Eurasian Journal of Analytical Chemistry
ISSN: 1306-3057 OPEN ACCESS 2018 13 (SP): 413-420

Prophecy Oxygen Levels at High Altitudes Using Fuzzy C-means Clustering

Dr.V. Vijayasamundeeswari, Dr.V. Savithri

Received 08 November 2018 ▪ Revised: 30 November 2018 ▪ Accepted: 06 December 2018

Abstract: Data Analysis is an efficient technique to implement the analysis on an immense data, which can be structured, semi-structured or even unstructured. Various advanced data analysis techniques from statistics, artificial intelligence and others can be used for analyzing data in the areas of medical diagnosis, user pattern extraction, image extraction, market research, cell segmentation and spatial data extraction. Predictive analysis is one of the methods in data analysis used to identify the predictions on future happenings which are not known in advance exactly. The key benefits of predictive analysis are preventing risks, managing resources, and strategic decision making. This paper focuses on the benefit of preventing risks factors by analyzing environment at high altitude areas through predicting the oxygen levels using Fuzzy C Means algorithm. The percentage of oxygen level is not same at sea level as it is in hilly areas which have high altitudes. Less oxygen in atmosphere may lead to short of breath causes chronic illness to people in all age groups. This study focuses on the collection and processing of data, identifying the prediction model, the results and the improvisation in the future work.

Keywords: Data Analysis, Predictive Model, Data Set, Fuzzy C-Means, Environmental Analysis.

INTRODUCTION

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains. In today's business, data analysis is playing a role in making decisions more scientific and helping the business achieve effective operation. Before one attempts to extract useful knowledge from data, it is important to understand the overall approach. The process defines the sequence of steps (with eventual feedback loops) that should be followed to discover knowledge in data[9] The amount of data collected will go unusable, if any of the pattern or knowledge is not extracted from the data. The study focuses to discover and investigate highly affected disease based on seasonal changes by fuzzy inference mechanism. [1] Difference two versions of K-means clustering compared and proved fuzzy algorithm is better for optimal partitions.[10] This study reveals with GPS, rice crops growth rate predicted using clustering. Predictive analysis is one of the techniques, which is used to analyze the data and predict the future happenings. Predictive Analysis can be implemented with the following steps.

1. Defining the Research/business objective
2. Collecting the required data using the appropriate method
3. Processing the data

Dr.V. Vijayasamundeeswari, Assistant Professor, Department of Computer Science, Women's Christian College, Chennai. E-mail: vijivicky@yahoo.co.in

Dr.V. Savithri, Assistant Professor, Department of Computer Science, Women's Christian College, Chennai. E-mail: dr.savithri.v@gmail.com

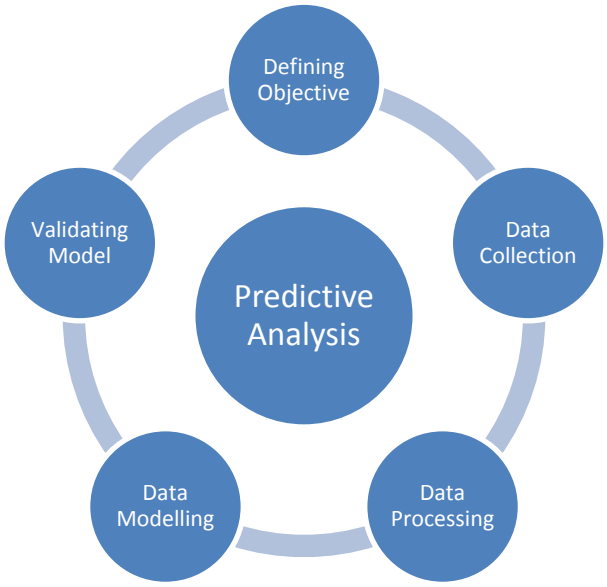

Fig. 1: Predictive Analysis

The predictive analytical model can be implemented in health care, education, governance, consumer oriented datasets to predict patterns/trends, analyze the performance, and derive the decisions and thus predicting the behavior of the model. Types of various techniques of predictive analysis are illustrated in the diagram given below.

