

IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK MENDETEKSI RESIKO TINGGI DIABETES MELITUS PADA IBU HAMIL (STUDI KASUS : PUSKESMAS KABUPATEN MALANG)

Fatmawati^{*1}, Satrio Agung Wicaksono², Satrio Hadi Wijoyo³

^{1, 2, 3}Universitas Brawijaya, Malang

Email: ¹fatmawati@ub.ac.id, ²satrio@ub.ac.id, ³satriohadi@ub.ac.id

^{*}Penulis Korespondensi

(Naskah masuk: 13 Juli 2022, diterima untuk diterbitkan: 26 Juli 2023)

Abstrak

Diabetes pada ibu hamil dapat meningkatkan berbagai risiko, baik maternal maupun neonatus. Terdapat gangguan homeostasis glukosa pada ibu hamil. Terjadinya malformasi kongenital, keguguran, risiko preeklampsia, CPD (Cepalo Pelvic Dispropotion), kelahiran prematur, kelainan letak, plasenta previa dan hipoglikemia neonatus. Oleh karena itu, Perhatian dan penanganan menyeluruh bagi ibu hamil yang mengalami diabetes. Data mining dapat digunakan untuk deteksi risiko tinggi diabetes mellitus pada ibu hamil. Data yang digunakan seperti nama, usia, umur kehamilan, gravida, para, riwayat kehamilan yang lalu, riwayat penyakit yang pernah diderita, faktor risiko, riwayat persalinan yang lalu untuk deteksi risiko tinggi diabetes mellitus pada ibu hamil. Tidak semua kehamilan dapat berjalan dengan normal atau fisiologis pada saat proses persalinannya ada faktor risiko yang dapat mempengaruhinya. Pada penelitian ini dapat mendeteksi risiko yang akan terjadi kepada ibu hamil dan bayi dalam kandungannya. Nilai akurasi tertinggi terdapat pada pengujian ke 4 sebesar 82.4324% dan terendah nilai akurasi pada pengujian ke 2 sebesar 75%. Nilai presisi tertinggi terdapat di uji coba ke 3 sebesar 79.2% dan nilai presisi terendah di uji coba ke 2 sebesar 76.3%. Nilai recall tertinggi terdapat di uji coba ke 4 sebesar 82.4% dan nilai recall terendah di uji coba ke 2 sebesar 75%. Nilai F-Measure tertinggi terdapat di uji coba ke 3 sebesar 79.2% dan nilai F-Measure terendah di uji coba ke 2 sebesar 74.8%.

Kata Kunci : *Algoritma Naïve Bayes, Deteksi Risiko Tinggi Diabetes Mellitus Pada Ibu Hamil*

IMPLEMENTATION OF NAVE BAYES ALGORITHM FOR DETECTING HIGH RISK OF DIABETES MELLITUS IN PREGNANT WOMEN (CASE STUDY: MALANG REGENCY PUSKESMAS)

Abstract

Diabetes in pregnant women can increase various risks, both maternal and neonatal. There is a disturbance of glucose homeostasis in pregnant women. Occurrence of congenital malformations, miscarriage, risk of preeclampsia, CPD (Cepalo Pelvic Disproportion), premature birth, position abnormalities, placenta previa and neonatal hypoglycemia. Therefore, attention and comprehensive treatment for pregnant women with diabetes. Data mining can be used to detect high risk of diabetes mellitus in pregnant women. The data used are name, age, gestational age, gravida, para, past pregnancy history, history of previous illness, risk factors, past delivery history to detect high risk of diabetes mellitus in pregnant women. Not all pregnancies can run normally or physiologically at the time of delivery there are risk factors that can affect it. In this study, it can detect the risks that will occur to pregnant women and their babies in the womb. In addition, recommendations from the system can support a midwife's decision making in taking action to pregnant women. The highest accuracy value is found in the 4th test of 82,4324% and the lowest accuracy value in the 2nd test is 75%. The highest precision value was found in the 3rd trial of 79.2% and the lowest precision value in the 2nd trial of 76.3%. The highest recall value was found in the 4th trial of 82.4% and the lowest recall value in the 2nd trial of 75%. The highest F-Measure value was found in the 3rd trial of 79.2% and the lowest F-Measure value in the 2nd trial of 74.8%.

