Performance Analysis of Machine Learning Algorithms in Identifying the Risk of Readmission for Inpatients at the Langsa Regional General Hospital

Fachrurazy, Rian Farta Wijaya

Abstract
This paper presents a preliminary descriptive analysis and an evaluation design for Random Forest and Extreme Gradient Boosting (XGBoost) to identify unplanned inpatient readmission risk at RSUD Langsa. Aggregate counts were extracted from the research proposal for index discharges between July 16, 2024 and June 16, 2026. The screened cohort comprised 34,122 episodes from 25,821 patients, including 2,870 unplanned readmissions occurring more than 24 hours and up to 30 days after discharge. The readmission proportion was 8.41%, compared with 13.63% among chronic-disease episodes and 6.94% among nonchronic episodes. Patients aged 60–74 years had a proportion of 13.78%. An analytical all-negative classifier would achieve 91.59% accuracy while detecting no readmissions. This supports evaluating precision–recall performance, recall, F2-score, and calibration rather than accuracy alone. The planned experiment uses patient-level splitting, nested cross-validation, and comparisons of unadjusted training, SMOTE-NC, and class weighting. As trained-model outputs and individual predictions are unavailable in the proposal, empirical superiority of either algorithm cannot yet be established. This stage contributes a descriptive outcome profile and an evaluation framework that accounts for class imbalance and predictor availability at discharge.
***Keywords**: Readmission; Random Forest; XGBoost; Class Imbalance; RSUD Langsa.*

Fachrurazy
Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
e-mail: oxxy.keren@gmail.com

Rian Farta Wijaya
e-mail: rianfartawijaya@dosen.pancabudi.ac.id
3rd International Conference Epicentrum of Economic Global Framework (ICEEGLOF)
Theme: Cross-Sector and Cross-Disciplinary Strategies for Achieving the SDGS
https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF

Introduction

Unplanned readmission warrants attention because it creates a renewed need for inpatient care after discharge. Risk prediction can help identify patients who require a review of discharge plans or more intensive follow-up. Model quality, however, depends on the outcome definition, completeness of service histories, and evaluation procedures. Reviews of readmission models document methodological variation and limitations in predictive performance across studies [1], [2].

Fachrurazy's research proposal reports preliminary observations involving 41,642 index episodes from 29,595 patients at Langsa Regional General Hospital (RSUD Langsa). After sequential screening, the main cohort comprised 34,122 episodes from 25,821 patients [3]. Unplanned readmission occurred in 2,870 episodes. These aggregate counts form the basis of the descriptive analysis in this paper; individual hospital information system records were unavailable for reanalysis.

Random Forest combines decision trees using bootstrap sampling and random feature selection [4]. XGBoost builds tree models sequentially through optimization and regularization [5]. Both approaches can model nonlinear relationships, but this capability does not guarantee superior performance in every hospital. Sharma and colleagues demonstrate the need to evaluate machine learning against conventional comparators in the readmission setting [6].

An Indonesian study of patients with heart failure provides context for applying machine learning to readmission outcomes [7]. Findings from a disease-specific cohort, however, do not directly apply to populations spanning multiple diagnoses. This paper aims to analyze the aggregate profile of RSUD Langsa, explain the implications of class imbalance, and define a comparison of Random Forest and XGBoost that can proceed when individual records become available. The results distinguish completed descriptive calculations from the planned model evaluation.

Literature Review

A readmission definition must specify the observation window, type of readmission, and facilities covered. The algorithm developed by Horwitz and colleagues provides a basis for identifying planned readmissions in claims data [8]. Its adaptation to RSUD Langsa requires mapping local diagnoses and procedures, particularly scheduled services. The LACE index is an established comparator for predicting early death or unplanned readmission [9]. Since its original target is a composite outcome, any application to readmission alone requires explicit explanation.

