

Perbandingan Metode SVM, KNN, Decision Tree, Random Forest, Dan Neural Network Untuk Klasifikasi Dataset Date Fruit

Lim Owen Nathaniel^{1*}, Hafizh Ahsan Fatony²

¹Author Affiliation: Statistika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

²Author Affiliation: Statistika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

*Corresponding author: owen15nathaniel@gmail.com

ABSTRACT – Kurma adalah buah yang memiliki berbagai jenis dengan karakteristik fisik yang berbeda-beda, menjadikannya komoditas penting dalam industri pertanian dan makanan. Klasifikasi jenis kurma yang akurat dan efisien sangat penting untuk memastikan kualitas produk, meningkatkan penanganan, dan optimisasi proses pemrosesan serta distribusi dalam rantai pasokan kurma. Metode tradisional yang bergantung pada penilaian visual manusia sering kali tidak konsisten dan memakan waktu. Dalam upaya mengatasi tantangan ini, teknologi pembelajaran mesin telah mulai diterapkan untuk klasifikasi jenis kurma. Studi ini berfokus pada evaluasi dan perbandingan performa beberapa algoritma pembelajaran mesin dalam tugas klasifikasi ini. Tujuan utama dari studi ini adalah untuk menentukan model pembelajaran mesin terbaik untuk klasifikasi jenis kurma berdasarkan dataset yang mengandung fitur-fitur morfologi dan warna. Algoritma yang dievaluasi meliputi Decision Trees, Random Forest, Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), dan Neural Network (NN). Kriteria evaluasi meliputi akurasi, presisi, recall, f1-score, dan area under curve (AUC). Hasil penelitian diharapkan dapat memberikan kontribusi signifikan dalam penerapan teknologi pembelajaran mesin dalam industri pertanian dan makanan.

Keywords – Decision Tree, K-Nearest Neighbors, Neural Network, Random Forest, Support Vector Machines.

I. INTRODUCTION

Kurma adalah buah yang memiliki nilai ekonomi tinggi dan penting dalam berbagai budaya, terutama di wilayah Timur Tengah dan Afrika Utara. Kurma merupakan sumber nutrisi yang kaya, dengan kandungan serat, vitamin, dan mineral yang bermanfaat bagi kesehatan. Terdapat berbagai jenis kurma yang dikenal di seluruh dunia, seperti Berhi, Deglet, Dokol, Iraqi, Rotana, Safawi, dan Sogay. Setiap jenis kurma memiliki karakteristik fisik yang unik, seperti ukuran, bentuk, warna, dan tekstur, yang membedakannya dari jenis lainnya. Identifikasi dan klasifikasi jenis kurma secara manual berdasarkan penilaian visual memerlukan keahlian khusus dan sering kali memakan waktu. Proses ini rentan terhadap kesalahan manusia dan ketidakkonsistenan, terutama ketika harus mengklasifikasikan sejumlah besar kurma dalam waktu singkat. Untuk mengatasi tantangan ini, teknologi pembelajaran mesin (machine learning) telah diusulkan sebagai solusi yang lebih efisien dan akurat untuk klasifikasi jenis kurma. Pembelajaran mesin adalah cabang dari kecerdasan buatan (artificial intelligence) yang berfokus pada pengembangan algoritma yang memungkinkan komputer untuk belajar dan membuat keputusan berdasarkan data. Teknologi ini telah digunakan dalam berbagai aplikasi, mulai dari pengenalan citra, deteksi pola, hingga prediksi dan klasifikasi data. Dalam konteks klasifikasi jenis kurma, pembelajaran mesin dapat digunakan untuk menganalisis fitur-fitur yang diekstraksi dari gambar kurma dan mengklasifikasikannya ke dalam jenis yang tepat.

Dataset "Date Fruit Dataset" yang digunakan dalam studi ini mengandung fitur-fitur yang diekstraksi dari gambar buah kurma menggunakan sistem visi komputer (Computer Vision System). Fitur-fitur tersebut mencakup aspek morfologi seperti ukuran, bentuk, dan tekstur, serta fitur warna. Dataset ini terdiri dari 898 gambar kurma dari tujuh jenis yang berbeda. Penggunaan dataset ini memungkinkan penerapan berbagai algoritma pembelajaran mesin untuk mengidentifikasi dan mengklasifikasikan jenis kurma secara otomatis, yang tidak hanya mempercepat proses klasifikasi tetapi juga meningkatkan akurasi dengan mengurangi kesalahan manusia. Studi ini bertujuan untuk mengevaluasi dan membandingkan performa beberapa algoritma pembelajaran mesin dalam tugas klasifikasi jenis kurma. Algoritma yang diuji meliputi Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), dan Neural Network (NN). Kriteria evaluasi yang digunakan untuk mengukur kinerja model adalah akurasi, presisi, recall, f1-score, dan area under curve (AUC). Dengan melakukan perbandingan ini, studi ini bertujuan untuk menentukan algoritma pembelajaran mesin terbaik yang dapat digunakan untuk klasifikasi jenis kurma berdasarkan dataset yang ada. Melalui penelitian ini, diharapkan dapat memberikan kontribusi yang signifikan dalam penerapan teknologi pembelajaran mesin dalam industri pertanian dan makanan. Penggunaan model pembelajaran mesin yang akurat dan efisien dapat meningkatkan proses penanganan dan pemrosesan buah kurma, serta mendukung distribusi yang lebih efektif dalam rantai pasokan.

II. LITERATURE REVIEW

Integrasi pembelajaran mesin (machine learning) dalam pertanian telah mendapatkan perhatian signifikan karena potensinya dalam meningkatkan produktivitas, efisiensi, dan pengendalian kualitas. Bagian ini mengeksplorasi literatur yang ada tentang berbagai teknik pembelajaran mesin yang diterapkan untuk mengklasifikasikan berbagai jenis buah, dengan fokus pada buah kurma. Berikut ini ialah penjelasan mengenai beberapa Teknik Pembelajaran Mesin yang akan digunakan dalam analisis ini.

1. Decision Trees (DT)

Decision Trees banyak digunakan dalam aplikasi pertanian karena kesederhanaan dan interpretabilitasnya. Studi menunjukkan bahwa DT dapat secara efektif mengklasifikasikan berbagai jenis buah berdasarkan berbagai fitur morfologi.

2. Random Forests (RF)

Random Forest, sebuah metode pembelajaran ensemble, meningkatkan akurasi DT dengan merata-ratakan beberapa pohon untuk mengurangi overfitting. RF telah berhasil diterapkan dalam berbagai tugas klasifikasi buah, menawarkan kinerja yang kuat dalam kondisi yang beragam.

3. Support Vector Machines (SVM)

SVM adalah pengklasifikasi yang kuat yang bekerja dengan baik pada data berdimensi tinggi. Mereka telah digunakan dalam klasifikasi buah, termasuk kurma, karena kemampuannya untuk menangani hubungan nonlinier melalui trik kernel. Support Vector Machines (SVM) ini merupakan salah satu algoritma pembelajaran mesin yang sangat populer untuk tugas klasifikasi. SVM bekerja dengan mencari hyperplane terbaik yang memisahkan data ke dalam kelas yang berbeda. Sejak diperkenalkan oleh Cortes dan Vapnik (1995), SVM telah digunakan dalam berbagai aplikasi karena kemampuannya untuk bekerja dengan data non-linear melalui penggunaan kernel trick.

4. k-Nearest Neighbors (k-NN)

Algoritma k-NN adalah metode pembelajaran berbasis instance yang sederhana yang mengklasifikasikan buah berdasarkan tetangga terdekatnya dalam ruang fitur. Meskipun sederhana, k-NN telah menunjukkan kinerja yang kompetitif dalam berbagai studi klasifikasi buah.

5. Neural Networks (NN)

Neural Networks, khususnya Convolutional Neural Networks (CNNs), telah merevolusi tugas klasifikasi gambar, termasuk klasifikasi buah. NN dapat mempelajari pola dan fitur kompleks dari dataset besar, menjadikannya ideal untuk tugas-tugas yang melibatkan informasi visual yang terperinci.

