

Komparasi Metode Evaluasi Pada *Credit Scoring Data Mining*

Hermawan
Prodi Sistem Informasi STMIK GI MDP
Jl. Rajawali No.14 Palembang 30113
Sumatera Selatan, Indonesia
E-mail: hermawan@mdp.ac.id

Yoannita
Teknik Informatika STMIK GI MDP
Jl. Rajawali No.14 Palembang 30113
Sumatera Selatan, Indonesia
E-mail: yoannita@mdp.ac.id

Abstract— *Credit Scoring* is a procedure that exists in all loan company. This procedure will give the result whether an applicant is eligible to receive loan or not. Data mining approach has been a well known method to assist this procedure. Based on data, this approach helps to measure an applicant's credit worthiness. This research is using Classification and Regression Tree (CART) as a part of classification task in credit scoring. As a matter a fact that CART is recognized as one of the best algorithms that can be used in data mining. The output of this classification algorithm is validate using evaluation methods which are usually used to measure the performance of classification model result. These evaluation methods that commonly used in this research area are k-fold validation method and holdout method. This experiment using CART Algorithm along with credit scoring public dataset. The experiment result shows that along with CART algorithm, using cross validation method gives a better performance in term of accuracy and error rate.

Keywords: *data mining, cart, hold out method, cross validation*

I. PENDAHULUAN

Era teknologi informasi yang dimanfaatkan di berbagai bidang kehidupan manusia, pada awalnya ditujukan untuk membantu proses operasional suatu sistem yang sudah ada. Seiring dengan pemanfaatan teknologi informasi pada sistem informasi, data yang dihasilkan dari sisi operasional akan menghasilkan suatu basis data. Basis data ini yang awalnya hanya bertugas sebagai media pencatatan saja, sudah bertransformasi menjadi informasi yang berharga dengan metode pengolahan basis data yang lazim dikenal dengan data mining. *Credit scoring data mining* adalah bagian dari *data mining*. *Credit scoring* adalah proses penilaian kredit yang wajib dilakukan oleh pihak lembaga kredit. Melalui proses ini, akan ditentukan apakah calon debitur yang mengajukan kredit diklasifikasikan sebagai calon debitur yang layak untuk diberikan pinjaman atau sebaliknya. Kesalahan dalam proses *credit scoring*, pada akhirnya akan mengakibatkan kerugian dari lembaga keuangan tersebut. Kesalahan proses yang umum terjadi adalah kesalahan hasil dari prosedur *credit scoring* tersebut. Kesalahan yang dinamakan *error* tipe I dan tipe II [1]. *Error* tipe satu didefinisikan sebagai kesalahan output dari

credit scoring, yang mana calon debitur potensial yang seharusnya layak sebagai debitur menjadi tidak layak. *Error* ini akan memberikan dampak yaitu kerugian bagi lembaga kredit tersebut. Sebaliknya, *error* tipe dua, dimana calon debitur yang tidak layak, setelah melalui prosedur *credit scoring*, diklasifikasikan sebagai calon debitur potensial. Tentunya ini menjadi kerugian yang lebih besar bagi lembaga kredit tersebut, karena akan berpotensi menjadi sumber kredit macet. Kemudian kredit macet ini pada akhirnya akan mengurangi keuntungan perusahaan dan mempengaruhi tingkat *cash flow* dari lembaga kredit tersebut.

Berbagai metode dan algoritma klasifikasi populer sudah pernah digunakan dalam penelitian sebelumnya, seperti algoritma *Support Vector Machine*, algoritma *Artificial Neural Network*, algoritma *Multivariate Adaptive Regression Splines* [2]. Adapun salah satu algoritma klasifikasi terbaik yang bisa digunakan untuk data mining adalah algoritma *Classification and Regression Tree* (CART). Algoritma CART adalah salah satu dari sepuluh algoritma terbaik untuk digunakan di dalam *data mining* [3]. Ini yang mendasari penulis memilih untuk menggunakan algoritma ini.

Sedangkan metode untuk mengevaluasi model klasifikasi *credit scoring* yang dihasilkan oleh algoritma-algoritma tersebut tidak banyak, hanya ada beberapa saja. Metode evaluasi yang populer dan umum digunakan untuk *credit scoring data mining* adalah metode holdout dan K-fold validation.

