International Journal Software Engineering and Computer Science (IJSECS)
5 (1), 2025, 53-62
Published Online April 2025 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v5i1.3759.

RESEARCH ARTICLE Open Access

K-Means Clustering Analysis of Poverty Data in Cilacap District

Kiki Setiawan *
Information Systems Study Program, Faculty of Computer Science, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Corresponding Email: ki2djoaz@stikomcki.ac.id.

Kastum
Information Systems Study Program, Faculty of Computer Science, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Email: kastum@gmail.com.

Yuliya Putri Pratama
Information Systems Study Program, Faculty of Computer Science, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, East Jakarta City, Special Capital Region of Jakarta, Indonesia.
Email: yuliyaputri@gmail.com.

Received: December 31, 2024; Accepted: January 15, 2025; Published: April 1, 2025.

Abstract: Poverty stands as a complex structural obstacle within social development frameworks. The COVID-19 pandemic intensified poverty dynamics in Indonesia which saw poverty rates increase by 9.78% in March and reach 10.19% by September. Local Bureau of Statistics data shows that the poverty rate in Cilacap Regency dropped to 10.99% (around 191,000 people) in March 2024 from 10.68% (186,080 people) in March 2023. The study uses k-means clustering methodology for analysis and maps poverty-prone areas utilizing QGIS software. The analysis revealed 12 sub-districts and 14 neighborhood units (RW) alongside a single community unit (RT) that show unique poverty characteristics. The silhouette coefficient evaluation produced a 0.55 score which showed a moderate cluster structure and acceptable cluster placement. The research provides empirical evidence about poverty distribution which shows how data mining methods can enhance spatial socioeconomic studies. The study presents a detailed analysis of poverty stratification across Cilacap Regency through the application of sophisticated computational methods.

Keywords: Poverty; Clustering; K-Means.

1. Introduction

Regional development strategies throughout Indonesia face persistent challenges from the multifaceted socio-economic issue of poverty which particularly affects districts such as Cilacap. The complex structure of economic marginalization requires thorough analytical methods to explore its fundamental mechanisms and develop effective mitigation strategies [1]. The understanding among researchers and policymakers has evolved to identify poverty as a complex problem that involves multiple social, economic, and structural factors beyond simple statistical representation [2]. From March 2014 to March 2024 Cilacap Regency saw a complex pattern of poverty reduction which included significant variations during the COVID-19 pandemic. The district's count for poor individuals stood at 186,060 in March 2024 which was 4,920 fewer than the count recorded in March 2023. The decrease in poverty levels from 10.99% to 10.68% reveals subtle socio-economic changes that merit further analysis according to sources [3][4].

Multiple interconnected factors contribute to a complex structure defining poverty conditions in Cilacap. The labor market continues to be a significant factor in the struggle to find stable work because it is marked by ongoing difficulty in securing employment, extensive job insecurity among workers and widespread low-paying jobs in the informal economy. The economic limitations face additional burdens from wider socioeconomic restrictions which manifest as low income levels alongside unexpected job losses and insufficient social assistance systems. Educational limitations further exacerbate the poverty cycle. Insufficient education results in major obstacles to economic advancement because it prevents people from obtaining higher-paying positions and developing essential professional skills. The COVID-19 pandemic significantly increased these pre-existing vulnerabilities by uncovering systemic weaknesses in regional economic resilience and producing unprecedented challenges for vulnerable populations [5][8].

The Poverty Line functions as an essential measure for analyzing socio-economic divisions. This benchmark reached IDR 441,093 per capita monthly in March 2024 after a 5.17% growth from the previous year and a remarkable 14.58% elevation since March 2022. A detailed analysis of the Poverty Line's compositional structure reveals telling insights: The Food Poverty Line (GKM) makes up 73.43% of total spending while the Non-Food Poverty Line (GKNM) accounts for 26.57% of expenditures [6]. Urban-rural disparities represent an essential component in the understanding of poverty dynamics. The living cost structures and economic opportunities differences between geographical areas result in a higher poverty line in urban areas (IDR 885,655) than in rural areas (IDR 799,327). The differences across locations demonstrate the necessity for tailored poverty reduction strategies that consider specific regional characteristics.

