



## Classification of Poverty Status using the Random Forest Algorithm

Syaidatussalihah<sup>a\*</sup>, Abdurahim<sup>b</sup>

<sup>a</sup> Politeknik Medica Farma Husada Mataram, Jalan Medica Farma No.1, Mataram, Indonesia, 83116. Email: [syaidatussalihah11@gmail.com](mailto:syaidatussalihah11@gmail.com)

<sup>b</sup> Politeknik Medica Farma Husada Mataram, Jalan Medica Farma No.1, Mataram, Indonesia, 83116. Email: [aim.math13@gmail.com](mailto:aim.math13@gmail.com)

### ABSTRACT

Poverty is a fundamental problem because it deals with the basic needs of society. In NTB Province, many households are living below the poverty line. One reason is that the government's efforts to reduce poverty are not optimal. Therefore, it is necessary to classify the factors that affect the poverty level so it can be used as a reference in making policies to reduce poverty. One of the classification methods is the Random Forest method. The Random Forest method with the optimal mtry and ntree scores, i.e.,  $m = 3$  and  $n = 1000$ , respectively, obtained an accuracy rate of 81.3%. This means that the accuracy of the Random Forest classification method for this data is very good. The income variable is the most influential factor in determining poverty status based on Random Forest analysis, with a Mean Decrease Accuracy score of 23.92%. It has the highest Mean Decrease Accuracy value among other attribute variables.

*Keywords: Poverty, Mean Decrease Accuracy, Random Forest*

Diserahkan: 14-06-2022; Diterima: 08-07-2022;

Doi: <https://doi.org/10.29303/emj.v5i1.133>

### 1. Pendahuluan

Kemiskinan merupakan situasi penduduk atau sebagian penduduk yang hanya dapat memenuhi makanan, pakaian, dan perumahan yang sangat diperlukan untuk mempertahankan tingkat kehidupan yang minimum (KBBI, 2022). Lebih jauh, kemiskinan merupakan persoalan mendasar karena menyangkut pemenuhan kebutuhan dasar masyarakat. Kemiskinan juga merupakan permasalahan global karena dihadapi oleh negara-negara belahan dunia lainnya (Yacoub, 2012). Kemiskinan dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan maupun non makanan yang

diukur dari sisi pengeluaran (Tarigan et al., 2020). Salah satu komponen yang mempengaruhi tingkat kemiskinan adalah Indeks Pembangunan Manusia (IPM), karena IPM berbanding lurus dengan kemampuan ekonomi masyarakat. IPM didasarkan pada tiga dimensi yaitu tingkat harapan hidup saat lahir, tingkat melek huruf, tahun sekolah rata-rata, dan paritas daya beli (Juliannisa & Siswanti, 2019).

Provinsi Nusa Tenggara Barat (NTB) merupakan salah satu contoh daerah yang masih menghadapi permasalahan kemiskinan dan penanggulangan kemiskinan. Provinsi NTB tergolong dalam kemiskinan yang cukup tinggi dikarenakan angkanya tercatat lebih besar jika dibandingkan dengan rata-rata

\* Corresponding author.

Alamat e-mail: [syaidatussalihah11@gmail.com](mailto:syaidatussalihah11@gmail.com)

nasional. Lebih jauh, Provinsi NTB termasuk dalam kelompok 2 berdasarkan indikator kemiskinan, yaitu indeks keparahan kemiskinan, indeks kedalaman kemiskinan dan presentase penduduk miskin (Ferezagia, 2018).

Menurut data Badan Pusat Statistik NTB, Kabupaten Lombok Timur merupakan daerah dengan jumlah penduduk miskin tertinggi di Provinsi Nusa Tenggara Barat pada tahun 2021, yaitu 190.840 jiwa.

