

**PENERAPAN *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE*
(SMOTE) UNTUK MENGATASI DATA *IMBALANCED* DALAM
KLASIFIKASI KEJADIAN HUJAN MENGGUNAKAN
METODE REGRESI LOGISTIK BINER**

SKRIPSI

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

Oleh :

VERTI MONA DESPALIA

08011182025018



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

2025

LEMBAR PENGESAHAN

**PENERAPAN *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE*
(SMOTE) UNTUK MENGATASI DATA *IMBALANCED* DALAM
KLASIFIKASI KEJADIAN HUJAN MENGGUNAKAN
METODE REGRESI LOGISTIK BINER**

SKRIPSI

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

Oleh:

VERTI MONA DESPALIA

NIM. 08011182025018

Pembimbing Kedua



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

**Indralaya 18 Maret 2025
Pembimbing Utama**



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

Mengetahui,

Ketua Jurusan Matematika



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

PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : Verti Mona Despalia

NIM : 08011182025018

Fakultas/Jurusan : Matematika dan Ilmu Pengetahuan Alam/Matematika

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 didalam 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, 18 Maret 2025

Penulis



Verti Mona Despalia
NIM. 08011182025018

HALAMAN PERSEMBAHAN

MOTTO

“Nikmati saja prosesnya, fase jatuh bangunnya, perkara hasil? ya *maybe not today, but someday*”

(Fiersa Besari)

“Tuhan tahu waktu yang tepat, tempat yang tepat dan jawaban yang tepat untuk semua do’a kita”

(Rony Parulian)

Skripsi ini kupersembahkan kepada:

- 1. Allah SWT**
- 2. Orang Tuaku**
- 3. Saudaraku**
- 4. Keluarga Besarku**
- 5. Semua Dosen dan Guruku**
- 6. Sahabatku**
- 7. Almamaterku**

KATA PENGANTAR

Bismillahirrahmanirrahim

Assalamu'alaikum warahmatullahi Wabarakatuh

Segala puji dan syukur kehadirat Allah SWT atas segala kasih sayang, rahmat dan karunia-Nya penulis bisa menyelesaikan skripsi dengan judul **“Penerapan *Synthetic Minority Oversampling Technique* (SMOTE) untuk Mengatasi Data *Imbalanced* dalam Klasifikasi Kejadian Hujan Menggunakan Metode Regresi Logistik Biner”** dengan baik sebagai salah satu syarat memperoleh gelar Sarjana Sains bidang Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (MIPA) Universitas Sriwijaya. Shalawat serta salam semoga senantiasa tercurahkan kepada Nabi Muhammad SAW beserta keluarga, sahabat, dan seluruh pengikutnya hingga akhir zaman.

