IJAI (Indonesian Journal of Applied Informatics)
Vol. 9 No. 1 Tahun 2024 pISSN: 2548-3846, eISSN: 2598-5981

Pengaruh Metode Seleksi Fitur terhadap Akurasi Model SVM dalam Klasifikasi Customer Churn pada Perusahaan Telekomunikasi

Mayke Andani Rohmaniar [1]*, Roni Habibi[1], Syafrial Fachri[1]
[1]Teknik Informatika, Vokasi, Universitas Logistik dan Bisnis Internasional, Bandung, Jawa Barat
*Email: maykeandani5@gmail.com

Info Artikel

Kata Kunci :
customer churn; support vector machine; pemilihan fitur; correlation matrix; ANOVA; PCA; genetic algorithm

Keywords :
customer churn; support vector machine; feature selection; correlation matrix; ANOVA; PCA; genetic algorithm

Tanggal Artikel
Dikirim : 2 September 2024
Direvisi : 13 November 2024
Diterima : 17 November 2024

Abstrak

Penelitian ini menganalisis pengaruh metode seleksi fitur terhadap akurasi model *Support Vector Machine* dalam memprediksi pelanggan di industri telekomunikasi. Empat metode seleksi fitur (*Correlation Matrix*, PCA, dan GA) dan empat kernel (*Linear*, *Polynomial*, RBF, dan *Sigmoid*) dibandingkan menggunakan dataset pelanggan telekomunikasi dari Kaggle dengan 7043 entri dan 33 fitur. Metodologi CRISP-DM digunakan, meliputi Pemahaman Bisnis, Pemahaman Data, Persiapan Data, Pemodelan, Evaluasi, dan Implementasi. Hasil penelitian menunjukkan bahwa metode seleksi fitur menggunakan *Correlation Matrix* dengan kernel Linear memberikan kinerja terbaik. Model ini mencapai akurasi tertinggi sebesar 92,48%, dengan *precision* 0,93, recall 0,97, dan *f1-score* 0,95. Metode seleksi fitur lainnya, seperti PCA dan GA, memberikan hasil yang lebih rendah dibandingkan dengan *Correlation Matrix*. Implementasi model prediksi yang akurat diharapkan dapat membantu perusahaan telekomunikasi mengembangkan strategi retensi pelanggan yang lebih efektif.

Abstract

This study examines the impact of various feature selection methods on the accuracy of the Support Vector Machine (SVM) model in predicting customer behavior within the telecommunications sector. Specifically, the research compares four feature selection techniques: Correlation Matrix, Principal Component Analysis (PCA), and Genetic Algorithm (GA). Additionally, it evaluates the performance of four SVM kernels: Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid. Utilizing a telecom customer dataset from Kaggle, which comprises 7043 entries and 33 features, the study adheres to the CRISP-DM methodology. This methodology includes phases such as Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Implementation. The findings indicate that the Correlation Matrix feature selection method, when paired with the Linear kernel, provides the best performance. This particular configuration achieves the highest accuracy rate of 92.48%, along with a precision score of 0.93, a recall score of 0.97, and an F1-score of 0.95. In contrast, other feature selection methods, such as PCA and GA, result in lower performance metrics. These findings underscore the effectiveness of the Correlation Matrix and Linear kernel combination in enhancing the predictive accuracy of SVM models.

1. PENDAHULUAN

Industri telekomunikasi merupakan salah satu sektor yang sangat dinamis dan kompetitif[1]. Perusahaan dalam industri ini berusaha keras untuk mempertahankan pelanggan mereka di tengah persaingan yang ketat[2]. Salah satu tantangan terbesar yang dihadapi adalah *churn* pelanggan, yaitu ketika pelanggan memutuskan untuk berhenti menggunakan layanan suatu perusahaan dan beralih ke penyedia layanan lain[3][4]. Fenomena *churn* ini tidak hanya menyebabkan hilangnya pendapatan, tetapi juga meningkatkan biaya akuisi pelanggan baru[5][6].

Industri telekomunikasi Indonesia telah mengalami pertumbuhan pesat dalam beberapa tahun terakhir[7]. Ekspansi industri juga tercermin dalam studi perilaku pelanggan, yang menyoroti pentingnya harga yang kompetitif, kualitas layanan, dan atribut jaringan dalam mempertahankan pelanggan, menunjukkan lingkungan pasar yang matang dan kompetitif[8]. Namun, perubahan ini membawa masalah baru. Salah satu masalah tersebut adalah masuknya Starlink, layanan internet satelit global yang memberikan akses internet cepat dan stabil di banyak negara, termasuk Indonesia[9]. Kehadiran Starlink menarik pelanggan, terutama di daerah yang sebelumnya tidak terjangkau oleh layanan telekomunikasi konvensional[10]. Akibatnya, banyak pelanggan beralih ke Starlink, menyebabkan peningkatan kehilangan pelanggan[11].