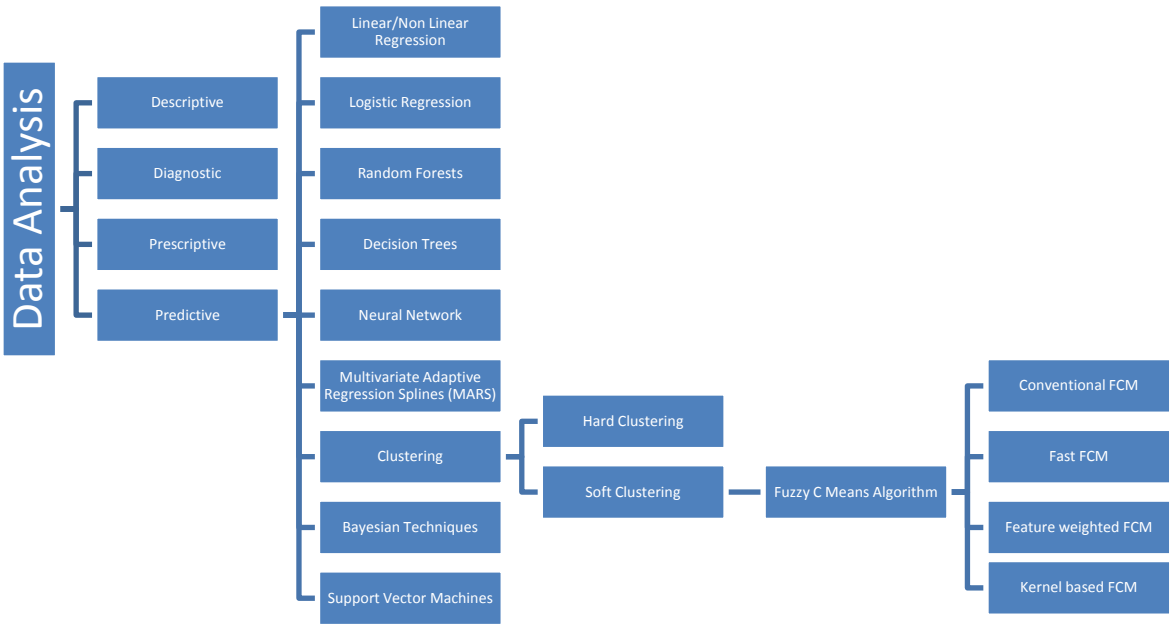

Fig. 2: Types of Data Analysis

FUZZY C-MEANS CLUSTERING ALGORITHM

Fuzzy C-Means is one of the soft clustering algorithm, which is used create 'n' clusters having the data points in each cluster will have certain degree. More precisely, a data point, which is close to the centroid, will have the high degree of membership where as a data point, which is far away from the centroid, will have the low degree of membership. The advantage of Fuzzy C-Means is the data point may be a part of more than one cluster.

In FCM clustering, the initial approximation for each cluster center is identified by calculating the mean location of each cluster. To improvise the correctness, the algorithm updates the membership grade of data points, thus moving the data center to the precise point. The iteration can be repeated with the objective of minimizing the distance between any data point to the cluster center data point. The procedural steps for Fuzzy C-Means algorithm is given below.

