

**PEMODELAN KLASIFIKASI PREDIKSI DIABETES
MENGUNAKAN PENDEKATAN ALGORITMA NAÏVE BAYES**

SKRIPSI

Nama : Imam Wahyudiarto NIM : 20230050010
Nama : Adisti Ridha Ramadhan NIM : 20210050129
Nama : Nadva Azhar Nabila NIM : 20210050060



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK KOMPUTER DAN DESAIN
UNIVERSITAS NUSA PUTRA
2024**

**PEMODELAN KLASIFIKASI PREDIKSI DIABETES
MENGUNAKAN PENDEKATAN ALGORITMA NAÏVE BAYES**

SKRIPSI

*Diajukan Untuk Memenuhi Salah Satu Syarat Dalam Menempuh
Gelar Sarjana Komputer*

Nama : Imam Wahyudiarto NIM : 20230050010
Nama : Adisti Ridha Ramadhan NIM : 20210050129
Nama : Nadva Azhar Nabila NIM : 20210050060



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK KOMPUTER DAN DESAIN
UNIVERSITAS NUSA PUTRA**

2024

PERNYATAAN PENULIS

**JUDUL : PEMODELAN KLASIFIKASI PREDIKSI DIABETES
MENGUNAKAN PENDEKATAN ALGORITMA NAÏVE
BAYES**

NAMA PENULIS 1 : IMAM WAHYUDIARTO

NIM PENULIS 1 : 20230050010

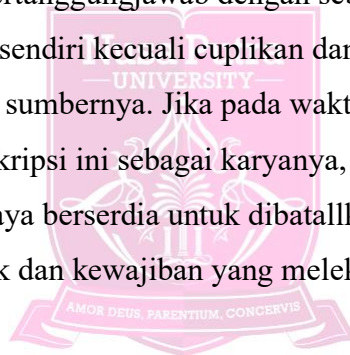
NAMA PENULIS 2 : ADISTI RIDHA RAMADHAN

NIM PENULIS 2 : 20210050129

NAMA PENULIS 3 : NADYA AZHAR NABILA

NIM PENULIS 3 : 20210050060

“Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai dengan bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut”.



Sukabumi, 18 Desember 2024

Penulis 1

Penulis 2

Penulis 3

Imam Wahyudiarto

20230050010

Adisti Ridha Ramadhan

20210050129

Nadya Azhar Nabila

20210050060

PENGESAHAN SKRIPSI

**JUDUL : PEMODELAN KLASIFIKASI PREDIKSI DIABETES
MENGUNAKAN PENDEKATAN ALGORITMA NAÏVE
BAYES**

NAMA PENULIS 1 : IMAM WAHYUDIARTO

NIM PENULIS 1 : 20230050010

NAMA PENULIS 2 : ADISTI RIDHA RAMADHAN

NIM PENULIS 2 : 20210050129

NAMA PENULIS 3 : NADYA AZHAR NABILA

NIM PENULIS 3 : 20210050060

Skripsi ini telah diujikan dan dipertahankan di depan Dewan Penguji pada Sidang Skripsi pada tanggal 27 Desember 2024. Menurut pandangan kami, skripsi ini memadai dari segi kualitas untuk tujuan penganugerahan gelar Sarjana Komputer.

Sukabumi, 27 Desember 2024

Pembimbing I

Sudin Saepudin, M.Kom
NIDN. 0414088608

Ketua Penguji

Pembimbing II

Carti Irawan, M.Kom
NIDN. 0401108606

Ketua Program Studi Sistem Informasi

Habi Baturahmah, M.Kom
NIDN.0414069701

Adhitia Erfina, S.T., M.Kom
NIDN. 0417049102

Plh. Dekan Fakultas Teknik Komputer dan Desain

Ir. Paikun, S.T., IPM., ASEAN, Eng
NIDN. 0402037401

Abstrack

Diabetes Mellitus is a global chronic disease according to WHO whose prevalence continues to increase massively, making early detection and risk prediction crucial for effective prevention and treatment. The dataset used in this study consists of medical and demographic parameters of a number of patients who have been diagnosed with diabetes as well as those who have not. The Naïve Bayes method was chosen for its ability to handle complex data with the assumption that each feature is independent of the other. Data pre-processing steps including normalization, removal of missing values, and feature selection were performed prior to the application of the classification algorithm. Model evaluation is performed using performance metrics such as accuracy, Classification Error, AUC, precision, recall, F-measure, and sensitivity. The experimental results show that the Naïve Bayes model is able to provide a fairly good prediction of diabetes status, with an average accuracy of 95.5%. These findings demonstrate the potential of using the Naïve Bayes algorithm approach in the development of diabetes prediction systems that can be used in clinical practice to support more timely and accurate decision making.

