

**KLASIFIKASI MAHASISWA *DROP OUT* MENGGUNAKAN METODE  
*NAIVE BAYES* DENGAN SELEKSI FITUR *INFORMATION GAIN***

**SKRIPSI**



**UIN SUNAN AMPEL  
S U R A B A Y A**

Disusun Oleh

**ROUDLOTUN NIKMAH**

**H72219033**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL  
SURABAYA**

**2024**

## PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : ROUDLOTUN NIKMAH

NIM : H72219033

Program Studi : Matematika

Angkatan : 2019

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "KLASIFIKASI MAHASISWA *DROP OUT* MENGGUNAKAN METODE *NAIVE BAYES* DENGAN SELEKSI FITUR *INFORMATION GAIN*".

Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 04 Januari 2024

Yang menyatakan,



**ROUDLOTUN NIKMAH**

**NIM. H72219033**

## LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : ROUDLOTUN NIKMAH  
NIM : H72219033  
Judul skripsi : *KLASIFIKASI MAHASISWA DROP OUT  
MENGUNAKAN METODE NAIVE BAYES DENGAN  
SELEKSI FITUR INFORMATION GAIN*

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



Nurissaidah Wlinuha, M.Kom  
NIP. 199011022014032004

Pembimbing II



Dr. Lutfi Hakim, M.Ag  
NIP. 197312252006041001

Mengetahui,  
Ketua Program Studi Matematika  
UIN Sunan Ampel Surabaya



Yuniar Farida, M.T  
NIP. 197905272014032002

## PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama : ROUDLOTUN NIKMAH  
NIM : H72219033  
Judul Skripsi : KLASIFIKASI MAHASISWA *DROP OUT*  
MENGUNAKAN METODE *NAIVE BAYES* DENGAN  
SELEKSI FITUR *INFORMATION GAIN*

Telah dipertahankan di depan Tim Penguji  
pada tanggal 04/01/2024

Mengesahkan,  
Tim Penguji

Penguji I



Aris Farani, M.Kom  
NIP. 198701272014031002

Penguji II



Ahmad Hanif Asyhar, M.Si  
NIP. 198601232014031001

Penguji III



Nurissadiah Izzinnuha, M.Kom  
NIP. 199011022014032004

Penguji IV



Dr. Lutfi Hakim, M.Ag  
NIP. 197312252006041001

Mengetahui,

Dekan Fakultas Sains dan Teknologi  
UIN Sunan Ampel Surabaya



Hamdani, M.Pd  
NIP. 196507312000031002



**KEMENTERIAN AGAMA**  
**UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA**  
**PERPUSTAKAAN**

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300  
E-Mail: perpus@uinsby.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI**  
**KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Roudlotun Nikmah  
NIM : H72219033  
Fakultas/Jurusan : Sains dan Teknologi / Matematika  
E-mail address : roudlotunnikmah05@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi     Tesis     Desertasi     Lain-lain (.....)

yang berjudul :

KLASIFIKASI MAHASISWA DROP OUT MENGGUNAKAN METODE NAIVE BAYES  
DENGAN SELEKSI FITUR INFORMATION GAIN

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara **fulltext** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 16 Januari 2024

Penulis

( **Roudlotun Nikmah** )



