

Penerapan Algoritma *Support Vector Machine* (SVM) Dalam Memprediksi Calon Mahasiswa Baru Pendidikan Matematika

Hanum Sal Sabila¹, Atika Zahra^{2*}

^{1,2}UIN Raden Fatah Palembang, Palembang

*Email korespondensi: Atikazahra_uin@radenfatah.ac.id

ABSTRAK

Banyaknya jumlah mahasiswa baru yang masuk dijadikan sebagai kesuksesan atau kemajuan dari suatu perguruan tinggi. Semakin meningkat jumlah mahasiswa baru, maka semakin baik keberadaan perguruan tinggi tersebut. Oleh karena itu, penting untuk memprediksikan jumlah mahasiswa baru pada tahun berikutnya. Namun dalam melakukan pemrediksian metode yang digunakan harus minim dengan kesalahan peramalannya sendiri. Salah satu metode prediksi yang dapat digunakan adalah *Support Vector Machine* (SVM). Dari hasil penerapan algoritma *Support Vector Machine* (SVM) dengan menggunakan kernel *Radial Basic Function* (RBF) dan parameter $C = 1$ dapat disimpulkan bahwa banyaknya calon mahasiswa pada tahun 2024 adalah 104 calon mahasiswa. Berdasarkan hasil evaluasi yang dilakukan dengan menggunakan metode *forecasting Mean Absolute Percentage Error* (MAPE) dimana nilai yang didapat adalah sebesar 17%. Maka dapat dikatakan bahwa *Support Vector Machine* (SVM) memiliki kemampuan model peramalan yang baik dalam memprediksi calon mahasiswa baru pendidikan matematika.

Kata kunci: Support Vector Machine, Mahasiswa Baru, MAPE.

PENDAHULUAN

Perguruan tinggi merupakan penyelenggara akademik bagi mahasiswa, yang diharapkan dapat menghasilkan sumberdaya manusia yang berilmu, cakap dan kreatif untuk mendukung tercapainya pembangunan nasional. Tujuan dari perguruan tinggi merupakan tempat akademi untuk mempersiapkan para mahasiswa menjadi masyarakat yang memiliki pengetahuan yang lebih luas dan dapat mengimplementasikan ilmu yang sudah dimiliki. Keberhasilan atau prestasi belajar mahasiswa sering dijadikan sebagai kesuksesan dan keunggulan dari perguruan tinggi (Nurhayati et al., 2015). Banyaknya jumlah mahasiswa baru yang masuk juga dijadikan sebagai kesuksesan atau kemajuan dari perguruan tinggi. Semakin meningkat jumlah mahasiswa baru, maka semakin baik keberadaan perguruan tinggi tersebut (Kesuma et al., 2022). Oleh karena itu, penting untuk memprediksikan jumlah mahasiswa baru pada tahun berikutnya. Namun dalam melakukan pemrediksian metode yang digunakan harus minim dengan kesalahan peramalannya sendiri. Salah satu metode prediksi yang dapat digunakan adalah *Support Vector Machine* (SVM).

Support Vector Machine (SVM) adalah algoritma dengan pencarian hyperlane terbaik yang memisahkan dua kelas dari input space (Saputro, 2015). Algoritma adalah urutan langkah-langkah untuk memecahkan masalah yang disusun secara sistematis dan logis (Dianta, 2021 dalam Thoyyibah et al., 2014). Algoritma SVM ini dianggap sebagai salah satu teknik klasifikasi yang paling efektif dalam memprediksi pengklasifikasian potensial pada data (Hovi et al., 2022). Sebagai salah satu teknik klasifikasi data, SVM

telah banyak digunakan. Antara lain digunakan sebagai klasifikasi akreditasi sekolah, klasifikasi kualitas pengelasan *shield metal arc welding*, klasifikasi penyimpangan tumbuh kembang anak dan klasifikasi pasang usia subur (Damasela et al., 2022). Selain itu, SVM juga telah digunakan untuk mendeteksi atau diagnosis penyakit diabetes, mendeteksi penyalahgunaan narkoba dan memprediksi harga saham.

