CUSTOMER SEGREGATION IN BANKING ORGANISATION USING KNOWLEDGE MANAGEMENT

Asha. N*
School of information technology, VIT University, Vellore.
Email: nasha@vit.ac.in

Received on 12-07-2016
Accepted on 28-08-2016

Abstract

Banking institutions are backbone of any country’s economy. The financial institutions are a means of providing finance to the needy. Recent technologies have helped the organizations to grow globally. Through all this development the one thing that hasn’t changed is customer segregation. Knowledge Management Systems help an organization cope with turnover, rapid change and downsizing by making the expertise of the organization’s human capital widely accessible. Data mining techniques can also be widely used to segregate different services from bank. We apply KNN and decision tree algorithms and attempt to analyze bank customer’s data set and reduce the amount of data required to make better decision efficiently. The result is to segregate customer based on particular criteria to make the better decision to offer a better service.

Keywords: Knowledge Management, data mining segregate customers, KNN, decision tree, banking.

Introduction

Banks deals with number of customer, bank has to maintain and update customer record for future reference, and banks also provide number of facilities to the customer not to all the customers only to those who full fill terms and conditions. Before data mining it’s hard to retrieve customer detail who follow or satisfy banks conditions. Because it’s hard to retrieve some information from big data. For this we use data mining concept.

Datamining

Datamining is the technique of extracting useful or meaning full data from the big data set. For extraction useful data from big data set firstly we have to examine the nature of data set i.e. what
Columns are available? What we have to find? For this we have to make pattern or we have to find relation between attributes. There are number of algorithms are present, we have to apply some of algorithms as we think that give suitable results.

Basic steps involved in Data Mining:

1. **Segmentation**: In this steps data is grouped i.e. different columns are group together to get efficient result.

2. **Clustering**: It is basically grouping of similar data.

3. **Prediction**: In this step we predict the future, based on past data set

4. **Estimation**: By applying algorithm on data set we get an estimate value i.e. up to which extent our algorithm gives result

**Rstudio**

To implement our idea we use R-language because this language has different inbuilt functions which are helpful to us to get efficient results.

**Algorithm**

We are using **k-Nearest Neighbors algorithm** (or **k-NN** for short), k-NN is instance based learning or lazy learning. k-NN works on distance, for this we have to normalize our data set in numeric form, after normalization our data set is in 0,1 form, now next task is to jumble the data, by jumbling whole rows are jumbled because we are not applying algorithm to whole big data, we apply it on 60%,70% of big data i.e. test data. Now we make confusion matrix to see which amount of data is accurate.

**Knn graph:**

1. **Reading Dataset**

   ![Knn plotting](image)
2. Convert whole data set in integer form

3. Normalization

4. Extract Test Data From Normalize Data Set

Now the normalized data is divided into two types 1. test data 2. train data

Test data contains 60% of jumbled data
Train data contains 40% of jumbled data.

5. Apply Knn On Different Values Of K(1, 3, 5, 7, 9.....)

Now we apply KNN algorithm on the test data to get the results

By applying algorithm we can get four types of results they are

<table>
<thead>
<tr>
<th>True positive</th>
<th>True negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>False positive</td>
<td>False negative</td>
</tr>
</tbody>
</table>

This above matrix is called confusion matrix

We get four types of results in which two results are favorable

We calculate accuracy by using this formula

\[
\text{ACCURACY} = \frac{T.P + T.N}{T.P + T.N + F.P + F.N}
\]

We will apply this formula at different k-values like 1, 3, 5, 7..... Because even numbers may give 50-50 results .we will try till we get good result i.e. above 89%

Conclusion

We apply k-nn to banking data set, maximum accuracy of k-nn is 89.908,it means our algorithm is satisfied because 89 % good result no algorithm gives result more than 90-95%. So with the help of k-nn we easily able to retrieve appropriate results i.e. we have to find how much number of percentage of customer are satisfy the term and condition for loan. By this we can say that k-nn is best algorithm used for banking data set.

Future work

In future we try to implement by combining some algorithms like K-NN, decision tree, and Bayes algorithm and try to get efficient results than this

References

4. Petra Hunziker, Andreas Maier, Alex Nippe, Markus Tresch, Douglas Weers, and Peter.

5. Zemp, Data Mining at a major bank: Lessons from a large marketing application retrieved 5th January, 2006.


**Corresponding Author:**

**Asha. N**,  
**Email:** nasha@vit.ac.in