The evaluation of inequality offers deeper insight that extends past basic poverty numbers. The World Bank's methodology revealed low inequality levels in Cilacap for March 2023 with bottom 40% expenditure recorded at 19.27%. However, subtle yet concerning trends emerged: The expenditure share of the bottom 40% slightly declined from 19.39% in 2022 to 19.27% in 2023 whereas the top 20% of the population saw their share rise from 44.08% to 45.03%. The gradual changes in data indicate potential enduring socio-economic stratification threats which require immediate action from government authorities. The study stands out by using advanced data mining methods, especially k-means clustering, to deliver detailed insights into how poverty spreads across Cilacap. The study utilizes computational methods to detect spatial poverty patterns while creating precise intervention strategies and offering backing for data-driven policy decisions. The examination includes complete datasets from 2019 to 2024 which provide a long-term view on poverty patterns in the region [1][5].

An effective analysis of poverty requires surpassing basic numerical representations to achieve deeper insights. Economic marginalization needs to be addressed through a comprehensive strategy that examines the complex connections between social conditions and economic structures. Policymakers who use advanced analytical frameworks can create detailed and situation-specific solutions to address the complex reality of socio-economic problems. This research holds value due to its potential to shape focused intervention methods rather than solely its ability to describe conditions. The study provides detailed poverty characteristic maps that enable better resource allocation and policy development as well as community-focused economic development strategies.

2. Related Work

Over recent decades poverty research has evolved from basic income-based models to sophisticated multidimensional analyses [14][15]. The development of research methods shows an expanded awareness of how intricate social and economic exclusion is which resists measurement by just one financial metric. Empirical investigations have established poverty as a complex condition that arises from structural, social and individual interactions. Recent progress in computational technology has produced innovative approaches to the analysis of poverty. Researchers can investigate underlying patterns within intricate socio-economic data

sets thanks to the capabilities of data mining and machine learning techniques [18][19]. Research now acknowledges the K-Means algorithm as a robust tool for discovering poverty clusters and their geographical distributions. Studies by Alsharkawi *et al*. (2021) in Jordan [14] and Ramadhan *et al*. (2023) The research conducted by Ramadhan *et al*. in Indonesia [18], illustrates the superior analytical capabilities of computational approaches compared to conventional methods.

The spatial dimension in poverty research is increasingly gaining significant attention. Geographic Information Systems (GIS) have become a key instrument in mapping the geographical distribution of poverty [21][22]. Nugraha *et al*. (2022) study in Central Java and Yogyakarta revealed uneven patterns of poverty distribution, providing empirical evidence of the geographic complexity of economic inequality [21]. Hidayati *et al*. (2022) used K-Means to classify villages based on poverty indicators in Surabaya, demonstrating the potential of clustering methodology in understanding local economic diversity [22]. Demographic factors play a fundamental role in poverty analysis. Contemporary research explores how variables such as education, employment opportunities, and social characteristics interact to shape economic vulnerability [17][24]. Erda *et al*. (2023) used K-Means to classify regions based on poverty levels in Indonesia, highlighting the need for approaches tailored to specific demographic contexts [24]. The COVID-19 pandemic has been a turning point in poverty research, exposing and exacerbating existing socio-economic fault lines [23]. Fitriana and Mabruri used the Self-Organizing Map algorithm to analyze the impact of the pandemic on poverty, demonstrating the need for adaptive and responsive research methodologies [23].

Methodological innovation continues to evolve, with researchers increasingly adopting interdisciplinary approaches. Collaboration between economists, sociologists, data scientists, and policy makers has become a key feature of contemporary research [20][25]. Riyono and Pujiastuti (2022), emphasize the importance of the Elbow method in determining the optimal number of clusters, demonstrating methodological sophistication in poverty analysis [25]. Participatory approaches that place the experiences of marginalized communities at the center of research have replaced traditional extractive models. This reflects a shift towards more empowering and inclusive methodologies [14][18].

Going forward, poverty research will increasingly rely on the integration of multimethodological approaches, the use of advanced computational techniques, and holistic interdisciplinary perspectives. Artificial intelligence and machine learning offer unprecedented capabilities in analyzing complex socio-economic data, enabling predictive and proactive approaches to understanding economic vulnerability [16][17]. Ardini and Sirait (2023) and others continue to develop and compare clustering algorithms, indicating that poverty research methodologies are in a phase of continuous innovation [26]. The increasing complexity of the global economic system requires increasingly sophisticated, adaptive, and responsive analytical tools. Poverty research has evolved from mere statistical measurement to a complex discipline that combines technology, ethics, and deep social understanding. This multidimensional approach not only provides richer insights into the dynamics of poverty, but also provides a foundation for more effective and equitable interventions [14][18][24].

