SKRIPSI

ANALISIS SENTIMEN *REVIEW* RESTORAN PADA KOMENTAR YOUTUBE

Disusun Oleh:

DILA AMALIA

D421 15 307

DEPARTEMEN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
MAKASSAR

2021

LEMBAR PENGESAHAN SKRIPSI

ANALISIS SENTIMEN *REVIEW* RESTORAN

PADA KOMENTAR YOUTUBE

Disusun dan diajukan oleh

DILA AMALIA

D421 15 307

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin pada tanggal 23 Februari 2021 dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,	Pembimbing Pendamping
Dr. Amil Ahmad Ilham, S.T., M.IT	Iqra Aswad, S.T.,M.T
Nip. 19731010 199802 1 001	Nip. 19901128 201904 3 001

Ketua Program Studi,

Dr. Amil Ahmad Ilham, S.T., M.IT

Nip. 19731010 199802 1 001

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : DILA AMALIA

NIM : D421 15 307

Program Studi : TEKNIK INFORMATIKA

Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

ANALISIS SENTIMEN *REVIEW* RESTORAN PADA KOMENTAR YOUTUBE

Adalah karya tulis saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 23 Februari 2021

Yang Menyatakan

DILA AMALIA

ABSTRAK

Di era web seperti sekarang, sejumlah informasi kini mengalir melalui jaringan. Karena berbagai konten web meliputi opini subjektif serta informasi yang objektif, sekarang umum bagi orang-orang untuk mengumpulkan informasi tentang produk dan jasa yang mereka ingin beli. Namun karena cukup banyak informasi yang ada dalam bentuk teks tanpa ada skala numerik, sulit untuk mengklasifikasikan evaluasi informasi secara efisien tanpa membaca teks secara lengkap. Analisa sentimen bertujuan untuk mengatasi masalah ini dengan secara otomatis mengelompokkan *review*. Sentimen analisis merukapakan proses klasifikasi dokumen tekstual ke dalam beberapa kelas seperti sentiment positif, negatif, dan netral. Pada penelitian ini dibahas analisis sentiment *review* restoran dari komentar Youtube. Penelitian ini menggunakan data komentar dari Platform Youtube berbahasa Indonesia. Fitur yang digunakan adalah TF-IDF dan algoritma *Support Vector Machine (SVM)*. Hasil pengujian menunjukkan akurasi sentiment analisis sebesar 80%.

Kata Kunci: Analisis Sentimen, TF-IDF, *Support Vector Machine (SVM)*.

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya sehingga tugas akhir yang berjudul " ANALISIS SENTIMEN *REVIEW* RESTORAN PADA KOMENTAR YOUTUBE " ini dapat diselesaikan sebagai salah satu syarat dalam menyelesaikan jenjang Strata-1 pada Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin.

Penulis menyadari bahwa dalam penyusunan dan penulisan laporan tugas akhir ini tidak lepas dari bantuan, bimbingan serta dukungan dari berbagai pihak, dari masa perkuliahan sampai dengan masa penyusunan tugas akhir. Oleh karena itu, penulis dengan senang hati menyampaikan terima kasih kepada:

1. Kedua Orang tua dan Saudara penulis, Bapak Moh. Ervin Laha S.Sos, MM, Ibu Chadijah S.E., Ulfah Ervita S.KM, Wildan Maulana, dan Nur Ramadhani yang selalu memberikan dukungan, doa, dan semangat serta selalu sabar dalam mendidik penulis sejak kecil;
2. Bapak Dr. Amil Ahmad Ilham,S.T., M.IT. selaku pembimbing I dan Bapak Iqra Aswad, S.T., M.T., selaku pembimbing II yang selalu menyediakan waktu, tenaga, pikiran dan perhatian yang luar biasa untuk mengarahkan penulis dalam penyusunan tugas akhir;
3. Bapak Dr. Amil Ahmad Ilham, ST., M.IT., selaku Ketua Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin atas bimbingannya selama masa perkuliahan penulis;

