



# Evaluasi Performa Support Vector Machine Classifier Terhadap Penyakit Mental

Mhd Furqan\*, Rakhmat Kurniawan, Kiki Iranda HP

<sup>a</sup>Departemen Ilmu Komputer, Universitas Islam Negeri Sumatera Utara

Naskah Diterima : 25 Oktober 2020; Diterima Publikasi : 3 Desember 2020  
DOI: 10.21456/vol10iss2pp203-210

## Abstract

Expression of genes found in the brains of autism, bipolar, and schizophrenia patients identified as overlapping. The overlap is a state in which the values of genes are similar. This paper aims to determine the best performance of support vector machines algorithm in classifying autism, bipolar, and schizophrenia based on the expression of genes using genome-wide association studies data. Using three support vector machine kernels, this study evaluates the performance of gaussian, laplacian, and sigmoid for genome-wide association studies datasets. The datasets were obtained from Psychiatric Genomics Consortium publications, where 660 data were taken with each disorder consisting of 220 data. This study proposes an optimal kernel for one-against-one and one-against-all multiclass support vector machine, and the performance is evaluated using accuracy. The study results show that the Gaussian kernel has the best accuracy performance compared to other support vector machines kernels in classifying genome-wide association studies data of autism, bipolar, and schizophrenia as early diagnosis.

**Keywords:** Autism; Bipolar; Classification; Schizophrenia; Support Vector Machine

## Abstrak

Eksresi gen yang ditemukan pada otak penderita autisme, bipolar, dan skizofrenia diidentifikasi sebagai tumpang tindih. Tumpang tindih adalah keadaan di mana nilai-nilai gen memiliki kemiripan. Tujuan penelitian ini menentukan kinerja terbaik algoritma *support vector machine* dalam mengklasifikasikan autisme, bipolar, dan skizofrenia berdasarkan ekspresi gen menggunakan data *Genome-Wide Association Studies*. Dengan menggunakan tiga kernel *support vector machine*, penelitian ini mengevaluasi kinerja gaussian, laplacian, dan sigmoid dari kumpulan data *Genome-Wide Association Studies*. Kumpulan data ini diperoleh dari publikasi *Psychiatric Genomics Consortium*, di mana sebanyak 660 data diambil dengan setiap gangguan yang terdiri dari sebanyak 220 data. Penelitian ini mengusulkan kernel optimal untuk *support vector machine* menggunakan *one-against-one* dan *one-against-all*, kemudian kinerjanya dievaluasi menggunakan akurasi. Hasil penelitian menunjukkan bahwa kernel gaussian memiliki kinerja akurasi terbaik dibandingkan dengan mesin vektor pendukung lainnya kernel dalam mengklasifikasikan data *Genome-Wide Association Studies* dari penyakit autisme, bipolar, dan skizofrenia sebagai diagnosis dini.

**Kata kunci:** Autisme; Bipolar; Klasifikasi; Skizofrenia; *Support Vector Machine*

## 1. Pendahuluan

Autisme, bipolar, dan skizofrenia, secara umum, adalah tiga jenis penyakit mental yang diberikan risiko sebesar 80% dari genetik atau turun-temurun, dan sebesar 20% dipengaruhi oleh lingkungan (Carroll & Owen, 2009). Adanya penyakit mental terkait dengan gangguan dalam fungsi otak sehingga ada perbedaan cara berpikir, merasakan sesuatu, dan cara perilaku yang ditunjukkan oleh penderita (Novianty & Rochman Hadjam, 2017). Oleh karena itu, gangguan kesehatan mental tidak bisa dianggap remeh. Menurut WHO, ada sekitar 450 juta orang yang mengalami gangguan mental di berbagai wilayah di dunia, yang sebagian besar berada di India (World Health Organization, 2013). Sementara di

Indonesia, ada sekitar 9.162.886 kasus atau sebesar 3,7% dari data sumber kependudukan dan terus meningkat menjadi 6% hingga saat ini (Kementerian Kesehatan Republik Indonesia, 2018). Menurut penelitian ilmiah yang dilakukan oleh Geschwind, menghasilkan bahwa ekspresi gen yang ditemukan pada otak pasien autisme, bipolar, dan skizofrenia diidentifikasi mengalami tumpang tindih. Tumpang tindih merupakan kondisi di mana nilai-nilai dari gen terdapat yang serupa satu sama lain (Geschwind *et al.*, 2018).

Autisme, gangguan bipolar serta skizofrenia secara diagnostiknya memiliki kecenderungan bersifat heterogen yang disertai batas kesewenangan gimana tergantung oleh gejala gejala selanjutnya yang dirasakan oleh pasien (Carroll & Owen, 2009). Agar

\*) Penulis korespondensi: [mfurqan@uinsu.ac.id](mailto:mfurqan@uinsu.ac.id)

dapat diketahui risiko genetik yang diwariskan sejak dini bisa diupayakan dengan memeriksa struktur DNA yang memanfaatkan penggunaan data meta analisis *Genome-Wide Association Studies* (GWAS) (Bush & Moore, 2012). GWAS adalah data hasil pindaian Polimorfisme Nukleotida Tunggal pada keseluruhan genom dari banyak manusia sebagai upaya pencarian penanda genetik yang dapat dimanfaatkan dalam melakukan prediksi adanya suatu penyakit atau gangguan (Walker, 2013). Data meta-analisis dilakukan pengumpulan agar mesin dapat mempelajarinya untuk mendapatkan klasifikasi yang dihasilkan dari tiap-tiap nilai dalam data. Pelaksanaan klasifikasi dengan memanfaatkan penggunaan mesin pembelajaran dari sekumpulan data agar melakukan proses pelatihan yang mana data tersebut didapatkan melalui publikasi dari *Psychiatric Genomics Consortium* (PGC). PGC adalah hasil *experimental* paling besar dari banyak ahli psikiater di dunia melalui pengumpulan data yang didasari oleh gen seseorang sebagai upaya kemajuan penemuan perspektif genetik pada biologis maupun klinis (PGC, 2019).