**Tabel 1** – Jumlah Penduduk Miskin di NTB Tahun 2021

Kabupaten/Kota	Jumlah Penduduk Miskin (Ribu Jiwa)
Kab. Lombok Barat	105,24
Kab Lombok Tengah	131,94
Kab. Lombok Timur	190,84
Kab. Sumbawa	66,00
Kab Dompu	33,26
Kab. Bima	75,49
Kab. Sumbawa Barat	21,51
Kab. Lombok Utara	61,70
Kota Mataram	44,45
Kota Bima	16,22

**Sumber :** Badan Pusat Statistik NTB

Berdasarkan Tabel 1 menunjukkan bahwa masih terdapat masyarakat Provinsi NTB yang hidup di bawah garis kemiskinan. Penyebab salah satunya adalah belum optimalnya upaya pemerintah dalam menurunkan tingkat kemiskinan. Untuk itu perlu diklasifikasi faktor-faktor yang mempengaruhi tingkat kemiskinan sehingga dapat digunakan sebagai acuan dalam mengambil kebijakan untuk mengurangi tingkat kemiskinan. Salah satu metode untuk klasifikasi adalah algoritma *Random Forest*. Misalnya pada penelitian yang dilakukan oleh (Mauludiyah & others, 2020) untuk mengklasifikasi indeks pembangunan manusia menggunakan algoritma *Random Forest*. Metode serupa juga digunakan pada bidang kedokteran, yaitu untuk mengklasifikasi penderita HIV/AIDS (Otok & others, 2015). Selain itu, dalam bidang kependudukan digunakan untuk mengklasifikasi apakah penduduk berstatus bekerja atau tidak sesuai dengan kriteria (Christy et al., 2021).

Selain untuk klasifikasi, metode ini juga dapat digunakan untuk prediksi. Dalam penelitian yang dilakukan oleh (Rianto & Yunis, 2021) untuk memprediksi jumlah mahasiswa baru dengan model *Random Forest* (RF) dengan tingkat akurasi 99,8%. Penelitian lain juga digunakan untuk memprediksi Bit Coin menggunakan RF dengan tingkat akurasi 98,50% (Saadah & Salsabila, 2021).

Selain algoritma *Random Forest*, terdapat beberapa metode yang bisa digunakan untuk

klasifikasi, misalnya *Regresi Logistik* (RL), *Support Vector Machines* (SVM), dan *Classification and Regression Trees* (CART). Pada penelitian terdahulu juga pernah dilakukan perbandingan metode-metode tersebut. Misalnya mengklasifikasi faktor-faktor yang mempengaruhi prestasi mahasiswa menggunakan Regresi Logistik Ordinal dan RF. Sebelum itu juga pernah dilakukan kajian perbandingan metode Regresi Logistik dan *Random Forest* untuk klasifikasi data *imbalanced* (Purwa, 2019). Pengklasifikasian rumah tangga miskin atau tidaknya dapat dilakukan dengan metode SVM dan CART (Nuzula et al., 2020). Sesuai dengan hasil kajian perbandingan tiga metode, yaitu CART konvensional, *Bagging*, dan RF, algoritma RF yang menghasilkan klasifikasi terbaik (Jatmiko et al., 2019). Kajian serupa juga pernah dilakukan untuk klasifikasi penyakit Hepatitis C dengan metode SMOTE RF dan SMOTE XGBOOST, di mana secara keseluruhan algoritma RF memiliki akurasi yang lebih tinggi (Syukron et al., 2020).

Berdasarkan latar belakang di atas, maka akan dianalisis faktor-faktor yang paling berpengaruh terhadap kemiskinan menggunakan algoritma *Random Forest* (Studi Kasus : Data Kemiskinan di Desa Paok Lombok). Metode ini digunakan untuk melihat hubungan variabel prediktor dengan variabel penjelas dan di dalam *Random Forest* juga terdapat nilai kepentingan masing-masing variabel penelitian sehingga bisa dilihat analisis faktor – faktor yang paling berpengaruh terhadap kemiskinan.

