahas tentang penyakit anemia, penyakit anemia suatu kondisi kesehatan yang ditandai oleh kadar hemoglobin atau jumlah sel darah merah yang rendah dalam tubuh, memiliki dampak serius terhadap kesejahteraan dan produktivitas pasien. Diagnosa yang tepat dan cepat menjadi kunci dalam memberikan perawatan yang efektif dan mencegah komplikasi yang dapat terjadi. Namun, diagnosa anemia sering kali melibatkan analisis kompleks dari berbagai parameter darah dan mempertimbangkan sejumlah faktor risiko yang bervariasi. Diagnosis anemia melibatkan evaluasi berbagai parameter darah, termasuk kadar hemoglobin, volume sel darah merah, dan indeks sel darah merah. Proses ini bisa rumit dan memerlukan pemahaman mendalam terhadap interaksi kompleks antara berbagai faktor yang mempengaruhi keseimbangan darah. Gejala anemia dapat bervariasi, termasuk kelelahan, pusing, dan kulit pucat. Selain itu, penyebab anemia juga beragam, mulai dari defisiensi zat besi hingga masalah genetik.

Sel darah adalah elemen hidup yang terdiri dari sel darah merah, sel darah putih, dan plasma darah. Salah satu bentuk penyakit yang melibatkan sel darah adalah anemia, yang umumnya dikenal oleh masyarakat sebagai "Kurang Darah." Anemia adalah kondisi di mana jumlah sel darah merah dalam darah berada di bawah kadar normal, mengakibatkan terhambatnya distribusi oksigen ke organ tubuh. Hal ini dapat menghambat fungsi pernapasan sel dan mengurangi aktivitas tubuh secara optimal[1]. Akibatnya, penderita anemia seringkali mengalami gejala seperti kebingungan, kelelahan, rasa dingin, dan kulit yang pucat. Sayangnya, banyak orang tidak menyadari tanda-tanda kecil ini, dan seringkali hanya mengunjungi rumah sakit ketika anemia telah menjadi penyakit kronis.

Dataset ini mencakup informasi tentang anemia yang terkait dengan jenis kelamin, hemoglobin, MCHC, MCV, KIA, dan hasilnya. Buku ini menyajikan sejumlah besar data yang dapat dimanfaatkan tanpa adanya data tambahan. Algoritma klasifikasi biner akan digunakan dalam analisis ini. Jenis kelamin direpresentasikan sebagai 0 untuk laki-laki dan 1 untuk perempuan. Hemoglobin, sebagai protein dalam sel darah, bertanggung jawab membawa oksigen dari paru-paru ke organ dan jaringan tubuh. KIA (Mean Corpuscular Hemoglobin) adalah ukuran rata-rata hemoglobin dalam setiap sel darah merah, yang membawa oksigen ke seluruh tubuh. MCH (Mean Corpuscular Hemoglobin Concentration) mengacu pada konsentrasi rata-rata hemoglobin dalam sel darah. Terdapat perbedaan dalam kadar hemoglobin pada individu dengan tekanan darah tinggi, dan MCV (Mean Corpuscular Volume) juga memiliki kadar hemoglobin yang tinggi, meningkatkan risiko tertentu. Hasil pada dataset ini dikategorikan sebagai 0 untuk tidak anemia dan 1 untuk anemia. Anemia memiliki dampak yang signifikan, terutama pada populasi dengan tekanan darah tinggi dan kadar hemoglobin yang tinggi, meningkatkan risiko kesehatan. Menurut data WHO hingga tahun 2015, lebih dari 20% atau 2 miliar penduduk dunia menderita anemia. Prevalensi anemia lebih tinggi pada kelompok usia 9 hari hingga 43 hari. Wanita usia subur dan wanita hamil juga memiliki prevalensi anemia yang cukup tinggi, yaitu 30 persen dan 42 persen, secara berturut-turut. Dengan tingginya jumlah penderita anemia, penerapan teknologi menjadi solusi yang membantu. Teknologi digunakan untuk mendiagnosis kemungkinan anemia pada seseorang, terutama untuk mendukung solusi perawatan[2].

