

**IMPLEMENTASI KNN-SMOTE-ROS-RUS DALAM
MENGKLASIFIKASIKAN KEJADIAN HUJAN
MENGUNAKAN METODE *NAÏVE BAYES***

SKRIPSI

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

Oleh :

SYAKIRA ADININGRUM

08011282025035



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

2024

LEMBAR PENGESAHAN

**IMPLEMENTASI KNN-SMOTE-ROS-RUS DALAM
MENGKLASIFIKASIKAN KEJADIAN HUJAN
MENGUNAKAN METODE *NAÏVE BAYES***

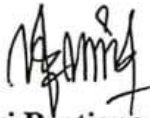
SKRIPSI

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

Oleh

**SYAKIRA ADININGRUM
NIM. 08011282025035**

Pembimbing Kedua



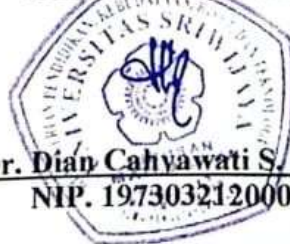
**Novi Rustiana Dewi, S.Si., M.Si
NIP. 197011131996032002**

**Indralaya, Mei 2024
Pembimbing Utama**



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

**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 : Syakira Adiningrum

NIM : 08011282025035

Fakultas/Jurusan : MIPA/Matematika

Menyatakan bahwa skripsi ini adalah hasil karya saya sendiri dan karya ilmiah ini belum pernah diajukan sebagai penentuan persyaratan untuk 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, 22 Mei 2024



Syakira Adiningrum
NIM. 08011282025035

LEMBAR PERSEMBAHAN

Motto:

**“Dan sesungguhnya telah Kami berikan hikmat kepada Luqman, yaitu:
"Bersyukurlah kepada Allah. Dan barangsiapa yang bersyukur (kepada
Allah), maka sesungguhnya ia bersyukur untuk dirinya sendiri; dan
barangsiapa yang tidak bersyukur, maka sesungguhnya Allah Maha Kaya
lagi Maha Terpuji””**

(QS. Al-Luqman : 12)

“Selesaikan yang sudah dimulai”

Skripsi ini kupersembahkan kepada:

- 1. Allah SWT**
- 2. Kedua Orangtuaku**
- 3. Saudara-saudaraku**
- 4. Semua Dosen dan Guruku**
- 5. Teman-temanku**
- 6. Almamaterku**

KATA PENGANTAR

“Allhamdulillah” penulis panjatkan atas kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Implementasi KNN-SMOTE-ROS-RUS dalam Mengklasifikasikan Kejadian Hujan Menggunakan Metode *Naïve Bayes***”.

Penulis menyadari bahwa terdapat berbagai kendala dalam proses penyusunan skripsi ini, tanpa bantuan dan bimbingan dari pihak-pihak terkait kendala tersebut tidak dapat diatasi. Penulis dengan hormat dan rendah hati mengucapkan terima kasih kepada Bapak **Sukatno** dan Ibu **Windiarti** atas semua doa dan dukungan selama seumur hidup penulis.

Dalam proses penyelesaian skripsi ini, penulis ingin mengucapkan terima kasih kepada pihak-pihak terkait berikut ini:

1. Bapak **Prof. Hermansyah, S. Si., M. Si., Ph. D** selaku Dekan FMIPA Universitas Sriwijaya atas ilmu dan bimbingan yang diberikan selama perkuliahan.
2. Ibu **Dr. Dian cahyawati Sukanda, S. Si., M. Si** selaku ketua jurusan Matematika FMIPA Universitas Sriwijaya yang telah memberikan ilmu dan bimbingan selama perkuliahan.
3. Ibu **Des Alwine Zayanti, S.Si., M.Si** selaku sekretaris jurusan Matematika FMIPA Universitas Sriwijaya dan Dosen Pembimbing Akademik yang telah memberikan bimbingan, arahan serta nasihat selama menempuh perkuliahan.

