

**PERBANDINGAN METODE PREDIKSI GALAT DALAM
PEMODELAN KLASIFIKASI DENGAN METODE
CLASSIFICATION AND REGRESSION TREE
UNTUK DATA SEIMBANG**

SKRIPSI

*Diajukan sebagai salah satu persyaratan untuk memperoleh gelar
Sarjana Statistika*



Oleh

FITRIA PANCA RAMADHANI

NIM 18337046

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2023**

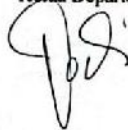
PERSETUJUAN SKRIPSI

PERBANDINGAN METODE PREDIKSI GALAT DALAM PEMODELAN KLASIFIKASI DENGAN METODE *CLASSIFICATION AND REGRESSION TREE* UNTUK DATA SEIMBANG

Nama : Fitria Panca Ramadhani
NIM : 18337046
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 21 Agustus 2023

Mengetahui:
Ketua Departemen Statistika



Dodi Vionanda, M.Si., Ph.D
NIP. 197906112005011002

Disetujui Oleh:
Pembimbing



Dodi Vionanda, M.Si., Ph.D
NIP. 197906112005011002

PENGESAHAN LULUS UJIAN SKRIPSI


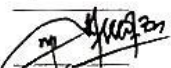
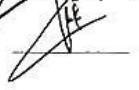
Nama : Fitria Panca Ramadhani
NIM : 18337046
Program Studi : SI Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

PERBANDINGAN METODE PREDIKSI GALAT DALAM PEMODELAN KLASIFIKASI DENGAN METODE *CLASSIFICATION AND REGRESSION TREE* UNTUK DATA SEIMBANG

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Padang

Padang, 21 Agustus 2023

Tim Penguji

	Nama	Tanda Tangan
Ketua	: Dodi Vionanda, Ph.D	
Anggota	: Dr. Syafriandi, M.Si	
Anggota	: Admi Salma, S.Pd, M.Si	

SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertandatangan di bawah ini:

Nama : Fitria Panca Ramadhani
NIM : 18337046
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul **“Perbandingan Metode Prediksi Galat dalam Pemodelan Klasifikasi dengan Metode *Classification and Regression Tree* untuk Data Seimbang ”** adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan.

Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,
Ketua Departemen Statistika,



Dodi Yonanda, Ph.D
NIP. 197906112005011002

Saya yang menyatakan,



Fitria Panca Ramadhani
NIM. 18337046

Perbandingan Metode Prediksi Galat Dalam Pemodelan Klasifikasi Dengan Metode *Classification and Regression Tree* Untuk Data Seimbang

Fitria Panca Ramadhani

ABSTRAK

Classification and Regression Tree (CART) adalah salah satu algoritma klasifikasi dalam metode *decision tree*. Model yang dibentuk dalam CART adalah pohon yang terdiri dari *root node*, *internal node*, dan *terminal node*. Keakuratan model akan dihitung melalui tingkat kesalahan prediksi pada model. Hal ini dapat dilakukan dengan menggunakan metode prediksi laju galat. Terdapat tiga metode dalam metode prediksi laju galat yaitu *Leave One Out Cross Validation* (LOOCV), *Hold Out* (HO), dan *K-Fold Cross Validation*. Ketiga metode tersebut memiliki kinerja yang berbeda dalam membagi data menjadi data training dan data testing sehingga masing-masing metode memiliki kelebihan dan kekurangan masing-masing. Oleh karena itu, dilakukan perbandingan untuk ketiga metode prediksi laju galat tersebut dengan tujuan menentukan metode prediksi laju galat yang tepat untuk algoritma CART.

Penelitian ini menggunakan bilangan acak berdistribusi normal dengan mempertimbangkan beberapa faktor seperti variasi rata-rata, jumlah variabel, dan korelasi. Hasil perbandingan nantinya diamati menggunakan boxplot dengan melihat median laju galat dan variasi terendah. Data yang dibangkitkan dibagi menjadi data *training* dan data *testing* sesuai dengan kinerja masing-masing metode prediksi laju galat setelahnya data *training* membentuk model dengan menggunakan algoritma CART. Model ini nantinya diuji menggunakan data *testing* dengan menghitung banyaknya galat yang terdapat model.

