

Customer Cluster Model to Determine Business Opportunity by Hierarchical Method

Eneng Tita Tosida, Mulyati, Cani Nur Rahmawati,

Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Pakuan, Indonesia
enengtitatosida@unpak.ac.id; mulyati@unpak.ac.id; caninurrahmawati@gmail.com

Abdul Talib Bon

Department of Production and Operations, University Tun Hussein Onn Malaysia, Malaysia
talibon@gmail.com

Abstract

The hotel and resort business is a business that has been very strongly affected by the Covid-19 pandemic. The decline in occupancy rates encourages entrepreneurs to be able to map business opportunities and business strategies that are fast and on target. The main objective of this research is to create a customer cluster model for a hotel and resort in Sukabumi, West Java, Indonesia. The cluster model is used to map business opportunities and determine business strategies. The customer cluster model is carried out using the Hierarchical Clustering method. The attributes involved in this study include program, type of transportation, type of package, nationality, age, customer classification and visit status. The most optimal number of customer clusters is 3 clusters, with a Davies Bouldin Index value of 0.262. Based on the customer cluster model, 3 business opportunities are selected : basic service compliment with dinner or yoga, basic service compliment with hiking to village or crater, and basic service compliment with body treatment. The business strategies that have been mapped are on-line promotion, remarketing, regional marketing, incentives, and customer loyalty programs.

Keywords:

Business opportunity, Customer Cluster, Hierarchical Clustering, Business Strategy

1. Introduction

Javana Spa & Resort is one of the resorts under PT. Sarana Prima Budaya Raga is located on a secluded mountain slope, a two and a half hour drive from Jakarta situated in the cool, refreshing Cangkuang mountains, surrounded by rainforest and seven waterfalls as well as views of large and elegant bungalows with beautiful Japanese gardens. The programs offered by Havana Spa are A Day Visit, Room Only, Camping Ground, Orientation, Javana Gateway, and Grand gateway. However, since the past few years until now, the number of customers at Javana Spa has decreased. In 2015, the total number of customers who visited was 2,341 people, in 2016 there were 1556 people, in 2017 the customers who visited were 2930 people, in 2018 the customers who visited were 2090 people and in 2019 the customers who visited were 506 people. The decline is due to the fact that the resort has not been able to determine a suitable and targeted marketing strategy, resulting in a bad economy in the company (Sugraheni, 2019). Therefore, for business sustainability, the right resources and formulation of business strategies are needed (Kumar & Reinartz, 2006). One of the most effective marketing strategies is customer segmentation (Hosseini M & Shabani M. 2015). Customer segmentation is a process where consumers of a business entity are divided into several groups according to their preferences, characteristics and purchasing behavior (Qian J & Gao C, 2011). Diverse customer segmentation has different potential benefits for companies (Payne & Holt, 2001). One method for customer segmentation is the Hierarchical method. Hierarchical clustering has been widely used for segmentation purposes because of its ability to provide visual results (O. Maimon & L. Rokach, 2005).

Several studies on hierarchical methods have been carried out, including MingHsu F, et al., (2012) regarding customer segmentation based on transaction data using the Hierarchy concept. The results of his research show that the hierarchical method gives better results than other traditional methods. Anika S, et al. (2017) with research to determine the types of people who are, or who might be, visiting each mall using the k-means method and

hierarchical techniques. Shreya T, et al. (2018) regarding clustering in customer segmentation using K-means and Hierarchical clustering methods. The results showed that the K-mean method has better performance for large data and hierarchical clustering has the ability to handle small data. Other research by Rodríguez, J, at al. (2018) on Tourism Market segmentation by Unsupervised hierarchical approach, the results showed that by using hierarchical clustering travelers can divide the population into homogeneous groups and meaningful. Phan DH, et al. (2019) also did research on customer segmentation, credit card data set to determine the right marketing strategy. Based on the results obtained, the analyst can promote the right marketing strategy so that it is more profitable. Other research by Neethu CM and Anitha A (2019) on customer segmentation based on transaction data using k-mean and herarchical clustering. The results showed that the k-mean method is recommended for large data, whereas hierarchical method is suitable for small data. Idowu O, et al. (2019) conducted a customer segmentation model based RFM using k-means algorithm, Hierarchical and Fuzzi C-Means, of some of these methods was concluded that the hierarchical model of the best performing with Dunn index of 1.58. Chandrasekaran, S & Kumar, A (2019) also perform clustering for Customer Billing predictions in the Mall with a machine learning approach that is k-means and Hierarchical Clustering. This study aims to create a customer cluster model to determine business opportunities using the hierarchical method.

