

PREDIKSI LOYALITAS PELANGGAN TELEKOMUNIKASI MENGUNAKAN *LOGISTIC REGRESSION* DENGAN SELEKSI FITUR *PARTICLE SWARM OPTIMIZATION*

Stefanus Santosa¹, dan Fenilinas Adi Artanto²
Pascasarjana Universitas Dian Nuswantoro

Abstract

For many companies, finding a reason to lose customers, measurement of customer loyalty and regain customers have become essential, including for telecommunication companies. The telecommunications company is one of the industry, where the customer really needs special attention because it is very influential in maintaining the stability of the company's revenue. The telecommunications industry has always faced the threat of financial loss resulting from customer loyalty. The customer who leaves the service is usually called churners. Find churners can help telecommunications companies in retaining customers and keep the company financially. This study used Logistic Regression algorithm with feature selection Particle Swarm Optimization to predict customer loyalty telecommunications. The test results obtained using ANN algorithm accuracy value amounted to 94.80%, and Logistic Regression Algorithm with Particle Swarm Optimization feature selection shows the value of accuracy of 97.65%, and the AUC value of 0.99, then the Logistic Regression algorithm with feature selection Particle Swarm Optimization can improve the accuracy of prediction telecommunications customer loyalty

Keyword : Data Mining, feature selection, Particle Swarm Optimization, Logistic Regression, Telecommunication Customers Churn, ..

1. PENDAHULUAN

1.1. Latar Belakang Masalah

Bagi banyak perusahaan, menemukan alasan kehilangan pelanggan, pengukuran loyalitas pelanggan dan mendapatkan kembali pelanggan telah menjadi hal yang penting termasuk bagi perusahaan telekomunikasi. Perusahaan telekomunikasi merupakan salah satu industri dengan pelanggan yang sangat membutuhkan perhatian khusus karena sangat berpengaruh dalam mempertahankan kestabilan pendapatan perusahaan. Industri telekomunikasi selalu menghadapi ancaman kerugian finansial yang diakibatkan dari loyalitas pelanggan.

Para pelanggan yang berhenti menggunakan produk biasanya disebut churners. Menemukan churners dapat membantu perusahaan mempertahankan pelanggan [1]. Kondisi dari loyalitas pelanggan akan mempengaruhi persaingan antarperusahaan telekomunikasi. Dengan begitu banyak kompetisi, suatu perusahaan harus fokus menjaga pelanggan yang sudah ada dengan memenuhi layanan yang dibutuhkan, karena biaya menarik pelanggan baru, biasanya lebih besar daripada biaya untuk mempertahankan pelanggan saat ini [2]. Jika suatu perusahaan dapat memprediksi loyalitas pelanggan sebelum menutup akun atau berhenti, perusahaan bisa melakukan cara untuk mempertahankan pelanggan tersebut [3]. Beberapa Algoritma data mining bisa digunakan untuk memprediksi loyalitas pelanggan.

Algoritma data mining dapat digunakan untuk membantu dalam pemilihan variabel dan membangun model untuk memprediksi loyalitas pelanggan, misalnya Algoritma Genetika (AG), Regresi, Artificial Neural Networks (ANN), *Decision Tree* (DT), Markov Model, Analisis Cluster, dan optimasi [4]. Salah satu algoritma yang dapat diterapkan dalam melakukan prediksi adalah *Logistic Regression*.

Pada penelitian sebelumnya yang dilakukan oleh Irvan, prediksi loyalitas pelanggan telekomunikasi dilakukan dengan menggunakan algoritma backpropagation yang dioptimasi dengan *Particle Swarm*

Optimization (PSO). Hasil penelitian menunjukkan adanya peningkatan akurasi sebesar 0.57% dan nilai AUC sebesar 0.106 menjadi 86.05% dan nilai AUC 0.637[5].

