SJME (Supremum Journal of Mathematics Education)
Vol.4, No.2, July 2020, pp. 147-158
ISSN: 2548-8163 (online)| ISSN: 2549-3639 (print)
DOI :

Pembelajaran Vektor Untuk Klasifikasi Data Pada Bidang

Hanna Arini Parhusip
Universitas Kristen Satya Wacana
hanna.parhusip@uksw.edu

Bambang Susanto
Universitas Kristen Satya Wacana
bambang.susanto@uksw.edu

Lilik Linawati
Universitas Kristen Satya Wacana
lilik.linawati@uksw.edu

Suryasatriya Trihandaru
Universitas Kristen Satya Wacana
suryasatriya@uksw.edu

Yohanes Sardjono
Badan Tenaga Nuklir Nasional
sardjono.batan@gmail.com

Informasi Artikel

Sejarah artikel:

Diterima 22 April 2020
Direvisi 22 Juni 2020
Disetujui 13 Juli 2020

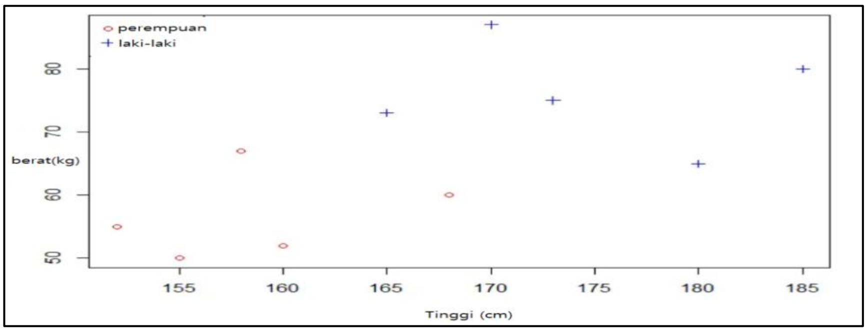
Kata kunci:
Vektor Normal, Proyeksi Ortogonal, Hyperplane , Margin.

ABSTRAK

Tujuan penelitian ini adalah penyusunan hyperplane untuk memisahkan data yang mempunyai 2 kelas dan bersifat linear pada bidang datar sebagai pembelajaran vektor untuk klasifikasi data. Adapun metode yang digunakan adalah pre-Support Vector Machine (SVM). Metode ini mencari garis (hyperplane) terbaik yang memisahkan data dan memberi ruang antar 2 kelas data dimana ruang pemisah tersebut tidak boleh memuat data serta ruang tersebut merupakan margin maksimal. Langkah awal adalah menduga garis pemisah (hyperplane) awal melalui titik O. Dengan mengambil salah satu titik data yang menjadi titik referensi, disusun vektor dari O terhadap titik referensi dan garis melalui titik referensi sebagai batas pertama margin. Kemudian dibentuk vektor arah dari titik O yang tegak lulus terhadap garis awal (hyperplane). Selanjutnya vektor proyeksi dibentuk dari titik referensi terhadap vektor arah sehingga vektor arah dan vektor proyeksi berhimpit (searah). Penyusunan margin diperoleh dengan menyusun garis yang pararel terhadap garis awal sebagai hyperplane serta berjarak 2 kali dengan panjang vektor proyeksi tersebut. Hyperplane terbaik diperoleh secara manual dengan mengatur batas kedua dari margin yang diperoleh dengan menggambar garis melalui suatu titik data pada kelas ke-2 dengan jarak terdekat dan pararel terhadap garis yang melalui titik referensi dari data kelas ke-1.