Keywords: *Naïve Bayes Algorithm, Detecting High Risk of Diabetes Mellitus in Pregnant Women*

1. PENDAHULUAN

Diabetes melitus merupakan kelainan metabolik karena kurangnya insulin atau ketidakmampuan tubuh untuk memanfaatkan insulin. Akibat defisiensi sekresi hormon insulin, aktivitas insulin, atau keduanya menyebabkan *insulin resistance* dengan simtoma berupa hiperglikemia kronis dan gangguan metabolisme. Seperti gangguan metabolisme karbohidrat, lemak, protein, dan sebagainya.

Data Survei Penduduk Antar Sensus (SUPAS) AKI (Angka Kematian Ibu) tahun 2015 sangat tinggi yaitu 305 per 100.000 kelahiran hidup. Padahal target AKI Indonesia pada tahun 2015 adalah 102 per 100.000 kelahiran hidup (Kemenkes RI, 2019). Angka kesakitan dan kematian ibu hamil semakin meningkat dikarenakan tingginya komplikasi dalam kehamilan dan persalinan.

Menurut World Health Organization (WHO), penderita diabetes melitus akan selalu meningkat di Indonesia. Data WHO menunjukkan penderita diabetes melitus sebesar 8,4 juta di Indonesia pada tahun 2000. Penderita diabetes melitus akan selalu meningkat sekitar 21,3 juta pada tahun 2030 (PERKENI, 2015). International Diabetes Federation (IDF) Atlas 2015 sepakat dengan WHO. Jumlah penderita diabetes di Indonesia diprediksi oleh IDF sebesar 10 juta dari tahun 2015 akan mengalami peningkatan menjadi sebesar 16,2 juta pada tahun 2040. Penderita diabetes dalam rentang usia 20-79 tahun. Penderita diabetes di Indonesia diprediksi peringkat ke-6 di dunia pada tahun 2040. Penderita diabetes di Indonesia naik satu peringkat dari urutan ke-7 di dunia dari data IDF pada tahun 2015 (IDF, 2015).

Diabetes pada ibu hamil dapat meningkatkan berbagai risiko, baik maternal maupun neonatus. Terdapat gangguan homeostasis glukosa pada ibu hamil. Terjadinya malformasi kongenital, keguguran, risiko preeklampsia, CPD (Cepalo Pelvik Dispropotion), kelahiran prematur, kelainan letak, plasenta previa dan hipoglikemia neonatus. Oleh karena itu, Perhatian dan penanganan menyeluruh bagi ibu hamil yang mengalami diabetes (Hidayati, R., dkk, 2018).

Jika gejala-gejala penyakit dapat diketahui sejak dini maka pencegahan penyakit diabetes mellitus dapat dilakukan oleh ibu hamil. Masyarakat sering mengalami kesulitan dalam mengetahui gejala-gejala penyakit diabetes mellitus karena kurangnya jumlah pakar dalam penyakit ini. Selain itu, mahalnya biaya serta rumitnya proses mendiagnosa juga menjadi kendala. Para teknisi dan masyarakat perlu dibantu dalam pendeteksian penyakit ini. Teknisi kesehatan dalam ketanggapan pendeteksian penyakit ini dapat dibantu oleh pengetahuan para pakar atau ahli dalam bidang penyakit diabetes melitus. Teknisi kesehatan dengan

bantuan teknologi saat ini tanpa harus mendatangi pakar secara langsung dapat mendeteksinya.

Data mining dapat dimanfaatkan untuk mendeteksi resiko tinggi diabetes mellitus pada ibu hamil. Beberapa penelitian telah melakukan implementasi data mining di bidang kesehatan. Penelitian sebelumnya melakukan pembuatan sebuah sistem pakar berbasis web yang dapat mengatasi nilai derajat kepercayaan atau faktor kepastian data yang diperoleh dari hasil konsultasi dengan pasien melalui metode *certainty factor*. Ujicoba terhadap 8 pasien didapatkan prosentase ketepatan dari sistem adalah 62.5% (berdasarkan fakta dan gejala) dan hasil anaisa sistem 100% (berdasarkan kadar gula darah) (Saputro, B., C., 2011). Optimasi untuk mempercepat konvergensi atau mengurangi jumlah iterasi. Penelitian tersebut menggunakan algoritma *Fuzzy C-Menas* dan Algoritma Genetika. Dataset yang digunakan ciri-ciri mata yang mengalami *diabetic retinopathy* (Rivan, M., E., A., dkk, 2020) Implementasi algoritma CNN (*Convolution Neural Network*) dengan melakukan uji coba pengaruh berbagai macam ruang warna untuk mendapatkan ruang warna yang optimal. Data yang digunakan dalam penelitiannya adalah database diabetic (Dewi, C., dkk, 2021).

