

## PERBANDINGAN KINERJA NAIVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK PREDIKSI HERREGISTRASI

Selvy Megira<sup>1)</sup>, Kusrini<sup>2)</sup>, Emha Taufiq Luthfi<sup>3)</sup>,

<sup>1)</sup>STMIK Lembah Dempo

<sup>2,3)</sup>Universitas Amikom Yogyakarta

Jalan H.Kol M.Nuh, Desa Airlaga, Pagar Alam

Pos-el : [selvymegi96@gmail.com](mailto:selvymegi96@gmail.com), [kusrini@amikom.ac.id](mailto:kusrini@amikom.ac.id)

### ABSTRACT

*Research with the title “Performace Comparison of Naive Bayes and Support Vector Machine for Herregistration Prediction” aims to determine the comparison of naive bayes and support vector machine in herregistration prediction with accuracy parameters and AUC using test scenarios with split validation, which will later be used as a reference for parties universities to carry out policies for students especially those who have the potential to experience non-registration. In this study, only herregistration prediction of prospective new students at the computer science faculty for the class of 2015 to 2017 using of the naive bayes algorithm and support vector machine. The accuracy obtained in the Naive Bayes method is 93.54% and AUC 0.946 while the support vector machine method is 92.67% and AUC 0.877 uses the RBF kernel with parameter cost (C) 1.0 and Epsilon 0.0. In addition, performance accuracy and AUC are very influential when deleting one of the variables used.*

**Keywords:** AUC, Naive Bayes, Support Vector Machine

### ABSTRAK

Penelitian dengan judul “ Perbandingan Kinerja Naive Bayes dan Support Vector Machine untuk Prediksi Herregistrasi “ bertujuan untuk mengetahui perbandingan kinerja *naive bayes* dan *support vector machine* dalam melakukan prediksi herregistrasi dengan parameter akurasi dan AUC menggunakan skenario pengujian dengan *split validation* yang nantinya akan dapat dijadikan acuan bagi pihak universitas untuk melakukan kebijakan bagi para mahasiswa khususnya yang berpotensi mengalami tidak melakukan herregistrasi. Pada penelitian ini hanya melakukan prediksi herregistrasi calon mahasiswa baru pada fakultas ilmu komputer untuk angkatan tahun 2015 sampai dengan tahun 2017 dengan menggunakan metode algoritma *naive bayes* dan *support vector machine*. Akurasi yang dihasilkan pada metode *naive bayes* yaitu 93,54% dan AUC 0,946 sedangkan pada metode *support vector machine* yaitu 92,67% dan AUC 0,877

menggunakan kernel RBF dengan parameter *cost* (C) 1,0 dan Epsilon 0,0. Selain itu performa akurasi dan AUC sangat berpengaruh apabila dilakukan penghapusan terhadap salah satu variabel yang digunakan.

Kata kunci: AUC, Naive Bayes, Support Vector Machine

## I. PENDAHULUAN

Perguruan tinggi sebagai institusi pendidikan akademik, dalam perguruan tinggi terdapat lima jenis bentuk yakni universitas, institut, sekolah tinggi, akademi, politeknik serta akademi komunitas. Universitas sebagai lembaga pendidikan tinggi yang terdiri dari beberapa fakultas yang mengadakan pendidikan ilmiah dan/atau professional. Berdasarkan data yang tercatat dari Pangkalan Data Pendidikan Tinggi di Provinsi Bengkulu terdapat 8 perguruan Tinggi Negeri dan perguruan Tinggi Swasta yang berjumlah 23.

Dengan jumlah perguruan swasta yang tidak sedikit maka terjadi persaingan untuk menarik calon mahasiswa baru agar mendaftar pada Perguruan Tinggi Swasta. Sebagian besar perguruan tinggi memiliki data akademik dan administrasi dalam jumlah yang besar, karena data dari tahun ke tahun akan terakumulasi. Namun hanya sebagian kecil yang

memanfaatkan data tersebut khususnya dalam penyusunan evaluasi diri (Suprawoto, 2016).

Universitas Dehasen Bengkulu sebagai salah satu perguruan tinggi swasta di Provinsi Bengkulu menjadi bagian dari persaingan untuk mendapatkan calon mahasiswa baru setiap tahunnya, setiap penerimaan calon mahasiswa baru sering terjadi ada calon mahasiswa yang tidak melakukan herregistrasi. Jika kemungkinan calon mahasiswa baru yang tidak melakukan herregistrasi dapat di ketahui lebih awal maka pihak pengelola dapat melakukan tindakan-tindakan untuk mempertahankan calon mahasiswa baru yang mungkin potensial.

Salah satu cara untuk mengetahui calon mahasiswa baru yang tidak herregistrasi dengan melakukan prediksi apakah seorang calon mahasiswa baru akan cenderung herregistrasi atau tidak menggunakan

*data mining* dengan membandingkan metode klasifikasi. Klasifikasi merupakan suatu pekerjaan yang melakukan pelatihan atau memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia yang menghasilkan suatu model kemudian disimpan sebagai memori (Prasetyo, 2012). Metode klasifikasi yang umum digunakan antara lain *decision tree*, *k-nearest neighbor* (Prasetyo, 2016), *naïve bayes*, *neural network* dan *support vector machines* (Prasetyo, 2014).

