

**PENGUNAAN METODE *UNDERSAMPLING* PADA KLASIFIKASI
PERSONALITY CALON SISWA MADRASAH DENGAN
ALGORITMA *NAIVE BAYES***

SKRIPSI



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

Disusun Oleh:

**ASIQOTUL KHUSNIYAH
H96219043**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA
2023**

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini,

Nama : Asiqotul Khusniyah

NIM : H96219043

Program Studi : Sistem Informasi

Angkatan : 2019

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi yang berjudul “PENGUNAAN METODE UNDERSAMPLING PADA KLASIFIKASI PERSONALITY CALON SISWA MADRASAH DENGAN ALGORITMA NAIVE BAYES”. Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Surabaya, 9 Juli 2023

Yang Menyatakan,



(Asiqotul Khusniyah)

NIM: H96219043

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi Oleh

NAMA : ASIQOTUL KHUSNIYAH

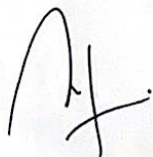
NIM : H96219043

JUDUL : PENGGUNAAN METODE UNDERSAMPLING PADA
KLASIFIKASI *PERSONALITY* CALON SISWA
MADRASAH DENGAN ALGORITMA NAIVE BAYES

Ini telah diperiksa dan disetujui untuk diujikan

Surabaya, 10 Juli 2023

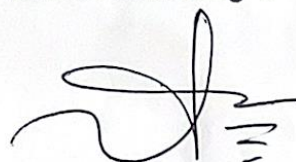
Dosen Pembimbing 1



Ahmad Yusuf. M. Kom.

NIP. 199001202014031003

Dosen Pembimbing 2



Noor Wahyudi, M. Kom.

NIP. 198403232014031002

LEMBAR PENGESAHAN

JUDUL : PENGGUNAAN METODE UNDERSAMPLING PADA
KLASIFIKASI *PERSONALITY* CALON SISWA
MADRASAH DENGAN ALGORITMA NAIVE BAYES

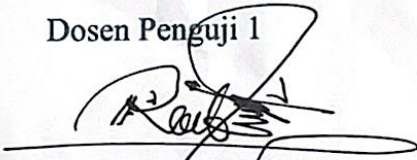
NAMA : ASIQOTUL KHUSNIYAH

NIM : H96219043

Skripsi tersebut telah dipresentasikan pada Sidang Skripsi didepan Dosen Penguji
pada, 14 Juli 2023.


Menyetujui,

Dosen Penguji 1



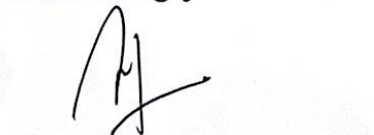
Mujib Ridwan, S.Kom., M.T.
NIP. 198604272014031004

Dosen Penguji 2



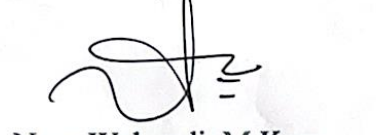
Dwi Rolliawati, M.T.
NIP. 197909272014032001

Dosen Penguji 3



Ahmad Yusuf, M.Kom.
NIP. 199001202014031003

Dosen Penguji 4



Noor Wahyudi, M.Kom.
NIP. 198403232014031002

Mengetahui,

Dekan Fakultas Sains dan Teknologi
Universitas Gunadarma Ampel Surabaya



ABUL HAMDANI, M.Pd.
NIP. 196507312000031002



UIN SUNAN AMPEL
SURABAYA

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 : Asiqotul Khusniyah
NIM : H96219043
Fakultas/Jurusan : Sains dan Teknologi / Sistem Informasi
E-mail address : H96219043@student.uinsby.ac.id

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

Skripsi Tesis Desertasi Lain-lain (.....)
yang berjudul :

PENGUNAAN METODE *UNDERSAMPLING* PADA KLASIFIKASI *PERSONALITY*
CALON SISWA MADRASAH DENGAN ALGORITMA *NAIVE BAYES*

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, 15 Juli 2023

Penulis

(Asiqotul Khusniyah)
nama terang dan tanda tangan

ABSTRAK

PENGGUNAAN METODE *UNDERSAMPLING* PADA KLASIFIKASI *PERSONALITY* CALON SISWA MADRASAH DENGAN ALGORITMA *NAIVE BAYES*

Oleh:

Asiqotul Khusniyah

Kepribadian setiap individu dipengaruhi oleh berbagai faktor, seperti aspek biologis, lingkungan fisik, kebudayaan, dan pengalaman-pengalaman yang dialami. Dengan adanya klasifikasi dan juga dilakukan *feature selection* dalam menentukan fitur atau faktor paling berpengaruh pada kepribadian. Pada data yang digunakan memiliki masalah salah satunya penanganan data yang tidak seimbang dalam klasifikasi. Untuk mengatasi masalah tersebut, peneliti menggunakan teknik *resampling* data *undersampling*. Teknik yang digunakan dalam mengatasi permasalahan *imbalanced* dataset yaitu *Random undersampling* dan *Tomek links*. Penelitian ini bertujuan mencari teknik *resampling* data *undersampling* terbaik. Penerapan teknik *undersampling* terbukti dapat meningkatkan kinerja klasifikasi pada data yang tidak seimbang. Terlihat adanya peningkatan dalam nilai presisi, recall dan f1-score setelah dilakukan metode *undersampling* pada teknik *random undersampling* dibandingkan sebelum dilakukan metode *undersampling*. Terlihat adanya peningkatan dalam nilai presisi pada skenario *random undersampling* sebesar 14% dan pada skenario *tomek links* 2%, nilai recall mengalami peningkatan pada skenario *random undersampling* sebesar 6% dan pada skenario *tomek links* sebesar 1%, dan pada nilai f1-score pada skenario *random undersampling* meningkat sebesar 10% dan pada skenario *tomek links* 2% dibandingkan dengan skenario *no resampling*. Oleh karena itu, dapat disimpulkan bahwa metode *undersampling resampling* data teknik *random undersampling* memiliki kemampuan yang lebih baik dalam menyeimbangkan data dibandingkan teknik *tomek links*.

Kata kunci: *Data Mining, Klasifikasi, Resampling, Undersampling, K-fold cross validation, Naive bayes.*

ABSTRACT

THE USE OF UNDERSAMPLING METHOD IN THE PERSONALITY CLASSIFICATION OF PROSPECTIVE MADRASAH STUDENTS WITH THE NAÏVE BAYES ALGORITHM

By:

Asiqotul Khusniyah

The personality of each individual is influenced by various factors, such as biological aspects, physical environment, culture, and experiences. With the classification and also carried out feature selection in determining the most influential features or factors on personality. The data used has problems, one of which is unbalanced data handling in classification. To overcome this problem, researchers use undersampling data resampling techniques. The techniques used in overcoming imbalanced dataset problems are Random undersampling and Tomek links. This study aims to find the best undersampling data resampling technique. The application of undersampling techniques is proven to improve classification performance on unbalanced data. There was an increase in precision, recall and f1-score values after the undersampling method was carried out in the random undersampling technique compared to before the undersampling method. There was an increase in the precision value in the random undersampling scenario by 14% and in the tokek links scenario by 2%, the recall value increased in the random undersampling scenario by 6% and in the tokek links scenario by 1%, and in the f1-score value in the random undersampling scenario increased by 10% and in the tokek links scenario by 2% compared to the no sampling scenario. Therefore, it can be concluded that the undersampling method of data resampling random undersampling technique has a better ability to balance data than the tokek links technique.

Keyword: Data Mining, Classification, Resampling, Undersampling, K-fold cross validation, Naive bayes.

DAFTAR ISI

HALAMAN JUDUL.....	ii
LEMBAR PERSETUJUAN PEMBIMBING	iii
LEMBAR PENGESAHAN	iv
PERNYATAAN KEASLIAN.....	v
MOTTO DAN PERSEMBAHAN	vi
KATA PENGANTAR	viii
ABSTRAK	ix
<i>ABSTRACT</i>	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
BAB 2 TINJAUAN PUSTAKA	5
2.1 Tinjauan Penelitian Terdahulu	5
2.2 Teori Dasar	9
2.2.1. <i>Personality</i>	9
2.2.2. Faktor yang mempengaruhi <i>Personality</i>	10
2.2.3. Data Mining	11
2.2.4. Klasifikasi	13
2.2.5. <i>Naive bayes</i>	13
2.2.6. <i>Feature selection</i>	15
2.2.7. LASSO Regression	15
2.2.8. <i>Balancing data</i>	16
2.2.9. <i>Undersampling</i>	17
2.2.10. <i>Random undersampling</i>	18
2.2.11. <i>Tomek links</i>	19

