

BAB 2

TINJAUAN PUSTAKA

2.1 Grafologi

Grafologi berasal dari bahasa Yunani, *graph* yang berarti tulisan dan logos yang berarti ilmu. Grafologi sendiri adalah sebuah cabang dari ilmu psikologi, ilmu grafologi ini sangat bermanfaat untuk menilai dan menganalisa karakter seseorang berdasarkan tulisan tangannya.

Grafologi adalah sebuah seni yang dapat menilai atau memprediksi karakter dan kepribadian seseorang dengan cara melihat tipe tulisan tangan. Menurut Sir Willian Herschel, tulisan tangan mencerminkan kepribadian seseorang, seperti sidik jari yang membuka identitas seseorang. Sementara Pirre Janet (1859-1947) mengemukakan bahwa grafologi sebagai “ilmu masa depan”, karena tulisan tangan mempunyai bentuk dan gaya yang tidak sama antar individu, maka tulisan tangan juga dapat menunjukkan ciri khas kepribadian seseorang.

Grafologi didasarkan pada hubungan antara pikiran bawah sadar seseorang, sementara tulisan tangan dihasilkan dari refleksi pikiran bawah sadar seseorang, sehingga keduanya menjadi dua hal yang dapat dihubungkan. Manfaat dari ilmu grafologi sendiri tidak hanya sebatas menilai dan menganalisa karakter seseorang berdasarkan tulisan tangannya. Di Amerika Serikat dan Eropa grafologi banyak digunakan untuk memilih dan menilai calon karyawannya. Tulisan tangan yang dapat diprediksi sendiri terbagi menjadi beberapa macam, ada tulisan tangan yang berbentuk tanda tangan dan tulisan tangan dalam bentuk kalimat.

2.2 Sejarah Grafologi

Grafologi memiliki sejarah yang telah berumur panjang dengan latar belakang filosofis berkaitan dengan hakikat manusia dan perilaku. Pada tahun 300 M, filsuf Yunani Aristoteles

telah menyampaikan pandangannya mengenai hubungan antara tulisan tangan, kata-kata, dan kepribadian.

Pada abad ke-2 M, C.Suetonius Tranquillus sudah mencatat keunikan dari tulisan tangan Caesar. Pada abad ke-11 M, bangsa Tiongkok juga mencatat hubungan antara tulisan tangan dan

kepribadian. Francois Demelle, seorang berkebangsaan Prancis pertama kali mengenalkan teori yang menyatakan bahwa tulisan tangan merupakan perwujudan dari karakter seseorang. Pada tahun 1622, Camilo Baldi menulis buku tentang Grafologi di Italia, Camilo Baldi adalah doktor Italia di bidang Kedokteran dan Filsafat serta professor di Universitas Bologna.

Ilmu analisis tulisan tangan, atau grafologi yang formal, muncul dari dua pastor Prancis pada abad ke-19, yaitu Abbe Flandrin dan Abbe Michon, pada waktu itulah istilah “Grafologi” mulai diperkenalkan dan digunakan untuk menyebut ilmu yang mempelajari kepribadian berdasarkan tulisan tangan, dari hasil penelitiannya, ditemukan banyak korelasi dengan karakteristik kepribadian penulisnya.

Penelitian Abbe Michon mengenai analisis tulisan tangan dipublikasikan pada tahun 1872. Berawal dari waktu tersebut dan selanjutnya, minat di bidang Grafologi menyebar ke seluruh Eropa. Jadi, grafologi berkembang sebelum kelahiran psikologi.

Penelitian grafologi terus-menerus dilakukan oleh para penyelidik secara individual dan program-program di bawah arahan bantuan organisasi grafologi profesional seperti American Handwriting Analysis Foundation.

2.3 Manfaat Grafologi

Saat ini penggunaan di Indonesia lebih umum menggunakan grafologi di kalangan praktisi SDM (Human Resource) untuk rekrutmen, tes kepribadian, assessment test, dan tes untuk promosi. Namun grafologi dapat digunakan lebih luas untuk pengenalan bakat dan minat,

konseling, kecocokan antarpasangan, konsultasi karier, analisis tindak kriminal, analisis penggunaan narkoba atau alkohol, dan lain-lain.

2.4 Fitur Tulisan Tangan Dalam Grafologi

Analisis tulisan tangan mempunyai enam fitur untuk menggambarkan apakah terdapat kemunculan kepribadian berdasarkan ciri yang ditinjau.

2.4.1 Margin

Margin antara tulisan individu satu dengan yang lainnya berbeda-beda. Ada tulisan dengan margin yang terlihat begitu lebar, tetapi ada juga yang terlihat begitu sempit. Margin dikategorikan menjadi empat, yaitu margin kiri, atas, kanan, dan bawah.

