

**EVALUASI KINERJA MODEL KLASIFIKASI KEJADIAN
HUJAN KABUPATEN OGAN ILIR DENGAN METODE
REGRESI LOGISTIK BINER DAN *K-NEAREST NEIGHBOR*
SEBELUM DAN SESUDAH PENERAPAN
*SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE***

SKRIPSI

Oleh:

NICHO SAPUTRA KUSNADI

08011282025053



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA
2025**

LEMBAR PENGESAHAN

**EVALUASI KINERJA MODEL KLASIFIKASI
KEJADIAN HUJAN KABUPATEN OGAN ILIR DENGAN METODE
REGRESI LOGISTIK BINER DAN *K-NEAREST NEIGHBOR*
SEBELUM DAN SESUDAH PENERAPAN
*SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE***

SKRIPSI

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Studi Matematika**

Oleh :

NICHO SAPUTRA KUSNADI

NIM. 08011282025053

**Indralaya, 16 September 2025
Pembimbing Utama**

Pembimbing Kedua



**Dr. Endang Sri Kresnawati, S.Si., M.Si.
NIP. 197702082002122003**



**Prof. Yulia Resti, S.Si., M.Si., Ph.D.
NIP. 197307191997022001**

**Mengetahui,
An. Ketua**

Sekretaris Jurusan



**Drs. Alwine Zayanti, S.Si., M.Si.
NIP. 197012041998022001**

HALAMAN PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : Nicho Saputra Kusnadi

NIM : 08011282025053

Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Menyatakan bahwa skripsi ini adalah hasil karya ilmiah saya sendiri dan karya ilmiah ini belum pernah diajukan sebagai pemenuhan persyaratan untuk memperoleh gelar kesarjanaan strata satu (S1) dari Universitas Sriwijaya maupun perguruan tinggi lain. Semua informasi yang dimuat di dalam skripsi ini yang berasal dari penulis lain baik yang dipublikasi atau tidak telah diberikan penghargaan dengan mengutip nama sumber penulis baik yang secara benar. Semua isi dari skripsi ini sepenuhnya menjadi tanggung jawab saya sebagai penulis.

Demikianlah surat pernyataan ini saya buat dengan sebenarnya.

Indralaya, 16 September 2025

Penulis



Nicho Saputra Kusnadi

NIM. 08011282025053

HALAMAN PERSEMBAHAN

“Stay hungry, stay foolish”

-Steve Jobs-

“The more I learn, the more I realize how much I don’t know”

-Albert Einstein-

“Man is born free, and everywhere he is in chains”

-Jean Jacques Rousseau-

Skripsi ini kupersembahkan kepada :

- 1. Allah Subhanahu wa ta’ala**
- 2. Rasulullah Muhammad SAW**
- 3. Kedua orangtuaku**
- 4. Saudara-saudaraku**
- 5. Keluarga besarku**
- 6. Semua Guru dan Dosenku**
- 7. Teman-temanku**
- 8. Almamaterku**

KATA PENGANTAR

Assalamu’alaikum Warahmatullahi Wabarakatuh.

Puji syukur saya panjatkan ke hadirat Allah Subhanahu Wa Ta’ala atas nikmat dan karunianya, sehingga saya dapat menyelesaikan skripsi yang berjudul **“Evaluasi Kinerja Model Klasifikasi Kejadian Hujan Kabupaten Ogan Ilir Menggunakan Metode Regresi Logistik Biner dan *K-Nearest Neighbor* Sebelum dan Sesudah Penerapan *Synthetic Minority Over-Sampling Technique*”** sebagai salah satu syarat memperoleh gelar Sarjana Sains pada Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Dengan penuh hormat dan kerendahan hati, penulis mengucapkan terimakasih yang mendalam kepada kedua orang tua penulis, Ibu **Rika Yunani** dan Bapak **Iwan Kusnadi** atas didikan, nasihat, perhatian, dorongan semangat, serta doa demi kebaikan penulis. Penulisan skripsi ini juga tidak lepas dari dukungan berbagai pihak, sehingga penulis ingin mengucapkan terimakasih kepada:

1. Bapak **Prof. Hermansyah, S.Si., M.Si., Ph.D.** selaku Dekan Fakultas MIPA Universitas Sriwijaya.
2. Ibu **Dr. Dian Cahyawati Sukanda, S.Si., M.Si.** selaku Ketua Jurusan Matematika.
3. Ibu Des **Alwine Zayanti, S.Si., M.Si.** selaku Sekretaris Jurusan Matematika.
4. Ibu **Prof. Yulia Resti, S.Si., M.Si., Ph.D.** selaku Pembimbing Utama atas bimbingan dan arahannya..
5. Ibu **Dr. Endang Sri Kresnawati, S.Si., M.Si.** selaku Pembimbing Kedua atas arahan dan bantuannya..
6. Bapak **Dr. Bambang Suprihatin, S.Si., M.Si.** selaku Pembahas Pertama sekaligus Ketua Pelaksana, atas saran dan masukan yang diberikan.
7. Ibu **Novi Rustiana Dewi, S.Si., M.Si.** selaku Pembahas Kedua atas masukan yang berharga.
8. Ibu **Dr. Indrawati, S.Si., M.Si.** selaku Dosen Pembimbing Akademik.

9. Seluruh dosen dan staf di Jurusan Matematika yang telah berbagi ilmu dan membantu selama studi.
10. Bapak **Irwansyah** dan Ibu **Hamidah** yang telah membantu penulis dalam hal administrasi di Jurusan Matematika FMIPA UNSRI.
11. Teman seperjuangan selama masa perkuliahan yaitu **Samuel Hariadi Lumban Gaol, Stefanie Fortunita Candra, Chinesia Febriana Wilim, Abdan Sakuro, Tarishah Azzahra, Tias Ahmadi Wijaya, Dido Fivalianda, Stephany Puspa Haryanti, Fitria Romadhona, Gilang Hanifah Al'kautsar,** dan **Ilham Fazar** serta teman-teman angkatan 2020 yang telah memberikan motivasi dan dukungannya.
12. Keluarga besar yang selalu mendoakan dan mendukung saya.
13. Semua pihak yang telah membantu secara langsung maupun tidak langsung hingga skripsi ini selesai.

Saya menyadari sepenuhnya bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu saya terbuka terhadap kritik dan saran yang membangun. Semoga skripsi ini dapat bermanfaat bagi pembaca terutama di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Indralaya, Juli 2025

Penulis

**EVALUATION OF THE PERFORMANCE OF RAINFALL EVENT
CLASSIFICATION IN OGAN ILIR REGENCY WITH
BINARY LOGISTIC REGRESSION AND K-NEAREST NEIGHBOR
BEFORE AND AFTER THE APPLICATION OF
*SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE***

By

NICHO SAPUTRA KUSNADI

08011282025053

ABSTRACT

Rainfall is a key component of the water cycle and plays an important role in the ecosystem, especially in tropical regions such as Indralaya Regency. Rain affects various operational aspects in the field, particularly in agriculture and plantations. Binary Logistic Regression and K-Nearest Neighbor (KNN) can be used to classify rainfall events. One common issue in rainfall event classification is class imbalance. To address this problem, a class-balancing technique called Synthetic Minority Over-Sampling Technique (SMOTE) is used, which helps balance the data by generating synthetic samples for the minority class. This study aims to obtain accuracy, precision, recall, and f-score values for Binary Logistic Regression and KNN both before and after applying SMOTE, and to compare model performance to identify the best model. The data used are secondary rainfall event data for Ogan Ilir Regency from 2018 to 2023, obtained from the visualcrossing.com website, with 14 independent variables and 1 dependent variable. The results show that Binary Logistic Regression after SMOTE provides the best performance with an accuracy of 83.56%, precision of 87.97%, recall of 81.44%, and f-score of 84.58%. Other results are Binary Logistic Regression before SMOTE (accuracy 79.18%, precision 74.14%, recall 95.79%, f-score 83.58%), KNN before SMOTE (accuracy 78.77%, precision 77.73%, recall 86.39%, f-score 81.83%), and KNN after SMOTE (accuracy 74.38%, precision 80.73%, recall 70.54%, f-score 75.29%).

