



**FEATURE EXTRACTION UNTUK KLASIFIKASI PENGENALAN  
WAJAH MENGGUNAKAN SUPPORT VECTOR MACHINE DAN  
K-NEAREST NEIGHBOR**

Yoga Religia

Program Studi Teknik Informatika, STT Pelita Bangsa  
Korespondensi email: [yoga.religia@pelitabangsa.ac.id](mailto:yoga.religia@pelitabangsa.ac.id)

---

**Abstrak:**

*Feature Extraction* adalah teknik pengambilan ciri / *feature* dari suatu bentuk yang nantinya nilai yang didapatkan akan dianalisis untuk proses selanjutnya. Klasifikasi adalah proses untuk menyatakan suatu objek ke dalam salah satu kategori yang sudah didefinisikan sebelumnya. Dalam penelitian ini membahas tentang pengambilan ciri dari suatu *feature* dengan dataset yang berisi kumpulan foto wajah yang akan digunakan untuk klasifikasi pengenalan wajah pria atau wanita. *Dataset* yang digunakan merupakan *global dataset* yang diambil dari “<http://www.cs.umass.edu/lfw>”. Dari *Feature Extraction* yang dilakukan diperoleh 6 atribut *independent* dan 1 atribut *dependent*. Tujuan dari penelitian ini adalah untuk membandingkan *performance* dari 2 algoritma yaitu *Support Vector Machine* dan *k-Nearest Neighbor*. Dari percobaan yang sudah dilakukan menunjukkan nilai rata-rata akurasi dari *Support Vector Machine* sebesar 88.13% sedangkan nilai rata-rata akurasi dari *k-Nearest Neighbor* sebesar 84.40%.

---

**Informasi Artikel:**

Diterima : 22 Juni 2019  
Direvisi : 12 Juli 2019  
Dipublikasikan: 9 September 2019

---

**Kata kunci :**

*Feature Extraction*, Klasifikasi,  
*face recognition*,  
k-NN, SVM

## I. Pendahuluan

*Image processing* merupakan suatu teknik pengolahan citra / gambar yang mentransformasikan citra input menjadi citra output yang memiliki kualitas lebih baik dari citra yang diinputkan. Dalam *image processing* terdapat ilmu yang mempelajari mengenai pengenalan wajah atau dikenal dengan *face recognition*. *Face recognition* adalah metode yang dapat digunakan untuk berbagai kontrol dari akses keamanan, *artificial intelligence*, otentikasi identitas, dll. Kebutuhan *face recognition* semakin penting dengan meningkatnya kesadaran akan perlunya keamanan dan privasi [1]. Salah satu teknik yang dapat digunakan untuk pengenalan wajah yaitu dengan menggunakan *Feature Extraction* [2].

*Feature Extraction* merupakan suatu proses untuk mendapatkan nilai unik, penting dan tidak berduplikasi dalam citra biometrik dalam pengenalan individu [3]. Dalam penelitian ini *Feature Extraction* akan digunakan untuk memperoleh  $\mu$ , deviasi, *skewness*, energi (keseragaman), entropi, dan *smoothness* untuk dikategorikan kedalam kategori pria dan wanita. Dari ekstraksi yang diperoleh akan digunakan untuk klasifikasi. Klasifikasi merupakan teknik yang digunakan untuk menentukan item dari suatu dataset kedalam kategori atau kelas tertentu. Tujuan klasifikasi adalah memprediksi kelas target secara akurat disetiap kasus yang terdapat dalam data [4].

*Support Vector Machine* (SVM) merupakan algoritma yang bekerja secara *linier classifier* yang dapat digunakan untuk proses klasifikasi. SVM pernah digunakan untuk pengenalan ekspresi wajah menggunakan metode *Geometric and*

*Appearance Features*. Pengenalan wajah dilakukan dengan membandingkan ekspresi wajah netral dengan ekspresi wajah yang ekspresif. Percobaan dilakukan dengan *Cohn-Kanade database*, dimana dalam penelitian tersebut menghasilkan akurasi berkisar 95% [5].

