

Penerapan *Decision Tree* dan *Naïve Bayes* dalam Perancangan Sistem Prediksi Jenis Golongan Darah

Febrilia Kamila Ahmad & Mia Kamayani

Program Studi Teknik Informatika, Fakultas Teknik – Universitas Prof. Dr. Hamka (UHAMKA)
Jl. Tanah Merdeka No.6, Pasar Rebo, Jakarta Timur, Daerah Khusus Ibukota Jakarta
Telp: (021) 87782739; E-mail: febri9kamila@gmail.com, mia.kamayani@uhamka.ac.id

Abstrak

Golongan darah seseorang dapat diketahui dengan menggunakan uji darah, dapat pula diprediksi melalui tekstur kulit [1] dan kepribadian [2]-[3]. Tujuan penelitian ini adalah menerapkan dan membandingkan performa Decision Tree dan Naïve Bayes untuk prediksi golongan darah serta mengetahui atribut yang signifikan untuk prediksi golongan darah. Selain prediksi dibuat pula antarmuka untuk berinteraksi dengan pengguna. Akurasi prediksi dengan Decision Tree yaitu 75.38% dan Naïve Bayes 55.88%. Untuk meningkatkan performa, dilakukan Rule Induction dengan akurasi yang lebih baik dari Decision Tree dan Naïve Bayes yaitu 76.93%. Atribut yang paling efektif untuk menentukan golongan darah yaitu jenis kulit, jenis kelamin dan jenis kepribadian. Jenis kepribadian yang berpengaruh signifikan adalah 10 dari 13 jenis kepribadian.

Kata kunci: *Golongan Darah, Tekstur Kulit, Kepribadian, Decision Tree, Naïve Bayes, Rule Induction*

Abstract

A person's blood type can be known by using a blood test, it can also be predicted through skin texture [1] and personality [2]-[3]. The purpose of this study was to apply and compare the performance of Decision Tree and Naïve Bayes for blood type prediction and to find out the significant attributes for blood type prediction. In addition to predictions, an interface is also made to interact with users. Prediction accuracy with Decision Tree is 75.38% and Naïve Bayes 55.88%. To improve performance, Rule Induction is performed with better accuracy than Decision Tree and Naïve Bayes, which is 76.93%. The most effective attributes for determining blood type are skin type, gender and personality type. The type of personality that has a significant effect is 10 out of 13 personality types.

Keyword: *Blood Type, Skin Texture, Personality, Decision Tree, Naïve Bayes, Rule Induction*

1 PENDAHULUAN

Golongan darah adalah zat yang dapat memicu respons imun jika asing bagi tubuh. Pemeriksaan golongan darah ABO di Dusun Jambu guna meneliti banyaknya frekuensi golongan darah dan setengah dari 367 penduduk Dusun Jambu belum mengetahui jenis golongan darah mereka [2]. Pemilihan judul berdasarkan metode *Decision Tree* dan *Naïve Bayes* untuk memprediksi kejadian dan pengambilan keputusan golongan darah seseorang berdasarkan tekstur kulit dan kepribadian manusia. Penelitian sebelumnya tentang prediksi jenis golongan darah indikator tekstur kulit berminyak, kering, normal, atau kombinasi [8]. Adapun penelitian-penelitian lain tentang prediksi jenis golongan darah indikator kepribadian yang akan digabungkan dan menghasilkan kesimpulan bahwa kepribadian golongan darah A pekerja keras, berorientasi pada detail, dan sangat terorganisir, golongan darah B seorang individualis dan melakukan sesuatu sesuai keinginan, [1] golongan darah AB seorang yang kreatif, rasional, tenang, terkontrol, kritis,

dan mudah bergaul, dan golongan darah O berambisi, mandiri, percaya diri, punya semangat yang tinggi [9]. Metode *Naïve Bayes* dan *Decision Tree* dipilih sebagai metode prediksi golongan darah karena memiliki akurasi yang baik dan bisa digunakan untuk data kuantitatif maupun kualitatif.

