Asia Pacific Journal of Information Systems
Vol. 32 No. 3 (September 2022), 496-513

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2022.32.3.496

Application of AI-based Customer Segmentation in the Insurance Industry

Kyeongmin Yum[a], Byungjoon Yoo[b], Jaehwan Lee[c,*]

[a] *Ph.D. Candidate, College of Business Administration, Seoul National University, Korea*
[b] *Professor, College of Business Administration, Seoul National University, Korea*
[c] *Ph.D., Institute of Management Research, Seoul National University, Korea*

A B S T R A C T

Artificial intelligence or big data technologies can benefit finance companies such as those in the insurance sector. With artificial intelligence, companies can develop better customer segmentation methods and eventually improve the quality of customer relationship management. However, the application of AI-based customer segmentation in the insurance industry seems to have been unsuccessful. Findings from our interviews with sales agents and customer service managers indicate that current customer segmentation in the Korean insurance company relies upon individual agents' heuristic decisions rather than a generalizable data-based method. We propose guidelines for AI-based customer segmentation for the insurance industry, based on the CRISP-DM standard data mining project framework. Our proposed guideline provides new insights for studies on AI-based technology implementation and has practical implications for companies that deploy algorithm-based customer relationship management systems.

Keywords: Customer Segmentation, Artificial Intelligence, Interview, Guideline Proposal

Ⅰ. Introduction

The finance industry is well known for its reliance on digital technology for efficient operation, new product development, and customer management. By digitalizing most of its operations, big data can be acquired more effortlessly than ever, allowing companies to apply artificial intelligence technology to increase competitiveness. AI provides businesses with unprecedented opportunities in the insurance sector. According to a review study by Boodhun (2017), three different data analytical approaches have been conducted in the insurance industry: Fraud Detection, Risk Prediction, and Customer Analytics. Customer analytics aims to predict what customers want from the insurance company by studying their behavior.

*Corresponding Author. E-mail: jaehwan@snu.ac.kr

Since Customer Relationship Management (CRM) is crucial in the insurance industry, it is expected to be one of the most digitalized and active in using state-of-the-art Artificial Intelligence (AI) technology. However, simply applying AI technology to business does not always guarantee success. According to Deloitte's (2018) report, insurance companies face various challenges in the middle of data analytics solutions. These challenges include a gap between data expertise and business sense, the undefined value of the data project, and unclear maintenance. For driving successful data analytics solutions or AI-based projects, it is necessary to investigate the current situation and develop a solid process guideline. By conducting interviews with customer service managers/sales agents working on the front lines of Korea's insurance field, we have found that insurance companies depend upon individual agent's heuristic-based customer segmentation standards rather than segmentation driven by the legacy data-based approach.

The success or failure of a data mining project often depends on the engaged person. Even if the project's goals are achieved, successful practices are rarely replicated or shared across the company. A standard approach specialized in these tasks will assist with problem description, data transformation, and model evaluation (Wirth and Hipp, 2000). Therefore, we propose an application guideline for AI-based customer segmentation in the insurance sector. The primary objective of our study is to inform IS researchers and practitioners on how to carry out AI-based customer segmentation in the insurance industry. We develop guidelines for preparing, conducting, and evaluating AI-based customer segmentation.

The remainder of this paper is organized as follows. Chapter 2 consists of a literature review of research related to the use of a large set of CRM data in different finance domains and theoretical factors that may affect customers' financial decisions. In Chapter 3, we present the results of interviews on the ways in which Korean insurance companies manage and classify their customers. In Chapter 4, concepts and examples about customer segmentation models and application practice are proposed. Finally, in Chapter 5, we suggest directions for future research and conclude by discussing the implications of our results.

Ⅱ. Related Work

This section reviews the literature on customer segmentation and financial decision-making.

2.1. Customer Segmentation

Customer segmentation has been studied extensively in customer relationship management in the finance industry. One example is Namvar et al. (2010) structured framework. Their model was developed to apply Recency, Frequency, and Monetary (RFM) and customers' Lifetime Value (LTV) into customer segmentation models. Their work utilizes demographic data from the banking industry. K-Means clustering technique and a self-organizing map method are applied.

Khajvand and Tarokh (2011) proposed a customer segmentation framework based on a banking service customer's LTV. Prior studies of customer segments were based upon customers' requirements or preferences; however, different approaches utilizing customers' lifetime value resulted in more efficient factors. The framework segments the customers and calculates each segment's lifetime value, estimating each element's future value based upon customer transaction data. K-means and two-step clustering algorithms were implemented for the model. The

RFM model was utilized to calculate the customer's lifetime value. They also used a time series method to predict the future value for each customer segment.

Baradaran et al. (2011) conducted bank customer segmentation based on customer behavior. Their goal was to provide retention strategies and attract new customers. They utilized customers' demographic information, transaction records, and bank card information. K-means clustering techniques were used to segment customers.

Goonetilleke and Caldera (2013) used insurance company-driven CRM data to analyze the customer and avoid churn. The authors used an extended amount of demographic data such as term of policy, sum assured, premium, and agent. Decision tree, neural network, and logistic regression were used for segmentation. Models generated were evaluated using receiver operating characteristic curves and area under the curve values.

Qadadeh and Abdallah (2018) utilized K-Means and Self-Organized Map (SOM) method with The Insurance Company (TIC) dataset. Their results showed that SOM had outperformed K-Means in speed, quality, clustering, and visualization. The authors argue that customer segmentation based upon demographic or behavioral data will lead to more effective marketing strategies and campaigns.

Zhuang et al. (2018) argued that decision-making based on customer segmentation results in the auto insurance field has been failing. They indicate that utilizing a single algorithm for segmentation leads to lacked validation. To resolve this issue, the authors propose customer segmentation results based upon mixed-type data clustering algorithms (k-prototypes, improved k-prototypes, and similarity-based agglomerative clustering) that complement each other.

