

PERBANDINGAN KINERJA KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) DAN REGRESI LOGISTIK BINER DALAM MENGLASIFIKASIKAN KETEPATAN WAKTU KELULUSAN MAHASISWA FMIPA UNTAD

I. T. Utami¹

¹Program Studi Statistika, Jurusan Matematika

¹Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Tadulako

¹triotami.iut@gmail.com

ABSTRAK

Evaluasi kinerja klasifikasi dapat ditentukan berdasarkan persentase besarnya kesalahan klasifikasi (misclassification rate atau MCR). Penelitian ini bertujuan membandingkan kinerja klasifikasi ketepatan waktu kelulusan mahasiswa FMIPA UNTAD dengan menggunakan metode support vector machine dan regresi logistik biner. Hasil penelitian diperoleh bahwa kesalahan klasifikasi dengan menggunakan metode Support Vector Machine (SVM) dan regresi logistik biner masing-masing sebesar 16.84% dan 19.3%. Berdasarkan perbandingan kinerja kedua metode tersebut, metode dengan kesalahan klasifikasi terkecil adalah metode Support Vector Machine. Metode tersebut dapat digunakan untuk mengklasifikasikan ketepatan waktu kelulusan mahasiswa FMIPA UNTAD.

Kata Kunci : Support Vector Machine, Regresi Logistik Biner, Kesalahan Klasifikasi.

I. PENDAHULUAN

Perguruan tinggi dituntut untuk menghasilkan lulusan yang berkualitas, handal dan mempunyai keahlian dibidangnya. Dalam pencapaian gelar S1 di suatu perguruan tinggi diperlukan waktu normal untuk menyelesaikan kuliah sekitar empat tahun atau kurang. Pada prakteknya banyak mahasiswa yang menyelesaikan kuliah lebih dari empat tahun. Hal ini akan menyebabkan menurunnya penilaian kualitas suatu perguruan tinggi apabila kebanyakan mahasiswanya lulus tidak tepat waktu. Penelitian ini bertujuan untuk membandingkan kinerja klasifikasi dari ketepatan waktu kelulusan mahasiswa FMIPA UNTAD dengan menggunakan metode Support Vector Machine (SVM) dan regresi logistik biner. Metode yang memiliki kesalahan klasifikasi terkecil dapat digunakan untuk mengklasifikasikan ketepatan waktu kelulusan mahasiswa FMIPA UNTAD.

Metode regresi logistik biner sering digunakan dalam berbagai penelitian karena mudah diaplikasikan. Metode ini termasuk metode parametrik yang memerlukan asumsi diantaranya adalah tidak terdapat multikolinieritas antar peubah penjelas. Support Vector Machine termasuk dalam metode nonparametrik. Metode ini merupakan metode alternatif untuk mengatasi masalah pelanggaran asumsi tertentu dalam mengklasifikasikan data. Kelebihan metode ini dapat digunakan pada ukuran data yang besar dengan peubah penjelas yang banyak dan kekar terhadap pencilaan (Steinberg & Colla 1995). Evaluasi kinerja kedua metode dapat diukur dengan melihat persentase besarnya kesalahan klasifikasi (misclassification rate atau MCR). Pengujian dilakukan dengan matriks konfusi.

Fiastantyo (2014) pernah meneliti tentang perbandingan kinerja metode Naïve Bayes dan algoritma C4.5 untuk memprediksi ketepatan waktu kelulusan mahasiswa yang menghasilkan algoritma C4.5 mempunyai akurasi yang lebih baik daripada Naïve Bayes. Suwitno (2017) juga telah melakukan penelitian dengan judul "Perancangan Aplikasi Prediksi Kelulusan Mahasiswa Tepat Waktu pada Universitas Buddhi Dharma dengan Menggunakan Perbandingan Algoritma C4.5 dan K-NN" yang mendapatkan hasil algoritma C4.5 memiliki akurasi lebih tinggi dibanding K-NN. Priati (2016) telah melakukan kajian perbandingan teknik klasifikasi algoritma C4.5, Naïve bayes dan CART untuk melakukan prediksi kelulusan mahasiswa di STMIK Rosm Karawang yang menghasilkan algoritma C4.5 dan CART dengan nilai akurasi yang sama tinggi dibandingkan dengan metode Naïve Bayes.