## ABSTRAK

### KLASIFIKASI MAHASISWA *DROP OUT* MENGGUNAKAN METODE *NAIVE BAYES* DENGAN SELEKSI FITUR *INFORMATION GAIN*

Salah satu indikator penting yang perlu diperhatikan dalam upaya meningkatkan keberlangsungan perguruan tinggi adalah mahasiswa *drop out*. Banyak faktor yang menyebabkan terjadinya *drop out* pada mahasiswa. Tingginya angka mahasiswa yang mengalami *drop out* dapat mempengaruhi kualitas pembelajaran dan akreditasi perguruan tinggi. Akreditasi sangat penting untuk keberlangsungan suatu institusi. Penelitian ini bertujuan untuk mengklasifikasikan mahasiswa *drop out* dan mengidentifikasi variabel yang paling berpengaruh pada klasifikasi mahasiswa *drop out* dengan menerapkan algoritma *Naive Bayes* dan algoritma *feature selection Information Gain*. Data yang digunakan terdiri dari data mahasiswa aktif dan *drop out* tahun 2015-2021 di UIN Sunan Ampel Surabaya. Hasil klasifikasi *Naive Bayes* menunjukkan akurasi 42.67%, presisi 65.14% dan recall 61.83%. Seleksi fitur *Information Gain* menghasilkan 3 fitur yang paling berpengaruh yaitu SKS, IPS dan Nominal UKT. Dengan menggunakan fitur yang terpilih tersebut hasil klasifikasi *Naive Bayes* dengan seleksi fitur *Information Gain* menunjukkan akurasi 89.78%, presisi 88.61% dan recall 83.04%. Hal ini menunjukkan bahwa seleksi fitur dengan *Information Gain* dapat mengidentifikasi fitur-fitur penting yang berkontribusi terhadap kualitas model klasifikasi dan terjadi peningkatan akurasi yang awalnya 42.67% menjadi 89.78%.

**Kata kunci:** *Drop Out, Information Gain, Naive Bayes, Klasifikasi*

## ABSTRACT

### CLASSIFICATION OF DROP OUT STUDENTS USING NAIVE BAYES METHOD WITH INFORMATION GAIN FEATURE SELECTION

One important indicator that needs to be considered to improve the sustainability of higher education is student drop out. Many factors cause students to drop out. The high number of students who drop out can affect the quality of learning and university accreditation. Accreditation is very important for the sustainability of an institution. This study aims to classify drop out students and identify the most influential variables in the classification of drop out students by applying the Naive Bayes algorithm and algorithm feature selection Information Gain. The data used consists of data of active and drop out students in 2015-2021 at UIN Sunan Ampel Surabaya. Naive Bayes classification results show 42.67% accuracy, 65.14% precision and 61.83% recall. Information Gain feature selection produces 3 features, namely SKS, IPS and Nominal UKT. By using the selected features, the Naive Bayes classification results with the Information Gain feature selection show an accuracy of 89.67%, precision of 88.61% and recall of 83.04%. This shows that feature selection with Information Gain can identify important features that contribute to the quality of the classification model and there is an increase in accuracy from 42.67% to 89.78%.

**Keywords:** *Classification, Drop Out, Information Gain, Naive Bayes*

UIN SUNAN AMPEL  
SURABAYA

## DAFTAR ISI

<b>HALAMAN JUDUL</b>	i
<b>LEMBAR PERSETUJUAN PEMBIMBING</b>	ii
<b>PENGESAHAN TIM PENGUJI SKRIPSI</b>	iii
<b>HALAMAN PERNYATAAN KEASLIAN</b>	iv
<b>MOTTO</b>	v
<b>HALAMAN PERSEMBAHAN</b>	vi
<b>KATA PENGANTAR</b>	vii
<b>DAFTAR ISI</b>	ix
<b>DAFTAR TABEL</b>	xii
<b>DAFTAR GAMBAR</b>	xiv
<b>ABSTRAK</b>	xv
<b>ABSTRACT</b>	xvi
<b>I PENDAHULUAN</b>	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	12
1.3. Tujuan Penelitian	13
1.4. Manfaat Penelitian	13
1.5. Batasan Masalah	14
1.6. Sistematika Penulisan	14
<b>II TINJAUAN PUSTAKA</b>	16
2.1. <i>Data Mining</i>	16
2.2. Klasifikasi	17
2.3. <i>Drop Out</i>	18
2.4. Semester	19
2.5. Satuan Kredit Semester (SKS)	19
2.6. Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK)	20
2.7. Jalur Masuk	20



