

PENGGOLONGAN UANG KULIAH TUNGGAL MENGGUNAKAN *SUPPORT VECTOR MACHINE*

I Gede Seka Suyoga^{1§}, I Putu Eka Nila Kencana², I Komang Gde Sukarsa³

¹Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: seka.suyoga@outlook.com]

²Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: i.putu.enk@unud.ac.id]

³Program Studi Matematika, Fakultas MIPA – Universitas Udayana [Email: sukarsakomang@yahoo.com]

[§]Corresponding Author

ABSTRACT

Tuition fee is the payment of tuition fees each semester borne by each student based on their economic capabilities. Tuition fee is divided into five groups from tuition fee group 1 to tuition fee group 5. This research aims to find the accuracy of the classification of tuition fee using Support Vector Machine (SVM). SVM is a method used for classification of the concept to find hyperplane (separator function) that can separate the data into a predetermined class. In this research, SVM is used to determine the accuracy of tuition fee classification. The variables used are income parents, father's occupation, mother's occupation, home ownership status, building, land area, electricity cost, water cost, phone cost, saving accounts, jewelry ownership, and a premium ownership. The results obtained are five hyperplanes to separate tuition fee with accuracy of the classification of tuition fee was 59,69%.

Keywords: Tuition Fee, Classification, Support Vector Machine

1. PENDAHULUAN

Dunia pendidikan tinggi di Indonesia seringkali menjadi sorotan masyarakat karena adanya persaingan antar perguruan tinggi negeri untuk menjadi yang terbaik, salah satu efek sampingnya menyebabkan biaya kuliah meningkat secara perlahan dari tahun ke tahun. Hal ini menimbulkan pandangan masyarakat bahwa hanya kalangan atas (mampu secara ekonomi) yang bisa memperoleh pendidikan lebih tinggi, sedangkan kalangan bawah (kurang mampu) hanya memperoleh pendidikan seadanya. Oleh karena itu perguruan tinggi negeri harus lebih bijak dalam menentukan biaya kuliah per semester karena pendidikan juga menjadi tolak ukur dalam mendapatkan pekerjaan dimana semakin meningkatnya syarat pendidikan yang harus dimiliki oleh calon karyawan pada suatu perusahaan atau instansi. Tentu hal ini menyebabkan adanya dorongan minat masyarakat untuk mengenyam pendidikan setinggi-tingginya.

Sistem Uang Kuliah Tunggal (UKT) merupakan salah satu terobosan pemerintah Indonesia selain beasiswa bidik misi dalam penentuan biaya kuliah. Pemerintah Indonesia melalui Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia menetapkan UKT adalah sebagian biaya kuliah tunggal yang ditanggung setiap mahasiswa berdasarkan kemampuannya, dimana biaya kuliah tunggal yang dimaksud adalah keseluruhan biaya operasional yang terkait langsung dengan proses pembelajaran mahasiswa per semester pada program studi di Perguruan Tinggi Negeri (PTN).

Salah satu perguruan tinggi negeri yang ada di Indonesia dan telah menggunakan sistem UKT adalah Universitas Udayana (UNUD). Sistem pembayaran UKT telah diberlakukan di Universitas Udayana sejak tahun ajaran 2013/2014. Kelompok UKT yang diberlakukan oleh Universitas Udayana terdiri dari lima kelompok mulai dari yang terendah hingga tertinggi yaitu kelompok UKT 1, UKT 2, UKT

