

# EXTRACTION OF ANALYSED DATA SET BY USING ILLUSTRATION NOMINATOR.

*Hemangi Mirani*, V.V.P. Engineering College, Rajkot.

**Prof. Darshana Patel**, Hod(It Department), Vvp Engineering College, Rajkot

## Abstract:

Clustering is one of the most useful and popular data mining techniques. Clustering is having very wide domain of applications and almost applicable in all the fields of science, technology and business analysis applications. There are number of Clustering techniques and Clustering tools available with which data mining process applied in various activities and processes. There are some applications which requires the cluster to be implemented in graphical selectors may be for some geographical area, may be for Land usage, may be some area to be mined in selected graphical selector shape which is called illustrator nominator. Thus to effectively use these Clustering techniques and tools it is required that user must have high prerequisite knowledge about how to perform the Clustering tasks also the output of the process is required to be understandable easily so it must be in user friendly way. As most of the Clustering, tools automatically generate the cluster for the given data set with data mining patterns applied by the users and if user wants to create the cluster of their desired data then it may not be possible for them using existing tools. Thus, the research presents an idea/ model about how user can effectively use the Clustering techniques without understanding the complex algorithms and that to with visualize approach by using user-friendly SVG illustrator nominator in easy to use of pointer device inputs from user as graphical selector.

## Introduction:

**Cluster** is a group of objects that belongs to

the same class. In other words, similar objects are grouped in one **cluster** and dissimilar objects are grouped in another **cluster**.

**Association rule mining** finds interesting associations and relationships among large sets of **data** items.

**Clustering** is an unsupervised machine learning method that attempts to uncover the natural groupings and statistical distributions of data.

**Clustering technique** is one of the most admirable techniques of the **data mining** techniques. There are number of prominent tools available those provide the implementation of clustering techniques to analyze various kinds of business data. Some of the business application, scientific and engineering application required that cluster should be formed in rounded or circular shape or polygon shape. They want the objects or resources to be distributed in particular shaped way. There are number of clustering techniques available which performs the effective data mining processes but it is required that user must be having deeper understanding of such clustering techniques and in-depth knowledge of algorithm that is required to use such clustering tool. Also very few techniques are designed to perform the clustering tasks in desired shape and most of the clustering tools generate the clusters automatically once user provides the required data. Thus, user may not be able to generate the cluster in rounded or circular shape. Due to all these above, define

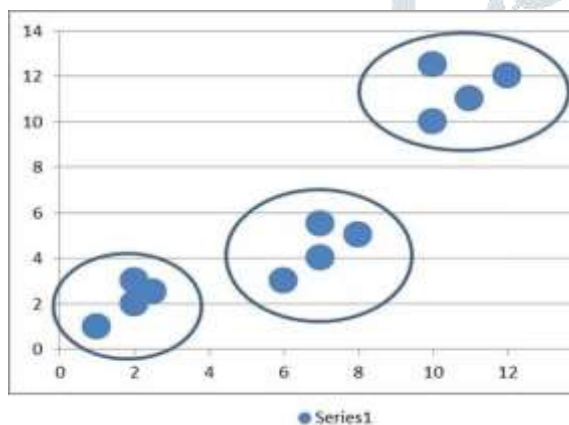
**BSON (Binary Script Object Notation)** It is a binary form for representing simple or complex data in unstructured format without any normalization applied on data including associative array format (also known as name-value pairs), integer indexed arrays, and a suite of fundamental scalar types. Generally, it used in huMONGOus Data (Big Data) with mongoDB.

**Illustrator Nominator** is a combination of two words Illustrator + Nominator also known as graphical selector. User who draws or creates shapes as per his/her requirement to select the

region called Illustrator. Region generates on geo mapping by illustrator to finalize the cluster on map called Nominator.

**Problems in current scenario and real time solution or working of research,** end-user may not be able to use the clustering technique efficiently. Thus study is carried out in this paper focuses on how to provide the effective clustering model that provides the utmost flexibility to end-user to create the cluster according to his/her requirement and convenience in shape as per their requirement by using illustrator nominator. User can applied different data mining patterns on BSON data set with huge number of BSON object and extract analysed data set as per the requirement and used this kind of analysed data set for forecasting related to business development, marketing strategies, employees core out comes, future impact of different government policies etc.

