

PENERAPAN ALGORITMA ID3 DAN SVM PADA KLASIFIKASI PENYAKIT DIABETES MELITUS TIPE 2

Reza Pahlevi¹, Ketut Queena Fredlina², Nengah Widya Utami³

^{1,2,3}Program Studi Teknik Informatika STMIK Primakara

e-mail :¹rezapahlevi2505@gmail.com,²naa.queena@gmail.com,³widya@primakara.ac.id,

ABSTRACT

Diabetes mellitus type 2 is a disease often suffered by people today and labeled as a "silent killer disease" because the sufferer didn't even realize it for years. The current habits of people who like to eat junk food, sleep irregularly, and lack of exercise cause type 2 diabetes mellitus to infect many people easily. Based on these problems, we need a system that can classify type 2 diabetes mellitus with predetermined requirements. To classify type 2 diabetes mellitus by applying Iterative Dichotomiser 3 (ID3) and Support Vector Machine (SVM) algorithms. The classification results only be early identification in predicting type 2 diabetes mellitus and the two algorithms can be used as a comparison, which algorithm is more effective in carrying out the classification of type 2 diabetes mellitus. The results of classification with the 725 total data with 75% train data and 25% test data using the ID3 algorithm, Linear SVM and RBF SVM, the most effective algorithm is Linear SVM algorithm with an accuracy of 78.5%, 79% of precision, 56% of recall, and 66% of f1 score. For the ROC / AUC method, Linear SVM gets a result of 74% and is included in the "Fair Classification" category.

Keywords : ID3, SVM, type 2 diabetes mellitus

INTISARI

Diabetes Melitus tipe 2 merupakan salah satu penyakit yang sering diidap oleh masyarakat saat ini dan dilabeli sebagai "silent killer disease" karena penderitanya bahkan tidak menyadarinya sampai bertahun-tahun. Kebiasaan masyarakat saat ini yang gemar mengonsumsi junk food, tidur tidak beraturan, serta kurangnya olahraga menyebabkan penyakit diabetes melitus tipe 2 banyak menjangkiti penderita usia muda maupun tua. Berdasarkan permasalahan tersebut, maka diperlukan sebuah sistem yang dapat mengklasifikasi penyakit diabetes melitus tipe 2 dengan persyaratan yang telah ditetapkan. Untuk melakukan klasifikasi penyakit diabetes melitus tipe 2 yaitu dengan menerapkan algoritma Iterative Dichotomiser 3 (ID3) dan Support Vector Machine (SVM). Hasil klasifikasi nanti hanya berupa identifikasi awal dalam memprediksi penyakit diabetes melitus tipe 2 serta kedua algoritma tersebut dapat digunakan sebagai pembandingan antara algoritma mana yang lebih efektif dalam menjalankan klasifikasi diabetes melitus tipe 2. Hasil dari klasifikasi dengan jumlah data sebanyak 725 data dengan 75% data train dan 25% data test menggunakan algoritma ID3, SVM kernel Linear dan RBF, didapat algoritma yang paling efektif jatuh pada algoritma SVM Linear dengan hasil accuracy sebesar 78,5%, precision sebesar 79%, recall sebesar 56%, dan f1 score sebesar 66%. Untuk metode penilaian ROC/AUC, SVM Linear mendapat hasil sebesar 74% dan masuk golongan "Fair Classification".

Kata kunci : ID3, SVM, diabetes melitus tipe 2

1. PENDAHULUAN

Diabetes melitus atau yang sering disebut dengan penyakit kencing manis adalah penyakit gangguan metabolik menahun yang ditandai dengan kadar glukosa dalam darah yang melebihi batas nilai normal (Efendi & Subhan Wibawa, 2018). Dari sekian banyaknya jenis diabetes melitus yang ada, diabetes melitus tipe 2 paling banyak diderita oleh masyarakat dunia, sehingga dilabeli dengan sebutan "silent killer disease", karena pengidap penyakit tersebut banyak yang tidak menyadari bahwa dirinya terkena penyakit diabetes melitus hingga diketahui sudah kronis. Selain karena gaya hidup yang kurang sehat, diabetes juga bisa disebabkan oleh faktor turunan.

Diabetes melitus tipe 2 terjadi ketika pankreas mampu memproduksi insulin yang memadai, namun sel-sel tubuh tidak dapat menggunakan insulin secara efisien untuk mengubah glukosa menjadi energi (Rahmay et al., 2017). Di Indonesia, berdasarkan data dari *International Diabetes Federation (IDF) Atlas tahun 2017*, jumlah penderita diabetes di Indonesia mencapai 10,3 juta jiwa. Tidak heran jumlah penderita diabetes melitus di Indonesia masuk ke dalam urutan ke-6. Berdasarkan data dari Riset Kesehatan Dasar 2018, angka prevalensi penderita diabetes melitus pada orang dewasa tahun 2013 mencapai 6,9%, dan tahun 2018, angka prevalensinya meningkat mencapai 8,5%. Dan uniknya, 90% dari penderita diabetes pada tahun 2013 mengidap penyakit diabetes melitus tipe 2 (Putri, 2018).

Berkaca dari data tersebut, sudah seharusnya dilakukan langkah-langkah pencegahan serta pengobatan pada penyakit diabetes melitus sedini mungkin. Namun sayangnya, pola hidup sebagian masyarakat yang cenderung tidak sehat membuat jumlah penderita penyakit diabetes melitus semakin meningkat dari tahun ke tahun.

