

BAB 2

LANDASAN TEORI

2.1. Grafologi

Grafologi berasal dari bahasa Yunani, *Graph* artinya tulisan tangan dan *logos* yang artinya ilmu. Grafologi sendiri merupakan salah satu cabang dari ilmu psikologi. Ilmu grafologi sangat berguna untuk mengevaluasi dan menganalisis kepribadian seseorang berdasarkan tulisan tangan seseorang.[1].

Grafologi adalah sebuah seni yang dapat menilai atau memprediksi karakter dan kepribadian seseorang dengan cara melihat tipe tulisan tangan[11]. Menurut Sir William 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[11].

Menurut Sen[12] Grafologi adalah suatu ilmu kuno yang dapat digunakan untuk menganalisis ciri-ciri kepribadian seseorang dengan menggunakan rujukan dari tulisan tangan dan tanda tangan, pada tulisan tangan seseorang dapat dikenali melalui Ciri tulisan tangan seperti garis dasar, ukuran huruf, dan lainnya. Sementara Hashemi[13] mengemukakan bahwa grafologi adalah suatu metode ilmiah untuk mengidentifikasi, mengevaluasi, dan memahami kepribadian seseorang melalui suatu goresan dan pola yang diungkapkan melalui tulisan tangan dan tanda tangan.

Penelitian grafologi terus dilakukan oleh peneliti dan program individu dengan bantuan organisasi grafologi profesional (seperti *American Handwriting Analysis Foundation*).

2.2. Tulisan Tangan dalam ilmu grafologi

Tulisan tangan merupakan salah satu cabang dari ilmu grafologi yang dapat menentukan ciri khas yang dimiliki manusia, gaya tulisan manusia berbeda dengan orang lain, biasanya tulisan tangan digunakan untuk memperoleh

informasi tentang kepribadian dan karakter dari seseorang[1]. Tulisan tangan sering dianggap sebagai indikasi ciri kepribadian yang diwakili oleh pola saraf di otak. Tulisan tangan dapat memberikan gambaran bagi masing-masing penulis. Tulisan tangan sulit dikenali oleh mesin, karena bentuk dan corak tulisan tangan pasti berbeda dengan yang lain, sehingga tulisan tangan dapat menunjukkan ciri kepribadiannya. Inilah salah satu alasan mengapa tulisan tangan dapat dijadikan bahan penelitian untuk memahami kepribadian seseorang. Dengan menganalisis tulisan tangan, seorang ahli grafologi dapat mengetahui kepribadian seseorang.

Menurut P.Joshi dkk[4] mengemukakan bahwa tulisan tangan adalah salah satu ciri khas yang dimiliki manusia dalam mengidentifikasi ciri dari individu dalam menentukan kepribadian seseorang, analisis tulisan tangan dilakukan oleh ahli grafologi dalam menentukan dan memprediksi ciri-ciri kepribadian penulis. Tulisan tangan juga dapat digunakan dalam menentukan kandidat saat proses rekrutmen karyawan karena dalam tulisan tangan memiliki keakuratan yang baik saat dianalisis[11]. Sedangkan S. Hashemi dkk[13] mengemukakan bahwa tulisan tangan dapat mengungkapkan kepribadian seseorang selain itu tulisan tangan dapat mengungkap sifat seperti emosional, ketakutan, kejujuran, dan banyak lainnya.

Menurut A. Bal dkk[14] mengemukakan bahwa analisis tulisan tangan merupakan suatu teknik ilmu yang dapat dikaji secara ilmiah dalam menentukan kepribadian seseorang dari goresan tangan dan pola tulisan tangan penulis, Analisis tulisan tangan termasuk tulisan otak karena seorang penulis dapat mengendalikan pikirannya tentang apa yang ditulis akan tetapi seorang penulis tidak dapat mengontrol cara menulis dengan analisis tulisan tangan, selain itu tulisan tangan dapat memberikan sebuah informasi semua mengenai karakteristik unik yang dimiliki manusia, tulisan tangan juga dapat memberikan informasi tentang keadaan fisik, mental, dan emosional penulis, dalam menentukan suatu kepribadian seseorang penulis harus dibantu oleh pihak grafologi.

2.3. Ciri tulisan tangan dalam ilmu grafologi

Setiap tulisan yang digoreskan memiliki arti yang dapat menunjukkan kepribadian yang berbeda dari setiap penulis. Hal ini dapat digunakan sebagai media untuk

menganalisis kondisi psikologis atau kepribadian seseorang. Ada banyak bentuk tulisan tangan, masing-masing dengan ciri khasnya sendiri. Namun demikian, banyak tulisan yang dapat ditelusuri kembali ke ciri atau Ciri pada tulisan tangan yang dapat menggambarkan sifat dan kepribadian tertentu[1]. Berikut Ciri tulisan tangan yang akan dijelaskan pada subbagian berikut ini:

2.3.1. Spasi

Di dalam tulisan, pemisah atau jarak antar kata disebut dengan spasi. Jarak atau spasi dalam tulisan dapat dibedakan menjadi tiga, yaitu jarak antar kata dan jarak antar baris. Ketiga macam spasi atau jarak tersebut memiliki penafsiran yang berbeda[1].

2.3.1.1. Spasi Antar kata

Spasi antar kata adalah jarak antara satu kata dengan kata berikutnya. Spasi antarkata merupakan cerminan, bagaimana pikiran mengalir begitu saja (alami), seperti sedang berbicara ada jeda dalam pembicaraan. Spasi antarkata juga dapat mengungkap interaksi penulis dengan lingkungannya (pergaulan maupun sosial). Spasi antar kata terdiri atas kategori[1].

Tabel 2. 1 Kepribadian berdasarkan spasi antar kata

| Spasi antar kata | |
|-------------------------|---|
| Ciri | Kepribadian |
| Spasi antar kata sempit | Penulis dapat cepat menjalin relasi dengan orang lain, juga mudah akrab dengan orang lain, pandai bergaul dan disukai di lingkungan sosialnya.. |
| Spasi antar kata medium | Penulis berada di lingkungan sosial sesuai kebutuhan atau menyendiri bila diperlukan, dapat menyesuaikan diri dengan kedua kondisi tersebut. |
| Spasi antar kata lebar | Penulis tidak menyukai kegiatan sosial, tidak suka keramaian, tidak merasa nyaman dengan orang banyak, dan mengikuti kegiatan sosial dengan keterpaksaan. |

2.3.1.2. Spasi Antar huruf

Spasi antar huruf adalah jarak antara satu huruf dengan huruf berikutnya. Spasi antar huruf mengungkap kebutuhan interaksi atau keakraban secara fisik pada penulisnya. Spasi antar huruf terdiri atas kategori[11].

Tabel 2. 2 Kepribadian berdasarkan spasi antar huruf

| Spasi antar huruf | |
|--------------------------------|---|
| Ciri | Kepribadian |
| Spasi antar baris huruf medium | Penulis dapat menerima kontak fisik secara akrab, namun dengan hanya dengan orang-orang yang membuatnya nyaman. |
| Spasi antar huruf sempit | Penulis senang akrab secara fisik, merangkul atau sebaliknya. |
| Spasi antar huruf lebar | Penulis tidak suka kontak atau akrab secara fisik, menginginkan ada ruang antara dirinya dan orang lain. |

2.3.1.3. Spasi Antar baris

Spasi antarbaris adalah ruang diantara baris tulisan di atas dengan baris tulisan dibawah-nya. Spasi antarbaris ini menggambarkan kemampuan pengorganisasian, pengaturan, dan kerapian penulisnya. Spasi antarbaris terbagi atas kategori[11].

Tabel 2. 3 Kepribadian berdasarkan spasi antar baris

| Spasi antar baris | |
|---------------------------|--|
| Ciri | Kepribadian |
| Spasi antar baris teratur | Penulis merupakan perencana yang baik (<i>well planer</i>), dan orang yang rapi. |
| Spasi antar baris tidak | Penulis mempunyai sifat kurang rapih, dan bukan perencana yang baik. |

| | |
|---------|--|
| teratur | |
|---------|--|

2.3.2. Kemiringan Tulisan

Kemiringan tulisan adalah menulis ke kiri atau kanan, atau tidak vertikal dalam bahasa umum. Sudut kemiringan huruf ke kiri atau ke kanan saat ditulis tangan biasanya tidak lebih dari 45 derajat. Pada saat yang sama, tulisan yang sangat miring lebih dari 45 derajat disebut kemiringan ekstrem[1].

Tabel 2. 4 Kepribadian berdasarkan arah kata(kemiringan)

| Kemiringan | |
|---------------------|--|
| Ciri | Kepribadian |
| Kemiringan ke kanan | penulis senang bergabung dalam kelompok sosial, senang menjalin hubungan dengan orang lain, menganggap hubungan sosial itu penting. Penulis mudah mengekspresikan dirinya, ekstrovert, mudah berempati. |
| Kemiringan tegak | penulis merupakan orang yang rasional, mengutamakan logika daripada perasaan, mampu mengontrol perasaan, mampu mengambil keputusan dengan rasional dan mengabaikan perasaan. |
| Kemiringan ke kiri | penulis membutuhkan kenyamanan dan ruang bagi dirinya, menjaga perasaan dan pikirannya dari orang lain. Ini dapat berbentuk isolasi dari dalam diri sendiri, represi, dan abstraksi. Penulis merupakan introvert dan berorientasi pada kebutuhan diri. |

2.3.3. 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[11].