4. Bapak dan Ibu dosen Departemen Teknik Elektro dan Informatika Universitas Hasanuddin atas bimbingan, nasehat dan wejangan terkait perkuliahan dan kehidupan;
5. Gibril, Diki Wahyudi, Kak Sofyan, Alfina Sulfiana telah memberikan begitu banyak bantuan selama penelitian, pengambilan data dan diskusi *progress* penyusunan tugas akhir;
6. Bayazid Sustami, Alfina Sulfiana, Nur Arifa Isnaeni, Fadhillah Armin, A. Ardiansyah, Nazila Riza dan seluruh sahabat lovely telah memberikan semangat untuk menyelesaikan tugas akhir ini;
7. Teman-teman Hypervisor FT UH atas dukungan dan semangat yang diberikan selama ini.
8. Segenap Staf Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah membantu penulis.
9. Orang-orang berpengaruh lainnya yang tanpa sadar telah menjadi inspirasi penulis.

Akhir kata, penulis berharap semoga Allah SWT. berkenan membalas segala kebaikan dari semua pihak yang telah banyak membantu. Semoga Tugas Akhir ini dapat memberikan manfaat bagi pengembangan ilmu. Aamiin.

Wassalam

Makassar, Februari 2021

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan internet yang begitu pesat saat ini telah membawa interaksi manusia yang intensif di dunia Internet ke era media sosial. Media sosial dapat didefinisikan sebagai kelompok dari aplikasi berbasis Internet yang berkumpul berdasarkan ideologi dan perkembangan teknologi Web 2.0 yang memperbolehkan adanya pembentukan dan pertukaran konten yang di buat oleh pengguna. Salah satu media sosial yang popular saat ini adalah Youtube. (Hu Tao, 2017)

YouTube adalah salah satu sumber informasi video yang memiliki pengguna aktif terbesar, dimana pengguna dapat berinteraksi dengan berbagi video. YouTube juga memfasilitasi pengguna untuk menanggapi video dengan cara memberikan komentar. Pengguna YouTube telah memanfaatkan YouTube untuk memberikan *review* terhadap suatu produk tertentu misalnya menu makanan di restoran. Sebagai contoh, link YouTube https://www.youtube.com/watch?v=0HTrKRJtd-k telah mendapatkan komentar sebanyak 17000.

Review dari customer menjadi informasi yang sangat berguna bagi pemilik restoran dan pelanggan. Bagi pelanggan, *review* dari customer dapat menjadi referensi pemilihan restoran yang akan dikunjungi. Bagi pemilik restoran, *review* dari customer dapat menunjukkan kualitas dari menu makanan yang disajikan dan hasil analisis *review* ini dapat digunakan untuk memperbaiki kualitas makanan

atau fasilitas untuk restoran sehingga dapat meningkatkan keuntungan restoran tersebut.

Masalah yang dihadapi adalah kesulitan untuk mendapatkan informasi dan *knowledge* secara manual dari *review* yang berjumlah besar dan beragam karena dibutuhkan waktu yang banyak untuk membaca dan memahami setiap *review*.

Untuk mengatasi masalah tersebut, pada tugas akhir ini diusulkan pengembangan sistem analisis sentimen yang dapat mengolah *review* dalam jumlah banyak dan menghasilkan informasi yang berguna.

1.2 Rumusan Masalah

1. Bagaimana mendapatkan informasi dan *knowledge* dengan lebih mudah dari *review* suatu produk yang berjumlah besar dan beragam.
2. Bagaimana menerapkan analisis sentimen yang optimal pada *review* yang berjumlah besar dan beragam

1.3 Tujuan Penelitian

1. Untuk mendapatkan informasi dan *knowledge* dari *review* suatu produk yang berjumlah besar dan beragam dengan lebih mudah.
2. Untuk menerapkan analisis sentimen yang optimal dengan jumlah *review* besar.

