

무인 스터디카페 거래 데이터 분석 및 머신러닝 분류 모델 기반 고객세분화*†

천 세 연(제1저자)

아주대학교 비즈니스애널리틱스학과 (석사)

박 옥 영(공동저자)

아주대학교 경영학과

강 주 영(교신저자)

아주대학교 e-비즈니스학과 (교수)

Customer Segmentation using Machine Learning Classification Model based on Unmanned Study Cafe Transaction Data Analysis

Cheon, Se Yeon(First Author)

Graduate School of Business, Ajou University (Master)

Park, Ok Young(Co Author)

Ajou University

Kang, Ju Young(Corresponding Author)

College of Business Administration, Ajou University (Professor)

* 본고는 한국서비스경영학회 2022년도 춘계학술대회발표에서 우수발표논문상으로 선정되었음.

† 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2021S1A5A2A01062441)

Abstract

A study cafe is a new form of learning environment that combines a study room and a cafe. Regarding the COVID-19's impact, the unmanned platform has become the new normal. The number of unmanned study cafes, launched by individuals, is rapidly increasing due to lower labor costs. Despite strong growth and promising market potential, little research on customer relationship management in unmanned study cafes has been conducted. Unlike franchises, the founder of an individual business must learn customers from the ground up. As a result, the purpose of this research is to provide a framework for individual founders to use in developing CRM strategies for their own unmanned study cafe. To accomplish this, analyze transaction data to better understand customers' untact transaction behavior and usage patterns, and also provide an example of customer segmentation approach using machine learning classification model so founders can plan practical CRM strategies based on the customer groups.

Keywords : Unmanned Study Cafe, CRM(Customer Relationship Management), Customer Segmentation, Machine Learning Classification Model

접수일(2022년 05월 23일), 수정일(2022년 06월 11일), 게재확정일(2022년 06월 12일)

I. 서 론

국내 독서실 산업은 불과 몇 년 전까지 하락세가 뚜렷했다(장상진, 2021). 한국교육개발원에 따르면, 전국 독서실 수는 2015년까지 하락세를 보였다. 이후 ‘카공족’의 등장과 함께 눈치 보지 않고 공부할 수 있는 카페인 ‘스터디카페’가 등장하며 가파르게 증가하기 시작했다. 스터디카페는 초기에 음료 제조, 서빙, 안내 등을 담당하는 카페 직원이 상주하는 형태에서 카페존(Cafe Zone) 셀프서비스와 타블렛PC, 키오스크 등 무인기기를 활용하는 ‘무인 스터디카페’로 변화했다.

점차 영향력을 키워가던 무인 스터디카페는 코로나19 팬데믹으로 언택트 문화가 만연하게 된 시장에서 더욱 그 기세를 펼쳤다. 무인 매장 프로그램 제공 기업 코보시스에 따르면, 코로나19 발생 이후 모든 무인 스터디카페 가맹점의 매출이 전년 동기간

대비 2021년 4월에 280% 증가한 것으로 집계되었다. 2022년 현재는 거리 두기 정부지침이 완화되며 등교하는 학생들이 늘어났음에도 스터디카페의 수요는 유지되어 코로나19 이후 국내 학습공간의 새로운 표준, 즉 뉴노멀로 자리매김했다고 말할 수 있다.

이러한 무인 스터디카페는 쉬운 창업 난이도와 소자본의 이점으로 1인 창업자에게 꾸준히 사랑받고 있는 아이템이다. 하지만 개인 창업자는 프랜차이즈 본사의 도움을 얻을 수 없어 팬데믹 등과 같은 사회, 시장의 큰 변화에도 적극적으로 대처하기 힘든 실정이다. 무인 스터디카페의 창업 관련 사이트를 살펴보면 입지선정, 인테리어 등 오픈 준비와 관련된 조언은 많은 것에 비해 오픈 이후 운영전략에 대한 조언은 찾아보기 힘든 것이 그 단적인 예이다.

또한, 무인 스터디카페의 운영과 관련하여 현재까지 진행된 고객관계관리 관련 연구는 거의 찾아볼 수 없었다. 이일한과 한수진(2019)의 ‘공간대여 서비스와 소비자의 일치성에 관한 연구’ 및 강경식(2021)의 ‘24시 스터디카페 이용 고객의 고객 만족과 재방문에 미치는 영향요인에 관한 연구’ 두 가지가 고객관계관리 연구 전반에 해당한다. 유통 산업 내에서는 무인 시스템과 고객 관련 연구가 학계와 산업계에서 다양하게 이루어지고 있는 것에 비하면, 공간임대업인 무인 스터디카페 연구는 그 수가 절대적으로 부족한 상황이다.

따라서 본 연구는 설문 조사 기반의 선행연구와의 차별성을 두어 무인 스터디카페 거래 데이터를 분석, 고객들의 비대면 거래 행동과 사용 패턴을 파악하고자 한다. 더 나아가 고객 특성 변수를 도출하고, 이를 기반으로 고객 그룹을 분류하여 프로모션 전략 등에 활용할 수 있는 고객세분화 정보를 제공할 것이다. 이는 고객들의 온라인 리뷰 데이터 분석 결과와 함께 창업자들을 위한 고객관계관리 전략의 가이드라인으로 제공하고자 하는 것이 연구의 최종 목적이다.

이를 위해 본 논문의 구성은 다음과 같다. 2절에서는 무인 점포와 서비스, 공간대여 서비스, 고객관계관리 및 데이터마이닝의 선행연구를 참고하여 그 개념을 정립하고자 한다. 3절에서는 무인 스터디카페 거래 데이터 수집과 현황 분석을 기반으로 한 고객 특성 변수 도출 과정, 머신러닝 분류 모델 성능 평가지표를 다루며, 추가로 리뷰 데이터의 활용도 및 수집 과정을 다룬다. 4절에서는 학습 및 검증 데이터 준비와 머신러닝 분류 모델 적용, 성능 평가와 더불어 온라인 리뷰 데이터 분석을 통해 무인 스터디카페에 대한 고객들의 평가 기준을 정립한다. 마지막 5절에서 연구 결과의 요약과 시사점 및 한계점을 언급하며 향후 연구 방향을 제시한다.

II. 관련 연구

2.1 무인 점포와 서비스

무인 점포는 디지털 정보 기술의 발전을 기반으로 이용자들의 편리성, 결제 간편성 등 이용 효율성 달성을 목적으로 개설이 확산되고 있으며, 인건비 상승과 코로나 확산에 따른 비대면 요건이 무인 점포 개점의 당위성을 강화하는 요인으로 작용하고 있다(이제원, 2021). 무인 점포는 셀프서비스기술을 기반한 비대면 방식으로 정보를 제공하고 상품과 서비스를 거래하는 방식을 의미하며 그 형태로는 키오스크, 태블릿 PC 등이 있다(안재영 등, 2020).

국내 키오스크 시장은 1999년 약 100억 원 규모에서 2021년 약 3000억 원 이상의 규모로 연평균 약 30%씩 성장하고 있다. 키오스크를 설치하면 직원이 최소 1.5명이 줄어드는 효과를 기대할 수 있기에 인건비 부담을 줄일 수 있다. 또한, 키오스크는 24시간 운영을 가능하게 하고 매출 집계의 전산화에 용이하기에 창업자들의 선호 창업 아이템으로 손꼽히고 있다(임하성 등, 2020). 더불어 코로나19로 인한 언택트 문화의 확산 등과 맞물려 무인화 추세를 더욱 가속했다. 2016년 타임카페 ‘타스’가 국내 최초로 24시 스터디카페로 출발하였고, 이후 조금씩 키오스크가 확산하기 시작했다(강경식, 2021).

이후 키오스크 기계의 단점을 해결한 모바일 플랫폼도 등장하기 시작했다. 고스카(GOSCA) 공식 홈페이지에 따르면 ‘고스카’는 전국 스터디카페 및 스터디룸의 좌석을 예약할 수 있는 서비스이다. 기존의 키오스크 방식이 아닌 태블릿형을 채택해 키오스크 기계의 가격부담까지 덜어 점주들에게 스터디카페 무인화 운영 솔루션을 제공한다. 또한, 고객들은 모바일에서 모든 예약과 좌석선택이 가능하며, 비밀번호로 무인 스터디카페에 입 출입이 가능하게 하는 등 더욱 편리한 서비스를 제공한다.