Oebidat R, et al [6], juga melakukan penelitian tentang prediksi loyalitas pelanggan telekomunikasi dengan data set yang sama menggunakan *K-Means Clustering* yang digabung dengan *Genetic Programming*. Pertama *K-Means* digunakan untuk menyaring dataset dari outlier dan *Genetic Programming* diterapkan untuk membangun pohon klasifikasi yang mampu mengklasifikasikan pelanggan, lalu dibandingkan dengan C4.5 dan ANN. Dari hasil akurasi didapatkan nilai akurasi *K-Means + Genetic Programming* 91.40%, C4.5 94.10%, dan ANN 94.80%.

Verbeke et al[7], melakukan penelitian tentang prediksi loyalitas pelanggan telekomunikasi dengan menghitung keuntungan maksimum yang dapat dihasilkan, menggunakan 21 algoritma klasifikasi yang diterapkan pada 11 dataset dari operator telekomunikasi di seluruh dunia. Percobaan yang dilakukan digunakan untuk mengevaluasi dampak dari pemilihan input, oversampling, dan teknik klasifikasi terhadap kinerja model prediksi loyalitas pelanggan. Penelitian ini menganjurkan bahwa keuntungan maksimum harus digunakan untuk mengavaluasi model prediksi loyalitas pelanggan. Pada penelitian ini algoritma yang menunjukkan kinerja yang terbaik secara keseluruhan adalah *Decision Trees*, meskipun sejumlah algoritma lainnya tidak signifikan mengungguli.

B.Q. Huang, et.al [8] Penelitian ini tentang loyalitas pengguna jasa telekomunikasi menggunakan algoritma ANN, SVM, dan DT. Dataset yang digunakan 47.391 pengguna dipilih secara acak dari beberapa kantor telekomunikasi di Ireland. Hasil penelitian menunjukkan algoritma-algoritma tersebut lebih efisien daripada yang sudah ada, namun teknik pemodelan SVM tidak cocok untuk aplikasi yang digunakan.

Guangli Nie et. All [3] penelitian tentang loyalitas pelanggan menggunakan dua algoritma (*Decision Tree* dan *Logistic Regression*) yang diterapkan pada 135 variabel untuk meringkas perilaku dan pilihan nasabah kartu kredit. Seleksi variabel telah menunjukkan bahwa informasi demografis membuat sedikit kontribusi terhadap hasil prediksi. Penelitian Nie ini membahas tentang pohon keputusan yang telah banyak digunakan di berbagai bidang. Algoritma *Decision Tree* mempunyai keandalan, namun pohon keputusan yang dihasilkan tidak optimal dan tidak bisa menggunakan sampel yang lebih besar.

Dari penelitian terdahulu akurasi terbaik prediksi loyalitas pelanggan telekomunikasi dicapai oleh algoritma ANN dengan nilai akurasi sebesar 94.8% [6]. Meskipun demikian algoritma ANN memiliki kelemahan, yakni rentan terhadap overfitting. ANN memiliki struktur yang kompleks, hidden node saat digunakan dalam jaringan yang besar cenderung overfit karena hidden node mendeteksi hampir semua interaksi yang mungkin. Oleh sebab itu diperlukan pendekatan lain yang dapat mengurangi kelemahan tersebut, dalam hal ini digunakan *Logistic Regression*. Pemilihan fitur pelanggan menggunakan *Particle Swarm Optimization* dan prediksi loyalitas pelanggan menggunakan *Logistic Regression*.

1.2. Prediksi Loyalitas Pelanggan Telekomunikasi Menggunakan *Logistic Regression* dengan Pemilihan Fitur *Particle Swarm Optimization*

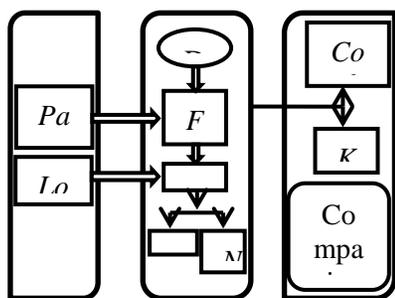
Penelitian-penelitian yang sudah dibahas sebelumnya menunjukkan bahwa seleksi variabel memiliki pengaruh yang signifikan terhadap akurasi prediksi. Pemilihan variabel secara manual adalah tugas yang melelahkan dan dapat mengabaikan parameter penting. Teknik optimasi yang sering digunakan untuk memfasilitasi pencarian variabel yang relevan adalah *Particle Swarm Optimization*[9]. *Particle Swarm Optimization* memiliki kelebihan dibanding algoritma yang lain untuk pemilihan fitur dan memiliki kinerja lebih unggul untuk banyak masalah optimasi, lebih cepat prosesnya, dan memiliki tingkat pencapaian konvergensi yang lebih stabil [9].