Keywords: Rainfall, SMOTE, Binary Logistic Regression, KNN

**EVALUASI KINERJA MODEL KLASIFIKASI KEJADIAN HUJAN
KABUPATEN OGAN ILIR DENGAN METODE
REGRESI LOGISTIK BINER DAN *K-NEAREST NEIGHBOR*
SEBELUM DAN SESUDAH PENERAPAN
*SYNTHETHIC MINORITY OVER-SAMPLING TECHNIQUE***

Oleh

NICHO SAPUTRA KUSNADI

08011282025053

ABSTRAK

Hujan merupakan komponen utama dalam siklus air dan memiliki peran penting dalam ekosistem, terutama di wilayah beriklim tropis seperti kabupaten Indralaya. Hujan memengaruhi berbagai aspek operasional di lapangan, khususnya bidang pertanian dan perkebunan. Regresi Logistik Biner dan K-Nearest Neighbor (KNN) dapat digunakan untuk mengklasifikasikan kejadian hujan. Salah satu masalah yang sering muncul adalah ketidakseimbangan kelas klasifikasi. Untuk mengatasinya digunakan teknik penyeimbang kelas yaitu Synthetic Minority Over-Sampling Technique (SMOTE), yang membantu menyeimbangkan data dengan membuat data sintesis pada kelas minoritas. Penelitian ini bertujuan memperoleh nilai accuracy, precision, recall, dan fscore pada Regresi Logistik Biner dan KNN baik sebelum maupun sesudah SMOTE serta membandingkan kinerja model untuk mendapatkan model terbaik. Data yang digunakan merupakan data sekunder kejadian hujan kabupaten Ogan Ilir tahun 2018–2023 yang diperoleh dari laman [visualcrossing.com](https://www.visualcrossing.com) dengan 14 variabel bebas dan 1 variabel terikat. Hasil penelitian menunjukkan Regresi Logistik Biner sesudah SMOTE memberikan kinerja terbaik dengan accuracy 83,56%, precision 87,97%, recall 81,44%, dan fscore 84,58%. Adapun hasil lainnya adalah Regresi Logistik Biner sebelum SMOTE (accuracy 79,18%, precision 74,14%, recall 95,79%, fscore 83,58%), KNN sebelum SMOTE (accuracy 78,77%, precision 77,73%, recall 86,39%, fscore 81,83%), dan KNN sesudah SMOTE (accuracy 74,38%, precision 80,73%, recall 70,54%, fscore 75,29%).

Kata Kunci: Hujan, SMOTE, Regresi Logistik Biner, KNN

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN KARYA ILMIAH	ii
HALAMAN PERSEMBAHAN	iii
KATA PENGANTAR	iv
ABSTRACT	vi
ABSTRAK	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	6
2.1 Hujan.....	6
2.2 Klasifikasi	6
2.3 <i>Statistical Machine Learning</i>	7
2.4 <i>Min-Max Normalization</i>	7
2.5 SMOTE	8
2.6 Regresi Logistik Biner	8
2.7 <i>Odds Ratio</i>	12
2.8 <i>K-Nearest Neighbors (KNN)</i>	13
2.9 <i>Confusion Matrix</i>	14
BAB III METODOLOGI PENELITIAN	16
3.1 Tempat.....	16
3.2 Waktu	16
3.3 Metode Penelitian	16
BAB IV HASIL DAN PEMBAHASAN.....	19
4.1 Deskripsi Data.....	19

