

Penerapan Naive Bayes Untuk Prediksi Penyakit Diabetes dengan Menggunakan Rapid Miner

Irwan Reza Pahlevi

Perbanas, Indonesia

E-mail: irwanrezap@gmail.com

Article History:

Received: 17 Maret 2025

Revised: 09 Mei 2025

Accepted: 20 Mei 2025

Keywords: *Diabetes, Naïve Bayes, Data Mining, Prediksi, RapidMiner.*

Abstract: *Diabetes merupakan penyakit kronis yang menjadi salah satu masalah kesehatan utama di Indonesia, dengan jumlah penderita yang terus meningkat setiap tahunnya. Deteksi dini menjadi langkah penting dalam upaya pencegahan dan penanganan penyakit ini. Penelitian ini bertujuan untuk mengimplementasikan algoritma Naïve Bayes dalam memprediksi kemungkinan seseorang menderita diabetes berdasarkan dataset yang diperoleh dari Kaggle. Metode yang digunakan dalam penelitian ini meliputi pengumpulan data, preprocessing, pemodelan menggunakan RapidMiner, serta evaluasi model dengan metrik akurasi, presisi, dan recall. Hasil penelitian menunjukkan bahwa model klasifikasi berbasis Naïve Bayes memiliki tingkat akurasi sebesar 80,52%, yang menunjukkan kinerja yang cukup baik dalam memprediksi diabetes. Dengan demikian, pendekatan ini dapat digunakan sebagai alat bantu dalam mendukung deteksi dini diabetes dan membantu tenaga medis dalam pengambilan keputusan.*

PENDAHULUAN

Berdasarkan Pusat Data Dan Informasi Kementerian Kesehatan Republik Indonesia (Kementerian Kesehatan RI., 2020) bahwa pada tahun 2020 terdapat 10,7 juta orang Indonesia menderita penyakit diabetes sehingga menyebabkan Indonesia menempati peringkat ke 7 di dunia sebagai negara yang memiliki jumlah orang yang menderita diabetes terbanyak. Diabetes adalah penyakit kronis yang menimbulkan gangguan metabolisme dan dikenali melalui kadar gula darah yang tinggi. Ini menjadi penyebab utama masalah jantung, kebutaan dan kegagalan ginjal. (Kementerian Kesehatan RI., 2020). Menurut Mustafa & Simpen (2019), bahwa penyakit diabetes yang juga dikenal sebagai penyakit gula adalah kondisi kronis yang berkembang sebagai akibat dari anomali sekresi insulin dalam peningkatan kadar glukosa yang tidak menentu. Diabetes menyebabkan kenaikan persentase gula pada darah yang meningkatkan risiko konsekuensi seperti penyakit stroke, jantung, kebutaan, gagal ginjal dan kematian. Adapun ia menyebutkan pengertian lain bahwa diabetes adalah sejenis penyakit yang mematikan dan memperburuk kondisi kadar gula darah. Penyebab utamanya terkait dengan gaya hidup yang tidak rajin olahraga dan pola makan yang salah dengan asupan kalori dan lemak tinggi tetapi rendah serat. Kemudian menurut (Siallagan & Fitriyani, 2021) menyatakan bahwa diabetes terjadi disebabkan penambahan angka gula pada darah yang terjadi dalam waktu yang lama.

Kondisi ini dapat berakibat komplikasi serius seperti gagal ginjal, kebutaan, dan serangan jantung jika tidak dikontrol dan diterapi dengan baik.

Adapun teknik diagnosis diabetes yang sering dilakukan adalah dengan pengukuran tingkat glukosa dalam darah. Dimana apabila konsentrasi glukosa dalam darah melampaui batas normal maka orang tersebut akan dikategorikan sebagai penderita diabetes. Namun pada penyakit diabetes terdapat fase asimtomatik yang cukup lama yaitu kondisi dimana penyakit sudah terdeteksi tetapi belum menunjukkan gejala klinis pada pasien (Karyadiputra & Setiawan, 2022).

