

빅데이터를 활용한 은행권 고객 세분화 기법 연구

장민석* · 김형중

고려대학교 정보보호대학원 빅데이터응용 및 보안학과

A Customer Segmentation Scheme Base on Big Data in a Bank

Min-Suk Chang* · Hyoung Joong Kim

Division of Information Security Graduate School of Information Security, Korea University, Seoul 02841, Korea

[요 약]

대부분의 은행은 고객 세분화를 위해 성별, 나이, 직업, 주소 등 인구통계정보만을 사용하고 있으나, 이는 고객의 다양한 금융행동 패턴을 반영하지 못하는 단점이 있다. 본 연구에서는 은행 내 다양한 빅데이터를 융합하여 문제점을 해결함과 동시에 향후 많은 은행에서 폭넓게 활용될 수 있는 고객 세분화 방법을 개발하는 것을 목표로 한다.

본 연구에서 제안한 블록을 만들어 이 블록을 클러스터링하는 상향식 방식의 세분화는 기법을 제안한다. 이 방식은 기존의 인구통계정보 뿐만 아니라 다양한 거래패턴, 채널접촉패턴에 기반을 둔 고객의 다양한 금융니즈를 정교하게 반영할 수 있다는 장점이 있다. 세분화를 통해 고객의 금융니즈를 보다 정교하게 반영한 적정 동료그룹을 찾아 이를 기반으로 상품추천, 금융니즈 등급 산출, 고객이탈 예측 등 다양한 마케팅 모델을 개발하여 실제 농협은행 마케팅에 활용할 것이다.

[Abstract]

Most banks use only demographic information such as gender, age, occupation and address to segment customers, but they do not reflect financial behavior patterns of customers. In this study, we aim to solve the problems by using various big data in a bank and to develop customer segmentation method which can be widely used in many banks in the future.

In this paper, we propose an approach of segmenting clustering blocks with bottom-up method. This method has an advantage that it can accurately reflect various financial needs of customers based on various transaction patterns, channel contact patterns, and existing demographic information. Based on this, we will develop various marketing models such as product recommendation, financial need rating calculation, and customer churn-out prediction based on this, and we will adapt this models for the marketing strategy of NH Bank.

색인어 : 고객 세분화, 빅데이터, 클러스터링, 자기조직화지도, 고객관리

Key word : Customer Segmentation, Big data, Clustering, Self-Organizing Map, CRM

<http://dx.doi.org/10.9728/dcs.2018.19.1.85>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 18 November 2017 ; **Revised** 23 January 2018

Accepted 29 January 2018

***Corresponding Author; Min-Suk Chang**

Tel: +82-02-2080-7751

E-mail: jerryzzang@nonghyup.com

I. 서론

최근 디지털기기의 비약적인 발전으로 기존의 공급자 중심의 마케팅이 수요자 중심의 마케팅으로 급속히 변하고 있다. 금융 소비자들은 갈수록 다양한 수단과 채널을 통해 자신들의 복잡한 니즈를 표현하고 있으며, 금융회사는 이러한 소비자의 니즈를 경영전략에 반영하지 않고서는 경쟁에서 이길 수 없음을 인식하고 있다. 이에 따라 금융회사들은 다양한 채널에서 고객이 남기는 디지털 흔적을 수집하고 분석하여 개인 금융니즈에 맞는 개인화 마케팅으로 전략을 변화시켜 나가고 있다.

고객 세분화는 고객에 대한 이해도를 제고하여 새로운 고객을 유치하거나 기존 고객의 만족도 및 충성도를 높이기 위한 마케팅 분야의 핵심적 개념이다[1]. 다양한 빅데이터를 활용하여 고객의 니즈를 정교하게 반영한 은행권의 새로운 고객 세분화 모형을 제시하고자 한다.

기존 은행권의 세분화를 살펴보면 변수로써 성별, 나이, 직업, 거주지 등 인구통계정보만 활용한 것이 대부분이었다[2]. 하지만 이러한 세분화는 고객의 선호채널, 선호상품, 금융니즈 등을 반영하지 못한 한계를 갖고 있으며, 인터넷뱅킹 및 스마트뱅킹 등 비대면 거래가 은행 전체 거래의 대부분을 차지하는 현 상황에서 고객에 대한 이해를 왜곡시킬 가능성이 크다.

