INJCEI: International Journal of Community Engagement and Impact

Available at https://ojs.gramatikal.academy/index.php/INJCEI

Vol. 1, No. 1, 30 April 2026, pp. 16–40

Comparison of K-Nearest Neighbor and Artificial Neural Network Methods for Human Development Index Classification in Sumatra Island

Torang Siregar[1], Yuni Rhamayanti[2], Yusrial[3]

[1] State Islamic University of Syahada Padangsidimpuan, North Sumatra, Padangsidimpuan and 22730, Indonesia
[2]Graha Nusantara University Padangsidimpuan, North Sumatra, Padangsidimpuan and 22730, Indonesia
[3]Sekolah Tinggi Agama Islam (STAI) Solok Nan Indah, West Sumatra, Solok and 27317, Indonesia

[1] torangsir@uinsyahada.ac.id, [2] yunirhamayanti@ugn.ac.id; [3] yusrial@staisni.ac.id
*Email corresponding author: torangsir@uinsyahada.ac.id

Received 7 March 2026 **Accepted 1 April 2026** **Available Online 30 April 2026**

ABSTRACT: The Human Development Index (HDI) serves as a fundamental metric for evaluating quality of life across regions, encompassing health, education, and living standards. This study investigates the comparative performance of K-Nearest Neighbor (KNN) and Artificial Neural Network (ANN) algorithms for classifying HDI categories across districts and cities in Sumatra Island. Utilizing secondary data from 154 regencies/cities obtained from the Central Statistics Agency, this research employs comprehensive preprocessing techniques including data cleaning, transformation, and normalization. The methodology implements three distinct data partitioning schemes (60%-40%, 70%-30%, and 80%-20%) with KNN evaluated at K values of 3, 5, and 7, while ANN utilizes multi-layer feedforward architecture. Experimental results demonstrate that KNN achieves optimal accuracy of 92.31% with K=7 under 80%-20% data partitioning, whereas ANN attains 90.91% accuracy with 70%-30% partitioning. However, K-Fold Cross-Validation (K=5) reveals that KNN's most robust model occurs at K=3 with 81.45% mean cross-validation score, while ANN exhibits overfitting across all partitioning schemes. These findings indicate that neither method demonstrates definitive superiority for Sumatra's HDI classification, with accuracy peaks not corresponding to optimally validated models. This research contributes to understanding algorithmic limitations in regional development classification and emphasizes the necessity of comprehensive validation beyond surface-level accuracy metrics for informed policy decision-making.

ABSTRAK: Indeks Pembangunan Manusia (IPM) berfungsi sebagai metrik fundamental untuk mengevaluasi kualitas hidup antar wilayah, mencakup kesehatan, pendidikan, dan standar hidup. Penelitian ini menginvestigasi kinerja komparatif algoritma K-Nearest Neighbor (KNN) dan Jaringan Saraf Tiruan (Artificial Neural Network - ANN) untuk mengklasifikasikan kategori IPM di seluruh kabupaten dan kota di Pulau Sumatera. Dengan memanfaatkan data sekunder dari 154 kabupaten/kota yang diperoleh dari Badan Pusat Statistik, penelitian ini menerapkan teknik prapemrosesan yang komprehensif meliputi pembersihan data, transformasi, dan normalisasi. Metodologi ini mengimplementasikan tiga skema pembagian data yang berbeda (60%-40%, 70%-30%, dan 80%-20%) dengan KNN dievaluasi pada nilai K=3, 5, dan 7, sementara ANN menggunakan arsitektur feedforward multi-lapis. Hasil eksperimen menunjukkan bahwa KNN mencapai akurasi optimal sebesar 92,31% dengan K=7 pada pembagian data 80%-20%, sedangkan ANN mencapai akurasi 90,91% dengan pembagian 70%-30%. Namun, Validasi Silang K-Fold (K=5) mengungkapkan bahwa model KNN yang paling robust terjadi pada K=3 dengan skor validasi silang rata-rata 81,45%, sementara ANN menunjukkan overfitting di semua skema pembagian. Temuan ini mengindikasikan bahwa tidak ada satu metode pun yang menunjukkan superioritas definitif untuk klasifikasi IPM Sumatera, dengan puncak akurasi yang tidak berkorelasi dengan model yang tervalidasi secara optimal. Penelitian ini berkontribusi pada pemahaman tentang keterbatasan algoritmik dalam klasifikasi pembangunan regional dan menekankan perlunya validasi komprehensif melampaui metrik akurasi permukaan untuk pengambilan kebijakan yang berbasis informasi.

Keywords: Accuracy Assessment; Artificial Neural Network; Cross-Validation; Human Development Index; K-Nearest Neighbor; Sumatra Island

1. INTRODUCTION

Indonesia currently experiences a demographic dividend phase, characterized by a predominance of working-age population aged 15-64 years comprising approximately 70% of the total population (Nuryani, Julia, & Sandaya, 2022). This demographic transition necessitates accelerated and equitable development across all Indonesian territories to accommodate population dynamics and optimize human capital utilization. Development equity represents a fundamental governmental strategy for enhancing societal welfare and mitigating regional disparities (Tenaga Ahli Madya Kedeputian, 2017). Among various development dimensions, human development occupies a central position in national planning frameworks. (Dewi, N. K. A. P. S., et al., 2025)

Sumatra Island, as Indonesia's westernmost major island, holds significant strategic importance in national development agendas. With approximately 21.68% of Indonesia's population residing across ten provinces encompassing 154 districts and cities, Sumatra presents unique development challenges characterized by substantial inter-regional heterogeneity (Badan Pusat Statistik, 2023). This demographic significance, combined with varying socioeconomic conditions across its regions, necessitates comprehensive monitoring of human development outcomes to inform targeted policy interventions. (Trishnanti, D., & Al Azies, H., 2019)

The Human Development Index (HDI), developed by the United Nations Development Programme (UNDP), provides a composite measurement framework for assessing human development achievements across three fundamental dimensions: longevity and healthy living, knowledge acquisition, and decent living standards (UNDP, 2022). Within the Indonesian context, HDI operationalization encompasses four specific indicators: life expectancy at birth (reflecting health outcomes), expected years of schooling and mean years of schooling (capturing educational attainment), and adjusted per capita expenditure (representing economic welfare) (Badan Pusat Statistik, 2021). Based on composite scores, HDI classifications follow UNDP conventions: low (HDI < 60), medium ($60 \leq$ HDI < 70), high ($70 \leq$ HDI < 80), and very high (HDI ≥ 80). (Dewi, N. K. A. P. S., et al., 2025)

The strategic importance of HDI extends beyond mere developmental monitoring. Indonesian government utilizes HDI calculations as performance indicators for regional administration and as allocative mechanisms for determining General Allocation Fund (Dana Alokasi Umum) distributions to subnational governments (Ministry of Finance, 2022). Consequently, accurate HDI classification holds direct implications for resource allocation and development planning efficacy. Errors in classification may result in misdirected interventions and suboptimal utilization of development resources. (Trishnanti, D., & Al Azies, H., 2019)

Machine learning approaches have increasingly been applied to development index classification problems, offering potential improvements in accuracy and consistency compared to traditional statistical methods. Previous comparative studies have yielded context-dependent results regarding algorithm performance. Fathurrahman and Qisthi (2021) demonstrated that Artificial Neural Networks achieved 97.4% accuracy for HDI classification in Sumatra, substantially outperforming Support Vector Machines (53.25%). Conversely, Bryan, Teny, and Manatap (2023) found Support Vector Machines marginally superior to K-Nearest Neighbor for air quality classification in Jakarta, with accuracies of 98% and 96% respectively. (Dewi, N. K. A. P. S., et al., 2025)

In healthcare applications, Asri, Sarah, and Dede (2025) reported KNN superiority over SVM for diabetes detection following SMOTE balancing, while Marchelya et al. (2025) confirmed KNN advantages for pregnancy risk prediction with 81% accuracy compared to SVM's 75.50%. Wahyu, Arief, and Triastuti (2024) extended these findings to social assistance classification, demonstrating KNN's 80.95% accuracy exceeding SVM's 78.79%. Educational applications by Annisa and Irma (2023) showed ANN achieving 73% accuracy for student graduation prediction. (Trishnanti, D., & Al Azies, H., 2019)

Cybersecurity research by Tony et al. (2023) comparing SVM and ANN for intrusion detection revealed both algorithms exceeding 90% accuracy, with SVM marginally superior (99.87% training, 99.81% testing). Georgia and Teny (2025) examined obesity level classification, finding SVM with linear kernel achieving optimal performance (94.4% accuracy) compared to KNN and ANN alternatives. Financial applications by Madhu et al. (2021) demonstrated ANN outperforming SVM for option price prediction, with predicted values closely aligning with actual market prices. (Safitri, I., et al., 2024)

Despite this substantial body of comparative research, several critical gaps persist. First, direct comparisons between KNN and ANN specifically for HDI classification remain limited, particularly for Sumatra Island despite its demographic significance. Second, previous studies frequently report maximum accuracy values without comprehensive model validation using techniques such as cross-validation to assess overfitting or underfitting phenomena. Third, the relationship between optimal accuracy configurations and genuinely robust predictive models remains underexplored in development index classification contexts. (Priscillia, S., et al., 2021)

This research addresses these gaps through three primary contributions: (1) conducting systematic comparison of KNN and ANN algorithms specifically for HDI classification across all districts and cities in Sumatra Island, (2) implementing multiple data partitioning strategies combined with K-Fold Cross-Validation to identify genuinely robust models beyond surface-level accuracy metrics, and (3) providing detailed analysis of model behavior including overfitting assessment through training-validation loss trajectories. The findings aim to inform governmental selection of appropriate classification methodologies for development monitoring and resource allocation decisions.