3.4 Manfaat Penelitian

1. Bagi pelanggan, *review* dari customer dapat menjadi referensi pemilihan restoran yang akan dikunjungi. Bagi pemilik restoran,

review dari *customer* dapat menunjukkan kualitas dari menu makanan yang disajikan**.**

2. Hasil analisis *review* ini dapat digunakan untuk memperbaiki kualitas makanan atau fasilitas untuk restoran sehingga dapat meningkatkan keuntungan restoran tersebut

1.5 Batasan Masalah

1. Komentar yang berbahasa Indonesia dan diambil dari media social YouTube.
2. Dataset yang digunakan adalah file yang berekstensi csv.
3. *Scrapping* data menggunakan selenium web *driver*.

1.6 Sistematika Penulisan

Untuk memberikan gambaran singkat mengenai isi tulisan secara keseluruhan, maka diuraikan beberapa tahapan dari penulisan secara sistematis, yaitu:

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini akan dijelaskan mengenai teori-teori yang menunjang percobaan yang dilakukan.

BAB III METODOLOGI PENELITIAN

Bab ini berisi membahas alur pengerjaan tugas akhir, meliputi kebutuhan sistem dan ta perancangan sistem yang akan diimplementasikan dalam sentiment analisis ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi pengujian dan analisis dari hasil implementasi terhadap penelitian yang dilakukan.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian dan saran yang diperlukan untuk pengembangan pada penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

2.1 YouTube

YouTube adalah sebuah situs web video sharing (berbagi video) populer dimana para pengguna dapat memuat, menonton, dan berbagi klip video secara gratis. Umumnya video-video di YouTube adalah klip musik (video klip), film, TV, serta video buatan para penggunanya sendiri. Format yang digunakan video-video di YouTube adalah yang dapat diputar di penjelajah web yang memiliki plugin Flash Player.

Nielsen dan Adobe melaporkan terjadi peningkatan aktivitas menonton secara online di dunia yang tak hanya melalui komputer, tetapi juga ponsel pintar, konsol permainan, serta televisi pintar. Berdasarkan data tersebut, ada 38,2 miliar orang yang menyaksikan video online secara gratis pada kuartal kedua 2014. Jumlah ini mengalami peningkatan sebesar 43 persen dari kuartal yang tahun lalu. Bahkan, angka pengguna yang kembali berkunjung (unique visitor) dalam satu bulan meningkat sebesar 146 persen.

2.2 Sentimen Analisis

Sentimen analisis atau opinion mining mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining* yang bertujuan menganlisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenaan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu.

Tugas dasar dalam analisis sentimen adalah mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan dalam kaliamat atau dokumen tersebut apakah bersifat positif, negatif atau netral. Sentimen analisis juga dapat menyatakan perasaan emosional positif, negatif, dan netral.

Ekspresi atau sentiment mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subject yang berbeda. Oleh karena itu pada beberapa penelitian, terutama pada *review* produk, pekerjaan didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses opinion mining.

2.3 *Web Scraping*

Web scraping adalah proses ekstrasi data dari sebuah website. Salah satu contoh web *scraping* adalah meng-copy daftar contact dari sebuah direktori web. Memang Anda bisa saja melakukan ini secara manual denagn meng-copy paste data ke excel, misalnya. Tetapi bagaimana kalau datanya banyak? Untuk ini, Anda membutuhkan automation yang bisa membantu proses *web scraping* Anda lebih cepat dan mudah.