본 세분화의 목표는 은행 내 다양한 데이터를 활용하여 고객의 니즈를 보다 정교하게 반영한 적정 Peer Group을 찾아 개인 고객에 대한 이해를 제고하는 것이다. 특히 타 은행에서는 시도할 수 없는 카드데이터를 활용하여 세분화의 차별성을 높였다. 세분화 결과를 토대로 금융니즈 등급 산정 모형, 맞춤형 상품추천 모형, 고객 이탈 예측 모형 등을 개발하여 농협은행 마케팅에 직접 활용할 예정이다.

II. 선행 연구

2-1 고객 세분화(Customer Segmentation)

고객 세분화는 고객 관점(Single Customer View)의 다양한 변수들을 활용하여 비슷한 특징을 갖는 고객들을 소수의 집단으로 분류하는 것을 의미한다. 고객 세분화는 마케팅의 핵심과정으로 기업 내 자원의 분배와 실적평가의 근본이 되며, 장기적으로는 기업 경쟁력 향상의 기회를 제공한다[3].

또한 고객에게 적절한 대응과 제안을 할 수 있게 하고 기업의 특정 마케팅 활동의 대상이 되는 핵심 고객군 선정을 위한 효과적인 타겟팅을 가능하게 한다. 따라서 효과적인 세분화를 위해서는 분석 시 기준 변수를 잘 선택하는 것이 매우 중요하다[4].

세분화의 기준으로 접근 가능성, 그룹 내에서는 동질성(Homogeneity), 그룹 간에는 이질성(Heterogeneity), 그룹의 충분한 크기(Substantial)가 있다. 효과적인 세분화의 기준을 자세히 살펴보면 아래와 같다.

표 1. 효과적인 세분화를 위한 기준
Table 1. Criteria for Effective Segmentation

Division	Contents
Homogeneity	Customers in the same group should have similar characteristics
Heterogeneity	Customers in other groups should expect differentiated responses to marketing activities
Substantial	The number of customers in each group should be sufficient to enable marketing

2-2 클러스터링(Clustering)

클러스터링은 주어진 항목들을 성격이 비슷한 것들끼리 모으는 작업을 말한다. 즉 동일한 그룹의 개체가 다른 그룹의 개체보다 서로 더 유사하도록 개체 집합을 그룹화하는 작업이다.

분류도 성격이 비슷한 항목들을 구분하는 작업인데, 클러스터링과 분류가 다른 점은 분류에서는 나눌 카테고리가 미리 정해진 반면 클러스터링에서는 나눌 카테고리가 미리 정해져 있지 않다는 것이다[5,6]. 고객을 특정 기준으로 나누고 할 때 어떤 유형이 많을지 미리 알지 못하므로 고객 세분화를 할 때에는 클러스터링을 주로 활용한다.

클러스터링은 계층적 클러스터링과 비계층적 클러스터링으로 나누며, 차이점은 자료의 크기에 있다. 본 연구에서는 자료의 크기에 제한이 없는 비계층적 클러스터링을 활용하였으며, 이 중 인공신경망 기법의 학습 방법론을 보유하여 보다 심층적인 클러스터링이 가능한 Self-Organizing Map 기법을 활용하였다.

2-3 Self-Organizing Map(SOM)

본 연구에서 사용한 기법인 SOM은 핀란드 헬싱키 공과대학(Helsinki Technology University)의 코호넨(Teuvo Kohonen)에 의해 1982년 개발되었다. 이는 인간 뇌의 대뇌 피질 중 시가피질을 모델화한 비지도 학습(unsupervised learning) 인공신경망 모형의 일종이다.

SOM은 사람이 눈으로 볼 수 있는 2차원 또는 3차원 격자에 고차원 데이터의 각 개체들이 대응하도록 학습을 통해 군집을 도출해내는 알고리즘이다. 고차원의 데이터 원공간에서 유사한 개체들은 저차원에 인접한 격자들과 연결되고, 저차원 격자에서의 유사도는 고차원 입력 공간에서의 유사도를 최대한 보존하도록 학습하게 되어 있다[6].