2.2.12. <i>K-Fold Cross Validation</i>	20
2.2.13. <i>Confusion matrix</i>	21
2.2.14. <i>Python</i>	24
2.3 Integrasi Keilmuan	25
BAB 3 METODOLOGI PENELITIAN	27
3.1 Kerangka Penelitian	27
3.1.1 Perumusan Masalah	27
3.1.2 Studi Pustaka.....	28
3.1.3 Pengumpulan Data	28
3.1.4 Pengolahan Data	29
3.1.5 Analisis Hasil	30
BAB 4 HASIL DAN PEMBAHASAN	31
4.1 Hasil Penelitian	31
4.1.1 Pengumpulan data.....	31
4.1.2 Pengolahan Data	32
4.1.3 Skenario Model Klasifikasi	34
4.1.4 Evaluasi Model	37
4.2 Analisis dan Pembahasan	42
4.2.1 Analisis hasil.....	42
4.2.2 Pembahasan hasil	44
BAB 5 PENUTUP	46
5.1 Kesimpulan.....	46
5.2 Saran.....	47
DAFTAR PUSTAKA	48

DAFTAR GAMBAR

Gambar 2.1 Undersampling vs oversampling	17
Gambar 2.2 Flowchart random undersampling.....	19
Gambar 2.3 Flowchart tokek links	20
Gambar 2.4 Ilustrasi 10-fold cross validation	21
Gambar 4.1 Perbandingan data pada atribut target	32
Gambar 4.2 perbandingan data atribut target setelah cleaning data.....	33



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu.....	5
Tabel 2.2 Confusion matrix.....	22
Tabel 3.1 Keterangan kelas target.....	28
Tabel 3.2 Deskripsi atribut dataset snpdb 2022	28
Tabel 3.3 Skenario resampling data	30
Tabel 4.1 Data yang mengandung <i>missing value</i>	33
Tabel 4.2 Hasil <i>feature selection</i> lasso regression	34
Tabel 4.3 Hasil <i>confusion matrix</i> skenario <i>no resampling</i>	38
Tabel 4.4 Hasil <i>10-fold cross validation</i> klasifikasi skenario <i>no resampling</i>	39
Tabel 4.5 Hasil <i>confusion matrix</i> skenario <i>random undersampling</i>	39
Tabel 4.6 Hasil <i>10-fold cross validation</i> klasifikasi skenario <i>random undersampling</i>	40
Tabel 4.7 Hasil <i>confusion matrix</i> skenario <i>tomek links</i>	41
Tabel 4.8 Hasil <i>10-fold cross validation</i> klasifikasi skenario <i>tomek links</i>	42
Tabel 4.9 Analisis skenario <i>no resampling</i>	43
Tabel 4.10 Analisis skenario <i>random undersampling</i>	43
Tabel 4.11 Analisis skenario <i>tomek links</i>	43
Tabel 4.12 Perbandingan analisis tiga skenario	44

UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR PUSTAKA

- Agustika, R. (2021). *Penerapan Kombinasi SMOTE dan Tomek Links untuk Klasifikasi Data Tidak Seimbang dengan Metode Random Forest* [Universitas Gadjah Mada]. <http://etd.repository.ugm.ac.id/penelitian/detail/199065>
- Al-Ashoor, A., & Abdullah, S. (2022). Examining Techniques to Solving Imbalanced Datasets in Educational Data Mining Systems. *International Journal of Computing*, 21(2), 205–213. <https://doi.org/10.47839/ijc.21.2.2589>
- Anam, K., Ismail, H., Hanggara, F. S., Avian, C., & Worsito, S. B. (2021). Cross Validation Configuration on k-NN for Finger Movements using EMG signals. *2021 International Conference on Instrumentation, Control, and Automation (ICA)*. <https://www.semanticscholar.org/paper/Cross-Validation-Configuration-on-k-NN-for-Finger-Anam-Ismail/cad9cb101025d567958a6af62ac57b4af15caf50>
- Apriliyani, E., & Salim, Y. (2022). Analisis performa metode klasifikasi Naïve Bayes Classifier pada Unbalanced Dataset. *Indonesian Journal of Data and Science*, 3(2), Article 2. <https://doi.org/10.56705/ijodas.v3i2.45>
- Araka, E., Oboko, R., Maina, E., & Gitonga, R. (2022). Using Educational Data Mining Techniques to Identify Profiles in Self-Regulated Learning: An Empirical Evaluation. *The International Review of Research in Open and Distributed Learning*, 23(1), 131–162. <https://doi.org/10.19173/irrodl.v22i4.5401>
- Arisandi, R. R. R., Warsito, B., & Hakim, A. R. (2022). Aplikasi Naïve Bayes Classifier (NBC) pada Klasifikasi Status Gizi Balita Stunting dengan Pengujian K-Fold Cross Validation. *Jurnal Gaussian*, 11(1), 130–139. <https://doi.org/10.14710/j.gauss.v11i1.33991>
- Artana, I. M., & Utami, N. W. (2022). Penerapan Data Mining untuk Menentukan Strategi Promosi Produk Industri Kreatif UMKM Kota Denpasar Pasca Pandemi Covid 19. *Jurnal Informatika Teknologi dan Sains*, 4(2), Article 2. <https://doi.org/10.51401/jinteks.v4i2.2032>

- Bagui, S., & Li, K. (2021). Resampling imbalanced data for network intrusion detection datasets. *Journal of Big Data*, 8(1), 6. <https://doi.org/10.1186/s40537-020-00390-x>
- Choirunnisa, S. (2021). *Metode Hibrida Oversampling dan Undersampling untuk Menangani Ketidakseimbangan Data Kegagalan Akademik Universitas XYZ* [Masters, Institut Teknologi Sepuluh Nopember]. <https://repository.its.ac.id/60454/>
- Costanzo, P., & Orphanou, K. (2022). *Performance, Transparency and Time. Feature selection to speed up the diagnosis of Parkinson's disease* (arXiv:2206.03716). arXiv. <https://doi.org/10.48550/arXiv.2206.03716>
- Dablain, D., Krawczyk, B., & Chawla, N. V. (2021, Mei 5). *DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data*. ArXiv.Org. <https://arxiv.org/abs/2105.02340v1>
- Durugkar, S. R., Raja, R., Nagwanshi, K. K., & Kumar, S. (2022). Introduction to Data Mining. Dalam *Data Mining and Machine Learning Applications* (hlm. 1–19). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119792529.ch1>
- Esposito, C., Landrum, G. A., Schneider, N., Stiefl, N., & Riniker, S. (2021). GHOST: Adjusting the Decision Threshold to Handle Imbalanced Data in Machine Learning. *Journal of Chemical Information and Modeling*, 61(6), 2623–2640. <https://doi.org/10.1021/acs.jcim.1c00160>
- Fiorentini, N., & Losa, M. (2020). Handling Imbalanced Data in Road Crash Severity Prediction by Machine Learning Algorithms. *Infrastructures*, 5(7), 61. <https://doi.org/10.3390/infrastructures5070061>
- Firmansyach, W. A., Hayati, U., & Wijaya, Y. A. (2023). Analisa Terjadinya Overfitting dan Underfitting pada Algoritma Naive Bayes dan Decision Tree dengan Teknik Cross Validation. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), Article 1. <https://doi.org/10.36040/jati.v7i1.6329>
- Fitria, R. I., & Ujianto, N. T. (2022). Komparasi Algoritma Logistic Regression dan Naive Bayes untuk Penerimaan Siswa Baru pada SMK di Brebes. *Journal of Informatics, Artificial Intelligence and Internet of Thing*, 1(1), Article 1.