Keywords: Diabetes Mellitus, Classification, Naïve Bayes algorithm, Prediction

Abstrak

Diabetes Melitus merupakan penyakit kronis global menurut WHO yang prevalensinya terus meningkat secara masif, sehingga deteksi dini dan prediksi risiko menjadi krusial untuk pencegahan dan penanganan yang efektif. Kumpulan data yang digunakan dalam penelitian ini terdiri dari parameter medis dan demografi sejumlah pasien yang telah didiagnosis menderita diabetes serta mereka yang belum. Metode Naïve Bayes dipilih karena kemampuannya menangani data yang kompleks dengan asumsi bahwa setiap fitur bersifat independen satu sama lain. Langkah-langkah pra-pemrosesan data termasuk normalisasi, penghapusan nilai yang hilang, dan pemilihan fitur dilakukan sebelum penerapan algoritma klasifikasi. Evaluasi model dilakukan dengan menggunakan metrik kinerja seperti akurasi, *Classification Error*, AUC, presisi, *recall*, *F-measure*, dan sensitivitas. Hasil eksperimen menunjukkan bahwa model Naïve Bayes mampu memberikan prediksi status diabetes yang cukup baik, dengan akurasi rata-rata 95,5 %. Temuan ini menunjukkan potensi penggunaan pendekatan algoritma Naïve Bayes dalam pengembangan sistem prediksi diabetes yang dapat digunakan dalam praktik klinis untuk mendukung pengambilan keputusan yang lebih tepat waktu dan akurat.

Kata Kunci : Diabetes Melitus, Klasifikasi, Algoritma Naïve Bayes, Prediksi

KATA PENGANTAR

Puji syukur kami panjatkan kehadiran Allah SWT Tuhan Yang Maha Esa atas limpahan nikmat, rahmat dan karunianya kepada kita semua, shalawat dan salam semoga tetap tercurah limpahkan kepada baginda Nabi Muhammad SAW beserta keluarganya, para sahabatnya dan kita sebagai para pengikutnya. Alhamdulillah dengan ini penulis dapat menyelesaikan skripsi yang berjudul **“PEMODELAN KLASIFIKASI PREDIKSI DIABETES MENGGUNAKAN PENDEKATAN ALGORITMA NAÏVE BAYES”**

Penulisan Skripsi ini merupakan tahap akhir dan salah satu syarat untuk meraih gelar Strata 1 Program Studi Sistem Informasi.

Dengan ini sesungguhnya penulis menyadari terdapat banyak kekurangan dalam penulisan skripsi ini. Namun dalam penelitian ini penulis berharap dapat memenuhi syarat untuk meraih gelar sarjana ilmu komputer dan semoga skripsi yang dibuat ini mampu memberikan banyak manfaat bagi siapa pun yang membacanya. Tentunya proses pembuatan skripsi ini tidak dapat terselesaikan dengan baik tanpa adanya arahan, bimbingan dan dukungan dari berbagai pihak. Oleh karena itu, dengan kerendahan hati penulis mengucapkan banyak-banyak terima kasih kepada :

1. Bapak Dr. Kurniawan ST., M.Si., MM Selaku Rektor Universitas Nusa Putra.
2. Bapak Anggy Pradiftha Junfithrana, S.Pd., MT Selaku Wakil Rektor I Bidang Akademik Universitas Nusa Putra.
3. Bapak Ir. Paikun, S.T., IPM., ASEAN, Eng Selaku Dekan Fakultas Teknik Komputer dan Desain Universitas Nusa Putra.
4. Bapak Adhitia Erfina, ST., M.Kom selaku Ketua Program Studi Sistem Informasi Universitas Nusa Putra.
5. Bapak Sudin Saepudin, M.Kom selaku Dosen Pembimbing I yang telah meluangkan waktu, tenaga dan pikiran untuk memberikan bimbingan, arahan, pengetahuan, dan petunjuk yang sangat bermanfaat bagi penulis dalam penyusunan skripsi ini.

