http://www.jsebs.org
pISSN 2288-3908
eISSN 2765-3846

The Journal of Society for e-Business Studies
Vol.27, No.2, May 2022, pp.21-36
https://doi.org/10.7838/jsebs.2022.27.2.021

클러스터링 기법을 활용한 이커머스 사용자 리뷰에 따른 시장세분화 연구

A Study on Market Segmentation Based on E-Commerce User Reviews Using Clustering Algorithm

김민경(Mingyeong Kim)*, 허재석(Jaeseok Huh)**, 사애진(Aejin Sa)***,
전아름(Ahreum Jun)****, 이한별(Hanbyeol Lee)*****

초 록

최근 코로나로 인해 이커머스 시장이 확대되면서 인터넷 쇼핑몰 이용률 증가와 함께 다양한 형태의 소비 패턴을 보이는 고객들이 나타나고 있다. 기업은 고객 리뷰를 통해 고객의 의견과 정보를 얻을 수 있기 때문에 온라인 플랫폼에서의 고객 리뷰 관리에 대한 연구의 필요성이 증가하고 있다. 본 연구에서는 고객들을 군집화하고 분석하였으며, 이커머스 시장에 존재하는 고객 유형을 정의하고 시장세분화를 수행하였다. 구체적으로, 본 연구는 온라인 쇼핑몰 위메프(Wemakeprice)의 고객 리뷰 데이터를 수집하여 K-means 클러스터링을 진행하였으며, 그 결과로 6개의 군집이 도출되었다. 이후 6개의 군집으로 시장세분화 된 결과를 분석하여 각 군집의 특징을 정의하고 고객관리 방안까지 함께 제시하였다. 본 연구 결과는 이커머스 시장의 고객 유형 파악과 고객관리를 용이하게 하는 자료로 사용될 것이며, 다양한 온라인 플랫폼의 고객관리 비용 절감과 수익 창출에 기여할 것으로 기대된다.

ABSTRACT

Recently, as COVID-19 has made the e-commerce market expand widely, customers who have different consumption patterns appear in the market. Because companies can obtain opinions and information of customers from reviews, they increasingly face the requirements of managing customer reviews on online platform. In this study, we analyze customers and carry out market segmentation for classifying and defining type of customers in e-commerce. Specifically, K-means clustering was conducted on customer review data collected from Wemakeprice online shopping platform, which leads to the result that six clusters were derived. Finally, we define the characteristics of each cluster and propose a customer management plan. This paper is possible to be used as materials which identify types of customers and it can reduce the cost of customer management and make a profit for online platforms.

키워드 : 이커머스, 고객리뷰, 시장세분화, K-means 클러스터링
E-Commerce, Customer Reviews, Market Segmentation, K-means Clustering

이 논문은 2021년 과학기술정보통신부의 재원으로 정보통신산업진흥원의 지원을 받아 수행된 연구임(S0317-21-1002)

* First Author, Undergraudate Student, Department of Business Administration, Tech University of Korea(kmg8561@tukorea.ac.kr)
** Corresponding Author, Assistant Professor, Department of Business Administration, Tech University of Korea(jshuh@tukorea.ac.kr)
*** Co-Author, Undergraudate Student, Department of IT Management, Tech University of Korea(aejin518@tukorea.ac.kr)
**** Co-Author, Undergraudate Student, Department of IT Management, Tech University of Korea(j2665138@tukorea.ac.kr)
***** Co-Author, Undergraudate Student, Department of IT Management, Tech University of Korea(star004@tukorea.ac.kr)
Received: 2022-02-09, Review completed: 2022-03-02, Accepted: 2022-03-22

1. 서 론

인터넷과 모바일 서비스의 발달로 온라인 플랫폼에서의 거래는 보편화되고 있으며, 이커머스 시장과 온라인 플랫폼의 규모는 점점 확대되고 있다. 통계청에 따르면, 2021년 온라인 쇼핑몰의 규모는 약 18조 원에 달하였으며, 규모는 매년 16%가량 성장하고 있는 것으로 확인되었다[24]. 이처럼, 온라인 플랫폼 이용이 활성화됨에 따라 기업들은 방대하고 다양해지는 온라인 고객관리를 위한 정보 수집과 분석에 힘쓰고 있다.

고객들은 온라인 내에서의 제품 구매를 위해 정보탐색의 과정을 거치는데 이 과정에서 고객 리뷰에 쉽게 노출된다[33]. 고객 리뷰는 한 명 또는 다수의 고객 구매 의사결정에 많은 영향을 미칠 뿐만 아니라 기업과 제품에 대한 이미지에도 큰 영향을 미친다. 따라서, 고객 리뷰 분석을 통한 고객관리는 온라인 플랫폼 기업들의 필수적인 과제가 되었다[23]. 또한, 이커머스 시장에는 서로 다른 특성과 니즈(needs)를 가진 고객들이 존재하고, 이는 고객이 작성한 리뷰를 통해서 파악될 수 있다.

한편, 시장세분화는 다양한 특성을 지닌 고객들을 분류할 수 있게 해주며, 기업의 고객에 대한 이해도를 향상시킬 수 있게 해준다. 효과적인 시장세분화는 기업이 새로운 고객을 유치하고 기존 고객의 만족도와 충성도를 높일 수 있는 맞춤형 서비스를 제공할 수 있게 하며, 이는 기업의 매출 증가로 이어질 수 있다.

따라서 본 연구는 고객 분석을 통한 효과적인 고객관리 방안 도출을 위해 위메프(Wemakeprice) 온라인 쇼핑몰의 고객들을 대상으로 이커머스 시장의 시장세분화를 진행하였다. 고객 정보에 대한 데이터와 고객들이 작성한 리뷰를 수집하여 감성분석과 전처리 과정을 진행하였다. 고객의 특성이 되는 23개 입력변수를 정의하여 K-means 클러스터링을 진행하였으며 6개의 군집을 도출하였다. 각 군집 특성을 분석하고 이에 맞는 군집 이름을 명명함으로써 군집을 유형화하고 그에 맞는 마케팅 방안을 제시하였다.

