

IMPLEMENTATION OF RANDOM FOREST AND SMOTE METHODS FOR ECONOMIC STATUS CLASSIFICATION IN CIREBON CITY

Neneng Nur Sholihah^{*1}, Arief Hermawan²

^{1,2}Informatics, Faculty of Science and Technology, Universitas Teknologi Yogyakarta, Indonesia
Email: ¹neneng.5200411216@student.uty.ac.id, ²ariefdb@uty.ac.id

(Article received: June 16, 2023; Revision: July 09, 2023; published: Desember 23, 2023)

Abstract

The Indonesian government has tried various methods to eradicate poverty throughout the country, one of which is the fair distribution of social assistance to households based on their economic status classification. The determination of social assistance recipients can be influenced by political factors or personal relationships and can be misused, leading to assistance being given to individuals with specific political connections or support. This research aims to develop a household economic status classification system in Cirebon City using the Random Forest algorithm to address these issues. The data used in this research experienced an imbalance in the number of class instances, where the high-class instances were much fewer than the low and medium classes. To address this, the Synthetic Minority Oversampling Technique (SMOTE) was employed. In this study, various testing scenarios were conducted to obtain the best model for accurately predicting household economic status. Based on the research results, the best testing was achieved using Random Forest and SMOTE with a random state of 0, obtaining an accuracy of 93% and excellent performance in classifying each class. When testing unlabeled data, Random Forest successfully predicted 24 out of 30 actual data, resulting in an accuracy rate of 80%. Although this accuracy is lower than what was achieved by Random Forest and SMOTE with a random state of 0, it can be said that the application successfully classifies household economic status in Cirebon City effectively.

Keywords: classification, economic status, random forest, SMOTE.

IMPLEMENTASI METODE RANDOM FOREST DAN SMOTE UNTUK KLASIFIKASI STATUS EKONOMI DI KOTA CIREBON

Abstrak

Pemerintah Indonesia telah mencoba berbagai metode untuk memberantas kemiskinan di seluruh negeri, salah satunya adalah distribusi bantuan sosial yang adil untuk rumah tangga sesuai dengan klasifikasi status ekonomi mereka. Penentuan penerima bantuan sosial dapat dipengaruhi oleh faktor politik atau hubungan personal dapat disalahgunakan sehingga diberikan kepada individu-individu yang memiliki koneksi atau dukungan politik tertentu. Penelitian ini bertujuan untuk membangun sistem klasifikasi status ekonomi rumah tangga di Kota Cirebon menggunakan algoritma *Random Forest* untuk mengatasi permasalahan tersebut. Data yang digunakan pada penelitian ini mengalami ketidakseimbangan jumlah kelas dimana kelas tinggi jauh lebih sedikit dari kelas rendah dan sedang sehingga digunakan metode *Synthetic Minority Oversampling Technique* (SMOTE). Pada penelitian ini, berbagai skenario pengujian dilakukan guna mendapatkan model terbaik untuk memprediksi status ekonomi rumah tangga secara akurat. Berdasarkan hasil penelitian, pengujian terbaik diperoleh menggunakan *Random Forest* dan SMOTE dengan *random state* 0 yang mendapatkan akurasi 93%, dan performa yang sangat baik dalam mengklasifikasikan setiap kelas. Pada pengujian terhadap data yang tidak memiliki label, *Random Forest* berhasil memprediksi benar 24 dari 30 data aktual sehingga memperoleh tingkat akurasi sebesar 80%. Meskipun masih di bawah akurasi yang diperoleh *Random Forest* dan SMOTE dengan *random state* 0, namun dapat dikatakan bahwa aplikasi berhasil mengklasifikasikan status ekonomi rumah tangga di Kota Cirebon dengan baik.

Kata kunci: klasifikasi, random forest, status ekonomi, SMOTE.

1. PENDAHULUAN

The International Society of Gynecological Endocrinology (ISGE) dalam kongres dunia ke-18

mengklasifikasikan Indonesia sebagai salah satu dari 145 negara berkembang di dunia [1]. Salah satu karakteristik umum yang menyebabkan suatu negara

diklasifikasikan sebagai negara berkembang adalah adanya permasalahan kemiskinan di negara tersebut [2]. Sejumlah problem sosial, politik, dan ekonomi yang terjadi khususnya di negara-negara berkembang termasuk Indonesia, disebabkan oleh kemiskinan [3]. Kurangnya pendapatan dan aset menjadi salah satu sebab kemiskinan, di mana kebutuhan dasar seperti makanan, pakaian, perumahan, tingkat kesehatan, dan pendidikan yang dapat diterima tidak dapat dipenuhi [4]. Oleh karena itu, pemerintah di setiap daerah perlu memiliki pemahaman terkait pertumbuhan ekonomi berkelanjutan di daerah mereka. Pengetahuan yang mendalam tentang kondisi ekonomi suatu wilayah merupakan landasan yang krusial dalam merancang kebijakan dan program pemerintah yang efektif untuk meningkatkan kesejahteraan masyarakat [5]. Melalui pemahaman yang mendalam tentang situasi ekonomi, langkah-langkah strategis dapat dirancang untuk menanggulangi kemiskinan, menciptakan lapangan kerja, meningkatkan akses pendidikan dan kesehatan, serta mendorong pertumbuhan ekonomi yang inklusif [6].

Salah satu kota di Indonesia yang juga menghadapi tantangan kemiskinan adalah Kota Cirebon. Terletak di Provinsi Jawa Barat, Cirebon merupakan salah satu kota penting di pesisir utara pulau Jawa. Kota Cirebon memiliki laju pertumbuhan ekonomi yang positif dan cenderung mengalami peningkatan. Pada tahun 2021, ekonomi Kota Cirebon mengalami pertumbuhan 3,04 persen dan meningkat menjadi 5,10 persen pada tahun 2022 [7][8]. Namun, meskipun memiliki potensi ekonomi yang signifikan, kondisi ini tidak diimbangi dengan menurunnya jumlah penduduk miskin di Kota Cirebon. Berdasarkan data yang dipublikasikan oleh Badan Pusat Statistik, jumlah penduduk miskin di Kota Cirebon mengalami peningkatan selama periode Maret 2019 hingga Maret 2021 [9][10]. Jumlah penduduk miskin periode Maret 2019 sebanyak 26,8 ribu jiwa mengalami peningkatan menjadi 30,6 ribu jiwa pada Maret 2020. Pada Maret 2021, bertambah sebanyak 1,4 ribu jiwa sehingga jumlah penduduk miskin menjadi 32 ribu jiwa [9]. Selain itu, garis kemiskinan di Kota Cirebon mengalami peningkatan dari Rp444,57 ribu per kapita per bulan pada periode Maret 2019 dan terus meningkat menjadi Rp467,25 ribu per kapita per bulan pada periode Maret 2021 [10]. Oleh karena itu, diperlukan kajian lebih mendalam terhadap berbagai program bantuan pemerintah agar dapat memberikan dampak positif yang signifikan bagi peningkatan kesejahteraan masyarakat miskin.