4. Ibu **Prof. Yulia Resti, S.Si., M.Si. Ph. D** selaku Dosen Pembimbing Pertama yang telah mencurahkan waktu, bimbingan, arahan dan saran yang sangat berharga bagi penulis dalam menyelesaikan skripsi dengan optimal.
5. Ibu **Novi Rustiana Dewi, S.Si., M.Si** selaku Dosen Pembimbing Kedua yang telah mencurahkan waktu, bimbingan, arahan dan saran yang sangat berharga bagi penulis dalam menyelesaikan skripsi dengan optimal.
6. Ibu **Dra. Ning Eliyati, M.Pd** selaku Dosen Penguji Pertama yang telah memberikan masukan, kritik dan saran yang konstruktif dalam penyelesaian skripsi ini.
7. Ibu **Dr. Yuli Andriani, S.Si., M.Si** selaku Dosen Penguji Kedua yang telah memberikan masukan, kritik dan saran yang konstruktif dalam penyelesaian skripsi ini.
8. Ibu **Irmeilyana, S.Si., M.Si** selaku Ketua Seminar dan Ibu Dr. Dian Cahyawati, S.Si., M.Si selaku Sekretaris Seminar atas upaya dan kontribusi dalam memastikan kelancaran seminar skripsi penulis.
9. **Seluruh Bapak/Ibu Dosen dan staff di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sriwijaya**, atas sumbangsih ilmu pengetahuan, bimbingan, nasihat, serta motivasi yang telah diberikan kepada penulis selama perkuliahan.
10. **Selfi** yang dengan kasih sayang selalu memberikan dukungan dan doa.
11. Sahabatku **Ainun** yang selalu mendukung dan memberikan saran serta motivasi bagi penulis dalam proses penyelesaian skripsi ini.

12. Sahabat seperjuangan **Khoiriyah** dan **Pransiska** yang telah memberikan dukungan, semangat serta motivasi bagi penulis dalam menyelesaikan skripsi.
13. Semua pihak yang tidak dapat disebutkan satu persatu atas bantuan dan dukungannya sehingga penulis dapat menyelesaikan skripsi ini.
14. *I am grateful to myself for never giving up, working hard, taking care of myself, looking for inspiration all the time and always being willing to learn new things.*

Penulis berharap skripsi ini dapat membantu pihak-pihak yang ingin mempelajari dan memperdalam pemahaman mengenai KNN, SMOTE, RUS dan ROS serta klasifikasi *Naïve Bayes* terutama dengan menggunakan data kejadian hujan.

Indralaya, Mei 2024

Penulis

**IMPLEMENTATION OF KNN-SMOTE-ROS-RUS IN
CLASSIFYING RAINFALL EVENTS USING
NAÏVE BAYES METHOD**

Oleh:

Syakira Adiningrum

08011282020535

ABSTRACT

Topographically, Palembang city consists mostly of swamps and lowlands. The city's condition of being surrounded by rivers and its moderate annual rainfall of 2500-3000 mm makes Palembang prone to flooding. Therefore, accurately predicting rainfall events is crucial to minimize the impact on daily activities. The rainfall prediction classification process, missing data and imbalanced data were encountered in the rainfall event dataset. Missing data was addressed using KNN (K=17) and standardization was performed to standardize the data scale. Subsequently, to address imbalanced data, SMOTE, ROS and RUS techniques were applied. Thus, for the imbalanced data, an accuracy of 98.77%, precision of 97.83%, and recall of 100% were obtained. Furthermore, for the balanced data using SMOTE, an accuracy of 98.22%, precision of 96.89%, and recall of 100% were achieved, while ROS yielded an accuracy of 98.90%, precision of 98.06%, and recall of 100%, and RUS yielded an accuracy of 99.31%, precision of 98.78%, and recall of 100%. Based on these findings, the best rainfall event prediction classification for Palembang city was obtained by addressing imbalanced data using the RUS technique.