2 LANDASAN TEORI

A. Golongan Darah

Penelitian di sebuah dusun yang dilakukan dengan memeriksa penduduk guna mengetahui jenis golongan darah [2].

B. Tekstur Kulit

Terdapat data klasifikasi golongan darah berdasarkan tekstur kulit yang diambil dari 170 orang [8].

C. Jenis Kepribadian

Penggolongan darah berdasarkan kepribadian seseorang berdasarkan dibagi menjadi empat tipe [9], seperti pada Tabel 1.

Tabel 1 Tabel Jenis Kepribadian.

Golongan Darah	Kepribadian
A	Pekerja keras, berorientasi pada detail, dan sangat terorganisasi.
B	Individualis dan melakukan sesuatu sesuai keinginan.
AB	Kreatif, rasional, tenang, terkontrol, kritis, dan mudah bergaul.
O	Berambisi, mandiri, percaya diri, punya semangat yang tinggi.

D. *Decision Tree* algoritma untuk *supervise* pembelajaran mesin untuk memecahkan masalah regresi dan klasifikasi dengan cara membagi data secara berulang-ulang tergantung pada variabel tertentu. Data dibagi menjadi *node* dan daun pohon mewakili keputusan akhir [12].

- Memilih atribut sebagai akar. Gain tertinggi akan menjadi akar pertama seperti pada rumus (1).

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S) \tag{1}$$

Nilai entropy dapat diperoleh dari rumus (2).

$$Entropy(S) = \sum_{i=0}^n -p_i * \log_2 p_i \tag{2}$$

- Kemudian dilakukan pengulangan metode rumus (1) dan (2) sampai semua data masuk ke kelas yang sama.
- Adapun *RasioGain* sebagai perbandingan antara *Decision Tree* dengan ID3, rumus (3) sebagai berikut:

$$RasioGain(s, j) = \frac{Gain(s, j)}{SplitInfo(s, j)} \tag{3}$$

Keterangan:

s: Himpunan kasus

j: fitur ke-j

Split info dapat diperoleh dengan rumus (4).

$$SplitInfo(s, j) = \sum_{i=1}^k p(V_i | s) \log_2 p(V_i | s) \tag{4}$$

Keterangan:

k: jumlah pemecahan

E. *Naive Bayes* adalah probabilitas dan metode statistik berdasarkan algoritma klasifikasi. Efisiensi komputasi sama dengan

kesederhanaan ini, membuat pendekatannya menarik dan sesuai untuk berbagai bidang seperti yang dijelaskan pada rumus (5).

$$P(A|B) = (p(B|A) * p(A)) / p(B) \tag{5}$$

Keterangan:

P(A|B): Angka *Posterior* A ketika B

p(B|A): Angka *Likelihood* B ketika A

p(A): Angka *Prior* pada kelas A

p(B): Angka *Evidence* suatu kelas

Probabilitas A sebagai B diperoleh rumus (6).

$$\frac{P(C|F_1 \dots F_n)}{P(B_1 \dots B_n)} = \frac{P(A|B_1 \dots B_n) * P(A)}{P(B_1 \dots B_n)} \tag{6}$$

F. *Confusion Matrix*

Metode tabulasi silang dimana data hasil prediksi direpresentasikan pada kolom klasifikasi pada **Tabel 2**. Demikian pula, jumlah contoh positif yang diklasifikasikan secara akurat direpresentasikan sebagai keluaran “TP” Positif Benar. Jumlah sebenarnya contoh negatif diklasifikasikan sebagai positif dilambangkan sebagai istilah “FP” Positif Palsu dan jumlah contoh positif yang sebenarnya diklasifikasikan sebagai negatif dilambangkan sebagai nilai Negatif Palsu “FN”.

Tabel 2 Tabel Confusion Matrix

	Classified Positive	Classified Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

G. *Recall*

Penarikan kembali pada dasarnya adalah rasio deteksi yang benar atas jumlah total sampel uji seperti pada rumus (8).

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

H. *Precision*

Menghitung jumlah relevan dari total seperti pada rumus (9).