4.2	Normalisasi Data.....	20
4.3	Partisi Data.....	21
4.4	SMOTE pada Data Latih	22
4.5	Metode Regresi Logistik Biner.....	22
4.5.1	Uji Serentak	22
4.5.2	Uji Parsial.....	26
4.5.3	Evaluasi Model Menggunakan AIC.....	33
4.5.4	Odds Ratio	34
4.5.5	Model Terbaik	38
4.5.6	<i>Confusion Matrix</i>	38
4.6	Metode <i>K-Nearest Neighbor</i>	40
4.6.1	Contoh Perhitungan Manual Metode <i>K-Nearest Neighbor</i>	40
4.6.2	<i>Confusion Matrix</i>	42
4.7	Analisis Hasil	44
BAB V KESIMPULAN DAN SARAN		47
5.1	Kesimpulan	47
5.2	Saran	48
DAFTAR PUSTAKA		49

DAFTAR TABEL

Tabel 2.1 <i>Confusion matrix</i>	14
Tabel 3.1 Variabel Data Kejadian Hujan.....	16
Tabel 4.1 Deskripsi variabel.....	19
Tabel 4.2 Data contoh untuk normalisasi data	20
Tabel 4.3 Tabel normalisasi data contoh	21
Tabel 4.4 Data latih	21
Tabel 4.5 Data uji	22
Tabel 4.6 Data latih setelah SMOTE	22
Tabel 4.7 Tabel estimasi model regresi logistik biner sebelum SMOTE	23
Tabel 4.8 Hasil uji serentak sebelum SMOTE	24
Tabel 4.9 Tabel estimasi model regresi logistik biner sesudah SMOTE	25
Tabel 4.10 Hasil uji serentak sesudah SMOTE.....	26
Tabel 4.11 Hasil uji parsial sebelum SMOTE.....	27
Tabel 4.12 Hasil uji parsial sesudah SMOTE	30
Tabel 4.13 <i>Confusion matrix</i> metode regresi logistik biner sebelum SMOTE	38
Tabel 4.14 <i>Confusion matrix</i> metode regresi logistik biner sesudah SMOTE.....	38
Tabel 4.15 Tabel data contoh untuk metode KNN	40
Tabel 4.16 Klasifikasi data contoh ke-10.....	41
Tabel 4.17 <i>Confusion matrix</i> metode KNN sebelum SMOTE.....	42
Tabel 4.18 <i>Confusion matrix</i> metode KNN sesudah SMOTE	43
Tabel 4.19 Perbandingan tingkat akurasi	44

BAB I

PENDAHULUAN

1.1 Latar Belakang

Hujan merupakan salah satu komponen utama dalam siklus air dan memiliki peran penting dalam ekosistem, terutama di wilayah yang beriklim tropis. Beberapa faktor yang mempengaruhi terjadinya hujan meliputi suhu, tekanan udara, kelembapan, lama penyinaran matahari, arah angin, jenis awan, dan jumlah awan yang terbentuk (Sunarmi et al., 2022).

Proses terjadinya hujan dimulai ketika air di permukaan bumi menguap menjadi uap air karena panas matahari. Uap air yang naik ke atmosfer akan mendingin dan membentuk awan. Pada tahap selanjutnya, terjadi kondensasi, yang menyebabkan uap air dalam awan berubah menjadi tetesan air. Ketika tetesan air mencapai ukuran yang cukup besar, maka butiran air tersebut jatuh ke bumi sebagai hujan (Nirwani et al., 2021).

Kabupaten Ogan Ilir, salah satu kabupaten di Provinsi Sumatera Selatan, memiliki luas wilayah 2.666,07 km² dengan jumlah penduduk sebanyak 431.558 jiwa pada tahun 2022. Sebagai wilayah beriklim tropis, Ogan Ilir mengalami kejadian hujan yang cukup signifikan sehingga dapat mempengaruhi sektor penting, khususnya operasional pertanian dan perkebunan yang bergantung pada kejadian hujan (Ruminta et al., 2018).