*K-Nearest Neighbor* (k-NN) adalah algoritma klasifikasi yang melakukan pembelajaran berdasarkan jarak terdekat dari nilai k. Algoritma k-NN pernah dipakai untuk pengujian antara hubungan berat badan selama kehamilan, berat badan sebelum kehamilan, serta indek massa tubuh (BMI) yang berkaitan dengan keguguran kehamilan. Dengan memakai algoritma k-NN dapat diketahui mana kehamilan yang baik atau keguguran. Dari pengujian yang dilakukan diperoleh akurasi sekitar 95% [6].

Berdasarkan penelitian sebelumnya, penelitian ini akan menggunakan algoritma *Support Vector Machine* dan *k-Nearest Neighbor* untuk melakukan klasifikasi pengenalan wajah menggunakan *Feature Extraction*.

## II. Landasan Teori

### 2.1. Feature Extraction

*Feature Extraction* adalah salah satu cara yang dapat digunakan untuk mengenali suatu objek berdasarkan pada histogram khusus yang dimiliki objek tersebut. *Feature Extraction* bertujuan untuk melakukan perhitungan serta perbandingan yang dapat dipakai untuk klasifikasi suatu citra berdasarkan ciri-ciri histogram yang dimiliki [7]. Fitur yang dapat dikenali menggunakan histogram seperti: rerata entitas ( $\mu$ ), deviasi, *skewness*, energi (keseragaman), entropi, dan *smoothness*

(kehalusan). Adapun setiap komponen tersebut dapat dihitung dengan persamaan [8]:

- a. Rerata entitas (mu)

$$m = \sum_{i=0}^{L-1} i \cdot p(i)$$

- b. Deviasi

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (i - m)^2 p(i)}$$

- c. *Skewness*

$$\text{Skewness} = \sum_{i=0}^{L-1} (i - m)^3 p(i)$$

- d. Energi (keseragaman)

$$\text{energi} = \sum_{i=0}^{L-1} [p(i)]^2$$

- e. Entropi

$$\text{entropi} = - \sum_{i=0}^{L-1} p(i) \log_2(p(i))$$

- f. *Smoothness* (kehalusan)

$$R = 1 - \frac{1}{1 + \sigma^2}$$

## 2.2. Support Vector Machine

SVM atau *Support Vektor Machine* merupakan algoritma klasifikasi dengan tujuan untuk menemukan fungsi pemisah yang dapat memisahkan 2 set data dari 2 kelas yang berbeda [9]. Metode SVM bekerja berdasarkan prinsip RSM (*Structural Risk Minimization*) yang tujuannya untuk menemukan *hyperplane* terbaik dengan memisahkan dua buah kelas pada input space.

Cara kerja *Support Vektor Machine* adalah dengan prinsip *linier classifier*, lalu kemudian dikembangkan

agar bisa bekerja pada kasus non-linear dengan menggunakan konsep kernel pada ruang kerja berdimensi tinggi. Ada beberapa jenis kernel yang dapat digunakan yaitu: Linear, polinomial, RBF (*Radial Basis Function*) dan *Sigmoid*. Persamaan dari masing-masing kernel adalah sebagai berikut [10]:

- a. Linear

$$K(x_i, x_j) = x_i^T x_j$$

- b. Polinomial

$$K(x_i, x_j) = (y x_i^T x_j + r)^d, y > 0$$

- c. Radial Basis Function

$$K(x_i, x_j) = \exp(-y x_i - x_j^2), y > 0$$

- d. Sigmoid

$$K(x_i, x_j) = \tanh(y x_i^T x_j + r)$$

## 2.3. k-Nearest Neighbor

*k-Nearest Network* (k-NN) adalah algoritma klasifikasi yang bekerja berdasarkan *k instances* terdekat dengan *query instances* yang diberikan, kemudian melakukan pemilihan antara *k tetangga* terdekat untuk memperoleh keluaran label dari *query instances* [11]. k-NN menyimpan semua instances pada tempat yang sama, dimana *n* merupakan fitur *instances* yang telah didefinisikan sebelumnya. Matrik *distances* yang dipakai untuk mengukur jarak antara instances.

