BAREKENG: Journal of Mathematics and Its Applications
December 2025 Volume 19 Issue 4 Page 2693–2708
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19no4pp2693-2708

DISTRIBUTION MODEL OF HUMAN DEVELOPMENT INDEX IN PAPUA PROVINCE BASED ON REGIONAL CLUSTERING

Jonathan K. Wororomi[1*], Alvian M. Sroyer[2], Henderina Morin[3], Felix Reba[4], Ishak S. Beno[5], Oscar O. O. Wambrauw[6]

[1,2,4,5]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Cenderawasih*
[3]*Department of Government Science, Faculty of Social and Political Sciences, Universitas Cenderawasih*
[6]*Department of Management, Faculty of Economics and Business, Universitas Cenderawasih*
Jln. Kamp Wolker Yabansai, Jayapura, Papua, 99351, Indonesia

Corresponding author's e-mail: * **jkwororomi@gmail.com**

Article History:

Received: 14th February 2025
Revised: 20th March 2025
Accepted: 27th May 2025
Available online: 1st September 2025

Keywords:

Clustering;
Goodness of Fit (GoF);
Human Development Index;
Papua;
Probability Distribution.

ABSTRACT

Modeling the distribution of Human Development Index (HDI) components is essential to uncover underlying disparities and guide targeted policy interventions. This study aims to analyze HDI data, focusing on the average length of schooling across 26 districts in Papua Province from 2010 to 2023, to identify the most suitable probability distribution model. Using the k-means clustering method, two main groups were identified based on the average length of schooling. Cluster 1 includes 11 districts with a Weibull distribution, characterized by a scale parameter of 8.9931 and a shape parameter of 16.1272, indicating significant variation in education duration. Cluster 2 consists of 15 districts with a scale parameter of 3.73006 and a shape parameter of 8.07662, showing a distribution with a long tail and greater variability. This study provides insights into the distribution patterns of education duration in Papua, which could aid policymakers in making more targeted decisions and allocating resources efficiently. The findings also highlight regional disparities and the need for specific educational interventions. These results are valuable for government entities, NGOs, researchers, and international donors interested in improving educational outcomes in underdeveloped areas. However, the analysis is limited by the scope of available data and the assumption of homogeneity within clusters.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License (https://creativecommons.org/licenses/by-sa/4.0/).

How to cite this article:

J. K. Wororomi, A. M. Sroyer, H. Morin, F. Reba, I. S. Beno, and O. O. O. Wambrauw, "DISTRIBUTION MODEL OF HUMAN DEVELOPMENT INDEX IN PAPUA PROVINCE BASED ON REGIONAL CLUSTERING", *BAREKENG: J. Math. & App.*, vol. 19, iss. 4, pp. 2693-2708, December, 2025.

Copyright © 2025 Author(s)
Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

The Human Development Index (HDI) is a crucial indicator used to measure the level of human development in a region, encompassing three main dimensions: health, education, and a decent standard of living [1], [2], [3]. In Indonesia, HDI is often used by the government and international institutions to evaluate and compare welfare levels across regions [4]. Papua, as one of the provinces facing significant development challenges, exhibits striking disparities in the education dimension, particularly in terms of the average length of schooling across different districts. Despite improvements in recent years, regional disparities remain a key issue in efforts to enhance the quality of education in Papua [5], [6]. Previous studies have analyzed HDI using various statistical and computational approaches [1]–[6], but they have generally focused on national or provincial-level aggregates. Few have explored the use of probabilistic and clustering methods for sub-provincial educational disparities in Papua, which presents unique challenges due to its geographical and socio-political context.

This study focuses on analysing the average length of schooling data as an essential component of the HDI for 26 districts in Papua from 2010 to 2023. Through a clustering approach, this research aims to identify patterns or groups of regions with similar characteristics based on the average length of schooling [7], [8], and to determine the most appropriate probability distribution model for the data [9], [10]. Probability distribution methods, such as the Weibull, Generalized Extreme Value (GEV), and Nakagami distributions, are employed to depict the variability patterns in education in Papua. The choice of these three distributions is based on their ability to accommodate different statistical characteristics of educational data. Weibull is particularly well-suited for modeling duration-type variables such as schooling due to its flexibility in shape and scale [24], [25], [26]. GEV is used to model extreme values and long-tailed distributions, which are important for capturing outlier regions with exceptionally high or low education indicators [41], [42]. Meanwhile, Nakagami is valuable in modeling asymmetric and highly variable data patterns that reflect the diverse conditions across Papua's districts [30], [43], [45].

Despite the frequent use of HDI in evaluating education policies, limited studies have combined clustering techniques with probability distribution modeling to address within-region disparities. This study fills that gap by applying statistical modeling to sub-provincial HDI indicators in Papua. The primary objective of this research is to provide deeper insights into the educational variation in Papua and offer a scientific basis for more accurate policy decisions. With the findings obtained, the government and other stakeholders can develop more effective policies to address educational disparities in underdeveloped areas, thereby enabling a more optimal allocation of resources [11], [12], [13], [14], [15].

It is also hoped that this research will attract the attention of these parties so that they can design more effective intervention programs. The novelty of this study lies in its integration of probability distribution modeling with sub-provincial clustering analysis to uncover intra-regional educational disparities in Papua—an approach that has rarely been explored in previous literature.

2. RESEARCH METHODS

The methodology employed in this study integrates various statistical techniques for data analysis to identify the most suitable probability distribution model for the average length of schooling across 26 districts in Papua Province. The dataset includes the average years of schooling in each district from 2010 to 2023, collected through official surveys conducted by the Central Bureau of Statistics (BPS) of Papua Province, including the National Socioeconomic Survey (SUSENAS). The data were validated to ensure consistency and completeness, making it a reliable source for further statistical analysis.

K-Means clustering is utilized to classify data into distinct groups based on similarity, as it is computationally efficient, scalable, and suitable for continuous numerical data such as schooling duration. It also allows for clear interpretation and performs well when cluster boundaries are distinct, making it appropriate for this study's objective of detecting region-based patterns. The optimal number of clusters is determined using the Silhouette Coefficient and Davies-Bouldin Index, which assess cluster validity.

Once the optimal clustering structure is established, Maximum Likelihood Estimation (MLE) is applied for parameter estimation to ensure robust statistical modeling. Subsequently, Goodness-of-Fit (GoF) testing using the Anderson-Darling statistic is conducted to evaluate the suitability of various univariate probability distributions. The best-fitting distribution is further assessed through information criteria, such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), to select the most appropriate model by balancing model complexity and GoF. This systematic approach ensures the selection of the most suitable probability distribution model for understanding educational disparities across districts, which is crucial for regional policy-making.

2.1 Data Sources

The primary dataset in this study was obtained from the Central Bureau of Statistics (BPS) of Papua Province and consists of the average length of schooling from 26 districts over the period 2010 to 2023. The data were sourced from systematic censuses and sample surveys, including the National Socioeconomic Survey (SUSENAS), which periodically measures education indicators in each region.

In this study, the data collection and preparation were carried out as follows:

1. Secondary Data Collection: Data were retrieved from official BPS publications, annual statistical reports, and open-access datasets curated for academic and research purposes.
2. Data Validation and Consistency: The dataset was checked for completeness and accuracy by comparing values with national education records and official reports from the Papua Department of Education. Outliers were identified using interquartile range analysis, and missing values were handled using linear interpolation for short gaps or listwise deletion for entries with significant missingness to maintain data integrity.
3. Data Processing: The data were organized in a time-series format across districts. Prior to analysis, the data were standardized using z-score normalization to ensure comparability between districts with varying scales and to improve the performance of clustering and distribution modeling methods.

This approach enabled a rigorous and reliable evaluation of schooling distribution patterns across Papua and ensured that the data used in this study were consistent, cleaned, and ready for statistical modeling.

