

## DATA MINING UNTUK KLASIFIKASI PENDERITA DIABETES MENGUNAKAN ALGORITMA NAÏVE BAYES

Celvin Juliansya[1], Ahmad Marsehan[2]

Program Studi Teknologi Informasi, Universitas PGRI Silampari  
e-mail: [celvinjuliansya5@gmail.com](mailto:celvinjuliansya5@gmail.com)

**Abstract-** Diabetes merupakan penyakit metabolik kronik yang ditandai dengan meningkatnya kadar glukosa darah (atau gula darah), yang seiring waktu dapat mengakibatkan kerusakan serius pada jantung, pembuluh darah, mata, ginjal, dan saraf. Diabetes tipe 2, yang paling sering terjadi pada orang dewasa, muncul ketika tubuh menentang terhadap insulin atau tidak menghasilkan insulin yang memadai. Kurun tiga dekade belakang, prevalensi diabetes tipe 2 telah meningkat secara signifikan di negara-negara dengan berbagai tingkat pendapatan. Diabetes tipe 1, sebelumnya dikenal sebagai diabetes remaja atau diabetes yang tergantung insulin, adalah kondisi kronis di mana pankreas memproduksi sedikit atau tidak sama sekali insulin. Bagi penderita diabetes, akses terhadap pengobatan yang terjangkau, termasuk insulin, berpengaruh untuk keberlangsungan hidup mereka. Ada target global yang disetujui untuk mengontrol peningkatan diabetes dan obesitas pada tahun 2025. Saat ini, sekitar 422 juta orang di seluruh dunia menderita diabetes, dengan sebagian besar penderita tinggal di negara-negara berpenghasilan rendah dan menengah. Setiap tahun, diabetes menyebabkan sekitar 1,5 juta kematian[1]. Jumlah kasus dan prevalensi diabetes terus meningkat selama beberapa dekade. Dalam penelitian untuk klasifikasi diabetes, digunakan algoritma Naïve Bayes. Naïve Bayes merupakan metode klasifikasi data berdasarkan peluang yang mungkin terjadi di masa depan. Alat yang digunakan untuk mengevaluasi akurasi adalah RapidMiner versi 10. Riset ini memakai metode Knowledge Discovery in Database (KDD). Tujuannya adalah menentukan apakah hasil pemeriksaan pasien menunjukkan positif atau negatif diabetes. Hasil penelitian menunjukkan tingkat akurasi sebesar 71,14%. Pada dataset penderita diabetes yang dianalisis dengan algoritma Naïve Bayes, kategori outcome (target) "0" (NO) lebih banyak dibandingkan dengan outcome (target) "1" (YES).

**Keywords:** Klasifikasi, Diabetes, Naive Bayes, Kdd

### PENDAHULUAN

Diabetes mellitus, atau sering disebut sebagai diabetes, terjadi akibat masalah pada pankreas yang memproduksi hormon insulin, yang berperan dalam membantu tubuh menggunakan glukosa sebagai sumber energi[1]. Diabetes mellitus dapat diklasifikasikan menjadi tipe 1 dan tipe 2, tergantung pada faktor penyebabnya.

Penyebab diabetes seringkali terkait dengan gaya hidup, seperti kurangnya aktivitas fisik, stres, dan konsumsi makanan tinggi gula, yang berkontribusi signifikan terhadap perkembangan penyakit ini. Selain itu, faktor genetik dan obesitas yang tidak dikelola dengan baik juga berperan besar dalam meningkatkan risiko diabetes tipe 2[2].

Diabetes mellitus merupakan penyakit kronik yang dapat mengakibatkan masalah kesehatan lebih lanjut, seperti gangguan penglihatan mata, katarak, dan masalah jantung. Prevalensi diabetes mellitus di Indonesia mencapai 1,5% berdasarkan diagnosa dokter, dan 2,1% berdasarkan gejala[3].

Penyakit ini biasanya dikatakan sebagai “penyakit kencing manis” karena gejala yang paling sering

ditemukan adalah ketidakpastian kencing. Penderita diabetes mellitus biasanya harus mengikuti diet seimbang dan olah raga teratur, serta menggunakan obat-obat anti-diabetes untuk mengatur kadar glukosa dalam darah[4].

