



Penerapan Algoritma *Support Vector Machine* Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu

Emy Haryatmi¹, Sheila Pramita Hervianti²

¹Magister Teknik Elektro, Direktorat Teknologi dan Rekayasa, Universitas Gunadarma

²Magister Manajemen Sistem Informasi, Direktorat Teknologi dan Rekayasa, Universitas Gunadarma

¹emy_h@staff.gunadarma.ac.id, ²sheilapramita@gmail.com

Abstract

A University can have many student data in their database because many students did not graduate on time. Data mining technique can be used to process student data to predict student graduation on time. Support Vector Machine (SVM) algorithm is one of data mining techniques. Objectives of this research was implementation of SVM algorithm to model the prediction of student graduation on time in private university in Indonesia. This research was conducted using CRISP-DM (Cross Industry Standard Process for Data Mining) method. There are five steps in that method such as understanding business to predict student graduation in time which is not available, data understanding by choosing the right attribute for the next step, data preparation includes cleaning the null data and transforming data into category which has been specified, modeling was used by implementing data training and data testing on SVM algorithm and evaluation to validate and measure the accuracy of the model. The result of this research shown that accuracy value of data testing was 94,4% using 90% data training and 10% data testing. This concluded SVM algorithm can be used to model the prediction of student graduation on time.

Keywords: SVM, graduation, students, algorithm, machine learning

Abstrak

Penumpukan data mahasiswa dapat terjadi jika mahasiswa tidak lulus tepat waktu. Teknik *data mining* dapat digunakan untuk mengolah data mahasiswa sehingga dapat menghasilkan prediksi kelulusan mahasiswa tepat waktu. Salah satu teknik data mining yang dapat digunakan adalah algoritma *Support Vector Machine* (SVM). Tujuan dari penelitian ini adalah menerapkan algoritma SVM untuk menghasilkan model prediksi kelulusan mahasiswa tepat waktu pada Fakultas Teknik Universitas Swasta di Indonesia. Metode yang digunakan dalam penelitian ini adalah CRISP-DM (*Cross Industry Standard Process for Data Mining*) dimulai dari pemahaman bisnis dimana belum dapat diprediksi mengenai kelulusan mahasiswa tepat waktu, pemahaman data dengan memilih atribut yang digunakan untuk tahap selanjutnya, pengolahan data yang terdiri dari pembersihan dari data yang isinya null dan transformasi data yang merupakan penentuan kategori dari setiap data, permodelan dilakukan dengan menerapkan algoritma SVM dari data *training* dan data *testing* dan evaluasi untuk memvalidasi dan mengukur keakuratan model. Hasil pengujian kelompok pertama dengan jumlah data *training* sebanyak 90% dan data *testing* sebanyak 10% menunjukkan bahwa algoritma SVM memberikan nilai akurasi yang sangat baik yaitu 94,4%.

Kata kunci: SVM, kelulusan, mahasiswa, algoritma, *machine learning*

1. Pendahuluan

Kelulusan mahasiswa tepat waktu merupakan hal yang paling penting dari suatu universitas. Semakin banyak mahasiswa yang masuk pada suatu universitas maka dengan jumlah yang sama juga harus ada mahasiswa yang lulus tepat waktu. Apabila banyak mahasiswa yang tidak lulus tepat waktu maka akan terjadi peningkatan jumlah data pribadi mahasiswa dan data akademis dari semua mahasiswa yang masih terdaftar.

Data dalam jumlah besar tersebut apabila diolah menggunakan beberapa metode tertentu akan memberikan informasi baru yang dapat membantu pihak perguruan tinggi dalam mengambil keputusan atau kebijakan. Salah satu kebijakannya adalah dapat memprediksi kelulusan secara dini untuk meningkatkan jumlah kelulusan mahasiswa dan meminimalisir jumlah mahasiswa yang tidak lulus tepat waktu sesuai dengan masa studinya. Salah satu teknik melakukan prediksi yang dapat digunakan adalah dengan teknik penggalian

data atau *data mining*. *Data mining* merupakan suatu kumpulan teknik yang digunakan sebagai bagian dari pengetahuan dalam pencarian di basis data. Tujuan dari *data mining* adalah untuk ekstraksi data yang berguna dari kumpulan data yang tersedia. *Data mining* merupakan proses yang menggunakan satu atau lebih teknik pada *machine learning* untuk melakukan analisis dan ekstraksi secara otomatis dari jumlah data yang banyak. *Data mining* dapat digunakan secara luas dalam seluruh penyimpanan informasi [1]. Beberapa faktor yang menentukan kriteria dari *data mining* seperti *data mining* adalah proses pencarian nilai yang telah ditambahkan ke kumpulan data yang telah lampau, objek *data mining* merupakan data dalam jumlah banyak dan kompleks dan tujuan dari *data mining* mencari hubungan atau bentuk yang dapat memberikan indikasi yang berguna [2].