2.4.1.1 Margin Kiri

Margin kiri adalah tepi bagian kiri kertas tempat penulis memulai tulisannya. Sebenarnya menulis dari tepi mana saja tidak dipermasalahkan, tetapi ketidaksadaran diri dalam memilih titik awal penulisan turut mempengaruhi pilihan tersebut. Margin kiri terbagi atas kategori:

- 1) Margin kiri normal mengindikasikan penulis dapat menyikapi masa lalunya dengan wajar, jadi bahan evaluasi, dan menjadi penyemangat untuk hidup yang sekarang.
- 2) Margin kiri lebar dari tepi kertas atau masuk terlalu jauh hampir mendekati tengah kertas, mengindikasikan kebaikan, memiliki kesadaran akan nilai-nilainya sendiri yang jadi suatu kebanggaan.
- 3) Margin kiri sempit hampir mendekati tepi kertas mengindikasikan adanya masa lalu yang membebani dirinya sampai kini, bercermin pada masa lalu dalam bertindak.
- 4) Margin kiri menyempit mengindikasikan penulis punya kemauan dan usaha untuk meninggalkan masa lalunya, upaya bergerak maju untuk masa depan lebih baik, tetapi terkadang masih teringat akan masa lalu.

- 5) Margin kiri melebar mengindikasikan penulis ini berusaha meninggalkan masa lalunya terutama pengalaman-pengalaman buruk agar tidak memengaruhi dirinya sekarang dan nanti.
- 6) Margin kiri tidak teratur mengindikasikan penulis ingin melawan norma, tidak mau mengikuti aturan, tidak stabil, senang dengan permusuhan, tingkat kedisiplinan yang buruk.

2.4.1.2 Margin Kanan

Margin kanan adalah jarak atau tepi bagian kanan kertas dengan huruf terakhir pada kertas tersebut dituliskan. Margin kanan mengungkap masa depan penulis. Margin kanan terbagi atas kategori:

- 1) Margin kanan sempit, mengindikasikan penulis memikirkan rencana-rencana masa depan, yakin masa depan cemerlang, komunikatif, mudah beradaptasi, aktif di lingkungan sosial, dan terbuka.
- 2) Margin kanan lebar, mengindikasikan penulis memiliki kekhawatiran terhadap masa depannya, tidak yakin masa depan sesuai harapan, ragu-ragu, pemalu, menarik diri dari sosial, mudah curiga, takut akan kegagalan.
- 3) Margin kanan terbentur, mengindikasikan penulis sangat bersemangat, menuruti kata hatinya, tidak terkontrol untuk mendapatkan keinginannya, tidak pernah belajar dari kesalahan, tidak peduli sekitar, suka menunda atau menolak yang akhirnya merugikan diri sendiri.

2.4.2 Spasi

Jarak atau spasi dalam tulisan dapat dibedakan menjadi dua, yaitu jarak antarkata dan jarak antarbaris. Kedua macam jarak atau spasi tersebut memiliki penafsiran yang berbeda.

2.4.2.1 Spasi Antarkata

Spasi antarkata adalah jarak antara satu kata dengan kata berikutnya. Spasi antarkata mengungkap interaksi penulis dengan lingkungannya (pergaulan maupun sosial). Spasi antar huruf terdiri atas kategori:

- 1) Spasi antarkata medium, mengindikasikan penulis berada di lingkungan sosial sesuai kebutuhan atau menyendiri bila diperlukan, dapat menyesuaikan diri dengan kedua kondisi tersebut.
- 2) Spasi antarkata sempit, mengindikasikan penulis menyukai hubungan sosial yang intens, nyaman dengan orang banyak.
- 3) Spasi antarkata lebar, mengindikasikan penulis tidak menyukai kegiatan sosial, tidak suka keramaian, tidak merasa nyaman dengan orang banyak, dan mengikuti kegiatan sosial dengan keterpaksaan.

2.4.2.2 Spasi Antarhuruf

Spasi antarhuruf adalah jarak antara satu huruf dengan huruf berikutnya. Spasi antarkata mengungkap kebutuhan interaksi atau keakraban secara fisik pada penulisnya. Spasi antarhuruf terdiri atas kategori:

- 1) Spasi antarhuruf medium, mengindikasikan penulis dapat menerima kontak fisik secara akrab, namun dengan hanya dengan orang-orang yang membuatnya nyaman.
- 2) Spasi antarhuruf sempit, mengindikasikan penulis senang akrab secara fisik, merangkul atau sebaliknya.
- 3) Spasi antarhuruf lebar, mengindikasikan penulis tidak suka kontak atau akrab secara fisik, menginginkan ada ruang antara dirinya dan orang lain.