2. Research Methodology

The object being analyzed

The object used in this study is the Resort PT Sarana Prima Budaya Raga, namely Javana Spa & Resort. This research was conducted in order to determine the appropriate business opportunity (Tosida et al. 2020) and marketing strategy.

Steps of research

The steps taken during the study are broken down as follows:

Stage 1: Transfer of Javana Spa & Resort knowledge

Transfer of knowledge about Javana Spa & Resort performed by interviewing HRD Javana Spa & Resort to get information about the program is on offer, the price, and the implementation of marketing strategies in Javana Spa & Resort.

Stage 2: Information Data Processing

The information that has been received is processed to conceptualize customer data to build a cluster. Who later can find out the number of each containing the customer and package selected. The results obtained will be used as a business opportunity and determine the appropriate marketing strategy.

Stage 3: Building system

The system is built using the KDD (Knowledge Discovery and Data Mining) method which consists of the basic concepts of Data Cleaning and Data Integration, Data Selection and Data Transformation, Pattern Evaluation, Knowledge Presentation as for the details of this stage are :

- 1) Data Cleaning and Data Integration is the process of removing noise and inconsistent or irrelevant data. Data integration is a combination of data from several sources. The data used in this study are real data from Javana Spa & Resort.
- 2) Data Selection and Data Transformation. Often, not all of the existing data selections are used in the database, therefore only data suitable for analysis will be taken from the database. Data transformation is data that will be transformed or combined into a format suitable for processing in data mining.
- 3) Pattern Evaluation In this stage the results of data mining techniques in the form of distinctive patterns and analysis models are evaluated to assess whether the existing hypothesis has indeed been achieved. If it turns out that the results obtained do not match the hypothesis there are several alternatives that can be taken, such as making feedback to improve the data mining process, or accepting this result as an unexpected result that might be useful.
- 4) Knowledge Presentation (Knowledge Presentation). Presentation of knowledge is a visualization and presentation of knowledge about the methods used to obtain knowledge obtained by users. The last stage of the data mining process is how to formulate decisions or actions from the results of the analysis obtained. In this presentation, the results are knowledge of how close the clusters are to become a single cluster.

Stage 4: Evaluation System

The evaluations were conducted in this study are: The first evaluation is done by setting the number 3 clusters in the data as much as 6800 and conducted searches DBI (Davies Bouldin Index) on the number 3,4,5 and 6 clusters.

2.1 Hierarchical Clustering

Hierarchical Clustering is a group analysis method that seeks to build a hierarchy of data groups. There are generally 2 types of grouping strategies, namely Agglomerative (Bottom-Up) and Devisive (Top-Down). In this study, agglomerative is used because the method is simple and the process only requires 4 stages which will be repeated (Jianfu, L, et al., 2011). The steps for Agglomerative Hierarchical Clustering are:

- 1) Compute the distance between data matrices, Form a distance matrix, with Manhattan Distance:

$$D = \sum_{i=1}^n |b_i - a_i| \quad (1)$$

or using *Euclidian Distance* :

$$D(a, b) = \sqrt{\sum_{i=1}^n (b_i - a_i)^2} \quad (2)$$

- 2) Combine the two closest groups based on the parameters specified proximity

Some Agglomerative Hierarchical clustering methods:

- a. Single Linkage

$$d_{uv(w)} = \max (d_{uw}, d_{vw}) \quad (4)$$

- b. Complete Linkage

$$d_{uv(w)} = \frac{\sum_i \sum_j d_{ij}}{n_{(uv)} n_w} \quad (5)$$

- 3) Update distance between data matrix to represent closeness between the new group and the remaining groups.

- 4) Repeat steps 2 and 3 until only one group remains.