Pemasalahannya Penyakit anemia dapat memiliki dampak yang signifikan pada kesehatan dan kualitas hidup seseorang. Dampak ini dapat bervariasi tergantung pada tingkat keparahan anemia, penyebabnya diantara dampak yang umumnya terkait dengan penyakit anemia yaitu, anemia dapat mengurangi kemampuan darah untuk membawa oksigen ke seluruh tubuh. dapat menyebabkan kelelahan yang berkepanjangan, kekurangan energi, dan penurunan daya tahan tubuh. Kurangnya oksigen dapat mempengaruhi fungsi otak, menyebabkan masalah kognitif seperti kesulitan berkonsentrasi, kebingungan, dan penurunan daya ingat. Anemia dapat menyebabkan peningkatan frekuensi dan intensitas pernapasan, terutama selama aktivitas fisik. Hal ini terjadi karena tubuh berusaha untuk meningkatkan pasokan oksigen. Penting untuk diingat bahwa dampak anemia dapat beragam dan setiap individu mungkin mengalami gejala yang berbeda. Pengobatan dan manajemen anemia akan tergantung pada penyebab spesifik dan tingkat keparahan kondisi tersebut.

Pada permasalahan tersebut anemia dapat berdampak signifikan pada kesehatan dan kualitas hidup seseorang. Efeknya dapat bervariasi tergantung pada tingkat keparahan anemia, karena salah satu dampak umum yang terkait dengan anemia adalah kemampuan darah yang menurun dalam membawa oksigen ke seluruh tubuh. Ini dapat menyebabkan kelelahan yang berkepanjangan, penurunan energi, dan berkurangnya stamina. Kekurangan oksigen juga dapat memengaruhi fungsi otak, menyebabkan masalah kognitif seperti kesulitan berkonsentrasi, kebingungan, dan kehilangan ingatan. Anemia juga dapat menyebabkan peningkatan frekuensi dan intensitas pernapasan, terutama saat melakukan aktivitas fisik, karena tubuh berusaha meningkatkan pasokan oksigen[3]. Penting untuk diingat bahwa dampak anemia dapat bervariasi, dan setiap individu mungkin mengalami gejala yang berbeda. Pengobatan dan penanganan anemia bergantung pada penyebab khusus dan tingkat keparahan kondisi.

Metode klasifikasi berbasis pembelajaran machine learning digunakan dalam analisis penyakit anemia untuk memodelkan korelasi antara variabel input yang terkait dengan kondisi anemia dan variabel target, yang mungkin mencakup tingkat keparahan anemia atau parameter terkait lainnya. Dalam situasi ini, klasifikasi seringkali dipilih karena sederhana dan memberikan pemahaman yang baik tentang interaksi antar variabel. Klasifikasi merupakan cara pengelompokan benda berdasarkan ciri – ciri yang dimiliki oleh objek klasifikasi. Dalam prosesnya, klasifikasi dapat dilakukan dengan banyak cara baik secara manual ataupun dengan bantuan teknologi. Klasifikasi yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari algoritma cerdas komputer. Sedangkan klasifikasi yang dilakukan dengan bantuan teknologi, memiliki beberapa algoritma, diantaranya Naïve Bayes, Support Vector Machine, Decision Tree, Fuzzy dan Jaringan Saraf Tiruan[4].

Maka dari itu, tujuan dari penelitian ini adalah untuk mendapatkan estimasi hasil dan mengevaluasi metode mana yang menghasilkan nilai akurasi yang lebih tepat dengan membandingkan 3 model yang tercakup dalam analisis regresi, termasuk (1) Naïve Bayes, (2) Decision Tree, dan (3) K-Nearest Neighbors. Selain itu, penelitian ini akan menguji hasil menggunakan metrik Precision, Recall, dan F1-Score.

METODE

Metode penelitian dapat memuat hal-hal penting yang berkaitan dengan pelaksanaan penelitian dan tahapan-tahapan yang akan ditempuh dalam penelitian ini untuk mencapai hasil (output) yang diinginkan. Metode penelitiannya meliputi hal-hal sebagai berikut:

Peralatan Riset

Alat yang akan dipakai dalam penelitian ini melibatkan (1) sebuah komputer dengan spesifikasi yang memadai, (2) perangkat lunak yang akan diterapkan dalam proses pengolahan adalah aplikasi Jupyter Notebook yang dapat diakses melalui Anaconda Navigator. Python diadopsi sebagai bahasa pemrograman, dan Visual Studio Code berfungsi sebagai penyunting teks, sementara HTML digunakan sebagai bahasa markup untuk menciptakan dan menyebarluaskan aplikasi.

