JENTIK

OPEN ACCESS

Jurnal Pendidikan Teknologi Informasi dan Komunikasi
https://ejournal.1001tutorial.com/index.php/jentik

School Feasibility Analysis and Grade Improvement Strategies Using the Random Forest Algorithm

Farrel Rahma Aliyya[1*], Syahandhika Naufal Farizi[2*], Lala Septem Riza[3], Rani Megasari[4], Eki Nugraha[5], Asep Wahyudin[6]

[1,2,3,4,5,6]Pendidikan Ilmu Komputer, Universitas Pendidikan Indonesia
Jawa Barat, Indonesia
farrelrahmaa@upi.edu*

Article Information:
Received July 09, 2025
Revised July 25, 2025
Accepted December 27, 2025

Keywords:
Education; Machine Learning; Random Forest

Abstract

Background of Study: Educational disparities across Indonesian provinces persist, particularly in infrastructure, teacher quality, and dropout rates, necessitating data-driven analysis for equitable improvements.

Aims: This study investigates school feasibility and proposes strategies to enhance provincial education performance using the Random Forest algorithm.

Methods: Aggregated provincial education data covering student numbers, dropout rates, teacher qualifications, and classroom conditions were transformed into derivative indicators. A binary classification (Feasible/Not Feasible) based on national dropout median was applied. The model was developed using R with six systematic steps, including training and evaluation of a Random Forest model (ntree = 100, mtry = 3) using accuracy, sensitivity, and specificity.

Result: The model accurately classified school feasibility. Key predictors included teacher quality, student-teacher ratios, and classroom conditions. Several provinces were identified as “Not Feasible.”

Conclusion: Machine learning proves effective for education policy support. The study offers targeted recommendations such as improving infrastructure, enhancing teacher training, and reducing dropouts to promote equitable education in Indonesia.

A. Introduction

Education plays a crucial role in national development (Panoyo, 2024). High-quality education can produce superior, competitive, and adaptive human resources in response to the challenges of a rapidly evolving era (Delima et al., 2025; Nasution et al., 2025). However, many regions in Indonesia continue to face significant disparities in educational quality among schools (Madhakomala et al., 2025; Apriliansyah et al., 2025). These inequalities encompass various aspects such as the availability and quality of infrastructure, teacher competencies, student academic performance, and managerial support within schools. As a result, not all students have equitable access to quality education (Ashari et al., 2024; Muchtar et al., 2025).

Efforts to improve educational quality must be grounded in accurate evaluations of school feasibility (Gustiani et al., 2023; Astuti & Diantoro, 2021). School feasibility is a vital indicator reflecting the readiness of an educational institution to provide optimal learning services (Safrizal et al., 2022; Syofyan & Rosyid, 2024). However, the current evaluations are often subjective and inconsistent. Therefore, a data-driven approach is needed to offer a comprehensive view of school conditions and development potential (Wahidin & Hulbat, 2022; Jin et al., 2025).

How to Cite : Aliyya, F. R., Farizi, S. N., Riza, L. S., Megasari, R., Nugraha, E., & Wahyudin, A. (2025). School Feasibility Analysis and Grade Improvement Strategies Using the Random Forest Algorithm. *JENTIK: Jurnal Pendidikan Teknologi Informasi Dan Komunikasi*, 4(2), 101–110. https://doi.org/10.58723/jentik.v4i2.475
ISSN : 2963-1963
Published by : CV Media Inti Teknologi

Given the complexity of educational challenges, descriptive data alone is no longer sufficient for crafting effective policies. Deeper and more predictive analysis is required to uncover hidden patterns within large-scale and multidimensional educational datasets. Here, the role of analytic technologies becomes essential, particularly in assisting policymakers to explore variable relationships, identify risk factors, and comprehensively assess the effectiveness of educational programs (Paolucci et al., 2024; Kharis & Zili, 2022; Sunarya et al., 2025).

