Prosiding Seminar Nasional Sains dan Teknologi Seri III
Fakultas Sains dan Teknologi, Universitas Terbuka
Vol. 2 No. 1 (2025)
e-ISSN: 3047-6569

KLASIFIKASI POLA KONSUMSI DAUN KELOR DI KOTA SURABAYA MENGGUNAKAN METODE *RANDOM FOREST*

Desy Nuriatul Fajariyah[1*], Ika Nur Laily Fitriana[2]
[1]Mahasiswa Program Studi Statistika, Universitas Terbuka, Surabaya
[2]Dosen Program Studi Statistika, Universitas Terbuka, Surabaya

**Penulis korespondensi: desynfajariyah@gmail.com*

ABSTRAK

Pola konsumsi daun kelor di Kota Surabaya menjadi salah satu topik yang menarik untuk dianalisis. Hal ini karena, Surabaya merupakan salah satu kota besar di Indonesia dengan populasi yang besar dan karakteristik sosial, ekonomi, dan budaya yang beragam. Kelor sering dijulukin *superfood* karena manfaat dari akar hingga ke daunnya. Namun, manfaat daun kelor tidak diimbangin dengan pemahaman tentang pengolahan daun kelor sebagai bahan masakan. Sehinga, tujuan pada penelitian ini untuk mengklasifikasi pola konsumsi daun kelor dengan metode *random forest*. *Random forest* adalah metode klasifikasi dari gabungan pohon klasifikasi (CART) yang saling independen. Sampel yang digunakan sebanyak 60 orang dengan kriteria pernah konsumsi daun kelor. Pengambilan sampel dilakukan dengan sampling klaster dua tahap secara sampling acak sederhana (SAS). Data Jenis Kelamin, Umur, Pendidikan Terakhir, Pekerjaan, Ketersediaan Daun Kelor, dan Harga Daun Kelor diteliti sebagai parameter pola konsumsi daun kelor. Hasil analisis menunjukkan bahwa dengan menggunakan random forest memiliki nilai akurasi 78% dari data total. Kombinasi data pada penelitian ini sebanyak 80% dari data digunakan untuk training sedangkan 20% dari data digunakan untuk testing.

Kata kunci: Daun Kelor, Pola Konsumsi, dan *Random Forest*

1 PENDAHULUAN

Pola konsumsi masyarakat menjadi refleksi dari kompleksitas dinamika sosial, budaya, dan ekonomi yang membentuk suatu bangsa (Anggiat, 2022). Perubahan dalam pola konsumsi masyarakat bukan hanya sekedar fenomena. Namun, pergeseran makanan tradisional dengan makanan *junk food* dan *international food*. Pergeseran ini dipicu oleh banyak faktor, termasuk globalisasi, perkembangan teknologi, urbanisasi, dan transformasi gaya hidup masyarakat. Perkembangan globalisasi dan teknologi yang semakin memukau dengan kemudahan akses informasi perlahan memengaruhi pola konsumsi masyarakat. Hal ini, terjadi juga pada daun kelor yang sudah jarang dimanfaatkan sebagai bahan makanan.

Tanaman kelor dapat tumbuh dan berkembang di daerah tropis seperti Indonesia. Tanaman kelor dapat tumbuh mulai dari dataran rendah sampai ketinggian 700 m di atas permukaan laut. Tanaman kelor adalah salah satu tanaman perdu dengan ketinggian 7-11 meter, tahan terhadap musim kering dengan toleransi terhadap kekeringan sampai 6 bulan serta mudah dibiakkan dan tidak memerlukan perawatan yang intensif. Kelor di Indonesia memiliki beragam nama di beberapa wilayah di antaranya kelor (Jawa, Sunda, Bali, Lampung), maronggih (Madura), moltong (Flores), keloro (Bugis), ongge (Bima), murong atau barunggai (Sumatera) dan hau fo (Timur). Kelor merupakan spesies dari keluarga monogenerik yang paling banyak dibudidayakan, yaitu

