

Algoritma *Support Vector Machine* (SVM) untuk Klasifikasi Ekonomi Penduduk Penerima Bantuan Pemerintah di Kecamatan Simpang Raya Sulawesi Tengah

Anita Ahmad Kasim^{*1}, Muhammad Sudarsono²

^{*1}Jurusan Teknologi Informasi, Fakultas Teknik, Universitas Tadulako

²Program Studi Teknik Informatika, Fakultas Teknik Universitas Tadulako

e-mail: ^{*1}nita.kasim@gmail.com, ²mamadndeso123@gmail.com

Abstrak

*Klasifikasi ekonomi penduduk diperlukan dalam menentukan penerima bantuan pemerintah. Pemberian bantuan terkadang tidak sesuai sasaran. Hal ini disebabkan oleh proses penentuan penerima bantuan masih dilakukan secara manual sehingga terdapat penilaian subjektif dalam menentukan masyarakat yang berhak menerima atau tidak menerima bantuan pemerintah. Hal ini dianggap kurang baik sehingga mengakibatkan distribusi bantuan menjadi kurang merata dan mempengaruhi kesejahteraan masyarakat. Penelitian terkait klasifikasi ekonomi penduduk menggunakan parameter yang telah ditentukan oleh seperti pendidikan terakhir, dinding rumah, atap rumah, luas rumah dan sumber listrik. Parameter tersebut diolah menggunakan algoritma *Support Vector Machine* (SVM) dengan tujuan klasifikasi ekonomi penduduk bagi penerima bantuan dan bukan penerima bantuan pemerintah di Kecamatan Simpang Raya Kabupaten Banggai Provinsi Sulawesi Tengah. Penelitian ini menggunakan data yang terdiri dari data latih sebanyak 320 data dan data uji sebanyak 80 data. Berdasarkan hasil uji coba keakuratan algoritma SVM didapatkan hasil akurasi sebesar 98%.*

Kata kunci— Machine Learning, klasifikasi, Support Vector Machine

Abstract

*The economic classification of the population is needed in determining the recipients of government assistance. Providing assistance is sometimes inappropriate for potential beneficiaries. The process of determining beneficiaries is still done manually so that there is a subjective assessment in determining who is entitled to receive or not receive government assistance. This results in the distribution of aid being uneven and affecting the welfare of the community. Research on the economic classification of the population uses parameters that have been determined by such as recent education, house walls, roofs, house area and electricity sources. These parameters are processed using the *Support Vector Machine* (SVM) algorithm with the aim of classifying the economic population of beneficiaries and not recipients of government assistance in Kecamatan Simpang Raya, Kabupaten Banggai, Propinsi Sulawesi Tengah. This study uses data consisting of 320 training data and 80 test data. Based on the results of the SVM algorithm accuracy test results obtained an accuracy of 98%.*

Keywords— Machine Learning, classification, Support Vector Machine

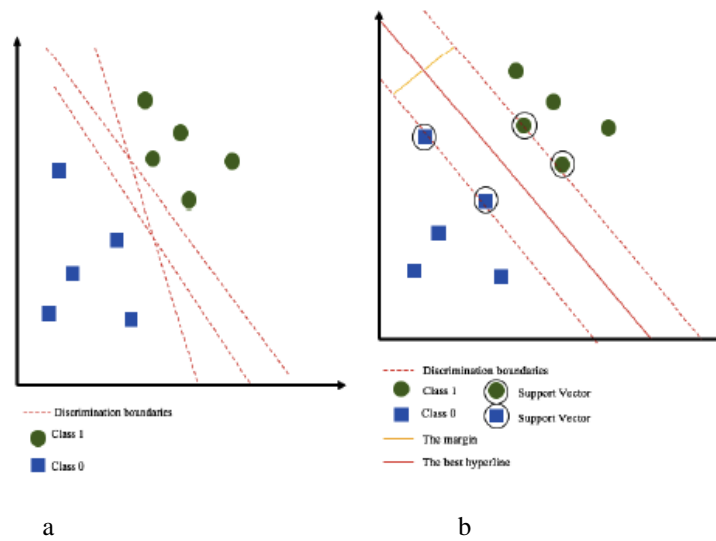
1. PENDAHULUAN

Status sosial ekonomi sebuah keluarga menggambarkan kondisi individu ditinjau dari segi ekonomi dengan gambaran seperti tingkat pendidikan, pendapatan dan pekerjaan[1]. Status sosial ekonomi merupakan posisi yang ditempati individu atau keluarga yang berkenaan dengan