본 논문의 구성은 다음과 같다. 제2장에서는 이커머스 리뷰, 시장세분화, 클러스터링에 대한 선행 연구를 다룰 것이다. 제3장에서는 데이터 수집, 데이터 전처리, 클러스터링 과정을 담은 연구 방법을 설명할 것이다. 제4장에서는 K-means 클러스터링을 통해 도출해 낸 6개 군집에 대한 분석과 시장세분화한 연구 결과를 다룰 것이며, 제5장에서는 결론 및 제언을 기술할 것이다.

2. 선행 연구

2.1 이커머스 시장의 고객 리뷰

이커머스 시장은 새롭고 다양한 플랫폼을 파생시키며 경제활동에 주축이 되는 하나의 시장으로 떠오르고 있다[28]. 이커머스 시장에서 기업이 경쟁우위를 선점하기 위해서는 고객을 효과적으로 관리하는 것이 중요하며, 이에 대해서 다양한 연구가 수행되었다[29]. 본 연구에서는 고객관리를 위해 온라인 플랫폼의 고객 리뷰를 이용하여 고객의 특성을 파악할 수 있는 방법론을 제안한다. 여기서, 고객 리뷰는 자신이 경험한 기업, 제품 및 서비스와 관련된 부정적 혹은 긍정적 메시지를 온라인 매체를 통해

다른 고객에게 전하는 것을 말한다[8].

Choi[4], Park[27]은 고객 리뷰가 제품 실물 확인이 불가능한 온라인에서 실제로 경험한 고객의 의견이나 사용정보를 포함하고 있기 때문에 다른 고객들이 신뢰하고자 하는 경향이 강하다고 강조하였다. 또한, 온라인 리뷰와 제품 구매 간의 상관관계를 설명하는 선행 연구가 많이 진행되어왔으며[2, 3], 나아가 Hur et al.[8]은 고객 리뷰의 어떤 변수들이 소비자의 긍정적인 구매의사결정을 유도하는지 연구하였다. 이처럼, 대부분의 기존 연구들은 마케팅 분야에서 고객 리뷰가 고객의 구매 의사나 행동에 미치는 영향력에 대해서 주로 다루어 왔다. 본 연구는 기존 연구들과 다르게 온라인 리뷰를 소비자 입장이 아닌, 기업 측면에서 바라보며 온라인 리뷰를 통해 고객을 바라보는 시각에 대한 연구 방법을 제공한다.

2.2 시장세분화

시장세분화의 개념은 Robinson[30]에 의해 최초로 이론화되었으며, Smith[32]에 의해 마케팅 부문에 소개된 이후 현재까지 마케팅에서 중요한 개념으로 인식되고 있다[1]. 기존 연구들은 시장세분화 개념을 소비자의 욕구를 정확하게 만족시키기 위해 유사한 소비 집단별로 시장을 나누고 각 집단에 맞게 마케팅 전략을 조정하는 과정으로 정의하였다[32, 34].

Kwon and Choi[19]는 시장세분화를 통해 세분화된 시장 속의 고객들은 같은 형태의 효익을 원하거나 문제해결을 바란다고 강조한다. 본 연구는 이러한 시장세분화의 특성을 활용하여 온라인 쇼핑몰의 다양한 고객들의 특성을 효율적으로 파악하고 각 고객에 적합한 마케팅 전략을 제안하고자 한다. 이에 더하여 본 논문은 리뷰 분석을 통해 기업의 현실적인 마케팅 방안을 제시함으로써 기업이 효율적으로 고객을 관리할 수 있는 방안을 제안한다는 의의가 있다.

2.3 클러스터링(Clustering)

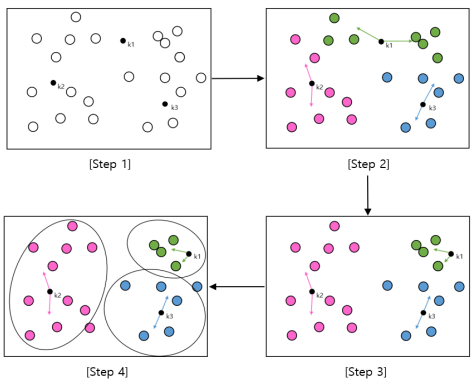
시장세분화를 위해 가장 널리 이용되고 있는 방법 중 하나는 클러스터링 기법을 이용하는 것이다[16]. 클러스터링은 주어진 데이터들을 특성에 따라 유사한 것끼리 그룹화함으로써 유형별 군집 특성을 분석하는 데이터 마이닝 기법으로[13] 군집분석(clustering analysis)이라고도 부른다.

본 연구는 비계층적 클러스터링 기법 중 K-means 클러스터링을 사용하였다. K-means 클러스터링은 군집 내 데이터 간의 차이를 최소화하고 군집 간 차이를 최대화하는 최적 군집 모델을 찾는 것을 목표로 하는 클러스터링 기법이다[22]. K-means 클러스터링의 절차는 다음과 같이 구성되어 있으며, <Figure 1>은 그 과정을 시각화하여 나타낸다[37].

단계 1: k개의 중심점을 각각 임의로 선택한다.
단계 2: 각 자료에서 가장 가까운 지점에 군집의 중심점을 설정한다.
단계 3: 각 군집 내의 자료들의 평균을 계산하여 군집의 중심을 바꿔준다.
단계 4: 모든 자료에 대해서 군집 중심의 변화가 없을 때까지 단계 2와 단계 3을 반복한다.

본 연구는 시장세분화 연구에 유용하게 쓰이

는 K-means 클러스터링 모델을 사용함으로써 최적 고객 군집을 정의하고자 한다.

