PERBANDINGAN METODE PREDIKSI LAJU GALAT DALAM PEMODELAN *CLASSIFICATION AND REGRESSION TREE* UNTUK KASUS DATA TIDAK SEIMBANG

SKRIPSI

LIFIA ZULLANI
NIM 18337053

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

PERSETUJUAN SKRIPSI

PERBANDINGAN METODE PREDIKSI LAJU GALAT DALAM PEMODELAN *CLASSIFICATION AND REGRESSION TREE* UNTUK KASUS DATA TIDAK SEIMBANG

Nama	: Lifia Zullani
NIM	: 18337053
Program Studi	: S1 Statistika
Departemen	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam

Padang, 06 Desember 2023

Mengetahui:
Ketua Departemen Statistika

Dr. Yenni Kurniawati, S.Si., M.Si
NIP. 198402232010122005

Disetujui Oleh:
Pembimbing

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

PENGESAHAN LULUS UJIAN SKRIPSI

Nama	: Lifia Zullani
NIM	: 18337053
Program Studi	: S1 Statistika
Departemen	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam

PERBANDINGAN METODE PREDIKSI LAJU GALAT DALAM PEMODELAN *CLASSIFICATION AND REGRESSION TREE* UNTUK KASUS DATA TIDAK SEIMBANG

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

Padang, 06 Desember 2023

Tim Penguji	Nama	Tanda Tangan
Ketua	: Dodi Vionanda, Ph.D	
Anggota	: Dr. Syafriandi, M.Si	
Anggota	: Dina Fitria, M.Si	

SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertandatangan di bawah ini:

Nama	: Lifia Zullani
NIM	: 18337053
Program Studi	: S1 Statistika
Departemen	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul **"Perbandingan Metode Prediksi Laju Galat dalam Pemodelan *Classification and Regression Tree* untuk Kasus Data Tidak 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,

Dr. Yenni Kurniawati, S.Si., M.Si
NIP. 19840223201012200005

Saya yang menyatakan,

Lifia Zullani
NIM. 18337053

Perbandingan Metode Prediksi Laju Galat dalam Pemodelan *Classification and Regression Tree* untuk Kasus Data Tidak Seimbang

Lifia Zullani

ABSTRAK

CART adalah salah satu algoritma klasifikasi berbasis pohon. Pohon CART terdiri dari simpul akar, simpul internal, dan simpul terminal. Akurasi model dalam CART dapat dihitung dengan mengukur kesalahan prediksi dalam model. Salah satu metode yang digunakan untuk memprediksi laju galat adalah *cross validation*. Terdapat tiga algoritma *cross validation*, yaitu *leave one out*, *hold out*, dan *k-fold cross validation*. Metode-metode ini memiliki kinerja yang berbeda dalam membagi data menjadi data pelatihan dan pengujian, sehingga terdapat kelebihan dan kekurangan pada setiap metode. *Hold out* tidak dapat menjamin bahwa himpunan pelatihan mewakili seluruh dataset, *leave one out* memerlukan perhitungan yang signifikan dan *k-fold cross validation* memerlukan waktu komputasi lebih lama. Dalam kenyataannya, data seringkali tidak seimbang. Data yang tidak seimbang merujuk pada data dengan jumlah observasi yang berbeda di setiap kelas.

Dalam CART, data tidak seimbang mempengaruhi hasil prediksi. Model CART menghasilkan pohon keputusan yang lebih cenderung memprediksi kelas mayoritas. Hal ini dapat mengurangi akurasi prediksi untuk kelas minoritas. Oleh karena itu, pemilihan metode *cross validation* yang sesuai sangat penting untuk memastikan bahwa c*ross validation* mempertahankan proporsi kelas yang benar selama evaluasi model. Penelitian ini berfokus pada perbandingan metode prediksi laju galat dalam model CART dengan data tidak seimbang. Penelitian ini menggunakan data simulasi bangkitan dengan pengaturan yang berbeda-beda.

Hasil yang diperoleh menunjukkan bahwa di antara ketiga metode *cross validation*, metode yang memiliki variasi *error rate* paling rendah adalah metode *k-fold cross validation*. Oleh karena itu, *k-fold cross validation* menjadi metode yang paling sesuai memprediksi laju galat pada CART dengan data tidak seimbang.