Untuk memahami dan mengklasifikasi kejadian hujan di wilayah ini, berbagai metode klasifikasi data yang dapat digunakan. Pada penelitian ini metode yang digunakan dalam klasifikasi adalah *K-Nearest Neighbors* (KNN) dan Regresi

Logistik Biner. KNN merupakan metode klasifikasi yang menilai objek berdasarkan kedekatan jarak dengan data yang diteliti (Manullang & Sianturi, 2021). Sedangkan metode Regresi Logistik Biner merupakan metode yang mengidentifikasi hubungan antara variabel bebas dan variabel terikat, dengan variabel terikat yang bersifat kategorik dan memiliki 2 kelas. Namun, salah satu masalah yang sering muncul dalam klasifikasi data cuaca seperti hujan adalah ketidakseimbangan data (Annisa & Rahkmawati, 2023). Ketika data kelas hujan atau tidak hujan tidak seimbang, hasil klasifikasi bisa menjadi bias terhadap kelas mayoritas (Cahyani *et al.*, 2024).

Penelitian yang dilakukan oleh Rahmawati *et al.* (2024) tentang klasifikasi curah hujan di Kota Bogor menunjukkan bahwa metode KNN memiliki tingkat akurasi sebesar 94,82%, yang lebih tinggi dibandingkan dengan Regresi Logistik Biner sebesar 92,75%. Namun, penelitian tersebut tidak menggunakan teknik penyeimbangan kelas.

Penelitian yang juga dilakukan (Sakkeer *et al.*, 2023), mengenai klasifikasi curah hujan negara India dengan menggunakan metode Regresi Logistik Biner, *Decision Tree*, KNN, dan *Random Forest*. Metode *Random Forest* yang dilakukan oleh peneliti tingkat akurasi yang paling tinggi yaitu 88.21%, KNN tingkat akurasi hanya mencapai 87.36%, Regresi Logistik Biner tingkat akurasinya 84.63%, dan *Decision Tree* memiliki tingkat akurasi yang paling rendah yaitu 73.67%. Namun penelitian ini belum menyeimbangkan data yang ada. Namun penelitian tersebut juga belum menerapkan teknik penyeimbangan kelas.

Martha et al. (2022) melakukan penelitian tentang kelayakan pemberian kredit menggunakan metode KNN, Regresi Logistik Biner dan *Decision Tree*. Metode *Decision Tree* yang mempunyai tingkat akurasi yang paling besar yaitu 87.68%, Regresi Logistik Biner tingkat akurasinya sebesar 83.06%, dan KNN tingkat akurasinya sebesar 82.62%. Namun pada penelitian ini tidak membahas kasus hujan dan juga tidak menggunakan teknik penyeimbang kelas.

Selain itu, ada kemungkinan kalau distribusi data tidak seimbang yang menyebabkan tingkat akurasi tidak tepat, sehingga teknik *Synthetic Minority Over-sampling Technique* (SMOTE) dapat digunakan (Cahyani et al., 2024). SMOTE membantu menyeimbangkan data dengan cara melakukan *oversampling* pada kelas minoritas, yakni kelas yang jumlah datanya lebih sedikit. Teknik ini menciptakan data sintetik baru berdasarkan data minoritas yang ada lalu menambahkan data minoritas yang kurang, sehingga distribusi data menjadi lebih seimbang (Yi et al., 2022).

Penelitian yang dilakukan Nasrullah et al. (2024), mengenai klasifikasi curah hujan berdasarkan iklim indonesia yang sudah diterapkan *resampling* SMOTE. Metode *Decision Tree* C4.5 yang dilakukan oleh peneliti tingkat akurasinya mencapai 81.42% saat data kasus tersebut diterapkan SMOTE, sedangkan metode KNN tingkat akurasi hanya mencapai 78.10% jika diterapkan SMOTE. Namun penelitian tersebut hanya menampilkan tingkat akurasi yang sudah diterapkan SMOTE, tanpa membandingkan dengan model sebelum diterapkan SMOTE.