Five publications inform the data-processing and algorithm-evaluation context. Aulia and colleagues used K-Means to group medical records by disease type [10]. Hamzah, Sitorus, and Khairul compared C4.5 and Naïve Bayes for diabetes identification [11]. Barutu and colleagues compared C4.5 and Random Forest for diarrheal disease prediction [12]. These studies provide health analytics context, but none evaluates readmission outcomes.

Hermawan, Iqbal, and Sarif compared Random Forest, XGBoost, and an ensemble using graduate tracer-study data [13]. Ernawati and colleagues evaluated Random Forest and K-Nearest Neighbors for poverty classification [14]. These two studies inform methodological choices and metric reporting rather than clinical evidence. Their numerical findings are not used to estimate model performance at RSUD Langsa.

Research Methodology

The study follows five stages: problem formulation, methodological design, data preparation, comparative evaluation, and reporting [3]. Figure 1 summarizes the activities and outputs. Figure 1 details the proposed analytical workflow and shows how the test set remains separate from model development. At this stage, the available evidence supports aggregate descriptive analysis; individual-record modeling remains pending.



Figure 1. Research Stages

The research consists of five stages: problem formulation, methodological design, data preparation, comparative evaluation, and reporting. These stages define the research objectives, establish the evaluation protocol, prepare patient data, compare Random Forest and XGBoost, and communicate the findings. Model training and performance testing remain planned.

1. **Planned Preprocessing and Data Partitioning**

Individual records will be split 80:20 at the patient level, stratified by whether a patient has experienced readmission. All episodes belonging to a patient must remain in the same partition. The 80:20 ratio refers to patient allocation, so actual episode counts need not follow that ratio exactly. Pseudonymized patient identifiers will serve as grouping keys and will not enter the models as predictors.

Imputation, rare-category grouping, categorical encoding, and resampling will be fitted only on the training portion of each fold. Numerical variables will use median imputation, while categorical variables will use mode imputation. SMOTE-NC will precede one-hot encoding to preserve valid categorical values [15], [16]. Leakage prevention follows the principle of separating model training from evaluation [17].

2. **Planned Experiments and Performance Measures**

Random Forest and XGBoost will use the same features and partitions, with logistic regression as a comparator. The proposal specifies nested cross-validation grouped by patient, using 10 outer folds and 5 inner folds. Random search will examine 60 configurations for each tree-based model, with PR-AUC as the selection criterion [18], [19]. Training without imbalance correction, SMOTE-NC, and class weighting will be evaluated separately. The test set will retain its original class distribution.

The operational model is distinguished from the post-coding model. The complete LACE index will serve as a comparator only when all its components are available. A score that omits comorbidity must be described as an adapted LACE score rather than the standard index. Comparisons must acknowledge the different information available to each version [3], [9].

Results

1. Cohort Screening and Outcome Distribution

Table 1. Sequential Screening of Episodes at RSUD Langsa

Stage	Excluded	Episodes remaining	Unique patients
Initial index episodes	-	41,642	29,595
Transfer out or discharge against medical advice	1,997	39,645	28,534
Scheduled procedures and thalassemia	2,516	37,129	28,033
Neonates aged ≤28 days	943	36,186	27,114
Discharge in November–December 2025	2,063	34,123	25,822
No diagnosis	1	34,122	25,821

A total of 7,520 episodes were excluded. Unique patients were recounted at each stage; therefore, changes in patient counts cannot be equated with excluded episodes. The final cohort contains 2,870 episodes labeled as readmission and 31,252 episodes without that label. The recalculated positive-outcome proportion is 2,870/34,122 × 100% = 8.41%.