## 2. Tinjauan Pustaka

### 2.1. Classification And Regression Trees (CART)

*Classification And Regression Trees* (CART) adalah suatu metode nonparametrik dari salah satu teknik eksplorasi data yang dikenal sebagai teknik pohon keputusan (*decision tree*). Pohon keputusan dibentuk secara biner dengan algoritma penyekatan rekursif (*binary recursive partitioning*) (Lewis, 2000). Adapun untuk Pohon klasifikasi dihasilkan saat peubah respons berupa data kategorik, sedangkan pohon regresi dihasilkan saat peubah respons berupa data numerik (Breiman & Ihaka, 1984)

Menurut Lewis (2000), CART merupakan suatu cara pemilahan sekelompok data dalam suatu ruang yang disebut simpul (*node*) menjadi dua simpul anak dan setiap simpul anak dapat dipilah lagi menjadi dua

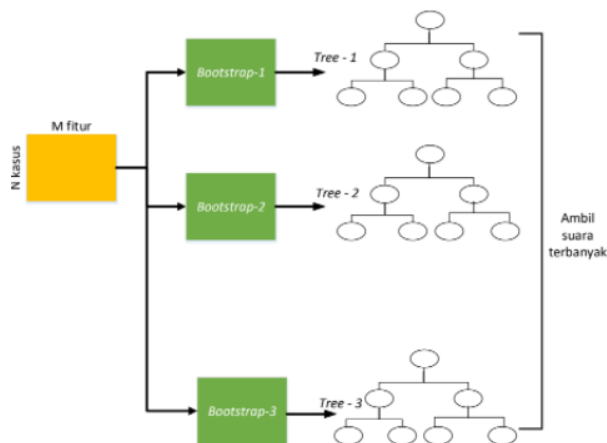
simpul anak berikutnya, begitu seterusnya dan berhenti jika telah mendapatkan sekelompok observasi yang relatif homogen.

**2.2. Random Forest**

*Random forest* adalah suatu algoritma yang digunakan pada klasifikasi data dalam jumlah yang besar. Klasifikasi *random forest* dilakukan melalui penggabungan pohon dengan melakukan training pada sampel data yang dimiliki. Penggunaan pohon yang semakin banyak akan mempengaruhi akurasi yang akan didapatkan menjadi lebih baik (Tahyudin et al., 2021).

Adapun pemilihan sejumlah data training pada metode *Random Forest* dilakukan dengan menggunakan teknik *bootstrapping*. Metode *bootstrapping aggregating* dalam pengambilan sampel untuk diimplementasikan dalam membangun setiap *decision tree* dengan kandidat atribut yang terpilih sebelumnya (Muflikhah et al., 2021). Secara garis besar, tahapan dari algoritma *Random Forest* adalah sebagaimana ditunjukkan dalam Gambar 1 berikut (Muflikhah et al., 2021):

- Ambil *bootstrap* dari data pelatihan
- Konstruksi pohon keputusan
- Setiap *node* dalam pemilihan fitur split, pilih hanya diantara fitur  $m < M$
- Buat pohon keputusan dari setiap sampel *bootstrap*
- Ambil suara terbanyak



**Gambar 1** – Ilustrasi Algoritma *Random Forest*

**2.3. Confusion Matrix**

Confusion matrix adalah tabulasi dari perhitungan yang didasari pada evaluasi kinerja model klasifikasi berdasarkan jumlah objek penelitian yang diprediksi dengan benar dan salah. Secara singkat confusion matrix memberikan perincian terkait kesalahan klasifikasi (Gorunescu, 2011). Berikut ini merupakan hasil dari *confusion matrix* (Sasongko, 2016).

**Tabel 2** – *Confusion Matrix*

Prediksi	Aktual	
	True	False
True	TP	FN
False	FP	TN

Tabel 2 menampilkan perbandingan kelas aktual (sebenarnya) dengan kelas prediksi yang dapat digunakan untuk menghasilkan metrik evaluasi seperti *Accuracy* (akurasi), *Precision* dan *Recall*. Adapun TP (*True Positive*) menunjukkan banyaknya sampel positif yang diklasifikasikan benar. TN (*True Negative*) menunjukkan banyaknya sampel negatif yang diklasifikasikan benar. FP (*False Positive*) menunjukkan banyaknya sampel negatif yang diklasifikasikan salah. Sementara, FN (*False Negative*) menunjukkan banyaknya sampel positif yang diklasifikasikan salah.

