

# Prediksi Kinerja Akademik Mahasiswa Menggunakan Machine Learning dengan Sequential Minimal Optimization untuk Pengelola Program Studi

Andi Iwan Nurhidayat<sup>1</sup>, Asmunin<sup>2</sup>, Dwi Fatrianto<sup>3</sup>

<sup>1,2,3</sup> Afiliasi (Jurusan Teknik Informatika, Universitas Negeri Surabaya)

<sup>1</sup>[andy134k5@unesa.ac.id](mailto:andy134k5@unesa.ac.id)

**Abstrak**— Kinerja akademik mahasiswa menjadi salah satu tolok ukur keberhasilan pengelolaan program studi. Kinerja akademik mahasiswa dapat dilihat pada nilai indek kumulatif (IPK) dan kelulusan tepat waktu. Penelitian ini dilakukan untuk membuat sebuah pemodelan yang bertujuan untuk Memprediksi Kinerja Mahasiswa dalam mengelola Program Studi di Universitas Negeri Surabaya pada Program Studi Teknik Informatika dan Sistem Informasi. Data yang digunakan pada penelitian ini berupa data mahasiswa Jurusan Teknik Informatika UNESA sejumlah 330 data. Metode *Sequential Minimal Optimization* (SMO) digunakan untuk membuat pemodelan yang digunakan untuk peramalan unjuk kerja akademik. Uji coba sistem diimplementasi menggunakan metode Percentage split 80% dengan mengukur keauratan output yang berupa kurasi, presisi dan *recall*. Berdasarkan pengujian yang telah dilakukan, penerapan SMO mendapatkan output perhitungan akurasi=93,94%, presisi=94,7% dan *recall*=94,7%

**Kata Kunci**— SVM, SMO, Akurasi, Presisi, Recall, Prediksi.

## I. PENDAHULUAN

Kinerja akademik mahasiswa umumnya dilihat dari index prestasi mahasiswa (IPK) dan kelulusan tepat waktu. Penelitian ini sangat berguna untuk dijadikan bahan pertimbangan untuk melihat unjuk kerja akademik mahasiswa yang ada. Untuk itu, penelitian ini mengajukan pemebuatan model peramalan unjuk kerja mahasiswa yang berguna dan bermanfaat untuk pengelolaan Universitas Negeri Surabaya (UNESA) umumnya dan di jurusan Teknik informatika FT Unesa. Untuk masalah prediksi selayaknya digunakan sebuah metode yang mampu mengklasifikasikan data sehingga *output* akhirnya yaitu *forecasting* bisa tercapai. Salah satu teknik yang akan diusulkan dalam mengelola informasi tersebut adalah dengan mengumpulkan segala informasi yang terdapat pada data akademik mahasiswa.

Pengelolaan atau pemrosesan data akademik mahasiswa dilakukan dengan tujuan mencapai prediksi yang diinginkan. Terdapat beberapa penelitian telah diimplementasikan diantaranya adalah sebagai berikut Penelitian yang dilakukan oleh David Hartanto Kamagi tentang implementasi data *mining* menggunakan Algoritma C4.5 (*Decision Tree*) yang bertujuan untuk meramalkan prosentase lulusan Mahasiswa. Adapun hasilnya adalah sebuah aplikasi *desktop* yang menghasilkan akurasi mahasiswa yang lulus dengan prosentase 87,5% dari 60 data *training* sedangkan data *testing* yang digunakan berjumlah 40 (Kamagi, 2014). Penelitian selanjutnya, dilakukan oleh Mustafa, yaitu membuat sebuah model yang diperuntukkan

untuk mahasiswa baru, yang bertujuan untuk meramalkan kemungkinan penyelesaian masa *study* tepat waktu dengan mengimplementasikan *K-Nearest Neighbor* (K-NN) *Algorithm*. Data yang digunakan untuk uji coba dengan mengimplementasikan algoritma K-NN adalah data lulusan yang telah wisuda pada tahun 2004 sampai dengan 2010 untuk kasus lama dan data lulusan yang wisuda pada tahun 2011 untuk kasus baru. Dari uji coba tersebut memperoleh keakuratan sebesar 83,36%. (Mustafa, 2014). Walaupun penelitian-penelitian sebelumnya sudah mencapai akurasi yang tinggi diatas 80%, terdapat kelemahan yang bisa disimpulkan yaitu waktu pemrosesan data yang kurang efisiensi.

Untuk data set memiliki ukuran yang cukup besar dengan atribut independen yang banyak menjadikan pengolahan data kurang efisien. Disamping itu terdapat kelemahan lain yaitu keakuratan yang diperoleh dari hasil uji coba masih rendah. Untuk itu peneliti mengusulkan penelitian menggunakan metode *Sequential Minimal Optimization* (SMO) yang diharapkan mendapat output yang lebih baik.

SVM merupakan metode pengelompokan yang mempunyai keakuratan yang tinggi dalam proses peramalan dan pengelompokan karakteristik dari data ada pada data-data akademik. Ini banyak dibuktikan pada beberapa permasalahan yang berkaitan dengan prediksi. Pada beberapa kasus SVM menghasilkan *ouput* yang lebih akurat dari pada metode sejenisnya, terutama pencapaian penyelesaian *ouput* yang diharapkan. SVM memberikan penyelesaian yang optimal secara *global*, sedangkan metode lain memiliki penyelesaian optimal yang bersifat *local* (Santosa, 2007). Adapun solusi optimal yang bersifat *local* atau disebut sebagai *local optimal* adalah penyelesaian yang dapat dituju oleh sebuah algoritma namun masih dalam *range* nilai tertentu yang dibatasi, sedangkan *global optimal* berarti penyelesaian yang diharapkan masih dalam *range* input data secara keseluruhan. Disamping itu SVM juga mempunyai kemampuan menyelesaikan permasalahan yang memiliki dimensi tinggi walaupun dengan data yang terbatas (Zhang, 2010).

Metode Multi-class SVM cocok digunakan pada permasalahan data yang bersifat geografik dan data mahasiswa yang bersifat non-linier yang digunakan pada penelitian ini. Penggunaan data tidak linier harus digandengkan dengan *Kernel Function* untuk menghasilkan *ouput* pengelompokan yang terbaik. Pengujian data dengan menggunakan teknik *Percentage Spli* sehingga penggunaan teknik tersebut diharapkan meberikan output yang diharapkan berupa

peramalan unjuk kerja akademik mahasiswa dengan keakuratan yang tinggi.