Support Vector Machine (SVM) mempunyai keunggulan yaitu memiliki error generelasi yang lebih kecil, memiliki landasan teori yang dapat dianalisis dengan jelas dan tidak bersifat black box, dan tingkat generalisasi yang diperoleh oleh SVM tidak dipengaruhi oleh dimensi dari input vector (Nugroho et al., 2003). Selain itu, kelebihan dari *Support Vector Machine* (SVM) ini adalah memiliki tingkat akurasi yang tinggi dan tingkat kesalahan yang relative kecil, kemampuan untuk mengatasi *overfitting*, tidak membutuhkan data yang terlalu besar dan dapat digunakan untuk melakukan prediksi (Harafani & Wahono, 2015).

Peneliti (Nurhayati et al., 2015) pernah meramalkan tentang memprediksi mahasiswa *drop out* menggunakan metode *Support Vector Machine*. Dalam penelitiannya, peneliti melakukan prediksi mahasiswa drop out dengan menggunakan variabel input data individu dan evaluasi IP dan IPK mahasiswa dengan variabel output mahasiswa drop out dan non-drop out. Dimana hasil dari penelitian adalah diperoleh jumlah prediksi yang benar sama sebanyak 304 set data dan jumlah prediksi yang tidak sama sebanyak 6 set data dengan tingkat akurasi sebesar 98,06% dan nilai error sebesar 0,0193.

Penelitian lainnya pernah dilakukan oleh (Saputro, 2015) yang menggunakan metode *Support Vector Machine* (SVM) untuk memprediksi harga emas yang membedakan antara variabel A dan B, diketahui bahwa variabel A menghasilkan nilai *Root Mean Squared Error* (RMSE) 4,695 dan variabel B nilai RMSE adalah 4,620. Dengan hasil tersebut *Root Mean Squared Error* (RMSE) yang telah didapat, maka variabel B (*open, high, low, close* dan *factory news*) dapat meningkatkan hasil prediksi.

Dari bahasan diatas peneliti tertarik untuk melakukan penelitian yang menjadi bahan seminar matematika dengan judul “Penerapan Algoritma *Support Vector Machine* (SVM) Dalam Memprediksi Calon Mahasiswa Baru Pendidikan Matematika UIN Raden Fatah Palembang”.

METODE

Jenis penelitian yang digunakan oleh peneliti adalah penelitian kuantitatif dan deskriptif. Dalam melakukan penelitian diperlukan beberapa tahapan yang harus dilakukan. Adapun tahapan yang digunakan pada penelitian ini adalah Knowledge Discovery Database (Nofriansyah & Nurcahyo, 2019) sebagai berikut:

1. Seleksi Data

Pada tahapan ini kita melakukan pengumpulan data yang di ambil dari Sistem Informasi Layanan Data Terpadu (SILADAS) Universitas Raden Fatah Palembang.

2. Pemilihan data (preprocessing)

Dari hasil data yang dikumpulkan masih berisikan null sehingga diperlukan proses pemilihan data. Pada tahapan dari data yang telah dikumpulkan dilakukan pemilihan atribut yang akan dipilih dan menghapus atribut yang tidak akan kita pakai. Memperbaiki kesalahan pada data, juga melakukan proses enrichment dimana data yang sudah ada ditambahkan dengan data atau informasi lain yang relevan.

3. Transformasi

Pada fase ini dilakukan proses mentransformasi data yang telah terpilih kedalam bentuk yang cocok untuk prosedur penggalian lebih lanjut dengan cara melakukan proses agregasi.

4. Data Mining

Data yang telah melewati tahapan preprocessing dilakukan processing data menggunakan Metode *Support Vector Machine* dengan menentukan nilai *Support vector* dengan bantuan dari aplikasi Orange3.

5. Evaluasi

Pada tahapan ini adalah pengolahan keseluruhan data menggunakan bantuan excel dengan menggunakan teknik *Forecast Errors* untuk mencari hasil dari MAE, RSME dan MAPE. Validitas hasil dari metode yang digunakan dievaluasi berdasarkan nilai-nilai tersebut.

