

BAB II

TINJAUAN PUSTAKA

2.1 Standarisasi Kualitas Udara dalam Mengatasi Pencemaran Udara

2.1.1 Indeks Standar Pencemar Udara (ISPU)

Kualitas udara pada umumnya dinilai dari konsentrasi parameter pencemaran udara yang terukur lebih tinggi atau lebih rendah. Indeks Standar Pencemar Udara (ISPU) adalah angka yang menggambarkan kondisi mutu udara ambien di lokasi tertentu yang didasarkan kepada dampak terhadap kesehatan manusia, nilai estetika dan makhluk hidup lainnya. Indeks Standar Pencemar Udara (ISPU) memiliki rentang nilai dimana semakin tinggi nilai Indeks Standar Pencemar Udara (ISPU), maka semakin tinggi tingkat polusi udara dan semakin besar pula dampaknya bagi kesehatan[17].

Tabel 2.1 Indeks Standar Pencemar Udara (ISPU)

Kategori	Rentang (parts per million)
Baik	0-50
Sedang	51-100
Tidak Sehat	101-199
Sangat Tidak Sehat	200-299
Berbahaya	300-lebih

(Sumber: Keputusan Kepala Bapedal No. 107 Tahun 1997)

Berdasarkan Tabel 2.1, berikut ini adalah penjelasan dari masing-masing tingkatan pada Indeks Standar Pencemar Udara (ISPU)[17]:

1. Baik

Nilai Indeks Standar Pencemar Udara (ISPU) berada diantara 0-50. Hal ini menunjukkan kualitas udara yang baik dengan dampak yang minim bagi kesehatan.

2. Sedang

Nilai Indeks Standar Pencemar Udara (ISPU) berada diantara 51-100. Kualitas udara cukup baik, namun terdapat polusi yang dapat mempengaruhi kesehatan sebagian kecil masyarakat. Masyarakat yang sensitif terhadap polusi partikel akan mengalami gangguan pernapasan.

3. Tidak Sehat

Nilai Indeks Standar Pencemar Udara (ISPU) berada diantara 101-199. Tingkat kualitas udara yang bersifat merugikan pada manusia ataupun kelompok hewan yang sensitif atau biasa menimbulkan kerusakan pada tumbuhan ataupun nilai estetika.

4. Sangat Tidak Sehat

Nilai Indeks Standar Pencemar Udara (ISPU) berada diantara 200-299. Dapat memicu peringatan kesehatan yang berarti seluruh masyarakat dapat merasakan dampak kesehatan yang serius.

5. Berbahaya

Nilai Indeks Standar Pencemar Udara (ISPU) berada diatas 300. Pada tingkat ini dapat memicu peringatan kesehatan dalam kondisi darurat. Seluruh populasi dapat merasakan dampak kesehatan yang serius.

2.1.2 Parameter Penyebab Pencemaran Udara

Menurut PP Nomor 41 Tahun 1999 Pasal 5[18], disebutkan bahwa setidaknya ada 5 parameter yang dapat menyebabkan pencemaran udara antara lain Karbon Monoksida (CO), Karbon Dioksida (CO₂), Hidro Karbon, Partikulat (PM₁₀), Suhu dan Kelembaban. Berikut ini adalah karakteristik dan dampak dari kelima parameter tersebut, yaitu :

1. Karbon Monoksida (CO)

Karbon monoksida adalah gas yang tidak berbau, pada suhu udara normal berbentuk gas yang tidak berwarna, dan tidak berasa. 90% dari CO di udara perkotaan berasal dari emisi kendaraan bermotor, sisanya berasal dari pembakaran sampah, pembakaran batubara, dan lain

sebagainya. Dampak karbon monoksida yang meningkat di berbagai perkotaan dapat mengganggu kesehatan.