Pada pengukuran jarak dapat menggunakan perhitungan *Euclidean distance* dan *Manhattan distance*. Misalkan *x* dan *y* adalah 2 *instances* yang didefinisikan sebagai  $\langle f_1, f_2, \dots, f_n \rangle$ , maka dengan menggunakan *Euclidean* dan *Manhattan* jaraknya didefinisikan sebagai  $d_1(x, y)$  dan  $d_2(x, y)$ , sehingga persamaannya dapat ditulis :

- a. *Euclidean*

$$d1(x, y) = \sqrt{\sum_{i=1}^n |f_i(x) - f_i(y)|^2}$$

b. Manhattan

$$d2(x, y) = \sum_{i=1}^n |f_i(x) - f_i(y)|$$

Yang paling berpengaruh pada algoritma k-NN adalah nilai k, dan nilai k terbaik pada algoritma k-NN bergantung pada data yang digunakan. Dengan memakai optimasi parameter dapat ditentukan nilai K yang sesuai. Pada penelitian ini nilai K diset 3. Pada tahap *training* algoritma k-NN hanya melakukan penyimpanan vektor-vektor fitur serta klasifikasi dari data *training*.

### III. Metode Penelitian

Pada penelitian ini akan dilakukan tahapan-tahapan penelitian sebagai berikut:

- Dataset* yang digunakan merupakan dataset global yang diambil dari “www.cs.umass.edu/lfw“. Dari dataset global tersebut terdapat 5749 data, yang terdiri dari gambar pria dan wanita dari berbagai negara [12].
- Melakukan pre-processing pada data dengan teknik human selection untuk memperoleh 1000 data yang berkualitas baik.
- Setelah dilakukan *pre-processing*, kemudian dengan data gambar tersebut akan dilakukan *Feature Extraction* menggunakan Matlab untuk memperoleh Mu, deviasi, *skewness*, energy (keseragaman), entropi, dan *smoothness* untuk dikategorikan kedalam kategori pria dan wanita.
- Dari hasil *Feature Extraction* akan dilakukan klasifikasi dengan RapidMiner menggunakan algoritma SVM dan k-NN.

e. Pengujian dilakukan dengan membagi data kedalam data training dan data testing. Pengujian akan dilakukan sebanyak 5 kali dengan setiap kali pengujian membagi data *training* dan data *testing* dengan proporsi yaitu : 90%:10%, 80%:20%, 70%:30%, 60%:40%, 50%:50%.

f. Hasil pengujian adalah berupa akurasi, *precision*, *recall* dan AUC.

- Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual. Dapat ditentukan dengan persamaan [13]:

$$akurasi = \frac{TP + TN}{TP + TN + FP + FN} * 100$$

- Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Dapat ditentukan dengan persamaan [13]:

$$precision = \frac{TP}{TP + FP} * 100$$

- Recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Dapat ditentukan dengan persamaan [13]:

$$recall = \frac{TP}{TP + FN} * 100$$

### IV. Dataset yang digunakan

*Dataset* yang digunakan dalam penelitian ini merupakan *dataset* global yang diambil dari *Labeled faces in the wild* yang berupa data gambar wajah orang dari berbagai belahan dunia [12]. Data yang diambil terdiri dari dua jenis gambar yaitu, gambar pria dan wanita. Jumlah data yang diambil adalah sebanyak 5749 gambar. Kemudian dari sejumlah data tersebut dipilih sebanyak 500 gambar pria dan 500 gambar wanita

dengan menggunakan teknik *Human Selection*.

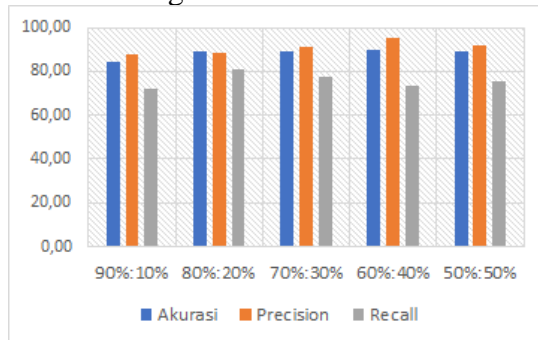
Setelah pemilihan data telah dilakukan, kemudian pada data yang sudah dipilih akan dilakukan ekstraksi fitur untuk memperoleh fitur diantaranya: Mu, deviasi, *skewness*, energi (keseragaman), entropi, dan *smoothness* untuk dikategorikan kedalam kategori pria dan wanita. Hasil ekstraksi fitur tersebutlah yang akan digunakan untuk klasifikasi pengenalan wajah.

