

PERBANDINGAN METODE JARINGAN SYARAF TIRUAN DAN POHON KEPUTUSAN UNTUK PREDIKSI CHURN

Dahnial

Program Studi Ilmu Administrasi Negara, STIA Satya Negara Palembang

e-mail: dahnial.dp@gmail.com

Abstract

Churn is the behavior in which customers stop using your product or service and move on to another company. Customer churns is a serious issue that must be addressed to ensure the survival of your business. Dealing with a customer churns starts with predicting which customers will churn. One approach to predicting a customer churn is data mining. This study compares his two classification methods in data mining: artificial neural networks and decision trees in the case of the banking industry. The results showed that the artificial neural network had an accuracy of 86% and a precision of 71%, which was better than the decision tree. However, the recall of decision trees is better than artificial neural networks, with a recall rate of 58%.

Keywords: data mining, churn, neural network, decision tree

Abstrak

Churn adalah proses dimana pelanggan berpindah ke perusahaan lain setelah mereka berhenti menggunakan produk atau layanan Anda. Perpindahan pelanggan merupakan masalah serius yang perlu diatasi demi keberlanjutan bisnis. Perjuangan melawan churn pelanggan dimulai dengan bisnis yang secara akurat memprediksi pelanggan mana yang kemungkinan besar akan berpindah. Salah satu cara dia memprediksi churn pelanggan adalah dengan melihat datanya. Penelitian ini membandingkan dua metode klasifikasinya dalam data mining: Jaringan Syaraf Tiruan dan Pohon Keputusan untuk Memprediksi Customer Churn di Sektor Perbankan. Studi tersebut menemukan bahwa jaringan saraf tiruan 86% lebih akurat dibandingkan pohon keputusan. Jaringan saraf tiruan memiliki tingkat akurasi 71% lebih tinggi dibandingkan pohon keputusan. Namun, dalam hal recall, pohon keputusan memiliki recall 58% lebih tinggi dibandingkan jaringan syaraf tiruan.

Kata Kunci: data mining, pemalsuan, jaringan syaraf tiruan, pohon keputusan

Diterima: 3 Desember 2023

Disetujui: 6 Desember 2023

Dipublikasi: 31 Desember 2023

Pendahuluan

Perkembangan teknologi yang pesat pada revolusi industri 4.0 yang kian mempengaruhi pola hidup masyarakat, turut menuntut bidang perbankan berinovasi agar tidak tertinggal. Terlebih lagi dunia perbankan dihadapkan dengan hadirnya pesaing dalam bidang finansial lain seperti industri *Financial* dan *Technology* (fintech) yang semakin berkembang pesat. Classification is the process of identifying and grouping an object into the same group or category Classification can be used to group a large-sized dataset, and some commonly used classification methods are CART (*Classification And Regression Tree*) and Naïve Bayes. This study discusses the comparison of CART and Naïve Bayes methods by measuring accuracy, precision, recall, and f1-score values with 3 scenarios of training and testing dataset distribution. Accuracy, precision, recall, and f1-score measurements are performed using a confusion matrix. The scenarios for training and testing dataset division are 70%, 80%, and 90% of the training dataset. From the results of the study, CART has the highest average accuracy and f1-score of 79.616% and 57.636% respectively, while the highest average accuracy and f1-score of Naïve Bayes are 75.104% and 62.004% respectively [1]. Pelanggan bagi perusahaan adalah aset utama, karenanya berbagai cara ditempuh oleh perusahaan agar pelanggan tidak menghentikan pembelian. Perusahaan juga harus mampu mempertahankan pelanggan potensial sehingga dapat mencegah pelanggan menghentikan pembelian dan berpindah ke perusahaan pesaing (*churn*). Hal tersebut tentunya tidak diinginkan oleh pihak perusahaan, sehingga berdasarkan permasalahan tersebut untuk mengetahui terlebih dahulu pelanggan yang memiliki kecenderungan untuk churn pada suatu waktu, yaitu dengan melakukan prediksi customer churn. Pada penelitian ini menggunakan metodologi CRISP- DM dengan model Recency, Frequency, dan Monetary (RFM) yang digunakan untuk mengidentifikasi karakter pelanggan dan membentuk segmentasi pelanggan, sehingga menghasilkan beberapa kelas pelanggan. Data yang digunakan untuk mendapatkan fitur RFM adalah data yang berasal dari data transaksi pelanggan. Data tersebut merupakan data yang digunakan untuk memprediksi customer churn dengan menggunakan algoritma Decision Tree (ID3) dan algoritma K- Nearest Neighbour (KNN). Penelitian ini dilakukan untuk mengetahui tingkat akurasi dan kecepatan kedua algoritma menggunakan data testing untuk memprediksi pelanggan yang melakukan churn berdasarkan segmentasi pelanggan perusahaan. Kata kunci: Pelanggan, Customer Churn, CRISP-DM, RFM, Segmentasi Pelanggan, Algoritma Decision Tree (ID3) dan Algoritma K-Nearest Neighbour (K-NN) [2]. Persaingan dalam bidang ini semakin ketat karena nasabah dapat memilih produk dan layanan yang dianggap lebih menguntungkan. Dampak persaingan ini adalah nasabah bisa beralih ke bank atau fintech lain. Dalam dunia bisnis, hal ini dikenal dengan istilah customer churn [3]. Churn adalah kecenderungan pelanggan untuk berhenti menggunakan produk atau layanan suatu perusahaan dan beralih ke produk atau layanan lain.