2. Karbon Dioksida (CO₂)

Karbon dioksida adalah salah satu bahan pencemar udara yang sangat berbahaya. Apabila gas karbon dioksida semakin banyak, maka akan menghalangi pemancaran panas dari bumi yang akan menyebabkan *global warming*. Selain itu, dapat menyebabkan *hypercapnia* yang menimbulkan asidosis apabila konsentrasinya mencapai >15000 ppm.

3. Hidro Karbon (HC)

Hidrokarbon adalah bahan pencemar udara yang dapat berbentuk gas, cairan maupun padatan. Sumber dari pencemaran hidrokarbon ini dapat berasal dari proses industri yang diemisikan ke udara dan dapat berasal dari transportasi dan kegiatan manusia yang lain. Dampak dari hidrokarbon ini, apabila masuk dalam paru-paru akan menimbulkan luka dan merangsang terbentuknya sel-sel kanker.

4. Partikulat (PM₁₀)

PM₁₀ adalah salah satu bahan pencemar udara primer, berbentuk padatan atau liquid di udara yang dapat tinggal di atmosfer dalam waktu yang lama. Di samping mengganggu estetika, partikulat juga memberikan dampak buruk terhadap kesehatan. Menurut WHO, ukuran partikulat debu yang berukuran 0,1 µm sampai dengan kurang dari 10 µm apabila terhirup secara terus menerus dapat mengakibatkan gangguan saluran pernapasan manusia.

5. Suhu dan Kelembaban (DHT11)

Kelembaban udara ditentukan dari banyaknya kandungan uap air di udara. Tingkat kelembaban bervariasi menurut suhu. Semakin hangat suhu udara, semakin banyak uap air yang dapat ditampung, dan sebaliknya. Jadi ketika udara hangat naik dan mulai mendingin, lama kelamaan akan kehilangan kemampuan untuk menahan/menampung uap air.

2.2 *Artificial Intelligence*

Artificial Intelligence atau kecerdasan buatan adalah ilmu yang mempelajari bagaimana membuat mesin/program (komputer) dapat melakukan pekerjaan seperti dan sebaik yang dilakukan oleh manusia bahkan bisa lebih baik daripada yang dilakukan manusia[19]. Teknologi non-fisik ini dapat membuat keputusan dengan cara menganalisis dan menggunakan data yang tersedia di dalam sistem. Proses yang terjadi dalam *Artificial Intelligence* mencakup *learning*, *reasoning*, dan *self-correction*[20]. Dalam dunia teknologi, *Artificial Intelligence* sudah banyak digunakan dimana salah satunya digunakan dalam model klasifikasi melalui algoritma.

2.3 Tahapan Klasifikasi Data pada *Machine Learning*

Machine learning adalah bagian dari *artificial intelligence* yang menyediakan sistem dengan kemampuan secara otomatis untuk belajar dan meningkatkan pengalaman tanpa diprogram secara eksplisit. *Machine learning* membutuhkan data sebagai bahan belajar dan menggunakan algoritma untuk proses komputasi data besar serta lebih cepat. Pada *machine learning*, ada berbagai algoritma yang dapat membantu menyelesaikan masalah terhadap data baru yang memungkinkan untuk mendukung pengambilan keputusan[21]. Salah satunya adalah algoritma *Support Vector Machine* yang digunakan pada penelitian ini.

Machine Learning terdiri dari tiga tahapan, yaitu[22]:

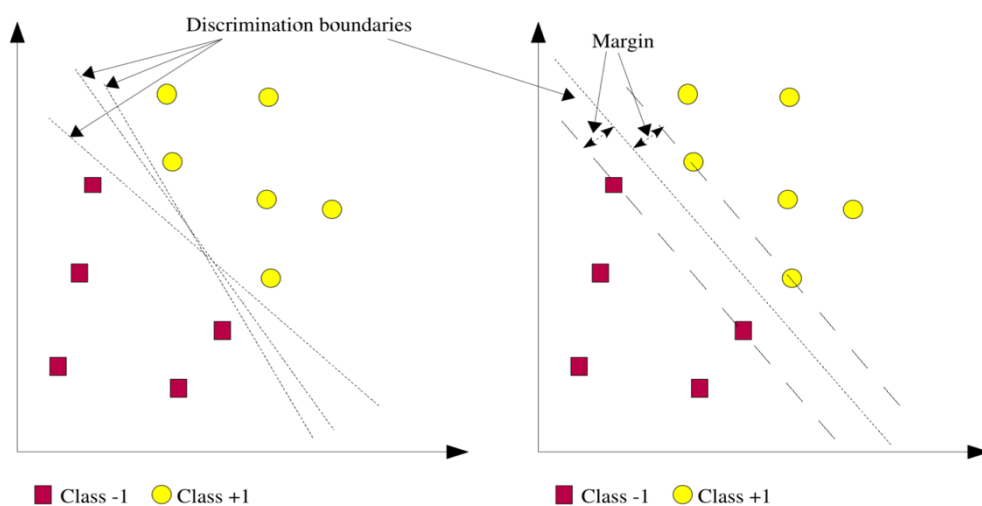
1. Model
Sistem yang membentuk prediksi atau identifikasi.
2. Parameter
Faktor yang digunakan oleh model untuk membentuk keputusannya.
3. Pembelajaran/Testing
Sistem yang menyesuaikan parameter dan model dalam prediksi versus hasil aktual.



Gambar 2.1 Tahapan dalam *Machine Learning*

2.4 Algoritma *Support Vector Machine (SVM)*

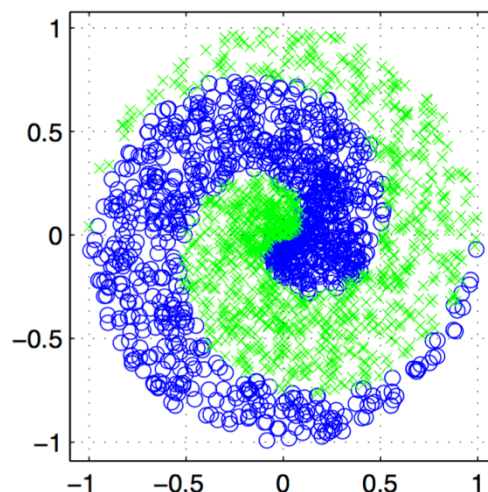
Support Vector Machine (SVM) adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi[23]. Namun, biasanya banyak digunakan pada kasus klasifikasi[13][14][15]. Konsep *Support Vector Machine (SVM)* dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada *input space*.



Gambar 2.2 SVM Berusaha Menemukan *Hyperplane* Terbaik yang Memisahkan Kedua Kelas

Masalah klasifikasi dapat diterjemahkan dengan usaha menemukan garis (*hyperplane*) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) yang ditunjukkan pada Gambar 2.2. *Hyperplane* pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur margin *hyperplane* tersebut dan mencari titik maksimalnya. Margin adalah jarak antara *hyperplane* tersebut dengan *pattern* terdekat dari masing-masing kelas. *Pattern* yang paling dekat ini disebut sebagai *support vector*. Garis solid pada Gambar 2.2 sebelah kanan menunjukkan *hyperplane* yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua kelas, sedangkan titik merah dan kuning yang berada dalam lingkaran hitam adalah *support vector*. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pembelajaran pada *Support Vector Machine* (SVM)[23].

Support Vector Machine (SVM) juga memiliki fungsi kernel yang terdiri atas linear, gaussian, polinomial, dan sigmoid. Pemilihan fungsi kernel yang tepat adalah hal yang sangat penting, karena fungsi kernel ini akan menentukan *feature space* dimana fungsi klasifier akan dicari. Pada Gambar 2.3, data ini sulit dipisahkan secara linear. Fungsi kernel adalah salah satu untuk mengatasinya. Dengan fungsi kernel, suatu data x di *input space* di mapping ke *feature space* F dengan dimensi yang lebih tinggi melalui map φ sebagai berikut $\varphi : x \rightarrow \varphi(x)$. Karena itu data x di *input space* menjadi $\varphi(x)$ di *feature space*[23].

