

**PENERAPAN METODE *GENERALIZED LINEAR MODEL* DENGAN
MISSING VALUE IMPUTATION, DISKRITISASI DATA, DAN
SMOTE UNTUK KLASIFIKASI KEJADIAN HUJAN
(Studi Kasus: Data Cuaca Harian Kota Palembang Tahun 2018-2023)**

DRAFT SKRIPSI

Oleh :

PRANSISKA YOLANDA AGUSTINA

NIM. 08011282025027



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

2024

LEMBAR PENGESAHAN

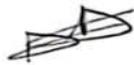
PENERAPAN METODE *GENERALIZED LINEAR MODEL* DENGAN
MISSING VALUE IMPUTATION, DISKRITISASI DATA, DAN SMOTE
UNTUK KLASIFIKASI KEJADIAN HUJAN

DRAFT SKRIPSI

Oleh

PRANSISKA YOLANDA AGUSTINA
NIM. 08011282025027

Pembimbing Kedua



Drs. Ali Amran, M.T
NIP. 196612131994021001

Indralaya, Desember 2024
Pembimbing Utama



Des Alwine Zavanti, S.Si., M.Si
NIP. 197012041998022001

Mengetahui,
Ketua Jurusan Matematika



Dr. Dian Cahyawati S. S.Si., M.Si.
NIP. 197303212000122001

HALAMAN PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan dibawah ini :

Nama Mahasiswa : Pransiska Yolanda Agustina

NIM : 08011282025027

Fakultas / Jurusan : Matematika dan Ilmu Pengetahuan Alam / Matematika

Menyatakan bahwa ini adalah hasil karya 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 dala skripsi ini berasal dari penulis lain baik yang dipublikasikan atau tidak telah diberikan penghargaan dengan mengutip nama sumber penulis secara benar. Semua isi dari skripsi ini sepenuhnya menjadi tanggung jawab saya sebagai penulis.

Demikianlah surat pernyataan ini saya buat dengan sebenarnya.

Indralaya, 21 Januari 2025



Pransiska Yolanda Agustina

NIM. 08011282025027

LEMBAR PERSEMBAHAN

Motto:

“Allah tidak akan membebani seseorang melainkan sesuai dengan kesanggupannya”

(Q.S Al-Baqarah : 286)

“Maka sesungguhnya bersama kesulitan itu ada kemudahan, Sesungguhnya bersama kesulitan itu ada kemudahan”

(Q.S Al-Insyirah : 5-6)

“Semua orang berlari pada tempo yang berbeda-beda, tetap semangat untuk menjadi diri sendiri dan mengejar mimpi, tanpa harus jadi orang lain”

-Freyanashifa Jayawardana-

Skripsi ini saya persembahkan kepada:

- 1. Allah SWT.**
- 2. Alm. Ayah dan Ibu Tercinta**
- 3. Kakak dan Adikku Tersayang**
- 4. Keluarga Besarku Terkasih**
- 5. Semua Dosen dan Guruku**
- 6. Sahabat-sahabatku**
- 7. Almamaterku**

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT yang Maha Pengasih Lagi Maha Penyayang atas berkat rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dan penelitian yang berjudul “**Penerapan Metode *Generalized Linear Model* Dengan *Missing Value Imputation*, Diskritisasi Data, dan SMOT Untuk Klasifikasi Kejadian Hujan**” dengan baik. Skripsi ini merupakan salah satu persyaratan yang harus dipenuhi oleh penulis untuk meraih gelar Sarjana Sains bidang Studi Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Penulis menyadari bahwa selama proses pembuatan skripsi telah banyak memperoleh bantuan baik pengajaran, bimbingan, dan pengarahan dari berbagai pihak baik secara langsung maupun tidak langsung. Oleh karena itu, dengan segala hormat dan kerendahan hati, penulis mengucapkan terimakasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini terkhusus kepada orang tua tercinta, **Alm. Ayah Sudiono** dan **Ibu Winda Satriani** yang tak pernah lelah mendidik, membimbing, menasehati, dan mendukung serta tak henti untuk mendoakan anaknya. Penulis juga mengucapkan terimakasih kepada:

1. Bapak **Prof. Hermansyah, S.Si., M.Si., Ph.D** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.
2. Ibu **Dr. Dian Cahyawati Sukanda, S.Si., M.Si** selaku Ketua Jurusan Matematika FMIPA Universitas Sriwijaya dan Dosen Pembimbing Akademik yang telah menasehati, membimbing, memotivasi dan memberikan arahan selama masa perkuliahan.

