

PERNYATAAN

Dengan ini saya menyatakan skripsi saya yang berjudul "Analisis Imbalance Multiclass Pada Klasifikasi Curah Hujan Dengan Metode SMOTE Neural Network" masih mengandung plagiat dibawah batas yang ditetapkan. Apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.

Bojonegoro, 8 Juli 2024



Wardiana Adinda Putri
NIM.2520200077

KATA PENGANTAR

Segala puji dan syukur kehadirat Allah SWT atas berkah, rahmat dan hidayah-nya yang senantiasa dilimpahkan kepada penulis, sehingga bisa menyelesaikan Proposal Skripsi ini dengan judul “Analisis *Imbalance Multiclass* Pada Klasifikasi Curah Hujan Dengan Metode *SMOTE Neural Network*” sebagai proses awal dalam pembuatan tugas akhir guna menyelesaikan pendidikan program strata satu pada Program Studi Statistika dan Sains Data di Perguruan Tinggi Universitas Nahdlatul Ulama Sunan Giri.

Dalam penyusunan Proposal Skripsi ini banyak hambatan namun pada akhirnya dapat melaluinya berkat ada bimbingan dan bantuan dari berbagai pihak baik secara moral maupun spiritual. Untuk itu pada kesempatan ini peneliti menyampaikan ucapan terima kasih kepada:

1. K. M. Jauharul Ma'arif selaku Rektor Universitas Nahdlatul Ulama Sunan Giri
2. M. Jauhar Vikri, M.Kom. selaku Dekan Fakultas Sains dan Teknologi Universitas Nahdlatul Ulama Sunan Giri
3. Nur Mahmudah, M.Stat. selaku Ketua Program Studi Statistika Sains dan Data Universitas Nahdlatul Ulama Sunan Giri
4. Nita Cahyani, M.Stat. dan Nur Mahmudah, M.Stat. selaku Pembimbing I dan Pembimbing II yang telah banyak memberikan bimbingan dan mengarahkan dalam penyusunan dan penulisan proposal skripsi ini.
5. Alif Yuanita Kartini, M.Si. Selaku Dosen Penasehat Akademi yang telah memberikan pengasahan dan konsultasi akademik dari awal perkuliahan sampai akhir perkuliahan.

Atas segala kekurangan dan ketidak sempurnaan Proposal Skripsi ini, penulis sangat mengharapkan masukan, kritik, dan saran yang bersifat membangun kearah perbaikan dan penyempurnaan skripsi ini agar dalam penyusunan karya tulis selanjutnya dapat lebih baik.

Penulis berharap semoga skripsi ini dapat bermanfaat bagi seluruh pihak dan semoga segala amal baik yang telah diberikan kepada penulis mendapat balasan dari Allah SWT.

Bojonegoro, Juli 2024

Penulis

Wardiana Adinda Putri



HALAMAN PERSETUJUAN

Nama = Wardiana Adinda Putri

NIM = 2520200077

Judul = Analisis *Imbalance Multiclass* pada klasifikasi curah hujan dengan metode *SMOTE Neural Network*

Telah di setujui dan dinyatakan memenuhi syarat untuk diajukan dalam sidang skripsi

Bojonegoro, 24 Juni 2024

Pembimbing I



Nita Cahyani, M.Stat.

NIDN: 0704038906

Pembimbing II



Nur Mahmudah, M. Stat.

NIDN: 0715039201

HALAMAN PENGESAHAN

Nama : Wardiana Adinda Putri

NIM : 2520200077

Judul : Analisis *Imbalance Multiclass* Pada Klasifikasi Curah Hujan dengan
Metode *SMOTE Neural Network*

Telah diujikan dalam sidang skripsi pada tanggal 8 Juli 2024

Dewan Penguji
Penguji I



Alif Yuanita Kartini, M.Si.
NIDN.0721048606

Tim Pembimbing
Pembimbing I



Nita Cahyani, M.Stat.
NIDN.0704038906

Penguji II



Dr. H. Ridlwan Hambali, Lc., M.A.
NIDN.2117056803

Pembimbing II



Nur Mahmudah, M.Stat.
NIDN.0715039201

Mengetahui,
Dekan Fakultas Sains dan Teknologi



M. Jauhar Vikri, M.Kom.
NIDN.0712078803

Mengetahui,
Ketua Program Studi Statistika



Nur Mahmudah, M.Stat.
NIDN.0715039201

FACULTAS
SAINS DAN TEKNOLOGI
UNUGIRI

MOTTO

“Lakukan apapun yang terbaik, maka tuhan akan membantumu menyelesaikannya.”