2.8. Uang Kuliah Tunggal (UKT)	21
2.9. Program Studi	22
2.10. Status Nikah	23
2.11. Penghasilan Ortu	24
2.12. Pendidikan Asal	25
2.13. Normalisasi Data	26
2.14. <i>Feature Selection</i> (Seleksi Fitur)	27
2.14.1. <i>Information Gain</i>	27
2.15. <i>Naive Bayes</i>	30
2.16. <i>Confusion Matrix</i>	36
2.17. Integrasi Keilmuan	38
<b>III METODE PENELITIAN</b>	<b>46</b>
3.1. Jenis Penelitian	46
3.2. Sumber Data	46
3.3. Variabel Penelitian	47
3.4. Tahapan Metode Penelitian	48
<b>IV HASIL DAN PEMBAHASAN</b>	<b>49</b>
4.1. Deskripsi Data	49
4.2. <i>Pre-processing Data</i>	55
4.3. Seleksi Fitur dengan <i>Information Gain</i>	58
4.3.1. Pemisahan Atribut	58
4.3.2. Perhitungan <i>Entropy</i>	60
4.3.3. Perhitungan <i>Information Gain</i>	67
4.3.4. Pemilihan Atribut	73
4.4. Perhitungan Algoritma <i>Naive Bayes</i>	73
4.4.1. Menghitung Nilai Mean	74
4.4.2. Menghitung Nilai Standar Deviasi	77
4.4.3. Menghitung Nilai Probabilitas	81
4.4.4. Menghitung Nilai <i>Gaussian</i>	82
4.5. Algoritma <i>Naive Bayes</i> dan <i>Information Gain</i>	93

4.6. Perbandingan Hasil	98
4.7. Integrasi Keilmuan	99
<b>V PENUTUP</b>	<b>102</b>
5.1. Kesimpulan	102
5.2. Saran	103
<b>DAFTAR PUSTAKA</b>	<b>103</b>



UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR TABEL

2.1 <i>Confusion Matrix</i> . . . . .	37
3.1 Data Penelitian . . . . .	47
4.1 Data Numerik . . . . .	50
4.2 Data Kategorik . . . . .	50
4.3 Sampel Data Sebelum di Konversi . . . . .	51
4.4 Sampel Data Sesudah di Konversi . . . . .	51
4.5 Distribusi Data . . . . .	54
4.6 Normalisasi Data Penelitian . . . . .	57
4.7 Jumlah Data Untuk Setiap Kelas Pada Program Studi . . . . .	58
4.8 Jumlah Data Untuk Setiap Kelas Pada Status Nikah . . . . .	58
4.9 Jumlah Data Untuk Setiap Kelas Pada Pendidikan Asal . . . . .	58
4.10 Jumlah Data Untuk Setiap Kelas Pada Semester . . . . .	58
4.11 Jumlah Data Untuk Setiap Kelas Pada SKS . . . . .	59
4.12 Jumlah Data Untuk Setiap Kelas Pada IPK . . . . .	59
4.13 Jumlah Data Untuk Setiap Kelas Pada IPS . . . . .	59
4.14 Jumlah Data Untuk Setiap Kelas Pada Jalur Masuk . . . . .	59
4.15 Jumlah Data Untuk Setiap Kelas Pada Nominal UKT . . . . .	60
4.16 Jumlah Data Untuk Setiap Kelas Pada Penghasilan Orang Tua . . . . .	60
4.17 Nilai <i>Entropy</i> Program Studi . . . . .	61
4.18 Nilai <i>Entropy</i> Status Nikah . . . . .	62
4.19 Nilai <i>Entropy</i> Pendidikan Asal . . . . .	62
4.20 Nilai <i>Entropy</i> Semester . . . . .	63
4.21 Nilai <i>Entropy</i> SKS . . . . .	64
4.22 Nilai <i>Entropy</i> IPK . . . . .	64
4.23 Nilai <i>Entropy</i> IPS . . . . .	65
4.24 Nilai <i>Entropy</i> Jalur Masuk . . . . .	66

