

IMPLEMENTASI KLASIFIKASI KEHAMILAN BERESIKO DENGAN METODE *NAIVE BAYES* PADA PUSKESMAS KELURAHAN MALAKA JAYA

Virginia Khoirunnisa ^{1*}, Sri Lestari ²

^{1,2} Program Studi Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia

Email: ginnyvirginia67@gmail.com ^{1*}, sri.lestari1203@gmail.com ²

Histori Artikel:

Dikirim 23 Juli 2023; Diterima dalam bentuk revisi 19 Agustus 2023; Diterima 25 Agustus 2023; Diterbitkan 10 September 2023. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Resiko kehamilan diklasifikasikan menjadi tiga kategori: Kehamilan Resiko Rendah (KRR), Kehamilan Resiko Tinggi (KRT), dan Kehamilan Resiko Sangat Tinggi (KRST). Data pemeriksaan yang tercatat dalam buku Kesehatan Ibu dan Anak (KIA) di Puskesmas digunakan untuk mengelompokkan resiko kehamilan. Meskipun angka kematian ibu telah mengalami penurunan, survei terbaru menunjukkan peningkatan yang signifikan. Penyebabnya antara lain kurangnya penyuluhan mengenai kesehatan obstetri dan ginekologi, kurangnya pelayanan emergensi obstetrik dan neonatal, serta penggunaan pengobatan tradisional yang tidak efektif. Untuk mengatasi masalah ini, diperlukan peningkatan pelayanan kesehatan ibu yang berkualitas. Upaya percepatan penurunan Angka Kematian Ibu (AKI) termasuk akses pelayanan kesehatan yang tepat, perawatan pasca persalinan, dan pelayanan keluarga berencana. Pengolahan data menggunakan teknik data mining dengan metode klasifikasi Naïve Bayes, yang memiliki akurasi tinggi dan waktu eksekusi yang pendek. Penelitian ini bertujuan untuk mengembangkan model klasifikasi Naïve Bayes dalam deteksi dini resiko kehamilan, sehingga dapat membantu mendiagnosa ibu hamil dengan resiko rendah atau tinggi.

Kata Kunci: Resiko Kehamilan; Naive Bayes; Angka Kematian Ibu (AKI); Persalinan; Klasifikasi.

Abstract

Pregnancy risk is classified into three categories: Low Risk Pregnancy (KRR), High Risk Pregnancy (KRT), and Very High Risk Pregnancy (KRST). Examination data recorded in the Maternal and Child Health (MCH) book at Puskesmas is used to categorize pregnancy risk. Although the maternal mortality rate has decreased, recent surveys show a significant increase. The causes include lack of counseling on obstetric and gynecological health, lack of emergency obstetric and neonatal care, and the use of ineffective traditional medicine. To overcome this problem, it is necessary to improve quality maternal health services. Efforts to accelerate the reduction of maternal mortality rate (MMR) include access to appropriate health services, postpartum care, and family planning services. Data processing uses data mining techniques with the Naïve Bayes classification method, which has high accuracy and short execution time. This study aims to develop a Naïve Bayes classification model in early detection of pregnancy risk, so that it can help diagnose pregnant women with low or high risk.

Keyword: Pregnancy Risk; Naive Bayes; Maternal Mortality Rate (MMR); Childbirth; Classification.

1. Pendahuluan

Masa kehamilan adalah masa yang membahagiakan bagi kehidupan rumah tangga setiap insan manusia. Kehamilan juga memiliki resiko yang harus diperhatikan dengan baik agar tidak membahayakan bayi maupun ibu hamil. Resiko kehamilan dapat diketahui dengan melakukan deteksi dini kehamilan dengan faktor resiko sehingga tenaga kesehatan dapat mengetahui penanganan yang lebih lanjut. Untuk menghindari atau menanggulangi adanya resiko maka perlu dilakukan pemeriksaan secara rutin atau berkala pada bidan atau dokter. Pemeriksaan rutin pada bidan dapat dilakukan di Pos Pelayanan Terpadu (Posyandu), puskesmas, atau lokasi praktek bidan. Penanganan resiko kehamilan merupakan kunci utama keberhasilan dalam penanggulangan penurunan angka kematian ibu dan bayi yang dilahirkan.

Deteksi dini resiko kehamilan diimplementasikan dengan mengelompokkan resiko kehamilan menjadi tiga yaitu Kehamilan Resiko Rendah (KRR), Kehamilan Resiko Tinggi (KRT) dan kehamilan Resiko sangat tinggi (KRST) (Rochyati, 2003). Resiko kehamilan dikelompokkan berdasarkan data pemeriksaan yang dilakukan di Puskesmas dan dicatat dalam buku Kesehatan Ibu dan Anak (KIA).

Angka kematian ibu di Indonesia pada tahun 1991 sampai dengan tahun 2007 telah mengalami penurunan dari 390 menjadi 228 per 100.000 kelahiran hidup. Pada tahun 2012 dalam Survei Demografi Kependudukan Indonesia kembali tercatat adanya kenaikan angka kematian ibu yang sangat signifikan, yaitu dari 228 jiwa menjadi 359 jiwa kematian ibu per 100.000 kelahiran hidup. Berdasarkan hasil Survei Demografi dan Kesehatan Indonesia pada tahun 2012, angka kematian Anak tahun 2012 mencapai sebesar 19 jiwa per 1.000 kelahiran hidup. Angka ini sama dengan AKN berdasarkan SDKI tahun 2007 dan hanya menurun 1 point dibanding SDKI tahun 2002-2003 yaitu 20 per 1.000 kelahiran hidup.