Pertumbuhan industri Fast Moving Consumer Goods (FMCG) masih menunjukkan double digit dan Indonesia menjadi pasar potensial bagi produk FMCG, sehingga persaingan antar perusahaan akan semakin ketat. Perusahaan harus berusaha untuk bertahan hidup, salah satu caranya adalah dengan menjaga loyalitas pelanggan. Teknik penambangan data dapat digunakan untuk memprediksi loyalitas pelanggan. Dalam pra-pemrosesan penambangan data, pemilihan fitur adalah salah satu hal penting untuk mengurangi jumlah fitur, menghilangkan kebisingan data yang tidak relevan, berlebihan, atau membawa efek langsung untuk aplikasi: mempercepat algoritma penambangan data, meningkatkan kinerja penambangan seperti keakuratan prediksi dan hasil yang komprehensif. Tulisan ini bertujuan untuk mengidentifikasi faktor-faktor relevan yang mempengaruhi kinerja klasifikasi loyalitas pelanggan dengan beberapa metode pemilihan fitur dan untuk membandingkan kinerja klasifikasi dalam prediksi

loyalitas pelanggan produk FMCG [4]. Perpindahan pelanggan (*Customer Churn*) adalah masalah serius dan perlu diatasi. Kegagalan untuk melakukannya akan berdampak besar pada bisnis Anda. Semakin banyak pelanggan yang pergi, maka semakin besar kemungkinan terjadi kesalahan pada produk atau jasa yang ditawarkan. Semakin dini hal ini diketahui, semakin cepat pula perusahaan dapat mengambil tindakan. Anda bisa menyusun strategi untuk memenangkan kembali pelanggan yang diprediksi akan pergi. Salah satu pendekatan untuk memprediksi churn pelanggan adalah dengan menggunakan pendekatan data mining.

Data mining adalah proses menemukan pola dan wawasan menarik dari sejumlah besar data [5]. Beberapa metode yang dapat digunakan untuk prediksi data mining antara lain pohon keputusan, jaringan syaraf tiruan, dan mesin vektor pendukung. Tujuan dari Prediksi Churn adalah untuk mengenali pola awal dan mengidentifikasi pelanggan yang kemungkinan besar akan melakukan churn. Dalam beberapa tahun terakhir, terdapat peningkatan minat untuk mempelajari churn di berbagai bidang seperti perbankan, telekomunikasi, dan asuransi [6]. Studi ini dilakukan oleh A Sharma dkk., yang menggunakan jaringan saraf tiruan untuk memprediksi churn pelanggan di industri seluler. Hasil penelitian ini memungkinkan untuk memprediksi churn pelanggan di industri telekomunikasi secara efektif dengan akurasi tinggi sebesar 92,35% [7]. Studi lain oleh Olivia et al. menggunakan pohon keputusan untuk menganalisis prediksi churn di industri seluler dan mencapai akurasi 94,43% [8]. Penelitian ini membandingkan implementasi teknik jaringan syaraf tiruan dan teknik pohon keputusan untuk memprediksi customer churn pada sektor perbankan. Churn nasabah menjadi permasalahan serius bagi Bank karena kehilangan nasabah berarti kehilangan pendapatan dan potensi kerugian jangka panjang. Untuk mencegah hal tersebut, diperlukannya teknologi masa kini untuk mengklasifikasi nasabah yang akan pergi (*churn*) atau bertahan melalui pemodelan machine learning algoritma Naïve Bayes Classifier dan algoritma Logistic Regression. Model klasifikasi dibuat untuk mempertahankan nasabah Bank berdasarkan data nasabah. Teknik SMOTE digunakan untuk mengatasi masalah imbalance class pada data yang digunakan, dan model di evaluasi dengan nilai accuracy, precision, recall, dan f1-score. Model terbaik yang dihasilkan ketika menggunakan teknik SMOTE ternyata mengalami penurunan nilai akurasi sebesar 13% pada algoritma Gaussian Naïve Bayes, model tanpa teknik SMOTE dengan akurasi sebesar 82% dan dengan teknik SMOTE sebesar 69%. Gaussian Naïve Bayes berhasil mengungguli jika dibandingkan dengan Complement Naïve Bayes dan Logistic Regression [9].

