



Komparasi Algoritma Nonparametrik untuk Klasifikasi Citra Wajah Berdasarkan Suku di Indonesia

Seno Hartono^{#1}, Herry Sujaini^{#2}, Anggi Perwitasari^{#3}

[#]Program Studi Informatika Fakultas Teknik Universitas Tanjungpura
Jl. Prof. Dr. H. Hadari Nawawi, Kota Pontianak, 78115

¹senohartono@student.untan.ac.id

²hs@untan.ac.id

³anggi.perwitasari@informatika.untan.ac.id

Abstrak— Klasifikasi merupakan metode *data mining* yang berfungsi untuk mengatur dan mengkategorikan data pada kelas yang berbeda-beda. Penelitian ini bertujuan untuk membandingkan dan menentukan algoritma nonparametrik terbaik dalam pengklasifikasian citra wajah. Dalam proses pengklasifikasian, penelitian ini menggunakan algoritma klasifikasi nonparametrik yaitu *k-Nearest Neighbor* (kNN), *Support Vector Machine* (SVM), *Decision Tree*, dan *AdaBoost* Untuk mengklasifikasikan citra wajah penduduk Indonesia yang berasal dari suku Batak, Dayak, Jawa, Melayu, dan Tionghoa. Penelitian ini menggunakan Orange Data Mining Tool sebagai alat bantu untuk melakukan proses *data mining*. Dari hasil pengklasifikasian dengan menerapkan algoritma *k-Nearest Neighbor*, *Support Vector Machine*, *Decision Tree*, dan *AdaBoost*, SVM memberikan nilai akurasi yang lebih baik dibanding algoritma lainnya. Rata-rata nilai *precision* keempat algoritma tersebut berturut-turut adalah *Support Vector Machine* 37.5%, diikuti oleh algoritma *k-Nearest Neighbor* 31.55%, *AdaBoost* 30.25%, dan untuk *Decision Tree* 29.75%.

Kata kunci— Klasifikasi, K-Nearest Neighbor, Support Vector Machine, Decision Tree, AdaBoost

I. PENDAHULUAN

Pengenalan wajah merupakan Salah satu cara alami bagi manusia untuk mengenali dan membedakan identitas suatu individu dari yang lainnya. Wajah manusia dapat memberikan berbagai informasi seperti jenis kelamin, umur, ekspresi, suku bangsa dan sebagainya. Penelitian tentang klasifikasi ras, negara, dan suku berdasarkan citra wajah menjadi topik yang populer diantara komunitas peneliti *face recognition* [1]. Dalam penerapannya pengklasifikasian citra dilakukan dengan mengimplementasikan algoritma klasifikasi untuk membedakan fitur-fitur yang dimiliki gambar digital.

Klasifikasi adalah proses untuk menemukan model atau pengetahuan untuk membedakan kelas data dengan memprediksikan kelas untuk data yang tidak diketahui

kelasnya [2]. Klasifikasi bertugas untuk mengatur dan mengkategorikan data pada kelas yang berbeda-beda.

Secara umum metode klasifikasi dibagi menjadi dua macam yaitu klasifikasi *supervised* dan *unsupervised*. Metode klasifikasi *supervised* adalah metode *learning* melalui bantuan dari supervisor (*learning* melalui contoh). Sedangkan metode klasifikasi *unsupervised* merupakan kebalikan dari metode *supervised* yakni *learning* tanpa supervisor (*learning* melalui observasi). Klasifikasi *unsupervised* dikenal juga sebagai *clustering*.

Algoritma klasifikasi citra parametrik dan nonparametrik berada dibawah *supervised learning*. Algoritma parametrik biasanya disertai dengan sejumlah asumsi tertentu, yang sepenuhnya telah ditentukan sebelum proses klasifikasi berlangsung. Misalnya sebagian *classifier* berasumsi bahwa data bersifat homogen atau terdistribusi berdasarkan pola tertentu. Karena itulah algoritma-algoritma ini dinilai memiliki keterbatasan ketika asumsi yang digunakan tidak sesuai dengan keadaan atau kondisi data [3].

Algoritma klasifikasi non-parametrik dinilai lebih siap untuk menghadapi data dengan berbagai kondisi. Algoritma klasifikasi non-parametrik memiliki asumsi dasar yang relatif lebih sedikit dibandingkan algoritma klasifikasi parametrik. Dalam proses klasifikasi, metode klasifikasi non-parametrik akan mempelajari dan menggunakan setiap fitur dimiliki sampel sebagai dasar untuk penunjukan kelas [3].

