

## IMPLEMENTASI *MACHINE LEARNING* DALAM DETEKSI RISIKO TINGGI DIABETES MELITUS PADA KEHAMILAN

Aryo Pinandito<sup>\*1</sup>, Satrio Agung Wicaksono<sup>2</sup>, Satrio Hadi Wijoyo<sup>3</sup>

<sup>1, 2, 3</sup>Universitas Brawijaya, Malang  
Email: <sup>1</sup>aryo@ub.ac.id, <sup>2</sup>satrio@ub.ac.id, <sup>3</sup>satriohadi@ub.ac.id  
<sup>\*</sup>Penulis Korespondensi

(Naskah masuk: 29 Desember 2022, diterima untuk diterbitkan: 25 Juli 2023)

### Abstrak

Diabetes dalam kehamilan dapat meningkatkan berbagai risiko, baik pada maternal maupun neonatus. Adanya gangguan homeostasis glukosa dalam kehamilan dapat meningkatkan terjadinya malformasi kongenital, keguguran, risiko preeklampsia, Cepalo Pelvik Dispropotion (CPD), kelahiran prematur, kelainan letak, plasenta previa dan hipoglikemia neonatus. Oleh karena itu, diperlukan perhatian dan penanganan menyeluruh bagi ibu hamil yang menderita diabetes. Teknologi *machine learning* dapat dimanfaatkan dalam berbagai hal di bidang kesehatan. Salah satu dari pemanfaatan *machine learning* di bidang kesehatan adalah kemampuannya untuk mendeteksi risiko tinggi diabetes mellitus pada ibu hamil melalui berbagai data dan informasi seperti nama, usia, umur kehamilan, gravida, para, riwayat kehamilan, riwayat penyakit yang pernah diderita, faktor risiko, dan riwayat persalinan yang lalu. Proses persalinan ibu hamil dipengaruhi oleh faktor fisiologis dan faktor-faktor risiko lain yang dapat mempengaruhinya. Penelitian ini menggunakan *dataset* yang diperoleh dari data pemeriksaan ibu hamil di Puskesmas Lawang dari bulan Januari 2021 sampai dengan Agustus 2021 dan menggunakan 16 atribut penting, yaitu: tanggal pengkajian, nama, tanggal HPHT, tanggal HPL, umur, GPA, usia kehamilan, KSPR, keterangan KSPR, IMT, kategori IMT, reduksi, albumin, hepatitis, HIV, dan IMS. Penelitian ini telah melakukan tahapan pengumpulan data, perancangan, implementasi, pengujian dan analisis data yang mengimplementasikan teknik *machine learning* K-Nearest Neighbor (KNN). Nilai akurasi tertinggi pada skenario pengujian pertama dengan atribut  $k=2$  adalah 70.27%. Nilai akurasi tertinggi pada skenario pengujian kedua dengan  $k=3$  adalah 75.68%. Nilai akurasi tertinggi pada skenario pengujian ketiga dengan  $k=4$  adalah 78.38%. Sedangkan Nilai akurasi tertinggi pada skenario pengujian keempat dengan  $k=5$  adalah 77.03%. Nilai akurasi tertinggi dicapai pada rasio sebesar 7:3 antara data latih dengan data testing.

**Kata kunci:** akurasi, diabetes melitus, *machine learning*, kehamilan, KNN

## MACHINE LEARNING IMPLEMENTATION IN DETECTING HIGH RISK OF DIABETES MELLITUS IN PREGNANCY

### Abstract

*Diabetes in pregnancy can increase various risks, both maternal and neonatal. Disorders of glucose homeostasis in pregnancy can increase the occurrence of congenital malformations, miscarriage, risk of preeclampsia, Cepalo Pelvic Dispropotion (CPD), premature birth, malformations, placenta previa and neonatal hypoglycemia. Thorough attention and treatment is needed for pregnancy with diabetes. Machine learning technology can be used to detect the risk of diabetes mellitus in pregnancy. Several data such as name, age, gestational age, gravida, para, past pregnancy history, past medical history, risk factors, and past birth history were used in the risk detection. Delivery process in pregnancy is affected by the physiologic of prospective mother and several other risk factors. The dataset used in this study was 248 examination data of pregnancy check up at the Lawang Health Center from January 2021 to August 2021. The study used 16 attributes in determining the risk, i.e., date of assessment, name of mother, date of HPHT, date of HPL, age, GPA, gestational age, KSPR, KSPR information, BMI, BMI category, reduction, albumin, hepatitis, HIV, and STIs. This study has follow the implementation stages of data collection, design, implementation, testing and data analysis. The highest accuracy for the first test scenario with  $k=2$  is 70.27%. The highest accuracy for the second test scenario with  $k=3$  is 75.68%. The highest accuracy for the third test scenario with  $k=4$  is 78.38%. While the highest accuracy for the fourth test scenario with  $k=5$  is 77.03%. Highest accuracies were achieved with the distribution ratio of 7:3 between training and testing data.*

**Keywords:** accuracy, diabetes mellitus, KNN, *machine learning*, pregnancy

## 1. PENDAHULUAN

Diabetes melitus adalah kelainan metabolik yang disebabkan oleh beberapa faktor penentu seperti kurangnya insulin atau ketidakmampuan tubuh untuk memanfaatkan insulin (*insulin resistance*), dengan gejala berupa hiperglikemia kronis dan gangguan metabolisme karbohidrat, lemak dan protein, sebagai akibat dari defisiensi sekresi hormon insulin, aktivitas insulin, atau keduanya.