Churn merupakan suatu kecenderungan pelanggan untuk berhenti dari suatu produk atau layanan dan kemudian berpindah ke perusahaan lain. Customer churn merupakan masalah yang serius yang harus ditangani untuk keberlangsungan suatu perusahaan. Penanganan customer churn berawal dari ketepatan suatu perusahaan memprediksi customer mana yang diduga akan melakukan churn. Salah satu pendekatan dalam memprediksi customer churn adalah data mining. Penelitian ini membandingkan dua metode klasifikasi dalam data mining yaitu jaringan syaraf tiruan dan pohon keputusan untuk memprediksi customer churn pada industri perbankan. Hasil penelitian menunjukkan bahwa jaringan syaraf tiruan memiliki tingkat akurasi lebih baik dibanding pohon keputusan yaitu 86%. Selain itu jaringan syaraf tiruan juga memiliki nilai precision yang lebih baik daripada pohon keputusan, yaitu sebesar 71%. Akan tetapi untuk nilai recall, pohon keputusan memiliki nilai recall lebih baik dibanding dengan jaringan syaraf tiruan yaitu sebesar 58% [10].

Metode Penelitian

Pada penelitian ini dilakukan pencarian pola customer bank / nasabah yang berpotensi melakukan *churn* dan tidak *churn*. Tahapan-tahapan yang dilakukan antara lain ditunjukkan dalam Gambar 1. Diagram blok alur penelitian.



Gambar 1. diagram blok alur penelitian

Berdasarkan Gambar. 1 terdapat 4 tahapan dalam penelitian ini antara lain pengumpulan data, praproses, pembuatan model klasifikasi, evaluasi model, dan analisis hasil.

1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari dataset Kaggle yaitu Bank Customer churns Modeling. Dataset ini terdiri dari 10.000 record dengan 13 atribut dan 1 atribut kelas. Atribut-atribut ini meliputi:

1. RowNumber: atribut yang menyatakan no urut baris data
2. CustomerID: kode unik nasabah
3. Surname: nama nasabah
4. CreditScore: reliabilitas nasabah
5. Geography: asal daerah nasabah
6. Gender: jenis kelamin nasabah, wanita atau pria
7. Age: umur nasabah
8. Tenure: Lama tahun menjadi nasabah di bank tersebut
9. Balance: sisa saldo pada rekening
10. NumberOfProduct: banyaknya produk yang diambil nasabah pada bank tersebut
11. HasCrCard: status kepemilikan kartu kredit
12. IsActiveMember: status keaktifan nasabah
12. EstimatedSalary: estimasi gaji berdasarkan pemasukan
13. Exited: atribut kelas yang menyatakan churn atau tidak (1 = churn, 0 = tidak churn)

Dari 10000 record dataset nasabah perbankan, terdapat 2037 record data dengan kelas 1 (churn) dan 7963 dengan kelas 0 (tidak churn). Selanjutnya, data dipisahkan untuk *training* dan testing. Untuk training diambil 80% dari banyaknya data dan 20% digunakan untuk testing.

2. Praproses

Tahap pengolahan prapemrosesan merupakan tahap pertama pengolahan data sebelum pemodelan. Prapemrosesan terdiri dari dua sublangkah: pemilihan atribut dan proses normalisasi data. Pemilihan atribut merupakan proses pemilihan atribut yang penting dan digunakan dalam proses pengolahan data. Langkah selanjutnya adalah langkah normalisasi. Normalisasi adalah proses transformasi yang menskalakan atribut numerik ke rentang kecil, seperti -1,0 hingga 1,0 atau 0,0 hingga 1,0 [11]. Pada penelitian ini kami melakukan normalisasi data dengan menggunakan metode normalisasi Min-Max. Gambar 2 menunjukkan rumus normalisasi min-max.