3, UKT 4, dan UKT 5. Penentuan UKT di UNUD dilakukan dengan menggolongkan data mahasiswa agar biaya kuliah yang dikenakan kepada mahasiswa lebih sesuai dengan kemampuan perekonomiannya. Adapun metode yang dapat diterapkan dalam penggolongkan (klasifikasi) adalah *support vector machine* (SVM). SVM bertujuan menemukan *hyperplane* terbaik untuk memisahkan dua set data dari dua kelas berbeda (Vapnik, 1995). Penelitian dengan menggunakan metode SVM telah dilakukan oleh Vanitha, *et al* (2015) yang meneliti tentang klasifikasi ekspresi gen menggunakan metode SVM pada data kanker, hasil yang diperoleh yakni SVM memiliki rata-rata akurasi tertinggi sebesar 0,9777 dibandingkan dengan metode klasifikasi lainnya. Penelitian lainnya dilakukan oleh Huang, *et al* (2005) yang menyelidiki prediktabilitas arah pergerakan keuangan, hasil penelitian menunjukkan bahwa SVM lebih unggul dari metode klasifikasi lainnya dengan akurasi sebesar 73%. Berdasarkan uraian tersebut maka penelitian ini bertujuan untuk mengetahui *hyperplane* UKT dan akurasi ketepatan penggolongkan UKT menggunakan SVM.

Ide dasar penggunaan SVM berawal dari masalah klasifikasi dua kelas yang dipisahkan secara linier (*linearly seperable*) ke dalam ruang fitur (Bishop, 2006). Diberikan masalah klasifikasi dua kelas (*binary classification*) dengan himpunan data *training* $G = \{(x_j, y_j), j = 1, 2, \dots, N\}$, $y_j \in \{-1, 1\}$ diasumsikan terpisah secara linier. *Hyperplane* adalah suatu fungsi bersifat linier yang dapat memecahkan masalah tersebut, diberikan sebagai berikut:

$$f(x) = w^T x + b \quad (1)$$

Diasumsikan himpunan data *training* dapat dipisahkan secara linier (*linearly seperable*) ke dalam ruang fitur sedemikian sehingga fungsi dari persamaan (1) memenuhi:

$$w^T x_j + b \geq 1, \text{ untuk } y_j = 1 \quad (2)$$

$$w^T x_j + b \leq -1, \text{ untuk } y_j = -1 \quad (3)$$

dimana $w \in R^n$ adalah vektor yang menjadi parameter bobot dan dimensinya bergantung dari n banyaknya fitur pemetaan ruang agar data

terpisah secara linier, $x_j \in R^n$ merupakan input vektor ke- j , $b \in R$ merupakan parameter bias, dan $y_j \in \{-1, 1\}$ merupakan nilai target atau kelas dari x_j . *Hyperplane* terbaik mempunyai *margin* yang optimal, maka untuk memaksimalkan *margin* dapat dilakukan dengan mengoptimalkan parameter w dan b . Pengoptimalan tersebut didapat dengan memecahkan:

$$\max_{w,b} \left\{ \frac{1}{\|w\|} \min_j [y_j (w^T x_j + b)] \right\} \quad (4)$$

Diambil faktor $\frac{1}{\|w\|}$ diluar optimasi daripada j karena w tidak bergantung pada j . memaksimalkan $\frac{1}{\|w\|}$ yang ekuivalen dengan meminimalkan $\|w\|^2$ sehingga diperoleh bentuk dari masalah optimasi tersebut adalah:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (5)$$

Asumsi data terpisah secara linier terkadang tidak dapat terpenuhi atau bisa dikatakan *linearly non-seperable*. *Soft margin* digunakan untuk menyelesaikan masalah tersebut dengan memodifikasi persamaan (5) sehingga memungkinkan beberapa data *training* tidak terklasifikasi dengan benar. Untuk merealisasikan *soft margin* ini maka diperkenalkan variabel *slack*, $\xi_j \geq 0$ dimana $j = 1, 2, \dots, N$. Berdasarkan bentuk dari masalah optimasi persamaan (5) maka bentuk masalah optimasi dari *soft margin* adalah