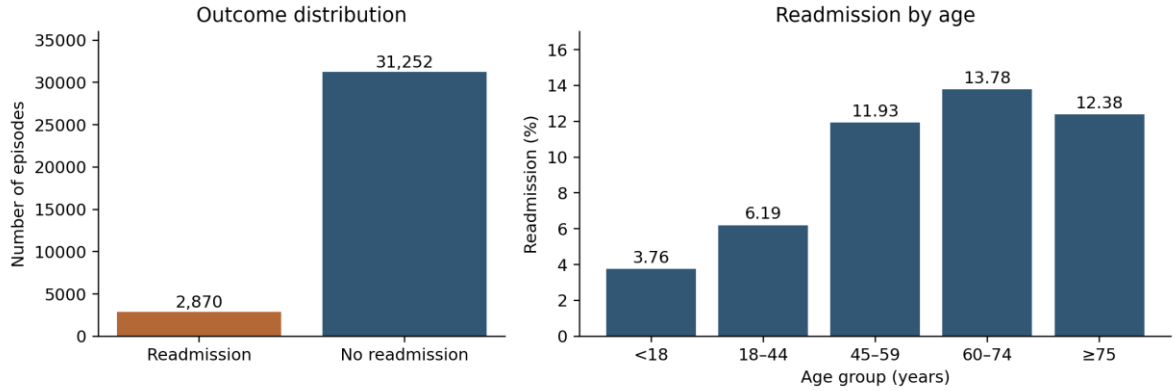


Figure 2. Outcome Distribution and Readmission Proportions

The figure shows that 2,870 of 34,122 inpatient episodes (8.41%) involved unplanned readmission, while 31,252 (91.59%) did not. This class imbalance emphasizes the need to assess recall and precision–recall performance alongside accuracy, ensuring that models identify readmission cases rather than mainly predicting the majority class.

Readmission proportions generally increased with age, from 3.76% among patients younger than 18 years to 6.19% at ages 18–44 and 11.93% at ages 45–59. The highest proportion occurred among patients aged 60–74 years (13.78%), followed by a slight decline among those aged 75 years or older (12.38%). This pattern supports examining age as a candidate predictor. However, these descriptive findings do not establish an independent effect of age because diagnoses, repeated episodes, and service characteristics have not been adjusted for.

2. Readmission by Chronic Disease Status

Table 2. Readmission Proportions by Age Group and Chronic Disease Status

Group	Episodes	Readmissions	Proportion (%)
Age <18 years	7,322	275	3.76
Age 18–44 years	12,418	769	6.19
Age 45–59 years	7,629	910	11.93
Age 60–74 years	5,727	789	13.78
Age ≥75 years	1,026	127	12.38
Chronic disease	7,521	1,025	13.63
Nonchronic	26,601	1,845	6.94

The table shows variation in readmission proportions by age and chronic disease status. Patients aged 60–74 years had the highest proportion (13.78%), while those younger than 18 years had the lowest (3.76%). Episodes involving chronic disease had a higher readmission proportion (13.63%) than nonchronic episodes (6.94%), approximately twice as high. These descriptive differences support examining both variables in predictive modeling, although their independent contributions require further analysis.