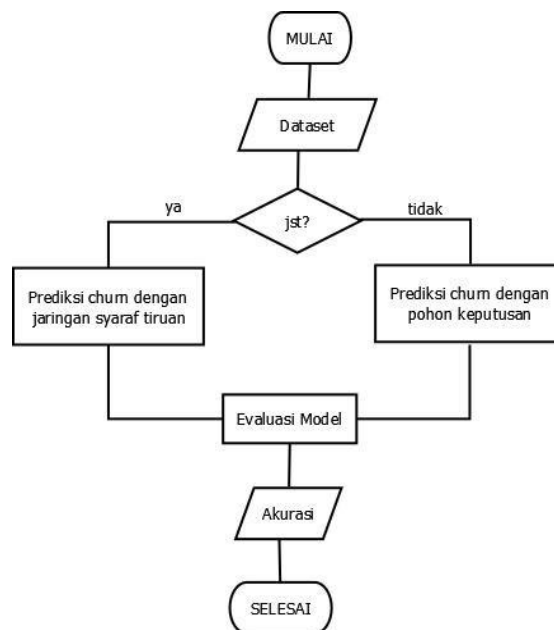
$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{new_max}_A - \text{new_min}_A) + \text{new_min}_A$$

Gambar 2. rumus min-max normalization

Untuk setiap fitur, nilai minimumnya diubah menjadi 0, nilai maksimumnya diubah menjadi 1, dan semua nilai lainnya diubah menjadi angka desimal antara 0 dan 1.

3. Pembuatan Model Klasifikasi

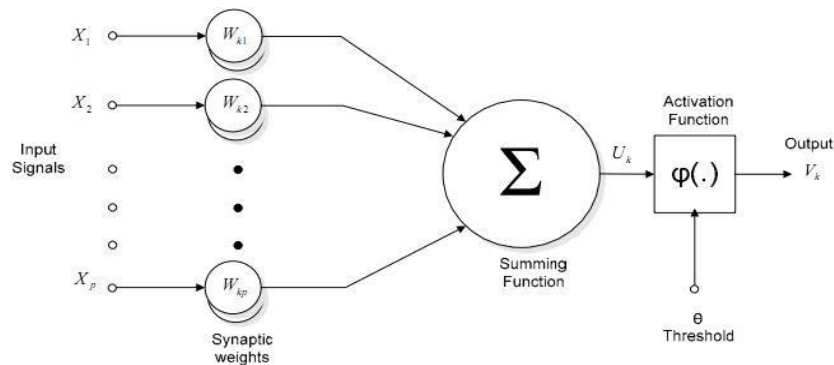
Model yang dibangun untuk prediksi churn pada penelitian ini menggunakan dua model sebagai perbandingan yaitu jaringan syaraf tiruan dan pohon keputusan. Bandingkan kedua model ini untuk melihat mana yang memiliki akurasi lebih baik. Diagram alir proses klasifikasi ditunjukkan pada Gambar 3



Gambar 3. flowchart proses klasifikasi

1) JST (Jaringan Syaraf Tiruan)

Jaringan syaraf tiruan (JST) adalah model jaringan kompleks yang terdiri dari neuron (node) yang disebut neuron. ANN terinspirasi oleh penelitian biologi pada jaringan otak dan sistem saraf manusia. ANN dapat digunakan untuk mensimulasikan aktivitas saraf selama pemrosesan informasi di otak manusia. Model jaringan saraf tiruan ditunjukkan Pada Gambar 4.



Gambar 4. Model struktur jaringan saraf tiruan [12]

Jaringan saraf terdiri dari kelompok neuron (misalnya) yang tersusun berlapis-lapis.

- Lapisan masukan: Lapisan masukan berfungsi sebagai koneksi jaringan ke dunia luar (sumber data).
- Lapisan tersembunyi: Suatu jaringan mungkin mempunyai beberapa lapisan tersembunyi atau tidak ada sama sekali.
- Lapisan keluaran: Keluaran neuron pada lapisan ini dianggap sebagai hasilnya.

