

BAB II STUDI PUSTAKA

2.1 Tinjauan Pustaka

Dunia computer vision, penelitian tentang ekspresi wajah telah dilakukan oleh *Chinese Academy of Sciences Micro-Expression (CASME)*. Dalam penelitian tersebut membuat basis data yang terdapat beberapa gambar ekspresi wajah. Bertujuan untuk membantu penelitian di bidang computer vision tentang ekspresi mikro. Penelitian computer vision tentang pengenalan emosi berdasarkan ekspresi mikro baru mencapai akurasi 59,51%.(Huang, Wang, Zhao, & Pietikäinen, 2015).

Proses pengolahan citra adalah menjadikan citra dengan kualitas yang lebih baik seperti citra digital. Pengertian citra itu sendiri adalah tiruan dari benda atau objek. Citra digital dapat dihasilkan oleh kamera digital berupa foto dan disusun dengan banyak piksel. Nilai piksel tersebut mempunyai nilai yang menunjukkan intensitasnya. Awal dari proses pengolahan citra disusun dari akuisisi data citra, pemrosesan hingga pengujian. (Andi Sri Irtawaty, 2014).

Beberapa penelitian menerima transformasi *wavelet* untuk penelitian tentang kompresi gambar. Di sini artefak dihindari pada rasio kompresi tinggi, tetapi pada gambar dengan isi yang berbeda, dekomposisi *wavelet* menghasilkan rasio kompresi yang buruk. Teknik kompresi baru yang disebut CALIC dikembangkan untuk menentukan konteks formasi, kuantisasi, dan pemodelan. Itu tidak mendapatkan popularitas karena kinerja yang buruk dan operasi yang kompleks pada mode biner dan terus menerus.

Metode yang disebutkan di atas tidak efisien karena mereka menggunakan banyak jumlah koefisien untuk merekonstruksi tepi sepanjang kurva dan untuk mengkarakterisasi diskontinuitas tepi sepanjang kurva. Jadi untuk mengatasi penarikan ini transformasi multi-resolusi baru yang disebut *Curvelet Transform* diperkenalkan (Anandan, P. and Sabeenian, R.S. 2014) Transformasi ini superior melalui transformasi *wavelet* dalam kasus-kasus yang diikuti:

- 1) Secara optimal jarang representasi objek dengan tepi.

- 2) Ini adalah rekonstruksi citra yang optimal dalam masalah yang berakibat buruk
- 3) Merupakan representasi sparse propaganda gelombang yang optimal.

Ada beberapa penelitian yang sudah dilakukan dalam hal analisis sentimen terhadap data yang tersedia, diantaranya (Hidayatullah dan Azhari, 2014) menggunakan algoritma *Naïve Bayes Classifier* dan *Support Vector Machine* dalam menganalisis sentimen pada data *tweet* pemilihan presiden tahun 2014. Dari percobaan yang dilakukan, disimpulkan bahwa algoritma *Support Vector Machine* memiliki akurasi yang lebih baik dari algoritma *Naïve Bayes Classifier*. Penelitian lainnya yaitu (Songbo Tan et al, 2008) membandingkan *Naive Bayes*, *centroid classifier*, *K-Nearest Neighbor (KNN)*, *winnow classifier*, dan *Support Vector Machine (SVM)* pada *dataset* dokumen berbahasa cina. Didapatkan bahwa SVM memiliki hasil yang terbaik. Dari penelitian diatas, *Support Vector Machine* memiliki tingkat akurasi yang lebih baik dibandingkan algoritma pembanding lainnya. Sehingga penelitian ini menggunakan *Support Vector Machine* untuk melakukan klasifikasi pada *review* film.

Pada penelitian pengenalan wajah yang dilakukan oleh (Asri, S. A., 2011) sistem membedakan mana senyum dan yang tidak senyum, dimana kelas yang digunakan sebagai perbandingan hanya ada dua kelas. Beberapa metode klasifikasi dapat digunakan, pertama Metode *K-Nearest Neighbor* yang karakteristiknya mengenali kesamaan dari dua buah objek, kedua Cascade biasanya metode klasifikasi Cascade berdampingan dengan ekstraksi fitur Haar dimana Cascade berciri khas untuk menemukan pola dari objek pada citra untuk pengenalan beberapa objek. Sementara itu SVM memiliki karakteristik membandingkan kedua buah kelas yaitu kelas senyum dan kelas tidak senyum serta mengenali sebuah pola, oleh karena itu pada penelitian ini akan digunakan metode SVM. Hal ini didukung dengan kemampuan SVM yang baik dalam mengenali objek, antara lain pada penelitian Penelitian (Cakara, B., & Dkk., 2017) ini menggunakan metode *Histogram of Oriented Gradient* untuk deteksi dan *Support Vector Machine (SVM)* untuk klasifikasi. Penulis menggunakan metode SVM dikarenakan mempunyai tingkat akurasi dan ketepatan literasi yang cukup tinggi untuk mengklasifikasikan data, sehingga *Support Vector Machine (SVM)* bisa dikatakan lebih baik