Karakteristik *Particle Swarm Optimization* dalam hal ini berkaitan dengan interaksi sosial yang mempromosikan pembagian informasi antarpartikel yang akan membantu dalam pencarian solusi yang optimal[10]. *Particle Swarm Optimization* memiliki beberapa parameter seperti posisi, kecepatan,

kecepatan maksimum, percepatan konstanta, dan berat inersia. Dalam *Particle Swarm Optimization* terdapat beberapa cara untuk melakukan optimasi diantaranya: meningkatkan bobot atribut (attribute *weight*) terhadap semua atribut atau variabel yang dipakai, menseleksi atribut (attribute selection), dan feature. Berdasarkan penjelasan tersebut penelitian ini menggunakan *Particle Swarm Optimization* untuk memilih atribut-atribut yang relevan dalam prediksi pelanggan yang tidak loyal.

Algoritma ANN memiliki kelebihan dalam memprediksi data set yang hubungan antara input dan output-nya tidak jelas, hal ini dikarenakan ANN memiliki hidden node. Tetapi data set dalam penelitian ini memiliki hubungan yang jelas, seperti banyaknya panggilan yang digunakan pada pagi, siang, dan malam hari yang akan mempengaruhi faktor loyalitas pelanggan. Pada sisi lain hidden node dalam ANN yang akan menganalisis data yang lebih jelas hubungannya akan menyebabkan terjadinya *over-fitting*, oleh karena itu penelitian ini tidak menggunakan ANN melainkan *Logistic Regression* untuk melakukan prediksi.

Logistic Regression dapat menganalisis data training yang kecil. *Logistic Regression* memiliki keuntungan yang berbeda dengan ANN. *Logistic Regression* mampu melakukan perhitungan yang sederhana untuk hasil prediksi. Dalam *Logistic Regression* pemilihan variabel harus dipilih dengan hati-hati sehingga hasil prediksi akan lebih akurat tapi tanpa *over-fitting* data. Disinilah *Particle Swarm Optimization* akan lebih memberikan manfaat kepada *Logistic Regression* daripada ANN, karena dataset yang digunakan memiliki hubungan antara atribut dan label yang jelas. Berikut ini adalah diagram penjelasan konsep tersebut di atas.



Gambar 1.

Gambar 1. Kerangka Pikir Prediksi Loyalitas Pelanggan Telekomunikasi Menggunakan *Logistic Regression* dengan Seleksi Fitur *Particle Swarm Optimization*

Dengan demikian dapat diharapkan bahwa pasangan kedua algoritma ini dengan fungsi masing-masing akan menghasilkan akurasi yang lebih baik daripada penelitian-penelitian sebelumnya.

2. METODE PENELITIAN

2.1. Pengumpulan data

Sumber data set prediksi pelanggan telekomunikasi yang hilang diperoleh dari sebuah website yaitu :

<http://www.informatics.buu.ac.th/~ureerat/321641/Weka/Data%20Sets/Churn/>

Jumlah dataset yang digunakan sebanyak 3333 record dan 20 atribut data tersebut dapat dilihat pada tabel 1 di bawah ini [5]:

Tabel 1 Atribut Dataset

NO	ATRIBUT	Keterangan
1	State	Untuk 50 negara bagian columbia
2	Account Legth	Layanan akun aktif
3	Area code	Kode area
4	Phone	Nomor telepon pelanggan
5	Int Plan	Perencanaan internasional
6	Vmail plan	Perencanaan pesan suara
7	Vmail Message	Jumah pesan suara
8	Day Mins	Jumlah layanan permenit dalam sehari pada siang hari
9	Day Calls	Jumlah panggilan dalam sehari dalam siang hari
10	Day Charge	Jumlah biaya panggilan dalam sehari dalam siang hari
11	Eve Mins	Jumlah layanan permenit dalam sehari pada sore hari
12	Eve Calls	Jumlah panggilan dalam sehari dalam sore hari
13	Eve Charge	Jumlah biaya panggilan dalam sehari dalam sore hari
14	Night Mins	Jumlah layanan permenit dalam sehari pada malam hari
15	Night Calls	Jumlah panggilan dalam sehari dalam malam hari
16	Night Charge	Jumlah biaya panggilan dalam sehari dalam malam hari
17	Intl Mins	Jumlah layanan permenit pada panggilan internasional
18	Intl Calls	Jumlah panggilan internasional
19	Intl Charge	Jumlah biaya pada panggilan internasional
20	CustServ Calls	Jumlah panggilan ke layanan pelanggan

2.2. Eksperimen

Dataset pelanggan telekomunikasi yang sudah didapat dengan jumlah record sebanyak 3333 record dan 20 atribut. Tahap awal dilakukan feature selection dengan *Particle Swarm Optimization* kemudian dilanjutkan dengan prediksi menggunakan *Logistic Regression*.

Dalam *Particle Swarm Optimization* dilakukan beberapa iterasi, mencari nilai fitness, jika nilai fitness lebih besar dari nilai best fitness maka nilai best fitness diperbaharui. Setelah itu nilai fitness dibandingkan dengan nilai *global fitness*, jika nilai fitness lebih besar dari nilai *global fitness* maka nilai *global fitness* harus diperbaharui, setelah itu diperbarui nilai kecepatannya dan posisinya. Dalam hal ini digunakan nilai *population size = 5*, *maximum number of generations = 30*, *inertia weight = 1*. *Local best weight = 1*, *global best weight = 1*, *min weight = 0* dan *max weight = 1*. Sebelum proses selesai, setelah memperbarui posisi partikel, hasil yang ada akan di uji lebih dahulu dengan *Logistic Regression*, baru setelah itu masuk terminasi sudah optimal atau belum. Hasil output atribut adalah atribut yang memiliki nilai *weight* mendekati 1.

Prediksi dengan *Logistic Regression* menggunakan *setting kernel type dot*, dengan nilai *kernel cache = 200*, *C = 1*, *convergence epsilon 0,0010*, dan *max iterations 100.000*. Selain itu juga digunakan fasilitas *nominal to numerik* sebelum melakukan proses prediksi *Logistic Regression*.

2.3. Evaluasi

Hasil eksperimen diuji dengan menggunakan *ROC Curve*. *ROC Curve* menggambarkan kelas positif dalam bentuk kurva. Pengujian dilakukan dengan menghitung nilai AUC (Area Under Curve), semakin tinggi nilai AUC dalam *ROC Curve*, maka semakin baik pula model klasifikasi yang terbentuk.

Setelah nilai *Accuracy* dari model prediksi *Logistic Regression* yang menggunakan feature selection *Particle Swarm Optimization* didapatkan, hasilnya dibandingkan dengan nilai *Accuracy* dari penelitian terdahulu yang menggunakan dataset yang sama namun dengan pendekatan yang berbeda ANN. Pada penelitian Oebidat R, et al [6], yang menggunakan ANN didapatkan nilai *Accuracy* sebesar 94,80%.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Eksperimen

Atribut-atribut yang diseleksi adalah State, Account Length, Area Code, Phone, Int Plan, Vmail Plan, Vmail Message, Day Mins, Day Calls, Day Charge, Eve Mins, Eve Calls, Eve Charge, Night Mins, Night Calls, Night Charge, Intl Mins, Intl Calls, Intl Charge, CustServ Calls, dan 1 atribut sebagai label yaitu Churn. Dengan menggunakan *Particle Swarm Optimization* diperoleh *weight* dari setiap atribut seperti yang tercantum dalam tabel berikut ini.

