

## Klasifikasi Diabetes Mellitus Menggunakan Support Vector Machine (Studi Kasus: Puskesmas Modopuro, Mojokerto)

Andharini Dwi Cahyani<sup>1\*)</sup>, Ari Basuki<sup>2)</sup>

<sup>1)</sup> Teknik Informatika, Universitas Trunojoyo Madura

<sup>2)</sup> Teknik Industri, Universitas Trunojoyo Madura

<sup>1\*)</sup> email: andharini.cahyani@trunojoyo.ac.id

DOI: <https://doi.org/10.21107/rekayasa.v12i2.19763>

### ABSTRAK

Diabetes Mellitus (DM) merupakan salah satu penyakit dimana tubuh seorang pasien tidak dapat mengontrol tingkat konsentrasi gula (glukosa). Ada beberapa indikator kondisi tubuh pasien yang menunjukkan adanya penyakit diabetes. Pada penelitian ini dilakukan pengelompokan pasien DM dengan menggunakan metode Machine Learning. Algoritma yang digunakan adalah Support Vector Machine (SVM) yang merupakan metode klasifikasi supervised learning. SVM bekerja dengan membagi dataset menjadi 2 bagian dengan menggunakan garis linear (hyperplane) yang optimal. Target dari klasifikasi digunakan untuk memprediksi apakah pasien dengan indikator kondisi tubuh tertentu merupakan pasien DM yang terdeteksi atau tidak. Dari data yang diperoleh dari Puskesmas Modopuro, terdapat beberapa missing value. Pada penelitian ini, missing value diatasi dengan cara melakukan imputasi. Ada 3 kernel yang digunakan pada metode SVM, yaitu linear, polynomial, dan sigmoid. Pengukuran kinerja 3 jenis kernel tersebut akan menggunakan 5 subset cross validation. Nilai akurasi cross validation terbaik yang didapatkan menggunakan kernel linear, polynomial, dan sigmoid masing-masing yaitu 62%, 64%, dan 54%. Berdasarkan hasil pengamatan kernel polynomial mendapatkan hasil akurasi yang lebih baik dari kernel lainnya.

**Kata kunci:** Diabetes Mellitus, Klasifikasi, Support Vector Machine

## Diabetes Mellitus Classification Using Support Vector Machine (Case Study: Puskesmas Modopuro, Mojokerto)

### ABSTRACT

*Diabetes Mellitus (DM) is a disease in which a patient's body cannot control the level of sugar (glucose) concentration. There are several indicators of the patient's body condition that indicate the presence of diabetes. In this study, DM patients were grouped using the Machine Learning method. The algorithm used is Support Vector Machine (SVM) which is a supervised learning classification method. SVM works by dividing the dataset into 2 parts using an optimal linear (hyperplane) line. The target of the classification is used to predict whether patients with certain indicators of body condition are detected DM patients or not. From the data obtained from the Modopuro Health Center, there are several missing values. In this study, missing values were overcome by imputation. There are 3 kernels used in the SVM method, namely linear, polynomial, and sigmoid. Measuring the performance of the 3 types of kernels will use 5 subsets of cross validation. The best cross validation accuracy values were obtained using linear, polynomial, and sigmoid kernels, respectively, namely 62%, 64%, and 54%. Based on the observations, the polynomial kernel has better accuracy than other kernels.*

*Keywords: Diabetes Mellitus, Classification, Support Vector Machine*

### PENDAHULUAN

Klasifikasi merupakan sebuah topik yang terdapat pada *data mining* dan *machine learning*. Klasifikasi sendiri yakni pengelompokan data menjadi beberapa kelas label atau target yang sudah ditentukan. Ada dua cara atau kategori yang digunakan untuk menyelesaikan permasalahan klasifikasi ini yaitu yang pertama *supervised learning*

(pembelajaran yang diawasi) dimana sebuah data telah memiliki sebuah label, *supervised learning* biasa dipakai pada proses klasifikasi dan juga regresi, contohnya kasus *object recognition*, *predictive analysis* dan juga *sentiment analysis*. Yang kedua yaitu *unsupervised learning* (pembelajaran tidak diawasi) dimana data yang digunakan berbeda dengan data dari *supervised learning* untuk

*unsupervised learning* data yang digunakan tidak memiliki label. *Unsupervised learning* biasa

digunakan untuk *clustering*, asosiasi, dan *dimensionality reduction* (Purwanto dkk., 2018).