Machine learning has emerged as a widely adopted approach in the education sector, enabling large-scale data analysis and producing accurate, objective insights (Ersozlu et al., 2024; Yağcı, 2022; Zhang & Fan, 2025). One effective method for processing educational data is the Random Forest algorithm, which excels at classification tasks and identifying the importance of individual variables in prediction processes (Putra & Harahap, 2024; Zawiyah et al., 2022; Saputri et al., 2025). In this context, the Random Forest algorithm can be utilized to identify the key factors influencing school feasibility. Moreover, clustering schools based on similar characteristics enables a grading process, categorizing schools into several feasibility levels. Schools within the same cluster are assigned similar grades, allowing for more focused identification of disparities and targeted recommendations for improvement (Ramadhan & Susetyo, 2021; Luo, 2023). A study by (Ramadhan et al., 2019) used the Random Forest algorithm to identify key factors in assessing education quality based on accreditation indicators. It found that infrastructure standards, human resource standards, and graduate competency standards played significant roles in determining educational quality.

Therefore, this study not only offers an innovative approach to evaluating school feasibility through the Random Forest algorithm, but also provides strategic, data-driven solutions to address real-world educational challenges in Indonesia. By identifying highly influential indicators such as dropout rates, the percentage of teachers with a bachelor's degree, and student repetition rates, this research enables more precise and context-sensitive policy recommendations aligned with regional characteristics.

B. Research Methods

This study employs the Random Forest method, which is capable of capturing nonlinear patterns and interactions among variables without assuming specific functional relationships. This makes it particularly suitable for analyzing aggregated educational data influenced by various structural factors. As such, this research contributes not only methodologically but also practically by offering data-driven recommendations to support educational quality improvement efforts at the provincial level.

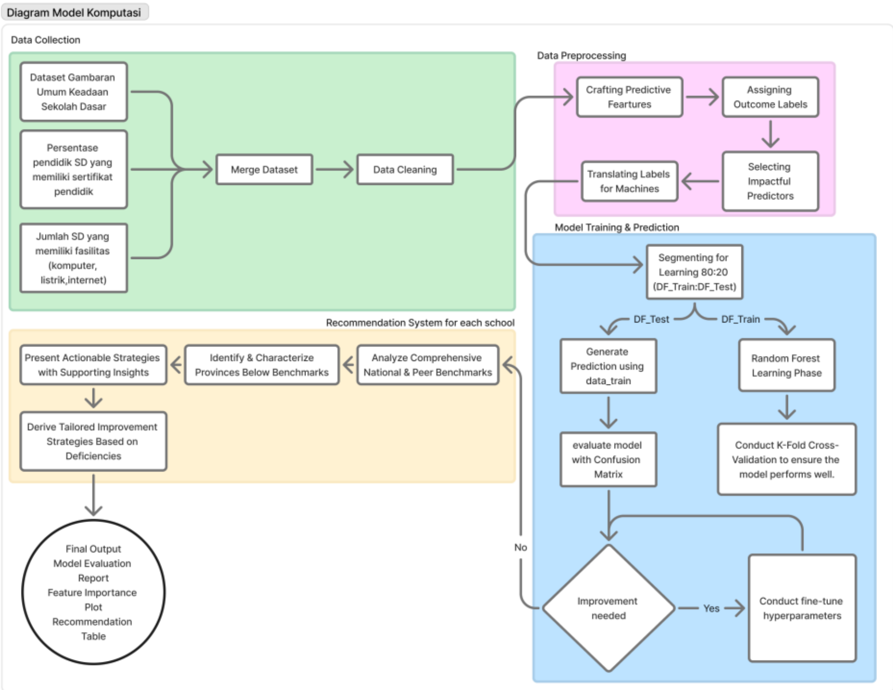

Figure 1. Computing Model

The computational model in this study consists of five main processes: data collection, data preprocessing, modeling, model refinement, and generating school-specific recommendation systems. An overview of the computational model is illustrated in Figure 1.

1. Data Collection

The data used in this study are secondary, aggregated educational data at the provincial level in Indonesia. Sources include official educational data portals and platforms such as Kaggle (e.g., the dataset titled *"kelayakan-pendidikan-indonesia.csv"*). The dataset comprises various educational indicators, such as student enrollment, dropout rates, teacher qualifications, and infrastructure conditions. It is assumed that the dataset has undergone prior validation, with irrelevant or invalid features either ignored or removed. If the data originated from multiple sources, they were merged into a single main dataset.

