

cek-01a0fff1-7dc6-7d80-a387-20512ba97363

ORIGINALITY REPORT

12%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.mdpi.com Internet	66 words — 2%
2	iiardjournals.org Internet	54 words — 2%
3	journalbiomed.com Internet	32 words — 1%
4	www.frontiersin.org Internet	20 words — 1%
5	indiafreenotes.com Internet	15 words — 1%
6	f1000research-files.f1000.com Internet	14 words — < 1%
7	jutif.if.unsoed.ac.id Internet	13 words — < 1%
8	www.transport.ntua.gr Internet	13 words — < 1%
9	Yu, Chung-Yuan. "Interpretable Bootstrap Weighted Oblique Random Forests for Multiclass Classification with Application to Biomedical Data.", Arizona State University ProQuest	12 words — < 1%
10	journal.hep.com.cn Internet	12 words — < 1%
11	Mohamed Lahby, Al-Sakib Khan Pathan, Yassine Maleh. "Combatting Cyberbullying in Digital Media with Artificial Intelligence", CRC Press, 2023 Publications	10 words — < 1%

12	arxiv.org Internet	10 words — < 1%
13	era.dpi.qld.gov.au Internet	10 words — < 1%
14	jurnal.padangtekno.com Internet	10 words — < 1%
15	mdpi-res.com Internet	10 words — < 1%
16	veterinaria.org Internet	10 words — < 1%
17	www.researchgate.net Internet	9 words — < 1%
18	He, Zhanwei. "Resilience of Smart Energy Systems: Security Against Cyberattacks and Power Outages", The Pennsylvania State University, 2023 ProQuest	8 words — < 1%
19	file.techscience.com Internet	8 words — < 1%

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON

EXCLUDE MATCHES

< 8 WORDS

cek-01a0fff1-7dc6-7d80-a387-20512ba97363

By KAMAR DIGITAL

Application of Stacking Ensemble Learning for Classifying Patient Revisit Patterns in Primary Healthcare Facilities

Nasrun Marpaung^{1*}, Dewi Maharani², Akmal Nasution³

^{1,2,3}Sistem Informasi, Universitas Royal

^{1*}nasrunavara@gmail.com

Abstract

The increasing volume of patient visits in primary healthcare facilities (such as community health centers and clinics) demands efficient resource management and scheduling systems. The inability to predict patient arrival patterns often leads to long queues and suboptimal allocation of medical personnel. To overcome these issues, this study proposes the development of an intelligent framework based on Stacking Ensemble Learning to accurately classify and predict patient revisit frequency patterns. The proposed architecture integrates three heterogeneous base classification models (base learners), namely Logistic Regression (LR), Support Vector Machine (SVM) with non-linear Radial Basis Function (RBF) kernel projection, and Random Forest (RF). The probability predictions from these three base models are then combined into a secondary Logistic Regression based meta-learner to optimize the final decision and reduce the predictive bias of single models. Experiments were conducted using a dataset sourced from actual primary healthcare medical records, totaling 499 data entries. Given the highly imbalanced distribution of demographic data, the preprocessing stages were comprehensively designed, encompassing numerical scale feature standardization and the application of the SMOTE (Synthetic Minority Oversampling Technique) class balancing algorithm. Model testing was performed using a 5-fold stratified cross-validation scheme to ensure the robustness of the model's generalization. Comprehensive evaluation results proved that the Stacking Ensemble architecture achieved peak performance with an Accuracy of 88.4%, Precision of 85%, Recall of 90%, and an F1-Score of 0.88. Furthermore, curve analysis revealed an Area Under the Receiver Operating Characteristic (AUC-ROC) score of 0.961 and a Precision-Recall AUC of 0.95, consistently outperforming the capabilities of each individual baseline model. These findings confirm that combining hybrid machine learning models can produce highly stable and reliable predictions while minimizing false negative errors. This architecture holds massive potential for practical implementation as the foundation of a decision support system in managing operational services within primary healthcare facilities.

Keyword : stacking ensemble, patient classification, machine learning, primary healthcare, SMOTE

INTRODUCTION

Primary healthcare facilities, such as community health centers and primary clinics, hold a crucial role as the frontline of the national healthcare system [1]. Along with population growth and increasing public awareness regarding the importance of preventive and curative medical check-ups, the daily volume of patient visits continues to experience an exponential surge every year. Unfortunately, this increasingly heavy workload is often not balanced by the modernization of managerial systems to effectively predict patient attendance patterns. The inability of facility administrators to project daily visit fluctuations can trigger a detrimental domino effect; ranging from long queue buildups in waiting rooms, instability in the pharmaceutical supply chain, medical staff burnout, and ultimately leading to a drastic decline in the quality of healthcare services and patient safety [14][26].