Beberapa penelitian tentang klasifikasi citra telah dilakukan sebelumnya. Perbedaan utama dari setiap penelitian pada umumnya terletak pada dataset dan metode klasifikasi yang digunakan.

Penelitian yang dilakukan oleh Laina Farsiah, Taufik Fuadi Abidin, dan Khairul Munadi, 2013, "Klasifikasi Gambar Berwarna Menggunakan *k-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM)". Penelitian ini membahas tentang algoritma penentuan kategori dari gambar berwarna menggunakan metode *k-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM).

Hasil pengujian menghasilkan nilai akurasi yang sama baiknya antara metode KNN dan SVM.

Penelitian yang dilakukan oleh Khairul Sani, Wing Wahyu Winarno, dan Silmi Fauziati, 2016, "Analisis Perbandingan Algoritma *Classification* Untuk *Authentication* Uang Kertas". Tahapan penelitian ini dimulai dengan pengumpulan data, lalu melakukan proses pengenalan *banknote* yang terdiri dari proses *image acquisition* dilanjutkan dengan proses konversi *grayscale*, hingga sampai tahap klasifikasi. Penelitian ini menggunakan aplikasi WEKA dengan metode *cross validation* terhadap data *banknote authentication*. Hasil pengujian memberikan algoritma *Decision Tree C4.5* sebagai algoritma terbaik dibanding dengan algoritma *naive bayes* dan *neural network*.

Penelitian yang dilakukan oleh Herry Sujaini, 2019, "Sistem Analisis Citra Alat Musik Tradisional dengan Metode *k-Nearest Neighbor*, *Random Forest*, dan *Support Vector Machine*". Masalah utama yang dikaji dalam penelitian ini adalah banyaknya perbedaan dalam kesimpulan dari hasil penelitian sebelumnya dalam menentukan algoritma klasifikasi yang terbaik. Penelitian ini membandingkan tiga buah metode nonparametrik yaitu Metode *K-Nearest Neighbor* (kNN), *Random Forest* (RF), dan *Support Vector Machine* (SVM). Hasil pengujian memberikan metode kNN sebagai metode yang paling baik dalam mengenali alat musik. Rata-rata nilai *precision* yang diperoleh dari ketiga algoritma yang dibandingkan adalah KNN dengan 92,1%, SVM dengan 85,4%, dan 69,4% untuk *Random Forest*.

Berdasarkan dari paparan hasil penelitian-penelitian sebelumnya, penelitian ini akan membandingkan algoritma yang memberikan hasil terbaik pada penelitian sebelumnya yaitu *K-Nearest Neighbor* (KNN), *Decision Tree*, *Support Vector Machine* (SVM), dan *Adaboost*.

Indonesia adalah negara kesatuan, karena itu Indonesia memiliki beragam suku bangsa, agama, bahasa, dan lain sebagainya. Menurut *Ensiklopedia Indonesia* Suku memiliki arti sebagai kelompok sosial yang memiliki atau mempunyai kedudukan tertentu yang disebabkan oleh faktor garis keturunan, adat, agama, bahasa, dan lain sebagainya. Biasanya setiap anggota suku itu mempunyai kesamaan dalam hal sejarah, bahasa, adat istiadat serta juga dalam tradisi. Setiap suku di Indonesia memiliki karakteristik wajah yang beragam dan memiliki keunikan masing-masing [2].

Pada penelitian ini dilakukan performa komparasi algoritma klasifikasi nonparametrik dalam mengklasifikasi citra wajah penduduk Indonesia yang berasal dari suku Jawa, Batak, melayu, Dayak dan Tionghoa. Metode yang digunakan yaitu *K-Nearest Neighbor* (KNN), *Decision Tree* *Decision tree* yang sangat kuat sehingga banyak diminati [4], *Support Vector Machine* (SVM), dan *Adaboost*.

KNN merupakan algoritma yang mampu memberikan hasil akurasi yang tinggi karena algoritma KNN bersifat fleksibel serta mampu bekerja dengan berbagai batasan keputusan [5], sedangkan pada SVM bekerja dengan

menemukan batas pemisah terbesar dengan mencari jarak maksimal dari keseluruhan titik data, sehingga apabila terjadi kesalahan pengukuran maka tidak menyebabkan kesalahan hasil kategori dan memberikan hasil klasifikasi dengan akurasi yang tinggi [6]. *Adaboost* yang merupakan algoritma *machine learning* yang tercipta dengan menggunakan kumpulan *weak classifier* yang dibentuk sehingga menjadi menjadi *strong classifier* [7] yang dimunculkan untuk mampu mengurangi kesalahan dalam proses pembelajaran [8].

