

Algoritma Genetika Optimalisasi Fitur Seleksi Pada Kismis Di Turki Menggunakan Model Random Forest

Trisna*, Hafidh Ashil, Serly Agustin, Yudi Ramadhani

Fakultas Teknik Informasi, Universitas Adhirajasa Reswara Sanjaya

E-mail Korespondensi : hadiyantitrisna@gmail.com

History Artikel

Diterima : 18 Januari 2023 Disetujui : 03 Maret 2023 Dipublikasikan : 24 April 2023

Abstract

There are many types of raisins, including golden and dark brown varieties. The two raisins are not the same, even though their first impressions are similar. Currently, more than 1 million tons of raisins are produced worldwide. In most parts of the world, raisins are consumed by people of all cultures and social classes. World raisin production is relatively high, especially in Indonesia. The purpose of this study is to classify the raisin dataset using data mining, so that the highest accuracy value of the five algorithms and two types of optimization will be compared which will later be compared. In this study, researchers used the classification method with the RF algorithm and optimizing the GA feature selection which was carried out to refine the training data and test the previous data using cross validation and validation with a split ratio of 0.5 – 0.9. The accuracy and AUC values produced by the RF method are 85.56% for accuracy and 0.922 for AUC. to evaluate the validation of split operators with split ratios ranging from 0.5 to 0.9 using the RF method. Based on the tests carried out, the average accuracy and AUC values were 84.52% and 0.916, respectively. Optimization of feature selection using the GA method utilizes data validation with 10-Fold-Validation to improve the performance of the accuracy value of the RF algorithm. This results in a value of 87.56% for accuracy and 0.928 for AUC.

Keywords: Data Mining; Genetic Algorithm; Raisin; Random Forest;

Abstrak

Ada banyak jenis kismis, termasuk varietas emas dan coklat tua. Kedua kismis itu tidak sama, meski kesan pertama mereka mirip. Saat ini, lebih dari 1 juta ton kismis diproduksi di seluruh dunia. Di sebagian besar dunia, kismis dikonsumsi oleh orang-orang dari semua budaya dan kelas sosial. Produksi kismis dunia relatif tinggi, khususnya di Indonesia. Tujuan dari penelitian ini yaitu untuk mengklasifikasi dataset raisin menggunakan data mining, sehingga memunculkan nilai akurasi yang paling tinggi dari lima algoritma dan dua tipe optimasi yang akan dikomparasi nantinya. Pada penelitian ini, peneliti menggunakan metode klasifikasi dengan algoritma RF dan optimasi fitur seleksi GA yang dilakukan pemisahan data training dan data testing sebelumnya menggunakan cross validation dan split validation rasio 0,5 – 0,9. Nilai akurasi dan AUC yang dihasilkan oleh metode RF adalah 85.56% untuk akurasi dan 0.922 untuk AUC. Untuk mengevaluasi operator validasi split dengan rasio split mulai dari 0,5 hingga 0,9 menggunakan metode RF. Berdasarkan pengujian yang dijalankan, rata-rata akurasi dan nilai AUC masing-masing adalah 84.52% dan 0.916. Optimalisasi fitur seleksi dengan metode GA memanfaatkan validasi data dengan 10-Fold-Validation dilakukan untuk meningkatkan performa nilai akurasi algoritma RF. Hal ini menghasilkan nilai sebesar 87.56% untuk akurasi dan 0.928 untuk AUC.

Kata Kunci: Data Mining; Algoritma Genetika; Kismis; Random Forest;

How to Cite: **Trisna** (2023). Algoritma Genetika Optimalisasi Fitur Seleksi Pada Kismis Di Turki Menggunakan Model Random Forest. *KOMPUTEK : Jurnal Teknik Universitas Muhammadiyah Ponorogo* Vol 7 (1): Halaman 52-62

© 2023 Universitas Muhammadiyah Ponorogo. All rights reserved

ISSN 2614-0985 (Print)

ISSN 2614-0977 (Online)

PENDAHULUAN

Kismis adalah sumber karbohidrat terkonsentrasi dan makanan ringan kaya zat besi, potasium, serat, dan antioksidan. (Karimi, 2011; Kemal Üniversitesi Ziraat Fakültesi Dergisi et al., n.d.) Salah satu negara yang memimpin dunia dalam produksi anggur adalah Turki. Turki memproduksi sekitar 30% dari anggur untuk meja, 37% untuk barang kering, 3% untuk anggur, dan 30% untuk produk lain (Kemal Üniversitesi Ziraat Fakültesi Dergisi et al., n.d.) Kismis salah satu buah kering yang paling sehat dan bergizi adalah kismis. Anggur kering Timur Tengah ini terus mendapatkan popularitas di negara-negara di seluruh Eropa. Nilai gizi kismis bermanfaat untuk menjaga kesehatan fisik. Selain itu, kismis populer di kalangan masyarakat umum karena mengandung nutrisi penting. Setiap orang yang mengonsumsi kismis bermanfaat sebagai sumber tenaga agar tubuh menjadi lebih kuat dan sehat, menurut Nutrisurvey.