본 연구에서는 서울시 관악구에 위치한 ‘M 무인 스터디카페(이하 M카페)’의 모든 거래 데이터를 고스카 예약 시스템 내 데이터베이스에서 추출하여 분석에 활용하고자 한다.

<그림 1> 스터디카페 좌석 예약 플랫폼 고스카(GOSCA 공식 홈페이지, 2022)



2.2 공간대여 서비스

공간대여 서비스의 대표적인 예로 독서실과 같은 문화 공간이 있다. 2012년 피투피 시스템즈의 ‘토즈 스터디센터’를 시작으로 프리미엄 공간대여 서비스의 사업이 확장되어왔다. 최근에는 개인 독서실을 넘어 프리미엄 공간의 개념으로 그 의미가 확장되어 더 좋은 시설과 서비스를 제공하고 있다(이일한과 한수진, 2019).

한편, 스터디카페는 보통 공간대여 서비스 업종으로 등록되지만, 창업자들의 선호에 따라 개인 서비스업으로도 등록이 가능한 것으로 파악되었다. 본 연구에서는 이용 고객이 일정한 시간 동안 공간 사용을 목적으로 대금 결제를 진행한다는 관점에서 스터디카페를 공간대여 서비스의 일종으로 고려하였다.

<표 1> 공간대여 서비스와 소비자 관련 연구

주제	설명	저자
공간대여 서비스와 소비자의 일치성에 관한 연구	프리미엄 독서실이라는 공간대여 서비스를 구성하는 요소를 파악하여 이용하는 소비자의 일치성이 시설 만족도와 고객 만족에 미치는 영향 분석	이일한, 한수지(2019)
24시 스터디카페 이용 고객의 고객 만족과 재방문에 미치는 영향요인에 관한 연구	무인 24시 스터디카페에 대한 이용 고객이 느끼는 지각된 위험이 고객 만족과 재방문에 미치는 영향 분석	강경식(2021)
스터디카페 색채배치에 관한 청년층 선호 경향 연구	프랜차이즈 스터디카페의 실내 공간 디자인을 위해 청년층 선호도 조사	임미(2021)
무인 스터디카페 관리 시스템	무인 스터디카페를 통합적으로 관리해주고 부가적인 기능을 통해 사용자에게 편의성을 제공하는 안드로이드 애플리케이션을 개발	윤하영 외(2019)

<표 1>은 공간대여 서비스와 관련된 소비자 관련 연구들을 나열하고 있다. 먼저 이일한과 한수진(2019)의 ‘공간대여 서비스와 소비자의 일치성에 관한 연구’에서는 연구 대상 스터디룸의 공간에 대한 접근성, 내부환경요인, 서비스 내의 시설이 시설 만족도에 미치는 영향과 소비자의 자아 이미지 일치성, 기능적 일치성, 가치 일치성이 고객 만족에 미치는 영향을 분석하였으며, 마지막으로 시설 만족도와 고객 만족이 이용공간에 대한 애호도에 미치는 영향을 분석하였다. 또한 강경식(2021)의 ‘24시 스터디카페 이용 고객의 고객 만족과 재방문에 미치는 영향요인에 관한 연구’에서는 24시 무인 스터디카페의 물리적 환경 요소와 다양한 이용 요금제, 그리고 지각된 위험이 이용 고객의 만족과 재방문에 미치는 영향을 분석하였다.

두 연구는 모두 공간대여 서비스의 구성 요소와 이용 고객의 만족도에 미치는 영향에 주목하여 연구를 진행했다. 이는 각 연구에서 자체 제작한 설문을 통해 정보를 수집하여 분석한 것으로 고객의 인지와 평가에 그 초점이 맞추어져 있다. 이와 달리, 본 연구는 고객이 인지하지 못하고 있는 그들의 비대면 거래 행동과 서비스 이용 패턴을

분석하기 위해 무인 스터디카페의 예약 시스템 내 데이터를 분석한다. 최초 방문부터 그들의 이용권 구매 습관, 선호도, 패턴과 더불어 방문 및 예약 주기, 체류 시간, 선호 좌석 등의 행동 자체의 분석에 주목하여 연구를 진행했다.

한 편, 임미(2021)의 ‘스터디카페 색채배치에 관한 청년층 선호 경향 연구’에서는 프랜차이즈 스터디카페의 공간 색채 측정과 청년층의 공간 색채 선호도 조사를 통해 스터디카페의 공간 색채배치의 발전 방향을 도출하였다. 이는 스터디카페라는 학습공간에 대한 실내 건축 디자인적 접근의 연구이다. 윤하영 외(2019)의 ‘무인 스터디카페 관리 시스템’ 연구는 말 그대로 무인 스터디카페를 통합적으로 관리해주고 부가적인 기능을 통해 사용자에게 편의성을 제공하는 안드로이드 애플리케이션 개발에 그 목적을 두었다.

이렇듯 고객관계관리 측면에서 무인 스터디카페의 거래 데이터를 활용한 연구는 본 연구가 유일하며, 이를 충분히 활용하여 공간대여 서비스를 이용하는 고객들의 비대면 거래 행동과 이용 패턴을 파악한다면 비즈니스에 활용하는 새로운 방향성을 제시할 수 있을 것이다.

2.3 고객관계관리(CRM)와 데이터 마이닝(Data mining)

고객관계관리(Customer relationship management, CRM)는 고객과 장기적이고 수익성이 높은 관계를 구축하기 위해 비즈니스 전략을 지원하는 시스템과 그 프로세스를 의미한다. 즉, 고객관계관리는 기업이 고객을 확보하고 유지하며 충성도 및 고객 수익성을 개선하기 위해 유의미한 고객 행동을 이해하는 접근법이다(Ling and Yen, 2001). 따라서, 기업들은 고객세분화를 통해 고객을 이해하고, 이를 바탕으로 고객관리 및 마케팅을 수행하고자 한다(진서훈, 2005). 고객세분화는 고객 관점의 다양한 변수들을 활용하여 비슷한 특징을 갖는 고객들을 그룹으로 분류하는 것을 말하며, 이를 통해 기업은 세분화된 고객 집단에 대해 특정 마케팅 활동을 할 수 있다(김담희와 안가경, 2018).

고객관계관리 관점에서 고객 행동과 패턴을 분석하고, 고객세분화 및 고객 유지 관리에 쓰이는 강력한 도구가 데이터 마이닝이다(Cheng and Chen, 2009; 강한구 등, 2003). 데이터 마이닝은 방대한 데이터베이스에 숨겨져 있는 고객의 특성과 행동을 분류, 클러스터링, 예측 등의 모델로 추출하거나 탐지함으로써 기업이 이루고자 하는 고객관계관리 목표를 수행하기 위해 사용된다(Ngai et al., 2009; Giraud-Carrier and Povel, 2003).

본 연구에서는 M카페의 거래 데이터를 분석하여 고객의 비대면 거래 행동 및 무인 스터디카페 이용 패턴을 파악하고, 파악한 고객 특성 변수에 따라 머신러닝 모델을

활용한 고객세분화 과정을 통해 맞춤형 고객관계관리 전략에 도움이 되는 가이드라인을 제시하고자 한다.

머신러닝 기반 고객관계관리 연구는 오래전부터 진행되어왔으며, 고객세분화 분야에서는 의사결정나무와 같은 지도학습 모델이나 K-means 클러스터링과 같은 비지도 학습모델이 주로 사용되었다(chang, 2005). 한 편, 김은미와 박지영(2019)의 연구에 따르면 지금까지의 고객관계관리 연구는 고객의 신용상태를 우량인지 불량인지를 판별하는 이진 분류 등의 문제가 대부분이었으며, 이런 신용상태를 이진으로 분류하는 것보다 신용상태의 정도에 따라 여러 등급으로 고객을 분류함으로써 보다 나은 의사결정을 할 수 있다고 시사한 바 있다.

