A Two-Layer Clustering Model For Mobile Customer Analysis

Chittiboina Venkateswarlu [1], M. Padmavathamma[2]
[1]Dept of Computer Science
[2]Professor, Dept of Computer Science
[1, 2] S.V.University, Tirupati, AP, India

Abstract- *With the rise of big data and the evolution of data mining technology, the mass storage of internal enterprise information can be analyzed effectively for hidden customer value. The promotion of marketing activities and customer relationship support are also based on an extensive precision-marketing model that evolved to target the customer base and obtain in-depth understanding to fit that base's needs. An increasingly important issue is how to integrate marketing resources and properly distribute and match individual customer interests and preferences with the most effective marketing activities, as well as mine data to determine those products or services most attractive to customers.Data clustering and clustering algorithms let us group highly homogeneous individuals and assign dissimilar individuals to the appropriate segments. Industry and academia have many examples of clustering analysis being used to establish cluster characteristics*

Keywords- Customer nalysis, clustering model ,search

I. INTRODUCTION

In this study, we propose a two-layer clustering model based on the analysis of customer attributes, customer contributions, and cluster segmentation. We cluster the value of mobile customers and execute customers' characteristics on a regular basis in a systematic way. Our model can also be applied to other business areas that track consumer behavior, such as membership cards used in retail sales (as with Costco) or bank-issued credit cards. Through such cards, organizations can record and identify customer consumption and use our proposed clustering model for business analysis. Preference analysis can help a company view changes in customer value and behavior and, at any time, adjust its product strategy to retain high-quality customers. Our model provides a way for companies to plan for long-term CRM and retain high-quality customers. In addition, short-term marketers can use this modeling approach to promote products or services accurately.

II. CUSTOMER CLUSTERING

We can divide clustering algorithms into the following general categories:

- *Hierarchical*. The data points are merged or split to form the target clusters.
- *Partitional*. The number of clusters to be formed is specified in advance, and the data points are assigned iteratively to the respective clusters.
- *Density-oriented.* Clusters are formed by concatenating the data space distribution density thresholds in line with the data points.
- *Grid-oriented*. The data space is quantized into a grid structure in accordance with grid-based units for clustering.
- *Model-based*. Existing models (often statistical ones) are used to cluster data points individually.4 Cluster analysis is widely used for multivariate data analysis in fields such as medicine, economics, text mining, and commercial applications.

There have been many studies on cluster analysis for separating data characteristics and detecting data-clustering phenomena. Business applications include targeted or direct marketing based on customer grouping and clustering, customization services, good CRM, and customer behavior, attributes, and preferences.

According to the 80/20 rule (or the Pareto principle), 5 80 percent of a company's profits come from the most important 20 percent of its customers, with the remaining 20 percent of profits coming from the ordinary 80 percent of customers.

If a company can fully comprehend its key 20 percent of customers, those customers can bring a substantial profit. The related research combines the concepts of *customer lifetime value* (CLV) and customer segmentation. In this study, customers form appropriate segments, which helps the company focus on its target customers and then develop CRM, marketing strategies, and promotional activities.

CLV refers to the total revenue that each customer can bring to the enterprise. It can be divided into the customer's historical value, current value, and potential value. Academic research on customer grouping has been conducted based on CLV, and three different models have been put forward according to customer contribution, basic attributes (such as age and gender), and preferred customer behavior.6 The results show that customer groups formed through multiple dimensions can differentiate customer attributes effectively.

Dividing the majority of customer groups into several special behavioral subgroups helps a company gain an in-depth understanding of its customer base.7

However, most customer grouping to date has either been based on rules of thumb or has used only the average revenue per user (ARPU) as a benchmark for customer segmentation. Only a few approaches have included other factors, such as customer lifecycle or overall customer contribution. 8,9 Vodafone, a British telecom operator, segments mobile users into many homogeneous clusters through customer segmentation and customer profiling to identify common features.