Perusahaan telekomunikasi membutuhkan teknik yang efektif untuk memprediksi kehilangan pelanggan yang mungkin, sehingga mereka dapat mengambil tindakan pencegahan yang tepat[12][13]. Prediksi kehilangan pelanggan yang akurat memungkinkan perusahaan untuk menemukan pelanggan yang berisiko dan membuat rencana retensi yang sesuai, seperti penawaran khusus atau peningkatan layanan[14].

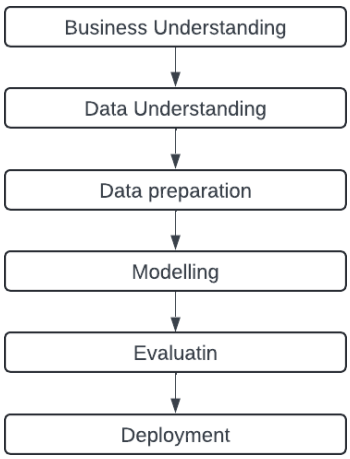
Teknik pembelajaran mesin telah banyak digunakan dalam prediksi *churn* pelanggan dalam beberapa tahun terakhir[15][16]. *Support Vector Machine* telah menunjukkan hasil yang menguntungkan dari berbagai metode saat ini[17]. SVM adalah algoritma klasifikasi yang kuat[18][19]. Dikenal karena kemampuan untuk menangani data dengan banyak fitur. Algoritma ini menghasilkan prediksi yang lebih akurat dengan menemukan *hyperplane* yang memisahkan kelas data dengan margin terbesar[20].

Penelitian ini bertujuan untuk menganalisis pengaruh metode seleksi fitur terhadap akurasi model SVM dalam klasifikasi *customer churn* pada perusahaan telekomunikasi. Empat metode seleksi fitur yang akan dibandingkan dalam penelitian ini adalah *Correlation Matrix (Pearson), Genetic Algorithm* (GA), ANOVA, dan PCA (*Principal Component Analysis*), [21][22][23][24]. Selain itu, empat kernel SVM yang akan dibandingkan adalah *linear*, *polynomial*, RBF (*Radial Basis Function*), dan *Sigmoid* serta dengan membandingkan 3 kinerja rasio yang berbeda diantaranya 80:20, 70:30, 90:10[25]. Dengan menggunakan metode seleksi fitur yang tepat dan kernel SVM yang optimal, diharapkan akurasi prediksi *churn* dapat ditingkatkan secara signifikan.

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam strategi manajemen pelanggan dan meningkatkan kinerja bisnis perusahaan telekomunikasi, khususnya dalam menghadapi tantangan baru yang dibawa oleh kehadiran Starlink di Indonesia.

2. METODE PENELITIAN

Metodologi penelitian ini dirancang untuk menganalisis pengaruh berbagai metode seleksi fitur terhadap akurasi *model Support Vector Machine* dalam memprediksi *churn* pelanggan di industri telekomunikasi. Penelitian ini menggunakan pendekatan *CRISP-DM (Cross-Industry Standard Process for Data Mining)* yang terdiri dari enam tahap utama: Pemahaman Bisnis, Pemahaman Data, Persiapan Data, Pemodelan, Evaluasi, dan *Deployment*[28]. Gambar 1 menunjukkan tahapan-tahapan tersebut secara visual untuk memberikan gambaran yang lebih jelas mengenai metodologi penelitian ini.



Gambar 1. *Research Methods*[26]

2.1 *Business Understanding*

Dalam kerangka penerapan metode *Crisp-DM* pada prediksi *customer churn* pada Perusahaan telekomunikasi dengan menggunakan algoritma classification, tahap Pemahaman Bisnis *(Business Understanding)* akan berfokus pada masalah yang ingin dipecahkan[26]. Dalam hal ini, tujuan bisnisnya adalah memprediksi dengan akurat *customer churn*.

2.2 *Data Understanding*

Setelah melewati tahap Pemahaman Bisnis, langkah berikutnya adalah memasuki tahap Pemahaman Data *(Data Understanding)* dalam konteks prediksi *customer churn*[26]. Pada tahap ini, data historis *customer churn* dikumpulkan dan dianalisis secara mendalam. Data ini mencakup variabel-variabel prediktif yang akan menjadi dasar dalam membangun model.

Data yang digunakan dalam analisis ini bersumber dari dataset *Kaggle* perusahaan telekomunikasi fiktif yang menyediakan telepon rumah dan layanan Internet kepada 7043 pelanggan di California. Pada data ini terdapat 33 fitur.

