Joinforcs: Journal of informatics and information system
Volume 1, (2) 2026, p. 64-70
E-ISSN : XXXX-XXXX

Expert System for Poverty Level Assessment Using the Certainty Factor Method

Riva Arieftaa[a,1,*], Aulia Putri Zahro[a,2]
[a] Department of Information System, STMIK Widya Cipta Dharma, Samarinda, Indonesia
[1] 2341037@wicida.ac.id, [2] 2341038@wicida.ac.id
* corresponding author

Article Info

Article History:

Received Jan 25, 2026
Revised Feb 19, 2026
Accepted Mar 06, 2026

Keywords:
Certainty Factor, Expert System, Human Development Index, Poverty Level Assessment, Socio-Economic

Abstract

Poverty is a multidimensional socio-economic problem that remains a major challenge in Indonesia's development process. The complexity of poverty is influenced by various interconnected factors, including economic conditions, education, human development, access to basic services, and employment opportunities. Accurate identification of poverty levels is therefore essential to support effective policy formulation and targeted poverty alleviation programs. This study aims to develop an expert system for assessing poverty levels in regencies and cities throughout Indonesia using the Certainty Factor (CF) method. The CF approach was selected because it is capable of representing expert knowledge and handling uncertainty in decision-making processes. The proposed system utilizes seven socio-economic indicators that are widely recognized as determinants of poverty, namely the percentage of the poor population, mean years of schooling, per capita expenditure, the Human Development Index (HDI), access to decent sanitation, access to safe drinking water, and the open unemployment rate. Expert certainty values were established through an extensive literature review and incorporated into the system as knowledge-based rules. These values were then combined with user certainty values derived from observational data collected from 514 regencies and cities across Indonesia. The integration of expert knowledge and empirical data enables the system to generate a poverty-level assessment that is both transparent and measurable. The performance of the developed expert system was evaluated by comparing the classification results with actual poverty data from the 514 regencies and cities. The evaluation results demonstrate that the system achieved an accuracy of 90.66%, precision of 58.75%, recall of 75.81%, and an F1-score of 66.20%. These findings indicate that the proposed approach is effective in identifying regions with varying poverty levels and has strong potential as a decision-support tool. The developed system can assist policymakers and stakeholders in prioritizing interventions, allocating resources more efficiently, and supporting evidence-based poverty reduction strategies across Indonesia.

1. Introduction

Poverty remains one of the fundamental issues that Indonesia still faces today. It can be defined as a state of limited capability to fulfill a decent standard of living, encompassing aspects of income, education, health, and access to basic facilities [1]. According to data from Statistics Indonesia (BPS), poverty rates across various regencies and cities in Indonesia exhibit significant variations, which are influenced by multiple factors such as education levels, per capita expenditure, the Human Development Index (HDI),

Joinforcs: Journal of informatics and information system
Volume 1, (2) 2026, p. 64-70
E-ISSN : XXXX-XXXX

access to decent sanitation and safe drinking water, and the open unemployment rate [2]. The complexity and interconnectedness of these factors make the process of determining a region's poverty status far from simple, thereby necessitating a system that can help evaluate poverty levels in a more systematic and measurable manner [3].

Conventional poverty assessments are generally conducted by setting a threshold on a single primary indicator, such as the percentage of the poor population (P0) or the poverty line [4]. This approach has limitations as it does not yet comprehensively account for the uncertainty and interrelatedness among indicators. An expert system offers a solution by mimicking the decision-making process of an expert based on a set of rules and facts [5]. One commonly used method to handle uncertainty in expert systems is the Certainty Factor (CF) method, which was introduced by Shortliffe and Buchanan during the development of the MYCIN system [6]. The CF method allows for the representation of the degree of belief in a hypothesis based on available evidence, making it highly suitable for poverty assessment cases that involve data with a certain degree of uncertainty.

Several previous studies have applied the Certainty Factor method to various diagnosis and classification problems, including human disease diagnosis [7], plant pest and disease diagnosis [8], fish disease diagnosis [9], and the early detection of illnesses in toddlers [10]. The results of these studies demonstrate that the CF method is effective in representing the uncertainty of expert knowledge and yields good accuracy compared to direct expert diagnoses [10]. However, the specific application of the CF method for assessing poverty levels based on socio-economic indicators at the regency and city level remains limited. Research related to poverty has historically leaned toward data-driven approaches, such as regression analysis or machine learning [11]. Consequently, the classification results are often less transparent when explained to users, especially local-level policymakers.

Based on these issues, this study aims to develop an expert system for assessing poverty levels using the Certainty Factor method by leveraging socio-economic data from regencies and cities in Indonesia. This data includes indicators such as the percentage of the poor population, mean years of schooling, per capita expenditure, the human development index, access to decent sanitation and drinking water, and the open unemployment rate. The contribution of this research lies in formulating the rules and certainty factor values for each indicator based on a literature review, as well as testing the system's classification results against actual data to measure the accuracy of the developed system. It is expected that this system can serve as a transparent and easily understood decision-support tool for stakeholders in identifying regions with high poverty levels.

