Data analytics and the P2P cloud: An integrated model for strategy formulation based on customer behaviour

H.Y. Lam[1], Y.P. Tsang[3], C.H. Wu[2,3], Valerie Tang[1]

ORCID: H.Y. Lam: 0000-0002-5858-1894; Y.P. Tsang: 0000-0002-6128-345X; C.H. Wu: 0000-0003-1259-4048; Valerie Tang: 0000-0002-4528-1091

[1]Department of Supply Chain and Information Management (SCM), The Hang Seng University of Hong Kong, Shatin, N.T., Hong Kong
[2]Big Data Intelligence Centre, School of Decision Sciences, The Hang Seng University of Hong Kong, Shatin, N.T., Hong Kong
[3]Department of Industrial and Systems Engineering (ISE), The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Acknowledgements

The authors would like to thank the Big Data Intelligence Centre of The Hang Seng University of Hong Kong, and the Department of ISE, The Hong Kong Polytechnic University for supporting the research.

Contact/Corresponding Authors:
Dr WU, Chun Ho
Big Data Intelligence Centre, The Hang Seng University of Hong Kong
Dept of ISE, The Hong Kong Polytechnic University
jackwu@ieee.org
Phone: +852 3963 5550

Dr LAM, Hoi Yan
Dept of SCM, The Hang Seng University of Hong Kong
cathylam@hsu.edu.hk
Phone: +852 3963 5639

Abstract
For companies to gain competitive advantage, an effective customer relationship management (CRM) approach is necessary. Based on customer purchase behaviour and ordering patterns, companies can be classified into different categories in terms of providing customised sales and promotions for customers. However, companies that lack an effective CRM strategy can only offer the same sales and marketing strategies to all customers. Furthermore, the traditional approach to managing customers is control via a centralised method, in which the information regarding customer segmentation is not shared among the customer network. Consequently, valuable customers may be neglected, resulting in the loss of customer loyalty and sales orders, and the weakening of trust in the customer–company relationship. This paper designs an integrated data analytic model (IDAM) in a peer-to-peer cloud, integrating RFM-based k-means clustering algorithm, analytical hierarchy processing and fuzzy logic to divide customers into different segments and hence formulate a customised sales strategy. A pilot study of IDAM is conducted in a trading company specialised in providing advanced manufacturing technology to demonstrate how IDAM can be applied to formulate an effective sales strategy to attract customers. Overall, this study explores the effective deployment of CRM into the peer-to-peer cloud so as to facilitate sales strategy formulation and trust between customers and companies in the network.

Keywords: data analytic, customer relationship management, sales strategy formulation, customer behaviour, peer-to-peer cloud

1. Introduction

In today’s highly competitive business environment, global trading has become a popular phenomenon. In recent decades, there has been a consistent increase in the international import and export of raw material and products. Customers are always the most important assets for a company, especially for a trading company (Kang et al., 2015). This is because customers will place further orders to a company if they are satisfied with the products or services provided. Customers have consistent expectations that the salesperson will provide all-round solutions with a minimum cost of service (Sawik, 2015). However, the marketing department of an organisation must determine an appropriate sales and marketing strategy to promote their products. The effective integration between the activities of the sales and marketing functions is deemed to be responsible for the success of the firm (Orman, 2000). In light of the above situation, there has been an increasing focus on customer relationship management (CRM) in recent years, which is regarded as an active research area in terms of the identification, attraction, development and retention of customers (Gummesson, 2011; Buttle and Maklan, 2019). Although there are a number of methods and technologies to achieve CRM, the practicality and usefulness of the sales strategy formulation to retain customers is a key challenge for trading companies. The effective deployment of CRM in the real-life business environment is still under-researched in terms of data collection and analytics. Without an effective deployment approach, the loss of customer satisfaction and sales profit may follow, due to ineffectiveness in managing the brand and sales performance in the trading industry (Mullins et al. 2014). Thus, trading companies are keen to explore an effective and practical model to analyse customer order patterns in order to learn customer buying behaviour so that an effective promotion strategy can be designed to retain existing, valuable customers. However, from the customers’ perspective, they traditionally have no knowledge of the methods and key figures required for CRM. Consequently, fairness and trust cannot be fully established in the customer network, which makes it difficult to maintain healthy competition between customers. Therefore, the aim of this research is to propose and develop an integrated data analytic model (IDAM) in the peer-to-peer (P2P) cloud to formulate sales strategy based on customer order patterns and their buying behaviour. Firstly, a customer behaviour analysis is conducted to classify customers into different segments so as to formulate an appropriate sales strategy. A hybrid artificial intelligence approach is then proposed to prioritise products for promotion sales and determine the pricing strategy based on customer values. Lastly, a customised sales promotion strategy is presented which aims to attract customers through a mobile application. To verify the practicality of the proposed model and illustrate its deployment, a case study on a trading company who provides manufacturing technology equipment is outlined.

The contribution of this paper is two-fold. First, the integration of data-driven approaches and artificial intelligence (AI) techniques, namely k-means clustering, analytical hierarchy processing (AHP) and fuzzy logic, generates the synergy for CRM. The proposed model provides an effective and systematic method to the aspects of customer classification and sales strategy formulation. Second, the proposed model is developed on the foundation of the P2P cloud, which achieves great flexibility in information sharing and transparency among the customer network. The customer themselves can understand the mechanism of customer classification and the average performance of the clusters so that healthy competition can be created in the business environment. This is regarded as a novel element in CRM that further strengthens fairness and trust among customers.

This paper is organized as follows. The literature related to CRM, data mining and AI techniques is reviewed in Section 2. Section 3 presents the design of IDAM. In Section 4, a case study is presented to demonstrate how IDAM is implemented. In Section 5, the results and discussion of the model are discussed, and a conclusion is drawn in Section 6.

2. Literature Review

Nowadays, customers have become the most important asset for most companies. Maintaining a good customer relationship has indisputable value for companies and facilitated the emergence and development of CRM (Batter & Batter, 2010; Pashchenko, 2020). CRM is derived from the concept of the customer-focused business strategy. It is recognised as a business processes that can capture, retain and create values for customers as well as stockholders (Lambert, 2010). According to Buttle and Maklan (2019), CRM is usually adopted to support organisations in managing their interactions with supply chain partners such as suppliers and customers, for the establishment of a good relationship. In addition, CRM acts as a communication channel to maintain and develop a long-term positive relationship with the current and targeted customers of the company (Bhat and Darzi, 2016).

In general, CRM can be divided into four dimensions: customer identification, customer attraction, customer retention and customer development (Tudor et al., 2011; Ngai et al., 2009). Customer identification involves targeting the potential and profitable population who have the potential to be customers, or the group of existing customers that may bring companies additional profits (Martinez & del Bosque, 2013). This involves the classification of customers into different segments to devise customised sales and marketing strategies. Customer attraction is the second phase, whereby organisations make the effort to promote and allocate resources so as to attract target customers, and maximise the effectiveness in direct marketing that motivates customers to purchase through various channels (Mars and Gouider, 2017). Customer retention aims to meet customers' expectations and increase their satisfaction based on a customised marketing approach, which is a critical step in building a long-term relationship with customers. For customised marketing, the marketing strategy is formulated based on customer needs and requirements, by detecting, analysing and predicting the changes in customer behaviour (Chen et al., 2017). The last stage, customer development, is the continuous expansion of customer orders, profitability and customer loyalty (Ang & Buttle, 2009). To manage the customer relationship effectively, companies have made efforts to develop company websites or apps to promote and sell products via the Internet. With the help of the CRM strategy, the preference, background and expectations of customers can be determined and placed into different categories with respective profitability to minimise the customer churn rate (Xiao et al., 2016). Although CRM helps to improve the branding and sales performance of a company, it lacks the ability to discover hidden patterns in large amounts of data so as to generate meaningful knowledge.

Before examining data analytics in the domain of CRM, it should be noted that information and communication technologies (ICTs) are promising avenues for achieving sustainable development in economic growth and industrial innovation (Wu et al., 2018). The adoption of ICTs can facilitate the process of data collection, gathering and aggregation in the orchestration centre. In terms of data management, the P2P cloud is a promising network to facilitate data collection and sharing between all related stakeholders (De et al., 2016; Naik and Keshavamurthy, 2019; Sun, 2020). The P2P cloud is defined as the implementation of cloud systems into the P2P network, which can prevent bottlenecks and a single point of failure in application systems. In contrast to the traditional cloud and client-server deployment, P2P cloud has the benefits of accessing shared resources so that information transparency and system reliability are greatly improved. Data mining is a process of obtaining useful information or knowledge by centralising and analysing a large amount of data (Yan et al., 2020;). It is a business-driven tool for organisations in knowledge management. Ranjan & Bhatnagar (2011) stated that data mining can help organisations conduct market basket analysis and customer segmentation in CRM. Yadav et al. (2013) also illustrated the application and benefits of CRM in terms of customer profit-making ability analysis and forecasting. Meanwhile, Ngai et al. (2009) summarised seven types of data mining models commonly used in CRM: classification, forecasting, clustering, association, sequence discovery, regression and visualization models. Furthermore, Shim et al. (2012) applied the association rule for customer behaviour and pattern analysis for the development of CRM strategies, and Hosseini et al. (2010) proposed the application of K-means clustering in customer knowledge management. Clustering customers into different groups can provide clustering analysis based on different customer characteristics such as spending habits, activities and spending psychology. Wu et al. (2016) found that the analytics methods proved promising for large amounts of data with respect to identifying hidden relationships and patterns to establish novel applications. This can help companies to develop more efficient and effective marketing strategies and enable them to choose suitable marketing channels for advertising campaigns to improve the customer relationships and assist future planning.