Pada masa sekarang ini dengan pesatnya perkembangan teknologi, peralatan optik, dan komputer berdampak pada pesatnya perkembangan ilmu taksonomi. Demikian pula sejalan dengan perkembangan ilmu biologi molekuler sangat mempengaruhi sistem klasifikasi modern. Metode-metode yang biasa digunakan dalam menyelesaikan klasifikasi yakni *Decision Tree*, KNN (*K-Nearest Neighbor*), *Naive Bayes*, *Neural Network*, dan masih banyak lagi metode klasifikasi yang bisa digunakan

Diabetes Mellitus (DM) merupakan masalah penyakit di mana tubuh pasien tidak dapat mengontrol tingkat konsentrasi glukosa pada darahnya. Penderita tidak dapat menghasilkan insulin pada jumlah yang cukup sehingga kadar gula dalam darah cukup tinggi. Jika kondisi ini berlanjut terus-menerus, maka akan menimbulkan bahaya bagi tubuh. DM merupakan penyakit degeneratif yang memerlukan penanganan yang tepat dan serius karena efeknya seperti penyakit serius lainnya (penyakit jantung, stroke, gagal ginjal, kerusakan pada sistem sensorik dan lain-lain) (Delima dkk., 2016). Berdasarkan data yang didapatkan di wilayah kerja Puskesmas Modopuro yang berada di Kecamatan Mojosari Kabupaten Mojokerto, saat ini masih banyak ditemukan masyarakat yang mengalami penyakit diabetes mellitus. Tujuan dari penelitian ini adalah untuk mengetahui diagnosa pasien penyakit diabetes mellitus di Puskesmas Modopuro. Dari hasil klasifikasi ini memprediksi apakah pasien diabetes mellitus tanpa suntik insulin mengalami komplikasi atau tidak. Dalam sistem prediksi Diabetes Mellitus ini terdapat 8 fitur yang digunakan dan sebuah label target, adapun fitur-fitur tersebut yakni jenis kelamin, umur, lama sakit, lingkaran perut, sistole, diastole, dan detak nadi (Purwanti dan Maghfirah, 2016).

Untuk memecahkan klasifikasi, berbagai jenis algoritma pada *machine learning* dapat diterapkan dalam penelitian ini, salah satunya adalah metode *Support Vector Machine* (SVM). (Singh dkk., 2016). SVM adalah perhitungan

dengan strategi pembelajaran yang diatur yang menyelidiki informasi dan memahami data dan pola. Keuntungan dari perhitungan SVM adalah sangat baik dapat digunakan untuk masalah pengelompokan dan kekambuhan dengan bit langsung atau potongan tidak lurus yang dapat menjadi kemampuan perhitungan pembelajaran untuk karakterisasi dan kekambuhan dengan presisi tinggi (Nayak *et al.*, 2015: 170).

### Algoritma SVM

SVM merupakan algoritma klasifikasi pada machine learning. Dalam metode SVM, proses klasifikasi dilakukan dengan pendekatan *supervised learning* untuk mencari nilai hyperplane yang memiliki margin paling optimal. SVM banyak diminati karena memiliki algoritma yang yang dapat menghasilkan akurasi yang bagus dengan daya komputasi yang sedikit.

Sesuai dengan penelitian Aulia dkk. (2015), perhitungan SVM adalah teknik klasik terbaik untuk klasifikasi. Teknik SVM dilakukan dengan memanfaatkan aturan *Underlying Gamble Minimization*, untuk mencari hyperplane terbaik dengan mengisolasi 2 kelas dan mencari garis tepi untuk vektor bantuan dan batas hyperplane. Teknik SVM mempartisi dataset menjadi 2 kelas. Bagian atas garis yang diisolasi oleh hyperplane bernilai 1, sedangkan kelas yang berbeda bernilai - 1.

Karakteristik SVM secara garis besar adalah sebagai berikut:

1. SVM merupakan linear classifier yang membagi dataset menjadi 2 bagian secara linear dengan menggunakan garis *hyperplane* (Carrera dkk., 2017).
2. SVM mengenali pola data dengan melakukan transformasi dimensi ruang input ke dimensi ruang yang lebih tinggi untuk melakukan optimasi. Pada metode pengenalan pola pada umumnya, optimasi dilakukan pada dimensi ruang yang lebih rendah daripada dimensi ruang input.
3. Menerapkan strategi *Structural Risk Minimization* (SRM) yang mencari

hyperplane terbaik yang memiliki nilai margin paling optimal.

Rumus Perhitungan SVM:

$$x_i W + b \geq 1 \text{ ketika } y_i = 1 \quad (1)$$

$$x_i W + b \leq -1 \text{ ketika } y_i = -1 \quad (2)$$

Keterangan:

$x_i$  = data ke -i

$W$  = nilai bobot support vector yang tegak lurus dengan hyperplane

$b$  = nilai bias

$Y_i$  = kelas data ke -i

Perhitungan menggunakan algoritma SVM pada data check up pasien diabetes melitus yang tidak melakukan suntik insulin memiliki *feature* ( $w_1, w_2, w_3, w_4 \dots n$ ) Formulasi yang akan digunakan adalah sebagai berikut :

$$\frac{1}{2} \|w\|^2 = \frac{1}{2} (w_1^2 + w_2^2 + w_3^2 + \dots w_n^2) \text{ dengan } (w^1, x^1 + b) \quad (3)$$

Keterangan:

$i$  merupakan indeks data ke 1,2,3,4...N

$y_i (w_1 x_1 + w_2 x_2 \dots n) = +1$  untuk yang tidak terdeteksi diabetes mellitus

$y_i (w_1 x_1 + w_2 x_2 \dots n) = -1$  untuk yang terdeteksi diabetes mellitus

## METODE PENELITIAN

### Data/Parameter yang digunakan

Dalam penelitian ini terdapat 8 fitur yang digunakan dan sebuah label target, adapun fitur-fitur tersebut yakni jenis kelamin, umur, lama sakit, lingkar perut, sistole, diastole, dan detak nadi. Berikut ini penjelasan masing – masing atribut data penelitian.

