Prediksi Kualitas Kopi Dengan Algoritma Random Forest Melalui Pendekatan Data Science

Kalvintirta Ciptady[1)*], Mawaddah Harahap[2)], Jonvin[3)], Yonata Ndruru[4)], Ibadurrahman[5)]
[1,2,3,4,5)] Universitas Prima Indonesia,

Received: 5 Sep 2022 Accepted: 7 Sep 2022 Published: 8 Sept 2022

***kalvintirta@gmail.com**

Kata Kunci: Kualitas Kopi, Prediksi Kualitas, Random Forest, Data Science

Abstrak : Perusahaan yang bergerak dibidang produksi kopi, selalu mementingkan kualitas kopi untuk menghasilkan produk yang dapat bersaing dengan kompetitor lainnya. Adapun sistem yang dapat dibangun untuk mengatasi permasalahan tersebut yaitu berupa prediksi dalam menentukan kualitas kopi. Data kopi yang digunakan pada penelitian ini didapatkan dari Coffee Quality Institute. Data ini memiliki 44 kolom atau variabel dengan jumlah data sebanyak 1339 data. Tahapan yang dilakukan dengan pendekatan data science dengan algoritma random forest terdiri dari proses pengumpulan data, preprocessing, pengumpulan data, split data, pemrosesan dengan algoritma Random Forest yang menghasilkan hasil prediksi, hingga yang terakhir adalah proses evaluasi performa algoritma Random Forest dalam memprediksi kualitas kopi. Variabel dependen yang diprediksi pada penelitian ini secara berurutan dari kualitas terbaik hingga terburuk antara lain adalah kualitas kopi Specialty Grade, Premium, Exchange, dan Below Standard. Kualitas kopi premium adalah hasil prediksi paling baik dengan hasil actual 135. Jumlah data yang cukup untuk memastikan algoritma random forest dapat memprediksi dengan cukup baik dengan akurasi yang mencapai 79% masih memiliki ruang untuk perkembangan sehingga bisa mendekati 100%. Kurang optimalnya nilai akurasi pada penelitian ini dapat diakibatkan oleh kurangnya variabel independen yang digunakan. Penelitian ini hanya menggunakan 8 variabel dari 43 variabel yang tersedia. Sehingga masih terdapat variabel yang berpotensi dapat meningkatkan akurasi. Tidak dilakukannya penyetelan parameter Random Forest yang tersedia untuk meningkatkan akurasi.

PENDAHULUAN

Perusahaan yang bergerak dibidang produksi kopi, selalu mementingkan kualitas kopi untuk menghasilkan produk yang dapat bersaing dengan kompetitor lainnya. Kopi merupakan salah satu hasil komoditi perkebunan yang memiliki ekonomis yang cukup tinggi dan berperan penting sebagai sumber penghasil devisa negara [1]. Perusahaan yang bergerak dibidang produksi kopi, selalu mementingkan kualitas kopi untuk menghasilkan produk yang dapat bersaing dengan kompetitor lainnya. Dalam penentuan kualitas kopi, Perusahaan menggunakan persepsi personal dan analisa secara manual. Hal ini pun menjadikan proses pemilihan kopi berkualitas menjadi proses yang cukup memakan waktu dan tenaga. Sedangkan untuk memenuhi permintaan pasar dan meningkatkan kinerja produksi, dibutuhkan proses yang lebih cepat dan akurat. Peran teknologi pun dinilai dapat membantu perusahaan dalam meningkatkan efisiensi dan kualitas produksi. Dalam menentukan kopi berkualitas, perlu adanya sistem yang tepat untuk menganalisa permasalahan tersebut [2].

Adapun sistem yang dapat dibangun untuk mengatasi permasalahan tersebut yaitu berupa prediksi dalam menentukan kualitas kopi. Prediksi adalah proses perkiraan tentang sesuatu yang akan terjadi pada waktu yang akan datang berdasarkan data yang ada [3]. Prediksi dapat diterapkan dalam berbagai bidang

termasuk salah satunya prediksi terkait kualitas biji kopi.

Dalam beberapa tahun terakhir, beberapa penelitian telah dilakukan dalam menentukan dan memprediksi kualitas kopi. Contohnya seperti penelitian yang dilakukan oleh Y. Christian & I. Darmawan dalam penelitian [4] yang melakukan prediksi kopi dengan menerapkan metode General Regression Neural Network (GRNN). Penelitian lain yang dilakukan oleh A. Ginanjar dalam penelitian [5] dengan merancang sistem deteksi jenis cacat biji kopi dengan menerapkan algoritma K-Nearest Neighbor (KNN). Kemudian penelitian lain yang dilakukan oleh I. Ilhamsyah dkk. dalam penelitian [6] yang melakukan klasifikasi terhadap kualitas biji kopi menggunakan MultilayerPerceptron berbasis fitur warna LCH.

Dalam melakukan pemeriksaan kopi secara manual menyebabkan proses yang cukup memakan waktu. Berdasarkan hal tersebut dilakukan prediksi untuk mengetahui kualitas kopi. Dengan demikian, terdapat permasalahan bagaimana tingkat akurasi dari metode yang digunakan dalam memprediksi kualitas kopi. Adapun manfaat dari penelitian ini berdasarkan hasil yang didapat nantinya dapat digunakan sebagai acuan dalam menyaring kopi sesuai hasil kualitas yang diprediksi secara sistem sehingga tidak memakan banyak waktu untuk mengetahui kualitas kopi dan juga dapat menjadi solusi untuk membantu meningkatkan pemasaran kopi yang berkualitas.