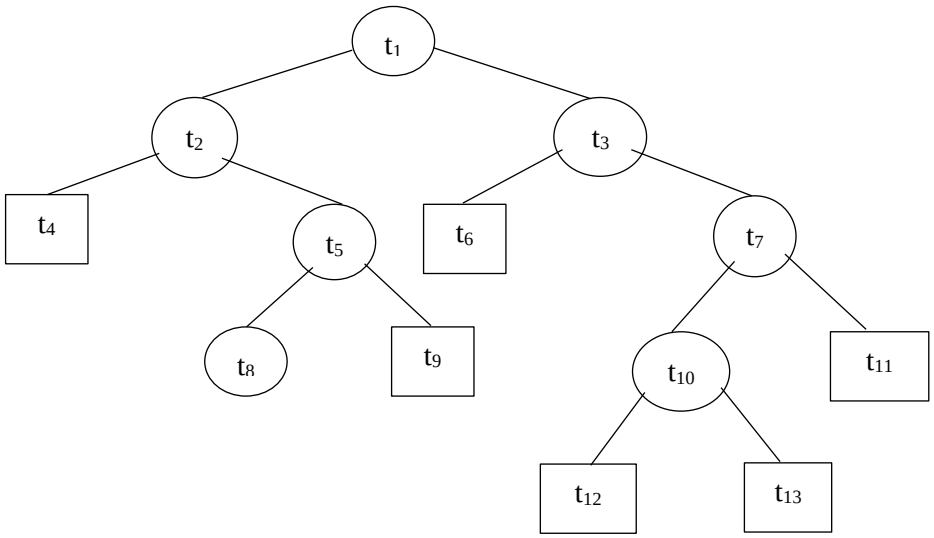
Moringaceae yang berasal dari India sub- Himalaya, Pakistan, Bangladesh dan Afghanistan (Luluk, 2021).

Moringa oleifera L. dijuluki sebagai "*The Miracle Plant*" karena memiliki banyak manfaat pada semua bagian tanamannya. Bagian akar kelor bermanfaat sebagai anti *scorbutic* dapat mengurangi iritasi. Bagian daun dapat dimanfaatkan sebagai antitumor, menurunkan tekanan darah, antioksidan, anti inflamasi, radio-protektif, dan bersifat diuretik. Tanaman kelor mengandung 46 jenis antioksidan dan lebih dari 90 nutrisi. Selain itu, ada 36 senyawa anti inflamasi (Sulasmi *et. al*., 2023). Kandungan nutrisi daun kelor yang begitu besar namun, tidak diimbangin dengan pemanfaatan daun kelor secara optimal sebagai pelengkap masakan sehari-hari. Hal ini, disebabkan oleh perubahan pola konsumsi terutama pada sayuran. Penelitian ini, bertujuan untuk mengklasifikasin pola konsumsi daun kelor bagi masyarakat di Kota Surabaya dengan menggunakan *random forests*.

Random forests adalah suatu metode klasifikasi yang terdiri dari gabungan pohon klasifikasi (CART) yang saling independen. Prediksi klasifikasi diperoleh melalui proses voting (jumlah terbanyak) dari pohon-pohon klasifikasi yang terbentuk. *Random forests* merupakan pengembangan dari metode *ensemble* yang pertama kali dikembangkan oleh Leo Breiman (2001) dan digunakan untuk meningkatkan ketepatan klasifikasi. Bila dalam proses *bagging* digunakan resampling bootstrap untuk membangkitkan pohon klasifikasi dengan banyak versi yang kemudian mengkombinasikannya untuk memperoleh prediksi akhir, maka dalam *random forests* proses pengacakan untuk membentuk pohon klasifikasi tidak hanya dilakukan untuk data sampel saja melainkan juga pada pengambilan variabel prediktor. Sehingga, proses ini akan menghasilkan kumpulan pohon klasifikasi dengan ukuran dan bentuk yang berbeda-beda. Hasil yang diharapkan adalah suatu kumpulan pohon klasifikasi yang memiliki korelasi kecil antar pohon. Korelasi yang kecil akan menurunkan hasil kesalahan prediksi *random forests* (Breiman, 2001).

