

Graphical abstract



DATA MINING KLASIFIKASI KELULUSAN MAHASISWA MENGGUNAKAN METODE NAÏVE BAYES

¹Andik, ¹Syarli, ²Cipta Riang Sari

¹Universitas Al Asyariah Mandar.

**Corresponding author*

Andhyka035@gmail.com

Abstract

A student's ability to complete a study according to a designated time is an important factor in assessing university accreditation. A good accreditation shows the image of a university. The problem that arises is that many students miss the completion of their studies, which hinders the certification of their learning programs. The purpose of this study is to design an application program that can support student graduation-related decisions. This study applies the Nave Bayesian method, which can predict future opportunities based on past experience at the University of Al-Ashalia Mander, School of Computer Science, and Information Systems Research Program. This study is a data mining application program for student graduation classification using the Nave Bayes method, which is expected to help educators complete their studies on time and develop their ability to increase accreditation in their research programs. The design was successful.

Keywords: *Data Mining, Classification, Nave Bayes.*

Abstrak

Kemampuan seorang mahasiswa untuk menyelesaikan studi sesuai dengan waktu yang ditentukan merupakan faktor penting dalam menilai akreditasi universitas. Akreditasi yang baik menunjukkan citra sebuah universitas. Permasalahan yang muncul adalah banyaknya mahasiswa yang ketinggalan dalam menyelesaikan studinya sehingga menghambat sertifikasi program pembelajarannya. Tujuan dari penelitian ini adalah untuk merancang program aplikasi yang dapat mendukung keputusan terkait kelulusan mahasiswa. Penelitian ini menerapkan metode Nave Bayesian, yang dapat memprediksi peluang masa depan berdasarkan pengalaman masa lalu di Universitas Al-Ashalia Mander, Fakultas Ilmu Komputer, dan Program Penelitian Sistem Informasi. Penelitian ini merupakan program aplikasi data mining klasifikasi kelulusan mahasiswa dengan menggunakan metode Nave Bayes, yang diharapkan dapat membantu para pendidik menyelesaikan studinya tepat waktu dan mengembangkan kemampuannya untuk meningkatkan akreditasi pada program penelitiannya.

Kata kunci: *Data Mining, Klasifikasi, Naive Bayes*

Article history

DOI: <http://dx.doi.org/10.35329/jp.v4i1.2358>

Received : 25 Agust 2021 | Received in revised form : 27 Mei 2022 | Accepted : 31 Mei 2022

1. PENDAHULUAN

Dalam sistem pendidikan tinggi, mahasiswa merupakan aset penting lembaga pendidikan. Naik turunnya kemampuan seorang mahasiswa untuk menyelesaikan studi tepat waktu menjadi salah satu faktor penilaian akreditasi sebuah universitas. (Sutoyo & Almaarif, 2020). Semakin tinggi nilai akreditasi maka semakin baik pula pandangan pihak luar terhadap kualitas program pembelajaran dan perguruan tinggi, sehingga perlu mempertimbangkan tingkat kelulusan mahasiswa dan berjalan tepat waktu.

Setiap universitas memiliki database yang menyimpan berbagai macam data, seperti biodata mahasiswa dan data akademik. Dengan memanfaatkan melimpahnya data yang dimiliki oleh universitas, Anda dapat menemukan informasi tersembunyi dengan mengolah data mahasiswa. Basis data dapat digunakan untuk mempelajari pola dan perilaku siswa untuk meminimalkan keterlambatan kelulusan. (Setiyani et al., 2020).

Salah satu metode yang dapat digunakan untuk menganalisis suatu database adalah teknologi data mining. Teknologi data mining atau penambangan data adalah proses memperoleh pengetahuan atau pola dari kumpulan data. (Sutoyo & Almaarif, 2020). Data mining memecahkan masalah dengan menganalisis data yang sudah ada di database. Ada banyak cara untuk mengolah data dalam data mining. Salah satunya adalah dengan menerapkan metode *Naive Bayes* untuk mengklasifikasikan data.