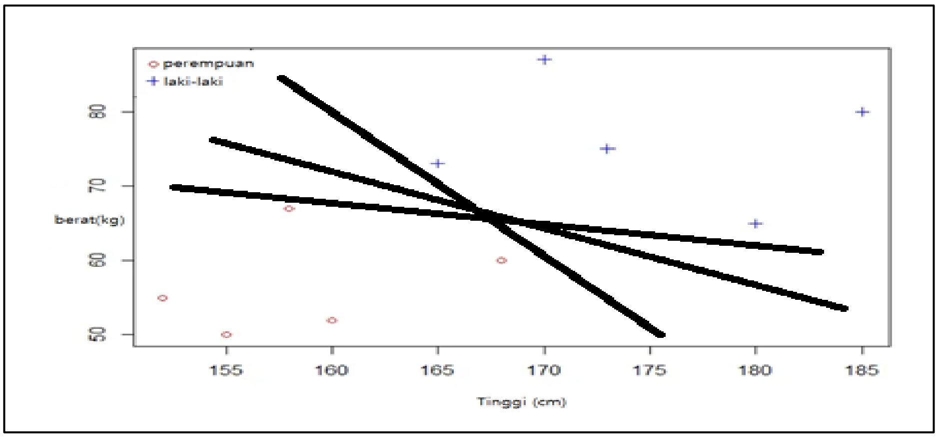
PENDAHULUAN

Pertanyaan yang sering muncul pada siswa dan dalam belajar matematika adalah aplikasi matematika yang dipelajari. Siswa pada saat ini seringkali menghendaki korelasi yang langsung terlihat dengan sekelilingnya. Untuk itu pada artikel ini diperkenalkan bagaimana siswa dapat belajar tentang vektor yang mendasari dalam pengolahan data ketika siswa pada dunia nyata dimana pengolahan data dewasa ini berupa *big data*. Pada tingkat pendidikan sekolah menengah atas, vektor lebih banyak diperkenalkan manfaatnya pada bidang fisika sekalipun siswa juga mengalami kesulitan (Poluakan & Runtuwene, 2018) sehingga salah satu pendekatan juga memperkenalkan vektor secara geometri (Claudia Orozco, Rodríguez Erla M. Morales & Gonçalves da Silva Cordeiro, 2015). Akan tetapi cara pengolahan data belum diperkenalkan kepada siswa. Sedangkan dengan teknik-tekniknya yang berkembang dalam pengolahan data yang disebut *Data Science* memerlukan ketrampilan matematika, pemodelan dan komputerisasi ataupun dalam melakukan *coding* . Beberapa teknik dalam pengolahan *big data* berkembang cepat dan salah satu teknik disebut *Support Vector Machine* **(**SVM) yang merupakan algoritma untuk mengklasifikasi yang berarti menggolongkan data pada suatu kelompok tertentu oleh karena data dalam beberapa kelompok yang berbeda dengan data besar (*big data*) (Demidova et al., 2016) . Misalkan diberikan data pada Gambar 1 yang menunjukkan setiap koordinat titik ditandai lingkaran (o) dan tambah (+) yang menunjukkan hubungan tinggi dan berat badan orang perempuan dan laki-laki.



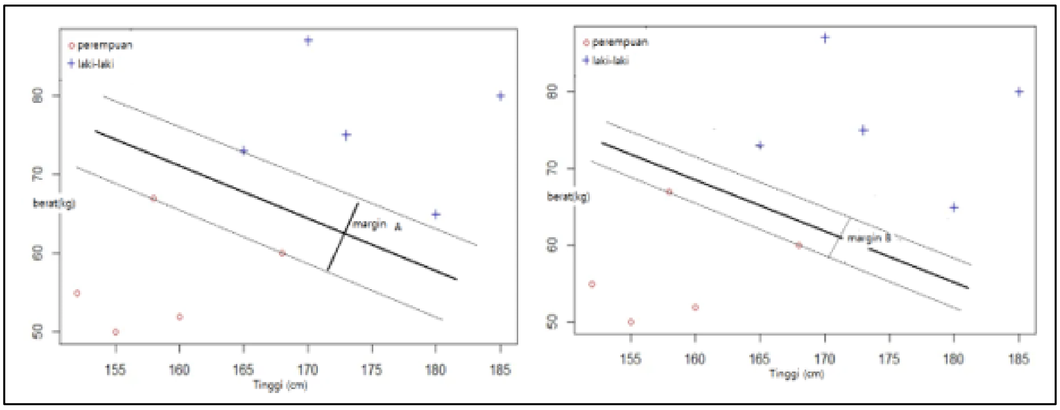
Gambar 1. Ilustrasi data berlabel

Pada gambar 1 merupakan tinggi dan berat beberapa orang dari laki-laki dan perempuan. Dengan himpunan data tersebut, selanjutnya SVM akan membantu menjawab pertanyaan : jika ada data berat dan tinggi, apakah orang itu laki-laki atau perempuan. Misal, jika seseorang dengan ukuran tinggi 175 cm dan berat 80 cm apakah dia laki-laki atau perempuan ?. Situasi ini sebenarnya situasi sehari-hari yang dapat muncul di sekitar siswa. Jika data orang yang diobservasi sangat banyak, maka hal ini tidak mungkin dilakukan secara manual. Siswa perlu melakukan pemrograman atau *coding*. Oleh karena itulah, pada artikel ini ditunjukkan bagaimana siswa diperkenalkan dalam penyusunan suatu algoritma yang terkait sehingga siswa dapat memahami perlunya belajar vektor dan operasi aljabar vektor yang terkait yang diperlukan. Kemungkinan siswa belum dapat memahami bagaimana SVM secara lengkap, tetapi pada bagian ini ditunjukkan manfaat vektor yang terkait dengan pengolahan data khususnya dalam melakukan klasifikasi dimana kelak siswa juga akan menjumpai data dalam skala besar. Kita dapat melihat data pada Gambar 1 bahwa terjadi 2 kelompok yang berbeda. Misalkan kita akan memasangkan garis yang memisahkan laki-laki dan perempuan, dimana garis **(***hyperplane***)** tersebut disebut garis **(***hyperplane***)** pemisah.



Gambar 2. Ilustrasi contoh garis pemisah yang mungkin

Dari Gambar 2, , terdapat 3 garis (bisa lebih) yang memungkinkan untuk merupakan pemisah 2 kelompok. Pertanyaan yang muncul : manakah yang terbaik ? Mengapa diberi nama *hyperplan*e ? Pada Gambar data memang hanya informasi yang dibangun dari pasangan data dalam 2 dimensi. Kata *hyperplane* merupakan generalisasi dari suatu bidang. Dalam 1 dimensi , suatu *hyperplane* adalah titik. Dalam 2 dimensi, *hyperplane* adalah garis, dalam 3 dimensi *hyperplane* adalah bidang dan dimensi lebih dari 3 disebut *hyperplane*. Agar data 2 kelompok terpisahkan dengan baik maka diperlukan margin (ruang) antara kelompok sehingga perlu penyusunan 2 garis optimal dan pemisah 2 ruang tersebut disebut margin. Diberikan suatu margin maka diharapkan tidak ada data dalam margin tersebut. Akan tetapi dapat terjadi data memuat informasi yang tidak jelas *(noise* data) sehingga perlu adanya margin memberikan batas keterpisahan 2 kelompok data. Gambar 3 menunjukkan 2 contoh margin dari data pada Gambar 1.