따라서 본 연구에서는 다분류 머신러닝 모델을 고객세분화 기법으로 활용하고자 하였다. 이를 위해 고객관계관리 연구에서 머신러닝 다분류 모델을 활용한 다음의 선행 연구들을 살펴보았다.

강한구 등(2003)은 의사결정나무(Decision Tree) 모델을 통해 보험회사의 고객 이탈 여부 분석을 진행하고자 ‘이탈확신’ 고객부터 ‘유지확신’ 고객까지 고객을 다섯 그룹으로 분류하여 서로 다른 마케팅 전략을 제시하였다. 의사결정나무는 모든 데이터의 집합인 루트 노드와 내부 노드, 그리고 리프 노드로 구성되어 있으며, 나무에서 정의한 의사결정에 따라 순차적으로 세분화되고, 리프 노드에 따라 각각의 클래스 레이블이 할당하는 분류 모델이다(Friedl and Brodley, 1997).

Ullah et al(2019)의 연구에서는 랜덤포레스트(Random Forest) 모델을 활용하여 통신 부문에서 유사한 고객 그룹을 파악하고, 각 그룹에 경쟁력 있는 서비스를 제공함으로써 고객을 유지하는 등 고객관계관리를 위한 이탈 예측 연구를 진행하였다. 랜덤포레스트는 여러 개의 랜덤 트리를 만들어서 그 결과들을 결합하여 데이터를 예측하거나 분류하는 기법이며 의사결정나무를 무작위로 만들어 과적합을 막고, 예측 오차를 줄어줄게 하는 장점을 가진 모델이다(Breiman, 2001).

Machado et al(2019)은 XGBoost와 LightGBM 모델을 이용하여 금융 기업의 신용카드 고객 충성도 분류하고, 예측하는 연구를 진행했다. 이 연구에서는 고객 충성도 예측뿐만 아니라 변수 중요도도 추가로 도출하였으며, 이를 바탕으로 기업이 효과적인 고객 충성도 프로그램을 설계하고 구현하는 데 도움이 된다고 밝혔다. XGBoost는 eXtreme Gradient Boosting의 약자로 Gradient Boosting은 잘못 예측된 데이터에 대해 경사 하강법을 이용하여 가중치를 업데이트하는 부스팅 알고리즘이다. 따라서 XGBoost는 높은 정확도와 과적합 문제를 해결하고 규모가 큰 데이터셋의 훈련 속도를 향상시킨다는 장점을 가지고 있다(Chen and Guestrin, 2016). LightGBM은 다른 일반 GBM 계열의 트리 분할 방법과 다르게 리프 중심 트리 분할(Leaf Wise) 방식을 사용하며, 균형 트리 분할 방식 (Level Wise)보다 예측 오류 손실을 최소화하여 빠른

훈련 속도와 더 높은 효율성을 장점으로 개선된 정확도를 제공한다(Ke et al., 2017).

신지안 외(2021)의 연구에서는 설명 가능한 정기에금 가입 여부 예측을 위해 우수한 예측 성능을 도출하는 의사결정나무 기반의 앙상블 학습 모델인 랜덤포레스트, GBM, XGBoost, LightGBM을 활용하였다. 해당 연구는 앙상블 학습 모델을 같은 연구 과제에 비교 적용함으로써 최적 모델을 찾고자 했다.

이처럼 고객관계관리 분야에서는 머신러닝 분류 모델을 활용하여 고객 그룹화 및 분류, 이탈 예측 등의 연구가 이루어지고 있다. 본 연구는 무인 스터디카페의 거래 데이터 중 고객 특성 변수를 추출하여 고객세분화를 위한 분류 모델 구축에 목표를 두고 있다. 앞선 선행연구들을 바탕으로 우수한 예측 성능을 도출하는 분류 모델인 의사결정나무, 랜덤포레스트, XGBoost, LightGBM 최종 모델로 선정하여 연구 과제에 비교 적용함으로써 무인 스터디카페의 고객세분화를 위한 최적의 모델을 찾고자 한다.

2.4 온라인 리뷰 분석

구매를 앞둔 고객은 소비에 대한 확신을 갖기 위해 다양한 정보를 탐색하여 지각되는 위험을 줄이려고 하며, 이를 위해 이미 제품이나 서비스를 경험해 본 사람이 제공하는 리뷰를 적극적으로 반영한다(Peterson, R. A. and Maria, C.M., 2003). 이에 온라인이라는 공개적이고 집단적으로 전달이 가능하다는 특성으로 고객들이 쉽게 리뷰에 접근 가능해지면서, 온라인 리뷰는 구매를 계획하고 있는 사람들에게 상품의 품질이나 서비스에 대한 기대를 형성한다(하현수 등, 2016).

또한, 고객이 자발적으로 작성하는 리뷰는 기업 입장에서 고객에게 개선된 서비스를 제공하기 위한 현재 서비스의 만족도나 기대사항 등을 파악하는 효과적인 전략 수단 중 하나이다(김도훈, 2019). 이러한 온라인 리뷰에 대한 연구로 조연주와 윤성준(2021)은 뷰티 소비자들의 온라인 리뷰를 통해 쇼핑 경험 가치가 제품 구매 시 긍정적인 태도 형성에 미치는 영향력을 살펴보기 위해 실증분석을 시행하였으며, 김도훈(2019)의 연구에서는 키즈폰 제품에 대한 리뷰 데이터를 다양한 사이트에서 수집하여 소비자의 전반적인 생각과 제품의 트렌드를 파악하고 종합적으로 고객이 느끼는 장단점을 파악하여 제품 개선 방향을 제시했다.

이처럼 온라인 리뷰를 통해 제품과 서비스에 대한 분석을 진행한 연구는 다수 존재하며, 이 연구들이 시사하는 바는 온라인 리뷰가 다른 고객이 제품이나 서비스 등을 선택하는 데 있어 중요한 영향을 준다는 점과 운영자 관점에서 서비스 개선을 위한 방향성을 제안해주는 지표가 된다는 점이다.

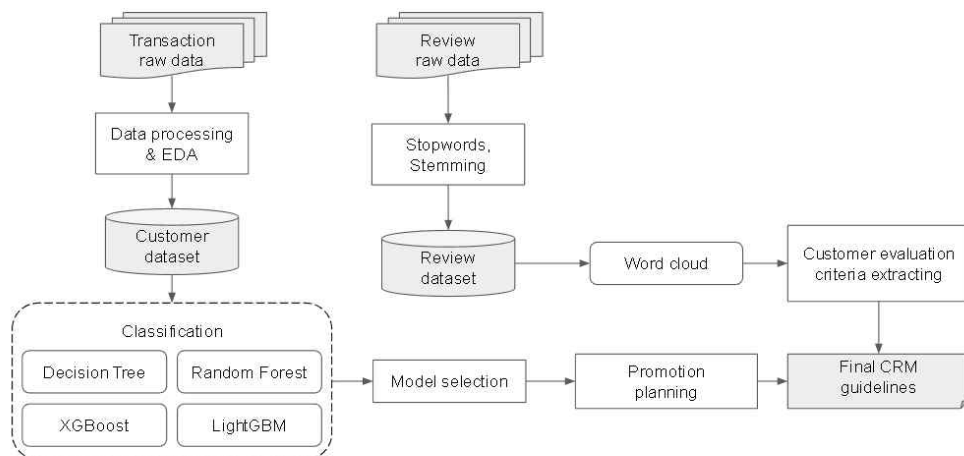
따라서 본 연구에서도 연구 대상 스타디카페와 경쟁사 스타디카페의 온라인 리뷰를 수집하여 고객들의 스타디카페 평가 기준을 파악하고, 개선 방안을 고안하는 데 활용하고자 한다.

Ⅲ. 연구방법

3.1 연구 개요

2장에서 언급한 바와 같이 시장의 괄목할만한 성장세와 긍정적 전망에도 무인 스타디카페의 실질적인 고객관계관리 관련 연구는 찾아보기 힘들다. 본 연구의 목표는 실제 무인 스타디카페를 운영하는 개인 창업자들을 위해 거래 데이터를 분석하여 고객들의 비대면 거래 행동을 파악하고, 고객 특성 변수들을 도출하여 머신러닝 분류 모델을 구축하며, 고객세분화 결과와 온라인 고객 리뷰 분석 결과를 함께 제시하여 고객관계관리 전략의 가이드라인으로써 제공하는 것이다. 본 연구는 <그림 2>와 같은 절차로 진행되며, 다음의 세 가지 주제를 다루고 있다.

