Indian Journal of Science and Technology, Vol 9(4), DOI: 10.17485/ijst/2016/v9i4/87030, January 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Bi-Analysis Recommendation for Utilization of ICT in E-Governance Services for BSNL using Data Analytics

C . Periasamy[1*] and N. Rama[2]

[1]Department of Electronics and Communication Engineering, Vedavyasa Institute of Technology, Karadparamba - 673632, Malappuram District, Kerala, India; cperiasamy@gmail.com
[2]Department of Computer Science, Presidency College, Chennai - 600005, Tamil Nadu, India; nramabalu@gmail.com

Abstract

Background: The main objective of this research work is to develop an E-Governance model based on bi-analysis results in standard database environment with aid of SWOT analysis along with cluster analysis. **Methods:** A model based questionnaire survey was conducted in all BSNL centers of Tamil Nadu, to know the infrastructure facilities like computer, software, hardware, network structure and automation components etc. A separate questionnaire was framed to know the present status of existing problems persisting throughout the BSNL center of Tamil Nadu, India. This research covers the entire research process including: formulating research questions, performing bi-analysis through SWOT analysis and Cluster analysis and modeling best recommendation model for BSNL public sector. **Findings:** SWOT analysis based Hierarchal clustering model is drawn as one among the best of other comparable models. Also produce 96% as success rate and 4% as weaken rate will promote this as opportunities for others to avail the ICT in BSNL organization. **Applications:** This model will provide a smooth flow of information, commands, requests and reporting between the government (BSNL administration) and the citizens (BSNL staff and Public), so as to enhance the speed and quality of internal functioning, as well as to provide a user-friendly access to outsiders. Also recommends the users to avail the ICT facility in public sector with maximum of utilization.

Keywords: Cluster Model, Data Mining, E-Governance, ICT (Information and Communication Technology), Recommendation System, SWOT Analysis

1. Introduction

The process of utilizing information and communication technologies is to simplify the automation process with minimum cost and time is termed as E-Governance. This brings effectiveness and interactions among different stake holders. The main objective of this research work is to develop an E-Governance model based on bi-analysis results in standard database environment by performing SWOT analysis with cluster analysis. This model will provide a smooth flow of information, commands, requests and reporting between the government (BSNL administration) and the citizens (BSNL staff and Public), so as to enhance the speed and quality of internal functioning, as well as to provide a user-friendly access to outsiders.

Indian BSNL organization treats the concept of office automation as base layer for demonstrating how the information technology tools could lead to definite improvements in productivity, efficiency and customer satisfaction. This results in substantial reduction in use and movement of paper, as well as reduces the need for people

* Author for correspondence

movement for the information searching, avoids delay in processing and finally saves cost based on environmental conversations. These initiations are tried in Tamil Nadu to enhance the internal and external activities of the BSNL which makes this as marketable product usable by few hundreds of similar organizations in our country.

The objective of achieving E-Governance[1,2] and transforming BSNL goes far beyond mere computerization of standalone back office operations. The result of this research work shows significant benefits after the implementation of E-Governance[3,4] in BSNL center.

With the above specified motivations this research work is formulated by using Perceived Characteristics of Innovations (PCI), constructs to test a model of E-Governance adoption. Also it provides a benchmarking methodology related to the area of E-Governance[5,6] of BSNL[7,8]. This research work discusses the aspects of knowledge management system, concept based E-Governance for BSNL center and information management system.

2. Research Initiatives of Literature Review

Digital democracy is actually E-politics rather than E-Government which is leveraging the Internet to simplify the election process. It is important that, these terms not jumble the objectives of E-Government[9,10]. Also the terms data resale, digital democracy, E-politics etc. are commonly mentioned within the same pant of E-Government. These all terms complies with the principle of leveraging the Internet to simplify government.

The rise of E-Government[11] has been one of the most striking developments of the web. The evolution of Internet Digital communities tends to bring into importance certain individuals at the national and global level, which poses a number of challenges and opportunities to the governments. Firstly, the elected representatives should have unlimited access to information and communication resources. They need to elicit the aspirations of their electorate and help to translate into reality. It is necessary for them to interact and listen to the people of their constituencies and communicate with the masses.

The rise of E-Government[11] has been one of the most striking developments of the web. The evolution of Internet Digital communities tends to bring into importance certain individuals at the national and global level, which poses a number of challenges and opportunities to the governments. Firstly, the elected representatives should have unlimited access to information and communication resources. They need to elicit the aspirations of their electorate and help to translate into reality. It is necessary for them to interact and listen to the people of their constituencies and communicate with the masses.