TINJAUAN LITERATUR

A. Majid dkk. (2022) pada [7] melakukan penelitian mengenai "Identifikasi Kualitas fisik Pada Biji Kopi Menggunakan Tekonologi Pengolahan Citra Dengan Metode Neural Network" dengan memperoleh hasil berupa aplikasi untuk identifikasi kualitas biji kopi yang terbukti mampu untuk menghasilkan seleksi dari biji kopi dan hasil akurasi dari uji coba yang dilakukan terhadap citra pada gambar yaitu warna biru menunjukkan kualitas fisik biji bagus dengan akurasi 80,41%, dan frame yang berwarna hijau menunjukkan kualitas fisik biji sedang dengan akurasi 90,99% sedangkan frame yang berwarna merah menunjukkan kualitas fisik biji rusak dengan akurasi 99,75%. S. Sidehabi dkk. (2022) pada [8] melakukan penelitian mengenai "Alat Penyortir Biji Kopi otomatis dengan menggunakan Mikrokotroller Arduino Nano pada Usaha Mikro Kecil dan Menengah (UMKM) Toraja Kawaa Roasters" dengan kesimpulan bahwa pemanfaatan alat sortir biji kopi dapat membantu mempercepat proses klasifikasi kualitas biji kopi yang akan digunakan.

F. Ruzaik (2022) pada [9] melakukan penelitian mengenai "Prediksi Kualitas Biji Kopi Menggunakan Metode Segmentasi Dan Content Based Image Retrieval(Cbir)" dengan perolehan dari hasil pengujian, sistem content based image retrieval dapat menemukan citra yang dicari berdasarkan perhitungan euclidean distance vector citra query dengan vector citra yang ada didalam database, dengan rata-rata persentase precission tingkat keberhasilan sistem sebesar 87% dan nilai Recall sebesar 7,096% dan waktu komputasi rata-rat 2,0202 detik dengan menggunakan 50 data uji terdiri dari 25 citra bagus dan 25 citra buruk.

METODE PENELITIAN

Langkah-langkah yang dilakukan dalam penelitian ini dapat dijabarkan sebagai berikut:

1. Pengumpulan Dataset
   Pengumpulan data dilakukan dengan cara melakukan pengambilan data yang berkaitan dengan objek penelitian. Data yang diambil berupa biji kopi yang selanjutnya akan diolah menjadi minuman kopi. Dan data-data algoritma dari jurnal yang sudah dipelajari.
2. Pengolahan Dataset
   Data kopi yang digunakan pada penelitian ini didapatkan dari Coffee Quality Institute. Data ini juga telah diteliti oleh peneliti lainnya untuk memprediksi kualitas kopi menggunakan algoritma Neural Network atau Jaringan Saraf Tiruan [4]. Data ini memiliki 44 kolom atau variabel dengan jumlah data sebanyak 1339 data. Variabel dependen yang diprediksi pada penelitian ini secara berurutan dari kualitas terbaik hingga terburuk antara lain adalah kualitas kopi Specialty Grade, Premium, Exchange, dan Below Standard. Ekspektasi dalam penelitian ini adalah berhasil menggunakan data variabel independen untuk memprediksi kualtias kopi tersebut.
3. Pre-processing

Pre-processing data dilakukan dengan memilih variabel independen dan dependen, mengganti tipe data sesuai dengan kebutuhan dan membagi data menjadi data training dan data testing.

HASIL PENELITIAN

Hasil pengumpulan data kopi menghasilkan 1.339 data spesifikasi kopi yang memiliki 44 kolom atau variable. Variabel dependen pada data tersebut adalah Kualitas Kopi sedangkan sisanya akan dibersihkan dan diproses untuk digunakan sebagai variabel independen. Data tersebut didapatkan dari Coffee Quality Institute yang juga telah diteliti oleh peneliti sebelumnya menggunakan teknik Deep Learning. Dengan meneliti data spefisikasi kopi, maka perusahaan produsen kopi dapat mulai mengumpulkan data speifikasi kopi yang serupa ataupun berbeda untuk menghasilkan prediksi kualitas kopi.

Langkah selanjutnya setelah data kopi berhasil dikumpulkan adalah mengikuti alur proses yang telah dibahas sebelumnya, antara lain proses pengumpulan data, preprocessing, pengumpulan data, split data, pemrosesan dengan algoritma Random Forest yang menghasilkan hasil prediksi, hingga yang terakhir adalah proses evaluasi performa algoritma Random Forest dalam memprediksi kualitas kopi. Hasil yang diharapkan adalah akurasi yang mendekati 100%, yakni menandakan bahwa kualitas kopi dapat diprediksi dengan sangat akurat menggunakan algoritma Random Forest yang dirancang dan diteliti pada penelitian ini.

