

IMPLEMENTASI DATA MINING KLASIFIKASI PENYAKIT DIABETES PADA PEREMPUAN MENGGUNAKAN NAÏVE BAYES

Amellia Veronica Agustin, Apriade Voutama

Program Studi Sistem Informasi S1, Fakultas Ilmu Komputer

Universitas Singaperbangsa Karawang, JL. HS. Ronggo Waluyo, Karawang, Jawa Barat, Indonesia.

2010631250003@student.unsika.ac.id

ABSTRAK

Diabetes atau kencing manis adalah penyakit yang disebabkan karena peningkatan kadar gula di dalam darah, ketika seseorang terkena diabetes maka insulin dalam tubuh tak bisa dihasilkan. Ada 3 tipe diabetes yang sering ditemukan yaitu tipe 1, tipe 2, dan diabetes gestasional yang hanya menyerang perempuan ketika hamil karena perubahan hormon. Penelitian ini bertujuan agar menghasilkan nilai akurasi dengan menggunakan algoritma klasifikasi yang dipakai untuk mengklasifikasikan penyakit diabetes pada perempuan agar membantu mendeteksi diabetes lebih dini. Data yang dipakai pada penelitian ini berasal dari Kaggle yang berjudul diabetes dataset dan terdiri dari 9 atribut. Proses klasifikasi pada kali ini menggunakan algoritma Naive Bayes dan tools yang dipakai RapidMiner untuk menguji nilai akurasi, class precision, dan class recall dari dataset yang digunakan. Sehingga hasil akurasi yang didapat pada penelitian ini yaitu sebesar 78.50%, nilai precision 85.24%, nilai recall sebesar 83.64%, serta nilai AUC sebesar 0.855. Jadi algoritma Naive Bayes yang digunakan untuk pengklasifikasian penyakit diabetes pada perempuan bisa dipakai karena proses dalam seleksinya cepat serta metodenya mudah untuk dipahami dengan nilai akurasi yang cukup baik.

Kata kunci: Diabetes, Naïve Bayes, Klasifikasi, RapidMiner, Kaggle.

1. PENDAHULUAN

Penyakit Tidak Menular (PTM) atau biasa dikenal penyakit *degenerative*. PTM merupakan salah satu penyebab masalah kesehatan bagi masyarakat, PTM tidak bisa menular dari yang menderitanya kepada oranglain, perkembangan pada penyakit ini terjadi dengan perlahan serta dengan jangka waktu yang lama [1]. Menurut [2] Penyebab kematian di dunia sebesar 74% disebabkan karena Penyakit tidak menular (PTM) contohnya kanker, pernapasan kronis, diabetes, dan kardiovaskular. Diabetes termasuk ke dalam penyakit tidak menular (PTM) yang menyebabkan kadar glukosa dalam darah menjadi meningkat karena tubuh yang tak dapat memproduksi insulin. [3] mengungkapkan bahwa peningkatan kadar glukosa dalam darah memberikan dampak dalam waktu jangka panjang yaitu merusak organ pada tubuh hingga dapat menyebabkan kegagalan fungsi jaringan dan organ. Penyakit diabetes terbagi jadi diabetes tipe 1, tipe 2, dan diabetes gestasional, beberapa peneliti mengungkapkan diabetes gestasional hanya terjadi pada perempuan yaitu pada saat kehamilan yang disebabkan oleh hormon yang berubah.

Penyakit diabetes telah menjadi masalah kesehatan yang signifikan di seluruh dunia. Ini adalah penyakit kronis yang mempengaruhi jutaan orang, termasuk perempuan. Penanganan penyakit diabetes menjadi penting karena komplikasi yang dapat terjadi jika tak ditanggulangi dengan benar. Oleh karena itu, pengembangan metode yang efektif dalam mendiagnosis penyakit diabetes pada perempuan menjadi sangat penting.

[4] mengungkapkan bahwa pada tahun 2021 jumlah penderita diabetes mencapai angka 537 juta jiwa atau sekitar 10,5% dari jumlah populasi.