It uses the analysis and description of user attributes to help management with decision-making and operational guidelines.10

Two-Layer Clustering Model

A mobile telecom can have as many as 10 million customers. Without customer segmentation in accordance with the business's nature, industry

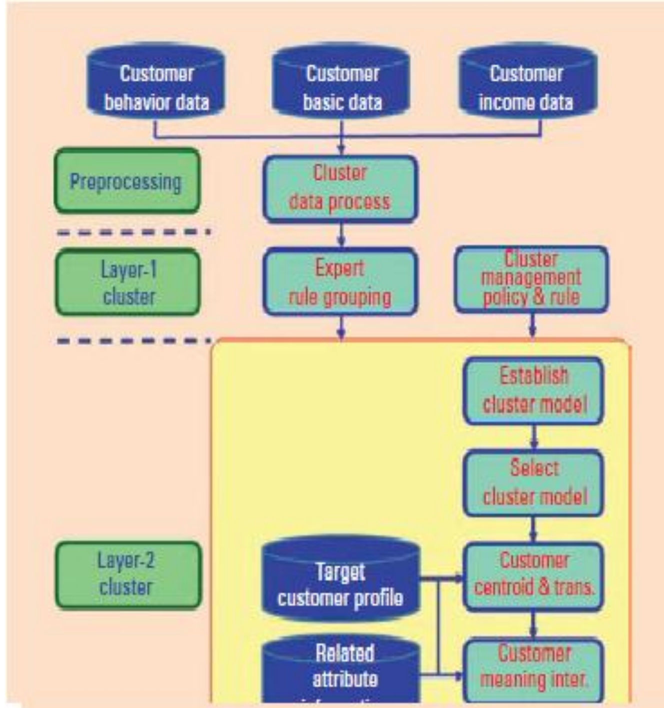

Figure 1. The two-layer clustering model. In the first layer, customers are clustered based on their contributions,

preferences, and other factors to develop a general customer relationship management strategy. In the second layer, clusters are divided into subgroups to further target customers for retention and effective marketing bile characteristics, and other factors, it is not only difficult to grasp dynamic changes in customer management, but customer retention will also have huge maintenance costs. Therefore, this study will provide telecom companies with an approach based on customers' operational needs (including voice, data, and other business) in accordance with the two-layer clustering model for customer segmentation. Figure 1 shows a flowchart of the two-layer clustering model. Our approach begins by collecting and segmenting individual customers' contributions, personal preferences, overall customer profile, and other factors.

Customer segmentation for a huge number of customers uses on the order of 105–106 clusters. These clusters are then used to develop a general strategy for CRM, which forms the first layer of clusters. After the first layer of the target has been clustered, the characteristics of the subgroups are described and interpreted by the subdivision of the second-layer clustering algorithm. The big data platform cross-analysis function maintains each group of customers and, along with effective marketing programs, forms the second layer of customer clustering analysis.

The aims of the proposed two-layer clustering model are as follows:

- provide real-time, diverse, and rich customer information through preplanned pre-analysis to strengthen the target customer base and reduce the workload of marketing staff;
- evaluate the customer segmentation strategy for each group to improve the effectiveness of activity planning and the CRM strategy; and
- use data mining technology to tap potential target customers, increase the feasibility of marketing products and services, and improve the accuracy of precision marketing.

In our proposed model, the first layer examines customer value, and the second layer uses consumer-behavior features for further grouping.

In practice, the definition of customer value varies by industry. Even the same industry can have different priorities, such as the amount of consumption, the number of consumers, the number of stores, and so on. So, when implementing this model, each industry must first define the customer value of each variable. In addition, because customer behavior can change, to make the marketing strategy more accurate, we must dynamically monitor the changes in these acts and cooperate with automatic mechanisms for observing customer-behavior trends and providing early warnings. We next describe our proposed two-layer clustering model in detail.