2.2 Davies Bouldin Index (DBI)

Davies Bouldin Index is one of the methods used to evaluate the cluster algorithm introduced by David L. Davies and Donald W. Bouldin in 1979 (Davies and Bouldin, 1979). Results Of the clustering process is then evaluated using Davies Bouldin index to determine the most optimal cluster number (Irhamni F, et al, 2014). Measurements with DBI maximizes the distance between the clusters and at the same time trying to minimize the distance between points in a cluster (Thakare & Mules, 2015). Overall IBD has better performance and also the complexity of the calculation time is smaller (Petrovic S, 2006) as for the steps of DBi as follows:

- 1) Compute the distribution in one class using formula:

$$S_i = \sqrt{\frac{1}{N} \sum (X_j - C_i)^2} \quad (6)$$

- 2) Finding how far the centroid class i to class j with formula:

$$M_{ij} = |C_i - C_j| \quad (7)$$

- 3) Compute the ratio of the distribution of classes to how far the centorids are between classes

$$R_j = \frac{S_i + S_j}{M_{ij}} \quad (8)$$

- 4) Compute the value of D

$$D_i = \text{Max } R_{ij}, \quad i \neq j \quad (9)$$

- 5) Compute DB Index

$$IDB = \frac{1}{N} \sum_{i=1}^n D_i \quad (10)$$

3. Result and Discussion

3.1 Results Of clustering

The results of clustering of 6800 customer data using the single linkage algorithm are presented in the form of a dendrogram which can be seen in Figure 1.

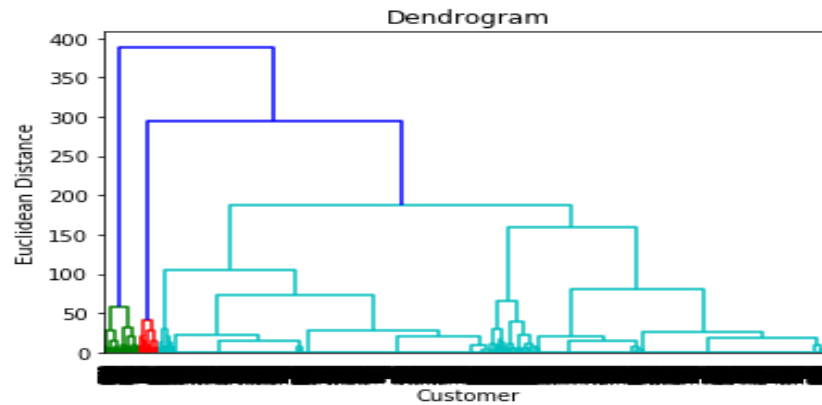


Figure 1. The clustering results in the form of dendrogram

Figure 1 is the result of customer clusters presented in a dendrogram where Cluster 0 is a customer who purchases the A day Way, A day Visit, Camping, Family Program, Meeting, Orientation, Room Only and Tent Only programs. Cluster 1 is a customer group buying orientation program. Cluster 2 is a customer who purchased the javana gateway and grand gateway program. The percentage of the number of objects in each cluster has a significant difference. About 92% of customers are in cluster zero, 5% are in cluster one, while cluster two is 3%. This percentage shows that most customers choose the one-day program.