Penelitian ini akan membandingkan pengaruh penggunaan kedua metode tersebut terhadap tingkat akurasi. Eksperimen akan dilakukan dengan menggunakan beberapa dataset publik yang sering digunakan untuk riset *credit scoring* yaitu *german.credit* dataset, *german.credit-numeric* dan *Australian* dataset. Hasil penelitian kemudian akan membahas metode validasi yang paling cocok untuk digunakan untuk *credit scoring*, dengan menggunakan beberapa poin pertimbangan. Penelitian ini akan terbagi ke dalam beberapa tahapan, yaitu diawali dengan studi literatur, kemudian penentuan metodologi penelitian, dilanjutkan dengan tahapan pelaksanaan penelitian dan diakhir dengan tahapan analisis data hasil penelitian untuk mengambil kesimpulan dan saran untuk penelitian lebih lanjut.

II. CLASSIFICATION AND REGRESSION TREE (CART)

Algoritma *Classification and Regression Tree* (CART) adalah algoritma yang dikembangkan oleh Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone. Algoritma ini dapat digunakan untuk membangun sebuah model pohon keputusan atau pohon regresi berdasarkan data-data historis yang digunakan sebagai objek observasi [4]. Beberapa ilmuwan berpendapat bahwa algoritma ini adalah bagian dari perkembangan beberapa bidang ilmu seperti kecerdasan buatan, *machine learning*, statistik non parametrik dan *data mining* [3]. Output dari algoritma ini tergantung dari variabel respon data yang ada. Apabila variabel respon datanya berupa tipe kategorik, mana outputnya adalah pohon keputusan, sedangkan apabila variabel respon datanya yang digunakan adalah data variabel dengan tipe data numerik maka akan mengeluarkan output berupa pohon regresi [5]. CART biasanya akan memberikan gambaran mengenai hubungan variabel respon dengan satu atau lebih dari beberapa variabel prediktor yang ada di dalam dataset. Implementasi dari algoritma ini digunakan di dalam berbagai bidang kehidupan manusia seperti kesehatan, pendidikan, finance dan sebagainya.

Algoritma Classification and Regression Tree (CART) terdiri dari beberapa tahapan utama yaitu [6]:

1. Tree Building

Tahapan awal dimulai dengan pembentukan *node root* pohon keputusan. Algoritma ini akan mencari variabel terbaik sebagai calon dari *node root* pohon keputusan, dengan cara melakukan pencarian semua kemungkinan variabel yang ada. Pemilahan kemudian dilakukan untuk memilih dua *node*, yaitu *node* kiri dan *node* kanan. Proses ini akan dilakukan secara berulang-ulang dan bersifat rekursif. Secara spesifik, kegiatan utama dari *tree building* adalah pemilahan, penentuan simpul/node dan penandaan label kelas (*class assignment*). Kriteria pemilahan umumnya menggunakan *impurity measure*. Adapun *impurity measure* yang bisa digunakan untuk pengukuran adalah *gini impurity*, *information gain* dan *twoing dan entropy*.

2. Stopping Tree Building Process

Pembuatan pohon keputusan menggunakan algoritma CART pada awalnya dibuat secara rekursif tanpa ada batasan, dan dilakukan sampai dengan tahapan tidak bisa dibuat lagi simpul atau node. Tahapan ini adalah tahapan penentuan *node* terminal, yang digunakan sebagai tempat diprediksikannya suatu objek pada *class* tertentu. Perkembangan selanjutnya dari algoritma ini, memungkinkan dibuatnya batasan-batasan untuk melakukan penghentian pembentukan pohon keputusan. Batasan tersebut, biasanya adalah penghentian pembentukan pohon keputusan jika tidak ada lagi penurunan keheterogenan yang berarti. Beberapa batasan lainnya digunakan adalah batasan minimum, batasan level pohon keputusan, dan batasan

kedalaman pohon keputusan.

3. Tree Pruning

Pembentukan pohon keputusan secara rekursif, akan menghasilkan juga pohon keputusan yang kompleks dengan banyak *node*. Kompleksitas akan menyebabkan pohon keputusan tersebut sulit untuk dipahami dan dimengerti. *Tree Pruning* /pemangkasan sebagai salah satu tahapan dari algoritma CART diperlukan untuk mengurangi kompleksitas dari model pohon keputusan yang sudah dibuat. Pemangkasan dikerjakan dengan mempertimbangkan tingkat akurasi yang baik. Metode pengukuran yang digunakan untuk memperoleh ukuran pohon keputusan yang baik tersebut dikenal dengan istilah *cost complexity*.

