# K Medoid Algorithm to Forcast the Lake Level

**Priyanka Dey[1] Mr. Avinash Dhole[2]**
[1]M.Tech. Scholar [2]Professor & Head of Department
[1]Department of Software Engineering [2]Department of Computer Science & Engineering
[1,2]RITEE, Raipur, India

*Abstract*— The shifting of water level in the lake creates fluctuation that results in the in balance of water label from its base level. The people who are leaving in the bank of these lakes remains in danger since the label of the lake increases or decreases and once it cross its upper limit it will create disaster. The basis of Remote sensing method there are various system has been designed to forecast this variation of water in the lakes. So it becomes very much important or it is a biggest challenge in order to identify the lake level so that disaster can we avoid. In this method we have designed an algorithm in MATLAB to detect the lake level so that the above-mentioned problem can overcome.

***Key words:*** Lake Level Change, Endangered Lakes, Meteorological Database, Clustering

## I. Introduction

Nowadays clustering technique become more important as the trend of data analysis, processing is increased through the computer. Clustering is having wide application in various fields such as recognition of pattern, artificial intelligence etc. Clustering of data is done under vigorous development. Cluster analysis has recently become a highly active topic in data mining research. As a branch of statistics, cluster analysis has been extensively studied for many years, focusing mainly on distance-based cluster analysis [1-4].

Accurate predictions and a reliable foresight of sea level behavior have always been important in water resource management scenarios. Lake level variations are complex outcomes of different site-specific geographical and meteorological variables, including precipitation, runoff, evaporation, temperature [3]. Data mining is a process which finds useful patterns from a large amount of data. The variables defining weather conditions vary continuously with time, prediction model can be developed either statistically or by using some other means like a decision tree, artificial neural networks, regression, clustering techniques of data mining. Weather prediction is a form of data mining which is concerned with finding hidden patterns inside largely available meteorological data [2].

## II. Objective

The object of this paper is to develop a method to obtain the lake level from the meteorological data by using k-Medoid algorithm.

## III. Basic K-means Algorithm for Finding K Clusters

1) Select K points as the initial centroids.
2) Assign all points to the closest centroid.
3) Recompute the centroid of each cluster.
4) Repeat steps 2 and 3 until the centroids don't change.

### A. *Implementation of Algorithm*

#### 1) *K-medoid Algorithm*

– Initialize

Randomly select k of the n data points as the medoids

– Assignment Step

Associate each data point to the closest medoid.

– Update Step

For each medoid m and each data point o associated to m swap m and o and compute the total cost of the configuration (that is, the average dissimilarity of o to all the data points associated to m).

Select the medoid o with the lowest cost of the configuration. Repeat alternating steps 2 and 3 until there is no change in the assignments.

– Eg. Dataset

Suppose Cluster the data set of ten objects into two clusters i.e. k = 2.

| | | |
|---|---|---|
| X1 | 2 | 6 |
| X2 | 3 | 4 |
| X3 | 3 | 8 |
| X4 | 4 | 7 |
| X5 | 6 | 2 |
| X6 | 6 | 4 |
| X7 | 7 | 3 |
| X8 | 7 | 4 |
| X9 | 8 | 5 |
| X10 | 7 | 6 |

1) Step 1

Initialize k centers. Let us assume x2 and x8 are selected as medoids, so the centers are c1 = (3, 4) and c2 = (7,4). Calculate distances to each center so as to associate each data object to its nearest medoid. Cost can be calculated as follows

$$\mathrm{cost}(x, c) = \sum_{i=1}^{d} |x_i - c_i|$$

| Cost (distance) to c1 | | | | | |
|---|---|---|---|---|---|
| i | c1 | | Data objects (Xi) | | Cost (distance) |
| 1 | 3 | 4 | 2 | 6 | 3 |
| 3 | 3 | 4 | 3 | 8 | 4 |
| 4 | 3 | 4 | 4 | 7 | 4 |
| 5 | 3 | 4 | 6 | 2 | 5 |
| 6 | 3 | 4 | 6 | 4 | 3 |
| 7 | 3 | 4 | 7 | 3 | 5 |
| 9 | 3 | 4 | 8 | 5 | 6 |
| 10 | 3 | 4 | 7 | 6 | 6 |

Then the clusters become:
Cluster1 = {(3,4)(2,6)(3,8)(4,7)}
Cluster2 = {(7,4)(6,2)(6,4)(7,3)(8,5)(7,6)}

 

Since the points (2,6) (3,8) and (4,7) are closer to c1 hence they form one cluster whilst remaining points for another cluster. So the total cost involved is 20.

$$\begin{aligned}\text{total cost} &= \{\text{cost}((3,4),(2,6)) + \text{cost}((3,4),(3,8)) + \text{cost}((3,4),(4,7))\} \\ &+ \{\text{cost}((7,4),(6,2)) + \text{cost}((7,4),(6,4)) + \text{cost}((7,4),(7,3)) \\ &+ \text{cost}((7,4),(8,5)) + \text{cost}((7,4),(7,6))\} \\ &= (3+4+4) + (3+1+1+2+2) \\ &= 20\end{aligned}$$

2) Step 2

Perform step 1 for another point then find optimal cost solution

## *B. Forecasting Using Regression Analysis*

In statistics, linear regression is an approach to modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X.

