

Sistem Informasi Prediksi Kelulusan Menggunakan Metode *Naive Bayes Classifier* **(Studi Kasus: Prodi Pendidikan Teknik Informatika)**

Kemal Refta Diska¹, Khairi Budayawan²

¹Program Studi Pendidikan Teknik Informatika, Universitas Negeri Padang

²Departement Teknik Elektronika, Universitas Negeri Padang

e-mail: kemalrd33@gmail.com

Abstrak

Prediksi (prediction) adalah kegiatan memprediksi apa yang akan terjadi di masa yang akan datang. Karena masalah pengambilan keputusan adalah masalah yang harus dihadapi, demikian juga peramalan. Karena prediksi sangat erat kaitannya dengan pengambilan keputusan. Prediksi kelulusan mahasiswa merupakan upaya untuk membantu mahasiswa mengetahui status kelulusannya. Metode yang digunakan yaitu metode prototyping merupakan salah satu metode pengembangan perangkat lunak yang menggunakan pendekatan untuk membuat rancangan dengan cepat dan bertahap sehingga dapat segera dievaluasi oleh calon pengguna/klien. Aplikasi prediksi mahasiswa ini diuji menggunakan data mahasiswa PTI angkatan 2014 dan 2015 sebagai data training berjumlah 94 data, dan data mahasiswa angkatan 2016 sebagai data testing berjumlah 46 data. Pada aplikasi ini *klasifikasi naive bayes classifier* digunakan untuk melakukan klasifikasi data berdasarkan hasil studi 6 semester dan total sks. Hasil klasifikasi dari aplikasi ini menyatakan bahwa dari 46 data testing memperoleh hasil accuracy 82,61%, precision 91,66%, recall 61,11%. Aplikasi ini dapat membantu memprediksi hasil kelulusan mahasiswa tepat waktu atau tidak tepat waktu berdasarkan hasil studi.

Kata kunci: Prediksi, *Naive bayes classifier*, Metode *Prototype*.

Abstract

Prediction is the activity of predicting what will happen in the future. Because decision-making problems are problems that must be faced, so are forecasting. Because prediction is very closely related to decision making. Prediction of student graduation is an effort to help students find out their graduation status. The method used is the prototyping method which is a software development method that uses an approach to make designs quickly and gradually so that it can be immediately evaluated by potential users/clients. This student prediction application was tested using PTI student data batch 2014 and 2015 as training data. 94 data, and 2016 class student data as testing data totaling 46 data. In this application the naïve Bayes classifier classification is used to classify data based on the results of 6 semester studies and total credits. The results of the classification of this application stated that of the 46 testing data obtained results of accuracy 82.61%, precision 91.66%, recall 61.11%. This application can help predict student graduation results on time or not on time based on study results.

Keywords : *Prediction, Naïve Bayes Classifier, Method Prototype*

PENDAHULUAN

Di zaman modern ini, perkembangan teknologi informasi dan komunikasi telah membuka peluang besar bagi masyarakat untuk mendapatkan informasi. Masyarakat dapat mengakses informasi apapun dimanapun. Kebutuhan akan informasi yang akurat sangat

penting agar informasi menjadi bagian penting dari pembangunan sosial. Kebutuhan informasi yang tinggi terkadang tidak terpuaskan dengan menyajikan informasi yang tepat, seringkali informasi tersebut digali dari informasi yang jumlahnya sangat banyak. Banyak teknologi informasi telah menghasilkan data dari misalnya bisnis, ekonomi, pendidikan dan banyak bidang lainnya (Suryadi, 2015).

Prediksi (forecasting) adalah suatu kegiatan yang memperkirakan apa yang terjadi pada masa yang akan datang. Masalah pengambilan keputusan merupakan masalah yang dihadapi maka peramalan juga merupakan masalah yang harus dihadapi, karena peramalan berkaitan erat dengan pengambilan suatu keputusan (Ayuningtias et al., 2017).