6. Ibu Carti Irawan, M.Kom selaku Dosen Pembimbing II yang telah meluangkan waktu dan memberikan arahan penulisan skripsi kepada penulis.
7. Seluruh Dosen Program Studi Sistem Informasi Universitas Nusa Putra yang tidak dapat disebutkan satu persatu, yang telah berjasa memberikan ilmu pengetahuannya.
8. Rekan-rekan seperjuangan di jurusan Sistem Informasi Angkatan 2021.
9. *Last but not last*. Terima kasih untuk **Imam Wahyudiarto, Adisti Ridha Ramadhan** dan **Nadya Azhar Nabila**, apresiasi sebesar-besarnya karena telah mampu bertanggung jawab untuk menyelesaikan apa yang telah dimulai. Terima kasih sudah berusaha dan tidak menyerah serta senantiasa menikmati proses yang bisa di bilang tidak mudah. Terima kasih sudah kuat dan bertahan sejauh ini.
10. Serta masih banyak lagi pihak-pihak yang sangat berpengaruh dalam proses pembuatan dan penyelesaian skripsi yang tidak bisa penulis sebutkan satu persatu. Dengan ini semua semoga atas izin Allah SWT semua kebaikan pihak yang telah membantu proses penyelesaian skripsi ini dibalas oleh Allah SWT dengan pahala yang berlipat ganda serta diberikan kebahagiaan, keberkahan dan kelak dapat berkumpul di jannahnya. Semoga skripsi ini dapat memberikan manfaat dan kontribusi bagi dunia pendidikan.

Sukabumi, 18 Desember 2024

Penulis 1

Penulis 2

Penulis 3

Imam Wahyudiarto

20230050010

Adisti Ridha Ramadhan

20210050129

Nadya Azhar Nabila

20210050060

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik UNIVERSITAS NUSA PUTRA, saya yang bertanda tangan di bawah ini:

Nama	: <u>Imam Wahyudiarto</u>	NIM	: 20230050010
Nama	: <u>Adisti Ridha Ramadhan</u>	NIM	: 20210050129
Nama	: <u>Nadya Azhar Nabila</u>	NIM	: 20210050060

Program Studi : Sistem Informasi
Jenis karya : Skripsi

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Nusa Putra **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty- Free Right*) atas karya ilmiah kami yang berjudul :

PEMODELAN KLASIFIKASI PREDIKSI DIABETES MENGGUNAKAN PENDEKATAN ALGORITMA NAÏVE BAYES

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Nusa Putra berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama kami sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Sukabumi
Pada tanggal : 18 Desember 2024

Yang menyatakan :

Penulis 1

Penulis 2

Penulis 3

Imam Wahyudiarto
20230050010

Adisti Ridha Ramadhan
20210050129

Nadya Azhar Nabila
20210050060

DAFTAR ISI

PERNYATAAN PENULIS	ii
PENGESAHAN SKRIPSI	iii
ABSTRACK	iv
ABSTRAK	v
KATA PENGANTAR	vi
PERNYATAAN PERSETUJUAN PUBLIKASI	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
DAFTAR ISTILAH	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	4
2.1 Penelitian Terkait	4
2.2 Landasan Teori	5
2.2.1 Data Mining	5
2.2.2 Klasifikasi	6
2.2.3 Algoritma Naïve Bayes	6
2.2.4 Rapid Miner	7
2.3 Kerangka Teori	8
2.4 Hipotesis	10
2.5 Pemodelan	10
2.6 Diabetes	12
2.7 Kaggle	13
2.8 Spesifikasi <i>Software</i> dan <i>Hardware</i>	15
2.9 Perbandingan Model	16