4.25 Nilai <i>Entropy</i> Nominal UKT	66
4.26 Nilai <i>Entropy</i> Penghasilan Ortu	67
4.27 Nilai <i>Information Gain</i> Program Studi	68
4.28 Nilai <i>Information Gain</i> Status Nikah	68
4.29 Nilai <i>Information Gain</i> Pendidikan Asal	69
4.30 Nilai <i>Information Gain</i> Semester	69
4.31 Nilai <i>Information Gain</i> SKS	70
4.32 Nilai <i>Information Gain</i> IPK	70
4.33 Nilai <i>Information Gain</i> IPS	71
4.34 Nilai <i>Information Gain</i> Jalur Masuk	71
4.35 Nilai <i>Information Gain</i> Nominal UKT	72
4.36 Nilai <i>Information Gain</i> Penghasilan Ortu	72
4.37 Perhitungan Nilai <i>Information Gain</i>	73
4.38 Data <i>Training</i>	74
4.39 Data <i>Testing</i>	74
4.40 Perhitungan Nilai <i>Mean</i>	77
4.41 Perhitungan Nilai <i>Standar Deviasi</i>	81
4.42 Perhitungan Nilai <i>Gaussian</i> Aktif	85
4.43 Perhitungan Nilai <i>Gaussian Drop Out</i>	88
4.44 Perhitungan Nilai Probabilitas <i>Gaussian</i> Aktif dan <i>Drop Out</i>	89
4.45 Data Aktual dan Prediksi	90
4.46 Evaluasi Algoritma <i>Naive Bayes</i>	91
4.47 Uji Coba Perbandingan Data <i>Training</i> dan Data <i>Testing</i>	92
4.48 Hasil Pengujian Atribut Algoritma <i>Information Gain</i>	93
4.49 Data <i>Training</i>	94
4.50 Data <i>Testing</i>	94
4.51 Evaluasi Algoritma <i>Naive Bayes</i> dengan <i>Information Gain</i>	96
4.52 Uji Coba Perbandingan Data <i>Training</i> dan Data <i>Testing</i>	97
4.53 Perbandingan Hasil	98

## DAFTAR GAMBAR

3.1 Alur Penelitian	48
4.1 Sebaran Data	52
4.2 <i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i>	90
4.3 <i>Confusion Matrix</i> Algoritma <i>Naive Bayes</i> dengan <i>Information Gain</i>	95



UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR PUSTAKA

- Alban, M. and Mauricio, D. (2019). Neural Networks to Predict Dropout at The Universities. *International Journal of Machine Learning and Computing*, 9(2):149–153.
- Anwarudin, A., Andriyani, W., DP, B. P., and Kristomo, D. (2022). The Prediction on the Students' Graduation Timeliness Using Naive Bayes Classification and K-Nearest Neighbour. *J. Intell. Softw. Syst.*, 1(1):75.
- Awaludin, M., Yasin, V., and Wahyuningsih, M. (2022). Optimization of Naïve Bayes Algorithm Parameters for Student Graduation Prediction At Universitas Dirgantara Marsekal Suryadarma. *Journal of Information System, Informatics and Computing Issue Period*, 6(1):91–106.
- Azmi, A. U., Hadi, A. F., Anggraeni, D., and Riski, A. (2021). Naive Bayes Methods for Rainfall Prediction Classification in Banyuwangi. *Journal of Physics: Conference Series*, 1872(1):012028.
- Bachhal, P., Ahuja, S., and Gargrish, S. (2021). Educational Data Mining: A Review. *Journal of Physics: Conference Series*, 1950(1):012022.
- Bedregal Alpaca, N., Cornejo Aparicio, V., Zarate Valderrama, J., and Yanque Churo, P. (2020). Classification Models for Determining Types of Academic Risk and Predicting Dropout in University Students. *International Journal of Advanced Computer Science and Applications*, 11(1):266–272.
- Bimantoro, A. S., Athallah, R., Al-azhar, N. I., and Legowo, N. (2022).