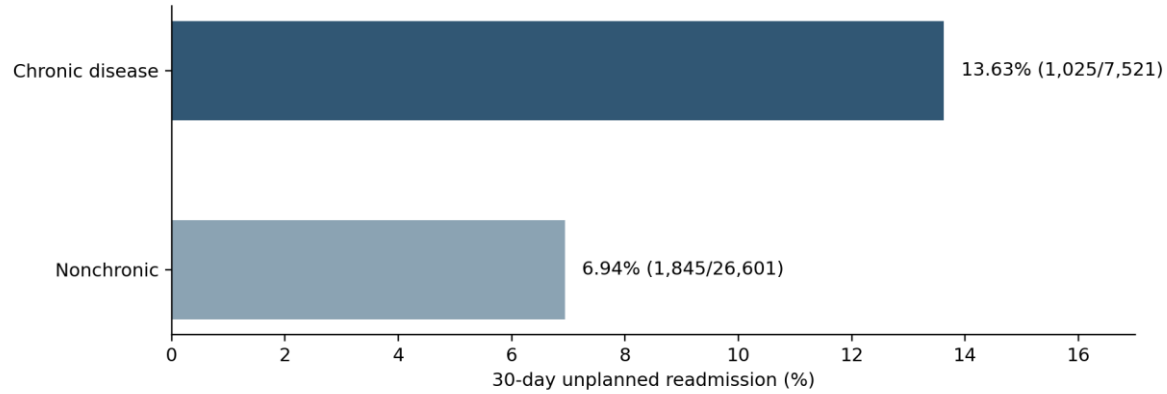


Figure 3. Readmission Proportions Among Chronic And Nonchronic Episodes

The readmission proportion was 1,025/7,521, or 13.63%, among chronic disease episodes and 1,845/26,601, or 6.94%, among nonchronic episodes. The difference was 6.69 percentage points, with a crude proportion ratio of approximately 1.96. This episode-level ratio does not account for dependence between episodes from the same patient. Accordingly, the paper does not report p-values or confidence intervals that assume independent episodes.

Discussion

The class distribution has direct implications for evaluation. A model optimized only for accuracy may overlook most readmissions. PR-AUC, recall, F2-score, calibration, and the number of patients flagged must be assessed to determine whether the model provides useful additional information. Comparison with logistic regression will test whether the complexity of tree-based algorithms yields a meaningful improvement on local data.

Differences between chronic and nonchronic episodes indicate outcome heterogeneity that warrants subgroup evaluation. However, the proposal assigns final diagnoses to the post-coding feature set. These descriptive findings do not justify including final diagnoses in an operational model without evidence that the information is available at prediction time. Maintaining this distinction aligns evaluation with intended use at discharge.

Observation within a single hospital limits the capture of readmissions across facilities. In addition, the youngest age category includes children beyond the neonatal period. LACE and

adult prediction models must not be assumed to be validated across the entire age range. Pediatric and adult performance should be evaluated separately or reported as subgroup analyses before clinical implementation.

Conclusion

The proposal cohort includes 34,122 episodes with 2,870 unplanned readmissions (8.41%). Readmission was more frequent among chronic disease episodes than nonchronic episodes. An all-negative comparator achieves 91.59% accuracy with zero recall. Evaluation must address positive outcomes, calibration, patient grouping, and predictor availability. The relative performance of Random Forest and XGBoost remains unknown. Experiments using individual records must be completed before reporting algorithm performance metrics and plots.

References

[1] D. Kansagara et al., "Risk prediction models for hospital readmission: A systematic review," JAMA, vol. 306, no. 15, pp. 1688–1698, 2011, doi: 10.1001/jama.2011.1515.

[2] A. Artetxe, A. Beristain, and M. Graña, "Predictive models for hospital readmission risk: A systematic review of methods," Computer Methods and Programs in Biomedicine, vol. 164, pp. 49–64, 2018, doi: 10.1016/j.cmpb.2018.06.006.

[3] Fachrurazy, "Komparasi algoritma Random Forest dan XGBoost dalam prediksi risiko readmisi pasien rawat inap di Rumah Sakit Umum Daerah Langsa," thesis research proposal, Master of Information Technology Program, Universitas Pembangunan Panca Budi, Medan, 2026. Author-supplied document, Chapters I–III (in Indonesian).

[4] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[6] Y. Sharma, C. Thompson, A. A. Mangoni, R. Shahi, C. Horwood, and R. Woodman, "Performance of machine learning models in predicting 30-day general medicine readmissions compared to traditional approaches in Australian hospital setting," Healthcare, vol. 13, no. 11, art. 1223, 2025, doi: 10.3390/healthcare13111223.

[7] F. E. Indriany et al., "Predicting the risk of severity and readmission in patients with heart failure in Indonesia: A machine learning approach," Healthcare Informatics Research, vol. 30, no. 3, pp. 253–265, 2024, doi: 10.4258/hir.2024.30.3.253.

