

# ANALISIS KINERJA ALGORITMA PEMBELAJARAN MESIN ENSEMBEL PADA DATASET MULTI KELAS CITRA JAFFE

## PERFORMANCE ANALYSIS OF ENSEMBLE MACHINE LEARNING ALGORITHMS ON JAFFE IMAGE MULTICLASS DATASETS

Huzain Azis<sup>1)</sup>, Alisma<sup>2)</sup>, Purnawansyah<sup>3)</sup>, Nirmala<sup>4)</sup>

<sup>1.) 2), 3), 4)</sup> Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muslim Indonesia  
Jl. Urip Sumoharjo No.km.5, Panaikang, Kec. Panakkukang, Kota Makassar, Sulawesi Selatan 90231  
Email : [huzain.azis@umi.ac.id](mailto:huzain.azis@umi.ac.id)<sup>1)</sup>, [13020200177@umi.ac.id](mailto:13020200177@umi.ac.id)<sup>2)</sup>, [purnawansyah@umi.ac.id](mailto:purnawansyah@umi.ac.id)<sup>3)</sup>,  
[malaaaanir.iclabs@umi.ac.id](mailto:malaaaanir.iclabs@umi.ac.id)<sup>4)</sup>

### Abstrak

Penelitian ini bertujuan untuk mengembangkan sistem pengenalan ekspresi wajah berbasis dataset JAFFE yang mencakup tujuh kelas ekspresi emosi, yakni ekspresi senang, sedih, marah, takut, jijik, dan netral. Langkah pertama yang dilakukan yakni segmentasi canny pada setiap dataset untuk menonjolkan tepi dan meningkatkan ketajaman gambar. Selanjutnya, dilakukan ekstraksi menggunakan metode hu moments guna mendapatkan pemahaman mendalam terhadap karakteristik penting dalam ekspresi wajah. Proses selanjutnya melibatkan ensemble vote dengan memanfaatkan lima metode klasifikasi, yaitu Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Gaussian Process Classifier (GPC), dan Decision Tree. Hasil dari kelima metode ini kemudian di ensemble menggunakan teknik voting, dan hasil akhirnya dievaluasi dengan menggunakan metrik kinerja seperti akurasi, presisi, recall, dan nilai F-1 score. Evaluasi dilakukan dengan membandingkan hasil akhir dengan data asli dari dataset JAFFE, dengan mengukur akurasi, presisi, recall, dan nilai F1 Score untuk mengevaluasi kinerja sistem. Meskipun pendekatan ensemble vote digunakan untuk meningkatkan performa, hasil penelitian menunjukkan bahwa metode ini tidak selalu memberikan peningkatan kinerja yang signifikan dibandingkan metode individu. Akurasi tertinggi diperoleh oleh Decision Tree sebesar 0,14, dengan presisi 0,15, recall 0,15, dan F1-Score 0,15, sedangkan model ensemble memperoleh akurasi 0,12, presisi 0,11, recall 0,12, dan F1-Score 0,09. Termasuk interaksi manusia dan komputer serta aplikasi di bidang kecerdasan buatan.

**Kata kunci:** Analisis Performa, Ensemble, Citra Jaffe, Klasifikasi, Multiclass

### Abstract

This research aims to develop a facial expression recognition system based on the JAFFE dataset which includes seven classes of emotional expressions, namely happy, sad, angry, afraid, disgusted and neutral expressions. The first step taken is canny segmentation on each dataset to maintain essential information on each face. Next, extraction was carried out using the hu moments method to gain an in-depth understanding of the important characteristics of facial expressions. The next process involves ensemble voting using five classification methods, namely Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Gaussian Process Classifier (GPC), and Decision Tree. The results of these five methods are then ensemble using voting techniques, and the final results are evaluated using performance metrics such as accuracy, precision, recall, and F-1 score. Evaluation is carried out by comparing the final results with the original data from the JAFFE dataset, by measuring accuracy, precision, recall, and F1 Score value to evaluate system performance. The results of this research show that the ensemble voting approach using a combination of classification methods is able to significantly improve facial expression recognition capabilities. The resulting accuracy, precision, recall, and F1 Score values provide a comprehensive picture of system performance. This research contributes to the development of facial emotion recognition technology and can be applied in various contexts. Includes human-computer interaction as well as applications in the fields of artificial intelligence.

**Keywords:** Performance Analysis, Ensemble, Jaffe Image, Classification, Multiclass.

## 1. PENDAHULUAN

Pengenalan ekspresi wajah manusia telah menjadi topik yang menarik dalam bidang *computer vision*, dengan potensi aplikasi yang luas dalam bidang seperti deteksi emosi, dan pengenalan kebohongan. Dataset JAFFE, yang mencakup gambar-gambar wajah manusia dengan label emosi seperti senang, sedih, marah, takut, terkejut, jijik, dan netral, telah menjadi subjek

penelitian yang penting dalam memahami ekspresi wajah manusia. Namun, untuk mengoptimalkan penggunaan dataset ini, diperlukan pendekatan yang sistematis dan efisien.

Meskipun banyak penelitian telah memanfaatkan dataset JAFFE, terdapat *gap* penelitian dalam hal pemanfaatan metode *ensemble* untuk meningkatkan performa klasifikasi. Metode *ensemble* seperti *ensemble voting* berpotensi mengatasi keterbatasan performa dengan menggabungkan hasil dari beberapa model klasifikasi. Pendekatan ini dipilih karena penelitian terdahulu yang hanya menggunakan metode individual sering kali tidak mencapai kinerja optimal. Dengan menggabungkan beberapa algoritma, diharapkan akurasi dan stabilitas model dapat ditingkatkan, terutama untuk pengenalan ekspresi wajah yang memiliki variasi emosi yang kompleks.

Penelitian ini bertujuan untuk menjawab pertanyaan – pertanyaan kritis terkait dengan efektivitas *ensemble vote* dalam meningkatkan kinerja pengenalan ekspresi wajah. Pertanyaan – pertanyaan tersebut melibatkan sejauh mana kombinasi metode klasifikasi dapat memberikan hasil yang lebih baik dibandingkan dengan pendekatan individual, serta sejauh mana sistem ini dapat diandalkan dalam mengenali berbagai ekspresi emosi pada wajah manusia [1]. Lingkup penelitian mencakup penggunaan *dataset JAFFE* dan penerapan metode *ensemble vote* dengan lima metode klasifikasi yang telah disebutkan sebelumnya. Namun, penelitian ini memiliki Batasan terkait dengan keterbatasan *dataset* dan perluasan model ke ekspresi wajah yang lebih kompleks.