- Implementation of Data Mining to Determine Payment Delays for Mall Shopping Center Tenants Using K-Means Clustering Method. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, 5(3):20614–20626.
- Cahyaningtyas, C., Manongga, D., and Sembiring, I. (2022). Algorithm Comparison and Feature Selection For Classification Of Broiler Chicken Harvest. *Jurnal Teknik Informatika (Jutif)*, 3(6):1717–1727.
- Casanova, J. R., Cervero, A., Núñez, J. C., Almeida, L. S., and Bernardo, A. (2018). Factors That Determine The Persistence and Dropout of University Students. *Psicothema*, 30(4):408–414.
- Charoenpong, J., Pimpunchat, B., Amornsamankul, S., Triampo, W., and Nuttavut, N. (2019). A Comparison of Machine Learning Algorithms and Their Applications. *International Journal of Simulation: Systems, Science & Technology*, 20(4):1–17.
- Chen, H., Hu, S., Hua, R., and Zhao, X. (2021). Improved Naive Bayes Classification Algorithm for Traffic Risk Management. *EURASIP Journal on Advances in Signal Processing*, 6(1):1–12.
- Cigdem, O. and Demirel, H. (2018). Performance Analysis of Different Classification Algorithms Using Different Feature Selection Methods on Parkinson's Disease Detection. *Journal of Neuroscience Methods*, 309:81–90.
- Dasi, H. and Kanakala, S. (2022). Student Dropout Prediction Using Machine Learning Techniques. *International Journal of Intelligent Systems and Applications in Engineering*, 10(4):408–414.
- Dewantoro, S., Herdiani, A., and Puspendari, D. (2019). Implementasi Information

- Gain sebagai Feature Selection pada Word Sense Disambiguation Bahasa Indonesia dengan Teknik Klasifikasi Decision List. *eProceedings of Engineering*, 6(3):10425–10435.
- Eriyani, D., Sholihah, V. M., Kusna, A., Mahardika, B. A., Nugraheni, D. R., Yulindasari, N. O., Taftania, S., and Nurabadi, A. (2020). Changes in the New Student Selection System as an Effort to Improve Prospective Student Achievements. In *1st International Conference on Information Technology and Education (ICITE 2020)*, volume 508, pages 446–452.
- Fahrudy, D. and ‘Uyun, S. (2022). Classification of Student Graduation by Naïve Bayes Method by Comparing between Random Oversampling and Feature Selections of Information Gain and Forward Selection. *JOIV: International Journal on Informatics Visualization*, 6(4):798–808.
- Firmansyah and Yulianto, A. (2021). Machine Learning Dengan Decision Tree untuk Prediksi Pembayaran Invoice, Case Study: Gramedia Jakarta. *Journal of Informatics and Telecommunication Engineering*, 5(1):167–175.
- Fitriani, A. S., Fajrillah, and Novarika, W. (2019). Implementation of Data Mining Using Naïve Bayes Classification Method To Predict Participation of Governor And Vocational Governor Selection in Jemirahan Village, Jabon District. *The IJICS (International Journal of Informatics and Computer Science)*, 3(2):66–79.
- Ghorbani, R. and Ghousi, R. (2019). Predictive Data Mining Approaches in Medical Diagnosis: A Review of Some Diseases Prediction. *International Journal of Data and Network Science*, 3(2):47–70.
- Gil, J. S., Delima, A. J. P., and Vilchez, R. N. (2020). Predicting Students’ Dropout Indicators in Public School using Data Mining Approaches. *International*

*Journal of Advanced Trends in Computer Science and Engineering*, 9(1):774–778.

Gosu, D. J. S. D., Priyadarsini, D. P., and Motupalli, R. K. (2021). A Hybrid Approach for the Analysis of Feature Selection using Information Gain and BAT Techniques on The Anomaly Detection. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(5):656–666.

Gunawan, D. (2020). Classification of Privacy Preserving Data Mining Algorithms: A Review. *Jurnal Elektronika dan Telekomunikasi*, 20(2):36–46.

Gunawan, M., Zarlis, M., and Roslina, R. (2021). Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbour Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Media Informatika Budidarma*, 5(2):513–523.

Harwati, H., Viridianawaty, R. I., and Mansur, A. (2016). Drop out Estimation Students based on the Study Period: Comparison between Naïve Bayes and Support Vector Machines Algorithm Methods. *IOP Conference Series: Materials Science and Engineering*, 105(1):012039.

Hasibuan, M. R. and Marji (2019). Pemilihan Fitur dengan Information Gain untuk Klasifikasi Penyakit Gagal Ginjal menggunakan Metode Modified K-Nearest Neighbour (MKNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(11):10435–10443.

Haut, J. M., Paoletti, M. E., Plaza, J., Li, J., and Plaza, A. (2018). Active Learning With Convolutional Neural Networks For Hyperspectral Image Classification Using A New Bayesian Approach. *IEEE Transactions on Geoscience and Remote Sensing*, 56(11):6440–6461.

