

**KOMPARASI METODE SMOTE DAN ADASYN DALAM
MENINGKATKAN PERFORMA KLASIFIKASI
HERREGISTRASI MAHASISWA BARU**

SKRIPSI



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

Disusun Oleh :

**RISKY AGUNG NURDIAN
NIM: H06217020**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA
2021**

PERNYATAAN KEASLIAN

Saya yang bertanda tangan dibawah ini:

Nama : Risky Agung Nurdian

NIM : H06217020

Program Studi : Sistem Informasi

Angkatan : 2017

Menyatakan bahwa saya tidak melakukan plagiasi dalam penulisan skripsi saya yang berjudul : “KOMPARASI METODE SMOTE DAN ADASYN DALAM MENINGKATKAN PERFORMA KLASIFIKASI HERREGISTRASI MAHASISWA BARU”. Apabila suatu saat nanti terbukti saya melakukan Tindakan plagiasi, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Bojonegoro, 31 Juli 2021

Yang menyatakan,

A 1000 Rupiah Indonesian postage stamp (Meterai Tempel) with a signature over it. The stamp features the Garuda Pancasila emblem and the text 'SEPLUH RIBU RUPIAH', '1000', 'METERAI TEMPEL', and '4068AAJX01111659'.

(Risky Agung Nurdian)

H06217020

LEMBAR PERSETUJUAN PEMBIMBING SKRIPSI

Skripsi oleh

NAMA : Risky Agung Nurdian

NIM : H06217020

JUDUL : **KOMPARASI METODE SMOTE DAN ADASYN DALAM
MENINGKATKAN PERFORMA KLASIFIKASI HERREGISTRASI
MAHASISWA BARU**

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 21 April 2021

Menyetujui,

Dosen pembimbing I



(Mujib Ridwan, S.Kom.,MT)

NIP.198604272014031004

Dosen pembimbing II



(Ahmad Yusuf, M.Kom)

NIP.199001202014031003

LEMBAR PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Risky Agung Nurdian ini telah dipertahankan
di depan tim penguji skripsi
di Surabaya, 10 Agustus 2021

Mengesahkan,
Dewan Penguji

Penguji I



(Dwi Rolliawati, MT)
NIP 197909272014032001

Penguji II



(Khalid, M. Kom)
NIP 197906092014031002

Penguji III



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

Penguji IV



(Ahmad Yusuf, M. Kom)
NIP 199001202014031003

Mengetahui,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya



(Dr. Evi Fatimatur Rusydiyah, M.Ag)
NIP 197312272005012003



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 : RISKY AGUNG NURDIAN
NIM : H06217020
Fakultas/Jurusan : SAINS DAN TEKNOLOGI/SISTEM INFORMASI
E-mail address : riskyagungnurdian@gmail.com

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 :

KOMPARASI METODE SMOTE DAN ADASYN DALAM MENINGKATKAN

PERFORMA KLASIFIKASI HERREGISTRASI MAHASISWA BARU

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, 17 Agustus 2021

Penulis

(RISKY AGUNG NURDIAN)

No.	Judul Penelitian	Hasil	Korelasi
3.	Integrasi <i>SMOTE</i> dan <i>Information Gain</i> pada <i>Naïve Bayes</i> Untuk Prediksi Cacat Software (Catal, 2011).	Hasil pengujian menunjukkan bahwa model NB <i>SMOTE+IG</i> memiliki nilai AUC lebih tinggi dibandingkan dengan model NB sebesar 0,798 sedangkan nilai AUC model NB sebesar 0,753.	Metode <i>SMOTE</i> pada algoritma <i>Naïve Bayes</i> dapat menangani <i>class imbalance</i> .
4.	Pengaruh <i>Oversampling</i> Pada Klasifikasi Hipertensi Dengan Algoritma <i>Naïve Bayes</i> , <i>Decision Tree</i> , dan <i>Artificial Neural Network (ANN)</i> (Chamidah et al., 2020).	Berdasarkan hasil pengujian metode <i>ADASYN</i> dapat digunakan untuk mengatasi ketidakseimbangan data pada <i>dataset</i> hipertensi.	Metode <i>ADASYN</i> dapat digunakan untuk mengatasi <i>class imbalance</i> .
5.	<i>Naïve Bayes Algorithm for Classification of Student Major's Specialization</i> (Syaputri et al., 2020).	Pengujian dilakukan dengan membagi <i>data training</i> dan <i>data testing</i> 70%:30%. Hasil confusion matrix menunjukkan akurasi <i>naïve bayes</i> sebesar 96.19%.	<i>Naïve bayes</i> digunakan untuk klasifikasi.
6.	<i>Predictive Modelling of Student Dropout Using Ensemble</i>	Penelitian ini menggunakan algoritma <i>SMOTE</i> untuk mengatasi <i>class imbalanced</i> . Hasil	Algoritma <i>SMOTE</i> untuk menangani <i>class imbalanced</i> ..