2.3 *Data Preparation*

Dalam tahap Data Preparasion, fokus utama adalah memastikan bahwa data yang akan digunakan untuk analisis telah disiapkan secara optimal agar sesuai untuk pelatihan model.[27] Proses persiapan data melibatkan serangkaian langkah kunci, di antaranya:

1) *Encoding data*: bertujuan untuk mengonversi data kategorikal menjadi format numerik yang dapat diproses oleh algoritma *machine learning*, sehingga meningkatkan kompatibilitas dan kinerja model[28].
2) Pembersihan Data: menemukan dan mengelola nilai yang hilang. Hal ini dapat melibatkan pengisian nilai yang hilang[29]

2.4 *Modelling*

Fase data modeling terdiri dari memilih teknik *modeling*, membangun *test case,* dan membuat model. Fase ini dimulai dengan perbandingan hasil pemilihan fitur yang selanjutnya diimplementasikan kepada pemodelan menggunakan algoritma *SVM*. Adapun penjelasannya sebagai berikut:

1) *Feature Selection*

Metode pemilihan fitur adalah teknik yang digunakan dalam *machine learning* untuk memilih subset fitur yang paling relevan dari kumpulan fitur asli untuk digunakan dalam model[30]. Pemilihan fitur bertujuan untuk meningkatkan performa model dengan mengurangi kompleksitasnya, menghilangkan fitur yang *redundant* atau tidak relevan, dan mengurangi *overfitting* [31].

Pada penelitian ini digunakan 4 Metode Pemilihan Fitur diantaranya adalah *correlation matrix, GA, ANOVA, dan PCA*. Adapun penjelasan masing masingnya sebagai berikut:

- *Correlation Matrix (Pearson):*
  Correlation Matrix (Pearson) mengukur hubungan *linear* antara fitur dalam dataset dengan koefisien korelasi berkisar antara -1 hingga 1. Metode ini mengidentifikasi fitur yang memiliki korelasi tinggi dengan variabel target dan mengeliminasi fitur yang redundant[32].
- *Genetic Algorithm (GA):*
  Genetic Algorithm adalah metode optimasi terinspirasi dari seleksi alam, menggunakan seleksi, crossover, dan mutasi untuk menemukan kombinasi fitur optimal dengan menjelajahi ruang solusi yang sangat besar[33].
- ANOVA *(Analysis of Variance):*
  ANOVA membandingkan rata-rata dari tiga kelompok atau lebih untuk menentukan apakah ada perbedaan signifikan di antara mereka. Dalam seleksi fitur, ANOVA mengukur hubungan antara fitur independen dan variabel dependen untuk mengidentifikasi fitur yang signifikan[22].

- PCA *(Principal Component Analysis):*
  PCA mereduksi dimensi data dengan mengubah data asli menjadi komponen utama, yang merupakan kombinasi linear dari variabel asli. Komponen utama ini disusun untuk menangkap variabilitas terbesar dalam data secara berurutan[34].

2) Rasio Pembagian Data

Rasio pembagian data adalah cara untuk membagi *dataset* menjadi dua bagian: set pelatihan (*training set*) dan set pengujian (*test set*)[35]. Dalam penelitian ini, digunakan tiga rasio pembagian data:

- Rasio 80:20: 80% data digunakan untuk melatih model, sedangkan 20% digunakan untuk menguji model.
- Rasio 70:30: 70% data digunakan untuk melatih model, sedangkan 30% digunakan untuk menguji model.
- Rasio 90:10: 90% data digunakan untuk melatih model, sedangkan 10% digunakan untuk menguji model.

3) SVM

Support Vector Machine adalah algoritma pembelajaran mesin yang kuat dan sering digunakan dalam tugas-tugas klasifikasi dan regresi[36]. Algoritma ini berfungsi dengan menemukan *hyperplane* optimal yang memisahkan data ke dalam kelas yang berbeda dengan margin maksimal[37]. SVM dapat bekerja dalam ruang fitur berdimensi tinggi dan sangat efektif dalam menangani data yang tidak dapat dipisahkan dengan penggunaan kernel *sigmoid, linear, polynomial, dan radial basis function* (RBF)[25]. Kernel pada *Support Vector Machine* adalah fungsi yang mengubah data menjadi ruang berdimensi lebih tinggi untuk memungkinkan pemisahan yang lebih baik antara kelas-kelas data[25].

Berikut adalah penjelasan masing-masing kernel yang umum digunakan:

- *Linear Kernel:*
  Kernel linear adalah fungsi kernel yang paling sederhana, digunakan ketika data dapat dipisahkan secara *linear*[26]. Adapun fungsinya sebagai berikut:

$$K(x, x') = sum\ (x.x')\ (1) \quad (1)$$

Rumus (1) menjelaskan K adalah fungsi kernel, x, x' adalah vektor input dari dua data yang berbeda, dan $sum\ (x.\ x')$adalah hasil pemjumlahan dari perkalian antar dua vektor input.

- *Polynomial Kernel:*
  Kernel polynomial memungkinkan pemisahan data dengan menggunakan *polynomial* dari derajat tertentu[26]. Adapun fungsinya sebagai berikut:

$$K(x, xi) = \left(sum\left(x.x'\right) + c\right)d \quad (2)$$

Rumus (2) menjelaskan K adalah fungsi kernel, x, x' adalah vektor input dari dua data yang berbeda, $sum\ (x.\ x')$adalah hasil penjumlahan dari perkalian antar dua vektor input, d adalah derajat *polynomial* dan c adalah konstanta yang dapat diatur.

