Research Article ISSN: 2833-0382

International Journal of Family Medicine & Healthcare

Machine Learning in Public Health: Cluster Analysis of Dengue Fever Periode 2012-2020 in Indonesia

Rachmah Indawati[1]* and Mochammad Bagus Qomaruddin[2]

Department Epidemiology, Biostatistics Population study and Health Promotion, Public Health School, Universitas Airlangga, East Java, Indonesia.

[*]Correspondence:

Rachmah Indawati, Department Epidemiology, Biostatistics Population study and Health Promotion, Public Health School, Universitas Airlangga, East Java, Indonesia.

Received: 24 Mar 2025; **Accepted:** 01 May 2025; **Published:** 09 May 2025

Citation: Indawati R, Qomaruddin MB. Machine Learning in Public Health: Cluster Analysis of Dengue Fever Periode 2012-2020 in Indonesia. Int J Family Med Healthcare. 2025; 4(1): 1-13.

ABSTRACT

Machine Learning offers a powerful tool for tracking diseases efficiently, quickly, and flexibly. This study aims to analyze sequential datasets 2012-2020 on Dengue fever, providing new insights into the latest trends and developments related to the disease. First, a review of health profile reports from East Java. Second, an examination of disease data, including the number of cases, deaths, and the Case Fatality Rate, CFR. Third, clustering analysis using the k-means method to identify regions with similar or distinct characteristics based on the number of cases, deaths, and CFR.

The results reveal that regions with low case counts are those with numbers below the East Java average. These low-case regions are more prevalent compared to those with moderate or high case counts. Similarly, clustering based on the number of deaths shows that regions with low death counts are more common. However, between 2012-2020, certain areas in East Java experienced high death counts and CFR clusters. Notably, some regencies/cities within high CFR clusters have seen repeated occurrences. A region's CFR cluster can change over time, shifting from low to moderate or high, and vice versa.

The findings highlight that the high number of regions with elevated death clusters underscores the serious challenges East Java faces in combating Dengue Fever. The recurrence of high CFR clusters in certain areas indicates the need for sustained vigilance against the severity of the disease. Additionally, the dynamic changes in cluster classifications across regions suggest that Dengue fever is highly contagious. The widespread distribution of the disease across nearly all regencies and cities in East Java implies potential spatial or regional factors influencing its spread.

Keywords

Clustering method, Dengue Fever, East Java, Machine learning.

Introduction

The rapid advancement of technology and information has created new possibilities for recording and storing data in databases. This phenomenon is known as "big data," which often refers to the massive volume of data generated. When properly analyzed, this vast amount of data can uncover new insights and support decision-making processes [1]. Recorded data is not merely stored but is actively analyzed and utilized to better understand public health challenges.

Public health issues remain a relevant and compelling topic. Today, with the support of Machine Learning, ML, it is possible to analyze data and generate actionable insights that can inform future policies [2]. Public health practitioners can benefit significantly from ML in analyzing complex datasets.

Several breakthroughs have been made by ML in public health research, particularly in exploring unlabeled data and enhancing large-scale case surveillance. Machine Learning is a technique that employs mathematical models to analyze patterns in data. ML is categorized into three main types: supervised learning, unsupervised learning, and semi-supervised learning. Supervised learning identifies patterns that link variables to outcomes, aiming to maximize prediction accuracy. Unsupervised learning explores the inherent properties of input data to detect trends and patterns. Semi-supervised learning combines these approaches, making predictions while primarily exploring input datasets. This method is often used in the data mining, DM phase [3].

ML is used to process data to uncover hidden information in Dengue Fever cases. Efforts to eradicate and control this disease are ongoing, yet cases persist. The number of cases fluctuates and sometimes increases significantly, leading to an Extraordinary Event [4-6].

Previous research has shown that the spread of the disease is influenced by changes in socio-demographic and economic conditions. Additionally, the progression of the disease can expand the distribution of mosquito serotypes, resulting in multiple Dengue virus serotypes being found in many countries. Factors such as population mobility between regions and countries, urban density, inadequate water supply, and waste management also contribute to the transmission of the Dengue virus. Other contributing factors include global trade, population growth, and urbanization [7,8]. Recent studies reinforce earlier findings that case numbers are linked to population growth [9] and that geographic areas can also influence disease incidence [10].

According to Indonesia's health profile, Dengue Fever is a recurring annual disease. East Java Province consistently reports higher case numbers compared to other regions. Data from 2006 to 2020 shows that cases are widespread across all regencies and cities [11-21].

In response to outbreaks in several areas of East Java, statistical and ML techniques can aid in studying Dengue Fever cases. ML techniques can analyze the progression of Dengue Fever and are expected to positively impact public health. ML can cluster cases and predict case numbers. Several studies have utilized data from database reports, considering various public health challenges. Cluster analysis is used to group data objects with similar characteristics into the same cluster. Two key outcomes of clustering algorithms are clustering performance and meaningful cluster descriptions [22].

Cluster analysis offers a way to organize and present complex datasets [23]. It is a primary technique for solving unsupervised learning problems, enabling the discovery of structures in unlabeled data [24].

This study aims to represent disease patterns by developing clustering algorithms based on the number of cases, deaths, and Case Fatality Rate, CFR, and to describe regions as low, medium, or high clusters.

Methods

The Machine Learning

The use of data in public health can help identify outcomes related to Dengue Fever, such as the number of cases, deaths, and CFR. However, these outcomes are often still in the form of raw datasets. The opportunity to utilize this data comes from the health profile reports of East Java. Epidemiological data on Dengue Fever cases reported annually from 2012 to 2020 for each regency and city in East Java were used [4-6,25-30].

The dataset has been recorded and reported in a database. To extract specific and nontrivial knowledge, a process known as Knowledge Discovery in Databases, KDD, is required. The first stage of using ML techniques to extract knowledge from the database is to identify disease patterns based on the stored data [31]. Since ML techniques can efficiently analyze datasets related to public health issues, it is crucial to apply these approaches to uncover new insights in public health.

Clustering Algorithm

ML techniques are used to identify patterns in data and build models that describe future outcomes of Dengue Fever. The k-means algorithm groups unlabeled data points into subgroups based on similarity. This algorithm is part of clustering models [32]. The k-means clustering analysis aims to categorize regions into low, medium, and high clusters based on the number of cases, deaths, and CFR.

K-means is an unsupervised learning algorithm. It is a straightforward method for solving clustering problems. The k-means procedure determines the centroid of a cluster by calculating the mean of the points within the cluster. First, the algorithm randomly selects k (central clusters) from various objects in the dataset, each representing a cluster center. Second, each remaining object is assigned to the cluster it is most similar to, based on the distance between the object and the cluster center [24]. Third, the k-means clustering algorithm iterates to refine the separation between clusters. For each cluster, the algorithm calculates a new mean using the objects assigned to the cluster in the previous iteration. All objects are then reassigned using the updated means as new cluster centers. The iteration continues until stable clustering is achieved [33].

The steps in this research include data collection, data preprocessing, and data mining [33]. In the data collection phase, data sources were obtained from the East Java health profile, accessed via the East Java Provincial Health Office website (https://dinkes.jatimprov.go.id/). The data preprocessing phase (data cleaning, data integration, data selection, and data transformation). Data cleaning (filling in missing values, correcting errors, identifying or removing outliers, and resolving inconsistencies). Missing measurement data is represented as empty values, which are imputed with zeros. Common methods for handling missing data in cross-sectional studies include removing missing values, imputing mean or median values, using the most frequent value or zero, or applying specific constants, as well as Multivariate Imputation by Chained Equations, MICE [34].

Data integration combines data from the East Java health profiles from 2012 to 2020. Data selection involves choosing only the data relevant to the analysis, such as the number of cases, deaths, and CFR. Data transformation is performed to support the analysis process using ML.

The data mining phase involves implementing the k-means clustering model, evaluating disease outcomes, and knowledge presentation (presenting the results of the data mining process).

