



Comparison of K-NN, Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction

Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir

**Aulia Putri^{1*}, Cindy Syaficha Hardiana², Elma Novfuja³, Farida Try Puspa Siregar⁴,
Rahmadden⁵, Yulia Fatma⁶, Refni Wahyuni⁷**

^{1,2,3,4,5}Teknik Informatika, STMIK Amik Riau, Indonesia, ⁶Teknik Informatika, Universitas Muhammadiyah Riau, Indonesia, ⁷Teknik Informatika, Universitas Hang Tuah Pekanbaru, Indonesia

E-Mail: ¹2010031802024@sar.ac.id, ²2010031802030@sar.ac.id, ³2010031802038@sar.ac.id,
⁴2010031802042@sar.ac.id, ⁵rahmadden@sar.ac.id, ⁶yuliafatma@umri.ac.id, ⁷refniabid@gmail.com

Received Jan 10th 2003; Revised Feb 19th 2003; Accepted Apr 15th 2003
Corresponding Author: Aulia Putri

Abstract

Final-year student is a student who is struggling to get a bachelor's degree and choose a life goal with a new task such as a job that suits his interests and talents. To get a good and timely graduation rate. Students are very dependent on the influence of factors inside and outside the campus. Selection and determination of data used, taken from public data. With 379 final year students as respondents. This test compares the K-NN, Naive Bayes, and SVM algorithms that better solve problems related to predicting graduate student graduation rates. Based on the comparison of these algorithms with data splitting techniques, it is found that the K-NN (K-Nearest Neighbor) Algorithm has a higher average than (NBC) Naive Bayes Classifier and SVM (Support Vector Machine) for predicting graduation rates of final-year students with 87.8% accuracy, 87.8% precision, and 84% recall.

Keywords: Comparison, Graduation, K-NN, Naive Bayes, Prediction, SVM

Abstrak

Mahasiswa tingkat akhir adalah seorang pelajar yang sedang berjuang demi mendapat gelar sarjana dan memilih tujuan hidup dengan tugas yang baru seperti pekerjaan yang sesuai dengan minat dan bakatnya. Untuk mendapatkan tingkat kelulusan dengan baik dan tepat waktu. Mahasiswa sangat bergantung pada pengaruh dari faktor dalam dan luar kampus. Pemilihan dan penentuan data yang digunakan, diambil dari data publik. Dengan 379 orang mahasiswa tahap akhir sebagai responden. Pengujian ini membandingkan algoritma K-NN, NBC, dan SVM yang lebih baik menyelesaikan masalah terkait prediksi tingkat kelulusan mahasiswa pascasarjana. Berdasarkan perbandingan algoritma tersebut dengan teknik splitting data, didapatkan bahwa Algoritma K-NN (K-Nearest Neighbor) memiliki rata-rata lebih tinggi dibandingkan (NBC) Naive Bayes Classifier dan SVM (Support Vector Machine) untuk prediksi kelulusan mahasiswa tingkat akhir dengan akurasi 87,8%, presisi 87,8%, dan recall 84%.

Kata Kunci: Kelulusan, Komparasi, K-NN, Naive Bayes, Prediksi, SVM

1. PENDAHULUAN

Mahasiswa adalah subjek penting bagi keberhasilan suatu perguruan tinggi baik swasta maupun negeri. Salah satu faktor penentu tingkat mutu suatu perguruan tinggi adalah kemampuan mahasiswa dalam menyelesaikan studinya tepat waktu. Tetapi pada hampir di setiap tahun ajaran yang baru, ketika jumlah mahasiswa bertambah, tidak semua mahasiswa menyelesaikan studinya tepat waktu, sehingga mengakibatkan peningkatan jumlah data mahasiswa [1].