Pemerintah Indonesia telah mencoba berbagai metode untuk memberantas kemiskinan di seluruh negeri. Salah satunya adalah distribusi bantuan sosial yang adil di seluruh kepulauan untuk rumah tangga yang dikategorikan layak menerima bantuan sosial [11]. Namun, faktanya pemerintah belum bisa menanggulangi kemiskinan secara keseluruhan. Pada

beberapa kesempatan, ada kebebasan bagi petugas bantuan sosial untuk mengevaluasi dan menilai kelayakan rumah tangga berdasarkan penilaian mereka sendiri [12]. Namun, hal ini dapat menyebabkan ketidakadilan dalam penentuan bantuan sosial. Prasangka atau diskriminasi pribadi pada tingkat individual bisa memengaruhi keputusan dalam menentukan penerima manfaat [13][14]. Contohnya, petugas mungkin memberikan prioritas kepada kelompok tertentu berdasarkan preferensi pribadi yang tidak relevan atau prasangka [15]. Dalam beberapa situasi, penentuan penerima bantuan sosial juga dapat dipengaruhi oleh faktor politik atau hubungan personal. Akibatnya, sistem ini dapat disalahgunakan, di mana bantuan diberikan kepada orang-orang yang memiliki koneksi atau dukungan politik tertentu [16], bukan berdasarkan kebutuhan yang sebenarnya.

Solusi untuk mengimplementasikan program penanggulangan kemiskinan dapat dilakukan melalui berbagai strategi, salah satunya dengan membangun sistem klasifikasi status ekonomi rumah tangga di Kota Cirebon dengan memanfaatkan *machine learning*. Tujuannya adalah untuk mengetahui status ekonomi suatu rumah tangga untuk mendapatkan bantuan dari pemerintah, mengidentifikasi variabel-variabel yang berpengaruh terhadap status ekonomi rumah tangga, dan mengembangkan model klasifikasi yang akurat dalam memprediksi status ekonomi suatu rumah tangga di Kota Cirebon. Klasifikasi merupakan salah satu jenis *supervised learning* atau pembelajaran terawasi [17]. Klasifikasi adalah proses pemberian label secara otomatis ke sampel data yang tidak memiliki label [18]. Pada masalah khusus klasifikasi, label adalah anggota dari himpunan kelas yang terbatas [19]. Apabila jumlah kelas terdiri dari dua set maka disebut dengan *binary classification*, sedangkan kasus klasifikasi dengan tiga kelas atau lebih disebut dengan *multiclass classification* [20]. Pada penelitian ini, klasifikasi status ekonomi rumah tangga di Kota Cirebon akan dibagi menjadi tiga kelas, yaitu rendah, sedang, dan tinggi.

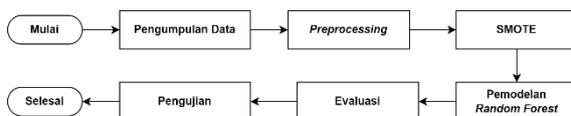
Pada literatur sebelumnya telah dilakukan beberapa penelitian terkait. Penelitian yang telah dilakukan pada tahun 2019 oleh Nurpadillah yang meneliti terkait status kemiskinan rumah tangga di provinsi Jawa Barat menggunakan metode *Random Forest* menghasilkan nilai akurasi sebesar 93.46% sebelum penggunaan SMOTE dan 86.01% setelah penggunaan SMOTE [21]. Kemudian penelitian pada tahun 2020 yang telah dilakukan oleh Nuzula *et al.* terkait klasifikasi status kemiskinan rumah tangga di Kabupaten Wonosobo memperoleh ketepatan klasifikasi untuk metode *Support Vector Machine* (SVM) sebesar 89.94% sedangkan ketepatan klasifikasi untuk metode *Classification And Regression Tree* (CART) sebesar 89.31% [22]. Kemudian penelitian pada tahun 2022 yang dilakukan oleh Hidayatulloh terkait klasifikasi status kemiskinan

di level kabupaten/kota dengan membandingkan kinerja metode *Random Forest* dan *Double Random Forest* menunjukkan bahwa tidak ada perbedaan hasil yang signifikan antara kedua metode tersebut dalam memprediksi status kemiskinan rumah tangga di Jawa Barat. Hal ini disebabkan oleh ukuran pohon yang dihasilkan oleh *Random Forest* sudah dianggap memadai untuk mencapai kinerja model terbaik dengan rata-rata *Area Under Curve* (AUC) sebesar 83,73% dan 83,77% [23]. Berikutnya penelitian pada tahun 2020 yang dilakukan oleh Nofriani terkait aplikasi *machine learning* untuk prediksi klasifikasi status kesejahteraan rumah tangga menyimpulkan bahwa algoritma *Random Forest* memberikan kinerja terbaik dibandingkan algoritma *machine learning* lain dalam memprediksi status kesejahteraan sosial rumah tangga dengan akurasi klasifikasi akhir sebesar 74,20% [11]. Kemudian penelitian pada tahun 2021 yang dilakukan oleh Umma *et al.* terkait klasifikasi status kemiskinan rumah tangga dengan algoritma C5.0 di Kabupaten Pemalang diperoleh bahwa penggunaan *Synthetic Minority Oversampling Technique* (SMOTE) dapat meningkatkan kemampuan *machine learning* dalam mengklasifikasikan data dari kelas mayor maupun minor dengan nilai rata-rata akurasi 82,20% meskipun hasilnya belum menunjukkan signifikansi yang cukup [24].

Berdasarkan latar belakang dan penelitian yang telah dilakukan sebelumnya, penelitian ini akan mengklasifikasikan status ekonomi rumah tangga di Kota Cirebon menggunakan metode *Random Forest* yang belum pernah dilakukan sebelumnya. Selain itu, karena jumlah kelas pada data yang digunakan tidak seimbang maka peneliti akan melakukan penanganan *imbalanced* data. Penanganan ketidakseimbangan data akan diatasi dengan *Synthetic Minority Oversampling Technique* (SMOTE) agar dapat memperoleh hasil prediksi kelas data dengan lebih akurat, baik untuk kelas rendah, sedang maupun kelas tinggi.

2. METODE PENELITIAN

Pada metodologi penelitian ini, dilakukan serangkaian proses pengolahan untuk mencapai hasil uji coba dengan performa metode yang optimal dalam mengklasifikasikan Status Ekonomi Kota Cirebon secara akurat. Metode dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian.