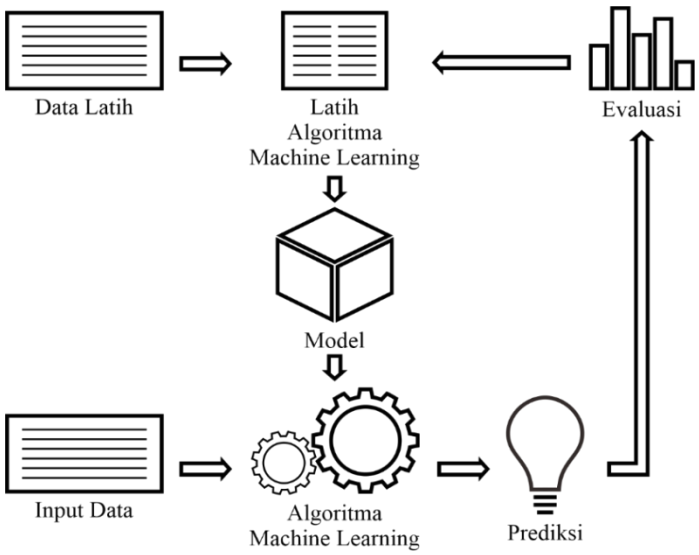
Web scraping dilakukan dengan menggunakan *web scraper, bot, web spider,* atau *web crawler*. *Web scraper* sendiri adalah program yang masuk ke halaman website, *download* kontennya, mengekstrak data dari konten, dan menyimpan data ke satu file atau database.

Selenium merupakan tools untuk *Automation Browser*, singkatnya browser akan dijalankan secara otomatis melalui program yang dirancang. Biasanya digunakan untuk melakukan testing dalam aplikasi sebuah website. Agar selenium dapat berjalan secara otomatis dibutuhkan *Web Driver* sesuai browser yang ingin digunakan. Setiap browser memiliki *driver* masing-masing. *Driver* inilah yang akan digunakan oleh *selenium* untuk menjalankan browser secara otomatis.

2.4 Machine Learning

Machine learning adalah disiplin ilmu dari *Artificial Intelligence* (Kecerdasan Buatan) yang menggunakan teknik statistika untuk menghasilkan suatu model otomatis dari sekumpulan data yang biasa disebut *dataset*, dengan tujuan memberikan komputer kemampuan untuk "belajar". Pembelajaran mesin atau machine learning memungkinkan komputer mempelajari sejumlah data (*learn from data*) sehingga dapat menghasilkan suatu model untuk melakukan proses input-output tanpa menggunakan kode program yang dibuat secara eksplisit.

Proses belajar tersebut menggunakan algoritma khusus yang disebut *machine learning algorithms*. Terdapat banyak algoritma *machine learning* dengan efesiensi dan spesifikasi kasus yang berbeda-beda. Tidak hanya individu yang belajar meningkatkan kecerdasan nya tetapi mesin juga membutuhkan hal tersebut untuk meningkatkan kecerdasannya dan memiliki kemampuan yang cerdas dan tidak dimiliki oleh mesin lainnya.



Gambar 2. 1 Konsep Machine Learning

Secara fundamental cara kerja *machine learning* adalah belajar seperti manusia dengan menggunakan contoh-contoh dan setelah itu barulah dapat menjawab suatu pertanyaan terkait. Proses belajar ini menggunakan data yang disebut *train dataset* (data latih). Berbeda dengan program statis, *machine learning* diciptakan untuk membentuk program yang dapat belajar sendiri. Gambaran konsep *machine learning* dapat dilihat pada Gambar 2.1.

Dari data tersebut, komputer akan melakukan proses belajar (*training*) untuk menghasilkan suatu model. Proses belajar ini menggunakan algoritma *machine learning* sebagai penerapan teknik statistika. Model inilah yang menghasilkan informasi, kemudian dapat dijadikan pengetahuan untuk memecahkan suatu permasalahan sebagai proses input-output. Model yang dihasilkan dapat melakukan klasifikasi ataupun prediksi kedepannya.

Untuk memastikan efesiensi model yang terbentuk, data akan dibagi menjadi data latih (*train dataset*) dan data uji (*test dataset*). Pembagian data yang digunakan bervariasi bergantung algoritma yang digunakan. Pada umumnya *train dataset* lebih banyak dari *test dataset*, misalnya dengan rasio 3:1. *Test dataset* digunakan untuk menghitung seberapa efesien model yang dihasilkan untuk melakukan klasifikasi atau prediksi kedepannya yang disebut *test score*. Semakin banyak data yang digunakan, *test score* yang dihasilkan semakin baik. Nilai *test score* bisa berada dalam rentang 0 sampai 1 atau -1 sampai 1.

