

IMPLEMENTASI MULTI-CLASS GRADIENT BOOSTING UNTUK MENGLASIFIKASIKAN JENIS HEWAN PADA KEBUN BINATANG

Sri Diantika¹⁾, Hiya Nalatissifa²⁾, Riki Supriyadi³⁾, Nurlaelatul Maulidah⁴⁾ dan Ahmad Fauzi⁵⁾

^{1,2,4,5)} Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

⁴⁾ Fakultas Teknologi Informasi, Universitas Nusa mandiri

¹⁾Jalan Kramat Raya No 98, Kwitang, Kec Senen, Kota Jakarta Pusat, Indonesia, kode pos: 10450

³⁾ Jl. Raya Jatiwaringin No.2, Cipinang Melayu, Makasar, Jakarta Timur, Daerah Khusus Ibukota Jakarta 13620

e-mail: sri.szd@bsi.ac.id¹⁾, hiya.hys@bsi.ac.id²⁾, riki.rsd@nusamandiri.ac.id³⁾, nurlaelatul.nlt@bsi.ac.id⁴⁾, ahmad.fzx@bsi.ac.id⁵⁾

Abstrak : Hewan merupakan salah satu makhluk hidup yang memiliki berbagai macam jenis. mengelompokkan jenis hewan berdasarkan kesamaan dan perbedaan ciri yang dimiliki menjadi salah satu kegiatan yang penting dilakukan guna mempermudah membandingkan, mengenali, mempelajari jenis hewan tertentu serta dapat mengetahui hubungan kekerabatan antar hewan, sehingga jika ditemukan hewan jenis baru yang belum memiliki nama maka kita kan lebih mudah memberikan nama pada hewan tersebut berdasarkan jenis dan berdaasarkan kelompoknya. Pada penelitian klasifikasi jenis hewan pada kebun binatang yang memiliki multi-class, klasifikasi terbaik didapatkan dengan menerapkan parameter gradient boosting dengan $n_estimators$ sebanyak 50, max_depth 3, sub_sample sebanyak 1.0, $learning_rate$ sebesar 0.1, dan menggunakan $criterion$ friedman Mse . Penelitian ini menerapkan teknik pembagian data $split_validation$ atau pembagian antara data training dan data testing, untuk besarnya adalah 80% untuk data training dan 20% untuk data testing. Hasil penelitian menyatakan bahwa model yang diusulkan yaitu model gradient boosting lebih baik dari beberapa model lain yang juga telah diuji dengan nilai akurasi sebesar 93.75%, $recal$ 94%, $presisi$ sebesar 96% dan MSE untuk mengukur besarnya rata-rata kesalahan dalam serangkaian klasifikasi sebesar 12,5%, semakin kecil nilai MSE maka akan semakin baik dalam mengklasifikasikan.

Kata Kunci—Klasifikasi, Jenis Hewan, Multi-class, Gradient Boosting, Algoritma

Abstract : Animals are one of the living things that have various types. Grouping types of animals based on similarities and differences in characteristics owned is one of the important activities carried out To make it easier to compare, recognize, study certain types of animals and be able to find out kinship relationships between animals, So if a new type of animal is found that does not yet have a name, it will be easier for us to give a name to the animal based on the type and based on the group. In research on the classification of animal species in zoos that have multi-class, the best classification is obtained by applying gradient boosting parameters with $n_estimators$ of 50, max_depth 3, sub_sample of 1.0, $learning_rate$ of 0.1, and using $criterion$ friedman Mse . And by implementing $Split_validation$ or division between training data by 80% for training data and 20% for testing data. The results stated that the proposed model was better than some other models that had also been tested with an accuracy value of 93.75%, $recal$ of 94%, $precision$ of 96% and MSE to measure the average magnitude of error in a series of classifications of 12.5%, the smaller the MSE value, the better it would be in classifying.