Gambar 3. Ilustrasi margin

Dari pengamatan Gambar 3 dapat disimpulkan bahwa jika suatu *hyperplane* sangat dekat pada suatu data, maka margin akan kecil. Semakin jauh suatu *hyperplane* dari data, maka margin makin besar. Hal inilah yang menyebabkan tujuan algoritma nantinya adalah memilih *hyperplane* pemisah optimal yang memaksimalkan margin data . Oleh karena itu pada metode ditunjukkan bagaimana melakukan perhitungan margin sehingga pembaca dapat memahami tentang manfaat belajar vektor dalam klasifikasi data.

METODE

Pada bagian ini ditunjukkan beberapa hal penting yang perlu dipahami dengan benar tentang vektor dan hal-hal yang terkait dengan vektor seperti : norm, arah , penambahan dan pengurangan vektor , perkalian dot dan proyeksi vektor pada vektor yang lain.

Langkah-langkah Analisis yang dilakukan

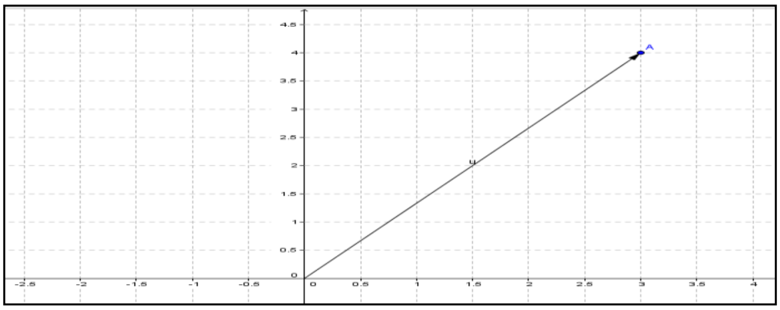
Untuk menyusun *hyperplane* dengan margin yang maksimal, maka dilakukan metode pre-*support vector machine* (SVM) dimana pada artikel ini hanya dibatasi untuk data pada bidang yang bersifat linear (dapat dipisah oleh garis), yaitu : 1) Penyusunan *hyperplane* awal melalui O sebagai pemisah 2 kelas data; 2) Menetapkan titik referensi dari data dan membuat garis pararel terhadap *hyperplane* awal melalui titik referensi dimana garis ini sebagai batas margin pertama; 3) Membuat vektor melalui O dan titik referensi; 4) Menyusun vektor arah ($\boldsymbol{w}$) dari titik O yang tegak lurus terhadap *hyperplane* awal; 5) Membentuk vektor proyeksi dari titik referensi terhadap vektor arah; 6) Sebutlah vektor $\boldsymbol{u}$ adalah vektor satuan vektor arah $\boldsymbol{w}$, dan $\boldsymbol{p}$ adalah vektor proyeksi dari titik data (***a***) pada $\boldsymbol{w}$, sehingga $\boldsymbol{z} = (\boldsymbol{u}.\boldsymbol{x})\boldsymbol{u}$ dan $\boldsymbol{p} = (\boldsymbol{u}.\boldsymbol{a})\boldsymbol{u}$**; 7)** Jarak dari suatu titik ke *hyperplane* adalah 2|| $\boldsymbol{p}$ ||; 8) Menyusun garis batas margin kedua yang melalui salah satu titik data pada kelas ke-2 dengan jarak terdekat terhadap hyperplane awal dan; 9) Melakukan penyesuain secara manual untuk mengatur *hyperplane* terbaik.

Langkah terakhir ini dilakukan karena pencarian hyperplane terbaik berarti memaksimalkan margin yang tidak dibahas dalam tulisan ini. Hal inilah yang menyebabkan metode ini disebut pre-SVM karena merupakan langkah awal dalam algoritma SVM yang dapat dipelajari pada tingkat yang lebih tinggi dalam bahasa mesin. Karena pada artikel ini mengutamakan pembelajaran vektor untuk klasifikasi data, maka langkah dalam algoritma lebih lanjut tidak ditunjukkan (Kowalczyk, 2017).

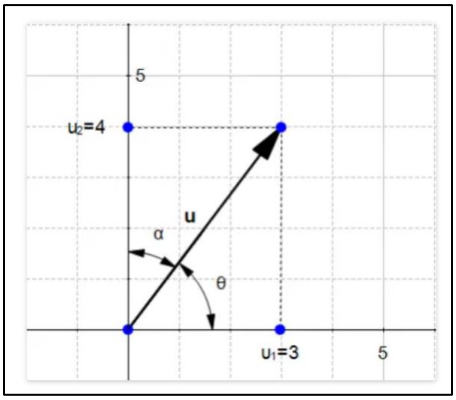
Hitung margin hyperplane

Contoh 1. Perhatikan titik A(3,4) di R^2. Jika kita tarik garis dari O(0,0) pada titik tersebut maka kita dapat mendefinisikan vektor OA dimana garis dari O ke-A merupakan vektor tersebut. Jadi vektor adalah obyek yang mempunyai arah dan besaran. Sedangkan panjang garis dari O ke A disebut panjang vektor yang disebut *norm* vektor atau disebut panjang OA. Ditulis panjang segmen OA dengan ||OA|| dimana kita dapat menghitungnya dengan rumus Pythagoras. Sedangkan arah vektor ditunjukkan dengan vektor satuan . Jika sembarang vektor adalah $\boldsymbol{u} = [u_1,u_2]^T$ maka vektor satuan $\boldsymbol{u} = [u_1,u_2]^T$ adalah

$$\boldsymbol{e} = \left[\frac{u_1}{\|\boldsymbol{u}\|}, \frac{u_2}{\|\boldsymbol{u}\|}\right]^T .$$