#### 1. Jenis Kelamin (L/P)

Jenis kelamin pada pria memiliki risiko diabetes yang lebih tinggi. Perbedaan risiko ini dipengaruhi oleh rasio sirkulasi otot dan lemak. Pada pria, pengumpulan lemak terkonsentrasi di sekitar bagian tengah tubuh, sehingga memicu peningkatan berat badan yang lebih berisiko memicu masalah metabolisme (Srikartika dkk., 2016).

#### 2. Umur (tahun)

Dalam keadaan umum manusia akan mengalami perubahan keadaan terhadap tubuh maupun jiwa yang cenderung menurun setelah

memasuki usia 40 tahun ke atas. Diabetes biasanya menampilkan sebuah gejala saat memasuki usia yang memiliki daya tahan tubuh yang mulai menurun secara signifikan yakni pada usia sekitar 45 tahun (Isnaini dan Ratnasari, 2018). Terdapat sebuah teori yang mengatakan bahwa seseorang yang memiliki usia menginjak 45 tahun memiliki risiko yang lebih besar untuk terkena diabetes apalagi ditambah dengan kelebihan berat badan.

#### 3. Lama Sakit (hari)

Lama sakit ini merupakan salah satu faktor penting, dikarenakan banyak warga yang menganggap bahwa sakit tersebut merupakan sakit biasa tanpa melakukan pengecekan secara berkala (Lathifah, 2017). Dari catatan yang ada rata-rata penderita mengalami sakit paling kecil 2 hari, dengan tambahan gejala.

#### 4. Lingkar Perut (cm)

Salah satu penyakit yang dikaitkan dengan perut buncit adalah diabetes, terutama diabetes tipe 2. Lingkar perut menjadi indikator diabetes yang lebih baik dibandingkan indeks massa tubuh (IMT). Ini karena indeks massa tubuh bisa dipengaruhi oleh massa otot, yang tidak ada hubungannya sama sekali dengan diabetes (Verma dan Hussain, 2017). Lemak visceral yang menumpuk di rongga perut secara signifikan bisa meningkatkan risiko diabetes tipe 2 karena dapat mempengaruhi sensitivitas insulin.

#### 5. Hasil IMT (ideal/ lebih/ kurang/ gemuk/ sangat gemuk)

IMT (Indeks Massa Tubuh) adalah cara untuk mengukur berat badan seseorang, apakah tipikal atau tidak. IMT adalah angka yang menunjukkan rasio antara berat badan (dalam kilogram) dan tinggi badan (dalam meter). Orang dengan IMT antara 25 dan 29,9 dikatakan kelebihan berat badan, sedangkan mereka yang memiliki IMT lebih dari 30 dikatakan obesitas. Skor IMT yang tinggi menimbulkan risiko diabetes. Diabetes terkait erat dengan obesitas, seperti yang ditunjukkan beberapa penelitian. Pada penderita diabetes, umumnya pankreas masih mampu menghasilkan insulin yang cukup untuk menjaga kadar gula darah pada tingkat normal, tetapi tubuh tidak mampu mengenalinya dengan baik sehingga penanganan glukosa tidak bisa dilakukan dengan baik.

#### 6. Sistole (mm)

Diabetes Melitus umumnya berasal dari gangguan metabolisme dalam tubuh karena

adanya kenaikan gula darah. Kenaikan ini terjadi karena adanya gangguan fungsi insulin maupun penurunan fungsi pankreas dalam mengeluarkan insulin (Setiyorini dkk., 2018). Dalam hasil penelitian Bouthoorn dkk. (2018), ditemukan banyak penderita diabetes mellitus yang juga rata-rata memiliki tekanan sistole yang tinggi. Sehingga Sistole juga menjadi salah satu faktor dari penyakit diabetes melitus, sistole ini memiliki hubungan yang cukup tinggi dengan hipertensi dan hipertensi juga masuk dalam faktor resiko penyakit diabetes melitus (Handayani dkk., 2018).

7. Diastole (Hg)

Hipertensi juga merupakan salah satu indikator diabetes mellitus yang sering dijumpai. Pada kondisi pasien yang memiliki hipertensi, biasanya aktivitas saraf simpatis memiliki terjadi abnormalitas lipid yang berpengaruh terhadap rangsangan dari resistensi tubuh terhadap insulin. (Sihombing, 2017). Sehingga, diastole juga merupakan salah satu indikator dari hipertensi dan juga diabetes melitus. Hal ini dikarenakan *diastole* menunjukkan tingkat hipertensi pada pasien diabetes melitus (Mokolomban, 2018).