Banyak masalah yang menjadi penyebab adanya angka kematian tinggi. Pertama, karena kurangnya penyuluhan informasi mengenai kesehatan obstetri dan ginekologi. Kedua, kurangnya pelayanan emergensi obstetrik dan neonatal dasar dan komprehensif yang dapat dijangkau secara tepat waktu oleh masyarakat yang membutuhkan (Profil Kesehatan Indonesia 2014). Ketiga, karena masih banyak yang mengatasi masalah kesehatan tersebut dengan pengobatan tradisional yang tidak jarang kontra produktif dengan keadaan tubuh, sehingga bila sudah mengalami gejala yang cukup serius baru datang ke dokter. Namun, hasil analisa menyimpulkan bahwa keadaannya sangat mencemaskan bagi perempuan Indonesia. Walaupun pelayanan antenatal (pemeriksaan sebelum kelahiran) dan pertolongan persalinan oleh tenaga kesehatan terlatih cukup tinggi, beberapa faktor seperti risiko tinggi pada saat kehamilan dan aborsi perlu mendapat perhatian. Ke depan, upaya peningkatan kesehatan ibu diprioritaskan pada perluasan pelayanan kesehatan berkualitas.

Berdasarkan penyebab, sebagian besar kematian ibu pada tahun 2021 terkait COVID-19 sebanyak 2.982 kasus, perdarahan sebanyak 1.330 kasus, dan hipertensi dalam kehamilan sebanyak 1.077 kasus. Upaya percepatan penurunan AKI dilakukan dengan menjamin agar setiap ibu mampu mengakses pelayanan kesehatan yang berkualitas, seperti pelayanan kesehatan ibu hamil, pertolongan persalinan oleh tenaga kesehatan terlatih di fasilitas pelayanan kesehatan, perawatan pasca persalinan bagi ibu dan bayi, perawatan khusus dan rujukan jika terjadi komplikasi, dan pelayanan keluarga berencana (KB) termasuk KB pasca persalinan.

Kesehatan ibu yang disajikan terdiri dari pelayanan kesehatan ibu hamil, pelayanan imunisasi Tetanus Difteri bagi Wanita Usia Subur (WUS), pemberian tablet tambah darah, pelayanan kesehatan ibu bersalin, pelayanan kesehatan ibu nifas, puskesmas melaksanakan kelas ibu hamil dan Program Perencanaan Persalinan dan Pencegahan Komplikasi (P4K), pelayanan kontrasepsi/Keluarga Berencana (KB), dan pemeriksaan HIV serta Hepatitis B. Berdasarkan prosentase penyebab kematian ibu menunjukkan bahwa kesehatan bagi ibu hamil sangatlah penting.

Pengolahan data dilakukan dengan teknik data mining dengan memperkirakan adanya kemungkinan yang terjadi berdasarkan kejadian sebelumnya. Data mining mengelola data yang pernah ada sehingga menghasilkan informasi baru untuk data baru. Salah satu metode data mining yang cocok untuk studi kasus ini adalah klasifikasi, metode ini memiliki beberapa algoritma salah satunya *naïve bayes*.

Klasifikasi *naïve bayes* digunakan untuk menghitung peluang yang terjadi dalam satu kelas dengan atribut yang telah ditentukan. *Naïve bayes* relatif memiliki akurasi yang tinggi dan waktu eksekusi yang pendek jika dibandingkan dengan metode klasifikasi lainnya, Penelitian ini akan membuat “Klasifikasi kehamilan beresiko Menggunakan Metode *Naïve Bayes*” yang dapat digunakan untuk mendiagnosa ibu hamil mengalami resiko rendah, atau resiko tinggi.

Penelitian terdahulu yang dilakukan oleh para peneliti telah mengaplikasikan Metode Naive Bayes dalam berbagai bidang penelitian. Arifin, Handoko, dan Efendi (2022) melakukan penelitian tentang implementasi Metode Naive Bayes untuk mengklasifikasikan penerima Program Keluarga Harapan [1]. Sementara itu, Putry (2022) melakukan komparasi antara algoritma K-Nearest Neighbors (KNN) dan Naive Bayes dalam klasifikasi diagnosis penyakit diabetes mellitus [2]. Firdaus dan Yanti (2022) menerapkan Metode Naive Bayes pada sistem pakar untuk mendiagnosa penyakit gigi berbasis web [3]. Penelitian lainnya oleh Rizky dan Hakim (2020) berkaitan dengan sistem pakar dalam menentukan penyakit hipertensi pada ibu hamil di RSUD Adjudarmo Rangkasbitung Provinsi Banten, yang juga menggunakan Metode Naive Bayes [4]. Apriyani dan Kurniati (2020) melakukan perbandingan antara Metode Naive Bayes dan Support Vector Machine dalam klasifikasi penyakit diabetes mellitus [5]. Zainiyah, Susanti, dan Setiawati (2021) fokus pada deteksi dini preeklamsi pada ibu hamil dengan berbagai parameter termasuk Indeks Massa Tubuh (IMT), Roll Over Test (ROT), dan Mean Arterial Pressure (MAP) [6].