Lama studi dipengaruhi oleh beberapa faktor antara lain motivasi, kemampuan siswa, serta minat pada faktor internal, lalu seperti lingkungan, daerah asal dan asal sekolah pada faktor eksternal. Sikap dan tingkah laku mahasiswa yang beragam juga dapat mempengaruhi masa studi. Inilah yang menarik untuk ditelaah guna meningkatkan mutu lulusan di masa selanjutnya dan mampu menjadi suatu solusi meningkatnya jumlah kegalan studi [2]. Masa lama waktu Pendidikan pada perguruan tinggi dapat diselesaikan dalam waktu

maksimal 4 tahun pada jenjang S1, maksimal 3 tahun pada jenjang D3, setelah itu mahasiswa dilaporkan telah melewati batas waktu [3]. Untuk Mengurangi tingkat kegagalan studi, dibutuhkan suatu sistem yang bisa memakai data akademik mahasiswa dalam memprediksi apakah mahasiswa mempunyai peluang lulus sempurna atau tidak, sehingga dapat mengambil keputusan bagi mahasiswa ketika melakukan prediksi tentang penyelesaian suatu gelar di kemudian hari.

Tingkat pengangguran di kalangan lulusan terus meningkat dari tahun ke tahun. Sebanyak 200.000 lulusan dari universitas dikeluarkan setiap tahun. 1 dari 4 lulusan tetap menganggur enam bulan setelah lulus dan mayoritas adalah pemegang gelar. Padahal banyaknya mahasiswa lulus tepat waktu, dapat meningkatkan kualitas akreditasi perguruan tinggi tersebut. [4]-[5].

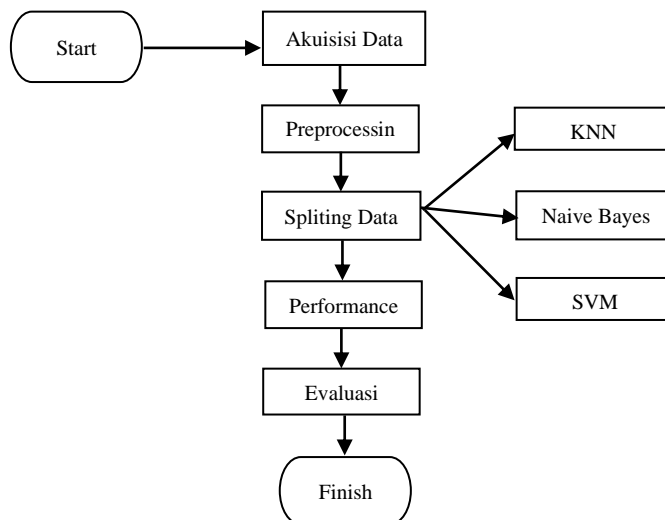
Pesatnya perkembangan data mining selalu berkaitan dengan perkembangan teknologi informasi dalam memproses kumpulan data dalam jumlah besar. Sehingga dengan pentingnya data mining, ada banyak algoritma untuk pemrosesan *big data*, salah satunya menggunakan algoritma klasifikasi [6]. Klasifikasi adalah proses yang dapat digunakan untuk menghubungkan masukan dengan masukan khusus yang disebut kelas dan label. Misalnya, catatan siswa dapat dinilai lulus atau tidak lulus. Beberapa algoritma yang cocok untuk mengklasifikasikan suatu data yaitu *SVM*, *NBC*, *K-NN*, *Discriminant Analysis*, *Decision Tree*, dan *Neural Networks* [7]–[9].

Pada penelitian sebelumnya untuk mendapatkan metode yang sesuai dalam prediksi pola tingkat kelulusan mahasiswa, hanya menggunakan dua model yaitu algoritma (*NBC*) *Naive Bayes Classifier* dan *SVM* (*Support Vector Machine*) [10]. Hasilnya menunjukkan bahwa metode *SVM* lebih baik dari *NBC*, dengan tingkat akurasi sebesar 69,15% untuk data yang digunakan dan secara keseluruhan, pemilihan kedua model tersebut memberikan akurasi yang baik dalam analisis prediksi yang dilakukan oleh peneliti. Selain itu penelitian lainnya memprediksi kinerja siswa dengan hanya menerapkan dua model algoritma yaitu yaitu *K-NN* (*K-Nearest Neighbor*) dan (*NBC*) *Naive Bayes Classifier* dengan hasil percobaan menunjukkan bahwa *NBC* lebih baik dari *K-NN* dengan mendapatkan nilai akurasi tertinggi yaitu 93,6% [11]. Berdasarkan beberapa hasil penelitian diatas didapatkan bahwa dengan membandingkan dua model algoritma klasifikasi masih menunjukkan perbedaan jenis kinerja model klasifikasi yang terbaik berdasarkan tingkat akurasi.