Berdasarkan data akademik Program Studi Pendidikan Teknik Informatika UNP, jumlah lulusan tepat waktu memperoleh rata-rata yang rendah. Rata-rata mahasiswa angkatan 2013 sampai 2015 yang lulus tepat waktu hanya sebesar 19,26%. Berdasarkan hasil wawancara terhadap mahasiswa prodi Pendidikan Teknik Informatika, dapat diketahui bahwa alasan belum selesainya skripsi dikarenakan ada kendala pada mahasiswa yang masih mengulang mata kuliah, kurang termotivasi untuk menyelesaikan skripsi, rasa malas, lingkungan yang tidak mendorong untuk membuat skripsi dengan semangat dan tekun, sulit menemui dosen pembimbing, faktor administrasi, kesibukan mahasiswa seperti kerja dan organisasi.

Prediksi kelulusan mahasiswa merupakan upaya untuk membantu mahasiswa mengetahui status kelulusannya. Sebelumnya belum ada yang membuat aplikasi untuk melakukan prediksi kelulusan mahasiswa di prodi Pendidikan Teknik informatika UNP. Dalam upaya prediksi ini dapat dilakukan dengan data mining. Metode data mining diantaranya yaitu klasifikasi, prediksi, klustering, estimasi, asosiasi. Algoritma data mining yang digunakan dalam penelitian ini yaitu Algoritma Naïve Bayes.

Klasifikasi adalah proses menemukan sekumpulan pola dan karakteristik yang menjelaskan dan memisahkan data ke dalam kategori tertentu. Tujuannya adalah menggunakan model untuk menentukan kelas dari suatu objek yang kelasnya tidak diketahui. Ada 2 proses dalam klasifikasi, yaitu proses pembelajaran/pelatihan. Bangun model menggunakan data pelatihan. Proses Pengujian Melakukan pengujian data dengan menggunakan model yang diperoleh dari proses pelatihan (Roring et al., 2022)

Berdasarkan penelitian yang dilakukan oleh (Sardi & Budayawan, 2020) yaitu tentang Klasifikasi Tingkat Kelulusan Mahasiswa Elektronika Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus : Pendidikan Teknik Informatika FT-UNP). Klasifikasi ini menggunakan 20 atribut sebagai atribut kontrol yang merupakan transkrip nilai semester 1 sampai dengan semester 3. Berdasarkan pengujian yang dilakukan terhadap mahasiswa tahun masuk 2014 diperoleh akurasi sebesar 79,07% dengan 43 sampel data dan pengujian yang dilakukan terhadap mahasiswa tahun masuk 2015 akurasi yang dihasilkan sebesar 68% dengan 50 sampel data. Tujuan pada penelitian ini adalah untuk mengetahui tingkat kelulusan mahasiswa prodi Pendidikan Teknik informatika UNP.

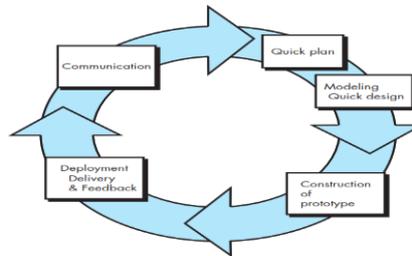
Algoritma Naive Bayes merupakan salah satu algoritma yang digunakan untuk Teknik pengklasifikasian data (Apridiansyah et al., 2021). Naive Bayes adalah classifier yang menggunakan metode probabilistik dan statistik yang diperkenalkan oleh ilmuwan Inggris Thomas Bayes yang memprediksi kemungkinan masa depan berdasarkan pengalaman masa lalu dengan cara yang dikenal sebagai teorema Bayes. Klausula tersebut diikuti oleh "naïve", yang mengasumsikan bahwa kondisi antar atribut tidak bergantung satu sama lain (Syarli & Muin, 2016).