Siringoringo (2018) melakukan penelitian pada *Dataset Credit Card Fraud* menunjukkan bahwa dengan melakukan metode KNN dan *resampling* SMOTE

meningkatkan nilai rata-rata *Precision* yang sebelumnya 60.9% menjadi 85.1%, dan *Fscore* yang sebelumnya 38.7% menjadi 81.8%, yang membuktikan bahwa SMOTE dapat meningkatkan kinerja model. Namun, pada penelitian tersebut hanya menggunakan metode KNN tanpa menggunakan metode Regresi Logistik Biner.

Selain itu, berdasarkan penelitian sebelumnya yang membahas metode Regresi Logistik Biner dan KNN, tidak ada yang membahas mengenai kejadian hujan dengan menggunakan metode Regresi Logistik Biner dan KNN dengan membandingkan sebelum dan sesudah diterapkan SMOTE di wilayah Kabupaten Ogan Ilir. Untuk itu, penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja model metode Regresi Logistik Biner dan KNN dalam mengklasifikasikan kejadian hujan di Kabupaten Ogan Ilir, baik sebelum maupun sesudah penerapan SMOTE.

1.2 Rumusan Masalah

Berikut merupakan rumusan masalah pada penelitian ini yaitu:

1. Bagaimana evaluasi kinerja model klasifikasi kejadian hujan dengan menggunakan metode Regresi Logistik Biner tanpa SMOTE, Regresi Logistik Biner dengan SMOTE, KNN tanpa SMOTE, dan KNN dengan SMOTE berdasarkan nilai *accuracy*, *precision*, *recall*, dan *fscore* ?
2. Bagaimana perbandingan nilai *accuracy*, *precision*, *recall*, dan *fscore* antara metode Regresi Logistik Biner tanpa SMOTE, Regresi Logistik Biner dengan SMOTE, KNN tanpa SMOTE, dan KNN dengan SMOTE ?

1.3 Batasan Masalah

Berikut merupakan batasan masalah pada penelitian ini yaitu:

1. Penelitian ini menggunakan data tentang hujan Kabupaten Ogan Ilir tahun 2018 sampai tahun 2023 yang diambil dari laman *visualcrossing.com*, dan variabel yang dipakai sebanyak 15 variabel dari 26 variabel yang ada pada data.
2. Kinerja model dibatasi oleh nilai *accuracy*, *precision*, *recall*, dan *f-score*.

1.4 Tujuan Penelitian

Berikut merupakan tujuan pada penelitian ini yaitu:

1. Mengevaluasi kinerja model klasifikasi menggunakan metode Regresi Logistik Biner tanpa SMOTE, Regresi Logistik Biner dengan SMOTE, KNN tanpa SMOTE, dan KNN dengan SMOTE.
2. Membandingkan hasil klasifikasi metode Regresi Logistik Biner tanpa SMOTE, Regresi Logistik Biner dengan SMOTE, KNN tanpa SMOTE, dan KNN dengan SMOTE berdasarkan nilai *accuracy*, *precision*, *recall*, dan *f-score*.

1.5 Manfaat Penelitian

Berikut merupakan manfaat pada penelitian ini yaitu:

1. Sebagai referensi bagi peneliti lain yang membahas tentang klasifikasi hujan.
2. Sebagai media pembelajaran terkait klasifikasi dengan metode Regresi Logistik Biner tanpa SMOTE, Regresi Logistik Biner dengan SMOTE, KNN tanpa SMOTE, dan KNN dengan SMOTE.

