

## METODE SUPPORT VECTOR MACHINE (SVM) UNTUK MEMPREDIKSI MASA STUDI MAHASISWA IST AKPRIND YOGYAKARTA

Riky Aldiman<sup>1</sup>, Kris Suryowati<sup>2</sup>, Rokhana Dwi Becti<sup>3</sup>

<sup>1,2,3</sup>Jurusan Statistika, Fakultas Sains Terapan, IST AKPRIND Yogyakarta

Email: <sup>1</sup>riky.aldi@gmail.com, <sup>2</sup>suryowati@akprind.ac.id, <sup>3</sup>rokhana@akprind.ac.id

**Abstract**—One of the statistical methods in data mining used to classify between two classes is the Support Vector Machine method by separating the two classes of data based on the optimum hyperplane line by maximizing the margin to minimize the error, in the case of the real world often found difficulty in separating the two classes in a linear manner it is necessary to use a kernel to be able to separate data linearly such as RBF, Linear, Polynomial kernels.

This study discusses the study period of IST Akprind graduate students in 2015-2017 using the Support Vector Machine method. This method serves to see the classification of the accuracy of the study period of students and predict the study period to be passed by the students based on predictor gender (X1), Hometown (X2), System Register (X3), Achievement Index Semester 2 (X4), and Faculty (X5). With the dependent variable the study period is categorized into two ie 1 and -1. With a value of 1 stating on time if the study period  $\leq 8$  semester and the value -1 states is not appropriate if  $> 8$  semesters. Based on available data obtained that 87.31% of students undergo the study period is not on time and 12.69% can finish on time.

From the results of classification analysis found that the best model is a classification model of support vector machine using linear kernel with parameter cost (C) of 0.1 and error of 0.126193. forecasting that all students will be studying for more than 4 years.

**Keywords** : Student Studi Period, Support Vector Machine, Kernel (RBF, Linear, Polynomial)

**Abstrak**—Salah satu metode statistik dalam data mining yang digunakan untuk mengklasifikasikan antar dua kelas adalah metode *Support Vector Machine* dengan memisahkan dua kelas data berdasarkan garis *hyperplane* optimum dengan cara memaksimalkan margin untuk meminimalkan *error*, dalam kasus dunia nyata sering ditemukan kesulitan dalam memisahkan dua kelas secara linier sehingga perlu digunakan sebuah kernel untuk dapat memisahkan data secara linier seperti kernel *RBF*, *Linear*, *Polynomial*.

Penelitian ini membahas masa studi mahasiswa lulusan IST Akprind pada tahun 2015-2017 dengan menggunakan metode *Support Vector Machine*. Metode ini berfungsi untuk melihat klasifikasi ketepatan masa studi mahasiswa dan memprediksi masa studi yang akan dilalui mahasiswa berdasarkan variabel prediktor jenis kelamin (X<sub>1</sub>), asal daerah (X<sub>2</sub>), jalur masuk (X<sub>3</sub>), nilai indeks prestasi semester 2 (IPS 2) (X<sub>4</sub>), dan fakultas (X<sub>5</sub>). Dengan variabel dependen masa studi dikategorikan menjadi dua yaitu 1 dan -1. Dengan nilai 1 menyatakan tepat waktu apabila masa studi  $\leq 8$  semester dan nilai -1 menyatakan tidak tepat apabila  $> 8$  semester. Berdasarkan data yang tersedia diperoleh bahwa 93% mahasiswa menjalani masa studinya tidak tepat waktu dan 7% dapat menyelesaikan tepat waktu.

Dari hasil analisis klasifikasi didapatkan bahwa model terbaik adalah model klasifikasi support vector machine dengan menggunakan kernel linier dengan parameter cost (C) sebesar 0.1 dan error sebesar 0.126193. Peramalan yang dihasilkan bahwa seluruh mahasiswa akan menempuh masa studi lebih dari 4 tahun.

Kata kunci: Masa Studi, Support Vector Machine, Kernel (RBF, Linear, Polynomial)

## 1. PENDAHULUAN

Pendidikan adalah kegiatan pembelajaran yang dilaksanakan melalui interaksi antara siswa dengan guru dan sumber belajar pada lingkungan belajar, pendidikan merupakan ukuran kemajuan suatu bangsa dalam meningkatkan kualitas sumber daya manusia agar dapat bersaing dengan negara lain. Sehingga pendidikan masyarakat menjadi salah satu dari indikator Indeks Pembangunan Manusia (IPM). Pendidikan tertinggi di Indonesia dilaksanakan oleh perguruan tinggi berdasarkan kebudayaan bangsa Indonesia yang mana dilaksanakan setelah masyarakat telah menempuh pendidikan menengah. Pendidikan tertinggi terdiri dari program diploma, program sarjana, program magister, program doktor, program profesi, dan program spesialis. Hal ini sesuai dengan peraturan menteri pendidikan dan kebudayaan Republik Indonesia Nomor 49 Tahun 2014 tentang standar nasional pendidikan tinggi.