2.4.2.3 Spasi Antarbaris

Spasi antarbaris adalah ruang diantara baris tulisan di atas dengan baris tulisan dibawahnya. Spasi antarbaris ini menggambarkan kemampuan pengorganisasian, pengaturan, dan kerapian penulisnya. Spasi antarbaris terbagi atas kategori:

- 1) Spasi antarbaris teratur, mengindikasikan bahwa penulis merupakan perencana yang baik (*well planer*), dan orang yang rapi.
- 2) Spasi antarbaris tidak teratur, mengindikasikan bahwa penulis bukan perencana yang baik, dan tidak rapi.

2.4.3 Kemiringan Huruf

Pola ayunan huruf menjadi arah kemana penulis mengekspresikan dirinya, apakah ke dalam atau ke luar diri penulis. Arah huruf juga mengekspresikan perasaan dan emosional penulisnya.

- 1) Arah huruf ke kanan, mengindikasikan penulis senang bergabung dalam kelompok sosial, senang menjalin hubungan dengan orang lain, menganggap hubungan sosial itu penting. Penulis mudah mengekspresikan dirinya, ekstrovert, mudah berempati.
- 2) Arah huruf tegak, mengindikasikan penulis merupakan orang yang rasional, mengutamakan logika daripada perasaan, mampu mengontrol perasaan, mampu mengambil keputusan dengan rasional dan mengabaikan perasaan.
- 3) Arah huruf ke kiri, mengindikasikan penulis membutuhkan kenyamanan dan ruang bagi dirinya, menjaga perasaan dan pikirannya dari orang lain. Penulis merupakan introvert dan berorientasi pada kebutuhan diri.

2.4.4 Garis Dasar

Garis dasar atau dapat disebut *baseline* adalah garis khayal atau imajiner yang terbentuk ketika menarik garis paling kiri menuju paling kanan. Garis dasar adalah pergerakan motivasi, termasuk ada emosional di dalamnya dari masa lalu, sekarang, menuju masa depan. Dalam hal ini garis yang terbentuk menunjukkan tingkat motivasi penulis dalam perjalanan menuju masa depan atau mencapai tujuan hidupnya.

- 1) Garis dasar lurus, mengungkapkan penulis merupakan seseorang yang stabil dalam berperilaku, konsisten, dan terkendali. Seringkali menunjukkan ketenangan, penulis juga memiliki kontrol yang baik terhadap dirinya, termasuk kontrol suasana hatinya.
- 2) Garis dasar menurun, mengungkapkan penulis merupakan seseorang yang lelah, depresi, sedih, atau pesimisme. Seringkali menunjukkan penurunan semangat dan daya juang, ada kekecewaan, keputusasaan, atau ketidakbahagiaan.
- 3) Garis dasar menaik, mengungkapkan semangat yang ceria, ambisi, optimisme, kejujuran, dan motivasi untuk sukses. Penulis mudah bergairah dan mudah mengambil tindakan.

2.4.5 Zona Tulisan Tangan

Zona tulisan tangan dapat dibagi menjadi tiga zona, yaitu zona atas, tengah, dan bawah. Pembagian zona tersebut mengacu pada prinsip antropomorfis, yaitu prinsip pembagian tubuh manusia yang meliputi tiga bagian, diantaranya kepala (atas), badan (tengah), dan kaki (bawah).

- 1) Zona atas dominan, mengindikasikan bahwa penulis memerhatikan pemikiran, gagasan, dan kehidupan spritual. Penulis lebih mengandalkan fikiran daripada fisik, suka berdiskusi tentang teori, konsep, dan gagasan.
- 2) Zona tengah dominan, mengindikasikan bahwa penulis lebih memerhatikan kehidupan, kondisi, penampilan diri saat ini, juga memiliki sosial yang aktif.

- 3) Zona bawah dominan, mengindikasikan bahwa penulis lebih memperhatikan kebutuhan dasar kehidupan seperti makanan, materi, dan seksualitas.

