
PERBANDINGAN KINERJA PENGKLASIFIKASI CITRA BUAH KAKAO SAKIT DAN SEHAT MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM) DAN K-NEAREST NEIGHBORS (KNN)

Yohanes Balawuri Blikon

Fakultas Teknik, Program Studi Ilmu Komputer
Universitas Katolik Widya Mandira Kupang
Email: yohanesbalawuriblikon@gmail.com

Dr. Adri Gabriel Sooai, S.T., M.T.

Fakultas Teknik, Program Studi Ilmu Komputer
Universitas Katolik Widya Mandira Kupang
Email: adrigabriel@unwira.ac.id

Ign. Pricher A. N. Samane, S.Si., M.Eng.

Fakultas Teknik, Program Studi Ilmu Komputer
Universitas Katolik Widya Mandira Kupang
Email: samane_pricher@yahoo.co.id

ABSTRAK

Kakao merupakan salah satu hasil bumi dibidang perkebunan. Perkebunan kakao dengan hasilnya yaitu biji kakao dapat diolah menjadi bahan dasar tepung atau coklat. Keberadaan perkebunan ini tentu perlu mendapat dukungan teknologi atau kecerdasan buatan untuk membantu proses pensortiran secara modern jika dilakukan penerapan *conveyer belt* atau model pemetikan otomatis masa depan menggunakan drone pemetik buah. Proses pensortiran yang dimaksud yaitu menggunakan model pengklasifikasian untuk mendeteksi *dataset* buah kakao sakit dan sehat. Penelitian ini membandingkan model klasifikasi *Support Vector Machine* (SVM) dan *k-Nearest Neighbors* (KNN) dengan tujuan untuk mengetahui kinerja pengklasifikasi yang lebih persisi. Dari hasil ujicoba yang dilakukan performa dari model klasifikasi *Support Vector Machine* (SVM) dengan *kernel RBF* dan *cross validation 2* mendapatkan hasil prediksi yang lebih tinggi yaitu sebesar 82,5% sedangkan model klasifikasi *k-Nearest Neighbors* (KNN) dengan *number of neighbors 5*, *metric euclidean* dan *weight distance* tingkat akurasinya sebesar 82,3%.

Kata kunci: *support vector machine* (SVM), *k-nearest neighbors* (KNN), *dataset* buah kakao, performa klasifikasi

ABSTRACT

Cocoa is one of the agricultural products in the plantation sector. Cocoa plantations with the result that cocoa beans can be processed into basic ingredients for flour or chocolate. The existence of this plantation certainly needs the support of technology or artificial intelligence to assist the modern sorting process if a conveyer belt is applied or future automatic picking models using fruit picking drones. The sorting process in question is using a classification model to detect datasets of sick and healthy cocoa pods. This study compares the Support Vector Machine (SVM) and k-Nearest Neighbors (KNN) classification models with the aim of knowing the performance of a more precise classifier. From the test results, the performance of the Support Vector Machine (SVM) classification model with the RBF kernel and cross validation 2 obtained a higher prediction result of 82.5% while the k-Nearest Neighbors (KNN) classification model with a number of neighbors 5, Euclidean metrics and weight distance have an accuracy rate of 82.3%.

Keywords: *support vector machine (SVM), k-nearest neighbors (KNN), cocoa pods dataset, classification performance*

1. PENDAHULUAN

Kakao merupakan salah satu hasil bumi dibidang perkebunan. Perkebunan kakao dengan hasilnya yaitu biji kakao dapat diolah menjadi bahan dasar makanan berbasis tepung atau coklat [1][2]. Namun tanaman kakao sangat mudah diserang penyakit busuk apabila tidak dilakukan perawatan dengan baik. Penyakit busuk ini tidak hanya menyerang buah kakao tetapi juga batang tanaman kakao [3]. Gejala serangan pada buah dapat dilihat pada perubahan kulit buah yang berwarna hijau-kuning atau seperti matang sebelum waktunya sedangkan serangan pada batang tanaman kakao dapat dilihat pada buah yang memiliki bercak hitam dan lama-kelamaan seluruh buah akan berwarna hitam gelap [4].