Layer-1 Clustering Architecture

The first-layer clustering architecture (the L1 cluster) uses the mean, and divides customers based on two 2D attributes (Figure 2): The horizontal axis divides the overall mobile customer base into the 0–99 range (that is, a total of 100 rankings) according to a customer's contribution to the company's revenue. The higher the value, the higher the customer's contribution, and the higher the customer's value to the company. The vertical axis is based on voice-leased monthly bills. The higher the voice-call monthly fee, the greater the customer's reliance on the telecom's mobile service, and the higher the demand for mobile calls. Finally, the behavior of special, data-oriented billed users (customers whose main need is for mobile Internet) becomes an independent group. In this study, we take the contribution and ARPU as the customer's grouping variables. As mentioned, in practical applications, the definition of customer value varies according to each industry. Even if the same industry may have different priorities, we can use the small-scale test to determine the first layer of the selected group variable. Mobile customers in the first layer are divided into a total of seven large groups, S1–S7. A separate overall strategy for CRM is developed for each group. For example, in accordance with the L1 cluster cutting, the overall strategy for maintaining the customer base should be driven by the performance of the S1 customer group as the main criterion.

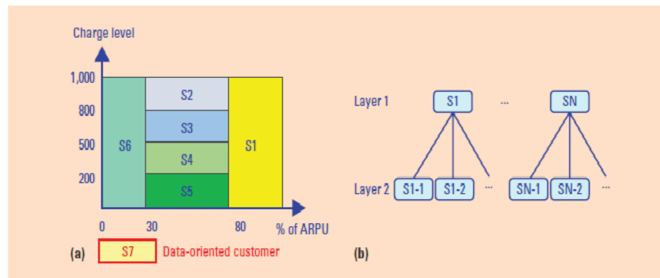

Figure 2. Schematic of layer-1 (L1) and layer-2 (L2) clustering. (a) In L1, customers are divided into groups according to 2D attributes. (b) In the second layer, they are further subdivided

Among the customer groups, S1 makes the highest contribution to the customer base. Focus should be committed to retaining this customer base and strengthening customer loyalty to stabilize the company's revenue. Meanwhile, the S6 group makes a low contribution to the customer base. The focus here should be on enhancing customer value and strengthening customers' dependence on mobile services to improve their contributions to the company's revenue. Due to the special behavior of the S7 customer base, the company should strengthen the promotion of its value-added services to drive customer demand for voice services.

Layer-2 Clustering Architecture

The second-level grouping (the L2 cluster) is structured under the L1 cluster of expert rules. Subgroups are subdivided for each L1 cluster, as Figure M2 shows. First, we aggregate for each L1 cluster by calculating customers' communication behavior to distinguish between their preferences and their business habits. The cluster variables used by the L2 cluster subdivision can vary depending on the L1 clusters the customer belongs to. Let's look at mobile-customer segmentation as an example—that is, postpaid customers in groups S1–S6. To find out the main sources of communication behavior and customer contribution, the inter- and intra network traffic minutes, the number of called objects, the ratio of each subitem to the total bill, and other variables are added.

In contrast, because the S7 group consists of customers whose main need is for mobile Internet (that is, a data-oriented type), the call behavior and use of other mobile value-added services among this group are significantly different from those of general, monthly postpaid customers. Variables assigned to the L2 cluster thus focus on the use of mobile data. For example, the proportion of downloads, the data transmission growth rate, and the time proportion of data transmission can be used to outline customers' online-behavior variables rather than voicebehavior variables.

To appropriately segment customers into the S1–S7 groups, we add where customers rank in their groupings to the group variable design. For example, international voice calls are telecom services that are often used by business customers or those who roam internationally, and are a source of mobile revenue. To differentiate the main users of international voice services into different cost bands and customer contributions, we investigate which customers are willing to use international voice roaming services and establish the difference in the dependence degree of international voice roaming groups. We summarize the number of minutes used by individual international voice customers in the overall customer segment rankings, using international phonetic variables to derive the variable type. Therefore, after the second-level grouping,

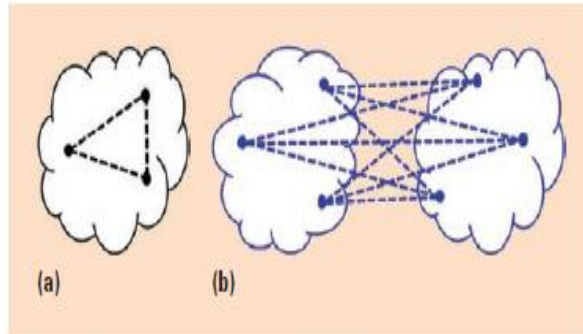

Figure 3. (a) Cohesion and (b) dispersion of groups. We use a silhouette coefficient to measure the distance between groups.