If the goal is forecasting then regression can be used to fit a predictive model to an observed data set of y and X values.

In linear regression the relationship between dependent and independent variable is assumed as:

y= a+ by ………………………….(I)

1) y – Dependent variable
2) a - y-intercept
3) b -Slope
4) x- Independent variable
5) a= $\bar{y}$-b$\bar{x}$
6) b=(Σxy - n$\bar{x}$ $\bar{y}$) / (Σx² - n $\bar{x}^2$)
7) x is x value of each data point
8) y is y value of each data point
9) $\bar{y}$ is arithmetic mean of all y values
10) $\bar{x}$ is arithmetic mean of all x values
11) n is number of data points

For forecasting the y value (dependent variable) we need to put the different value of x (independent variable) in eq. no (I).

| Cost (distance) to c2 | | | | | |
|---|---|---|---|---|---|
| i | c2 | | Data objects (Xi) | | Cost (distance) |
| 1 | 7 | 4 | 2 | 6 | 7 |
| 3 | 7 | 4 | 3 | 8 | 8 |
| 4 | 7 | 4 | 4 | 7 | 6 |
| 5 | 7 | 4 | 6 | 2 | 3 |
| 6 | 7 | 4 | 6 | 4 | 1 |
| 7 | 7 | 4 | 7 | 3 | 1 |
| 9 | 7 | 4 | 8 | 5 | 2 |
| 10 | 7 | 4 | 7 | 6 | 2 |

Here in our dataset Time is independent variable and rest ph, rain etc. are dependent variable.

## IV. RESULT

These are the following results of our proposed work.

