

# Perbandingan Algoritma *Machine Learning* menggunakan *Orange Data Mining* untuk Klasifikasi Jenis Kendaraan pada Sistem Tilang Digital

Egipta Pranadjaya<sup>1</sup>, Evan Sudira Pangestu<sup>2</sup>, Catherine Olivia Sereati<sup>3\*</sup>, Sandra Octaviani<sup>4</sup>, Marten Darmawan<sup>5</sup>

<sup>1,2,5</sup>Program Studi Teknik Mesin, Fakultas Teknik  
Universitas Katolik Indonesia Atma Jaya, Jakarta 12930, Indonesia

<sup>3,4</sup>Program Studi Teknik Elektro, Fakultas Teknik  
Universitas Katolik Indonesia Atma Jaya, Jakarta 12930, Indonesia

## Article Info

### Article history:

Received  
26 03 2024

Accepted  
24 04 2024

### Keywords:

Classification of Vehicle  
Types; Logistic  
Regression; Neural  
Network; Orange Data  
Mining; Support Vector  
Machine

## Abstract

*This paper discusses the application of the Orange Data Mining application to compare several machine learning algorithms for classifying vehicle types in digital ticket systems. This research compares and analyzes the logistic regression algorithm, Support Vector Machine (SVM) and Neural Network (NN) to solve vehicle classification problems in digital traffic tickets. The research results show that in the training process and based on the dataset used, the algorithms that have the highest level of accuracy are Logistic Regression, Neural Network and Support Vector Machine. Meanwhile, during the testing process, all algorithms in the model were able to carry out classification with 100% accuracy*

## Info Artikel

### Histori Artikel:

Diterima:  
26 03 2024

Disetujui:  
24 04 2024

### Kata Kunci:

Klasifikasi Jenis  
Kendaraan; Logistic  
Regression; Neural  
Network; Orange Data  
Mining; Support Vector  
Machine

## Abstrak

*Tulisan ini membahas tentang penerapan aplikasi Orange Data Mining untuk membandingkan beberapa algoritma pembelajaran mesin untuk mengklasifikasikan jenis kendaraan pada sistem tilang digital. Penelitian ini membandingkan dan menganalisa algoritma logistik regression, Support Vector Machine (SVM) dan Neural Network (NN) untuk memecahkan masalah klasifikasi kendaraan pada tilang digital. Data set diambil dari website Kaggle, dan metode validasi menggunakan Confussion matrix. Hasil penelitian menunjukkan bahwa pada proses training dan berdasarkan dataset yang digunakan, algoritma yang memiliki tingkat akurasi dari yang tertinggi adalah Logistic Regression, Neural Network dan Support Vector Machine. Sedangkan pada proses testing, seluruh algoritma pada model mampu melakukan klasifikasi dengan akurasi yang sempurna yaitu 100 %.*

## 1. PENDAHULUAN

Pada era digital ini terjadi perkembangan teknologi yang pesat dan beragam. Perkembangan teknologi digital didukung dengan teknologi kecerdasan artifisial atau *Artificial Intelligence* (AI), dapat memudahkan pekerjaan manusia di berbagai bidang. Pada sistem komputer teknologi AI yang sering digunakan adalah *Machine Learning*. *Machine learning* merupakan bagian dari AI yang membuat komputer berperilaku seperti manusia, yaitu meningkatkan pengetahuan dengan belajar dari pengalaman atau dengan berjalannya waktu secara otomatis [1]. Pada dasarnya *machine learning* lebih terfokus dalam peningkatan kinerja program komputer dalam mengolah data dan belajar dari data-data yang ada.

Salah satu contoh *tools machine learning* yang digunakan adalah *Orange Data Mining*. *Orange data mining* merupakan sebuah *software open source* yang memiliki fungsi untuk melakukan pengolahan *data analytics* atau *data mining*[2]. Pada *orange* terdapat banyak *widget* yang dapat diletakkan pada *drawing board* lalu dapat dihubungkan dengan banyak *widget* lainnya. *Software* ini memiliki keunggulan

\*Corresponding author.

Email address: [catherine.olivia@atmajaya.ac.id](mailto:catherine.olivia@atmajaya.ac.id)

dalam visualisasi atau *visual programming*. *Software* ini dapat juga digunakan untuk melakukan klasifikasi data berupa gambar. Sehingga *software* ini sangat membantu dalam penelitian ini untuk menganalisa gambar dan mengklasifikasi gambar [3].

Pada umumnya dengan adanya *software Orange Data Mining*, maka dapat melakukan sebuah pengklasifikasian [4]. Klasifikasi merupakan sebuah proses pengelompokan objek yang memiliki karakteristik atau ciri-ciri yang sama, lalu dimasukkan ke dalam beberapa kelas atau kelompok [5]. Pada penelitian ini dilakukannya klasifikasi gambar, yaitu pengelompokan gambar yang memiliki karakteristik atau ciri-ciri yang sama. Pada penelitian ini yang diklasifikasi adalah makanan, yaitu mengelompokkan makanan-makanan yang memiliki karakteristik atau ciri-ciri yang sama.

