

KLASIFIKASI VARIETAS ANGGUR BERDASARKAN CITRA DAUN MENGGUNAKAN VGG16-XGBOOST

Imeldha Elzandy, Eva Yulia Puspaningrum, Afina Lina Nurlaili

Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur
Jalan Raya Rungkut Madya, Gunung Anyar, Surabaya, Jawa Timur, Indonesia
evapuspaningrum.if@upnjatim.ac.id

ABSTRAK

Tanaman anggur memiliki varietas beragam dengan karakteristik daun yang cukup mirip antar varietas, hal ini menjadi tantangan tersendiri dalam proses pengenalan varietas secara manual, terutama bagi masyarakat awam. Penelitian ini bertujuan untuk melakukan klasifikasi varietas anggur berdasarkan citra daun menggunakan kombinasi metode CNN dengan arsitektur VGG-16 sebagai ekstraksi fitur dan XGBoost sebagai klasifikasi. Adapun empat varietas anggur yang digunakan pada penelitian ini, yaitu Jumer, Ninel, Ruby88, dan Transfigurasi. Dataset yang digunakan terdiri dari 560 citra untuk pelatihan dan 60 citra untuk pengujian, yang kemudian dilakukan proses ekstraksi fitur menggunakan VGG-16 dan diklasifikasikan menggunakan XGBoost. Model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score*. Hasil pada penelitian menunjukkan bahwa kombinasi metode VGG-16 dan XGBoost mampu menghasilkan *accuracy* klasifikasi dengan hasil 89.58%, *precision* = 89.64%, nilai *recall* = 89.58%, nilai *f1-Score* = 89.55% dengan estimasi waktu *training* selama 832.55 detik.

Kata kunci : daun anggur, klasifikasi, varietas anggur, vgg-16, xgboost.

1. PENDAHULUAN

Varietas pada tanaman anggur memiliki jumlah yang cukup beragam dan seringkali memiliki kemiripan dari bentuk daun maupun buahnya. Daun pada tanaman anggur dapat digunakan sebagai sumber informasi untuk melakukan pengenalan varietas anggur. Masing-masing varietas memiliki ciri khas atau karakteristik tersendiri, seperti dari segi bentuk, ukuran, warna, tekstur, urat daun, serta bentuk tepi daun yang mirip untuk setiap varietasnya [1]. Hal ini menjadi tantangan dalam proses pengenalan varietas anggur, terutama bagi masyarakat awam yang ingin melakukan budidaya tanaman anggur. Sehingga, diperlukan adanya pengenalan yang lebih akurat dan efisien untuk membantu masyarakat dalam menentukan varietas anggur yang sesuai dengan kebutuhan budidaya mereka, yaitu dengan melakukan klasifikasi pada citra daun anggur.

Citra daun anggur dapat digunakan untuk mengklasifikasikan berbagai varietas yang dimiliki oleh tanaman anggur. Klasifikasi menjadi aspek penting dalam pengenalan varietas karena setiap varietas memiliki karakteristik yang khas. Dengan adanya klasifikasi yang akurat, petani dan pelaku industri hortikultura dapat lebih mudah mengenali varietas yang tepat untuk dibudidayakan, sehingga dapat dengan mudah memahami karakteristik masing-masing varietas. Salah satu kombinasi algoritma yang dapat digunakan untuk melakukan klasifikasi varietas daun anggur adalah *transfer learning* dengan menggunakan arsitektur *Visual Geometry Group* atau VGG-16 dan *Extreme Gradient Boosting* atau XGBoost. Dimana VGG-16 akan digunakan sebagai ekstraksi fitur dan XGBoost akan digunakan sebagai klasifikasi.

Transfer learning termasuk ke dalam algoritma *deep learning* yang dikenal untuk melakukan *object detection* dan *object recognition* pada sebuah citra. Metode ini menggunakan masukan dua dimensi berupa citra atau gambar, sehingga model klasifikasi yang digunakan untuk menentukan pola pada suatu gambar akan digunakan untuk belajar. *Transfer learning* akan melakukan ekstraksi pada fitur penting yang terdapat pada citra daun serta mempelajari polanya untuk membuat suatu gambar memiliki label sesuai dengan varietas daun tersebut. Setelah citra daun diekstraksi, fitur-fitur tersebut akan menjadi masukan untuk algoritma XGBoost.

XGBoost atau extreme gradient boosting merupakan algoritma machine learning yang berbasis pada pohon keputusan. Algoritma ini dapat digunakan untuk melakukan klasifikasi maupun prediksi. Selain itu, dapat juga digunakan untuk meningkatkan efisiensi model dengan memanfaatkan teknik boosting. Algoritma ini bekerja dengan menambahkan pohon keputusan secara iteratif, dimana untuk setiap pohon baru akan memperbaiki kesalahan prediksi yang terdapat pada pohon sebelumnya, setelahnya model akan menghasilkan output berupa kelas-kelas dari setiap pohon [2].

Terdapat beberapa penelitian yang telah dilakukan sebelumnya dengan menggunakan bantuan teknologi komputer untuk melakukan klasifikasi. Penelitian sebelumnya dilakukan dengan menggunakan metode CNN, dataset yang digunakan oleh peneliti, yaitu berupa 150 data citra daun dengan tiga varietas anggur. Berdasarkan penelitian tersebut, hasil pengujiannya mendapatkan akurasi data latih 86% dan hasil akurasi pada data validasi mencapai 91%. Kelebihan pada penelitian tersebut adalah penggunaan arsitektur ResNet-50 dalam

meningkatkan akurasi. Namun, penelitian ini memiliki kekurangan berupa jumlah dataset yang terbatas, variasi kelas yang sedikit, dan evaluasi model yang hanya menggunakan metrik akurasi [3].

Penelitian lain juga dilakukan dengan menggunakan gabungan metode CNN dan SVM, jumlah dataset yang digunakan adalah 500 data citra daun dengan lima varietas anggur. Penelitian tersebut menggunakan metode CNN yang digunakan untuk melakukan ekstraksi dan SVM untuk klasifikasi, didapatkan hasil akurasi gabungan kedua metode tersebut adalah 97%. Kelebihan dari penelitian ini adalah kombinasi metode yang mampu meningkatkan akurasi klasifikasi. Namun, kekurangannya adalah dataset yang digunakan masih terbatas pada lima varietas local dari Turki, sehingga kemampuan model terhadap varietas lain belum dapat dipastikan [4].

