

Design of an Intelligent Customer Identification Model in e-Commerce Logistics Industry

C.C. Luk, K.L. Choy*, and H.Y. Lam

Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Abstract. The emergence of e-commerce in recent years has led to revolutionary changes in the logistics industry, as e-commerce relies heavily on efficient logistics to deliver the online goods to customers in a short period of time. Compared with traditional logistics, e-commerce orders, with a high variety of goods but small in quantity, are generally received from large number of customers worldwide. With a huge customer base, it is challenging for logistics service providers (LSPs) to provide satisfactory time-critical logistics services to meet the diversified customer requirements. In order to differentiate its services from others e-commerce LSPs, it is important to identify potential target groups of customers, and their behaviour so as to attract their attention. In this paper, an intelligent customer identification model (ICIM) is designed to support data analysis for managing customer relationships in a systematic way. The ICIM integrates the k-means clustering algorithm and the C4.5 classification algorithm in order to be able to deal with both continuous and discrete attributes for extracting valuable hidden knowledge. This effectively supports the identification of actual customer needs, and the classification of new customers in the future with minimum time for developing customer relationship management (CRM) recommendations to customers, thus improving business performance. Through a pilot study in a freight forwarding company in Hong Kong, it provides a real world demonstration and validation of data mining for CRM in the emerging e-commerce logistics industry.

1 Introduction

In the past two decades, the global e-commerce has started to become more and more popular [1]. Online shopping has become a modern trend around the world as it provides numerous benefits to customers in the form of the availability of foreign goods at lower cost and with wider choice. The rapid proliferation of the personal electronic devices, implementation of the internet and globalization are the key factors in driving the diffusion of e-commerce. Major advancements in online security assurance, and the commercial use of the internet and electronic communication systems also have contributed to the growth of e-commerce [2]. E-commerce is an international trading activity which heavily relies on logistics to ensure the online product is transported to the right customer's location at the right time. It also help in keeping pace with the demands of an increasingly world economy. Therefore, the boom in global business-to-business (B2B) and business-to-consumer (B2C) and business-to-business-to-consumer (B2B2C) e-commerce has led to an evolution of the retail model and brought revolutionary changes to the logistics industry [3]. The traditional logistics industry has had to change with it, with evolution of logistics practice, in e-commerce logistics.

Under the rapid growth in e-commerce logistics, many traditional logistics providers wish to enter the

profitable e-commerce logistics industry. As new entrants to e-commerce logistics industry, many of them have maintained their existing traditional logistics framework in the relatively new industry. Generally, e-commerce logistics has five characteristics, which are order fragmentation, low profit margin per each transaction, high volume, high product variety and time-critical logistics operation [4]. Therefore, it has a larger number of customers, both business customers and individuals, when compared with the traditional logistics industry. The need for managing customers becomes critical. To achieve the goal of managing the huge customer base in e-commerce logistics industry systematically, customer relationship management (CRM) is important supporting the business strategies for building long term, profitable relationships with specific customers, while maximizing the customer value to the organization with the help of business intelligence. Therefore, in this study, an intelligent customer identification model (ICIM) is designed to support data analysis for managing customer relationships in a systematic way.

This paper is organized as follows. Section 2 covers the past literature related to e-commerce logistics, CRM in e-commerce logistics, and data mining for CRM. Section 3 presents the design of the intelligent customer identification model (ICIM). Section 4 presents a case study to illustrate how the proposed model works,

* Corresponding author: kl.choy@polyu.edu.hk

followed by the results and discussion in Section 5. Finally, the conclusions are drawn in the Section 6.

2 LITERATURE REVIEW

2.1 Overview of E-Commerce Logistics

With the proliferation of personal computers and communication technology, and the rise in commercial use of the Internet, e-commerce has become the latest global retail model in recent years [5]. In 2017, worldwide retail e-commerce sales reached \$2.3 trillion US dollars and are expected to grow further over the years, reaching \$4.88 trillion US dollars in 2021 [6]. According to Laudon and Traver [7], e-commerce refers to commercial transactions between and among businesses and individuals using the Internet and the web. Among different types of e-commerce activities, B2B and B2C e-commerce concerns the direct transaction of goods and services through the electronic media between business establishments; between an organization and consumer respectively [8]. In the era of e-commerce, consumers are able to buy products from a manufacturer directly, and the products are then delivered to the end customers. It brings both new challenges and opportunities to logistics management as the cost of logistics and transportation has a large impact on a company's profitability. Under the e-commerce business environment, organizations are able to expand their marketplace greatly, as new channels are created for global marketing, hereby increasing their competitiveness internationally by connecting with trading partners to adopt just-in-time production and delivery [9][10]. Logistics has an essential role in e-commerce as the final trading heavily relies on logistics support, especially for the last mile delivery [11]. The role of e-commerce logistics can reduce the risk by ensuring the online ordered merchandize is delivered to the right place at the right time, and to specific customer in the global market [12].