SOM은 고차원으로 표현된 데이터를 저차원으로 변환해서 보는데 유용하며, 군집화와 시각화의 특성을 갖는다. 또한 인공신경망 기법의 학습 방법론을 보유하여 연속적인 학습이 가능하고 시간에 따라 입력 데이터의 분포가 변해도 변화에 잘 적응한다는 장점이 있다. 이에 따라 Micro 단위의 세밀한 고객 세분화와 시각화가 가능하고, 여러 변수를 활용한 Density Plot을 통해 잠재적 유사고객 파악이 가능하다는 장점이 있어 은행권 고객 세분화에 유용하게 활용될 수 있다.

III. 연구 방법

3-1 문제정의

캐나다의 RBC(Royal Bank of Canada)는 정교한 세분화를 통해 'Snowbirds'라는 세분집단을 발견함으로써 고객당 평균 수익률이 큰 폭으로 증가하고, 고객 이탈률 또한 45%나 감소하는 성과를 거둔 결과가 있다.

이렇게 최근에는 단순한 인구통계학적 방식을 넘어서는 정교한 마이크로 세그먼트화를 통해 수익성 높고 고객유지에 영향력 높은 그룹을 발견하여 맞춤형서비스를 제공하는 것이 금융업에서 강력한 경쟁력이 되고 있다. 하지만 기존 국내 은행권의 세분화 전략을 살펴보면 나이, 직업, 거주지, 소득추정 등 인구통계 정보를 주 변수로 활용하는 경우가 많았다.

인구통계정보만을 활용한 세분화는 고객의 거래패턴, 상품구매패턴, 선호채널 등 금융행동패턴을 정확하게 반영하지 못하는 치명적인 단점이 있다. 즉 직관적이고 쉽게 적용 가능하다는 장점이 있는 반면에 유사한 변수의 특성을 갖는 고객이 유사한 금융행동패턴을 나타내지 않을 수 있다는 단점이 존재한다.

본 연구는 고객의 다양한 금융행동패턴의 유사성을 기반으로 고객을 세분화하여 고객의 금융니즈를 보다 정확하게 파악하기 위한 분석 기반을 마련하였다. 이를 통해 향후 상품추천, EBM 등 마케팅 실행 시 타겟 고객 정교화와 차별화된 고객관리 기반을 통해 고객의 만족도를 향상시키는 것이 목표이다.

3-2 연구방법 및 범위

본 연구는 농협은행의 실제 다양한 데이터를 활용하여 고객 세분화에 대한 실증 연구를 진행했다. 인구통계정보 외에 요구불 거래내역, Web/App Log, 카드데이터 등을 다양하게 활용하였다. 특히 타행과 달리 카드사가 분리되지 않았기 때문에 타행과의 차별화를 위해 카드데이터를 적극 활용하고자 하였다.

또한 농협은행 2,200만 유효고객 중 200만 고객을 층화추출하여 분석에 활용하였으며, 전체 고객과의 인구통계학적 구성비가 유사함을 확인하여 층화추출에 대한 정확성을 검증

하였다.

클러스터링 기법 중 Self-Organizing Map 기법을 활용하여 다양한 군집개수를 테스트하여 활용도와 성능면을 충분히 고려하여 세그먼트수를 결정하였다.

IV. 연구과정 및 결과

본 연구에서 제안하는 고객 세분화 기법의 전체 프레임워크는 크게 5단계로 구성된다. 1단계는 은행 내 각 시스템에서 관리하고 있는 데이터를 수집하는 단계이며, 2단계에서는 고객변수의 선택 및 변수를 활용하여 개인별 금융프로파일을 만들었다. 3단계에서 직업군별 연령대, 거주지역, 거래성향 등을 고려하여 블록을 구성했으며, 4단계에서 이 블록을 클러스터링하여 고객 세그먼트를 완성하였다. 마지막 단계는 각 그룹의 고객 특성을 이해하고 기존 세분화 모형과 차별성을 찾는 것이다.