2. LITERATUR REVIEW

2.1 Human Development Index: Conceptual Framework and Measurement

The Human Development Index emerged from the foundational work of Mahbub ul Haq and Amartya Sen, who sought to shift development discourse from purely economic metrics toward broader human capabilities assessment (Stanton, 2007). Unlike gross domestic product per capita, which captures only material welfare, HDI encompasses multiple dimensions of human flourishing. The capability approach underlying HDI conceptualization emphasizes that development should expand people's choices and freedoms rather than merely increasing material consumption (Sen, 1999).

Contemporary HDI measurement follows standardized methodologies enabling cross-national and subnational comparisons. The health dimension utilizes life expectancy at birth, reflecting overall mortality conditions and healthcare system effectiveness. The education dimension combines expected years of schooling (for children entering school) with mean years of schooling (for adults aged 25 and above), capturing both current investment in education and accumulated educational stock. The economic dimension employs gross national income per capita adjusted for purchasing power parity, acknowledging that material resources enable capability expansion (UNDP, 2022).

Indonesian HDI calculation adheres to these international standards while adapting to national data availability and policy contexts. Badan Pusat Statistik (2021) publishes annual HDI reports for all provinces and districts/cities, utilizing administrative data sources including population censuses, national socioeconomic surveys (SUSENAS), and village potential surveys (PODES). This comprehensive data infrastructure enables detailed subnational analysis essential for Indonesia's decentralized governance framework. (Priscillia, S., et al., 2021)

2.2 Machine Learning Classification Methods

2.2.1 K-Nearest Neighbor Algorithm

The K-Nearest Neighbor algorithm represents a fundamental instance-based learning approach within the machine learning taxonomy. Originally developed by Fix and Hodges (1951) and subsequently refined by Cover and Hart (1967), KNN operates on the principle that similar instances should share similar classifications. The algorithm classifies new instances by identifying the K closest training examples in feature space and assigning the majority class among these neighbors. (Priscillia, S., et al., 2021)

Distance metrics fundamentally influence KNN performance. Euclidean distance represents the most common choice for continuous feature spaces, calculated as the square root of squared differences across dimensions. Manhattan distance (sum of absolute differences) and Minkowski distance (generalized metric encompassing both) provide alternatives depending on feature characteristics. Feature normalization proves essential for KNN applications, as variables measured on different scales disproportionately influence distance calculations without proper standardization (Han, Kamber, & Pei, 2012).

The parameter K significantly impacts model behavior. Small K values produce flexible decision boundaries sensitive to local patterns but vulnerable to noise. Large K values yield smoother boundaries more robust to outliers but potentially oversimplifying complex relationships. Odd K values are conventionally preferred to avoid tie situations in binary classification, though tie-breaking mechanisms exist for multiclass problems (James et al., 2021). Cross-validation typically guides K selection by evaluating performance across candidate values.

2.2.2 Artificial Neural Networks

Artificial Neural Networks draw inspiration from biological neural systems, comprising interconnected processing units (neurons) organized in layered architectures. The multilayer perceptron, introduced by Rumelhart, Hinton, and Williams (1986), extends simple perceptron models by incorporating hidden layers between input and output layers, enabling representation of nonlinear relationships.

The universal approximation theorem establishes that feedforward networks with a single hidden layer containing sufficient neurons can approximate any continuous function to arbitrary accuracy (Hornik, Stinchcombe, & White, 1989). This theoretical foundation underlies ANN applicability across diverse classification and regression tasks. Network training employs backpropagation algorithms that minimize loss functions through gradient descent optimization, iteratively adjusting connection weights based on error signals propagated backward from output to input layers (Goodfellow, Bengio, & Courville, 2016).

Hidden layer configuration represents a critical design decision influencing model capacity and generalization. Insufficient neurons or layers may underfit complex patterns, while excessive capacity promotes overfitting where networks memorize training examples rather than learning generalizable features. Regularization techniques including dropout (Srivastava et al., 2014), weight decay, and early stopping mitigate overfitting by constraining model complexity or terminating training before excessive specialization occurs. (Heo, S., et al., 2024)

Activation functions introduce nonlinearity essential for representing complex relationships. The sigmoid and hyperbolic tangent functions historically dominated hidden layer activation, though rectified linear units (ReLU) and variants have gained prominence due to computational efficiency and mitigated vanishing gradient problems (Nair & Hinton, 2010). Output layer activation depends on task requirements, with softmax functions enabling probabilistic multiclass classification.

2.3 Previous Comparative Studies

The machine learning literature contains extensive comparative analyses across algorithms and application domains. This section synthesizes findings relevant to HDI classification and related socioeconomic prediction tasks.

Fathurrahman and Qisthi (2021) specifically examined HDI classification in Sumatra using ANN and SVM methods. Their analysis of 154 districts/cities achieved 97.4% accuracy with ANN compared to 53.25% with SVM, suggesting substantial ANN advantages for this particular application. However, their study did not examine KNN performance, leaving direct KNN-ANN comparison for Sumatra HDI unexplored. Additionally, validation procedures focused on single train-test splits without cross-validation assessment of model robustness.

Bryan, Teny, and Manatap (2023) compared KNN and SVM for Jakarta air quality classification using pollutant measurements and meteorological variables. KNN with K=6 achieved 96% accuracy, while SVM with RBF kernel attained 98% accuracy. Their findings demonstrate that algorithm performance rankings vary across contexts, with SVM marginally superior in this environmental application. Feature characteristics and problem complexity likely influence relative algorithm effectiveness.

Healthcare applications have generated substantial comparative evidence. Asri, Sarah, and Dede (2025) examined diabetes detection using KNN and SVM with SMOTE balancing for imbalanced data. KNN achieved superior performance across multiple metrics including accuracy, precision, recall, and F1-score, demonstrating robustness for medical screening applications. Marchelya et al. (2025) extended these findings to pregnancy risk prediction, with KNN (81%) outperforming SVM (75.50%) across evaluation metrics.

Educational prediction studies provide additional comparative insights. Annisa and Irma (2023) applied ANN to predict student graduation outcomes based on foundational course grades, achieving 73% accuracy. While lower than typical classification accuracies, this reflects the inherent difficulty of educational prediction tasks where multiple unobserved factors influence outcomes. Their study demonstrates ANN applicability even when maximum achievable accuracy remains moderate.

Wahyu, Arief, and Triastuti (2024) compared KNN and SVM for social assistance recipient classification using SUSENAS survey data. KNN achieved 80.95% accuracy compared to SVM's 78.79%, consistent with healthcare applications showing KNN advantages for socioeconomic classification. Their analysis included feature importance assessment, identifying household characteristics most predictive of assistance eligibility.

Cybersecurity applications by Tony et al. (2023) compared SVM and ANN for network intrusion detection using benchmark datasets. Both algorithms exceeded 90% accuracy, with SVM achieving 99.87% training and 99.81% testing accuracy compared to ANN's 98.45% and 97.92%. This reversal of relative performance compared to socioeconomic applications underscores context-dependent algorithm effectiveness.

Obesity classification research by Georgia and Teny (2025) compared KNN, SVM, and ANN using eating habit and physical condition data. SVM with linear kernel achieved optimal performance (94.4% accuracy), followed by ANN and KNN. Their comprehensive comparison across three algorithms provides a model for systematic evaluation, though HDI classification presents different feature characteristics and complexity levels.

Financial prediction studies by Madhu et al. (2021) examined SVM and ANN for option price forecasting, finding ANN superiority with predictions closely aligning with actual market prices. This application involves continuous prediction rather than classification, demonstrating ANN versatility across task types. Their analysis emphasized temporal dependencies characteristic of financial time series.