Kismis memiliki banyak varian, ada jenis Kismis emas dan coklat tua. Meski sekilas tampak sama, kedua kismis itu tidak identik. Saat ini, produksi kismis dunia melebihi 1 juta ton, Konsumsi kismis terjadi di semua budaya dan berbagai struktur sosial di sebagian besar wilayah dunia. Produksi kismis cukup tinggi di seluruh dunia, termasuk Indonesia Kemungkinan bisa mengembangkan bahan olahan baru. Kismis yang kaya akan efek untuk kesehatan gigi, pengobatan anemia, dan sifat antibakteri dapat diolah dan digunakan sebagai bahan makanan yang dapat dimakan tanpa memandang usia. Kismis berpotensi menggantikan berbagai produk susu karena kandungan nutrisi kismis yang lebih tinggi berupa magnesium dan zat besi dibandingkan buah lainnya. (Rosyada, 2019)

Data Mining adalah proses mencari data melalui database yang sudah ada sebelumnya yang cukup besar untuk menghasilkan informasi baru. Teknologi yang kuat memiliki banyak potensi untuk membantu perusahaan berkonsentrasi pada informasi terpenting di data warehouse. Metode dan teknologi data mining akan

membantu bisnis menjadi lebih proaktif dan membuat keputusan berbasis pengetahuan yang lebih baik, yang akan membantu mereka meramalkan tren masa depan. Teknik data mining dapat memberikan jawaban atas masalah terkait bisnis yang sebelumnya terlalu memakan waktu untuk ditangani. (Osman, 2019) Volume data yang besar dapat diproses dan digunakan dengan lebih efektif berkat data mining, yang berakar pada Artificial Intelligence (AI), khususnya dalam Machine Learning (ML) dan analisis statistik untuk memecahkan masalah yang memerlukan prediksi, klasifikasi, dan segmentasi. (Meira, 2020)

Machine Learning bertujuan untuk memungkinkan komputer belajar tanpa pemrograman eksplisit. Penerapan Machine Learning (ML) mengajarkan komputer cara menangani data dengan lebih efektif. Terkadang, bahkan setelah melihat datanya, kita tidak dapat mengevaluasi atau mengekstrapolasi informasinya. Dalam situasi itu machine learning dapat digunakan untuk membantu mengevaluasi data agar lebih efektif. Ketersediaan datasets dalam jumlah besar telah meningkatkan kebutuhan untuk machine learning. Machine learning menggunakan berbagai teknik untuk mengatasi masalah data. Seperti misalnya, supervised learning, un-supervised learning, semi-supervised learning, reinforcement learning, multi-task learning, ensemble learning, neural network, dan instance based learning. (Masesh, 2019)

Klasifikasi adalah teknik yang digunakan untuk memperkirakan kelas atau properti dari setiap objek data. Setelah tahap persiapan data selesai dilakukan, selanjutnya dilakukan tahap klasifikasi. Duplikasi data diperiksa pada tahap preprocessing menggunakan metrik string Levenshtein, pemilihan fitur, dan penanganan missing value. (Faid et al., 2019)

Random Forest adalah prosedur machine learning populer yang dapat digunakan untuk mengembangkan model prediksi. Classification dan regression trees, yang merupakan model langsung yang menggunakan pemisahan biner pada variabel

prediktor untuk mendapatkan prediksi hasil, dikelompokkan bersama dalam random forest. Dalam pengaturan random forest, beberapa classification dan regression trees dibangun menggunakan dataset pelatihan yang dipilih secara acak dan subset variabel prediktor yang dipilih secara acak untuk hasil pemodelan. Untuk memberikan prediksi pada setiap pengamatan, hasil dari setiap pohon digabungkan. Hasilnya, dalam hal kategorisasi, random forest seringkali menawarkan akurasi yang lebih baik daripada model lainnya. (Speiser, 2019)

Feature selection adalah proses memilih subset dari fitur yang bersangkutan untuk digunakan dalam pembuatan model. Hal ini bertujuan untuk meningkatkan kualitas data. (Almomani, 2020) Ada berbagai teknik pemilihan fitur yang telah diusulkan dalam literatur. (Kumar, 2014) Feature Selection adalah Salah satu teknik untuk meningkatkan kinerja classifier. Teknik ini berpusat pada penyempitan ruang fitur yang sangat besar, khususnya dengan menghilangkan fitur yang tidak relevan dan dengan menggunakan algoritma pemilihan fitur terbaik untuk meningkatkan akurasi. Salah satu teknik optimasi yang dapat digunakan untuk masalah seleksi fitur adalah algoritma genetika (GA). (Somantri & Khambali, 2017)