Diperkirakan pada tahun 2030 sebanyak 643 juta jiwa akan mengidap diabetes atau sekitar 11,3% dari jumlah seluruh populasi, jika dibandingkan pada tahun 2000 prevalensi pengidap diabetes pada usia dewasa dengan rentang umur 20-79 tahun yang berjumlah 151 juta jiwa hal ini mendekati estimasi oleh WHO (*World Health Organization*) pada saat itu (150 juta jiwa) Maka dari itu, perkiraan oleh IDF membuktikan peningkatan yang mengkhawatirkan mengenai jumlah penderita diabetes jika tidak ada tindakan yang memadai untuk mengatasi situasi tersebut.

pada saat ini dengan perkembangan teknologi yang memberikan kemudahan kepada setiap orang untuk bertukar informasi dan data dengan jangkauannya yang tak terbatas, transformasi digital memberikan suatu perubahan cara untuk menangani pekerjaan dengan memakai teknologi informasi agar efisien dan efektif [5]. Salah satu contohnya yaitu penerapan teknologi dalam bidang kesehatan karena membutuhkan peralatan yang mampu memberikan diagnosa suatu penyakit dengan beberapa pertimbangan sehingga teknik data mining bisa digunakan untuk memberikan prediksi ataupun mengklasifikasikan suatu penyakit berdasarkan himpunan data yang terdapat pada rumah sakit maupun layanan kesehatan lain [6].

Penerapan *Naïve Bayes* telah dilakukan oleh beberapa penelitian terdahulu. [7] melakukan penelitian mengenai penerapan algoritma *Naïve Bayes* Untuk Klasifikasi Tindakan Jenis Abortus dengan menggunakan 10% *data training* sebanyak 14 data dan 90% *data testing* sebanyak 121 data, menghasilkan 99,26% nilai kebenaran. [8] melakukan penelitian untuk membandingkan Algoritma C4.5 dan *Naïve Bayes* untuk Deteksi Dini Gangguan Autisme pada

Anak yang hasilnya algoritma *Naïve Bayes* memiliki nilai akurasi yang lebih besar yaitu sebanyak 73.33%. [9] melakukan penelitian dengan menerapkan metode *Naïve Bayes* untuk menentukan penerima PIP.

Berdasarkan penjelasan tersebut, penelitian ini bertujuan untuk memperluas pemahaman kita tentang penggunaan data mining dalam klasifikasi penyakit diabetes pada perempuan. Dengan menggabungkan teknologi komputer dan informasi medis, diharapkan penelitian ini dapat memberikan manfaat yang signifikan dalam upaya pencegahan, diagnosis, dan penanganan penyakit diabetes. Penelitian ini menggunakan dataset yang didapatkan dari kaggle yang berjudul diabetes datasets sebanyak 9 atribut.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining merupakan teknik yang berfungsi untuk mengolah suatu data yang berukuran atau berjumlah besar yang didasarkan pada pola hingga bisa menghasilkan pengetahuan yang bermanfaat atau informatif serta menarik [10]. Data yang bisa diolah agar menghasilkan suatu informasi dapat diperoleh dari suatu *data warehouse*, *web*, *database* dan sumber lainnya. Dalam berbagai sektor penggunaan *data mining* dapat diterapkan karena tujuannya untuk meningkatkan penjualan dalam beberapa sektor, serta dapat dimanfaatkan untuk meningkatkan pengetahuan dan lain sebagainya.

2.2. Klasifikasi

Klasifikasi menjadi salah satu bagian data mining yang berfungsi untuk menggali informasi dari data yang diperoleh. Klasifikasi merupakan suatu proses yang berguna untuk mencari model dengan cara menganalisis kumpulan data pelatihan yang memberi gambaran serta membedakan kelas label ataupun konsep suatu data [11]. Dalam metode klasifikasi ada beberapa algoritma yang bisa digunakan, salah satunya yaitu algoritma *Naïve Bayes*. *Naïve Bayes* merupakan salah satu algoritma dari metode klasifikasi yang memanfaatkan statistik serta probabilitas dalam penggunaannya. *Naïve Bayes* berguna untuk memprediksi peluang di waktu yang akan datang dengan berdasarkan pada pengalaman di waktu lampau [12].