Berdasarkan keterangan-keterangan yang telah diuraikan pada pendahuluan maka rumusan masalah dalam penelitian ini adalah Bagaimana penerapan algoritma *Sequential Minimal Optimization* untuk meramalkan unjuk kerja mahasiswa. Bagaimana keakuratan hasil uji coba prediksi unjuk kerja mahasiswa menggunakan *Sequential Minimal Optimization* yang nantinya akan dibandingkan dengan data aktualnya.

Setelah merumuskan permasalahan seperti keterangan diatas, maka dapat ditentukan tujuan dilakukan penelitian ini. Adapun tujuan tersebut adalah sebagai berikut Membuat sebuah model prediksi unjuk kerja akademik yang dapat dijadikan sebagai alat dalam mengevaluasi prestasi mahasiswa yang berguna dalam pengelolaan jurusan Teknik Informatika di UNESA dengan algoritma *Sequential Minimal Optimization*. Melakukan evaluasi terhadap penggunaan algoritma *Support Vector Machine* dengan *Sequential Minimal Optimization* dalam melakukan peramalan unjuk kerja mahasiswa dengan mengukur keakuratan hasil ramalan menggunakan teknik *Percentage split*.

## II. TINJAUAN PUSTAKA

### A. Machine Learning

Menurut Gotama (2018) (link) *Machine Learning* (ML) adalah teknik untuk melakukan inferensi terhadap data dengan pendekatan matematis. Pada intinya *machine learning* merupakan teknik yang digunakan untuk membuat model (matematis) yang menggambarkan *patern-patern* pada data. *Inferensi* yang dimaksud pada *machine learning* lebih memfokuskan relasi antar atribut. Selain itu, *machine learning* merupakan bentuk penggambaran data/ilmu/pengetahuan dalam model rumusan matematika. Disebut model matematis karena keilmuan *machine learning* merupakan turunan dari rumusan matematika dan statistika. *Machine learning* ibarat sebuah “alat”, yang identik dengan rumus matematika. Adapun cara penggunaannya tergantung pada daerah permasalahan yang dihadapi. Penggunaan *machine learning* memiliki minimal 2 tujuan: meramalkan masa depan (unobserved event); dan/atau memperoleh ilmu pengetahuan (*knowledge discovery/discovering unknown structure*).

Dalam *machine learning* terdapat 2(dua) metode pembelajaran yang paling banyak digunakan. 2 metode pembelajaran tersebut adalah pembelajaran yang diawasi (*Supervised learning*) dan pembelajaran yang tidak terawasi (*Unsupervised learning*). *Supervised learning* merupakan sebuah sistem yang dilatih untuk mendapatkan tujuan output seperti (kelas/label) yang diharapkan, pada kasus ini output yang ingin dituju telah ditetapkan kriterianya. Ada beberapa langkah yang dilakukan untuk pembuatan *machine learning* langkah awal adalah menentukan Algoritma pembelajaran yang akan digunakan oleh mesin/sistem yang dapat memasukan satu rangkaian input bersamaan dengan kelompok output yang bernilai *true* dan sesuai, langkah kedua adalah algoritma yang telah ditetapkan diberi waktu untuk belajar atau mempelajari

sejumlah data (umumnya diistilahkan dengan training data), selanjutnya setelah mendapatkan output, output pembelajaran tersebut dibandingkan dengan output aktual untuk menemukan kesalahan. Setelah mendapatkan tingkat kesalahan (error rate) maka mode tersebut dimodifikasi sesuai kebutuhan dengan error rate yang semakin kecil. Metode yang digunakan untuk mencapai tujuan tersebut adalah pengelompokan, regresi, dan peramalan.

*Supervised learning* menggunakan *pattern* berguna untuk meramalkan tingkatan kelas pada data baru yang tidak memiliki label. Adapun data baru tersebut harus memiliki jumlah atribut yang sama dengan data sumber atau awal. *Supervised learning* umumnya diimplementasikan pada aplikasi yang memiliki data historis yang digunakan untuk meramalkan kemungkinan-kemungkinan yang terjadi di masa yang akan datang. Contoh penerapan metode *Supervised learning* adalah menggunakan untuk mencegah beberapa kemungkinan terjadinya penipuan atau pelanggaran pada transaksi kartu kredit dan pelanggan asuransi yang akan mengajukan klaim.

*Unsupervised Learning* diimplementasikan terhadap data yang tidak memiliki label historis. Sistem yang dibangun dengan *Supervised learning* tidak diberi tahu mana “jawaban yang benar” mana “jawaban yang salah”. Sehingga Algoritma tersebut akan mencari tahu sendiri data apa yang ingin ditunjukkan. Penggunaan metode *Unsupervised learning* bertujuan mendalami data dan mengidentifikasi struktur-struktur yang ada di dalam data tersebut. Sebagai contoh *Unsupervised Learning* umumnya digunakan untuk pengelompokan atau klasifikasi kelas. Pembelajaran jenis ini, bertujuan mencari kesamaan karakter dalam data latih yang diberikan. Misalnya, pengklasifikasian individu berdasarkan status sosial yang dapat mengelompokkan orang yang berada di satu kelompok dan orang kurang mampu dikelompok lain.

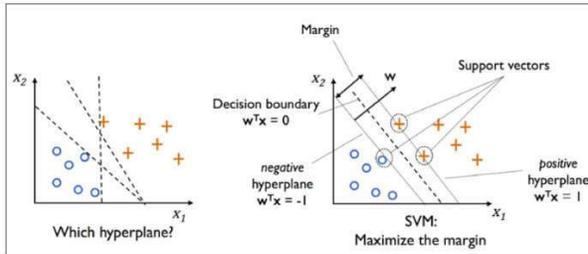
Untuk memahami struktur data dengan baik sebuah *machine learning* membutuhkan pendekatan model secara statistik, kemudian model tersebut menyesuaikan dengan data yang dipahami dengan baik yang telah terdistribusi secara teoretis. Akhirnya, dengan rumusan statistik menghasilkan teori yang terbukti benar secara matematika, selain itu data yang diberikan harus memenuhi asumsi-asumsi yang diinginkan. Selanjutnya pada proses pengujian output perhitungan dari hasil *testing* pada model yang di hasilkan oleh algoritma *machine learning* berbentuk keakuratan nilai akurasi atau nilai kesalahan (error). Proses belajar mesin sering diimplementasikan dengan pendekatan yang berulang-ulang untuk sebuah proses pembelajarannya dari 1 data set, sehingga pembelajaran dapat dengan mudah diotomatisasi.