Dengan menggabungkan Langkah – langkah segmentasi, ekstraksi fitur hu moments, dan pendekatan *ensemble vote*, penelitian ini diharapkan dapat memberikan kontribusi positif dalam menghadapi tantangan dalam pengenalan ekspresi wajah secara otomatis dan juga diharapkan dapat membuka pintu menuju pengembangan teknologi yang dapat secara efektif berinteraksi dengan manusia, mengenali dan memahami ekspresi emosi secara mendalam [2]. Kontribusi utama dari penelitian ini adalah mengetahui metode klasifikasi yang dapat mengenali ekspresi wajah manusia dengan tingkat akurasi, presisi, recall, dan nilai F1-Score yang tinggi. Selain itu, penelitian ini juga akan menyediakan wawasan tentang efektivitas *ensemble voting* dalam kinerja klasifikasi. Struktur artikel ini akan membahas langkah – langkah segmentasi, ekstraksi fitur hu moments serta pendekatan *ensemble vote*, menyajikan temuan utama, dan menguraikan kesimpulan serta arah penelitian mendatang [3][4].

Artikel ini akan terus membahas langkah-langkah implementasi, temuan signifikan, dan memberikan wawasan mendalam ke dalam hasil penelitian ini, dengan harapan dapat menjadi landasan bagi pengembangan lebih lanjut dalam pemahaman dan pengenalan ekspresi wajah secara otomatis.

## 2. DASAR TEORI

Pengenalan ekspresi wajah dalam *computer vision* memanfaatkan dataset seperti JAFFE, yang mencakup berbagai ekspresi dasar untuk keperluan klasifikasi emosi. Pada proses ini, segmentasi tepi (seperti Canny) digunakan untuk menonjolkan fitur wajah, sementara *hu moments* membantu dalam ekstraksi pola unik dari setiap ekspresi. Berbagai algoritma pembelajaran mesin—termasuk Naive Bayes, Support Vector Machine, K-Nearest Neighbors, Gaussian Process Classifier, dan Decision Tree—dapat digunakan untuk klasifikasi multikelas, dan pendekatan *ensemble vote* sering diandalkan untuk menggabungkan prediksi guna meningkatkan akurasi secara keseluruhan.

### 2.1. SEGMENTASI GAMBAR CANNY

*Canny segmentation* atau adalah teknik pemrosesan citra yang berfungsi mendeteksi tepi atau kontur penting dalam gambar. Proses ini melibatkan penghalusan gambar untuk mengurangi noise, menghitung gradien intensitas untuk menemukan perubahan signifikan yang menandakan tepi, dan menerapkan ambang batas ganda untuk memisahkan tepi kuat dan lemah. Dengan metode ini, hanya kontur penting yang dipertahankan [5].

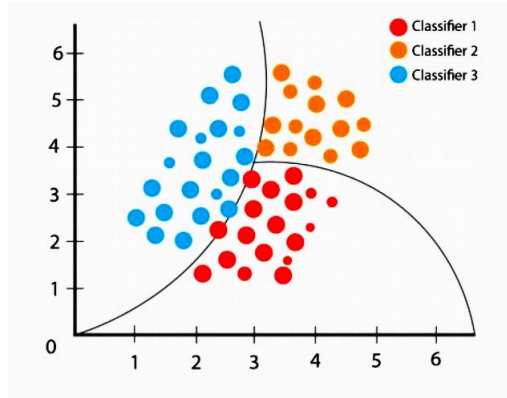
**2.2. EKSTRAKSI FITUR HU MOMENT**

Ekstraksi fitur *Hu moments* adalah metode yang digunakan untuk menangkap karakteristik bentuk dan pola dalam citra secara invarian terhadap perubahan posisi, ukuran, dan rotasi. Teknik ini menghasilkan tujuh momen statistik yang unik, yang dapat mewakili karakteristik geometris penting dari objek dalam gambar. Dalam konteks pengenalan ekspresi wajah, *Hu moments* membantu menggambarkan pola-pola khas dari berbagai ekspresi wajah dalam bentuk numerik, memungkinkan model pembelajaran mesin mengenali perbedaan halus antar ekspresi [6].

**2.3. ALGORITMA PEMBELAJARAN**

**2.3.1. NAIVE BAYES**

*Naive Bayes* adalah suatu metode klasifikasi probabilistik yang berdasarkan pada teorema Bayes, yang mengasumsikan independensi antara setiap fitur dalam *dataset*, meskipun kondisi ini mungkin tidak sepenuhnya terpenuhi dalam dunia nyata [7]. *Naive Bayes* bekerja dengan menghitung probabilitas masing-masing kelas (ekspresi emosi) berdasarkan fitur-fitur yang telah diekstraksi. Model ini kemudian membandingkan probabilitas kelas-kelas tersebut dan menentukan kelas dengan probabilitas tertinggi sebagai prediksi akhir [8][9].



**Gambar 1.** *Naive Bayes*

Pertimbangan dasar *Naive Bayes* (NB) adalah teorema Bayes, yang menyatakan bagaimana probabilitas suatu hipotesis berubah berdasarkan bukti baru. Dalam konteks klasifikasi, hipotesis adalah kelas atau label, dan bukti baru adalah fitur-fitur yang diamati. Algoritma *Naive Bayes* mencoba menghitung probabilitas setiap kelas berdasarkan distribusi fitur-fiturnya [10]. Formula dapat dilihat pada persamaan 1.