Pada kemajuan teknologi sekarang, sistem penilaian secara manual mulai ditinggalkan, dan mulai beralih ke sistem penilaian digital melalui kamera dan cctv untuk meringankan pekerjaan penegak hukum. Dengan adanya tilang digital, maka diharapkan mampu memecahkan masalah pungutan liar yang telah merajalela, sehingga masyarakat dapat terhindar dari pungutan liar yang meresahkan masyarakat [6]. Tilang digital sudah menjadi solusi yang cukup populer dan efisien dalam penegakan hukum lalu lintas, namun tidak dapat dipungkiri jika tilang digital memiliki beberapa kendala seperti, kemacetan lalu lintas yang menyebabkan pelanggaran tidak dapat tertangkap oleh kamera karena terhalang kendaraan lain, dan juga kesalahan dalam identifikasi jenis kendaraan.

Tulisan ini bertujuan untuk melakukan analisa perbandingan beberapa algoritma *machine learning* untuk klasifikasi jenis tilang. *Tools* yang digunakan untuk mengolah data adalah *Orange Data Mining*. Dengan melakukan analisis perbandingan ini diharapkan akan didapatkan solusi yang efektif dan inovatif untuk mengklasifikasi jenis kendaraan yang ada, sehingga nantinya dapat mengoptimalkan tilang digital.

## 2. LANDASAN TEORI

*Logistic regression* merupakan sebuah algoritma klasifikasi linear yang berfokus pada masalah klasifikasi multi kelas dan telah terbukti menghasilkan klasifikasi yang *powerful*. Algoritma ini sering digunakan ketika variabel respon mengarah pada dua nilai, contohnya subjek memiliki karakteristik khusus atau tidak memiliki karakteristik khusus. Keuntungan dari algoritma ini adalah algoritma ini telah dipelajari dan didalami secara ekstensif disamping perkembangannya tentang penerapan *truncate newton*[7].

*Support Vector Machines (SVM)* merupakan algoritma yang menggunakan ruang hipotesis dengan bentuk fungsi-fungsi linier dalam sebuah ruang fitur (feature space) yang memiliki dimensi tinggi dan dapat mengimplementasikan learning bias yang berasal dari teori pembelajaran statistik dengan pelatihan yang berasal dari algoritma pembelajaran. Algoritma ini juga banyak digunakan pada klasifikasi otomatis dengan menggunakan sebuah metode diskriminatif yang sangat cocok untuk digunakan pada masalah klasifikasi, yaitu dengan menggunakan structural risk minimization sebagai dasar untuk bekerja dalam pembelajaran mesin[8]. Keuntungan dari algoritma ini adalah algoritma ini dapat bekerja dengan baik untuk data-data yang tidak dapat dipisahkan dengan cara linear. Algoritma ini juga termasuk algoritma yang populer dan banyak digunakan untuk mengklasifikasikan dokumen.

Neural Network atau NN merupakan algoritma yang diciptakan karena terinspirasi oleh kinerja dari neuron pada otak manusia bekerja, dimana setiap neuron pada otak manusia saling terkoneksi dan setiap informasi yang ada mengalir melalui setiap neuron yang ada pada otak manusia [9]. Kemudian pada tahun 1943, sebuah model matematika yang dibentuk dari sebuah penyederhanaan yang berasal dari struktur sel saraf manusia, diperkenalkan oleh Mc. Culloch dan Pitts. Algoritma NN terdiri dari unit-unit yang dapat mengkombinasikan input, kemudian unit-unit tersebut dapat diklasifikasikan menjadi tiga layer sederhana, yaitu *input layer*, *output layer*, dan *middle layer* [10].

## 3. METODOLOGI PENELITIAN

Penelitian dilakukan menggunakan software Orange Data Mining untuk melakukan proses klasifikasi pada gambar. Gambar yang dipilih merupakan gambar dari beberapa jenis transportasi seperti sepeda, sepeda motor, sepeda motor skuter, mobil, bus, dan truk. Data set tersebut didapatkan dari situs *Kaggle* dan berisi tiga kategori data berupa training, dan testing. Data set dengan kategori training memiliki jumlah gambar yang paling banyak diantara yang lainnya, hal ini bertujuan untuk melakukan pelatihan terhadap model machine learning dengan jumlah data pembelajaran yang banyak sehingga mampu memberikan hasil sesuai dengan tujuan yang ingin dicapai, pada penelitian ini yaitu mampu mengklasifikasikan jenis makanan cepat saji. Dataset testing merupakan data akhir untuk menguji secara nyata apakah model yang dibuat telah dilatih secara benar menggunakan data yang sesuai. Semua gambar yang ada pada dataset ini nantinya akan dimasukkan kedalam software Orange Data Mining dan