2. Data Preprocessing

The data preprocessing stage aims to clean, transform, and prepare the dataset for analysis using the Random Forest algorithm. This step is critical for ensuring data quality and achieving accurate model performance. The following procedures were carried out:

2.1 Environment Setup and Data Loading

a. Loading the required libraries in R, such as tidyverse for data manipulation, randomForest for algorithm implementation, readr for reading data, caret for data splitting and model evaluation, knitr for report generation, magrittr for pipe operators, and jsonlite for JS output

b. Loading the main dataset using the read_csv() function from the readr library. A file existence check was also performed

```
df <- read_csv("kelayakan-pendidikan-indonesia.csv")
```

2.2 Data Cleaning and Feature Transformation (Crafting Predictive Features)Data Cleaning and Feature Transformation (Crafting Predictive Features):

a. Entity Name Transformation: The 'Provinsi' column in the dataset was transformed into unique school initials using the generate_school_initials() function for anonymization or standardization of the unit of analysis.

```
generate_school_initials <- function(n) {
  sapply(1:n, function(i) {
    res_letter <- ""
    temp_val <- i
    while(temp_val > 0) {
      temp_val <- temp_val - 1
      res_letter <- paste0(LETTERS[temp_val %% 26 + 1], res_letter)
      temp_val <- floor(temp_val / 26)
    }
    return(paste("Sekolah", res_letter))
  })
}

if (nrow(df) > 0) {
  df$Provinsi <- generate_school_initials(nrow(df))
} else {
  stop("Dataset kosong, tidak dapat menghasilkan inisial sekolah.")
}
```

b. Derived Indicator Construction: Based on the basic variables in the dataset, a number of more informative derived indicators were created for analysis. This process was conducted using the mutate() function from tidyverse.

c. Handling Invalid Values: Inf or NaN values that may result from division by zero or other numerical operations were converted to NA (Not Available).

2.3 Assigning Outcome Labels (Translating Labels for Machines):

a. The target variable "Feasibility" (Accreditation) was determined based on percentiles of the Persentase_putus_sekolah variable.

b. The quantile() and case_when() functions were used to classify schools into five feasibility categories ("A" to "E").Variabel target Kelayakan (Akreditasi) ditentukan berdasarkan persentil dari Persentase_putus_sekolah.

2.4 Final Feature Selection for Modeling:

a. The df_model_input dataset was prepared by selecting relevant predictor features and the target variable.

b. Rows containing NA values were removed, and the target variable was converted into a factor. Dataset df_model_input disiapkan dengan memilih fitur-fitur prediktor yang relevan dan variabel target.

3. Model Training & Prediction

This stage involves building a classification model using the Random Forest algorithm to predict school feasibility.

3.1 Segmenting for Learning 80:20:

The df_model_input dataset was split into training data (80%) and testing data (20%) using the create Data Partition () function from the caret package.

```
trainIndex <- createDataPartition(df_model_input$Kelayakan, p = 0.8,
list = FALSE, times = 1)
df_train <- df_model_input[trainIndex, ]
df_test <- df_model_input[-trainIndex, ]
```

3.2 Random Forest Learning Phase:

The Random Forest model is trained using training data (df_train) with the Random Forest() function.

```
num_predictors <- ncol(df_train) - 1
mtry_val <- if (num_predictors > 0) min(3, num_predictors) else 1

if (length(unique(df_train$Kelayakan)) > 1 && nrow(df_train) > 0 &&
num_predictors > 0) {
  cat("\nMemulai pelatihan model Random Forest...\n")
  rf_model <- randomForest(Kelayakan ~ .,
                           data = df_train,
                           ntree = 100,
                           mtry = mtry_val,
                           importance = TRUE,
                           na.action = na.omit)
  cat("Pelatihan model selesai.\n")
} else {
  warning("Model Random Forest tidak dilatih karena kondisi data
training tidak memadai.")
  rf_model <- NULL
}
```

