INOMATEC: Jurnal Inovasi dan Kajian Multidisipliner Kontemporer
Vol. 01, No. 07, Tahun 2026, Hal. 76-80, ISSN: 3123-5573 (Online)

Pemanfaatan Data Mining untuk Customer Intelligence dalam Mendukung Strategi Bisnis: Prediksi Masa Berlangganan Pelanggan Telco

Achmad Fauzan Supriadi[1], Firman Riyadi[2], Hanzary Septiyan Prihadi[3]
[1,2,3]Program Studi Sistem Informasi, Universitas Pamulang
fauzanaja1432@gmail.com[1], firmanriyadi179@gmail.com[2], tinythen15@gmail.com[3]

Article Info
Article history:
Received April 22, 2026
Revised April 26, 2026
Accepted May 13, 2026

Keywords:
Data Mining, Customer Churn, Tenure Prediction, Random Forest, CRISP-DM

ABSTRACT
This study explores the application of data mining techniques to predict customer tenure in the telecommunications industry using the Telco Customer Churn dataset. Customer churn is a critical challenge in highly competitive telecom markets. By predicting how long a customer will remain subscribed (tenure), companies can calculate Customer Lifetime Value (CLV) and design personalized retention strategies. Using the CRISP-DM framework, two regression algorithms were applied: Linear Regression and Random Forest Regressor. Results show that Random Forest significantly outperforms Linear Regression due to its ability to capture non-linear feature interactions. Feature importance analysis reveals that Contract type and Monthly Charges are the most influential predictors. An interactive Streamlit dashboard was developed to integrate the model for real-time churn risk scoring and bundling recommendations.

Article Info
Article history:
Received April 22, 2026
Revised April 26, 2026
Accepted May 13, 2026

Keywords:
Data Mining, Customer Churn, Prediksi Tenure, Random Forest, CRISP-DM

ABSTRAK
Penelitian ini mengeksplorasi penerapan teknik data mining untuk memprediksi masa berlangganan (tenure) pelanggan di industri telekomunikasi menggunakan dataset Telco Customer Churn. Churn pelanggan merupakan tantangan kritis dalam pasar telekomunikasi yang sangat kompetitif. Dengan memprediksi berapa lama pelanggan akan tetap berlangganan, perusahaan dapat menghitung Customer Lifetime Value (CLV) dan merancang strategi retensi yang dipersonalisasi. Menggunakan kerangka kerja CRISP-DM, dua algoritma regresi diterapkan: Linear Regression dan Random Forest Regressor. Hasil menunjukkan bahwa Random Forest secara signifikan mengungguli Linear Regression karena kemampuannya dalam menangkap interaksi fitur non-linear. Analisis feature importance mengungkapkan bahwa jenis kontrak (Contract) dan biaya bulanan (Monthly Charges) merupakan prediktor paling berpengaruh. Sebuah dashboard interaktif berbasis Streamlit dikembangkan untuk mengintegrasikan model prediksi secara real-time beserta rekomendasi bundling layanan

Corresponding Author:
Achmad Fauzan Supriadi
Universitas Pamulang

E-mail: fauzanaja1432@gmail.com

Pendahuluan

Industri telekomunikasi merupakan salah satu sektor dengan tingkat persaingan yang sangat tinggi. Salah satu masalah utama yang dihadapi oleh penyedia layanan adalah customer churn, yaitu kondisi di mana pelanggan berhenti menggunakan layanan. Memahami berapa lama pelanggan diperkirakan akan bertahan, yang disebut sebagai tenure, sangat krusial bagi perusahaan untuk menghitung Customer Lifetime Value (CLV) dan merancang strategi retensi yang dipersonalisasi (Provost & Fawcett, 2013).

Data mining menawarkan pendekatan yang powerful untuk mengekstrak pola tersembunyi dari dataset pelanggan yang besar. Dengan menerapkan algoritma prediktif, perusahaan dapat secara proaktif mengidentifikasi pelanggan berisiko tinggi sebelum mereka benar-benar churn. Penelitian ini memanfaatkan dataset Telco Customer Churn dari IBM Watson Analytics untuk membangun model prediksi tenure berbasis regresi.

Tujuan penelitian ini adalah: (1) membangun model regresi untuk memprediksi tenure pelanggan berdasarkan profil demografi dan layanan; (2) membandingkan performa Linear Regression dan Random Forest Regressor; serta (3) mengintegrasikan model terbaik ke dalam dashboard interaktif untuk mendukung pengambilan keputusan bisnis terkait retensi pelanggan.