Results

From 2012 to 2020 in East Java, the number of Dengue Fever cases fluctuated significantly. East Java consistently reported higher case numbers compared to the national average, indicating a high prevalence of Dengue Fever in the region. The annual mean and standard deviation of cases were as follows: (222.95 ± 204.74); (405.82 ± 406.73); (250.84 ± 208.29); (563.03 ± 377.38); (684.19 ± 411.08); (212.46 ± 131.29); (254.46 ± 199.35); (496.57 ± 416.53); and (231.24 ± 268.29). The wide standard deviations suggest substantial variability in case numbers across different regions, highlighting the uneven distribution of the disease. This variability underscores that Dengue Fever is influenced by social, cultural, and environmental factors. As such, comprehensive public health measures are essential.

To identify patterns in the disease, it is necessary to categorize the data into three clusters. This clustering aims to distinguish regions with low, medium, and high case counts, thereby determining the severity of infectious diseases and fatalities across different regions annually. Additionally, the widespread distribution of cases across a large area confirms that East Java is an endemic region.

Administratively, East Java comprises 37 regencies and cities, including 28 regencies and 9 cities. Each year, clustering was performed to categorize the regions. The results divided the case counts into low, medium, and high clusters. Regions in the low cluster had case numbers below the East Java average, while those in the medium and high clusters had case numbers above or significantly higher than the East Java average, respectively.

From 2012 to 2020, regions identified with high case clusters included Jember, Pacitan, Sumenep, Bondowoso, Trenggalek, Malang, Sidoarjo, Mojokerto, Kediri, Sampang, Jombang, Bojonegoro, Ngawi, Ponorogo, Surabaya City, and Batu City.

Regions consistently classified in the low case cluster included 6 areas (2 regencies and 4 cities): Mojokerto City, Madiun City, Pasuruan City, Kediri City, Madiun Regency, and Tuban Regency. The number of regions in the low case cluster was significantly higher than those in the medium and high clusters. The remaining regencies and cities fell into the medium case cluster.

From 2012 to 2020 in East Java, there were fluctuations in the number of deaths caused by Dengue Fever. Each year, an average of 3 to 24 individuals died from the disease in a given region, with the number of deaths often exceeding the East Java average. In

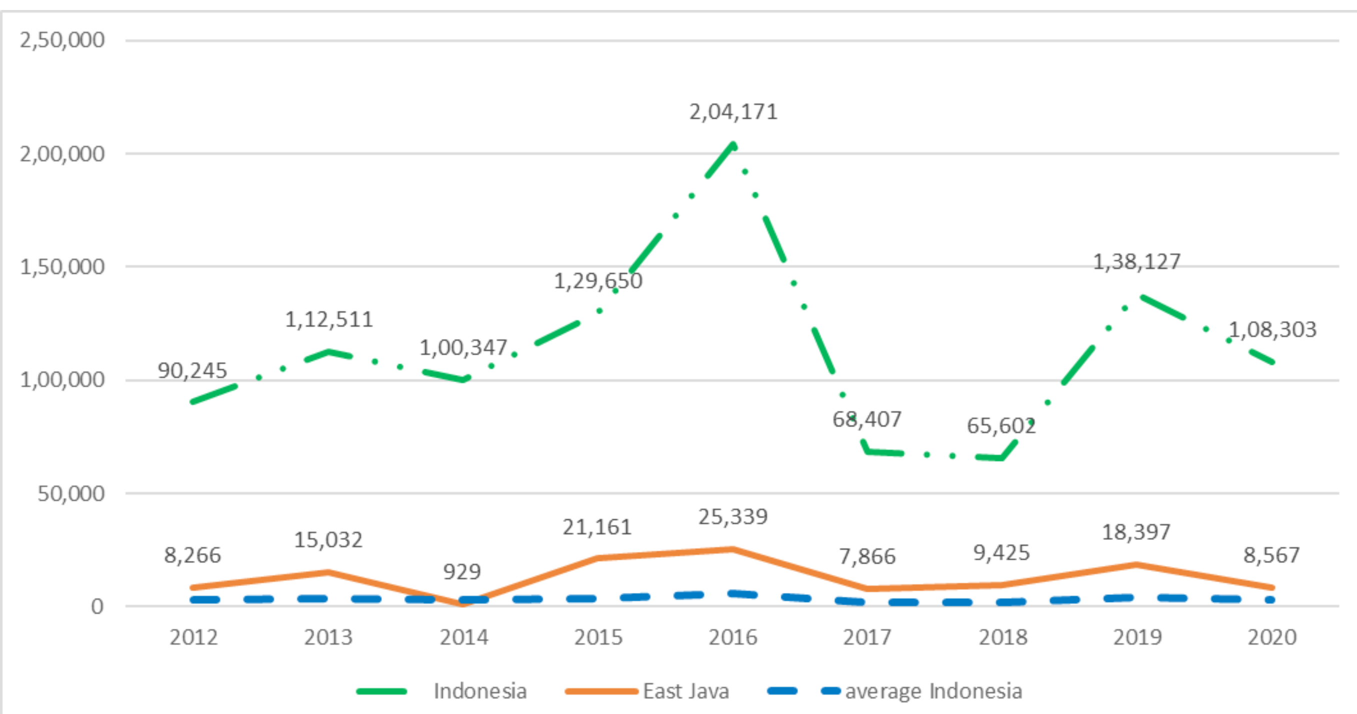


Figure 1: Comparison of the Number of Individuals Affected by Dengue Fever per year in East Java and Indonesia (2012-2020).

Table 1: Regions Categorized by Low, Medium, and High Case Clusters in East Java (2012-2020).