Tabel 2. Hasil Seleksi Atribut

Atribut	Weight
State	1.000
Account Length	0.000
Area Code	0.000
Phone	0.025
Intl Plan	0.929
Vmail Plan	0.000
Vmail Message	0.100
Day Mins	0.130
Day Calls	1.000
Day Charge	1.000
Eve Mins	0.000
Eve Calls	1.000
Eve Charge	1.000
Night Mins	0.000
Night Calls	0.706
Night Charge	0.000
Intl Mins	0.000
Intl Calls	0.000
Intl Charge	1.000
Custserv Calls	0.031

Dari 20 variabel setelah dilakukan seleksi atribut diperoleh 12 atribut yaitu: State, Phone, Intl Plan, Vmail Message, Day Mins, Day Calls, Day Charger, Eve Calls, Eve Charger, Night Calls, Intl Charge, Custserv Calls yang digunakan untuk memprediksi loyalitas pelanggan dengan nilai bobotnya yang lebih dari 0 atau mendekati 1. Sedangkan atribut lain dengan bobot = 0.00 seperti Account Length, Area Code, Vmail Plan, Eve Mins, Night Mins, Night Charger, Intl Mins, Intl Calls tidak digunakan dalam proses prediksi. Atribut State, Day Call, Day charger, Eve Call, Eve Charger dan Intl Charger memiliki bobot tertinggi, yakni sebesar 1.000.

Hasil evaluasi menggunakan *Confusion Matrix* ditunjukkan oleh tabel di bawah ini.

Tabel 3. Hasil *Confusion Matrix* untuk Algoritma *Logistic Regression* dengan Feature Selection *Particle Swarm Optimization*

Classification	Predicted Class		
		True no	True yes
Observed Class	Pred. No	2846	74
	Pred. Yes	4	409

Dari tabel di atas diperoleh jumlah *True Positive* (TP) adalah 2846 diklasifikasikan sebagai 1 sesuai dengan prediksi yang dilakukan dengan Algoritma *Logistic Regression*. *False Negative* (FN) sebanyak 74 data diprediksi sebagai 1 tetapi ternyata 2, kemudian *True Negative* (TN) sebanyak 409 data sebagai 2 sesuai dengan prediksi, dan *False Positive* (FP) sebanyak 4 data diprediksi 2 ternyata 1.

Tingkat akurasi yang dihasilkan dengan menggunakan algoritma *Logistic Regression* dengan fitur seleksi *Particle Swarm Optimization* adalah sebesar 97.65%, selain itu nilai, *sensitivity*, *specificity*, *ppv*, dan *npv* bisa diperoleh melalui persamaan di bawah ini.

$$\begin{aligned}
 \text{Accuracy} &= (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) = (2846 + 409) / (2846 + 409 + 4 + 74) = 0.9765 \\
 \text{Sensitivity} &= \text{TP} / (\text{TP} + \text{FN}) = 2846 / (2846 + 74) = 0.9746 \\
 \text{Specificity} &= \text{TN} / (\text{TN} + \text{FP}) = 409 / (409 + 4) = 0.9903 \\
 \text{PPV} &= \text{TP} / (\text{TP} + \text{FP}) = 2846 / (2846 + 4) = 0.9985 \\
 \text{NPV} &= \text{TN} / (\text{TN} + \text{FN}) = 409 / (409 + 74) = 0.8467
 \end{aligned}$$