Tabel 2. 5 kepribadian berdasarkan garis dasar

| Garis Dasar | |
|---------------------|--|
| Ciri | Kepribadian |
| Garis dasar lurus | 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. |
| Garis dasar menaik | Mengungkapkan semangat yang ceria, ambisi, optimisme, kejujuran, dan motivasi untuk sukses. Penulis mudah bergairah dan mudah mengambil tindakan. |
| Garis dasar menurun | Penulis merupakan seseorang yang lelah, depresi, sedih, atau pesimisme. Seringkali menunjukkan penurunan semangat dan daya juang, ada kekecewaan, keputusasaan, atau ketidak bahagiaan. |

2.4. Pengertian Citra

Piksel adalah contoh pemandangan, yang berisi intensitas citra atau gambar yang dinyatakan sebagai bilangan bulat. Citra itu sendiri dapat didefinisikan sebagai fungsi dari dua variabel misalnya ax, y di mana a adalah amplitudo atau kecerahan, sedangkan x, y adalah koordinat yang membentuk citra atau koordinat bayangan[15].

Dengan kata lain citra juga dapat diartikan sebagai citra pada salah satu komponen multimedia yang memiliki karakteristik yang tidak dimiliki oleh data teks. Secara harfiah, citra adalah citra atau saluran cahaya dalam bidang dua dimensi (dua dimensi) dari sudut pandang matematis[15].

Secara umum, terdapat tiga jenis citra yang digunakan dalam pemrosesan citra. Ketiga jenis citra tersebut adalah citra berwarna (RGB), citra berskala keabuan (*grayscale*), dan citra biner[15].

2.4.1. Pengolahan Citra

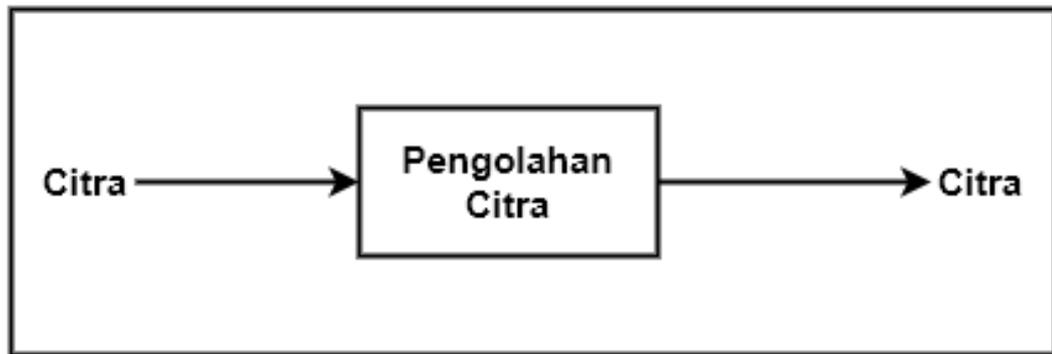
Menurut definisi U. Ahmad[16] pengolahan citra atau *Image Processing* merupakan bidang yang mandiri, karena masyarakat mengenal bahwa komputer tidak hanya dapat mengolah data teks, tetapi juga data citra, sehingga perkembangannya cukup besar. Teknologi pemrosesan gambar biasanya digunakan untuk melakukan konversi dari satu gambar ke gambar lainnya, dan tugas pemulihan informasi diselesaikan dengan kompilasi algoritmik. Pengolahan citra merupakan pengolahan suatu citra menjadi citra lain, dengan tujuan untuk memperoleh kualitas citra yang lebih baik. Selain itu pada pengolahan citra terdapat operasi dalam pengolahan citra yaitu: Perbaikan kualitas citra (*Image enhancement*), yang meliputi:

- a. Pengaturan kontras
 - b. Penegasan tepian objek (*edge enhancement*)
 - c. Penajaman (*sharpening*)
 - d. Pemberian warna semu (*pseudo coloring*)
 - e. Pemfilteran derau (*noise filtering*)
1. Restorasi citra, yang meliputi:
 - a. Penghilangan kesamaran (*deblurring*)
 - b. Penghilangan derau (*derau*)
 2. Kompresi citra yaitu suatu operasi yang memanfaatkan ukuran memori citra dengan tetap mempertahankan kualitas pada sebuah citra (gambar).
 3. Segmentasi citra (*image segmentation*) yaitu suatu operasi dalam memecahkan sebuah citra menjadi beberapa segmen yang berdasarkan pada kriteria tertentu. Jenis operasi ini sangat berkaitan dengan pengenalan pola.
 4. Analisa citra yaitu suatu citra yang bertujuan untuk menghitung besaran kuantitatif dari sebuah citra gambar untuk menghasilkan deskripsinya. Teknik pengolahan citra dapat mengekstraksi ciri-ciri tertentu yang dapat membantu dalam mengidentifikasi sebuah objek. Proses segmentasi

kadangkala diperlukan untuk membatasi sebuah objek yang diinginkan dari sekelilingnya. Operasi pada analisis citra dapat meliputi:

- a. Pendeteksian tepi objek (*edge detection*)
- b. Ekstraksi batas
- c. Representasi daerah

Adapun teknik-teknik dalam pengolahan citra yaitu untuk mentransformasikan citra menjadi citra lain. Jadi, jika masukannya berupa citra (gambar) maka keluaran yang dihasilkan adalah berupa citra (gambar) juga, untuk gambaran dari sebuah teknik pengolahan citra dapat dilihat pada Gambar 2.1 seperti berikut:



Gambar 2. 1 Teknik Pengolahan Citra

2.4.2. Pengolahan Citra Digital

Definisi Pengolahan Citra Digital menurut U. Ahmad[16] Pengolahan citra digital adalah penggunaan komputer digital untuk memproses citra digital melalui suatu algoritma. Sebagai subkategori atau bidang pengolahan sinyal digital, pengolahan citra digital memiliki banyak keunggulan dibandingkan pengolahan citra analog. Hal ini memungkinkan algoritma yang jauh lebih luas untuk diterapkan pada data input dan dapat menghindari masalah seperti penumpukan kebisingan dan distorsi selama pemrosesan. Karena gambar didefinisikan melalui dua dimensi, pemrosesan gambar digital dapat dimodelkan dalam bentuk sistem multidimensi. Generasi dan perkembangan pemrosesan citra digital terutama dipengaruhi oleh tiga faktor: pertama, perkembangan komputer; kedua, perkembangan matematika; ketiga, permintaan untuk berbagai aplikasi di lingkungan, pertanian, militer, industri dan ilmu kedokteran telah meningkat.

Definisi Pengolahan Citra Digital menurut R. Munir[17] adalah sebuah ilmu yang mempelajari tentang teknik-teknik mengolah citra digital dengan melakukan operasi-operasi pemrosesan sinyal dengan menggunakan komputer. Citra digital adalah suatu representasi citra melalui pencuplikan (*sampling*) secara ruang dan waktu, dimana pencuplikan secara ruang yaitu berdasarkan titik koordinat sinyal (x, y) , sedangkan pencuplikan secara waktu yaitu sederetan citra yang bergerak atau sering disebut dengan video digital.

2.4.3. Jenis Citra

Jenis citra dibagi menjadi tiga jenis citra, pada umumnya ada tiga jenis citra yang sering digunakan dalam pengolahan citra digital. Ada tiga Jenis citra seperti berikut:

1. Citra berwarna

Citra berwarna atau biasa disebut juga sebagai citra RGB, citra ini memiliki tiga komponen warna. Warna-warna tersebut adalah Red (merah), Green (hijau), dan Blue (biru)[15]. Dimana masing-masing warnanya menggunakan nilai 8bit yang berarti nilainya berkisar antara 0 sampai 255. Sehingga nilai yang dapat dihasilkan oleh citra berwarna adalah $255 \times 255 \times 255$ atau sebanyak 16.581.375 warna. Dapat dilihat pada Tabel 2. 6 berikut yang telah dijelaskan mengenai contoh citra warna RGB[17].

Tabel 2. 6 Contoh warna RGB

| Warna | R | G | B |
|--------|-----|-----|-----|
| Merah | 255 | 0 | 0 |
| Hijau | 0 | 255 | 0 |
| Biru | 0 | 0 | 255 |
| Hitam | 0 | 0 | 0 |
| Putih | 255 | 255 | 255 |
| Kuning | 0 | 255 | 255 |

2. Citra berskala keabuan

Citra berskala keabuan, memiliki nilai yang serupa dengan citra berwarna, hanya saja pada citra berskala keabuan nilai yang ada menyatakan nilai keabuan nya, dimana nilai 0 yang berarti hitam, dan nilai 255 yang berarti putih[17].