2.2 K-Means Clustering

To address the issue of schooling distribution in Papua, we employed the k-means clustering method, which divided the 26 districts into two clusters based on the average length of schooling from 2010 to 2023. This clustering process was conducted to identify significant differences between regions in terms of access to education. The k-means method was selected due to its ability to group regions with similar characteristics based on educational variables. The k-means clustering approach was used to classify the districts according to their similarities in the average length of schooling. The steps involved in k-means are as follows: first, the number of clusters, k, is determined, and centroids are randomly selected. The distance between each data point and the centroid is calculated using the Euclidean distance formula [16], [17], [18], [19], [20], [21]:

$$d_{ik} = \sqrt{\sum_{j=1}^{m} \left(x_{kj} - c_{ij}\right)^2} \tag{1}$$

Where d_{ik} is the distance between the i-th data point and the centroid of the cluster k, x_{kj} is the coordinate of the k-th data point, and c_{ij} is the coordinate of the i-th centroid. Each data point is then assigned to the cluster with the nearest centroid. After that, a new centroid is calculated as the mean of the data points in the cluster:

$$C_{ij} = \frac{\sum_{k=1}^{p} x_{kj}}{p} \tag{2}$$

This process is repeated until no further significant changes occur in the position of the centroids, indicating that convergence has been reached. Convergence here refers to the condition where the centroids stabilize, and no more data points switch clusters in subsequent iterations. K-means was chosen for its efficiency in handling large datasets and its suitability for identifying hidden structures in numerical data such as education indicators. The selection of two clusters was based on evaluation using the Silhouette Coefficient (SC) and the Davies-Bouldin Index (DBI), where both indices indicated optimal values when $k = 2$, suggesting that two clusters represented the best grouping structure for the available data.

2.3 Optimal k Estimation

The estimation of the optimal k value was carried out using the K-Means Clustering method, evaluated by two indices: the Silhouette Coefficient (SC) and the Davies-Bouldin Index (DBI). SC evaluates cluster compactness, while DBI measures the degree of separation between clusters. A high SC value indicates well-separated and compact clusters, whereas a low DBI suggests minimal overlap and greater distance between centroids. These two indices are used in combination to assess the quality of different cluster configurations by computing SC and DBI values for several values of k (e.g., 2 to 5), and selecting the k that yields the highest SC and lowest DBI simultaneously. The final decision for $k = 2$ was based on this dual-criteria evaluation, where both SC and DBI showed optimal values at $k = 2$, supporting the identification of two main schooling groups in Papua. These results reinforce the probability distribution analysis and are useful for developing localized educational policies. The Silhouette Coefficient (SC) is calculated using the formula [22], [23]:

$$\bar{S} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)}\right) \tag{3}$$

Where $a(i)$ is the average distance from the data point to its cluster centroid, and $b(i)$ is the average distance to the nearest centroid of another cluster. The Davies-Bouldin Index (DBI) is calculated as:

$$\mathrm{IDB} = \frac{1}{k}\sum_{i=1}^{k} R_i \tag{4}$$

Where R_i is the ratio between the average distance of each data point to its cluster centroid and the distance between centroids. Once the optimal k value is selected, each data point is assigned to the nearest centroid based on Euclidean distance, and the centroids are iteratively updated until no further changes occur in assignments—indicating convergence. SC and DBI thus provide a quantitative and statistical basis for finalizing cluster formation and ensuring that the grouping is optimal, valid, and meaningful.

2.4 Parameter Estimation with MLE

Parameter estimation was conducted using the Maximum Likelihood Estimation (MLE) method due to its ability to produce efficient and consistent parameter estimates. MLE aims to maximize the likelihood function to obtain distribution parameters that best fit the observed data in each cluster. This method was chosen for the Weibull, Generalized Extreme Value (GEV), and Nakagami distributions to describe the variation in schooling duration patterns across clusters. For the Weibull distribution, MLE estimates the shape and scale parameters by fitting the exponential growth behavior typically observed in duration data. In the case of the GEV distribution, MLE estimates the location, scale, and shape parameters, allowing it to accurately model extreme values in the data distribution tails. For the Nakagami distribution, MLE is used to estimate the shape (m) and spread (Ω) parameters, which are well-suited for modeling asymmetric and highly variable data patterns. Using MLE, the resulting parameters reflect the actual conditions of the average schooling duration distribution, providing a strong basis for evaluating the probability models used. MLE was selected because it offers efficient and consistent parameter estimates. The likelihood function for a random sample $X_1, X_2, \dots, X_n$ from $f(x;\theta)$ is given by [24], [25], [26], [27], [28]:

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) \tag{5}$$

MLE aims to find the parameter value θ that maximizes this likelihood function. The estimation is carried out using optimization techniques to ensure that the resulting parameters are the most likely based on the observed data.

2.5 Goodness-of-Fit (GOF)

To evaluate the fit of the distribution models to the observed data, we used the Goodness-of-Fit (GoF) test, particularly the Anderson-Darling (AD) test. The AD test was chosen for its sensitivity to tail differences in the distribution, which is critical in analyzing schooling duration, especially for capturing extreme values in districts with limited access to education. The hypothesis tested in this method is that the observed data follow the theoretical distribution model (e.g., Weibull, GEV, or Nakagami). The decision-making process is based on comparing the AD test statistics across candidate distributions; the distribution with the lowest AD statistic is considered to have the best fit. The GoF test results indicate that the Weibull distribution provides the best fit for representing the schooling duration distribution patterns in both clusters, compared to the GEV and Nakagami distributions. This preference is supported by the fact that the Weibull distribution effectively captures positively skewed and bounded duration data, which is typical for educational indicators such as years of schooling. While GEV and Nakagami models are suitable for modeling extremes and asymmetric patterns, the AD test statistics consistently favored Weibull in both clusters, highlighting its superior performance in modeling the overall distribution.

The Anderson-Darling (AD) test was used to evaluate the fit of the probability distributions to the observed data. The AD test statistic is calculated using the formula [29], [30], [31], [32], [33]:

$$A_n^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\log F(x_i) + \log(1 - F(x_{n+1-i})] \tag{6}$$

Where $F(x_i)$ is the cumulative distribution function of the hypothesized distribution, and n is the sample size. The AD test was chosen due to its high sensitivity to deviations in the tails of the distribution, which is relevant in analyzing the schooling duration distribution that may show extreme values.

2.6 Univariate Probability Distribution

Several probability distributions were used in this study, including Weibull, Generalized Extreme Value (GEV), and Nakagami, to describe the schooling duration distribution in Papua. The Weibull distribution is particularly suitable due to its flexibility in modeling positively skewed duration data, which is common in educational contexts. Its probability density function is defined as,

$$f(x) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} \exp\left(-\left(\frac{x}{a}\right)^{b}\right), x \geq 0, a > 0, b > 0 \tag{7}$$

where a is the scale parameter controlling spread, and b is the shape parameter determining tail behavior and skewness [33]–[39].

The Generalized Extreme Value (GEV) distribution is used to model extreme values in schooling duration, particularly in districts with very high or very low educational attainment. Its probability density function is given by:

$$f(x) = \frac{1}{\sigma}\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \tag{8}$$

defined for $1 + \xi\left(\frac{x-\mu}{\sigma}\right) > 0$, where μ is the location parameter, $\sigma > 0$ is the scale parameter, and ξ is the shape parameter that controls the tail thickness [40], [41], [42].

The Nakagami distribution is appropriate for modeling asymmetric and highly variable distributions of schooling duration, reflecting substantial disparities in access to education across different districts. Its probability density function is

$$f(x) = 2\left(\frac{\mu}{\omega}\right)^{\mu} \frac{1}{\Gamma(\mu)} x^{(2\mu-1)} exp\left(\frac{-\mu}{\omega}x^2\right); \quad x > 0 \tag{9}$$

where $\mu \geq 1/2$ is the shape parameter, $\omega > 0$ is the spread parameter, and $\Gamma(\mu)$ is the gamma function [30], [43]–[47].

The selection of these distributions was based on their theoretical ability to represent diverse schooling duration behaviors: Weibull for general flexible modeling, GEV for tail-sensitive extreme values, and Nakagami for representing asymmetric and variable patterns in the data.