2.3. Naïve Bayes

Model *Naive Bayes* menyajikan cara untuk menyatukan peluang dahulu berdasarkan syarat atas kemungkinan menjadi suatu formula yang bisa dipakai untuk menghitung suatu peluang dari setiap kemungkinan yang akan terjadi [13]

Bentuk umum :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Keterangan:

A = Data beserta kelas yang belum diketahui

B = Hipotesis data A merupakan suatu kelas spesifik

$P(A|B)$ = Probabilitas hipotesis A berdasarkan kondisi B (posteriori probability)

$P(A)$ = Probabilitas hipotesis A (prior probability)

$P(B|A)$ = Probabilitas B berdasarkan kondisi pada hipotesis A

$P(B)$ = Probabilitas dari B

2.4. Rapidminer

Rapidminer memiliki fungsi untuk penelitian pendidikan ataupun berbagai aplikasi karena bisa diterapkan untuk menyelesaikan permasalahan mengenai data mining di kehidupan sehari-hari khususnya klasifikasi. RapidMiner mudah dipakai dan diterapkan di tingkat yang berbeda. Rapidminer menyediakan tools yang berguna untuk Visualisasi, preprocessing data, klastering, klasifikasi, Regresi, dan Asosiasi. RapidMiner bisa digunakan untuk processing data, menginputnya ke dalam skema pembelajaran serta menganalisa classifier yang dihasilkan dengan performansinya. Hal tersebut bisa dilakukan tanpa harus menulis kode program[3]

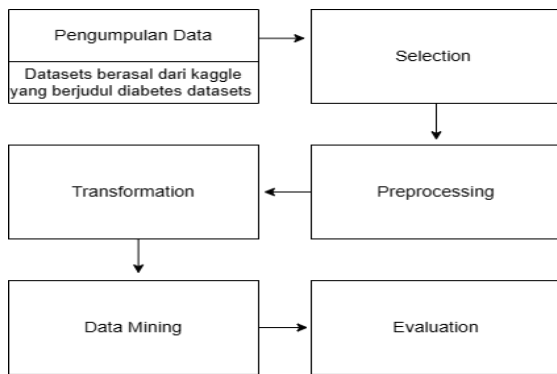
3. METODE PENELITIAN

3.1. Pengumpulan Data

Untuk mendukung penelitian pada kali ini dilakukan proses mengumpulkan data. Pada penelitian ini dalam pengumpulan data dilakukan untuk mendapatkan sumber dataset yang akan digunakan dengan pengumpulan data sekunder yang berasal dari web penyedia datasets yaitu *Kaggle* dengan judul diabetes datasets. Data sekunder merupakan data yang bisa didapatkan oleh peneliti tanpa harus terjun langsung ke lapangan untuk pengambilan datanya karena datanya sudah diolah dan dikumpulkan oleh peneliti lain sebelumnya. Jumlah data sampel yang dipakai pada penelitian ini ialah sebanyak 767 data dan terdiri dari 9 atribut.

3.2. Memproses Data

Metodologi yang digunakan pada penelitian ini yaitu *Knowledge Discovery on Database (KDD)*. KDD ialah suatu proses komputasi yang memakai perhitungan matematika berfungsi untuk mengekstraksi suatu informasi serta membuat kalkulasi probabilistic yang berasal dari probabilitas tindakan di waktu mendatang [14]. Metode *Knowledge Discovery on Database* berisi tahapan *data selection*, *data preprocessing*, *data transformation*, *data mining*, dan *Evaluation* [15]



Gambar 1. Proses KDD

1) Selection

Pada tahap seleksi akan dilakukan seleksi atribut lalu hasilnya akan diintegrasikan sehingga menghasilkan suatu dataset. Proses dari pembangunan dataset ialah proses yang penting dikarenakan dari proses pembelajaran data mining serta penemuan pola yang baru berdasar dataset yang sudah dibentuk [14]

2) Preprocessing

Selanjutnya pada tahap *preprocessing* akan dilakukan penghilangan duplikasi data, memperbaiki dari kesalahan data, *handling missing value*, dan lainnya [10]

3) Transformation

Pada tahap selanjutnya yaitu tahap transformation yang merupakan proses transformasi atau bisa disebut untuk mengubah bentuk data yang mempunyai entitas yang tidak jelas agar diubah menjadi bentuk data yang valid untuk diproses ke tahap berikutnya [16]

4) Data Mining

Data mining adalah pusat dari keseluruhan proses ini. Di tahap ini merupakan penerapan dari metode ataupun algoritma berdasarkan informasi yang ingin dicari. Metode ataupun algoritma tersebut bisa klasifikasi, prediksi, klastering, estimasi, dan sebagainya. Untuk penelitian pada kali ini memakai algoritma *Naïve Bayes*.

