



PERBANDINGAN ALGORITMA NAÏVE BAYES, SVM DAN TREES J48 PADA PENGENALAN PENGARUH SUARA KONSONAN TERHADAP VOKAL

Muhammad Makmun Effendi

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa
effendiyan@gmail.com

Abstrak

Suara merupakan alat komunikasi paling mendasar yang dimiliki manusia. Perkembangan teknologi yang semakin tinggi sehingga kebutuhan manusia dalam menjalani kehidupannya dengan praktis dan otomatis, maka perkembangan teknologi ke arah bidang biometrica. Penelitian ini dilakukan untuk mengetahui apakah antara jenis suara pria dan wanita dapat dikenali dengan merekam sebatas huruf konsonan yang telah di pengaruhi vokal, dengan membandingkan algoritma Naive Bayes, algoritma SVM, dan J48. Metode Penelitian yang digunakan dalam membangun sistem pengenalan suara ini berdasarkan konsonan yang telah di pengaruhi oleh vokal dengan membandingkan algoritma Naive Bayes, SVM dan Trees J48. Hal ini dilakukan untuk mengetahui dari kedua algoritma yang dibandingkan yang mana yang lebih baik dalam mengenali suara konsonan yang telah dipengaruhi oleh suara vokal.

Kata kunci : **Suara , Naïve Bayes, SVM dan Trees J48**

Abstract

Sound is the most basic communication tool that humans have. Technological developments are increasingly high so that human needs in living his life with practical and automatic, then the development of technology toward the field biometrica. This study was conducted to determine whether between the male and female voice types can be recognized by recording the consonant letters that have been influenced by vowels, by comparing the Naive Bayes algorithm, SVM algorithm, and J48. Research methods used in building this voice recognition system based on consonants that have been influenced by vowels by comparing the Naïve Bayes, SVM and Trees J48 algorithms. This is done to find out from the two algorithms compared which are better at recognizing consonant sounds that have been affected by the vowel sound.

Keywords : *Sound, Naïve Bayes, SVM and Trees J48*

1. Pendahuluan

1.1 Latar Belakang

Komunikasi dapat dilakukan dengan berbagai cara diantaranya melalui percakapan, bahasa verbal, mimik muka, gambaran atau lukisan, karya seni, ataupun teknologi. Seiring kemajuan zaman dimana teknologi semakin berkembang, maka manusia tidak

hanya berinteraksi dengan manusia saja, tetapi seringkali juga harus berinteraksi dengan perangkat-perangkat teknologi.

Ide dasar dari penelitian ini adalah untuk membandingkan nilai akurasi dari algoritma naïve bayes ,algoritma SMO,dan J48 apakah kedua algoritma tersebut dapat mengenali suara pria dan wanita dajn juga ingin mengetahui algoritma mana yang lebih baik dari kedua algoritma yang akan dibandingkan dalam penelitian ini.

1.2 Rumusan Masalah

Untuk Melakukan penelitian apakah fitur yang diperoleh dari suara konsonan dalam Bahasa Indonesia dapat dikenali dan dapat mengenali suara pria dan wanita dengan menggunakan metode Naïve Bayes, SVM, J48 dan juga untuk mengetahui berapa tingkat akurasi yang dapat dicapai.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut :

- 1) Data Rekaman yang digunakan sebagai data pelatihan dan data testing
- 2) Menggunakan Metode *Naïve Bayes, dan SVM* serta J48
- 3) Fokus ucapan yang dikenali adalah dengan menggunakan bahasa Indonesia dengan menggunakan suara wanita dan pria.

1.4 Tujuan Penelitian

Tujuan utamanya adalah untuk mengetahui nilai akurasi pengaruh konsonan terhadap vokal terhadap suara pria dan wanita

1.5 Manfaat Penelitian

Untuk menunjang sistem pengenalan ucapan, dan juga untuk mengetahui seberapa besar pencapaian akurasi suara pria dan wanita dengan membandingkan algoritma Naïve Bayes, SVM dan J48.