Antisipasi secara dini merupakan hal yang tepat untuk dapat mencegah penyakit diabetes. Dengan memanfaatkan teknologi informasi dan komunikasi maka saat ini proses diagnosis secara dini penyakit diabetes dapat dilakukan dengan cara mengolah dataset rekam medis orang yang memiliki penyakit diabetes dengan parameter gejala umum dan tidak umum yang ditemukan pada fase awal sampai fase diagnosa dan hasil dari pengelolaan data tersebut bisa digunakan guna membuat suatu sistem prediksi penyakit diabetes. Salah satu cara pengolahan data yang dapat dimanfaatkan dalam mengklasifikasi penyakit diabetes yaitu dengan menerapkan Data mining untuk memperoleh pola pengetahuan yang berguna pada data. Menurut penelitian Purwadi et al., (2019), Data mining yaitu tindakan mendapatkan informasi pola dalam sekumpulan data dengan menerapkan ilmu kecerdasan buatan, teknik statistik dan machine learning guna mendapatkan pengetahuan yang berguna. Proses penggalian informasi yang berasal dari sejumlah besar data menggunakan berbagai pendekatan termasuk pembelajaran mesin, statistik, dan metode matematika disebut data mining.

Salah satu algoritma data mining yaitu Algoritma Naive bayes yang merupakan algoritma yang biasa diterapkan dalam data mining untuk proses klasifikasi yang berguna untuk prediksi kelas data. Algoritma ini melakukan klasifikasi probabilistik yang digunakan untuk menentukan kelas suatu item berdasarkan beberapa fitur yang melekat pada item tersebut. Algoritma ini mengasumsikan bahwa setiap fitur adalah independen antara satu dan lainnya dan memiliki pengaruh yang serupa terhadap kelas akhir (Herfandi et al., 2021). Penggunaan metode Naive bayes sudah banyak digunakan dalam proses klasifikasi baik pada bidang kesehatan, pendidikan, ekonomi dan lain - lain.

Tujuan dari penelitian ini adalah untuk mengimplementasikan algoritma naive dalam mengklasifikasikan dan memprediksi apakah pasien terjangkit penyakit diabetes atau tidak.

LANDASAN TEORI

Gejala Diabetes

Menurut Lestari et al., (2021), beberapa gejala yang terjadi pada orang yang menderita diabetes yaitu sebagai berikut:

- a. Polydipsia adalah keadaan dimana seseorang merasa sangat haus, meskipun sudah minum air atau cairan lainnya.
- b. Polyuria adalah kondisi dimana seseorang mengeluarkan banyak air kecil dalam sehari. Hal ini disebabkan oleh kadar gula darah tinggi, yang menyebabkan ginjal harus bekerja keras untuk mengeluarkan gula dari dalam tubuh.
- c. Polyphagia adalah keadaan dimana seseorang merasa sangat lapar, meskipun sudah makan. Ini terjadi karena tubuh sulit mengubah gula menjadi tenaga sehingga terus meminta makan.
- d. Sudden Weight Loss adalah penurunan berat badan yang drastis dan tiba-tiba. Ini dapat terjadi karena tubuh membakar lemak dan otot sebagai sumber energi, karena gula darah

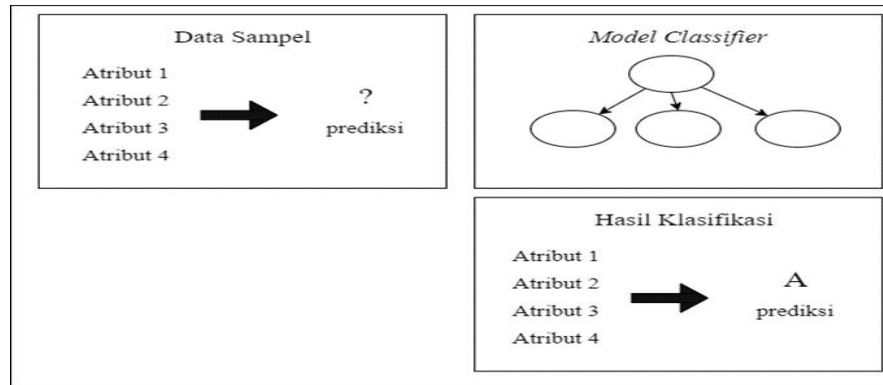
- tidak dapat digunakan dengan benar.
- e. Weakness adalah rasa lelah dan lemah yang tidak wajar. Ini dapat terjadi disebabkan tubuh tidak punya energi dari gula dalam darah.
 - f. Visual Blurring adalah pandangan kabur atau buram. Ini dapat terjadi karena kerusakan pada pembuluh darah dan jaringan mata akibat kadar gula darah yang tinggi.
 - g. Delayed Healing adalah kondisi saat proses penyembuhan luka memakan waktu yang lama atau sering terkena infeksi. Hal ini dapat disebabkan oleh kerusakan pada pembuluh darah dan jaringan akibat tingginya kadar gula darah.
 - h. Acanthosis Nigricans adalah perubahan warna kulit menjadi gelap dan kasar, biasanya pada bagian leher, siku, dan lipatan paha. Ini dapat menjadi tanda dari masalah kesehatan, termasuk diabetes.