Selain itu, Hidayah (2020) mengembangkan sistem monitoring kondisi kesehatan sebelum dan sesudah olahraga menggunakan Metode Naive Bayes [7]. Putro, Vlandari, dan Saptomo (2020) mengaplikasikan Metode Naive Bayes dalam klasifikasi pelanggan [8], sementara Nurdiana, Rodiyansyah, dan Algifari (2020) melakukan studi komparasi antara algoritma ID3 dan Naive Bayes dalam klasifikasi penyakit diabetes mellitus [9]. Hidayati dan Utomo (2019) mengaplikasikan Metode Naive Bayes dalam pemeriksaan kesehatan [10]. Penelitian lainnya termasuk penggunaan Metode Naive Bayes dalam klasifikasi berita hoax oleh Mustofa dan Mahfudh (2019) [11], deteksi dini preeklamsi pada ibu hamil oleh Setiawati dan Lailiyah (2020) [12], analisis sentimen terhadap penerapan sistem plat nomor ganjil/genap pada Twitter oleh Ruhyana (2019) [13], klasifikasi kelayakan keluarga penerima beras Rastra oleh Fadlan, Ningsih, dan Windarto (2018) [14], serta analisis publik terhadap kebijakan lockdown Pemerintah Jakarta menggunakan algoritma SVM oleh Isnain *et al.* (2021) [15]. Selanjutnya, prediksi tingkat kelancaran pembayaran sewa teras UMKM oleh Rachman, Handayani, dan Artikel (2021) [16], prediksi penyakit lambung oleh Jefa *et al.* (2021) [17], perkiraan waktu studi mahasiswa oleh Fadrial (2021) [18], dan prediksi heregistrasi calon mahasiswa baru oleh Wibowo dan Manan (2022) [19]. Semua penelitian ini mencerminkan beragam aplikasi Metode Naive Bayes dalam berbagai bidang dan konteks. Dengan demikian, penelitian Implementasi Klasifikasi Kehamilan Beresiko Dengan Metode Naive Bayes Pada Puskesmas Kelurahan Malaka Jaya terkait erat dengan penelitian-penelitian terdahulu yang telah menggunakan Metode Naive Bayes dalam berbagai konteks kesehatan dan klasifikasi. Hasil penelitian ini diharapkan dapat memberikan kontribusi positif dalam upaya identifikasi kehamilan berisiko tinggi dan perawatan yang lebih baik bagi ibu hamil di Puskesmas Kelurahan Malaka Jaya.

2. Metode Penelitian

2.1 Penerapan Metodologi

Dalam penelitian ini penulis menggunakan beberapa metode untuk memproses data dan algoritma yang digunakan adalah algoritma *Naive Bayes*. Proses klasifikasi dibagi menjadi dua fase yaitu learning/training dan testing/ classify (Ginting & Trinanda, 2013). Pada tahap pembelajaran, model perkiraan digunakan untuk beberapa data yang kelas datanya diketahui. Kemudian model estimasi yang dibuat pada tahap pengujian diuji terhadap data lain untuk mengetahui keakuratan model.

Persamaan 1 Teorema Bayes:

$$P(H.X) = \frac{P(X.H)xP(H)}{P(X)} \tag{1}$$

Keterangan

- X = Data dengna kelas yang belum diketahui
- H = Hipotesis data X merupakan suatu kelas spesifik
- $P(H.X)$ = Probabulitas hipotesis H berdasarkan kondisi X (*Posterior* probability)
- $P(H)$ = Probabilitas hipotesis H (prior probability)
- $P(X.H)$ = Probabilitas X berdasarkan kondisi hipotesis H
- $P(X)$ = Probabilitas X

Untuk menjelaskan teorema *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas yang cocok bagi sampel yang dianalisis tersebut (Bustami, 2014). Karena itu, teorema bayes di atas disesuaikan seperti persamaan 2.

$$P(C, F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n.C)}{P(F_1 \dots F_n)} \tag{2}$$

Dimana variabel C merepresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, sering kali disebut prior), dikali dengan peluang kemunculan karakteristik – karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik – karakteristik sampel secara global (disebut juga *evidence*) (Bustami, 2014).

Dimana variabel C merepresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, sering kali disebut prior), dikali dengan peluang kemunculan karakteristik – karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik – karakteristik sampel secara global (disebut juga *evidence*) (Bustami, 2014). Nilai bukti selalu sama untuk setiap kelas sampel. Nilai *Posterior* kemudian dibandingkan dengan nilai *Posterior* kelas lain untuk menentukan kelas sampel yang akan diklasifikasi. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan $(C|F_1, \dots, F_n)$ menggunakan aturan perkalian (Bustami, 2014), seperti pada persamaan 3.

$$\begin{aligned} P(C, F_1 \dots F_n) &= P(C, F_1 \dots C) \\ &\hat{=} P(C)P(F_1, C) P(F_2 \dots F_n, C, F_1) \\ &\hat{=} P(C)P(F_1, C) P(F_2, C, F_1)P(F_3 \dots F_n, C, F_1, F_2) \\ &\hat{=} P(C)P(F_1, C) P(F_2, C, F_1)P(F_4 \dots F_n, C, F_1, F_2, F_3) \\ &\hat{=} P(C)P(F_1, C) P(F_2, C, F_1) \dots P(F_n, C, F_1, F_2, F_3 \dots F_n=1) \end{aligned} \tag{3}$$

Dari penjabaran diatas dapat dilihat semakin banyak dan semakin kompleksnya faktor–faktor yang mempengaruhi nilai probabilitas, yang hampir tidak mungk in untuk dianalisa satu persatu. Akibatnya perhitungan tersebut menjadi sulit untuk dilakukan, dari pernyataan itulah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing–masing petunjuk ($F_1, F_2 \dots F_n$) saling bebas (independen) satu sama lain (Bustami, 2014). Selanjutnya, penjabaran $P(C|F_1, \dots, F_n)$ dapat disederhanakan menjadi persamaan 4.

langkah selanjutnya adalah melakukan evaluasi model secara deskriptif. Evaluasi model bertujuan untuk mengukur kinerja model dalam melakukan klasifikasi berdasarkan data yang diberikan.