Tingginya komplikasi dalam kehamilan dan persalinan dapat meningkatkan angka kesakitan dan kematian ibu hamil. Sampai saat ini Angka kematian ibu (AKI) sangat tinggi target AKI Indonesia pada tahun 2015 adalah 102 per 100.000 kelahiran hidup berdasarkan data Survei Penduduk Antar Sensus (SUPAS) AKI tahun 2015 yaitu 305 per 100.000 kelahiran hidup (Kemenkes, 2019).

Penderita Diabetes Melitus di Indonesia menurut World Health Organization (WHO), akan selalu mengalami peningkatan. Hal ini dapat dilihat dari data yang dimiliki WHO bahwa penderita Diabetes Melitus di Indonesia pada tahun 2000 yaitu 8,4 juta dan akan mengalami peningkatan pada tahun 2030 yaitu sekitar 21,3 juta (Perkeni, 2021). International Diabetes Federation (IDF) Atlas 2015 sepakat dengan WHO bahwa IDF juga memprediksi jumlah penderita diabetes di Indonesia dari tahun 2015 sebesar 10 juta akan menjadi 16,2 juta pada tahun 2040 dengan rentang usia 20-79 tahun. Berdasarkan data tersebut, Indonesia menempati urutan ke-6 di dunia pada tahun 2040, dibandingkan dengan data IDF pada tahun 2015, Indonesia naik satu peringkat dari urutan ke-7 di dunia (International Diabetes Federation, 2015).

Diabetes dalam kehamilan dapat meningkatkan berbagai risiko, baik maternal maupun neonatus. Adanya gangguan homeostasis glukosa dalam kehamilan dapat meningkatkan terjadinya malformasi kongenital, keguguran, risiko preeklampsia, Cepalo Pelvik Dispropotion (CPD), kelahiran prematur, kelainan letak, plasenta previa dan hipoglikemia neonatus. Oleh karena itu, diperlukan perhatian dan penanganan menyeluruh bagi ibu hamil yang menderita diabetes melitus (Hidayati, Setyorini, & Nuari, 2018).

Pemantauan gula darah rutin, aktivitas fisik, dan terapi nutrisi menjadi manajemen non-farmakologi yang penting bagi ibu hamil dengan diabetes. Metformin dapat digunakan sebagai penanganan farmakologis untuk diabetes dalam kehamilan, tetapi jika gula darah tidak dapat terkontrol, pilihan terbaik adalah penggunaan insulin. Pemberian aspirin dosis rendah juga direkomendasikan bagi ibu hamil dengan diabetes untuk mengurangi risiko preeklampsia. Oleh karena itu pelayanan *antenatal care* harus dilakukan secara rutin sesuai standar terutama pada ibu hamil yang memiliki riwayat penyakit diabetes melitus untuk dilakukan skrining dan diagnosis yang tepat (Wojtyla, Stanierowski, Ciebiera, & Wojtyla, 2021).

Pencegahan penyakit diabetes melitus dapat dilakukan jika gejala-gejala penyakit dapat diketahui

sejak dini. Dalam masyarakat umumnya, orang mengalami kesulitan untuk mengetahui gejala-gejala penyakit diabetes melitus. Hal ini dikarenakan kurangnya jumlah pakar dalam penyakit ini. Selain itu, mahalnya biaya serta rumitnya proses diagnosis juga menjadi kendala. Oleh sebab itu, para tenaga medis dan masyarakat umum, khususnya ibu hamil perlu dibantu dalam mendeteksi risiko kehamilan atas penyakit ini sehingga risiko kegagalan dalam proses persalinan dapat dicegah dan dihindari sejak dini. Pengetahuan yang dimiliki oleh para pakar atau ahli dalam bidang penyakit diabetes melitus dapat membantu tenaga kesehatan untuk tanggap dalam mendeteksi penyakit ini. Pengetahuan yang dimiliki oleh pakar ini dapat diperoleh dari teknisi kesehatan dengan bantuan teknologi saat ini tanpa harus mendatangi pakar secara langsung.

*Machine learning* atau biasa disebut dengan ML adalah sebuah pendekatan dalam bidang kecerdasan buatan atau *artificial intelligence* (AI) yang digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah ataupun melakukan optimalisasi permasalahan tertentu. *Machine learning* menggunakan komputer untuk memprediksi dan memahami sifat ataupun ciri dari objek yang tidak dikenal dengan mengidentifikasi pola atau *pattern* dalam sebuah kumpulan data atau *dataset*. Ciri dari proses dalam *machine learning* adalah adanya proses pelatihan, pembelajaran, atau training dari program komputer sedemikian rupa sehingga program komputer tersebut dapat. Oleh karena itu *machine learning* membutuhkan data untuk dipelajari sebagai data latih atau data *training*-nya, dalam hal ini yaitu data pasien penyakit diabetes.

Di bidang kesehatan, *machine learning* dapat dimanfaatkan untuk mendeteksi risiko penyakit diabetes melitus pada ibu hamil. Beberapa penelitian telah menerapkan *machine learning* pada bidang kesehatan. Penelitian sebelumnya mengidentifikasi suatu penyakit gangguan kesehatan di mana kadar gula dalam darah seseorang menjadi tinggi karena kekurangan insulin atau reseptor insulin tidak dapat berfungsi baik. Dari hasil uji coba terhadap 8 pasien didapatkan persentase ketepatan sistem sebesar 62.5% (berdasarkan fakta dan gejala) dan hasil analisis sistem 100% (berdasarkan kadar gula darah). Oleh karena itu, dapat disimpulkan bahwa penyakit diabetes melitus merupakan salah satu penyakit yang berbahaya yang memerlukan pencegahan dan perawatan dini (Budi, Rosa, & Joko, 2011).