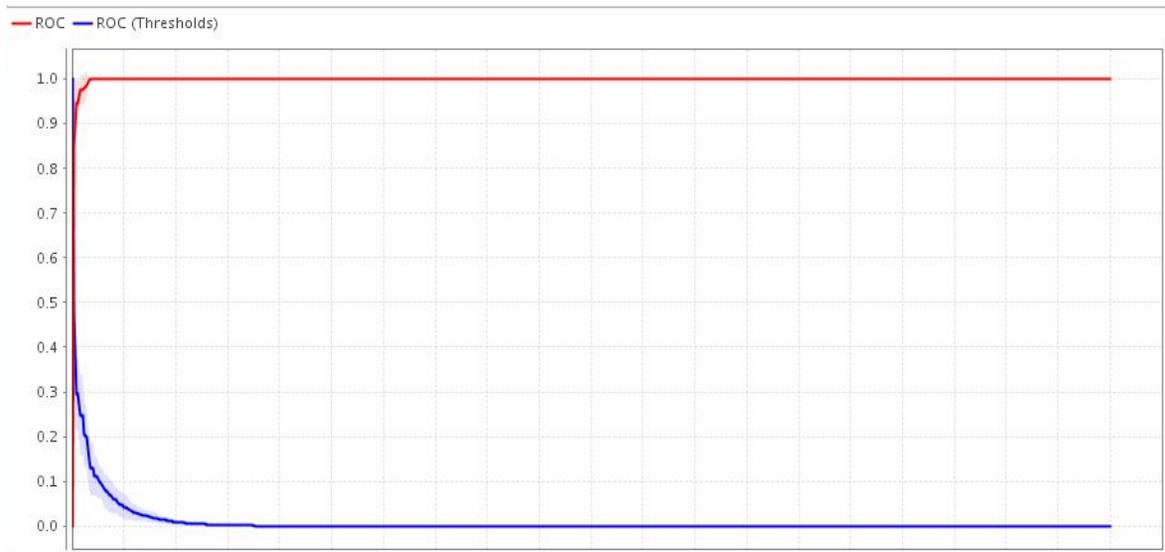
Hasil perhitungan terlihat pada tabel di bawah ini :

Tabel 4 Nilai *Accuracy*, *Sensitivity*, *Specificity*, *PPV* dan *NPV* Algoritma *Logistic Regression* dan *Particle Swarm Optimization*

	Nilai (%)
<i>Accuracy</i>	97.65
<i>Sensitivity</i>	97.46
<i>Specificity</i>	99.03
PPV	99,85
NPV	84.67

Hasil perhitungan divisualisasikan juga dengan kurva ROC. Perbandingan kedua class bisa dilihat pada gambar berikut yang merupakan kurva ROC untuk algoritma *Logistic Regression* dengan *Particle Swarm Optimization* sebagai seleksi fiturnya. Kurva ROC pada gambar tersebut mengekspresikan *Confusion Matrix* dari tabel 4. Garis horizontal adalah *false positives* dan garis vertikal *true positives*.

Nilai AUC (Area Under Curve) yang diperoleh adalah sebesar 0.99 .



Gambar 2. Kurva ROC Prediksi Menggunakan *Logistic Regression* dengan Pemilihan Fitur *Particle Swarm Optimization*

3.2. Pembahasan Hasil

Pada penelitian sebelumnya Muzakir Irfan [5], melakukan penelitian untuk memprediksi loyalitas pelanggan telekomunikasi dengan menggunakan algoritma backpropagation dengan pemilihan fitur *Particle Swarm Optimization* pada data set yang sama. Hasil dari penelitiannya mendapatkan nilai *Accuracy* 85,48 % dan nilai AUC 0,531. Sedangkan pengujian dengan menggunakan backpropagation berbasis *Particle Swarm Optimization* dilakukan seleksi atribut dan penyesuaian pada parameter didapatkan nilai *Accuracy* 86,05% dan nilai AUC adalah 0,637.

Oebidat R, et all [6], juga melakukan penelitian tentang prediksi loyalitas pelanggan telekomunikasi dengan data set yang sama menggunakan K-means Clustering digabung dengan *Genetic Programming*. Pertama K-means digunakan untuk menyaring dataset dari outlier dan *Genetic Programming* diterapkan untuk membangun pohon klasifikasi yang mampu mengklasifikasikan pelanggan, lalu dibandingkan dengan C4.5 dan ANN. Dari hasil akurasi didapatkan nilai akurasi K-means + GP 91.40% , GP 92.10%, C4.5 94.10%, dan ANN 94.80%.