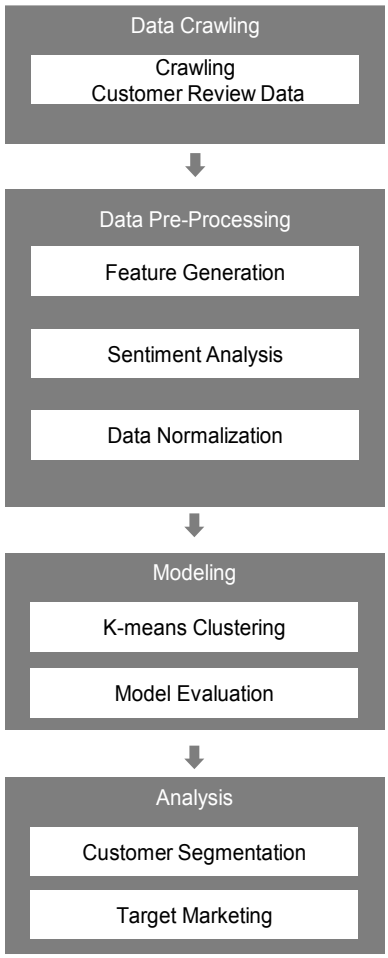


〈Figure 1〉 The Procedure of K-means Clustering

3. 연구 방법

본 연구는 특정 온라인 플랫폼 사이트를 지정하여 사이트 내 고객 리뷰를 통해 고객의 특성을 분석하고 시장세분화를 진행하였다. 연구 대상 사이트는 위메프(https://front.wemakeprice.com/main)라는 이커머스 플랫폼으로, 고객들의 다양한 상품 구매 리뷰들이 작성되어 있다. 또한, 위메프는 한 명의 고객이 작성한 모든 리뷰들을 한 페이지에서 확인하여 정보를 수집 및 분석할 수 있다는 점에서 본 연구에 적합한 연구 대상 사이트라고 판단하였다.

본 연구는 크게 4단계의 과정으로 진행되었다. 첫 번째 단계에서는 위메프 내 고객 리뷰 데이터를 웹 크롤링 기법을 이용하여 수집하였다. 두 번째 단계는 데이터 전처리 과정으로, 수집한 데이터 내에서 고객의 특성들을 나타내는 입력변수들을 추출하고 가공하였다. 세 번째 단계에서는 이러한 입력변수들을 이용하여 K-means 클러스터링을 통한 고객 군집화를 진행하였다. 고객 분석에 가장 최적화된 모델을 찾기 위해 군집 모델의 성능을 평가할 수 있는 지표인 실루엣 점수를 이용하여 최종 군집 모델을 도출하였다. 마지막 단계에서는 각 군집의 특성을 분석하고 시장세분화하여 각 군집에 맞는 마케팅 전략을 제시하였다. 위 내용을 도식화하면 <Figure 2>과 같다.



〈Figure 2〉 Flow Diagram of Proposed Method

3.1 데이터 수집

본 연구에서는 위메프를 이용하는 고객을 무작위로 추출하여 연구 대상으로 지정하였다. 특정 기간 내의 고객의 특성을 비교분석 하기 위해 2020년 1월부터 2021년 3월까지 작성한 리뷰들을 수집하였다.

데이터를 수집하기 위해서 웹 크롤링 기법을 사용하였다. 웹 크롤링은 웹 문서를 제공하는 웹 사이트의 내용을 자동으로 수집하는 기술이다[12]. 많은 양의 데이터를 쉽게 수집하기에 적합한 기법으로, 본 연구에서는 파이썬(Python) 셀레니움(Selenium) 라이브러리[36]를 이용하여 위메프 사이트 내의 리뷰 데이터를 수집하였다.

본 연구는 고객 한 명에 대한 10가지 정보를 하나의 데이터로 간주하고 1,250명의 고객데이터를 수집 및 저장하였다. 한 고객에 대한 수집 항목은 <Table 1>과 같다.

〈Table 1〉 Items Collected by Crawling

No.	Items
1	Customer ID
2	Customer ranking
3	The total number of reviews
4	The total number of helpful reviews
5	Product name
6	Product rating
7	The date of a review
8	The whether to register the image(o/x)
9	The contents of a review
10	The helpful number of a review

3.2 데이터 전처리

크롤링이 완료되면 수집한 고객데이터를 가공하는 작업이 진행된다. 가공된 데이터들은 모델링에 사용되는 입력변수들이며, 각 고객이 가지는 특성을 수치화한 값을 의미한다. K-means 클러스터링을 진행할 때 한 고객의 특성을 나타내는 입력변수 값들을 기준으로 군집화가 진행되기 때문에 입력변수 선정이 매우 중요하다[17].

본 연구에서는 <Table 1>에 나타나 있는 수집항목으로부터 도출할 수 있는 고객의 특성을 정의하여 입력변수를 선정하였다. 구체적으로, 고객에 대한 선호도(3개), 리뷰 별점 정보(6개), 리뷰 작성 주기 정보(3개), 리뷰의 신뢰도에 대한 정보(4개), 리뷰의 감성분석에 대한 정보(7개)와 관련된 23개의 입력변수가 이에 해당한다. 상세한 내용은 <Table 2>와 같다.

본 연구에서는 각 고객의 리뷰를 감성 분석하는 것이 고객의 중요한 특징이 될 수 있다고 판단하였다. 따라서, 리뷰 내용에 대한 감성 분석 결과를 활용하여 입력변수를 도출하고 전처리하였다. 감성 분석이란 텍스트 마이닝 기술의 한 분야로 온라인 텍스트 속에 내포된 감성을 추측하고 분류하기 위해 사용되는 기술이다[10].

본격적인 감성 분석을 진행하기 이전에 수집된 리뷰 데이터를 가공하기 위해 <Figure 3>와 같이 텍스트 전처리 작업을 수행하였다.

먼저, 한글 맞춤법 검사기 hanspell 라이브러리를 활용하여 문장 내에서 한글을 제외한 모든 문자를 제거한 뒤 잘못된 맞춤법 또는 띄어쓰기를 교정하였다. 이후 각 형태소에 따른 품사를 태깅하여 유의미한 단어 토큰을 선별하였

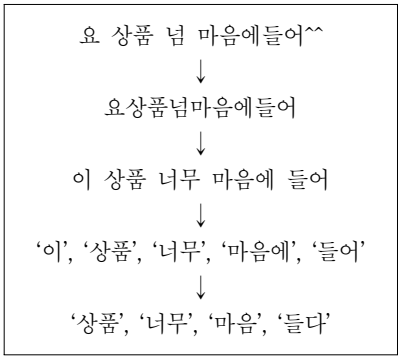
고 기존 단어를 표제어 처리하였다. 위의 단계를 거친 후, 전처리 된 텍스트를 감성 분석하기 위해 한글 말뭉치 사전인 KOSAC(한국어감성분석코퍼스) 감성사전을 이용하였다[11].