Pencegahan dini dapat dilakukan jika penderita diabetes melitus dapat mengetahui gejala-gejala yang dialami penderita. Diketahui bahwa gejala penyakit diabetes melitus sangat banyak dan beragam. Untuk mengidentifikasi gejala penyakit diabetes melitus sejak dini diperlukan bantuan dokter atau tenaga kesehatan yang ahli di bidangnya. Namun, untuk itu perlu didatangkan seorang dokter atau tenaga kesehatan untuk berkonsultasi serta diperlukan waktu dan biaya yang tidak sedikit di mana tidak semua

penderita diabetes melitus mampu untuk datang dan berkonsultasi dengan dokter. Dalam mengatasi permasalahan tersebut, diusulkan sebuah program aplikasi komputer yang mampu memberikan informasi deteksi dini terhadap risiko penyakit diabetes melitus yang dapat digunakan secara pribadi dan personal.

Terdapat beberapa teknik *machine learning* yang dapat diimplementasikan dalam penelitian ini, yaitu dengan metode K-Nearest Neighbor (KNN) dan Naïve Bayes (Han & Kamber, 2006; Cortez & Silva, 2008; Zhang, Oussena, & Hvensook, 2010). Namun dalam penelitian ini dibatasi pada implementasi metode KNN dalam proses klasifikasi datanya. Sehingga, objek kajian yang diteliti dalam penelitian ini adalah mengetahui sejauh mana tingkat akurasi metode KNN dapat diimplementasikan dalam melakukan proses klasifikasi data risiko penyakit diabetes melitus pada ibu hamil yang digunakan dalam penelitian ini serta mengetahui rasio data latih dan data uji yang optimal untuk digunakan dalam program aplikasi yang diimplementasikan.

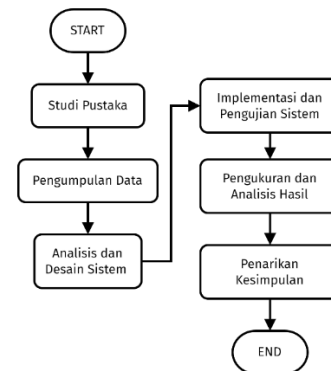
Berdasarkan permasalahan yang telah diuraikan sebelumnya serta dengan dukungan hasil penelitian yang telah dilakukan sebelumnya terkait implementasi pendekatan *machine learning* dalam mendeteksi risiko penyakit diabetes melitus pada ibu hamil dapat menghasilkan sebuah aplikasi komputer untuk mengidentifikasi tingkat risiko penyakit Diabetes Melitus ibu hamil kepada bayi yang tengah berada dalam kandungan seperti risiko kemungkinan bayi besar, hipoglikemia, perdarahan dan waktu persalinan yang lama (Fadilah, Saraswati, & Adi, 2016; Ahmad & Bialangi, 2021). Sehingga dengan sistem ini ibu hamil dapat mendeteksi risiko-risiko yang mungkin terjadi terhadap ibu dan bayi yang dikandungnya. Selain itu, sistem yang diimplementasikan diharapkan dapat mendukung dokter kandungan, bidan, maupun tenaga ahli di bidang kesehatan dalam memberikan perawatan, mengambil keputusan, dan melakukan tindakan preventif yang dapat diberikan sejak dini kepada ibu hamil untuk mencegah dan mengurangi risiko yang akan dihadapi dalam proses persalinan.

## 2. METODE PENELITIAN

Dalam menjawab rumusan masalah penelitian ini yang diangkat dalam penelitian ini, dilalui beberapa aktivitas yang tersusun atas dua tahapan utama. Aktivitas dalam tahap pertama utamanya adalah mengimplementasikan teknik *machine learning* dengan metode KNN menjadi sebuah aplikasi yang kemudian dilanjutkan dengan tahap pengujian sistem dengan menggunakan data penelitian yang telah diperoleh dari lapangan guna mengetahui tingkat akurasi metode yang dipilih untuk mengklasifikasi informasi yang diberikan ke dalam sistem menjadi beberapa klaster tingkatan risiko diabetes melitus terhadap ibu hamil.

Tahapan implementasi sistem ini pada dasarnya mengacu pada tahapan pengembangan perangkat

lunak yang umum digunakan dengan tujuan utama menghasilkan sebuah aplikasi yang mampu mengidentifikasi dan mendeteksi tingkat risiko kehamilan dengan diabetes melitus terhadap bayi dan proses persalinan. Proses klasifikasi yang dijalankan oleh disusun atas model yang terbentuk dari data masukan yang diberikan berdasarkan data yang diperoleh dari lapangan. Secara garis besar, aktivitas yang dilakukan dalam penelitian yang bersifat implementatif ini diperlihatkan dalam Gambar 1.



Gambar 1. Tahapan Penelitian

Studi pustaka yang dilakukan di awal penelitian ini mengumpulkan informasi dan referensi untuk dijadikan acuan dalam mengimplementasikan metode KNN dalam teknik *machine learning* pada data ibu hamil yang dikumpulkan menjadi sebuah aplikasi yang dapat digunakan pengguna. Pustaka yang diacu sebagai referensi bersumber dari jurnal, buku, maupun artikel-artikel teknis implementasi *machine learning* dari Internet yang meliputi:

- Parameter dan faktor-faktor yang berpengaruh dalam mendeteksi risiko tinggi Diabetes Melitus pada ibu hamil (Soewondo & Pramono, 2011).
- Teknik-teknik dalam bidang ilmu kecerdasan buatan sebagai bagian dalam teknik *machine learning* yang dapat digunakan untuk mengklasifikasi data yang diperoleh guna mendeteksi risiko kehamilan bagi penderita Diabetes Melitus (Cortez & Silva, 2008; Zhang, Oussena, & Hvensook, 2010).