Years	Number of Cases		
	Low	Medium	High
2012	Batu City, Lumajang, Probolinggo City, Mojokerto City, Blitar, Probolinggo, Mojokerto, Pamekasan, Madiun City, Magetan, Pasuruan City, Kediri City, Banyuwangi, Malang City, Madiun, Pasuruan, Bondowoso, Trenggalek, Malang, Ngawi, Tulungagung, Situbondo, Pacitan, Tuban, Nganjuk, Sidoarjo, Ponorogo, Blitar City, Jember	Lamongan, Sumenep, Sampang, Gresik, Bangkalan, Kediri, Jombang, Bojonegoro	Surabaya City
2013	Mojokerto City, Mojokerto, Probolinggo City, Blitar City, Magetan, Pamekasan, Madiun City, Madiun, Ngawi, Probolinggo, Batu City, Lumajang, Pasuruan City, Blitar, Tuban, Nganjuk, Sidoarjo, Banyuwangi, Kediri City, Bojonegoro, Situbondo, Trenggalek, Pacitan, Ponorogo, Tulungagung	Pasuruan, Bondowoso, Jombang, Malang City, Sampang, Gresik, Bangkalan, Lamongan, Kediri, Sumenep, Jember, Malang	Surabaya City
2014	Mojokerto City, Mojokerto, Batu City, Magetan, Blitar City, Bojonegoro, Nganjuk, Pamekasan, Blitar, Pasuruan City, Lumajang, Kediri City, Lamongan, Madiun, Malang City, Kediri, Tuban, Sidoarjo, Ngawi, Madiun City, Pasuruan, Sampang, Pacitan, Probolinggo, Jombang, Tulungagung, Situbondo, Trenggalek, Gresik	Bangkalan, Sumenep, Probolinggo City, Ponorogo, Banyuwangi, Bondowoso	Surabaya City, Malang, Jember
2015	Gresik, Mojokerto City, Blitar City, Batu City, Magetan, Lumajang, Tuban, Madiun City, Pasuruan City, Probolinggo City, Kediri City, Malang City, Mojokerto, Madiun, Nganjuk, Blitar, Pamekasan, Situbondo, Probolinggo	Bojonegoro, Sidoarjo, Sampang, Surabaya City, Lamongan, Jombang, Trenggalek, Pasuruan, Kediri, Ponorogo, Ngawi, Bondowoso, Tulungagung, Jember, Banyuwangi, Sumenep, Bangkalan	Malang, Pacitan
2016	Mojokerto City, Batu City, Magetan, Tuban, Blitar City, Madiun City, Nganjuk, Madiun, Blitar, Lumajang, Pasuruan City, Banyuwangi, Kediri City, Pamekasan, Mojokerto, Malang City, Probolinggo, Probolinggo City, Bojonegoro, Lamongan	Sampang, Pasuruan, Ngawi, Bangkalan, Gresik, Situbondo, Ponorogo, Surabaya City, Kediri, Tulungagung	Sumenep, Bondowoso Jombang, Trenggalek Malang, Jember Pacitan, Sidoarjo
2017	Madiun, Blitar, Pasuruan City, Banyuwangi, Malang City, Magetan, Bangkalan, Mojokerto City, Tuban, Bojonegoro, Bondowoso, Lamongan, Ponorogo, Pacitan, Probolinggo City, Tulungagung, Probolinggo, Kediri City	Lumajang, Jombang, Nganjuk, Situbondo, Pasuruan, Madiun City, Malang, Blitar City, Gresik, Pamekasan, Ngawi, Sumenep, Sidoarjo	Jember, Batu City, Trenggalek, Mojokerto, Kediri, Surabaya City, Sampang
2018	Mojokerto City, Probolinggo City, Lumajang, Banyuwangi, Pasuruan City, Malang City, Madiun City, Batu City, Probolinggo, Kediri City, Bangkalan, Situbondo, Madiun, Lamongan, Blitar City, Tuban, Magetan, Pamekasan, Pasuruan, Gresik, Sampang, Nganjuk	Bondowoso, Pacitan, Sidoarjo, Trenggalek, Sumenep, Surabaya City, Blitar, Jember,Mojokerto, Ponorogo, Tulungagung, Kediri	Jombang, Bojonegoro, Malang, Ngawi
2019	Mojokerto City, Batu City, Pasuruan City, Banyuwangi, Lumajang, Bangkalan, Pasuruan, Probolinggo City, Kediri City, Madiun City, Blitar City, Sampang, Mojokerto, Surabaya City, Madiun, Nganjuk, Pamekasan, Jombang, Sidoarjo, Sumenep, Bondowoso, Lamongan, Tuban	Probolinggo, Gresik, Situbondo, Magetan, Trenggalek, Bojonegoro, Malang City, Pacitan, Blitar, Tulungagung, Jember	Kediri, Ngawi, Malang, Ponorogo
2020	Mojokerto City, Nganjuk, Pasuruan City, Tuban, Madiun City, Batu City, Surabaya City, Probolinggo City, Bojonegoro, Gresik, Blitar City, Madiun, Mojokerto, Bangkalan, Lamongan, Pasuruan, Sumenep, Jombang, Sidoarjo, Ponorogo, Lumajang, Kediri City, Probolinggo, pamekasan	Blitar, Banyuwangi, Sampang, Ngawi, Tulungagung, Magetan, Bondowoso, Malang City, Situbondo, Kediri, Trenggalek	Pacitan, Jember, Malang

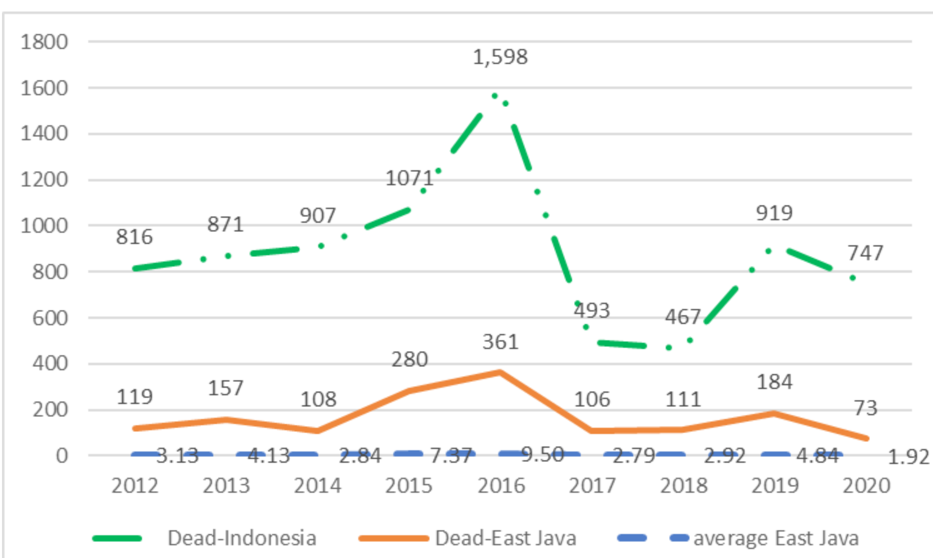


Figure 2: Comparison of the Number of Deaths per Year in East Java and Indonesia (2012-2020).

Table 2: Regions Categorized by Low, Medium, and High Death Clusters in East Java (2012-2020).

Year	Number of Death		
	Low	Medium	High
2012	Ponorogo, Magetan, Sampang, Malang City, Probolinggo City, Pasuruan City, Mojokerto City, Batu City, Pacitan, Trenggalek, Tulungagung, Lumajang, Banyuwangi, Probolinggo, Sidoarjo, Kediri City, Mojokerto, Lamongan	Blitar, Situbondo, Pasuruan, Tuban, Pamekasan, Madiun, Bojonegoro, Madiun City, Jember, Bondowoso, Gresik, Sumenep	Bangkalan, Surabaya City, Malang, Kediri, Jombang, Nganjuk, Blitar City, Ngawi
2013	Pacitan, Trenggalek, Probolinggo City, Pasuruan City, Mojokerto City, Madiun City, Tulungagung, Tuban, Bangkalan, Blitar City, Batu City, Banyuwangi, Sidoarjo, Lamongan, Kediri City, Malang City, Situbondo, Probolinggo, Magetan, Sumenep	Ponorogo, Lumajang, Bondowoso, Mojokerto, Madiun, Pamekasan, Jombang, Sampang, Ngawi, Bojonegoro, Gresik, Blitar, Nganjuk, Pasuruan	Kediri, Jember, Surabaya City, Malang
2014	Pacitan, Trenggalek, Kediri, Mojokerto, Pamekasan, Sumenep, Blitar City, Pasuruan City, Mojokerto City, Batu City, Banyuwangi, Sidoarjo, Magetan, Tuban, Lamongan, Kediri City, Malang City, bangkalan	Ponorogo, Tulungagung, Nganjuk, Madiun, Bojonegoro, Sampang, Madiun City, Blitar, Malang, Situbondo, Probolinggo, Ngawi, Lumajang, Bondowoso, Jombang, Pasuruan, Gresik, Jember, Probolinggo City	Surabaya City
2015	Lumajang, Gresik, Pasuruan City, Mojokerto City, Batu City, Tuban, Kediri City, Blitar City, Madiun City, Pacitan, Magetan, Malang City, Probolinggo, Probolinggo City, Ponorogo, Situbondo, Madiun, Ngawi, bangkalan, Trenggalek, Kediri, Bojonegoro, Lamongan	Pamekasan, Banyuwangi, Nganjuk, Tulungagung, Malang, Blitar, Sumenep, Surabaya City, Bondowoso, Mojokerto, Sampang, Jember, Jombang	Pasuruan, Sidoarjo,
2016	Tuban, Mojokerto City, Pacitan, Gresik, Kediri City, Blitar City, Malang City, Trenggalek, Magetan, Lamongan, Madiun City, Batu City, Madiun, Sampang, Pamekasan, Sumenep, Pasuruan City, Blitar, Lumajang, Surabaya City	Jember, Bangkalan, Ponorogo, Probolinggo, Mojokerto, Nganjuk, Ngawi, Probolinggo City, Bondowoso, Malang, Situbondo, Jombang, Kediri, Bojonegoro, Banyuwangi	Tulungagung, Pasuruan, Sidoarjo
2017	Pacitan, Blitar, Malang, Lumajang, Probolinggo, Sidoarjo, Mojokerto, Madiun, Tuban, Kediri City, Malang City, Pasuruan City, Batu City, Ponorogo, Jombang, Bangkalan, Mojokerto City, Trenggalek, Banyuwangi, Pamekasan, Sumenep	Bondowoso, Situbondo, Magetan, Lamongan, Probolinggo City, Tulungagung, Kediri, Bojonegoro, Jember, Pasuruan, Madiun City, Nganjuk, Blitar City, Surabaya City	Gresik, Sampang, Ngawi
2018	Jember, Mojokerto, Tuban, Bangkalan, Sumenep, Kediri City, Blitar City, Pasuruan City, Mojokerto City, Madiun City, Batu City, Pacitan, Lumajang, Situbondo, Madiun, Gresik, Malang City, Probolinggo City, Surabaya City, Ponorogo, Jombang, Magetan, Lamongan, Sampang, Trenggalek, Malang, Banyuwangi	Probolinggo, Ngawi, Sidoarjo, Blitar, Bondowoso, Nganjuk, Pasuruan, Pamekasan, Kediri	Bojonegoro, Tulungagung
2019	Situbondo, Mojokerto, Mojokerto City, Pacitan, Bondowoso, Bangkalan, Blitar City, Lumajang, Pasuruan, Tuban, Gresik, Sampang, Madiun City, Batu City, Jombang, Ngawi, Lamongan, Kediri City, Malang City, Probolinggo City, Pasuruan City, Surabaya City, Madiun	Trenggalek, Probolinggo, Banyuwangi, Nganjuk, Magetan, Sumenep, Jember, Blitar, Malang, Bojonegoro, Sidoarjo, Pamekasan	Ponorogo, Tulungagung, Kediri
2020	Pasuruan, Mojokerto, Magetan, Tuban, Lamongan, Sampang, Sumenep, Blitar City, Malang City, Mojokerto City, Surabaya City, Blitar, Probolinggo, Nganjuk, Bojonegoro, Gresik, Bangkalan, pamekasan, Kediri City, Probolinggo City, Pasuruan City, Batu City	Ponorogo, Lumajang, Sidoarjo, Madiun, Madiun City, Trenggalek, Jember, Jombang, Ngawi, Bondowoso	Tulungagung, Banyuwangi, Malang, Situbondo, Pacitan, Kediri