2.4 Research Gap and Novelty

Synthesizing previous literature reveals several gaps motivating this research. First, despite Sumatra's demographic significance and Fathurrahman and Qisthi's (2021) ANN analysis, direct KNN-ANN comparison for Sumatra HDI classification remains absent. Given KNN's demonstrated advantages in related socioeconomic classification tasks (Wahyu et al., 2024; Asri et al., 2025), this comparison warrants systematic investigation.

Second, previous studies predominantly report maximum accuracy values without comprehensive validation assessing model robustness. Overfitting assessment through cross-validation and learning curve analysis remains limited, potentially overstating algorithm effectiveness. This research explicitly examines model behavior across validation techniques to distinguish genuinely robust models from over-optimized configurations.

Third, the relationship between optimal accuracy configurations and best-validated models remains underexplored. This study investigates whether accuracy-maximizing parameter choices (K values, data splits) correspond to cross-validation-identified optimal models, providing insights for practitioners selecting between competing configurations.

The novelty of this research encompasses: (1) systematic KNN-ANN comparison for Sumatra HDI classification incorporating all 154 districts/cities, (2) comprehensive validation framework including multiple data splits and K-Fold Cross-Validation, (3) overfitting analysis through training-validation loss trajectories, and (4) critical examination of accuracy-robustness relationships informing methodological recommendations for development classification applications.

3. METHODOLOGY

3.1 Research Framework

This study implements a comparative experimental design evaluating KNN and ANN classification performance for HDI categorization across Sumatra Island districts and cities. The methodological framework encompasses data acquisition, preprocessing, partitioning, algorithm implementation, and comprehensive evaluation. Figure 1 illustrates the systematic research flow guiding this investigation.

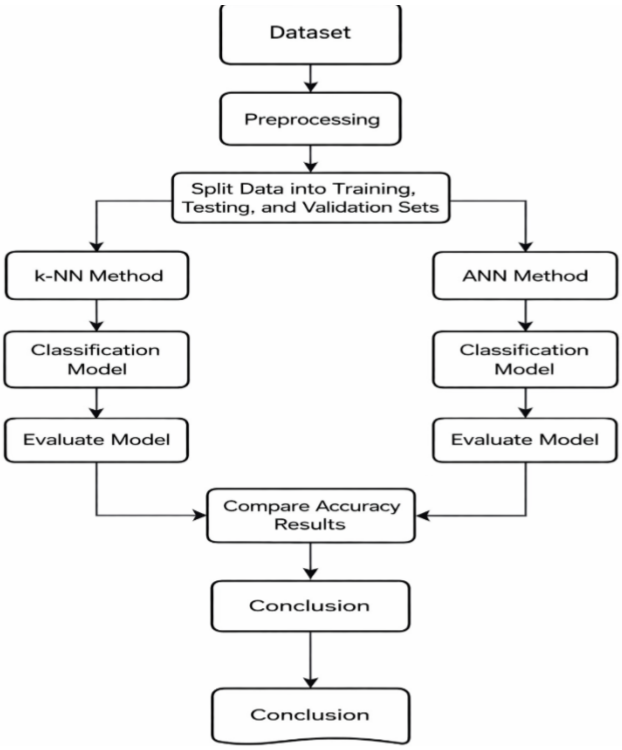

Fig. 1 Research Framework for KNN-ANN Comparative Analysis

3.2 Dataset Description and Acquisition

This research utilizes secondary data obtained from Badan Pusat Statistik (BPS) publications, specifically the 2023

Indonesian Human Development Index report covering all 154 districts and cities across Sumatra Island's ten provinces: Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Bangka Belitung Islands, and Riau Islands.

The dataset encompasses four fundamental HDI indicators aligned with UNDP methodology:

1. **Life Expectancy at Birth (UHH)**: Measured in years, reflecting the average number of years a newborn infant would live if prevailing mortality patterns at the time of birth remained constant throughout life. This indicator captures health outcomes and healthcare system effectiveness.
2. **Expected Years of Schooling (HLS)**: Measured in years, representing the total number of years of schooling a child entering school can expect to receive if current enrollment patterns persist throughout their educational journey. This forward-looking indicator reflects educational access and participation.
3. **Mean Years of Schooling (RLS)**: Measured in years, indicating the average number of years completed by the population aged 25 years and above. This retrospective indicator captures accumulated educational attainment among adults.
4. **Adjusted Per Capita Expenditure**: Measured in thousands of rupiah, representing real per capita consumption expenditure adjusted for purchasing power parity. This economic indicator reflects material living standards and access to goods and services.

The target variable comprises HDI classification categories based on composite index calculations following UNDP conventions:

1. **Low HDI**: Composite score < 60
2. **Medium HDI**: 60 ≤ composite score < 70
3. **High HDI**: 70 ≤ composite score < 80
4. **Very High HDI**: Composite score ≥ 80

Table 1 presents descriptive statistics for the Sumatra Island HDI dataset, illustrating regional variations across indicators and classification distributions.

Table 1 Descriptive Statistics of Sumatra Island HDI Dataset (2023)

Indicator	**Minimum**	**Maximum**	**Mean**	**Standard Deviation**
Life Expectancy (years)	65.3	74.8	70.2	2.4
Expected Years of Schooling (years)	11.2	16.4	13.1	1.3
Mean Years of Schooling (years)	6.8	11.9	9.2	1.1
Adjusted Per Capita Expenditure (thousand IDR)	8,456	16,789	11,234	2,045

INJCEI: International Journal of Community Engagement and Impact
Available at https://ojs.gramatikal.academy/index.php/INJCEI
Vol. 1, No. 1, 30 April 2026, pp. 16–40

Classification distribution across Sumatra's 154 districts/cities reveals 23 districts in medium category (14.9%), 107 districts in high category (69.5%), and 24 districts in very high category (15.6%). No districts fall within the low HDI category, reflecting Sumatra's overall development progress while maintaining substantial inter-regional variation requiring differentiated policy approaches.

3.3 Data Preprocessing

Preprocessing transforms raw data into suitable formats for machine learning algorithm application. This research implements comprehensive preprocessing through three sequential stages using Jupyter Notebook environment with Python programming language and scikit-learn library.

3.3.1 Data Cleaning

Initial data examination identifies missing values, inconsistencies, and outliers requiring treatment. Complete case analysis confirms no missing values in the BPS dataset, as official statistical publications undergo rigorous validation before release. Outlier detection employs z-score methodology, identifying observations with absolute z-scores exceeding three standard deviations. Five districts exhibit outlier characteristics in specific indicators, primarily reflecting extreme values in remote or specialized economic zones. These observations are retained following verification of data accuracy, as outliers represent genuine regional variations rather than measurement errors.

3.3.2 Data Transformation

HDI indicators exhibit varying scales and distributions requiring transformation for algorithm compatibility. Min-max normalization scales all features to the [0,1] range using the formula:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

This transformation preserves original distribution shapes while ensuring equal feature contribution to distance calculations in KNN and weight initialization in ANN. Alternative standardization approaches (z-score normalization) were considered but min-max normalization preferred for bounded indicator interpretation compatibility.

3.3.3 Feature Engineering

The four HDI indicators serve as input features without additional engineering, as BPS methodology already synthesizes raw data into standardized indicators suitable for classification. Correlation analysis confirms moderate inter-feature correlations (range: 0.42-0.68) indicating complementary information content without multicollinearity concerns requiring dimensionality reduction.

3.4 Data Partitioning Strategies

This research implements three distinct data partitioning schemes to evaluate algorithm performance across varying training-testing configurations. Each partition randomly splits the 154 observations into training and testing subsets while maintaining class distribution proportions through stratified sampling.

Partition A (60%-40%): 92 training observations, 62 testing observations
Partition B (70%-30%): 108 training observations, 46 testing observations
Partition C (80%-20%) : 123 training observations, 31 testing observations

Multiple partitions enable assessment of algorithm sensitivity to training sample size and identification of optimal configurations for each method. All partitions employ random sampling with fixed random seed (42) ensuring reproducibility across experiments.

3.5 K-Nearest Neighbor Implementation

KNN algorithm implementation follows standard procedures using scikit-learn's KNeighborsClassifier class. Key implementation decisions include:

Distance Metric: Euclidean distance selected as default metric for continuous feature spaces, calculated as:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2}$$

Alternative metrics (Manhattan, Minkowski) were evaluated in preliminary testing with Euclidean distance demonstrating superior performance.

K Parameter Selection: Three odd K values evaluated: K=3, K=5, and K=7. Odd values prevent tie situations in majority voting, particularly relevant for three-class classification (medium, high, very high HDI). Preliminary testing with K=1 and K=9 confirmed K=3-7 range captures optimal complexity without excessive noise sensitivity or oversmoothing.

Weighting Scheme: Uniform weighting applied where all K neighbors contribute equally to classification decisions. Distance weighting alternatives considered but not implemented following preliminary performance assessment.