II. TINJAUAN PUSTAKA

A. Citra Digital

Citra digital adalah sebuah gambar yang direpresentasikan secara digital dalam bentuk fungsi 2 dimensi, $f(x, y)$, yang terdiri dari x dan y sebagai koordinat horizontal dan vertikal, dan f sebagai nilai intensitas atau derajat keabuan dari x dan y . Setiap nilai dinyatakan dalam skala tertentu berdasarkan citra yang ditampilkan. Citra merupakan fungsi dua dimensi yang digambarkan oleh perangkat keras (komputer). Citra yang seharusnya berwujud tiga dimesi kemudian digambarkan dalam bentuk dua dimensi yang berupa titik data diskrit akibat ketidak mampuan computer dalam memproses gambar yang sebenarnya berwujud tiga dimensi. Titik yang ditampilkan disebut juga dengan piksel [9].

B. Suku

Suku bangsa merupakan kelompok manusia atau kelompok sosial dengan kesamaan ciri, budaya, suku bangsa berkaitan erat dengan asal-usul, tempat asal dan kebudayaan. Di Indonesia, wilayah yang terbentuk atas kepulauan dan lautan tentu mempengaruhi keanekaragaman budaya dan suku bangsa Indonesia.

Setiap suku bangsa mempunyai keunikan dan ciri-ciri tertentu. Suku bangsa memiliki ciri fisik, adat istiadat, bahasa, kebudayaan, dan kesenian. Warna kulit di Indonesia cenderung berwarna sawo matang, bentuk atau jenis rambut cenderung lurus dan ikal, dan bentuk wajah yang bermacam-macam, merupakan pembeda satu suku bangsa dengan suku bangsa lainnya [2].

C. Data Mining

Data mining adalah sebuah proses untuk menganalisa kumpulan data menggunakan beberapa bidang ilmu yaitu *machine learning*, teknik statistik, pengenalan pola, kecerdasan buatan, dan visualisasi. *Data mining* bertujuan untuk menemukan dan menghasilkan sebuah informasi dari himpunan data yang sangat banyak sehingga menghasilkan pengetahuan (*knowledge*) yang bermanfaat. [10].

D. Orange Data Mining Tools

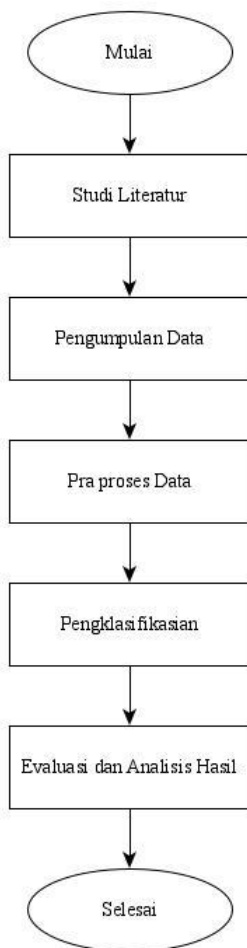
Data mining tools yang digunakan untuk membantu dalam penelitian ini adalah Orange. Orange merupakan sebuah *tools open source* untuk pengolahan data mining. Orange memiliki *widget* yang berfungsi sebagai unit

komputasi untuk membaca, memproses, melakukan visualisasi, melakukan analisis, mengeksplorasi data, dan lain-lain. *Widget* disusun sedemikian rupa sehingga membentuk workflow (alur kerja) dan berkomunikasi satu sama lain di lingkungan Orange. *Widget* data memungkinkan Orange untuk manipulasi data text atau gambar[11].

Dalam menganalisis gambar menggunakan Orange, diperlukan sebuah *add-ons* tambahan yaitu *image analytics*. Orange mengubah data gambar menjadi representasi vektor menggunakan *deep neural network* yang telah dilatih pada banyak gambar sehingga menghasilkan data yang dapat diproses dan memungkinkan *machine learning* [11].