(Syarli & Muin, 2016) Kami melakukan studi tentang metode *Naive Bayes* untuk prediksi kelulusan (studi kasus: data mahasiswa baru). Menurut mereka, *Naive Bayes* dapat melakukan klasifikasi menggunakan metode statistik dan probabilitas. Dengan kata lain, Anda dapat memprediksi peluang masa depan berdasarkan pengalaman Anda sebelumnya. Nilai persentase akurasi menunjukkan efektifitas dataset penerimaan mahasiswa baru yang berlaku pada klasifikasi *Naive Bayes*. (Nikmatun & Waspada, 2019). Kami melakukan survey yang berjudul "Implementasi Data Mining Untuk Klasifikasi Periode Belajar Siswa Menggunakan Algoritma *K-Nearest Neighbor*". Kesimpulan dari penelitian ini adalah nilai akurasi terbaik didapatkan pada skenario yang menggunakan atribut elektif. Ini adalah 75,95%. Berdasarkan nilai akurasi tertinggi yang digunakan, semua mata kuliah pilihan dapat disimpulkan bahwa mata kuliah pilihan memiliki pengaruh yang signifikan terhadap masa belajar siswa. (Widaningsih, 2019) Kami melakukan survei dengan judul "Perbandingan Akurasi *Naive Bayes* dan *K-Nearest Neighbor* Pada Klasifikasi untuk Memprediksi Status Ketenagakerjaan Alumni ITBSTIKOM Bali". Dibandingkan dengan metode tetangga terdekat dengan akurasi 82,34% dan MAPE 17,66% dan nilai K 9, akurasi dan MAPE dari hasil yang diperoleh dengan *Naive Bayes* masing-masing adalah 83,83% dan 16,17%. (Hasan et al., 2018). Mereka melakukan penelitian yang berjudul "Perbandingan

Algoritma C4.5, KNN, dan *Naive Bayes* untuk Menentukan Model Klasifikasi Direksi *Entrepreneur Center BSI*." Pada penelitian ini melihat metode mana yang lebih tepat dan melihat nilai akurasi yang lebih tinggi, nilai akurasi metode C4.5 sebesar 73,33%, nilai akurasi metode KNN sebesar 70%, dan nilai akurasi *Naive Bayes* adalah 80%. Artinya, dalam tiga cara

Oleh karena itu, algoritma yang paling tepat digunakan untuk mengklasifikasikan sebagai Ketua Pusat Kewirausahaan BSI di setiap kampus Universitas Bina Sarana Informatika menggunakan metode *Naive Bayes* dengan menghasilkan nilai akurasi tertinggi. (Sabilla & Putri, 2017) Dalam penelitiannya yang berjudul "Memprediksi Ketepatan Waktu Kelulusan Mahasiswa Menggunakan Metode *K-Nearest Neighbor* dan *Naive Bayes Classifier*". Mereka menggunakan beberapa atribut untuk memprediksi ketepatan waktu lulusan. Yakni, nomor induk mahasiswa, nama mahasiswa, IPS semester satu sampai empat, jenis kelamin, dan kampung halaman mahasiswa. Kampung halaman siswa dapat dibagi menjadi dua kategori: kota dan pinggiran kota. Data tersebut kemudian dibagi menjadi dua kelas: kelulusan terlambat dan kelulusan tepat waktu.

Berdasarkan penjelasan yang diberikan, peneliti memimpin dalam membangun sistem yang dapat mengklasifikasikan kelulusan bagi mahasiswa yang lulus semester lima menggunakan teknik data mining menggunakan algoritma *Naive Bayes*. Atribut yang digunakan dalam data pelatihan adalah jenis kelamin, IPS 1, IPS 2, IPS 3, IPS 4, IPS 5, IPK semester 5, jumlah SKS 1-5, alamat (luar polman atau dalam polman) dan kelulusan. .. Pernyataan untuk atribut data pengujian: NIM, nama, jenis kelamin, IPS 1, IPS 2, IPS 3, IPS 4, IPS 5, IPK, jumlah SKS, dan alamat (luar polman atau dalam polman).