Pada penelitian ini menggunakan algoritma Naive Bayes. Algoritma Naive Bayes merupakan algoritma sederhana yang menerapkan teorema Bayes dengan menganggap semua fitur saling tidak berhubungan. Algoritma klasifikasi Naive Bayes memiliki kualitas yang bagus, proses komputasi yang sederhana dan akurasi analisis yang tinggi dari pada algoritma yang lain.

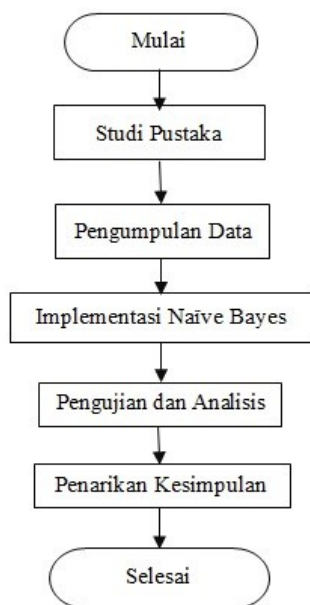
Penelitian sebelumnya terkait penerapan algoritma naïve bayes sudah banyak dilakukan. Pengklasifikasian teks sumber belajar dari website menggunakan algoritma naïve bayes. Klasifikasi menghasilkan Sembilan kelompok mata pelajaran produktif. Nilai akurasi yang diperoleh sebesar 81.48% (Herlambang, A., D., dan Satrio H. W., 2019). Klasifikasi siswa *slow learner* dan *non slow learne* menggunakan algoritma naïve bayes. Hasil pengujian diperoleh nilai akurasi sebesar 92.86%, presisi sebesar 94.74%, recall sebesar 97.29%, dan *F-measure* sebesar 96% (Wicaksono, A., H., dkk, 2022). Klasifikasi gejala awal penyakit *diabete mellitus* pada dataset UCI *Mechine Learning Repository*. Jumlah data sebanyak 520 dataset dengan 17 variabel. Algoritma naïve bayes menghasilkan nilai presisi sebesar 82.35% dan recall sebesar 87.50% (Ridwan, A. 2020). Algoritma naïve bayes digunakan untuk memprediksi penyakit *diabetes mellitus* berdasarkan gejala-gejala yang terjadi. Data *diabetes mellitus* yang digunakan pasien pada kasus positif dan negatif. Atribut yang digunakan umur, jenis kelamin, polyuria, polydipsia, berat badan turun, keletihan, polyphagia, iritasi

genital, pandangan kabur, gatal, mudah marah, penyembuhan lambat, partial paresis, otot kaku, rambut rontok, dan obesitas (Rumini, 2021).

Berdasarkan uraian permasalahan, penelitian ini bertujuan melakukan implementasi algoritma naïve bayes untuk mendeteksi resiko diabetes militus pada ibu hamil. Data skrining pada ibu hamil yang digunakan seperti nama, usia, umur kehamilan, gravida, para, riwayat kehamilan yang lalu, riwayat penyakit yang pernah diderita, faktor risiko, dan riwayat persalinan.

2. METODE PENELITIAN

Metodologi penelitian yang digunakan terkait implementasi algoritma naïve bayes dalam mendeteksi resiko tinggi diabetes mellitus pada ibu hamil. Gambar 1 menjelaskan tahapan pada bab metodologi penelitian ini.



Gambar 1. Tahapan Penelitian

Studi literatur untuk mempelajari konsep terjadinya resiko tinggi diabetes melitus pada ibu hamil dan faktor-faktornya. Konsep algoritme naïve bayes untuk deteksi juga dipelajari. Pengumpulan data dilakukan dengan dua cara yaitu wawancara dan skrining pada ibu hamil. Atribut data yang digunakan seperti nama, usia, umur kehamilan, gravida, para, riwayat kehamilan yang lalu, riwayat penyakit yang pernah diderita, faktor risiko, dan riwayat persalinan. Data primer dan sekunder diperoleh mulai bulan januari sampai dengan bulan agustus 2021 di puskesmas Kabupaten Malang. Jumlah data yang diperoleh sebanyak 200 pasien ibu hamil.