### B. Support Vector Machine (SVM)

*Support Vector Machine (SVM)* dikembangkan oleh Boser, Guyon, dan Vapnik, pertama kali diperkenalkan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. Konsep dasar metode SVM sebenarnya merupakan gabungan atau kombinasi dari teori-teori komputasi yang telah ada pada tahun sebelumnya, seperti *marginhyperplane* (Dyda dan Hart, 1973; Cover, 1965; Vapnik, 1964), kernel diperkenalkan oleh

Aronszajn tahun 1950, Lagrange Multiplier yang ditemukan oleh Joseph Louis Lagrange pada tahun 1766, dan demikian juga dengan konsep-konsep pendukung lain.

Menurut Fachrurrazi (2011) SVM merupakan suatu teknik untuk melakukan prediksi, baik prediksi dalam kasus regresi maupun klasifikasi. Teknik SVM digunakan untuk mendapatkan fungsi pemisah (hyperplane) yang optimal. Hyperplane ini dapat berupa line pada two dimension dan dapat berupa flat plane pada multiple dimension.



Gbr.1 Hyperplane dan Margin

pada Gbr.1 dapat dilihat bahwa *Hyperplane* yang ditemukan oleh metode SVM posisinya tepat ditengah diantara dua kelas, yang bisa diartikan bahwa jarak antara *hyperplane* dengan objek-objek data yang ada, berbeda dengan kelas yang berdekatan (terluar) dengan ditandai pada Gbr.1 adalah bulat tanpa isi dan berisi tanda *plus* . Pada metode SVM/SMO objek-objek data yang berada paling luar dan yang paling dekat dengan *hyperplane* disebut *support vector*. *Support vector* merupakan objek yang paling sulit dikelompokkan, karena posisi yang hampir bertindihan (*overlap*) dengan kelas lain. Karena hanya *support vector* yang memiliki karakteristik yang diperlukan, maka objek ini yang dijadikan sebagai perhitungan untuk menemukan *hyperplane* yang paling optimal oleh SVM.

Menurut pandangan (Nugroho, dkk, 2003), karakteristik SVM secara global dapat disimpulkan sebagai berikut:

- 1) Pada dasarnya SVM merupakan bagian dari *linear classifier*.
- 2) Mengalih bentuk data pada *input space* ke ruang yang berdimensi lebih tinggi, dan melakukan optimisasi pada ruang *vector* yang baru tersebut dengan menggunakan *Pattern recognition*. Hal ini yang menjadikan SVM berbeda dari output *pattern recognition* pada umumnya.
- 3) Mengimplementasikan teknik *Structural Risk Minimization* (SRM).
- 4) Cara kerja SVM pada prinsipnya hanya mampu menangani pengelompokan dua kelas.

Sebagaimana diketahui untuk memudahkan proses pembelajaran dengan metode SVM diperlukan sebuah trick. Dimana trick tersebut dikenal dengan istilah *kernel trick* yang digunakan untuk menetapkan *support vector*. Dengan mengetahui *Kernel Function* yang digunakan, maka kebutuhan bentuk dari fungsi non-linear tidak diperlukan lagi. Pandangan (Karatzoglou, dkk, 2004) memiliki empat *Kernel Function*

yang acap kali dimanfaatkan dalam pustaka SVM antara lain sebagai berikut:

- 1) *Kernel linear* merupakan kernel yang paling sederhana dari semua fungsi kernel. Sering kali Kernel ini diimplementasikan dalam permasalahan pengelompokan teks.

$$K(x_i, x_j) = x_i^T x_j \quad (1)$$

- 2) *Kernel Polynomial* merupakan kernel yang banyak dimanfaatkan pada permasalahan pengelompokan *image*.

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^p, \gamma > 0 \quad (2)$$

- 3) *Kernel Radial Basis Gaussian* merupakan kernel yang pada umumnya dimanfaatkan pada data-data yang benar-benar sudah valid (*available*).

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (3)$$

- 4) *Kernel Tangent Hyperbolic* merupakan kernel yang sering dimanfaatkan pada jaringan saraf tiruan.

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (4)$$

### C. Sequential Minimal Optimization

Algoritma *Sequential Minimal Optimization* (SMO) ditemukan oleh (Platt, 2000) pada saat melakukan riset atau penelitian di perusahaan microsoft. Menurut (Platt, 1998) SMO merupakan bagian dari algoritma pengelompokan yang mudah diimplementasi dan bekerja secara sederhana. penggunaan SMO dilakukan untuk mengurangi atau menutupi kekurangan dari algoritma yang sudah ada yaitu algoritma *Support Vector Machine* (SVM) yang memiliki permasalahan pada *Quadratic Programming* (QP). SMO termasuk dalam metode *Decomposition* yang bekerja berdasarkan prinsip "*Working Set*" (Fan, Chen, & Lin, 2005). Tidak seperti metode sebelumnya, SMO memungkinkan menyelesaikan masalah kecil yang mungkin terjadi pada setiap prosesnya. Standart permasalahan pada *Quadratic Programming* (QP) adalah masalah optimasi pada lgoritma SMO yang melibatkan dua *Lagrange Multipliers*. SMO melibatkan *Working Set* berelemen dua sehingga penetapan penyelesaian yang optimal dapat ditetapkan secara analitis. Penyelesaian tersebut tentu akan mengakibatkan jumlah perulangan semakin bertambah, akan tetapi karena waktu yang dibutuhkan dalam setiap iterasi sangat kecil.

Secara singkat algoritma SMO bekerja sebagai berikut: pertama menemukan lagrange multiplier yang melanggar aturan karush-kuhn-tucker (KKT) dalam optimasi, kedua memilih multiplier kedua dan mengoptimalkan pasangannya, yang ketiga mengulangi poin 1 dan 2 hingga menjadi konvergen.

Jika semua lagrange multipliers mencapai kondisi KKT yang sesuai maka permasalahan dapat diselesaikan. SMO merupakan algoritma yang dapat menyelesaikan permasalahan *quadratic programming* dengan jalan bagaimana mendapatkan nilai  $a$  dengan menggunakan *analytical quadratic programming solver* pada setiap langkah sehingga waktu

training yang dibutuhkan bisa lebih cepat dari algoritma SVM. Algoritma ini dapat memecahkan masalah dengan cara memanfaatkan 2 buah data pada tiap-tiap iterasinya sehingga penelusuran penyelesaian yang terbaik dapat dilakukan. Kondisi ini menyebabkan jumlah proses perulangan bertambah, meskipun begitu waktu total yang dibutuhkan untuk pelatihan menjadi lebih singkat.