Oleh karena itu penelitian ini melakukan komparasi tiga algoritma sekaligus yaitu *K-Nearest Neighbor*, *Naive Bayes Classifier*, dan *Support Vector Machine* untuk melihat algoritma model mana yang pengklasifikasiannya memiliki nilai akurasi, pressi, dan recall yang lebih baik, juga diharapkan penelitian ini nantinya akan menjadi suatu tolak ukur sejauh mana tingkat kelulusan mahasiswa terutama pada mahasiswa tingkat akhir agar dapat kembali dilakukan evaluasi oleh tiap perguruan tinggi yang ada.

2. METODOLOGI PENELITIAN

Penelitian dilakukan dengan berbagai tahapan-tahapan. Berawal dari pengumpulan data yang bertujuan untuk menerapkan dan menganalisis isi yang dimaksud, data disusun dengan cara membagi data yang telah diolah. Kemudian merancang algoritma yang akan digunakan yaitu *KNN*, *NBC*, dan *SVM* Kemudian eksplorasi model dengan data uji yang akan memprediksi hasil dan membandingkan tingkat kesalahan dalam proses pengujian data. Pengolahan data dibantu dengan Microsoft Excel dan Jupiter notebook. Penelitian dilakukan sebagai pembanding cara kerja metode *K-NN*, *Naive Bayes*, dan *SVM* menggunakan splitting data 90:10, 80:20, 70:30, 60:40, dan 50:50.



Gambar 1. Metodologi Penelitian

Tahapan Metodologi Penelitian:

1. Akuisisi Data

Akuisisi data merupakan tahap mengumpulkan data yang akan diolah. Data dapat diperoleh melalui berbagai metode seperti wawancara langsung, kuisioner, observasi atau pengambilan data dari sumber yang ada. Metode yang digunakan untuk memperoleh data harus konsisten dengan tujuan penelitian dan asumsi yang telah ditentukan sebelumnya. Data juga dapat diperoleh dengan cara mengambil data dari sumber seperti data di internet, data pemerintah, data rumah sakit dan lain-lain [12]. Dataset yang dipakai dalam penelitian ini diperoleh dari *Kaggle* sebanyak 379 data publik. Data-data ini memiliki 6 fitur, dapat dilihat deskripsi fitur data pada tabel 1.

Tabel 1. Deskripsi Fitur Data

No	Fitur	Deskripsi
1.	JENIS_KELAMIN	Jenis Kelamin
2.	STATUS_MAHASISWA	Status Mahasiswa
3.	UMUR	Umur
4.	IPS_(1-8)	Indeks Prestasi Semester 1 hingga 8
5.	IPK	Indeks Prestasi Kumulatif
6.	STATUS_KELULUSAN	Status Kelulusan

2. Preprocessing

Preprocessing merupakan adalah tahap yang digunakan untuk mempersiapkan data sebelum dilakukan analisis. Tahap ini meliputi beberapa tahap seperti pembersihan data, pengubahan data, dan pengintegrasian data. Tahap ini adalah tahap awal sebelum dilakukannya uji model algoritma yang digunakan. *Dataset* yang ada akan diubah menjadi data siap uji, yang bertujuan agar sifat data dapat diuji terhadap model algoritma klasifikasi, yang terdiri dari tahap pembersihan data yaitu terjadi proses penghapusan data kosong dan data yang tidak diperlukan dan transformasi data yaitu terjadi perubahan struktur karakteristik data untuk memenuhi kebutuhan penelitian [12].