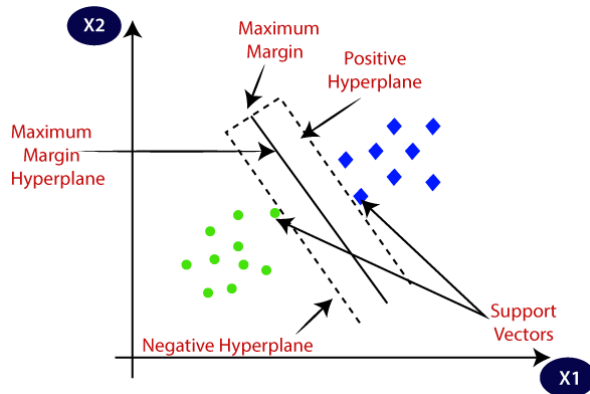
$$P(C | X) = \frac{P(C | X) \cdot P(C)}{P(X)} \tag{1}$$

Pada formula ini,  $P(C | X)$  disebut sebagai probabilitas posterior, yang menunjukkan probabilitas suatu kelas  $C$  diberikan fitur-fitur  $X$  yang diamati. Komponen  $P(X | C)$  adalah *likelihood*, yaitu probabilitas bahwa fitur  $X$  muncul jika kelas yang benar adalah  $C$ . Selanjutnya,  $P(C)$  adalah prior probability atau probabilitas awal dari kelas  $C$  dalam dataset, yang mencerminkan seberapa umum kelas tersebut. Terakhir,  $P(X)$  adalah probabilitas total dari fitur  $X$ , berfungsi sebagai faktor normalisasi, meskipun dalam banyak kasus tidak perlu dihitung karena tujuan utama adalah membandingkan nilai relatif antar kelas.

**2.3.2. SUPPORT VECTOR MACHINE**

*Support Vector Machine* (SVM) sangat efektif digunakan untuk klasifikasi dan *regresi*. Metode ini sangat efektif dalam menangani masalah klasifikasi dan *regresi* pada *dataset* yang kompleks [11]. Ide dasar dari SVM adalah mencari *hyperplane* terbaik yang dapat memisahkan dua kelas dalam ruang fitur. *Hyperplane* ini dipilih sedemikian rupa sehingga jarak antara *hyperplane* dan *instance* terdekat dari kedua kelas, yang disebut *vector* dukungan (*support*

vectors), maksimum. Dengan kata lain, SVM berusaha untuk mencari *hyperplane* yang memiliki *margin* maksimum antara dua kelas [12].



**Gambar 2.** Support Vektor Machine

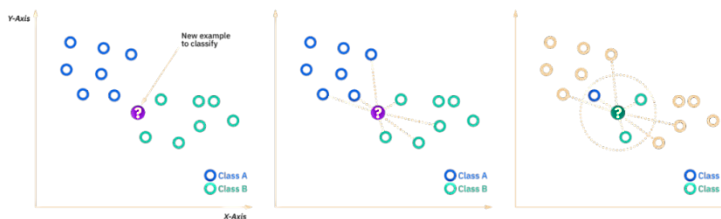
Pada prakteknya, solusi SVM sering melibatkan menggunakan teknik optimasi untuk menemukan nilai optimal dari  $W$  dan  $b$ . Metode ini melibatkan fungsi Lagrange dan dapat melibatkan penggunaan kernel untuk menangani masalah klasifikasi *non-linear*. Solusi yang optimal memerlukan penyelesaian masalah optimasi tertentu yang dapat diselesaikan dengan metode seperti SMO (*Sequential Minimal Optimization*) atau metode optimasi *convex*. Rumus lengkapnya dapat menjadi lebih kompleks tergantung pada penggunaan kernel dan masalah spesifik yang dihadapi [6]. Formula dapat dilihat pada persamaan 2.

$$f(x) = \omega \cdot x + b = 0 \tag{2}$$

Pada formula  $y_i(\omega \cdot x_i + b) \geq 1$  untuk setiap  $i$ ,  $y_i$  merupakan label kelas dari titik data  $x_i$ , dengan  $y_i = +1$  untuk kelas positif dan  $y_i = -1$  untuk kelas negatif.

### 2.3.3. K-NEAREST NEIGHBORS

*K-Nearest Neighbors* (KNN) merupakan algoritma klasifikasi yang bekerja dengan mengambil sejumlah nilai  $K$  data terdekat (tetangganya) sebagai acuan untuk menentukan kelas dari data baru [13]. KNN mengklasifikasikan data berdasarkan *similarity* atau kemiripan atau kedekatannya terhadap data lainnya. Algoritma KNN ini bersifat *lazy learning* yang berarti tidak menggunakan titik data *training* untuk membuat model. Singkatnya pada algoritma KNN sangat minim ada fase *training* [14][15].



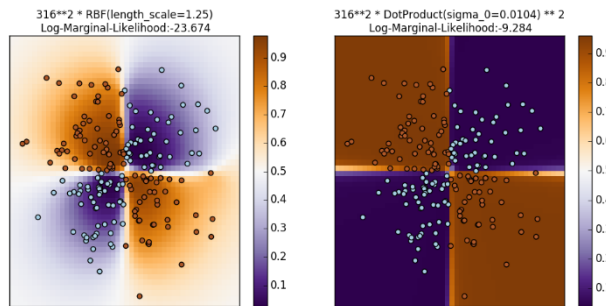
**Gambar 3.** K-Nearest Neighbor

*K-Nearest Neighbors* (KNN) merupakan algoritma yang berfungsi untuk melakukan klasifikasi suatu data yang diambil dari  $k$  tetangga terdekatnya (*nearest neighbors*). Dengan  $k$  merupakan banyaknya tetangga terdekat. *K-nearest neighbors* melakukan klasifikasi dengan proyeksi data pembelajaran pada ruang berdimensi banyak. Ruang ini dibagi menjadi bagian-bagian yang merepresentasikan kriteria data pembelajaran. Setiap data pembelajaran direpresentasikan menjadi titik-titik  $c$  pada ruang dimensi banyak [16]. Formula dapat dilihat pada persamaan 3.

$$d(x, x_i) = \sqrt{1 \sum_{j=1}^n (x_j - x_{ij})^2} \tag{3}$$

**2.3.4. GAUSSIAN PROCESS CLASSIFIER**

*Gaussian Process Classifier* (GPC) adalah metode dalam *machine learning* yang menggunakan *Gaussian processes* untuk tugas klasifikasi. Dalam *Gaussian Process*, kita memandang model sebagai distribusi probabilitas atas fungsi yang mungkin memetakan input ke output. GPC memberikan distribusi probabilitas atas kelas yang mungkin untuk setiap titik data. GPC sering digunakan dalam situasi di mana ketidakpastian prediksi dan pemodelan distribusi probabilitas atas kelas-kelas yang diinginkan. Kelebihan GPC melibatkan kemampuannya menghasilkan prediksi yang bersifat probabilitas dan memperhitungkan ketidakpastian pada data [17].