3. Citra Biner

Citra biner, citra jenis ini banyak digunakan dalam pemrosesan citra, misalnya untuk memperoleh tepi bentuk suatu objek. Citra ini hanya memiliki dua jenis nilai, yaitu nilai 0 yang berarti hitam dan nilai 1 yang berarti putih, karena hanya ada dua warna untuk setiap piksel, maka hanya perlu 1 bit per piksel (0 dan 1) atau apabila dalam 8 bit (0 dan 255), sehingga sangat efisien dalam hal penyimpanan. Gambar yang direpresentasikan dengan biner sangat cocok untuk teks (dicetak atau tulisan tangan), sidik jari (finger print), atau gambar arsitektur[17].

2.5. Penelitian Sebelumnya

Dalam penelitian ini akan membahas mengenai identifikasi kepribadian seseorang berdasarkan tulisan tangan yang terkomputerisasi, dalam penelitian awal pada sistem grafologi yang terkomputerisasi pernah dilakukan oleh Sheikholeslami dkk pada tahun 1995, dalam penelitiannya yang telah ditulis oleh mereka mengembangkan suatu sistem grafologi yang disebut *Computer Aided Graphology (CAG)*, dalam sistem tersebut terdiri dari beberapa modul yaitu: Pemindaian, Preprocessing, Ciri ekstraksi, dan analisis Ciri. Dalam penelitian yang sudah pernah dikaji mengenai tulisan tangan ini ada beberapa penulis yang menjadi rujukan dalam penelitian ini seperti yang pernah dilakukan oleh S.Prasad[2] yang mengulas sebuah metode otomatis yang dapat memprediksi psikologis kepribadian seseorang, pada penelitian ini menggunakan metode segmentasi yang digunakan untuk menghitung Ciri dari tulisan tangan dan dilatih melalui metode klasifikasi SVM yang bertujuan untuk menghasilkan suatu perilaku atau kepribadian penulis, Ciri yang di fokuskan pada penelitian[2] adalah ukuran huruf, kemiringan huruf dan kata-kata, baseline, tekanan pena, jarak antara

huruf dan kata dalam dokumen. Metode yang diusulkan dapat memberikan tingkat akurasi sebesar 94% dari 100 dataset tulisan tangan.

Selain itu Fadhilla dkk[18] mengusulkan sebuah sistem untuk pengenalan kepribadian berdasarkan beberapa Ciri tulisan tangan yaitu seperti margin kiri, margin kanan, ukuran, kemiringan, dan bentuk huruf. Selain itu pada penelitian [18] bertujuan untuk membandingkan dua metode antara metode Jaringan Saraf Tiruan dan *Learning Vector Quantization*, Kepribadian dari Ciri tulisan tangan lalu diklasifikasikan berdasarkan *Myers-Briggs Type Indicator* (MBTI) menggunakan metode *Back Propagation* dan *Learning Vector Quantization* (LVQ). Metode yang diusulkan dapat memberikan tingkat akurasi sebesar 90% untuk metode LVQ dan 82% untuk metode *Black Propagation*.

Pada tahun 2019, Fatimah dkk[19] mengusulkan sebuah penelitian untuk identifikasi kepribadian seseorang berdasarkan tulisan tangan yang menggunakan metode CNN untuk klasifikasi kepribadian dari Ciri yang dikaji, pada penelitian [19] memfokuskan enam Ciri yang dikaji yaitu seperti margin, spasi antar garis, spasi antara kata, kemiringan, dan zona dominan, beserta empat huruf spesifik seperti a, g, s, t. metode yang diusulkan mendapatkan tingkat akurasi sebesar 98%.

Pada tahun 2015, Joshi dkk[4] mengusulkan alat pembelajaran mesin dengan Klasifikasi KNN dengan pembelajaran tambahan untuk meningkatkan efisiensi analisis tulisan tangan. Pada penelitian [4] memfokuskan Ciri yang dikaji adalah Ciri dasar, miring huruf, tinggi “t” bar, margin digunakan untuk analisis kepribadian. Baseline dihitung menggunakan metode poligon, margin dihitung menggunakan metode pemindaian vertikal, ketinggian t-bar pada batang alfabet “t” dan kemiringan kata dihitung menggunakan *template matching*. Akan tetapi metode yang diusulkan tidak menyebutkan nilai akurasi nya.

Pada tahun 2019, Chitlangia dkk[3] mengusulkan sistem yang terkomputerisasi untuk memprediksi sifat-sifat karakteristik manusia melalui gaya tulisan tangan penulis, metode yang digunakan pada penelitian ini untuk klasifikasi menggunakan metode SVM. SVM mengambil hasil nilai dari Ciri yang diekstraksi menggunakan metode ekstraksi HOG dan mengklasifikasikan ciri-ciri kepribadian penulis menjadi 5 ciri kepribadian (*Energic, Extrovert, Introvert,*

Sloppy dan Optimistic). Hasil akurasi dari metode usulan mendapatkan nilai akurasi sebesar 80%.

Pada tahun 2018, Asra dkk[20] mengusulkan dalam penelitiannya mengenai sebuah sistem untuk pengakuan perilaku manusia berdasarkan tulisan tangan dengan menggunakan metode klasifikasi SVM dengan pendekatan Ciri zoning, Ciri yang dikaji yaitu huruf "O", data set yang digunakan pada penelitian ini sebanyak 500 sampel data tulisan dan mendapatkan tingkat akurasi sebesar 86,66%.

Adapun penelitian yang sudah pernah dilakukan menggunakan metode SSVM, namun penelitian tersebut diterapkan pada kasus lain seperti prediksi perilaku pasien psikiatris dalam pencegahan bunuh diri yang dilakukan oleh Indrawan dkk[21], pada penelitian ini menggunakan 30.660 rekam medis pasien dari lima tahun terakhir. Setelah itu terdapat pembersihan data dan mendapatkan 2665 data yang relevan untuk penelitian ini, termasuk 111 pasien yang memiliki perilaku terkait bunuh diri dan dalam perawatan aktif. Data yang dibersihkan kemudian ditransformasikan menjadi sepuluh variabel prediktor dan variabel respons. Pelatihan pemisahan dan pengujian data pada data yang diubah tersebut dilakukan untuk membangun dan evaluasi akurasi model metode. Berdasarkan percobaan, menggunakan metode yang diusulkan mendapatkan nilai akurasi rata-rata terbaik pada 63% dapat diperoleh dengan menggunakan 30% dari data yang relevan.

Pada tahun 2017, Fauzi dkk[6] melakukan penelitian menggunakan metode SSVM pada kasus pengklasifikasian indeks pembangunan manusia kabupaten/kota se-Indonesia, pada penelitiannya dengan metode yang diusulkan peneliti mendapatkan nilai tingkat keakuratan prediksi pada metode ini sebesar 84,77%.

Pada tahun 2017, Nugroho dkk[5] mengusulkan metode SSVM dalam penelitiannya pada kasus klasifikasi pasien diabetes mellitus, dari hasil metode yang diusulkan hasilnya menunjukkan bahwa klasifikasi SSVM menggunakan fungsi kernel Gaussian RBF, dapat mengklasifikasikan 98 dari 110 data pasien Diabetes Mellitus dengan benar sesuai dengan kelas aslinya dan mendapatkan

akurasi yang sangat baik yaitu sebesar 97,03%. Akan tetapi penerapan pada metode SSVM ini belum diterapkan pada kasus prediksi kepribadian berdasarkan tulisan tangan.

Adapun penelitian yang menggunakan metode ekstraksi Ciri DBFE, yang pernah diteliti oleh J. Pradeep dkk[7] mengusulkan metode DBFE dalam kasus pengenalan tulisan tangan alfabet dengan kombinasi metode klasifikasi ANN yang menguji 570 data karakter alfabet tulisan tangan yang berbeda karakter, dalam pengujian menggunakan metode usulan peneliti mendapatkan akurasi sebesar 97,8% dari 54 Ciri dan 98,5% dari 69 Ciri yang diuji.

Pada tahun 2017, K. Vijayalakshmi dkk[22] mengusulkan suatu sistem pengenalan karakter tulisan tangan dengan metode DBFE, data sampel yang digunakan untuk diuji pada metode ini yaitu sebanyak 26 sampel huruf alfabet dari 26 Ciri, akan tetapi pada penelitian ini tidak disebutkan tingkat akurasi yang didapatkan. Sama halnya dengan metode SSVM, metode DBFE pun belum pernah diterapkan pada kasus prediksi kepribadian berdasarkan tulisan tangan. Berikut tabel hasil penelitian sebelumnya yang sudah dikaji dapat dilihat pada Tabel 2. 7.

