

# Klasifikasi Citra Cuaca Menggunakan Inception-V3 dan K-Nearest Neighbors

Iqbal Giffari Ritonga<sup>1</sup>, Rika Rosnelly<sup>2</sup>, Pius Deski Manalu<sup>3</sup>, Teresa Tamba<sup>4</sup>, Kristine Wau<sup>5</sup>

Address : Universitas Potensi Utama, Ilmu Komputer, Indonesia<sup>1,2,3,4,5</sup>

Email : iqbalgiffari97@gmail.com<sup>1</sup>, rika@potensi-utama.ac.id<sup>2</sup>, piusdeski@gmail.com<sup>3</sup>, teresatamba3@gmail.com<sup>4</sup>, wau\_kristine@gmail.com<sup>5</sup>

## Abstrak

Gambaran cuaca memiliki peran krusial di berbagai sektor, seperti penerbangan, kelautan, dan pertanian. Keadaan cuaca memiliki dampak besar dalam aktivitas di bidang-bidang tersebut dan sangat berpengaruh terhadap operasionalnya. Mengklasifikasikan gambaran cuaca dapat dilakukan dengan menganalisis data citra cuaca, yang bisa digunakan untuk meramalkan jenis cuaca yang mungkin terjadi. Hasil prediksi tentang cuaca ini memiliki nilai yang signifikan dalam pengambilan keputusan sehari-hari dalam berbagai sektor tersebut. Salah satu metode klasifikasi citra cuaca dapat dilakukan dengan cara mengekstraksi fitur citra cuaca terlebih dahulu menggunakan *Inception-V3* yang selanjutnya dilakukan perhitungan menggunakan metode *K-Nearest Neighbors*. Penelitian ini menggunakan 1748 citra cuaca dengan 4 kategori untuk melakukan pelatihan yang menghasilkan model dengan *Accuracy 91%*, *F1 91%*, *Recall 91%*, *Precision 91%*, serta menggunakan 8 citra cuaca dengan 4 kategori untuk melakukan pengujian yang menghasilkan klasifikasi bernilai benar semua pada setiap citranya.

**Keywords** – *Weather Datasets, Inception-V3, K-Nearest Neighbors*

## Abstract

Weather imagery has a crucial role in various sectors, such as aviation, maritime and agriculture. Weather conditions have a big impact on activities in these fields and greatly influence operations. Classifying weather images can be done by analyzing weather image data, which can be used to predict the type of weather that may occur. The results of these weather predictions have significant value in daily decision making in these various sectors. One method for classifying weather images can be done by first extracting weather image features using *Inception-V3* which is then calculated using the *K-Nearest Neighbors* method. This research uses 1748 weather images with 4 categories to carry out training which produces a model with *Accuracy 91%*, *F1 91%*, *Recall 91%*, *Precision 91%*, and uses 8 weather images with 4 categories to carry out testing which produces classifications with all correct values. every image.

**Keywords** – *Weather Datasets, Inception-V3, K-Nearest Neighbors*

## 1. Latar Belakang

Citra cuaca memiliki peran penting di beberapa bidang seperti bidang penerbangan, kelautan, dan pertanian. Cuaca dapat mempengaruhi bidang tersebut dan sangat berdampak dalam kegiatannya. Citra Cuaca dapat diklasifikasikan dengan mengidentifikasi setiap dataset citra cuaca yang dapat menghasilkan prediksi klasifikasi citra cuaca. Hasil prediksi citra cuaca dapat bermanfaat pada bidang tersebut dalam menentukan keputusan pada kegiatan sehari – harinya.

Identifikasi adalah suatu cara yang berperan penting dalam proses pengenalan ataupun pembeda antara

suatu hal seperti gambar/foto ataupun tanda tangan [1], [2].

Metode yang digunakan dalam penelitian ini adalah penelitian kuantitatif. Metode kuantitatif adalah metode yang digunakan untuk memecahkan masalah dengan menggunakan metode ilmiah untuk mengukur variabel penelitian dan hubungan antar variabel [3].