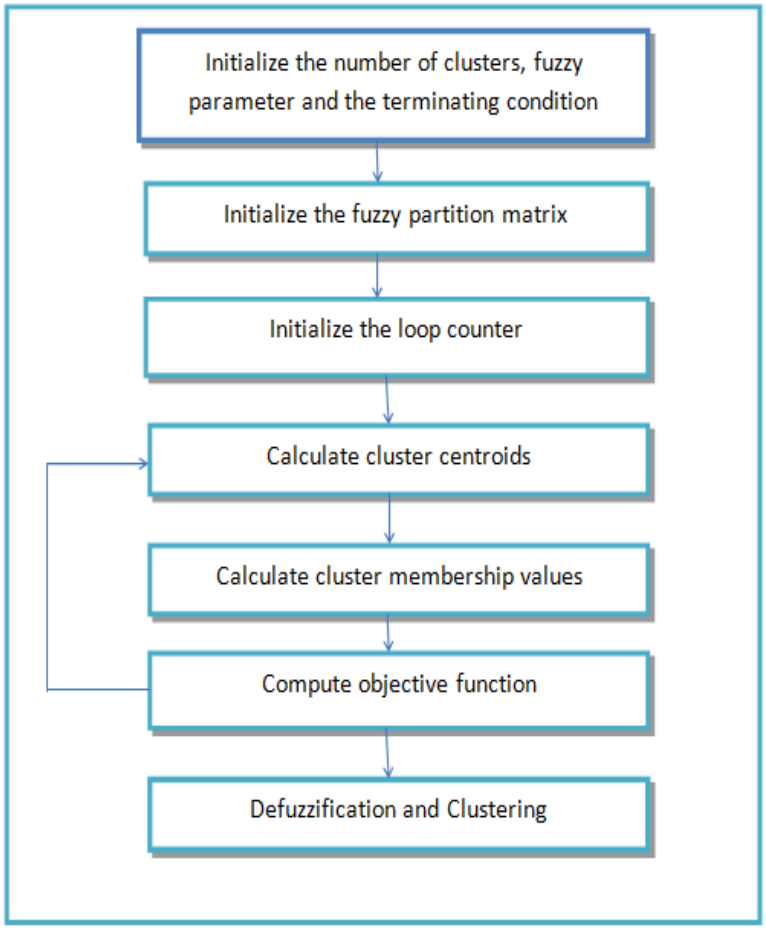

Fig. 3: Fuzzy C-means Clustering

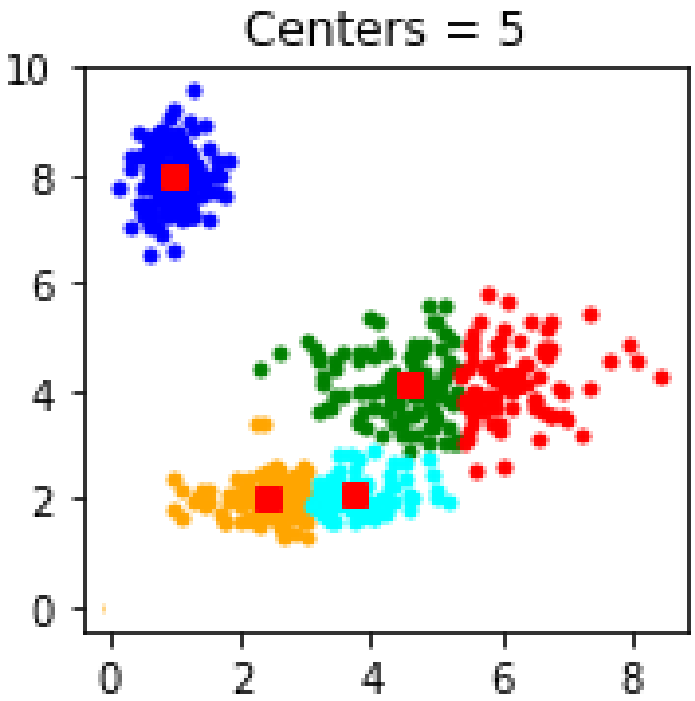

Fig. 4: Illustration of Fuzzy Clusters

The following literature survey has been done for the various data sets for the technique Fuzzy C Means to derive the predictions.

LITERATURE REVIEW

Comparative Analysis of different techniques

Table 1: Analysis

Name of the Author & Year	Data Set	Technique	Prediction
Yong Y. et al (2004)[3]	Synthetic Image Data Set	Spartially Weighted Fuzzy C-Means Clustering	To incorporate spartialneighbourhood information
GuangeliXiong. et al (2006)[4]	Cellular Image Data Set	FCM	To segment cell
M.HasanzaDeh.et al(2010)[5]	Multispectral Image Data Set	Size-Weighted Fuzzy Clustering Method.	To detect small regions that often appear in remote sensed images
Suresh, Kallam.et al(2011)[6]	Navigation Data Set	FCM	To find user access patterns
SoumiGhosh.et al (2013)[7]	Iris Plant Dataset	KNN & FCM	To analyze the efficiency
Jingfeng Yan.et al (2014)[8]	Remote Sensing Image Data Set	FCM	To improve visual interpretation of segmentation results
KapilChaturvedi. et al (2014)[1]	Clinical Data Set	FCM -ARM	To discover and investigate highly effected disease in a particular season

METHODOLOGY

This study has been carried out with the following steps in data analysis to predict oxygen levels at high altitudes.

