

## PREDIKSI PENYAKIT DIABETES MENGGUNAKAN MACHINE LEARNING DENGAN ALGORITMA NAÏVE BAYES

**Achmad Supandi<sup>1)\*</sup>, Ahmad Faqih<sup>2)</sup>, Fadhil Muhammad Basysyar<sup>3)</sup>**

<sup>1,2)</sup> Teknik Informatika, STMIG IKMI Cirebon, Kota Cirebon, Jawa Barat

<sup>3)</sup> Sistem Informasi, STMIG IKMI Cirebon, Kota Cirebon, Jawa Barat

email: [supandi168@gmail.com](mailto:supandi168@gmail.com)<sup>1)</sup>, [ahmadfaqih367@gmail.com](mailto:ahmadfaqih367@gmail.com)<sup>2)</sup>,  
[fadhil.m.basysyar@gmail.com](mailto:fadhil.m.basysyar@gmail.com)<sup>3)</sup>

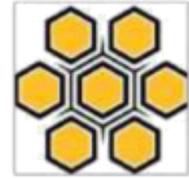
### Abstrak

Diabetes Melitus (DM) merupakan penyakit kronis yang ditandai dengan hiperglikemia dan intoleransi glukosa yang terjadi karena kelenjar pankreas tidak dapat memproduksi insulin secara adekuat atau karena tubuh tidak dapat menggunakan insulin yang diproduksi secara efektif. data di Apotik Harapan Mulya yang dipengaruhi oleh jumlah dari pasien yang melakukan pemeriksaan kesehatan seperti penyakit diabetes melitus sehingga berpengaruh dalam hal klasifikasi data yang akan menyulitkan pihak Apotik. Maka dengan memanfaatkan data mining, pengklasifikasian untuk menentukan pasien yang telah melakukan pemeriksaan termasuk penderita penyakit diabetes atau tidak. Pada penelitian ini bertujuan untuk membuat model prediksi menggunakan Algoritma Naïve Bayes yang menghasilkan klasifikasi dan prediksi diagnosa penyakit diabetes melitus yang dilakukan dengan menggunakan Rapidminer agar pencegahan terhadap penyakit diabetes dapat dilakukan segera mungkin. akurasi klasifikasi diagnosa penyakit diabetes melitus menggunakan algoritma Naïve Bayes sebesar 95,94% dengan nilai true positif mengidap penyakit Diabetes dengan nilai precision adalah 97,64% dan false negative 94,54% dari 1035 diagnosa penyakit diabetes. Metode Algoritma Naïve Bayes yang digunakan, maka model prediksi yang dihasilkan memiliki class recall true negative 93,63% dan false negative 97,99%. Nilai accuracy dari model Algoritma Naïve Bayes untuk prediksi diagnose penyakit diabetes adalah 95,94%, Data yang diprediksi positif sebanyak 487 true positif dan 548 true negatif.

**Kata Kunci :** Data mining, DM, Naïve Bayes, Klasifikasi.

### Abstract

*Diabetes Mellitus (DM) is a chronic disease characterized by hyperglycemia and glucose intolerance that occurs because the pancreatic glands are unable to produce insulin adequately or because the body cannot use insulin produced effectively. Data at Harapan Mulya Pharmacy is influenced by the number of patients who perform health checks such as diabetes mellitus so that it has an effect in terms of data classification that will make it difficult for the dispensary. By utilizing data mining. classification to determine patients who have done the examination including diabetics or not. This study aims to create a predictive model using the Naïve Bayes Algorithm which produces classification and prediction of diagnosis of diabetes mellitus done using Rapidminer so that prevention against diabetes can be done as soon as possible. accuracy of diagnostic classification. Diabetes mellitus using Naïve Bayes algorithm of 95,94% with a true positive score with diabetes with a precision value is 97,64% and false negative 94,54% of 1035 diagnoses of diabetes. Naïve Bayes Algorithm*



*method used, then the resulting prediction model has a true negative recall class of 93,63% and a false negative of 97,99%. Nilai accuracy dari model Algoritma Naïve Bayes for prediksi diagnose penyakit diabetes adalah 95,94%, Data yang diprediksi positif sebanyak 487 true positif dan 548 true negative.*