ukuran rata-rata yang umum berlaku tentang kepemilikan budaya, pendapatan efektif, kepemilikan barang dan partisipasi dalam aktivitas kelompok dan komunitasnya. Upaya pemerintah untuk meningkatkan status ekonomi sebuah keluarga dilakukan dengan pemberian bantuan bagi masyarakat dengan status ekonomi rendah/miskin. Penentuan status sosial ekonomi penduduk yang dilakukan secara manual memungkinkan keputusan yang bersifat subjektif. Penelitian yang pernah dilakukan untuk menentukan status ekonomi pernah dilakukan menggunakan metode K-Nearest Neighbour (KNN). Penelitian ini dilatar belakangi proses pendataan dan klasifikasi tingkat ekonomi warga Desa Pacewetan yang masih manual, sehingga prosesnya kurang efisien. Pengontrolan kesejahteraan masyarakat pada Desa Pacewetan juga masih belum dilakukan secara mendetail, maka tingkat kesejahteraan warganya tidak terkontrol dengan baik[2]. Selain KNN, algoritma lain yang digunakan untuk klasifikasi adalah *Naive Bayes Classifier*. Penggunaan algoritma *Naive Bayes Classifier* digunakan untuk klasifikasi kesejahteraan rumah tangga menggunakan variabel target/kelas berupa kategori miskin dan tidak miskin. Hasil pengujian akurasi yang diperoleh algoritma Naive Bayes sebesar 85.80%[3]. Algoritma Naive Bayes memiliki beberapa kekurangan diantaranya Nilai k bias, Komputasi kompleks, keterbatasan memori, dan memiliki beberapa atribut yang tidak relevan[4]. Berdasarkan hal tersebut, penelitian ini bertujuan melakukan klasifikasi ekonomi penduduk dengan menggunakan algoritma Support Vector Machine (SVM). *Support Vector Machines* (SVM) adalah suatu metode yang handal dalam menyelesaikan masalah klasifikasi data. Penggunaan model SVM mengolah data menjadi data latih dan data uji. Data latih digunakan dalam membentuk model SVM, sementara nilai parameter bebasnya dipilih dari data awal. Selanjutnya model SVM yang dihasilkan digunakan untuk mengklasifikasi data uji. Data yang digunakan dalam penelitian ini adalah data penduduk di Kecamatan Simpang Raya Kabupaten Banggai Sulawesi Tengah. Artikel ini terdiri dari lima bagian yaitu pendahuluan, metode penelitian, hasil dan pembahasan, kesimpulan dan saran.

2. METODE PENELITIAN

Langkah-langkah yang dilakukan dalam penelitian ini diawali dengan input data parameter untuk klasifikasi ekonomi penduduk. Parameter yang diinputkan meliputi pendidikan terakhir, dinding rumah, atap rumah, luas rumah dan sumber listrik. Penelitian dilakukan melalui simulasi komputer menggunakan model SVM linear. Metode Mesin Vector Support berfungsi untuk memisahkan dua kelas pada data input. Gambar 1 menunjukkan data dari dua kelas yaitu data kelas 1 dan data kelas 0 yang dilambangkan dengan kotak biru dan lingkaran hijau. Masalah klasifikasi dapat diselesaikan dengan metode klasifikasi SVM. Hal ini dilakukan dengan menemukan garis yang dapat memisahkan dua kelompok data. Separator hyperplane terbaik dapat diperoleh dengan mengukur margin hyperplane dan titik maksimumnya. Margin adalah jarak antara hyperplane dan data terdekat dari masing-masing kelas. Data yang paling dekat dengan hyperplane terbaik disebut Support Vector. Garis solid pada Gambar 1 menunjukkan *hyperplane* terbaik. Hyperplane paling baik terletak tepat di kelas kedua (kelas 0)[5]. Model support Vector Machine dapat dilihat pada Gambar 1.