Salah satu penilaian dari tujuh standar utama yang dilakukan BAN PT untuk melihat kesuksesan dan kualitas perguruan tinggi menciptakan lulusan adalah ketepatan waktu lulus mahasiswa, dimana mahasiswa akan dikatakan lulus dengan tepat waktu jika dapat menyelesaikan studi dalam kurun waktu 4 tahun atau selama 8 semester untuk jenjang studi S1. Namun pada praktiknya, mahasiswa tidak selalu dapat menyelesaikan studinya tersebut dalam waktu normal. Hal ini ditunjukkan dengan dibaginya masa studi menjadi dua kategori yaitu kurang dari sama dengan delapan semester dan lebih besar dari delapan semester.

Faktor-faktor yang diduga dapat mempengaruhi masa studi mahasiswa tersebut adalah jenis kelamin, asal daerah, jalur masuk, nilai indeks prestasi semester 2, dan fakultas. Sehingga perlu dilakukan kegiatan analisis dan studi kasus untuk melihat klasifikasi ketepatan waktu lulus mahasiswa.

Klasifikasi adalah suatu proses pengelompokan objek berdasarkan kaidah atau standar tertentu yang ditetapkan. Dalam statistika banyak metode yang dapat digunakan dalam mengklasifikasikan suatu objek seperti *decision tree*, *naïve bayes classifier*, *k-nearest neighbor*, *support vector machine*, dan lainnya. Namun dalam beberapa penelitian menghasilkan kesimpulan bahwa metode *support vector machine* merupakan metode yang paling efektif dalam melakukan klasifikasi, seperti pada penelitian yang dilakukan oleh Niwayan Sumartini Saraswati (2013) yang melakukan analisis sentiment dengan menggunakan metode *naïve bayes* dan *support vector machine* dan menghasilkan klasifikasi *support vector machine* lebih baik dibandingkan dengan metode *naïve bayes*.

Berdasarkan deskripsi latar belakang di atas peneliti tertarik untuk melakukan analisis klasifikasi terhadap masa studi mahasiswa di IST AKPRIND dengan tujuan penelitian ini adalah:

1. Untuk mengetahui kelulusan mahasiswa program sarjana IST AKPRIND Yogyakarta Pada Tahun 2012-2017.

2. Untuk menerapkan metode *Support Vector Machine* (SVM) dalam pengklasifikasian dan memprediksi ketepatan waktu kelulusan mahasiswa sarjana IST AKPRIND Yogyakarta.
3. Melihat seberapa akurat penggunaan metode ini dalam pengklasifikasian.
4. Menentukan model terbaik dalam mengklasifikasikan ketepatan waktu kelulusan mahasiswa sarjana IST AKPRIND Yogyakarta.

## 2. LANDASAN TEORI DAN METODE

### 2.1. Klasifikasi

Klasifikasi merupakan pekerjaan memprediksi dan clustering pada data mining yaitu suatu pekerjaan menilai objek data untuk memasukkannya kedalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (1) pembangunan model sebagai prototype untuk disimpan sebagai memory dan (2) penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek data lain agar diketahui dikelas mana objek data tersebut dalam model yang sudah disimpannya.

Menurut Zaki (2014). Klasifikasi adalah proses memprediksi label/kelas yang akan dilakukan pada data yang belum diberikan label/kelas berdasarkan fungsi atau pola yang telah dibentuk, biasanya berdasarkan sifat obyek atau konsep yang bersangkutan. Metode klasifikasi ditujukan untuk pembelajaran fungsi-fungsi berbeda yang memetakan masing-masing data terpilih kedalam salah satu dari kelompok kelas yang telah ditetapkan sebelumnya. Proses klasifikasi didasarkan pada komponen :

#### a. Kelas (*class*)

Variabel dependen dari model yang merupakan kategori variabel yang mewakili label-label yang diletakkan pada obyek setelah pengklasifikasian. Contoh: kelas bintang dan kelas gempa bumi.

#### b. Prediktor (*predictors*)

Variabel independen dari model yang diwakili oleh karakteristik atau atribut dari data yang diklasifikasikan berdasarkan klasifikasi yang dibuat. Contoh: tekanan darah, status perkawinan dan musim.

#### c. Dataset Pelatihan (*training dataset*)

Merupakan dataset yang berisi dua komponen nilai yang digunakan untuk pelatihan untuk mengenali model yang sesuai dengan kelasnya, berdasarkan prediktor yang ada. Contoh: database penelitian gempa, database badai dan database pelanggan supermarket.

#### d. Dataset Pengujian (*testing dataset*)

Merupakan dataset baru yang akan diklasifikasikan oleh model yang dibangun sehingga dapat dievaluasi hasil akurasi klasifikasi tersebut.

Sebuah sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua set data dengan benar, tetapi tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar sehingga sebuah sistem klasifikasi juga harus kinerjanya. Umumnya, pengukuran kinerja klasifikasi dilakukan dengan matriks konfusi (*confussion matrix*).