Variabel yang Terdaftar dalam Riset

Dalam ranah machine learning, penelitian ini masuk dalam kategori supervised learning, suatu metode pemodelan di mana data yang akan dianalisis memiliki label atau kelas. Tujuan utamanya adalah untuk memahami hubungan sebab-akibat antara variabel independen (variabel x) dan variabel dependen (variabel y), yang dalam konteks ini disebut sebagai target atau label. Oleh karena itu, variabel yang akan diuji dalam penelitian ini terdiri dari variabel independen atau x, yang melibatkan Gender, hemoglobin, KIA, MCHC, sementara variabel dependennya adalah MCV.

Tahapan – Tahapan Riset

Tahapan yang akan ditempuh dalam penelitian ini digambarkan dengan flowchart sehingga dapat terlihat dengan jelas tahapan yang akan dilakukan.

Data Riset

Data riset yang akan digunakan adalah dataset yang berasal dari kaggle.com pada tahun 2022. Data tersebut berjumlah sebanyak 1422 baris data. Terdapat 6 kolom (Gender, Hemoglobin, MCH, MCHC, MCV, Result) yang akan digunakan untuk pengujian dan 1 kolom (Result) sebagai kolom target/label

Pembagian Data

Pembagian data dalam penelitian ini menggunakan data training dan data testing dengan rasio sebesar 90:10 Dimana data training memiliki kisaran data sebanyak 90 dari dataset serta data testing sebanyak 10 dari dataset. Algoritma yang menggunakan rasio 90:10 memiliki performa yang lebih baik dibandingkan dengan algoritma lainnya dimana tingkat akurasi dapat diperoleh sebesar 99,40%. Adapun rasio 80:20, 70:30, dan 60:40 yang digunakan dalam penelitian ini sehingga rasio pembagian data lebih bervariasi dimana tujuan yang diharapkan adalah agar memperoleh model regresi manakah yang memiliki hasil estimasi terbaik.

Estimasi dengan 7 Model Regresi

Klasifikasi dalam informasi mining ialah tata cara pendidikan informasi buat memprediksi nilai dari sekelompok atribut. Algoritma klasifikasi hendak menciptakan sekumpulan ketentuan yang disebut rule yang hendak digunakan selaku penanda buat bisa memprediksi kelas dari informasi yang mau diprediksi.

HASIL DAN PEMBAHASAN

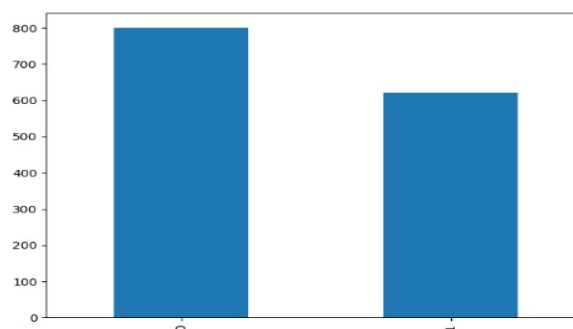
Pada pembahasan dalam penelitian ini akan dilakukan visualisasi data agar data dapat lebih mudah untuk diamati serta dapat mengetahui hubungan antara tiap variabel x terhadap variabel y . Selanjutnya, data akan diuji dengan menggunakan metode *K-Nearest Neighbors*. Hal tersebut bertujuan untuk menemukan model klasifikasi yang paling baik dengan menghasilkan nilai akurasi yang paling akurat untuk digunakan dalam estimasi pada penyakit anemia. Tahapan pembahasan yang dilakukan adalah sebagai berikut :

Visualisasi Data

Visualisasi data bertujuan untuk memudahkan seseorang dalam mengamati dan memahami sebuah data. Terdapat berbagai model yang dapat digunakan untuk menyajikan sebuah data kedalam bentuk visualnya, Visualisasi data disajikan dalam bentuk diagram yang menggunakan warna yang berbeda agar lebih mudah dipahami dan mudah diingat. Visualisasinya data tersebut sebagai berikut ini :

Diagram Batang

Berikut ini tampilan visualisasi data pada Label penyakit anemia.