II. METODE PENELITIAN

2.1. Support Vector Machine (SVM)

SVM adalah pencarian hyperplane terbaik yang berfungsi sebagai pemisah dua kelas data. Hyperplane terbaik antara kedua kelas terletak di tengah-tengah diantara dua bidang kendala kelas dan untuk mendapatkan hyperplane terbaik dilakukan dengan cara memaksimalkan jarak antara hyperplane dari titik terdekat (support vector) dari masing-masing kelas. Andaikan M adalah jarak tersebut, didefinisikan besaran margin sebagai $2M$.

Misalkan sebuah gugus data berisi n pasangan data pengamatan yang dinotasikan sebagai $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$ dengan $\mathbf{x}_i \in \mathbb{R}^p$ untuk $i = 1, 2, \dots, n$. Label kelas dinotasikan sebagai: $y_i \in \{-1, 1\}$. Bentuk umum hyperplane adalah

$$\mathbf{w} \bullet \mathbf{x} + b \dots\dots\dots (1)$$

dengan \mathbf{w} adalah vektor normal dengan ukuran $1 \times p$ dan tegak lurus dengan hyperplane dan \mathbf{x} merupakan vektor data pengamatan yang berukuran $p \times 1$. Skalar b disebut dengan simpangan. Fungsi klasifikasi didefinisikan sebagai:

$$f(\mathbf{x}) = \text{sign}(\mathbf{w} \bullet \mathbf{x} + b) \dots\dots\dots(2)$$

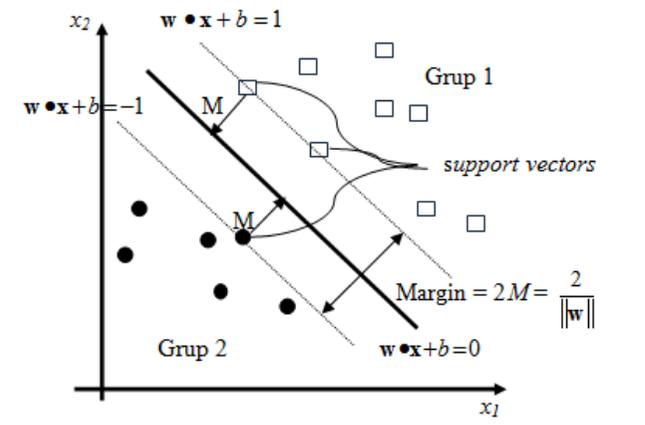
Keckman (2005) menyatakan bahwa apabila $f(\mathbf{x}) > 0$, maka data pengamatan diklasifikasikan ke dalam Grup 1, sedangkan apabila $f(\mathbf{x}) < 0$, maka data pengamatan diklasifikasikan ke dalam Grup 2. Hyperplane dengan margin maksimum diperoleh dengan menemukan solusi dari permasalahan primal (Cortes & Vapnik 1995) yaitu :

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \dots\dots\dots (3)$$

dengan kendala

$$y_i (\mathbf{w} \bullet \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, n \dots\dots\dots(4)$$

Notasi \mathbf{w} , b merupakan parameter-parameter yang akan dicari nilainya. Ilustrasi SVM pada struktur data terpisahkan linier secara sempurna dapat dilihat pada Gambar 1.



Gambar 1 : Ilustrasi SVM Pada Struktur Data Terpisahkan Linier Secara Sempurna.