2. METODE PENELITIAN

Metode penelitian untuk penelitian ini adalah dengan mengklasifikasikan data siswa menggunakan metode *Naive Bayes*, yaitu menggunakan teknik data mining.

2.1 Data mining

Data mining adalah proses memberikan metode pembelajaran satu atau lebih sistem komputer tugas untuk menganalisis dan mengambil pengetahuan secara langsung. Ini juga mencakup serangkaian proses untuk mengeksplorasi nilai tambah secara manual dari sekumpulan data yang tidak diketahui. (Sinaga & Handoko, 2021).

Data mining memiliki serangkaian proses yang perlu dilakukan sebelum mendapatkan informasi baru. Yaitu, pembersihan data, pembersihan proses penghilangan noise dan data yang tidak konsisten, proses integrasi data yang menggabungkan data dari sumber data yang berbeda, pemilihan data, dan proses seleksi. Data yang telah diambil untuk digunakan dalam proses data mining, meninggalkan data yang tidak digunakan. Konversi data adalah proses mengubah data ke dalam format yang dapat digunakan dalam perhitungan algoritmik dan evaluasi pola. merupakan

Proses analisis hasil proses penambangan menggunakan satuan pengukuran, dan penyajian pengetahuan, adalah proses menampilkan hasil proses penambangan. (Saputro & Sari, 2020).

2.2 Klasifikasi

Klasifikasi adalah proses menemukan model atau fungsi yang menggambarkan dan membedakan kelas data, dan dimaksudkan untuk memungkinkan Anda menggunakan model untuk memprediksi kelas objek dari label kelas yang tidak diketahui. Klasifikasi yang dihasilkan harus membedakan setiap kelas dari yang lain dan menyajikan diagram terorganisir dari dataset. Tujuan dari klasifikasi adalah untuk memudahkan dalam menemukan model, huruf, dan konsep dari dataset untuk setiap kelas untuk membantu Anda membuat keputusan. Studi klasifikasi dalam berbagai disiplin ilmu telah dilakukan dengan cara yang berbeda. Metode Bayesian telah dipelajari di berbagai bidang untuk klasifikasi, klasifikasi pola adalah salah satu bidang kecerdasan buatan yang paling penting, dan dalam beberapa tahun terakhir jaringan saraf tiruan telah digunakan untuk klasifikasi pola karena kemampuannya yang luar biasa. pemodelan sistem, belajar mandiri, penyesuaian diri, dll. (Syarli, 2017).

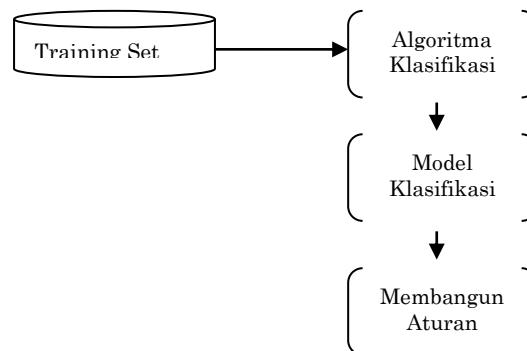
Algoritma data mining dapat dibagi menjadi tiga paradigma: terawasi, tidak terawasi, dan semi terawasi. (Widaningsih, 2019)

- a. Pembelajaran terawasi adalah jenis pembelajaran mesin di mana model ini menyediakan data pelatihan berlabel. Dalam bahasa Indonesia, pembelajaran terawasi berarti pembelajaran terawasi karena memiliki "label" yang menunjukkan bagian mana yang merupakan "hasil".
- b. Pembelajaran tanpa pengawasan bahasa Indonesia adalah "pembelajaran tanpa pengawasan". Unsupervised bertujuan untuk mengidentifikasi pola yang bermakna dalam data. Jika pembelajaran yang diawasi belajar dari data berlabel, mesin yang tidak diawasi harus belajar dari kumpulan data yang tidak berlabel.
- c. Algoritma kereta semi-diawasi menggunakan data berlabel dan tidak berlabel. Biasanya, sejumlah kecil data berlabel digunakan dan sejumlah besar data tidak berlabel digunakan. Pembelajaran semi terawasi tidak jauh berbeda dengan pembelajaran terawasi.