<그림 2> 연구절차도



먼저 무인 스타디카페 거래 데이터를 통해 고객들의 비대면 거래 행동 및 이용 패

턴을 파악하고자 한다. 독서실, 카페 등 기존의 스터디 공간과 달리 무인 스터디카페 고객의 경우 모든 구매, 예약, 좌석 이용 데이터가 예약 시스템상에 기록된다. 이를 통해 비대면 공간대여 서비스에서 고객들의 거래 행동과 이용 패턴을 발견할 수 있다면, 이는 고객관계관리 차원에서 무인 스터디카페 운영자에게 큰 인사이트로 작용할 것이다. 거래 행동 및 이용 패턴 분석을 진행하기 위해 M카페가 사용하고 있는 예약 시스템 고스카의 데이터베이스에서 1년간의 데이터를 추출하였다. 거래 데이터 내 변수들을 분석하고, 선행연구를 참고하여 고객 그룹 분류에 활용될 고객 특성 변수들을 선정하였다.

다음으로 앞서 진행한 데이터 분석과 선행연구를 기반으로 고객세분화를 위한 고객 특성 변수를 도출하여 머신러닝 분류 모델을 구축하는 것이다. 다양한 선행연구에서 고객관계관리와 데이터마이닝의 중요성에 대해 언급하였고, 고객세분화 및 전략 수립에서 머신러닝 분류 모델을 활용한 사례들을 발견할 수 있다. 이를 참고하여 다른 고객 그룹을 구분 짓는 고객 특성 변수를 도출하고 네 가지 머신러닝 분류 모델에 적용한 뒤 비교하여 최적의 분류 모델을 선정하였다. 본 연구에서는 매출 기여도를 기준으로 고객 그룹 분류를 진행하여 세분화된 고객 그룹별 프로모션 차별화 전략을 수립하는 등 실질적인 도움이 될 것으로 예상된다.

마지막으로 온라인상의 고객 리뷰 수집 및 분석을 통해 무인 스터디카페에 대한 전반적인 평가 기준과 본 연구의 대상인 M카페 및 경쟁사 고객의 피드백을 반영하여 고객관계관리 측면에서의 전략을 제시한다. 선행연구를 통해 온라인 리뷰가 구매를 계획하고 있는 사람들에게 서비스에 대한 기대를 형성하고, 기업 입장에서는 현재의 제품이나 서비스를 개선하기 위한 방향성을 제안해주는 하나의 지표가 된다는 것을 파악할 수 있었다. 따라서 본 연구에서는 전국 무인 스터디카페와 M카페 및 경쟁사의 고객 온라인 리뷰를 수집하고, 해당 리뷰들을 정제한 뒤 워드클라우드를 표현하여 리뷰에 쓰인 주요 키워드들을 확인하고자 한다. 이를 통해 무인 스터디카페 내부 거래 데이터로 파악한 고객들의 비대면 거래 행동 및 이용 패턴과 외부의 온라인 리뷰 분석 결과를 종합하여 개인 창업자가 고객관계관리 측면에서 본 연구를 가이드라인으로 활용할 수 있을 것이라 기대한다.

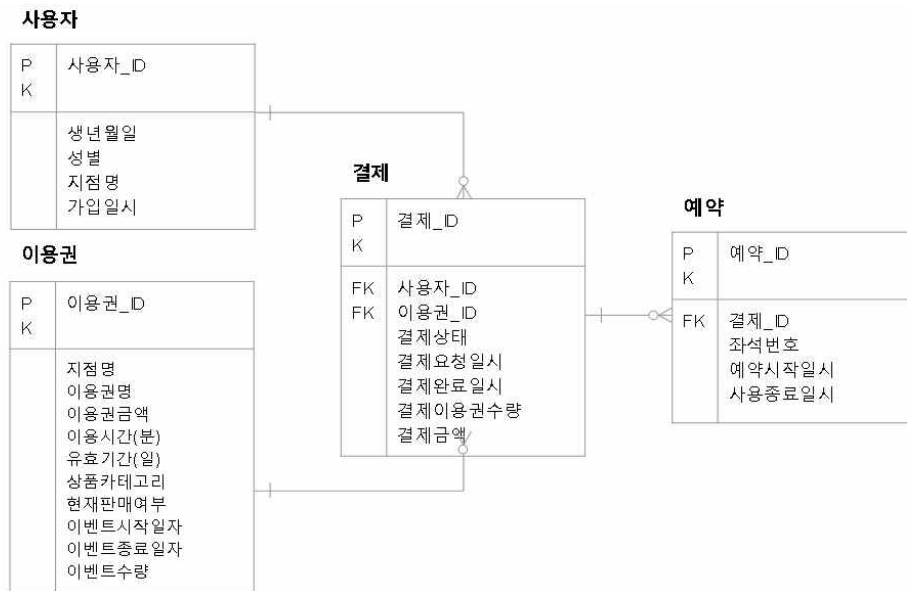
3.2 데이터 수집 및 변수 생성

3.2.1 M카페 DB 데이터

M카페 예약 시스템 데이터베이스에서 추출 가능한 모든 데이터를 첫 번째로 수집하였다. 데이터베이스에는 회원 정보, 구매 정보, 좌석 예약 정보, 이용권 정보의 총 네 가지 테이블이 존재한다. 2020년 4월 1일부터 2021년 3월 31일까지 M카페 오픈

초기 약 1년간의 데이터를 SQL로 추출했으며, 미가공 데이터(Raw data) 기준 각각 약 1200명의 회원, 약 3000건의 구매, 약 20000건의 좌석 예약, 약 40개의 이용권 정보를 확보하였다.

<그림 3> M카페 예약 시스템 개체 관계도



<그림 3>은 M카페 예약 시스템 내 테이블 간 관계를 알 수 있는 개체 관계도(Entity Relationship Diagram, ERD)이다. ‘사용자’ 테이블은 ‘사용자_ID’를 기본 키(Primary Key)로 가지며, 카페를 1회 이상 이용한 사용자의 생년월일, 성별, 가입일시, 지점명 정보가 있다. ‘이용권’ 테이블은 ‘이용권_ID’를 기본 키로 하며 이용권명, 금액, 이용시간, 유효 기간, 상품 카테고리 및 현재 판매 여부와 그 밖에 이벤트 관련 정보를 확인할 수 있다. ‘결제’ 테이블은 사용자 및 이용권 테이블의 기본 키가 외래 키(Foreign Key)로 연결되어 어떤 사용자가 어떤 이용권을 결제했는지, 그 결제 상태와 결제 요청·완료 일시, 수량 및 결제 금액을 확인할 수 있다. 결제 테이블의 기본 키인 ‘결제_ID’는 ‘예약_ID’를 기본 키로 한 ‘예약 테이블’의 외래 키로 연결되어 고객이 예약한 좌석과 예약 시작 및 사용 종료 일시를 확인할 수 있다. 각 테이블의 데이터를 더욱 자세히 살펴보고, M카페 이용자들의 거래 현황 분석을 진행했다.

3.2.2 M카페 고객 거래 데이터 현황

확보한 M카페 데이터는 인구통계학적 정보가 Null 값인 회원 정보를 기준으로 정제되었으며, 이상치를 제거한 후 고객 거래 현황 분석을 진행했다. <표 2>는 정제된 M카페 데이터의 네 가지 테이블을 활용한 현황 분석 결과이다.