3. Ibu **Des Alwine Zayanti, S.Si., M.Si** selaku sekretaris jurusan Matematika dan Dosen Pembimbing Utama yang telah meluangkan waktu, pikiran, dan tenaga untuk membimbing, menasehati, memberi saran serta mengarahkan penulis hingga skripsi ini dapat selesai dengan baik.
4. Bapak **Drs, Ali Amran, M.T** selaku Dosen Pembimbing Kedua yang telah meluangkan waktu, pikiran dan tenaga untuk membimbing, menasehati, memberi saran serta mengarahkan penulis hingga skripsi ini dapat selesai dengan baik.
5. Bapak **Dr. Bambang Suprihatin, S.Si., M.Si** dan Ibu **Novi Rustiana Dewi, S.Si., M.Si** selaku Dosen Pembahas yang telah meluangkan waktu untuk memberikan tanggapan, ilmu, saran, dan kritikan yang bermanfaat dalam perbaikan dan penyelesaian skripsi ini.
6. Ibu **Prof. Yulia Resti, S.Si., M.Si., Ph.D** selaku ketua seminar dan Bapak **Drs. Endro Setyo Cahyono, S.Si., M.Si** selaku Sekretaris Seminar Skripsi yang telah meluangkan waktu untuk memberikan arahan, saran, membimbing penulis, serta mengatur jalannya seminar sehingga dapat berjalan dengan baik.
7. **Bapak dan Ibu Dosen Jurusan Matematika FMIPA Universitas Sriwijaya**, terima kasih atas ilmu, nasehat dan motivasinya serta bimbingan yang telah diberikan untuk penulis selama masa perkuliahan.
8. Bapak **Irwansyah** dan Ibu **Hamidah** selaku Pegawai tata usaha Jurusan Matematika FMIPA yang telah membantu penulis dalam hal administrasi selama perkuliahan.

9. Kakak dan Adik penulis tersayang **Pipit Safitri dan Najjla Yolanda Afifa**, serta semua keluarga besar yang selalu memberikan dukungan kepada penulis.
10. Sahabat dan teman seperjuangan **Khoiriyah, dan Syakira** atas segala dukungan, motivasi, dan semua bantuan yang telah diberikan selama masa perkuliahan hingga selesainya skripsi ini.
11. Semua pihak yang tidak dapat disebutkan satu persatu atas bantuan dan dukungannya sehingga penulis dapat menyelesaikan skripsi ini.

Penulis menyadari bahwa penulisan skripsi ini masih memiliki kekurangan sehingga penulis sangat mengharapkan kritik dan saran yang membangun agar skripsi ini dapat menjadi lebih baik.

Indralaya, Oktober 2024

Penulis

**APPLICATION OF THE GENERALIZED LINEAR MODEL METHOD
WITH MISSING VALUE IMPUTATION, DATA DISCRETIZATION,
AND SMOTE FOR CLASSIFICATION OF RAIN EVENTS**

By:

PRANSISKA YOLANDA AGUSTINA

NIM. 08011282025027

ABSTRACT

Palembang City often faces erratic weather especially during the rainy season, which can cause flooding, disrupt daily activities, and damage infrastructure. So by classifying rainfall events can anticipate the impacts that occur due to rain and by utilizing weather data, the Palembang City government can be more aware of natural threats, protect the local economy, and maintain a sustainable city environment. The event data used for classification has missing data and imbalanced data. To overcome missing data, KNN ($k = 5$) is used and data discretization is done to convert continuous data into discrete data. Furthermore, imbalanced data is overcome with SMOTE. Thus, the accuracy on imbalanced data is 72.603%, precision 68.864% and recall 92.611%. Furthermore, on balanced data, the accuracy is 72.192%, precision 74.224%, and recall 76.601%.