Sistem pohon keputusan (*decision trees*) berbeda proses yang terjadi pada metode regresi logistik, pohon keputusan (*decision trees*) bukanlah model parametrik sehingga tidak memiliki parameter. Pohon disajikan berupa rangkaian yang terdiri dari pilihan dan keputusan. Keputusan digambarkan seperti pohon asli namun ditampilkan secara terbalik. Ada dua macam pohon keputusan yakni pohon klasifikasi dan pohon regresi. Pada penelitian ini akan digunakan pohon klasifikasi.

Pohon klasifikasi seperti yang diilustrasikan pada Gambar 1. dibuat dari kumpulan simpul yang saling terhubung. Masing-masing simpul mewakili sebuah keputusan, dimana keputusan didasarkan pada salah satu variabel prediktor. Untuk setiap keputusan, proses akan turun satu level dari pohon ke simpul lainnya hingga mencapai simpul terminal yang dinamakan daun (*leaf*). Daun berisi level dari respon dan diakhiri dengan menentukan kelas.



Gambar 1. Ilustrasi Pohon Klasifikasi

Pembentukan pohon klasifikasi dilakukan dengan memisahkan sebuah masalah menjadi sub masalah. Keputusan pada simpul teratas akan dipisahkan menjadi dua simpul dimana kedua simpul tersebut terdiri dari pernyataan benar dan salah. Algoritma yang digunakan untuk memilih keputusan misalnya indeks Gini yang didefinisikan sebagai berikut:

$$G = \sum_{i=1}^{I} p_{mi}(1 - p_{mi}) \tag{1}$$

dimana p_{mi} adalah pengamatan pada kelas i dan simpul ke m (Kalmar, 2014). Persamaan (1) merupakan ukuran keheterogenan simpul. Himpunan bagian yang dihasilkan dari proses pemilihan harus lebih homogen dibandingkan simpulinduknya. Kemudian Langkah selanjutnya adalah menentukan kriteria *goodness of split*. Cara yang dilakukan adalah dengan mengoptimumkan fungsi indeks Gini.

Persamaan (1) adalah pernyataan derajat kedua sehingga memiliki satu titik ekstrim. Dengan menurunkan persamaan (1) hingga turunan kedua, titik ekstrim akan bisa diperoleh. Tunan pertama dari persamaan (1) adalah sebagai berikut:

$$\frac{d}{dp_{mi}} p_{mi}(1 - p_{mi}) = 1 - 2p_{mi} \tag{2}$$

saat persamaan (2) sama dengan 0, maka akan diperoleh titik ekstrim, yakni:

$$1 - 2p_{mi} = 0 \tag{3}$$

$$2p_{mi} = 1$$

$$p_{mi} = 0{,}5$$

sehingga 0,5 adalah posisi dari titik ekstrim. Turunan kedua dari persamaan (1) adalah