Penulis menyadari bahwa dalam proses penyusunan skripsi ini masih ada kekurangan, serta banyaknya rintangan dan tantangan dalam mengerjakannya. Dengan penuh rasa hormat dan kerendahan hati, penulis mengucapkan terima kasih yang sebesar-besarnya khusus kepada kedua orang tua tercinta, yaitu Bapak **Balson Ferry** dan Ibu **Santini** untuk seluruh kasih sayang, cinta, pengorbanan, didikan, nasihat, motivasi, perhatian dan do'a yang tidak pernah putus dipanjatkan kepada penulis. terselesaikannya skripsi ini juga tidak terlepas dari bantuan berbagai pihak, baik secara langsung maupun tidak langsung. Untuk itu penulis menyampaikan ucapan terima kasih yang tak terhingga 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 dan Ilmu Pengetahuan Alam Universitas Sriwijaya yang telah memotivasi dan memberikan arahan selama masa perkuliahan.
3. Ibu **Des Alwine Zayanti, S.Si., M.Si.**, selaku Sekretaris Jurusan Matematika sekaligus Dosen Pembimbing Kedua yang telah meluangkan waktu di tengah kesibukannya untuk membimbing serta memberikan saran dalam pengerjaan skripsi ini.
4. Ibu **Prof. Yulia Resti, S.Si., M.Si., PhD.**, selaku Dosen Pembimbing Utama yang telah meluangkan waktu, tenaga dan pikiran untuk bimbingan, memberikan motivasi, nasehat dan saran serta kesabaran dalam memberi bimbingan terbaik kepada penulis selama pengerjaan skripsi ini.
5. Ibu **Dr. Ir. Herlina Hanum, M.Si.**, selaku Dosen Pembimbing Akademik yang telah sangat baik membimbing dan mengarahkan urusan akademik kepada penulis di setiap semester selama perkuliahan.
6. Ibu **Novi Rustiana Dewi, S.Si., M.Si.** dan Ibu **Dr. Endang Sri Kresnawati, S.Si., M.Si.**, selaku Dosen Penguji yang telah meluangkan waktu untuk memberikan kritik dan saran yang bermanfaat dalam pengerjaan skripsi ini.
7. **Semua Dosen di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya**, atas semua ilmu, bimbingan dan motivasi untuk penulis selama masa perkuliahan.
8. Pak **Irwansyah** dan Ibu **Hamidah**, selaku Admin dan Pegawai Tata Usaha di Jurusan Matematika yang telah membantu penulis dalam segala hal administrasi sehingga bisa menyelesaikan skripsi ini.

9. Kakak penulis **Insani Fijri Pratama**, memberikan dukungan dan bantuan baik dalam bentuk material maupun fisik sehingga penulis bisa menyelesaikan skripsi ini.
10. Adik penulis **Syahbani Try Desantra, Versan Syabri Alkon Seprianda** dan **Verni Quinnzha Desthalia**, atas dukungan dan kasih sayang hingga terselesaikannya skripsi ini.
11. **Keluarga Besarku**, atas segala dukungan dan kasih sayang yang telah diberikan kepada penulis.
12. Teman di perkuliahan **Desfarina Fitriani, Vivi Clara Dita, Eggyana Deri Hernanda, Miska Nurulita, Dhiya Qatrunnada**, dan **Muhammad Rizky**, terima kasih atas bantuan, dukungan dan kebersamaan selama kuliah.
13. Sahabat penulis **Inez Vermatasari**, atas waktu yang telah diluangkan untuk mendengarkan cerita dan keluh kesah penulis, memberikan dukungan, kasih sayang, motivasi, dan semua bantuan yang telah diberikan selama ini.
14. Semua pihak yang tidak bisa penulis sebutkan satu persatu yang terlibat dan memberikan bantuan dalam menyelesaikan skripsi ini.

Semoga skripsi ini bisa menambah pengetahuan dan bermanfaat bagi mahasiswa/mahasiswi Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya dan semua pihak yang memerlukan.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Indralaya, 18 Maret 2025

Penulis

**APPLICATION OF SYNTHETIC MINORITY OVERSAMPLING
TECHNIQUE (SMOTE) TO OVERCOME IMBALANCED
DATA IN CLASSIFICATION OF RAIN EVENTS USING
BINARY LOGISTIC REGRESSION METHOD**

By

**VERTI MONA DESPALIA
NIM.08011182025018**

ABSTRACT

Classification on imbalanced data can affect the accuracy value of the classification and tends to ignore the minority class so that the prediction results will tend to the category. To overcome the problem of imbalanced data, synthetic minority oversampling technique (SMOTE) can be applied. This study aims to obtain an increase in classification accuracy by applying SMOTE to overcome imbalanced data in the classification of rain events using the binary logistic regression method. The data used is secondary data from the weather query builder dataset, namely daily data on rain events in Prabumulih City. The application of SMOTE to the binary logistic regression method for the classification of rain events resulted in an increase in the classification accuracy value in accuracy, precision and fscore, namely 0.27%, 0.91% and 0.04%, this shows that the application of SMOTE for the classification of rain events provides a better level of classification accuracy in accuracy, precision and fscore. While the level of classification accuracy in recall after the application of SMOTE decreased by 0.74%, this was caused by overfitting of synthetic data generated by SMOTE in the non-rain class which could make the model too focused on recognizing train data patterns, thus losing the ability to recognize patterns in test data.