No.	Judul Penelitian	Hasil	Korelasi
	<i>Classifier Method in Higher Education</i> (Hutagaol & Suharjito, 2019).	pengujian menunjukkan bahwa nilai <i>accuracy K-NN</i> sebesar 98.20% dan <i>naïve bayes</i> sebesar 98.24%.	
7.	ALGORITMA BAYESIAN CLASSIFICATION UNTUK MEMPREDIKSI HEREGISTRASI MAHASISWA BARU DI STMIK WIDYA PRATAMA (Sugianti, 2012).	Hasil penelitian menunjukkan bahwa tingkat keakuratan untuk melakukan klasifikasi dari algoritma <i>bayesian</i> adalah 78%.	Algoritma <i>naïve bayes</i> digunakan untuk melakukan klasifikasi daftar ulang calon mahasiswa baru.
8.	Memprediksi Daftar Ulang Mahasiswa Baru Menggunakan Algoritma Bayesian Classification di Universitas XYZ (Saputro & Jumasa, 2018).	Hasil dari penelitian ini memaparkan bahwa algoritma <i>naïve bayes</i> dapat digunakan untuk melakukan prediksi mengenai data calon mahasiswa yang akan melakukan daftar ulang dan data mahasiswa yang tidak melakukan daftar ulang.	Algoritma <i>naïve bayes</i> digunakan untuk melakukan klasifikasi mengenai daftar ulang calon mahasiswa baru.
9.	Komparasi Model Prediksi Daftar Ulang Calon Mahasiswa Baru	Penelitian ini memaparkan bahwa algoritma <i>decision tree</i> merupakan algoritma yang	Algoritma <i>decision tree</i> cocon digunakan untuk melakukan

No.	Judul Penelitian	Hasil	Korelasi
	Menggunakan Metode Decision Tree dan Adaboost (Rabbani et al., 2021).	memiliki performa lebih baik dibandingkan dengan algoritma lain.	klasifikasi calon mahasiswa baru.
10.	KOMPARASI KINERJA ALGORITMA C.45 DAN NAÏVE BAYES UNTUK PREDIKSI KEGIATAN PENERIMAAN MAHASISWA BARU (STUDI KASUS: UNIVERSITAS STIKUBANK SEMARANG) (Yahya & Jananto, 2019).	Algoritma <i>decision tree</i> memiliki akurasi sebesar 88.74% dan algoritma <i>naïve bayes</i> memiliki akurasi sebesar 87.24%.	Algoritma <i>decision tree</i> dan algoritma <i>naïve bayes</i> dapat digunakan untuk melakukan prediksi calon mahasiswa baru.
11.	Integrasi Metode <i>Decision Tree</i> dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas (Franseda et al., 2020).	Dari semua skenario pengujian yang dilakukan memaparkan bahwa SMOTE+ <i>decision tree</i> memiliki performa klasifikasi yang lebih baik dibandingkan algoritma	SMOTE dapat menangani <i>class imbalanced</i> dan dapat meningkatkan performa klasifikasi.