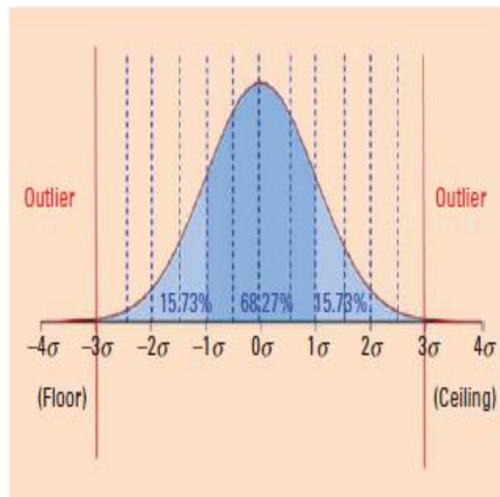

Figure 4. Class center-level distance conversion diagram. The distribution of the cluster centers of all L2 clusters is converted into five classes according to the mean value (μ) and standard deviation (σ) of the parent variables.

we can analyze the voice call behavior of relative subgroups within the overall group.

We take the following steps to clean up (preprocess) the input variables of the L2 cluster:

- *correlation coefficient analysis*—eliminate the dependency on high-value variables to avoid excessively similar weightings that result in subgroup errors; and
- *outlier processing*—replace the outliers using the ceiling or floor method to avoid bias due to extremes

that affect the clustering algorithm in determining customer attributes.

After determining the two-layer clustering model and segmenting the L1 cluster into clusteringvariables, we use SPSS Clementine 13.0 to implement L2 clustering. Clustering algorithms such as *k*-means, two-step, and Kohonen are used to establish the cluster model according to customer attributes, behavior preferences, and contract status. Each L1 cluster is split into five-to-seven different L2 clusters. The optimal clustering model is then selected as the basis of the final L2 clustering, based on conditions such as the maximum and minimum cluster ratio, the silhouette coefficient, and readability for marketers.

Clustering Analysis Results

Currently, many indicators are used to evaluate clustering results.11,12 In general, a good clustering method should match the criteria that the clusters be of similar size, with a long distance from an object to a neighboring cluster and a short distance between objects in the same cluster.

We used the following criteria to select the optimal cluster in the L2 grouping:

- *Cluster number and size*. The number of L2 clusters and the number of customers that belong to each cluster should be based on the range of activities that the company's marketers can promote. It should avoid over-segmentation of customers,
- and clusters that are either too large or too small.

- *Value of the silhouette coefficient*. The silhouette coefficient is a measure of the distance between groups, as Figure 3 shows. It is combined with cohesion and dispersion to determine the quality of the subgroup of indicators. Generally, the silhouette coefficient is greater than 0.2; if it is greater than 0.5, the cluster is more effective at distinguishing between heterogeneous or homogeneous individuals.

- *Maximum and minimum cluster ratio.* The maximum– minimum cluster ratio represents the ratio of the maximum number of clusters to the number of individuals in the smallest cluster.

A lower maximum–minimum cluster ratio means that the size of each subgroup is closer to the average, which avoids concentrating all individuals into a large group, resulting in bias. This study used a lower ratio.

- *Cluster readability*. The characteristics of a cluster should make marketing sense, be intuitive, and be aligned with the company's products or services to allow marketing staff to market directly and effectively.