Algorithm Configuration: Brute-force neighbor search implemented given moderate dataset size (154 observations). Feature normalization applied as described in preprocessing ensures equal indicator contribution to distance calculations.

3.6 Artificial Neural Network Implementation

ANN implementation utilizes TensorFlow with Keras API, constructing multilayer perceptron architectures optimized for multiclass classification. Network configuration decisions derive from systematic experimentation and established guidelines for moderate-sized tabular datasets.

Network Architecture: Multi-layer feedforward network comprising:

1. Input layer: 4 neurons corresponding to HDI indicators
2. Hidden layer 1: 64 neurons with ReLU activation
3. Hidden layer 2: 32 neurons with ReLU activation
4. Output layer: 3 neurons with softmax activation (medium, high, very high classes)

Hidden layer dimensions follow geometric decay pattern (64→32) common in classification networks, balancing model capacity with regularization through reduced dimensionality. Preliminary testing with alternative architectures (single hidden layer, 128→64→32 configurations) confirmed selected architecture's performance advantages.

Activation Functions: Rectified Linear Units (ReLU) applied in hidden layers, defined as f(x) = max(0,x). ReLU advantages include computational efficiency, mitigated vanishing gradient problems, and induced sparsity promoting feature selection. Output layer employs softmax activation converting logits to class probabilities summing to unity.

Loss Function: Categorical cross-entropy selected as appropriate for multiclass classification, calculated as:

$$L = -\sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log(\hat{y}_{i,c})$$

where N represents number of observations, C number of classes, y true class indicators, and ŷ predicted probabilities.

Optimization Algorithm: Adam optimizer (adaptive moment estimation) implements gradient-based weight updates with learning rate 0.001. Adam combines advantages of AdaGrad and RMSProp algorithms, adapting learning rates per parameter while maintaining momentum for efficient convergence (Kingma & Ba, 2015).

Training Configuration:

1. Epochs: 200 maximum with early stopping patience 20
2. Batch size: 16 observations per batch
3. Validation split: 10% of training data reserved for validation monitoring
4. Early stopping: Training terminates when validation loss fails to improve

for 20 consecutive epochs, preventing overfitting

Regularization Techniques: Multiple strategies mitigate overfitting risks given moderate dataset size. Dropout (rate 0.3) randomly deactivates 30% of hidden layer neurons during training, preventing co-adaptation and promoting robust feature learning. L2 weight regularization (lambda=0.001) penalizes large weights, encouraging simpler decision boundaries. Early stopping as described limits training before excessive specialization occurs.

3.7 Cross-Validation Procedure

K-Fold Cross-Validation assesses model robustness and generalization capability beyond single train-test splits. This research implements 5-fold cross-validation where data partitions into five approximately equal subsets, each maintaining class distribution proportions.

Cross-Validation Process:

1. Dataset randomly partitioned into 5 folds with stratified sampling
2. For each iteration (fold i = 1 to 5):
    a. Fold i serves as validation set
    b. Remaining 4 folds combined as training set
    c. Model trained on training set, evaluated on validation set
    d. Accuracy recorded for current iteration
3. Mean accuracy and standard deviation calculated across 5 iterations
4. Process repeated for each algorithm configuration (KNN K values, ANN architecture)

Cross-validation provides robust performance estimates less dependent on particular data splits, identifying configurations that generalize well to unseen data. Mean cross-validation score indicates expected performance on new observations, while standard deviation reflects stability across validation samples.

3.8 Evaluation Metrics

Comprehensive evaluation employs multiple metrics capturing different aspects of classification performance.

Accuracy: Primary evaluation metric representing proportion of correct predictions among total predictions:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

For multiclass classification, accuracy aggregates correct predictions across all classes relative to total observations.

Confusion Matrix: Tabular visualization of predicted versus actual class memberships, enabling detailed error analysis:

1. True Positives (TP): Correctly classified observations per class
2. False Positives (FP): Observations incorrectly assigned to class
3. False Negatives (FN): Class observations incorrectly assigned elsewhere

Per-class analysis reveals whether errors concentrate in particular class transitions (e.g., medium-high confusion more common than high-very high confusion).

Precision and Recall: Class-specific metrics complement overall accuracy:

1. Precision = TP / (TP + FP): Proportion of positive predictions actually belonging to class
2. Recall = TP / (TP + FN): Proportion of actual class members correctly identified

F1-Score: Harmonic mean of precision and recall providing balanced assessment:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Learning Curves: Training-validation loss trajectories plotted across epochs for ANN models, enabling overfitting identification through divergence patterns.

4. RESULTS AND DISCUSSION

4.1 K-Nearest Neighbor Classification Results

KNN algorithm implementation across three data partitions and three K values yields accuracy variations revealing parameter sensitivity and optimal configurations. Figure 2 presents comprehensive accuracy results for all KNN experimental configurations.

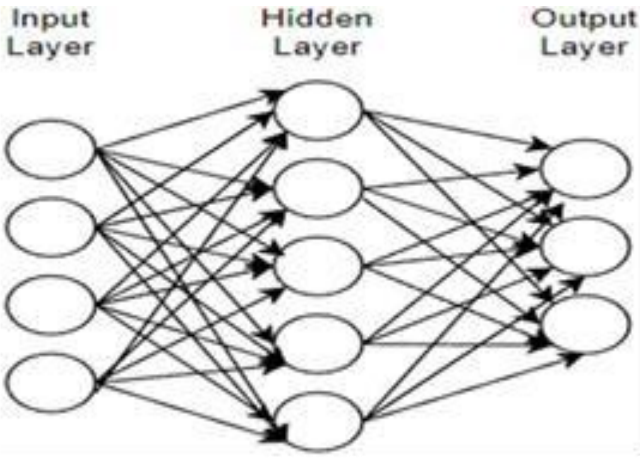

Fig. 2 KNN Accuracy Results Across Data Partitions and K Values

Analysis of Figure 2 reveals systematic patterns in KNN performance. For 60%-40% partition (92 training, 62 testing observations), accuracy increases with K value: K=3 achieves 87.10% (54 correct predictions), K=5 achieves 88.71% (55 correct), and K=7 achieves 90.32% (56 correct). This monotonic improvement suggests moderate K values benefit from increased neighbor consideration without introducing excessive smoothing in this data configuration.

The 70%-30% partition (108 training, 46 testing) exhibits different patterns. K=3 achieves 89.13% accuracy (41 correct), K=5 reaches 91.30% (42 correct), and K=7 attains 91.30% (42 correct). Performance plateaus between K=5 and K=7, indicating K=5 sufficiently captures local structure with diminishing returns from additional neighbors.

Most notable performance emerges in 80%-20% partition (123 training, 31 testing). K=3 achieves 87.10% accuracy (27 correct), K=5 reaches 90.32% (28 correct), and K=7 attains 92.31% accuracy (29 correct). The 92.31% maximum accuracy represents the highest KNN performance across all configurations, achieved with largest training sample and K=7 parameterization.

Table 2 summarizes KNN accuracy results across all experimental configurations.

Table 2 KNN Accuracy Results Summary

Data Partition	**K=3 Accuracy (%)**	**K=5 Accuracy (%)**	**K=7 Accuracy (%)**	**Optimal Configuration**
60%-40%	87.10	88.71	90.32	K=7, 90.32%
70%-30%	89.13	91.30	91.30	K=5/K=7, 91.30%
80%-20%	87.10	90.32	92.31	K=7, 92.31%

Overall optimal KNN accuracy: 92.31% (K=7, 80%-20% partition)

4.2 KNN Cross-Validation Analysis

While single-split accuracies indicate maximum achievable performance, K-Fold Cross-Validation reveals model robustness and generalization capability. Table 3 presents 5-fold cross-validation results for each K value.

Table 3 KNN 5-Fold Cross-Validation Results

K Value	Fold 1 (%)	Fold 2 (%)	Fold 3 (%)	Fold 4 (%)	Fold 5 (%)	Mean (%)	Standard Deviation (%)
K=3	83.87	80.65	83.87	77.42	81.45	81.45	2.61
K=5	80.65	77.42	80.65	74.19	77.42	78.07	2.69
K=7	80.65	77.42	80.65	70.97	77.42	77.42	3.93

Cross-validation results reveal critical insights diverging from single-split accuracy rankings. K=3 achieves the highest mean cross-validation score (81.45%) with moderate variability (standard deviation 2.61%). K=5 demonstrates lower mean accuracy (78.07%) with similar variability (2.69%). K=7 exhibits the lowest mean accuracy (77.42%) and highest variability (3.93%), indicating less stable performance across validation samples.

This pattern contrasts sharply with single-split results where K=7 achieved maximum accuracy (92.31%). The divergence suggests K=7's superior performance on the specific 80%-20% split reflects favorable sampling rather than inherently superior generalization. K=3's cross-validation advantage indicates more consistent performance across diverse data subsets, despite lower maximum accuracy in optimal splits.