5) Evaluation

Tahap evaluasi yang merupakan tahap untuk menerjemahkan atau mengartikan pola yang dihasilkan data mining. Pola dari informasi yang telah dihasilkan harus disajikan dengan bentuk yang gampang dimengerti. Pada tahap ini dilakukan pemeriksaan terhadap pola dan informasi yang bertentangan dengan hipotesis serta fakta yang telah ada sebelumnya [17].

4. HASIL DAN PEMBAHASAN

Penelitian ini pada proses klasifikasi dengan memakai *software RapidMiner* dan algoritma yang dipakai adalah *Naïve Bayes*. Tujuan penelitian kali ini untuk menghasilkan nilai akurasi pada algoritma klasifikasi yang dipakai untuk mengklasifikasikan penyakit diabetes. Pada tabel 1 menunjukkan deskripsi atribut dari dataset yang dipakai untuk penelitian ini.

Tabel 1. Deskripsi atribut datasets

No	Atribut	Keterangan
1	<i>Pregnancies</i>	Menyatakan Jumlah kehamilan
2	<i>Glucose</i>	Kadar Glukosa dalam darah
3	<i>Blood Pressure</i>	Pengukuran tekanan darah
4	<i>Skin Thickness</i>	Menyatakan ketebalan kulit
5	<i>Insulin</i>	Memperlihatkan tingkat Insulin dalam darah
6	<i>BMI</i>	Menyatakan indeks massa tubuh
7	<i>Diabetes Pedigree Function</i>	Persentase diabetes
8	<i>Age</i>	Menyatakan umur
9	<i>Outcome</i>	Menyatakan hasil akhir 1 adalah Ya dan 0 adalah Tidak

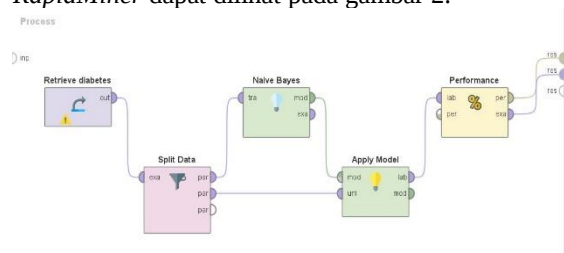
4.1 Selection

Pada tahap *selection* data, dilakukan pemilihan pada atribut (kolom) yang akan digunakan untuk penelitian ini. Atribut yang terpilih untuk digunakan pada penelitian ini ada 7 atribut yaitu *pregnancies*, *glucose*, *blood pressure*, *insulin*, *BMI*, *diabetes pedigree function*, *age*, dan 1 atribut yang digunakan sebagai label yaitu atribut *Outcome*. Atribut yang digunakan ditunjukkan pada tabel 2.

Tabel 2. Atribut yang dipilih

No	Atribut	Keterangan
1	<i>Pregnancies</i>	Menyatakan Jumlah kehamilan
2	<i>Glucose</i>	Kadar Glukosa dalam darah
3	<i>Blood Pressure</i>	Pengukuran tekanan darah
4	<i>Insulin</i>	Memperlihatkan tingkat Insulin dalam darah
5	<i>BMI</i>	Menyatakan indeks massa tubuh
6	<i>Diabetes Pedigree Function</i>	Persentase diabetes
7	<i>Age</i>	Menyatakan umur
8	<i>Outcome</i>	Menyatakan hasil akhir 1 adalah Ya dan 0 adalah Tidak