1.6 Nilai Kebaruan Penelitian

Penelitian disini adalah membandingkan algoritma Naïve Bayes dengan algoritma SVM, dan J48 pada pengaruh suara konsonan terhadap vokal untuk pria dan wanita

2. Studi Pustaka

2.1 Suara

Bunyi atau suara adalah sesuatu yang dihasilkan oleh getaran dari benda yang bergetar. Suara yang dihasilkan oleh suatu sumber getar dapat sampai ke telinga karena suara merambat sebagai gelombang. Bunyi terjadi secara berkelanjutan dikarenakan adanya gelombang analog. Untuk merubah gelombang analog ke dalam komputer dilakukan dengan cara melakukan digitalisasi gelombang analog tersebut

2.2 J48, Naïve Bayes dan SVM

Metode penelitian yang akan digunakan sebagai pembandingan adalah Algoritma J-48 yang merupakan implementasi c4.5 di WEKA. Keunggulannya adalah mampu menangani atribut dengan type diskrit dan kontinu., menangani atribut kosong dan memotong cabang.

Metode selanjutnya adalah SVM Boser bersama Guyon dan Vapnik mengembangkan SVM dan pada tahun 1992 pertama kali di presentasikan di *Annual Workshop on Computational Learning Theory*.

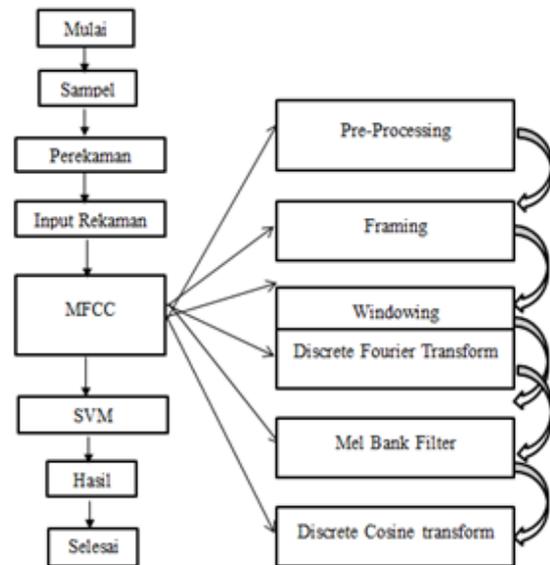
Secara sederhana konsep SVM dipresentasikan sebagai usaha dalam mencari Hyperplane yang lebih baik dan untuk memisahkan dua buah kelas pada input space. Anggota yang terdiri dari dua buah kelas masing masing kelas +1 dan kelas -1 dan berbagi alternative garis pemisah salah satu anggotanya adalah pattern. Untuk jarak antara *hyperplane* tersebut dari masing masing kelas dengan *patern* yang terdekat disebut sebagai *support vektor* adalah *margin*.

Adapun beberapa keuntungan dari metode SVM adalah sebagai *Generalisasi*, *Curse of dimensionality* dan *Feasibility*.

Selanjutnya adalah Metode yang menjadikan perbandingan adalah naive Bayes, pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai **teorema Bayes**.

3. Metode penelitian

Langkah-langkah yang akan digunakan dalam penelitian yaitu sebagai berikut:



Gambar 1. Langkah Penelitian

4. Hasil Penelitian

Pada bagian ini akan dijelaskan hasil penelitian mulai dari deskripsi data, uji persyaratan analisis, pengujian hipotesa, pembahasan dan berbagai permasalahan yang ditemui dalam penelitian lapangan.