Mihova and Pavlov (2018) utilized data set of borrowers from a commercial bank to conduct customer segmentation. The study applied three features (loan amount, business relationship length with the bank, number of missed payments in recent 12 months, and utilized the k-means algorithm for analysis (initial variable, standardized variable, and two-step clustering applied).

Bansal and Shukla (2020) presented a workflow for analyzing online insurance businesses. The authors emphasize the importance of conducting customer segmentation and propose a flow of customer segmentation based on the k-means algorithm.

Nandapala et al. (2020) proposed a new micro-segmentation process by applying demographic segmentation based on RFM analysis. The authors argue that their new customer segmentation method provides deeper insight into understanding customer behavior in the insurance industry.

Guo et al. (2021) presented the customer segmentation concept based upon customer loyalty and contribution to the company instead of prevalent product-oriented segmentation. The authors present a model concept based on the k-means algorithm.

Abdul-Rahman et al. (2021) conducted customer segmentation with K-Modes Clustering and Decision Tree Classifier method utilizing customer data from a Malaysian insurance company.

Khamesian et al. (2021) utilized data set of Iranian insurance company customers to conduct customer segmentation. The authors applied five features (age, gender, third party liability coverage limit, premium paid for car insurance other than third party liability, and premium for life insurance) to the k-means algorithm for analysis.

Kumar and Oommen Philip (2022) presented customer segmentation results that differ from traditional studies by analyzing B2B insurance client data. The traditional RFM analysis considered customers as individuals. However, the authors included organizations as the customer in their approach. They conduct clustering analysis using the K-Means algorithm.

In the insurance industry, rapid environmental change is common, and customers expect better service, which naturally creates fierce competition among companies to improve the customer experience. Applying the AI-based customer segmentation method can be an ideal solution. However, most prior studies are limited to the banking sector or examinations limited to customer attrition. Studies regarding insurance sectors are primarily from the non-business field, which focuses on model performance and the algorithm itself. Those results emphasizing clustering algorithm leads to less consideration in the business context.

Others approach classifying customers based on the likelihood of churn. Assessing the reason why customers churn may be a necessary business objective. However, a broad universal customer segmentation method that understands customers is needed to increase customer satisfaction and bring in new customers.

Therefore, insurance companies strive to perform customer segmentation. Their most common criteria are demographic information such as customer gender, age, occupation, and the composition and size of their insurance products. However, unlike AI, this method is outdated.

2.2. Financial Decision

Several theoretical factors, especially the selection of insurance services may comprise customers' financial decisions. Representatively: expected return and risks, experiential knowledge, informativeness, reliability, and degree of involvement can be considered significant factors. The following theories and concepts constitute consumer personalities in the insurance sector.

2.2.1. Expected Return and Risks

Consumers' choice of financial investment products is intended to maximize profits and minimize losses (Zhou and Pham, 2004). Therefore, expected returns and risks are critical factors in the choice of insurance products (Markowitz, 1952). Since it is necessary to select an optimal investment alternative that considers both the expected return and risk of the choice to establish a financial product, observing how customers perceive and understand the relationship between expected return and risk and utilizing it in decision-making is essential (Ha and Kim, 2011; Huh et al., 2010).

Risk refers to uncertainty caused by expected returns from financial products subject to investment. Because individual investors base their investment decisions on only partial knowledge, not expertise, purchasing the financial product is a decision made under risk (Huh and Yoo, 2012). Ganzach (2000) defines risk as a key characteristic of alternatives with uncertain outcomes. Ganzach added that an individual's perception of risk is one of the most important features to consider when assessing an individual's choices in financial decision-making, recruitment, and accepting new technologies.

Prior studies related to risk focus on risk aversion and wealth. According to Campbell (2006), individuals with more wealth were more risk-averse, while those with less wealth showed were more risk-averse. Warner and Cramer (1995) also suggested that risk-seeking tendencies become stronger as asset levels increase. Bertaut and Starr-McCluer (2000) reported, based on their empirical results, that the top 1% of income groups are investing in high-risk assets. Similar results were found in the Korean consumer sample (Huh et al., 2010; Joo, 2008).

2.2.2. Experiential Knowledge

Experiential knowledge and investment experience can influence current or future investment (Andersen, 1993). For example, if a customer has already received benefits from insurance products, that customer can rely upon other similar products based upon that experience. Johanson and Vahlne (1977) also argued that current market decisions depend on previous investment experience knowledge. Min and Song (2014) found that experience is the primary variable used in allocating financial assets. It is also important to note that experience can improve the understanding of the financial product (Choe and Cho, 2011).

2.2.3. Informativeness

Informativeness is defined as the degree of information provided by companies in the financial sector (Joo, 2011). Information asymmetry occurs in financial transactions, including those that involve insurance products, and these characteristics can serve as a significant factor in the customer segment. Since financial products are the combined result of both product and service attributes, these factors can play an important role in investment decision making, and consequently, consumers' information search behavior when purchasing finance products may differ from consumers' information search behavior during general product purchase (Choi et al., 2016). In addition, because financial products are intangible, consumers cannot immediately experience their value because they must judge and determine products based on given information. This may affect consumer characteristics from a long-term perspective.

Financial products such as insurance are complex. Therefore, customers are expected to have asymmetric information. Those with an elementary knowledge about finance will achieve information dominance and choose the product they need. However, if customers lack this knowledge, they will make investments based on instinct or on limited information. According to Joo's (2011) experimental result, consumers over 50 years old tended to rely on information from financial companies or sales associates when deciding whether or not to purchase financial products. At the same time, they showed less baseline understanding than younger investors about the fund investment. Therefore, consumers in 50s and older may be considered a vulnerable group in financial product purchases, indicating that informativeness may be a critical feature of customer segmentation regarding insurance products. Mishra and Kumar (2011) claim that more knowledge investors have, the more likely they are to show more information search behavior, culminating in increased investment activity. This result may indicate that prior information is critical in a customer's financial activity.