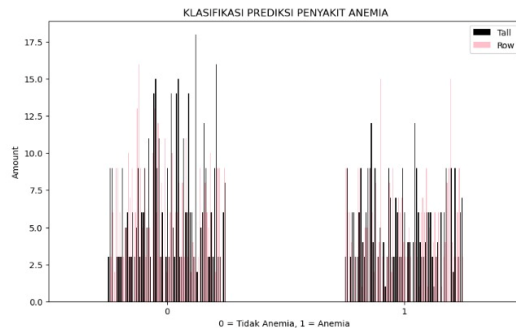


Gambar 1. Diagram Batang penyakit anemia

Diagram batang tersebut menunjukkan perbandingan antara jumlah pasien yang tidak anemia (0) dan yang anemia (1) berdasarkan jenis kelamin. Batang biru mewakili laki-laki, sedangkan batang merah mewakili perempuan. Diagram tersebut menunjukkan bahwa ada lebih banyak perempuan yang anemia daripada laki-laki, baik pada kelompok yang tidak anemia maupun yang anemia.

Visualisasi Data Diagram Batang

Berikut ini adalah tampilan visualisasi.

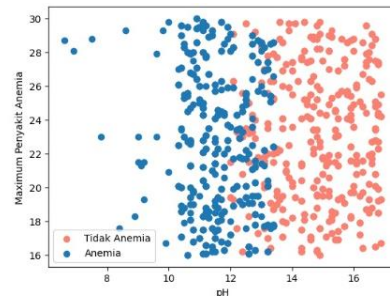


Gambar 2. Diagram

Diagram tersebut menunjukkan klasifikasi prediksi penyakit anemia. Terdapat dua jenis kategori, yakni “Tidak Anemia” dan “Anemia”, yang direpresentasikan oleh batang-batang vertikal dengan dua warna berbeda. Batang-batang berwarna ungu mewakili kategori “Tidak Anemia”, sedangkan batang-batang berwarna merah mewakili kategori “Anemia”. Grafik ini menunjukkan jumlah kasus “Tidak Anemia” dan “Anemia” pada sumbu Y, dan tiga kategori pada sumbu X. Namun, tanpa konteks tambahan, tidak jelas apa yang mewakili tiga kategori pada sumbu X.

Diagram Scater

Berikut ini adalah tampilan data diagram scater pada penyakit anemia.



Gambar 3. Diagram Scater

Pada gambar scater tersebut menunjukkan dua kelompok data yang diwakili oleh titik-titik biru dan merah. Sumbu x mengukur “Hemoglobin” dan sumbu y mengukur “Eritrosit juta sel/mm³”. Titik-titik biru mewakili data untuk “Tidak Anemia” dan titik-titik merah untuk “Anemia”. Dari grafik, terlihat bahwa pasien dengan anemia cenderung memiliki hemoglobin yang lebih rendah dan jumlah eritrosit yang lebih sedikit dibandingkan dengan pasien tanpa anemia.

Pengujian Data dengan 3 Model

Pada penelitian ini menggunakan metode klasifikasi yang akan dilakukan perbandingan dari metode agar klasifikasi diperoleh model dengan nilai hasil akurasi terbaik untuk digunakan dalam proses prediksi penyakit anemia. Penelitian ini menggunakan pembagian data untuk diuji sebanyak bentuk diantaranya pembagian rasio pertama yaitu 90:10, kedua 80:20, ketiga 70:30 dan rasio keempat yaitu 60:40. Pengujian dengan data rasio yang bervariasi untuk mencari serta menentukan model klasifikasi manakah yang paling sering muncul dalam menghasilkan nilai prediksi yang terbaik.

Berikut pembahasan penelitian untuk menemukan model klasifikasi terbaik yang diuraikan dalam bentuk rincian poin sebagai berikut :

Pengujian Data Dengan 3 Model Klasifikasi 90:10

Pengujian data pertama menggunakan rasio 90:10, yaitu data training sebesar 90% dari total data dan untuk data testing sebesar 10% menggunakan 3 model klasifikasi diperoleh hasil seperti pada tabel dibawah ini

Tabel 1. Pengujian Akurasi Data Rasio 90:10

MODEL	Akurasi	Precision	Recall	F1-SCORE
Decision Tree	1.0	1.0	.0	1.0
Naïve Bayes	0.96	0.97	.97	0.97
K-Nearst Neghtbors	0.90	0.92	.91	0.917

Pada table tersebut tampaknya merupakan perbandingan akurasi dan presisi dari tiga model atau data set yang berbeda, Pada uji coba awal dengan pembagian rasio 90:10 menggunakan tiga model klasifikasi, yakni Decision Tree, Naive Bayes, dan K-Nearest Neighbors, hasil evaluasi menunjukkan bahwa Decision Tree unggul dalam kinerja. Decision Tree berhasil mencapai nilai akurasi, presisi, recall, dan F1-Score sebesar 1.0, yang merupakan prestasi tertinggi dibandingkan dengan metode klasifikasi lainnya. Dengan demikian, berdasarkan observasi terhadap tabel pengujian akurasi dengan rasio data 90:10, dapat disimpulkan bahwa Decision Tree menonjol sebagai metode yang efektif, karena mampu mencapai nilai klasifikasi terbaik untuk keempat parameter pengukuran, yakni akurasi, presisi, recall, dan F1-Score.