## V. Hasil Pengujian

Pengujian dilakukan menggunakan 2 algoritma yaitu *Support Vector Machine* dan *k-Nearest Neighbor*. Hasil yang diperoleh akan digunakan untuk membandingkan *performance* dari kedua algoritma tersebut.

### 5.1. Hasil Pengujian SVM

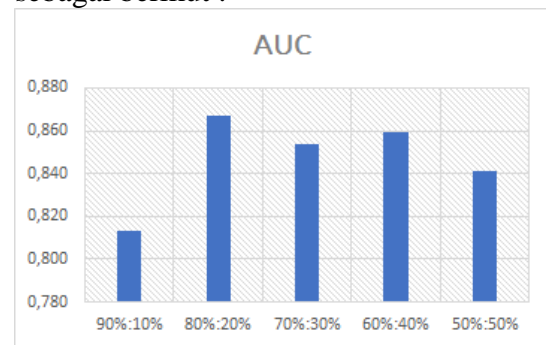
Berdasarkan hasil pengujian SVM menggunakan *tools* RapidMiner 5.3, maka diperoleh akurasi, *precision*, dan *recall* sebagai berikut :



Gambar 5.1. Hasil akurasi, precision dan recall SVM

Pada gambar 5.1 menunjukkan hasil dari pembagian data 90% *training* dan 10% *testing* memperoleh akurasi 84%, *precision* 87.89%, dan *recall* 42.31%. Hasil dari pembagian data 80% *training* dan 20% *testing* memperoleh akurasi 89%, *precision* 88.33%, dan *recall* 81.04%. Hasil dari pembagian data 70% *training* dan 30% *testing*

memperoleh akurasi 89.33%, *precision* 91.40%, dan *recall* 77.30%. Hasil dari pembagian data 60% *training* dan 40% *testing* memperoleh akurasi 89.50%, *precision* 95.19%, dan *recall* 73.35%. Hasil dari pembagian data 50% *training* dan 50% *testing* memperoleh akurasi 88.80%, *precision* 91.69%, dan *recall* 75.60%. Sedangkan hasil AUC menggunakan algoritma SVM adalah sebagai berikut :



Gambar 5.2. Hasil AUC SVM

Pada gambar 5.2 menunjukkan hasil dari pembagian data 90% *training* dan 10% *testing* memperoleh AUC 0.813. Hasil dari pembagian data 80% *training* dan 20% *testing* memperoleh AUC 0.867. Hasil dari pembagian data 70% *training* dan 30% *testing* memperoleh AUC 0.854. Hasil dari pembagian data 60% *training* dan 40% *testing* memperoleh AUC 0.859. Hasil dari pembagian data 50% *training* dan 50% *testing* memperoleh AUC 0.841. Adapun hasil pengujian secara keseluruhan menggunakan algoritma SVM dapat dilihat pada table 5.1:

Tabel 5.1. Hasil Pengujian SVM

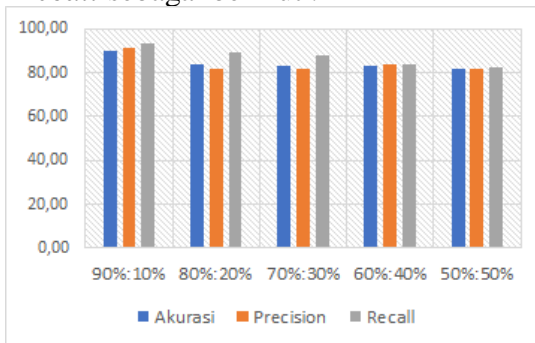
Training : Testing	Akurasi	Precision	Recall	AUC
90%:10%	84,00	87,89	72,31	0,813
80%:20%	89,00	88,33	81,04	0,867
70%:30%	89,33	91,40	77,30	0,854
60%:40%	89,50	95,19	73,35	0,859

50%:50%	88,80	91,96	75,60	0,841
---------	-------	-------	-------	-------

Dari table 5.1 dapat diperoleh rata-rata akurasi sebesar 88.13%, *precision* sebesar 90.95%, *recall* sebesar 75.92%, dan AUC sebesar 0.847.