Kata kunci : KNN, SMOTE, ROS, RUS, Rain

**IMPLEMENTASI KNN-SMOTE-ROS-RUS DALAM
MENGKLASIFIKASIKAN KEJADIAN HUJAN
MENGUNAKAN METODE NAÏVE BAYES**

Oleh

Syakira Adiningrum

08011282020535

ABSTRAK

Secara topografi, Kota Palembang sebagian besar berupa rawa dan dataran rendah. Kondisi wilayah yang dikelilingi oleh sungai-sungai serta curah hujan dalam kategori sedang antara 2500-3000 mm pertahun membuat Kota Palembang rentan terhadap banjir. Sehingga dengan melakukan klasifikasi kejadian hujan diharapkan dapat meminimumkan dampak ketika terjadi hujan dalam kehidupan sehari-hari. Dalam proses klasifikasi ditemukan *missing data* dan *imbalanced data* pada data kejadian hujan. *Missing data* diatasi menggunakan KNN (K=17) dan standarisasi dilakukan untuk menyeragamkan skala pada data. Selanjutnya untuk mengatasi *imbalanced data* diterapkan SMOTE, ROS dan RUS. Dengan demikian diperoleh akurasi pada *imbalanced data* sebesar 98,77%, presisi 97,83% dan *recall* 100%. Selanjutnya pada *balanced data* menggunakan SMOTE diperoleh akurasi 98,22%, presisi 96,89% dan *recall* 100% sedangkan ROS diperoleh akurasi 98,90%, presisi 98,06% dan *recall* 100% dan RUS diperoleh akurasi 99,31%, presisi 98,78% dan *recall* 100%. Berdasarkan hal-hal tersebut diperoleh hasil klasifikasi kejadian hujan di Kota Palembang terbaik dengan mengatasi *imbalanced data* menggunakan RUS.

Kata kunci : KNN, SMOTE, ROS, RUS, Hujan

DAFTAR ISI

| | |
|---|-------------|
| LEMBAR PENGESAHAN..... | ii |
| LEMBAR PERSEMBAHAN | iii |
| KATA PENGANTAR..... | iv |
| ABSTRACT | vii |
| ABSTRAK | viii |
| DAFTAR ISI..... | ix |
| DAFTAR TABEL | xi |
| DAFTAR GAMBAR..... | xiii |
| DAFTAR LAMPIRAN | xiv |
| 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 <i>Data mining</i> | 6 |
| 2.2 <i>Missing Data</i> | 6 |
| 2.3 <i>K-Nearest Neighbors</i> | 8 |
| 2.4 <i>Imbalanced Data</i> | 10 |
| 2.5 <i>Sythetic Minority Oversampling Technique</i> | 11 |
| 2.6 <i>Random Oversampling</i> | 12 |
| 2.7 <i>Random Undersampling</i> | 13 |
| 2.8 Klasifikasi..... | 13 |
| 2.9 Metode <i>Naïve Bayes</i> | 14 |
| 2.10 <i>Confusion Matrix</i> | 16 |
| 2.11 Hujan | 17 |
| BAB III METODOLOGI PENELITIAN | 18 |
| 3.1 Tempat..... | 18 |
| 3.2 Waktu | 18 |
| 3.3 Alat | 18 |
| 3.4 Data penelitian..... | 18 |