Berdasarkan hal yang telah dijelaskan pada latar belakang, maka penelitian ini bertujuan untuk melakukan analisis terhadap pengaruh penggunaan kolaborasi metode VGG-16 dengan metode XGBoost terhadap akurasi klasifikasi varietas anggur berdasarkan citra daun. Berdasarkan kombinasi kedua metode tersebut, penelitian ini akan melakukan pengujian untuk meningkatkan akurasi pada proses klasifikasi dengan menggunakan objek citra daun anggur. Penelitian ini menggunakan empat varietas daun anggur yang akan di klasifikasikan, yaitu Jumer, Ninel, Ruby88, dan Transfigurasi. Keempat varietas tersebut dipilih karena memiliki struktur daun dan buah yang cenderung mirip. Dengan demikian penelitian ini diharapkan dapat memberikan solusi yang efektif dan efisien dalam melakukan klasifikasi pada berbagai varietas anggur berdasarkan daunnya, sehingga dapat dengan tepat dirawat sesuai dengan kebutuhan varietas tersebut.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Adapun penelitian terdahulu yang digunakan sebagai acuan pembelajaran oleh penulis, yaitu literasi yang berjudul Deep Learning in Grapevine Leaves Varieties Classification Based on Dense Convolutional Network untuk mengembangkan model klasifikasi daun anggur menggunakan arsitektur DenseNet201 untuk meningkatkan akurasi dalam identifikasi varietas. Dataset terdiri dari 500 citra daun dari 5 varietas, yang diperluas menjadi 2.800 gambar melalui augmentasi guna mengurangi *overfitting*. Hasil menunjukkan bahwa *fine-tuning* pada lapisan ke-15 dan ke-30 DenseNet201 memberikan performa terbaik dengan akurasi mencapai 98%. Studi ini membuktikan bahwa DenseNet201 dengan strategi *fine-tuning* efektif dalam klasifikasi daun anggur secara akurat dan efisien [1].

Penelitian yang berjudul Convolutional Neural Network Pada Identifikasi Varian Tanaman Anggur Menggunakan Resnet-50 untuk mengembangkan sistem identifikasi varietas anggur menggunakan arsitektur ResNet-50. Dataset terdiri dari 150 citra

daun dari tiga varietas, yaitu Nizina, Jupiter, dan Isabella. Proses *preprocessing* mencakup resize, penghapusan latar belakang, dan augmentasi data untuk meningkatkan variasi. Model dibangun dengan pendekatan *transfer learning* menggunakan bobot *pre trained*. Hasil menunjukkan akurasi 86% pada data *training* dan 91% pada data validasi [3].

Penelitian yang berjudul A CNN-SVM Study Based On Selected Deep Features For Grapevine Leaves Classification untuk mengembangkan sistem klasifikasi varietas daun anggur dengan pendekatan *hybrid* menggunakan CNN dan SVM. Dataset terdiri dari 500 citra daun dari 5 varietas, yang diperluas menjadi 2.500 gambar melalui augmentasi. CNN digunakan untuk ekstraksi fitur, kemudian diklasifikasikan menggunakan SVM. Hasil menunjukkan bahwa kombinasi CNN-SVM menghasilkan akurasi sebesar 97,60% [4].

Penelitian yang berjudul Klasifikasi Penyakit Daun Anggur dengan Menggunakan Convolutional Neural Network dan Transfer Learning dari VGG16 untuk mengembangkan sistem klasifikasi penyakit daun anggur menggunakan arsitektur VGG16 untuk membantu deteksi dini dan mengurangi kesalahan identifikasi manual. Dataset terdiri dari 4.062 gambar daun dengan tiga kelas penyakit : Busuk Hitam, Campak Hitam, dan Hawar Daun. Proses awal mencakup segmentasi dan *thresholding*. Model mencapai akurasi 98% pada data *training* dan 92% pada validasi. Hasil menunjukkan bahwa VGG16 efektif dalam meningkatkan akurasi klasifikasi penyakit daun anggur serta dapat menjadi solusi praktis bagi petani [5].

2.2. Varietas Anggur

Tanaman anggur memiliki berbagai macam varietas, dimana setiap varietas memiliki perbedaan pada bentuk, ukuran, warna, tekstur daun, rasa, serta kandungan nutrisinya. Pengenalan varietas anggur menjadi langkah penting dalam menentukan bibit yang ditanam, karena perbedaan varietas dapat memengaruhi metode perawatan, produktivitas, dan ketahanan terhadap lingkungan. Biasanya, varietas anggur baru dapat dikenali setelah tanaman mulai berbuah, dan membutuhkan waktu 8 bulan bahkan lebih sejak masa tanam. Oleh karena itu, diperlukan proses klasifikasi untuk mengidentifikasi dan membedakan varietas anggur secara tepat. Dalam penelitian ini, terdapat empat jenis varietas anggur yang digunakan untuk proses klasifikasi, yaitu Jumer, Ninel, Ruby88, dan Transfigurasi dengan masing-masing varietas memiliki citra daun yang berbeda.

a. Varietas Jumer



Gambar 1. Varietas Jumer

Buah anggur varietas Jumer memiliki buah berwarna hijau apabila telah masak pohon. Bentuk buahnya cenderung bulat dan sedikit lonjong, serta memiliki daging buah yang bening dan berair. Rasa dari anggur Jumer cenderung manis dengan sedikit rasa asam yang khas. Varietas Jumer dapat dikonsumsi secara langsung sebagai buah segar, namun biasanya juga digunakan untuk membuat olahan minuman anggur. Bentuk citra daun pada varietas Jumer seperti pada gambar 1 memiliki ukuran daun yang relatif sedang dan bentuk daunnya cenderung menjari dengan ujung yang simetris. Selain itu, pada bagian sisi kanan dan kiri daun memiliki tepi yang bergerigi relative teratur dengan ujung pada setiap sisinya berbentuk oval. Daun ini memiliki warna hijau tua dengan urat daun yang terlihat cukup jelas [6].