Klasifikasi adalah suatu teknik yang dapat digunakan untuk memetakan masukan menjadi keluaran diskrit yang dinamakan dengan label dan kategori. Kategori untuk suatu observasi diprediksi oleh pemetaan fungsi. Sebagai contoh, performa dari suatu kumpulan mahasiswa dapat diklasifikasikan sebagai “lulus” atau “tidak lulus”. Teknik klasifikasi terdiri dari *Support Vector Machine*, *Naïve Bayes*, *Discriminant Analysis*, *Nearest Neighbor*, *Neural Network*, dan *Naïve Bayes* [3]. Salah satu metode yang dapat digunakan untuk klasifikasi adalah *Support Vector Machine* (SVM). SVM [4] berawal dari klasifikasi dua kelas yang membutuhkan *training set* positif dan negatif. SVM berusaha mencari *hyperplane* (pemisah) terbaik yang berfungsi untuk memisahkan dua kelas dan mendapatkan margin terbaik antara dua kelas tersebut. Beberapa peneliti telah menggunakan SVM untuk melakukan prediksi waktu kelulusan mahasiswa seperti yang dilakukan oleh [5] sebagai metode klasifikasi. Algoritma SVM digunakan untuk menentukan model prediksi masa studi mahasiswa dari fakultas ilmu komputer. Berdasarkan hasil yang didapatkan dari pengujian, algoritma SVM memberikan nilai akurasi yang tinggi yaitu 83,64% pada percobaan pertama dan 77% pada percobaan kedua. Pengujian juga dilakukan terhadap algoritma *decision tree* dan *Naïve Bayes*, namun hasil dari algoritma SVM masih lebih baik dibandingkan dengan kedua algoritma tersebut.

Penelitian yang dilakukan oleh [6] menggunakan sepuluh algoritma *machine learning* dalam menentukan model untuk memprediksi kelulusan tepat waktu bagi calon mahasiswa, salah satunya adalah SVM. Berdasarkan pengujian dengan menggunakan data yang diimplementasikan pada setiap algoritma *machine learning*, maka algoritma SVM memberikan nilai akurasi paling tinggi yaitu 65% dibandingkan algoritma lainnya. Hal ini menunjukkan bahwa model yang dihasilkan dari algoritma SVM mampu memprediksi kelulusan tepat waktu bagi calon mahasiswa. Penelitian

yang juga menggunakan SVM untuk menentukan model terbaik dalam prediksi kelulusan mahasiswa adalah [4] dengan menggunakan *parameter* yang telah ditentukan. Perancangan yang digunakan pada penelitian ini adalah memasukkan data, melakukan normalisasi data, melakukan *training SVM*, melakukan *testing SVM* dan mendapatkan hasil klasifikasi. Data yang digunakan sebanyak 188 data dengan menggunakan jumlah data latih yang berbeda, pengujian pengaruh nilai *parameter* pada *Sequential Training SVM* dan *kernel*. Berdasarkan pengujian didapatkan pengaruh penggunaan *kernel Gaussian RBF* pada SVM memberikan hasil yang terbaik dengan tingkat akurasi sebesar 80,55%. Penelitian mengenai penerapan algoritma SVM untuk mendapatkan model prediksi kelulusan mahasiswa tepat waktu juga dilakukan oleh [7]. Pada penelitian ini, dibandingkan juga dengan algoritma *machine learning* lainnya yaitu *Naïve Bayes*. Data yang digunakan dalam penelitian ini seperti GPA, data profil, asal sekolah SMA dan tempat tinggal. Berdasarkan hasil pengujian dengan data-data tersebut, algoritma SVM memberikan nilai akurasi yang lebih baik dibandingkan dengan algoritma *Naïve Bayes* yaitu 69,15%. Model prediksi kelulusan mahasiswa tepat waktu yang dihasilkan oleh algoritma SVM lebih baik untuk digunakan dengan data tersebut.