Hasil penelitian ini menunjukkan bahwa *K-Fold Cross Validation* memiliki median laju galat dan variasi yang paling rendah dibandingkan metode LOOCV dan HO, sehingga metode prediksi laju galat yang paling sesuai digunakan untuk algoritma CART adalah metode *K-Fold Cross Validation*.

Kata kunci: CART, *Hold Out*, *K-Fold Cross Validation*, *Leave One Out Cross Validation*, Prediksi Laju Galat

Comparison of Error Rate Prediction Methods in Classification Modeling with Classification and Regression Tree Methods for Balanced Data

Fitria Panca Ramadhani

ABSTRACT

Classification and Regression Tree (CART) is one of the classification algorithms in the decision tree method. The model formed in CART is a tree consisting of root nodes, internal nodes, and terminal nodes. The accuracy of the model will be calculated through the prediction error rate in the model. This can be done using the error rate prediction method. This method works by dividing the data into training data and testing data. There are three methods in the error rate prediction method, such as Leave One Out Cross Validation (LOOCV), Hold Out (HO), and K-Fold Cross Validation. These methods have different performance in dividing data into training data and testing data, so there are advantages and disadvantages to each method. Therefore, a comparison was made for the three error rate prediction methods with the aim of determining the appropriate method for the CART algorithm.

This study uses normally distributed random numbers by considering several factors such as the average variation, number of variables, and correlation. The comparison results will be observed using a boxplot by looking at the median error rate and the lowest variance. The generated data will be divided into training data and testing data according to the performance of each error rate prediction method after which the training data will form a model using the CART algorithm. This model will later be tested using data testing by calculating the number of errors in the model.

The results of this study indicate that the K-Fold Cross Validation has the median error rate and the lowest variance compared to the LOOCV and HO methods, so the most suitable error prediction method used for the CART method is the K-Fold Cross Validation method.

Keyword: Classification and Regression Tree, Hold Out, K-Fold Cross Validation, Leave One Out Cross Validation, Error Rate Prediction

KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada Allah Subhanaahu Wa Ta'ala atas segala rahmat-Nya dan berbagai kesempatan yang telah diberikan sehingga skripsi dengan judul Perbandingan Metode Prediksi Galat dalam Pemodelan Klasifikasi dengan Metode *Classification and Regression Tree* untuk Data Seimbang ini dapat diselesaikan. Shalawat serta salam juga penulis haturkan untuk Nabi Muhammad SAW.

Dalam penyusunan dan penulisan skripsi ini penulis mendapatkan banyak bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh sebab itu, dalam kesempatan ini penulis dengan senang hati menyampaikan banyak terima kasih kepada:

1. Bapak Dodi Vionanda, Ph.D., selaku pembimbing akademik dan skripsi yang telah membimbing dan memberikan banyak saran, arahan, dan dukungan kepada penulis selama pengerjaan skripsi ini.
2. Bapak Dr. Syafriandi, M. Si., dan Ibu Admi Salma, S.Pd, M.Si., selaku penguji skripsi yang telah memberikan saran dan masukan positif guna kesempurnaan skripsi ini.
3. Bapak/Ibu Dosen, Staf Pengajar, Karyawan Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
4. Teristimewa untuk kedua orang tua, Ayah Samson dan Ibu Islamidar yang telah memberikan kasih sayang, do'a dan dukungannya dalam bentuk moril maupun materil kepada penulis.

5. Kakak Meria Rezki, S.E., Ridha Aulia, S.Pd., Ir.Syamsul Akmal, S.T., dan Alvi Zichri, S.T., yang juga telah memberikan kasih sayang, do'a, dan dukungannya dalam bentuk moril maupun materil kepada penulis.
6. Semua sahabat, teman, dan rekan-rekan yang telah banyak membantu dan memberi dukungan kepada penulis.