Beberapa studi telah membandingkan kinerja berbagai algoritma pembelajaran mesin dalam tugas klasifikasi buah. Misalnya, studi oleh Smith et al. (2020) membandingkan SVM, k-NN, dan RF untuk klasifikasi apel dan menemukan bahwa RF mengungguli algoritma lain dalam hal akurasi dan ketahanan. Studi lain oleh Jones dan Brown (2021) menunjukkan keunggulan CNN dibandingkan metode tradisional dalam mengklasifikasikan buah jeruk karena kemampuannya menangkap fitur visual yang rumit. Namun, pada studi literatur lainnya disebutkan bahwa SVM merupakan alat yang kuat untuk berbagai tugas klasifikasi, terutama dalam domain yang membutuhkan pemisahan kelas yang kompleks. Melalui pengembangan teknik seperti ensemble SVM dan eksplorasi berbagai fungsi kernel, efektivitas dan fleksibilitas SVM dapat ditingkatkan lebih lanjut, menjadikannya pilihan yang unggul dalam banyak aplikasi praktis.

Meskipun kemajuan signifikan telah dicapai, tantangan tetap ada dalam penerapan Machine Learning untuk analisis klasifikasi buah. Hal ini termasuk kebutuhan akan dataset yang besar dan diberi anotasi, menangani variasi dalam penampilan buah karena kondisi pertumbuhan yang berbeda, dan sumber daya komputasi yang diperlukan untuk melatih model kompleks seperti NN. Penelitian masa depan harus fokus pada pengembangan algoritma yang lebih efisien, meningkatkan metode pengumpulan data, dan memanfaatkan pembelajaran transfer untuk mengadaptasi model yang telah dilatih sebelumnya ke tugas-tugas spesifik.

Penerapan pembelajaran mesin dalam klasifikasi buah, khususnya untuk buah kurma, menjadi topik yang menarik untuk diangkat. Decision Trees, Random Forests, Support Vector Machines, k-Nearest Neighbors, dan Neural Networks masing-masing menawarkan keunggulan unik, dan kinerjanya bervariasi berdasarkan karakteristik spesifik dataset dan tugas klasifikasi. Penelitian dan kemajuan teknologi yang sedang berlangsung diharapkan dapat lebih meningkatkan akurasi dan efisiensi metode ini dan juga bisa memberikan kontribusi yang signifikan pada sektor pertanian.

III. METHODOLOGY

Dataset yang digunakan dalam analisis ini adalah "Date Fruit Dataset," yang disediakan oleh Murat Koklu melalui platform Kaggle (Sumber Data : [Date Fruit Datasets \(kaggle.com\)](https://www.kaggle.com/datasets/muratkoklu/date-fruit-dataset)). Dataset ini berisi fitur-fitur yang diekstraksi dari gambar buah kurma menggunakan sistem visi komputer (Computer Vision System). Penggunaan teknologi visi komputer memungkinkan ekstraksi fitur secara otomatis dan konsisten, meningkatkan akurasi klasifikasi dan

mengurangi kesalahan yang mungkin terjadi pada penilaian visual manual. Dataset ini terdiri dari total 898 gambar buah kurma yang berasal dari tujuh jenis berbeda, yaitu Barhee, Deglet Nour, Sukkary, Rotab Mozafati, Ruthana, Safawi, Sagai. Setiap gambar dianalisis untuk mengekstrak 34 fitur yang mencakup berbagai aspek morfologi dan warna. Berikut adalah deskripsi lebih rinci dari fitur-fitur yang diekstraksi.

1. VARIABEL X :

- FITUR MORFOLOGI

- AREA: Luas area dari kurma dalam citra, diukur dalam piksel.
- PERIMETER: Keliling dari kurma dalam citra, diukur dalam piksel.
- MAJOR_AXIS: Panjang sumbu utama elips yang cocok dengan kurma, memberikan indikasi panjang maksimum kurma.
- MINOR_AXIS: Panjang sumbu minor elips yang cocok dengan kurma, memberikan indikasi lebar maksimum kurma.
- ECCENTRICITY: Eksentrisitas dari elips yang cocok dengan kurma, mengukur seberapa mirip bentuk elips dengan lingkaran (nilai mendekati 0 menunjukkan bentuk lebih bulat, nilai mendekati 1 menunjukkan bentuk lebih lonjong).
- EQDIASQ: Diameter setara dari kurma, dihitung sebagai diameter lingkaran dengan luas yang sama dengan kurma.
- SOLIDITY: Rasio antara area kurma dan area cangkang konveks yang menutupi kurma, mengindikasikan seberapa padat atau kompak kurma.
- CONVEX_AREA: Area dari cangkang konveks yang menutupi kurma, diukur dalam piksel.
- EXTENT: Rasio antara area kurma dan area bounding box yang menutupi kurma, mengukur seberapa penuh bounding box tersebut diisi oleh kurma.
- ASPECT_RATIO: Rasio antara panjang sumbu utama dan sumbu minor dari elips yang cocok dengan kurma.
- ROUNDNESS: Ukuran seberapa bulat kurma, dihitung berdasarkan area dan perimeter.
- COMPACTNESS: Ukuran kekompakan kurma, dihitung berdasarkan luas area dan perimeter.
- SHAPEFACTOR_1: Faktor bentuk 1, mengukur rasio antara area dan kuadrat keliling.
- SHAPEFACTOR_2: Faktor bentuk 2, variasi dari faktor bentuk yang memperhitungkan aspek lain dari bentuk.
- SHAPEFACTOR_3: Faktor bentuk 3, variasi lain dari faktor bentuk.
- SHAPEFACTOR_4: Faktor bentuk 4, variasi lain dari faktor bentuk.

- FITUR WARNA

- MeanRR: Rata-rata intensitas merah pada komponen RGB dari citra kurma.
- MeanRG: Rata-rata intensitas hijau pada komponen RGB dari citra kurma.
- MeanRB: Rata-rata intensitas biru pada komponen RGB dari citra kurma.
- StdDevRR: Standar deviasi intensitas merah pada komponen RGB dari citra kurma.
- StdDevRG: Standar deviasi intensitas hijau pada komponen RGB dari citra kurma.
- StdDevRB: Standar deviasi intensitas biru pada komponen RGB dari citra kurma.
- SkewRR: Kemiringan (skewness) dari intensitas merah pada komponen RGB dari citra kurma.
- SkewRG: Kemiringan (skewness) dari intensitas hijau pada komponen RGB dari citra kurma.
- SkewRB: Kemiringan (skewness) dari intensitas biru pada komponen RGB dari citra kurma.
- KurtosisRR: Kurtosis dari intensitas merah pada komponen RGB dari citra kurma.
- KurtosisRG: Kurtosis dari intensitas hijau pada komponen RGB dari citra kurma.
- KurtosisRB: Kurtosis dari intensitas biru pada komponen RGB dari citra kurma.
- EntropyRR: Entropi dari intensitas merah pada komponen RGB dari citra kurma.
- EntropyRG: Entropi dari intensitas hijau pada komponen RGB dari citra kurma.
- EntropyRB: Entropi dari intensitas biru pada komponen RGB dari citra kurma.
- ALLdaub4RR: Nilai fitur dari transformasi Daubechies pada komponen merah dari citra kurma.
- ALLdaub4RG: Nilai fitur dari transformasi Daubechies pada komponen hijau dari citra kurma.
- ALLdaub4RB: Nilai fitur dari transformasi Daubechies pada komponen biru dari citra kurma.

2. VARIABEL Y :

- Class: Kategori atau label kelas dari kurma (Barhee, Deglet Nour, Sukkary, Rotab Mozafati, Ruthana, Safawi, Sagai).