Penelitian yang dilakukan oleh [8] melakukan perbandingan tiga algoritma *machine learning* dalam menghasilkan model yang mampu memprediksi kelulusan mahasiswa tepat waktu di Fakultas Teknik Universitas Jenderal Soedirman. Algoritma yang digunakan adalah *decision tree*, SVM dan *artificial neural network* (ANN). Berdasarkan hasil pengujian dengan menggunakan data yang diterapkan pada tiga algoritma tersebut menghasilkan bahwa algoritma SVM mampu memberikan tingkat akurasi yang paling baik yaitu 90,55% dibandingkan algoritma *decision tree* dan ANN. Model yang dihasilkan dari SVM dapat digunakan sebagai peringatan dini dalam kelulusan mahasiswa tepat waktu. Penelitian [9] membahas mengenai presentase kesalahan klasifikasi yang dihasilkan dari algoritma dalam menentukan model kelulusan mahasiswa tepat waktu di FMIPA UNTAD. Dua algoritma digunakan untuk membandingkan nilai kesalahan yang dihasilkan dari masing-masing algoritma yaitu SVM dan regresi logistik biner. Hasil pengujian didapatkan bahwa nilai kesalahan klasifikasi yang dihasilkan oleh SVM sebesar 16,84% sedangkan nilai kesalahan klasifikasi dari regresi logistik biner sebesar 19,3%. Algoritma SVM memiliki nilai kesalahan yang lebih kecil dibandingkan dengan regresi logistik biner sehingga SVM dapat digunakan pada proses klasifikasi kelulusan mahasiswa tepat waktu FMIPA UNTAD.

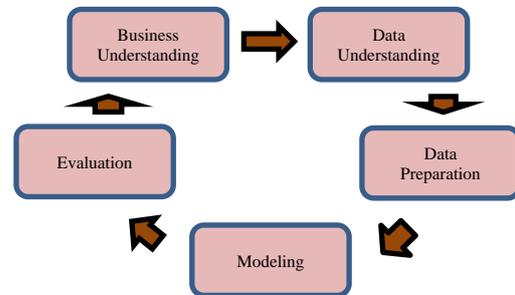
Penelitian mengenai penggunaan SVM juga dilakukan oleh [10]. Pada penelitian ini, algoritma SVM dibandingkan dengan dua algoritma lainnya yaitu KNN

dan *Decision Tree* dalam mendapatkan model terbaik untuk memprediksi mahasiswa yang dapat lulus tepat waktu. Berdasarkan hasil pengujian didapatkan bahwa SVM memberikan tingkat akurasi yang paling tinggi yaitu sebesar 95% sedangkan KNN memberikan akurasi sebesar 92% dan *Decision Tree* memberikan akurasi sebesar 93% berdasarkan dari data yang digunakan. Model yang dihasilkan oleh SVM dapat digunakan untuk memprediksi kelulusan mahasiswa tepat waktu. Penelitian yang dilakukan oleh [11] membandingkan algoritma *support vector machine* (SVM), *neural network* (NN) dan *decision tree* model (DT) untuk menghasilkan model yang dapat memprediksi waktu kelulusan dari mahasiswa pada salah satu universitas swasta di Indonesia. Akurasi yang didapatkan dari ketiga algoritma tersebut menunjukkan bahwa penggunaan algoritma SVM memberikan tingkat akurasi yang paling tinggi yaitu 85,18%, sementara algoritma NN memiliki tingkat akurasi sebesar 84,68% dan akurasi dari DT sebesar 84,96%. Berdasarkan hasil tersebut, penggunaan algoritma SVM mampu memberikan model yang paling baik dalam memprediksi waktu kelulusan karena SVM memiliki tingkat akurasi yang paling tinggi.

Penelitian yang dilakukan oleh [12] juga menggunakan SVM sebagai algoritma dalam menghasilkan model untuk memprediksi kelulusan mahasiswa dengan menggunakan pembelajaran elektronik. Berdasarkan hasil pengujian yang telah dilakukan terhadap penggunaan *data log* dari mahasiswa terhadap algoritma SVM diketahui bahwa algoritma SVM mampu memberikan tingkat akurasi sebesar 85,02%. Hal ini membuktikan bahwa model yang dihasilkan dari algoritma SVM mampu memprediksi kelulusan mahasiswa pada pembelajaran elektronik. Penelitian mengenai review penggunaan algoritma machine learning dilakukan oleh [3]. Penelitian ini mengulas dari penelitian-penelitian yang telah diteliti oleh beberapa peneliti mengenai penggunaan algoritma machine learning dalam menghasilkan model untuk memprediksi waktu kelulusan mahasiswa berdasarkan faktor utama dari masing-masing penelitian. Hasil review memperlihatkan bahwa faktor yang paling dominan dalam menghasilkan model untuk prediksi waktu kelulusan mahasiswa adalah hasil ujian mahasiswa. Algoritma yang paling akurat dalam menghasilkan model prediksi waktu kelulusan mahasiswa di Malaysia adalah SVM dan *neural network* dibandingkan algoritma lainnya. Berdasarkan data yang dimiliki dan penelitian terkait yang telah menggunakan algoritma SVM yang menghasilkan tingkat akurasi lebih baik dibandingkan dengan algoritma lainnya dalam menentukan model prediksi kelulusan mahasiswa tepat waktu, maka tujuan dari penelitian ini adalah menerapkan algoritma SVM untuk menghasilkan model prediksi kelulusan mahasiswa tepat waktu pada Fakultas Teknik Universitas Swasta di Indonesia.

