

PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK PREDIKSI WAKTU TUNGGU ALUMNI MENDAPATKAN PEKERJAAN

I Made Budi Adnyana

Program Studi Sistem Informasi
Institut Teknologi Dan Bisnis STIKOM Bali
budi.adnyana@stikom-bali.ac.id

ABSTRACT

The application of the website-based Tracer Study system at ITB STIKOM Bali is useful for tracking and getting feedback from alumni. The waiting time for alumni to get a job is one of the assessments of the Tracer Study system at STIKOM Bali. The data obtained from this assessment has not been optimally processed by STIKOM Bali. If you dig deeper, you can get a pattern to predict the waiting time for alumni to get a job. In this study, we tried to apply the Support Vector Machine (SVM) algorithm in classifying alumni work waiting time. SVM is a method in supervised learning which is usually used for classification (such as Support Vector Classification) and regression (Support Vector Regression). In classification modeling, SVM has a more mature and clearer concept mathematically compared to other classification techniques. Sample data were obtained from the STIKOM Bali Tracer Study system, with a total of 1240 data consisting of 7 categorical type attributes. The trial of this SVM algorithm model uses WEKA tools with 10 folds cross-validation. The test results show an accuracy rate of 48.629%, where out of 1240 test data, 603 data were successfully classified correctly

Keywords: tracer study, alumni, svm, waiting time for work.

ABSTRAK

Penerapan sistem Tracer Study berbasis website di ITB STIKOM Bali berguna untuk melakukan pelacakan dan mendapatkan umpan balik dari alumni. Waktu tunggu alumni dalam memperoleh pekerjaan merupakan salah satu penilaian dari sistem Tracer Study di STIKOM Bali. Data yang diperoleh dari penilaian ini belum diolah secara optimal oleh pihak STIKOM Bali. Jika digali lebih mendalam, maka bisa didapatkan suatu pola untuk memprediksi waktu tunggu alumni dalam memperoleh pekerjaan. Pada penelitian ini mencoba menerapkan algoritma Support Vector Machine (SVM) dalam melakukan proses klasifikasi waktu tunggu kerja alumni. SVM merupakan salah satu metode dalam *supervised learning* yang biasanya digunakan untuk klasifikasi (seperti Support Vector Classification) dan regresi (Support Vector Regression). Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matang dan lebih jelas secara matematis dibandingkan dengan teknik-teknik klasifikasi lainnya. Data sampel diperoleh dari sistem Tracer Study STIKOM Bali, dengan jumlah data 1240 terdiri dari 7 atribut bertipe kategorikal. Uji coba model algoritma SVM ini menggunakan tools WEKA dengan 10 folds cross-validation. Hasil uji coba menunjukkan tingkat akurasi yang dihasilkan sebesar 48.629%, dimana dari 1240 data uji, 603 data berhasil diklasifikasi dengan benar

Kata Kunci : tracer study, alumni, svm, waktu tunggu kerja.

PENDAHULUAN

Tujuan utama dari perguruan tinggi adalah pengembangan sumber daya manusia yang berkualitas, mampu beradaptasi dengan perkembangan teknologi dan siap menghadapi tantangan dalam dunia industri. Setiap perguruan tinggi memiliki strategi dan sistem tersendiri untuk membentuk lulusan yang berkualitas. Kualitas lulusan merupakan salah satu fokus utama dalam Indikator Kinerja Utama Perguruan Tinggi (IKU PTN) yang diatur dalam Keputusan Menteri Pendidikan dan Kebudayaan Nomor 754/P/2020 [1]. Terserapnya lulusan di dunia kerja dan mendapatkan pekerjaan yang layak merupakan salah satu indikator dari keberhasilan perguruan tinggi. Menurut Direktorat Jenderal Pembelajaran dan Kemahasiswaan [2] perguruan tinggi agar melakukan Tracer Study secara reguler setiap tahun untuk memenuhi kebutuhan data akreditasi, pengembangan kurikulum dan perbaikan pembelajaran di perguruan tinggi. Tracer study dinilai penting karena menjadi alat evaluasi kinerja perguruan tinggi dan sekarang telah dijadikan salah satu syarat kelengkapan akreditasi oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT)