Lalu terdapat juga faktor penyebab dari penyebab penyakit diabetes menurut Fatimah yang dikutip pada penelitian Siallagan & Fitriyani (2021) yaitu sebagai berikut :

- a. Obesitas (kegemukan) yaitu orang yang mempunyai indeks massa tubuh yang tinggi (BMI) memiliki risiko lebih tinggi untuk menderita diabetes karena obesitas meningkatkan resistensi insulin dalam tubuh.
- b. Hipertensi yaitu tekanan darah tinggi dapat mempengaruhi kerja insulin dan memicu peningkatan kadar glukosa dalam darah.
- c. Riwayat Keluarga Diabetes Mellitus yaitu faktor genetik memainkan peran penting dalam kemungkinan menderita diabetes, jika ada riwayat keluarga yang menderita diabetes, risiko untuk menderita diabetes meningkat.
- d. Dislipidemia yaitu tingginya kadar lemak dalam darah dapat mempengaruhi kerja insulin dan meningkatkan risiko menderita diabetes.
- e. Umur yaitu bahwa risiko menderita diabetes meningkat seiring bertambahnya usia.
- f. Riwayat Persalinan yaitu bahwa ibu hamil yang memiliki riwayat persalinan dengan berat lahir besar atau memiliki gestational diabetes mempunyai tingkat lebih tinggi untuk terkena diabetes.
- g. Faktor Genetik yaitu beberapa variasi genetik mempunyai tingkat yang lebih tinggi untuk terkena diabetes.
- h. Alkohol dan Rokok yaitu bahwa mengonsumsi alkohol dan merokok dapat meningkatkan risiko menderita diabetes dan memperburuk kondisi jika sudah menderita diabetes.

Klasifikasi

Menurut Wijaya dan Dwiasnati (2020), proses identifikasi suatu fungsi atau model yang dapat menggambarkan dan membedakan data menjadi berbagai kelas atau konsep. Proses ini melibatkan evaluasi terhadap karakteristik dari objek dan menentukan objek pada kelas yang sudah ditetapkan sebelumnya. Contohnya seorang calon penerima beasiswa dikategorikan sebagai layak menerima atau tidak layak menerima. Berikut bagan proses klasifikasi data sampel atau training menggunakan model klasifikasi untuk memprediksi suatu kasus.



Gambar 1. Bagan Proses Klasifikasi

Naive Bayes

Seorang ilmuwan Inggris bernama Thomas Bayes menyampaikan bahwa Naive Bayes adalah algoritma yang memakai metode probabilitas atau kemungkinan dan statistik. Algoritma ini memiliki cara yang sangat sederhana untuk mengklasifikasikan data dengan asumsi sederhana tentang atribut klasifikasi. Naive Bayes sering digunakan dalam memecahkan masalah dalam belajar mesin dan dikenal mempunyai akurasi yang lebih baik meskipun dengan menggunakan perhitungan yang sederhana (Hayuningtyas, 2019). Rumus persamaan Naive bayes yaitu sebagai berikut:

$$p(H|X) = \frac{p(X|H) p(H)}{p(X)}$$

Keterangan:

X = Data Uji

H = Hipotesis Bahwa Data X Termasuk Ke Dalam Kelas Tertentu

P(H|X)= Probabilitas Hipotesis H dengan acuan X

P(H) = Probabilitas anggapan dasar H

P(X|H)= Probabilitas Data X dengan acuan Kondisi Hipotesis H

P(X) = Probabilitas Data X

Confusion Matrix

Menurut Hasanah et al., (2019), Confussion matrix adalah metode penghitungan yang membandingkan hasil klasifikasi dengan data sebenarnya. Matrik ini menunjukkan tingkat akurasi dalam persentase dan berguna sebagai acuan untuk menilai performa algoritma klasifikasi.