Implementasi naïve bayes untuk mendeteksi diabetes melitus pada ibu hamil menggunakan bahasa pemrograman Weka. Pengujian dilakukan dengan beberapa skenario untuk mengetahui seberapa baik hasil optimasi. Analisis dilakukan

untuk menentukan seberapa baik algoritma naïve bayes dalam bekerja dengan cara menentukan nilai akurasi paling baik sebagai representasi solusi

3. HASIL DAN PEMBAHASAN

Data yang akan digunakan dalam penelitian diperoleh dari database Puskesmas Lawang Kabupaten melalui bidan. Data yang akan digunakan berupa riwayat pemeriksaan pasien atau ibu hamil. Data yang didapatkan berjumlah 248 data yang terdiri dari Tanggal Pengkajian, Nama Ibu Hamil, Tanggal HPHT, Tanggal HPL, Umur, GPA, Usia Kehamilan, KSPR, Keterangan KSPR, IMT, Kategori IMT, Reduksi, Albumin, Hepatitis, HIV, dan IMS.

Gambar 2 merupakan contoh data yang diperoleh dari Puskemas Lawang Kabupaten Malang. Data yang diperoleh mulai bulan Januari 2021 sampai dengan Agustus 2021.

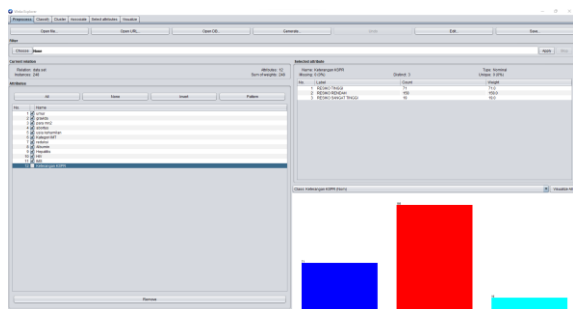
| No | Tanggal Pengkajian | Nama Ibu Hamil | Umur | GPA | Usia Kehamilan | KSPR | Keterangan KSPR | IMT | Kategori IMT | Reduksi | Albumin | Hepatitis | HIV | IMS |
|----|--------------------|----------------|------|--------|----------------|------|-----------------|------|--------------|---------|---------|-----------|-----|-----|
| 1 | 2021-01-15 | Fitriyanti | 25 | 150/90 | 32 | 100 | Normal | 24.5 | Tinggi | 0 | 0 | 0 | 0 | 0 |
| 2 | 2021-01-20 | Fitriyanti | 25 | 140/80 | 33 | 100 | Normal | 23.5 | Tinggi | 0 | 0 | 0 | 0 | 0 |
| 3 | 2021-01-25 | Fitriyanti | 25 | 130/70 | 34 | 100 | Normal | 22.5 | Tinggi | 0 | 0 | 0 | 0 | 0 |
| 4 | 2021-01-30 | Fitriyanti | 25 | 120/60 | 35 | 100 | Normal | 21.5 | Tinggi | 0 | 0 | 0 | 0 | 0 |
| 5 | 2021-02-05 | Fitriyanti | 25 | 110/50 | 36 | 100 | Normal | 20.5 | Tinggi | 0 | 0 | 0 | 0 | 0 |

Gambar 2. Contoh Dataset yang digunakan

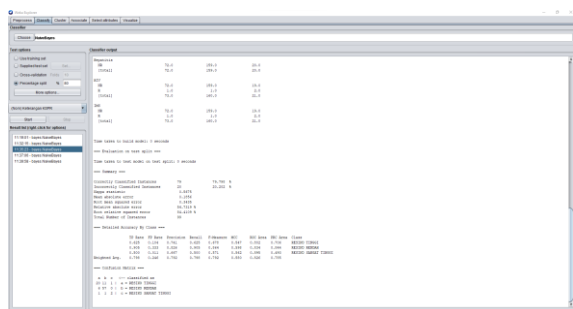
Proses seleksi data dilakukan dengan melakukan wawancara dr. Julia Rosana sebagai Kepala Puskesmas dan Lilik Agustin, S.ST., Bd sebagai Bidan koordinator untuk mengetahui atribut apa saja yang sekiranya memiliki pengaruh terhadap Deteksi Resiko Tinggi Diabetes Mellitus Pada Ibu Hamil. Pada penelitian ini, atribut yang akan digunakan berasal dari data ibu hamil yang melakukan pemeriksaan pada puskesmas Lawang.