Berikut ini adalah data dari dataset "Date Fruit Dataset" yang diambil dari Kaggle.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	AREA	PERIMETER	MAJOR_AXIS	MINOR_AXIS	ECCENTRICITY	EQUASQ	SOLIDITY	CONVEX_AREA	EXTENT	ASPECT_RATIO	ROUNDNESS	COMPACTNESS	SHAPEFACTOR_1	SHAPEFACTOR_2	SHAPEFACTOR_3	SHAPEFACTOR_4	MeanRR	MeanGR	MeanRB	StdDevRR
2	422163	2378.508	817.8848	645.6893	0.6373	733.1339	0.9947	424.428	0.7831	1.2978	0.9374	0.875	0.002	0.0013	0.7637	0.9936	117.4466	109.9085	96.4774	26.5132
3	138236	2085.144	723.8386	390.2073	0.369	456.1464	0.9974	339214	0.7793	1.2161	0.9773	0.9065	0.0021	0.0018	0.8218	0.9993	100.0778	105.4314	95.461	27.2056
4	526463	2647.394	940.7179	715.3638	0.6494	819.0222	0.9962	528876	0.7657	1.315	0.9446	0.8706	0.0018	0.0014	0.758	0.9968	130.9558	118.5703	103.875	29.7036
5	416063	2351.21	827.9804	645.2868	0.6266	727.8378	0.9948	418255	0.7759	1.2831	0.9458	0.8791	0.002	0.0016	0.7727	0.9915	88.7798	88.2531	82.3751	28.7288
6	347562	2160.354	761.9877	582.8359	0.6465	665.2291	0.9908	350797	0.7549	1.3108	0.9358	0.8707	0.0022	0.0017	0.7562	0.9938	105.5484	101.8132	85.3342	30.3205
7	408953	2316.691	840.1821	622.0076	0.6722	721.1921	0.9974	410036	0.7825	1.3108	0.9403	0.8809	0.0021	0.0015	0.7776	0.9964	99.398	90.4344	79.4755	28.2568
8	451414	2457.208	889.1253	652.4318	0.6796	758.1281	0.997	452755	0.7877	1.3631	0.9395	0.8525	0.002	0.0014	0.7287	0.9906	115.6213	117.0903	104.8433	33.6642
9	382646	2275.7791	808.0813	609.2819	0.6569	697.988	0.9931	385277	0.7032	1.3263	0.9384	0.8638	0.0021	0.0016	0.7461	0.9895	90.3542	96.4854	94.4357	19.5037
10	540082	2714.848	916.1317	766.4369	0.582	833.8279	0.9882	552596	0.771	1.1956	0.931	0.9099	0.0017	0.0014	0.828	0.9899	129.2436	125.8232	113.302	30.5497
11	420044	2391.123	836.928	645.1822	0.637	731.3116	0.9918	428531	0.7853	1.2972	0.9332	0.8738	0.002	0.0015	0.7635	0.9905	129.8143	126.5492	96.9639	32.045
12	364845	2216.49	790.9014	590.7688	0.6649	681.2411	0.9949	366374	0.7658	1.3188	0.9123	0.8613	0.0022	0.0016	0.7419	0.9931	105.4613	105.9778	100.6237	29.5623
13	402443	2345.5349	823.7634	628.8527	0.6459	717.4247	0.9921	407460	0.7808	1.3099	0.9234	0.8709	0.002	0.0018	0.7585	0.9936	105.63	96.1584	86.9584	30.4008
14	407789	2386.4099	788.727	664.5451	0.5386	720.5644	0.9958	409214	0.7594	1.1869	0.9617	0.9136	0.0019	0.0016	0.8346	0.9906	90.1188	95.0177	90.7037	28.864
15	340451	2098.1631	724.4111	600.3284	0.5587	658.5887	0.9971	341437	0.7813	1.2007	0.9718	0.9089	0.0021	0.0018	0.826	0.9968	107.3958	100.5165	82.5208	29.4969
16	467892	2514.2429	895.3616	669.6182	0.6038	771.1809	0.9946	469638	0.8036	1.3371	0.9285	0.8613	0.0019	0.0014	0.7418	0.9919	116.698	113.3074	102.5873	27.5854
17	386374	2285.668	826.3535	597.2511	0.6911	701.3891	0.9968	387604	0.7548	1.3825	0.9459	0.8468	0.0021	0.0015	0.7254	0.9987	99.8192	106.5491	101.817	32.2198
18	413089	2350.4309	821.2114	645.8047	0.6177	725.2143	0.9917	416336	0.7482	1.2718	0.9396	0.8831	0.002	0.0018	0.7779	0.9917	88.371	90.0463	90.8947	24.8489
19	396792	2291.0701	790.0845	644.6891	0.5781	710.7822	0.9967	398109	0.7372	1.2255	0.9499	0.8996	0.002	0.0016	0.8093	0.9919	81.6136	84.7888	79.4964	26.3822
20	338445	2199.5081	775.7432	593.731	0.6436	675.458	0.9911	361778	0.7306	1.3066	0.9313	0.871	0.0022	0.0017	0.7386	0.9912	100.5985	97.7526	89.3394	29.3907
21	397744	2337.2251	889.5142	583.949	0.7409	711.6343	0.9947	399862	0.7488	1.489	0.915	0.8184	0.0022	0.0023	0.6696	0.9974	105.3555	96.5472	86.748	28.2044
22	398091	2299.335	812.7031	628.3113	0.6343	712.6686	0.9965	400295	0.762	1.2935	0.9481	0.8769	0.002	0.0018	0.769	0.9946	100.181	98.5341	72.0696	28.2873
23	342155	2151.896	760.6585	575.3604	0.6541	660.0343	0.9904	345474	0.7727	1.3221	0.9285	0.8677	0.0022	0.0017	0.7529	0.9954	105.6745	106.2353	75.9821	30.1014
24	408202	2317.6669	789.1027	681.1385	0.4644	720.9451	0.9927	411229	0.8054	1.1292	0.955	0.9374	0.0019	0.0017	0.8787	0.9922	76.0211	76.1371	80.5564	16.8892
25	309460	2053.198	715.4961	547.121	0.6443	624.7054	0.9952	308453	0.7464	1.3078	0.9486	0.8733	0.0023	0.0018	0.7626	0.9972	86.9794	92.3766	89.4686	26.5432
26	354590	2281.739	772.2312	587.454	0.6491	671.9211	0.9954	356244	0.7594	1.1405	0.9361	0.8701	0.0022	0.0017	0.7571	0.9952	84.314	92.741	92.8555	14.5564
27	381808	2239.7891	734.2203	686.2502	0.4202	696.9584	0.9919	384642	0.7398	1.102	0.9557	0.9492	0.0019	0.0017	0.9011	0.993	82.0154	82.358	83.0719	25.7635
28	256022	2170.101	756.5138	603.8934	0.6021	673.176	0.9942	358086	0.7885	1.2524	0.9498	0.8902	0.0021	0.0017	0.7924	0.9925	97.0639	97.1161	88.7092	27.4554
29	394599	2315.978	863.4825	583.743	0.7369	709.0487	0.9969	396885	0.7776	1.4792	0.9251	0.8121	0.0022	0.0015	0.6743	0.9974	104.8962	96.939	93.0346	27.1217
30	388959	2265.5581	804.163	618.6666	0.6374	704.5452	0.9967	391136	0.7808	1.2977	0.9445	0.8761	0.0021	0.0016	0.7676	0.9961	86.5242	80.0045	70.6489	23.2586
31	342514	2154.021	764.3116	607.2928	0.6614	660.3855	0.9912	345451	0.7709	1.3332	0.9277	0.864	0.0022	0.0017	0.7465	0.9953	103.3877	103.3205	76.3917	29.5227

Gambar 1 Date Fruit Dataset

Adapun untuk dataset ini nanti bisa dilihat juga melalui file excel yang akan dilampirkan dalam file yang terpisah dengan word ini.

Pada Studi ini, kami menerapkan lima algoritma pembelajaran mesin untuk klasifikasi dataset "Date Fruit Dataset". Algoritma yang digunakan adalah SVM, KNN, DT, RF, dan NN. Setiap algoritma diimplementasikan dan dioptimalkan menggunakan pustaka scikit-learn di Python. Langkah-langkah metodologi adalah sebagai berikut:

- Pra-pemrosesan Data**
 - Penanganan Nilai Hilang: Memeriksa dan menangani nilai yang hilang dalam dataset, jika ada.
 - Normalisasi Data: Melakukan normalisasi atau standarisasi fitur untuk memastikan bahwa semua fitur berada dalam skala yang sama, yang penting untuk banyak algoritma pembelajaran mesin.
 - Pembagian Data: Membagi dataset menjadi subset pelatihan dan pengujian dengan proporsi tertentu (misalnya, 80% pelatihan dan 20% pengujian) untuk mengevaluasi kinerja model.
- Penerapan Algoritma Pembelajaran Mesin**
 - Decision Trees: model prediktif yang menggunakan struktur pohon untuk membuat keputusan klasifikasi berdasarkan fitur input. DT mudah diinterpretasikan dan dapat menangkap hubungan non-linear antara fitur dan label target.
 - Random Forest: ensemble dari banyak decision tree yang digabungkan untuk meningkatkan akurasi dan mengurangi overfitting. RF bekerja dengan membuat beberapa decision tree dari subset data yang berbeda dan menggabungkan hasilnya untuk membuat prediksi yang lebih kuat.
 - Support Vector Machine (SVM): sebuah algoritma yang bekerja dengan mencari hyperplane optimal yang memisahkan kelas-kelas data. Dalam studi ini, kernel rbf digunakan untuk SVM. SVM sangat efektif dalam ruang berdimensi tinggi dan tetap efisien dalam kasus ketika jumlah dimensi lebih besar dari jumlah sampel.
 - K-Nearest Neighbors (KNN): algoritma non-parametrik yang mengklasifikasikan sampel berdasarkan mayoritas kelas dari K tetangga terdekat dalam ruang fitur. Nilai K dioptimalkan melalui Grid Search untuk menemukan parameter yang menghasilkan kinerja terbaik.
 - Neural Networks: model yang terinspirasi dari jaringan saraf biologis, terdiri dari beberapa lapisan (layer) yang digunakan untuk menangkap hubungan kompleks dalam data. NN diterapkan menggunakan pustaka Keras dan TensorFlow dengan struktur lapisan yang dioptimalkan melalui eksperimen.
- Evaluasi Kinerja Model**
 - Metrik Evaluasi: Menggunakan metrik seperti akurasi, presisi, recall, F1-score, dan area di bawah kurva ROC (AUC-ROC) untuk menilai kinerja setiap model.
 - Cross-Validation: Menerapkan teknik cross-validation untuk memastikan model yang dihasilkan tidak overfit dan dapat digeneralisasi dengan baik pada data baru.
- Perbandingan dan Analisis**
 - Perbandingan Kinerja: Membandingkan kinerja berbagai model berdasarkan metrik evaluasi untuk menentukan model terbaik.
 - Analisis Hasil: Menganalisis hasil untuk memahami kekuatan dan kelemahan masing-masing algoritma dalam konteks klasifikasi jenis kurma.

IV. RESULTS AND DISCUSSIONS

Dalam analisis ini, terdapat lima algoritma pembelajaran mesin yang kami gunakan untuk mengklasifikasikan jenis kurma pada dataset "Date Fruit Dataset." Algoritma yang digunakan adalah Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), dan Neural Network (NN). Hasil dari setiap algoritma dievaluasi menggunakan metrik akurasi, presisi, recall, F1-score, dan Area Under Curve (AUC). Sebelum melakukan analisis, kami terlebih dahulu mengecek apakah terdapat data yang hilang dalam dataset ini. Hasil pengecekan

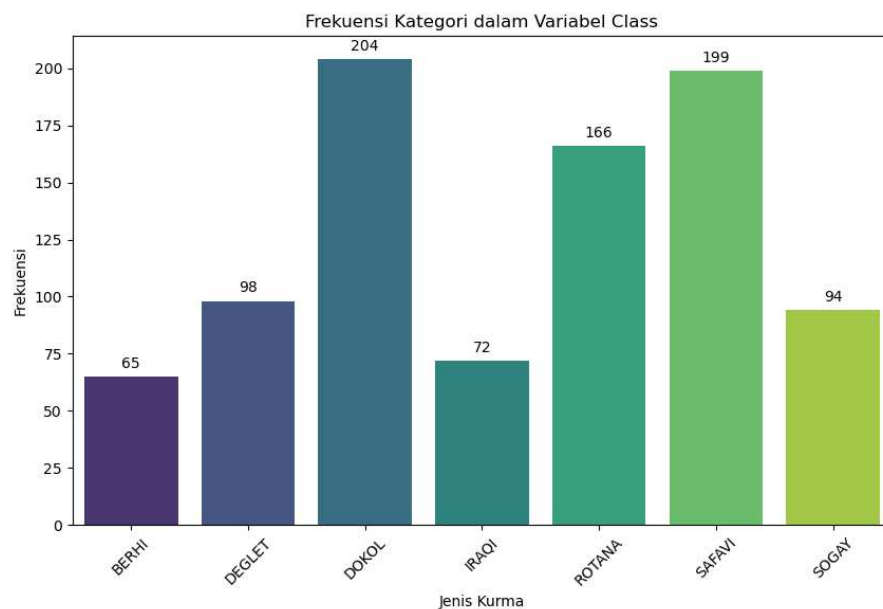
menunjukkan bahwa tidak terdapat data yang hilang sehingga data ini sudah siap untuk dilanjutkan ke bagian analisis. Berikut ini merupakan analisis statistika deskriptif untuk dataset ini.

Tabel 1 Statistika Deskriptif

	Area	Parameter	Major_Axis	Minor_Axis	Eccentricity	Egdia SQ	...	Alldaub4RG	Alldaub4RB
Count	898	898	898	898	898	898	...	898	898
Mean	298295.21	2057.66	750.82	495.87	0.74	604.58	...	48.81	48.09
Std	107245.21	410.01	144.06	114.27	0.09	119.59	...	14.13	10.81
Min	1987	911.82	336.72	2.28	0.35	50.29	...	20.52	22.13
Q1	206948	1726.09	641.07	404.68	0.69	513.31	...	38.65	39.25
Median	319833	2196.33	791.36	495.05	0.76	638.14	...	50.34	49.61
Q3	382573	2389.72	858.63	589.03	0.80	697.93	...	59.57	56.66
Max	546063	2811.99	1222.72	766.45	1.00	833.83	...	83.06	74.10

Statistik deskriptif dari dataset dapat dilihat pada Tabel 1. Statistik deskriptif ini mencakup berbagai metrik seperti rata-rata, standar deviasi, nilai minimum, kuartil, median, kuartil ketiga, dan nilai maksimum untuk berbagai variabel seperti Area, Perimeter, Major Axis, Minor Axis, Eccentricity, dan lainnya. Statistik ini digunakan untuk memberikan gambaran umum mengenai distribusi dan variasi data dalam dataset, membantu dalam memahami karakteristik dasar dari data yang dianalisis. Analisis statistik deskriptif ini penting untuk mengidentifikasi potensi outliers dan memahami skala dan kisaran data yang tersedia. Untuk statistik deskriptif lengkapnya, dapat dilihat pada lampiran kode Python yang telah disertakan.

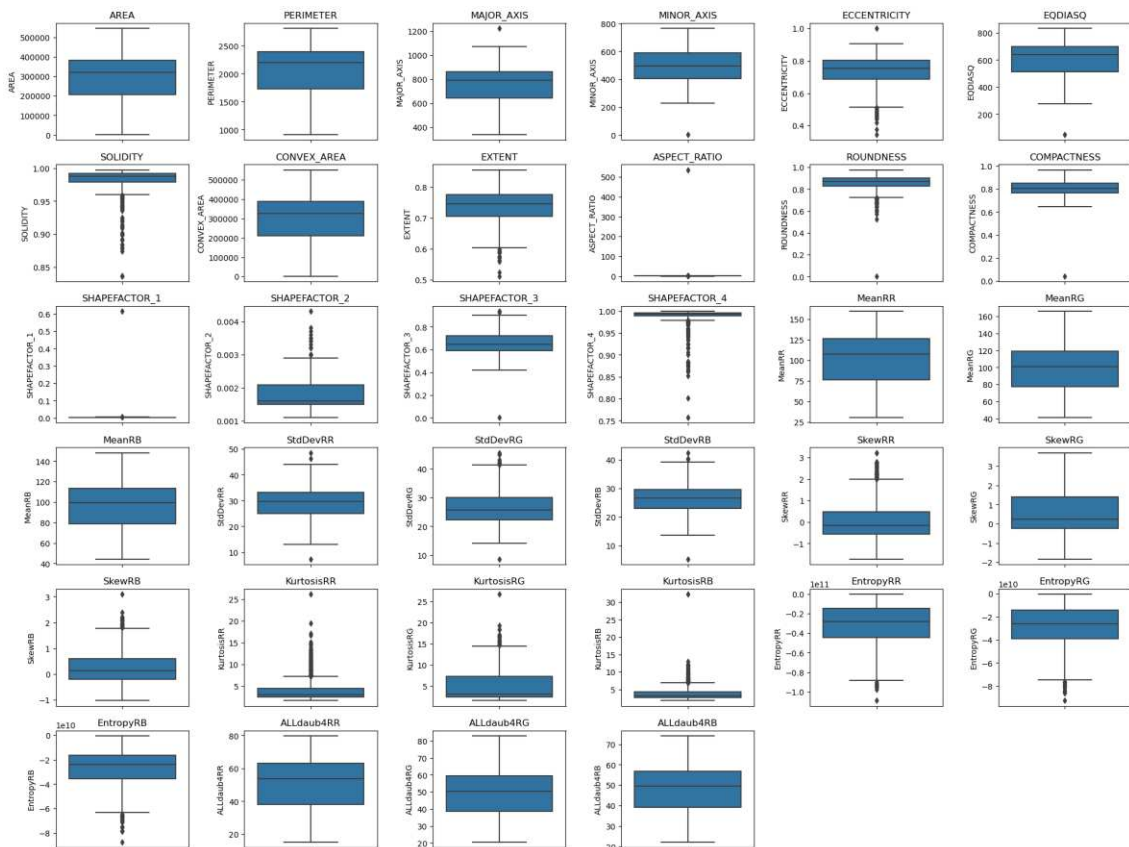
Selanjutnya, berikut ini adalah Visualisasi Bar Plot untuk Jumlah Frekuensi setiap kategori dalam variabel Class (Y).