Pada penelitian ini diperlukan suatu metode dalam mengolah data pelatihan melalui pemanfaatan algoritma pengklasifikasi yakni *support vector machine* (SVM). Penggunaan SVM dalam penelitian ini disebabkan kemampuan yang dimilikinya telah memiliki pembuktian dalam penelitian pada bidang bioinformatika lain yang turut terkait pada bidang pada penelitian ini. Pada teoritisnya, algoritma ini melakukan pemeriksaan terhadap nilai kromosom dari sampel, lalu ditemukan *hyperplane* paling baik agar data penyakit mental yang telah dikumpulkan dapat diklasifikasikan (Affymetrix, 2006).

Sampai saat ini, penelitian yang sudah ada masih belum mengevaluasi kinerja kernel SVM untuk ketiga gangguan mental ini secara bersamaan (Grove *et al.*, 2019). Tiga kernel SVM yang dievaluasi dalam penelitian ini adalah gaussian, laplacian, dan sigmoid. Agar dapat diketahui seberapa optimalnya SVM melalui penggunaan tiga jenis kernel pada tahapan pembelajaran untuk mengklasifikasikan autisme, bipolar serta skizofrenia dengan data GWAS sebagai upaya mengatasi permasalahan tumpang tindih, penelitian ini dilakukan untuk mengevaluasi kinerja pengklasifikasian dengan menggunakan akurasi. Informasi dari akurasi tersebut akan menentukan kelayakan SVM sebagai pengklasifikasi dari tiga jenis penyakit mental yang mengalami *overlapping* untuk upaya diagnosis dini sebagai bagian dari aplikasi di bidang bioinformatika.

## 2. Kerangka Teori

### 2.1. Penyakit Mental

Penyakit mental adalah gangguan yang mempengaruhi kontrol kognitif, emosional, dan perilaku, yang secara substantif merusak kemampuan

seseorang untuk melakukan sesuatu (Choresyo *et al.*, 2015). Penyakit mental pada umumnya bermula semenjak permulaan dalam kehidupan manusia, yang melibatkan adanya hubungan antar pengaruh genetik dan non-genetik. Pengaruh dari jenis kelamin bisa saja memungkinkan dalam sejumlah kasus yang terjadi, yakni laki-laki memiliki risiko mengidap autisme, hiperaktif, dan penggunaan obat-obat terlarang yang lebih besar. Sebaliknya, perempuan lebih memungkinkan dengan kecenderungan mengalami keputusasaan, gangguan makan maupun gangguan kecemasan (Hyman *et al.*, 2006).

Kelainan gen dan morfologi otak yang terkait dengan autisme, bipolar, dan skizofrenia masih terus diidentifikasi untuk menentukan gen yang menjadi penyebab terjadinya gangguan ini semakin rentan dan memberi pengaruh klinis dalam berkontribusi untuk meningkatkan diagnostik serta terapi yang dapat dilakukan (Hyman, 2000). Penyakit mental bukan saja menyebabkan kesusahan untuk penderitanya namun turut bagi keluarga karena perbedaan yang dimiliki, terutama adanya faktor memerlukan biaya yang lebih dikarenakan pengobatan yang tinggi, akses yang tidak sama, kekurangan dukungan secara sosial, pandangan, serta diskriminasi, dan hak-hak individu yang mengidap penyakit mental masih tidak diberi seoptimal mungkin dalam lingkungan bermasyarakat (Hyman *et al.*, 2006).

### 2.2. Autism

Autisme dari kata “*auto*” yang memiliki arti “sendiri” karena orang dengan autis merasakan bahwa ia memiliki kehidupan pada dunia mereka sendiri. Autisme bermula dengan diperkenalkan oleh Leo Kanner pada tahun 1943, meskipun gangguan ini pada dasarnya telah ada sejak lama (Aprilia *et al.*, 2014). Penyebab autisme berdasarkan banyak penelitian, mengungkapkan bahwa autisme telah ada jauh sebelum bayi lahir. Menurut Almandil, secara genetik menemukan keluarga dengan 2,5-3% saudara kandung autis, di mana ada 50-100 kali lebih mungkin daripada populasi umum. Ini menunjukkan bahwa ada risiko genetik yang memiliki peran dalam gangguan autis. Awalnya, diagnosis autisme didasarkan pada karakteristik gejala yang dilakukan anak-anak ketika mereka mulai merasa sulit melakukan komunikasi, sosialisasi, serta pertumbuhan yang mengalami keterlambatan. Banyaknya ahli menyarankan bahwa jika penyebab autisme yang paling mungkin adalah risiko warisan genetik (Almandil *et al.*, 2019). Namun, hingga saat ini, belum diketahui kromosom mana yang membawa sifat autisme. Dengan demikian diperlukan pembelajaran terhadap sejumlah data kromosom yang dimiliki orang pengidap autisme. GWAS paling banyak yang dilakukan pemindaian terhadap pengidap autisme sampai sekarang kurang lebih yakni 7.387 kasus, serta pengidentifikasian terhadap

satu varian dalam genom secara keseluruhan (Grove *et al.*, 2019).

