

**PENERAPAN *SYNTHETIC MINORITY OVER-SAMPLING
TECHNIQUE* PADA PENGKLASIFIKASIAN KEJADIAN
HUJAN KOTA PRABUMULIH MENGGUNAKAN
METODE *FUZZY NAÏVE BAYES* DAN
*K-NEAREST NEIGHBOR***

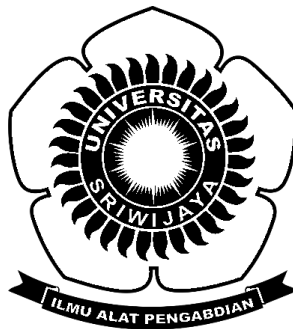
SKRIPSI

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

Oleh :

VIVI CLARA DITA

08011382025116



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

2024

LEMBAR PENGESAHAN

**PENERAPAN *SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE*
PADA PENGKLASIFIKASIAN KEJADIAN HUJAN KOTA
PRABUMULIH MENGGUNAKAN METODE *FUZZY NAÏVE BAYES* DAN
*K-NEAREST NEIGHBOR***

SKRIPSI

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

Oleh

**VIVI CLARA DITA
NIM.08011382025116**

Pembimbing Kedua



**Drs. Endro Setyo Cahyono, M.Si
NIP. 196409261990021002**

**Indralaya, 21 Mei 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**

PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan dibawah ini :

Nama : Vivi Clara Dita
NIM : 08011382025116
Fakultas/Jurusan : MIPA/Matematika

Menyatakan bahwa skripsi 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 dalam 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.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Indralaya, 27 Mei 2024



Vivi Clara Dita

LEMBAR PERSEMBAHAN

“Maka, ingatlah kepada-Ku, Aku pun akan ingat kepadamu. Bersyukurlah kepada-Ku dan janganlah kamu ingkar kepada-Ku” - QS. Al Baqarah (152)

“Do the things you want to do. Do the things that you think are right.

Do the things you love. ”

-J.J.H-

“Life is a journey to be experienced, not a problem to be solved”

~R.W.S~

Kupersembahkan skripsi ini kepada :

♥ Allah SWT.

♥ Papa dan Mama Tercinta

♥ Kakakku dan Adikku Tersayang

♥ Keluarga Besarku Terkasih

♥ Semua Guru dan Dosenku

♥ Sahabat-Sahabatku

♥ Almamaterku

KATA PENGANTAR

Puji syukur ucapkan atas kehadiran Allah SWT dengan berkat rahmat, karunia, dan hidayah-Nya penulis dapat menyelesaikan skripsi sebagai salah satu syarat memperoleh gelar Sarjana Sains bidang Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (MIPA) Universitas Sriwijaya dengan judul “**Penerapan *Synthetic Minority Over-Sampling Technique* Pada pengklasifikasian Kejadian Hujan Kota Prabumulih Menggunakan Metode *Fuzzy Naïve Bayes* Dan *K-Nearest Neighbor*”**”. Shalawat serta salam selalu turunkan kepada Rasulullah SAW, keluarga, sahabat, dan pengikutnya.

Penulis menyadari dengan segala kerendahan dan ketulusan hati bahwa penyusunan skripsi ini tidak akan berhasil tanpa adanya dukungan, bimbingan, bantuan, dan kerjasama dari pihak lain. Penulis mengucapkan terima kasih yang sebesar-besarnya kepada kedua orang tua tercinta, yaitu **Papa IPDA Akhmad Meidi** dan **Mama Eni Juharia, SKM** yang telah merawat, membesarkan, dan mendidik dengan penuh cinta dan kasih sayang. Terima kasih untuk kesabaran dan kebesaran hati dalam menghadapi penulis. Terima kasih atas segala pengorbanan, doa, nasihat, dan dukungan yang diberikan dari penulis kecil hingga saat ini.

Dengan kerendahan hati dan rasa hormat, penulis mengucapkan terima kasih kepada yang terhormat:

1. Ibu **Dr. Dian Cahyawati Sukanda, S.Si., M.Si.**, selaku Dosen Pembimbing Akademik sekaligus Ketua Jurusan Matematika Fakultas Matematika dan Ilmu

Pengetahuan Alam Universitas Sriwijaya yang telah menasehati, membimbing, memotivasi dan memberikan arahan selama masa perkuliahan.

