



## Penerapan data mining dalam analisa profil mahasiswa menggunakan *metode support vector machines (SVM)*

Luki Hernando<sup>\*1</sup>, Vitri Aprilla Handayani<sup>2</sup>, Deosa Putra Caniago<sup>3</sup>, Nadia Widari Nasution<sup>4</sup>

Email: <sup>1</sup>luki@iteba.ac.id, <sup>2</sup>vitri@iteba.ac.id, <sup>3</sup>Deosa@iteba.ac.id, <sup>4</sup>Widari@iteba.ac.id

<sup>1,2,4</sup>Teknik Komputer, Fakultas Teknologi Informasi, Institut Teknologi Batam

<sup>2</sup>Matematika, Fakultas Teknologi Informasi, Institut Teknologi Batam

Diterima: 30 Juni 2023 | Direvisi: 06 September 2023 | Disetujui: 06 September 2023

©2023 Program Studi Teknik Informatika Fakultas Ilmu Komputer,

Universitas Muhammadiyah Riau, Indonesia

### Abstrak

Penerimaan calon mahasiswa baru setiap tahun semakin meningkat sehingga menghasilkan data yang cukup besar berupa data profil mahasiswa selama ini data profil mahasiswa yang mendaftar belum dimanfaatkan sebagaimana mestinya. Dengan banyaknya pendaftar di setiap tahunnya menimbulkan penumpukan data yang cukup besar sehingga terjadinya kesusahan dalam melakukan identifikasi terhadap profil mahasiswa yang mendaftar. Hal tersebut menimbulkan masalah yang harus segera diselesaikan oleh institusi atau pihak kampus dalam menentukan strategi promosi sehingga target mahasiswa yang diinginkan tercapai. Data mining merupakan proses analisa data dengan jumlah yang besar dalam menemukan pola-pola yang mempunyai arti bagi pendukung keputusan atau dalam pengambilan keputusan. Algoritma *Support Vector Machine* merupakan salah satu metode dalam *supervised learning* yang biasanya digunakan untuk klasifikasi seperti *Support Vector Classification* dan *regresi support Vector Regression*. Metode SVM merupakan metode terbaik dalam melakukan prediksi berdasarkan kerangka kerja statistik. Dalam penelitian ini algoritma SVM digunakan dalam klasifikasi profil data mahasiswa Institut Teknologi Batam, yang dapat digunakan sebagai bahan pertimbangan dalam menentukan dan pertimbangan dalam pengambilan keputusan guna membuat strategi daerah tujuan promosi yang efektif dan efisien.

**Kata kunci:** Data Mining, Knowledge Discovery Database, Support Vector Machine, Klasifikasi, Profil Mahasiswa

**The application of data mining in the analysis of student profiles using the vector machines (SVM) method**

### Abstract

*The acceptance of new student candidates every year is increasing so as to produce large enough data in the form of student profile data so far the profile data of students who register has not been utilized properly. With so many registrants every year, it causes a large accumulation of data so that there are difficulties in identifying the profile of students who register. This causes problems that must be resolved immediately by the institution or campus in determining the promotion strategy so that the desired student target is achieved. Data mining is the process of analyzing large amounts of data in finding patterns that have meaning for decision support or in decision making. The Support Vector Machine algorithm is one of the methods in supervised learning which is usually used for classification such as Support Vector Classification and Support Vector Regression. The SVM method is the best method in making predictions based on a statistical framework. In this research, the SVM algorithm is used in the classification of student data profiles of Batam Institute of Technology, which can be used as a consideration in determining and considering decision making in order to make an effective and efficient promotion destination strategy.*

**Keywords:** Data Mining, Knowledge Discovery Database, Support Vector Machines, Classification, Student Profile

## 1. PENDAHULUAN

Teknologi semakin berkembang seiring bertambahnya tahun, perkembangan teknologi dapat dilihat dari beberapa aspek seperti di bidang teknologi infomasi, sosial, ekonomi dan budaya. Perkembangan teknologi informasi sangatlah pesat di era sekarang ini yang dapat membantu dalam mempermudah pekerjaan dalam pengelolaan informasi [1]. Penerapan teknologi dapat membantu masyarakat dalam mengolah data sehingga menjadi informasi yang lebih akurat [2]. Dengan berkembangnya teknologi informasi sangat memberikan dampak atau keuntungan bagi masyarakat seperti masyarakat dengan sangat mudah dan cepat dalam mendapatkan informasi yang diinginkan [3]. Perkembangan teknologi semakin mengarahkan perkembangan ke dunia digital dalam hal ini perkembangan teknologi mencakup beberapa aspek salah satunya teknologi internet merupakan salah wajib dalam kemajuan zaman saat ini [4].