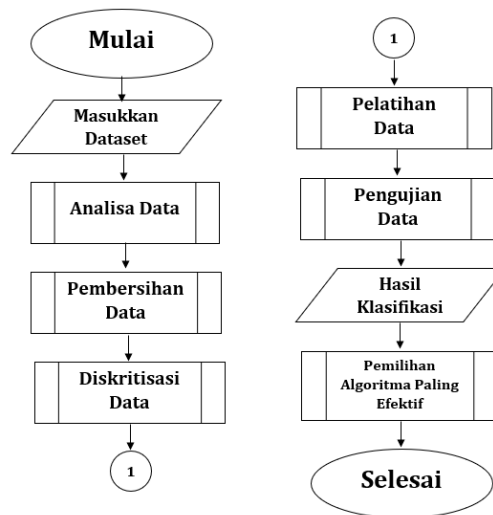
Ada berbagai macam cara untuk mengklasifikasikan seseorang terkena penyakit diabetes melitus tipe 2, salah satunya adalah dengan menerapkan metode *data mining* yang merupakan metode yang digunakan untuk menggali data dan hasil akhirnya berupa informasi baru (Widiyati et al., 2018). Hasil akhir tersebut juga dapat digunakan untuk membuat keputusan baru (Utami et al., 2019). Untuk menggunakan metode *data mining*, terdapat berbagai macam algoritma atau metode yang digunakan. Beberapa diantaranya seperti *Iterative Dichotomiser 3* (ID3) dan *Support Vector Machine* (SVM). ID3 merupakan algoritma yang mengandalkan bantuan *decision tree* untuk pengambilan keputusannya (Mutmainah, 2017), sedangkan SVM sendiri merupakan algoritma pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah fitur berdimensi tinggi (Puspitasari et al., 2018).

Tujuan diimplementasikan algoritma ID3 dan SVM untuk penelitian klasifikasi penyakit diabetes melitus tipe 2 untuk dapat membantu melakukan identifikasi awal penyakit diabetes melitus tipe 2 menggunakan algoritma ID3 dan SVM menggunakan *dataset* berisi data penderita penyakit diabetes melitus tipe 2. *Dataset* tersebut diolah, kemudian menghasilkan nilai prediksi penyakit diabetes melitus tipe 2. Setelah nilai prediksi dari ID3 dan SVM keluar, selanjutnya dilakukan perbandingan penilaian kinerja dari algoritma ID3 dan SVM untuk menentukan mana algoritma yang paling efektif diantara ID3 dan SVM. Hasil dari perbandingan algoritma tersebut nantinya akan menjadi patokan dalam melakukan identifikasi awal dalam identifikasi penyakit diabetes melitus tipe 2.

2. METODOLOGI PENELITIAN

2.1 Skema Alur Penelitian

Skema alur penelitian memuat tentang gambaran besar/konsep dari sebuah penelitian. Alur penelitian yang akan diimplementasikan pada hasil akhir penelitian akan dijabarkan melalui *flowchart* dibawah ini.



Gambar 1. Flowchart Penelitian

Berdasarkan *flowchart* diatas, tahap pertama adalah mempersiapkan *dataset* yang akan digunakan. Setelah itu akan dilakukan analisa data untuk mencari nilai 0 pada *dataset*. Setelah nilai 0 pada *dataset*, kemudian dilakukan pembersihan data untuk menghilangkan nilai 0 pada *dataset* tersebut. Setelah *dataset* tersebut bersih dari nilai 0, tahap selanjutnya adalah diskritisasi data untuk mempermudah klasifikasi menggunakan algoritma ID3 dan SVM.

Tahap selanjutnya adalah pelatihan data menggunakan algoritma ID3 dan SVM. Setelah *dataset* tersebut dilatih, tahap berikutnya adalah melakukan *testing* data pada algoritma ID3 dan SVM. Terakhir, tahap pemilihan algoritma yang efektif akan meliputi penilaian kerja, seperti *Accuracy*, *Precision*, *Recall*, *F1-Score*, nilai ROC/AUC dan waktu eksekusi algoritma (Widiyati et al., 2018). Implementasi penelitian ini akan menggunakan *software* Jupyter Notebook versi 6.0.0 yang sudah dilengkapi dengan bahasa pemrograman Python versi 3.7.

2.2 Pengumpulan Data

Metode pengumpulan data primer dilakukan melalui wawancara terbuka oleh narasumber dari Bagian Pendidikan dan Penelitian Rumah Sakit Sanglah Denpasar. Data sekunder diperoleh melalui pencarian *dataset* yang diperoleh melalui *website UC Irvine Machine Learning Repository* dan studi pustaka. Studi pustaka dilakukan untuk mendapatkan referensi sebanyak-banyaknya terkait penelitian tentang penyakit Diabetes Melitus Tipe 2 menggunakan algoritma ID3 dan SVM.