Perhitungan *Accuracy*, *Precision*, dan *Recall* dengan *confusion matrix* adalah sebagai berikut (Mayadewi & Rosely, 2015) :

Rumus *Accuracy* :

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Rumus *Precision*:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Rumus *Recall* :

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

**2.4. Mean Decrease Accuracy (MDA)**

MDA merupakan metode untuk menghitung besarnya nilai dari *variable importance*. Perhitungan tingkat *variable importance* ini dengan permutasi dan

menggunakan OOB (*Out Of Bag*) untuk mempartisi data sampelnya. Di mana OOB adalah teknik yang digunakan untuk menghitung kesalahan prediksi dan kemudian mengevaluasi variabel importance. Secara sederhana, MDA adalah metode yang menggambarkan penurunan akurasi model dari permutasi setiap fitur. Rumus MDA adalah sebagai berikut

$$MDA(x_j) = \frac{1}{n_{tree}} \sum_{t=1}^{n_{tree}} \frac{\sum_{i \in OOB} I(y_i = f(x_i)) - \sum_{i \in OOB} I(y_i = f(x_i^j))}{|OOB|} \quad (4)$$

di mana:

- $MDA(x_j)$  : *variable importance* dari  $j$   
 $j$  :  $1, 2, 3, \dots, M$   
 $M$  : banyaknya variabel di *tree*  $t$   
 $t \in (1, 2, 3, \dots, n_{tree})$   
 $y_i = f(x_i)$  : *error* OOB pada *tree*  $t$  sebelum nilai permutasi  $x_i$   
 $y_i = f(x_i^j)$  : *error* OOB pada *tree*  $t$  sesudah nilai permutasi  $x_i$

(Shafi et al., 2020)

### 3. Metode Penelitian

#### 3.1 Jenis dan Sumber Data

Penelitian ini menggunakan data primer yang diambil dari rumah tangga di dusun Paok Lombok Timur Kecamatan Suralaga Lombok Timur NTB tahun 2022 dengan jumlah sampel 150 rumah tangga. Data tersebut diperoleh melalui pengisian kuesioner oleh 150 rumah tangga dari berbagai *background* pekerjaan yang berbeda-beda pada tahun 2022. Dari data yang diperoleh, sebanyak 99 rumah tangga masuk kategori miskin (66%) dan sisanya hanya sebanyak 51 rumah tangga saja yang masuk ketegori tidak miskin (34%). Dengan demikian data kemiskinan tersebut merupakan data *imbalanced*.

#### 3.2 Variabel penelitian

Variabel penelitian yang digunakan dalam penelitian ini adalah sebanyak 13 variabel prediktor. Adapun variabel respon dan prediktor yang digunakan adalah sebagai berikut:

**Tabel 3 – Variabel Penelitian**

Variabel	Nama	Keterangan
Y	Status Kemiskinan	0 : Penduduk Tidak Miskin 1 : Penduduk Tidak Miskin
X <sub>2</sub>	Jumlah ART	1: <4 2: >=4
X <sub>3</sub>	Pendidikan Tertinggi KRT	1: SMP ke bawah 2: SMA ke atas
X <sub>4</sub>	Kemampuan Membaca dan Menulis	1: Dapat 2: Tidak Dapat
X <sub>5</sub>	Mordibitas	1: Ada 2: Tidak Ada
X <sub>6</sub>	Cara Berobat	1: Pengobatan Sendiri 2: Puskemas 3: Rumah Sakit
X <sub>7</sub>	Fasilitas Air Bersih	1: Ada 2: Tidak Ada
X <sub>8</sub>	Fasilitas Sanitasi	1: Ada 2: Tidak Ada/Umum
X <sub>9</sub>	Luas Lantai	1: <10 m 2: >=10 m
X <sub>10</sub>	Status Pekerjaan	1: Bekerja Sendiri 2: Buruh 3: Pengusaha 4: Karyawan/PNS
X <sub>11</sub>	Lapangan Usaha	1: Jasa-Jasa 2: Lain-Lain
X <sub>12</sub>	Sektor Lapangan	1: Formal 2: Informal
X <sub>13</sub>	Status Perkawinan	1: Kawin 2: Belum Kawin 3: Janda 4: Duda