$$\frac{d}{dp_{mi}} 1 - 2p_{mi} = -2 \tag{4}$$

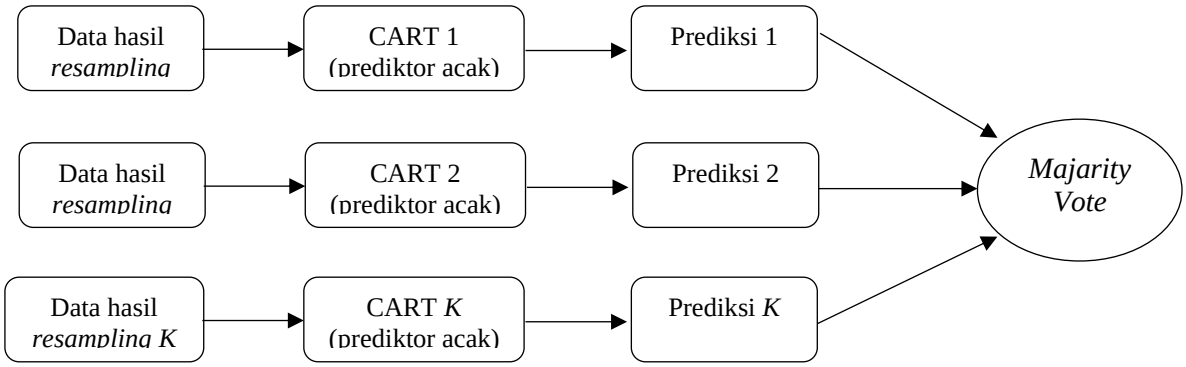
Nilai dari turunan kedua adalah negative sehingga titik p_{mi} = 0,5 adalah titik maksimum. Nilai rasio harus berada pada interval 0 sampai 1 sehingga persamaan diatas akan minimum dengan nilai yang mendekati 0 atau 1. Hal ini menunjukkan bahwa dengan rasio yang tinggi atau rendah, indeks Gini akan bernilai paling kecil. Simpul dengan nilai rasio yang tinggi atau rendah akan memiliki indeks Gini yang kecil. Pemisah dengan indeks Gini yang kecil akan menjadikan simpul yang memisahkan pengamatan menjadi simpul yang homogen.

Pengembangan random forests yang dilakukan dari proses bagging yaitu terletak pada proses pemilihan pemilah. Pada random forests, pemilihan pemilah hanya melibatkan beberapa variabel prediktor yang terambil secara acak. Algoritma Random Forests dijelaskan Langkah-langkah sebagai berikut:

a. Mengambil *n* data sampel dari *dataset* awal dengan menggunakan teknik *resampling bootsrap* dengan pengambilan
b. Menyusun pohon klasifikasi dengan setiap *dataset* hasil *resampling bootstrap*, dengan penrntuan pemilihan terbaik didasarkan pada variabel prediktor yang diambil secara acak. Jumlah variabel yang diambil secara acak dapat mennetukan melalui perhitungan $\log_2 (M+1)$, dimana *M* adalah banyak variabel prediktor (Breiman, 2001) atau menggunakan $\sqrt{p}$, dimana *p* adalah banyak variabel prediktor (Genuer *et al*., 2009).
c. Melakukan prediksi klasifikasi data sampel berdasarkan pohon klasifikasi yang terbentuk.
d. Mengulangi Langkah a sampai Langkah c hingga diperoleh sejumlah pohon klasifikasi yang diinginkan. Perulangan dilakukan sebanyak *K* kali.
e. Melakukan prediksi klasifikasi data sampel akhir dengan mengombinasikan hasil prediksi pohon klasifikasi yang diperoleh berdasarkan aturan *majority vote*.

Analisis dengan menggunakan metode *random forests* dimulai dari pengambilan data dengan teknik resampling bootstrap. Bootstrap adalah suatu metode yang dapat bekerja tanpa membutuhkan asumsi distribusi karena sampel asli digunakan sebagai populasi. Bootstrap digunakan untuk mencari distribusi sampling dari suatu estimator dengan prosedur resampling dengan pengembalian dari data asli (Sungkono, 2013). Algoritma dari resampling bootstrap diilustrasikan pada Gambar 2. dengan langkah-langkah sebagai berikut:

a. Mengkonstruksi distribusi empiris $\hat{F}_n$dari suatu sampel dengan memberikan probabilitas 1/*n* pada setiap X_i dimana $i = 1,2,\ldots,n$.
b. Mengambil sampel bootstrap berukuran *n* secara random dengan pengembalian dari distribusi empiris $\hat{F}_n$ sebut sebagai sampel bootsrap pertaman X^{*1}.
c. Menghitung statistik $\hat{\theta}$ yang diinginkan dari sampel bootstrap X^{*1} sebut sebagai $\hat{\theta}_1^*$.
d. Mengulangi Langkah b dan Langkah c hingga *B* kali diperoleh $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots \hat{\theta}_B^*$.
e. Mengkonstruksi suatu distribusi probabilitas dari $\hat{\theta}_B^*$ dengan memberikan probabilitas 1/*B* pada setiap $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots \hat{\theta}_B^*$. Distribusi tersebut merupakan estimator bootsrap untuk distribusi sampling $\hat{\theta}$ dan dinotasikan dengan $\hat{F}^*$.
f. Pendekatan estimasi bootsrap adalah $\hat{\theta}^* = \sum_{b=1}^{B} \hat{\theta}_b^* \frac{1}{B}$



Gambar 2. *Random Forest*

2 METODE

2.1 Sumber Data

Data yang digunakan dalam penelitian kali ini diperoleh dari data primer yakni survei ke masyarakat Surabaya dengan kriteria pernah konsumsi daun kelor. Survei dilaksanakan pada tanggal 14 Juni – 15 Agutus 2024 dan jumlah responden sebanyak 60 orang.

Gambar 3. Peta Kota Surabaya (Novriyadi., 2023)

2.2 Variabel Penelitian

Variabel penelitian dalam dataset akan ditunjukkan pada Tabel 1 adalah sebagai berikut:

Tabel 1. Variabel Penelitian

Variabel	Indikator	Kategori	Skala
Pola Konsumsi Daun Kelor (Y)	Frekuensi Konsumsi Daun Kelor, Waktu Konsumsi Daun Kelor, dan Jenis Olahan Daun Kelor	1 : Cukup (Total Skor 5-7) 2 : Baik (Total Skor 8-11) 3 : Lebih Baik (Total Skor 12-15)	Ordinal
Jenis Kelamin (X_1)			Nominal
Umur (X_2)			Rasio
Pendidikan Terakhir (X_3)			Ordinal
Pekerjaan (X_4)			Nominal
Ketersedian Kelor (X_5)			Nominal
Harga (X_6)			Rasio

2.3 Ukuran Sampel

Penelitian ini merupakan penelitian eksploratori karena dilakukan untuk pertama kalinya. Sehingga ukuran populasi dan sampel sama yakni 60 responden. Ukuran tersebut berdasarkan teori limit pusat. Teknik pengambilan dengan klister dua tahap secara acak sederhana. Dimana kerangka sampling yang diambil sebagai berikut:

Tabel 2. Kerangka Sampling

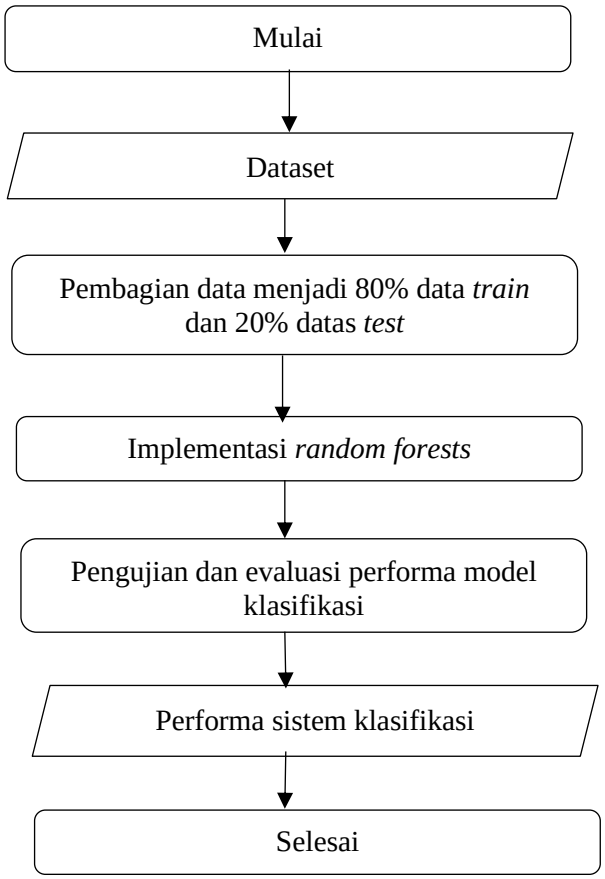
Klaster 1 (Kecamatan)	**Klaster 2 (Kelurahan)**
Dukuh Pakis	Dukuh Kupang
	Dukuh Pakis
	Gunung Sari
	Pradah Kalikendal
Gubeng	Airlangga
	Baratajaya
	Gubeng
	Kertajaya
	Mojo
	Pucang Sewu
Sukolilo	Gebang Putih
	Keputih
	Klampis Ngasem
	Medokan Semampir
	Menur Pumpungan
	Nginden Jangkungan
	Semolowaru