KOSAC 감성 사전은 뉴스 및 기사를 기반으로 개발된 사전으로서 n-gram 표제어 16,362개에 대하여 [긍정, 부정, 복합, 중립, 없음]으로 감성을 5가지로 구분한 후, 각 감성을 수치화하

〈Table 2〉 Description of the Features

Cateogry	Feature	Description
Basic preference	Customer ranking score	Whether a customer is in top 1000 rankings (0 or 1)
	Number of reviews	The number of reviews written by a customer
	Number of helpful	The number of helpfuls received by a customer
Rating	Share of 1 star	The share of 1-star rated reviews by a customer
	Share of 2 star	The share of 2-star rated reviews by a customer
	Share of 3 star	The share of 3-star rated reviews by a customer
	Share of 4 star	The share of 4-star rated reviews by a customer
	Share of 5 star	The share of 5-star rated reviews by a customer
	Average star rating	The average star ratings of reviews by a customer
Cycle of writing	The ratio of Share of reviews without ones written on the same date	The value obtained by dividing reviews without ones written on the same date by the number of reviews written by a customer
	Average writing cycle	The average cycle of writing reviews
	Period of reviews not written	A period during which a customer has not left reviews as of Mar. 31 2021 (When the latest review was posted)
Reliability	Share of reviews with helpfuls	The share of a customer's reviews given helpfuls of all his/her reviews
	Share of reviews with images attached	The share of reviews with images attached
	Share of reviews with no content	The share of reviews with no content
	Average review length	The average review length (Letter count)
Emotional analysis of reviews	Share of positive reviews	The share of positive reviews out of a customer's all reviews
	Share of negative reviews	The share of negative reviews out of a customer's all reviews
	Share of neutral reviews	The share of neutral reviews out of a customer's all reviews
	Positivity of reviews	The degree of positivity of all reviews by a customer
	Negativity of reviews	The degree of negativity of all reviews by a customer
	Neutrality of reviews	The degree of neutrality of all reviews by a customer
	Share of positivity/ negativity mismatch between ratings and comments	The share of 1- or 2-star rated reviews that turn out to be positive; and the share of 4- or 4-star rated reviews that turn out to be negative

여 확률분포로 나타낸다. 예를 들어 '기쁨'이라는 단어의 경우, [1,0,0,0,0]으로 나타내며, 긍정적인 감성을 강하게 내포하고 있음을 알 수 있다. 본 연구에서는 기존 KOSAC 감성 사전에 이커머스 리뷰에 자주 등장하는 500개 이상의 단어(반품, 파손, 불량품 등)를 별도로 추가하여 변형된 KOSAC 감성 사전을 구축하였다. 이 사전에 근거하여 리뷰 내의 한 문장에 대한 긍정, 부정, 중립 비율을 산출하였다. 이러한 감성 분석 결과들을 이용하여 총 7개의 '리뷰 내용 감성 분석' 입력변수 값들을 정의하였다.



〈Figure 3〉 The Example of a Text Tokenization Process

23개 입력변수에 대한 전처리를 완료한 후 모델링을 위한 데이터 정규화 작업을 진행하였다. 데이터 정규화란 데이터의 값을 비슷한 범위의 값들로 조정해주는 것을 의미한다[26]. 최대-최소 정규화 기법을 활용하여 모든 입력변수의 데이터들의 값을 [0,1] 범위 안으로 설정하였다. 본 연구에서는 식 (1)을 사용하여 정규화가 필요한 7개 입력변수(총 도움이 돼요 수, 별점 평균, 작성 주기 평균, 리뷰 작성하지 않은 기간, 리뷰 길이 평균(글자수), 리뷰어의 전체 댓글 수)에 대해 정규화 작업을 진행하였다.

$$x_i^{'} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

구체적으로, 한 개의 변수(컬럼) $x = [x_1, x_2, \ldots, x_n]$가 주어졌을 때, 한 요소 x_i와 x의 최소값의 차이를 x의 최대값과 최소값의 차이로 나누어 줌으로써 모든 데이터를 0과 1사이의 값으로 정규화하였다[9].

3.3 클러스터링

전처리가 완료된 입력변수 23개를 사용하여 클러스터링을 진행하였다. K-means 클러스터링 모델을 사용했으며, 다양한 군집 개수의 모델을 도출하고 결과를 확인하였다.

모델링을 진행한 후 모델링이 잘 되었는지 확인하기 위해 클러스터링 모델 성능 평가 지표인 실루엣(Silhouette) 점수를 사용하였다. 실루엣 점수란 군집화된 모델의 각 군집 간의 거리가 얼마나 효율적인지를 나타내는 지표로 -1.0~1.0 사이의 값을 가진다[7]. K-means 클러스터링에 사용되는 각각의 데이터는 하나의 실루엣 계수를 가지는데 이 값들의 평균이 모델의 실루엣 점수를 의미한다[31]. 하나의 데이터(i)가 가지는 실루엣 계수 $s(i)$를 구하는 방법은 식 (2)와 같다.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad -1 \le s(i) \le 1 \quad (2)$$

여기서, $a(i)$는 데이터(i)가 속해있는 군집 내의 데이터들과의 거리 평균을 의미하며, $b(i)$는 데이터(i)가 속해있지 않은 외부 군집 중에 가장 가까운 군집 내의 데이터들과의 거리 평균을 의미한다[15]. 예를 들면, 데이터(i)가 속해있지 않은 외부 군집 A, B가 있고, 데이터(i)

와 각 군집 내 데이터들과의 거리 평균값을 $A(i)$, $B(i)$라고 할 때, $A(i)$, $B(i)$ 중 최소값이 $b(i)$가 된다.

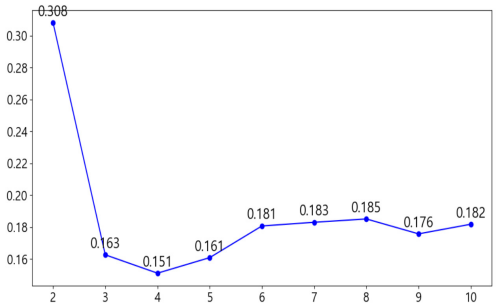
데이터(i)에 대한 $a(i)$와 $b(i)$값의 차이를 $a(i)$와 $b(i)$ 두 개의 값 중 큰 값으로 나누어주면 데이터(i)에 대한 실루엣 계수 $s(i)$를 구할 수 있다. 이렇게 구해진 모델 내의 모든 데이터의 $s(i)$ 값들의 평균값이 모델의 실루엣 점수가 되며, K-means 클러스터링 모델을 평가할 수 있다[15]. 한 집단 내의 동질성이 높고 서로 다른 집단 간의 이질성이 높을수록 실루엣 점수가 높으며, 점수가 높을수록 성능이 좋은 모델로 평가된다.