**Keywords:** *Data Mining, Learning Type, clustering, FCM, Naïve Bayes, Student learning outcomes.*

## PENDAHULUAN

Kesehatan manusia bisa sangat bergantung dari banyak faktor, antara lain karena faktor lingkungan tempat tinggal ataupun karena faktor keturunan/genetik. Kesehatan inilah yang menjadi kekuatan utama manusia untuk melaksanakan aktivitas hidupnya. Salah satu penyakit yang mengakibatkan tingginya angka kematian adalah Diabetes Mellitus (DM). Pemeriksaan kesehatan yang berkaitan dengan DM dalam dunia medis dapat dilakukan dengan cara diagnosa penyakit yang menghasilkan data hasil uji laboratorium dan rekam medis gejala sakit. Guna menekan angka kematian dari penyakit DM ini, para pakar kesehatan harus melakukan diagnosa penyakit sedini mungkin. Namun dalam mendiagnosa penyakit tersebut hanya mengacu pada rekam medis. Diabetes melitus merupakan penyakit kronis yang terjadi akibat kadar glukosa didalam darah yang terlalu tinggi sehingga tidak adanya insulin. Penyakit diabetes merupakan masalah kesehatan di seluruh dunia dengan perkiraan dari 120 juta penderita. Angka tersebut akan semakin meningkat jika adanya ketidaktahuan masyarakat umum tentang faktor-faktor yang dapat memicu terkena penyakit diabetes.[1]

Penyakit diabetes dapat mengakibatkan komplikasi penyakit, yang tentunya sangat berbahaya terhadap penderita diabetes. Oleh karena itu, sangat diperlukanya suatu teknologi yang dapat mendeteksi penyakit

diabetes dengan tingkat analisis yang akurat, sehingga penyakit diabetes dapat ditangani lebih awal untuk mengurangi jumlah penderita, kecacatan, dan kematian. Beberapa tahun terakhir, penelitian terhadap penyakit diabetes sudah dilakukan dengan menggunakan berbagai macam metode klasifikasi untuk mendeteksi diabetes. Berikut beberapa penelitian yang terkait dengan pengujian data diabetes. [2] untuk itu peneliti menganggap pentingnya dilakukan penelitian terhadap pasien penderita diabetes yang ada di apotek harapan mulya dengan pendekatan algoritma Naïve Bayes.

Diabetes adalah penyebab kematian keempat di sebagian besar negara maju. Gejala awal penyakit diabetes yaitu sering buang air kecil, meningkatnya rasa haus, dan lapar. Berdasarkan etiologi, diabetes diklasifikasikan menjadi 3 jenis: diabetes tipe 1, diabetes tipe 2, dan diabetes gestasional. Diabetes tipe 1 terjadi karena adanya kerusakan sel autoimun pada pankreas sehingga menyebabkan hilangnya produksi insulin. Sedangkan, diabetes tipe 2 disebabkan karena adanya resistensi insulin yaitu hilangnya kemampuan tubuh seseorang dalam merespon hormon insulin. Sehingga, hormon insulin tidak dapat membantu penyerapan glukosa dan menyebabkan penumpukan gula dalam darah. Penyakit diabetes dapat menimbulkan komplikasi seperti stroke, gagal ginjal, penyakit arteri coroner, amputasi, dan kebutaan yang bisa



mengakibatkan kecacatan. Karena efek diabetes dapat menyebabkan komplikasi bahkan kematian, dibutuhkan model prediksi untuk mengklasifikasikan seseorang mengidap penyakit diabetes untuk mengetahui seseorang mengidap penyakit diabetes atau tidak secara cepat.[3]

Diskritisasi atribut juga berfungsi untuk meningkatkan akurasi hasil klasifikasi yang akan dilakukan. Diskritisasi atribut dimulai dari atribut Pregnancies yang dibagi menjadi 3 kelas, yaitu Rendah (0, 1), Sedang (2-5), dan Tinggi (>6). Atribut Glucose dibagi menjadi 3 kelas, yaitu Rendah (140). Atribut BloodPressure dibagi menjadi 3 kelas, yaitu Normal (90). Atribut BMI dibagi menjadi 4 kelas, yaitu Rendah (35). Atribut DPF dibagi menjadi 2 kelas, yaitu Rendah (0,5275). Atribut Age dibagi menjadi 3 kelas, yaitu Muda (60). Sedangkan atribut Class dibagi menjadi 2, yaitu Positif (1), dan Negatif (0). Penjabaran berupa tabel dari diskritisasi atribut dapat dilihat pada tabel 3 dibawah ini.[4]

Tabel 1. Diskritisasi atribut

Atribut	Diskritisasi	Keterangan
Pregnancies	(0,1) (2-5) (>6)	Rendah Sedang Tinggi
Glucose	(<95) (95-140) (>140)	Rendah Sedang Tinggi
BloodPressure	(<80) (80-90) (>90)	Normal Normal ke Tinggi Tinggi
BMI	(<24.9) (25-29.9) (30-34.9) (>35)	Rendah Normal Obesitas