### 4. Hasil dan Pembahasan

#### 4.1 Persiapan Data Training dan Data Testing

Data yang akan diolah terdiri dari 2 kelas, yaitu Penduduk Miskin dan Penduduk Tidak Miskin dengan jumlah masing-masing kelas sebanyak 99 dan 51. Perbandingan data antara 2 kelas tersebut cukup jauh, sehingga data dikatakan *imbalance* atau tidak seimbang. Untuk melakukan *balancing* data menggunakan metode *Random Over Sampling* (ROS)

yaitu dengan cara mengambil sampel secara acak (*random sampling*) sebanyak data pada kelas yang lebih banyak. Pembagian jumlah data untuk data training dan data testing dapat dilihat dari Tabel 4 berikut:

**Tabel 4** – Data *Training* dan Data *Testing* Setelah *Balancing Data*

Keterangan	Data <i>Training</i>	Data <i>Testing</i>	Total
Jumlah	139	59	198
Presentase	70%	30%	100%

Berdasarkan Tabel 4, total data yang digunakan pada penelitian ini adalah 198, dengan komposisi 70% untuk data *training* yaitu sebesar 139 data, dan sisanya 59 untuk data *testing*. Pembagian data *training* dan *testing* tersebut dilakukan secara random dengan bantuan *Software R*.

Tahapan pertama sebelum melakukan analisis adalah penyiapan data penelitian, yaitu mengklasifikasikan setiap rumah tangga ke kategori miskin atau tidak dengan cara membandingkan nilai pengeluaran per kapita per bulan dengan Garis Kemiskinan (GK) Kabupaten Lombok Timur tahun 2021, yaitu sebesar Rp. 472.304,-. Jika pengeluaran per kapita per bulan suatu rumah tangga dibawah GK maka rumah tangga tersebut masuk kategori miskin. Selanjutnya dilakukan pengkodean ulang (*recode*) pada variabel prediktor yang bertipe kategorik sesuai Tabel 3.

#### 4.2 Analisis Algoritma *Random Forest*

Analisis di sini dilakukan dengan menentukan berapa banyak pohon yang akan terbentuk (*n<sub>tree</sub>*). Sebelum menentukan pohon, terlebih dahulu menentukan berapa banyak random sample yang diambil untuk setiap percobaan (*m<sub>try</sub>*). *M<sub>try</sub>* digunakan sebagai kandidat pemisah dalam pembentukan pohon. Penentuan nilai *m<sub>try</sub>* yang digunakan dalam penelitian ini adalah :

$$\begin{aligned} Mtry_1 &= \frac{\sqrt{\text{banyak variabel}}}{2} \\ &= \frac{\sqrt{12}}{2} \\ &= 2 \end{aligned}$$

$$\begin{aligned} Mtry_2 &= \sqrt{\text{banyak variabel}} \\ &= \sqrt{12} \end{aligned}$$

$$= 3$$

$$Mtry_3 = \sqrt{\text{banyak variabel}} \times 2$$

$$= \sqrt{12} \times 2$$

$$= 6$$

Dari rumus di atas, didapatkan bahwa nilai *m<sub>try</sub>* adalah 2, 3 dan 6 yang akan digunakan untuk pengujian dalam klasifikasi ini.