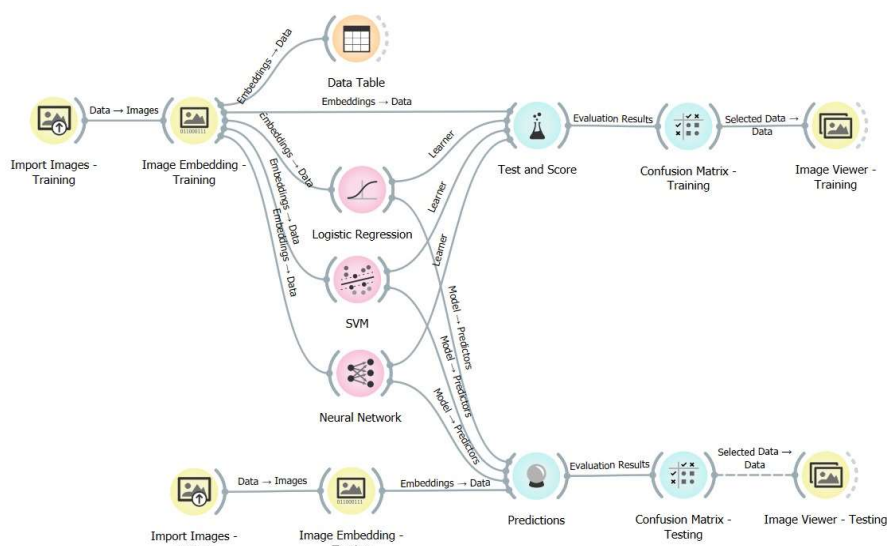
dilakukan proses *Image Embedding* terlebih dahulu, dikarenakan algoritma pada machine learning tidak dapat memproses gambar secara langsung. Proses *Image Embedding* ini akan menghasilkan suatu vektor yang merepresentasikan gambar tersebut.

Pada software Orange Data Mining, digunakanlah dua algoritma dan dua juga metode validasi. Algoritma yang digunakan adalah Logistic Regression dan juga Support Vector Machine. Metode validasi yang digunakan adalah Confusion Matrix dan prediction. Confusion matrix digunakan untuk melihat hasil klasifikasi dari model yang telah dibuat sehingga kita mampu melakukan penilaian terhadap kinerja dari model itu sendiri dengan cara melihat dimana model membuat kesalahan dan juga jenis kesalahannya. Sedangkan prediction digunakan untuk melakukan prediksi atas klasifikasi dari dataset testing nantinya berdasarkan algoritma yang telah dilatih.

Proses klasifikasi menggunakan software Orange Data Mining dimulai dengan menyiapkan dataset training untuk memulai pelatihan model dan memproses gambar yang ada dengan Image Embedding. Kemudian akan dilanjutkan dengan pengolahan data yang akan dilakukan oleh algoritma yang telah dipilih. Hasil dari algoritma dapat dilihat tingkat akurasi pada bagian *test & score*. Untuk melihat dimana letak kesalahan algoritma dalam melakukan klasifikasi gambar, maka digunakanlah bantuan Confusion Matrix dan juga Image Viewer. Proses ini merupakan proses pelatihan model. Untuk melakukan pengujian atau testing pada model yang telah dibuat, maka dataset testing harus disiapkan dan juga dilakukan proses Image Embedding. Selanjutnya, klasifikasi akan dilakukan oleh algoritma yang sebelumnya telah dibuat dan dilatih yang kemudian akan dihubungkan kepada metode validasi prediction untuk memprediksi hasil klasifikasi berdasarkan model yang telah dibuat. Untuk melihat performa dan juga letak kesalahan dari model, metode validasi berupa Confusion Matrix dan juga Prediction dapat ditambahkan.

#### 4. HASIL DAN PEMBAHASAN

Proses pelatihan dan juga pengujian model yang akan digunakan untuk klasifikasi nantinya akan dilakukan pada software Orange Data Mining. Terdapat beberapa langkah yang diperlukan agar model dapat dibuat dengan tujuan tertentu dan hasil yang maksimal. Secara keseluruhan, pada Gambar 1 merupakan rangkaian yang telah dibuat pada Orange Data Mining dalam membentuk suatu model.



