

**PENERAPAN *K-NEAREST NEIGHBORS* DAN *RANDOM  
OVERSAMPLING* PADA KLASIFIKASI KEJADIAN HUJAN  
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE***

**SKRIPSI**

**Sebagai Salah Satu Syarat Untuk Memperoleh Gelar  
Sarjana Sains Bidang Studi Matematika**

Oleh:

**KHOIRIYAH APRIANTI**

**NIM. 08011282025055**



**JURUSAN MATEMATIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS SRIWIJAYA**

**2024**

**LEMBAR PENGESAHAN**

**PENERAPAN K-NEAREST NEIGHBORS DAN RANDOM  
OVERSAMPLING PADA KLASIFIKASI KEJADIAN HUJAN  
MENGUNAKAN METODE SUPPORT VECTOR MACHINE**

**SKRIPSI**

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

**Oleh**

**KHOIRIYAH APRIANTI  
NIM. 08011282025055**

**Pembimbing Kedua**



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

**Indralaya, Juli 2024  
Pembimbing Utama**



**Dra. Ning Eliyati, M.Pd  
NIP. 195911201991022001**

**Mengetahui,  
Ketua Jurusan Matematika**



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

## SURAT PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan dibawah ini:

Nama Mahasiswa : Khoiriyah Aprianti

NIM : 08011282025055

Fakultas/Jurusan : MIPA/Matematika

Menyatakan bahwa skripsi ini adalah hasil karya saya sendiri dan karya ilmiah ini belum pernah diajukan sebagai penentuan persyaratan memperoleh gelar kesarjanaan strata satu (S1) dari Universitas Sriwijaya maupun perguruan tinggi lain.

Semua informasi yang dimuat dalam skripsi ini berasal dari penulisan lain baik yang dipublikasikan atau tidak telah diberikan penghargaan dengan mengutip nama sumber penulis secara benar. Semua isi dari skripsi sepenuhnya menjadi tanggung jawab saya sebagai penulis.

Demikianlah surat pernyataan ini saya buat dengan sebenarnya.

Indralaya, 17 Juli 2024



Khoiriyah Aprianti  
NIM. 08011282025055

## **LEMBAR PERSEMBAHAN**

**“Boleh jadi kamu membenci sesuatu, padahal itu baik bagimu dan boleh jadi kamu menyukai sesuatu, padahal itu buruk bagimu. Allah mengetahui, sedangkan kamu tidak mengetahui”**

**(Q.S Al-Baqarah: 216)**

**“Apabila sesuatu yang kau senangi tidak terjadi maka senangilah apa yang terjadi”**

**-Ali bin Abi Thalib-**

**Skripsi ini saya persembahkan kepada:**

- 1. Allah SWT.**
- 2. Kedua Orangtuaku Tercinta**
- 3. Saudara-Saudaraku Tersayang**
- 4. Keluarga Besarku Tersayang**
- 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 *K-Nearest Neighbors* dan *Random Oversampling* Pada Klasifikasi Kejadian Hujan Menggunakan *Support Vector Machine*”** dengan baik. Skripsi ini merupakan salah satu persyaratan yang harus dipenuhi penulis untuk meraih gelar Sarjana Sains bidang Studi Matematika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Selama menyelesaikan skripsi, penulis 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 ingin mengucapkan terima kasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini terkhusus kepada orang tua tercinta, yaitu Bapak **M. Torik** dan Ibu **Nilawati** atas segala doa terbaik, kasih sayang dan nasihat yang selalu diberikan kepada penulis. Penulis juga mengucapkan terima kasih 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 S. S.Si., M.Si.** selaku Ketua Jurusan Matematika FMIPA Universitas Sriwijaya yang telah memberikan bimbingan, motivasi dan ilmu selama masa perkuliahan.