2. Ibu **Des Alwine Zayanti, S.Si., M.Si.**, selaku Sekretaris Jurusan Matematika sekaligus Dosen Pembimbing Utama yang telah meluangkan waktu, pikiran, dan tenaga ditengah kesibukan, untuk membimbing, menasehati, memberi saran serta mengarahkan penulis hingga skripsi ini dapat selesai dengan baik.
3. Bapak **Drs. Endro Setyo Cahyono, M.Si.**, selaku Dosen Pembimbing Kedua yang telah meluangkan waktu ditengah kesibukan, pikiran, dan tenaga untuk membimbing, menasehati, memberi saran serta mengarahkan penulis hingga skripsi ini dapat terselesaikan dengan baik.
4. Ibu **Dr. Ir. Herlina Hanum, M.Si.**, dan Ibu **Endang Sri Kresnawati, S.Si., M.Si.**, selaku Dosen Penguji yang telah bersedia meluangkan waktu, memberikan tanggapan, saran, dan masukkan yang bermanfaat kepada penulis untuk perbaikan penyelesaian skripsi ini.
5. Ibu **Prof. Yulia Resti, S.Si., M.Si., Ph.D.**, selaku Ketua Pelaksana dan Ibu **Dr. Evi Yuliza, S.Si., M.Si.**, selaku Sekretaris Pelaksana yang telah bersedia meluangkan waktu ditengah kesibukan dalam seminar penulis.
6. **Semua Dosen di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya** atas semua ilmu yang bermanfaat, bimbingan, dan arahan untuk penulis selama masa perkuliahan.
7. Pak **Irwansyah** dan Ibu **Hamidah** selaku Admin dan Pegawai Tata Usaha di Jurusan Matematika yang telah membantu penulis dalam segala administrasi sehingga dapat menyelesaikan skripsi ini dengan lancar.

8. Kakak dan Adik penulis tersayang **Shinta Dista Wulan Pratiwi** dan **Thirza Aziel Abhista Nabil** serta semua keluarga besar atas dukungan dan motivasi hingga terselesaikannya skripsi ini.
9. Sahabat penulis **Inez Vermatasari, Verti Mona, Safta Dwi S, Dhiya Qatrunnada, Miska Nurulita, Eggyana, Desfa, Pasma, Indah, Astri, Natasya, Wahyuni, Yolana, Nanda, Fia, Jaehyun, Justin, Rafael, Nathan, Elkan,** dan **Ivar**, atas dukungan, motivasi, dan semua bantuan yang telah diberikan selama masa perkuliahan dari awal hingga selesainya skripsi ini.
10. Sahabat dan teman seperjuangan yang tersayang **Chindy Putri Army** selaku tim penelitian atas kerja sama dukungan, motivasi, dan bantuannya dalam penyelesaian skripsi serta semua bantuan yang telah diberikan selama masa perkuliahan.
11. Semua pihak yang terlibat dan turut serta membantu dalam menyelesaikan skripsi ini yang tidak dapat penulis sebutkan satu persatu.

Penulis berharap agar skripsi ini dapat bermanfaat bagi semua pihak yang membutuhkan terutama mahasiswa Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Indralaya, Mei 2024

Penulis

**APPLICATION OF SYNTHETIC MINORITY OVER-SAMPLING
TECHNIQUE IN THE CLASSIFICATION OF RAINFALL EVENTS IN
PRABUMULIH CITY USING FUZZY NAÏVE BAYES AND
K-NEAREST NEIGHBOR METHODS**

By

**VIVI CLARA DITA
NIM.08011382025116**

ABSTRACT

Imbalanced class data can affect classification performance. An imbalanced class occurs when the minority class is less than the majority class. Imbalanced class can be overcome by resampling one of them using synthetic minority over-sampling technique (SMOTE). This research aims to balance the class by applying SMOTE to the Prabumulih city rainfall event classification using the Fuzzy Naïve Bayes and K-Nearest Neighbor methods. The data used is rainfall event data in Prabumulih city sourced from the visual crossing website with 17 variables, totalling 2556 data. Train data is 71,44% or 1826 data and test data is 28,56% or 730 data. The accuracy of rainfall event classification before the application of SMOTE using Fuzzy Naïve Bayes method resulted in accuracy, precision, recall, and f-score of 73,42%, 71,43%, 65,63%, and 68,40%, respectively. After the application of SMOTE using the Fuzzy Naïve Bayes method, the accuracy and precision decreased by 2,6% and 7,46%. In contrast, recall and f-score increased by 10,93% and 1,3%. Meanwhile, before the application of SMOTE using the K-Nearest Neighbor method, the accuracy, precision, recall, and f-score were 69,45%, 75,66%, 44,68%, and 69,93%. After the application of SMOTE using the K-Nearest Neighbor method, the accuracy, recall, and f-score increased by 3,84%, 26,57%, and 0,12%, respectively. On the other hand, precision decreased by 6,78%. The application of SMOTE using the K-Nearest Neighbor method has a better classification accuracy, seen from the accuracy, recall, and f-score values that increase compared to the application of SMOTE using the Fuzzy Naïve Bayes method.