Masalah SVM dapat diperoleh dengan menyelesaikan pemrograman kuadratik dengan menggunakan fungsi optimasi Lagrangian sebagai berikut :

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i \{y_i (\mathbf{w} \bullet \mathbf{x}_i) - 1\} \dots\dots\dots(5)$$

dengan α_i adalah pengali Lagrange non negatif.

$$L(\mathbf{w}, b, \alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \left(\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \bullet \mathbf{x}_j) \right) \dots\dots\dots(6)$$

dengan kendala $\sum_{i=1}^n \alpha_i y_i = 0$, $\alpha_i \geq 0$, $i, j = 1, \dots, n$. Persamaan (6) merupakan fungsi masalah pengoptimuman dual. Nilai α_i didapatkan dengan cara memaksimumkan fungsi Lagrange pada persamaan (8). Gugus data yang memiliki nilai $\alpha_i > 0$ dinamakan support vector. Gugus data tersebut akan digunakan untuk menghitung bobot $\mathbf{w} = \sum_{i=1}^{nSV} \alpha_i y_i \mathbf{x}_i$ dan $b = (\mathbf{w} \bullet \mathbf{x}_i) - y_i$ untuk $i = 1, \dots, nSV$, dengan nSV adalah banyaknya support vector.

Masalah klasifikasi sesungguhnya muncul pada ruang dimensi tinggi terutama pada data yang terpisahkan linier tidak secara sempurna. Struktur data yang terpisahkan linier tidak secara sempurna adalah data yang berada di dalam margin atau berada pada sisi yang salah dari batas keputusan. Hal ini menyebabkan proses optimisasi tidak dapat diselesaikan, karena tidak ada \mathbf{w} dan b yang memenuhi pertidaksamaan (5).

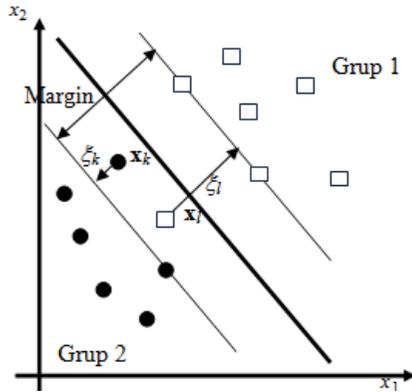
Persamaan (3) dan (4) akan dimodifikasi dengan memasukkan peubah slack ξ_i ($\xi_i > 0$), sehingga menjadi :

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \dots\dots\dots(7)$$

dengan kendala :

$$y_i (\mathbf{w} \bullet \mathbf{x} + b) \geq 1 - \xi_i, \quad \xi_i \geq 0; \quad i = 1, \dots, n$$

C adalah parameter yang menentukan besar penalti akibat kesalahan klasifikasi. Nilai C yang besar akan menghasilkan kesalahan klasifikasi yang kecil. Pada struktur data yang dapat dipisahkan linier secara tidak sempurna, peubah slack didefinisikan sebagai penyimpangan dari batas margin. Ilustrasi SVM pada struktur data terpisahkan linier secara tidak sempurna dapat dilihat pada Gambar 2.



Gambar 2 : Ilustrasi SVM Pada Struktur Data Yang Terpisahkan Linier Secara Tidak Sempurna.

Dua titik data x_i dan x_k pada Gambar 2 memperlihatkan dua titik yang menggambarkan kasus pada data terpisahkan linier secara tidak sempurna dengan penambahan peubah slack ξ_i dan ξ_k . Titik x_i adalah kesalahan klasifikasi karena berada di sisi yang salah batas keputusan. Titik x_k merupakan titik yang berada di dalam margin tetapi diklasifikasikan dengan benar.

Hyperplane yang optimal mempunyai kemampuan generalisasi yang baik termasuk pada dimensi yang tinggi. Gugus data yang tidak dapat dipisahkan secara linier dapat menghasilkan hyperplane yang optimal tetapi belum tentu sebagai alat klasifikasi dengan kemampuan generalisasi yang baik. SVM dapat memetakan data ke ruang dimensi lebih tinggi dengan menggunakan metode kernel sehingga data pada ruang tersebut dapat dipisahkan secara linier dengan transformasi non linier ϕ (Burgess 1998).