Keywords: KNN, SMOTE, Rain

**PENERAPAN METODE *GENERALIZED LINEAR MODEL* DENGAN
MISSING VALUE IMPUTATION, DISKRITISASI DATA, DAN SMOTE
UNTUK KLASIFIKASI KEJADIAN HUJAN**

Oleh:

PRANSISKA YOLANDA AGUSTINA

NIM. 08011282025027

ABSTRAK

Kota Palembang sering menghadapi cuaca yang tidak menentu terutama saat musim hujan, dimana bisa menyebabkan banjir, mengganggu aktivitas sehari-hari, dan merusak infrastruktur. Sehingga dengan mengklasifikasikan kejadian hujan dapat mengantisipasi dampak-dampak yang terjadi akibat hujan dan dengan memanfaatkan data cuaca, pemerintah Kota Palembang bisa lebih waspada terhadap ancaman alam, melindungi ekonomi lokal, dan menjaga lingkungan kota agar tetap berkelanjutan. Pada data kejadian yang digunakan untuk klasifikasi memiliki *missing data* dan *imbalanced data*. Untuk mengatasi *missing data* digunakan KNN ($k = 5$) dan diskritisasi data dilakukan untuk mengubah data kontinu menjadi data diskrit. Selanjutnya *imbalanced data* diatasi dengan SMOTE. Dengan demikian diperoleh akurasi pada *imbalanced data* sebesar 72,603%, presisi 68,864% dan *recall* 92,611%. Selanjutnya pada *balanced data* diperoleh akurasi 72,192%, presisi 74,224%, dan *recall* 76,601%.

Kata Kunci: KNN, SMOTE, Hujan

DAFTAR ISI

LEMBAR PENGESAHAN	ii
LEMBAR PERSEMBAHAN	iv
KATA PENGANTAR.....	v
ABSTRACT	viii
ABSTRAK	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB 1	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.....	5
BAB II	6
TINJAUAN PUSTAKA	6
2.1 Hujan.....	6
2.2 <i>Generalized Linear Model</i>	6
2.3 Penduga Parameter	8
2.4 Distribusi Binomial	11
2.5 <i>K-Nearest Neighbor Imputation</i>	12
2.6 Diskritisasi Data	13
2.7 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	14
2.8 <i>Confusion Matrix</i>	15
BAB III.....	17
METODE PENELITIAN	17
3.1 Tempat.....	17
3.2 Waktu.....	17

3.3	Variabel Penelitian	17
3.4	Metode Penelitian.....	18
BAB IV		20
HASIL DAN PEMBAHASAN		20
4.1	Deskripsi Data	20
4.2	Imputasi Data Hilang	22
4.3	Diskritisasi Data	25
4.4	Data Training dan Data Testing.....	28
4.5	Menyeimbangkan Kelas Dengan SMOTE.....	30
4.6	Klasifikasi GLM.....	34
4.6.1	Klasifikasi GLM dengan <i>Imbalanced data</i>	34
4.6.2	Klasifikasi GLM dengan <i>Balanced data</i>	37
4.7	Analisis Perbandingan Hasil Klasifikasi <i>GLM Binomial</i>	40
BAB V.....		41
KESIMPULAN DAN SARAN		41
5.1	Kesimpulan.....	41
5.2	Saran	41
DAFTAR PUSTAKA.....		20