Klasifikasi termasuk dalam pembelajaran terawasi. Dalam pembelajaran terawasi, algoritma bekerja untuk sekumpulan data yang diberi label atau kelasnya diketahui. Proses klasifikasi dibagi menjadi dua tahap (Annasaheb & Verma, 2016) yaitu :

a. Tahap pembuatan model

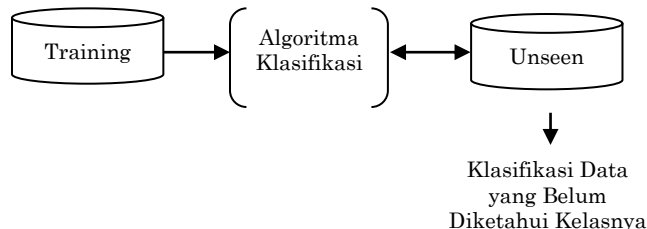
Data sampel yang digunakan disebut data uji atau data latih (training set). Pada tahap ini, model klasifikasi dibangun berdasarkan data yang ditentukan oleh kelas. Proses ini disebut proses induksi yang ditunjukkan pada Gambar 1.



Gambar 1 Tahapan Membangun Model

b. Tahapan menggunakan model klasifikasi

Pada tahap ini, model diterapkan pada data kelas yang tidak diketahui. Proses penerapan model klasifikasi dan penggunaan data uji (test set) untuk memprediksi label kelas dari data dalam himpunan disebut estimasi. Gambar 2.



Gambar 2 Tahapan Menggunakan Model

2.3 Algoritma Naive Bayes

Naive Bayes adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk pembelajaran mesin dan penambangan data. Performa naive bayes yang bersaing dalam proses klasifikasi (tidak ada hubungan antar atribut), bahkan saat menggunakan asumsi independensi atribut. Asumsi independensi atribut dalam data ini jarang terjadi dalam praktik, tetapi meskipun melanggar asumsi independensi atribut, pengklasifikasi naive Bayes berkinerja sangat baik, sebagaimana dibuktikan oleh berbagai studi empiris. (Syarli & Muin, 2016).

Keuntungan menggunakan Nave Bays adalah metode ini hanya membutuhkan sedikit data latih (training data) untuk menentukan estimasi parameter yang dibutuhkan oleh proses klasifikasi. (Riang Sari, 2015).

Naive Bayes adalah metode yang bagus karena mudah dibuat dan tidak memerlukan skema estimasi parameter atau iterasi yang rumit. Artinya, ini dapat diterapkan pada kumpulan data yang besar. Naive Bayes adalah klasifikasi statistik dan probabilistik yang diusulkan oleh ilmuwan Inggris Thomas Bayes, yang dikenal sebagai teorema Bayes karena memprediksi peluang masa depan berdasarkan pengalaman masa lalu. (Syarli & Muin, 2016). Nave Bayes mudah diimplementasikan, efisien dan memiliki kinerja yang sangat baik. Algoritma Nave Bayes memproses data

menggunakan probabilitas yang dapat terjadi antara atribut data pelatihan. (Dwi et al., 2019).