Jumlah Data dalam penelitian ini adalah sbb:

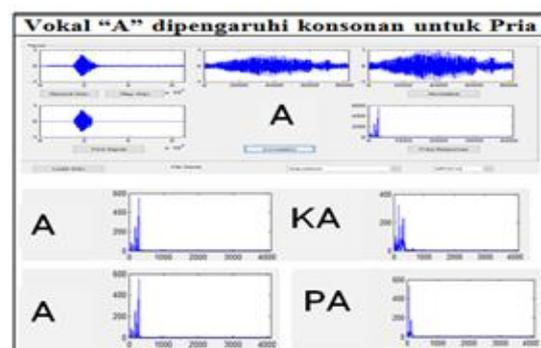
2 Pria dan 1 orang wanita. Adapun data yang didapat 21 konsonan dikalikan dengan 5x perulangan dikalikan dengan 6 vokal maka hasil frekuensi yang didapat adalah

$21 \times 5 \times 6 = 630$ data /orang, untuk 3 orang = 1890 data frekuensi.

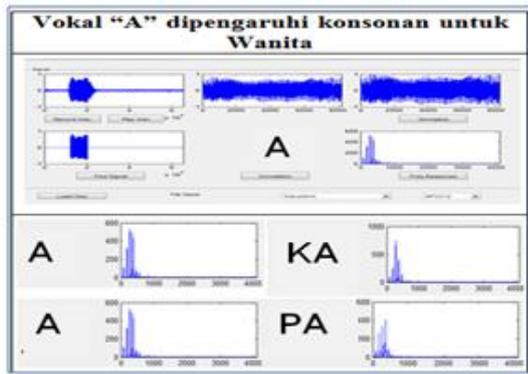
1x suara menghasilkan besaran frekuensi 4097 maka data yang dihasilkan 4097 x 1890.

Untuk data *coefisien corelation* adalah jumlah huruf koefisien kolerasi konsonan terhadap vokal x jumlah pengucapan x jumlah huruf vokal $(105 \times 5 \times 6) = 3150$, dikalikan 3 orang = 9450.

Hasil gambar sinyal 4.1 untuk pria dan wanita, dibawah ini adalah huruf vokal "A" setelah dipengaruhi oleh konsonan "K" dan "P", maka ada perubahan pada *frekuensi respond* spektrum seperti pada gambar dibawah ini, Ada Perbedaan Vokal "A" dengan konsonan "Ka", Vokal "A" dengan konsonan "Pa" dan lihat adanya perbedaan juga Spektrum konsonan "Ka" dan "Pa",



Gambar 2. Sinyal Suara Pria



Gambar 3. Sinyal Suara Wanita

Untuk selanjutnya hasil dari Weka untuk pengenalan genre pria dan wanita pada konsonan yang dipengaruhi vokal adalah sebagai berikut :

Tabel 1. Hasil Training SVM, Naives Bayes dan Trees-J48

Classification	SVM (AU)		SVM (wD)		Naive Bayes		Trees-J48		
	Laki (a)	Perempuan (b)	Laki (a)	Perempuan (b)	Laki (a)	Perempuan (b)	Laki (a)	Perempuan (b)	
Instances	504		504		1260		1260		
Attributes	409	AVG	409	AVG	409	AVG	409	AVG	
Success	501/99.4048%		481/95.4365%		1087/79.5230%		1081/85.7921%		
Subst	3/1.8925%		41/12.8946%		253/20.0794%		197/14.2636%		
TP	1	0.988	0.994	0.825	0.921	0.873	0.884	0.744	0.799
FP	0.012	0	0.006	0.079	0.179	0.127	0.256	0.246	0.201
Precision	0.998	1	0.994	0.912	0.841	0.874	0.77	0.838	0.77
Recall	1	0.988	0.994	0.825	0.921	0.873	0.884	0.744	0.799
F-Measure	0.994	0.994	0.994	0.867	0.879	0.873	0.81	0.788	0.799
ROC	0.994	0.994	0.994	0.873	0.873	0.873	0.799	0.81	0.804
a / Pria	252	0	20	50	44	58	92	58	92
b / Perempuan	3	240	20	232	161	469	141	469	
	0.988		0.746		0.984		0.749		
	0.066		0.117		0.208		0.145		
	0.972		0.360		0.448		0.372		
	1.00%		25.40%		40.16%		29.60%		
	15.43%		71.27%		89.62%		74.68%		
	504		504		1260		1260		



Gambar 4. Kinerja pengaruh konsonan terhadap vokal Trees J48



Gambar 5. Kinerja pengaruh konsonan terhadap vokal SVM



Gambar 6. Klasifikasi Pengaruh Konsonan - vokal pada Pria dan



Gambar 7. Kinerja pengaruh konsonan terhadap vokal Naive

Dari data diatas bisa dilihat bahwa hasil uji klasifikasi sampel terhadap tiga metode (SVM, Naives Bayes dan Trees-J48). Metode SVM memiliki tingkat pencapaian yang baik jika dibandingkan dengan metode lain yaitu sebesar 93.35%.