Different from data mining, AI is a group of techniques that can simulate the thinking of human beings. It not only indicates the characteristics of data but is also capable of making decisions and solving problems. AHP and fuzzy logic are typical examples of AI techniques. AHP is a well-established multi-criteria decision-making method to tackle problems involving both intuitive and rational factors (Barker & Zabinsky, 2011). This method helps choose the best alternative from a wide range of options which are assessed regarding select criteria (Saaty & Vargas, 2012). Nepal et al. (2010) proposed a fuzzy-AHP approach to prioritise subjective customer satisfaction attributes in order to improve the vehicle design. Ho et al. (2011) considered the company stakeholder requirements and preferences with AHP when making decisions on supplier sourcing in an automobile manufacturing company. Meanwhile, Singh (2013) applied AHP to prioritise the strategic factors for managing a coordinated supply chain. Different from AHP, fuzzy logic can be applied to a decision support system to deal with imprecise, vague and subjective data, as the fuzzy variables allow users to indicate the fuzzy characteristics of the goods between 0 and 1 (Chui & Ip, 2017; Jin, 2020). Fuzzy inference has been applied to the customer requirement information system (Pourjavad and Mayorga, 2019). For example, Kusan et al. (2010) applied fuzzy logic to predict the selling price of house building, whereas Lin et al. (2011) proposed an agent-based price negotiation system based on fuzzy logic to customise the price negotiation strategies on the Internet. Chakraborty et al. (2013) determined the price discount for selling deteriorating seasonal products in order to increase the storage capacity of a warehouse.

In summary, the above review indicates that effective CRM can assist companies in attracting new customers and retaining valuable ones by providing customised services and an appropriate sales and marketing strategy. However, the existing CRM approach can only collect customer buying behaviour; the means of analysing data for resources allocation is neglected. Therefore, by integrating cloud technology, data analytic tools, and artificial intelligence techniques including AHP and the fuzzy logic approach, hidden customer data can be detected to deploy marketing strategies that can target, attract and maintain long term profitable relationships with customers.

3. Design of the Integrated Data Analytic Model

The design of the IDAM is shown in Figure 1. Four modules are designed, including: (i) Module 1: cloud-based information services module (CISM), (ii) Module 2: customer behaviour analysis module (CBAM), (iii) Module 3: sales strategy formulation module (SSFM); and (iv) Module 4: mobile CRM app implementation module (MAIM).

3.1. Module 1: Cloud-based Information Services Module

To run the IDAM successfully, the data is uploaded to the cloud-based platform and stored in the P2P network, which is updated on a regular basis. Instead of storing everything in the centralised cloud database, the proposed model is developed in a P2P cloud environment in which information can be shared in the customer network free from conflict of interests and confidentiality concerns. Such shared information can motivate customers to engage in transactions and interactions with companies so as to achieve a mutual beneficial relationship. In this study, the discounts of promotional products are the incentives to drive the engagement of customers in the P2P cloud environment. Simple cloud computing as the foundation upon which to deploy the proposed model is efficient at managing a large amount customer and product data, but it is difficult to establish trust in the customer network. Therefore, the P2P cloud, which hybridises the P2P network and cloud as a whole, is designed to achieve effective data sharing and trust establishment. The data uploaded includes product ID, product name, customer ID, customer name and transactions records. The process of extraction, transformation and loading (ETL) is then required. During the ETL process, the preferred data are identified and extracted from internet-enabled devices and enterprise social media. The incomplete and incorrect data are sorted out and the data in an inappropriate format are reformatted. The extracted data are then physically loaded into the data warehouse for further data analytics.

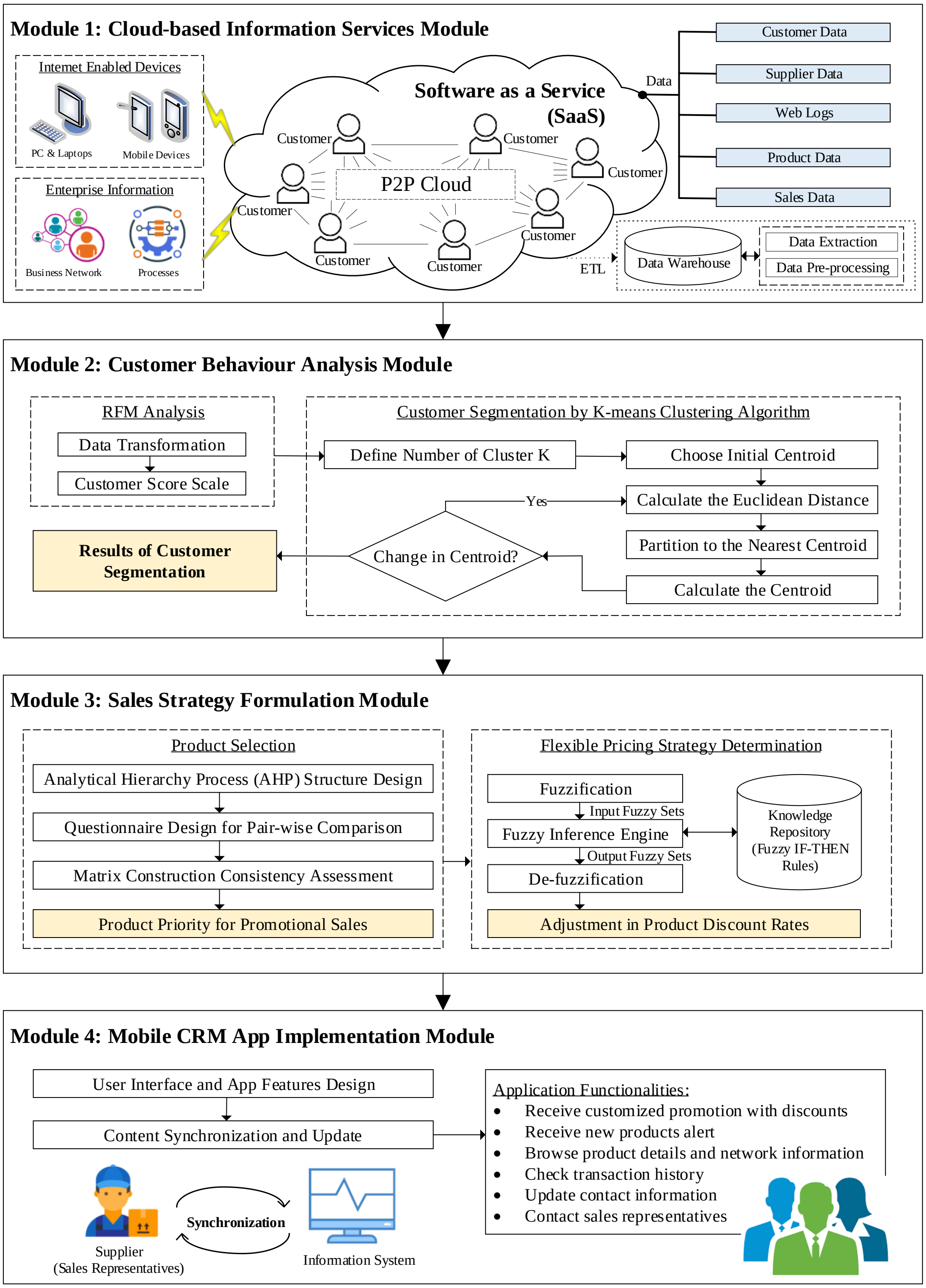

Figure 1. Modular framework of the integrated data analytic model

3.2. Module 2: Customer Behaviour Analysis Module

After the data are retrieved from the centralised data warehouse, an RFM analysis is conducted to establish the customer loyalty to the company. As each set of data are different, a tailored ranking score structure, i.e. a customer quintiles method, is applied to normalise the data. After obtaining the RFM score of each customer, customer segmentation with an RFM-based clustering algorithm is done to partition the customers into different sectors. Hence, the trading company can propose the most appropriate marketing strategies to each sector of customers. RFM analysis is an analytical technique than analyses customer behaviour by evaluating the data in terms of three attributes: Recency (R), Frequency (F) and Monetary (M). Recency is the time interval between the latest purchasing behaviour of the customer and the present; frequency is the number of transactions made by the customer in the specific period of time; and monetary is the amount of money spent by the customer in the particular time period. In recent years, RFM analysis has been found to show promise in analysing customers' patterns and behaviour (Dursun and Caber, 2016; Stormi et al., 2019). Particular to the CRM, the RFM analysis is capable of discovering the characteristics of customers for specific business sectors, and thus it is selected in this study to be incorporated into the k-means clustering for the customer segmentation. As each set of data is different, a tailored ranking score structure is obtained by normalising the data collected and transforming the value to a 5-point scale. Table 1 shows the notation table of RFM analysis, and Table 2 shows the conversion of RFM data to score.