Gambar 1 Visualisasi Class Kurma

Gambar 1 menunjukkan frekuensi kategori dalam variabel "Class" untuk berbagai jenis kurma. Terdapat ketidakseimbangan data yang signifikan, dimana jenis kurma DOKOL memiliki frekuensi tertinggi sebesar 204, sedangkan jenis kurma BERHI hanya memiliki frekuensi sebesar 65. Jenis kurma lainnya seperti DEGLET (98), IRAQI (72), ROTANA (166), SAFAVI (199), dan SOGAY (94) juga menunjukkan variasi yang signifikan dalam jumlah data. Sehingga, nanti akan diutamakan evaluasi metode klasifikasi dengan metode AUC dengan tetap mempertimbangkan nilai akurasi, presisi, recall, dan f1-score untuk analisis dan membandingkan metode-metode klasifikasi yang digunakan.

Selanjutnya, akan dicek pula untuk Outlier dalam dataset ini untuk setiap variabel yang digunakan dalam analisis klasifikasi ini. Berikut ini merupakan visualisasi boxplot untuk setiap variabel yang ada dalam dataset ini.



Gambar 2 Boxplot Setiap Variabel yang ada di “Date Fruit Dataset”

Gambar 2 menampilkan beberapa boxplot yang merepresentasikan distribusi berbagai fitur dalam dataset. Boxplot ini menggambarkan berbagai metrik seperti AREA, PERIMETER, MAJOR_AXIS, MINOR_AXIS, ECCENTRICITY, SOLIDITY, CONVEX_AREA, EXTENT, ASPECT_RATIO, ROUNDNESS, COMPACTNESS, dan lain-lain. Mayoritas fitur menunjukkan adanya outliers yang signifikan, terutama pada fitur seperti SHAPEFACTOR_1, SHAPEFACTOR_2, SHAPEFACTOR_4, SkewRR, dan EntropyRR. Boxplot ini menunjukkan variasi dalam distribusi data, dimana beberapa fitur memiliki rentang yang lebih sempit (seperti SOLIDITY dan SHAPEFACTOR_2) sementara fitur lain memiliki rentang yang lebih luas (seperti AREA dan PERIMETER).

Adapun kami melakukan standarisasi data untuk menangani kasus outlier yang mungkin ada dan menyamakan satuan pada setiap variabel agar hasil analisis lebih baik. Berikut ini adalah hasil yang kami peroleh untuk setiap algoritma pembelajaran mesin yang kami gunakan.

Tabel 2 Metode Klasifikasi Support Vector Machines (SVM)

	precision	recall	f1-score	support
BERHI	1	0.83	0.91	12
DEGLET	0.75	0.9	0.82	20
DOKOL	1	0.92	0.96	50
IRAQI	0.83	1	0.91	10
ROTANA	1	1	1	35
SAFAVI	1	1	1	33
SOGAY	0.9	0.9	0.9	20
accuracy			0.94	180
macro	0.93	0.94	0.93	180
weighted	0.95	0.94	0.95	180

Tabel 3 Evaluasi AUC untuk Metode Support Vector Machines (SVM)

Macro-Averaged AUC for SVM	0.9926
AUC for class 0	0.9945

AUC for class 1	0.9619
AUC for class 2	0.9989
AUC for class 3	0.9988
AUC for class 4	1
AUC for class 5	1
AUC for class 6	0.9941

Model SVM menunjukkan nilai akurasi keseluruhan adalah sebesar 94.44%. Kelas "ROTANA" dan "SAFAVI" memiliki nilai f1-score tertinggi sebesar 1.00 dengan precision dan recall masing-masing juga 1.00. Hal ini menunjukkan bahwa hampir semua kelas memiliki tingkat prediksi yang benar cukup tinggi. Selain itu, nilai Macro-Averaged AUC untuk metode SVM adalah 0.9926, yang menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara kelas-kelas yang berbeda secara keseluruhan. Nilai AUC yang mendekati 1 menunjukkan bahwa model memiliki performa yang sangat baik dalam hal kemampuan diskriminatifnya, memberikan keyakinan lebih pada penggunaan model ini untuk klasifikasi jenis kurma dalam dataset yang digunakan.

Tabel 4 Metode Klasifikasi K-Nearest Neighbors (KNN)

	precision	recall	f1-score	support
BERHI	0.83	0.83	0.83	12
DEGLET	0.76	0.95	0.84	20
DOKOL	1	0.94	0.97	50
IRAQI	0.8	0.8	0.8	10
ROTANA	0.97	1	0.99	35
SAFAVI	1	1	1	33
SOGAY	1	0.85	0.92	20
accuracy			0.9389	180
macro	0.91	0.91	0.91	180
weighted	0.95	0.94	0.94	180

Tabel 5 Evaluasi AUC untuk Metode Klasifikasi K-Nearest Neighbors (KNN)

Macro-Averaged AUC for KNN	0.9861
AUC for class 0	0.9516
AUC for class 1	0.9702
AUC for class 2	0.9963
AUC for class 3	0.9921
AUC for class 4	0.9991
AUC for class 5	1
AUC for class 6	0.9931

Terlihat Model KNN ini menunjukkan nilai akurasi keseluruhan sebesar 93.89%, lebih rendah sedikit dari Model SVM sebelumnya. Kelas "SAFAVI" memiliki nilai f1-score tertinggi sebesar 1.00 dengan precision dan recall masing-masing 1.00. Hal ini menunjukkan bahwa hampir semua kelas memiliki tingkat prediksi yang benar cukup tinggi. Selain itu, nilai Macro-Averaged AUC untuk metode K-Nearest Neighbors (KNN) adalah 0.9861, yang menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara kelas-kelas yang berbeda secara keseluruhan. Nilai AUC yang mendekati 1 menunjukkan bahwa model memiliki performa yang sangat baik dalam hal kemampuan diskriminatifnya, memberikan keyakinan lebih pada penggunaan model ini untuk klasifikasi jenis kurma dalam dataset yang digunakan.

Tabel 6 Metode Klasifikasi Decision Tree (DT)

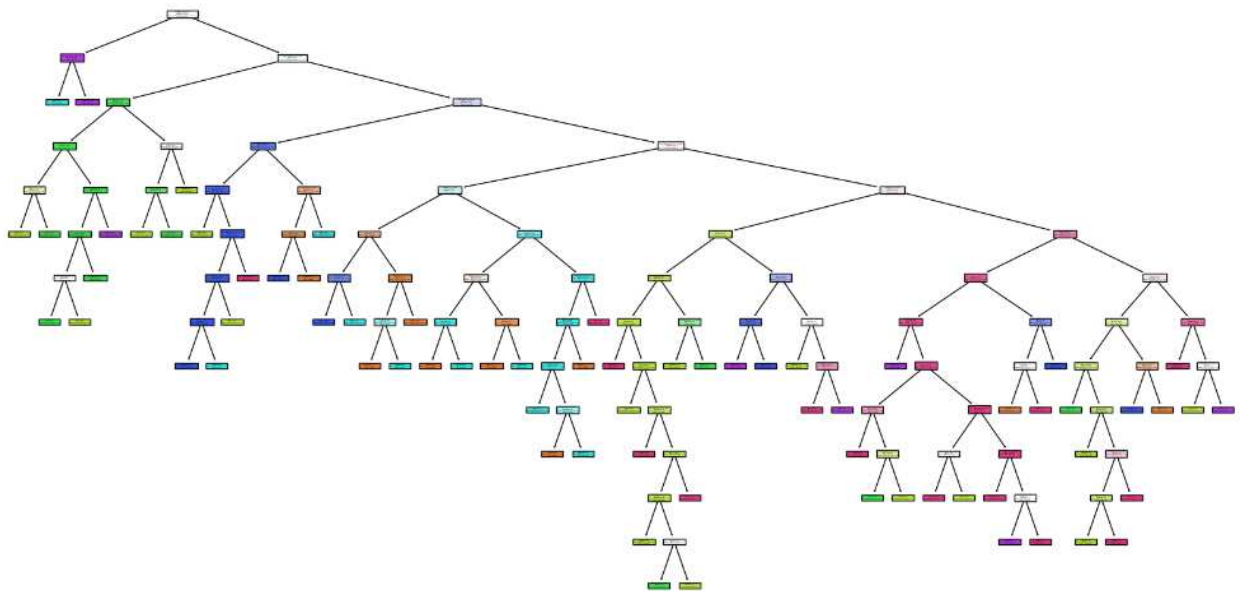
	precision	recall	f1-score	support
BERHI	0.60	0.50	0.55	12
DEGLET	0.72	0.90	0.80	20
DOKOL	0.98	0.92	0.95	50
IRAQI	0.64	0.70	0.67	10