D. Evaluasi Unjuk Kerja Klasifikasi

Menurut (Nugroho, dkk, 2003) keakuratan klasifikasi adalah ukuran akurasi klasifikasi yang menunjukkan unjuk kerja teknik klasifikasi tersebut. Dengan tingginya tingkat keakuratan klasifikasi berarti unjuk kerja klasifikasi tersebut juga semakin tinggi. Keakuratan masalah pengelompokan data disajikan pada Tabel 1

TABEL 1  
AKURASI PENGELOMPAKAN DATA

TRUE VALUE	PREDIKSI	
	TRUE	FALSE
TRUE	TP	FN
FALSE	FP	TN

Dari Tabel 1 dapat dijelaskan sebagai berikut, pengukuran unjuk kerja menggunakan *confusion matrix*, memiliki 4 (empat) istilah sebagai gambaran hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)* dan *False Negative (FN)*. Nilai *True Negative (TN)* adalah jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive (FP)* merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, *True Positive (TP)* adalah data positif yang terdeteksi benar. *False Negative (FN)* adalah kebalikan dari *True Positive*, sehingga data positif, namun terdeteksi sebagai data negatif. Berdasarkan Tabel 1 perhitungan evaluasi keakuratan sebuah metode pengelompokan dapat dilakukan dengan formula sebagai berikut:

$$Akurasi = \frac{TN+TP}{TN+TP+FN+FP} \times 100\% \quad (5)$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

$$f - Measure = 2 * \frac{Precision*Recall}{Precision+Recall} \times 100\% \quad (8)$$

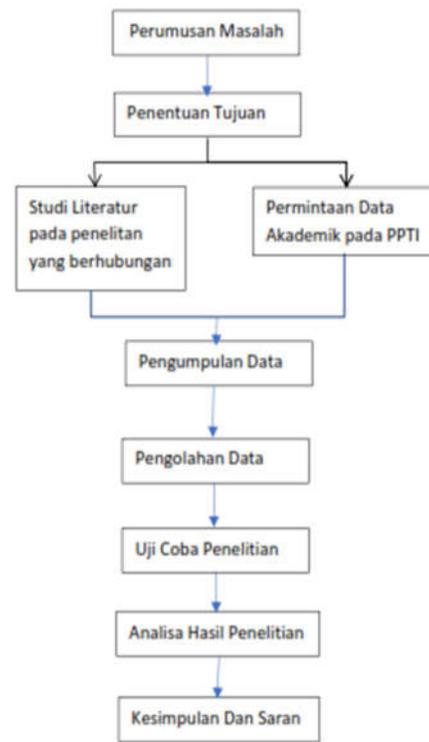
III. METODE PENELITIAN

Output yang dihasilkan dari penelitian ini merupakan sebuah model yang nantinya bisa diimplementasikan dalam pengelompokan data untuk peramalan unjuk kerja akademik yang ada di suatu jurusan Teknik Informatika Universitas Negeri Surabaya. Berdasarkan hasil studi pustaka yang telah dilakukan maka metode SVM dengan *Sequential Minimal Optimization* adalah yang paling baik untuk meramalkan produktivitas yang bisa dilakukan oleh Jurusan maupun

Universita. Data-data akademik yang didapat dari institusi ini awalnya masih bersifat kategorikal yang tidak bisa digunakan untuk metode SMO. Untuk itu dilakukan proses konversi data dari kategorikal menjadi data numerikal. Setelah diolah dengan SMO maka akan menghasilkan prediksi klasifikasi unjuk kerja akademik mahasiswa.

A. Skema Alur Penelitian

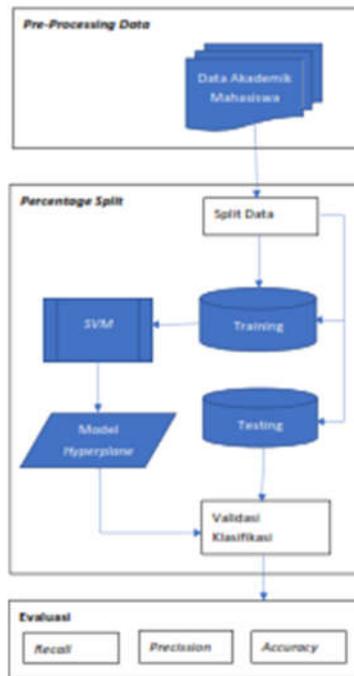
Untuk melakukan penelitian yang mencapai target diperlukan suatu langkah-langkah yang harus dijalankan. Langkah-langkah tersebut akan menjadi sandaran dalam melakukan penelitian hingga mencapai tujuan yang diinginkan. Berikut adalah alur proses secara umum yang akan dilakukan dalam penelitian Peramalan unjuk kerja Akademik menggunakan *Sequential Minimal Optimization* untuk pengelolaan Program Studi di Jurusan Informatika Universitas Negeri Surabaya yang dapat dilihat pada Gbr. 2



Gbr. 2 Tahapan Penelitian

Pada Gbr. 2 menunjukkan langkah pertama adalah melakukan perumusan masalah. Selanjutnya menetapkan tujuan dari penelitian yaitu pembuatan sebuah model prediksi akademik berdasarkan data ada pada UNESA, sehingga output dari penelitian ini dapat dimanfaatkan oleh pimpinan Jurusan khususnya dan pimpinan universitas pada umumnya dengan menggunakan metode *Sequential Minimal Optimization*. Tahap berikutnya melakukan kajian pustaka tentang penelitian-penelitian sebelumnya yang terkait atau berhubungan dengan penelitian ini. Setelah melakukan permintaan data-data

akademik yang dibutuhkan tahap selanjutnya adalah melakukan pengumpulan data-data dan menyimpannya dalam database. Tahap selanjutnya adalah pengolahan data yang ditunjukkan pada Gbr.3 diantaranya memilah atribut-atribut yang dibutuhkan, menghapus atribut-atribut yang tidak dibutuhkan. Tahapan terakhir adalah melakukan uji coba dengan skenario yang dirancang. Tahapan selanjutnya adalah melakukan analisa terhadap hasil uji coba yang telah dilakukan. Setelah melakukan analisa terhadap hasil tahap selanjutnya adalah menarik kesimpulan dan saran untuk penelitian.