### 2.1. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan teknik dokumentasi. Data yang akan digunakan dalam penelitian diperoleh dari *database* yang dimiliki Pusat Kesehatan Masyarakat (Puskesmas) Lawang Kabupaten Malang melalui seorang tenaga kesehatan bidan. Data yang digunakan berupa riwayat pemeriksaan pasien atau ibu hamil, baik telah yang memiliki riwayat penderita Diabetes Melitus maupun tidak. Data yang diperoleh berjumlah 248 yang meliputi informasi-informasi rekam medis seperti tanggal pengkajian, nama pasien/ibu hamil, tanggal HPHT, tanggal HPL, umur, GPA, usia kehamilan, KSPR, keterangan KSPR, IMT, kategori IMT, reduksi, albumin, hepatitis, HIV, dan IMS.

Tabel 1. Sampel data dari Puskesmas Lawang Kabupaten Malang

Umur	Gravida	Para	Abortus	Usia Kehamilan	Kategori IMT	Reduksi	Albumin	Hepatitis	HIV	IMS	Keterangan KSPR
22	2	1	0	9	Normal	Negatif	Negatif	NR	NR	NR	Risiko Rendah
36	1	0	0	8	Berlebih	Positif 1	Negatif	NR	NR	NR	Risiko Tinggi
25	1	0	0	5	Normal	Negatif	Negatif	NR	NR	NR	Risiko Rendah
30	3	1	1	10	Normal	Negatif	Negatif	NR	NR	NR	Risiko Sangat Tinggi
30	3	1	1	12	Normal	Negatif	Negatif	NR	NR	NR	Risiko Tinggi
28	2	1	0	6	Normal	Negatif	Negatif	NR	NR	NR	Risiko Tinggi
24	1	0	0	10	Normal	Negatif	Negatif	NR	NR	NR	Risiko Rendah
32	3	2	0	12	Normal	Negatif	Negatif	NR	NR	NR	Risiko Tinggi
25	1	0	0	10	Normal	Negatif	Negatif	NR	NR	NR	Risiko Rendah
27	2	1	0	12	Normal	Negatif	Negatif	NR	NR	NR	Risiko Rendah

## 2.2. Pengujian dan Analisis

Guna memudahkan dalam proses pengujian dan analisis hasil pengujian untuk menjawab rumusan masalah penelitian ini, sebelum parameter-parameter teknik *machine learning* diimplementasikan ke dalam sebuah program aplikasi komputer, terlebih dahulu diujikan dengan menggunakan perangkat lunak generik *machine learning* Waikato Environment for Knowledge Analysis (Weka) (Frank, Mark, Hall, & Witten, 2016) untuk mendapatkan nilai-nilai parameter dari metode KNN yang optimal untuk diterapkan pada program aplikasi *machine learning* yang akan diimplementasikan dalam penelitian ini.

Akurasi dari diukur dari sistem deteksi risiko kehamilan dengan metode KNN dalam penelitian ini dilakukan dengan merancang beberapa skenario dengan mengubah nilai beberapa atribut yang digunakan sebagai parameter dalam implementasinya sedemikian rupa sehingga tingkat akurasi hasil deteksi dengan klasifikasi menggunakan KNN yang optimal dapat dicapai. Sehingga tahap analisis yang dilakukan untuk menentukan seberapa baik tingkat akurasi metode KNN dalam mengklasifikasikan data risiko kehamilan pada penderita Diabetes Melitus ini dapat diketahui dengan menghitung nilai-nilai *precision*, *recall*, dan *F-measure* dari hasil klasifikasi data yang digunakan dalam penelitian ini.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Data

Dari hasil proses pengumpulan data melalui proses wawancara dan visitasi di lapangan diperoleh lebih dari 260 rekaman data rekam medis pada pemeriksaan ibu hamil di Puskesmas Lawang Kabupaten Malang. Beberapa sampel data yang diperoleh dan digunakan dalam penelitian ini diperlihatkan pada Tabel 1. Data yang digunakan adalah data yang diperoleh dari hasil pemeriksaan pasien ibu hamil pada bulan Januari 2021 sampai dengan bulan Agustus 2021.

Sebelum data tersebut digunakan pada tahap implementasi dan pengujian, dilakukan proses *preprocessing* yang melibatkan dua proses, yaitu seleksi dan pembersihan data. Proses seleksi data atau

*data selection* merupakan pemilihan himpunan data atau seleksi data dari sekumpulan data. Selanjutnya dilakukan pembersihan data atau *data cleaning* di mana data-data yang salah atau tidak lengkap diperbaiki dan konsistensinya diperiksa. Data yang rusak tidak disertakan dalam analisis dan pengujian. Format data yang diperoleh kemudian disesuaikan melalui proses *data transformation* menjadi nilai dengan format-format tertentu yang dapat dipahami oleh program komputer.

Proses seleksi data dilakukan dengan cara melakukan observasi dengan jurnal-jurnal yang memiliki topik dan bahasan yang serupa dengan penelitian ini. Data yang diperoleh diverifikasi dan divalidasi dengan melakukan wawancara langsung dengan dokter dan bidan yang terkait dengan data yang digunakan dalam penelitian ini. Para tenaga kesehatan yang diwawancarai dalam validasi data penelitian ini juga diminta untuk menentukan atribut apa saja yang memiliki pengaruh pada risiko kehamilan dengan penyakit diabetes melitus.