8. Denyut Jantung (/menit)

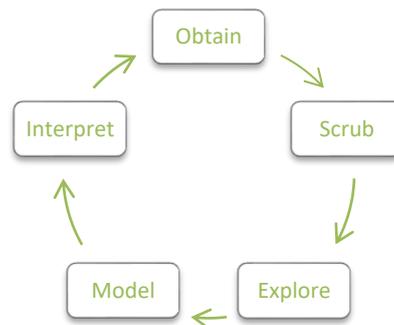
Menurut Aquarista (2017), salah satu faktor penyebab diabetes adalah detak jantung yang relatif cepat. Denyut nadi bervariasi dari orang ke orang dan dapat dipengaruhi oleh berbagai faktor. Faktor-faktor tersebut antara lain usia, aktivitas fisik, tingkat kebugaran, suhu, emosi, posisi, tinggi badan, dan penggunaan obat-obatan tertentu. Namun, denyut nadi orang dewasa biasanya berkisar antara 60 hingga 100 denyut per menit.

**Pengumpulan Data**

Data yang digunakan adalah data check up penyakit diabetes dari Puskesmas Modopuro Kabupaten Mojokerto yang diperoleh dari hasil tes laboratorium. Puskesmas Modopuro adalah salah satu pusat kesehatan di kabupaten Mojokerto yang menerima pengecekan laboratorium pasien diabetes. Dataset tersebut memiliki 1163 record data dalam periode 1 tahun 4 bulan (2 Januari 2021 - 20 April 2022). Namun pada data tersebut terdapat banyak missing value, sehingga data yang digunakan pada proses klasifikasi tersebut sebanyak 700 record data yang memiliki nilai lengkap.

**Alur Penelitian**

Tahapan pada penelitian ini menggunakan metode OSEMNI yang banyak digunakan dalam ilmu sains data untuk menyusun model data yang sedang diteliti (Nantasemat, 2018). Tahapan pada metode OSEMNI ini adalah *cycle* yang terdiri dari 5 tahap, yaitu: Obtain, Scrub, Explore, Model, dan iNterpret. Berikut ini adalah diagram metode OSEMNI (Gambar 1).



Gambar 1. Metode OSEMNI

Tahap Obtain

Pada tahap Obtain, peneliti mengumpulkan data – data yang akan diteliti dengan memastikan tipe data yang sesuai dan sumber data yang terpercaya. Pada tahap ini kami menyiapkan surat-surat yang dibutuhkan untuk melakukan pengambilan data diabetes di puskesmas modopuro. Setelah surat-surat tersebut terkumpul kami dapat datang ke puskesmas Modopuro untuk melakukan pengambilan data diabetes.

Tahap Scrub

Data yang telah berhasil dikumpulkan pada tahap ini nantinya akan dilanjutkan ke tahap pembersihan data. Seluruh kumpulan data akan diperiksa oleh peneliti. Mulai dari format data, keseragaman, jenis, atribut, bahkan anomali data pada dataset yang kita miliki. telah dikumpulkan. Peneliti menggunakan proses Scrub untuk menangani data yang hilang, memperbaiki nilai yang tidak valid, menghilangkan duplikat, mengatur data, dan menggunakan rekayasa fitur.

Tahap Explore

Pindah ke tahap eksplorasi atau eksplorasi. Peneliti diberi tugas untuk melihat pola data yang disediakan pada tahap kedua, juga dikenal sebagai data scrubbing, dalam fase ini. Temuan eksplorasi pola harus dilihat oleh peneliti dari perspektif pengetahuan domain atau ilmu bisnis. Hal ini dilakukan agar peneliti dapat

menggunakan pola yang telah mereka lihat untuk menghasilkan ide atau solusi yang akurat dan bermanfaat. Di antara langkah-langkah yang dilakukan peneliti selama proses Eksplorasi adalah memeriksa korelasi dan memeriksa distribusi nilai untuk setiap kolom.

**Tahap Model**

Pada tahap ini, spesialis melakukan siklus yang menampilkan informasi atau mendemonstrasikan informasi tentang konsekuensi dari contoh investigasi yang telah diselesaikan pada tahap sebelumnya, khususnya Investigate. Hanya data yang berkaitan dengan prediksi hasil akhir yang digunakan selama proses. Regresi dan prediksi dapat digunakan untuk meramalkan, klasifikasi dapat digunakan untuk mengidentifikasi, dan pengelompokan dapat digunakan untuk mengelompokkan atau mengklasifikasikan data yang diamati. Kami menggunakan tiga kernel yang berbeda untuk mengklasifikasikan klasifikasi SVM dalam penelitian ini yaitu kernel linier, polinomial, dan sigmoid (Jung, 2018).

**Tahap Interpret**

Pada tahap ini, peneliti menginterpretasikan model yang sudah dibuat. Peneliti menganalisa data dengan menggunakan model yang sudah dibuat. Hasil analisa ini kemudian dijadikan kesimpulan tentang karakteristik data yang telah diolah sesuai dengan metode yang digunakan dalam penelitian.