2. Metode Penelitian

Metode penelitian yang digunakan adalah model CRISP-DM (*Cross Industry Standard Process for Data Mining*). Penggambaran metode tersebut terdiri dari fase pemahaman bisnis, pemahaman data, pengolahan data, permodelan dan evaluasi dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

2.1. Pemahaman Bisnis (*Business Understanding*)

Jumlah kelulusan mahasiswa tepat waktu belum dapat diprediksi. Total mahasiswa dalam sistem akademik sebanyak 2192 mahasiswa pada Fakultas Teknik Universitas Swasta di Indonesia dari tahun 2009-2016 yang bersumber dari Bagian Kemahasiswaan Fakultas Teknik. Data ini masih bercampur antara mahasiswa yang sudah lulus tepat waktu, mahasiswa yang belum lulus tepat waktu dan mahasiswa yang masih dalam masa studi.

2.2. Pemahaman Data (*Data Understanding*)

Atribut data mahasiswa yang digunakan adalah usia, jenis kelamin, IPK, IPS 1, IPS 2, IPS 3, IPS 4, jurusan, SKS dan lama masa studi. Usia mahasiswa dihitung dari tanggal kelahiran sampai tanggal mahasiswa mendaftarkan diri di Universitas. Jenis kelamin yang digunakan adalah laki-laki dan perempuan. IPK yang digunakan antara 2.00 – 4.00. IPS 1, 2, 3, 4 yang digunakan antara 2.00 – 4.00. Jurusan yang digunakan adalah seluruh jurusan dalam Fakultas Teknik. SKS yang menjadi nilai acuan untuk kelulusan adalah 144 SKS. Lama studi kurang dari 4 tahun artinya mahasiswa lulus tepat waktu. Lama studi lebih dari 4 tahun artinya mahasiswa lulus tidak tepat waktu.

2.3. Pengolahan Data (*Data Preparation*)

Data yang dikumpulkan masih banyak yang berisikan null sehingga diperlukan tahap *pre-processing*. Tahap *preprocessing* data meliputi proses pembersihan data (*cleaning*) dan transformasi data. Proses pembersihan data dilakukan terhadap atribut data yang isinya null. Atribut ini dihapus dan tidak digunakan dalam proses selanjutnya seperti terlihat pada Tabel 1. Setelah dilakukan proses pembersihan, data yang dapat digunakan sebanyak 2181 dari 2192 data.

Tabel 1. Contoh Hasil Pembersihan Data Akademik

Jurusan	Umur	JK	IPS1	IPS2	IPS3	IPS4	IPK	SKS
Teknik Elektro	17	L	2,11	2,11	2,17	2,17	2,70	159
Teknik Elektro	18	L	1,26	1,26	0,94	0,00	0,94	44
Teknik Elektro	22	L	3,27	3,27	2,95	2,95	2,77	148
Teknik Elektro	18	L	1,74	1,74	1,48	1,48	1,48	57
Teknik Elektro	18	L	0,00	0,00	0,00	0,00	0,50	19
Teknik Elektro	18	L	0,00	0,00	0,00	0,00	0,87	19
Teknik Elektro	19	L	2,92	2,92	2,62	2,62	1,94	170
Teknik Elektro	18	L	0,61	0,61	0,40	0,00	0,32	57
Teknik Elektro	18	L	2,73	2,73	2,69	2,69	2,63	163
Teknik Elektro	18	L	1,00	1,00	0,00	0,00	1,00	19
Teknik Elektro	19	L	2,08	2,08	1,86	1,86	2,28	169
Teknik Elektro	21	L	2,55	2,55	2,57	2,57	2,91	157
Teknik Elektro	20	L	2,42	2,42	0,00	0,00	2,42	40
Teknik Elektro	19	L	2,08	2,08	0,00	0,00	2,08	37
Teknik Elektro	37	L	1,38	1,38	1,04	0,00	1,04	47