DAFTAR TABEL

Tabel 2.1 Confusion matrix.....	15
Tabel 4.1 Deskripsi Variabel	20
Tabel 4.2 Data Cuaca Harian Kota Palembang Tahun 2018 – 2023	21
Tabel 4.3 Variabel yang Memiliki Missing Data.....	22
Tabel 4.4 Observasi yang mengandung missing data.	22
Tabel 4.5 Observasi yang tidak mengandung missing data.	22
Tabel 4.6 Jarak Euclidean antara observasi ke-1139 dengan observasi lengkap ..	24
Tabel 4.7 Jarak Terkecil antara Observasi ke-1139 dengan Observasi Lengkap .	24
Tabel 4.8 Diskritisasi Data Cuaca Harian Kota Palembang Tahun 2018 – 2023 .	26
Tabel 4.9 Pembagian Jumlah Data Training dan Data Testing	29
Tabel 4.10 Data Training.....	29
Tabel 4.11 Data Testing.....	30
Tabel 4.12 Contoh Data Untuk SMOTE	30
Tabel 4.13 Penentuan Nearest Neighbors untuk Proses SMOTE	31
Tabel 4.14 Hasil Data Sintetik Pertama untuk Contoh Data Sampel SMOTE	31
Tabel 4.15 Hasil Data Sintetik Kedua untuk Contoh Sampel SMOTE.....	32
Tabel 4.16 Hasil Data Sintetik Ketiga untuk Contoh Sampel SMOTE	33
Tabel 4.17 Data Sampel Balanced Setelah Menggunakan SMOTE	33
Tabel 4.18 Nilai konstanta dan koefisien variabel bebas	34
Tabel 4.19 Hasil Klasifikasi GLM Binomial dengan Imbalanced Data.....	36
Tabel 4.20 Confusion Matrix Klasifikasi GLM Binomial dengan Imbalanced Data	36
Tabel 4.21 Nilai konstanta dan koefisien variabel bebas balanced data	37
Tabel 4.22 Hasil Klasifikasi GLM Binomial dengan Imbalanced Data.....	39
Tabel 4.23 Confusion Matrix Klasifikasi GLM Binomial dengan Balanced Data	39
Tabel 4.24 Perbandingan Hasil Klasifikasi GLM Binomial	40

DAFTAR GAMBAR

Gambar 4.1 Nilai Akurasi Klasifikasi GLM Binomial Dengan Imputasi KNN ..	23
Gambar 4.2 distribusi Kelas Variabel Y Data Training	29

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kota Palembang adalah sebuah kota metropolitan di Indonesia yang terletak di antara 20°52' hingga 30°5' Lintang Selatan dan 104°37' hingga 104°52' Bujur Timur, dengan ketinggian rata-rata 8 meter di atas permukaan laut. Kota ini terbagi menjadi dua bagian utama, Seberang Ulu dan Seberang Ilir, oleh Sungai Musi.. Cuaca di Kota Palembang hampir sama dengan cuaca di seluruh wilayah Indonesia, yaitu musim hujan dan musim kemarau. (Badan Perencanaan Pembangunan Daerah Penelitian dan Pengembangan, 2019). Kota Palembang sering menghadapi cuaca tidak menentu terutama saat musim hujan, dimana hujan deras bisa menyebabkan banjir, mengganggu aktivitas warga, dan merusak infrastruktur. Dampak hujan di Palembang mempengaruhi kehidupan sehari-hari, seperti sektor ekonomi dan transportasi, dengan kerugian seperti kerusakan jalan dan jembatan serta terhentinya aktivitas ekonomi (Ariska et al., 2022). Oleh karena itu, penting mengklasifikasikan kejadian hujan untuk mengantisipasi dampak-dampak yang terjadi akibat hujan. Informasi pola hujan juga penting untuk mengurangi dampak bencana, memungkinkan respons cepat terhadap banjir. Dengan memanfaatkan data cuaca secara efektif, pemerintah Kota Palembang bisa lebih waspada terhadap ancaman alam, melindungi ekonomi lokal, dan menjaga lingkungan kota tetap berkelanjutan.