3.3 Generate Prediction using data_test:

The trained model is used to predict the test data (df_test) using predic t().

```
if (!is.null(rf_model) && nrow(df_test) > 0) {
  cat("\nMelakukan prediksi pada data test...\n")
  rf_predictions <- predict(rf_model, newdata = df_test)
  cat("Prediksi selesai.\n")
```

```
} else {
  rf_predictions <- NULL
  warning("Prediksi tidak dilakukan karena model tidak dilatih atau
data test kosong.")
}
```

3.4 Evaluate model with Confusion Matrix:
The performance of the model was evaluated using the confusion matrix of the caret package. The importance of a feature is analyzed with importance ().

3.5 Conduct K-Fold Cross-Validation, Conduct fine-tune hyperparameters:
If the model evaluation shows unsatisfactory performance ("Improvement needed? -> Yes"), steps such as K-Fold Cross-Validation and fine-tuning hyperparameters can be performed. This logic is important in the model development cycle. If the performance is good ("Improvement needed? -> No"), then proceed to the next stage.

C. Results and Discussion

1. Results

The result of this study is the presentation of the results of analysis and recommendations. The main output includes the Recommendation System for each school. Once the model has been evaluated and important features identified, the next stage is to build a recommendation system. This process involves: Analyze Comprehensive National & Peer Benchmarks, Identify & Characterize Provinces Below Benchmarks, i.e. calculating the national median for each key numerical indicator, for each school, the value of the indicator compared to the national median. At this stage, a comparison is made between the values of education indicators in each school or province with the national median value. The indicators compared include dropout rates, the ratio of teachers with S1 qualifications, the quality of classroom infrastructure, and other relevant indicators. The goal is to identify areas that are below national standards and require special attention. Furthermore, the preparation of Data-Based Recommendations (Derive Tailored Improvement Strategies Based on Deficiencies, Present Actionable Strategies with Supporting Insights), detailed recommendations (Rekomendasi_Detail) are prepared automatically for each school.

```
rekomendasi_data <- df_rekomendasi_input %>%
  rowwise() %>%
  mutate(
    Rekomendasi_Detail = list({
      rekomendasi_items <- list()
      deviasi_putus_sekolah <- Persentase_putus_sekolah -
feature_medians$Persentase_putus_sekolah
      rekomendasi_items <- add_rekomendasi(rekomendasi_items,
                                           !is.na(Persentase_putus_sekolah)
&& Persentase_putus_sekolah > feature_medians$Persentase_putus_sekolah,
                                           "Persentase Putus Sekolah",
Persentase_putus_sekolah, feature_medians$Persentase_putus_sekolah,
                                           paste("Tingkat putus sekolah (",
round(Persentase_putus_sekolah, 2), "%) lebih tinggi ",
round(deviasi_putus_sekolah, 2), "% dari median nasional (",
round(feature_medians$Persentase_putus_sekolah, 2), "%).", sep=""),
                                           "Fokus pada program intervensi
untuk mencegah siswa putus sekolah, identifikasi faktor penyebab utama di
tingkat lokal, dan perkuat dukungan bagi siswa berisiko."
      )
      current_kelayakan <- Kelayakan
      if (length(rekomendasi_items) == 0 && !is.na(current_kelayakan)) {
      } else if (length(rekomendasi_items) == 0 && is.na(current_kelayakan))
{
```

```
      }
      rekomendasi_items
    })
  ) %>%
  ungroup() %>%
  select(Provinsi, Kelayakan,
         P_Mengulang = Persentase_mengulang,
         P_Putus_Sekolah = Persentase_putus_sekolah,
         P_Guru_S1 = Persentase_guru_S1,
         Rasio_Siswa_Guru = Performa_guru,
         Siswa_Per_Rombel = Siswa_rombel,
         P_Kualitas_Ruang = Persentase_kualitas_ruang,
         Rekomendasi_Detail)
```

Based on the results of the comparison, the system automatically generates customized recommendations for each school or province. For example, if a region has a dropout rate that is higher than the national median, the system will provide suggestions such as: "The out-of-school rate in this region is higher than the national average. It is recommended to strengthen intervention programs for students at risk of dropping out, identify local causes, and improve learning support." This kind of recommendation is prepared individually for each educational entity, taking into account deviations or differences from ideal values. In addition to compiling recommendations, the system also generates model evaluation reports that include information on the classification accuracy, sensitivity, specificity, and relative contribution of each variable to the prediction results. This information is presented in the form of tables and visualizations to make it easier for policymakers or end users to understand. To complement that, each indicator used in the analysis is also given its own description, so that readers can better understand the meaning of each data and its interpretation.