Pada proses ini, data yang tidak relevan atau data yang nilainya kosong akan dihilangkan. Data yang akan digunakan sebagai data training dan data testing adalah data Data yang akan digunakan berupa riwayat pemeriksaan pasien atau ibu hamil. *Learning process* dalam penelitian dilakukan pada program menggunakan data *training*. Hasil dari *learning process* mendapatkan model yang digunakan untuk Deteksi Resiko Tinggi Diabetes Mellitus Pada Ibu Hamil. Selain dataset digunakan untuk data *training*, dataset digunakan untuk data *testing* juga. Data *testing* berguna untuk proses Deteksi. Implementasi algoritma menggunakan Weka terlihat pada Gambar 3 dan 4 berikut.

Evaluasi algoritma naïve bayes dilakukan dengan memanfaatkan *confusion matrix* dan kurva ROC yang sudah ada pada tool weka. *Test option* atau model pengujian dilakukan menggunakan beberapa skenario pengujian yang berbeda-beda.



Gambar 3. Visualisasi dari Semua Variabel



Gambar 4. Tampilan Algoritme Naïve Bayes pada Weka

Data yang sudah memiliki kelas untuk diuji dengan proses prediksi apakah sesuai dengan kelas sebenarnya.

1. Skenario Pengujian 1

Skenario pengujian 1 adalah 248 data dibagi menjadi 40% sebagai data training dan 60% sebagai data testing. Evaluasi algoritmanya menggunakan nilai akurasi, precision, recall, dan f-measure. Hasil akurasi dari skenario pengujian 1 sebesar 78.5235%. Nilai tertinggi untuk Precision pada kelas Resiko Rendah sebesar 84,5% sedangkan nilai tertinggi Recall juga pada kelas Resiko Rendah sebesar 89.7%. Nilai rata-rata precision pada pengujian 1 sebesar 77,9%, nilai rata-rata recall sebesar 78.5%, dan nilai rata-rata F-Measure sebesar 78.1%. Hasil evaluasi algoritma pada pengujian 1 dapat dilihat pada Tabel 1.

Tabel 1. Hasil Evaluasi Algoritma Pada Pengujian 1

| Class | Precision | Recall | F-Measure |
|----------------------|-----------|--------|-----------|
| RESIKO TINGGI | 0.692 | 0.6 | 0.643 |
| RESIKO RENDAH | 0.845 | 0.897 | 0.87 |
| RESIKO SANGAT TINGGI | 0.429 | 0.429 | 0.429 |
| Rata-rata | 0.779 | 0.785 | 0.781 |

Tabel 2 menunjukkan hasil confusion matrix pada pengujian 1. Data testing 60% dari 248 adalah 149 data. Terdapat 27 data yang hasil prediksinya tepat pada kelas resiko tinggi, 87 data yang hasil prediksinya tepat pada kelas resiko rendah, dan 3 data yang hasil prediksinya tepat pada kelas resiko sangat tinggi. Sehingga total data yang hasil prediksinya benar sebanyak 117 data. Total data yang hasil prediksinya salah sebanyak 32 data.

Tabel 2. Hasil Confusion Matrix Pada Pengujian 1

| a | b | c | classified as |
|----|----|---|--------------------------|
| 27 | 15 | 3 | a = RESIKO TINGGI |
| 9 | 87 | 1 | b = RESIKO RENDAH |
| 3 | 1 | 3 | c = RESIKO SANGAT TINGGI |

2. Skenario Pengujian 2

Skenario pengujian 2 adalah 248 data dibagi menjadi 50% sebagai data training dan 50% sebagai data testing. Evaluasi algoritmanya menggunakan nilai akurasi, precision, recall, dan f-measure. Hasil akurasi dari skenario pengujian 2 sebesar 75%. Nilai tertinggi untuk Precision pada kelas Resiko Rendah sebesar 81% sedangkan nilai tertinggi Recall juga pada kelas Resiko Rendah sebesar 88.3%. Nilai rata-rata precision pada pengujian 2 sebesar 76,3%, nilai rata-rata recall sebesar 75%, dan nilai rata-rata F-Measure sebesar 74.8%. Hasil evaluasi algoritma pada pengujian 2 dapat dilihat pada Tabel 3.