Aturan berikut berlaku untuk algoritma naive Bayes (Dwi et al., 2019):

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

di mana :

X: Data kelas yang tidak diketahui

H: Data hipotesis adalah kelas tertentu

P (H | X): Probabilitas hipotesis H berdasarkan kondisi X (probabilitas posterior)

P (H): Probabilitas hipotesis H (probabilitas sebelumnya)

P(X | H): Probabilitas X berdasarkan kondisi hipotesis H

P (X): Probabilitas X

Naive Bayes adalah penyederhanaan dari metode Bayesian. Teorema Bayes disederhanakan sebagai berikut.

$$P (H | X) = P (H) P (X)$$

Gunakan teorema Bayes untuk menghitung probabilitas dan posterior dari data sebelumnya. Hasil akhirnya adalah mendapatkan informasi post-mortem dan periferal untuk menghasilkan probabilitas menggunakan Nave Bayes.

3. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan hasil berupa program aplikasi data mining klasifikasi kelulusan mahasiswa menggunakan metode Nave Bayes. Hal ini memungkinkan siswa untuk mengkategorikan peluang kelulusan mereka, apakah mereka tepat waktu atau terlambat dalam menyelesaikan studi mereka. Dengan menggunakan data mahasiswa yang lulus tepat waktu atau terlambat di Arsip Program Penelitian Sistem Informasi Fakultas Ilmu Komputer Universitas Al Ashalia Mander sebagai data pelatihan. Semakin tinggi tingkat akurasi perhitungan, semakin tergantung pada jumlah data latih yang digunakan.

Beberapa fitur dari aplikasi yang berhasil dibuat adalah::

a. Form Login

Bagian ini mengharuskan pengguna untuk memasukkan nama pengguna dan kata sandi sebelum menggunakan program. Seperti yang Anda lihat di gambar 3.

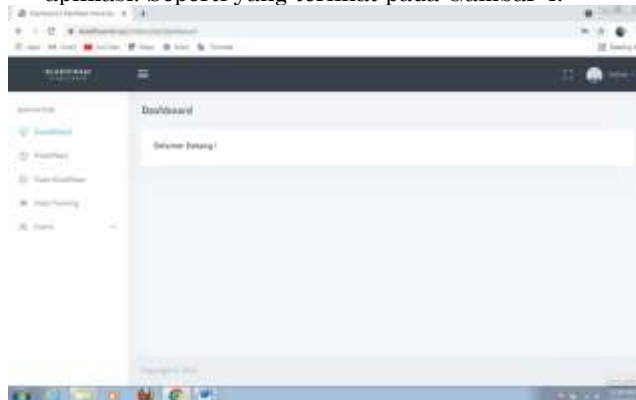


Gambar 3 Form Login

Jika nama pengguna dan kata sandi yang dimasukkan benar, pengguna akan dibawa ke halaman menu utama aplikasi, tetapi dengan nama pengguna dan atau *password* salah maka akan muncul peringatan mengenai hal tersebut.

b. Form Menu Utama

Halaman selanjutnya adalah halaman pertama yang dilihat pengguna ketika berhasil login ke aplikasi. Seperti yang terlihat pada Gambar 4.

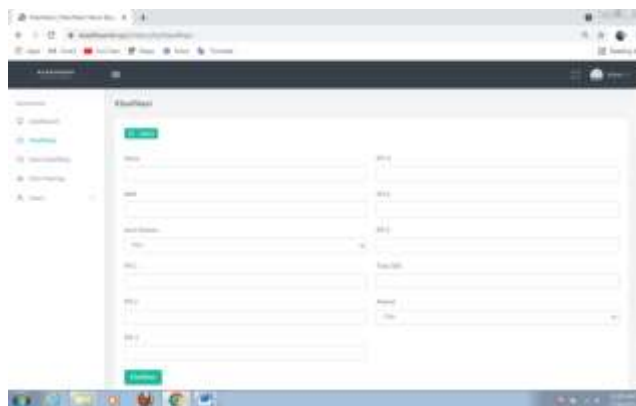


Gambar 4 Form Menu Utama

Aplikasi ini memiliki lima menu utama: Dashboard, Classification, Classification Data, Training Data, dan User.

c. Form Klasifikasi

Pada bagian ini, pengguna dapat mengkategorikan kelulusan siswa seperti yang ditunjukkan pada Gambar 5.