dibandingkan metode yang lain penelitian ini mendapatkan hasil akurasi sebesar 82,20% (RBF Pasien 1), 81,63% (RBF Pasien 2) dan 78,73% (Linear Pasien 1), 79,55% (Linear Pasien 2), kemudian nilai specificity sebesar 89,91% (RBF Pasien 1), 92,18% (RBF Pasien 2) dan 88,06% (Linear Pasien 1), 91,34% (Linear Pasien 2), dan nilai sensitivity 15,45% (RBF Pasien 1), 12,97% (RBF Pasien 2) dan 13,33% (Linear Pasien 1), 12,50% (Linear Pasien 2).

Penelitian Jaeger, (2014), ini menggunakan dua set ciri yang digunakan sebagai deskriptor dalam proses klasifikasi. Set ciri pertama adalah ciri: bentuk, tepi, dan tekstur; sedangkan set kedua adalah ciri: intensitas, tepi, tekstur, dan momen bentuk (Jaeger, 2014). Pada proses klasifikasi, peneliti ini menggunakan SVM (*Support Vector Machine*), yang mengklasifikasikan ciri hasil kalkulasi komputer sebagai ciri yang menandakan citra normal atau abnormal. Uji coba dengan dua dataset uji menghasilkan akurasi 78,3% untuk data Montgomery County X-ray set dan 84% untuk data Shenzhen Hospital X-ray set.

Citra digital bisa diolah oleh komputer, maka citra digital harus mempunyai format tertentu. Format citra digital yang dipakai adalah citra skala keabuan format citra ini disebut skala keabuan karena pada umumnya warna yang dipakai warna hitam sebagai warna minimal (0) dan warna putih (255) sebagai warna maksimalnya, sehingga warna antaranya adalah abu-abu (Indraani, Jumaddina, & Sinaga, 2014).

Aplikasi pengolahan citra negatif menjadi citra true color dengan memanfaatkan perangkat lunak MATLAB. MATLAB merupakan bahasa tingkat tinggi dan lingkungan interaktif yang memungkinkan untuk melakukan tugas-tugas komputasi secara intensif lebih cepat dibandingkan dengan bahasa pemrograman tradisional seperti C, C++, dan Fortran. Matlab adalah sebuah lingkungan komputasi numerikal dan bahasa pemrograman komputer generasi keempat. Matlab memungkinkan manipulasi matriks, pem-plot-an fungsi dan data, implementasi algoritma, pembuatan antarmuka pengguna, dan pengantarmukaan dengan program dalam bahasa lainnya. Meskipun hanya bernuansa numerik, sebuah toolbox yang menggunakan mesin simbolik MuPAD, memungkinkan akses terhadap kemampuan aljabar komputer. Sebuah paket tambahan, Simulink, menambahkan simulasi grafis multiranah, dan Desain berdasar-Model untuk sistem terlekat dan dinamik (Sukardi, 2015)

Tujuan utama dari kompresi gambar adalah untuk memudahkan redundansi spektral dan spasial untuk menyimpan atau untuk mengomunikasikan informasi. Tujuan lain adalah untuk menjaga kualitas gambar bahkan pada bit rate yang lebih rendah mewakili sebuah gambar dan merekonstruksinya tanpa menurunkan kualitas visualnya. Dalam kompresi *lossless* skema, gambar yang direkonstruksi akan mirip dengan gambar aslinya. Dalam skema kompresi *lossy*, direkonstruksi gambar mungkin tidak mirip dengan gambar aslinya. Namun, skema kompresi *lossy* memberikan kompresi yang lebih tinggi rasio dari skema kompresi *lossless*. Skema yang diusulkan adalah semacam kompresi *lossless*. Dalam skema yang diusulkan, koefisien diperoleh menggunakan generasi kedua dari *Curvelet Transform*. Dalam kuantisasi panggung, koefisien yang membawa informasi paling sedikit dibulatkan. Kompresi berbasis transformasi secara spasial mendistribusikan energi ke sejumlah sampel data sehingga tidak ada informasi yang hilang. Gambar terkompresi direkonstruksi ke dalam domain spasial oleh proses transformasi (Ezhilarasi, P. dan Nirmalkumar, P. 2014)