Semoga semua bimbingan, bantuan, dan do'anya dibalas oleh Allah Subhanahu Wa Ta'ala sebagai amal ibadah. Penulis menyadari bahwa masih terdapat kekurangan dan kesalahan dalam penulisan skripsi ini. Oleh karena itu, penulis meminta maaf atas segala kesalahan yang dibuat dalam penulisan skripsi ini serta penulis menerima semua kritik dan saran yang bersifat membangun. Semoga penulisan skripsi ini bermanfaat bagi seluruh pihan, Aamiin.

Padang, Agustus 2023

Penulis

Fitria Panca Ramadhani

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR GAMBAR	vi
DAFTAR TABEL	vii
DAFTAR LAMPIRAN	viii
BAB I PENDAHULUAN	1
A.Latar Belakang	1
B.Batasan Masalah.....	7
C.Rumusan Masalah	7
D.Tujuan Penelitian.....	7
E.Manfaat Penelitian	8
BAB II TINJAUAN PUSTAKA	9
A. <i>Classification and Regression Tree (CART)</i>	9
B.Langkah Metode Cart.....	11
C.Prediksi Laju Galat.....	15
D.Boxplot.....	20
BAB III METODOLOGI PENELITIAN	23
A.Jenis Penelitian.....	23
B.Jenis dan Sumber Data.....	23
C.Pembangkitan Bilangan Acak.....	24
D.Teknik Analisis.....	30
BAB IV HASIL DAN PEMBAHASAN	33
A.Hasil Penelitian.....	33
B.Pembahasan.....	42
BAB V PENUTUP	45
A.Kesimpulan	45
B.Saran.....	45
DAFTAR PUSTAKA	46
LAMPIRAN	48

DAFTAR GAMBAR

Gambar	Halaman
1. Struktur Decision Tree.....	10
2. Ilustrasi Pemangkasan Pohon.....	15
3. Ilustrasi Hold out.....	17
4. Ilustrasi LOOCV.....	18
5. Ilustrasi K-Fold Cross Validation.....	20
6. Boxplot.....	21
7. Flowchart Tahapan Penelitian.....	32
8. Scatterplot untuk Pengaturan 1 Data Bivariat ketika (a) Tanpa Korelasi, (b) Korelasi Sedang, dan (c) Korelasi Tinggi.....	34
9. Scatterplot Data Multivariat untuk (a) Pengaturan Data 6, (b) Pengaturan Data 7, dan (c) Pengaturan Data 9	35
10. Perbandingan Perbedaan Rataan pada Kasus 1 dan 2 untuk Metode (a) LOOCV, (b) HO, dan (c) k-fold cv.....	37
11. Perbandingan Pengaturan Rataan untuk data multivariat pada (a) LOOCV, (b) HO, (c) K-fold cv dengan korelasi tinggi	38
12. Perbandingan Boxplot k-fold cv berdasarkan Korelasi pada (a) Pengaturan Data 1 dan (b) Pengaturan Data 2.....	39
13. Perbandingan Rataan pada Data Multivariat untuk Metode <i>K-fold CV</i>	40
14. Perbandingan Boxplot LOOCV berdasarkan Korelasi pada Data Multivariat untuk Pengaturan 1.....	41
15. Perbandingan Boxplot LOOCV berdasarkan Korelasi pada Data Multivariat untuk Pengaturan 2.....	41

DAFTAR TABEL

Tabel	Halaman
1. Pengaturan Rataan untuk Kasus Univariat.....	24
2. Pengaturan Rataan untuk Kasus Bivariat	25
3. Ketentuan Korelasi untuk Kasus Bivariat.....	26
4. Pengaturan Rataan untuk Kasus Data Multivariat.....	28
5. Ketentuan Korelasi untuk Kasus Multivariat.....	29
6. Nilai Median dan IQR untuk Boxplot pada Gambar 8.....	37

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Syntak Simulasi Pembangkitan Data Acak.....	48
2. Perbandingan Boxplot untuk Data Bivariat berdasarkan Korelasi	51
3. Perbandingan Boxplot untuk Data Multivariat berdasarkan Korelasi.....	53
4. Nilai Boxplot	55

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Klasifikasi merupakan metode dalam statistika yang bertujuan untuk memprediksi data baru menggunakan data lama. Dalam data mining, terdapat metode *unsupervised learning* dan *supervised learning*. *Unsupervised learning* merupakan metode yang bertujuan untuk mengelompokkan data berdasarkan kriterianya tanpa pemberian label sedangkan *supervised learning* yaitu metode yang membangun model antara variabel input dengan variabel target. Variabel input yang dimaksud yaitu variabel independen, sedangkan variabel target yaitu variabel dependen. Dalam hal ini, klasifikasi termasuk ke dalam metode *supervised learning*. (Rokach and Maimon, 2014: 9).