| | | |
|--|--|-----------|
| 3.5 | Metode Penelitian | 19 |
| BAB IV HASIL DAN PEMBAHASAN | | 21 |
| 4.1 | Deskripsi Data | 21 |
| 4.2 | Imputasi Data Menggunakan KNN | 22 |
| 4.3 | Standarisasi Data | 30 |
| 4.4 | <i>Training Data</i> dan <i>Testing Data</i> | 30 |
| 4.5 | Penerapan SMOTE untuk <i>Imbalanced Data</i> | 31 |
| 4.6 | ROS untuk <i>Imbalanced Data</i> | 37 |
| 4.7 | RUS untuk <i>Imbalanced Data</i> | 37 |
| 4.8 | Klasifikasi <i>Naïve Bayes</i> Saat <i>Imbalanced Data</i> | 38 |
| 4.9 | Klasifikasi <i>Naïve Bayes</i> dengan SMOTE, ROS, RUS | 41 |
| 4.10 | Perbandingan Hasil Klasifikasi | 43 |
| BAB V KESIMPULAN DAN SARAN | | 46 |
| 5.1 | Kesimpulan..... | 46 |
| 5.2 | Saran | 46 |
| DAFTAR PUSTAKA | | 47 |
| LAMPIRAN..... | | 50 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2. 1 <i>Confusion Matrix</i> | 16 |
| Tabel 3. 1 Variabel yang Digunakan dalam Penelitian..... | 18 |
| Tabel 3. 1 (Lanjutan) Variabel yang Digunakan dalam Penelitian | 19 |
| Tabel 4. 1 Deskripsi Variabel | 21 |
| Tabel 4. 2 Variabel yang Memiliki <i>Missing Data</i> | 22 |
| Tabel 4. 3 Contoh Data Sampel Perhitungan RMSE dengan K=3 | 23 |
| Tabel 4. 4 Perbandingan Nilai RMSE Terhadap K..... | 24 |
| Tabel 4. 5 Contoh Data Sampel untuk Imputasi KNN..... | 25 |
| Tabel 4. 6 Jarak Euclidean Antara Sampel-Sampel | 26 |
| Tabel 4. 7 Mengurutkan K untuk Imputasi KNN | 27 |
| Tabel 4. 8 Nilai Observasi untuk Variabel <i>Wind Gust</i> | 27 |
| Tabel 4. 8 (Lanjutan) Nilai Observasi untuk Variabel <i>Wind Gust</i> | 28 |
| Tabel 4. 9 Nilai Observasi untuk Variabel <i>Severe Risk</i> | 28 |
| Tabel 4. 9 (Lanjutan) Nilai Observasi untuk Variabel <i>Severe Risk</i> | 29 |
| Tabel 4. 10 <i>Training Data</i> | 31 |
| Tabel 4. 11 <i>Testing Data</i> | 31 |
| Tabel 4. 12 Contoh Data Sampel untuk SMOTE..... | 33 |
| Tabel 4. 13 Penentuan <i>Nearest Neighbors</i> untuk Proses SMOTE..... | 34 |
| Tabel 4. 14 Hasil Data Sintetik Pertama untuk Contoh Data Sampel SMOTE | 34 |
| Tabel 4. 15 Hasil Data Sintetik Kedua untuk Contoh Data Sampel SMOTE..... | 35 |
| Tabel 4. 16 Data Sampel <i>Balanced</i> Setelah Menggunakan SMOTE..... | 36 |
| Tabel 4. 17 Distribusi Kelas Setelah ROS | 37 |

| | |
|---|----|
| Tabel 4. 18 Distribusi Kelas Setelah RUS | 37 |
| Tabel 4. 19 $P(X_1, X_2, \dots, X_{17} Y)$ | 38 |
| Tabel 4. 20 <i>Confusion Matrix</i> Klasifikasi <i>Naïve Bayes</i> Saat <i>Imbalanced Data</i> .. | 40 |
| Tabel 4. 21 <i>Confusion Matrix</i> Setelah SMOTE | 41 |
| Tabel 4. 22 <i>Confusion Matrix</i> Setelah ROS..... | 42 |
| Tabel 4. 23 <i>Confusion Matrix</i> Setelah RUS..... | 42 |
| Tabel 4. 24 Perbandingan Hasil Penelitian dengan Penelitian Sebelumnya | 43 |
| Tabel 4. 24 (Lanjutan) Perbandingan Hasil Penelitian dengan Penelitian Sebelumnya..... | 44 |

DAFTAR GAMBAR

| | |
|---------------------------------------|----|
| Gambar 2. 1 Ilustrasi Proses ROS..... | 12 |
| Gambar 2. 2 Ilustrasi Proses RUS..... | 13 |