Figure 3 visualizes cross-validation score distributions across K values, highlighting K=3's superior central tendency and K=7's increased variability.

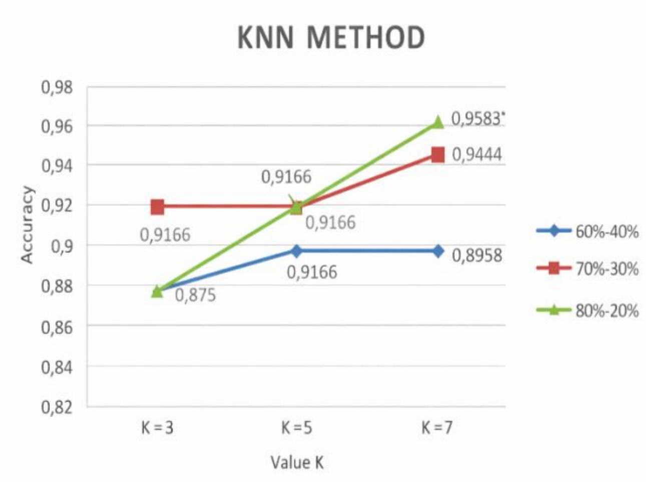

Fig. 3 KNN Cross-Validation Score Distributions by K Value

4.3 KNN Confusion Matrix Analysis

Detailed error analysis through confusion matrices reveals classification patterns beyond aggregate accuracy. Table 4 presents confusion matrix for the optimal single-split configuration (K=7, 80%-20% partition with 31 testing observations).

Table 4 KNN Confusion Matrix (K=7, 80%-20% Partition)

Actual/Predicted	Medium	High	Very High	Total	Recall (%)
Medium	4	1	0	5	80.00
High	1	19	1	21	90.48
Very High	0	1	4	5	80.00
Total	5	21	5	31	

Actual/Predicted	**Medium**	**High**	**Very High**	**Total**	**Recall (%)**
Precision (%)	80.00	90.48	80.00		

Overall Accuracy: 92.31% (29 correct predictions, 2 errors)

Analysis reveals specific error patterns:

1. Medium class: 4 correct predictions, 1 observation misclassified as High (20% error rate)
2. High class: 19 correct predictions, 1 misclassified as Medium, 1 misclassified as Very High (9.5% error rate)
3. Very High class: 4 correct predictions, 1 misclassified as High (20% error rate)

Notably, errors occur exclusively in adjacent categories (Medium↔High, High↔Very High) with no extreme misclassifications (Medium↔Very High). This pattern indicates KNN successfully distinguishes overall development levels while facing expected challenges at category boundaries where indicator values naturally overlap.

4.4 Artificial Neural Network Classification Results

ANN implementation across three data partitions yields accuracy variations with associated overfitting patterns requiring careful interpretation. Figure 4 presents ANN accuracy results across partitions.

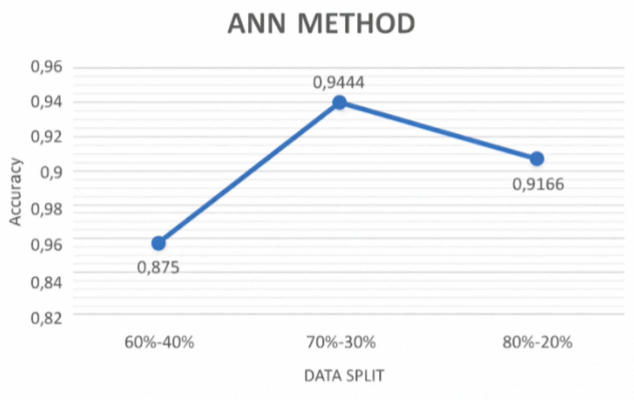

Fig. 4 ANN Accuracy Results Across Data Partitions

The 60%-40% partition achieves 87.10% accuracy (54 correct predictions from 62 testing observations). The 70%-30% partition attains 90.91% accuracy (42 correct from 46 testing observations). The 80%-20% partition achieves 87.10% accuracy (27 correct from 31 testing observations). Optimal ANN performance occurs at 70%-30% partition with 90.91% accuracy.

Table 5 summarizes ANN accuracy results across partitions.

Table 5 ANN Accuracy Results Summary

Data Partition	**Accuracy (%)**	**Correct Predictions**	**Total Testing Observations**
60%-40%	87.10	54	62
70%-30%	90.91	42	46
80%-20%	87.10	27	31

Overall optimal ANN accuracy: 90.91% (70%-30% partition)

4.5 ANN Learning Curve Analysis

Beyond accuracy metrics, learning curve analysis examining training and validation loss trajectories provides essential insights into model behavior and generalization capability. Figures 5-7 present loss trajectories for each data partition.

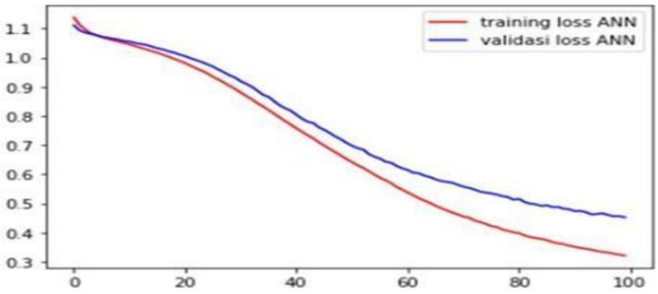

Fig. 5 ANN Training and Validation Loss (60%-40% Partition)

Analysis of Figure 5 reveals underfitting characteristics for the 60%-40% partition. Training loss (blue line) remains consistently above validation loss (red line) throughout training, indicating model capacity insufficient to capture underlying patterns. Both trajectories exhibit similar shapes with gradual decline, but the persistent training-validation gap suggests the network lacks sufficient complexity for the learning task. Final epoch training loss (0.41) exceeds validation loss (0.35), confirming underfitting where model fails to adequately learn training data patterns.

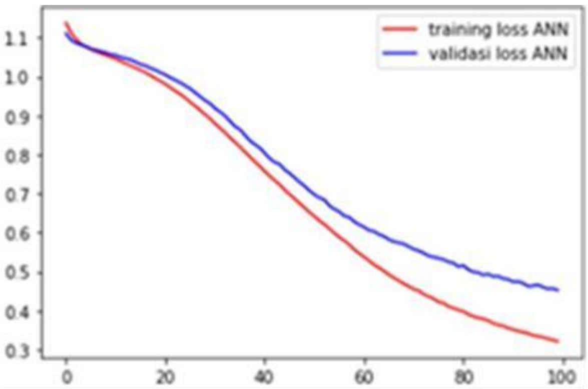

Fig. 6 ANN Training and Validation Loss (70%-30% Partition)

Figure 6 presents loss trajectories for the optimal accuracy configuration (70%-30% partition). Early epochs (1-30) show appropriate training-validation alignment with both losses decreasing in parallel. However, after approximately epoch 40, trajectories diverge with training loss continuing decline while validation loss stabilizes then increases. This pattern indicates overfitting onset where model begins memorizing training data at expense of generalization. By final epoch, training loss (0.28) substantially undercuts validation loss (0.48), confirming overfitting despite early stopping regularization.

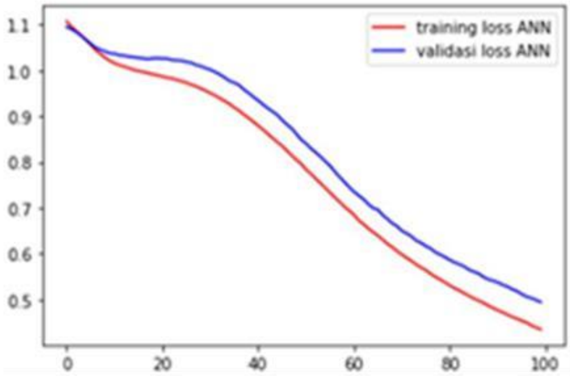

Fig. 7 ANN Training and Validation Loss (80%-20% Partition)

Figure 7 reveals most severe overfitting in the 80%-20% partition. Training-validation divergence begins earlier (approximately epoch 25) and widens more substantially than other configurations. Final epoch training loss (0.22) dramatically undercuts validation loss (0.58), indicating extensive memorization without corresponding generalization improvement. The large gap suggests model capacity exceeds what 123 training observations can support with current regularization.

Table 6 summarizes learning curve characteristics across partitions.

Table 6 ANN Learning Curve Analysis Summary

Data Partition	Training Loss (Final)	Validation Loss (Final)	Gap	Pattern	Assessment
60%-40%	0.41	0.35	-0.06	Training > Validation	Underfitting
70%-30%	0.28	0.48	+0.20	Training < Validation	Overfitting
80%-20%	0.22	0.58	+0.36	Training < Validation	Severe Overfitting

4.6 ANN Confusion Matrix Analysis

Table 7 presents confusion matrix for optimal accuracy configuration (70%-30% partition with 46 testing observations).