Tabel 2. 7 Hasil Penelitian sebelumnya

| No | Penulis dan Tahun | Metode | Study Kasus | Dataset | Akurasi | Ciri yang dikaji | Persamaan | Perbedaan |
|----|-------------------------|------------------------|---|-------------------------|------------------|--|---|---|
| 1. | Shita Prasad, (2010) | - Segmentasi, - SVM | - Pengenalan kepribadian berdasarkan tulisan tangan | 100 data tulisan tangan | 94% | - Ukuran huruf - Kemiringan kata - Tekanan huruf - Spasi antar baris - Spasi antar kata - Garis dasar | - Study kasus - Jumlah Ciri tulisan tangan - 3 Ciri (kemiringan kata, tekanan huruf, garis dasar) | - 3 Ciri (coretan awal, coretan akhir, zona) - Metode klasifikasi - Metode ekstraksi Ciri - Jumlah dataset |
| 2. | Mutia Fadhillah, (2017) | - JST/ANN - LVQ. | - Pengenalan kepribadian berdasarkan tulisan | 100 data tulisan tangan | - 90 % - 82 % | - margin kiri, - margin kanan, - ukuran, | - Study kasus - 1 Ciri (kemiringan) | - Jumlah Ciri - 5 Ciri tulisan tangan (coretan awal, coretan |

| | | | | | | | | |
|----|------------------------------------|-------|---|-------------------------------|--------------|---|---|--|
| | | | tangan | | | <ul style="list-style-type: none"> - kemiringan - bentuk huruf | kata) | <ul style="list-style-type: none"> akhir, zona, tekanan, garis dasar) - Metode klasifikasi - Metode ekstraksi Ciri. - Jumlah dataset |
| 3. | Siti Hastuti Fatimah, (2019) | - CNN | - Pengenalan kepribadian berdasarkan tulisan tangan | 100 data tulisan tangan | - 98.03 % | <ul style="list-style-type: none"> - margin, - spasi antar garis, - spasi antara kata, - kemiringan, - zona dominan, | <ul style="list-style-type: none"> - Study Kasus - Jumlah Ciri - 2 Ciri (zona, kemiringan kata) | <ul style="list-style-type: none"> - 4 Ciri (coretan awal, coretan akhir, tekanan, garis dasar) - Metode klasifikasi - Metode ekstraksi Ciri |

| | | | | | | | | |
|----|---------------------------|----------------|---|-------------------------|--------------------|---|---|---|
| | | | | | | - huruf spesifik seperti a, g, s, t. | | |
| 4. | Prachi Joshi, (2015) | - KNN | - Pengenalan kepribadian berdasarkan tulisan tangan | 100 data tulisan tangan | - Tidak disebutkan | - Garis dasar, - Kemiringan huruf, - Ketinggian huruf "t" - Margin | - Study Kasus - 2 Ciri (garis dasar, kemiringan huruf) | - Jumlah Ciri yang dikaji. - 4 Ciri (coretan awal, coretan akhir, tekanan, zona) |
| 5. | Aditya Chitlangia, (2019) | - SVM - HOG | - Pengenalan kepribadian berdasarkan tulisan tangan | 50 data tulisan tangan | - 80% | - 4 Ciri tulisan tangan yang dikaji yaitu: - Arah tulisan, - Jarak tulisan (tulisan melebar), | - Study Kasus - 1 Ciri (kemiringan kata) | - Jumlah Ciri yang dikaji - 5 Ciri (Coretan Awal, Coretan Akhir, kemiringan kata, zona, tekanan, dan garis dasar). |

| | | | | | | | | |
|----|---------------------|--------|---|---|-----------|---|----------------------|---|
| | | | | | | <ul style="list-style-type: none"> - Ukuran huruf - Kemiringan kata | | <ul style="list-style-type: none"> - Metode Klasifikasi - Metode Ekstraksi Ciri |
| 6. | Syeda Asra, (2017) | - SVM | - Pengenalan kepribadian berdasarkan tulisan tangan | 500 data tulisan tangan huruf "O" | - 86,66 % | - 1 Ciri tulisan tangan yang dikaji yaitu Huruf "O" | - Study kasus | <ul style="list-style-type: none"> - Metode klasifikasi - Metode ekstraksi Ciri - Jumlah dataset - Ciri yang dikaji |
| 7. | G. Indrawan, (2018) | - SSVM | - Prediksi Perilaku Terkait Bunuh Diri | 30.660 rekam medis pasien dari lima tahun terakhir. | - 63% | - | - Metode Klasifikasi | <ul style="list-style-type: none"> - Study Kasus - Metode ekstraksi Ciri - Output penelitian - Jumlah dataset |

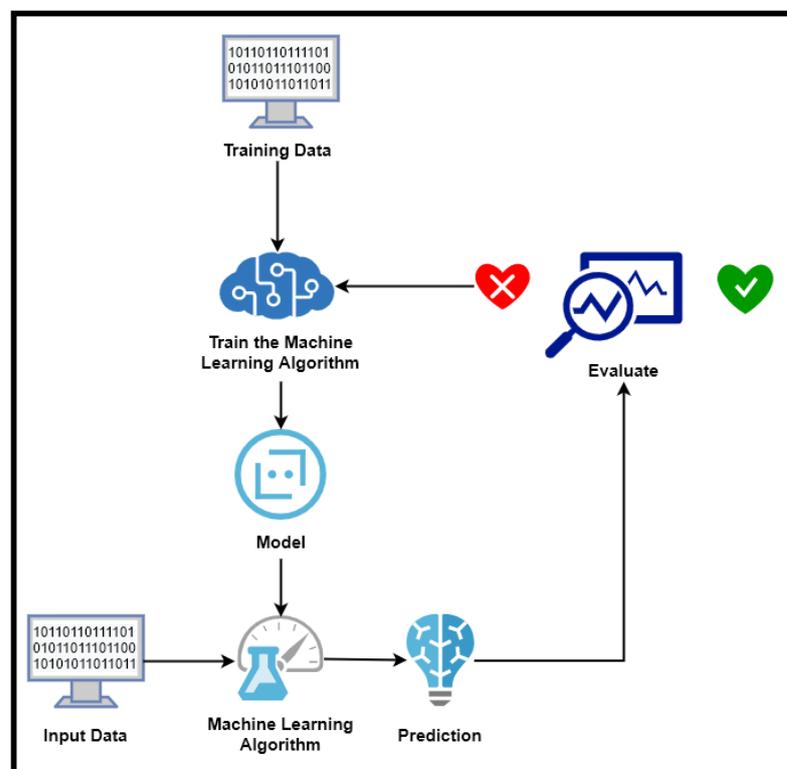
| | | | | | | | | |
|----|------------------------------|--------|--|--|------------|---|----------------------|--|
| | | | | | | | | - Dataset yang digunakan dalam penelitian |
| 8. | Fatkhurokhaman Fauzi, (2017) | - SSVM | - Pengklasifikasian indeks pembangunan manusia di kabupaten/kota se-Indonesia. | Jumlah data pada penelitian ini tidak disebutkan | - 84,77 %. | - | - Metode Klasifikasi | - Study Kasus - Metode ekstraksi Ciri - Output penelitian - Jumlah dataset - Dataset yang digunakan dalam penelitian |
| 9. | Rizky Adhi Nugroho, (2017) | - SSVM | - Klasifikasi Pasien Diabetes Mellitus Menggunakan | 110 data pasien Diabetes Mellitus | - 97% | - | - Metode Klasifikasi | - Study Kasus - Metode ekstraksi Ciri - Output penelitian |

| | | | | | | | | |
|-----|--------------------------|-----------------|--|--|--------------------|------------------------|-------------------------|---|
| | | | n Metode <i>Smooth</i> <i>Support</i> <i>Vector</i> <i>Machine</i> | | | | | - Jumlah dataset - Dataset yang digunakan dalam penelitian |
| 10. | Jayabala Pradeep, (2011) | - NN - DBFE | - Pengenalan karakter tulisan alfabet | 570 data karakter alfabet tulisan tangan | - 97,8% - 98,5% | - 54 Ciri - 69 Ciri | - Metode Ekstraksi Ciri | - Study Kasus - Metode Klasifikasi - Output penelitian - Jumlah dataset - Dataset yang digunakan dalam penelitian |
| 11. | K. Vijayalakshmi, (2017) | - ANN - DBFE | - Pengenalan karakter tulisan | 26 data sampel huruf | - Tidak disebutkan | - 29 Ciri | - Metode Ekstraksi Ciri | - Study Kasus - Metode |

| | | | | | | | | |
|--|--|--|---------|---------------------|--|--|--|---|
| | | | alfabet | karakter alfabet | | | | Klasifikasi - Output penelitian - Jumlah dataset - Dataset yang digunakan dalam penelitian |
|--|--|--|---------|---------------------|--|--|--|---|