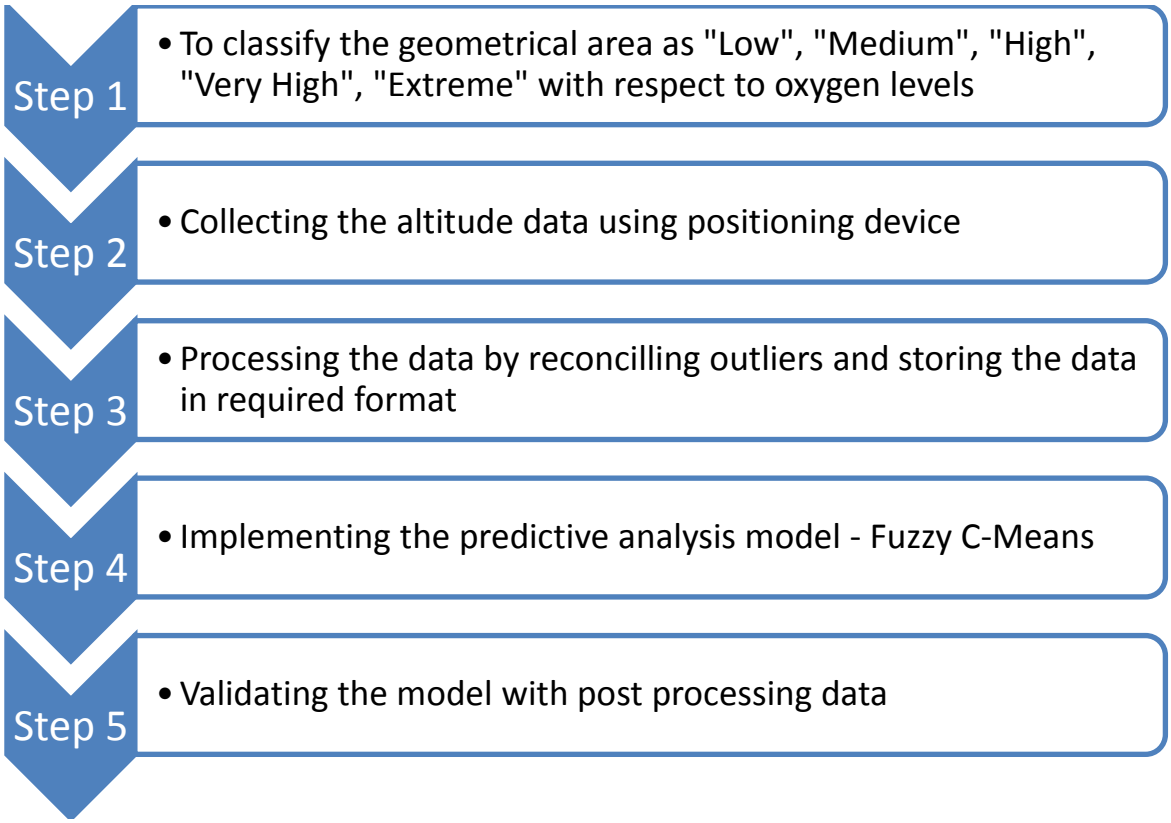

Fig. 5: Proposed Methodology

Initially, the location data were collected with the parameters Latitude, Longitude, Altitude and time stamp in the hilly area Ooty, Tamilnadu India. The collected data were in xml format with .GPX extension. The data were transformed to MySQL database using Java after processing outliers[11]. Once the data were processed, the sample data were taken for the implementation process. The data were divided into three data sets to analyze the results.