All final results, including model recommendations and evaluations, are packaged in JSON digital format. This format was chosen because it is flexible and easy to use in various technology applications, such as dashboards, education monitoring systems, or web-based reporting.

```
deskripsi_indikator <- list(
  list(nama_kolom = "Provinsi", deskripsi = "Nama inisial unik...",
interpretasi_nilai = "..."),
  # ... (semua deskripsi indikator seperti di UTS.R) ...
  list(nama_kolom = "Rekomendasi_Detail.saran", deskripsi = "Saran
tindakan...", interpretasi_nilai = "Netral")
)

output_json_final <- list(
  deskripsi_indikator = deskripsi_indikator,
  data_sekolah = rekomendasi_data
)

json_output <- toJSON(output_json_final, pretty = TRUE, auto_unbox = TRUE, na
= "null")
write(json_output, "data_rekomendasi.json")

cat("\nAnalisis selesai. Hasil evaluasi model (jika dilatih) telah
dicetak.\n")
```

2. Discussion

The findings of the model evaluation not only prove the high level of accuracy, but also open up the space for further analysis of the data, in particular in identifying key variables and regional patterns that affect the feasibility of Education. The Random Forest model used to classify educational eligibility status shows very high performance. Based on the confusion matrix, the model was able to classify all test data correctly, namely 3 provinces in the "Eligible" category and 4 provinces in the "Not Eligible" category, without any classification errors at all. This results in a perfect accuracy value (Accuracy = 1.0), with a Kappa value = 1, which indicates a level of perfect agreement between the model's prediction and reality.

The model performance table further shows a sensitivity and specificity of 1.0000, and positive and negative predictive values also reach 1.0000. With a Balanced Accuracy value = 1.0000, this model can be said to have high performance stability in classifying both classes with balanced proportions. A p-value of 0.01989 indicates that the model's accuracy is statistically significantly better than random guesses (no information rate = 0.5714). In other words, the model is not only empirically accurate, but also statistically valid.

The feature importance analysis reinforces the finding that the Dropout Percentage is the variable with the highest contribution in the model. The Mean Decrease Accuracy value of this feature is 15.49, while its Mean Decrease Gini value is 9.61, far above the other variables. This means that when these features are omitted or their values are randomized, the accuracy of the model decreases significantly, indicating the dominance of its influence in the classification process.

Furthermore, the variables S1 Teacher Percentage and Repeat Percentage also showed a fairly high contribution, with Mean Decrease Accuracy of 4.27 and 3.02, respectively. In contrast, features such as Students per Volume and Space Quality Percentage show low importance values, even negative values in the context of influence on accuracy. This negative value indicates that the presence of such features in the model does not strengthen accuracy, and in some cases actually weakens.

The analysis of provinces classified as “Not Feasible” reveals that weak educational indicators are widely distributed, yet follow a consistent pattern

a. Educational Disparities in Papua and Eastern Indonesia. Provinces such as Papua, South Papua, and Highland Papua exhibit a combination of poor indicators, including the highest dropout rates (up to 0.98), high repetition rates (up to 5.26), low percentages of bachelor-qualified teachers (even below 76%), and extreme student-teacher ratios (more than 28 students per teacher). These conditions indicate an urgent need for holistic interventions that address human resource quality, curriculum relevance, and educational infrastructure.
b. Repetition Rates and Teacher Quality Indicators. West Papua and Gorontalo stand out due to high student repetition rates, at 2.86 and 1.33 respectively, indicating system-level failures in curriculum absorption or fundamental skill gaps among students. In Gorontalo’s case, although the proportion of bachelor-qualified teachers is high (98.15%), the persistence of repetition suggests ongoing challenges in teaching effectiveness.
c. Physical Infrastructure Remains a Challenge. In several provinces such as West Sulawesi (77.7%) and Highland Papua (79.7%), the low percentage of classrooms in good condition reflects weak infrastructure support. The lack of a physically conducive learning environment directly affects student outcomes and motivation.