Salah satu tolok ukur kualitas lulusan dalam tracer study adalah waktu tunggu mendapatkan pekerjaan. Tidak jarang terdapat lulusan yang sampai menunggu bertahun-tahun untuk mendapatkan pekerjaan, namun ada juga yang mendapatkan pekerjaan dengan begitu cepat setelah kelulusan. Terdapat beberapa faktor yang dapat mempengaruhi lama lulusan dalam mendapatkan pekerjaan, seperti Indeks Prestasi Kumulatif (IPK) dan kemampuan soft skills yang dimiliki oleh alumni. Faktor internal lain yang juga dapat mempengaruhi adalah lama studi, pengalaman selama kuliah, pengetahuan, cara yang ditempuh dalam mendapatkan pekerjaan dan sebagainya [3]. Hal ini dapat berpengaruh terhadap penilaian untuk sebuah perguruan

tinggi dalam membentuk lulusan yang berkualitas sehingga mudah dalam memperoleh pekerjaan. Dengan memanfaatkan data yang terdapat pada sistem tracer study di ITB STIKOM Bali, dapat diterapkan teknik data mining untuk melakukan prediksi waktu tunggu alumni mendapatkan pekerjaan. Output dari model ini dapat digunakan untuk membantu manajemen perguruan tinggi dalam membuat kebijakan dan strategi dalam meningkatkan kualitas lulusan.

Salah satu teknik data mining yang dapat digunakan untuk melakukan prediksi adalah teknik klasifikasi. Klasifikasi merupakan metode yang melakukan pengkategorian item data kedalam label kelas yang sudah ditetapkan, membangun model klasifikasi dari kumpulan data input, serta dapat digunakan untuk memprediksi tren di masa depan [4]. Metode klasifikasi yang populer digunakan antara lain bayesian networks, neural networks, Support Vector Machines (SVM), decision trees, dan instance based. Penelitian yang dilakukan oleh Amrinda [5] menerapkan algoritma SVM dan Naïve Bayes untuk melakukan klasifikasi waktu tunggu kerja. Hasil penelitiannya menunjukkan algoritma SVM menghasilkan tingkat akurasi yang lebih baik daripada algoritma Naïve Bayes. Algoritma SVM juga telah digunakan untuk melakukan klasifikasi di berbagai studi kasus permasalahan seperti prediksi kelulusan mahasiswa [6], prediksi kinerja mahasiswa [7], dan klasifikasi data akreditasi sekolah dasar [8]. Pada penelitian ini menerapkan algoritma SVM untuk melakukan prediksi waktu tunggu kerja khususnya untuk alumni ITB STIKOM Bali diharapkan mampu memperoleh tingkat akurasi yang baik dengan metode yang digunakan dan outputnya dapat dijadikan sebagai acuan manajemen perguruan tinggi dalam meningkatkan kualitas lulusan

kedepannya.

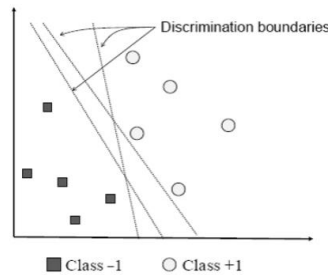
TINJAUAN PUSTAKA

Teknik Klasifikasi Data Mining

Klasifikasi merupakan bagian dalam machine learning dimana klasifikasi merupakan sebuah proses pembelajaran dari sebuah f sebagai fungsi target yang mampu memetakan tiap himpunan atribut x ke dalam salah satu label class dependen yang sudah didefinisikan sebelumnya [9]. Tahapan dalam klasifikasi antara lain adalah training dan testing. Proses training merupakan proses pembelajaran menggunakan data training. Proses testing merupakan proses melakukan pengujian terhadap model yang diperoleh pada proses training dengan menggunakan data testing.

Algoritma Support Vector Machine

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 teori-teori komputasi yang telah ada puluhan tahun sebelumnya, seperti margin hyperplane, kernel, dan juga dengan konsep-konsep pendukung yang lainnya.