3. Splitting data

Splitting data merupakan proses pembagian data yang digunakan dalam penelitian menjadi dua atau lebih bagian, biasanya digunakan untuk menguji model atau algoritma. *Dataset* umumnya dibagi menjadi data (*training*) dan data uji (*testing*). Data latih digunakan melatih algoritma, sementara data uji untuk evaluasi kinerja pada algoritma. [12]. Oleh karena itu pada penelitian ini data dibagi menjadi dua tahap. *Splitting data* pada penelitian ini menggunakan rasio 90:10, 80:20, 70:30, 60:40, dan 50:50.

4. K-Nearest Neighbor (K-NN)

K-NN yaitu teknik pengklasifikasian sekumpulan data dimana sebelumnya telah diklasifikasi. Keakuratan algoritma *K-NN* sangat dipengaruhi oleh berbagai karakteristik, jika nilai fitur ini tidak sesuai dengan perkiraan. Beberapa studi yang menggunakan algoritma *K-NN* menangani hampir secara eksklusif pemilihan fitur dan pembobotan untuk meningkatkan efisiensi algoritma dalam klasifikasi.

Algoritma *K-NN* merupakan *instance-based learning* dimana data training disimpan agar dapat ditemukannya klasifikasi record yang baru, belum terklasifikasi dengan cara melakukan perbandingan pada sebagian besar persamaan atau kesamaan pada data training. *K-NN* berperan untuk mencari jarak terdekat antar data, yang kemudian dilakukan pengevaluasian tetangga terdekat (K) pada data latih. *Data training* ditampilkan dalam ruang multidimensi, masing-masing menggambarkan karakteristik data. Langkah-langkah perhitungan *K-NN* yaitu:

- Menentukan Nilai K yang berupa bilangan bulat positif
- Menghitung jarak antar data baru dan dengan data training

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (1)$$

Diketahui:

- x : data sampel
- y : data pengujian
- D : jangkauan

- Menentukan urutan jarak (jangkauan) paling dekat dengan jarak min pada K
- Label tetangga terdekat digunakan untuk memprediksi data baru. Sedangkan persamaan *K-NN* dalam menghitung nilai prediksi.

$$Y = \frac{1}{K} \sum_{i=1}^k y_i \tag{2}$$

Diketahui:

- Y : dugaan awal
- K : total tetangga terdekat
- y_i : keluaran tetangga terdekat

Algoritma *K-NN* ini memiliki kelebihan berupa keefektifan pada data dengan jumlah banyak dan menghasilkan data lebih akurat. Selain itu *K-NN* ini memiliki kekurangan yaitu kita perlu menentukan nilai k yang optimal dan harus mengeluarkan biaya komputasi yang tinggi untuk menghitung jarak pada query instance.

5. Naïve Bayes Classifier

NBC yaitu algoritma yang memprediksi probabilitas, didasarkan kepada penggunaan teorema bayes dimana antar fitur satu sama lain pada data yang tak terkait satu sama lain. Algoritma ini termasuk satu dari beberapa bentuk sederhana pengklasifikasian suatu persamaan. Metode Naive Bayes ini dapat mengolah data secara cepat walaupun jumlah data sedikit atau tidak banyak. Jika probabilitasnya nol maka prediksi juga akan nol.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{3}$$

Diketahui:

- X : data (label)
- H : hipotesa X adalah label
- P(H) : peluang H
- P(X) : peluang X
- P(X|H) : peluang X pada H
- P(H|X) : peluang H pada X

Terdapat beberapa tahapan pada algoritma Naive Bayes diantaranya yaitu :

- a. Mengitung jumlah label yang ada (P(H))
- b. Menghitung jumlah kasus dari masing-masing label (P(X|H))
- c. Mengalikan semua variabel label (P(H|X) . P(H))

