

Klasifikasi Diabetes Pada Wanita Menggunakan Metode Naive Bayes Classifier

Mohammad Faisal Fahrul^a, Wiwien Hadikurniawati^b

^aUniversitas Stikubank, Jl. Tri Lomba Juang No.1, Semarang 50241, Indonesia

^bUniversitas Stikubank, Jl. Tri Lomba Juang No.1, Semarang 50241, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 09 Januari 2022

Revisi Akhir: 15 Februari 2022

Diterbitkan Online: 01 Maret 2022

KATA KUNCI

Klasifikasi

Diabetes Wanita

Naive Bayes Classifier

KORESPONDENSI

E-mail: f4isalfahrul@gmail.com

ABSTRACT

The report from Riskesdas shows that there is a 2x increase in diabetes every year in Indonesia. This is due to an increase in factors such as human population, age, obesity, irregular eating patterns and lack of physical activity. The increase in a factor that causes diabetes in Indonesia must be prevented. The first step in preventing diabetes is to detect the risk factors for diabetes that may occur. Influencing factors include behavioral factors and sociodemographic factors. The increase in diabetes in a country is due to late identified factors. The number of factors that are collected in order to detect whether a person has diabetes or not requires a fairly large data processing system. The data used in this study are diabetes data obtained from the Pima Indian Diabetes Database with attributes of pregnant, glucose, diastolic, triceps, insulin, BMI, history of diabetes, age and 300 data output. The Naive Bayes Classifier method can be used to classify diabetes in women based on pregnant, glucose, diastolic, triceps, insulin, BMI, history of diabetes, age and output. The accuracy result of the Naive Bayes Classifier method in classifying diabetes in women is 84% of 300 data which is divided into 2, namely 275 data as training data and 25 data as test data..

1. PENDAHULUAN

Salah satu penyakit yang paling kronis adalah penyakit diabetes. Diabetes berhubungan erat dengan Pankreas, ketika Pankreas menghasilkan hormone alami insulin namun terganggu menyebabkan gula darah meningkat. Peningkatan gula darah di dalam tubuh manusia akan merusak fungsi organ seperti, ginjal, jantung dan otak. Hal ini menyebabkan seseorang menderita diabetes. Penyakit diabetes pada usia dewasa dapat diatasi dengan menjaga pola makan/diet dan berolahraga secara teratur.

Laporan dari WHO menyatakan bahwa diabetes berkontribusi besar atas kematian orang di seluruh dunia yakni 82%. Penyakit diabetes diderita sekitar 450 juta orang menurut data WHO. Pada tahun 2014, 8,5% orang di atas usia 18 tahun di seluruh dunia menderita diabetes. Pada tahun 2016, 1.6 juta orang meninggal karena penyakit diabetes. Tahun 2012 tingginya gula darah menjadi penyebab 2,2 juta meninggal dunia. Di antara tahun 2000 sampai 2016 terdapat kenaikan 5% angka kematian dini akibat penyakit diabetes. Di negara yang maju angka kematian dini akibat penyakit diabetes menurun dari tahun 2000 sampai 2010,

namun di negara yang menengah ke bawah, angka kematian dini akibat penyakit diabetes meningkat dari tahun 2000 sampai 2010. Laporan dari Riset Kesehatan Dasar (Riskesdas) tahun 2020 adanya peningkatan diabetes sebanyak 2x lipat setiap tahunnya di negara Indonesia. Hal ini disebabkan karena bertambahnya suatu faktor seperti populasi manusia, usia, obesitas, pola makan tidak teratur dan kurangnya aktifitas fisik. Bertambahnya suatu faktor yang menyebabkan diabetes di negara Indonesia harus dicegah. Langkah pencegahan diabetes dilakukan dengan deteksi faktor-faktor resiko diabetes. Faktor yang mempengaruhi antara lain faktor perilaku dan faktor sosiodemografi

Peningkatan diabetes di suatu negara disebabkan faktor-faktor yang terlambat diidentifikasi. Banyaknya faktor-faktor yang dikumpulkan guna mendeteksi seseorang apakah menderita diabetes atau tidak membutuhkan sistem olah data yang cukup besar dengan data mining. Data mining merupakan ilmu atau metode dari pengetahuan yang diekstrak atau ditemukannya pola dari data yang besar [1]. Naive Bayes Classifier yaitu algoritma untuk meminimalkan kesalahan dari semua metode klasifikasi yang lain[2].