b. Varietas Ninel



Gambar 2. Varietas Ninel

Buah anggur varietas Ninel memiliki buah berwarna merah tua hingga ungu gelap ketika telah matang sempurna. Bentuk buahnya bulat sedikit lonjong, serta memiliki daging buah yang berair dan lembut, dengan sedikit biji, sehingga menjadikannya lebih cocok untuk dikonsumsi secara langsung. Rasa dari anggur Ninel ini cenderung manis, namun terdapat sedikit rasa asam. Varietas buah ini sering dikonsumsi langsung sebagai buah segar, maupun produk olahan anggur lainnya. Bentuk citra daun pada varietas Ninel seperti pada gambar 2 memiliki ukuran daun yang relatif sedang, bentuk daunnya cenderung menjari menjadi 3 hingga 5 cabang dengan ujung daun yang runcing mengarah ke kanan atau kiri. Selain itu, pada sisi kanan dan kiri daun memiliki tepi yang bergerigi abstrak dan runcing diseluruh permukaannya. Daun ini memiliki warna hijau tua dengan urat daun yang terlihat cukup jelas [7].

c. Varietas Ruby88



Gambar 3. Varietas Ruby88

Buah anggur varietas Ruby88 memiliki buah berwarna hijau saat matang sempurna. Bentuk buahnya cenderung bulat ketika masak dan lonjong saat masa pertumbuhan, serta memiliki daging buah yang berair, lembut, dan sedikit biji, sehingga menjadikannya lebih nyaman untuk dikonsumsi secara langsung tanpa perlu mengkhawatirkan bijinya. Rasa dari anggur Ruby88 ini cenderung manis dan sedikit

asam. Buah anggur Ruby88 dapat secara langsung dikonsumsi sebagai buah segar, maupun produk olahan lainnya. Bentuk citra daun pada varietas Ruby88 seperti pada gambar 3 memiliki ukuran daun yang relatif sedang, bentuk daunnya cenderung menjari dengan 3 hingga 5 cabang dengan ujung daun agak meruncing dan simetris. Selain itu, pada bagian tepi daun memiliki bentuk bergerigi abstrak. Daun ini memiliki warna hijau tua dengan urat daun yang terlihat cukup jelas [6].

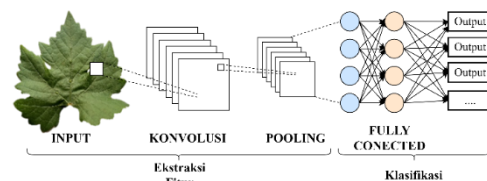
d. Varietas Transfigurasi



Gambar 4. Varietas Transfigurasi

Buah anggur varietas Transfigurasi memiliki buah berwarna hijau kemerahan ketika telah matang sempurna. Buah anggur Transfigurasi berbentuk lonjong, serta memiliki daging buah yang cenderung padat dan berair, dengan kondisi buah yang memiliki sedikit biji, sehingga menjadikannya lebih cocok untuk dikonsumsi secara langsung tanpa perlu mengkhawatirkan bijinya. Rasa dari anggur Transfigurasi ini cenderung manis. Buah ini sering dikonsumsi secara langsung sebagai buah segar karena cita rasanya. Bentuk citra daun pada varietas Transfigurasi seperti pada gambar 4 memiliki ukuran daun relatif sedang, bentuk daunnya cenderung menjari menjadi 5 cabang dengan ujung yang runcing menyerong ke arah kanan atau kiri. Selain itu, pada sisi kanan dan kiri daun memiliki tepi yang bergerigi teratur dengan ujung cabang daun yang relatif tumpul atau oval diseluruh permukaannya. Daun ini memiliki warna hijau tua dengan urat daun yang terlihat cukup jelas, serta memiliki batang berwarna kemerahan [7].

2.3. Convolutional Neural Network



Gambar 5. Lapisan CNN

Convolutional Neural Network atau CNN merupakan algoritma *deep learning* yang meniru cara kerja otak manusia dalam mengenali informasi visual pada objek citra. Seperti otak yang terdiri dari banyak sel, CNN terdiri dari banyak *neuron* yang saling terhubung dan melakukan komputasi sederhana untuk mengambil keputusan. CNN tidak hanya memproses citra, tetapi juga mengenali pola visual melalui operasi konvolusi, sehingga dapat digunakan untuk deteksi, klasifikasi, segmentasi, dan tugas visual lainnya [8]. Algoritma CNN bekerja dengan melakukan operasi

konvolusi menggunakan kernel yang digerakkan melintasi setiap bagian citra. Proses ini memungkinkan komputer mengenali fitur penting seperti warna, garis, dan tekstur dari piksel citra. Operasi ini dilakukan berulang kali untuk membentuk representasi fitur yang semakin kompleks [9]. Seperti pada gambar 5 yang menjelaskan mengenai lapisan-lapisan pada CNN, terdapat dua tahapan utama, yaitu ekstraksi fitur dan klasifikasi. Pada tahap pertama, citra diproses melalui lapisan konvolusi dan pooling untuk menyederhanakan dimensi citra sambil mempertahankan informasi penting. Kemudian, fitur yang diperoleh diklasifikasikan melalui fully connected layer, di mana seluruh neuron saling terhubung untuk menghasilkan output akhir berupa prediksi kelas.

2.4. Arsitektur VGG-16

Convolutional Neural Network secara sederhana memiliki 3 lapisan utama, yaitu *Input Layer*, *Hidden Layer*, dan *Output Layer*. VGG atau *Visual Geometry Group* merupakan salah satu arsitektur yang dimiliki oleh algoritma CNN dan dianggap sebagai salah satu arsitektur model terbaik pada *computer vision*. Arsitektur ini diusulkan oleh Karen Simonyan dan Andrew Zisserman dari Universitas Oxford, arsitektur ini dikembangkan oleh *Visual Engineering Group*. Arsitektur VGG memiliki dua jenis model berdasarkan jumlah lapisan yang dimiliki, yaitu VGG-16 dan VGG-19. Pada jenis model arsitektur VGG-16 memiliki struktur lapisan yang terbagi menjadi 13 untuk *convolutional layer* dan 3 untuk *fully connected layer*, hal ini yang membuat model menjadi kompleks [10].