Kata kunci: CART, *cross validation*, data tidak seimbang, prediksi laju galat.

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

Lifia Zullani

ABSTRACT

CART is one of the tree-based classification algorithms. The CART tree consists of root nodes, internal nodes, and terminal nodes. The model's accuracy in CART can be calculated by measuring prediction errors in the model. One of the methods used to predict the error rate is cross-validation. There are three cross- validation algorithms: leave-one-out, hold-out, and k-fold cross-validation. These methods have different performances in dividing data into training and testing data,thus having their own advantages and disadvantages. Hold-out cannot guarantee that the training set represents the entire dataset, leave-one-out requires significant computation, and k-fold cross-validation takes more computational time. In reality, data is often imbalanced. Imbalanced data refers to data with varying observations in each class.

In CART, imbalanced data affects the prediction outcomes. The CART model tends to produce decision trees that predict the majority class more frequently. This can reduce the accuracy of predictions for minority classes. Therefore, the choice of an appropriate cross-validation method is crucial to ensure that cross-validation maintains the correct class proportions during model evaluation. This study focuses on comparing error rate prediction methods in CART models with imbalanced data. The research uses simulated data with various settings.

The results obtained indicate that among the three cross-validation methods,the method with the lowest error rate variation is k-fold cross-validation. Therefore,k-fold cross-validation is the most suitable method for predicting error rates in CART with imbalanced data.

Keywords: CART, cross validation, error rate prediction, imbalanced data.

KATA PENGANTAR

Puji syukur penulis ucapkan kepada Allah SWT karena berkat rahmat dan karunia-Nya, penulis dapat melakukan penelitian dan menyelesaikan penulisan Skripsi yang berjudul **Perbandingan Metode Prediksi Laju Galat dalam Pemodelan *Classification and Regression Tree* untuk Kasus Data Tidak Seimbang**. Shalawat beriring salam, penulis haturkan untuk Nabi Muhammad SAW yang telah membawa kita ke zaman yang penuh ilmu pengetahuan. Skripsi ini ditulis untuk memenuhi salah satu persyaratan dalam menyelesaikan pendidikan dan untuk memperoleh gelar Sarjana Statistika pada Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.

Dalam penulisan skripsi ini penulis banyak mendapat bantuan dari berbagai pihak, untuk itu penulis mengucapkan terima kasih kepada :

1. Ibuk Dr. Yenni Kurniawati, S.Si., M.Si., selaku Ketua Departemen Statistika dan Ketua Program Studi Sarjana Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang yang telah memberikan izin dan kemudahan dalam penyusunan skripsi ini.
2. Bapak Dodi Vionanda, S.Si, M.Si, Ph.D., selaku Pembimbing Akademik yang telah menyediakan waktu, tenaga, fikiran dan kesabaran untuk membimbing serta mengarahkan penulis dalam menyelesaikan penulisan skripsi ini.
3. Bapak Dr. Syafriandi, M.Si., dan Ibu Dina Fitria, M.Si., selaku dosen penguji skripsi yang telah memberi masukan, kritikan, dan petunjuk demi penyempurnaan skripsi ini.
4. Seluruh dosen dan pegawai tata usaha Departemen Statistika yang telah memberikan bantuan dan kemudahan dalam penyelesaian skripsi ini.

5. Orang tua dan keluarga yang telah memberikan doa, nasihat, dorongan secara moril maupun materil, motivasi serta dukungan pada penulis dalam penyelesaian skripsi ini.
6. Sahabat seperjuangan yang telah memberikan masukan, semangat dan bantuan dalam penulisan skripsi ini.
7. Seluruh rekan-rekan seperjuangan mahasiswa Departemen Statistika dan seluruh pihak yang telah memberikan dorongan demi penyelesaian skripsi ini.

Tiada kata yang dapat penulis persembahkan selain doa kepada Allah SWT mudah-mudahan segenap bantuan, bimbingan yang diberikan bernilai ibadah disisi Allah SWT dan mendapat balasan setimpal. Aamiin.