#### 4.3 Penentuan Nilai *m<sub>Optimum</sub>*

Pemilihan nilai *m<sub>optimum</sub>* didasarkan pada nilai *error* yang paling kecil sebagaimana dapat dilihat pada tabel 5 berikut ini:

**Tabel 5** – Nilai *Error* untuk masing-masing *m<sub>try</sub>*

<i>m<sub>try</sub></i>	Nilai <i>Error</i>
2	0,2518
3	0,2158
6	0,259

Berdasarkan Tabel 5, didapatkan bahwa nilai *m<sub>try</sub>* optimal adalah sebesar 3. Tahap berikutnya adalah mencari nilai *n<sub>tree</sub>* optimal dengan memasukkan nilai *m<sub>try</sub>* yaitu 2.

#### 4.4 Menumbuhkan Jumlah Pohon

Nilai *n<sub>tree</sub>* yang akan dicobakan yaitu 25, 50, 100, 500 dan 1000 (Yahya et al, 2018). Nilai *n<sub>tree</sub>* tersebut dapat dilihat pada Tabel 6 berikut.

**Tabel 6** – Nilai *Error* Masing-Masing *N<sub>tree</sub>*

<i>n<sub>tree</sub></i>	Nilai <i>Error</i>
25	0,1151
50	0,0791
100	0,0719
500	0,0719
1000	0,0647

Berdasarkan Tabel 6, didapatkan nilai *n<sub>tree</sub>* optimal adalah sebesar 1000 (karena memiliki nilai *error* terkecil yakni 0,0647). Setelah mendapatkan nilai *m<sub>try</sub>* dan *n<sub>tree</sub>* yang optimal, maka langkah berikutnya adalah menggunakan nilai *m<sub>try</sub>* dan *n<sub>tree</sub>* optimal tersebut untuk menentukan prediksi *Random*

Forest pada data *training* dan berikutnya menguji tingkat akurasi dari *Random Forest* dengan menggunakan data *testing*.

#### 4.5 Hasil Prediksi Random Forest

Hasil prediksi dari *Random Forest* dengan menggunakan data *testing* dapat dilihat pada tabel 7 berikut.

**Tabel 7** – Hasil prediksi *Random Forest* dengan Data *Testing*

Data Testing	Prediksi	
	Penduduk Miskin	Penduduk Tidak Miskin
Penduduk Miskin	17	8
Penduduk Tidak Miskin	3	31

Berdasarkan Tabel 7, didapatkan bahwa banyaknya penduduk miskin dari data testing yaitu sebanyak 25, hasil prediksi benar untuk penduduk miskin yaitu sebanyak 17 dan prediksi yang salah sebanyak 8. Selanjutnya banyaknya penduduk tidak miskin dari data testing yaitu sebanyak 34, adapun hasil prediksi benar untuk penduduk tidak miskin sebanyak 31 dan prediksi yang salah yaitu sebanyak 3.

Adapun hasil akurasi untuk data *testing* dapat dihitung sebagai berikut:

$$\begin{aligned}
 \text{Total accuracy rate} &= \frac{\sum(\text{prediksi benar})}{\sum(\text{semua prediksi})} \times 100\% \\
 &= \frac{17 + 31}{59} \times 100\% \\
 &= 81,3\%
 \end{aligned}$$

Berdasarkan perhitungan di atas, hasil akurasi untuk data testing yang diukur melalui *total accuracy rate* yaitu sebesar 81,3%. Ini menunjukkan bahwa algoritma *Random Forest* bisa dikatakan sangat tepat untuk melakukan prediksi dari data *testing* maupun data baru.

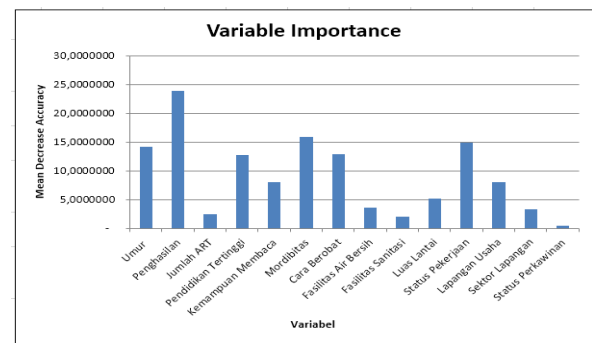
#### 4.6 Importance Variable

Dengan *Random Forest*, dapat pula dilihat variabel-variabel apa saja yang paling berpengaruh sebagaimana Tabel 8 berikut ini.