Metode dan teknik *data mining* sangatlah bervariasi, *data mining* merupakan sebuah proses yang digunakan dalam mencari sebuah pola juga bisa mencari infomasi yang menarik terhadap data yang terpilih [5]. *Data Mining* juga bertujuan menarik abstrak ataupun informasi yang didapatkan dari *database* yang besar dan mampu menganalisa data dalam jumlah yang sangat besar sehingga infomasi yang didapatkan berupa pola yang mempunyai nilai dalam mengambil keputusan [6].

*Data mining* memiliki konsep dasar yaitu menentukan informasi yang tersembunyi dari *database* yang merupakan bagian dari KDD (*Knowledge Discovery in Databases*) yang bertujuan dalam menentukan informasi dan sebuah pola yang berguna dalam data. Data Mining dapat mencari sekumpulan informasi data dengan menggunakan teknologi komputer dan manusia yang bersifat iteratif yang melalui proses manual maupun otomatis [7].

KDD (*Knowledge Discovery In Databases*) merupakan proses terorganisir yang digunakan untuk identifikasi pola data baru, valid data berguna dan dapat mengerti dari kumpulan data yang sangat besar dan kompleks. Intinya data mining merupakan sebuah tahapan yang menyimpulkan dan melibatkan algoritma yang dapat mengeksplorasi data dan mengembangkan mode juga pola pola sebelumnya yang belum diketahui [8].

KDD (*Knowledge Discovery Databases*) sering terhalang dikarenakan tantangan dari suatu navigasi dan investasi data-data yang berbeda [9]. *Preprocessing* dan *post processing* merupakan serangkaian langkah-langkah dari perubahan KDD. KDD merupakan sebuah proses yang digunakan dalam menemukan infomasi, pengetahuan dan manfaat dari suatu data yang terpilih [10].

Institut Teknologi Batam (ITEBA) merupakan salah satu perguruan tinggi yang menerima mahasiswa di setiap tahunnya dari berbagai sekolah untuk program studi Teknik Industri, Manajemen Rekayasa, Matematika, Desain Komunikasi Visual, Sistem Informasi dan Teknik Komputer. Dari tahun 2018 sampai dengan tahun 2022 jumlah mahasiswa baru untuk masing-masing program studi cukup bervariasi di setiap tahunnya, ada yang mengalami kenaikan dan ada yang mengalami penurunan. Jumlah kenaikan mahasiswa baru tentunya akan membawa dampak baik untuk perguruan tinggi Institut Teknologi Batam untuk ke depannya. Sedangkan untuk penurunan jumlah mahasiswa baru tentunya membawa dampak bagi Institut Teknologi Batam sebab tidak memenuhi kuota mahasiswa baru sesuai dengan target. Penelitian ini bertujuan untuk mengetahui asal sekolah mahasiswa yang masuk ke Institut Teknologi Batam dengan menggunakan *Data Mining* dengan metode *Support Vector Machines* (SVM) guna untuk memperoleh profil mahasiswa ITEBA, sehingga mendapatkan profil mahasiswa yang lebih akurat sehingga dapat menjadi acuan dalam pertimbangan guna pengambilan keputusan yang lebih akurat dalam penentuan sekolah untuk target promosi.