2.1. Pengumpulan Data

Data yang digunakan untuk penelitian ini diperoleh dari halaman Sistem Informasi Layanan

Statistik (SILASTIK) milik Badan Pusat Statistik (BPS). Data yang digunakan merupakan data mikro Survei Sosial Ekonomi Nasional (SUSENAS) yang dilakukan oleh BPS untuk wilayah Kota Cirebon tahun 2021. Pada penelitian ini digunakan 1059 data rumah tangga yang terdiri dari status ekonomi rendah 667 rumah tangga, sedang 373 rumah tangga, dan tinggi 19 rumah tangga. Variabel dalam penelitian ini terdiri atas 1 variabel respon (Y) dan 36 variabel penjelas (X). Informasi terperinci mengenai kelas dan variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Kelas dan variabel yang digunakan dalam penelitian

Variabel	Keterangan
Status Ekonomi	0: Rendah 1: Sedang 2: Tinggi
X1	Perkotaan/Perdesaan
X2	Kehabisan makanan
X3	Status kepemilikan rumah
X4	Luas lantai rumah (m ²)
X5	Memiliki rumah lain
X6	Jenis atap rumah
X7	Jenis dinding rumah
X8	Jenis lantai rumah
X9	Kepemilikan fasilitas BAB
X10	Jenis kloset
X11	Sumber utama air minum
X12	Kekurangan air minum
X13	Sumber air mandi/cuci/dll
X14	Sumber penerangan rumah
X15	Bahan bakar memasak
X16	Memiliki lemari es/kulkas
X17	Memiliki AC
X18	Memiliki pemanas air
X19	Memiliki telepon rumah
X20	Memiliki komputer/laptop
X21	Memiliki emas/perhiasan (min 10gr)
X22	Memiliki sepeda motor
X23	Memiliki perahu
X24	Memiliki perahu motor
X25	Memiliki mobil
X26	Memiliki televisi layar datar (min 30 inch)
X27	Memiliki tanah/lahan
X28	Sumber penghasilan terbesar
X29	Menerima KKS (Kartu Keluarga Sejahtera)
X30	Menerima PKH (Program Keluarga Harapan)
X31	Menerima BPNT (Bantuan Pangan Non Tunai)
X32	Menerima bantuan sosial/subsidi pemerintah
X33	Jumlah anggota rumah tangga
X34	Umur KRT
X35	Ijazah/STTB tertinggi
X36	Status pekerjaan kepala rumah tangga

2.2. Preprocessing

Pada tahap *preprocessing* dilakukan serangkaian langkah untuk mempersiapkan data sebelum dilakukan proses klasifikasi. Tujuan dari tahap pra-pemrosesan ini adalah untuk memastikan kualitas dan konsistensi data yang akan digunakan dalam proses klasifikasi berikutnya [25]. Langkah-langkah ini meliputi *remove duplicate*, *check missing values*, *feature selection*, *one hot encoding*, *label encoding*, dan penanganan ketidakseimbangan data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE).

1. Remove Duplicate

Remove duplicate berfungsi untuk menghapus teks yang identik atau duplikat. Tujuannya adalah untuk mencegah data terisi dengan teks yang sama yang dapat memperlambat kinerja untuk menganalisis model [26]. Menghapus data duplikat sangat penting karena apabila terdapat teks yang identik atau duplikat dalam data dapat menyebabkan beberapa masalah. Pertama, jika data dipenuhi dengan teks yang sama, akan ada redundansi yang tidak perlu yang dapat menghabiskan ruang penyimpanan yang berharga [27]. Hal ini juga dapat mengganggu keberlanjutan proses analisis, karena jumlah data yang perlu diproses akan menjadi lebih besar daripada yang sebenarnya.

2. *Check Missing Values*

Missing values adalah informasi yang tidak tersedia dalam suatu data atau kasus. *Missing values* terjadi ketika informasi mengenai sebuah data tidak diberikan, sulit ditemukan, atau memang tidak tersedia [28]. *Missing values* dapat memiliki dampak yang serius pada analisis data, karena dapat mengganggu keakuratan dan keandalan hasil yang diperoleh. Untuk mengatasi *missing values*, terdapat beberapa pendekatan yang dapat digunakan. Salah satunya adalah dengan melakukan imputasi. Imputasi merupakan sebuah metode untuk mengisi nilai yang hilang dengan nilai yang dapat diprediksi berdasarkan pola atau karakteristik data yang ada [29]. Metode untuk mengatasi *missing values* dalam penelitian ini adalah *median imputation*. *Median imputation* mengatasi *missing values* dengan menghitung nilai median dari variabel tersebut kemudian menggantikan *missing values* dengan nilai median yang telah dihitung [30].

3. *Feature Selection*

Feature selection adalah metode yang digunakan untuk mengoptimalkan model dalam proses preprocessing data. Cara kerja *feature selection* adalah menemukan subset fitur dari himpunan fitur yang ada guna meningkatkan performa model yang diterapkan [31]. Selain itu, pemilihan fitur dapat membantu menghilangkan fitur yang tidak relevan dan berlebihan dalam model yang digunakan sehingga model tersebut dapat mencapai tingkat optimal.

4. *One Hot Encoding*

One Hot Encoding adalah metode yang digunakan untuk mengubah variabel diskrit atau categorical menjadi bentuk biner sehingga dapat digunakan lebih efektif dalam algoritma klasifikasi [32]. Beberapa algoritma tidak dapat secara langsung mengoperasikan variabel diskrit sebagai input sehingga perlu dilakukan transformasi pada variabel diskrit tersebut agar dapat digunakan dalam proses komputasi algoritma tersebut.

5. *Label Encoding*

Label Encoding adalah teknik dalam pemrosesan data yang digunakan untuk mengubah nilai-nilai kategorikal menjadi nilai numerik [33]. Setiap kategori atau label unik dalam variabel

kategorikal diberikan nilai numerik secara berurutan mulai dari 0 hingga n-1 di mana n merupakan jumlah kategori yang ada. *Dataset* yang digunakan dalam penelitian ini terdiri dari tiga kelas, yaitu "rendah", "sedang" dan "tinggi" sehingga label kelas tersebut diubah menjadi 0, 1, dan 2.

2.3. *Synthetic Minority Oversampling Technique (SMOTE)*

SMOTE adalah sebuah metode untuk menyeimbangkan kelas yang berbeda dengan menggunakan oversampling. Pendekatan SMOTE membuat duplikasi data pada kelas minoritas agar seimbang dengan data pada kelas mayoritas [34]. Ketidakeimbangan *dataset* dapat mengakibatkan hasil klasifikasi yang keliru, di mana data pada kelas minoritas sering kali salah diklasifikasikan sebagai kelas mayoritas [35]. Pada penelitian ini, jumlah kelas rendah dan sedang jauh lebih banyak daripada kelas tinggi. Oleh karena itu, akan dilakukan penanganan menggunakan SMOTE dengan membuat data sintesis untuk kelas tinggi agar jumlahnya seimbang dengan dua kelas lainnya.