Gambar 1 Model Support Vector Machine

Pemahaman sederhana konsep SVM digambarkan sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas[6]. Gambar 1 a dan 1 b memperlihatkan beberapa pola yang merupakan anggota dari dua buah kelas yaitu 1 dan 0. Pola yang tergabung pada kelas 1 digambarkan dengan lingkaran hijau sedangkan pola pada kelas 0 digambarkan dengan kotak biru. Masalah klasifikasi dapat dijabarkan dengan usaha menemukan *hyperplane* yang memisahkan dua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) ditunjukkan pada Gambar 1 a. *Hyperplane* pemisah yang terbaik diantara kedua kelas ditemukan dengan cara mengukur margin *hyperplane* dan mencari titik maksimalnya. Margin adalah jarak antara *hyperplane* dengan pola terdekat dari setiap kelas. Pola yang paling dekat ini disebut sebagai *support vector*. Garis solid pada Gambar 1 b menunjukkan *hyperplane* yang terbaik, yaitu yang terletak tepat pada tengah-tengah kedua kelas, sedangkan titik hijau dan biru yang berada dalam lingkaran hitam adalah *support vector*. Usaha untuk mencari lokasi *hyperplane* ini merupakan inti dari proses pembelajaran pada SVM. Data dinotasikan sebagai $\vec{x}_i \in R^2$ sedangkan label masing-masing dinotasikan $y_i \in \{1,0\}$ untuk $i = 1,2, \dots, l$ yang mana 1 adalah banyaknya data. Asumsi kedua kelas 1 dan 0 dapat terpisah secara sempurna oleh *hyperplane* berdimensi d yang didefinisikan pada Persamaan 1

$$\vec{w} \cdot \vec{x} + b = 0 \tag{1}$$

Pola \vec{x}_i yang termasuk kelas 1 dapat dirumuskan sebagai pola yang memenuhi pertidaksamaan (2)

$$\vec{w} \cdot \vec{x}_i + b \leq 1 \tag{2}$$

Sedangkan pola \vec{x}_i yang termasuk kelas 0 dirumuskan dengan pertidaksamaan (3)

$$\vec{w} \cdot \vec{x}_i + b \geq -1 \tag{3}$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya dengan persamaan (4)

$$\frac{1}{\|\vec{w}\|} \tag{4}$$

Hal ini dapat dirumuskan sebagai *quadratic programming problem* yaitu mencari titik minimal persamaan 5 dengan memperhatikan constraint persamaan 6.

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \quad (5)$$

$$y_i(\vec{w} \cdot \vec{x}_i) - 1 \geq 0, \forall i \quad (6)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi di antaranya *Langrange Multiplier*, seperti ditunjukkan pada Persamaan (7).

$$L(\vec{w}, b, \alpha) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^l \alpha_i (y_i (\vec{w} \cdot \vec{x}_i + b) - 1), i = 1, 2, \dots, l \quad (7)$$

α_i adalah Langrange multipliers, yang bernilai nol atau positif $\alpha_i \geq 0$. Nilai optimal dari persamaan (8) dapat dihitung dengan meminimalkan L terhadap \vec{w} dan b dan memaksimalkan L terhadap α_i . Berdasarkan sifat bahwa pada titik optimal $L = 0$, persamaan (9) dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung α_i .

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j \quad (8)$$

dari hasil perhitungan diperoleh α_i yang kebanyakan bernilai positif. Data yang berkorelasi dengan α_i yang positif inilah yang disebut *support vector*.

$$\alpha_i \geq 0 (i = 1, 2, \dots, l) \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (9)$$

3. HASIL DAN PEMBAHASAN

Proses klasifikasi ekonomi penduduk dilakukan dengan cara mengolah data parameter ekonomi penduduk di Kecamatan Simpang Raya Sulawesi Tengah dengan parameter meliputi pendidikan terakhir, dinding rumah, atap rumah, luas rumah dan sumber listrik. Data yang diperoleh dibagi menjadi dua bagian yaitu data latih dan data uji dengan persentase 80% data latih dan 20% data uji. Tiga ratus dua puluh data dilatih dengan menggunakan metode SVM, dan diperoleh sebuah model SVM yang mampu mengklasifikasi masyarakat penduduk Kecamatan Simpang Raya sebagai penerima atau bukan penerima bantuan pemerintah. Model SVM yang telah diperoleh selanjutnya diujikan pada 80 data uji. Proses pengujian ini dilakukan untuk mengetahui akurasi model klasifikasi dengan metode SVM, seberapa akurat metode SVM mampu mengklasifikasi penduduk yang berhak menerima bantuan atau tidak berhak menerima bantuan. Dalam proses pelatihan, terdapat beberapa variabel yang mempengaruhi formula seperti nilai *alpha*, bias, dan *weight*. Nilai ini memiliki keterkaitan yang sangat erat karena untuk menentukan nilai bias dibutuhkan nilai *weight* nilai *alpha* sedangkan nilai *alpha* sangat berkaitan dengan nilai yang terkandung dalam data. Dari proses ini tercipta formula yang memiliki tingkat klasifikasi yang cukup baik untuk digunakan dalam pengujian data. Hasil pengujian dari 80 data uji tidak dapat mencapai 100%. Hal ini terjadi karena salah satu parameter yang digunakan memiliki nilai yang tinggi. Misalnya sebuah keluarga memiliki rumah yang terdiri dari dinding rumah menggunakan kayu, atap rumah menggunakan seng, lantai rumah menggunakan semen dan luas rumah 68m². Pada umumnya, penduduk yang memiliki dinding kayu, atap seng dan lantai semen diklasifikasikan sebagai penduduk yang dapat menerima bantuan. Namun, karena memiliki tanah yang luas sehingga penduduk ini terklasifikasi sebagai penduduk yang tidak dapat menerima bantuan dan begitu juga sebaliknya. Pengujian ini dilakukan untuk mendapatkan nilai *error rate* dan *accuracy*. Nilai akurasi diperoleh melalui tabel *confusion matrices*, seperti yang