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

I. *Accuracy*

Tingkat kedetakan antara nilai prediksi dengan nilai actual seperti pada rumus (10).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

J. *F-Measure*

F-Measure digunakan untuk memilih yang terbaik algoritma mengevaluasi kinerja seperti pada rumus (11).

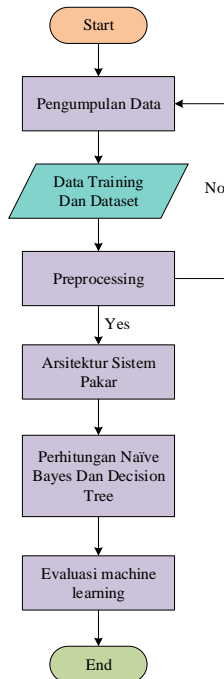
$$F_1 = 2 \cdot \frac{Precision * Recall}{Precision + Recall} \tag{11}$$

K. *Root Mean Squared Error (RMSE)*

Menghitung varian antara data yang diselidiki dengan data akurat yang diamati seperti pada rumus (12)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \tag{12}$$

3 METODOLOGI PENELITIAN



Gambar 1 Alur Penelitian Machine Learning

A. Pengumpulan data

Pengumpulan data dilakukan dengan metode studi literatur dan kuisisioner melalui *Google Form* yang dituju kepada 221 orang yang sudah mengetahui jenis golongan darah mereka.

B. Data Training dan Dataset

Terdapat 221 data survey yang akan diambil 80% sebagai data training (data latih) dan 20% sebagai data testing (data uji) dengan atribut jenis kelamin, tekstur kulit, dan jenis kepribadian.

C. Preprocessing

1) *Data Cleaning*

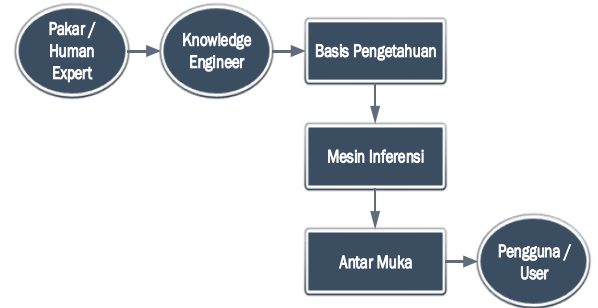
Tahap menemukan kelancaran data *noise*, nilai yang hilang, mengenali

outlier serta mengoreksi ketidakkonsistenan.

2) *Missing Value*

Tahap pengecekan *missing value* pada *dataset* yang digunakan. Untuk mempermudah proses klasifikasi, data polinomial diubah menjadi data binomial.

D. Arsitektur Sistem Pakar



Gambar 2 Arsitektur Sistem Pakar [3]

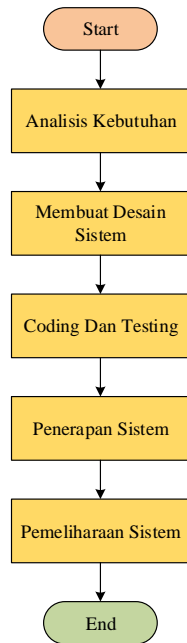
Basis pengetahuan berisi fakta, aturan, dan objek dalam domain-spesifik dan pengetahuan berkualitas tinggi. Pengetahuan diperlukan untuk dipamerkan kecerdasan. Keberhasilan sistem pakar mana pun sangat bergantung pada kumpulan pengetahuan yang sangat akurat dan tepat. Mesin Inferensi yaitu perangkat lunak yang mencocokkan input pengguna dengan data yang ada di basis pengetahuan. Mesin Inferensi akan memberi dan mengatur pengetahuan dari dasar pengetahuan untuk sampai pada solusi tertentu. Antarmuka pengguna yaitu tampilan pertanyaan kepada pengguna dan terima masukan dari mereka. Pengguna sistem pakar tidak harus ahli dalam kecerdasan buatan [3].

