

Pemisahan Sampah Menggunakan Metode Klasifikasi Regresi Logistik

Abdul Majid

Magister Teknik Informatika, Universitas Pamulang

e-mail: kangabdulpati@gmail.com

Abstrak—Pengolahan citra merupakan salah satu metode pengambilan data agar suatu pekerjaan dalam otomatisasi, terutama untuk pengolahan data dalam jumlah besar. Pengolahan data itu sendiri biasa disebut data mining. Dengan memanfaatkan pengolahan citra, data dapat dikumpulkan dan diproses untuk selanjutnya digunakan dalam sebuah system otomatisasi. Dengan contoh data set gambar beberapa jenis sampah, selanjutnya dilakukan proses klasifikasi menggunakan metode Regresi Logistik yang disimulasikan menggunakan aplikasi data mining *Orange* memiliki akurasi hingga 0,836. Selain itu, aplikasi *orange* memiliki fitur *image analytics* sehingga sangat memudahkan dalam melakukan simulasi dengan data citra.

Kata Kunci—Pemisahan Sampah, Regresi Logistik, Orange.

I. PENDAHULUAN

Sampah dan masalah pengelolaannya adalah perhatian global yang semakin mendalam seiring dengan pertumbuhan populasi dan urbanisasi yang pesat. Identifikasi dan pengelompokan sampah dari gambar merupakan langkah kunci dalam upaya otomatisasi proses pengelolaan sampah. Penelitian ini bertujuan untuk mengembangkan sistem yang dapat mengklasifikasikan sampah dalam gambar ke dalam kategori yang relevan, dengan menggunakan metode regresi logistik.

Penelitian ini melibatkan pengumpulan dataset gambar sampah yang mencakup berbagai jenis sampah, termasuk kardus, plastik, kertas, logam, dan kaca. Setelah preprocessing dan ekstraksi fitur dari gambar-gambar tersebut, metode regresi logistik diterapkan untuk melakukan klasifikasi. Keakuratan klasifikasi diukur dan dievaluasi dengan berbagai metrik performa, seperti akurasi, presisi, recall, dan F1-score.

Hasil eksperimen menunjukkan bahwa metode regresi logistik mampu memberikan hasil yang obyektif dan akurat dalam klasifikasi sampah berdasarkan gambar. Selain itu, pendekatan ini memberikan landasan yang kokoh untuk pengembangan sistem pengelolaan sampah yang lebih efisien dan otomatis. Kemampuan untuk mengenali dan mengelompokkan sampah dari gambar-gambar ini berpotensi memberikan kontribusi positif dalam upaya pelestarian lingkungan dan pengurangan dampak negatif dari pengelolaan sampah yang tidak efisien. Hasil penelitian ini dapat digunakan sebagai landasan bagi penelitian lanjutan dalam bidang pengelolaan sampah berbasis gambar dan memperkaya pemahaman kita tentang aplikasi metode regresi logistik dalam konteks klasifikasi sampah.

II. METODE PENELITIAN

Perhitungan *image analytics* dan klasifikasi Regresi Logistik maupun visualisasi data menggunakan aplikasi *Orange*.

A. Persiapan Data Set

Data set gambar sampah diperoleh secara *online* yang dibuat terbuka untuk kebutuhan umum [1].

B. Image Analytics

Image Analytics merupakan ekstraksi informasi-informasi yang berarti dari gambar, umumnya dari gambar digital dalam cakupan Teknik pemrosesan gambar [2].

1. *Preprocessing* mencakup peningkatan kualitas gambar, menghilangkan noise, meningkatkan kontras gambar.
2. *Segmentasi* memisahkan bagian-bagian gambar digital menjadi beberapa bagian/segmen [3].
3. Ekstraksi Fitur melibatkan identifikasi fitur-fitur penting dalam gambar yang akan digunakan untuk analisis lebih lanjut.
4. Pengenalan Pola teknik-teknik seperti pengenalan wajah, pengenalan tulisan tangan, atau klasifikasi objek digunakan untuk mengidentifikasi dan mengklasifikasikan pola yang ada dalam gambar.