**Gambar 4.** *Gaussian Process Classifier*

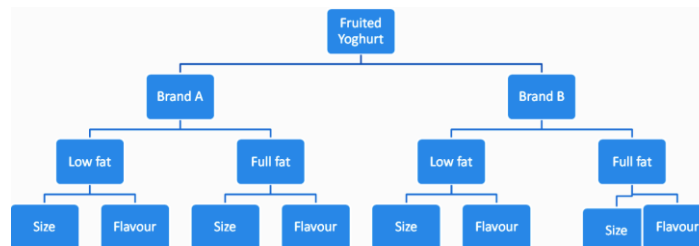
Dalam konteks GPC, kita dapat mengasumsikan bahwa kelas data pada setiap titik mengikuti distribusi Bernoulli, yang dapat dimodelkan dengan menggunakan fungsi logit atau sigmoid. Sebagai catatan, penulisan rumus GPC bisa menjadi cukup rumit dan melibatkan konsep matematis seperti distribusi multivariat Gaussian [18]. Formula dapat dilihat pada persamaan 4.

$$f(x) \sim \mathcal{G}\rho(m(x), k(x, x')) \tag{4}$$

$m(x)$  adalah fungsi mean dari proses Gaussian, sering kali diasumsikan nol ( $m(x) = 0$ ).  $k(x, x')$  adalah fungsi covariance atau kernel yang mengukur kesamaan antara dua titik data  $x$  dan  $x'$ . Kernel umum termasuk *Radial Basis Function* (RBF), *Matern kernel*, dan *linear kernel*.

**2.3.5. DECISION TREE**

Metode *Decision Tree* (pohon keputusan) adalah salah satu algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi dan *regresi*. Pohon keputusan memodelkan keputusan berdasarkan serangkaian aturan *if-else* yang membentuk struktur berhirarki berbentuk pohon. Setiap *node* dalam pohon mewakili suatu tes pada fitur data, cabang-cabang dari setiap *node* merupakan hasil dari tes tersebut dan daun-daun (*leaf nodes*) mewakili kelas atau nilai *regresi* [19]. Proses pembentukan pohon ini dilakukan secara rekursif, dengan memilih aturan pengelompokan yang paling informatif pada setiap Langkah [20].



**Gambar 5.** *Decision Tree*

Formula pada metode *Decision Tree* dapat dilihat pada persamaan 5.

$$Entropy(S) = - \sum_{i=1}^e p_i \log_2(p_i) \tag{5}$$

$S$  adalah himpunan data pada node tersebut.  $c$  adalah jumlah kelas yang berbeda.  $p_i$  adalah proporsi dari data dalam kelas  $i$  di node tersebut.

## 2.4. PERFORMANCE COMPARISON ANALYSIS

Penelitian ini melibatkan analisis metrik kinerja yang diperoleh dari masing-masing metode *ensemble*. Metrik evaluasi standar seperti akurasi, presisi, perolehan, dan pengukuran *F-1 score* digunakan untuk mengevaluasi performa model. Metrik ini memberikan wawasan tentang kemampuan model dalam mengklasifikasikan objek dengan benar dan mengukur kualitas prediksi yang dihasilkan. Selanjutnya dilakukan perbandingan kinerja diantara keempat metode *ensemble* yang digunakan untuk mengetahui kelebihan dan kelemahan masing-masing metode dalam hal akurasi, presisi, *recall*, dan *F-I score*. Hasil perbandingan ini memberikan wawasan penting mengenai metode *ensemble* mana yang paling efektif dalam tugas klasifikasi objek pada dataset gambar yang digunakan [21].

Akurasi mengukur kemampuan model secara keseluruhan untuk mengklasifikasikan data dengan benar dan ditentukan oleh.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{6}$$

Presisi mengukur sejauh mana model mengidentifikasi hal positif dengan benar dibandingkan dengan semua prediksi positifnya, seperti yang didefinisikan dalam.

$$Presisi = \frac{TP}{(TP + FP)} \tag{7}$$

*Recall (Sensitivity atau True Positive Rate)* mengukur sejauh mana model mampu mengklasifikasikan *true* positif dengan benar, sebagaimana didefinisikan dalam.

$$Recall = \frac{TP}{(TP + FN)} \tag{8}$$

*F-1 Score* adalah rata-rata harmonik dari presisi dan perolehan. Ini digunakan untuk menggabungkan presisi dan perolehan menjadi nilai komprehensif, seperti yang didefinisikan dalam

$$f - 1 \text{ score} = \frac{2(\text{presisi} \times \text{recall})}{(\text{presisi} + \text{recall})} \tag{9}$$

Keterangan :

TP (*True Positive*) adalah jumlah positif yang di prediksi dengan benar.

TN (*True Negative*) adalah jumlah prediksi negative yang benar.

FP (*False Positive*) adalah jumlah prediksi positif yang salah.

FN (*False Negative*) adalah jumlah prediksi negative yang salah.