2.2 Customer Relationship Management in E-commerce Logistics

Compared with traditional logistics, fragmented and high volumes of orders in e-commerce lead to a larger number of customers in e-commerce logistics. With the huge customer base, it is important to manage customers and relationships in a systematic way [13]. CRM is defined as a chain of processes and enabling systems for creating and maintaining relationships with customers, improving customer acquisition, customer retention, customer loyalty and customer profitability [14][15]. In general, a customer life cycle is divided into four stages including customer identification, customer attraction, customer retention and customer development [16]. Due to the high variance and wide range of customer orders and preferences in the e-commerce business, the prediction of customer behaviour becomes a strategically important and challenging issue [17]. By focusing on CRM in e-commerce logistics, companies are able to understand customer behaviour which is important in decision

making, and hence in improving competitiveness [18][19]. Pan et al. [20] predicted customer behaviour by using customer-related data to optimize e-grocery home delivery plans in order to reduce the rate of failed delivery. Wong and Wei [21] developed a customer online behaviour analysis tool to enhance customer relationships by segmenting high-value customers, analysing their online purchasing behaviour and finally predicting their next purchases.

2.3 Data Mining for CRM

In order to understand more about customers, data mining techniques have been adopted to discover the hidden customer information and behaviours of customers [22]. Data mining is a useful process for discovering knowledge and hidden patterns from the large data sets by the intersection of artificial intelligence, machine learning and statistics [23]. Clustering and decision trees are typical data mining tools in analysis.

Clustering is a ubiquitous tool for finding the intrinsic structure of the data and organizing data objects into groups, based on similarity [24]. It is an unsupervised classification for grouping data objects into meaningful clusters to determine hidden data concepts. It is important and useful in exploratory pattern-analysis, grouping, decision-making machine-learning situations and data mining [25]. Amine et al. [26] presented a clustering-based approach to evaluate customers' values of e-commerce websites in Morocco for developing effective marketing strategies. Cao et al. [27] analysed customers' purchase data and evaluated customers in terms of recency, frequency and monetary (RFM) variables based on ordered weighting averaging and the k-means cluster algorithm. Wang et al. [28] classified customers into clusters using k-means algorithm so as to formulate different services and marketing strategies to different types of customers.

The decision tree is another common tool in data mining and is mainly used for model classification and prediction. According to Tan et al. [29], classification is the task of learning a model that maps an input attribute set into its predefined class label. Classification using decision trees helps in discovering and deriving a set of models which distinguishes the data into classes, based on the analysis of a set of training data. It is supervised learning as the set of training data involves the class-labelled data objects. Ma [30] targeted improving customer loyalty by customer segmentation of e-commerce website using decision trees. Guo and Qin [31] analysed the features of customer churns in e-commerce based on the decision tree algorithm. Xu et al. [32] designed a decision tree-based clustering model for supporting mobile services in CRM.

In summary, it is found that e-commerce is an emerging industry and hence the applications of data mining for CRM have not been exploited in the industry. By integrating k-means clustering and the decision tree, a practical marketing strategy can be formulated by the e-commerce logistics service provider so as to provide customized services to different group of customers.

Since the clustered results vary, depending on the scale of the company, it can provide subjective segmentation which makes the data more applicable for decision tree analysis.

3. DESIGN OF AN INTELLIGENT CUSTOMER IDENTIFICATION MODEL (ICIM)

ICIM is proposed to improve the current CRM implementation in e-commerce logistics. The architecture of ICIM is shown in Figure 1. The system consists of three tiers which are (i) data management, (ii) customer classification analysis and (iii) CRM recommendation development.