3. Research Method

The study utilizes the KNN Algorithm together with the CRISP-DM method for data mining. The k-nearest neighbors (KNN) algorithm functions as a non-parametric supervised learning classifier that categorizes individual data points based on proximity measurements. The CRISP-DM method consists of six distinct phases which include business understanding followed by data understanding data processing modeling evaluation and deployment. The k-means algorithm performs clustering during the Modeling phase of the process. This study provides a detailed description of each step within the CRISP-DM framework:

1) Business Understanding
   In the first phase, a problem analysis or understanding of what problems can be raised in this research is carried out. This phase has three processes, including:
   a. Determine business objectives, understand the goals to be achieved by conducting an interview with the Central Statistics Agency.
   b. Assess the situation, this process is carried out by analyzing the facts that occur in the field, by knowing what analysis has been carried out in the Cilacap Regency Government and whether poverty mapping has been carried out in Cilacap Regency.
   c. Determine data mining goals, this process is carried out by determining the process technically, namely determining the data mining method that will be used to achieve the research objectives to be achieved.

2) Data Understanding
In the second phase, namely data preparation, data collection is carried out, describing or depicting data, then exploring which data might be useful for the Cilacap Regency Government, then identifying problems that will be carried out related to the data owned and studying the data obtained that will be used in the study. Data was obtained from BPS regarding poverty data based on reports in the Cilacap Regency area for five years, namely 2019 to 2023.

3) Data Preparation
At this stage, preprocessing is carried out where this phase process has several stages:
   a. Data selection, this process is the process of selecting data and also selecting attributes that will be used according to data mining objectives.
   b. Data preprocessing, this process is done by cleaning data or cleansing data by dealing with outliers, noisy data and missing values. At this stage also ensures the data has good quality.
   c. Transformation, this process is the process of grouping attributes into new data, then integrating the data, transmitting data that is in accordance with the objectives to then be processed in data mining.

4) Modeling
The modeling phase includes selecting the model followed by implementing data mining techniques with appropriate algorithms and tools through a Jupyter Notebook utilizing Python. The study uses the clustering method with the k-means algorithm to perform data mining on poverty data specifically from Central Java Province. The model results will be displayed using QGIS. The QGIS method employs Quantum Geographic Information System (QGIS) software to handle and visualize geospatial data while also providing analysis capabilities. Geographic Information System (GIS) software QGIS is available without cost and operates under open-source licensing. The open-source GIS software QGIS serves multiple functions including mapping tasks and spatial analysis as well as managing geospatial data.

5) Evaluation
In this evaluation phase, an analysis is carried out on the results of the data learning process. This phase is the process of interpreting the results of the data mining modeling used. The evaluation is carried out using the Silhouette Coefficient method, which is a method that tests the quality of the resulting cluster. This evaluation is carried out to determine whether applied modeling is appropriate and suitable to be applied to this research case and is in accordance with the objectives to be achieved. Then, from the results of the evaluation, determine the next steps whether they can be continued to the next stage or repeated from the beginning because they are not in accordance with the objectives.

6) Deployment
In the sixth phase, namely the dissemination phase by making reports or presentations of the knowledge obtained from the results of modeling and evaluation in the data mining process. The results obtained are given to the Central Java Provincial Government which are used to determine the right decisions in overcoming and preventing poverty against the results of mapping poverty-prone areas in Central Java Province.

4. Result and Discussion

4.1 Results

Data Mining Research Results The results of the research conducted were to analyze the mapping of poverty-prone areas in Central Java Province using the k-means clustering algorithm. The results of the clustering were evaluated and visualized using QGIS software.