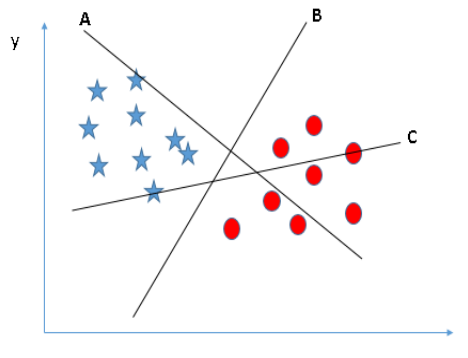


Gambar 2.3 Data Spiral yang Menggambarkan Ketidaklinearan

2.4.1 Algoritma *Support Vector Machine* dalam Melakukan Klasifikasi Data

Berikut ini adalah bagaimana *Support Vector Machine* mengidentifikasi *hyperplane* sehingga didapatkan klasifikasi kelas-kelas terbaik[24]:

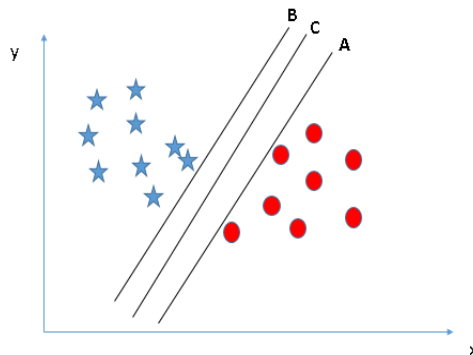
1. Identifikasi *hyperplane* yang tepat (Skenario 1). Dibawah ini terdapat 3 *hyperplane* (A, B, dan C). Lalu, identifikasikan *hyperplane* yang tepat untuk mengklasifikasikan bintang dan lingkaran.



Gambar 2.4 Identifikasi *Hyperplane* (Skenario 1)

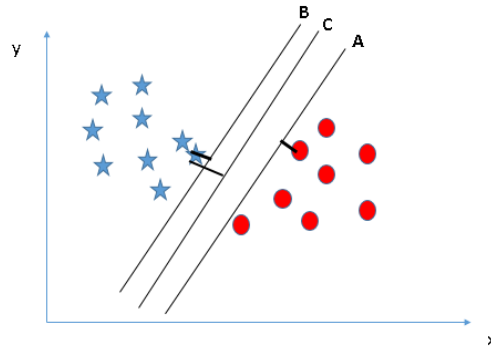
Perlu diingat mengenai aturan praktis untuk mengidentifikasi *hyperplane* yang tepat. Pilih *hyperplane* yang memisahkan dua kelas dengan lebih baik. Pada skenario ini, *hyperplane* “B” telah melakukan pekerjaan ini dengan sangat baik.

2. Identifikasi *hyperplane* yang tepat (Skenario 2). Disini, kita mempunyai 3 *hyperplanes* (A, B, dan C) dan semua memisahkan kelas dengan baik.



Gambar 2.5 Identifikasi *Hyperplane* (Skenario 2)

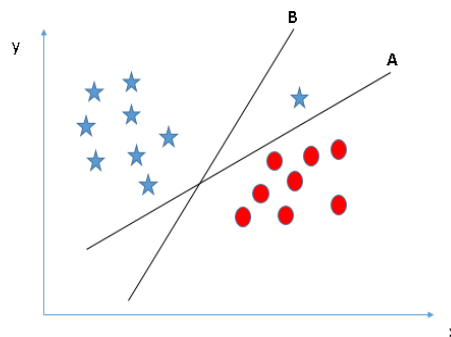
Disini, memaksimalkan jarak antara titik kelas terdekat, dan *hyperplane* akan membantu kita untuk memutuskan *hyperplane* yang tepat.