4-1 데이터 수집

본 연구에서 실제 활용한 데이터는 농협은행에서 자체적으로 구축하여 운영하는 시스템에서 관리하는 데이터로 정보보호를 위해 철저한 접근 통제가 이루어지고 있다. 실제 마케팅에 활용할 모형을 만들 목적으로 데이터에 접근이 허용되었으며, 농협은행 DW 및 계정계에 있는 데이터를 수집 및 전처리 후 임시서버에 저장하고 고객별로 데이터를 통합하여 분석을 진행하였다.

활용한 데이터는 앞서 밝힌 바와 같이 농협은행 200만 고객을 대상으로 하였다. 성별, 나이, 직업, 하나로 등급 등이 샘플고객의 분포와 전체고객의 분포가 동일하도록 설계하였으며, 전체고객 구성비와의 비교 검증을 통해 분포가 거의 유사함을 확인 후 확정하였다.

해당 데이터 중에는 농협은행 데이터 구조적인 문제로 사망자(21,575명)와 연령 미상의 고객(7,305명)이 있어서 이를 제외한 1,971,120명을 대상으로 분석을 실시하였다.

데이터의 기간은 2015년 1월 1일부터 2016년 12월 31일까지로 2년 치를 활용하였다. 데이터 수집 과정에서 고객의 채널접촉정보를 파악할 수 있는 Web/App 로그데이터는 데이터 용량이 방대하여 정보계의 효율적인 운영 상 3개월 보관 후 삭제하고 있었다. 이와 같은 이유로 본 연구에서는 미리 준비하여 해당 데이터는 6개월치를 활용할 수 있었으며, 데이터 수집기간의 문제는 농협은행의 데이터 관리 부분에서 해결해야 할 시급한 과제로 보인다.

데이터 종류에는 인구통계정보, 거래 및 접촉이력, 카드데이터 등이 있으며, 인구통계정보는 CRM시스템에서 기본적으로 관리되는 정보가 모두 포함되었다. 은행업에서 판매되는 모든 상품군에 대한 거래정보를 반영하였으며, 채널접촉 정보로는 업무원가 산정을 위한 사무량 데이터를 비롯하여

Web/App 로그데이터, 캠페인 정보 등을 수집하였다. 데이터 수집 현황은 표 2와 같다.

본 연구는 보안 문제에 대해 매우 엄격한 은행 내부 규정상 Python과 R과 같은 오픈소스 프로그램 사용이 금지되어 있는 관계로 주 분석 솔루션인 SAS Enterprise Guide, Viya와 Text Analytics(TA)를 활용하여 데이터 핸들링 및 분석을 실시하였다.

표 2. 데이터 수집 현황
Table 2. Data collection status

Source	Type	Details
Population Stat	structured	Sex, Age, Occupation, Address, Customer rating etc.
Deposit	structured	Balance, Transactions etc.
Loan	structured	Application, Review, Redemption etc.
Bancassurance	structured	Contract and status etc.
Card	structured	Products, Payment, Franchisee, Credit information etc.
Work volume	structured	Channel contact history
Demand deposit	unstructured	Transactions, Records etc.
Campaign	unstructured	Contacts, Reaction etc.
Web/App Log	unstructured	Contact history of Internet Banking and Smart Banking
e Marketing	structured	Purchase at Internet Banking and Smart Banking

4-2 개인별 금융 프로파일 산출

이번 단계는 수집된 데이터를 정제하여 고객별 금융 프로파일 산출과 각 프로파일의 상태를 정의하는 것이다. 가장 기본적인 고객 프로파일은 인구통계정보이며, 이는 일반적인 CRM시스템에서 얻을 수 있다. 대표적인 변수는 나이, 성별, 주소, 연락처 등이다.

인구통계정보 외에 고객이 은행에서 활동하는 저축, 대출 등의 거래 이력과 영업점, 인터넷뱅킹, 스마트뱅킹 등 거래를 위해 활용하는 채널접촉 이력이 있다. 또한 급여이체, 카드사용 등 소득소비 이력이 고객 프로파일로 활용될 수 있다. 특히 최근에는 대부분의 거래가 비대면 채널(인터넷뱅킹, 스마트뱅킹 등)에서 이루어지기 때문에 이에 대한 변수를 효과적

으로 만드는데 주의를 기울였다[7,8]. 또한 타행에서는 활용하지 못하고 있는 카드데이터를 적극 활용해 차별성을 갖도록 했다.