6. Support Vector Machine (SVM)

SVM yaitu teknik pengklasifikasian dimana konsep utama *SVM* adalah mencari *hyperplane* terakurat sebagai pemisah 2 label data. Sederhananya *SVM* berusaha semaksimal mungkin dalam margin (pemisah antar label data). Beberapa tahapan algoritma *SVM* diantaranya :

- a. Memvisualisasikan data
- b. Meminimalisir nilai pada margin
- c. Mencari persamaan hyperplane
- d. Memvisualisasikan hyperplane
- e. Menguji data
- f. Mengklasifikasikan objek

Dari beberapa kasus yang ada dan diteliti, beberapa data yang tidak dapat diklasifikasikan dengan metode *SVM*, kemudian fungsi kernel untuk mengklasifikasikan data dengan bentuk nonlinier dikembangkan. Pelatihan berurutan (*Sequential Training*) digunakan karena mudah dan cepat, yaitu sebagai berikut:

1. Pemberian inisial dan juga parameter lainnya, misal $\lambda=0,7$, $\gamma=0,03$, $c=1$, Iterasi Maksimal=100, dan $\epsilon=0,001$. Selanjutnya perhitungan matrik *Hessia*. Rumus dari persamaan tersebut yaitu.

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \tag{4}$$

untuk data ke-i hingga j, dengan rumus:

- a. $E_i = \sum_{j=1}^n \alpha_j D_{ij}$
- b. $\delta a_i = \min \{ \max[\gamma(1 - E_i), -a_i], c - a_i \}$
- c. $a_i = a_i + \delta a_i$

2. Proses selanjutnya yaitu melakukan secara berkala sampai kondisi pada iterasi tertinggi memenuhi target. Kemudian terbentuk nilai *support vector* (*SV*) yang berdasarkan bekal-kali pengujian, umumnya dengan menggunakan $\text{Threshold} > 0$. Lalu melakukan testing proses agar didapatkan hasil. fungsi hasil keputusan kemudian dihitung menggunakan rumus.

$$f(x) = \sum_{i=1}^m a_i y_i K(x_i, x) + b \quad (5)$$

dengan nilai b:

$$-\frac{1}{2} [\sum_{i=1}^m a_i y_i K(x_i, x^+) + [\sum_{i=1}^m a_i y_i K(x_i, x^-)]]$$

7. Performance

Performance merupakan adalah ukuran atau metrik yang digunakan untuk mengevaluasi kualitas atau keakuratan dari model atau algoritma yang digunakan dalam penelitian [12]. Performance dapat diukur dengan confusion matrix pada berbagai metrik seperti akurasi, presisi, recall, lain-lain sesuai dengan tujuan penelitian yang telah ditentukan dan dikonsepsikan.

Tabel 2. *Confusion Matrix*

Prediksi	Aktual	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Dengan:

TP : True Positive
FP : False Positive
FN : False Negative
TN : True Negative

- a. Akurasi, ini adalah rasio jumlah total prediksi benar dibagi dengan jumlah total prediksi.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (6)$$

- b. Precision, adalah rasio dari jumlah prediksi benar yang positif dibandingkan dengan jumlah total prediksi positif.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

- c. Recall, yaitu rasio dari jumlah prediksi benar yang positif dibandingkan dengan jumlah yang benar positif

$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

8. Evaluasi

Evaluasi adalah tahap yang digunakan untuk mengevaluasi kualitas atau keakuratan model atau algoritma pada penelitian, yang dapat dilakukan melalui berbagai metode seperti uji coba dengan data uji. Metode yang digunakan harus sesuai hipotesis tujuan dan penelitian yang telah ditentukan sebelumnya [12].

3. HASIL DAN ANALISIS

Data diambil atau ditentukan berupa dataset yang bersumber dari data Mahasiswa tahap akhir pada website *Kaggle*. Dengan 379 orang Mahasiswa tahap akhir sebagai responden. Berikut data yang didapatkan dari sumber <https://www.kaggle.com/datasets/permanaraga/graduation-student>. Setelah data didapatkan, langkah berikutnya yaitu melakukan *preprocessing data* untuk mempersiapkan data yang meliputi beberapa tahap seperti pembersihan data, perubahan data, dan pengintegrasian data pada algoritma *machine learning*. Berikut adalah hasil dari *preprocessing data*.