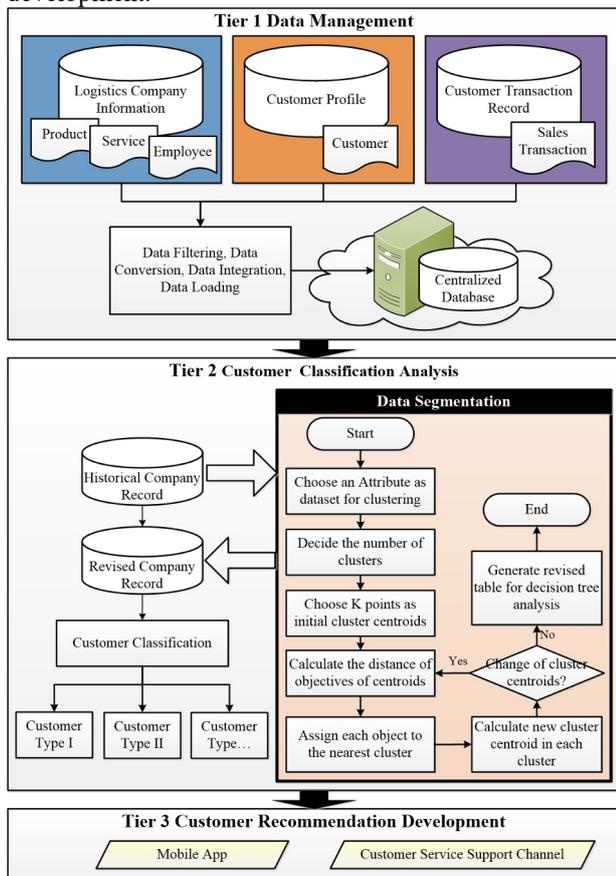


Fig. 1. Architecture of ICIM

3.1 Data Management

The existing problems of an e-commerce logistics company are first identified. Interviews were conducted with the company managers to gain the general understanding of company background, role, value, business nature, organizational culture, goal and operation. A review and evaluation of the current CRM implementation in e-commerce logistics should also be carried out in the interview. Meanwhile, the daily operation processes and constraints of operating and marketing e-commerce logistics should also be investigated. After identifying the problems of the company, three categories of data are collected for ICIM. They are logistics company information, customer

profiles and customer transaction records. Table 1 shows the sample data type required for data collection.

Table 1. Sample data type required for data collection.

Data Source	Data Type	
Logistics Company Information	<ul style="list-style-type: none"> Service ID Service Description Service Price Product ID Product Name Product Description Product Dimension 	<ul style="list-style-type: none"> Product Unit Price Available Quantity Employee ID Employee Name Position Year of Experience Contact Number
Customer Profiles	<ul style="list-style-type: none"> Customer ID Company Name Staff Name Contact Number Foundation Year Number of Staff Highest Order Value Last Order Date 	<ul style="list-style-type: none"> Experience in e-Commerce Monetary Value Per Month Transaction Frequency Per Month Types of service used Bad Debt China/Overseas
Customer Transaction Records	<ul style="list-style-type: none"> Order number Customer ID Order Date Service ID Delivery Date Product ID Order's Total Price Monetary Value Order Method 	<ul style="list-style-type: none"> Total Value of Transported Goods Unit of Products Sold Destination Country Types of Transported Good Employee ID Number of SKU in the order

3.2 Customer Classification Analysis

In the customer classification tier, two data mining techniques are selected and applied for studying the existing customers in extracting valuable customer knowledge. They are k-means clustering and the decision tree. As there are large sets of customer data in e-commerce logistics, the calculation processes of the two data mining techniques become complicated. The results of the data mining are analysed and used for further development of CRM recommendations in the company. Therefore, data mining is the major section in the methodology. Table 2 shows the notation for customer classification analysis.

Table 2. Notation table for customer classification analysis.

Notation	Definition
n	Total number of data objects in the dataset
x_e	Tata objects where $e = 1, 2, \dots, n$
k	Number of clusters
K_j	Clusters where $i = 1, 2, \dots, k$
m_j	Cluster centroids where $j = 1, 2, \dots, k$
$d(x_e, m_j)$	Euclidean distance between x_e and m_j
T	A set of training samples
$ T $	The total number of samples in T
T_1, T_2, \dots, T_n	Subsets of T
C_1, C_2, \dots, C_k	Possible classes
$freq(C_i, T)$	The total number of samples in T which belongs to class C_i

3.2.1 K-means Clustering for Segmentation

K-means clustering is used to divide data objects into k clusters by a distance measure, the Euclidean Distance. One data object belongs to only one cluster. In this paper, k-means clustering helps to cluster the large customer data into groups and extract the hidden customer data relationships, depending on the scale of the company. After generating the k-means clustering results, it also generates revised table with modified data, providing subjective segmentation which makes the data more applicable for decision tree analysis. For example, a group of customers with high monetary value per month in the company can be identified. Even if a customer spent the same amount of money in two different companies in the same month, it may mean different things to the two companies as their business models and scales are different. The clustered results are varied depending on the scale of the company. Euclidean distance, as shown in Eq. (1) and (2), is applied to calculate the distance between each object to the cluster centroid.

$$d(x_e, m_j) = \sqrt{(x_e - m_j)^2} = |x_e - m_j| \quad (1)$$

$$m_j = \frac{\text{total value of data objects in the cluster}}{\text{total number of data objects in the cluster}} \quad (2)$$

3.2.2 Decision Tree for Classification

The decision tree consists of rules, and contributes in the classification and prediction of decision support. Hence, relevant historical data in the database and the revised table generated in the previous stage are used to construct a decision tree for the C4.5 Algorithm. The existing customers have been classified into different classes, and are class-labelled data objects. The C4.5 algorithm is used to learn the classification model which distinguishes the customer data into classes based on the analysis of a set of training data. There are eight steps in constructing a decision tree.