2.5 Preprocessing

Text preprocessing merupakan tahap awal dari *text mining* dimana data text akan dibersihkan sehingga text menjadi lebih terstruktur sebelum masuk ketahap berikutnya untuk diolah lebih lanjut. Sekumpulan karakter yang bersambungan (teks) harus dipecah-pecah menjadi lebih berarti. Hal tersebut dapat dilakukan dalam beberapa tingkatan yang berbeda. Suatu dokumen dapat di pecah menjadi suatu bab, paragraf, kalimat dan kata. Tahapan *Text preprocessing* dalam penelitian ini meliputi: *Tokenizing, Case Folding, Filtering*, dan *Stopword.*

2.6 Ekstraksi Fitur menggunakan tf idf

Term Frequency (TF) dan *Inverse Document Frequency (IDF)* merupakan pembobotan yang sering digunakan dalam penelusuran informasi dan *text mining*. TF-IDF dapat digunakan untuk pembobotan kata sebagai fitur untuk klasifikasi sentimen kalima. TF merupakan pembobotan yang sederhana dimana penting tidaknya sebuah kata diasumsikan sebanding dengan jumlah kemunculan kata tersebut dalam dokumen, sementara IDF adalah pembobotan yang mengukur

seberapa penting sebuah kata dalam dokumen bila dilihat secara global pada seluruh dokumen. Nilai pembobotan TF x IDF akan tinggi jika nilai TF besar dan kata yang diamati tidak ditemukan di banyak dokumen. Nilai TF dihitung menggunakan fungsi berikut:

$$TF(d,t) = f(d, t) \quad \text{……..…} (1)$$

Dimana f(d,t) adalah jumlah kemunculan kata t pada dokumen d. IDF mempertimbangkan frekuensi kata pada seluruh dokumen yang ada. Pembobotan IDF menganggap bahwa bobot sebuah kata akan besar jika kata tersebut sering muncul dalam sebuah dokumen tetapi tidak banyak dokumen yang mengandung kata tersebut. Nilai IDF dihitung menggunakan fungsi berikut:

$$IDF(t) = \log(N/df(t)) \quad \text{……...(2)}$$

Dimana df(t) adalah jumlah dokumen yang memiliki kata t. Hasil kajian sebelumnya memperlihatkan bahwa pembobotan TF x IDF dapat meningkatkan performansi secara lebih baik. Nilai TF x IDF dihitung menggunakan fungsi berikut ini. (Taufiq M.Isa. 2013)

$$TFIDF(d,t) = TF(d,t) \times IDF(t) \quad \text{……...…(3)}$$

Pada term frequency (tf), terdapat beberapa jenis formula yang digunakan, yaitu: (Akbar, Martha, 2012)

a. Tf biner (binery tf), yang hanya memperhatikan apakah suatu kata ada atau tidak ada dalam dokumen, jika ada diberi nilai satu, jika tidak diberi nilai nol.

b. Tf murni (raw tf), nilai tf diberikan berdasarkan jumlah kemunculan suatu kata di dokumen. Contohnya jika muncul lima kali maka kata tersebut akan bernilai lima.
c. Tf logaritmik, hal ini untuk menghindari dominansi dokumen yang mengandung sedikit kata dalam query, namun mempunyai frekuensi yang tinggi.
d. Tf normalisasi, menggunakan perbandingan antara frekuensi sebuah kata dengan jumlah keseluruhan kata pada dokumen