Arsitektur VGG-16 biasanya digunakan untuk melakukan klasifikasi pada suatu objek, seperti objek gambar. Kelebihan yang dimiliki oleh arsitektur ini adalah memiliki kemampuan untuk menghasilkan fitur-fitur yang kuat dan sesuai karena memiliki jumlah parameter yang cukup banyak. Sebanyak 138 juta parameter dimiliki oleh VGG-16, sehingga dapat menghasilkan *output* yang sangat baik saat melakukan pengenalan gambar, namun karena jumlah parameter yang cukup banyak membuat VGG-16 memerlukan daya komputasi yang tinggi saat akan melakukan pelatihan [11].

2.5. Extreme Gradient Boosting

XGBoost atau *eXtreme Gradient Boosting* adalah model machine learning yang dikembangkan dari metode GBM atau *Gradient Boosting Machine*, metode ini dirancang untuk melakukan klasifikasi dan regresi. *Boosting* merupakan metode pembelajaran yang bekerja dengan cara menggabungkan beberapa model lemah atau *weak learners* menjadi suatu model yang kuat atau *strong learners*, hal tersebut dapat digunakan untuk meningkatkan kinerja model dalam melakukan klasifikasi. Teknik *boosting* bekerja dengan cara menggabungkan hasil klasifikasi dari seluruh model yang lemah menggunakan metode

voting untuk menghasilkan model akhir yang lebih kuat dan akurat [12].

Model XGBoost menggunakan teknik *boosting* yang bekerja dengan membangun pohon keputusan secara iteratif atau berulang. Pada setiap iterasi, nilai prediksi ditentukan oleh kombinasi hasil dari beberapa pohon keputusan yang dibangun. Setiap pohon baru yang dibangun memiliki tugas untuk memperbaiki kesalahan prediksi yang terdapat pada pohon sebelumnya, sehingga model prediksi akan terus berfokus untuk memperbaiki *residual* atau kesalahan prediksi yang dilakukan pada model sebelumnya, teknik ini biasanya disebut dengan *gradient boosting*. Residual akan dihitung sebagai selisih antara nilai aktual dengan prediksi awal. Berikut adalah persamaan matematis untuk menghitung *residual* atau nilai kesalahan dari prediksi :

$$\hat{y}_i = y_i - f_0(x) \quad (1)$$

Keterangan :

$\hat{y}_i$ = *residual* atau nilai kesalahan

y_i = nilai atau variabel prediksi sebenarnya

$f_0(x)$ = nilai prediksi model pada iterasi sebelumnya

Adapun fungsi objektif yang digunakan model XGBoost salah satunya, yaitu *multi:softprob*. Fungsi ini dapat digunakan untuk melakukan perhitungan *loss* untuk mengatasi masalah klasifikasi multi kelas dengan menghasilkan probabilitas di setiap kelasnya. Untuk menghitung probabilitas dari setiap kelas digunakan fungsi aktivasi *Softmax*, berikut merupakan persamaan matematisnya :

$$s(o_i) = \frac{e^{o_i}}{\sum_{j=1}^K e^{o_j}} \quad (2)$$

Keterangan :

$s(o_i)$ = probabilitas kelas i

e^{o_i} = nilai eksponen kelas i

K = jumlah kelas

$\sum_{j=1}^K e^{o_j}$ = jumlah nilai eksponen seluruh kelas

2.6. Parameter

Model XGBoost memiliki sejumlah hyperparameter yang berperan penting dalam meningkatkan kinerjanya, seperti *learning_rate*, *max_depth*, *n_estimators*, dan *subsample*. *learning_rate* mengatur kecepatan model dalam belajar, *max_depth* menentukan kedalaman maksimum pohon keputusan, *n_estimators* menunjukkan jumlah pohon yang dibangun, dan *subsample* mengatur proporsi data yang digunakan untuk setiap pohon. Kombinasi parameter ini bertujuan untuk menjaga keseimbangan model agar menghasilkan performa yang optimal serta mencegah *overfitting* maupun *underfitting* [13].

2.7. Evaluasi Model

Evaluasi model menjadi salah satu proses yang cukup penting dalam hal mengolah gambar maupun mengenali suatu pola pada gambar. Tujuan dari evaluasi model adalah untuk mengukur kinerja dari

model yang digunakan saat melakukan eksekusi pada data yang tersedia, sehingga dapat memberikan gambaran yang jelas mengenai seberapa baik dan akurat model tersebut melakukan prediksi. Hasil gambar yang diolah akan diproses pada bagian ini menggunakan berbagai matriks sebagai parameter. Model CNN dengan arsitektur VGG-16 dan XGBoost yang digunakan akan dievaluasi menggunakan beberapa parameter, yaitu *Confusion Matrix*, *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. *Accuracy* merupakan metrik evaluasi yang mengukur kinerja model berdasarkan rasio prediksi yang benar terhadap seluruh prediksi. Secara matematis, akurasi dirumuskan sebagai :

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (3)$$

Precision merupakan metrik evaluasi yang mengukur sejauh mana model benar dalam memprediksi kelas positif, dengan menghitung rasio antara prediksi positif yang benar terhadap seluruh prediksi positif. Secara matematis, presisi dirumuskan sebagai :

$$Precision = \frac{TP}{(TP+FP)} \quad (4)$$

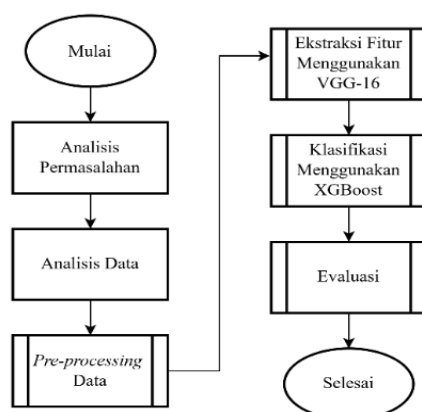
Recall merupakan metrik evaluasi yang mengukur kemampuan model dalam mengidentifikasi kelas positif secara benar, yaitu rasio antara prediksi positif yang benar dengan seluruh data aktual yang benar. Secara matematis, recall dirumuskan sebagai :

$$Recall = \frac{TP}{(TP+FN)} \quad (5)$$

F1-Score merupakan metrik evaluasi yang mengukur keseimbangan antara *precision* dan *recall*, terutama berguna ketika data tidak seimbang. Nilai ini dihitung sebagai rata-rata harmonik dari *precision* dan *recall*. Secara matematis, *F1-Score* dirumuskan sebagai :

$$F1\ Score = 2 \times \frac{Recall \times Presisi}{Recall + Presisi} \quad (6)$$