Beberapa penelitian mengenai klasifikasi data menggunakan metode *naïve bayes* ataupun *support vector machine* dilakukan oleh Jannieca Camba menunjukkan bahwa *support vector machine* terbukti dapat mengklasifikasi berkurangnya mahasiswa di bidang seni dan ilmu pengetahuan, profesi medis, keperawatan dan farmasi yang disebabkan faktor pekerjaan, wilayah dan kelas (Jannieca Camba et al, 2016).

Penelitian selanjutnya menunjukkan bahwa kinerja *support vector machine* dengan menggunakan teknik

validasi *7-fold cross* dapat membantu prediksi kinerja siswa serta memberikan informasi pada institusi dalam menempatkan siswa ke berbagai program fakultas yang sesuai dengan tingkat akurasi 72,25 % (Mojisola G. Asogbon, 2016). Pada penelitian selanjutnya *support vector machine* terbukti memiliki akurasi lebih tinggi pada data *set wine* sebesar 84.15% dibandingkan dengan *k-nearest neighbor* sebesar 95,52% (Prasetyo, 2016).

Berdasarkan paparan diatas, peneliti bermaksud untuk melakukan perbandingan kinerja algoritma *naïve bayes* dan *support vector machine* dalam melakukan prediksi herregistrasi calon mahasiswa baru.

### **1.1 Rumusan Masalah**

Berdasarkan latar belakang yang telah disampaikan sebelumnya, maka penulis merumuskan masalah pada penelitian ini adalah sebagai berikut.

1. Diantara dua algoritma klasifikasi *Naïve Bayes* dan *Support Vector Machines*, algoritma manakah yang memiliki kinerja terbaik dalam

prediksi herregistrasi calon mahasiswa baru?

2. Apakah variabel atau atribut yang digunakan dapat mempengaruhi kinerja akurasi, dan *Area Under Curve (AUC)* apabila dilakukan penghapusan terhadap salah satu variabel tersebut?

### 1.2 Batasan Variabel Penelitian

Untuk dapat mencapai tujuan dari penelitian ini agar lebih terarah, maka ditetapkan batasan masalah sebagai berikut:

1. Penelitian ini melakukan perbandingan pada algoritma *Naïve Bayes* dan *Support Vector Machine* untuk prediksi herregistrasi calon mahasiswa baru.
2. Data set yang digunakan sebagai sampel data penelitian (data uji dan data training) adalah data calon mahasiswa baru yang sudah dinyatakan lulus seleksi dari Universitas Dehasen Bengkulu dengan tahun 2015 sampai 2017 pada saat Penerimaan Mahasiswa Baru.
3. Atribut data yang akan digunakan meliputi nilai UN, jenis kelamin,

gelombang pendaftaran, pilihan jurusan 1, pilihan jurusan 2, jurusan lulus, pekerjaan orang tua, penghasilan orang tua, biaya kuliah dan status herregistrasi.

4. *Output* dari penentuan parameter tiap algoritma berupa tingkat akurasi terhadap parameter yang digunakan.
5. Metode pengukuran validasi kinerja algoritma *Naïve Bayes* dan *Support Vector Machine* menggunakan metode *Split Validation*.

### 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengetahuidi antara algoritma *naïve bayes* dan *support vector machine* mana yang memiliki kinerja terbaik apabila digunakan dalam prediksi herregistrasi calon mahasiswa baru.
2. Mengetahui apakah variabel penentu dapat mempengaruhi kinerja akurasi, dan *Area Under Curve (AUC)* apabila dilakukan penghapusan terhadap salah satu variabel penentu?

**1.4 Manfaat Penelitian**

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Dapat memberikan evaluasi dalam menentukan algoritma yang lebih baik dalam prediksi herregistrasi calon mahasiswa baru.
2. Diharapkan dengan adanya prediksi herregistrasi calon mahasiswa baru dapat membantu sebagai bahan acuan dalam mengevaluasi calon mahasiswa baru yang tidak melakukan registrasi di Universitas Dehasen Bengkulu. Sehingga dapat menentukan strategi yang bermanfaat dalam jumlah target mahasiswa baru.
3. Menambah pengalaman dalam melakukan penelitian khususnya dibidang Teknologi Informatika.

**II. LANDASAN TEORI**

**1.1 Naive Bayes**

Naïve bayes adalah teknik prediksi yang berbasis probabilistik sederhana dengan penerapan teorema Bayes (aturan Bayes) dengan asumsi independensi (ketidaktergantungan)

yang kuat (Prasetyo, 2012). Klasifikasi naïve bayes diasumsikan bahwa ada atau tidaknya ciri spesifik dari sebuah kelas tersebut tidak ada hubungannya dengan ciri dari kelas lainnya. Berdasarkan ciri alami dari sebuah model probabilitas, klasifikasi Naïve Bayes dapat dibuat lebih efisien sebagai bentuk pembelajaran. Adapun parameter yang dapat digunakan untuk perhitungan Naïve Bayes dapat menggunakan metode maximum likelihood atau kemiripan tertinggi (Prasetyo, 2012). Untuk menghitung prediksi dengan naïve bayes digunakan formula (Kusrini & Emha T, 2009) :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{1}$$