Pengujian akurasi 80:20

Pada pengujian kedua dengan pembagian data rasio 80:20, terdapat 80% data digunakan untuk pelatihan (training), dan 20% sisanya digunakan untuk pengujian (testing) dari dataset. Hasilnya dapat dilihat dalam tabel berikut:

Tabel 2. Pengujian Akurasi Data Rasio 80:20

MODEL	Akurasi	Precision	Recall	F1-SCORE
Decision Tree	1.0	1.0	.0	1.0
Naïve Bayes	0.95	0.96	.96	0.96
K-Nearst Neghtbors	0.92	0.93	.93	0.93

Tabel pengujian 80:20 merujuk pada metode pembagian data di mana 80% data digunakan untuk pelatihan model dan 20% sisanya digunakan untuk pengujian atau validasi model. Ini membantu dalam mengevaluasi kinerja model dalam hal akurasi, presicision, recall, dan F1-Score. Pada uji coba kedua dengan pembagian rasio 80:20 menggunakan tiga model klasifikasi, yakni Decision Tree, Naive Bayes, dan K-Nearest Neighbors, hasil evaluasi menunjukkan bahwa Decision Tree unggul dalam kinerja. Decision Tree berhasil mencapai nilai akurasi, presisi, recall, dan F1-Score sebesar 1.0, yang merupakan prestasi tertinggi dibandingkan dengan metode klasifikasi lainnya. Dengan demikian, berdasarkan observasi terhadap tabel pengujian akurasi dengan rasio data 80:20, dapat disimpulkan bahwa Decision Tree menonjol sebagai metode yang efektif, karena mampu mencapai nilai klasifikasi terbaik untuk keempat parameter pengukuran, yakni akurasi, presisi, recall, dan F1-Score.

Pengujian akurasi 70:30

Pada pengujian kedua dengan pembagian data rasio 70:30 , terdapat 70% data digunakan untuk pelatihan (training), dan 30% sisanya digunakan untuk pengujian (testing) dari dataset. Hasilnya dapat dilihat dalam tabel berikut:

Tabel 3. Pengujian Akurasi Data Rasio 70:30

MODEL	Akurasi	Precision	Recall	F1-SCORE
Decision Tree	1.0	1.0	.0	1.0
Naïve Bayes	0.94	0.95	.95	0.95
K-Nearst Neghtbors	0.91	0.92	.91	0.91

Tabel pengujian 70:30 merujuk pada metode pembagian data di mana 70% data digunakan untuk pelatihan model dan 30% sisanya digunakan untuk pengujian atau validasi model. Ini membantu dalam mengevaluasi kinerja model dalam hal akurasi, presicision, recall, dan F1-Score. Pada uji coba ketiga dengan pembagian rasio 70:30 menggunakan tiga model klasifikasi, yakni Decision Tree, Naive Bayes, dan K-Nearest Neighbors, hasil evaluasi menunjukkan bahwa Decision Tree unggul dalam kinerja. Decision Tree berhasil mencapai nilai akurasi, presisi, recall, dan F1-Score sebesar 1.0, yang merupakan prestasi tertinggi dibandingkan dengan metode klasifikasi lainnya. Dengan demikian, berdasarkan observasi terhadap tabel pengujian akurasi dengan rasio data 70:30, dapat disimpulkan bahwa Decision Tree menonjol sebagai metode yang efektif, karena mampu mencapai nilai klasifikasi terbaik untuk keempat parameter pengukuran, yakni akurasi, presisi, recall, dan F1-Score.