3. METODE PENELITIAN



Gambar 6. Tahapan Penelitian

Tahapan penelitian yang ada pada gambar 6 dimulai dengan melakukan proses analisis permasalahan yang terdapat pada varietas citra daun anggur. Setelah itu, melakukan proses analisis data untuk memahami karakteristik data yang akan digunakan. Selanjutnya, melakukan pra pemrosesan data untuk mempersiapkan data sehingga layak dan dapat digunakan pada proses selanjutnya. Setelah data melalui berbagai tahapan pada pra pemrosesan, data akan digunakan untuk membangun model, yaitu CNN dan XGBoost. Tahap berikutnya yaitu melakukan evaluasi untuk mengukur performa kedua model sehingga didapatkan model yang lebih unggul dalam memecahkan masalah tersebut. Penelitian diakhiri pada tahap selesai yang mana semua proses telah dilakukan dan mendapatkan hasil.

3.1. Analisis Permasalahan

Tanaman anggur memiliki banyak varietas dengan karakteristik unik seperti warna, bentuk, ukuran, dan rasa pada buah maupun daunnya. Kemiripan antar varietas sering menyulitkan proses identifikasi, terutama bagi masyarakat umum yang awam mengenai varietas tanaman anggur yang beragam. Proses pengenalan secara manual memerlukan waktu dan keahlian khusus, sehingga penerapan algoritma machine learning menjadi solusi potensial untuk membantu masyarakat mengenali varietas anggur secara mandiri dan mendukung praktik budidaya yang lebih tepat.

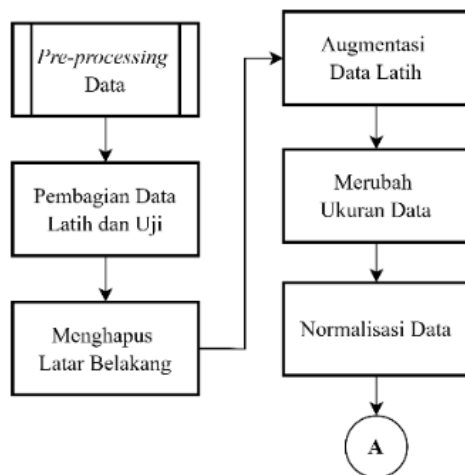
3.2. Analisis Data

Tahap awal penelitian ini dimulai dengan pengumpulan dataset citra daun anggur yang diperoleh secara langsung dari perkebunan anggur di Jawa Timur dengan proses pengambilan gambar menggunakan kamera ponsel dengan jarak 21 cm dari objek citra daun anggur, proses pengambilan gambar dilakukan pada siang hari dengan kondisi pencahayaan alami. Dataset yang digunakan merupakan dataset primer yang mencakup empat varietas yang populer karena perawatannya mudah dan produktivitasnya tinggi, yaitu Jumer, Ninel, Ruby88, dan Transfigurasi. Keempat varietas ini diklasifikasikan ke dalam empat kelas berdasarkan perbedaan bentuk, ukuran, warna, dan pola urat daunnya. Dataset berformat .png dan berjumlah total 2.400 citra pada empat kelasnya.

3.3. Pre-processing Data

Berdasarkan gambar 7 yang menjelaskan mengenai proses yang terdapat saat melakukan *Pre-processing* Data pada setiap citranya. Prosesnya dimulai dari melakukan *split* data atau membagi data kedalam 2 bagian, yaitu data latih, dan data uji. Setelah melakukan *split* data, data berupa citra akan dilakukan augmentasi pada data latih. Selanjutnya data akan disamakan dimensi ukurannya sehingga menjadi seragam, lalu proses terakhir adalah melakukan normalisasi. Adapun penjelasan dari proses pada

diagram alir *Pre-processing* Data diatas, yaitu sebagai berikut :



Gambar 7. *Pre-processing* Data

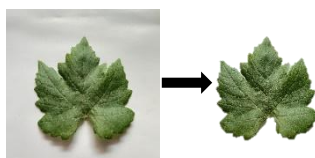
a. Pembagian Data

Terdapat sebanyak 2.400 data gambar yang mewakili keempat varietas daun anggur, yaitu Jumer, Ninel, Ruby88, dan Transfigurasi. Data gambar daun anggur tersebut akan dibagi kedalam dua subset yang berbeda, yaitu untuk pelatihan dan pengujian. Data latih digunakan untuk melakukan pelatihan pada model yang memungkinkan model akan mempelajari pola pada data citra. Data uji akan digunakan untuk melakukan pengujian pada model yang dapat meningkatkan kinerja model dan dikembangkan dalam melakukan evaluasi akhir.

Pembagian rasio dataset yang digunakan pada penelitian ini, yaitu 90:10, pembagian ini bertujuan untuk memberikan lebih banyak data latih untuk melakukan pelatihan sehingga dapat meningkatkan kemampuan model dalam belajar data citra daun anggur. Setelah proses pelatihan selesai, data pengujian akan digunakan untuk melakukan evaluasi performa model terhadap data yang belum pernah dilihat sebelumnya.

b. Menghapus Latar Belakang

Proses penghapusan latar belakang dilakukan untuk memfokuskan data citra daun dengan menghilangkan elemen-elemen lain yang tidak berkaitan langsung dengan objek daun.