**Tabel 8** Variable Importance

Variabel	Mean Decrease Accuracy
Umur	14,1782139
Penghasilan	23,9151956
Jumlah ART	2,4094209
Pendidikan Tertinggi	12,7671813
Kemampuan Membaca	7,9959033
Mordibitas	15,9307306
Cara Berobat	12,8459863
Fasilitas Air Bersih	3,6025410
Fasilitas Sanitasi	1,9918810
Luas Lantai	5,0937474
Status Pekerjaan	14,9363359
Lapangan Usaha	7,9935241
Sektor Lapangan	3,3592008
Status Perkawinan	0,4466311

Berikut ini adalah gambar diagram batang yang menggambarkan nilai *Variable Importance* dengan  $m = 3$  dan  $n_{tree} = 1000$ .



**Gambar 2** – Diagram Batang *Variable Importance*

Besarnya pengaruh sebuah variabel dapat diukur dari nilai *Mean Decrease Accuracy* pada tabel 8 sekaligus Gambar 2 di atas. Semakin besar angka *Mean Decrease Accuracy*, maka semakin besar pula peran variabel tersebut dalam mempengaruhi hasil analisis. Adapun variabel yang pengaruhnya paling besar dalam menentukan status kemiskinan seseorang adalah penghasilan, dengan nilai *Mean Decrease Accuracy* sebesar 23,92. Sementara variabel yang pengaruhnya paling kecil adalah status perkawinan, dengan nilai *Mean Decrease Accuracy* sebesar 0,45.

## 5. Kesimpulan dan Saran

Berdasarkan hasil analisis dan pembahasan pada bab sebelumnya, dapat diambil beberapa kesimpulan sebagai berikut.

- Kasus data *imbalanced* pada penelitian ini ditangani dengan metode *Random Over Sampling* (ROS), dengan memilih kelas yang paling banyak yaitu sebanyak 99.
- Klasifikasi menggunakan *Random Forest* dengan nilai *mtry* dan *n tree* optimal masing-masing yaitu  $m = 3$  dan  $n = 1000$  menghasilkan tingkat akurasi sebesar 81,3% yang berarti ketepatan metode klasifikasi *Random Forest* untuk data ini sudah sangat baik. Pengujian algoritma *Random Forest* ini menggunakan data training sebesar 139 dan data *testing* sebesar 59.
- Variabel yang paling berpengaruh dalam menentukan status kemiskinan di dusun Paok Lombok Timur Kec. Suralaga Lombok Timur NTB tahun 2022 berdasarkan analisis *Random Forest* adalah variabel penghasilan dengan dengan nilai *Mean Decrease Accuracy* sebesar 23,92% paling tinggi diantara variabel atribut yang lainnya.

## Ucapan Terima Kasih

Terima kasih disampaikan kepada rekan-rekan dosen khususnya Prodi D3 Statistika Terapan Politeknik Medica Farma Husada Mataram yang telah berkontribusi dalam penulisan jurnal penelitian ini.

