

K-means Clustering Algorithm for Dengue Disease Detection using Tanagra Tool

P.Yogapriya^{1*}, P.Geetha²

¹Department of Computer Science, Dr.Umayal Ramanathan College For Women, Karaikudi, India

²Department of Computer Science, Dr.Umayal Ramanathan College For Women, Karaikudi, India

*Corresponding Author: saiyoga2896@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v7i2.765768> | Available online at: www.ijcseonline.org

Accepted: 18/Feb/2019, Published: 28/Feb/2019

Abstract— Dengue disease caused over the tropical and sub-tropical area which spread by Aedes mosquitoes. Dengue has to turn into a severe healthiness problem occurs frequently in the humid and sub-tropical region. The scientists use the data mining algorithm for preventing and protecting different diseases like Dengue disease. This analysis of the attack of Dengue fever in different districts mainly Puducherry, Tamil Nadu. This paper helps to apply the algorithm for clustering of Dengue fever. After that, the construction of the clustering algorithm depends on the graph-based dataset. The K-Means clustering algorithm is applied to detect Dengue fever.

Keywords— Dengue, Clustering, K-Means.

I. INTRODUCTION

Data mining is the process of extracting useful information. Sometimes it is called Knowledge Discovery in Databases (KDD). Data mining is about finding new information from pre-existing databases. The information obtained from data mining is optimistically both new and useful. It is the procedure of mining knowledge from data. In many cases, data is stored, so it can be used later. The data mining is the process of retrieving hidden information from the huge sets of data.

Dengue is also known as ‘breakbone’ fever. Dengue is a mosquito-borne viral infection which is extended and affected by Aedes mosquitoes. Dengue has to turn into a severe healthiness problem occurs frequently in the humid and sub-tropical region.

The clustering is one of the techniques used in data mining which is the method of finding related groups of objects or data in a data set. Clustering is the process of separating a set of objects or data into a group of relevant sub-classes. Clustering analysis has been a prominent issue in data mining due to its heterogeneity of applications.

According to this analysis, It deals with the clustering algorithm namely, K-Means propose to split-up “n” objects or observation into “k” clusters. It is an unsupervised learning algorithm that is used to interpret the clustering

problem. K-Means creates K-groups from a set of the object then the member of the object is more similar. K-Means clustering analysis is a popular clustering analysis technique for exploring dataset. First, exhibit the overall performance of dengue in the district Puducherry. Then describe with the help of geographical pictures. After that construction of the K-Means clustering algorithm depends on the graph-based data set. Based on these performance K-Means algorithm performed better in terms of clustering accuracy and execution time.

In this paper, it contains the introduction of the data mining concept. It contains the related work of the clustering algorithm mainly use as the partitioning clustering method. To detect some measures of Dengue fever with their accuracy. The architecture mainly focuses on collecting the database and formatting the database to initializing the K-Means clustering algorithm. This paper proposes the following methodology which is in Figure 1. The first step deals with data collection using a dataset from tropical and sub-tropical areas in Puducherry district. The input data have been collected from urban areas. The second step deals with data preparation using dengue sufferer of data possessed from domestic clustering of data. The dataset was gathered in Puducherry district. In data preparation method the attributes of the dataset which contains EPID, Fever, Bleeding, Myalgia, Flu, Fatigue, Results to identify the type of fever.

To detect the Dengue fever with Dengue-related dataset using K-Means clustering algorithm. In this analysis of K-Means clustering can describe Univariate discrete stat in order to measure of Gini index and Distribution. The performance of the algorithm can increase the recall accuracy of the results. This analysis is the most effective to predict the Dengue fever with their results. In this prediction shows 60% of results caused by Dengue fever with positive values. In the future, To implement the hierarchical clustering method to predict Dengue fever.

II. RELATED WORK

P.Manivannan, Dr. P. Isakki et al presented a paper has been proposed four stages namely preprocessing, attribute selection, K-Medoid clustering and predicting the dengue fever. R 3.3.2 tool is used for preprocessing the dengue data set and compare the results produced from those above-mentioned algorithms [1].

Sahanaa C et al evaluate the condition of being diseased and the state of being to death of dengue for a period of five years. The data were sourced from the National Health Profile 2017, Survey of secondary data was done, Graph was devised to study the trend of the disease which has been implemented [2].

Shoukat et al presented to analyzes the attack of dengue fever in the district Jhelum, Pakistan. Dataset obtained from Executive District Officer(EDO) in Jhelum. K-Means, K-Medoids, DBSCAN, Optics are used. Contrast the results build from those algorithms [3].