Proses jaringan saraf tiruan dapat dirangkum dalam 7 langkah besar antara lain:

1. Inisialisasi bobot secara acak, bobot merupakan angka kecil mendekati 0, tetapi bukan 0.
2. Masukkan input pertama di *input layer*, setiap fitur dimasukkan untuk satu simpul.
3. *Forward-Propagation*: dari kiri ke kanan, neuron diaktifkan dengan cara dampak dari setiap aktivasi neuron dibatasi oleh bobot dan berjalan hingga mendapatkan bobot yang sesuai
4. Bandingkan hasil prediksi dengan hasil yang sebenarnya. Ukur error yang dihasilkan
5. *Back-propagation*: dari kanan ke kiri. Perbarui bobot sesuai dengan seberapa besar error yang dihasilkan dari tahap sebelumnya. Tingkat pembelajaran menentukan seberapa banyak bobot diperbaharui.
6. Ulangi langkah 1 hingga 5 dan perbarui bobot setelah setiap pengamatan.
7. Ketika seluruh pelatihan ditetapkan melewati JST, lakukan sebanyak *epoch* yang telah ditentukan.

1) *Decision Tree* (Pohon Keputusan)

Decision Tree (Pohon Keputusan) merupakan teknik klasifikasi dengan memodelkan pohon atau struktur hierarki. Proses pembelajaran pohon keputusan menugaskan setiap kumpulan atribut ke suatu kelas dari kelas-kelas yang telah ditentukan sebelumnya [12]. Salah satu algoritma yang digunakan untuk membangun pohon keputusan adalah ID3. Algoritma ID3 menggunakan entropi dan perolehan informasi. Entropi mengukur jumlah informasi yang terkandung dalam suatu atribut, dan perolehan informasi mewakili pengurangan entropi. Secara umum algoritma ID3 untuk membangun pohon keputusan adalah:

1. Menghitung perolehan informasi untuk setiap atribut
2. Memilih atribut dengan nilai perolehan informasi maksimum
3. Bentuk node yang memuat atribut tersebut
4. Proses penghitungan informasi Penguatan tersebut berlanjut untuk diimplementasikan sampai data dimasukkan ke dalam kelas yang sama.
5. Atribut yang dipilih tidak lagi dimasukkan dalam perhitungan nilai information gain.

Evaluasi

Proses yang dilakukan pada tahap evaluasi adalah perhitungan presisi, presisi, dan recall. Karena churn merupakan klasifikasi biner (churn atau tidak), maka nilai akurasi ditentukan dengan membagi jumlah data pengujian yang benar yang terdiri dari true positif (TP) dan true negative (TN) dengan jumlah total data pengujian. Rumus perhitungan akurasi ditunjukkan pada Gambar 5.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Gambar 5. Rumus menghitung akurasi

TP: True positive

TN: True negative

FP: False positive

FN: False Negative

Selain presisi, presisi dan recall juga sering digunakan sebagai evaluasi pada permasalahan klasifikasi. Akurasi adalah perbandingan ketepatan antara informasi yang diminta dengan jawaban yang diberikan [13]. Dalam hal ini, proporsi pelanggan yang diprediksi dengan benar akan cacat (churn) sebanding dengan jumlah total pelanggan yang diprediksi akan cacat berdasarkan model klasifikasi. Gambar 6 menunjukkan rumus akurasi:

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

Gambar 6. Rumus menghitung precision

Recall adalah rasio elemen relevan yang dipilih dengan jumlah total elemen relevan yang tersedia [8]. Dalam hal ini, persentase churn pelanggan yang diprediksi dengan benar oleh sistem dibandingkan dengan jumlah total churn pelanggan sebenarnya di data Anda. Gambar 7 menunjukkan rumus menghitung *recall*.

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

Gambar 7. Rumus menghitung recall

Hasil dan Pembahasan

“Hasil dan Pembahasan” memaparkan hasil penelitian disertai penjelasan.

1. *Preprocessing*

Hasil dari preprocessing berupa data dengan atribut yang dipilih dan dinormalisasi menggunakan normalisasi min-max. Atribut yang dipilih dari 13 atribut tersebut adalah 11 atribut (CreditScore, Geography, Gender, Age, Tenure, Balance, NumberofProduct, HasCrCard, IsActiveMember, claimSalary) dan atribut Exited sebagai kelas atributnya. Untuk atribut yang nilainya non-numerik, tipe datanya diubah menjadi tipe data numerik. Misalnya, dalam Geografi, Perancis diwakili oleh 1, Spanyol diwakili oleh 2, dan Jerman diwakili oleh 3. Namun, untuk gender, 0 mewakili laki-laki dan 1 mewakili perempuan.