Decision tree merupakan salah satu metode klasifikasi yang tidak membutuhkan asumsi-asumsi parameter tertentu. Menurut Gupta (2017) *decision tree* memiliki kelebihan yaitu tipe data kategorik dan numerik dapat digunakan pada *decision tree* sedangkan metode lainnya seperti regresi logistik hanya menggunakan salah satu tipe data saja untuk variabel datanya.

Gupta dkk (2017) menyebutkan bahwa dalam *decision tree* terdapat berbagai macam algoritma di dalamnya seperti *Iterative Dichotomiser 3 (ID3)*, *C4.5*, *Classification and Regression Tree (CART)* dan *Chi-squared Automatic Interaction Detector (CHAID)*. Metode ID3 hanya memakai variabel dengan tipe kategorik sementara itu C4.5 dan CART sama-sama menggunakan variabel diskrit dan kontinu dan kedua metode ini juga

melakukan pemangkasan pada pohon, namun perbedaan CART dan CHAID yaitu CART merupakan algoritma *decision tree* dengan pohon biner. Sedangkan metode CHAID untuk variabel berskala nominal dan berfungsi untuk mendeteksi variabel dependen dari variabel yang dikategorikan dari suatu dataset.

Sebelumnya penelitian tentang perbandingan metode-metode algoritma ini telah dilakukan oleh sejumlah peneliti seperti oleh Patel dan Purvi (2018) menyebutkan bahwa metode CART memiliki akurasi paling tinggi dibandingkan dengan akurasi yang dihasilkan oleh ID3 dan C4.5 dengan tingkat akurasi yang dihasilkan oleh CART yaitu sebesar 97,11%. Sementara itu, penelitian yang dilakukan Nazar pada tugas akhirnya tentang klasifikasi preeklampsia (2018) diperoleh kesimpulan bahwa metode CART memiliki akurasi yang lebih tinggi daripada metode CHAID dengan akurasi pada metode CART sebesar 74% sedangkan pada CHAID sebesar 67%.

Berdasarkan penjelasan di atas, maka metode yang digunakan pada penelitian ini yaitu metode CART karena CART memiliki tingkat akurasi paling tinggi dibanding metode lainnya. Breiman dkk (1984) menyebutkan bahwa algoritma CART merupakan algoritma pohon keputusan dengan pelabelan berdasarkan kelas pada variabel dependen. Model pohon yang dihasilkan oleh CART bergantung pada tipe variabel dependen yang digunakan, variabel numerik akan menghasilkan pohon regresi sementara itu variabel kategorik akan menghasilkan pohon klasifikasi. Perbedaan pohon regresi dan pohon klasifikasi terletak pada kriteria pemisahannya. Kriteria pemisahan pada pohon regresi yaitu pohon regresi menggunakan perhitungan

Mean Squared Error (MSE) pada setiap simpul yang dihasilkan sedangkan pohon klasifikasi menggunakan nilai index gini maksimal pada pemisahan simpul yang dihasilkan. Pada penelitian ini, pohon yang digunakan yaitu pohon klasifikasi dikarenakan tipe variabel dependen yang digunakan berupa variabel kategorik.