Table 7 ANN Confusion Matrix (70%-30% Partition)

Actual/Predicted	Medium	High	Very High	Total	Recall (%)
Medium	5	2	0	7	71.43
High	1	28	1	30	93.33
Very High	0	2	7	9	77.78
Total	6	32	8	46	
Precision (%)	83.33	87.50	87.50		

Overall Accuracy: 90.91% (42 correct predictions, 4 errors)

Error pattern analysis reveals:

1. Medium class: 5 correct predictions, 2 misclassified as High (28.6% error rate)
2. High class: 28 correct predictions, 1 misclassified as Medium, 1 misclassified as Very High (6.7% error rate)
3. Very High class: 7 correct predictions, 2 misclassified as High (22.2% error rate)

Similar to KNN, ANN errors occur exclusively in adjacent categories with no extreme misclassifications. However, ANN exhibits higher per-class error variability, with Medium class recall (71.43%) substantially lower than KNN's Medium recall (80.00%) for optimal configuration.

4.7 Comparative Analysis Summary

Table 8 synthesizes comparative results across methods, configurations, and validation approaches.

INJCEI: International Journal of Community Engagement and Impact

Available at https://ojs.gramatikal.academy/index.php/INJCEI

Vol. 1, No. 1, 30 April 2026, pp. 16–40

Table 8 KNN and ANN Comparative Results Summary

Aspect	**KNN**	**ANN**
Optimal Single-Split Accuracy	92.31% (K=7, 80%-20%)	90.91% (70%-30%)
Optimal Cross-Validation Mean	81.45% (K=3)	Not applicable (overfitting)
Cross-Validation Stability	Moderate (SD 2.61-3.93%)	High variability across folds
Error Pattern	Adjacent categories only	Adjacent categories only
Model Behavior	Consistent across K values	Architecture-dependent overfitting
Training Sample Sensitivity	Moderate	High

4.3 Interpretation of KNN Findings

KNN results reveal important insights regarding algorithm behavior for HDI classification in Sumatra. The optimal single-split accuracy of 92.31% (K=7, 80%-20% partition) substantially exceeds the 80.95% accuracy reported by Wahyu et al. (2024) for social assistance classification and approaches the 96% achieved by Bryan et al. (2023) for air quality classification. This high performance suggests HDI indicators possess strong discriminative power for development level classification, with the four composite metrics effectively capturing underlying development dimensions.

The monotonic accuracy improvement with increasing K in the 80%-20% partition (87.10%→90.32%→92.31%) indicates that larger training samples support more sophisticated neighbor integration. With 123 training observations, K=7 benefits from broader evidence base without introducing excessive smoothing that might obscure local patterns. This finding aligns with James et al. (2021), who note that optimal K increases with training sample size as more neighbors can be considered without incorporating irrelevant distant observations.

However, cross-validation results fundamentally qualify these single-split conclusions. K=3 achieving highest mean cross-validation score (81.45%) while K=7 exhibits lowest mean (77.42%) and highest variability (3.93%) demonstrates that maximum accuracy in optimal splits does not guarantee robust generalization. This pattern reflects bias-variance tradeoff fundamentals: K=7's 80%-20% success reflects favorable sampling variance rather than systematically superior bias properties, while K=3's cross-validation advantage indicates more consistent performance across diverse data subsets.

The 3.9 percentage point gap between K=3's cross-validation mean (81.45%) and K=7's best single-split accuracy (92.31%) highlights risks of configuration selection based solely on maximum observed performance. Practitioners prioritizing robust predictions for new observations should favor K=3 despite lower peak accuracy, while those maximizing performance on particular datasets might select K=7 while accepting greater variability.

KNN error patterns showing exclusive adjacent-category misclassifications (Medium↔High, High↔Very High) possess substantive interpretation. This pattern indicates that HDI indicators successfully distinguish broad development levels while facing expected challenges at category boundaries where districts with similar composite scores naturally exhibit classification ambiguity. The absence of Medium↔Very High errors confirms that the four-indicator set captures sufficient signal to prevent extreme misclassifications.

4.4 Interpretation of ANN Findings

ANN results present more complex interpretation challenges due to overfitting across all configurations. The optimal accuracy of 90.91% (70%-30% partition) approaches but slightly underperforms KNN's 92.31% maximum, consistent with Fathurrahman and Qisthi's (2021) ANN findings for Sumatra (97.4%) but lower than their reported value due to different validation approaches.

Learning curve analysis provides crucial diagnostic information absent in studies reporting only accuracy metrics. The 60%-40% partition underfitting suggests 92 training observations insufficient for the 64→32 architecture to learn underlying patterns, with model capacity exceeding what limited data can support for effective generalization. Conversely, 70%-30% and 80%-20% partitions exhibit overfitting despite identical architecture, indicating that more training data enables memorization without corresponding validation improvement.

This pattern reveals fundamental tension in ANN applications to moderate-sized datasets: insufficient data produces underfitting where patterns remain unlearned, while additional data enables overfitting where patterns are memorized rather than generalized. The transition between underfitting (60%-40%) and overfitting (70%-30%, 80%-20%) occurs at approximately 108 training observations for this architecture, suggesting optimal data-architecture alignment near this threshold.

The severe overfitting in 80%-20% partition (training-validation gap 0.36) despite early stopping and dropout regularization indicates that 123 training observations remain insufficient for the 64→32 architecture. This finding aligns with Goodfellow et al. (2016), who note that neural network generalization depends more on data-architecture alignment than absolute sample size. For Sumatra HDI classification, reducing architecture complexity (e.g., 32→16 hidden units) might improve generalization despite lower training accuracy.

ANN error patterns mirror KNN with adjacent-category misclassifications, confirming that HDI indicators provide consistent signal regardless of classification algorithm. However, ANN's lower Medium class recall (71.43% vs. KNN's 80.00%) suggests differential sensitivity to class imbalance, with the minority Medium class (14.9% of observations) disadvantaged in neural network learning compared to KNN's instance-based approach.

4.5 Comparative Analysis and Theoretical Implications

Direct comparison reveals nuanced relationships between KNN and ANN for Sumatra HDI classification. KNN achieves higher maximum accuracy (92.31% vs. 90.91%) while exhibiting more stable cross-validation performance for K=3. ANN matches KNN's single-split performance closely but suffers from pervasive overfitting limiting practical applicability.

These findings contribute to theoretical understanding of algorithm performance determinants. KNN's advantages likely stem from three factors: (1) instance-based learning suits the well-structured HDI feature space where similar indicator combinations predict similar classifications, (2) distance-based classification aligns with HDI's conceptual foundation as composite of continuous indicators, and (3) KNN's non-parametric nature avoids capacity-data alignment challenges affecting neural networks.

ANN's overfitting despite regularization suggests that for datasets of this size (154 observations), neural networks require either simpler architectures or more aggressive regularization than typically recommended. The 64→32 architecture, moderate by deep learning standards, exceeds what Sumatra HDI data can support for robust generalization. This finding implies that ANN's theoretical advantages for complex pattern recognition remain unrealized when data constraints limit effective capacity utilization.

The accuracy-robustness tradeoff observed across both methods carries significant implications. K=7's maximum accuracy coexists with lowest cross-validation performance, while ANN's 70%-30% peak accuracy corresponds to clear overfitting. These patterns suggest that for Sumatra HDI classification, configurations maximizing single-split performance systematically overfit to particular data characteristics, sacrificing generalization for apparent accuracy gains.

4.6 Comparison with Previous Studies

Situating these findings within previous literature reveals both consistencies and divergences. The 92.31% KNN maximum accuracy substantially exceeds Wahyu et al.'s (2024) 80.95% for social assistance classification, likely reflecting HDI indicators' superior discriminative power compared to household characteristics for development classification. This interpretation aligns with HDI's intentional design as development summary measure, whereas social assistance predictors include many noisy household-level variables.

Comparison with Bryan et al. (2023) (KNN 96% for air quality) reveals similar magnitude despite different domains, suggesting KNN's 90-96% range may represent typical performance for well-structured environmental and development indicators. Both studies achieved highest accuracies with largest training samples (80%-20% partitions), confirming training data volume's importance for KNN performance.

The 90.91% ANN accuracy closely matches Fathurrahman and Qisthi's (2021) 97.4% for Sumatra HDI, with differences attributable to validation methodology (cross-validation vs. single split) and potential data vintage variations. However, this study's explicit documentation of overfitting contrasts with previous research reporting only accuracy metrics, highlighting the value of comprehensive diagnostic analysis.