BAB III METODOLOGI PENELITIAN	18
3.1 Tahapan Penelitian	18
3.1.1 Identifikasi Masalah	18
3.1.2 Studi Literatur	18
3.1.3 Pengumpulan Data	19
3.2 Pengumpulan Data	19
3.2.1 Observasi	21
3.2.2 Wawancara	22
3.2.3 Studi Pustaka	22
3.3 Metode Penelitian	22
3.3.1 Metode <i>Knowledge Discovery on Database</i> (KDD)	22
3.3.2 Tahapan <i>Knowledge Discovery on Database</i> (KDD)	23
3.4 Data dan Bahan Penelitian	25
3.4.1 Data Penelitian	25
3.4.2 Bahan Penelitian	26
BAB IV HASIL DAN PEMBAHASAN	27
4.1 Hasil Penelitian	27
4.2 Pembahasan	28
4.2.1 Akurasi	28
4.2.2 <i>Performance</i>	29
4.2.3 Prediksi	31
4.2.4 Korelasi	34
4.2.5 Hasil Analisis dan Perbandingan Data	37
BAB V KESIMPULAN DAN SARAN	40
5.1 Kesimpulan	40
5.2 Saran	41
DAFTAR PUSTAKA	42
LAMPIRAN	44

DAFTAR TABEL

Tabel 1.1 Software (Perangkat Lunak)	15
Tabel 2.2 <i>Hardware</i> (Perangkat Keras)	15
Tabel 2.3 <i>Minimum Requirements</i> (Persyaratan Minimal)	15
Tabel 3.1 Atribut Data Diabetes	20
Tabel 3.2 <i>Minimum Requirements</i> (Persyaratan Minimal)	28



DAFTAR GAMBAR

Gambar 1.1 Kumpulan Data Diabetes	1
Gambar 3.1 Proses Data Mining Menggunakan Metode KDD	23
Gambar 4.1 Potongan Dataset	27
Gambar 4.2 Hasil Pengolahan Data	28
Gambar 4.3 Performance	29
Gambar 4.4 Hasil Prediksi	31
Gambar 4.5 Korelasi	34
Gambar 4.6 Hasil Korelasi	36
Gambar 4.7 Hasil Prediksi dan Perbandingan Data	38



DAFTAR LAMPIRAN

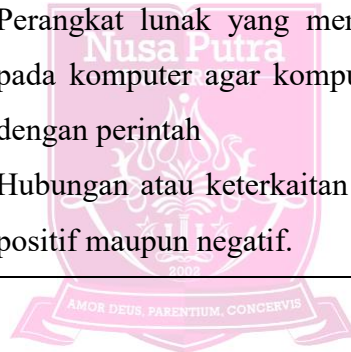
Lampiran 1. Dokumentasi Penelitian Ke kantor Kemenkes dengan Pak Menteri	44
Lampiran 2. Dokumentasi Pengambilan Data PTM Bersama Wakil Bupati Sukabumi ke Kantor Kemenkes (Dirjen P2P)	45
Lampiran 3. Wawancara Bersama pejabat atau staf di Kementerian Kesehatan (Ditjen Pencegahan dan Pengendalian Penyakit).	46



DAFTAR ISTILAH

Kata	Arti Kata
Dataset	: Kumpulan data yang terstruktur dan digunakan untuk analisis atau pelatihan model.
Algoritma	: Proses atau langkah sistematis untuk memecahkan masalah atau melakukan perhitungan.
Atribut	: Ciri yang melekat pada suatu objek
Naive Bayes	: Algoritma klasifikasi berbasis probabilitas dengan asumsi independensi antar variabel.
<i>Clustering</i>	: Proses pengelompokkan data ke dalam beberapa <i>Cluster</i> yang memiliki kemiripan maksimum.
<i>Cleaning Data</i>	: Proses mendeteksi dan memperbaiki data set, tabel dan database yang tidak relevan atau tidak akurat
<i>Data Mining</i>	: Proses menemukan pola dan informasi dari kumpulan data besar.
Kaggle	: Platform untuk berbagi dataset, kompetisi data, dan pelatihan <i>machine learning</i> .
<i>Hardware</i>	: Komponen dari sebuah komputer yang sifatnya dapat dilihat dan di sentuh secara langsung
<i>Machine Learning</i>	: Cabang kecerdasan buatan yang memungkinkan komputer belajar dari data tanpa pemrograman eksplisit.
<i>Literatur Review</i>	: Bahan bacaan yang digunakan dari berbagai sumber
Diabetes	: Penyakit kronis yang ditandai oleh kadar gula darah tinggi akibat gangguan produksi atau penggunaan insulin.
BMI	: Indeks Massa Tubuh, ukuran untuk menilai status berat badan berdasarkan tinggi dan berat badan.
<i>Heart Disease</i>	: Kondisi yang memengaruhi fungsi atau struktur jantung, seperti penyakit arteri koroner.