Table 1. Notation table of RFM analysis

Notation	Description
R_n	Recency of the n^{th} customer
F_n	Frequency of the n^{th} customer
M_n	Monetary value of the n^{th} customer
R_m	Maximum recency of N customers, where N={1,2,…,n}, i.e. R_m=max{$R_1,R_2,\ldots,R_n$}
F_m	Maximum frequency of N customers, where N={1,2,…,n}, i.e. F_m=max{$F_1,F_2,\ldots,F_n$}
M_m	Maximum monetary value of N customers, where N={1,2,…,n}, i.e. M_m=max{$M_1,M_2,\ldots,M_n$}
R_s	Minimum recency of N customers, where N={1,2,…,n}, i.e. R_s=max{$R_1,R_2,\ldots,R_s$}
F_s	Minimum frequency of N customers, where N={1,2,…,n}, i.e. F_s=max{$F_1,F_2,\ldots,F_s$}
M_s	Minimum monetary value of N customers, where N={1,2,…,n}, i.e. M_s=max{$M_1,M_2,\ldots,M_s$}
$\overline{R}$	Average recency value, where $\overline{R}$ = (R_m - R_m)/5
$\overline{F}$	Average frequency value, where $\overline{F}$ = (F_m - F_m)/5
$\overline{M}$	Average monetary value, where $\overline{M}$ = (M_m - M_m)/5

Table 2. A conversion table of RFM data to scores

Score	R	F	M
5	$\mathrm{R}_S \sim \overline{R} \times 1$	$\overline{F} \times 4 \sim F_m$	$\overline{M} \times 4 \sim M_m$
4	$\overline{R} \times 1 \sim \overline{R} \times 2$	$\overline{F} \times 3 \sim \overline{F} \times 4$	$\overline{M} \times 3 \sim \overline{M} \times 4$
3	$\overline{R} \times 2 \sim \overline{R} \times 3$	$\overline{F} \times 2 \sim \overline{F} \times 3$	$\overline{M} \times 2 \sim \overline{M} \times 3$
2	$\overline{R} \times 3 \sim \overline{R} \times 4$	$\overline{F} \times 1 \sim \overline{F} \times 2$	$\overline{M} \times 1 \sim \overline{M} \times 2$
1	$\overline{R} \times 4 \sim R_m$	$F_s \sim \overline{F} \times 1$	$M_S \sim \overline{M} \times 1$

After finishing the table, the RFM score of each customer can be found. In total, there are 125 (= 5^3) types of customers based on the results of the RFM analysis. The most loyal and valuable customer receives a score of 555, showing that a recent purchase, a large number of purchase orders within the time period, and high monetary orders were placed. However, it would be time-consuming to design the tailor-made marketing strategies for all 125 types of customer. Hence, a RFM-based clustering algorithm is used to partition the customers into several segments to develop the most appropriate marketing strategies for different customer segments. In this paper, the number of clusters was set at 8, whereby the RFM values are divided into high and low level. A letter "h" or "l" is assigned when the RFM scores of each segment are higher than or lower than the mean values respectively. Table 3 shows the eight clusters defined by the RFM-based clustering algorithm.

Table 3. Eight clusters by RFM-based clustering algorithm

Cluster	RFM Arrangement	Customer Type
1	$R_hF_hM_h$	Loyalty
2	$R_hF_hM_l$	Shopper
3	$R_hF_lM_h$	Valuable
4	$R_hF_lM_l$	New
5	$R_lF_hM_h$	Old
6	$R_lF_hM_l$	Frequent
7	$R_lF_lM_h$	Spender
8	$R_lF_lM_l$	Reminder

3.3. Module 3: Sales Strategy Formulation Module

There are two tiers in this module: product selection and flexible pricing strategy determination.

3.3.1. Product Selection

Product Selection is designed by using the AHP to prioritise promotional products in sequence so as to attract customers. AHP is a multi-criteria decision-making technique for selecting alternatives based on a set of criteria. In this phase, the structure of AHP is first designed by identifying the goal, criteria and alternatives. Decision makers who are experienced in product selection for promotion should be selected carefully, as their opinion will be used to construct the AHP framework. They should have clear knowledge about the types of products that the company has on hand, the characteristics of each product, customer information, and the trend of product demand. To construct the structure, decision makers should first identify the alternatives to be classified and determine the most crucial criteria for consideration when evaluating the alternatives. The alternatives could be defined based on a selected set of promotional products. The criteria could be defined so as to allow the decision makers to evaluate and identify the type of products for promotion in a given period.

To facilitate decision makers to make pair-wise comparisons, a questionnaire should be designed. The aim of the questionnaire is to ask decision makers how they would express the importance ratio when comparing two criteria or alternatives. Considering the convenience of decision makers, the questionnaire should be designed to be user-friendly, that is, comparison results are expressed by a nine-point scale. After the questionnaires are collected, an (n x n) evaluation matrix A is constructed to present the result of the pairwise comparison, where n is the total number of criteria α_{ij}, $\mathrm{ij} = (1, 2 \ldots, \mathrm{n})$ is the quotient of the weights of the criteria.

$$A = [\alpha_{ij}] = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{\pi 1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{bmatrix}, \qquad \alpha_{ii} = 1, \alpha_{ji} = \frac{1}{\alpha_{ij}}, \alpha_{ij} \neq 0$$

By transforming the matrix into the standard matrix, a priority vector which shows the importance of each criterion can be calculated. The consistency ratio is checked in this step to ensure that inconsistency between each comparison due to subjective perception can be reduced.

3.3.2. Flexible Pricing Strategy Determination

To retain customers and get orders, a trading company usually provides different levels of discount to their potential customers. In this model, one of the common artificial intelligence techniques, fuzzy logic, is adopted to determine the discount rate that should be offered to the customers. Firstly, sales managers have to identify the input and output variables for determining the discount rate. Subsequently, it is essential to convert the input data set into fuzzy sets with the process of fuzzification by determining the universe of discourse and membership functions of the variables, as this is one of the critical steps for operating the fuzzy system. Fuzzy sets are defined for listing the linguistic variables, with a range of possible values of each variable. Membership function helps determine the characteristics of fuzzy set, whereas the universe of discourses is used to represent their corresponding membership functions based on different predicates. The predicates have special shapes and styles, such as triangle and trapezoid, to represent their membership functions. After finishing the fuzzification operation, fuzzy sets are prepared and the process proceeds to the inference process. This is a process which converts the input fuzzy set into output fuzzy set with the use of the fuzzy inference engine. A knowledge repository is developed to store the fuzzy rules in the form of an IF-THEN structure to define how the predicates of the output fuzzy set vary with the combinations of input dimensions' predicates. The final step is the defuzzification setting. The output solution has to process a reverse process to obtain crisp values. The defuzzification result is expressed in both numerical and graphical forms, indicating the centre of area. To support the decision making, the users may have to rely on the linguistic terms conversion table to convert the abbreviation shown in the system to clear wording, to ensure the full understanding of the calculated result. Based on the output crisp value, the company is able to decide on the discount rate for their customers. Through the fuzzy logic process, a flexible pricing strategy with a discount rate can be formulated to attract customers in the ordering process.

3.4. Module 4: Mobile CRM App Implementation Module

After the RFM analysis, different marketing strategies for different customer segments are developed. Therefore, this tool allows the user to tailor their app content for specific customer segments. Different customer segments can view different content according to that set by the content manager. The information in the app format is synchronised with IDAM. There are two types of user accounts with different interfaces, i.e. customer and staff accounts. Apart from the general customer information, promotions and discounts will be allocated based on the results of the customer data analysis in the customer account, whereas the customer order can be viewed in the staff account.