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{j=1}^N \xi_j \quad (6)$$

dengan syarat

$$y_j (w^T x_j + b) \geq 1 - \xi_j, \xi_j \geq 0, j = 1, 2, \dots, N$$

dimana parameter $C > 0$ mengontrol *trade-off* antara penalti variabel *slack* dan *margin*. Penyelesaian masalah optimasi persamaan (6) dapat dilakukan dengan mencari bentuk dual dari masalah tersebut menggunakan pengali Lagrange dan menghitung turunan parsialnya terhadap w, b dan ξ_j sehingga diperoleh bentuk dual dari masalah optimasi dari *soft margin*:

$$\begin{aligned} \max_a L(a) \\ = \sum_{j=1}^N \alpha_j - \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N \alpha_j \alpha_k y_j y_k x_j^T x_k \end{aligned} \quad (7)$$

dimana $\alpha \geq 0$ merupakan pengali Lagrange. Bentuk dual yang dihasilkan merupakan masalah pemrograman kuadrat (*quadratic programming*). Misalkan solusi masalah pemrograman kuadrat pada persamaan (7) adalah α^* , maka hanya data *training* dengan $\alpha^* > 0$ yang berperan dalam pembentukan *hyperplane* atau disebut sebagai *support vector*. Sehingga untuk suatu data baru x , diberikan *hyperplane* sebagai berikut:

$$f(x) = \sum_{j \in S} \alpha_j^* y_j x_j x + b \quad (8)$$

dimana S adalah indeks dari *support vector*.

SVM pada dasarnya dirancang untuk masalah klasifikasi dua kelas (*binary classification*). Dalam prakteknya masalah klasifikasi *multi-class* ($k > 2$) biasanya didekomposisi menjadi serangkaian masalah biner sehingga standar SVM dapat diterapkan secara langsung. Salah satu skema yang digunakan pada *multi-class* SVM yaitu *one against all* yang mengkonstruksi k penggolong terpisah untuk klasifikasi sebanyak k kelas. Suatu fungsi penggolong ke- m dilatih menggunakan kelas ke- m sebagai sampel positif dan kelas $k - 1$ sisanya sebagai sampel negatif. Selama pengujian, label kelas ditentukan oleh penggolong yang memberikan nilai output maksimal (Wang & Xue, 2014).

Pada prakteknya, ketika data *linearly non-seperable* maka data dapat dipetakan ke dalam ruang fitur berdimensi tinggi yang didefinisikan dengan fungsi kernel (Schölkopf & Smola, 2002). SVM memiliki *hyperplane* dengan *margin* maksimal pada ruang fitur berdimensi tinggi tersebut dan ada empat kernel dasar yang dapat ditemukan yaitu kernel *linier*, kernel *polynomial*, kernel *gaussian radial basis function* (RBF), dan kernel *sigmoid* (Hsu, et al., 2003). Pada penelitian ini digunakan fungsi kernel RBF yang diberikan sebagai berikut:

$$K(x_j, x_k) = \exp(-\gamma \|x_j - x_k\|^2), \gamma > 0 \quad (9)$$

dimana γ adalah parameter kernel, maka persamaan (8) dapat ditulis:

$$f(x) = \sum_{j \in S} \alpha_j^* y_j K(x_j, x) + b \quad (10)$$

2. METODE PENELITIAN

A. Jenis Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data kelompok UKT di Universitas Udayana untuk mahasiswa S1 tahun 2016. Data yang diperoleh dipartisi secara acak menjadi dua yaitu 70% sebagai data *training* dan 30% sebagai data *testing* menggunakan metode *holdout* dengan pedekatan *random subsampling* karena jumlah keseluruhan data 4954 maka penggunaan data *training* sebanyak 3468 dan data *testing* sebanyak 1486.

Variabel yang digunakan pada penelitian adalah $Y =$ golongan UKT, $Z_1 =$ penghasilan orang tua, $Z_2 =$ pekerjaan ayah, $Z_3 =$ pekerjaan ibu, $Z_4 =$ status kepemilikan rumah, $Z_5 =$ luas bangunan, $Z_6 =$ luas tanah, $Z_7 =$ biaya rekening air, $Z_8 =$ biaya rekening listrik, $Z_9 =$ biaya rekening telepon, $Z_{10} =$ kepemilikan tabungan, $Z_{11} =$ kepemilikan perhiasan, dan $Z_{12} =$ kepemilikan premi.