Data mining adalah proses mengumpulkan, mengolah, dan menganalisis data untuk mendapatkan informasi yang berguna. Klasifikasi merupakan salah satu metode dalam data mining yang digunakan untuk menentukan hasil dari dataset yang diperoleh ke dalam suatu kelas tertentu. Pada penelitian terkait penerapan data mining pada klasifikasi penyakit diabetes mellitus, Fungsi dari klasifikasi data mining dalam penelitian ini ialah untuk memprediksi class yang dicari pada data penderita diabetes. Algoritma Naïve Bayes digunakan untuk mendapatkan akurasi yang tinggi yang dimana hasil dari penelitian tersebut digunakan untuk mencari target atau menentukan positif atau negatif dari pemeriksaan pasien dari dataset tersebut.

Penulis melakukan penelitian secara ilmiah dengan metode yang digunakan adalah klasifikasi algoritma Naïve Bayes yang memiliki kelebihan bekerja dengan baik meskipun dengan dataset yang relatif kecil,

karena ia mengandalkan estimasi probabilitas yang mudah dihitung dari data terbatas. agar diketahui nilai Perbandingan antara kambuh atau tidak kambuhnya penyakit kanker tersebut. Nantinya hasil dari klasifikasi ini nantinya dapat dimanfaatkan untuk mengetahui jumlah pasien penderita penyakit Diabetes. Kesimpulan yang dapat ditarik dari hasil penelitian Nadya Meilani, Odi Kurniawan yaitu pasien pemeriksaan penyakit Diabetes dengan outcome "0" (NO) lebih banyak dari pada "1" (YES). Analisa yang dilakukan dengan menggunakan algoritma K-NN Menunjukkan tingkat keakuratan dari perhitungannya yaitu cukup besar dapat dilihat Dari persentase yang mencapai lebih dari 70%. Dengan nilai sebesar 72,62% bisa Juga disebabkan oleh kurangnya kekompleksan data yang menyebabkan model Dapat memprediksi secara akurat. Berdasarkan jurnal terdahulu yang dikutip dapat Diketahui tingkat keakuratan algoritma K-NN menunjukkan persentase yang Cukup besar maka dibuatlah penelitian mengenai ketepatan algoritma NAÏVE BAYES pada dataset melakukan penelitian kanker payudara untuk mengetahui apakah algoritma NAÏVE BAYES Dapat menunjukkan tingkat keakuratan lebih tinggi daripada algoritma K-NN [5].

Algoritma Naïve Bayes adalah salah satu teknik umum dalam data mining yang sering dipakai untuk pengklasifikasian data. Algoritma Naïve Bayes memiliki keunggulan berupa tingkat kesalahan yang lebih rendah ketika digunakan pada dataset yang besar. Selain itu, akurasi dan kecepatan algoritma ini juga lebih tinggi saat diaplikasikan pada dataset dengan jumlah yang lebih besar

Dalam data mining, klasifikasi digunakan untuk memprediksi class yang dicari secara akurat pada setiap kasus dalam data. Pada penelitian metode klasifikasi algoritma K-NN dalam jurnal rujukan digunakan untuk memprediksi penyakit kanker payudara menghasilkan persentase hasil nilai sebesar 72,62% [6].

## METODOLOGI PENELITIAN

### A. Data Mining

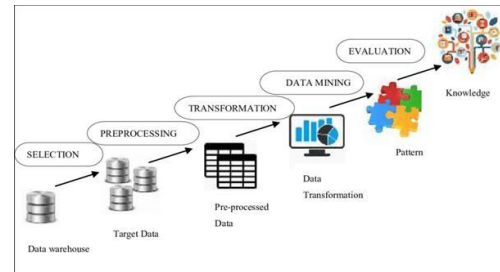
Data mining adalah proses mengumpulkan data bermakna dari sejumlah besar data atau big data. Proses ini sering menggunakan berbagai metode seperti matematika, statistik, dan teknologi kecerdasan buatan. Istilah lain yang digunakan untuk data mining adalah Knowledge Discovery in Database (KDD) dan Data Analysis.

Penelitian ini menggunakan metode KDD (Knowledge Discovery in Database Process).

Knowledge Discovery in Database Process (KDD) adalah salah satu metode yang digunakan untuk data mining. Fayyad et al. (1996) mendefinisikan KDD sebagai proses awal yang menggunakan teknik data mining untuk menemukan

informasi berharga dan pola dalam data, termasuk algoritma untuk mengidentifikasi pola-pola tersebut. Dunham (2003) merangkum tahapan-tahapan dalam proses KDD sebagai berikut: pemilihan data, pemrosesan data, transformasi data, penggalian data, serta interpretasi dan evaluasi akhir.