한편 기존의 연구를 살펴 보면 서로 다른 고객 프로파일의 선정은 다른 세분화 결과를 도출할 수 있기 때문에 선정 자체가 매우 중요한 문제였다. 하지만 본 연구의 분석과정에서는 의미 없는 고객 프로파일을 활용한다고 해도 블록과 클러스터링 과정에서 통합되거나 큰 영향을 미치지 못하므로 기존 세분화에 비해 중요도가 떨어지는 면이 있었다.

수집한 데이터를 기반으로 인구통계, 저축성향, 소득소비성향, 접촉성향, 대출성향의 5가지 이력 구분을 만들어 개인별 금융 프로파일을 정의하였다. 정의한 내역은 아래와 같다.

표 3. 금융 프로파일 산출 결과
Table 3. Result of financial profile

Type	Contents
Population Stat	<ul style="list-style-type: none"> ■ Age ■ Address ■ Occupation
Saving Tendency	<ul style="list-style-type: none"> ■ Monthly deposit/withdrawal amount ■ Number of accounts ■ Average balance ■ Fluctuation of Average Balance etc ■ Amount of deposit by product group ■ Average balance ratio ■ Average withdrawal amount to another bank ■ Product views from Internet Banking and Smart Banking
Income / Consumption tendency	<ul style="list-style-type: none"> ■ Income : Salary ■ Consumption : Card payment amount, consumption type, cash service, etc.
Channel Tendency	<ul style="list-style-type: none"> ■ Average contact count of Internet Banking and Smart Banking, Views per menu etc. ■ Number of transactions per channel, Channel contact ratio, etc.
Loan Tendency	<ul style="list-style-type: none"> ■ Maximum loan amount, Average loan amount, Average monthly balance etc ■ Principal and interest, Prepayment ratio etc.

4-3 금융 프로파일 기반 블록 생성

본 연구에서 블록은 각 고객 프로파일 별로 모두 유사한 속성 값을 갖는 고객들의 집합을 의미한다. 블록을 생성하는 이유는 개인 단위로 세그먼트화할 경우 개별 고객의 데이터 존재 유무에 따라 고객 세그먼트가 이뤄질 가능성이 커져 정확한 고객 금융 유사그룹 산출이 어렵기 때문이다. 따라서 생성된 금융 프로파일을 기반으로 개인과 가장 비슷하게 만들기 위해 블록을 만들었으며, 이 과정에서는 변수 선택이 전체 세분화에

미치는 영향이 클 수 있어 변수의 분포 및 상관계수 등을 면밀히 검토한 후 블록변수를 선택하였다[9,10].

직업군별 연령대, 거주지역, 거래성향 등을 고려하였으며, 채널접촉성향 등 은행이 보유하고 있는 다양한 데이터를 융합하여 생성하였다.

우선 18,478개의 블록을 만든 후 생성된 블록 중 활동 고객이 30명 이상인 6,459개 블록을 클러스터링 대상으로 선정하였다. 여기서 활동고객은 은행에서 총수신 및 외화저축성 평잔과 방카슈랑스 보험료의 합산 금액이 30만원 이상인 경우이거나 총 여신 평잔이 0원 이상인 고객을 말한다. 블록 생성 결과를 요약하면 아래와 같다.

표 4. Block 생성 결과
Table 4. Result of block creation

Block variable	Number of groups
Age by occupation group	67
Residence	3
Average number of bank account transactions	2
Average credit card usage per month	2
Call center contact history	2
Non-face to face1: Average Monthly Transactions of Tele Banking, Internet Banking and Smart Banking	3
Non-facing2: Average Monthly Transactions of ATM	2
Face to face: Visits	4
Number of blocks created	18,478
Number of blocks to be clustering (More than 30 active customers)	6,459

4-4 고객 세그먼트 생성

고객 프로파일과 각 프로파일의 상태 정의에 따라 고객 블록을 구분한 이후 각 블록을 클러스터링한다. 다시 말하면 각 블록 간의 유사도를 기반으로 유사도가 높은 블록들을 하나로 통합하는 과정이다. 본 연구에서는 Kohonen SOM기법을 활용하여 클러스터링을 실시하였다.