Supported by the model’s high performance and the strong correlation of selected features with educational status, this study provides a solid foundation for evidence-based policy making. Variables such as dropout rates, repetition, and teacher qualifications should be prioritized in the design of educational interventions, both at the national and regional levels. The model also demonstrates the potential of using machine learning for dynamic monitoring and evaluation of education policies, enabling simulations of various intervention scenarios. Moreover, the findings suggest that educational inequality in Indonesia is not only spatial but also structural in nature, requiring cross-sectoral and inter-ministerial collaboration.

2.1 Implications

This study provides a strong basis for policymakers to develop evidence-based educational interventions. Since it has been proven that reducing dropout rates and improving teacher qualifications (bachelor's degree) are the main factors that determine the quality of education in a region, the main focus should be on these two factors. By using the Random Forest algorithm, school feasibility evaluations can be carried out dynamically and objectively rather than through subjective assessments, which are often inconsistent. In practical terms, the findings of this study emphasize that in order to overcome structural and spatial educational inequalities, resources must be allocated further to eastern Indonesia, such as Papua.

2.2 Research contribution

Methodologically, this study shows that the Random Forest algorithm is highly effective in classifying educational eligibility at the provincial level. Practically, this study helps by building an automated recommendation system that can provide specific recommendations for each educational entity based on the national median. In addition, this study successfully converted raw education data into more useful derivative indicators, providing a new perspective for stakeholders examining large-scale education data in Indonesia.

2.3 Limitations

By using aggregate data at the provincial level, this study cannot describe the variations in conditions between schools within a province comprehensively. In addition, the relatively small amount of test data (approximately 7 samples) allows the model to achieve perfect accuracy. However, to ensure the overall stability of the model, further testing on a larger dataset is required. The percentage of classroom quality is one feature that shows low or even negative importance in this model. This may be due to the quality of the underlying data or the relevance of this feature in binary classification.

2.4 Suggestions

Further research is recommended to use more detailed data, such as school or district level data. To improve the model's resilience to data variation, it is recommended to test the use of additional machine learning algorithms such as XGBoost or Neural Networks. To support digital transformation in public decision-making in the Indonesian education sector, the JSON-based recommendation system generated in this study should be integrated into a real-time education monitoring dashboard.

D. Conclusion

This study demonstrates that a classification approach using the Random Forest algorithm can accurately and significantly differentiate Indonesian provinces based on the feasibility of their educational status, achieving perfect accuracy (100%) and strong statistical validation ($p = 0.01989$). The model successfully identified dropout rate percentage, percentage of bachelor-qualified teachers, and student repetition rate as the most influential variables in determining a region’s educational feasibility. These findings indicate that issues related to access, teacher quality, and instructional effectiveness remain major challenges within the national education system particularly in Eastern Indonesia. The results underscore the importance of integrating data-driven analytic technologies with education policy to produce more targeted and evidence-based interventions.Machine learning approaches such as Random Forest not only perform well in classification tasks but also offer transparency in evaluating the contribution of each variablean essential advantage for designing policies based on prioritized indicators. Therefore, this method can be more widely adopted in the formulation and evaluation of education policies in Indonesia as part of the broader digital transformation in public decision-making.

E. Acknowledgment

The authors would like to express their sincere gratitude to the Department of Computer Science Education, Universitas Pendidikan Indonesia, for providing the academic support and research facilities used in this study. Special thanks are extended to the supervising lecturers LSR, AW, RM, and EN for their invaluable guidance, encouragement, and constructive feedback throughout the research process. The authors also acknowledge the contributions of colleagues and classmates who provided insight and support during the preparation of this manuscript.