1) Business Understanding
The business understanding stage focuses on understanding the objectives of the needs based on business assessment. Then the plan is transformed into an initial data mining plan designed to achieve the objectives. There are 3 stages of business understanding, namely:
   a. Determining Business Objectives, if in this initial phase, observations were made by searching for data on the West Java opendata website to find out the potential that can be done in research according to needs so that it can be useful and can be used. The focus of this observation is on poverty data that occurs in Cilacap Regency, Central Java.
   b. Assessing the Situation, the following conclusions were drawn:

1. The West Java Provincial Government has data that can be used as information or knowledge. The data is data on the distribution of the number of poor people based on districts/cities, where both data are available annually. The data obtained is data from 2019 to 2023.
2. The annual data collection has not been processed to produce better and more useful information.
3. The form of information delivery regarding poverty in Cilacap Regency, Central Java, which is on the Central Java Statistics Agency website is only a list of the total number of poor people, there is no visual form of information delivery to the public to find out the distribution of existing poverty data.

c. Determining the Data Mining Goals In overcoming the problem of observation results in the facts that occur in the Cilacap Regency Government, Central Java, the purpose of data mining that can be carried out in this study is to cluster poverty-prone areas using the clustering method by applying the k-means algorithm which then makes it easier for the public to find out the intensity of existing crimes, it will be visualized in the form of a map with the help of one of the geographic information system software.

2) Data Understanding
The initial data understanding phase, getting to know and understand the data you have and analyzing what can be done on the data. The initial stage is the process of collecting data obtained from the Central Java open data portal website, namely poverty data from 2019 to 2023. The data obtained from the Cilacap Regency open data portal is in the form of poverty data in excel. Where there are sub-districts in 6 years. The initial poverty dataset has attributes of sub-district name, village name, number of poor people, units and years.

3) Data Preparation
Data preparation includes all activities in building a dataset that will be included in the modeling from the initial raw data. K-means is the algorithm used in processing dataset modeling. There are several stages in data preparation, namely data selection, data preprocessing and data transformation.

a. Data Selection
The data taken is about poverty data, namely data on the number of poor people based on Cilacap Regency in Central Java Province. This dataset is stored in excel or csv format. Then the data is cleaned by deleting attributes that are not needed in this study. The selected data is data related to poverty in 2019 to 2023. The attributes used are the name of the district/city, the number of poor people and the year.

b. Data Preprocessing
In this process, a dataset is created from raw data to data that is ready to be used in the data mining modeling process. What is done is to delete unused attributes such as sub-district names and units. Then make changes to the number of poor people who initially had units of thousands of people to units of people. The process can be seen in the following image.

c. Data Transformation
The data set development stage is the process of transforming data according to the needs of the modeling process. The development of this new dataset is done by transforming data where the values are calculated by summarizing information. In the modeling process, the number of poor population data is needed and grouped by district/city. Data changes are made by changing the data that was previously in one column showing the number of poor people each year, at this stage changing the number of poor people into several columns that are grouped by year. Then add the number attribute that shows the number of poor people during 2019 to 2023. This data transformation produces 27 data containing the names of sub-districts in Cilacap which are in the sub-district column. In each RT there is data on the number of poor people according to the year from 2019 to 2023 which is located in the columns 2019, 2020, 2021, 2022, 2023. In the last column, the total is the overall value of the number of poor people in Cilacap Regency, Central Java.

4) Modeling
Modeling is a phase that directly involves data mining techniques, selection of data mining techniques, algorithms and determining parameters with optimal values. Clustering is a data mining technique used in this study. The algorithm chosen is the k-means algorithm which will later be modeled on the transformed data. The results obtained from the clustering technique and the k-means algorithm are data groups used to group poverty-prone areas that are not vulnerable, vulnerable, to very vulnerable in Central Java Province, especially in Cilacap Regency. The clusters that will be created are 3 clusters, the determination of the clusters will be divided into 3, namely the non-vulnerable cluster (C1), the vulnerable

cluster (C2), and the very vulnerable cluster (C3). Determination of the midpoint or initial centroid is carried out from the data of each attribute by taking the smallest value or minimum value for the non-vulnerable cluster (C1), then the average value for the vulnerable cluster (C2), and the largest value in the data or maximum for the very vulnerable cluster (C3). The attributes that are used as references in clustering are the entire district/city. The following is the k-means clustering process using python.

```
# mulai proses kmeans nya
X = np.array(nilai_centroid_awal, np.float64)
km = KMeans(n_clusters=3, init=X, n_init=1).fit(clear_dataset)
centers = km.cluster_centers_
print(centers)
```

```
[[ 68075.          65575.          64825.          57108.33333333
   53900.          62203.33333333]
 [223685.71428571 210492.78571429 207392.78571429 179671.42857143
  168383.57142857 193437.14285714]
 [487100.         490800.         487300.         415000.
  395030.         465670.        ]]
```