- Herfandi, H., Zaen, M. T. A., Yuliadi, Y., Julkarnain, M., and Hamdani, F. (2021). Application of Information Gain to Select Attributes in Improving Naïve Bayes Accuracy in Predicting Customer's Payment Capability. *JISA(Jurnal Informatika dan Sains)*, 4(2):155–163.
- Hidayanti, I., Kurniawan, T. B., and Afriyudi, A. (2020). Perbandingan Dan Analisis Metode Klasifikasi Untuk Menentukan Konsentrasi Jurusan. *Jurnal Ilmiah Informatika Global*, 11(1):16–21.
- Hozairi, H., Anwari, A., and Alim, S. (2021). Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbour, Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, 6(2):133–144.
- Irmayanti, D., Muhyidin, Y., and Nurjaman, D. A. (2021). Prediksi Mahasiswa Berpotensi Drop Out Dengan Metode Iteratif Dichotomiser 3 (ID3). *Jurnal Teknologi Informasi*, 5(2):103–113.
- Iswanto, I., Tulus, T., and Poltak, P. (2022). Comparison of Feature Selection To Performance Improvement of K-Nearest Neighbor Algorithm in Data Classification. *Jurnal Teknik Informatika (Jutif)*, 3(6):1709–1716.
- Jain, D. and Singh, V. (2018). Feature Selection and Classification Systems for Chronic Disease Prediction: A Review. *Egyptian Informatics Journal*, 19(3):179–189.
- Kabathova, J. and Drlik, M. (2021). Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques. *Applied Sciences*, 11(7):3130.

- Kadafi, M. (2020). Metode Naïve Bayes Classifier (Nbc) Untuk Memprediksi Tingkat Kelulusan Mahasiswa Tepat Waktu. *JSI: Jurnal Sistem Informasi (E-Journal)*, 12(2).
- Kamel, H., Abdulah, D., and Al-Tuwaijari, J. M. (2019). Cancer Classification Using Gaussian Naive Bayes Algorithm. *Proceedings of the 5th International Engineering Conference, IEC 2019*, pages 165–170.
- Khotimah, A. C. and Utami, E. (2022). Comparison Naïve Bayes Classifier, K-Nearest Neighbour and Support Vector Machine in the Classification of Individual on Twitter Account. *Jurnal Teknik Informatika (Jutif)*, 3(3):673–680.
- Khotimah, B. K., Miswanto, M., and Suprajitno, H. (2020). Optimization of Feature Selection Using Genetic Algorithm in Naïve Bayes Classification for Incomplete Data. *International Journal of Intelligent Engineering and Systems*, 13(1):334–343.
- Kurniabudi, K., Harris, A., and Rahim, A. (2020). Seleksi Fitur dengan Information Gain untuk Meningkatkan Deteksi Serangan DDoS Menggunakan Random Forest An Information Gain Feature Selection to Improve DDoS Detection using Random Forest. *Techno. Com*, 19(1):56–66.
- Laili, U. F., Umatin, C., and Ridwanulloh, M. U. (2023). Analisis Potensial Drop Out Mahasiswa Dengan K-Means++ Clustering Dalam Upaya Peningkatan Kualitas IAIN Kediri. *Paedagoria: Jurnal Kajian, Penelitian dan Pengembangan Kependidikan*, 14(2):145–153.
- Li, Z., Jing, Y., Yao, L., and Gan, B. (2019). Unsupervised Feature Selection Algorithm Based on Information Gain. In *2019 International Conference on Big*

*Data, Electronics and Communication Engineering (BDECE 2019)*, volume 94, pages 63–67.

Manrique, R., Nunes, B. P., Marino, O., Casanova, M. A., and Nurmikko-Fuller, T. (2019). An Analysis of Student Representation, Representative Features and Classification Algorithms to Predict Degree Dropout. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 401–410.

Marzuqi, A., Laksitowening, K. A., and Asror, I. (2021). Temporal Prediction on Students' Graduation using Naïve Bayes and K-Nearest Neighbour Algorithm. *Jurnal Media Informatika Budidarma*, 5(2):682–686.