III. METODOLOGI

Metodologi penelitian adalah suatu kegiatan untuk merencanakan penelitian yang akan dilakukan dengan cara sistematis dan ilmiah. Penelitian ini menggunakan metode penelitian komparatif. Metode analisis yang akan dipakai adalah metode *data mining* yaitu klasifikasi dengan menggunakan *tools software* Orange. Penelitian ini dilakukan dengan beberapa tahapan yang diilustrasikan pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

A. Studi Literatur

Langkah pertama yang dilakukan adalah studi literatur yaitu mencari referensi dan pengetahuan dari berbagai sumber seperti buku, jurnal dan artikel yang berhubungan dengan penelitian yang akan dilakukan. Tujuan utama studi literatur dalam penelitian ini adalah untuk mempelajari sumber-sumber terdahulu sebagai bahan panduan untuk mendukung penelitian ini. Studi Literatur dimulai dengan mempelajari penelitian sebelumnya yang terkait tentang *data mining*, pengklasifikasian, Algoritma KNN, SVM, Decision Tree, dan Adaboost, Kemudian mempelajari cara evaluasi data menggunakan k-fold Cross Validation dan confusion matrix.

B. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dengan melakukan wawancara dan penelusuran internet. Wawancara adalah sebuah proses tanya jawab antara dua orang atau lebih untuk mendapatkan informasi tertentu. Proses tanya jawab itu dilakukan oleh dua pihak, yang pertama pewawancara (*interviewer*) yaitu orang yang memberikan pertanyaan dan responden yang memberikan jawaban atas pertanyaan yang diberikan [12]. Melalui wawancara ini peneliti akan menggali informasi dan mendapatkan data berupa foto (*image*) dari wajah responden yang nantinya akan digunakan dalam proses klasifikasi. Wawancara dilakukan kepada masyarakat sekitar atau pendatang di kota Pontianak yang berasal dari suku yang terkait dengan penelitian yaitu suku Jawa, Batak, Melayu, Dayak, dan Tionghoa.

Secara Keseluruhan, koleksi data yang dikumpulkan berjumlah 500 buah citra wajah yang terdiri dari 5 kategori suku. Masing masing suku diwakili dengan 100 buah citra. Koleksi gambar yang telah dikumpulkan akan dibagi berdasarkan kelompok sukunya.

C. Framework Penelitian

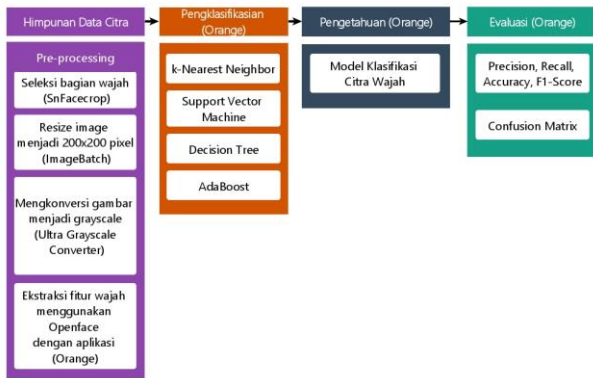
Tahapan-tahapan penelitian yang dilakukan terdiri dari pre-processing, pengklasifikasian, dan evaluasi yang dilakukan menggunakan bantuan beberapa aplikasi yaitu SnFacecrop, ImageBatch, Ultra Grayscale Converter, dan aplikasi Orange. Tahapan-tahapan ini diilustrasikan pada Gambar 2.

D. Pengujian

Dalam penelitian ini pengujian dilakukan sebagai tahap penerapan dalam membandingkan tingkat keakuratan algoritma yang diuji. Adapun pengujian dalam penelitian ini adalah dengan melakukan perhitungan *Precision*, *recall*, *F1-Score* dan *accuracy*. Pengujian dilakukan menggunakan *widget Test and score* pada aplikasi Orange.

Metode uji coba yang diterapkan dalam penelitian ini adalah *10-fold cross validation* dan *random sampling* menggunakan 75% data latih. Analisa akan dilakukan terhadap nilai rata-rata dari hasil perhitungan menggunakan kedua metode ini.