(Penulis)

“Tidak ada mimpi yang tidak bisa digapai mau setinggi apapun itu jika melibatkan tuhan dalam setiap prosesnya.”

(Penulis)

PERSEMBAHAN

Skripsi ini penulis persembahkan kepada:

1. Seluruh Dosen Statistika yang telah membimbing dari awal perkuliahan berjalan sampai akhir perkuliahan dengan baik.
2. Kepada kedua orang tua saya bapak Wardji dan Ibu Armiyati yang telah memberikan kasih sayang, semangat, nasehat, dukungan dan doa yang tiada terputus sehingga skripsi ini dapat selesai
3. Kepada kakak saya, Ardian Yoga Aritianto yang juga memberika semangat dan motivasi kepada saya.
4. Kepada teman-teman seperjuangan program studi Statistika 2020 khususnya Jami'atul Khoirina, Tisa Dwi Julianti Huda dan Hidayatul Fitri'ah yang ikut berjuang bersama-sama dari awal perkuliahan sampai akhir perkuliahan.

UNUGIRI

ABSTRAK

Putri, W. A. 2024. *Analisis Imbalance Multiclass Pada Klasifikasi Curah Hujan Dengan Metode SMOTE Neural Network*. Skripsi. Jurusan Statistika Fakultas Sains dan Teknologi Universitas Nahdlatul Ulama Sunan Giri Bojonegoro. Pembimbing Utama Nita Cahyani, M.Stat. dan Pembimbing Pendamping Nur Mahmudah, M.Stat. Curah hujan memengaruhi aktifitas manusia, khususnya dalam bercocok tanam maupun kehidupan sehari-hari. Oleh karena itu, sangat penting untuk mengklasifikasikan curah hujan agar kita dapat memprediksi jumlah curah hujan yang akan datang. Penelitian ini menggunakan metode klasifikasi *Multilayer perceptron (MLP)*, salah satu dari algoritma *Neural Network*. Dataset yang diperoleh dari *website BMKG* terjadi *imbalance class* sehingga perlu dilakukan teknik *oversampling SMOTE*. Penelitian ini membandingkan performa MLP tanpa SMOTE dan MLP dengan SMOTE. Berdasarkan hasil analisis, model terbaik yang dihasilkan ketika menggunakan SMOTE, MLP tanpa SMOTE diperoleh nilai akurasi 75%, *sensitivity* 40,34%, *specificity* 86,15% dan nilai AUC 63,25% sedangkan MLP dengan SMOTE diperoleh nilai akurasi 71,27%, *sensitivity* 71,14%, *specificity* 90,30% dan nilai AUC 80,72%. meskipun nilai akurasi mengalami penurunan namun secara keseluruhan evaluasi mengalami peningkatan terutama nilai AUC sehingga dapat disimpulkan bahwa teknik SMOTE mampu mengatasi masalah *imbalance class*.

Kata kunci: *curah hujan, klasifikasi, imbalance class, multilayer perceptron, SMOTE*

UNUGIRI

ABSTRACT

Putri, W. A. 2024. Multiclass Imbalance Analysis in Rainfall Classification Using the SMOTE Neural Network Method. Thesis. Department of Statistics, Science and Technology, University Nahdlatul Ulama Sunan Giri Bojonegoro. Main Supervisor Nita Cahyani, M.Stat. and Assistant Supervisor Nur Mahmudah, M.Stat. Rainfall affects human activities, especially in farming and daily life. For that reason, it is important to classify rainfall so we can predict the amount of rainfall to come. This study in a multilayer perceptron (MLP) classification, one of an algorithm neural network. Dataset obtained from bmkg website occurs imbalance class so it is necessary to do oversampling technique SMOTE. This study compared the performance of MLP without SMOTE and MLP with SMOTE. Based on analysis results, MLP without SMOTE got accuracy value 75%, sensitivity 40,34%, specificity 86,15% and AUC 63,25% while MLP with SMOTE acquired accuracy value 71,27%, sensitivity 71,14%, specificity 90,30% and AUC 80,72%. Despite the accuracy dropped but the whole evaluation AUC increased especially of technique so that it can be concluded that SMOTE able to overcome the problems of imbalance class.