## 2. METODE PENELITIAN

Di penelitian ini, peneliti menggunakan metode Support Vector Machines (SVM) untuk klasifikasi profil mahasiswa Institut Teknologi Batam dengan metodologi Knowledge Discovery in database (KDD). Adapun alur dari penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

## 2.1. Metode SVM (*Support Vector Machines*)

Metode SVM adalah metode terbaik yang digunakan dalam melakukan prediksi berdasarkan kerangka kerja statistik yang dikembangkan oleh seorang peneliti yang bernama Chervonenkis dan Vapnik. SVM di dalam machine learning merupakan salah satu algoritme pembelajaran yang terawasi sehingga dapat digunakan dalam melakukan analisis untuk kebutuhan klasifikasi dan regresi[11].

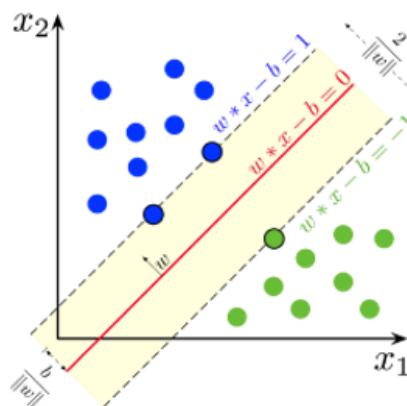
Pada tahun 1992 Vapnik merupakan seorang peneliti yang pertama kali memperkenalkan metode *support vector machines*, ia mengemukakan metode SVM adalah salah satu metode dalam *machine learning* yang bekerja menggunakan prinsip SRM (*Structural Risk Minimization*) yang bertujuan dalam menemukan *hyper plane* yang terbaik dalam memisahkan antara dua buah *class* yang terdapat di dalam *input space*. Metode ini sangat bergantung kepada fungsi parameter dan kernel yang digunakan [12].

## 2.2 Linear Support Vector Machines (SVM)

SVM adalah salah satu dari algoritma terbaik dalam penyelesaian masalah klasifikasi. Metode ini bekerja dengan cara membagi kelas data menggunakan garis pemisah atau *hyper plane* yang dinyatakan dalam persamaan.

$$w \cdot x + b = 0 \quad (1)$$

Hyper Plane dalam metode ini terdiri dari satu garis pemisah utama dan dua garis pendukung yang dapat dilihat pada gambar



Gambar 2. Hyper Plane Pada Metode Support Vector Machines

Metode *Support Vector Machines* membagi dataset menjadi dua kelas yaitu kelas pertama yang dipisah oleh hyper plane bernilai 1, sedangkan kelas lainnya bernilai -1.

$$x_i \cdot w + b \geq 1, \text{ Jika } y_i = 1 \quad (2)$$

$$x_i \cdot w + b \geq -1, \text{ Jika } y_i = -1 \quad (3)$$

Keterangan :

$x_i$  = Data ke-i

$w$  = Nilai bobot support vector yang tegak lurus dalam *hyper plane*

$b$  = Nilai Bias

$y_i$  = Kelas data ke -i

Bobot dari vector ( $w$ ) merupakan garis vector yang tegak lurus di antara titik pusat koordinat dengan garis hyper plane. Dan bias ( $b$ ) adalah koordinat garis yang relative terhadap titik koordinat. Untuk menghitung nilai  $b$  bisa dilihat pada rumus persamaan 4, sedangkan untuk mencari nilai  $w$  bisa dilihat di persamaan 5.

$$b = -\frac{1}{2} (w \cdot x^+ + w \cdot x^-) \quad (4)$$

$$w = \sum_{i=1}^n a_i y_i x_i \quad (5)$$

Keterangan :

$b$  = nilai bias

$w \cdot x^+$  = nilai bobot untuk kelas data positif

$w \cdot x^-$  = nilai bobot untuk kelas data negatif

$a_i$  = nilai bobot data ke-i

$y_i$  = nilai bobot data ke-i

$x_i$  = data ke-i

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Analisa Data

Pengumpulan data merupakan sebuah tahapan yang terpenting dalam metode riset. Dalam proses pengumpulan data dengan mengumpulkan dan merakap data mahasiswa yang mendaftar dari tahun 2019 sampai 2022.