Figure 1. K-Means Data Processing

K-Means Process Image The data above shows the centroid value at the last iteration in the k-means clustering process using the library in python. For the process of dividing or grouping into 3 clusters using python.

```
# masukan nilai cluster ke dalam dataframe
df['cluster'] = km.labels_
a = km.labels_
mapping = {0:'Tidak Rawan', 1:'Rawan', 2:'Sangat Rawan'}
a = [mapping[i] for i in a]
df['keterangan'] = a
```

```
new_df = df[['kabupaten', 'cluster', 'keterangan']]
new_df
```

Figure 2. Clustering Process

The results of poverty data clustering using k-means were obtained for the cluster of areas with a non-prone poverty rate or C1, namely South Cilacap Regency, then all areas recorded in South Cilacap Regency. Then for the cluster of areas with a vulnerable poverty rate or C2, namely Central Regency. The third cluster is an area with a very vulnerable poverty rate or C3, namely North Cilacap Regency. Evaluation At this evaluation stage, an interpretation phase is carried out on the data mining results which are carried out in depth with the aim that the results at the modeling stage are in accordance with the targets to be achieved in the previous business understanding stage and to find out the extent of the quality of the applied modeling. Evaluation Result The next modeling result is evaluated using the silhouette coefficient. This test is carried out to determine the closeness of the relationship between objects and how far apart the clusters are, so that the quality of a cluster is known. If the value is close to 1 or positive and ai is close to 0, this results in a maximum value of 1 with ai = 0, then the structure of the cluster falls into the good category, or the cluster is in the right cluster. Then if the silhouette coefficient value = 0 then the cluster structure falls into the unclear category. While if the silhouette coefficient value = -1 then the structure of the cluster falls into the overlapping category. The following tests are carried out using the library in python.

```
from sklearn.metrics import silhouette_score
```

```
silhouette_score(clear_dataset, km.labels_)
```

```
0.5539929079261192
```

Figure 3. Evaluation process

The test results using the silhouette coefficient or silhouette score shown in Figure 3. produce an index value of 0.55. The quality of the clusters produced by this k-means algorithm is included in the medium

structure category with a reasonable cluster placement interpretation. Determine Next Step At this stage, it is the stage of determining the next steps to be taken based on the test results. There are two options in determining the next steps, namely returning to the business understanding stage or continuing to the final stage in the crisp-dm methodology, namely deployment. This stage can be continued if the test results are in accordance with the objectives, if not, then return to the initial stage (business understanding). In this study, the test results showed that the Cluster is included in the medium structure category. This category can be said to be a good cluster, and the research results are in accordance with the initial objectives, so it can be continued to the next step. Deployment After the modeling process by applying data mining clustering techniques, the next stage is the preparation of reports and the process of reporting results to Central Java. Provincial Government which is expected to allow the Government to determine the right decisions in overcoming and preventing poverty. The final report is made after all data mining processes have been completed. The report that will be given to the Central Java Provincial Government is made in the form of mapping visualization using QGIS. The results of the clustering in the modeling process are visualized in the form of mapping using QGIS so that they are easy to understand in mapping poverty-prone areas in Central Java Province.

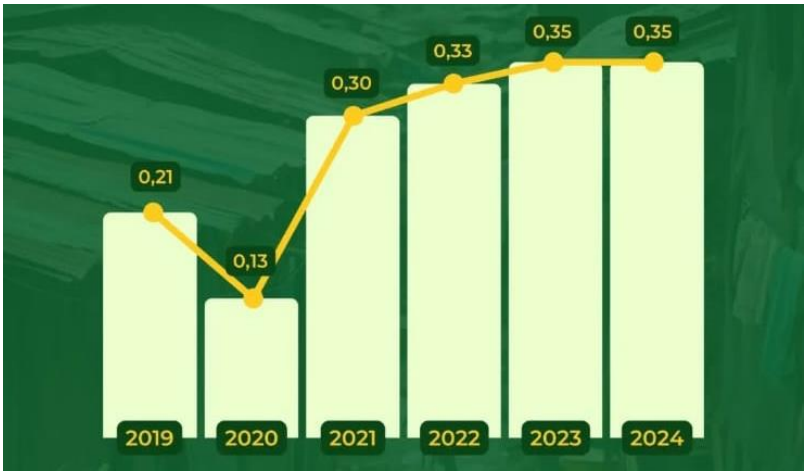

Figure 4. 6 Is the Visualization Result of Mapping Poverty-Prone Areas from 2019 to 2023