Visualization clustering results shown by the graph as shown in Figure 2-8

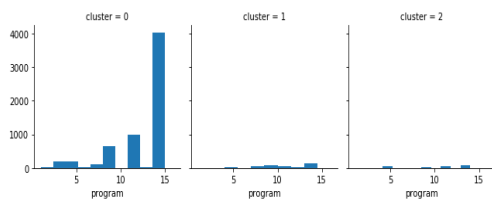


Figure 2. Visualization cluster with program attributes

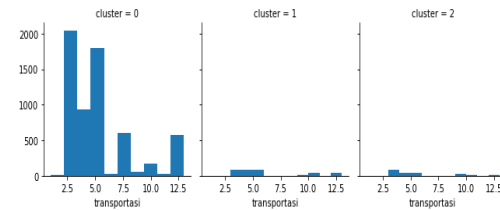


Figure 3. Visualization cluster with transport attributes

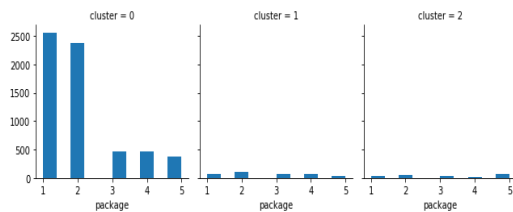


Figure 4. Cluster visualization with the package attribute

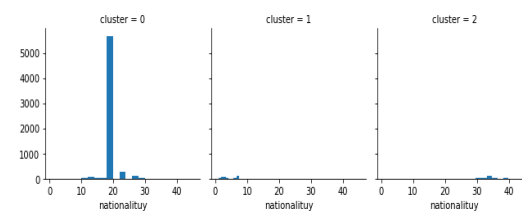


Figure 5. Visualization of clusters with nationality attributes

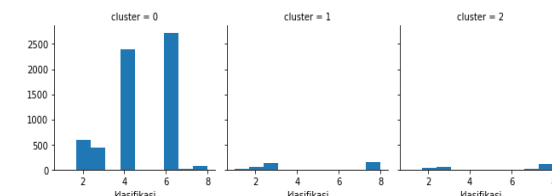
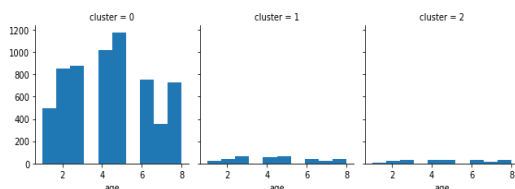


Figure 6. Cluster visualization with program age

Figure 7. Cluster visualization with classification attributes

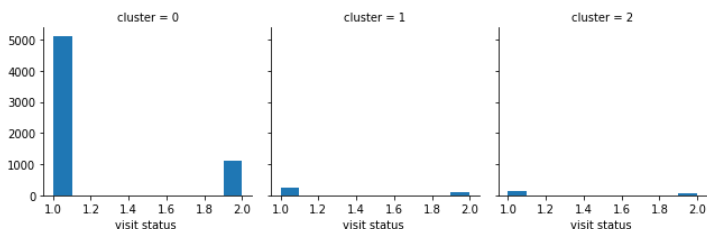


Figure 8. Visualization of clusters with visit status attributes

Figure 2-8 shows that the program chosen by customers in cluster 0 is the A day Way, A day Visit, Camping, Family Program, Meeting, Orientation, Tent Only program but the program that is very popular is the Room Only program. Most of the customers who come use by on car transportation or use personal transportation, the costs incurred to buy the program are Rp.450,000 to Rp.650,000. Customers who come are predominantly Indonesian citizens with various ages from 25th-75th but most of them are aged 43th-48th and are visiting for the first time.

Cluster 1 is a customer who chooses the Grand Gadeway and Orientation program but prefers the transportation orientation package used is by on car or private vehicle and the costs incurred to buy the program are Rp. 850,000 to Rp. 1,100,000. Customers who come are American, Arabian, Australian, aged 25th-75th but mostly 31th-36th and they are coming for the first time.

Cluster 2 is a customer who chooses the Grand gradeway and Javana Gateway program, the transportation used is by on car or private vehicle and the cost incurred to buy the program is Rp. 2,500,000 to Rp. 3,500,000. Most of the customers who come are Singaporean citizens aged 25th-75th but the dominant age is 43th-48th and they are coming for the first time. Through this attribute analysis shows conformity with the results of interviews with the hotel.

3.2 Interpretation of Knowledge

The interpretation of the resulting knowledge to determine the business opportunity is as follows:

1. C1 is a group of 6272 customers who purchase the A day Way, A day Visit, Camping, Family Program, Meeting, Room Only and Tent Only programs. These programs are 1 day programs
2. C2 is a group of 348 customers buying orientation programs. The orientation program is a 1-day stay program with a meal & Spa package.
3. C3 is a group of 180 customers who purchased the javana gateway and grand gateway program. The program is a multi-day stay program where all treatment packages can be chosen freely.
4. In determining business opportunities, there are several ways that will be implemented and improved.
 - a. For C1 the strategy used is to provide a compliment, namely light Spa treatment or dinner or yoga.
 - b. For C2, the strategy used is to provide compliement, namely hiking to vilage or crater.
 - c. For C3, the strategy used is to provide compliment body treatment.
 - d. Easy to find online
 - e. Remarketing
 - f. Market the area
 - g. Provide incentives
 - h. Build a customer loyalty program
 - i. Collaborating with SMEs around the hotel to produce souvenirs and tourism services based on local wisdom
 - j. Empowerment of the community around the hotel for the creation of tourism activities based on local wisdom and promoted digitally (Tosida et al. 2019)

3.3 Evaluation Result

The evaluations carried out in this study are:

1. The first evaluation is done by setting the number of 3 clusters in the data as much as 6800 and DBI search is carried out on the number of 3,4,5 and 6 clusters. The results show that the smallest DBI is in the 3rd cluster with a value of 0.262. The smaller the DBI value, the better the cluster evaluation results
2. The second evaluation conducted an interview with Javana Spa & Resort and the results obtained were as desired.
3. In determining a good business opportunity, we need to pay attention to where the opportunity lies from the results of research conducted that the largest group is customers who buy packages per-1 day which in practice need to be improved in the service sector, providing promo-pomo interesting, make advertisements in social media, and others so that customer interest in the resort increases and can also stay longer.

