

Klasifikasi Ucapan Kata dengan Support Vector Machine

Sukmawati Nur Endah, Dinar Mutiara KN
Program Studi Teknik Informatika FSM
Universitas Diponegoro
sukma_ne@yahoo.co.id; dinar.mutiara@undip.ac.id

Menurut Undang-Undang No 32 Tahun 2002, Peraturan Komisi Penyiaran Indonesia No 02/P/KPI/12/2009 tentang Pedoman Perilaku Penyiaran dan Peraturan Komisi Penyiaran Indonesia No 03/P/KPI/12/2009 tentang Standar Program Siaran, diantaranya disebutkan bahwa program siaran tidak boleh memaki dengan kata-kata kasar, tidak boleh melecehkan, menghina atau merendahkan kelompok masyarakat minoritas dan marginal. Namun untuk mensensor kata-kata tersebut secara otomatis belum ada. Sehingga perlu adanya suatu perangkat lunak cerdas yang dapat mensensor kata-kata tersebut secara otomatis. Untuk dapat mensensor, perangkat lunak harus mampu mengenali kata-kata yang dimaksud. Penelitian ini mengusulkan pengklasifikasian ucapan kata dalam dua kelas, yaitu kelas kata-kata yang berisi makian kasar dan kelas kata yang bukan termasuk kata makian dengan menggunakan metode *Support Vector Machine* (SVM). Data ucapan kata yang diambil merupakan data suara untuk kelompok umur 19 – 21 tahun yang terdiri dari 19 pria dan 8 wanita. Masing-masing orang mengucapkan 25 kata makian kasar dan 25 kata yang bukan termasuk kata makian. Berdasarkan hasil eksperimen dengan melibatkan 25 data training dan 25 data testing, untuk suara pria menghasilkan tingkat keakuratan sebesar 78,4% dan untuk suara wanita menghasilkan tingkat keakuratan sebesar 80%.

Kata kunci: *Support Vector Machine*, klasifikasi ucapan kata

I. PENDAHULUAN

Berdasarkan Undang-Undang No 32 Tahun 2002 tentang penyiaran, dalam pasal 4 disebutkan bahwa penyiaran sebagai kegiatan komunikasi massa mempunyai fungsi sebagai media informasi, pendidikan, hiburan yang sehat, kontrol dan perekat sosial [1]. Untuk memenuhi tujuan tersebut, isi siaran telah diatur dalam Undang-Undang No 32 Tahun 2002 Pasal 35 dan Pasal 36 serta dipertegas di Peraturan Komisi Penyiaran Indonesia No 02/P/KPI/12/2009 tentang Pedoman Perilaku Penyiaran dan Peraturan Komisi Penyiaran Indonesia No 03/P/KPI/12/2009 tentang Standar Program Siaran. Dalam peraturan di atas, diantaranya disebutkan bahwa program siaran tidak boleh memaki dengan kata-kata kasar, tidak boleh melecehkan, menghina atau merendahkan kelompok masyarakat minoritas dan marginal seperti kelompok dengan pekerjaan tertentu (pekerja rumah tangga, hansip, satpam), kelompok yang memiliki penyimpangan (waria), kelompok dengan ukuran dan bentuk fisik di luar normal (gigi tongos, gemuk, cebol, mata juling), kelompok yang memiliki cacat fisik (tuli, buta, bisu), kelompok yang memiliki cacat atau keterbelakangan mental (autis, idiot), dan kelompok pengidap penyakit tertentu (HIV/AIDS, kusta, epilepsi, alzheimer, latah) [2,3].

Sampai saat ini, untuk program-program siaran yang tidak ditayangkan secara *live* (langsung), kata-