DAFTAR LAMPIRAN

Lampiran 1. Rata-rata Curah Hujan Tahunan di Indonesia Periode 1991-2020... 50

BAB I

PENDAHULUAN

1.1 Latar Belakang

Hujan merupakan bentuk tetesan air yang jatuh dari langit ke bumi. Curah hujan menjadi salah satu parameter meteorologi yang paling berpengaruh dalam berbagai aspek kehidupan sehari-hari (Barrera-Animas *et al.* 2022). Curah hujan diukur menggunakan alat penakar hujan. Ketinggian air hujan diukur di suatu daerah datar, tidak menyerap, tidak meresap dan mengalir. Hujan memiliki dampak secara langsung terhadap berbagai industri seperti pertanian, pengelolaan sumber daya air, produksi pembangkit listrik tenaga air, dan lain-lain. Hal-hal tersebut berkaitan erat dengan bidang politik, sosial, ekonomi, dan kesehatan daerah tersebut (Dawoodi and Patil, 2023).

Indonesia memiliki karakteristik curah hujan tinggi antara 460,5 hingga 4.627,4 milimeter (Ruhiat, 2022 *dalam* Statistik Indonesia, 2021). Kondisi ini menjadikan Indonesia rentan terhadap bencana banjir. Menurut Badan Nasional Penanggulangan Bencana (BNPB) suatu peristiwa disebut banjir ketika terdapat peningkatan volume air ke suatu daerah atau daratan yang mengakibatkan daerah tersebut tenggelam. Curah hujan yang tinggi, debit air yang melebihi kapasitas drainase, minimnya wilayah penyerapan air, penebangan liar, membuang sampah ke sungai dan berbagai faktor lainnya berkontribusi terhadap terjadinya banjir (Sholihah *et al.* 2020).

Secara topografi, sebagian besar wilayah Kota Palembang berupa rawa dan dataran rendah. Kota Palembang dikelilingi bukit-bukit rendah dan dialiri sungai-sungai besar maupun kecil. Curah hujan tahunan di Kota Palembang termasuk dalam kategori sedang berkisar antara 2500 hingga 3000 milimeter per tahun (BMKG, 2022). Kondisi tanah yang rendah serta dialiri sungai-sungai besar maupun kecil membuat Kota Palembang rentan terjadi banjir saat terjadinya hujan. Diperkirakan terdapat 68 sungai kecil yang berada di Seberang Ilir berfungsi sebagai drainase daerah perkotaan (Badan Perencanaan Pembangunan Daerah Penelitian dan Pengembangan, 2019).

Mengklasifikasikan kejadian hujan secara tepat merupakan hal yang penting dalam salah satu upaya meminimumkan dampak yang ditimbulkan oleh hujan di wilayah Kota Palembang. Eksistensi rawa yang berfungsi sebagai sumber air, penampungan air serta pengendalian banjir memiliki dampak terhadap perekonomian masyarakat dalam pertanian, perikanan, peternakan, kehutanan, industri dan pariwisata Kota Palembang sebagaimana diatur dalam Peraturan daerah nomor 5 tahun 2008 tentang Pembinaan dan Retribusi Pengendalian dan Pemanfaatan Rawa pasal 3 dan 8. Selain itu upaya konservasi dan pelestarian sungai menjadi faktor yang perlu diperhatikan. Data historis dapat dianalisis menggunakan metode *machine learning* dalam *data mining* sehingga membutuhkan *training data* dan *testing data* untuk memperoleh dan mengevaluasi hasil klasifikasi (Pawluszek-Filipiak and Borkowski, 2020). Pada penelitian ini digunakan teknik klasifikasi untuk memprediksi kejadian hujan di Kota Palembang.