4. Case Study

In the trading industry which provides manufacturing technology equipment, once a purchase order is received from the customer or dealer, the trading company then contacts its supplier to produce and ship the ordered product to them. The product is then delivered to its customer or dealer to complete the order. Under such a practice, there is no value-added service provided by the trading company. The only service that the trading company can offer is the after sales service, including training and maintenance. Therefore, customers are easily lost, as they can contact the supplier directly and purchase the product. To retain the customer and fulfil their needs, the trading company usually has to offer different selling prices to customers. The discount rate changes periodically, which requires the salesperson to calculate the price based on past experience and the existing market environment. However, due to the lack of decision support in such a flexible pricing strategy, companies can only offer the same sales and marketing strategies to each customer. Valuable customers may be neglected, which results in the loss of customer loyalty and even the loss of a sales order. Thus, it is crucial to develop a system which can determine flexible pricing strategy to retain valuable customers. The implementation roadmap of IDAM is divided into five stages, namely (i) data collection by cloud technology, (ii) customer behaviour analysis, (iii) product selection, (iv) flexible pricing strategy determination, and (v) mobile CRM app implementation.

4.1. Stage 1: Data collection by P2P cloud technology

The P2P Cloud-based platform allow real-time information transmission and sharing. All data retrieved from multiple stores are required to be pre-processed by data extraction, transformation, and loading (ETL) which is a general process in database usage for discovering the required data for further processing, especially in data warehousing. Firstly, all the product and customer data have to be identified and extracted from the existing sales system in the company. Next, as different salespeople may use different data formats for data storing, the retrieved data may be incomplete, in a wrong file size and in an inappropriate format. The data format, size and the primary key should be revised and amended so that the correlated data can be linked and then foreign keys determined for data extraction in various tables in the database. In addition, the incomplete and incorrect data are sorted and the data that are in an inappropriate format are reformatted. The data are also transformed into the correct size. Lastly, the extracted data are physically loaded to the data warehouse for further processing.

4.2. Stage 2: Customer behaviour analysis

To analyse customer behaviour, 218 customer records were extracted and evaluated with RFM analysis. By calculating the RFM scores of each customer, an RFM-based clustering algorithm was then used to divide the customers into eight segments. Figure 2 shows the RFM-based clustering result. The findings show that 17% of customers are classified as new customers, indicating they are new to the company and have placed an order recently. The company should propose a new scheme to keep and develop this kind of customer, in order to form a long-term business relationship. Subsequently, both loyal customers and frequent customers total 14.2% of the customers, showing that they have placed large orders recently and frequently. These are the most valuable customers to the company. Approximately 12.4% of the customers belong to the segment. Spenders are the customers willing to pay for the product. However, they may not place orders frequently and recently. Most customers who target the purchase of large equipment may belong in this group, as they do not need to buy large machines frequently. Meanwhile, nearly 12% of customers can be classified as old customers and shoppers. This segment indicates that although old customers have not placed orders recently, they placed large orders frequently. The company should therefore provide incentives such as discounts to persuade both loyal customers and old customers to place orders again. That leaves shoppers as the customers who always place orders, implying that the target products they need are small instruments or accessories.

After the customers are divided into eight segments, sales and marketing staff can start to develop and propose suitable marketing strategies for each type of customer. The marketing strategies proposed according to the customer type can increase the customer response rate and encourage the customer to visit and repeat the order. For instance, the company can send promotion items to the customer of the old customer type to get their attention and encourage them to place the order again. For the loyalty customer type, the company can offer certain special discounts and some follow-up advice to thank them for their support for the company and maintain a good long-term relationship with them.

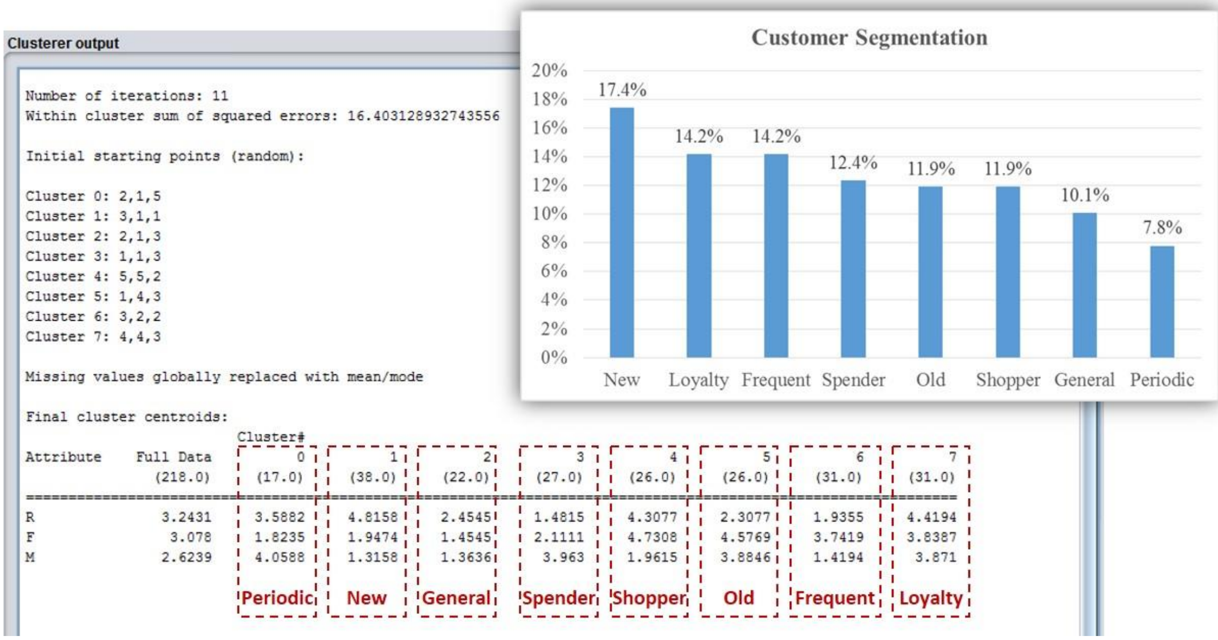

Figure 2. Clustering results by using CBAM

4.3. Stage 3: Product selection

In order to select the product for promotion, the AHP structure is designed, as shown in Figure 3. This is a 2-level hierarchy which consists of goal (level 0), criteria (level 1) and choices (level 2). Based on the goal of prioritising products for promotion, decision makers including top management staff have to discuss the potential criteria, and they agree that they should use those which would have a positive impact on profit earning. Therefore, five criteria – profit (C1), customer needs (C2), flexibility (C3), functionality (C4) and competition (C5) – are defined. The description of the five criteria is given in Table 4.

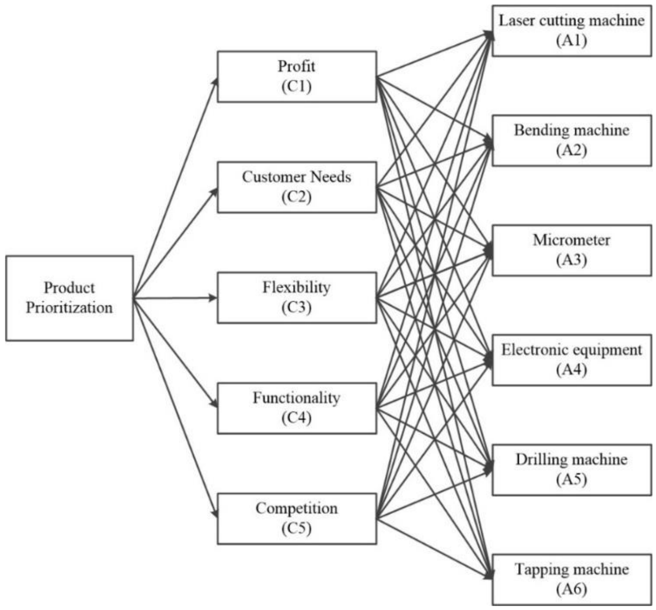

Figure 3. Hierarchical structure of the product prioritisation for AHP

Table 4. Description of production prioritisation criteria

Criteria	Description
Profit (C1)	This criterion evaluates whether the goods can bring profit to the company, especially when the company has to make a decision on whether to promote it to its customers at a promotional price.
Customer needs (C2)	Customer needs refer to whether the product selected for promotion has met the customer demand. If the number of enquiries for a product increases within a period of time, it shows that customers are looking for similar kind of products in the market. Hence, the company should consider whether promotions should be placed to persuade the customers who are interested to place the order.
Flexibility (C3)	Flexibility measures whether the goods are always available from the supplier due to a different supply lead time; it is vital to ensure that the goods are delivered to the customer without delay.
Functionality (C4)	Functionality refers to the function of goods. It represents whether the function and specification of the goods suit the needs of customers. If there is an increasing number of enquiries for a product regarding its new function, it shows that an increasing number of customers are searching for the product with such new function. Hence, functionality becomes one of the important criteria in product selection for promotion.
Competition (C5)	Competition refers to whether the product has severe competition in the market currently. If there is more than one competitor in the market who sell the same types of product to customers, the company should consider whether a promotion should be made for such a product to attract customers in the market.