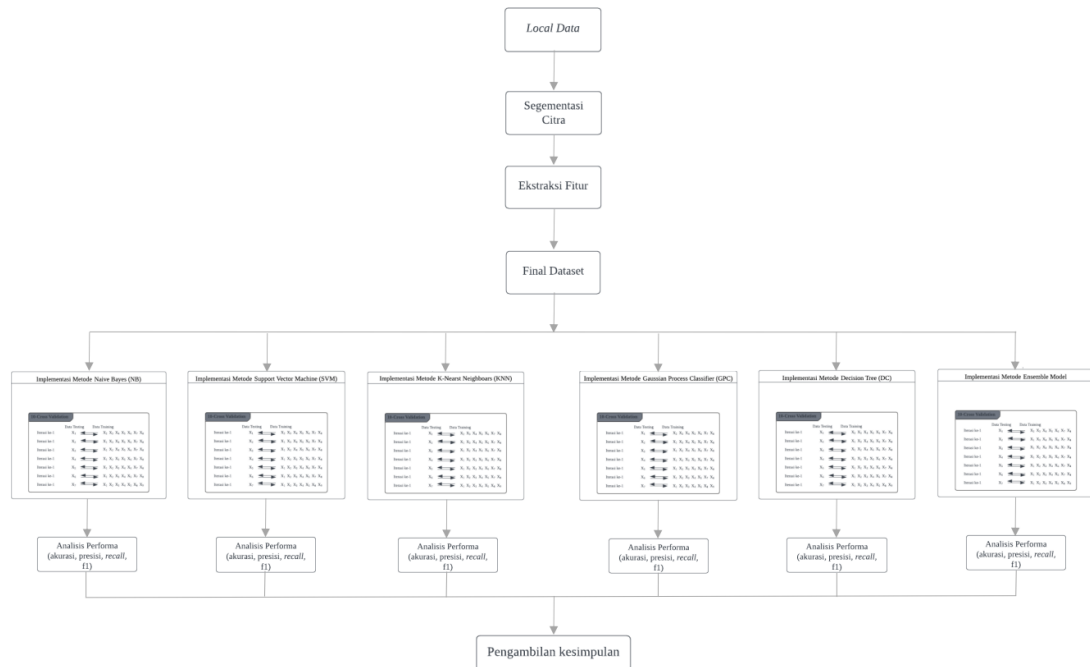
## 3. METODE PENELITIAN

### 3.1. DESAIN PENELITIAN

Desain penelitian ini terdiri dari lima tahapan utama, dimulai dari pengumpulan *Local Data* dengan menggunakan dataset JAFFE sebagai dasar. Tahap berikutnya adalah *Segmentasi Citra*, di mana metode Canny digunakan untuk menonjolkan tepi pada citra wajah, diikuti oleh *Ekstraksi Fitur* menggunakan *hu moments* untuk menangkap karakteristik penting dari ekspresi wajah. Setelah fitur diekstraksi, tahapan *Implementasi Model Klasifikasi* dilakukan dengan berbagai algoritma, yaitu Naive Bayes, SVM, KNN, Gaussian Process Classifier, Decision Tree,



serta model *ensemble*, yang masing-masing menghasilkan prediksi terhadap ekspresi wajah. Langkah akhir adalah *Analisis Performa*, di mana akurasi, presisi, recall, dan F1-Score digunakan untuk mengevaluasi setiap model, diikuti dengan pengambilan kesimpulan untuk menentukan metode terbaik dalam pengenalan ekspresi wajah. Detail dari tahapan dapat dilihat pada Gambar 6.

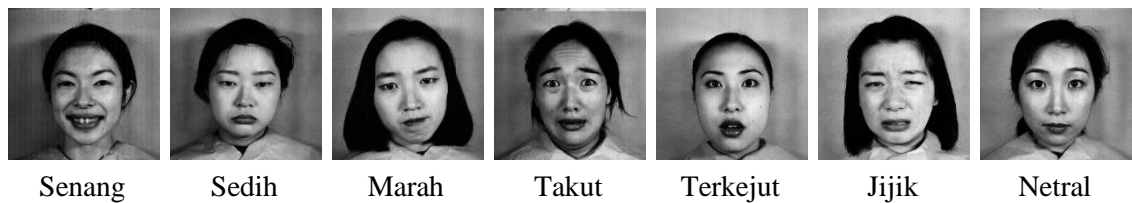


**Gambar 6.** Tahapan Umum Desain Penelitian

### 3.2. DATASET MULTI KELAS

*Dataset multiclass* merujuk pada kumpulan data yang mencakup lebih dari dua kelas atau kategori. Dalam konteks klasifikasi, setiap sampel dapat dikategorikan ke dalam salah satu dari beberapa kelas yang berbeda [22].

*Dataset multiclass* yang digunakan dalam penelitian ini mengenai informasi ekspresi wajah dalam konteks tujuh kelas emosional yang berbeda, yaitu senang, sedih, marah, takut, terkejut, jijik, dan netral. Contoh dataset yang digunakan dapat dilihat pada Gambar 7. Setiap kelas tersebut mewakili sebuah kategori yang membedakan variasi ekspresi emosional pada wajah manusia [23].



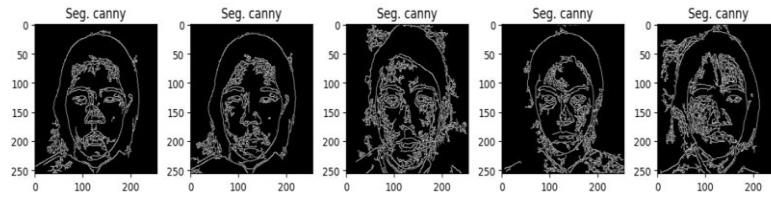
**Gambar 7.** Sample Dataset Jaffe

*Dataset multiclass* menjadi kunci dalam pengembangan model yang mampu memahami dan mengenali berbagai variasi ekspresi emosional pada wajah manusia, dengan potensi aplikasi dalam pengenalan wajah otomatis dan analisis sentimen [24].

### 3.3. DATASET PREPROCESSING

Dalam tahap awal penelitian ini, dilakukan *Dataset Preprocessing* untuk mempersiapkan *dataset JAFFE* sebelum menjalani proses pengenalan ekspresi wajah. Langkah pertama yang dilakukan adalah melakukan segmentasi canny pada setiap citra pada *dataset* [25]. Proses

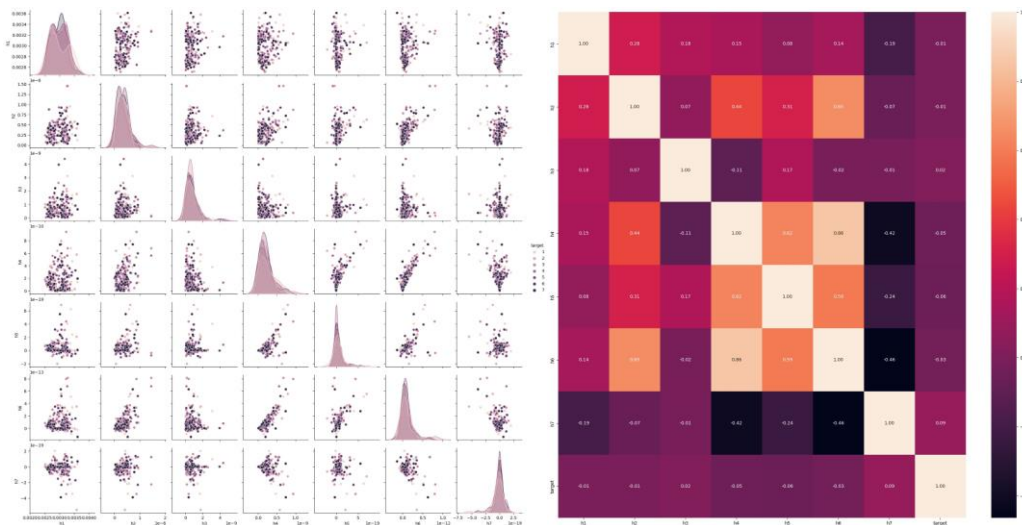
segmentasi ini bertujuan untuk mengekstraksi tepi-tepi yang signifikan pada wajah, mempertahankan informasi esensial yang diperlukan untuk pengenalan ekspresi [26].