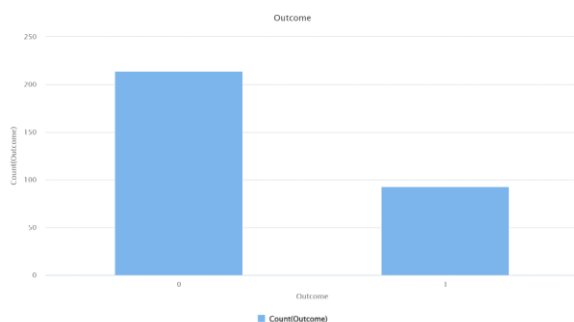
Atribut yang terpilih tersebut sesuai dengan gejala penderita diabetes pada perempuan, atribut yang terpilih diolah menggunakan algoritma *Naïve Bayes* dengan *tools RapidMiner* yang akan menghasilkan nilai akurasi, recall, precision, confusion matrix, kurva ROC dari *RapidMiner*. Proses pengolahan data dengan menggunakan algoritma *Naïve Bayes* dan *tools RapidMiner* dapat dilihat pada gambar 2.



Gambar 2. Pengolahan data dengan naïve bayes

Dataset yang telah dipilih atributnya masuk ke tahap *preprocessing* untuk memastikan bahwa tidak ada *missing value* agar mempermudah pengolahan dataset. Pada penelitian ini atribut dataset diabetes yang digunakan tidak ditemukan *missing value* dan sudah lengkap. Setelah memastikan data yang digunakan sudah lengkap tidak terdapat *missing value* maka selanjutnya masuk ke tahap mentransformasikan data dengan mengubahnya sesuai dengan format yang dibutuhkan, perubahan pada data dilakukan dengan menggunakan *RapidMiner*, pada tipe data atribut *Outcome* diubah menjadi binomial agar data dapat digunakan.

Setelah transformasi data maka selanjutnya melakukan Split Data menggunakan *RapidMiner*, dengan membagi data menjadi data latih dan data uji, dengan presentase data latih sebesar 60% dan data uji sebesar 40% lalu tahap selanjutnya yaitu Data Mining yang akan melakukan klasifikasi dengan algoritma Naive Bayes untuk menghasilkan nilai akurasi yang dipakai untuk mengklasifikasikan penyakit diabetes pada perempuan agar membantu mendeteksi diabetes lebih dini. Berikut ini merupakan contoh visualisasi data dalam bentuk bar chart berdasarkan atribut *Outcome* ditunjukkan pada gambar 3.



Gambar 3. Visualisasi data dengan chart bar

Setelah melakukan *split data*, maka data latih dan data uji akan diproses klasifikasi menggunakan *RapidMiner*. Perhitungan evaluasi yang dilakukan menggunakan *confussion matrix* untuk menghasilkan performa klasifikasi. Nilai *confussion matrix* yang dihasilkan dengan menggunakan algoritma *naive bayes* ditunjukkan oleh tabel 3.

Tabel 3. Nilai confusion matrix

	True 1	True 0	Class precision
Pred. 1	62	35	63.92%
Pred. 0	31	179	85.24%
Class recall	66.67%	83.64%	

Keterangan:

1 = Positive

0 = Negative

Umumnya confusion matrix mempunyai 3 metode untuk pengujian, sebagai Berikut:

1) Accuracy

Merupakan metode untuk menguji suatu algoritma menurut tingkat kedekatan antara nilai prediksi dengan nilai sebenarnya. Rumus untuk menghitung nilai akurasi adalah :

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$= \frac{62+179}{62+179+35+31} = \frac{241}{307} = 0.7850 = 78.50\%$$

2) Precision

Metode tes algoritma yang membandingkan data benar yang diperoleh sistem dengan seluruh jumlah data yang diperoleh sistem yang benar atau salah. Rumus untuk menghitung nilai precision adalah:

$$\text{Precision} = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

$$= \frac{62}{62+31} = \frac{62}{93} = 0.6667 = 66.67\%$$

3) Recall

Merupakan salah satu proses pengujian dari suatu Algoritma yang bisa digunakan untuk membandingkan data yang tepat yang didapat sistem dengan seluruh data yang tepat diambil atau tidak diambil oleh sistem. Rumus untuk menghitung nilai recall adalah :