No	Tanggal Dafta	Nama Lengkap	Prodi Pilihan	No Handphone	Email	Nama Sekolah	Provinsi Asal Sekolah
1	14/10/2018	Angelia	Desain Komunikasi Visual	089619953465	<a href="mailto:angelia202998@gmail.com">angelia202998@gmail.com</a>	SMA Yos Sudarso	Kepulauan Riau
2	15/10/2018	Calvin Januardy	Desain Komunikasi Visual	082294334452	<a href="mailto:calvinjanuardy01@gmail.com">calvinjanuardy01@gmail.com</a>	SMK Maitreyawira	Kepulauan Riau
3	27/10/2018	Riky Murtiadi	Desain Komunikasi Visual	087894246952	<a href="mailto:murtiadirkky@gmail.com">murtiadirkky@gmail.com</a>	SMK Kartini Batam	Kepulauan Riau
4	29/10/2018	Rahima Aisyah Putri	Desain Komunikasi Visual	085103024288	<a href="mailto:rahimahaisyah29@gmail.com">rahimahaisyah29@gmail.com</a>	SMK Kartini Batam	Kepulauan Riau
5	02/11/2018	Levianny	Desain Komunikasi Visual	082390979610	<a href="mailto:levichen1808@gmail.com">levichen1808@gmail.com</a>	SMK Harmoni	Kepulauan Riau
6	02/11/2018	Ferdinna Yunicia	Desain Komunikasi Visual	0895355802001	<a href="mailto:ferdinnavync@gmail.com">ferdinnavync@gmail.com</a>	SMK Kartini Batam	Kepulauan Riau
7	04/12/2018	Rio Alvian Trihandanu	Desain Komunikasi Visual	089673964619	<a href="mailto:uzumaki.danu@gmail.com">uzumaki.danu@gmail.com</a>	SMKN 4 Batam	Kepulauan Riau
8	09/01/2019	Imam Akbar Syah Yasmindaniz Nasution	Desain Komunikasi Visual	085977821559	<a href="mailto:iyasmindaniz@gmail.com">iyasmindaniz@gmail.com</a>	SMK Kartini Batam	Kepulauan Riau
9	13/01/2019	Deniyan Saputra	Desain Komunikasi Visual	08136506702	<a href="mailto:saputradeniyan@gmail.com">saputradeniyan@gmail.com</a>	SMK Ibu Sina Batam	Kepulauan Riau
10	29/01/2019	Elsi Susanti	Desain Komunikasi Visual	081908302983	<a href="mailto:susanti.elsi2710@gmail.com">susanti.elsi2710@gmail.com</a>	SMKN 4 Batam	Kepulauan Riau
11	30/01/2019	Antony	Desain Komunikasi Visual	082388555521	<a href="mailto:zeenxlash@gmail.com">zeenxlash@gmail.com</a>	SMK Permata Harapan	Kepulauan Riau
12	31/01/2019	Leo Bryan Arief	Desain Komunikasi Visual	082173281881	<a href="mailto:bryanarie22@gmail.com">bryanarie22@gmail.com</a>	SMK Kartini Batam	Kepulauan Riau
13	18/02/2019	M. Rizqiansyah Haqiqi Ali	Desain Komunikasi Visual	087791101128	<a href="mailto:rizqiansyah.haqiqi@gmail.com">rizqiansyah.haqiqi@gmail.com</a>	SMK Kartini Batam	Kepulauan Riau
14	18/02/2019	Alexander Daniel K	Desain Komunikasi Visual	0895629421038	<a href="mailto:danielkroberick32@gmail.com">danielkroberick32@gmail.com</a>	SMK Kartini Batam	Kepulauan Riau
15	18/02/2019	Dendita Tarigan	Desain Komunikasi Visual	0895365936935	<a href="mailto:dendita9905@gmail.com">dendita9905@gmail.com</a>	SMK Kartini Batam	Kepulauan Riau
16	20/02/2019	Akbar Muhammad Lelang	Desain Komunikasi Visual	081267915559	<a href="mailto:akbarmuhammad2334@gmail.com">akbarmuhammad2334@gmail.com</a>	SMK Kartini Batam	Kepulauan Riau