F. Author Contribution Statement

FRA and SNF contributed to data collection, preprocessing, and initial model implementation. LSR, AW, RM, and EN provided supervision, guidance on research design, and validation of results. All authors contributed to the discussion of results, manuscript revision, and approved the final version of the manuscript.

References

Apriliansyah, F., Ahmad, M., & Rochimah, H. (2025). Analisis Kesenjangan Pembiayaan Pendidikan Di

Indonesia : Studi Literatur Terhadap Perbedaan Akses Dan Kualitas Antara Sekolah Negeri Dan Swasta Di Daerah Perkotaan Dan Pedesaan. *Jurnal Manajemen Pendidikan*, *10*(2), 753–758. https://doi.org/10.34125/jmp.v10i2.651

Ashari, N. R., Gunawan, S., Irwan, W., & Kunci, K. (2024). Pengaruh Pendidikan, Pengangguran , dan Kemiskinan terhadap Ketimpangan Pendapatan di Indonesia. *Jurnal Intelektualitas: Keislaman, Sosial, Dan Sains*, *13*(2), 293–302. https://doi.org/10.19109/intelektualita.v13i2.25139

Astuti, P. Y., & Diantoro, F. (2021). Evaluasi sekolah dan madrasah melalui sistem akreditasi dalam kebijakan pendidikan di indonesia. *Jurnal Kajian Kependidikan Islam*, *6*(2), 149–164. https://doi.org/10.22515/attarbawi.v6i2.4282

Delima, Suci, A., & Hasibuan, E. I. (2025). Kontribusi Pendidikan terhadap Kualitas Human Capital dan Dampaknya pada Indeks Pembangunan Manusia menghasilkan sumber daya manusia (SDM) yang kompeten , kreatif , dan berdaya saing global , budaya , sehingga pendidikan harus dirancang untuk membentuk. *Jurnal Pendidikan Dan Kewarganegara Indonesia*, *2*(4), 181–191. https://doi.org/10.61132/jupenkei.v2i4.887

Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. *Education and Information Technologies*, *29*, 22125–22145. https://doi.org/10.1007/s10639-024-12704-0

Gustiani, F., Barus, I. C., & Elfrianto. (2023). Monitoring and Evaluation to Improve School Quality. *Indonesian Journal Education*, *2*(4), 99–103. https://doi.org/10.30596/jcositte.v1i1.xxxx

Jin, R., Peng, Y., Wang, Z., Wang, J., Tang, J., & Zhang, M. (2025). Data-Driven Educational Decision-Making : How to Enhance Educational Quality and Management Efficiency. *Journal of Higher Education Research*, *50*(6), 550–553. https://doi.org/10.32629/jher.v5i6.3385

Kharis, S. A. A., & Zili, A. H. A. (2022). Learning Analytics dan Educational Data Mining pada Data Pendidikan. *Jurnal Riset Pembelajaran Matematika Sekolah*, *6*(1), 12–20. https://doi.org/10.21009/jrpms.061.02

Luo, Y. (2023). Identifying Factors Influencing China Junior High Students ' Cognitive Ability through Educational Data Mining : Utilizing LASSO , Random Forest , and XGBoost. *Proceedings of ICMEIM*. https://doi.org/10.4108/eai.8-9-2023.2340189

Madhakomala, R., Ayu, P. P., Septiani, T., & Butarbutar, P. H. (2025). Education for All: A Systematic Analysis of Equity Policies and Challenges in Indonesia. *International Journal Of Humanities Education and Social Sciences*, *4*(4), 1706–1717. https://doi.org/10.55227/ijhess.v4i4.1453

Muchtar, C. O., Purba, A. A., Sintya, J., Siagian, M. F., Harahap, N. M., Ginting, R. A., Bangun, M. B., & Muchtar, Z. (2025). Systematic Literature Review : Examining Indonesia ' s Educational Inequality Factors and Government Equity Policies. *International Journal of Educational Practice and Policy*, *3*(1), 8–16. https://doi.org/10.61220/ijepp.v3i1.0257

Nasution, R. A., Fauzi, F., Harahap, I., & Anggara, W. (2025). The Role Of Education In Human Resource Development In The Context Of Economic Development : A Review. *Proceeding International Seminar on Islamic Studies*, *6*(1), 2028–2035. https://doi.org/10.3059/insis.v0i0.23615