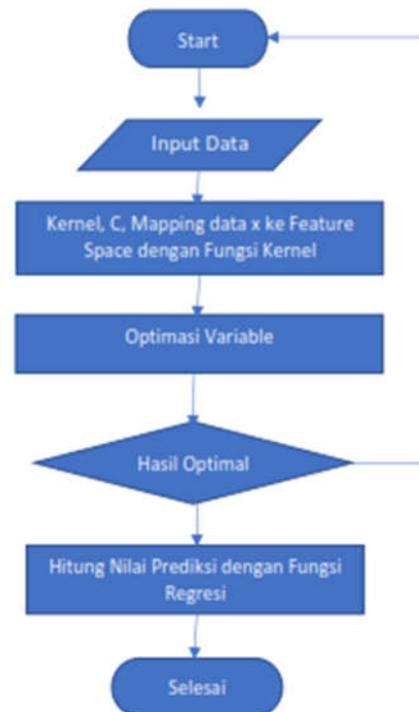
Gbr. 3 Pengolahan Data

Tahap awal sebelum data dapat digunakan oleh algoritma SVM/SMO adalah Pre-processing data. Pada pre-processing data, data-data akan diolah dengan memilah-milah variabel atribut yang digunakan atau yang dihilangkan. Sehingga data mentah tersebut menjadi data yang bisa digunakan sebagai input pada aplikasi WEKA. Selain itu pada proses ini, akan melakukan analisis distribusi data yang bertujuan untuk memilah sekumpulan data yang akan dijadikan sebagai variabel *input* dan variabel *output*. Disamping itu pada proses ini akan dilakukan normalisasi data. Dimana data mentah yang diperoleh akan dirubah menjadi bentuk data yang dapat digunakan dan data yang siap untuk diolah. Proses seleksi terhadap data ini menghasilkan atribut input yang digunakan untuk proses pengelompokan dengan uji korelasi pada data yang ada.

Proses berikutnya merupakan langkah ini yaitu langkah pembuatan model data dengan mengimplementasikan algoritma *Support Vector Machine* (SVM) dengan *Sequential Minimal Optimization* dengan menggunakan teknik *Percentage Split*. Teknik *Percentage Split* merupakan metode validasi yang

membagi data menjadi dua bagian yang sudah ditentukan besarnya, yaitu data latih (*training*) dan data pembelajaran (*testing*). Dari data-data yang sudah dilakukan proses pelatihan selanjutnya akan masuk ke proses pembuatan model klasifikasi menggunakan *SMO* hingga menghasilkan sebuah model *hyperplane*. Setelah model *hyperplane* didapat selanjutnya model tersebut akan diimplementasikan pada data testing. Untuk lebih jelas dan detail proses tersebut dapat digambarkan sebagaimana pada Gbr 4.

Proses terakhir merupakan proses inti dari penelitian ini yaitu proses evaluasi hasil dari tahap sebelumnya. Adapun evaluasi yang digunakan untuk pengukuran klasifikasi ini yaitu perhitungan *Recall*, *Precision* dan Akurasi.



Gbr. 4 Flowchar Algoritma SVM

## B. Akumulasi Data

Akumulasi data yang didapat pada penelitian ini bertujuan untuk mencari atribut-atribut yang mempengaruhi unjuk kerja akademik. Selanjutnya pada bagian ini akan diterangkan atribut-atribut mana yang ditetapkan sebagai input dan atribut-atribut mana yang ditetapkan sebagai output.

### 1) Atribut Masukan (*Input*)

Data mentah yang diperoleh pada penelitian ini merupakan data akademik mahasiswa Jurusan Teknik Informatika Universitas Negeri Surabaya yang diambil dari data lulusan tahun 2015 sampai dengan 2017 sebanyak 330 data terdiri dari 174 mahasiswa dari Prodi Teknik Informatika (82 lulus tepat waktu dan 92 tidak lulus tepat waktu) dan 156 mahasiswa Prodi Sistem Informasi (87 lulus dengan kriteria tepat waktu dan 69 dengan kriteria tidak lulus tepat waktu).

2) Atribut Keluaran (Output)

Atribut keluaran yang dimaksud pada penelitian merupakan data kategori yang berisi ujuk kerja dan keberhasilan mahasiswa. Performansi unjuk kerja dalam penelitian ini digunakan untuk mengukur keberhasilan seorang mahasiswa dalam menempuh pembelajaran di jurusan Teknik Informatika, Fakultas Teknik UNESA. Kriteria atribut keluaran yang didapat berdasarkan arahan dari pada Buku Panduan Akademik UNESA Tahun 2020 serta tanya jawab dengan pimpinan jurusan yaitu Ketua Program Studi dan Sekretaris Jurusan di Jurusan Teknik Informatika. Kriteria unjuk kerja mahasiswa dapat disebabkan karena factor nilai dari IPS/IPK. Kriteria unjuk kerja akademik sebagai *output* pengelompokan data yang akan dihasilkan dalam penelitian ini, menggunakan status lulus “Tepat Waktu” dan “Tidak Tepat Waktu”

IV. HASIL DAN LUARAN YANG DICAPAI

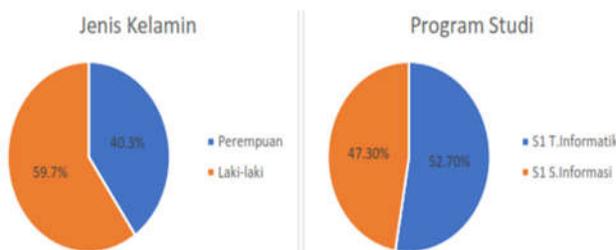
Hasil dan luaran yang dihasilkan dari penelitian ini akan dijabarkan dan diterangkan secara ringkas pada bagian ini. Untuk alat bantu yang digunakan peneliti untuk mengolah data, adalah *software* WEKA. Pada bagian ini hasil dan luaran dibagi menjadi 5 bagian yakni Pembagian Data, Implementasi *Sequential Minimal Optimization* pada WEKA, Uji dan Coba, serta Analisis penggunaan metode SMO.