Tabel 2 menjelaskan bahwa tahapan klaster dua tahap secara sampling acak sederhana adalah:

- Pada tahap pertama, dari 31 kecamatan di Surabaya terambil 3 kecamatan, yakni Dukuh Kupang, Gubeng, dan Sukolilo.
- Pada tahap kedua, dari 17 kelurahan di tiga kecamatan terpilih, terambil 6 kelurahan, yakni Dukuh Kupang, Airlangga, Mojo, Keputih, Klampis Ngasem, dan Semolowaru. Dimana setiap kelurahan diambil 10 responden sebagai perwakilan dari populasi.

2.4 Implementasi Model Klasifikasi

Setelah pengumpulan data primer dari survei selanjutnya, proses ini terdiri dari pembersihan data yang tidak diperlukan dalam model klasifikasi dan memanipulasi data yang bernilai kosong atau data yang tidak terukur menggunakan nilai mean atau median berdasarkan distribusinya. Untuk mempermudah tahapan metode *random forests* ditampilkan diagram alir berikut:



Gambar 4. Diagram Alir Implementasi *Random Forest*

2.5 Pengujian dan Evaluasi

Pengujian dan evaluasi untuk melihat seberapa baik model bekerja pada data pola konsumsi daun kelor. Proses pengujian pada sistem dialkukan dengan mengklasifikasikan seluruh data pengujian ke dalam tiga kategori pola konsumsi daun kelor. Evaluasi akan dilakukan dengan memanfaatkan pengukuran dengan konsep perhitungan Confusion Matrix. Konsep ini akan membandingkan keluaran model dengan label data aktual, sehingga menghasilkan nilai metrik akurasi, presisi, recall, dan f1-score dari masing-masing model sesuai dengan persamaan (5) hingga persamaan (8).

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Presisi = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

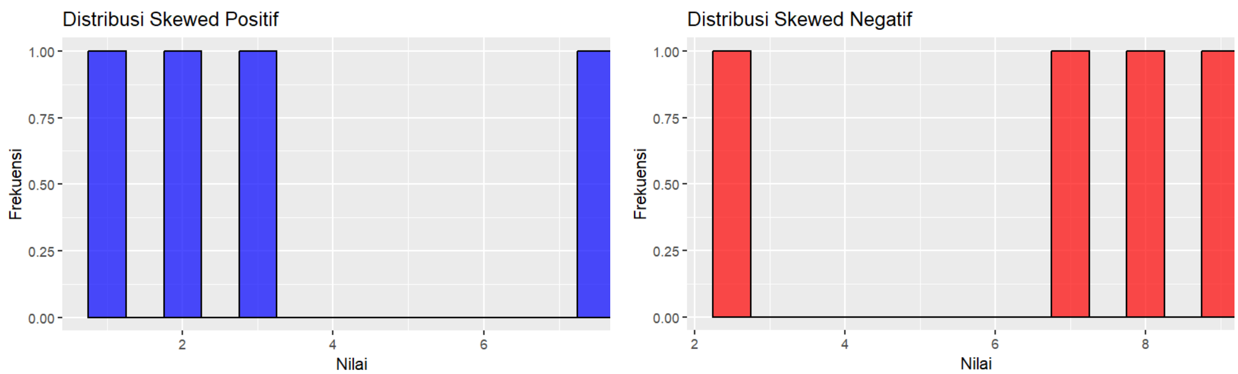
$$F - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{8}$$