3. Ibu **Des Alwine Zayanti, S.Si., M.Si.** selaku Sekretaris Jurusan Matematika sekaligus Dosen Pembimbing Kedua yang telah bersedia meluangkan waktu, tenaga dan pikiran untuk memberikan bimbingan, saran, ilmu, motivasi, dan pengarahan yang sangat bermanfaat bagi penulis dalam penyusunan skripsi ini.
4. Ibu **Dra. Ning Eliyati, M.Pd.** selaku Dosen Pembimbing Utama yang telah bersedia meluangkan waktu, tenaga dan pikiran untuk memberikan bimbingan, saran, ilmu, motivasi, dan pengarahan yang sangat bermanfaat bagi penulis dalam penyusunan skripsi ini.
5. Bapak **Drs. Ali Amran, M.T** dan Ibu **Endang Kresnawati, S.Si., M.Si** selaku Dosen Penguji yang telah bersedia meluangkan waktu untuk memberikan tanggapan, ilmu, kritik dan saran yang bermanfaat untuk perbaikan dan penyelesaian skripsi ini.
6. Ibu **Indrawati, S.Si., M.Si** selaku Dosen Pembimbing Akademik yang telah memberikan semangat, motivasi dan pengarahan kepada penulis selama masa perkuliahan.
7. Ibu **Prof. Yulia Resti, S.Si., M.Si., Ph.D.** selaku Ketua Seminar dan Ibu **Yuli Andriani, S.Si., M.Si** selaku Sekretaris Seminar Skripsi yang telah bersedia meluangkan waktu untuk memberikan arahan, saran, membimbing penulis, serta mengatur jalannya seminar sehingga dapat berjalan dengan baik.

8. **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.
9. Bapak **Irwansyah** dan Ibu **Hamidah** selaku Pegawai tata usaha Jurusan Matematika FMIPA yang telah membantu penulis dalam hal administrasi selama perkuliahan.
10. Kakak dan Adik penulis tersayang **RA. Robiatul Adawiyah, Soleha, Fadila Dwi Putri dan M. Yasin Hadi**, serta semua keluarga besar yang selalu memberikan dukungan kepada penulis.
11. Sahabat dan teman seperjuangan **Syakira, Lingling dan Pransiska** atas segala dukungan, motivasi dan semua bantuan yang telah diberikan selama masa perkuliahan hingga selesainya skripsi ini.
12. 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, Juli 2024

**Penulis**

**APPLICATION OF K-NEAREST NEIGHBORS AND RANDOM  
OVERSAMPLING ON RAINFALL EVENT CLASSIFICATION USING  
SUPPORT VECTOR MACHINE METHOD**

**By:**

**KHOIRIYAH APRIANTI**

**NIM. 08011282025055**

**ABSTRACT**

Before performing classification, it is important to ensure that the dataset used does not contain missing data and imbalanced data. Missing data is a condition where some information or data from the dataset is not available. Imbalanced data is a condition where the number of observations in one class in a dataset is much greater than the number of observations in other classes. The purpose of this research is to classify rainfall events using linear *SVM* method by applying *KNN* ( $K=2$ ) and *ROS*. The level of classification accuracy with imbalanced data produces an accuracy value of 83.29%, precision of 78.06%, and recall of 97.29%. While on balanced data by applying *ROS* produces an accuracy value of 92.74%, precision 100%, and recall 86.95%. The results showed that the application of *ROS* succeeded in increasing the accuracy value by 9.45% and precision by 21.94%.

Keywords: *KNN*, Random Oversampling, *SVM*

**PENERAPAN *K-NEAREST NEIGHBORS* DAN *RANDOM  
OVERSAMPLING* PADA KLASIFIKASI KEJADIAN HUJAN  
MENGUNAKAN METODE *SUPPORT VECTOR MACHINE***

**Oleh:**

**KHOIRIYAH APRIANTI**

**NIM. 08011282025055**

**ABSTRAK**

Sebelum melakukan klasifikasi, penting untuk memastikan bahwa dataset yang digunakan tidak mengandung data hilang dan *imbalanced data*. Data hilang adalah kondisi di mana sebagian informasi atau data dari dataset tidak tersedia. *Imbalanced data* adalah kondisi di mana jumlah observasi di satu kelas dalam sebuah dataset jauh lebih banyak dibandingkan dengan jumlah observasi di kelas lainnya. Tujuan penelitian ini adalah mengklasifikasi kejadian hujan menggunakan metode *SVM linear* dengan menerapkan *KNN* ( $K=2$ ) dan *ROS*. Tingkat ketepatan klasifikasi dengan *imbalanced data* menghasilkan nilai akurasi sebesar 83,29%, presisi 78,06%, dan recall 97,29%. Sedangkan pada *balanced data* dengan menerapkan *ROS* menghasilkan nilai akurasi sebesar 92,74%, presisi 100%, dan recall 86,95%. Hasil penelitian menunjukkan bahwa penerapan *ROS* berhasil meningkatkan nilai akurasi sebesar 9,45% dan presisi sebesar 21,94%.