2. Research Methodology

The research workflow is illustrated in Figure 1.

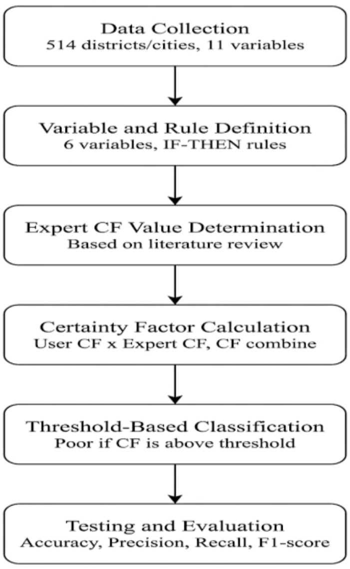

Figure 1. Research Workflow

Joinforcs: Journal of informatics and information system
Volume 1, (2) 2026, p. 64-70
E-ISSN : XXXX-XXXX

This research was conducted through several stages: data collection, determination of variables and rules, determination of expert certainty factor values, implementation of certainty factor calculations, and testing and evaluation of results. The data used in this study is secondary data consisting of 514 data points from regencies and cities across Indonesia. It comprises 11 input variables, including the percentage of the poor population (P_0), mean years of schooling for the population aged 15+, adjusted per capita expenditure, the Human Development Index (HDI), life expectancy, the percentage of households with access to decent sanitation, the percentage of households with access to decent drinking water, the open unemployment rate (TPT), the labor force participation rate (TPAK), and GRDP at constant prices, as well as one target variable in the form of a poverty classification label (0 = not poor, 1 = poor).

Out of the 11 input variables, six were selected as parameters for the expert system based on their relevance to poverty conditions: the percentage of the poor population (P0), mean years of schooling, the Human Development Index (HDI), the percentage of access to decent sanitation, the percentage of access to decent drinking water, and the open unemployment rate (TPT). Each variable is categorized into three classes (low, medium, high) based on threshold values determined through descriptive statistical analysis of the data. Furthermore, IF-THEN rules were constructed to connect the condition of each variable with the poverty level hypothesis, as shown in Table 1.

Table 1. Rules and Expert Certainty Factor Values

No.	Variable	Condition	Category	User CF	Expert CF
1	Percentage of poor population (P0)	P0 > 15%	High	0.8	0.9
2	Percentage of poor population (P0)	10% < P0 ≤ 15%	Medium	0.5	0.9
3	Percentage of poor population (P0)	P0 ≤ 10%	Low	0.1	0.9
4	Mean years of schooling	< 8 years	Low	0.8	0.7
5	Mean years of schooling	8-9,5 years	Medium	0.5	0.7
6	Mean years of schooling	≥ 9,5 years	High	0.1	0.7
7	Human Development Index (HDI)	< 65	Low	0.8	0.8
8	Human Development Index (HDI)	65-70	Medium	0.5	0.8
9	Human Development Index (HDI)	≥ 70	High	0.1	0.8
10	Access to decent sanitation	< 60%	Low	0.8	0.5
11	Access to decent sanitation	60-80%	Medium	0.4	0.5
12	Access to decent sanitation	≥ 80%	High	0.1	0.5
13	Access to decent drinking water	< 70%	Low	0.8	0.5
14	Access to decent drinking water	70-85%	Medium	0.4	0.5
15	Access to decent drinking water	≥ 85%	High	0.1	0.5
16	Open unemployment rate (TPT)	> 7%	High	0.7	0.6

Joinforcs: Journal of informatics and information system
Volume 1, (2) 2026, p. 64-70
E-ISSN : XXXX-XXXX

No.	Variable	Condition	Category	User CF	Expert CF
17	Open unemployment rate (TPT)	5-7%	Medium	0.4	0.6
18	Open unemployment rate (TPT)	≤ 5%	Low	0.1	0.6

The Certainty Factor (CF) method utilizes two types of values: the expert CF (CF expert) and the user CF (CF user). The user CF is obtained by mapping the observed data values into low, medium, or high categories for each variable, whereas the expert CF is determined based on the level of confidence regarding the contribution of each indicator to poverty conditions. The rule CF value is calculated using the following equation:

$$CF(rule) = CF[user] \times CF[pakar] \ldots (1)$$

To combine the CF values from multiple rules into a single final CF value, the following combination equation is applied:

$$CFcombine(CF1, CF2) = CF1 + CF2 \times (1 - CF1), \text{ untuk } CF1,\ CF2 \geq 0 \ldots (2)$$

The CF calculation was implemented using the Python programming language on the Google Colaboratory platform. Each regency/city is classified as "Poor" if the final CF value meets or exceeds a certain threshold, and "Not Poor" otherwise. The threshold was determined by testing several values (ranging from 0.80 to 0.95) against the data, and the value yielding the highest accuracy was selected. The system's classification results were then compared with the actual labels in the dataset using a confusion matrix to calculate accuracy, precision, recall, and F1-score.