Gambar 4. Contoh penampilan vektor dari O ke A(3,4)

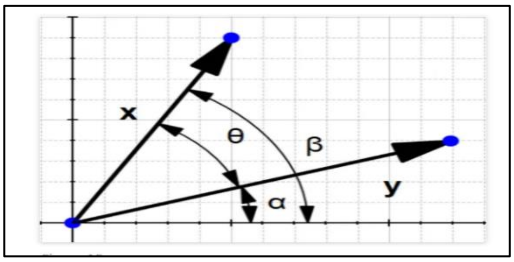


Gambar 5. Ilustrasi vektor ***u*** dengan sudut yang dibentuk

Mendefinisikan arah hyperplane

Untuk mendefinisikan arah vektor *hyperplane* diperlukan terlebih dahulu definisi arah vektor. Kita ketahui bahwa vektor pada 2 dimensi tersebut membentuk 2 sudut yaitu sudut terhadap sumbu horizontal dan sudut terhadap sumbu vertikal. Oleh karena itu vektor arah ditentukan oleh arah vektor $\boldsymbol{u} = [u_1, u_2]^T$ terhadap sumbu x dan sumbu y yang didefinisikan sebagai

$$cos(\theta) = \frac{u_1}{||\boldsymbol{u}||}; \quad cos(\alpha) = \frac{u_2}{||\boldsymbol{u}||}.$$



Gambar 6. Sudut yang dibentuk antara 2 vektor

Tetapi diketahui bahwa

$$\cos(\theta)=\cos(\beta-\alpha) = \cos(\beta)\cos(\alpha) + \sin(\beta)\sin(\alpha)$$

$$\cos(\theta) = \frac{x_1\, y_1}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} + \frac{x_2\, y_2}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}$$

$$\cos(\theta) = \frac{x_1\, y_1 \ + x_2\, y_2}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} = \frac{\boldsymbol{x}.\boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}$$

Jika dikalikan kedua ruas dengan $||x||||y||$ diperoleh :

$$\cos(\theta)\,\|\boldsymbol{x}\|\|\boldsymbol{y}\| = x_1\, y_1 \ + x_2\, y_2$$

Yang sama artinya dengan :

$$\|\boldsymbol{x}\|\|\boldsymbol{y}\|\cos(\theta) = \boldsymbol{x}.\boldsymbol{y}$$

Selanjutnya diasumsikan bahwa penjumlahan dan pengurangan antara 2 vektor secara geometri telah dipahami. Kemudian , diperlukan proyeksi ortogonal vektor yang ditunjukkan berikut ini.

Proyeksi ortogonal suatu vektor

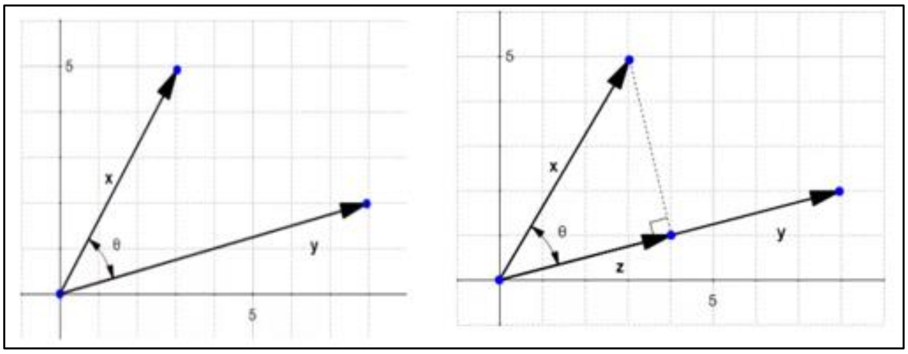
Telah disebutkan di atas bahwa salah satu algoritma yang muncul dalam pengelompokan data yaitu SVM yang penyusunannya sangat dilatarbelakangi pengetahuan tentang vektor dan proyeksi ortogonal vektor dimana salah satu penggunaan algoritma ini adalah untuk klasifikasi kanker panyudara (Moh. Yamin Darsyah, 2013) (Kathija & Shajun,

2016) (Min-Wei et al., 2018). Proyeksi ortogonal ini juga digunakan pada algoritma yang lain seperti CAD (Kwanghee & Takis, 2014), juga dalam persiapan pemrosesan data pada algoritma Support Vector for Regression(SVR) dan metode *chemometry* (Fatma et al., 2019) serta untuk kalibrasi data (Boulet & Roger, 2018). Untuk itulah bagian ini ditunjukkan bagaimana proses proyeksi ortogonal vektor. Anggaplah vektor $\boldsymbol{x}$ diproyeksikan ortogonal ke vektor $\boldsymbol{y}$ (pada Gambar 7) maka diperoleh vektor proyeksi searah $\boldsymbol{y}$ dan sebutlah $\boldsymbol{z}$. Dari Gambar 7 diperoleh hubungan

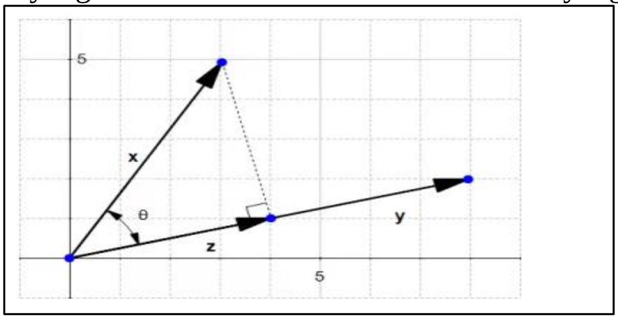
$$\cos(\theta) = \frac{\|\boldsymbol{z}\|}{\|\boldsymbol{x}\|}$$

yang berarti pula

$$\|\boldsymbol{z}\| = \|\boldsymbol{x}\| \cos(\theta).$$



Gambar 7. Sudut yang dibentuk antara 2 vektor dan vektor yang diproyeksikan



Gambar 8. Hubungan vektor yang diproyeksikan dan proyeksinya.