Based on the hierarchy structure, questionnaires are designed and distributed to the decision makers for pairwise comparisons. Figure 4 shows the pair-wise comparison of criteria with respect to the goal “Project Prioritisation”. 10 questions are included to weight the five criteria with respect to the goal. Based on the results of the questionnaires, the matrix regarding the comparison of the goal to each criterion is shown in Table 5. Take the criteria “Customer Needs (C2)” and

"Competition (C5)" as an example. An upward arrow with a value of 5 is assigned, indicating that Customer Needs (C2) is 5 times more important than Competition (C5) when considering the goal of product prioritisation. By conducting the pair-wise comparison of criteria with respect to the five criteria, the matrix of alternatives with respect to the five criteria are also constructed respectively.

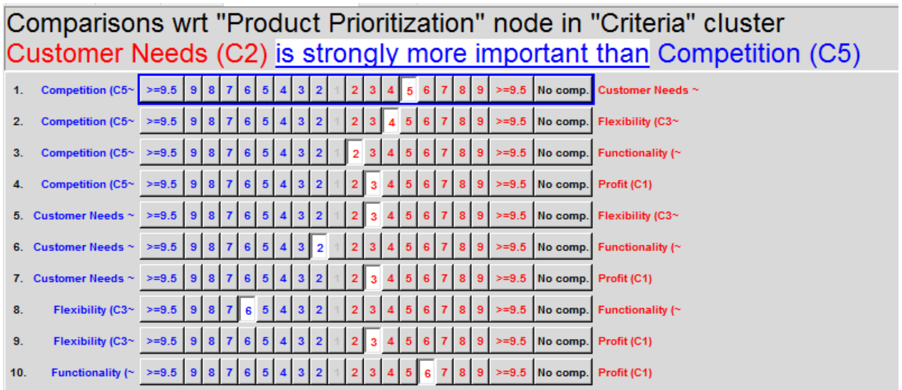

Figure 4. Pair-wise Comparison of criteria with respect to goal "Project Prioritisation"

Table 5. Matrix of criteria with respect to goal

	Customer Needs (C2)	Flexibility (C3)	Functionality (C4)	Competition (C5)
Profit (C1)	← 3	← 3	← 6.48	← 3.33
Customer Needs (C2)		↑ 2.72	← 2	← 5
Flexibility (C3)			← 6	← 3.74
Functionality (C4)				← 2

After the construction of the matrix, the weights of each criterion and each alternative with respect to each criterion are calculated respectively. The consistency ratio is found to be 9.9%, which, as it is smaller than 10%, shows that the result of the pair-wise comparison has acceptable consistency. The criterion "Profit" has the highest weight among all criteria, accounting for 43.8% of the total weight. Therefore, it is obvious that "Profit" is emphasised by decision makers. This implies that the company would pay more attention to analysing whether more profit can be gained if a product is selected for promotion. To obtain the overall result, the six alternatives are compared based on the five criteria. Figure 5 shows the overall weights of product and their final ranking. The result shows that Drilling Machine has the highest weighting of 24.5%, followed by Bending Machine and Laser Cutting Machine, which account for 18.2% and 15.2% respectively. This suggests that, after consideration of profit, customer needs, flexibility, functionality and competition, Drilling Machine should be selected for promotion so as to attract more customers.

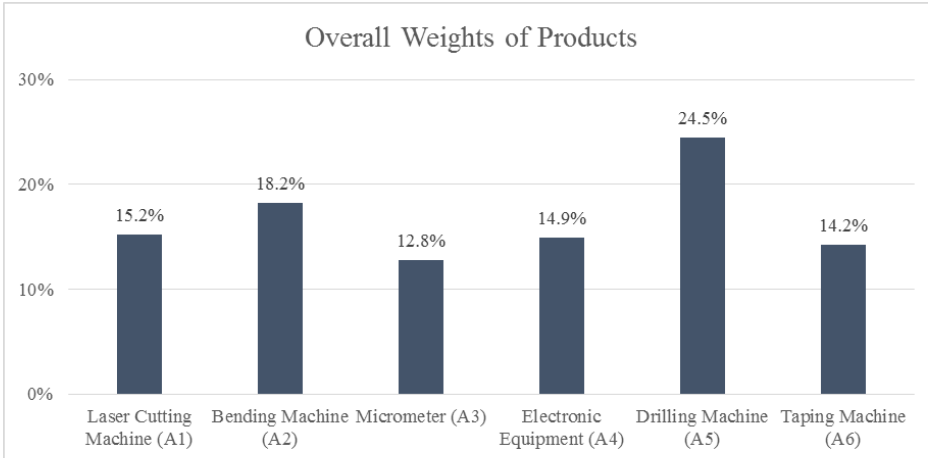

Figure 5. Overall weights of products

4.4. *Stage 4: Flexible pricing strategy formulation*

After investigating the key decision-making criteria, the findings indicate that the discount rating is determined based on five major criteria: (i) customer rating, (ii) order volume, (iii) change in supplier price, (iv) change in exchange rate, and (v) change in forecasted demand. The company classifies its customers into different grades, such that valued customers have a higher rating and enjoy more benefits. As the five key decision-making criteria affect the discount rate that should be given, they become the input parameters in the module, whereas the required change in discount rate is defined as the output parameter. Figure 6 shows the membership functions of the input and output parameters in terms of determining the discount rate.

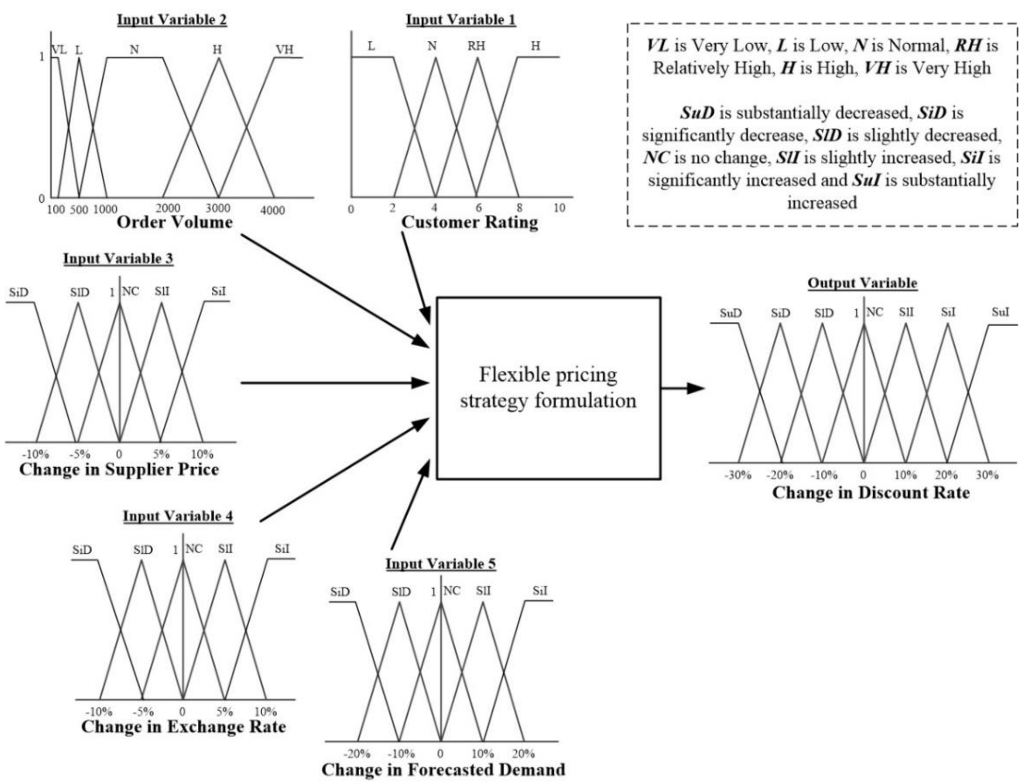

Figure 6. Membership functions for flexible pricing strategy determination

After defining the fuzzy sets and membership functions of the five input variables and the output of change in discount rate, the fuzzy rules also need to be defined in the knowledge repository so that successful rules can be fired to obtain the solution. Given the input crisp values of the five input variables, the membership values of each fuzzy set can be determined. By firing the related fuzzy rules, the crisp values of the output parameters are calculated to determine the consequent fuzzy region of each output fuzzy set. Figure 7 shows the result of a change in discount rate, with the result that a discount of 15% is suggested for this customer.