Algoritma Naïve Bayes merupakan algoritma data mining yang dapat digunakan untuk melakukan klasifikasi, dan merupakan algoritma yang akan memberikan hasil kemungkinan yang akan terjadi di masa akan datang dari hasil klasifikasi (Gunawan et al., 2021). *Naive bayes classifier* merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan bayes dengan asumsi ketidak tergantungan yang kuat. Selain itu *Naive bayes classifier* juga dapat menganalisa variabel-variabel yang paling mempengaruhinya dalam bentuk peluang.

Berdasarkan uraian sebelumnya peneliti tertarik untuk membuat sebuah sistem informasi prediksi kelulusan mahasiswa menggunakan metode naïve bayes classifier.

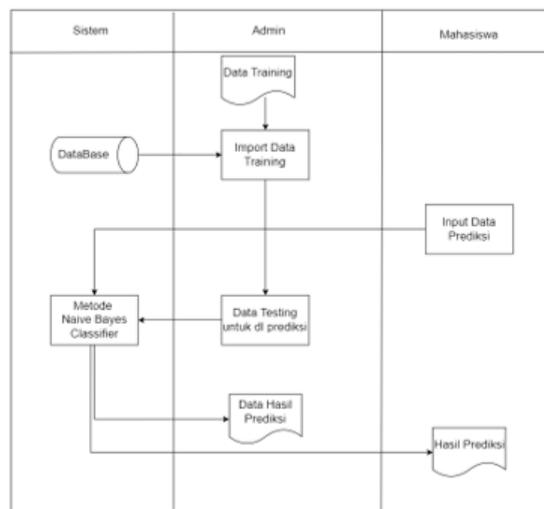
METODE

Penelitian sistem informasi prediksi kelulusan dengan menggunakan metode klasifikasi Naive Bayesian. Prototipe adalah teknik untuk mengumpulkan informasi spesifik dengan cepat tentang kebutuhan informasi pengguna. Fokus pada penyajian aspek perangkat lunak yang terlihat oleh pelanggan atau pengguna. Prototype tersebut akan dievaluasi oleh pelanggan/pemakai dan dipakai untuk menyaring kebutuhan perancangan perangkat lunak (Susanto & Andriana, n.d.).