C. Regresi Logistik

Regresi logistik (kadang disebut model logistik atau model logit), dalam statistika digunakan untuk prediksi probabilitas kejadian suatu peristiwa dengan mencocokkan data pada fungsi logit kurva logistik [4]. Fungsi logistik adalah sebagai berikut:

$$p(y = 1) = \frac{1}{(1 + e^{-z})}$$

Di sini, z adalah kombinasi linear dari fitur dan parameter model:

$$z = b_0 + (b_1 * x_1) + (b_2 * x_2) + \dots + (b_n * x_n)$$

Di mana $b_0, b_1, b_2, \dots, b_n$ adalah parameter model yang akan dipelajari selama proses pelatihan.

D. Test and Score

Evaluasi hasil model dengan mendapatkan data numerik.

E. Classification Accuracy (CA)

Menghitung tingkat akurasi klasifikasi

F. F1

Rata-rata terimbang dari presisi dan daya ingat dengan rumus:

$$F1 = \frac{2 * (precision * recall)}{(precision + recall)}$$

G. Precision

Kemampuan *classifier* untuk tidak memberikan label positif kepada sampel negatif begitupun sebaliknya. *Precision* memiliki rumus sebagai berikut:

$$Precision = \frac{TP}{TP + FP}$$

Keterangan:

TP : Jumlah positif sebenarnya

FP : Jumlah positif palsu

H. Recall

Kemampuan *classifier* untuk menemukan dan menggolongkan semua sampel yang bernilai positif.

I. Confusion Matrix

Dalam dunia *Machine Learning* dan terutama pada permasalahan klasifikasi statistik, *confusion matrix* dikenal juga dengan *error matrix* [5]. Dimana pengukuran performa masalah klasifikasi data mining dengan keluaran berupa dua kelas atau lebih.

J. ROC Analysis

Analisa hasil dari alat ukur *performance* untuk mengklasifikasikan masalah dalam menentukan *threshold* dari suatu model.

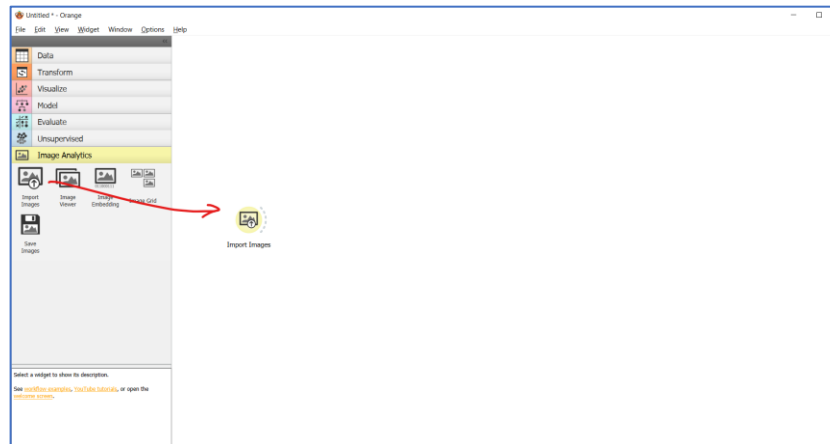
III. HASIL DAN PEMBAHASAN

A. Data Set

Data set gambar sampah diperoleh secara online yang dibuat terbuka untuk kebutuhan umum.

B. Orange Workflow

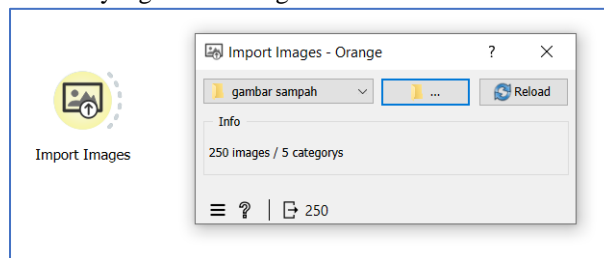
1. Import data set



Gambar 1.