Tabel 2. Contoh Hasil Transformasi Data Akademik

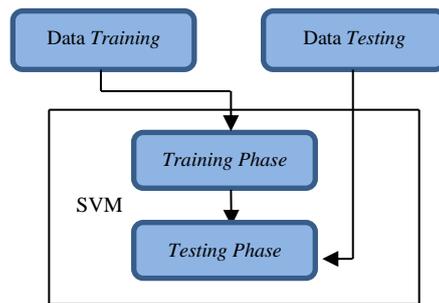
Jurusan	Umur	JK	IPS1	IPS2	IPS3	IPK	SKS	Status
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	BANYAK	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	SEDIKIT	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	CUKUP	CUKUP	CUKUP	CUKUP	TEPAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	SEDIKIT	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	SEDIKIT	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	BANYAK	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	SEDIKIT	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	BANYAK	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	SEDIKIT	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	BANYAK	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	CUKUP	BANYAK	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	SEDIKIT	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	SEDIKIT	TERLAMBAT
Teknik Elektro	LEBIH	L	RENDAH	RENDAH	RENDAH	RENDAH	SEDIKIT	TERLAMBAT

Proses transformasi merupakan tahapan selanjutnya setelah proses pembersihan data. Proses transformasi dari atribut data yang memiliki angka dilakukan untuk memudahkan penambahan data. Proses transformasi dari atribut data yang digunakan ditentukan dengan membuat kategori dari setiap data. Data IPK lebih besar dari 3,50 ditentukan sebagai kategori pujian. IPK antara 3,01 sampai 3,50 ditentukan sebagai kategori sangat memuaskan. IPK antara 2,75 sampai 3,00 ditentukan sebagai kategori memuaskan. Data IPS sama dengan atau lebih besar dari 3,50 ditentukan sebagai kategori tinggi. IPS antara 2,75 sampai 3,50 ditentukan sebagai kategori cukup. IPS sama dengan atau kurang dari 2,75 ditentukan sebagai kategori rendah. Total SKS yang telah ditempuh sama dengan atau lebih dari 144 SKS sebagai kategori tercapai. Total SKS yang telah ditempuh kurang dari 144 SKS sebagai kategori tidak tercapai. Usia sama dengan atau kurang dari 18 tahun sebagai kategori sesuai. Usia lebih dari 18 tahun sebagai

kategori lebih. Contoh hasil transformasi data dapat dilihat pada Tabel 2.

2.4. Permodelan (Modeling)

Tahap permodelan menggunakan algoritma Support Vector Machine (SVM). Sebelum menerapkan data-data dalam algoritma SVM, data training dan data testing ditentukan terlebih dahulu dari data yang merupakan hasil transformasi. Data *training* dimasukkan ke dalam fase *training* pada algoritma SVM. Setelah didapatkan modelnya, data *testing* dimasukkan ke dalam fase *testing* pada algoritma SVM. Tahapan permodelan dapat dilihat pada Gambar 2.



Gambar 2. Tahap Permodelan

Persentase data *training* dan data *testing* yang digunakan dibagi dalam 3 kelompok yang dapat dilihat pada Tabel 3.

Tabel 3. Kelompok Data *Training* dan Data *Testing*

Kelompok	Data Training	Jumlah Data	Data Testing	Jumlah Data
1	90%	1963	10%	218
2	80%	1745	20%	436
3	70%	1527	30%	654

Kelompok 1 menggunakan 90% data *training* dengan jumlah data sebanyak 1963 dan 10% data *testing* dengan jumlah data sebanyak 218. Data *training* yang digunakan pada kelompok 2 sebesar 80% dengan jumlah data sebanyak 1745 dan 20% data *testing* dengan jumlah data adalah 436. Kelompok terakhir yaitu kelompok 3 menggunakan 70% data *training* dimana jumlah data yang digunakan sebanyak 1527 serta 30% data *testing* dimana jumlah data yang digunakan sebanyak 654. Data setiap kelompok dimasukkan dalam algoritma SVM secara bergantian. Kelompok pertama merupakan data pertama yang dimasukkan dalam algoritma SVM. Setelah memasukkan data, memasukkan nilai C (*cost*). Nilai C yang digunakan adalah 1. Selanjutnya dilakukan perhitungan SVM secara linier. Kelompok kedua dan ketiga mengikuti prosedur yang sama dengan kelompok pertama.