2012, the estimated percentage of deaths due to Dengue Fever in a specific region was 3,13/119= 0.03 or (3%). This indicates that Dengue Fever contributed to 3% of highly infectious disease in East Java.

Regions with deaths caused by Dengue Fever were divided into three clusters: low, medium, and high. The number of regions with low death counts was significantly higher compared to those with medium or high death counts. From 2012 to 2020, 19 regions were identified as having high death clusters, including 17 regencies and 2 cities. These regions were Ngawi, Nganjuk, Bangkalan, Kediri, Jombang, Malang, Sidoarjo, Sampang, Pasuruan, Tulungagung, Jember, Gresik, Bojonegoro, Ponorogo, Banyuwangi, Situbondo, Pacitan, Blitar, and Surabaya city. The high number of regions with elevated death clusters underscores the serious public health challenges East Java faces in combating Dengue Fever.

From 2012 to 2020 in East Java, the Case Fatality Rate served as a measure of disease severity. The CFR graph shows fluctuations, with East Java's CFR consistently exceeding the national average (Indonesia's CRF). If the national target for CFR is ≤1%, East Java's CFR has consistently surpassed this threshold, indicating a high severity of Dengue Fever within the population.

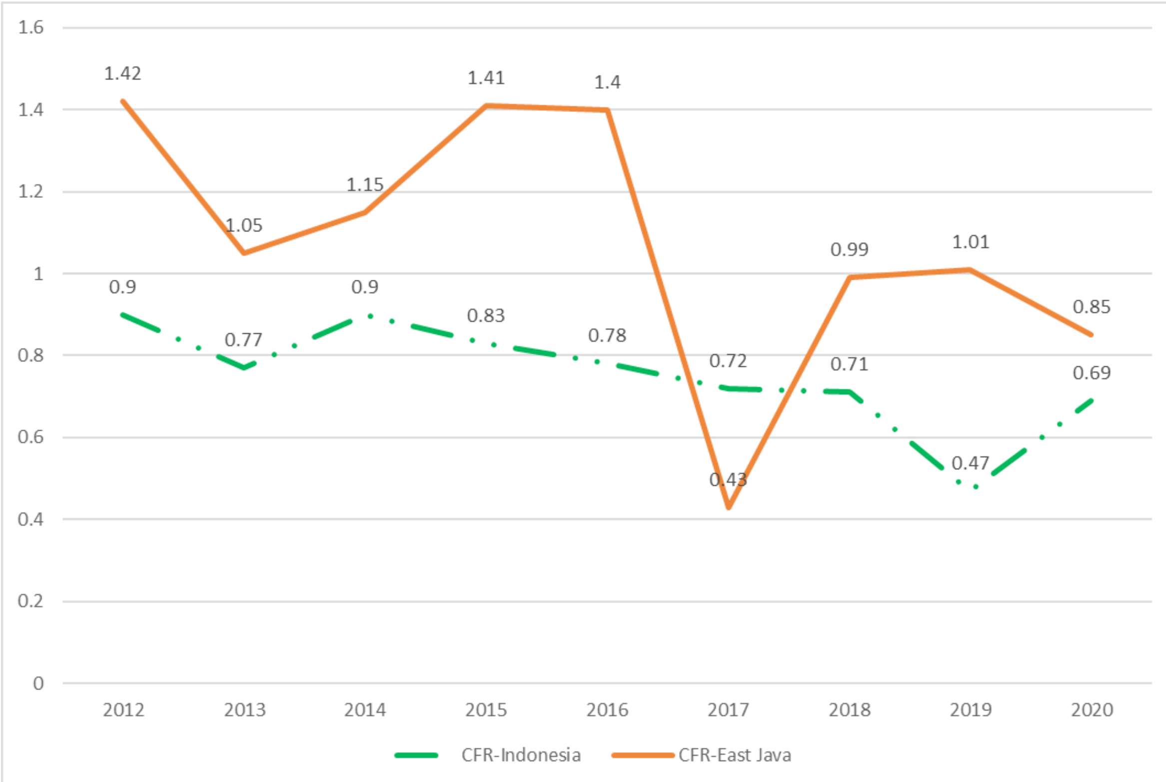


Figure 3: Comparison of Case Fatality Rates in Indonesia and East Java, 2012-2020.

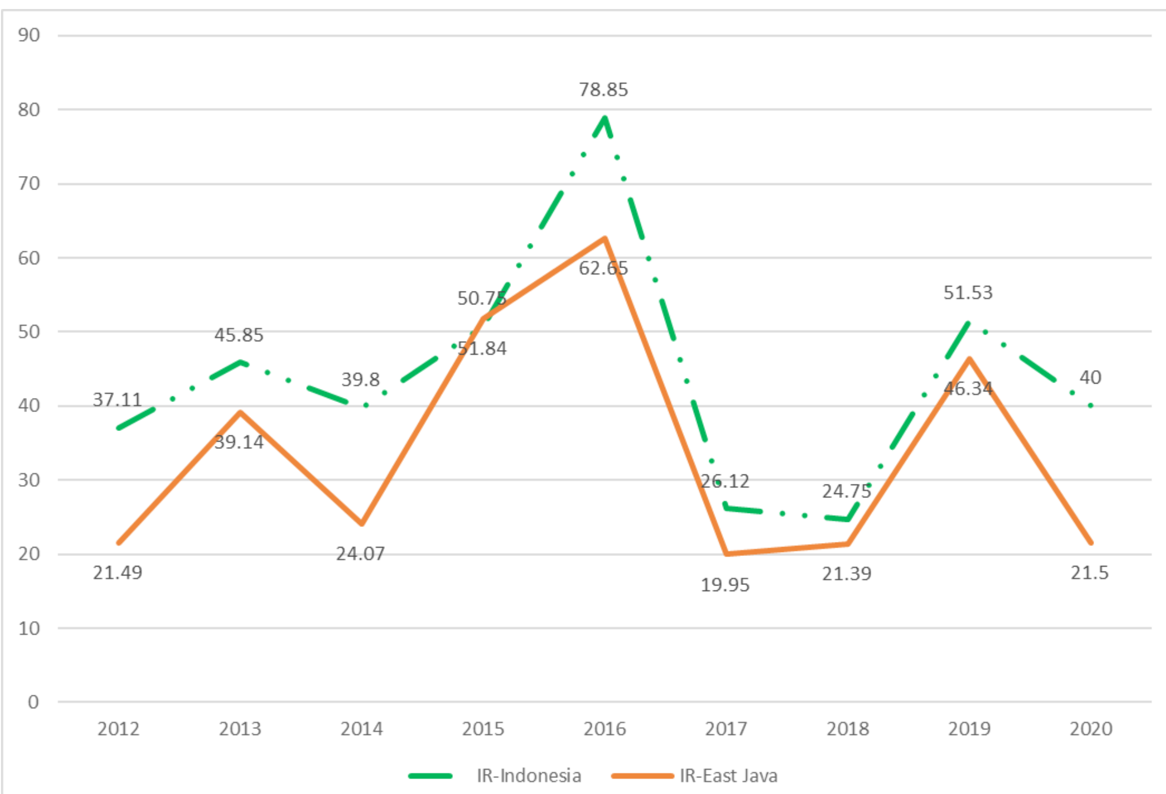


Figure 4: Comparison of Incidence Rates in Indonesia and East Java, 2012-2020.