Gambar 1. Pemodelan Orange Data Mining

Proses pelatihan model dimulai dengan memberikan dataset training ke dalam proses *Image Embedding*. Data set *training* terdiri atas gambar dari beberapa jenis transportasi seperti sepeda, sepeda motor, sepeda motor skuter, mobil, bus, dan truk. Hasil dari Image Embedding berupa tabel yang berisikan kategori, nama, ukuran, dan juga vektor dari gambar yang dapat dilihat pada Gambar 2.

id	category	image name	image	size	width	height	n0	n1	n2	n3	n4	n5	n6
1	bike	2Q_(1)	bike/2Q_(1).jpg	3313	100	100	0.146475	0.081787	0.132593	0.0130032	0.163899	0.456262	0.4
2	bike	2Q_(10)	bike/2Q_(10).jpg	8419	225	225	0.0199737	0.534995	0.069383	0.129696	0.280257	0.43248	0.2
3	bike	2Q_(2)	bike/2Q_(2).jpg	2815	100	100	0.15066	0.515355	0.0509276	0.0649397	0.11069	0.664263	0.4
4	bike	2Q_(3)	bike/2Q_(3).jpg	7600	283	178	0.188239	0.0484139	0.182873	0.0514845	0.0634647	0.494944	0.2
5	bike	2Q_(4)	bike/2Q_(4).jpg	11098	264	191	0.102041	0.0983159	0.148821	0.209999	0.0490858	0.252591	0.2
6	bike	2Q_(5)	bike/2Q_(5).jpg	10210	260	189	0.361221	0.267122	0.0998827	0.094422	0.20279	0.324236	0.2
7	bike	2Q_(6)	bike/2Q_(6).jpg	6361	259	194	0.105682	0.141173	0.344732	0.0592444	0.0960305	0.317113	0.1
8	bike	2Q_(7)	bike/2Q_(7).jpg	4993	270	183	0.125571	0.0549007	0.729198	0.149886	0.421131	0.287799	0.4
9	bike	2Q_(8)	bike/2Q_(8).jpg	8002	259	194	0.396353	0.474271	0.347039	0.0240661	0.131478	0.272374	0.
10	bike	2Q_(9)	bike/2Q_(9).jpg	10306	300	168	0.266325	0.320608	0.0656123	0.066326	0.0690781	0.745377	0.
11	bike	2Q_	bike/2Q_.jpg	2638	100	100	0.0845386	0.159908	0.130709	0.0586469	0.136753	0.311492	0.
12	bike	9k_(1)	bike/9k_(1).jpg	2627	100	100	0.114812	0.0693661	0.787307	0.0935103	0.139245	0.291434	0.5
13	bike	9k_(10)	bike/9k_(10).jpg	7865	259	194	0.235295	0.34526	0.15754	0.0154766	0.116783	0.203247	0.1
14	bike	9k_(11)	bike/9k_(11).jpg	9122	251	201	0.200086	0.355668	0.221786	0.0815166	0.189297	0.769634	0.3
15	bike	9k_(2)	bike/9k_(2).jpg	2777	100	100	0.323571	0.122212	0.139004	0.172983	0.0916815	0.268383	0.
16	bike	9k_(3)	bike/9k_(3).jpg	3056	100	100	0.0667729	0.166319	0.309619	0.314635	0.0337325	0.326227	0.2
17	bike	9k_(4)	bike/9k_(4).jpg	6875	286	176	0.204966	0.253146	0.192032	0.177829	0.37419	0.187234	0.4
18	bike	9k_(5)	bike/9k_(5).jpg	8306	225	225	0.12365	0.131886	0.076609	0.457735	0.0199485	0.288014	0.1
19	bike	9k_(6)	bike/9k_(6).jpg	6971	275	183	0.169505	0.018389	0.308999	0.0924429	0.265481	0.205092	0.3
20	bike	9k_(7)	bike/9k_(7).jpg	9911	225	225	0.493402	0.214598	0.153599	0.108568	0.102009	0.531761	0.08
21	bike	9k_(8)	bike/9k_(8).jpg	6963	225	225	0.238741	0.155466	0.0115185	0.303568	0.116242	0.326013	0.1
22	bike	9k_(9)	bike/9k_(9).jpg	8177	225	225	0.0810219	0.236149	0.370258	0.2338	0.333204	0.209066	0.3
23	bike	9k_	bike/9k_.jpg	3191	100	100	0.137487	0.0859161	0.229911	0.0386151	0.178487	0.497155	0.4
24	bike	images (1)	bike/images (1).jpg	2746	100	100	0.145422	0.285604	0.097022	0.0425125	0.0215208	0.34242	0.3
25	bike	images (10)	bike/images (10).jpg	3231	100	100	0.120413	0.10164	0.274309	0.158259	0.294541	0.165163	0.6
26	bike	images (11)	bike/images (11).jpg	3097	100	100	0.327246	0.269913	0.0778005	0.11216	0.0663871	0.45528	0.2
27	bike	images (12)	bike/images (12).jpg	2638	100	100	0.0845386	0.159908	0.130709	0.0586469	0.136753	0.311492	0.
28	bike	images (13)	bike/images (13).jpg	3522	100	100	0.263545	0.139217	0.224738	0.037802	0.107669	0.471772	0.
29	bike	images (14)	bike/images (14).jpg	3191	100	100	0.137487	0.0859161	0.229911	0.0386151	0.178487	0.497155	0.4
30	bike	images (15)	bike/images (15).jpg	8700	259	194	0.114439	0.188964	0.231296	0.283379	0.179893	0.122026	0.3
31	bike	images (16)	bike/images (16).jpg	7339	275	183	0.394814	0.155832	0.421313	0.0812343	0.147704	0.361434	0.
32	bike	images (17)	bike/images (17).jpg	10433	245	206	0.0900351	0.393884	0.182298	0.226313	0.27463	0.373451	0.1
33	bike	images (18)	bike/images (18).jpg	8284	287	176	0.195038	0.424041	0.183867	0.0795676	0.319001	0.188959	0.3
34	bike	images (19)	bike/images (19).jpg	8557	225	225	0.0521804	0.157271	0.633982	0.238467	0.162927	0.113797	0.
35	bike	images (2)	bike/images (2).jpg	3554	100	100	0.563915	0.149795	0.165255	0.111635	0.132245	0.482223	0.