Gambar 1. Metode *Prototype*

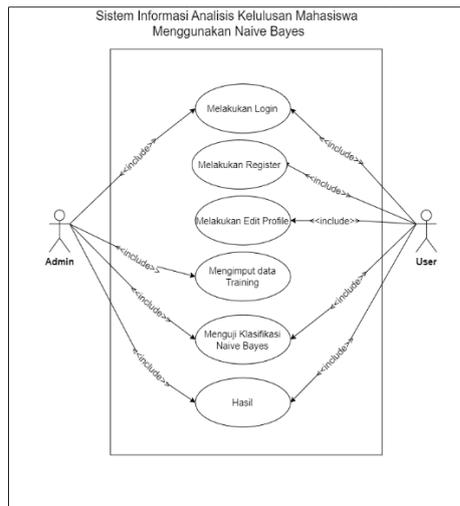
Analisis Sistem yang diusulkan



Gambar 2. Flowmap Analisis yang diusulkan

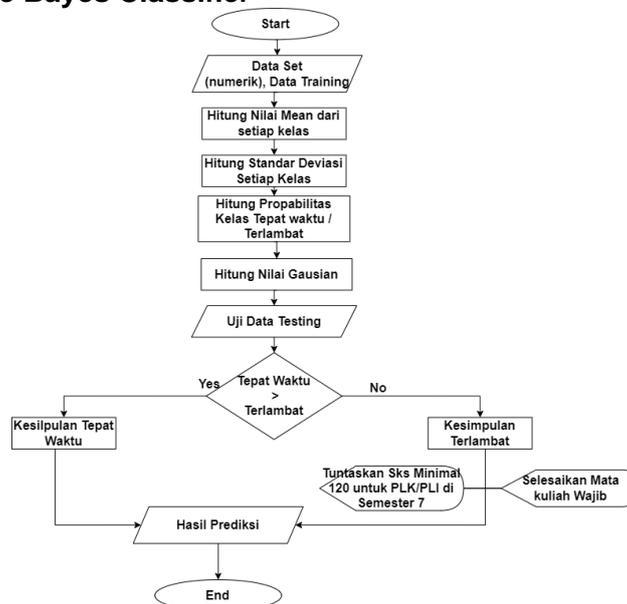
Use Case Diagram

Use case adalah deskripsi tentang bagaimana sistem bekerja dari sudut pandang atau sudut pandang pengguna sistem. Use case mendefinisikan apa yang ditangani oleh sistem dan komponennya. Use case bekerja dengan skenario, yang merupakan deskripsi urutan atau langkah-langkah yang menjelaskan apa yang dilakukan pengguna dengan sistem dan sebaliknya. Kasus penggunaan mengidentifikasi fungsionalitas sistem, interaksi pengguna dengan sistem, dan hubungan antara pengguna dan fungsionalitas sistem (Setiyani, 2021).



Gambar 3. Use Case Diagram Sistem Prediksi Kelulusan Mahasiswa

Metode Usulan Naïve Bayes Classifier

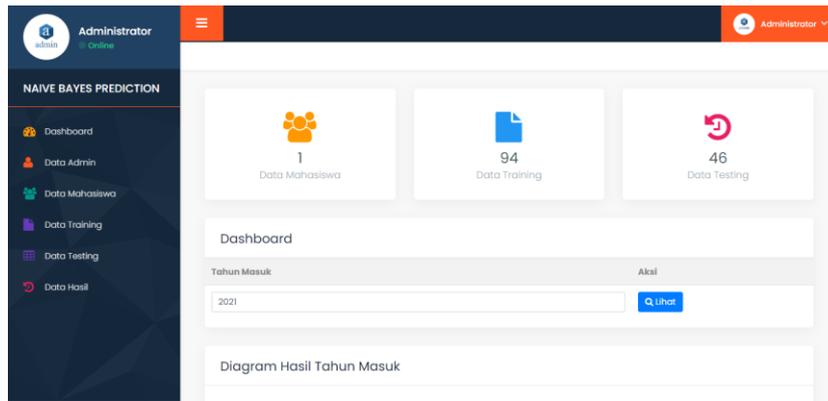


Gambar 4. Flowchart Perhitungan Naive Bayes

HASIL DAN PEMBAHASAN

Dalam penelitian ini, terdapat dua user yang dapat mengakses aplikasi yaitu admin dan mahasiswa. Pada user admin, nantinya dapat mengelola data mahasiswa, data training, data testing, dan data hasil berdasarkan proses data mining dengan metode *naive bayes classifier*.

Pada gambar 2, ditampilkan halaman dashboard admin, dimana pada dashboard terdapat informasi jumlah data yang sudah terinput ke sistem.



Gambar 5. Halaman Dashboard Admin

Dalam penelitian ini menggunakan metode naïve bayes dalam proses perhitungannya, rumus yang digunakan dalam proses naïve bayes ini sebagai berikut:

Rumus Perhitungan Nilai Mean untuk data:

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \tag{1}$$

Rumus Perhitungan Nilai Standart Deviasi untuk data:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \tag{2}$$

Rumus Perhitungan Gaussian Classifier:

$$F(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x-\mu}{2\sigma^2}} \tag{3}$$

Data Training

Data training adalah data yang digunakan sebagai acuan untuk membangun model klasifikasi (Darwis et al., 2021). Data training digunakan untuk tabel propabilitas.