Keterangan:

- X :data dengan *class* yang belum diketahui
- H :hipotesis data X merupakan suatu *class* spesifik
- P(H|X) :probabilitas hipotesis H berdasar kondisi X (*posteriori probability*)

P(H) :probabilitas hipotesis H  
 (*prior probability*)  
 P(X|H) :probabilitas X berdasar  
 kondisi pada hipotesis H  
 P(X) : probabilitas dari X

$i = 1, 2, \dots, N$ , dan  $x_i = \{x_{i1}, x_{i2}, \dots, x_{iq}\}^T$  merupakan atribut (fitur) set untuk data latih ke-i. sedangkan label masing-masing dinotasikan  $y_i \in \{-1,+1\}$ . Diasumsikan kedua class  $-1$  dan  $+1$  dapat terpisah secara sempurna oleh hyperplane berdimensi  $d$ , yang didefinisikan :

$$w \cdot x_i + b = 0 \quad (2)$$

$w$  dan  $b$  adalah parameter model,  $w \cdot x_i$  merupakan *inner-product* antara  $w$  dan  $x_i$ . Data  $x_i$  yang termasuk class  $-1$  (sampel negatif) dapat dirumuskan sebagai pattern yang memenuhi pertidaksamaan :

$$w \cdot x_i + b \leq -1 \quad (3)$$

### 1.2 Support Vector Machine

Support Vector Machine merupakan metode yang berakar dari teori pembelajaran statistic yang hasilnya sangat menjanjikan dalam memberikan hasil yang lebih baik daripada metode lain, metode ini bekerja dengan baik pada set data dengan dimensi yang tinggi, bahkan Support Vector Machine yang menggunakan teknik kernel harus memetakan data asli dari dimensi asalnya sehingga menjadi dimensi lain yang relative lebih tinggi (Prasetyo, 2014). Teknik Support Vector Machine menarik orang dalam bidang data mining maupun machine learning dikarenakan performansinya yang meyakinkan dalam melakukan prediksi kelas suatu data baru (Susanto, dkk, 2010).

Secara mate-matika, formulasi SVM untuk kasus klasifikasi linier Setiap data latih dinyatakan  $(x_i, y_i)$  dengan

### 1.3 ROC Curve

Kurva ROC dibagi dalam dua dimensi, dimana tingkat TP diplot pada sumbu Y dan tingkat FP diplot pada sumbu X. akan tetapi dalam melakukan represntasi grafis untuk menentukan klasifikasi mana yang lebih baik maka digunakan metode yang melakukan penghitungan luas daerah dibawah kurva ROC yang disebut AUC (*Area Under the ROC Curve*). Dan dapat diartikan sebagai *probabilitas*, dalam melakukan pengukuran kinerja AUC

memperkirakan probabilitas *output* sampel yang dipilih secara acak dari populasi positif dan negative. Semakin besar AUC maka semakin kuat klasifikasi yang digunakan. AUC dapat disebut dengan kata lain bagian dari daerah unit persegi yang nilainya berkisar antara 0.0 dan 1,0 (Pudjo, et al, 2013).

#### **1.4 Split Validation**

*Split Validation* merupakan teknik validasi dengan cara membuat data terbagi menjadi dua bagian secara acak, dimana data tersebut dibagi menjadi data training dan data testing. Data training adalah data yang akan dipakai dalam melakukan pembelajaran. Sedangkan data testing ialah data yang digunakan sebagai pengujian dari model yang telah dibuat. Dengan menggunakan split validation maka akan dilakukan percobaan pada data training berdasarkan pada split ratio yang telah ditentukan. Setelah itu sisa dari split ratio pada data training digunakan untuk data testing, misalkan jika split ratio pada data training 80% maka data testing

sebesar 20% begitupun seterusnya (I. H. Witten, E. Frank, and M. A. Hall, 2011 ).

#### **1.5 Jenis, Sifat, dan Pendekatan Penelitian**

Jenis penelitian yang dilaksanakan menggunakan penelitian eksperimen, karena data yang digunakan bersifat kuantitatif. Penelitian eksperimen yaitu melakukan pengujian tingkat akurasi, dan AUC pada algoritma Naïve Bayes dan Support Vector Machine untuk prediksi herregistrasi.

Sifat dari penelitian ini menggunakan penelitian deksriptif. Yang digunakan untuk mengetahui perbandingan tingkat *accuracy*, dan *Area Under Curve* (AUC) menggunakan algoritma Naïve Bayes dan Support Vector Machine dalam memprediksi herregistrasi calon mahasiswa baru dan mengetahui apakah jumlah variabel dan record mempunyai pengaruh terhadap performa kedua algoritma tersebut.

Penelitian ini menggunakan pendekatan kuantitatif, dimana pendekatan kuantitatif pada dasarnya merupakan suatu pengamatan yang

melibatkan suatu ciri tertentu, berupa perhitungan, angka atau kuantitas. Pengumpulan data dilakukan melalui hasil eksperimen kemudian data tersebut akan dilakukan perbandingan dan analisis, untuk melihat perbandingan mana yang lebih baik dalam memprediksi herregistrasi calon mahasiswa baru maka disajikan data dalam bentuk tabel dan diagram.