2.6. Machine Learning

Definisi pembelajaran mesin atau *Machine Learning* menurut D. Michie[23] adalah suatu prosedur komputasi yang berbasis dari operasi logika atau biner, adapun pengertian lain mengenai machine learning yaitu suatu disiplin ilmu pada kecerdasan buatan (*Artificial Intelligence*) yang menggunakan teknik pendekatan statistika untuk menghasilkan suatu model otomatis dari sekumpulan data, dengan tujuan memberikan komputer kemampuan untuk "belajar" dalam melakukan proses input-output tanpa menggunakan kode program yang dibuat secara eksplisit, proses pembelajaran yang dilakukan oleh *machine learning* yaitu menggunakan suatu kumpulan data yang disebut *train dataset*[23]. Machine Learning bertujuan untuk menghasilkan suatu ekspresi klasifikasi yang cukup sederhana agar dapat dimengerti dengan mudah oleh manusia. Mereka harus meniru penalaran manusia secara memadai untuk memberikan wawasan tentang proses pengambilan keputusan[24]. Adapun cara kerja pada *machine learning* dapat dilihat pada Gambar 2. 2, seperti berikut:



Gambar 2. 2 Cara kerja Machine Learning

2.7. Pre-Processing

Tahap *preprocessing* merupakan penjelasan tahap awal dalam alur kerja untuk menyiapkan citra sebagai data input yang siap diolah lebih lanjut oleh sistem. Proses pada pra-proses yang digunakan pada penelitian ini sebagai berikut:

2.4.4 2.7.1. Grayscale

Citra *grayscale* atau juga citra berskala keabuan. Gunanya adalah merubah citra berwarna menjadi citra berskala keabuan. Secara umum perubahan citra berwarna menjadi citra keabuan dengan menggunakan rumus:

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

Dimana (R) adalah nilai pada warna merah, (G) adalah nilai pada warna hijau, dan (B) adalah nilai pada warna biru. Sementara untuk (a,b,c) adalah nilai tetap, nilai tetap yang biasa dipakai untuk (a,b,c) adalah:

$$a = 0.2989$$

$$b = 0.5870$$

$$c = 0.1141$$

Sehingga didapatkan persamaan 2.1 untuk merubah citra berwarna menjadi citra berskala keabuan adalah sebagai berikut [16].

$$\text{Grayscale} = 0.2989 * R + 0.5870 * G + 0.1141 * B \quad (2.1)$$

Dimana nilai R menyatakan nilai piksel merah, G menyatakan piksel hijau, dan B menyatakan nilai komponen biru. Contoh hasil dari teknik *grayscale* sendiri telah tergambar pada Gambar 2. 3.



Gambar 2. 3 Contoh hasil teknik grayscale

2.4.5 2.7.2. Threshold

Threshold atau biasa disebut sebagai pengambangan citra merupakan teknik untuk menghasilkan citra biner, yaitu citra yang hanya mempunyai dua tingkat keabuan (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 digunakan sebuah titik tengah atau pengambangan global. Dimana pengambangan global dapat dilakukan dengan persamaan berikut[16]:

$$\mathbf{threshold}_{xy} = \begin{cases} 255 & \text{jika } a_{xy} \geq 127 \\ 0 & \text{jika tidak} \end{cases} \quad (2.2)$$

Contoh dari gambar *threshold* sendiri telah tergambarkan pada Gambar 2. 4



Gambar 2. 4 Contoh hasil teknik threshold

2.4.5.1 2.7.2.1. Metode Otsu Threshold

Definisi Metode Otsu menurut N. Otsu[25] dalam jurnal ilmiahnya adalah salah satu metode untuk segmentasi pada citra digital dengan menggunakan nilai ambang secara otomatis, yaitu mengubah nilai citra berwarna abu menjadi citra hitam putih berdasarkan perbandingan nilai ambang dengan nilai warna pada piksel digital. Tahapan dari proses perhitungan menggunakan metode otsu sebagai berikut:

1. Nilai ambang dinyatakan dengan k . Nilai k berkisar antara 1 sampai dengan L , dengan nilai $L = 255$
2. Dalam menentukan probabilitas pada setiap piksel ke- i , dinyatakan pada persamaan 2.3.

$$p_i = \frac{n_i}{N} \tag{2.3}$$

Keterangan:

p_i = Probabilitas piksel ke i

n_i = Jumlah piksel dengan tingkat keabuan i

N = Total Jumlah piksel pada citra

3. Mencari nilai *Zeroth cumulative moment* terdapat pada persamaan 2.4.

$$\omega(k) = \sum_{i=1}^k p_i \tag{2.4}$$

4. Mencari nilai *First cumulative moment* terdapat pada persamaan 2.5.

$$\mu(k) = \sum_{i=1}^k i \cdot p_i \tag{2.5}$$

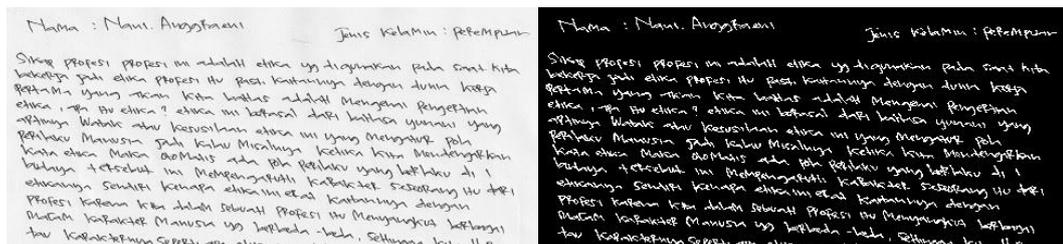
5. Mencari nilai *mean* terdapat pada persamaan 2.6.

$$\mu_T(k) = \sum_{i=0}^L i \cdot p_i \tag{2.6}$$

6. Mencari nilai ambang k ditentukan dengan memaksimumkan persamaan 2.7.

$$\sigma_B^2(k) = \frac{[\mu_T \omega(k) - \mu(k)]^2}{\omega(k)[1 - \omega(k)]} \tag{2.7}$$

Untuk hasil dari metode *otsu thresholding* dapat dilihat pada Gambar 2. 5, berikut:



Gambar 2. 5 Contoh penerapan metode otsu threshold

2.4.6 2.7.3. Segmentasi

Segmentasi adalah suatu teknik yang membagi suatu citra menjadi beberapa daerah sesuai dengan kondisi tertentu, dengan kata lain merupakan proses membagi suatu citra menjadi daerah-daerah yang memiliki karakteristik yang sama (grayscale, tekstur, warna, atau bentuk gerak). Dari pemahaman tersebut,

tujuan segmentasi adalah untuk menemukan ciri-ciri khusus dari citra. Oleh karena itu, pengenalan pola perlu dilakukan segmentasi, jika proses segmentasi nya baik maka kualitas pengenalan pola akan semakin baik. Pada penelitian ini akan menggunakan segmentasi baris[16].

Segmentasi baris dilakukan untuk mencari Ciri baris pada tulisan, segmentasi baris pada penelitian ini dijadikan acuan untuk penentuan Ciri yang lainnya. Segmentasi baris dilakukan dengan pencarian batas atas baris dan batas bawah dengan cara segmentasi horizontal. Piksel putih yang ditemukan pertama akan dijadikan batas atas, dan piksel putih akan dijadikan batas bawah jika setelahnya hanya ditemukan baris berisi piksel hitam, isi karakter diantara batas awal dan akhir akan dinyatakan sebagai baris, proses ini terus diulang hingga tidak ditemukan lagi piksel putih.

2.7.4. Resize

Resize atau penskalaan adalah sebuah operasi geometri yang digunakan untuk memperbesar atau memperkecil ukuran dari sebuah citra sesuai dengan ukuran yang dibutuhkan. Untuk melakukan penskalaan pada penelitian ini tidak digunakan metode khusus. Untuk mendapatkan ukuran yang diinginkan maka akan dilakukan perbandingan ukuran antara citra segmentasi dengan target ukuran citra dengan menggunakan persamaan 2.8 untuk mendapatkan nilai dari titik x dan persamaan 2.9 untuk mendapatkan nilai dari titik y . Dimana kedua persamaan tersebut adalah sebagai berikut:

$$x = \frac{pb * pp}{pa} \quad (2.8)$$

Keterangan:

x = Posisi x baru

pb = Ukuran panjang dari matriks baru.

pp = Posisi piksel x lama.

pa = Ukuran panjang dari matriks lama.

$$y = \frac{lb * pp}{la} \quad (2.9)$$

Keterangan:

y = Posisi y baru

lb = Ukuran lebar dari matriks baru.

pp = Posisi piksel y lama.

la = Ukuran lebar dari matriks lama.

2.8. Ekstraksi Ciri

Definisi Ekstraksi Ciri menurut Xu[24] dalam bukunya mengungkapkan bahwa Ekstraksi Ciri adalah dasar untuk intrusi kinerja tinggi deteksi menggunakan penggalian data karena model deteksi harus dioptimalkan berdasarkan pada pemilihan ruang Ciri. Jika Ciri tidak dipilih dengan benar, kinerja tertinggi model deteksi akan banyak dipengaruhi.

Ekstraksi Ciri adalah proses mengubah data masukan menjadi kumpulan Ciri untuk mengambil representasi terkecil dari data masukan. Ekstraksi Ciri adalah proses memperoleh Ciri yang terdapat dalam objek citra. Dalam proses ini, pertama-tama perlu mendeteksi tepi semua objek pada citra, kemudian menghitung atribut objek terkait sebagai Ciri, kemudian menggunakan teknologi ekstraksi Ciri untuk mengurangi dimensi data sampel masukan. Pengurangan output direpresentasikan sebagai vektor Ciri[3].