HASIL DAN PEMBAHASAN

1. Seleksi Data

Peneliti mengumpulkan data yang diambil dari Sistem Informasi Layanan Data Terpadu (SILADAS) Universitas Islam Negeri Raden Fatah Palembang. SILADAS adalah portal data terpadu dari Universitas Islam Negeri Raden Fatah Palembang yang menyajikan data-data dari seluruh prodi dan unit kerja Universitas Islam Negeri Raden Fatah Palembang. Data yang dapat diakses melalui SILADAS secara publik sangat banyak sehingga kita memerlukan penseleksian data. Dalam penelitian ini data yang dibutuhkan adalah data mahasiswa baru dari prodi pendidikan matematika. Data yang didapatkan terdiri dari tahun, kuota, daftar, lulus dan registrasi yang dapat dilihat pada tabel dibawah ini.

Tabel 1. Data mahasiswa baru pendidikan matematika 2013 – 2023

No	Tahun	Kuota	Daftar	Lulus	Registrasi
1.	2013	0	397	134	90
2.	2014	0	792	175	121
3.	2015	0	238	166	130
4.	2016	0	644	137	113
5.	2017	80	344	142	107
6.	2018	80	283	184	138
7.	2019	140	299	147	115
8.	2020	120	292	121	101
9.	2021	90	150	104	80
10.	2022	120	229	81	63
11.	2023	120	396	88	69

2. Pemilihan data

Data yang diperoleh dari SILADAS sebanyak 12 data, namun tidak semua data dan atribut digunakan maka dilakukan pemilihan data dan pembersihan data. Pemilihan dan pembersihan data sesuai dengan kebutuhan penelitian, di karena pada pada penelitian ini peneliti berfokus pada data ketika masa covid sampai sekarang sehingga data yang diambil hanya pada data tahun 2019 – 2023 dan dengan menghapus atribut registrasi karena pada penelitian ini kita akan memprediksi yang lulus. Pada tabel 2 merupakan data yang sudah melalui proses *cleaning*.

Tabel 2. Data setelah proses *cleaning*

No	Tahun	Kuota	Daftar	Lulus
----	-------	-------	--------	-------

1.	2019	140	299	147
2.	2020	120	292	121
3.	2021	90	150	104
4.	2022	120	229	81
5.	2023	120	396	88

3. Transformasi data

Transformasi data yang dilakukan merupakan proses pengubahan data ke dalam format CSV sesuai dengan format dari Orange3. Data yang di dapatkan melalui website SILADAS masih dalam format xls kemudian di ubah ke dalam format CSV, format CSV dapat di lihat pada gambar 1 dibawah ini.

	A	B	C	D	E
1	Tahun,Kuota,Daftar,Lulus				
2	2019,140,299,147				
3	2020,120,292,121				
4	2021,90,150,104				
5	2022,120,229,81				
6	2023,120,396,88				

Gambar 1 Data dengan format CSV

4. Data Mining

Proses data mining ini dilakukan dengan bantuan aplikasi Orange3 yang merupakan perangkat lunak untuk melakukan proses data mining atau data analytic melalui konsep *visual programming*. Proses data mining dimulai dengan membagi data menjadi dua, yaitu data *training* dan data *testing*. Pada data *training* yang terdiri dari beberapa variabel dan kelas target yang dijadikan sebagai data sampel untuk membentuk model SVM. Sedangkan data *testing* digunakan untuk melihat keakuratan model klasifikasi. Pembagian data dibagi menjadi 80% data *training* dan 20% data *testing*. Berdasarkan keseluruhan 6 data terdapat 5 data *training* dan 1 data *testing*.

	A	B	C
	Tahun,Kuota,Daftar,Lulus		
	2019,140,299,147		
	2020,120,292,121		
	2021,90,150,104		
	2022,120,229,81		
	2023,120,396,88		

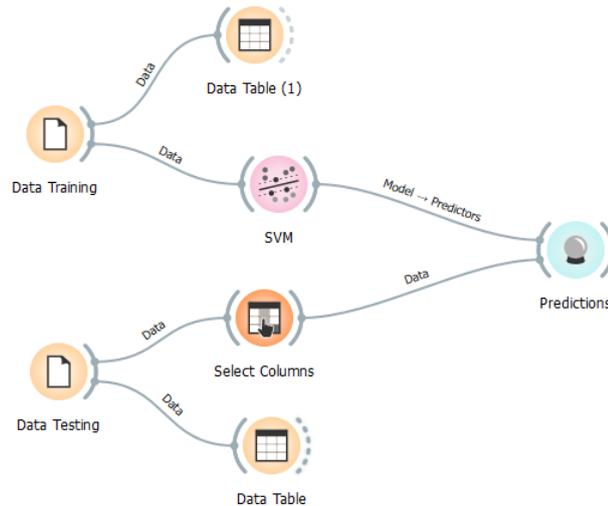
(a)