2.7 Distribution Selection Criteria

To determine the best-fitting probability distribution, various information criteria were used, including Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Corrected Akaike Information Criterion (AICc), Consistent Akaike Information Criterion (CAIC), and Hannan-Quinn Criterion (HQC). The model with the lowest criteria value is considered the most appropriate for the data [48]–[51]. The formulas for each criterion are as follows:

1. Akaike Information Criterion

$$AIC = -2\ln(L) + 2k \tag{10}$$

2. Bayesian Information Criterion

$$BIC = -2\ln(L) + k\ln(n) \tag{11}$$

3. Corrected Akaike Information Criterion

$$AICc = AIC + \left(\frac{2k(k+1)}{n-k-1}\right) \tag{12}$$

4. Consistent Akaike Information Criterion

$$CAIC = -2\ln(L) + k\left(1 + \ln\left(\frac{n}{k}\right)\right) \tag{13}$$

5. Hannan-Quinn Criterion

$$HQC = -2\ln(L) + 2k\ln(\ln(n)) \tag{14}$$

Using multiple criteria ensures that the chosen distribution model not only fits the data well but is also not overly complex, avoiding overfitting.

3. RESULTS AND DISCUSSION

This study analyzed the average schooling duration data across 26 districts in Papua Province to identify the most suitable probability distribution model. Following clustering with the k-means method and model evaluation using information criteria, the findings are discussed in several stages.

3.1 Regional Clustering Based on Average Schooling Duration

The k-means clustering method was used to group 26 districts in Papua based on the average years of schooling from 2010 to 2023. This process generated two main clusters: Cluster 1 consists of districts with higher average schooling durations (7.53 to 9.67 years), indicating more advanced educational access; while Cluster 2 consists of districts with lower average schooling durations (2.61 to 4.32 years), reflecting educational challenges. As shown in Table 1, the optimal number of clusters was determined to be $k = 2$, supported by the highest Silhouette score (0.8612) and the lowest Davies-Bouldin Index (0.3787), indicating well-separated and compact clusters.

This clustering structure reveals educational disparities between regions. Districts in Cluster 1 benefit from better school infrastructure, teacher availability, and access to technology, whereas those in Cluster 2 face multiple constraints such as geographic remoteness, limited resources, and a lack of qualified personnel. These findings highlight the need for targeted policies and interventions to address the gaps, especially for regions in Cluster 2 where educational improvement is most urgently needed.

Table 1. Clustering Results and Optimal k Estimation Using Silhouette and DBI

k	Silhouette	DBI
2	**0.8612**	**0.3787**
3	0.7859	0.4760
4	0.7103	0.5597
5	0.6967	0.5489

Cluster 1

Year													
2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
7.53	7.76	7.97	8.2	8.48	8.59	8.67	8.76	8.94	9.11	9.3	9.32	9.53	9.67

Cluster 2													
Year													
2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
2.61	2.77	3	3.17	3.26	3.34	3.45	3.6	3.71	3.79	3.91	4.05	4.2	4.32

Based on the data in **Table 1** and **Figure 1**, it can be concluded that Cluster 1 represents districts with a higher average length of schooling, whereas Cluster 2 consists of districts with a lower average length of schooling. Cluster 1 represents areas with higher educational attainment and better infrastructure. Districts in this cluster have an average length of schooling close to the national standard, indicating that government efforts to improve education quality in these areas have yielded positive results. To assess the clustering quality, the Davies-Bouldin Index (DBI) was used, which evaluates the clustering structure based on the distance between centroids and the variability within clusters. Policy interventions for Cluster 1 could focus on maintaining and further enhancing educational quality.

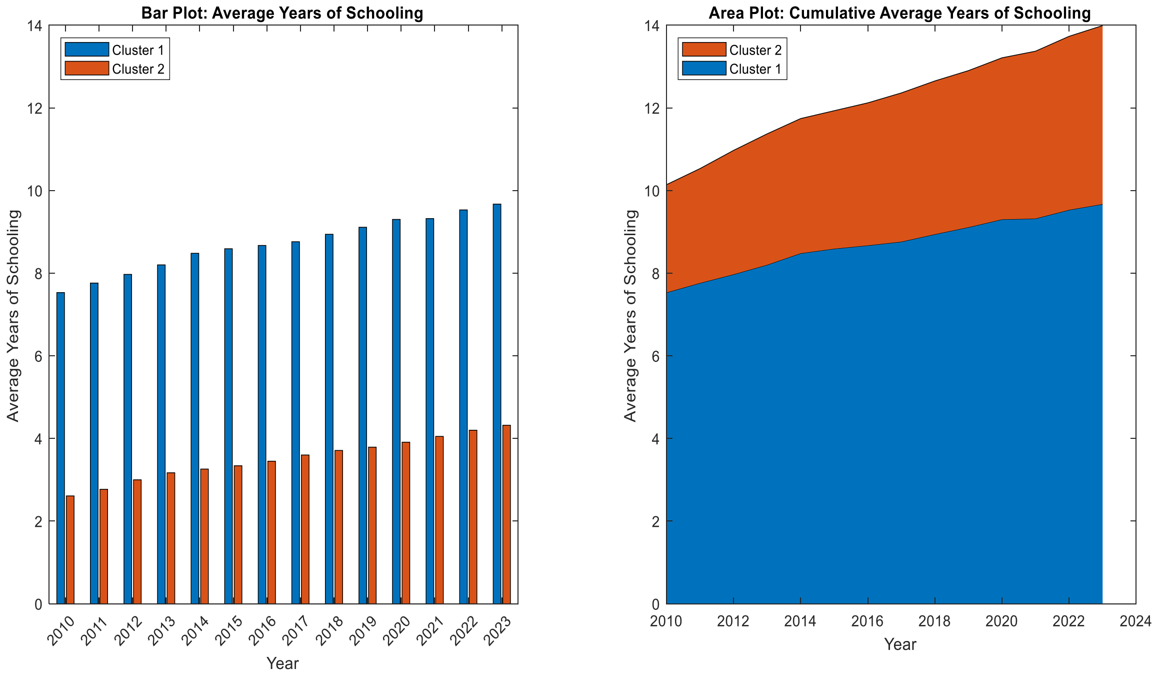

Figure 1. Bar and Area Chart of Clusters 1 and 2
Generated using MATLAB 2024b (Statistics and Machine Learning Toolbox)

Conversely, Cluster 2 highlights regions facing greater challenges. The lower average length of schooling in this cluster suggests barriers to accessing education, potentially due to geographic limitations, infrastructure gaps, or limited human resources. These findings underscore the need to focus additional attention on Cluster 2 districts by improving educational facilities, providing teacher training, and supporting educational programs in these underserved areas.

3.2 Parameter Estimation and Goodness-of-Fit (GoF) Test

After clustering using the k-means method, the next step involved estimating parameters for each hypothesized probability distribution: Generalized Extreme Value (GEV), Nakagami, and Weibull. Parameter estimation was conducted using Maximum Likelihood Estimation (MLE), known for its efficiency and consistency in identifying probability distribution parameters. Once parameters for each distribution were estimated, the Anderson–Darling (AD) Test was used as a Goodness-of-Fit (GoF) test to assess how well each hypothesized distribution fit the observed data.

In this study, the selected distribution models—GEV, Nakagami, and Weibull—were chosen for their ability to capture the variability patterns in the schooling data from Papua. GEV is effective for capturing extreme values, which may appear in districts with either very high or very low lengths of schooling. Nakagami is suitable for reflecting high variability, which aligns with the diverse educational conditions across districts. Weibull, with its flexibility, is ideal for modeling duration-based variables like schooling length, the primary focus of this research. **Table 2** presents the parameter estimation results for each distribution and the GoF test outcomes for each cluster (Cluster 1 and Cluster 2).