P.Sathya, A.Sumathi et al proposed toward prediction of dengue infection using clustering techniques. The dataset was gathered from Lotus and 24care hospitals. Estimate the implementation of all the techniques separately based on tables and graphs be controlled upon the dataset [4].

III. METHODOLOGY

EXISTING METHODOLOGY

Many existing, industrialized, investigate dataset contains missing values. They are introduced due to a variety of reasons, such as physical data access, tools errors and inaccurate quantifications. The finding of imperfect data is easy in most cases, looking for null values in a dataset. Three types of harms generally associated with missing values such as loss of efficiency, an obstacle in managing and analyzing the data, unfairness resulting from differences among missing and complete data. In existing the K-Means clustering algorithm with Tanagra using some different dataset such as a Car, Bank dataset. In that dataset, It can compare cluster K-means1 and cluster K-Means 2 with discrete and continuous value for finding mean standard deviation and recall accuracy for that dataset.

PROPOSED METHODOLOGY

To predict the attack of dengue fever in different districts mainly Puducherry. This paper proposes the following methodology which is in Figure 1. The first step deals with data collection using a dataset from tropical and sub-tropical areas in Puducherry district. The input data have been collected from urban areas. The second step deals with data preparation using dengue sufferer of data possessed from domestic clustering of data. The dataset was gathered in Puducherry district. In data preparation method the attributes of the dataset which contains EPID, Fever, Bleeding, Myalgia, Flu, Fatigue, Results to identify the type of fever. To construct the K-Means clustering algorithm for initializing K clusters into N partitions of the dataset. To overcome the existing method using Tanagra tool to predict the attack of dengue fever in the form of the accuracy of a dataset.

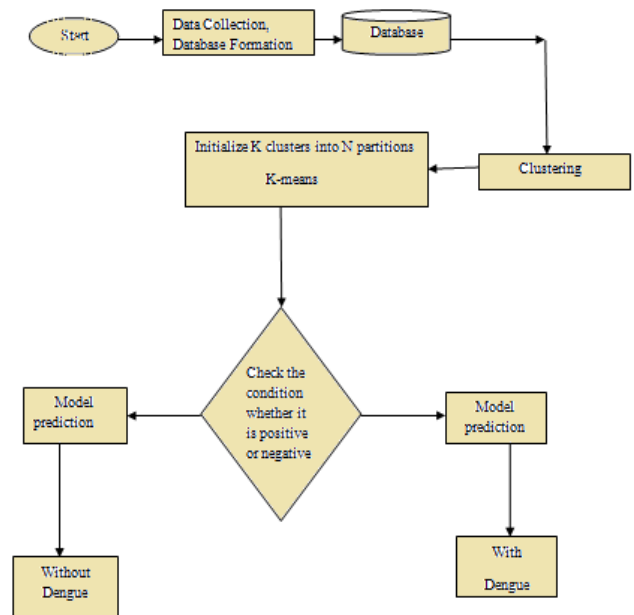


Figure.1 Dataflow Diagram for Proposed Methodology

DATASET CONSTRUCTION

(a) TOOL DESCRIPTION

Tanagra tool was written as an aid to education and research on data mining developed by Ricco Rakotomalala. In the Tanagra tool, the data can be imported and exported in the text files. Analyses available in the Tanagra tool which includes Association Rule Mining, Clustering, Classification. Tanagra is an open source software designed primarily for research use. In Tanagra software, the dataset can be imported into different data formats such as text files, ARFF files, CSV files, as well as Microsoft Excel, IBM Visual Warehouse, and Oracle Express formats. In this analysis of

Dengue fever, the dataset can be imported into Microsoft Excel file formats.

(b)DATASET DESCRIPTION

In the dataset description, Dengue dataset contains 7 attributes and 115 examples. Attribute includes EPID, Fever, Bleeding, Myalgia, Flu, Fatigue, Results. The attributes categorized into two types of values such as Discrete value, Continue value. In this Dengue dataset, 6 attributes can be classified into Discrete values and the remaining one as continue value.