Figure a': Clusters derived by using clustering process.



## Related Work

### Literature Review

Eli Packer [1] provided the approach in which User can select among several possible visual shapes for the clusters as well as Clustering is conducted fully automatically with immediate visual responses and also user is provided with a meaningful setting from which to start the clustering exploration. Patrick Oesterling [2] proposed the split the visual analysis of high-dimensional clustering into two separated phases. In which they previously introduced topology-/ density based analysis to identify the global clustering, but present this knowledge in a novel landscape profile from which structure can be extracted more precisely and with less user interaction efforts. CHANG Dongxia

[3] presents a clustering algorithm based on dynamic niching with data attraction for clustering problems with an unknown number of clusters. The algorithm automatically finds the optimal number of clusters as well as the cluster centers. Since the number of clusters is not known a priori in most practical cases, this clustering algorithm is very useful. Witold Pedrycz [4] introduced a certain knowledge guided scheme of fuzzy clustering in which domain knowledge is represented in the form of so-called viewpoints. Viewpoints capture a way in which the user introduces his/her point of view at the data by identifying some representatives, which, being treated as externally introduced prototypes, have to be included in the clustering process. Oliver Ru'bel [5] present an integrated visualization and analysis framework that supports user-guided data clustering to aid exploration of these new complex data sets. The interplay of data visualization and clustering-based data classification leads to improved visualization and enables a more detailed analysis than previously possible. Robert S. H. Istepanian [6] proposed a Fractal-based Cluster Hierarchy Optimization (FCHO) algorithm that integrates the cluster similarity with the cluster shape and the cluster distribution to construct cluster hierarchy tree from the disjoint initial clusters. Guoping Lin [7] presented a comparative analysis of two genomic signal processing methods for robust micro-array data clustering based on Fractal Dimension and Discrete Wavelet Decomposition with Vector Quantization are validated for standard data sets.

### Review Finding

As it is observed that clusters are automatically created by the software [1,3] where once user gives input and thus not allowing user to create the cluster of his/ her desired objects. Also to use the clustering tool or techniques user must be having the in-depth knowledge [2,4,6]. However visualization in terms of creating cluster helps user to utilize the technique to generate the cluster according to his / her desired objects. Generally it is very difficult to understand the know-how and techniques of clustering due to its complexity [1,5,7] for the end-user. Some application requires creating the cluster or distribution of the objects in rounded or circular shape and according to his/her desire and convenience, which is not allowed in the most of the tools and techniques. Thus, research work is



the size of graphical selector to form the cluster. (4) All the records selected by the user through graphical selector are then fetched from the database and stored in form of table or data grid like structure. At last, user can view all the records graphically selected by him / her with the Circular graphical selector in the data grid like tabular format.

**Development of the Clustering Model to extract plotted data in the form of table by taking a Circle as graphical selector.**

To provide clustering process in circular way Researcher design & develop a model to extract the plotted data in the form of some table by taking a Circle shape as graphical selector to create a cluster.

**Components of the Clustering Model to extract plotted data in the form of table by taking a Circle as graphical selector.**

**Sales Table to store sample records of customer’s sales transaction**

Researcher created a Cluster Database as sample database to store the data related to clustering activities. Here researcher has created cluster table from which records are fetched and displayed on the Plotting area. Also all the records selected by graphical selector are first compared with the field lx and ly of cluster table and then stored into the data grid. Cluster table is having following structure.

Researcher created a Cluster Database as

Cluster Table			
Customer_no	lx	ly	sales
21	560	240	10
22	80	320	15
23	160	320	20
24	240	320	25
25	320	320	35
26	400	320	10
27	480	320	15
28	560	320	20
29	80	400	25
30	160	400	35
31	240	400	10
32	320	400	15
33	400	400	20
34	480	400	25
35	560	400	35
36	80	480	10
37	160	480	15
38	240	480	20
39	320	480	25
40	400	480	35
26	400	320	10
27	480	320	15
28	560	320	20
29	80	400	25
30	160	400	35
31	240	400	10
31	240	400	10
32	320	400	15
33	400	400	20
34	480	400	25
35	560	400	35
36	80	480	10
37	160	480	15
38	240	480	20
39	320	480	25
40	400	480	35
41	480	480	10
42	560	480	15
43	80	560	20
44	160	560	25
45	240	560	35
46	320	560	10
47	400	560	15
48	480	560	20
49	560	560	25