Kata Kunci: *KNN, Random Oversampling, SVM*

## DAFTAR ISI

<b>LEMBAR PENGESAHAN .....</b>	<b>ii</b>
<b>LEMBAR PERSEMBAHAN .....</b>	<b>iii</b>
<b>KATA PENGANTAR.....</b>	<b>iv</b>
<b>ABSTRACT .....</b>	<b>vii</b>
<b>ABSTRAK .....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>ix</b>
<b>DAFTAR TABEL .....</b>	<b>xi</b>
<b>DAFTAR GAMBAR.....</b>	<b>xii</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
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
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>6</b>
2.1. Data hilang.....	6
2.2. <i>K-Nearest Neighbor (KNN)</i> .....	7
2.3. Standarisasi Data.....	8
2.4. <i>Imbalanced Data</i> .....	10
2.5. <i>Random Oversampling (ROS)</i> .....	11
2.6. <i>Machine Learning</i> .....	12
2.7. Klasifikasi .....	13
2.8. <i>Support Vector Machine (SVM)</i> .....	13
2.9. <i>Hard Margin SVM: Linear and Separable Case</i> .....	15
2.10. <i>Soft Margin SVM: Linear and Non-Separable Case</i> .....	19
2.11. <i>Confusion Matrix</i> .....	22
2.12. Variabel Penelitian.....	24
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>25</b>
3.1. Tempat Penelitian .....	25
3.2. Waktu Penelitian.....	25
3.3. Alat .....	25
3.4. Metode Penelitian .....	25

<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>28</b>
4.1. Deskripsi Data.....	28
4.2. Imputasi Data Hilang.....	29
4.3. Standarisasi Data.....	32
4.4. Pembagian <i>Data training</i> dan <i>Data Testing</i> .....	33
4.5. Menyeimbangkan Kelas Dengan ROS .....	35
4.6. Klasifikasi <i>SVM Linear</i> .....	36
4.6.1. Klasifikasi <i>SVM Linear</i> dengan <i>Imbalanced data</i> .....	36
4.6.2. Klasifikasi <i>SVM Linear</i> dengan <i>Balanced data</i> .....	42
4.7. Analisis Perbandingan Hasil Klasifikasi <i>SVM linear</i> .....	47
<b>BAB V KESIMPULAN .....</b>	<b>48</b>
5.1. Kesimpulan .....	48
5.2. Saran .....	48
<b>DAFTAR PUSTAKA .....</b>	<b>49</b>

## DAFTAR TABEL

Tabel 2.1 Confusion Matrix .....	23
Tabel 2.2 Variabel Penelitian.....	24
Tabel 4.1 Data Cuaca Harian Kota Palembang .....	28
Tabel 4.2 Variabel-Variabel yang Mengandung Data Hilang.....	29
Tabel 4.3 Observasi yang Mengandung Data Hilang ( $x_{aj}$ ).....	29
Tabel 4.4 Observasi Lengkap yang Tidak Mengandung Data Hilang ( $x_{bj}$ ) .....	29
Tabel 4.5 Jarak Euclidean antara Observasi ke-1139 dengan Observasi Lengkap 31	
Tabel 4.6 Jarak Terkecil antara Observasi ke-1139 dengan Observasi Lengkap ..	31
Tabel 4.7 Hasil Standarisasi Data Menggunakan Nilai Z-Score.....	33
Tabel 4.8 Pembagian Jumlah Data training dan Data Testing.....	33
Tabel 4.9 Data Training.....	34
Tabel 4.10 Data Testing.....	34
Tabel 4.11 Data Training Hasil Proses ROS .....	36
Tabel 4.12 Support Vector dari SVM linear dengan Imbalanced Data .....	37
Tabel 4.13 Nilai Bias per Support Vector untuk Imbalanced Data .....	39
Tabel 4.14 Hasil Klasifikasi SVM Linear dengan Imbalanced Data .....	40
Tabel 4.15 Confusion Matrix Klasifikasi SVM linear dengan Imbalanced Data .	41
Tabel 4.16 Support Vector dari SVM Linear dengan Balanced Data.....	42
Tabel 4.17 Nilai Bias per Support Vector untuk Balanced Data .....	44
Tabel 4.18 Hasil Klasifikasi SVM Linear dengan Balanced Data .....	45
Tabel 4.19 Confusion Matrix Klasifikasi SVM linear dengan Balanced Data .....	46
Tabel 4.20 Perbandingan Hasil Klasifikasi SVM Linear .....	47