Keyword : SMOTE, Imbalanced class, Binary Logistic Regression.

**PENERAPAN *SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE*
(SMOTE) UNTUK MENGATASI DATA *IMBALANCED* DALAM
KLASIFIKASI KEJADIAN HUJAN MENGGUNAKAN
METODE REGRESI LOGISTIK BINER**

Oleh

**VERTI MONA DESPALIA
NIM.08011182025018**

ABSTRAK

Klasifikasi pada data *imbalanced* dapat mempengaruhi nilai ketepatan pada klasifikasi dan cenderung mengabaikan kelas minoritas sehingga hasil prediksi akan cenderung kepada kategori. Untuk mengatasi masalah data *imbalanced* dapat menerapkan *synthetic minority oversampling technique* (SMOTE). Penelitian ini bertujuan untuk memperoleh peningkatan ketepatan klasifikasi dengan penerapan SMOTE untuk mengatasi data *imbalanced* dalam klasifikasi kejadian hujan menggunakan metode regresi logistik biner. Data yang digunakan berupa data sekunder dari *dataset weather query builder* yakni data harian kejadian hujan Kota Prabumulih. Penerapan SMOTE pada metode regresi logistik biner untuk klasifikasi kejadian hujan menghasilkan peningkatan nilai ketepatan klasifikasi pada *accuracy*, *precision* dan *fscore* yakni 0,27%, 0,91% dan 0,04%, hal ini menunjukkan bahwa penerapan SMOTE untuk klasifikasi kejadian hujan memberikan tingkat ketepatan klasifikasi yang lebih baik pada *accuracy*, *precision* dan *fscore*. Sedangkan tingkat ketepatan klasifikasi pada *recall* setelah penerapan SMOTE mengalami penurunan sebesar 0,74%, hal ini disebabkan oleh *overfitting* data sintesis yang dihasilkan oleh SMOTE pada kelas tidak hujan yang dapat membuat model terlalu fokus dalam mengenali pola data *train*, sehingga kehilangan kemampuan mengenali pola pada data *test*.

Kata Kunci : SMOTE, *Imbalanced class*, Regresi Logistik Biner.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERSEMBAHAN	iii
KATA PENGANTAR.....	iv
ABSTRACT	viii
ABSTRAK	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 <i>Statistical Machine Learning</i>	5
2.2 Klasifikasi	5
2.3 Diskritisasi	7
2.4 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	7
2.5 Regresi Logistik Biner	9
2.5.1 Estimasi Parameter.....	10
2.5.2 Uji Serentak.....	11
2.5.3 Uji Parsial.....	12
2.5.4 Uji Kesesuaian Model.....	13
2.5.5 <i>Odds Rasio</i>	14
2.6 <i>Confusion Matrix</i>	15
BAB III METODOLOGI PENELITIAN	17
3.1 Tempat.....	17
3.2 Waktu	17
3.3 Alat.....	17
3.4 Metode Penelitian.....	17
3.4.1 Data.....	17
3.4.2 Langkah Penelitian	17
BAB IV HASIL DAN PEMBAHASAN	20
4.1 Deskripsi Data	20
4.2 Diskritisasi Data	21
4.3 Partisi Data	23
4.4 Klasifikasi dengan Regresi Logistik Biner pada Data <i>Imbalanced</i>	24
4.4.1 Uji Serentak Data <i>Imbalanced</i>	24
4.4.2 Uji Parsial Data <i>Imbalanced</i>	26
4.4.3 Uji Kesesuaian Model Data <i>Imbalanced</i>	32
4.4.4 <i>Odds Rasio</i> Data <i>Imbalanced</i>	33