2.5 Pengertian Citra

Sebuah piksel adalah sampel dari pemandangan yang mengandung intensitas citra yang dinyatakan dalam bilangan bulat. Sebuah citra adalah kumpulan dari piksel-piksel yang disusun dalam larik dua-dimensi. Indeks baris dan kolom (x,y) dari sebuah piksel dinyatakan dalam bilangan bulat. Piksel (0,0) terletak pada sudut kiri atas pada citra, indeks x bergerak ke kanan dan indeks y bergerak ke bawah. Konvensi ini dipakai merujuk pada cara penulisan larik yang digunakan dalam pemrograman komputer. [pengolahan citra digital]

Ada tiga jenis citra yang umum digunakan dalam pemrosesan citra. Ketiga jenis citra tersebut adalah citra berwarna, citra berskala keabuan, dan citra biner.[teori pengolahan citra digital]

1) Citra berwarna

Citra berwarna, atau disebut juga sebagai citra RGB, merupakan jenis citra yang menyajikan warna dalam bentuk komponen yang meliputi *Red* (merah), *Green* (hijau), dan *Blue* (biru). Setiap komponen warna menggunakan nilai delapan bit yang berarti nilainya berkisar antara 0 sampai 255. Oleh karena itu, kemungkinan warna yang dapat dihasilkan oleh mencapai $255 \times 255 \times 255$ atau sebanyak 16.581.375 warna.

2) Citra berskala keabuan

Sesuai dengan namanya, citra berskala keabuan hanya menangani gradasi warna hitam dan putih, yang menghasilkan efek warna abu-abu. Pada citra berskala keabuan, warna dinyatakan dengan intensitas. Dalam hal ini, intensitas berkisar antara 0 sampai dengan 255. Nilai 0 menyatakan hitam, sedangkan 255 menyatakan putih.

3) Citra biner

Citra biner adalah citra dengan setiap piksel hanya dinyatakan dengan sebuah nilai dari dua kemungkinan, yaitu antara nilai 0 dan 1. Nilai 0 menyatakan warna hitam dan nilai 1 menyatakan warna putih. Citra jenis ini banyak dipakai untuk pemrosesan citra, misalnya untuk kepentingan memperoleh tepi bentuk suatu objek.

2.6 Pengolahan Citra Digital

Secara umum, menurut Jain, istilah pengolahan citra digital menyatakan “pemrosesan gambar berdimensi-dua melalui komputer digital. Sedangkan menurut Efford pengolahan citra adalah istilah umum untuk berbagai teknik yang keberadaannya untuk memanipulasi dan memodifikasi citra dengan berbagai cara. Foto adalah contoh gambar berdimensi dua yang dapat diolah dengan mudah. Setiap foto dalam bentuk citra digital dapat diolah melalui perangkat lunak tertentu. [teori dan app pengolahan citra].

2.6.1 Grayscale

Citra grayscale seringkali disebut sebagai citra keabuan. Gunanya adalah merubah citra berwarna menjadi citra berskala keabuan, mengingat banyak pemrosesan citra yang bekerja pada skala keabuan. Secara umum citra berwarna dapat dikonversi ke citra berskala keabuan melalui rumus:

$$Grayscale = a * R + b * G + c * B$$

Dimana nilai R menyatakan nilai komponen merah, G menyatakan komponen hijau, dan B menyatakan nilai komponen biru. Sementara untuk nilai a,b,c merupakan nilai tetap. Nilai yang biasa digunakan untuk a,b,c adalah:

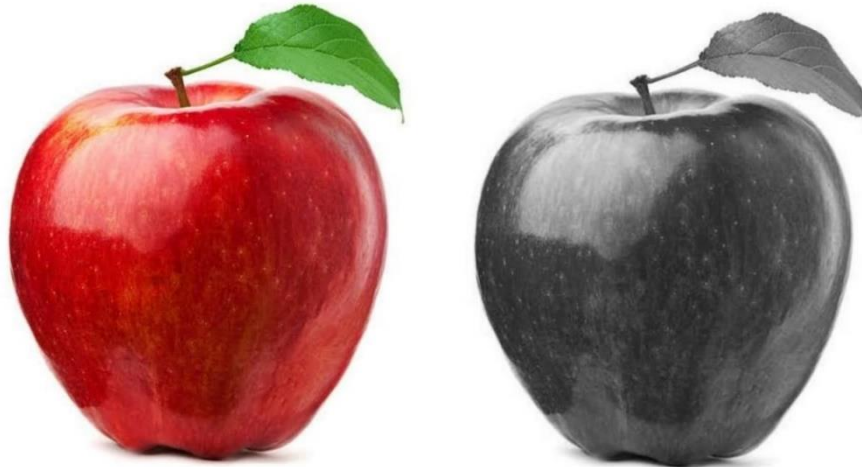
$$a = 0,2989$$

$$b = 0,5870$$

$$c = 0,1141$$

Sehingga didapatkan persamaan (n) yang biasa dipakai untuk merubah citra berwarna menjadi citra berskala keabuan.