2.1 Review on Data Analytics

Data mining is the analysis of data in novel ways to make data useful to the data owner. This results in derivation of model or patterns as an output. Some of the models are such as clustering, classification[12–14] and association rules. Data mining[15,16] typically deals with discovering knowledge from data warehouse. For this reason, data mining is often referred to as secondary data analysis. Data mining is also referred as one of the essential step in knowledge discovery process. The real world dataset consists of various instances and attributes. Due to irrelevant and redundant attributes and instances, the performance of the learners gets degraded. In order to avoid such situation, the prominent features are selected as opportunities by an SWOT analysis which provides the required and important features for the recommendation of any model based on success and opportunities.

Clustering[12,15,16] is the process of partitioning a group of data points into a small number of clusters. For instance, the items in a supermarket are clustered in categories (butter, cheese and milk are grouped in dairy products). A quantitative approach would be to measure certain features of the products, say percentage of milk and others and products with high percentage of milk would be grouped together. In general, for n data points $x_{i,I} = 1...n$ that have to be partitioned in k clusters and the goal is to assign a cluster to each data point. K-means[11,20] is a clustering method that aims to find the positions $\mu_{i,I} = 1...k$ of the clusters that minimize the distance from the data points to the cluster[11,20].

3. The Interacting Roles of Statistics and Data Mining

3.1 Distance Measures

Distance measures are used as statistical measure for cluster analysis, nearest neighbor classification methods and multidimensional scaling methods. They can be obtained directly from the objects. On the other hand, measures of similarity may be obtained indirectly from

vectors of measurements or characteristics describing each object. There are two essential ways to obtain measures of similarity.

The similarity and dissimilarity between different objects are computed using the Equation (1-3) respectively.

$$d(i,j=1-s(i,j) \quad (1)$$

$$d(i,j)=\sqrt{2(1-s(i,j))} \quad (2)$$

$$d_E(i,j)=\left(\sum_{k=1}^{p}(x_k(i)-x_k(j))^2\right)^{\frac{1}{2}} \quad (3)$$

The standard deviation for the k^{th} variable X_k can be estimated as in Equation (4) and Equation (5).

$$d_E(i,j)=\left(\sum_{k=1}^{p}(x_k(i)-x_k(j))^2\right)^{\frac{1}{2}} \quad (4)$$

$$\sigma_k=\left(\frac{1}{n}\sum_{i=1}^{}(x_k(i)-\ _k\ ^2\right)^{\frac{1}{2}} \quad (5)$$

Where μ_k is the mean for variable X_k, which (if unknown) can be estimated using the sample mean Equation (6).

$$\overline{x_k}=\frac{1}{n}\sum_{i=1}^{n}x_k(i)$$

$$x_k^r=x_k/\sigma_k r \quad (7)$$

Thus Equation (7) removes the effect of scale as captured by σ_k

In accumulation, some idea of the relative importance accorded with each variable weighted by standardization to yield the weighted Euclidean distance measure as in Equation (8).

$$d_{WE}(i,j)=\left(\sum_{k=1}^{p}(w_k(x_k(i)-x_k(j))^2\right)^{\frac{1}{2}} \quad (8)$$

3.2 Cluster Analysis

Clustering[17,18,21] plays its vital role in several exploratory pattern-analysis, decision-making for recommending and machine-learning situations by including data mining, document retrieval, image segmentation and pattern classification. All statistical models perform the task of decision maker with few assumptions about the data available[22]. With these appropriate assumptions the clustering methodology makes an assessment among the data points[19].

3.2.1 Components of a Clustering Task

Clustering mechanisms actively involves the following steps:

- Collecting the unlabelled datasets.
- Grouping similar objects together.
- Recommendation based on cluster formation .

Figure 1 depicts a first three steps of clustering activity including a feedback path.

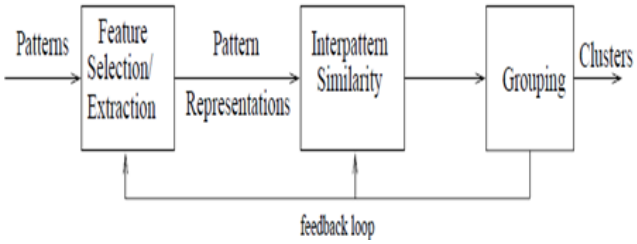

Figure 1. Clustering stages.