Keywords—Classification, Animal Type, Gradient Boosting Algorithm

I. PENDAHULUAN

KEBUN BINATANG merupakan sebuah tempat untuk mengkonservasi berbagai jenis binatang, kebun binatang memiliki visi yang sangat penting yaitu untuk melindungi dan melestarikan berbagai macam hewan agar masyarakat mendapatkan pembelajaran dan edukasi yang aman dan sehat mengenai dunia Fauna [1]

Beberapa hal yang diperlukan untuk mempelajari dunia fauna ini adalah salah satunya dengan cara menggolongkan atau mengklasifikasikan jenis hewan agar mempermudah proses mempelajari hewan serta mengetahui hubungan kekerabatan antara hewan hingga kita juga dapat membedakan ciri-ciri antar hewan yang satu dengan yang lainnya dan juga dapat dengan mudah memberikan nama hewan yang belum diketahui berdasarkan pengelompokannya.

Pengelompokan atau klasifikasi merupakan salah satu teknik pengolahan dalam data mining yang digunakan untuk mengelompokkan suatu objek yang memiliki karakteristik atau ciri yang sama ke dalam beberapa kelas [2]. Pengelompokan atau klasifikasi hewan ini dapat didasarkan pada karakteristik maupun kriteria yang dimiliki oleh masing-masing hewan, antara lain seperti berdasarkan makanannya, habitat, berdasarkan ada atau tidak adanya tulang belakang, alat geraknya sampai pada cara berkembang biak pada hewan tersebut.

Penelitian mengenai klasifikasi hewan pada kebun binatang juga pernah dilakukan sebelumnya oleh K. Nugroho and S. Murdowo [3] Penelitian dilakukan menggunakan algoritma Neural Network untuk melakukan prediksi jenis hewan yang terdapat pada kebun binatang mendapatkan akurasi sebesar 98% dengan tingkat presisi sebesar 98.1% dan Recall sebesar 98 %. Kemudian pada penelitian lain, Hewan juga dapat diidentifikasi menggunakan citra atau gambar, seperti pada penelitian [4] pengklasifikasian jenis hewan yang dilakukan menggunakan pendekatan segmentasi dan tanpa segmentasi sebagai tahapan awal, Metode K-Nearest Neighbors (K-NN) dipilih guna mengklasifikasi jenis hewan mamalia berdasarkan fitur HOG dari citra saliency, hasilnya Nilai k yang menghasilkan klasifikasi baik yaitu k=5

Berdasarkan penelitian tersebut, peneliti memilih menerapkan data mining menggunakan algoritma *multi-class gradient boosting* untuk mengklasifikasikan jenis hewan pada kebun binatang berdasarkan karakteristik dan kriteria yang dimiliki hewan. Metode *gradient boosting* dipilih karena dalam beberapa studi kasus dapat dengan baik mengatasi data yang memiliki banyak kelas atau *multi-class*. Untuk dataset yang digunakan merupakan dataset publik yang kemudian datanya ditambahkan dengan data pribadi yang dimiliki peneliti. Untuk judul penelitian yang peneliti ambil adalah “Implementasi Algoritma *Multi-Class Gradient Boosting* Untuk Mengklasifikasikan Jenis Hewan Pada Kebun Binatang”

II. TINJAUAN PUSTAKA

Tinjauan pustaka yang dijabarkan oleh peneliti dalam penelitian ini menggunakan beberapa referensi, seperti dari buku, prosiding dan jurnal. Berikut tinjauan pustaka yang mendukung teoritis dari penelitian ini

A. Data Mining

Proses analitis untuk meninjau kumpulan data dengan tujuan untuk menemukan adanya hubungan atau pola yang tidak terduga kemudian meringkas data tersebut dengan cara berbeda dengan sebelumnya agar dapat dipahami dan berguna bagi pemilik dari data, hal ini dapat dikatakan sebagai data mining [5] Data mining melakukan eksplorasi pada basis data guna menemukan pola yang tersembunyi, mencari informasi pemprediksi yang dimungkinkan terlewatkan atau terlupakan oleh para pelaku bisnis karena terletak diluar perkiraan mereka [6]

Teknik atau metode yang bisa digunakan dalam proses data mining jumlahnya sangatlah bervariasi, sehingga pemilihan teknik, algoritma atau metode yang pas dan sesuai sangat bertumpu pada proses dan tujuan *Knowledge discovery in database* (KDD) secara menyeluruh. *Data mining* dilakukan dengan melewati 3 tahap [7]:

1) Eksplorasi

Sebelum melakukan penelitian Hal pertama yang dilakukan yaitu menyiapkan data dalam jumlah yang banyak, kemudian dilakukan proses *preprocessing* terhadap data agar sinkron dengan kebutuhan penelitian, hal ini juga dilakukan guna membuang data yang berjumlah ganda. Data yang tersisa setelah melewati proses *preprocessing* ini adalah data yang benar-benar dapat digunakan sebagai objek penelitian.