Matrik konfusi merupakan tabel pencatat hasil kerja klasifikasi. Tabel konfusi dapat dicontohkan pada permasalahan 2 kelas ( 1 dan 0 ) seperti berikut:

**Tabel 2.1. Matrik konfusi untuk klasifikasi dua kelas**

$f_{ij}$	Kelas Hasil Prediksi (j)
----------	--------------------------

		Kelas = 1	Kelas = -1
Kelas Asli (i)	Kelas = 1	$f_{1,1}$	$f_{1,-1}$
	Kelas = -1	$f_{-1,1}$	$f_{-1,-1}$

Berdasarkan isi matriks konfusi, dapat diketahui jumlah data masing-masing kelas yang diprediksi secara benar, yaitu ( $f_{1,1} + f_{-1,-1}$ ), dan data yang diklasifikasikan secara salah, yaitu ( $f_{1,-1} + f_{-1,1}$ ). Kuantitas matriks konfusi dapat diringkas menjadi dua nilai, yaitu akurasi dan laju error. Dengan mengetahui jumlah data yang diklasifikasikan secara benar, dapat dihitung tingkat akurasi hasil klasifikasi, dan dengan mengetahui jumlah data yang diklasifikasikan secara salah dapat dihitung laju error atau tingkat error klasifikasi yang dihasilkan.

Untuk menghitung seberapa besar ketepatan klasifikasi terdapat beberapa metode, salah satunya adalah *Apperant Error Rate* (APER). APER adalah persentase kesalahan yang dikelompokkan salah. APER dihitung berdasarkan persamaan sebagai berikut:

$$APER = \frac{\text{jumlah prediksi yang salah}}{\text{jumlah prediksi yang dilakukan}} \times 100\%$$

$$APER = \frac{F_{1,-1} + F_{-1,1}}{\sum_{j=1}^2 \sum_{i=1}^2 F_{ij}} \times 100\% \quad (2.1)$$

Untuk menghitung seberapa besar ketepatan klasifikasi dengan menggunakan *Correct Classification Rate* (CCR). CCR merupakan persentase ketepatan nilai amatan dan dugaannya, CCR di hitung dengan persamaan sebagai berikut :

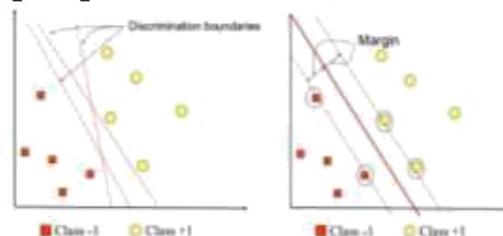
$$CCR = \frac{\text{jumlah prediksi yang tepat}}{\text{jumlah prediksi yang dilakukan}} \times 100\%$$

$$CCR = \frac{F_{1,1} + F_{-1,-1}}{\sum_{j=1}^2 \sum_{i=1}^2 F_{ij}} \times 100\% \quad (2.2)$$

## 2.2. Support Vector Machines (SVM)

Teknik SVM digunakan untuk mendapatkan fungsi pemisah (*hyperplane*) yang optimal untuk memisahkan observasi yang memiliki nilai variabel target yang berbeda (Cortes, 2011). *Hyperplane* ini dapat berupa line pada two dimension dan dapat berupa flat plane pada multiple dimension.

Saat ini *SVM* telah berhasil diaplikasikan dalam problem dunia nyata dan secara umum memberikan solusi yang lebih baik dibandingkan metode konvensional. proses cara kerja *SVM* dapat dilihat pada gambar berikut:



Gambar 2.1. Proses Kerja SVM

Pada gambar diatas memperlihatkan bagaimana konsep dasar dari SVM, penyebaran data ditunjukkan oleh warna merah (kotak) dan warna kuning (lingkaran). Data berwarna merah merupakan anggota dari kelas -1 dan data berwarna kuning

merupakan anggota dari kelas. Masalah utama dari klasifikasi adalah mencari *hyperplane* pemisah antara kedua kelas.

Misalkan data yang tersedia direpresentasikan dalam bentuk vektor :

$$\vec{d} := \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\},$$

Dengan  $X_i \in R$  dan  $y_i \in \{-1, 1\}$

Diasumsikan data tersebut terpisah secara sempurna kedalam kelas -1 dan 1 oleh *hyperplane*, yang didefinisikan (Zaki, 2014):

$$\mathbf{w}^T \cdot \mathbf{x} + b = 0 \quad (2.3)$$

Dimana  $\mathbf{w}$  merupakan matrik vektor normal pada *hyperplane* tersebut, dan  $b$  adalah jarak dari *hyperplane* ke titik pusat. Sehingga data yang termasuk kelas 1 dan -1 adalah data yang memenuhi persamaan :

$$y_i = 1, \text{ jika } \mathbf{w}^T \cdot \mathbf{x} + b > 0$$

$$y_i = -1, \text{ jika } \mathbf{w}^T \cdot \mathbf{x} + b < 0$$

*Support vector* adalah vektor terluar dari masing-masing kelas yang terdekat dengan *hyperplane* pemisah, maka fungsi keputusan (2.4) yang digunakan dalam klasifikasi :

$$f(x, y) = \text{sign}(\mathbf{w}^T \cdot \mathbf{x} + b) \quad (2.4)$$

Dengan

$\mathbf{w}^T$  = Parameter model (vektor normal pada *hyperplane*)

$b$  = Jarak *hyperplane* ke titik pusat

*Sign* = Notasi + / -

Jika nilai  $f(x, y) > 0$  maka data akan dimasukkan ke kelas 1, sedangkan jika  $f(x, y) < 0$  maka data akan dimasukkan ke kelas -1.