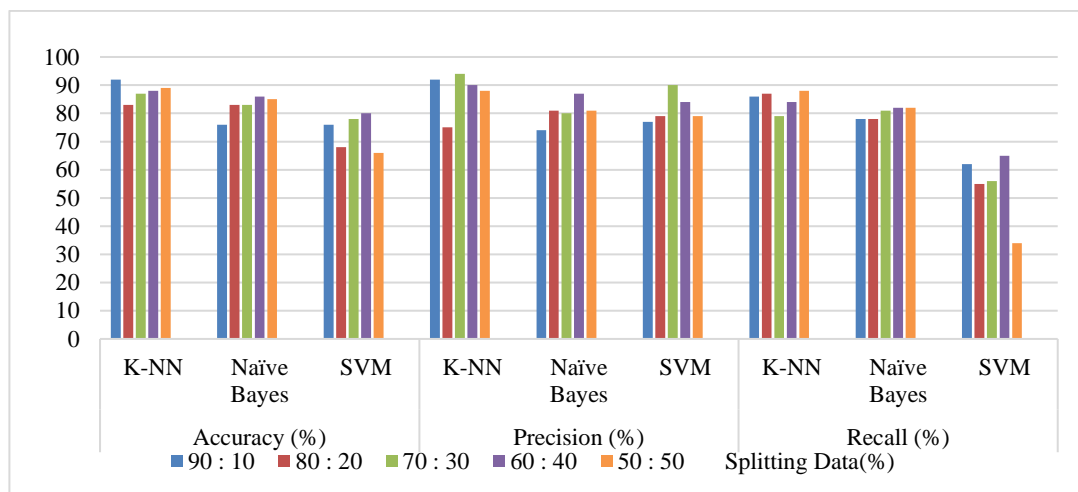
Proses *preprocessing data* yang dilakukan dimulai dengan mengubah jenis kelamin menjadi bernilai 0 adalah laki-laki dan 1 adalah perempuan, Proses kedua yaitu mengubah status mahasiswa menjadi bernilai 0

untuk mahasiswa dan 1 untuk yang sudah bekerja, Lalu proses yang terakhir yaitu mengubah status kelulusan menjadi o untuk yang terlambat dan 1 untuk yang lulus tepat waktu.

Tabel 3. Hasil *Splitting Data*

Data Training (%)	Data Testing (%)	Accuracy (%)			Precision (%)			Recall (%)		
		K-NN	Naïve Bayes	SVM	K-NN	Naïve Bayes	SVM	K-NN	Naïve Bayes	SVM
90	10	92	76	76	92	74	77	86	78	62
80	20	83	83	68	75	81	79	87	78	55
70	30	87	83	78	94	80	90	79	81	56
60	40	88	86	80	90	87	84	84	82	65
50	50	89	85	66	88	81	79	88	82	34
Nilai rata-rata		87,8	82,6	73,6	87,8	80,6	81,8	84,8	80,2	54,4

Penelitian ini dilakukan sebagai pembandingan cara kerja metode *K-NN*, *NBC*, dan *SVM* menggunakan *splitting data* 90:10, 80:20, 70:30, 60:40, dan 50:50. Berdasarkan hasil pengujian masing-masing algoritma yang dilakukan, menghasilkan tingkat akurasi, presisi dan recall dengan nilai rata-rata tiap algoritma yang cukup baik. Pada algoritma *K-NN*, *NBC*, dan *SVM* didapatkan nilai akurasi paling tinggi menggunakan *K-NN* yang mencapai rata-rata 87,8%, untuk nilai presisi tertinggi menggunakan *K-NN* dengan rata-rata 87,8%, dan nilai recall tertinggi menggunakan *K-NN* pula dengan rata-rata 84%.