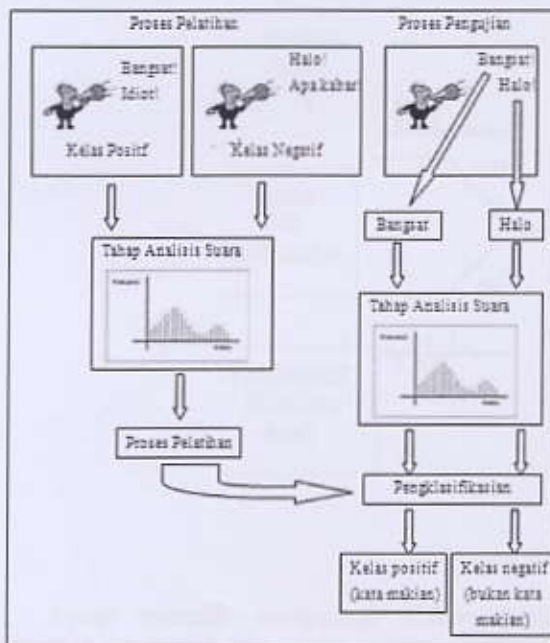
kata makian yang kasar, melecehkan ataupun menghina akan disensor dengan memberikan suara *bip* pada kata-kata tersebut. Namun, untuk siaran langsung hal tersebut tidak dapat dilakukan. Hal ini merupakan kekhawatiran Komisi Penyiaran Indonesia (KPI) yang disampaikan oleh Nina Mutmainah, Anggota KPI Pusat yang merangkap Koordinator Isi Siaran dalam Diskusi KPI dengan Grup Trans Corp di Gedung Trans Corp, Senin, 13 Februari 2012 [4]. Untuk itu, diperlukan suatu perangkat lunak cerdas yang mampu mensensor kata-kata yang tidak diinginkan secara otomatis dan *realtime*. Supaya dapat mensensor kata-kata yang tidak diinginkan, perangkat lunak harus mampu mengenali terlebih dulu kata yang terucap termasuk kata-kata makian atau bukan. Salah satu cara untuk mengenali kata yang terucap, dapat dilakukan pengklasifikasian kata, apakah termasuk kata-kata makian atau bukan.

Dalam beberapa penelitian sebelumnya, metode untuk melakukan pengklasifikasian ada beberapa macam, diantaranya jaringan syaraf tiruan untuk klasifikasi data aplikasi kredit di bank dan data mining [5,6], metode boosting untuk klasifikasi deteksi objek manusia [7], hidden markov model, Bayesian classification untuk klasifikasi dalam pengenalan suara [8], dan *support vector machine* untuk klasifikasi pose manusia [9,10]. Dari hasil

penelitian tersebut, pengklasifikasian dengan support vector machine mempunyai tingkat keakuratan yang tinggi. Sehingga dalam penelitian ini dilakukan pengklasifikasian ucapan kata dalam dua kelas yaitu kelas kata-kata makian yang kasar yang perlu disensor, dan kelas yang bukan kata-kata makian, dengan menggunakan metode *support vector machine*.

II. GAMBARAN UMUM SISTEM

Gambaran umum sistem yang dibangun adalah sebagai berikut.



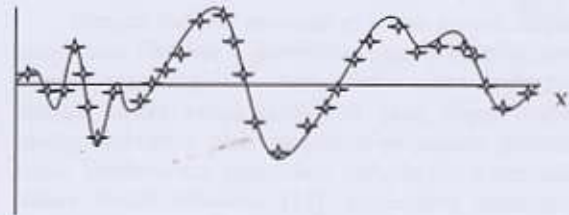
Gambar 1. Gambaran umum sistem

Tahap analisis suara merupakan tahap untuk mengambil nilai pitch dari ucapan kata yang telah direkam. Sedangkan untuk proses pengklasifikasiannya menggunakan metode SVM.

III. SINYAL SUARA

Definisi dari sinyal suara yaitu suatu sinyal yang mewakili dari suara [11]. Sinyal suara dibentuk dari kombinasi berbagai frekuensi pada berbagai amplitudo dan fasa. Untuk dapat melakukan proses pengolahan sinyal suara langkah pertama yaitu penyamplingan. Penyamplingan / pencuplikan adalah mengukur masukan sinyal suara pada sela waktu tertentu, kemudian mengkonversinya ke dalam skala tersendiri, dan menyimpannya. Gambar 2 menunjukkan sinyal suara yang dicuplik pada frekuensi 8 KHz. Ini berarti bahwa terjadi pengukuran masukan sinyal suara sebanyak 8000 kali

selama 1 detik. Fungsi pencuplikan yaitu agar sinyal suara dapat disimpan dalam bentuk digital.