DAFTAR PUSTAKA

- Annisa, S., & Rahkmawati, Y. (2023). Resampling techniques in rainfall classification of Banjarbaru using decision tree method. *TIERS Information Technology Journal*, 4(2), 122–128. <https://doi.org/https://doi.org/10.38043/tiers.v4i2.5069>
- Ariyanto, D., Sofro, A., Hanifah, A. N., Prihanto, J. B., Maulana, D. A., & Romadhonia, R. W. (2024). Logistic and probit regression modeling to predict the opportunities of diabetes in prospective athletes. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 18(3), 1391–1402. <https://doi.org/10.30598/barekengvol18iss3pp1391-1402>
- Azhari, M. F., & Fitriani, F. A. (2022). Coronary heart disease risk prediction using binary logistic regression based on principal component analysis. *Enthusiastic International Journal of Statistics and Data Science*, 2(1). <https://journal.uui.ac.id/ENTHUSIASTIC>
- Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. In *Critical Care* (Vol. 9, Issue 1, pp. 112–118). <https://doi.org/https://doi.org/10.1186/cc3045>
- Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, 14, 1–16. <https://doi.org/https://doi.org/10.1186/1471-2105-14-106>
- Cahyani, N., Putri, W. A., & Irsyada, R. (2024). Improving multiclass rainfall prediction with multilayer perceptron and SMOTE: addressing class imbalance challenges. *Brilliance: Research of Artificial Intelligence*, 4(2), 901–908. <https://doi.org/https://doi.org/10.47709/brilliance.v4i2.5203>
- 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>
- Kasanah, A. N., Muladi, M., & Pujianto, U. (2019). Penerapan teknik SMOTE untuk mengatasi imbalance class dalam klasifikasi objektivitas berita online menggunakan algoritma KNN. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2), 196–201. <https://doi.org/10.29207/resti.v3i2.945>
- Malefaki, S., Markatos, D., Filippatos, A., & Pantelakis, S. (2025). A Comparative Analysis of Multi-Criteria Decision-Making Methods and Normalization Techniques in Holistic Sustainability Assessment for Engineering

- Applications. *Aerospace*, 12(2).
<https://doi.org/https://doi.org/10.3390/aerospace12020100>
- Manullang, R. A., & Sianturi, F. A. (2021). Penerapan algoritma K-Nearest Neighbour untuk memprediksi kelulusan mahasiswa. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, 4(2), 42–50.
<https://ejournal.sisfokomtek.org/index.php/jikom/article/view/155/146>
- Martha, S., Andani, W., & Rizki, S. W. (2022). Perbandingan metode K-Nearest Neighbor, regresi logistik biner, dan pohon klasifikasi pada analisis kelayakan pemberian kredit. *Euler : Jurnal Ilmiah Matematika, Sains Dan Teknologi*, 10(2), 262–273. <https://doi.org/https://doi.org/10.34312/euler.v10i2.16751>
- Nasrullah, M. F., Saedudin, Rd. R., & Hamami, F. (2024). Perbandingan akurasi algoritma C4.5 dan K-Nearest Neighbors untuk klasifikasi curah hujan berdasarkan iklim indonesia. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(2), 628–638.
<https://doi.org/https://doi.org/10.29100/jupi.v9i2.4655>
- Nawawi, M. I. (2021). Hubungan jenis kelamin dan gejala terinfeksi COVID-19. *Jurnal Kesehatan Tadulako*, 7(3), 118–127. <https://jurnal.fk.untad.ac.id/index.php/htj/article/view/331>
- Nirwani, L., Amamuhti, B. R., Adzkiya, M. A. Z., & Radhia Pradana. (2021). Komparasi metode amonium fosfomolibdat (AMP) dan metoda kalium ferosianida ($K_4Fe(CN)_6$) pada analisis ^{137}Cs dalam sampel air hujan. *Annual Nuclear Safety Seminar*, 105–110. <https://inis.iaea.org/records/navpv-p3p17>
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes classifier dan confusion matrix pada analisis sentimen berbasis teks pada twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711.
- Nysa, J. U. U., Mahmudi, A., & Auliasari, K. (2023). Perbandingan jarak Euclidean, Manhattan, Chebyshev pada klasifikasi status gizi balita menggunakan metode K-Nearest Neighbors (KNN). *Jurnal Mahasiswa Teknik Informatika (JATI)*, 7(4), 2443–2450.
- Panoto, A., Utami, Y. R. W., & Laksito, W. Y. (2017). Penerapan algoritma K-Nearest Neighbors untuk prediksi kelulusan mahasiswa pada STMIK Sinar Nusantara Surakarta. *Jurnal Teknologi Informasi Dan Komunikasi Sinar Nusantara (TIKomSiN)*, 5(1), 27–31. <https://p3m.sinus.ac.id/jurnal/index.php/TIKomSiN/article/view/284/256>
- Pratiwi, L. P. S., & Wijaya, I. M. P. P. (2024). Binary logistic regression analysis on the spread of dengue fever in Bali province. In *J Statistika* (Vol. 17, Issue 1). <https://doi.org/https://doi.org/10.36456/jstat.vol17.no1.a8955>