$$\text{Recall} = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

$$= \frac{62}{62+35} = \frac{62}{97} = 0.63917 = 63.92\%$$

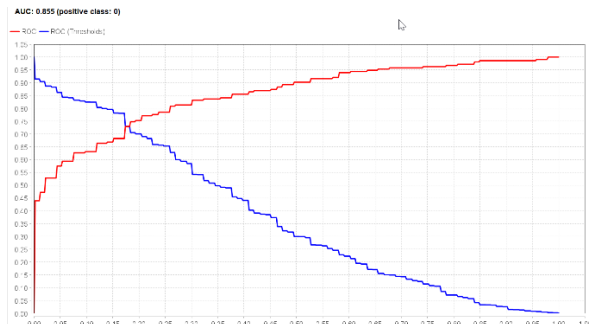
Pada table 3 diketahui bahwa nilai *class precision* sebesar 85.24% dan nilai *class recall* sebesar 83.64%. Maka dari evaluasi klasifikasi menggunakan algoritma Naïve Bayes menghasilkan nilai akurasi yang bisa dilihat pada *PerformanceVector* di gambar 4.

PerformanceVector

```
PerformanceVector:
accuracy: 78.50%
ConfusionMatrix:
True: 1 0
1: 62 35
0: 31 179
precision: 85.24% (positive class: 0)
ConfusionMatrix:
True: 1 0
1: 62 35
0: 31 179
recall: 83.64% (positive class: 0)
ConfusionMatrix:
True: 1 0
1: 62 35
0: 31 179
AUC (optimistic): 0.855 (positive class: 0)
AUC: 0.855 (positive class: 0)
AUC (pessimistic): 0.855 (positive class: 0)
```

Gambar 4 performance vector

Dapat dilihat berdasarkan gambar tersebut didapatkan nilai akurasi sebesar 78.50% untuk mengklasifikasikan dataset diabetes pada perempuan. Hasil dari pengujian pada data yang digunakan pun menghasilkan kurva ROC yang bisa dilihat pada gambar 5.



Gambar 5. Kurva ROC

Kurva ROC atau Receiver Operating Characteristic pada gambar tersebut menunjukkan hasil klasifikasi algoritma Naïve Bayes memiliki nilai AUC sebanyak 0.855 yang artinya good (baik) dari hasil penelitian, menunjukkan bahwa Naïve Bayes untuk mengklasifikasikan penyakit diabetes pada perempuan memberikan akurasi yang cukup baik dengan dataset yang digunakan.

5. KESIMPULAN DAN SARAN

Diabetes merupakan Penyakit Tidak Menular (PTM) yang menjadi penyebab kadar glukosa dalam darah meningkat karena tubuh yang tidak bisa memproduksi insulin. Peningkatan kadar glukosa dalam darah memberikan dampak waktu jangka panjang bisa merusak organ tubuh serta menyebabkan kegagalan fungsi jaringan dan organ. Pada penelitian ini terdapat 8 atribut yang digunakan untuk mengklasifikasikan dataset diabetes pada perempuan, yaitu atribut *pregnancies*, *glucose*, *blood pressure*, *insulin*, *BMI*, *diabetes pedigree function*, *age*, dan *outcome*. Algoritma yang digunakan pada penelitian ini yaitu Naïve Bayes untuk mengklasifikasikan dataset diabetes pada perempuan. Pada proses pengklasifikasian permodelan menggunakan algoritma Naïve Bayes pada dataset yang digunakan menghasilkan nilai *Class precision* sebesar = 85.24%, dan nilai *Class recall* sebesar = 83.64%. Hasil dari evaluasi pada klasifikasi ini menunjukkan hasil nilai akurasi untuk pengklasifikasian dataset sebesar 78.50% yang artinya sudah cukup baik, tetapi lebih baik nilai akurasinya ditingkatkan dengan cara atau *Assemble* yang lain. Kurva ROC (*Receiver Operating Characteristic*) yang dihasilkan menggunakan klasifikasi dari permodelan yang memakai Naïve Bayes menunjukkan nilai AUC sebanyak = 0.855 artinya model yang dihasilkan termasuk good (baik). Adapun saran untuk penelitian berikutnya dapat melakukan komparasi atau perbandingan dengan

beberapa algoritma klasifikasi yang lain, agar menghasilkan nilai akurasi yang lebih baik.