Panoyo. (2024). Pendidikan sebagai Pilar Pembangunan Bangsa Indonesia Menuju Masa Depan Berkelanjutan. *Jurnal Studi Pendidikan Dasar*, *2*(1), 33–44. https://doi.org/10.54180/jsped.v2i1.499

Paolucci, C., Vancini, S., Bex, R. T., Cavanaugh, C., Salama, C., & Araujo, Z. De. (2024). Heliyon Review article A review of learning analytics opportunities and challenges for K-12 education. *Heliyon*, *10*(4), e25767. https://doi.org/10.1016/j.heliyon.2024.e25767

Putra, M., & Harahap, E. (2024). Machine Learning pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Random Forest. *Jurnal Riset Matematika (JRM)*, *4*(3), 127–136. https://doi.org/10.29313/jrm.v4i2.5102

Ramadhan, A., & Susetyo, B. (2021). Classification Modelling of Random Forest to Identify the Important Factors in Improving the Quality of Education. *Int. J. Adv. Sci. Eng. Inf. Technol*, *11*(2), 501–507. https://doi.org/10.18517/ijaseit.11.2.8878

Ramadhan, A., Susetyo, B., & Indahwati. (2019). Penerapan Metode Klasifikasi Random Forest Dalam Mengidentifikasi Faktor Penting Penilaian Mutu Pendidikan Application Of Random Forest Classification Method In Identi- Fying Important Factors Of Educational Quality. *Jurnal Pendidikan Dan Kebudayaan*, *4*(2), 169–182. https://doi.org/10.24832/jpnk.v4i2.1327

Safrizal, Yulia, R., & Jumiarti, D. (2022). School readiness analysis to implement an inclusive education policy : A case study at elementary school. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *26*(1), 1–11. https://doi.org/10.21831/pep.v26i1.45294

Saputri, R. A., Asrianda, & Rosnita, L. (2025). A Random Forest-Based Predictive Model for Student Academic Performance : A Case Study in Indonesian Public High Schools. *Journal of Applied*

Informatics and Computing, *9*(3), 1042–1049. https://doi.org/10.30871/jaic.v9i3.9460

Sunarya, P. A., Asri, M., Azizah, N., & Lim, C. P. (2025). Evaluasi Sistem Informasi Pendidikan untuk Pengelolaan Data dan Keputusan Evaluation of Educational Information Systems for Data and Decision. *Jurnal MENTARI: Manajemen, Pendidikan Dan Teknologi Informasi*, *3*(2), 118–126. https://doi.org/10.33050/mentari.v3i2.738

Syofyan, H., & Rosyid, A. (2024). Mengukur Kesiapan Sekolah Untuk Meningkatkan Efektifitas Penerapan Kurikulum Merdeka Di Sekolah Dasar. *Jurnal Ilmiah Pendidikan Dasar*, *09*(03), 268–293. https://doi.org/10.23969/jp.v9i3.18765

Wahidin, & Hulbat, R. (2022). Pendekatan Evaluasi Berbasis Data Dan Teknologi Untuk Meningkatkan Kualitas Pendidikan. *PINISI Discretion Review*, *6*(1), 621–629. https://doi.org/10.26858/pdr.v6i1.36986

Yağcı, M. (2022). Educational data mining : prediction of students ' academic performance using machine learning algorithms. *Smart Learn. Environ*, *9*(11), 1–19. https://doi.org/10.1186/s40561-022-00192-z

Zawiyah, S., Qodriyah, L., & Tamam, M. B. (2022). Klasifikasi Prestasi Akademik Mahasiswa Menggunakan Metode Random Forest. *Journal of Digital Business and Information Technology*, *1*(2), 61–71. https://doi.org/10.23971/jobit.v1i2.317

Zhang, J., & Fan, Y. (2025). Machine Learning Based Big Data Analytics for Education in Curriculum Reforms. *Applied Mathematics and Nonlinear Sciences*, *10*(1), 1–13. https://doi.org/10.2478/amns-2025-0135