### 2.3. Bipolar

Gangguan bipolar adalah gangguan mental yang memiliki sifat kronik atau episode, dimana terjadi sesekali atau tidak teratur (Mattes, 2010). Gangguan bipolar memberikan kemungkinan bagi individu berperasaan secara berlebihan yang meliputi dua periode yaitu manik dan depresi. Periode manik adalah kondisi di mana pasien memiliki semangat yang menggebu-gebu, merasakan kesenangan yang berlebihan saat mendapatkan kebahagiaan, dan bicara yang terlalu cepat. Sementara dalam periode depresi, pengidap bipolar dapat merasakan sangat malang, sensitif, mudah mendapatkan keputusan, serta bersikap tidak peduli dengan lingkungan (Reig, 2013). Menurut penelitian, seseorang yang mempunyai gangguan mental bipolar memiliki paling tidak satu anggota keluarganya yang juga mengidap gangguan yang sama. Faktor genetik menyebabkan 60-80% gangguan bipolar, sementara faktor lainnya yang menjadi penyebab dari gangguan ini yaitu faktor dari lingkungan (Choresyo *et al.*, 2015). Struktur otak yang dimiliki oleh orang-orang dengan gangguan bipolar berbeda dari orang lainnya secara umum, ada perbedaan dalam struktur neurotransmitter ataupun senyawa pada tubuh yang bertanggung jawab dalam penyampaian informasi antara neuron menuju sel saraf yang ditargetkan (Mattes, 2010). Gangguan bipolar dapat dilihat saat individu telah memiliki kemampuan pemikiran, oleh karena itu belum dapat didiagnosis melalui gejala ketika masih bayi, karena mereka belum dapat mengekspresikan perasaan (National Institute of Mental Health, 2018). Diagnosis dini dibutuhkan sebagai upaya penentuan terapi yang benar dalam mengendalikan gangguan ini dikarenakan jika tidak ada perawatan secara tepat dan benar dapat memberikan peningkatan risiko kecenderungan bunuh diri pada pasien yang mengidap bipolar (Reig, 2013).

### 2.4. Skizofrenia

Skizofrenia merupakan jenis penyakit mental akut yang memiliki tanda-tanda penderitanya seringkali berhalusinasi, tingkah laku yang tidak normal disertai kepribadian tunggal, dan delusi (Choresyo *et al.*, 2015). Gangguan skizofrenia memiliki penyebab dikarenakan kelainan struktur otak dan sistem saraf pusat. Penelitian yang diterbitkan dalam jurnal *Biology Psychiatry* mengatakan bahwa gen menyebabkan sebesar 80% faktor risiko skizofrenia (Hyman, 2000). Orang dengan riwayat keluarga skizofrenia akan rentan terhadap gangguan ini. Penderita gangguan skizofrenia hendaknya melakukan konsultasi pada dokter dengan teratur serta minum obat pada dosis maupun jarak waktu yang telah disesuaikan dengan rekomendasi dokter (Grove *et al.*, 2019). Gejala yang jelas pada

umumnya muncul ketika berusia 15-25 tahun, sedangkan gejala permulaan dapat dilihat yakni mudah malu, menjauhkan dirinya dari lingkungan, serta kurang ramah. Hal ini menyebabkan respon dari lingkungan yang kurang baik dan dapat menjadikan kesehatan mental dari pengidap skizofrenia semakin parah (Purcell *et al.*, 2009).

### 2.5. Genome Wide Association-Studies

Genetika manusia bertujuan sebagai pengidentifikasian dari faktor risiko genetik yang dapat diwarisi seperti penyakit mental termasuk autisme, gangguan bipolar, dan skizofrenia (Hyman, 2000). Terdapat sejumlah penelitian dan analisis yang digunakan dalam pengidentifikasian faktor risiko genetik. Salah satunya adalah *Genome-Wide Association Studies* (GWAS), yang melakukan pengukuran dan analisis varian dalam pengurutan DNA semua genom seseorang untuk mengidentifikasikan faktor risiko genetik dari penyakit pada populasi (Bush & Moore, 2012). Tujuan inti dari GWAS yakni penggunaan faktor risiko genetik sebagai prediksi terhadap seseorang yang memiliki risiko terkena penyakit dan pengidentifikasian secara mendasar terhadap biologis dari kerentanan penyakit yang dapat menjadi upaya untuk mengembangkan strategi pencegahan baru dan perawatan baru (Rietveld *et al.*, 2013). Mengetahui dasar biologis faktor genetik dapat memainkan peranan yang esensial pada pengembangan terapi farmakologis paling baru saat ini (Maciukiewicz *et al.*, 2017).