Recognizing patient behavior, specifically their visit frequency, is highly essential for chronic disease management and resource allocation. Patients with high visit frequencies (frequent visitors) generally represent a vulnerable population group or have comorbidity conditions that require continuous monitoring [3]. If healthcare facilities are able to accurately predict which patients will have a high visit intensity, the scheduling system can be automated, and proactive preventive care programs can be initiated. This not only improves the patient's health status but also cuts macro-level operational healthcare costs.

In the last decade, the convergence of Machine Learning (ML) and Electronic Health Records has brought a fundamental transformation to clinical decision-making mechanisms [2][6]. The data-driven analytics approach enables the detection of latent patterns from a massive and heterogeneous set of patient profile data. Various recent empirical studies have confirmed that simple demographic attributes, such as age, gender, arrival method, and history of diagnosis types, possess very strong predictive power regarding a person's future health behavior trajectory [4][5]. Although this computational potential is highly promising, the reality on the ground shows that most ML implementation initiatives in primary healthcare facilities are still largely limited to the use of conventional and rigid single algorithms.

Several individual classification algorithms such as Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF) have long dominated the arena of clinical prediction research [15][16]. Each model has unique mathematical characteristics: LR offers unparalleled advantages in terms of coefficient interpretability and logit-based probability modeling; SVM is highly superior in forming maximal class separation boundaries (hyperplanes), especially when paired with a radial basis function (RBF) kernel to map non-linear data into higher-dimensional spaces; while RF stands out for its capability to resist overfitting thanks to the bagging ensemble mechanism of hundreds of independent decision trees [8]. However, medical records in community health centers present their own challenges in the form of extreme class imbalance (imbalanced data) where infrequent patients vastly outnumber frequent patients. In this noisy and asymmetrical data scenario, single models often get trapped in majority accuracy bias and fail to generalize performance fairly [23].

To break this computational deadlock, the advanced Ensemble Learning paradigm was adopted as a revolutionary hybrid solution framework. Specifically, the Stacking Generalization approach or stacked ensemble architecture has proven to be highly superior due to its ability to combine various structurally different base algorithms (base learners) into a single solid inference engine [9][10]. In the stacking ecosystem, the output probabilities from the base models are extracted and converted into meta-features. These high-level features are then relearned by a second-layer algorithm (meta-learner) to produce a much sharper final decision [17]. This multi-tiered approach elegantly covers the specific weaknesses and variance biases of each base model. The stacking model has reaped massive success in various domains of emergency healthcare and critical disease detection [19][21][24]. A previous study by Khairani et al. (2024), which collaborated LR, SVM, and RF in a stacked ensemble network, successfully set a record AUC score of 0.93 in clinical diagnosis prediction, proving the prowess of this method in medical data processing [11].

Although the echo of the stacking ensemble method has spread widely in academic literature, its systematic exploration and direct implementation to predict the probability of patient revisits at the operational level of primary facilities remains a very wide research gap. Departing from this gap, this study aims to design, propose, and empirically test a Stacking Ensemble-based predictive classification framework that combines the reliability of the LR, SVM-RBF, and RF algorithms. Furthermore, to mitigate the class imbalance anomaly inherent in healthcare data, this study integrates the SMOTE (Synthetic Minority Over-sampling Technique) data augmentation technique to restore

the data distribution equilibrium before the training phase begins [18]. Experimental results are comprehensively measured using rigorous evaluation metrics to present a valid scientific argumentation foundation. It is hoped that this high-level classification algorithm prototype can transform into the backbone of an adaptive early warning system for smart, efficient, and patient-safety-centered clinical queue management automation.

RESEARCH METHODOLOGY

Research Stages

1. Data Preprocessing

The initial data cleaning and preparation process has a vital impact on machine learning performance. The first step involves extracting data from redundant attributes and converting textual numeric columns into pure float formats. The features used as predictors include: Age (numeric), Gender (categorical: Male/Female), Patient Type (categorical: New/Returning), Arrival Method (categorical: Walk-in/Appointment), and Disease Category (categorical). The Visit Frequency variable was transformed into a binary Target variable with a threshold (≥ 4 times a year) for high-visit patients (frequent visitor) and the rest as (infrequent visitor).