Gambar 2. Hasil Kategorisasi *Image Embedding*

Setelah gambar pada dataset training telah mendapatkan properti tambahan melalui proses Image Embedding, maka proses akan dilanjutkan dengan klasifikasi menggunakan tiga buah algoritma yaitu Logistic Regression, Support Vector Machine, dan Neural Network. Tiga algoritma ini dipilih karena mampu menghasilkan klasifikasi dengan akurasi yang paling tinggi dibandingkan dengan algoritma lainnya. Untuk melihat perbandingan dari akurasi yang dihasilkan, maka kita dapat menggunakan Test and score yang dapat dilihat pada Gambar 3.

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	1.000	0.981	0.981	0.981	0.981	0.977
Neural Network	0.999	0.961	0.961	0.962	0.961	0.954
SVM	0.999	0.958	0.959	0.964	0.958	0.950

Gambar 3. Hasil Perbandingan akurasi

Dapat dilihat bahwa dari ketiga model yang digunakan, angka CA(Classification Accuracy) dari yang paling tinggi ke yang paling rendah adalah Logistic Regression (98,1%), Neural Network (96,1%), dan Support Vector Machine (95,8%). Perbedaan akurasi ini dapat ditimbulkan karena Logistic Regression cenderung membutuhkan fitur yang lebih sedikit dan tidak kompleks dan mampu mencegah model menjadi terlalu kompleks jika dibandingkan dengan Support Vector Machine. Neural Network mampu melakukan ekstraksi secara otomatis terhadap data penting dengan berbagai fitur yang terdapat pada gambar, dimana Support Vector Machine sulit atau membutuhkan langkah manual untuk melakukan hal yang sama.

Dikarenakan model yang dipilih tidak mampu menghasilkan tingkat akurasi yang sempurna (100%), maka tentu terdapat kesalahan didalamnya. Letak dari kesalahan tersebut dapat kita ketahui melalui metode validasi berupa Confusion Matrix. Pada Confusion Matrix, kita mampu melihat hasil klasifikasi yang telah dilakukan oleh model. Dapat dilihat pada Gambar 4 bahwa angka selain yang berada pada garis diagonal merupakan kesalahan klasifikasi yang dilakukan oleh model. Dapat dilihat bahwa sumbu-X pada Confusion Matrix merupakan kategori aktual dari gambar, sedangkan sumbu-Y pada Confusion Matrix merupakan prediksi klasifikasi terhadap kategori gambar yang dilakukan oleh model.

		Predicted						$\Sigma$
		bike	bus	car	cycle	scooty	truck	
Actual	bike	48	0	0	0	0	0	48
	bus	0	46	1	0	0	0	47
	car	0	0	55	0	0	1	56
	cycle	1	0	0	58	0	0	59
	scooty	1	0	0	0	57	0	58
	truck	0	2	0	0	0	40	42
$\Sigma$		50	48	56	58	57	41	310

(a)

		Predicted						$\Sigma$
		bike	bus	car	cycle	scooty	truck	
Actual	bike	48	0	0	0	0	0	48
	bus	0	44	1	0	0	2	47
	car	0	1	53	0	0	2	56
	cycle	1	0	0	58	0	0	59
	scooty	2	0	0	0	56	0	58
	truck	0	2	1	0	0	39	42
$\Sigma$		51	47	55	58	56	43	310