Pada proses tahap *data cleaning*, data yang tidak relevan atau data yang nilainya kosong dihilangkan dari *dataset* menghasilkan data yang siap digunakan dalam pengujian dan analisis. Data yang telah melalui proses *cleaning* ini digunakan sebagai data latih dan data uji dalam penelitian ini. Atribut data yang digunakan dari data pemeriksaan ibu hamil dalam penelitian ini mencakup tanggal pemeriksaan, nama ibu hamil, tanggal HPHT, tanggal HPL, umur, Gravida-Para-Abortus (GPA), usia kehamilan, IMT, kategori IMT, KSPR, Keterangan KSPR, Reduksi, Albumin, Hepatitis, HIV, dan IMS seperti yang diperlihatkan pada Tabel 1.

Tabel 2. Hasil Pengujian Akurasi Proses Klasifikasi Klaster Risiko dengan KNN

Kelas	Precision					Recall					F-measure (F <sub>1</sub> )				
	Proporsi Data Latih Terhadap Data Uji														
	90%	80%	70%	60%	50%	90%	80%	70%	60%	50%	90%	80%	70%	60%	50%
<b>Metode KNN k=2</b>															
RT	30%	35%	48.1%	55.3%	53.1%	75%	58.3%	65%	65.6%	61.9%	42.9%	43.8%	55.3%	60%	57.1%
RR	93%	83.3%	83%	80%	74.3%	80%	67.6%	75%	76.2%	71.4%	80%	74.6%	78.8%	78%	72.8%
RST	-	-	-	0%	0%	-	0%	0%	0%	0%	-	-	-	0%	0%
<b>Metode KNN k=3</b>															
RT	28.6%	42.9%	57.1%	60.7%	59.4%	50%	50%	60%	53.1%	45.2%	36.4%	46.2%	58.5%	56.7%	51.4%
RR	88.9%	83.3%	84.6%	78.3%	76.7%	80%	81.1%	84.6%	85.7%	85.7%	84.2%	82.2%	84.6%	81.8%	81%
RST	-	-	0%	0%	16.7%	0%	0%	0%	0%	20%	-	-	0%	0%	18.2%
<b>Metode KNN k=4</b>															
RT	28.6%	46.2%	61.9%	66.7%	67.6%	50%	50%	65%	56.3%	59.5%	36.4%	48%	63.4%	61%	63.3%
RR	88.9%	83.8%	86.5%	79.7%	80.7%	80%	83.8%	86.5%	87.3%	87%	84.2%	83.8%	86.5%	83.3%	83.8%
RST	-	-	0%	33.3%	25%	0%	0%	0%	25%	20%	-	-	0%	28.6%	22.2%
<b>Metode KNN k=5</b>															
RT	16.7%	41.7%	60%	62.5%	69%	25%	41.7%	60%	46.9%	47.6%	20%	41.7%	60%	53.6%	56.3%
RR	84.2%	81.6%	84.9%	76.4%	75%	80%	83.8%	86.5%	87.3%	89.6%	82.1%	82.7%	85.7%	81.5%	81.7%
RST	-	-	0%	66.7%	66.7%	0%	0%	0%	50%	40%	-	-	0%	57.1%	50%

\*Kelas RT: Risiko Tinggi, RR: Risiko Rendah, RST: Risiko Sangat Tinggi

### 3.2. Implementasi

Implementasi algoritme apriori dengan tools *machine learning* Weka dalam penelitian ini menggunakan data sebanyak 248 data dengan beberapa skenario pengujian untuk mengetahui seberapa besar tingkat akurasi dari metode KNN yang digunakan untuk mendeteksi tingkat risiko kehamilan dengan diabetes melitus.

Melalui perangkat lunak Weka yang digunakan dalam analisis data penelitian menggunakan konfigurasi classifier Lazy IBk sebagai metode yang digunakan untuk analisis klasifikasi yang menggunakan metode KNN.

Dataset dengan jumlah 248 data pemeriksaan ibu hamil dalam proses klasifikasi untuk risiko diabetes dibagi menjadi data latih dan data uji. Data latih digunakan untuk mendapatkan model atau pola pembelajaran *machine learning* sedangkan data uji digunakan untuk menguji kesesuaian model atau pola yang didapat dari data latih. Penelitian ini menggunakan metode KNN sebagai metode klasifikasi untuk deteksinya.

### 3.3. Pengujian

Pengujian yang dilakukan dalam penelitian ini untuk mendapatkan nilai akurasi metode KNN yang diterapkan dalam *machine learning* dirancang beberapa skenario pengujian dengan penentuan besarnya nilai parameter ketetangaan k dalam KNN yang berbeda-beda. Jumlah klaster data tingkat risiko kehamilan yang ada dalam data pemeriksaan yang diperoleh adalah 3 (risiko rendah, tinggi, dan sangat tinggi) sehingga dalam pengujian yang dilakukan dalam penelitian ini menggunakan nilai k=2, 3, 4, dan 5 untuk menentukan tingkat kedekatan anggota suatu klaster dengan anggota lainnya. Selain diuji dengan

menggunakan nilai k yang berbeda-beda, pengujian dalam penelitian ini juga dilakukan dengan melakukan pembagian data latih dengan data uji sebesar 90%, 80%, 70%, 60%, 50%, dan 40%. Di mana nilai persentase yang dipilih adalah proporsi yang ditentukan untuk data latih, sedangkan sisanya digunakan sebagai data uji untuk mengukur tingkat akurasi sistem yang diimplementasikan.

Dalam proses pengujian sistem yang diimplementasikan dalam penelitian ini, terdapat empat skenario pengujian yang dibuat berdasarkan jumlah klaster k yang telah ditentukan sebelumnya.

#### Hasil Uji Pada Skenario Pengujian

Pada pengujian skenario pertama dilakukan dengan nilai ketetangaan k=2 dengan pembagian data latih 90%, 80%, 70%, 60%, 50%, dan 40%, pengujian ini mendapatkan nilai akurasi tertinggi sebesar 70.27% dengan rasio data latih dan data uji 7:3. Nilai *precision* tertinggi pada kelas Risiko Rendah dengan data latih 90% sebesar 93%. Nilai *recall* terbesar pada kelas Risiko Rendah dengan data latih 90% yaitu 80%, dan nilai *F-measure* terbesar pada kelas Risiko Rendah dengan data latih 90% yaitu 80%.