2.3 Persiapan Data

Data yang digunakan pada penelitian ini adalah *dataset Pima Indians Diabetes*. *Dataset* ini berisi data-data informasi tentang pasien yang diduga memiliki penyakit diabetes. Keseluruhan data pada *dataset* ini berjumlah 768 records. Atribut-atribut pada *dataset* yang akan digunakan sebagai berikut:

1. *Number of times pregnant* : Jumlah kehamilan
2. *Plasma glucose concentration for 2 hours after eating* (mg/dl) : Kadar glukosa 2 jam setelah makan
3. *Diastolic blood pressure* (mmhg) : Tekanan darah
4. *Triceps skin fold* (mm) : Ketebalan lipatan kulit *triceps*
5. *2-Hour Serum Insulin* (mu U/ml) : Banyaknya insulin dalam darah
6. *Body Mass Index* (kg/m²) : Berat Badan
7. *Diabetes pedigree function* : Peluang mengalami diabetes berdasarkan turunan/gen
8. *Age* (Year) : Umur
9. *Class Variable* (0 or 1) \square target

2.4 Tahap Pra Pemrosesan

Dalam penelitian ini, atribut-atribut yang akan digunakan sudah dijabarkan sebelumnya, namun ada perubahan dalam penyebutan atribut yang akan dipakai dalam implementasi penelitian ini ke dalam *software*. Perubahan penyebutan atribut ini menyesuaikan nama atribut yang akan dipakai dalam *dataset*, diantaranya adalah:

1. *Number of times pregnant* diubah menjadi **Pregnancies**.
2. *Plasma glucose concentration for 2 hours after eating* diubah menjadi **Glucose**.
3. *Diastolic blood pressure* diubah menjadi **BloodPressure**.
4. *Triceps skin fold* diubah menjadi **SkinThickness**.
5. *2-Hour Serum Insulin* diubah menjadi **Insulin**.
6. *Body Mass Index* diubah menjadi **BMI**.
7. *Diabetes pedigree function* diubah menjadi **DPF**.
8. Untuk atribut *Age* dan *Class* tidak mengalami perubahan penyebutan atribut.

Langkah pertama untuk melakukan klasifikasi data adalah dengan melakukan *data analyzing*. *Data analyzing* berfungsi untuk mencari apakah masih ada data yang kosong pada *dataset* tersebut. Atribut-atribut yang memiliki nilai kosong dijabarkan dalam table 1 dibawah ini.

Tabel 1. Tabel *Dataset* Sebelum Dibersihkan

No	Nama Atribut	Keterangan	Jumlah Data Nol
1	<i>Pregnancies</i>	Jumlah Kehamilan	111
2	<i>Glucose</i>	Kadar Glukosa 2 Jam Setelah Makan	5
3	<i>Blood Pressure</i>	Tekanan Darah	35
4	<i>Skin Thickness</i>	Ketebalan Lipatan Kulit Triceps	227
5	<i>Insulin</i>	Banyaknya Insulin dalam Darah	374
6	<i>BMI</i>	Berat Badan	11
7	<i>DPF</i>	Peluang Mengalami Diabetes Berdasarkan Gen	0
8	<i>Age</i>	Umur	0
9	<i>Class</i>	Target Variabel	0

Ada beberapa cara yang bisa dipakai untuk mengatasi nilai yang tidak lengkap pada masing-masing atribut (Lesmana, 2012), yaitu:

1. Menghapus data yang tidak memiliki nilai lengkap. Tetapi dengan cara ini akan menyebabkan informasi penting pada atribut lain yang terkandung dalam *dataset* tersebut banyak yang hilang.
2. Menggunakan nilai *mean* sebagai pengganti nilai yang hilang pada *dataset* tersebut. Tetapi karena nilai yang hilang tersebut ada banyak, cara ini akan membuat *dataset* tersebut menjadi tidak relevan dengan kondisi aslinya.
3. Mengganti semua nilai yang tidak ada dengan 0. Tetapi, ini akan berpengaruh besar terhadap hasil klasifikasi nanti dan hasilnya akan condong ke hasil yang tidak baik.

- Mengganti nilai yang tidak ada dengan nilai dari atribut lain yang memiliki jarak atau kemiripan terdekat atau bisa disebut dengan *euclidian distance*. Cara ini memang lebih baik, tetapi karena banyaknya nilai yang hilang, *dataset* tersebut nantinya tidak akan bisa mencerminkan kondisi yang sebenarnya.

Berdasarkan kondisi yang terjadi diatas dapat dilakukan kombinasi dari keempat cara diatas untuk mengatasi data/nilai yang hilang dari *dataset* yang akan digunakan. Berikut adalah aturan yang akan dipakai dalam melakukan *data cleaning* pada *dataset* yang akan digunakan (Lesmana, 2012).

- Nilai 0 pada atribut *Pregnancy* kita akan anggap pasien tersebut belum pernah melahirkan.
- Untuk atribut *Glucose*, *BloodPressure*, dan *BMI*, nilai 0 tersebut dapat dihilangkan karena jumlah data yang nilainya 0 tidak begitu banyak.
- Untuk atribut *SkinThickness* dan *Insulin*, nilai 0 yang ada pada atribut tersebut sangat banyak dan jumlahnya melebihi 10% dari total data pada *dataset*, maka Penulis memutuskan untuk menghapus atribut *SkinThickness* dan *Insulin* pada *dataset* tersebut.

Dari pembersihan data yang sudah dilakukan sebelumnya, atribut-atribut yang akan digunakan dijabarkan pada tabel 2 dibawah ini.