실루엣 점수를 확인한 후 모델의 시각화를 위해 주성분 분석(Principal component analysis; PCA) 차원축소를 진행하였다. 주성분 분석은 고차원의 데이터를 저차원의 데이터로 변환하는 기법을 말한다[35]. 데이터 프레임의 크기는 행과 열로 표현되는데 이때 하나의 "열(Column)"은 하나의 차원을 가리킨다. 주성분 분석은 전체 데이터의 유의미한 주요 특성을 보고자 하거나, 데이터의 일부 특성만으로 데이터의 차원의 수를 줄여 시각화하기 위해 활용된다[6]. 본 연구에서는 군집화 결과를 시각화하기 위해 모델링에 사용된 23개의 다차원 입력변수를 3차원으로 축소하였다.

4. 연구 결과

4.1 최종모델선정

본 연구에서는 시장세분화 최종 모델에 적합한 최적 군집 수를 선정하기 위한 연구를 진행하였다. <Figure 4>는 군집모델별 실루엣 점수(k = 2~10)를 나타낸다. <Figure 4>에 의하면, 군집 2개인 모델의 실루엣 점수가 가장 높은 것을 확인할 수 있다. 그러나 2개의 군집에 대한 시장세분화는 고객들의 특징을 구분 지어 정보를 제공하고자 하는 본 연구의 목적에 적합하지 않다고 판단하여 최종 군집 개수 후보에서 제외하였다. 따라서 실루엣 점수가 비교적 높은 군집 6개, 7개, 8개, 9개, 10개 모델을 최종 모델 후보로 고려하였다.



〈Figure 4〉 Silhouette Scores for Each Cluster Model(k=2~10)

최종 모델 후보군집 중 적절한 군집 개수를 선정하기 위해 k-군집모델에 속한 각 군집이 가지는 실루엣 점수의 표준편차를 고려하였다. 각 군집의 실루엣 점수의 표준편차가 낮다는 것은 군집들의 동질성의 편차가 크지 않다는 것을 의미하기 때문이다.

본문 3.3 절에서 설명한 것처럼 하나의 데이터가 가지는 실루엣 계수를 통해 모델의 실루엣 점수를 구할 수 있는 것처럼 모델 내 각 군집이 가지는 실루엣 점수를 구할 수 있다. k-군집모델 내의 각 군집(j)이 가지는 실루엣 점수를 S_j 라고 하고, S_j들의 평균을 $\overline{S}$ 라고 할 때, k-군집모델의

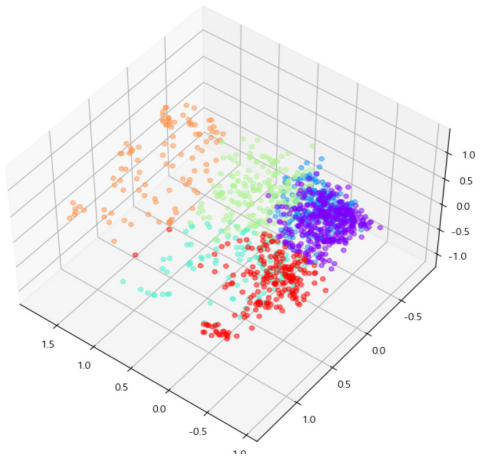
군집간 실루엣 점수 표준편차(SD_k)를 구하는 방법은 식 (3)과 같다.

$$SD_k = \sqrt{\frac{\sum_{j=1}^{k}(S_j - \overline{S})^2}{k-1}} \quad (3)$$

군집 개수가 6개, 7개, 8개, 9개, 10개일 때 SD_k의 값은 <Table 3>과 같으며 표준편차가 가장 낮은 군집 6개 모델이 시장세분화를 하기에 가장 적절한 모델이라고 판단하였다. 따라서 최종 군집 개수는 6개이며 주성분 분석을 통한 최종 모델 시각화 결과는 <Figure 5>과 같다.

〈Table 3〉 The Standard Deviation of the Silhouette Score for Each Model

k	6	7	8	9	10
SD_k	0.078	0.083	0.101	0.094	0.092



〈Figure 5〉 Visualization of Six Clusters via PCA

4.2 군집분석

본 연구에서는 최종 모델에 대한 각 군집의 특징을 정의하기 위해 입력변수별로 각 군집의 평균값을 산출하였다. 또한, <Table 4>와 같이 리뷰를 정량적으로 분석함으로써 뚜렷하게 나타나는 특징점을 찾아 군집의 특성을 파악하고 정의하였다.

제 1 군집에는 총 368명의 고객이 소속되었다. 해당 군집은 대체로 별점이 5점인 리뷰의 비율이 높지만, 전반적으로 리뷰 길이가 짧으며 내용이 없는 리뷰가 종종 발견되었다. 또한, '도움이 돼요' 받은 비율이 낮은 것으로 보아, 다른 고객들에게 의미 있는 정보를 제공하지 못하는 리뷰들이 많이 작성된 것으로 파악되었다.

제 2 군집에는 총 246명의 고객이 소속되었다. 해당 군집은 전반적으로 별점이 2, 3, 4점인 리뷰의 비율이 높고, 이미지 첨부 비율이 낮았다. 특히, '중복 날짜 제거 작성 비율'이 비교적 낮게 나타난 것으로 보아, 같은 날짜에 많은 리뷰를 작성하는 것으로 파악되었다.

제 3 군집에서는 총 135명의 고객이 소속되었다. 해당 군집은 전반적으로 별점이 3, 4점인 리뷰의 비율이 높고, 별점 5점인 리뷰의 비율이 비교적 낮은 것이 특징이었다. 또한, 리뷰 작성 수 비율이 낮지 않고 이미지 첨부 비율이 높은 것으로 보아 이 군집에 속한 고객은 상품의 구매 빈도가 높지만, 상품에 대해 만족하지 못하고 있는 것으로 파악되었다.