Pada pengujian skenario kedua dengan nilai konfigurasi jumlah nilai ketetangaan k=3 metode KNN dengan rasio pembagian data latih dan data uji yang sama dengan skenario pertama, didapat hasil pengujian skenario kedua ini dengan nilai akurasi tertinggi sebesar 75.68% pada rasio pembagian data latih dengan data uji sebesar 7:3. Nilai *precision* tertinggi pada kelas Risiko Rendah dengan data latih 90% sebesar 88.9%. Nilai *recall* terbesar pada kelas Risiko Rendah dengan data latih 60% dan 50% yaitu 85.7%, dan *F-measure* terbesar pada kelas Risiko Rendah dengan data latih 70% yaitu sebesar 84.6%.

Pada pengujian skenario ketiga dengan nilai ketetangaan  $k=4$  metode KNN pada dataset dengan pembagian data latih 90%, 80%, 70%, 60%, 50%, dan 40%, pengujian skenario ketiga ini mampu mendapatkan nilai akurasi tertinggi sebesar 78.38% pada pembagian data latih 70% dan data testing 30%. Nilai *precision* tertinggi pada kelas Risiko Rendah dengan data latih 90% sebesar 88.9%. Nilai *recall* terbesar pada kelas Risiko Rendah dengan data latih 60% yaitu 87.3% dan F-measure terbesar pada kelas Risiko Rendah dengan data latih 70% yaitu 86.5%.

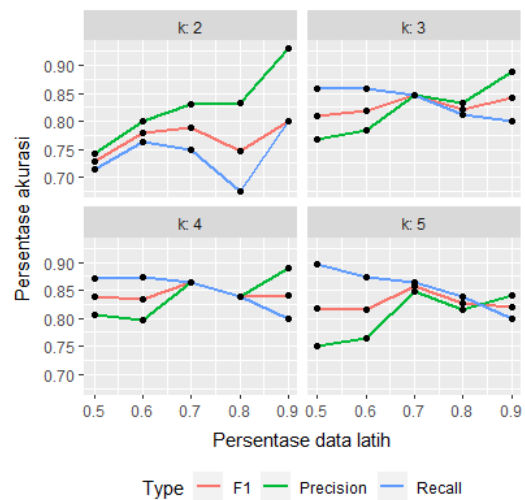
Pada pengujian skenario keempat dengan jumlah nilai ketetangaan  $k=5$  dengan pembagian data latih sebesar 90%, 80%, 70%, 60%, 50%, dan 40%, pengujian skenario keempat ini mampu mendapatkan nilai akurasi tertinggi sebesar 77.03% pada pembagian data latih 70% dan data testing 30%. Nilai *precision* tertinggi pada kelas Risiko Rendah dengan data latih 70% sebesar 84.9%. Nilai *recall* terbesar pada kelas Risiko Rendah dengan data latih 50% yaitu 89.6% dan F-measure terbesar pada kelas Risiko Rendah dengan data latih 70% yaitu 85.7%. Hasil pengujian akurasi dengan jumlah data dalam dataset sebanyak 248 data untuk setiap skenario pengujian dalam penelitian ini diperlihatkan pada Tabel 2.

Grafik yang memperlihatkan akurasi metode KNN pada machine learning untuk deteksi risiko kehamilan dengan diabetes melitus risiko rendah, risiko tinggi, dan risiko sangat tinggi diperlihatkan dalam Gambar 2, Gambar 3, dan Gambar 4 secara berurutan. Secara umum, capaian akurasi tinggi diperoleh untuk pada jumlah data latih sebesar 70% (0,7). Namun khusus untuk klaster risiko sangat tinggi, nilai akurasi metode KNN tidak bisa dicapai dengan baik dikarenakan oleh jumlah data yang tidak memadai untuk memenuhi jumlah rasio yang ada.

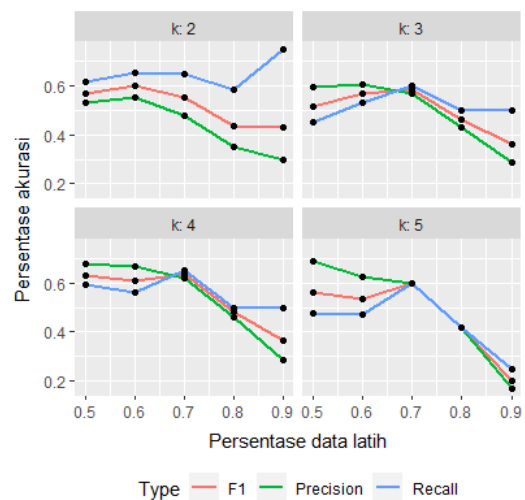
**Perbandingan Tingkat Akurasi Tiap Skenario Pengujian**

Dari pengujian yang telah dilakukan pada empat skenario pengujian dengan nilai ketetangaan  $k$  dari metode KNN yang berbeda-beda, yaitu nilai  $k=2, 3, 4,$  dan  $5$  dengan jumlah proporsi pembagian data latih 90%, 80%, 70%, 60%, dan 50% dari jumlah total data keseluruhan sebanyak 248 data, metode KNN yang digunakan untuk mendeteksi tingkat risiko diabetes melitus pada kehamilan dapat dievaluasi melalui hasil pengukuran nilai-nilai akurasi, *precision*, *recall*, dan *F-measure*. Hasil perbandingan skenario pengujian dapat dilihat pada Tabel 2 dan Tabel 3.