Gambar 2. Sinyal suara yang dicuplik pada frekuensi pencuplikan 8 KHz [11]

Pada Gambar 3 menunjukkan sinyal suara yang sudah ditampilkan dalam bentuk digital, dimana bentuk sinyal bukanlah sinyal kontinu tetapi merupakan sinyal diskret.



Gambar 3. Sinyal suara dalam bentuk diskret [11]

Menganalisis suatu suara dapat diartikan menganalisis suara dalam komponen frekuensi dan waktu dari suara tersebut. Komponen frekuensi menentukan tinggi rendahnya suara sedangkan komponen waktu sangat menentukan kapan suatu suara dibunyikan. Dengan mengetahui komponen-komponen tersebut sinyal suara dapat dianalisis dengan rinci.

Semua informasi isyarat analog agar tidak hilang setelah mengalami pencuplikan, maka proses pencuplikan tersebut harus mengikuti teorema pencuplikan, yaitu jika komponen frekuensi tertinggi didalam suatu sinyal adalah f_{maks} , maka sinyal tersebut harus dicuplik pada frekuensi paling sedikit $2f_{maks}$ agar sampel dapat menggambarkan sinyal asli kembali secara penuh. Frekuensi cuplik dapat dinyatakan dengan $F_s \geq 2f_{maks}$, dimana F_s adalah frekuensi pencuplikan atau disebut laju Nyquist. Oleh

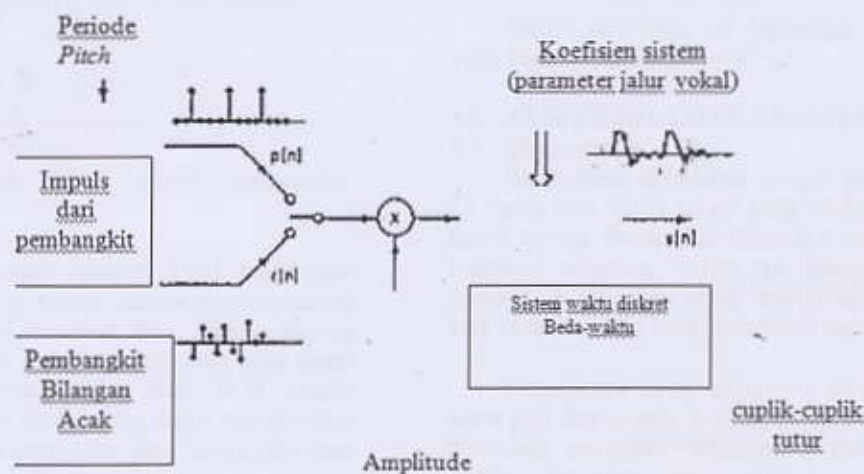
karena itu, jika komponen frekuensi maksimum dalam sinyal analog adalah 4 KHz, maka agar bisa mendapatkan semua informasi dalam sinyal tersebut, suara harus dicuplik dengan kecepatan frekuensi 8 KHz atau lebih.

IV. PITCH

Salah satu parameter dari sinyal suara adalah frekuensi fundamental. Frekuensi fundamental dalam istilah instrumen musik dikenal sebagai *pitch* atau nilai frekuensi dari suatu jenis nada. *Pitch* atau tinggi nada [12] adalah hasil akustik dari kecepatan getaran pita suara. Semakin cepat getaran pita suara, semakin

tinggi, tinggi nadanya. Sehingga *pitch* ini dapat digunakan sebagai ciri bersuara.

Dengan melihat pemodelan untuk sintesa sinyal suara pada Gambar 4, generator impuls memberikan sumber pembangkitan untuk sinyal suara berbunyi berupa fonem vokal (a/e/i/o/u) yang dapat diatur selang waktunya oleh parameter-parameter periode *pitch*. Berdasarkan penelitian, periode *pitch* berkisar antara 10-20 milidetik [13]. Disamping generator impuls, generator derau acak juga berfungsi sebagai sumber pembangkitan untuk sinyal suara tidak bersuara berupa fonem konsonan (h/k/s/t).