Tabel 2.1. Confusion Matrix

Kelas	Terklarifikasi Positif	Terklarifikasi Negatif
<i>Positif</i>	<i>TP (True Positif)</i>	<i>FN (False Negatif)</i>
<i>Negatif</i>	<i>FP (False Positif)</i>	<i>TN (True Negatif)</i>

Sumber : (Marcoulides, 2005)

Menurut Ainurrohman (2021), Confusion Matrix dalam kinerjanya dapat dinilai dengan empat parameter: TP, FN, FP, dan TN. TP (True Positive) adalah sejumlah data yang terbukti benar dan dideteksi oleh sistem sebagai benar. FN (False Negative) adalah sejumlah data yang terbukti salah dan dideteksi oleh sistem sebagai salah. FP (False Positive) adalah sejumlah data terbukti benar namun dideteksi oleh sistem sebagai salah. TN (True Negative) ialah sejumlah data terbukti salah dan dideteksi oleh sistem sebagai benar.

Rumus untuk menentukan tingkat keakuratan sistem klasifikasi data adalah dengan menjumlahkan True Positive dan True Negative, lalu membagi jumlah tersebut dengan total data yang digunakan dalam sistem klasifikasi. Hasil dari perhitungan ini dapat dinyatakan dalam bentuk persentase.

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

Rumus untuk mengukur sejauh mana sistem memperoleh nilai kebenaran (positif) dari dataset yang bernilai benar (positif) dengan tepat. Semakin banyak nilai salah negatif dalam hasil yang dihasilkan oleh sistem maka nilai recall yang diperoleh semakin kecil (Nurjanah et al., 2017).

$$TPRate (Recall) = \frac{TP}{TP+FN}$$

Rumus untuk mengukur tingkat keakuratan dalam klasifikasi sesuai dengan nilai kebenaran yang sebenarnya. Semakin banyak kesalahan yang terjadi pada hasil klasifikasi, maka semakin rendah tingkat precision yang dihasilkan (Nurjanah et al., 2017).

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

Keterangan :

TP : *True Positif*

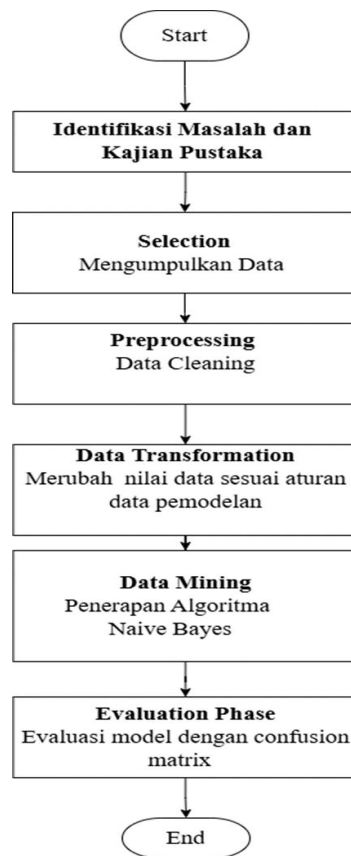
TN : *True Negatif*

FN : *False Negatif*

FP : *False Positif*

METODE PENELITIAN

Dalam melakukan analisa terkait penerapan naive bayes dengan menggunakan rapid miner, penulis mempunyai beberapa prosedur tahapan analisa yang ditunjukkan pada alur dibawah ini:



Gambar 2. Alur Analisa

Penjelasan dari prosedur Analisa yang terdapat pada alur tersebut diatas antara lain:

1. Identifikasi masalah dan tujuan

Pada tahapan ini penulis melakukan identifikasi masalah yang akan diangkat dan tujuan penelitian

2. Pengumpulan data

Pada tahapan selanjutnya adalah tahapan pencarian dan pengumpulan dataset penyakit diabetes dari Kaggle dan melakukan identifikasi dataset yaitu kegiatan mengidentifikasi jenis atribut yang akan digunakan.

3. Tahapan Preprocessing dan Data Transformation

Pada tahap ini dilakukan pembersihan data dengan membersihkan data atau mengganti data yang tidak valid, kosong, dan duplikat menjadi data yang siap untuk diolah. Jika data telah sesuai maka tahapan ini dilewat.