We next discuss cluster interpretation and marketing policy development. To gain a deeper understanding of cluster characteristics, the cluster center (centroid) of each L2 cluster is output after grouping to provide a marketing policywe can analyze the voice call behavior of relative subgroups within the overall group. We take the following steps to clean up (preprocess) the input variables of the L2 cluster:

- *correlation coefficient analysis*—eliminate the dependency on high-value variables to avoid excessively similar weightings that result in subgroup errors; and

- *outlier processing*—replace the outliers using the ceiling or floor method to avoid bias due to extremes that affect the clustering algorithm in determining customer attributes.

After determining the two-layer clustering model and segmenting the L1 cluster into clustering variables, we use SPSS Clementine 13.0 to implement L2 clustering. Clustering algorithms such as *k*-means, two-step, and Kohonen are used to establish the cluster model according to customer attributes, behavior preferences, and contract status. Each L1 cluster is split into five-to-seven different L2 clusters. The optimal clustering model is then selected as the basis of the final L2 clustering, based on conditions such as the maximum and minimum cluster ratio, the silhouette coefficient, and readability for marketers.

Clustering Analysis Results

Currently, many indicators are used to evaluate clustering results.11,12 In general, a good clustering method should match the criteria that the clusters be of similar size, with a long distance from an object to a neighboring cluster and a short distance between objects in the same cluster. We used the following criteria to select the optimal cluster in the L2 grouping:

• *Cluster number and size*. The number of L2 clusters and the number of customers that belong to each cluster should be based on the range of activities that the company's marketers can promote.

It should avoid over-segmentation of customers, and clusters that are either too large or too small.

- *Value of the silhouette coefficient*. The silhouette coefficient is a measure of the distance between groups, as Figure 3 shows. It is combined with cohesion and dispersion to determine the quality of the subgroup of indicators. Generally, the silhouette coefficient is greater than 0.2; if it is greater than 0.5, the cluster is more effective at distinguishing between heterogeneous or homogeneous individuals.

- *Maximum and minimum cluster ratio.* The maximum– minimum cluster ratio represents the ratio of the maximum number of clusters to the number of individuals in the smallest cluster.

A lower maximum–minimum cluster ratio means that the size of each subgroup is closer to the average, which avoids concentrating all individuals into a large group, resulting in bias.

This study used a lower ratio.

- *Cluster readability*. The characteristics of a cluster should make marketing sense, be intuitive, and be aligned with the company's products or services to allow marketing staff to market directly and effectively.

We next discuss cluster interpretation and marketing policy development. To gain a deeper understanding of cluster characteristics, the cluster center (centroid) of each L2 cluster is output after grouping to provide a marketing policy

Table 1. Cluster-center transformation and description for L2 mobile clusters.

Variable name	S1-1		S1-2		S1-3		S1-4		S1-5	
Monthly charge/amount (%)	376.65	ML	427.98	ML	431.85	ML	539.35	M	599.65	M
International bill/amount (%)	1.20	M	27.35	M	**84.76**	H	1.04	M	0.80	M
Data bill/amount (%)	516.71	H	518.73	H	162.93	M	36.45	M	13.07	M
VAS bill	23.18	M	1,192.66	H	45.32	M	10.06	M	6.04	M
Ranking of amount charge	902.00	H	829.00	H	910.00	H	868.00	H	694.00	MH
Proportion of international network calls	434.37	M	446.58	M	431.87	M	438.84	M	526.51	M
Total no. of calling objects	22.12	M	29.38	MH	42.23	H	47.95	H	19.83	M
Ranking of local call minutes	580.00	M	645.00	MH	778.00	H	853.00	H	404.00	M
Ranking of international call minutes	0.00	N	64.00	M	967.00	H	0.00	N	2.00	M