3. Hasil dan Pembahasan

3.1 Dataset Collection

Data *Collection* adalah proses pengumpulan informasi atau data dari berbagai sumber untuk tujuan analisis atau penelitian. Tujuan utama dari data collection adalah untuk mengumpulkan informasi yang relevan dan akurat yang dapat digunakan untuk mengambil keputusan atau menghasilkan wawasan baru. Proses data collection dapat melibatkan berbagai metode, termasuk survei, wawancara, observasi, pengumpulan data sekunder, atau menggunakan teknologi seperti sensor atau perangkat lunak khusus untuk mengumpulkan data secara otomatis.

Data *Collection* sangat penting dalam berbagai bidang, termasuk ilmu sosial, ilmu komputer, bisnis, kesehatan, dan banyak lagi. Contohnya, dalam penelitian ilmiah, data *collection* adalah langkah awal yang kritis untuk mengumpulkan data yang diperlukan untuk menjawab pertanyaan penelitian dan menguji hipotesis. Namun, penting untuk diingat bahwa data collection harus dilakukan dengan memperhatikan etika dan privasi data. Pengumpulan data harus dilakukan dengan izin yang tepat dan mematuhi regulasi privasi yang berlaku untuk melindungi informasi pribadi individu atau organisasi yang terlibat.

Data yang digunakan dalam penelitian ini adalah data kesehatan ibu hamil pada tahun 2022 yang diperoleh dari izin Puskesmas Malaka Jaya. Dataset ini terdiri dari 832 baris dan 25 kolom. Dataset ini mencakup beberapa variabel prediktor medis (variabel independen) dan satu variabel target atau hasil (variabel dependen). Variabel target yang digunakan adalah tingkat risiko kehamilan, yang terdiri dari tiga kelas: tidak beresiko, beresiko, dan sangat beresiko. Penelitian ini bertujuan untuk melakukan klasifikasi tingkat risiko kehamilan pada ibu hamil menggunakan dataset ini.

3.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah suatu pendekatan atau proses untuk menganalisis data dengan tujuan memahami karakteristik dan pola yang terdapat dalam data. EDA biasanya dilakukan sebelum penerapan model statistik atau pemodelan yang lebih lanjut. Tujuan utama EDA adalah untuk mendapatkan pemahaman yang mendalam tentang data dan mengidentifikasi pola, anomali, ketergantungan, atau tren yang mungkin ada di dalamnya. Metode EDA sering melibatkan penggunaan visualisasi data, seperti grafik, diagram, dan plot, untuk menggambarkan data dengan cara yang mudah dipahami. Berikut adalah beberapa hasil yang dapat diperoleh melalui *exploratory data analysis* (EDA).

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 832 entries, 0 to 831
Data columns (total 25 columns):
#   Column                                     Non-Null Count  Dtype
---  ---                                     -
0   NO                                         832 non-null    int64
1   TANGGAL                                   202 non-null    object
2   NO REGISTER                               826 non-null    object
3   PASIEN BARU/LAMA                         831 non-null    object
4   NAMA BUMIL                               832 non-null    object
5   UMUR (TAHUN)                             830 non-null    object
6   USIA KEHAMILAN (MINGGU)                 797 non-null    float64
7   ALAMAT LENGKAP                          830 non-null    object
8   KTP DKI /NON /NIK                       829 non-null    object
9   BPJS /NON BPJS / NO                    829 non-null    object
10  NAMA SUAMI                              830 non-null    object
11  PEKERJAAN SUAMI                        830 non-null    object
12  ( G.. / P.. / A.. )                   831 non-null    object
13  BB (kg)                                 828 non-null    object
14  TB (cm)                                 830 non-null    object
15  LiLA (cm)                              829 non-null    object
16  HB                                       805 non-null    object
17  KETERANGAN K1 MURNI                    728 non-null    object
18  KETERANGAN K1 AKSES                    736 non-null    object
19  HPHT                                    776 non-null    object
20  TP                                       774 non-null    object
21  TT                                       775 non-null    object
22  FE                                       816 non-null    object
23  TINGKAT RESIKO KEHAMILAN                722 non-null    object
24  KETERANGAN                              817 non-null    object
dtypes: float64(1), int64(1), object(23)
memory usage: 162.6+ KB
```

Gambar 2. Informasi Dataset

Berdasarkan data di atas, memberikan informasi tentang dataset yang meliputi jumlah data yang tidak terisi (*missing values/null*) dan jenis data dari setiap kolom.

3.3 Data Preprocessing

Data *Preprocessing* adalah proses persiapan data sebelum dilakukan analisis atau pemodelan. Tujuannya adalah untuk membersihkan, mengubah, dan mengorganisir data sehingga lebih siap digunakan dalam proses analisis lebih lanjut. Data *preprocessing* penting untuk memastikan kualitas dan kecocokan data sebelum diterapkan pada algoritma atau model.

Setelah proses data *preprocessing*, dataset ini terdiri dari 730 baris dan 11 kolom. Terdapat 10 fitur (variabel independen) yang akan digunakan dalam membangun model, yaitu:

- 1) Pasien baru/lama
- 2) Umur
- 3) Usia kehamilan
- 4) Pekerjaan suami
- 5) Berat badan
- 6) Tinggi badan
- 7) Lingkar lengan
- 8) Hemoglobin
- 9) TT (Tetanus Toxoid)
- 10) Fe (Suplemen Zat Besi)

Selain itu, terdapat 1 variabel target yaitu "Tingkat Resiko Kehamilan". Hasil dari data preprocessing mencakup pemrosesan dan transformasi data seperti menangani missing values, pembersihan data, pemskaalan fitur, pengkodean variabel kategorikal, seleksi fitur, dan transformasi data. Hal ini bertujuan untuk mempersiapkan dataset agar siap digunakan dalam analisis atau pemodelan selanjutnya. Berikut merupakan hasil yang diperoleh setelah data *preprocessing*:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 730 entries, 0 to 831
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                ---
0   PASIEN BARU/LAMA                       730 non-null    int64
1   UMUR (TAHUN)                           730 non-null    int64
2   USTA KEHAMILAN (MINGGU)                730 non-null    float64
3   PEKERJAAN SUAMI                        730 non-null    int64
4   BB (kg)                                 730 non-null    float64
5   TB (cm)                                 730 non-null    float64
6   LiLA (cm)                              730 non-null    float64
7   HB                                       730 non-null    float64
8   TT                                       730 non-null    int64
9   FE                                       730 non-null    int64
10  TINGKAT RESIKO KEHAMILAN               730 non-null    int64
dtypes: float64(5), int64(6)
memory usage: 68.4 KB
```