No.	Field Name	Data Type	Description
1	Customer_No	Number	Customer Number + Primary Key
2	Lx	Number	X coordinates Value
3	Ly	Number	Y Coordinates Value
4	Sales	Number	Sales Taken Place at Lx and Ly Coordinates

sample database to store the data related to clustering activities. Here researcher has created cluster table from which records are fetched and displayed on the Plotting area. Also all the records selected by graphical selector are first compared with the field lx and ly of cluster table and then stored into the data grid.

**(a) Cluster Database:**

Cust. No.	lx	ly	sales
1	80	80	10
2	160	80	15
3	240	80	20
4	320	80	25
5	400	80	35
6	480	80	45
7	560	80	30
8	80	160	20
9	160	160	25
10	240	160	10
11	320	160	10
12	400	160	15
13	480	160	20
14	560	160	25
15	80	240	35
16	160	240	10
17	240	240	15
18	320	240	20
19	400	240	25
20	480	240	35

### (b) Plotting area:

Plotting area is the component of the model that used to display all the records of the cluster table on the screen with geo spatial database integration with markers shown on the live map for each records as per longitude and latitude.. Records displayed on plotting area at Specific top- left values where top is dependent on ly field value and left is dependent on lx field value of the cluster table.



Figure b, Plotting area, the component of the Clustering Model to extract plotted data in the form of table by taking a Circle as graphical selector.



Figure b', on plotting area, user can see the exact statistics also on the click event of marker.



Figure c, More than one Graphical Selector or Illustration Nominator can work on single plotted area on live map with geo spatial data integration.

**(c) Illustrator Nominator OR Graphical Selector:**

Illustrator Nominator OR Graphical Selector is the component of the model that is used to select records from the plotted area. Researcher has provided the polygon as Illustrator Nominator tool. End-user has to click on the plotting area where he/she wants to create cluster. Here Circular shape will appear once user click left mouse button and it will move in the direction of mouse pointer. Its size can be also increased by clicking right mouse button. This movement will be continued until and unless end-user double click the left mouse button. Illustrator Nominator is creates by using geometry library of google map.



Figure c, Graphical Selector or Illustration Nominator is GEO SPATIAL data set extractor that is used to extract the accurate raw material (subset or cluster data set) from HUMONGOUS DATA or BIG DATA for data mining process input to generate the accurate time series analysis and forecast in very user-friendly manner.

**(d) Data Populator**

Data Populator is the component of the model that performs the process to display records on the plotting area. When user clicks on this component, it starts the process of fetching records from the table and displays the records on the plotting area.



Figure d, Data Populator, the component of the Clustering Model to extract plotted data in the form of table from specific region selected by illustrator nominator on live map with geo spatial database integration.



Figure d, Data populator populate the Clustering Model to extract plotted data in tabular data format with all the extract information regarding any business analysis for any products from specific region selected by illustrator nominator on live map with geo spatial database integration.

**(e) Data Grid**

Data Grid is the component of the model that is responsible to store & display the records selected by graphical selector. When user completes the process of selecting records with graphical selector and when user releases the mouse, immediately the process of storing and displaying starts in the data grid and it is populated with the selected records.

Customer_no	lx	ly	sales
1	80	80	10
2	160	80	15
3	240	80	20
4	320	80	25
5	400	80	35
6	480	80	45
7	560	80	30
8	80	160	20
9	160	160	25
10	240	160	10
11	320	160	10
12	400	160	15
13	480	160	20
14	560	160	25
15	80	240	35
16	160	240	10
17	240	240	15
18	320	240	20
19	400	240	25
20	480	240	35

Figure e, Data Grid, the component of the Clustering Model to extract plotted data in the form of table by taking illustrator nominator as a graphical selector.

**(f) Illustrator Nominator Generates Cluster from Live Map with GEO Spatial database integration:**



Figure f, Illustrator Nominator with editable or drag gable characteristics used by user to select the specific record markers from the live geo map with geo spatial database integrity for cluster dataset generation.