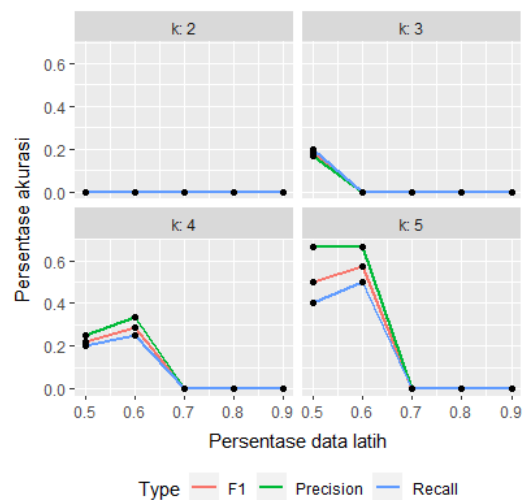
Nilai akurasi tertinggi yang diperoleh dari skenario pengujian pertama dengan nilai  $k=2$  adalah sebesar 70.27%. Nilai akurasi tertinggi pada skenario pengujian kedua dengan  $k=3$  adalah sebesar 75.68%. Nilai akurasi tertinggi pada skenario pengujian ketiga dengan nilai  $k=4$  adalah sebesar 78.38%. Sedangkan Nilai akurasi tertinggi pada skenario pengujian keempat dengan jumlah nilai ketetangaan  $k=5$



Gambar 2. Akurasi metode KNN klaster Risiko Rendah



Gambar 3. Akurasi metode KNN klaster Risiko Tinggi



Gambar 4. Akurasi metode KNN klaster Risiko Sangat Tinggi

adalah sebesar 77.03%. Semua nilai akurasi tertinggi diperoleh dengan rasio pembagian data latih dan data uji sebesar 7:3.

Walaupun demikian, dikarenakan oleh adanya batasan dari jumlah dan proporsi jumlah data untuk tiap kategori risiko kehamilan yang digunakan dalam penelitian ini, terdapat beberapa hal yang terjadi pada saat melakukan pengujian. Seperti nilai akurasi kosong atau sebesar 0% dikarenakan pada saat pengambilan data latih dan data uji dilakukan secara acak oleh perangkat lunak analisis *machine learning* Weka yang digunakan, terdapat beberapa data pada kelas risiko sangat tinggi tidak diambil sebagai data latih maupun data ujinya. Selain itu, jumlah data untuk tiap kategori risiko yang ada di dalam *dataset* yang diperoleh tidak seimbang jumlahnya antara data risiko tinggi, risiko rendah, dan risiko sangat tinggi. Dari *dataset* yang digunakan, jumlah data risiko tinggi sebesar 71, risiko rendah sebesar 158, dan risiko sangat tinggi hanya terdapat 19 data dari 248 total jumlah data yang digunakan. Hasil perbandingan skenario pengujian akurasi untuk nilai ketetanggan metode KNN yang berbeda-beda dalam penelitian ini dapat dilihat pada Tabel 3.

Tabel 3. Perbandingan Akurasi Tiap Skenario Pengujian

Data Latih	Data Uji	Nilai Akurasi			
		$k=2$	$k=3$	$k=4$	$k=5$
90%	10%	68%	72%	70%	68%
80%	20%	64%	72%	74%	72%
70%	30%	70.27%	75.68%	78.38%	77.03%
60%	40%	69.70%	71.72%	74.75%	72.73%
50%	50%	65.32%	69.35%	75%	73.39%

#### 4. KESIMPULAN DAN SARAN

Berdasarkan pembahasan penelitian yang telah dilakukan, maka penelitian disimpulkan bahwa dalam mendeteksi risiko kehamilan diabetes melitus pada ibu hamil menggunakan *machine learning* dapat dilakukan dengan menggunakan metode K-Nearest Neighbor (KNN), Perangkat lunak yang digunakan untuk mengukur akurasi metode klasifikasi KNN yang digunakan adalah Weka, sehingga prosedur implementasi klasifikasi yang digunakan mengacu pada perangkat lunak tersebut. Jumlah data dalam *dataset* yang digunakan dalam penelitian ini adalah sebanyak 248 data pemeriksaan ibu hamil yang diperoleh dari puskesmas Lawang mulai dari bulan Januari sampai dengan bulan Agustus tahun 2021. Penelitian ini hanya menggunakan 16 atribut penting dalam penentuan tingkat risiko kehamilan diabetes melitus pada ibu hamil yaitu: tanggal pemeriksaan, nama ibu hamil, tanggal HPHT, tanggal HPL, umur, GPA, IMT, kategori IMT, usia kehamilan, KSPR, keterangan KSPR, Reduksi, Albumin, Hepatitis, HIV, dan IMS. Dari hasil implementasi dan pengujian yang dilakukan, dapat diketahui bahwa teknik *machine learning* menggunakan metode KNN dapat

digunakan untuk mendeteksi tingkat risiko kehamilan diabetes melitus pada ibu hamil.

Dari hasil pengujian yang dilakukan dengan empat kali skenario pengujian. Nilai akurasi tertinggi pada skenario pengujian pertama dengan nilai ketetanggan  $k=2$  sebesar 70.27%. Nilai akurasi tertinggi pada skenario pengujian kedua dengan nilai ketetanggan  $k=3$  adalah sebesar 75.68%. Nilai akurasi tertinggi pada skenario pengujian ketiga dengan nilai ketetanggan  $k=4$  adalah 78.38%. Sedangkan Nilai akurasi tertinggi pada skenario pengujian keempat dengan nilai ketetanggan  $k=5$  adalah sebesar 77.03%. Semua nilai akurasi tertinggi dicapai dengan proporsi pembagian data latih dan data uji sebesar 7:3. Sehingga dapat ditarik suatu kesimpulan bahwa rasio perbandingan data latih dan data uji yang optimal untuk mencapai tingkat akurasi tertinggi dari proses klasifikasi *machine learning* dengan metode KNN untuk jenis data ini adalah sebesar 7:3.