Keberadaan perkebunan kakao tentu perlu mendapat dukungan teknologi kecerdasan buatan untuk membantu proses pensortiran secara modern jika dilakukan penerapan *conveyer* atau model pemetikan otomatis masa depan menggunakan drone pemetik buah. Proses pensortiran dalam hal ini menggunakan algoritma pengklasifikasian *Support Vector Machine (SVM)* dan *k-Nearest Neighbors (KNN)*.

Penelitian ini bertujuan untuk untuk mengetahui kinerja algoritma pengklasifikasi yang lebih persisi dari dua alternatif *Support Vector Machine (SVM)* dan *k-Nearest Neighbors (KNN)* terhadap *dataset* buah kakao sakit dan sehat sehingga dapat diterapkan pada berbagai industri.

2. METODOLOGI PENELITIAN

Tahap metodologi penelitian ini terdiri dari pengumpulan *dataset*, pembangunan model klasifikasi, *training* model klasifikasi, *testing* dan perhitungan performa.

2.1 Pengumpulan Dataset Buah Kakao

Dataset citra buah kakao sakit dan sehat didapatkan pada *website kaagle* dengan jumlah *dataset* sebanyak 4.390 citra. *Dataset* buah kakao sakit dan sehat memiliki 3 kelas yang dapat dilihat pada tabel 1. Tahapan ini tidak menggunakan pemrosesan awal seperti membuang *background* pada gambar dikarenakan menggunakan *dataset* sekunder yang sudah jadi untuk siap digunakan [5]. *Dataset* citra buah kakao pada gambar 1 akan digenerate menjadi sebuah angka-angka yang akan dibaca oleh mesin menggunakan *image embedding* dengan *server inception v3* agar dapat diklasifikasikan menggunakan model klasifikasi [6].

Tabel 1. Kelas Dataset

| <i>Kelas</i> | <i>Jumlah</i> |
|----------------------|---------------|
| <i>Black Pod Rod</i> | 943 |
| <i>Healthy</i> | 3344 |
| <i>Pod Borer</i> | 103 |



Black Pod Rod



Healthy

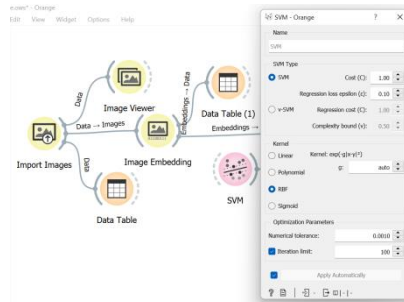


Pod Borer

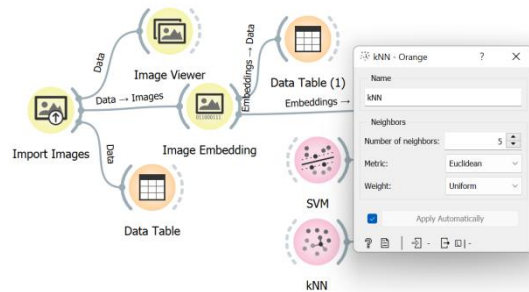
Gambar 1. Citra buah kakao

2.2 Pembangunan Model Klasifikasi

Model klasifikasi yang digunakan yaitu *Support Vector Machine (SVM)* dan *k-Nearest Neighbors (KNN)*. Pada pembangunan model klasifikasi SVM menggunakan *cost 1* dan *kernel Radial Basis Function (RBF)* yang dapat dilihat pada gambar 2 sedangkan pembangunan model klasifikasi KNN menggunakan *number of neighbors 5*, *metric euclidean* dan *weight distance* dapat dilihat pada gambar 3.



Gambar 2. Kernel Radial Basis Function (RBF)



Gambar 3. SEQ Gambar * ARABIC 3. Model

2.3 Training dan testing

Tahap *training* dan *testing* model klasifikasi menggunakan *cross validation* yang mana *cross validation* akan membandingkan data secara acak. Penelitian ini menggunakan *cross validation number of folds 2*, yang mana 4.390 citra akan *ditesting* sebanyak 2 citra hingga setiap kelompok diperlakukan sebagai validasi dan tersisa sebagai data *training*. Tujuan dari *cross validation* sendiri yaitu untuk memperoleh hasil akurasi yang maksimal [7].