4. Optimal Tree Selection

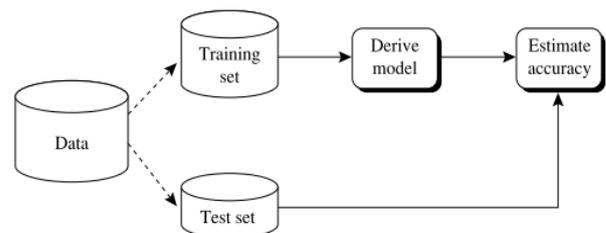
Tahapan ini akan dilakukan pemilahan pohon keputusan yang paling optimum dan terbaik. Metode pengukuran dilakukan dengan melakukan dengan *test sample estimate* dan *V cross validation estimate*.

III. METODE EVALUASI

Pengukuran performa dilakukan untuk mengetahui seberapa handal model klasifikasi yang dibentuk oleh algoritma tertentu, dalam peneliti ini yaitu algoritma CART. Evaluasi dikerjakan dengan metode yang sudah dipilih oleh peneliti. Metode yang populer di kalangan *researcher* dan yang akan digunakan di dalam penelitian ini adalah metode *hold out validation* dan *cross validation method*. Penelitian ini akan membandingkan hasil dari kedua metode evaluasi tersebut untuk kebutuhan *credit scoring data mining*.

A. Hold Out Validation Method

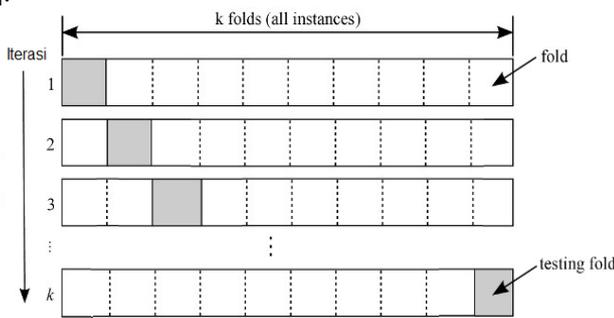
Metode ini akan membagi *dataset* secara acak ke dalam dua kategori dataset yaitu *dataset training* dan *dataset testing*. *Dataset training* akan digunakan oleh algoritma CART untuk membentuk pohon klasifikasi, sedangkan *dataset testing* digunakan untuk menguji model klasifikasi yang dibentuk. Umumnya untuk alokasi *data training* adalah sebesar 2/3 dari *dataset* dan sisanya sebesar 1/3 dari *dataset data testing*. Estimasi dan prediksi akan dilakukan selanjutnya untuk menilai kehandalan dari model klasifikasi yang dibentuk berdasarkan data training [7].



Gambar 1. Model Evaluasi Validation Method

B. Cross Validation

Metode pengukuran performa model lain yang bisa digunakan adalah metode *K-Fold cross validation*. Nilai K yang sering digunakan adalah nilai 5 dan nilai 10. Metode ini membagi dataset menjadi beberapa subset data (D_1, D_2, \dots, D_k) sesuai dengan nilai K, yang ukuran setiap subset datanya sama. Evaluasi model klasifikasi akan diulangi sebanyak K kali. Pada iterasi pertama, subset data D_1 akan digunakan sebagai data testing, sedangkan sisa subset lainnya (D_2, \dots, D_k) akan digunakan sebagai data training. Pada iterasi kedua, subset D_2 akan digunakan sebagai data testing, sedangkan sisa subset lainnya (D_1, D_3, \dots, D_k) akan digunakan sebagai data training. Hal ini akan berlanjut terus menerus sampai dengan nilai iterasi k, yang sudah ditentukan sebelumnya sebagai konstanta. Untuk model klasifikasi, tingkat akurasi yang dihasilkan adalah nilai akurasi rata-rata yang dihitung akumulasi total dari nilai akurasi yang didapatkan dari setiap iterasi dibagi dengan jumlah subset /K [7].