Penggunaan data historis menjadi penting, ketika mengklasifikasikan kejadian hujan. Namun, data yang digunakan sering memiliki *missing data* (data tidak

lengkap) dan *imbalanced data* (data tidak seimbang). *Missing data* disebabkan oleh kesalahan pada saat pengumpulan data, kesalahan pada saat *entri* data dan ketidakmampuan responden dalam memberikan jawaban yang akurat (Mukarromah et al., 2015). *Missing data* dapat menyebabkan penurunan keakuratan dan kualitas data pada saat data diolah (Irawan et al., 2017). *Imbalanced data* merupakan kondisi dimana jumlah kejadian suatu kelas lebih banyak dibandingkan dengan jumlah kejadian kelas lainnya. *Imbalanced data* menyebabkan metode klasifikasi lebih mudah mengklasifikasikan kelas mayoritas dibandingkan kelas minoritas (Hairani et al., 2020). Untuk mengatasi permasalahan yang dihadapi dalam penelitian ini dilakukan dengan metode *K-Nearest Neighbors Imputation (KNN-I)* untuk mengatasi data yang hilang. *Synthetic Minority Oversampling Technique (SMOTE)* digunakan untuk mengatasi *imbalanced data*.

Generalized Linier Model (GLM) merupakan salah satu metode yang dapat digunakan dalam mengklasifikasikan kejadian hujan. Ketika distribusi variabel respon adalah distribusi eksponensial, GLM telah banyak digunakan untuk memodelkan berbagai jenis data. Model regresi GLM menggambarkan struktur variabel prediktor, dan fungsi keterkaitan menjelaskan hubungan antara model regresi dan nilai yang diharapkan dari variabel respon. (Santi et al., 2021). Berdasarkan penelitian Rismawati dkk (2022) dengan judul Klasifikasi Data Diagnosis *COVID-19* Menggunakan Metode *Support Vector Machine (SVM)* dan *Generalized Linear Model (GLM)*, menghasilkan persentase akurasi untuk metode GLM sebesar 95,64%. Penelitian lain dilakukan oleh Herman dan Christian (2022) dengan judul Analisis Performa Akademik Mahasiswa Menggunakan *Distributed*

Random Forest, diperoleh hasil akurasi *Distributed Random Forest* sebesar 99,8% sedangkan hasil akurasi GLM sebesar 99,6%.

Penelitian terdahulu dengan judul *Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE dan K-Nearest Neighbor* oleh Siringoringo (2018) menunjukkan bahwa metode SMOTE mampu menangani ketidakseimbangan kelas dataset Credit Card Fraud dengan menghasilkan nilai G-Mean dan F-Measure yang lebih tinggi dibandingkan dengan KNN. Penelitian lain dilakukan oleh Halim dan Azmi (2023) dengan judul *Analisis Perbandingan Klasifikasi dan Penerapan SMOTE Dalam Imbalanced Data pada Credit Card Default*, menunjukkan bahwa setelah penerapan SMOTE metode *Random Forest* dan *Extreme Gradient Boosting* mengalami peningkatan nilai AUC.

Berdasarkan penelitian-penelitian terdahulu, peneliti mencoba untuk mengklasifikasi kejadian hujan pada dataset cuaca di Palembang menggunakan metode *Generalized Linear Model* berdistribusi binomial dan teknik *SMOTE*. Peneliti menggunakan metode GLM karena dapat menangani distribusi respon dari berbagai jenis, termasuk distribusi binomial, yang umum digunakan dalam klasifikasi hujan.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah:

1. Bagaimana ketepatan klasifikasi *Generalized Linear Model* dalam mengklasifikasi kejadian hujan ?
2. Bagaimana ketepatan klasifikasi kejadian hujan menggunakan *Generalized Linear Model* dengan SMOTE ?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini sebagai berikut:

1. Data yang digunakan dalam penelitian ini diambil dari *website VisualCrossing*.
2. Metode *Generalized Linear Model* yang digunakan adalah *Generalized Linear Model binomial*.
3. Evaluasi ketepatan klasifikasi dilakukan dengan menggunakan akurasi, presisi, dan recall.
4. Data yang digunakan sekitar 2191 yang terdiri dari 17 variabel bebas yaitu *Tempmax, Tempmin, Temp, Feelslike, Dew, Humidity, Precip, Windgust, Windspeed, Winddir, Sealevelpressure, Cloudcover, Visibility, Solarradiation, Solarenergy, Severerisk*, dan *Moonphase* serta 1 variabel terikat yaitu *Icon*. Variabel terikat memiliki 2 kelas yaitu hujan dan tidak hujan.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mendapatkan hasil ketepatan klasifikasi kejadian hujan menggunakan metode *Generalized Linear Model*.
2. Mendapatkan hasil ketepatan klasifikasi kejadian hujan menggunakan metode *Generalized Linear Model* dengan SMOTE.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini sebagai berikut :

1. Hasil penelitian ini bisa digunakan oleh Pemerintah Kota Palembang dan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) untuk meningkatkan akurasi prediksi cuaca di Palembang, sehingga dapat memberikan informasi yang lebih tepat dan cepat kepada masyarakat, agar dapat mengurangi dampak yang akan terjadi.
2. Penelitian ini diharapkan dapat menambah pengetahuan tentang penerapan metode *Generalized Linear Model* dalam klasifikasi kejadian hujan, termasuk penggunaan *missing value* untuk menangani nilai yang hilang, diskritisasi data, dan *imbalanced data*.

DAFTAR PUSTAKA

- Agresti, A. (2007). An Introduction to Categorical Data Analysis Second Edition. In *Wiley Series In Probabbility And Statistics* (Second Edition).
- Ariska, M., Akhsan, H., Muslim, M., Romadoni, M., & Putriyani, F. S. (2022). Prediksi Perubahan Iklim Ekstrem di Kota Palembang dan Kaitannya dengan Fenomena El Niño-Southern Oscillation (ENSO) Berbasis Machine Learning. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 6(2), 79–86. <https://doi.org/10.30599/jipfri.v6i2.1611>
- Astuti, F. D., & Lenti, F. N. (2021). Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN. *Jurnal JUPITER*, 13, 89–98.
- Badan Perencanaan Pembangunan Daerah Penelitian dan Pengembangan. (2019). RKPd KOTA PALEMBANG TAHUN 2020. *Badan Perencanaan Pembangunan Daerah Penelitian Dan Pengembangan*.
- Bain L.J & Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics*. Second Edition. California : Duxbury.
- Fadlil, A., Herman, & Praseptian M, D. (2022). K Nearest Neighbor Imputation Performance on Missing Value Data Graduate User Satisfaction. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(4), 570–576. <https://doi.org/10.29207/resti.v6i4.4173>
- Hairani, Saputro, K. E., & Fadli, S. (2020). K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes. *Jurnal Teknologi Dan Sistem Komputer*, 8(2), 89–93. <https://doi.org/10.14710/jtsiskom.8.2.2020.89-93>
- Halim, S. F. N., & Azmi, U. (2023). Analisis Perbandingan Klasifikasi dan Penerapan SMOTE Dalam Imbalanced Data Pada Card Default. *Jurnal Sains Dan Seni*, 12, d127–d134.
- Herman, & Christian, Y. (2022). Analisis Performa Akademik Mahasiswa Menggunakan Distributed Random Forest. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 6, Issue 2). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Hogg, R., & Craig, A. 1995. *Introduction to mathematical (5th ed)*. Prentice Hall Internasional, New Jersey.
- Irawan, N. D., Wijono, W., & Setyawati, O. (2017). Perbaikan Missing value Menggunakan Pendekatan Korelasi Pada Metode K-Nearest Neighbor. *JURNAL INFOTEL*, 9(3). <https://doi.org/10.20895/infotel.v9i3.286>