### **1.6 Metode Pengumpulan Data**

Data yang akan digunakan dalam penelitian ini adalah data mahasiswa yang dipilih dengan rentang tahun 2015 sampai 2017 dari Penerimaan Mahasiswa Baru dengan atribut yang dibutuhkan nilai UN, jenis kelamin, gelombang pendaftaran, pilihan jurusan 1, pilihan jurusan 2, jurusan lulus, pekerjaan orang tua, penghasilan orang tua, biaya kuliah dan status herregistrasi. Adapun teknik pengumpulan data yang digunakan adalah observasi, wawancara dan tinjauan pustaka. Metode wawancara yaitu melakukan wawancara terhadap pihak-pihak yang berkaitan langsung dengan permasalahan yang sedang dibahas pada penelitian ini untuk

mendapatkan gambaran dan penjelasan secara mendasar. Metode observasi yaitu melakukan pengamatan ke Universitas Dehasen Bengkulu untuk mendapatkan informasi dan data yang dibutuhkan. Dataset yang digunakan pada penelitian ini yaitu *dataset* Penerimaan Mahasiswa Baru tahun angkatan 2015 sampai dengan 2017. Metode studi pustaka yaitu pada tahap ini dilakukan pengumpulan referensi dari literature-literature yang dapat mendukung penelitian ini sebagai landasan teori dan dasar pedoman dalam pembuatan laporan.

### **1.7 Metode Analisis Data**

Metode analisis data pada penelitian ini dilakukan dengan *Split Validation*. Dengan terlebih dahulu data yang telah dikumpulkan dari Penerimaan Mahasiswa Baru dilakukan penyeleksian dan pembersihan data. Atribut yang tidak dibutuhkan akan dihapus kemudian data tersebut akan di transformasikan dalam format *xlsx*. setelah didapatkan data yang sesuai dengan yang dibutuhkan maka data tersebut akan dilakukan pengujian menggunakan metode *naïve bayes*

dan *support vector machine* dalam melakukan prediksi herregistrasi dengan teknik *split validation*. Sedangkan metode pengevaluasian dilakukan dengan algoritma *naïve bayes* dan *support vector machine*, pengukuran tingkat *accuracy*, dan *Area Under Curve (AUC)* digunakan sebagai acuan untuk menentukan hasil manakah algoritma yang direkomendasikan pada penelitian ini.

### Proses Prediksi

Metode yang digunakan untuk memprediksi herregistrasi calon mahasiswa baru adalah *Naive Bayes* dan *Support Vector Machine*. Variabel input yang digunakan dalam penelitian ini yaitu data calon mahasiswa baru yang sudah dinyatakan lulus seleksi. Jumlah kategori yang akan diprediksi memiliki 2 (dua) kategori *output* adalah mahasiswa yang Herregistrasi dan Tidak Herregistrasi. Tahapan algoritma *Support Vector Machine* sebagai berikut:

- a. Melakukan proses transformasi data sesuai dengan format *Support Vector Machine*.

- b. Menentukan fungsi kernel yang akan digunakan yaitu menggunakan tipe kernel *radial basis function* (RBF).
- c. Menentukan nilai-nilai parameter kernel dan parameter cost (C) untuk melakukan optimasi.
- d. Memilih parameter terbaik untuk optimasi data training dan untuk prediksi data testing.
- e. Menampilkan hasil prediksi.

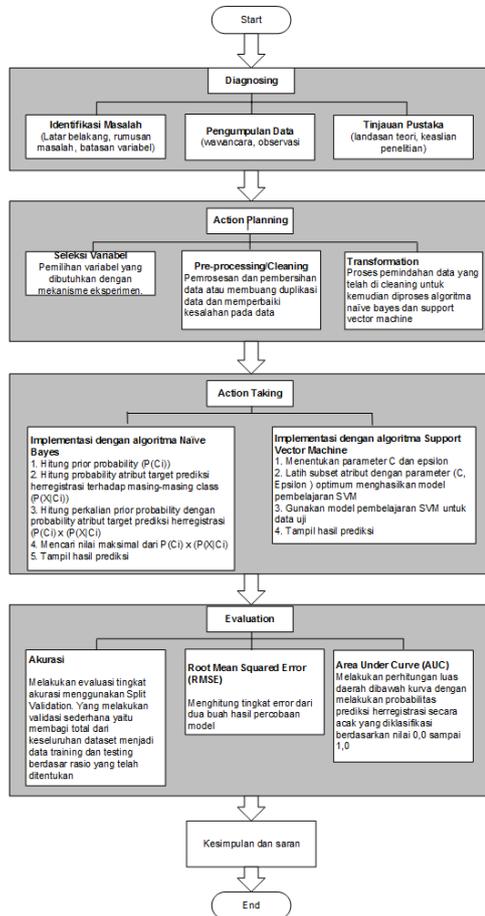
Sedangkan tahapan algoritma *Naive Bayes* sebagai berikut:

- a. Menghitung *prior probability* ( $P(C_i)$ ).
- b. Hitung probability atribut target prediksi herregistrasi terhadap masing-masing class ( $P(X|C_i)$ ).
- c. Hitung perkalian *prior probability* dengan *probability* atribut target prediksi herregistrasi ( $P(C_i) \times P(X|C_i)$ ).
- d. Mencari nilai maksimal dari  $P(C_i) \times P(X|C_i)$ .
- e. Menampilkan hasil prediksi.