Keywords : Imbalanced class, SMOTE, Fuzzy Naïve Bayes, K-Nearest Neighbor.

**PENERAPAN *SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE*
PADA PENGKLASIFIKASIAN KEJADIAN HUJAN KOTA
PRABUMULIH MENGGUNAKAN METODE *FUZZY NAÏVE BAYES* DAN
*K-NEAREST NEIGHBOR***

Oleh

**VIVI CLARA DITA
NIM.08011382025116**

ABSTRAK

Data kelas tidak seimbang dapat mempengaruhi kinerja klasifikasi. Kelas tidak seimbang atau *imbalanced class* terjadi saat kelas minoritas lebih sedikit dari kelas mayoritas. *Imbalanced class* dapat diatasi dengan resampling salah satunya menggunakan *synthetic minority over-sampling technique* (SMOTE). Penelitian ini bertujuan untuk *balance class* dengan menerapkan SMOTE pada pengklasifikasian kejadian hujan kota Prabumulih menggunakan metode *Fuzzy Naïve Bayes* dan *K-Nearest Neighbor*. Data yang digunakan merupakan data kejadian hujan di kota Prabumulih bersumber dari website visual crossing dengan 17 variabel, sebanyak 2556 data. Data *train* sebesar 71,44% atau sebanyak 1826 data dan data *test* sebesar 28,56% atau sebanyak 730 data. Tingkat ketepatan klasifikasi kejadian hujan sebelum penerapan SMOTE menggunakan metode *Fuzzy Naïve Bayes* menghasilkan *accuracy*, *precision*, *recall*, dan *f-score* secara terurut sebesar 73,42%, 71,43%, 65,63%, dan 68,40%. Setelah penerapan SMOTE menggunakan metode *Fuzzy Naïve Bayes* mengalami penurunan pada *accuracy* dan *precision* sebesar 2,6% dan 7,46%. Sebaliknya, *recall* dan *f-score* mengalami peningkatan sebesar 10,93% dan 1,3%. Sedangkan, sebelum penerapan SMOTE menggunakan metode *K-Nearest Neighbor* menghasilkan *accuracy*, *precision*, *recall*, *f-score* sebesar 69,45%, 75,66%, 44,68%, dan 69,93%. Setelah penerapan SMOTE menggunakan metode *K-Nearest Neighbor* menghasilkan secara berurut *accuracy*, *recall*, dan *f-score* yang meningkat sebesar 3,84%, 26,57%, dan 0,12%. Sebaliknya pada *precision* mengalami penurunan sebesar 6,78%. Penerapan SMOTE menggunakan metode *K-Nearest Neighbor* mempunyai tingkat ketepatan klasifikasi yang lebih baik, dilihat dari nilai *accuracy*, *recall*, dan *f-score* yang meningkat dibandingkan penerapan SMOTE menggunakan metode *Fuzzy Naïve Bayes*.

Kata Kunci : *Imbalanced class*, SMOTE, *Fuzzy Naïve Bayes*, *K-Nearest Neighbor*.