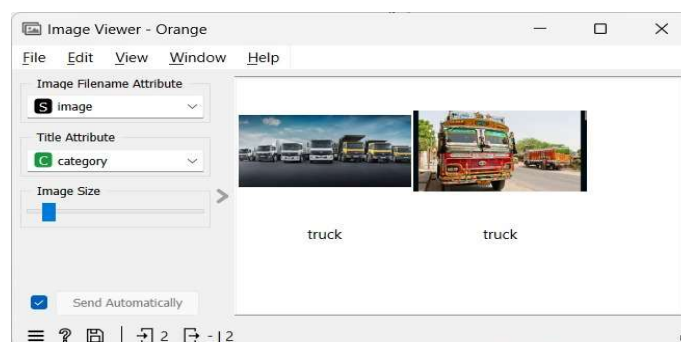
(b)

		Predicted						$\Sigma$
		bike	bus	car	cycle	scooty	truck	
Actual	bike	47	1	0	0	0	0	48
	bus	0	46	1	0	0	0	47
	car	0	2	54	0	0	0	56
	cycle	0	3	0	56	0	0	59
	scooty	1	1	0	0	56	0	58
	truck	0	4	0	0	0	38	42
$\Sigma$		48	57	55	56	56	38	310

(c)

Gambar 4 . Hasil Confusion Matrix (a) Logistic Regression (b) SVM (c) NN

Untuk mengetahui kesalahan model, digunakanlah referensi berdasarkan Confusion Matrix, dengan cara menganalisa prediksi yang error. Contoh kesalahan prediksi dapat dilihat pada hasil pemodelan *Logistic Regression*. Dapat dilihat bahwa terdapat dua buah gambar dengan kategori aktual yaitu truck yang diprediksi termasuk kedalam kategori bus. Untuk melihat gambar yang dimaksud, maka kita harus mengklik dua kali pada *cell* yang kita tuju lalu menggunakan bantuan *Image Viewer* maka gambar tersebut dapat muncul, seperti pada Gambar 5. Kesalahan tersebut dapat terjadi karena gambar kendaraan yang ada pada *data* hanya berfokus kepada tampak depan, dimana tampak depan ini memiliki kemiripan dengan gambar bus yang ada pada dataset sehingga membuat model keliru dalam mengkategorikannya



Gambar 5. Gambar hasil kesalahan prediksi dari Confusion Matrix

Langkah selanjutnya merupakan proses pengujian dengan dataset testing yang berisikan gambar dengan kategori yang serupa namun yang sebelumnya tidak pernah diberikan kepada model. Proses pengujian diawali dengan langkah yang sama yaitu *Image Embedding*. Setelah itu, proses klasifikasi akan menggunakan model dengan algoritma yang sama seperti yang sudah dilatih untuk menguji apakah model sudah berjalan dengan baik. Untuk metode validasi digunakanlah Prediction. Hal ini bertujuan agar model melakukan prediksi terhadap dataset yang diberikan berdasarkan dataset saat proses pelatihan. Hasil pada Predictions dapat dilihat pada Gambar 6 merupakan rincian bagaimana setiap algoritma memprediksi kategori dari setiap gambar yang ada pada dataset testing dan juga error yang terbaca oleh tiap-tiap algoritma. Dari hasil Predictions dapat dilihat bahwa akurasi klasifikasi yang dihasilkan oleh seluruh model, baik yang menggunakan algoritma Logistic Regression, Neural Network, dan yang menggunakan algoritma Support Vector Machine berada di angka 100%. Hal ini menunjukkan bahwa proses pelatihan model berjalan dengan baik, sehingga model mampu melakukan prediksi terhadap dataset testing dengan akurasi yang sempurna.