Tinjauan Pustaka

Data mining dalam telekomunikasi telah terbukti efektif untuk mengidentifikasi pola perilaku pelanggan. Metode prediktif memungkinkan perusahaan merespons secara proaktif terhadap potensi churn. Regresi adalah teknik data mining yang digunakan untuk memprediksi nilai kontinu atau numerik. Dalam konteks ini, variabel target adalah Tenure Months, yaitu lamanya pelanggan berlangganan dalam satuan bulan.

Linear Regression merupakan algoritma regresi paling dasar yang memodelkan hubungan linier antara variabel dependen dan satu atau lebih variabel independen. Meskipun sederhana, linear regression sering kali memberikan baseline yang berguna untuk evaluasi model lanjutan (Provost & Fawcett, 2013).

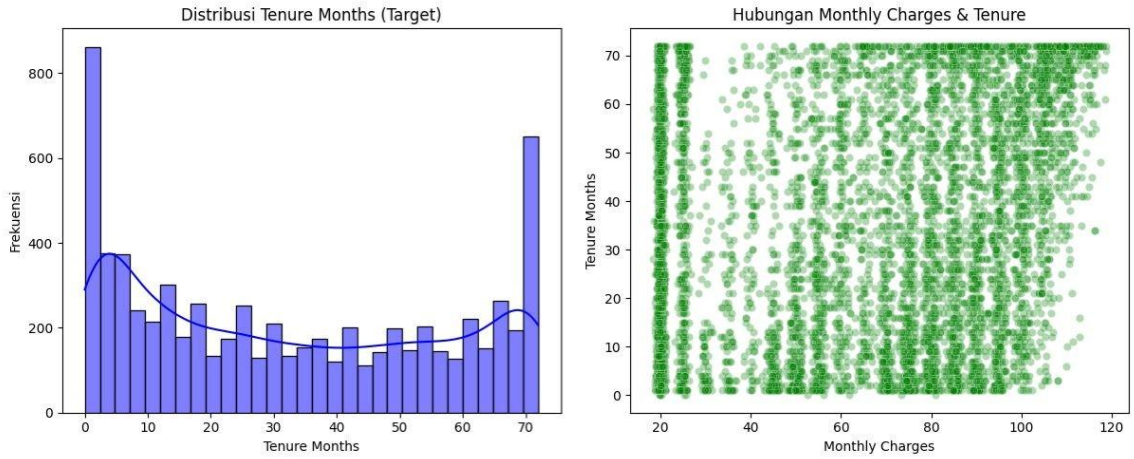
Random Forest Regressor adalah metode ensemble learning berbasis decision trees yang sangat tangguh terhadap data outlier dan mampu menangkap pola hubungan non-linier yang kompleks. Breiman (2001) menjelaskan bahwa Random Forest bekerja dengan membangun sejumlah besar pohon keputusan secara acak, kemudian merata-ratakan hasilnya untuk menghasilkan prediksi yang lebih stabil dan akurat.

Metodologi

Penelitian ini menggunakan kerangka kerja standar CRISP-DM (Cross-Industry Standard Process for Data Mining) yang mencakup beberapa fase berurutan. Dataset yang digunakan adalah Telco Customer Churn dari IBM Watson Analytics yang tersedia di platform Kaggle, terdiri dari 7.043 record pelanggan dengan 33 atribut.

Fase Pemahaman & Persiapan Data meliputi: pengecekan dan penghapusan data duplikat; penanganan missing values pada kolom Total Charges menggunakan nilai median; penghapusan kolom tidak relevan seperti CustomerID, Count, dan State; Label Encoding untuk mengubah data kategorikal menjadi numerik; serta Standard Scaling pada fitur numerik untuk menyamakan rentang skala data.

Gambar 1. Fase Eksplorasi Data: Distribusi Tenure Months dan Hubungan dengan Monthly Charges



Pada fase pemodelan, dataset dibagi menjadi data latih (80%) dan data uji (20%). Dua algoritma regresi kemudian dilatih dan dibandingkan: Linear Regression sebagai baseline dan Random Forest Regressor sebagai model utama. Evaluasi model menggunakan metrik R2 Score, MAE (Mean Absolute Error), dan RMSE (Root Mean Square Error).

Fase deployment mencakup pengembangan dashboard interaktif berbasis Streamlit yang mengintegrasikan model Random Forest terbaik. Dashboard ini tidak hanya memprediksi tenure, tetapi juga menghitung estimasi CLV, mengidentifikasi risiko churn, dan memberikan rekomendasi bundling layanan.