Gambar 2.6 Garis C sebagai *hyperplane* yang tepat

Pada Gambar 2.6, kita dapat melihat bahwa margin untuk *hyperplane* C lebih tinggi dibandingkan dengan A dan B. Oleh karena itu, garis C adalah *hyperplane* yang tepat. Alasan untuk memilih *hyperplane* dengan margin yang lebih tinggi adalah untuk ketahanan. Jika kita memilih *hyperplane* yang memiliki margin lebih rendah, maka ada kemungkinan besar kesalahan klasifikasi.

3. Identifikasi *hyperplane* yang tepat (Skenario 3). Gunakan aturan seperti yang dibahas di bagian sebelumnya untuk mengidentifikasi *hyperplane* yang tepat.

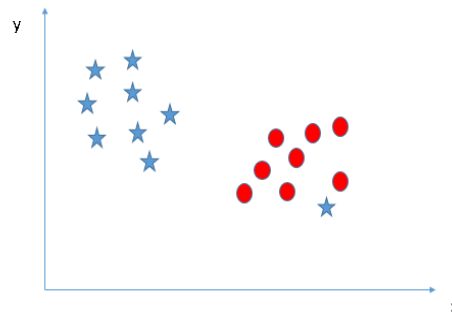


Gambar 2.7 Identifikasi *hyperplane* (Skenario 3)

Mungkin kita memilih *hyperplane* B, karena memiliki margin yang lebih tinggi dibandingkan dengan A. Namun inilah hasilnya, SVM memilih *hyperplane* yang mengklasifikasikan kelas secara akurat

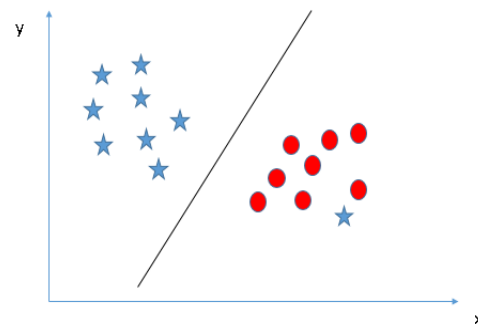
sebelum memaksimalkan margin. Disini, *hyperplane* B memiliki kesalahan klasifikasi dan A telah mengklasifikasikan semua dengan benar. Oleh karena itu, *hyperplane* yang tepat adalah A.

4. Bisakah kita mengklasifikasikan dua kelas? (Skenario 4). Pada Gambar 2.8, tidak dapat memisahkan dua kelas menggunakan garis lurus, karena salah satu bintang terletak di wilayah kelas (lingkaran) lainnya.



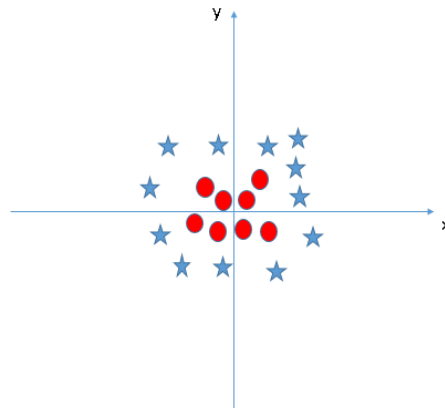
Gambar 2.8 Klasifikasi 2 Kelas (Skenario 4)

Seperti yang telah disebutkan, satu bintang di ujung yang lain seperti *outlier* untuk kelas bintang. *Support Vector Machine* (SVM) memiliki fitur untuk mengabaikan *outlier* dan menemukan *hyperplane* yang memiliki margin maksimum.



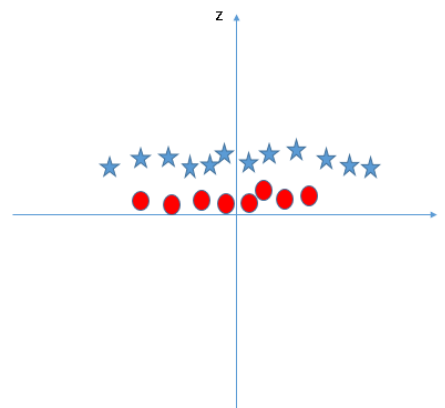
Gambar 2.9 Menemukan *hyperplane* dengan mengabaikan *outlier*

5. Temukan *hyperplane* untuk dipisahkan ke kelas (Skenario 5). Dalam skenario dibawah ini, kita tidak dapat memiliki *hyperplane* linier antara dua kelas.