## DAFTAR GAMBAR

Gambar 2.1 Random Oversampling.....	11
Gambar 2.2 Konsep Hyperplane pada SVM Linear .....	14
Gambar 2.3 Hard Margin Hyperplane .....	16
Gambar 2.4 Soft Margin Hyperplane.....	20
Gambar 4.1 Nilai Akurasi Klasifikasi SVM Linear Dengan Imputasi KNN.....	30
Gambar 4.2 Distribusi Kelas Variabel Y Data Training.....	34
Gambar 4.3 Proses Random Oversampling .....	35

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Data hilang adalah kondisi di mana sebagian informasi atau data dari dataset tidak tersedia (Eldiyana, et al., 2021). Data hilang dapat menjadi masalah yang sangat penting karena dapat mengakibatkan adanya bias dan ketidakakuratan dalam melakukan prediksi terhadap respon dan amatan (Oktavianti dan Yanti, 2022). Adapun cara untuk mengatasi data hilang adalah dengan melakukan imputasi data.

Imputasi data bertujuan untuk menggantikan atau mengisi data yang hilang dalam dataset, sehingga dataset yang awalnya tidak lengkap menjadi lengkap tanpa mengurangi jumlah unit data yang di observasi. Hal ini memungkinkan dataset tersebut untuk digunakan dalam analisis lebih lanjut (Fadillah dan Puspita, 2020). Salah satu metode imputasi data yaitu *K-Nearest Neighbor (KNN)*.

Berdasarkan penelitian Prasetya et al. (2023), imputasi data dengan metode *KNN*, *Neural Network*, nilai mean, dan nilai maksimum digunakan untuk mengatasi masalah data hilang. Hasil penelitian menunjukkan bahwa metode *KNN* adalah metode paling efektif karena menghasilkan nilai *Root Mean Squared Error (RMSE)* terkecil. Penelitian lain dilakukan oleh Aljrees (2024) membuktikan bahwa, dibandingkan imputasi data menggunakan *Principal Component Analysis (PCA)* dan penghapusan data hilang, imputasi data dengan metode *KNN* berhasil mencapai akurasi tertinggi sebesar 99,41% dalam memprediksi kanker serviks menggunakan metode *ensemble* berdasarkan *Extreme Gradient Boosting (XGB)*, *Extra Tree Classifier (ETC)* dan *Random Forest (RF)*.

Pada penelitian ini, peneliti akan menggunakan data cuaca harian Kota Palembang tahun 2018-2023 untuk melakukan klasifikasi kejadian hujan menggunakan metode *Support Vector Machine (SVM)*. Klasifikasi merupakan proses pengelompokan data berdasarkan variabel-variabel tertentu kedalam kelas-kelas yang berbeda. Melalui proses klasifikasi, memungkinkan data menjadi lebih mudah dipahami dan terorganisir, sehingga mempermudah dalam melakukan pencarian dan analisis data (Nasution dan Fatonah, 2023). Tujuan utama metode *SVM* adalah untuk menemukan *hyperplane* terbaik yang berfungsi sebagai pemisah optimal antara kelas-kelas yang terlibat (Fikri, et al., 2020).

Beberapa penelitian yang menggunakan metode *SVM* diantaranya adalah penelitian Laia dan Setyawan (2020). Pada penelitiannya digunakan metode *SVM* dan *Naïve Bayes Classifier (NBC)* untuk mengklasifikasi curah hujan. Hasilnya menunjukkan bahwa akurasi metode *SVM* mencapai 79.45%, sementara metode *NBC* mencapai 65.75%. Penelitian lain dilakukan oleh Dawoodi and Patil (2023) menggunakan metode yang sama, hasilnya menunjukkan bahwa metode *SVM* mencapai akurasi 93%, lebih tinggi dibandingkan *Naïve Bayes* yang mencapai 92%. Sementara penelitian Hudnurkar and Rayavarapu (2022) diperoleh hasil akurasi metode *SVM* sebesar 82,1% sedangkan metode *Artificial Neural Network (ANN)* sebesar 82,8%.