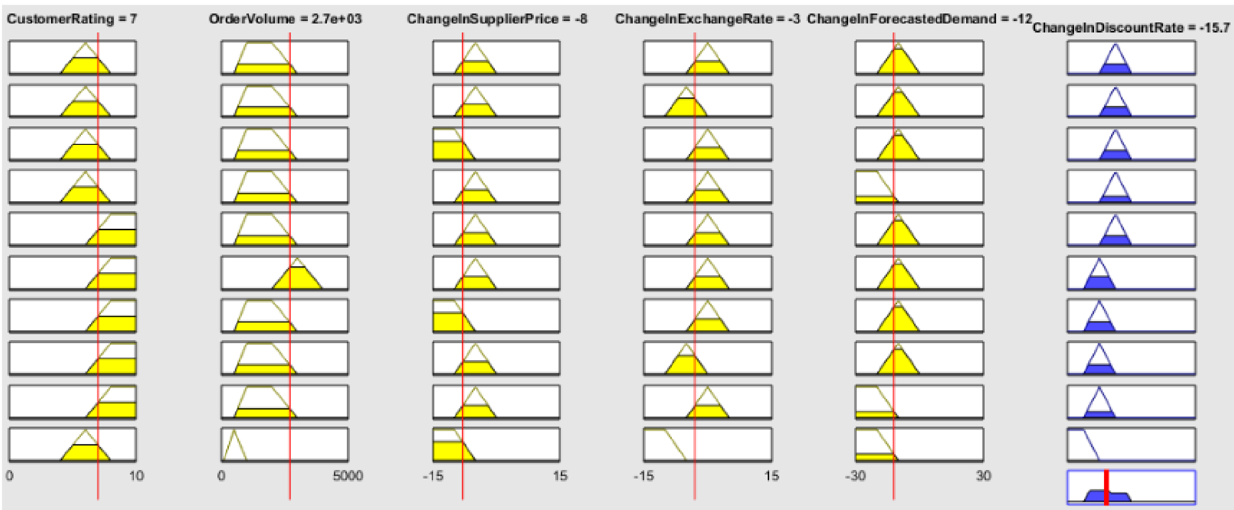

Figure 7. Changes in discount rate by using fuzzy logic

4.5. Stage 5: Mobile CRM app implementation

In this stage, the data analysed results, customer and product information are integrated into the mobile app platform. Users who have downloaded the app can receive updated information about the company and can check the latest updated products and news. They can also browse the detail of upcoming promotions. Moreover, customers can contact the sales staff for sales enquiries or technical support and can check their purchase history and after sales service information. Furthermore, they can receive their tailored discounts or promotions in the VIP offer interface according to the result of the customer data analysis, as shown in Figure 8.

5. Results and Discussion

5.1. Results and Discussion of the Customer Behaviour Analysis Module

The CBAM is used to divide the customers into different groups so as to formulate a customised sales strategy. By calculating the RFM value for each customer, the RFM-based clustering algorithm is then applied to partition the customers into eight sectors: loyalty, shoppers, spenders, periodic, frequent, old, new and general. To adopt an RFM-based clustering algorithm, the number of clusters should first be defined. To evaluate the performance of the CBAM, tests are conducted to determine a suitable setting for the number of clusters, i.e. the number of clusters are set as 2, 4, 6, 8 and 10. Tables 6 to 10 show the RFM clustering results for 2, 4, 6, 8 and 10 clusters respectively.

Figure 8. User-interfaces of the proposed model for VIPs

Table 6. RFM clustering result for 2 clusters

Cluster	**Size**	**Recency (R)**	**Frequency (F)**	**Monetary (M)**	**RFM Pattern**	**Customer Type**
C1	98	2.8061	3.2755	3.9388	$R_hF_hM_h$	Loyalty
C2	120	3.6	2.9167	1.55	$R_hF_hM_l$	Shopper
Overall	**218**	**3.2431**	**3.078**	**2.6239**		
Within cluster sum of squared errors: 55.73						

Table 7. RFM clustering result for 4 clusters

Cluster	**Size**	**Recency (R)**	**Frequency (F)**	**Monetary (M)**	**RFM Pattern**	**Customer Type**
C1	51	3.8039	3.5098	4.2549	$R_hF_hM_h$	Loyalty
C2	54	4.3333	1.8889	1.3333	$R_hF_lM_l$	Periodic
C3	65	3.0462	4.4462	2.1538	$R_hF_hM_l$	Shopper
C4	48	1.6875	2.1042	2.9792	$R_lF_lM_h$	Spender
Overall	**218**	**3.2431**	**3.078**	**2.6239**		
Within cluster sum of squared errors: 34.68						

Table 8. RFM clustering result for 6 clusters

Cluster	**Size**	**Recency (R)**	**Frequency (F)**	**Monetary (M)**	**RFM Pattern**	**Customer Type**
C1	27	3.963	2.2222	4	$R_hF_lM_h$	Periodic
C2	37	4.8108	1.9459	1.2703	$R_hF_lM_l$	New
C3	43	2.2326	2.3953	1.2791	$R_lF_lM_l$	General
C4	27	1.4815	2.1111	3.963	$R_lF_lM_h$	Spender
C5	45	4.4222	4.4444	2.8889	$R_hF_hM_h$	Loyalty
C6	39	2.2308	4.5897	3.2051	$R_lF_hM_h$	Old
Overall	**218**	**3.2431**	**3.078**	**2.6239**		
Within cluster sum of squared errors: 23.17						

The results of the analysis reveal that customers are classified into different customer types when the number of clusters defined change. The company may find it difficult to formulate appropriate strategies if the number of customer types is too small. Figure 9 summarises the customer types in different number of clusters. As is shown, loyalty is a persistent customer type even though the number of clusters increases, and spender and shopper are suggested in most cases when setting a different number of clusters. In addition, the number of customer types is limited to eight even though the number of clusters is set as ten. Therefore, the number of clusters should not be set to more than eight, as some groups will have the same pattern and will ultimately be merged into one group.

Table 9. RFM clustering result for 8 clusters

Cluster	**Size**	**Recency (R)**	**Frequency (F)**	**Monetary (M)**	**RFM Pattern**	**Customer Type**
C1	17	3.5882	1.8235	4.0588	$R_hF_lM_h$	Periodic
C2	37	4.8108	1.9189	1.2973	$R_hF_lM_l$	New
C3	22	2.4545	1.4545	1.3636	$R_lF_lM_l$	General
C4	27	1.4815	2.1111	3.963	$R_lF_lM_h$	Spender
C5	26	4.3462	4.6538	1.9231	$R_hF_hM_l$	Shopper
C6	30	2.2333	4.6333	3.5	$R_lF_hM_h$	Old
C7	26	1.9231	3.5385	1.3462	$R_lF_hM_l$	Frequent
C8	33	4.3636	3.8788	3.8788	$R_hF_hM_h$	Loyalty
Overall	**218**	**3.2431**	**3.078**	**2.6239**		
Within cluster sum of squared errors: 16.4						

Table 10. RFM clustering result for 10 clusters

Cluster	**Size**	**Recency (R)**	**Frequency (F)**	**Monetary (M)**	**RFM Pattern**	**Customer Type**
C1	11	3.9091	1.4545	4.4545	$R_hF_lM_h$	Periodic
C2	38	4.8158	1.9474	1.3158	$R_hF_lM_l$	New
C3	18	2.4444	1.3333	1.2222	$R_lF_lM_l$	General
C4	17	1.3529	1.5882	3.7647	$R_lF_lM_h$	Spender
C5	11	5	4.9091	2.0909	$R_hF_hM_l$	Shopper
C6	23	1.6522	4.1739	1.8261	$R_lF_hM_l$	Frequent
C7	20	2.4	2.65	2	$R_lF_hM_l$	Frequent
C8	39	4.2051	4	3.6923	$R_hF_hM_h$	Loyalty
C9	15	3.6	4.4	1.4667	$R_hF_hM_l$	Shopper
C10	26	2.1154	4.0385	4.4615	$R_lF_hM_h$	Old
Overall	**218**	**3.2431**	**3.078**	**2.6239**		
Within cluster sum of squared errors: 15.34						