### Alur Penelitian

Alur penelitian merupakan tahapan-tahapan yang harus dilalui dalam penelitian ini untuk menghasilkan

output yang diinginkan. Adapun alur dalam penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Diagram Alur Penelitian

**Data selection**

Pada tahapan ini proses seleksi variabel data set menggunakan metode Eksperimen. Proses seleksi variabel data set dilakukan dengan mencocokkan variabel/atribut yang awal dengan variabel/atribut yang ada pada basis kasus. Analisis ini digunakan untuk mengelompokkan peubah-peubah penting dalam menemukan fenomena, sekaligus mempelajari struktur dan melihat

hubungan antar variabel calon mahasiswa yang tidak herregistrasi.

**Data cleaning**

Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsesten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*). Akan tetapi dilakukan juga proses *enrichment* yang merupakan proses “memperkaya” data yang sudah ada dengan data ataupun informasi lain yang relevan dan diperlukan (Kusrini, Luthfi, 2009).

**Transformation**

Untuk mengubah dataset maka digunakan transformasi data sehingga informasi terbaik diambil dan dengan melakukan pengurangan ataupun pengubahan tipe data standar sehingga data siap digunakan untuk dipresentasikan ke teknik data mining menggunakan *Rapidminer*.

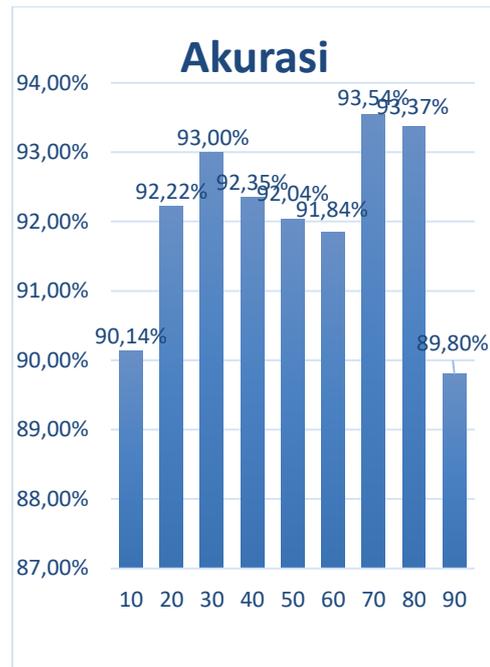
**III. HASIL DAN PENGUJIAN**

Pengujian yang dilakukan dengan menggunakan Bahasa Pemrograman *Rapid Miner Studio 7.6*, data yang digunakan di representasikan dalam bentuk proses pengujian dengan penerapan *algoritma Naive Bayes* dan *Support Vector Machine*. Pengujian dilakukan dengan menggunakan data yang tersimpan

dalam *Microsoft Excel*. Dimana sampel data yang digunakan terdapat 980 data dan 10 variabel. Data yang terdapat pada file excel awal mulanya di import ke dalam *Rapidminer*, setelah itu menentukan indikator kedalam masing – masing atribut. Atribut di rubah menjadi label karena penentu, dan field no (nomor urut) tanda centang di hilangkan karena tidak diperlukan. Gambar 2 merupakan hasil pengaturan dan potongan dataset pada proses import kedalam *Rapidminer*

**Prediksi dengan Naive bayes**

Hasil pengujian akurasi pada Naive Bayes dengan *split validation* dapat dilihat pada Gambar 2:

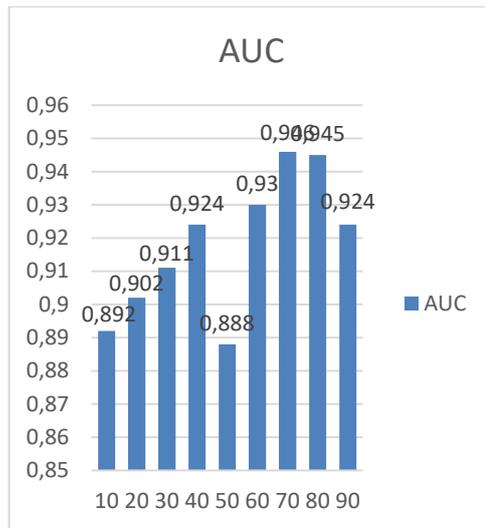


Gambar 2. Grafik Akurasi Naive Bayes

Berdasarkan gambar 2. dapat disimpulkan hasil tinjauan yang dilakukan oleh peneliti, yaitu

1. Hasil Akurasi terendah yaitu 89,90% dengan rasio data training 90%
2. Rata-rata Akurasi yaitu 92,03%
3. Hasil Akurasi tertinggi yaitu 93,54% dengan rasio data training 70%

Sedangkan hasil pengujian AUC pada naive bayes dapat dilihat pada gambar 3.