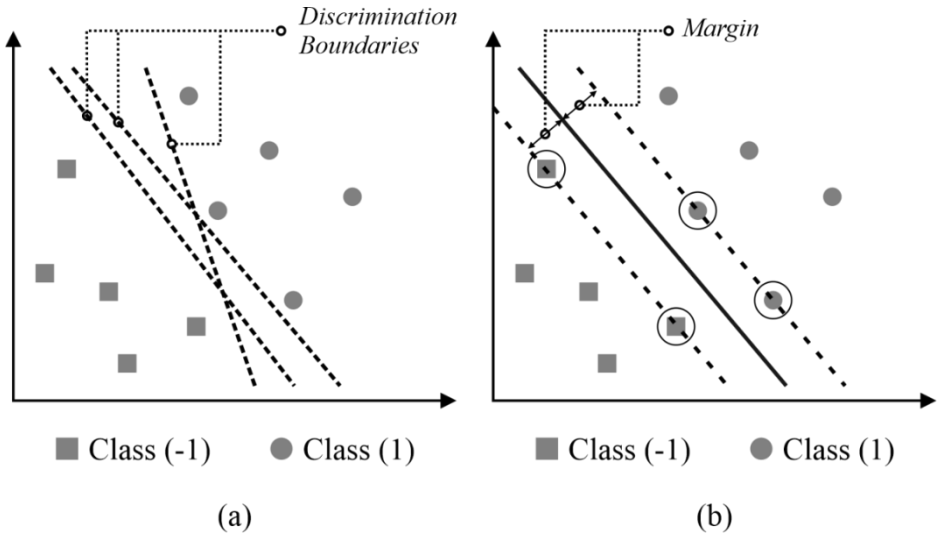
2.7 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu metode yang akhir-akhir ini banyak mendapat perhatian. *Support Vector Machine* (SVM) dikembangkan oleh Boser, Guyon, Vapnik, dan pertama kali dipresentasikan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. Konsep dasar SVM sebenarnya merupakan kombinasi harmonis dari teoriteori komputasi yang telah ada puluhan tahun sebelumnya, seperti *margin hyperplane* (Duda & Hart tahun 1973, Cover tahun 1965, Vapnik 1964, dsb.), kernel diperkenalkan oleh Aronszajn tahun 1950, dan demikian juga dengan konsep-konsep pendukung yang lain. Akan tetapi hingga tahun 1992, belum pernah ada upaya merangkaikan komponen-komponen tersebut. Prinsip dasar SVM adalah *linear classifier*, dan selanjutnya dikembangkan agar dapat bekerja pada problem *non-linear*. dengan memasukkan konsep *kernel trick* pada ruang kerja berdimensi tinggi.

Support Vector Machine (SVM) juga dikenal sebagai teknik pembelajaran mesin (*machine learning*) paling mutakhir setelah pembelajaran mesin

sebelumnya yang dikenal sebagai *Neural Network (*NN). baik SVM maupun NN tersebut telah berhasil digunakan dalam pengenalan pola. Pembelajaran dilakukan dengan menggunakan pasangan data input dan data output berupa sasaran yang diinginkan. Pembelajaran dengan cara ini disebut dengan pembelajaran terarah (*supervised learning*). Dengan pembelajaran terarah ini akan diperoleh fungsi yang menggambarkan bentuk ketergantungan input dan outputnya. Selanjutnya, diharapkan fungsi yang diperoleh mempunyai kemampuan generalisasi yang baik, dalam arti bahwa fungsi tersebut dapat digunakan untuk data input di luar data pembelajaran. diperoleh mempunyai kemampuan generalisasi yang baik, dalam arti bahwa fungsi tersebut dapat digunakan untuk data input di luar data pembelajaran.

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah *class* pada *input*



space. Gambar 2.2.

Gambar 2. 2 SVM Berusaha Menemukan Hyperplane Terbaik

Bagian (a) memperlihatkan beberapa *pattern* yang merupakan anggota dari dua buah *class*: positif (dinotasikan dengan 1) dan negatif (dinotasikan dengan -1). *Pattern* yang tergabung pada *class* negatif disimbolkan dengan kotak, sedangkan *pattern* pada *class* positif, disimbolkan dengan lingkaran. Proses pembelajaran dalam problem klasifikasi diterjemahkan sebagai upaya menemukan garis (*hyperplane*) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) ditunjukkan pada gambar 2.2. bagian (a).