IV. HASIL DAN ANALISIS



Gambar 2. Framework penelitian

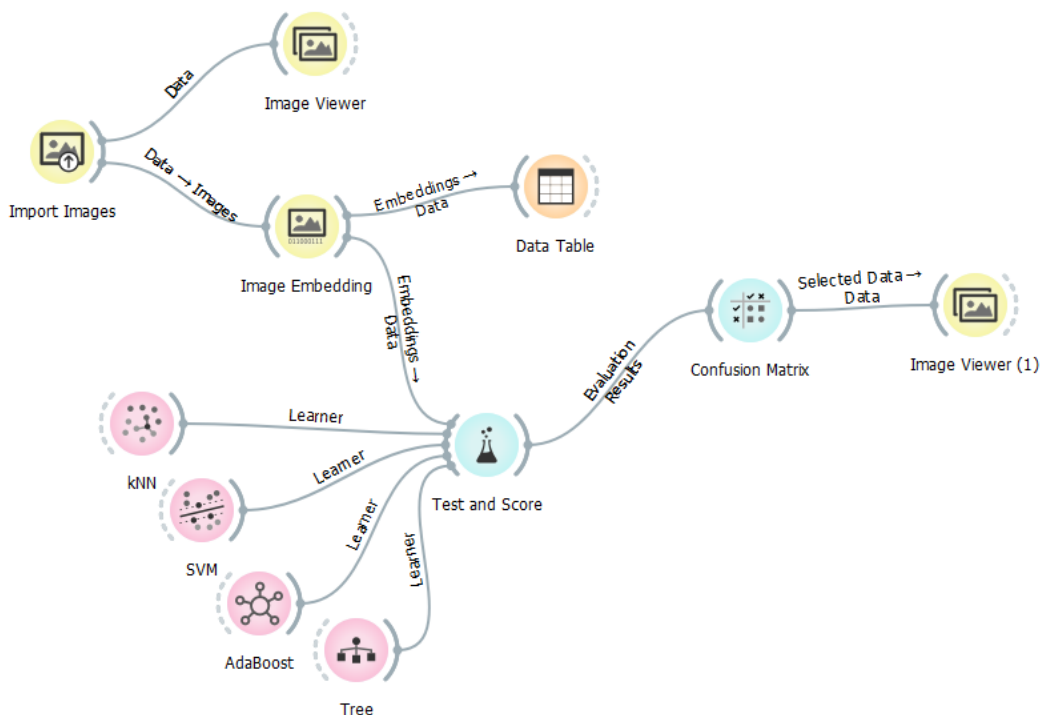
A. Hasil Penelitian

Dari perancangan *workflow* menggunakan Orange Data Mining Tools, dihasilkan sebuah model pengklasifikasi yang dapat mengklasifikasikan citra menggunakan algoritma *K-Nearest Neighbors*, *Support Vector Machine*, *Decision Trees* dan *Adaboost* yang bisa diamati pada Gambar 3.

B. Implementasi Pelatihan Data

1) Dataset

Dataset yang digunakan dalam penelitian ini berupa file gambar yang didapat dengan melakukan penelusuran internet melalui google dan media sosial facebook dan instagram dan wawancara terhadap masyarakat sekitar atau pendatang di kota Pontianak yang berasal dari suku yang terkait dengan penelitian.



Gambar 3. Workflow klasifikasi citra wajah

Setelah dilakukan penelusuran internet dan wawancara, diperoleh data awal dari suku Batak berjumlah 131 file, dari suku Dayak berjumlah 114 file, suku Jawa berjumlah 123 file, suku Melayu berjumlah 118 file, dan suku Tionghoa berjumlah 108 file. Sehingga total seluruh data awal dalam penelitian ini berjumlah sebanyak 594 file gambar yang

kemudian akan dikategorikan dan diberi label berdasarkan masing-masing suku.

2) Implementasi Pre-preprocessing

Dalam penelitian ini terdapat beberapa langkah *pre-processing* yaitu:

- Melakukan seleksi dan *crop* pada bagian wajah pada file gambar menggunakan aplikasi *sfacecrop*.
- Melakukan konversi gambar dari RGB ke *Grayscale* yang dilakukan menggunakan aplikasi *Ultra Grayscale Converter*. Citra yang awalnya berbentuk RGB akan ditransformasikan terlebih dahulu menjadi bentuk aras keabuan atau *grayscale* dengan nilai dari 0- 255 [13].
- Melakukan *resize* terhadap semua file gambar menjadi 200x200 pixel menggunakan aplikasi *Imagebatch*.

3) *Implementasi Ekstraksi Fitur*

Ekstraksi fitur dilakukan untuk mendapatkan vektor representasi dari fitur terpenting pada bagian wajah dari setiap citra. Ekstraksi fitur dilakukan dengan *widget image embedding* dan menggunakan *embedder OpenFace*.