Tabel 2. Tabel *Dataset* Setelah Dibersihkan

No	Nama Atribut	Keterangan
1	<i>Pregnancies</i>	Jumlah Kehamilan
2	<i>Glucose</i>	Kadar Glukosa 2 Jam Setelah Makan
3	<i>Blood Pressure</i>	Tekanan Darah
4	<i>BMI</i>	Berat Badan
5	<i>DPF</i>	Peluang Mengalami Diabetes Berdasarkan Gen
6	<i>Age</i>	Umur
7	<i>Class</i>	Target Variabel

Diskritisasi atribut bertujuan untuk mempermudah pengelompokan nilai berdasarkan kriteria yang telah ditetapkan (Lesmana, 2012). Selain itu juga, diskritisasi atribut juga berfungsi untuk meningkatkan akurasi hasil klasifikasi yang akan dilakukan.

Diskritisasi atribut dimulai dari atribut *Pregnancies* yang dibagi menjadi 3 kelas, yaitu Rendah (0, 1), Sedang (2-5), dan Tinggi (>6). Atribut *Glucose* dibagi menjadi 3 kelas, yaitu Rendah (<95), Sedang (95-140), dan Tinggi (>140). Atribut *BloodPressure* dibagi menjadi 3 kelas, yaitu Normal (<80), Normal ke Tinggi (80-90), dan Tinggi (>90). Atribut *BMI* dibagi menjadi 4 kelas, yaitu Rendah (<24,9), Normal (25-29,9), Obesitas (30-34,9), dan Obesitas Akut (>35). Atribut *DPF* dibagi menjadi 2 kelas, yaitu Rendah (<0,5275) dan Tinggi (>0,5275). Atribut *Age* dibagi menjadi 3 kelas, yaitu Muda (<40), Dewasa (40-59), dan Tua (>60). Sedangkan atribut *Class* dibagi menjadi 2, yaitu Positif (1), dan Negatif (0) (Lesmana, 2012). Penjabaran berupa tabel dari diskritisasi atribut dapat dilihat pada tabel 3 dibawah ini.

Tabel 3. Tabel Diskritisasi Atribut

Atribut	Diskritisasi	
<i>Pregnancies</i>	Rendah	(0,1)
	Sedang	(2-5)
	Tinggi	(>6)
<i>Glucose</i>	Rendah	(<95)
	Sedang	(95-140)
	Tinggi	(>140)
<i>BloodPressure</i>	Normal	(<80)
	Normal ke Tinggi	(80-90)
	Tinggi	(>90)
<i>BMI</i>	Rendah	(<24.9)
	Normal	(25-29.9)
	Obesitas	(30-34.9)
	Obesitas Akut	(>35)
<i>DPF</i>	Rendah	(< 0.5275)
	Tinggi	(> 0.5275)

Age	Muda	(<40)
	Dewasa	(40-59)
	Tua	(>60)
Class	Positif	(1)
	Negatif	(0)

2.5 Tahap Pelatihan Data

Penelitian ini menggunakan *dataset Pima Indians Diabetes* yang sudah dibersihkan dari nilai 0 pada setiap atribut yang awalnya berjumlah 768 data menjadi 725 data. Data tersebut kemudian akan dibagi menjadi data *training* dan data *test* sebesar 75% dari data *training* dan 25% dari data *test*. Selanjutnya data tersebut akan masuk ke dalam tahap pelatihan menggunakan algoritma ID3 dan SVM. Tahap pelatihan pada algoritma ID3 menggunakan perhitungan manual dan implementasi pada *software*, sedangkan algoritma SVM diimplementasikan langsung menggunakan *software*. Pelatihan data menggunakan algoritma ID3 memiliki tahapan-tahapan sebagai berikut:

1. Langkah pertama adalah hitung nilai entropi secara keseluruhan.
2. Setelah itu, hitung nilai entropi untuk setiap atribut yang digunakan pada *dataset* tersebut.
3. Setelah itu, hitung nilai *information gain* untuk masing-masing atribut.
4. Setelah menghitung nilai *information gain*, kumpulkan semua nilai *information gain* pada setiap atribut pada sebuah tabel. Nilai *information gain* yang tertinggi akan dijadikan *root node*/node akar. Dalam penelitian ini, atribut *Glucose* memiliki nilai *information gain* tertinggi.

Tabel 4. Tabel *Information Gain*

Atribut	<i>Information Gain</i>
<i>Pregnancies</i>	0,008
<i>Glucose</i>	0,039
<i>Blood Pressure</i>	0,002
<i>BMI</i>	0,026
<i>DPF</i>	0,006
<i>Age</i>	0,009

Dalam penelitian ini, tahapan pelatihan algoritma ID3 juga bisa langsung diimplementasikan menggunakan *software*. *Decision Tree* yang dihasilkan oleh *software* juga mengacu pada *dataset* yang digunakan.

Dalam pelatihan data menggunakan algoritma SVM langsung diimplementasikan pada *software*. *Kernel SVM* yang dipakai dalam penelitian ini adalah *kernel Linear* sebagai *kernel* yang umum digunakan untuk melakukan klasifikasi menggunakan algoritma SVM, serta *kernel RBF/Gaussian* sebagai contoh pembandingan pada kernel SVM Linear.