Pada tahap awal algoritma SMOTE, perhitungan dilakukan dengan mengambil selisih nilai antara vektor fitur pada kelas minoritas dan nilai tetangga terdekat dari kelas minoritas. Kemudian, nilai selisih tersebut dikalikan dengan angka acak yang berada dalam rentang antara 0 hingga 1. Selanjutnya, hasil perhitungan tersebut ditambahkan kembali ke vektor fitur asli sehingga menghasilkan vektor fitur yang baru [30].

$$X_{new} = X_i + (-X_i) \times \delta \quad (1)$$

Keterangan:

X_i = vektor variabel kelas minoritas

\hat{X}_i = *k-nearest neighbors* untuk X_i

δ = nilai acak antara 0 sampai 1

2.4. *Pemodelan Random Forest*

Random Forest adalah sebuah teknik yang menggunakan kumpulan pohon keputusan sebagai model dasar untuk melakukan klasifikasi atau regresi [36]. *Random Forest* merupakan salah satu metode *ensemble learning* yang terdiri dari kumpulan pohon keputusan yang digunakan untuk membuat prediksi yang lebih akurat dan stabil [37]. Dalam klasifikasi menggunakan *Random Forest*, metode ini menggunakan pendekatan *voting* untuk mengambil keputusan mayoritas berdasarkan hasil dari pohon-pohon yang telah dibentuk [38]. Pendekatan menggunakan *ensemble learning* memiliki beberapa keuntungan, yaitu dapat digunakan baik untuk klasifikasi maupun regresi, mampu mencapai tingkat akurasi yang tinggi, dan cocok untuk menganalisis *dataset* yang besar dengan banyak dimensi [39]. Metode *Random Forest* melakukan pemilihan secara acak variabel penjelas untuk mengurangi korelasi

antar *tree* yang terbentuk. Tahapan proses klasifikasi dalam menggunakan *Random Forest* pada data latih yang terdiri atas n amatan dan p variabel penjelas adalah sebagai berikut [40]:

1. Proses *bootstrap*. Penarikan contoh acak berukuran n dari data latih yang disertai dengan pengembalian.
2. Proses *random feature selection*. Menyusun pohon tanpa pemangkasan berdasarkan data hasil *bootstrap* sampai ukuran maksimum. Di setiap proses pemisahan, memilih acak m variabel penjelas dimana $m < p$ kemudian dilakukan pemisahan yang terbaik.
3. Mengulangi langkah 1-2 sebanyak k kali sampai diperoleh k pohon acak.
4. Memprediksi respon suatu amatan dengan menggabungkan hasil prediksi k pohon. Untuk menentukan hasil prediksi akhir menggunakan *majority vote* (mengambil suara terbanyak).

Pada pembentukan pohon, setiap tahapan pemisahan harus mempertimbangkan nilai entropi untuk kemudian menghitung *information gain*. Variabel dengan nilai *information gain* tertinggi akan dipilih sebagai pemisah atau partisi terbaik. Misalkan *dataset A* dibagi menjadi beberapa bagian (A_1, A_2, \dots, A_k). Berikut adalah rumus yang dikembangkan oleh Claude Shannon dalam teori informasi [41] untuk menghitung entropi dan *information gain*.

$$Entropi(A) = - \sum_{i=1}^k -p_i \log_2(p_i) \quad (2)$$

$$IG = Entropi(A) - \sum_{i=1}^k \frac{|A_i|}{|A|} \times Entropi(A_i) \quad (3)$$

Keterangan:

- A = Gugus data
- k = Jumlah partisi A
- p_i = Proporsi dari A_i terhadap A
- IG = *Information gain*
- $|A_i|$ = Jumlah observasi pada partisi ke- i
- $|A|$ = Jumlah observasi pada A

Selain entropi, terdapat ukuran lain yang dapat digunakan dalam proses pemisahan, yaitu indeks *gini*. Pada setiap tahapan pemisahan, nilai *gini* diperhatikan untuk menghitung *gini decrease*. Berikut adalah rumus perhitungan yang digunakan untuk menghitung nilai tersebut.

$$Gini(A) = 1 - \sum_{i=1}^k (p_i^2) \quad (4)$$

$$GD = Gini(A) - \sum_{i=1}^k \frac{|A_i|}{|A|} \times Gini(A_i) \quad (5)$$

Keterangan:

- A = Gugus data
- k = Jumlah partisi A
- p_i = Proporsi dari A_i terhadap A
- GD = *Gini decrease*
- $|A_i|$ = Jumlah observasi pada partisi ke- i
- $|A|$ = Jumlah observasi pada A

2.5. Evaluasi

Pada tahap ini, akan dilakukan proses penilaian kualitas dan kinerja algoritma *Random Forest* untuk mengukur sejauh mana model yang dikembangkan mampu melakukan klasifikasi dengan benar. Metrik evaluasi yang digunakan diantaranya adalah akurasi, *precision*, *recall*, dan *f1-score*. Evaluasi akan ditampilkan dalam bentuk *classification report* agar dapat dilihat kinerja model *Random Forest* dalam memprediksi setiap kelas.

2.6. Pengujian

Ada 4 skenario pengujian yang akan dilakukan. Pengujian pertama adalah penggunaan algoritma *Random Forest* tanpa SMOTE dengan *random state* 42. Pengujian kedua adalah penggunaan algoritma *Random Forest* dengan SMOTE menggunakan *random state* 42. Pengujian ketiga penggunaan algoritma *Random Forest* dengan SMOTE menggunakan *random state* 0. Sedangkan untuk pengujian yang terakhir, akan memanfaatkan model terbaik di antara 3 skenario sebelumnya untuk melakukan prediksi kelas yang tepat untuk setiap data rumah tangga yang belum diberi label. Pengujian ini bertujuan untuk melihat sejauh mana model yang dikembangkan mampu melakukan generalisasi pada data baru yang belum pernah dilihat sebelumnya.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data yang terdiri dari 1059 rumah tangga dan 37 variabel. Proporsi data rumah tangga terbagi menjadi 667 kelas rendah, 373 kelas sedang, dan 19 kelas tinggi. Dapat dilihat pada Tabel 2 terdapat ketidakseimbangan data dimana jumlah rumah tangga dengan status ekonomi kelas tinggi lebih sedikit dibandingkan dengan dua kelas lainnya.