Tabel 1. Tabel Data Training

No	C1	IP 1	IP 2	IP 3	IP 4	IP 5	IP 6	C8	Hasil
1	2	3.23	3.45	3.57	3.54	3.28	3.61	135	Tepat Waktu
2	1	2.65	2.8	3.5	3.45	3.16	3.31	128	Tepat Waktu
3	1	2.99	3.56	3.43	3.63	3.17	3.31	133	Tepat Waktu
4	1	2.58	3.08	3.18	3.41	2.73	3.08	128	Terlambat

90	2	3.37	3.35	3.47	3.36	3.78	3.55	135	Tepat Waktu
91	2	3.2	3.24	3.38	3.41	3.53	3.02	138	Tepat Waktu
92	2	1.05	1.97	1.7	2.25	2.54	1.89	97	Terlambat
93	1	3.02	3.18	3.38	3.25	3.24	3.39	135	Tepat Waktu
94	2	3.25	3.2	3.29	2.97	3.31	2.89	138	Tepat Waktu

Tabel 1 merupakan table data training untuk dijadikan sampel pengolahan yang nantinya akan dijadikan acuan untuk melakukan testing menggunakan data testing.

Data Testing

Data testing adalah data yang digunakan untuk menguji performa dari model klasifikasi tersebut (Darwis et al., 2021). Data testing digunakan untuk menguji tabel probabilitas yang telah terbentuk.

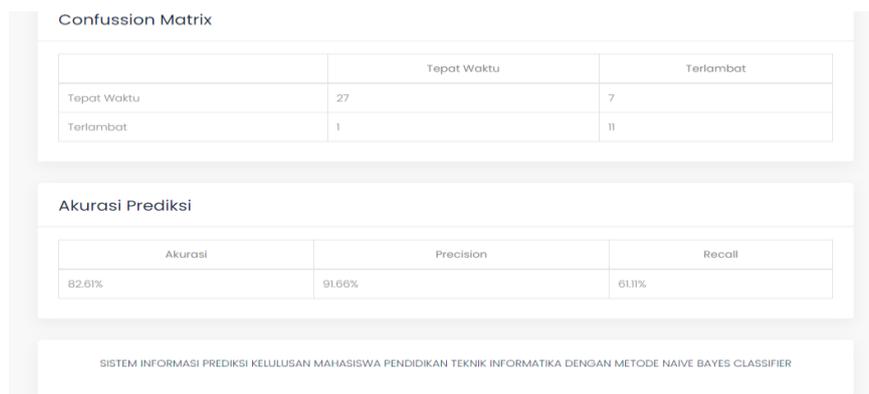
Tabel 2. Data Testing

No	C1	IP 1	IP 2	IP 3	IP 4	IP 5	IP 6	C8	Aktual	Prediksi
1	1	2.97	2.82	2.71	3.15	3.33	3.53	127	Terlambat	Tepat Waktu
2	2	2.96	3.18	2.36	3.38	3.35	3.2	125	Terlambat	Terlambat
3	1	3.56	3.76	3.51	3.7	3.8	3.72	135	Tepat Waktu	Tepat Waktu
4	2	3.26	3.35	2.94	3.46	3.39	3.43	133	Terlambat	Tepat Waktu
5	1	2.92	2.71	2.98	2.7	2.73	2.57	120	Terlambat	Terlambat
6	1	2.78	3	2.98	3.37	3.39	2.78	125	Terlambat	Tepat Waktu

43	2	3.1	3.45	3.2	3.21	3.37	2.97	135	Tepat Waktu	Tepat Waktu
44	2	3.23	2.96	2.7	3.14	2.6	3	131	Tepat Waktu	Terlambat
45	1	2.64	2.49	3	2.69	3.07	2.25	120	Terlambat	Terlambat
46	2	3.13	3.19	3.03	3.13	3.57	3.27	138	Tepat Waktu	Tepat Waktu

Tabel 2 merupakan data testing yang sudah didapat hasil prediksi sesuai kolom prediksi.