### 2.5. Support Vector Machine

Algoritma SVM melakukan pembelajaran dengan memberikan masukan data yang dilatih untuk menentukan jenis label pada objek (Vapnik, 1998). SVM juga telah banyak berhasil diterapkan pada berbagai aplikasi biologis untuk memprediksi kategori dari suatu penyakit (Evgeniou & Pontil, 2001). SVM memiliki keuntungan menentukan jarak antara dua set data dari dua kelas yang berbeda. SVM dapat menentukan *hyperplane* yang memisahkan secara optimal dan maksimum sebagai solusi dalam permasalahan dari pembelajaran mesin. Formulasi dari SVM yang termudah yaitu linier, di mana *hyperplane* terletak di tengah ruang data *input* (Furqan, 2011).

Masalah klasifikasi dapat diselesaikan dengan menentukan garis *hyperplane* yang memisahkan antara dua kelompok dengan menghitung margin antara kedua kelas, kemudian menghasilkan titik *hyperplane* yang paling optimal (Affymetrix, 2006). Margin adalah jarak antara *hyperplane* dan anggota terdekat dari setiap kelas. Hasil penentuan *hyperplane* adalah inti dari proses pembelajaran di SVM (Weston & Ben-Hur, 2010).

Persamaan *hyperplane* dapat dirumuskan:

$$\bar{w} \cdot \bar{x} + b = 0 \quad (1)$$

Anggota  $\bar{x}$  yang termasuk dalam kelas plus (+) dapat dirumuskan:

$$\bar{w} \cdot \bar{x} + b \geq +1 \quad (2)$$

Anggota  $\bar{x}$  yang termasuk dalam kelas minus (-) dapat dirumuskan:

$$\bar{w} \cdot \bar{x} + b \leq -1 \quad (3)$$

Perumusan klasifikasi SVM:

$$\min_f \|f\|_k^2 + C \sum_{i=1}^l |1 - y_i f(x_i)| \quad (4)$$

di mana C adalah parameter regulasi yang berfungsi menjadi kontrol dari margin serta *misclassification* (Evgeniou & Pontil, 2001).

Dalam permasalahan multiklas SVM terdapat dua jenis metode yakni *one-against-one* dan *one-against-all*. Metode *one-against-one* mengubah masalah multiklas menjadi serangkaian sub-tugas biner dengan  $g = c(c-1)/2$  yang dapat dilatih oleh SVM biner. Metode *one-against-all* dalam kernel SVM mengubah masalah multiklas menjadi serangkaian sub-tugas biner  $c$  yang dapat dilatih oleh SVM biner. Ketika ada permasalahan dari data yang belum bisa terpisahkan dengan cara linier di ruang masukan, margin umumnya belum bisa mendapatkan *hyperplane* yang benar untuk meminimalkan kesalahan klasifikasi titik data dan secara umum secara akurat. Oleh karena itu, kernel digunakan, yang membuat vektor latih dilakukan pemetaan dalam ruang dimensional yang lebih besar (tidak adanya batasan) dari fungsi  $\phi$  (pemetaan nonlinear). Lalu algoritma ini dapat memperoleh pemisah *hyperplane* yang memiliki margin maksimum di ruang dimensional yang besar tersebut.  $C > 0$  merupakan parameter penalti pada *misclassification* (Awad & Khanna, 2015).

Kernel Gaussian RBF:

$$K(x_i, x_j) = \exp(-\gamma \|x_i, x_j\|^2), \gamma > 0 \quad (5)$$

Parameter dari sigma yang bisa dilakukan penyesuaian adalah peranan utama pada performa kernel ini dan wajib disesuaikan berdasarkan masalah yang dimiliki. Bila berlebihan, dengan demikian eksponensial menunjukkan perilaku yang hampir linier, serta proyeksi dari dimensional yang lebih besar dapat menyebabkan kekuatan non linier yang dimiliki semakin hilang.

Kernel Laplacian:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{\sigma}\right) \quad (6)$$

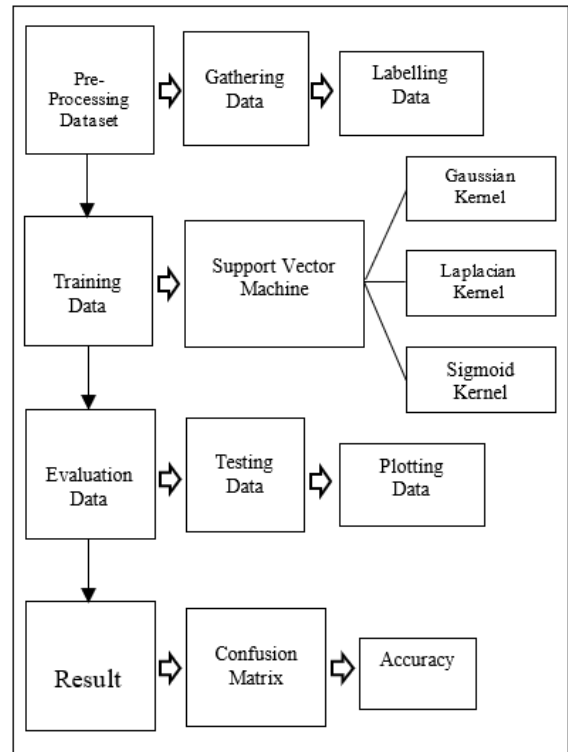
Kernel Laplacian adalah kelas kernel yang termasuk ke dalam *radial basis function* (RBF). Penggunaan parameter sigma dalam kernel gaussian juga berlaku untuk kernel laplacian.