Genetic algorithm adalah model komputasi evolusi biologis. Baik sebagai teknik pencarian masalah maupun untuk

mensimulasikan sistem evolusioner, algoritme genetik sangat membantu. (Forrest, 1996) Algoritma genetika adalah algoritma komputasi yang terinspirasi oleh proses evolusi untuk menemukan solusi optimal secara alami. Karena dapat menangani berbagai persoalan dengan ruang pencarian yang kompleks dan menggunakan model matematis dengan nilai akurasi yang tinggi, maka pendekatan ini banyak digunakan. (Dharma et al., 2020)

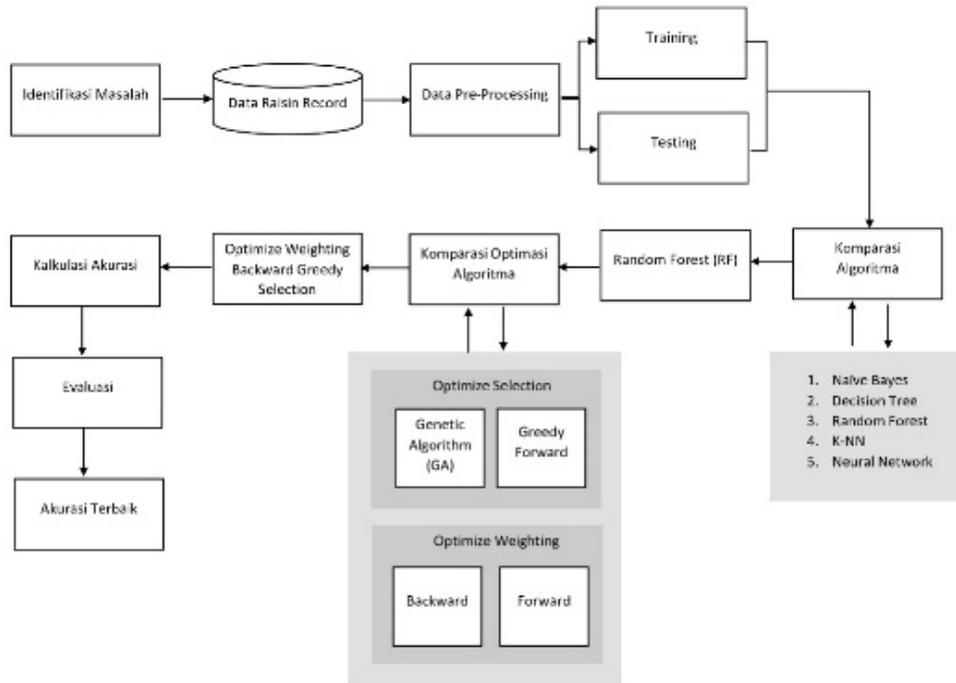
Penelitian ini menggunakan fitur optimasi selection dengan algoritma GA pada Random Forest untuk membuat model prediksi kismis kecimen dan besni.

METODE PENELITIAN

Tahapan Penelitian

Dalam penelitian ini, beberapa teknik digunakan untuk mengatasi masalah klasifikasi, termasuk pelatihan pemisahan data dan pengujian data menggunakan dua metodologi (Cross Validation dan Split validation) Untuk memilih algoritma klasifikasi yang optimal, dilakukan juga perbandingan algoritma klasifikasi. Klasifikasi selanjutnya akan ditingkatkan dengan meningkatkan fitur dan bobot dataset. Tahapan penelitian ini diakhiri dengan evaluasi untuk memilih algoritma klasifikasi dan strategi optimasi yang dapat meningkatkan nilai klasifikasi. Langkah-langkah penelitian ditunjukkan pada Gambar 1. (Ramdhani et al., 2023)

Gambar 1. Desain Penelitian



a. Identifikasi Masalah

Kismis tersedia dalam berbagai macam. Model algoritma yang memadai diperlukan untuk mengkategorikan jenis kismis ini, yang akan membantu para peneliti dalam klasifikasi variasi kismis. Pada penelitian ini, varietas kismis yang ditanam di Turki dikategorikan sebagai varietas kismis Kecimen dan Besni. (Ramdhani et al., 2023)

a. Dataset

Data set yang digunakan dalam penelitian ini adalah Raisin Dataset, yang memiliki 900 data record dan 8 atribut, dan diterbitkan pada tahun 2020 oleh UCI Repository (<http://archive.ics.uci.edu/ml>). Dua kelas kismis Turki dalam dataset adalah kismis Kecimen dan Besni. Ekstensi file dataset yang digunakan adalah *.csv.

b. Preproscesing (Split & Cross)

Pada tahap data preprocessing, data hasil penelitian akan dipecah menjadi data training dan data testing. Memanfaatkan cross validation dan split validation, data sharing akan dilakukan. Split validation digunakan untuk menguji model tertentu, sedangkan data sharing dengan cross validation digunakan untuk mengidentifikasi model pengujian mana yang akan berkinerja terbaik secara keseluruhan.