E. Perhitungan *Decision Tree* dan *Naive Bayes*

Proses perhitungan algoritma *Decision Tree* terhadap *Data Training* dan *Dataset* dilakukan menggunakan *Software Microsoft Excel* berdasarkan persamaan rumus (1) sampai dengan (4). Sedangkan algoritma *Naive Bayes* berdasarkan persamaan rumus (5) sampai dengan (7).

F. Evaluasi Machine Learning

Terdapat 5 jenis evaluasi *Machine Learning* seperti *Confusion Matrix*, *Recall Precision*, *Accuracy*, *F-Measure*, *Root Mean Squared Error (RMSE)* dimana seluruh proses perhitungan evaluasi dilakukan menggunakan *Software RapidMiner*.



Gambar 3 Alur Perancangan Sistem

Tahap perancangan sistem yang menggunakan metode *Waterfall* sebagai berikut:

- A. Analisis Kebutuhan
Tahap analisis kebutuhan sistem prediksi jenis golongan darah dengan membuat kuisioner dan disebarluaskan kepada orang-orang secara acak serta studi literatur. Tahapan ini menghasilkan data yang berhubungan dengan keinginan *user* terhadap sistem.
- B. Membuat Desain Sistem
Tahap menerjemahkan data kebutuhan *user* sebelum melakukan koding. Proses ini dilakukan menggunakan *Software Balsamic Mockup 3*.
- C. Coding and testing
Proses coding dilakukan menggunakan *Software Text Editor Sublime Text 3* dan tampilan sistem dirancang menggunakan *Framework Bootstrap*. Kemudian dilakukan pengujian atau *testing* terhadap sistem prediksi golongan darah.
- D. Penerapan Sistem
Tahap penerapan sistem prediksi jenis golongan darah oleh *user* melalui *localhost*.
- E. Pemeliharaan Sistem
Dengan rencana penambahan indikator prediksi golongan darah, sistem masih dapat dikembangkan

menjadi lebih variatif dan edukatif dengan tambahan informasi seputar golongan darah.

4 HASIL DAN PEMBAHASAN

Terdapat 177 data training yang akan diolah dengan metode *Naive Bayes* dan *Decision Tree* menggunakan *software Microsoft Excel* dan *RapidMiner* untuk mengetahui *Accuracy*, *Precision*, *Recall*, *F-Measure*, dan *RMSE*. Hasil pengujian ditunjukkan pada Tabel 3 berikut:

Tabel 3 Hasil Evaluasi Machine Learning Naive Bayes dan Decision Tree

	Naive Bayes	Decision Tree
<i>Accuracy</i>	36.36%	51.23%
<i>Precision</i>	29.25%	50.34%
<i>Recall</i>	30.92%	40.28%
<i>F-Measure</i>	30.06%	38.33%
<i>RMSE</i>	0.694	0.649

Sebagai alternatif untuk meningkatkan nilai performa metode *Decision Tree* dan *Naive Bayes* di atas, maka dilakukan Eksperimen *Rule Induction* dimana dilakukan pengujian tingkat performa terbaik antara *Rule Induction*, *Decision Tree*, dan *Naive Bayes* menggunakan Grafik ROC terhadap masing-masing golongan darah A, B, AB, dan O. Dengan memisahkan label *polynomial* pada data uji menjadi label *binomial* menambahkan operator *Bagging* pada *Cross Validation*, maka diperoleh hasil pengujian ditunjukkan pada Tabel 4 berikut:

Tabel 4 Hasil Evaluasi Machine Learning Rule Induction, Naive Bayes dan Decision Tree

	Rule Induction	Decision Tree	Naive Bayes
<i>Accuracy</i>	76.93%	75.38%	55.88%
<i>Precision</i>	79.93%	79.83%	79.07%
<i>Recall</i>	90.78%	87.49%	58.38%
<i>F-Measure</i>	84.00%	83.02%	62.84%
<i>RMSE</i>	0.398	0.401	0.524

5 SIMPULAN

1. Performa *Decision Tree* lebih baik daripada *Naive Bayes*. Hal ini ditunjukkan pada hasil perhitungan *Accuracy* *Naive Bayes* sebesar 36.36% sedangkan *Decision Tree* sebesar 51.23%.