Kernel Sigmoid:

$$K(x_i, x_j) = \tanh(\gamma x_i, x_j + c) \quad (7)$$

### 3. Metode

Pada Gambar 1 menunjukkan alur dalam penelitian ini untuk mencapai hasil data yang diproses dalam tahapan pelatihan dan kemudian dievaluasi untuk menentukan kinerja kernel SVM dalam pembelajaran untuk dapat memberikan hasil yang optimal.



Gambar 1. Diagram penelitian

#### 3.1. Pre-processing Dataset

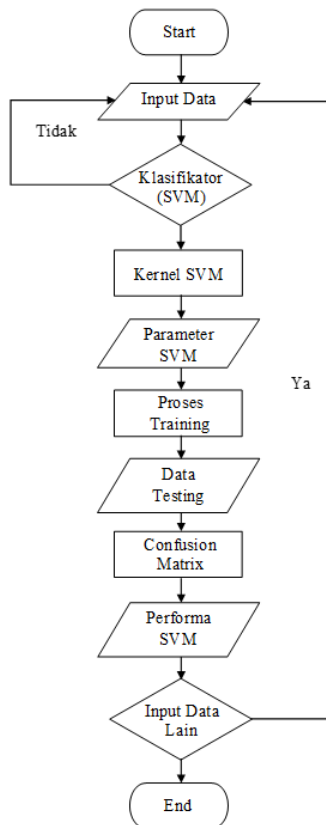
Data diambil dari publikasi PGC yang dapat diakses dan diunduh melalui [www.med.unc.edu/pgc](http://www.med.unc.edu/pgc). Data diberi label dalam format *file .xls* sesuai dengan label yang terkandung dalam set data yang telah diunduh. Label data terdiri dari kromosom, nilai kromosom, dan target, di mana 660 data diambil dengan setiap gangguan yang terdiri dari 220 data.

#### 3.2. Training Data

Pelatihan data memiliki tujuan dalam melakukan pelatihan terhadap mesin dengan tahapan pembelajaran sehingga bisa mempelajari data yang masih belum diketahui jenisnya serta mendapatkan akurasi yang dihasilkan seoptimal mungkin. Implementasi data pelatihan pada penelitian ini dengan menggunakan kernel SVM melalui pemanfaatan *library Accord.NET* sebagai alat untuk pembelajaran mesin dalam bahasa pemrograman C# di Visual Studio.

3.2. Evaluation Data

Pengujian atau evaluasi hasil memiliki tujuan agar dapat diketahui seberapa baiknya kinerja dari mesin pada tahapan pembelajaran menggunakan kernel SVM. Evaluasi kinerja menggunakan teknik *confusion matrix* dan lalu dihasilkan nilai akurasi yang merupakan informasi paling penting dari kesuksesan algoritma yang lebih berkelas pada pemrosesan data yang diklasifikasi. Melalui penggunaan tiga jenis kernel yang dihasilkan dilakukan perbandingan yang bertujuan agar dapat diketahui kernel yang paling baik dari SVM pada data penyakit mental. *Flowchart* SVM tertera pada Gambar 2.



Gambar 2. Flowchart SVM

Parameter kinerja *confusion matrix* merupakan akurasi yang memiliki definisi menjadi rasio banyaknya diagonal utama terhadap banyaknya total objek. Apabila akurasinya mendekati 100%, hasil dari kinerja algoritma hampir sempurna. *Confusion matrix* hanya memiliki dua *output* kelas dengan empat istilah sebagai representasi dari hasil proses klasifikasi, termasuk *true positive* (TP), *true negative* (TN), *false positive* (FP), *false negative* (FN). Berdasarkan nilai empat representasi tersebut, nilai dari *accuracy*, *precision*, dan *recall* dapat diperoleh. Rumus *confusion matrix* dalam menghitung nilai akurasi:

$$accuracy = \frac{\sum_{x=1}^l \frac{TP_x + TN_x}{TP_x + TN_x + FP_x + FN_x}}{l} \times 100\% \quad (8)$$

Tingkat kesalahan klasifikasi (*misclassification*), merupakan persentase dari seberapa besar kesalahan yang terjadi oleh klasifikasi yang dihasilkan. Semakin kecil persentase kesalahan, semakin baik kinerja yang dimiliki algoritma yang digunakan.

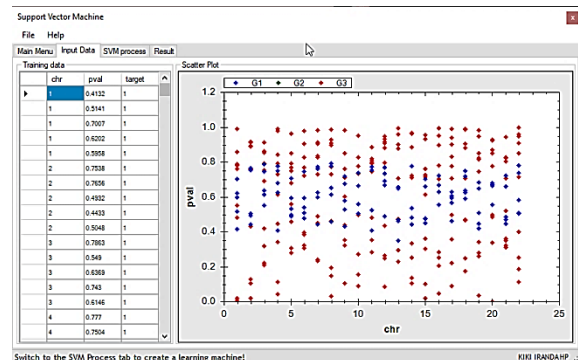
$$error\ rate = 1 - accuracy \quad (9)$$

*Recall* atau sensitivitas merupakan perbandingan dari banyaknya objek yang dilakukan klasifikasi dengan benar menjadi bagian dari kelas x pada sistem dengan banyaknya objek dari kelas yang diprediksi sebagai kelas x.

$$recall = \frac{\sum_{x=1}^l TP_x}{\sum_{x=1}^l (TP_x + FN_x)} \quad (10)$$

4. Hasil dan Pembahasan

Data yang dimasukkan dalam proses pelatihan akan menampilkan distribusi data dalam plot persebaran yang ditunjukkan pada Gambar 3 dalam sistem yang telah dibangun.