- Putra, F., Tahiyat, H. F., Ihsan, R. M., Rahmadden, & Efrizoni, L. (2024). Penerapan algoritma K-Nearest Neighbor menggunakan wrapper sebagai preprocessing untuk penentuan keterangan berat badan manusia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 273–281. <https://doi.org/10.57152/malcom.v4i1.1085>
- Rahmawati, F., Amanah, F., & Fallo, S. I. (2024). Studi komparasi regresi logistik biner dan K-Nearest Neighbor pada kasus prediksi curah hujan. *Statistika*, 24(1), 20–30. <https://doi.org/https://doi.org/10.29313/statistika.v24i1.2739>
- Ruminta, Handoko, & Nurmala, T. (2018). Indikasi perubahan iklim dan dampaknya terhadap produksi padi di Indonesia (Studi kasus : Sumatera Selatan dan Malang Raya). *Jurnal AGRO*, 5(1), 48–60. <https://doi.org/https://doi.org/10.15575/1607>
- Sakkeer, J., Ragavendiran, S. D. P., Ravindran, F., & Anitha, E. (2023). Rainfall detection using machine learning. *IJSRD-International Journal for Scientific Research & Development*, 11(7), 2321–0613. www.ijssrd.com
- Parsaulian, A. S., Tarno, & Ispriyanti, D. (2021). Analisis faktor-faktor yang mempengaruhi penerima beras raskin menggunakan regresi logistik biner dengan GUI R. *Jurnal Gaussian*, 10(1), 31–43. <https://ejournal3.undip.ac.id/index.php/gaussian/>
- Siringoringo, R. (2018). Klasifikasi data tidak seimbang menggunakan algoritma SMOTE dan K-Nearest Neighbor. *Journal Information System Development (ISD)*, 3(1), 44–49.
- Sunarmi, N., Kumailia, E. N., Nurfaiza, N., Nikmah, A. K., Aisyah, H. N., Sriwahyuni, I., & Lailly, S. N. (2022). Analisis faktor unsur cuaca terhadap perubahan iklim di Kabupaten Pasuruan pada tahun 2021 dengan metode principal component snalysis. *Newton-Maxwell Journal of Physics*, 3(2), 56–64. <https://doi.org/10.33369/nmj.v3i2.23380>
- Taye, M. M. (2023). Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions. In M. A. A. Dewan & A. K. Sahu (Eds.), *Computers* (Vol. 12, Issue 5, p. 1). MDPI. <https://doi.org/https://doi.org/10.3390/computers12050091>
- The Royal Society. (2017). *Machine Learning: The Power and Promise of Computers That Learn by Example*. The Royal Society. <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>
- Van Den Goorbergh, R., Van Smeden, M., Timmerman, D., & Van Calster, B. (2022). The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. *Journal of the American*

Medical Informatics Association (JAMIA), 29(9), 1525–1534.
<https://doi.org/https://doi.org/10.1093/jamia/ocac093>

Yi, X., Xu, Y., Hu, Q., Krishnamoorthy, S., Li, W., & Tang, Z. (2022). ASN-SMOTE: a synthetic minority oversampling method with adaptive qualified synthesizer selection. *Complex and Intelligent Systems*, 8(3), 2247–2272.
<https://doi.org/https://doi.org/10.1007/s40747-021-00638-w>