<표 2> M카페 고객 거래 데이터 현황

구분		빈도(명)	비율(%)
성별	남자	307	47.82
	여자	335	52.18
연령	10대	169	26.32
	20대	339	52.8
	30대	106	16.51
	40대	21	3.27
	50대	5	0.78
	60대	2	0.31
월별 실제 이용 고객 수	2020년 4월	66	4.65
	2020년 5월	127	8.95
	2020년 6월	166	11.7
	2020년 7월	187	13.18
	2020년 8월	138	9.73
	2020년 9월*	101	7.12
	2020년 10월	154	10.78
	2020년 11월	131	9.23
	2020년 12월	117	8.25
	2020년 1월	84	5.92
	2020년 2월	72	5.07
	2020년 3월	77	5.43
	200회 이상	3	0.47
총 예약 횟수	150회 이상 200회 미만	8	1.25
	100회 이상 150회 미만	14	2.18
	50회 이상 100회 미만	48	7.48
	10회 이상 50회 미만	183	28.5
	10회 미만	386	60.12
이용권 총 구매 횟수	일일권	1,464	51.06
	시간권	489	11.68
	기간권	579	22.99
	사물함	335	16.42
	일일권 및 사물함	77	22.99
	시간권 및 사물함	55	16.42
	기간권 및 사물함	203	60.6
	이벤트 이용권	72	3.56

M카페 회원은 1회 이상 이용을 기준으로 해당 기간 약 642명으로 파악되었다. 여자 회원 수가 남자 회원 수의 110%를 차지하며, 연령대에서는 20대가 절반 이상으로 가장 많았고 그 뒤로 10대, 30대가 뒤를 이었다. 월별 실제 이용 회원 수는 2020년

4월 개업 이후 5, 6, 7월에 187명까지 늘어났다. 방학 기간인 8월에 138명, 9월에는 사회적 거리 두기 정부지침의 영향으로 101명까지 내려갔다가 10, 11, 12월까지는 다시 100명 이상을 유지했다. 2021년 1, 2월은 방학과 코로나 19 확진자 수 급증의 영향으로 70명대까지 내려갔으며 3월에는 상승이 예상되었으나 2월과 비슷한 수준을 유지했다. 동 기간 내 총 예약횟수는 10회 미만이 60% 이상으로 가장 많았으며, 10회 이상 50회 미만이 약 28% 이상을 차지했다. 동 기간 내 이용권 총 구매횟수는 일일권이 약 51%로 가장 많았으며 기간권이 약 20%, 시간권이 약 17%로 가장 적었다. 사물함을 구매한 횟수는 335건으로 모두 다른 이용권과 함께 구매한 것으로 파악되었다. 함께 구매한 이용권은 기간권이 약 60% 이상으로 가장 많았고, 그다음이 약 23%의 일일권이며 시간권이 약 16%로 가장 적은 것을 알 수 있었다.

3.2.3 최종 변수 구성

본 연구의 분석 모형 구축을 위한 변수 및 데이터셋 구축에는 고객관계관리 선행연구를 참고하였다. 이영진 등(2010)의 연구에서는 고객관계관리 분석을 통한 고객세분화 모형 수립에 필요한 변수로 기본적인 고객 번호, 성별, 나이 등의 인구통계학적 변수와 최근 구매 시기, 총 구매 수량, 평균 구매 금액을 설정하여 최종 분석 데이터셋을 구축하였다고 밝혔다. 또한, 고객세분화 작업 시 가장 흔하게 활용되는 점수부여방식의 고객관계관리 분석 기법인 RFM을 참고하였다(조광현·박희창 2012; 정운필·염창선, 2013). 해당 기법은 얼마나 최근에 구매했는지(Recency), 얼마나 자주 구매했는지(Frequency), 얼마나 구매했는지(Monetary) 등 세 가지 기준으로 고객의 수익성을 평가하고, 세분화할 수 있는 모델이다(김동석, 2021).

본 연구에서는 선행연구들의 고객세분화 변수를 참고하여 <표 3>과 같이 머신러닝 분류 모델을 위한 최종 변수를 구성하였다.

<표 3> 최종 변수 구성

변수명	변수 설명	변수 예제	선행연구
uid	사용자 id	19,460	이영진 외(2010)
age	나이	17	이영진 외(2010)
usex	성별	1	이영진 외(2010)
last_payment	가장 최근 결제일로부터 지난 시간(일)	74	조광현 외(2012), 김동석(2021)
last_visitday	가장 최근 방문일로부터 지난 시간(일)	69	조광현 외(2012), 김동석(2021)
total_pdmoney	총 구매 금액	13,500	조광현 외(2012) 김동석(2021)
month	시험 기간 방문(월) 횟수	1	M카페 EDA
event	이벤트 이용권 구매 여부	0	이재용 외(2012)

먼저 ‘사용자 id’를 비롯하여 기본 인구통계학적 변수인 ‘나이’, ‘성별’을 포함하였다. 또한, RFM 기법을 참고하여 사용자의 ‘가장 최근 결제 일로부터 지난 시간(일)’을 R(Recency), ‘총 구매 금액’을 M(Monetary)으로 보고 분석 변수를 생성하였다. F(Frequency)의 경우 사용자의 이용권 구매 횟수로 고려할 수 있지만, M(Monetary)의 ‘총 구매 금액’과 관련도가 높아 최종 변수에서 제외하였다. 그 밖에도 결제 행위에 이어 예약 행위가 발생 된다는 점, 무인 스터디카페의 이용 패턴에서 가장 최근 방문 일자가 고객의 충성도를 판단하기에 더욱 적합하다는 점에 착안하여 R(Recency)에 해당하는 변수로 ‘가장 최근 방문 일로부터 지난 시간(일)’을 추가 생성하였다.

‘시험 기간 방문(월) 횟수’ 변수는 앞선 탐색적 데이터 분석에서 파악한 월별 이용 빈도 차이가 고객의 무인 스터디카페 이용 패턴에 따라 달라진다는 점을 들어 추가로 생성하였다. 이는 학교에 다니며 시험 기간인 4~6월과 10~11월에 방문 빈도, 체류 시간이 늘어나는 중·고등학생 및 대학생과 나머지 직장인 혹은 불특정 다수의 패턴을 보이는 고객들의 무인 스터디카페 이용 패턴에 차이가 있다는 점을 근거로 하였다.

마지막으로 M카페에서는 기존 이용권의 가격할인, 기존 이용권과 사물함 이용권을 함께 구매 시 할인하는 이벤트 이용권을 기간 한정 및 선착순으로 판매한 이력이 존재한다. 이재용 등(2012)의 연구에서는 기간 한정 및 선착순 가격할인이 고객으로 하여금 구매 의도를 증가시킨다고 밝힘에 따라, 본 연구에서는 해당 이벤트 이용권이 고객 구매에 미치는 영향을 고려하기 위해 ‘이벤트 이용권 구매 여부를 변수로 설정하였다.

3.3 분류 모델 성능 평가 지표

본 연구에서는 의사결정트리, 랜덤포레스트, XGBoost와 LightGBM 모델을 통해 고객 등급을 분류하고자 하며, 과적합을 막고 높은 정확도를 가진 모델을 구축하고자 각 모델의 튜닝 파라미터를 조정하여 최적의 모델을 선정하였다. 그 후에는 모델을 평가하는 지표를 통해 각 모델들을 비교하여 고객 등급 분류를 가장 잘한 모델을 선정하고자 한다.

머신러닝 분류 모델의 평가는 오차행렬(Confusion matrix)를 기반으로 구성되며, 이는 구축된 모델로 예측을 수행한 후에 종속변수 값이 올바르게 예측된 경우와 틀리게 예측된 경우의 개수를 나타낸 표이다. 이 오차행렬을 바탕으로 <표 4>와 같이 분류 모델의 성능 평가 지표인 적중률, 정밀도, 민감도, f1-score를 구할 수 있고, 이를 통해 가장 우수한 모델을 선정하고자 한다(이재식, 2020).

<표 4> 분류 모델 성능 평가 지표

Measure	Fomula
적중률(Accuracy)	$\frac{TP + TN}{P + N}$
정밀도(Precision)	$\frac{TP}{TP + FP}$
민감도(Recall)	$\frac{TP}{P}$
f1-score	$\frac{2 \times precision \times recall}{precision + recall}$

3.4 온라인 리뷰 데이터

본 연구는 무인 스터디카페 고객들의 거래 데이터 분석과 함께 온라인에서 고객 리뷰 데이터를 수집하고 분석하여 운영자에게 고객관계관리 관점에서의 전략을 제공하는 것을 목표로 한다.