DATASET DESCRIPTION

Dataset description
 7 attribute(s)
 115 example(s)

Attribute	Category	Informations
EPID	Continue	-
Fever	Discrete	3 values
Bleeding	Discrete	2 values
Myalgia	Discrete	2 values
Flu	Discrete	2 values
Fatigue	Discrete	2 values
Results	Discrete	2 values

Group characterization 1

Parameters

Normalization : 0

Results

Description of "Cluster_KMeans_1"

Cluster_KMeans_1=c_kmeans_1				Cluster_KMeans_1=c_kmeans_2					
Examples	[85.1 %] 57			Examples	[14.9 %] 10				
Att - Desc	Test value	Group	Overall	Att - Desc	Test value	Group	Overall		
Continuous attributes : Mean (StdDev)				Continuous attributes : Mean (StdDev)					
Discrete attributes : [Recall] Accuracy				Discrete attributes : [Recall] Accuracy					
Bleeding=No	8.12	[100.0 %]	100.0 %	85.1 %	Bleeding=Yes	8.12	[100.0 %]	100.0 %	14.9 %
Myalgia=No	2.14	[100.0 %]	33.3 %	28.4 %	Myalgia=Yes	2.14	[20.8 %]	100.0 %	71.6 %
Fatigue=Yes	1.50	[89.4 %]	73.7 %	70.1 %	Fatigue=No	1.50	[25.0 %]	50.0 %	29.9 %
Fever=No	0.86	[100.0 %]	7.0 %	6.0 %	Fever=Yes	1.07	[16.4 %]	100.0 %	91.0 %
Flu=Yes	0.74	[100.0 %]	5.3 %	4.5 %	Flu=No	0.74	[15.6 %]	100.0 %	95.5 %
Fever=yes	0.60	[100.0 %]	3.5 %	3.0 %	Fever=yes	-0.60	[0.0 %]	0.0 %	3.0 %
Flu=No	-0.74	[84.4 %]	94.7 %	95.5 %	Flu=Yes	-0.74	[0.0 %]	0.0 %	4.5 %
Fever=Yes	-1.07	[83.6 %]	89.5 %	91.0 %	Fever=No	-0.86	[0.0 %]	0.0 %	6.0 %

Figure.2 Group Characterization of K-Means Clustering

In Figure.2 To detect the attack of Dengue fever using K-Means clustering algorithm can be used different components in order to find the accuracy results. In K-means clustering split up into "n" objects into "K" clusters. In this analysis, the dataset can cluster the results of cluster K-Means1 which can split into cluster c_K-Means_1 and cluster c_K-Means_2 with their recall accuracy using discrete and continue value.

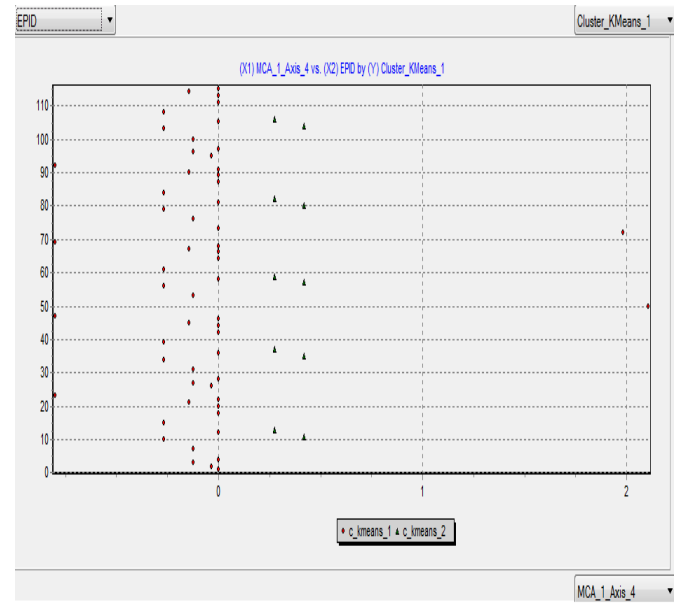


Figure.3 Scatter plot of K-Means

In Figure.3 The scatter plot of K-Means clustering describes as several attributes in this analysis the cluster_K-Means_1 can plot the points into different shapes and colors. Therefore c_K-Means_1 in rounded shape and c_K-Means_2 in a triangle shape. Finally, exporting the dataset in K-Means clustering with Tanagra.

IV. RESULTS AND DISCUSSION

As per the secondary analysis of the data available on Dengue, The Dengue case reported in Puducherry District. Dengue is a break borne viral disease. To detect the Dengue fever with Dengue-related dataset using K-Means clustering algorithm. In this analysis of K- Means clustering can describe Univariate discrete stat in order to measure of Gini index and Distribution. A univariate discrete stat which expressed the relative mean absolute difference is a measure of the average absolute difference of two independent values. Gini index is also known as the Gini coefficient which is used to measure statistical distribution. This paper proposed a Univariate discrete stat with clusters K-Means1 and cluster K-Means2 in order to measure the statistical difference of two independent values.