B. Teknik Analisis Data

Pada penelitian ini analisis SVM menggunakan bantuan software MATLAB R2017a. Adapun langkah-langkah penelitian sebagai berikut:

1. Melakukan transformasi data pada variabel bebas yang berskala ordinal menggunakan *method of successive interval* (MSI).
2. Melakukan proses analisis data *training* dan data *testing* dengan *multi-class* SVM *one against all* untuk mendapatkan *hyperplane* UKT dengan langkah-langkah berikut:
 - i. Membagi data menjadi data *training* dan data *testing* secara acak.
 - ii. Menentukan nilai parameter C dan parameter kernel yang digunakan sebagai pencarian *hyperplane* SVM.

- iii. Mendapatkan nilai α dan b melalui *quadratic programming*.
 - iv. Menghitung akurasi penggolongan data *training* menggunakan *confusion matrix*.
 - v. Melakukan penggolongan pada data *testing*.
 - vi. Menghitung akurasi penggolongan data *testing* menggunakan *confusion matrix*.
 - vii. Ulangi langkah (i) sampai (vi) sebanyak 10 kali.
3. Memilih parameter-parameter SVM dengan akurasi terbesar pada data *testing*.
 4. Membentuk lima *hyperplane* menggunakan persamaan (10).

3. HASIL DAN PEMBAHASAN

Pada penelitian ini menggunakan analisis *multi-class SVM* dengan pendekatan *one againts all* karena terdapat lima kelompok UKT yaitu kelompok UKT 1, kelompok UKT 2, kelompok UKT 3, kelompok UKT 4, dan kelompok UKT 5. Rincian data yaitu UKT 1 berjumlah 315, UKT 2 berjumlah 311, UKT 3 berjumlah 2894, UKT 4 berjumlah 792, dan UKT 5 berjumlah 642. Proporsi pengacakan dari 3468 data *training* yang digunakan sebagai berikut:

- UKT 1 = $315 \times 0,7 = 221$
- UKT 2 = $311 \times 0,7 = 218$
- UKT 3 = $2894 \times 0,7 = 2026$
- UKT 4 = $792 \times 0,7 = 554$
- UKT 5 = $642 \times 0,7 = 449$

Sedangkan proporsi pengacakan dari 1486 data *testing* yang digunakan sebagai berikut:

- UKT 1 = $315 \times 0,3 = 94$
- UKT 2 = $311 \times 0,3 = 93$
- UKT 3 = $2894 \times 0,3 = 868$
- UKT 4 = $792 \times 0,3 = 238$
- UKT 5 = $642 \times 0,3 = 193$

Pada tahap *training* digunakan nilai parameter $C=1$, $C=5$, dan $C=10$ pada setiap pengulangan dan fungsi kernel *Gaussian RBF* dengan nilai parameter kernel yang diperoleh menggunakan bantuan aplikasi MATLAB R2017a karena menurut Wu & Wang (2009)

tidak ada metode teoritis untuk menentukan parameter fungsi kernel dan tidak ada pengetahuan umum untuk menetapkan parameter C . Tahap *training* menggunakan SVM menghasilkan nilai a dan nilai bias. Nilai-nilai tersebut menentukan bentuk *hyperplane* dan hasil klasifikasi menggunakan data *testing*. Proses perhitungan nilai a dilakukan melalui perhitungan *quadratic programming*. Setiap data *training* akan memiliki nilai a tersendiri.

Hyperplane terpilih hanya *hyperplane* yang memiliki akurasi penggolongan data *testing* terbesar dengan parameter yang tepat pada salah satu pengulangan. Akurasi pada data *testing* diperoleh melalui jumlah penggolongan yang tepat dibagi jumlah keseluruhan data *testing*. Tabel 1 menunjukkan hasil akurasi penggolongan terbesar data *testing* diperoleh pada pengulangan ke-7 dengan nilai parameter $C=1$ yaitu sebesar 59,69%. Tabel 1 juga menunjukkan bahwa hasil akurasi penggolongan terbesar pada data *training* belum tentu menghasilkan akurasi klasifikasi terbesar pada data *testing*.