Gambar 8. Proses Menghapus Latar Belakang

Dengan menghapus latar belakang seperti pada gambar 8, citra yang digunakan akan lebih bersih dan hanya menampilkan bagian daun saja, sehingga meminimalkan kemungkinan model melakukan ekstraksi fitur dari area yang tidak diperlukan seperti

bayangan hingga gangguan visual lainnya. Hal ini sangat penting dalam proses klasifikasi karena fitur-fitur yang diekstraksi model akan lebih terfokus pada karakteristik khas daun, seperti bentuk, tekstur, dan pola warna.

c. Augmentasi Data

Augmentasi data merupakan tahapan yang dilakukan untuk memperluas *dataset*, sehingga data tampak memiliki kombinasi yang beragam tanpa perlu menambahkan citra asli. Tujuan dari augmentasi data adalah untuk mengurangi resiko *overfitting* serta memastikan bahwa model mampu mengenali pola untuk berbagai kondisi pada citra. Tahap augmentasi dilakukan dengan menerapkan berbagai transformasi pada citra asli untuk menciptakan variasi pada citra baru. Data pada penelitian ini mulanya terdiri dari kumpulan citra daun yang diperoleh tanpa proses modifikasi tambahan yang berjumlah 150 data daun. Namun, jumlah data tersebut masih terbatas, sehingga untuk memperluas jumlah dan variasi data, dilakukan proses augmentasi berdasarkan skenario proporsi pembagian data 90:10, dengan 90 citra mentah diaugmentasi menjadi 540 citra per kelas, dengan sisa 60 citra mentah digunakan sebagai data uji. Total citra per kelas adalah 600 data yang akan digunakan untuk proses pelatihan dan pengujian model.

d. Merubah Ukuran Data

Tahap perubahan ukuran gambar atau *resizing image* pada penelitian ini digunakan untuk menyesuaikan dimensi inputan gambar dengan arsitektur model CNN yang digunakan, yaitu VGG16, hal ini dapat membantu model dalam melakukan pemrosesan gambar yang konsisten. Tahap ini menjadi tahap yang cukup penting pada penelitian ini, karena sebagian besar model *machine learning* membutuhkan ukuran data yang sama atau tetap termasuk model CNN.



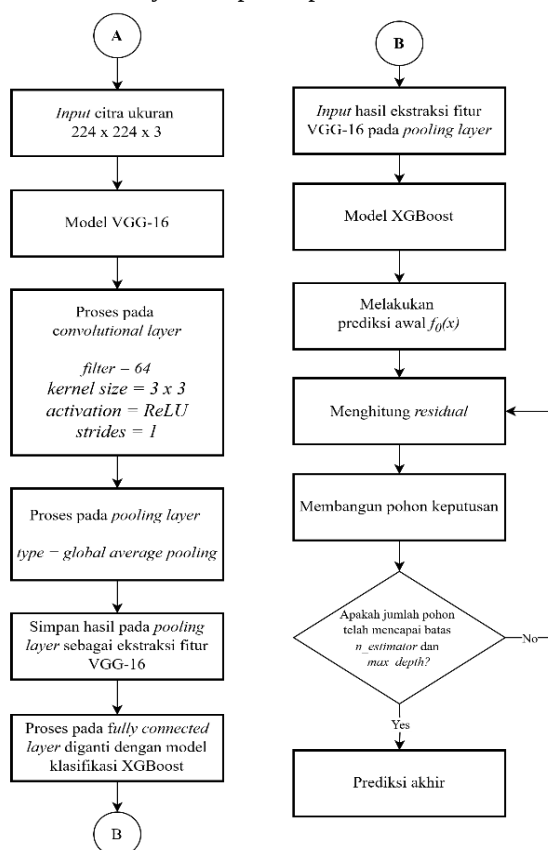
Gambar 9. Proses Merubah Ukuran Data

Pada penelitian ini, gambar daun pada mulanya memiliki ukuran yang seragam, yaitu 500 x 500 piksel. Namun, ukuran tersebut diubah menjadi 224 x 224 piksel, seperti pada gambar 3.5, sesuai dengan ukuran inputan standar yang dibutuhkan oleh arsitektur VGG-16. Adanya perubahan ukuran pada gambar yang seragam membantu mempercepat proses *training* dan *testing* karena dimensi *input* yang konsisten akan mengurangi beban model saat melakukan komputasi.

3.4. Model CNN dan XGBoost

Berdasarkan gambar 10 berikut menjelaskan mengenai arsitektur VGG-16 pada model CNN untuk melakukan pemrosesan pada data citra untuk

melakukan ekstraksi fitur. Proses tersebut akan dilanjutkan dengan model XGBoost yang akan digunakan untuk melakukan klasifikasi berdasarkan hasil dari ekstraksi fitur. Terdapat beberapa lapisan CNN untuk melakukan pemrosesan citra menggunakan VGG-16, seperti *input layer*, *convolutional layer*, dan *pooling layer*. Proses model CNN dimulai dengan melakukan *input* yang menerima gambar berukuran 224 x 224 piksel dengan 3 channel warna, yaitu *red*, *green*, *blue*. Setelahnya, citra *input* akan melalui lapisan konvolusi yang didalamnya berisi jumlah filter yang berbeda, yaitu 64, 128, 256, 512, 512, memiliki ukuran kernel yang sama, yaitu 3 x 3, menggunakan aktivasi ReLU, dengan *strides* nya adalah 1. Proses akan dilanjutkan pada *pooling layer* menggunakan *global average pooling*, sehingga dimensi hasilnya merupakan piksel berdimensi 512.



Gambar 10. Model CNN dan XGBoost

Pada model XGBoost prosesnya dimulai dengan menginputkan hasil dari proses ekstraksi fitur pada *pre-trained* model, yaitu menggunakan VGG-16. Pada tahap awal, model XGBoost akan dibangun dengan melakukan prediksi awal yang dilakukan dengan cara menghitung rerata nilai pada label target dari setiap kelas yang digunakan. Setelah prediksi awal dilakukan, selanjutnya model akan menghitung *residual* antara selisih dari nilai aktual dengan prediksi awal yang sebelumnya telah dilakukan. Berdasarkan nilai yang dihasilkan oleh *residual* ini, pohon keputusan baru akan dibangun untuk memperbaiki kesalahan prediksi pada pohon sebelumnya. Proses ini

akan dilakukan secara berulang, dimana model akan terus memperbarui nilai *residual* serta membangun pohon keputusan baru hingga jumlah pohon sesuai dengan batas *n_estimator* yang telah di inputkan. Jika jumlah pohon belum mencapai batas *n_estimator*, maka model akan mengulang proses pada perhitungan residual hingga membangun pohon keputusan baru, sedangkan jika jumlah pohon telah mencapai batas *n_estimator*, maka model akan melakukan prediksi akhir dan menyelesaikan prosesnya.

3.5. Skenario Pengujian Model

Skenario pengujian model merupakan bagian akhir yang terdapat pada penelitian ini, di mana program yang telah dibuat akan melakukan pengujian menggunakan sekumpulan data yang telah disiapkan sebagai data uji. Pada skenario ini, pengujian akan menggunakan 4 kelas data dari citra daun anggur. Tahapan pengujiannya dimulai dengan melakukan pembagian data 90:10. Pemilihan skenario ini didasarkan pada standar umum dalam *machine learning* serta merujuk pada rasio pembagian data yang telah digunakan dalam penelitian-penelitian sebelumnya.