Import data set/image ke dalam aplikasi orange

Dilanjutkan dengan memilih folder yang berisi data gambar.

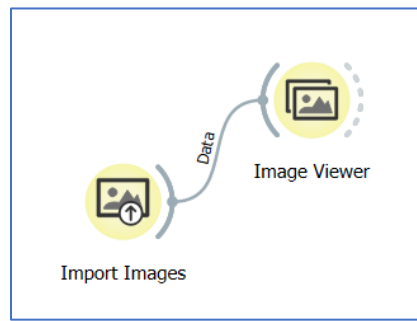


Gambar 2.

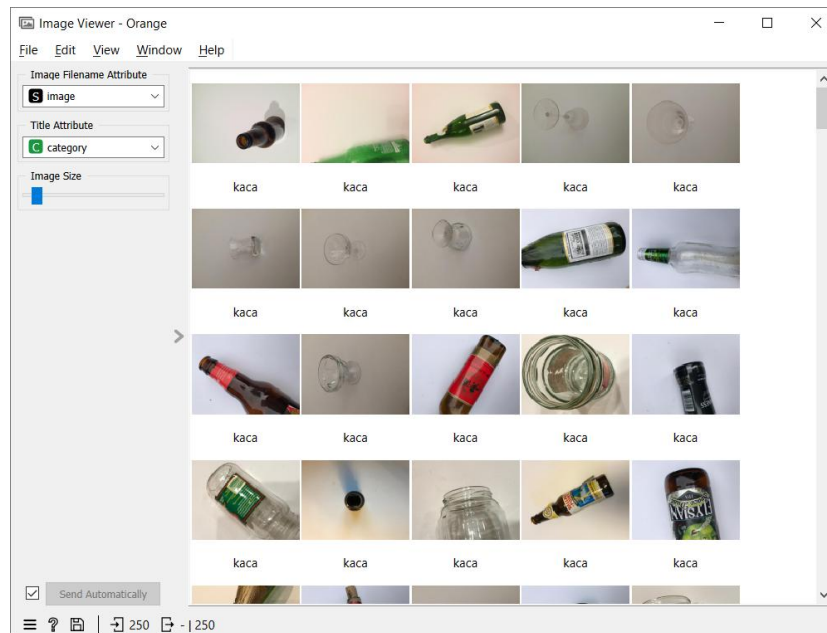
Memilih folder yang berisi data set gambar

2. Image Viewer

Menampilkan gambar yang sudah di-import ke dalam aplikasi.

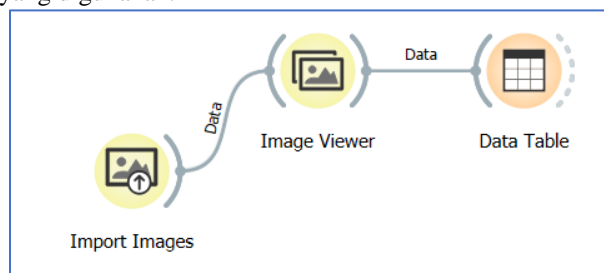


Gambar 3.
Image Viewer untuk menampilkan data gambar



Gambar 4.
 Data set gambar yang ditampilkan dalam *Image Viewer*

3. *Data Table*
 Menampilkan list data set yang digunakan.



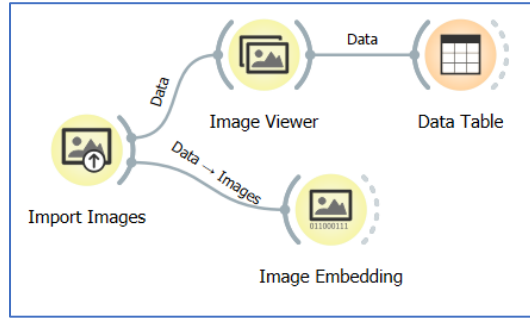
Gambar 5.
Data Table menampilkan list data set