http://www.garmin.com/xmlschemas/WaypointExtension/v1
http://www8.garmin.com/xmlschemas/WaypointExtensionv1.xsd
http://www.garmin.com/xmlschemas/TrackPointExtension/v1
http://www.garmin.com/xmlschemas/TrackPointExtensionv1.xsd"><metadata><link
href="http://www.garmin.com"><text>Garmin International</text></link><time>2018-07-
20T03:00:12Z</time></metadata><trk><name>2018-07-20
08:30:03</name><extensions><gpxx:TrackExtension><gpxx:DisplayColor>Cyan</gpxx:Displ
ayColor></gpxx:TrackExtension><gpxtrkx:TrackStatsExtension><gpxtrkx:Distance>13193</
gpxtrkx:Distance><gpxtrkx:TimerTime>1633</gpxtrkx:TimerTime><gpxtrkx:TotalElapsedTi
me>1655</gpxtrkx:TotalElapsedTime><gpxtrkx:MovingTime>1631</gpxtrkx:MovingTime>
<gpxtrkx:StoppedTime>3</gpxtrkx:StoppedTime><gpxtrkx:MovingSpeed>8</gpxtrkx:Movi
ngSpeed><gpxtrkx:MaxSpeed>15</gpxtrkx:MaxSpeed><gpxtrkx:MaxElevation>1962</gpxt
rkx:MaxElevation><gpxtrkx:MinElevation>1359</gpxtrkx:MinElevation><gpxtrkx:Ascent>6
00</gpxtrkx:Ascent><gpxtrkx:Descent>1</gpxtrkx:Descent><gpxtrkx:AvgAscentRate>0</g
pxtrkx:AvgAscentRate><gpxtrkx:MaxAscentRate>1</gpxtrkx:MaxAscentRate><gpxtrkx:Avg
DescentRate>0</gpxtrkx:AvgDescentRate><gpxtrkx:MaxDescentRate>-
0</gpxtrkx:MaxDescentRate><gpxtrkx:Calories>582</gpxtrkx:Calories></gpxtrkx:TrackStat
sExtension></extensions><trkseg><trkpt lat="11.3354882039"
lon="76.8149852660"><ele>1358.87</ele><time>2018-07-
20T02:32:37Z</time></trkpt><trkpt lat="11.3354825880"
lon="76.8149607908"><ele>1358.87</ele><time>2018-07-
20T02:32:38Z</time></trkpt><trkpt lat="11.3354291115"
lon="76.8147922307"><ele>1358.95</ele><time>2018-07-
20T02:32:42Z</time></trkpt><trkpt lat="11.3353253435"
lon="76.8145634886"><ele>1359.49</ele><time>2018-07-
20T02:32:47Z</time></trkpt><trkpt lat="11.3352513313"
lon="76.8142156396"><ele>1360.66</ele><time>2018-07-20T02:32:53Z</time></trkpt><

Fig. 6: Raw positioning data in Xml

Table 2: Processed data

Timestamp	**Latitude**	**Longitude**	**Elevation**
2018-07-20 02:32:37	11.3354882039	76.8149852660	1358
2018-07-20 02:32:38	11. 3354825880	76. 8149607908	1358
2018-07-20 02:32:42	11. 3354291115	76. 8147922307	1358
2018-07-20 02:32:47	11. 3353253435	76. 8145634886	1359
2018-07-20 02:32:53	11. 3352513313	76. 8142156396	1360

RESULT ANALYSIS

The proposed Fuzzy C-means model is implemented in Spyder 3.2.4, scientific python development environment. Initially the following assumptions were made.

Table 3: Category 1

Altitude in meters	Effective oxygen %	Risk Category
0-1500	17 -21	Low
1500-1750	15-17	Medium
1751-3600	12-15	High

Further, the number of clusters increased with respect to the risk category.