3. Results and Discussion

Testing was conducted on 514 data points from regencies and cities in Indonesia using six variables as parameters for the expert system, namely the percentage of the poor population, mean years of schooling, the human development index, access to decent sanitation, access to decent drinking water, and the open unemployment rate. Prior to the final testing, an experiment was carried out to determine the best classification threshold value by testing several options: 0.80, 0.85, 0.90, 0.92, and 0.95. The comparison results of these five threshold values are presented in Table 2.

Table 2. Comparison of Testing Results Across Various Threshold Values

Threshold	Akurasi	Precision	Recall	F1-Score
0,80	52,92%	20,20%	98,39%	**33,52%**
0,85	66,93%	26,32%	96,77%	**41,38%**
0,90	78,02%	34,55%	91,94%	**50,22%**
0,92	82,30%	39,42%	87,10%	**54,27%**
0,95	90,66%	58,75%	75,81%	**66,20%**

Based on Table II, the threshold value of 0.95 yields the highest accuracy compared to the other threshold values, reaching 90.66%, with a precision of 58.75%, recall of 75.81%, and an F1-score of 66.20%. Therefore, the threshold value of 0.95 was selected as the final classification threshold for the developed expert system. The system's classification results at the 0.95 threshold are presented as a confusion matrix in Figure 2.

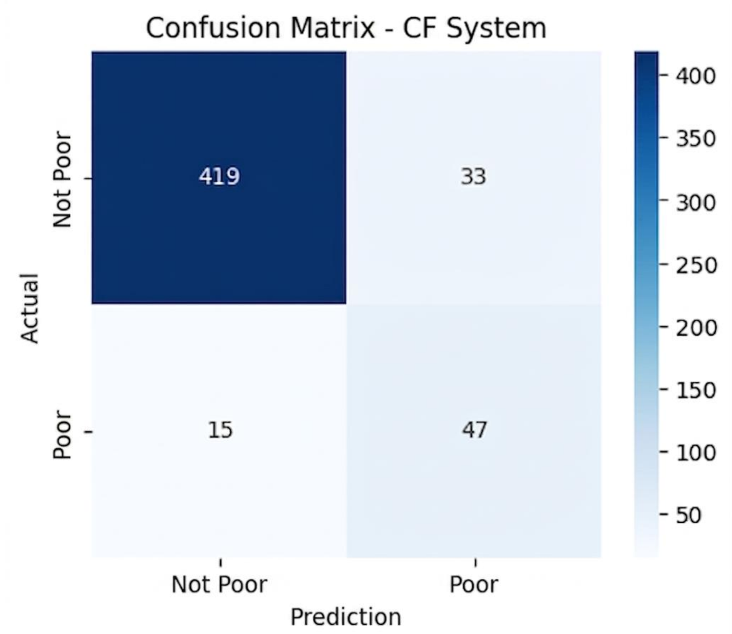

Figure 2. Confusion Matrix of System Classification Results at CF Threshold = 0.95

Figure 2 shows that out of 452 regencies/cities labeled as "Not Poor," the system successfully classified 419 correctly, while 33 were misclassified as "Poor" (false positives). Meanwhile, out of 62 regencies/cities labeled as "Poor," the system correctly identified 47, and 15 were misclassified as "Not Poor" (false negatives). The high number of false positives indicates that several regencies/cities with socio-economic conditions near the poverty threshold tend to be classified as "Poor" by the system, even though the actual data indicates otherwise. As an illustration, Table III displays examples of the certainty factor calculation results for ten sample regencies/cities, along with the system classification and actual classification results.

Table 3. Examples of Certainty Factor Calculation Results on Sample Data

No.	Province	Regency/City	CF Value	System Prediction	Actual Classification
1	East Nusa Tenggara	Manggarai	0.947	Not Poor	Poor
2	Papua	Yahukimo	0.985	Poor	Poor
3	Gorontalo	Gorontalo	0.947	Not Poor	Not Poor
4	Riau Islands	Tanjung Pinang City	0.466	Not Poor	Not Poor
5	Papua	Tolikara	0.985	Poor	Poor
6	Lampung	Way Kanan	0.934	Not Poor	Not Poor
7	Central Java	Pati	0.811	Not Poor	Not Poor
8	Papua	Yalimo	0.985	Poor	Poor

Joinforcs: Journal of informatics and information system
Volume 1, (2) 2026, p. 64-70
E-ISSN : XXXX-XXXX

No.	Province	Regency/City	CF Value	System Prediction	Actual Classification
9	Kalimantan Barat	Kubu Raya	0.894	Not Poor	Not Poor
10	East Java	Ngawi	0.919	Not Poor	Not Poor

As an example, in Table 3, Manggarai Regency (East Nusa Tenggara) has a CF value of 0.947, slightly below the 0.95 threshold, causing it to be classified as "Not Poor" by the system, even though the actual data classifies the regency as "Poor." This case demonstrates that a very narrow threshold value at high CF scores (above 0.90) can lead to misclassifications for regencies and cities with socio-economic conditions sitting on the borderline.