- Liao, S. G., Lin, Y., Kang, D. D., Chandra, D., Bon, J., Kaminski, N., Scieurba, F. C., & Tseng, G. C. (2014). Missing value imputation in high-dimensional phenomic data: Imputable or not, and how? *BMC Bioinformatics*, *15*(1). <https://doi.org/10.1186/s12859-014-0346-6>
- Mukarromah, Martha, S., & Ilhamsyah. (2015). *PERBANDINGAN IMPUTASI MISSING DATA MENGGUNAKAN METODE MEAN DAN METODE ALGORITMA K-MEANS*.
- Nikita, E. (2014). The use of generalized linear models and generalized estimating equations in bioarchaeological studies. *American Journal of Physical Anthropology*, *153*(3), 473–483. <https://doi.org/10.1002/ajpa.22448>
- Pangestu, I. T. (2020). *ANALISIS KORELASI DATA CURAH HUJAN BMKG DENGAN TRMM (STUDI KASUS STASIUN BMKG DI SUMATRA UTARA)*.
- Perdana, D. A., Zakaria, A., & Sumiharni. (2015). *Studi Pemodelan Curah hujan sintetik dari beberapa stasiun di wilayah Pringsewu* (Vol. 3, Issue 1).
- Putra, D., & Wibowo, A. (2020). Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, *2*, 84–92.
- Rismawati, Y., Made Tirta, I., Setia Dewi, Y., Matematika, J., Mipa, F., & Jember Jl Kalimantan, U. (2022). *KLASIFIKASI DATA DIAGNOSIS COVID-19 MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DAN GENERALIZED LINEAR MODEL (GLM) (Classification of Covid-19 Diagnosis Data Using Support Vector Machine (SVM) and Generalized Linear Model (GLM) Methods)*. <https://magestic.unej.ac.id/>
- Santi, V. M., Wiyono, A., & Sudarwanto. (2021). Pemodelan Jumlah Kasus Malaria di Indonesia Menggunakan Generalized Linear Model. *Jurnal Statistika Dan Aplikasinya*, *5*(1).
- Setyaningsih, A., Gunawan, M. I., Fauzi, L., & Taher, R. A. A. M. (2021). Metode Binomial Mengenai Keberhasilan Pemerintah Dalam Mengatasi Kemacetan Di ibu Kota Jakarta. *Bulletin of Applied Industrial Engineering Theory*, *2*.
- Siregar, S. Y., Toharudin, T., & Tantular, B. (n.d.). *PERFORMA METODE K NEAREST NEIGHBOR IMPUTATION (KNNI) UNTUK MENANGANI MULTIVARIATE MISSING DATA*.
- Siringoringo, R. (2018). *KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN k-NEAREST NEIGHBOR* (Vol. 3, Issue 1).

- Susanto, G. M., Kosasi, S., David, Gat, & Kuway, S. M. (2017). Sistem Referensi Pemilihan Smartphone Android Dengan Metode Fuzzy C-Means dan TOPSIS. *Jurnal Rekayasa Sistem Dan Teknologi Informasi*, 5.
- Wijaya, J., Soleh, A. M., & Rizki, A. (2018). *Penanganan Data Tidak Seimbang pada Pemodelan Rotation Forest Keberhasilan Studi Mahasiswa Program Magister IPB* (Vol. 2, Issue 2).
- WIJAYANTI, N. P. Y. T., N. KENCANA, E., & SUMARJAYA, I. W. (2021). SMOTE: POTENSI DAN KEKURANGANNYA PADA SURVEI. *E-Jurnal Matematika*, 10(4), 235. <https://doi.org/10.24843/mtk.2021.v10.i04.p348>
- Yamin Darsyah, M., & Haryo Ismunarti, D. (2013). *DISTRIBUSI UNIFORM DAN DISTRIBUSI BINOMIAL* (Vol. 1, Issue 1). <http://jurnal.unimus.ac.id>
- Zaki, M. J., & Meira Jr, W. (n.d.). *Data Mining and Analysis: Fundamental Concepts and Algorithms*.