2.2.4. Reliability

Even within the service industry, the financial industry recognizes the importance of customers, perceives customer service as a core competency of the industry itself, and recognizes the significant value of customer satisfaction. In addition, the quality of service for customer satisfaction dramatically influences reliability. Service quality should be consistent even during fierce competition, and when the company is in a critical situation (Lee, 2012). Reliability is essential because the public trusts companies with good reputations (Lyon and Cameron, 2004). Anderson (1982) provided evidence that the importance of the head office is essential when consumers decide which finance branch to choose. Therefore, tangible products and intangible services significantly influence the reputation of financial companies. Naturally, consumers' psychological attributes related to intangible services in the financial sector are used in the development of service models (Hur and Kim, 2001). Several studies have demonstrated that the reliability of a financial institution, meaning its image or reputation, affects consumers' behavior. Therefore, it can be as a significant variable in customer segmentation.

2.2.5. Degree of Involvement

The market can be classified according to the degree of consumers' involvement in the purchase decision. Ha (1996) organized bank customers into three groups based on transaction amount and desire for personalized services. The first group is customers with large amounts of transactions who want highly personalized services. Members of the second group have smaller transactions and prefer convenience to personalization. The last group is somewhere in the middle; their transactions are moderate in size, and they seek both personalized services and the convenience of using financial institutions.

Bae (2004) examined financial services that segment consumers and provide different marketing strategies from a corporate perspective. According to the study, banks classify their customers with deposits, loans, credit card use, automatic transfers, electronic finance, foreign exchange, and length of business relationship. The banks provide differentiated services based on customer segmentation, meaning that more involved customers receive more benefits.

Hahn et al. (2000) attempted to determine whether the market can be subdivided by measuring customers' risk attitudes as shown in the purchase of financial products. They measured individuals' risk attitudes in the financial product market, where risk factors are important, and subdivided the market into two categories: risk-averse and risk-seeking. Based on the result, each individual's choices and degree of involvement in situations where profits or losses may occur are studied.

Ⅲ. Current Practice

To explore the practice of the insurance industry in Korea, we interviewed ten customer service managers and sales agents working for several insurance companies in Korea. The interview process information is briefly described in the <Table 1>. In Korea, there is no clear distinction between service managers and sales agents; they usually perform both roles simultaneously. In other words, they contact customers to sell insurance products while assisting customers who have questions and concerns.

We asked the interviewees to explain their strategy of customer management. We also asked them how they segment customers and persuade them to sign a new insurance contract. The interviews were conducted in Korean and then translated into English.

<Table 1> Interview Process Information

Interview Period	July 23rd, 2020～July 28th 2020
Interview Methods	In person or online meetings for introduction and questioning
Interviewee Information (Anonymous)	All interviewees reside in South Korea
	Three interviewees with career of under 5 years
	Two interviewees with career of 5 to 20 years
	Five interviewees with career over 20 years

3.1. Customer Segmentation Practice

Insurance agents manage a designated list of customers; each one agent has an average of 40 customers. All the agents that we interviewed had their own expertise in the customer segmentation criteria. We have also asked agents whether they utilize the customer segmentation provided by the insurance company. Many agents replied that they do not use it, because it does not align well with their customers, and they have more trust in their own heuristic methods. Their standards can be categorized as demographic and contract-based segmentation.

Many agents use demographic information such as age, gender, occupation, household, and income for segmentation. The most popular age segmentation rule defines young customers as those in their 20s or 30s. Eight out of ten interviewees replied that most of their customers were older than 40. However, the number of customers in their 30s has increased in recent years. Several agents reported that these younger customers prefer the direct type of insurance without the intervention of insurance agents. This allows customers to make their own contracts and manage insurance services online. Household composition is a critical factor in customer relationship management. According to one manager, *"It is important to know whether the customer has a child and their child's age. Sometimes, I save a customer's phone number under their child's name."*

Another agent preferred managing customers by household. By showing concern about customers' family, agents get to know their customers and earn their trust. This practice makes it likely to sign additional insurance contracts for the customer or other family members.

Customers' income or financial condition is another notable factor mentioned in the interviews. Some agents pay special attention to their high-value customers who retain policy loans or expensive insurance products with premiums over $400. When considering customers' occupations, the agents noted a preference for face-to-face communication and an understanding of insurance service. One agent noted:

Teachers, researchers and other professionals are difficult to schedule appointments with. Most of them have a good knowledge of policy conditions or insurance clauses. Even though they do not prefer face-to-face communication, once we have met with them, the chances of selling them new contracts are high.

Another segmentation method used by insurance agents is contract-based. In contrast to the demographic category, the contract-based method focuses on the insurance plans that the customer already has. For example, the amount of insurance premium and the type of insurance product are considered unique characteristics. According to one agent:

Some customers only keep insurance services with low insurance premiums. They do recognize the future risk in their lives; however, they do not want to pay large amounts for that risk. I avoid recommending expensive insurance services to those types of customers.

Another agent explained that there are two reasons to take out an insurance policy, and customers' preference depends on whether they are interested in protection or in savings. The customer's insurance retention rate may also be considered. Some customers have a low retention rate, which means they cancel the insurance policy soon after taking it out. Some agents reported that some customers tended to make and cancel several insurance contracts.

3.1.1 Summary

- Many agents do not utilize customer segmentation provided by the insurance company.
- Demographic or contract-based segmentation method is widely used.
- Younger customers prefer to purchase insurance directly, without going through an agent
- Knowing the customer's family composition is essential.
- Customer's income is a notable factor.
- Some customers tend to stick with low insurance premiums despite the risk.
- Customers' preferences depend on whether they are interested in protection or in savings.

3.2. Customer Persuasion

The purpose of customer relationship management is to improve customer service, retain customers, and increase sales. Many agents insisted on the importance of long-term relationships with customers. Below are some interviewee comments about these relationships.

I wait until I have bonded with a new customer. At that point, the customer signs the contract.

New customers rarely decide to sign insurance contracts immediately. Sometimes I have to wait two or three years for their decision.

I build trust with customers by talking with them on the phone several times before the appointment. It is also very effective to call them once in a while and ask how they are doing.