Tabel 2. Data Status Ekonomi Rumah Tangga Kota Cirebon

Status Ekonomi	Frekuensi	Persentase
Rendah	667	62,98%
Sedang	373	35,22%
Tinggi	19	1,79%

Tabel 3 menunjukkan jumlah data setelah dilakukan proses *split*, yaitu diperoleh proporsi 741 untuk data *training* dan 318 untuk data *testing*. Namun, terdapat ketidakseimbangan dalam perbandingan persentase antara kelas tinggi dengan kelas rendah dan sedang pada kedua jenis data tersebut. Jumlah data kelas tinggi jauh lebih sedikit dibandingkan dengan kedua kelas lainnya yang menyebabkan ketidakseimbangan dalam distribusi kelas. Fenomena ini perlu diperhatikan karena dapat mempengaruhi performa dan akurasi model klasifikasi yang akan dibangun. Ketidakseimbangan dalam distribusi kelas dapat mempengaruhi kemampuan model untuk mengenali dan memprediksi data pada kelas yang lebih sedikit. Hal ini dapat mengakibatkan hasil klasifikasi yang bias

terhadap kelas mayoritas sehingga kelas minoritas cenderung diabaikan atau memiliki akurasi yang lebih rendah. Oleh karena itu, penanganan masalah ketidakseimbangan data menjadi penting dalam proses pemodelan klasifikasi ini untuk memastikan hasil yang lebih akurat dan seimbang dalam memprediksi status ekonomi rumah tangga.

Tabel 3. Data status ekonomi rumah tangga kota Cirebon

Kelas	Data Training		Data Testing	
	Jumlah	%	Jumlah	%
Rendah	463	62.48%	204	64.15%
Sedang	263	35.49%	110	34.59%
Tinggi	15	2.02%	4	1.26%
Total	741	100.00%	318	100.00%

Berdasarkan Tabel 3 persentase data *training* status ekonomi rendah, dan sedang berturut-turut 62.48% dan 35.49%. Namun, persentase status ekonomi tinggi pada data *training* jauh lebih sedikit, yaitu 2.02%. Masalah ketidakseimbangan data dapat ditangani menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) dengan menerapkannya pada data *training*. Sebelum *oversampling* dengan SMOTE jumlah kelas rendah, sedang, dan tinggi berturut-turut sebanyak 463, 263, dan 15. Begitupun dengan data *testing* yang juga mengalami ketidakseimbangan. Jumlah kelas rendah, sedang, dan tinggi berturut-turut sebanyak 64.15% atau 204, 34.59% atau 110, dan 1.26% atau 4. Dengan demikian, pada data *testing* perlu juga diterapkan metode SMOTE.

Tabel 4. Komposisi data sebelum dan setelah SMOTE

Kelas	Data Training		Data Testing	
	Sebelum SMOTE	Setelah SMOTE	Sebelum SMOTE	Setelah SMOTE
Rendah	463	463	204	204
Sedang	263	463	110	204
Tinggi	15	463	4	204
Total	741	1389	318	612

Pada Tabel 4 menunjukkan komposisi data setelah pembuatan data sintesis pada data *training* dan data *testing* untuk mengatasi ketidakseimbangan data. Setelah menggunakan SMOTE komposisi data pada setiap kelas untuk data *training* adalah sama rata berjumlah 463 data. Sementara untuk data *testing* setelah SMOTE adalah sama rata berjumlah 204 data.

Tabel 5. Classification Report tanpa SMOTE

Classification Report				
	Precision	Recall	F1-Score	Support
Rendah	91%	95%	93%	204
Sedang	88%	84%	86%	110
Tinggi	100%	50%	67%	4
Accuracy	90%			318

Berdasarkan Tabel 5 hasil klasifikasi status ekonomi Kota Cirebon menggunakan metode *Random Forest* tanpa SMOTE dengan *random state* 42 memperoleh tingkat akurasi yang tinggi, yaitu mencapai 90%. Nilai *precision* yang diperoleh termasuk sangat baik karena mencapai 91% untuk status ekonomi kelas rendah, 88% untuk kelas

sedang, dan 100% untuk kelas tinggi. Namun, *recall* menunjukkan nilai yang bervariasi. Untuk status ekonomi kelas rendah, *recall* mencapai 95%, sedangkan untuk kelas sedang dan tinggi, nilai *recall* masing-masing adalah 84% dan 50%. Hasil *recall* untuk status ekonomi kelas tinggi menunjukkan performa yang kurang baik karena hanya 50% rumah tangga dengan status ekonomi tinggi yang terklasifikasikan dengan benar.

Tabel 6. Classification Report dengan SMOTE

Classification Report				
	Precision	Recall	F1-Score	Support
Rendah	91%	94%	92%	204
Sedang	81%	91%	86%	204
Tinggi	100%	86%	92%	204
Accuracy	90%			612

Pada Tabel 6 ditunjukkan hasil klasifikasi status ekonomi Kota Cirebon menggunakan metode *Random Forest* dengan SMOTE memperoleh tingkat akurasi yang tinggi juga, yaitu mencapai 90%. Begitupun dengan perolehan nilai *precision* yang mencapai 91% untuk status ekonomi kelas rendah, 81% untuk kelas sedang, dan 100% untuk kelas tinggi. Namun, setelah penggunaan SMOTE nilai *recall* yang diperoleh menjadi lebih baik. Untuk status ekonomi kelas rendah, *recall* mencapai 94%, sedangkan untuk kelas sedang dan tinggi, nilai *recall* masing-masing adalah 91% dan 86%. Hal itu menunjukkan bahwa penggunaan SMOTE mampu meningkatkan performa algoritma *Random Forest* dalam mengklasifikasikan rumah tangga dengan status ekonomi tinggi dibandingkan sebelumnya tanpa penggunaan SMOTE.

Pengujian menggunakan nilai parameter yang berbeda dapat dilakukan agar bisa menghasilkan tingkat akurasi dan klasifikasi yang lebih baik. Pada proses pembagian data, SMOTE dan pemodelan *Random Forest*, nilai parameter *random state* diatur menjadi 0. Perubahan nilai parameter *random state* sedikit mengubah proporsi jumlah kelas pada data *training* dan data *testing*. Namun, jumlah data *training* yang diperoleh dari proses *split* data dengan *random state* 0 tidak mengalami perbedaan dengan *random state* 42. Perbandingan kelas tinggi pada data *training* maupun data *testing* dengan *random state* 0 tidak terlalu jauh seperti saat penggunaan *random state* 42, yaitu sebanyak 11 pada data *training* dan sebanyak 8 pada data *testing*.

Tabel 7. Kinerja klasifikasi Random Forest dengan random state 0

Classification Report				
	Precision	Recall	F1-Score	Support
Rendah	95%	96%	95%	210
Sedang	88%	93%	90%	210
Tinggi	97%	91%	94%	210
Accuracy	93%			630

Pada Tabel 7 dapat dilihat bahwa hasil percobaan dengan mengubah parameter *random state* menjadi 0 mendapatkan persentase akurasi yang lebih tinggi dibandingkan penggunaan *random state* 42,

21	Tinggi	1
22	Tinggi	1
23	Tinggi	1
24	Sedang	0
25	Tinggi	1
26	Tinggi	1
27	Tinggi	1
28	Sedang	0
29	Sedang	0
30	Tinggi	1

Dari 30 entri, model memprediksi benar 24 data aktual sehingga memperoleh tingkat akurasi sebesar 80%. Model berhasil memprediksi benar 9 dari 10 data aktual untuk kelas rendah, 8 dari 10 prediksi benar untuk kelas sedang, dan 7 dari 10 prediksi benar untuk kelas tinggi. Untuk akurasi dari masing-masing kelas, kelas rendah memiliki akurasi tertinggi, yaitu 90%. Selanjutnya diikuti oleh kelas sedang sebesar 80% dan kelas tinggi 70%. Meskipun masih di bawah akurasi yang diperoleh sebelumnya oleh *Random Forest* dan SMOTE dengan *random state* 0 yang mencapai 93% namun dapat dikatakan bahwa aplikasi berhasil mengimplementasikannya. Oleh karena itu, algoritma ini cocok untuk mengklasifikasikan status ekonomi rumah tangga di Kota Cirebon.