DAFTAR PUSTAKA

- [1] Kemenkes, "Penyakit Tidak Menular (PTM)," 2022. https://yankes.kemkes.go.id/view_artikel/761/penyakit-tidak-menular-ptm
- [2] World Health Organization, *Noncommunicable diseases PROGRESS MONITOR 2022*. 2022.
- [3] A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.
- [4] I. D. Federation, *IDF Diabetes Atlas 10th edition*, vol. 102, no. 2. 2021. doi: 10.1016/j.diabres.2013.10.013.
- [5] M. Danuri, "Perkembangan Dan Transformasi Teknologi Digital," *Infokam*, 2020.
- [6] L. U. Khasanah, Y. N. Nasution, and F. D. T. Amijaya, "Klasifikasi Penyakit Diabetes Melitus Menggunakan Algoritma Naïve Bayes Classifier," *BASIS (Jurnal Ilmiah Matematika)*, vol. 1, no. 1, pp. 41–50, 2022.
- [7] I. D. Lestari, T. Setiadi, and L. Zahrotun, "Penerapan Data Mining Menggunakan Metode Naïve Bayes Untuk Klasifikasi Tindakan Jenis Abortus Di RSUD Duta Mulya," *Jurnal Sarjana Teknik Informatika*, vol. 6, no. 2, pp. 60–68, 2018, [Online]. Available: <http://journal.uad.ac.id/index.php/JSTIF>
- [8] B. Sugara, D. Adidarma, and S. Budilaksono, "Perbandingan Akurasi Algoritma C4.5 dan Naïve Bayes untuk Deteksi Dini Gangguan Autisme pada Anak," *Jurnal IKRA-ITH Informatika*, vol. 3, no. 1, pp. 119–128, 2019.
- [9] A. Pebdika et al., "KLASIFIKASI MENGGUNAKAN METODE NAIVE BAYES UNTUK MENENTUKAN CALON PENERIMA PIP," vol. 7, no. 1, pp. 452–458, 2023.
- [10] B. Nurseptia, A. Voutama, N. Haeryana, and J. HSRonggo Waluyo, "Penerapan Algoritma K-Means Untuk Mengelompokkan Kabupaten/Kota Dalam Upaya Pemetaan Lapangan Pekerjaan Baru," *Jurnal Teknologi Informasi*, vol. 6, no. 2, pp. 181–186, 2022.
- [11] A. Yoga Pratama, Y. Umaidah, and A. Voutama, "Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 897–910, 2021.
- [12] M. F. Andriansyah, D. Yusup, and A. Voutama, "SISTEM PAKAR DETEKSI DINI COVID-19 MENGGUNAKAN METODE NAIVE BAYES BERBASIS WEBSITE," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 4, no. 2, pp. 446–455, 2021.

-
- [13] S. Arif and M. Fitriana, "IMPLEMENTASI DATA MINING DENGAN METODE KLASIFIKASI NAIVE BAYES UNTUK MEMPREDIKSI STOK BAHAN JADI," *SNAsTekS*, vol. 10, no. 2, 2019.
- [14] A. Fabiano Milan and V. Apriade, "PERBANDINGAN METODE DATA MINING UNTUK MEMPREDIKSI PRESTASI AKADEMIK SISWA," vol. 15, no. 1, pp. 61–66, 2023, doi: 10.30736/jt.v15i1.929.
- [15] A. Abdussalam, P. Intam, S. Betha Nurina, Garno, and V. Apriade, "Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang)," *Jurnal Informatika dan Rekayasa Elektronik*, vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.
- [16] P. Naisah Marito and S. Betha Nurina, "Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus," *EVOLUSI: Jurnal Sains dan Manajemen*, vol. 10, no. 1, 2022, doi: 10.31294/evolusi.v10i1.12514.
- [17] N. Nurdiana and A. Algifari, "Studi Komparasi Algoritma Id3 Dan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *INFOTECH journal*, vol. 6, no. 2, pp. 18–23, 2020.