Secara umum metode kernel memiliki dua bagian utama. Bagian pertama adalah sebuah modul yang mentransformasikan data dari ruang awal ke dalam ruang baru yang berdimensi tinggi. Bagian kedua adalah suatu algoritma yang berfungsi untuk menemukan pola linier di dalam ruang baru yang terbentuk (Cristianini & Shawe-Taylor 2004).

Fungsi keputusan pada ruang berdimensi tinggi adalah:

$$f(\mathbf{x}) = \text{sign}(\mathbf{w} \bullet \phi(\mathbf{x}) + b) \dots\dots\dots (8)$$

Pada struktur data terpisahkan linier secara tidak sempurna, vektor \mathbf{w} merupakan kombinasi linier dari support vector di ruang berdimensi tinggi. Hal ini berarti :

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \phi(\mathbf{x}_i) \dots\dots\dots (9)$$

Fungsi klasifikasi $f(\mathbf{x})$ pada persamaan (8) bergantung pada hasil kali dalam $\phi(\mathbf{x}_i)$ dan $\phi(\mathbf{x})$ yaitu :

$$f(\mathbf{x}) = \text{sign}(\mathbf{w} \bullet \phi(\mathbf{x}) + b) = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle + b\right)$$

SVM dibangun berdasarkan bentuk umum dari hasil kali dalam ruang Hilbert (Anderson & Bahadur 1966) yaitu :

$$\phi(u) \bullet \phi(v) = K(u, v)$$

Fungsi pemetaan $\phi(\mathbf{x})$ yang memetakan ruang awal ke dalam ruang berdimensi tinggi memenuhi:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \bullet \phi(\mathbf{x}_j)$$

dengan $\phi(\mathbf{x}_i)$ dan $\phi(\mathbf{x}_j)$ gambaran dari ruang berdimensi tinggi dan vektor \mathbf{x}_i dan \mathbf{x}_j sebagai ruang awal. Keuntungan menggunakan fungsi kernel adalah memperlihatkan transformasi non linier ϕ secara eksplisit. Teknik ini biasa dikenal kernel trick. Gunakan kernel trick untuk memaksimalkan masalah dual pada ruang berdimensi tinggi yaitu :

$$\max L(\alpha) = \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \bullet \mathbf{x}_j) \quad (9)$$

dengan kendala : $\sum_{i=1}^n \alpha_i y_i = 0$, $0 \leq \alpha_i \leq C$ untuk $i = 1, \dots, n$. Fungsi keputusan dari masalah dual adalah

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{nSV} \alpha_i y_i K(\mathbf{x}_i \bullet \mathbf{x}) + b\right) \quad (10)$$

dengan b diperoleh dari :

$$b = y_i - \sum_{i=1}^{nSV} \alpha_i y_i K(\mathbf{x}_i \bullet \mathbf{x}) \quad (11)$$

Fungsi kernel $K(\mathbf{x}_i, \mathbf{x}_j)$ yang biasa digunakan dalam SVM (Meyer 2013) adalah :

1) Polinomial:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \left(\mathbf{x}_i \bullet \mathbf{x}_j + 1\right)^d$$

dengan d adalah derajat polinom. Pada R dengan package e1071 digunakan default dengan $d = 3$.

2) Radial basis function (RBF) :

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$$

dengan γ merupakan parameter positif yang mengontrol radius. Pada software R dengan package e1071 menggunakan default $\gamma = (1/\text{dimensi})$.

3) Tangent hyperbolic (sigmoid):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh[v(\mathbf{x}_i \bullet \mathbf{x}_j) + b]$$

untuk nilai parameter v , b telah ditentukan. Pada software R dengan package `e1071` menggunakan default dengan $v = (1/\text{dimensi})$ dan $b = 0$.