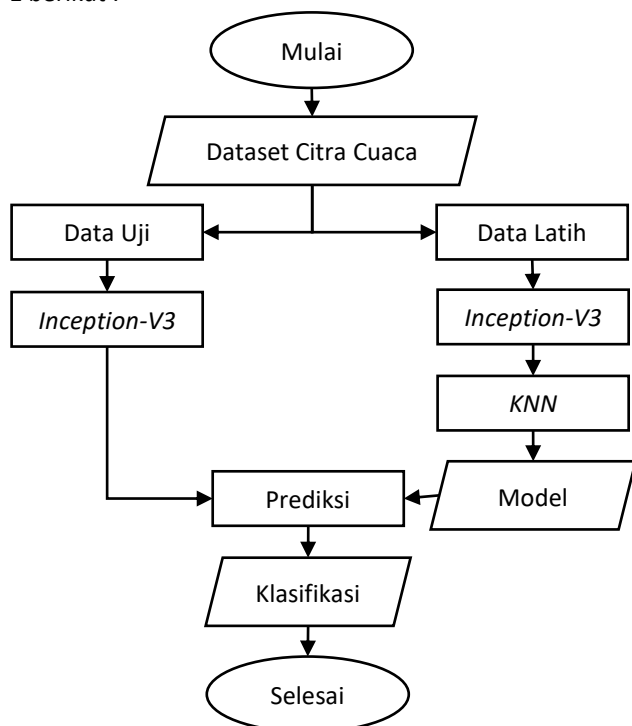
*Inception-V3* adalah jaringan yang didasarkan algoritma *Convolutional Neural Network* terlatih yang dapat digunakan untuk mengesktrak fitur gambar berdasarkan kumpulan dataset sebelumnya [4], [5].

Satu dari klasifikasi yang populer *K-Nearest Neighbors* adalah metode klasifikasi yang tidak membutuhkan pengetahuan sebelumnya, sampel label baru hanya ditentukan oleh tetangga terdekatnya [6], [7].

Berdasarkan latar belakang diatas ekstrasi fitur citra akan dilakukan dengan metode *Inception-V3* dan hasil dari ekstrasi akan dilakukan klasifikasi menggunakan *K-Nearest Neighbors* yang diharapkan dapat meningkatkan ketepatan dalam klasifikasi citra cuaca.

## 2. Metode

Rancangan metode penelitian dapat dilihat pada Gambar 1 berikut :



Gambar 1. Rancangan Metode Penelitian

### 2.1. Dataset Citra Cuaca

Dataset yang digunakan bersumber dari [8]. Dataset ini memiliki 11 kategori dan yang digunakan hanya berjumlah 4 kategori yaitu : *rain*, *rainbow*, *lightning*, dan *snow*.

### 2.2. Data Latih dan Data Uji

Dataset kemudian akan dibagi menjadi 2 kategori yaitu : data latih dan data uji. Data latih dan data uji dapat dilihat pada Tabel 1 dan Tabel 2 sebagai berikut :

Tabel 1. Data Latih

No	Kategori	Jumlah
1	<i>Rain</i>	524 Citra
2	<i>Rainbow</i>	230 Citra

No	Kategori	Jumlah
3	<i>Lightning</i>	375 Citra
4	<i>Snow</i>	619 Citra
<b>Total</b>		1748 Citra

Tabel 2 Data Uji

No	Kategori	Jumlah
1	<i>Rain</i>	2 Citra
2	<i>Rainbow</i>	2 Citra
3	<i>Lightning</i>	2 Citra
4	<i>Snow</i>	2 Citra
<b>Total</b>		8 Citra

### 2.3. Inception-V3

Kombinasi dengan 9 dan 22 modul *convolutional layer*, *IV3* menggunakannya untuk memproses fitur dari sebuah gambar [4], [9].

Sebagai salah satu variasi *GoogleNet*, *IV3*, berdasarkan faktorisasi konvolusi  $7 \times 7$ , yang terdiri dari 2 atau 3 *convolution layer* dengan  $3 \times 3$  dimensi [4], [10].