c. Komparasi Algoritma

Perbandingan algoritma dilakukan pada penelitian, algoritma dibandingkan untuk menentukan metode mana yang dianggap terbaik. Berbagai algoritma diuji selama tahap perbandingan algoritma. Lima algoritma yang digunakan dalam penelitian ini adalah Naive Bayes, K-Nearest Neighbor (K-NN), Decision Tree, Neural Network, dan Random Forest. Melalui perbandingan nilai akurasi yang dihasilkan oleh masing-masing dari lima algoritma ini, akan dimungkinkan untuk menentukan algoritma mana yang terbaik dalam mengidentifikasi raisin dataset.

d. Algoritma Random Forest

Pada tahap klasifikasi Varietas Kismis ini, algoritma RF muncul sebagai model terbaik. Model RF dipilih berdasarkan 6 algoritma yang mengklasifikasikan biji kismis dengan akurasi dan AUC tertinggi. Cross validation dan split validation akan digunakan untuk mengevaluasi algoritma untuk model. Untuk mendapatkan hasil rata-rata, model akan dievaluasi menggunakan parameter rasio split 0,5 hingga 0,9 dan parameter rasio cross 10 fold.

e. Komparasi Optimasi

Pada tahap perbandingan algoritma optimasi, dilakukan pengujian terhadap fitur optimasi. Penelitian ini menggunakan 2 fitur optimasi yaitu Optimize Selection dengan

algoritma GA, Greedy Forward dan Optimize Weight dengan algoritma Backward dan Forward.

f. Genetic Algorithm

Istilah "Algoritma Genetika" (GA) mengacu pada metode pencarian optimasi dan heuristik yang terinspirasi dari proses seleksi alam.(McCall, 2005) Dengan menggunakan pendekatan komputasi rendah, dihasilkan pembangkitan acak, populasi, atau sejumlah besar kromosom. Setiap kali iterasi GA dijalankan, nilai fitness kromosom populasi dinilai, dan kromosom dengan fitness tertinggi dicatat. Kromosom populasi optimal ditentukan menjadi solusi terpilih setelah

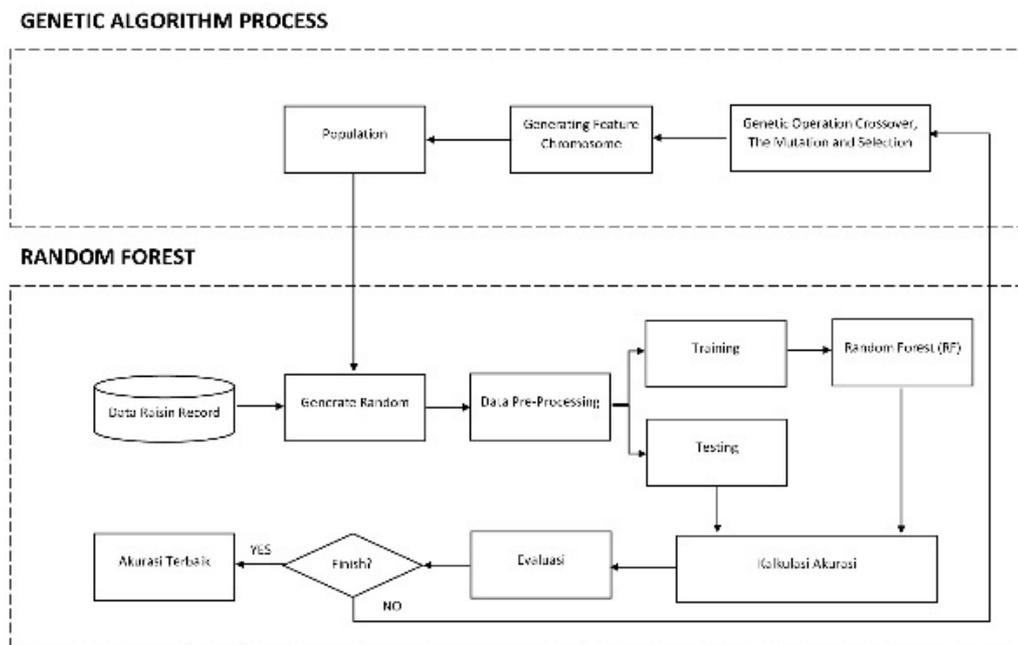
sejumlah iterasi yang signifikan.(Qiu et al., 2015)

g. Evaluasi (AUC&Akurasi)

Pada tahap evaluasi ini, akan diketahui nilai akurasi dan AUC terbaik dari klasifikasi Raisin Dataset. Untuk mengetahui apakah ada perbedaan antara sebelum dan sesudah optimasi, peneliti membandingkan hasil akurasi dan AUC yang diperoleh dari algoritma Random Forest split validation dengan rasio split 0,5 hingga 0,9 dan cross validation, dengan yang diperoleh dari algoritma Random Forest berdasarkan Optimize Selection Genetic Algorithm.