DAFTAR ISI

HALAMAN JUDUL	
LEMBAR PENGESAHAN	ii
LEMBAR PERSEMBAHAN	iii
KATA PENGANTAR	iv
ABSTRACT	vii
ABSTRAK	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 Kejadian Hujan.....	7
2.2 <i>Statistical Machine Learning</i>	7
2.3 Klasifikasi.....	8
2.4 <i>Synthetic Minority Over-Sampling Technique (SMOTE)</i>	8
2.5 Probabilitas	9
2.6 Metode <i>Naïve Bayes</i>	10
2.7 <i>Laplace Smoothing</i>	11
2.8 Himpunan <i>Fuzzy</i>	12
2.9 Fungsi Keanggotaan	13
2.10 Metode <i>Fuzzy Naïve bayes</i>	16
2.11 Normalisasi.....	17
2.12 Metode <i>K-Nearest Neighbor (K-NN)</i>	17
2.13 <i>Confusion Matrix</i>	19
BAB III METODOLOGI PENELITIAN	21

3.1	Tempat	21
3.2	Waktu.....	21
3.3	Alat	21
3.4	Metode Penelitian.....	21
3.4.1	Sumber Data.....	21
3.4.2	Variabel Penelitian	21
3.4.3	Metode Analisis.....	22
BAB IV HASIL DAN PEMBAHASAN.....		25
4.1	Deskripsi Data	25
4.2	Menentukan Himpunan Universal <i>Fuzzy</i>	26
4.3	Menentukan Nilai Keanggotaan <i>Fuzzy</i>	31
4.4	Partisi Data Nilai Keanggotaan <i>Fuzzy</i>	35
4.5	Pengklasifikasian Metode <i>Fuzzy Naïve Bayes</i>	35
4.5.1	Menghitung Probabilitas <i>Prior, Likelihood, dan Posterior</i>	35
4.5.2	<i>Confusion Matrix</i> Metode <i>Fuzzy Naïve Bayes</i>	40
4.6	Pengklasifikasian Metode <i>Fuzzy Naïve Bayes</i> dengan SMOTE	42
4.7.1	Penerapan SMOTE Pada Data Train.....	42
4.7.2	Menghitung Probabilitas <i>Prior, Likelihood, dan Posterior</i>	43
4.7.3	<i>Confusion Matrix</i> Metode <i>Fuzzy Naïve Bayes</i> dengan SMOTE.....	47
4.7	Metode <i>K-Nearest Neighbor</i>	49
4.7.1	Perhitungan Metode <i>K-Nearest Neighbor</i>	49
4.7.2	Normalisasi <i>Dataset</i> Penelitian Metode <i>K-Nearest Neighbor</i> menggunakan <i>Machine Learning</i>	52
4.7.3	Evaluasi <i>Dataset</i> Penelitian Metode <i>K-Nearest Neighbor</i> menggunakan <i>Machine Learning</i>	54
4.7.4	Evaluasi <i>Dataset</i> Penelitian dengan penerapan SMOTE Metode <i>K-Nearest Neighbor</i> menggunakan <i>Machine Learning</i>	55
4.8	Analisis Hasil	58
BAB V KESIMPULAN DAN SARAN		60
5.1	Kesimpulan.....	60
5.2	Saran.....	61
DAFTAR PUSTAKA.....		62
LAMPIRAN.....		65

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	19
Tabel 4.1 Variabel Data	25
Tabel 4.2 Dataset Penelitian	26
Tabel 4.3 Interval Himpunan Fuzzy Variabel X_2	27
Tabel 4.4 Himpunan Fuzzy Variabel X_2	27
Tabel 4.5 Interval Himpunan Fuzzy Fase Bulan X_{16}	29
Tabel 4.6 Himpunan Fuzzy Variabel X_{16}	29
Tabel 4.7 Nilai Keanggotaan Terbesar Variabel X_2	32
Tabel 4.8 Nilai Keanggotaan Terbesar Variabel X_{16}	33
Tabel 4.9 Nilai Keanggotaan Terbesar Semua Variabel Prediktor	34
Tabel 4.10 Data Train Fuzzy	35
Tabel 4.11 Data Test Fuzzy	35
Tabel 4.12 Hasil Perhitungan Nilai Probabilitas Likelihood	37
Tabel 4.13 Hasil Posterior Fuzzy Naïve Bayes	40
Tabel 4.14 Hasil Confusion Matrix Metode Fuzzy Naïve Bayes.....	40
Tabel 4.15 Hasil SMOTE	42
Tabel 4.16 Hasil Perhitungan Nilai Probabilitas Likelihood	44
Tabel 4.17 Hasil Posterior Fuzzy Naïve Bayes dengan SMOTE.....	47
Tabel 4.18 Hasil Confusion Matrix Metode Fuzzy Naïve Bayes dengan SMOTE.....	47
Tabel 4.19 Dataset Metode K-NN.....	49
Tabel 4.20 Hasil Normalisasi Metode K-NN	50
Tabel 4.21 Hasil SMOTE Metode K-NN.....	50
Tabel 4.22 Klasifikasi Data ke-13	52
Tabel 4.23 Normalisasi Dataset Penelitian Metode K-NN	52
Tabel 4.24 Data Train Metode K-NN.....	53
Tabel 4.25 Data Test Metode K-NN.....	53
Tabel 4.26 Hasil Confusion Matrix Metode K-NN.....	54
Tabel 4.27 Hasil SMOTE Data Penelitian Metode K-NN	55
Tabel 4.28 Hasil Confusion Matrix Metode K-NN dengan SMOTE.....	56
Tabel 4.29 Analisis Hasil.....	58