Preprocessing data sangat penting karena mencegah berkurangnya akurasi algoritma dalam memprediksi kualitas kopi. Jadi, sebelum mengolah data, kita harus memastikan bahwa data yang akan kita gunakan merupakan data "bersih". Mengamati kekosongan data penting untuk mencegah kebingungan pada algoritma Machine Learning dalam pelatihan untuk memprediksi kualitas kopi. Kekosongan data dapat diakibatkan mulai dari kesalahan manusia selama entri data, pembacaan sensor yang salah, hingga bug perangkat lunak dalam jalur pemrosesan data. Data yang hilang mungkin dapat mempengaruhi rendahnya performa ataupun akurasi dari algoritma Machine Learning.

Kekosongan data pasti sering ditemui ketika saat sedang melakukan sebuah penelitian. Kekosongan data dapat ditemukan pada beberapa variabel independen, seperti Lot. Number, Mill, Company, dan sebagainya. Pada penelitian ini penulis mengatasi masalah kekosongan data dengan menghapus data yang kosong. Kemudian memilih sebagian besar data untuk diteliti. Penelitian data dengan variabel yang lebih banyak merupakan diluar dari ruang lingkup penelitian ini dan dapat dijadikan sebagai penelitian selanjutnya dengan tingkat penelitian yang lebih rumit dan detil. Lantas demikian Tabel 1 menampilkan variabel-variabel yang dipilih pada penelitian ini.

Tabel 1. Variabel-Variabel Data Kopi yang Digunakan

Nama variabel yang digunakan	**Pengertian**
Country.of.Origin	Negara asal kopi diproduksi
Harvest.Year	Tahun panen
Variety	Varietas kopi (contoh: Arabika atau Robusta)
Processing.Method	Metode pemrosesan kopi
Category.One.Defects	Terdapat satu kecacatan pada biji kopi
Category.Two.Defects	Terdapat dua kecacatan pada biji kopi
altitude_mean_meters	Rata-rata ketinggian pada kebun tempat kopi ditanam
Total.Cup.Points	Total poin kualitas kopi

Selanjutnya adalah mengamati tipe data, hal ini dilakukan dengan mengamati data setelah masalah kekosongan data diatasi seperti pada Gambar 1 Terlihat ada negara asal, tahun panen, varietas kopi, metode pemrosesan kopi, kecacatan pada kopi, rata-rata ketinggian tumbuhan kopi, dan total poin kualitas yang merupakan variabel-variabel yang harus ada untuk menguji kualitas kopi.

```
df.head()
```

	Country.of.Origin	Harvest.Year	Variety	Processing.Method	Category.One.Defects	Category.Two.Defects	Quakers	altitude_mean_meters	Total.Cup.Points
0	Ethiopia	2014	Other	Washed / Wet	0	1	0.0	2075.0	89.92
1	Ethiopia	2014	Other	Washed / Wet	0	2	0.0	2075.0	88.83
2	Ethiopia	2014	Other	Natural / Dry	0	4	0.0	1822.5	88.25
3	United States	2014	Other	Washed / Wet	0	0	0.0	1872.0	87.92
4	United States	2014	Other	Washed / Wet	0	0	0.0	1943.0	87.92