4.4.5 Model Terbaik Data <i>Imbalanced</i>	35
4.4.6 Probabilitas Data <i>Imbalanced</i>	36
4.4.7 <i>Confusion Matrix</i> Data <i>Imbalanced</i>	37
4.5 Penerapan <i>Synthetic Minority Oversampling Technique</i> (SMOTE) ...	39
4.6 Klasifikasi dengan Regresi Logistik Biner pada Data <i>Balanced</i>	40
4.6.1 Uji Serentak Data <i>Balanced</i>	40
4.6.2 Uji Parsial Data <i>Balanced</i>	42
4.6.3 Uji Kesesuaian Model Data <i>Balanced</i>	48
4.6.4 <i>Odds Ratio</i> Data <i>Balanced</i>	49
4.6.5 Model Terbaik Data <i>Balanced</i>	52
4.6.6 Probabilitas Data <i>Balanced</i>	52
4.6.7 <i>Confusion Matrix</i> Data <i>Balanced</i>	54
4.7 Analisis Hasil	56
BAB V KESIMPULAN DAN SARAN	58
5.1 Kesimpulan.....	58
5.2 Saran.....	58
DAFTAR PUSTAKA	59
LAMPIRAN.....	61

DAFTAR TABEL

Tabel 2.1 Nilai Model Regresi Logistik	14
Tabel 2.2 <i>Confusion Matrix</i>	15
Tabel 4.1 Deskripsi Variabel	20
Tabel 4.2 Nilai <i>Max</i> , <i>Min</i> dan <i>Range</i>	21
Tabel 4.3 Diskritisasi Data	21
Tabel 4.4 Hasil Diskritisasi	23
Tabel 4.5 Data <i>Train</i>	23
Tabel 4.6 Data <i>Test</i>	24
Tabel 4.7 Estimasi Model Regresi Logistik Biner Data <i>Imbalanced</i>	25
Tabel 4.8 Hasil Uji Serentak Data <i>Imbalanced</i>	25
Tabel 4.9 Hasil Uji Parsial Data <i>Imbalanced</i>	27
Tabel 4.10 Hasil Uji <i>Hosmer</i> dan <i>Lemeshow</i> Data <i>Imbalanced</i>	32
Tabel 4.11 Hasil Klasifikasi Regresi Logistik Biner Data <i>Imbalanced</i>	37
Tabel 4.12 Ketepatan Hasil Klasifikasi Regresi Logistik Biner Data <i>Imbalanced</i>	37
Tabel 4.13 Data <i>Train</i> SMOTE.....	39
Tabel 4.14 Estimasi Model Regresi Logistik Biner Data <i>Balanced</i>	41
Tabel 4.15 Hasil Uji Serentak Data <i>Balanced</i>	42
Tabel 4.16 Hasil Uji Parsial Data <i>Balanced</i>	43
Tabel 4.17 Hasil Uji <i>Hosmer</i> dan <i>Lemeshow</i> Data <i>Balanced</i>	49
Tabel 4.18 Hasil Klasifikasi Regresi Logistik Biner Data <i>Balanced</i>	53
Tabel 4.19 Ketepatan Hasil Klasifikasi Regresi Logistik Biner Data <i>Balanced</i>	54
Tabel 4.20 Analisis Hasil	56

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Klasifikasi.....	6
Gambar 4.1 Proporsi Kejadian Hujan	40

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknik *machine learning* berupa proses mengevaluasi data untuk mengategorikannya ke dalam kelas yang telah ditentukan dapat disebut dengan klasifikasi (Rahayu *et al.*, 2024). Dalam klasifikasi, terdapat dua hal utama yang dilakukan yakni memperoleh model dan menguji model yang telah diperoleh dengan data lain untuk mengetahui seberapa akurat model tersebut. Adapun metode pemodelan yang bisa diterapkan dalam pengklasifikasian data yaitu metode regresi logistik biner.