2. TINJAUAN PUSTAKA

Penelitian oleh [3] menggunakan metode C4.5 untuk melakukan klasifikasi penyakit diabetes pada rumah sakit Unhas. Hasil tingkat akurasi dari algoritma C4.5 menunjukkan rata-rata prediksi sebesar 98,5 %.

Penelitian oleh [4] juga melakukan identifikasi penyakit diabetes dengan metode C4.5. Hasil pohon keputusan menghasilkan 3 aturan.

Penelitian oleh [5] melakukan pebandingan antara metode *Naive Bayes Classifier* dengan KNN dengan hasil metode *Naive Bayes Classifier* lebih baik dari metode KNN.

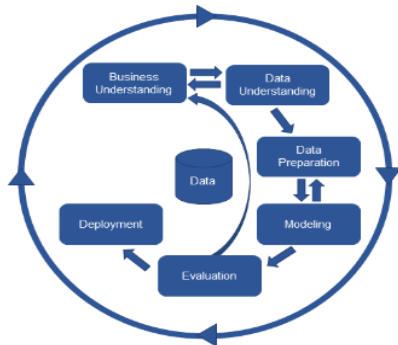
Penelitian oleh [6] melakukan pengujian klasifikasi menggunakan NBC dengan hasil akurasi 96.24%, presisi 93.14 %, dan *recall* 98.96%.

Penelitian oleh [7] berhasil meningkatkan akurasi *Naive Bayes Classifier* dan *forward selection* dari 64.77% ke 78.08%.

Penelitian oleh [8] menyebutkan bahwa algoritma NBC telah menghasilkan akurasi 89.65% dalam klasifikasi data.

3. METODOLOGI

Metode analisis data digunakan metode CRISP-DM (*Cross Industry Standart Process for Data Mining*) seperti gambar 1.



Gambar 1. CRISP-DM

Tahap CRISP-DM dalam penelitian ini yaitu

1. Pemahaman Bisnis (*Bussiness Understanding*)

Tahap ini memahami proses klasifikasi penyakit diabetes pada wanita yang diharapkan mampu melakukan penanganan terbaik terhadap penyakit diabetes. Hal tersebut untuk mengurangi tingkat penyakit diabetes yang dapat mengakibatkan kematian..

2. Pemahaman Data (*Data Understanding*)

Dataset didapatkan dari Pima Indian Diabetes Database. Dataset penyakit diabetes ini berjumlah 300 *record* data. Dalam dataset tersebut terdapat 8 atribut dan 1 label dengan tipe data *integer* dan *binominal*. Atribut dalam dataset penyakit diabetes dijabarkan sebagai tabel 1.

Tabel 1. Atribut Penyakit Diabetes

| No | Atribut | Tipe Data |
|----|------------------|-----------|
| 1. | <i>Pregnant</i> | Integer |
| 2. | <i>Glucose</i> | Integer |
| 3. | <i>Diastolic</i> | Integer |
| 4. | <i>Tricept</i> | Integer |
| 5. | <i>Insulin</i> | Integer |
| 6. | BMI | Integer |
| 7. | Diabetes | Integer |
| 8. | Age | Integer |
| 9. | Hasil | Binominal |

3. Pengolahan Data (*Data Preparation*)

Pada tahap ini dilakukan konversi dataset penyakit diabetes yang berbentuk bilangan integer dan binominal akan di konversi menjadi bentuk permissalan.

Mohammad Faisal Fahrul

4. Pemodelan (*Modelling*)

Tahap ini dilakukan klasifikasi data uji dari penyakit diabetes dengan metode NBC.

5. Evaluasi (*Evaluation*)

Tahap ini model yang terbentuk diuji menggunakan diagram *confusion matrix* untuk mengetahui nilai akurasi dengan menggunakan rumus akurasi = $\frac{TP+TN}{TP+TN+FP+FN} \times 100\%$.

4. HASIL DAN PEMBAHASAN

Penelitian akan menguji tingkat akurasi dari algoritma NBC dalam melakukan klasifikasi penyakit diabetes pada wanita. Hasil akurasi NBC adalah sebesar 84 % dari 300 data yang diperlihatkan seperti gambar 2.