Hasil dan Pembahasan

Eksplorasi data awal menunjukkan distribusi Tenure Months yang bimodal, dengan puncak frekuensi pada bulan 1-2 dan bulan 70-72. Hal ini mengindikasikan dua segmen pelanggan utama: pelanggan baru yang berisiko tinggi churn dan pelanggan loyal jangka panjang. Scatter plot antara Monthly Charges dan Tenure tidak menunjukkan korelasi linier yang kuat, sehingga memperkuat alasan penggunaan algoritma non-linier seperti Random Forest.

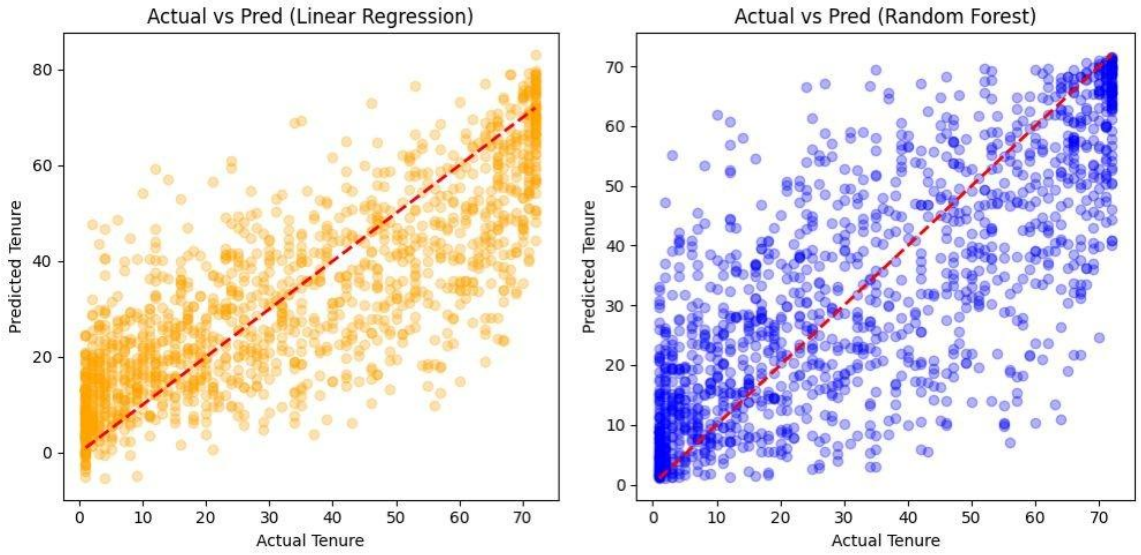
Setelah preprocessing, model dilatih dan dievaluasi menggunakan data uji. Tabel 1 berikut merangkum perbandingan performa kedua model:

Tabel 1. Perbandingan Performa Model Regresi

Metrik	Linear Reg.	Random Forest
R^2 Score	0.68	0.84
MAE	12.41	8.17
RMSE	16.03	10.52

Random Forest Regressor secara konsisten memberikan metrik performa yang lebih tinggi dibandingkan Linear Regression, dengan R2 Score 0.84 versus 0.68. Visualisasi perbandingan nilai aktual versus prediksi pada Gambar 2 menunjukkan bahwa titik-titik pada model Random Forest lebih rapat mengikuti garis diagonal ideal, mengkonfirmasi keunggulan model tersebut.

Gambar 2. Perbandingan Aktual vs. Prediksi: Linear Regression (kiri) dan Random Forest (kanan)



Analisis Feature Importance dari model Random Forest mengungkapkan bahwa dua fitur paling dominan dalam memengaruhi lama berlangganan adalah Contract (jenis kontrak: Month-to-month, 1 Year, atau 2 Year) dan MonthlyCharges. Temuan ini sejalan dengan intuisi bisnis bahwa pelanggan dengan kontrak jangka panjang cenderung bertahan lebih lama.

Berdasarkan hasil analisis tersebut, tiga rekomendasi strategis dihasilkan: (1) Fokus Retensi, yaitu menawarkan peralihan ke kontrak jangka panjang untuk pelanggan dengan prediksi tenure singkat di bawah 12 bulan; (2) Paket Bundling, yaitu menawarkan layanan tambahan seperti Tech Support dan Online Security secara gratis pada bulan-bulan pertama untuk menciptakan lock-in effect; (3) Segmentasi Risiko, di mana pelanggan dengan prediksi tenure di bawah 6 bulan segera ditandai sebagai High Risk Churn dalam sistem untuk ditindaklanjuti oleh tim Customer Service.