Metode yang di usulkan

Gambar 2. Desain Metode yang diusulkan



Pada penelitian ini dikemukakan teknik klasifikasi Raisin, GA sebagai seleksi fitur dan algoritma Random Forest (RF) sebagai klasifikasi Raisin Dataset. Akuisisi dataset Raisin adalah langkah pertama dalam proyek ini. Dataset kemudian akan diperiksa, dan data akan dinormalisasi. Langkah selanjutnya adalah membagi data menjadi data training dan data testing (test data). Algoritma Random Forest (RF) digunakan sebagai data pelatihan untuk membuat model, sedangkan dataset pengujian digunakan untuk membuat angka akurasi. Algoritma kemudian

harus dibandingkan. Untuk mendapatkan model algoritma yang optimal, algoritma dibandingkan satu sama lain saat mengklasifikasikan.

Dalam penelitian ini, GA digunakan untuk memilih fitur. Untuk meningkatkan performa nilai akurasi klasifikasi dari varietas Raisin, Genetic Algorithm membuat populasi yang terdiri dari beberapa individu terpilih dengan nilai yang paling relevan dengan klasifikasi.(Ramadhani et al., 2020)

Peneliti menjelaskan struktur teknik yang diusulkan untuk mengklasifikasikan

Varietas Kismis. Hasil evaluasi model klasifikasi varietas Kismis yang diusulkan dengan menggunakan optimasi fitur GA memberikan pengaruh paling besar terhadap kemampuan algoritma Random Forest (RF) dalam mengklasifikasikan varietas Kismis ke dalam kelas Kecimen dan kelas Besni dengan hasil klasifikasi terbesar.

HASIL DAN PEMBAHASAN

Pada tahap ini ditampilkan hasil percobaan uji klasifikasi dataset kismis. Langkah pertama adalah mengidentifikasi masalah. Diketahui bahwa dalam mengklasifikasikan kismis dibutuhkan suatu metode atau algoritma dengan model terbaik. Berdasarkan hal tersebut maka dilakukan penelitian tentang klasifikasi varietas Kismis.

Dataset penelitian Raisin digunakan dalam penelitian ini, dan diunduh dari situs web UCI Machine Learning Repository. Dataset penelitian terdiri dari 900 record data untuk jenis kismis yang berbeda, masing-masing dengan 8 atribut dan 1 label yang terdiri dari dua kelas yaitu kelas Kecimen dan kelas Besni. Ditunjukkan pada Tabel 1.

Tabel 1. Atribut

No	Data Input	Detail
1	Area	Memberikan jumlah piksel dalam batas kismis.
2	Perimeter	Ini mengukur lingkungan dengan menghitung jarak antara batas kismis dan piksel di sekitarnya.
3	MajorAxisLength	Memberikan panjang sumbu utama, yang

		merupakan garis terpanjang yang dapat ditarik pada kismis.
4	MinorAxisLength	Memberikan panjang sumbu kecil, yang merupakan garis terpendek yang dapat ditarik pada kismis.
5	Eccentricity	Ini memberikan ukuran eksentrisitas elips, yang memiliki momen yang sama dengan kismis.
6	ConvexArea	Memberikan jumlah piksel dari cangkang cembung terkecil dari wilayah yang dibentuk oleh kismis.
7	Extent	Memberikan rasio wilayah yang dibentuk oleh kismis terhadap total piksel dalam kotak pembatas.
8	Class	Memberikan rasio wilayah yang dibentuk oleh kismis terhadap total piksel dalam kotak pembatas.

Peneliti melakukan preprocessing data setelah mengumpulkannya. Untuk menentukan apakah ada data yang salah pada saat ini, nilai yang hilang diperiksa pada data. Selanjutnya, hilangkan duplikat untuk memastikan tidak ada dua bagian data yang identik, dan normalkan menggunakan teknik Z-Transformation. Kisaran nilai untuk data yang dinormalisasi akan tetap antara 0 dan 1. Hasil dari proses normalisasi yang digambarkan pada Tabel 2 adalah sebagai berikut.

Tabel 2. Normalisasi Data

No	Area	Perimeter	MajorAxisLength	MinorAxisLength	Eccentricity	ConvexArea	Extent	Class
1	87524.0	442.2460114	253.291155	0.819738392	90546.0	0.758650579	1184.04	Kecimen
2	75166.0	406.690687	243.0324363	0.801805234	78789.0	0.68412957	1121.786	Kecimen
...
...
900	85609.0	512.0817743	215.2719758	0.907345395	89197.0	0.632019963	1272.862	Besni

Setelah preprocessing data selesai, selanjutnya membandingkan metode. Keenam algoritma yang digunakan dalam penelitian ini dibandingkan menggunakan

algoritma Naive Bayes, K-Nearest Neighbor (K-NN), Decision Tree, Random Tree, dan Random Forest. Proses validasi data dilakukan dengan menggunakan metode

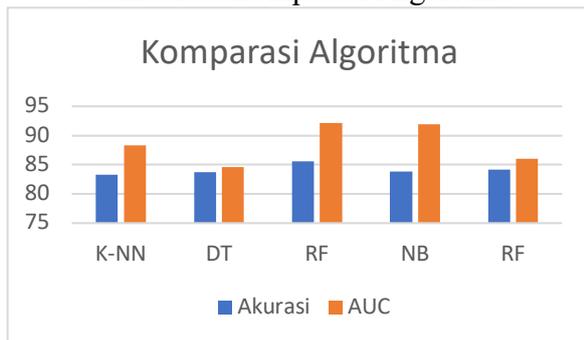
validasi 10-Fold yang menghasilkan nilai akurasi, presisi, recall, dan AUC untuk menilai kinerja dari keenam algoritma tersebut. Nilai akurasi dan AUC yang dihasilkan oleh masing-masing metode tercantum di bawah ini dan dapat ditemukan pada Tabel 3.