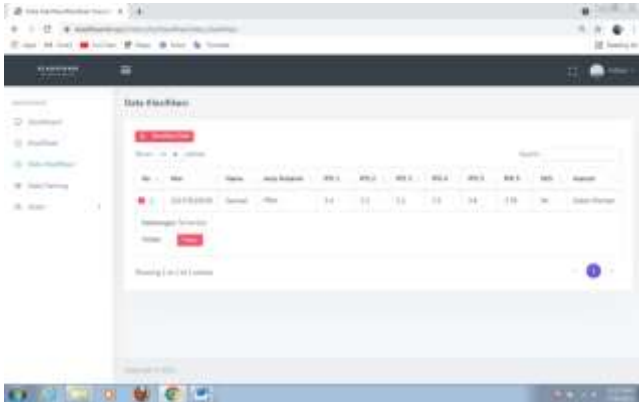


Gambar 5 Form Klasifikasi

Ada dua cara untuk mengklasifikasikan. Artinya, pengguna dapat memasukkan data siswa secara manual atau mengimpor data.

d. Form Data Klasifikasi

Disini kita dapat melihat hasil klasifikasi yang pernah kita lakukan sebelumnya. Seperti yang terlihat pada gambar 6.

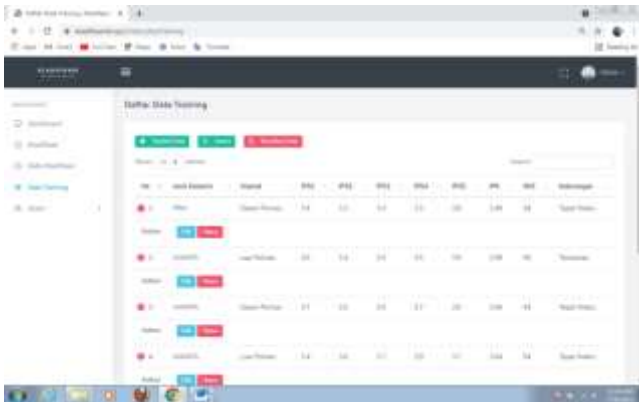


Gambar 6 Form Data Klasifikasi

Ada fungsi untuk membersihkan data klasifikasi dan fungsi pencarian untuk mempersempit data yang ada.

e. *Form Data Training*

Pada bagian ini, pengguna diharuskan memasukkan data pelatihan sebelum melakukan klasifikasi. Seperti yang ditunjukkan pada Gambar 7.

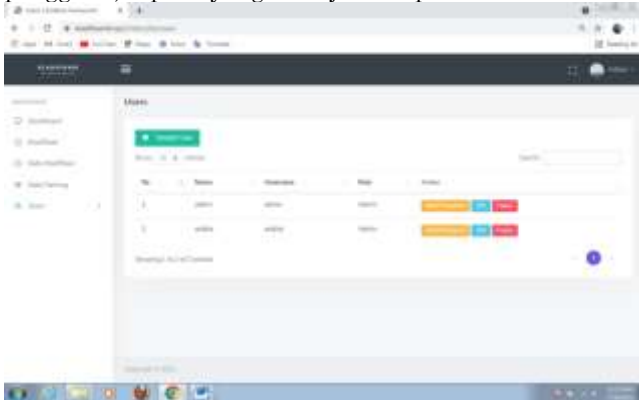


Gambar 7 Form Data Training

Ada dua cara untuk menambahkan data pelatihan ke aplikasi Anda, baik secara manual atau dengan menggunakan fitur impor data..

f. *Form Users*

Bagian ini memiliki form untuk menambahkan pengguna, seperti yang ditunjukkan pada Gambar 8.



Gambar 8 Form Users

Akun pengguna yang dibuat dapat sepenuhnya menggunakan aplikasi dan tidak ada yang dapat menggunakannya, Perbedaan antara satu pengguna

dengan pengguna lainnya. Akun pengguna yang dibuat akan dihapus atau diedit.