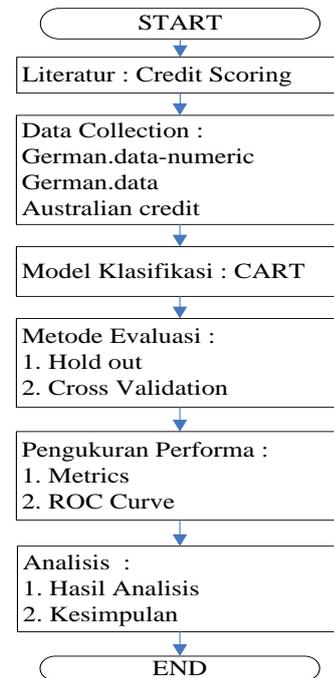
Gambar 2. Model Evaluasi Cross Validation

C. Instrumen Penelitian

Penelitian akan menggunakan *notebook* dengan spesifikasi yaitu processor Intel core i5, memory dengan kapasitas 10 GB dan harddisk 500 GB. Lingkungan sistem operasi yang digunakan adalah Windows 7 Home Edition 64 bit.

IV. METODE PENELITIAN

Gambar 3 menunjukkan metodologi dan tahapan yang digunakan oleh peneliti. Penelitian akan dimulai dengan studi literatur terhadap bidang ilmu yang akan diteliti, yaitu *credit scoring data mining*, termasuk di dalamnya algoritma yang digunakan dan metode evaluasi yang menjadi subjek di dalam penelitian. Pengunduhan juga dilakukan untuk mengumpulkan dataset penelitian yang umum digunakan oleh peneliti dari repository University of California. Tahapan selanjutnya adalah pengumpulan data hasil penelitian dengan melakukan penelitian dengan menggunakan algoritma CART dan menggunakan metode evaluasi *holdout* dan *cross validation*. Berdasarkan data hasil penelitian kemudian akan dilakukan performa model klasifikasi dengan menggunakan metrik yang sudah ditentukan sebelumnya. Tahapan terakhir adalah melakukan analisis terhadap data hasil penelitian dan kemudian menarik kesimpulan.



Gambar 3. Metodologi Penelitian

V. HASIL DAN PEMBAHASAN

Penelitian ini akan menggunakan beberapa metrik sebagai indikator yang menunjukkan performa dari model klasifikasi yang sudah terbentuk. Metrik tersebut adalah *accuracy*, *Error rate*, *Sensitivity*, *Specificity*, *Precision*, dan *AUC (Area Under Curve)*. Beberapa parameter lain yang digunakan adalah nilai *random seed* yang diinisialisasi dengan nilai 3. Kemudian parameter lain adalah persentase untuk metode evaluasi adalah 66% dan 33%. Persentase *data training* sebesar 66% atau 2/3 dari *dataset* dan persentase untuk *data testing* sebesar 33% atau 1/3 dari *dataset*. Parameter selanjutnya adalah nilai k untuk *k-fold validation*. Sesuai dengan literatur sebelumnya, nilai k diinisialisasi dengan nilai yang umum untuk digunakan peneliti yaitu nilai 10.

TABEL I
HASIL PENELITIAN DATASET GERMAN.DATA

Metrik	Hold Out	10-Fold Cross Validation
Accuracy (%)	71,47	71,7
Error rate (%)	28,53	28,3
Sensitivity (%)	93,20	85,6
Specificity (%)	23,60	39,3
Precision (%)	72,90	76,7
AUC	0,7368	0,707

TABEL III
HASIL PENELITIAN DATASET GERMAN.DATA-NUMERIC

Metrik	Hold Out	10-Fold Cross Validation
Accuracy (%)	74,11	75,2
Error rate (%)	25,88	24,8
Sensitivity (%)	88,9	89,1
Specificity (%)	41,5	42,7

Precision (%)	77	78.4
AUC	0,7657	0.7196

TABEL III
HASIL PENELITIAN DATASET AUSTRALIAN CREDIT

MetriK	Hold Out	10-Fold Cross Validation
Accuracy (%)	84,26	85,36
Error rate (%)	15,74	14,63
Sensitivity (%)	85,7	84
Specificity (%)	83,2	86,4
Precision (%)	78,5	83,2
AUC	0,8994	0,8771