Menurut penelitian (Hamami dan A. Dahlan, 2022) salah satu masalah dalam klasifikasi adalah ketidakseimbangan data (*imbalance data*). Kemungkinan terjadinya *false negative* atau *false positive* menjadi cukup tinggi di kelas minoritas jika data tiap kelas tidak tersebar secara merata atau seimbang. Keadaan ini dapat

menyebabkan *classifier* kurang mampu untuk memahami dengan baik kelas minoritas, sehingga prediksinya cenderung lebih condong ke arah kelas mayoritas. Kelas minoritas merupakan kelas dengan kelompok data yang lebih sedikit, sedangkan kelas dengan kelompok data lainnya disebut kelas mayoritas (Qadrini, et al., 2022). Teknik *oversampling* dapat digunakan untuk mengatasi *imbalance data*. Salah satu teknik *oversampling* yang banyak digunakan yaitu *Random Oversampling (ROS)*.

Pada algoritma *ROS*, data dari kelas minoritas secara acak akan dipilih, setelah itu ditambahkan ke dalam dataset pelatihan. Proses ini diulang hingga jumlah data dari kelas minoritas setara dengan jumlah data dari kelas mayoritas (Yulianti, 2022). Penelitian terdahulu yang menggunakan *ROS* adalah penelitian Ramadhan, et al., (2023) berjudul “*a hybrid ROS-SVM for detecting target multiple drug types*”. Hasil penelitiannya menunjukkan bahwa *SVM* kernel *linear* mencapai 97% dengan penerapan *ROS*, sedangkan tanpa *ROS* sebesar 78%. Penelitian lain dilakukan oleh Hayaty, et al., (2021), berdasarkan penelitiannya diketahui bahwa penerapan *Synthetic Over-Sampling (SOS)* dan *ROS* berhasil meningkatkan akurasi yang semula 12% menjadi 100% pada klasifikasi donor darah menggunakan metode *KNN*.

Berdasarkan penjelasan yang telah disebutkan sebelumnya, penelitian terdahulu memperlihatkan bahwa metode *SVM* menghasilkan akurasi yang lebih tinggi dibandingkan *naïve bayes*. Oleh karena itu, peneliti memilih metode *SVM* untuk mengklasifikasikan kejadian hujan Kota Palembang tahun 2018-2023. Namun, untuk meningkatkan kinerja model *SVM*, peneliti akan terlebih dahulu

mengatasi data hilang dan menyeimbangkan jumlah data antar kelas dengan menggunakan metode *KNN* dan teknik *ROS*. Penelitian ini diharapkan dapat menghasilkan suatu sistem cerdas yang mampu memberikan prediksi apakah akan terjadi hujan atau tidak berdasarkan data yang diberikan dengan menggunakan metode *SVM*.

## 1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah:

1. Bagaimana tingkat ketepatan klasifikasi kejadian hujan menggunakan metode *SVM linear* dengan data tidak seimbang (*imbalanced data*)?
2. Bagaimana tingkat ketepatan klasifikasi kejadian hujan menggunakan metode *SVM linear* dengan data seimbang (*balanced data*) menggunakan *ROS*?
3. Apakah penyeimbangan kelas menggunakan *ROS* berhasil meningkatkan kinerja model *SVM linear*?

## 1.3. Batasan Masalah

Batasan pada penelitian ini, yaitu:

1. Data yang digunakan merupakan data cuaca harian Kota Palembang dari tanggal 1 Januari 2018 sampai dengan 31 Desember 2023.
2. Penelitian ini menggunakan metode *SVM linear*.
3. Tingkat ketepatan klasifikasi mencakup nilai akurasi, presisi, dan *recall*.