- RBF *(Radial Basis Function) Kernel:*
  RBF *atau Gaussian kernel*, adalah salah satu kernel yang paling umum digunakan karena fleksibilitasnya dalam menangani data yang sangat kompleks dan *non-linear*. Adapun fungsinya sebagai berikut:

$$K(x, x') = exp(-\gamma||x - x'||^2) \quad (3)$$

Rumus (3) menjelaskan K adalah fungsi kernel, x, x' adalah vektor input dari dua data yang berbeda, exp adalah hasil pemjumlahan dari perkalian antar dua vektor input, γ\gamma adalah parameter yang mengontrol lebar *Gaussian* dan $|x - x'|^2$ adalah kuadrat dari jarak *euclidean* antara dua vektor input.

- *Sigmoid Kernel:*
  Kernel sigmoid berfungsi mirip dengan fungsi aktivasi dalam jaringan saraf dan cocok untuk data yang memiliki karakteristik tertentu[26]. Adafun fungsinya sebagai berikut:

$$K(x, x') = tanh(\sigma sum(x.x') + c) \quad (4)$$

Rumus (4) menjelaskan K adalah fungsi kernel, x, x' adalah vektor input dari dua data yang berbeda, $tanh$ adalah fungsi tangens hiperbolik, $sum(x.x')$adalah hasil penjumlahan dari perkalian dot product antara dua vektor input, α dan c adalah parameter yang dapat diatur.

2.5 *Evaluation*

Evaluasi dalam kode melibatkan pelatihan model menggunakan data pelatihan, prediksi pada data pengujian, dan penilaian kinerja model melalui metrik seperti *accuracy, precision, recall, dan f1-score* laporan klasifikasi untuk membandingkan kinerja model SVM dimasing masing kernel dan rasio. Metrik- metrik ini memiliki rentang nilai dari 0 hingga 1, dengan nilai yang mendekati 1 menunjukkan kinerja model yang semakin baik[40]. Berikut ini mengenai Metrik Evaluasi:

- *Accuracy:* proporsi dari jumlah prediksi benar terhadap total prediksi.

$$Accuracy = \frac{\sum_{i+1}^{n} 1(\widehat{(yi)}=yi)}{n} \quad (5)$$

Rumus (5) menjelaskan yi adalah nilai aktual pada waktu ke-i. $\hat{y}\imath$ adalah nilai prediksi pada waktu ke-i. n adalah jumlah pengamatan. 1 adalah fungsi indikator yang bernilai 1 jika argumennya benar, dan 0 jika salah.

- *Precision:* proporsi dari prediksi positif yang benar terhadap total prediksi positif.

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives(FP)} \quad (6)$$

Rumus (6) menjelaskan *TP (True Positive)* adalah jumlah prediksi positif yang benar. *FP (False Positive)* adalah jumlah prediksi positif yang salah.

- *Recall:* proporsi dari prediksi positif yang benar terhadap total kasus positif sebenarnya.

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negative(FN)} \quad (7)$$

Rumus (7) menjelaskan *TP (True Positive)* adalah jumlah prediksi positif yang benar. *FN (False Negative) a*dalah jumlah prediksi positif yang terlewat.

- *F1-Score:* rata-rata harmonik dari *precision* dan *recall*

$$F1\text{-}Score = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad (8)$$

Rumus (8) *Precision* adalah proporsi dari prediksi positif yang benar terhadap total prediksi positif. *Recall* adalah proporsi dari prediksi positif yang benar terhadap total kasus positif sebenarnya.

2.6 *Deployment*

Dalam tahap pengimplementasian, model prediksi *churn* pelanggan yang telah diverifikasi dan dinilai akan diimplementasikan dalam operasional bisnis. Model ini menjadi alat kunci dalam mengoptimalkan pengelolaan pelanggan dan membantu dalam mengidentifikasi pelanggan yang berpotensi *churn* dengan lebih akurat.

3. HASIL DAN PEMBAHASAN

Penelitian ini menganalisis pengaruh metode seleksi fitur terhadap akurasi model *Support Vector Machine* dalam memprediksi *churn* pelanggan dengan menggunakan empat kernel SVM *(Linear, Polynomial,* RBF*, dan Sigmoid)* dan empat metode seleksi fitur (*Correlation Matrix,* PCA*,* ANOVA*, dan Genetic Algorithm).* Hasil pemilihan fitur ditampilkan dalam tabel 1 dibawah ini:

Tabel 1. Hasil Pemiliihan Fitur

Metode	Correlation Matrix	*GA*	ANOVA	PCA
Fitur yang dipilih	City, Zip Code, Longitude,Senior Citizen, Phone Service, Multiple Lines, Paperless Billing, Payment Method, MonthlyCharges, Total Charges, Churn Value, Churn Score, Churn Reason	CustomerID, State, City, Zip Code, Longitude, Partner, Dependents, Phone Service, Multiple Lines, Internet Service, Online Security, Online Backup, Device Protection, Payment Method, Monthly Charges, Churn Label, Churn	Dependents, Tenure Months, Online Security, Online Backup, Tech Support, Contract, Monthly Charges, Churn Label, Churn Score, Churn Reason	CustomerID, Count, Country,State, City, Zip Code, Lat Long, Latitude, Longitude, Gender
Jumlah	12	18	10	10