Regresi logistik biner adalah jenis regresi logistik nonlinier yang mempunyai variabel dependen berupa variabel biner atau bersifat dikotomi yakni terdiri dari dua kategori. (Roflin *et al.*, 2023). Regresi logistik biner diterapkan agar memperoleh model terbaik yang bisa menggambarkan keterkaitan antara variabel dependen dan variabel independen. Metode pemodelan ini sering digunakan pada data *balanced* karena ketika diterapkan pada data *imbalanced* hasil prediksinya bisa menjadi keliru.

Menurut (Zhang *et al.*, 2021), analisis data menggunakan regresi logistik biner pada data *imbalanced* sebagian besar bias terhadap kelas mayoritas dan tidak dapat memprediksi kelas minoritas secara akurat sehingga menghasilkan ketepatan klasifikasi yang kurang baik. Untuk mengatasi masalah data *imbalanced* dapat menerapkan *synthetic minority oversampling technique* (SMOTE). SMOTE

dapat membangun data minoritas sebanyak data mayoritas dan tepat dalam mengatasi suatu permasalahan data yang *overfitting* dalam tahapan *oversampling*.

Penelitian tentang Regresi Logistik Biner berjudul *A Binary Logistic Regression Model as a Tool to Predict Craft Beer Susceptibility to Microbial Spoilage* (Saavedra *et al.*, 2021) diterapkan metode Regresi Logistik Biner memperoleh nilai sensitivitas dan kesesuaian yang tinggi masing-masing 87% dan 83,4%. Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima KIP Kuliah Menggunakan Regresi Logistik Biner (Susetyoko *et al.*, 2022) menghasilkan rata-rata nilai *F-Score* sebesar 92,40%, nilai *recall* sebesar 87,93%, nilai *accuracy* sebesar 88,01%, nilai *precision* sebesar 97,92%, dan terakhir nilai AUC sebesar 84,6%. Kemudian penelitian tentang Ketepatan Klasifikasi Pemberian Kartu Keluarga Sejahtera di Kota Semarang Menggunakan Metode Regresi Logistik Biner dan Metode *Chaid* (Suhendra *et al.*, 2020) menjelaskan bahwa dengan penggunaan metode Regresi Logistik Biner, beberapa faktor yang mempengaruhi yaitu jumlah anggota keluarga dan jenjang pendidikan kepala keluarga yang nilai ketepatan klasifikasinya diperoleh sebesar 88% sedangkan kesalahan sebesar 12%.

Penelitian yang dilakukan oleh (Mualfah *et al.*, 2022) Teknik SMOTE untuk Mengatasi *Imbalance Data* pada Deteksi Penyakit Stroke Menggunakan Algoritma *Random Forest* memperoleh nilai akurasi yakni 0.91, *precision* yakni 0.92, *recall* yakni 0.91, dan *f1-score* yakni 0.91. Walaupun adanya penurunan terhadap nilai akurasi setelah penerapan SMOTE, hal itu dinilai wajar terjadi saat melakukan *balance* pada *class* yang tidak seimbang karena penurunan tingkat

ketepatan klasifikasi ini dapat disebabkan oleh beberapa faktor, termasuk *overfitting* pada data sintesis yang dihasilkan oleh SMOTE, peningkatan *noise* dalam data *train*, atau ketidakmampuan algoritma tertentu untuk beradaptasi dengan distribusi data yang telah diubah. Demikian juga penelitian yang diperoleh oleh (Arifiyanti & Wahyuni, 2020) berjudul SMOTE: Metode Penyeimbang Kelas pada Klasifikasi Data *Mining* dengan menggunakan metode regresi logistik yang telah di SMOTE menghasilkan *accuracy* 74,8%, *precision* sebesar 0,70%, *recall* sebesar 70,3% dan F-Score 12,8%.