In the first phase all the features (facts) which pay way for more opportunities and success are identified through SWOT analysis. The result obtained in this first phase is given for the next phase in which cluster analysis is done for recommendation of ICT[23] utilization in Public sectors like BSNL. Clustering can be done in number of ways. Some of the familiar clustering models are fuzzy clustering, hierarchical clustering and so on. The hierarchical methods execute a hierarchical decomposition of the data and produce a nested series of partitions based on a criterion for merging or splitting clusters based on similarity. The model-based methods formulate a model hypothesis for each cluster and find the best fit of the data to the model. Identifying the partition which optimizes a clustering criterion can be found by partitional clustering algorithms. The partition methods build a set of partitions on the data in which each partition represents a cluster. Some other additional techniques for the clustering operations include graph-theoretic and probabilistic clustering methods. The next step in clustering is data abstraction which is the process of extracting a simple and compact representation of a data set. This is either human oriented approach or automatic analysis. The cluster prototypes include the

output of clustering algorithm, good and bad clustering result for any particular type of problem.

It is essential for the user of a clustering algorithm to not only have a thorough understanding of the particular technique being utilized, but also to know the details of the data gathering process and to have some domain expertise; the more information the user has about the data at hand, the more likely the user would be able to succeed in assessing its true class structure[24.] This domain information can also be used to improve the quality of feature extraction, similarity computation, grouping and cluster representation.

Appropriate constraints on the data source can be incorporated into a clustering procedure. One example of this is mixture resolving, wherein it is assumed that the data are drawn from a mixture of an unknown number of densities (often assumed to be multivariate Gaussian). The clustering problem here is to identify the number of mixture components and the parameters of each component. The concept of density clustering and a methodology for decomposition of feature spaces have also been incorporated into traditional clustering methodology, yielding a technique for extracting overlapping clusters.

4. Proposed Model

The research work in this paper is carried out with two analysis phase namely pre-analysis and post–analysis. The pre-analysis is done with the help of SWOT[25] analysis as preprocessing task which is further analyzed by the clustering analysis using WEKA tool. The SWOT analysis is done to bring out the decisions for the recommendation of the utility of ICT under the BSNL organizations. Initially the analysis is carried out by self-analysis. But this proves to be unworthy for all the situations. Meanwhile self-analysis is considered as the complicated way of finding the relevant requirement for the recommendation. Alternatively the SWOT is analysis is an important step towards finding life and career direction. Based on the SWOT analysis the recommended factors are further post analyzed by the data mining task such as cluster analysis. Figure 2 shows the model of the proposed recommendation system to the society for the effective utilization of ICT in the Public organization.

In this proposed model, as a pre analysis task SWOT analysis is performed[26,27]. The feature based on four factors strength, weakness, opportunities and threats are analyzed and particularly strengthen features with more opportunities are decided .Then these selected features are post analyzed by the data mining task such as cluster analysis to promote recommendation for the public to highly include the ICT utilization in the BSNL sector.

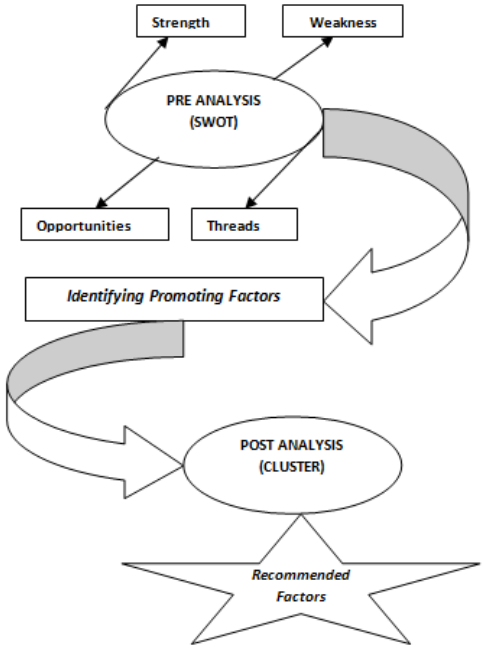

Figure 2. Proposed recommendation model.

4.1 Pre Analysis based Recommendation by SWOT Analysis

A strategic planning tool named SWOT is used in many concerns for their enhancement and better growth. An organization identifies, determines and analyses the facts as internal and external factors for developing the

ng the socio-economic [28].