Step 1. Identify training set T , non-categorical and categorical attributes, total number of samples ($|T|$), the class label (C_1, C_2, \dots, C_k) based on the historical data in the database and the revised table. Count the number of samples in each class. It represents the data preparation before the calculation.

Step 2. Calculate $Info(T)$, which is the entropy of T by equation (3) to get the information content before splitting. Also, identify the leaf nodes (outcomes) of each possible attribute and calculate the $Info_x(T)$ for each possible attribute, which is the entropy after T has been split according to the attribute in equation (4).

$$Info(T) = -\sum_{i=1}^k \frac{freq(C_i, T)}{|T|} \log_2 \left(\frac{freq(C_i, T)}{|T|} \right) \quad (3)$$

$$Info_x(T) = \sum_{i=1}^n \frac{|T_{il}|}{|T|} Info(T_i) \quad (4)$$

Step 3. Calculate the information gain of each possible attribute using equation (5).

$$Gain(x) = Info(T) - Info_x(T) \quad (5)$$

Step 4. Calculate the gain ratio of each possible attribute using equations (6) and (7).

$$\text{Split-Info}(x) = -\sum_{i=1}^n \frac{|T_{il}|}{|T|} \log_2 \left(\frac{|T_{il}|}{|T|} \right) \quad (6)$$

$$\text{Gain-Ratio}(x) = \frac{Gain(x)}{\text{Split-info}(x)} \quad (7)$$

Step 5. Compare the gain ratio for each possible attribute and select the attribute with the highest gain ratio as the split (node attribute) in the construction of the decision tree.

Step 6. Repeat Steps 1-5 for the remaining node(s) which has no attribute for classification.

Step 7. Develop a decision tree when there is no more unclassified node.

Step 8. Evaluate and validate the decision tree by using a set of testing data for ensuring accuracy.

After implementing the above steps, the derived model is represented in a flow-chart-like tree structure. It can be easily converted into rules for classification. Therefore, the company can predict or assign a class label to new customers according to the derived decision tree or the classification rules. When customers can be identified, corresponding customized marketing plans are developed for each customer segment, based upon the result of customer analysis.

3.3 CRM Recommendation Development

In this study, ICIM is designed to improve the CRM in the e-commerce logistics. Effective interactions with customers can provide better insight into customers' needs. Companies should emphasize and improve the effectiveness of the interactions with their customers. In order to manage the relationships with customers, it is essential to formulate a cost-efficient approach for interacting and communicating with customers. The efforts put into the interactions help in building close relationships with customers. Furthermore, the logistics company should customize and tailor its e-commerce logistics services and marketing campaigns based on the needs and values of different customers. The company can put more resources and efforts to customers who are identified being valuable customers, and serve them in more profitable ways.

4 CASE STUDY

ABC Company is a freight forwarding company, which is a traditional logistics service provider. It specializes in handling the transportation of time-critical air and sea freight. It offers professional logistics services to eight different industrial business segments, which are aviation, automotive, chemical, healthcare, project cargo, technology, retail and fashion. It aims at providing customized logistics solutions to fit different needs of different sizes and types of customers.

Nowadays, there is rapid growth in both e-commerce and e-commerce logistics. To cope with the increasing trend of e-commerce, ABC Company established a new

division of e-commerce logistics in mid June in 2017, providing cross-border parcel delivery services, and developing a global e-commerce logistics business. It provides several types e-commerce logistics services to meet the newly increasing demand of customers. ABC Company is willing to offer a set of one-stop, customized, comprehensive e-commerce logistics solutions for supporting a customer's global business, with total monitoring. As Hong Kong is a logistics hub and the gateway to China, possesses many advantages in bringing foreign goods into the mainland market, it has lead to an increase of e-commerce logistics customers in ABC Company.

In order to enter the e-commerce market, ICIM was proposed to improve the current CRM implementation of e-commerce logistics in the case company. Fig. 2 shows the implementation procedures of ICIM, which are divided into three stages: data collection, data pre-processing, data mining.