### 2.4. KNN

Beberapa cara yang digunakan untuk melakukan perhitungan jarak antara data uji dengan data latih, yaitu *manhattan distance* (*city block distance*) dan *euclidean distance*. Perhitungan jarak yang sering digunakan adalah *euclidean distance*. Jarak *euclidean distance* dapat dilihat pada persamaan 1 berikut [11], [12]:

$$d = \sqrt{\sum_{i=1}^p (x_1 - x_2)^2} \quad (1)$$

Keterangan :

- $d$  = jarak
- $l$  = data variabel
- $p$  = data dimensi
- $x_1$  = data sampel
- $x_2$  = data uji

Algoritma *K-Nearest Neighbor* memiliki kelebihan yaitu tangguh terhadap data latih yang terdapat banyak *noise* dan data yang jumlah besar. Kekurangan algoritma *K-Nearest Neighbor* adalah perlu penentuan jumlah tetangga terdekat dari data target, yang disimbolkan dengan parameter  $K$ , data latih didasarkan pada hasil perhitungan jarak kurang akurat dikarenakan harus memilih, mencoba, serta menentukan jenis jarak yang digunakan, atribut mana yang digunakan untuk menghasilkan yang memiliki perhitungan jarak yang terbaik, dan biaya komputasi yang tinggi karena memerlukan perhitungan jarak dari setiap *query instance* pada keseluruhan contoh data *training* [11].

**2.5. Model**

*Confusion matrix* adalah matriks yang cukup intuitif serta mudah untuk mengetahui tingkat ketepatan dan akurasi dari model yang dihasilkan [13].

Berdasarkan *confusion matrix*, jumlah total prediksi yang benar diperlihatkan oleh variabel *TP (True Positive)* dan *TN (True Negative)*. Sedangkan jumlah prediksi yang salah diperlihatkan oleh variabel *FP (False Positive)* dan *FN (False Negative)*. Perhitungan indikator performa didapatkan dengan cara menghitung nilai *accuracy*, *precision*, *recall*, dan *F1-Score* dapat dilihat pada persamaan (2), (3), (4), dan (5) berikut [13]:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 - Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (5)$$

*TP* = jumlah data kelas *positive* (1) yang benar diprediksi sebagai kelas *positive* (1).

*TN* = jumlah data kelas *negative* (0) yang benar diprediksi sebagai kelas *negative* (0).

*FP* = jumlah data kelas *negative* (0) yang salah diprediksi sebagai kelas *positive* (1).

*FN* = jumlah data kelas *positive* (1) yang salah diprediksi sebagai kelas *negative* (0).

**2.6. Prediksi**

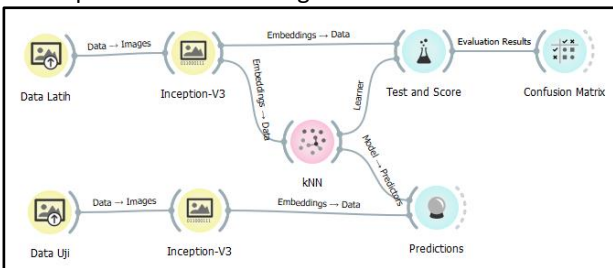
Hasil dari model akan dilakukan prediksi dengan menggunakan data uji.

**2.7. Klasifikasi**

Hasil dari prediksi akan menghasilkan klasifikasi citra cuaca dari setiap data uji.