Strengths, Weaknesses, ngth and Weakness are nities and Threats are ls the regions of SWOT

Figure 3. SWOT analysis.

Once we have completed our SWOT analysis, the next step is to develop recommendations for BSNL organization which could be implemented to, 1) Leverage/build upon strengths, 2) Fix weaknesses, 3) Take advantage of the opportunities and 4) Reduce or mitigate threats.

The dataset initially includes 215 reports with 112 features related to the ICT. By SWOT analysis the features on success rate is projected for post analysis by various clustering algorithms. This recommends only 24 as relevant features which are considered as best opportunities for recommending the efficient utilization on ICT in BSNL organization. The recommended features by SWOT analysis is specified in Table 1.

Table 1. Recommending factors of ICT

S. No.	Utilization Data	Description
1	SEDM	Strength- Democratization
2	SMOD	Strength-Modern Image
3	SSPP	Strength-Skilled people
4	STRA	Strength-Transparency
5	SLEG	Strength-Legacy
6	SNEW	Strength-Everything is new
7	WBUD	Weak- Budget
8	WCLN	Weak- Cyber laws not available
9	WSDM	Weak- Slow decision making process
10	WHIO	Weak- Hierarchy in organizations
11	WLAN	Weak- Different languages
12	WHET	Weak- Heterogeneous data
13	OCED	Opportunities- Competitive edge
14	OEMI	Opportunities- Employment increases
15	OESI	Opportunities- Education system improves
16	OPOI	Opportunities- Promotion of internet
17	OBUS	Opportunities- New business
18	OUOS	Opportunities- Use one standard
19	TPIR	Threats- Piracy
20	TROP	Threats- Resistance of people
21	TDID	Threats- Digital divide
22	TPRI	Threats- Privacy
23	TDOT	Threats- Dependency of technology
24	TCOR	Threats- Corruption

4.2 Intentional Post Analysis by Cluster Analysis

Clustering techniques[22,29,30] identify the similarities and dissimilarities between data and classify the data into groups with similar characteristics. Such techniques have been successfully used in market research, astronomy, psychiatry and transportation. A framework is proposed for clustering data based on fundamental recommended flow variables.

The compactness and separability are the two measures used for evaluating clustering results. The compactness is a property that expresses how much the cluster elements are close. Compactness of the cluster will be greater for the lesser variance value. Intra cluster distance is very useful for this compactness calculation. The separability evaluates the diversity of the cluster. This can be assessed by the inter cluster distance that will be the greater possible so the clusters are better.

In this paper five types of clustering techniques are investigated: 1) Cobweb clustering, 2) EM clustering, 3) Farthest clustering, 4) Hierarchal clustering and 5) k-means.

Cobweb generates hierarchical clustering, where clusters are described probabilistically. Doing this automatically through the "Classes to clusters" option does not make much sense for hierarchical clustering, because of the large number of clusters. Sometimes we need to evaluate particular clusters or levels in the clustering hierarchy.

Hierarchical clustering is a method of cluster analysis which follows to build a hierarchy of clusters. Hierarchical cluster analysis or hierarchical clustering is a general approach to cluster analysis, in which the object is to group together objects or records that are close to one another. A key component of the analysis is repeated calculation of distance measures between objects and between clusters once objects begin to be grouped into clusters. The outcome is represented graphically as a dendrogram. The dendrogram is a graphical representation of the results of hierarchical cluster analysis.

EM algorithm is also an important algorithm of data mining. We used this algorithm when we are satisfied the result of k-means methods. Expectation-Maximization (EM) algorithm is an iterative method for finding maximum likelihood or Maximum A Posteriori (MAP).This also estimates the parameters in statistical models, whereas the model depends on unobserved latent variables. The EM iteration alternates between performing an Expectation (E) step, which computes the expectation of the log likelihood which evaluated by using the current estimate for the parameters and Maximization (M) step, which is computed by parameters maximizing the expected log-likelihood found on the E step.

K-means clustering in data mining is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. This results into a

partitioning of the data space into various cells. K-means is one of the well known clustering problem and also treated as simplest unsupervised learning algorithms.

Farthest first is a variant of K means that places each cluster centre in turn at the point furthest from the existing cluster centers. This point must lie within the data area. This greatly sped up the clustering in most cases since less reassignment and adjustment is needed.