2.4 Perhitungan Performa

Perhitungan performa dari kedua model klasifikasi berupa *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *Area Under Curve (AUC)*. Studi kasus pada penelitian ini adalah klasifikasi *multiclass* sehingga untuk perhitungannya menggunakan *weighted metric* [8]. TP (*True Positif*), TN (*True Negatif*), FP (*False Positive*), FN (*False Negative*), x (*total precision, recall, dan f1-score* tiap-tiap kelas), n_a (jumlah sampel *dataset* tiap-tiap kelas), n (jumlah keseluruhan *dataset*).

$$accuracy = \frac{TP}{\text{Jumlah Dataset}} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \times 100\% \quad (4)$$

$$W_{multiclass} = \frac{x_a \cdot n_a + x_b \cdot n_b + x_c \cdot n_c}{n} \times 100\% \quad (5)$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil Image Embedding

Hasil *embedder inception v3* dengan *image embedding* berhasil mengenerate *dataset* citra buah kakao sakit dan sehat sebanyak 2048 fitur. Proses *image embedding* diperlukan sebagai proses awal data gambar untuk dapat dengan mudah untuk melakukan klasifikasi. Hasil *embedder inception v3* dapat dilihat pada gambar 4.

| hidden | n2038 | n2039 | n2040 |
|--------|----------|----------|---------|
| 1 | 0.921333 | 0.247986 | 0.42250 |
| 2 | 0.445029 | 0.240353 | 0.03764 |
| 3 | 0.547821 | 0.297145 | 0.34404 |
| 4 | 0.376455 | 0.836076 | 0.43513 |
| 5 | 0.697804 | 0.266252 | 0.44737 |
| 6 | 0.994428 | 0.407531 | 0.42677 |
| 7 | 1.11266 | 0.518102 | 0.32901 |

Gambar 4. Hasil *embedder inception v3*

3.2 Hasil Cross Validation

Training dan *testing* model klasifikasi menggunakan *cross validation 2* dengan *widjet test dan score* mendapatkan hasil yang ada pada gambar 5.

| Model | AUC | CA | F1 | Precision | Recall |
|-------|-------|-------|-------|-----------|--------|
| kNN | 0.810 | 0.823 | 0.788 | 0.824 | 0.823 |
| SVM | 0.891 | 0.825 | 0.823 | 0.841 | 0.825 |

Gambar 5. *widjet test dan score*

Berdasarkan 4390 *dataset* citra buah kakao sakit dan sehat yang telah diuji, diperoleh hasil perhitungan *Accuracy*, *Precision*, *Recall*, *F1-Score* dan *AUC* dari model klasifikasi *Support Vector Machine* (SVM) dan *k-Nearest Neighbors* (KNN) seperti pada gambar 5. Hasil klasifikasi

model SVM dan KNN menunjukkan bahwa nilai akurasi SVM sebesar 82,5% sedangkan nilai akurasi dari KNN yaitu 82,3%.

Pada gambar 5 juga memperlihatkan perbandingan nilai AUC pada model SVM dan KNN, diketahui bahwa nilai AUC yang paling tinggi adalah model SVM yaitu sebesar 0.891. AUC digunakan untuk mengukur kinerja diskriminatif dengan memperkirakan *probabilitas output* dari ilustrasi yang diseleksi secara acak dari populasi positif maupun negatif, semakin besar AUC, semakin baik hasil klasifikasi yang digunakan.

3.3 Perhitungan performa *Support Vector Machine* (SVM) dan *k-Nearest Neighbors* (KNN)

Perhitungan performa model klasifikasi *support vector machine* (SVM) dan *k-nearest neighbors* (KNN) menggunakan *widjet confusion matrix*.

3.3.1 *Confusion Matrix Support Vector Machine* (SVM)

Perhitungan performa model klasifikasi *Support Vector Machine* (SVM) menggunakan *confusion matrix* dapat dilihat pada gambar 6.

| | | Predicted | | | Σ |
|--------|----------------------------|----------------------------|----------------------|------------------------|------|
| | | cacao_photos/black_pod_rot | cacao_photos/healthy | cacao_photos/pod_borer | |
| Actual | cacao_photos/black_pod_rot | 688 | 254 | 1 | 943 |
| | cacao_photos/healthy | 418 | 2926 | 0 | 3344 |
| | cacao_photos/pod_borer | 57 | 38 | 8 | 103 |
| Σ | | 1163 | 3218 | 9 | 4390 |