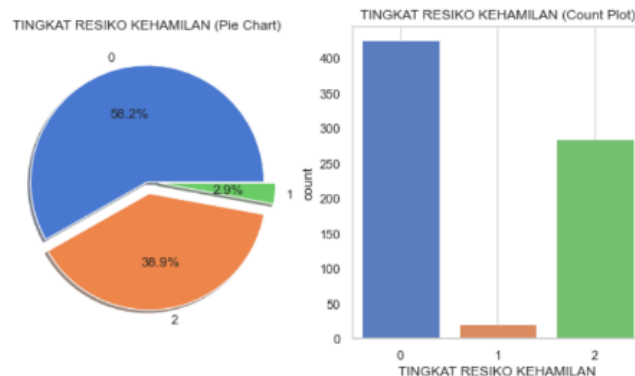
Gambar 3. Hasil Data *Preprocessing*

Berdasarkan data diatas merupakan informasi hasil dari preprocessing data sehingga tidak ada lagi missing values/null dan jenis data yang sudah ditransformasi sudah bukan object lagi sehingga dapat diproses dalam membangun model.

3.4 Data *Visualization*

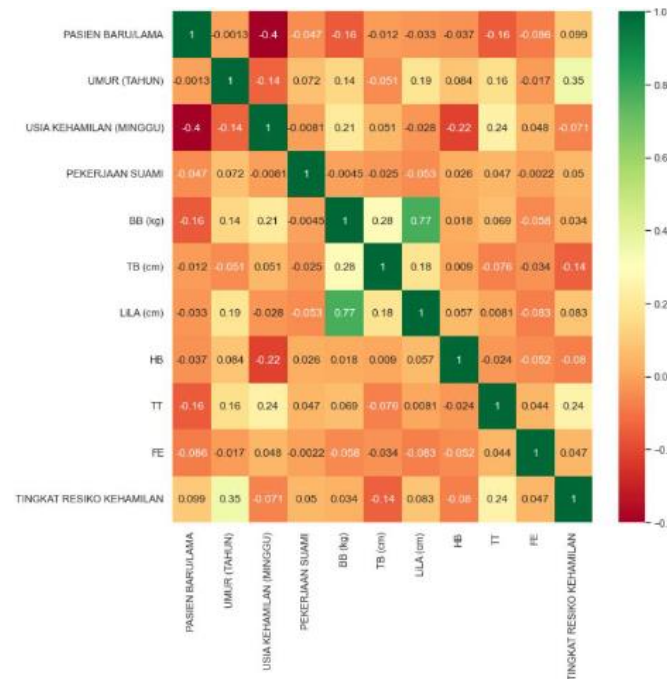
Data *Visualization* adalah proses representasi grafis atau visualisasi data dan informasi menggunakan berbagai jenis grafik, diagram, dan plot. Tujuannya adalah untuk menyajikan data dengan cara yang mudah dipahami. Berikut merupakan gambaran tingkat resiko kehamilan dari dataset yang digunakan:

- 1.) Tidak Beresiko (0): 425
- 2.) Beresiko (1): 21
- 3.) Sangat Beresiko (2): 284



Gambar 4. Pie Chart dan Histogram Variabel Target

Berdasarkan grafik, dapat dilihat bahwa persentase variabel target adalah sebagai berikut: Tidak Beresiko sebesar 58.2%, Beresiko sebesar 2.9%, dan Sangat Beresiko sebesar 38.9%. Diagram tersebut juga menggambarkan frekuensi variabel target, yaitu Tidak Beresiko sebanyak 425, Beresiko sebanyak 21, dan Sangat Beresiko sebanyak 284. Selanjutnya merupakan gambaran korelasi antar variabel yang digunakan dalam membangun model. Berikut grafik korelasinya:



Gambar 5. Tingkat Korelasi antar Variabel

Korelasi antar variabel mengacu pada hubungan statistik antara dua atau lebih variabel dalam sebuah dataset. Ini memberikan informasi tentang sejauh mana perubahan dalam satu variabel berhubungan dengan perubahan dalam variabel lainnya. Korelasi dapat menggambarkan arah (positif atau negatif) dan kekuatan hubungan antara variabel-variabel tersebut. Korelasi diukur menggunakan koefisien korelasi, yang paling umum adalah koefisien korelasi Pearson. Koefisien korelasi Pearson berkisar dari -1 hingga 1, dengan nilai 1 menunjukkan hubungan linier positif sempurna, nilai -1 menunjukkan hubungan linier negatif sempurna, dan nilai 0 menunjukkan tidak adanya hubungan linier antara variabel.