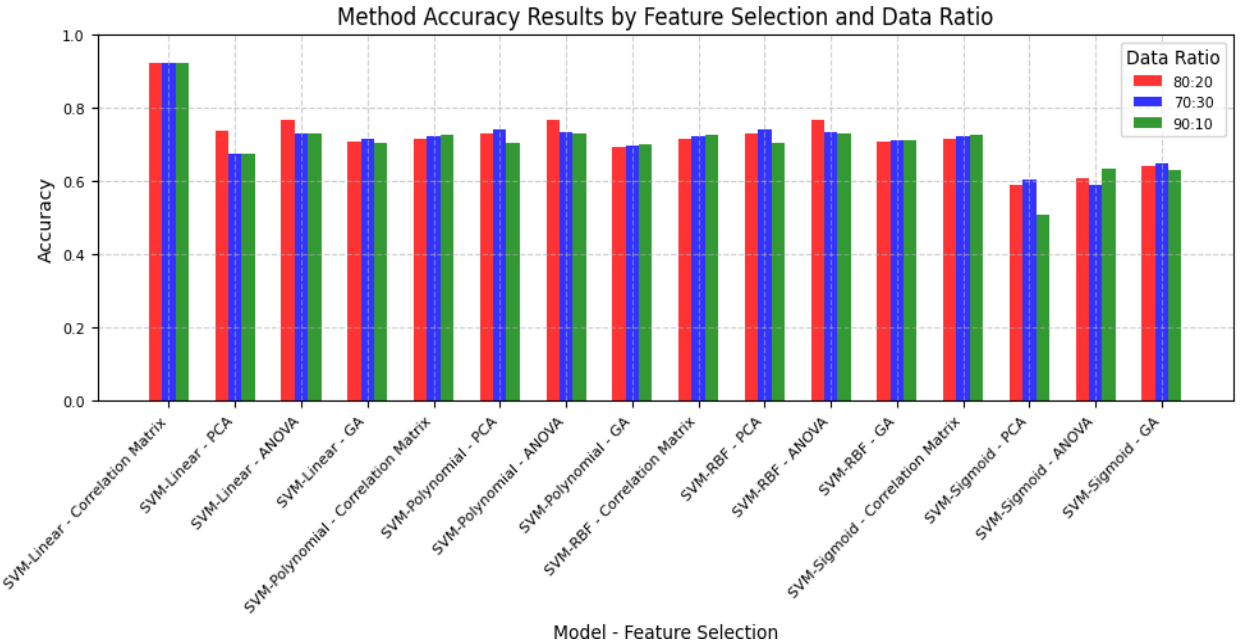
Setelah pemilihan fitur menggunakan *metode Correlation Matrix,* GA, ANOVA*, dan* PCA, pemodelan dilakukan dengan algoritma SVM menggunakan empat kernel: *linear, polynomial,* RBF*, dan sigmoid.* Evaluasi dilakukan menggunakan tiga rasio pembagian data: 80:20, 70:30, dan 90:10, untuk memastikan konsistensi kinerja model. Rasio ini membantu menentukan stabilitas model dalam berbagai skenario pembagian data. Kinerja model dievaluasi melalui metrik akurasi*, precision, recall,* dan *f1-score*, memberikan gambaran menyeluruh tentang seberapa baik model memprediksi *churn* pelanggan. Hasil dari berbagai kombinasi kernel dan rasio disajikan dalam Tabel 2, menunjukkan kinerja model pada masing-masing kombinasi, sehingga memudahkan pemilihan model yang paling optimal.

Tabel 2. *Comparison of evaluation matrix on classification methods.*

No	model	feature selection	Ratio											
			80:20				70:30				90:10			
			Acc	Prec	Rec	f1-S	Acc	Prec	Rec	f1-S	Acc	Prec	Rec	f1-S
1	svm-*linear*	Corellation matrix	0.9233	0.91	0.99	0.95	0.9233	0.92	0.98	0.95	0.9248	0.93	0.97	0.95
		GA	0.7088	0.71	1.00	0.83	0.7162	0.72	1.00	0.83	0.7062	0.71	1.00	0.83
		ANOVA	0.7659	0.77	1.00	0.87	0.7323	0.73	1.00	0.85	0.7323	0.73	1.00	0.85
		PCA	0.7375	0.75	0.96	0.84	0.6760	0.71	0.92	0.80	0.6760	0.71	0.92	0.80
2	svm-poly	Corellation matrix	0.7161	0.72	1.00	0.83	0.7217	0.72	0.82	0.84	0.7276	0.73	1.00	0.84
		GA	0.6938	0.71	0.97	0.82	0.6984	0.72	0.96	0.82	0.7023	0.71	0.98	0.82
		ANOVA	0.7102	0.71	1.00	0.83	0.7140	0.71	1.00	0.83	0.7159	0.72	1.00	0.83
		PCA	0.7121	0.72	1.00	0.84	0.7176	0.72	1.00	0.84	0.7031	0.70	1.00	0.83