Another manager noted that face-to-face communication is efficient for customer acquisition while online communication is better for customer management and retention. The manager also emphasized the importance of acknowledging national holidays of customers' anniversaries is essential.

To persuade customers to purchase a new insurance policy, agents appeal to their reason and to their emotions. To appeal to their reason, news articles on recent changes to health insurance and banks' interest rates can be used to recommend insurance products. Statistical data presented by the insurance company is also a great option. To appeal to their emotions, int can be effective to emphasize risk. According to one agent:

No one can be sure that they won't ever find themselves in a bad situation. After someone has been diagnosed with a serious illness, it is almost impossible to obtain an insurance policy. After explaining those risks, customers decide to take out additional insurance policies.

A customized recommendation can also be persuasive. According to another agent:

It is important to figure out what a customer needs. If a customer is in a good financial situation, I emphasize the inheritance function of insurance products. I recommend saving types of pf insurances like pension insurance only to customers who never express concern about their health.

3.2.1. Summary

- Long-term relationships with customers are essential.
- Face-to-face communication is useful for customer acquisition, but online communication is useful for customer management and retention

- Customers can be persuaded by appeals to their reason or to their emotions.
- Customized recommendations of insurance products based on the customer's financial status are recommended.

3.3. Discussion

The interviews revealed that most agents have their heuristics in customer segmentation, but the heuristics are based on agent's personal experiences. This makes their customer management methods subjective, making it difficult to deliver, share, and transfer to other agents. Another shortcoming is that the heuristics were derived from a sample of only about 40 customers, making it vulnerable to generalization.

The insurance company provides customer segmentation results based upon non-AI-based legacy methods. However, artificial intelligence technologies allow analysis of more extensive customer data features to segment customers. By adopting artificial intelligence technologies, insurance companies can enhance their customer service, management, and recommendation to increase revenue. However, it is challenging for companies lacking information technologies to construct and operate an artificial intelligence structure. In the next section, we describe a concise conceptual guideline of AI-based customer segmentation for insurance companies and show possible practical examples.

Ⅳ. Application for Customer Segmentation

A structured experimental approach can be used to develop machine learning algorithms for customer segmentation. This section proposes a customer segmentation guideline based on the Cross Industry Standard Process for Data Mining (CRISP-DM) framework for the insurance industry. Our guideline focuses on customer segmentation for customer relationship management and insurance product recommendations.

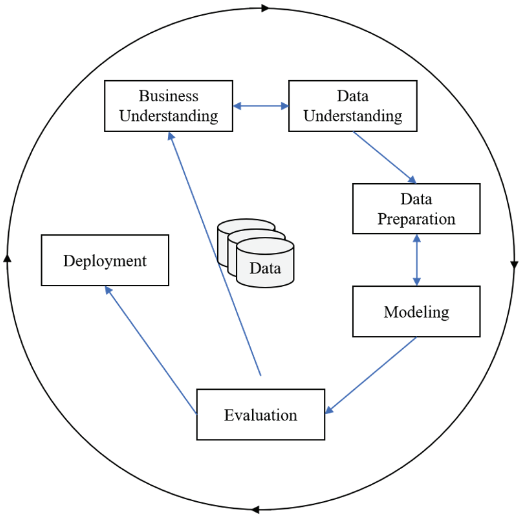
4.1. CRISP-DM Model

The CRISP-DM model is a comprehensive process model for data mining projects (Wirth and Hipp, 2000). It provides an overview of the life cycle of a data mining project. Built on previous knowledge discovery methodologies, the model breaks down the life cycle of a data mining project into six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment (<Figure 1>). The sequence of the phases presented in <Figure 1> is not strict; the arrows only indicate the most important and frequent dependencies among the phases. The purpose of each phase is described below.

(1) Business Understanding: To understand the project objectives and requirements
(2) Data Understanding: To become familiar with the data, to identify data quality problems, or to discover first insights into the data
(3) Data Preparation: All activities to construct the final dataset
(4) Modeling: To select and apply modeling techniques and to optimize model parameters
(5) Evaluation: To evaluate the model more thoroughly and review the model construction steps
(6) Deployment: To gain knowledge from the user, to organize and present the knowledge

4.2. Our Guideline

Drawing on the CRISP-DM framework, we pro-



<Figure 1> Phases of the CRISP-DM Process Model

pose a six-step customer segmentation framework for the insurance industry. This section describes the framework.

4.2.1. Business Understanding

Business understanding consists of identifying the problem, the goal of the customer segmentation, and the characteristics of the insurance industry. It is necessary to comprehend how the insurance company communicates with customers, what it knows about customers, and what customers want. We suggest interviews with customers and sales agents to arrive at this comprehension.

It is also essential to specify the goal of the segmentation model. The objective can be acquiring new customers, selling customers additional services, or persuading customers to change their insurance policy. For customer acquisition, the segmentation model should assign customers to well-defined categories.

4.2.2. Data Understanding

Data for customer segmentation in the insurance sector contains sensitive information. In many countries, personal information data is strictly regulated, so it is not available to the public. One exception is TIC benchmark dataset (Putten and Someren, 2000). The TIC dataset contains 5,822 customer records, and each record has 86 attributes. Each record includes information on product ownership and sociodemographic data. However, even sociodemographic data is based only on ZIP codes.

Insurance companies are limited to using data from their own customers, not from those of other companies. This restriction might create selection bias. If needed, insurance companies can contact their customers to obtain more detailed information. However, it is costly, and the newly gathered data can be imbalanced.

4.2.3. Data Preparation

The format of the data should be decided before the final dataset is constructed. Two formats are wide and long. A wide dataset (<Table 2>) has a single row for each individual. Thus, a wide dataset consists of many columns, and each individual has the same amount of data. The long dataset (<Table 3>) has several rows.