Tabel 3. Hasil perbandingan algoritma

Algoritma	Validasi	Akurasi	AUC
k-NN	Cross	83.33%	0.883
Decision Tree	Cross	83.78%	0.846
Random Forest	Cross	85.56%	0.922
Naïve Bayes	Cross	83.89%	0.919
Random Tree	Cross	84.22%	0.860

Sebuah grafik harus dibuat untuk lebih jelas menggambarkan varians dalam akurasi Kinerja AUC algoritma perbandingan. Prosedur perbandingan algoritma yang digambarkan pada Gambar 4 memiliki 6 algoritma yang direpresentasikan pada grafik di bawah ini.

Gambar 3. Komparasi Algoritma



Tabel 4. Confusion Matrix RF

	True kecimen	True besni	Class precision
Pred. kecimen	395	79	83.33%
Pred. besni	55	371	87.09%
Class recall	87.78%	82.44%	

Berdasarkan perbandingan algoritma tersebut, dapat disimpulkan bahwa algoritma RF memiliki performa terbaik secara keseluruhan, dengan akurasi dan nilai AUC masing-masing sebesar 85.56% dan 0.922. Tabel 4 menampilkan klasifikasi yang dihasilkan oleh metode RF menggunakan Confusion Matrix. Nilai pengujian menunjukkan bahwa model algoritma RF berkinerja sangat baik dalam hal klasifikasi. Rapidminer juga menghasilkan kurva

Receiver Operating Characteristic (ROC). Validasi terpisah digunakan untuk memvalidasi data setelah ditentukan bahwa algoritma RF mengklasifikasikan Varietas Raisin dengan akurasi terbaik. Hasil validasi data dengan validasi split menggunakan rasio split 0,5 sampai 0,9 ditunjukkan pada Tabel 5.

Tabel 5. Split ratio 0.5 - 0.9 RF

Algoritma	Validasi	Ratio	Akurasi	AUC
RF	Split	0.5	86.89%	0.932
RF	Split	0.6	86.11%	0.930
RF	Split	0.7	85.19%	0.910
RF	Split	0.8	84.44%	0.931
RF	Split	0.9	80.00%	0.880
Rata-Rata			84.52%	0.916

Tabel 6. Confusion Matrix split 0.5 RF

	True kecimen	True besni	Class precision
Pred. kecimen	206	40	83.74%
Pred. besni	19	185	90.69
Class recall	91.56%	82.22%	

Tabel 5 menunjukkan bahwa algoritma RF dengan rasio split 0,5 hingga 0,9 memiliki nilai akurasi rata-rata sebesar 84.52% dan nilai AUC sebesar 0,916. Validasi paling akurat memiliki AUC sebesar 0,932 dan akurasi sebesar 86.89% bila menggunakan rasio split 0,5. Tabel 6 menampilkan Confusion Matrix yang dihasilkan pada tabel berikut. Menurut hasil pengujian, nilai AUC model algoritma RF dengan rasio split 0,5 adalah 0,932. Klasifikasi yang sangat baik dicapai dengan menggunakan model algoritma RF dengan rasio pemisahan 0,5, sesuai dengan nilai pengujian. Rapidminer juga menghasilkan kurva Receiver Operating Characteristic (ROC). Fitur Optimization digunakan untuk menaikkan nilai akurasi algoritma RF. Fitur pengoptimalan, Optimize Selection and Optimize Weight, dikontraskan dalam penelitian ini. Metode validasi Cross 10 fold digunakan untuk melakukan validasi. Nilai akurasi dan AUC dari setiap fitur optimasi yang terdapat pada Tabel 7 tercantum di bawah ini.

Tabel 7. Fitur optimaze selection dan optimaze weight

Algoritma	Fitur	Metode	Validasi	Akurasi	AUC
-----------	-------	--------	----------	---------	-----

RF	Selecti on	GA	Cross	87.56 %	0.928
RF	Selecti on	Greedy Forward	Cross	87.00 %	0.933
RF	Weighting	Forward	Cross	87.22 %	0.928
RF	Weighting	Backward	Cross	86.33 %	0.924