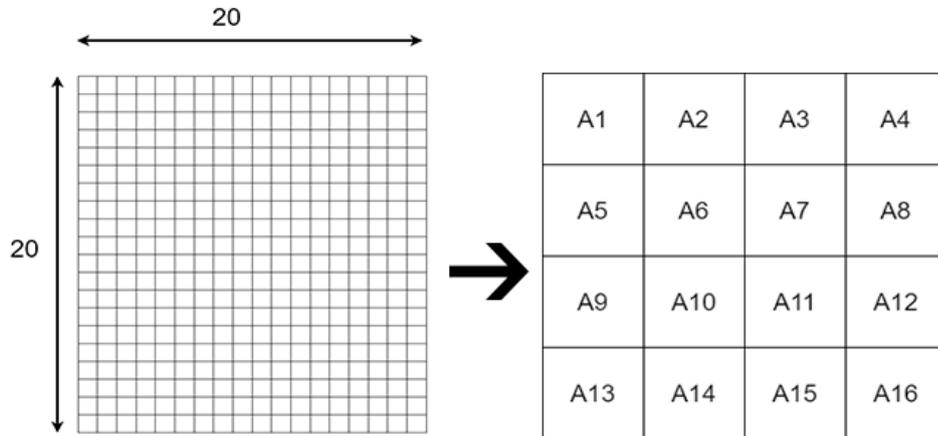
Karakteristik Ciri yang baik sebisa mungkin memenuhi persyaratan sebagai berikut[3]:

- a. Dapat membedakan suatu objek dengan yang lainnya.
- b. Kompleksitas komputasi yang tidak terlalu rumit
- c. Tidak terikat invariant terhadap transformasi
- d. Jumlahnya sedikit

2.8.1. Diagonal Based Feature Extraction

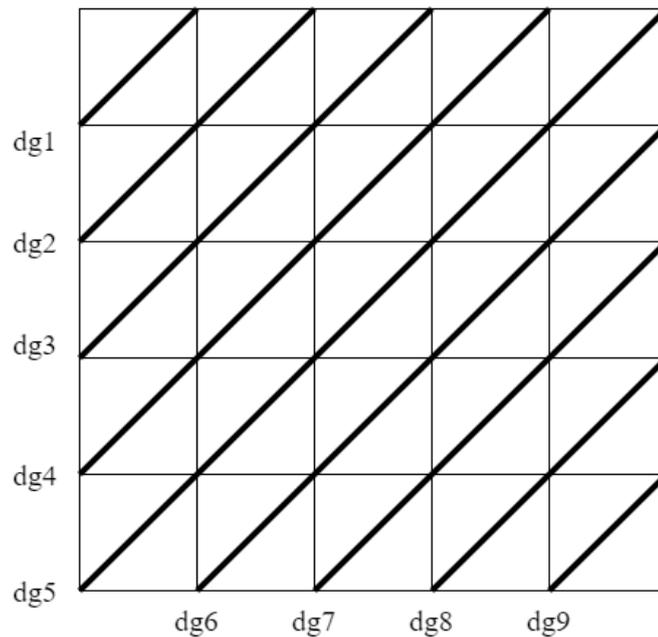
Definisi *Diagonal Based Feature Extraction* menurut Pradeep[7] adalah algoritma ekstraksi Ciri yang dapat membagi ukuran piksel gambar menjadi piksel yang lebih kecil dan seragam. Misalnya membagi gambar karakter dengan ukuran 100x100 piksel menjadi 100 bidang yang sama. Ukuran setiap area adalah 10x10 piksel, dan Ciri diekstraksi dari setiap area dengan bergerak secara diagonal dari setiap 10x10 piksel. Ada 19 diagonal di setiap area, dan piksel latar depan di setiap diagonal ditambahkan bersama untuk mendapatkan sub fungsi. 19 nilai sub-Ciri ini akan diringkas untuk mendapatkan nilai Ciri tunggal dan ditempatkan di area yang sesuai. Prosedur ini diulangi untuk semua zona atau wilayah[7].

Sedangkan menurut Vijayalakshmi[22] *Diagonal Based Feature Extraction* adalah ekstraksi Ciri berbasis diagonal yang digunakan untuk mendapatkan informasi berupa piksel gambar di sepanjang dua diagonal yang diekstraksi dan digunakan untuk melatih jaringan, jaringan yang dilatih dengan demikian akan memiliki kemampuan dalam mengenali karakter yang dimasukkan. Akan ada beberapa zona yang diagonal nya kosong dari foreground pixels. Nilai ciri untuk zona tersebut adalah nol. 100 ciri telah di ekstraksi untuk masing-masing karakter. Selain itu, 100 ciri didapatkan dari merata-ratakan nilai yang ditempatkan pada tiap zona baris dan kolom[22]. *Diagonal Based Feature Extraction* dilakukan dengan cara membagi sebuah citra menjadi beberapa zona yang sama besar. Pada setiap zona, dilakukan perhitungan ciri titik pada setiap diagonal nya[8]. Contoh pada sebuah citra yang berukuran 20x20 piksel, dibagi menjadi 16 area sama besar yang dapat dilihat pada Gambar 2. 6.



Gambar 2. 6 Contoh pembagian area pada citra

Untuk setiap area yang berjumlah 16 memiliki ukuran 5x5 piksel, untuk setiap area didapat sejumlah 9 diagonal yang dapat dilihat pada Gambar 2. 7 berikut:



Gambar 2. 7 Contoh perhitungan diagonal pada area yang berukuran 5x5 piksel

Untuk mendapatkan diagonal pada area yang berukuran $n \times n$ dapat dihitung dengan persamaan berikut:

$$\text{Subfeature (Area)} = \frac{1}{2n-1} \sum_{i=1}^{2n-1} dg_i \quad (2. 10)$$

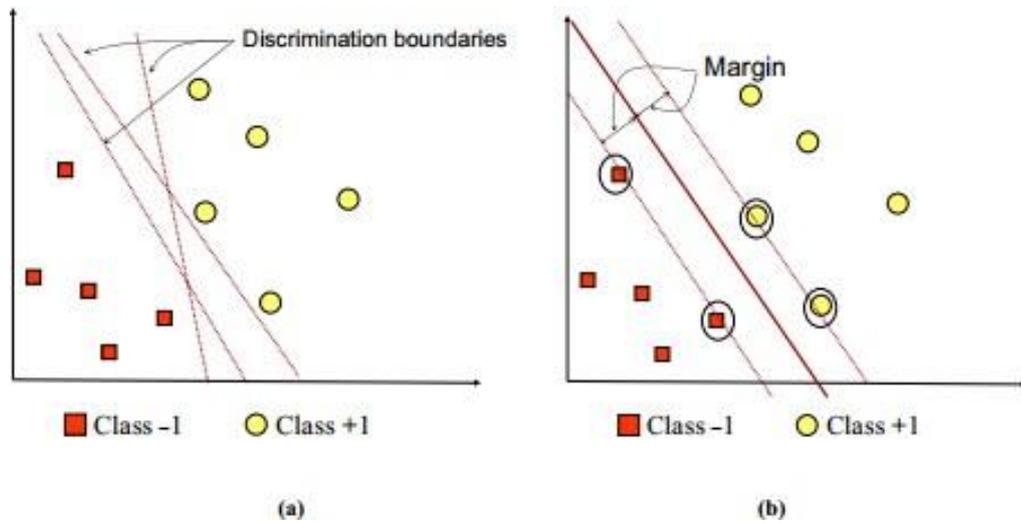
Keterangan:

n = Jumlah Piksel pada citra

dg_i = Jumlah garis diagonal ke i

2.9. Support Vector Machine

Definisi *Support Vector Machine* menurut Nugroho[26] adalah suatu Metode pembelajaran mesin yang menggunakan ruang hipotesis berupa fungsi linier pada ruang Ciri berdimensi tinggi, dan melatihnya menggunakan algoritme pembelajaran berdasarkan teori Optimasi dengan mengimplementasikan pembelajaran bias yang diturunkan dari teori pembelajaran statis. *Support Vector Machine (SVM)* diperkenalkan oleh Vapnik pada tahun 1992 sebagai suatu teknik klasifikasi yang efisien untuk masalah nonlinier. SVM berbeda dengan teknik klasifikasi di era 1980-an, seperti decision tree dan ANN, yang secara konsep kurang begitu jelas dan seringkali terjebak pada optimum lokal. SVM memiliki konsep yang jauh lebih matang, lebih jelas secara matematis, dibanding teknik – teknik klasifikasi sebelum era 1990-an. SVM berusaha menemukan hyperplane dengan memaksimalkan jarak antar kelas. Dengan cara ini, SVM dapat menjamin kemampuan generalisasi yang tinggi untuk data – data yang akan datang[26]. Konsep dari SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbalik yang berfungsi sebagai pemisah dua buah *class* pada *input space*. Dapat dilihat pada Gambar 2. 8-a dimana beberapa *pattern* memperlihatkan anggota dari 2 buah *class*: +1 dan -1. *Pattern* yang tergabung pada *class -1* disimbolkan dengan warna merah (kotak). Sedangkan *pattern* pada *class +1*, disimbolkan dengan warna kuning (lingkaran).