(Larasati, *et al*. 2012).

3 HASIL DAN PEMBAHASAN

3.1 Distribusi Data

Manipulasi data dilakukan sesuai dengan karakteristik dari masing-masing parameter pola konsumsi daun kelor. Nilai-nilai yang dimanipulasi pada parameter dengan kemiringan "skewed positif" menggunakan distribusi gamma dan "skewed negative" dengan 10 – nilai skewed positif.



Gambar 5. Distribusi Parameter Data

Gambar 5 menunjukkan bahwa skewed positif sebagian besar berada di rentang rendah, tetapi ada beberapa nilai tinggi. Sedangkan, skewed negative berbanding terbalik dengan skewed positif. Sehingga, ukuran data *training* sebesar 80% dan data *test* sebesar 20%.

3.2 *Random Forests*

Proses selanjutnya adalah analisis dataset dengan metode *random forests* diinisialisasi di awal pembuatan model. Berikut *output* analisis *random forests* dengan menggunaka *software* RStudio.

```
Call:
 randomForest(formula = Pola.Konsumsi ~ ., data = train_data,       ntree = 100, mtry = 3, importance = TRUE)
               Type of random forest: classification
                     Number of trees: 100
No. of variables tried at each split: 3

        OOB estimate of  error rate: 14.17%
Confusion matrix:
            Baik Cukup  Baik  Sangat Baik  class.error
Baik            23        1        16         0.425
Cukup Baik       0       40         0         0.000
Sangat Baik      0        0        40         0.000
```

Gambar 6. Ketepatan Klasifikasi Model

Gambar 6 menjelaskan bahwa nilai OOB (*Out of Bag*) sebesar 14,17% artinya model memprediksi dengan benar 85,83% dari data *training*. Sedangkan pada matriks *confusion* yang memiliki presentase kesalahan tertinggi pada kategori "Baik" sebesar 42,5%. Hal ini, menunjukkan bahwa prediksi untuk kategori "Baik" tidak akurat sebesar 42,5% namun, pada kategori "Cukup Baik"

dan “Sangat Baik” diprediksi sangat baik. Sehingga, model perlu diperbaiki dalam memprediksi kategori “Baik”.

Confusion Matrix and Statistics

	Reference		
Prediction	Baik	Cukup Baik	Sangat Baik
Baik	4	0	0
Cukup Baik	1	9	0
Sangat Baik	4	0	9

Overall Statistics

Accuracy : 0.8148
95% CI : (0.6192, 0.937)
No Information Rate : 0.3333
P-Value [Acc > NIR] : 3.789e-07

Kappa : 0.7222

Mcnemar's Test P-Value : NA

Statistics by Class:

	Class: Baik	Class: Cukup Baik	Class: Sangat Baik
Sensitivity	0.4444	1.0000	1.0000
Specificity	1.0000	0.9444	0.7778
Pos Pred Value	1.0000	0.9000	0.6923
Neg Pred Value	0.7826	1.0000	1.0000
Prevalence	0.3333	0.3333	0.3333
Detection Rate	0.1481	0.3333	0.3333
Detection Prevalence	0.1481	0.3704	0.4815
Balanced Accuracy	0.7222	0.9722	0.8889

Gambar 7. Akurasi Model

Gambar 7 menunjukkan nilai akurasi sebesar 0,8148 artinya model berhasil memprediksi dengan benar sebesar 81,48% dari keseluruhan data dengan interval kepercayaan untuk akurasi model berada di antara 61,92% dan 93,7%. Nilai Kappa sebesar 0,7222 yang berkisar antara -1 hingga 1 yang menunjukkan bahwa model baik.