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 975 entries, 0 to 974
Data columns (total 9 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Country.of.Origin     975 non-null    object
 1   Harvest.Year          975 non-null    object
 2   Variety               975 non-null    object
 3   Processing.Method     975 non-null    object
 4   Category.One.Defects  975 non-null    int64
 5   Category.Two.Defects  975 non-null    int64
 6   Quakers               975 non-null    float64
 7   altitude_mean_meters  975 non-null    float64
 8   Total.Cup.Points      975 non-null    float64
dtypes: float64(3), int64(2), object(4)
memory usage: 68.7+ KB
```

Gambar 1. Hasil pengamatan data setelah dibersihkan dari data kosong

Langkah selanjutnya adalah mengatasi tipe data yang kurang sesuai seperti yang ditampilkan pada Tabel 2. Untuk mengurangi variasi data negara yang terlalu beragam, penulis mengkelompokan data negara asal menjadi 'Others' untuk negara yang memiliki data dibawah 10.

Tabel 2 Pemrosesan Data

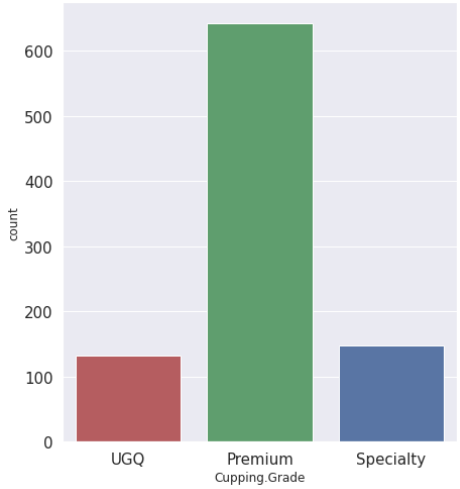
Variabel	**Jenis Pemrosesan Data**
Country of Origin	Mengelompokkan data minoritas menjadi 'other'
Harvest Year	Label encoding dari tipe data string menjadi integer
Variety	Mengelompokkan data minoritas menjadi 'other'
Processing	Label encoding dari tipe data string menjadi integer
Category	-
Quakers	-
Altitude	Menghapus data yang terlalu besar atau rendah
Total Cup Points	Transformasi data menjadi tiga kualitas kopi (1=Specialty Quality, 2 = Premium Quality, 3 = Usually Good Quality)

Perubahan menjadi bentuk One-hot encoding (misalnya 2018 bernilai 1, dan 2017 bernilai 0) tidak dilakukan karena data tahunan tersebut memiliki hirarki antara yang lebih besar dan lebih tinggi. Selain itu tipe data pada variabel harvest year juga perlu dikonversikan menjadi integer. Hal ini perlu dilakukan karena algoritma Machine Learning tidak mengerti tipe data berbentuk teks. Sehingga data tipe integer lebih diprioritaskan agar dapat mempermudah penilaian.

Kemudian data Country of Origin atau negara asal pembuatan kopi juga perlu disederhanakan. Hal ini dilakukan dengan mengubah data yang berjumlah hanya 1 data agar diubah menjadi 'others'. Hal ini perlu dilakukan karena jumlah informasi mengenai data tersebut terlalu sedikit sehingga tidak dapat memberikan pembelajaran yang signifikan kepada algoritma Machine Learning. Namun apabila dikategorikan menjadi 'other' maka akan terdapat banyak data yang memiliki nilai serupa. Misalnya adalah terdapat 10 negara yang hanya memiliki 1 data kopi, apabila data tersebut diubah menjadi 'other' maka akan terdapat 10 data yang memiliki nilai serupa.

Selanjutnya menghapus data altitude atau ketinggian dataran perkebunan tempat menanam kopi yang terlalu tinggi ataupun terlalu rendah. Melakukan perbaikan tipe data pada variabel dalam dataset. Kemudian

yang terakhir adalah membuat variabel dependen grading biji kopi berdasarkan data defects, quakers, dan Total Cupping Point. Defects merupakan tingkat kecacatan kopi, dan quakers merupakan kopi yang secara abnormal tidak berubah warna menjadi kecokelatan saat dipanggang. Keduanya digunakan untuk menilai kopi yang berkualitas rendah. Sedangkan Total Cupping Point yang tinggi digunakan untuk menandakan kopi berkualitas tinggi. Cupping Grade merupakan variabel dependen yang digunakan pada penelitian ini dengan empat tingkat secara berurutan dari yang tertinggi hingga terendah antara lain Specialty, Premium, Exchange, dan Below Standard. Setelah data diproses dengan maka akan menghasilkan seperti pada Gambar 10 Dapat dilihat pada tipe data sudah menjadi klasifikasi. Namun pada gambar tersebut dapat dilihat bahwa terdapat ketidak-seimbangan data. Hal ini terlihat dari data Premium Quality yang terlalu banyak dibandingkan data lainnya. Untuk mengatasi ketidak-seimbangan data seperti ini, maka dapat dilakukan Random Over Sampling.

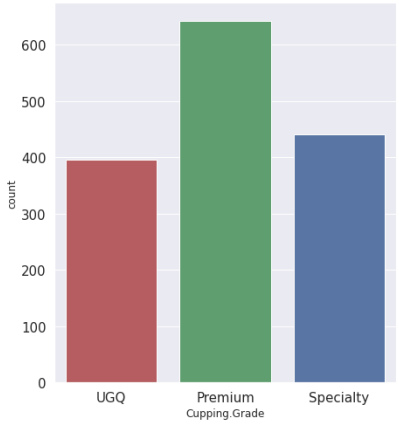


Gambar 2. Hasil pengamatan variabel dependen yang tidak seimbang

Oversampling merupakan metode pembangkitan data minoritas sebanyak data mayoritas. Metode ini merupakan metode yang populer diterapkan dalam rangka menangani ketidak seimbangan kelas. Teknik ini mensintesis sampel baru dari kelas minoritas untuk menyeimbangkan dataset dengan cara membuat instance baru dari minority class dengan pembentukan convex kombinasi dari instances yang saling berdekatan. Random Oversampling melibatkan duplikasi secara acak contoh dari kelas minoritas dan menambahkannya ke dataset pelatihan. Contoh dari dataset pelatihan dipilih secara acak dengan penggantian.

Ini berarti bahwa sampel data dari kelas minoritas dapat dipilih dan ditambahkan ke kumpulan data pelatihan baru yang “lebih seimbang” beberapa kali. Data tersebut dipilih dari kumpulan data pelatihan asli, ditambahkan ke kumpulan data pelatihan baru, dan kemudian dikembalikan atau "diganti" dalam kumpulan data asli, memungkinkan mereka untuk dipilih kembali.

Hasil pengamatan tersebut ditampilkan pada Gambar 3. Dapat dilihat bahwa jarak antara data kopi Premium dengan yang lainnya sudah tidak terlalu jauh. Dalam beberapa kasus, mencari distribusi yang seimbang untuk dataset yang sangat tidak seimbang dapat menyebabkan algoritma yang terpengaruh menjadi overfit pada kelas minoritas, yang menyebabkan peningkatan kesalahan generalisasi. Efeknya bisa berupa kinerja yang lebih baik pada Train Data, tetapi kinerja yang lebih buruk pada Test Data. Setelah Random Oversampling dilakukan maka langkah selanjutnya adalah mengamati kembali apakah Random Over Sampling tersebut bekerja dengan baik.



Gambar 3. Hasil pengamatan setelah data telah diseimbangkan

Jumlah data dan kualitas data dapat menentukan performa algoritma Random Forest dalam memprediksi kualitas kopi. Jumlah data yang digunakan penting karena akan menentukan performa hasil prediksi. Data yang sedikit, contohnya 50-100 data dapat menghasilkan performa yang kurang optimal sedangkan data yang cukup, contohnya 500-1000 data dapat menghasilkan performa yang optimal. Kemudian kualitas data juga perlu diperhatikan agar data dapat diproses dan dimengerti oleh algoritma Machine Learning.

Hasil prediksi yang tepat adalah hasil prediksi yang sesuai dengan hasil aktual. Misalnya apabila hasil aktual adalah 2, 0, dan 2 secara berurutan maka hasil prediksi yang tepat adalah 2, 0, dan 2. Dengan demikian hasil prediksi pada penelitian ini seperti yang ditampilkan pada Gambar 5. Area yang dilingkari menunjukkan hasil aktual yang sesuai dengan hasil prediksi. Hal tersebut berarti bahwa algoritma berhasil memprediksi kualitas kopi sesuai dengan kualitas kopi yang sebenarnya. Namun hal tersebut mungkin tidak berjalan dengan baik dan terdapat beberapa prediksi yang berebeda dari hasil yang sebenarnya.