### 2.3. Support Vector Machine Linier Separable

Data dapat dinyatakan terpisah sempurna secara linier oleh *hyperplane*, jika didapat pasangan nilai  $(\mathbf{w}, b)$  sedemikian hingga data dapat dipisahkan ke dalam kelas 1 dan kelas -1, yang memenuhi kondisi berikut :

$$\mathbf{w}^T \cdot \mathbf{x} + b \leq -1, \text{ jika } y_i = -1$$

$$\mathbf{w}^T \cdot \mathbf{x} + b \geq 1, \text{ jika } y_i = 1$$

Kedua kondisi diatas dapat disederhanakan :

$$y_i(\mathbf{w}^T \cdot \mathbf{x} + b) \geq 1, i = 1, 2, 3, \dots, N \quad (2.5)$$

Terdapat banyak pasangan yang dapat membentuk *hyperplane*, dalam hal ini *hyperplane* yang dicari adalah yang paling optimal, yang memiliki margin maksimum antar kelas. Margin didefinisikan sebagai  $d = d_1 + d_2$ , sehingga margin akan memiliki nilai maksimal jika  $d_1 = d_2$

$$d = d_1 + d_2 = \frac{1}{\|\mathbf{w}\|} (|\mathbf{w}^T \cdot \mathbf{x}_1 + b| |\mathbf{w}^T \cdot \mathbf{x}_2 + b|) = \frac{2}{\|\mathbf{w}\|} \quad (2.6)$$

Berdasarkan persamaan diatas, maka mencari margin maksimal sama dengan meminimumkan nilai  $\|\mathbf{w}\|^2$ , secara matematis dinyatakan sebagai :

$$\text{Min } \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.7)$$

$$y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) \geq 1, i = 1, 2, 3, \dots, N$$

Dengan metode lagrange, masalah klasifikasi tersebut dapat dinyatakan sebagai masalah minimumkan fungsi *Lagrange*.

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i y_i (\mathbf{w}^T \cdot \mathbf{x}_i + b) - \sum_{i=1}^N \alpha_i \quad (2.8)$$

Dimana  $\alpha_i$  adalah pengali *Lagrange*

Karena,

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^N \alpha_i y_i = 0$$

Maka masalah *Lagrange* untuk klasifikasi dapat dinyatakan :

$$\text{Min } L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i y_i ((\mathbf{x}_i \cdot \mathbf{w}^T + b)) - \sum_{i=1}^N \alpha_i \quad (2.9)$$

#### 2.4. Support Vector Machine Non-Linier Separable

Dalam masalah klasifikasi kebanyakan sampel data tidak terpisah secara linier sehingga jika digunakan SVM linier maka hasil yang diperoleh tidak optimal dan mengakibatkan hasil klasifikasi yang buruk. SVM linier dapat diubah menjadi SVM non-linier dengan menggunakan metode kernel. Metode ini bekerja dengan cara memetakan data input ke ruang *feature* yang dimensinya lebih tinggi. Diharapkan data input hasil pemetaan ke ruang *feature* akan terpisah secara linier sehingga dapat dicari *hyperplane* yang optimal Tomasouw(2012).

Fungsi kernel dapat menghitung inner product pada feature space secara langsung dari ruang input tanpa secara eksplisit menghitung kordinat proyeksi masing-masing vektor input pada feature space.

Inner product adalah operasi yang sangat penting karena sangat erat kaitanya dengan persoalan geometri dari data ruang fitur, seperti untuk menghitung jarak. Beberapa fungsi kernel pada SVM Karatzoglu (2004) :

**Tabel 2.2. Jenis-jenis Kernel Pada SVM**

Nama Kernel	Definisi Fungsi
Linear	$K(x, x') = x \cdot x'$
Polinomial	$K(x, x') = (scale \cdot x \cdot x' + offset)^d$
Gaussian RBF	$K(x, x') = \exp(-\sigma \ x - x'\ ^2)$

Pada masing-masing kernel pada tabel 2.4, nilai parameter  $d$ ,  $scale$ ,  $offset$ ,  $\sigma > 0$ .

#### 2.5. Sumber Data dan Variabel

Data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data yang diperoleh dari Badan Perencanaan, Pengembangan dan Pelayanan Sistem Informasi (BP3SI) yang merupakan salah satu lembaga pelaksana akademis Institut Sains & Teknologi AKPRIND berupa data yang sudah diolah dan dapat menjadi sumber informasi.