DAFTAR GAMBAR

Gambar 2.1 Representasi Kurva Segitiga	14
Gambar 2.2 Kurva-S Penyusutan.....	15
Gambar 2.3 Kurva-S Pertumbuhan.....	16

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada pengklasifikasian menggunakan *machine learning* terdapat permasalahan distribusi kelas tidak seimbang yang dapat mempengaruhi kinerja klasifikasi. *Imbalance class dataset* terjadi pada saat kelas minoritas lebih sedikit dari kelas mayoritas. Model yang dibuat dengan menggunakan data yang tidak seimbang akan menghasilkan tingkat ketepatan klasifikasi pada minoritas yang rendah sehingga dapat menyebabkan sistem klasifikasi memiliki batas-batas keputusan yang bias (Ambarsari, 2020). *Imbalanced class dataset* dapat diatasi dengan metode *sampling* atau *resampling*, salah satunya menggunakan SMOTE (*Synthetic Minority Over-Sampling Technique*) (Syukron & Subekti, 2018).

Penelitian terdahulu oleh Hunafa & Hermawan, (2023) berjudul “*Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Imbalance Class Dataset Penyakit Diabetes*” mendapatkan hasil akurasi, presisi, *recall*, dan *f-score* metode Naïve Bayes dengan SMOTE berurut sebesar 71,66%, 32,27%, 72,66%, dan 44,7%, dan Naïve Bayes tanpa SMOTE sebesar 76,03%, 34,42%, 57,58%, dan 43,08%. Sedangkan, metode *K-Nearest Neighbor* dengan SMOTE sebesar 80,47%, 37,93%, 37,57%, dan 37,75%, dan metode *K-Nearest Neighbor* tanpa SMOTE sebesar 83,02%, 43,14%, 24,4%, dan 31,17%.

Penelitian terdahulu oleh Kasanah *et al.*, (2019) berjudul “*Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN*” dengan menerapkan nilai $k=1$ sampai

dengan $k=10$ mendapatkan hasil akurasi yang terbaik pada nilai $k=9$ sebesar 88%. Penelitian lainnya berjudul “*Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN*” dengan menerapkan $k=1$ mendapatkan hasil akurasi sebesar 65,92% dan $k=5$ sebesar 57,41% (Nasution, Khotimah, & Chamidah, 2019).

Metode *Fuzzy Naïve Bayes* dan *K-Nearest Neighbor* merupakan salah satu metode klasifikasi dari *Statistical Machine Learning*. Metode *Fuzzy Naïve Bayes* merupakan kombinasi dari himpunan *Fuzzy* dan *Naive Bayes* di dalam perhitungan. Himpunan *Fuzzy* bekerja menggunakan derajat keanggotaan dari sebuah nilai yang kemudian digunakan untuk menentukan hasil yang diharapkan, berdasarkan spesifikasi yang telah ditentukan (Sastrawan *et al.*, 2019). *K-Nearest Neighbor* merupakan metode yang digunakan dalam klasifikasi terhadap data *train* yang memiliki jarak terdekat dengan data *test*. Prediksi pada metode *K-Nearest Neighbor* berdasarkan mayoritas kelas dari tetangga terdekat tersebut.

Penelitian terdahulu oleh Putri *et al.*, (2021) berjudul “*Sentiment Analysis using Fuzzy Naïve Bayes Classifier on Covid-19*” mendapatkan hasil akurasi sebesar 83,1%. Penelitian lainnya yang dilakukan oleh oleh Antofa (2023) tentang “*Komparasi Metode Fuzzy Naïve Bayes dan Deep Neural Network Untuk Klasifikasi Indeks Harga Saham Gabungan (IHSG) Bursa Efek Indonesia Tahun 2012-2022*” *fuzzy naïve bayes* mendapatkan akurasi sebesar 57,32%.