		Corellation matrix	0.7161	0.72	1.00	0.83	0.7217	0.72	1.00	0.84	0.7276	0.73	1.00	0.84
3	svm-rbf	*GA*	0.7088	0.71	1.00	0.83	0.7232	0.99	0.83	0.83	0.7075	0.71	1.00	0.83
		ANOVA	0.7102	0.71	1.00	0.83	0.7140	0.71	1.00	0.83	0.7159	0.72	1.00	0.83
		PCA	0.7215	0.72	1.00	0.84	0.7176	0.72	1.00	0.84	0.7031	0.70	1.00	0.83
		Corellation matrix	0.7161	0.72	1.00	0.83	0.7217	0.72	1.00	0.84	0.7276	0.73	1.00	0.84
4	svm - sigmanoid	*GA*	0.6429	0.73	0.80	0.76	0.6488	0.73	0.80	0.77	0.6292	0.72	0.79	0.75
		ANOVA	0.6107	0.72	0.73	0.73	0.6160	0.73	0.74	0.73	0.6136	0.73	0.73	0.73
		PCA	0.6051	0.73	0.72	0.73	0.5945	0.72	0.71	0.72	0.5852	0.71	0.71	0.71

Penelitian ini menunjukkan bahwa metode seleksi fitur *correlation matrix* memberikan hasil terbaik dalam meningkatkan akurasi model SVM untuk prediksi *churn* pelanggan di industri telekomunikasi. Evaluasi dilakukan dengan menggunakan empat jenis kernel SVM *(linear, polynomial,* RBF*, dan sigmoid)* dan tiga rasio pembagian data (80:20, 70:30, dan 90:10). Hasil terbaik dicapai dengan kernel *linear*, terutama pada rasio 90:10, di mana model dengan *Correlation Matrix* mencapai akurasi tertinggi sebesar 92.48%, *precision* 0.93, *recall* 0.97, dan *f1-score* 0.95. Hal ini menunjukkan bahwa pemilihan fitur yang tepat dan penggunaan kernel *linear* secara konsisten meningkatkan kinerja model SVM dalam memprediksi *churn* pelanggan. Dengan demikian, *correlation matrix* dan kernel *linear* dapat dianggap sebagai kombinasi optimal untuk tugas ini, memberikan panduan penting bagi perusahaan telekomunikasi dalam strategi retensi pelanggan. Dapat kita lihat juga pada visualisasi *accuracy* pada gambar 2 berikut ini:



Gambar 2. *Visualization of experimental results*

Pada gambar 2 menunjukkan akurasi berbagai metode SVM dengan beberapa metode seleksi fitur dan rasio data pelatihan atau pengujian (80:20, 70:30, dan 90:10), di mana *Correlation Matrix* secara konsisten memberikan akurasi tertinggi pada semua rasio data untuk kernel *Linear* dengan hasil terbaik dicapai pada rasio 90:10 dengan akurasi 92,48%. *PCA* dan *GA* menunjukkan akurasi lebih rendah dibandingkan dengan *Correlation Matrix* pada kernel *Linear* dan *Polynomial,* sementara kernel *Sigmoid* memiliki performa terendah secara keseluruhan, terutama dengan PCA dan ANOVA. Secara umum, kernel *Linear* dengan *Correlation Matrix* menunjukkan performa terbaik, diikuti oleh kernel *Polynomial* dan RBF, yang menegaskan pentingnya pemilihan metode seleksi fitur yang tepat untuk meningkatkan akurasi prediksi *churn* pelanggan.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa metode seleksi fitur memiliki pengaruh signifikan terhadap akurasi model *Support Vector Machine* dalam memprediksi *churn* pelanggan pada industri telekomunikasi. Dari hasil evaluasi metriks pada tabel 2, ditemukan bahwa metode seleksi fitur *correlation matrix (Pearson)* secara konsisten memberikan hasil terbaik dibandingkan dengan metode seleksi fitur lainnya. Model *SVM* dengan kernel *linear* menunjukkan performa tertinggi, terutama pada rasio data 90:10, mencapai akurasi 0.92, *recall* 0.97, *precision* 0.93, dan *F1-Score* 0.95. Metode seleksi fitur *PCA* dan *GA* memberikan hasil yang lebih rendah, sedangkan kernel Sigmoid secara keseluruhan menunjukkan performa terendah. Hasil ini menegaskan bahwa pemilihan metode seleksi fitur yang tepat sangat penting untuk meningkatkan akurasi model prediksi *churn*.

UCAPAN TERIMA KASIH

Saya ingin mengucapkan terima kasih yang sebesar-besarnya kepada Program Studi Teknik Informatika, Fakultas Vokasi, Universitas Logistik dan Bisnis Internasional, Bandung yang telah memberikan ilmu dan pengalaman yang sangat berarti.