Mayadewi, P. and Rosely, E. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining. In *Seminar Nasional Sistem Informasi Indonesia*, volume 2, pages 329–334.

Mohammad Suhaimi, N., Abdul-Rahman, S., Mutalib, S., Abdul Hamid, N. H., and Hamid, A. (2019). Review on Predicting Students' Graduation Time Using Machine Learning Algorithms. *International Journal of Modern Education and Computer Science*, 11(7):1–13.

Muqorobin, M., Kusri, K., and Luthfi, E. T. (2019). Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah. *Jurnal Ilmiah SINUS*, 17(1):1–14.

Mustafid, M. F., Wedi, A., and Adi, E. P. (2020). Perbedaan Indeks Prestasi Kumulatif (IPK) Berdasarkan Gaya Belajar Pada Mahasiswa Jurusan Teknologi Pendidikan Universitas Negeri Malang Angkatan 2017. *Jurnal Inovasi dan Teknologi Pembelajaran (JINOTEP): Kajian dan Riset Dalam Teknologi Pembelajaran*, 6(2):119–128.



- Nagy, M. and Molontay, R. (2018). Predicting Dropout in Higher Education Based on Secondary School Performance. In *2018 IEEE 22nd international conference on intelligent engineering systems (INES)*, pages 000389–000394. IEEE.
- Nalim, N., Dewi, H. L., and Safii, M. A. (2021). Analisis Faktor-Faktor yang Mempengaruhi Keberhasilan Studi Mahasiswa di PTKIN Provinsi Jawa Tengah. *Jurnal Kependidikan: Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran*, 7(4):1003–1013.
- Nasrullah, A. H. (2018). Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out. *ILKOM Jurnal Ilmiah*, 10(2):244–250.
- Norhalimi, M. and Siswa, T. A. Y. (2022). Optimasi Seleksi Fitur Information Gain pada Algoritma Naïve Bayes dan K-Nearest Neighbour. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 7(3):237–255.
- Nurhachita and Negara, E. S. (2021). A Comparison Between Deep Learning, Naïve Bayes and Random Forest for The Application of Data Mining On The Admission of New Students. *IAES International Journal of Artificial Intelligence*, 10(2):324–331.
- Oktanisa, I. and Supianto, A. A. (2018). Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(5):567.
- Palete, S. and Hidayat, A. (2021). Perhitungan Biaya Kuliah Tunggal Program Studi Diploma Universitas Fajar. *YUME: Journal of Management*, 4(3):87–94.
- Peling, I. B. A., Arnawan, I. N., Arthawan, I. P. A., and Janardana, I. G. N. (2017). Implementation of Data Mining To Predict Period of Students Study Using

- Naive Bayes Algorithm. *International Journal of Engineering and Emerging Technology*, 2(1):53.
- Qisthiano, M. R., Kurniawan, T. B., Negara, E. S., and Akbar, M. (2021). Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(3):987–994.
- Rahmawati, S., Silviah, M. N., and Husna, N. S. (2021). Implementation of Data Mining in Shopping Cart Analysis using the Apriori Algorithm. *Internasional Journal of Data Science, Engineering, and Analytics*, 1(1):30–36.
- Ramadhani, L. K. and Widyaningrum, B. N. (2022). Perbandingan Metode Klasifikasi Naive Bayes Dan Support Vector Machine Pada Predikat Kelulusan Mahasiswa. *LEDGER: Journal Informatic and Information Technology*, 1(3):150–154.
- Ratniasih, N. L. (2020). Penerapan Algoritma Klasifikasi Untuk Penentuan Mahasiswa Berpotensi Drop Out. *Jurnal Teknologi Informasi dan Komputer*, 6(3):21315101.
- Razaque, F., Soomro, N., Shaikh, S. A., Soomro, S., Samo, J. A., Kumar, N., and Dharejo, H. (2017). Using Naive Bayes Algorithm to Students' Bachelor Academic Performances Analysis. In *2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, volume 2018-Janua, pages 1–5.
- Rismayati, R., Ismarmiaty, I., and Hidayat, S. (2022). Ensemble Implementation for Predicting Student Graduation with Classification Algorithm. *Int. J. Eng. Comput. Sci. Appl.*, 1(1):35–42.