	A	D
	Tahun,Kuota,Daftar	
	2024,84,194	

(b)

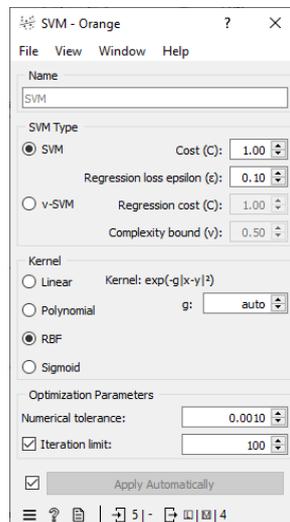
Gambar 2 (a) Data training dengan format csv (b) Data testing dengan format csv

Dari gambar 2 dapat diketahui bahwa variabel-variabel yang digunakan pada data *training* yaitu banyaknya pendaftar dan banyaknya kuota, serta banyaknya yang lulus sebagai target. Sedangkan pada data testing digunakan variabel tahun, kuota dan daftar. Kemudian dilakukan proses model SVM dengan fungsi kernel yang digunakan adalah kernel RBF dan parameter *cost* (C) yang digunakan adalah 1. Proses model SVM dapat dilihat pada gambar 3 dibawah ini.



Gambar 3 Proses model SVM

Dari gambar 3 diatas dapat dijelaskan bahwa proses untuk mencari hasil prediksi dengan algoritama support vector machine dan menggunakan bantuan dari perangkat lunak orange3 adalah pertama mengimport Data training dan data testing dari excel, data tabel digunakan menampilkan data yang telah di import, SVM digunakan sebagai model yang dipakai dalam memprediksi data yang disambungkan dari data training dimana model dapat dilihat pada gambar 4 dibawah ini, select columes digunakan untuk menyambungkan antara data testing dengan prediction dan prediction digunakan untuk menampilkan hasil prediksi lulus pada tahun 2024.



Gambar 4 Model SVM

Setelah model SVM terbentuk dengan menggunakan Orange3, maka diperoleh hasil prediksi sebagai berikut:

SVM	Tahun	Kuota	Daftar
1 104	2024	84	194

Gambar 5 Hasil Prediksi Tahun 2024

Dari gambar diatas diketahui bahwa yang lulus pada tahun 2024 sebanyak 104 calon mahasiswa.

5. Evaluasi

Perhitungan keakuratan model SVM menggunakan kernel *Radial Basic Function* (RBF) dengan parameter $C = 1$ dapat dianalisis dengan mencari RMSE dan MAPE dari data set yang merupakan gabungan data *training* dan data *testing* berikut:

Tabel 3. Hasil *root mean squared error* (RMSE)

Tahun	Nilai Aktual	Nilai Prediksi	Selisih	Selisi Kuadrat
2019	147	110	37	1369
2020	121	109	12	144
2021	104	107	-3	9
2022	81	108	-27	729
2023	88	107	-19	361
2024	112	104	8	64
Rata-Rata				446
RMSE				21.11871208

Berdasarkan tabel 3 diketahui bahwa hasil yang diperoleh dengan menggunakan RSME sebesar 21.11871208. Perhitungan RMSE di atas dilakukan dengan cara menghitung rata-rata dari selisih kuadrat antara nilai aktual dan nilai prediksi, kemudian dilakukan pengakaran kuadrat pada hasil rata-rata.

Tabel 4. Hasil *mean absolute percentage error* (MAPE)

Tahun	Nilai Aktual	Nilai Prediksi	MAPE
2019	147	110	0.25170068
2020	121	109	0.099173554
2021	104	107	0.028846154
2022	81	108	0.333333333
2023	88	107	0.215909091
2024	112	104	0.071428571
Jumlah			1.000391384
MAPE			0.166731897
MAPE (dalam %)			17%