4 KESIMPULAN

Data Pola Konsumsi Daun Kelor yang meliputi Jenis Kelamin, Umur, Pendidikan Terakhir, Pekerjaan, Ketersedian Daun Kelor, dan Harga dapat digunakan untuk mengklasifikasi pola konsumsi daun kelor. Model *random forests* dengan ukuran sampel 60 responden yang tersebar di enam kelurahan di Kota Surabaya menghasilkan akurasi sebesar 81,48%. Untuk meningkatkan

penelitian ini agar digunakan dalam kondisi nyata dan memberikan kontribusi di lapangan. Saran untuk peneliti selanjutnya adalah dengan menambahkan variabel lain untuk melengkapi pola konsumsi daun kelor.

UCAPAN TERIMAKASIH

Ucapan terimakasih kepada Alm bapak, ibu, dan adek yang menanamkan kepribadian positif. Dan seluruh keluarga yang pengertian dengan kondisi peneliti. Serta tidak lupa kepada dosen jurusan statistika UT yakni ibu Ika yang memberikan arahan dan bimbingan hingga rampung penelitian ini.

DAFTAR PUSTAKA [TNR 12, CETAK TEBAL, KAPITAL DAN RATA KIRI KANAN]

Anggiat, H. (2022). Analisis Perubahan Pola KOnsumsi Masyrakata dan Implikasinya Terhadap Sektor Perekonomian. *Jurnal Price : Ekonomi dan Akutansi Vol 01, No 01*. https://ejournal.seaninstitute.or.id/index.php/jecoa/

Breiman, L. (2001). *Random Forests. Machine Learning*. 45(1). 5-32.

Breiman, L., Friedman, J. H., Olshen, R. A.,dan Stone, C. J.. (1993). *Classification and Regression Trees*. New York : Chapman Hall.

Darmawan, A. (2014). *Deteksi Dini Penyakit Kanker Leher Rahim (Serviks) di Kota Bogor Mengunakan Support Vector Machine (SVM)*. Surabaya : Institut teknologi Sepuluh Nopember.

Geneur, R., Poggi, J. M.,dan Malot, C., T. (2009). *Variable Selection using Random Forests. France : Laboratoire de Mathematiques*, Universite Paris-Sud 11, Bat. 425, 91405.

Larasati., Mardhani. (2012). Klasifikasi Level Banjir Menggunakan Random Forests dan Support Vector Machine. *Indonesia Journal of Electronics and Instrumentation System (IJEIS). https://etd.repository.ugm.ac.id/home/detail_pencarian_downloadfiles/1375893*

Luluk, Sutji Marhaeni. (2021). Daun Kelor (Moringa Oleifera) Sebagai Sumber Pangan Fungsional Dan Antioksidan. *Jurnal Agrisia- Vol. 13 No. 2*. https://ejournal.borobudur.ac.id/index.php/3/article/view/882

Novriyadi. (2023). Peta Surabaya Lengkap dengan Nama Kota. https://www.lamudi.co.id/journal/peta-surabaya/.

Sulasmi., Alfina., Bunga., *et al*. (2023). Pemanfaatan Daun Kelor (Moringa Oleifera L) Untuk Mengatasi Anemia. *Jurnal Pengemas Kesehatan Vol 02 No 01*. https://jpk.stikesbup.ac.id/index.php/jpk/article/view/18

Sungkono, J. (2013). *Resampling Bootstrap Pada R. Magistra* No. 84, Th. XXV. ISSN : 0215-9511.