Gambar 2. 8 Konsep SVM dalam mencari hyperplane

Masalah klasifikasi dapat diterjemahkan dengan usaha menemukan garis (*hyperplane*) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) ditunjukkan pada Gambar 2. 8-a. *Hyperplane* pemisah terbaik antara kedua *class* dapat ditemukan dengan mengukur *margin hyperplane* tsb. dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* tersebut dengan pattern terdekat dari masing-masing *class*. *Pattern* yang paling dekat ini disebut sebagai *support vector*.

Garis solid pada Gambar 2. 8 -b menunjukkan *hyperplane* yang terbaik yaitu yang terletak tepat pada tengah-tengah kedua *class*, 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 SVM. Data yang tersedia dinotasikan sebagai $\vec{x}_i \in R^d$ sedangkan label masing-masing dinotasikan $y_i \in \{-1+1\}$ untuk $i = 1, 2, \dots, n$, yang mana n adalah banyaknya data. Di asumsikan kedua *class* -1 dan $+1$ dapat terpisah secara sempurna oleh *hyperplane* berdimensi d , yang didefinisikan sebagai berikut:

$$\vec{w} * \mathbf{x} + \mathbf{b} = \mathbf{0} \quad (2.11)$$

Pattern \vec{x}_i yang termasuk *class* -1 (sampel negatif) dapat dirumuskan sebagai:

$$\vec{w} * \mathbf{x} + \mathbf{b} \leq -1 \quad (2.12)$$

Sedangkan *pattern* \vec{x}_i yang termasuk *class* +1 (sampel positif) adalah sebagai berikut:

$$\vec{w} * \mathbf{x} + \mathbf{b} \geq +1 \quad (2.13)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu $1 / \|\vec{w}\|$. Hal ini dapat dirumuskan sebagai *Quadratic Programming (QP) problem*, yaitu mencari titik minimal persamaan (2.11), dengan memperhatikan *constraint* persamaan (2.12)

$$\min \tau(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.14)$$

$$\mathbf{y}_i(\mathbf{w} * \mathbf{x}_i + \mathbf{b}) - 1 \geq 0, (i = 1, \dots, n) \quad (2.15)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, di antaranya *Lagrange Multiplier*

$$\mathbf{L}(\mathbf{w}, \mathbf{b}, \mathbf{a}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i \mathbf{y}_i(\mathbf{w}^T + \mathbf{b}) + \sum_{i=1}^n \alpha_i \quad (2.16)$$

α_i adalah *Langrange multipliers*, yang bernilai nol atau positif ($\alpha_i \geq 0$). Nilai optimal dari persamaan (2.12) dapat dihitung dengan meminimalkan L terhadap \vec{w} dan b , dan memaksimalkan L terhadap α_i . Dengan memperhatikan sifat bahwa pada titik optimal *gradient* $L = 0$, persamaan (2.12) dapat dimodifikasi sebagai maksimalisasi *problem* yang hanya mengandung α_i , sebagaimana persamaan (2.13) dibawah.

Maximize:

$$\text{Max} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^n \alpha_i \alpha_j \mathbf{y}_i \mathbf{y}_j^T \mathbf{x}_i \mathbf{x}_j^T \quad (2.17)$$

Dimana $\alpha_i \geq 0$ ($i=1, 2, \dots, l$) dan $\sum \alpha_i \mathbf{y}_i = 0$ $l=1$. Dari hasil perhitungan ini diperoleh α_i yang kebanyakan bernilai positif. Data yang berkorelasi dengan α_i yang positif inilah yang disebut *support vector*.

Untuk mendapatkan nilai α_i , langkah pertama adalah mengubah setiap *feature* menjadi nilai vektor (*support vector*) = xy . Kemudian vektor akan dimasukkan kedalam persamaan *kernel trick phi phi* yaitu sebagai berikut

$$\varphi \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{cases} \sqrt{\mathbf{Xn}^2 + \mathbf{Yn}^2} > 2, \left[\begin{array}{l} 4 - \mathbf{x} + |\mathbf{x} - \mathbf{y}| \\ 4 - \mathbf{x} + |\mathbf{x} - \mathbf{y}| \end{array} \right] \\ \sqrt{\mathbf{Xn}^2 + \mathbf{Yn}^2} \leq 2, \text{ maka } \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \end{cases} \quad (2.18)$$

Kemudian untuk mencari nilai α_i didapatkan dari persamaan sebagai berikut.

$$\sum_{i=1, j=1}^n \mathbf{a}_i \mathbf{S}_i^T \mathbf{S}_j \quad (2.19)$$

$$\sum_{i=1, j=1}^n \mathbf{a}_i \mathbf{S}_i^T \mathbf{S}_j = \mathbf{y}_i \quad (2.20)$$

Setelah parameter \mathbf{a}_i didapatkan kemudian dimasukkan ke persamaan 2.17 berikut.

$$\sum_{i=1}^n \mathbf{a}_i \mathbf{S}_i \quad (2.21)$$

Terakhir akan dicari nilai w dan b untuk menemukan *hyperplane* sebagai patokan proses klasifikasi dengan persamaan sebagai berikut.

$$\mathbf{y} = \mathbf{w}\mathbf{x} + \mathbf{b} \quad (2.22)$$

Nilai s_i dari persamaan 2.17 merupakan nilai *support vector* yang telah dihitung sebelumnya. Dengan demikian proses klasifikasi selesai dengan memperhatikan *hyperplane* nya.

2.10. Smooth Support Vector Machine

SSVM merupakan pengembangan SVM dengan menggunakan teknik smoothing. Metode ini pertama kali diperkenalkan oleh Lee pada tahun 2001[27]. SVM memanfaatkan Optimasi dengan quadratic programming, sehingga untuk data berdimensi tinggi dan data jumlah besar SVM menjadi kurang efisien. Oleh karena itu dikembangkan smoothing technique yang menggantikan plus function SVM dengan integral dari fungsi sigmoid neural network yang selanjutnya dikenal dengan SSVM.

Berdasarkan sifat-sifat ini, solusi unik dari SSVM akan konvergen terhadap solusi unik problem Optimasi SVM bila smooth parameter \square pada SSVM mendekati tak

hingga. Notasi-notasi yang digunakan dalam algoritma SSVM adalah sebagai berikut:

$x^T y$ = Hasil skalar (inner product) dari dua vektor kolom x dan y pada dimensi real n .

x_+ = plus function, didefinisikan sebagai $(x_+)_i = \max\{0, x_i\}, i = 1, 2, \dots, n$

A = matriks bilangan real berukuran $m \times n$

e = vektor satu pada ruang dimensi real tertentu

I = matriks identitas di ruang dimensi real tertentu

$$\text{sign}(x) : \text{sign}(x) = \begin{cases} -1, & \text{jika } x < 0 \\ 0, & \text{jika } x = 0 \\ 1, & \text{jika } x > 0 \end{cases}$$

$\nabla f(x)$ = gradien dari f pada x dan merupakan sebuah vektor baris pada R^n

$\nabla^2 f(x)$ = turunan parsial kedua dari f terhadap x

$\Psi_{n \times n}$ = Matriks Hessian berukuran $n \times n$ yang merupakan hasil $\nabla^2 f(x)$

Diberikan masalah klasifikasi dari n objek dalam ruang dimensi R^p sehingga susunan data berupa matriks A berukuran $n \times p$ dan keanggotaan tiap titik terhadap kelas $\{+1\}$ atau $\{-1\}$ yang didefinisikan pada diagonal matriks D berukuran, problem Optimasi nya adalah

$$\min_{w, b, \xi} \frac{c}{2} \xi' \xi + \frac{1}{2} (w' w + b^2) \quad (2.23)$$

dengan kendala,

$$D(Aw + eb) + \xi \geq 0 \quad (2.24)$$

Solusi problem adalah

$$\xi = (e - D(Aw + eb)) \quad (2.25)$$

Dimana ξ adalah variabel *slack* yang mengukur kesalahan klasifikasi.