(Khoje, Bodhe, & Adsul, 2013) dalam penelitiannya "Cacat Kulit Otomatis Sistem Identifikasi untuk Grading Buah Berdasarkan Discrete *Trasnformasi Curvelet* " dalam penilaian kuliatas buah secara objektif dengan analisis tekstur berdasarkan *transformasi curvelet*. Metode *curvelet* ini dapat mendeteksi permukaan buah yang memiliki piksel rendah maupun tinggi untuk mengekstrak detail global dan lokal permukaan buah. Ekstraksi ciri menggunakan 4 nilai yaitu *energy, entropy, mean, and standard deviation* yang digunakan mengkarakterisasi tekstur permukaan buah. Fitur klasifikasi yang digunakan dalam penelitian permukaan buah adalah *Support Vector Machine (SVM)*. Hasil dengan klasifikasi SVM memperoleh nilai terbaik dengan akurasi 96 %. Dapat disimpulkan bahwa metode transformasi *curvelet* memberikan kerusakan kulit buah.

Pengertian *Support Vector Machine (SVM)* yaitu sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi – fungsi linier dalam sebuah fitur yang berdimensi tinggi dan dilatih dengan menggunakan algoritma pembelajaran yang didasarkan pada teori optimasi. SVM pertama kali diperkenalkan pada tahun 1992 oleh Vapnik sebagai rangkaian dari beberapa konsep – konsep unggulan dalam bidang *pattern recognition* (Elly Susilowati, 2015)

Data Mining adalah penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Terdapat tiga teknik Data Mining yaitu Asosiasi, *Clustering* dan Klasifikasi. Dalam penelitian ini penulis menggunakan teknik klasifikasi. “Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek” (Bustami, 2014). Salah satu teknik klasifikasi yaitu *Support Vector Machine* (SVM).

Penelitian yang menerapkan algoritma *Support Vector Machine* (SVM) adalah klasifikasi tingkat kerusakan jalan pada rel kereta api yang dilakukan oleh Munawaroh. Klasifikasi yang dihasilkan berupa rusak berat, rusak sedang dan rusak ringan. Langkah algoritma yang dilakukan adalah normalisasi, menghitung kernel polynomial d, perhitungan matriks hessian, *sequential training* SVM, menghitung $f(x)$ serta melakukan klasifikasi dengan strategi one-against-all. Hasil akurasi yang didapat pada penelitian ini sebesar 91,24% (Munawaroh, 2014).

Proses klasifikasi dalam penelitian ialah suatu proses yang bertujuan untuk menentukan suatu obyek kedalam suatu kelas atau kategori yang sudah ditentukan sebelumnya (Elly Susilowati, 2015). Proses dari pembangunan terhadap suatu model yang mengklasifikasi suatu objek sesuai dengan atribut – atributnya. Klasifikasi data ataupun dokumen juga dapat dimulai dari membangun aturan klasifikasi tertentu yang menggunakan data *training* yang sering disebut sebagai tahapan pembelajaran dan pengujian digunakan sebagai data *testing*. (Winarko, 2014).

Clustering adalah metode yang digunakan dalam data mining yang cara kerjanya mencari dan mengklompokkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya yang telah diperoleh. Ciri khas dari teknik data mining ini adalah mempunyai sifat tanpa arahan (unsupervised), yang dimaksud adalah teknik ini diterapkankan tanpa perlunya data *training* dan tanpa ada teacher serta tidak memerlukan target *output* (J. O. Ong, 2013).

Akurasi adalah nilai ukur untuk validasi pada suatu metode. Akurasi adalah perbedaan antara harapan hasil tes dan nilai referensi yang diterima. Akurasi biasanya dinyatakan sebagai presentase. Akurasi dapat ditentukan melalui metode

penambahan baku (standard addition method). Dalam metode penambahan baku selisih dari jumlah keseluruhan data dan jumlah salah dibandingkan dengan jumlah keseluruhan data (Riyanto, 2014).

2.2 Landasan Teori

2.2.1 Pengenalan Wajah

Teknologi pengenalan wajah semakin berkembang, hasil penelitian dari pengenalan wajah banyak diaplikasikan dalam sistem pengenalan biometrik, pencarian dan pengindeksan database citra dan video digital misal untuk sistem keamanan, konferensi video, dan interaksi manusia dengan komputer. Pendeteksian wajah (*face detection*) merupakan salah satu tahap awal yang sangat penting sebelum dilakukan proses pengenalan wajah (*face recognition*).