| ID | PREGNANT | GLUCOSE | DIASTOLIC | TRICEPT | INSULIN | BMI | DIABETES | AGE | HASIL | KLASIFIKASI |
|----|------------------|---------|-----------|---------|---------|---------------|---------------|------|-------|-------------|
| 1 | Multigravida | Baik | Rendah | Tinggi | Normal | Sangat Rendah | Sedang | Muda | 0 | 0 |
| 2 | Grande Multipara | Baik | Rendah | Tinggi | Rendah | Gemuk | Rendah | Muda | 0 | 0 |
| 3 | Primipara | Baik | Rendah | Tinggi | Normal | Gemuk | Rendah | Muda | 0 | 0 |
| 4 | Multigravida | Baik | Rendah | Rendah | Normal | Normal | Sedang | Tua | 0 | 0 |
| 5 | Multigravida | Baik | Rendah | Tinggi | Rendah | Gemuk | Tinggi | Muda | 0 | 0 |
| 6 | Primipara | Sedang | Rendah | Rendah | Rendah | Sangat Rendah | Rendah | Muda | 1 | 1 |
| 7 | Grande Multipara | Baik | Rendah | Tinggi | Normal | Sangat Rendah | Rendah | Muda | 0 | 0 |
| 8 | Grande Multipara | Baik | Normal | Normal | Normal | Gemuk | Rendah | Muda | 0 | 0 |
| 9 | Grande Multipara | Sedang | Normal | Rendah | Rendah | Gemuk | Sangat Rendah | Tua | 1 | 1 |
| 10 | Multigravida | Baik | Normal | Rendah | Rendah | Gemuk | Rendah | Tua | 0 | 0 |
| 11 | Grande Multipara | Baik | Rendah | Tinggi | Normal | Gemuk | Sedang | Tua | 0 | 0 |
| 12 | Multigravida | Sedang | Normal | Tinggi | Tinggi | Sangat Rendah | Sedang | Muda | 0 | 0 |
| 13 | Primipara | Baik | Normal | Tinggi | Tinggi | Sangat Rendah | Tinggi | Muda | 1 | 1 |
| 14 | Multigravida | Baik | Rendah | Tinggi | Normal | Normal | Rendah | Muda | 0 | 0 |
| 15 | Multigravida | Baik | Rendah | Tinggi | Normal | Sangat Rendah | Rendah | Muda | 0 | 0 |
| 16 | Primipara | Baik | Normal | Tinggi | Normal | Sangat Rendah | Rendah | Muda | 0 | 0 |

Gambar 2. Hasil Klasifikasi NBC

Diketahui data uji dari klasifikasi penyakit diabetes pada wanita yang diperlihatkan seperti tabel 2.

Tabel 2. Data Uji

| No | Pregnant | Glucose | Diastolic | Tricept | Insulin | BMI | Diabetes | Age | Hasil |
|----|------------------|---------|-----------|---------|---------|---------------|---------------|------|-------|
| 1 | Multigravida | Baik | Rendah | Tinggi | Normal | Sangat Rendah | Sedang | Muda | 0 |
| 2 | Grande Multipara | Baik | Rendah | Tinggi | Rendah | Gemuk | Rendah | Muda | 0 |
| 3 | Primipara | Baik | Rendah | Tinggi | Normal | Gemuk | Rendah | Muda | 0 |
| 4 | Multigravida | Baik | Rendah | Rendah | Rendah | Normal | Sedang | Tua | 0 |
| 5 | Multigravida | Baik | Rendah | Rendah | Tinggi | Gemuk | Tinggi | Muda | 0 |
| 6 | Primipara | Sedang | Rendah | Rendah | Rendah | Sangat Rendah | Rendah | Muda | 1 |
| 7 | Grande Multipara | Baik | Rendah | Tinggi | Normal | Sangat Rendah | Rendah | Muda | 0 |
| 8 | Grande Multipara | Baik | Normal | Normal | Normal | Gemuk | Rendah | Muda | 0 |
| 9 | Grande Multipara | Sedang | Normal | Rendah | Rendah | Gemuk | Sangat Rendah | Tua | 1 |
| 10 | Multigravida | Baik | Normal | Rendah | Rendah | Gemuk | Rendah | Tua | 0 |
| 11 | Grande Multipara | Baik | Rendah | Tinggi | Normal | Gemuk | Sedang | Tua | 0 |
| 12 | Multigravida | Sedang | Normal | Tinggi | Tinggi | Sangat Rendah | Sedang | Muda | 0 |
| 13 | Primipara | Baik | Normal | Tinggi | Tinggi | Sangat Rendah | Tinggi | Muda | 1 |
| 14 | Multigravida | Baik | Rendah | Tinggi | Normal | Normal | Rendah | Muda | 0 |
| 15 | Multigravida | Baik | Rendah | Tinggi | Normal | Sangat Rendah | Rendah | Muda | 0 |
| 16 | Primipara | Baik | Normal | Tinggi | Normal | Sangat Rendah | Rendah | Muda | 0 |