Dalam analisis korelasi, beberapa skenario yang mungkin terjadi antara dua variabel adalah sebagai berikut:

- 1) Korelasi Positif: Jika koefisien korelasi positif (nilai mendekati 1), maka perubahan dalam satu variabel cenderung berhubungan dengan perubahan yang searah dalam variabel lainnya. Misalnya, jika kenaikan dalam variabel A sering terjadi bersamaan dengan kenaikan dalam variabel B, maka keduanya memiliki korelasi positif.
- 2) Korelasi Negatif: Jika koefisien korelasi negatif (nilai mendekati -1), maka perubahan dalam satu variabel cenderung berhubungan dengan perubahan yang berlawanan arah dalam variabel lainnya. Misalnya, jika peningkatan dalam variabel A cenderung terjadi bersamaan dengan penurunan dalam variabel B, maka keduanya memiliki korelasi negatif.
- 3) Korelasi Nol: Jika koefisien korelasi mendekati 0, maka tidak ada hubungan linier yang jelas antara variabel-variabel tersebut. Artinya, perubahan dalam satu variabel tidak berkaitan dengan perubahan dalam variabel lainnya. Analisis korelasi membantu dalam pemahaman hubungan antar variabel dalam dataset. Hal ini berguna dalam menemukan pola atau tren, mengidentifikasi faktor-faktor yang saling terkait, atau memahami bagaimana variabel-variabel tersebut berinteraksi. Namun, penting untuk diingat bahwa korelasi tidak menyiratkan hubungan sebab-akibat, tetapi hanya menggambarkan hubungan statistik antara variabel-variabel tersebut.

3.5 Build Model

Build model adalah proses pembuatan model prediktif atau analitis yang digunakan untuk melakukan prediksi, klasifikasi, atau analisis pada data. Dalam membangun suatu mode diperlukan pembagian data frame menjadi fitur (*features*) dan label, variabel prediktor medis (variabel independen)

dapat digunakan sebagai fitur, sedangkan variabel target "Tingkat Resiko Kehamilan" (variabel dependen) dapat dijadikan sebagai label. Setelah melakukan *preprocessing* data, dilakukan pembagian data yang sudah diproses. Salah satu metode yang dapat digunakan untuk mengevaluasi performa model *machine learning* adalah train/test split. Metode ini membagi dataset menjadi dua bagian: data latih (training data) dan data uji (testing data) dengan proporsi tertentu. Dalam penelitian ini, proporsi pembagian yang digunakan adalah 80% untuk data latih dan 20% untuk data uji. Random state yang digunakan dalam pembagian data adalah 10326. Dengan menggunakan pembagian tersebut, model dapat dilatih (di-train) menggunakan data latih, dan kemudian diuji (di-test) menggunakan data uji untuk mengevaluasi kinerja model yang telah dibangun. Berikut merupakan *accuracy* dari hasil model yang telah dibuat :

Tabel 1. Hasil Model Akurasi

Algoritma	Train Score	Train Score (%)	Test Score	Test Score (%)
Naïve Bayes	0.7055	70.55%	0.8562	85.62%

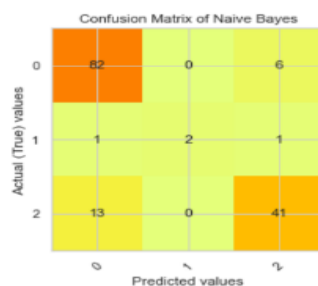
3.6 Evaluasi Model

Evaluasi model merupakan proses untuk mengukur kinerja atau keefektifan model *machine learning* dalam melakukan prediksi atau pengklasifikasian terhadap data yang belum pernah dilihat sebelumnya. Evaluasi model bertujuan untuk menentukan sejauh mana model dapat memahami pola, hubungan, atau tren yang ada dalam data dan melakukan prediksi atau klasifikasi dengan akurasi yang tinggi.

Confusion Matrix adalah tabel yang digunakan untuk menggambarkan kinerja masalah klasifikasi. Ini memvisualisasikan keakuratan classifier dengan membandingkan nilai prediksi dengan nilai sebenarnya. Istilah yang digunakan dalam *Confusion Matrix* adalah *True positive* (TP), *true negative* (TN), *false positive* (FP) dan *false negative* (FN).

- 1) True Positive: Hasil yang diprediksi adalah positif, meskipun diberi label positif.
- 2) False Positive: Hasil yang diprediksi adalah positif, meskipun diberi label negatif. Itu juga menyebut Kesalahan Tipe I.
- 3) False Negative: Hasil yang diprediksi adalah negatif, meskipun diberi label positif. Itu juga menyebut Kesalahan Tipe II.
- 4) True Negative: Hasil prediksi negatif, sementara itu diberi label negatif.

Berikut merupakan hasil evaluasi model algoritma naive bayes menggunakan confusion matrix :



Gambar 6. Confusion Matrix

Berdasarkan confusion matrix diatas dapat dijelaskan sebagai berikut:

- 1) True Positive (TP): Jumlah data yang benar diprediksi sebagai positif (kelas positif) oleh model. Pada tabel ini, TP memiliki nilai 82, 1, dan 41 untuk setiap kolom dari atas ke bawah.
- 2) True Negative (TN): Jumlah data yang benar diprediksi sebagai negatif (kelas negatif) oleh model. Pada tabel ini, TN memiliki nilai 1 dan 41 untuk setiap kolom dari atas ke bawah.

- 3) False Positive (FP): Jumlah data yang salah diprediksi sebagai positif oleh model. Dalam tabel ini, FP memiliki nilai 0 dan 0 untuk setiap kolom.
- 4) False Negative (FN): Jumlah data yang salah diprediksi sebagai negatif oleh model. Dalam tabel ini, FN memiliki nilai 6, 2, dan 13 untuk setiap kolom.