Table 3: Regions Categorized into Low, Medium, and High CFR Clusters in East Java, 2012-2020.

Year	Number of Death		
	Low	Medium	High
2012	Ponorogo, Magetan, Sampang, Malang City, Probolinggo City, Pasuruan City, Mojokerto City, Batu City, Sidoarjo, Pacitan, Surabaya City, Tulungagung, Trenggalek, Bojonegoro, Lamongan, Banyuwangi, Kediri City	Gresik, Tuban, Bangkalan, Jombang, Sumenep, Kediri, Situbondo, Probolinggo, Pasuruan, Jember, Madiun, Mojokerto, Bondowoso, Blitar City	Nganjuk, Malang, Lumajang, Pamekasan, Ngawi, Madiun City, Blitar
2013	Pacitan, Trenggalek, Probolinggo City, Pasuruan City, Mojokerto City, Madiun City, bangkalan, Tulungagung, Lamongan, Sumenep, Malang City, Tuban, Surabaya City, Batu City, Kediri City, Banyuwangi, Bondowoso, Sidoarjo, Situbondo, Sampang, Ponorogo, Jombang, Jember Gresik, Blitar City, Kediri, Malang	Pasuruan, Bojonegoro, Probolinggo, Lumajang, Blitar, Madiun, Nganjuk, Magetan, Pamekasan	Ngawi, Mojokerto
2014	Pacitan, Trenggalek, Kediri, Mojokerto Pamekasan, Sumenep, Blitar City, Pasuruan City, Mojokerto City, Batu City, Banyuwangi, Malang, Sidoarjo, Tuban, Malang City, Lamongan, Bangkalan, Kediri City, Ponorogo, Jember, Bondowoso	Tulungagung, Magetan, Sampang, Situbondo, Madiun City, Probolinggo, Madiun, Surabaya City, Probolinggo City, Jombang, Ngawi	Surabaya City
2015	Kediri, Lumajang, Lamongan, Gresik, Pasuruan City, Mojokerto City, Batu City, Pacitan, Kediri City, Tuban, Madiun City, Ponorogo, Bangkalan, Ngawi, Malang, Banyuwangi, Probolinggo, Trenggalek, Blitar City, Malang City, Situbondo, Bojonegoro, Sumenep, Tulungagung	Bondowoso, Jember, Madiun, Probolinggo City, Magetan, Surabaya City, Sampang, Pamekasan, Jombang, Nganjuk, Blitar	Pasuruan, Mojokerto, Sidoarjo
2016	Tuban, Mojokerto City, Pacitan, Gresik, Trenggalek, Kediri City, Sumenep, Malang City, Jember, Lamongan, Surabaya City, Blitar City, Sampang, Malang, Ponorogo, Bondowoso, Bangkalan	Ngawi, Situbondo, Jombang, Madiun City, Pamekasan, Madiun, Kediri, Pasuruan City, Probolinggo, Probolinggo City, Lumajang, Blitar, Tulungagung, Magetan, Sidoarjo, Mojokerto	Bojonegoro, Batu City, Pasuruan, Nganjuk, Banyuwangi
2017	Pacitan, Blitar, Malang, Lumajang, Probolinggo, Sidoarjo, Mojokerto, Madiun, Tuban, Kediri City, Malang City, Pasuruan City, Batu City, Trenggalek, Jombang, Sumenep, Pamekasan, Ponorogo, Kediri	Situbondo, bangkalan, Mojokerto City, Jember, Surabaya City, Sampang, Madiun City, Pasuruan, Probolinggo City, Blitar City, Nganjuk	Bondowoso, Lamongan, Tulungagung, Gresik, Magetan, Bojonegoro, Ngawi, Banyuwangi
2018	Jember, Mojokerto, Tuban, Bangkalan, Sumenep, Kediri City, Blitar City, Pasuruan City, Mojokerto City, Madiun City, Batu City, Surabaya City, Pacitan, Malang, Jombang, Ponorogo, Ngawi, Gresik, Situbondo, Madiun, Sampang, Trenggalek, Magetan, Lamongan, Malang City, Blitar, Sidoarjo, Kediri, Bojonegoro	Bondowoso, Lumajang, Tulungagung, Nganjuk, Probolinggo City, Pasuruan, Pamekasan, Probolinggo	Banyuwangi
2019	Situbondo, Mojokerto, Mojokerto City, Pacitan, Ngawi, Bondowoso, Blitar City, Malang, Tuban, Gresik, Bangkalan, Malang City, Jember, Ponorogo, Lamongan, Sampang, Madiun City, Jombang, Trenggalek, Probolinggo, Pasuruan, Surabaya City, Blitar, Lumajang, Madiun, Magetan, Kediri City, Probolinggo City, Bojonegoro, Sumenep	Tulungagung, Kediri, Nganjuk, Sidoarjo, Pamekasan, Pasuruan City, Banyuwangi	Batu City
2020	Pasuruan, Mojokerto, Magetan, Tuban, Lamongan, Sampang, Sumenep, Blitar City, Malang City, Mojokerto City, Surabaya City, Jember, Malang, Blitar, Pamekasan, Probolinggo, Kediri City	Trenggalek, Pacitan, Bangkalan, Ngawi, Gresik, Ponorogo, Lumajang, Bojonegoro, Probolinggo City, Bondowoso, Sidoarjo, Batu City, Tulungagung, Situbondo	Kediri, Jombang, Banyuwangi, Madiun, Pasuruan City, Nganjuk, Madiun City

Another measure used to assess the risk of Dengue Fever is the Incidence Rate, IR per 100,000 population. The IR trend in East Java fluctuates similarly to the national IR in Indonesia. If the target IR for Indonesia is ≤49, East Java remains below this threshold, indicating a lower risk level compared to the national average. The consistent presence of Dengue cases every year in this population suggests that East Java is an endemic region for Dengue fever.

The division of regions based on CFR aims to identify areas with varying levels of disease severity (low, medium, and high clusters). The results of the k-means algorithm reveal that regions in the low CFR cluster have a CFR of ≤1. This indicates alignment between the clustering results and the national target. Meanwhile, regions in the medium and high CFR clusters exceed the national target.

In 2013, 2015, and 2018, there were exceptions where the k-means algorithm identified low CFR clusters with a CFR of ≥2. Specifically, these years highlight that the severity of the disease in certain areas of East Java was significantly higher, or the disease was highly infectious.

Between 2012 and 2020, 23 regions (comprising 20 regencies and 3 cities) were identified as having high CFR clusters at some point. These regions include Lumajang, Pamekasan, Malang, Ngawi, Mojokerto, Bojonegoro, Nganjuk, Blitar, Pasuruan, Magetan, Bondowoso, Lamongan, Tulungagung, Gresik, Madiun, Sidoarjo, Jombang, Banyuwangi, Kediri, Madiun City, Pasuruan City, and Batu City. Although these regions did not consistently have high CFRs throughout the entire period, the majority of areas in the high

CFR cluster were regencies. Some regencies/cities experienced repeated entries into the high CFR cluster during 2012-2020, underscoring the need for continuous vigilance against the severity of Dengue fever.