Tabel 3. Hasil Evaluasi Algoritma Pada Pengujian 2

| Class | Precision | Recall | F-Measure |
|----------------------|-----------|--------|-----------|
| RESIKO TINGGI | 0.733 | 0.524 | 0.611 |
| RESIKO RENDAH | 0.81 | 0.883 | 0.845 |
| RESIKO SANGAT TINGGI | 0.3 | 0.6 | 0.4 |
| Rata-rata | 0.763 | 0.75 | 0.748 |

Tabel 4 menunjukkan hasil confusion matrix pada pengujian 2. Data testing 50% dari 248 adalah 124 data. Terdapat 22 data yang hasil prediksinya tepat pada kelas resiko tinggi, 68 data yang hasil prediksinya tepat pada kelas resiko rendah, dan 3 data yang hasil prediksinya tepat pada kelas resiko sangat tinggi. Sehingga total data yang hasil prediksinya benar sebanyak 93 data. Total data yang hasil prediksinya salah sebanyak 31 data.

Tabel 4. Hasil Confusion Matrix Pada Pengujian 2

| a | b | c | classified as |
|----|----|---|--------------------------|
| 22 | 15 | 5 | a = RESIKO TINGGI |
| 7 | 68 | 2 | b = RESIKO RENDAH |
| 1 | 1 | 3 | c = RESIKO SANGAT TINGGI |

3. Skenario Pengujian 3

Skenario pengujian 3 adalah 248 data dibagi menjadi 60% sebagai data training dan 40% sebagai data testing. Evaluasi algoritmanya menggunakan nilai akurasi, precision, recall, dan f-measure. Hasil akurasi dari skenario pengujian 3 sebesar 79.798%. Nilai tertinggi untuk Precision pada kelas Resiko Rendah sebesar 82.6% sedangkan nilai tertinggi Recall juga pada kelas Resiko Rendah sebesar 90.5%. Nilai rata-rata precision pada pengujian 3 sebesar 79,2%, nilai rata-rata recall sebesar 79.8%, dan nilai rata-rata F-Measure sebesar 79.2%. Hasil evaluasi algoritma pada pengujian 3 dapat dilihat pada Tabel 5.

Tabel 5. Hasil Evaluasi Algoritma Pada Pengujian 3

| Class | Precision | Recall | F-Measure |
|----------------------|-----------|--------|-----------|
| RESIKO TINGGI | 0.741 | 0.625 | 0.678 |
| RESIKO RENDAH | 0.826 | 0.905 | 0.864 |
| RESIKO SANGAT TINGGI | 0.667 | 0.5 | 0.571 |
| Rata-rata | 0.792 | 0.798 | 0.792 |

Tabel 6 menunjukkan hasil confusion matrix pada pengujian 3. Data testing 40% dari 248 adalah 99 data. Terdapat 20 data yang hasil prediksinya tepat pada kelas resiko tinggi, 57 data yang hasil prediksinya tepat pada kelas resiko rendah, dan 2 data yang hasil prediksinya tepat pada kelas resiko sangat tinggi. Sehingga total data yang hasil prediksinya benar sebanyak 79 data. Total data yang hasil prediksinya salah sebanyak 20 data.

Tabel 6. Hasil Confusion Matrix Pada Pengujian 3

| a | b | c | classified as |
|----|----|---|--------------------------|
| 20 | 11 | 1 | a = RESIKO TINGGI |
| 6 | 57 | 0 | b = RESIKO RENDAH |
| 1 | 1 | 2 | c = RESIKO SANGAT TINGGI |

4. Skenario Pengujian 4

Skenario pengujian 4 adalah 248 data dibagi menjadi 70% sebagai data training dan 30% sebagai data testing. Evaluasi algoritmanya menggunakan nilai akurasi, precision, recall, dan f-measure. Hasil akurasi dari skenario pengujian 4 sebesar 82.4324%. Nilai tertinggi untuk Precision pada kelas Resiko Rendah sebesar 85.7% sedangkan nilai tertinggi Recall juga pada kelas Resiko Rendah sebesar 82.4%. Nilai rata-rata recall pada pengujian 4 sebesar 82.4%. Hasil evaluasi algoritma pada pengujian 4 dapat dilihat pada Tabel 7.