Hyperplane pemisah terbaik antara kedua class dapat ditemukan dengan mengukur margin hyperplane dan mencari titik maksimalnya. Margin adalah jarak antara hyperplane tersebut dengan data terdekat dari masing-masing class. Subset data training set yang paling dekat ini disebut sebagai support vector. Garis solid pada Gambar 2.2. bagian (b) menunjukkan hyperplane yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua class, sedangkan titik kotak dan lingkaran yang berada dalam lingkaran hitam adalah support vector. Upaya mencari lokasi hyperplane optimal ini merupakan inti dari proses pembelajaran pada SVM.

Salah satu metode statistik yang dapat diterapkan untuk melakukan klasifikasi adalah Support Vector Machine (SVM). SVM merupakan suatu teknik untuk menemukan hyperplane yang bisa memisahkan dua set data dari dua kelas yang berbeda (Vapnik, 1999). SVM memiliki kelebihan diantaranya adalah dalam menentukan jarak menggunakan support vector sehingga proses komputasi

menjadi cepat (Vapnik, 1995). Penelitian tentang SVM telah dilakukan oleh Rustam, et al (2003) yaitu membandingkan metode klasifikasi K-Nearest Neighbor (KNN) dengan metode SVM diperoleh kesimpulan bahwa SVM memiliki kinerja yang lebih unggul, karena telah mampu 100% mengklasifikasikan data aroma berdasarkan kelas yang tepat. Selain itu Rachman dan Purnami (2012) yang melakukan penelitian mengenai klasifikasi tingkat keganasan kanker dengan menggunakan metode regresi logistik dan SVM yang akhirnya diperoleh hasil bahwa tingkat akurasi menggunakan SVM lebih tinggi, yaitu sebesar 98,11%.

2.7.1 Karakteristik SVM

Berikut ini karakteristik yang dimiliki oleh algoritma Support Vector Machine (SVM):

1. Secara prinsip SVM adalah *linear classifier*
2. *Pattern recognition* dilakukan dengan mentransformasikan data pada *input space* ke ruang yang berdimensi lebih tinggi, dan optimisasi dilakukan pada ruang vector yang baru tersebut. Hal ini membedakan SVM dari solusi *pattern recognition* pada umumnya, yang melakukan optimisasi parameter pada ruang hasil transformasi yang berdimensi lebih rendah daripada dimensi *input space*.
3. Menerapkan strategi *Structural Risk Minimization* (SRM)
4. Prinsip kerja SVM pada dasarnya hanya mampu menangani klasifikasi dua class.

2.7.2 Kelebihan dan Kelemahan SVM

Dalam memilih solusi untuk menyelesaikan suatu masalah, kelebihan dan kelemahan masing-masing metode harus diperhatikan. Selanjutnya metode yang tepat dipilih dengan memperhatikan karakteristik data yang diolah. Dalam hal SVM, walaupun berbagai studi telah menunjukkan kelebihan metode SVM dibandingkan metode konvensional lain, SVM juga memiliki berbagai kelemahan.

a. Kelebihan SVM

1) Generalisasi

Generalisasi didefinisikan sebagai kemampuan suatu metode untuk mengklasifikasikan suatu *pattern*, yang tidak termasuk data yang dipakai dalam fase pembelajaran metode itu. Vapnik menjelaskan bahwa generalization error dipengaruhi oleh dua faktor: error terhadap training set, dan satu faktor lagi yang dipengaruhi oleh dimensi VC (*Vapnik-Chervokinensis*). Strategi pembelajaran pada *neural network* dan umumnya metode *machine learning* difokuskan pada usaha untuk meminimalkan error pada training-set. Strategi ini disebut *Empirical Risk Minimization* (ERM). Adapun SVM selain meminimalkan error pada training-set, juga meminimalkan faktor kedua. Strategi ini disebut *Structural Risk Minimization* (SRM), dan dalam SVM diwujudkan dengan memilih hyperplane dengan margin terbesar. Berbagai studi empiris menunjukkan bahwa pendekatan SRM pada SVM memberikan error generalisasi yang lebih kecil

daripada yang diperoleh dari strategi ERM pada neural network maupun metode yang lain.