4. Tahapan Data Mining atau Pemodelan

Tahapan ini merupakan tahapan penerapan algoritma yang dipilih untuk melakukan pemodelan. Adapun Teknik pemodelan pada penelitian ini menerapkan algoritma Naive Bayes.

5. Tahapan Evaluasi

Pada tahapan ini melakukan evaluasi model apakah hasil pemodelan sudah sesuai dengan tujuan

pada fase awal. Langkah dalam tahapan ini adalah dengan melakukan uji accuracy, recall, precision.

HASIL DAN PEMBAHASAN

Pada tahapan pemahaman bisnis terdapat dua aktifitas, yaitu identifikasi masalah dan kajian pustaka. Adapun penulis menemukan permasalahan bahwa pada tahun 2020 terdapat 10,7 juta orang Indonesia menderita penyakit diabetes sehingga menyebabkan Indonesia menempati peringkat ke 7 di dunia sebagai negara yang memiliki jumlah orang yang menderita diabetes terbanyak. Maka dari hal tersebut penulis tertarik meneliti masalah diabetes. Setelah itu penulis melakukan kajian pustaka untuk memperoleh pemahaman lebih mendalam mengenai masalah tersebut dapat diberikan solusi data mining dengan memprediksi lebih dini penyakit menggunakan teknik klasifikasi dan algoritma naive bayes.

Setelah melakukan kajian pustaka maka selanjutnya penulis melakukan pengumpulan data yang dibutuhkan yang berasal dari kaggle berupa data penyakit diabetes. Berikut adalah sebagian data yang dipergunakan pada penelitian ini

Tabel 1 Data

Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Label
6	148	72	35	0	33,6	0,627	50	1
1	85	66	29	0	26,6	0,351	31	0
8	183	64	0	0	23,3	0,672	32	1
1	89	66	23	94	28,1	0,167	21	0
0	137	40	35	168	43,1	2,288	33	1
5	116	74	0	0	25,6	0,201	30	0
3	78	50	32	88	31	0,248	26	1
10	115	0	0	0	35,3	0,134	29	0
2	197	70	45	543	30,5	0,158	53	1
8	125	96	0	0	0	0,232	54	1
4	110	92	0	0	37,6	0,191	30	0
10	168	74	0	0	38	0,537	34	1
10	139	80	0	0	27,1	1,441	57	0
1	189	60	23	846	30,1	0,398	59	1
5	166	72	19	175	25,8	0,587	51	1
7	100	0	0	0	30	0,484	32	1
0	118	84	47	230	45,8	0,551	31	1
7	107	74	0	0	29,6	0,254	31	1
1	103	30	38	83	43,3	0,183	33	0
1	115	70	30	96	34,6	0,529	32	1
3	126	88	41	235	39,3	0,704	27	0
8	99	84	0	0	35,4	0,388	50	0
7	196	90	0	0	39,8	0,451	41	1
9	119	80	35	0	29	0,263	29	1
11	143	94	33	146	36,6	0,254	51	1

Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Label
10	125	70	26	115	31,1	0,205	41	1
...	...	...	...	...	...	...	...	...
1	93	70	31	0	30,4	0,315	23	0

Preprocessing

Tahapan preprocessing yaitu tahapan mengolah dataset mentah yang telah diperoleh menjadi data yang siap untuk diolah atau dimodelkan. Tahapan ini meliputi tahapan data cleaning, data transformation, normalisasi dan data reduction. Namun dalam penelitian ini dataset yang telah diperoleh sudah lengkap tidak ada nilai yang hilang (missing values), atribut yang digunakan sudah cukup untuk membangun model dan semua nilai sudah numerik maka penulis dapat menyederhanakan tahapan pengolahan data dan hanya memerlukan normalisasi untuk data siap untuk dimodelkan dengan algoritma seperti Naive Bayes. Normalisasi nilai variabel digunakan agar terjadi keseimbangan nilai perbandingan antara nilai data sebelum dinormalisasi dengan nilai data yang telah dinormalisasi. Pada penelitian ini menggunakan range dari 0 dan 1 yang berarti nilai variabel setiap atribut tidak boleh bernilai lebih dari 1 dan tidak boleh kurang dari 0. Berikut adalah hasil normalisasi nilai atribut asal sekolah dengan min-max Normalization.