Kota Prabumulih dengan kondisi iklim tropisnya yang mendukung perkebunan sepanjang tahun, salah satunya pada perkebunan nanas. Hasil produksi dari perkebunan nanas di kota Prabumulih menduduki tempat terbanyak ketiga

setelah Ogan Ilir dan Muara Enim. Berdasarkan data pada Badan Pusat Statistik, luas lahan panen nanas di kota Prabumulih mencapai sekitar 18,110 Ha dengan jumlah produksi 1.969 ton pada tahun 2022.

Kejadian hujan sangat berpengaruh terhadap pertumbuhan buah nanas. Nanas dapat tumbuh dengan baik di daerah dataran rendah yang memerlukan hujan dengan kelembaban yang cukup selama lebih dari enam bulan dalam satu tahun (Srilestari & Suwardi, 2021). Kejadian hujan yang sering dengan curah tinggi dapat mempengaruhi area lahan nanas, dimana akan membuat lahan tergenang air yang menyebabkan tanaman nanas mudah terkena hama dan penyakit bahkan dapat menyebabkan gagal panen.

Pentingnya klasifikasi kejadian hujan yang tepat sangat berpengaruh terhadap sektor pertanian. Kejadian hujan memiliki pola yang tidak pasti sehingga akan sulit diprediksi secara manual, tetapi hal ini dapat diperkirakan menggunakan *statistical machine learning* dengan metode klasifikasi (Nurmaulida *et al.*, 2023). Klasifikasi merupakan pengelompokkan objek atau data baru ke dalam kelas berdasarkan atribut tertentu dengan melihat variabel dari kelompok data yang ada (Nasution *et al.*, 2019).

Berdasarkan hasil dari penelitian terdahulu menggunakan metode *Fuzzy Naïve Bayes* dan *K-Nearest Neighbor*, maka penulis bertujuan untuk membandingkan ketepatan nilai klasifikasi dengan menerapkan SMOTE pada data kejadian hujan menggunakan metode *Fuzzy Naïve Bayes* dan *K-Nearest Neighbor*.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan pada penelitian ini, maka rumusan masalah yang akan peneliti bahas dalam penelitian ini :

1. Berapa tingkat ketepatan klasifikasi dalam memprediksi kejadian hujan sebelum dan sesudah penerapan SMOTE menggunakan metode *Fuzzy Naïve Bayes* dan apakah dengan penerapan SMOTE pada metode *Fuzzy Naïve Bayes* akan meningkatkan tingkat ketepatan klasifikasi?
2. Berapa tingkat ketepatan klasifikasi dalam memprediksi kejadian hujan sebelum dan sesudah penerapan SMOTE menggunakan metode *K-Nearest Neighbor* dan apakah dengan penerapan SMOTE pada metode *K-Nearest Neighbor* akan meningkatkan tingkat ketepatan klasifikasi?
3. Metode mana yang lebih baik setelah penerapan SMOTE pada pengklasifikasian data kejadian hujan menggunakan metode *Fuzzy Naïve Bayes* dan *K-Nearest Neighbor*?

1.3 Batasan Masalah

Batasan masalah yang akan peneliti bahas dalam penelitian ini :

1. Data yang digunakan pada penelitian ini ialah data kejadian hujan Kota Prabumulih, Sumatera Selatan selama 7 tahun yaitu pada tahun 2017-2023 dari website <https://www.visualcrossing.com/weather/weather-data-services>.
2. Data yang digunakan terdiri dari, 16 variabel prediktor bertipe data numerik yang terdiri dari *temperature* minimum, *temperature* maksimum, *temperature*, *feels like* minimum, *feels like* maksimum, *feels like*, embun,

kelembaban udara, kecepatan angin, arah angin, lapisan awan, visibilitas, penyinaran surya, *solar energy*, *moonphase*, dan Uv Index, serta 1 variabel respon bertipe data kategori, yaitu kejadian hujan, pada kategori 0 merupakan tidak hujan dan kategori 1 merupakan hujan.

3. Data yang diobservasi sejumlah 2556 data, dimana data *train* sebesar 71,44% dari tanggal 1 Januari 2017 sampai dengan 31 Desember 2021 dengan data sebanyak 1826 data dan data *test* sebesar 28,56% dari tanggal 1 Januari 2022 sampai dengan 31 Desember 2023 dengan data sebanyak 730 data.
4. Ketepatan klasifikasi dibatasi dengan nilai *accuracy*, *precision*, *recall*, dan *f-score*.