A wide dataset stores data in a horizontally wide table, and a long dataset stores data in a vertically long table. A wide format is more efficient if each observation can be stored in a fixed length. Insurance customers usually have several insurance plans. For example, one customer may hold ten insurance policies while another customer owns only one. Thus, when storing product ownership of insurance customers, a wide format is not efficient. The table should

<Table 2> An Example of Wide Dataset

	Name	Birth	Sex	Income	…	Product_1	Product_2	Product_3
1	Jay	1987.2.6	Male	A		P01034	P01334	null
2	Grace	1990.3.10	Female	E		D11023	null	null
3	Lily	1995.6.17	Female	D		C02234	null	null

<Table 3> An Example of Long Dataset

	Name	Birth	Product Code	Start Date	End Date	Premium
1	Jay	1987.2.6	P01034	2022.03.20	2042.03.19	44,000
2	Jay	1987.2.6	P01334	2018.06.01	9999.12.31	95,000
3	Grace	1990.3.10	D11023	2020.12.01	2021.11.31	725,100
4	Lily	1995.6.17	C02234	2014.01.01	2018.12.31	98,560

contain many columns for outlier customers with several insurance plans; the values are null for most customers.

In addition, the observation on insurance customers can be made at different time points. For instance, customers might have contacted the customer center several times: last year, three months ago, and last week. Moreover, the premium of customers' insurance can change over time. Given these time variables, we recommend constructing the final dataset in a long format.

4.2.4. Modeling

The objective of the customer segmentation model is likely to vary. In this paper, three goals of the model are assumed. The first goal of the model is designed to sell additional products to existing customers. This model is similar to the general recommendation model. Thus, we can evaluate this model with precision, recall, and f1-score.

In addition, the model is intended to persuade customers to change their insurance policies. Sometimes insurance companies want to discontinue unprofitable products. The customer switch model is like the customer portfolio expansion model, but the recommendation mechanism is different. Customers tend to prefer a broad range of insurance coverage to duplicate insurance policies. In other words, the customer portfolio expansion model seeks new insurance products not covered under their current plans. However, the customer switch model seeks insurance products that are like existing plans but more profitable to the company.

Finally, the purpose of the model is to acquire new customers. Usually, the information of prospective customers is limited. Statistically, there are many missing values in the data. Therefore, while building the model, its robustness needs to be considered. The algorithm finds similar existing customers and recommends the most suitable insurance service. Precision, recall, and f1-score are still utilized to evaluate the model.

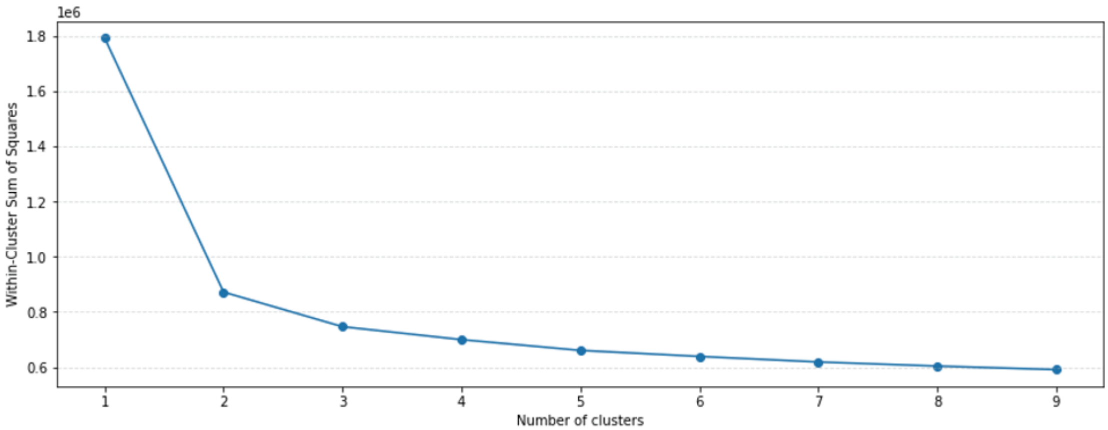
Methodologically, both classification and clustering methods are possible for customer segmentation. Classification and clustering are popular pattern identification methods in machine learning. Both methods estimate similarities among elements and categorize them. The difference is that classification uses predefined categories, but clustering does not. In other words, the classification method is a type of supervised learning, but the clustering method is a type of unsupervised learning.

To use the classification approach, categories must be defined. One possible category is customer lifetime value, which can be high, medium or low. Classification algorithms such as logistic regression, naive bayes, decision tree, or support vector machine can be applied.

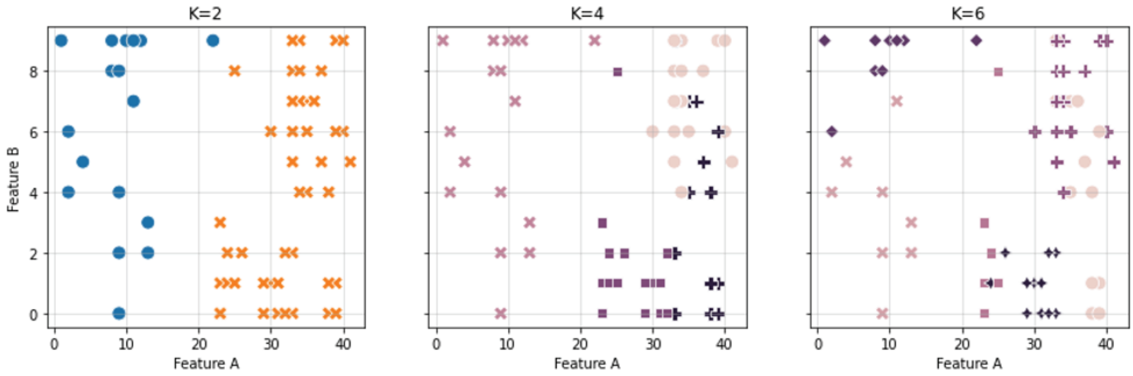
When using the clustering approach, the names of clusters should be based on the characteristics of each cluster. For example, when the clustering was performed with features of {*age, income, job, household, usage of customer service*}, the following labels for clusters are possible.