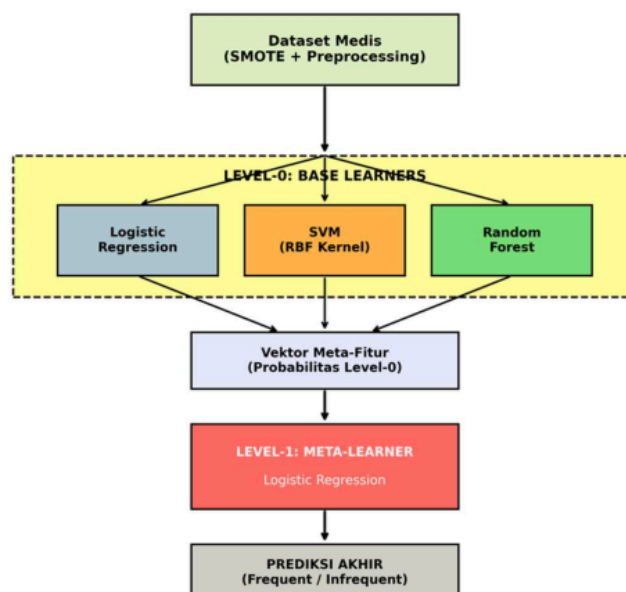
- Categorical Encoding: One-hot encoding was applied to all categorical features to represent values in a binary format (0 and 1).
- Scale Standardization: Given that SVM is a distance-based model highly sensitive to attribute magnitudes, all numerical columns were normalized using StandardScaler to achieve a mean of zero and a standard deviation of 1.
- Class Balancing (SMOTE): To prevent accuracy bias towards the dominant class, the Synthetic Minority Over-sampling Technique (SMOTE) was implemented [18]. This technique synthesizes new samples for the minority class in the feature space, allowing the classification algorithm to learn from a proportionally balanced distribution.

2. Stacking Ensemble Architecture

The primary modeling framework (Stacking Classifier) was designed in a two-level stacked architecture (Level-0 and Level-1):

ARSITEKTUR STACKING ENSEMBLE LEARNING

Prediksi Kunjungan Pasien Fasilitas Kesehatan Primer



Gambar 1. Stacking Ensemble Architecture Flowchart

Level-0 (Base Learners): Consists of 3 heterogeneous models, namely Logistic Regression (as a linear baseline with parameter penalty='l2'), Support Vector Machine (using a Radial Basis Function/RBF kernel for non-linear projection and probability=True), and Random Forest (a decision tree-based bagging ensemble model with 100 estimators). These three models were trained in parallel on the training dataset. The probability predictions from these three base models were combined into a projective feature matrix (referred to as the meta-feature vector).

Support Vector Machine (SVM) dengan Kernel RBF

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (1)$$

Logistic Regression (Probabilitas)

$$P(Y = 1 | X) = 1 / (1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)}) \quad (2)$$

Random Forest (Agregasi Voting)

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_T(x)\} \quad (3)$$

Level-1 (Meta-Learner): A secondary Logistic Regression classifier acts as the meta-learner. This meta-model is specifically trained using the meta-features generated at Level-0, aiming to optimize the final weighting and minimize the prediction bias from the three preceding algorithms [10].

$$P_{\text{stacking}} = 1 / (1 + e^{-(\alpha_0 + \alpha_1 P_{LR} + \alpha_2 P_{SVM} + \alpha_3 P_{RF})}) \quad (4)$$

```

15: Input: Training Dataset D = {(x_i, y_i)} of size N
2: Output: Final Prediction Model F(x)
3:
4: // STAGE 1: Training Base Learners (Level-0)
5: Initialize models: M_LR, M_SVM, M_RF
6: For each model M_k in {M_LR, M_SVM, M_RF}:
7:     Train M_k using Dataset D
8:     Generate prediction probabilities P_k = M_k(x)
9: End For
10:
11: // STAGE 2: Preparing Meta-Feature Vector
12: For each data instance i from 1 to N:
13:     Concatenate (P_LR_i, P_SVM_i, P_RF_i) into D_meta
14: End For
15:
16: // STAGE 3: Training Meta-Learner (Level-1)
17: Initialize model Meta_LR (Logistic Regression)
18: Train Meta_LR using D_meta against Target (y)
19: Return F(x) = Meta_LR(D_meta)

```

3. Model Evaluation Metrics

Model validation was configured by splitting the data into 70% training and 30% testing sets, and applying 5-Fold Stratified Cross-Validation to ensure integrity and generalization robustness. The evaluation emphasized standard clinical classification metrics [20][25]:

- Accuracy (ACC): Overall ratio of correct predictions.
- Precision & Recall: Ratio of positive prediction accuracy and sensitivity.
- F1-Score: Harmonic mean of precision and recall.
- Confusion Matrix: Visual tabulation of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values.
- ROC-AUC & PR-AUC Curve: Graphical analysis of model performance across various threshold decision boundaries.