Penggunaan parameter *learning_rate*, *max_depth*, *n_estimators*, dan *subsample*, dengan bobot nilai parameter *learning_rate* = 0.1, *max_depth* = 5, *n_estimators* = 100, dan *subsample* = 1 digunakan sebagai nilai *default* dari model XGBoost, sedangkan pada bobot nilai untuk variasi nilai lain *learning_rate* = 0.5, *max_depth* = 7, *n_estimators* = 250, dan *subsample* = 0.8, diperoleh berdasarkan penelitian sebelumnya yang melakukan *tuning* parameter pada model XGBoost untuk klasifikasi. Pemilihan ini bertujuan untuk membandingkan performa antara penggunaan *default setting* dan *setting* yang diperoleh dari penelitian terdahulu, sehingga dapat dianalisis pengaruh perubahan parameter terhadap akurasi model. Pada tabel 1 merupakan skenario pengujian yang digunakan pada penelitian ini, sebagai berikut :

Tabel 1. Skenario Pengujian

Skenario	Parameter			
	<i>Learning rate</i>	<i>Max depth</i>	<i>N estimators</i>	<i>subsample</i>
1.	0.1	5	100	0.8
2.	0.1	5	100	1
3.	0.1	5	250	0.8
4.	0.1	5	250	1
5.	0.1	7	100	0.8
6.	0.1	7	100	1
7.	0.1	7	250	0.8
8.	0.1	7	250	1
9.	0.5	5	100	0.8
10.	0.5	5	100	1
11.	0.5	5	250	0.8
12.	0.5	5	250	1
13.	0.5	7	100	0.8
14.	0.5	7	100	1
15.	0.5	7	250	0.8
16.	0.5	7	250	1

4. HASIL DAN PEMBAHASAN

Pada bagian ini akan dipaparkan mengenai hasil dari skenario pengujian menggunakan model yang telah dibangun dengan kombinasi proporsi pembagian data hingga penggunaan berbagai parameter yang dimiliki oleh model XGBoost. Pengujian ini bertujuan untuk melakukan klasifikasi pada varietas daun anggur berdasarkan citra daun menggunakan model CNN dan XGBoost. Selain itu, juga digunakan mengukur performa kombinasi model CNN dan XGBoost dalam melakukan klasifikasi. Pengujian ini dilakukan

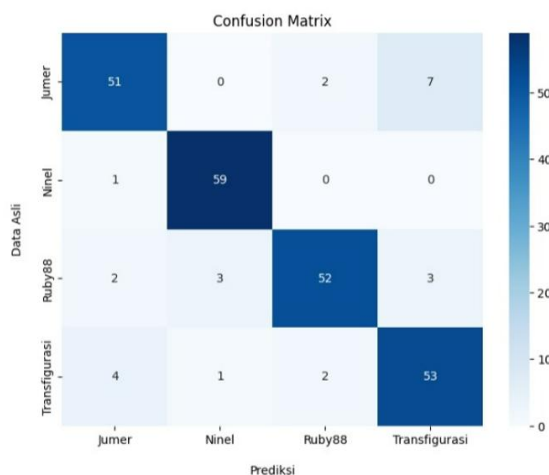
berdasarkan skenario pengujian pada tabel 1 yang memaparkan mengenai berbagai kombinasi skenario pengujian. Skenario pembagian data yang digunakan pada penelitian ini, yaitu 90:10, dengan 2.160 data untuk pelatihan dan 240 data untuk pengujian. Pembagian ini bertujuan agar model memperoleh lebih banyak data latih sehingga mampu mengenali pola dengan lebih baik. Setelah pelatihan, data uji digunakan untuk mengevaluasi performa model terhadap data yang belum pernah dilihat sebelumnya.

Tabel 2. Hasil Skenario Pembagian Data 90:10

SK	Parameter				waktu train (detik)	Evaluasi			
	LR	MD	NE	S		Accuracy	Precision	Recall	F1-Score
1	0.1	5	100	0.8	583.32	86.67%	87.46%	86.67%	86.65%
2	0.1	5	100	1	906.07	87.50%	88.54%	87.50%	87.48%
3	0.1	5	250	0.8	1710.19	88.75%	89.10%	88.75%	88.75%
4	0.1	5	250	1	1798.13	89.17%	89.42%	89.17%	89.11%
5	0.1	7	100	0.8	1431.83	87.08%	88.02%	87.08%	87.11%
6	0.1	7	100	1	1734.41	85.42%	86.58%	85.42%	85.34%
7	0.1	7	250	0.8	2112.2	87.08%	87.84%	87.08%	87.07%
8	0.1	7	250	1	2094.63	87.08%	87.74%	87.08%	87.00%
9	0.5	5	100	0.8	499.55	88.75%	88.89%	88.75%	88.72%
10	0.5	5	100	1	524.24	86.25%	87.08%	86.25%	86.22%
11	0.5	5	250	0.8	832.55	89.58%	89.64%	89.58%	89.55%
12	0.5	5	250	1	1433.55	85.42%	86.30%	85.42%	85.37%
13	0.5	7	100	0.8	556.72	85.83%	86.60%	85.83%	85.86%
14	0.5	7	100	1	580.61	88.75%	89.15%	88.75%	88.70%
15	0.5	7	250	0.8	924.52	85.83%	86.35%	85.83%	85.82%
16	0.5	7	250	1	941.54	88.75%	89.15%	88.75%	88.70%

Setelah melakukan pengujian dengan menggunakan 16 skenario pengujian seperti pada tabel 2 dengan SK (Skenario), LR (*learning_rate*), MD (*max_depth*), NE (*n_estimators*) dan S(*subsamples*), didapatkan performa model terbaik terdapat pada skenario pengujian yang ke – 11 dengan kombinasi parameter *learning_rate* = 0.5, *max_depth* = 5, *n_estimators* = 250, dan *subsamples* = 0.8 dengan hasil *accuracy* = 89.58%, *precision* = 89.64%, nilai *recall* = 89.58%, nilai *f1-Score* sebesar 89.55% dengan estimasi waktu *training* selama 832.55 detik.

