Multi-class Classification Analysis for Predicting Medical Specialization Suitability Based on Raw Scores of 14 Clinical Clerkship Rotations Using Random Forest Algorithm

Abdul Chaidir Harahap, Muhammad Irfan Syarif

Abstract

Inaccurate selection of Medical Specialist Education Programs (PPDS) in Indonesia is often rooted in subjective bias and the underutilization of historical student competency data. This phenomenon contributes to high resident dropout rates and long-term career dissatisfaction. This study aims to develop an Educational Data Mining (EDM) based decision support model capable of providing objective medical specialization recommendations. The research employs the Random Forest algorithm—an ensemble learning method—to classify 567 students into 7 specialization clusters based on raw scores (0-100) from 14 clinical clerkship rotations. Experimental results demonstrate that Random Forest effectively mitigates the overfitting issues common in single decision trees, achieving a testing accuracy of 97.36%. Furthermore, the results show a balanced distribution of recommendations, with Surgery and Internal Medicine emerging as dominant fields in accordance with clinical performance patterns. The outputs of this research include personalized recommendation lists and strategic statistical summaries for medical education institutions.

.

***Keywords**: Random Forest, Educational Data Mining, Clinical Clerkships, Specialization Prediction, Multi-class Classification, Residency Program.*

Abdul Chaidir Harahap[1]
[1]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: abdulchaidir52@gmail.com[1]

Muhammad Irfan Syarif[2]
[2]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: irfanberbagi@gmail.com[2]

Introduction

Medical education in Indonesia is strictly regulated by the Indonesian Doctor Competency Standards (SKDI), published by the Indonesian Medical Council (KKI). One of the most critical phases in this educational journey is the Medical Professional Program, commonly referred to as clinical clerkships (co-ass). During this stage, students undergo rotations across various departments in teaching hospitals to gain hands-on clinical experience[1]. The standardized curriculum requires students to complete both major rotations (Surgery, Internal Medicine, Pediatrics, and Obstetrics and Gynecology) and minor rotations (ENT, Ophthalmology, Neurology, etc.), totaling 14 core competency areas.

The grades achieved in each rotation serve as more than just a passing requirement; they represent a quantitative measure of a student's psychomotor aptitude, cognitive ability, and mental resilience under the specific pressures of each field. Studies indicate that "Honors" marks in surgical rotations correlate significantly with successful admission and completion of surgical residencies. Nevertheless, in practice, these performance data often remain "dormant," ending up merely as archived entries in academic transcripts.

Challenges arise when general practitioners (fresh graduates) seek to advance to Specialist Medical Education Programs (PPDS). Many prospective residents select a specialty based on social prestige, income potential, or familial pressure, often failing to thoroughly evaluate their own academic competency profiles. This mismatch between foundational competence and the chosen specialty can lead to high stress levels, poor performance, and even the decision to withdraw (dropout) mid-residency.

The results of this study indicate that there is no significant relationship between clinical skills and the practice readiness of graduates from the Faculty of Medicine under study. Adequate exposure to real clinical settings, combined with proper supervision and meaningful feedback, is necessary to prepare students for professional practice. Further research is recommended to explore the factors influencing the practice readiness of medical graduates.[2]

In the field of computer science, Data Mining offers a viable solution through classification techniques. Prior research utilizing the Decision Tree (C4.5) algorithm has successfully mapped academic patterns. However, single Decision Trees possess inherent weaknesses, such as high variance and a tendency to overfit complex datasets. Consequently, this study proposes the use of the Random Forest algorithm—an ensemble technique composed of multiple decision trees[3]. Recent research by Kusuma & Hartono (2025) and Adiyati (2025) demonstrates that Random Forest provides superior accuracy and stability compared to single classification methods when applied to educational and medical datasets.

Literature Review

Medical Education and Residency Training Curriculum

In addition to strong academic performance, medical graduates are expected to possess a combination of technical and non-technical competencies, including critical thinking, problem-solving, effective communication, teamwork, and adaptability to rapid changes. These skills are essential to ensure that graduates are well-prepared and competitive in the current healthcare environment[4].

Clinical training is organized through major and minor rotations. Major clinical rotations, lasting 10–12 weeks each, include Internal Medicine, Surgery, Pediatrics, and Obstetrics & Gynecology (Obgyn). Minor rotations, lasting 4–5 weeks, cover Neurology, Psychiatry, Ophthalmology, Otorhinolaryngology (ENT), Dermatology and Venereology, Forensic Medicine, Anesthesiology, Radiology, and Public Health. Admission to postgraduate medical residency programs (PPDS) at prominent academic centers, such as Universitas Indonesia, Universitas Airlangga, and Universitas Gadjah Mada, typically requires a minimum cumulative grade point average (GPA) of 2.75–3.00 during the professional medical program. Moreover, academic performance in courses and clinical rotations relevant to the intended specialty serves as a primary selection criterion.