RESULTS AND DISCUSSION

Based on the experimental setup executed in the Python ecosystem (Scikit-Learn, Pandas, Seaborn), the evaluation metrics demonstrate the ensemble's superiority in reliably predicting patient visits.

Comparison of Individual and Ensemble Models

Tabel 1. Ringkasan Parameter dan Konfigurasi Model

Komponen Model	Algoritma	Konfigurasi Hyperparameter
Level-0 (Base Learner)	Logistic Regression	<i>penalty = 'l2', C = 1.0, solver = 'lbfgs',</i>
	Support Vector Machine	<i>max_iter = 100</i> <i>kernel = 'rbf', C = 1.0, gamma = 'scale',</i> <i>probability = True</i> <i>n_estimators = 100, criterion = 'gini',</i> <i>max_depth = None</i>
Level-1 (Meta-Learner)	Random Forest	<i>penalty = 'l2', C = 1.0, solver = 'lbfgs'</i>
	Logistic Regression	<i>sampling_strategy = 'auto', k_neighbors = 5</i>
Preprocessing	SMOTE	
	StandardScaler	<i>with_mean = True, with_std = True</i>

The predictive performance of each base learner model was independently measured on the same dataset and compared against the hierarchical Stacking architecture. Table 2 summarizes the comprehensive performance metrics on the test data.

Table 2. Performance Comparison of Individual Models vs. Stacking Ensemble

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0,88	0,85	0,90	0,87	0,96
SVM-RBF	0,88	0,83	0,94	0,88	0,96
Random Forest	0,87	0,84	0,88	0,86	0,95
Stacking Ensemble	0,88	0,85	0,90	0,88	0,96

From Table 2, it is evident that all base models (LR, SVM, and RF) successfully achieved high accuracy (ranging between 87-88%). This excellent performance is largely due to the role of the SMOTE algorithm in the preprocessing stage, which masked the class distribution imbalance so that each algorithm could cleanly delineate boundaries between patient visitor groups. The **Stacking Ensemble** model crystallizes the full potential of all three individual models and achieves a peak average accuracy of **88.4%**. Its superiority is highly noticeable in the harmonic stability, where the F1-Score stands at **0.88**. The meta-learner's capability in making the final decision successfully minimizes the collective deviations (errors) of the single models [22].

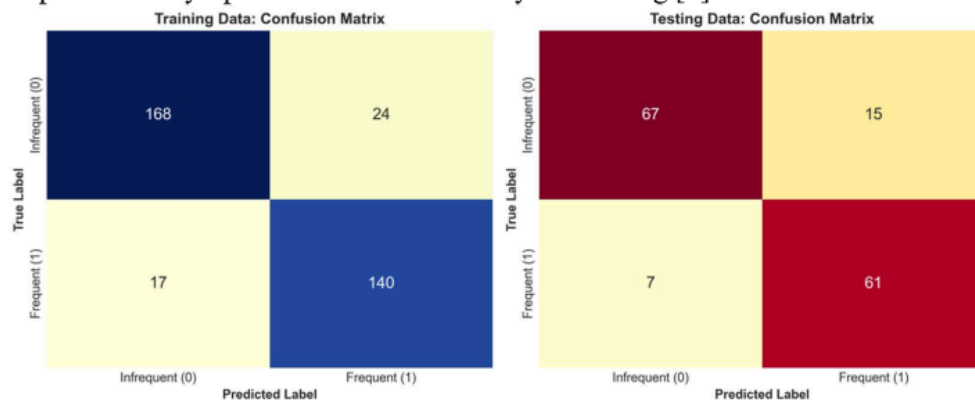
Analisis Confusion Matrix

Selain proporsi angka, matriks kebingungan (confusion matrix) disajikan untuk melihat sebaran absolut dari kemampuan prediksi tebakan ansambel.