No region remained in the low CFR cluster for all nine consecutive years (2012-2020). However, some areas were classified as low CFR clusters in certain years, and similarly, some were identified as medium CFR clusters annually. Regions in the low CFR cluster were more numerous compared to those in the medium or high clusters.

The findings indicate that a region's CFR cluster category can change over time. For example, a region might start with a low CFR in one year and shift to a medium or high CFR cluster in another year, or vice versa. This variability may be influenced by regional factors such as population density, inter-regional mobility, environmental conditions, and the availability of adequate health infrastructure.

Next, a graphical plot was created to analyze annual data for each region based on the following criteria: 1. Low number of cases, deaths and CFR 2. High number of cases, deaths, and CFR 3. Low number of cases but high number of deaths and/or high CFR, 4. High number of cases but low number of deaths and low CFR.

Based on the plotted graphs, several regencies and cities were identified as having clusters with a low number of cases, low number of deaths, and low CFR. Among these regions, some maintained a low cluster status throughout 2012-2020. However, no region remained in the low cluster category for all nine consecutive years.

Regions classified in the high cluster for the number of cases, number of deaths, and CFR were also identified on some regencies/cities. On average, two regencies/cities (approximately 5% of all regions in East Java) fell into this high cluster category each year.

Table 4: Regions with High Clusters for Number of Cases, Number of Deaths, and CFR, 2012-2020.

Year– Region	Total
2012- Blitar	1 Regency
2013- Mojokerto	1 Regency
2014- Lumajang	1 Regency
2015- Sidoarjo	1 Regency
2016- Banyuwangi, Nganjuk, Pasuruan	3 Regencies
2017- Bojonegoro, Banyuwangi, Ngawi	3 Regencies
2018- Banyuwangi	1 Regency
2019- Batu City	1 City
2020- Madiun City, Nganjuk, Pasuruan City	2 Regencies and 1 City

There is no regency or city in East Java fell into the cluster with a low number of cases but a high number of deaths and/or high CFR. Similarly, no region was identified as having a high number of cases but a low number of deaths and low CFR.

Findings from 2012 to 2020 indicate that regions with a low number of cases tend to have a low or moderate number of deaths and/or CFR. Conversely, regions with a high number of cases are more likely to have a high number of deaths and/or CFR. It can be concluded that an increase in the number of cases is likely to lead to a corresponding rise in the number of deaths and CFR.

Regions experiencing high numbers of deaths and/or high case fatality rates (CFR) naturally make significant efforts to reduce these figures. By implementing preventive programs to lower the number of cases, regions with high case numbers, deaths, and/or CFR can transition to lower clusters over time.

Between 2012 and 2020, a total of 123,014 cases of Dengue Fever were reported in East Java. This figure was calculated by summing the annual case totals over the nine-year period. The year 2016 saw a notable surge in cases, accounting for 20.6% of all cases reported during 2012–2020. This percentage was derived by dividing the total cases in 2016 by the total cases over the period of 2012-2020, resulting in 0.206 (25,339/123,014 = 0.206).

East Java is one of Indonesia's most populous provinces, despite covering only 2.49% of the country's total land area. The high number of dengue cases in East Java can be attributed to its significantly higher population density compared to other provinces in Indonesia [13]. Additionally, the region's conducive environment for mosquito breeding, coupled with its dense population and human activity, creates ideal conditions for disease transmission. As such, areas with higher population densities tend to exhibit higher rates of disease transmission. This highlights the strong correlation between disease spread and environmental, spatial, and socioeconomic factors. All these factors within a region can influence the health outcomes of its population and the emergence of diseases [35].

Clustering

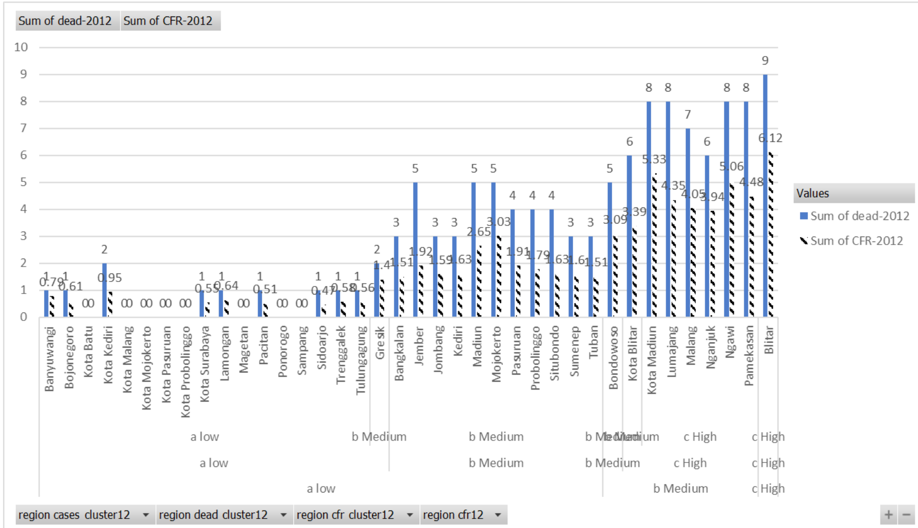
Clustering is a data analysis method that involves grouping data into distinct clusters. Data points within a cluster share similar characteristics, while those in different clusters exhibit distinct traits. This approach is particularly useful for identifying and categorizing previously undefined objects within a dataset.

The clustering model organizes data into separate groups based on shared attributes. By applying clustering to case numbers, deaths, and CFR, we can identify disease patterns tailored to specific regions. This model efficiently categorizes regions into similar groups based on their case numbers, deaths, and CFR characteristics.

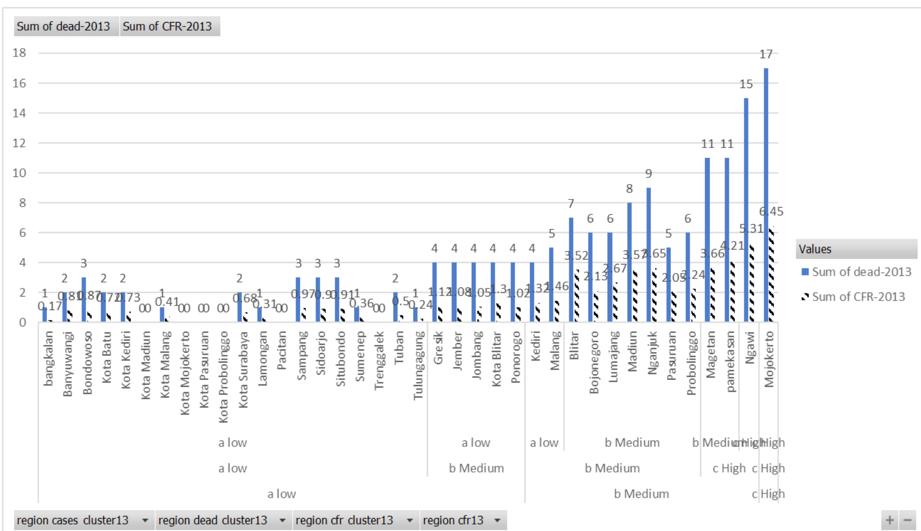
Analytics Algorithms

The high-speed algorithm used for clustering in this study is k-means. The k-means method groups unlabeled data points into distinct clusters based on their similarities. It identifies the characteristics of objects within the data and organizes them into cohesive sets.