Penulis menyadari bahwa dalam penulisan skripsi ini masih banyak terdapat berbagai kekurangan. Oleh karena itu, diharapkan kritikan dan saran yang membangun dari pembaca guna kesempurnaan skripsi ini serta penelitian lanjutan untuk menyempurnakan segala kekurangan.

Padang, Desember 2023

Penulis,

Lifia Zullani
NIM. 18337053

DAFTAR ISI

DAFTAR TABEL

Tabel **Halaman**

DAFTAR GAMBAR

Gambar **Halaman**

DAFTAR LAMPIRAN

Lampiran **Halaman**

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Pengklasifikasian merupakan salah satu metode statistik untuk mengelompokkan atau mengklasifikasikan suatu data yang disusun secara sistematis. Metode klasifikasi dapat dilakukan dengan pendekatan parametrik dan nonparametrik (Amaliyyah, 2021). Salah satu metode klasifikasi dengan pendekatan nonparametrik adalah *Decision Tree* (Pohon Keputusan). Pohon keputusan adalah suatu metode eksplorasi yang berstruktur pohon untuk melihat hubungan antar variabel prediktor dan variabel respon. Beberapa metode yang dapat digunakan dalam metode pohon keputusan antara lain CHAID (*Chi-Squared Automatic Interaction Delection Analysis*), QUEST (*Quick, Unbiased Eefficient, Statistical Tree*), CART (*Classification and Regression Tree*), C4.5 dan lain-lain di mana masing-masing metode tersebut memiliki kekuatan dan kelemahan masing-masing.

Metode CART dan QUEST merupakan metode pohon keputusan yang menghasilkan struktur pohon biner, dimana sebuah pohon yang setiap simpulnya dipilah menjadi dua simpul yang terpisah (Breiman, *et al,* 1993). Sedangkan metode CHAID dan C4.5 merupakan metode pohon keputusan yang menghasilkan struktur pohon non-biner, dimana setiap simpul dipilah menjadi dua atau lebih simpul yang terpisah.

Selain perbedaan struktur pohon yang dihasilkan, metode-metode pohon keputusan memiliki perbedaan dalam kriteria pemilihan fitur dan nilai ambang

batas yang digunakan untuk pemisahan. Metode CART menggunakan *Gini Index* yang mengukur tingkat ketidakpastian atau keacakan dari suatu kumpulan data. Metode QUEST menggunakan uji F berdasarkan analisis varians untuk membandingkan rata-rata kelas target pada setiap pemisahan yang mungkin. Metode CHAID menggunakan uji *Chi-Square* untuk menentukan apakah ada hubungan yang signifikan antara fitur prediktor dan variabel target. Sedangkan metode C4.5 menggunakan kriteria *Gain Ratio* untuk mengukur seberapa besar informasi yang diperoleh dari suatu fitur untuk mengklasifikasikan data.

Metode-metode ini telah banyak digunakan sebelumnya dalam beberapa penelitian diantaranya adalah penelitian yang dilakukan oleh Pratiwi (2008) yang membandingkan metode QUEST dengan metode CHAID diperoleh hasil bahwa metode CHAID lebih baik dalam mengklasifikasikan debitur kredit konsumtif dikarenakan kelemahan metode QUEST pada analisis diskriminan kuadratik yang diterapkan menghasilkan bias apabila menggunakan peubah-peubah bebas kategorik dengan kesesuaian klasifikasi CHAID sebesar 70% sedangkan klasifikasi QUEST hanya sebesar 68,4%. Sedangkan Nazar (2018) yang menerapkan metode CHAID dan CART dalam klasifikasi Preeklamsia diperoleh hasil bahwa ketepatan hasil klasifikasi dengan algoritma CART sebesar 74% ini lebih tinggi dibandingkan dengan metode CHAID yang sebesar 67%. Kemudian penelitian oleh Darmawan (2017) melakukan perbandingan metode C.45 dan CART dalam klasifikasi lama masa studi mahasiswa diperoleh hasil bahwa kinerja algoritma CART lebih baik yaitu menghasilkan tingkat akurasi 60% dibandingkan algoritma C4.5 yang menghasilkan tingkat akurasi 40%.