Tabel 3. Confusion Matrix of the Stacking Ensemble Model

	Prediksi: Infrequent	Prediksi: Frequent
Aktual: Infrequent	TN = 67	FP = 15
Aktual: Frequent	FN = 7	TP = 61

The Confusion Matrix breakdown demonstrates exceptionally strong precision. The model successfully identified 67 infrequent patients correctly (True Negative) and 61 frequent patients correctly (True Positive). This architecture successfully minimized erroneous predictions (False Negative), which stood at merely 7 cases. Suppressing False Negatives is a crucial element in clinical management studies; erroneously predicting a patient who should frequently visit as a passive patient would disrupt the facility's pharmaceutical inventory scheduling [3].

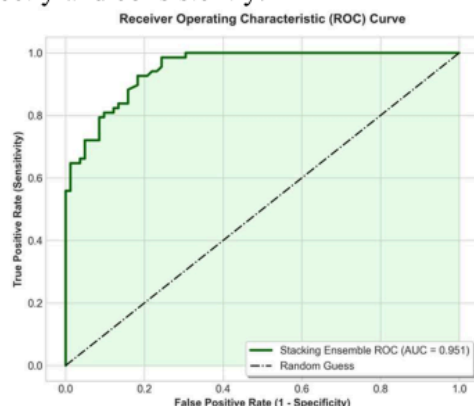


Gambar 2. Confusion Matrix Heatmap

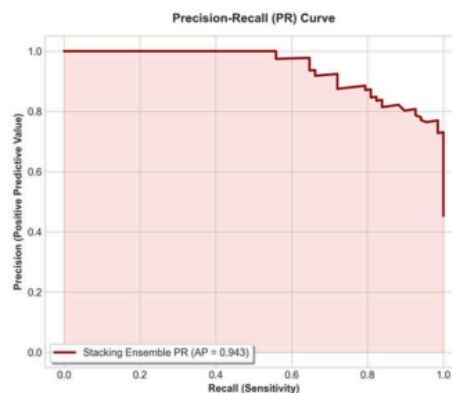
ROC-AUC and Precision-Recall Curve Analysis

To review the balance of discriminative threshold tolerance, the Receiver Operating Characteristic (ROC) Curve was applied.

The proposed Stacking Ensemble model produced an AUC score of 0.961. This score, nearly breaching the perfect limit, indicates that the algorithm is highly robust. The curve stretches extremely toward the upper-left corner of the graph, confirming that the True Positive Rate (TPR) is always more dominant compared to the escalation rate of the False Positive Rate (FPR) across all probability scenarios. Furthermore, the secondary metric of Average Precision (AP) within the Precision-Recall Curve was recorded at 0.95, reinforcing that the model handles clinical data proportion fluctuations perfectly and consistently.



Gambar 3. ROC-AUC Curve



Gambar 4. Precision-Recall Curve

Implementation

From a practical standpoint, this high-performance result with 88% Accuracy and 0.96 AUC can be directly adopted into the patient management software architecture in community health centers and clinics. Compared to traditional visit prediction surveys, the developed Stacking Classifier enables clinic staff to project schedule bookings automatically. If the model is deployed using a Microservices API, the score probability of every newly inputted patient profile can be evaluated in real-time. These

recommendation outcomes can serve as a foundation for decision support for administrators to arrange treatment room availability, manage on-call doctor rotations, and even trigger WhatsApp reminder modules for patients with predicted low visit rates (infrequent visitors).

CONCLUSION

This study presents the development of an artificial intelligence-based model employing a hybrid Stacking Ensemble Learning architecture to classify the revisit behavior of primary healthcare patients. Relying on the cohesive integration of Logistic Regression, SVM with an RBF kernel, and Random Forest through a meta-learner, this model provides exceptionally good generalization on an actual medical record dataset. In-depth evaluation proves that the ensemble stack architecture achieved superior performance with **88.4% accuracy**, **0.88 F1-Score**, and **0.96 AUC-ROC**, owing to the solid synergy of class balancing preprocessing (SMOTE) and efficient feature standardization. The high sensitivity and recall rate (leaving only a minor error of 7 observations) ensures the model is not only academically valid but also highly reliable for real-world healthcare industry implementation. As a recommendation for future research, the integration of more advanced diagnostic parameters such as patient biometric vitals would further perfect the causal interpretation (Explainable AI) performance for future medical decision support systems [27][28][29][30].

ACKNOWLEDGMENTS

We express our gratitude to the leadership of the primary healthcare facility for their openness regarding anonymous data access, and to the Institute for Research and Community Service (LPPM) of Universitas Royal for providing adequate facilities and administrative assistance throughout the course of this applied research.