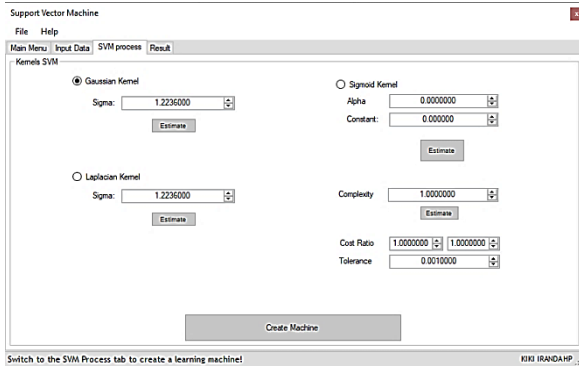


Gambar 3. Training Data Process

Dalam menguji data, algoritma ini diimplementasikan ke dalam sistem yang dibuat menggunakan bahasa pemrograman C# di Visual Studio dengan menggunakan *library* Accord.NET yang menyediakan algoritma SVM. Dalam *form* pengujian data, proses SVM pemrosesan SVM menyediakan tampilan opsi kernel yang nantinya dimanfaatkan pada tahapan proses data menggunakan algoritma SVM. Pemilihan kernel SVM yang dimanfaatkan pada penelitian ini meliputi kernel gaussian, laplacian, dan sigmoid. Sesudah memilih kernel, langkah lebih lanjut yaitu memperkirakan parameter. Pada sistem yang dibangun ini, parameter diperkirakan dengan otomatis sesudah dianalisis data menggunakan algoritma SVM melalui *machine learning library* di Accord.NET. Sesudah parameter diperkirakan melalui penggunaan *estimate button*, lalu klik *Create Machine button* agar mesin melakukan pemrosesan yang mempelajari data melalui pemanfaatan algoritma SVM.

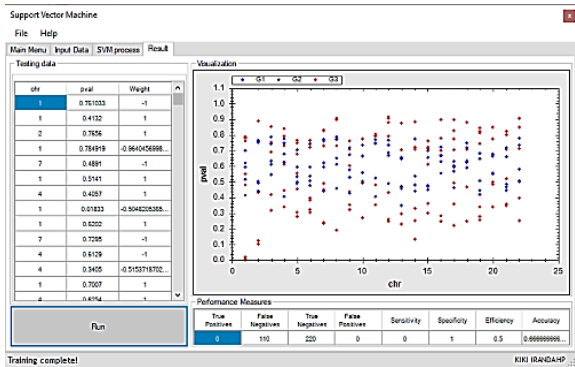
Setiap kernel mempunyai rumusan persamaan yang berbeda-beda ada untuk melakukan pembelajaran terhadap algoritma SVM, sehingga mempengaruhi kinerja algoritma dalam proses pembelajaran set data. Berdasarkan tahapan

pembelajaran data dari penyakit mental yang memanfaatkan penggunaan algoritma SVM dalam penelitian ini, kemudian dilakukan pengimplementasian berbentuk sistem, hasil tes yang didapatkan dalam bentuk akurasi dari algoritma kernel SVM yang melatih data dengan mengusulkan metode *one-against-one* dan *one-against-all* tertera pada Gambar 4.



Gambar 4. SVM Kernels Process

Pada Gambar 5 menunjukkan data pengujian dan data plotting yang akan diuji pada kinerja kernel SVM menggunakan *confusion matrix*.



Gambar 5. Testing Data dan evaluasi performa

Pada Tabel 1 dan Tabel 2 menunjukkan pengujian yang dihasilkan dari tiap-tiap data penyakit mental, yaitu autisme, bipolar, dan skizofrenia, memanfaatkan penggunaan algoritma svm dengan tiga jenis kernel yang berbeda-beda di antaranya gaussian, laplacian, dan sigmoid.

Tabel 1. Performa Kernel SVM menggunakan metode *One-Against-One*

Dataset	SVM Kernel	Recall	Accuracy	Error Rate
Autism vs Bipolar	Gaussian	0.63	80.9%	19.1%
	Laplacian	0.83	50.9%	49.1%
	Sigmoid	0.95	50%	50%
Bipolar vs Schizophrenia	Gaussian	0.69	84.5%	15.5%
	Laplacian	0.76	86.3%	13.7%
	Sigmoid	0.96	48.1%	51.9%
Schizophrenia vs Autism	Gaussian	0.1	88.1%	11.9%
	Laplacian	0.1	89.1%	10.9%
	Sigmoid	0.45	50%	50%

Tabel 2. Performa Kernel SVM Menggunakan Metode *One-Against-All*

Dataset	SVM Kernel	Recall	Accuracy	Error Rate
Autism vs Bipolar	Gaussian	0.30	60%	40%
	Laplacian	0.83	40%	60%
Schizophrenia vs Autism	Sigmoid	0.95	34.8%	65.2%
	Gaussian	0.33	66%	34%
	Laplacian	0.30	60%	40%
Schizophrenia vs Autism	Sigmoid	0.963	30%	70%
	Gaussian	0.45	60%	40%
	Laplacian	0.49	60%	40%
Bipolar	Sigmoid	0.455	63%	37%

Melalui hasil tes ini dapat dikenal sebagai hasil kinerja algoritma SVM dalam bentuk akurasi, yang merupakan tujuan dari penelitian ini agar menentukan kelayakan SVM dalam melakukan pembelajaran terhadap penyakit mental autisme, bipolar, dan skizofrenia.