4.2 Discussion

This study presents an innovative approach in analyzing poverty distribution using the K-Means clustering algorithm, which allows a comprehensive understanding of the complexity of the poverty phenomenon in Cilacap Regency [14][17][18]. The applied data mining methodology does not merely produce statistical figures, but also reveals hidden spatial and temporal patterns in poverty data during the 2019-2023 period [24][25]. The clustering analysis produces three main clusters that reflect the gradation of economic vulnerability in the Cilacap region. The first cluster, including South Cilacap, is identified as a non-poverty-prone area with relatively stable socio-economic characteristics [28]. This is consistent with the findings of Yuningsih and Aryani (2024) who emphasize the importance of spatial identification in poverty analysis [28]. The economic stability of this region can be associated with infrastructure accessibility, economic opportunities, and the effectiveness of empowerment programs that have been implemented [20].

The Central Cilacap area is placed in the vulnerable category, reflecting the complexity of socio-economic dynamics that require strategic interventions [29][30]. The clustering approach allows the identification of transition zones with higher potential poverty risks. Rahman *et al.*'s (2021) study emphasized the importance of multidimensional analysis in understanding economic vulnerability, which is in line with the findings in this study [30]. A comprehensive approach is needed that includes capacity building, skills training, and structured economic support [23]. The critical cluster, North Cilacap, is categorized as an area with very high poverty vulnerability. In-depth analysis revealed multidimensional challenges, including limited access to education, minimal economic opportunities, and the complexity of structural problems [17][18]. Alfiah *et al.* (2022) emphasized the importance of an adaptive approach in poverty mitigation, which is in line with the need for intervention in this area [29].

The evaluation method using the silhouette coefficient produces an index of 0.55, which indicates the quality of clustering in the moderate but significant category [24][25]. This value provides methodological confidence that the approach used is able to produce valid and scientifically significant groupings. This is consistent with the study of Erda *et al.* (2023) which emphasizes the importance of the silhouette coefficient in clustering analysis [24]. The main contribution of the study lies in the transformation of raw data into strategic information for policy making [14][18]. The resulting mapping visualization serves as a diagnostic tool that allows for spatial understanding of the distribution of poverty. This approach is in line with the

recommendations of Ramadhan *et al*. (2023) on the use of big data in poverty analysis. However, the study has methodological limitations [18]. The focus on a limited period and dependence on the quality of source data are major challenges [15][17]. Therefore, continuous validation and model development are important recommendations for further research. This study confirms the significant potential of the data mining approach in understanding poverty dynamics [14][27]. By integrating computational technology and socio-economic analysis, this study provides a strong scientific basis for strategic decision making in poverty alleviation efforts in Cilacap Regency.

5. Conclusion

This study implements a clustering method using the K-Means algorithm to analyze poverty data in Cilacap Regency during the period 2019-2023, producing comprehensive and in-depth findings on the distribution of regional economic vulnerability. Through a sophisticated data analysis approach, the study successfully identified three clusters that illustrate the structural complexity of poverty with a significant level of precision. The clustering results reveal a complex and diverse pattern of poverty distribution. The first cluster, consisting of 12 Rukun Warga (RW), is characterized as a relatively economically stable area, showing adequate social and infrastructure resilience. The second cluster, covering 14 RW, is categorized as an area vulnerable to poverty, indicating a transition zone that requires strategic and sustainable intervention. The third cluster, covering one district/city, is identified as an area with extreme poverty vulnerability, presenting multidimensional challenges that require comprehensive attention. Methodological evaluation using the silhouette coefficient produces an index of 0.55, which is in the medium structure category with a scientifically acceptable interpretation of cluster placement. This value not only provides mathematical validation of the clustering method used but also shows the model's ability to extract hidden patterns in the poverty dataset. This proves the effectiveness of the K-Means approach in analyzing complex socio-economic phenomena. The visualization of clustering results in the form of a poverty level map with color differentiation for each region is an innovative contribution to this study. Maps serve as diagnostic tools which transform raw data into strategic insights to support evidence-based decisions beyond their traditional role as graphical data representations. The map uses different colors to display poverty vulnerability levels which helps people comprehend poverty distribution complexity in Cilacap Regency.