**Skenario Ujicoba**

Untuk mengurangi bias, maka dalam penelitian ini uji coba dilakukan dengan menggunakan 5-fold cross validation. Metode k-fold cross validation merupakan metode yang umumnya digunakan untuk mengukur kinerja algoritma klasifikasi dengan membagi data menjadi k subset data. Salah satu subset data digunakan sebagai data testing yang digunakan untuk menguji model, sedangkan subset data lainnya digunakan pada tahap pembuatan model (data training). Pembagian data pada 5-fold cross validation yang digunakan dalam penelitian ini tampak pada Gambar 2.

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
			Data Pengujian	
			Data Pelatihan	

Gambar 2. Metode 5-Fold Cross Validation

Setelah menguji model, maka selanjutnya dilakukan perhitungan kinerja model atau algoritma klasifikasi yang dilakukan. Kinerja suatu model / algoritma bisa dilihat dari tingkat akurasi, presisi, recall, dan f-measure yang diperoleh dari *confusion matrix* (Fernanda dkk., 2017). Berikut ini langkah-langkah yang dilakukan untuk menyusun confusion matrix

1. Masukan hasil pengujian dari klasifikasi yang telah diperoleh dalam sebuah confusion matrix seperti pada tabel di bawah ini.

Tabel 1. Confusion Matrix

		Kelas hasil prediksi		
		Ya	Tidak	Jml
Kelas Aktual	Ya	TP	FN	P
	Tidak	FP	TN	N

Keterangan:

- True Positive (TP) : jumlah kasus pasien diabetes yang tidak melakukan suntik insulin dengan komplikasi yang diklasifikasikan sebagai pasien diabetes yang tidak melakukan suntik insulin dengan komplikasi.
- False Positive (FP) : jumlah kasus pasien diabetes yang tidak melakukan suntik insulin tanpa komplikasi yang diklasifikasikan sebagai pasien diabetes yang tidak melakukan suntik insulin dengan komplikasi.
- True Negative (TN) : jumlah kasus pasien diabetes yang tidak melakukan suntik insulin tanpa komplikasi yang diklasifikasikan sebagai pasien diabetes yang tidak melakukan suntik insulin tanpa komplikasi.

- False Negative (FN) : jumlah kasus pasien diabetes yang tidak melakukan suntik insulin dengan komplikasi yang diklasifikasikan sebagai pasien diabetes yang tidak melakukan suntik insulin tanpa komplikasi.
- Positive (P) : jumlah kasus pasien diabetes yang tidak melakukan suntik insulin dengan komplikasi.
- Negative (N) : jumlah kasus pasien diabetes yang tidak melakukan suntik insulin tanpa komplikasi.

2. Hitung nilai akurasi menggunakan persamaan 4 berikut ini.

$$Accuracy = \frac{TP+TN}{P+N} \times 100\% \tag{4}$$

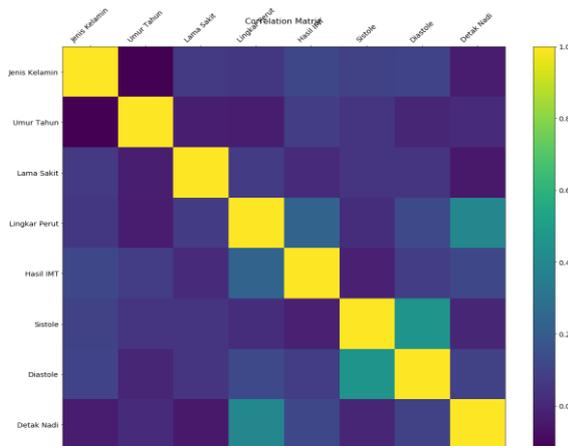
**HASIL DAN PEMBAHASAN**

**HASIL**

Pada penelitian ini, kami tertarik untuk melihat indikator kondisi tubuh yang saling berkorelasi erat. Karena itu, kami melakukan analisa korelasi antar variabel yang ditunjukkan pada Tabel 2 tentang nilai Korelasi tiap indikator dan Gambar 3 tentang visualisasi heatmap dari data yang diperoleh.

Tabel 2. Nilai Korelasi

	Jenis Kelamin	Umur Tahun	Lama Sakit	Lingkar Perut	Hasil IMT	Sistole	Diastole	Detak Nadi
Jenis Kelamin	1.000000	-0.123056	0.088625	0.059019	0.119652	0.095016	0.100864	-0.035890
Umur Tahun	-0.123056	1.000000	-0.027164	-0.033129	0.080845	0.051639	-0.000792	0.013084
Lama Sakit	0.088625	-0.027164	1.000000	0.070077	0.014759	0.046294	0.051726	-0.045400
Lingkar Perut	0.059019	-0.033129	0.070077	1.000000	0.236119	0.024246	0.125428	0.395029
Hasil IMT	0.119652	0.080845	0.014759	0.236119	1.000000	-0.021526	0.080617	0.120021
Sistole	0.095016	0.051639	0.046294	0.024246	-0.021526	1.000000	0.460295	-0.003725
Diastole	0.100864	-0.000792	0.051726	0.125428	0.080617	0.460295	1.000000	0.093974
Detak Nadi	-0.035890	0.013084	-0.045400	0.395029	0.120021	-0.003725	0.093974	1.000000