Gambar 4. Model sintesis suara [14]

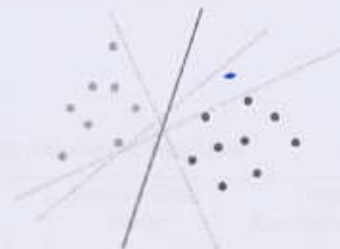
Setiap manusia mempunyai kisaran *pitch* tersendiri, tergantung dari bentuk pangkal tenggorok yang dimilikinya[13]. Untuk pria memiliki kisaran *pitch* sebesar 50-250 Hz. Dan untuk wanita memiliki kisaran *pitch* sebesar 120-500 Hz. Tinggi rendahnya nilai *pitch* tergantung pada intonasi suara, dan tingkat emosi dari manusia [13].

V. SVM

Support Vector Machine (SVM) dikembangkan oleh Boser, Guyon, Vapnik dan pertama kali dipresentasikan pada tahun 1992 di Annual Workshop on Computational Learning Theory. Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah kelas pada input space [15]. Hyperplane dalam ruang vector berdimensi d adalah affine subspace berdimensi $d-1$ yang membagi ruang vector tersebut ke dalam dua bagian yang masing-masing berkorespondensi pada kelas yang berbeda [16].

Masalah klasifikasi merupakan usaha menemukan garis (hyperplane) yang memisahkan

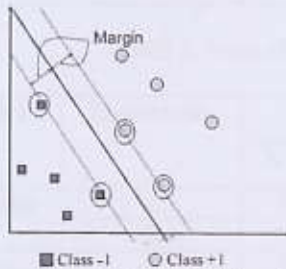
suatu objek dalam dua kelas. Seperti contoh pada Gambar 5 terlihat banyak kemungkinan garis klasifikasi linier yang dapat digunakan untuk memisahkan data, akan tetapi hanya ada satu garis yang memaksimalkan margin (jarak maksimum antara data terdekat pada setiap kelas). Garis klasifikasi linier tersebut dinamakan hyperplane pemisah yang optimal [17].



Gambar 5. Hyperplane pemisah yang optimal (Garis yang tebal)

Hyperplane pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur margin hyperplane tersebut dan mencari titik maksimalnya. Margin adalah jarak antara hyperplane tersebut

dengan pattern terdekat dari masing-masing kelas. Pattern yang terdekat ini disebut dengan support vector. Garis solid pada Gambar 6 menunjukkan hyperplane yang terbaik, yaitu yang terletak tepat pada tengah-tengah kelas kedua kelas. 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 [15].



Gambar 6. Usaha SVM dalam menemukan hyperplane terbaik

Klasifikasi dengan metode SVM melibatkan data *training* dan data testing dimana masing-masing terdiri dari beberapa masukan data. Masing-masing masukan dalam data *training* berisi satu nilai target dan beberapa atribut. Tujuan dari SVM adalah memproduksi sebuah model yang dapat memprediksi nilai target dari data pengujian yang hanya diberikan nilai atributnya.

Jika diberikan himpunan data pelatihan (x_i, y_i) , $i = 1, 2, \dots, l$ dimana $x_i \in \mathbb{R}^n$ dan $y \in \{-1, 1\}$, SVM membutuhkan solusi dari problem optimasi berikut,

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i$$

$$\text{subject to } y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0.$$

Disini vektor pelatihan x_i dipetakan ke dimensi yang lebih tinggi dengan suatu fungsi ϕ . Kemudian SVM akan mencari sebuah hyperplane yang

memisahkan secara linier dengan margin maksimal dalam dimensi yang lebih tinggi tersebut. $C > 0$ merupakan parameter pinalti dari error. Selanjutnya, $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ disebut fungsi kernel. Fungsi kernel ini ada beberapa macam, diantaranya fungsi linier, polynomial, radial basis function (RBF) dan sigmoid. [18]

- linier: $K(x_i, x_j) = x_i^T x_j$
- polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0.$
- radial basis function (RBF): $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0.$
- sigmoid: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

Dalam penelitian ini digunakan fungsi kernel yang linier yaitu $K(x_i, x_j) = x_i^T x_j$.