2.2. Regresi Logistik Biner

Regresi logistik biner merupakan salah satu metode statistika untuk menganalisis peubah respon (dependen) yang mempunyai skala pengukuran nominal dan terdiri dari dua kategori. Peubah penjelas (independen) yang dapat disertakan dalam model dapat berupa data kategori atau kontinu yang berjumlah dua peubah atau lebih.

Pengujian parameter digunakan untuk mengetahui apakah variabel prediktor yang terdapat dalam model yang diperoleh berpengaruh secara signifikan terhadap variabel respon. Pengujian parameter terdiri dari uji individu dan uji serentak.

Uji individu atau uji parsial adalah pengujian yang dilakukan untuk memeriksa signifikansi parameter β terhadap variabel respon secara individu. Pengujian signifikansi parameter β secara individu menggunakan uji Wald. Statistik uji yang digunakan adalah sebagai berikut:

$$W_1^2 = \frac{(\hat{\beta}_1)^2}{\{SE(\hat{\beta}_1)\}^2} \quad (12)$$

Tolak H_0 jika $W_1^2 > \chi_{(\alpha, db)}^2$.

Uji serentak adalah pengujian yang dilakukan untuk memeriksa signifikansi parameter β terhadap variabel respon secara keseluruhan. Statistik uji yang digunakan adalah sebagai berikut :

$$G = -2 \ln \left[\frac{\binom{n_1}{n} \binom{n_0}{n}^{n_0}}{\prod_{i=1}^n (\pi_i)^{y_i} (1-\pi_i)^{(n-y_i)}} \right] \quad (2.14)$$

Tolak H_0 jika $G > \chi_{(\alpha, db)}^2$.

2.3. Evaluasi Kinerja Klasifikasi

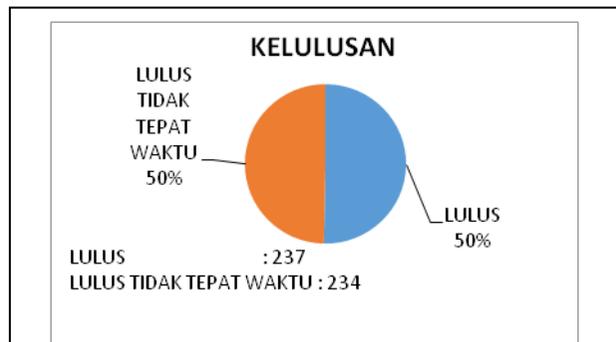
Evaluasi ketepatan klasifikasi model adalah suatu evaluasi yang bertujuan untuk melihat kesalahan klasifikasi yang dilakukan oleh suatu fungsi klasifikasi.

Tabel 1 : Ketepatan Klasifikasi

Hasil Observasi	Prediksi	
	Positive = class 0	Negative = class 1
Positive = class 0	True Positive (TP)	False Positive (FP)
Negative = class 1	False Negative (FN)	True Negative (TN)

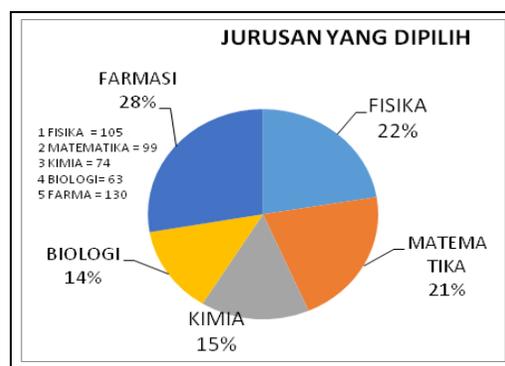
III. HASIL DAN PEMBAHASAN

Setiap tahun Universitas Tadulako menyelenggarakan tiga kali wisuda yaitu di bulan April, Agustus dan Desember. Jumlah wisudawan FMIPA UNTAD tahun lulus 2014-2015 sebanyak 471 mahasiswa. Analisis karakteristik status kelulusan mahasiswa FMIPA UNTAD tahun lulus 2014-2015 dapat ditunjukkan pada Gambar 1. Status kelulusan mahasiswa FMIPA UNTAD tahun 2014-2015 berbeda sangat tipis, mahasiswa yang lulus tepat waktu menggungguli mahasiswa yang lulus tidak tepat waktu dengan selisih 3 mahasiswa. Sehingga persentase status kelulusan tepat waktu dan tidak tepat waktu masing-masing sebesar 50%.