Mendeteksi wajah dengan input video, yang kemudian dibuat klasifikasi. Algoritma deteksi wajah yang biasa digunakan dan terbukti cepat dan akurat adalah deteksi wajah *Viola-Jones*. Ditambah Haar cascade classifier yang kemudian menemukan wajah frontal (tampak depan). Masing-masing dari wajah ini kemudian akan diberi gambar kotak pada wajah (Priambodo, 2014).

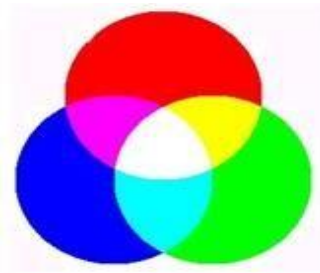
Pengenalan wajah (*face recognition*) merupakan salah satu teknologi biometric yang sekarang telah diterapkan untuk banyak aplikasi dalam bidang keamanan, antara lain Access security system, Authentication system, hingga sebagai alat bantu dalam pelacakan pelaku kriminal. Namun dalam perkembangannya masih terdapat beberapa permasalahan, selain masalah komputasi dan kapasitas penyimpanan data, kondisi citra wajah yang menjadi masukan (*input*) sistem juga merupakan masalah yang penting. Beberapa aspek penting yang

2.2.2 Citra Digital

Citra adalah suatu representasi (gambaran), kemiripan, atau imitasi dari suatu objek. Citra sebagai keluaran suatu sistem perekaman data dapat bersifat optik berupa foto, bersifat analog berupa sinyal-sinyal video seperti gambar pada *monitor* televisi, atau bersifat digital yang dapat langsung disimpan pada suatu media penyimpanan. (Kadir Abdul. 2013)

2.2.3 Citra RGB

Red (Merah), *Green* (Hijau) dan *Blue* (Biru) merupakan warna dasar yang dapat diterima oleh mata manusia. Setiap piksel pada citra warna mewakili warna yang merupakan kombinasi dari ketiga warna dasar RGB. Setiap titik pada citra warna membutuhkan data sebesar 3 byte. Setiap warna dasar memiliki intensitas tersendiri dengan nilai minimum nol (0) dan nilai maksimum 255 (8 bit). RGB didasarkan pada teori bahwa mata manusia peka terhadap panjang gelombang 630nm (merah), 530 nm (hijau), dan 450 nm (biru).



Gambar 2. 1 *Representasi Warna pada RGB* Pada

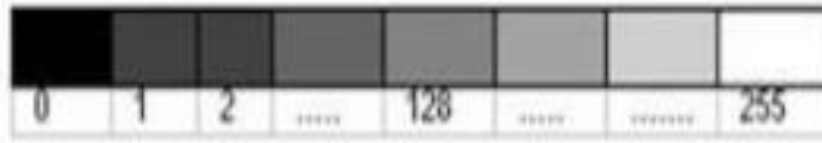
gambar 2.1 di atas dapat diambil beberapa kesimpulan yaitu:

1. RGB terdiri dari tiga warna utama, yaitu merah, hijau, dan biru.
2. Campuran dua warna pada RGB menghasilkan warna baru, yaitu kuning = merah + hijau, cyan = hijau + biru, dan magenta = biru + merah.
3. Bila seluruh warna merah, hijau, dan biru dicampur akan menghasilkan warna putih.
4. Bila warna merah, hijau, dan biru tidak dicampur maka akan menghasilkan warna hitam.

Jenis warna lain akan dihasilkan oleh variasi campuran warna dan intensitas campuran setiap warna

2.2.4 Citra *Grayscale*

Format citra ini disebut skala keabuan karena pada umumnya warna yang dipakai warna hitam sebagai warna minimal (0) dan warna putih (255) sebagai warna maksimalnya, sehingga warna antaranya adalah abu-abu, seperti ditunjukkan pada gambar :



Gambar 2. 2 Proses warna keabuan

Derajat keabuan sendiri sebenarnya memiliki beberapa nilai, tidak hanya skala 0 sampai 255. Hal ini tergantung pada nilai kedalaman piksel yang dimiliki oleh citra. Beberapa pembagian nilai derajat keabuan yang hubungannya dengan kedalaman pixel ditunjukkan pada tabel berikut ini:

Tabel 2. 1 Derajat keabuan

GrayScale	Scale (0, L)	Depth
$2^1(2)$	0,1	1 bit
$2^2(4)$	0 sampai 3	2 bit
$2^4(16)$	0 sampai 15	4 bit
$2^8(256)$	0 sampai 255	8 bit

Nilai 1,2,4,8 adalah bilangan bulat positif pada proses kuantisasi citra. Proses kuantisasi citra adalah salah satu bentuk dari proses digitalisasi citra yaitu proses untuk merepresentasikan citra dari fungsi malar (kontinyu) menjadi nilai-nilai diskrit. Proses kuantisasi membagi skala keabuan (*grayscale*) (0, L) menjadi sejumlah level, dinotasikan dengan G dan nilainya berupa bilangan bulat (*integer*), biasanya G merupakan hasil perpangkatan dari dua: $G = 2^m$ dengan $G =$ Derajat keabuan (*grayscale*) dan $m =$ Bilangan bulat positif.

2.2.5 Transformasi *Curvelet*

Transformasi *Curvelet* dapat menjadi ciri khas sinyal-sinyal berdimensi tinggi yang memiliki garis, kurva atau singularitas pesawat hiper. (Vijay Kumar et al, 2014). Transformasi *curvelet* telah dikembangkan untuk mengatasi keterbatasan wavelet meskipun transformasi wavelet telah dieksplorasi secara luas di berbagai cabang pemrosesan gambar, ia gagal untuk merepresentasikan objek yang berisi sisi dan kurva yang berorientasi acak karena tidak pandai mewakili singularitas garis. *Filter Gabor* ditemukan berfungsi lebih baik daripada transformasi wavelet dalam

merepresentasikan tekstur dan mengambil gambar karena pendekatan multiorientasi. Namun, karena hilangnya informasi spektral dalam filter mereka tidak dapat secara efektif mewakili gambar. Ini memengaruhi kinerja CBIR. Konsekuensinya, diperlukan mekanisme yang lebih kuat untuk meningkatkan kinerja CBIR. Untuk mencapai cakupan domain spektral yang lengkap dan untuk menangkap lebih banyak informasi orientasi, transformasi *curvelet* telah dikembangkan.

2.2.6 Ekstraksi Ciri

Setelah dilakukan normalisasi baseline, dilakukan ekstraksi ciri. Ekstraksi ciri pada *electronic nose* adalah proses untuk mengekstraksi informasi penting dan relevan dari sinyal sensor yang dapat mewakili ciri respon sensor secara keseluruhan (Rosyad, F., 2015).

Ekstraksi ciri yang digunakan adalah nilai rata-rata(*mean*), standar deviasi, *energy* dan *entropy* yang dihasilkan dari perhitungan matriks koefisien *curvelet* dari masing-masing citra sehingga nantinya ekstraksi ciri yang dipilih ini akan digunakan untuk menyelidiki dan memberikan penilaian dari kualitas buah manggis. Empat ekstraksi ciri seperti mean, standar deviasi, *energy* dan *entropy* menunjukkan kekuatan diskriminasi yang tinggi dalam pemeriksaan kualitas buah (Khoje, S. A., dkk. 2013).

Perhitungan nilai ekstraksi ciri citra menggunakan rumus sebagai berikut:

1. Rata-rata

Persamaan rata – rata :

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i \quad (2.1)$$

Nilai mean yaitu menghitung rata- rata dari jumlah nilai piksel mulai piksel ke-1 hingga ke-N. selanjutnya dibagi dengan jumlah yang ada.

2. Standar Deviasi

Persamaan standar deviasi:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (2.2)$$

Nilai standar deviasi yaitu menghitung dengan nilai yang terurai setiap piksel ke-1 hingga ke-N. Selanjutnya hasil jumlah akan dikudratkan guna mendapatkan nilai.

3. Entropy

Nilai *entropy* yaitu nilai ukuran yang secara tidak teratur dalam sebuah bentuk. Nilai H menunjukkan nilai besar jika digunakan citra transisi keabuan yang rata dan nilai kecil dengan struktur citra yang tidak teratur. Nilai *entropy* menghasilkan distribusi derajat keabuan dari citra secara acak. Semakin bervariasi distribusi derajat keabuannya semakin banyak juga nilai *entropy* yang di dapat. Proses menghitung *entropy* pada persamaan berikut:

$$H = - \sum_{i=1}^G p(d_i) \cdot \log_2 p(d_i) \quad (2.3)$$

4. Energy

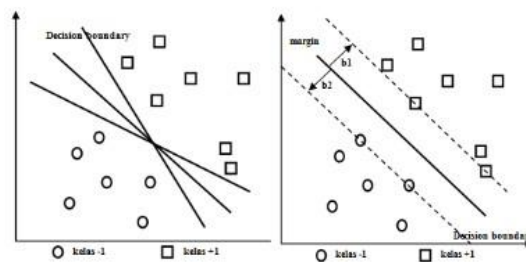
Nilai *energy* yaitu fitur yang menggunakan ukuran konsentrasi pasangan intensitas dalam sebuah matrik co-occurrence. Hasil nilai tersebut berupa nilai besar distribusi *level gray* citra yang memiliki bentuk secara konstan. Semakin tinggi nilai *entropy* maka nilai *energy* akan semakin rendah nilai *energy* karena nilai *energy* digambarkan dengan keteraturan penyebaran derajat keabuan citra. Jadi bisa disebut *energy* adalah invers dari nilai *entropy*. Proses menghitung *energy* persamaan berikut:

$$E_k = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N |x_k(i, j)| \quad (2.4)$$

2.2.7 Support Vector Machine (SVM)

SVM merupakan metode klasifikasi yang kini banyak dikembangkan dan diterapkan. Metode ini berasal dari teori pembelajaran statistik yang menjanjikan dan memberikan hasil yang lebih baik dibandingkan dengan metode yang lainnya. SVM bekerja sangat baik pada himpunan data berdimensi tinggi. SVM yang menggunakan teknik kernel harus memetakan data asli dari dimensi asalnya menjadi dimensi lain yang relatif lebih tinggi. Pada metode ANN, selama proses pelatihan semua data latih akan dipelajari. Berbeda dengan SVM yang hanya

sejumlah data terpilih sajalah yang digunakan dalam klasifikasi yang akan dipelajari. Berbeda juga dengan *Nearest Neighbor* yang menyimpan semua data latih yang akan digunakan pada saat prediksi, SVM hanya menyimpan sebagian kecil data latih untuk digunakan pada saat prediksi. Hal inilah yang menjadi kelebihan SVM karena tidak semua data latih dilibatkan pada saat pelatihan (Prasetyo, 2014). Ide dasar dari SVM adalah memaksimalkan batas hyperplane yang dijelaskan pada Gambar 2.3



Gambar 2.3 *Hyperplane SVM*

Nugroho (2007) dalam Eko Prasetyo (2014) menjelaskan konsep klasifikasi dengan SVM adalah sebagai usaha mencari hyperplaneterbaik yang berfungsi sebagai pemisah dua buah kelas data pada input space. Beberapa data yang merupakan anggota dari dua buah kelas data yakni +1 dan -1. Bentuk lingkaran merupakan simbol data yang tergabung dalam kelas -1, sedangkan bentuk bujur sangkar merupakan simbol data yang tergabung dalam kelas +1. Hyperplane pemisah terbaik antara kedua kelas tersebut dapat ditemukan dengan mengukur margin hyperplane dan mencari titik maksimalnya. Margin merupakan jarak antara hyperplane tersebut dengan data terdekat dari masing-masing kelas. Data yang paling dekat inilah yang disebut dengan *support vector*. Pada gambar b di atas menunjukkan hyperplane terbaik karena terletak di tengah-tengah kedua kelas. Data lingkaran dan bujur sangkar yang dilewati garis putus-putus (garis batas margin) adalah *support vector*. Inti dari proses pelatihan pada SVM ini adalah usaha untuk mencari lokasi *hyperplane* (Prasetyo, 2014)

2.2.8 K-Fold Cross Validation

Cross Validation adalah salah satu metode yang digunakan untuk memvalidasi keakuratan sebuah model yang dibangun berdasarkan dataset tertentu.

Pembuatan model biasanya bertujuan untuk melakukan klasifikasi terhadap suatu data baru yang belum pernah muncul di dalam dataset. Data yang digunakan dalam proses pembangunan data model disebut data *training*, sedangkan data yang akan digunakan untuk validasi model disebut dengan data *testing*. Salah satu metode *cross-validation* yang populer adalah *K-Fold Cross Validation*. Dalam teknik ini dataset dibagi menjadi sejumlah K-buah partisi secara acak. Kemudian dilakukan sejumlah K-kali eksperimen, dimana masing-masing eksperimen menggunakan data partisi ke-K sebagai data testing dan memanfaatkan sisa partisi lainnya sebagai data training (Muafiq, 2016).