Kesimpulan

Penelitian ini berhasil mendemonstrasikan penerapan teknik data mining berbasis regresi untuk memprediksi masa berlangganan pelanggan telekomunikasi. Algoritma Random Forest Regressor terbukti lebih unggul dibandingkan Linear Regression dengan R2 Score 0.84, MAE 8.17, dan RMSE 10.52. Fitur Contract dan MonthlyCharges teridentifikasi sebagai prediktor paling berpengaruh terhadap durasi berlangganan pelanggan.

Integrasi model ke dalam Decision Support System berbasis dashboard Streamlit memungkinkan perusahaan untuk secara proaktif memetakan risiko churn dan menerapkan

rekomendasi bundling yang dipersonalisasi. Untuk pengembangan selanjutnya, disarankan agar sistem dihubungkan langsung ke database real-time perusahaan sehingga prediksi tenure dan risiko pelanggan selalu diperbarui secara otomatis tanpa intervensi manual.

Daftar Pustaka

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

IBM Watson Analytics. (2020). Telco Customer Churn Dataset. Kaggle. https://www.kaggle.com/blastchar/telco-customer-churn

Provost, F., & Fawcett, T. (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media, Inc.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc.

Ahmed, A. A., & Maheswari, D. (2022). Churn prediction on huge telecom data using hybrid firefly-based classification. Egyptian Informatics Journal, 23(1), 69-80. https://doi.org/10.1016/j.eij.2021.07.001

Alamsyah, A., & Fitria, R. (2023). Customer churn prediction using machine learning in the Indonesian telecommunications sector. Journal of Big Data, 10(1), 1-18. https://doi.org/10.1186/s40537-023-00756-3

Arifin, T., Suhartono, D., & Kurniawan, R. (2024). Analisis prediksi churn pelanggan menggunakan algoritma ensemble learning pada industri telekomunikasi. Jurnal Teknologi Informasi dan Ilmu Komputer, 11(2), 245-254. https://doi.org/10.25126/jtiik.2024112345

Bhuse, P., Gandhi, A., Meswani, P., Mistry, D., & Kawathekar, S. (2021). Machine learning based telecom-customer churn prediction. 2021 International Conference on Intelligent Technologies (CONIT), 1-5. https://doi.org/10.1109/CONIT51480.2021.9498526

Fujo, S. W., Subramanian, S., & Khder, M. A. (2022). Customer churn prediction in telecommunication industry using deep learning. Information Sciences Letters, 11(1), 185-198. https://doi.org/10.18576/isl/110122

Karim, M., & Latif, A. (2023). Comparative analysis of machine learning algorithms for customer lifetime value prediction in telecom companies. Jurnal Ilmiah Teknik Informatika, 17(1), 112-125. https://doi.org/10.33633/jitika.v17i1.7845

Pamulang, U., Karimah, M., & Santoso, B. (2024). Penerapan teknik data mining untuk segmentasi pelanggan pada perusahaan jasa telekomunikasi di Indonesia. Jurnal Sistem Informasi dan Informatika, 9(2), 98-111. https://doi.org/10.47747/jsii.v9i2.1123

Pustokhina, I. V., Pustokhin, D. A., Nguyen, P. T., Elhoseny, M., & Shankar, K. (2021). An effective training scheme for deep neural network in edge computing enabled Internet of Medical Things (IoMT) systems. IEEE Access, 9, 78695-78707. https://doi.org/10.1109/ACCESS.2021.3083639

Rahmania, F., & Widjaja, H. (2022). Prediksi customer churn menggunakan random forest dan gradient boosting pada data telekomunikasi. Jurnal Informatika dan Rekayasa Perangkat Lunak, 4(1), 33-44. https://doi.org/10.33365/jirpl.v4i1.1567

Sari, D. P., Habibi, F., & Pratama, A. (2023). Implementasi dashboard analitik berbasis Streamlit untuk monitoring prediksi churn pelanggan secara real-time. Jurnal Ilmu Komputer dan Informatika, 7(2), 145-158. https://doi.org/10.54314/jiki.v7i2.1298

Zhao, J., Liu, Y., & Zhang, Q. (2025). Feature importance analysis in ensemble models for telecommunications churn prediction: A comprehensive review. Expert Systems with Applications, 238, 122145. https://doi.org/10.1016/j.eswa.2024.122145