DAFTAR PUSTAKA

[1] M. E. Meena and J. Geng, "Dynamic Competition in Telecommunications: A Systematic Literature Review," *SAGE Open*, vol. 12, no. 2, 2022, doi: 10.1177/21582440221094609.

[2] Y. Khan, S. Shafiq, A. Naeem, S. Hussain, S. Ahmed, and N. Safwan, "Customers churn prediction using Artificial Neural Networks (ANN) in telecom industry," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, pp. 132–142, 2019, doi: 10.14569/ijacsa.2019.0100918.

[3] Q. Tang, G. Xia, and X. Zhang, "A hybrid classification model for churn prediction based on customer clustering," *J. Intell. Fuzzy Syst.*, vol. 39, no. 1, pp. 69–80, 2020, doi: 10.3233/JIFS-190677.

[4] R. Sharma, "Customer Churn Analysis in Telecom Industry using Logistics Regression in Machine Learning with Kaplan–Meier and Cox Proportional Hazards Model," *Interantional J. Sci. Res. Eng. Manag.*, vol. 08, no. 04, pp. 1–5, 2024, doi: 10.55041/ijsrem30745.

[5] O. J. Ogbonna, G. I. O. Aimufua, M. U. Abdullahi, and S. Abubakar, "Churn Prediction in Telecommunication Industry: A Comparative Analysis of Boosting Algorithms," *Dutse J. Pure Appl. Sci.*, vol. 10, no. 1b, pp. 331–349, 2024, doi: 10.4314/dujopas.v10i1b.33.

[6] S. R. K. S. P, "Customer Churn Prediction Using Ensemble Techniques on Telco Dataset," vol. 10, no. 11, pp. 376–383, 2023, doi: 10.53555/kuey.v30i6.6126.

[7] P. D. A. N. Pencegahan, R. Hartini, and F. Azzahra, "OLIGOPOLI DAN PERSEKONGKOLAN OLEH KPPU (STUDI KASUS PT . TELEKOMUMIKASI," vol. 5, no. 1, pp. 98–107, 2024.

[8] R. I. Sujono *et al.*, "Maintaining sustainable use of the Indonesian telecommunications provider," *J. Stud. Komun. (Indonesian J. Commun. Stud.*, vol. 8, no. 1, pp. 042–052, 2024, doi: 10.25139/jsk.v8i1.6246.

[9] R. Hooda, "Starlink: A Revolution in Global Satellite Internet Communication," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 11, pp. 2622–2628, 2023, doi: 10.22214/ijraset.2023.57105.

[10] C. I. Samuels, N. R. Syambas, Hendrawan, I. J. M. Edward, Iskandar, and W. Shalannanda, "Service level measurement based on Uptime data monitoring for rural internet access services in Indonesia," *Proceeding 2017 11th Int. Conf. Telecommun. Syst. Serv. Appl. TSSA 2017*, vol. 2019-Janua, pp. 1–5, 2019, doi: 10.1109/TSSA.2017.8272951.

[11] P. W. Nudan, P. Widodo, and M. Affifudin, "Navigating the Starlink Era of Personal Data Protection in Indonesia," vol. 3, no. 7, pp. 1447–1458, 2024.

[12] M. Z. Alotaibi and M. A. Haq, "Customer Churn Prediction for Telecommunication Companies using Machine Learning and Ensemble Methods," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 3, pp. 14572–14578, 2024, doi: 10.48084/etasr.7480.

[13] T. R. Noviandy, G. M. Idroes, I. Hardi, M. Afjal, and S. Ray, "A Model-Agnostic Interpretability Approach to Predicting Customer Churn in the Telecommunications Industry," *Infolitika J. Data Sci.*, vol. 2, no. 1, pp. 34–44, 2024, doi: 10.60084/ijds.v2i1.199.

[14] V. Chang, K. Hall, Q. Xu, F. Amao, M. Ganatra, and V. Benson, "Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models," *Algorithms*, vol. 17, no. 6, p. 231, 2024, doi: 10.3390/a17060231.

[15] V. Geetha, C. K. Gomathy, C. S. Ganesh, and S. Aravind, "The customer churn prediction using machine learning," *AIP Conf. Proc.*, vol. 3028, no. 1, pp. 614–619, 2024, doi: 10.1063/5.0212569.

[16] M. A. Al Rahib, N. Saha, R. Mia, and A. Sattar, "Customer data prediction and analysis in e-commerce using machine learning," *Bull. Electr. Eng. Informatics*, vol. 13, no. 4, pp. 2624–2633, 2024, doi: 10.11591/eei.v13i4.6420.

[17] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review," *Inf.*, vol. 15, no. 4, 2024, doi: 10.3390/info15040235.

[18] S. WANG, "Svm-Based Support Vector Type Recognition Machine for Smart Things in Soccer Training Motion Recognition," *Scalable Comput.*, vol. 25, no. 4, pp. 2519–2531, 2024, doi: 10.12694/scpe.v25i4.2923.