Adapun hasil visualisasi *confusion matrix* pada skenario pengujian model ke-11 menunjukkan hasil klasifikasi empat varietas daun anggur: Jumer, Ninel, Ruby88, dan Transfigurasi. Setiap baris pada matrix merepresentasikan label asli, sementara kolom menunjukkan hasil prediksi. Hasilnya menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Sebagai contoh, dari 60 data daun varietas Ninel, sebanyak 59 berhasil diprediksi dengan tepat, dan hanya 1 data yang salah diklasifikasikan sebagai Jumer.



Gambar 11. *Confusion Matrix* Skenario Pengujian Ke – 11 (90:10)

5. KESIMPULAN DAN SARAN

Penelitian ini mengklasifikasikan empat varietas anggur, yaitu varietas Jumer, Ninel, Ruby88, dan Transfigurasi berdasarkan citra daun yang dikumpulkan dari perkebunan di Jawa Timur. Data diproses melalui tahapan *pre-processing* dan diuji pada rasio pembagian data 90:10, 80:20, 70:30, dan 60:40. Metode yang digunakan merupakan kombinasi CNN dengan arsitektur VGG-16 untuk ekstraksi fitur dan XGBoost sebagai klasifikasi. Hasil terbaik diperoleh pada rasio 90:10, yaitu pada skenario ke – 11 dengan *accuracy* 89,58%, *precision* 89,64%, *recall* 89,58%, *f1-score* 89,55%, dan nilai *loss* sebesar 0,34793.

Untuk pengembangan selanjutnya, disarankan menambah jumlah data dan varietas anggur, selain itu

mencoba menggunakan algoritma lain di luar kombinasi CNN dan XGBoost, serta mengoptimalkan parameter XGBoost guna meningkatkan performa model klasifikasi.

DAFTAR PUSTAKA

- [1] H. A. Ahmed, H. M. Hama, S. I. Jalal, and M. H. Ahmed, "Deep Learning in Grapevine Leaves Varieties Classification Based on Dense Convolutional Network," *J. Image Graph. Kingdom*, vol. 11, no. 1, pp. 98–103, 2023, doi: 10.18178/joig.11.1.98-103.
- [2] G. Poornimia and D. D. S. Sakari, "SEGMENTATION AND CLASSIFICATION OF SKIN LESIONS IN DERMOSCOPIC IMAGES USING DEEP TRANSFER LEARNING TECHNIQUES," *NeuroQuantology*, vol. 20, no. 10, pp. 4499–4525, 2022, doi: 10.14704/nq.2022.20.10.NQ55433.
- [3] M. A. N. Hidayat, Kusri, and Hanafi, "Convolutional Neural Network Pada Identifikasi Varian Tanaman Anggur Menggunakan Resnet-50," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 3, pp. 61–70, 2023, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/download/4789/1531>
- [4] M. Koklu, M. F. Unler, I. A. Ozkan, M. F. Aslan, and K. Sabanci, "A CNN-SVM study based on selected deep features for grapevine leaves classification," *Meas. J. Int. Meas. Confed.*, vol. 188, no. November 2021, p. 110425, 2022, doi: 10.1016/j.measurement.2021.110425.
- [5] A. Kusuma Whardana, D. Febriyanto, M. Jagad Katanka, N. Andina Oktavia, and T. Desta Loria, "Klasifikasi Penyakit Daun Anggur dengan Menggunakan Convolutional Neural Network dan Transfer Learning dari VGG16," *J. Sist. Komput. dan Kecerdasan Buatan*, vol. 7, no. 2, pp. 198–202, 2024.
- [6] A. P. Asia, "Tanaman Buah Anggur : Klasifikasi, Ciri Ciri, Jenis, Dan Syarat Tumbuh," [alatpertanian.asia](https://www.alatpertanian.asia). [Online]. Available: <https://www.alatpertanian.asia/2023/10/tanaman-buah-anggur-klasifikasi-ciri.html>
- [7] R. Tani, "Jenis Anggur Berdasarkan Bentuk Daun," www.jualbenihmurah.com. Accessed: Apr. 29, 2025. [Online]. Available: [https://www.jualbenihmurah.com/blog/karakteristik-bentuk-daun-anggur#:~:text=Varietas](https://www.jualbenihmurah.com/blog/karakteristik-bentuk-daun-anggur#:~:text=Varietas%20anggur%20ninel%20dikenal%20dengan%20tandan%20buahnya%20yang%20ukuran%20daunnya%20cukup%20lebar%20meskipun%20batangnya%20berukuran%20kecil)
- [8] Arnita, F. Marpaung, F. Aulia, N. Suryani, and R. C. Nabila, *Computer Vision Dan Pengolahan Citra Digital*. Surabaya, Jawa Timur, Indonesia: Pustaka Aksara, 2022. [Online]. Available: www.pustakaaksara.co.id
- [9] Q. Lina, "Apa itu Convolutional Neural Network?," medium.com. Accessed: Nov. 20, 2024. [Online]. Available: <https://medium.com/@16611110/apa-itu-convolutional-neural-network-836f70b193a4>
- [10] N. H. Hameed and W. A. Araheemah, "Using an Improved Visual Geometry Group Neural Network for Skin Cancer Detection," *Al-Rafidain J. Eng. Sci.*, vol. 2, no. 1, pp. 265–273, 2024, doi: 10.61268/xankz410.
- [11] I. M. Karo Karo, D. Kiswanto, S. Panggabean, and A. Perdana, "Hair Disease Classification Using Convolutional Neural Network (CNN) Algorithm with VGG-16 Architecture," *Sinkron*, vol. 7, no. 4, pp. 2786–2793, 2023, doi: 10.33395/sinkron.v8i4.13110.
- [12] S. E. Herni Yulianti, Oni Soesanto, and Yuana Sukmawaty, "Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit," *J. Math. Theory Appl.*, vol. 4, no. 1, pp. 21–26, 2022, doi: 10.31605/jomta.v4i1.1792.
- [13] Jan Melvin Ayu Soraya Dachi and Pardomuan Sitompul, "Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit," *J. Ris. Rumpun Mat. Dan Ilmu Pengetah. Alam*, vol. 2, no. 2, pp. 87–103, 2023, doi: 10.55606/jurrimipa.v2i2.1470.