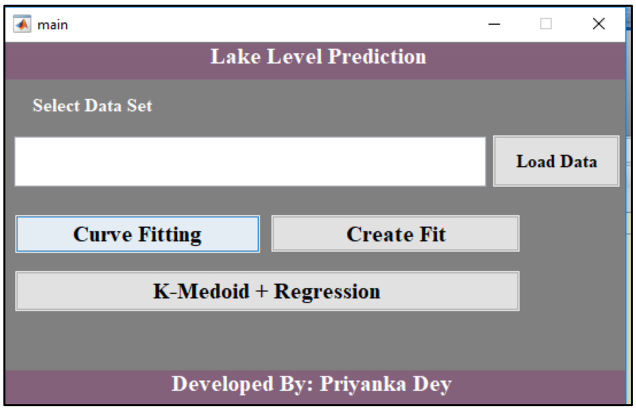


Fig. 2: GUI of Proposed Work

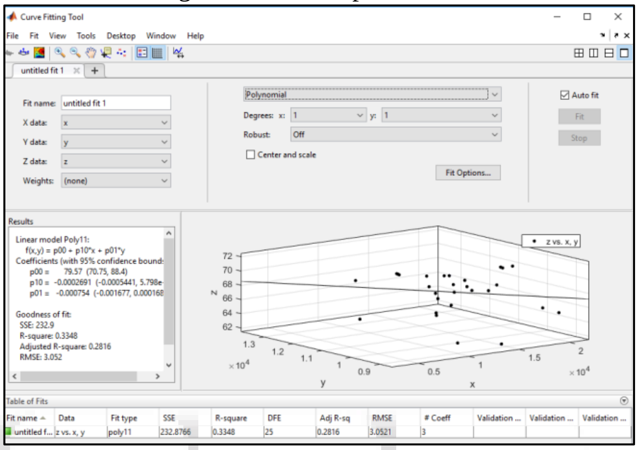

Fig. 3: Curve Fitting Tool Output

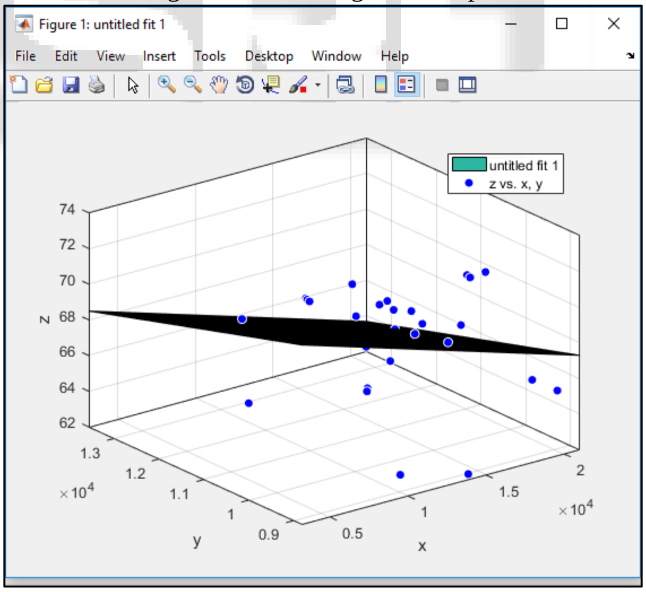


Fig. 4: Curve Fitting

 

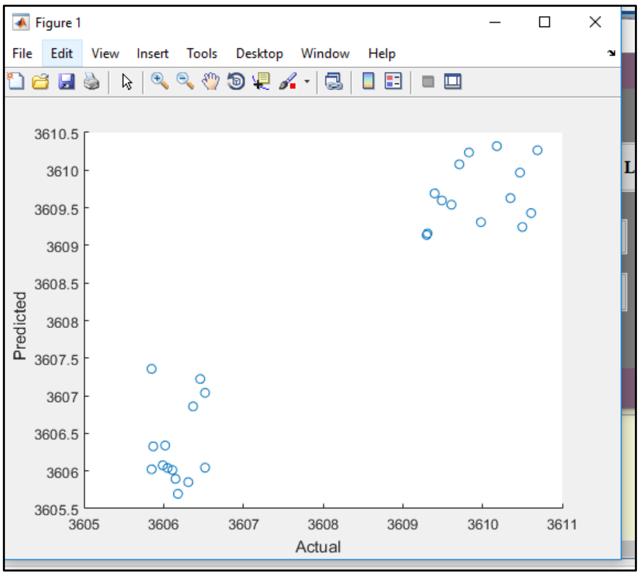


Fig. 5: Prediction

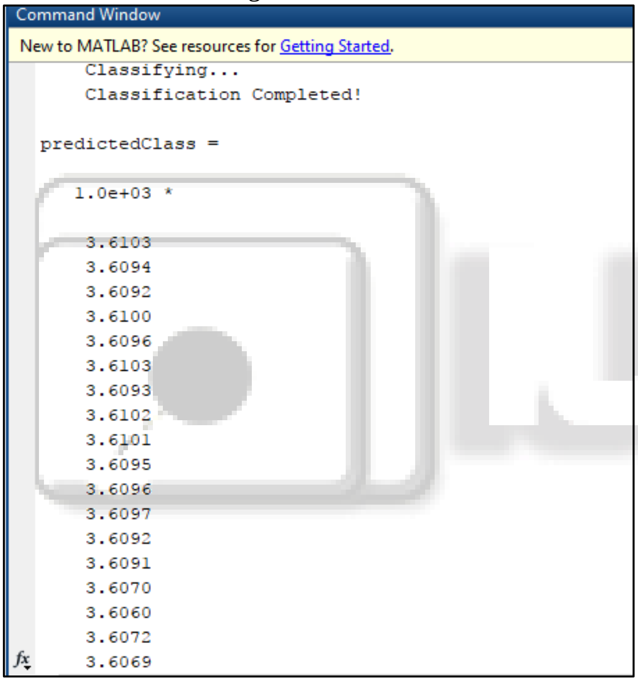


Fig. 6: Command Prediction Value

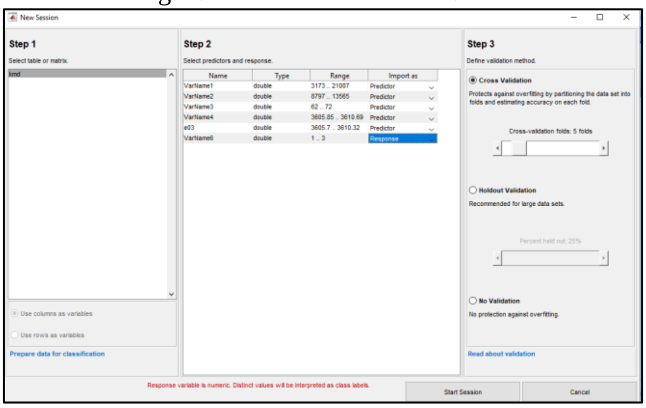

Fig. 7: Variables Value In Step 2 (1)