제 4 군집에는 총 191명의 고객이 소속되었다. 해당 군집은 별점 리뷰 비율과 긍정/부정/중립 비율이 비교적 다양한 것이 특징이었다. 또한, 작성 주기 평균이 높은 것으로 보아, 리뷰를 작성하는 빈도수가 낮으며 대체로 솔직한 리뷰를 작성한 것으로 파악되었다.

제 5 군집은 총 105명의 고객이 소속되었다.

해당 군집은 별점이 1, 2점인 리뷰의 비율과 부정 비율이 높은 것이 특징이었다. 또한, 리뷰를 작성하지 않은 기간이 길고, '중복 날짜 제거 작성 비율'이 높았으며, 전체 리뷰 개수가 많지 않았다. 이 군집에 속한 고객은 구매한 상품에 대한 만족도가 매우 낮을 때마다 리뷰를 작성한 것으로 파악되었다.

제 6 군집은 총 205명의 고객이 소속되었다.

〈Table 4〉 Average Value of Features by Cluster

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Customer ranking score	0.03	0.01	0.06	**0 (↓)**	**0 (↓)**	**0.14 (↑)**
Number of reviews	0.07	**0.08 (↑)**	0.05	0.02	**0.01 (↓)**	0.06
Number of helpful	0.01	0.01	0.02	**0 (↓)**	**0 (↓)**	**0.03 (↑)**
Share of 1 star	**0.02 (↓)**	0.05	0.16	0.24	**0.59 (↑)**	**0.02 (↓)**
Share of 2 star	0.03	0.07	0.09	0.18	**0.22 (↑)**	**0.02 (↓)**
Share of 3 star	0.05	**0.32 (↑)**	0.2	0.15	0.14	**0.05 (↓)**
Share of 4 star	0.09	0.4	**0.41 (↑)**	**0.11**	**0.02 (↓)**	**0.08**
Share of 5 star	0.8	0.16	0.14	0.32	**0.03 (↓)**	**0.83 (↑)**
The ratio of Share of reviews without ones written on the same date	0.42	**0.35 (↓)**	0.66	0.65	**0.94 (↑)**	0.64
Average writing cycle	0.13	0.1	**0.08 (↓)**	**0.14 (↑)**	**0.08 (↓)**	0.1
Period of reviews not written	**0.39 (↓)**	0.41	0.46	0.47	**0.59 (↑)**	0.45
Share of reviews with images attached	0.1	**0.05 (↓)**	0.77	0.13	0.33	**0.79 (↑)**
Share of reviews with no content	**0.03 (↑)**	0.02	0.01	0.02	**0 (↓)**	0.02
Average review length	**0.06 (↓)**	**0.06 (↓)**	0.14	0.11	0.16	**0.18 (↑)**
Share of positive reviews	0.82	0.81	0.85	0.57	**0.11 (↓)**	**0.86 (↑)**
Share of negative reviews	**0.09 (↓)**	0.11	0.11	0.32	**0.85 (↑)**	0.1
Share of neutral reviews	0.04	0.04	0.03	**0.06 (↑)**	0.04	**0.02 (↓)**

해당 군집은 별점이 5점인 리뷰의 비율이 높았으며, 긍정 리뷰의 비율이 높은 것이 특징이었다. 또한, 리뷰어 랭킹 점수가 높고, 이미지 첨부 비율이 높은 것으로 보아 상품 구매에 대한 만족도가 매우 높은 것으로 나타났다. 이 군집의 고객들은 다른 고객에게 유의미한 정보를 제공하는 리뷰를 다수 작성한 것으로 파악되었다.

4.3 시장세분화

본 연구에서는 군집을 분석한 결과를 바탕으로 고객을 총 6개의 시장으로 세분화하였으며, 각 시장에 적합한 마케팅 방안을 구상하였다. 시장세분화 결과는 <Table 5>와 같다.

첫 번째 세분 시장은 높은 별점을 주로 부여하지만, 반복되고 성의 없는 리뷰가 많았다는 점에 주목하였으며, '게으른 고객'으로 명명하였다. 해당 세분 시장에는 리뷰 자동완성 기능 또는 글자 수 제한 기능(최소 10글자 이상 작성)과 같은 간략한 원칙을 통해 리뷰 작성을 권고하는 마케팅을 제시한다[21].

두 번째 세분 시장은 리뷰를 작성 마감일에 몰아서 작성하고, 대체로 이미지 첨부를 하지 않는다는 점에 주목하였으며, '마감임박 벼락치기 고객'으로 명명하였다. 해당 세분 시장에는 구매 후 리뷰 작성 시점 또는 이미지 첨부의 여부에 따라 포인트를 차별적으로 지급하는 보상 정책을 실시하면, 양질의 리뷰를 얻을 수 있을 것이다[14].

세 번째 세분 시장은 별점 5점을 부여하는 것에 인색하지만 별점 3, 4점 부여 비율이 높고 리뷰작성에 비교적 성실하다는 점에 주목하여, '미리잡자 잠재 고객'으로 명명하였다. 해당 세분 시장에는 인공지능 기반 맞춤형 상품 추천 시스템 등을 제공하여 고객이 만족할 수 있는 상품을 구매할 수 있도록 유도하는 것이 중요하다[5].

네 번째 세분 시장은 별점은 다양하게 부여하며, 다양한 감정의 리뷰를 골고루 작성한다는 점에 주목하여, '논리정연 스마트 고객'으로 명명하였다. 해당 세분 시장을 제품 체험단으로 선정하여 다른 고객에게 유용하고 신뢰도 높은 리뷰를 제공하도록 하며 이에 따른 혜택을 제공하는 마케팅 방안을 제시한다[18].

다섯 번째 세분 시장은 낮은 별점을 주로 부여하며, 부정적인 리뷰를 주로 작성한다는 점에 주목하여, '주의경보 이탈 고객'으로 명명하였다. 해당 세분 시장에게는 VOC(Voice of Customer) 시스템을 적극적으로 이용하게 하여 실시간으로 피드백을 제공하면 고객의 이탈

〈Table 5〉 The Results of Market Segmentation

	Type	Characteristics
Segment 1	Laidback and easygoing	High star ratings, less thoughtful reviews
Segment 2	Procrastinating and cramming	Multiple reviews at all once, fewer images attached
Segment 3	Marketing prospect/lead	Less likely to comment "Helpful" and give 5 stars
Segment 4	Reasonable and well-informed	Large spread of star ratings and opinions shared
Segment 5	Highly likely to chrun	Low star ratings, critical reviews
Segment 6	Loyal and faithful	High star ratings, positive reviews

을 줄일 수 있을 것이다[20].