Overall, the expert system developed using the Certainty Factor method is capable of classifying the poverty levels of regencies and cities with a reasonably good accuracy of 90.66%. However, the recall of 75.81% indicates that the system still misses a portion of regencies and cities that should be classified as "Poor," while the precision of 58.75% points to a tendency of the system to overestimate the "Poor" classification. This could be caused by the class imbalance within the dataset (where the "Poor" class accounts for only about 12% of the total data) and the expert CF values, which remain preliminary estimates derived from the literature review. Further validation with experts in the socio-economic field is required to refine the precision of the CF values for each rule, thereby enabling the system's classification results to align more closely with actual conditions.

4. Conclusion

This study developed an expert system for assessing the poverty levels of regencies and cities in Indonesia using the Certainty Factor method based on six socio-economic indicators: the percentage of the poor population, mean years of schooling, the human development index, access to decent sanitation, access to decent drinking water, and the open unemployment rate. The testing results on 514 regency/city data points show that the system is capable of classifying poverty levels with an accuracy of 90.66% at a CF threshold of 0.95, demonstrating that the Certainty Factor method can be utilized as a transparent approach to assist in identifying regions with high poverty levels. Future research can be expanded by addressing the data imbalance problem, validating expert CF values through direct interviews with socio-economic experts, and comparing the results with other uncertainty handling methods, such as Dempster-Shafer or fuzzy logic.

REFERENCES

[1] A. Andriani, *Expert System Programming*. Yogyakarta: Mediakom, 2016.

[2] A. Riadi, "Application of the Certainty Factor Method for a Website-Based Expert System to Diagnose Carp Diseases (Case Study: UPTD Fish Seed Center, Binjai City)," Kumpulan Karya Ilmiah Mahasiswa Fakultas Sains dan Teknologi, vol. 3, no. 1, pp. 41–41, 2021.

[3] Isna Isnaini, I. K. Iqbal, And M. I. Ihsan, " Application Of The Certainty Factor Method For Diagnose Palm Oil Disease Web-Based ", J. Tek. Inform. (Jutif), Vol. 3, No. 3, Pp. 581–590, Jun. 2022, doi: https://doi.org/10.20884/1.Jutif.2022.3.3.242

[4] A. Riadi, "Application of the Certainty Factor Method for an Expert System to Diagnose Diabetes Mellitus at RSUD Bumi Panua, Pohuwato Regency," *ILKOM Jurnal Ilmiah*, vol. 9, no. 3, pp. 309-316, 2017, doi: https://doi.org/10.33096/ilkom.v9i3.161.309-316.

[5] D. Adellia, A. C. Siregar, and S. P. Alkadri, "Application of the Certainty Factor Method in an Expert System for Diagnosing Pests and Diseases in Tomato Plants," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 8, no. 3, pp. 451–458, 2022, doi: https://doi.org/10.26418/jp.v8i3.54143.

[6] G. Hamidi, "Comparative Analysis of Expert Systems Using the Certainty Factor Method and the Dempster-Shafer Method on Rabbit Diseases," *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 27–31, 2017, doi: 10.26418/justin.v5i1.18129.Kata, S., Pabboju, S., Babu, V., & Medishetti, A. (2018). Snap and split: an android application for bill payment using tesseract OCR. *International Journal of Engineering & Technology*, *7*(4.5), 634. doi: https://doi.org/10.14419/ijet.v7i4.5.21175

[7] M. Arhami, *Basic Concepts of Expert Systems*. Yogyakarta: Andi, 2013.

[8] S. Maharani, "Application for Diagnosing Fever Diseases in Toddlers Using Certainty Factor (CF) and Artificial Neural Networks (ANN)," Master's thesis, Information Systems Department, Diponegoro Univ., Semarang, 2012.A. Andriani, *Expert System Programming*. Yogyakarta: Mediakom, 2016.

[9] SMERU Research Institute, "Toward Pro-poor Policy through Research," SMERU Working Paper.

[10] Statistics Indonesia, "Human Development Index," BPS, 2023.

[11] Statistics Indonesia, "Poverty Line by Regency/City," BPS, 2025.