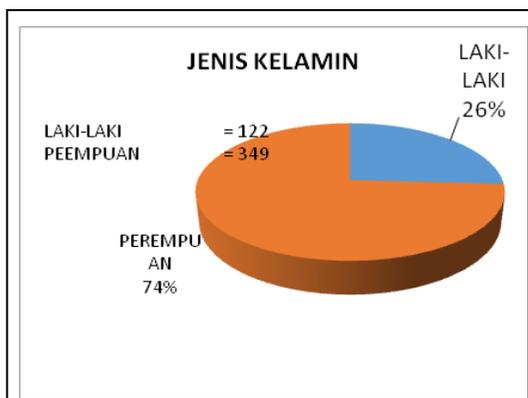
Gambar 3 : Status Kelulusan Mahasiswa FMIPA UNTAD tahun lulus 2014-2015

FMIPA Universitas Tadulako memiliki lima jurusan dan enam program studi. Fisika adalah jurusan pertama yang didirikan oleh FMIPA UNTAD dan Statistika merupakan program studi termuda. Pada tahun 2014-2015 program studi Statistika belum menghasilkan wisudawan. Pada Gambar 3. terlihat bahwa wisudawan terbanyak adalah jurusan Farmasi dan yang paling sedikit menghasilkan wisudawan adalah jurusan Biologi.



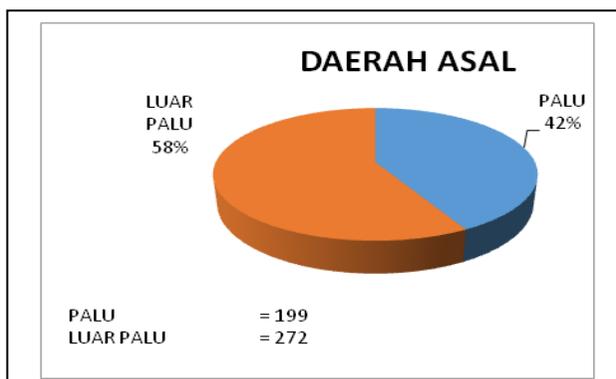
Gambar 4 : Jurusan yang Dipilih Mahasiswa FMIPA UNTAD tahun lulus 2014-2015

Pada Gambar 5 terlihat bahwa wisudawan tahun lulus 2014-2015 didominasi oleh perempuan sebanyak 349 mahasiswa dengan persentase sebesar 74%. Jumlah ini hampir tiga kali lipat dari jumlah wisudawan dengan jenis kelamin laki-laki.



Gambar 5 : Jenis Kelamin Mahasiswa FMIPA UNTAD tahun lulus 2014-2015

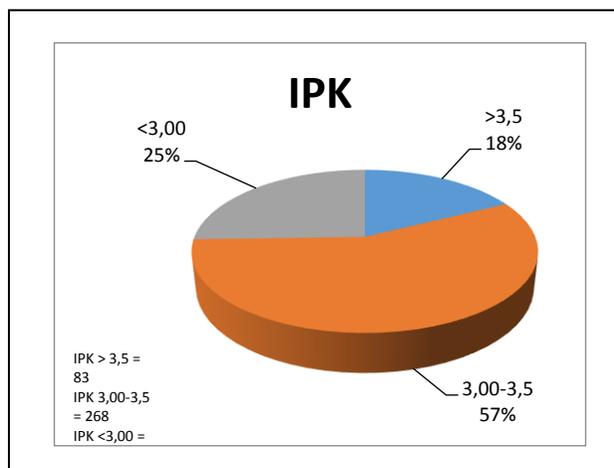
Sulawesi Tengah terdiri dari 12 kabupaten dan 1 kotamadya yaitu kabupaten Toli-toli, Banggai, Banggai kepulauan, Banggai Laut, Buol, Donggala, Morowali, Morowali Utara , Parigi Moutong, Poso, Sigi, Tojo Una-una dan Kota Palu. Mahasiswa yang diterima di Universitas Tadulako berasal dari berbagai kabupaten-kabupaten di Sulawesi Tengah dan berasal dari daerah di luar propinsi Sulawesi Tengah. Pada penelitian ini daerah asal wisudawan tahun lulus 2014-2015 dikategorikan menjadi dua yaitu dari Kota Palu dan luar Palu. Pada Gambar 6 terlihat bahwa daerah asal wisudawan tahun lulus 2014-2015 terbanyak adalah luar Palu dengan persentase sebesar 58%.