[8] L. I. Horwitz et al., "Development and validation of an algorithm to identify planned readmissions from claims data," Journal of Hospital Medicine, vol. 10, no. 10, pp. 670–677, 2015, doi: 10.1002/jhm.2416.

[9] C. van Walraven et al., "Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community," CMAJ, vol. 182, no. 6, pp. 551–557, 2010, doi: 10.1503/cmaj.091117.

[10] W. Aulia, A. P. U. Siahaan, L. Marlina, Khairul, and M. Iqbal, "K-Means clustering algorithm analysis for grouping patient medical record data based on disease type," Jurnal Info Sains: Informatika dan Sains, vol. 14, no. 4, pp. 832–843, 2024. https://ejournal.seaninstitute.or.id/index.php/InfoSains/article/view/6055

[11] I. Hamzah, Z. Sitorus, and Khairul, "Analisa classification decision tree C45 dan Naïve Bayes pada indikasi penyakit diabetes menggunakan Rapid Miner," Jurnal Nasional Teknologi Komputer, vol. 4, no. 1, pp. 25–33, 2024, doi: 10.61306/jnastek.v4i1.126.

[12] S. Barutu, M. Iqbal, Khairul, and D. Nasution, "Comparative analysis of the C4.5 and Random Forest algorithms for the prediction of diarrheal disease," Jurnal Ilmiah Multidisiplin Indonesia (JIM-ID), vol. 4, no. 7, pp. 625–635, 2025. https://ejournal.seaninstitute.or.id/index.php/esaprom/article/view/7168

[13] R. Hermawan, M. Iqbal, and M. I. Sarif, "Analisis prediksi kesesuaian lulusan dengan dunia kerja menggunakan model ensemble Random Forest dan XGBoost berdasarkan data tracer study Universitas Royal," JUKTISI, vol. 5, no. 2, pp. 1513–1522, 2026, doi: 10.62712/juktisi.v5i2.1398.

[14] A. Ernawati, Khairul, Z. Sitorus, M. Iqbal, and D. Nasution, "Penerapan data mining untuk klasifikasi penduduk miskin di Kabupaten Labuhanbatu menggunakan Random Forest dan K-Nearest Neighbors," Bulletin of Information Technology (BIT), vol. 6, no. 2, pp. 23–35, 2025. https://journal.fkpt.org/index.php/BIT/article/view/1783

[15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[16] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning," Journal of Machine Learning Research, vol. 18, no. 17, pp. 1–5, 2017. https://jmlr.org/papers/v18/16-365.html

[17] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," ACM Transactions on Knowledge Discovery from Data, vol. 6, no. 4, art. 15, 2012, doi: 10.1145/2382577.2382579.

[18] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," BMC Bioinformatics, vol. 7, art. 91, 2006, doi: 10.1186/1471-2105-7-91.

[19] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," Journal of Machine Learning Research, vol. 13, pp. 281–305, 2012. https://jmlr.org/papers/v13/bergstra12a.html

[20] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," PLOS ONE, vol. 10, no. 3, e0118432, 2015, doi: 10.1371/journal.pone.0118432.

[21] R. van den Goorbergh, M. van Smeden, D. Timmerman, and B. Van Calster, "The harm of class imbalance corrections for risk prediction models: Illustration and simulation using logistic regression," Journal of the American Medical Informatics Association, vol. 29, no. 9, pp. 1525–1534, 2022, doi: 10.1093/jamia/ocac093.

[22] A. J. Vickers and E. B. Elkin, "Decision curve analysis: A novel method for evaluating prediction models," Medical Decision Making, vol. 26, no. 6, pp. 565–574, 2006, doi: 10.1177/0272989X06295361.

[23] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems 30, 2017, pp. 4765–4774. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[24] G. S. Collins et al., "TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods," BMJ, vol. 385, e078378, 2024, doi: 10.1136/bmj-2023-078378.