표 5. 클러스터링 활용 변수
Table 5. Variables used in clustering

Group	Final variable
Extra funds (3)	<ul style="list-style-type: none"> ■ Monthly average balance ■ Deposit utilization rate ■ Share
Making funds (7)	<ul style="list-style-type: none"> ■ Average monthly deposit amount ■ Usage rate of installment savings ■ Share of installment savings ■ Usage rate of subscription savings ■ Share of subscription savings ■ Usage rate of mutual fund and trust ■ Share of mutual fund and trust
Loan (7)	<ul style="list-style-type: none"> ■ Loan total balance ■ Average monthly redemption ■ Average monthly prepayment ■ Credit loan utilization rate ■ Credit loan share ■ Mortgage loan utilization rate ■ Mortgage loan share
Card (2)	<ul style="list-style-type: none"> ■ Average monthly card usage ■ Monthly average card loan amount ■ Monthly average cash service amount
Demand deposit (3)	<ul style="list-style-type: none"> ■ Average monthly deposit amount ■ Average monthly withdrawal amount ■ Average yearly average balance

우선 생성된 블록의 금융활동 유사성을 판단하기 위해 총 6개(여유자금, 목돈마련, 대출, 외화, 카드, 요구불) 영역에서의 고객의 성향을 기반으로 변수를 추출하였다. 여유자금 7개, 목돈마련 13개, 대출 19개, 외화 2개, 카드 6개, 요구불 4개의 후보변수를 만든 후 이 변수들 간의 상관계수를 확인하여 클러스터링에 활용할 최종 변수를 22개를 산정하였으며, 현황은 표 5와 같다.

최종 선택 변수에 대해서 표준화 작업을 진행한 후 SOM 알고리즘 기반의 클러스터링 작업을 진행하였다. 여기서 적절한 k값(군집수)을 찾기 위한 성능판단지표로 CCC(Cubic Clustering Criterion), 군집 간 표준편차(Total STD), 군집 내 표준편차(Within STD), R-Squared(RSQ), RSQ Ratio를 활용했다. 군집 내 표준편차는 작을수록 좋으며, 그 외 지표는 클수록 클러스터링이 잘 되었다고 판단할 수 있다.

k값은 마케팅 관점에서 활용하는데 유용할 것으로 판단되는 세분화 개수인 50~80개 사이에서 처음에는 10개부터 10개 단위로 클러스터링을 해보았을 때 70개 내외에서 판단지표가 가장 좋게 나오는 것을 확인하고 70개 전후에서는 k값을 1씩 조정하며 클러스터링했다. 그 결과 70개일 때 성능판단지표가 고르게 좋게 나와 선택하였다. k값별로 각 지표의 값은 표 6과 같다.

표 6. k값에 따른 모형 성능 평가

Table 6. Evaluation of model performance according to k value

K	CCC	TOTAL STD	WITHIN STD	RSQ	RSQ RATIO
50	1013.49	0.989	0.512	0.735	2.770
60	983.03	0.976	0.488	0.752	3.040
65	960.69	0.962	0.472	0.762	3.208
69	961.91	0.968	0.467	0.770	3.348
70	1086.84	0.989	0.466	0.781	3.562
71	1015.08	0.989	0.462	0.785	3.641
75	982.76	0.969	0.455	0.782	3.592
80	989.43	0.969	0.454	0.784	3.624

군집의 특성을 하나씩 살펴보면, 이 중 13개 군집에서 고등학생 이하, 농협 임직원, 외국인, 공무원/군인 직업군으로 주로 구성되어 있었다. 이들 직업군은 다른 고객들과 분석결과나 비즈니스 관점에서도 전혀 다른 금융성향을 보인다. 특히 공무원/군인의 경우 일반직장인과 비슷할 것이라고 판단할 수 있으나, 지자체와 공공기관과 거래가 많은 농협은행의 특징 상 다른 거래성향을 보인다. 따라서 해당 직업군을 4개로 새롭게 군집화하여 결국 61개의 세그먼트를 생성하였다.