ditunjukkan pada Tabel 1. Persamaan (10) dan persamaan (11) digunakan untuk menghitung nilai akurasi dan nilai *error rate*. Hasil pengujian dapat dilihat pada Tabel 2

Tabel 1 Tabel Confusion Matrix.

Observasi	Klasifikasi		
		Penerima	Tidak Penerima
	Penerima	$T_{Penerima}$	$F_{Penerima}$
Tidak Penerima	$F_{Tidak_Penerima}$	$T_{Tidak_Penerima}$	

$$Akurasi = \frac{T_{Penerima} + T_{Tidak_Penerima}}{Penerima + Tidak_Penerima} \times 100\% \quad (10)$$

$$Error\ Rate = 100\% - Akurasi \quad (11)$$

Tabel 2 Hasil Pengujian dengan Confusion matrix

Observasi	Klasifikasi		
		Penerima	Tidak Penerima
	Penerima	39	1
Tidak Penerima	0	40	

$$Akurasi = \frac{39 + 40}{40 + 40} \times 100\% = 98\%$$

$$Error\ Rate = 100\% - 98\% = 2\%$$

4. KESIMPULAN

Berdasarkan pengujian dan analisis rancang bangun Sistem klasifikasi ekonomi penduduk Menggunakan Algoritma *Support Vector Machine* (SVM), maka dapat diambil kesimpulan sebagai berikut:

1. Sistem ini mampu mengklasifikasikan data penduduk yang terdapat pada Pemerintah menggunakan algoritma *Support Vector Machine* (SVM).
2. Hasil Pengujian *Confusion Matrix* menjelaskan bahwa sistem ini memiliki rata-rata akurasi sebesar 98% yang berarti bahwa sistem ini dapat mengklasifikasikan data penduduk dengan baik.

5. SARAN

Dari hasil penelitian ini masih terdapat beberapa kekurangan, sehingga masih diperlukan perbaikan untuk pengembangan lebih lanjut, sebagai berikut:

1. Penelitian selanjutnya yang membahas klasifikasi data penduduk pengembangan sistem dapat berbasis *website* atau Android.
2. Penelitian selanjutnya dapat menggunakan algoritma lainnya seperti Naïve Bayes, K-NN dan algoritma lainnya.

DAFTAR PUSTAKA

- [1] E. S. Indrawati, "Status Sosial Ekonomi Dan Intensitas Komunikasi Keluarga Pada Ibu Rumah Tangga Di Pangung Kidul Semarang Utara," *J. Psikol. Undip*, vol. 14, no. 1, pp. 52–57, 2015.
- [2] D. I. Permatasari, "Klasifikasi Status Ekonomi Keluarga Dengan Menggunakan Algoritma K-Nearest Neighbor Di Desa Pacewetan Kecamatan Pace Kabupaten Nganjuk," *Simki-Techsin*, vol. 01, no. 01, pp. 1–7, 2017.
- [3] E. Karyadiputra, "Analisis Algoritma Naive Bayes Untuk Klasifikasi Status Kesejahteraan

- Rumah Tangga Keluarga Binaan Sosial,” *Technologia*, vol. 7, no. 4, pp. 199–208, 2016.
- [4] S. Mutro N, A. Izzah, A. Kurniawardhani, and M. Masrur, “Optimasi Teknik Klasifikasi Modified K Nearest Neighbor Menggunakan Algoritma Genetika,” *J. Gamma*, vol. 10, no. 1, pp. 130–134, 2014.
- [5] Nuraedah, M. Bakri, and A. A. Kasim, “Quadratic support vector machine for the bomba traditional textile motif classification,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 11, no. 3, pp. 1004–1014, 2018.
- [6] A. S. Nugroho, A. B. Witarto, and D. Handoko, “Support Vector Machine ‘Teori dan Aplikasinya dalam Bioinformatika,”” in *Proceeding of Indonesian Scientific Meeting in Central Japan, December 20, 2003*, 2003.