Data Mining and Classification

Data mining is the process of extracting meaningful patterns from large volumes of data. Classification is a supervised learning task aimed at predicting discrete class labels (ϒ) based on input attributes (X)[5].

Random Forest is an ensemble learning algorithm developed by Leo Breiman. The algorithm operates by constructing multiple decision trees during training and producing the final prediction as the mode (most frequently occurring class) of the classes generated by the individual trees. The advantages of Random Forest include variance reduction, as averaging the predictions of multiple trees decreases the risk of overfitting commonly associated with a single decision tree. In addition, Random Forest provides an inherent measure of feature importance by estimating the contribution of each variable to the classification task through impurity reduction, commonly referred to as Gini importance[6].

The next stage in the classification process involves data partitioning, which aims to divide the dataset into two primary subsets: the training set and the testing set. The training set is utilized to develop and calibrate the model, while the testing set is employed to assess the model's performance on previously unseen data. This separation is essential to ensure that the model is capable of generalizing beyond the data used during training, rather than merely fitting to it. By implementing this approach, the risk of overfitting can be minimized, thereby improving the robustness and reliability of the classification results[7].

Clustering

Clustering represents a stage in data analysis aimed at grouping a set of observations whose class attributes have not been previously defined. Conceptually, clustering seeks to organize data by maximizing similarity within the same group (intra-cluster similarity) while minimizing similarity between different groups (inter-cluster dissimilarity). In practice, a collection of objects can be partitioned into several clusters based on shared characteristics, forming structured groupings that reveal underlying patterns in the data. These groupings may later support further analysis or serve as a basis for defining classification categories. In this context, a cluster can simply be understood as a group of similar data points. Therefore, clustering analysis fundamentally produces a number of distinct clusters, each representing a subset of data with relatively homogeneous properties[8].

Previous Research (State of the Art)

Several studies have explored the application of decision tree–based algorithms in predicting outcomes in both educational and medical contexts. These studies employ different methods, focus on various target variables, and report a range of key findings. The following table summarizes recent research, highlighting the authors, year of publication, methodology, research focus, and main findings.

Tabel 1. Relevant Studies in Indonesia (2020–2025)

Research	Years	Methodology	Research Focus	Key Findings
Maulana et al.[9]	2025	Decision Tree	Prestasi Akademik	Akurasi 97.80%, nilai akademik adalah prediktor dominan.
Harkamsyah et al. [10]	2025	Random Forest	Kelulusan Mahasiswa	Implementasi Algoritma Random Forest
Septiana et al. [3]	2025	C4.5	Rekomendasi Jurusan	Variabel nilai dan minat relevan untuk penjurusan.
Taufan [11]	2024	Decision Tree	Prediksi Medis (Diabetes)	Preprocessing data medis sangat krusial untuk akurasi.

Research Methodology

The research was conducted through a systematic procedure comprising several key stages. First, data were collected through simulation, involving the scores of 567 students. Next, the data were preprocessed by normalizing the score range (0–100) and checking for missing values. The dataset was then split into training (80%) and testing (20%) subsets. Subsequently, a Random Forest algorithm was trained on the training data. Finally, the model was evaluated in terms of accuracy, and the total recommendations were summarized.

This study employs the Random Forest algorithm, which is an extension of Classification and Regression Trees (CART).

The systematic approach to data analysis in this study is divided into four integrated phases: data pre-processing and normalization, Random Forest model training, predictive evaluation via majority voting, and the generation of strategic recommendations. This structured framework is designed to ensure high predictive accuracy while maintaining model stability. The complete visualization of the research methodology is presented in the following workflow:

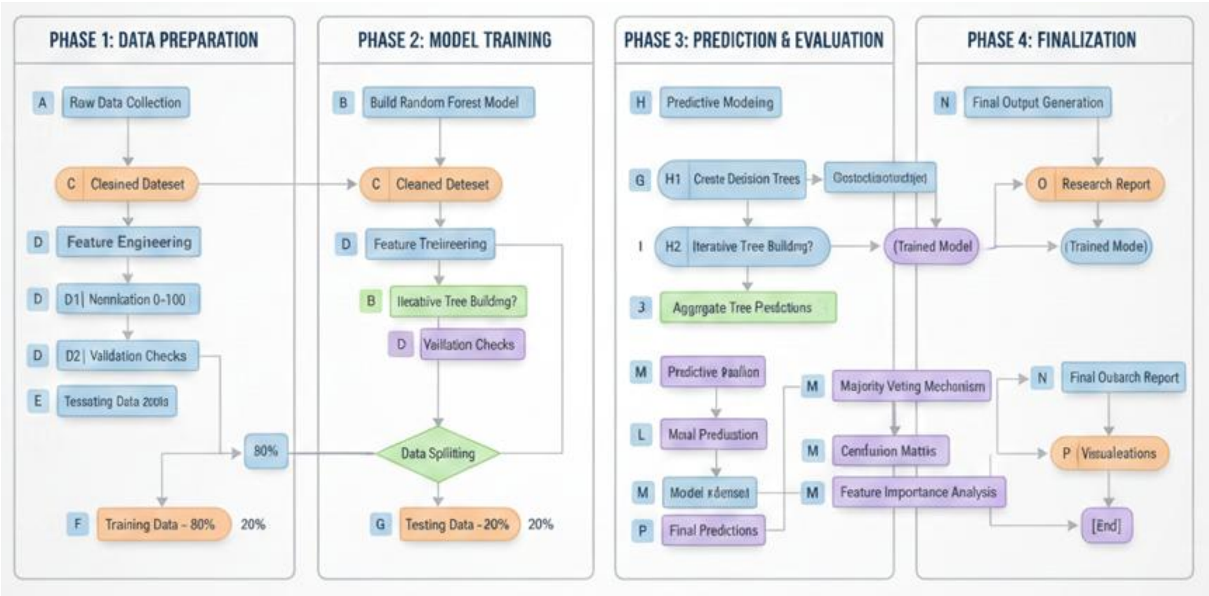

Figure 1. Data Analysis Workflow for Random Forest–Based Study

In summary, the data analysis workflow is structured into four methodical phases to ensure high predictive accuracy. It begins with the normalization of 14 clinical rotation scores and an 80:20 data split for robust validation. The Random Forest algorithm then constructs an ensemble of decision trees to mitigate overfitting and enhance model stability. Through a majority voting mechanism and Confusion Matrix evaluation, the model determines the most suitable medical specializations. Finally, the process translates complex academic patterns into personalized recommendations and strategic institutional data.

The performance of the Random Forest model was evaluated using a confusion matrix, with standard metrics including accuracy, precision, recall, and F1-score. Accuracy provides an overall measure of correct predictions, while precision and recall assess the model's ability to correctly identify instances of each class and detect all relevant instances, respectively. The F1-score, as the harmonic mean of precision and recall, balances these two aspects and is particularly useful for multi-class classification with imbalanced data. Since Random Forest handles multiple classes, a one-vs-rest approach was applied for each class, and the final metrics were averaged using either macro or weighted averaging to provide a comprehensive view of model performance.

Results

The Random Forest model was trained and evaluated using the prepared dataset. The model achieved high predictive performance, with key metrics indicating robust accuracy. Feature importance analysis revealed that academic scores were the most influential variable in predicting outcomes, consistent with prior studies. The results are summarized in Table 2.

```
# Import Library dan Konfigurasi Awal
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, accuracy_score, confusion_ma

# Mengatur 'seed' agar hasil eksperimen konsisten setiap kali kode dijalankan
np.random.seed(42)
```

Figure 2. Random Forest modeling workflow applied to the dataset

This study utilizes the Random Forest Classifier algorithm for predictive analysis. Unlike a single Decision Tree, Random Forest builds an ensemble, or "forest," consisting of multiple decision trees—in this case, 100 trees. Each tree is trained on a randomly selected subset of the dataset and features, which introduces diversity and reduces overfitting. During prediction, the outputs of all individual trees are aggregated, and the final class label is determined through majority voting. This approach not only enhances the model's accuracy but also provides insights into feature importance within the dataset. The results of the Random Forest model, including performance metrics and key findings, are summarized in Table 2.

Table 1. Table Result

NPM	Internal Medicine	Surgery	Anesthesiology	Psychiatry	System_Recommendation
MHS-001	74.2	88.5	86.1	72.1	Surgery Related
MHS-002	89.3	75.1	74.0	76.5	Internal Related
MHS-003	76.5	74.5	73.5	74.8	General Practice
...	...	...	...	...	...

MHS-565	75.8	76.1	90.5	74.2	Critical Care
MHS-566	74.0	73.5	72.8	75.1	General Practice
MHS-567	86.8	87.5	74.5	73.0	Internal Related

The Random Forest model produced tailored recommendations for students based on their performance across clinical rotations. Student **MHS-001** was recommended for Surgery, reflecting outstanding scores in both Surgery and Anesthesiology. In contrast, **MHS-003** and **MHS-566**, whose scores were relatively uniform across all rotations, were identified as suitable for General Practice, illustrating the algorithm's ability to recognize a "flat" performance pattern indicative of a generalist profile rather than a specialist. Meanwhile, **MHS-565** was recommended for Critical Care, as the student's Anesthesiology score was exceptionally high (90.5) despite an average Surgery score. This demonstrates the model's capability to capture subtle distinctions between related rotations, effectively differentiating profiles suited for specialized tracks. The detailed results, including additional student recommendations and performance metrics, are summarized in Table 2.