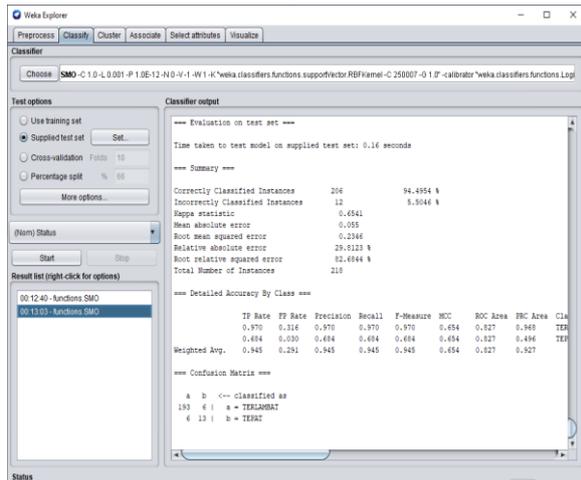
2.5 Evaluasi (Evaluation)

Pengolahan keseluruhan data menggunakan aplikasi WEKA dengan menggunakan teknik *confusion matrix* untuk menghasilkan nilai *recall*, *precision*, *F-measure*, *accuracy* dan *ROC Curve* (AUC). Validitas dan

keakuratan hasil dari model yang telah didapatkan dievaluasi berdasarkan nilai-nilai tersebut.

3. Hasil dan Pembahasan

Berdasarkan perancangan yang telah dijelaskan sebelumnya, maka hasil pengujian dibagi dalam 3 kelompok. Hasil pengujian 1 dapat dilihat pada Gambar 3. Hasil pengujian kelompok 1 dengan menggunakan data *training* sebanyak 1963 mahasiswa dan data *testing* sebanyak 218 mahasiswa.



Gambar 3. Hasil Pengujian Kelompok 1

Hasil dari pengujian 1 dengan menggunakan persentase data *training* 90% dan data *testing* 10%, maka nilai akurasi data *training* sebesar 92,5%, nilai akurasi data *testing* sebesar 94,4%, nilai *precision* data *testing* sebesar 94,5%, nilai *recall* data *testing* sebesar 94,5% dan nilai *F-measure* data *testing* sebesar 94,5%. *Confusion matrix* yang dihasilkan dari pengujian 1 dapat dilihat pada Tabel 4.

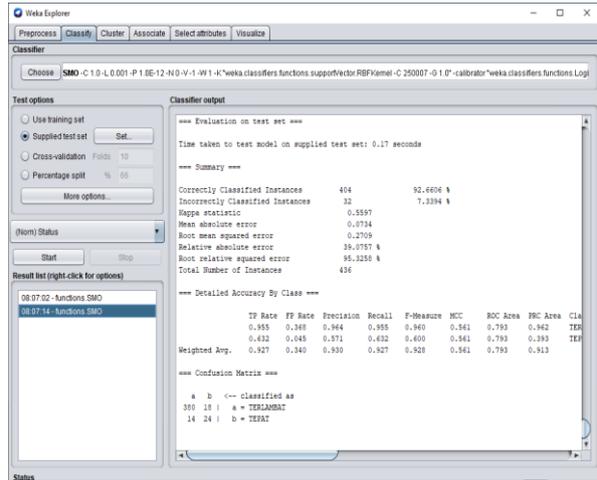
Tabel 4. Confusion Matrix Pengujian 1

	Kondisi Positif (Tidak Tepat Waktu)	Kondisi Negatif (Tepat Waktu)
Prediksi Kondisi Positif (Tidak Tepat Waktu)	193	6
Prediksi Kondisi Negatif (Tepat Waktu)	6	13

Berdasarkan hasil yang diberikan oleh *confusion matrix* dari pengujian pertama, maka sebanyak 193 mahasiswa yang diprediksi lulus tidak tepat waktu dan kenyataannya mahasiswa tersebut lulus tidak tepat waktu. Kemudian, sebanyak 6 mahasiswa yang diprediksi lulus tepat waktu dan kenyataannya mahasiswa tersebut tidak lulus tepat waktu. Selanjutnya, sebanyak 6 mahasiswa yang diprediksi tidak lulus tepat waktu dan kenyataannya mahasiswa tersebut dapat lulus tepat waktu. Kemudian, sebanyak 13 mahasiswa yang

diprediksi lulus tepat waktu dan kenyataannya mahasiswa tersebut dapat lulus tepat waktu. Berdasarkan jumlah data mahasiswa yang digunakan sebagai data *testing* yaitu 218, maka yang lulus tidak tepat waktu yaitu 199 mahasiswa dan yang lulus tepat waktu sebanyak 19 mahasiswa.

Hasil pengujian kedua menggunakan kelompok 2 yaitu 1745 mahasiswa sebagai data *training* dan 436 mahasiswa sebagai data *testing*. Hasil pengujian dapat dilihat pada Gambar 4.



Gambar 4. Hasil Pengujian Kelompok 2

Pengujian kelompok 2 menggunakan 80% data *training* dan 20% data *testing*. Nilai akurasi data *training* yang dihasilkan sebesar 92,7%, nilai akurasi data *testing* yang dihasilkan sebesar 92,5%, nilai *precision* data *testing* yang dihasilkan sebesar 93%, nilai *recall* data *testing* yang dihasilkan sebesar 92,7% dan nilai *F-measure* data *testing* yang dihasilkan sebesar 92,8%. *Confusion matrix* yang dihasilkan dari pengujian kelompok 2 dapat dilihat pada Tabel 5.