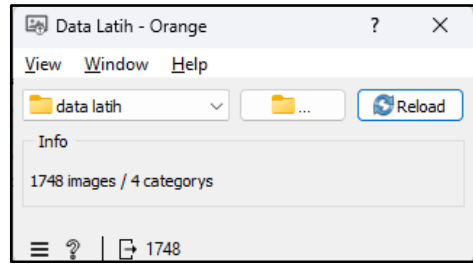
**3. Hasil**

Rancangan metode penelitian dilakukan dengan menggunakan *Software Orange Data Mining* dapat dilihat pada Gambar 2 sebagai berikut :



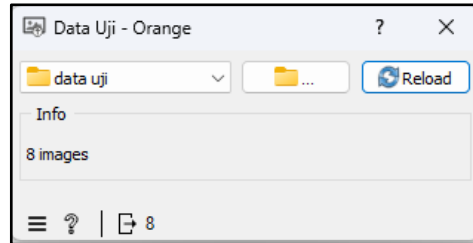
Gambar 2. Rancangan Metode Penelitian *Orange Data Mining*

Data latih yang digunakan berjumlah 1748 citra dengan 4 kategori dapat dilihat pada Gambar 3 berikut :



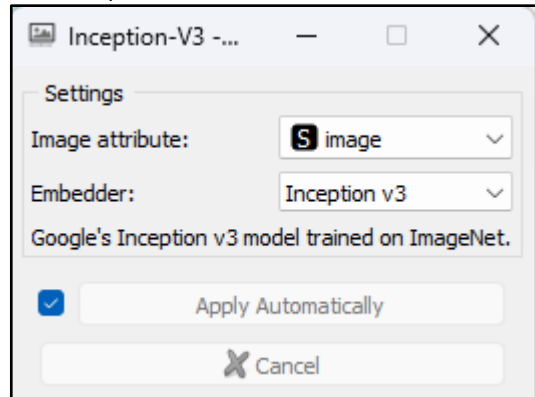
Gambar 3. Data Latih

Data uji yang digunakan berjumlah 8 citra dengan 4 kategori dapat dilihat pada Gambar 4 berikut :



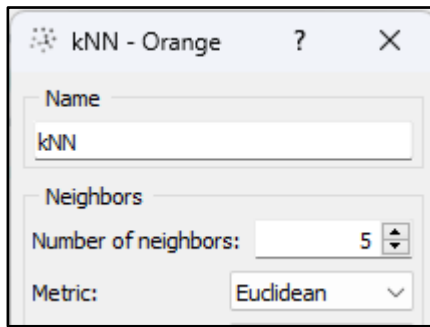
Gambar 4. Data Uji

Data latih dan data uji akan dilakukan proses mengekstrasi fitur gambar menggunakan *Inception-V3* dapat dilihat pada Gambar 5 berikut :



Gambar 5. *Inception-V3*

Setelah ekstrasi fitur *image* dilakukan maka akan dilakukan perhitungan metode *K-Nearest Neighbors* dengan jumlah *K=5* dan jarak yang digunakan adalah *euclidean* yang dapat dilihat pada Gambar 6 sebagai berikut :



Gambar 6. K-Nearest Neighbors

Hasil dari perhitungan model *K-Nearest Neighbors* akan menghasilkan *Confusion Matrix* dapat dilihat pada Gambar 7 berikut :

		Predicted				Σ
		lightning	rain	rainbow	snow	
Actual	lightning	371	0	2	2	375
	rain	18	457	5	44	524
	rainbow	13	0	217	0	230
	snow	8	56	6	549	619
Σ		410	513	230	595	1748

Gambar 7. Confusion Matrix

Hasil dari *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dapat dilihat pada Gambar 8 berikut :

Model	CA	F1	Prec	Recall
kNN	0.912	0.911	0.912	0.912

Gambar 8. Accuracy, Precision, Recall, dan F1-Score

Selanjutnya model *K-Nearest Neighbors* akan dilakukan pengujian untuk mendapatkan hasil klasifikasi menggunakan data uji dapat dilihat pada Gambar 9 berikut :

	kNN	image name
1	lightning	test lightning 1
2	lightning	test lightning 2
3	rain	test rain 1
4	rain	test rain 2
5	rainbow	test rainbow 1
6	rainbow	test rainbow 2
7	snow	test snow 1
8	snow	test snow 2

Gambar 9. Pengujian

#### 4. Kesimpulan

Berdasarkan hasil penelitian dari 1748 citra cuaca 4 kategori yang diekstraksi fitur menggunakan *Inception-V3* dan dilakukan perhitungan dengan *K-Nearest Neighbors* didapatkan model dengan *Accuracy 91%*, *F1 91%*, *Recall 91%*, *Precision 91%*.