VI. EKSPERIMEN DAN ANALISA HASIL

6.1. Eksperimen

Eksperimen dilakukan dengan merekam suara 27 orang usia 19-21 tahun yang terdiri dari 19 pria dan 8 wanita. Perekaman dilakukan dengan rata-rata frekuensi sampling 44100 Hz, channel mono dan resolusi 16 bit. Tiap orang mengucapkan 25 (lima) kata kasar dan 25 (lima) kata tidak kasar.

Eksperimen yang dilakukan dibedakan untuk suara pria dan wanita. Masing-masing kategori suara dilakukan pengujian sebanyak 5 kali untuk data training dan data testing yang sama yaitu masing-masing 25 data dan sebanyak 2 kali untuk jumlah data training dan data testing yang berbeda. Setiap eksperimen mempunyai data kata yang berbeda. Data yang digunakan sebagai data training juga berbeda dengan data yang digunakan untuk data testing. Kelas positif merupakan kelas untuk kata-kata makian yang kasar, sedangkan kelas negatif merupakan kelas untuk kata yang bukan makian. Tabel 1 merupakan hasil eksperimen dengan jumlah data training sama dengan data testing, yaitu 25 data, sedangkan Tabel 2 merupakan hasil eksperimen dengan jumlah data training tidak sama dengan data testing.

Tabel 1. Hasil eksperimen dengan jumlah data training dan data testing sama

Suara	Eksperimen	Data Training			Data Testing			Jumlah Data Error	Tingkat Keakuratan (%)
		K +	K -	Σ	K +	K -	Σ		
Wanita	1	12	13	25	13	12	25	6	76%
	2	13	12	25	12	13	25	3	88%
	3	12	13	25	13	12	25	5	80%
	4	13	12	25	12	13	25	7	72%
	5	12	13	25	13	12	25	4	84%
Rata-rata									80%

Suara	Eksperimen	Data Training			Data Testing			Jumlah Data Error	Tingkat Keakuratan (%)
		K +	K -	Σ	K +	K -	Σ		
Pria	1	12	13	25	13	12	25	5	80%
	2	13	12	25	12	13	25	1	96%
	3	13	12	25	12	13	25	6	76%
	4	13	12	25	12	13	25	7	72%
	5	12	13	25	13	12	25	8	68%
Rata-rata									78,4%

Tabel 2. Hasil eksperimen dengan jumlah data training dan data testing tidak sama

Suara	Eksperimen	Data Training			Data Testing			Jumlah Data Error	Tingkat Keakuratan (%)
		K +	K -	Σ	K +	K -	Σ		
Wanita	1	5	5	10	20	20	40	11	72,5%
	2	20	20	40	5	5	10	2	80%
Pria	1	5	5	10	20	20	40	3	92,5%
	2	20	20	40	5	5	10	2	80%

6.2. Analisa Hasil

Dari tabel 1 terlihat bahwa rata-rata untuk suara pria menghasilkan tingkat keakuratan sebesar 78,4% dan untuk suara wanita menghasilkan tingkat keakuratan sebesar 80%. Adanya *error* dalam pengklasifikasian ini kemungkinan disebabkan oleh adanya data-data nilai pitch yang bernilai 0. Nilai 0 pada pitch dapat terjadi karena salah pemotongan kata pada saat proses perekaman ucapan kata. Dalam hal ini kata dipotong lebih maju untuk beberapa waktu sebelumnya. Sehingga frekuensi awal yang didapat adalah frekuensi sebelum seseorang berbicara.

Tingkat keakuratan pada suara wanita untuk data training yang lebih kecil dari data testing lebih rendah dibanding tingkat keakuratan untuk data training yang lebih semakin besar dari data testing. Hal ini karena semakin banyak data training, SVM semakin dapat membentuk hyperplane dengan sempurna. Kondisi yang ada kebalikan dengan data suara pria. Ini disebabkan karena atribut suara pria lebih banyak dari pada suara wanita.

VII. KESIMPULAN

Berdasarkan pembahasan sebelumnya dapat disimpulkan bahwa metode SVM dapat digunakan untuk mengklasifikasikan ucapan kata yang termasuk kata-kata makian dan kata yang bukan makian dengan tingkat keakuratan yang tinggi, yaitu berkisar 72% - 95%. Tinggi nada atau nilai pitch suatu ucapan kata dapat digunakan sebagai ciri atau atribut untuk proses pengklasifikasian. Hasil penelitian ini dapat digunakan untuk membuat sebuah sistem cerdas

otomatisasi penyensor ucapan kata makian yang kasar.