Model yang terbentuk pada metode CART penting untuk diukur akurasi modelnya, Breiman dkk (1984) menyebutkan salah satu cara untuk melihat seberapa akurat suatu pengklasifikasi adalah dengan menguji pengklasifikasi pada kasus berikutnya dengan model klasifikasi yang telah terbentuk. Menurut Molinaro dkk (2005) akurasi model ini dapat dihitung dengan menggunakan laju galat. Laju galat dihitung dari perbandingan banyaknya galat dengan jumlah amatan yang digunakan dalam analisis. Perhitungan prediksi laju galat ini dilakukan dengan mempertimbangkan suatu metode mungkin menjadi yang paling baik dalam memprediksi suatu gugus data namun belum tentu metode tersebut dapat memprediksi gugus data yang berbeda dengan baik. Oleh karena itu, penting untuk melakukan prediksi laju galat pada metode yang digunakan.

Dalam mengukur akurasi model CART terdapat dua tipe metode dalam menghitung *error rate* yang dapat digunakan yaitu *training error rate* dan *test error rate*. *Training error rate* atau disebut juga dengan *resubstitution estimation* merupakan metode *error* yang menggunakan data yang sama untuk membentuk model dan menguji akurasi model. Metode ini sering digunakan untuk menghitung akurasi model namun metode *resubstitution estimate* merupakan metode galat paling tidak akurat diantara metode *error*

lainnya. Metode kedua yaitu metode *test error rate*. *Test error rate* yaitu metode yang membagi data menjadi dua bagian dengan masing-masing bagian berfungsi untuk membentuk model dan bagian lainnya berfungsi untuk menguji model yang telah terbentuk. Breiman dkk (1984) menyebutkan bahwa *test error rate* merupakan perkiraan klasifikasi yang sebenarnya. Berdasarkan penjelasan di atas, maka metode *error rate* yang digunakan untuk menguji model pada algoritma CART yaitu metode *test error rate*.

Metode *test error rate* yaitu metode prediksi *error rate* dengan membagi data menjadi dua bagian yaitu data *testing* dan data *training*. Metode ini juga disebut sebagai metode *cross validation*. Menurut Hastie dkk (2008: 241), *cross validation* merupakan salah satu metode yang paling sederhana dan paling banyak digunakan untuk memperkirakan prediksi *error* pada model. *Cross validation* bekerja dengan membagi data menjadi data *training* dan data *testing*. Data *training* merupakan data set yang digunakan untuk pemodelan pada data, sementara itu data *testing* digunakan sebagai pengujian atau validasi pada model yang terbentuk. Menurut James dkk (2013: 176), terdapat tiga metode dalam metode *cross validation*, yaitu *Hold Out Estimation*, *Leave One Out Estimation* (LOOCV), dan *K-Fold Cross validation*.

Estimasi *Hold Out* bekerja dengan membagi data secara acak menjadi dua kelompok dengan dua per tiga data menjadi data *training* dan sepertiga data tersisa menjadi data *testing*. Kinerja metode ini menyebabkan ketidakefisienan penggunaan data karena sepertiga datanya tidak digunakan untuk membentuk model pada data *training* (Kohavi, 1995).

Berbeda dengan *Hold Out*, pada LOOCV tidak terdapat pengacakan pada data. Prosedur pada LOOCV menyebabkan jumlah pengamatan sama dengan jumlah kasus LOOCV yang terbentuk sehingga setiap pengamatan berkesempatan menjadi data training dan data testing. Hasil akhir dari LOOCV yaitu berupa rataan *error* dari semua model yang terbentuk dari data *training* dan data *testing* (James, dkk, 2013: 179). Kelebihan dari metode LOOCV yaitu hasil akurasi rataan yang selalu konstan. Hal ini disebabkan oleh sistem beraturan pada LOOCV untuk setiap pengambilan data *testing* dan data *training* serta tidak adanya pengacakan dalam metode LOOCV (Wong, 2015). Selain kelebihan, LOOCV juga mempunyai kelemahan yang disebutkan oleh Efron (1983) yaitu metode ini memakan terlalu banyak waktu dan biaya jika digunakan pada data berukuran besar dan memiliki hasil perkiraan dengan nilai variasi yang sangat besar.