2) Pemodelan atau identifikasi pola

Pemodelan merupakan bagian penting dari proses data mining yang memiliki tujuan untuk mengevaluasi sehingga akan menghasilkan mana prediksi yang akurat dan terbaik. Proses ini biasanya membutuhkan waktu yang lama karena menerapkan bisa jadi menerapkan beberapa model yang berbeda pada kumpulan data yang sama dan dilakukan berulang untuk nenatinya dibandingkan hasilnya dan dipilih mana yang lebih baik.

3) Penerapan

Untuk mengetahui apakah model yang dibangun dapat dengan baik melakukan suatu prediksi maka dilakukan pengujian model terhadap data pembelajaran dan data pengujian.

B. Klasifikasi

Klasifikasi merupakan sebuah teknik dari data mining guna memetakan sebuah basis data membentuk kelompok atau kelas yang sudah ditentukan. Klasifikasi digolongkan ke dalam metode *supervised learning*, yaitu sebuah metode yang memerlukan data uji atau *training* yang memiliki label untuk menghasilkan suatu aturan agar dapat mengklasifikasikan data uji atau *training* tersebut masuk kedalam kelompok atau kelas yang sudah ditentukan tadi [8]. Klasifikasi adalah proses yang dilakukan untuk menemukan model atau pola untuk menggolongkan suatu data ke dalam kelompok atau kelas data tertentu, sehingga klasifikasi dalam data mining ini dapat digunakan untuk memperkirakan atau memprediksi suatu objek atau data yang belum diketahui kelasnya atau kelompoknya [9]

Algoritma klasifikasi yang saat ini banyak digunakan data mining, yaitu *Decision* atau *classification trees*, *analisa statistic*, *Bayesian classifiers* atau sering disebut *Naïve Bayes classifiers*, *k-nearest neighbor*, *Neural networks*, *Algoritma Genetika*, *Rough sets*, *Memory based reasoning* kemudian ada Metode *Rule Based* serta *Support vector machines (SVM)*

C. Multi-Class

Multi-class data merupakan kondisi dimana sebuah dataset memiliki lebih dari dua label atau kelas (*multi-class*) [10] Klasifikasi *multi-class* merupakan kegiatan mengklasifikasikan setiap data pada basis data kedalam kelas yang berbeda, dimana terdapat lebih dari 2 kelas data. Terdapat beberapa pendekatan pengklasifikasian *multi-class*, diantaranya yaitu *One Against One* dan *One Against All* [10]

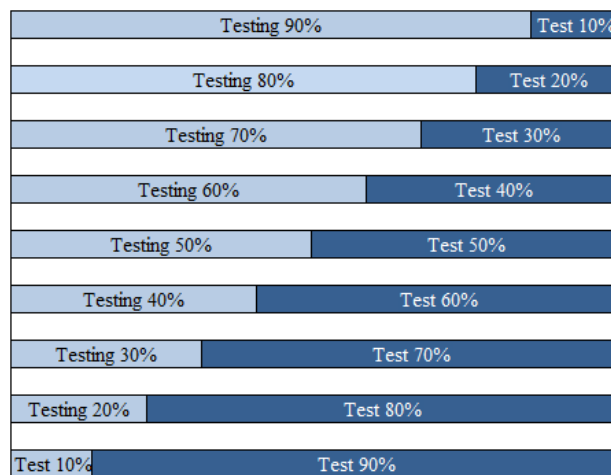
D. Gradient Boosting

Boosting adalah jenis dari ensemble learning yang dapat mengubah suatu model yang lemah (*weak Classifier*) sehingga dapat menjadi model yang kuat (*strong classifier*). Beberapa algoritma *boosting* yang populer diantaranya: *extreme gradient boosting (XGBoost)*, *Light gradient boosting machine (LightGBM)*, *Adaptive voosting (Adaboost)* serta *gradient boosting machine (GBM)* [11]