Tabel 3.1 Struktur Data

Atribut	Jenis Variabel	Contoh
Nama Prodi	Nominal	“Ekonomi Syariah”, “Hukum Ekonomi Syariah”
Skor Peserta	Numerik	“550,612”, “396,964”
Akreditasi	Ordinal	“A”, “B”
Pekerjaan Ayah	Nominal	“Petani”, “Karyawan”
Pekerjaan Ibu	Nominal	“Petani”, “Tidak Ada”
Penghasilan Ayah	Interval	“1.000.000 s/d 2.000.000”
Penghasilan Ibu	Interval	“1.000.000 s/d 2.000.000”
Jenis Sekolah	Nominal	“SMA”, “MA”
Kepemilikan	Nominal	“Negeri”, “Swasta”
Kabupaten	Nominal	“KABUPATEN BOJONEGORO, KABUPATEN LAMONGAN”
Provinsi	Nominal	“JAWA TIMUR”, “RIAU”
Rumpun Prodi	Nominal	“Bahasa”, “Agama”
Daftar Ulang	Biner	“1”, “0”

3.1.4 Pengolahan Data

Setelah data yang digunakan diperoleh, langkah selanjutnya adalah pengolahan data. Penelitian ini menggunakan 3 tahap pengolahan data yaitu *pre-processing*, klasifikasi, dan *evaluating*.

1. *Pre-processing*

Pada tahap ini *atribut* yang memiliki korelasi lemah dengan *class* atau *label* berdasarkan dari nilai korelasi yang dihitung dengan menggunakan *pearson correlation* akan dibuang atau tidak digunakan. Selanjutnya akan dilakukan pemilihan fitur dan *class* (label) hal ini dilakukan untuk mempersingkat dalam

Tabel 4.4 Statistik Herregistrasi Berdasarkan Pekerjaan Ayah

No.	Pekerjaan Ayah	2019			2020			2021		
		1	0	%	1	0	%	1	0	%
1.	Karyawan Swasta	70	38	45,7	139	79	36,2	110	62	36
2.	Buruh	124	103	45,3	21	36	63,1	22	18	45
3.	Pedagang	32	26	44,8	43	23	34,8	28	23	45,09
4.	Peternak	15	12	44,4				3	0	0
5.	Nelayan	89	54	37,7	7	4	57,1			
6.	Petani	116	64	35,5	49	39	44,3	60	26	30,2
7.	Tidak Ada	182	100	35,4	42	26	38,2	19	15	44,1
8.	Wirausaha	276	136	33,1				24	14	36,8
9.	TNI/POLRI	148	72	32,7	8	3	27,2	46	37	44,5
10.	Pensiunan	60	22	26,8				15	10	33,3
11.	Wiraswasta	50	10	16,6				133	78	36,9
12.	Guru/Dosen Non PNS				11	2	15,38	3	7	70
13.	Lainnya				52	57	52,2	42	30	41,6
14.	Pengusaha				59	44	40,3			
15.	PNS Guru/Dosen				23	16	41,02			
16.	PNS Non Guru				30	21	41,7			
17.	Sudah Meninggal							39	23	37,09
18.	Tim Ahli/Konsultan							0	2	100

Ket :

1 = Daftar ulang

0 = Tidak daftar ulang

% = Persentase tidak daftar ulang

Tabel 4.5 merupakan statistik dari pekerjaan ibu calon mahasiswa UINSA yang melakukan daftar ulang dan tidak melakukan daftar ulang. Pada tahun 2019 calon mahasiswa UINSA yang tidak melakukan daftar ulang berdasarkan pekerjaan ibu, pekerjaan Pedagang memiliki presentase paling tinggi sebesar 51,3% dan pekerjaan Petani memiliki presentase paling rendah sebesar 22,3%.