1.4 Tujuan Penelitian

Tujuan penelitian ini, bertujuan untuk :

1. Memperoleh tingkat ketepatan klasifikasi kejadian hujan sebelum dan sesudah penerapan SMOTE pada metode *Fuzzy Naïve Bayes*.
2. Memperoleh tingkat ketepatan klasifikasi kejadian hujan sebelum dan sesudah penerapan SMOTE pada metode *K-Nearest Neighbor*.
3. Memperoleh metode terbaik dengan penerapan SMOTE pada pengklasifikasian data kejadian hujan menggunakan metode *Fuzzy Naïve Bayes* dan *K-Nearest Neighbor*.

1.5 Manfaat Penelitian

Manfaat pada penelitian ini :

1. Media pembelajaran dalam mengklasifikasi kejadian hujan.

2. Menjadi bahan referensi untuk peneliti yang menggunakan himpunan fuzzy dengan fungsi keanggotaan kurva sigmoid dan kurva segitiga.
3. Menjadi bahan referensi untuk peneliti lain yang berhubungan dengan penerapan SMOTE pada pengklasifikasian kejadian hujan menggunakan metode *Fuzzy Naïve Bayes* dan *K-Nearest Neighbor*.

DAFTAR PUSTAKA

- Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning. *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)*, 2(1), 1–6.
- Ambarsari, D. (2020). *Metoda Random Over-Under Sampling Dan Random Forest Untuk Klasifikasi Prediksi Stroke*. Tesis. Universitas Nusamandiri
- Ambarwari, A., Adrian, Q. J., & Herdiyeni, Y. (2020). Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman. *JURNAL RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(1), 117–122. <http://jurnal.iaii.or.id>
- Anas, Tempola, F., & Amal, K. (2019). Hybrid fuzzy dan Naive Bayes Dalam Penentuan Status UKT. *Jurnal PROtek, 06 No.01*, 6-11.
- Antofa, Fitriana. (2023). Komparasi Metode *Fuzzy Naïve Bayes* dan *Deep Neural Network* Untuk Klasifikasi Indeks Harga Saham Gabungan (IHSG) Bursa Efek Indonesia Tahun 2012-2022. Skripsi. Universitas Sriwijaya.
- Arabzadeh Jamali, M., & Pham, H. (2023). Statistical Machine Learning. *Springer Handbooks*, 865–886. https://doi.org/10.1007/978-1-4471-7503-2_42
- Astriyani, M., Laela, I. N., Lestari, D. P., Anggraeni, L., & Astuti, T. (2023). Analisis Klasifikasi Data Kualitas Udara Dki Jakarta Menggunakan Algoritma C.45. *JuSiTik : Jurnal Sistem Dan Teknologi Informasi Komunikasi*, 6(1), 36–41. <https://doi.org/10.32524/jusitik.v6i1.790>
- Astuti, F. D., & Lenti, F. N. (2021). Implementasi SMOTE untuk mengatasi Imbalanced Class pada Klasifikasi Car Evolution menggunakan K-NN. *Jurnal JUPITER, Vol.13 No.1*, 89-98.
- G.S, T., Hariprasad, Y., Iyengar, S., Sunitha, N., Badrinath, P., & Chennupati, S. (2022). An extension of Synthetic Minority Oversampling Technique based on Kalman Filter for Imbalanced datasets. *Machine Learning With Application*, 1-12
- Hariyani, & Sugiyanto. (2020). Equal-Width Interval Discretization in Naive Bayes (Case Study : Classification TBC Patients). *Jurnal AdMathEdu*.
- Hasnain, M., Pasha, M. F., Ghani, I., Imran, M., Alzahrani, M. Y., & Budiarto, R. (2020). *Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking*. *IEEE Access*, 8, 90847–90861.