Table 2. GEV, Nakagami, and Weibull (Clusters 1 and 2)

Distribution	Cluster 1			Cluster 2		
	Parameter	**p-Value**	**GoF AD Test**	**Parameter**	**p-Value**	**GoF AD Test**
GEV	Shape: -0.591 Scale: 0.7145 Location: 8.567	0.9946	H_0 received	Shape: -0.485 Scale: 0.5417 Location: 3.3861	0.9988	H_0 received
Nakagami	Shape: 46.293 Scale: 76.135	0.9967	H_0 received	Shape: 12.217 Scale: 12.593	0.9995	H_0 received
Weibull	Shape: 16.1273 Scale: 8.993	0.9872	H_0 received	Shape: 8.077 Scale: 3.7301	0.9992	H_0 received

After reviewing the results in Table 2, it can be concluded that all three probability distributions—GEV, Nakagami, and Weibull—demonstrate a good fit to the average schooling length data. This conclusion is supported by the high p-values obtained from the Anderson–Darling (AD) Test, all of which exceed 0.98. In statistical terms, a high p-value indicates a failure to reject the null hypothesis (H_0), which implies that there is no significant difference between the observed data and the theoretical distribution. Therefore, the distributions successfully represent the schooling duration data. Although all three distributions show satisfactory results in terms of p-values, the distribution parameters provide additional insight into the shape and characteristics of the data. For example, the GEV distribution exhibits negative shape parameters in both clusters, which corresponds to a left-skewed distribution—an indication that a majority of districts in both clusters have relatively longer schooling durations, but some extremely low values exist. This is particularly evident in Cluster 2. The Nakagami distribution, characterized by a high shape parameter, indicates a narrow and symmetric distribution with low skewness in Cluster 1. However, in Cluster 2, the lower shape parameter reflects higher variability, which aligns with the broader spread of schooling data in underserved districts. The Weibull distribution demonstrates flexibility in capturing both narrow and wide spreads, with high shape values indicating low variability in Cluster 1 and moderate variability in Cluster 2. The relatively high shape parameters also suggest a right-skewed distribution, especially in Cluster 2, where lower education levels dominate.

Overall, the combination of high p-values and interpretable parameter estimates (shape, scale, and location) support the conclusion that these three distributions are statistically appropriate for modeling schooling length. Their ability to reflect skewness, variability, and tail behavior in the data provides a strong basis for further modeling and policy interpretation in subsequent sections.

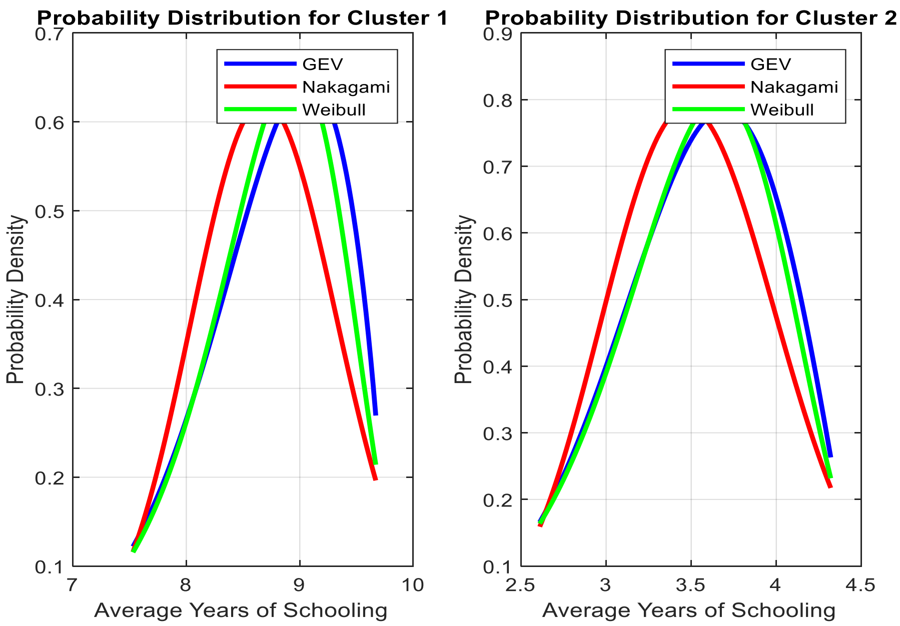

Figure 2. Probability Models of GEV, Nakagami, and Weibull for Clusters 1 and 2
Generated using MATLAB 2024b (Statistics and Machine Learning Toolbox)

Figure 2 presents the probability distribution models of Generalized Extreme Value (GEV), Nakagami, and Weibull applied to Cluster 1 and Cluster 2, which consist of districts with higher and lower average schooling durations, respectively. In Cluster 1, which includes areas with better educational access, the GEV distribution depicts a pattern with a longer left tail, indicating lower variability in schooling duration across some districts, even though the majority have higher values. This reflects that while educational access in Cluster 1 is generally better, there are still some regions facing challenges. The Nakagami distribution in Cluster 1 appears more concentrated around the median value, suggesting relatively low variability, indicating greater educational stability. The Weibull distribution is the most suitable for describing long and stable education durations in Cluster 1, as shown by its long right tail representing higher schooling lengths.

Conversely, Cluster 2 includes districts with lower average schooling durations. The GEV distribution effectively captures extreme values in this cluster, particularly in areas with very low educational attainment. Its long-left tail indicates many districts with persistently low schooling durations, though a few outliers show improvement. The Nakagami distribution captures high heterogeneity in Cluster 2 with a wider spread, indicating significant variation among districts. Meanwhile, the Weibull distribution is less suitable in Cluster 2, evidenced by its shorter tail and limited representation of variation. Overall, GEV and Nakagami distributions perform better in modeling schooling duration in Cluster 2, while Weibull is more appropriate for Cluster 1, which is characterized by a more stable distribution.

3.3 Criteria for the Best Distribution Model

After estimating parameters and conducting the Goodness-of-Fit (GoF) tests for each distribution, the next step is to evaluate the probability distribution models based on information criteria to determine which model best fits the schooling duration data in Cluster 1 and Cluster 2. **Table 3** presents the values of AIC, BIC, AICc, CAIC, and HQC for each probability distribution (GEV, Nakagami, and Weibull) in both clusters. AIC and AICc focus more on model fit with lighter penalties for the number of parameters, while BIC and CAIC impose greater penalties on model complexity, especially with larger sample sizes. HQC provides a balanced evaluation, particularly when dealing with larger sample sizes, with more moderate penalties compared to BIC. These information criteria values are used to determine which model best fits the data and to ensure that the chosen model is not overfit, which often occurs if the model is too complex.

Table 3. AIC, BIC, AICc, CAIC, and HQC (Clusters 1 and 2)

		Cluster 1			
Distribution	**AIC**	**BIC**	**AICc**	**CAIC**	**HQC**
GEV	31.6286	33.5458	34.0286	34.5076	31.4512
Nakagami	31.2364	32.5145	32.3273	32.5559	31.1181
Weibull	30.7368	32.0149	31.8277	32.0563	30.6184
		Cluster 2			
Distribution	**AIC**	**BIC**	**AICc**	**CAIC**	**HQC**
GEV	25.5903	27.5074	27.9903	28.4692	25.4128
Nakagami	24.5479	25.8260	25.6388	25.8674	24.4296
Weibull	24.3574	25.6355	25.4483	25.6769	24.2391

Based on the results in **Table 3**, it is evident that for Cluster 1, the Weibull distribution has the lowest information criteria values across all categories (AIC, BIC, AICc, CAIC, HQC), indicating that this model is the best fit for describing the schooling duration distribution in areas with better educational access. The lower AIC and AICc values for Weibull suggest that this model provides an optimal balance between fit and parameter complexity, while the lower BIC and CAIC values compared to other models indicate that Weibull is not overly complex for the sample size used. Therefore, Weibull is considered the best model for Cluster 1, as it captures the stable and long education duration patterns effectively.