여섯 번째 세분 시장은 주로 높은 별점을 부여하며, 긍정적인 리뷰를 작성하고, 리뷰어 랭킹 점수가 높다는 점에 주목하여, '한결같은 충성고객'으로 명명하였다. 해당 세분 시장에는 멤버십 서비스를 통해 혜택을 제공하는 마케팅 방안을 제시한다[25].

5. 결 론

본 연구의 목적은 이커머스 시장의 고객들을 시장세분화하여 시장에 존재하는 고객의 유형을 정의하고, 각 시장에 적절한 고객관리 방안을 제시하는 것이다. 이를 위해 위메프 온라인 쇼핑 플랫폼의 고객 정보를 수집하여 가공하고, K-means 클러스터링을 통해 군집화를 진행하였다. 그 결과를 바탕으로 고객의 특징을 분석하여 총 6개로 시장세분화 하였다. 본 연구의 결과를 정리하면 다음과 같다.

첫째, 이커머스 시장의 고객은 K-means 클러스터링 모델을 사용하여 6개의 군집으로 군집화가 가능하다. 둘째, 6개의 군집은 그 특징에 따라 '게으른 고객', '마감임박 벼락치기 고객', '미리 잡자 잠재 고객', '논리정연 스마트 고객', '주의경보 이탈 고객', '한결 같은 충성 고객'으로 시장세분화 될 수 있다. 셋째, 본 연구의 시장세분화 결과를 바탕으로 충성고객, 잠재고객, 이탈고객에 대한 효과적인 마케팅 방안을 도출할 수 있다.

따라서 본 연구 결과는 다음과 같은 기대효과와 시장성을 가진다. 판매자인 기업은 세분화된 이커머스 시장의 고객 유형 정보를 바탕으로 효과적이고 다양한 고객관리 방안 도출이 가능하다. 이로 인해 기업은 고객관리를 위한 연구비용 감소와 잠재고객 발굴을 통한 맞춤형 고객관리 서비스 제공을 통한 수익을 창출할 수 있다.

하지만 본 연구는 다음과 같은 추후 연구가 필요할 것으로 보인다. 먼저, 감성 분석 결과에 대한 명확한 검증이 필요하다. 본 연구에서의 감성 분석은 감성 사전을 활용하여 도출된 긍정/부정 판별 결과이다. 감성 사전이 얼마나 정확한지와 한 댓글에 대한 감성 분석 결과가 실제 긍/부정 정도와 일치하는지에 대한 추가 연구가 필요하다. 다음으로 모델링 결과에 대한 더욱 명확하고 근거 있는 평가가 필요하다. 본 연구의 모델평가에 사용된 실루엣 지표만으로는 모델 성능을 평가하는 데에 한계가 있었다. 클러스터링 모델을 평가하는 다른 지표의 활용을 통해 모델 성능 평가에 명확한 근거를 추가해야 한다. 또한, 새로운 고객 정보가 모델링 결과와 얼마나 일치하는지에 대한 실증적인 연구가 추후 필요하다.

References

[1] Ahn, K. H., Limg, B. H., and Lee, Y. H., "The study of the selection of optimal variables and clustering method for the market segmentation," Journal of Marketing Management Research, Vol. 14, No. 3, pp. 157-176, 2009.

[2] Chatterjee, P., "Online reviews: Do consumers use them?," NA - Advances in Consumer Research, Vol. 28, pp. 129-133,

2001.

[3] Chevalier, J. A. and Mayzlin, D., “The effect of word of mouth on sales: Online book reviews,” Journal of Marketing Research, Vol. 43, No. 3, pp. 345-354, 2006.

[4] Choi, E. H., “The effect of afternote online on fashion brand attitude and brand equity,” The Korean Society of Clothing and Textiles, pp. 4-92, 2006.

[5] Choi, J. W. and Lee, H. J., “An integrated perspective of user evaluating personalized recommender systems: Performance-driven or user-centric,” The Journal of Society for e-Business Studies, Vol. 17, No. 3, pp. 85-103, 2012.

[6] Choi, K. B. and Nam, K. W., “Analysis of shopping website visit types and shopping pattern,” Journal of Intelligence and Information Systems Vol. 25, No. 1, pp. 85-107, 2019.

[7] Godwin, O. and Ugwoke, F. N., “Clustering algorithm for a healthcare dataset using silhouette score value,” International Journal of Computer Science & Information Technology (IJCSIT), Vol. 10, No. 2, pp. 27-37, 2018.

[8] Hur, S. H., Ryoo, S. Y., and Jeon, S. H., “Determinants of online review adoption: Focusing on online review quality and consensus,” Journal of Information Technology Applications & Management, Vol. 16, No. 4, pp. 41-58, 2009.

[9] Im, S. W. and Kim, B. S., “A study on the dimensionality reductional algorithm base on normalized mean impact value algorithm for regression models,” The Institute of Electronics and Information Engineers, pp. 1835-1837, 2020.

[10] Jeon, W. J., Lee, Y. B., and Geum, Y. J., “Airline service quality evaluation based on customer review using machine learning approach and sentiment analysis,” The Journal of Society for e-Business Studies, Vol. 26, No. 4, pp. 15-36, 2021.

[11] Kang, J. G. and Lee, K. S., “Study on yemeni refugees in Jeju Island viewed through text-mining: Focusing on Naver News comment,” Journal of Multi-Cultural Contents Studies, Vol. 30, pp. 103-135, 2019.

[12] Kang, K. S. and Park, S. M., “Keyword analysis of KCI Journals on business administration using web crawling and machine learning,” Korean Journal of Business Administration, Vol. 32, No. 4, pp. 597-615, 2019.

[13] Kim, J. W. and Choi, H. J., “Identification of playing styles for K-League football clubs through cluster analysis,” The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science, Vol. 23, No. 1, pp. 1-9, 2021.