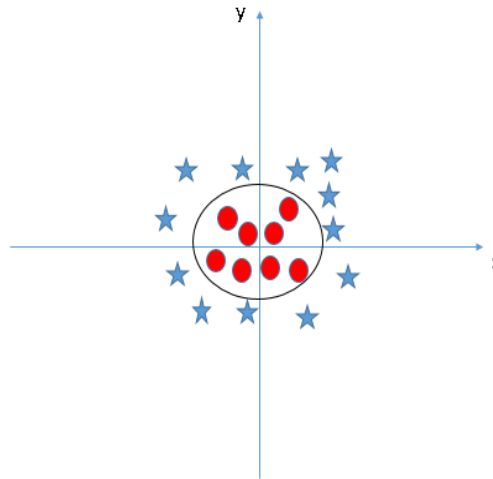
Gambar 2.10 Identifikasi *hyperplane* linier

Support Vector Machine (SVM) dapat mengatasi masalah ini, yaitu dengan memperkenalkan fitur tambahan. Pada skenario ini, akan ditambahkan fitur baru $z = x^2 + y^2$. Sekarang, kita plot titik kelas pada sumbu x dan z :



Gambar 2.11 Plot Titik Kelas

Dalam plot diatas, poin yang perlu dipertimbangkan adalah semua nilai untuk z akan selalu positif, karena z adalah jumlah kuadrat dari x dan y . Dalam plot asli, lingkaran merah tampak dekat dengan asal sumbu x dan y , yang mengarah ke nilai z yang lebih rendah dan bintang yang relatif jauh dari hasil asal ke nilai z yang lebih tinggi. Berikut ini adalah saat melihat *hyperplane* di ruang input asli, terlihat seperti lingkaran :



Gambar 2.12 Hasil Identifikasi *hyperplane* di ruang input asli

2.4.2 Implementasi *Support Vector Machine* (SVM)

Support Vector Machine (SVM) telah banyak diterapkan dalam menyelesaikan masalah-masalah nyata dalam kehidupan sehari-hari. Baik dalam masalah seperti finansial, cuaca hingga bidang kedokteran[23]. Terbukti dalam banyak implementasi[14][25][26], *Support Vector Machine* (SVM) memberi hasil yang lebih baik terutama dalam hal solusi yang dicapai.

Penerapan *Support Vector Machine* (SVM) dalam diagnosis pasien kanker payudara[14], mampu membangun suatu sistem klasifikasi yang dapat mendiagnosa keadaan pasien secara optimal berdasarkan hasil mamografi dan faktor-faktor yang mempengaruhi kanker payudara. Selain itu, *Support Vector Machine* (SVM) diimplementasikan pada sistem prediksi kualitas udara yang mampu menghasilkan tingkat akurasi yang lebih baik, hal ini disebabkan karena pengaruh pengaturan nilai parameter yang optimal dan memiliki waktu komputasi yang cepat[25]. Berbeda dari sebelumnya, penggunaan algoritma *Support Vector Machine* (SVM) menghasilkan kurva trend yang sama persis dengan pola kurva online trans forex, sehingga sangat membantu para *trader* dalam pengambilan keputusan untuk beli (*buy*) atau jual (*sell*)[26]. Penggunaan metode *Support Vector Machine* (SVM) menunjukkan hasil yang lebih baik dari segi ketepatan maupun kecepatan pembuatan model[27].