[19] A. Kar, N. Nath, U. Kemprai, and Aman, "Performance Analysis of Support Vector Machine (SVM) on Challenging Datasets for Forest Fire Detection," *Int. J. Commun. Netw. Syst. Sci.*, vol. 17, no. 02, pp. 11–29, 2024, doi: 10.4236/ijcns.2024.172002.

[20] C. Kaushik, A. D. McRae, M. A. Davenport, and V. Muthukumar, "New Equivalences Between Interpolation and SVMs: Kernels and Structured Features," pp. 1–22, 2023, [Online]. Available: http://arxiv.org/abs/2305.02304

[21] P. Chen, F. Li, and C. Wu, "Research on Intrusion Detection Method Based on Pearson Correlation Coefficient Feature Selection Algorithm," *J. Phys. Conf. Ser.*, vol. 1757, no. 1, 2021, doi: 10.1088/1742-6596/1757/1/012054.

[22] R. Babatunde, S. O. Abdulsalam, O. A. Abdulsalam, and M. O. Arowolo, "Classification of customer churn prediction model for telecommunication industry using analysis of variance," *IAES Int. J. Artif. Intell.*, vol. 12, no. 3, pp. 1323–1329, 2023, doi: 10.11591/ijai.v12.i3.pp1323-1329.

[23] F. Song, Z. Guo, and D. Mei, "Feature selection using principal component analysis," *Proc. - 2010 Int. Conf. Syst. Sci. Eng. Des. Manuf. Informatiz. ICSEM 2010*, vol. 1, pp. 27–30, 2019, doi: 10.1109/ICSEM.2010.14.

[24] L. Huang, X. Zhao, and K. Huang, "Globaltrack: A simple and strong baseline for long-term tracking," *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 11037–11044, 2020, doi: 10.1609/aaai.v34i07.6758.

[25] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.

[26] S. Amri, "Perbandingan kerangka model klasifikasi untuk pemilihan metode kontrasepsi dengan pendekatan CRIPS-DM," *Inf. Sci. Libr.*, vol. 1, no. 1, pp. 14–23, 2020.

[27] T. A. R. Akbar and C. Apriono, "Machine Learning Predictive Models Analysis on Telecommunications Service Churn Rate," *Green Intell. Syst. Appl.*, vol. 3, no. 1, pp. 22–34, 2023, doi: 10.53623/gisa.v3i1.249.

[28] K. N. R. Srinivas, K. S. S. Manikanta, T. Prem Jacob, G. Nagarajan, and A. Pravin, "Customer Stress Prediction in Telecom Industries Using Machine Learning," *Lect. Notes Electr. Eng.*, vol. 691, no. 4, pp. 491–498, 2021, doi: 10.1007/978-981-15-7511-2_48.

[29] A. M. Rahmani *et al.*, "Machine learning (MI) in medicine: Review, applications, and challenges," *Mathematics*, vol. 9, no. 22, pp. 1–52, 2021, doi: 10.3390/math9222970.

[30] A. Kumar, A. Kaur, P. Singh, M. Driss, and W. Boulila, "Efficient Multiclass Classification Using Feature Selection in High-Dimensional Datasets," *Electron.*, vol. 12, no. 10, 2023, doi: 10.3390/electronics12102290.

[31] X. Xu *et al.*, "Spectral preprocessing combined with feature selection improve model robustness for plastics samples classification by LIBS," *Front. Environ. Sci.*, vol. 11, no. May, pp. 1–13, 2023, doi: 10.3389/fenvs.2023.1175392.

[32] I. M. Nasir *et al.*, "Pearson correlation-based feature selection for document classification using balanced training," *Sensors (Switzerland)*, vol. 20, no. 23, pp. 1–18, 2020, doi: 10.3390/s20236793.

[33] C. L. Huang and C. J. Wang, "A GA-based feature selection and parameters optimizationfor support vector machines," *Expert Syst. Appl.*, vol. 31, no. 2, pp. 231–240, 2019, doi: 10.1016/j.eswa.2005.09.024.

[34] Y. Lu, I. Cohen, X. S. Zhou, and Q. Tian, "Feature selection using principal feature analysis," *Proc. ACM Int. Multimed. Conf. Exhib.*, pp. 301–304, 2007, doi: 10.1145/1291233.1291297.

[35] N. Nurzilla, "Prediksi Pertumbuhan Tumor Kanker Payudara Menggunakan Model Regresi Linear Berbasis Machine Learning," *J. Artif. Intell. Appl.*, vol. 1, no. 1, pp. 28–35, 2024.

[36] A. Febrisa Sidabutar, R. Habibi, and W. Isti Rahayu, "Perbandingan Metode Klasifikasi Untuk Pengelompokan Risiko Magang Mahasiswa," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 3, pp. 2071–2076, 2023, doi: 10.36040/jati.v7i3.7026.

[37] H. Shamsudin, U. K. Yusof, Y. Haijie, and I. S. Isa, "an Optimized Support Vector Machine With Genetic Algorithm for Imbalanced Data Classification," *J. Teknol.*, vol. 85, no. 4, pp. 67–74, 2023, doi: 10.11113/jurnalteknologi.v85.19695.