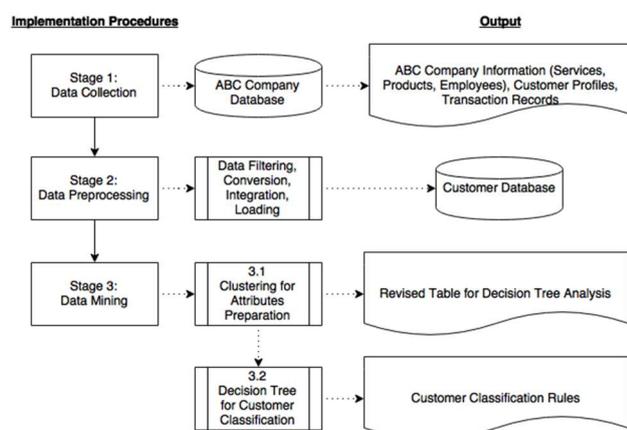


Fig. 2. Implementation procedures of ICIM in ABC Company

4.1 Stage 1: Data Collection

The data used in the case study covered the period June to December 2017, and was provided by ABC Company. It was collected by visiting the company in the beginning of 2018. Three types of data collected for ICIM and originally stored in its cloud database, including company information, customer profiles, and customer transaction records, as inputted by various departments. Company information includes the details of ABC Company's services (service ID, service name, service price), products (product ID, product name, product description, product dimension, unit price and quantity), and employee (employee ID, employee name, position, year of experience). A customer profile is created when each new customer makes an e-commerce logistics order in ABC Company. The customer profile includes data on customer ID, company name, name of in-charge staff, contact number, foundation year, etc. A total of fifty e-commerce logistics customers were extracted from the original company database as the preliminary data for ICIM. Customer transaction records include data on order number, customer ID, order date, order total price and so on. ABC Company's staff members key in the transaction data into the structured database immediately after each

single transaction is completed. The transaction records from June 2017 to December 2017 were used in the case study.

4.2 Stage 2: Data Pre-processing

Data pre-processing in Stage 2 was to prepare data and ensure the data quality for the data mining analysis in Stage 3. After data collection, those repeated, missing and useless data was pre-processed by data filtering, data conversion, data integration and data loading before storing into the e-commerce logistics customer database. As there may be missing values and double data entries, only the relevant e-commerce logistics customers' data were used in the case study. The pre-processed data were stored in a new structured database.

4.3 Stage 3: Data Mining

The software, WEKA was adopted to demonstrate the data mining in the case study. The main purpose of the data mining was to study the existing customers and extract the hidden valuable customer knowledge. Two data mining techniques, k-means clustering and decision tree were applied in this stage for clustering customers, identifying valuable customers and finding out customer classification rules. In this stage, some useful historical data in the customer database were retrieved to cluster the customers, and ultimately build a well-defined decision tree for classifying new customers in the future with minimum time and human resources and further developing CRM recommendations to customers. Table 3 shows the extracted customer data attributes.

4.3.1 K-means Clustering for Segmentation

Before building a decision tree, k-means clustering was applied for clustering customers in the selected data attributes, subjective segmentation which makes the data more applicable for the decision tree analysis in the next part. In this case, three data attributes were selected to be further processed by k-means clustering as the clustered results in these attributes may be varied depends on the scale of the company. They were transaction frequency per month, monetary value per month and number of SKU in orders. K-means clustering helps to divide the existing e-commerce logistics customers who had similar values of transaction frequency per month, monetary value per month and number of SKU in the orders, into 3 groups respectively. For the data attributes of transaction frequency per month, customers were divided into three groups: rare (K_0), occasional (K_1) and frequent (K_2). For monetary value per month, customers were divided into uncertain (K_0), good (K_1) and best (K_2). For the number of SKU in orders, customers were divided into small (K_0), medium (K_1) and large (K_2). The clustering results were generated by using WEKA. Based on the output, the final cluster centroids of M_0 , M_1 , M_2 are 6.5714, 21.3 and 37 respectively. 21 customers were clustered as rare (K_0), 20 customers occasional (K_1), 9 customers frequent (K_2).

Table 3. Extracted customer data attributes

Customer ID	Types of Transported Goods	Experience in e-Commerce	Chain/Overseas	Bad Debt	Transaction Frequency per Month	Monetary Value per Month	Number of SKU in Orders	Company Decision
1	Toys	No	Overseas	No	2	65	5	Non-valuable
2	Food	Yes	China	Yes	4	145	7	Non-valuable
3	Food	No	China	No	5	187	25	Non-valuable
4	Cosmetic	Yes	Overseas	No	20	855	43	Premium
5	Furnishing	No	Overseas	Yes	8	677	32	Normal
6	Electronic Devices	Yes	China	Yes	24	977	149	Premium
7	Food	Yes	China	No	12	665	30	Normal
8	Cosmetic	Yes	China	No	25	1400	189	Premium
9	Clothes	Yes	Overseas	Yes	38	2500	350	VIP
10	Clothes	Yes	China	Yes	47	3100	378	VIP

The visualized result of clustering in transaction frequency per month is displayed in Fig. 3. The X-axis represents the value of transaction frequency per month, and the Y-axis represents the Customer ID. Fig. 4 shows the partial modification of data in the transaction frequency per month, before and after applying k-means clustering. The clustering results were generated by using WEKA. Based on the output, the final cluster centroids of M_0 , M_1 , M_2 are 6.5714, 21.3 and 37 respectively. 21 customers were clustered as rare (K_0), 20 customers occasional (K_1), 9 customers frequent (K_2). The visualized result of clustering in transaction frequency per month is displayed in Fig. 3. The X-axis represents the value of transaction frequency per month, and the Y-axis represents the Customer ID. Fig. 4 shows the partial modification of data in the transaction frequency per month, before and after applying k-means clustering.