4. SIMPULAN

4.1 Kesimpulan

Data mining menggunakan metode Naive Bayes Berdasarkan hasil survey yang dilakukan dari awal hingga proses pengujian mengenai penerapan klasifikasi kelulusan siswa, semakin banyak data training atau data training yang digunakan maka semakin akurat prediksinya.

4.2 Saran

Berdasarkan uraian sistem yang dibuat, saran untuk pengembangan sistem ini dapat diberikan sebagai berikut:

Aplikasi ini tidak dapat menganalisis mengapa siswa terlambat dalam menyelesaikan pembelajarannya dan hanya menggunakan satu metode yaitu teorema Bayes, sehingga tidak dapat dibandingkan dengan metode data mining lainnya. Untuk pengembangannya diharapkan memberikan kemampuan kepada mahasiswa untuk menganalisis penyebab keterlambatan penyelesaian studinya, dan disarankan untuk menggunakan beberapa metode data mining.

Karena saran-saran dari penulis, kami berharap saran-saran tersebut dapat dijadikan bahan masukan khususnya bagi penulis, dan umumnya bagi mahasiswa yang sedang mengedit.

DAFTAR PUSTAKA

(Annasaheb & Verma, 2016). (n.d.).

- Dwi, R., Pambudi, Afif, A., Supianto, & Setiawan, N. Y. (2019). Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 2196, 3(3), 2194–2200. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/4655/2154>
- Hasan, F. N., Hikmah, N., & Utami, D. Y. (2018). Perbandingan Algoritma C4.5, KNN, dan Naive Bayes untuk Penentuan Model Klasifikasi Penanggung jawab BSI Entrepreneur Center. *Jurnal Pilar Nusa Mandiri*, 14(2), 169. <https://doi.org/10.33480/pilar.v14i2.908>
- Nikmatun, I. A., & Waspada, I. (2019). Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432.
- Riang Sari, C. (2015). *Teknik Data Mining Menggunakan Classification Dalam Sistem Penunjang Keputusan Peminatan SMA*.
- Sabilla, W. I., & Putri, T. E. (2017). Prediksi Ketepatan Waktu Lulus Mahasiswa dengan k- Nearest Neighbor dan Naive Bayes Classifier (Studi Kasus Prodi D3 Sistem Informasi Universitas Airlangga). *Jurnal Komputer Terapan*, 3(2), 233–240. <https://jurnal.pcr.ac.id/index.php/jkt/article/download/1544/699>
- Saputro, I. W., & Sari, B. W. (2020). Uji Performa

- Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, 6(1), 1. <https://doi.org/10.24076/citec.2019v6i1.178>
- Setiyani, L., Wahidin, M., Awaludin, D., & Purwani, S. (2020). Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review. *Faktor Exacta*, 13(1), 38–47. <https://doi.org/10.30998/faktorexacta.v13i1.5548>
- Sinaga, K., & Handoko, K. (2021). Implementasi Data Mining Untuk Memprediksi Kelulusan Siswa Dengan Metode Naïve Bayes. *Antimicrobial Agents and Chemotherapy*, 04. <https://doi.org/10.1128/AAC.03728-14>
- Sutoyo, E., & Almaarif, A. (2020). Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(1), 95–101. <https://doi.org/10.29207/RESTI.V4I1.1502>
- Syarli, S., & Muin, A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmiah Ilmu Komputer*, 2(1), 22–26. <https://fikom-unasman.ac.id/ejournal/index.php/jikom/article/download/76/52>
- Syarli, S., & Qashlim, A. (2017). Pemetaan Pemerataan Pendidikan Menggunakan Self Organizing Maps (SOM) Terintegrasi Sistem Informasi Geografis. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, 3(2), 27-34.
- Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, 13(1), 16–25. <https://doi.org/10.36787/jti.v13i1.78>