Buku Wisuda Sarjana dan Ahli Madya serta website resmi Institut Sains & Teknologi (IST) AKPRIND menjadi pendukung penulis untuk melakukan penelitian ini. Dalam penelitian ini yang akan diteliti oleh penulis adalah klasifikasi masa studi mahasiswa program sarjana Institut Sains & Teknologi AKPRIND tahun 2012-2017.

Adapun variabel-variabel dalam penelitian ini adalah :

No	Variabel	Tipe	Keterangan
Variabel Dependen			
1	Masa Studi	Nominal	1 = Tepat Waktu -1 = Tidak Tepat Waktu

Variabel Independen			
2	Jenis Kelamin	Nominal	1 = Laki-laki -1 = Perempuan
3	Asal Daerah	Nominal	1 = Dalam Daerah Yogyakarta -1 = Luar Daerah Yogyakarta
4	Jalur Masuk	Nominal	1 = Tanpa Test 2 = Camabapress 3 = PMB 4 = Reguler

**Variabel Penelitian**

**Tabel 2.3. Variabel Penelitian (Lanjutan)**

No	Variabel	Tipe	Keterangan
5	Nilai Indek Prestasi Semester 2	Interval	0-4
6	Asal Fakultas	Nominal	1 = FST (Fakultas Sain Terapan) 2 = FTM (Fakultas Teknologi Mineral) 3 = FTI (Fakultas Teknologi Industri)

Langkah-langkah analisis dalam penelitian ini yaitu:

1) Melakukan analisis klasifikasi dengan SVM (*Support Vector Machine*) dengan tahapan sebagai berikut :

- a. Memasukkan data sesuai format software.
- b. Membagi data training dan data testing dengan proporsi data training sebesar 90 (944 mahasiswa)% dari populasi sebanyak 1072 mahasiswa dan sisanya sebesar 10% (128 mahasiswa) sebagai data testing.

- c. Data training dan data testing merupakan data lulusan mahasiswa tahun 2012-2017, Sedangkan data untuk ramalan adalah data mahasiswa tahun akademik 2016.
  - d. Menentukan fungsi kernel terbaik yang akan digunakan sebagai permodelan *hyperplane* SVM yaitu kernel Polynomial, Linier, RBF Gaussian, Analisis dilakukan pada data training.
  - e. Menentukan nilai parameter C dan nilai-nilai parameter kernel yang akan digunakan sebagai permodelan *hyperplane* SVM . Penentuan nilai adalah dengan menggunakan grid search dan bernilai positif, semakin besar nilai C akan mempengaruhi pelanggaran yang akan dikenakan untuk tiap klasifikasi. Untuk nilai parameter kernel linier C akan digunakan adalah 0.1, 0.5, 1, 5, 10, 50, 100, untuk parameter kernel polynomial akan digunakan nilai degree = 2, 3, 4, 5. Dan untuk parameter kernel rbf gaussian nilai sigma akan digunakan adalah 0.1, 0.2, 0.3, 0.4, 0.5.
  - f. Mendapatkan nilai alpha dan b
  - g. Membentuk persamaan *hyperplane*
  - h. Melakukan prediksi klasifikasi pada data testing.
- 2) Evaluasi performansi model klasifikasi menggunakan matriks konfusi.
    - a. Menghitung akurasi klasifikasi hasil prediksi berdasarkan nilai APER dan nilai CCR.
    - b. Memilih nilai parameter dan fungsi kernel terbaik
  - 3) Melakukan peramalan dengan menggunakan data mahasiswa tahun akademik 2016.

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Analisis Deskriptif

Lulusan IST AKPRIND Yogyakarta pada periode 2012-2017 adalah sebanyak 1072 lulusan dengan rincian Fakultas Sains dan Terapan (FST) sebanyak 106 lulusan, Fakultas Teknologi Industri (FTI) sebanyak 757 lulusan, dan Fakultas Teknologi Mineral (FTM) sebanyak 208 lulusan. Keadaan lulusan mahasiswa IST AKPRIND sangat kurang seimbang, ini dapat dilihat dari perbandingan lulusan yang menyelesaikan studi tepat waktu ( $\leq 8$  Semester) dengan lulusan yang menyelesaikan studi tidak tepat waktu ( $> 8$  Semester). Jumlah lulusan yang tepat waktu paling kecil terdapat pada Fakultas Teknologi Mineral dengan jumlah 5 lulusan, sedangkan lulusan yang tidak tepat waktu sebanyak 203 lulusan, hal serupa juga dialami oleh Fakultas Teknologi Industri yaitu jumlah lulusan tepat waktu hanya sebanyak 102 lulusan dari 757 lulusan, dan jumlah lulusan yang tidak tepat waktu sebanyak 655 lulusan, Fakultas Sains dan Terapan juga memiliki lulusan dengan tepat waktu sedikit yaitu sebanyak 29 lulusan, dan lulusan yang tidak tepat waktu sebanyak 77 lulusan.