A. Pembagian Data

Distribusi data diperlukan untuk mengetahui gambaran penyebaran data secara umum. Penyebaran yang dilakukan pada penelitian ini memuat penyebaran data berdasarkan jenis kelamin dan program studi. Penyebaran ini memuat prosentase distribusi data pada setiap atribut *input* dan atribut *output*. Penggambaran distribusi data dapat dilihat pada Tabel 2 dan Gbr. 5

TABEL 2  
DISTRIBUSI DATA VARIABLE MASUKAN

Variable Masukan	Jumlah	Prosentase
<b>Gender</b>	Laki-laki	197 59,7%
	Perempuan	133 40,3%
<b>Program Studi</b>	S1 T. Informatika	174 52,7%
	S1 Sistem Informasi	156 47,3%
<b>Total</b>	330	100%



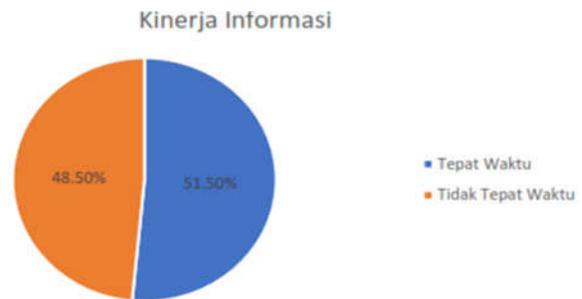
Gbr. 5 Grafik Distribusi Data Masukan

Sebagaimana yang telah dijelaskan sebelumnya atribut output yang merupakan hasil prediksi dari penelitian terbagi 2 (dua)

kriteria tingkatan unjuk kerja. Adapun kedua kriteria tersebut bisa terlihat dan tersajikan Tabel 3 dan Gbr 6.

TABEL 3  
DISTRIBUSI DATA KELUARAN

Variable Keluaran	Jumlah	Prosentase
<b>Tepat Waktu</b>	170	51,5%
<b>Tidak Tepat Waktu</b>	160	48,5%
<b>Total</b>	330	100%



Gbr. 6 Grafik Pembagian Data Output

Pada Tabel 3 dan Gbr 6 dapat data hasil unjuk kerja performansi akademik Jurusan Informatika Universitas Negeri Surabaya yang dijadikan bahan penelitian didominasi oleh mahasiswa yang selesai studinya “Tepat Waktu” dengan prosentase 51,5% dan 48,5% mahasiswa yang selesai studinya “Tidak Tepat Waktu” Dari 330 data yang berhasil dikumpulkan.

B. Implementasi SVM

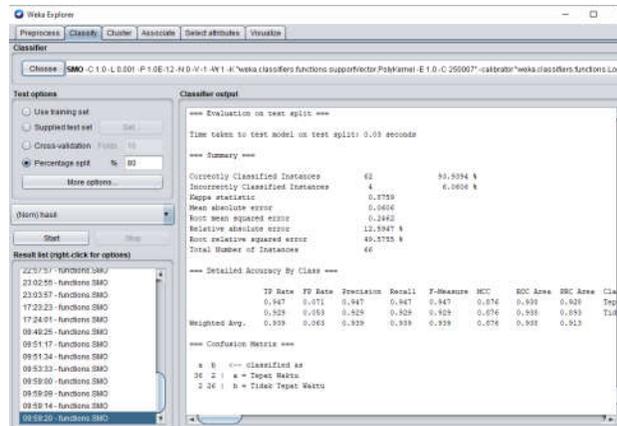
Data mentah yang didapat berupa file excel seperti pada Gbr 7 akan diubah strukturnya seperti Gbr 8 yang kemudian disimpan menjadi file csv.

A	B	C	D	E	F	G	H	I	J	
nipd	nm_pd	tmpt_lahir	tgl_lahir	jk	nm_prodi	id_smt	sks_smt	ipk	ips	
1	1505120400	FARIKHATUR RO'ISA	LAMONGAN	1998-01-03	P	S1Teknik Informatika	20151	19	3,29	3,29
2	1505120400	FARIKHATUR RO'ISA	LAMONGAN	1998-01-03	P	S1Teknik Informatika	20152	20	3,12	2,95
3	1505120400	FARIKHATUR RO'ISA	LAMONGAN	1998-01-03	P	S1Teknik Informatika	20161	20	3,3	3,65
4	1505120400	FARIKHATUR RO'ISA	LAMONGAN	1998-01-03	P	S1Teknik Informatika	20162	22	3,37	3,57
5	1505120400	FARIKHATUR RO'ISA	LAMONGAN	1998-01-03	P	S1Teknik Informatika	20171	20	3,42	3,63
6	1505120400	FARIKHATUR RO'ISA	LAMONGAN	1998-01-03	P	S1Teknik Informatika	20172	20	3,46	3,66
7	1505120400	FARIKHATUR RO'ISA	LAMONGAN	1998-01-03	P	S1Teknik Informatika	20181	17	3,5	3,79
8	1505120400	FARIKHATUR RO'ISA	LAMONGAN	1998-01-03	P	S1Teknik Informatika	20182	6	3,52	4
9	1505120400	FARIKHATUR RO'ISA	LAMONGAN	1998-01-03	P	S1Teknik Informatika	20191	0	3,52	0
10	1505120400	FARIKHATUR RO'ISA	LAMONGAN	1998-01-03	P	S1Teknik Informatika	20191	0	3,52	0
11	1505120400	DIAZ ARDIAN ALVIANTO	PONOROGO	1997-04-23	L	S1Teknik Informatika	20151	19	3,14	3,14
12	1505120400	DIAZ ARDIAN ALVIANTO	PONOROGO	1997-04-23	L	S1Teknik Informatika	20152	20	3,01	2,89
13	1505120400	DIAZ ARDIAN ALVIANTO	PONOROGO	1997-04-23	L	S1Teknik Informatika	20161	20	3,13	3,35
14	1505120400	DIAZ ARDIAN ALVIANTO	PONOROGO	1997-04-23	L	S1Teknik Informatika	20162	22	3,26	3,61
15	1505120400	DIAZ ARDIAN ALVIANTO	PONOROGO	1997-04-23	L	S1Teknik Informatika	20171	20	3,31	3,51
16	1505120400	DIAZ ARDIAN ALVIANTO	PONOROGO	1997-04-23	L	S1Teknik Informatika	20172	20	3,36	3,63
17	1505120400	DIAZ ARDIAN ALVIANTO	PONOROGO	1997-04-23	L	S1Teknik Informatika	20181	15	3,42	3,85
18	1505120400	DIAZ ARDIAN ALVIANTO	PONOROGO	1997-04-23	L	S1Teknik Informatika	20182	2	3,42	4
19	1505120400	DIAZ ARDIAN ALVIANTO	PONOROGO	1997-04-23	L	S1Teknik Informatika	20191	0	3,42	0
20	1505120400	DIAZ ARDIAN ALVIANTO	PONOROGO	1997-04-23	L	S1Teknik Informatika	20192	0	3,42	0