Confusion matrix membantu dalam mengukur akurasi dan performa model klasifikasi dengan menghitung berbagai metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Dengan menggunakan nilai-nilai di dalam confusion matrix, kita dapat menghitung metrik evaluasi tersebut untuk mengevaluasi kinerja model.

	precision	recall	f1-score	support
0	0.8542	0.9318	0.8913	88
1	1.0000	0.5000	0.6667	4
2	0.8542	0.7593	0.8039	54
accuracy			0.8562	146
macro avg	0.9028	0.7304	0.7873	146
weighted avg	0.8582	0.8562	0.8528	146

Gambar 7. *Classification Report*

Berdasarkan *Classification Report* dari Algoritma *Naive Bayes* yang diperoleh, berikut adalah penjelasan dari setiap metrik evaluasi:

1) Precision

Precision (presisi) mengukur sejauh mana model dapat memberikan prediksi yang benar untuk kelas tertentu. *Precision* dihitung dengan membagi *True Positive* (TP) dengan jumlah prediksi positif yang dilakukan oleh model (TP + False Positive (FP)).

- a) Untuk kelas 0, presisi adalah 0.8542, yang berarti 85.42% prediksi yang diklasifikasikan sebagai kelas 0 adalah benar.
- b) Untuk kelas 1, presisi adalah 1.0000, yang berarti 100% prediksi yang diklasifikasikan sebagai kelas 1 adalah benar.
- c) Untuk kelas 2, presisi adalah 0.8542, yang berarti 85.42% prediksi yang diklasifikasikan sebagai kelas 2 adalah benar.

2) Recall

Recall (*recall*) mengukur sejauh mana model dapat mengidentifikasi atau mendeteksi dengan benar kelas tertentu dari data aktual. Recall dihitung dengan membagi *True Positive* (TP) dengan jumlah data aktual dari kelas tertentu (TP + False Negative (FN)).

- a) Untuk kelas 0, recall adalah 0.9318, yang berarti model dapat mengidentifikasi 93.18% data aktual yang merupakan kelas 0.
- b) Untuk kelas 1, recall adalah 0.5000, yang berarti model hanya dapat mengidentifikasi 50% data aktual yang merupakan kelas 1.
- c) Untuk kelas 2, recall adalah 0.7593, yang berarti model dapat mengidentifikasi 75.93% data aktual yang merupakan kelas 2.

3) F1-Score

F1-score adalah rata-rata harmonik dari precision dan recall. F1-score menggabungkan presisi dan recall menjadi satu metrik yang berguna untuk mengevaluasi performa model secara keseluruhan.

- a) Untuk kelas 0, F1-score adalah 0.8913, yang merupakan rata-rata harmonik dari presisi dan recall untuk kelas 0.
- b) Untuk kelas 1, F1-score adalah 0.6667, yang merupakan rata-rata harmonik dari presisi dan recall untuk kelas 1.

c) Untuk kelas 2, F1-score adalah 0.8039, yang merupakan rata-rata harmonik dari presisi dan recall untuk kelas 2.

4) Support

Support adalah jumlah data aktual dalam masing- masing kelas. Selain itu, terdapat metrik evaluasi lainnya pada *Classification Report* seperti *accuracy*, *macro avg*, dan *weighted avg*. *Accuracy* mengukur sejauh mana model dapat memprediksi dengan benar secara keseluruhan. *Macro avg* adalah rata-rata dari metrik evaluasi untuk setiap kelas, sementara *weighted avg* adalah rata-rata yang memberikan bobot berdasarkan jumlah sampel dalam masing- masing kelas. Dengan menggunakan *Classification Report*, kita dapat mengevaluasi kinerja Algoritma *Naive Bayes* dalam melakukan klasifikasi pada dataset yang digunakan.

4. Kesimpulan dan Saran

Berdasarkan hasil evaluasi model, ditemukan bahwa model *Naive Bayes* berhasil mencapai tingkat akurasi sebesar 85.62% dalam memprediksi tingkat risiko kehamilan pada dataset yang telah melalui proses pemrosesan. Analisis lebih lanjut mengungkapkan bahwa model ini memiliki presisi yang tinggi untuk kelas 0 dan 1, namun presisi yang sedikit lebih rendah untuk kelas 2. Hal ini menandakan kemampuan model dalam mengklasifikasikan tingkat risiko kehamilan yang tidak beresiko dan beresiko, namun mungkin menghadapi beberapa kesulitan dalam mengidentifikasi tingkat risiko yang sangat beresiko. Meskipun begitu, perlu diperhatikan bahwa model memiliki recall yang rendah untuk kelas 1, menunjukkan kurangnya kemampuan dalam mengenali dengan tepat data yang sebenarnya termasuk kelas 1. Secara keseluruhan, meskipun model *Naive Bayes* telah menunjukkan kinerja yang cukup baik dalam memprediksi tingkat risiko kehamilan, namun perlu diingat bahwa model ini memiliki keterbatasan dalam mengidentifikasi data aktual yang termasuk kelas 1. Evaluasi mendalam terhadap performa model ini memberikan wawasan yang berharga dalam klasifikasi, yang dapat digunakan sebagai landasan bagi pengembangan atau perbaikan lebih lanjut pada model di masa depan.