Tabel 1. Hasil Training dan Testing SVM pada Pengulangan ke-7

One Againts All	Training			Testing
	γ	C	Akurasi	Akurasi
UKT 1 vs non-UKT 1	53185	1	73,50%	59,69%
UKT 2 vs non-UKT 2	74488			
UKT 3 vs non-UKT 3	78936			
UKT 4 vs non-UKT 4	167246			
UKT 5 vs non-UKT 5	143933			
UKT 1 vs non-UKT 1	57182	5	76,80%	58,08%
UKT 2 vs non-UKT 2	58096			
UKT 3 vs non-UKT 3	102302			
UKT 4 vs non-UKT 4	174172			
UKT 5 vs non-UKT 5	219696			
UKT 1 vs non-UKT 1	74398	10	79,50%	56,33%
UKT 2 vs non-UKT 2	66537			
UKT 3 vs non-UKT 3	88422			
UKT 4 vs non-UKT 4	139345			
UKT 5 vs non-UKT 5	201589			

Sumber: data diolah tahun 2017

SVM dapat menggolongkan secara tepat sebanyak 5 data pada UKT 1, 5 data pada UKT 2, 810 data pada UKT 3, 45 data pada UKT 4, dan 22 data pada UKT 5 yang dapat dilihat pada Tabel 2.

Tabel 2. Confusion Matrix Hasil Testing Pengulangan ke-7 dengan C=1

		Kelompok Prediksi					Total
		UKT 1	UKT 2	UKT 3	UKT 4	UKT 5	
Kelompok Asli	UKT 1	5	2	86	0	1	94
	UKT 2	1	5	87	0	0	93
	UKT 3	11	3	810	30	14	868
	UKT 4	0	1	180	45	12	238
	UKT 5	1	0	146	24	22	193
Total		18	11	1309	99	49	1486

Sumber: data diolah tahun 2017

Selanjutnya dibentuk *hyperplane* dengan menggunakan persamaan (10). *Hyperplane* yang dibentuk pada penelitian ini hanya *hyperplane* dari hasil tahap *training* pengulangan ke-7 dengan parameter C=1 dan jumlah *support vector* yaitu 2091 untuk kelompok UKT 1, 1692 untuk kelompok UKT 2, 2900 untuk kelompok UKT 3, 1569 untuk kelompok UKT 4, dan 1575 untuk kelompok UKT 5. Jadi suatu data baru (x) digolongkan ke dalam kelompok UKT 1, UKT 2, UKT 3, UKT 4, atau UKT 5 dilakukan dengan mencari nilai *hyperplane* yang maksimal. Jika nilai UKT dari $x \equiv f_{UKT k}(x)$, $k = 1, 2, 3, 4$, atau 5, maka data baru ke- x dengan 12 variabel tersebut akan terdistribusi ke dalam UKT k , dimana

(i) *Hyperplane* untuk UKT 1

$$f_{UKT 1}(x) = \sum_{j=1}^{2091} \alpha_{(1)j} y_{(1)j} \left(\exp \left(-53185 \|x_{(1)j} - x\|^2 \right) \right) + (-0,9763)$$

(ii) *Hyperplane* untuk UKT 2

$$f_{UKT 2}(x) = \sum_{j=1}^{1692} \alpha_{(2)j} y_{(2)j} \left(\exp \left(-74488 \|x_{(2)j} - x\|^2 \right) \right) + (-0,9829)$$

(iii) *Hyperplane* untuk UKT 3

$$f_{UKT 3}(x) = \sum_{j=1}^{2900} \alpha_{(3)j} y_{(3)j} \left(\exp \left(-78936 \|x_{(3)j} - x\|^2 \right) \right) + (-0,0622)$$