Pada tahun 2020 calon mahasiswa UINSA yang tidak melakukan daftar ulang berdasarkan pekerjaan ibu, pekerjaan buruh memiliki persentase paling tinggi yaitu

Atribut	Perlakuan	Ekspektasi Hasil
Pekerjaan Ayah	Menghapus data yang tidak sesuai dan tidak valid	Tidak Ada, Petani, Pedagang,.....
Pekerjaan Ibu	Menghapus data yang tidak sesuai dan tidak valid	Tidak Ada, Petani, Pedagang,.....
Penghasilan Ayah	Menghapus data yang tidak sesuai dan tidak valid	Tidak Ada, <1.000.000,.....
Penghasilan Ibu	Menghapus data yang tidak sesuai dan tidak valid	Tidak Ada, <1.000.000,.....
Jenis Sekolah	Menghapus data yang tidak sesuai dan tidak valid	SMA, MA, MAS,....
Kepemilikan Sekolah	Menghapus data yang tidak sesuai dan tidak valid	Negeri, Swasta
Provinsi	Menghapus data yang tidak sesuai dan tidak valid	ACEH, JAWA TIMUR, JAWA BARAT,
Kabupaten	Menghapus data yang tidak sesuai dan tidak valid	KABUPATEN LAMONGAN, KABUPATEN BOJONEGORO,
Rumpun Prodi	Menghapus data yang tidak sesuai dan tidak valid	Bahasa, Agama, IPA,.....

Tabel 4.21 Hasil Pelatihan Model Data SPAN PTKIN 2020

Data Daftar Ulang Tahun 2020						
<i>k-fold</i>	<i>Decision Tree</i>	<i>SMOTE Decision Tree</i>	<i>ADASYN Decision Tree</i>	<i>Naïve Bayes</i>	<i>SMOTE Naïve Bayes</i>	<i>ADASYN Naïve Bayes</i>
Accuracy						
3	59%	55%	50%	60%	53%	52%
5	58%	53%	52%	60%	54%	53%
7	60%	54%	53%	60%	54%	53%
Roc_Auc						
3	53%	56%	51%	57%	51%	49%
5	52%	54%	54%	58%	54%	52%
7	54%	55%	56%	58%	54%	52%
Precision						
3	59%	54%	54%	60%	52%	52%
5	60%	58%	53%	60%	51%	53%
7	60%	58%	54%	59%	52%	52%
Recall						
3	95%	62%	65%	95%	90%	90%
5	87%	80%	89%	96%	90%	92%
7	95%	73%	65%	96%	92%	93%

Pada Gambar 4.4 dapat dilihat bahwa data normal memiliki ketidakseimbangan kelas yang signifikan 539 daftar ulang dan 340 tidak daftar ulang. Metode *SMOTE* dapat melakukan *oversampling* daftar ulang sebanyak 539 dan tidak daftar ulang sebanyak 539 yang artinya jumlah daftar ulang dan tidak daftar ulang sama dan metode *ADASYN* dapat melakukan *oversampling* daftar ulang sebanyak 539 dan tidak daftar ulang sebanyak 609.

Tabel 4.22 Hasil Pelatihan Model Data SPAN PTKIN 2021

Data Daftar Ulang Tahun 2021						
<i>k-fold</i>	<i>Decision Tree</i>	<i>SMOTE Decision Tree</i>	<i>ADASYN Decision Tree</i>	<i>Naïve Bayes</i>	<i>SMOTE Naïve Bayes</i>	<i>ADASYN Naïve Bayes</i>
Accuracy						
3	65%	57%	54%	64%	49%	52%
5	64%	51%	53%	64%	50%	52%
7	65%	55%	56%	64%	52%	52%
Roc_Auc						
3	61%	56%	53%	62%	56%	56%
5	61%	50%	51%	61%	57%	57%
7	61%	55%	53%	60%	56%	54%
Precision						
3	65%	55%	43%	64%	53%	50%
5	64%	51%	39%	64%	53%	50%
7	64%	49%	50%	64%	53%	49%
Recall						
3	95%	43%	34%	96%	81%	51%
5	95%	59%	40%	96%	86%	48%
7	95%	59%	43%	96%	90%	51%

Pada Gambar 4.5 dapat dilihat bahwa data normal memiliki ketidakseimbangan kelas yang signifikan 2154 daftar ulang dan 1321 tidak daftar ulang. Metode *SMOTE* dapat melakukan *oversampling* daftar ulang sebanyak 2154 dan tidak daftar ulang sebanyak 2154 yang artinya jumlah daftar ulang dan tidak daftar ulang sama dan metode *ADASYN* dapat melakukan *oversampling* daftar ulang sebanyak 2154 dan tidak daftar ulang sebanyak 2183.