Hasil normalisasi data preprocessing menggunakan normalisasi min-max ditunjukkan pada Tabel 1.

Tabel 1 *Data hasil praproses*

CreditScore	Geography	Gender	Age	...	Exited
0.538	0	0	0.324	...	1
0.516	0.5	0	0.311	...	0
0.304	0	0	0.324	...	1
0.698	0	0	0.284	...	0
1.000	0.5	0	0.338	...	0
0.590	0.5	1	0.351	...	1
0.944	0	1	0.432	...	0
0.052	1	0	0.149	...	1
0.302	0	1	0.351	...	0
...	...	...	...	...	...
0.884	0	0	0.135	...	0

1. Evaluasi dari Model Klasifikasi

Informasi tersebut dipartisi menjadi informasi pengujian dan informasi persiapan. Informasi persiapan yang digunakan menyumbang 80% dari kumpulan data, 8000 catatan. Informasi pengujian yang digunakan adalah 20 record, 2000 record. Persiapan klasifikasi dilakukan dengan menggunakan dua strateginya: strategi pohon pilihan dan susunan saraf palsu. Parameter yang digunakan untuk membentuk tampilan susunan saraf yang diproduksi adalah dua lapisan tertutup dan karya penerapan Relu, dan ID3 digunakan untuk pegangan klasifikasi menggunakan pohon pilihan.

Selanjutnya, persiapan penilaian dilakukan dengan menggunakan sistem saraf buatan dan pohon pilihan, yang bertujuan untuk menentukan ketepatan, ketepatan, dan peninjauan persiapan klasifikasi. Hasil dari strategi pengorganisasian saraf palsu ditunjukkan pada Tabel 2.

Tabel 2 *Confusion Matrix dengan JST*

		CLASSIFIER PREDICTION	
		0	1
ACTUAL VALUE	0	1513	82
	1	202	203

Berdasarkan *disarray framework* di Tabel 2 dapat disimpulkan bahwa terdapat 1513 customer yang tidak churn yang diprediksi tepat, 203 *client churn* diprediksi tepat, 82 client yang ternyata tidak churn yang diprediksi menjadi churn, dan 202 client yang churn namun dianggap oleh *classifier* sebagai tidak churn. Dari *confussion lattice* tersebut dapat dihitung nilai akurasi yaitu 86%, accuracy 71%, dan review 50%. Sedangkan hasil dari metode pohon keputusan disajikan pada Tabel 3.

Tabel 3 *Confusion Matrix dengan Pohon Keputusan*

		CLASSIFIER PREDICTION	
		0	1
ACTUAL VALUE	0	1381	214
	1	171	234

Berdasarkan *disarray network* di Tabel 3 dapat disimpulkan bahwa terdapat 1381 customer yang tidak churn yang diprediksi tepat, 234 *client churn* diprediksi tepat, 214 *client* yang tidak *churn* yang namun diprediksi menjadi churn, dan 171 *client* yang churn namun dianggap oleh classifier sebagai tidak *churn*. Dari *confussion lattice* tersebut dapat dihitung nilai akurasi yaitu 81%, precision-nya 52%, dan nilai recall-nya 58%. Perbandingan nilai akurasi, exactness, dan review dideskripsikan pada Tabel 4.

Tabel 4 *Evaluasi JST dan Pohon Keputusan*

Hasil Evaluasi	JST	Pohon Keputusan
Akurasi	86%	81%
Precision	71%	52%
Recall	50%	58%

Berdasarkan Tabel 4, hasil akurasi dan accuracy dari show jaringan saraf tiruan lebih tinggi dibanding menggunakan pohon keputusan. Jaringan saraf tiruan memiliki persentase ketepatan lebih besar untuk memprediksi nasabah churn dibanding dengan pohon keputusan. Selain itu, dalam hal presisi, jaringan saraf tiruan memiliki proporsi yang lebih baik untuk memprediksi nasabah churn yang juga positif churn pada information yang sebenarnya dibanding dengan pohon keputusan.

Akan tetapi nilai review dari jaringan saraf tiruan lebih kecil daripada nilai review pohon keputusan. Hal ini menunjukkan bahwa show jaringan saraf tiruan dapat memisahkan nasabah yang churn lebih sedikit dan masih banyak *client churn* yang lolos tidak terprediksi dibanding dengan pohon keputusan.