Tahapan-tahapan yang harus dilakukan ketika melakukan klasifikasi algoritma SVM *Linear* dan *RBF* adalah:

1. Masukkan *dataset* yang akan diolah. Dalam penelitian ini menggunakan *dataset Pima Indian Diabetes* yang sudah dibersihkan sebelumnya.
2. Pisahkan *Feature Value* dan *Target Value*. *Target Value* memiliki nilai 0 dan 1 pada kolom atributnya. Sehingga atribut *Class* dijadikan sebagai *Target Value*.
3. Bagi *dataset* tersebut menjadi 2 bagian, yaitu *Train Data* yang memuat 75% jumlah data keseluruhan, dan *Test Data* yang memuat 25% jumlah data keseluruhan.
4. Lakukan *Feature Scaling* pada *software* untuk menormalisasikan *dataset* yang digunakan. *Dataset* yang tidak ternormalisasikan akan berakibat penilaian kerja menjadi buruk.
5. Setelah semua langkah-langkah sebelumnya terpenuhi, maka proses eksekusi atau pelatihan menggunakan algoritma SVM bisa dilaksanakan. Buat sintaks eksekusi algoritma SVM *Linear* dan SVM *RBF* di baris kode yang terpisah.
6. Jika proses eksekusinya berjalan dengan baik, maka akan didapat sebuah tabel *Confusion Matrix*. *Confusion Matrix* inilah yang bisa memberikan hasil penilaian kerja yang diperlukan, yaitu *Accuracy*, *Precision*, *Recall*, *F1 Score*, nilai ROC/AUC, dan waktu eksekusi algoritma.

3. HASIL DAN PEMBAHASAN

Eksekusi algoritma ID3 maupun SVM menggunakan *software* sama-sama menghasilkan *Confusion Matrix* yang dapat dijadikan acuan dalam proses penilaian kerja yang meliputi *Accuracy*, *Precision*, *Recall*, *F1 Score*, nilai ROC/AUC, dan waktu eksekusi algoritma. Dari hasil klasifikasi algoritma ID3 menggunakan *software* dan bahasa pemrograman Python, *Confusion Matrix*-nya ditunjukkan oleh tabel 5 sebagai berikut:

Tabel 5. Tabel *Confusion Matrix* Hasil Klasifikasi ID3

Hasil Pengujian ID3	Gangguan Diabetes	
	Negatif	Positif
Negatif	109	6
Positif	42	25

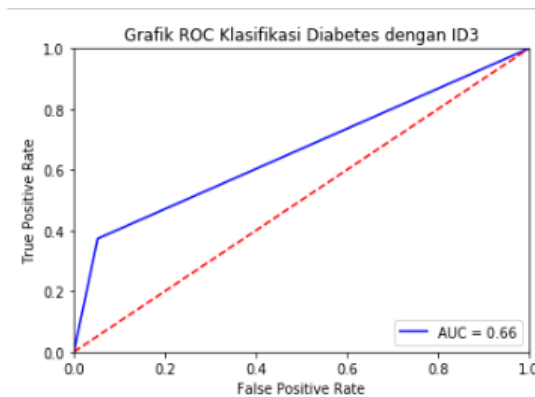
Adapun perhitungan manual dari nilai *Accuracy*, *Recall*, *Precision* dan *F1 Score* adalah:

Tabel 6. Tabel *Accuracy*, *Precision*, *Recall*, Dan *F1 Score* Hasil Klasifikasi ID3

<i>Accuracy</i>	$Accuracy = (109+25) / (109+6+42+25) = 74\%$
<i>Precision</i>	$Precision = 25 / (25+6) = 81\%$
<i>Recall</i>	$Recall = 25 / (25+42) = 37\%$
<i>F1 Score</i>	$F1\ Score = 2*(37\%*81\%) / (37\%+81\%) = 51\%$

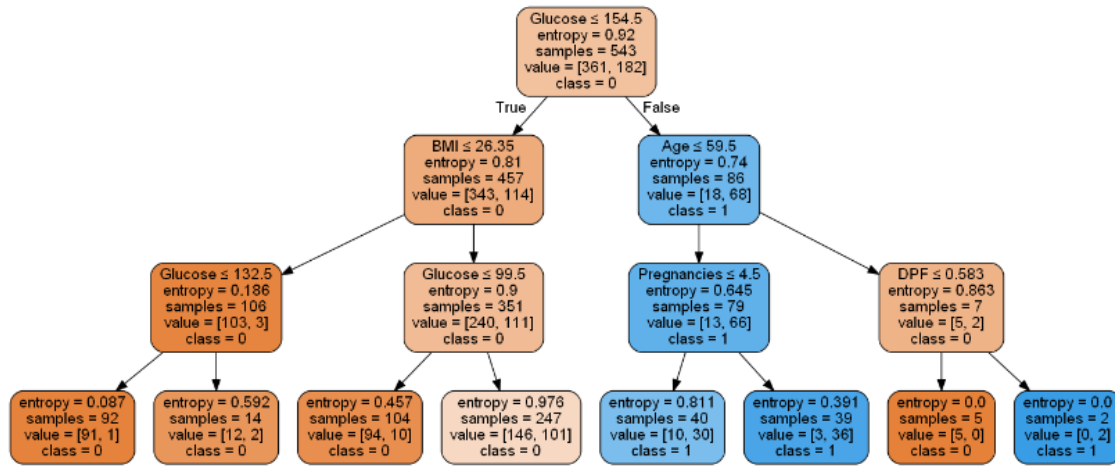
Berdasarkan pada tabel 5, terlihat 109 pasien dikenali dengan benar tidak memiliki gangguan diabetes melitus tipe 2. Sebanyak 6 pasien yang salah sebenarnya tidak mengidap diabetes melitus tipe 2. Sebanyak 42 pasien yang salah sebenarnya mengidap penyakit diabetes melitus tipe 2. Sisanya, sebanyak 25 pasien dikenali dengan benar mengidap penyakit diabetes melitus tipe 2.