Gambar 3. Grafik AUC Naive Bayes

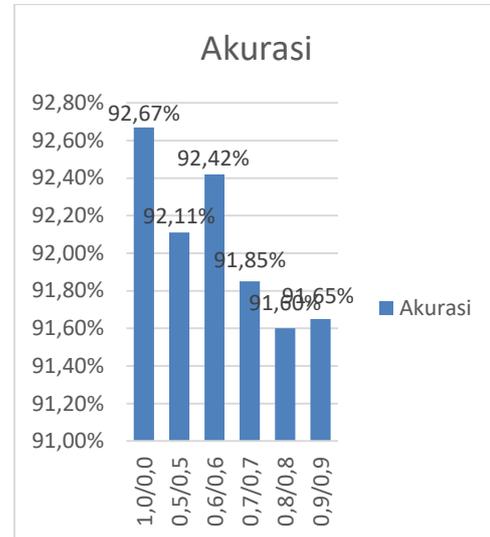
Berdasarkan gambar 3. dapat disimpulkan hasil tinjauan yang dilakukan oleh peneliti, yaitu

1. Hasil AUC terendah yaitu 0,888 dengan rasio data training 50% dengan diagnosa *Good Classification*
2. Rata-rata AUC yaitu 0,918 dengan diagnosa *Excellent Classification*
3. Hasil AUC tertinggi yaitu 0,945 dengan rasio data training 70% dengan diagnosa *Excellent Classification*

### Support Vector Machine

Dari pengujian akurasi *split validation* metode Support Vector Machine mempunyai rata-rata

akurasi dapat dilihat pada Gambar 4. berikut :

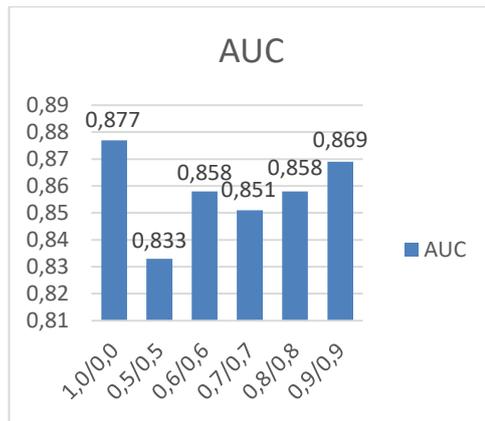


Gambar 4. Grafik Akurasi Support Vector Machine

Berdasarkan gambar 4 dapat disimpulkan hasil tinjauan yang dilakukan oleh peneliti, yaitu

1. Hasil Akurasi terendah yaitu 91,60% dengan nilai C=0,8 dan Epsilon=0,8
2. Rata-rata Akurasi yaitu 92,05%
3. Hasil Akurasi tertinggi yaitu 92,67% dengan nilai C=1,0 dan Epsilon=0,0

Sedangkan hasil pengujian AUC pada *Support Vector Machine* yaitu pada gambar 5.



Gambar 5. Grafik AUC Support Vector Machine

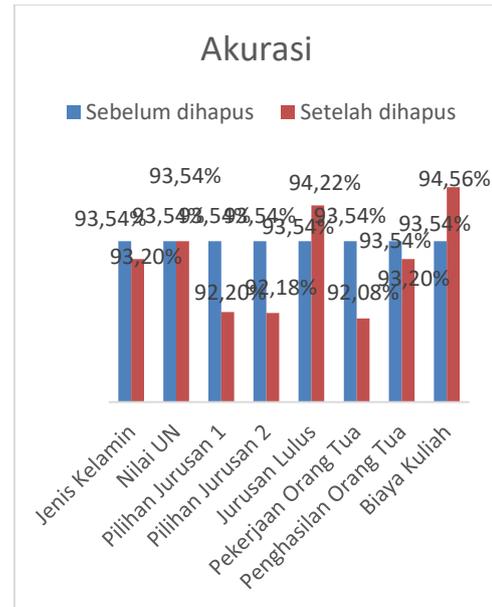
Berdasarkan gambar 5. dapat disimpulkan hasil tinjauan yang dilakukan oleh peneliti, yaitu

1. Hasil AUC terendah yaitu 0,833 dengan nilai  $C=0,5$  dan  $\text{Epsilon}=0,5$  % dengan diagnosa *Good Classification*
2. Rata-rata AUC yaitu 0,858 % dengan diagnosa *Good Classification*
3. Hasil AUC tertinggi yaitu 0,877 dengan nilai  $C=1,0$  dan  $\text{Epsilon}=0,0$  % dengan diagnosa *Good Classification*

**Pengujian Hapus Variabel Dengan Naive Bayes**

Pada tahapan pemilihan rasio yang dipilih untuk pengujian Naive Bayes yaitu 70% yang mana ini merupakan akurasi, dan AUC terbaik dari

pengujian Naive Bayes sebelumnya. Pengujian dilakukan dengan menghapus salah satu variabel secara bergantian maka didapatkan hasil seperti pada Gambar 6.