- Santoso, P. H., Fauziah, and Nurhayati (2020). Application of Data Mining Classification for Covid-19 Infected Status Using Algoritma Naïve Method. *Jurnal Mantik*, 4(1):267—275.
- Sari, N. W. W., Mirwansyah, D., Fahrullah, F., Tandi, I. L., and Sopyani, A. H. H. (2022). Prediction of Students Drop Out With Naïve Bayes. *Jurnal Mantik*, 6(3):3858–3863.
- Saritas, M. M. and Yasar, A. (2019). Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification. *International journal of intelligent systems and applications in engineering*, 7(2):88–91.
- Sepriansyah, R. and Purnamasari, S. D. (2022). Prediction of Student Graduation Using Naïve Bayes. *Budapest International Research and Critics Intitute-Journal (BIRCI-Journal)*, 5(3):24255–24268.
- Setiyani, L., Wahidin, M., Awaludin, D., and Purwani, S. (2020). Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review. *Faktor Exacta*, 13(1):35–43.
- Sinaga, D., Solaiman, E. J., and Kaunang, F. J. (2021). Penerapan Algoritma Decision Tree C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop out Di Universitas Advent Indonesia. *TeIKa*, 11(2):167–173.
- Singh, D. and Singh, B. (2020). Investigating The Impact of Data Normalization On Classification Performance. *Applied Soft Computing*, 97:105524.
- Somantri, O. and Khambali, M. (2017). Feature Selection Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes dan Algoritme Genetika. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 6(3):301–306.

- Suryadi, A. (2019). Integration of Feature Selection with Data Level Approach for Software Defect Prediction. *Sinkron: jurnal dan penelitian teknik informatika*, 4(1):51–57.
- Suryanegara, G. A. B., Adiwijaya, and Purbolaksono, M. D. (2021). Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(1):114–122.
- Sutoyo, E. and Almaarif, A. (2020). Educational Data Mining for Predicting Student Graduation Using the Naïve Bayes Classifier Algorithm. *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, 4(1):95–101.
- Tanjung, R. H., Yunus, Y., and Nurcahyo, G. W. (2023). Perbandingan Algoritma C4. 5 dan Naive Bayes Dalam Prediksi Kelulusan Mahasiswa. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1):193—199.
- Triayudi, A. and Widyarto, W. O. (2021). Educational Data Mining Analysis Using Classification Techniques. In *Journal of Physics: Conference Series*, volume 1933, page 012061.
- Utari, M., Warsito, B., and Kusumaningrum, R. (2020). Implementation of Data Mining for Drop Out Prediction using Random Forest Method. In *2020 8th International Conference on Information and Communication Technology (ICoICT)*, pages 1–5.
- Vangara, R. V. B., Thirupathur, K., and Vangara, S. P. (2020). Opinion Mining Classification using Naive Bayes Algorithm. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9(5):495–498.

- Wanti, N., Sari, W., Mirwansyah, D., Tandil, I. L., and Sopyani, H. H. (2022). Prediction of Students Drop Out With Naïve Bayes. *Jurnal Mantik*, 6(3):582–586.
- 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. *J. Tekno Insentif*, 13(1):16–25.
- Yaacob, W. F. W., Sobri, N. M., Nasir, S. A., Yaacob, W. F. W., Norshahidi, N. D., and Wan Husin, W. Z. (2020). Predicting Student Drop Out in Higher Institution Using Data Mining Techniques. In *Journal of Physics: Conference Series*, volume 1496, page 012005.
- Yualinda, S., Wijaya, D. R., and Hernawati, E. (2020). Aplikasi Berbasis Dataset E-Commerce Untuk Prediksi Kemiskinan Menggunakan Algoritma Naïve Bayes, Xgboost Dan Similarity Based Feature Selection. *e-Proceeding of Applied Science*, 6(2):9–11.
- Yunianto, I., Kurniawan, A., and Malik Mutoffar, M. (2022). Comparison of Decision Tree, KNN and Naïve Bayes Methods In Predicting Student Late Graduation In the Informatics Engineering Department, Institute Business XYZ. *Adpebi International Journal of Multidisciplinary Sciences*, 1(1):374–383.