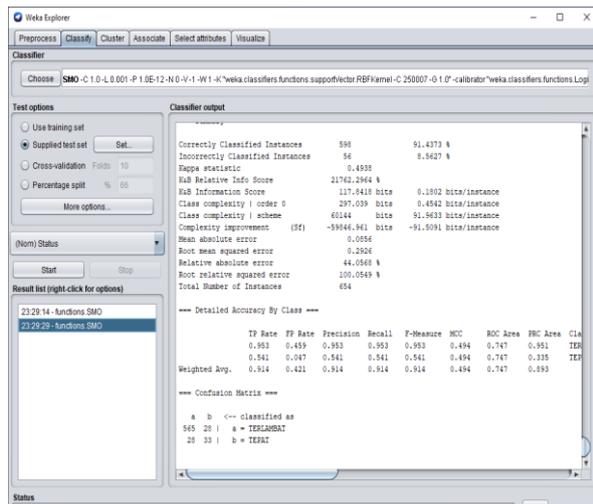
Tabel 5. Confusion Matrix Pengujian 2

	Kondisi Positif (Tidak Tepat Waktu)	Kondisi Negatif (Tepat Waktu)
Prediksi Kondisi Positif (Tidak Tepat Waktu)	380	18
Prediksi Kondisi Negatif (Tepat Waktu)	14	24

Confusion matrix yang dihasilkan terhadap pengujian kedua yaitu sebanyak 380 mahasiswa yang diprediksi lulus tidak tepat waktu dan kenyataannya mahasiswa tersebut lulus tidak tepat waktu. Selanjutnya, sebanyak 14 mahasiswa yang diprediksi lulus tepat waktu dan kenyataannya mahasiswa tersebut tidak lulus tepat waktu. Kemudian, sebanyak 18 mahasiswa yang diprediksi tidak lulus tepat waktu dan kenyataannya mahasiswa tersebut lulus tepat waktu. Selanjutnya,

sebanyak 24 mahasiswa yang diprediksi lulus tepat waktu dan kenyataannya mahasiswa tersebut lulus tepat waktu. Pengujian kedua dengan data *testing* yang digunakan adalah 436, maka lulus tidak tepat waktu sebanyak 394 mahasiswa dan 42 mahasiswa dinyatakan lulus tepat waktu.

Pengujian ketiga menggunakan data kelompok 3 yaitu 1527 mahasiswa sebagai data *training* dan 654 mahasiswa sebagai data *testing*. Hasil pengujian dapat dilihat pada Gambar 5.



Gambar 5. Hasil Pengujian Kelompok 3

Penggunaan presentase data *training* sebesar 70% dan data *testing* sebesar 30% menghasilkan nilai akurasi data *training* sebesar 93,1%, nilai akurasi data *testing* sebesar 91,4%, nilai *precision* data *testing* sebesar 91,4%, nilai *recall* sebesar 91,4% dan nilai *F-measure* sebesar 91,4%. Hasil pengujian juga menghasilkan confusion matrix seperti yang terlihat pada Tabel 6.

Tabel 6. Confusion Matrix Pengujian 3

	Kondisi Positif (Tidak Tepat Waktu)	Kondisi Negatif (Tepat Waktu)
Prediksi Kondisi Positif (Tidak Tepat Waktu)	565	28
Prediksi Kondisi Negatif (Tepat Waktu)	28	33

Hasil dari confusion matrix pada pengujian ketiga, sebanyak 565 mahasiswa yang diprediksi tidak dapat lulus tepat waktu dan kenyataannya tidak lulus tepat waktu. Sebanyak 28 mahasiswa diprediksi dapat lulus tepat waktu dan kenyataannya tidak dapat lulus tepat waktu. Sebanyak 28 mahasiswa diprediksi tidak dapat lulus tepat waktu dan kenyataannya lulus tepat waktu. Sebanyak 33 mahasiswa diprediksi dapat lulus tepat waktu dan kenyataannya dapat lulus tepat waktu. Data *testing* yang digunakan sebanyak 654 memperlihatkan

bahwa 593 mahasiswa tidak dapat lulus tepat waktu dan 61 mahasiswa dapat lulus tepat waktu.