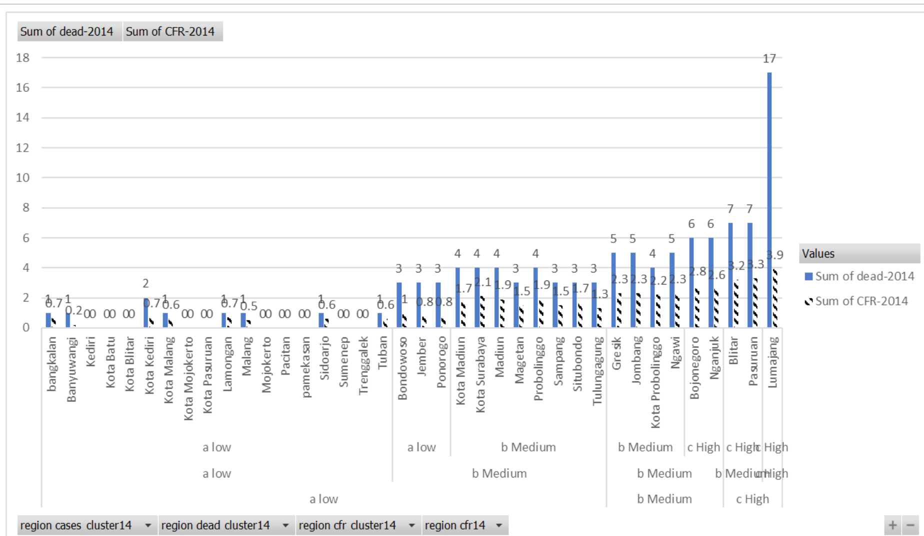
2012



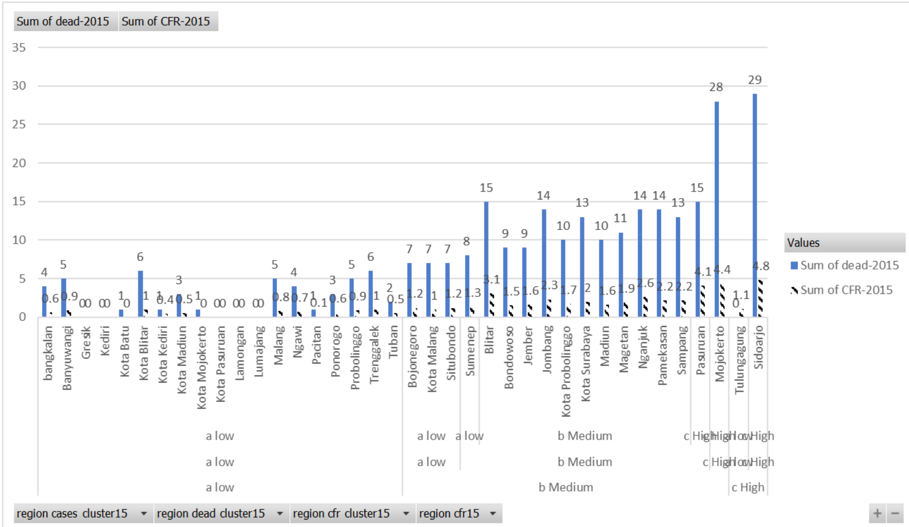
2013



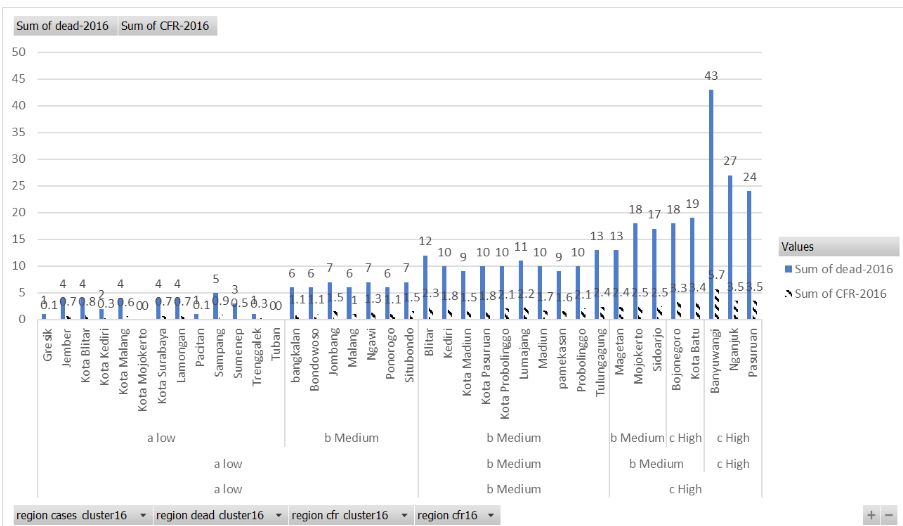
2014



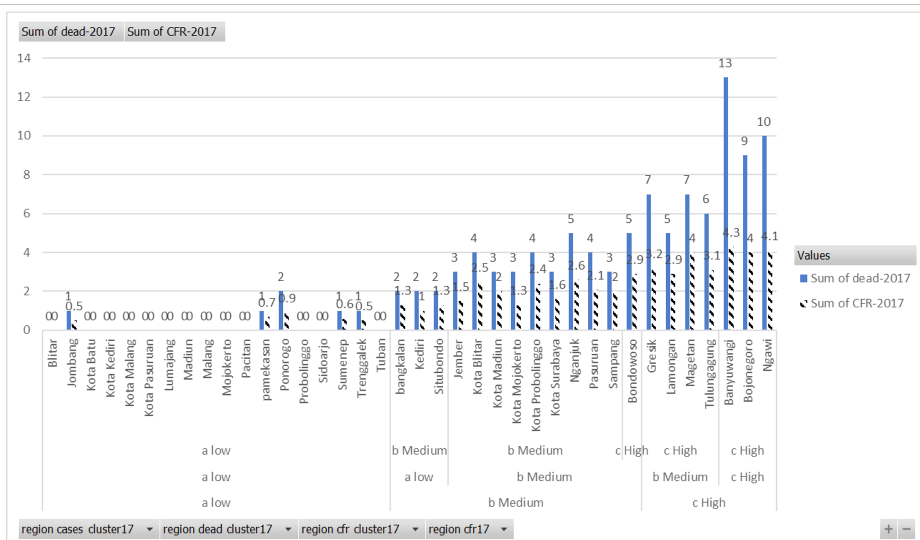
2015



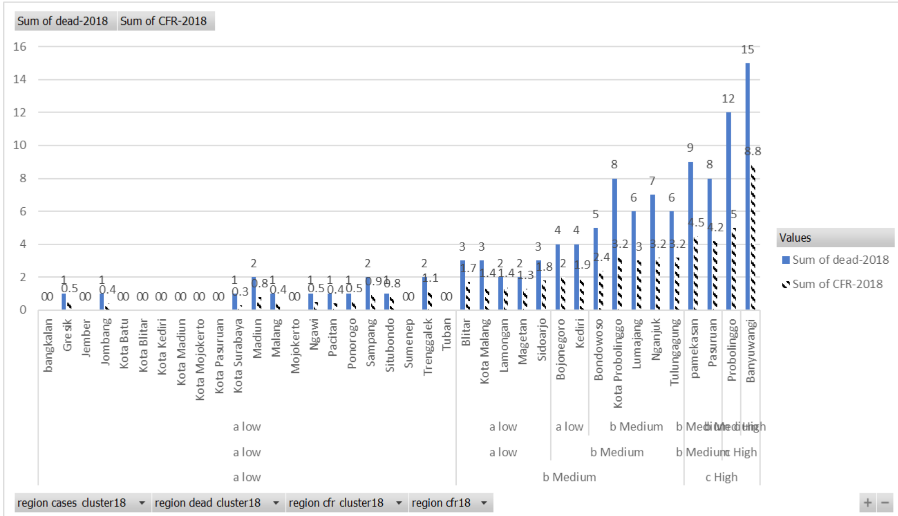
2016



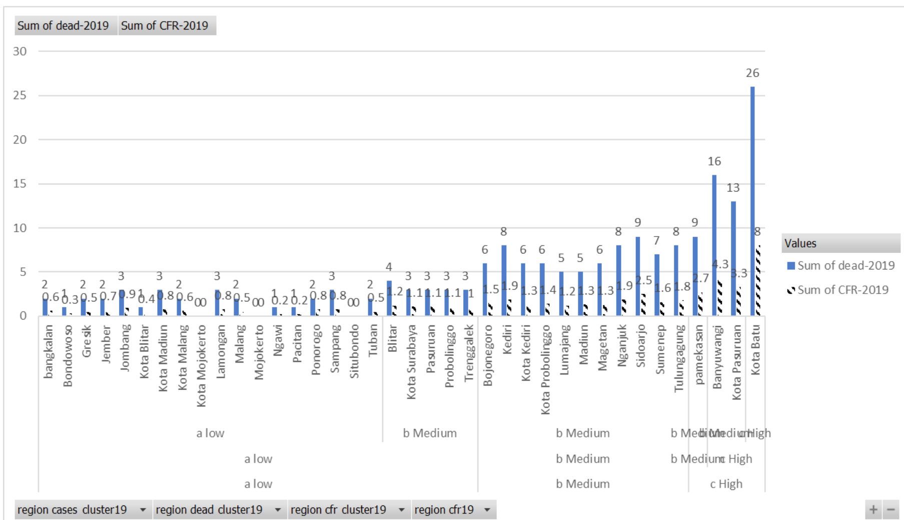
2017



2018



2019



2020

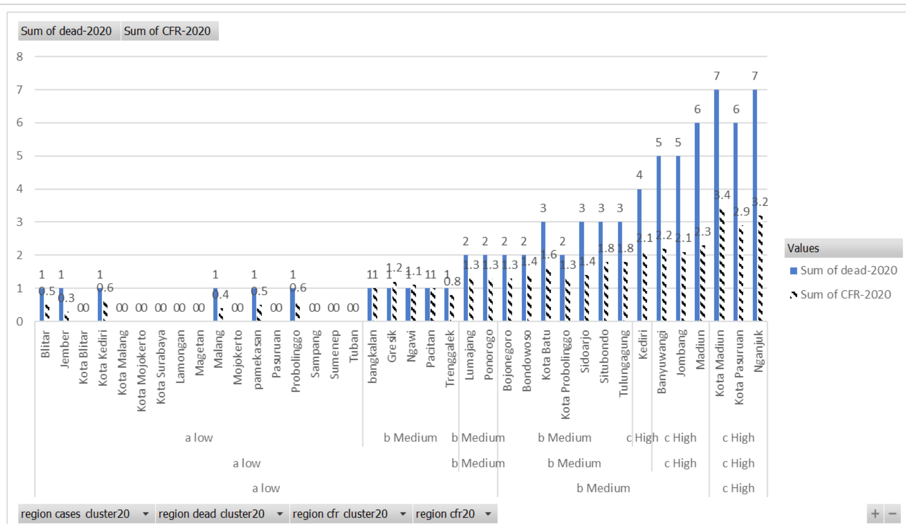


Figure 5: Regions with Low, Medium, and High Clusters of Case Numbers, Deaths, and CFR, 2012–2020.

The clustering results reveal that a region's cluster status can fluctuate over time. For example, a region may shift from a low to a medium or high cluster, or vice versa. It was also observed that maintaining a low cluster status for case numbers, deaths, or CFR is challenging for any region. Furthermore, each year in East Java, at least two regions are classified as high clusters for case numbers, deaths, and CFR. Regions with high case numbers also tend to have high death tolls and/or CFR.

Infectious Disease

Based on the findings using machine learning (ML), the rise in dengue cases is closely linked to regional factors and the adaptability of mosquito vectors. To address this, prevention efforts should focus on individual and household levels, with an emphasis on early detection through environmental monitoring. Public health providers play a critical role in raising community awareness, conducting regional surveillance, and controlling mosquito breeding. Therefore, the key to protecting public health lies in preventing outbreaks through active collaboration between communities and government agencies to reduce disease risks.

Conclusion

This study applies data analytics technology to the control of dengue fever in communities. Data related to dengue fever and its associated characteristics such as case numbers, deaths, and CFR were identified and analyzed. By leveraging this data, public health issues can be uncovered, enabling effective preventive and control measures. Data collection was conducted using official government websites as primary sources. Based on the gathered information, a technical framework for data analysis was developed.

The findings reveal that dengue fever remains a serious public health issue in East Java. Regions continue to struggle to maintain low clusters for case numbers, deaths, and CFR. Effective eradication and control of the disease require the involvement of all stakeholders, including communities, healthcare providers, and government agencies.

Acknowledgement

We extend our gratitude to the Ministry of Health of the Republic of Indonesia and the East Java Health Office for providing dengue fever data to support this research. The data can be accessed directly through their official websites.

References

1. Torrecilla JL, Romo J. Data learning from big data. Stat Probab Lett. 2018; 136: 15-19.
2. Bansal S, Chowell G, Simonsen L, et al. Big Data for infectious disease surveillance and modeling. J Infect Dis. 2016; 214: 375-379.
3. Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning, and privacy. Annu Rev Public Health. 2018; 39: 95-112.
4. East Java Public Health Office. Health profile of East Java 2013. 2014.