```
# Menggabungkan data hasil prediksi yang benar
hasil_prediksi.loc[hasil_prediksi['Hasil Prediksi'] == hasil_prediksi['Hasil Aktual']]
```

	Hasil Prediksi	Hasil Aktual
0	2	2
2	0	0
3	2	2
4	1	1
5	1	1
...	...	...
438	1	1
439	0	0
440	2	2
442	0	0
443	1	1

349 rows × 2 columns

Gambar 4. Contoh prediksi yang benar

Gambar 5 menunjukkan hasil prediksi yang tidak akurat. Hasil aktual menunjukan nilai 0 dan 1 secara berurutan sedangkan hasil prediksi menunjukkan nilai 2. Kesalahan prediksi tersebut dapat berakibat buruk apabila terjadi dalam jumlah yang banyak karena dapat memberikan penilaian yang salah. Contohnya adalah apabila algoritma memprediksi kualitas kopi yang bagus dan ternyata kualitas sebenarnya adalah buruk, maka pelanggan akan kecewa dan memberikan efek negatif pada perusahaan. Untuk dapat mengevaluasi performa algoritma Random Forest yang digunakan pada penelitian ini secara keseluruhan, tentu saja akan memakan waktu yang lama apabila dilakukan secara manual.

Evaluasi secara otomatis dapat dilakukan dengan fungsi-fungsi dan rumus yang biasa dipakai untuk mengevaluasi performa Machine Learning dalam melakukan klasifikasi. Hal tersebut berguna untuk memberikan gambaran umum terhadap seberapa baik algoritma Machine Learning dalam memprediksi kualitas kopi secara tepat. Tentunya harapan terbaik dalam memprediksi kualitas kopi adalah akurasi mendekati sempurna atau dengan kata lain mendekati 100%, namun pada nyatanya akan butuh berbagai penelitian lebih lanjut untuk dapat mencapai nilai akurasi yang mendekati sempurna tersebut.

```
# Menggabungkan data hasil prediksi yang salah
hasil_prediksi.loc[hasil_prediksi['Hasil Prediksi'] != hasil_prediksi['Hasil Aktual']]
```

	Hasil Prediksi	Hasil Aktual
1	2	0
7	2	1
9	2	0
13	0	2
22	2	1
...	...	...
421	0	1
427	2	0
428	2	0
431	1	0
441	0	2

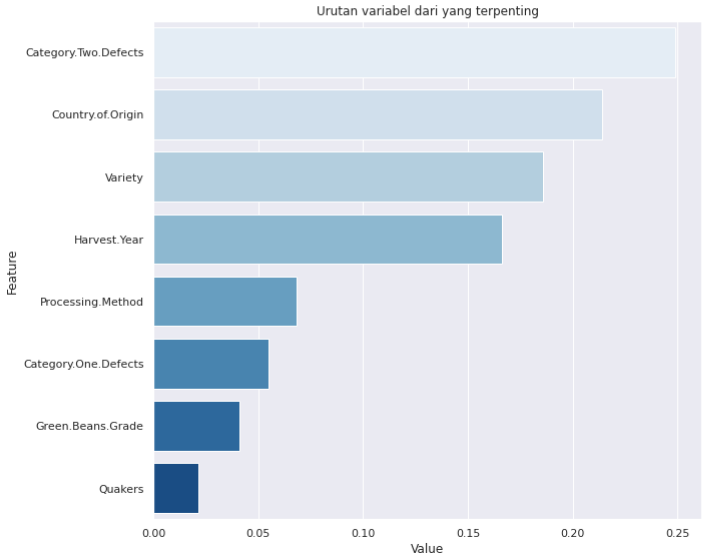
95 rows × 2 columns

Gambar 5. Contoh Prediksi Yang Salah

Kemudian selanjutnya adalah mengobservasi feature importance. Hal ini diperlukan untuk. Hasil observasi feature importance ditampilkan pada Gambar 5. Dapat dilihat bahwa variabel atau fitur yang paling berpengaruh pada prediksi kualitas kopi adalah 'Category.Two.Defects', 'Country.of.Origin', 'Variety', dan 'Harvest.Year'. Keempat data tersebut secara berurutan merupakan kecacatan biji kopi, negara asal, jenis varietas, dan tahun panen. Evaluasi hasil produksi sangat penting dilakukan untuk menentukan apakah kualitas itu baik atau buruk. Untuk 'Harvest.Year' agar kita mengetahui kapan kopi dipetik dari tangkainya dan semua ada perbandingannya masing-masing.