Hasil klasifikasi dari 8 citra cuaca 4 kategori menghasilkan klasifikasi yang bernilai benar semua.

Pada penelitian selanjutnya diharapkan agar menambahkan jumlah dataset agar meningkatkan keakuratan dari model yang dihasilkan.

#### References

- [1] D. Darwish, "Simulation and Evaluation of Signature Recognition Techniques," *vol*, vol. 37, pp. 65–74, 2013.
- [2] M. S. SIMANJUNTAK, "Identifikasi Tanda Tangan menggunakan Metode Fitur Ekstraksi Biner dan K Nearest Neighbor," *CSRID (Computer Science Research and Its Development Journal)*, vol. 12, no. 3, p. 191, Mar. 2021, doi: 10.22303/csrid.12.3.2020.191-200.
- [3] R. Rosnelly and L. Wahyuni, "Tropical Diseases Identification Using Neural Network Adaptive Resonance Theory 2," in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, IEEE, Aug. 2018, pp. 1–4. doi: 10.1109/CITSM.2018.8674364.
- [4] I. Firmansyah, R. Rosnelly, and Wanayumini, "Inception-V3 Versus VGG-16 in Rice Classification Using Multilayer Perceptron," in *2nd International Conference on Information Science and Technology Innovation (ICoSTEC)*, 2023.
- [5] W. Shen and X. Li, "Facial expression recognition based on bidirectional gated recurrent units within deep residual network," *International Journal of Intelligent Computing and Cybernetics*, vol. 13, no. 4, pp. 527–543, Nov. 2020, doi: 10.1108/IJICC-07-2020-0088.
- [6] D. Pan, Z. Zhao, L. Zhang, and C. Tang, "Recursive clustering K-nearest neighbors algorithm and the application in the classification of power quality disturbances," in *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*, IEEE, Nov. 2017, pp. 1–5. doi: 10.1109/EI2.2017.8245652.
- [7] R. M. F. Lubis, Z. Situmorang, and R. Rosnelly, "ANALISIS VARIATION K-FOLD CROSS VALIDATION ON CLASSIFICATION DATA METHOD K-NEAREST NEIGHBOR," *JURNAL IPTEKS TERAPAN*, vol. 14, no. 3, pp. 206–211, 2020, doi: https://doi.org/10.22216/jit.v14i3.98.
- [8] Jehan Bhatena, "Weather Image Recognition," Kaggle. Accessed: Oct. 23, 2023. [Online].

Available:

<https://www.kaggle.com/datasets/jehanbathena/weather-dataset>

- [9] R. A. Pratiwi, S. Nurmaini, D. P. Rini, M. N. Rachmatullah, and A. Darmawahyuni, "Deep ensemble learning for skin lesions classification with convolutional neural network," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 3, p. 563, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp563-570.
- [10] A. Lumini, L. Nanni, and G. Maguolo, "Deep learning for plankton and coral classification," *Applied Computing and Informatics*, vol. 19, no. 3/4, pp. 265–283, Jun. 2023, doi: 10.1016/j.aci.2019.11.004.
- [11] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 6, no. 2, pp. 118–127, 2021, doi: <https://doi.org/10.31294/ijcit.v6i2.10438>.
- [12] N. L. G. P. Suwirmayanti, "Penerapan Metode K-Nearest Neighbor Untuk Sistem Rekomendasi Pemilihan Mobil," *Techno.Com*, vol. 16, no. 2, pp. 120–131, Feb. 2017, doi: 10.33633/tc.v16i2.1322.
- [13] A. F. Hidayatullah, A. A. Fadila, K. P. Juwairi, and R. A. Nayoan, "Identifikasi Konten Kasar pada Tweet Bahasa Indonesia," *Jurnal Linguistik Komputasional (JLK)*, vol. 2, no. 1, p. 1, Mar. 2019, doi: 10.26418/jlk.v2i1.15.