Gambar 1. SVM menemukan titik hyperplane terbaik

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah class pada input space. Gambar 1 memperlihatkan beberapa pattern yang merupakan anggota dari dua buah class : +1 dan -1. Pattern yang tergabung pada class -1 disimbolkan dengan warna hitam (kotak),

sedangkan pattern pada class +1, disimbolkan dengan warna putih (lingkaran). Problem klasifikasi dapat diterjemahkan dengan usaha menemukan garis (hyperplane) yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (discrimination boundaries) ditunjukkan pada Gambar 1. Data yang tersedia dinotasikan sebagai $\vec{x}_i \in \mathfrak{R}_d$ sedangkan label masing-masing dinotasikan $y_i \in \{-1, +1\}$ untuk $i = 1, 2, \dots, l$, yang mana l adalah banyaknya data. Diasumsikan kedua class -1 dan +1 dapat terpisah secara sempurna oleh hyperplane berdimensi d , yang didefinisikan

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

w = Bidang normal

b = Posisi bidang relatif terhadap pusat koordinat

Pattern \vec{x}_i yang termasuk class -1 (sampel negatif) dapat dirumuskan sebagai pattern yang memenuhi pertidaksamaan

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

Sedang \vec{x}_i yang termasuk class + 1 (sampel positif)

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

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara hyperplane dan titik terdekatnya, yaitu $1/w$. Hal ini dapat dirumuskan sebagai Quadratic Programming (QP) problem, yaitu mencari titik minimal persamaan (4), dengan memperhatikan constraint persamaan (5).

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \tag{4}$$

$$y_i(\vec{w} \cdot \vec{x} + b) - 1 \geq 0, \forall_i \tag{5}$$

Masalah ini dapat dipecahkan dengan berbagai teknik komputasi, di antaranya Lagrange Multiplier

$$L(\vec{w}, b, \alpha) =$$

$$\frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^l \alpha_i (y_i(\vec{w} \cdot \vec{x} + b) - 1) \tag{6}$$

$(i = 1, 2, \dots, l)$

α_i adalah Lagrange multipliers, yang bernilai nol atau positif ($\alpha_i \geq 0$). Nilai optimal dari persamaan (6) dapat dihitung dengan meminimalkan L terhadap w dan

memaksimalkan L terhadap α_i . Dengan memperhatikan sifat bahwa pada titik optimal gradient $L=0$, persamaan (6) dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung saja α_i , sebagaimana persamaan (7) di bawah.

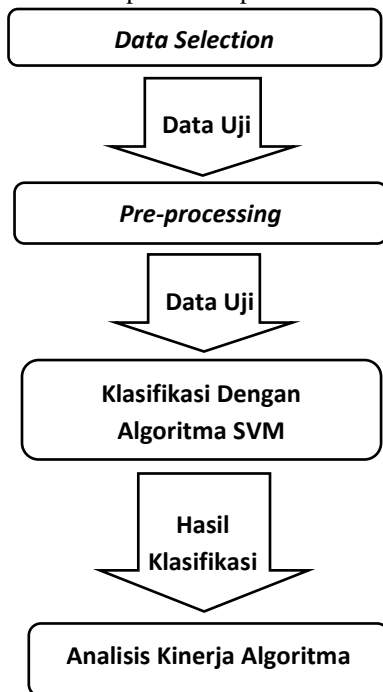
$$\text{Maximize:} \\ \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \vec{x}_i \cdot \vec{x}_j \quad (7)$$

$$\text{Subject to:} \\ \alpha_i \geq 0 (i = 1, 2, \dots, l) \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (8)$$

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

METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan dimulai dari data selection, data preprocessing, proses klasifikasi dengan algoritma SVM, dan terakhir analisis kinerja dari hasil uji coba klasifikasi algoritma SVM terhadap data waktu tunggu alumni dalam memperoleh pekerjaan. Gambaran umum alur penelitian ini dapat dilihat pada Gambar 2.