Berdasarkan jenis kelamin Fakultas Teknologi Industri mempunyai jumlah lulusan berjenis kelamin laki-laki paling banyak, yaitu sebanyak 640 lulusan, sedangkan lulusan yang berjenis kelamin perempuan hanya sebanyak 117 lulusan. Keadaan berbeda terjadi di Fakultas Sains dan Terapan yang memiliki proporsi lulusan yang hampir seimbang antara laki-laki dan perempuan, jumlah lulusan yang berjenis kelamin laki-laki sebanyak 50 lulusan dan perempuan sebanyak 56 lulusan. Sedangkan pada Fakultas Teknologi Mineral lulusan berjenis kelamin laki-laki juga

lebih mendominasi daripada lulusan berjenis kelamin perempuan dengan jumlah laki-laki sebanyak 170, dan perempuan sebanyak 38.

Berdasarkan asal daerah Pada Fakultas Sains dan Terapan, proporsi sedikit lebih seimbang antara mahasiswa yang berasal dari dalam daerah dengan mahasiswa yang berasal dari luar daerah yaitu dari dalam daerah sebanyak 47 dan dari luar daerah sebanyak 59. Sedangkan pada Fakultas Teknik Industri jumlah mahasiswa dari dalam daerah sebanyak 169 mahasiswa dan dari luar daerah sebanyak 588 mahasiswa. Pada Fakultas Teknologi Mineral jumlah lulusan dari dalam daerah sebanyak 17 mahasiswa dan yang berasal dari luar daerah sebanyak 191 mahasiswa.

Untuk jalur masuk Dari total 106 mahasiswa yang lulus dari Fakultas Sains dan Terapan, 33 diantaranya merupakan mahasiswa yang masuk dengan jalur masuk tanpa test, 16 mahasiswa merupakan mahasiswa jalur masuk camabapres, 46 mahasiswa merupakan mahasiswa dengan jalur masuk PMB dan sisanya sebanyak 11 mahasiswa merupakan mahasiswa dengan jalur masuk reguler. Pada Fakultas Teknologi Industri, sebanyak 346 mahasiswa dari total 757 mahasiswa merupakan lulusan dengan jalur masuk tanpa test, 136 mahasiswa menempuh kuliah dengan jalur masuk camabapres, 228 mahasiswa merupakan mahasiswa dengan jalur masuk ke IST AKPRIND dengan cara PMB, dan sisanya sebanyak 47 mahasiswa merupakan mahasiswa dengan jalur masuk Reguler. Dan pada Fakultas Teknologi Mineral, sebanyak 120 mahasiswa merupakan mahasiswa dengan jalur masuk tanpa test, 52 mahasiswa dengan jalur masuk camabapres, 34 mahasiswa adalah mahasiswa dengan jalur masuk PMB, dan hanya 21 orang mahasiswa yang merupakan mahasiswa reguler.

Indeks Prestasi Semester 2 pada masing masing fakultas sangat berbeda, pada Fakultas Sains Dan Terapan nilai Indeks Prestasi Semester 2 maksimum yaitu 4, nilai minimum dari seluruh mahasiswa yaitu 2,27 dengan rata-rata dari keseluruhan mahasiswa lulusan dari Fakultas Sains dan Terapan yaitu 3,321, Pada Fakultas Teknologi Mineral nilai maksimum dari Indeks Prestasi Semester 2 yang didapat dari seluruh lulusan adalah 3,55, nilai minimum Indeks Prestasi Kumulati yaitu 0,75 dengan rata-rata keseluruhan 2,80. Sedangkan pada Fakultas Teknologi Industri nilai maksimum yang diraih mahasiswa lulusannya yaitu 4 dengan nilai minimum 0,75 dan rata-rata keseluruhan yaitu 2,82.

### **3.2. Analisis Support Vector Machines**

#### **3.2.1. Analisis Support Vector machine Kernel Linier**

Analisis SVM pada kernel linier digunakan beberapa parameter cost yaitu 0.1, 0.5, 1, 5, 10, 50, 100. Parameter ini diterapkan pada data training sehingga didapatkan error klasifikasi sebagai berikut :

**Tabel. 2.4. Nilai Error Klasifikasi Pada Model Kernel Linier**

Nilai Cost	Error	Jumlah Support Vector
1	0.126193	242
5	0.126193	252
10	0.126193	258
50	0.126193	263
100	0.126193	271

Berdasarkan tabel 2.4 diatas didapat model terbaik yaitu pada nilai parameter cost 0.1, model ini diterapkan pada data testing sehingga didapatkan matrik konfusi sebagai berikut :

**Tabel 2.5. Tabel Matrik Konfusi Model Klasifikasi Kernel Linier**

		Ramalan	
		Tepat Waktu (1)	Tidak Tepat Waktu (-1)
Data Aktual	Tepat Waktu (1)	0	17
	Tidak Tepat Waktu (-1)	0	111

Selanjutnya kita lakukan pengujian ketepatan model dengan menghitung nilai APER dan nilai CCR.

$$APER = \frac{0 + 17}{128} \times 100\% = 13,28 \%$$

$$CCR = \frac{0 + 111}{128} \times 100\% = 86,71\%$$

Berdasarkan matrik konfusi diatas dapat disimpulkan bahwa dengan menggunakan kernel trick linier. Model yang digunakan dapat dipercaya sebesar 86,71% dengan tingkat kesalahan sebesar 13,28%.