Berdasarkan tabel 6, tingkat akurasi yang dihasilkan oleh algoritma ID3 mencapai 74%. Sedangkan tingkat *Precision* menghasilkan nilai sebanyak 81%, *Recall* sebanyak 37% dan *F1 Score* sebanyak 51%. Grafik ROC/AUC yang dihasilkan sebesar 66% dan masuk ke dalam kategori “*Poor Classification*”.



Gambar 2. Grafik ROC Hasil Klasifikasi ID3

Decision Tree yang dihasilkan oleh algoritma ID3 dengan menggunakan bantuan *software* digambarkan pada gambar 3 dibawah ini.



Gambar 3. Tampilan *Decision Tree* dari Hasil Klasifikasi ID3

Dari hasil klasifikasi algoritma SVM *Linear* menggunakan *software* dan bahasa pemrograman Python, *Confusion Matrix*-nya ditunjukkan oleh table 7 sebagai berikut:

Tabel 7. Tabel *Confusion Matrix* Hasil Klasifikasi SVM *Linear*

Hasil Pengujian SVM <i>Linear</i>	Gangguan Diabetes	
	Negatif	Positif
Negatif	105	10
Positif	29	38

Sedangkan *Confusion Matrix* yang dihasilkan oleh algoritma SVM RBF ditunjukkan pada table 8.

Tabel 8. Tabel *Confusion Matrix* Hasil Klasifikasi SVM RBF

Hasil Pengujian SVM RBF	Gangguan Diabetes	
	Negatif	Positif
Negatif	106	9
Positif	31	36

Adapun perhitungan manual dari nilai *Accuracy* dengan rumus (7), *Recall* dengan rumus (8), *Precision* dengan rumus (9), dan *F1 Score* menggunakan rumus (10) adalah:

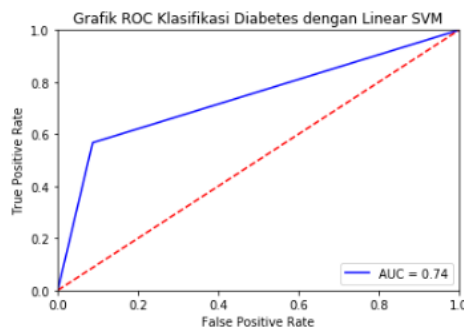
Tabel 9. Tabel *Accuracy*, *Precision*, *Recall*, Dan *F1 Score* Hasil Klasifikasi SVM *Linear* dan RBF

	SVM <i>Linear</i>	SVM RBF
<i>Accuracy</i>	78,5%	78%
<i>Precision</i>	79%	80%
<i>Recall</i>	56%	54%
<i>F1 Score</i>	66%	64%

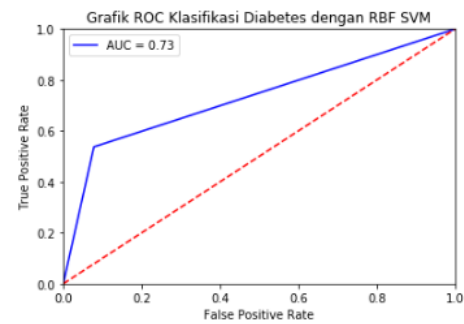
Berdasarkan pada tabel 7, terlihat 105 pasien dikenali dengan benar tidak memiliki gangguan diabetes melitus tipe 2. Sebanyak 10 pasien yang salah sebenarnya tidak mengidap diabetes melitus tipe 2. Sebanyak 29 pasien yang salah sebenarnya mengidap penyakit diabetes melitus tipe 2. Sisanya, sebanyak 38 pasien dikenali dengan benar mengidap penyakit diabetes melitus tipe 2.

Sedangkan pada tabel 8, 106 pasien memang tidak mengidap diabetes melitus tipe 2. Sebanyak 9 yang tadinya memiliki gangguan diabetes melitus tipe 2, sebenarnya mereka tidak memiliki gangguan tersebut. Kemudian, 31 pasien diprediksi tidak ada gangguan diabetes, sebenarnya mereka terkena diabetes melitus tipe 2. Terakhir, 36 pasien terdeteksi mengidap penyakit diabetes melitus tipe 2.

Berdasarkan tabel 9, tingkat akurasi yang dihasilkan oleh algoritma SVM *Linear* mencapai 78,5%. Sedangkan tingkat *Precision* menghasilkan nilai sebanyak 79%, *Recall* sebanyak 56% dan *F1 Score* sebanyak 66%. Sedangkan algoritma SVM RBF memiliki akurasi sebesar 78%, sedikit lebih kecil daripada SVM *Linear*. *Precision* yang dihasilkan oleh algoritma SVM RBF ternyata lebih besar dari SVM *Linear*, yaitu 80%. *Recall* yang dihasilkan sebesar 54%, dan *F1 Score*-nya sebesar 64%. Jadi, bisa dipastikan algoritma SVM *Linear* sedikit lebih unggul daripada SVM RBF. Grafik ROC/AUC yang dihasilkan oleh algoritma SVM *Linear* adalah 74% dan masuk kategori “*Fair Classification*”. Sedangkan SVM RBF mendapat nilai 73% dari grafik ROC/AUC dan masuk kategori “*Fair Classification*”.