Dari 20 variabel dilakukan seleksi atribut dengan algoritma *Particle Swarm Optimization* sehingga menghasilkan 12 atribut yang digunakan yaitu : State, Phone, Intl Plan, Vmail Message, Day Mins, Day Calls, Day Charger, Eve Calls, Eve Charger, Night Calls, Intl Charge, Custserv Calls nantinya akan digunakan sebagai atribut untuk memprediksi algoritma *Logistic Regression* karena nilai bobotnya lebih dari 0 atau mendekati 1. Sedangkan atribut yang nilain bobotnya 0.00 seperti Account Length, Area Code, Vmail Plan, Eve Mins, Night Mins, Night Charger, Intl Mins, Intl Calls tidak digunakan dalam prediksi. Bobot atribut yang paling tinggi yaitu State, Day Call, Day charger, Eve Call, Eve Charger, dan Intl Charger dengan bobot hasil sebesar 1.000.

Berikut ini adalah pembahasan terpilihnya 12 atribut dalam *Particle Swarm Optimization* tersebut di atas.

- a. State (Negara bagian Columbia): mungkin budaya dan kebiasaan masyarakat di Colombia yang menjadikan alasan kenapa banyak pelanggan yang loyal.

- b. Phone (nomor telepon): bisa disebabkan karena nomor telepon merupakan aspek utama terjadinya komunikasi.
- c. Intl Plan (perencanaan internasional): pendaftaran layanan international, bila sudah melakukan pendaftaran biasanya penggunaan layanan telekomunikasi masih akan berlanjut untuk menggunakan layanan yang sudah didaftarkan.
- d. Vmail Message (jumlah pesan suara): adanya layanan pesan suara yang sering digunakan, artinya pengguna merasa terbantu dengan layanan pesan suara.
- e. Day Mins (jumlah layanan permenit dalam sehari di siang hari): semakin sering menggunakan layanan berarti, pengguna nyaman dengan layanan yang diberikan.
- f. Day Calls (jumlah panggilan dalam sehari di siang hari): biasanya panggilan di siang hari ada hubungannya dengan pekerjaan, artinya bila layanan layanan sudah digunakan untuk bekerja maka jarang sekali pengguna meninggalkan layanan.
- g. Day Charger (jumlah biaya dalam sehari di siang hari): semakin banyak biaya yang dikeluarkan untuk penggunaan layanan telekomunikasi, artinya pengguna nyaman dengan layanan yang diberikan.
- h. Eve Calls (jumlah panggilan dalam sehari di sore hari): panggilan dalam sore hari biasanya digunakan pelanggan untuk mengisi waktu luang mereka, semakin sering menggunakan panggilan di sore hari, artinya pelanggan biasa menggunakan layanan telekomunikasi untuk mengisi waktu mereka di luar saat penting ketika bekerja.
- i. Eve Charger (jumlah biaya dalam sehari di sore hari): semakin banyak biaya yang dikeluarkan untuk menggunakan layanan telekomunikasi di sore hari, artinya pengguna terbiasa mengisi waktu mereka di luar saat penting ketika bekerja.
- j. Night Calls (jumlah panggilan dalam sehari di malam hari): biasanya panggilan di malam hari digunakan untuk menghubungi rekan, teman, atau saudara artinya bila layanan layanan sudah digunakan untuk berkomunikasi dengan teman, rekan, atau saudara maka jarang sekali pengguna meninggalkan layanan karena adanya keterikatan komunikasi.
- k. Intl Charge (jumlah biaya panggilan internasional): bila pengguna banyak mengeluarkan biaya untuk menggunakan panggilan internasional, biasanya pengguna nyaman dengan fasilitas panggilan internasional yang diberikan layanan telekomunikasi.
- l. Custserv Calls (jumlah panggilan ke layanan pelanggan): semakin sering pengguna menghubungi layanan pelanggan berarti semakin banyak hal yang ingin diketahui pelanggan tentang layanan yang bisa digunakan oleh pelanggan, kesediaan berkonsultasi bila mengalami masalah dalam komunikasi. Hal ini cenderung positif dibandingkan dengan pelanggan yang tidak mau berkonsultasi, namun tiba-tiba memutuskan diri untuk tidak berlangganan ke perusahaan yang bersangkutan.