$$Grayscale = 0,2989 * R + 0,5870 * G + 0,1141 * B \quad (0.1)$$

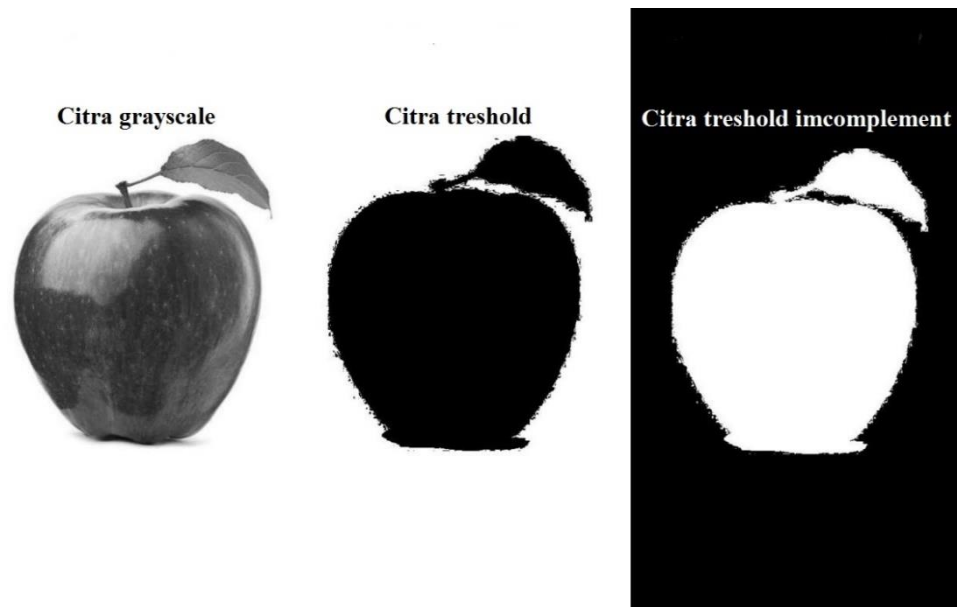


merupakan perbandingan citra RGB dengan citra berskala keabuan.

2.6.2 Threshold

Threshold atau pengambangan merupakan teknik untuk menghasilkan citra biner, yaitu citra yang mempunyai dua tingkat keabuan yaitu hitam dan putih saja. Saat ini terdapat dua teknik pengambangan, yaitu pengambangan global (*global threshold*) dan pengambangan secara lokal adaptif (*local adaptive threshold*), pada penelitian ini akan menggunakan titik tengah atau pengambangan global. Dimana pengambangan global dapat dilakukan dengan persamaan berikut:

$$threshold_{x,y} = \begin{cases} 255 & \text{jika } a_{x,y} \geq 127 \\ 0 & \text{jika tidak} \end{cases} \quad (0.2)$$



Perbandingan citra berskala keabuan dengan citra berskala hitam putih.

2.6.3 Segmentasi

Segmentasi adalah sebuah teknik untuk mengelompokkan citra menjadi beberapa wilayah (region) dengan kriteria tertentu, atau dengan kata lain merupakan proses pembagian citra ke dalam wilayah yang mempunyai kesamaan fitur berupa *grayscale*, *texture*, *color* atau *motion*.

Dari pengertiannya, segmentasi mempunyai tujuan untuk menemukan karakteristik khusus yang dimiliki oleh suatu citra. Oleh karenanya, segmentasi dibutuhkan untuk pengenalan pola dimana jika proses segmentasi baik maka kualitas dalam pengenalan polanya juga akan semakin baik.

2.6.4 Ekstraksi Fitur

Ekstraksi fitur (*feature extraction*) merupakan sebuah bagian fundamental dari analisis citra. Fitur dapat didefinisikan sebagai suatu karakteristik unik dari objek. [D.putra] Ekstraksi fitur juga biasa disebut sebagai pencirian, yang merupakan proses mengubah tiap-tiap sampel sinyal menjadi vektor-vektor data. [jrnl zng]

Fitur yang baik sebisa mungkin harus memenuhi persyaratan berikut: [D.putra]

- 1) Dapat membedakan suatu objek dengan yang lain (*discrimination*).
- 2) Memperhatikan kompleksitas komputasi dalam memperoleh fitur. Kompleksitas komputasi yang tinggi tentu akan menjadi beban tersendiri dalam menemukan suatu fitur.
- 3) Tidak terikat (*independence*) dalam arti bersifat invarian terhadap berbagai transformasi (rotasi, penskalaan, pergeseran, dan lain sebagainya).
- 4) Jumlahnya sedikit, fitur yang jumlahnya sedikit akan dapat menghemat waktu komputasi dan ruang penyimpanan untuk proses selanjutnya.