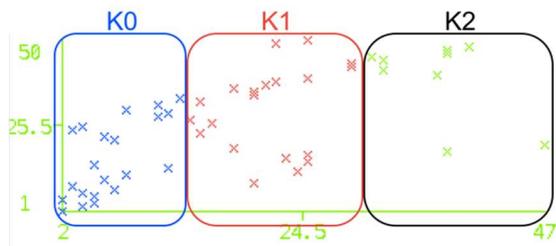


Fig. 3. The visualized result of clustering in transaction frequency per month

Before clustering:				After clustering:			
Customer ID	Transaction Frequency Per Month	Monetary Value Per Month	Number of SKU in orders	Customer ID	Transaction Frequency Per Month	Monetary Value Per Month	Number of SKU in orders
1	2	65	5	1	Rare	Uncertain	Small
2	4	145	7	2	Rare	Uncertain	Small
3	5	187	25	3	Rare	Uncertain	Small
4	20	855	43	4	Occasional	Good	Small
5	8	677	32	5	Rare	Good	Small
6	24	977	149	6	Occasional	Good	Medium
7	12	665	30	7	Rare	Good	Small
8	25	1400	189	8	Occasional	Good	Medium
9	38	2500	350	9	Frequent	Best	Large
10	47	3100	378	10	Frequent	Best	Large

Fig. 4. The partial modification of data before and after applying k-means clustering

The clustering result for the three data attributes is summarized in Table 4. For the monetary value per month, 14 customers were clustered as uncertain (K_0), 23 customers good (K_1) and 13 customers best (K_2). For the number of SKU in the orders, 26 customers were clustered as small (K_0), 16 customers medium (K_1) and 8 customers

large (K_2). After clustering fifty customers in the three selected attributes one by one, a revised table with the modified data was generated.

Table 4. Number of customers in each cluster in each selected attribute

Cluster	Transaction Frequency Per Month	Monetary Value Per Month	No. of SKU in orders
K_0	21	14	26
K_1	20	23	16
K_2	9	13	8

4.3.2 Decision Tree

In this case study, the decision tree was used for discovering and deriving a set of models which distinguished the customers into four classes based on the analysis of a set of data from fifty customers. Customer classification rules can also be generated. With the customer classification rules, ABC Company can classify e-commerce logistics new comers with minimum time and human resources in the future. According to the historical company decisions which were made by managers, existing customers were classified into four different groups, including non-valuable, normal, premium, VIP. In the view of ABC Company, premium and VIP customers are the most valued. Seven data attributes of the types of transported goods, experience in e-commerce, China/ overseas, bad debts, transaction frequency per month, monetary value per month and number of SKU in the orders were selected in the decision tree analysis for classifying the existing e-commerce logistics customers.

WEKA was implemented to handle large data for constructing the decision tree. The classifier, J48 was chosen to perform the above calculation process. With 10-fold cross-validation, the dataset was divided into 10 groups. The training data is an induction for learning algorithm and trying to determine the target attribute value of new comers. The ultimate decision tree is pruned to reduce the prediction error. The minimum number of instances per leaf was two to ensure no splitting if the nodes became very small. The gain ratio was calculated for those seven attributes in WEKA. By comparing the gain ratio for the seven attributes, the attribute of the monetary value per month had the highest gain ratio, the

best differentiates sample data, is selected as the first split in the construction of the decision tree. There are three branches {Uncertain, Good, Best} created by the split. The attribute with the largest gain ratio was selected as optimal solution. In WEKA, the entire calculation process of optimization was repeated for each child node. The result of decision tree is shown in Fig. 5. From the result, the correct classification is higher than 80%, which is acceptable. The visualized decision tree and customer classification rules are generated, as shown in Fig. 6.

```

Number of Leaves : 6
Size of the tree : 9

Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 42 84 %
Incorrectly Classified Instances 8 16 %
Kappa statistic 0.7872
Mean absolute error 0.1868
Root mean squared error 0.26
Relative absolute error 28.3813 %
Root relative squared error 59.884 %
Total Number of Instances 50

=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
1.000 0.953 0.857 1.000 0.923 0.901 0.958 0.788 Non-valuable
0.917 0.132 0.688 0.917 0.786 0.719 0.921 0.658 Premium
0.462 0.027 0.857 0.462 0.600 0.549 0.723 0.641 Normal
1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 VIP
Weighted Avg. 0.840 0.051 0.854 0.840 0.826 0.792 0.897 0.772

=== Confusion Matrix ===
a b c d ←- classified as
12 0 0 0 | a = Non-valuable
0 11 1 0 | b = Premium
2 5 6 0 | c = Normal
0 0 0 13 | d = VIP
    
```