Keywords: Rainfall, classifications, imbalance class, multilayer perceptron, smote

UNUGIRI

DAFTAR ISI

	Halaman
PERNYATAAN	i
KATA PENGANTAR	ii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN	v
MOTTO	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	6
2.1 Tinjauan Pustaka.....	6
2.2 Dasar Teori	11
2.2.1 Curah Hujan	11
2.2.2 Faktor Yang Mempengaruhi Curah Hujan	12
2.2.3 Geografis Kabupaten Tuban.....	13
2.2.4 Statistik Deskriptif	14

2.2.5	Standarisasi Data.....	14
2.2.6	<i>Feature Selection</i>	15
2.2.7	<i>Neural network</i>	15
2.2.8	<i>Multilayer Perceptron (MLP)</i>	20
2.2.9	<i>Imbalance Class</i>	21
2.2.10	<i>Synthetic Minority Oversampling Technique (SMOTE)</i>	22
2.2.11	Klasifikasi	24
2.2.12	Cross Validation	24
2.2.13	Evaluasi ketepatan klasifikasi	25
BAB III METODE PENELITIAN		29
3.1	Sumber Data	29
3.2	Variabel Penelitian.....	29
3.3	Langkah-langkah Penelitian.....	30
3.3.1	Langkah-langkah metode <i>Multilayer Perceptron</i> tanpa SMOTE ...	31
3.3.2	Langkah-langkah metode <i>Multilayer Perceptron</i> dengan SMOTE	32
BAB IV HASIL DAN PEMBAHASAN.....		35
4.1	Pra-Pemrosesan Data.....	35
4.2	Statistik Deskriptif.....	37
4.2.1	Temperatur Minimum	37
4.2.2	Temperatur Maksimum	37
4.2.3	Temperatur Rata-Rata	38
4.2.4	Kelembapan Rata-Rata.....	38
4.2.5	Lamanya Penyinaran Matahari	39
4.2.6	Kecepatan Angin Maksimum	40
4.2.7	Arah Angin Saat Kecepatan Maksimum	40
4.2.8	Kecepatan Angin Rata-rata	41

4.3	Standarisasi Data	41
4.4	Feature Selection	43
4.5	Analisis Klasifikasi Neural Network.....	44
4.5.1	Pembagian data <i>Holdout cross validation</i>	46
4.5.2	Pembagian data <i>K-Fold cross validation</i>	48
4.6	Perbandingan Pengujian Klasifikasi MLP	52
4.7	Proses SMOTE	53
4.8	Analisis Klasifikasi Neural Network dengan SMOTE.....	54
4.8.1	Pembagian data <i>Holdout cross validation 80%:20%</i>	55
4.9	Perbandingan Hasil Klasifikasi	56
4.9.1	Perbandingan Nilai akurasi	56
4.9.2	Perbandingan nilai <i>sensitivity</i>	57
4.9.3	Perbandingan nilai <i>specificity</i>	57
4.9.4	Perbandingan nilai AUC	58
4.9.5	Perbandingan Evaluasi Model MLP dan MLP+SMOTE.....	59
BAB V KESIMPULAN DAN SARAN		62
5.1	Kesimpulan.....	62
5.2	Saran.....	63
DAFTAR PUSTAKA.....		64
LAMPIRAN.....		83