id	Logistic Regression	error	SVM	error	Neural Network	error	category	image name	Image	size	width	height	error
1	1.00:0.00:0.00:0.00:0.00:0.00	bike	0.005	0.96:0.09:0.00:0.00:0.02:0.00	bike	0.038	1.00:0.00:0.00:0.00:0.00:0.00	bike	0.000				0.00
2	0.94:0.00:0.00:0.02:0.03:0.00	bike	0.061	0.91:0.09:0.00:0.04:0.04:0.00	bike	0.067	1.00:0.00:0.00:0.00:0.00:0.00	bike	0.000				0.00
3	0.99:0.00:0.00:0.00:0.01:0.00	bike	0.010	0.97:0.09:0.00:0.00:0.03:0.00	bike	0.035	1.00:0.00:0.00:0.00:0.00:0.00	bike	0.000				0.00
4	0.96:0.00:0.00:0.00:0.04:0.00	bike	0.040	0.88:0.09:0.00:0.01:0.10:0.00	bike	0.121	1.00:0.00:0.00:0.00:0.00:0.00	bike	0.000				0.00
5	1.00:0.00:0.00:0.00:0.00:0.00	bike	0.001	0.99:0.00:0.00:0.01:0.00:0.00	bike	0.010	1.00:0.00:0.00:0.00:0.00:0.00	bike	0.000				0.00
6	1.00:0.00:0.00:0.00:0.00:0.00	bike	0.003	0.96:0.00:0.00:0.01:0.02:0.00	bike	0.037	1.00:0.00:0.00:0.00:0.00:0.00	bike	0.000				0.00
7	1.00:0.00:0.00:0.00:0.00:0.00	bike	0.004	0.92:0.01:0.02:0.03:0.01:0.01	bike	0.075	1.00:0.00:0.00:0.00:0.00:0.00	bike	0.000				0.00
8	0.00:0.87:0.00:0.00:0.00:0.13	bus	0.133	0.01:0.76:0.03:0.02:0.01:0.17	bus	0.241	0.00:0.94:0.00:0.00:0.00:0.06	bus	0.055				0.00
9	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.000	0.00:0.99:0.00:0.00:0.00:0.00	bus	0.014	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.000				0.00
10	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.001	0.01:0.97:0.01:0.01:0.01:0.01	bus	0.033	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.000				0.00
11	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.002	0.01:0.97:0.00:0.01:0.01:0.01	bus	0.030	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.000				0.00
12	0.00:0.98:0.00:0.00:0.00:0.02	bus	0.021	0.01:0.91:0.01:0.01:0.01:0.05	bus	0.093	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.000				0.00
13	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.001	0.01:0.93:0.02:0.01:0.01:0.01	bus	0.072	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.000				0.00
14	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.000	0.00:0.98:0.00:0.00:0.00:0.00	bus	0.018	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.000				0.00
15	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.000	0.01:0.95:0.01:0.01:0.01:0.01	bus	0.048	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.000				0.00
16	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.001	0.02:0.83:0.06:0.03:0.03:0.04	bus	0.174	0.00:1.00:0.00:0.00:0.00:0.00	bus	0.000				0.00
17	0.00:0.99:0.00:0.00:0.00:0.00	car	0.007	0.01:0.96:0.01:0.01:0.01:0.01	car	0.035	0.00:1.00:0.00:0.00:0.00:0.00	car	0.000				0.00
18	0.00:0.90:1.00:0.00:0.00:0.00	car	0.004	0.01:0.91:0.96:0.01:0.00:0.01	car	0.035	0.00:0.00:1.00:0.00:0.00:0.00	car	0.000				0.00
19	0.00:0.00:0.99:0.00:0.00:0.00	car	0.009	0.02:0.06:0.85:0.02:0.02:0.03	car	0.148	0.00:0.00:1.00:0.00:0.00:0.00	car	0.000				0.00
20	0.00:0.00:1.00:0.00:0.00:0.00	car	0.001	0.01:0.01:0.97:0.01:0.00:0.00	car	0.028	0.00:0.00:1.00:0.00:0.00:0.00	car	0.000				0.00
21	0.00:0.00:1.00:0.00:0.00:0.00	car	0.002	0.02:0.01:0.94:0.01:0.01:0.01	car	0.060	0.00:0.00:1.00:0.00:0.00:0.00	car	0.000				0.00
22	0.00:0.00:1.00:0.00:0.00:0.00	car	0.000	0.01:0.02:0.94:0.01:0.01:0.01	car	0.064	0.00:0.00:1.00:0.00:0.00:0.00	car	0.000				0.00
23	0.00:0.00:1.00:0.00:0.00:0.00	car	0.001	0.02:0.04:0.89:0.02:0.02:0.02	car	0.113	0.00:0.00:1.00:0.00:0.00:0.00	car	0.000				0.00
24	0.03:0.02:0.93:0.00:0.00:0.02	car	0.066	0.04:0.16:0.66:0.03:0.04:0.07	car	0.242	0.00:0.00:1.00:0.00:0.00:0.00	car	0.000				0.00
25	0.00:0.00:0.00:1.00:0.00:0.00	cycle	0.001	0.01:0.02:0.02:0.92:0.01:0.01	cycle	0.076	0.00:0.00:0.00:1.00:0.00:0.00	cycle	0.000				0.00
26	0.00:0.00:0.00:1.00:0.00:0.00	cycle	0.003	0.01:0.01:0.01:0.94:0.01:0.01	cycle	0.056	0.00:0.00:0.00:1.00:0.00:0.00	cycle	0.000				0.00
27	0.00:0.00:0.00:1.00:0.00:0.00	cycle	0.001	0.01:0.00:0.00:0.98:0.00:0.00	cycle	0.020	0.00:0.00:0.00:1.00:0.00:0.00	cycle	0.000				0.00

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	1.000	1.000	1.000	1.000	1.000	1.000
SVM	1.000	1.000	1.000	1.000	1.000	1.000
Neural Network	1.000	1.000	1.000	1.000	1.000	1.000