Metode yang akan digunakan untuk pencirian atau ekstraksi fitur dalam penelitian ini adalah dengan menggunakan metode *zoning*.

2.6.4.1 Metode Zoning

Metode *zoning* memiliki beberapa keunggulan jika dibandingkan dengan metode lainnya, diantaranya adalah penciriannya yang sederhana, kompleksitas yang rendah dan memiliki perhitungan yang cepat.. Menurut Rajashekararadya dan Ranjan (2008) ada beberapa variasi algoritma untuk metode ekstraksi zoning, yaitu ICZ (*image centroid and zone*), ZCZ (*zone centroid zone*), dan algoritma gabungan ICZ dan ZCZ. Pada penelitian ini metode yang digunakan menggunakan metode ZCZ, berikut adalah tahapan metode ZCZ: [14]

- 1) Dengan menggunakan algortima *zone centroid zone* pertama-tama citra masukkan akan dibagi menjadi n zona yang sama. Kemudian, menghitung *zone centroid* dari masing-masing zona menggunakan persamaan 2.7 dan 2.8 berikut:

$$X_c = \frac{\sum_{i=1}^x f(x,y)x_i}{\sum f(x,y)} \quad (0.3)$$

$$y_c = \frac{\sum_{i=1}^y f(x,y)y_i}{\sum f(x,y)} \quad (0.4)$$

Dengan $f(x, y)$ adalah nilai piksel dari citra pada posisi tertentu.

- 2) Kemudian setelah dilakukan perhitungan keseluruhan piksel, hitung jarak antara *centroid* citra dengan setiap piksel yang ada dalam zona menggunakan persamaan 2.9 berikut:

$$d(P, C) = \sqrt{(x_p - x_c)^2 + (y_p - y_c)^2} \quad (0.5)$$

Dengan d adalah jarak antara dua titik, P adalah koordinat titik berat, C adalah koordinat piksel, x_p = koordinat x titik berat, y_p = koordinat y titik berat, x_c adalah koordinat x piksel, dan y_c adalah koordinat y titik berat.

- 3) Setelah keseluruhan perhitungan jarak antara *centroid* dengan setiap piksel dalam zona sudah dilakukan, berikutnya adalah mencari nilai fitur dengan cara merata-ratakan total jarak tersebut dengan persamaan:

$$\text{Rata - rata} = \frac{\text{jarak 1} + \text{jarak 2} + \text{jarak 3} + \dots + \text{jarak n}}{\text{Banyak titik}} \quad (0.6)$$

2.7 Relevance Vector Machine

Pada *supervised learning*, sekumpulan data latih yang terdiri dari himpunan vektor $\{x_n\}_{n=1}^N$ digunakan bersamaan dengan target yang sesuai $\{t_n\}_{n=1}^N$ dimana pada dasarnya target merupakan nilai sebenarnya dalam hal regresi dan label kelas pada proses klasifikasi. Tujuan dari *supervised learning* adalah membuat sebuah model dari target dengan masukan dari *training set*, sehingga prediksi yang akurat terhadap target t dapat dicapai untuk nilai x yang belum diketahui sebelumnya. Secara umum prediksi tersebut dapat dihitung dengan menggunakan persamaan $y(x)$, yaitu:

$$y(x; w) = \sum_{i=1}^M w_i \phi_i(x) + w_0 = w^T \varphi(x) + w_0$$

dimana $w = (w_1, w_2, w_3, \dots, w_m)$ merupakan vektor bobot (*weight vector*) dan $\phi_i(x)$ merupakan fungsi kernel terhadap data x , dan w_0 merupakan *bias*.

Proses belajar atau pelatihan pada dasarnya merupakan sebuah proses untuk menentukan parameter dari fungsi $y(x)$, dalam konteks ini menentukan parameter bobot $w = (w_1, w_2, w_3, \dots, w_m)$ Untuk himpunan dari pasangan data latih $\{x_n, t_n\}_{n=1}^N$ tugas dari proses pelatihan ini adalah untuk mencari nilai dari bobot $w = (w_1, w_2, w_3, \dots, w_m)^T$ sehingga fungsi $y(x)$ dapat digeneralisasi secara cukup baik terhadap data yang baru, dan bobot yang

dihasilkan memiliki elemen nilai bukan nol (*nonzero*) yang sedikit. Memiliki beberapa elemen nilai yang bukan nol dapat menghasilkan representasi vektor yang *sparse*, dengan keuntungan dapat menyediakan implementasi yang cepat .