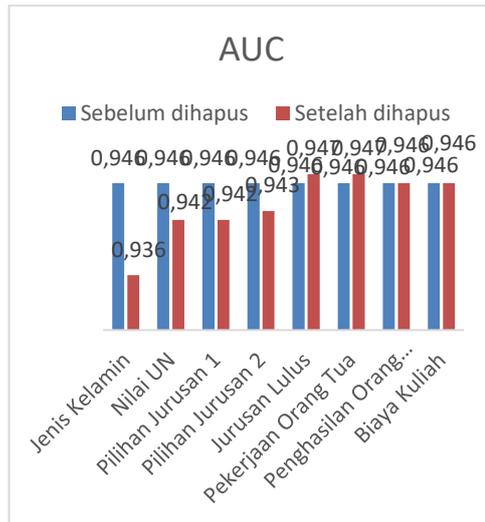


Gambar 6. Grafik Akurasi sebelum dan setelah dihapus variabel

Berdasarkan gambar 6 dapat disimpulkan hasil tinjauan yang dilakukan oleh peneliti, yaitu

1. Hasil akurasi akan semakin tinggi apabila variabel Biaya Kuliah dihapus yang sebelumnya 93,54% menjadi 94,56%
2. Hasil akurasi akan semakin rendah apabila variabel Pekerjaan Orang Tua dihapus yang sebelumnya 93,54% menjadi 92,08%

Selanjutnya hasil pengujian kurva AUC sebelum dan setelah dilakukan penghapusan variabel yang dapat dilihat pada gambar 7.



Gambar 7. Grafik hasil AUC sebelum dan setelah dihapus variabel

Berdasarkan gambar 7 dapat disimpulkan hasil tinjauan yang dilakukan oleh peneliti, yaitu

1. Hasil AUC akan semakin tinggi apabila variabel Jurusan Lulus dan Pekerjaan Orang Tua dihapus yang sebelumnya 0,946 menjadi 0,947.
2. Hasil AUC akan semakin rendah apabila variabel Jenis Kelamin dihapus yang sebelumnya 0,946 menjadi 0,936.

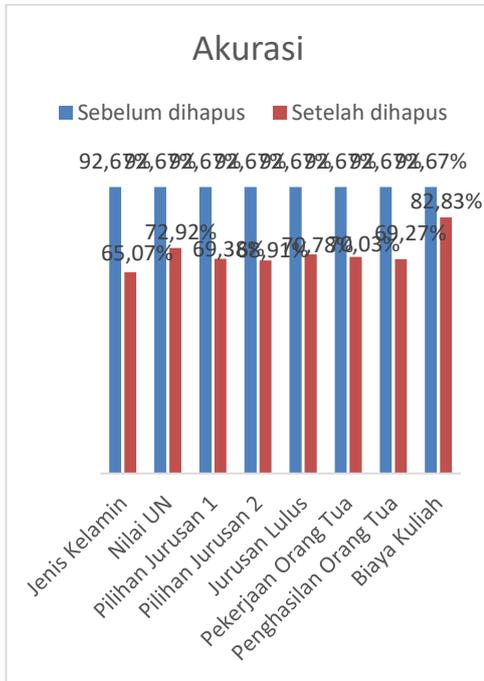
**Pengujian Hapus Variabel dengan support vector machine**

Pemilihan untuk skenario pengujian pada SVM yaitu menggunakan parameter C=1,0 dan epsilon=0,0 dikarenakan pengujian menggunakan parameter tersebut menghasilkan akurasi, dan AUC terbaik. Hasil akurasi tertinggi yaitu apabila dilakukan penghapusan variabel Biaya Kuliah yang memiliki akurasi 82,83%, sedangkan untuk akurasi terendah apabila dilakukan penghapusan terhadap variabel Jenis Kelamin dengan akurasi 65,07%. Untuk performa AUC tertinggi yaitu 0,766 yang dihasilkan dari penghapusan variabel Pekerjaan Orang Tua dan untuk performa AUC terendah yaitu 0,58 yang dihasilkan dari penghapusan variabel Biaya Kuliah.

Berdasarkan gambar 8 dapat disimpulkan hasil tinjauan yang dilakukan oleh peneliti, yaitu

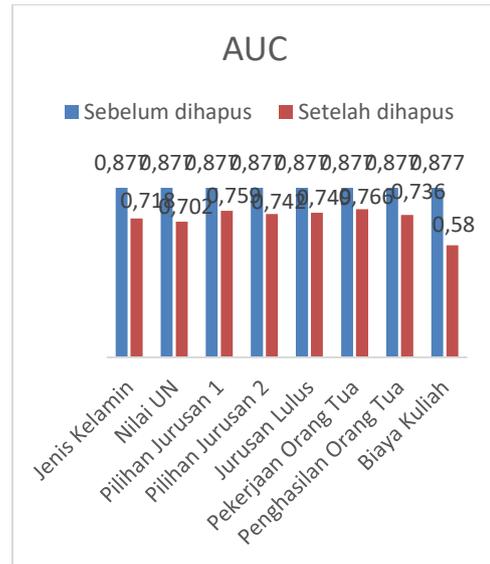
1. Hasil akurasi akan semakin rendah apabila dilakukan penghapusan terhadap salah satu variabel dibandingkan dengan sebelum variabel dihapus.
2. Hasil akurasi paling rendah apabila variabel Jenis Kelamin

dihapus yang sebelumnya  
 92,67% menjadi 65,07%



Gambar 8. Grafik Akurasi sebelum dan setelah dihapus variabel

Selanjutnya hasil pengujian kurva AUC sebelum dan setelah dilakukan penghapusan variabel yang dapat dilihat pada gambar 9.