Pada tahap klasifikasi kejadian hujan Kota Palembang, penggunaan data historis menjadi sangat penting. Namun, data yang digunakan sering ditemukan adanya *missing data* (data yang tidak lengkap) dan *imbalanced data* (data tidak seimbang). Masalah data yang hilang dapat menyebabkan berbagai masalah seperti penurunan performa, masalah analisis data dan adanya bias antara data yang lengkap dengan data yang memiliki *missing values* (Ayilara *et al.* 2019). *Imbalanced data* dapat menyebabkan bias terhadap kelas mayoritas bahkan kemungkinan terburuk tidak menggambarkan kelas minoritas (Johnson and Khoshgoftaar, 2019). Distribusi kelas yang persentasenya lebih kecil dari kelas lainnya disebut kelas minoritas, sedangkan kelas yang persentasenya besar disebut kelas mayoritas (Ustyannie and Suprpto, 2020). Upaya mengatasi permasalahan yang dihadapi dalam penelitian ini dilakukan dengan metode *K-Nearest Neighbors* (KNN) untuk mengatasi data yang hilang. *Synthetic Minority Oversampling Technique* (SMOTE), *Random Oversampling* (ROS) dan *Random Undersampling* (RUS) diaplikasikan untuk mengatasi *imbalanced data*.

Naïve Bayes Classifier (NBC) merupakan istilah umum yang sering digunakan untuk menggambarkan metode klasifikasi *Naïve Bayes*. Metode *Naïve Bayes* menggunakan teorema Bayes yang memiliki algoritma simpel dan efisiensi komputasi yang tinggi. Metode *Naïve Bayes* mengasumsikan variabel bersifat independen sehingga tidak memerlukan *covariance matrix* secara keseluruhan. Hal ini membuat NBC memerlukan metode yang lebih sederhana untuk mengestimasi parameter setiap variabel (Chen *et al.* 2021). Berdasarkan penelitian yang dilakukan oleh Prasetya (2020) dibandingkan algoritma *Classification Tree*

dan *K-Nearest Neighbors* (KNN), metode *naïve bayes* mendapatkan hasil terbaik dengan akurasi sebesar 77,1% dalam memprediksi cuaca jangka pendek dan memiliki potensi untuk digunakan secara operasional. Menurut Dawoodi and Patil (2023) diperoleh hasil akurasi algoritma *Super Vector Machine* (SVM) sebesar 93% sedangkan algoritma *Naïve Bayes* sebesar 92%. Sehingga penelitian ini menggunakan metode *Naïve Bayes* untuk mengukur model klasifikasi kejadian hujan di Kota Palembang.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana hasil evaluasi klasifikasi kejadian hujan di Kota Palembang menggunakan metode *Naïve Bayes* saat *missing data* diimputasi dengan KNN pada *imbalanced data*?
2. Bagaimana hasil evaluasi klasifikasi kejadian hujan di Kota Palembang menggunakan metode *Naïve Bayes* saat *missing data* diimputasi dengan KNN dan diperoleh *balanced data* melalui SMOTE, ROS dan RUS?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data cuaca atau kejadian hujan di Kota Palembang diperoleh melalui *website* visualcrossing.com
2. Data yang digunakan mulai dari 01 Januari 2018 sampai 31 Desember 2023 berjumlah 2191 data meliputi 17 variabel independen dan 1 variabel dependen.

3. Data terbagi menjadi *training data* sebanyak 4 tahun dimulai dari 01 Januari 2018 sampai 31 Desember 2021 dan *testing data* sebanyak 2 tahun dimulai dari 01 Januari 2022 sampai 31 Desember 2023. Persentase *training data* sebesar 66,68% sedangkan *testing data* 33,32%.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Memperoleh hasil evaluasi model klasifikasi melalui proses imputasi KNN pada data kejadian hujan di Kota Palembang menggunakan metode *Naïve Bayes*.
2. Memperoleh hasil evaluasi model klasifikasi melalui proses KNN untuk *missing data* dan melalui SMOTE, ROS dan RUS sehingga *balanced*. Selanjutnya mengklasifikasikan data kejadian hujan di Kota Palembang menggunakan metode *Naïve Bayes*.