고객 유입 및 서비스 개선이라는 고객관계관리의 전략을 제시하기 위해 본 연구에서도 온라인 리뷰데이터를 수집하고 분석함으로써 고객의 무인 스터디카페에 대한 전반적인 생각과 트렌드를 파악하고, 서비스 개선 방향 제시 및 타 사업장과의 차별점을 부각시키고자 한다.

고객 온라인 리뷰 데이터셋을 확보하기 위해 먼저 M카페의 네이버 리뷰와 경쟁사 스터디카페 리뷰 전체를 수집하였으며 각각 105건, 136건을 확인하였다. 또한, 본 연구의 대상인 M카페가 개업한 2020년 4월부터 거래 데이터 분석에 활용한 기간인 1

년 후까지를 기준으로 ‘스터디카페’를 키워드로 한 네이버 블로그 및 네이버 카페에 올라온 리뷰 총 1000여 개를 수집하였다.

수집된 각각의 리뷰들을 워드 클라우드로 만들어 가장 많이 등장하는 키워드 빈도를 통해 고객의 평가 기준을 확인하였다. 또한, 스터디카페의 리뷰가 올라온 네이버 카페 이름도 함께 수집하여 카페 이름으로 대표되는 스터디카페 이용 고객 자체의 특성을 함께 조사하였다. 수집된 블로그와 카페의 리뷰는 연구 결과에 워드클라우드 이미지로 첨부하였다.

IV. 연구 결과

4.1 분류 모델 구축 및 결과

4.1.1 학습 및 검증 데이터 준비

고객 분류를 위해서 앞서 구축한 데이터셋의 고객 나이, 성별, 시험 기간 방문 횟수, 이벤트 이용권 구매 여부, 최근 결제일로부터 지난 시간, 최근 방문일로부터 지난 시간을 독립변수로 설정하였으며, 고객 매출순으로 고객 등급을 나누어 등급별 모형을 생성했다는 윤성준(2005)의 연구에 따라 본 연구에서는 구매 금액별 고객 등급을 종속변수로 설정하였다. M카페 운영자 의견에 따라 구매 금액 상위 20%의 고객을 등급 1로, 구매 금액 상위 50%에 속하는 고객을 등급 2로, 구매 금액 하위 30%의 고객을 등급 3으로 설정하였다. 고객 등급별 분포 현황은 아래 <표 5>를 통해 확인할 수 있다.

<표 5> 고객 등급별 분포 현황

고객 등급	정의	고객 수	비율(%)
1	구매 금액 180,500 이상인 상위 20% 고객	137	21.77
2	구매 금액 180,500 미만, 38,500 이상인 상위 50% 고객	316	50.11
3	구매 금액 38,500 미만인 하위 30% 고객	177	28.12

머신러닝 모델 구축 단계에 앞서, 데이터 스케일링을 통해 데이터 전처리를 시행하였다. 그 후, 모델 학습을 위한 학습 데이터셋과 모델 성능을 평가하는 검증 데이터셋을 만들고자 전처리된 데이터셋을 7:3의 비율로 나누었으며, 학습 데이터는 442개,

검증 데이터는 189개로 구성되어 있다.

4.1.2 머신러닝 분류 모델 적용 및 성능 평가

2장에서 언급한 고객관계관리 선행연구들을 통해 의사결정트리, 랜덤포레스트, XGBoost, LightGBM의 네 가지 분류 모델을 선정하였다. 학습 데이터셋을 활용하여 네 가지 분류 모델을 각각 학습한 모델을 구축했으며, 검증 데이터셋을 적용한 예측 결과를 아래 <표 6>의 오차행렬로 정리했다. 또한, 분류 모델별 성능 평가 지표를 적용한 결과는 다음의 <표 7>과 같다.

<표 6> 머신러닝 분류 모델별 오차행렬

모델	의사결정나무			랜덤포레스트			XGBoost			LightGBM		
예측 실제	등급 1	등급 2	등급 3	등급 1	등급 2	등급 3	등급 1	등급 2	등급 3	등급 1	등급 2	등급 3
등급 1	23	7	0	22	8	0	22	8	0	23	7	0
등급 2	12	75	7	11	75	8	8	66	20	9	65	20
등급 3	1	19	45	1	17	47	1	11	53	0	18	47

<표 7> 머신러닝 분류 모델별 성능 평가 지표 적용

모델	accuracy	precision	recall	f1-score
의사결정나무	0.7566	0.7489	0.7523	0.7451
랜덤포레스트	0.7619	0.7505	0.7523	0.7481
XGBoost	0.7460	0.7374	0.7503	0.7423
LightGBM	0.7143	0.7142	0.7271	0.72019

본 연구에서는 모델의 분류 정도를 나타내는 지표로 적중률, 정밀도, 민감도, f1-score를 선택하였다. 적중률은 총 데이터에서 실제 고객 등급과 모델이 예측한 고객 등급이 일치한 데이터 수의 비율이다. 정밀도는 모델이 예측한 데이터 중에 실제값으로 정확히 분류한 지표이고, 민감도는 실제 데이터값 중 모델이 올바르게 예측한 지표이며, f1-score는 정밀도와 민감도를 활용한 모델 평가 지표이다. 이러한 지표들을 중

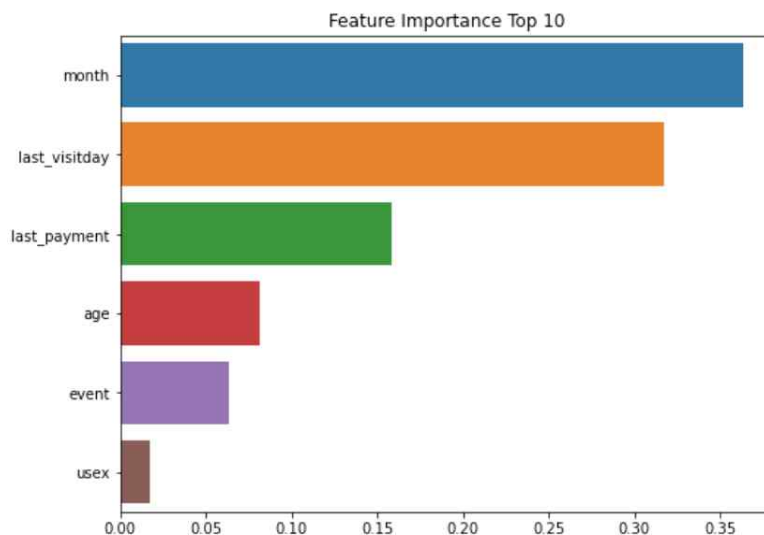
합하여 모델을 평가하였을 때, 랜덤포레스트, 의사결정나무, XGBoost, LightGBM 순으로 나타났으며 이에 따라 랜덤포레스트를 최종 모델로 선정하였다.

아래 <그림 4>는 랜덤포레스트의 변수 중요도를 나타낸 그래프이다. 이 그래프를 통해 고객 등급 분류에 사용된 변수 중요도는 시험 기간 방문(월)횟수가 약 35% 이상으로 가장 높았음을 알 수 있다. 나이는 고객 등급 분류에 기여한 정도는 비교적 낮지만, 스터디카페를 방문하는 연령대 중 10~20대가 가장 높았음을 고려하여 운영자는 시험 기간과 관련된 고객 프로모션을 고안할 수 있다.

방문일로부터 지난 시간(일)과 결제일로부터 지난 시간(일)이 그다음으로 모델에 영향을 많이 끼친 것으로 나타났다. 이는 이용권을 구매하고 소진하는 고객의 무인 스터디카페의 구매 및 이용 행동에 따라 결정되는 변수들이다. 두 변수는 RFM 기법에서 R(Recency)에 해당하는 지표이다. 즉, 마지막 방문일이 최근일수록 무인 스터디카페 방문주기가 짧다고 예상할 수 있으며, 자연스레 이용권 소진 횟수가 늘어남에 따라 이를 위한 이용권 구매가 뒤따라 매출에 영향을 미친다고 해석할 수 있다.