Pada Penelitian ini pengujian *Accuracy* dan AUC menggunakan algoritma *Logistic Regression* dan pemilihan fitur *Particle Swarm Optimization* mendapatkan hasil *Accuracy* sebesar 97,65% dan dengan menggunakan kurva ROC mendapatkan nilai AUC sebesar 0.99. waktu yang digunakan untuk memproses algoritma *Logistic Regression* berbasis *Particle Swarm Optimization* membutuhkan waktu proses selama 2 jam 31 menit dengan 3333 record yang digunakan.

Dibandingkan dengan nilai *Accuracy* dari ANN+PSO sebesar 86.05%, K-means+GP sebesar 91.40% , GP sebesar 92.10%, C4.5 sebesar 94.10%, dan ANN sebesar 94.80%, maka penelitian ini memiliki hasil yang lebih akurat dengan menggunakan *Logistic Regression* dengan *Particle Swarm Optimization*, yakni didapatkan nilai *Accuracy* sebesar 97,65%.

4. KESIMPULAN

Dari hipotesis yang sudah dituangkan peneliti dalam kerangka pemikiran penelitiannya yang menyatakan bahwa penggunaan feature selection *Particle Swarm Optimization* sebelum dilakukannya prediksi, dan penggunaan *Logistic Regression* untuk prediksi loyalitas (churn) pelanggan telekomunikasi akan lebih akurat dibanding penelitian- penelitian sebelumnya, telah terbukti.

Penelitian- penelitian sebelumnya nilai *Accuracy* yang diperoleh dari model prediksi ANN+PSO

sebesar 86.05%, K-means+GP sebesar 91.40% , GP sebesar 92.10%, C4.5 sebesar 94.10%, dan ANN sebesar 94.80%. Penelitian ini memiliki hasil yang lebih akurat dengan menggunakan model prediksi *Logistic Regression* dengan *Particle Swarm Optimization* sebagai perangkat feature selection, yakni didapatkan nilai *Accuracy* sebesar 97,65%. Bila digunakan kurva ROC didapatkan nilai AUC sebesar 0.99.

DAFTAR PUSTAKA

- [1] G. Xia and W. Jin, "Model of Customer Churn Prediction on Support Vector Machine," *Systems Engineering - Theory & Practice*, vol. 28, no. 1, pp. 71–77, Jan. 2008.
- [2] C.-S. Lin, G.-H. Tzeng, and Y.-C. Chin, "Combined rough set theory and flow network graph to predict customer churn in credit card accounts," *Expert Systems with Applications*, vol. 38, no. 1, pp. 8–15, Jan. 2011.
- [3] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, "Credit card churn forecasting by *Logistic Regression* and decision tree," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15273–15285, Nov. 2011.
- [4] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1414–1425, Jan. 2012.
- [5] Muzakir Irvan, "Peningkatan Akurasi Algoritma Backpropagation dengan Seleksi Fitur *Particle Swarm Optimization* dalam Prediksi Pelanggan Telekomunikasi yang Hilang ," *Tesis Program pasca sarjana Magister teknik informatika Universitas Dian Nuswantoro 2013*.
- [6] Obiedat ruba, Alkasassbeh M, Faris H, Hrfoushi O, "Customer Churn Prediction Using a Hybrid Genetic Programming Approach", *Scientific Research Vol. 8(27)*, pp. 1289-1295, 18 july, 2013.
- [7] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," *European Journal of Operational Research*, vol. 218, no. 1, pp. 211–229, Apr. 2012.
- [8] B. Q. Huang, T.-M. Kechadi, B. Buckley, G. Kiernan, E. Keogh, and T. Rashid, "A new feature set with new window techniques for customer churn prediction in land-line telecommunications," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3657–3665, May 2010.
- [9] T.-S. Park, J.-H. Lee, and B. Choi, "Optimization for Artificial Neural Network with Adaptive inertial weight of *Particle Swarm Optimization*," *2009 8th IEEE International Conference on Cognitive Informatics*, pp. 481–485, Jun. 2009.
- [10] X. Shao, "Based on Two Swarm Optimized Algorithm of Neural Network to Prediction the Switch's Traffic of Coal," *2011 International Symposium on Computer Science and Society*, pp. 299–302, Jul. 2011.