Gambar 8. Hasil Gambar Segmentasi Canny

### 3.4. DATASET PREPROCESSING

Tahap selanjutnya yaitu proses Ekstraksi fitur hu moments yang memiliki tujuan untuk memahami karakteristik wajah yang lebih mendalam, menciptakan representasi fitur yang lebih abstrak dan signifikan dari setiap citra [27]. Gambar 9 merupakan hasil ekstraksi fitur dari dataset JAFFE yang telah melalui proses segmentasi Canny untuk menonjolkan tepi pada wajah. Plot di sebelah kiri menunjukkan distribusi dan korelasi antar fitur yang diekstraksi, membantu mengidentifikasi pola tertentu yang mungkin mendukung klasifikasi ekspresi wajah. Sementara itu, peta panas (heatmap) di sebelah kanan memperlihatkan korelasi antar fitur, di mana warna yang lebih gelap menunjukkan korelasi yang lebih kuat antara fitur-fitur tertentu. Proses ini memberikan landasan yang kuat untuk mengenali pola – pola khas yang berkaitan dengan ekspresi emosi pada wajah [28].



Gambar 9. Ekstraksi Fitur Hu Moments

### 3.5. ALGORITHM IMPLEMENTATION

Metode *ensemble vote* adalah suatu pendekatan dalam pembelajaran mesin di mana hasil prediksi dari beberapa model atau algoritma berbeda digabungkan untuk menghasilkan keputusan akhir. Dalam konteks klasifikasi, metode *ensemble vote* bekerja dengan mengumpulkan prediksi yang dihasilkan oleh beberapa model klasifikasi, kemudian menentukan hasil akhir berdasarkan mayoritas atau bobot tertentu dari prediksi tersebut [29]. Sebagai contoh, jika metode ini beberapa algoritma untuk mengklasifikasikan data, keputusan akhir diambil dari prediksi yang paling banyak dipilih (mayoritas suara) atau dengan bobot tertentu pada setiap model. Dengan cara ini, *ensemble vote* dapat menghasilkan prediksi yang lebih akurat dan stabil.



### 3.6. PENGAMBILAN KEPUTUSAN

Berdasarkan analisis dan perbandingan kinerja, diambil keputusan mengenai metode *ensemble* mana yang memberikan kinerja terbaik dalam tugas klasifikasi objek pada kumpulan data gambar ini. Keputusan ini didasarkan pada evaluasi kinerja yang dilakukan dan dapat digunakan sebagai panduan untuk memilih metode *ensemble* yang optimal dalam konteks penelitian ini [30].

## 4. HASIL DAN PEMBAHASAN

Dalam penelitian ini dilakukan evaluasi kinerja lima metode *ensemble* yaitu *Naive Bayes* (NB), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Gaussian Process Classifier* (GPC), dan *Decision Tree* (DC). Visualisasi hasil eksperimen dapat ditemukan dalam Tabel 1, yang menyajikan metrik evaluasi seperti akurasi, presisi, *recall*, dan nilai *F-1 score* untuk setiap metode [31].

Tabel 1. Hasil Perbandingan Kinerja

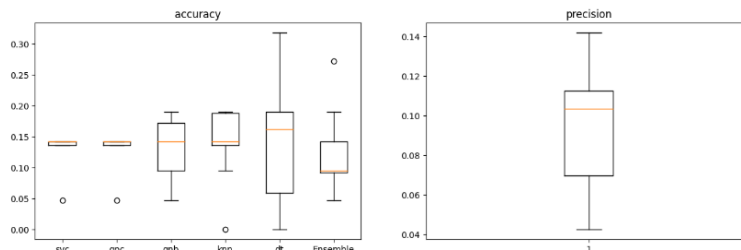
Metode Klasifikasi	Akura si	Presis i	Recal l	Nilai F-1 Score
Naive Bayes (NB)	0,13	0,11	0,13	0,12
Support Vector Machine (SVM)	0,13	0,04	0,13	<b>0,05</b>
K-Nearest Neighbors (KNN)	0,14	0,11	0,14	0,11
Gaussian Process Classifier (GPC)	0,13	0,06	0,13	0,06
Decision Tree (DT)	0,14	<b>0,15</b>	<b>0,15</b>	<b>0,15</b>
Ensemble Model	0,12	0,11	0,12	0,09

Pada metode *ensemble*, keputusan akhir dihasilkan melalui teknik voting di mana prediksi dari setiap model individu tiap algoritma dikumpulkan dan kemudian dianalisis. *Ensemble voting* ini menggunakan pendekatan mayoritas, di mana kelas yang paling sering diprediksi oleh model-model individu akan menjadi hasil prediksi akhir. Dengan demikian, keputusan akhir dari *ensemble* ini mengandalkan konsensus dari beberapa algoritma untuk memberikan hasil yang lebih stabil dan akurat dibandingkan hasil model tunggal.

Hasil penelitian menunjukkan metode NB, SVM, dan GPC untuk metrik akurasi memiliki nilai 0,13 sedangkan KNN dan DT memiliki nilai akurasi 0,14 dan juga merupakan nilai tertinggi, sedangkan *ensemble model* mendapatkan nilai terendah dengan nilai 0,12. Jika melihat tabel di atas metode DT merupakan metode yang cocok digunakan untuk klasifikasi pada dataset JAFFE ini dengan nilai metrik akurasi, presisi, *recall* dan *F-1 Score* tertinggi meskipun nilai nya juga masih terbilang rendah. Berbanding terbalik dengan *ensemble model* yang seharusnya memberikan evaluasi yang baik, nyatanya mendapatkan nilai yang cenderung rendah, bahkan pada akurasi dan *recall* merupakan nilai terendah dibandingkan dengan metode yang lain.