4. DISKUSI

Untuk memperluas sudut pandang dan sebagai faktor perbandingan dengan beberapa jurnal, peneliti akan memilih beberapa jurnal yang membahas terkait algoritma *Random Forest*. Hal ini bertujuan untuk mendapatkan wawasan yang lebih luas tentang performa dari algoritma *Random Forest* dalam konteks yang berbeda. Selain itu, perbandingan ini juga dapat membantu mengevaluasi keunggulan dan kelemahan dari algoritma *Random Forest* dan memberikan pemahaman yang lebih komprehensif tentang potensi penggunaan algoritma *Random Forest* dalam aplikasi yang berbeda.

Penelitian pertama dilakukan oleh [42] berjudul “Prediction Of Stunting Prevalence In East Java Province With Random Forest Algorithm”. Tujuan dari penelitian ini adalah untuk memprediksi prevalensi stunting di Provinsi Jawa Timur menggunakan algoritma *Random Forest*. Pendekatan kuantitatif deskriptif yang digunakan melibatkan penggunaan algoritma *Random Forest*. Penelitian menunjukkan bahwa dari 20 faktor kandidat, hanya 12 faktor yang diduga sebagai faktor penyebab stunting berdasarkan nilai korelasinya. Hasil dari penelitian ini menunjukkan bahwa *Multi-Linear Regressor* (MLR) dengan nilai *Mean Absolute Error* (MAE) sebesar 0,93 dan nilai MSE sebesar 1,34 lebih unggul daripada *Random Forest Regression* dengan nilai MAE sebesar 1,02 dan nilai MSE sebesar 1,64.

Penelitian kedua dilakukan oleh [43] berjudul “Implementation of The Random Forest Algorithm in Classifying The Accuracy of Graduation Time for Computer Engineering Students at Dian Nuswantoro

University”. Masalah yang diteliti dalam penelitian ini adalah tingkat akurasi dalam memprediksi waktu kelulusan mahasiswa program Teknik Komputer di Universitas Dian Nuswantoro. Tujuan penelitian ini adalah untuk menciptakan model klasifikasi terbaik menggunakan algoritma *Random Forest* untuk memprediksi akurasi waktu kelulusan mahasiswa yang akan bermanfaat dalam pengambilan kebijakan di masa depan. Hasil dari penelitian ini menunjukkan bahwa model klasifikasi menggunakan algoritma *Random Forest* memiliki akurasi sebesar 93% untuk data *training* dan 92% untuk data *testing*.

Penelitian ketiga dilakukan oleh [44] berjudul “Studi Komparasi *Naive Bayes*, *K-Nearest Neighbor*, dan *Random Forest* untuk Prediksi Calon Mahasiswa yang Diterima atau Mundur”. Metode klasifikasi yang digunakan dalam penelitian ini untuk memprediksi apakah calon mahasiswa diterima atau mundur. Tujuan dari penelitian ini adalah mendapatkan model prediksi terbaik dari data Penerimaan Mahasiswa Baru tahun 2014 hingga 2019 dengan membandingkan performa algoritma *Naive Bayes*, *K-Nearest Neighbor*, dan *Random Forest*. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* merupakan yang terbaik dengan akurasi sebesar 73,61%, dibandingkan dengan *K-Nearest Neighbor* yang memiliki akurasi sebesar 72,08%, dan *Naive Bayes* dengan akurasi sebesar 70,47%. Disimpulkan juga bahwa optimasi model menggunakan teknik *hyperparameter* dapat menghasilkan peningkatan nilai akurasi.

Penelitian keempat dilakukan oleh [45] berjudul “Klasifikasi Citra Menggunakan Metode *Random Forest* dan *Sequential Minimal Optimization* (SMO)”. Masalah yang ingin dipecahkan dalam penelitian ini adalah memisahkan jenis-jenis kategori dan label citra bunga yang telah dibuat. Metode segmentasi digunakan untuk menghilangkan noise pada latar belakang dan mendapatkan fitur citra yang akan digunakan dalam klasifikasi, seperti *eccentricity*, *perimeter*, *metric*, dan *area*. Hasil penelitian menunjukkan bahwa metode *Random Forest* memiliki akurasi terbaik dengan nilai 100% dalam kedua skenario, sementara *Sequential Minimal Optimization* (SMO) mendapatkan nilai akurasi sebesar 92,68% dalam skenario pembagian 66%. Penelitian ini menyimpulkan bahwa metode *Random Forest* lebih unggul dalam klasifikasi citra bunga dibandingkan dengan SMO.

Berdasarkan penelitian yang telah dilakukan terkait penggunaan algoritma *Random Forest*, dapat disimpulkan bahwa algoritma ini memiliki performa yang baik dalam berbagai konteks dan aplikasi yang berbeda. Hasil observasi pada penelitian-penelitian tersebut menunjukkan bahwa algoritma *Random Forest* memiliki keunggulan dalam mengatasi permasalahan klasifikasi dan prediksi dalam berbagai bidang, baik itu dalam analisis data kesehatan, pendidikan, maupun pengolahan citra. Algoritma ini dapat menghasilkan model prediksi yang akurat

dengan kemampuan untuk mengekstraksi fitur penting dari *dataset*. Selain itu, *Random Forest* juga mampu mengatasi masalah *overfitting* dan dapat bekerja dengan baik pada *dataset* yang kompleks dan besar.

5. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penggunaan metode *Random Forest* dalam mengklasifikasikan status ekonomi rumah tangga di Kota Cirebon memberikan hasil yang sangat baik. Penerapan metode *Synthetic Minority Oversampling Technique* (SMOTE) untuk menangani ketidakseimbangan data mampu meningkatkan performa dalam mengklasifikasikan kelas minoritas, terutama pada kelas tinggi yang jumlahnya lebih sedikit dari kelas rendah dan sedang. Pengujian terbaik diperoleh menggunakan *Random Forest* dan SMOTE dengan *random state 0* yang mendapatkan akurasi 93% dan performa yang lebih baik dalam mengklasifikasikan setiap kelas. Pada pengujian terhadap data yang tidak memiliki label, *Random Forest* berhasil memprediksi benar 24 dari 30 data aktual sehingga memperoleh tingkat akurasi sebesar 80%. Meskipun masih di bawah akurasi yang diperoleh *Random Forest* dan SMOTE dengan *random state 0* namun dapat dikatakan bahwa aplikasi berhasil mengklasifikasikan status ekonomi rumah tangga di Kota Cirebon dengan baik. Dengan keunggulan seperti kemampuan mengatasi ketidakseimbangan data, identifikasi variabel penting, tingkat akurasi yang tinggi, dan penanganan *overfitting*, *Random Forest* merupakan algoritma yang efektif dan dapat digunakan untuk analisis status ekonomi rumah tangga di Kota Cirebon.

UCAPAN TERIMA KASIH

Penulis mengungkapkan penghargaan yang tulus kepada semua pihak yang telah memberikan kontribusi berarti dalam penelitian ini. Terima kasih kepada Badan Pusat Statistik yang telah memberikan akses data yang penting dalam mendukung penelitian kami. Penulis juga ingin menyampaikan rasa terima kasih kepada Universitas Teknologi Yogyakarta yang telah menyediakan fasilitas dan sumber daya yang berharga dalam penelitian ini. Kami juga ingin menyampaikan rasa terima kasih kepada keluarga dan teman-teman atas dukungan, dan semangat yang diberikan selama pelaksanaan penelitian ini. Semoga hasil penelitian ini dapat memberikan sumbangsih yang berarti dalam pengembangan kebijakan dan program penanggulangan kemiskinan di Kota Cirebon.

DAFTAR PUSTAKA

- [1] The International Society of Gynecological Endocrinology (ISGE), "List of developing countries | ISGE 2018," 2018. <https://isge2018.isgesociety.com/registration/list-of-developing-countries/> (accessed Jul. 08, 2023).
- [2] V. Houlden, G.-M. Tsarouchi, and N. Walmsley, "The Impact of Climate Change on the Achievement of the Post 2015 Sustainable Development Goals," in *Climate and Development Knowledge Network*, South Africa: Climate and Development Knowledge Network (CDKN), 2015.
- [3] D. Seftiana, O. D. Arleina, S. Dewi, R. Amalia, and F. Fachrunisah, "Klasifikasi Rumah Tangga Miskin di Kabupaten Jombang dengan Pendekatan Random Forest Cart," in *Pekan Ilmiah Mahasiswa Nasional Program Kreativitas Mahasiswa-Penelitian 2014*, Indonesian Ministry of Research, Technology and Higher Education, 2014.
- [4] N. Halimatussa'diyah, G. W. Sasmito, and D. Apriliani, "Aplikasi Klasifikasi Status Kemiskinan Warga di Kabupaten Tegal," *Komputika: Jurnal Sistem Komputer*, vol. 10, no. 1, pp. 61–67, 2021, doi: 10.34010/komputika.v10i1.3701.
- [5] D. Maryani and R. R. E. Nainggolan, *Pemberdayaan Masyarakat*. Deepublish, 2019.
- [6] O. Purba et al., *Dasar Hukum & Analisis Tata Kelola Ibu Kota Negara dari Berbagai Bidang*. PT. Sonpedia Publishing Indonesia, 2023.
- [7] Badan Pusat Statistik, *Pertumbuhan Ekonomi Kota Cirebon Tahun 2021*. Cirebon, 2022.
- [8] Badan Pusat Statistik, *Pertumbuhan Ekonomi Kota Cirebon Tahun 2022*. Cirebon, 2023.
- [9] Badan Pusat Statistik, *Indikator Kesejahteraan Rakyat Kabupaten Cirebon Tahun 2022*. Cirebon, 2022.
- [10] Badan Pusat Statistik, *Profil Kemiskinan Kota Cirebon Tahun 2022*. Cirebon, 2022.
- [11] N. Nofriani, "Machine Learning Application for Classification Prediction of Household's Welfare Status," *JITCE (Journal of Information Technology and Computer Engineering)*, vol. 4, no. 02, pp. 72–82, Sep. 2020, doi: <https://doi.org/10.25077/jitce.4.02.72-82.2020>.
- [12] M. Rosfadhila, N. Toyamah, B. Sulaksono, S. Devina, R. J. Sodo, and M. Syukri, "Kajian Cepat Pelaksanaan Program Bantuan Langsung Tunai (BLT) 2008 dan Evaluasi Penerima Program BLT 2005 di Indonesia," *SMERU Research Institute*, pp. 1–107, 2011.
- [13] F. Taruh, *Motivasi Kerja (Meniti Suara Hati Menolak Perilaku Korupsi)*, Cetakan Pertama. Yogyakarta: Deepublish, 2020.
- [14] M. C. Ulum, *Perilaku Organisasi Menuju*