Atribut	Diskritisasi	Keterangan
		Obesitas akut
DPF	(< 0.5275) (> 0.5275)	Rendah Tinggi
Age	(<40) (40-59) (>60)	Muda Dewasa Tua
Kelas	1 0	Positif Negatif

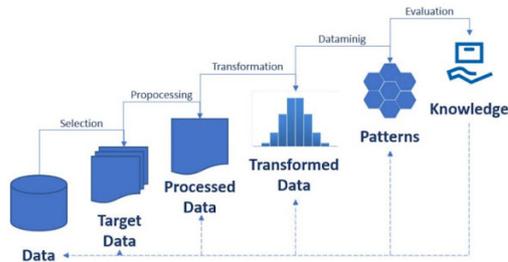
### METODE PENELITIAN

Pada penelitian ini peneliti menggunakan pendekatan deskriptif kuantitatif. Dalam pengertiannya penelitian deskriptif adalah penelitian yang menggambarkan suatu keadaan atau kondisi yang terjadi disuatu daerah, yang pemecahan masalahnya secara tersistem berdasarkan data – data yang bersikap fakta yang ada. dan data kualitatif adalah jenis data berupa angka yang dapat digunakan dalam proses operasi matematika, data ini berasal dari data fakta yang ada. Jenis data yang digunakan dalam penelitian ini bersifat rasio yang merupakan data yang berbentuk angka sebenarnya.[5] Pada Penelitian ini penulis menggunakan metode penelitian kuantitatif. Data Kuantitatif adalah data yang diperoleh dari diagnose penyakit diabetes yang ada di apotik harapan mulya kemudian data tersebut akan di analisis menggunakan pendekatan metode algoritma naives bayes.

Tahapan penelitian merupakan suatu proses memperoleh atau mendapatkan suatu pengetahuan atau memecahkan permasalahan yang dihadapi, yang dilakukan secara ilmiah, sistematis dan



logis. Tahapan penelitian pada penelitian adalah sebagai berikut :



Gambar 1. Tahapan penelitian

## HASIL DAN PEMBAHASAN

### Data Selection

Pada tahapan ini data yang digunakan akan diseleksi dengan cara melihat , kesesuaian data dengan topik atau judul penelitian yang akan di teliti, dalam hal ini data yang diperoleh dari Apotek Harapan Mulya sudah sesuai dengan format data yang terdiri dari atribut Pregnancies(NP), Glucose(GTT), BloodPressure(DBP), SkinThickness(TSF), Insulin(HIS), BMI, DiabetesPedigreeFunction(DPF), Age, Outcome, kelas.

### Preprocessing Data

Tahapan preprocessing data dalam penelitian ini menggunakan operator normalize pada tool rapidminer. operator Normalize akan melakukan Transformasi z (juga dikenal sebagai Standardisasi) yang menghasilkan nilai rata-rata 0 dan deviasi standar 1 untuk setiap atribut. Dengan kata lain, semua atribut berada pada skala yang sama setelah normalisasi dan dapat dibandingkan satu sama lain.

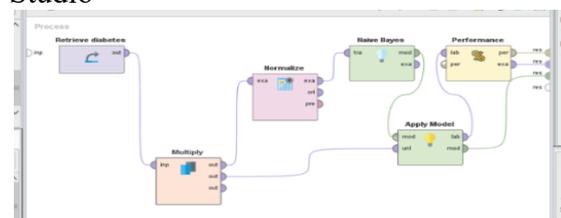
### Transpormasi data

Ada pun pada tahap ini data akan diubah menjadi bentuk yang sesuai untuk proses data mining. Karena dalam penelitian ini akan dilakukan uji coba secara teoritis dan menggunakan software data

mining yaitu RapidMiner, maka data yang telah melalui proses sebelumnya akan di transformasi agar dapat sesuai dengan algoritma yang dipakai yaitu algoritma Naïve Bayes. Pada tahapan ini atribut yang di pakai akan diberi label mengikuti kondisi data - data pada atribut tersebut. data set terdiri dari 1035 record yang sudah melalui proses transformasi data.