2) Curse of dimensionality

Curse of dimensionality didefinisikan sebagai masalah yang dihadapi suatu metode *pattern recognition* dalam mengestimasikan parameter (misalnya jumlah hidden neuron pada neural network, stopping criteria dalam proses pembelajaran dsb.) dikarenakan jumlah sampel data yang relatif sedikit dibandingkan dimensional ruang vektor data tersebut. Semakin tinggi dimensi dari ruang vektor informasi yang diolah, membawa konsekuensi dibutuhkannya jumlah data dalam proses pembelajaran. Pada kenyataannya seringkali terjadi, data yang diolah berjumlah terbatas, dan untuk mengumpulkan data yang lebih banyak tidak mungkin dilakukan karena kendala biaya dan kesulitan teknis. Dalam kondisi tersebut, jika metode itu "terpaksa" harus bekerja pada data yang berjumlah relatif sedikit dibandingkan dimensinya, akan membuat proses estimasi parameter metode menjadi sangat sulit. *Curse of dimensionality* sering dialami dalam aplikasi di bidang biomedical engineering, karena biasanya data biologi yang tersedia sangat terbatas, dan penyediaannya memerlukan biaya tinggi. Vapnik membuktikan bahwa tingkat generalisasi yang diperoleh oleh SVM tidak dipengaruhi oleh dimensi dari input vector. Hal ini merupakan alasan mengapa SVM merupakan salah satu metode yang tepat

dipakai untuk memecahkan masalah berdimensi tinggi, dalam keterbatasan sampel data yang ada.

3) Landasan teori

Sebagai metode yang berbasis statistik, SVM memiliki landasan teori yang dapat dianalisa dengan jelas, dan tidak bersifat kuliah umum.

4) *Feasibility*

SVM dapat diimplementasikan relative mudah, karena proses penentuan support vector dapat dirumuskan dalam QP problem. Dengan demikian jika kita memiliki library untuk menyelesaikan QP problem, dengan sendirinya SVM dapat diimplementasikan dengan mudah. Selain itu dapat diselesaikan dengan metode sekuensial sebagaimana penjelasan sebelumnya.

b. Kelemahan SVM

SVM memiliki kelemahan atau keterbatasan, antara lain:

1. Sulit dipakai dalam *problem* berskala besar. Skala besar dalam hal ini dimaksudkan dengan jumlah sampel yang diolah.
2. SVM secara teoritik dikembangkan untuk *problem* klasifikasi dengan dua class. Dewasa ini SVM telah dimodifikasi agar dapat menyelesaikan masalah dengan *class* lebih dari dua, antara lain strategi *One versus rest* dan strategi *Tree Structure*. Namun

demikian, masing-masing strategi ini memiliki kelemahan, sehingga dapat dikatakan penelitian dan pengembangan SVM pada *multiclass-problem* masih merupakan tema penelitian yang masih terbuka.

2.8 Python

Python adalah Bahasa pemrograman interpretatif yang dianggap mudah dipelajari serta berfokus pada keterbacaan kode. Dengan kata lain, Python diklaim sebagai bahasa pemrograman yang memiliki kode-kode pemrograman yang sangat jelas, lengkap, dan mudah untuk dipahami. Python secara umum berbentuk pemrograman beriorentasi objek, pemrograman imperatif, dan pemrograman fungsional. Python dapat digunakan untk berbagai keperluan perangkat lunak dan dapat berjalan di berbagai platform system operasi. Python memiliki beberapa fitur dan kelebihan adalah(Jubilee Enterprise, 2017):

1. Memiliki koleksi perpustakaan yang banyak, itu artinya telah tersedia modul-modul siap pakai untuk berbagai keperluan.
2. Memiliki sturuktur bahasa yang jelas, sederhana, dan mudah dipelajari.
3. Beriorentasi objek.
4. Memiliki system pengelolaan memori otomatis(*garbage collection)* seperti halnya java.
5. Bersifat modular sehingga mudah dikembangkan dengan menciptakan modul-modul baru, baik dengan bahasa Python maupun C/C++.