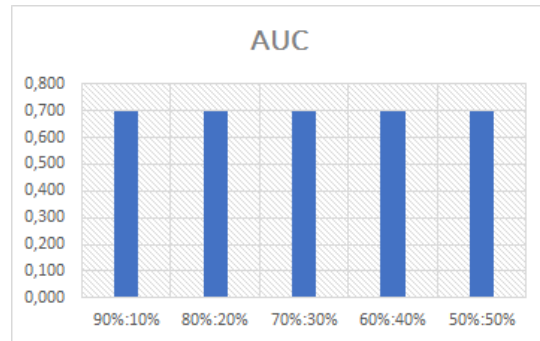
### 5.2. Hasil Pengujian k-NN

Berdasarkan hasil pengujian k-NN menggunakan *tools* RapidMiner 5.3, maka diperoleh akurasi, *precision*, dan *recall* sebagai berikut :



Gambar 5.3. Hasil akurasi, precision dan recall k-NN

Pada gambar 5.3 menunjukkan hasil dari pembagian data 90% *training* dan 10% *testing* memperoleh akurasi 90%, *precision* 91.11%, dan *recall* 93.46%. Hasil dari pembagian data 80% *training* dan 20% *testing* memperoleh akurasi 84%, *precision* 81.82%, dan *recall* 89.38%. Hasil dari pembagian data 70% *training* dan 30% *testing* memperoleh akurasi 83%, *precision* 82.12%, dan *recall* 88.11%. Hasil dari pembagian data 60% *training* dan 40% *testing* memperoleh akurasi 83%, *precision* 83.66%, dan *recall* 84.19%. Hasil dari pembagian data 50% *training* dan 50% *testing* memperoleh akurasi 82%, *precision* 81.97%, dan *recall* 82.80%. Sedangkan hasil AUC menggunakan algoritma k-NN adalah sebagai berikut :



Gambar 5.4. Hasil AUC k-NN

Pada gambar 5.4 menunjukkan hasil dari pembagian data 90% *training* dan 10% *testing* memperoleh AUC 0.7. Hasil dari pembagian data 80% *training* dan 20% *testing* memperoleh AUC 0.7. Hasil dari pembagian data 70% *training* dan 30% *testing* memperoleh AUC 0.7. Hasil dari pembagian data 60% *training* dan 40% *testing* memperoleh AUC 0.7. Hasil dari pembagian data 50% *training* dan 50% *testing* memperoleh AUC 0.7. Adapun hasil pengujian secara keseluruhan menggunakan algoritma k-NN dapat dilihat pada table 5.2:

Tabel 5.2. Hasil Pengujian k-NN

Training : Testing	Akurasi	Precision	Recall	AUC
90%:10%	90,00	91,11	93,46	0,70
80%:20%	84,00	81,82	89,38	0,70
70%:30%	83,00	82,12	88,11	0,70
60%:40%	83,00	83,66	84,19	0,70
50%:50%	82,00	81,97	82,80	0,70

Dari tabel 5.2 dapat diperoleh rata-rata akurasi sebesar 84.4%, *precision* sebesar 84.13%, *recall* sebesar 87.59%, dan AUC sebesar 0.70.

### 5.3. Analisa Hasil Pengujian

Dari pengujian yang sudah dilakukan menggunakan algoritma SVM dan k-NN dapat diperoleh hasil rata-rata yang dapat dilihat pada tabel 5.3 berikut:

Table 5.3. Hasil Rata-Rata Pengujian

	SVM	k-NN	Unggul
Akurasi	88.13	84.40	SVM
Precision	90.95	84.13	k-NN
Recall	85.92	87.59	k-NN
AUC	0.847	0.700	SVM

Dapat dilihat bahwa dari table 5.3 diperoleh algoritma SVM lebih unggul pada akurasi dan AUC, sedangkan algoritma unggul pada *precision* dan *recall*.