Tabel 7. Hasil Evaluasi Algoritma Pada Pengujian 4

| Class | Precision | Recall | F-Measure |
|----------------------|-----------|--------|-----------|
| RESIKO TINGGI | 0.722 | 0.65 | 0.684 |
| RESIKO RENDAH | 0.857 | 0.923 | 0.889 |
| RESIKO SANGAT TINGGI | ? | 0 | ? |
| Rata-rata | ? | 0.824 | ? |

Tabel 8 menunjukkan hasil confusion matrix pada pengujian 4. Data testing 30% dari 248 adalah 74 data. Terdapat 13 data yang hasil prediksinya tepat pada kelas resiko tinggi, 48 data yang hasil prediksinya tepat pada kelas resiko rendah, dan 0 data yang hasil prediksinya tepat pada kelas resiko sangat tinggi. Sehingga total data yang hasil prediksinya benar sebanyak 61 data. Total data yang hasil prediksinya salah sebanyak 13 data.

Tabel 8. Hasil Confusion Matrix Pada Pengujian 4

| a | b | c | classified as |
|----|----|---|--------------------------|
| 13 | 7 | 0 | a = RESIKO TINGGI |
| 4 | 48 | 0 | b = RESIKO RENDAH |
| 1 | 1 | 0 | c = RESIKO SANGAT TINGGI |

5. Skenario Pengujian 5

Skenario pengujian 5 adalah 248 data dibagi menjadi 80% sebagai data training dan 20% sebagai data testing. Evaluasi algoritmanya menggunakan nilai akurasi, precision, recall, dan f-measure. Hasil akurasi dari skenario pengujian 5 sebesar 80%. Nilai tertinggi untuk Precision pada kelas Resiko Rendah sebesar 86.8% sedangkan nilai tertinggi Recall juga pada kelas Resiko Rendah sebesar 89.2%. Nilai rata-rata recall pada pengujian 5 sebesar 80%. Hasil

evaluasi algoritma pada pengujian 5 dapat dilihat pada Tabel 9.

Tabel 9. Hasil Evaluasi Algoritma Pada Pengujian 5

| Class | Precision | Recall | F-Measure |
|----------------------|-----------|--------|-----------|
| RESIKO TINGGI | 0.583 | 0.583 | 0.583 |
| RESIKO RENDAH | 0.868 | 0.892 | 0.88 |
| RESIKO SANGAT TINGGI | ? | 0 | ? |
| Rata-rata | ? | 0.8 | ? |

Tabel 10 menunjukkan hasil confusion matrix pada pengujian 5. Data testing 20% dari 248 adalah 50 data. Terdapat 7 data yang hasil prediksinya tepat pada kelas resiko tinggi, 33 data yang hasil prediksinya tepat pada kelas resiko rendah, dan 0 data yang hasil prediksinya tepat pada kelas resiko sangat tinggi. Sehingga total data yang hasil prediksinya benar sebanyak 40 data. Total data yang hasil prediksinya salah sebanyak 10 data.

Tabel 10. Hasil Confusion Matrix Pada Pengujian 5

| a | b | c | classified as |
|---|----|---|--------------------------|
| 7 | 5 | 0 | a = RESIKO TINGGI |
| 4 | 33 | 0 | b = RESIKO RENDAH |
| 1 | 0 | 0 | c = RESIKO SANGAT TINGGI |

6. Analisa Perbandingan Hasil Pengujian

Analisa perbandingan hasil dari kelima pengujian untuk membandingkan nilai akurasi, precision, recall, dan F-Measure. Hasil perbandingan nilai akurasi untuk setiap percobaan dapat lihat pada Tabel 11.

Tabel 11. Hasil Perbandingan Akurasi Pengujian

| Pengujian ke-i | Akurasi | Precision | Recall | F-Measure |
|----------------|---------|-----------|--------|-----------|
| 1 | 78.5235 | 0.779 | 0.785 | 0.781 |
| 2 | 75 | 0.763 | 0.75 | 0.748 |
| 3 | 79.798 | 0.792 | 0.798 | 0.792 |
| 4 | 82.4324 | ? | 0.824 | ? |
| 5 | 80 | ? | 0.8 | ? |

Nilai akurasi tertinggi terdapat pada pengujian ke 4 sebesar 82.4324% dan terendah nilai akurasi pada pengujian ke 2 sebesar 75%. Nilai precision tertinggi terdapat pada pengujian ke 3 sebesar 79.2% dan terendah nilai precision pada pengujian ke 2 sebesar 76.3%. Nilai recall tertinggi terdapat pada pengujian ke 4 sebesar 82.4% dan terendah nilai recall pada pengujian ke 2 sebesar 75%. Nilai F-Measure tertinggi terdapat pada pengujian ke 3 sebesar 79.2% dan terendah nilai F-Measure pada pengujian ke 2 sebesar 74.8%.