origin	category	image name	image	size	width	height	Selected
1	kaca	kaca1	kaca/kaca1.jpg	8296	512	384	No
2	kaca	kaca10	kaca/kaca10.jpg	8722	512	384	No
3	kaca	kaca11	kaca/kaca11.jpg	11447	512	384	No
4	kaca	kaca12	kaca/kaca12.jpg	7488	512	384	No
5	kaca	kaca13	kaca/kaca13.jpg	7656	512	384	No
6	kaca	kaca14	kaca/kaca14.jpg	7908	512	384	No
7	kaca	kaca15	kaca/kaca15.jpg	9076	512	384	No
8	kaca	kaca16	kaca/kaca16.jpg	9424	512	384	No
9	kaca	kaca17	kaca/kaca17.jpg	19880	512	384	No
10	kaca	kaca18	kaca/kaca18.jpg	11869	512	384	No

Gambar 6.
 Tampilan data set dalam *Data Table*

4. *Image Embedding*

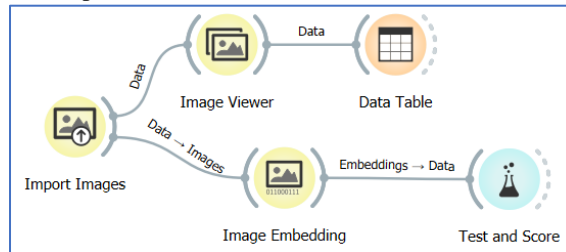
Konversi gambar menjadi vektor fitur numerik yang dapat digunakan dalam analisis data lebih lanjut.



Gambar 7.
Proses Image Embedding

5. *Test and Score*

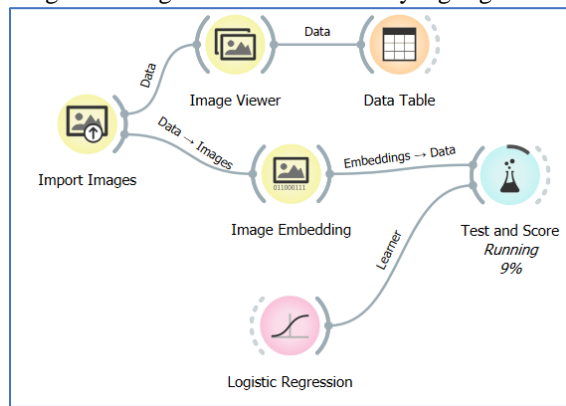
Test and score untuk menghasilkan prediksi dan evaluasi hasil.



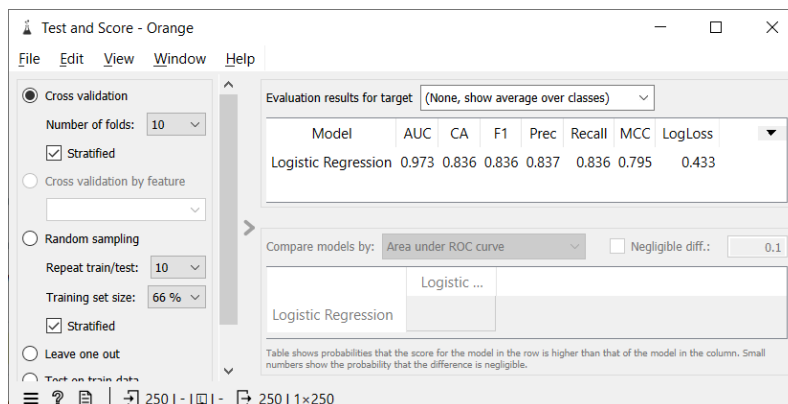
Gambar 8.
Proses *Test and Score*

6. Regresi Logistik

Penambahan metode Regresi Logistik sebagai metode klasifikasi yang digunakan.