Gambar 4. Grafik ROC pada Algoritma SVM *Linear*



Gambar 5. Grafik ROC pada Algoritma SVM *RBF*

Adapun tabel perbandingan hasil klasifikasi Diabetes Melitus Tipe 2 menggunakan *dataset Pima Indians Diabetes* pada algoritma ID3 dan SVM. Klasifikasi ID3 dan SVM yang sudah dilakukan menggunakan 75% *data train* dan 25% *data test*.

Tabel 10. Tabel Evaluasi Hasil Klasifikasi ID3 dan SVM

Algoritma	Akurasi	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>	ROC / AUC
ID3	74%	81%	37%	51%	66%
SVM <i>Linear</i>	78,5%	79%	56%	66%	74%
SVM RBF	78%	80%	54%	64%	73%

Tabel 11. Tabel Evaluasi Waktu Eksekusi Rata-Rata ID3 dan SVM

Algoritma	Percobaan Waktu Eksekusi					Rata-Rata
	I	II	III	IV	V	
ID3	6,98 ms	6,95 ms	7,94 ms	7,95 ms	6,98 ms	7,35 ms
SVM Linear	112 ms	105 ms	127 ms	111 ms	101 ms	111,2 ms
SVM RBF	20,9 ms	18 ms	18,9 ms	22,9 ms	20 ms	15,63 ms

Berdasarkan tabel perbandingan diatas, nilai akurasi tertinggi dimiliki oleh algoritma SVM *Linear* dengan nilai akurasi 78,5%, disusul oleh SVM RBF dengan akurasi 78% dan terakhir oleh ID3 dengan akurasi 66%. Nilai *Precision* tertinggi jatuh pada algoritma ID3 dengan nilai 81%, disusul dengan SVM RBF sebanyak 80% dan terakhir oleh SVM *Linear* dengan nilai 79%. Untuk *Recall*, SVM *Linear* memperoleh nilai tertinggi dengan nilai 56%, disusul dengan SVM RBF sebanyak 54% dan ID3 di posisi terakhir dengan nilai 37%.

Sementara itu, untuk *F1 Score* tertinggi dipegang oleh SVM *Linear* dengan nilai 66%, disusul dengan SVM RBF dengan nilai 64% dan terakhir ID3 sebanyak 51%. Untuk nilai *ROC/AUC* tertinggi dipegang oleh SVM *Linear* dengan nilai 74%, SVM RBF di posisi kedua dengan nilai 73%, dan ID3 di urutan terakhir dengan nilai 66%. Terakhir, untuk waktu eksekusi tercepat jatuh pada ID3 dengan waktu 7,35 ms, disusul dengan SVM RBF dengan waktu 15,63 ms dan terakhir SVM *Linear* dengan nilai 111,2 ms. Hasil ini berdasarkan tabel 4.8 diatas yang menampilkan nilai rata-rata waktu eksekusi dengan percobaan sebanyak 5 kali.

Dari penjabaran tabel diatas, SVM *Linear* unggul di 4 kriteria, yaitu pada Akurasi, *Recall*, *F1 Score*, dan *ROC/AUC*. Algoritma ID3 unggul di 2 kriteria, yaitu pada *Precision* dan waktu eksekusi algoritma. Sedangkan SVM RBF menduduki peringkat 2 di semua kriteria yang digunakan untuk proses klasifikasi penyakit Diabetes Melitus Tipe 2.

Hasil klasifikasi pada algoritma ID3 dan SVM bukan merupakan acuan utama dalam penanganan penyakit diabetes melitus tipe 2. Hasil klasifikasi ini hanya sebagai identifikasi awal dalam membantu para *user*/pengguna memprediksi apakah dia terkena penyakit diabetes melitus tipe 2 atau tidak. Jika didiagnosis terkena penyakit diabetes melitus tipe 2 oleh sistem hasil klasifikasi diatas, maka pasien harus segera menemui dokter atau pakar yang menangani penyakit diabetes melitus tipe 2 agar mendapat penanganan lebih lanjut.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penerapan *data mining* pada *dataset* penderita Diabetes Melitus Tipe 2 menggunakan algoritma *Iterative Dichotomiser 3* (ID3) dan *Support Vector Machine* (SVM), maka dapat disimpulkan:

1. Algoritma ID3 dan SVM dapat diimplementasikan pada proses klasifikasi penderita diabetes melitus tipe 2 dengan nilai akurasi sebesar 70% keatas. Penelitian ini menggunakan *dataset Pima Indians Diabetes* yang semula berjumlah 768 data dan setelah dibersihkan menjadi 725 data dengan atribut *Pregnancies*, *Glucose*, *BloodPressure*, *BMI*, *DPF*, *Age* sebagai *Feature Value* dan *Class* sebagai *Target Value*. Penelitian ini menggunakan algoritma ID3 dan SVM *linear* dan RBF dengan perbandingan *data train* sebesar 75% dan *data test* sebesar 25%. Metode penilaian kerja yang digunakan adalah *Accuracy*, *Recall*, *Precision*, *F1 Score* yang berpatokan pada *Confusion Matrix*, dan grafik *ROC/AUC*.
2. Pada proses klasifikasi algoritma menggunakan ID3 dan SVM *linear* dan RBF dengan menggunakan 543 *data train* dan 182 *data test*, hasil klasifikasi paling baik diperoleh SVM *linear* dengan akurasi 78,5%, disusul oleh algoritma SVM RBF dengan akurasi 78% dan hasil klasifikasi yang terendah oleh ID3 dengan akurasi 74%. Dari hasil evaluasi klasifikasi pada algoritma ID3, SVM *linear* dan SVM RBF melalui proses *testing data*, algoritma klasifikasi terbaik jatuh pada algoritma SVM *Linear*. Algoritma SVM *Linear* memperoleh nilai *True Negative*=109, *False Negative*=9, *False Positive*=42, dan *True Positive*=25 melalui perhitungan *Confusion Matrix*. Dari hasil *Confusion Matrix* tersebut diperoleh nilai akurasi sebesar 78,5%, *Precision* sebesar 79%, *Recall* sebesar 56% dan *F1 Score* sebesar 66%. Hasil dari penggunaan algoritma SVM *Linear* sudah cukup baik karena nilai *Accuracy*, *Precision*, *Recall* dan *F1 Score* berada di atas 50%. Lalu pada grafik *ROC*, nilai *AUC* yang didapat sebesar 74% dan masuk ke dalam kelompok "*Fair Classification*". Waktu eksekusinya sebesar 111,2 ms.

Adapun saran yang dapat diberikan untuk peningkatan hasil penelitian ini adalah:

1. Sebaiknya memperbanyak jumlah *dataset* yang akan digunakan agar hasil klasifikasi yang didapat lebih bagus dan akurat.

2. Diharapkan menggunakan *dataset* yang jumlah nilai yang hilang pada setiap atribut berjumlah lebih sedikit dan jika memungkinkan tidak ada toleransi bagi *dataset* yang masih memiliki nilai 0 atau NaN (*Not as Number*) pada setiap atribut.
3. Untuk pengembangan sistem kedepannya diharapkan dapat memasukkan *dataset* dari berbagai sumber, misalnya rumah sakit tertentu agar hasil klasifikasinya bisa lebih baik lagi dan hasilnya bisa lebih bervariasi lagi.
4. Dalam penelitian berikutnya diharapkan dapat menghasilkan sebuah sistem informasi berdasarkan hasil klasifikasi yang sudah dilakukan. Sistem informasi ini dapat memuat hasil klasifikasi yang lebih dinamis sehingga *user*/pengguna dapat menyesuaikan *dataset*, *parameter*, serta poin diskritisasi yang akan digunakan terhadap sistem informasi tersebut.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak Rumah Sakit Sanglah atas ketersediaan waktunya untuk melakukan wawancara terbuka untuk mendapatkan wawasan baru tentang diabetes melitus tipe 2. Tidak lupa juga para akademisi yang telah banyak membantu dalam penyusunan jurnal ini sehingga jurnal ini diharapkan dapat berkontribusi besar bagi masyarakat luas dalam identifikasi awal menggunakan algoritma ID3 dan SVM.

DAFTAR PUSTAKA

- Efendi, M., & Subhan Wibawa, H. A. (2018). *Prediksi Penyakit Diabetes Menggunakan Algoritma ID3 dengan Pemilihan Atribut Terbaik*.
- Lesmana, I. P. D. (2012). *PERBANDINGAN KINERJA DECISION TREE J48 DAN ID3 DALAM PENGKLASIFIKASIAN DIAGNOSIS PENYAKIT DIABETES MELLITUS*.
- Mutmainah, P. A. M. (2017). *DECISION TREE MENGGUNAKAN ALGORITMA ID3 UNTUK MELAKUKAN DETEKSI PENYAKIT KANKER PAYUDARA*.
- Puspitasari, A. M., Ratnawati, D. E., & Widodo, A. W. (2018). *Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine*.
- Putri, A. S. (2018). *Tahun 2018 Penderita Diabetes di Indonesia Meningkat*. <https://www.fimela.com/lifestyle-relationship/read/3739252/tahun-2018-penderita-diabetes-di-indonesia-meningkat>
- Rahmay, M. F., Darmawidjadja, M. I., & Alamsah, D. (2017). *KLASIFIKASI UNTUK DIAGNOSA DIABETES MENGGUNAKAN METODE BAYESIAN REGULARIZATION NEURAL NETWORK (RBNN)*.
- Utami, N. W., Sukajaya, I. N., Candiasa, I. M., & Dewi, E. G. A. (2019). The Implementation of Data Mining to Show UKT (Students' Tuition) Using Fuzzy C-Means Algorithm : (Case Study: Universitas Pendidikan Ganesha). In *The Implementation of Data Mining to Show UKT (Students' Tuition) Using Fuzzy C-Means Algorithm : (Case Study: Universitas Pendidikan Ganesha): Vol. pp. 101-10*.
- Widiyati, D. K., Wati, M., & Pakpahan, H. S. (2018). *Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara*.