For Cluster 2, the Weibull distribution also has the lowest information criteria values compared to GEV and Nakagami, particularly in AIC, AICc, and HQC. Although GEV and Nakagami provide good fits in the GoF tests, the Weibull distribution shows superior overall performance in capturing the schooling duration distribution pattern in Cluster 2, with lower complexity penalties. The low AIC and AICc values for Weibull in Cluster 2 indicate that this distribution can describe the extreme variability present in the data, but with a simpler model compared to the others.

Overall, based on the evaluation results using information criteria, Weibull is the best probability model for both clusters. While GEV and Nakagami perform well in capturing extreme patterns and high variability, Weibull provides a better balance between fit and complexity, making it a more optimal model for describing schooling duration distributions in regions with diverse educational characteristics, in both Cluster 1 and Cluster 2. These findings support the decision to use Weibull as the primary model in this analysis, as it offers a clearer representation of education duration patterns and their implications for educational policy in Papua.

3.4 Weibull Model of Education Duration in Papua Province (2010-2023) Based on Regional Clusters

The Weibull distribution has been identified as the most suitable model for describing the average schooling duration in Papua Province from 2010 to 2023, covering both Cluster 1 and Cluster 2**.** The selection of the Weibull distribution is based on its effectiveness in modeling schooling duration variables**,** particularly in capturing the variability in education across different regencies. This model is characterized by scale and shape parameters**,** which enable a more precise representation of educational trends and disparities.

To further illustrate the disparities in schooling duration across Papua Province, **Figure 3** presents a spatial clustering of the 26 regencies. The k-means clustering algorithm classifies the districts into two main groups based on their average years of schooling (2010–2023)**:**

- Cluster 1 (Blue): Represents regencies with higher average schooling duration, indicating better educational infrastructure and accessibility**.**
- Cluster 2 (Red): Represents districts with lower schooling duration, which suggests limited access to education, inadequate infrastructure, and educational disparities**.**

This clustering provides a clearer visualization of the regional disparities in education and serves as an important reference for targeted policy interventions. Although this study attempted to employ area-colored maps to enhance the spatial differentiation of clusters, the final map visualization was generated using MATLAB's `Mapping Toolbox`, which has limited support for full polygon-based area color rendering using shapefile data under certain attribute conditions. Specifically, the color assignment relies on the accuracy of region name matching in the shapefile attributes and the rendering capabilities of MATLAB's default plotting engine. As a result, the current visualization retains region boundaries and labeled cluster points, while polygon fill color may not appear due to technical limitations in the shapefile's attribute integration. Nonetheless, the clustering patterns remain visually interpretable and align with the analytical results presented in subsequent sections.

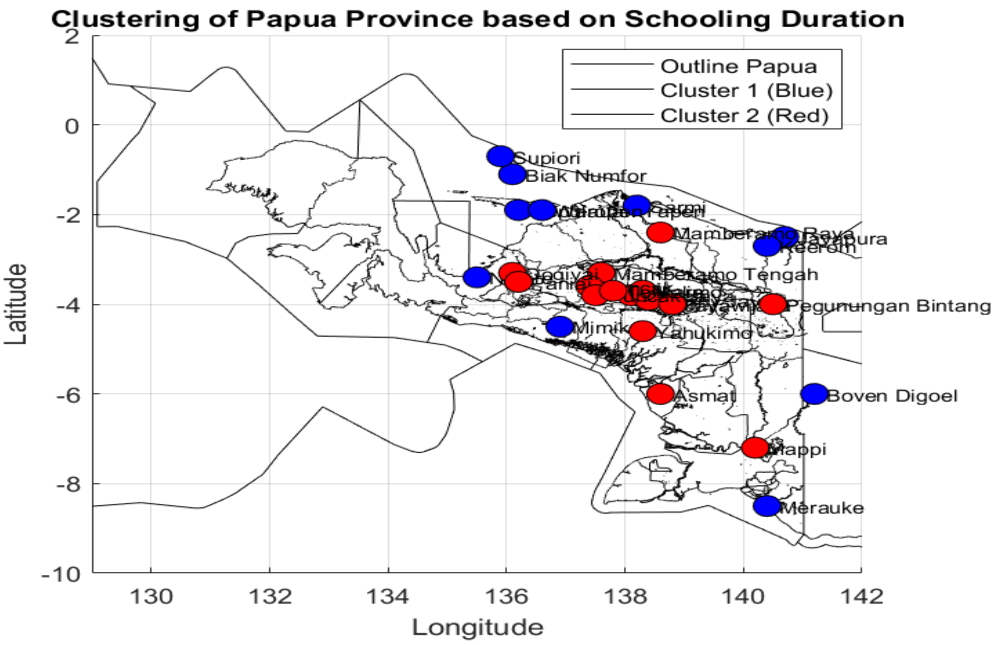

Figure 3. Clustering of Papua Province based on Schooling Duration
Generated using MATLAB 2024b (Mapping Toolbox)

Figure 3 highlights the spatial patterns of educational disparities in Papua Province, emphasizing the clear separation between the two clusters**.** The blue-marked regencies demonstrate stronger educational performance**,** while the red-marked regencies struggle with lower schooling durations. These visual insights reinforce the statistical findings presented in **Table 4**, which provides a quantitative comparison between the two clusters based on key statistical indicators such as mean, standard deviation, minimum, maximum, and median schooling duration.

Table 4. Descriptive Statistics Table

Cluster	Mean	Standard Deviation	Minimum	Maximum	Median
Cluster 1 (Higher schooling duration)	8.70	0.66	7.53	9.67	8.71
Cluster 2 (Lower schooling duration)	3.51	0.52	2.61	4.32	3.52

Table 4 provides statistical evidence of the disparities observed in **Figure 3**:

- Cluster 1 has a higher mean schooling duration (8.70 years) with a relatively low standard deviation (0.66), indicating homogeneity in educational outcomes. The minimum and maximum values range from 7.53 to 9.67 years, with a median of 8.71 years, suggesting that most regencies in this cluster align with or slightly exceed the national average schooling duration, which is approximately 8.5 years (BPS, 2023).
- Cluster 2 consists of regencies with lower schooling duration, exhibiting a mean of 3.51 years and a standard deviation of 0.52, which still reflects noticeable disparities in educational attainment. The distribution of schooling duration in this cluster ranges from 2.61 to 4.32 years, with a median of 3.52 years.

These findings underscore significant educational inequalities, highlighting the necessity of targeted policies in Cluster 2 to enhance education access and bridge disparities among regencies.

For Cluster 1, which includes districts with higher average schooling durations, such as Merauke, Jayapura, Nabire, Kepulauan Yapen, Biak Numfor, Mimika, Boven Digoel, Sarmi, Keerom, Waropen, and Supiori, the Weibull distribution has a shape parameter β of 16.1272 and a scale parameter η of 8.9931.

$$f(x) = \frac{16.1272}{8.9931}\left(\frac{x}{8.9931}\right)^{16.1272-1} \exp\left(-\left(\frac{x}{8.9931}\right)^{16.1272}\right) \tag{15}$$

The large shape parameter indicates that this distribution has a long tail, illustrating the variability among individuals in completing their education, even though the majority of the population in this area has a relatively long schooling duration. Meanwhile, the large-scale parameter signifies that the average schooling duration in Cluster 1 is relatively high. In other words, most districts in this cluster exhibit stability in access to and quality of education, reflecting better infrastructure and educational support compared to other regions. For Cluster 2, which includes areas with lower average schooling durations, such as Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, Puncak, Dogiyai, Jayawijaya, Paniai, Puncak Jaya, Mappi, Asmat, Yahukimo, Pegunungan Bintang, and Tolikara, the Weibull distribution also provides a good fit, but with different parameters. In this cluster, the shape parameter β of the Weibull distribution is 8.07662, while the scale parameter η is 3.73006.

$$f(x) = \frac{8.07662}{3.73006}\left(\frac{x}{3.73006}\right)^{8.07662-1} \exp\left(-\left(\frac{x}{3.73006}\right)^{8.07662}\right) \tag{16}$$

The smaller shape parameter compared to Cluster 1 indicates greater variability in educational duration among these districts. With a long right tail in the distribution, the Weibull model captures a more diverse pattern of schooling duration, where some individuals take significantly longer to complete their education. The lower scale parameter also reflects that the average schooling duration in Cluster 2 is generally lower, indicating that educational access in these areas is still limited and that there are significant challenges regarding educational infrastructure and the quality of teaching staff.