```
# Mengobservasi variabel penting prediksi kualitas kopi
feature_imp = pd.DataFrame(sorted(zip(rf.feature_importances_,X.columns)), columns=['Value','Feature'])

plt.figure(figsize=(10, 8))
sns.barplot(x="Value", y="Feature", data=feature_imp.sort_values(by="Value", ascending=False), palette='Blues')
plt.title('Urutan variabel dari yang terpenting')
plt.tight_layout()
plt.show()
```

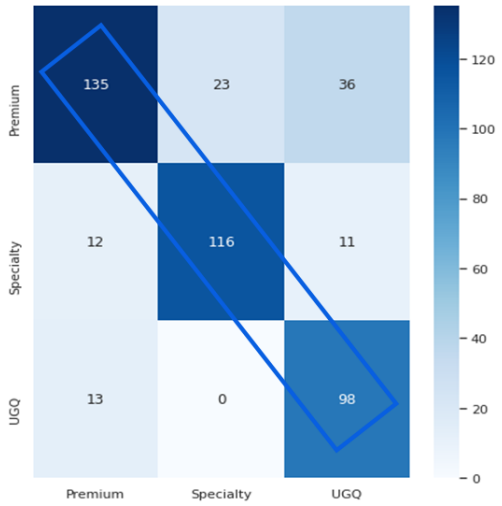


Gambar 6. Urutan variabel dari yang terpenting

PEMBAHASAN

Akurasi dapat dihitung dengan rumus akurasi seperti yang di bahas pada Bab 2, dengan memasukkan nilai hasil prediksi yang telah dihasilkan. Dikarenakan terdapat tiga kemungkinan output pada variabel dependen, yakni ketiga jenis kualitas kopi yang hendak diprediksi maka akan terdapat 9 kemungkinan hasil prediksi. Hasil prediksi pada penelitian ini terbagi menjadi dua jenis, yang benar dan yang salah. Gambar 7 menampilkan hasil jumlah hasil prediksi yang benar, antara lain:

1. 135 hasil aktual Premium yang diprediksi sebagai Premium
2. 116 hasil aktual Specialty yang diprediksi sebagai Specialty
3. 98 hasil aktual UGQ yang diprediksi sebagai UGQ

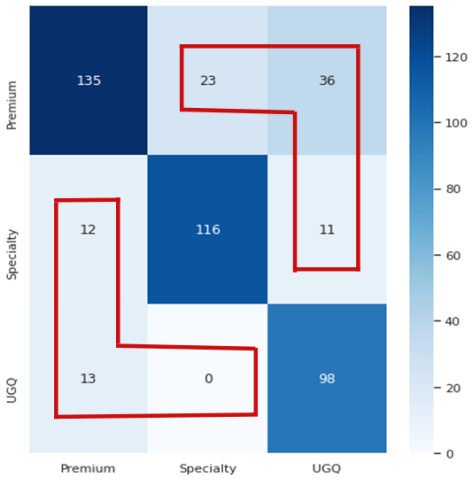


Gambar 7 Hasil Prediksi yang diprediksi Benar

Terlihat hasil prediksi akan tepat jika akurasi tingkat penelitian tersebut tinggi, sehingga jarang terjadi adanya kesalahan dissat penelitian.

Kemudian Gambar 8 menampilkan jumlah hasil prediksi yang salah, antara lain:

1. 23 hasil aktual Premium yang diprediksi sebagai Specialty
2. 36 hasil aktual Premium yang diprediksi sebagai UGQ
3. 12 hasil aktual Specialty yang diprediksi sebagai Premium
4. 11 hasil aktual Specialty yang diprediksi sebagai UGQ
5. 13 hasil aktual UGQ yang diprediksi sebagai Premium
6. Tidak ada hasil aktual UGQ yang diprediksi sebagai Specialty



Gambar 8. Hasil Prediksi yang diprediksi Salah

Dengan menggunakan hasil tersebut maka dapat dihitung jumlah akurasi pada penelitian ini, yakni sebagai berikut:

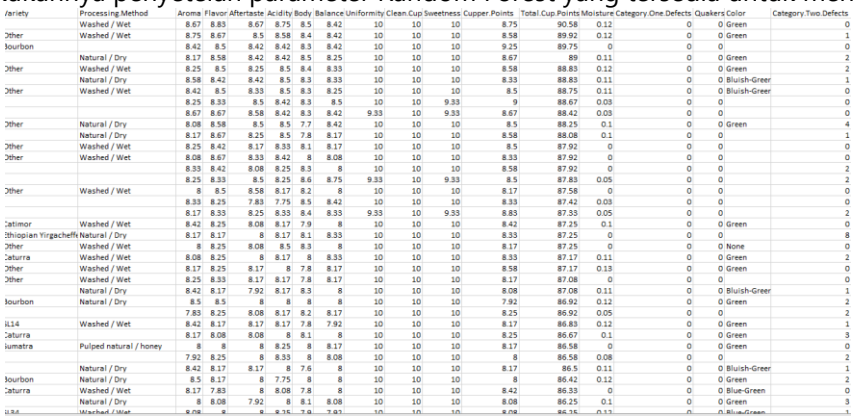
$$Akurasi = \frac{Jumlah\ data\ kualitas\ kopi\ yang\ diprediksi\ benar}{Jumlah\ data\ kopi\ secara\ keseluruhan} = \frac{349}{444} = 79\% \qquad (1)$$

Berdasarkan penelitian prediksi kualitas kopi dengan Random Forest yang dilakukan maka dapat disimpulkan hal-hal yang mendukung keberhasilan prediksi kualitas kopi pada penelitian ini antara lain:

1. Proses preprocessing yang berhasil mengeliminasi data kosong dan tipe data yang tidak sesuai sehingga algoritma Random Forest dapat memproses data spesifikasi kopi
2. Jumlah data yang cukup untuk memastikan algoritma random forest dapat memprediksi dengan cukup baik

Kemudian di lain sisi, akurasi yang mencapai 79% masih memiliki ruang untuk perkembangan sehingga bisa mendekati 100%. Kurang optimalnya nilai akurasi pada penelitian ini dapat diakibatkan oleh:

1. Kurangnya variabel independen yang digunakan. Penelitian ini hanya menggunakan 8 variabel dari 43 variabel yang tersedia. Sehingga masih terdapat variabel yang berpotensi dapat meningkatkan akurasi.
2. Tidak dilakukannya penyetelan parameter Random Forest yang tersedia untuk meningkatkan akurasi.



/ariety	Processing.Method	Aroma	Flavor	Aftertaste	Acidity	Body	Balance	Uniformity	Clean.Cup	Sweetness	Cupper.Points	Total.Cup.Points	Moisture	Category.One.Defects	Quakers	Color	Category.Two.Defects
	Washed / Wet	8.67	8.83	8.67	8.75	8.5	8.42	10	10	10	8.75	90.58	0.12	0	0	Green	0
)ther	Washed / Wet	8.75	8.67	8.5	8.58	8.4	8.42	10	10	10	8.58	89.92	0.12	0	0	Green	1
lourbon		8.42	8.5	8.42	8.42	8.3	8.42	10	10	10	9.25	89.75	0	0	0		0
	Natural / Dry	8.17	8.58	8.42	8.42	8.5	8.25	10	10	10	8.67	89	0.11	0	0	Green	2
)ther	Washed / Wet	8.25	8.5	8.25	8.5	8.4	8.33	10	10	10	8.58	88.83	0.12	0	0	Green	2
	Natural / Dry	8.58	8.42	8.42	8.5	8.3	8.33	10	10	10	8.33	88.83	0.11	0	0	Bluish-Greer	1
)ther	Washed / Wet	8.42	8.5	8.33	8.5	8.3	8.25	10	10	10	8.5	88.75	0.11	0	0	Bluish-Greer	0
		8.25	8.33	8.5	8.42	8.3	8.5	10	10	9.33	9	88.67	0.03	0	0		0
		8.67	8.67	8.58	8.42	8.3	8.42	9.33	10	9.33	8.67	88.42	0.03	0	0		0
)ther	Natural / Dry	8.08	8.58	8.5	8.5	7.7	8.42	10	10	10	8.5	88.25	0.11	0	0	Green	4
	Natural / Dry	8.17	8.67	8.25	8.5	7.8	8.17	10	10	10	8.58	88.08	0.1	0	0		1
)ther	Washed / Wet	8.25	8.42	8.17	8.33	8.1	8.17	10	10	10	8.5	87.92	0	0	0		0
)ther	Washed / Wet	8.08	8.67	8.33	8.42	8	8.08	10	10	10	8.33	87.92	0	0	0		0
		8.33	8.42	8.08	8.25	8.3	8	10	10	10	8.58	87.92	0	0	0		2
		8.25	8.33	8.5	8.25	8.6	8.75	9.33	10	9.33	8.5	87.83	0.05	0	0		2
)ther	Washed / Wet	8	8.5	8.58	8.17	8.2	8	10	10	10	8.17	87.58	0	0	0		0
		8.33	8.25	7.83	7.75	8.5	8.42	10	10	10	8.33	87.42	0.03	0	0		0
		8.17	8.33	8.25	8.33	8.4	8.33	9.33	10	9.33	8.83	87.33	0.05	0	0		2
Catimor	Washed / Wet	8.42	8.25	8.08	8.17	7.9	8	10	10	10	8.42	87.25	0.1	0	0	Green	0
Ethiopian Yirgacheffe	Natural / Dry	8.17	8.17	8	8.17	8.1	8.33	10	10	10	8.33	87.25	0	0	0		8
)ther	Washed / Wet	8	8.25	8.08	8.5	8.3	8	10	10	10	8.17	87.25	0	0	0	None	0
Caturra	Washed / Wet	8.08	8.25	8	8.17	8	8.33	10	10	10	8.33	87.17	0.11	0	0	Green	2
)ther	Washed / Wet	8.17	8.25	8.17	8	7.8	8.17	10	10	10	8.58	87.17	0.13	0	0	Green	0
)ther	Washed / Wet	8.25	8.33	8.17	8.17	7.8	8.17	10	10	10	8.17	87.08	0	0	0		0
	Natural / Dry	8.42	8.17	7.92	8.17	8.3	8	10	10	10	8.08	87.08	0.11	0	0	Bluish-Greer	1
lourbon	Natural / Dry	8.5	8.5	8	8	8	8	10	10	10	7.92	86.92	0.12	0	0	Green	2
		7.83	8.25	8.08	8.17	8.2	8.17	10	10	10	8.25	86.92	0.05	0	0		2
ıL14	Washed / Wet	8.42	8.17	8.17	8.17	7.8	7.92	10	10	10	8.17	86.83	0.12	0	0	Green	1
Caturra		8.17	8.08	8.08	8	8.1	8	10	10	10	8.25	86.67	0.1	0	0	Green	3
Sumatra	Pulped natural / honey	8	8	8	8.25	8	8.17	10	10	10	8.17	86.58	0	0	0	Green	0