Berdasarkan uraian di atas dapat dilihat bahwa metode CART memiliki ketepatan hasil klasifikasi yang lebih baik dibandingkan metode yang lain. CART dikembangkan untuk melakukan analisis klasifikasi pada variabel respon baik yang nominal, ordinal maupun kontinu. CART juga dapat menyeleksi variabel-variabel dan interaksi-interaksi variabel yang paling penting dalam menentukan hasil atau variabel prediktor. Menurut Saranya (2018), tujuan utama CART adalah untuk mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. Prinsip dari metode pohon klasifikasi ini adalah memilah seluruh amatan menjadi dua gugus amatan dan memilah kembali gugus amatan berikutnya, hingga diperoleh jumlah amatan minimum pada tiap-tiap gugus amatan berikutnya.

Metode CART memiliki kekurangan dan kelebihan tersendiri dibandingkan dengan metode lainnya. Kelebihan CART jika dibandingkan dengan metode klasifikasi yang lain adalah CART lebih mudah diinterpretasikan dan mempunyai tingkat akurasi yang tinggi. CART juga dapat melakukan penanganan terhadap variabel dalam jumlah banyak dengan skala variabel campuran melalui prosedur pemilihan biner (Hariati, 2018). Kekurangan yang dimiliki CART adalah hasil akhir tidak didasarkan pada model probabilistik, tidak ada tingkat probabilitas atau selang kepercayaan yang berhubungan dengan dugaan yang didapat dari pohon CART untuk mengelompokkan data baru (Ma, 2018:12). Tingkat kepercayaan dalam keakurasian CART benar-benar didasarkan pada keakuratan saat membuat pohon keputusan.

Akurasi pohon keputusan yang telah dibentuk menggunakan model CART dapat dihitung dengan menggunakan metode prediksi laju galat. Galat atau *error*

adalah perbedaan antara nilai yang diukur atau dihitung dengan nilai sebenarnya. Galat merupakan ukuran yang digunakan untuk menilai keakuratan pohon klasifikasi yang dibangun menggunakan data set dalam memprediksi data baru. Untuk menghitung prediksi galat terdapat dua metode yang dapat digunakan yaitu *training error rate* dan *test error rate. Training error rate* dan *testing error rate* adalah dua konsep penting dalam pembelajaran mesin yang digunakan untuk mengevaluasi kinerja model yang dilatih.

*Training error rat*e adalah tingkat kesalahan atau *error* yang dihasilkan oleh model saat dilatih pada data pelatihan (*training data*). Ini menunjukkan seberapa baik model dapat mempelajari pola dan hubungan dalam data pelatihan. Semakin rendah *training error rate*, semakin baik model dapat mencocokkan data pelatihan. Namun, *training error rate* yang terlalu rendah dapat mengarah pada *overfitting*, dimana model terlalu cocok dengan data pelatihan dan tidak dapat menggeneralisasi dengan baik pada data baru. *Testing error rate* adalah tingkat kesalahan atau *error* yang dihasilkan oleh model saat diuji pada data pengujian (*testing data*). Data *testing* adalah data yang tidak digunakan dalam proses pelatihan model. *Testing error rate* memberikan indikasi tentang seberapa baik model dapat menggeneralisasi dan memprediksi data baru yang belum pernah dilihat sebelumnya. Ini adalah ukuran yang lebih akurat untuk mengevaluasi kinerja model dalam dunia nyata.

Perbedaan utama antara *training error rate* dan *testing error rate* adalah bahwa *training error rate* hanya mengevaluasi kinerja model pada data pelatihan, sedangkan *testing error rate* mengevaluasi kinerja model pada data yang belum pernah dilihat sebelumnya. Oleh karena itu, *testing error rate* dianggap sebagai

ukuran yang lebih baik untuk menilai seberapa baik model dapat menggeneralisasi dan memberikan prediksi yang akurat pada data baru.