Penerapan SMOTE untuk mengatasi data *imbalanced* dalam klasifikasi menggunakan metode regresi logistik biner dapat diterapkan di berbagai kasus salah satunya kasus kejadian hujan. Menurut (Hasyimzoem *et al.*, 2019), hujan adalah fenomena alam dimana jatuhnya butiran air dari langit di suatu wilayah pada waktu tertentu. Fenomena alam di beberapa daerah Indonesia yang terjadi akibat curah hujan yaitu bencana banjir dan kekeringan. Dampak tersebut memberikan tantangan bagi masyarakat Indonesia. Oleh sebab itu, informasi terkait cuaca dan iklim dalam periode pendek terkhusus kejadian hujan bisa menjadi langkah awal untuk mencegah dampak negatif yang disebabkan oleh perubahan cuaca dan iklim khususnya masyarakat Kota Prabumulih.

Masyarakat Kota Prabumulih sebagian bekerja sebagai petani salah satunya usaha tani melon. Berdasarkan data Badan Pusat Statistika (BPS), pada tahun 2022 hasil produksi semusim melon mencapai 2.555 kuintal. Menurut (Zikri & Iskandar, 2020), musim hujan mempengaruhi pengairan budidaya melon karena areal budidaya melon tidak boleh tergenang air sehingga perlu aktif untuk

melakukan pembersihan dan pengecekan irigasi. Saat musim hujan tanaman melon rentan terkena hama dan penyakit, bahkan bisa sampai gagal panen.

Berdasarkan penelitian-penelitian sebelumnya dan permasalahan terkait kejadian hujan di Indonesia membuat penulis tertarik dalam melakukan penelitian terkait penerapan *synthetic minority oversampling technique* (SMOTE) untuk mengatasi data *imbalanced* dalam klasifikasi kejadian hujan menggunakan metode regresi logistik biner.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini yakni apakah penerapan SMOTE untuk mengatasi data *imbalanced* dalam klasifikasi kejadian hujan menggunakan metode regresi logistik biner dapat meningkatkan ketepatan klasifikasi.

1.3 Batasan Masalah

Batasan masalah dari penelitian ini yakni tingkat ketepatan klasifikasi berupa nilai *accuracy*, nilai *precision*, nilai *recall*, dan nilai *fscore*.

1.4 Tujuan Penelitian

Tujuan penelitian ini yakni agar memperoleh peningkatan ketepatan klasifikasi dengan penerapan SMOTE untuk mengatasi data *imbalanced* dalam klasifikasi kejadian hujan menggunakan metode regresi logistik biner.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah dapat menjadi media pembelajaran dalam penerapan SMOTE untuk mengatasi data *imbalanced* dalam klasifikasi kejadian hujan menggunakan metode regresi logistik.

DAFTAR PUSTAKA

- Apriliah, W. Kurniawan, I. Baydhowi, M. & Haryati, T. (2022). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. *SISTEMASI: Jurnal Sistem Informasi*, 10(1), 163-171.
- Arifiyanti, A.A. & Wahyuni, E.D. (2020). SMOTE: Metode Penyeimbang Kelas Pada Klasifikasi Data Mining. *SCAN - Jurnal Teknologi Informasi dan Komunikasi*, 15(1), 34-39.
- Badan Pusat Statistika. (2023). *Kota Prabumulih dalam Angka Prabumulih Municipality In Figures*. Prabumulih.
- Hasyimzoem, E.F., Zakaria, A. & Sumiharni (2019). Perbandingan Analisis Data Curah Hujan yang Hilang Menggunakan Metode Normal Ratio, Inversed Square Distance, Rata-Rata Al-jabar, dan Regresi Berganda. *JRSDD*, 7(1), 155-162.
- Hosmer, D.W., Lemeshow, S. 2000. *Applied Logistic Regression*, Second Edition. Canada : John Wiley and Sons Inc.
- Mualfah, D., Fadila, W. & Firdaus, R. (2022). Teknik SMOTE untuk mengatasi imbalance data pada deteksi penyakit stroke menggunakan algoritma random forest. *Jurnal Computer Science and Information Technology (CoSciTech)*, 3(2), 107-113.
- Nanda, D.M., Pudjiantoro, T.H., & Sabrina, P.N. (2022). Metode K-Nearest Neighbor (KNN) dalam memprediksi curah hujan di Kota Bandung. *Seminar nasional teknik elektro, sistem informasi dan teknik informatika*, 387–393.
- Novaliendry, D. (2015). *Statistika untuk pembelajaran mesin*. Jilid 1, CV. Sarnu Untung, Purwodadi, Indonesia.
- Permana, A.A., Wahyuddin, Santoso, L.W. *et al.* 2023. *Machine Learning*, PT. Global Eksekutif Teknologi, Padang, Indonesia.
- Purwa, T. (2019). Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced. *Jurnal Matematika, Statistika & Komputasi*, 16(1), 58-73.
- Putra, J.W.G. (2020). *Pengenalan pembelajaran mesin dan deep learning*. Edisi 1.4, Tokyo.