Kesimpulan

Berdasarkan perbandingan jaringan saraf tiruan dan pohon keputusan untuk prediksi churn pelanggan pada kumpulan data 10 atribut dan 10.000 catatan (8.000 catatan pelatihan dan 2.000 catatan pengujian), jaringan saraf tiruan lebih baik. Kita dapat menyimpulkan. Akurasi kinerjanya lebih tinggi dibandingkan pohon keputusan, yaitu 86%. Selain itu, jaringan saraf tiruan juga memiliki nilai akurasi yang lebih baik (71%) dibandingkan pohon keputusan. Namun dari segi nilai evaluasi, pohon keputusan mempunyai nilai evaluasi yang lebih baik dibandingkan jaringan syaraf tiruan yaitu sebesar 58%. Namun, perbedaan antara nilai evaluasi jaringan syaraf tiruan dan pohon keputusan tidak terlalu besar, dan masih ada ruang untuk perbaikan. Saran untuk penelitian selanjutnya adalah menangani dataset yang tidak seimbang sebelum melakukan proses klasifikasi. Hal ini diharapkan dapat meningkatkan nilai presisi, presisi, dan evaluasi.

Daftar Pustaka

- Adhitya, R. R., Witanti, W., & Yuniarti, R. Perbandingan Metode Cart Dan Naïve Bayes Untuk Klasifikasi Customer Churn. *Infotech Journal*, 9(2), 307-318. 2023
- Agustina Indriyani, H. *Algoritma Decision Tree (Id3) Dan K-Nearest Neighbour (K-Nn) Dalam Prediksi Costumer Churn Berdasarkan Segmentasi Pelanggan (Studi Kasus Radity Craft Tasikmalaya)* (Doctoral Dissertation, Universitas Siliwangi). 2019.
- X. Yu, S. Guo, J. Guo, And X. Huang, “An Extended Support Vector Machine Forecasting Framework For Customer Churn In E-Commerce,” *Expert Syst. Appl.*, Vol. 38, No. 3, Pp. 1425–1430, 2011.
- Satrio, B. (2022). Analisis Perbandingan Metode Pemilihan Atribusi Untuk Memprediksi Loyalitas Pelanggan. *Jurnal Portal Data*, 2(8). 2022
- J. Han, J. Pei, And M. Kamber, *Data Mining: Concepts And Techniques*. Elsevier, 2011.
- T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, And K. C. Chatzisavvas, “A Comparison Of Machine Learning Techniques For Customer Churn Prediction,” *Simul. Model. Pract. Theory*, Vol. 55, Pp. 1–9, 2015.
- A. Sharma, D. Panigrahi, And P. Kumar, “A Neural Network Based Approach For Predicting Customer Churn In Cellular Network Services,” *Arxiv Prepr. Arxiv1309.3945*, 2013.
- C. Olivia, I. Indwiarti, And Y. Sibaroni, “Analisis Prediksi Churn Menggunakan Metode Logistic Regression Dan Algoritma Decision Tree,” *Eproceedings Eng.*, Vol. 2, No. 2, 2015.
- Mulia, C., & Kurniasih, A. (2023, December). Teknik Smote Untuk Mengatasi Imbalance Class Dalam Klasifikasi Bank Customer Churn Menggunakan Algoritma Naïve Bayes Dan Logistic Regression. In *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer Dan Aplikasinya* (Vol. 4, No. 2, Pp. 552-559). 2023
- Irmanda, H. N., Astriratma, R., & Afrizal, S. Perbandingan Metode Jaringan Syaraf Tiruan Dan Pohon Keputusan Untuk Prediksi Churn. *Jsi: Jurnal Sistem Informasi (E-Journal)*, 11(2). 2019
- H. Junaedi, H. Budianto, I. Maryati, And Y. Melani, “Data Transformation Pada Data Mining,” *Pros. Konf. Nas. “Inovasi Dalam Desain Dan Teknol. Ideatech*, 2011.
- Z. Azmi And M. Dahria, “Decision Tree Berbasis Algoritma Untuk Pengambilan Keputusan,” *J. Ilm. Saintikom*, Vol. 12, 2013.
- P. Mayadewi And E. Rosely, “Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining,” *Sesindo 2015*, Vol. 2015, 2015.