1.5 Manfaat Penelitian

1. Penelitian ini diharapkan dapat memberikan visualisasi mengenai pentingnya imputasi data terutama menggunakan metode KNN dan mengatasi *missing data* saat kondisi *imbalanced data* terhadap hasil evaluasi model klasifikasi data kejadian hujan menggunakan metode *Naïve Bayes*.
2. Sebagai rujukan bagi peneliti lain dalam mengimplementasikan KNN untuk mengatasi *missing data* serta menerapkan SMOTE, ROS dan RUS sehingga tercapai kondisi *balanced data* melalui hasil evaluasi model klasifikasi data kejadian hujan menggunakan metode *Naïve Bayes*.

DAFTAR PUSTAKA

- Arsyadani, F., & Purwinarko, A. (2023). Implementation of Synthetic Minority Oversampling Technique and Two-phase Mutation Grey Wolf Optimization on Early Diagnosis of Diabetes using K-Nearest Neighbors. *Recursive Journal of Informatics*, 1(1), 9–17.
- Asniar, Maulidevi, N. U., & Surendro, K. (2022). SMOTE-LOF for noise identification in imbalanced data classification. *Journal of King Saud University - Computer and Information Sciences*, 34(6), 3413–3423.
- Ayilara, O. F., Zhang, L., Sajobi, T. T., Sawatzky, R., Bohm, E., & Lix, L. M. (2019). Impact of missing data on bias and precision when estimating change in patient-reported outcomes from a clinical registry. *Health and Quality of Life Outcomes*, 17(1), 1–9.
- AT, E., M, A., F, A. M., & M, S. (2016). Classification of Imbalance Data using Tomek Link (T-Link) Combined with Random Under-sampling (RUS) as a Data Reduction Method. *Global Journal of Technology and Optimization*, 01(S1), 1–11.
- Badan Perencanaan Pembangunan Daerah Penelitian dan Pengembangan. (2019). *Rencana Kerja Pemerintah Daerah (RKPD) Tahun Anggaran 2020*. Palembang.
- Balaji, V. R., Suganthi, S. T., Rajadevi, R., Krishna Kumar, V., Saravana Balaji, B., & Pandiyan, S. (2020). Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naïve Bayes classifier. *Measurement: Journal of the International Measurement Confederation*, 163, 107922.
- Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, 7, 100204.
- Badan Meteorologi, Klimatologi dan Geofisika (BMKG). (2021). Peta Rata-rata Curah Hujan dan Hari Hujan Periode 1991-2020 Indonesia. Pusat informasi perubahan iklim BMKG. Jakarta.
- Chen, H., Hu, S., Hua, R., & Zhao, X. (2021). Improved Naïve Bayes classification algorithm for traffic risk management. *Eurasip Journal on Advances in Signal Processing*, 2021(1), 1–12.
- Dawoodi, H. H., & 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.
- Gautam, R., & Latifi, S. (2023). Comparison of Simple Missing Data Imputation Techniques for Numerical and Categorical Datasets. *Journal of Research in Engineering and Applied Sciences*, 8(1), 468–475.