Gradient boosting merupakan salah satu algoritma yang menggunakan teknik *ensamble* dari *decision tree* dan termasuk ke dalam *supervised learning* [12] Cara kerja dari Algoritma *gradient boosting* adalah bergantian secara urut atau sekuensial menambahkan prediktor sebelumnya yang dinilai kurang tepat dengan prediksi ke *ensemble*, kemudian memerikasa apakah kesalahan yang ada sebelumnya sudah diperbaiki atau belum. *Gradient boosting* memulai tahapannya dengan membangun pohon klasifikasi awal kemudian melakukan iterasi secara berulang dan memperbaiki pohon klasifikasi sebelumnya dengan membuat model baru. Proses ini dilakukan sampai model yang dihasilkan memenuhi kriteria tertentu [13] Kelebihan dari algoritma *Gradient boosting* diantaranya dapat digunakan untuk semua jenis basis data, memiliki ketahanan ketika menangani atau menyelesaikan data pencilan (*outlier*) dan memiliki hasil prediksi yang baik [14]

E. Split Validation

Split validation merupakan sebuah proses membagi (*split*) basis data secara acak menjadi beberapa bagian seperti bagian data untuk pembelajaran dan data pengujian untuk mengevaluasi model yang diusulkan. *Split validation* juga dapat melakukan validasi untuk mengukur performa dari algoritma atau metode yang digunakan. Pengukuran performa pada umumnya dilakukan untuk menilai seberapa besar nilai akurasi sebuah model dalam memprediksi data sesuai dengan kelas data yang sebenarnya [15]



Gambar 1. Ilustrasi split validation

F. Confussion Matrix

Untuk mengetahui performa sebuah model, peneliti dapat melihat pada *Confusion matrix*, karena di dalam sebuah *Confusion matrix* berisi banyaknya baris data pengujian yang diprediksi benar dan tidak benar oleh model penelitian yang digunakan, tabel ini diperlukan untuk menentukan kinerja suatu model klasifikasi agar dapat ditentukan mana model yang memiliki performa terbaik [16] Dalam Confussion Matrix ada 4 istilah yang dapat merepresentasikan hasil dari klasifikasi, yaitu True Negatif (TN) merupakan jumlah data negatif yang dideteksi dengan benar, *False Negatif* (FN) merupakan jumlah data positif terdeteksi sebagai data *negative*, *True Positif* (TP) merupakan nilai positif yang dideteksi dengan benar dan *False Postif* (FP) merupakan data *negatif* namun terdeteksi secara *positif* [16]

Tabel 1. Pengujian confusion matrix

Klasifikasi		Hasil Prediksi	
		Benar	Salah
Data Actual	Benar	<i>True Positif</i> (TP)	<i>False Negative</i> (FN)
	Salah	<i>False Positif</i> (FP)	<i>True Negatif</i> (TN)

Dengan data confusion matrix, maka akan didapatkan sebuah data yang lain yang pastinya akan sangat berguna untuk mengukur performa sebuah algoritma atau model yang digunakan, adapun data tersebut antarlain:

1) Akurasi

Sebuah Parameter yang menunjukkan seberapa tepat algoritma mengklasifikasikan data yang ada, persamaan yang digunakan untuk menghitung nilai akurasi [16]

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

2) Presis

Tingkat keakuratan hasil dari klasifikasi dengan jumlah total pengenalan yang dilakukan sistem. Untuk Formula menghitung nilai presisi dapat dituliskan menggunakan rumus persamaan berikut:

$$\text{Presisi} = \frac{\text{True positif (TP)}}{\text{True positif (TP)+False positif (FP)}}$$

3) Recall

Recall menunjukkan total data yang betul diklasifikasi dalam sebuah kelas dibagi dengan jumlah total dalam kelas tersebut. Untuk Formula menghitung nilai *recall* dapat dituliskan menggunakan rumus persamaan berikut:

$$\text{Recall} = \frac{\text{True positif (TP)}}{\text{True positif (TP) + False negatif (FN)}}$$