- 40s, office worker, middle-income family, teenage children
- Late 40s, highly paid specialized job, little customer services experience
- 30s, production worker, practical insurance plans

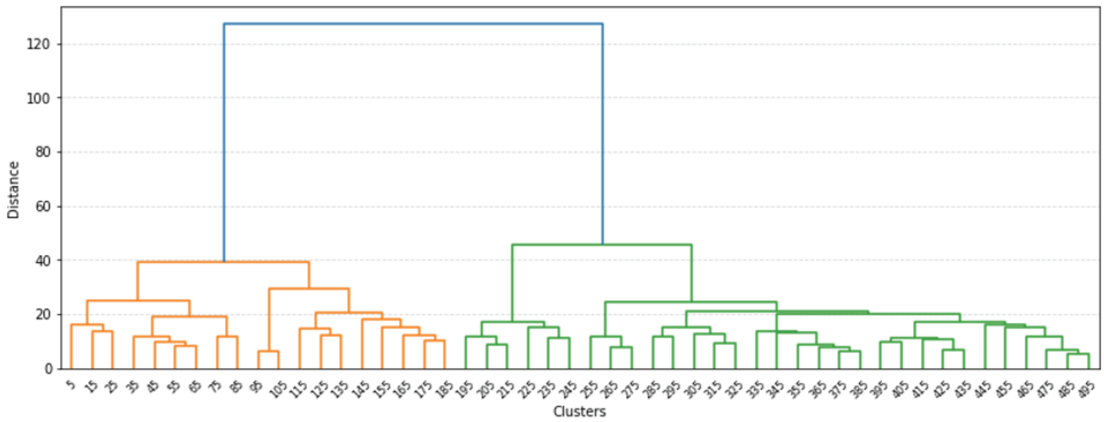
Methods like Affinity Propagation, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), K-means, or mean shift can be used with the clustering algorithms. Among clustering algorithms, we present simple and efficient example of the K-means. The K-means divides N observations into K clusters by minimizing the within-cluster sum of squares (Hartigan and Wong, 1979). The K-means algorithm needs to iterate; thus, it also needs stopping criteria. The computational difficulty of the K-means algorithm is NP-hard. Efficient heuristic algorithms are usually used to make the algorithm converge quickly to a local optimum. The within-cluster sum of squares is affected by the size of K itself, as <Figure 2> shows. Thus, finding optimal K is another problem. <Figure 3> shows the different results of K-means clustering on different Ks. When there are many variables for observations (i.e., high dimension), hierarchical clustering can reduce the number of clusters. <Figure 4> shows an example of a dendrogram of hierarchical clustering. A dendrogram shows the hierarchical relationship between the clusters. Starting from 60 clusters, they eventually converge into a single cluster. The ideal number of clusters is determined



<Figure 2> An Example of Within-cluster Sum of Squares (Different Ks)



<Figure 3> An Example of K-means Clustering (Different Ks)



<Figure 4> An Example of the Dendrogram (Hierarchical Clustering)

by the distance (i.e., the difference in height). In the case of <Figure 4>, the optimal number of clusters is two.

4.2.5. Evaluation

A thorough evaluation of the model and model construction steps review is required before deployment. Many studies found that not every information technology implementation succeeds, and one of the key implementation issues is resistance (Lapointe and Rivard, 2005). The evaluation step is essential to convince end users (i. e., sales agents and service managers) of the segmentation model. Resistance can be reduced by conducting a pilot test. End users should learn how to use the customer segmentation model and how it works. We can optimize the education phase by starting with a small number of test participants. Responses from customers are also helpful to verify the generalizability of the model. Furthermore, unexpected side effects of the model are expected to appear before final deployment.

4.2.6. Deployment

The creation of the model is not the end. After

<Table 4> Summary of Guidelines for AI-based Customer Segmentation in Insurance Field

Phase	Key Task
Business Understanding	- To interview customers and sales agents - To specify the goal of the segmentation model
Data Understanding	- To be aware of privacy laws - To obtain detailed information by direct contact with the customers
Data Preparation	- To decide the storing format of the dataset (We recommend using a long data format)
Modeling	- To optimize model parameters with precision, recall, and f1-score - To choose clustering or classification by the purpose of the model
Evaluation	- To conduct pilot tests to reduce the resistance of end users
Deployment	- To gain knowledge from sales agents, service managers, and customers - To improve the model with feedback

deploying the model, feedback from the users can be used to improve the model. In the deployment step, the knowledge can be gained from sales agents, service managers, and customers. This knowledge helps to understand business, process data, building and evaluate models.

To highlight the differences between our guideline and the CRISP-DM framework, we summarized the key aspects of each phase in <Table 4>.

Ⅴ. Conclusion

In this study, we have provided literature and theoretical factors that may affect customers' financial decisions. From the interview with insurance agents, we have found that the current customer segmentation practice is unreliable and merits further exploration. In contrast to general expectations, the AI-based customer segmentation has not been sufficiently investigated and there have been few attempts to adapt it to the insurance industry. With the legacy segmentation method, most of the data are not analyzed. As a result, the result does not match the real-world customer segments, and agents are turning away from the provided results. Therefore, we proposed a customer segmentation framework that is better suited to the insurance field, and that would improve business performance.

Our study applied the CRISP-DM methodology in the insurance industry and proposed a framework and guideline for AI-based customer segmentation. We interviewed customer service managers and sales agents to investigate critical issues in customer segmentation methods. Our results from the interviews and proposed framework will contribute to the success of AI-based innovation in insurance companies. This research will benefit future studies concerning AI-based data mining projects in insurance companies.

From a practical perspective, our findings can be helpful for managers in insurance companies. Berente et al. (2021) addressed that managers make vital decisions related to AI-based projects. It is the managers that oversee the development, implementation, and deployment of the project. Thus, the IS community needs to understand the process of AI-related activities. We hope our guidelines assist practitioners in finishing their AI-related projects successfully.