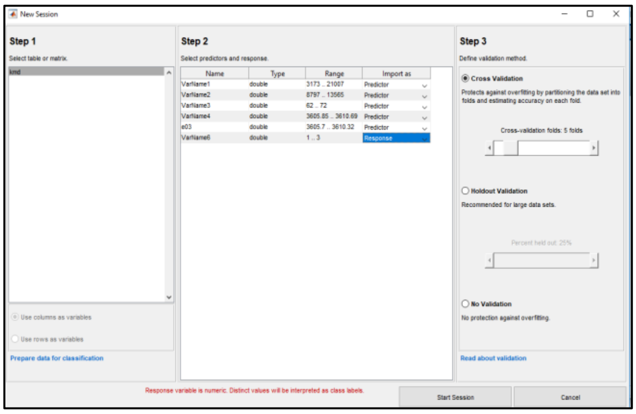

Fig. 8: Variables Value In Step 2 (2)

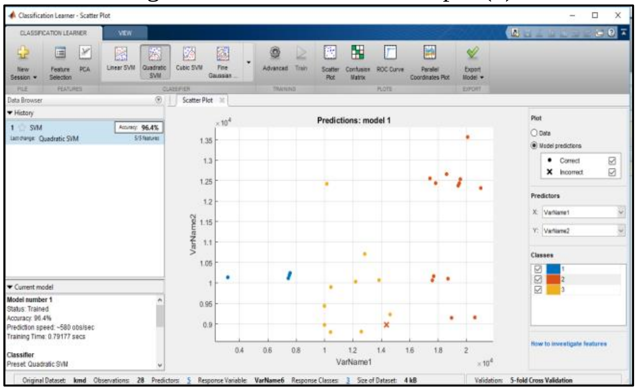


Fig. 9: Final Prediction Output (1)

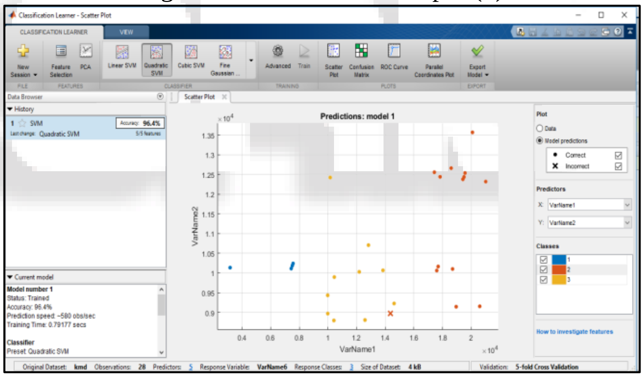


Fig. 10: Final Prediction Output (2)

## V. CONCLUSION

This proposed model is successfully designed and the prediction of the lake is done.

## REFERENCES

[1] Abeyou Wale (2008). Hydrological water balance of Lake Tana, Upper Blue Nile, Ethiopia.

[2] Unpublished M.Sc. thesis, International Institute for Geo-Information science Earth observation, Enschede, The Netherlands.

[3] Alemayehu HaileMicael and Solomon Raju, A.J. (2014). Towards sustainable management of Ethiopia's Lake Chamo Biodiversity resource: A Geospatial supported approach. Water Reservoirs in Ethiopia: Ecology and Their Benefits. In: proceeding 6th Annual conference of the Ethiopian Fisheries and Aquatic Science Association (EFASA), Mekelle, Editors: Brook Lemma and Seyoum

Mengistou, 81–108 pp. AAU printing press, Addis Ababa, Ethiopia.
[4] Alemayehu Mengistu (2006). Country Pasture/Forage Resource Profiles, Ethiopia, FAO. Alesheikh, A. A., Ghorbanali, A., and Nouri, N. (2007). Coastline change detection using remote sensing. Int. J. Environ. Sci. Tech. 4: 61–66.
[5] Ashenafi Burka (2008). Land Use /Land Cover Dynamics in Prosopis juliflora invaded the area of Metehara and the Surrounding Districts Using Remote Sensing & GIS Techniques. Unpublished M.Sc. thesis, Addis Ababa University, Ethiopia.
[6] Assefa Melesse, Weng, Q., Thenkabail, P. S. and Gabriel Senay (2007). Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling. Sensors 7: 3209–3241.
[7] Assefa Mitike (2014). Fish Production, Consumption, and Management in Ethiopia. Research Journal of Agriculture and Environmental Management 3(9): 460–466.
[8] Aksoy, H., Unal, N. E., Eris, E., and Yuce, M. I. (2013). Stochastic modeling of Lake Van water level time series with jumps and multiple trends. Hydrology and Earth System Sciences, 17: 2297–2303.
[9] Azeb Belete (2009). Climate change impact on Lake Abaya Water Level. Unpublished M.Sc. thesis, Addis Ababa University, Ethiopia.
[10] Berihu Abadi (2007). Brittle Fracture and Lake Level Change at Beseka: Main Ethiopian Rift, Metehara Area. Unpublished M.Sc. thesis, Addis Ababa University, Ethiopia.