HbA1 Level	: Rata-rata kadar gula darah selama 2–3 bulan terakhir, digunakan untuk memantau diabetes.
<i>Blood Glucose Level</i>	: Jumlah glukosa dalam darah, indikator utama diabetes.
<i>Hypertension</i>	: Tekanan darah tinggi yang meningkatkan risiko penyakit jantung dan komplikasi lainnya.
Big Data	: Kumpulan data berukuran besar, kompleks, dan beragam yang membutuhkan teknologi khusus untuk pengelolaannya.
Akurasi	: Ukuran ketepatan model dalam memprediksi hasil yang benar.
<i>Performance</i>	: Kinerja model atau sistem, biasanya diukur dengan metrik seperti akurasi, presisi, atau <i>recall</i> .
<i>Software</i>	: Perangkat lunak yang memiliki komponen penting pada komputer agar komputer dapat berjalan sesuai dengan perintah
Korelasi	: Hubungan atau keterkaitan antara dua variabel, baik positif maupun negatif.

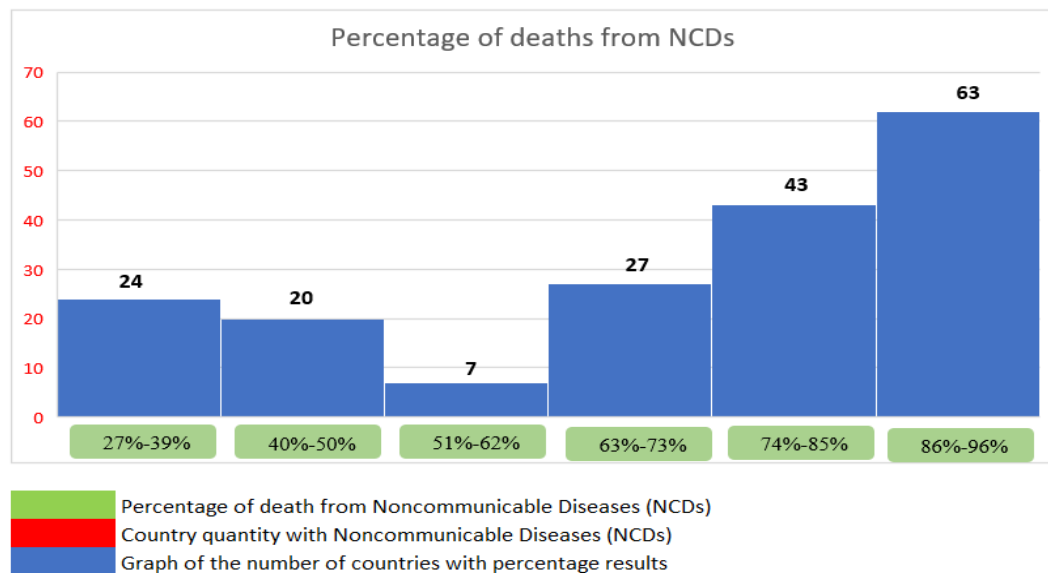


BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit Tidak Menular (PTM) atau yang biasa dikenal dengan penyakit degeneratif. PTM merupakan salah satu penyebab masalah kesehatan bagi masyarakat di seluruh dunia, PTM tidak dapat menular dari penderita kepada orang lain, perkembangan penyakit ini terjadi secara perlahan dan dalam jangka waktu yang lama [1]. Menurut *World Health Organizations*, penyebab kematian di dunia 70% disebabkan oleh penyakit tidak menular (PTM) meliputi kanker, penyakit pernapasan kronis, diabetes, dan kardiovaskular [2]. Diabetes merupakan salah satu penyakit tidak menular (PTM) yang menyebabkan kadar glukosa darah meningkat karena tubuh tidak dapat memproduksi insulin [3]. Diabetes melitus (DM) merupakan penyakit metabolik kronis yang ditandai dengan kadar glukosa darah tinggi (*hiperglikemia*) akibat kelainan produksi insulin, kerja insulin, atau keduanya [4].