Gambar 9. Grafik hasil AUC sebelum dan setelah dihapus variabel

Berdasarkan gambar 9 dapat disimpulkan hasil tinjauan yang dilakukan oleh peneliti, yaitu :

1. Hasil AUC akan semakin rendah apabila dilakukan penghapusan terhadap salah satu variabel dibandingkan dengan sebelum variabel dihapus.
2. Hasil AUC paling rendah apabila variabel Biaya Kuliah dihapus yang sebelumnya 0,877 menjadi 0,580.

#### IV. Kesimpulan dan Saran

##### 4.1 Kesimpulan

Berdasarkan hasil pengujian metode *naive bayes* dan *support vector machine* (SVM) terhadap prediksi herregistrasi yang telah dilakukan

maka dapat diambil beberapa kesimpulan, antara lain:

1. Atribut yang digunakan untuk memprediksi herregistrasi calon mahasiswa baru adalah jenis kelamin, nilai ujian nasional, gelombang pendaftaran, pilihan jurusan 1, pilihan jurusan 2, jurusan lulus, pekerjaan orang tua, penghasilan orang tua dan biaya kuliah.
2. Bahwa hasil pengujian metode *naive bayes* merupakan metode dengan kinerja terbaik dibandingkan metode *support vector machine* berdasarkan parameter Akurasi sebesar 93,54% dan AUC 0,946.
3. Dari pengujian yang sudah dilakukan bahwa performa akurasi, dan AUC sangat berpengaruh apabila dilakukan penghapusan terhadap salah satu variabel yang digunakan.
4. Pengujian data dalam penelitian ini adalah sebanyak 980 data, 720 data digunakan untuk data training dan 260 data untuk data testing.

Adapun kesimpulan dari penelitian ini sebagai berikut:

1. Tahapan penelitian yang dapat dilakukan guna memanfaatkan *command center record journal entry* dan *transaction journal* dengan menggunakan MyOB Accounting pada Home Industry Putra Abadi yaitu Tahapan studi pendahuluan dan Tahapan pengembangan.
2. Penelitian ini merupakan salah satu proses integrasi antara MyOB Accounting berbasis komputer dengan sistem pencatatan transaksi pada Home Industry Putra Abadi yang telah berjalan.

#### **4.2 Saran**

Dengan adanya penelitian tentang perbandingan SVM dan Naive Bayes, maka diharapkan mampu memberikan solusi bagi penelitian selanjutnya. Serta dapat membantu universitas dalam memprediksi herregistrasi bagi calon mahasiswa baru. Namun dalam penelitian ini penulis berharap dalam pengembangan penelitian ini dapat menambahkan beberapa variabel serta jumlah data dengan rentang tahun yang lebih banyak lagi. Selain itu dapat juga dilakukan perbandingan metode algoritma

lainnya yang akan mendukung pengujian data yang ada, sehingga mendapatkan tingkat akurasi yang lebih akurat.

## DAFTAR PUSTAKA

### Pustaka Buku

- Kusrini, & Emha T. Luthfi., 2009, *Algoritma Data Mining*, Andi Offset, Yogyakarta
- Prasetyo, E., 2012, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*, Andi Offset, Yogyakarta
- Prasetyo, E., 2014, *Data Mining: Mengolah Data menjadi Informasi Menggunakan Matlab*, Andi Offset, Yogyakarta
- Pudjo, W, P, Herlawati, Trias, H, R., 2013, *Penerapan Data Mining dengan Matlab*, Rekayasa Sains, Bandung
- Susanto, S & Suryadi, D., 2010, *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*, Andi Offset, Yogyakarta
- Witten, I, H, Frank, E, Hall, A, M., 2011, *Data Mining : Practical Machine Learning Tools and Technique*, 3<sup>rd</sup> ed., Asma Sthepan and Burlington, Eds. Morgan Kaufman, United States of America.

### Pustaka Majalah, Jurnal Ilmiah atau Prosiding

- Asogbon, G. M, Samuel, W. O, Omisore, O. M, Ojokoh, A. B, 2016, *A Multi-class Support Machine Approach for Students Academic Performance*

- Prediction, International Journal of Multidisciplinary and Current Reasearch*, ISSN: 2321-3124, Vol 4 March/April 2016
- Camba, J, David, E. R, Betan, A, Lagman, A, Caro, DL. J, 2016, *Student Analytics Using Support Vector Machines*, Information, Intelligence, Systems & Applications (IISA), **Doi: [10.1109/IISA.2016.7785425](https://doi.org/10.1109/IISA.2016.7785425)**
- Harwati, Ikha, V. R, Mansur, A, 2016, *Drop out Estimation Students based on the Study Period: Comparison between Naïve Bayes and Support Vector Machines Algorithm Methods*, IOP Conf. Series: Materials Science and Engineering, Doi: 10.1088/1757-899X/105/1/012039
- Prasetyo, E., 2016, *K-Support Vector Nearest Neighbor: Classification Method, Data Reduction, and Performance Comparison*, Journal of Electrical Engineering and Computer Sciences, ISSN: 2528-0260, Vol 1. 1 June, 2016
- Suprawoto, T., 2016, *Klasifikasi Data Mahasiswa Menggunakan Metode K-Means untuk Menunjang Pemilihan Strategi Pemasaran*, Jurnal Informatika dan Komputer, Vol 1, No 1, Februari 2016