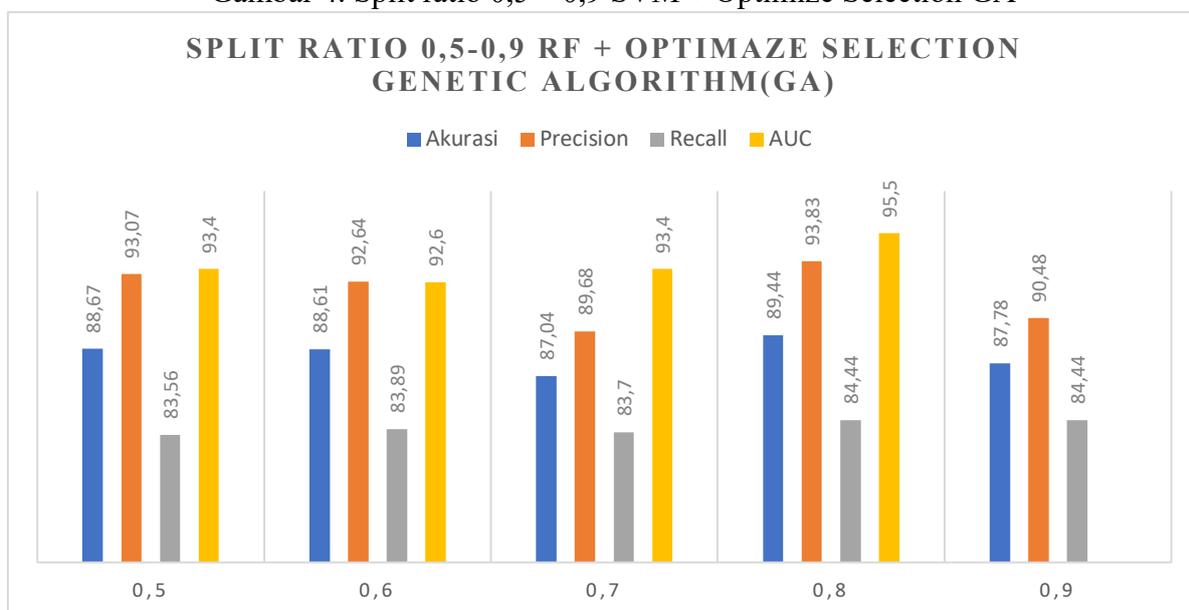
Tabel 7 menunjukkan bahwa teknik Optimize Selection dan Optimize Weight berhasil meningkatkan akurasi algoritma RF dan nilai AUC pada klasifikasi Varietas Kismis. Dibandingkan dengan fitur teknik

pengoptimalan lainnya, fitur optimize selection dengan GA memiliki akurasi dan nilai AUC tertinggi. Selain itu nilai akurasi dan AUC yang dibuat memiliki nilai yaitu 87,56% untuk akurasi dan 0,928 untuk AUC. Peneliti memutuskan untuk menguji algoritma RF berdasarkan Optimize Selection GA dalam mengidentifikasi Varietas Raisin berdasarkan hal tersebut. Tabel 8 menampilkan nilai akurasi, presisi, recall, dan AUC secara keseluruhan untuk setiap rasio parameter split 0,5 hingga 0,9 menggunakan algoritma berbasis RF pada GA optimize selection.

Tabel 8. Hasil Akurasi, Precision, Recall, AUC SVM + Optimize selection GA

Algoritma	Metode	Ratio	Akurasi	Precision	Recall	AUC
RF	GA	0.5	88.67%	93.07%	83.56%	0.934
RF	GA	0.6	88.61%	92.64%	83.89%	0.926
RF	GA	0.7	87.04%	89.68%	83.70%	0.934
RF	GA	0.8	89.44%	93.83%	84.44%	0.955
RF	GA	0.9	87.78%	90.48%	84.44%	0.930
Rata-rata			88.30%	91.94%	84.00%	0.935

Gambar 4. Split ratio 0,5 – 0,9 SVM + Optimize Selection GA



Grafik pengujian untuk metode RF, yang didasarkan pada Optimize Selection GA dan memiliki rasio pemisahan 0,5 hingga 0,9, ditunjukkan di bawah ini (Gambar 5). Nilai rata-rata yang diketahui untuk akurasi, presisi, daya ingat, dan AUC untuk algoritme RF berdasarkan Optimize Selection GA dengan rasio parameter 0,5 hingga 0,9 masing-masing adalah 88,30%,

84,00%, dan 0,935. Jika dibandingkan dengan rasio split lainnya, validasi dengan rasio split 0,8 menghasilkan hasil akhir yang tinggi. Perbandingan akurasi algoritma RF dan RF menggunakan Optimize Selection GA dengan cross 10 fold ditunjukkan pada Tabel 9.

Tabel 9. Akurasi RF dan RF + GA

Validasi	Akurasi Algoritma RF	Akurasi Algoritma RF + GA
Cross	85.56%	87.56%

Tabel 10. AUC RF dan RF + GA

Validasi	AUC algoritma RF	AUC algoritma RF + GA
Cross	0.922	0.928

Algoritma RF dan RF berdasarkan Optimize Selection GA dengan cross 10 fold Perbandingan pada Tabel 10 dalam hal AUC.