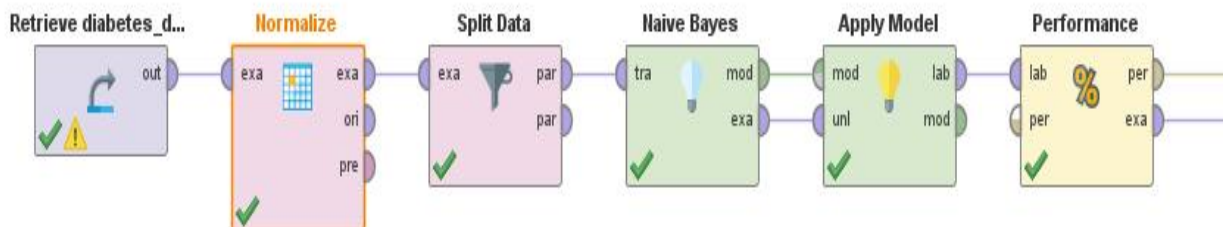
Row No.	Label	Pregnancies	Glucose	BloodPress...	SkinThickne...	Insulin	BMI	DiabetesPe...	Age
1	Ya	0.353	0.744	0.590	0.354	0	0.501	0.234	0.483
2	Tidak	0.059	0.427	0.541	0.293	0	0.396	0.117	0.167
3	Ya	0.471	0.920	0.525	0	0	0.347	0.254	0.183
4	Tidak	0.059	0.447	0.541	0.232	0.111	0.419	0.038	0
5	Ya	0	0.688	0.328	0.354	0.199	0.642	0.944	0.200
6	Tidak	0.294	0.583	0.607	0	0	0.382	0.053	0.150
7	Ya	0.176	0.392	0.410	0.323	0.104	0.462	0.073	0.083
8	Tidak	0.588	0.578	0	0	0	0.526	0.024	0.133
9	Ya	0.118	0.990	0.574	0.455	0.642	0.455	0.034	0.533
10	Ya	0.471	0.628	0.787	0	0	0	0.066	0.550
11	Tidak	0.235	0.553	0.754	0	0	0.560	0.048	0.150
12	Ya	0.588	0.844	0.607	0	0	0.566	0.196	0.217
13	Tidak	0.588	0.698	0.656	0	0	0.404	0.582	0.600
14	Ya	0.059	0.950	0.492	0.232	1	0.449	0.137	0.633

ExampleSet (769 examples, 1 special attribute, 9 regular attributes).

Gambar 3 Hasil Normalisasi

Tahapan Pemodelan

Dalam proses pemodelan penelitian ini akan menggunakan rapid miner dalam membantu proses pemodelan. Berikut adalah langkah – langkah proses pemodelan menggunakan rapid miner:



Gambar 4 Proses Klasifikasi

Gambar 4 menunjukkan proses pemodelan pada rapid miner dengan menggunakan algoritma naive bayes. Langkah – langkah lebih detail dari proses pada Gambar 5 yaitu sebagai berikut :

1. Langkah pertama adalah membuka software Rapidminer kemudian pilih menu untuk membuat proses baru atau new process.
2. Langkah kedua yaitu memasukan data yang akan diproses dengan mengklik import data kemudian cari lokasi dari data diabetes. label menjadi Label (atribut target).
3. Langkah ketiga yaitu seret data yang telah dimasukan di bagan repository tersebut kedalam bagan process. Lalu gunakan operator normalize untuk membuat range nilai data seragam.
4. Selanjutnya masukan operator split data untuk membagi data menjadi data testing dan data training dengan perbandingan 1:9 untuk data uji dan data latih. Hubungkan operator ini dengan operator naive bayes dan apply model.
5. Setelah itu masukan operator performance dan hubungkan dengan operator apply model untuk mengetahui performa model berdasarkan accuray, precission, recall.
6. Langkah terakhir adalah mengklik tombol play untuk melakukan proses klasifikasi.

Tahapan Evaluasi

Tahapan evaluasi dalam adalah langkah di mana penulis mengukur kinerja model yang telah dibangun untuk menentukan seberapa baik model tersebut dapat memprediksi atau mengklasifikasikan data baru. Adapun pada proses pemodelan rapid miner dapat dilihat terdapat operator performance untuk mengetahui performa model berdasarkan accuray, precission dan recall. Berikut adalah hasilnya model performa pada penelitian ini:

accuracy: 80.52%			
	true Ya	true Tidak	class precision
pred. Ya	16	7	69.57%
pred. Tidak	8	46	85.19%
class recall	66.67%	86.79%	

Gambar 5 Model Performance

Gambar 5 menunjukkan hasil *model performance* penelitian ini dimana menunjukkan hasil kinerja akurasi yang baik dalam memprediksi penyakit diabetes dengan akurasi 80,52% yang menunjukkan bahwa model dapat mengklasifikasikan pasien diabetes dengan benar.