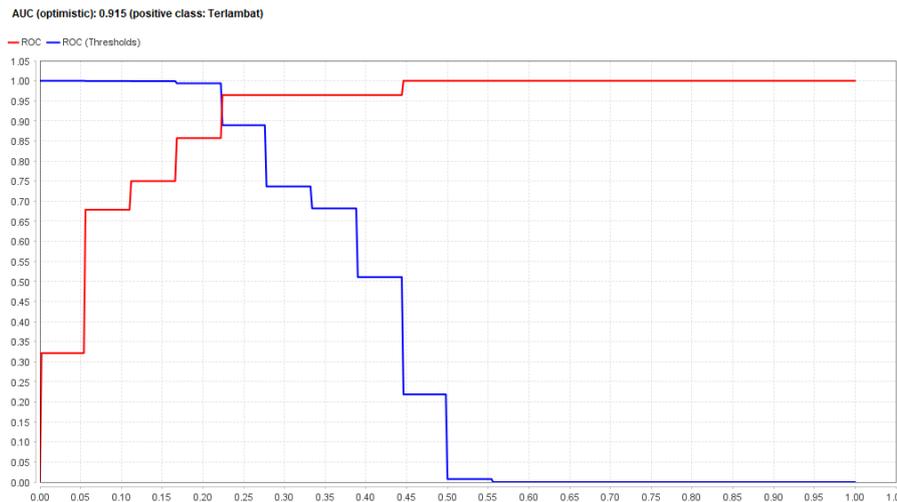
Berdasarkan rumus dan data training yang sudah diolah, lalu mendapatkan hasil berdasarkan data testing, maka pada gambar 3 dijelaskan bahwa:



Gambar 6. Confussion Matrix dan Performance

Dari gambar 3 di dapat hasil confusion matrix. Confusion Matrix adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah (Normawati & Prayogi, 2021). Jumlah data hasil prediksi yang di lakukan dengan sampel data testing sebanyak 46 memperoleh data prediksi tepat waktu dan actual tepat waktu sebanyak 27 mahasiswa, prediksi tepat waktu actual terlambat sebanyak 7 mahasiswa, prediksi terlambat actual tepat waktu sebanyak 1 mahasiswa, dan prediksi terlambat actual terlambat sebanyak 11 mahasiswa.

Dari gambar 3 juga di simpulkan hasil accuracy sebesar 82,6%, precision 91,66%, recall 61,11%.



Gambar 7. Gambar ROC/AUC

Berdasarkan gambar 4. Diperoleh hasil ROC/AUC. Kurva ROC merupakan kurva yang biasa digunakan oleh para peneliti untuk mengevaluasi hasil prediksi. Kurva ROC menggambarkan kinerja klasifikasi terlepas dari distribusi atau kesalahan kelas, di mana sumbu vertikal mewakili nilai positif (TP) dan sumbu horizontal mewakili nilai negatif. Nilai (FP) (Hadianto et al., 2019). berdasarkan kelas False Positive sebesar 0,915. ROC memiliki tingkat diaknosa yaitu: Classification akurasi bernilai 0,900 – 1,000 = Excellent, 0,800 – 0,900 = good, 0,700 – 0,800 = fair, 0,600 – 0,700 = poor, 0,500 – 0,600 = failure.

Hasil dari Classification akurasi pada data training dan data testing yang dilakukan pada penelitian kali ini masuk kedalam kategori Excellent (0,915) untuk melakukan classification Prediksi kelulusan mahasiswa.

Berdasarkan hasil dari perhitungan, aplikasi yang dirancang layak untuk melakukan proses prediksi mahasiswa dikarenakan memperoleh hasil akurasi 82,6% dan ROC sebesar 0,915 yang mumpuni untuk dilakukan proses prediksi dengan menggunakan metode *naïve bayes classifier*.