Gambar 6. Hasil *Confusion Matrix SVM*

Pada gambar 6 mendapat hasil bahwa nilai *True Positive* (TP) adalah 3622, *True Negative* (TN) adalah 8012, *False Positive* (FP) adalah 768, dan *False Negative* (FN) adalah 768. Maka perhitungan nilai *accuracy*, *precision*, *recall* dan *f1-score* adalah sebagai berikut:

$$Accuracy = \frac{TP}{Jumlah\ dataset} \times 100\%$$

$$Accuracy = \frac{3622}{4390} \times 100\%$$

Nilai *Accuracy* adalah 0.825%

$$Precision = \frac{Prec_{BPR} \cdot jumlah_{BPR} + Prec_{Healthy} \cdot jumlah_{Healthy} + Prec_{PB} \cdot jumlah_{PB}}{4390} \times 100\%$$

$$Precision = \frac{0.592 \cdot 943 + 0.909 \cdot 3344 + 0.889 \cdot 103}{4390} \times 100\%$$

Nilai *precision* adalah 0.840%

$$Recall = \frac{Recall_{BPR} \cdot jumlah_{BPR} + Recall_{Healthy} \cdot jumlah_{Healthy} + Recall_{PB} \cdot jumlah_{PB}}{4390} \times 100\%$$

$$Recall = \frac{0.730 \cdot 943 + 0.875 \cdot 3344 + 0.078 \cdot 103}{4390} \times 100\%$$

Nilai *recall* adalah 0.825%

$$F1 - Score = \frac{F1_{BPR} \cdot jumlah_{BPR} + F1_{Healthy} \cdot jumlah_{Healthy} + F1_{PB} \cdot jumlah_{PB}}{4390} \times 100\%$$

$$F1 - Score = \frac{0.654 \cdot 943 + 0.892 \cdot 3344 + 0.143 \cdot 103}{4390} \times 100\%$$

Nilai *f1-score* adalah 0.823%

3.3.2 Confusion Matrix *k*-Nearest Neighbors (KNN)

Perhitungan performa model klasifikasi *Support Vector Machine* (SVM) menggunakan *confusion matrix* dapat dilihat pada gambar 7.

| | | Predicted | | | Σ |
|--------|----------------------------|----------------------------|----------------------|------------------------|------|
| | | cacao_photos\black_pod_rot | cacao_photos\healthy | cacao_photos\pod_borer | |
| Actual | cacao_photos\black_pod_rot | 299 | 638 | 6 | 943 |
| | cacao_photos\healthy | 43 | 3299 | 2 | 3344 |
| | cacao_photos\pod_borer | 9 | 79 | 15 | 103 |
| Σ | | 351 | 4016 | 23 | 4390 |

Gambar 7. Hasil perhitungan performa model klasifikasi *Support Vector Machine* (SVM)

Pada gambar 7 mendapatkan hasil nilai *True Positive* (TP) adalah 3613, *True Negative* (TN) adalah 8003, *False Positive* (FP) adalah 777, dan *False Negative* (FN) adalah 713. Maka perhitungan nilai *accuracy*, *precision*, *recall* dan *f1-score* adalah sebagai berikut:

$$Accuracy = \frac{TP}{jumlah\ dataset} \times 100\%$$

$$Accuracy = \frac{3613}{4390} \times 100\%$$

Nilai *accuracy* adalah 0.823%

$$Precision = \frac{Prec_{BPR} \cdot jumlah_{BPR} + Prec_{Healthy} \cdot jumlah_{Healthy} + Prec_{PB} \cdot jumlah_{PB}}{4390} \times 100\%$$

$$Precision = \frac{0.852 \cdot 943 + 0.821 \cdot 3344 + 0.652 \cdot 103}{4390} \times 100\%$$

Nilai *precision* adalah 0.824%

$$Recall = \frac{Recall_{BPR} \cdot jumlah_{BPR} + Recall_{Healthy} \cdot jumlah_{Healthy} + Recall_{PB} \cdot jumlah_{PB}}{4390} \times 100\%$$

$$Recall = \frac{0.317 \cdot 943 + 0.987 \cdot 3344 + 0.146 \cdot 103}{4390} \times 100\%$$

Nilai *recall* adalah 0.823%

$$F1 - Score = \frac{F1_{BPR} \cdot jumlah_{BPR} + F1_{Healthy} \cdot jumlah_{Healthy} + F1_{PB} \cdot jumlah_{PB}}{4390} \times 100\%$$

$$f1 - score = \frac{0.462 \cdot 943 + 0.896 \cdot 3344 + 0.239 \cdot 103}{4390} \times 100\%$$

Nilai *f1-score* adalah 0.788%

Berdasarkan hasil *confusion matrix* dari model klasifikasi *Support Vector Machine* (SVM) dan *k-Nearest Neighbors* (KNN) dapat diperoleh nilai perbandingan *accuracy*, *precision*, *recall*, dan *f1-score* seperti terlihat pada tabel 2.