Dari hasil di atas dengan pengetahuan sebelumnya diperoleh :

$$\cos(\theta) = \frac{\boldsymbol{x}.\boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} \text{ dan karena } \|\boldsymbol{z}\| = \|\boldsymbol{x}\| \cos(\theta)$$

diperoleh

$$\|\boldsymbol{z}\| = \|\boldsymbol{x}\| \frac{\boldsymbol{x}.\boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} = \frac{\boldsymbol{x}.\boldsymbol{y}}{\|\boldsymbol{y}\|}$$

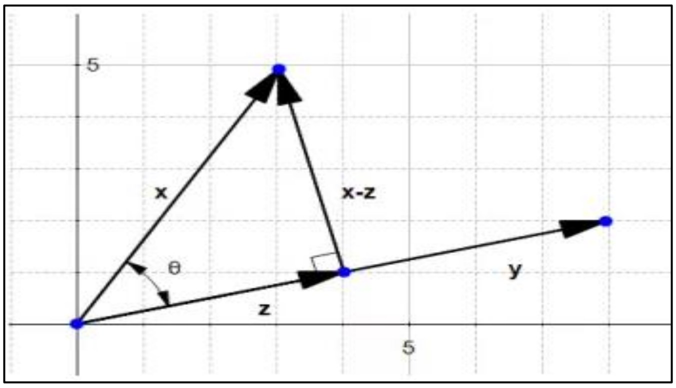
Padahal

$$\boldsymbol{u} = \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} \text{ sehingga } \|\boldsymbol{z}\| = \boldsymbol{u}.\boldsymbol{x}$$

Perhatikan bahwa $\boldsymbol{z}$ searah dengan $\boldsymbol{y}$ sehingga

$$\boldsymbol{u} = \frac{\boldsymbol{z}}{\|\boldsymbol{z}\|} \text{ sehingga } \boldsymbol{z} = \boldsymbol{u}\|\boldsymbol{z}\|$$

Substitusikan $\|\boldsymbol{z}\| = \boldsymbol{u}.\boldsymbol{x}$ pada $\boldsymbol{z} = \boldsymbol{u}\|\boldsymbol{z}\|$ diperoleh $\boldsymbol{z} = (\boldsymbol{u}.\boldsymbol{x})\boldsymbol{u}$
Jadi dapat dituliskan $\boldsymbol{z} = (\boldsymbol{u} \,.\, \boldsymbol{x})\, \boldsymbol{u}$ dimana $\boldsymbol{z}$ adalah vektor proyeksi ortogonal $\boldsymbol{x}$ ke $\boldsymbol{y}$.



Gambar 9. Vektor selisih .

Dengan mengetahui hal itu dapat dihitung jarak vektor ***x*** dan ***z*** yaitu dibentuk dari panjang vektor ***x -z***. Jadi diperoleh :

$$\|x - z\| = \sqrt{(3-4)^2 + (5-1)^2} = \sqrt{17}$$

Contoh di atas adalah cara untuk menghitung margin dari 1 titik pada bidang. Jadi dengan metode dapat dilakukan perhitungan margin untuk sekelompok data pada bidang 2 dimensi.

HASIL DAN PEMBAHASAN

Penyusunan *Hyperplane*

Pada bagian ini , kita mengikuti pada literatur bagaimana penyusunan hyperplane pada bidang (Kowalczyk, 2017). Kita mengetahui persamaan garis yaitu $y=ax+b$. Sedangkan jika dalam notasi vektor kita dapat menyusun persamaan umum *hyperplane* adalah

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} = 0.$$

Demikian pula garis $y=ax+b$ dapat ditulis $y-ax-b = 0$. Atau dapat ditulis :

$$y*1+(-a)*x+(-b)*1=0.$$

Persamaan tersebut dapat ditulis dalam notasi vektor. Sebutlah 2 vektor

$$\boldsymbol{w} = \begin{bmatrix} -b \\ -a \\ 1 \end{bmatrix} \text{ dan } \boldsymbol{x} = \begin{bmatrix} 1 \\ x \\ y \end{bmatrix}$$

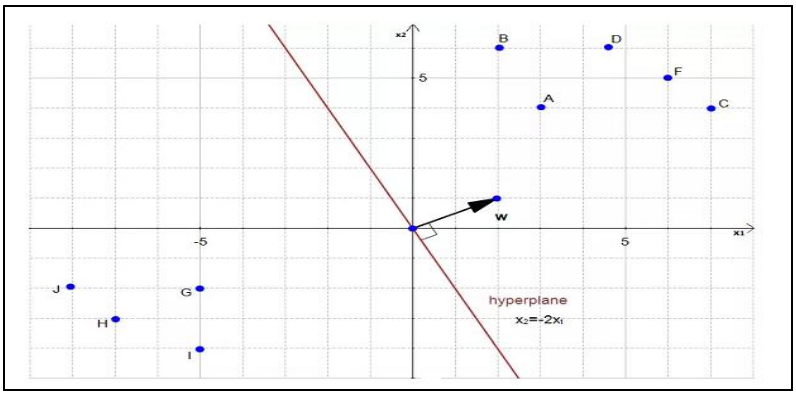
$$\boldsymbol{w}^T\boldsymbol{x} = -b\mathrm{x}(1) + (-a)\ \mathrm{x}\ x + (-b)\mathrm{x}1 = 0$$