Several limitations in the current study can be addressed in future research. Although we have developed guidelines based on the grounded model and genuine interviews with managers, our guidelines are not demonstrated with real-world data. We have presented some examples with simulated data but not with real data. Future studies can utilize actual data from the insurance company or conduct case studies. If actual insurance customer data can be acquired, we will be able to understand more about insurance customers in terms of their characteristics, preference, and behavior by finding more detailed tasks for each phase. Despite limitations, our study proposes specific guidelines for AI-based customer segmentation in the insurance industry that would help the success of the data mining project. Our work also contributes to understanding the current practice of customer segmentation in the Korean insurance industry.

Acknowledgements

This study was supported by the Institute of Management Research at Seoul National University.

<References>

[1] Abdul-Rahman, S., Arifin, N. F. K., Hanafiah, M., and Mutalib, S. (2021). Customer segmentation and profiling for life insurance using K-modes clustering and decision tree classifier. *International Journal of Advanced Computer Science and Applications*, *12*(9), 434-444. https://doi.org/10.14569/IJACSA.2021.0120950

[2] Andersen, O. (1993). On the internationalization process of firms: A critical analysis. *Journal of International Business Studies*, *24*(2), 209-231. https://doi.org/10. 1057/palgrave.jibs.8490230

[3] Anderson, P. F. (1982). Marketing, strategic planning and the theory of the firm. *Journal of Marketing*, *46*(2), 15-26. https://doi.org/10.1177/002224298204600203

[4] Bae, M. (2004). A study of marketing strategies as a customer segmentation in domestic bank. *Korean Journal of Human Ecology*, *13*(3), 453-466.

[5] Bansal, A., and Shukla, A. (2020). Online insurance business analytics approach for customer segmentation. *International Journal of Advance Science and Technology*, *29*(10S), 2299-2305. Retrieved from https://www.researchgate.net/profile/Appoorva-Bansal-2/publication/348630350_Online_Insurance_Business_Analytics_Approach_for_Customer_Segmentation/links/60083d47a6fdccdcb8691146/Online-Insurance-Business-Analytics-Approach-for-Customer-Segmentation.pdf

[6] Baradaran, R., Zadeh, K., Faraahi, A., and Mastali, A. (2011). Profiling bank customers behaviour using cluster analysis for profitability. *Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management*, 458-467.

[7] Berente, N., Gu, B., Recker, J., and Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3). https://doi.org/10.25300/MISQ/2021/16274

[8] Bertaut, C. C., and Starr-McCluer, M. (2000). Household portfolios in the United States. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.234154

[9] Boodhun, N. (2017). A review of data analytical approaches in the insurance industry. *Journal of Applied Technology and Innovation*, *1*(1), 58-73.

[10] Campbell, J. Y. (2006). Household finance. *Journal of Finance*, *61*(4), 1553-1604. https://doi.org/10.1111/j.1540-6261.2006.00883.x

[11] Choe, H. C., and Cho, H. J. (2011). A study on the effects of financial education and financial experiences on the financial literacy of undergraduate students. *Journal of Consumer Studies*, *22*(4), 55-75.

[12] Choi, A., Jang, Y. J., and Han, S. M. (2016). A study on financial consumer's online information search behavior according to their perceived difficulties.

Journal of Consumer Studies, 27(2), 203-231.

[13] Deloitte. (2018). *Data Analytics within the Dutch Insurance industry*. Retrieved from https://www2.deloitte.com/nl/nl/pages/financial-services/articles/data-analytics-within-the-dutch-insurance-industry.html

[14] Ganzach, Y. (2000). Judging risk and return of financial assets. *Organizational Behavior and Human Decision Processes*, *83*(2), 353-370. https://doi.org/10.1006/OBHD.2000.2914

[15] Goonetilleke, T. L. O., and Caldera, H. A. (2013). Mining life insurance data for customer attrition analysis. *Journal of Industrial and Intelligent Information*, *1*(1), 52-58. https://doi.org/10.12720/jiii.1.1.52- 58

[16] Guo, W., Liu, F., and Zhang, X. (2021). Research on Insurance Customer Segmentation Model and Marketing Strategy Based on Big Data and Machine Learning. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3469213.3471326

[17] Ha, Y. W. (1996). A study on the segmentation of financial markets: Focusing on the 1customer tax division of household deposits. *Sogang Journal of Business*, *7*, 519-535.

[18] Ha, Y. W., and Kim, Y. D. (2011). The influence of regulatory focus on financial consumers' investment behavior: The moderating role of regulatory focus in the disposition effect and post-disposition risky alternative choice behavior. *Korean Management Review*, *40*(3), 605-631.

[19] Hahn, M. H., Shin, S. K., and Ahn, S. A. (2000). Risk attitude and risk: Aversion purchase behavior in the financial market. *Korean Management Review*, *29*(3), 337-352.

[20] Hartigan, J. A., and Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *28*(1), 100-108. http://dx.doi.org/10.2307/2346830

[21] Huh, K., Lee, H., and Choi, M. (2010). Analysis on financial asset portfolio structure in relation to household heads tendency for financial risk tolerance. *Financial Planning Review*, *3*(2), 105-134.

[22] Huh, K., and Yoo, S. (2012). Investigation of discriminating variables between three groups divided by risk acceptance in financial investment and finding differences in financial characteristics of household. *Financial Planning Review*, *5*(1), 133-159.

[23] Hur, J. O., and Kim, C. H. (2001). A study on the marketing strategy for enhancing service quality of local bank. *Journal of Marketing Management Research*, *6*(3), 157-187. https://doi.org/10.37727/jkdas. 2019.21.6.3043

[24] Johanson, J., and Vahlne, J. E. (1977). The internationalization process of the firm—a model of knowledge development and increasing foreign market commitments. *Journal of International Business Studies*, *8*(1), 23-32. https://doi.org/10.1057/palgrave.jibs.8490676

[25] Joo, S. H. (2008). Workplace financial education and retirement planning. *Journal of Consumer Studies*, *19*(1), 73-91.