Relevance Vector Machine adalah sebuah metode pembelajaran mesin yang diperkenalkan oleh Mike Tipping pada tahun 2001, yang diadaptasi dari *Bayesian Framework* dan memiliki bentuk model fungsi yang mirip dengan *Relevance Vector Machine* . Sama halnya dengan *Relevance Vector Machine*, *Relevance Vector Machine* digunakan dalam proses klasifikasi dan regresi. *Relevance Vector Machine* diperkenalkan untuk menutupi beberapa kelemahan yang dimiliki oleh *Relevance Vector Machine*. Kelemahan tersebut antara lain .

- a. Walaupun *Relevance Vector Machine* menghasilkan vektor yang relatif *sparse*, namun jumlah *Relevance Vector* yang ada akan secara linear bertambah seiring dengan ukuran himpunan data latih.
- b. Prediksi yang tidak bersifat probabilistik, dan *Relevance Vector Machine* tidak cocok untuk proses klasifikasi dimana peluang *posterior* dari keanggotaan kelas dianggap penting.
- c. *Relevance Vector Machine* memerlukan proses estimasi parameter *error/margin* C, yang secara umum memerlukan prosedur *cross-validation* yang dapat memperlambat proses komputasi.
- d. Fungsi *kernel* yang digunakan dalam *Relevance Vector Machine* harus memenuhi kondisi *Mercer*.

Pada klasifikasi untuk dua label kelas (*binary classification*), semua target dapat diklasifikasikan kedalam dua kelas, yang dapat direpresntasikan dengan nilai 0 dan 1, seperti $t_n \in \{0,1\}$. Distribusi *bernoulli* dapat diadopsi untuk peluang $p(t|x)$ dalam *framework* probabilistik karena hanya dua nilai kelas yang mungkin, yaitu 0 dan 1 [15]. Dalam kasus klasifikasi, model prediksi mengambil bentuk kombinasi linear dari fungsi basis/kernel yang diubah oleh fungsi *logistic sigmoid* .

$$y(x, w) = \sigma(x; w) = \sigma(w^T \varphi(x))$$

dimana $\sigma(\cdot)$ Merupakan fungsi *logistic sigmoid* yang didefinisikan dengan persamaan

$$\sigma(y) = \frac{1}{1 + \exp(-y)}$$

Berdasarkan definisi dari distribusi *Bernoulli*, *likelihood* terdefiniskan sebagai berikut

$$p(t|w) = \prod_{t=1}^N \sigma\{y(x_n; w)\}^{t_n} [1 - \sigma\{y(x_n; w)\}]^{1-t_n}$$

untuk target $t_n \in \{0,1\}$.

Persamaan *likelihood* dilengkapi dengan sebuah *prior* terhadap parameter (bobot) dalam bentuk

$$p(w|a) = \prod_{t=1}^N \frac{\sqrt{a_i}}{\sqrt{2\pi}} \exp\left(-\left(\frac{a_i w_i^2}{2}\right)\right)$$

dimana $a = (a_1, a_2, a_3, \dots, a_n)^T$ merupakan *hyperparameter* yang diperkenalkan untuk mengontrol kekuatan dari *prior* terhadap parameter bobot yang diasosiasikannya.

Untuk nilai α tertentu, distribusi *posterior* bobot terhadap data dapat dihitung menggunakan aturan *Bayes*, dengan persamaan

$$p(w|t, a) = \frac{p(t|w)p(w|a)}{p(t|a)}$$

dimana $p(t|w)$ adalah *likelihood*, $p(w|a)$ adalah *prior*, dan $p(t|a)$ adalah *evidence*.

Parameter bobot model tidak dapat diperoleh dengan cara analitik, sehingga aproksimasi *Laplacian* digunakan. Sejak $p(w|t, a)$ secara linear proporsional untuk $p(t|w) \times p(w|a)$, dapat dimungkinkan untuk mencari maksimum dari persamaan.

$$\begin{aligned} \ln p(w|t, a) &= \ln\{p(t|w)p(w|a)\} + \ln p(t|a) \\ &= \sum_{n=1}^N \{t_n \ln y_n + (1 - t_n) \ln(1 - y_n)\} + \frac{1}{2} w^T A w \end{aligned}$$

untuk parameter bobot yang paling mungkin w' , dengan $y_n = \sigma\{y(x_n; w)\}^{t_n}$ dan $A = \text{diag}(a_1, a_2, \dots, a_n)$ yang terdiri dari nilai α yang diinisialisasi. Untuk mencari parameter w' pada proses reestimasi, metode *Iteratively Reweight Least-Square* dapat digunakan.