4. Variabel yang digunakan terdiri dari 1 variabel *dependent* yaitu *icon* dan 17 variabel *independent*, yaitu *Temperature Max*, *Temperature Min*, *Temperature*, *Feelslike*, *Dew*, *Humidity*, *Precipitation*, *Wind Gust*, *Wind Speed*, *Wind Direction*, *Sea Level Pressure*, *Cloud Cover*, *Visibility*, *Solar Radiation*, *Solar Energy*, *Severe Risk*, dan *Moon Phase*.

#### **1.4. Tujuan Penelitian**

Berikut beberapa tujuan dari penelitian ini:

1. Menentukan tingkat ketepatan klasifikasi kejadian hujan menggunakan metode *SVM linear* dengan data tidak seimbang (*imbalanced data*)
2. Menentukan tingkat ketepatan klasifikasi kejadian hujan menggunakan metode *SVM linear* dengan data seimbang (*balanced data*) menggunakan *ROS*
3. Mengetahui apakah penyeimbangan kelas menggunakan *ROS* berhasil meningkatkan kinerja model *SVM linear*

#### **1.5. Manfaat Penelitian**

Berikut beberapa manfaat dari penelitian ini:

1. Mendapatkan pemahaman tentang efektivitas metode *SVM linear* dalam mengklasifikasikan kejadian hujan ketika dihadapkan pada *imbalanced data*.
2. Mendapatkan pemahaman tentang efektivitas teknik *ROS* dalam meningkatkan kinerja model klasifikasi dengan metode *SVM linear*.

## DAFTAR PUSTAKA

- Ahmad, H, dkk. (2022). Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kebakaran Hutan dan Lahan Berdasarkan Sebaran Titik Panas. *TEKNIMEDIA: Teknologi Informasi dan Multimedia*, 2(2), 83–89.
- Aljrees, Turki. (2024). Improving prediction of cervical cancer using *KNN* imputer and multi-model ensemble learning. *PLOS ONE*, 19(1).
- Alfarizi, M.R. Sirfatullah, dkk. (2023). Penggunaan Phyton Sebagai Bahasa Pemrograman Untuk Machine Learning dan Deep Learning. *Karimah Tauhid*, 2(1).
- Amir, R.F., Sobari, I.A., dan Rousyati. (2020). Penerapan PSO Over Sampling Dan Adaboost Random Forest Untuk Memprediksi Cacat Software. *IJSE-Indonesian Journal on Software Engineering*, 6(2), 230–239.
- Chazar, C. (2020). Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine. *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, 12(1), 67–80.
- Dawoodi, H.H. and Patil, M.P. (2023). Rainfall Prediction for North Maharashtra, India Using Advanced Machine Learning Models. *Indian Journal Of Science And Technology*, 16(13), 956–966.
- Demidova, L.A. (2021). Two-Stage Hybrid Data Classifiers Based on *SVM* and *kNN* Algorithms. *Symmetry*, 13(4), 615.
- Eldiyana, I., Nurlaelah, E., dan Herrhyanto, N. (2021). Estimasi Missing Data dengan Metode Multivariate Imputation by Chained Equation (Mice) untuk Membentuk Persamaan Regresi *Linear* berganda. *Jurnal EurekaMatika*, 9(1), 95–106.
- Fadillah, I.J. dan Puspita, C.D. (2021). Pemanfaatan Metode Weighted K-Nearest Neighbor Imputasi (Weighted *KNNI*) Untuk Mengatasi Missing Data. *Seminar Nasional Official Statistics, 2020(1)*, 511–518.
- Fikri, M.I., Sabrila, T.S., dan Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *STIKI Informatika Jurnal*, 10(2), 71–76.
- Hadnurkar, S. and Rayavarapu, N. (2022). Binary classification of rainfall time-series using machine learning algorithms. *IJECE: International Journal of Electrical and Computer Engineering*, 12(2), 1945–1954.