- Hakim, L., Kristanto, S. P., Yusuf, D., & Afia, F. N. (2022). Pengenalan Motif Batik Banyuwangi Berdasarkan Fitur Grey Level Co-Occurrence Matrix. *Jurnal Teknoinfo*, 16(1), 1. <https://doi.org/10.33365/jti.v16i1.1320>
- Hao, J., & Ho, T. K. (2019). Machine Learning Made Easy: A Review of *Scikit-learn* Package in Python Programming Language. *Journal of Educational and Behavioral Statistics*, 44(3), 348–361. <https://doi.org/10.3102/1076998619832248>
- Hayami, R., Soni, & Gunawan, I. (2022). Klasifikasi Jamur Menggunakan Algoritma Naïve Bayes. *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(1), 28–33. <https://doi.org/10.37859/coscitech.v3i1.3685>
- Helmus, J. J., & Collis, S. M. (2016). The Python ARM Radar Toolkit (Py-ART), a Library for Working with Weather Radar Data in the Python Programming Language. *Journal of Open Research Software*, 4(1), 25. <https://doi.org/10.5334/jors.119>
- Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. *IEEE Access*. <https://www.semanticscholar.org/paper/MLCM%3A-Multi-Label-Confusion-Matrix-Heydarian-Doyle/ded74d65b18b067c34262ceff633feb476590aee>
- Indrawati, A. (2021). Penerapan Teknik Kombinasi Oversampling dan Undersampling untuk Mengatasi Permasalahan Imbalanced Dataset. *JIKO (Jurnal Informatika dan Komputer)*, 4(1), Article 1. <https://doi.org/10.33387/jiko.v4i1.2561>
- Istighfar, F., Negara, A. B. P., & Tursina, T. (2023). Klasifikasi Bidang Keahlian Mahasiswa Menggunakan Algoritma Naive Bayes. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, 11(1), Article 1. <https://doi.org/10.26418/justin.v11i1.52402>
- Jeong, D.-H., Kim, S.-E., Choi, W.-H., & Ahn, S.-H. (2022). A Comparative Study on the Influence of Undersampling and Oversampling Techniques for the Classification of Physical Activities Using an Imbalanced Accelerometer Dataset. *Healthcare*, 10(7). <https://doi.org/10.3390/healthcare10071255>

- Kaur, H., Pannu, H. S., & Malhi, A. K. (2020). A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions. *ACM Computing Surveys*, 52(4), 1–36. <https://doi.org/10.1145/3343440>
- Khushi, M., Shaukat, K., Alam, T. M., Hameed, I. A., Uddin, S., Luo, S., Yang, X., & Reyes, M. C. (2021). A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data. *IEEE Access*, 9, 109960–109975. <https://doi.org/10.1109/ACCESS.2021.3102399>
- Kumar, P., Bhatnagar, R., Gaur, K., & Bhatnagar, A. (2021). Classification of Imbalanced Data: Review of Methods and Applications. *IOP Conference Series: Materials Science and Engineering*, 1099(1), 012077. <https://doi.org/10.1088/1757-899X/1099/1/012077>
- Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem. *The 14th Pervasive Technologies Related to Assistive Environments Conference*. <https://www.semanticscholar.org/paper/Multi-Class-Confusion-Matrix-Reduction-method-and-Markoulidakis-Kopsiaftis/1173b572eddb66b76f10046102cd4cf91f1381ed>
- Mathews, L., & Hari, S. (2019). Learning From Imbalanced Data. *Advances in Computer and Electrical Engineering*. <https://www.semanticscholar.org/paper/Learning-From-Imbalanced-Data-Mathews-Hari/6a7364f6ed2846ea2b705336a4c49dd287102a50>
- Minarno, A. E., Wandani, L. R., & Azhar, Y. (2022). Classification of Breast Cancer Based on Histopathological Image Using EfficientNet-B0 on Convolutional Neural Network. *International Journal of Emerging Technology and Advanced Engineering*. <https://www.semanticscholar.org/paper/Classification-of-Breast-Cancer-Based-on-Image-on-Minarno-Wandani/1d23b596281e98cb51835d25f2b2d109b3439e15>
- Mufarrihah, I., Andriani, A., & Lazulfa, I. (2022). Mengenalai Kepribadian dan Kaitannya dalam Pemilihan Jurusan Kuliah bagi Siswa SMA. *DINAMIS:*