Gambar 3 Rekapitulasi Pendaftaran

#### 3.2 Atribut dalam penelitian

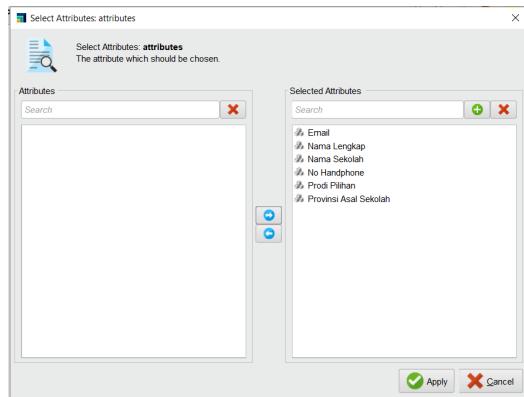
Atribut yang digunakan di dalam penelitian ini ada 7 atribut yaitu tanggal daftar, nama lengkap, prodi pilihan, no headphone, email, nama sekolah, provinsi asal sekolah. Deskripsi dari data set tersebut dapat yang dapat dilihat pada tabel 1.

Tabel 1. Atribut data pendaftaran mahasiswa baru

No	Attribute	Type	Description
1.	Tanggal Daftar	Teks	Tanggal Pendaftaran Mahasiswa
2.	Nama Lengkap	Teks	Nama Mahasiswa Baru
3.	Prodi Pilihan	Teks	Prodi Yang dipilih mahasiswa
4.	No Headphone	Numeric	No Headphone Mahasiswa
5.	Email	Teks	Email mahasiswa
6.	Nama Sekolah	Categorical	Asal Sekolah mahasiswa
7	Provinsi Asal Sekolah	Teks	Provinsi asal sekolah mahasiswa

#### 3.3 Data Selection Process

Data profil mahasiswa akan dilakukan pengujian dengan cara pengumpulan data, pemecahan jenis data dengan menggunakan *cross validation*. Pada tahapan pemecahan data dibagi menjadi dua jenis data yaitu data yang digunakan dalam melatih algoritme dan menguji agoritma tersebut. Setalah itu dilakukan pengujian data, setelah itu dilanjutkan ke tahap analisa data secara terstruktur dengan mengenali pola-pola yang kompleks dan lengka, kemudian dilakukan pemilihan data yang dibutuhkan setelah data yang dibutuhkan didapatkan maka dilanjutkan dengan menghilangkan data yang *missing value*.