Fungsi *logistic log-likelihood* dapat diturunkan sebanyak dua kali untuk memperoleh matriks *Hessian* dengan persamaan

$$H = \nabla \nabla \ln p(w|t, a) = \varphi_n^T B \varphi_n + A$$

Dimana φ merupakan matriks kernel dan $B = \text{diag}(\beta_1, \beta_2, \dots, \beta_n)$ merupakan matriks diagonal yang elemen matriksnya diperoleh dengan persamaan

$$\beta_n = \sigma\{y(x_n; w')\} \{1 - \sigma\{y(x_n; w')\}\}$$

dimana w' merupakan parameter bobot yang akan direestimasi. Proses reestimasi parameter bobot w didapat dengan persamaan

$$w' = w_n + \lambda \Delta$$

dimana λ merupakan koefisien estimasi, dan Δ merupakan vektor yang diperoleh dari persamaan

$$\Delta = H^{-1} \cdot g$$

dimana H^{-1} adalah invers dari matriks hessian dan g adalah gradien negatif yang diperoleh dari penurunan fungsi *logistic log-likelihood* sebanyak satu kali dengan persamaan

$$g = -\nabla \ln p(z|t, a) = \varphi_n^T e - Aw_n$$

Penjelasan proses estimasi parameter bobot w secara rinci dalam bentuk diagram *flowchart* dapat dilihat pada **Error! Reference source not found.** Setelah estimasi parameter bobot w dilakukan, selanjutnya *hyperparameter* α_j dilakukan reestimasi dengan menggunakan persamaan

$$\alpha_j^{new} = \frac{y_j}{w_j^2}$$

Dimana y_j didefinisikan dengan persamaan

$$y_j = 1 - \alpha_j \sum_{jj}$$

2.8 Perhitungan Akurasi

Perhitungan akurasi digunakan untuk mengetahui seberapa besar performansi yang ada terhadap hasil pengujian. Dalam penelitian ini, pengujian performansi dilakukan dengan menghitung akurasi dan *F1 score*, sementara untuk mendapatkan *precision* dan *recall* akan

menggunakan *Confussion matrix*. *Confussion matrix* adalah salah satu metode yang dipakai untuk menganalisis *tuple* dari data yang berbeda [17].

	<i>Positive</i>	<i>Negative</i>	
<i>Positive</i>	<i>TP</i>	<i>FN</i>	<i>TP+FN</i>
<i>Negative</i>	<i>FP</i>	<i>TN</i>	<i>FP+TN</i>
	<i>TP+FP</i>	<i>FN+TN</i>	

Gambar 0.1 *Confussion matrix*

- 1) Akurasi merupakan proporsi dari total prediksi *true* semua data. Adapun untuk menghitung akurasi fitur dapat menggunakan persamaan berikut: [18]

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \quad (0.7)$$

Keterangan:

TP : *True positive*

FN : *False Negative*

FP : *False Positive*

TN : *True Negative*

- 2) *Precision* merupakan ketepatan dari hasil suatu model. Persamaanya merupakan perbandingan antara *true positive* dengan total data dengan label positif. Adapun untuk menghitung *precision* dapat menggunakan persamaan berikut:

$$Precision = \frac{TP}{TP+FP} \quad (0.8)$$

- 3) *Recall* adalah ukuran kelengkapan dari suatu model. Persamaannya merupakan perbandingan antara *true positive* dan total contoh yang benar-benar positif. Adapun untuk menghitung *recall* dapat menggunakan persamaan berikut:

$$Recall = \frac{TP}{TP+FN} \quad (0.9)$$

- 4) *F-Measure* merupakan *harmonic mean* dari *precision* dan *recall*. Adapun untuk menghitungnya dapat menggunakan persamaan berikut:

$$F1\ Score = \frac{2*(Precision*recall)}{Precision+recall} \quad (0.10)$$

Jika dalam suatu pengujian jumlah kelas yang diklasifikasikan lebih dari dua kelas, maka untuk menghitung *precision* dan *recall* dapat menggunakan persamaan berikut [21]:

$$Precision = \frac{\sum_i^L \frac{TP_i}{TP_i+FP_i}}{L} * 100\% \quad (0.11)$$

$$Recall = \frac{\sum_i^L \frac{TP_i}{TP_i+FN_i}}{L} * 100\% \quad (0.12)$$

Keterangan:

L :Jumlah keseluruhan data yang di uji/jumlah kelas

2.9 Pengertian Python

Python dikembangkan oleh Guido van Rossum pada tahun 1990 di CWI, Amsterdam sebagai kelanjutan dari bahasa pemrograman ABC. Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python merupakan salah satu bahasa pemrograman yang populer di dunia kerja Indonesia.

Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan dengan sitaks kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Pada umumnya Python mendukung multi paradigma pemrograman, seperti pemrograman berorientasi objek, pemrograman imperatif dan pemrograman fungsional