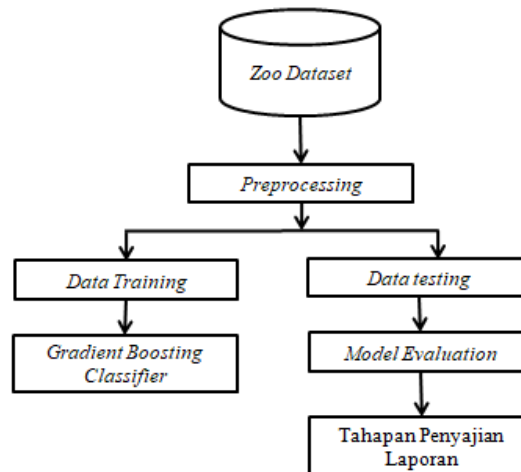
4) *F1-Score*

F1 score digunakan untuk menilai rata-rata *precision* dan *recall* hasil klasifikasi. Perhitungan *F1 Score* dapat dinyatakan dalam bentuk formula sebagai berikut:

$$F1\ score = 2 * \frac{(Recall * Precision)}{(Recall + Precision)}$$

III. METODE PENELITIAN

Pada bagian metode penelitian ini akan diuraikan secara detail, mulai dari dataset yang digunakan pada penelitian hingga bagaimana alur atau tahapan dari penelitian klasifikasi jenis hewan pada kebun binatang. Untuk digram alur penelitian dapat dilihat pada gambar 2 berikut ini:



Gambar 2. Diagram alur penelitian

A. *Dataset*

Dataset atau basis data yang dijadikan sebagai bahan atau objek dari penelitian ini adalah data hewan pada kebun binatang yang didapat dari data publik dengan url <https://www.kaggle.com/uciml/zoo-animal-classification>. *Dataset* tersebut berisi kumpulan kumpulan hewan yang ada di kebun binatang. Jumlah data sebanyak 101 kemudian ditambahkan dengan data privat sehingga jumlah data yang digunakan dalam penelitian ini adalah 130 data hewan. Setiap sampel memiliki 18 parameter dan label kelas yang mengidentifikasinya ada sebanyak 7 kelas data, yaitu jenis mamalia, burung, reptile, ikan, amfibi, serangga dan kelas data invertebrata.

B. *Preprocessing*

Preprocessing data merupakan suatu kegiatan yang harus dilakukan dalam proses data mining sebelum tahap klasifikasi dimulai, hal ini dilakukan dikarenakan tidak semua data atau atribut data dalam data digunakan dalam proses data mining. Proses ini juga dilakukan agar data yang akan digunakan sesuai dengan kebutuhan penelitian [17] Adapun tahapannya seperti melakukan pengisian data yang kosong, membuang data yang ganda atau *double*, memeriksa ketidakkonsistenan data, pembersihan data serta membenahi kesalahan yang mungkin terdapat pada data

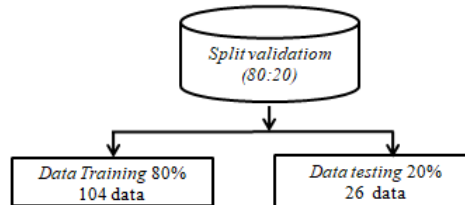
Salah satu proses *preprocessing* data yang dilakukan dalam penelitian ini adalah proses *cleaning*, proses ini dijalankan untuk membersihkan data yang nilainya hilang (*missing value*) atau *noise* pada data. Proses *cleaning* ini dilakukan dengan menjalankan perintah `fillna ()` yang terintegrasi pada alat uji yang digunakan, kemudian membersihkan data yang ganda.

Tools atau *executable document* yang digunakan dalam penelitian ini untuk menulis program adalah *google colabration* menggunakan bahasa pemrograman Python serta dengan menerapkan model *gradient boosting classifier*.

C. Split validation

Langkah yang selanjutnya dilakukan setelah melalui tahap *preprocessing* adalah membagi data menjadi 2 bagian yaitu satu bagian dijadikan sebagai data *training* dan yang lainnya sebagai data *testing*. Data *training* ini digunakan untuk mengajarkan model sedangkan data *testing* digunakan untuk memverifikasi model yang telah dibangun.

Setelah melakukan *preprocessing* banyaknya *instance* yang dapat digunakan dalam penelitian ini adalah sebanyak 130 data, kemudian data di bagi menjadi 2 bagian, yaitu 80% data untuk mengajarkan model mengenal data dan 20% untuk menguji model apakah sudah benar dalam proses mengenali suatu data, dalam penelitian ini didapatkan data *training* sebanyak 104 data dan data *testing* sebanyak 26 data.