- Jurnal Pengabdian Kepada Masyarakat*, 2(2), 102–106.
<https://doi.org/10.33752/dinamis.v2i2.703>
- Mukti, M. W., & Saputri, T. A. (2022). *Prediksi Kelulusan Tepat Waktu Mahasiswa Menggunakan Algoritma C4.5 pada STMIK Dharma Wacana*. 6(2), 25–29.
- Mutmainah, S. (2021). Penanganan Imbalance Data pada Klasifikasi Kemungkinan Penyakit Stroke. *Jurnal SNATi*, 1(1).
- Nugroho, A., & Religia, Y. (2021). *Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging*.
<https://www.semanticscholar.org/paper/Analisis-Optimasi-Algoritma-Klasifikasi-Naive-Bayes-Nugroho-Religia/94816fef2eaea9ce345797bf049826f58d0e4f67>
- Nurdian, R. A., Ridwan, M., & Yusuf, A. (2022). Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru. *Jurnal Teknik Informatika Dan Sistem Informasi*, 8(1), Article 1. <https://doi.org/10.28932/jutisi.v8i1.4004>
- Pikriyansah, R., Umbara, F. R., & Sabrina, P. N. (2022). Klasifikasi Daftar Ulang Calon Mahasiswa Baru Dengan Menggunakan Metode Klasifikasi Naive Bayes. *Informatics and Digital Expert (INDEX)*, 4(2), Article 2. <https://doi.org/10.36423/index.v4i2.912>
- Prasetyo, D. A. B. (2022). Implementasi Information Schema Database Pada PostgreSQL untuk Pembuatan Tabel Informasi dengan Menggunakan Python Di PT XYZ. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(3), 1961–1972. <https://doi.org/10.35957/jatisi.v9i3.2221>
- Pratiwi, R., Hayati, M. N., & Prangga, S. (2020). *Perbandingan Klasifikasi Algoritma C5.0 dengan Classification and Regression Tree (Studi Kasus: Data Sosial Kepala Keluarga Masyarakat Desa Teluk Baru Kecamatan Muara Ancalong Tahun 2019)*.
<https://www.semanticscholar.org/paper/PERBANDINGAN-KLASIFIKASI-ALGORITMA-C5.0-DENGAN-AND-Pratiwi-Hayati/246a98950fb27c4b2f5aff63f1a45dbad69ed501>
- Rahayu, L. D. (2020). *Kritik Entitas Homogen Masyarakat Adat pada Putusan MK Nomor 35/PUU-X/2012 Melalui Feminist Political Ecology Criticism Of*

The Homogeneous Entities of Indigenous Peoples in the Constitutional Court Decision No. 35/PUUX/2012 Through Feminist Political Ecology. 1(8).

- Rahman, Md. M., Watanobe, Y., Matsumoto, T., Kiran, R. U., & Nakamura, K. (2022). Educational Data Mining to Support Programming Learning Using Problem-Solving Data. *IEEE Access*, 10, 26186–26202. <https://doi.org/10.1109/ACCESS.2022.3157288>
- Raschka, S. (2017). *Naive Bayes and Text Classification I - Introduction and Theory* (arXiv:1410.5329). arXiv. <https://doi.org/10.48550/arXiv.1410.5329>
- Rostami, M., Berahmand, K., & Forouzandeh, S. (2020). *Review of Swarm Intelligence-based Feature Selection Methods* (arXiv:2008.04103). arXiv. <https://doi.org/10.48550/arXiv.2008.04103>
- Safitri, A. R., & Muslim, M. A. (2020). Improved Accuracy of Naive Bayes Classifier for Determination of Customer Churn Uses SMOTE and Genetic Algorithms. *Journal of Soft Computing Exploration*, 1(1), Article 1. <https://doi.org/10.52465/josce.v1i1.5>
- Saifudin, A., & Wahono, R. S. (2015). Pendekatan Level Data Untuk Menangani Ketidakseimbangan Kelas Pada Prediksi Cacat Software—Neliti. *Journal of Software Engineering*, 1(2), 76–85.
- Sáiz-Manzanares, M. C., Rodríguez-Díez, J. J., Díez-Pastor, J. F., Rodríguez-Arribas, S., Marticorena-Sánchez, R., & Ji, Y. P. (2021). Monitoring of Student Learning in Learning Management Systems: An Application of Educational Data Mining Techniques. *Applied Sciences*, 11(6), 2677. <https://doi.org/10.3390/app11062677>
- Saputro, E., & Rosiyadi, D. (2022). Penerapan Metode Random Over-Under Sampling pada Algoritma Klasifikasi Penentuan Penyakit Diabetes. *Jurnal Khatulistiwa Informatika*, 10(1), 42–47. <https://doi.org/10.31294/bi.v10i1.11739>
- Sarwido, Wibowo, G. W. N., & Manan, M. A. (2022). Penerapan Algoritma Naive Bayes untuk Prediksi Heregistrasi Calon Mahasiswa Baru. *JTINFO : Jurnal Teknik Informatika*, 1(1), Article 1. <https://doi.org/10.02220/jtinfo.v1i1.126>