Overall, the Weibull distribution provides deep insights into the educational characteristic differences between Cluster 1 and Cluster 2. In Cluster 1, the Weibull model indicates stability and a long duration of education, with the distribution being more concentrated around higher schooling durations. Conversely, in Cluster 2, the Weibull model reveals a wider distribution with high variability, indicating disparities in access to and quality of education in this region. These findings have important implications for policymakers, particularly in designing educational improvement programs in Cluster 2 areas, such as Yahukimo, Nduga, and Pegunungan Bintang, which require more significant interventions in terms of accessibility and educational facilities.

3.5 Implications for Policy Based on Analysis Results

Based on the analysis of the probability distribution of average schooling duration in Papua Province, there are significant policy implications to improve the quality and accessibility of education. Cluster 1, which encompasses areas with better educational access and higher average schooling durations, indicates that policy interventions in this region can focus on enhancing the quality of education and infrastructure, such as teacher training and educational technology. Conversely, Cluster 2, which includes districts with lower average schooling durations and significant challenges in educational access, requires more targeted policies to improve physical infrastructure and enhance accessibility, such as building schools and transportation facilities in remote areas. The government and related institutions should allocate resources more effectively, particularly in areas that exhibit high educational variability, such as Yahukimo, Nduga, and Pegunungan Bintang. Sustainable policies, involving collaboration between the central government, regional authorities, and international agencies, are crucial to addressing existing educational disparities and achieving overall improvements in educational quality in Papua.

4. CONCLUSION

Based on the analysis of the Human Development Index (HDI) data regarding average schooling duration in Papua Province from 2010 to 2023, classified using the k-means clustering method, this study successfully identified two main clusters that reflect significant differences in educational characteristics across 26 districts. Cluster 1, consisting of districts with higher average schooling durations such as Merauke, Jayapura, and Nabire, shows a more uniform and stable educational distribution pattern. The Weibull distribution model applied to this cluster indicates better educational access, with the majority of the population able to complete education over a longer duration. In contrast, Cluster 2, which includes districts with lower average schooling durations such as Mamberamo Raya, Nduga, and Yahukimo, presents more serious challenges related to access and educational quality. The Weibull distribution is also suitable for this cluster, but with parameters reflecting greater educational variability, where some individuals take longer to complete their education, and the overall average schooling duration is lower compared to Cluster 1. This model demonstrates good Goodness-of-Fit (GoF) results, indicating that Weibull is the most appropriate model to describe educational distribution in both clusters. These findings are significant for policymakers, providing guidance for designing more effective and targeted educational interventions, particularly in Cluster 2, which still lags in access and quality. By leveraging these results, the government and related institutions are expected to allocate educational resources more efficiently, especially in districts requiring special attention, such as Yahukimo, Nduga, and Pegunungan Bintang, to enhance educational quality and achieve better educational equity in Papua.

AUTHOR CONTRIBUTIONS

Jonathan K. Wororomi: Conceptualization, Methodology, Formal Analysis, Writing – Original Draft, Supervision. Alvian M. Sroyer: Resources, Supervision, Writing – Review and Editing. Henderina Morin: Data Curation, Investigation, Validation. Felix Reba: Visualization, Software, Writing – Review and Editing. Ishak S. Beno: Methodology, Validation, Supervision. Oscar O. O. Wambrauw: Investigation, Project Administration, Writing – Review and Editing. All authors discussed the results and contributed to the preparation of the final manuscript.

FUNDING STATEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The entire research was self-funded by the authors.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Universitas Cenderawasih for the academic environment and administrative support during the course of this research. This study was conducted using the authors' resources without institutional funding.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this study.

REFERENCES

[1] R. Karagiannis and G. Karagiannis, "CONSTRUCTING COMPOSITE INDICATORS WITH SHANNON ENTROPY: THE CASE OF HUMAN DEVELOPMENT INDEX," *Socio-Economic Planning Sciences*, 2020, doi: https://doi.org/10.1016/j.seps.2019.03.007.

[2] C. Türe and Y. Türe, "A MODEL FOR THE SUSTAINABILITY ASSESSMENT BASED ON THE HUMAN DEVELOPMENT INDEX IN DISTRICTS OF MEGACITY ISTANBUL (TURKEY)," *Environment, Development and Sustainability*, 2021, doi: https://doi.org/10.1007/s10668-020-00735-9.

[3] Y. Jiang and C. Shi, "ESTIMATING SUSTAINABILITY AND REGIONAL INEQUALITIES USING AN ENHANCED SUSTAINABLE DEVELOPMENT INDEX IN CHINA," *Sustainable Cities and Society*, 2023, doi: https://doi.org/10.1016/j.scs.2023.104555.

[4] S. Kwatra, A. Kumar, and P. Sharma, "A CRITICAL REVIEW OF STUDIES RELATED TO CONSTRUCTION AND COMPUTATION OF SUSTAINABLE DEVELOPMENT INDICES," 2020. doi: https://doi.org/10.1016/j.ecolind.2019.106061.

[5] W. Afalia, I. Hamda, S. A. Adriana, A. F. Alamsyah, and N. L. Wafiroh, "DETERMINANTS OF HUMAN DEVELOPMENT INDEX IN PAPUA PROVINCE 2012-2021," *Wiga : Jurnal Penelitian Ilmu Ekonomi*, 2023, doi: https://doi.org/10.30741/wiga.v13i2.1076.

[6] D. P. Rahmawati, I. N. Budiantara, D. D. Prastyo, and M. A. D. Octavanny, "MODELING OF HUMAN DEVELOPMENT INDEX IN PAPUA PROVINCE USING SPLINE SMOOTHING ESTIMATOR IN NONPARAMETRIC REGRESSION," in *Journal of Physics: Conference Series*, 2021. doi: https://doi.org/10.1088/1742-6596/1752/1/012018.

[7] A. E. Ezugwu *et al.*, "A COMPREHENSIVE SURVEY OF CLUSTERING ALGORITHMS: STATE-OF-THE-ART MACHINE LEARNING APPLICATIONS, TAXONOMY, CHALLENGES, AND FUTURE RESEARCH PROSPECTS," 2022. doi: https://doi.org/10.1016/j.engappai.2022.104743.

[8] A. Rajabi, M. Eskandari, M. J. Ghadi, L. Li, J. Zhang, and P. Siano, "A COMPARATIVE STUDY OF CLUSTERING TECHNIQUES FOR ELECTRICAL LOAD PATTERN SEGMENTATION," *Renewable and Sustainable Energy Reviews*, 2020, doi: https://doi.org/10.1016/j.rser.2019.109628.

[9] J. Li and J. Liu, "A MODIFIED EXTREME VALUE PERSPECTIVE ON BEST-PERFORMANCE LIFE EXPECTANCY," *Journal of Population Research*, 2020, doi: https://doi.org/10.1007/s12546-020-09248-8.

[10] M. S. Dhanoa *et al.*, "A STRATEGY FOR MODELLING HEAVY-TAILED GREENHOUSE GASES (GHG) DATA USING THE GENERALISED EXTREME VALUE DISTRIBUTION: ARE WE OVERESTIMATING GHG FLUX USING THE SAMPLE MEAN?," *Atmospheric Environment*, 2020, doi: https://doi.org/10.1016/j.atmosenv.2020.117500.

[11] H. Omrani, A. Alizadeh, and M. Amini, "A NEW APPROACH BASED ON BWM AND MULTIMOORA METHODS FOR CALCULATING SEMI-HUMAN DEVELOPMENT INDEX: AN APPLICATION FOR PROVINCES OF IRAN," *Socio-Economic Planning Sciences*, 2020, doi: https://doi.org/10.1016/j.seps.2019.02.004.