Sementara itu, pada metode *K-Fold cross validation*, pengamatan akan dikelompokkan secara acak terlebih dahulu dan setelahnya membagi kelompok yang terbentuk menjadi data *training* dan data *testing*. Dengan pengelompokkan terlebih dahulu, maka kelebihan dari metode *k-fold cv* yaitu dapat menghemat waktu dan biaya jika digunakan pada data berukuran besar. Namun, kelemahan dari *K-Fold Cross validation* adalah metode ini tidak cocok digunakan untuk data dengan sampel berukuran kecil (James, dkk, 2013: 181).

Perbandingan metode prediksi laju galat ini menggunakan data yang seimbang. Data seimbang yaitu data yang memiliki perbandingan kelas yang hampir sama. Shelke (2017) menyebutkan bahwa data seimbang sebagian

besar bekerja lebih baik pada algoritma data mining. Oleh karena itu pada penelitian ini digunakan data dengan proporsi kelas seimbang. Selain itu, korelasi juga memberikan pengaruh pada hasil prediksi. Pada CART, variabel yang berkorelasi dapat menyebabkan reduksi dalam informasi yang dimuat pada pohon. Klusowski (2020) menyebutkan bahwa korelasi berpengaruh terhadap variabel independen pada setiap node. Korelasi menyebabkan proses *splitting* yang kurang efektif sehingga berakibat pada berkurangnya kemampuan prediksi pohon serta hal ini juga menyebabkan *overfitting* yang berakibat pada generalisasi yang buruk pada pohon.

Pada penelitian terdahulu seperti penelitian yang dilakukan oleh Kohavi (1995) yang membandingkan *hold out*, *LOOCV*, *k-fold cross validation*, dan *bootstrap* didapatkan kesimpulan bahwa *ten-fold cross validation* merupakan metode terbaik dalam seleksi model. Kesimpulan yang sama juga disebutkan oleh Payam, dkk (2016) pada penelitiannya yang meneliti tentang perbandingan metode *resubstitution validation*, *hold-out*, *LOOCV*, dan *k-fold cross validation* dalam menyeleksi model klasifikasi.

Penelitian ini bertujuan untuk membandingkan metode prediksi laju galat sehingga mendapatkan metode prediksi laju galat yang sesuai untuk algoritma CART. Alat perbandingan yang digunakan yaitu *boxplot* dengan kriteria perbandingan nilai variasi terkecil dari masing-masing metode prediksi laju galat. Berdasarkan pemaparan di atas, maka dalam penelitian ini akan dibahas tentang “Perbandingan Metode Prediksi laju galat dalam Pemodelan Klasifikasi dengan Metode *Classification and Regression Tree* untuk Data Seimbang”.

B. Batasan Masalah

Berdasarkan pemaparan di atas maka batasan masalahnya yaitu sebagai berikut:

1. Algoritma klasifikasi yang digunakan yaitu algoritma *Classification and Regression Tree* (CART).
2. Perbandingan metode prediksi laju galat dalam mengukur akurasi model klasifikasi pada metode CART menggunakan tiga metode prediksi laju galat *cross validation* yaitu *Hold Out Estimation*, *Leave One Out Cross validation* (LOOCV), dan *K-Fold Cross validation*.

C. Rumusan Masalah

Rumusan masalah pada penelitian ini yaitu sebagai berikut:

1. Manakah metode *cross validation* yang menghasilkan dugaan prediksi laju galat yang paling baik diantara metode prediksi laju galat LOOCV, *Hold Out*, dan *K-Fold Cross validation* ?
2. Apakah korelasi antar variabel independen berpengaruh terhadap kinerja metode prediksi laju galat?

D. Tujuan Penelitian

Tujuan penelitian ini yaitu sebagai berikut:

1. Untuk mengetahui dugaan prediksi laju galat yang paling baik diantara metode LOOCV, *Hold Out*, dan *K-Fold Cross validation* untuk metode CART.
2. Untuk mengetahui pengaruh korelasi antar variabel independen terhadap kinerja metode prediksi laju galat.

E. Manfaat Penelitian

Manfaat dalam penelitian ini yaitu sebagai berikut:

1. Untuk penulis, memperluas pengetahuan penulis tentang penggunaan prediksi error menggunakan algoritma CART.
2. Untuk peneliti selanjutnya, menjadi referensi untuk memperluas cakupan hasil penelitian.