Gambar 2. Alur Penelitian

Data Selection adalah suatu proses seleksi data atau pemilihan himpunan data dari sekumpulan data. Kegiatan yang dilakukan pada tahap ini adalah pemilihan data alumni STIKOM Bali serta atribut-atributnya yang relevan terhadap proses analisis yang akan dikerjakan berikutnya. Data hasil seleksi ini lebih lanjut akan digunakan sebagai data uji, disimpan terpisah dengan database operasional STIKOM Bali.

Preprocessing bertujuan untuk pembersihan data, integrasi data, dan transformasi data agar data alumni yang akan digunakan pada proses klasifikasi lebih tepat dan akurat. Tahap preprocessing ini dilakukan dengan beberapa langkah, seperti mengisi data yang hilang/kosong, mengurangi data noise yang mengandung kesalahan atau data yang outlier, dan menangani data yang tidak konsisten.

Tahapan klasifikasi menggunakan algoritma SVM dilakukan dengan tools aplikasi WEKA. Aplikasi ini merupakan salah satu aplikasi yang populer digunakan pada data mining karena kehandalannya. Hasil uji coba klasifikasi menggunakan aplikasi WEKA selanjutnya dianalisis tingkat akurasi dan kesalahannya. Pengukuran tingkat akurasi menggunakan pendekatan Correctly Classified Instances, sedangkan tingkat kesalahan menggunakan pendekatan RMSE.

HASIL DAN PEMBAHASAN

Dataset yang digunakan pada penelitian ini diperoleh dari sistem Tracer Study STIKOM Bali. Data yang digunakan adalah lulusan jenjang S1 lima tahun terakhir sebanyak 1240 data alumni. Data ini terdiri dari 7 atribut bertipe kategorikal, yaitu: Program Studi, Jenis Kelamin, Status Kelas, Konsentrasi, Lama Studi, IPK, dan Lama Memperoleh Pekerjaan. Output class dari model yang dikembangkan ini adalah berupa klasifikasi "Lama Memperoleh Pekerjaan", terdiri dari 4 kelas, yaitu: "0 – 2 Bulan", "3 – 6 Bulan", "7 – 12 Bulan", dan "Lebih Dari 12 Bulan". Format dataset alumni yang digunakan sebagai data uji ditunjukkan pada Tabel 1.

Tabel 1. Format Dataset Alumni

Nama Atribut	Type	Value
Program Studi	Categorical	<ul style="list-style-type: none"> Sistem Komputer Sistem Informasi
Jenis Kelamin	Categorical	<ul style="list-style-type: none"> Laki Perempuan
Status Kelas	Categorical	<ul style="list-style-type: none"> Reguler Karyawan
Konsentrasi	Categorical	<ul style="list-style-type: none"> NTCS RADE ES GM IS
Lama Studi	Categorical	<ul style="list-style-type: none"> Kurang dari 4 tahun 4 – 5 tahun Lebih dari 5 tahun
IPK	Categorical	<ul style="list-style-type: none"> IPK 2.50 – 2.99 IPK 3.00 – 3.49 IPK 3.50 – 4.00
Lama Memperoleh Pekerjaan	Categorical	<ul style="list-style-type: none"> 0 – 2 bulan 3 – 6 bulan 7 – 12 bulan Lebih dari 12 bulan

Agar data uji yang diperoleh ini dapat digunakan pada aplikasi WEKA, maka format data uji tersebut perlu dirubah kedalam bentuk ARFF (Attribute Relation File Format). Proses konversi format data ini dapat dilakukan pada aplikasi WEKA menggunakan fitur ArffViewer. Tampilan data sampel pada ArffViewer dapat dilihat pada Gambar 2. Visualisasi data sampel untuk setiap atributnya dapat dilihat pada Gambar 3.