C. Hasil Implementasi Algoritma

1) *Implementasi Menggunakan K-Nearest Neighbor*

Pada perhitungan menggunakan algoritma *K-nearest neighbour* digunakan nilai *number of neighbors = 5*, hal ini dikarenakan jika nilai *k* terlalu kecil maka akan semakin mudah terpengaruh oleh data *noise* dan nilai *k* yang terlalu besar akan membuat keakuratan menjadi berkurang.

2) *Implementasi Menggunakan Support Vector Machine*

Eksperimen dengan menggunakan SVM, untuk menghindari *overfitting* maka digunakan nilai *cost* kecil yaitu *C = 1*. *Regression loss epsilon = 0,1* fungsinya adalah untuk mendefinisikan *lost function* pada jarak tertentu dari nilai sebenarnya sehingga nilai yang dimasukkan kecil.

3) *Implementasi Menggunakan Decision Tree*

Eksperimen menggunakan *Decision Tree* Untuk mendapatkan hasil yang optimal jumlah minimal instances in leaves = 2, tidak membagi subset kurang dari 5, max tree depth = 100.

4) *Implementasi Menggunakan Adaboost*

Eksperimen menggunakan *Adaboost*, digunakan parameter *base estimator = Tree*, estimator yang digunakan adalah default dengan jumlah *estimator = 50*, jika sudah ditemukan perfect fit, maka prosedur learning akan berhenti lebih awal sebelum mencapai 50.

D. Pengujian

1) *Pengujian 10-fold cross validation*

Hasil perhitungan *10-fold cross validation* terhadap kelas suku menggunakan algoritma *K-Nearest Neighbour, Support Vector Machine, Decision Tree*, dan *AdaBoost* dapat dilihat pada Tabel 1.

TABEL I
PERBANDINGAN PENGUJIAN 10-FOLD CROSS VALIDATION PADA MASING-MASING ALGORITMA

Algoritma	Precision	Recall	Accuracy	F1-Score
kNN	32.6%	32.6%	32.6%	32%
SVM	37%	37.6%	37.6%	37.1%
Decision Trees	27.8%	28.2%	28.2%	27.9%
Adaboost	33.7%	33.8%	33.8%	33.7%

2) *Pengujian menggunakan Random Sampling*

Pembagian *data training* dan *testing* tersebut dilakukan secara *random* atau acak dengan widget *Test and score*. Hasil perhitungan *random sampling* terhadap kelas suku menggunakan algoritma *K-Nearest Neighbour, Support Vector Machine, Decision Tree*, dan *AdaBoost* dapat diamati pada Tabel 2

TABEL II
PERBANDINGAN PENGUJIAN RANDOM SAMPLING PADA MASING-MASING ALGORITMA

Algoritma	Precision	Recall	Accuracy	F1Score
kNN	30.4%	31.3%	31.3%	30.4%
SVM	38%	38.5%	38.5%	38.2%
Decision Trees	31.7%	30.6%	30.6%	31%
Adaboost	29.4%	29.3%	29.3%	29.3%

E. Perbandingan 4 Algoritma

1) *Precision*

K-Nearest Neighbour mendapatkan nilai rata-rata precision 31.55%, *Support Vector Machine* mendapatkan nilai precision tertinggi diantara 4 algoritma yang dipilih, yaitu 37.5%, lalu *Decision Trees* 29.75%, dan yang terakhir adalah algoritma *AdaBoost* yang mendapatkan nilai precision 31.55%.

2) *Recall*

K-Nearest Neighbour mendapatkan nilai rata-rata *recall* 31.95%, *Support Vector Machine* mendapatkan nilai *recall* tertinggi diantara 4 algoritma yang dipilih, yaitu 38.05%, lalu *Decision Trees* 29.45%, dan *AdaBoost* yang mendapatkan nilai 31.55%.

3) *Accuracy*

K-Nearest Neighbour mendapatkan nilai rata-rata *Accuracy* 31.95%, *Support Vector Machine* mendapatkan nilai *Accuracy* tertinggi diantara 4 algoritma yang dipilih, yaitu 38.05%, lalu *Decision Trees* 29.4%, dan *AdaBoost* 31.55%.