**(g) Extraction of analyzed data set**  
 Extract analyzed data set from geo spatial database based on selected distributors or braches or sales person, or customers types or geo graphical regions from illustrator nominator for any business firm in graphical manner with time series analysis and association rule data sets.

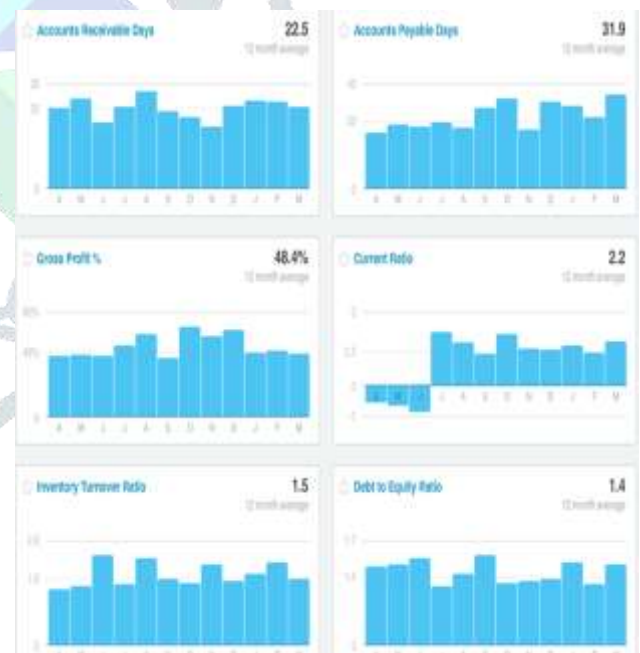


Figure g, Extract analyzed data set from geo spatial database represents in chart format with accurate extract and abstract represents for decision maker regarding all kind of business analysis.



Figure g', Exact extraction and abstraction with time series analysis graphs for the decision maker regarding the profit loss statement with sales margin summarization in graphical manner for the accurate business feasibility and risk management analysis.



Figure g'', Market share based on region and manufacturer is shown with different time series analysis so that market sentiment analysis, cross selling and upselling can be precisely identified for decision maker.

## Reference

- [1] Eli Packer, Peter Bak, Mikko Nikkil, Valentin Polishchuk, Harold J. Ship : Visual Analytics for Spatial Clustering: Using a Heuristic Approach for Guided Exploration in IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 19, NO. 12, DECEMBER 2013.
- [2] Patrick Oesterling, Christian Heine, Gunther H. Weber, Gerik Scheuermann : Visualizing nD Point Clouds as

Topological Landscape Profiles to Guide Local Data Analysis in IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 19, NO. 3, MARCH 2013.

[3] CHANG Dongxia , ZHANG Xianda : Dynamic Niching Genetic Algorithm with Data Attraction for Automatic Clustering in TSINGHUA SCIENCE AND TECHNOLOGY ISSN11007-02141107/1611pp718-724 Volume 14, Number 6, December 2009.

[4] Witold Pedrycz, Vincenzo Loia and Sabrina Senatore: Fuzzy Clustering With Viewpoints at IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 18, NO. 2, APRIL 2010.

[5] Oliver Rubel, Gunther H. Weber, Min-Yu Huang, E. Wes Bethel, Mark D. Biggin, Charles C. Fowlkes, Cris L. Luengo Hendriks, Soile V.E. Keränen, Michael B. Eisen, David W. Knowles, Jitendra Malik, Hans Hagen, and Bernd Hamann : Integrating Data Clustering and Visualization for the Analysis of 3D Gene Expression Data in IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, VOL. 7, NO. 1, JANUARY-MARCH 2010

[6] Using Cluster Similarity to Detect Natural Cluster Hierarchies by GuangHui Yan ,ZhanHuai Li in Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007) 0-7695-2874-0/07 \$25.00 © 2007 IEEE

[7] Fractal Dimension and Wavelet Decomposition for Robust Microarray Data Clustering by Robert S. H. Istepanian, Ala Sungoor, and Jean- Christophe Nebel in 978-1-4244-1815- 2/08/\$25.00 ©2008 IEEE.