4. Conclusion

Based on the research that has been done, it is found that the Hierarchical method can be used as an option in solving customer clustering problems, namely at Javana Spa & Resort. The results of the analysis were obtained using 3 clusters and the number of customers was 6800, namely cluster 0 contains 6272 customers, cluster 1 contains 348 customers, cluster 2 contains 180 customers. Cluster 0 is the dominant customer who chooses A day Way, A day Visit, Camping, Family Program, Meeting, Room Only and Tent Only programs. With a vulnerable package of Rp. 450,000 to Rp. 1,100,000. Cluster 1 is the dominant customer who buys the orientation program, which is a 1-day stay program with dining and spa facilities with a package of Rp. 1,700,000 to Rp. 2,500,000. Cluster 2 is the customer who predominantly buys the Javana Gateway and Javana Gateway packages, namely a several-day stay program with dining and treatment facilities and can be freely chosen with a package of Rp. 1.700.000 - Rp. 3,700,000. Based on the results of the cluster, it can be used as a reference in determining business opportunities by improving marketing and service strategies and evaluating other programs so that customers are interested in staying longer at the resort in order to compete with existing resorts. The strategy used for C1 will be given a compliment, namely light Spa treatment or dinner or yoga, C2 will be given a compliment, namely hiking to vilage or crater and C3 will be given a compliment body treatment. Evaluation was carried out using the Davies Bouldin Index which obtained a value of 0.262 with 3 clusters, the smaller the DBI value, the better the results of the cluster evaluation.

References

- Anika Singh Prasanna Balaji Venkataraman. 2017. Customer Segmentation through K-Means and Hierarchical Clustering Techniques. *International Journal for Scientific Research & Development*. 5,{7}. ISSN (online): 2321-0613.
- Chandrasekaran, S. Kumar, A. (2019). A Clustering Approach for Customer Billing Prediction in Mall: A Machine Learning Mechanism. *Journal of Computer and Communications*, 7: 55-66. doi: 10.4236/jcc.2019.73006.
- Davies, D. L.; Bouldin, D. W. 1979. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2): 224-227. **Doi:** 10.1109/TPAMI.1979.4766909
- Hosseini M, Shabani M. 2015. New approach to customer segmentation based on changes in customer value. *Jurnal of Marketing Analytics* 3(3): 110-121. Doi: 10.1057/jma.2015.10
- Idowu S, Annam A, Rangaraja E, Kattukottai S, 2019. Customer Segmentation Based on RFM Model Using K-Means, Hierarchical and Fuzzy C- Means Clustering Algorithms. Doi 10.13140/RG.2.2.15379.71201.
- Irhamni F, Damayanti F, Khusnul B K, Miftachul A. 2014. Optimization of Subdistrict Grouping Based on Educational Indicators using Clustering Methods and Davies Bouldin Index. (Optimalisasi Pengelompokan Kecamatan Berdasarkan Indikator Pendidikan Menggunakan Metode Clustering Dan Davies Bouldin Index). *Prosiding semnastek 2014*. ISSN : 2407 - 1846
- Jianfu, L., Jianshuang, L., Huaiqing, H. (2011). A Simple and Accurate Approach to Hierarchical Clustering. *Journal of Computational Information Systems*, 7(7), 2577-2584.
- Jumadi D. D.S, Salim O S, Poltak S. 2018. Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm. *The 3rd International Conference on Computing and Applied Informatics*. doi:10.1088/1742-6596/1235/1/012015.
- Kumar, V., & Reinartz, W. (2006). *Customer relationship management: A database approach*. New York, NY: John Wiley.