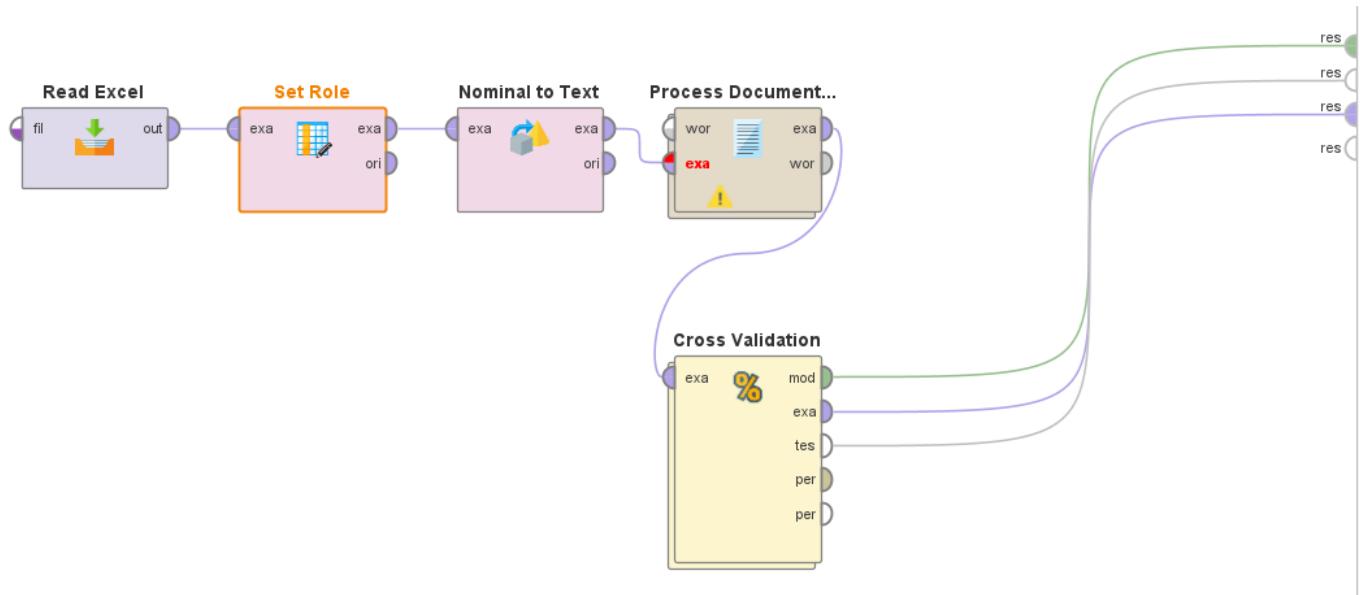


Gambar 4. Proses pemilihan Atribut

Proses pemilihan data yang dilakukan menggunakan *parameter*, dimana langkah yang dilakukan adalah memilih kolom attribute yang akan digunakan sebagai target.

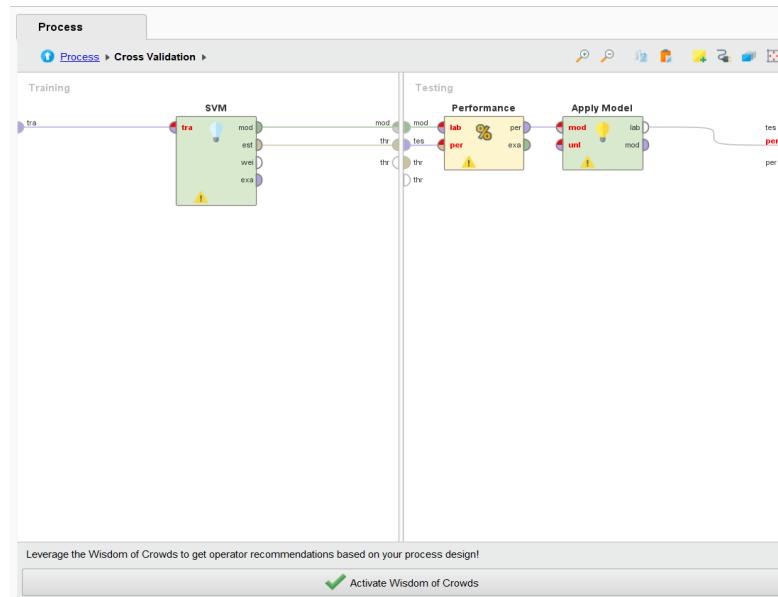
### 3.4 Pengujian Metode *Support Vector Machines* (SVM)

Data yang digunakan dalam analisa profil mahasiswa merupakan data dari periode 2019 sampai dengan periode 2022. Dalam penelitian ini menggunakan *software* dalam menganalisa profil mahasiswa menggunakan metode *Support Vector Machine* (SVM). Proses pengujian Metode SVM dapat dilihat pada gambar 4.