- Hunafa, M. R., & Hermawan, A. (2023). Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor Pada Imbalance Class Dataset Penyakit Diabetes. *Media Online*, 4(3), 1551–1561. <https://doi.org/10.30865/klik.v4i3.1486>
- Juleha, Rismalinda, & Rahmi, A. (2016). Analisa Metode Intensitas Hujan Pada Stasiun Hujan Rokan Iv Koto, Ujung Batu, Dan Tandun Mewakili Ketersediaan Air Di Sungai Rokan. *Jurnal Mahasiswa Teknik UPP*, 1(1), 1–8.
- Kasanah, A. N., Muladi, & Pujiyanto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi *Imbalance Class* dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma K-NN. *Jurnal Rekayasa Sistem dan Teknologi Informasi*, Vol. 3 No.2, 196-201.
- Kurniati, V., Triyanto, D., & Rismawan, T. (2017). Penerapan Logika Fuzzy Dalam Sistem Prakiraan Cuaca Berbasis Mikrokontroler. *Jurnal Coding, Sistem Komputer Untan*, Vol.5, No.2, 119-128.
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal*, 4(1), 78. <https://doi.org/10.24114/cess.v4i1.11458>
- Noto, A. P., & Saputro, D. R. S. (2022). *Classification Data Mining with Laplacian Smoothing on Naïve Bayes Method*. *AIP Conference Proceedings*, 2566. <https://doi.org/10.1063/5.0116519>
- Nurmaulida, I., Sunge, A. S., & Zy, A. T. (2023). Penggunaan Naïve Bayes dalam Implementasi Prediksi Tingkat Curah Hujan. *Jurnal Ilmiah Mahasiswa Pendidikan Sejarah*, 8(3), 3149–3157. <https://jim.usk.ac.id/sejarah>
- Pratama, A. R. I., Latipah, S. A., & Sari, B. N. (2022). Optimasi Klasifikasi Curah Hujan Menggunakan Support Vector Machine (Svm) Dan Recursive Feature Elimination (Rfe). *JIPi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 7(2), 314–324. <https://doi.org/10.29100/jipi.v7i2.2675>
- Putri, Z., Sugiyarto, & Salafudin. (2021). *Sentiment Analysis using Fuzzy Naïve Bayes Classifier on Covid-19*. *Desimal: Jurnal Matematika*, 4(1), 13–20. <https://doi.org/10.24042/djm>
- Rachmaniar, N. S. (2017). Stabilitas Sistem Dinamik Pertumbuhan Sel Kanker Dengan Terapi Radiasi. *Math Unesa*, 3(6), 80-89.
- Salmi, N., & Rustam, Z. (2019). *Naïve Bayes Classifier Models for Predicting the Colon Cancer*. *IOP Conference Series: Materials Science and Engineering*, 546(5). <https://doi.org/10.1088/1757-899X/546/5/052068>

- Sabilla, W. I., & Vista, C. B. (2021). Implementasi SMOTE dan Under Sampling pada Imbalanced Dataset Untuk Prediksi Kebangkrutan Perusahaan. *Jurnal Politeknik Caltex Riau, Vol.7, No. 2*, 329-339.
- Sastrawan, A. S., Gunadi, I. G. A., & Sukajaya, I. N. (2019). Perbandingan Kinerja Algoritma Dempster Shafer Dan Fuzzy-Naive Bayes Dalam Klasifikasi Penyakit Demam Berdarah Dan Tifus. *Jurnal Ilmu Komputer Indonesia, 4(2)*, 24–32. <https://ejournal-pasca.undiksha.ac.id/index.php/jik/article/view/3125>
- Srilestari, R., & Suwardi. (2021). *Pascapanen Nanas*. (Dedi, Ed.) Sleman, Yogyakarta: LPPM UPN "VETERAN" YOGYAKARTA.
- Syukron, A., & Subekti, A. (2018). Penerapan Metode Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit. *Jurnal Informatika, 5(2)*, 175–185. <https://doi.org/10.31311/ji.v5i2.4158>
- Triyanto, S., Sunyoto, A., & Arief, M. R. (2021). Analisis Klasifikasi Bencana Banjir Berdasarkan Curah Hujan Menggunakan Algoritma Naïve Bayes. *JOISIE (Journal Of Information Systems And Informatics Engineering), 5(2)*, 109–117. <https://doi.org/10.35145/joisie.v5i2.1785>
- Walpole, R. E. (2019). *Pengantar Statistika Edisi ke-3* (3 ed.). (B. Sumantri, Trans.) Jakarta: PT Gramedia Pustaka Utama.
- Wicaksono, A. R., & Sulaiman, R. (2020). Operasi dan Relasi Pada Himpunan Fuzzy Intuisionistik. *Jurnal Ilmiah Matematika, 69-78*.