Gambar 3. Heatmap dari visualisasi data dengan 8 parameter

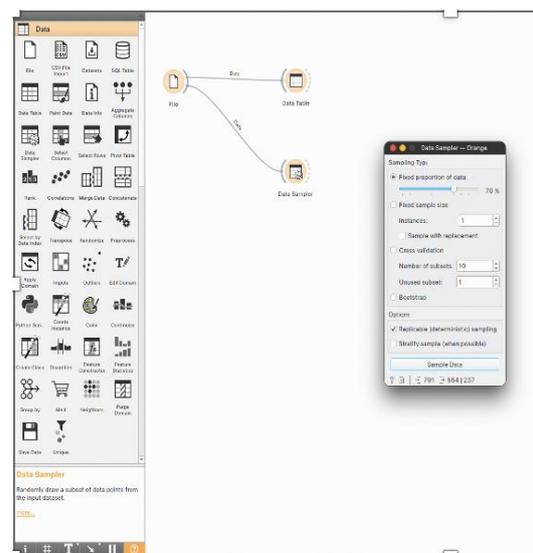
Kekuatan korelasi antara dua variabel dinyatakan dengan koefisien korelasi, dalam hal

ini dimisalkan dengan huruf k yang memiliki rentang nilai  $1 < k < -1$ .

- Jika dua variabel berkorelasi erat, maka nilai koefisien mendekati nilai 1 dan -1. Nilai koefisien yang mendekati nilai 1 menunjukkan adanya korelasi positif, dan sebaliknya, jika mendekati nilai -1, maka menunjukkan adanya korelasi negatif.
- Jika dua variabel memiliki hubungan korelasi yang rendah, maka nilai koefisien mendekati angka 0.
- Jika nilai koefisien korelasi = 0, maka kedua variabel tidak memiliki hubungan korelasi.

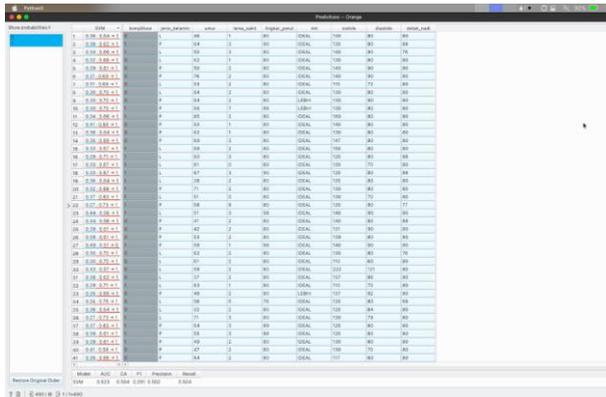
Dari gambar diatas ini didapatkan bahwa diantara fitur yang memiliki korelasi tertinggi adalah sistole dengan diastole dengan tingkat korelasi 0.46 dan juga detak nadi dengan lingkar perut dengan tingkat korelasi sebesar 0.39.

Selanjutnya, untuk mendapatkan hasil yang diperoleh dari SVM, pertama buka aplikasi orange terlebih dahulu. Lalu masukkan data, dan tarik ke data table untuk melihat data. Kemudian tarik ke data sampler untuk membuat sample data dan set data sample sebanyak 80%. Untuk menjalankan metode SVM pada data sample, tarik ke fitur SVM yang terdapat pada side bar orange. Dari SVM dan data sampel, tarik ke prediction.



Gambar 4. Set data sample yang digunakan sebagai data training

Hasil yang diperoleh dari SVM adalah sebagai berikut



Gambar 5. Hasil SVM menggunakan tools Orange

**PEMBAHASAN**

Dalam data Diabetes Melitus (DM) memiliki 700 record data yang clean. Dalam data ini memiliki rata-rata umur dari penderita Diabetes Melitus dengan komplikasi adalah 58 tahun dengan usia tertinggi 85 tahun dan usia terendah 20 tahun, sedangkan pada penderita penyakit Diabetes Melitus tanpa komplikasi diderita dengan rata-rata usia 56 tahun dengan usia tertinggi 85 tahun dan usia terendah 32 tahun. Rata-rata pasien memiliki Riwayat lama sakit sekitar 2 hari dengan lama sakit terlama hingga 30 hari dan paling cepat 1 hari saja.

Dataset yang sudah bersih dilakukan pengecekan informasi tentang dataset menggunakan pandas profiling. Dapat disimpulkan bahwa dataset tersebut memiliki 10 kolom dan 700 baris. Terdapat 7 kolom dengan nilai numerik dan 3 kolom dengan nilai kategorik. Tidak terdapat missing value dan baris duplikat dalam dataset tersebut. Kolom jenis\_kelamin memiliki nilai P berjumlah 446 dan L berjumlah 254. Dalam dataset tersebut pasien perempuan lebih banyak daripada pasien laki-laki. Kolom umur memiliki nilai berjumlah 49 dengan rentang 20 sampai 85. Rata-rata nilai dari kolom umur adalah 57,78. Kolom lama\_sakit memiliki nilai berjumlah 13 dengan rentang 0 sampai 20. Rata-rata nilai dari kolom lama\_sakit adalah 2.292857143.