4-4 세그먼트별 주요 특성 이해

표 7. 고객 세분화 결과

Table 7. Result of customer segmentation

Channel	Preferred product groups		
Face to face (14)	All (3)	Card & Making funds (3)	Extra funds & Making funds (3)
	Subscription savings (3)	Extra funds (2)	
All (12)	All (5)	Card & Making funds (5)	Making funds (2)
Non-face to face (31)	Making funds & Loan (2)	Loan (3)	Subscription savings (3)
	Extra funds (1)	Subscription savings (6)	Non activity (16)

클러스터링 결과를 기반으로 각 군집별로 어떤 채널 및 상품을 선호하는지를 알아보기 위해 고등학생 이하, 농협 임직원, 외국인, 공무원/군인 군집을 제외한 57개 군집을 대면/비대면 채널 선호 여부와, 상품선호를 기준으로 군집 특성을 정의하였다. 표 7은 정리한 결과이다.

본 연구의 결과 몇몇 세그먼트에서 그 동안 인지하지 못했던 고객군의 특성을 알아낼 수 있었다. 대출 니즈가 많은 고객은 비대면 채널을 선호하는 집단으로 분류되었는데, 비대면 채널도 많이 선호하는 것을 알 수 있었다. 이 결과를 통해 은행은 고객이 편의성을 고려하여 인터넷이나 스마트뱅킹으로 대출할 수 있도록 기반을 신속하게 만들어야 경쟁에서 앞설 수 있다는 것을 보여준다.

또한 60대 이상 고객은 비대면 채널을 많이 활용하지 않을 것이라는 등의 선입견을 깰 수 있는 기회가 되고, 타겟 고객에 대한 재설정할 수 있는 근거를 마련하였다. 예를 들면 60대 이상의 고객 군 중 대면채널을 주로 이용하는 고객 군과 영업점 방문을 통해 거래하는 고객을 구분하여 타겟 고객을 선정할 수 있게 되어 마케팅 성공률을 높일 수 있을 것으로 기대한다.

군집별로 주요 특성은 세그먼트별로 정의하여 많은 직원이 한 눈에 볼 수 있도록 프로파일을 정의하여 각 채널 및 부서에 정기적으로 업데이트하여 제공할 예정이다. 그림 1은 예시 화면이며, 각 그룹별로 이와 같은 화면을 CRM시스템에 개발하여 전 직원이 고객에 대해 쉽게 파악할 수 있도록 할 예정이다.

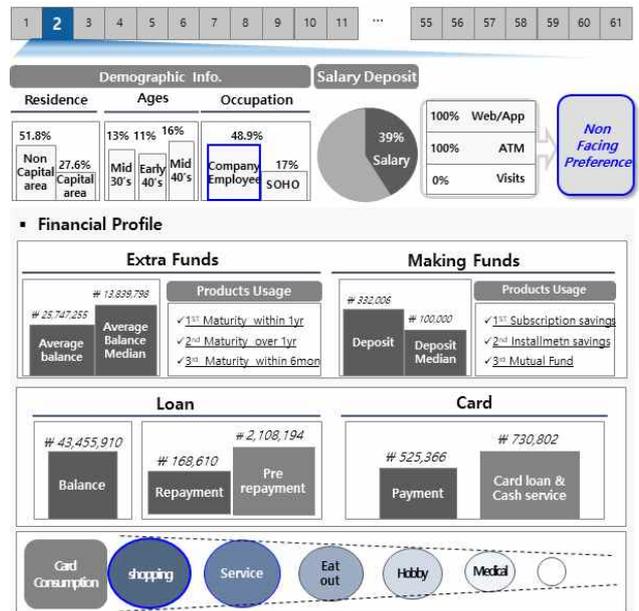


그림 1. Segment별 고객 금융 프로파일 화면 예시
Fig. 1. Customer Finance Profile Screen Example by Segment

V. 결론 및 향후 연구계획

본 연구에서는 고객에 대한 심층적 이해를 위해 은행권에서 활용할 수 있는 다양한 정형 및 비정형 데이터를 활용하여 고객세분화를 실시하였다. 이 과정에서 특정 고객군에 대한 잘못된 인식을 확인할 수 있었고, 이에 대한 대응책도 제시하는 계기가 되었다. 또한 특정 데이터를 활용하여 타겟 고객군을 만들어 내는 효율적인 방법도 찾아내는 소득도 있었다.