Fig. 5. Result of J48 algorithm in decision tree generation

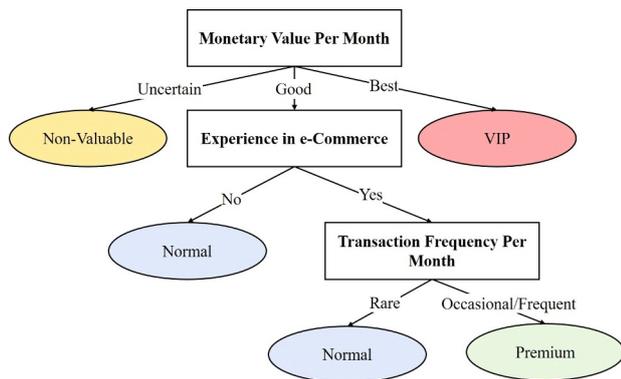


Fig. 6. The visualized decision tree for customer classification in ABC Company

4.4 CRM Recommendation Development

In the view of ABC Company, premium and VIP customers are more valued, having higher monetary value or transaction frequency per month, so the company wishes to establish good relationships with these valued customers. As ABC is a new entrant to e-commerce logistics, and is also an SME, having limited investment in CRM systems, it did not have a practical and effective CRM system to support the building of a close relationships with different customer segments. According to the customer classification results, customized service plans were developed for each customer segment in ABC Company for attracting and building close relationships. Table 5 shows the customized services for each customer segment, in the areas of mobile apps and customer service support channels.

Table 5. The customized services for each customer segment

Customer Segment	Mobile App	Customer Service Support Channel
Non-valuable	✓ General functions	Office hours: ✓ CS Hotline ✓ Email
Normal	✓ General functions	Office hours: ✓ CS Hotline ✓ Email
Premium	✓ General functions ✓ Online ordering system ✓ Auto reminder for every completed delivery ✓ Online chatroom	Office hours: ✓ CS Hotline ✓ Email ✓ Online chatroom
VIP	✓ General functions ✓ Online ordering system ✓ Auto reminder for every completed delivery ✓ Online chatroom	Office hours: ✓ CS Hotline ✓ Email ✓ Online chatroom 24/7: ✓ Dedicated Emergency Team

5. RESULTS AND DISCUSSION

To evaluate the performance of ICIM in ABC Company, surveys were conducted online with those 50 e-commerce logistics customers, before and after the implementation of ICIM respectively. The main purpose was to collect information on how the customers experienced ABC Company's e-commerce logistics, including their satisfaction level with the service plans, willingness of establishing close relationship with ABC Company and likely future order behaviour, before and after the implementation. The survey invitation was sent to those fifty e-commerce logistics customer through email, with an overall response rate of 90% in both surveys. The results and comparison in the two online surveys are summarized in Table 6. It shows the implementation of ICIM led to three improvements in the case company: (i) customer satisfaction enhancement, (ii) close relationship establishment and (iii) future order behaviour improvement.

Table 6. Results and comparison in the two online surveys

	Before	After	Percentage Change
Overall Satisfaction	55%	75%	+36.4%
Willing to Establish Close Relationship	20	30	+50%
Expected Order Frequency (per month)	5	8	+60%
Expected Order Spending Amount (per month)	HKD 500	HKD 2000	+300%

(i) Customer Satisfaction Enhancement

After adopting five area of customized service plans on each customer segment, there was a 36.4% increase in the overall satisfaction, including their satisfaction level with the service plans.

(ii) Close Relationship Establishment

After providing customized service plans to the corresponding customer segments, more existing customers were satisfied with the services and were willing to establish a closer relationship with ABC Company. The number of customers who were willing to establish a close relationship increased from 20 to 30 customers, which has a 50% increase.

(iii) Future Order Behaviour Improvement

Before implementing ICIM, existing customers expected to place 5 e-commerce logistics orders per month. Now, they expect to place 8 orders per month, with a 60% increase in the expected order frequency. Also, there was a 300% increase in the expected order spending amount. It shows existing customers are willing to order and spend more in e-commerce logistics services in the future. Their intention for future ordering behaviour has improved.

6. CONCLUSIONS

The logistics industry has seen revolutionary changes as e-commerce has become heavily reliant on logistics to deliver the online ordered merchandize to customers in a certain period of time. In this paper, an intelligent customer identification model (ICIM) is presented for e-commerce logistics businesses to manage the large customer bases, and build long-term and profitable relationships. The ICIM consists of a historical view and analysis of all the existing or potential customers. This supports the effective identification of actual customer needs, classification of new customers in the future with minimum time and human resources, and the development of CRM recommendations to customers, thus improving business performance.