Klasifikasi Diabetes Pada Wanita

| | | | | | | | | | |
|----|------------------|--------|--------|--------|--------|--------|--------|------|---|
| 17 | Primipara | Baik | Rendah | Tinggi | Normal | Sangat | Sedang | Muda | 0 |
| 18 | Multigravida | Baik | Rendah | Tinggi | Tinggi | Sangat | Sangat | Muda | 0 |
| 19 | Primipara | Baik | Rendah | Tinggi | Tinggi | Sangat | Sedang | Muda | 1 |
| 20 | Primipara | Sedang | Rendah | Rendah | Rendah | Normal | Rendah | Tua | 0 |
| 21 | Grande Multipara | Sedang | Rendah | Tinggi | Normal | Sangat | Sedang | Muda | 0 |
| 22 | Multigravida | Sedang | Rendah | Tinggi | Tinggi | Gemuk | Rendah | Muda | 1 |
| 23 | Primipara | Baik | Normal | Tinggi | Tinggi | Gemuk | Rendah | Muda | 0 |
| 24 | Grande Multipara | Baik | Rendah | Tinggi | Tinggi | Sangat | Rendah | Tua | 1 |
| 25 | Grande Multipara | Baik | Rendah | Rendah | Rendah | Normal | Tinggi | Tua | 0 |

Sebelum melakukan perhitungan dengan metode *Naive Bayes Classifier*, tentukan jumlah kasus data latih yang digunakan yaitu

- a. Jumlah Data : 275
- b. Kasus diabetes (1) : 104
- c. Kasus non diabetes (0) : 171

Data uji yang dihitung adalah data nomor 276 dengan kriteria sebagai berikut:

- Tricept : Multigravida
- Glucose : Baik
- Diastolic : Rendah
- Tricept : Tinggi
- Insulin : Normal
- BMI : Sangat Gemuk
- Diabetes : Sedang
- Age : Muda
- Hasil : 0

Proses perhitungan data uji dengan algoritma *Naive Bayes Classifier* sebagai berikut:

$$P(\text{Status} = 1) = \frac{104}{275}$$

$$P(\text{Tricept} = \text{Multigravida} \mid \text{Hasil} = 1) = \frac{64}{104}$$

$$P(\text{Glucose} = \text{Baik} \mid \text{Hasil} = 1) = \frac{59}{104}$$

$$P(\text{Diastolic} = \text{Rendah} \mid \text{Hasil} = 1) = \frac{69}{104}$$

$$P(\text{Tricept} = \text{Tinggi} \mid \text{Hasil} = 1) = \frac{64}{104}$$

$$P(\text{Insulin} = \text{Normal} \mid \text{Hasil} = 1) = \frac{25}{104}$$

$$P(\text{BMI} = \text{Sangat Gemuk} \mid \text{Hasil} = 1) = \frac{45}{104}$$

$$P(\text{Diabetes} = \text{Sedang} \mid \text{Hasil} = 1) = \frac{32}{104}$$

$$P(\text{Age} = \text{Muda} \mid \text{Hasil} = 1) = \frac{81}{104}$$

$$P(X \mid \text{Hasil} = 1) = \frac{104}{275} \times \frac{64}{104} \times \frac{59}{104} \times \frac{69}{104} \times \frac{64}{104} \times \frac{25}{104} \times \frac{45}{104} \times \frac{32}{104} \times \frac{81}{104} = 0,0006718193$$

$$P(\text{Status} = 0) = \frac{171}{275}$$

$$P(\text{Tricept} = \text{Multigravida} \mid \text{Hasil} = 0) = \frac{102}{171}$$

$$P(\text{Glucose} = \text{Baik} \mid \text{Hasil} = 0) = \frac{151}{171}$$

$$P(\text{Diastolic} = \text{Rendah} \mid \text{Hasil} = 0) = \frac{132}{171}$$

$$P(\text{Tricept} = \text{Tinggi} \mid \text{Hasil} = 0) = \frac{102}{171}$$

$$P(\text{Insulin} = \text{Normal} \mid \text{Hasil} = 0) = \frac{61}{171}$$

$$P(\text{BMI} = \text{Sangat Gemuk} \mid \text{Hasil} = 0) = \frac{38}{171}$$

$$P(\text{Diabetes} = \text{Sedang} \mid \text{Hasil} = 0) = \frac{33}{171}$$

$$P(\text{Age} = \text{Muda} \mid \text{Hasil} = 0) = \frac{153}{171}$$

$$P(X \mid \text{Hasil} = 0) = \frac{171}{275} \times \frac{102}{171} \times \frac{151}{171} \times \frac{132}{171} \times \frac{102}{171} \times \frac{61}{171} \times \frac{38}{171} \times \frac{33}{171} \times \frac{153}{171} = 0,0016392659$$