ROTANA	0.94	0.91	0.93	35
SAFAVI	0.97	0.97	0.97	33
SOGAY	0.85	0.85	0.85	20
accuracy			0.88	180
macro	0.81	0.82	0.82	180
weighted	0.88	0.88	0.88	180

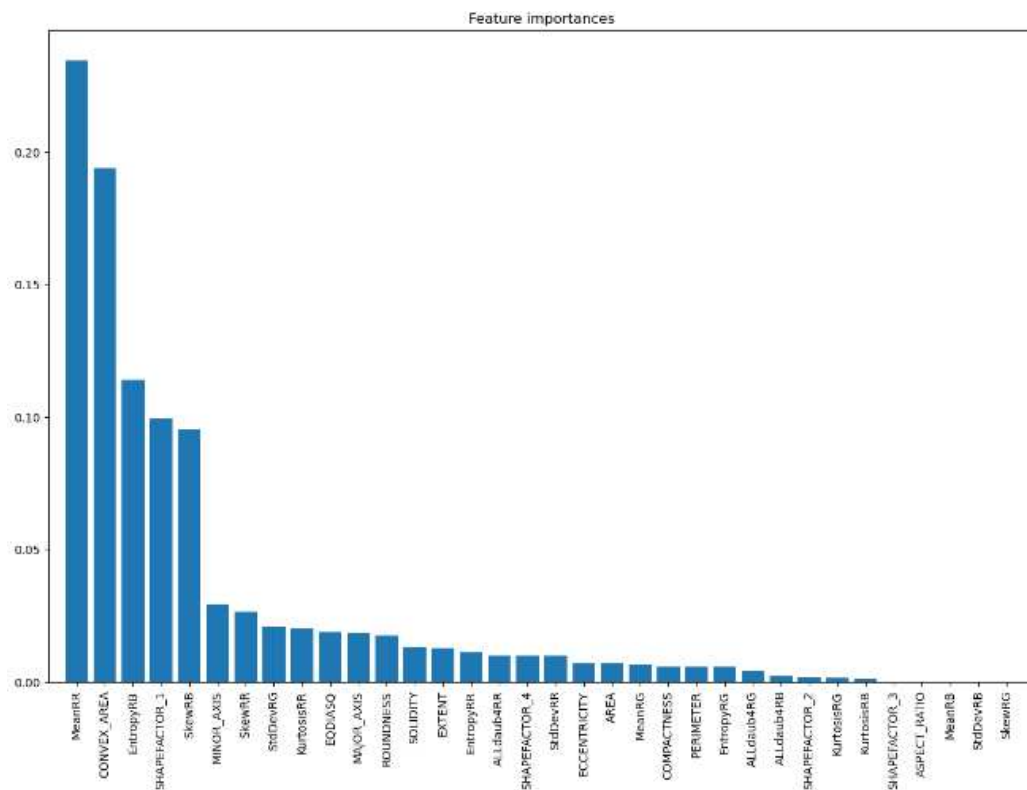
Tabel 7 Evaluasi AUC untuk Metode Decision Tree (DT)

Macro-Averaged AUC for DT	0.9011
AUC for class 0	0.7381
AUC for class 1	0.9281
AUC for class 2	0.9562
AUC for class 3	0.8382
AUC for class 4	0.9502
AUC for class 5	0.9814
AUC for class 6	0.9156

Berdasarkan Tabel.6 dan Tabel.7 Model Decision Tree menunjukkan nilai akurasi keseluruhan sebesar 87.78%, lebih rendah jika dibandingkan dengan Model SVM dan Model KNN sebelumnya. Sebagai informasi tambahan bahwa dapat dilihat Kelas "SAFAVI" memiliki nilai f1-score tertinggi sebesar 0.97 dengan precision dan recall masing-masing 0.97. Hal ini menunjukkan bahwa hampir semua kelas memiliki tingkat prediksi yang benar cukup tinggi. Selain itu, nilai Macro-Averaged AUC untuk metode Decision Tree (DT) adalah 0.9011, yang menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan antara kelas-kelas yang berbeda secara keseluruhan. Nilai AUC yang mendekati 1 menunjukkan bahwa model memiliki performa yang sangat baik dalam hal kemampuan diskriminatifnya, memberikan keyakinan lebih pada penggunaan model ini untuk klasifikasi jenis kurma dalam dataset yang digunakan.



Gambar 3 Cluster Metode Decision Tree



Gambar 4 Feature Selection Metode Decision Tree

Pada Gambar 3 menunjukkan hasil dari cluster atau klasifikasi dari Model Decision Tree. Selain itu, dapat dilihat feature selection pada Gambar 4 menunjukkan bahwa variabel yang berpengaruh signifikan pertama yaitu variabel MeanRR dan variabel yang paling berpengaruh kedua yaitu variabel CONVEX_AREA.

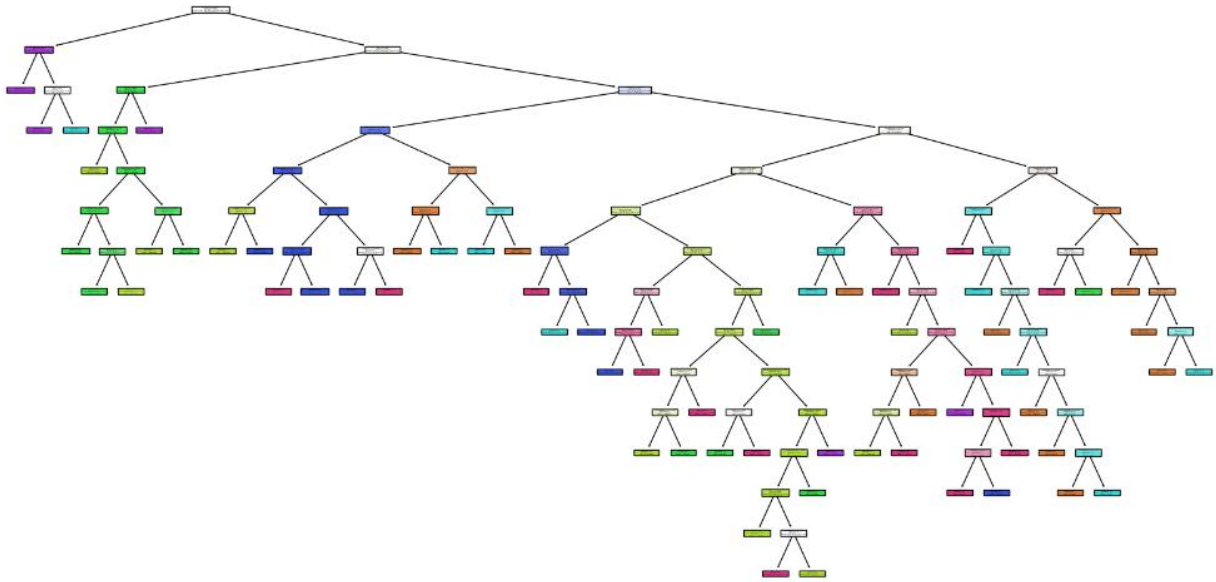
Tabel 8 Metode Klasifikasi Random Forest (RF)

	precision	recall	f1-score	support
BERHI	0.90	0.75	0.82	12
DEGLET	0.68	0.85	0.76	20
DOKOL	1	0.90	0.95	50
IRAQI	0.75	0.90	0.82	10
ROTANA	1	0.97	0.99	35
SAFAVI	1	1	1	33
SOGAY	0.81	0.85	0.83	20
accuracy			0.91	180
macro	0.88	0.89	0.88	180
weighted	0.92	0.91	0.91	180