Tabel 2. Perbandingan Kinerja SVM dan KNN

| <i>Metode</i> | <i>Accuracy</i> | <i>Precision</i> | <i>Recall</i> | <i>F1-Score</i> |
|---------------|-----------------|------------------|---------------|-----------------|
| SVM | 0.825% | 0.841% | 0.824% | 0.823% |
| KNN | 0.823 | 0.824% | 0.823% | 0.788% |

Berdasarkan Tabel 2 dapat diketahui bahwa kinerja dari model *Support Vector Machine* (SVM) lebih baik dari model *K-Nearest Neighbors* (KNN) walaupun perbandingannya tidak terlalu signifikan. Akurasi klasifikasi tidak mencapai hasil yang sempurna karena ada nilai error. Hal tersebut diakibatkan oleh data uji dan data latih yang begitu banyak dalam melakukan proses simulasi.

4. KESIMPULAN

Tingkat akurasi klasifikasi dataset buah kakao sakit dan sehat yang diukur menggunakan model klasifikasi *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF) dan *Cross Validation 2* mendapatkan hasil prediksi yang lebih tinggi yaitu sebesar 82,5% sedangkan model klasifikasi *k-Nearest Neighbors* (KNN) dengan *number of neighbors 5*, *metric Euclidean* dan *weight distance* mendapatkan tingkat akurasi sebesar 82,3%.

DAFTAR PUSTAKA

- [1] E. Ariningsih, H. J. Purba, J. F. Sinuraya, S. Suharyono, and K. S. Septanti, "Kinerja Industri Kakao di Indonesia," forum penelit. agro ekon., vol. 37, no. 1, p. 1, Jul. 2020, doi: 10.21082/fae.v37n1.2019.1-23.
- [2] S. Y. Tyasmoro, P. N. Permanasari, and A. Saitama, *Teknologi Produksi Tanaman Perkebunan*. Universitas Brawijaya Press, 2021.
- [3] S. Alim, P. P. Lestari, and R. Rusliyawati, "SISTEM PAKAR DIAGNOSA PENYAKIT TANAMAN KAKAO MENGGUNAKAN METODE CERTAINTY FACTOR PADA KELOMPOK TANI PT OLAM INDONESIA (COCOA) CABANG LAMPUNG," *Jurnal Data Mining dan Sistem Informasi*, vol. 1, no. 1, Art. no. 1, Aug. 2020, doi: 10.33365/jdmsi.v1i1.798.
- [4] "Buku saku hama dan penyakit kakao.pdf." Accessed: Nov. 14, 2022. [Online]. Available: <http://repository.pertanian.go.id/bitstream/handle/123456789/9785/Buku%20saku%20hama%20dan%20penyakit%20kakao.pdf?sequence=1>
- [5] P. Purwono and I. Mangkunegara, *Belajar Data Science dengan Kaggle Dilengkapi dengan Praktikum Latihan Data Science Untuk Pemula*. 2022.
- [6] A. J. Lado et al., "Comparison of Neural Network and Random Forest Classifier Performance on Dragon Fruit Disease," in 2021 International Electronics Symposium (IES), Sep. 2021, pp. 287–291. doi: 10.1109/IES53407.2021.9593992.
- [7] B. Prasetyo, V. Suryani, and D. R. Anbiya, "Analisis Deteksi Malware Pada Aplikasi Android Fintech Berdasarkan Permissions Dengan Menggunakan Naive Bayes Dan Random Forest," *eProceedings of Engineering*, vol. 8, no. 5, Art. no. 5, Oct. 2021, Accessed: Nov. 15, 2022. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15749>

- [8] M. F. Naufal, "Analisis Perbandingan Algoritma Svm, Knn, Dan Cnn untuk Klasifikasi Citra Cuaca.," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 8, no. 2, Art. no. 2, Apr. 2021.