Various combinations of ICT variables were investigated for all five clustering techniques. The results indicated that the hierarchal clustering is an effective way to partition ICT data. The performance of K-means and hierarchical clustering techniques were comparable to each other and they outperformed well on comparing with other clustering methods. The results of five clustering algorithms are shown in Table 2.

Table 2. Results of cluster analysis

Cluster Model	Recommended Factors (Success Rate)	Unrecommended Factors (Weakness Rate)
Cobweb	77 %	23%
EM	78%	22%
Farthest	83%	17%
Hierarchal	96%	4%
K-Means	83 %	17%

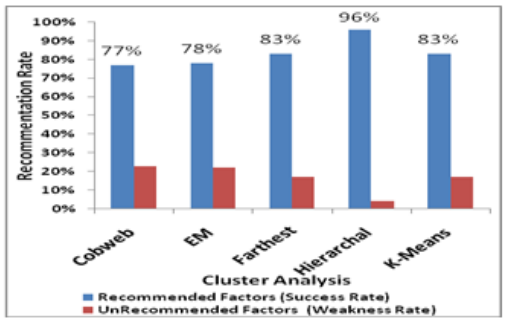

Figure 4. Comparative analysis of cluster algorithm.

The Figure 4 reveals the best clustering model for the ICT recommendation in BSNL sector.

The cluster formed by cluster algorithms is numerically measured by the Euclidean distance with maximum runs from 2 to 11. Performing SWOT as pre analysis task the more promoting factors for high utilization of ICT in BSNL is considered which is again subjected for the post analysis by various cluster algorithms. Based on the result obtained by the various cluster algorithms, the best suit cluster model with high percentage of recommendation is considered as socio-economic model for the future generations.

On performing SWOT as pre analysis task the more promoting factors for high utilization of ICT in BSNL is considered which is again subjected for the post analysis by various cluster algorithms. Based on the result obtained by the various cluster algorithms, the best suit cluster model with high percentage of recommendation is considered as socio-economic model for the future generations.

This recommendation rate of 96% also recommends the remaining 4% to utilize the benefits of ICT technology in E-Governance implementation of BSNL.

5. Conclusion and Discussion

This study indicates that not only users are benefitted in terms of ease of access to information, transparency, centralization, improved intra-and inter-departmental connectivity, removal of duplicate information, reduced transportation resulting in saving time and cost, incorporation of advanced technologies, higher accountability and customer delight but also gives suggestions for improvement upon the existing systems.

Adequate utilization of ICT technology in E-Governance implementation in BSNL is found for the motivation for the future generations. This also helps in attaining the employee job satisfaction. Implementing strategies and policies in BSNL with ICT avoids switching of more employees to other companies in the near future.

6. Recommendations

- Entrepreneurs should be more enterprising and build positively on the existing technological and information explosion in science technology and other disciplines for continual renewal of the globe or world.
- Governments should concentrate on the human capital development of their people and in general common humanity for enjoyment.
- Encouraging the establishment of a globe sovereign authority to ensure strict observance of the rule of law.

The collected data are subjected to show that majority of the respondents favors the implementation of E-Governance with significant improvements in BSNL in respect of the following:

- Expectations of the customers.

- Provision of proper and amicable public services.
- Ensuring consistent and accurate services.
- Assuring more transparency in matters.
- Contributing towards economic growth.
- Incorporation of advanced technology.
- Sharing knowledge.
- Providing services for anyone, anywhere and anytime.

The proposed E-Governance approach provide effective communication among Local, Circle and HQ so that all the possible policies are implemented in tune with all stake holders.