2.5 Uji Akurasi Pemodelan Klasifikasi dengan Metode Statistik

Kecermatan (akurasi) adalah ukuran yang menunjukkan kedekatan hasil analisis dengan kadar analit yang sebenarnya. Akurasi dinyatakan sebagai persen perolehan nilai kesalahan klasifikasi (*error*) terhadap total sampel yang diperoleh selama pengukuran. Akurasi hasil klasifikasi sangat tergantung kepada sebaran galat sistematis di dalam keseluruhan tahapan analisis. Oleh karena itu untuk mencapai akurasi klasifikasi yang tinggi hanya dapat dilakukan dengan cara mengurangi galat sistematis tersebut seperti menggunakan peralatan yang telah dikalibrasi, pelaksanaannya yang cermat, dan lain sebagainya[28].

Berikut ini adalah rumus statistik perhitungan performansi akurasi klasifikasi yang ditetapkan sebagai berikut:

$$\% \text{ Akurasi Klasifikasi} = \frac{\text{Error}}{\text{Total sampel yang diperoleh dari pengujian}} \times 100$$

2.6 Perbandingan Penelitian

Tabel 2.2 Perbandingan Penelitian Sebelumnya

Penulis (Tahun)	Algoritma/ Teknik Pemodelan	Tipe Klasifikasi	Kelebihan	Kekurangan	Tujuan
Purwantaka, Ricki Indra (2011)[9]	<i>Maximum Likelihood Estimator (MLE)</i>	Metode Regresi Logistik	Model regresi logistik lebih mudah dan efisien.	- <i>Error</i> klasifikasi cukup besar. - Sedikitnya variabel prediktor.	Untuk mendapatkan faktor yang mempengaruhi kanker payudara.
Intansari, dkk (2012)[10]	Algoritma <i>bagging</i>	Metode <i>Bagging Logistic Regression</i>	- Mampu menyelesaikan masalah ketidakseimbangan kelas. - Terdapat 4 variabel prediktor. - Replikasi bootstrap meningkatkan ketepatan klasifikasi sebesar 1,11%.	- Lebih baik menggunakan metode Naive Bayes, <i>decision tree</i> , atau K-NN. - Ketepatan klasifikasi sebesar 66,7%.	Untuk meningkatkan akurasi klasifikasi data medis.
Putra, I Wayan Suartika Eka (2016)[11]	Algoritma <i>Correspondence</i>	Metode <i>Convolutional Neural Network</i>	- Menggunakan <i>confussion matrix</i> .	- Hasil akurasi sebesar 20-50%. - Waktu penyelesaian lama.	Penyelesaian klasifikasi citra objek dengan <i>Convolutional Neural Network</i> .
Sugiarto, Bambang dan Sustika, Rika (2016)[12]	Algoritma C4.5, Algoritma <i>decision tree</i>	Metode <i>decision tree</i>	- Menggunakan WSN untuk mengirimkan semua data ke BSC. - Menggunakan <i>confussion matrix</i> . - Implementai algoritma si simpul BSC untuk menghasilkan status bahaya sistem peringatan.	- Ketepatan klasifikasi sebesar 85,71%.	Untuk klasifikasi data kualitas udara.
Octaviani, dkk (2014)[13]	Algoritma <i>Support Vector Machine</i>	Metode <i>Support Vector Machine</i>	- Menghasilkan akurasi klasifikasi sempurna yaitu 100%. - Data sesuai kelas asli dengan kernel RBF.	Akan lebih baik menggunakan database.	Untuk penerapan metode <i>Support Vector Machine</i>
Novianti, Fourina Ayu dan Purnami, Santi Wulan (2012)[14]	Algoritma <i>Support Vector Machine</i> , <i>Maximum Likelihood Estimator (MLE)</i>	Metode regresi logistik dan <i>Support Vector Machine</i>	SVM memiliki performansi akurasi klasifikasi terbaik yaitu 94,34%.	- Terdapat <i>missing value</i> . - Metode regresi logistik memiliki performansi akurasi sebesar 84,90%.	Untuk melakukan perbandingan klasifikasi data antara metode regresi logistik dan <i>Support Vector Machine</i> .