[14] Kim, J. Y., Hou, W. S., and Kahn, H. S., “The power of online review: consumer evaluation based on online review types,” Journal of Product Reserach, Vol. 38, No. 4, pp. 21-30, 2020.

[15] Kim, S. S., Baek, J. Y., and Kang, B. S., “Group search optimization data clustering using silhouette,” The Korean Operations Research and Management Science Society, Vol. 42, No. 3, pp. 25-34, 2017.

[16] Kim, Y. C. and Lee, D. H., “Who are the internet shoppers?,” Journal of Consumer Studies, Vol. 13, No. 1, pp. 233-256, 2002.

[17] Kim, Y. J. and Park, H. W., “Cluster analysis of Players through Korean Women’s professional golf game records,” The Korean Society of Sports Science, Vol. 30, No. 2, pp. 1025-1032, 2021.

[18] Ko, S. S. and Kim, S. E., “Relationship among experiential marketing, brand trust and brand loyalty,” A Journal of Brand Design Association of Korea, Vol. 18, No. 2, pp. 5-16, 2020.

[19] Kwon, H. I. and Choi, Y. S., “A study on on-line game market segmentation classification and discrimination variable,” Journal of The Korean Society for Computer Game, Vol. 4, No. 14, pp. 53-61, 2008.

[20] Lee, J. H., “A study on service improvement through a Hotel VOC,” Ewha Womans University, pp. 1-86, 2019.

[21] Lee, K. A., “A study on measures to improve online consumer review systems,” Journal of Policy Analysis, pp. 1-96, 2016.

[22] Lee, S. W., “Comparison of initial seeds methods for K-means clustering,” Journal of Korean Society for Internet Information, Vol. 13, No. 6, pp. 1-8, 2012.

[23] Mudambi, S. M., and Schuff, D., “Research note: What makes a helpful online review? A study of customer reviews on Amazon.com,” MIS quarterly, Vol. 34, No. 1, pp. 185-200, 2010.

[24] Online shopping trends in December (including overseas direct online sales and purchase in the fourth quarter of 2021)”, Statistics Korea, last modified Feb 3. 2022, accessed Feb 7 2022, https://kostat.go.kr/portal/korea/kor_nw/1/12/3/index.board?bmode=read&bSeq=&aSeq=416587&pageNo=1&rowNum=10&navCount=10&currPg=&searchInfo=&sTarget=title&sTxt=.

[25] Paik, C. H., Kim, C. M., and Byun, H. J., “A study on the relationship among service quality of membership programs, customer satisfaction, and customer loyalty in Korean mobile telecommunications,” Korean Management Science Review, Vol. 23, No. 1, pp. 115-133, 2006.

[26] Park, B. R. and Ha, J. Y., “Geographical accessibility of seoul youth employment and welfare services according to the concentrated areas of youth,” Seoul Studies, Vol. 22, No. 1, pp. 17-38, 2021.

[27] Park, K. O., “A study on the effect of on-line customer review on purchase intention,” Pukyong National University, pp. 1-104, 2008.

[28] Park, K. W., Kim, D. W., and Ahn, H. S., “The impact of mobile commerce quality on customer satisfaction and repurchase intention: Focusing on moderat-

ing effect mobile familiarity," Journal Of Advanced Information Technology and Convergence, Vol. 15, No. 7, pp. 149-162, 2017.

[29] Priem, R. L., Li, S., and Carr, J. C., "Insights and new directions from demand-side approaches to technology innovation, entrepreneurship, and strategic management research," Journal of Management, Vol. 38, No. 1, pp. 346-374, 2012.

[30] Robinson, J., "The Economics of Imperfect Competition," Springer, 1969.

[31] Rousseeuw, P. J., "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, Vol. 20, pp. 53-65, 1987.

[32] Smith, W., "Product differentiation and market segmentation as alternative marketing strategies," Journal of Marketing, Vol. 21, pp. 3-8, 1956.

[33] Son, S. B. and Chun, J. H., "Product feature extraction an rating distribution using user reviews," The Journal of Society for e-Business Studies, Vol. 22, No. 1, pp. 65-87, 2017.

[34] Wind, Y., "Issues and advances in segmentation theory," Journal of Marketing Research, Vol. 15, No. 3, pp. 317-337, 1978.

[35] Wold, S., Esbensen, K., and Geladi, P., "Principal component analysis," Chemometrics and Intelligent Laboratory Systems, Vol. 2, No. 1-3, pp. 37-52, 1987.

[36] Wu, H., Liu, F., Zhao, L., Shao Y., and Cui, R., "Application research of crawler and data analysis based on python," International Journal of Advanced Network, Monitoring and Controls, Vol. 5, No. 2, pp. 68-74, 2020.

[37] Zhao, T., Nehorai, A., and Porat, B., "K-Means clustering-based data detection and symbol-timing recovery for burst-mode optical receiver," IEEE transactions on Communications, Vol. 54, No. 8, pp. 1492-1501, 2006.

저 자 소 개

김민경 (E-mail: kmg8561@tukorea.ac.kr)
2017년～현재 한국공학대학교 경영학부 IT경영학과 (학사)
한국공학대학교 스마트팩토리학과 (학사)
관심분야 데이터마이닝, 기계학습

허재석 (E-mail: jshuh@tukorea.ac.kr)
2013년 서울대학교 산업공학과 (학사)
2019년 서울대학교 산업공학과 (박사)
2019년～현재 한국공학대학교 경영학부 조교수
관심분야 스케줄링/디스패칭, 강화학습, 메타휴리스틱, 기계학습 산업응용

사애진 (E-mail: aejin518@tukorea.ac.kr)
2017년～현재 한국공학대학교 경영학부 IT경영학과 (학사)
한국공학대학교 스마트팩토리학과 (학사)
관심분야 그로스해킹, 퍼포먼스 마케팅, CRM 마케팅

전아름 (E-mail: j2665138@tukorea.ac.kr)
2017년～현재 한국공학대학교 경영학부 IT경영학과 (학사)
관심분야 데이터마이닝, 기계학습, 웹 개발

이한별 (E-mail: star004@tukorea.ac.kr)
2018년～현재 한국공학대학교 경영학부 IT경영학과 (학사)
한국공학대학교 스마트팩토리학과 (학사)
관심분야 데이터마이닝, 마케팅