Kesimpulan dari data di atas yaitu jika data uji memiliki kriteria

- Tricept : Multigravida
- Glucose : Baik
- Diastolic : Rendah
- Tricept : Tinggi
- Insulin : Normal
- BMI : Sangat Gemuk
- Diabetes : Sedang
- Age : Muda
- Hasil : 0

Hasil klasifikasi penyakit diabetes pada wanita dengan klasifikasi 0, karena nilai *posterior probability* dari penyakit diabetes pada wanita dengan hasil 0 lebih tinggi dibanding dengan nilai *posterior probability* dari penyakit diabetes pada wanita dengan dengan hasil 1.

Dari perhitungan dengan menggunakan seluruh data uji, selanjutnya dilakukan pengujian menggunakan diagram *confusion matrix* untuk mengetahui nilai akurasi yang diperoleh dari algoritma *Naive Bayes Classifier* yang ditunjukkan seperti tabel 3.

Tabel 3. Hasil *Confusion Matrix* Algoritma NBC

| | True 1 | True 0 |
|---------|--------|--------|
| Pred. 1 | 6 | 4 |
| Pred. 0 | 0 | 15 |

Data uji yang berjumlah 25 menghasilkan hasil klasifikasi sebagai berikut:

1. *True positif* menghasilkan 6 data
2. *False positif* menghasilkan 4 data
3. *False negatif* menghasilkan 0 data
4. *True negatif* menghasilkan 15 data

Hasil dari tabel *confusion matrix* 5.3 didapatkan nilai akurasi sebesar 88 % dengan perhitungan sebagai berikut:

$$\text{Akurasi} = \frac{6+15}{6+4+0+15} \times 100\% = 84\%$$

5. KESIMPULAN DAN SARAN

Metode NBC dapat melakukan klasifikasi diabetes pada wanita berdasarkan pregnant, glucose, diastolic, triceps, insulin, BMI, riwayat diabetes, age dan output. Hasil akurasi NBC dari 300 data dengan 25 data uji adalah sebesar 84 %

DAFTAR PUSTAKA

- [1] F. A. Hermawati, *Data Mining*, Yogyakarta: Andi, 2016.
- [2] Kusrini dan E. T. Luthfi, *Algoritma Data Mining*, Yogyakarta: Andi, 2015.
- [3] D. R. Ente, S. A. Thamrin, H. Kuswanto, S. Arifin dan Andreza, "Klasifikasi Faktor Penyebab Penyakit Diabetes Melitus Di Rumah Sakit UNHAS Menggunakan Algoritma C4.5," *Indonesian Journal of Statistics and Its Applications*, vol. IV, no. 1, pp. 80-88, 2020.
- [4] F. Aris dan Benyamin, "Penerapan Data Mining Untuk Identifikasi Penyakit Diabetes Melitus Dengan Menggunakan Metode Klasifikasi," *Router Research*, vol. I, no. 1, pp. 1-6, 2019.
- [5] R. N. Devita, H. W. Herwanto dan A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. V, pp. 427-434, 2018.
- [6] M. Siddik, R. N. Putri dan Y. Desnelita, "Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi Menggunakan Algoritma Naive Bayes," *Journal of Information Technology and Computer Science*, vol. III, pp. 162-166, 2020.
- [7] Y. P. Astuti, U. Sudibyo, A. W. Kurniawan dan Y. Rahayu, "Algoritma Naive Bayes Dengan Fitur Seleksi Untuk Mengetahui Hubungan Variabel Nilai Dan Latar Belakang Pendidikan," *Simetris*, vol. IX, pp. 597-602, 2018.
- [8] D. W. Nugraha, A. E. Dodu dan N. Chandra, "Klasifikasi Penyakit Stroke Menggunakan Metode Naive Bayes Classifier (Studi Kasus Pada Rumah Sakit Umum Daerah Undata Palu)," *Semantik*, vol. III, pp. 13-22, 2017.

BIODATA PENULIS



Mohammad Faisal Fahrul
Mahasiswa Universitas Stikubank Semarang
Program Studi Teknik Informatika
Email: f4isalfahrul@gmail.com



Wiwien Hadikurniawati, S.T., M.Kom
Dosen Universitas Stikubank Semarang
Program Studi Teknik Informatika
Email: wiwien@edu.unisbank.ac.id