Gbr. 7 Contoh Data Mentah

nipd	jk_Prodi	ips1	ips2	ips3	ips4	ips5	ips6	ips7	ips8	ips9	ips10	ips11	ips12	ips13	jml_sms	jml_sks	hasil
15051204001P	SI Teknik Informatik	3,3	2,95	3,7	3,6	3,6	3,7	3,8	4	0	0	0	0	0	9	144	Tepat Waktu
15051204001L	SI Teknik Informatik	3,1	2,89	3,4	3,6	3,5	3,6	3,9	4	0	0	4	0	0	12	144	Tidak Tepat Wa
15051204002P	SI Teknik Informatik	3,5	3,29	3,7	3,5	3,6	3,5	3,8	4	0	0	0	0	0	9	144	Tepat Waktu
15051204004L	SI Teknik Informatik	3,2	3,1	3,4	3,4	3,3	3,6	3,8	3,8	0	0	0	0	0	13	138	Tidak Tepat Wa
15051204005P	SI Teknik Informatik	3,4	3,28	3,7	3,7	3,7	3,7	3,8	4	0	0	0	0	0	9	144	Tepat Waktu
15051204006L	SI Teknik Informatik	3,3	2,31	3	3,2	3,5	3,6	3,5	3,3	0	0	4	0	0	11	153	Tidak Tepat Wa
15051204007L	SI Teknik Informatik	3	3,1	3,5	3,4	3,5	3,6	3,3	1	0	0	0	4	0	12	173	Tidak Tepat Wa
15051204008L	SI Teknik Informatik	2,7	3,05	3,5	3,5	3,2	3,5	3,3	0,9	0	0	0	0	0	13	179	Tepat Waktu
15051204009P	SI Teknik Informatik	3,5	3,35	3,8	3,7	3,7	3,7	3,8	4	0	0	0	0	0	8	144	Tepat Waktu
15051204010L	SI Teknik Informatik	4	3,69	3,9	3,9	3,9	3,8	4	4	0	0	0	0	0	9	144	Tepat Waktu
15051204011L	SI Teknik Informatik	2,8	3,2	3,5	3,2	3,4	3,3	3,8	4	0	0	0	0	0	10	141	Tidak Tepat Wa
15051204012L	SI Teknik Informatik	3,4	3,06	3,1	3,1	3,5	2,7	3,6	0	0	0	0	0	0	13	138	Tidak Tepat Wa
15051204013L	SI Teknik Informatik	3,2	3,03	2,7	1,6	1,4	0,4	1	0	0	0	0	0	0	12	127	Tidak Tepat Wa
15051204014P	SI Teknik Informatik	3,7	3,34	3,8	3,8	3,7	3,6	2,7	4	0	0	0	0	0	8	150	Tepat Waktu
15051204015P	SI Teknik Informatik	3,1	3,3	3,8	3,6	3,5	3,7	3,8	0	3,8	0	0	0	0	9	144	Tidak Tepat Wa

Gbr. 8 Contoh Data Setelah di olah



Gbr.9 Proses Klasifikasi

Gbr. 9 menunjukkan bahwa, proses klasifikasi menggunakan teknik *Percentage split 80%* yang artinya data sebanyak 330 digunakan sebagai data latih sebanyak 80% x 330=264 data, sisanya sebanyak 66 data digunakan sebagai data uji coba.

Metode *Support Vector Machine /SMO* pada aplikasi WEKA secara default dan digunakan pada penelitian ini menggunakan fungsi *Polynomial Kernel* serta menggunakan *Regresi Logistik* sebagai metode pembandingan

### C. Perhitungan Akurasi

Salah satu cara untuk menghitung keakuratan akurasi prediksi unjuk kerja akademik mahasiswa yang mengimplementasikan metode *Support Vector Machine (SVM)* dapat digunakan juga untuk sistem yang dibangun dengan metode *Sequential Minimal Optimization (SMO)*. Untuk menghitung keakuratan prediksi dalam penelitian ini digunakan confusion matriks sebagai alat bantu perhitungannya. Sebagaimana diketahui Confusion matriks biasanya diimplementasikan pada data minning. Adapun cara perhitungan akurasinya adalah dengan membandingkan data hasil peramalan yang telah didapat dengan data sebenarnya yang sudah ada. Pada confusion matik terdapat 3 (tiga) keluaran, yaitu akurasi, presisi dan recall

Terdapat 3 (tiga) pelabelan untuk hasil prediksi kriteria unjuk kerja mahasiswa, yaitu : label TP (*True Positive*), yang artinya jika ouput uji coba yang dilakukan untuk peramalan

kriteria unjuk kerja akademik benar, tepat sesuai dengan data sebenarnya. Selanjutnya label FP (*False Positive*) dihasilkan dengan cara mnghitung output peramalan yang tidak sama atau tidak sesuai dengan data yang sebenarnya, akan tetapi kriteria unjuk kerja yang diramalkan atau diprediksi masih berada di bawah kriteria unjuk kerja yang sebenarnya. Dalam beberapa kasus seperti salah satu tujuan dari penelitian ini sebagai pendeteksian dini untuk kinerja mahasiswa, label FP (*False Positive*) dapat dipertimbangkan sebagai system peringatan deteksi dini bagi ketua jurusan di perguruan tinggi. Sehingga, jika pendeteksian kriteria unjuk kerja yang menghasilkan ouput dibawah kriteria aktual (data sebenarnya) akan dipandang sebagai suatu kesalahan prediksi data yang bersifat positif (*FalsePositive*).

Selanjutnya label FN (*False Negatif*) merupakan output prediksi kriteria unjuk kerja akademik yang berada di atas data level unjuk kerja akademik yang sebenarnya akan dipandang sebagai kesalahan peramalan data yang bersifat negatif. Setelah *output* TP, FP dan FN didapat langkah selanjutnya adalah mengolah atau menghitung menjadi data-data prosentase antara rumus yang sudah diterangkan pada bagian sebelumnya. Setelah melakukan pengolahan data lulusan melalui software WEKA dengan melakukan transformasi data untuk menseleksi subset fitur yang dilakukan dengan mencoba uji validasi *Percentage Split 80%* maka selanjutnya menggunakan model SVM didapat hasil Confusion Matrix seperti terlihat di tabel 4.