- Putri, D.L.W., Mariani, S. & Sunarmi. (2021). Peningkatan Ketepatan Klasifikasi Model Regresi Logistik Biner dengan Metode Bagging (Bootstrap Aggregating). *Indonesian Journal of Mathematics and Natural Sciences*, 44(2), 61-72.
- Rahayu, P.W., Sudipa, I.G.I., Suryani, *et al.* 2024. *Buka Ajar Data Mining*, PT. Sonpedia Publishing Indonesia, Jambi, Indonesia.
- Roflin, E., Riana, F., Pariyana, E.M. & Liberty, I. A. 2023. *Regresi Logistik Biner dan Multinomial*, PT. Nasya Expanding Management, Pekalongan, Indonesia.
- Rohmi, A. L. (2017). Analisis regresi logistik multinomial pada jenis pelanggaran lalu lintas di kota Surabaya. Skripsi Fakultas Vokasi Institut Teknologi Sepuluh Nopember Surabaya.
- Saavedra, M.R., Revelo, K.P., Valero, A., Arribas, M.V.M. & Llano, D.G.D. (2021). A Binary Logistic Regression Model as a Tool to Predict Craft Beer Susceptibility to Microbial Spoilage. *Foods*, 10, 1-15.
- Sabilah, W.I. & Vista, C.B. (2021). Implementasi SMOTE dan Under Sampling pada Imbalanced Dataset untuk Prediksi Kebangkrutan Perusahaan. *Jurnal Komputer Terapan*, 7(2), 329-339.
- Suhendra, M.A., Ispriyanti, D. & Sudarno (2020). Klasifikasi Pemberian Kartu Keluarga Sejahtera di Kota Semarang Menggunakan Metode Regresi Logistik Biner dan Metode Chaid. *Jurnal Gaussian*, 9(1), 64-74.
- Susetyoko, R., Yuwono, W. & Purwantini, E. (2022). Model Klasifikasi Pada Seleksi Mahasiswa Baru Penerima KIP Kuliah Menggunakan Regresi Logistik Biner. *JIP (Jurnal Informatika Polinema)*, 8(10), 31-40.
- Wibowo, M. & Ramadhani, R. (2021). Perbandingan Metode Klasifikasi Data Mining Untuk Rekomendasi Tanaman Pangan. *Jurnal Media Informatika Budidarma*, 5(3), 913-921.
- Zhang, L., Geisler, T., Ray, H. & Xie, Y. (2021). Improving Logistic Regression on The Imbalanced Data by a Novel Penalized Log-likelihood Function. *Jurnal of Applied Statistics*, 49(13), 3257-3277.
- Zikri, M. & Iskandar, S. (2020). Studi Agribisnis dan Tingkat Keuntungan Melon Putih Milik Aji Santoso di Desa Pangkul Jawa Kecamatan Cambai Kota Prabumulih. *Societa*, 9(2), 64-73.