Pembagian data *testing* dan *training* pada *testing error rate* adalah salah satu aspek penting dalam proses pemodelan mesin pembelajaran. Tujuan utama dari pembagian data adalah untuk mengevaluasi kinerja model yang dilatih pada data *training* dan menguji kemampuannya dalam memprediksi data baru (data *testing*). Dengan memisahkan data menjadi data *training* dan data *testing*, dapat memperoleh estimasi yang lebih akurat tentang bagaimana model akan berfungsi pada data baru yang belum pernah dilihat sebelumnya. Metode pembagian data dilakukan dengan menggunakan *cross validation. Cross validatian* merupakan metode pembagian data menjadi data *training* dan data *testing*. Dalam *cross validation* terdapat berbagai algoritma yang dapat digunakan dimana masing-masing algoritma memiliki karakteristik yang berbeda, yaitu *hold out, leave one out cross validation,* dan *k-fold cross validation.*

Hold out membagi data ke dalam dua himpunan data yang diberi nama data *training* dan data *testing*. Model klasifikasi dihasilkan dari data *training* dan kinerjanya dievaluasi menggunakan data *testing*. Proporsi data yang dicadangkan umumnya menggunakan 2/3 untuk data *training* dan 1/3 untuk data *testing* (Suyanto, 2017). Metode *hold out* memiliki beberapa keterbatasan, diantaranya lebih sedikit pengamatan yang tersedia untuk data *training* karena beberapa pengamatan digunakan sebagai data *testing*.

Leave one out cross validation (LOOCV) sama halnya dengan *hold out validation,* membagi data menjadi dua bagian yaitu data *training* dan data *testing*. Pada metode *leave one out* hanya satu pengamatan saja yang dijadikan sebagai

data *testing*, sehingga *n-1* pengamatan lainnya digunakan sebagai data *training*. Dengan kata lain, dalam setiap iterasi hampir semua data kecuali satu pengamatan digunakan untuk pelatihan, dan model diuji pada observasi tunggal tersebut. Estimasi akurasi yang diperoleh dengan menggunakan LOOCV diketahui hampir tidak bias, tetapi memiliki varians yang tinggi, yang mengarah pada estimasi yang tidak reliabel.

K-fold cross validation, sesuai dengan namanya metode ini membagi himpunan data secara acak menjadi *k-fold* yang saling bebas. Dalam metode *k-fold cross validation*, 1 *fold* (atau satu lipatan) mengacu pada satu bagian dari data yang digunakan sebagai data *testing*, sementara bagian lainnya digunakan sebagai data *training*. Proses *k-fold cross validation* dilakukan dengan cara membagi data secara acak menjadi *K* bagian (*fold*) yang sama besar. Untuk setiap iterasi, satu *fold* digunakan sebagai data *testing*, sementara *K-1 fold* lainnya digunakan sebagai data *training*. Proses ini diulang sebanyak *K* kali, sehingga setiap *fold* akan menjadi data *testing* tepat satu kali.

Molinaro dkk (2005), membandingkan berbagai metode resampling data untuk kumpulan data berdimensi tinggi, yang umum dalam bioinformatika. Temuan mereka menunjukkan bahwa LOOCV, *k-10 cross validation*, dan .632+ bootstrap memiliki bias terkecil. Dimana, perkiraan kesalahan prediksi hampir tidak bias dalam *k-10 cross validation*.

Kohavi (1995) dalam penelitiannya membandingkan beberapa pendekatan untuk memperkirakan akurasi, *cross validation* (termasuk *hold out validation, leave one out cross validation* dan *k-fold cross validation*) dan *bootstrap*, dan

merekomendasikan *10-fold cross validation* sebagai metode pemilihan model terbaik, karena cenderung memberikan estimasi akurasi yang kurang bias.