Gambar 5 Proses Split Validation

D. Modelling

Pada penelitian klasifikasi jenis hewan pada kebun binatang ini, peneliti mengusulkan model algoritma *gradient boosting classifier*. Pada model *gradient boosting classifier* menggunakan pendekatan iteratif, di mana setiap iterasi bertujuan untuk meningkatkan model sebelumnya dengan menambahkan model baru. Proses tersebut dilakukan secara berulang hingga model yang dihasilkan memenuhi kriteria tertentu, seperti nilai loss function yang cukup kecil. *Parameter* yang digunakan dalam model *gradient boosting classifier* ini dapat dilihat pada tabel 2

Tabel 2. Parameter Gradient Boosting Classifier

Parameter	Nilai
<i>N Estimator</i>	50
<i>Max Depth</i>	3
<i>Sub Sample</i>	1.0
<i>Learning Rate</i>	0.1
<i>Criterion</i>	Friedman Mse
<i>Min Samples Split</i>	2
<i>Min Samples Leaf</i>	1
<i>Min Weight Fraction Leaf</i>	11
<i>Random State</i>	None
<i>Verbose</i>	0
<i>Max Features</i>	None

E. Evaluation

Evaluasi yang digunakan untuk menguji kinerja hasil klasifikasi adalah dengan confusion matrices. Dengan confusion matrices ini maka dapat diketahui kebenaran data prediksi terhadap data aktual. Serta dapat diperoleh nilai MSE, Recall, dan F1-Score, serta Akurasi.

IV. HASIL DAN PEMBAHASAN

Untuk menunjukkan bahwa model yang direkomendasikan mempunyai kinerja yang signifikan dalam mengklasifikasikan website phishing, maka pada subbab ini akan ditampilkan hasil pengujian menggunakan beberapa model dan menggunakan teknik resampling yang sama. Perbandingan ini dilakukan untuk meyakinkan bahwa model yang diusulkan adalah model yang terbaik. Perbandingan beberapa metode tersebut diukur menggunakan kinerja metode klasifikasi yang meliputi akurasi, recall, f1-score dan ROC.

A. Akurasi Gradient Boosting Classifier

Untuk mengetahui seberapa baik model dalam mengklasifikasikan dengan benar jenis hewan sesuai dengan kelasnya dibandingkan dengan keseluruhan data, maka bisa kita lihat dari besaran nilai akurasi yang dihasilkan dari model yang diusulkan. Pada tabel 3 akan di paparkan hasil pengujian model gradient boosting classifier menggunakan beberapa *learning rate* dan variasi pembagian data atau split data untuk menyakinkan bahwa model yang diusulkan merupakan model yang paling cocok dengan data. Penelitian dilakukan dengan mencoba beberapa *learning rate* dan variasi *split data* yang kemudian diuji menggunakan model *gradient boosting classifier*, didapatkan nilai akurasi tertinggi sebesar 93,75%. Nilai ini didapatkan dari parameter *learning rate* sebesar 0,1 dan pembagian data sebesar 20% untuk *data testing* dan 80% untuk *data training*.

Tabel 3. Hasil Pengujian Nilai Akurasi

Akurasi				
Data training	Data Testing	Learning Rate		
		0.01	0.1	1
50%	50%	81.57%	86.84%	81.57%
60%	40%	74.2%	87.09%	80.64%
70%	30%	73.91%	82.60%	86.95%
80%	20%	68.75%	93.75%	87.5%

B. Recall Gradient Boosting Classifier

Untuk mengetahui seberapa *actual* model dalam mengklasifikasikan dengan benar jenis hewan sesuai dengan kelasnya dibandingkan dengan kelas hewan tersebut, maka bisa kita lihat dari besaran nilai *recall* yang dihasilkan dari model yang diusulkan. Pada tabel 3 akan di paparkan hasil pengujian model *gradient boosting classifier* menggunakan beberapa *learning rate* dan variasi pembagian data atau split data untuk menyakinkan bahwa model yang diusulkan merupakan model yang paling cocok dengan data. Penelitian dilakukan dengan mencoba beberapa *learning rate* dan variasi *split data* yang kemudian diuji menggunakan model *gradient boosting classifier*, didapatkan nilai akurasi tertinggi sebesar 94%. Nilai ini didapatkan dari parameter *learning rate* sebesar 0,1 dan pembagian data sebesar 20% untuk *data testing* dan 80% untuk *data training*.