Hasil evaluasi kinerja dalam Tabel 1 dan Tabel 2 didapatkan melalui penggunaan *confusion matrix* di mana memberi penggambaran prediksi dan keadaan secara nyata berdasarkan pembelajaran yang dihasilkan dari algoritma SVM. Akurasi menjadi referensi kinerja yang dimanfaatkan dalam penelitian ini. Hal tersebut dikarenakan akurasi menjadi suatu perbandingan prediksi kebenaran yang positif ataupun negatif melalui semua data yang telah menjadi masukan dan dilakukan pembelajaran oleh algoritma.

Kernel Gaussian berkaitan dengan data yang memiliki distribusi probabilitas bersyarat dengan pendekatan fungsi gaussian. Kernel Gaussian umumnya berkinerja lebih baik daripada kernel linear dan kernel polinomial dalam multikelas. Kelemahan kernel ini adalah sulit untuk menemukan parameter sigma yang optimal dan setara C yang memberikan hasil yang lebih baik untuk masalah tertentu. Namun, dalam evaluasi tiga kernel SVM, kernel gaussian memberikan akurasi terbaik dari kernel lain dengan rata-rata 84,5% menggunakan metode *one-against-one* dan 62% menggunakan metode *one-against-all*.

Kernel Laplacian dibangun di atas kerangka kerja SVM dan metode kernel yang telah ditentukan. Fungsi klasifikasi *output* oleh kernel laplacian identik dalam formulasi SVM tradisional, sehingga menyederhanakan aplikasi kernel Laplacian. Kernel Laplacian mendefinisikan fungsi klasifikasi logis untuk semua ruang input, menjadikannya solusi praktis untuk berbagai aplikasi. Dalam data ini, kernel laplacian memiliki akurasi rata-rata 75,43% menggunakan metode *one-against-one* dan 53,33% menggunakan metode *one-against-all*, yang masih lebih rendah dari kinerja kernel gaussian.

Kernel Sigmoid pada penelitian ini tidak efisien seperti fungsi kernel lainnya, karena tidak memiliki kondisi kernel yang valid yang diperlukan. Jumlah parameter yang memadai harus dipilih dengan benar untuk akurasi klasifikasi tinggi. Dalam penelitian ini, kernel sigmoid memiliki akurasi 50,63% menggunakan metode *one-against-one* dan 42,6%



menggunakan metode *one-against-all*, di mana hasil ini adalah kinerja terendah kernel SVM.

## 5. Kesimpulan

Sistem pembelajaran mesin yang menggunakan algoritma SVM dengan tiga kernel yang dibuat dalam penelitian ini dapat membantu menentukan kemampuan SVM untuk bekerja dalam proses pembelajaran data, khususnya studi autisme, gangguan bipolar, dan skizofrenia berdasarkan data GWAS. Melalui pengujian data penyakit mental ke dalam sistem yang telah dibuat, hasilnya menunjukkan bahwa algoritma SVM dengan kernel gaussian memiliki akurasi rata-rata terbaik daripada dua kernel lainnya, hasil akurasinya adalah 62% menggunakan metode *one-against-all* dan 84,5% menggunakan metode *one-against-one*. Namun, hasil ini masih memerlukan pengembangan kelayakan kernel menggunakan data GWAS dalam penelitian lebih lanjut melalui penggunaan lebih banyak data dan kernel SVM lainnya.

## Ucapan Terima Kasih

Penelitian ini didukung oleh Departemen Ilmu Komputer, dan akademisi Universitas Islam Negeri Sumatera Utara.