Gambar 2. Komparasi Algoritma *K-NN*, *Naïve Bayes*, dan *SVM*

Sesuai dengan hasil komparasi diatas menggunakan beberapa *splitting data* bahwa Algoritma *K-NN* lebih baik dari pada *NBC* dan *SVM* dalam memprediksi kelulusan mahasiswa tingkat akhir dengan menggunakan data public sebanyak 379 data yang ada.

4. KESIMPULAN

Berdasarkan hasil pengolahan data, pengujian hingga analisis yang dilakukan untuk melakukan perbandingan algoritma *K-NN*, *NBC*, dan *SVM* dalam prediksi kelulusan mahasiswa tingkat akhir dengan menggunakan data public sebanyak 379 data, diperoleh kesimpulan bahwa algoritma *K-Nearest Neighbor (K-NN)* memiliki rata-rata yang lebih tinggi berdasarkan *splitting data* 90:10, 80:20, 70:30, 60:40, dan 50:50 yaitu dengan tingkat akurasi 87,8%, presisi 87,8%, dan recall 84%. Hasil pengujian menunjukkan bahwa algoritma *K-Nearest Neighbor (K-NN)* dianggap memiliki rata-rata lebih tinggi (baik) dibandingkan *Naïve Bayes Classifier (NBC)* dan *Support Vector Machine (SVM)* terkait prediksi tingkat kelulusan mahasiswa tahap akhir (pascasarjana).

REFERENSI

- [1] Irwan, A. Sauddin, Dan N. Ida, "Penerapan Pohon Keputusan Dalam Memprediksi Masa Studi Mahasiswa Uin Alauddin Makassar," *Jurnal Instek*, Vol. 7, No. 2, Hlm. 201–210, Okt 2022, Doi: <https://doi.org/10.24252/Instek.V7i2.31390>.
- [2] W. Agwil, H. Fransiska, Dan N. Hidayati, "Analisis Ketepatan Waktu Lulus Mahasiswa Dengan Menggunakan Bagging Cart," *Fibonacci: Jurnal Pendidikan Matematika Dan Matematika*, Vol. 6, No. 2, Hlm. 155–166, Des 2020, Doi: 10.24853/Fbc.6.2.155-166.