Berdasarkan dari hasil pengujian terhadap 3 kelompok persentase data *training* dan data *testing*, algoritma SVM memberikan nilai akurasi paling tinggi terhadap data *testing* yaitu sebesar 94,4% dengan persentase data *training* dan data *testing* sebesar 90% dan 10%. Nilai akurasi dari presentase data *training* dan data *testing* sebesar 80% dan 20% adalah 92,6%. Nilai akurasi dari presentase data *training* dan data *testing* sebesar 70% dan 30% adalah 91,4%. Hal ini menandakan bahwa tingkat kedekatan antara nilai prediksi dan nilai aktual mendekati 100% sehingga data *training* dan data *testing* yang digunakan dapat dikatakan valid.

4. Kesimpulan

Penerapan algoritma SVM terhadap atribut data mahasiswa yang digunakan seperti usia, jenis kelamin, IPK, IPS 1, IPS 2, IPS 3, IPS 4, jurusan, SKS dan lama masa studi menghasilkan nilai akurasi yang cukup tinggi yaitu 94,4%. Hasil ini didapatkan dari jumlah data *training* yang digunakan sebanyak 1963 mahasiswa dan data *testing* yang digunakan sebanyak 218 mahasiswa dari total data mahasiswa sebanyak 2181 mahasiswa setelah dilakukan pembersihan data. Hal ini menandakan bahwa model yang dihasilkan dari algoritma SVM dapat digunakan untuk memprediksi tingkat kelulusan mahasiswa tepat waktu dari Fakultas Teknik Universitas Swasta di Indonesia.

Daftar Rujukan

- [1] N. Mayasari, "Comparison of Support Vector Machine and Decision Tree in Predicting On-Time Graduation (Case Study : Universitas Pembangunan Panca Budi)," *Int. J. Recent Trends Eng. Res.*, vol. 2, no. 12, pp. 140–151, Dec. 2016.
- [2] L. Marlina, M. Lim, and A. P. Utama Siahaan, "Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms)," *Int. J. Eng. Trends Technol.*, vol. 38, no. 7, pp. 380–383, Aug. 2016.
- [3] N. Mohammad Suhaimi, S. Abdul-Rahman, S. Mutalib, N. H. Abdul Hamid, and A. Hamid, "Review on Predicting Students' Graduation Time Using Machine Learning Algorithms," *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, no. 7, pp. 1–13, Jul. 2019.
- [4] A. Pratama, R. C. Wihandika, and D. E. Ratnawati, "Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 4, pp. 1704–1708, Mar. 2018.
- [5] T. Handhayani and L. Hiryanto, "Predicting And Analyzing The Student's Length of Studi-Time Using Support Vector Machine," *ComTech Comput. Math. Eng. Appl.*, vol. 8, no. 2, pp. 107–114, Mar. 2017.
- [6] A. Saifudin, "Metode Data Mining Untuk Seleksi Calon Mahasiswa Pada Penerimaan Mahasiswa Baru di Universitas Pamulang," *J. Teknol.*, vol. 10, no. 1, pp. 25–36, Jul. 2018.
- [7] A. Kesumawati and D. T. Utari, "Predicting Patterns of Student Graduation Rates Using Naïve Bayes Classifier and Support Vector Machine," *AIP Conf. Proc.*, vol. 2021, no. 1, pp. 060005-1-060005-10, Oct. 2018.
- [8] A. Fadli, M. I. Zulfa, and Y. Ramadhani, "Performance Comparison of Data Mining Classification Algorithms for Early Warning System of Students Graduation Timeliness," *J. Teknol. dan Sist. Komput.*, vol. 6, no. 4, pp. 158–163, Oct. 2018.

- [9] I. T. Utami, "Perbandingan Kinerja Klasifikasi Support Vector Machine (SVM) Dan Regresi Logistik Biner Dalam Mengklasifikasikan Ketepatan Waktu Kelulusan Mahasiswa Fmipa Untad," *J. Ilm. Mat. Dan Terap.*, vol. 15, no. 2, pp. 256–267, Dec. 2018.
- [10] S. Wiyono and T. Abidin, "Comparative Study of Machine Learning KNN, SVM, and Decision Tree Algorithm To Predict Student's Performance," *Int. J. Res. -GRANTHAALAYAH*, vol. 7, no. 1, pp. 190–196, Jan. 2019.
- [11] V. Riyanto, A. Hamid, and R. Ridwansyah, "Prediction of Student Graduation Time Using the Best Algorithm," *Indones. J. Artif. Intell. Data Min.*, vol. 2, no. 1, pp. 1–9, Mar. 2019.
- [12] R. A. Permana and S. Sahara, "Metode Support Vector Machine Sebagai Penentu Kelulusan Mahasiswa pada Pembelajaran Elektronik," *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 50–58, Jun. 2019.