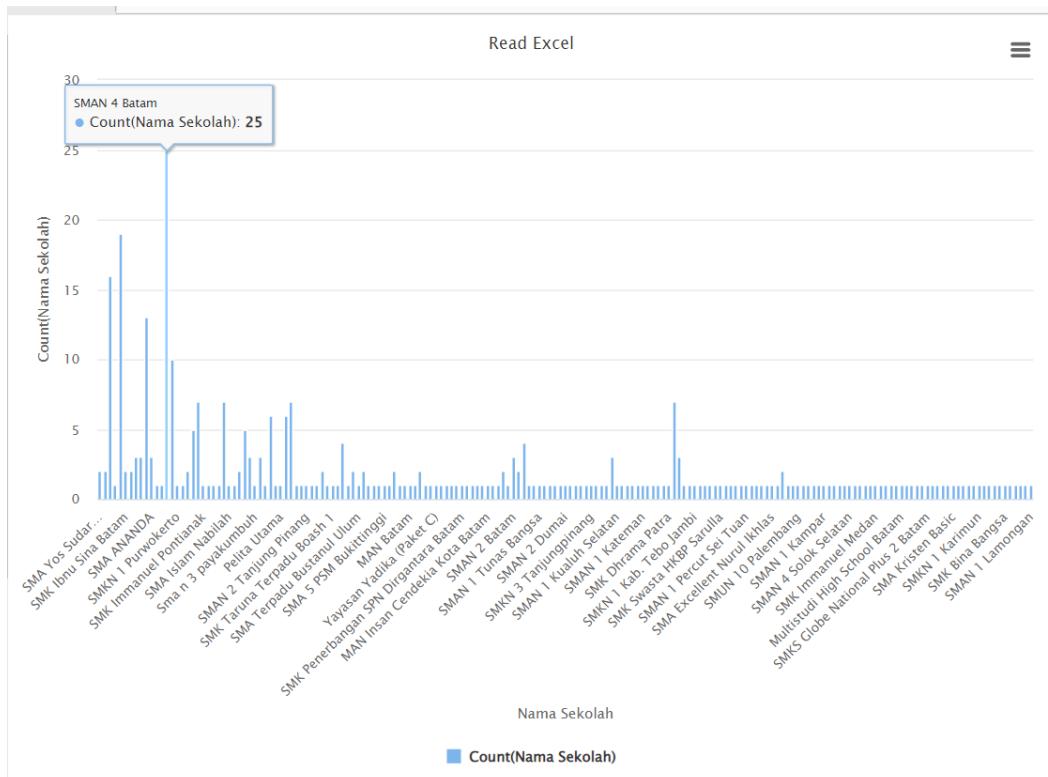
Gambar 4. Proses Pengujian Metode SVM

Data yang sudah di import tersebut dilakukanlah validasi dengan menggunakan *Cross validation*, teknik validasi data yang bertujuan agar tidak terjadi bias terhadap data serta juga dapat meningkatkan kinerja akurasi model aplikasi prediktif yang digunakan untuk menganalisa.



Gambar 5. Model Proses Algoritma SVM

Proses yang berada di dalam cross validation yaitu algoritma SVM yang digunakan dalam menganalisa profil mahasiswa dimana model SVM akan melakukan proses model data testing yang digunakan sebagai data prediksi.

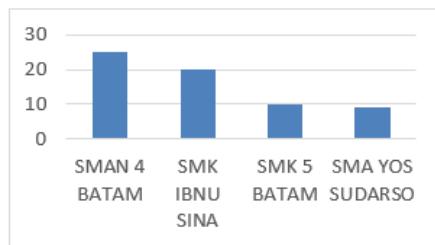


Gambar 6. Analisa Grafik Asal Sekolah

Analisa profil mahasiswa menggunakan metode *Support Vector Machine* (SVM) dengan menggunakan target atribut asal sekolah. Berdasarkan hasil dari pengolahan data dengan menggunakan metode SVM maka didapatkan beberapa sekolah yang akan menjadi target promosi. Analisa dengan SVM berdasarkan hasil analisis maka didapatkan data sekolah yang mengalami peningkatan dari 25 asal sekolah mahasiswa yaitu SMAN 4 Batam, SMK Ibnu sina dan SMA Yos Sudarso. Dari hasil analisa ini dapat disimpulkan bahwasanya ada peningkatan asal sekolah mahasiswa setiap tahunnya. SMAN 4 Batam mengalami peningkatan yang signifikan yaitu 75 % .

### 3.5 Evaluasi Akurasi SVM

Menginformasikan area under curve (AUC) dari semua algoritma klasifikasi yang digunakan dimana dapat dilihat bahwa algoritma SVM memiliki nilai AUC paling tinggi yaitu 25, data akurasi SVM diambil sampel 4 sekolah tertinggi yang bisa di lihat pada grafik AUC dari Perhitungan Metode SVM.