- Orientasi Pemberdayaan*, Cetakan Pertama. Malang: Universitas Brawijaya Press, 2016.
- [15] A. Liliweri, *Prasangka, Konflik, dan Komunikasi Antarbudaya*, Edisi Kedua. Prenada Media, 2018.
- [16] R. Febari, *Politik Pemberantasan Korupsi*, Cetakan Pertama. Jakarta: Yayasan Pustaka Obor Indonesia, 2015.
- [17] A. C. Müller and S. Guido, *Introduction to Machine Learning with Python*, First Edition. Sebastopol: O'Reilly Media, Inc., 2016.
- [18] A. Rachmat and Y. Lukito, "Sentipol: Dataset Sentimen Komentar pada Kampanye Pemilu Presiden Indonesia 2014 dari Facebook Page," *Konferensi Nasional Teknologi Informasi dan Komunikasi 2017*, pp. 218–228, 2016.
- [19] D. H. Fudholi, "Presensi Kelas Berbasis Pola Wajah dan Tersenyum Menggunakan Deep Learning," Tesis, Universitas Islam Indonesia, Yogyakarta, 2023.
- [20] A. Burkov, *The Hundred-Page Machine Learning*, vol. 1. Andriy Burkov Quebec City, QC, Canada, 2019.
- [21] W. Nurpadillah, "Metode Ensemble pada Pohon Klasifikasi Tunggal untuk Klasifikasi Status Kemiskinan Rumah Tangga di Provinsi Jawa Barat," Skripsi, Institut Pertanian Bogor, Bogor, 2019.
- [22] L. Nuzula, A. Prahutama, and A. R. Hakim, "Klasifikasi Status Kemiskinan Rumah Tangga dengan Metode Support Vector Machines (SVM) dan Classification and Regression Trees (CART) Menggunakan GUI R (Studi Kasus di Kabupaten Wonosobo Tahun 2018)," *Jurnal Gaussian*, vol. 9, no. 4, pp. 525–534, 2020, doi: <https://doi.org/10.14710/j.gauss.v9i4.29449>.
- [23] N. G. T. Hidayatulloh, "Perbandingan Kinerja Random Forest dan Double Random Forest untuk Klasifikasi Status Kemiskinan di Level Kabupaten/Kota," Skripsi, Institut Pertanian Bogor, Bogor, 2022.
- [24] F. N. Umma, B. Warsito, and I. M. Di Asih, "Klasifikasi Status Kemiskinan Rumah Tangga dengan Algoritma C5. 0 di Kabupaten Pematang," *Jurnal Gaussian*, vol. 10, no. 2, pp. 221–229, 2021, doi: <https://doi.org/10.14710/j.gauss.10.2.221-229>.
- [25] L. B. Adzy, A. Asriyanik, and A. Pambudi, "Algoritma Naïve Bayes untuk Klasifikasi Kelayakan Penerima Bantuan Iuran Jaminan Kesehatan Pemerintah Daerah Kabupaten Sukabumi," *Jurnal Mnemonic*, vol. 6, no. 1, pp. 1–10, May 2023, doi: <https://doi.org/10.36040/mnemonic.v6i1.5714>.
- [26] H. Hermanto, A. Mustopa, and A. Y. Kuntoro, "Algoritma Klasifikasi Naive Bayes dan Support Vector Machine dalam Layanan Komplain Mahasiswa," *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, vol. 5, no. 2, pp. 211–220, Feb. 2020, doi: <https://doi.org/10.33480/jitk.v5i2.1181>.
- [27] E. S. Eriana and S. Farizy, *Sistem Informasi Manajemen*, Cetakan Pertama. Tangerang Selatan: Unpam Press, 2021.
- [28] D. A. Rusdah and H. Murfi, "XGBoost in Handling Missing Values for Life Insurance Risk Prediction," *SN Appl Sci*, vol. 2, pp. 1–10, Jul. 2020, doi: 10.1007/S42452-020-3128-Y.
- [29] Y. R. Dewi, "Analisis Perbandingan Akurasi Imputasi Data Mean dan Median untuk Data Impor Pupuk," *AGRIMAS*, vol. 6, no. 1, pp. 44–45, Jun. 2022.
- [30] M. L. Suliztia, "Penerapan Analisis Random Forest pada Prototype Sistem Prediksi Harga Kamera Bekas Menggunakan Flask," Skripsi, Universitas Islam Indonesia, Yogyakarta, 2020.
- [31] O. Somantri, W. E. Nugroho, and A. R. Supriyono, "Penerapan Feature Selection pada Algoritma Decision Tree untuk Menentukan Pola Rekomendasi Dini Konseling," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, pp. 272–279, Dec. 2022, doi: <http://dx.doi.org/10.30865/json.v4i2.5267>.
- [32] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creative Information Technology Journal*, vol. 6, no. 1, pp. 1–11, 2020.
- [33] E. Andini, M. R. Faisal, R. Herteno, R. A. Nugroho, and F. Abadi, "Peningkatan Kinerja Prediksi Cacat Software dengan Hyperparameter Tuning pada Algoritma Klasifikasi Deep Forest," *Jurnal Mnemonic*, vol. 5, no. 2, pp. 119–127, Aug. 2022, doi: <https://doi.org/10.36040/mnemonic.v5i2.4793>.
- [34] M. Mustaqim, B. Warsito, and B. Surarso, "Combination of synthetic minority oversampling technique (Smote) and backpropagation neural network to handle imbalanced class in predicting the use of contraceptive implants," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 5, no. 2, pp. 116–127, Jul. 2019, doi: 10.26594/register.v5i2.1705.
- [35] L. Qadrini, H. Hikmah, and M. Megasari, "Oversampling, Undersampling, SMOTE SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejava Timur Tahun

- 2017,” *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 386–391, Sep. 2022, doi: <https://doi.org/10.47065/josyc.v3i4.2154>.
- [36] A. Primajaya and B. N. Sari, “Random Forest Algorithm for Prediction of Precipitation,” *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 1, no. 1, pp. 27–31, 2018, doi: <http://dx.doi.org/10.24014/ijaidm.v1i1.4903>.
- [37] D. Ismafillah, T. Rohana, and Y. Cahyana, “Analisis Algoritma Pohon Keputusan untuk Memprediksi Penyakit Diabetes Menggunakan Oversampling SMOTE,” *INFOTECH: Jurnal Informatika & Teknologi*, vol. 4, no. 1, pp. 27–36, Jun. 2023, doi: <https://doi.org/10.37373/infotech.v4i1.452>.
- [38] M. W. B. Santoso, R. C. Wihandika, and M. A. Rahman, “Ekstraksi Ciri untuk Klasifikasi Jenis Kelamin berbasis Citra Wajah menggunakan Metode Compass Local Binary Patterns,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 11, pp. 10556–10563, Jan. 2019.
- [39] I. A. Dahlan, “Klasifikasi Cuaca Provinsi DKI Jakarta Menggunakan Algoritma Random Forest dengan Teknik Oversampling,” *Jurnal Teknoinfo*, vol. 16, no. 1, pp. 87–92, 2022, doi: <https://doi.org/10.33365/jti.v16i1.1533>.
- [40] J. Widiastuti, “Klasifikasi Pembiayaan Warung Mikro Menggunakan Metode Random Forest dengan Teknik Sampling Kelas Imbalanced (Studi Kasus: Data Nasabah Pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi),” Tugas Akhir, Universitas Islam Indonesia, Yogyakarta, 2018.
- [41] R. G. Gallager, “Claude E. Shannon: a retrospective on his life, work, and impact,” *IEEE Trans Inf Theory*, vol. 47, no. 7, pp. 2681–2695, 2001, doi: 10.1109/18.959253.
- [42] M. S. Haris, M. Anshori, and A. N. Khudori, “Prediction of Stunting Prevalence in East Java Province with Random Forest Algorithm,” *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 1, pp. 11–13, 2023, doi: <https://doi.org/10.20884/1.jutif.2023.4.1.614>.
- [43] D. A. Rachmawati, N. A. Ibadurrahman, J. Zeniarja, and N. Hendriyanto, “Implementation of The Random Forest Algorithm in Classifying The Accuracy of Graduation Time for Computer Engineering Students at Dian Nuswantoro University,” *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 3, pp. 565–572, 2023, doi: <https://doi.org/10.52436/1.jutif.2023.4.3.920>.
- [44] P. Sejati, M. Munawar, M. Pilliang, and H. Akbar, “Studi Komparasi Naive Bayes, K-Nearest Neighbor, dan Random Forest untuk Prediksi Calon Mahasiswa yang Diterima atau Mundur,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 7, pp. 1341–1348, 2022, doi: <http://dx.doi.org/10.25126/jtiik.201743299>.
- [45] P. Rosyani, “Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO),” *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, vol. 9, no. 2, pp. 132–134, 2021, doi: 10.26418/justin.v9i2.44120.