4) *F1 score*

K-Nearest Neighbour mendapatkan nilai rata-rata *F1-Score* 31,2%, *Support Vector Machine* mendapatkan nilai tertinggi diantara 4 algoritma yang dipilih, yaitu 37.65%, lalu *Decision Trees* 29,45%, dan *AdaBoost* 31.5%.

F. Confusion Matrix

Confusion matrix memberikan informasi data aktual (actual) dan prediksi (predicted) berdasarkan klasifikasi yang dilakukan. Dengan Menggunakan *confusion matrix* kita dapat mengamati performa

setiap algoritma pada model klasifikasi yang dibuat berdasarkan jumlah data yang diprediksi dengan benar dan salah [14].

Gambar dibawah ini menampilkan *confusion matrix* dari hasil penelitian dengan 4 kombinasi nilai prediksi dan nilai aktual yang berbeda. Terdapat 4 istilah sebagai representasi hasil proses klasifikasi pada *confusion matrix*. Keempat istilah tersebut adalah *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)* dan *False Negative (FN)*.

Dengan menggunakan *confusion matrix*, citra yang tidak terklasifikasi dengan baik dari hasil pengklasifikasian bisa diamati pada gambar 4, 5, 6, dan 7.

		Predicted					
		Batak	Dayak	Jawa	Melayu	Tionghoa	Σ
Actual	Batak	33	27	10	17	13	100
	Dayak	17	40	14	9	20	100
	Jawa	25	27	19	16	13	100
	Melayu	25	28	15	22	10	100
	Tionghoa	8	25	6	12	49	100
Σ		108	147	64	76	105	500

Gambar 4. Hasil klasifikasi citra KNN

		Predicted					
		Batak	Dayak	Jawa	Melayu	Tionghoa	Σ
Actual	Batak	32	25	15	19	9	100
	Dayak	23	39	15	6	17	100
	Jawa	19	26	28	19	8	100
	Melayu	22	15	22	26	15	100
	Tionghoa	11	9	9	8	63	100
Σ		107	114	89	78	112	500

Gambar 5. Hasil klasifikasi citra SVM

		Predicted					
		Batak	Dayak	Jawa	Melayu	Tionghoa	Σ
Actual	Batak	24	22	23	14	17	100
	Dayak	22	24	18	17	19	100
	Jawa	29	20	21	18	12	100
	Melayu	24	17	22	22	15	100
	Tionghoa	15	16	8	11	50	100
Σ		114	99	92	82	113	500

Gambar 6. Hasil klasifikasi citra Tree

		Predicted					
		Batak	Dayak	Jawa	Melayu	Tionghoa	Σ
Actual	Batak	27	19	16	26	12	100
	Dayak	21	28	23	12	16	100
	Jawa	19	13	29	28	11	100
	Melayu	23	12	17	34	14	100
	Tionghoa	13	17	13	6	51	100
Σ		103	89	98	106	104	500

Gambar 7. Hasil klasifikasi citra AdaBoost

G. Analisis Pengujian

Hasil perhitungan *precision*, *recall*, *accuracy* dan F1 menunjukkan hasil yang konsisten dengan algoritma terbaik yaitu SVM, diikuti KNN, *AdaBoost* dan *Decision Tree*. SVM dirancang tidak sensitif terhadap noise sehingga nilai performa yang diperoleh dapat menjadi lebih tinggi [15]. Hal ini dibuktikan dengan *Support Vector Machine* yang memiliki nilai tertinggi dan dapat menangani data yang tidak seimbang lebih baik dibanding algoritma lainnya.

Dari informasi yang ditampilkan *confusion matrix* memperlihatkan bahwa terdapat banyak kesalahan dalam pengklasifikasian data. Pada hasil terbaik yang didapat dari algoritma *Support Vector Machine* dapat diamati bahwa satu kelas citra yang dapat diprediksi lebih baik dibanding yang lainnya yaitu kelas Tionghoa yang terprediksi sebanyak 60%. Hal ini dikarenakan wajah suku Tionghoa memiliki keunikan yang menjadi ciri khas dan berbeda dengan wajah dari suku lainnya.