- MingHsu F, PangLu L MinLin C. 2012 .Segmenting customers by transaction data with concept hierarchy. Elsevier, *Expert Systems with Applications*. 39(6): 6221-6228.
- Nikhil R. Pal, J. Biswas. 1997. Cluster validation using graph theoretic concepts. Elsevier. *Pattern Recognition* 30(6): 847-857.
- O. Maimon and L. Rokach, 2005. Clustering methods in Data Mining and Knowledge Discovery Handbook. Boston: Springer US : 321-352
- Payne, A., & Holt, S. (2001). Integrating the value process and relationship marketing. *British Journal of Marketing*, 12, 159–182.
- Petrovic S. 2006. A Comparison Between the Silhouette Index and the Davies-Bouldin Index in Labelling IDS Clusters. *Proceeding 11th Nordic Workshop Secure IT System*: 53-64,
- Phan Duy Hung, Nguyen Thi Thuy Lien, Nguyen Duc Ngoc. 2019. Customer Segmentation Using Hierarchical Agglomerative Clustering. *Proceedings of the 2019 2nd International Conference on Information Science and Systems*: 33–37. doi.org/10.1145/3322645.3322677.
- Qian J & Gao C. 2011. The application of Data Mining in CRM. 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC). Doi:10.1109/AIMSEC.2011.6010697.
- Rodríguez R, Semajski I Gautama S, Weghe N V D, Ochoa D. 2018. Unsupervised Hierarchical Clustering Approach for Tourism Market Segmentation Based on Crowdsourced Mobile Phone Data. *Sensors (Basel)*. 18(9): 2972.
- Shreya Tripathi, Aditya Bhardwaj, Poovamma. 2018. Approaches to Clustering in Customer Segmentation. *International Journal of Engineering & Technology*. 7 (3.12) : 802 -807
- Sugraheni. (2019). Javana Spa & Resort. <https://www.javana-Spa.com/>, 29 februari 2020.
- Thakare, Y.S. & Bagal, S.B. 2015. Performance evaluation of k-means clustering algorithm with various distance metrics. *International Journal of Computer Applications* 110(11): 12-16.
- Tosida ET, F Andria, I Wahyudin, R Widiyanto, M Ganda, RR Lathif. 2019. A Hybrid Data Mining Model for Indonesian Telematics SMEs Empowerment. *IOP Conf. Series : Material Science and Engineering* 567 (2019) 012001. DOI : 10.1088/1757-899X/567/1/012001
- Tosida ET, Wahyudi I, Andria F, Djatna T, Ningsih WK, Lestari DD. 2020. Business Intelligence of Indonesian Telematics Human Resource : Optimization of Customer and Internal Balaced Scorecards. *Journal of Southwest Jiaotong University*, Vol. 55 No. 2, Apr. 2020. DOI : 10.35741/issn.0258-2724.55.2.7.

Acknowledgements

Acknowledgments are Institute for Research and Community Service of Pakuan University, The Computer Science Department of the Faculty of Mathematics and Natural Sciences, Pakuan University, which has provided both moral and material support.

Biographies

Eneng Tita Tosida is a lecturer it the Department of Computer Sciences, Faculty of Mathematics and Natural Sciences, Universitas Pakuan. She teaches in Simulation Techniques and Data Mining, Linear Programming and Optimization Models and research methods. She leads research group of Decision Support System (DSS) and Socio Informatic, and actives on educational digital media base on game, Augmented Reality and Virtual Reality researchs. She also actives on Indonesian Operations Research Association (IORA) as Secretary. Now is serving as head of Community Services Center, Universitas Pakuan.

Cani Nur Rahmawati is a graduate Computer Science Department, Faculty of Mathematics and Natural Sciences, Universitas Pakuan. She actives also as a part time employee in a hotel and resort in Sukabumi District, Jawa Barat, Indonesia

Mulyati is a lecturer in the Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Pakuan. She teaches in mathematic, statistic, optimization and also actives on artificial intelligence research.

Abdul Talib Bon is a professor of Production and Operations Management in the Faculty of Technology Management and Business at the Universiti Tun Hussein Onn Malaysia since 1999. He has a PhD in Computer Science, which he obtained from the Universite de La Rochelle, France in the year 2008. His doctoral thesis was on topic Process Quality Improvement on Beltline Moulding Manufacturing. He studied Business Administration in the Universiti Kebangsaan Malaysia for which he was awarded the MBA in the year 1998. He's bachelor degree and diploma in Mechanical Engineering which his obtained from the Universiti Teknologi Malaysia. He received his postgraduate certificate in Mechatronics and Robotics from Carlisle, United Kingdom in 1997. He had published more 150 International Proceedings and International Journals and 8 books. He is a member of MSORSM, IIF, IEOM, IIE, INFORMS, TAM and MIM.