KESIMPULAN

Berdasarkan hasil penelitian ini mengenai implementasi algoritma Naive Bayes dalam prediksi penyakit diabetes menunjukkan hasil kinerja akurasi yang baik dengan akurasi 80,52%. Hal ini menunjukkan bahwa model dapat mengklasifikasikan pasien diabetes dengan benar dalam sebagian besar kasus. Adapun hasil pemodelan ini dapat menjadi alat yang berguna dalam membantu mendeteksi dan memprediksi penyakit diabetes dimana saat ini prevalensi diabetes yang tinggi di Indonesia seperti yang dijelaskan dalam bgaian permasalahan. Penggunaan algoritma Naïve Bayes dapat menjadi solusi dann mendukung upaya pencegahan dan pengelolaan penyakit diabetes dengan memberikan prediksi yang cepat dan efisien.

DAFTAR REFERENSI

- Ainurrohmah. (2021). Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka. PRISMA, Prosiding Seminar Nasional Matematika, 4, 493–499. <https://doi.org/ISSN2613-9189>
- Boussaa, M., Atouf, I., Atibi, M., & Bennis, A. (2016). ECG signals classification using MFCC coefficients and ANN classifier. Proceedings of 2016 International Conference on Electrical and Information Technologies, ICEIT 2016, 5(4), 480–484. <https://doi.org/10.1109/EITech.2016.7519646>
- Hasanah, R. L., Hasan, M., Pangesti, W. E., Wati, F. F., & Gata, W. (2019). Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor). Jurnal Techno Nusa Mandiri, 16(1), 1–6. <https://doi.org/10.33480/techno.v16i1.25>
- Hayuningtyas, R. Y. (2019). Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita. 6(1), 18–22.
- Herfandi, H., Zaen, M. T. A., Yuliadi, Y., Julkarnain, M., & Hamdani, F. (2021). Application of Information Gain to Select Attributes in Improving Naïve Bayes Accuracy in Predicting Customer's Payment Capability. JISA(Jurnal Informatika Dan Sains), 4(2), 155–163. <https://doi.org/10.31326/jisa.v4i2.1044>
- Karyadiputra, E., & Setiawan, A. (2022). Penerapan data mining untuk prediksi awal kemungkinan terindikasi diabetes. 221–232.
- Kementerian Kesehatan RI. (2020). Infodatin tetap produktif, cegah, dan atasi Diabetes Melitus 2020. In Pusat Data dan Informasi Kementerian Kesehatan RI (pp. 1–10).
- Lestari, Zulkarnain, Sijid, & Aisyah, S. (2021). Diabetes Melitus: Review Etiologi, Patofisiologi, Gejala, Penyebab, Cara Pemeriksaan, Cara Pengobatan dan Cara Pencegahan. UIN Alauddin Makassar, 1(2), 237–241. <http://journal.uin-alauddin.ac.id/index.php/psb>
- Marcoulides, G. A. (2005). Discovering Knowledge in Data: an Introduction to Data Mining. In Journal of the American Statistical Association (Vol. 100, Issue 472). <https://doi.org/10.1198/jasa.2005.s61>
- Mustafa, M. S., & Simpen, I. W. (2019). Implementasi Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba. Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi, VIII(1), 1–10. <https://ejurnal.dipaneegara.ac.id/index.php/sisiti/article/view/1-10>
- Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. 1(12), 1750–1757.
- Purwadi, P., Ramadhan, P. S., & Safitri, N. (2019). Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang. Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer), 18(1), 55. <https://doi.org/10.53513/jis.v18i1.104>
- Siallagan, R. A., & Fitriyani. (2021). Prediksi Penyakit Diabetes Mellitus. Jurnal Responsif, 3(1), 45–46.
- Wijaya, H. D., & Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. Jurnal Informatika, 7(1), 1–7. <https://doi.org/10.31311/ji.v7i1.6203>