2. Atribut yang signifikan yaitu jenis kelamin dan tekstur kulit, sedangkan jenis kepribadian hanya 10 dari 13 kepribadian yang signifikan untuk memprediksi golongan darah.
3. Pada hasil Eksperimen *Rule Induction*, dengan mengubah label *polynomial* dalam data training menjadi label *binomial* dan menambahkan operator *Bagging* pada *Cross Validation*.
4. *Rule Induction* menghasilkan *Accuracy* 76.93%, *Decision Tree* menghasilkan *Accuracy* 75.38%, dan *Naive Bayes* menghasilkan *Accuracy* 55.88%.

KEPUSTAKAAN

- [1] F. C. Permana, A. C. Padmasari, and S. Sylviani, "Rancang Bangun Aplikasi Pendeteksi Jenis Golongan Darah Berdasarkan Konsep Kepercayaan Rakyat Jepang (Minkan Shinkō)," *Edsence J. Pendidik. Multimed.*, vol. 1, no. 1, 2019, doi: 10.17509/edsence.v1i1.17933.
- [2] R. Pebrina, M. Thomisnancy, B. Sherly, and S. Rassajati, "Pendataan Golongan Darah Warga Dusun Jambu sebagai Upaya Persiapan Pembentukan Desa Siaga Donor Darah," *Semin. Nas. Has. Pengabd. Kpd. Masy. Univ. Ahmad Dahlan*, no. September, pp. 761–768, 2019.
- [3] L. F. Samhan, A. H. Alfarra, and S. S. Abu-Naser, "An Expert System for Knee Problems Diagnosis," *Int. J. Acad. Inf. Syst. Res.*, vol. 5, no. 4, pp. 59–66, 2021, [Online]. Available: www.ijeais.org/ijaisr.
- [4] D. Marutho, "Perbandingan Metode Naive Bayes , KNN , Decision Tree Pada Laporan Water Level Jakarta," *Manaj. Inform. AMIK JTC Semarang*, vol. 15, no. 2, 2019.
- [5] O. D. Madeeh and H. S. Abdullah, "An Efficient Prediction Model based on Machine Learning Techniques for Prediction of the Stock Market," *J. Phys. Conf. Ser.*, vol. 1804, no. 1, 2021, doi: 10.1088/1742-6596/1804/1/012008.
- [6] G. Divva, M. Zulma, and N. Chamidah, "Perbandingan Metode Klasifikasi Naive Bayes , Decision Tree Dan K- Nearest Neighbor Pada Data Log Firewall," no. April, pp. 679–688, 2021.
- [7] J. Zhong and Y. Li, "Design and implementation of intelligent guide system based on LBS," *2019 IEEE Int. Conf. Consum. Electron. - Taiwan, ICCE-TW 2019*, vol. 13, no. 3, pp. 92–103, 2019, doi: 10.1109/ICCE-TW46550.2019.8991682.
- [8] M. Justina Gamache, "What is the role of blood cultures in bacterial pneumonia?," no. January, 2019, [Online]. Available: www.medscape.com/answers/300157-19144/what-is-the-role-of-blood-cultures-in-bacterial-pneumonia.
- [9] A. A. Haqq, "Analisis Sikap Matematis Berdasarkan Golongan Darah," *Unswagati*, 2018.
- [10] D. melina Agustina and Wijanarto, "Analisis Perbandingan Algoritma ID3 Dan C4 . 5 Untuk Klasifikasi Penerima Hibah Pemasangan Air Minum pada PDAM Kabupaten Kendal," *J. Appl. Intell. Syst.*, vol. 1, no. 3, pp. 234–244, 2016.
- [11] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [12] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
- [13] Y. Mohammed, S. Murad, and B. Tahir, "Air Temperature Prediction Using Different Datamining Approaches In Sulaymaniyah City In Iraq," *Passer*, vol. 3, no. 2, pp. 1–9, 2021, doi: 10.24271/psr.21.