(iv) *Hyperplane* untuk UKT 4

$$f_{UKT 4}(x) = \sum_{j=1}^{1569} \alpha_{(4)j} y_{(4)j} \left(\exp \left(-167246 \|x_{(4)j} - x\|^2 \right) \right) + (-0,6669)$$

(v) *Hyperplane* untuk UKT 5

$$f_{UKT 5}(x) = \sum_{j=1}^{1575} \alpha_{(5)j} y_{(5)j} \left(\exp \left(-143933 \|x_{(5)j} - x\|^2 \right) \right) + (-0,6911)$$

dengan $x_{(i)j}$, $i = 1, 2, 3, 4, 5$ adalah *support vector* data ke- j pada UKT ke- i , $y_{(i)j}$, $i = 1, 2, 3, 4, 5$ adalah kelas *support vector* data ke- j pada UKT ke- i .

4. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil yang diperoleh maka di dapat diambil kesimpulan bahwa ada lima *hyperplane* yang terbentuk untuk menggolongkan UKT mahasiswa S1 Universitas Udayana tahun 2016 ke dalam UKT 1, UKT 2, UKT 3, UKT 4, atau UKT 5. SVM menghasilkan akurasi sebesar 59,69% dalam menggolongkan mahasiswa ke dalam kelompok UKT 1, UKT 2, UKT 3, UKT 4, atau UKT 5.

B. Saran

Oleh karena metode *multi-class SVM* yang digunakan pada penelitian ini adalah *one againts all*, maka disarankan untuk penelitian lebih lanjut dapat menggunakan *multi-class SVM one againts one*.

Pemetaan data input ke ruang yang berdimensi lebih tinggi pada penelitian ini hanya menggunakan fungsi kernel *gaussian radial*

basic function (RBF), maka disarankan untuk penelitian lebih lanjut dapat menggunakan fungsi kernel lainnya seperti fungsi kernel *linier*, fungsi kernel *polynomial*, atau fungsi kernel *sigmoid*.

Mengingat hasil penggolongan UKT pada penelitian ini kurang dari 60%, hal ini kemungkinan disebabkan data UKT yang digunakan tidak tergolong dengan benar. Oleh karena itu kepada pihak terkait agar lebih objektif dalam menggolongkan UKT dan untuk informasi pengisian data dilakukan oleh lembaga independen yang mengetahui kondisi dan keadaan ekonomi calon mahasiswa.

Pada penelitian lebih lanjut disarankan melakukan standarisasi data yang bertujuan untuk mempermudah analisis data dan interpretasi hasil.

DAFTAR PUSTAKA

- Bishop, C. M., 2006. *Pattern Recognition and Machine Learning*. Cambridge: Springer.
- Hsu, C.-W., Chang, C.-C. & Lin, C.-J., 2003. *A Practical Guide to Support Vector Classification*, Taipei: Departement of Computer Science National Taiwan University.
- Huang, W., Nakamori, Y. & Wang, S.-Y., 2005. Forecasting Stock Market Movement Direction with Support Vector Machine. *Computers & Operations Research*, Volume 32, pp. 2513-2522.
- Schölkopf, B. & Smola, A. J., 2002. *Learning with Kernels*. Cambridge: The MIT Press.
- Vanitha, D. A., Devaraj, D. & Venkatesulu, M., 2015. Gene Expression Data Classification using Support Vector Machine and Mutual Information-based Gene Selection. *Procedia Computer Science*, Volume 47, pp. 13-21.
- Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. New York: Springer.
- Wang, Z. & Xue, X., 2014. Multi-Class Support Vector Machine. In: *Support Vector Machine Applications*. Switzerland: Springer, pp. 23-48.
- Wu, K.-P. & Wang, S.-D., 2009. Choosing The Kernel Parameters for Support Vector Machine by The Inter-Cluster Distance in The Feature Space. *Pattern Recognition*, Issue 42, pp. 710-717.