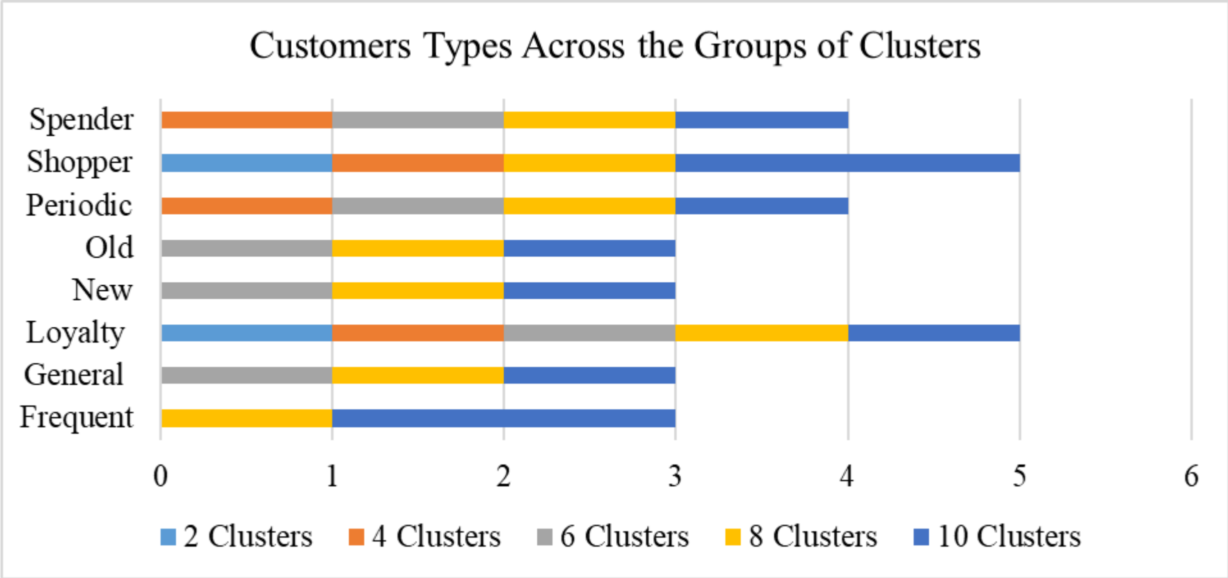

Figure 9. Customer types in different number of clusters

Another comparison in terms of defining the number of clusters is the within-cluster sum of squared errors, which is the measurement of the variability of the observations within each cluster. This calculates the distance between each customer's RFM values and the corresponding cluster centre. The smaller the figure of the within-cluster sum of squared error is, the better the performance of customer segmentation is. As shown in Figure 10, the within cluster sum of squared errors decreases when the number of clusters increases. The value decreases significantly from 2 clusters to 6 clusters, whereas the value becomes steady from 8 clusters to 10 clusters. Summarising the result using customer types and the within cluster sum of squared errors indicates that the number of clusters should be set as eight to obtain a better performance.

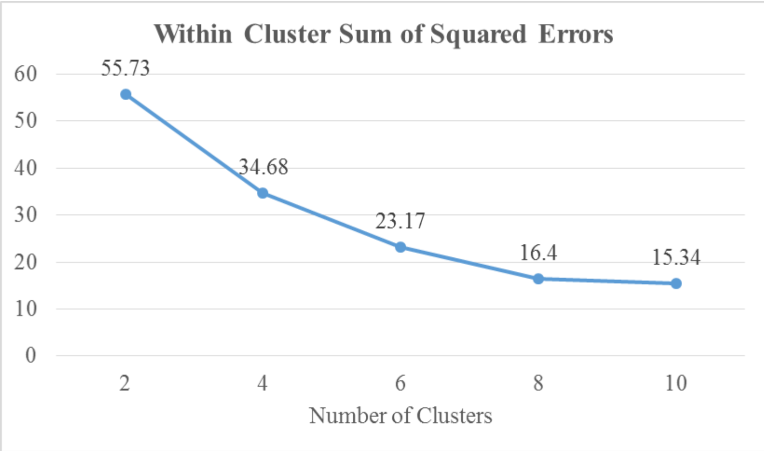

Figure 10. Comparison of *within cluster sum of squared errors* across the different number of clusters

5.2. Results and Discussion of the Sales Strategy Formulation Module

As the first part of the sales strategy formulation module, product selection is designed using AHP to prioritise and select products for promotion. Five criteria which affect the decision on the selection of promotional products are identified: profit, customer needs, flexibility, functionality and competition. To study whether the change in criteria weighting will affect the choice of alternatives, a sensitivity analysis of the five criteria is conducted, as shown in Figure 11. As shown in Figure 11, the current weight of profit, customer needs, flexibility, functionality and competition are 0.438, 0.156, 0.274, 0.07 and 0.061 respectively. As profit has the highest weighting among the five criteria, any change in the weighting of profit would result in a significant change in product selection. For instance, if profit becomes more important and the weighting of profit increases to a value larger than 0.53, the ranking of alternative products would change. The ranking of tapping machine and electronic equipment would be swapped, but no adjustment would be made to the first choice, i.e. drilling machine. However, if profit becomes less important and the weighting drops below 0.438, the result would be significant changes to the product ranking.

As the second part of the SSFM, a flexible pricing strategy is determined by using the fuzzy logic approach. Fuzzy logic can manage and deal with uncertainties effectively, and is used to determine the discount rate that should be offered to the customers. Following the result generated in the CBAM, customers are divided into different segments according to their purchase record. Customers with high scores in RFM values are classified as loyalty customers. This kind of customer may require high levels of attention from the company and sustainable business cooperation, and hence the company should offer them a higher discount to maintain a long-term, good relationship with them. By using the fuzzy logic approach, rules can be transferred to useful information that can be applied directly to the decision-making processes. Before using this method, managers calculated the discount rate solely by experience, which is hard to justify. Fuzzy logic can generate a suggested discount rate for reference and help to improve the quality of the decision-making process.

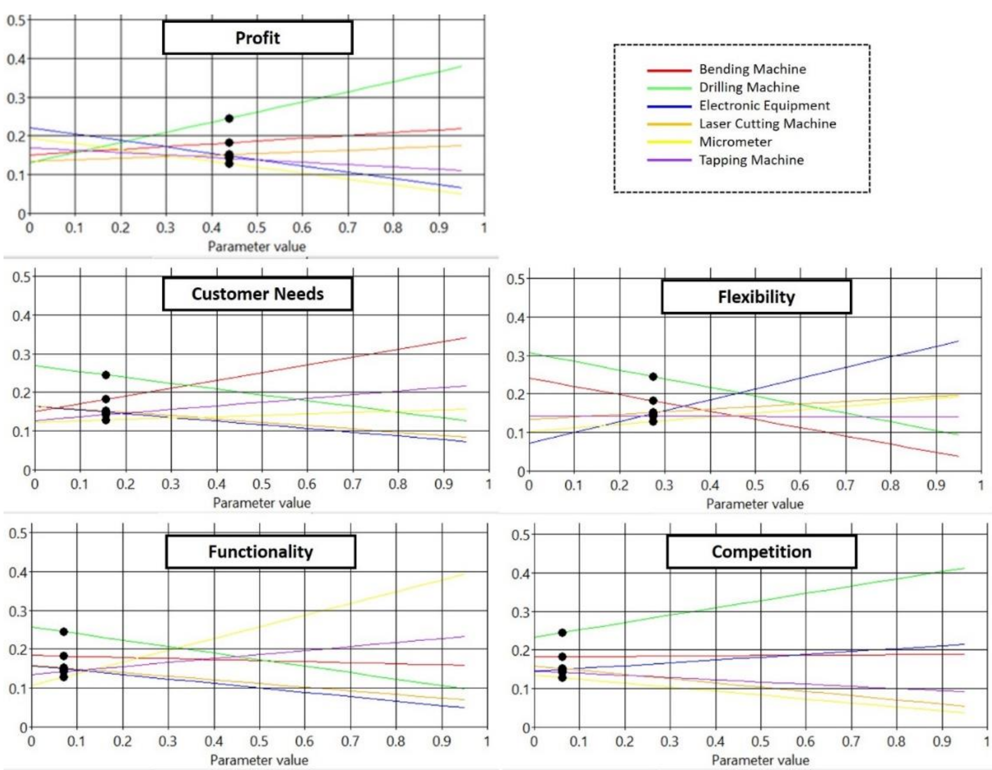

Figure 11. Sensitivity analysis for the five criteria

5.3. Managerial Implications

The deployment of the proposed model in the trading business environment provides an effective measure of sales strategy formulation according to the results of customer segmentation. In contrast to traditional business practice, sales strategies generated from the proposed model are customised to the nature of customers in terms of RFM aspects, for which promotional products and customised discounts are evaluated accordingly. On the one hand, by using the proposed model, companies can abandon the complicated and repetitive business processes required to investigate customers and products for the creation of sales strategies. In addition, the sales strategies become customised to customers, rather than a fixed strategy for all customers. Subsequently, the relationship between companies and high-value customers can be effectively developed through such customised sales strategies, and the mutual benefits can be maintained. On the other hand, the model formulation in the P2P cloud environment makes the application a pervasive service in the customer network. Inside the network, companies share business information, for example (i) mechanisms of customer segmentation and strategy formulation process, (ii) performance index of customer segmentation, (iii) list of promotional products, and (iv) customised strategies, except for sharing private and confidential information. Consequently, the transparency of the strategy formulation process is high, with the result that trust and fairness can be easily established in the customer network. With respect to different sizes of the customer network, the proposed model is developed as a Strategy Formulation as a Service (SFaaS) so that the system resources and specifications can be deployed on demand to satisfy multitenancy needs. To extend the proposed model, trading companies can implement the proposed SFaaS-based model to facilitate their own decision-making process for customer segmentation and product prioritisation, while the target companies should be facing challenges in customisation of sales strategies. Regarding the sales strategy formulation, customers and products are two essential factors to be considered to formulate customised strategies. Even though trading companies have different number of customer sizes, types, and product types, the proposed model is flexible and applicable to establish the most appropriate sales strategy in an intelligent manner. According to the business scales, the companies can deploy the proposed SFaaS-based model to achieve the customisation of sales strategies through the customer segmentation by using k-means clustering, product prioritisation by using AHP, and product discount evaluation by using fuzzy logic. Consequently, a smart trading business mode can be boosted and evolved in the peer-to-peer cloud network.