$$\boldsymbol{w}^T\boldsymbol{x} = y - ax - b$$

Notasi tersebut digunakan agar diketahui bahwa vektor ***w*** adalah vektor yang selalu merupakan vektor normal ke h*yperplane* dimana disini selalu didefinisikan $w_0 = 0$. Pada Gambar 10 terdapat *hyperplane* yang memisahkan 2 grup data dimana persamaannya adalah :

$$x_2 = 2x_1$$

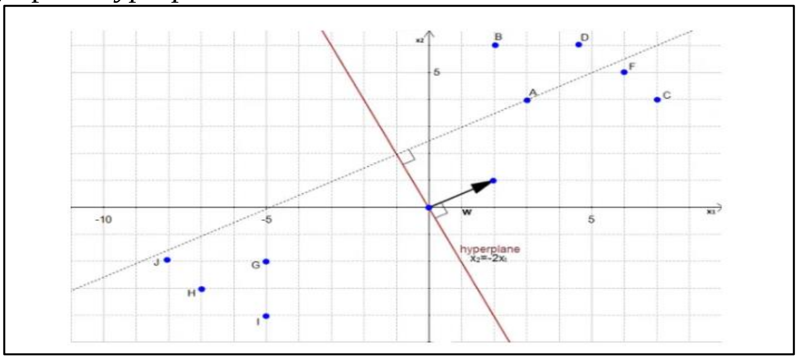
$$\boldsymbol{w}^T\boldsymbol{x} = 0$$



Gambar 10. Ilustrasi pemisahan data.

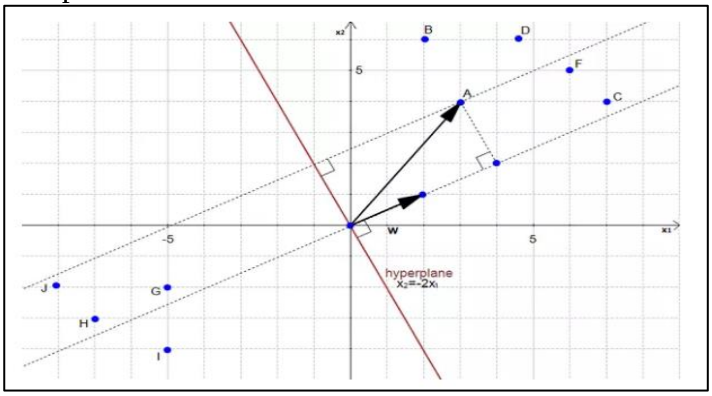
dimana $\boldsymbol{w} = \begin{bmatrix} \mathbf{2} \\ \mathbf{1} \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Catat bahwa vektor ***w*** ditunjukkan pada Gambar 10 dimana ***w*** bukan data. Selanjutnya akan dihitung jarak titik A(3,4) dan *hyperplane*. Akan ditentukan jarak dari A dan proyeksinya pada *hyperplane* .



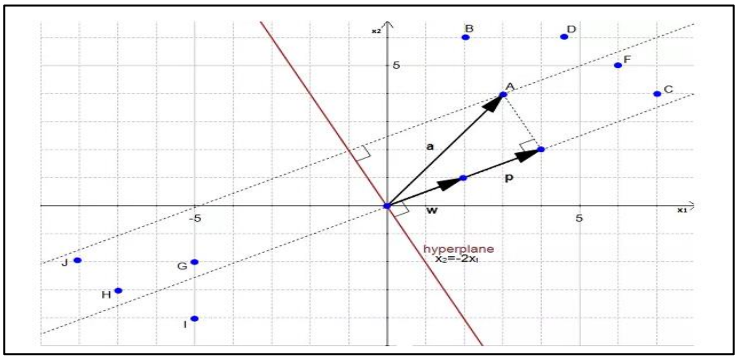
Gambar 11**.** Ilustrasi pemisahan data lanjutan

Kemudian vektor dari A ke titik O dapat disusun dan proyeksikan A normal ke vektor ***w*** sebagaimana pada Gambar12.



Gambar 12. Ilustrasi pemisahan data dengan vektor yang diproyeksikan serta *hyperplane* yang diperoleh

Hasilnya diperoleh vektor ***p*** yang ditunjukkan pada Gambar 13.



Gambar 13. Ilustrasi pemisahan data dengan vektor yang diproyeksikan serta *hyperplane* yang diperoleh dan vektor hasil proyeksi

Tujuan selanjutnya adalah menghitung jarak dari titik A ke *hyperplane*. Jarak itu adalah panjang vektor $\boldsymbol{p}$ yaitu $\|\boldsymbol{p}\|$. Diketahui bahwa $\boldsymbol{w}$= $[2,1]^T$ adalah vektor normal ke *hyperplane* dan $\boldsymbol{a}$=$[3,4]^T$ adalah vektor dari O ke titik A. Sebutlah vektor $\boldsymbol{u}$ adalah vektor arah $\boldsymbol{w}$, dan $\boldsymbol{p}$ adalah proyek $\boldsymbol{a}$ pada $\boldsymbol{w}$,sehingga

$$\|\boldsymbol{w}\| = \sqrt{4+1} = \sqrt{5} \text{ dan vektor satuanya adalah } \boldsymbol{u} = \left[\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right]^T.$$

Jelas bahwa dari formulasi $\boldsymbol{z} = (\boldsymbol{u}.\boldsymbol{x})\boldsymbol{u}$ dapat disusun