### 3.2.2. Analisis Support Vector Machine Kernel Polynomial

Analisis SVM pada kernel polynomial digunakan beberapa parameter cost yaitu 0.1, 0.5, 1, 5, 10, 50, 100. Dan parameter degree 2, 3, 4, 5. Parameter ini diterapkan pada data training sehingga didapatkan error klasifikasi sebagai berikut :

**Tabel 2.6. Nilai Error Klasifikasi Pada Model Kernel Polynomial**

Nilai C	Degree	Error	Jumlah Support Vector
0.1	2	0.126193	286
	3	0.121951	265
	4	0.111347	299
	5	0.101803	288
0.5	2	0.126193	307
	3	0.121951	265
	4	0.107105	324
	5	0.099682	278
1	2	0.126193	323
	3	0.121951	265
	4	0.107105	323
	5	0.099682	275
5	2	0.126193	369
	3	0.121951	266
	4	0.121951	277

	5	0.098621	289
10	2	0.126193	395
	3	0.126951	265
	4	0.123012	291
	5	0.115589	288
	50	2	0.126193
50	3	0.133619	272
	4	0.14316	304
	5	0.170732	287
100	2	0.126193	320
	3	0.230117	284
	4	0.199364	285
	5	0.181336	297

Berdasarkan tabel 2.6 diatas didapat model terbaik yaitu pada nilai parameter cost 5 degree 5, model ini diterapkan pada data testing sehingga didapatkan matrik konfusi sebagai berikut :

**Tabel 2.7. Tabel Matrik Konfusi Model Klasifikasi Kernel Polynomial**

		Ramalan	
		Tepat Waktu (1)	Tidak Tepat Waktu (-1)
Data Aktual	Tepat Waktu (1)	0	17
	Tidak Tepat Waktu (-1)	6	105

Selanjutnya kita lakukan pengujian ketepatan model dengan menghitung nilai APER dan nilai CCR.

$$APER = \frac{6 + 17}{128} \times 100\% = 14,84 \%$$

$$CCR = \frac{0 + 105}{128} \times 100\% = 85,15\%$$

Berdasarkan matrik konfusi diatas dapat disimpulkan bahwa dengan menggunakan kernel trick polynomial. Model yang digunakan dapat dipercaya sebesar 85,15% dengan tingkat kesalahan sebesar 14,84%.

### 3.2.3. Analisis Support Vector Machine Kernel RBF Gaussian

Analisis SVM pada kernel RBF Gaussiidigunakan beberapa parameter cost yaitu 0.1, 0.5, 1, 5, 10, 50, 100. Dan parameter sigma 0.1, 0.2, 0.3, 0.4, 0.5. Parameter ini diterapkan pada data training sehingga didapatkan error klasifikasi sebagai berikut :

**Tabel 2.8. Nilai Error Klasifikasi Pada Model Kernel RBF Gaussian**

Nilai Cost	Gamma	Error	Jumlah Support Vector
0.1	0.1	0.126193	286
	0,2	0.126193	303

	0.3	0.126193	294
	0.4	0.126193	310
	0.5	0.126193	307
0.5	0.1	0.126193	298
	0,2	0.126193	299
	0.3	0.126193	342
	0.4	0.126193	328
	0.5	0.126193	330
1	0.1	0.126193	305
	0,2	0.126193	316
	0.3	0.124072	332
	0.4	0.121951	347
	0.5	0.121951	351
5	0.1	0.126193	295
	0,2	0.121951	327
	0.3	0.120891	348
	0.4	0.115589	333

**Tabel. 2.8. Nilai Error Klasifikasi Pada Model Kernel RBF Gaussian (Lanjutan)**

Nilai Cost	Gamma	Error	Jumlah Support Vector
	0.5	0.112407	309
10	0.1	0.120891	292
	0,2	0.116649	328
	0.3	0.115589	339
	<b>0.4</b>	<b>0.08165</b>	<b>298</b>
	0.5	0.107105	300
50	0.1	0.11877	303
	0,2	0.111347	329
	0.3	0.108165	291
	0.4	0.106045	282
	0.5	0.101803	294
100	0.1	0.115589	315
	0,2	0.108165	298
	0.3	0.104984	285
	0.4	0.101803	286

Berdasarkan tabel 2.6 diatas didapat model terbaik yaitu pada nilai parameter cost 10 sigma 0.4, model ini diterapkan pada data testing sehingga didapatkan matrik konfusi sebagai berikut :

**Tabel 2.9. Tabel Matrik Konfusi Model Klasifikasi Kernel RBF Gaussian**

		Ramalan	
		Tepat Waktu (1)	Tidak Tepat Waktu (-1)
Data Aktual	Tepat Waktu (1)	0	17
	Tidak Tepat Waktu (-1)	2	109

Selanjutnya kita lakukan pengujian ketepatan model dengan menghitung nilai APER dan nilai CCR.

$$APER = \frac{2 + 17}{128} \times 100\% = 17,96 \%$$

$$CCR = \frac{0 + 109}{128} \times 100\% = 82,03\%$$

Berdasarkan matrik konfusi diatas dapat disimpulkan bahwa dengan menggunakan kernel trick RBF Gaussian. Model yang digunakan dapat dipercaya sebesar 82,03% dengan tingkat kesalahan sebesar 17,96%.