Gambar 6 : Daerah Asal Mahasiswa FMIPA UNTAD tahun lulus 2014-2015

Pada Gambar 6, sebaran data IPK mahasiswa Wisudawan tahun lulus 2014-2015 lebih jelas terlihat apabila dikategorikan menjadi tiga yaitu $IPK > 3.50$, $3.00 - 3.50$ dan < 3.00 , hal ini terlihat pada

Gambar 5. IPK antara 3.00 dan 3.50 memiliki persentase terbesar yaitu 57%, nilai ini lebih besar jika dibandingkan dengan IPK > 3.50 yang digabungkan dengan IPK < 3.00.



Gambar 7 : IPK Mahasiswa FMIPA UNTAD tahun lulus 2014-2015

Penerapan metode Support Vector Machine (SVM) pada data wisudawan FMIPA UNTAD tahun lulus 2014-2015 dilakukan dengan menggunakan software R. Software R dapat diperoleh secara bebas pada The Comprehensive R Archive Network (CRAN) di alamat <http://cran.r-project.org/>. Paket yang digunakan untuk mengolah data dengan menggunakan SVM adalah “e1071”, caret, caTools dan kernlab.

Pada langkah awal penelitian, pemilihan kernel dilakukan pada pembelajaran SVM. Perbandingan dilakukan antara kernel dasar (linier) dan kernel RBF SVM [10]. Dengan menggunakan metode sampling 10-fold cross validation diperoleh parameter dan kinerja terbaik dari masing-masing fungsi kernel. Pada kernel linier diperoleh parameter terbaik adalah cost = 10 dan gamma = 0.25 dengan banyak support vector = 168.

Pemilihan kernel SVM menghasilkan kesalahan klasifikasi sebesar 25.79%. Kernel RBF SVM menghasilkan parameter terbaik dengan cost = 10 dan gamma = 0.25 dengan banyaknya support vector = 143. Dengan menggunakan kernel RBF SVM diperoleh kesalahan klasifikasi sebesar 18.02%. Perbandingan hasil akurasi SVM yang diperoleh dengan kernel RBF SVM dan linier menghasilkan akurasi kernel RBF SVM lebih baik dari pada kernel linier dengan data mahasiswa FMIPA UNTAD tahun lulus 2014-2015. Dengan demikian SVM dengan kernel RBF SVM akan dicari lebih lanjut parameter yang kedua yaitu nilai parameter cost dan gamma untuk dapat memperoleh kesalahan klasifikasi yang lebih kecil.

Parameter SVM yang optimal akan diperoleh dengan tuning SVM. Data akan dibagi menjadi data training dan data tes. Data training akan digunakan untuk membangun model sedangkan data tes untuk memprediksi klasifikasi data baru. Pada penelitian ini data training menggunakan 80% dari keseluruhan data dan sisanya sebagai data tes.