Kemudian dilakukan substitusi dan konversi, sehingga persamaan dapat ditulis sebagai berikut:

$$\min_{w, b} \frac{c}{2} \| (e - D(Aw - eb)) \|_2^2 + \frac{1}{2} (w' w + b^2) \quad (2.26)$$

Fungsi objektif dalam persamaan tidak memiliki turunan kedua. Teknik smoothing yang diusulkan dilakukan dengan mengganti fungsi plus dengan $p(x, a)$ yaitu integral dari fungsi sigmoid *neural network* atau dapat dituliskan sebagai berikut:

$$p(x, a) = x + \frac{1}{a} \log(1 + e^{-ax}), a > 0 \quad (2.27)$$

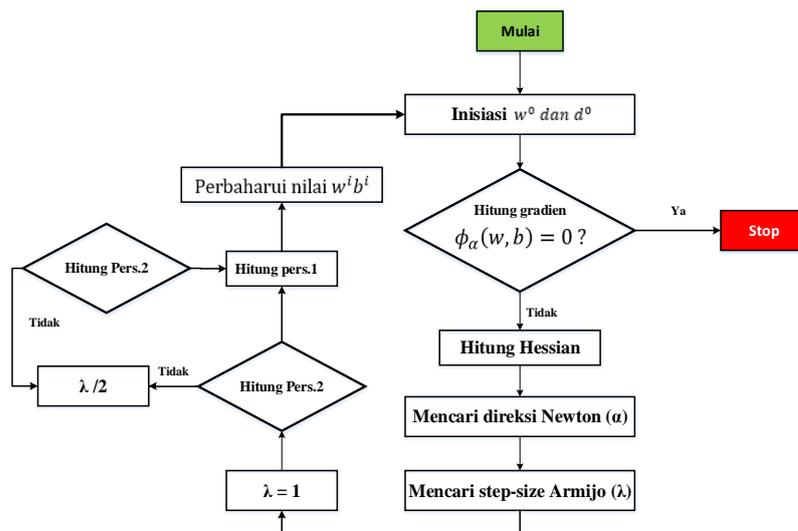
dimana α adalah parameter *smoothing*. Dengan menggantikan fungsi plus dengan $p(x, a)$ maka diperoleh model SSVM sebagai berikut:

$$\min_{(w,b) \in \mathbb{R}^{p+1}} \phi_{\alpha}(w, b) = \min_{(w,b) \in \mathbb{R}^{p+1}} \frac{c}{2} \|p(e - D(Aw - eb))\|_2^2 + \frac{1}{2}(w'w + b^2) \quad (2.8)$$

Secara umum, problem Optimasi SSVM dapat ditulis sebagai berikut:

$$\min_{(w,b) \in \mathbb{R}^{p+1}} \phi_{\alpha}(w, b) = \min_{(w,b) \in \mathbb{R}^{p+1}} \frac{c}{2} \|p(e - D(K(x_i, x_j)Dw - eb))\|_2^2 + \frac{1}{2}(w'w + b^2) \quad (2.9)$$

Yang diselesaikan dengan iterasi Newton Armijo yang dapat dilihat pada Gambar 2.9 dan $K(X_i, X_j)$ merupakan fungsi kernel yang dalam penelitian ini digunakan kernel Gaussian atau bisa dirumuskan berikut $K(X_i, X_j) = \exp(-\gamma(\|X_i, X_j\|^2))$ dengan parameter kernel γ



Gambar 2.9 Diagram analisis algoritma Newton-Armijo

Algoritma Newton-Armijo dimulai dengan menetapkan *starting value* $(\mathbf{w}^0, \mathbf{y}^0) \in R^{n+1}$. Proses akan berhenti apabila gradien dari fungsi objektif SSVM, yaitu $\nabla \Psi_\alpha(\mathbf{w}^i, \mathbf{y}^i) = \mathbf{0}$. Selain itu, $(\mathbf{w}^{i+1}, \mathbf{y}^{i+1})$ akan dihitung dengan langkah-langkah sebagai berikut.

1. Newton Direction (\mathbf{d}^i) = dengan menggunakan persamaan berikut:

$$\nabla^2 \Psi_\alpha(\mathbf{w}^i, \mathbf{y}^i) \mathbf{d}^i = -\nabla \Psi_\alpha(\mathbf{w}^i, \mathbf{y}^i) \quad (2.10)$$

2. Armijo *stepsize* = memilih sebuah *stepsize* $\lambda_i \in R$ sedemikian hingga

$$(\mathbf{w}^{i+1}, \mathbf{y}^{i+1}) = (\mathbf{w}^i, \mathbf{y}^i) + \lambda_i \mathbf{d}^i \quad (2.11)$$

Dimana $\lambda_i = \max\{1, 1/2, 1/4\}$ sehingga

$$\Psi_\alpha(\mathbf{w}^i, \mathbf{y}^i) - \Psi_\alpha((\mathbf{w}^i, \mathbf{y}^i) + \lambda_i \mathbf{d}^i) \geq -\delta \lambda_i \nabla \Psi_\alpha(\mathbf{w}^i, \mathbf{y}^i) \mathbf{d}^i \quad (2.12)$$

Dengan $\delta \in (0, \frac{1}{2})$

Saat $\nabla \Psi_\alpha(\mathbf{w}^i, \mathbf{y}^i) = \mathbf{0}$, iterasi pada algoritma Newton-Armijo berhenti dan diperoleh nilai \mathbf{w} dan \mathbf{y} yang konvergen. Pada proses Optimasi ini, matriks Hessian diberikan oleh Lee & Mangasarian.

$$\lim_{\alpha \rightarrow \infty} \nabla^2 \Psi_\alpha(\mathbf{w}, \mathbf{y}) = \begin{bmatrix} \frac{\partial^2 \Psi_\alpha(\mathbf{w}, \mathbf{y})}{\partial^2 \mathbf{w}^2} & \frac{\partial^2 \Psi_\alpha(\mathbf{w}, \mathbf{y})}{\partial \mathbf{w} \partial \mathbf{y}} \\ \frac{\partial^2 \Psi_\alpha(\mathbf{w}, \mathbf{y})}{\partial^2 \partial \mathbf{w}} & \frac{\partial^2 \Psi_\alpha(\mathbf{w}, \mathbf{y})}{\partial^2 \mathbf{y}^2} \end{bmatrix} \quad (2.13)$$

$$= \mathbf{I} + \mathbf{v} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$$

Dimana

$$H_{11} = A^T \text{diag}(S_\infty(e - D(A\mathbf{w} - e\gamma)))A$$

$$H_{12} = -A^T \text{diag}(S_\infty(e - D(A\mathbf{w} - e\gamma)))e$$

$$H_{21} = -e^T \text{diag} (S_\infty(e - D(Aw - e\gamma)))A$$

$$H_{22} = e^T \text{diag} (S_\infty(e - D(Aw - e\gamma)))e$$

Dan

$$S_\infty(x) = \frac{1 + \text{sign}(x)}{2}$$

Sementara itu, gradien dari matriks Hessian memiliki formulasi sebagai berikut

$$\lim_{\alpha \rightarrow \infty} \nabla \Psi_\alpha(\mathbf{w}, \gamma) = \begin{bmatrix} \mathbf{w} - \mathbf{v}A^T \mathbf{D}(e - \mathbf{D}(A\mathbf{w} - e\gamma)) \\ \gamma - \mathbf{v}e^T \mathbf{D}(e - \mathbf{D}(A\mathbf{w} - e\gamma)) \end{bmatrix} \quad (2.14)$$

Nilai \mathbf{w} dan γ optimum kemudian membentuk classifier dari SSVM linier yang memiliki formulasi berikut.

$$\mathbf{f}(\mathbf{x}) = \text{sign}(\mathbf{g}(\mathbf{x})) \quad (2.15)$$

Dengan

$$\mathbf{g}(\mathbf{x}) = \mathbf{x}^T \mathbf{w} - \gamma \quad (2.16)$$

2.11. Pengujian Akurasi

Pengujian akurasi dilakukan untuk mengetahui seberapa besar akurasi atau nilai kebenarannya yang didapatkan terhadap metode yang digunakan, adapun rumus dari pengujian akurasi yang digunakan adalah sebagai berikut:

$$\text{Akurasi} = \frac{\text{Jumlah Kepribadian Yang Tepat}}{\text{Jumlah Data Yang Diuji}} * 100\% \quad (2.35)$$

2.12. Diagram Konteks

Menurut Muslihudin pada tahun 2016 mendefinisikan *Diagram Context* atau Diagram konteks merupakan diagram yang menggambarkan cara proses pencatatan data. Grafik konteks terdiri dari proses konversi, sumber data, dan target data yang menerima atau mengirim data langsung dari proses konversi. Tujuan utama pembuatan diagram konteks adalah menjadi jembatan komunikasi dengan stakeholders, bukan untuk menjelaskan arus data kepada developer [29].

2.13. Data Flow Diagram (DFD)

Diagram arus data atau Data Flow Diagram (DFD) adalah model sistem yang digunakan untuk menggambarkan pembagian sistem menjadi modul yang lebih kecil. Menggunakan DFD memiliki banyak keuntungan, salah satunya adalah memudahkan pengguna atau pengguna yang belum mengenal komputer untuk memahami arti dari sistem yang digunakan[29]. Karena DFD memberikan pandangan secara menyeluruh mengenai sistem yang ditangani, menunjukkan fungsi-fungsi utama atau proses yang ada, aliran data, dan external entity

2.14. Pengujian Black Box

Pengujian *black box* dilakukan untuk menemukan kesalahan serta kekurangan yang tidak dapat ditemukan sebelumnya, pengujian ini dikhususkan untuk menguji fungsionalitas serta kinerja sistem yang telah dibangun.