TABLE 4  
CONFUSION MATRIX SVM

	Tepat Waktu	Tidak Tepat Waktu
Data Aktual Tepat Waktu	36	2
Data Aktual Tidak Tepat Waktu	2	26

Dengan menggunakan rumus penentuan akurasi yang telah dijelaskan pada bab 2, maka didapat nilai sebagai berikut :

$$akurasi = \frac{36+26}{36+2+2+27} \times 100\% = 93,94\% \quad (1)$$

$$Precision = \frac{36}{36,2} \times 100\% = 94,7\% \quad (2)$$

$$Recall = \frac{36}{36+3} \times 100\% = 94,7\% \quad (3)$$

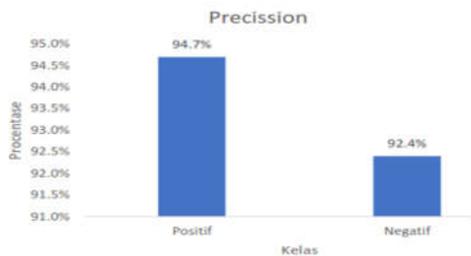
$$f - Measure = 2 * \frac{0,947 * 0,947}{0,947 + 0,947} * 100\% = 94,7\% \quad (4)$$

Dari perhitungan di atas dapat disimpulkan sebagaimana dijelaskan pada tabel 5

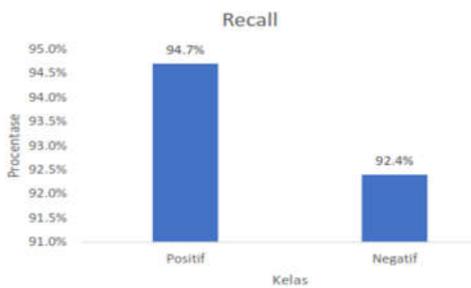
TABEL 5  
 HASIL PRESSION, RECALL, F-MEASURE

Klass Target	Precision (%)	Recall (%)	F-Measure (%)
Tepat Waktu	94,7%	94,7%	94,7%
Tidak Tepat Waktu	92,4%	92,4%	92,4%

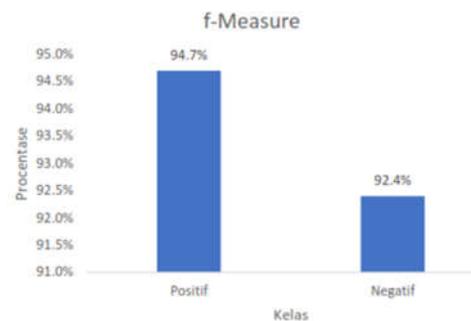
Pada gambar 10, 11, dan 12 tergambar grafik presisi, recall, dan *f-measure* pada tiap kelas dengan SVM. Adanya grafik agar kita dapat lebih mudah mengetahui perbedaan presentase tiap kelas.



Gbr. 10 Grafik Precision



Gbr. 11 Grafik Recall



Gbr. 12 Grafik f-Measure

Hasil *precision*, *recall* dan *f-measure* uji data antara kelas positif dan negative memiliki nilai atau bentuk grafik yang sama yaitu kelas positif bernilai 94,7% dan kelas negative bernilai 92,4%. Kelas positif lebih tinggi dari negatif terjadi karena jumlah data pada kelas negatif lebih sedikit dibanding kelas positif

## V. KESIMPULAN DAN SARAN

Berdasarkan hasil uji coba dan luaran yang didapat maka dapat disimpulkan bahwa : Pembuatan peramalan unjuk kerja akademik menggunakan metode *Sequential Minimal Optimization* berhasil diimplementasi untuk mengukur level performansi mahasiswa tersebut. Berdasarkan hasil *testing* yang telah dijelaskan pada bagian hasil dan keluaran dengan menggunakan data akademik mahasiswa Tahun 2015-2017 mendapatkan prosentase akurasi sebesar 93,94%, presisi sebesar 94,7% dan recall sebesar 94,7%. Sehingga model prediksi menggunakan *Support Vector Machine* (SVM) dengan *Sequential Minimal Optimization* (SMO) dapat digunakan sebagai salah satu metode prediksi.

Meskipun penggunaan metode SVM/SMO pada penelitian ini mendapatkan hasil dengan akurasi yang tinggi, ada beberapa saran yang diperlukan untuk penelitian tahap selanjutnya yaitu : Untuk penelitian yang akan datang perlu menggunakan dataset yang lebih besar sehingga keakuratan tingkat akurasi kinerja model lebih terjamin. Diperlukan menambahkan parameter-parameter selain IPS/IPK, seperti asal sekolah, asal tempat dan lain sebagainya, sehingga hasil yang didapat lebih detil.

## REFERENSI

- [1] Kamagi. 2014. Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. Tangerang: Universitas Multimedia Nusantara.
- [2] A. Karatzoglou, A. Smola, K. Hornik, and Zeileis, A. (2004) kernlab – An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, vol. 11(9), pp. 1-20.
- [3] Mustafa. (2014) Perancangan Aplikasi Prediksi Kelulusan Tepat Waktu Bagi Mahasiswa Baru Dengan Teknik Data Mining (Studi Kasus : Data Akademik Mahasiswa STMIL Dipanegara Makassar). Makassar: Citec Jurnal Vol : 1. STMik Dipanegara
- [4] Zhang, Y dan Wang, W. (2010) Pattern Classification of Electroencephalography from the Typical Specialized Students. *Education Technology and Computer Science (ETCS), International Workshop.*, vol. 1, pp. 836-839.
- [5] Fachrurazi. 2011. "Penerapan Pembelajaran Berbasis Masalah untuk Meningkatkan Kemampuan Berpikir Kritis dan Komunikasi Matematis Siswa Sekolah Dasar". *Forum Penelitian*, Edisi khusus No. 1: 76-89.
- [6] Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003) Support Vector Machine dan Aplikasinya Dalam Bioinformatika. Diakses 15 Januari 2017, dari <http://www.ilmukomputer.com>.
- [7] C. Platt, John. 2000. Fast Training of Support Vector Machine using Sequential Minimal Optimization. *SMO-book*. Pages 41-65.