## DAFTAR PUSTAKA

- Breiman, L., & Ihaka, R. (1984). *Nonlinear discriminant analysis via scaling and ACE*. Department of Statistics, University of California Davis One Shields Avenue.
- Christy, E., & Suryowati, K. (2021). Analisis Klasifikasi Status Bekerja Penduduk Daerah Istimewa Yogyakarta Menggunakan Metode Random Forest. *Jurnal Statistika Industri dan Komputasi*, 6(1), 69-76.
- Ferezagia, D. V. (2018). Analisis Tingkat Kemiskinan di Indonesia. *Jurnal Sosial Humaniora Terapan*, 1(1), 1-6.
- Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12). Springer Science & Business Media.
- Jatmiko, Y. A., Padmadisastra, S., & Chadidjah, A. (2019). Analisis Perbandingan Kinerja Cart Konvensional, Bagging Dan Random Forest pada Klasifikasi Objek: Hasil Dari Dua Simulasi. *Media Statistika*, 12(1), 1.
- Juliannisa, I. A., & Siswantini, T. (2019). Improving Human Development Index (HDI) by Illiteracy Eradication (Case Study in Cibadak Village). *Jurnal Samudra Ekonomi Dan Bisnis*, 11(1), 29-42.
- KBBI. (2022). *Kamus Besar Bahasa Indonesia (KBBI)*. <https://kbbi.kemdikbud.go.id/>
- Lewis, R. J. (2000). An Introduction to Classification And Regression Tree (CART) Analysis. *Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California*, 14.
- Mauludiyah, K. (2020). Klasifikasi Indeks Pembangunan Manusia Kabupaten/Kota di Indonesia Menggunakan Metode Random Forest. *Doctoral Dissertation*, Muhammadiyah University, Semarang.
- Mayadewi, P., & Rosely, E. (2015). Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining. *SESINDO*.
- Muflikhah, L., Mahmudy, W. F., & others. (2021). *Machine Learning dalam Bioinformatika*. Universitas Brawijaya Press.
- Nuzula, L., Prahutama, A., & Hakim, A. R. (2020). 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*, 9(4), 525-534.
- Otok, B. W., & others. (2015). Random Forest dan Multivariate Adaptive Regression Spline (MARS) Binary Response untuk Klasifikasi Penderita HIV/AIDS di Surabaya. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 3(1).
- Purwa, T. (2019). Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017). *Jurnal Matematika, Statistika dan Komputasi*, 16(1), 58.
- Rianto, M., & Yunis, R. (2021). Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest. *Paradigma - Jurnal Komputer Dan Informatika*, 23(1).
- Saadah, S., & Salsabila, H. (2021). Jurnal Politeknik

- Caltex Riau Prediksi Harga Bitcoin Menggunakan Metode Random Forest (Studi Kasus: Data Acak Pada Awal Masa Pandemic Covid-19). *Jurnal Komputer Terapan* (Vol. 7, Issue 1).
- Sasongko, T. B. (2016). Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA). *Jurnal Teknik Informatika Dan Sistem Informasi*, 2(2).
- Shafi, A. S. M., Molla, M. M., Jui, J. J., & Rahman, M. M. (2020). Detection of colon cancer based on microarray dataset using machine learning as a feature selection and classification techniques. *SN Applied Sciences*, 2(7), 1–8.
- Syukron, M., Santoso, R., & Widiharih, T. (2020). Perbandingan Metode Smote Random Forest dan Smote Xgboost untuk Klasifikasi Tingkat Penyakit Hepatitis C pada Imbalance Class Data. *Jurnal Gaussian*, 9(3), 227–236.
- Tahyudin, I., Putra, I. M., & Syafa'at, A. Y. (2021). *Data Mining Dan Data Warehouse Menggunakan Aplikasi KNIME* (Vol. 1). Zahira Media Publisher.
- Tarigan, H., Sinaga, J. H., & Rachmawati, R. R. (2020). Dampak Pandemi Covid-19 Terhadap Kemiskinan di Indonesia. *Pusat Sosial Ekonomi Dan Kebijakan Pertanian*, 3, 457–479.
- Yacoub, Y. (2012). Pengaruh Tingkat Pengangguran Terhadap Tingkat Kemiskinan Kabupaten/Kota di Provinsi Kalimantan Barat. *Jurnal Eksos*, 8(3).
- Yahya, S. A., & others. (2018). *Klasifikasi Ketepatan Lama Studi Mahasiswa Menggunakan Metode Support Vector Machine dan Random Forest (Studi Kasus: Data Lama Studi Alumni Universitas Islam Indonesia Tahun Kelulusan 2000-2017)*.