5. East Java Public Health Office. Health profile of East Java 2017. 2018.
6. East Java Public Health Office. Health profile of East Java 2020. 2021.
7. Gubler DJ. Dengue and dengue hemorrhagic fever. Clin Microbiol Rev. 1998; 11: 480-496.
8. Abramson JH, Abramson ZH. Making sense of data: A self instruction manual on the interpretation of epidemiological data. New York: Oxford University Press. 1988; 1-326.
9. Indawati R, Hendrati LH, Widati S. The early vigilance of dengue hemorrhagic fever outbreak in the community. Jurnal Kesehatan Masyarakat. 2021; 16: 366-376.
10. Tipayamongkholgu M, Lisakulruk S. Socio-geographical factors in vulnerability to dengue in Thai villages: A spatial regression analysis. Geospat Health. 2011; 5: 191-198.
11. Health Ministry of Republic Indonesia. Indonesia health profile 2010. 2011.
12. Health Ministry of Republic Indonesia. Indonesia health profile 2011. 2012.
13. Health Ministry of Republic Indonesia. Indonesia health profile 2020. 2021.
14. Health Ministry of Republic Indonesia. Indonesia health profile 2012. 2013.
15. Health Ministry of Republic Indonesia. Indonesia health profile 2013. 2014.
16. Health Ministry of Republic Indonesia. Indonesia health profile 2014. 2015.
17. Health Ministry of Republic Indonesia. Indonesia health profile 2015. 2016.
18. Health Ministry of Republic Indonesia. Indonesia health profile 2016. 2017.
19. Health Ministry of Republic Indonesia. Indonesia health profile 2017. 2018.
20. Health Ministry of Republic Indonesia. Indonesia health profile 2018. 2019.
21. Health Ministry of Republic Indonesia. Indonesia health profile 2019. 2020.
22. Gothai E, Balasubramanie P. An efficient way for clustering using alternative decision tree. Am J Appl Sci. 2012; 9: 531-534.
23. Wang H, Song M. Ckmeans.1d.dp: optimal k-means clustering in one Dimension by dynamic programming. RJ. 2011; 3: 29-33.
24. Tayal MA, Raghuwanshi MM. Review on various clustering methods for the image data. J Emerg Trends Comput Inf Sci. 2011; 2: 34-38.
25. East Java Public Health Office. Health profile of East Java 2012. 2013.
26. East Java Public Health Office. Health profile of East Java 2014. 2015.

27. East Java Public Health Office. Health profile of East Java 2015. 2016.
28. East Java Public Health Office. Health profile of East Java 2016. 2017.
29. East Java Public Health Office. Health profile of East Java 2018. 2019.
30. East Java Public Health Office. Health profile of East Java 2019. 2020.
31. Lara JA, Lizcano D, Martinez MA, et al. A System for knowledge discovery in e-learning environments within the european higher education area – application to student data from open university of Madrid, UDIMA. Comput Educ. 2014; 72: 23-36.
32. Seyedan M, Mafakheri F. Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. J Big Data. 2020; 7: 1-22.
33. Han J, Kamber M. Data mining: data mining concepts and techniques. Elsevier. 2012; 1-740.
34. Rubin DB. Inference and missing data. Biometrika. 1976; 63: 581-592.
35. Martins ABS, Alencar CH. Ecoepidemiology of Dengue in Brazil: From the virus to the environment. One Heal Implement Res. 2022; 2: 1-14.