Kolom lingkar\_perut memiliki nilai berjumlah 28 dengan rentang 57 sampai 135. Rata-rata nilai dari kolom lingkar\_perut adalah 82.28. Kolom imt memiliki nilai IDEAL berjumlah 628, LEBIH berjumlah 62, GEMUK berjumlah 5, dan KURANG berjumlah 5. Dalam dataset tersebut pasien dengan hasil imt ideal lebih banyak

daripada pasien dengan hasil imt lainnya. Kolom sistole memiliki nilai berjumlah 65 dengan rentang 87 sampai 222. Rata-rata nilai dari kolom sistole adalah 130.5771429. Kolom sistole memiliki nilai berjumlah 40 dengan rentang 40 sampai 256. Rata-rata nilai dari kolom sistole adalah 79.85428571. Kolom detak\_nadi memiliki nilai berjumlah 12 dengan rentang 58 sampai 90. Rata-rata nilai dari kolom detak\_nadi adalah 80.49714286. Kolom komplikasi memiliki nilai YA berjumlah 366 dan TIDAK berjumlah 334. Dalam dataset tersebut pasien terdeteksi DM lebih banyak daripada pasien yang tidak terdeteksi DM.

Tabel 3. SVM dengan Kernel Linear

	Linear				
	1	2	3	4	5
fit_time	0.01794887	0.01370645	0.01157856	0.0110476	0.01132464
score_time	0.00547099	0.00573993	0.00565886	0.00525928	0.00516653
test_accuracy	0.53571429	0.50714286	0.62142857	0.55714286	0.54285714
test_f1	0.50381679	0.48888889	0.61313869	0.56338028	0.53623188
test_precision	0.57894737	0.53225806	0.65625	0.57971014	0.56923077
test_recall	0.44594595	0.45205479	0.57534247	0.54794521	0.50684932

Tabel 4. SVM dengan Kernel Polynomial

	Polynomial				
	1	2	3	4	5
fit_time	0,0301886	0,02020621	0,02167058	0,02034712	0,02297711
score_time	0,0082991	0,00816154	0,00830603	0,0083158	0,00859189
test_accuracy	0,5357143	0,49285714	0,64285714	0,55	0,54285714
test_f1	0,5185185	0,4964539	0,66666667	0,61349693	0,53623188
test_precision	0,5737705	0,51470588	0,64935065	0,55555556	0,56923077
test_recall	0,472973	0,47945205	0,68493151	0,68493151	0,50684932

Tabel 5. SVM dengan Kernel Sigmoid

	Sigmoid				
	1	2	3	4	5
fit_time	0,0437582	0,04032564	0,04067445	0,04536104	0,04009557
score_time	0,0130155	0,01273966	0,01334882	0,01663566	0,01464438
test_accuracy	0,5428571	0,54285714	0,52142857	0,48571429	0,49285714
test_f1	0,5555556	0,46666667	0,68544601	0,64705882	0,44094488
test_precision	0,5714286	0,59574468	0,52142857	0,50381679	0,51851852
test_recall	0,5405405	0,38356164	1.	0,90410959	0,38356164

Kinerja dari model klasifikasi SVM dengan kernel linear memiliki kinerja terbaik pada subset 3. Nilai akurasi dari model klasifikasi SVM dengan kernel linear pada subset 3 adalah 62%. Kinerja dari model klasifikasi SVM dengan kernel polynomial memiliki kinerja terbaik pada subset 3. Nilai akurasi dari model klasifikasi SVM dengan kernel polynomial pada subset 3 adalah 64%. Kinerja dari model klasifikasi SVM dengan kernel sigmoid memiliki kinerja terbaik pada subset 1 dan 2. Nilai akurasi dari model klasifikasi SVM dengan kernel linear pada subset 1 dan 2 adalah 54%. Sehingga dapat

disimpulkan model tersebut memiliki kinerja yang kurang baik.

## KESIMPULAN DAN SARAN

Kasus penyakit diabetes di Indonesia merupakan salah satu yang terbesar di dunia. Berdasarkan data yang diperoleh dari laboratorium Puskesmas Modopuro, jumlah pasien di Kecamatan Mojosari Kabupaten Mojokerto ini yang mengalami komplikasi diabetes cukup tinggi. Tujuan dari penelitian ini adalah untuk mengetahui diagnosa pasien penyakit diabetes melitus di Puskesmas Modopuro. Dari hasil klasifikasi ini memprediksi apakah pasien diabetes mellitus tanpa suntik insulin mengalami komplikasi atau tidak.