DAFTAR TABEL

Tabel 2. 1 Penelitian terdahulu	6
Tabel 2. 2 Kategori Curah Hujan.....	12
Tabel 2. 3 contoh confusion matrix 4x4	26
Tabel 2. 4 kategori nilai AUC	28
Tabel 3. 1 Variabel Penelitian	29
Tabel 3. 2 Struktur Data.....	30
Tabel 4. 1 Data curah hujan tidak terukur	35
Tabel 4. 2 Statistik Deskriptif Temperatur Minimum.....	37
Tabel 4. 3 Statistik Deskriptif Temperatur Maksimum	37
Tabel 4. 4 Statistik Deskriptif Temperatur Rata-Rata.....	38
Tabel 4. 5 Statistik Deskriptif Kelembapan Rata-Rata.....	39
Tabel 4. 6 Statistik Deskriptif Lamanya Penyinaran Matahari.....	39
Tabel 4. 7 Kecepatan Angin Maksimum	40
Tabel 4. 8 Statistik Deskriptif Arah Angin Saat Kecepatan.....	40
Tabel 4. 9 Statistik Deskriptif Arah Angin Saat Kecepatan.....	41
Tabel 4. 10 Data Sebelum Standarisasi	42
Tabel 4. 11 Uji Information Gain Feature Selection.....	44
Tabel 4. 12 Uji Coba penentuan jumlah neuron pada hidden layer.....	45
Tabel 4. 13 Confusion matrix Klasifikasi MLP Pengujian 1	46
Tabel 4. 14 Confusion matrix Klasifikasi MLP Pengujian 2.....	47
Tabel 4. 15 Confusion matrix Klasifikasi MLP Pengujian 3.....	49
Tabel 4. 16 Confusion matrix Klasifikasi MLP Pengujian 4.....	50
Tabel 4. 17 Confusion matrix Klasifikasi MLP Pengujian 5.....	51
Tabel 4. 18 Perbandingan Pengujian Klasifikasi MLP	52
Tabel 4. 19 Persentase Perbandingan Data Awal dan Data SMOTE.....	54
Tabel 4. 20 Confusion matrix MLP+SMOTE	55
Tabel 4. 21 Perbandingan Hasil Nilai Akurasi	56
Tabel 4. 22 Perbandingan Hasil Nilai Sensitivity.....	57
Tabel 4. 23 Perbandingan Hasil Nilai Specificity.....	58
Tabel 4. 24 Perbandingan Hasil Nilai AUC.....	58

Tabel 4. 25 Evaluasi Model MLP dan MLP+SMOTE 59
Tabel 4. 26 nilai bobot Masing-masing Variabel 61



DAFTAR GAMBAR

Gambar 2. 1 model <i>single layer network</i>	16
Gambar 2. 2 Model <i>multi layer network</i>	17
Gambar 2. 3 Model competitive layer network.....	17
Gambar 2. 4 fungsi sigmoid	18
Gambar 2. 5 fungsi tanH.....	18
Gambar 2. 6 fungsi ReLu	19
Gambar 2. 7 fungsi aktivasi Linear	19
Gambar 2. 8 Arsitektur Multilayer Perceptron	20
Gambar 2. 9 ROC Curve	27
Gambar 3. 1 flowchart penelitian.....	34
Gambar 4. 1 Alur Proses Data Curah Hujan.....	35
Gambar 4. 2 Score Plot PCA dan Pie chart Curah Hujan.....	36
Gambar 4. 3 Data Setelah Standarisasi.....	43
Gambar 4. 4 Perbandingan Data Awal dan Data SMOTE.....	53
Gambar 4. 5 Grafik Perbandingan Nilai Akurasi	56
Gambar 4. 6 Grafik Perbandingan Nilai Sensitivity.....	57
Gambar 4. 7 Grafik Perbandingan Nilai <i>Specificity</i>	58
Gambar 4. 8 Grafik Perbandingan Nilai AUC.....	59
Gambar 4. 9 Model train Neural Net.....	60

UNUGIRI

DAFTAR LAMPIRAN

Lampiran 1 Data penelitian sebelum di standarisasi	83
Lampiran 2 Data penelitian setelah standarisasi	84
Lampiran 3 Data penelitian setelah SMOTE	85
Lampiran 4 Statistik deskriptif SPSS	86
Lampiran 5 Information Gain Feature Selection.....	86
Lampiran 6 model pengujian MLP terbaik	87



UNUGIRI