The research expands knowledge about poverty dynamics by merging advanced computational techniques with thorough socio-economic examination. The approach taken expands past standard techniques to provide an all-encompassing view that encompasses multiple dimensions. The study recognizes research limitations due to the time span and source data quality which future research should investigate. The study's findings deliver precise poverty distribution details for Cilacap Regency while establishing a replicable analytical framework for broader applications. The primary contribution of this research is its integration of computational technology with statistical analysis and socio-economic context understanding to help create effective poverty alleviation interventions.

References

[1] Solichin, A., & Khairunnisa, K. (2020). Klasterisasi persebaran virus Corona (Covid-19) di DKI Jakarta menggunakan metode K-Means. *Fountain of Informatics Journal, 5*(2), 52-59.

[2] Alfina, T., Santosa, B., & Barakbah, A. R. (2012). Analisa perbandingan metode hierarchical clustering, k-means dan gabungan keduanya dalam cluster data (studi kasus: Problem kerja praktek teknik industri its). *Jurnal Teknik ITS, 1*(1), A521-A525.

[3] Aria, P. (2020, April 7). Globalization and the World Supply Chain Locked by the Covid-19 Pandemic.

[4] Anggraini, A. A., Muharom, L. A., & Si, M. (2017). Pengelompokan Kecamatan Menggunakan Metode K-Means Cluster. Universitas Muhammadiyah Jember. http://repository.unmuhjember.ac.id/591

[5] Nosra, A., Arifianto, D., & Rahman, M. (2022). Penerapan Metode Cosine Similarity Untuk Meningkatkan Kinerja K-Means Pada Pengelompokan Wilayah Penanganan Covid Di DKI Jakarta. *Jurnal Smart Teknologi, 3*(5), 560-566.

[6] Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, (2)*, 224-227. https://doi.org/10.1109/TPAMI.1979.4766909

[7] Nurjannah, M., Hamdani, H., & Astuti, I. F. (2016). Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) untuk Text Mining. *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer, 8*(3), 110-113. http://dx.doi.org/10.30872/jim.v8i3.113

[8] Noviyanto, N. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia. *Jurnal Khatulistiwa Informatika, 22*(2), 183-188. https://doi.org/10.31294/p.v21i2

[9] Noviyanto, N., & Ekasari, P. (2022). Algoritma K-Means Untuk Klasterisasi Jabatan Fungsional Dosen Pada Perguruan Tinggi Swasta Di Lingkungan LLDikti Wilayah III. *Paradigma - Jurnal Komputer Dan Informatika, 24*(1), 103-107. https://doi.org/10.31294/paradigma.v24i1.1112

[10] Fitri, R., & Asyikin, A. N. (2015). Aplikasi penilaian ujian essay otomatis menggunakan metode cosine similarity. *Poros Teknik, 7*(2), 88-94. https://doi.org/10.31961/porosteknik.v7i2.218

[11] Siahaan, D., & Christina, S. (2015). Using Semantic Similarity for Identifying Relevant Page Numbers for Indexed Term of Textual Book. In R. Intan, C. H. Chi, H. Palit, & L. Santoso (Eds.), *Intelligence in the Era of Big Data* (pp. XX-XX). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-46742-8_17

[12] Triharseno, W., Dhuhita, W. M. P., & Priadana, A. (2020). Sistem Pendukung Keputusan Penerimaan Programer Software House Menggunakan Metode Simple Additive Weighting (SAW). *JURNAL PILAR TEKNOLOGI Jurnal Ilmiah Ilmu Ilmu Teknik, 5*(1). https://doi.org/10.33319/piltek.v5i1.52

[13] Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., ... & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems, 14*, 1-37. https://doi.org/10.1007/s10115-007-0114-2

[14] Alsharkawi, A., Al-Fetyani, M., Dawas, M., Saadeh, H., & Yaman, M. (2021). Poverty classification using machine learning: The case of Jordan. *Sustainability, 13*(3), 1412. https://doi.org/10.3390/su13031412

[15] Li, G., Cai, Z., Qian, Y., & Chen, F. (2021). Identifying urban poverty using high-resolution satellite imagery and machine learning approaches: Implications for housing inequality. *Land, 10*(6), 648. https://doi.org/10.3390/land10060648

[16] Lopes, M., Silva, L., Silva, S., Lima, W., Barbosa, R., & Fernandes, M. (2023). Self-organising maps applied to the stratification of preterm birth risk in Brazil with socioeconomic data. https://doi.org/10.21203/rs.3.rs-3676862/v1