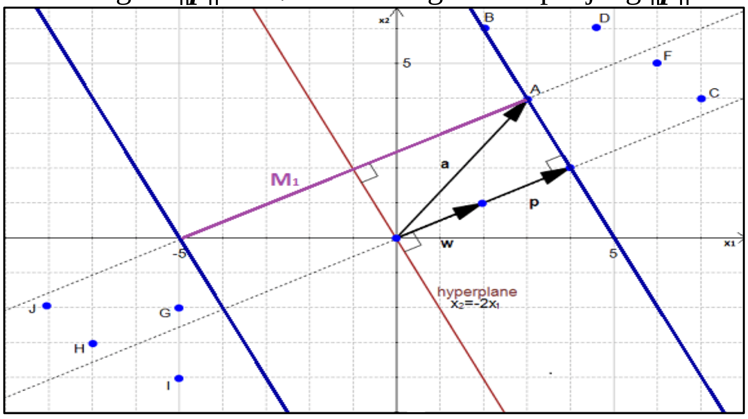
$$\boldsymbol{p} = (\boldsymbol{u}.\boldsymbol{a})\boldsymbol{u} = \left(3\text{x}\frac{2}{\sqrt{5}} + 4\text{x}\frac{1}{\sqrt{5}}\right)\left[\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right]^T$$

$$\boldsymbol{p} = (\boldsymbol{u}.\boldsymbol{a})\boldsymbol{u} = \frac{10}{\sqrt{5}}\left[\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right]^T = \left[\frac{20}{5}, \frac{10}{5}\right]^T$$

$$\boldsymbol{p} = [4,2]^T \text{dan } \|\boldsymbol{p}\| = \sqrt{4^2+2^2} = 2\sqrt{5} .$$

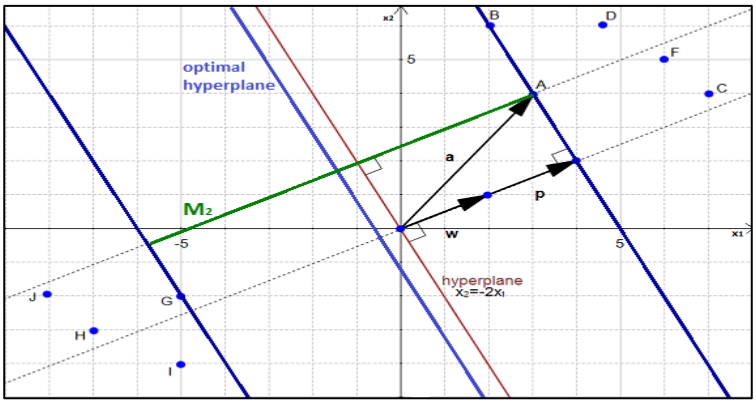
Hitung margin hyperplane

Diketahui bahwa $\|\boldsymbol{p}\|$ adalah jarak antara *A* dan *hyperplane*, sehingga margin didefinisikan sebagai $2\|\boldsymbol{p}\| = 4\sqrt{5}$. Pada bagian ini panjang $\|\boldsymbol{p}\|$ sudah dihitung.



Gambar 14. Ilustrasi yang menyediakan *hyperplane*

Diketahui bahwa *hyperplane* yang maksimal adalah *hyperplane* yang memaksimalkan margin dari data. Pada Gambar 14 diperoleh margin terhadap titik A dan sebutlah M_1. Tetapi M_1 bukanlah margin terbesar yang memisahkan data terbaik. Margin terbesar adalah M_2 pada Gambar 15. Hal ini dapat diperoleh dengan menarik garis yang pararel dengan garis yang melalui A dan mencari garis yang melalui titik terdekat terhadap h*yperplane* dengan garis baru tersebut yaitu G. Kemudian carilah garis optimalyang membagi 2 jarak garis baru yang melalui G dengan garis melalui A sehingga diperoleh Gambar 15.



Gambar 15. Ilustrasi *hyperplane*

Jadi dapat disimpulkan bagaimana dalam mendapatkan *hyperplane* terbaik yaitu: 1) Pilih himpunan data; 2) Pilih h*yperplane* yang memisahkan data dengan tidak ada titik-titik diantaranya dan; 3) Maksimalkan jarak (marginnya). Daerah yang dibatasi oleh *hyperplane* akan merupakan margin terbesar. Jadi pada step 1, dipunyai suatu himpunan data *D* dan akan dilakukan klasifikasi. Data terdiri dari vektor $\boldsymbol{x_i}$. Setiap $\boldsymbol{x_i}$ berasosiasi dengan nilai $\boldsymbol{y_i}$ yang mengindikasikan bahwa jika elemen-elemennya tergolong dalam 1 kelompok yang sama maka bertanda +1 dan sebaliknya akan bertanda -1. Jadi nilai $\boldsymbol{y_i}$ hanya bernilai +1 atau -1. Sedangkan pada step 2, ingin dibuat *hyperplane* yang memisahkan data sehingga ruang diantara *hyperplane* tidak memuat titik. Untuk data pada bidang 2 dimensi , maka data dapat digambar dengan pensil. Akan tetapi pada umumnya data berdimensi *p* sehingga tidak mungkin digambar. Jadi *hyperplane* dapat selalu ditemukan untuk memisahkan data dalam 2 dimensi. Hal ini hanya bisa dilakukan jika data terpisah secara linear. Hasil ini dapat menjadi dasar bagi siswa dalam melakukan pengelompokan data dalam 2 kelompok dengan pengetahuan vektor yang sudah dipelajari. Penyusunan *hyperplane* dalam algoritma untuk klasifikasi data yang lebih besar telah dibahas oleh Hakan dimana hal ini berlaku pada kasus yang lebih kompleks (Hakan, 2016).