KESIMPULAN

Dataset Raisin (Raisin Varieties), yang disediakan dari UCI Machine Learning Repository, digunakan dalam penelitian ini untuk memodelkan algoritma RF dan RF berdasarkan Optimize Selection GA. Di antara lima algoritme; Naive Bayes, K-Nearest Neighbor (K-NN), Decision Tree, Random Tree, dan Random Forest, dengan validasi data 10-Fold-Validation pada Algoritma RF tampil di akurasi tertinggi. Nilai akurasi dan AUC yang dihasilkan oleh metode RF adalah 85.56% untuk akurasi dan 0.922 untuk AUC. untuk mengevaluasi operator validasi split dengan rasio split mulai dari 0,5 hingga 0,9 menggunakan metode RF. Berdasarkan pengujian yang dijalankan, rata-rata akurasi dan nilai AUC masing-masing adalah 84.52% dan 0.916. Tingkat akurasi dan AUC terbesar untuk validasi data dengan split ratio 0,5 adalah 86.89 untuk akurasi dan 0.932 untuk AUC. Optimalisasi fitur seleksi dengan metode GA memanfaatkan validasi data dengan 10-Fold-Validation dilakukan untuk meningkatkan performa nilai akurasi algoritma RF.

Hal ini menghasilkan nilai akurasi dan AUC sebesar 87.56% untuk akurasi dan 0.928 untuk AUC. Operator validasi split dengan rasio split 0,5 hingga 0,9 digunakan untuk menguji algoritma RF. berdasarkan hasil pengujian. Rata-rata nilai akurasi, presisi, recall, dan AUC masing-masing adalah 88.30%, 91.94%, 84.00%, dan 0.935. Keakuratan algoritme dan AUC mencapai maksimumnya saat diuji dengan rasio pemisahan 0,8; masing-masing, mereka adalah 89.44% untuk akurasi dan 0,955 untuk AUC.

REFERENSI

- Almomani, O. (2020). A feature selection model for network intrusion detection system based on pso, gwo, ffa and ga algorithms. *Symmetry*, 12(6), 1–20. <https://doi.org/10.3390/sym12061046>
- Dharma, F., Noviana, A., Tahir, M., Hendrastuty, N., & Author, C. (2020). Prediction of Indonesian Inflation Rate Using Regression Model Based on Genetic Algorithms. *Jurnal Online Informatika*, 5(1), 45–52. <https://doi.org/10.15575/join>
- Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi. *Teknika*, 8(1), 11–16. <https://doi.org/10.34148/teknika.v8i1.95>
- Forrest, S. (1996). Genetic Algorithms. In *ACM Computing Surveys* (Vol. 28, Issue 1).
- Karimi. (2011). Modelling of raisin berries by some physical and statistical characteristics. *ResearchGate*, 141–147.
- Kemal Üniversitesi Ziraat Fakültesi Dergisi, M., Semerci, A., Kiziltuğ, T., Duran ÇELİK, A., & Ali KİRACI, M. (n.d.). *Türkiye Bağcılığının Genel Durumu*.
- Kumar, V. (2014). Feature Selection: A literature Review. *The Smart Computing Review*, 4(3). <https://doi.org/10.6029/smarter.2014.03.007>
- Masesh, B. (2019). *Machine Learning Algorithms-A Review*. <https://doi.org/10.21275/ART20203995>
- McCall, J. (2005). Genetic algorithms for modelling and optimisation. *Journal of Computational and Applied Mathematics*, 184(1), 205–222. <https://doi.org/10.1016/j.cam.2004.07.034>
- Meira, W. (2020). Data Mining and Machine Learning: Fundamental Concepts and Algorithms. *Researchgate*.
- Osman, A. S. (2019). *Data Mining Techniques: Review*. <https://www.educba.com/7-data->
- Qiu, M., Ming, Z., Li, J., Gai, K., & Zong, Z. (2015). Phase-change memory optimization for green cloud with genetic algorithm. *IEEE Transactions on Computers*, 64(12), 3528–3540.
- Ramadhani, Y., Mubarak, A., Hidayatullah, S., & Wiguna, W. (2020). *Attribute Optimization: Genetic Algorithms and Neural Network for Voice Analysis Classification of Parkinson's*

- Disease*. 3074–3079.
<https://doi.org/10.5220/0009947030743079>
- Ramdhani, Y., Apra, D. F., & Alamsyah, D. P. (2023). Feature selection optimization based on genetic algorithm for support vector classification varieties of raisin. *Indonesian Journal of Electrical Engineering and Computer Science*, 30(1), 192. <https://doi.org/10.11591/ijeecs.v30.i1.pp192-199>
- Rosyada, F. A. N. H. K. M. (2019). BOKIS (BAHAN_OLAHAN_KISMIS)-1[1]. *OSFPREPRINTS*.
- Somantri, O., & Khambali, M. (2017). Feature Selection Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes dan Algoritme Genetika. In *JNTETI* (Vol. 6, Issue 3). <http://cerpenmu.com/>.
- Speiser, J. L. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *ELSEVIER*, 134, 93–101.