- Sembiring, M. T., & Tambunan, R. H. (2021). Analysis of graduation prediction on time based on student academic performance using the Naïve Bayes Algorithm with data mining implementation (Case study: Department of Industrial Engineering USU). *IOP Conference Series: Materials Science and Engineering*, 1122(1), 012069. <https://doi.org/10.1088/1757-899X/1122/1/012069>
- Shino, Y., Durachman, Y., & Sutisna, N. (2022). Implementation of Data Mining with Naive Bayes Algorithm for Eligibility Classification of Basic Food Aid Recipients. *International Journal of Cyber and IT Service Management*, 2(2), 154–162. <https://doi.org/10.34306/ijcitsm.v2i2.114>
- Simanullang, T. (2021). Pengaruh Tipe Kepribadian The Big Five Model Personality Terhadap Kinerja Aparatur Sipil Negara (Kajian Studi Literatur Manajemen Keuangan). *JURNAL MANAJEMEN PENDIDIKAN DAN ILMU SOSIAL*, 2(2), 747–753. <https://doi.org/10.38035/jmpis.v2i2.634>
- Sinaga, L. M., Sawaluddin, & Suwilo, S. (2020). Analysis of classification and Naïve Bayes algorithm k-nearest neighbor in data mining. *IOP Conference Series: Materials Science and Engineering*, 725(1), 012106. <https://doi.org/10.1088/1757-899X/725/1/012106>
- Sir, Y. A., & Soepranoto, A. H. H. (2022). Pendekatan Resampling Data Untuk Menangani Masalah Ketidakseimbangan Kelas. *Jurnal Komputer dan Informatika*, 10(1), 31–38. <https://doi.org/10.35508/jicon.v10i1.6554>
- Subarkah, P. (2020). Penerapan Algoritme Klasifikasi Classification And Regression Trees (CART) Untuk Diagnosis Penyakit Diabetes Retinopathy. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(2), Article 2. <https://doi.org/10.30812/matrik.v19i2.676>
- Suprpto, S., & Nikmah, Y. L. (2023). Ridge and Lasso Regression for Feature Selection of Overlapping Ibuprofen and Paracetamol UV Spectra. *Moroccan Journal of Chemistry*, 11(1), Article 1. <https://doi.org/10.48317/IMIST.PRSM/morjchem-v11i1.31466>
- Tadlaoui, M. A., & Chekou, M. (2021). A blended learning approach for teaching python programming language: Towards a post pandemic pedagogy.

- International Journal of Advanced Computer Research*, 11(52), 13–22.
<https://doi.org/10.19101/IJACR.2020.1048120>
- Tan, S. W. B., Naraharisetti, P. K., Chin, S. K., & Lee, L. Y. (2020). Simple Visual-Aided Automated Titration Using the Python Programming Language. *Journal of Chemical Education*, 97(3), 850–854.
<https://doi.org/10.1021/acs.jchemed.9b00802>
- Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(2), 785–795.
<https://doi.org/10.35957/jatisi.v9i2.1835>
- Tempola, F., Muhammad, M., & Khairan, A. (2018). Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(5), Article 5. <https://doi.org/10.25126/jtiik.201855983>
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.
<https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Yuniarti, W. D., Faiz, A. N., & Setiawan, B. (2020). *Identifikasi Potensi Keberhasilan Studi Menggunakan Naïve Bayes Classifier*. 2(1), 1–12.
- Zhafirah, S., Masnuna, M., & Yani, A. R. (2021). Buku Ilustrasi 4 Macam Kepribadian Remaja 16-18 Tahun. *Citradirga - Jurnal Desain Komunikasi Visual dan Intermedia*, 2(02), 1–12. <https://doi.org/10.33479/cd.v2i02.323>
- Zhang, Y., Zhao, Z., Zhao, S., Liu, Y., & He, K. (2020). A New Embedded Feature Selection Method using IBALO mixed with MRMR criteria. *Journal of Physics: Conference Series*, 1453(1), 012027. <https://doi.org/10.1088/1742-6596/1453/1/012027>