Berikut deskripsi dan detail proses KDD:



Gambar 1.  
Sumber : (Beon 2020)

1. **Data Cleansing:** Proses di mana data diolah dan dipilih data yang dianggap bisa digunakan.
2. **Data Integration:** Proses menggabungkan data yang dianggap berulang menjadi satu kesatuan.
3. **Selection:** Proses memilih data yang relevan untuk analisis.
4. **Data Transformation:** Proses mengubah data terpilih ke dalam bentuk yang sesuai untuk prosedur mining.
5. **Data Mining:** Proses menerapkan berbagai teknik untuk mengekstrak pola-pola potensial yang menghasilkan data berguna.
6. **Pattern Evaluation:** Proses menilai pola-pola yang telah diidentifikasi berdasarkan ukuran tertentu.
7. **Knowledge Presentation:** Tahap terakhir dalam proses KDD, di mana data yang sudah diproses divisualisasikan agar lebih mudah dipahami oleh pengguna dan memungkinkan pengambilan tindakan berdasarkan analisis tersebut [7].

### B. Klasifikasi

Klasifikasi adalah metode yang digunakan ketika atributnya berupa numerik atau nominal, namun labelnya harus nominal. Dalam metode data mining ini, dilakukan pengelompokan atau pengklasifikasian berdasarkan hubungan antara variabel kriteria dengan variabel target [8].

### C. Algoritma Naïve Bayes

Naïve Bayes adalah algoritma machine learning yang digunakan untuk klasifikasi atau pengelompokan data. Algoritma ini berdasarkan pada teorema probabilitas yang diperkenalkan oleh Ilmuwan Inggris Thomas Bayes mengembangkan Naïve Bayes, yang berfungsi memprediksi probabilitas kejadian di masa depan berdasarkan pengalaman dari data sebelumnya, sehingga berguna untuk pengambilan keputusan [9].

Rumus Algoritma Naive bayes

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

## HASIL DAN PEMBAHASAN

### A. Implementasi Klasifikasi Penderita Diabetes Menggunakan Algoritma Naive Bayes.

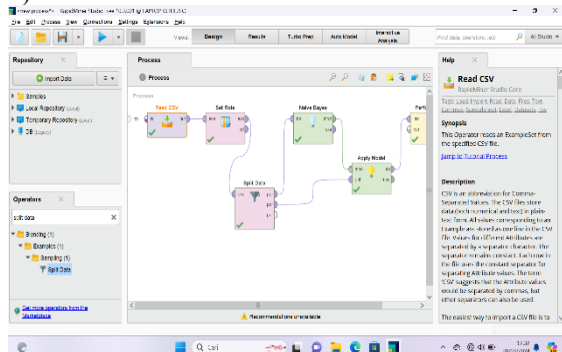
#### 1) Data Selection

Proses select data ini dilakukan dengan memilih data yang akan digunakan dan kemudian akan dianalisis yang nantinya diharapkan akan memberikan keterangan penilaian pada Penderita Diabetes Dengan Kategori "0"(NO) atau "1"(YES).

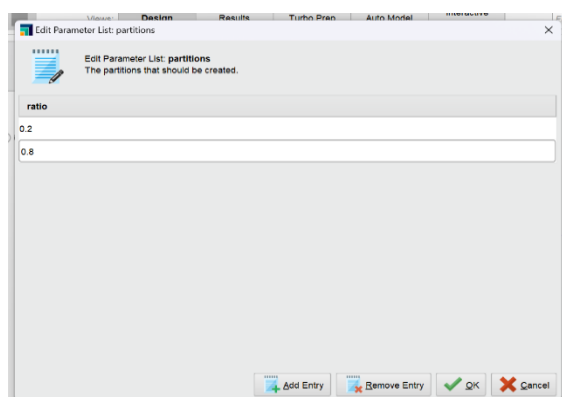
Tabel 1. Dataset Kanker Payudara

|   | A         | B       | C         | D          | E       | F        | G         | H   | I       |
|---|-----------|---------|-----------|------------|---------|----------|-----------|-----|---------|
| 1 | Pregnanci | Glucose | BloodPres | SkinThickr | Insulin | BMI      | DiabetesP | Age | Outcome |
| 2 | 9         | 120     | 72        | 22         | 56      | 20.08    | 0,509028  | 48  | 0       |
| 3 | 1         | 71      | 62        | 0          | 0       | 21.08    | 0,288889  | 26  | 0       |
| 4 | 8         | 74      | 70        | 40         | 49      | 35.03.00 | 0,489583  | 39  | 0       |