6. Conclusion

To conclude, trading companies should provide customised sales and marketing strategies to different types of customer, to retain and build long term relationships with them. Meanwhile, a customised sales strategy encourages loyalty from the customers, whereas the adoption of a flexible pricing strategy allows valued customers to enjoy special discounts when ordering. Therefore, in this paper, an IDAM was designed for enhancing customer satisfaction. By integrating cloud technology and integrated data analytic tools, the trading company can collect and analyse the searching behaviour of customers based on the data collected. Firstly, a RFM-based k-means clustering approach was designed to divide customers into different segments according to the RFM values of purchase orders. In this approach, customers with a combination of high RFM values are classified as valuable customers, and customised marketing strategies can be offered to maintain good long-term relationships with them. Secondly, the AHP approach was applied to prioritise and select products for promotion. Five criteria – profit, customer needs, flexibility, functionality and competition – were identified in the decision-making process. Thirdly, the fuzzy logic approach was designed to determine the discount rate for various type of customers so as to promote products that are of interest to the customer at attractive prices. To conclude, the developed model can provide a clear road map for exploring potential business opportunities in the trading industry, enabling trading companies to retain customers by building long term relationships with them.

References

Ang, L., & Buttle, F. (2009). Customer development strategies for exceeding expectations–An exploratory study. *Journal of Database Marketing & Customer Strategy Management*, *16*(4), 267-275.

Barker, T. J., & Zabinsky, Z. B. (2011). A multicriteria decision making model for reverse logistics using analytical hierarchy process. *Omega*, *39*(5), 558-573.

Battor, M., & Battor, M. (2010). The impact of customer relationship management capability on innovation and performance advantages: testing a mediated model. *Journal of marketing management*, *26*(9-10), 842-857.

Bhat, S. A., & Darzi, M. A. (2016). Customer relationship management. *International Journal of Bank Marketing*.

Buttle, F., & Maklan, S. (2019). *Customer relationship management: concepts and technologies*. Routledge.

Chakraborty, N., Mondal, S., & Maiti, M. (2013). A deteriorating multi-item inventory model with price discount and variable demands via fuzzy logic under resource constraints. *Computers & Industrial Engineering*, *66*(4), 976-987.

Chen, Q., Zhang, M., & Zhao, X. (2017). Analysing customer behaviour in mobile app usage. *Industrial Management & Data Systems*,117(2), 425-438.

Chui, A. B., & Ip, W. H. (2017). Improving merger and acquisition decision-making using fuzzy logic and simulation. *International Journal of Engineering Business Management*, *9*, 1847979017711521.

De, S., Barik, M. S., & Banerjee, I. (2016). Goal Based Threat Modeling for Peer-to-Peer Cloud. *Procedia Computer Science*, *89*, 64-72.

Dursun, A., & Caber, M. (2016). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism management perspectives*, *18*, 153-160.

Gummesson, E. (2011). *Total relationship marketing*. Routledge.

Ho, W., Dey, P. K., & Lockström, M. (2011). Strategic sourcing: a combined QFD and AHP approach in manufacturing. *Supply Chain Management: An International Journal, 16*(6), 446-461.

Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, *37*(7), 5259-5264.

Jin, S. (2020). A Bidirectional Reasoning Based on Fuzzy Interpolation. *International Journal of Software Science and Computational Intelligence (IJSSCI)*, 12(1), 1-14.

Kang, J., Alejandro, T. B., & Groza, M. D. (2015). Customer–company identification and the effectiveness of loyalty programs. *Journal of Business Research*, *68*(2), 464-471.

Kuşan, H., Aytekin, O., & Özdemir, İ. (2010). The use of fuzzy logic in predicting house selling price. *Expert systems with Applications*, *37*(3), 1808-1813.

Lambert, D. M. (2010). Customer relationship management as a business process. *Journal of Business & Industrial Marketing*.

Lin, C. C., Chen, S. C., & Chu, Y. M. (2011). Automatic price negotiation on the web: An agent-based web application using fuzzy expert system. *Expert Systems with Applications*, *38*(5), 5090-5100.

Mars, A., & Gouider, M. S. (2017). Big data analysis to Features Opinions Extraction of customer. *Procedia computer science*, *112*, 906-916.

Martínez, P., & Del Bosque, I. R. (2013). CSR and customer loyalty: The roles of trust, customer identification with the company and satisfaction. *International Journal of Hospitality Management*, *35*, 89-99.

Mullins, R. R., Ahearne, M., Lam, S. K., Hall, Z. R., & Boichuk, J. P. (2014). Know your customer: How salesperson perceptions of customer relationship quality form and influence account profitability. *Journal of Marketing*, *78*(6), 38-58.

Naik, A. R., & Keshavamurthy, B. N. (2019). Next level peer-to-peer overlay networks under high churns: a survey. *Peer-to-Peer Networking and Applications*, 1-27.

Nepal, B., Yadav, O. P., & Murat, A. (2010). A fuzzy-AHP approach to prioritization of CS attributes in target planning for automotive product development. *Expert Systems with Applications*, *37*(10), 6775-6786.

Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, *36*(2), 2592-2602.

Orman, L. V. (1998). A model management approach to business process reengineering. *Journal of Management Information Systems*, *15*(1), 187-212.

Pashchenko, D. (2020). Digitalization in Software Engineering and IT Business. *International Journal of Software Science and Computational Intelligence (IJSSCI)*, 12(2), 1-14.

Pourjavad, E., & Mayorga, R. V. (2019). A comparative study and measuring performance of manufacturing systems with Mamdani fuzzy inference system. *Journal of Intelligent Manufacturing*, *30*(3), 1085-1097.

Ranjan, J., & Bhatnagar, V. (2011). Role of knowledge management and analytical CRM in business: data mining based framework. *The Learning Organization*. *18*(2), 131-148.

Saaty, T. L., & Vargas, L. G. (2012). *Models, methods, concepts & applications of the analytic hierarchy process* (Vol. 175). Springer Science & Business Media.

Sawik, T. (2015). On the fair optimization of cost and customer service level in a supply chain under disruption risks. *Omega*, *53*, 58-66.

Shim, B., Choi, K., & Suh, Y. (2012). CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns. *Expert Systems with Applications*, *39*(9), 7736-7742.

Singh, R. K. (2013). Prioritizing the factors for coordinated supply chain using analytic hierarchy process (AHP). *Measuring Business Excellence, 17*(1), 80-97

Stormi, K., Lindholm, A., Laine, T., & Korhonen, T. (2019). RFM customer analysis for product-oriented services and service business development: an interventionist case study of two machinery manufacturers. *Journal of Management and Governance*, 1-31.

Sun, P. (2020). Research on cloud computing service based on trust access control. *International Journal of Engineering Business Management, 12*, 1847979019897444.

Tudor, A., Bara, A., & Botha, I. (2011). Data mining algorithms and techniques research in CRM systems. *Recent Researches in Computational Techniques, Non-Linear Systems and Control*, 265-269.

Wu, J., Guo, S., Huang, H., Liu, W., & Xiang, Y. (2018). Information and communications technologies for sustainable development goals: state-of-the-art, needs and perspectives. *IEEE Communications Surveys & Tutorials*, *20*(3), 2389-2406.

Wu, J., Guo, S., Li, J., & Zeng, D. (2016). Big data meet green challenges: Big data toward green applications. *IEEE Systems Journal*, *10*(3), 888-900.

Xiao, J., Jiang, X., He, C., & Teng, G. (2016). Churn prediction in customer relationship management via gmdh-based multiple classifiers ensemble. *IEEE Intelligent Systems*, *31*(2), 37-44.

Yadav, S., Desai, A., & Yadev, V. (2013). Knowledge management in CRM using data mining technique. *International Journal of Scientific & Engineering Research*, *4*(7), 465.

Yan, Y., Huang, C., Wang, Q., & Hu, B. (2020). Data mining of customer choice behavior in internet of things within relationship network. *International Journal of Information Management*, *50*, 566-574.