		7.92	8.25	8	8.33	8	8.08	10	10	10	8	86.58	0.08	0	0		2
	Natural / Dry	8.42	8.17	8.17	8	7.6	8	10	10	10	8.17	86.5	0.11	0	0	Bluish-Greer	1
Bourbon	Natural / Dry	8.5	8.17	8	7.75	8	8	10	10	10	8	86.42	0.12	0	0	Green	2
Caturra	Washed / Wet	8.17	7.83	8	8.08	7.8	8	10	10	10	8.42	86.33	0	0	0	Blue-Green	0
	Natural / Dry	8	8.08	7.92	8	8.1	8.08	10	10	10	8.08	86.25	0.1	0	0	Green	3

Gambar 9. Data hasil kopi yang telah diteliti

Sesuai dengan gambar 9, merupakan gambar hasil dari data yang telah diteliti. Yang diteliti ada varieti, metode pemrosesan penelitian, aroma, rasa, tingkat keasaman, kecacatan, dan masih ada banyak lainnya. Dengan demikian hasil penelitian semakin banyak data nya semakin bukti kuat hasil penelitiannya. Serta akurasi jga membantu agar semakin tinggi persentase kebenarannya dan terhindar dari kesalahan.

KESIMPULAN

Berdasarkan hasil berikut beberapa hal yang dapat disimpukan:

1. Penelitian ini berhasil meneliti 1.339 data spesifikasi kopi dan merancang algoritma Random Forest untuk memprediksi kualitas kopi
2. Akurasi algoritma Random Forest dalam memprediksi kualitas kopi pada penelitian ini adalah sebesar 79%

REFERENCES

[1] "Kopi Berdasarkan Penentuan Lama Sangrai Variasi Derajat Sangrai Menggunakan Model Warna Rgb Pada Pengolahan Citra Digital (Digital Image ProcessinG)." https://eprints.umm.ac.id/79727/2/BAB 1.pdf (accessed Feb. 09, 2022).

[2] A. Pradana, "'Prediksi Kualitas Biji Kopi Menggunakan Metode Segmentasi Dan Content Based Image Retrieval(CBIR),'" Sep. 2021.

[3] Petrus Katemba and R. Koro, "Prediksi Tingkat Produksi Kopi Menggunakan Regresi Linear," J. Ilm. Flash, vol. 1, no. 3, pp. 42–51, 2017.

[4] Y. Christian and I. D. M. B. A. Darmawan, "Specialty Coffee Cupping Score Prediction with General Regression Neural Network (GRNN)," JELIKU (Jurnal Elektron. Ilmu Komput. Udayana), vol. 9, no. 2, p. 185, 2020, doi: 10.24843/jlk.2020.v09.i02.p04.

[5] A. R. Ginanjar, "Sistem Deteksi Jenis Cacat Biji Kopi dengan Algoritma K-Nearest Neighbor," 2019.

[6] I. Ilhamsyah, A. Y. Rahman, and I. Istiadi, "Klasifikasi Kualitas Biji Kopi Menggunakan MultilayerPerceptron Berbasis Fitur Warna LCH," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 5, no. 6, pp. 1008–1017, 2021, doi: 10.29207/resti.v5i6.3438.

[7] A. M. Majid, U. Khairat, and A. Qaslim, "Identifikasi Kualitas fisik Pada Biji Kopi Menggunakan Tekonologi Pengolahan Citra Dengan Metode Neural Network," J. Peqguruan Conf. Ser., vol. 4, no. 1, 2022.

[8] S. W. Sidehabi, N. Jabir, and M. Ilyas, "Alat Penyortir Biji Kopi otomatis dengan menggunakan Mikrokotroller Arduino Nano pada Usaha Mikro Kecil dan Menengah (UMKM) Toraja Kawaa Roasters," vol. 1, no. 1, pp. 37–43, 2022.

[9] F. Ruzaik, "Prediksi Kualitas Biji Kopi Menggunakan Metode Segmentasi Dan Content Based Image Retrieval(CBIR)," pp. 61–64, 2022.