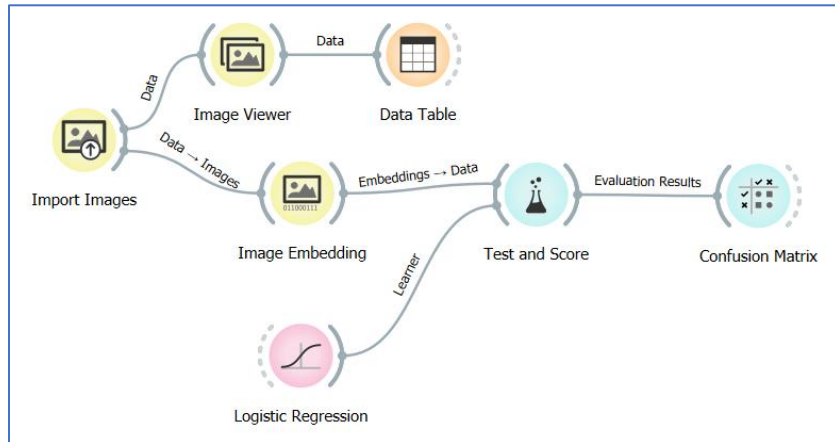
Gambar 9.
Penambahan metode klasifikasi



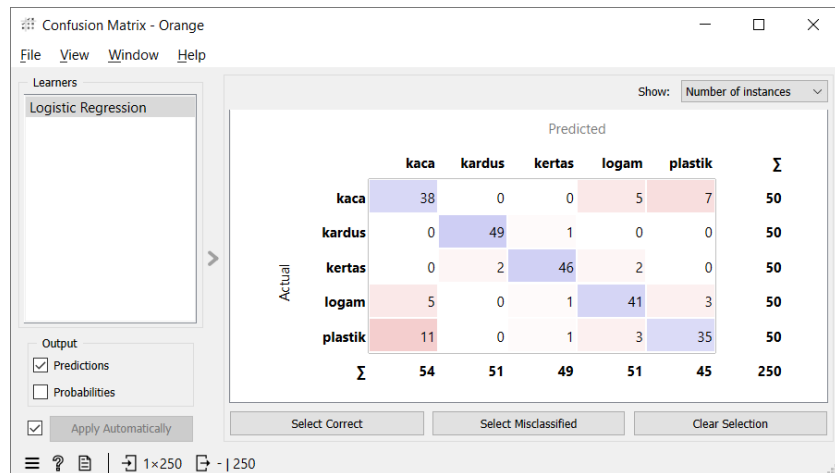
Gambar 10.
Hasil *Test and Score* setelah ditambahkan metode klasifikasi

7. *Confusion Matrix*

Confusion matrix dapat dibuat setelah hasil dari test and score didapatkan. Cara membuat *confusion matrix* adalah dengan menghubungkan *widget test and score* ke *widget confusion matrix* seperti gambar 11.

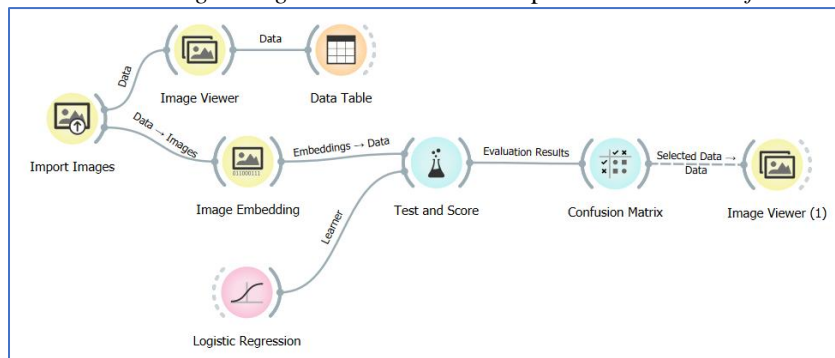


Gambar 11.
 Penambahan proses *Confusion Matrix*



Gambar 12.
 Hasil dari *Confusion Matrix*

Selanjutnya bisa ditambahkan *widget image viewer* untuk menampilkan hasil dari *confusion matrix*.