Adapun kekurangan dari pengujian yang dilakukan dalam penelitian ini adalah adanya beberapa nilai data dalam *dataset* yang tidak seimbang sehingga besar kemungkinan terjadi bahwa data yang digunakan pada kelas Risiko Sangat Tinggi tidak digunakan sebagai data latih untuk proses training pada metode KNN *machine learning* yang digunakan. Selain itu, *dataset* yang diperoleh tidak seimbang antara risiko tinggi, risiko rendah, dan risiko sangat tinggi di mana jumlah data risiko tinggi sama dengan 71, risiko rendah sama dengan 158, dan risiko sangat tinggi sama dengan 19. Hal tersebut dapat berimbas pada rendahnya hasil pengukuran akurasi yang didapat dalam penelitian ini. Sehingga, hasil implementasi metode KNN pada teknik implementasi *machine learning* untuk mendeteksi tingkat risiko kehamilan ibu hamil dengan diabetes melitus hanya dapat digunakan sebagai dasar pertimbangan, *second opinion*, deteksi dini secara personal, sehingga hasil deteksi yang diperoleh hanya bersifat sementara dan bukan sebagai pengganti diagnosis dan keputusan dari para dokter, ahli, dan tenaga kesehatan yang berwenang dalam mengambil keputusan.

Berdasarkan batasan dan kekurangan yang ada dari penelitian ini, maka adapun saran yang dapat diberikan adalah perlunya dilakukan pengumpulan data yang lebih banyak dan seimbang untuk setiap kelas risiko kehamilan yang dideteksi. Penelitian selanjutnya juga dapat diimplementasikan dengan menggunakan metode atau algoritme klasifikasi *machine learning* lainnya yang mampu menangani data yang jumlahnya tidak seimbang guna mencapai nilai akurasi dan hasil deteksi yang lebih baik.

#### 5. APRESIASI

Ucapan terima kasih disampaikan kepada dr. Julia Rosana sebagai Kepala Puskesmas kota Lawang Kabupaten Malang dan Lilik Agustin, S.ST., Bd. sebagai bidan koordinator di Puskesmas kota Lawang Kabupaten Malang. Penelitian ini didanai dari dana program hibah penelitian untuk profesor dan doktor

tahun anggaran 2022 Fakultas Ilmu Komputer Universitas Brawijaya, No: 2128/UB10.F15/PN/2022.

#### DAFTAR PUSTAKA

- KEMENKES RI, 2019. Profil Kesehatan Indonesia tahun 2018
- PERKENI, 2021. Pedoman Pengelolaan dan Pencegahan Diabetes Melitus Tipe 2 Dewasa di Indonesia. PB PERKENI.
- INTERNATIONAL DIABETES FEDERATION (IDF), 2015. Diabetes Atlas 7th Edition, International Diabetes Federation, Brussels, Belgium. <http://www.diabetesatlas.org>
- HIDAYATI, R., SETYORINI, D., & NUARI, N.A., 2018. Differences Complications During Perinatal in Hiatory of Woman With Diabetes Mellitus And Obesity Gestational. 148-160
- WOJTYLA, C., STANIROWSKI, P., CIEBIERA, M.P., & WOJTYLA A., 2021. Perinatal Outcomes in a Population of Diabetic and Obese Pregnant Women—The Results of the Polish National Survey, 18(516), 1-14
- HAN, J. & KAMBER, M., 2006. Data Mining Concepts and Techniques Second Edition. San Francisco: Morgan Kaufmann.
- BUDI CAHYO SAPUTRO, ROSA DELIMA, & JOKO PURWADI., 2011. Sistem Diagnosa Penyakit Diabetes Melitus Menggunakan Metode Certainty Factor. Jurnal Informatika Vol 7(11). <http://dx.doi.org/10.21460/inf.2011.71.92>
- FRANK, E, A, MARK. HALL, & WITTEN, I.H., 2016. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition.
- CORTEZ, P. & SILVA, A.M.G., 2008. Using Data Mining to Predict Secondary School Student Performance. In: Brito, A., Teixeira, J. (eds.) Proceedings of 5th Future Business Technology Conference (FUBUTEC 2008). Porto, Portugal
- ZHANG, Y., OUSSENA, S., CLARK, T., & HYENSOOK, K., 2010. Using Data Mining to Improve Student Retention in Higher Education: A Case Study. In: 12th International Conference on Enterprise Information Systems (ICEIS), pp. 1-8. Madeira. Portugal
- SOEWONDO, P. & PRAMONO, 2011. L.A: Prevalence, Characteristic, and Predictors of PreDiabetes in Indonesia. Jakarta: Departement of Internal Medicine, Faculty of Medicine of University Indonesia. *Med.J.Indonesia*. 20, 283-294
- FADILAH, N. A., SARASWATI, L. D., & ADI, M. S., 2016. Gambaran karakteristik dan faktor-faktor yang berhubungan dengan kejadian diabetes melitus tipe 2 pada wanita (Studi di RSUD Kardinah Kota Tegal). *Jurnal Kesehatan Masyarakat (e-Journal)*, 4(1), 176-183
- AHMAD, F., & BIALANGI, S., 2021. Hubungan Riwayat Keluarga dan Perilaku Sedentari terhadap Kejadian Diabetes Melitus. *Jambura Journal*, 3(1), 103–114.