Gambar 1.1 Kumpulan Data Diabetes

Berdasarkan grafik di atas, terdapat kematian yang disebabkan oleh penyakit tidak menular terutama Diabetes Melitus berupa komplikasi antara lain serangan jantung dan stroke, infeksi kaki berat (menyebabkan gangren yang dapat berujung pada amputasi), gagal ginjal terminal dan disfungsi seksual [5].

Metode yang digunakan dalam penelitian ini adalah *Knowledge Discovery on Database* (KDD). *Knowledge Discovery in Database* (KDD) merupakan

proses menemukan pola-pola yang menarik, pengetahuan baru, dan informasi yang berguna dari data yang tersimpan dalam database. Tujuan utama KDD adalah mengubah data mentah menjadi pengetahuan yang dapat digunakan untuk pengambilan keputusan yang lebih baik [6].

Berdasarkan Gambar 1 di atas, penelitian ini bertujuan untuk memperluas pengetahuan kita tentang penggunaan data mining dalam klasifikasi penyakit diabetes. Dengan menggabungkan teknologi komputer dan informasi medis, diharapkan penelitian ini dapat memberikan manfaat yang signifikan dalam upaya pencegahan, diagnosis, dan pengobatan penyakit diabetes. Penelitian ini menggunakan dataset yang diperoleh dari *UCI Machine Learning* yang berjudul *datasets_diabetes* sebanyak 9 atribut.

Pada penelitian sebelumnya terdapat perbedaan dengan penelitian yang saat ini dilaksanakan, terutama pada dataset yang digunakan dan fokus penelitian. Dalam penelitian ini, dataset yang digunakan adalah dataset diabetes dari Kaggle, yang memiliki jumlah data lebih besar serta atribut yang lebih variatif dibandingkan dataset dari penelitian sebelumnya. Peneliti memilih metode *Naïve Bayes Classification*, karena metode ini sederhana, mudah dipahami, dan mampu mengolah data dengan jumlah besar serta memberikan hasil prediksi yang cukup akurat dengan asumsi independensi antar fitur. Dengan pendekatan ini, penelitian diharapkan memberikan kontribusi dalam peningkatan akurasi prediksi dan interpretasi variabel-variabel utama yang berpengaruh terhadap klasifikasi penyakit diabetes.

Penelitian ini bertujuan untuk mengembangkan model klasifikasi yang dapat memprediksi risiko diabetes pada populasi tertentu menggunakan pendekatan algoritma *Naïve Bayes*. Penelitian ini menggunakan kumpulan data yang terdiri dari parameter medis dan demografi sejumlah pasien yang telah didiagnosis menderita diabetes serta mereka yang belum.

Hasil penelitian menunjukkan bahwa model *Naïve Bayes* mampu memberikan prediksi status diabetes yang cukup baik, dengan rata-rata akurasi sebesar 95,5%. Temuan ini menunjukkan potensi penggunaan pendekatan algoritma *Naïve Bayes* dalam pengembangan sistem prediksi diabetes yang dapat

digunakan dalam praktik klinis untuk mendukung pengambilan keputusan yang lebih tepat waktu dan akurat.

1.2 Rumusan Masalah

1. Bagaimana algoritma Naive Bayes dapat diunakan untuk memprediksi status diabetes berdasarkan data pasien?
2. Apa saja faktor yang paling memengaruhi risiko diabetes berdasarkan hasil analisis data?
3. Bagaimana cara mendeteksi dini dan memprediksi faktor-faktor utama risiko diabetes secara akurat untuk membantu pencegahan dan penanganan yang lebih efektif?
4. Bagaimana ke akuratan sistem berbasis data yang memanfaatkan *machine learning* untuk membantu pencegahan dan pengobatan diabetes secara efektif?

1.3 Tujuan Penelitian

1. Mengembangkan model prediksi risiko diabetes menggunakan algoritma Naïve Bayes.
2. Mengidentifikasi faktor-faktor yang berpengaruh terhadap klasifikasi penyakit diabetes.
3. Mengevaluasi performa model prediksi berdasarkan metrik seperti akurasi, presisi, dan sensitivitas.