DAFTAR PUSTAKA

- [1] Soekarnoputri, M (Presiden RI). 2002. Undang-Undang No 32 Tahun 2002 Tentang Penyiaran. Lembaran Negara Republik Indonesia Tahun 2002 No 139. Jakarta
- [2] Sendjaja, S.D. (Ketua Komisi Penyiaran Indonesia Pusat). 2009. Peraturan Komisi Penyiaran Indonesia No 02/P/KPI/12/2009 tentang Pedoman Perilaku Penyiaran. Jakarta
- [3] Sendjaja, S.D. (Ketua Komisi Penyiaran Indonesia Pusat). 2009. Peraturan Komisi Penyiaran Indonesia No 03/P/KPI/12/2009 tentang Standar Program Siaran. Jakarta
- [4] RG. 2012. Diskusi KPI-Trans Corp: Hindari Tayangan Sensitif, http://www.kpi.go.id/index.php?option=com_content&view=article&id=30378%3Akpi-hindari-tayangan-sensitif&catid=14%3Adalam-negeri-umum&lang=id, 7 Maret 2012
- [5] Dhauswara, G dan V.S Moertini. 2004. Jaringan Syaraf Tiruan Propagasi Balik untuk Klasifikasi Data. Integral. FMIPA Unpar. Vol 9 No 3, Nov 2004
- [6] Maharani, W. 2009. Klasifikasi Data Menggunakan JST Back Propagation

- Momentum dengan Adaptive Learning Rate. Seminar Nasional Informatika 2009 (Semnas IF 2009), UPN Veteran Yogyakarta, 23 Mei 2009. ISSN 1979-2328
- [7] Fatihah, C., 2007. Tesis : Metode Boosting dengan Fungsi Pengklasifikasi Dasar Kernel untuk Deteksi Objek Manusia. Universitas Indonesia. Jakarta
- [8] Fung, P, and Roth, D., 2005. Guest Editor Introduction: Machine Learning in Speech and Language Technologies. Springer Science+Business Media. Netherland
- [9] Widyanto, M.R., Endah, S.N, and Hirota, K., 2010. Human Behavior Classification Using Thinning Algorithm and Support Vector Machine. Journal of Advanced Computational Intelligence and Intelligent Informatics. Volume 14 No. 1 Page 28-34, Januari 2010, ISSN 1343-0130, Jepang
- [10] Endah, S.N, dan Widyanto, M.R., 2008. Klasifikasi Pose Skeleton Manusia dengan Menggunakan Support Vector Machine. Jurnal Ilmu Komputer dan Informasi, Volume 1 Nomer 2 Halaman 27-30, Juni 2008, ISSN 1979-0732. Fasilkom UI
- [11] Wismono, R. Yudhi, 2002. *Identifikasi Jenis Tingkatan Suara Manusia Dengan Metode Real Cepstrum*, Teknik Elektro Universitas Diponegoro, Semarang, 2002.
- [12] Chaer, Abdul, 1994. *Linguistik Umum*, Cetakan Pertama, Penerbit PT. Rineka Cipta, Jakarta.
- [13] Deller Jr, John R., dan Proakis, John G., 1993. *Discrete – Time Processing of Speech Signal*, Mac Millan Publising Company, New York.
- [14] Oppenheim, Alan V., 1998. *Digital Signal Processing*, Prentice-Hall of India.
- [15] Nugroho, A.S., Witarto, A.B, Handoko, D., Application of Support Vector Machine in Bioinformatics, Proceeding of Indonesian Scientific Meeting in Central Japan, December 20, 2003, Gifu-Japan
- [16] Cristianini N., Taylor J.S., *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge Press University, 2000
- [17] Gunn, Steve R., 1998. Technical Report: Support Vector Machine for Classification and Regression. Faculty of Engineering, Science and Mathematics, University of Southampton, 14 May 1998
- [18] Hsu, C., Chang, C., and Lin, C., 2008. *A Practical Guide to Support Vector Classification*. Department of Computer Science, National Taiwan University, Taiwan