customer segmentation application model based on data mining technology, CRM

I need customer segmentation with in store data (DEMOGRAPHIC SEGMENTATION)

I need geographical market segmentation and demographic to match with the customer in store and see where it is based

Then can do trade area analysis (see where other stores are)/age and nostalgia

(need to remember that new product change change vastly the clusters)

Store--> Cluster (40s with no family)

Istat --> Cluster demographic (to match) with location(to find)

Lots use RFM or similar to classify: probably don’t have it and cannot do that yet.

Others use behavioral data from questionnaire: still don’t have it.

Paper 2006: Many companies capture this information routinely. If yours does not, **you can use qualitative research to explore underlying motives** and needs propelling current purchases and use quantitative research to understand competitive strengths and vulnerabilities. **(This for market segmentation = get new customers)**

**Segment change:** need to take into account, especially when new product (ex. Gucci makes a bag appreciated by older woman, new clusters)

**Different variables for different products**

**Don’t use demographic**

Yellow = algorithms NON HO CODICE

1. INTRO: different segmentations: Demographic, Geographic, …, BEHAVIOURAL: don’t think I have the data for this, but maybe I can add an OCEAN classifier to IoT cameras, unique ID to faces to get frequency/recency.
2. IHPSO-KM (2022)
3. RFM (2021 c) don't think is applicable, need to have Recency, Frequency and Monetary value of customers (ok if have fidelity card but not with IoT data of BS4, only have data of customers but not money spent(or not always) and not frequency and recency(same customer not recognized)).  
   Can use KMeans, Fuzzy C-Means, RFM KMeans
4. RFM review (2021): RFM usually without personal data (demographic), useful to say how many do same things, *To segment the customers, the majority of studies used clustering (88%); the rest used CLV(Customer Lifetime Value), such as using RFM ranking scores [30] and calculating CLV values based on weighted RFM [35]. Grouping based on clustering has more accuracy than the CLV method [27].*
5. ALPSO -KM (2021): Same as IHPSO but older (and seems to be less efficient)
6. GPHC (2021): use customer data from questionnaire. don’t think is applicable. Also talk about SOM, GA, ABC
7. RFM DBSCAN (2020):
8. UKMEANS (2020 c): talks about XMEANS, C-FS
9. Elbow method kmeans (2018 c)
10. (2018 c) NQ-DBSCAN: mentions other dbscans (ρ-Approximate DBSCAN)
11. (2018 c) two phase clustering: *Although customer segmentation and market segmentation have been considered similarly in the literature, there are some critical differences regarding data availability for their clustering mechanisms. Market segmentation usually aims at acquiring new customers, and deals with the first step of CRM, customer acquisition, using socio-demographic data. While customer* *segmentation works at all steps of CRM using both socio demographic and transactional data.* *“We can imagine that customer cultivation and retention are more important than customer acquisition, because lack of information on new customers makes it difficult to select target customers and this will cause inefficient marketing efforts”*
12. (2018 c) cust seg using kmeans: aggl and mean shift worse
13. (2017) geomarketing: example of what I have to do
14. (2017) Cs Centroid Based-density IEEE: both similar
15. 2016 Fuzzy cmeasn ga = uses rfm
16. 2016 Clustering ensembles
17. 2016 Differentiated: uses RFM, *15%fortune 500*
18. 2015 New approach: consider customer behaviour change, two approaches based on demographic and rfm
19. 2014 ccccccccccc: based on density peaks
20. 2013 cc review on determine #k
21. 2013 Customer Segmentation Using Clustering and Data: uses kmeans
22. 2013 Fuzzy clustering for geographic
23. 2012 cc Overview on clustering methods: *DBSCAN does not deal very well with clusters of*