1.4 Manfaat Penelitian

1. Bagi Praktisi Kesehatan: Menyediakan alat bantu untuk mendeteksi diabetes secara cepat, akurat, dan efisien, sehingga membantu dalam pengambilan keputusan medis.
2. Bagi Pengembangan Ilmu Pengetahuan: Memberikan kontribusi dalam penelitian data mining, khususnya penggunaan algoritma Naïve Bayes untuk klasifikasi penyakit.
3. Bagi Masyarakat: Meningkatkan kesadaran dan deteksi dini terhadap risiko diabetes, sehingga dapat mencegah komplikasi lebih lanjut.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan pada proses tahapan penelitian yang dilakukan mengenai analisis Pemodelan Klasifikasi Prediksi Diabetes Menggunakan Pendekatan Algoritma Naïve Bayes, maka peneliti dapat menarik kesimpulan sebagai berikut:

1. Diabetes melitus (DM) merupakan salah satu penyakit kronis yang semakin meningkat prevalensinya di seluruh dunia. Oleh karena itu, deteksi dini dan prediksi risiko diabetes sangat penting untuk pengelolaan dan pencegahan penyakit ini. Penelitian ini bertujuan untuk mengembangkan model klasifikasi yang dapat memprediksi status diabetes pada individu menggunakan algoritma Naïve Bayes, dengan memanfaatkan data medis dan demografi pasien.
2. Dari hasil penelitian yang dilakukan, didapatkan bahwa model Naïve Bayes dapat memberikan hasil yang cukup baik dalam memprediksi status diabetes, dengan akurasi mencapai 95,5%. Model ini mampu melakukan klasifikasi antara pasien yang terdiagnosis diabetes dan yang tidak, berdasarkan sembilan atribut yang mencakup data medis dan demografi, seperti jenis kelamin, usia, hipertensi, riwayat penyakit jantung, riwayat merokok, BMI, kadar HbA1c, kadar glukosa darah, dan status diabetes.
3. Penelitian ini juga menunjukkan bahwa beberapa atribut memiliki pengaruh yang signifikan terhadap prediksi diabetes, di antaranya kadar glukosa darah, tingkat HbA1c, usia, dan indeks massa tubuh (BMI). Kadar glukosa darah dan tingkat HbA1c terbukti menjadi faktor yang paling dominan dalam menentukan status diabetes, yang mengindikasikan bahwa pengelolaan kadar glukosa darah menjadi salah satu langkah utama dalam pencegahan dan pengobatan diabetes.
4. Penggunaan algoritma Naïve Bayes untuk prediksi diabetes dalam penelitian ini menunjukkan potensi besar dalam penerapannya di bidang medis, khususnya dalam sistem pendukung keputusan klinis. Dengan akurasi yang tinggi, model ini dapat digunakan oleh tenaga medis untuk

membantu dalam proses pengambilan keputusan yang lebih cepat dan lebih tepat, terutama dalam hal deteksi dini dan pemantauan kondisi pasien. Selain itu, model ini juga dapat diadaptasi untuk diterapkan pada berbagai data pasien dengan karakteristik yang berbeda di masa depan.

5. Secara keseluruhan, penelitian ini memberikan kontribusi signifikan dalam pengembangan sistem prediksi diabetes berbasis data mining yang dapat diintegrasikan dalam praktik klinis untuk mendukung pencegahan, diagnosis, dan pengobatan diabetes yang lebih efektif. Dengan adanya sistem prediksi yang lebih akurat, diharapkan dapat mengurangi jumlah kasus diabetes yang tidak terdiagnosis dan memberikan pengelolaan kesehatan yang lebih baik untuk pasien di masa depan.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, maka peneliti menyadari masih terdapat banyak keterbatasan dan kekeliruan pada penelitian ini. Namun dengan adanya penelitian ini, diharapkan mampu memberikan urgensi yang bermanfaat terutama bagi masyarakat. Maka dari itu, peneliti memberikan saran sebagai berikut:

1. Selain algoritma Naïve Bayes, penelitian lebih lanjut dapat mencoba algoritma machine learning lain, seperti Random Forest, Support Vector Machine (SVM), atau Deep Learning, untuk membandingkan performa model dan menentukan algoritma yang paling optimal.
2. Penelitian ini menggunakan dataset dengan jumlah sampel dan atribut yang terbatas. Untuk meningkatkan generalisasi dan akurasi model, disarankan untuk menggunakan dataset yang lebih besar dan mencakup populasi dengan karakteristik yang beragam, termasuk variasi etnis, geografis, dan pola hidup.
3. Disarankan untuk menambahkan variabel lain yang relevan dalam analisis risiko diabetes, seperti faktor genetik, tingkat aktivitas fisik, pola makan, tingkat stres, dan kondisi lingkungan. Dengan menambahkan variabel tersebut, model dapat memberikan prediksi yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] Pusat Data dan Informasi Kemenkes RI, “Penyakit Tidak Menular,” *Bul. Jendela Data dan Inf. Kesehat.*, vol. 2, pp. 1–14, 2012.
- [2] WHO, *Noncommunicable Diseases: Progress Monitor*. 2022. [Online]. Available:
<https://apps.who.int/iris/bitstream/handle/10665/353048/9789240047761-eng.pdf>
- [3] M. R. Hunafa and A. Hermawan, “KLIK: Kajian Ilmiah Informatika dan Komputer Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Imbalance Class Dataset Penyakit Diabetes,” *Media Online*, vol. 4, no. 3, pp. 1551–1561, 2023, doi: 10.30865/klik.v4i3.1486.
- [4] S. Rokhanah, A. Hermawan, and D. Avianto, “Pengaruh Principal Component Analysis Pada Naïve Bayes dan K-Nearest Neighbor Untuk Prediksi Dini Diabetes Melitus Menggunakan Rapidminer,” *EVOLUSI J. Sains dan Manaj.*, vol. 11, no. 1, 2023, doi: 10.31294/evolusi.v11i1.14728.
- [5] R. Hellmund, R. Weitgasser, and D. Blissett, “Cost calculation for a flash glucose monitoring system for UK adults with type 1 diabetes mellitus receiving intensive insulin treatment,” *Diabetes Res. Clin. Pract.*, vol. 138, pp. 193–200, 2018, doi: 10.1016/j.diabres.2018.01.028.
- [6] D. Muhammad, “Perkembangan Dan Transformasi Teknologi Digital,” *Infokam*, vol. 15, no. 2, pp. 116–123, 2019.
- [7] A. Ridwan, “Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020, doi: 10.47970/siskom-kb.v4i1.169.
- [8] H. Nugroho, G. E. Yuliasuti, and A. Firman, “Klasifikasi Diagnosis Diabetes Melitus Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Backward Elimination,” *J. Ilm. NERO*, vol. 8, no. 2, p. 2023, 2023, [Online]. Available: <https://journal.trunojoyo.ac.id/nero/article/view/21110>
- [9] Bisola Oluwafadekemi Adegoke, Tolulope Odugbose, and Christiana Adeyemi, “Data analytics for predicting disease outbreaks: A review of

- models and tools,” *Int. J. Life Sci. Res. Updat.*, vol. 2, no. 2, pp. 001–009, 2024, doi: 10.53430/ijlsru.2024.2.2.0023.
- [10] A. Veronica Agustin and A. Voutama, “Implementasi Data Mining Klasifikasi Penyakit Diabetes Pada Perempuan Menggunakan Naïve Bayes,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1002–1007, 2023, doi: 10.36040/jati.v7i2.6808.
- [11] A. Atthohiroh, R. Ayu, and S. Maharani, “Penerapan Metode Naive Bayes Untuk Memprediksi Penyakit Jantung,” *J. Tek.*, vol. 3, no. 1, p. 8, 2023, doi: 10.54314/teknisi.v3i1.1252.
- [12] M. Sholeh, E. K. Nurnawati, and U. Lestari, “Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, pp. 10–21, 2023, doi: 10.14421/jiska.2023.8.1.10-21.
- [13] S. Ucha Putri, E. Irawan, F. Rizky, S. Tunas Bangsa, P. A. -Indonesia Jln Sudirman Blok No, and S. Utara, “Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5,” *Januari*, vol. 2, no. 1, pp. 39–46, 2021.