Pengujian akurasi 60:40

Pada pengujian kedua dengan pembagian data rasio 60:40, terdapat 60% data digunakan untuk pelatihan (training), dan 40% sisanya digunakan untuk pengujian (testing) dari dataset. Hasilnya dapat dilihat dalam tabel berikut:

Tabel 4. Pengujian Akurasi Data Rasio 70:30

Model	Akurasi	Precision	Recal	F1-Score
Decision Tree	1.0	1.0	1.0	1.0
Naive Bayes	0.95	0.96	0.96	0.96
K-Nearest Neightboars	0.90	0.90	0.90	0.90

Tabel pengujian 60:40 merujuk pada metode pembagian data di mana 60% data digunakan untuk pelatihan model dan 40% sisanya digunakan untuk pengujian atau validasi model. Ini membantu dalam mengevaluasi kinerja model dalam hal akurasi, presicision, recall, dan F1-Score. Pada uji coba ketiga dengan pembagian rasio 60:40 menggunakan tiga model klasifikasi, yakni Decision Tree, Naive Bayes, dan K-Nearest Neighbors, hasil evaluasi menunjukkan bahwa Decision Tree unggul dalam kinerja. Decision Tree berhasil mencapai nilai akurasi, presisi, recall, dan F1-Score sebesar 1.0, yang merupakan prestasi tertinggi dibandingkan dengan metode klasifikasi lainnya. Dengan demikian, berdasarkan observasi terhadap tabel pengujian akurasi dengan rasio data 60:40, dapat disimpulkan bahwa Decision Tree menonjol sebagai metode yang efektif, karena mampu mencapai nilai klasifikasi terbaik untuk keempat parameter pengukuran, yakni akurasi, presisi, recall, dan F1-Score.

KESIMPULAN

Berdasarkan hasil pengujian dengan berbagai rasio data, dapat disimpulkan bahwa model Decision Tree memiliki kinerja yang konsisten dan unggul dalam memprediksi penyakit anemia. Decision Tree mencapai nilai akurasi, presisi, recall, dan F1-Score sebesar 1.0 pada semua rasio data yang diuji. Model Naive Bayes juga menunjukkan hasil yang baik, tetapi sedikit lebih rendah dibandingkan Decision Tree, dengan nilai akurasi sekitar 0.94 dan nilai presisi, recall, serta F1-Score sekitar 0.95. Model K-Nearest Neighbors, meskipun mendapatkan nilai yang sedikit lebih rendah, masih dapat dianggap sebagai hasil yang baik dengan akurasi sekitar 0.88-0.90 dan nilai presisi, recall, serta F1-Score sekitar 0.90. Selain itu berdasarkan hasil pengujian dan evaluasi, Decision Tree menjadi model klasifikasi terbaik untuk memprediksi penyakit anemia pada dataset yang digunakan. Model ini menunjukkan konsistensi dalam mencapai nilai akurasi, presisi, recall, dan F1-Score sebesar 1.0 pada berbagai rasio data. Meskipun Naive Bayes juga memberikan hasil yang baik, tetapi Decision Tree unggul dalam mencapai performa tertinggi. K-Nearest Neighbors, meskipun mendapatkan nilai yang lebih rendah, masih memberikan hasil yang baik dan dapat dianggap sebagai pilihan alternatif. Evaluasi model dengan berbagai metrik, seperti akurasi, presisi, recall, dan F1-Score, memberikan gambaran holistik tentang kemampuan model dalam melakukan klasifikasi dan prediksi penyakit anemia.

DAFTAR PUSTAKA

- [1] Sumarno and R. Pambudi, "Sistem Pakar Untuk Mendiagnosa Penyakit Anemia," Pros. Semin. Nas., vol. Volume 2, pp. 128–138, 2014.
- [2] A. M. Sarah, B. Kurniadi, and E. Warsini, "Implementasi Metode Regresi Linear Dalam Memprediksi Penyakit Anemia Secara Dini," J. Tek., vol. 3, no. 1, p. 14, 2023, doi: 10.54314/teknisi.v3i1.1253.
- [3] T. Kan, G. Buyuksalih, Y. Kaya, and A. P. Baskaraca, "The importance of digital methods in preservation of cultural heritage the example of Zirnikli Mansion," Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch., vol. 42, no. 5W1, pp. 175–186, 2017, doi: 10.5194/isprs-Archives-XLII-5-W1-175-2017.
- [4] D. S. Auladi, B. Hidayat, and S. Darana, "Identifikasi dan Klasifikasi Kemurnian Susu Sapi Berdasarkan Pemrosesan Sinyal Video Menggunakan Metode Gabor Wavelet dan Support Vector Machine Identification and Classification of Cow Milk Purenness Based on Video Signal Processing By Using Gabor Wavelet," e-Proceeding Eng., vol. 4, no. 2, pp. 1758–1765, 2017.