- Hamami, F. dan Dahlan, I.A. (2022). Klasifikasi Cuaca Provinsi DKI Jakarta Menggunakan Algoritma Eandom Forest dengan Teknik Oversampling. *Jurnal Teknoinfo*, 16(1), 87.
- Handayany, I., Kusnandar, D. dan Andani, W. (2023). Pendugaan Data Hilang Menggunakan Perbandingan Imputasi Hot-Deck dan K-Nearest Neighbor. *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, 319–324.
- Henderi, dkk. (2021). Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *IJIIS: International Journal of Informatics and Information Systems*, 4(1), 13–20.
- Huda, M. (2019). *Algoritma Data Mining: Analisis Data Dengan Komputer*. bisakimia.
- Hsu, C and Lin, C. (2002). A Comparison of Methods for Multi-class Support Vector Machines. *IEEE Transaction on Neural Network*, 13(2), 415–425.
- Indrawati, A. (2021). Penerapan Teknik Kombinasi Oversampling dan Undersampling Untuk Mengatasi Permasalahan *Imbalanced dataset*. *JIKO (Jurnal Informatika dan Komputer)*, 4(1), 38–43.
- Laia, M.L. dan Setyawan, Y. (2020). Perbandingan Hasil Klasifikasi curah Hujan menggunakan metode *SVM* dan NBC. *Jurnal Statistika Industri dan Komputasi*, 5(2), 51–61.
- Mestika, J.C., dkk. (2022). Menjelajahi Teknik-Teknik Supervised Learning untuk Pemodelan Prediktif Menggunakan Python. *Buletin Ilmiah Ilmu Komputer dan Multimedia (BIKMA)*, 99(99), 216–219.
- Nasution, A.L. dan Fatonah, R.N.S. (2023). *Klasifikasi Kondisi Peralatan Elektronik Metode Gaussian Naïve Bayes*. Penerbit Buku Pedia.
- Nurhopipah, A. dan Magnolia, C. (2022). Perbandingan Metode Resampling Pada *Imbalanced dataset* Untuk Klasifikasi Komentar Program MBKM. *Jurnal Publikasi Ilmu Komputer Multimedia*, 1(2), 9–22.
- Oktavianti, S.A. dan Yanti, T.S. (2022). Penanganan Data Hilang Menggunakan Metode MarkoviChain Monte Carlo (MCMC). *Bandung Conference Series: Statistics*, 2(2), 409–415.
- Prasetya, M.R.A., Priyatno, A.M. dan Nurhaeni (2023). Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining. *Jurnal Informasi dan Teknologi*, 5(2), 56–62.

- Qadrini, L., Hikmah dan Megasari (2022). Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejawah Timur Tahun 2017. *Journal of Computer System and Informatics (JoSYC)*, 3(4), 386–391.
- Ramadhan, N.G., *et al.* (2023). A Hybrid ROS-SVM Model for Detecting Target Multiple Drug Types. *JOIV: International Journal on Informatics Visualization*, 7(3), 794–800.
- Sallaby, A.F., and Azlan. (2021). Analysis of Missing Value Imputation Application with K-Nearest Neighbor (K-NN) Algorithm in Dataset. *The IJICS (International Journal of Informatics and Computer Science)*, 5(2), 141–144.
- Sang, A.I., dkk. (2021). Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma Decision Tree dan Support Vector Machine. *e-Proceeding of Engineering*, 8(5), 8954–8963.
- Suryanegara, G.A.B., Adiwijaya dan Purbolaksono, M.D. (2021). Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(1), 114–122.
- Suweleh, A.S., Susilowati, D. dan Hairani (2020). Aplikasi Penentuan Penerima Beasiswa Menggunakan Algoritma C4.5. *Jurnal BITE*, 2(1), 12–21.
- Swastika, R. Mukodimah, S., Susanto, F., Muslihudin, M., & Ipnuwati, S. (2023) *Implementasi Data Mining (Clustering, Association, Prediction, Estimation, Classification)*. Penerbit Adab.
- Werdiningsih, I., Novitasari, D.C.R. & Haq, D.Z. (2022). *Pengelolaan Data Mining dengan Pemrograman Matlab*. Airlangga University Press.
- Werdiningsih, I., Nuqoba, B. dan Muhammadun (2020). *Data Mining Menggunakan Android, Weka, dan SPSS*. Airlangga University Press.
- Whendasmoro, R.G. dan Joseph (2022). Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 872–876.
- Yu, D.J., *et al.* (2013). Improving protein-ATP binding residues prediction by boosting SVMs with random under-sampling. *Neurocomputing*, 104, 180–190.
- Yulianti (2022) *Penerapan Resampling dan Bagging Berbasis Naive Bayes*. Pascal

Zaki, M.J. and Meira Jr, W. (2014). *Data Mining and Analysis*. Cambridge University Press.