- [3] F. P. Widhianto, A. A. Supianto, Dan N. Y. Setiawan, "Aplikasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritme C4.5," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, Vol. 3, No. 6, Hlm. 5549–5555, Jun 2019.
- [4] R. S. Effendy, "Pengaruh Upah Minimum Terhadap Pengurangan Tingkat Pengangguran Terbuka Di Indonesia," *Fokus Ekon. J. Ilm. Ekon.*, Vol. 14, No. 1, Pp. 115–124, 2019, Doi: 10.34152/Fe.14.1.115-124.
- [5] V. No, J. S. Informasi, And I. Komputasi, "Optimasi Algoritma Naïve Bayes Parameter Kelulusan Siswa Prediction At Universitas Dirgantara," Vol. 6, No. 1, Pp. 91–106, 2022, Doi: 10.52362/Jisicom.V6i1.785
- [6] I. Parlina et al., "Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012031.
- [7] Y. P. Chiu, "Social Recommendations For Facebook Brand Pages," *Journal Of Theoretical And Applied Electronic Commerce Research*, Vol. 16, No. 1, Hlm. 71–84, Mei 2021, Doi: 10.4067/S0718-18762021000100106.
- [8] T. Hardiani, "Comparison Of Naive Bayes Method, K-Nn (K-Nearest Neighbor) And Decision Tree For Predicting The Graduation Of 'Aisiyah University Students Of Yogyakarta," *International Journal Of Health Science And Technology*, Vol. 2, No. 1, Hlm. 75–85, Jan 2021, Doi: 10.31101/Ijhst.V2i1.1829.
- [9] W. Wiguna Dan D. Riana, "Diagnosis Of Coronavirus Disease 2019 (Covid-19) Surveillance Using C4.5 Algorithm," *Jurnal Pilar Nusa Mandiri*, Vol. 16, No. 1, Hlm. 71, Mar 2020, Doi: 10.33480/Pilar.V16i1.
- [10] Kesumawati, A., & Utari, D. T. (2018). Predicting patterns of student graduation rates using Naïve bayes classifier and support vector machine. *AIP Conference Proceedings*, 2021, 1-1-. <https://doi.org/10.1063/1.5062769>
- [11] I. A. Abu Amra and A. Y. A. Maghari, "Students performance prediction using KNN and Naïve Bayesian," 2017 8th International Conference on Information Technology (ICIT), Amman, Jordan, 2017, pp. 909-913, doi: 10.1109/ICITECH.2017.8079967.
- [12] Akbar, F., Wira Saputra, H., Karel Maulaya, A., Fikri Hidayat, M., & Rahmadden. (2022). Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 61–67.
- [13] L. Setiyani, M. Wahidin, D. Awaludin, Dan S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review," *Faktor Exacta*, Vol. 13, No. 1, Hlm. 35, Jun 2020, Doi: 10.30998/Faktorexacta.V13i1.5548.
- [14] L. Abd, R. Hakim, A. A. Rizal, Dan D. Ratnasari, "Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-Nn)," 2019.
- [15] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, K-Nn Dan Svm," *Jurnal Tekno Insentif*, Vol. 13, No. 1, Hlm. 16–25, Apr 2019, Doi: 10.36787/Jti.V13i1.78.
- [16] A. Pratama, R. Cahya Wihandika, Dan D. E. Ratnawati, "Implementasi Algoritme Support Vector Machine (Svm) Untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa," *Jurnal Pilar Nusa Mandiri*, Vol. 2, No. 4, Hlm. 1704–1708, 2018, [Daring]. Available: [Http://J-Ptiik.Ub.Ac.Id](http://J-Ptiik.Ub.Ac.Id)
- [17] M. E. Lasulika, "Komparasi Naïve Bayes, Support Vector Machine Dan K-Nearest Neighbor Untuk Mengetahui Akurasi Tertinggi Pada Prediksi Kelancaran Pembayaran Tv Kabel," *Ilkom Jurnal Ilmiah*, Vol. 11, No. 1, Hlm. 11–16, Mei 2019, Doi: 10.33096/Ilkom.V11i1.408.11-16.
- [18] R. Puspita, S. Putri, I. Waspada, Dan U. D. Semarang, "Penerapan Algoritma C4.5 Pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," 2018.
- [19] S. A. Naufal, A. Adiwijaya, Dan W. Astuti, "Analisis Perbandingan Klasifikasi Support Vector Machine (Svm) Dan K-Nearest Neighbors (Knn) Untuk Deteksi Kanker Dengan Data Microarray," *Jurikom (Jurnal Riset Komputer)*, Vol. 7, No. 1, Hlm. 162, Feb 2020, Doi: 10.30865/Jurikom.V7i1.2014.
- [20] J. Zeniarja, A. Salam, Dan F. A. Ma'ruf, "Seleksi Fitur Dan Perbandingan Algoritma Klasifikasi Untuk Prediksi Kelulusan Mahasiswa," *Jurnal Rekayasa Elekrika*, Vol. 18, No. 2, Jul 2022, Doi: 10.17529/Jre.V18i2.24047.
- [21] D. Prajarini, S. Tinggi, S. Rupa, D. Desain, Dan V. Indonesia, "Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Kulit," 2016.
- [22] N. A. Sinaga, B. H. Hayadi, Dan Z. Situmorang, "Perbandingan Akurasi Algoritma Naïve Bayes, K-Nn Dan Svm Dalam Memprediksi Penerimaan Pegawai," *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, Vol. 5, No. 1, Hlm. 27, Jul 2022, Doi: 10.37600/Tekinkom.V5i1.446.