Pemilihan nilai cost dan gamma akan dipilih dengan melihat kesalahan klasifikasi paling kecil terhadap nilai cost dan gamma yang diberikan. Berikut adalah output yang dihasilkan oleh program R.

```

best parameters:
gamma cost
0.1 100
best performance: 0.2338549
Detailed performance results:
gamma cost error dispersion
1 1e-06 10 0.4635135 0.11242306
2 1e-05 10 0.4635135 0.11242306
3 1e-04 10 0.3432432 0.08438967
4 1e-03 10 0.2711949 0.07939785
5 1e-02 10 0.2817923 0.10001134
6 1e-01 10 0.2604552 0.08793228
7 1e-06 100 0.4635135 0.11242306
8 1e-05 100 0.3406117 0.08300443
9 1e-04 100 0.2711949 0.07939785
10 1e-03 100 0.2819346 0.10878251
11 1e-02 100 0.2576102 0.08112807
12 1e-01 100 0.2338549 0.08463680
  
```

Hasil percobaan nilai parameter dengan akurasi yang paling tinggi didapat pada nilai gamma = 0.1 dan cost = 100 dengan kesalahan klasifikasi 16,84%. Kesalahan klasifikasi ini lebih baik dibandingkan dengan SVM tanpa optimasi nilai parameter C. Parameter optimasi SVM dengan akurasi yang lebih baik yaitu dengan menggunakan kernel RBF SVM dengan parameter gamma = 0.1 dan cost = 100.

Evaluasi kinerja klasifikasi pada regresi logistik biner diperoleh kesalahan klasifikasi sebesar 19,3%. Hal ini dapat dilihat pada Tabel 2.

Tabel 2 : Ketepatan Klasifikasi Regresi Logistik Biner

Classification Table^a

Observed		Predicted		Percentage Correct
		KELULUSAN tidak	tepat	
Step 1	KELULUSAN tidak	193	44	81,4
	tepat	47	187	79,9
Overall Percentage				80,7

a. The cut value is ,500

IV. KESIMPULAN

Metode dengan kesalahan klasifikasi terkecil adalah metode Support Vector Machine sehingga metode tersebut dapat digunakan untuk mengklasifikasikan ketepatan waktu kelulusan mahasiswa FMIPA UNTAD.

DAFTAR PUSTAKA

- [1]. Vapnik, Vladimir, *Statistical Learning Theory*. New York: Wiley, 1998.
- [2]. Vapnik Vladimir, *The Nature of Statistical Learning Theory*. New York: Springer. 1995.
- [3]. R. Y. Gofur. "Klasifikasi Kelulusan Mahasiswa Fakultas Sains Dan Matematika Universitas Diponegoro Menggunakan Multivariate Adaptive Regression Spline (Mars)," [Skripsi], Universitas Diponegoro, Semarang, 2014.
- [4]. S. Wulansari, Soehono LA, Mitakda MB. Penerapan Metode Chaid (Chi-Squared Automatic Interaction Detection) Dan Analisis Regresi Logistik Biner Pada Klasifikasi Profil Mahasiswa Fmipa Ub [Skripsi]. Malang: Universitas Brawijaya.
- [5]. Y.S. Nugroho. *Data Mining Menggunakan Algoritma Naïve Bayes untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro*.
- [6]. Univeristas Tadulako, *Buku Alumni Wisuda Angkatan 74-77*. Palu, Sulawesi tengah: Universitas Tadulako, 2014.
- [7]. Univeristas Tadulako, *Buku Alumni Wisuda Angkatan 78-81*. Palu, Sulawesi tengah: Universitas Tadulako, 2015.
- [8]. D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*. New York: John Wiley and Sons. 2000.
- [9]. D. Meyer. 2013. Package 'e1071' [Internet]. Bogor (ID):cran. hlm 49-50; [diunduh 2014 Jan 2]. Tersediapada: cran.r-project.org/web/packages/e1071/e1071.pdf
- [10]. J. P. Vert, "Practical Session: Introduction to SVM in R," 2015. [Online]. Available: http://escience.rpi.edu/data/DA/svmbasic_notes.pdf [Diakses: Augus. 15, 2017].