Gambar 6. Hasil Prediction

## 5. SIMPULAN

Penelitian yang telah dilakukan menggunakan software *Orange Data Mining* untuk mengklasifikasikan gambar dari beberapa jenis kendaraan yang memiliki kategorinya masing-masing pada dataset yang ada. Hasil penelitian menunjukkan bahwa pada proses *training* dan berdasarkan dataset *training* yang digunakan, algoritma yang memiliki tingkat akurasi tertinggi dimiliki oleh *Logistic Regression* di angka 98,1% yang kemudian diikuti dengan *Neural Network* di 96,1% dan *Support Vector Machine* di 95,8%. Kesalahan klasifikasi yang dilakukan oleh model umumnya terjadi dikarenakan adanya kemiripan dari segi bentuk yang ada pada gambar. Hal ini dapat diminimalisasi dengan meningkatkan jumlah gambar pada dataset training sehingga algoritma pada model memiliki lebih banyak data yang akan menjadi referensi nantinya. Walaupun demikian, pada proses testing, seluruh model mampu melakukan klasifikasi berdasarkan kategori yang menjadi target dengan tingkat akurasi 100%. Tiga algoritma ini dipilih berdasarkan hasil pengujian karena mampu memberikan tingkat akurasi yang paling tinggi dibandingkan dengan yang lainnya. Meskipun demikian, penggunaan algoritma ini tidak dapat menjadi acuan pasti untuk penggunaan data yang berbeda, dikarenakan kinerja dari algoritma akan bergantung kepada data yang digunakan.

**DAFTAR PUSTAKA**

- [1] C. O. Sereati, H. R. W. I, C, L. W. Pandjaitan, and M. A. Kartwidjaja, "Music Genre Classification Using Machine Learning Method," *J. Ilm. Tek. Elektro*, vol. 21, no. 1, pp. 71–82, 2023, doi: <https://doi.org/10.25105/jetri.v21i1.15339>.
- [2] A. Sentiya and H. Suroyo, "ANALISIS TEXT CLUSTERING AKUN FANPAGE SHOPEE INDONESIA DENGAN KOMENTAR FOLLOWERS MENGGUNAKAN TOOLS ORANGE DATA MINING," in *Bina Darma Conference on Computer Science (BDCCS)*, 2019, pp. 1055–1067.
- [3] D. Vaishnav and B. R. Rao, "Comparison of Machine Learning Algorithms and Fruit Classification using Orange Data Mining Tool," in *2018 3rd International Conference on Inventive Computation Technologies (ICICT)*, 2018, pp. 603–607, doi: 10.1109/ICICT43934.2018.9034442.
- [4] S. Mittal, C. Monga, A. Bansal, and N. Singla, "Analyzing Data Scientist Salaries Dataset Through Machine Learning Algorithms Using Tool 'Orange,'" in *2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, 2023, pp. 1–6, doi: 10.1109/ICCUBEA58933.2023.10392037.
- [5] N. I. Widiastuti, E. Rainarli, and K. E. Dewi, "Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen," *J. INFOTEL (Informatics, Telecommunication, Electron.)*, vol. 9, no. 4, pp. 416–421, 2017, doi: <https://doi.org/10.20895/infotel.v9i4.312>.
- [6] B. Bakri, M. S. Nuh, and B. Badaru, "Efektifitas Penerapan Program E-Tilang Dalam Mewujudkan Penegakan Hukum Lalu Lintas," *J. LEX THEORY*, vol. 1, no. 1, pp. 82–98, 2020, doi: <https://doi.org/10.52103/jlt.v1i1.46>.
- [7] R. Prasetyo, I. Nawawi, A. Fauzi, and G. Ginabila, "Komparasi Algoritma Logistic Regression Dan Random Forest Pada Prediksi Cacat Software," *J. Tek. Inform. Unika St. Thomas*, vol. 6, no. 2, pp. 275–281, 2021, doi: 10.54367/jtiust.v6i2.1522.
- [8] B. Prihanto, C. O. Sereati, M. A. Kartawidjaja, and M. Siregar, "Attrition Analysis using XG Boost and Support Vector Machine Algorithm," *Int. J. Innov. Sci. Res. Technol.*, vol. 8, no. 6, pp. 2096–2112, 2023, doi: 10.38124/ijisrt.
- [9] M. S. Wibawa, "A Comparison Study Between Deep Learning and Conventional Machine Learning on White Blood Cells Classification," in *2018 International Conference on Orange Technologies (ICOT)*, 2018, pp. 1–6, doi: 10.1109/ICOT.2018.8705892.
- [10] M. Cheung, J. She, and L. Liu, "Deep learning-based online counterfeit-seller detection," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPs)*, 2018, pp. 51–56, doi: 10.1109/INFOCOMW.2018.8406896.