[26] Joo, S. H. (2011). Mutual fund investments literacy of mid-to old-age consumers. *Journal of Consumer Policy Studies*, *40*, 47-69. https://doi.org/10.15723/jcps..40.201110.47

[27] Khajvand, M., and Tarokh, M. J. (2011). Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. *Procedia Computer Science*, *3*, 1327-1332. https://doi.org/10.1016/j.procs.2011.01.011

[28] Khamesian, F., Khanizadeh, F., and Bahiraie, A. (2021). Customer segmentation for life insurance in Iran using K-means clustering. *International Journal of Nonlinear Analysis and Applications*, *12*(February), 633-642. https://doi.org/10.22075/IJNAA.2021.22324.2350

[29] Kumar, S. J., and Oommen Philip, A. (2022). Achieving market segmentation from B2B insurance client data using RFM and K-means algorithm. *SPICES 2022 - IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems*, 463-469. https://doi.org/10.1109/

SPICES5 2834.2022.9774051

[30] Lapointe, L., and Rivard, S. (2005). A multilevel model of resistance to information technology implementation. *MIS Quarterly*, *29*(3), 461-491. https://doi.org/10.2307/ 25148692

[31] Lee, S. S. (2012). A study on the relationship of service quality, customer satisfaction, and cross buying intension in bank service: Comparison of the Korea and China bank. *Korean Corporation Management Review*, *19*(1), 83-106.

[32] Lyon, L., and Cameron, G. T. (2004). A relational approach examining the interplay of prior reputation and immediate response to a crisis. *Journal of Public Relations Research*, *16*(3), 213-241. https://doi.org/10.1080/1532-754X.2004.11925128

[33] Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, *7*(1), 77-91. https://doi.org/10.2307/ 2975974

[34] Mihova, V., and Pavlov, V. (2018). A customer segmentation approach in commercial banks. *AIP Conference Proceedings*, *2025*(1), 030003. https://doi.org/10.1063/1.5064881

[35] Min, K., and Song, I. (2014). Measuring the distribution of Korean investors' risk aversion. *Journal of Consumption Culture*, *17*(1), 69-88. https://doi.org/10.17053/jcc.2014.17 .1.004

[36] Mishra, S. K., and Kumar, M. (2011). How mutual fund investors' objective and subjective knowledge impacts their information search and processing behaviour. *Journal of Financial Services Marketing 2011 16:1*, *16*(1), 27-41. https://doi.org/10.1057/ FSM.2011.1

[37] Namvar, M., Gholamian, M. R., and Khakabi, S. (2010). A two phase clustering method for intelligent customer segmentation. *ISMS 2010 - UKSim/AMSS 1st International Conference on Intelligent Systems, Modelling and Simulation*, 215-219. https://doi.org/ 10.1109/ISMS.2010.48

[38] Nandapala, E. Y. L., Jayasena, K. P. N., and Rathnayaka, R. M. K. T. (2020). Behavior segmentation based micro-segmentation approach for health insurance industry. *ICAC 2020 - 2nd International Conference on Advancements in Computing, Proceedings*, 333-338. https://doi.org/ 10.1109/ICAC51239.2020.9357282

[39] Putten, P. van der, and Someren, M. van. (2000). *CoIL Challenge 2000: The Insurance Company Case*. Sentient Machine Research, Amsterdam. Retrieved from http://www.liacs.nl/～putten/library/cc2000/

[40] Qadadeh, W., and Abdallah, S. (2018). Customers segmentation in the insurance company (TIC) dataset. *Procedia Computer Science*, *144*, 277-290. https://doi.org/10. 1016/J.PROCS.2018.10.529

[41] Warner, N., and Cramer, S. (1995). Saving behaviours: First wave baby boomers. *Journal of Consumer Studies & Home Economics*, *19*(1), 57-67. https://doi.org/10.1111/j.1470 -6431.1995.tb00532.x

[42] Wirth, R., and Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 29-39.

[43] Zhou, R., and Pham, M. T. (2004). Promotion and prevention across mental accounts: When financial products dictate consumers' investment goals. *Journal of Consumer Research*, *31*(1), 125-135. https://doi.org/10.1086/383429

[44] Zhuang, K., Wu, S., and Gao, X. (2018). Auto insurance business analytics approach for customer segmentation using multiple mixed-type data clustering algorithms. *Tehnicki Vjesnik*, *25*(6), 1783-1791. https://doi.org/10.17559/TV-20180720122815

◆ About the Authors ◆

Kyeongmin Yum

Kyeongmin Yum is a Ph.D. Candidate at Seoul National University. His research interests are in e-business, customer behavior in information systems, and the application of artificial intelligence to business and society. He has presented on these topics in conferences such as Workshop on Information Technology and Systems and Korean Chapter of the Association for Information Systems. He has been working with e-commerce companies on various customer data analysis projects.

Byungjoon Yoo

Byungjoon Yoo is Professor at the College of Business Administration at Seoul National University. Before he joined Seoul National University, he worked at Korea University and Hong Kong University of Science and Technology. Also, he is a board member of AI committee at Seoul National University. His research interests are in online marketplaces and management strategies of digital goods such as software products, online music songs and online games. He has published on these topics in journals such as *Management Science*, *Journal of Marketing*, *Journal of Management Information Systems*, *Decision Support Systems*, and *International Journal of Electronic Commerce*.

Jaehwan Lee

Jaehwan Lee is a visiting scholar at the Institute of Management Research in Seoul National University. He received a Ph.D. degree in Management Information Systems from the Seoul National University. He has teaching experience in management information systems (undergraduate) and big data programming (graduate). He worked as a post-doctoral researcher of the Ewha Brain Institute in Ewha Woman's University, South Korea. His research interests are in e-business, new technology, application of artificial intelligence to business and society, and user behavior in information systems.

Submitted: March 30, 2022; 1st Revision: July 18, 2022; Accepted: August 16, 2022