Table 4: Category 1

Altitude in meters	Effective oxygen %	Risk Category
0-1500	17 -21	Low
1500-1750	15-17	Medium
1751-2500	13-15	High
2501-3600	12-13	Very High

In this implementation, 3 data sets were analyzed to measure the risk category with respect to oxygen level and high altitude data. After implementing the fuzzy C-Means clustering algorithm, the data is grouped into three clusters with the risk category low, medium, high and into four clusters with the risk category low, medium, high and very high illustrated in Table 2 and 3

Data Set 1

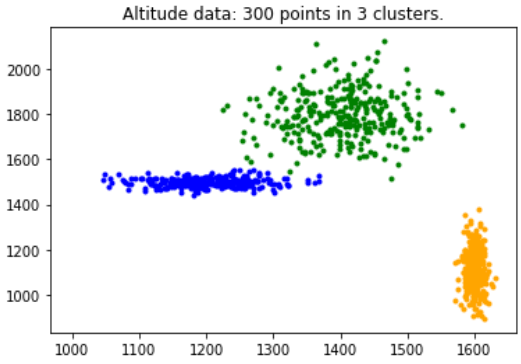

Fig. 7: Clustering analysis for Dataset1

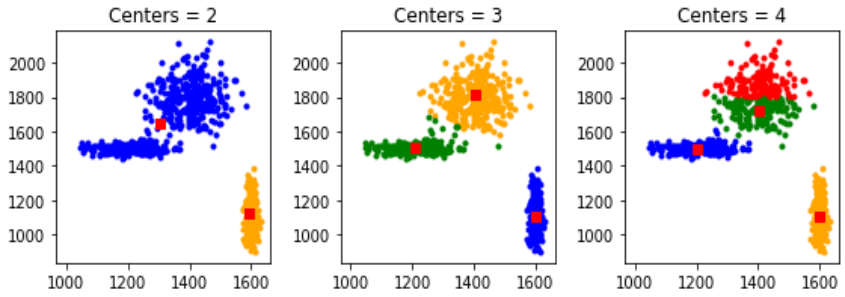

Fig. 8: Cluster Analysis with different centroid

Data Set 2

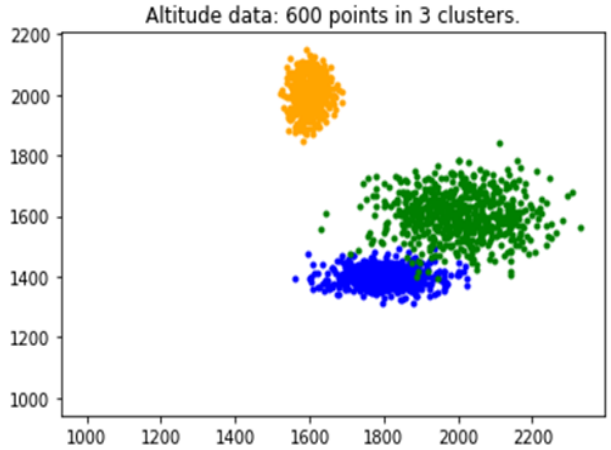

Fig. 9: Cluster Analysis for Dataset2

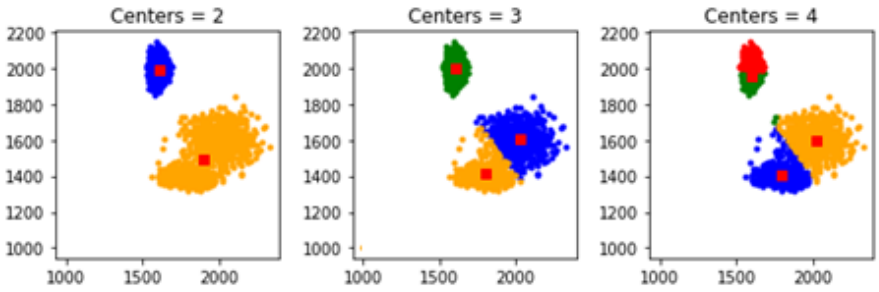

Fig. 10: Cluster Analysis with different centroid

Data Set 3

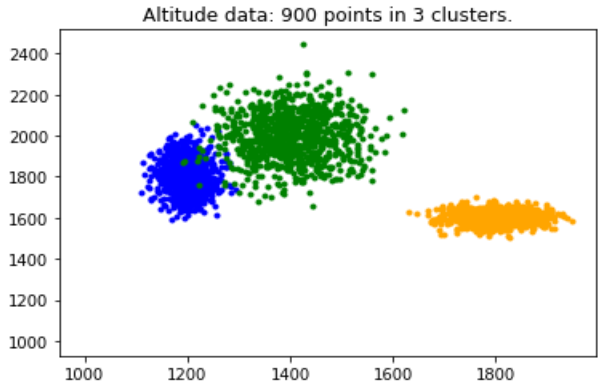

Fig. 11: Cluster Analysis for dataset3

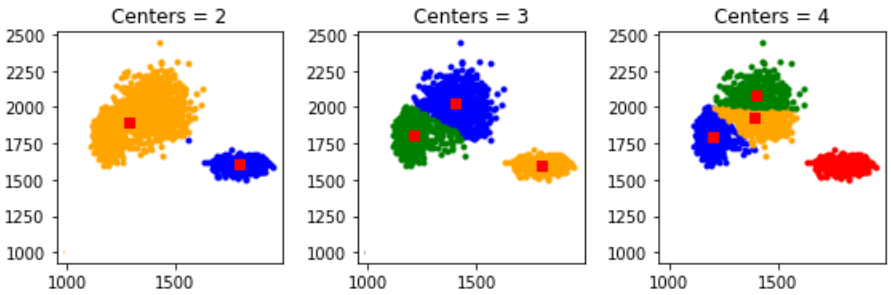

Fig. 12: Cluster Analysis with different centroid

Based on the sample data set applied on to the system the following facts were identified and represented in the table 4. According to the derived results, any geographical area can be identified with latitude, longitude, altitude, along with this prediction of oxygen level and thus the precautionary steps can be taken for the group of people in all age group with different type of disease with respect to breathless ness and so on.

Table 5: Results

Risk Category	Data set 1(Total Data:300)	Data set 2(Total Data:600)	Data set 3(Total Data:900)
Low	172	135	402
Medium	52	364	301
High	76	101	197

CONCLUSION