- Ismail, A. R., Abidin, N. Z., & Maen, M. K. (2022). Systematic Review on Missing Data Imputation Techniques with Machine Learning Algorithms for Healthcare. *Journal of Robotics and Control (JRC)*, 3(2), 143–152.
- Jassim, M. A., & Abdulwahid, S. N. (2021). Data Mining preparation: Process, Techniques and Major Issues in Data Analysis. *IOP Conference Series: Materials Science and Engineering*, 1090(1), 012053.
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1), 1–54.
- Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks. *Applied Sciences (Switzerland)*, 13(6), 1–34.
- Kabade, V., Hooda, R., Raj, C., Awan, Z., Young, A. S., Welgampola, M. S., & Prasad, M. (2021). Machine learning techniques for differential diagnosis of vertigo and dizziness: A review. *Sensors*, 21(22), 1–17.
- Kim, S., Seo, J., & Kim, S. (2024). Machine Learning Technologies in the Supply Chain Management Research of Biodiesel: A Review. *Energies*, 17(6).
- Khan, S. I., & Hoque, A. S. M. L. (2020). SICE: an improved missing data imputation technique. *Journal of Big Data*, 7(1), 1–21.
- Kumar, P., Bhatnagar, R., Gaur, K., & Bhatnagar, A. (2021). Classification of Imbalanced Data: Review of Methods and Applications. *IOP Conference Series: Materials Science and Engineering*, 1099(1), 012077.
- Le, V. T., Quan, N. H., Loc, H. H., Duyen, N. T. T., Dung, T. D., Nguyen, H. D., & Do, Q. H. (2019). A multidisciplinary approach for evaluating spatial and temporal variations in water quality. *Water (Switzerland)*, 11(4), 1–16.
- Mohammed, M. B., Zulkafli, H. S., Adam, M. B., Ali, N., & Baba, I. A. (2021). Comparison of Five Imputation Methods in Handling Missing Data in a Continuous Frequency Table. *AIP Conference Proceedings*. Malaysia.
- Pawluszek-Filipiak, K., & Borkowski, A. (2020). On the importance of train-test split ratio of datasets in automatic landslide detection by supervised classification. *Remote Sensing*, 12(18), 1–33.
- Prasetya, R. (2020). Data Mining Application on Weather Prediction Using Classification Tree, Naïve Bayes and K-Nearest Neighbors Algorithm With Model Testing of Supervised Learning Probabilistic Brier Score, Confusion Matrix and ROC. *Jaict*, 4(2), 25–33.
- Ruhiat, Y. (2022). Forecasting rainfall and potential for repeated events to predict flood areas in Banten province, Indonesia. *Journal of Measurements in Engineering*, 10(2), 68–80.
- Saputra, D. D., Gata, W., Wardhani, N. K., *et al.* (2020). Optimization Sentiments of Analysis from Tweets in myXLCare using Naïve Bayes Algorithm and

- Synthetic Minority over Sampling Technique Method. *Journal of Physics: Conference Series*, 1471(1), 1–7.
- Sasanya, B. F., Awodutire, P. O., Ufuoma, O. G., & Balogun, O. S. (2022). Modelling the Effects of Meteorological Factors on Maximum Rainfall Intensities Using Exponentiated Standardized Half Logistic Distribution. *Journal of Applied Mathematics*, 2022, 1–10.
- Sholihah, Q., Kuncoro, W., Wahyuni, S., Puni Suwandi, S., & Dwi Feditasari, E. (2020). The analysis of the causes of flood disasters and their impacts in the perspective of environmental law. *IOP Conference Series: Earth and Environmental Science*, 437(1), 1–7.
- Sirajul Munir. (2022). Prakiraan Daerah Potensi Banjir Provinsi Sumatera Selatan Periode Januari 2023 – Maret 2023 (Update 10 Desember 2022) [versi elektronik]. Tersedia pada <https://staklim-sumsel.bmkg.go.id/>. Diakses pada 17 Desember 2023.
- Supriyadi, D., Purwanto, P., & Warsito, B. (2023). Optimizing Neural Networks for Academic Performance Classification Using Feature Selection and Resampling Approach. *Mendel*, 29(2), 261–272.
- Tan, H. (2021). Machine Learning Algorithm for Classification. *Journal of Physics: Conference Series*, 1994(1), 1–5.
- Tumusiime, A. G., Eyobu, O. S., Mugume, I., & Oyana, T. J. (2023). A weather features dataset for prediction of short-term rainfall quantities in Uganda. *Data in Brief*, 50(October), 109613.
- Ustyannie, W., & Suprpto, S. (2020). Oversampling Method To Handling Imbalanced Datasets Problem In Binary Logistic Regression Algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14(1), 1–10.
- Wongvorachan, T., He, S., & Bulut, O. (2023). A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. *Information (Switzerland)*, 14(1).