7. References

1. Alamelu R , Motha LCS, Amudha R. Assessing HRD climate and outpace culture in BSNL. International Conference on IEEE - Science Engineering and Management Research (ICSEMR); 2014. p. 1–6.
2. Furuholt B, Wahid F. E-Government challenges and the role of political leadership in Indonesia: The case of Sragen. Proceedings of the 41st Annual Hawaii International Conference on System Sciences – IEEE; Waikoloa, HI. 2008 Jan 7-10. p. 411.
3. Scholl HJ. Electronic government: Information management capacity, organizational capabilities and the sourcing mix. Government Information Quarterly. 2006; 23(1):73–96.
4. Xiao-Hua L. New perspective for E-Governance performance. International Conference on IEEE (ICMECG), Management of E-Commerce and E-Government; Nanchang. 2009. p. 128–32.
5. Dwivedi SK, Bharathi AK. E-Governance in India – Problems and Acceptability. Journal of Theoretical and Applied Information Technology. 2010; 17(1):37–43.
6. Bhatnagar SC, Singh N. Assessing the impact of E-Government: A study of projects in India. Information Technologies and International Development. 2010; 6(2):109–27.
7. GuiSheng C. On integrated Governance of e-Government: Technology, institution, market and Government. International Conference on E-Business and E-Government (ICEE), IEEE; Guangzhou. 2010 May 7-9. p. 4212–5.
8. Uddin G. E-Governance of Bangladesh: Present scenario, expectation, ultimate target and recommendation. International Journal of Scientific and Engineering Research. 2012 Nov; 3(11):1–20.
9. Abdul Kalam APJ. A vision of citizen-centric E-Governance for India. New Delhi: Sage Publications; 2011.
10. Compendium of E-Governance Initiatives in India. Computer Society of India. University Press.
11. United Nations E-Government survey; 2014. p. 13.
12. Salton G. Developments in automatic text retrieval. Science. 1991 Aug; 253(5023):974–80.
13. Oehler KL, Gray RM. Combining image compression and classification using vector quantization. IEEE Transactions on Pattern Analysis Machine Intelligence. 1995 May; 17(1):461–73.
14. Michalski RS, Stepp RE, Diday E. Automated construction of classifications: Conceptual clustering versus numerical taxonomy. IEEE Transactions on Pattern Analysis Machine Intelligence.1983 Jul; PAMI5(4):396–10.
15. Witten IH, Frank E. Data mining: Practical machine learning tools and techniques. San Fransisco: Morgan Kaufmann Publishers; 2005.
16. Bruha I, Famili A. Post processing in machine learning and data mining. ACM SIGKDD Explorations Newsletter. 2000; 2(2):110–4.
17. Jain AK, Murthy MN, Flynn PJ. Data clustering: A Review. ACM Computing Surveys. 1999 Sep; 31(3):264–323.
18. Xu R, Wunsch D. Survey of clustering algorithms. IEEE Transactions on Neural Networks. 2005 May; 16(3):645–78.
19. Berkhin P. Survey of clustering Data Mining Techniques. Berlin Heidelberg: Springer-Verlag; 2006. p. 25–71.
20. Ailon N, Jaiswal R, Monteleoni C. Streaming k-means approximation. In NIPS; 2009. p. 10–8.
21. Anderberg MR. Cluster analysis for applications. New York: Academic Press; 1973.
22. Jain AK, Flynn PJ. Image segmentation using clustering. In Advances in Image Understanding: A Festschrift for Azriel Rosenfeld, N. Ahuja and K. Bowyer, Editors. Piscataway, NJ: IEEE Press; 1996. p. 65–83.
23. Lee SJ, Kim JM, Lee WG. Analysis of elementary students' ICT literacy and their self-evaluation according to their residential environments. Indian Journal of Science and Technology. 2015 Jan; 8(1):81–8.
24. Jain AK, Dubes RC. Algorithms for clustering data. Prentice-Hall Advanced Reference Series. Upper Saddle River, NJ: Prentice-Hall, Inc.; 1988.
25. Gonan Bozac M, Tipuric D. Top Management's Attitudes-based SWOT Analysis. Ekonomski Pregled. 2006 Aug; 57(7-8):429–74.
26. Al-Zoubi MR. The impact of intellectual capital on SWOT analysis among Jordanian banking industry "Empirical Study". International Journal of Business and Social Science. 2013 Feb; 4(2):123–37.
27. Houben G, Lenie K, Vanhoof K. A knowledge-based SWOT-analysis system as an instrument for strategic planning in small and medium sized enterprises. Decision Support Systems. 1999 Aug; 26(1):125–35.
28. Bendigo Bank Limited. SWOT analysis for the future. 2013. Available from: www.bendigobank.com.au/public/
29. Har-Peled S, Kushal A. Smaller core sets for k-median and k-means clustering. Proceedings of the Twenty-First Annual Symposium on Computational Geometry; 2005. p. 126–34.
30. Rasmussen E. Clustering algorithms. In information retrieval: Data structures and algorithms. W. B. Frakes and R. Baeza-Yates, Editors. Upper Saddle River, NJ: Prentice-Hall, Inc.; 1992. p. 419–42.