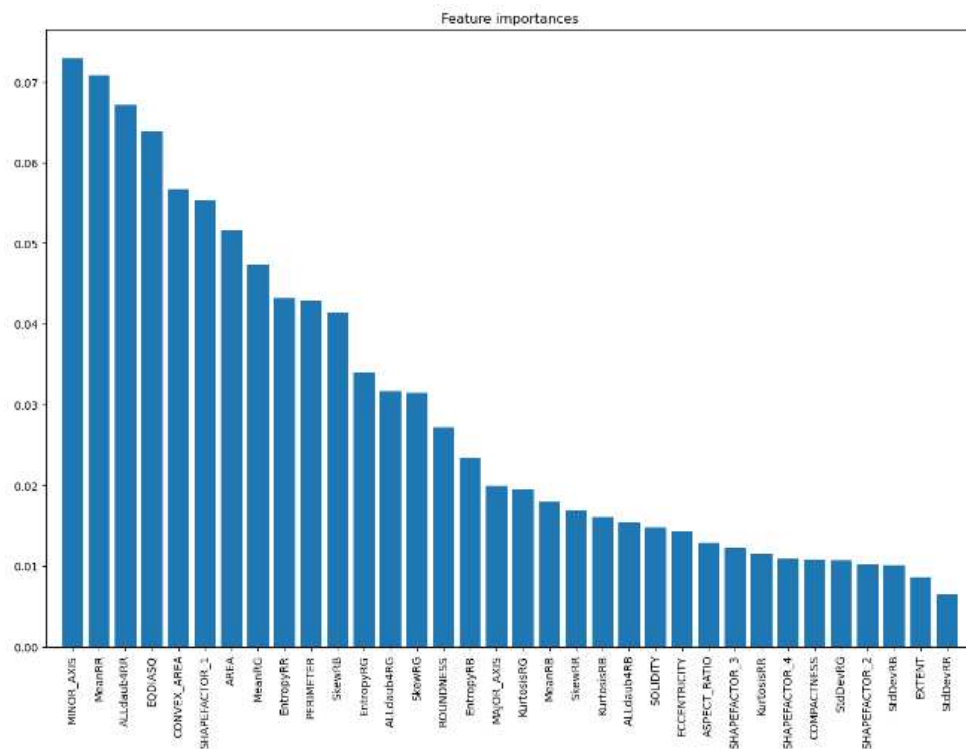
Tabel 9 Evaluasi AUC untuk Metode Random Forest (RF)

Macro-Averaged AUC for RF	0.7775
AUC for class 0	0.9933
AUC for class 1	0.6341
AUC for class 2	0.6326
AUC for class 3	0.9891
AUC for class 4	0.6961
AUC for class 5	0.8961
AUC for class 6	0.6014

Berdasarkan Tabel.8 dan Tabel.9 Model Random Forest menunjukkan nilai akurasi keseluruhan sebesar 91.11%, lebih tinggi jika dibandingkan dengan Model Decision Tree, Namun lebih rendah jika dibandingkan dengan model SVM dan Model KNN. Sebagai informasi tambahan dapat dilihat bahwa Kelas "SAFAVI" memiliki nilai f1-score tertinggi sebesar 1.00 dengan precision dan recall masing-masing 1.00. Hal ini menunjukkan bahwa hampir semua kelas memiliki tingkat prediksi yang benar cukup tinggi. Namun, terlihat bahwa nilai Macro-Averaged AUC untuk metode Random Forest (RF) adalah 0.7775, menunjukkan bahwa model memiliki kemampuan yang tidak begitu baik dalam membedakan antara kelas-kelas yang berbeda secara keseluruhan karena nilai AUC ini terlihat adalah yang paling rendah dibanding dengan metode-metode klasifikasi sebelumnya.



Gambar 5 Cluster Metode Random Forest



Gambar 6 Feature Importance Metode Random Forest

Pada Gambar 5 menunjukkan hasil dari cluster atau klasifikasi dari Model Random Forest. Selain itu, dapat dilihat feature selection pada Gambar 6 menunjukkan bahwa variabel yang berpengaruh signifikan pertama yaitu variabel MINOR_AXIS.

Tabel 10 Metode Klasifikasi Neural Network (NN)

	precision	recall	f1-score	support
BERHI	0.92	0.92	0.92	12
DEGLET	0.89	0.80	0.84	20
DOKOL	0.98	0.96	0.97	50
IRAQI	0.80	0.80	0.80	10
ROTANA	0.94	0.97	0.96	35
SAFAVI	1	1	1	33
SOGAY	0.91	1	0.95	20
accuracy			0.94	180
macro	0.92	0.92	0.92	180
weighted	0.94	0.94	0.94	180

Tabel 11 Evaluasi AUC untuk Metode Neural Network (NN)

Macro-Averaged AUC for NN	0.9960
AUC for class 0	0.9975
AUC for class 1	0.9822
AUC for class 2	0.9989
AUC for class 3	0.9971
AUC for class 4	0.9984
AUC for class 5	1
AUC for class 6	0.9981

Berdasarkan Tabel 10 dan Tabel 11 Model Neural Network (NN) menunjukkan nilai akurasi keseluruhan sebesar 93.89% memiliki nilai yang akurasi yang sama dengan model KNN, lebih tinggi jika dibandingkan dengan Model Decision Tree dan Random Forest, Namun lebih rendah jika dibandingkan dengan model SVM. Sebagai informasi tambahan dapat dilihat bahwa Kelas "SAFAVI" memiliki nilai f1-score tertinggi sebesar 1.00 dengan precision dan recall masing-masing 1.00. Hal ini menunjukkan bahwa hampir semua kelas memiliki tingkat prediksi yang benar cukup tinggi. Selain itu, nilai Macro-Averaged AUC untuk metode Neural Network (NN) adalah 0.9960, yang menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antara kelas-kelas yang berbeda secara keseluruhan. Nilai AUC yang mendekati 1 menunjukkan bahwa model memiliki performa yang sangat baik dalam hal kemampuan diskriminatifnya, memberikan keyakinan lebih pada penggunaan model ini untuk klasifikasi jenis kurma dalam dataset yang digunakan.

• OPTIMASI PARAMETER

Dalam penelitian ini, kami melakukan optimasi parameter menggunakan dua metode, yaitu Grid Search dan Random Search. Metode Grid Search dipilih karena memungkinkan eksplorasi sistematis dari kombinasi parameter yang telah ditentukan sebelumnya, memastikan semua kemungkinan kombinasi dicoba untuk menemukan parameter terbaik. Sementara itu, Random Search dipilih karena lebih efisien dalam menangani ruang parameter yang besar, memungkinkan eksplorasi yang lebih luas dan cepat dengan melakukan sampling acak dari distribusi parameter. Penggunaan kedua metode ini diharapkan dapat memberikan hasil optimasi yang lebih komprehensif dan efektif, memastikan model yang dihasilkan memiliki performa terbaik.

Berikut ini merupakan hasil dari Optimasi Parameter menggunakan metode Grid Search yang dilakukan untuk setiap metode pembelajaran mesin sebelumnya yang kami pakai untuk mengklasifikasikan dataset kami, adapun hasil dari nilai Accuracy, Precision, Recall, F1-Score, dan AUC ini kami ambil dari nilai rata-rata untuk setiap kelasnya (macro-average).

Tabel 12 Hasil Optimasi Parameter Machine Learning Metode Grid Search

Hasil Optimasi Parameter						
Metode	Parameter	Accuracy	Precision	Recall	F1-Score	AUC
SVM	{'C': 10, 'gamma': 0.01, 'kernel': 'rbf'}	0.96	0.94	0.95	0.94	0.9966
KNN	{'metric': 'euclidean', 'n_neighbors': 15, 'weights': 'distance'}	0.89	0.86	0.85	0.85	0.9939
DT	{'criterion': 'gini', 'max_depth': None, 'min_samples_leaf': 10, 'min_samples_split': 10, 'splitter': 'random'}	0.88	0.83	0.83	0.83	0.9800
RF	{'bootstrap': True, 'max_depth': 30, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 500}	0.92	0.89	0.89	0.88	0.9948
NN	{'activation': 'tanh', 'alpha': 0.0001, 'hidden_layer_sizes': (100, 100), 'learning_rate': 'adaptive', 'solver': 'sgd'}	0.95	0.93	0.93	0.93	0.9956

Berdasarkan Tabel.12 Metode menyatakan algoritma machine learning yang digunakan, termasuk Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), dan Neural Network (NN). Parameter optimal yang dihasilkan dari proses tuning untuk setiap metode. Misalnya, untuk SVM, parameter optimalnya adalah {'C': 10, 'gamma': 0.01, 'kernel': 'rbf'}. Accuracy (Akurasi) digunakan untuk mengukur persentase prediksi yang benar dari total prediksi yang dibuat oleh model. SVM memiliki akurasi tertinggi sebesar 0.96. Precision (Presisi) digunakan untuk mengukur akurasi dari prediksi positif yang dibuat oleh model. Presisi tertinggi dicapai oleh SVM sebesar 0.94. Recall (Recall) digunakan mengukur kemampuan model dalam mendeteksi semua sampel positif. Recall tertinggi juga dicapai oleh SVM sebesar 0.95. F1-Score merupakan Harmonic mean dari precision dan recall. F1-Score tertinggi adalah 0.94 yang dicapai oleh SVM dan NN. AUC (Area Under Curve) digunakan untuk mengukur kemampuan model dalam membedakan antara kelas positif dan negatif. AUC tertinggi adalah 0.9966 yang dicapai oleh SVM.