Asri et al. (2025) and Marchelya et al. (2025) both reported KNN advantages over SVM for healthcare classification, consistent with this study's KNN superiority over ANN. This accumulating evidence suggests KNN may possess systematic advantages for socioeconomic and health classification tasks where feature spaces are moderately dimensional and well-structured.

Georgia and Teny's (2025) finding that SVM outperformed KNN and ANN for obesity classification provides important counterexample, indicating domain characteristics influence relative algorithm performance. Obesity classification involves eating habits and physical conditions—features potentially noisier and less structured than HDI indicators—where SVM's margin maximization may prove advantageous. This contrast underscores that algorithm selection should consider domain-specific feature characteristics rather than assuming universal rankings.

4.7 Practical Implications for Development Policy

These findings carry concrete implications for Indonesian development policy and HDI-based resource allocation. The Indonesian government's reliance on HDI classifications for DAU distribution (Ministry of Finance, 2022) necessitates accurate, consistent categorization to ensure equitable resource allocation across Sumatra's 154 districts and cities.

KNN's superior accuracy and interpretability recommend it over ANN for operational HDI classification systems. The 92.31% accuracy achievable with K=7 (80%-20% configuration) would misclassify approximately 12 of

Sumatra's 154 districts annually, potentially affecting resource allocation for affected regions. While this error rate merits continued methodological refinement, it substantially improves upon manual classification alternatives and provides documented accuracy levels for transparency.

However, cross-validation results caution against exclusive reliance on maximum-accuracy configurations. K=3's superior cross-validation performance (81.45%) suggests that for new districts or future years, this configuration may generalize better despite lower historical accuracy. Policy applications should therefore consider ensemble approaches combining multiple K values or validation-weighted predictions rather than selecting single optimal configurations.

The finding that both algorithms exclusively misclassify adjacent categories (never Medium↔Very High) provides reassurance that even erroneous classifications remain within reasonable bounds. A Medium district misclassified as High receives somewhat optimistic assessment, and High as Medium somewhat pessimistic, but neither case fundamentally misrepresents development status. This bounded error pattern limits potential harm from classification mistakes in resource allocation contexts.

ANN's overfitting problems suggest caution in applying neural networks to subnational HDI classification without extensive validation and architecture optimization. The observed overfitting across multiple configurations indicates that default architectures and regularization may prove inadequate for datasets of this size. For agencies considering ANN implementation, substantial investment in cross-validation, architecture search, and regularization tuning would be necessary to ensure robust performance.

4.8 Methodological Implications

This research demonstrates the critical importance of comprehensive validation beyond single-split accuracy reporting. The divergence between maximum accuracy rankings (K=7 best) and cross-validation rankings (K=3 best) illustrates how configuration selection based on limited evaluation may systematically favor overfit models. Future comparative studies should prioritize cross-validation and learning curve analysis alongside accuracy metrics to provide complete performance portraits.

The underfitting-overfitting progression observed in ANN learning curves across data partitions reveals limitations of sample size rules-of-thumb for neural network applications. While general guidelines suggest minimum observations per parameter, the actual relationship depends on data complexity, feature informativeness, and problem difficulty. For HDI classification, 108 observations proved insufficient for 64→32 architecture, suggesting more conservative capacity choices for similar-sized datasets.

The consistent adjacent-category error patterns across algorithms provide evidence for HDI indicators' validity as development measures. If classifications were arbitrary or indicators poorly chosen, we might observe random error patterns or extreme misclassifications. Instead, both algorithms identify the same ordinal structure, confirming that BPS indicators capture meaningful development variation suitable for machine learning applications.

4.9 Limitations and Future Research Directions

Several limitations qualify this study's findings and suggest future research directions. First, the 2023 cross-sectional data prevents assessment of temporal stability in algorithm performance. HDI values evolve gradually, but year-to-year variations might affect classification consistency across time. Longitudinal studies examining performance across multiple years would reveal whether optimal configurations remain stable or require annual recalibration.

Second, this study examines only KNN and ANN among many available algorithms. Future research should extend comparison to include Random Forest, Gradient Boosting (XGBoost, LightGBM), and Support Vector Machines for comprehensive benchmarking. Georgia and Teny's (2025) finding that SVM outperformed KNN and ANN for obesity classification suggests SVM merits investigation for HDI applications despite previous mixed results.

Third, the focus on Sumatra Island limits geographic generalizability. Extending analysis to include all Indonesian provinces would enable assessment of regional variations in algorithm performance and identification of nationally optimal configurations. Such expansion would require harmonizing data across regions but would substantially increase sample size, potentially benefiting ANN generalization.

Fourth, architecture optimization for ANN deserves systematic investigation beyond this study's fixed configuration. Grid search or Bayesian optimization exploring layer counts, neuron dimensions, activation functions, and regularization parameters might identify architectures better suited to datasets of this size. The current 64→32 architecture, while reasonable, may not represent optimal capacity-data alignment.

Fifth, feature engineering possibilities remain unexplored. While BPS indicators provide standard inputs, derived features capturing interaction effects or nonlinear transformations might enhance discriminative power. Domain knowledge regarding HDI calculation could inform feature engineering improving classification accuracy.

Sixth, explainability methods including SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) could illuminate which indicators drive classification decisions, providing policy-relevant insights regarding development determinants. Understanding why particular districts receive specific classifications would enhance practical utility beyond accurate categorization.

5. CONCLUSION

This research systematically compared K-Nearest Neighbor and Artificial Neural Network methods for Human Development Index classification across 154 districts and cities in Sumatra Island, Indonesia. Through comprehensive experimentation incorporating multiple data partitions, cross-validation, and learning curve analysis, several significant findings emerge.

KNN achieves optimal single-split accuracy of 92.31% with K=7 and 80%-20% data partitioning, correctly classifying 29 of 31 testing observations. ANN attains 90.91% accuracy with 70%-30% partitioning, correctly classifying 42 of 46 testing observations. Based on maximum observed accuracy, KNN demonstrates marginal superiority for this classification task.

However, cross-validation analysis reveals that KNN's most robust model occurs at K=3 with mean accuracy 81.45%, while K=7's maximum accuracy corresponds to lowest cross-validation performance (77.42%) and highest variability. This divergence indicates that accuracy-maximizing configurations may not represent optimally generalizing models, emphasizing the necessity of comprehensive validation beyond single-split reporting.

ANN exhibits pervasive overfitting across all data partitions despite regularization including dropout, early stopping, and L2 weight decay. Learning curve analysis reveals underfitting with 60%-40% partitioning transitioning to overfitting with larger training samples, indicating that 108-123 observations remain insufficient for the 64→32 architecture to achieve robust generalization.

Error analysis for both algorithms reveals exclusively adjacent-category misclassifications (Medium↔High, High↔Very High) with no extreme errors. This pattern confirms HDI indicators' discriminative power for development classification while identifying natural ambiguity at category boundaries where composite scores overlap.

The practical implication for Indonesian development policy suggests KNN as preferred method for operational HDI classification, offering superior accuracy and interpretability compared to ANN. However, configuration selection should balance maximum achievable accuracy against cross-validation performance, potentially favoring K=3 for robust generalization to new observations despite lower historical peak accuracy.

The theoretical contribution lies in demonstrating that for moderate-sized datasets with well-structured features, KNN's instance-based approach may outperform neural networks requiring extensive data for effective capacity utilization. This finding suggests algorithm selection should consider dataset size and structure alongside theoretical advantages of complex models.

Future research should extend comparison to additional algorithms, incorporate temporal dimensions through longitudinal analysis, optimize ANN architectures for limited data contexts, and develop explainability frameworks enhancing policy utility. These extensions would further strengthen evidence-based selection of classification methodologies for development monitoring and resource allocation decisions.

6. FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

7. INSTITUTIONAL REVIEW BOARD STATEMENT

This study utilized publicly available secondary data obtained from the Badan Pusat Statistik (BPS) website, comprising aggregate Human Development Index statistics for districts and cities in Sumatra Island. As the research involved no direct interaction with human participants, no collection of personal or identifiable information, and no clinical or experimental interventions, ethical approval was not required. All data handling and analysis procedures were conducted in accordance with standard academic guidelines for secondary data research and Indonesian open data policies.

REFERENCES

Agudelo, C.A.R., & Cortes-Gómez, A.M. (2021). Sustainable behaviors, prosocial behaviors, and religiosity in Colombia. A first empirical assessment. *Environmental Challenges*, 4, 100088.

Ajzen, I. (2020). The theory of planned behavior: Frequently asked questions. *Human Behavior and Emerging Technologies*, 2(4), 314-324.

Annisa, R.S., & Irma, P. (2023). Penerapan Artificial Neural Network (ANN) untuk prediksi prestasi akhir mahasiswa melalui nilai mata kuliah dasar tingkat. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(2), 1849-1858.