Ada beberapa yang menjadi penyebab sehingga nilai metrik evaluasi di setiap metode maupun *ensemble model* sangat rendah yaitu dalam hal variasi dataset dan metode eksperimen. Oleh karena itu, rekomendasi untuk penelitian selanjutnya mencakup peningkatan jumlah *dataset*, eksplorasi fitur-fitur tambahan, dan eksperimen lanjutan untuk mencari tahu faktor-faktor yang mempengaruhi kinerja metode klasifikasi. Implikasi praktis dari penelitian ini dapat mendukung pengembangan sistem pengenalan ekspresi wajah yang lebih efektif dalam berbagai aplikasi, termasuk dalam konteks kecerdasan buatan dan teknologi pengenalan emosi.

Penampilan visual dari hasil penelitian ini diperkuat dengan menggunakan *boxplot* yang secara detail diilustrasikan dalam Gambar 10. *Boxplot* tersebut memberikan representasi grafis yang jelas terkait distribusi dan variasi kinerja masing-masing metode klasifikasi, mencakup akurasi, presisi, *recall*, dan nilai *F-1 score*. Melalui Gambar 10, pembaca dapat dengan mudah mengidentifikasi sebaran data, melihat adanya potensi *outlier*, serta membandingkan konsistensi dan performa relatif dari setiap metode. Visualisasi ini tidak hanya memperkaya pemahaman pembaca tentang hasil eksperimen, tetapi juga memberikan informasi tambahan yang dapat digunakan untuk evaluasi lebih mendalam terhadap efektivitas metode-metode yang diujikan [32].



**Gambar 10.** Hasil Visualisasi Boxplot

## 5. KESIMPULAN

Secara singkat, hasil penelitian ini menunjukkan bahwa metode KNN dan DT memiliki performa yang unggul dalam konteks klasifikasi ekspresi wajah, dan ensemble vote tidak sepenuhnya memberikan peningkatan yang diharapkan. Penelitian ini memberikan gambaran bahwa tidak semua metode klasifikasi memberikan hasil yang optimal dalam tugas klasifikasi ekspresi wajah pada *dataset JAFFE*. Hanya ada beberapa metode yang efektif digunakan untuk klasifikasi pada *dataset JAFFE* yaitu *Decision Tree (DT)*, *K-Nearest Neighbors (KNN)*.

Rekomendasi untuk penelitian lebih lanjut mencakup peningkatan jumlah *dataset*, eksplorasi fitur-fitur tambahan, dan eksperimen lebih lanjut dengan parameter metode klasifikasi. Evaluasi lebih mendalam terhadap *ensemble vote* juga diperlukan untuk memahami faktor-faktor yang dapat meningkatkan kinerjanya. Analisis hasil menunjukkan bahwa *Decision Tree* dan *K-Nearest Neighbors* memberikan performa tertinggi diantara akurasi yang didapat mencapai 0,14, sedangkan metode *ensemble vote* justru menunjukkan hasil yang lebih rendah dengan akurasi 0,12. Hal ini mengindikasikan bahwa penggunaan *ensemble* tidak selalu efektif, terutama pada *dataset* kecil seperti *JAFFE*, yang dapat mempengaruhi stabilitas hasil dari kombinasi model. Implikasi praktis dari penelitian ini dapat mendukung pengembangan sistem pengenalan ekspresi wajah yang lebih efektif dalam berbagai aplikasi. Dengan demikian, penelitian mendatang dapat memperluas pemahaman kita tentang klasifikasi ekspresi wajah dan menerapkan temuan ini dalam konteks praktik nyata.

## DAFTAR PUSTAKA

- [1] A. Rahmawati, Y. Rianto, and D. Riana, "Deteksi Defect Coffee Pada Citra Tunggal Green Beans Menggunakan Metode Ensemble Decision Tree.," *Techno. com*, vol. 20, no. 2, 2021.
- [2] N. Rismayanti and A. P. Utami, "Improving Multi-Class Classification on 5-Celebrity-Faces Dataset using Ensemble Classification Methods," *Indones. J. Data ...*, vol. 4, 2023, doi: 10.56705/ijodas.v4i2.78.
- [3] M. I. Aziz, A. Z. Fanani, and A. Affandy, "Analisis Metode Ensemble Pada Klasifikasi Penyakit Jantung Berbasis Decision Tree.," *J. Media Inform. Budidarma*, vol. 7, no. 1, pp. 1–12, 2023.
- [4] H. Azis, "Assessing the Performance of Logistic Regression in Heart Disease Detection through 5-Fold Cross-Validation," *Int. J. Artif. Intell. ...*, 2024, [Online]. Available:

<https://jurnal.yoctobrain.org/index.php/ijaimi/article/view/137>

- [5] V. Yevsieiev, S. Maksymova, and A. Abu-Jassar, "The Canny Algorithm Implementation for Obtaining the Object Contour in a Mobile Robot's Workspace in Real Time," 2024.
- [6] B. S. W. Poetro, E. Maria, H. Zein, and ..., "Advancements in Agricultural Automation: SVM Classifier with Hu Moments for Vegetable Identification," *Indones. J. ...*, vol. 5, 2024, doi: 10.56705/ijodas.v5i1.123.
- [7] L. Hablinawati and A. A. Dzikrullah, "Analisis Sentimen Pengguna Twiter terhadap Perubahan Kebijakan Skripsi sebagai Syarat Wajib Kelulusan menggunakan Metode Naïve Bayes Classifier," *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 3, pp. 1429–1439, 2024.
- [8] N. A'yunnisa, Y. Salim, and H. Azis, "Analisis performa metode Gaussian Naïve Bayes untuk klasifikasi citra tulisan tangan karakter arab," *Indones. J. Data ...*, 2022, [Online]. Available: <https://jurnal.yoctobrain.org/index.php/ijodas/article/view/54>
- [9] H. A. D. Rani, S. Zuhri, and S. Fuji, "Sistem Prediksi Kondisi Kelahiran Bayi Menggunakan Klasifikasi Naïve Bayes," *Joined J. (Journal Informatics Educ.*, vol. 3, no. 2, pp. 48–56, 2020.
- [10] L. Priyambodo *et al.*, "Klasifikasi Kematangan Tanaman Hidroponik Pakcoy Menggunakan Metode SVM," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 153–160, 2022.
- [11] H. Azis, L. Syafie, F. Fattah, and ..., "Unveiling Algorithm Classification Excellence: Exploring Calendula and Coreopsis Flower Datasets with Varied Segmentation Techniques," *2024 18th Int. ...*, 2024, doi: 10.1109/IMCOM60618.2024.10418246.
- [12] R. Maneno, B. Baso, P. G. Manek, and K. Fallo, "Deteksi Tingkat Kematangan Buah Pinang Menggunakan Metode Support Vector Machine Berdasarkan Warna Dan Tekstur," *J. Inf. Technol.*, vol. 3, no. 2, pp. 60–66, 2023.
- [13] D. Cahyanti, A. Rahmayani, and ..., "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indones. J. ...*, 2020, [Online]. Available: <https://www.jurnal.yoctobrain.org/index.php/ijodas/article/view/13>
- [14] A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree," *JATISI (Jurnal Tek. Inform. Dan Sist. Informasi)*, vol. 9, no. 2, pp. 1109–1119, 2022.
- [15] I. G. I. Sudipa, R. A. Azdy, I. Arfiani, and ..., "Leveraging K-Nearest Neighbors for Enhanced Fruit Classification and Quality Assessment," *Indones. J. ...*, 2024, [Online]. Available: <https://www.jurnal.yoctobrain.org/index.php/ijodas/article/view/125>
- [16] P. Putra, A. M. H. Pardede, and S. Syahputra, "Analisis Metode K-Nearest Neighbour (Knn) Dalam Klasifikasi Data Iris Bunga," *JTIK (Jurnal Tek. Inform. Kaputama)*, vol. 6, no. 1, pp. 297–305, 2022.
- [17] L. Yang, C. Heiselman, J. G. Quirk, and P. M. Djurić, "Class-imbalanced classifiers using ensembles of Gaussian processes and Gaussian process latent variable models," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 3775–3779.
- [18] Y. Chen, Q. Da, W. Liang, P. Xiao, B. Dai, and G. Zhao, "Bagged ensemble of gaussian process classifiers for assessing rockburst damage potential with an imbalanced dataset," *Mathematics*, vol. 10, no. 18, p. 3382, 2022.
- [19] M. Solahuddin, A. I. Purnamasari, and A. R. Dikananda, "Klasifikasi Kualitas Berita Pada Majalah Menggunakan Metode Decision Tree," *J. Teknol. Ilmu Komput.*, vol. 1, no. 2, pp. 48–54, 2023.

- [20] L. N. Rachmadi, A. P. Wibawa, and U. Pujiyanto, "Rekomendasi Jurusan dengan Menggunakan Decision Tree pada Sistem Penerimaan Peserta Didik Baru SMK Widya Dharma Turen," *Belantika Pendidik.*, vol. 4, no. 1, pp. 29–36, 2021.
- [21] K. R. Bhatele and S. S. Bhadauria, "Multiclass classification of central nervous system brain tumor types based on proposed hybrid texture feature extraction methods and ensemble learning," *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3831–3858, 2023.
- [22] B. Bramantyo, M. P. K. Putra, and N. Hendrastuty, "Implementasi Recurrent Neural Network Pada Multiclass Text Classification Judul Berita," *J. Media Borneo*, vol. 1, no. 1, pp. 1–11, 2023.
- [23] A. Faradibah, D. Widyawati, A. U. T. Syahar, and ..., "Comparison Analysis of Random Forest Classifier, Support Vector Machine, and Artificial Neural Network Performance in Multiclass Brain Tumor Classification," *Indones. J. ...*, 2023, [Online]. Available: <https://www.jurnal.yoctobrain.org/index.php/ijodas/article/view/73>
- [24] T. I. Z. M. Putra, S. Suprpto, and A. F. Bukhori, "Model Klasifikasi Berbasis Multiclass Classification dengan Kombinasi Indobert Embedding dan Long Short-Term Memory untuk Tweet Berbahasa Indonesia," *J. Ilmu Siber dan Teknol. Digit.*, vol. 1, no. 1, pp. 1–28, 2022.
- [25] M. Harim, H. A. T. Muslimin, and R. A. D. I. SAPUTRA, "Segmentasi Citra Telapak Tangan menggunakan Deteksi Tepi Prewitt, Sobel, Roberts, dan Canny," *JIMP-Jurnal Inform. Merdeka Pasuruan*, vol. 8, no. 1, pp. 9–16, 2023.
- [26] F. N. Cahya and R. Pebrianto, "Klasifikasi Buah Segar dan Busuk Menggunakan Ekstraksi Fitur Hu-Moment, Haralick dan Histogram," *J. Khatulistiwa Inform.*, vol. 6, no. 1, p. 490852, 2021.
- [27] R. Setiawan, A. M. A. K. Parewe, A. J. Latipah, N. R. D. P. Astuti, A. W. Murdiyanto, and F. P. Putra, "Assessing Bagging-meta Estimator in Imbalanced CT Kidney Disease Classification: A Focus on Sobel and Hu Moment Techniques," *Int. J. Artif. Intell. Med. Issues*, vol. 1, no. 2, pp. 64–73, 2023.
- [28] F. Aziz and B. L. E. Panggabean, "Klasifikasi Nasabah Potensial menggunakan Algoritma Ensemble Least Square Support Vector Machine dengan AdaBoost," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 269–274, 2023.
- [29] F. Dwiramadhan, M. I. Wahyuddin, and D. Hidayatullah, "Sistem Pakar Diagnosa Penyakit Kulit Kucing Menggunakan Metode Naive Bayes Berbasis Web," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 3, pp. 429–437, 2022.
- [30] É. Bédard, V. D. B. de Vazelhes, and G. Beaudoin, "Performance of predictive supervised classification models of trace elements in magnetite for mineral exploration," *J. Geochemical Explor.*, vol. 236, p. 106959, 2022.
- [31] M. Prameswari, P. E. Kania, I. G. De Ayu, and S. N. P. Harnoko, "Penerapan Metode Stacking Ensemble Untuk Klasifikasi Status Pinjaman Nasabah Bank," in *PROSIDING SEMINAR NASIONAL SAINS DATA*, 2024, pp. 802–811.
- [32] G. Antariksa, R. Muammar, and J. Lee, "Performance evaluation of machine learning-based classification with rock-physics analysis of geological lithofacies in Tarakan Basin, Indonesia," *J. Pet. Sci. Eng.*, vol. 208, p. 109250, 2022.