Dari eksplorasi dan analisa data diatas yang dilakukan untuk menjawab pertanyaan penelitian, maka ada beberapa kesimpulan yang dapat diambil, yaitu:

1. Pada penelitian ini dilakukan pengolahan data dengan metode SVM. Dengan menggunakan tool Orange, metode SVM yang diimplementasikan mampu melakukan klasifikasi penyakit DM dengan menggunakan 8 indikator. 2 Indikator yang saling berkorelasi erat dari 8 indikator yang ada adalah sistole dengan diastole dengan tingkat korelasi 0.46 dan juga detak nadi dengan lingkar perut dengan tingkat korelasi sebesar 0.39.
2. Ujicoba yang dilakukan menggunakan beberapa kernel (linear, polinomial, dan sigmoid). Hasilnya menunjukkan bahwa kernel polinomial yang memiliki akurasi tertinggi dibandingkan dengan dua metode lainnya.
3. .

## DAFTAR PUSTAKA

Alfian, R. (2015). Korelasi Antara Kepatuhan Minum Obat dengan Kadar Gula Darah pada Pasien Diabetes Melitus Rawat Jalan di RSUD Dr. H. Moch. Ansari Saleh Banjarmasin. *Jurnal Pharmascience*, 2(2), 15-23.

Aquarista, N. C. (2017). Perbedaan karakteristik penderita diabetes melitus tipe 2 dengan dan tanpa penyakit jantung koroner. *Jurnal Berkala Epidemiologi*, 5(1), 37-47.

Aulia, S., Hadiyoso, S., & Ramadan, D. N. (2015). Analisis Perbandingan KNN dengan SVM untuk Klasifikasi Penyakit Diabetes Retinopati berdasarkan Citra Eksudat dan Mikroaneurisma. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, 3(1), 75.

Bouthoorn, S., Valstar, G. B., Gohar, A., den Ruijter, H. M., Reitsma, H. B., Hoes, A. W., & Rutten, F. H. (2018). The prevalence of left ventricular diastolic dysfunction and heart failure with preserved ejection fraction in men and women with type 2 diabetes: A systematic review and meta-analysis. *Diabetes and Vascular Disease Research*, 15(6), 477-493.

Carrera, E. V., González, A., & Carrera, R. (2017, August). Automated detection of diabetic retinopathy using SVM. In *2017 IEEE XXIV international conference on electronics, electrical engineering and computing (INTERCON)* (pp. 1-4). IEEE.

Delima, D., Mihardja, L. K., & Ghani, L. (2016). Faktor risiko dominan penderita stroke di Indonesia. *Indonesian Bulletin of Health Research*, 44(1), 20146.

Fernanda, S. I., Ratnawati, D. E., & Adikara, P. P. (2017). Identifikasi Penyakit Diabetes Mellitus Menggunakan Metode Modified K-Nearest Neighbor (MKNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X*.

Handayani, F., Bintang, A. K., & Kaelan, C. (2018). Hubungan Hipertensi, Diabetes Mellitus dan Dislipidemia dengan Luaran Klinis Pasien Iskemik Stroke dengan Hipersomnia. *Healthy Tadulako Journal (Jurnal Kesehatan Tadulako)*, 4(1), 1-6.

Isnaini, N., & Ratnasari, R. (2018). Faktor risiko mempengaruhi kejadian diabetes mellitus tipe dua. *Jurnal Kebidanan Dan Keperawatan Aisyiyah*, 14(1), 59-68.

Jung, Y. (2018). Multiple predicting K-fold cross-validation for model selection. *Journal of Nonparametric Statistics*, 30(1), 197-215.

Lathifah, N. L. (2017). Hubungan durasi penyakit dan kadar gula darah dengan keluhan subyektif penderita diabetes melitus. *Jurnal berkala epidemiologi*, 5(2), 231-239.

Mokolomban, C. (2018). Kepatuhan Minum Obat Pada Pasien Diabetes Melitus Tipe 2

- Disertai Hipertensi Dengan Menggunakan Metode MMAS-8. *Pharmacon*, 7(4).
- Purwanto, A., & Darmadi, E. A. (2018). Perbandingan Minat Siswa Smu Pada Metode Klasifikasi Menggunakan 5 Algoritma. *ikraith-informatika*, 2(1), 43-47.
- Purwanti, L. E., & Maghfirah, S. (2016). Faktor risiko komplikasi kronis (kaki diabetik) dalam diabetes mellitus tipe 2. *The Indonesian Journal of Health Science*, 7(1).
- Setiyorini, E., Wulandari, N. A., & Efyuwinta, A. (2018). Hubungan kadar gula darah dengan tekanan darah pada lansia penderita Diabetes Tipe 2. *Jurnal Ners Dan Kebidanan (Journal of Ners and Midwifery)*, 5(2), 163-171.
- Singh, A., Thakur, N., & Sharma, A. (2016, March). A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 1310-1315). IEEE.
- Srikartika, V. M., Cahya, A. D., & Hardiati, R. S. W. (2016). Analisis faktor yang memengaruhi kepatuhan penggunaan obat pasien diabetes melitus tipe 2. *Jurnal Manajemen dan Pelayanan Farmasi*, 6(3), 205-212.
- Verma, S., & Hussain, M. E. (2017). Obesity and diabetes: an update. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 11(1), 73-79.