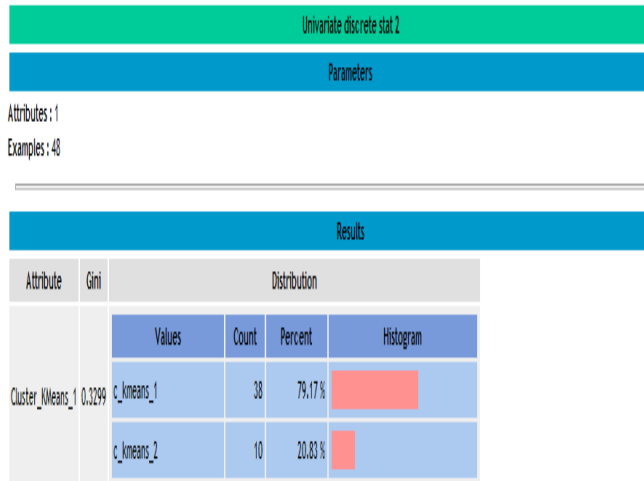


Figure.4 Univariate Discrete Stat Results

V. CONCLUSION AND FUTURE SCOPE

The proposed K-Means clustering algorithm with Tanagra software in Dengue dataset shows better recall accuracy in the results. The proposed K-Means clustering algorithm is used the efficient method of choosing the initial K-Means as n objects into K clusters. The performance of the algorithm can increase the recall accuracy of the results. This analysis is the most effective to predict the Dengue fever with their results. In this prediction shows 60% of results caused by Dengue fever with positive values. In the future, To implement the hierarchical clustering method to predict Dengue fever.

REFERENCES

- [1] P.Manivannan, P. Isakki @ Devi., “Dengue fever prediction using K-Medoid clustering Algorithm”, International Journal of Innovative Research in Computer and Communication Engineering, volume.5, Special Issue.1, March 2017.
- [2] Sahana C.*, Amit Kumar Mishra, Joy Bazroy., “Trend of Morbidity and Mortality of Dengue in Tamil Nadu and Puducherry, South India”, International Journal of Community Medicine and Public Health, volume.5, Special Issue.1, January 2018.
- [3] Kamran Shaukat et al., “Dengue Fever in Perspective of Clustering Algorithms”, International Journal of Data Mining in Genomics and Proteomics, volume.6, Special Issue.3, 2015.
- [4] P.Sathya, Mrs.A.Sumathi.,” Predicting Dengue Fever Using Data Mining techniques”, International Journal of Computer Science and Technology, volume.6, Special Issue.2, March-April 2018.
- [5] S.Muthukumar, E.Ramaraj., “A Multilayered Backpropagation Algorithm to Predict Significant Attributes of UG Pursuing Students Absenteeism at Rural Educational Institution”, International Journal of Computer Science and Engineering, volume.6, Issue.3, December 2018.
- [6] P.Manivannan, P.Isakki Devi., “Dengue Fever Prediction using K-Means Clustering Algorithm”, International Conference Techniques in Control, Optimization and Signal Processing, volume.5, Special Issue.1, March 2017.

- [7] KR.Sivabalan, E.Ramaraj., “Remote Sensing Satellites and its agricultural development technical aspects”, International Journal of Computer Science and Engineering, volume.6, Special Issue.9, September 2018.
- [8] Mohammed Shahadat Hossain, Ishrat Binte Habib, Karl Andersson, “A Belief Rule-Based Expert System to Diagnose Dengue Fever under Uncertainty”, IEEE Computing Conference, volume.1, Special Issue.5, July 2017.
- [9] Shamimul Hasan, Sami Faisal Jamdar et al., “Dengue Virus: A global human threat: Review of Literature”, Journal of International Society of Preventive and Community Dentistry, volume.6, Special Issue.1, Jan-Feb 2016.

Authors Profile

P.Yogapriya is currently pursuing the M.Phil Degree in computer science at the Department of Computer Science, Dr.Umayal Ramanathan College for Women, Karaikudi. Her research interest includes Dengue Disease and Data mining.



P.Geetha is working as the Associate Professor in the Department of Computer Science, Dr. Umayal Ramanathan College For Women, Karaikudi. She has the sound knowledge in many research fields especially in Data mining, Big Data, and Analytics. She has published 16 international journals.