Gambar 7.Grafik AUC klasifikasi Algoritma SVM

## 4. KESIMPULAN

Berdasarkan analisa profil mahasiswa di Institut Teknologi Batam maka dapat disimpulkan bawa dengan menggunakan metode data mining dengan algoritma *Support Vector Machine* (SVM) menghasilkan informasi mengenai asal sekolah mahasiswa mulai dari jumlah mahasiswa terbanyak sampai dengan jumlah yang paling sedikit. Hasil dari *Clustering* menunjukkan bahwa ada beberapa sekolah yang mendominasi mahasiswa terbanyak. Penelitian ini dapat menjadi pertimbangan bagi pimpinan dalam menentukan target promosi sekolah-sekolah, dari hasil evaluasi tersebut bisa diketahui sekolah mana yang paling banyak siswa yang kuliah di ITEBA, dari penelitian ini didapatkan SMA 4 Batam dengan jumlah siswa 30 orang, SMK Ibnu Sina 20 Orang, SMK 5 Batam 10 Dan SMA Yos Sudarso 9 Orang, dari data tersebut bisa didapatkan rata-rata mahasiswa yang kuliah di ITEBA banyak berasal dari SMAN 4 Batam.

## DAFTAR PUSTAKA

- [1] I. A. Huda, "Perkembangan Teknologi Informasi Dan Komunikasi (Tik) Terhadap Kualitas Pembelajaran Di Sekolah Dasar," *J. Pendidik. dan Konseling*, vol. 2, no. 1, pp. 121–125, 2020, doi: 10.31004/jpdk.v1i2.622.
- [2] Yuli Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ) . Jurnal Edik Informatika," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.
- [3] A. Maria, Luki, Yopy, "Perancangan aplikasi pengantaran barang berbasis mobile," vol. 3, no. 3, pp. 371–381, 2022.
- [4] D. P. Caniago, T. Komputer, and F. T. Informasi, "Jurnal Computer Science and Information Technology ( CoSciTech ) Aplikasi Internet of Things pada Kotak Cerdas Penerima Tugas Mahasiswa menggunakan Internet of Things application on Student Assignee Smart Box using ESP32-Cam," vol. 3, no. 3, pp. 479–486, 2022.
- [5] Di. P. Mulya, "Analisa Dan Implementasi Association Rule Dengan Algoritma Fp-Growth Dalam Seleksi Pembelian Tanah Liat (Studi Kasus Di Pt. Anveve Ismi Berjaya)," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 1, no. 1, pp. 47–57, 2019, doi: 10.47233/jtekisis.v1i1.6.
- [6] R. Asmara and D. Saputra, "Jurnal J – Click Jurnal J – Click," *J-Click*, vol. 6, no. 2, pp. 201–207, 2019.
- [7] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, "Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang," *J. Janitra Inform. dan Sist. Inf.*, vol. 2, no. 1, pp. 9–19, 2022, doi: 10.25008/janitra.v2i1.142.
- [8] A. Darmawan, N. Kustian, and W. Rahayu, "Implementasi Data Mining Menggunakan Model SVM untuk Prediksi Kepuasan Pengunjung Taman Tabebuya," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 2, no. 3, p. 299, 2018, doi: 10.30998/string.v2i3.2439.
- [9] I. K. Juni Arta, G. Indrawan, and G. R. Dantes, "Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di Stmk Denpasar Menggunakan Metode Technique for Others Reference By Similarity To Ideal Solution," *JST (Jurnal Sains dan Teknol.)*, vol. 5, no. 2, pp. 11–21, 2017, doi: 10.23887/jstundiksha.v5i2.8549.
- [10] Karsito and W. Monika Sari, "Prediksi Potensi Penjualan Produk Delifrance Dengan Metode Naive Bayes Di Pt. Pangan Lestari," *J. Teknol. Pelita Bangsa*, vol. 9, no. 1, pp. 67–78, 2018.
- [11] B. A. Nugroho, A. Kurnia, A. Pradana, and E. Nurfarida, "Prediksi Waktu Kedatangan Pelanggan Servis Kendaraan Bermotor Berdasarkan Data Historis menggunakan Support Vector Machine," vol. 7, no. 1, pp. 25–30, 2021.
- [12] I. M. Parapat and M. T. Furqon, "Penerapan Metode Support Vector Machine ( SVM ) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," vol. 2, no. 10, pp. 3163–3169, 2018.