Gambar 3. Data sampel pada ArffViewer

Uji coba dilakukan dengan menggunakan tools WEKA. Pembagian data sampel menggunakan teknik 10 folds cross-validation, dimana data sampel dibagi menjadi beberapa kombinasi dengan komposisi 9/10 sebagai data training dan 1/10 sebagai testing. Hasil uji coba klasifikasi SVM dengan menggunakan aplikasi WEKA dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pengukuran Akurasi

Pengukuran	Hasil
Correctly Classified Instances	812 (65.48 %)
Incorrectly Classified Instances	428 (34.52 %)
RMSE	0.4514
Total Number of Instances	1240

Dari hasil uji coba diatas, berdasarkan pada nilai Correctly Classified Instances menunjukkan bahwa tingkat akurasi yang dihasilkan adalah 65.48%, Sedangkan nilai Incorrectly Classified Instances adalah 34.52%. Hal ini menunjukkan dari 1240 data uji, 812 data berhasil diklasifikasi dengan benar, sedangkan 428 data diklasifikasi dengan tidak benar. Tingkat kesalahan (error) dapat dilihat dari pendekatan RMSE (Root Mean Squared Error) bernilai 0.4514. Dalam pengukuran ini model prediksi dikatakan paling baik apabila nilai RMSE adalah 0 (nol).

SIMPULAN

Berdasarkan hasil penelitian ini dapat disimpulkan model algoritma SVM yang digunakan berhasil melakukan prediksi/klasifikasi data waktu tunggu kerja alumni di ITB STIKOM Bali. Berdasarkan

hasil uji coba menggunakan tools WEKA menunjukkan tingkat akurasi yang dihasilkan adalah sebesar 65.48%, dari 1240 data uji hanya 812 data yang berhasil diklasifikasi dengan benar. Sedangkan tingkat kesalahan ditunjukkan dengan RMSE (Mean Absolute Error) yang bernilai 0.4514.

DAFTAR PUSTAKA

- [1] PDDikti, "Higher Education Statistics 2020," pp. 81–85, 2020, [Online]. Available: <https://pddikti.kemdikbud.go.id/publikasi>.
- [2] Ristekdikti, *Undang-undang Republik Indonesia Tahun Nomor 12 Tahun 2012 tentang Pendidikan Tinggi*. Indonesia: Ristekdikti, 2012.
- [3] Pretty Kurniawati, "Analisis Faktor-faktor yang Mempengaruhi Lama Mencari Kerja Lulusan S1 Fakultas Ekonomi Universitas Andalas Periode Wisuda Tahun 2013-2016," Universitas Andalas, Padang, 2017.
- [4] Yulia Rizki Amalia, "Penerapan Data Mining Untuk Prediksi Penjualan Produk elektronik Terlaris Menggunakan Metode K-Nearest Neighbor (Studi Kasus PT Bintang Multi Sarana Palembang)," Universitas Islam Negeri Raden Fatah, Palembang, 2018.
- [5] Gustiara Dayu Amrinda, "Analisis Klasifikasi Waktu Tunggu Kerja Dengan Metode Support Vector Machine Dan Naïve Bayes Classification (Studi Kasus: Waktu tunggu kerja Alumni Universitas Islam Indonesia)," Universitas Islam Indonesia, Yogyakarta, 2018.
- [6] Suhardjono, W. Ganda, and H. Abdul, "Prediksi Waktu Kelulusan Mahasiswa Menggunakan Svm Berbasis Pso," *Bianglala Inform.*, vol. 7, no. 2, pp. 97–101, 2019.
- [7] F. Hilmiyah, "Prediksi Kinerja Mahasiswa Menggunakan Support Vector Machine untuk Pengelola Program Studi di Perguruan Tinggi (Studi Kasus: Program Studi Magister Statistika ITS)," *Dep. Manaj. Teknol. Bid. Keahlian Manaj. Teknol. Inf. Fak. Bisnis Dan Manaj. Teknol. Inst. Teknol. Sepuluh Nop. Surabaya*, pp. 1–99, 2017, [Online]. Available: <http://repository.its.ac.id/46712/>.
- [8] D. I. Pushpita Anna Octaviani, Yuciana Wilandari, "Penerapan Metode SVM Pada Data Akreditasi Sekolah Dasar Di Kabupaten Magelang," *J. Gaussian*, vol. 3, no. 8, pp. 811–820, 2014.
- [9] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000, doi: 10.1093/bioinformatics/16.10.906.