성별은 고객 등급 분류에 기여한 정도가 약 2% 정도로 미비함을 알 수 있다. 이는 스터디카페라는 공간 특성을 미루어 보았을 때, 방문하는 고객이 성별에 따라 방문 의도가 달라지지 않음을 짐작할 수 있다. 따라서 운영자는 성별에 따른 프로모션 보다는 시험 기간 방문 횟수나 방문일, 결제일 주기 혹은 연령별로 제공하는 프로모션을 고안하여 매출 극대화에 기여할 수 있을 것이다.

<그림 4> 랜덤포레스트 모델 변수 중요도



4.2 리뷰 분석 결과

아래의 <표 8>은 M카페와 경쟁사 네이버 블로그 리뷰를 워드클라우드로 만든 결과이며 <표 9>는 ‘스터디카페’를 키워드로 한 네이버 블로그 및 카페 리뷰와 네이버 카페와 이름을 워드클라우드로 만든 결과이다.

<표 8> M카페와 경쟁사 네이버 블로그 리뷰



<표 9> 네이버 카페와 블로그 리뷰



M카페 블로그 리뷰에서는 ‘공부’, ‘독서실’, ‘공간’, ‘분위기’, ‘집중’, ‘존’, ‘인테리어’, ‘오픈’, ‘무인’, ‘좌석’, ‘학습’, ‘프리미엄’, ‘자리’, ‘처음’, ‘느낌’, ‘시설’, ‘가구’ 등 키워드를 확인하였다. 경쟁사 스터디카페 리뷰 분석 결과에서는 ‘노트북’, ‘공간’, ‘공부’, ‘존 (Zone)’, ‘자리’, ‘선택’, ‘화장실’, ‘어플’, ‘문’, ‘백색소음’, ‘무인’, ‘커피’ 등의 키워드를 확인하였다.

네이버에 ‘스터디카페’ 검색 시, ‘공무원’, ‘수만회’, ‘공사’, ‘경찰’, ‘특목고’, ‘신도시’, ‘학원’, ‘과학고’ 등이 카페 이름에 포함되어있는 네이버 카페에서 글이 많이 작성되었으며, 네이버 블로그 리뷰에서는 ‘팬찮은’, ‘독서실’, ‘추천’, ‘도서관’, ‘집중’, ‘코로나’, ‘좋은’, ‘시간’, ‘시설’ 등의 키워드를 확인할 수 있었다.

워드클라우드를 통해 파악한 고객들의 스터디카페 평가 기준은 크게 학습공간의 쾌적도 및 인테리어, 학습에 필요한 물품 제공 여부, 화장실 편의성, 카페 서비스의 네 가지로 구분할 수 있었다. 이는 모두 공부에만 집중할 수 있는 환경을 이루기 위해 고객들이 고려하는 사항으로 볼 수 있다. 경쟁사 리뷰를 통해 M카페에는 없거나 상대적으로 부족한 서비스를 파악하고, 현재 M카페의 리뷰에서 장단점을 보완한다면 고객들의 스터디카페 이용 만족도를 높이는 데 기여할 수 있을 것이다.

V. 결 론

무인 스터디카페 산업이 나날이 발전하는 것에 비해 이와 관련된 전략적인 연구나 논문은 찾아보기 어렵다. 이에 따라 본 연구는 무인 스터디카페 내 고객과의 최초 접점인 태블릿 예약 시스템 ‘고스카’를 통해 실증적인 데이터 분석을 진행하였으며, ‘M카페’ 이용 고객의 전반적인 특성을 파악하였다.

그 후, 파악한 고객 특성을 통해 ‘시험 기간 방문 횟수’와 ‘이벤트 이용권 구매 여부’ 변수를 생성하였으며, 고객세분화 기법인 RFM을 바탕으로 ‘최근 결제일로부터 지난 시간’, ‘최근 방문일로부터 지난 시간’, ‘총 구매 금액’이라는 변수를 생성하였다. 그 중, 고객의 총 구매 금액을 바탕으로 고객을 등급으로 나누었으며, 이를 종속변수로 설정하고 분류 모델을 이용하여 고객 그룹을 분류하는 방법을 제시했다. 분류 모델의 평가를 위해 오차행렬 기반 성능 평가 지표를 활용하였으며, 랜덤포레스트 모델의 정확도가 가장 높았다. 유의미한 변수로는 ‘시험 기간 방문 횟수’와 ‘최근 방문일로부터 지난 시간’, ‘최근 결제일로부터 지난 시간’을 꼽을 수 있었다. 이를 통해 M카페 운영자는 고객 등급별 프로모션에 차별성을 고안할 수 있을 것이다.

또한, 크롤링을 통해 본 연구의 대상인 M카페뿐만 아니라 경쟁업체와 전국 스터디카페의 리뷰를 수집하여 워드클라우드 시각화를 진행하였다. 워드클라우드를 통해 파악한 고객들의 스터디카페 평가 기준은 크게 학습공간의 쾌적도 및 인테리어, 학습에 필요한 물품 제공 여부, 화장실 편의성, 카페 서비스의 4가지로 구분할 수 있었다. 이러한 고객들의 스터디카페 평가 기준과 함께 M카페의 경쟁업체와 현시점의 ‘M’카페 서비스를 비교하여 객관적으로 평가할 수 있었다.

이를 통해 본 연구가 기여하는 바는 다음과 같다. 첫째, 무인 스터디카페의 실제 거래 데이터를 이용하여 고객들의 비대면 거래 행동과 사용 패턴을 파악할 수 있었다. 1년간의 거래 데이터 속 고객 현황, 인구통계학적 정보를 비롯하여 고객들의 이용권 구매, 좌석 예약 및 무인 스터디카페 사용 행동 특성을 살펴볼 수 있었다. 둘째, 고객 특성 변수를 추출하고 구매 금액별 고객 등급을 종속변수로 머신러닝 분류 모델을 적용하여 고객세분화를 결과를 확인하였다. 이를 기반으로 추후 매출 증대 및 충성 고객층을 생성을 위한 고객관계관리 전략에 차별성을 부여하는 등 무인 스터디카페 운영에 실질적 도움이 될 것이다. 특히, 본 연구에서 사용한 분류 모델의 독립변수로 새로운 고객 행동 특성 데이터를 추가 혹은 변경하거나, 새로운 고객 데이터 사례를 추가하여 예측 가능하다는 점에서 높은 활용 가능성을 내포하고 있다. 셋째, 네이버 블로그 및 네이버 카페 리뷰의 워드클라우드를 분석을 통해 고객이 M카페를 이용하며 만족하거나 아쉬웠던 점을 키워드로 볼 수 있었으며, 무인 스터디카페에 대한 전반적인 고객 평가 기준을 확인할 수 있었다. 운영자는 앞으로 무인 스터디카페 서비스 보완 및 유지를 위한 기준으로 삼고 안정적인 운영을 위한 비즈니스 전략을 수립할 수

있다.

반면, 다음과 같은 시기적 유의점이 있다. 본 연구에서 활용된 데이터는 M카페 오픈 초인 20년도 4월부터 21년 3월까지의 데이터이다. 이는 전 세계를 강타한 코로나 19 전염병의 발병이 얼마 되지 않아 사회 시스템의 혼란과 개인의 심적 혼란이 가중된 시점과 일치한다. 특히 데이터의 수집 기간 중 정부의 사회적 거리두기 조치로 스터디카페의 운영상 제한이 있었으며, 국내 확진자 현황에 따라 해당 지역구 내에서 수시로 거리 두기 내용에 변화가 있었음에 유의해야 한다. 특히 2020년 8월 31일부터 9월 13일까지 2주간은 M카페가 위치한 관악구에서 서울·수도권 사회적 거리 두기 2단계(5단계 거리 두기 개정 전) 및 강화된 사회적 거리두기 2단계가 각 1주씩 진행되었다. 스터디카페에 해당하는 M카페는 동기간 집합금지 조치를 진행하여, 해당 2주간의 데이터는 공란이며 코로나 19 이후 서울·수도권에서 가장 강한 거리 두기의 실시로 동기간 전후의 매출 감소 등의 이슈에는 고객들의 심리적 영향이 크게 작용했을 것으로 보인다. 시기적 유의점에도 불구하고, 사회적으로 소비자들의 행동 방식이 급격히 변화한 시기에 유의미한 분석을 통해 무인 스터디카페의 고객 행동 특성을 연구하고자 하였다. 한편 최근에는 정부지침 등으로 제한되었던 고객 행동반경이 다시 회복되며 연구 시점과 다소 다른 결과를 보일 수 있다.