Berdasarkan hasil pembahasan pada bab 4, maka hasil dari pengujian klasifikasi adalah sebagai berikut :

- 1) Pengaruh konsonan terhadap vokal sampel ke satu lebih signifikan jika dibandingkan dengan pengaruh konsonan pada sampel kedua
- 2) Untuk Naive Bayes, SVM dan Trees-J48 dilakukan pengujian menggunakan Tools Seka dan hasil darinya Pengaruh konsonan terhadap vokal dapat di prediksi.
- 3) SVM memiliki hasil yang lebih baik yaitu 93.35% dibanding dengan Naives Bayes dan Trees-J48
- 4) Trees-J4 menghasilkan 85.79% lebih baik dibandingkan dengan Naives Bayes yang berkisar diantar 79.92%
- 5) Hasil dari SVM, Naives Bayes maupun Trees-J48 dapat pada penelitian ini adalah bahwasanya pengaruh konsonan terhadap vokal dapat dikenali dengan menggunakan ketiga metode tersebut.

Daftar Pustaka

[1] Anil Kumar Vuppala, Saswat Chakrabarti, and K. Sreenivasa Rao, Effect of Speech Coding on

- Recognition of Consonant-Vowel (CV) Units 1G. S. Sanyal School of Telecommunications 2School of Information Technology Indian Institute of Technology Kharagpur Kharagpur – 721302, West Bengal, India.
- [2] Becchetti, Claudio dan Ricotti, Lucio Prina. 1999. *Speech Recognition Theory and C++ Implementation*. Chichester : John Willey & Sons.
- [3] Chou, Wu dan Juang, B. H., 2003. *Pattern Recognition in Speech and Language Processing*. Washington DC : CRC PRESS.
- [4] John. G, and Dimitris G. Manolakis, *Pemrosesan Sinyal Digital : Prinsip- prinsip, Algoritmadan Aplikasi*, PT Prenhallindo, Jakarta, 1995.
- [5] Lesmana, Dody Putu. 2012. 'Perbandingan Kinerja Decision Tree J48 dan ID3 Dalam Pengklasifikasian Diagnosis Penyakit Diabetes Mellitus'. *Jurnal 6 Teknologi dan Informatika*, Vol. 2, no. 2.
- [6] Pedoman Umum Ejaan yang disempurnakan / EYD oleh Menteri Pendidikan & Kebudayaan berdasarkan surat keputusannya pada tanggal 12 Oktober 1972, No.156/P/1972(Amran Halim,Ketua)
- [7] Rabiner, Lawrence and Biing-Hwang Juang, *Fundamentals of Speech Recognition*, PTR Prentince-Hall Inc., New Jersey, 1993.
- [8] Santra, A. K., Jayasudha, S., 2012, Classification of Web Log Data to Identify Interested Users Using Naive Bayesian Classification, *International Journal of Computer Science Issues (IJCSI)*, ISSN (online) : 1694-0814, Vol. 9 Issue.1 No. 2, January 2012.
- [9] T. Zachary Burkle , *Accepted by The Faculty of the Graduate School, Indiana University*,
- [10] .Widiastuti, Dwi. 2010. 'Analisa Perbandingan Algoritma SVM, Naive Bayes, dan Decision Tree dalam Mengklasifikasikan Serangan (Attacks) pada Sistem Pendeteksi Instrusi', *Jurnal Jurusan Sistem Informasi , Universitas Gunadarma*.