Sebagian besar data riil yang dihasilkan dalam penelitian merupakan data yang tidak seimbang. Data tidak seimbang adalah data yang memiliki jumlah kelas amatan yang berbeda. Ketidakseimbangan data berdampak pada hasil prediksi yang tidak stabil. Model CART memiliki kecenderungan untuk menghasilkan pohon keputusan yang lebih cenderung memprediksi kelas mayoritas karena ada lebih banyak contoh kelas mayoritas (Sari, 2019). Hal ini dapat mengurangi akurasi prediksi untuk kelas minoritas. sehingga pemilihan metode prediksi laju galat yang sesuai sangat penting untuk memastikan bahwa metode prediksi laju galat dapat mempertahankan proporsi kelas yang benar selama evaluasi model. Oleh karena itu, penelitian ini membahas tentang perbandingan metode prediksi galat *hold out,* LOOCV, dan *k-fold cross validation* dengan tujuan menentukan algoritma yang sesuai untuk diterapkan pada metode CART dengan data tidak seimbang. Bahan penduga yang baik adalah penduga dengan bias yang rendah dan variansinya terkecil.

Boxplot dapat digunakan untuk membandingkan metode estimasi dalam menentukan metode terbaik. Dalam hal ini, boxplot dapat membantu dalam membandingkan distribusi data, mengidentifikasi *outlier*, membandingkan presisi, dan membandingkan bias dari berbagai metode estimasi.

Boxplot dalam membandingkan distribusi data dapat membantu menilai perbedaan dalam distribusi data yang dihasilkan oleh masing-masing metode. Metode estimasi yang menghasilkan boxplot dengan median, kuartil, dan rentang yang lebih kecil atau terpusat dapat dianggap sebagai metode yang lebih baik.

Boxplot dalam mengidentifikasi *outlier*, dapat membantu melihat apakah ada perbedaan dalam jumlah atau posisi *outlier* yang dihasilkan oleh berbagai metode estimasi. Metode estimasi yang menghasilkan lebih sedikit *outlier* atau *outlier* yang lebih terpusat dapat dianggap sebagai metode yang lebih baik.

Boxplot dalam membandingkan presisi, dapat membantu menilai ukuran sebaran data. Metode estimasi yang menghasilkan boxplot dengan rentang interquartil (IQR) dan *whiskers* yang lebih kecil dapat dianggap sebagai metode yang lebih presisi.

Boxplot dalam membandingkan bias, posisi median dalam boxplot dapat memberikan informasi tentang bias metode estimasi. Jika median boxplot dari suatu metode estimasi berbeda secara signifikan dari nilai sebenarnya (atau nilai referensi), maka metode tersebut dapat dianggap memiliki bias yang lebih besar. Dengan menggunakan boxplot sebagai alat pembanding, maka dapat memilih metode estimasi yang paling sesuai untuk data yang dimiliki.

Berdasarkan uraian tersebut peneliti tertarik melakukan penelitian dengan judul "Perbandingan Metode Prediksi Laju Galat untuk Pemodelan Klasifikasi dengan Metode *Classification and Regression Tree* (CART) untuk Kasus Data Tidak Seimbang (*Imbalanced*)".

B. Batasan Masalah

Batasan masalah pada penelitian ini adalah data yang digunakan merupakan data tidak seimbang (*imbalanced*) bangkitan *Software R* dengan proporsi kelas data yang bervariasi.

C. Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja prediksi laju galat *leave one out cross validation, hold out,* dan *k-fold cross validation* pada kasus data tidak seimbang dengan proporsi kelas data yang bervariasi?
2. Diantara metode prediksi laju galat *leave one out cross validation, hold out,* dan *k-fold cross validation*, metode manakah yang lebih cocok digunakan pada CART dengan kasus data tidak seimbang?

D. Tujuan Penelitian

Tujuan penelitian dalam skripsi ini adalah sebagai berikut:

1. Untuk mengetahui kinerja prediksi laju galat *leave one out cross validation, hold out,* dan *k-fold cross validation* pada kasus data tidak seimbang dengan proporsi kelas data yang bervariasi.
2. Untuk mengetahui metode prediksi laju galat yang paling cocok digunakan pada CART dengan kasus data tidak seimbang.

E. Manfaat Penelitian

Penelitian ini diharapkan bisa memberikan manfaat sebagai berikut:

1. Bagi penulis, diharapkan dapat memperluas pengetahuan tentang penggunaan metode prediksi laju galat pada CART dengan kasus data tidak seimbang.
2. Bagi pembaca dan peneliti selanjutnya, diharapkan dapat menjadi pedoman dalam melakukan penelitian sejenis untuk melengkapi penelitian metode terkait.