Data Mining techniques in Planning Business Strategies

Abstract:-

"Role of Analytics is to learn from the past and know your future".

Business analytics is an important aspect of data mining. Any business includes simple as well as complex decision making problems. Here, businesses can use data mining for knowledge discovery and exploration of available data. This can help then to predict future trends, understand customer preferences and conduct a constructive market analysis. The paper discusses about application of one of the data mining techniques-"Cluster Analysis" with reference to banking industries. It involves identifying various distinct customer groups and clustering them with similar behavior and taste. This will help banks in designing strategy about different aspects such as retaining old customers, acquiring new improving bank services and providing people independent process.

Keywords: Data mining, K-means, Clustering.

Introduction:- Data mining is the computational process of discovering patterns in large data sets .The overall goal of data mining is to extract information from a data set and transform it into an understandable structure solving decision making problems. Data from data warehouses is collected, cleansed, consolidated, conformed and stored in one location. Because of this BI tools are able to concentrate on analyzing the data this in turn helps to design appropriate business strategies.

In today's competitive environment retaining current customers and acquiring new is a very big challenge for a Bank. So, it is important to analyze customer behavior which is based on bank databases. Here, Cluster analysis is an important data mining tool for customer segmentation .In this example objective is to suggest a method to identify customer segment leaving the bank with similar reasons. This will help the bank to develop strategies for retention of their current customers.

Clustering involves Hierarchical and Non-Hierarchical (K-means) techniques. K-means algorithm is the simplest clustering algorithm widely used for classification of data. This

algorithm has a parameter called K which indicates the number of clusters. In this method, the researcher should specify the number of clusters contrary to hierarchical clustering methods

Banks customers' classification:

Knowledge extraction from bank database includes,

I. 1st sub action – Data cleansing is the extraction of only that data considered useful for the analysis. Unnecessary data fields and records containing incomplete or missing data were removed from the data sets.
II. 1st sub action -Application of simple statistics to calculate an aggregate of new behavioral predicaters.

Behavior Scoring model

In this example, the bank is aiming at identifying the reasons for closures of accounts by different customers. We can generalize these clustering parameters as

Reason A: Service dissatisfaction

Reason B: Hidden Charges

Reason C: Salary account requirement

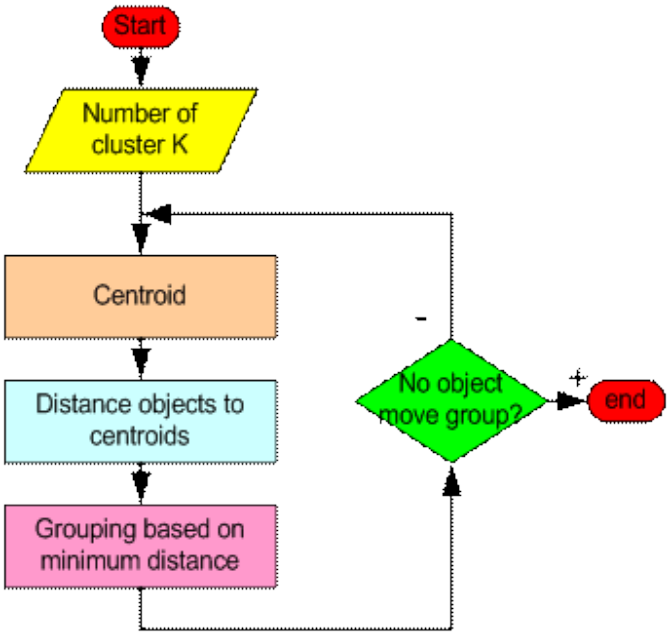
Reason D: Migration

Reason E: Others

To apply K-means methodologies will do following steps as below,

1. Determine the centroid co-ordinate
2. Define the distance function to measure distance between two customers.
3. Define threshold of the procedure(i.e iterate until stability-no object move from the group
4. Randomly select k data points as the centroids .
5. Compute distance of each data point with respect to centroid.
6. Assign cluster number of nearest centroid to that data point.
7. Re-compute centroid of the respective clusters based on new data points

8. Repeat above steps for all data points till the threshold is reached.



The parameters defined above can be converted into categorical data as follows

Service dissatisfaction	1
Hidden Charges	2
Migration	3
Others	4

Determined initial value of centroids. For Ex: 2,3,4

We can use different distance measures to find object-centroids distance such as Euclidean, Pearson, Manhattan, Squared Euclidean and squared Pearson methods. In this example we can use simple Euclidean measure given as, $\sqrt{(x1 - x2)\ 2 + (y1 - y2)2}$

This can be applied to multi-dimensional data also.

In this way continuing with K-mean algorithm , we can classify the customers in three clusters.

In these clusters, customers with similar behavior and taste are classified. Further these clusters can be analyzed to find