### Penerapan Algoritma Naïve Bayes

Pengujian dengan aplikasi yang telah ditentukan. Hasil dari analisis diuji lagi dengan menggunakan software RapidMiner Studio



Gambar 2. Model Proses Algoritma Naïve Bayes

### Evaluasi dan Interpretasi

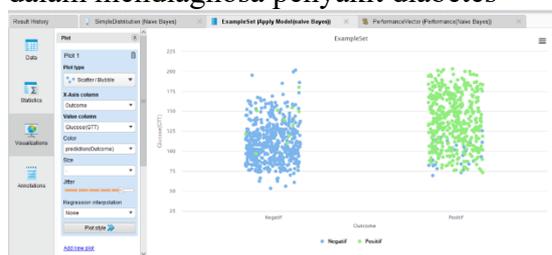
	true Negatif	true Positif	class precision
accuracy 93,6%			
pred Negatif	537	31	94,54%
pred Positif	11	468	97,99%
class recall	97,99%	93,63%	

Gambar 3. PerformanceVector

Dari gambar 3 menunjukkan bahwa nilai true positif mengidap penyakit Diabetes dengan nilai precesion adalah 97,64% dan false negative 94,54% dari 1035 diagnosa penyakit diabetes. Metode Algoritma Naïve Bayes yang digunakan, maka model prediksi yang dihasilkan memiliki class recall true negative 97,99% dan false negative 93,63%. Nilai accuracy dari model



Algoritma Naïve Bayes untuk prediksi diagnose penyakit diabetes adalah 95.94%, Data yang diprediksi positif sebanyak 456 true positif dan 11 true negatif. Sedangkan data yang diprediksi negatif sebanyak 537 true positif dan 31 true negatif. Hasil akurasi yang diperoleh dengan menggunakan algoritma Naïve Bayes dapat digunakan untuk membuat GUI atau aplikasi yang dapat digunakan untuk membantu dokter dan pasien diabetes dalam mendiagnosa penyakit diabetes



Gambar 4. Visualizations Bar

Dilihat dari hasil gambar 4 bahwa diagnose penyakit diabetes yang diprediksi positif berwarna biru muda ternyata ada juga yang negatif berwarna hijau, begitu juga sebaliknya data diagnose penyakit diabetes diprediksi negatif warna hijau ternyata ada juga yang positif warna biru muda.

### Tahapan Evaluasi dan Interpretasi

## PerformanceVector

```
PerformanceVector:
accuracy: 95.94%
ConfusionMatrix:
True:  Negatif Positif
Negatif:    537    31
Positif:    11    456
```

Gambar 5. Deskripsi PerformanceVector

### Evaluasi Performance

Evaluasi kinerja model klasifikasi yang dihasilkan, dilihat berdasarkan pada hasil pengujian objek yang diprediksi dengan benar dan salah. Model klasifikasi akan ditentukan nilai akurasi. Akurasi dalam klasifikasi menentukan persentase ketepatan record data yang di klasifikasikan secara benar. Hasil hitungan, ditabulasi kedalam bentuk confusion matrix yang memiliki jumlah nilai true positif (TP), false negative (FN) dan true negative (TN).

Perhitungan akurasi menggunakan rumus :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

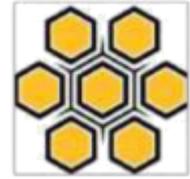
$$Accuracy = \frac{TP + TN}{TP + TN} * 100\%$$

$$Accuracy = \frac{993}{1035} * 100\%$$

$$Accuracy = \frac{993}{1035} * 100\%$$

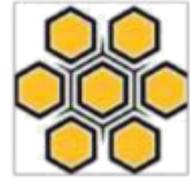
### KESIMPULAN

Hasil klasifikasi diagnosa penyakit diabetes melitus menggunakan metode algoritma Naïve Bayes adalah Data yang diprediksi positif sebanyak 456 true positif dan 11 true negatif. Sedangkan data yang diprediksi negatif sebanyak 537 true positif dan 31 true negatif. Klasifikasi terhadap diagnosa penyakit diabetes melitus di Apotik Harapan Mulya. Akurasi yang diperoleh dengan menggunakan algoritma Naïve Bayes dapat digunakan untuk membuat GUI atau aplikasi yang dapat digunakan untuk membantu dokter dan pasien diabetes dalam mendiagnosa penyakit diabetes. Tingkat akurasi klasifikasi diagnosa penyakit diabetes melitus menggunakan algoritma Naïve Bayes sebesar 95,94%.