Sedangkan hasil yang diperoleh dengan menggunakan MAPE adalah sebesar 17%. Dari perhitungan MAPE diatas dapat diartikan bahwa hasil pengurangan antara nilai actual dan nilai prediksi yang telah di absolute-kan dibagi dengan nilai actual per periode masing-masing, kemudian dilakukan penjumlahan terhadap hasil-hasil tersebut dan dibagi 5. Nilai MAPE lebih rendah menunjukkan bahwa model peramalan yang digunakan memiliki kemampuan yang lebih baik. Kategori nilai MAPE yang dapat digunakan sebagai ukuran kemampuan model peramalan sebagai berikut (Maricar, 2019):

Tabel 5. Kategori nilai MAPE

Nilai MAPE	Kategori
< 10%	Sangat Baik
10% - 20%	Baik
21% - 50%	Layak
> 50%	Buruk

Berdasarkan tabel 4.5 maka algoritma SVM dengan menggunakan kernel RBF ini dapat dikategorikan ke dalam model peramalan yang baik untuk digunakan dalam memprediksi calon mahasiswa baru pendidikan matematika.

KESIMPULAN

Dari hasil penerapan algoritma Support Vector Machine (SVM) menggunakan aplikasi Orange3 dengan fungsi kernel Radial Basic Function (RBF) dan parameter $C = 1$ dapat disimpulkan bahwa banyaknya calon mahasiswa pada tahun 2024 adalah 104 calon mahasiswa. Berdasarkan hasil evaluasi yang dilakukan dengan menggunakan metode forecasting Errors, Root Mean Squared Error (RSME) yang menghasilkan nilai 21.11871208 serta Mean Absolute Percentage Error (MAPE) dimana nilai yang didapat adalah sebesar 17%. Maka dapat dikatakan bahwa Support Vector Machine (SVM) memiliki kemampuan model peramalan yang baik dalam memprediksi calon mahasiswa baru pendidikan matematika.

DAFTAR PUSTAKA

- Damasela, R., Tomasouw, B. P., & Leleury, Z. A. (2022). Penerapan metode support vector machine (SVM) untuk mendeteksi penyalahgunaan narkoba. *PARAMETER: Jurnal Matematika, Statistika Dan Terapannya*, 1(2), 111–122. <https://doi.org/10.30598/parameter.v1i2pp111-122>
- Harafani, H., & Wahono, R. S. (2015). Optimasi parameter pada support vector machine berbasis algoritma genetika untuk estimasi kebakaran hutan. *Journal of Intelligent Systems*, 1(2), 82–90.
- Hovi, H. S. W., Id Hadiana, A., & Rakhmat Umbara, F. (2022). Prediksi penyakit diabetes menggunakan algoritma Support Vector Machine (SVM). *Informatics and Digital Expert (INDEX)*, 4(1). <https://doi.org/10.36423/index.v4i1.895>
- Kesuma, H. Di, Apriadi, D., Juliansa, H., & Etriyanti, E. (2022). Implementasi data mining prediksi mahasiswa baru menggunakan algoritma regresi linear berganda. *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya Lubuklinggau*, 4(2), 62–66. <https://doi.org/10.52303/jb.v4i2.74>
- Maricar, M. A. (2019). Analisa perbandingan nilai akurasi moving average dan exponential smoothing untuk sistem peramalan pendapatan pada perusahaan XYZ.

Jurnal Sistem Dan Informatika.

- Nofriansyah, D., & Nurcahyo, G. W. (2019). Algoritma data mining dan pengujian. In *Algoritma Data Mining dan Pengujian*. Deepublish.
- Nugroho, A. S., Witaro, A. B., & Handoko, D. (2003). Support vector machine – Teori dan aplikasinya dalam bioinformatika. *Kuliah Umum IlmuKomputer.Com*.
- Nurhayati, S., Luthfi, E. T., & Papua, U. Y. (2015). Prediksi mahasiswa drop out menggunakan metode support vector. *Prediksi Menggunakan SVM*, 3(6), 82–93.
- Saputro, N. D. (2015). Penerapan algoritma support vector machine untuk prediksi harga emas. *Jurnal Informatika Upgris*, 1(1), 10–19.
- Thoyyibah, Irwanto, D., Afqal, R. G., Purwandi, R., Supriatna, D., Suharwati, N., Quddustiani, H., Najwah, M., Suri, A. P., Rohim, D. A., Robbi, M. S., Rahmanto, A., Purnomo, G. E. D., & Sholihah, S. (2014). Algoritma Analisis. In *Science Signaling*.