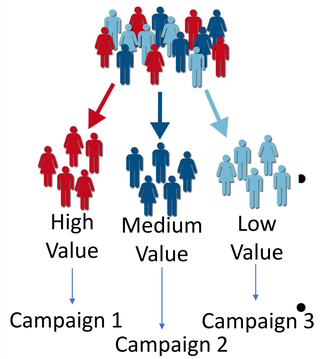
Customer Analytics in Retail Supermarket Chain

Akanksha Sharma & Namrita Gupta

Introduction



- Retail industry is customer centric so Customer Relationship Management (CRM) is crucial for its growth.
- Customer segmentation is one of the techniques to improve customer relations.
- Machine learning clustering technique helps in customer profiling.

Background

Insights from previous studies:

- Significant features affecting customers purchase behaviour was selected using Structural Equation Model (SEM) model.
- Integrated clustering technique included (Self Organized Map) SOM and K-means for segmenting customers.
- Evaluation of Clusters based on following matrices:

Fit indices		Recommended criteria	Model
Absolute fit indices	Chi-square/df	⩽5.00 (Hayduk, 1987)	1.892
	GFI	⩾0.90 (Jöreskog & Sörbom, 1993)	0.871
	RMSR	⩽0.08 (Jöreskog & Sörbom, 1993)	0.049
	RMSEA	0.05~0.08 (Jöreskog & Sörbom, 1993)	0.068

The integrated model iterated the results till it fulfilled the above mentioned criteria

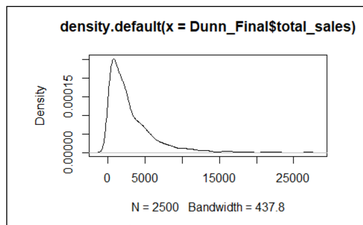
Approach

Data Importing and Cleaning → Modeling techniques → Evaluation and Results

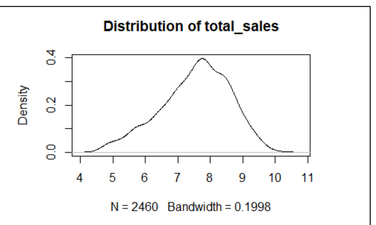
Data Preparation

Merging of Transactional data (2.5 Million) with product data (90K) to get product details.

Computation of additional features like weekend visit, recency, product brand and data rollup at house hold (HH) level resulting into 2500 distinct HH.

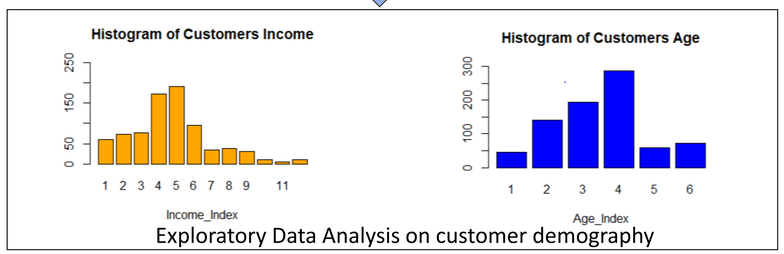


Before Data Treatment



After Data Treatment

EDA highlighted skewness in data which was treated by removing outliers and inaccurate values (<1%)



Exploratory Data Analysis on customer demography

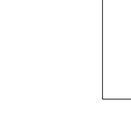
Final merging of the demographic data with the above treated data

Modeling & Evaluation

Attribute1	Attribute2	Attribute3	Attribute4	Attribute5	Attribute6
8059	6373	9249	7047	7974	7623
3397	8704	1396	4443	1660	5454
3733	5245	2561	9889	6336	4742
3060	9907	3980	6237	1960	2057
9700	1061	8047	1652	3003	6249
1125	9087	9886	8992	9297	4474
4137	9484	5766	7873	1666	1445
2079	5371	7924	4465	9635	7852

Feature selection using linear regression classifier in Weka

Customer Segments based on customer purchase behaviour using K-means and Expectation Maximization algorithms



Classification via clusters to predict sales using 1R&Naïve Bayes classifiers

Class to cluster evaluation method using “Customer spend” as class

Preliminary Results

- Clusters formed in initial stages gave performance accuracy of ~50% using class to cluster evaluation.
- Clusters predicted the customer spend giving the following accuracies across 1R and Naïve Bayes