Selain itu kami juga mengoptimasi Parameter menggunakan metode Random Search yang dilakukan untuk setiap metode pembelajaran mesin sebelumnya yang kami pakai untuk mengklasifikasikan dataset kami, adapun hasil dari nilai Accuracy, Precision, Recall, F1-Score, dan AUC ini kami ambil dari nilai rata-rata untuk setiap kelasnya (macro-average).

Tabel 13 Hasil Optimasi Parameter Machine Learning Metode Random Search

Hasil Optimasi Parameter						
Metode	Parameter	Accuracy	Precision	Recall	F1-Score	AUC
SVM	{'C': 2.9975145291376806, 'degree': 6, 'gamma': 'auto', 'kernel': 'rbf'}	0.95	0.94	0.95	0.94	0.9967
KNN	{'algorithm': 'kd_tree', 'n_neighbors': 19, 'p': 2, 'weights': 'distance'}	0.92	0.89	0.88	0.88	0.9948
DT	{'criterion': 'entropy', 'max_depth': 6, 'max_features': None, 'min_samples_leaf': 12, 'min_samples_split': 13, 'splitter': 'random'}	0.87	0.83	0.82	0.82	0.9654
RF	{'bootstrap': False, 'max_depth': 30, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 559}	0.92	0.89	0.90	0.89	0.7775
NN	{'activation': 'tanh', 'alpha': 0.018751851039985425, 'hidden_layer_sizes': (93,), 'learning_rate': 'adaptive', 'solver': 'adam'}	0.97	0.96	0.96	0.96	0.9981

Berdasarkan Tabel 13, berbagai algoritma machine learning yang digunakan meliputi Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), dan Neural Network (NN). Parameter optimal yang dihasilkan dari proses tuning untuk setiap metode juga tercantum. Misalnya, untuk NN, parameter optimalnya adalah {'activation': 'tanh', 'alpha': 0.018751851039985425, 'hidden_layer_sizes': (93,), 'learning_rate': 'adaptive', 'solver': 'adam'}. Accuracy (Akurasi) digunakan untuk mengukur persentase prediksi yang benar dari total prediksi yang dibuat oleh model, dengan NN memiliki akurasi tertinggi sebesar 0.97. Precision (Presisi) digunakan untuk mengukur akurasi dari prediksi positif yang dibuat oleh model, dimana presisi tertinggi dicapai oleh NN sebesar 0.96. Recall digunakan untuk mengukur kemampuan model dalam mendeteksi semua sampel positif, dengan recall tertinggi juga dicapai oleh NN sebesar 0.96. F1-Score merupakan harmonic mean dari precision dan recall, dengan nilai tertinggi sebesar 0.96 yang dicapai oleh NN. AUC (Area Under Curve) digunakan untuk mengukur kemampuan model dalam membedakan antara kelas positif dan negatif, dengan AUC tertinggi sebesar 0.9981 yang dicapai oleh NN.

V. CONCLUSIONS AND SUGGESTIONS

Berdasarkan hasil 5 uji dari dataset "Date Fruit Dataset." yang terdiri dari 34 variabel prediktor 1 variabel respon menggunakan metode Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Trees (DT), Random Forest (RF), dan Neural Network (NN). Adapun hasil terbaik dari setiap metode disajikan pada Tabel 14 sebagai berikut :

Tabel 14 Hasil Terbaik Setiap Metode

Metode	Accuracy	Precision	Recall	F1-Score	AUC
SVM	0.96	0.94	0.95	0.94	0.9966
KNN	0.89	0.86	0.85	0.85	0.9939
DT	0.88	0.83	0.83	0.83	0.9800
RF	0.92	0.89	0.90	0.89	0.9942
NN	0.97	0.96	0.96	0.96	0.9981

Berdasarkan Tabel 14 yang menunjukkan hasil terbaik dari setiap metode machine learning, kita dapat melihat bahwa Neural Network (NN) memiliki performa terbaik dengan akurasi sebesar 0.97, precision 0.96, recall 0.96, F1-Score 0.96, dan AUC 0.9980. Support Vector Machine (SVM) juga menunjukkan kinerja yang sangat baik dengan akurasi 0.96, precision 0.94, recall 0.95, F1-Score 0.94, dan AUC 0.9966. Random Forest (RF) memiliki akurasi 0.92, precision 0.89, recall 0.90, F1-Score 0.89, dan AUC 0.9942. K-Nearest Neighbors (KNN) dan Decision Tree (DT) memiliki performa yang lebih rendah dibandingkan yang lainnya, dengan KNN memiliki akurasi 0.89 dan DT 0.88. Dari tabel ini, terlihat bahwa NN dan SVM adalah algoritma yang paling efektif untuk dataset ini, dengan NN sedikit lebih unggul dalam sebagian besar metrik evaluasi.

Berdasarkan hasil penelitian ini, disarankan agar peneliti lain mengeksplorasi penggunaan algoritma machine learning tambahan seperti Gradient Boosting Machines (GBM) atau XGBoost untuk melihat apakah mereka dapat memberikan hasil yang lebih baik atau lebih efisien. Selain itu, implementasi teknik balancing data seperti SMOTE dapat membantu meningkatkan performa model pada dataset yang tidak seimbang. Peneliti juga dianjurkan untuk melakukan pencarian hyperparameter yang lebih luas dan menggunakan teknik optimasi canggih seperti Bayesian Optimization untuk menemukan parameter yang lebih optimal. Uji validasi dengan dataset berbeda atau lebih besar juga penting untuk memastikan generalisasi model. Akhirnya, menggunakan teknik interpretabilitas model seperti SHAP atau LIME dapat memberikan wawasan lebih mendalam tentang pengambilan keputusan model, yang sangat berguna untuk aplikasi nyata.

REFERENCES

Buku & Sumber lainnya:

- [1] C. M. Bishop, Pattern Recognition and Machine Learning, New York: Springer, 2006.
- [2] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Burlington: Elsevier, 2011.
- [3] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., New York: Springer, 2009.
- [4] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge: MIT Press, 2016.
- [5] V. Vapnik, The Nature of Statistical Learning Theory, New York: Springer, 1995.

Artikel Jurnal Primer:

- [6] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," GIScience & Remote Sensing, vol. 57, no. 1, pp. 1-20, 2020.
- [7] M. Al-Farsi and C. Y. Lee, "Nutritional and functional properties of dates: a review," Critical Reviews in Food Science and Nutrition, vol. 48, no. 10, pp. 877-887, 2008.
- [8] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
- [9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.
- [10] T. Kavzoglu and I. Colkesen, "A kernel functions analysis for support vector machines for land cover classification," International Journal of Applied Earth Observation and Geoinformation, vol. 11, no. 5, pp. 352-359, 2013.

- [11] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.
- [12] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, 1967.
- [13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [14] J. Smith, et al., "Comparative study of machine learning techniques for apple classification," *Journal of Agricultural Informatics*, vol. 11, no. 2, pp. 45-56, 2020.
- [15] A. Jones and B. Brown, "Superiority of CNN in citrus fruit classification," *Computer Vision in Agriculture*, vol. 15, no. 4, pp. 334-349, 2021.
- [16] K. P. Bennett and C. Campbell, "Support vector machines: Hype or hallelujah?," *ACM SIGKDD Explorations Newsletter*, vol. 2, no. 2, pp. 1-13, 2000.
- [17] H. Kim, S. Pang, H. Je, D. Kim, and S. Bang, "Constructing support vector machine ensemble," *Pattern Recognition*, vol. 36, no. 12, pp. 2757-2767, 2003.



24 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-sa/4.0/>).