Tabel 4. Hasil Pengujian Nilai Recall

Recall				
Data training	Data Testing	Learning Rate		
		0.01	0.1	1
50%	50%	82%	87%	82%
60%	40%	74%	87%	81%
70%	30%	74%	83%	87%
80%	20%	69%	94%	88%

C. Presisi Gradient Boosting Classifier

Untuk mengetahui perbandingan antara *true positif* dengan banyaknya data yang diprediksi positif, maka bisa kita lihat dari besaran nilai presisi yang dihasilkan dari model yang diusulkan. Pada tabel 3 akan di paparkan hasil pengujian model *gradient boosting classifier* menggunakan beberapa *learning rate* dan variasi pembagian data atau split data untuk menyakinkan bahwa model yang diusulkan merupakan model yang paling cocok dengan data. Dari beberapa *learning rate* dan variasi *split data* yang telah diuji menggunakan model *gradient boosting classifier*, didapatkan nilai akurasi tertinggi sebesar 96%. Nilai ini didapatkan dari parameter *learning rate* sebesar 0,1 dan pembagian data sebesar 20% untuk *data testing* dan 80% untuk *data training*.

Tabel 5. Hasil Pengujian Nilai Presisi

Presisi				
Data training	Data Testing	Learning Rate		
		0.01	0.1	1
50%	50%	84%	88%	84%
60%	40%	65%	87%	83%
70%	30%	71%	82%	91%
80%	20%	53%	96%	84%

Tabel 6. Hasil Pengujian Nilai Mean Absolute Error (MSE)

Mean Absolute Error (MSE)				
Data training	Data Testing	Learning Rate		
		0.01	0.1	1
50%	50%	0.447	0.263	0.368
60%	40%	0.677	0.225	0.483
70%	30%	0.65	0.391	0.260
80%	20%	0.75	0.125	0.3125

D. Mean Absolute Error (MSE) Gradient Boosting Classifier

Untuk mengetahui besarnya rata-rata kesalahan dalam serangkaian Kalsifikasi yang dilakukan model yang diusulkan. Pada tabel 3 akan di paparkan hasil pengujian model *gradient boosting classifier* menggunakan beberapa *learning rate* dan variasi pembagian data atau *split* data untuk menyakinkan bahwa model yang diusulkan merupakan model yang paling cocok dengan data. Penelitian dilakukan dengan mencoba beberapa *learning rate* dan variasi *split* data yang telah diuji menggunakan model *gradient boosting classifier*, didapatkan nilai akurasi tertinggi sebesar 12,5%. Nilai ini didapatkan dari parameter *learning rate* sebesar 0,1 dan pembagian data sebesar 20% untuk *data testing* dan 80% untuk *data training*.

V. KESIMPULAN DAN SARAN

Klasifikasi jenis hewan pada kebun binatang yang memiliki *multi-class* menggunakan model *gradient boosting classifier*, klasifikasi terbaik didapatkan dengan menerapkan parameter *boosting* dengan *n_estimators* sebanyak 50, *max_depth* 3, *sub_sample* sebanyak 1.0, *learning rate* sebesar 0.1, dan menggunakan *criterion friedman* Mse. Serta dengan menerapkan *Split validation* atau pembagian data sebesar 80% untuk *data training* dan 20% untuk *data testing*.