#### 2) Data Transformation



Gambar 2. Mengatur Target Role pada Operator Set Role

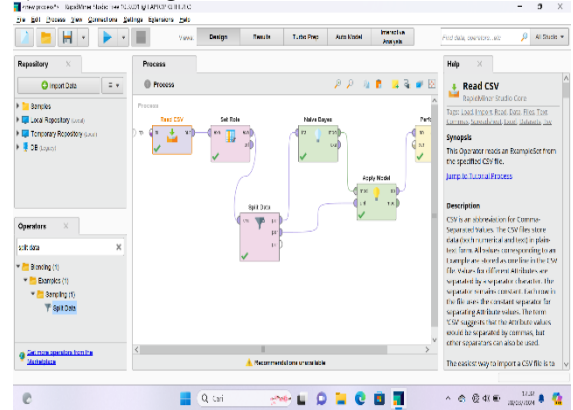


Gambar3. Mengatur Parameter pada Operator Set Data

Pada tahap ini data akan diubah ke bentuk yang sesuai dengan data mining, dan nantinya akan diberikan atribut agar skala pengukuran data asli biasa diubah kedalam bentuk lain. Atribut yang

digunakan adalah polynominal dan integer . Kemudian diatur role label pada atribut Class. Pada operator Split Data juga diatur agar dapat membagi dataset menjadi data training sebesar 80% dan data testing sebesar 20%.

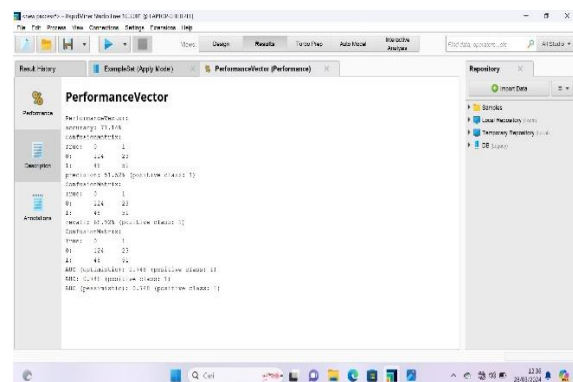
#### 3) Data Mining



Gambar 4. Proses Data Mining

Pada tahap ini akan dijalankan proses rancang desain dengan menggunakan aplikasi Rapidminer dan diterapkan model klasifikasi Naive Bayes untuk menghasilkan pengklasifikasian dan tingkat akurasi performancnya.

#### 5) Interpretation/Evaluation



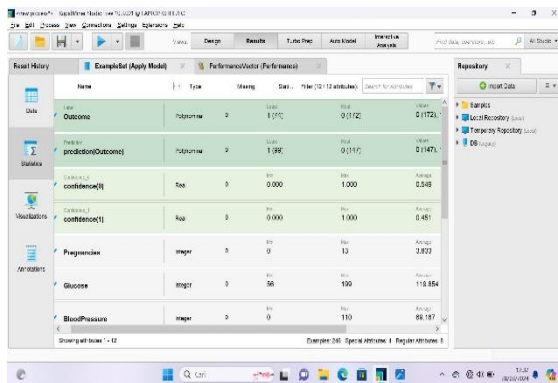
Gambar 5. Hasil Performance Vector

Hasil dari klasifikasi algoritma Naive Bayes dengan menggunakan dataset kanker payudara terdapat 246 sample data dengan 2 klasifikasinya dikategorikan kedalam "0"(NO) atau "1"(YES).

#### B. Akurasi Algoritma Naive Bayes

Hasilnya disini berisikan tentang hasil dari penelitian yang disesuaikan dengan tujuan penulisan jurnal dan mengacu pada tahapan perancangan.

## 1. Statistik Data

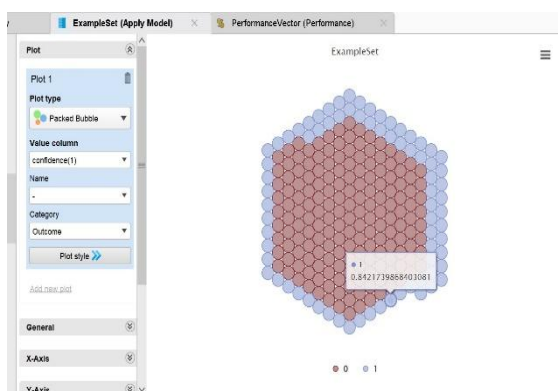


Gambar 6. Statistik Data Diabetes

Berdasarkan Gambar 4 dan 5 diketahui bahwa:

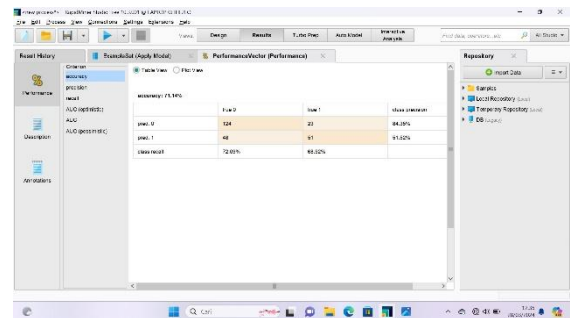
- Data label outcome dengan tipe polynominal terdapat 246 data, dengan rinciannya kedalam "0"(NO) sebanyak 172 data dan "1"(YES) 74 data.
- Data prediksi dengan tipe polynominal terdapat 246 data, dengan rinciannya kedalam "0"(NO) sebanyak 147 data dan "1"(YES) 99 data.
- "0"(NO) dengan nilai terkecil 0, terbesar 1,000 dan rata-rata yang didapat adalah sebesar 0,549.
- "1"(YES) dengan nilai terkecil 0, terbesar 1,000 dan rata-rata yang didapat adalah sebesar 0,451.

## 2. Visualisasi Data



Hasil dari visualisasi data dengan plot type packed bubble menghasilkan example set apply model data bahwa "0"(NO) lebih mendominasi daripada "1"(YES).

## 3. Nilai Akurasi Performance



Hasil dari akurasi performance vector yaitu sebesar 71,14% dengan rincian class precision adalah pred "0"(NO) 84,35%, pred "1"(YES) 51,52% dan rincian class recall true "0"(NO) sebesar 72,09% dan true "1"(YES) sebesar 68,92%.

Bagian ini merupakan tempat untuk menyajikan hasil penelitian secara rinci, jelas, dan teratur. Hasil penelitian disampaikan dalam bentuk tabel, grafik, atau ilustrasi lainnya, serta dilengkapi dengan pembahasan yang terstruktur dan sistematis. Penjelasan mengenai performa, kelemahan, dan kelebihan dari hasil penelitian harus disertakan.

## KESIMPULAN

Dapat disimpulkan bahwa dari data awal sebanyak 246 yang dibagi menjadi 20% data training dan 80% data testing dengan model algoritma klasifikasi Naive Bayes pada aplikasi Rapidminer versi 10 menghasilkan nilai accuracy sebesar 71,14% yang menunjukkan bahwa hasil klasifikasinya baik, dimana prediksi penderita Diabetes dengan outcome "1"(YES) lebih sedikit dibandingkan dengan prediksi penderita Diabetes dengan outcome "0"(NO).

## REFERENSI

- [1] WHO:Diabetes.(n.d.). <https://www.who.int/health-topics/diabetes>
- [2] Chitra Meytarizqi, C. M. (2021). FAKTOR-FAKTOR SOSIODEMOGRAFI PADA KEJADIAN DIABETES MELLITUS TIPE II (Doctoral dissertation, Poltekkes Kemenkes Yogyakarta). <http://eprints.poltekkesjogja.ac.id/5509/>
- [3] Algoritma Data Science School-Naive Bayes, Retrieved, (30 Maret 2022). <https://algoritma.blog/naive-bayes-2022/>
- [4] Wardana, Y., & Sari, D. P. (2023). Prediksi Penyakit Diabetes Dengan Naive Bayes. *Journal of Mathematics UNP*, 8(3), 89-97. <https://ejournal.unp.ac.id/students/index.php/mat/article/view/15070>

- [5] Meilani, N., & Nurdiawan, O. (2023). Data Mining untuk Klasifikasi Penderita Kanker Payudara Menggunakan Algoritma K-Nearest Neighbor. In *Jurnal Wahana Informatika (JWI)* (Vol. 2, Issue1).  
<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>
- [6] Rian Tineges, (23 Mei 2022). Mengenal Naive Bayes Sebagai Salah Satu Algoritma Data Science.  
<https://dqqlab.id/mengenal-naive-bayes-sebagai-salah-satu-algoritma-data-science#:~:text=Keuntungan%20algoritma%20naive%20bayes%20adalah,dataset%20yang%20jumlahnya%20lebih%20besar>.
- [7] Andi Fahlevi, (30 Sep 2021). Proses Data Mining KDD[7].  
<https://sis.binus.ac.id/2021/09/30/proses-data-mining-kdd/>
- [8] (30 Januari 2023). Apa Itu Metode Data Mining? Ini Klasifikasi dan Contohnya!  
<https://www.acerid.com/bisnis/metode-data-mining-dan-contohnya>
- [9] Chandra Henny, (11 januari 2022). Kenali Algoritma Naive Bayes Secara Lengkap[9].  
<https://id.linkedin.com/pulse/kenali-algoritma-naive-bayes-secara-lengkap-chandra-henny>