마지막으로, 본 연구는 M카페를 분석 대상으로 선정하였기에 단일 매장 데이터만을 활용한 분석을 진행하였다는 점에서 한계가 있다. 즉, M카페만의 고유한 상황에서 수집 가능한 데이터를 활용하여 분석 및 연구가 진행되었기 때문에 분석 결과 및 분류 모델을 타 매장에 적용하기 어렵다는 한계점이 있다. 추후 더욱 다양한 환경 속에서 운영되는 무인 스터디카페들을 대상으로 연구하여 또 다른 시사점을 제시할 여지가 있다.

참고문헌

- 강경식 (2021), 24시 스터디카페 이용고객의 고객 만족과 재방문에 미치는 영향요인에 관한 연구, 석사학위논문, 중앙대학교 산업창업경영대학원
- 강한구, 김규곤, 강창완, 최승배, 조성기 (2003), 보험 CRM에서 데이터마이닝 기법의 응용연구, *Journal of The Korean Data Analysis Society (JKDAS)*, 5(2), 307-321.
- 김담희, 안가경 (2018), 머신러닝을 이용한 고객세분화에 관한 연구, *융복합지식학회 논문지*, 6(2), 115-120.
- 김도훈, 차경진 (2019), AI 키즈폰의 소비자리뷰 분석을 통한 제품개선 전략에 대한 연구, *한국전자거래학회지*, 24(2), 71-89.
- 김동석 (2021), RFM 모형의 가중치 선택에 관한 연구, 석사학위논문, 제주대학교 대학원.
- 김은미, 박지영 (2019), 온라인 P2P 환경에서 딥러닝을 적용한 다분류 기반 개인신용등급 예측모형, *인터넷전자상거래연구*, 19(4), 43-57.
- 신지안, 문지훈, 노승민 (2021), 설명 가능한 정기예금 가입 여부 예측을 위한 앙상블 학습 기반 분류 모델들의 비교 분석, *한국전자거래학회지*, 26(3), 97-117.
- 안재영, 이중정, 배다은, 이소현 (2020), 언택트 서비스 사용에 관한 연구: 키오스크 사례 중심으로, *인터넷전자상거래연구*, 20(4), 49-73.
- 윤성준 (2005), 데이터마이닝 기법을 통한 백화점의 고객이탈예측 모형 연구, *ASIA MARKETING JOURNAL*, 6(4), 45-72.
- 윤하영, 민희수, 신동진, 이용수, 김정준 (2019), 무인 스터디카페 관리 시스템, *한국인터넷방송통신학회 논문지*, 19(3), 1-7.
- 이영진, 강창완, 김규곤, 최승배 (2010), CRFM 모형을 이용한 고객세분화와 모형평가, *Journal of The Korean Data Analysis Society*, 12(6), 3283-3293.
- 이일한, 한수진 (2019), 공간대여 서비스와 소비자의 일치성에 관한 연구: 매니저와 소비자의 비교를 중심으로, *예술인문사회 융합 멀티미디어 논문지*, 9(11), 1253-1267.
- 이재식 (2020), *데이터 애널리틱스 :예제로 배우는 모델링 기법, 데이터 분석, 모델 구축 과정*, 위키북스.
- 이재용, 이철성, 이호배 (2012), 가격할인제품판매에서의 회소성 메시지 효과: 수량한정 및 시간한정 가격할인을 중심으로, *경영학연구*, 41(6), 1591-1618.
- 이제원 (2021), 무인점포의 고객만족과 신뢰, 관계몰입이 충성도에 미치는 영향 연

- 구, *고객만족경영연구*, 23(2), 89-109.
- 임미 (2021), *스터디카페의 선택배치에 관한 청년층 선호 경향 연구*, 석사학위논문, 한양대학교.
- 임하성, 류두진, 박대현 (2020), 키오스크 산업 분석: 도입 효과와 시장 전망, *Korea Business Review*, 24(1), 21-48.
- 장상진 (2021), "독서실이 돌아왔다, 카페·스터디룸 갖춘 어른 공간으로", *조선일보*, 2021년 06월 22일자.
- 정운필, 염창선 (2013), 고객세분화를 통한 인터넷 쇼핑몰의 고객관계관리, *한국정보기술학회논문지*, 11(12), 159-167.
- 조광현, 박희창 (2012), CFD 기법을 이용한 부동산 회원 분류에 관한 연구, *Journal of The Korean Data Analysis Society*, 14(2), 717-726.
- 조연주, 윤성준 (2021), 뷰티 소비자의 온라인 경험가치가 소비자행복과 브랜드신뢰 및 구매의도에 미치는 영향: SNS 구전정보특성의 조절효과를 중심으로, *한국창업학회지*, 16(1), 23-48.
- 진서훈 (2005), 데이터마이닝에 의한 고객세분화 개발, *응용통계연구*, 18(3), 555-565.
- 하현수, 박연옥, 김건휘 (2016), 소비자의 긍정적 온라인 리뷰는 레스토랑의 선택속성과 재방문의도 간에 조절효과 연구, *관광레저연구*, 28(11), 357-373.
- Breiman, L. (2001), Random Forests, *Machine Learning*, 45(1), 5-32.
- Chang, N.-S. (2005), Improving the effectiveness of customer classification models: a pre-segmentation approach, *Information Systems Review*, 7(2), 23-40.
- Chen, T. and Guestrin, C. (2016), XGBoost: A Scalable Tree Boosting System, Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA.
- Cheng, C.-H. and Chen, Y.-S. (2009), Classifying the segmentation of customer value via RFM model and RS theory, *Expert Systems with Applications*, 36(3, Part 1), 4176-4184.
- Friedl, M. A., and Brodley, C. E. (1997), Decision tree classification of land cover from remotely sensed data, *Remote sensing of environment*, 61(3), 399-409.
- Giraud-Carrier, C. and Povel, O. (2003), Characterising Data Mining software,

- Intelligent Data Analysis*, 7, 181–192.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. (2017), Lightgbm: A highly efficient gradient boosting decision tree, *Advances in neural information processing systems*, 30, 3146–3154.
- Ling, R. and Yen, D. C. (2001), Customer Relationship Management: An Analysis Framework and Implementation Strategies, *Journal of Computer Information Systems*, 41(3), 82–97.
- Machado, M. R., Karray, S. and Sousa, I. T. d. (2019), *LightGBM: an Effective Decision Tree Gradient Boosting Method to Predict Customer Loyalty in the Finance Industry*, Paper presented at the 2019 14th International Conference on Computer Science & Education (ICCSE).
- Ngai, E. W. T., Xiu, L. and Chau, D. C. K. (2009), Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2, Part 2), 2592–2602.
- Peterson, R. A., and Maria, C. (2003), *Consumer Behavior*, In: Upper Saddle River, NJ: Prentice–Hall, Inc.
- Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U. and Kim, S. W. (2019), A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector, *IEEE Access*, 7, 60134–60149.

*** 저자소개 ***

· **천 세 연**(cjstpdu9@ajou.ac.kr)

아주대학교 경영대학원 비즈니스애널리틱스학과에서 재학 중에 있다. 주요 관심 분야는 빅데이터 분석, 머신러닝 등이다.

· **박 옥 영**(dhfpswntm@ajou.ac.kr)

아주대학교에서 경영학사를 취득하였다. 주요 관심분야는 서비스품질이다.

· **강 주 영**(jykang@ajou.ac.kr)

현재 아주대학교 경영대학 e-비즈니스학과 교수로 재직 중이며, 포항공과대학교 컴퓨터공학과에서 학사, 서울대학교 컴퓨터공학과에서 석사, 한국과학기술원 경영공학전공에서 공학박사학위를 취득하였다. 주요 관심분야는 빅데이터, 텍스트마이닝, 시맨틱 웹, 지능형 정보시스템 등이다.