Untuk megevaluasi model yang dibangun, Penulis menggunakan *performance metrics* seperti akurasi, *recall*, *F1-score*, *ROC curve* dan Mse. Hasil penelitian menyatakan bahwa model yang diusulkan lebih baik dari bebarapa model lain yang juga telah diuji dengan nilai akurasi sebesar 93.75%, *recall* 94%, presisi sebesar 96% dan MSE untuk mengukur besarnya rata-rata kesalahan dalam serangkaian klasifikasi sebesar 12,5%, semakin kecil nilai MSE maka akan semakin baik dalam mengklasifikasikan. Untuk perbaikan hasil yang lebih akurat, di harapkan pada penelitian berikutnya ada halhal yang perlu dicoba, seperti penambahan data amatan sehingga sebaran data dapat seimbang dan *representative*

DAFTAR PUSTAKA

- [1] A. P. Anugrah and A. Karya, “Produk Observasi Interaktif untuk Sarana Introduksi Hewan di Kebun Binatang,” *J. Tingkat Sarj. Seni Rupa dan Desain*, vol. 3, no. 1, pp. 376–382, 2014, [Online]. Available: <http://garuda.ristekbrin.go.id/documents/detail/270103>

- [2] N. I. Widiastuti, E. Rainarli, and K. E. Dewi, “Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen,” *J. Infotel*, vol. 9, no. 4, p. 416, 2017, doi: 10.20895/infotel.v9i4.312.
- [3] K. Nugroho and S. Murdowo, “Klasifikasi Jenis Hewan Pada Kebun Binatang Dengan Menggunakan Metode Deep Neural Network,” *J. Ilm. Infokam*, vol. 18, no. 1, pp. 46–51, 2022, doi: 10.53845/infokam.v18i1.317.
- [4] M. E. Al Rivian and Y. Yohannes, “Klasifikasi Mamalia Berdasarkan Bentuk Wajah Dengan k-NN Menggunakan Fitur CAS dan HOG,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 5, no. 2, pp. 169–176, 2019, doi: 10.35957/jatisi.v5i2.139.
- [5] D. P. Utomo and M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [6] A. M. Siregar and A. Puspabhuana, *Pengolahan Data Menjadi Informasi dengan Rapid Miner*. Surakarta: Kekata Group, 2017.
- [7] M. Arhami and M. Nasir, *Data Mining - Algoritma dan Implementasi*. Yogyakarta: ANDI, 2020.
- [8] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, “Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5,” *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
- [9] S. Febriani and H. Sulistiani, “Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5,” *89Jurnal Teknol. dan Sist. Inf.*, vol. 2, no. 4, pp. 89–95, 2021.
- [10] H. Azis, F. T. Admojo, and E. Susanti, “Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah,” *Angew. Chemie Int. Ed. 6(11)*, 951–952., pp. 2013–2015, 2020.
- [11] M. N. H. S. Muttaqin Muttaqin, Arsan Kumala Jaya, Sitti Harlina, Wahyuddin S, Lutfi Hakim, Mochammad Anshori, Agus Ambarwari, Fergie Joanda Kaunang, Green Arther Sandag, Harizahayu Harizahayu, Green Ferry Mandias, Maria F V Ruslau, Adhi Prasetio, Khaidir Rahman Nasir, *Data Science dan Pembelajaran Mesin*. Sumatera Utara: Yayasan Kita Menulis, 2023.
- [12] E. Ismanto and M. Novalia, “Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas,” *Techno.Com*, vol. 20, no. 3, pp. 400–410, 2021, doi: 10.33633/tc.v20i3.4576.
- [13] S. E. Suryana, B. Warsito, and S. Suparti, “Penerapan Gradient Boosting Dengan Hyperopt Untuk Memprediksi Keberhasilan Telemarketing Bank,” *J. Gaussian*, vol. 10, no. 4, pp. 617–623, 2021, doi: 10.14710/j.gauss.v10i4.31335.
- [14] U. L. Yuhana and A. Purwarianti, “Tuning Hyperparameter pada Gradient Boosting,” *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 1, pp. 134–139, 2022.
- [15] S. Suryono, E. Utami, and E. T. Luthfi, “Klasifikasi Sentimen Pada Twitter Dengan Naive Bayes Classifier,” *Angkasa J. Ilm. Bid. Teknol.*, vol. 10, no. 1, p. 89, 2018, doi: 10.28989/angkasa.v10i1.218.
- [16] J. J. Purnama, S. Rahayu, S. Nurdiani, T. Haryanti, and N. A. Mayangky, “Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Kanker Payudara,” *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 149–156, 2019, doi: 10.33480/pilar.v15i2.601.
- [17] R. R. Rerung, “Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk,” *J. Teknol. Rekayasa*, vol. 3, no. 1, p. 89, 2018, doi: 10.31544/jtera.v3.i1.2018.89-98.