4. KESIMPULAN DAN SARAN

Kesimpulan yang didapatkan berdasarkan hasil penelitian yang telah dilakukan adalah Nilai akurasi tertinggi terdapat pada pengujian ke 4 sebesar 82.4324% dan terendah nilai akurasi pada pengujian ke 2 sebesar 75%. Nilai precision tertinggi terdapat pada pengujian ke 3 sebesar 79.2% dan terendah nilai precision pada pengujian ke 2 sebesar 76.3%. Nilai recall tertinggi terdapat pada pengujian ke 4 sebesar 82.4% dan terendah nilai recall pada

pengujian ke 2 sebesar 75%. Nilai F-Measure tertinggi terdapat pada pengujian ke 3 sebesar 79.2% dan terendah nilai F-Measure pada pengujian ke 2 sebesar 74.8%.

Pada penelitian hanya sebatas implementasi algoritma naïve bayes untuk data diabetes mellitus pada ibu hamil. Saran penelitian ini dapat dilanjutkan pengembangan sistem prediksi resiko tinggi diabetes mellitus pada ibu hamil. Sistem dari sisi fungsionalitas agar sistem mampu untuk melakukan transformasi data, karena pada penelitian ini transformasi data masih dilakukan secara manual melalui microsoft excel. Pengembangan system prediksi lanjutan dapat berbasis website.

DAFTAR PUSTAKA

- DEWI, C., ANDRI S., INDRIATI, NADIA A. D., YOKE K. A., 2021. Evaluasi Performasi Ruang Warna Pada Klasifikasi Diabetic Retinopathy Menggunakan Convolution Neural Network. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*. Vol. 8, No. 3, Juni 2021, hlm. 619-624
- HERLAMBANG, A., D., DAN SATRIO H. W., 2019. Algoritma Naïve Bayes untuk Klasifikasi Sumber Belajar Berbasis Teks Pada Mata Pelajaran Produktif di SMK Rumpun Teknologi Informasi dan Komunikasi. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*. Vol. 6, No. 4, Agustus 2019, hlm. 431-436
- HIDAYATI, R., SETYORINI, D., NUARI, N.A. 2018. Differences Complications During Perinatal in Hiatory of Woman With Diabetes Mellitus And Obesity Gestational. 148-160
- IDF. 2015. *Diabetes Atlas (Seventh Edition)*. International Diabetes federation
- KEMENKES RI. 2019. *Profil Kesehatan Indonesia tahun 2018*
- PERKENI. 2015. *Konsensus Pengelolaan dan Pencegahan Diabetes Mellitus tipe 2 di Indonesia*. Jakarta. PB PERKENI
- RIDWAN, A. 2020. Penerapan Algoritma Naïve Bayes untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal Sistem Komputer dan Kecerdasan Buatan*. Vol. 4, No. 1, September 2020.
- RIVAN, M., E., A., STEVEN, DAN WILLIAM T., 2020. Optimasi Fuzzy C-Means dan K-Means Menggunakan Algoritma Genetika untuk Pengklasteran Dataset Diabetic Retinopathy. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*. Vol. 7, No. 5, Oktober 2022, hlm. 993-1000
- RUMINI, DAN AHMAD N., 2021. Prediksi Awal Penyakit Diabetes Mellitus Menggunakan Algoritma Naïve Bayes. *Jurnal ICT : Information Communication & Technology*. Vol. 20, No. 2, Desember 2021. pp. 246-253
- SAPUTRO, B., C., ROSA DELIMA, JOKO PURWADI. 2011. *Sistem Diagnosa Penyakit Diabetes Melitus Menggunakan Metode Certainty Factor*.
- WICAKSONO, A., H., AHMAD A. S., DAN SATRIO H. W., 2022. Klasifikasi Siswa Slow Learner untuk Mendukung Sekolah dalam Meningkatkan Pemahaman Siswa Menggunakan Algoritma Naïve Bayes. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*. Vol. 9, No. 3, Juni 2022, hlm. 589-596