Tabel 4.23 Hasil Pelatihan Model Data SPAN PTKIN 2019-2021

Data Daftar Ulang Tahun 2019-2021						
<i>k-fold</i>	<i>Decision Tree</i>	<i>SMOTE Decision Tree</i>	<i>ADASYN Decision Tree</i>	<i>Naïve Bayes</i>	<i>SMOTE Naïve Bayes</i>	<i>ADASYN Naïve Bayes</i>
Accuracy						
3	65%	53%	51%	64%	54%	53%
5	65%	52%	50%	64%	54%	51%
7	65%	49%	50%	64%	55%	52%
Roc_Auc						
3	57%	55%	52%	60%	58%	55%
5	57%	54%	50%	60%	57%	52%
7	56%	53%	49%	60%	56%	50%
Precision						
3	65%	52%	50%	64%	53%	52%
5	65%	52%	50%	64%	52%	50%
7	65%	48%	49%	64%	52%	51%
Recall						
3	95%	47%	52%	97%	90%	84%
5	95%	47%	53%	97%	90%	83%
7	95%	41%	48%	97%	91%	85%

Pada semua *dataset* berdasarkan statistik deskriptif semua atribut memiliki pengaruh terhadap calon mahasiswa yang tidak daftar ulang. Hal tersebut dapat dilihat masih banyaknya presentase calon mahasiswa yang tidak daftar ulang di tiap atributnya. Namun, setelah dilakukan perhitungan korelasi dengan *pearson correlation* semua atribut memiliki korelasi yang lemah karena nilai dari *pearson correlation* berada di bawah 0,2. Hal tersebut akan berpengaruh pada hasil *accuracy* dan kinerja model klasifikasi. *Dataset* yang digunakan dalam penelitian ini memiliki *class imbalanced* hal tersebut akan berpengaruh pada keakuratan klasifikasi.

- Catal, C. (2011). Software fault prediction: A literature review and current trends. *Expert Systems with Applications*, 38(4), 4626–4636. <https://doi.org/10.1016/j.eswa.2010.10.024>
- Chamidah, N., Santoni, M. M., & Matondang, N. (2020). *Pengaruh Oversampling pada Klasifikasi Hipertensi dengan Algoritma Naïve Bayes, Decision Tree, dan Artificial Neural Network (ANN)*. 4(4), 7.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Dhitama, F. S., & Bachtiar, F. A. (2020). *Penentuan Kelayakan Debitur Menggunakan Metode Decision Tree C4.5 Dan Oversampling Adaptive Synthetic (ADASYN)*. 10.
- Eska, J. (2018). *Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5* [Preprint]. INA-Rxiv. <https://doi.org/10.31227/osf.io/x6svc>
- Fitri, R. K. R., Rahmansyah, A., & Darwin, W. (2017). *Penggunaan Bahasa Pemrograman Python Sebagai Pusat Kendali Pada Robot 10-D*.
- Franseda, A., Kurniawan, W., & Anggraeni, S. (2020). *Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas*. 08(3), 9.
- Hairani, H. (2019, January 21). *Metode Klasifikasi Data Mining dan Teknik Sampling Smote Menangani Class Imbalance untuk Segmentasi Customer pada industri Perbankan*.

- Hartati, S., Wardoyo, R., & Harjoko, A. (2009). *PERBANDINGAN METODE NEAREST NEIGHBOR DAN ALGORITMA C4.5 UNTUK MENGANALISIS KEMUNGKINAN PENGUNDURAN DIRI CALON MAHASISWA DI STMIK AMIKOM YOGYAKARTA*. 10(1), 19.
- Hastuti, K. (2012). *ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI MAHASISWA NON AKTIF*. 9.
- Hendrian, S. (2018). Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan. *Faktor Exacta*, 11(3).
<https://doi.org/10.30998/faktorexacta.v11i3.2777>
- Hutagaol, N., & Suharjito, S. (2019). Predictive Modelling of Student Dropout Using Ensemble Classifier Method in Higher Education. *Advances in Science, Technology and Engineering Systems Journal*, 4(4), 206–211.
<https://doi.org/10.25046/aj040425>
- Kasanah, A. N., Muladi, & Pujiyanto, U. (2021). *Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN | Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*. 5(3).
<https://jurnal.iaii.or.id/index.php/RESTI/article/view/945>
- Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Data Preprocessing for Supervised Learning. *International Journal of Computer Science*, 1, 111–117.
- Mambang, M., & Marleny, F. D. (2015). PREDIKSI CALON MAHASISWA BARU MENGGUNAKAN METODE KLASIFIKASI DECISION TREE.

- Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *BINA INSANI ICT JOURNAL*, 7(2), 156–165.
- Rhomadhona, H., & Permadi, J. (2019). Klasifikasi Berita Kriminal Menggunakan Naïve Bayes Classifier (NBC) dengan Pengujian K-Fold Cross Validation. *Jurnal Sains dan Informatika*, 5(2), 108–117. <https://doi.org/10.34128/jsi.v5i2.177>
- Romadloni, N. T., & Pardede, H. F. (2017). Seleksi Fitur Berbasis Pearson Correlation Untuk Optimasi Opinion Mining Review Pelanggan | *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*. <https://jurnal.iaii.or.id/index.php/RESTI/article/view/1189>
- Saputro, W. T., & Jumasa, H. M. (2018). Memprediksi Daftar Ulang Mahasiswa Baru Menggunakan Algoritma Bayesian Classification Di Universitas XYZ. *INTEK: Jurnal Informatika Dan Teknologi Informasi*, 1(2), 73–82. <https://doi.org/10.37729/intek.v1i2.566>
- Sugianti, D. (2012). *ALGORITMA BAYESIAN CLASSIFICATION UNTUK MEMPREDIKSI HEREGISTRASI MAHASISWA BARU DI STMIK WIDYA PRATAMA*. 5.
- Sugianto, A. (2016). *JENIS-JENIS DATA VARIABEL (VARIABEL DISKRIT DAN VARIABEL KONTINYU)*. 6.
- Sumiran, K. (2018). *An Overview of Data Mining Techniques and Their Application in Industrial Engineering*. 2(2), 8.

- Syaputri, A. W., Irwandi, E., & Mustakim, M. (2020). Naïve Bayes Algorithm for Classification of Student Major's Specialization. *Journal of Intelligent Computing & Health Informatics*, 1(1), 17. <https://doi.org/10.26714/jichi.v1i1.5570>
- UINSA - Seleksi Penerimaan Mahasiswa Baru—Jalur seleksi. (2021). <https://pmb.uinsby.ac.id/jalur-seleksi>
- Wibawa, A. P., Kurniawan, A. C., Murti, D. M. P., Adiperkasa, R. P., Putra, S. M., Kurniawan, S. A., & Nugraha, Y. R. (2019). Naïve Bayes Classifier for Journal Quartile Classification. *International Journal of Recent Contributions from Engineering, Science & IT (IJES)*, 7(2), 91–99.
- Wibawa, A. P., Purnama, M. G. A., Akbar, M. F., & Dwiyanto, F. A. (2018). *Metode-metode Klasifikasi*. 3(1), 5.
- Wicaksana, P. D. (2015). *PERBANDINGAN ALGORITMA K-NEAREST NEIGHBORS DAN NAIVE BAYES UNTUK STUDI DATA “WISCONSIN DIAGNOSIS BREAST CANCER.”* UNIVERSITAS SANATA DHARMA YOGYAKARTA.
- Ws, A. M., & Nooraeni, R. (2020). PENERAPAN METODE RESAMPLING DALAM MENGATASI IMBALANCED DATA PADA DETERMINAN KASUS DIARE PADA BALITA DI INDONESIA (ANALISIS DATA SDKI 2017). *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, 8(1), 19–27. <https://doi.org/10.24252/msa.v8i1.13452>
- Wu, X., Kumar, V., Quinlan, R., Ghosh, J., Yang, Q., Motoda, H., Mclachlan, G., Ng, S. K. A., Liu, B., Yu, P., Zhou, Z.-H., Steinbach, M., Hand, D., &