SIMPULAN

Pada artikel ini ditunjukkan bagaimana data pada bidang 2 dimensi yang mempunyai 2 jenis kelompok diklasifikasi. Adapun pemisah data merupakan garis (hyperplane) yang mempunyai spasi antar data yang disebut margin. Margin ini perlu dimaksimalkan agar tidak memuat data yang berbeda menjadi tidak terklasifikasi dengan baik. Oleh karena itu pada artikel ini penyusunan hyperplane dan margin yang optimal ditunjukkan. Pembaca juga perlu mengetahui bahwa tidak selamanya data pada bidang 2 dimensi selalu dapat dipisahkan oleh 2 garis pararel dengan margin maksimal. Data yang dibahas pada bagian ini hanyalah data yang terpisah secara linear sehingga pemisah berupa garis (hyperplane) dengan margin antar garis dimungkinkan. Pemilihan margin optimal dengan hyperplane terbaik dilakukan manual karena langkah optimasi dalam algoritma tidak digunakan karena artikel ini memberikan contoh pemanfaatan vektor sebagai pembelajaran untuk klasifikasi data pada bidang dapat diperkenalkan.

DAFTAR PUSTAKA

Boulet, J., & Roger, J. (2018). A review of orthogonal projections for calibration. *Journal of Chemometric*. https://doi.org/https://doi.org/10.1002/cem.3045

Claudia Orozco, Rodríguez Erla M. Morales, M. F., & Gonçalves da Silva Cordeiro, M. (2015). Learning Objects and Geometric Representation for Teaching “Definition and

Applications of Geometric Vector.” *Journal of Cases on Information Technology (JCIT)*, *17*(1). https://doi.org/10.4018/JCIT.2015010102

Demidova, L., Nikulchev, E., & Sokolova, Y. (2016). Big Data Classification Using the SVM Classifiers with the Modified Particle Swarm Optimization and the SVM Ensembles. *International Journal of Advanced Computer Science and Applications*, *7*(5), 294–312. https://doi.org/10.14569/ijacsa.2016.070541

Fatma, F., Abdallah, Hany, W., & Darwish, I. A. (2019). Orthogonal projection to latent structures and first derivative for manipulation of PLSR and SVR chemometric models’ prediction: A case study. *Plos One*. https://doi.org/https://doi.org/10.1371/journal.pone.0222197

Hakan, C. (2016). Best Fitting Hyperplanes for Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1076–1088. https://doi.org/10.1109/TPAMI.2016.2587647

Kathija, & Shajun, N. (2016). Breast Cancer Data Classification Using SVM and Naïve Bayes Techniques. *International Journal of Innovative Research in Computer and Communication Engineering*, *3297*(6), 11449–11455. https://doi.org/10.15680/IJIRCCE.2016. 0412129

Kowalczyk, A. (2017). *Support Vector Machines Succintctly*. www.syncfusion.com.

Kwanghee, K., & Takis, S. (2014). Orthogonal projection of points in CAD/CAM applications: an overview. *Journal of Computational Design and Engineering*, *Volume 1*,(2), 116–127. https://doi.org/https://doi.org/10.7315/JCDE.2014.012

Min-Wei, H., Chih-Wen, C., Wei-Chao, L., Shih-Wen, K., & Chih-Fong, T. (2018). SVM and SVM Ensembles in Breast Cancer Prediction. *Plos One*, *12*. https://doi.org/10.1371/journal.pone.0161501

Moh. Yamin Darsyah. (2013). Menakar Tingkat Akurasi Support Vector Machine Study Kasus Kanker Payudara. *Statistika Universitas Muhammadiyah*, *1*(1), 15–20.

Poluakan, & Runtuwene, J. (2018). Students’ difficulties regarding vector representations in free-body system. *Journal of Physics Conference Series*, *1120*(1).

Learning Vector For Data Classification On A Plane

Hanna Arini Parhusip
Universitas Kristen Satya Wacana
hanna.parhusip@uksw.edu

Bambang Susanto
Universitas Kristen Satya Wacana
bambang.susanto@uksw.edu

Lilik Linawati
Universitas Kristen Satya Wacana
lilik.linawati@uksw.edu

Suryasatriya Trihandaru
Universitas Kristen Satya Wacana
suryasatriya@uksw.edu

Yohanes Sardjono
Badan Tenaga Nuklir Nasional
sardjono.batan@gmail.com

ABSTRACT

The purpose of this research was to arrange a hyperplane to separate data with 2 classes and linear on a plane as a vector learning to classify linear data. The used method is called pre-Support Vector Machine (SVM). This method is searching the best hyperplane separating data with a space called a margin which contain no data in between. By taking one point data as a reference to start with, the vector from O to this reference point is made and the line through this reference point is drawn as the first line to bound the margin. Futhermore, a vector is made from O normal to the first hyperplane as a direction vector. The projection vector is then computed from the reference point to the direction vector such that both vector on the same direction. Finally, the margin can be obtained by drawing an other line that is pararel to the first hyperplane with the distance twice of the length of projection vector. The best hyperplane is then maintained manually where the second bound for the margin is obtained by drawing a line passing through the point from the second class with shortest distance and pararel to the line passing to the reference point.

Keywords: Normal Vector; Orthogonal Projection, Hyperplane; Margin.

Received April 22nd, 2020
Revised June 22nd, 2020
Accepted July 13th, 2020