### 3.3. Pemilihan Model Terbaik

Dengan membandingkan hasil error pada setiap model yang diterapkan pada data testing, akan dipilih model yang menghasilkan nilai error paling rendah dan akan digunakan sebagai model untuk melakukan peramalan masa studi pada data mahasiswa tahun akademik 2016. Perbandingan nilai error masing masing model dapat dilihat pada Tabel dibawah ini :

### 2.10. Perbandingan Hasil Error Pada Data Test

Parameter	Kernel		RBF Gaussian
	Linier	Polynomial	
APER	13.28	14.84	17.96
CCR	86.71	85.15	82.03

Berdasarkan Tabel 2.10, model klasifikasi dengan kernel linier menghasilkan nilai APER yang paling rendah dengan parameter cost sebesar 0,1.

## 4. KESIMPULAN DAN SARAN

Dari hasil dan pembahasan dapat disimpulkan :

1. Berdasarkan data kelulusan tahun 2012-2017 diketahui bahwa sebanyak 136 mahasiswa (12.69%) mahasiswa menepuh studi dengan tepat waktu, dan sisa sebanyak 935 mahasiswa (87.31%) menempuh studi dengan tidak tepat waktu.
2. Analisis Kalsifikasi pada masing-masing kernel adalah sebagai berikut:
  - a. Pada analisis klasifikasi *Support Vector Machine* dengan menggunakan kernel linier didapatkan model terbaik pada data training yaitu pada nilai parameter cost sebesar 0.1 yang menghasilkan nilai error klasifikasi sebesar 0.126193.
  - b. Klasifikasi *Support Vector Machine* dengan menggunakan kernel polynomial, model terbaik untuk data training yaitu pada nilai parameter cost sebesar 5 dengan nilai parameter degree sebesar 5. Model ini menghasilkan nilai error klasifikasi sebesar 0.098621.
  - c. Kernel RBF Gaussian pada klasifikasi *Support Vector Machine* menghasilkan model terbaik pada data training yaitu pada nilai parameter cost sebesar 10 dan parameter sigma sebesar 0.4. Model ini menghasilkan nilai error klasifikasi sebesar 0.08165.
3. Adapun nilai APER dan CCR masing-masing kernel pada pengujian dengan menggunakan data testing adalah sebagai berikut:
  - a. Uji coba penggunaan model pada data testing menghasilkan nilai APER sebesar 0.1328 dengan tingkat akurasi (CCR) klasifikasi sebesar 0.8671 atau 86,71%.
  - b. Uji Coba penggunaan model pada data testing menghasilkan nilai APER sebesar 0.1484 dengan tingkat akurasi (CCR) klasifikasi sebesar 0.8515 atau 85,15%.
  - c. Uji coba penggunaan model pada data testing menghasilkan nilai APER sebesar 0.1796 dengan tingkat akurasi (CCR) klasifikasi sebesar 0.8203 atau 82,03%.

Dari ketiga kernel yang digunakan pada kasus masa studi mahasiswa tahun angkatan 2012-2017 dapat disimpulkan bahwa model terbaik menurut nilai error paling rendah pada pengujian data testing adalah model klasifikasi dengan penggunaan kernel linier dengan nilai cost sebesar 0,1. Model ini menghasilkan ramalan bahwa pada mahasiswa tahun angkatan 2016 akan menempuh masa studi lebih dari 8 semester (4 tahun) atau tidak tepat waktu

#### **DAFTAR PUSTAKA**

- Cortes, Vapnik, 1995, *Support Vector Network*, Machine Learning, 20, 273-297, Kluwer Academic Publisher, Boston.
- Karatzoglou, Alexandros, 2004, *Kernlab-An S4 Package for Kernel Methode in R*, Jurnal Of Statistic Software.
- Prasetyo, Eko , 2012, *Data Mining : Konsep dan Aplikasi Menggunakan MATLAB*, ANDI Offset, Yogyakarta.
- Sarawati, Niwayan Sumartini, 2013, *Naïve Bayes Classifier dan Support Vector Machine Untuk Sentimen Analisis*, Jurnal Seminar Nasional Sistem Informasi Indonesia, 2-4 Desember 2013. Denpasar, Bali.
- Zaki, 2014, *Data Mining And Analysis : Fundamental Concepts And Algorithms*, Cambridge University Press, 32 Avenue Of The America, New York.