[17] Gao, C., Fei, C., McCarl, B., & Leatham, D. (2020). Identifying vulnerable households using machine learning. *Sustainability, 12*(15), 6002. https://doi.org/10.3390/su12156002

[18] Ramadhan, R., Wijayanto, A., & Pramana, S. (2023). Geospatial big data approaches to estimate granular level poverty distribution in East Java, Indonesia using machine learning and deep learning regressions. *Proceedings of the International Conference on Data Science and Official Statistics, 2023*(1), 186-200. https://doi.org/10.34123/icdsos.v2023i1.359

[19] Nofriani, N. (2020). Machine learning application for classification prediction of household's welfare status. *Jitce (Journal of Information Technology and Computer Engineering), 4*(2), 72-82. https://doi.org/10.25077/jitce.4.02.72-82.2020

[20] Novitasari, M., & Iskandar, D. (2022). Spatial spillover impact of sectoral government expenditure on poverty alleviation in South Kalimantan Province. *The Journal of Indonesia Sustainable Development Planning, 3*(3), 207-221. https://doi.org/10.46456/jisdep.v3i3.361

[21] Nugraha, A., Prayitno, G., Nandhiko, L., & Nasution, A. (2022). Socioeconomic conditions on poverty levels a case study: Central Java Province and Yogyakarta in 2016. *Revista De Economia E Sociologia Rural, 60*(1). https://doi.org/10.1590/1806-9479.2021.233206

[22] Hidayati, S., Darmaliana, A., & Riski, R. (2022). Comparison of k-means, fuzzy c-means, fuzzy gustafson kessel, and DBSCAN for village grouping in Surabaya based on poverty indicators. *Jurnal Pendidikan Matematika (Kudus), 5*(2), 185. https://doi.org/10.21043/jpmk.v5i2.16552

[23] Fitriana, I., & Mabruri, M. (2022). Cluster analysis of COVID-19 impact on poverty in Indonesia using self-organizing map algorithm. *Jurnal Aplikasi Statistika & Komputasi Statistik, 14*(1), 85-94. https://doi.org/10.34123/jurnalasks.v14i1.389

[24] Erda, G., Gunawan, C., & Erda, Z. (2023). Grouping of poverty in Indonesia using K-means with silhouette coefficient. *Parameter Journal of Statistics, 3*(1), 1-6. https://doi.org/10.22487/27765660.2023.v3.i1.16435

[25] Riyono, J., & Pujiastuti, C. (2022). Simulation of the K-means clustering algorithm with the elbow method in making clusters of provincial poverty levels in Indonesia. *Jurnal Matematika Mantik, 8*(2), 113-123. https://doi.org/10.15642/mantik.2022.8.2.113-123

[26] Ardini, A., & Sirait, H. (2023). Comparison of K-medoids and CLARA algorithm in poverty clustering analysis in Indonesia. *Operations Research International Conference Series, 4*(4), 141-148. https://doi.org/10.47194/orics.v4i4.279

[27] Poerwanto, B. (2023). K-means – resilient backpropagation neural network in predicting poverty levels. *Jurnal Varian, 6*(2), 157-166. https://doi.org/10.30812/varian.v6i2.2756

[28] Yuningsih, I., & Aryani, M. (2024). Comparison of methods for applying the data mining clustering concept to Banten provincial government poverty data: Systematic review. *Jurnal Ilmu Sosial Mamangan, 13*(1), 121-134. https://doi.org/10.22202/mamangan.v13i1.8074

[29] Fitriana, I., & Mabruri, M. (2022). Cluster analysis of COVID-19 impact on poverty in Indonesia using self-organizing map algorithm. *Jurnal Aplikasi Statistika & Komputasi Statistik, 14*(1), 85-94. https://doi.org/10.34123/jurnalasks.v14i1.389

[30] Rahman, M., Sani, N., Hamdan, R., Othman, Z., & Bakar, A. (2021). A clustering approach to identify multidimensional poverty indicators for the bottom 40 percent group. *Plos One, 16*(8), e0255312. https://doi.org/10.1371/journal.pone.0255312

[31] Alfiah, F., Almadayani, A., Farizi, D., & Widodo, E. (2021). Analisis clustering K-medoids berdasarkan indikator kemiskinan di Jawa Timur tahun 2020. *Jurnal Ilmiah Sains, 22*(1), 1. https://doi.org/10.35799/jis.v22i1.35911.