[12] B. Giles-Corti, M. Lowe, and J. Arundel, "ACHIEVING THE SDGS: EVALUATING INDICATORS TO BE USED TO BENCHMARK AND MONITOR PROGRESS TOWARDS CREATING HEALTHY AND SUSTAINABLE CITIES," 2020. doi: https://doi.org/10.1016/j.healthpol.2019.03.001.

[13] S. A. Takyi, O. Amponsah, M. O. Asibey, and R. A. Ayambire, "AN OVERVIEW OF GHANA'S EDUCATIONAL SYSTEM AND ITS IMPLICATION FOR EDUCATIONAL EQUITY," *International Journal of Leadership in Education*, 2021, doi: 10.1080/13603124.2019.1613565.

[14] A. Sinha, D. Balsalobre-Lorente, M. W. Zafar, and M. M. Saleem, "ANALYZING GLOBAL INEQUALITY IN ACCESS TO ENERGY: DEVELOPING POLICY FRAMEWORK BY INEQUALITY DECOMPOSITION," *Journal of Environmental Management*, 2022, doi: https://doi.org/10.1016/j.jenvman.2021.114299.

[15] T. Ladi, A. Mahmoudpour, and A. Sharifi, "ASSESSING IMPACTS OF THE WATER POVERTY INDEX COMPONENTS ON THE HUMAN DEVELOPMENT INDEX IN IRAN," *Habitat International*, 2021, doi: https://doi.org/10.1016/j.habitatint.2021.102375.

[16] R. Zriaa and S. Amali, "A COMPARATIVE STUDY BETWEEN K-NEAREST NEIGHBORS AND K-MEANS CLUSTERING TECHNIQUES OF COLLABORATIVE FILTERING IN E-LEARNING ENVIRONMENT," in *Lecture Notes in Networks and Systems*, 2021. doi: https://doi.org/10.1007/978-3-030-66840-2_21.

[17] T. H. T. Nguyen, D. T. Dinh, S. Sriboonchitta, and V. N. Huynh, "A METHOD FOR K-MEANS-LIKE CLUSTERING OF CATEGORICAL DATA," *Journal of Ambient Intelligence and Humanized Computing*, 2023, doi: 10.1007/s12652-019-01445-5.

[18] A. Vouros and E. Vasilaki, "A SEMI-SUPERVISED SPARSE K-MEANS ALGORITHM," *Pattern Recognition Letters*, 2021, doi: https://doi.org/10.1016/j.patrec.2020.11.015.

[19] M. K. Gupta and P. Chandra, "AN EMPIRICAL EVALUATION OF K-MEANS CLUSTERING ALGORITHM USING DIFFERENT DISTANCE/SIMILARITY METRICS," in *Lecture Notes in Electrical Engineering*, 2020. doi: https://doi.org/10.1007/978-3-030-30577-2_79.

[20] D. A. Lusia, I. Salsabila, H. Kusdarwati, and S. Astutik, "CLUSTERING DISTRICTS/CITIES IN EAST JAVA PROVINCE BASED ON HIV CASES USING K-MEANS, AGNES, AND ENSEMBLE," *Barekeng: Jurnal Ilmu Matematika dan Terapan*, vol. 19, no. 1, pp. 63–72, March 2025, doi: https://doi.org/10.30598/barekengvol19iss1pp63-72.

[21] Z. A. Zain, N. Saif, M. Mussafi, and E. D. Supandi, "CLUSTER ANALYSIS OF K-MEANS AND WARD METHOD IN FORMING A ROBUST PORTFOLIO : AN EMPIRICAL STUDY OF JAKARTA ISLAMIC INDEX," *Barekeng*: Jurnal Ilmu Matematika dan Terapan, vol. 19, no. 1, pp. 537–546, March 2025,doi: https://doi.org/10.30598/barekengvol19iss1pp537-546

[22] M. Daraghmeh, A. Agarwal, and Y. Jararweh, "AN ENSEMBLE CLUSTERING APPROACH FOR MODELING HIDDEN CATEGORIZATION PERSPECTIVES FOR CLOUD WORKLOADS," *Cluster Computing*, 2024, doi: https://doi.org/10.21203/rs.3.rs-3246818/v1.

[23] Z. Halim, H. M. Sargana, Aadam, Uzma, and M. Waqas, "CLUSTERING OF GRAPHS USING PSEUDO-GUIDED RANDOM WALK," *Journal of Computational Science*, 2021, doi: https://doi.org/10.1016/j.jocs.2020.101281.

[24] Y. M. Kantar and I. Usta, "ANALYSIS OF THE UPPER-TRUNCATED WEIBULL DISTRIBUTION FOR WIND SPEED," *Energy Conversion and Management*, vol. 96, pp. 81–88, 2015, doi: https://doi.org/10.1016/j.enconman.2015.02.063.

[25] C. Ozay and M. S. Celiktas, "STATISTICAL ANALYSIS OF WIND SPEED USING TWO-PARAMETER WEIBULL DISTRIBUTION IN ALAÇATI REGION," *Energy Conversion and Management*, vol. 121, pp. 49–54, 2016, doi: https://doi.org/10.1016/j.enconman.2016.05.026.

[26] T. Arslan, Y. M. Bulut, and A. Altin Yavuz, "COMPARATIVE STUDY OF NUMERICAL METHODS FOR DETERMINING WEIBULL PARAMETERS FOR WIND ENERGY POTENTIAL," *Renewable and Sustainable Energy Reviews*, vol. 40, pp. 820–825, 2014, doi: https://doi.org/10.1016/j.rser.2014.08.009.

[27] P. A. Costa Rocha, R. C. de Sousa, C. F. de Andrade, and M. E. V. da Silva, "COMPARISON OF SEVEN NUMERICAL METHODS FOR DETERMINING WEIBULL PARAMETERS FOR WIND ENERGY GENERATION IN THE NORTHEAST REGION OF BRAZIL," *Applied Energy*, vol. 89, no. 1, pp. 395–400, 2012, doi: https://doi.org/10.1016/j.apenergy.2011.08.003.

[28] J. Then, F. J. Permana, and B. Yong, "PARAMETER ESTIMATION OF LOGNORMAL AND PARETO TYPE I DISTRIBUTIONS USING FREQUENTIST AND BAYESIAN INFERENCES," *Barekeng: Jurnal Ilmu Matematika dan Terapan*, vol. 19, no. 1, pp. 141–152, March 2025, doi: https://doi.org/10.30598/barekengvol19iss1pp141-152.

[29] E. Salvador, M. B. Gonzalez, F. Campabadal, J. Martin-Martinez, R. Rodriguez, and E. Miranda, "ASSESSMENT OF THE VARIABILITY OF THE I-V CHARACTERISTIC OF HFO2-BASED RESISTIVE SWITCHING DEVICES AND ITS SIMULATION USING THE QUASI-STATIC MEMDIODE MODEL," *Solid-State Electronics*, vol. 206, no. May, p. 108667, 2023, doi: https://doi.org/10.1016/j.sse.2023.108667.

[30] M. H. Samuh and A. M. Salhab, "DISTRIBUTION OF SQUARED SUM OF PRODUCTS OF INDEPENDENT NAKAGAMI-M RANDOM VARIABLES," *Communications in Statistics: Simulation and Computation*, 2023, doi: https://doi.org/10.1080/03610918.2023.2234668.

[31] H. A. Bayoud, "TESTS OF NORMALITY: NEW TEST AND COMPARATIVE STUDY," *Communications in Statistics: Simulation and Computation*, vol. 50, no. 12, pp. 4442–4463, 2021, doi: https://doi.org/10.1080/03610918.2019.1643883.

[32] D. K. Wijekularathna, A. B. W. Manage, and S. M. Scariano, "POWER ANALYSIS OF SEVERAL NORMALITY TESTS: A MONTE CARLO SIMULATION STUDY," *Communications in Statistics: Simulation and Computation*, vol. 51, no. 3, pp. 757–773, 2020, doi: https://doi.org/10.1080/03610918.2019.1658780.