H = high, M = middle, L = low, N = null

for defining the subgroups that are actually availablefor operation. The *cluster center* refers to the representative point of each L2 cluster, and can be expressed simply as the average value of each group variable. It explains the distribution of the L2 cluster variables. To facilitate the interpretation of the values, the distribution of the cluster centers of all L2 clusters is converted into five classes according to the mean value (μ) and standard deviation (s) of the parent variables (see Figure 4). This aids in comparing each L2 cluster for differences in overall customer characteristics. In addition, the integration of mobile customers' basic information (contract status, age, gender, and so on) and other cross-analysis results gives a basis for interpreting cluster characteristics. Table 1 gives an example of cluster-center transformation and a feature description for L2 mobile clusters. It lists each subgroup's cluster center and the cluster-center distance after conversion. For example, m = 7.20 for the total customer, and s = 47.14 for the global variable, which is the ratio of the international call charge to the bill amount. Therefore, the cluster center of S1-3 is 84.76 (bold in the table),

which is greater than 54.34 (that is, $m + s$). The center of the group is converted to "high" at the level at which the international telephone bill occupies the bill amount ratio according to the classification of the group center pole pitch conversion. After the group center is converted to a class distance, it can be used to describe the characteristics of each group and formulate a marketing strategy. For example, S1 is a large L1 group and can be divided into five L2 clusters, such as S1-1 to S1-5. However, there are differences in the nature of the subgroups, such as the S1-3 cluster, in which the ratio of international calls to invoices is significantly higher than for the other four subgroups. To get a more accurate target client, L2 selects the appropriate algorithm, such as 2-steps, *k*-means, or silhouette coefficient.

Moreover, with the basic cross-analysis of customer data, S1-3's use of international roaming accounted for 13.1 percent, customers aged 30–45 accounted for slightly more than all customers in this subgroup, and telephone rental for more than 10 years accounted for up to 40 percent of the subgroup, or 53.18 percent for customers with a smartphone. From this, we can conclude that the customer base is likely to be business travelers, for whom voice calls, international roaming concessions, and other related products are suitable. Looking at S1-4 of the customer base, we can see that there is no significant difference in the distribution of variables between this subgroup and the overall customer base. However, the rank transitions of the total bill amount, the total number of uttered objects, and the telephone voice talk time are displayed as "high." This shows that the S1-4 target group comprises a wide range of people, suggesting that sales or customer service could serve this group by, for example, employing large-scale use of telephone contact with this customer base. This group could thus be targeted for plans having a large number of call minutes and a discount voice package. To verify the correctness of the model, approximately 7.7 million mobile consumers were clustered in our approach. The original data was stored in a SQL server, and Clementine 13.0 was used as the implementation tool for L2 clustering. We used the correlation coefficient matrix to remove variables with high correlation coefficients, and analysis of variance (ANOVA) for cluster and input variables. With an S1-3 of about 160,000 customers, marketing staff based this subgroup's characteristics on CRM development of the corresponding policy, including recommended products containing new smartphone and international data-roaming day-rental products. In addition to retention, the plan for these customers is proactive marketing of new mobile phones and international telephone promotion rates, proactive notification to important customers about new smartphone sales, and proactive notification to key customers about international voice and data-roaming promotions.

our two-layer customer clustering model provides a macro and micro perspective to assist mobile CRM. Marketers can use preanalysis and data mining to target their customers and sell the company's products and services with accurate marketing. In addition, the expert-rule L1 subgroup can help companies to develop a general CRM direction and improve customer service. After establishing the clustering model and related strategies, we can track changes in the group structure periodically and systematically. This allows us to monitor trends in group movement, monitor size changes in each cluster, and adjust the group marketing policy and management strategy to enhance the effectiveness of the early warning mechanism.

At present, customer clustering is only included in cluster modeling through mobile voice, data usage behavior, customer contributions, and customer base data. In future work, we intend to increase the grouping of the customer-variables selection function. For different marketing or business needs, a customer-clustering model will be established to increase the flexibility of customer-clustering applications. In addition, in accordance with changing customer group structures to achieve a set threshold value, we aim to establish restart of the cluster modeling process or modify the marketing strategy of the warning mechanism to improve the dynamic feedback model grouping benefits.

REFERENCES

[1] N. Lu et al., "A Customer Churn Prediction Model in Telecom Industry Using Boosting," *IEEE Trans. Industrial Informatics*, vol. 10, no. 2, 2014, pp. 1659–1665.