## Daftar Pustaka

- Affymetrix, T., 2006. What is a support vector machine? *Nature Biotechnology*, 24(12), 1565–1567.  
[https://www.researchgate.net/publication/6639744\\_What\\_is\\_a\\_Support\\_Vector\\_Machine](https://www.researchgate.net/publication/6639744_What_is_a_Support_Vector_Machine)
- Almandil, N., Alkuroud, D., AbdulAzeez, S., AlSulaiman, A., Elaissari, A., & Borgio, J., 2019. Environmental and genetic factors in autism spectrum disorders. *Environmental Research and Public Health*, 16(4), 1–16.  
<https://doi.org/10.3390/ijerph16040658>
- Aprilia, D., Johar, A., & Hartuti, P., 2014. Sistem Pakar diagnosa autisme pada anak. *Jurnal Rekursif*, 93–98.  
<https://ejournal.unib.ac.id/index.php/rekursif/article/download/310/269>
- Awad, M., & Khanna, R., 2015. Efficient learning machines: Theories, concepts, and applications for engineers and system designers. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers* (1st ed.). Apress.  
<https://doi.org/10.1007/978-1-4302-5990-9>
- Bush, W.S., & Moore, J.H., 2012. Genome-wide association studies. *PLoS Computational Biology*, 8(12), 1–12.  
<https://doi.org/10.1371/journal.pcbi.1002822>
- Carroll, L.S., & Owen, M.J., 2009. Genetic overlap between autism, schizophrenia and bipolar disorder. *Genome Medicine*, 1(10).  
<https://doi.org/10.1186/gm102>
- Choresyo, B., Nulhaqim, S.A., & Wibowo, H., 2015. Kesadaran masyarakat terhadap penyakit mental. *Prosiding Penelitian dan Pengabdian Kepada Masyarakat*, 2(3), 301–444.  
<https://doi.org/10.24198/jppm.v2i3.13587>
- Evgeniou, T., & Pontil, M., 2001. Support vector machines: theory and applications. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2049 LNAI, 249–257.  
[https://doi.org/10.1007/3-540-44673-7\\_12](https://doi.org/10.1007/3-540-44673-7_12)
- Furqan, M., 2011. Face recognition using eigenfaces and smooth support vector machine [Universiti Malaysia Pahang].  
[http://umpir.ump.edu.my/id/eprint/2732/1/mhd\\_furqan.pdf](http://umpir.ump.edu.my/id/eprint/2732/1/mhd_furqan.pdf)
- Geschwind, D.H., Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., Van Bakel, H., Varghese, M., Wang, Y., Shieh, A.W., Haney, J., Parhami, S., Belmont, J., Kim, M., Losada, P. M., Khan, Z., Mleczko, J., Pinto, D., 2018. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*, 362(6420).  
<https://doi.org/10.1126/science.aat8127>
- Grove, J., Ripke, S., & Als, T.D., 2019. Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, 51(3), 431–444. <https://doi.org/10.1038/s41588-019-0344-8>
- Hyman, S., Chisholm, D., Kessler, R., Patel, V., & Whiteford, H., 2006. Mental Disorders. *Disease Control Priorities in Developing Countries*, 31, 605–625.
- Hyman, S.E., 2000. The genetics of mental illness implications for practice. *Bulletin of the World Health Organization*, 78(4), 455–463.  
<https://doi.org/10.1590/S0042-96862000000400007>
- Kementerian Kesehatan Republik Indonesia, 2018. Laporan Riset Kesehatan Dasar. Kementerian Kesehatan RI.
- Maciukiewicz, M., Geraci, J., Rotzinger, S., Foster, J. A., Kennedy, J.L., & Muller, D.J., 2017. GWAS-based machine learning approach to predict duloxetine response in major depressive disorder. *European Neuropsychopharmacology. Conference: 23rd World Congress of Psychiatric Genetics, WCPG. Canada*, 27(Supplement 2), S285–S286.  
<https://doi.org/10.1016/j.euroneuro.2015.09.010>
- Mattes, J.A., 2010. Bipolar disorder. *Archives of General Psychiatry*, 67(11), 1198.

- <https://doi.org/10.1001/archgenpsychiatry.2010.143>
- National Institute of Mental Health, 2018. Bipolar Disorder. National Institutes of Health Publication.
- Novianty, A., & Rochman Hadjam, M.N., 2017. Literasi Kesehatan Mental dan Sikap Komunitas sebagai Prediktor Pencarian Pertolongan Formal. *Jurnal Psikologi*, 44(1), 50. <https://doi.org/10.22146/jpsi.22988>
- Psychiatric Genomics Consortium, 2019. What is the PGC? <https://www.med.unc.edu/pgc/>
- Purcell, S.M, Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F. & Sklar P.. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 460(7256), 748–752. [https://www.researchgate.net/publication/232772602\\_Common\\_polygenic\\_variation\\_contributes\\_to\\_risk\\_of\\_schizophrenia\\_and\\_bipolar\\_disorder](https://www.researchgate.net/publication/232772602_Common_polygenic_variation_contributes_to_risk_of_schizophrenia_and_bipolar_disorder)
- Reig, R.B., 2013. Bipolar disorder. *FMC Formacion Medica Continuada En Atencion Primaria*, 20(10), 565–572. [https://doi.org/10.1016/S1134-2072\(13\)70667-4](https://doi.org/10.1016/S1134-2072(13)70667-4)
- Rietveld, C., Medland, S., Derringer, J., Yang, J., Esko, Y., Martin, N., Westra, H., Shakhbazov, K& Abdellaoui, A., 2013. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, 340(6139), 1467–1471. <https://doi.org/10.1126/science.1235488>
- Vapnik, V., 1998. *Statistical Learning Theory*. Wiley.
- Walker, J., 2013. *Methods in molecular biology*. In *Research in Microbiology* (Vol. 149, Issue 5). Springer Science Media. [https://doi.org/10.1016/s0923-2508\(98\)80442-7](https://doi.org/10.1016/s0923-2508(98)80442-7)
- Weston, J.& Ben-Hur, A., 2010. *A User's Guide to Support Vector Machine*. *Data Mining Techniques for the Life Sciences*, 609, 223–239. [https://doi.org/10.1007/978-1-60327-241-4\\_13](https://doi.org/10.1007/978-1-60327-241-4_13)
- World Health Organization, 2013. *Mental Health Action Plan 2013-2020*. World Health Organization.