## DAFTAR PUSTAKA

- [1] [1] D. A. K. Irfan Nurdiyanto, Odi Nurdiawan, Nining Rahaningsih, Ade Irfma Purnamasari, "Penentuan Keputusan Pemberian Pinjaman Kredit Menggunakan Algoritma C.45," *J. Data Sci. dan Inform.*, vol. 1, no. 1, pp. 16–20, 2021.
- [2] A. S. kaslani, Ade Irma Purnamasari, "Pengembangan Media Pembelajaran Interaktif Berbasis Android Pada Materi Hidrokarbon," *J. ICT Infirm. Comun. Technol.*, vol. 5, no. 1, p. 37, 2021, doi: 10.23887/jjpk.v5i1.33520.
- [3] I. A. Putri Saadah, Odi Nurdiawan, Dian Ade Kurnia, Dita Rizki Amalia, "Klasifikasi Penerima Beasiswa Dengan Menggunakan Algoritma," *J. DATA Sci. Inform. ( JDSI )*, vol. 1, no. 1, pp. 11–15, 2021.
- [4] I. A. Erliyana, Odi Nurdiawan, Nining R, Ade Irma Purnamasari, "Klasifikasi Penerima Beasiswa Dengan Menggunakan Algoritma," *J. DATA Sci. Inform. ( JDSI )*, vol. 1, no. 1, pp. 11–15, 2021.
- [5] D. Anggarwati, O. Nurdiawan, I. Ali, and D. A. Kurnia, "Penerapan Algoritma K-Means Dalam Prediksi Penjualan," *J. DATA Sci. Inform. ( JDSI )*, vol. 1, no. 2, pp. 58–62, 2021.
- [6] T. Hadi, N. Suarna, A. I. Purnamasari, O. Nurdiawan, and S. Anwar, "Game Edukasi Mengenal Mata Uang Indonesia ' Rupiah ' Untuk Pengetahuan Dasar Anak-Anak Berbasis Android," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 3, pp. 89–98, 2021, doi: 10.30865/jurikom.v8i3.3609.
- [7] O. Nurdiawan, R. Herdiana, and S. Anwar, "Komparasi Algoritma Naïve Bayes dan Algoritma K-Nearst Neighbor terhadap Evaluasi Pembelajaran Daring," *Smatika J.*, vol. 11, no. 02, pp. 126–135, 2021, doi: 10.32664/smatika.v11i02.621.
- [8] A. rinaldi D. Subandi, Husein Odi Nuriawan, "Augmented Reality dalam Mendeteksi Produk Rotan menggunakan Metode Multimedia Development Life Cycle ( MDLC )," *Means (Media Inf. Anal. dan Sist.*, vol. 6, no. 2, pp. 135–141, 2021.
- [9] H. S. Mr Agis, O. Nurdiawan, G. Dwilestari, and N. Suarna, "Sistem Informasi Penjualan Motor Bekas Berbasis Android Untuk Meningkatkan Penjualan di Mokascirebon.com," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 205–212, 2021, doi: 10.30865/jurikom.v8i6.3629.
- [10] D. Teguh, A. Ade, B. Riyan, T. Hartati, D. R. Amalia, and O. Nurdiawan, "Smart School Sebagai Sarana Informasi Sekolah di SDIT Ibnu Khaldun Cirebon," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 284–293, 2021, doi: 10.30865/jurikom.v8i6.3681.
- [11] I. Kepuasan, P. Informa, A. Febriyani, G. K. Prayoga, and O. Nurdiawan, "Index Kepuasan Pelanggan Informa dengan Menggunakan Algoritma C.45," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 330–335, 2021, doi: 10.30865/jurikom.v8i6.3686.
- [12] K. S. H. K. Al Atros, A. R. Padri, O. Nurdiawan, A. Faqih, and S. Anwar, "Model Klasifikasi Analisis Kepuasan Pengguna Perpustakaan Online Menggunakan K-Means dan Decision Tree," *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp.



- 
- 323–329, 2022, doi:  
10.30865/jurikom.v8i6.3680.
- [13] F. Febriansyah, R. Nining, A. I. Purnamasari, O. Nurdiawan, and S. Anwar, “Pengenalan Teknologi Android Game Edukasi Belajar Aksara Sunda untuk Meningkatkan Pengetahuan,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 336–344, 2021, doi: 10.30865/jurikom.v8i6.3676.
- [14] E. S. Nugraha, A. R. Padri, O. Nurdiawan, A. Faqih, and S. Anwar, “Implementasi Aplikasi Pengaduan Masyarakat Berbasis Android Pada Gedung DPRD,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, pp. 360–366, 2021, doi: 10.30865/jurikom.v8i6.3679.
- [15] R. Nurcholis, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, “Game Edukasi Pengenalan Huruf Hiragana Untuk Meningkatkan Kemampuan Berbahasa Jepang,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 338–345, 2021, doi: 10.47065/bits.v3i3.1091.
- [16] H. Putri, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, “Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN,” *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.