*different densities.*

1. 2012 customer segm arch: uses kmeans
2. 2012 c ikmeans best btwn considered (# clusters)
3. 2013 ccc initialization kmeans : see paper
4. 2011 review UTILE: Key Steps in a Needs-Based Market Segmentation Process, why segment market, demographic not enough. *The demographic variables of interest were age, gender, household size, occupation, education and level of income. Results of this study shows the demographic influence on choice of retail outlet is partial with household size, education and income having a significant effect on the choice of retail outlet selected. This study shows that some of the demographical factors like education, income and household size effect the choice of retail outlet and definitely the choice of brands also*
5. 2011 ecommerce: *We found that soft clustering produces more promising results for real-world applications than hard clustering, through which each customer can be assigned to only one cluster.*
6. 2013 c soft computing UTILE SC: Segmentability, segmentation variables, standarfization/normalization, reliability and validityChart, bar chart

   Description automatically generatedGraphical user interface

   Description automatically generated with medium confidence
7. FAPSO-ACO-K
8. 2008 c K–NM–PSO
9. 2008: uses GA + rfm and ltv
10. 2010 MMSEA
11. 2009 analysis: *However, a successful segmentation plan must produce market segments which meet the four basic criteria: “substantiality, identifiability, accessibility, and responsiveness*Steps in market segmentation
12. 2008 Kmeans + SOM: *segmentation oriented by product and the segmentation oriented by customer*
13. 2007 digital geomarketing mmh ok
14. 2005 kmeans selection: # clusters for kmeans, propose a new solution
15. 2004 cc initialization for Kmeans: CCIA
16. 2002 c review: kmeans most used
17. 1996 c nostalgia and age: can be used if found clusters with same age to design products that exploit this fact
18. 1998 customer value matrix: can be used
19. 2005 clustering alg for bank: most popular is kmeans, kmeans + hierarchical agglomerative
20. 2002 uses ANN: HK clustering method
21. 2006 customer segmentation intro

NP-hard to find an optimal segmentation

solution.

KMEANS BASED (most used based on paper #4, #39, bad with much noise)

Soft computing (also much used, overlapping, much more time consuming)

Density based (same eff)

Others

1. Used in Customer Segmentation:
   * Kmeans x
   * IHPSO -KM (with PSO, free init problem), need # clusters newer x
   * ALPSO-KM (with PSO, free init problem), need # clusters, older x
   * Kmeans with RFM (much used)
   * FAPSO-ACO-K (with RFM)
   * Kmeans + hierarchical agglomerative (detect outliers)
   * NN
   * Fuzzy c-means
   * Fuzzy c-means + GA (rfm)
   * GA
   * Kmeans + SOfM (seen in more than one)
   * SOM (often used as data preparation)
   * DBSCAN
   * MMSEA
   * HK ANN clustering
   * GPHC (need questionnaire)
   * Agglomerative, mean shift
2. Improvements (not used for CS):
   * Kmeans with elbow (optimal # clusters) cc
   * Kmeans with other optimal #clusters (#25) c
   * Kmeans with initialization (#26) ccl
   * Kmeans with # clusters (#37) cc
   * CCIA: initialization for kmeans c
   * XMEANS (automatic # clusters) ccc
   * UKMEANS (automatic # clusters + free initialization problem) cc
   * K–NM–PSO
   * PSO clustering
   * Xmeans ccc
   * Agglomerative fuzzy + #clusters
   * Kmeans categorical ccc
   * k means with putliers
   * k++
   * Density peaks
   * NQ-DBSCANS
   * ρ-Approximate DBSCAN
   * CONSENSUS CLUSTERING x
   * DENSITY BASED ccc
3. Used in Geographic Segmentation:
   * Direct analysis (max value)
   * Kmeans
   * Fuzzy c-means (for superimposed)

Questions:

* For iot data: Algorithms? Don’t have code for most of them
* For istat data: geomarketing (2 papers, one take majority one uses fuzzy clustering) can I use it?
* Alcuni dati per comune, altri per capoluogo, altri per regione, altri per gruppo di regioni, idee?
* Develop how? Interactive map like Sun/Javax or just printed maps or powerBI or ?

To do:

* 22/3: Leggere Small data + Trasformazione digitale, relazione algoritmi usati e perché (2 pagine)
* 29/3: Altri 2 libri, dati istat, partire da cluster demografici e vedere: 30-40 con figli con +2000mq, dove stanno? 40-50? ecc