이번 연구는 상향식방식의 세분화로 하향식 방식에서 변수의 유무에 따라 군집화되는 단점을 극복했다는 점에 의의가 있다. 또한 금융행동패턴이 유사한 고객 집단을 인구통계학적 일반변수를 기준으로 세분화시킬 수 있기 때문에 향후 많은 은행권에서 폭넓은 활용이 가능할 것으로 기대한다.

이번 세분화는 다양한 개인 맞춤형 분석모델로 활용이 가능하다. 향후 연구계획으로는 세분화 기준에 따라 고객 금융니즈 유형을 정의하고, 세그먼트별 금융니즈 분포에 따라 금융니즈 등급을 산출하는 모형을 개발할 예정이다.

또한 세분화와 금융니즈 등급 산출을 발전시켜 궁극적으로는 개인 맞춤형 상품추천 모형 및 고객이탈 예측 등 다양한 모형을 개발하여 농협은행 마케팅 전략에 실제 활용할 예정이다.

참고문헌

- [1] S.-B. Jeong, Y. H. Shin, S. R. Koo, and H.-S. Yoon, "A market segmentation scheme based on customer information and QAP correlation between product networks," *Journal of the Korea Society for Simulation*, vol. 24, no. 4, pp. 97-106, 2015.
- [2] E. L. Melnic, "How to strengthen customer loyalty, using customer segmentation?," *Bulletin of the Transilvania University of Brasov, Series V: Economic Science*, vol. 9, no.2, pp. 51-60, 2016.
- [3] Jongho Kim, "Application of data mining techniques for finding customer-oriented product market segments," *Journal of Digital Contents Society*, vol. 13, no. 3, pp. 385-392, 2012.
- [4] J. H. Seo, A Study on the Customer Behavioral Analysis Using Big Data of Distribution Industry, MS Thesis, Gachon University, Seongnam, 2017
- [5] G. Wang, G. Nie, P. Zhang, and Y. Shi, "Personal financial market segmentation based on clustering ensembles," in *Proceedings of WRI World Congress on Computer Science and Information Engineering*, Los Angeles, CA, vol. 3, pp. 694-698, 2009.
- [6] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Transaction on Neural Networks*, vol. 11, no 3, pp. 586-600, 2000.
- [7] P. Klaus and B. Nguyen, "Exploring the role of the online customer experience in firms' multi-channel strategy: An empirical analysis for the retail banking services sector," *Journal of Strategic Marketing*, vol. 21, no. 5, pp. 429-442, 2013.
- [8] W. C. Lee and S.-B. Jeong, "A research recommendation of financial goods using multi-profile ensemble technique: Focused on credit & check card," in *Proceedings of the Korean Academic Society Of Business Administration Annual Meeting*, Gwangju, pp. 681-696, 2017.
- [9] F.M. Hsu, L.P. Lu, and C.M. Lin, "Segmenting customers by transaction data with concept hierarchy," *Expert System with Applications*, vol. 39, no. 6, pp. 6221-6228, 2012.
- [10] T.C. Lu and K.Y. Wu, "A transaction pattern analysis system based on neural network," *Expert Systems with Applications*, vol. 36, no.3, pp. 6091-6099, 2009.



장민석(Min-suk Chang)

2008년 : 동국대학교 경제학과(학사)
2017년 : 고려대학교 정보보호대학원
(석사과정)

2006년~현재 : 농협은행

※ 관심분야 : 빅데이터, 데이터마이닝, 인공지능, 마케팅 등



김형중(Hyoung-Joong Kim)

1978년 : 서울대학교 전기공학과 학사
1986년 : 서울대학교 제어계측공학과
(공학석사)
1989년 : 서울대학교 제어계측공학과
(공학박사)

1989년~2006년 : 강원대학교 교수

2006년~현재 : 고려대학교 정보보호대학원 교수

※ 관심분야 : 컴퓨터보안, 패턴인식, 가역정보은닉, 머신러닝, 빅데이터 분석 등