The authors would like to thank the Engineering Doctorate Programme, Faculty of Engineering of The Hong Kong Polytechnic University for inspiring the development of this project.

References

1. M. Falk, E. Hagsten, *Int. J. Prod. Econ.*, **170**, 357-369 (2015)
2. R. Rahayu, J. Day, *Procedia Soc Behav Sci.*, **195**, 142-150 (2015).
3. Y. Yu, X. Wang, R.Y. Zhong, G.Q. Huang, *Proc. CIRP*, **52**, 179-185 (2016).
4. B. Xiao, I. Benbasat, *Handbook of Strategic e-Business Management* (pp. 403-431). Springer, Berlin, Heidelberg (2014).
5. eMarketer, Retail e-commerce sales worldwide from 2014 to 2021 (in billion U.S. dollars). In *Statista - The Statistics Portal*. Available at <https://www.statista.com/statistics/379046/worldwid-e-retail-e-commerce-sales/>
6. X. Chen, H. Lin, *Procedia Soc Behav Sci.*, **96**, 838-843 (2013).
7. K.C. Laudon, C.G. Traver, *E-commerce*. Pearson (2013).
8. R. Mangiaracina, G. Marchet, S. Perotti, A. Tumino, *Int. J. Phys. Distrib. Logist. Manag.*, **45**(6), 565-591 (2015).
9. S. Yin, L. Wang, J. Yang, *Int. J. Mob. Commun.*, **14**(4), 328-347 (2016).
10. T. Bui, I. Gavriliv, *Scientific Papers NaUKMA. Economics*, **172**, 8-14 (2016).
11. S.X. Xu, M. Cheng, G.Q. Huang, *Transp. Res. Pt. B-Methodol.*, **80**, 322-337 (2015).
12. Y. Yu, X. Wang, R.Y. Zhong, G.Q. Huang, *Ind. Manage. Data Syst.*, **117**(10), 2263-2286 (2017).
13. J. Bhattacharjya, A. Ellison, S. Tripathi, *Int. J. Phys. Distrib. Logist. Manag.*, **46**(6/7), 659-680 (2016).
14. E.W. Ngai, L. Xiu, D.C. Chau, *Expert Syst. Appl.*, **36**(2), 2592-2602 (2009).
15. R.S. Hassan, A. Nawaz, M.N. Lashari, F. Zafar, *Procedia Econ. Financ.*, **23**, 563-567 (2015).
16. N. Hill, J. Alexander, *The handbook of customer satisfaction and loyalty measurement*. Routledge (2017).
17. P.A. Sarvari, A. Ustundag, H. Takci, *Kybernetes*, **45**(7), 1129-1157 (2016).
18. S. Akter, S.F. Wamba, *Electron. Mark.*, **26**(2), 173-194 (2016).
19. I. Gamayanto, H. Christian, *Binus Bus. Rev.*, **9**(1), 29-38 (2018).
20. S. Pan, V. Giannikas, Y. Han, E. Grover-Silva, B. Qiao, *Ind. Manage. Data Syst.*, **117**(9), 1917-1933 (2017).
21. E. Wong, Y. Wei, *Int. J. Retail Distrib. Manag.*, **46**(4), 406-420 (2018).
22. I.T. Afolabi, O. Oladipupo, R.E. Worlu, I.O. Akinyemi, *Covenant Journal of Business and Social Sciences*, **7**(1), 41-60. (2016)
23. I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann (2016).
24. B. Liu, Y. Xia, P.S. Yu, *Foundations and advances in data mining* (pp. 97-124). Springer Berlin Heidelberg (2005).
25. X. Wu, X. Zhu, G.Q. Wu, W. Ding, *IEEE Trans. Knowl. Data Eng.*, **26**(1), 97-107 (2014).
26. A. Amine, B. Bouikhalene, R. Lbibb, R. International Journal of Computer, Electrical, Automation, Control and Information Engineering, **9**(8), 2000-2010 (2015).
27. J. Cao, X. Yu, Z. Zhang, *J. Syst. Sci. Complex.*, **28**(2), 381-392 (2015).
28. X. Wang, W. Qu, J. Dang, *ICIDEL 2017*, pp. 162-166 (2017).
29. P. N. Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining*, **1**, 145-205 (2006).
30. H. Ma, *American Journal of Industrial and Business Management*, **5**(12), 813-818 (2015).
31. F. Guo, H.L. Qin. 2015 International Conference on Computer Science and Applications (CSA), Wuhan, China, pp. 199-203 (2015).
32. L. Xu, H.C. Chu, *J. Comput.*, **26**(4), pp. 63-72 (2016)