[33] W. Zheng, D. Lai, and K. L. Gould, "A SIMULATION STUDY OF A CLASS OF NONPARAMETRIC TEST STATISTICS: A CLOSE LOOK OF EMPIRICAL DISTRIBUTION FUNCTION-BASED TESTS," *Communications in Statistics: Simulation and Computation*, vol. 52, no. 3, pp. 1133–1148, 2023, doi: https://doi.org/10.1080/03610918.2021.1874987.

[34] M. Krit, O. Gaudoin, M. Xie, and E. Remy, "SIMPLIFIED LIKELIHOOD BASED GOODNESS-OF-FIT TESTS FOR THE WEIBULL DISTRIBUTION," *Communications in Statistics: Simulation and Computation*, vol. 45, no. 3, pp. 920–951, 2016, doi: https://doi.org/10.1080/03610918.2013.879889.

[35] M. H. Ouahabi, H. Elkhachine, F. Benabdelouahab, and A. Khamlichi, "COMPARATIVE STUDY OF FIVE DIFFERENT METHODS OF ADJUSTMENT BY THE WEIBULL MODEL TO DETERMINE THE MOST ACCURATE METHOD OF ANALYZING ANNUAL VARIATIONS OF WIND ENERGY IN TETOUAN - MOROCCO," *Procedia Manufacturing*, vol. 46, no. 2019, pp. 698–707, 2020, doi: https://doi.org/10.1016/j.promfg.2020.03.099.

[36] A. K. Azad, M. G. Rasul, M. M. Alam, S. M. Ameer Uddin, and S. K. Mondal, "ANALYSIS OF WIND ENERGY CONVERSION SYSTEM USING WEIBULL DISTRIBUTION," *Procedia Engineering*, vol. 90, pp. 725–732, 2014, doi: https://doi.org/10.1016/j.proeng.2014.11.803.

[37] M. Nassar, A. Alzaatreh, M. Mead, and O. Abo-Kasem, "ALPHA POWER WEIBULL DISTRIBUTION: PROPERTIES AND APPLICATIONS," *Communications in Statistics - Theory and Methods*, vol. 46, no. 20, pp. 10236–10252, 2017, doi: https://doi.org/10.1080/03610926.2016.1231816.

[38] M. Krit, O. Gaudoin, and E. Remy, "GOODNESS-OF-FIT TESTS FOR THE WEIBULL AND EXTREME VALUE DISTRIBUTIONS: A REVIEW AND COMPARATIVE STUDY," *Communications in Statistics: Simulation and Computation*, vol. 50, no. 7, pp. 1888–1911, 2021, doi: https://doi.org/10.1080/03610918.2019.1594292.

[39] P. K. Chaurasiya, S. Ahmed, and V. Warudkar, "STUDY OF DIFFERENT PARAMETERS ESTIMATION METHODS OF WEIBULL DISTRIBUTION TO DETERMINE WIND POWER DENSITY USING GROUND BASED DOPPLER SODAR INSTRUMENT," *Alexandria Engineering Journal*, vol. 57, no. 4, pp. 2299–2311, 2018, doi: https://doi.org/10.1016/j.aej.2017.08.008.

[40] A. M. Basheer, "ALPHA POWER INVERSE WEIBULL DISTRIBUTION WITH RELIABILITY APPLICATION," *Journal of Taibah University for Science*, vol. 13, no. 1, pp. 423–432, 2019, doi: https://doi.org/10.1080/16583655.2019.1588488.

[41] A. O. Lima, G. B. Lyra, M. C. Abreu, J. F. Oliveira-Júnior, M. Zeri, and G. Cunha-Zeri, "EXTREME RAINFALL EVENTS OVER RIO DE JANEIRO STATE, BRAZIL: CHARACTERIZATION USING PROBABILITY DISTRIBUTION FUNCTIONS AND CLUSTERING ANALYSIS," *Atmospheric Research*, vol. 247, no. March 2020, p. 105221, 2021, doi: https://doi.org/10.1016/j.atmosres.2020.105221.

[42] A. Sulis, R. Cozza, and A. Annis, "EXTREME WAVE ANALYSIS METHODS IN THE GULF OF CAGLIARI (SOUTH SARDINIA, ITALY)," *Ocean and Coastal Management*, vol. 140, pp. 79–87, 2017, doi: https://doi.org/10.1016/j.ocecoaman.2017.02.023.

[43] Q. Han, S. Ma, T. Wang, and F. Chu, "KERNEL DENSITY ESTIMATION MODEL FOR WIND SPEED PROBABILITY DISTRIBUTION WITH APPLICABILITY TO WIND ENERGY ASSESSMENT IN CHINA," *Renewable and Sustainable Energy Reviews*, vol. 115, no. September, p. 109387, 2019, doi: https://doi.org/10.1016/j.rser.2019.109387.

[44] K. Mohammadi, O. Alavi, and J. G. McGowan, "USE OF BIRNBAUM-SAUNDERS DISTRIBUTION FOR ESTIMATING WIND SPEED AND WIND POWER PROBABILITY DISTRIBUTIONS: A REVIEW," *Energy Conversion and Management*, vol. 143, pp. 109–122, 2017, doi: https://doi.org/10.1016/j.enconman.2017.03.083.

[45] K. S. Guedes, C. F. de Andrade, P. A. C. Rocha, R. dos S. Mangueira, and E. P. de Moura, "PERFORMANCE ANALYSIS OF METAHEURISTIC OPTIMIZATION ALGORITHMS IN ESTIMATING THE PARAMETERS OF SEVERAL WIND SPEED DISTRIBUTIONS," *Applied Energy*, vol. 268, no. March, p. 114952, 2020, doi: https://doi.org/10.1016/j.apenergy.2020.114952.

[46] C. Jung and D. Schindler, "GLOBAL COMPARISON OF THE GOODNESS-OF-FIT OF WIND SPEED DISTRIBUTIONS," *Energy Conversion and Management*, vol. 133, pp. 216–234, 2017, doi: https://doi.org/10.1016/j.enconman.2016.12.006.

[47] M. A. ul Haq, G. S. Rao, M. Albassam, and M. Aslam, "MARSHALL–OLKIN POWER LOMAX DISTRIBUTION FOR MODELING OF WIND SPEED DATA," *Energy Reports*, vol. 6, no. May, pp. 1118–1123, 2020, doi: https://doi.org/10.1016/j.egyr.2020.04.033.

[48] A. M. Basheer, "ALPHA POWER INVERSE WEIBULL DISTRIBUTION WITH RELIABILITY APPLICATION," *Journal of Taibah University for Science*, vol. 13, no. 1, pp. 423–432, 2019, doi: https://doi.org/10.1080/16583655.2019.1588488.

[49] M. Mansoor, M. H. Tahir, G. M. Cordeiro, S. B. Provost, and A. Alzaatreh, "THE MARSHALL-OLKIN LOGISTIC-EXPONENTIAL DISTRIBUTION," *Communications in Statistics - Theory and Methods*, vol. 48, no. 2, pp. 220–234, 2019, doi: https://doi.org/10.1080/03610926.2017.1414254.

[50] I. Pobočíková, Z. Sedliačková, and M. Michalková, "APPLICATION OF FOUR PROBABILITY DISTRIBUTIONS FOR WIND SPEED MODELING," *Procedia Engineering*, vol. 192, pp. 713–718, 2017, doi: https://doi.org/10.1016/j.proeng.2017.06.123.

[51] T. B. M. J. Ouarda, C. Charron, and F. Chebana, "REVIEW OF CRITERIA FOR THE SELECTION OF PROBABILITY DISTRIBUTIONS FOR WIND SPEED DATA AND INTRODUCTION OF THE MOMENT AND L-MOMENT RATIO DIAGRAM METHODS, WITH A CASE STUDY," *Energy Conversion and Management*, vol. 124, pp. 247–265, 2016, doi: https://doi.org/10.1016/j.enconman.2016.07.012.