Kelas citra yang paling sedikit terprediksi secara benar terdapat pada citra suku melayu. Hal ini dikarenakan pada suku melayu memiliki banyak kemiripan terhadap Dayak, Batak, dan Jawa dan begitu juga sebaliknya, sehingga banyak ditemukan hasil prediksi yang salah dan tertukar.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan dari hasil evaluasi dan analisis sistem klasifikasi citra wajar berdasarkan suku di Indonesia menggunakan algoritma K-Nearest Neighbour, Support Vector Machine, Decision Tree, dan AdaBoost, maka dapat disimpulkan bahwa:

1. Dalam pengklasifikasian citra wajah, sistem yang menggunakan algoritma Support Vector Machine memiliki akurasi yang lebih baik dibanding algoritma lainnya. Rata-rata nilai *precision* keempat algoritma tersebut secara berturut-turut adalah Support Vector Machine 37.5%, diikuti algoritma K-Nearest Neighbour dan AdaBoost dengan *precision* 31.55%, untuk Decision Tree 29.75%.
2. Data yang digunakan memiliki variasi yang tinggi, sehingga proses klasifikasi dengan algoritma yang diterapkan memiliki kelemahan dalam keakuratan hasil data, namun hasil yang didapatkan tidak

mempengaruhi keakuratan algoritma atau rumus yang diterapkan.

3. Keunggulan Support Vector Machine terdapat pada penggunaan kernel. Dengan memilih kernel yang sesuai dengan kondisi data yaitu kernel RBF, SVM dapat menangani data yang tidak seimbang lebih baik dan nilai performa yang diperoleh menjadi lebih tinggi dibanding algoritma lainnya.

B. Saran

Dalam Adapun hal-hal yang menjadi saran dalam rangka untuk mengembangkan dan menyempurnakan penelitian ini adalah sebagai berikut:

1. Menambah jumlah suku yang diteliti dan menggunakan kondisi data yang lebih baik dan seragam.
2. Mengembangkan hasil penelitian ini menjadi aplikasi yang dapat melakukan klasifikasi citra wajah secara otomatis..

REFERENSI

- [1] Fu, S., He, H., & Hou, Z.-G., "Learning race from face: A survey". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(12), 2483–2509, 2014.
- [2] Gonzales, Rafael C., Woods, Richard E., "Digital Image Processing", New Jersey: *Prentice-Hall, Inc.*, 2002.
- [3] Liu, B., "Web Data mining: Exploring Hyperlinks, Contents, dan Usage Data" Berlin: *Springer*, 2007.
- [4] Manning, C. D., Schütze, Hinrich & Raghavan, Prabhakar, "An Introduction to Information Retrieval". *Cambridge Univeristy Press*, 2009.
- [5] Maulana, S. M., Sabarudin, R., Nugraha, W., "Prediksi Ketepatan Kelulusan Mahasiswa Diploma dengan Komparasi Algoritma Klasifikasi". *Jurnal Sistem dan Informasi (JUSTIN)*, 2019.
- [6] Melalatoa, M. Junus, "Ensiklopedi suku bangsa di Indonesia". *Direktorat Sejarah dan Nilai Tradisional Direktorat Jenderal Kebudayaan*, Jakarta, 1995.
- [7] Ruan, C. a. R. Q. a. L. X., "Real adaboost feature selection for face recognition. In *Signal Processing*", 2010 *IEEE 10th International Conference*, IEEE, 2010
- [8] Syamani, "Komparasi Algoritma Non-Parametrik k-Nearest Neighbour Classifier Menggunakan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Multispektral Tutupan Lahan", Universitas Lambung Mangkurat, 2008.
- [9] Sujaini, H. 2019. "Sistem Analisis Citra Alat Musik Tradisional dengan Metode K-Nearest Neighbor, Random Forest, dan Support Vector Machine" *Jurnal Sistem Informasi Bisnis*, vol. 9, 2009.
- [10] Turban, E., "Decision Support Systems and Intelligent Systems Edisi Bahasa Indonesia Jilid 1" Yogyakarta: *Andi*, 2005.
- [11] Waliansyah, R. R., Fitriyah, C., "Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naive Bayes dan k-Nearest Neighbor (k-NN)". *Jurnal Edukasi dan Penelitian Informatika (Jepin)* Vol. 5 no. 2, 2019.
- [12] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. Mclachlan, A. Ng, B. Liu, P. S. Yu, Z. Z. Michael, S. David, and J. H. Dan, "Top 10 algorithms in data mining" *Knowl Inf Syst*, vol 14, 2007.
- [13] Yoav Freund Robert E. Schapire, "A Short Introduction to Boosting", *Journal of Japanese Society for Artificial Intelligence*, 1999.
- [14] Moleong, Lexy J. (2010), Metodologi penelitian kualitatif, Remaja Rosdakarya, Bandung.