Hasil penelitian menunjukkan perbandingan antara metode evaluasi *hold out* dan metode *10-fold validation* pada *dataset credit scoring* yang ditentukan dalam bentuk persentase. Tabel I menunjukkan bahwa penggunaan algoritma CART dan metode evaluasi *holdout* untuk kepentingan *credit scoring* dengan dataset *german.data* memberikan hasil dengan tingkat akurasi sebesar 74,11 % dan tingkat error sebesar 28,53 %, sedangkan dengan menggunakan metode *10-Fold Validation* memperlihatkan bahwa tingkat akurasi yang didapat lebih baik yaitu dengan tingkat akurasi 71,7 % dan diiringi juga dengan tingkat error yang lebih rendah yaitu dengan 28,3 %. Tabel II menunjukkan bahwa penggunaan metode evaluasi *10-fold validation* memperlihatkan tingkat akurasi yang lebih baik bila dibandingkan dengan penggunaan metode evaluasi *holdout*, dengan selisih persentase sebesar 1,08 %. Hal ini juga diikuti dengan tingkat error yang lebih baik, yaitu tingkat kesalahan model klasifikasi memprediksi calon nasabah yang lebih baik, dengan selisih persentase yang sama sebesar 1,08 %. Hal yang sama juga terlihat pada tabel III, *credit scoring* dengan menggunakan dataset *Australian credit approval dataset*, tingkat akurasi terlihat lebih baik dengan menggunakan metode *10-fold validation* dengan tingkat akurasi sebesar 85,36 %, yang berarti nilai tersebut lebih baik sebesar 1,1 % dibandingkan dengan metode *holdout*.

Berdasarkan tabel hasil penelitian, diketahui bahwa penggunaan metode evaluasi *10-fold validation* menunjukkan performa yang lebih baik dibandingkan dengan penggunaan metode evaluasi *holdout*. Beberapa indikator utama yang terlihat bahwa tingkat akurasi yang lebih baik dan juga indikator tingkat *error* atau kesalahan prediksi yang turun, menunjukkan bahwa tingkat *error* yang lebih baik dibandingkan dengan metode *holdout*.

VI. KESIMPULAN

Credit scoring adalah bentuk penilaian yang dilakukan oleh lembaga keuangan terhadap calon debitur. Prosedur ini akan menentukan layak atau tidaknya calon debitur untuk menerima kredit dari lembaga keuangan tersebut. Data mining kemudian digunakan untuk penilaian tersebut yang berbasis data. CART adalah algoritma yang berbasis *decision tree* yang bisa digunakan untuk *credit scoring*. Model evaluasi kemudian digunakan untuk melakukan perhitungan performa dari model klasifikasi yang terbentuk oleh algoritma.

Dua model evaluasi yang populer digunakan adalah metode *holdout* dan *k-fold validation*.

Berdasarkan penelitian yang dilakukan, terlihat bahwa metode *k-fold validation* dengan nilai 10, lebih baik dibandingkan dengan metode *holdout*. Ini terlihat dari data hasil penelitian dengan menggunakan beberapa dataset publik yang sering digunakan untuk penelitian *credit scoring data mining*. Terlihat juga bahwa performa algoritma CART terlihat lebih baik dengan menggunakan *dataset* dengan atribut numerik dibandingkan dengan *dataset* dengan atribut ordinal/kategori. Indikasi ini terlihat dari *dataset* *german-data* yang mayoritas atribut prediksinya adalah atribut dengan tipe ordinal/kategori memberikan hasil yang lebih rendah dibandingkan dengan *dataset* *german.data-numeric*. *German.data-numeric* adalah *dataset* yang sama dengan *german-data*, akan tetapi sudah dimodifikasi menjadi dataset dengan atribut numerik.

DAFTAR PUSTAKA

- [1] X.-L. Li, "An Overview of Personal Credit Scoring: Techniques and Future Work," *Int. J. Intell. Sci.*, vol. 02, no. 24, pp. 182–190, 2012.
- [2] S. M. Sadatrasoul, M. Gholamian, M. Siami, and Z. Hajimohammadi, "Credit scoring in banks and financial institutions via data mining techniques: A literature review," *J. AI Data Mining Journal AI Data Min.*, vol. 1, no. 2, pp. 119–129, 2013.
- [3] X. Wu, V. Kumar, Q. J. Ross, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, *Top 10 algorithms in data mining*, vol. 14, no. 1, 2008.
- [4] Roman Timofeev, "Classification and Regression Trees (CART) Theory and Applications," 2004.
- [5] J. Morgan, "Classification and Regression Tree Analysis," *PM931 Dir. Study Heal. Policy Manag.*, no. 1, p. 16, 2014.
- [6] R. J. Lewis, "An Introduction to Classification and Regression Tree (CART) Analysis," *2000 Annu. Meet. Soc. Acad. Emerg. Med.*, no. 310, p. 14p, 2000.
- [7] H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.