## VI. Kesimpulan

Berdasarkan pengujian yang sudah dilakukan, pada penelitian ini diperoleh hasil bahwa :

1. Akurasi SVM lebih tinggi dari pada k-NN dengan perbandingan 88.13% (SVM) banding 84.40% (k-NN).
2. Hasil untuk *precision* diperoleh k-NN lebih unggul dari SVM dengan perbandingan 84.13 (k-NN) banding 90.95 (SVM).
3. Hasil untuk *recall* menunjukkan bahwa k-NN juga lebih unggul dari SVM dengan perbandingan 87.59% (k-NN) banding 85.92% (SVM).
4. Pada AUC diperoleh hasil bahwa SVM lebih unggul dari pada k-NN dengan perbandingan 0.847 (SVM) banding 0.7 (kNN).

## VII.Saran

Pada penelitian *Feature Extraction* untuk klasifikasi pengenalan wajah

menggunakan *support vector machine* dan *k-nearest neighbors* ini ada beberapa hal yang perlu diperhatikan agar menjadi lebih baik kedepannya, yaitu: penggunaan data gambar dengan *background* yang seragam agar *Feature Extraction* yang dilakukan lebih optimal, pemilihan data *feature* perlu memperhatikan pengambilan sudut yang sama serta pencahayaan yang cukup.

## Daftar Pustaka

- Z. Dol, R. A. Salam and Z. Zainol, "Face Feature Extraction using Bayesian Network," ACM 1-59593-64-/06/0011, pp. 261-264, 2006.
- D. R., A. Rath, K. Manikantan and S. Ramachandran, "Astroid Shaped DCT Feature Extraction For Enhanced Face Recognition," ACM 978-1-4503-1185-4/12/09, pp. 95-101, 2012.
- M. Kusban, "Verifikasi dan Identifikasi Telapak Tangan dengan Karnael Gabor," Jurnal Nasional Teknik Elektro Teknik Informatika (JNTETI), vol. 4, no. 2, 2015.
- G. Kesavaraj and S. Sukumaran, "A Study On Classification Techniques in Data Mining," Computing, Communications and Networking Technologies (ICCCNT), pp. 1-7, 4 July 2013.
- J. Chen, Y. Gong and K. Zhang, "Facial Expression Recognition using Geometric and Appearance Features," ACM 978-1-4503-1600-2/12/09, pp. 29-33, 9 September 2012.
- H. Qureshi and M. Khan, "Association of Pre-pregnancy Weight and Weight

- Gain with Perinatal Mortality,” ACM 978-1- 4503-0342-2/10/12, 23 Desember 2010. 978-1-4503-2469-4/14/03, pp. 674-679, 24 March 2014.
- R. S. Bahri and I. Maliki, “Perbandingan Algoritma Template Matching dan Feature Extraction pada Optical Character Recognition,” Jurnal Komputer dan Informatika (KOMPUTA), vol. 1, no. 1, 2012.
- S. Sergyan, “Color histogram features based image classification in content-based image retrieval Systems,” International Symposium on Applied Machine Intelligence and Informatics, IEEE Xplore, pp. 221-224, February 2008.
- F. Rachman and W. Purnami, “Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer Dengan Menggunakan Regresi Logistik Ordinal Dan Support Vector Machine (SVM),” JURNAL SAINS DAN SENI ITS, vol. 1, no. 1, pp. 130-135, 2012.
- S. M.M., N. H., O. H.A., N. Zaini and G. S.A., “Support Vector Machine for Classification of Stress Subjects using EEG Signals,” in IEEE Conference on Systems, Process and Control, Kuala Lumpur, 2014.
- A. Rane, N. Naik and J. Laxminarayana, “Performance Enhancement of K Nearest Neighbor Classification Algorithm Using 8-Bin Hashing and Feature Weighting,” ACM 978-1-4503-2908-8/14/08, 2014.<http://www.cs.umass.edu/lfw/>, Wild, Labeled Faces in the, 2016.
- S. Aminikhanghahi, W. Wang, S. Shin, S. H. Son and S. I. Jeon, “Effective Tumor Feature Extraction for Smart Phone Based Microwave Tomography Breast Cancer Screening,” ACM