Gambar 13.
 Penambahan *image viewer* setelah *confusion matrix*

IV. KESIMPULAN

Hasil dari percobaan pengelompokan sampah menggunakan metode regresi logistik dengan aplikasi data mining *Orange* dapat disimpulkan berdasarkan *Test and Score* sebagai berikut:

1. *AUC (Area Under the Curve)*: Nilai AUC adalah 0.9726. Ini mengukur area di bawah kurva Receiver Operating Characteristic (ROC) yang digunakan untuk mengukur kinerja model klasifikasi. Karena semakin dekat nilai AUC ke 1, semakin baik modelnya.
2. *CA (Classification Accuracy)*: Nilai CA adalah 0.836. Ini adalah akurasi klasifikasi, yang mengukur sejauh mana model benar dalam memprediksi kelas dengan benar. Nilai ini berarti model memprediksi dengan benar sekitar 83.6% dari keseluruhan pengamatan.
3. *F1-Score*: Nilai F1-Score adalah sekitar 0.8358. F1-Score adalah ukuran gabungan dari presisi dan recall. Nilai ini menunjukkan keseimbangan antara kemampuan model untuk mengidentifikasi positif dengan benar dan menghindari positif palsu.
4. *Precision (Prec)*: Nilai Presisi adalah sekitar 0.836. Presisi adalah rasio dari True Positives terhadap semua yang diprediksi positif. Ini mengukur sejauh mana model benar dalam memprediksi positif ketika ia mengklasifikasikan sesuatu sebagai positif.
5. *Recall*: Nilai Recall adalah sekitar 0.836. Recall, juga dikenal sebagai Sensitivitas, adalah rasio dari True Positives terhadap semua yang seharusnya positif. Ini mengukur sejauh mana model klasifikasi mampu mendeteksi semua kasus positif.
6. *MCC (Matthews Correlation Coefficient)*: Nilai MCC adalah sekitar 0.7954. MCC mengukur hubungan antara prediksi positif dan negatif serta kebenaran positif dan negatif. Nilai MCC berkisar antara -1 hingga 1, dengan 1 menunjukkan klasifikasi sempurna dan 0 menunjukkan klasifikasi acak.
7. *Log Loss (Logarithmic Loss)*: Nilai Log Loss adalah sekitar 0.4329. Log Loss mengukur ketidakpastian model dalam mengklasifikasikan instansi. Semakin rendah nilai Log Loss, semakin baik modelnya.
8. Secara keseluruhan, hasil "test and score" menunjukkan bahwa model regresi logistik ini memiliki kinerja yang sangat baik dengan nilai AUC yang tinggi dan akurasi yang baik (CA). F1-Score yang tinggi menunjukkan bahwa model ini baik dalam memprediksi kelas positif dan negatif. Namun, selalu penting untuk mempertimbangkan tujuan dan karakteristik masalah klasifikasi Anda saat mengevaluasi kinerja model.

DAFTAR PUSTAKA

- [1] cchanges, "Garbage Classification Dataset," <https://www.kaggle.com/datasets/asdasdasdas/garbage-classification>.
- [2] K. T. Maheswari, R. Srimathi, S. JaanaaRubavathy, P. Ponmuruga, K. Chandraprakash, and M. S. Dhayanithi, "Smart security and home automation using internet of things and artificial intelligence," *AIP Conf. Proc.*, vol. 2725, no. 1, p. 70002, Apr. 2023, doi: 10.1063/5.0125139.
- [3] L. G. Shapiro and G. C. Stockman, *Computer Vision*. New Jersey: Prentice-Hall, 2021.
- [4] J. Tolles and W. J. Meurer, "Logistic Regression: Relating Patient Characteristics to Outcomes," *JAMA*, vol. 316, no. 5, pp. 533–534, Aug. 2016, doi: 10.1001/jama.2016.7653.
- [5] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote Sens. Environ.*, vol. 62, no. 1, pp. 77–89, 1997, doi: [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7).