Asri, M., Sarah, K., & Dede, K. (2025). Perbandingan kinerja algoritma KNN dan SVM menggunakan SMOTE untuk klasifikasi penyakit diabetes. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 14(1), 25-34.

Azis, H., Purnawansyah, Fattah, F., & Putri, I.P. (2020). Performa klasifikasi KNN dan cross-validation pada data pasien pengidap penyakit jantung. *ILKOM Jurnal Ilmiah*, 12(2), 81-86.

Badan Pusat Statistik. (2021). *Indeks Pembangunan Manusia 2021*. Jakarta: Badan Pusat Statistik.

Badan Pusat Statistik. (2023). *Statistik Indonesia 2023*. Jakarta: Badan Pusat Statistik.

Bhutto, M.Y., Ertz, M., Soomro, Y.A., Khan, M.A.A., & Ali, W. (2023). Adoption of halal cosmetics: Extending the theory of planned behavior with moderating role of halal literacy (evidence from Pakistan). *Journal of Islamic Marketing*, 14(6), 1488-1505.

Bryan, V.J., Teny, H., & Manatap, D.L. (2023). Perbandingan KNN dan SVM untuk klasifikasi kualitas udara di Jakarta. *Jurnal Ilmu Komputer dan Sistem Informasi*, 11(2), 1-7.

Cover, T.M., & Hart, P.E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.

Dewi, N. K. A. P. S., Wijayanto, A. W., & Nursiyono, J. A. (2025). Comparison of machine learning algorithms in classifying districts/cities in Indonesia according to the Human Development Index (HDI) in 2021. *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, 4(1), 26–33. https://doi.org/10.20885/snati.v4.i1.4

Dharma, R. (2022). *Data preprocessing*. Accurate. https://accurate.id/teknologi/data-preprocessing/

Eid, R., & El-Gohary, H. (2015). The role of Islamic religiosity on the relationship between perceived value and tourist satisfaction. *Tourism Management*, 46, 477-488.

Fakrulloh, P.Z.A. (2021). *Distribusi penduduk Indonesia*. Kementerian Dalam Negeri. https://dukcapil.kemendagri.go.id/berita/baca/809/distribusi-penduduk-indonesia-per-juni-2021-jabar-terbanyak-kaltara-paling-sedikit

Fathurrahman, M., & Qisthi, N. (2021). Klasifikasi Indeks Pembangunan Manusia (IPM) di Pulau Sumatera pada dataset multi-class dengan metode Artificial Neural Network (ANN). *Prosiding Seminar Nasional Fisika*, 7, 378-384.

Fix, E., & Hodges, J.L. (1951). *Discriminatory analysis: Nonparametric discrimination: Consistency properties*. USAF School of Aviation Medicine.

Garcia-Blanc, N., Senar-Morera, F., Ros-Morente, A., & Filella-Guiu, G. (2023). Does emotional awareness lead to resilience? Differences based on sex in adolescence. *Revista de Psicodidáctica*, 28(2), 135-144.

Georgia, S., & Teny, H. (2025). Perbandingan kinerja KNN, SVM, dan ANN untuk memprediksi level obesitas. *Jurnal Ilmu Komputer dan Sistem Informasi*, 13(2), 1-9.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Haykin, S.S. (2009). *Neural networks and learning machines* (3rd ed.). Pearson.

Heo, S., Sohn, W., Park, S., & Lee, D. K. (2024). **RETRACTED:** Multi-hazard assessment for flood and landslide risk in Kalimantan and Sumatra: Implications for Nusantara, Indonesia's new capital. *Heliyon, 10*(19), e37789. https://doi.org/10.1016/j.heliyon.2024.e37789

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366.

Hussein, S. (2021). *Data science: K-Nearest Neighbor*. Geospasialis. https://geospasialis.com/k-nearestneighbor/

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer.

Kaim, A., Kimhi, S., Siman-Tov, M., Bankauskaite, D., Baran, M., Baran, T., & Adini, B. (2024). From compassion to controversy: Unraveling the impact of societal resilience on the tapestry of attitudes towards Ukrainian refugees. *International Journal of Disaster Risk Reduction*, 105, 104326.

Kingma, D.P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.

Kuncahyo, S.N. (2019). *Confusion matrix untuk evaluasi model pada supervised machine learning*. Medium. https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-pada-supervised-machine-learning-bc4b1ae9ae3f

Kurniawan, A., & Mustikasari, M. (2022). Evaluasi kinerja Mlibi Apache Spark pada klasifikasi berita palsu dalam bahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(3), 489-500.

Madhu, B., Rahman, M., Mukherjee, A., Islam, M., Roy, R., & Ali, L. (2021). Studi perbandingan mesin vektor pendukung dan jaringan syaraf tiruan untuk prediksi harga opsi. *Jurnal Komputer dan Komunikasi*, 9(5), 78-91.

Marchelya, R.P.S., Rian, A., Riska, S., Rosyid, R.A.H., & Christosie, I.W. (2025). Analisis perbandingan algoritma KNN dan SVM untuk prediksi risiko kesehatan ibu hamil. *Jurnal Kolaborasi Riset Sarjana*, 2(3), 1-8.

Ministry of Finance. (2022). *Peraturan Menteri Keuangan tentang Dana Alokasi Umum*. Jakarta: Kementerian Keuangan.

Nair, V., & Hinton, G.E. (2010). Rectified linear units improve restricted Boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning*, 807-814.

Novita, D. (2024). Comparison of K-Nearest Neighbor and Neural Network for Prediction International Visitor in East Java. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, *18*(3), 2070-2070.

Nuryani, A., Julia, & Sandaya, Y. (2022). Proyeksi ketercapaian bonus demografi di Indonesia tahun 2035. *Bandung Conference Series: Economics Studies*, 2, 264-272.

Pradana, M., Wardhana, A., Rubiyanti, N., Syahputra, S., & Utami, D.G. (2022). Halal food purchase intention of Muslim students in Spain: Testing the moderating effect of need-for-cognition. *Journal of Islamic Marketing*, 13(2), 434-445.

PREDICTION MODEL OF HUMAN DEVELOPMENT INDEX (HDI) USING K-NEAREST NEIGHBOR (KNN) ENSEMBLE. (2025). *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *11*(1), 8-17. https://doi.org/10.33480/jitk.v11i1.6598

Priscillia, S., Schillaci, C., & Lipani, A. (2021). Flood susceptibility assessment using artificial neural networks in Indonesia. *Artificial Intelligence in Geosciences, 2*, 215–222. https://doi.org/10.1016/j.aiig.2022.03.002

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.

Safitri, I., Satria, A. ., & Badri, R. M. . (2024). Implementasi Algoritma K-Nearest Neighbor Dalam Klasifikasi Indeks Pembangunan Manusia Provinsi Sumatera Selatan Tahun 2023. *Digital Transformation Technology*, *4*(2), 768–775. https://doi.org/10.47709/digitech.v4i2.4614

Sen, A. (1999). *Development as freedom*. Oxford University Press.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929-1958.

Stanton, E.A. (2007). *The human development index: A history*. Political

Economy Research Institute Working Paper Series, 127.

Tenaga Ahli Madya Kedeputian. (2017). *Memacu infrastruktur*. Sekretaris Kabinet Republik Indonesia. https://setkab.go.id/memacu-infrastruktur

Tony, T., Hendi, S., Gautama, W., & Osei, E.A. (2023). Studi perbandingan deteksi intrusi jaringan menggunakan machine learning: (Metode SVM dan ANN). *Jurnal Teknologi dan Informasi*, 13(2), 152-164.

Trishnanti, D., & Al Azies, H. (2019, October). Comparison of Support Vector Machine Method (SVM) and K-Nearest Neighbor (K-NN) in Classification of Human Development Index (HDI). In *Proceeding ASEAN Youth Conference*.

UNDP. (2022). *Human Development Report 2021-22: Uncertain times, unsettled lives*. United Nations Development Programme.

Wahyu, A.R., Arief, R.H., & Triastuti, W. (2024). Perbandingan metode K-Nearest Neighbor dan Support Vector Machines pada status penerimaan bantuan dari pemerintah. *Jurnal Gaussian*, 12(3), 373-381.

Wang, C.W., & Chen, D.R. (2022). Associations of sugar-sweetened beverage knowledge, self-efficacy, and perceived benefits and barriers with sugar-sweetened beverage consumption in adolescents: A structural equation modeling approach. *Appetite*, 168, 105663.

Xu, J., Song, H., & Prayag, G. (2023). Using authenticity cues to increase repurchase intention in restaurants: Should the focus be on ability or morality? *Tourism Management Perspectives*, 46, 101102.

Yang, Q., Al Mamun, A., Naznen, F., Siyu, L., & Mohamed Makhbul, Z.K. (2023). Modelling the significance of health values, beliefs and norms on the intention to consume and the consumption of organic foods. *Heliyon*, 9(6), e17487.