[2] T. Jiang and A. Tuzhilin, "Improving Personalization Solutions through Optimal Segmentation of Customer Bases," *IEEE Trans. Knowledge and Data Eng.*, vol. 21, no. 3, 2009; doi:10.1109/ICDM.2006.87.

[3] C. Qiuru et al., "Telecom Customer Segmentation Based on Cluster Analysis," *Proc. 2012 Int'l Conf. Computer Science and Information Processing* (CSIP), 2012, pp. 1179–1182.

[4] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed., Elsevier, 2006.

[5] R.S. Duboff, "Marketing to Maximize Profitability," *J. Business Strategy*, vol. 13, no. 6, 1992, pp. 10–13.

[6] S.Y. Kim et al., "Customer Segmentation and Strategy Development Based on Customer Lifetime Value: A Case Study," *Expert Systems with Applications*, vol. 31, 2006, pp. 101–107.

[7] S.H. Han et al., "Segmentation of Telecom Customers Based on Customer Value by Decision Tree Model," *Expert Systems with Applications*, vol. 39, 2012, pp. 3964–3973.

[8] S. Masood et al., "Customer Segmentation and Analysis of a Mobile Telecommunication Company of Pakistan using Two Phase Clustering Algorithm," *Proc. Int'l Conf. Digital Information Management* (ICDIM), 2013, pp. 137–142.

[9] J. Wu and Z. Lin, "Research on Customer Segmentation Model by Clustering," *Proc. 7th Int'l Conf. Electronic Commerce*, 2005, pp. 316–318.

[10] S.M.H. Jansen, *Customer Segmentation and Custome Profiling for a Mobile Telecommunications Company Based on Usage Behavior: A Vodafone Case Study*, 17 July 2007.

[11] D.M. Eler et al., "Simplified Stress and Simplified Silhouette Coefficient to a Faster Quality Evaluation of Multidimensional Projection Techniques and Feature Spaces," *Proc. 19th Int'l Conf. Information Visualisation*, 2015, pp. 133–139.

[12] H. Wang et al., "An Approach for Improving *K*-Means Algorithm on Market Segmentation," *Proc. Int'l Conf. System Science and Eng.*, 2010, pp. 368–372.

Chiung-I Chang is an assistant professor in the Department of Information Management at the National Taichung University of Science and Technology, Taiwan. Her research interests include information management, computer networks, big data analytics, the Internet of Things, and

applications in smart cities. Chang received an ME in electrical technology from Tokyo Gakugei University, Japan. Contact her at ccy@nutc.edu.tw.

Jui-Chih Ho is an assistant professor in the Department of Information Management at the National Taichung University of Science and Technology, Taiwan. Her research interests include organizational behavior, leadership, marketing management, and human resource management. Ho received an MBA from the National Taichung University of Science and Technology. Contact her at mandy@nutc.edu.tw.

AUTHOR'S PROFILE

I.CHITTIBOINA VENKATESWARLU, received Bachelor of Computer science degree from Yogi Vemana University, Kadapa in the year of 2011-2014. Pursuing Master of Computer Applications from Sri Venkateswara University, Tirupati in the year of 2015-2018. Research interest in the field of Computer Science in the area of Cryptography-Network Security, Data Mining, Cloud Computing and Software Engineering.

Prof. M. Padmavathamma, working as a Professor in Dept. of. Computer Science, Sri Venkateswara University College of Commerce Management and Computer Science, Tirupati (AP)-India. Received Master of Computer Applications and Doctorate in Computer Science from S.V.University, Tirupati. Total number of research articles published :142, Research papers published in National Journals : 10, Research papers published in International Journals: 65.Total Number of Books Edited : 5 ,Total Number of Books Published : 4. State Best Teacher Award (Inter University Level in Engineering Subject), announced by APSCHE, Hyderabad for the year 2014.Best Researcher Award in ICT, Kuwait 2010 Outstanding Researcher Award, ICSESS, China Conducted by IEEE 2011 Best Keynote Speaker, ICDIP, Malaysia, 2012.