SIMPULAN

Berdasarkan hasil penelitian, prediksi kelulusan mahasiswa menggunakan metode *naive bayes classifier* dapat di simpulkan bahwa, aplikasi yang dirancang mampu memproses prediksi dengan baik dan cukup akurat, penerapan model *naive bayes classifier* dengan menggunakan data training sejumlah 94 data mahasiswa dan data testing sejumlah 46 data mahasiswa memperoleh hasil accuracy sebesar 82,6%, precision sebesar 91,66%, dan recall sebesar 61,11%. Hasil dari pengujian ROC memperoleh hasil 0,915 dimana hasil itu masuk kedalam kategori excellent untuk melakukan classification prediksi kelulusan mahasiswa. Dari hasil ROC 0,915 disimpulkan penelitian ini dikategorikan excellent / sangat layak

Untuk penelitian selanjutnya, diharapkan mampu menggunakan metode lain selain *naive bayes classifier*, data yang digunakan lebih banyak lagi sehingga memperoleh hasil yang lebih akurat.

DAFTAR PUSTAKA

- Apridiansyah, Y., David, N., Veronika, M., Putra, E. D., Muhammadiyah, U., & Id, Y. A. (2021). Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes. *JSAI: Journal Scientific and Applied Informatics*, 4(2). <https://doi.org/10.36085>
- Ayuningtias, L. P., Irfan, M., & Jumadi, J. (2017). Analisa Perbandingan Logic Fuzzy Metode Tsukamoto, Sugeno, Dan Mamdani (Studi Kasus : Prediksi Jumlah Pendaftar Mahasiswa Baru Fakultas Sains Dan Teknologi Universitas Islam Negeri Sunan Gunung Djati

- Bandung). *Jurnal Teknik Informatika*, 10(1). <https://doi.org/10.15408/jti.v10i1.5610>
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/jtk.v15i1.744>
- Gunawan, M., Zarlis, M., & Roslina, R. (2021). Analisis Komparasi Algoritma Naive Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Media Informatika Budidarma*, 5(2), 513. <https://doi.org/10.30865/mib.v5i2.2925>
- Hadianto, N., Novitasari, H. B., & Rahmawati, A. (2019). *View of KLASIFIKASI PEMINJAMAN NASABAH BANK MENGGUNAKAN METODE NEURAL NETWORK*. <http://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/658/546>
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naive Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(2), 697–711.
- Roring, C. B., Mulyana, D. I., Lubis, Y. T., & Zamzami, A. R. (2022). *Klasifikasi Tingkat Kematangan Buah Jambu Bol Berdasarkan Warna Kulit Menggunakan Metode Naive Bayes*. <https://jptam.org/index.php/jptam/article/view/3335/2822>
- Sardi, H. Y., & Budayawan, K. (2020). Klasifikasi Tingkat Kelulusan Mahasiswa Elektronika Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus: Pendidikan Teknik Informatika FT-UNP). *Voteteknika (Vocational Teknik Elektronika Dan Informatika)*, 8(4), 147. <https://doi.org/10.24036/voteteknika.v8i4.110394>
- Setiyani, L. (2021). *Desain Sistem: Use Case Diagram*. <https://ejournal.rosma.ac.id/index.php/inotek/article/view/183/142>
- Suryadi, S. (2015). PERANAN PERKEMBANGAN TEKNOLOGI INFORMASI DAN KOMUNIKASI DALAM KEGIATAN PEMBELAJARAN DAN PERKEMBANGAN DUNIA PENDIDIKAN. *INFORMATIKA*, 3(3), 133–143. <https://doi.org/10.36987/INFORMATIKA.V3I3.219>
- Susanto, R., & Andriana, A. D. (n.d.). PERBANDINGAN MODEL WATERFALL DAN PROTOTYPING UNTUK PENGEMBANGAN SISTEM INFORMASI. *Majalah Ilmiah UNIKOM*, 14(1).
- Syarli, & Muin, A. A. (2016). *Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)*. <https://fikom-unasman.ac.id/ejournal/index.php/jikom/article/view/76/52>