The data were collected in the hilly areas with the range of altitude data from 630 meters to 2350 meters. Afterwards the collected data were preprocessed and stored in well-defined format. From the stored data, three datasets were chosen for environmental analysis with the objective of segregating the geometrical area with respect to oxygen levels in those areas. After implementing the Fuzzy C-means algorithm to the data set, it is clearly identified that the number of data those falls in the specific category. Furthermore, adding few parameters like temperature, the proposed algorithm can be improved to segregate the data into groups and the apps can be developed to alert the person, who is having the breath related issues based on environmental changes.

REFERENCES

[1] KapilChaturvedi K., Patel R. and Swami D.K., —An Inference Mechanism Framework for Association Rule Mining‖ International Journal of Advanced Research in Artificial Intelligence (IJARAI), 3(9), pp.1-8, 2014.

[2] Dunn J. C., "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics 3, pp.32-57, 1973.

[3] Yong, Y., Z. Chongxun. “A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding”, 2004. Measurement Science Review 4(1): 11-19.

[4] GuangleiXiong, X. Zhou,L. ji,. Bradley.S.Wong, “Automated segmentation of Drosophila RNAi fluorescence cellullar Images using deformable models”, IEEE Transactions on circuits andsystems.vol 53. No.11, pp.2415-2424, 2006.

[5] Hasanzadeh, M. and S. Kasaei, "A multispectral image segmentation method using size-weighted fuzzy clustering and membership connectedness", 2010. IEEE Geoscience and Remote Sensing Letters 7(3): 520-524.

[6] K. Suresh, R.Madana Mohan et el., 2011, IJCSi, vol.8, issue 1, pg. 42-46.

[7] SoumiGhosh, SanjaykumarDubey et el., 2013, Comparative Analysis of K-means and fuzzy C-means algorithm, IJACSA, vol.4, issue 4.

[8] Shufen Sun, Jingfengya et el., 2008, Chinese Science Bulletin,Vol.51, Supp. 212-220.

[9] https://en.wikipedia.org/wiki/Data_analysis

[10] Savithri.V, Anuradha.G,2018,Prognosis of Lush in rice crops and nourishing inadequate exerting multiclass SVM through GPS, IJCSE, Vol6., issue 1,pg no.114-119.

[11] Vijaya Samundeeswari.V, 2018, Identifying Error Sources in Elevation Data Using GPS, IJCER, Vol8, Issue 9, pgno 18-21