Aim

To clusterize the data for a bank on the basis of attributes.

SYNOPSIS

Attribute wise handling of null data

1.Age - about 35 years…its the median age group wherein approx most of the people ha e an account

2.job - self-employed

3.martial status - single.

4.education illeterate

5.default no

6.housing no

7.loan no

8.contact cellulat

9.month april (closing)

10.day friday(last day of week)

Clustering algoritnm

We would use a exclusive algorithm wherein all values of the original dataset are alloted a cluster

The no of clusters here will be 10, as defined before they are equal to the no of attributes

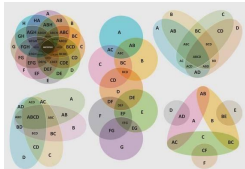Since one persons data can overlap in many attributes,

The most effective clusturing algorithm will be that of overlapping clusturing

This will allow one person to be a part of different clusters and make the clustering highly inclusive and ensure that all elements of the dataset are alloted a minimum of 1 cluster

Overlapping clustering will also make study of the database easier.

For total no of clusters

There will be minimum of 10 clusters , given that none of the elements identify with 2 attribute clusters, i.e. they are mutually exclusive

The image attached is an example of overlapping clustering.