Dalam rangka meningkatkan performa dan hasil prediksi yang lebih akurat, terdapat beberapa saran yang dapat diusulkan. Pertama, disarankan untuk mengembangkan model klasifikasi risiko kehamilan dengan mempertimbangkan lebih banyak atribut yang sesuai dengan pedoman kesehatan terkait risiko kehamilan, serta memperluas jumlah dataset yang digunakan. Kedua, alternatif pengelompokan status risiko kehamilan dapat dijajaki dengan menggabungkan metode *Naive Bayes* dengan metode klasifikasi lainnya, atau dengan menerapkan pendekatan yang berbeda secara keseluruhan. Dengan mengimplementasikan saran-saran ini, diharapkan kinerja model dapat ditingkatkan lebih lanjut, sehingga hasil prediksi dapat lebih mendekati kondisi sebenarnya.

5. Daftar Pustaka

- [1] Arifin, A. A. A., Handoko, W., & Efendi, Z. (2022). Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan. *J-Com (Journal of Computer)*, 2(1), 21-26. DOI: <https://doi.org/10.33330/j-com.v2i1.1577>.
- [2] Putry, N. M. (2022). Komparasi algoritma knn dan naïve bayes untuk klasifikasi diagnosis penyakit diabetes mellitus. *Evolusi: Jurnal Sains Dan Manajemen*, 10(1).
- [3] Firdaus, T., & Yanti, F. (2022). IMPLEMENTASI METODE NAÏVE BAYES PADA SISTEM PAKAR DIAGNOSA PENYAKIT GIGI BERBASIS WEB. DOI: <https://doi.org/10.35315/informatika.v14i2.9201>.

- [4] Rizky, R. R., & Hakim, Z. H. (2020). Sistem Pakar Menentukan Penyakit Hipertensi Pada Ibu Hamil Di RSUD Adjudarmo Rangkasbitung Provinsi Banten. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 9(1), 30-34. DOI: <https://doi.org/10.32736/sisfokom.v9i1.781>.
- [5] Apriyani, H., & Kurniati, K. (2020). Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus. *Journal of Information Technology Ampera*, 1(3), 133-143. DOI: 10.51519/journalita.volume1.issue3.year2020.page133-143.
- [6] Zainiyah, Z., Susanti, E., & Setiawati, I. (2021). Deteksi Dini Preeklampsia Pada Ibu Hamil Dengan Imt (Indeks Massa Tubuh), Rot (Roll Over Test) Dan Map (Mean Arteri Pressure). *GEMASSIKA: Jurnal Pengabdian Kepada Masyarakat*, 5(1), 22-30. DOI: <https://doi.org/10.30787/gemassika.v5i1.558>.
- [7] Hidayah, I. (2020). Sistem Monitoring Kondisi Kesehatan Sebelum Dan Sesudah Olahraga Menggunakan Pulse Sensor Dan Sensor Ds18b20 Dengan Metode Naive Bayes. *Publikasi Tugas Akhir S-1 PSTI FT-UNRAM*.
- [8] Putro, H. F., Vulandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi dan Komunikasi (TIKoSIN)*, 8(2). DOI: <http://dx.doi.org/10.30646/tikomsin.v8i2.500>.
- [9] Nurdiana, N., Rodiyansyah, S. F., & Algifari, A. (2020). Studi Komparasi Algoritma ID3 dan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *INFOTECH journal*, 6(2), 18-23.
- [10] Hidayati, N., & Utomo, V. G. (2019). Penerapan Metode Naïve Bayes Dalam Pemeriksaan Kesehatan. *Jurnal Informatika Upgris*, 5(2). DOI: <https://doi.org/10.26877/jiu.v5i2.4313>.
- [11] Mustofa, H., & Mahfudh, A. A. (2019). Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes. *Walisongo Journal of Information Technology*, 1(1), 1-12. DOI: 10.21580/wjit.2019.1.1.3915.
- [12] Setiawati, I., & Lailiyah, S. R. (2020). DETEKSI DINI PREEKLAMSI PADA IBU HAMIL DENGAN PENIMBANGAN BERAT BADAN DAN PENGUKURAN TEKANAN DARAH. *JURNAL PARADIGMA (PEMBERDAYAAN & PENGABDIAN KEPADA MASYARAKAT)*, 2(1), 16-27.
- [13] Ruhjana, N. (2019). Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil/Genap Pada Twitter Dengan Metode Klasifikasi Naive Bayes. *IKRA-ITTH Informatika: Jurnal Komputer dan Informatika*, 3(1), 94-99.
- [14] Fadlan, C., Ningsih, S., & Windarto, A. P. (2018). Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra. *J. Tek. Inform. Musirawas*, 3(1), 1.
- [15] Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm. *Jurnal Data Mining Dan Sistem Informasi*, 2(1), 31-37. DOI: <https://doi.org/10.33365/jdmsi.v2i1.1021>.
- [16] Rachman, R., Handayani, R. N., & Artikel, I. (2021). Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM. *J. Inform*, 8(2), 111-122.



- [17] Jefi, J., Hendri, H., Afni, N., Salim, A., & Maulana, Y. I. (2021). Penerapan Algoritma Naive Bayes Dalam Memprediksi Penyakit Lambung. *JISICOM (Journal of Information System, Informatics and Computing)*, 5(2), 524-531. DOI: <https://doi.org/10.52362/jisicom.v5i2.659>.
- [18] Fadrial, Y. E. (2021). Algoritma Naive Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa. *INTECOMS: Journal of Information Technology and Computer Science*, 4(1), 20-29. DOI: <https://doi.org/10.31539/intecom.v4i1.2219>.
- [19] Wibowo, G. W. N., & Manan, M. A. (2022). Penerapan Algoritma Naive Bayes Untuk Prediksi Heregistrasi Calon Mahasiswa Baru. *JTINFO: Jurnal Teknik Informatika*, 1(1), 1-10. DOI: <https://doi.org/10.02220/jtinfo.v1i1.126>.