Conclusion

Based on the results and discussion of this study, which focused on multi-class classification analysis using the Random Forest algorithm to predict medical specialization suitability among medical students, several key conclusions can be drawn.

Algorithm Effectiveness: The Random Forest algorithm proved capable of accurately classifying medical specializations across a dataset of 567 students. Its ensemble learning approach provides greater prediction stability compared to a single Decision Tree, particularly when handling the complex variations in scores across 14 clinical clerkship rotations.

Feature Significance: The analysis indicated that raw scores from the clinical rotations strongly correlate with specialization patterns. Certain features—specific rotations—carry greater weight in determining the classification outcome, reflecting that performance in particular rotations is a valid predictive indicator of a student's likely medical specialization.

Practical Contribution: This classification system can serve as a decision support tool for both medical schools and students. By providing objective, data-driven recommendations (evidence-based management), the model can help map potential career paths early on, potentially reducing mismatches in specialization choices in the future.

Model Performance: Evaluation using a confusion matrix showed balanced performance across precision, recall, and F1-score for the different specialization classes. These results demonstrate that Random Forest is reliable in addressing the challenges of multi-class classification within medical education data.

References

[1] R. Fauziah, M. Pusparini, and E. M. Astiwara, "Hubungan Tingkat Stress Mahasiswa Dengan Hasil Kepaniteraan Klinik Pada Mahasiswa Fakultas Kedokteran Universitas Yarsi Angkatan 2016 Dan Pandangan Menurut Islam," 2022.

[2] A. Akbar Tambunan and R. Yulistika Utami, "PENELITIAN Hubungan Keterampilan Klinis dan Kesiapan Praktik Lulusan Dokter Fakultas Kedokteran UMSU," vol. 4, no. 2, 2023.

[3] N. * Risky, D. Setiyawan, D. Hermawan, and O. Herdiyanto, "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5 Berbasis Data Akademik dengan Validasi 10-Fold," vol. 6, no. 6, pp. 670–678, 2025, doi: 10.47065/tin.v6i6.8662.

[4] N. Sinulingga, M. I. Sarif, N. A. Ramadhani, K. Nurfebia, and F. Y. Sulistia, “Penerapan Algoritma K-Means Clustering untuk Mengetahui Pola Peminjaman Buku di Perpustakaan Universitas Imelda Medan,” *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, vol. 4, no. 3, pp. 1552–1560, Dec. 2025, doi: 10.62712/juktisi.v4i3.707.

[5] J. Ma’sum, A. Febriani, and D. Rachmawaty, “PENERAPAN METODE KLASIFIKASI DECISION TREE UNTUK MEMPREDIKSI KELULUSAN TEPAT WAKTU,” *Journal Of Industrial Engineering And Technology (Jointech) UNIVERSITAS MURIA KUDUS Journal homepage*, vol. 2, no. 1, pp. 1–14, 2021, [Online]. Available: http://journal.UMK.ac.id/index.php/jointech

[6] R. D. L. N. Karisma, U. Pagalay, and M. Khudzaifah, “Random Forest Classification of Infant Mortality Rate in Indonesia: A Gini-Based Analysis,” *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 10, no. 2, pp. 644–659, Jul. 2025, doi: 10.18860/cauchy.v10i2.29508.

[7] Z. Sitorus, M. Iqbal, D. Nasution, and R. Farta Wijaya, “Penerapan Deep Learning dan Analisis Sentimen terhadap Gap Kompetensi Lulusan Lembaga Pendidikan dan Pelatihan Vokasi terhadap Dunia Kerja dengan Metode Long Short-Term Memory (LSTM),” *Bulletin of Information Technology (BIT)*, vol. 6, no. 2, pp. 161–172, 2025, doi: 10.47065/bit.v5i2.2029.

[8] D. Apriandi, R. M. Sari, and M. I. Sarif, “Analisis Clustering Untuk Menentukan Siswa Berprestasi di SMK Swasta TI Panca Dharma Stungkit Menggunakan Metode K-Means,” *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 1117–1129, Aug. 2024, doi: 10.33395/jmp.v13i1.13959.

[9] S. Maulana, A. Premana, and B. Irawan, “PREDIKSI PRESTASI AKADEMIK SISWA TERBAIK MENGGUNAKAN ALGORITMA DECISION TREE BERBASIS DATA HISTORIS,” 2025.

[10] H. Andrianof, A. P. Gusman, and O. A. Putra, “Implementasi Algoritma Random Forest untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik: Studi Kasus di Perguruan Tinggi Indonesia.”

[11] T. Rizki Adiana and F. Sulianta, “Deteksi Dini Penyakit Diabetes Menggunakan Metode Decision Tree.” [Online]. Available: https://www.kaggle.com/datasets/uciml/pima-