

온라인 무료 샘플 관측의 효과적 활용을 위한 기계학습 기반 고객분류예측 모형*

원하람** · 김무전*** · 안현철****

〈목 차〉

I. 서론	IV. 실증 분석
II. 이론적 배경	4.1 실험 데이터
2.1 고객관계관리	4.2 실험 설계
2.2 고객분류예측을 위한 기계학습기법	4.3 실험 결과
2.3 무료 샘플 관측	4.4 표적 마케팅 전략 모색
III. 제안 모형	V. 결론 및 시사점
	참고문헌
	<Abstract>

I. 서론

Kotler(2002)가 지적한 바와 같이 오늘날 네트워크 및 정보기술의 발달과 기업 간 기술 평준화로 인하여 시장 전체를 고객으로 삼는 대량생산시대는 종말을 맞이하고 있으며, 이제는 시장과 고객에 대한 적절한 세분화가 기업들의 주요한 과제로 부상하고 있다(조대현 등, 2009; 원지성, 2013). 정보기술의 발전은 기업 환경에도 영향을 끼쳤다. 고객 관련 정보와 거래실적

등을 데이터베이스로 구축하게 됨으로써 고객 성향, 구매실적 등을 분석할 수 있는 기반이 형성되었고, 고객의 구매행태 분석을 통하여 고객에 대한 차별적인 마케팅이 가능하게 되었다(김상수, 문준연, 1998; 조대현 등, 2009). 이러한 기업 환경과 사회의 변화에 따라 ‘어떻게 정보를 이용하여 새로운 고객을 발굴, 확보하며, 기존 고객의 이탈을 막고 유지할 수 있는가?’에 대한 질문이 기업에게 중요한 과제가 되었으며, 그것을 구현하는 방법론인 고객관계관리

* 본 논문은 2018 경영정보관련 춘계 통합학술대회에서 우수논문상을 수상한 저자들의 발표논문을 보완 및 발전시킨 논문임.

** 국민대학교 비즈니스IT전문대학원, haramy44@kookmin.ac.kr(주저자)

*** 국민대학교 비즈니스IT전문대학원, kim@dhckorea.com

**** 국민대학교 비즈니스IT전문대학원, hcahn@kookmin.ac.kr(교신저자)

(customer relationship management: CRM)가 오늘날 기업의 중요한 경영방식으로 자리잡고 있다(서현석, 주형준, 2011).

기업들은 CRM 시스템을 통하여 얻은 정보로부터 고객을 체계적으로 분석할 수 있으며, 이를 통하여 차별화된 마케팅 전략을 사용함으로써 기존 고객 유지 및 신규 고객 확보 등 긍정적 효과를 기대할 수 있다. 이러한 이유로 고객 관계관리에서는 고객의 정보를 분석하는 것이 중요한데, 이 때 분석 방법으로 널리 사용되는 방법이 데이터 마이닝(data mining) 및 기계학습(machine learning) 기법들이다(조대현 등, 2009). 이러한 데이터 마이닝이나 기계학습 기법들은 기업들로 하여금 방대한 기업의 데이터 베이스 속에 숨겨진 지식을 발견하고 그 속에 숨겨진 잠재 가치를 포착하여 이를 경영의사결정을 위한 정보로 활용할 수 있도록 지원한다(김신곤, 1999; Jiawei et al., 2011).

데이터 마이닝 및 기계학습 관점의 고객관계 관리에 있어 중요한 주제 중 하나는 ‘기업에서 팔고자 하는 상품을 구매할 가능성이 높은 잠재 구매자를 발굴’ 하는 고객 분류 모형(customer classification model)을 구축하는 것이다(안현철, 2013). 고객 분류 모형은 다양한 마케팅 전략 개발과 기회 창출에 활용될 수 있으며, 특히 세분화 개념에 기초하여 세분 집단의 다양한 욕구에 맞게 차별적인 마케팅을 제공하는 표적 마케팅(target marketing)에 유용하게 적용될 수 있다(홍태호, 박지영, 2009).

본 연구는 이러한 CRM의 관점에서 온라인 무료 샘플 판촉 대상을 효과적으로 발굴할 수 있는 기계학습 기법 기반의 고객분류예측 모형을 제안한다. 아울러 제안된 모형의 적용 가능

성을 확인하기 위해, 국내 한 화장품 회사의 사례에 제안 모형을 적용해 보고 그 성능을 실증적으로 검증해 보고자 한다.

본 논문은 다음과 같이 구성된다. 우선 2장에서는 관련된 기존 문헌들을 통하여 이론적 배경을 살펴보고, 3장에서는 본 연구의 제안 모형을 설명한다. 4장에서는 앞서 제시한 모형을 검증하기 위한 실험 데이터 구성과 실험 설계를 설명한 다음, 실험 결과도 함께 제시할 것이다. 끝으로 5장에서는 결론과 함께 시사점과 연구의 한계점이 함께 제시된다.

II. 이론적 배경

2.1 고객관계관리

오늘날 다양한 성향의 고객들이 다양한 니즈(needs)를 갖고 있는 상황에서, 기업의 경쟁력은 개별 고객이 원하는 제품과 서비스를 얼마나 신속하게 제공할 수 있는 지에 달려 있다. 따라서 시장 및 고객에 대한 깊이 있는 이해와 함께 고객이 원하는 것을 신속히 파악하여 구매로 연결하고, 연결된 고객과의 관계를 지속적으로 유지하는 것이 대단히 중요해졌다. 이처럼 개별 고객과의 접점 및 관계 유지를 중요시하는 방법론이 바로 고객관계관리이다(김주호, 1998; 박주석, 2000).

고객관계관리는 다수의 기업과 학자들에 의해 다음과 같이 정의되고 있다. 먼저 가트너 그룹(Gartner group)은 “신규 고객획득, 기존 고객 유지 및 고객 수익성을 증대시키기 위하여 지속적인 커뮤니케이션을 통해 고객 행동을 이

해하고 영향을 주기 위한 광범위한 접근”으로 정의하고 있으며, 프로세스의 관점에서 Rajiv et al.(2001)는 “판매자 또는 서비스 제공자가 고객과의 장기적인 관계를 구축하고 끊임없이 변화하는 고객의 니즈를 따라가기 위해 고객의 기대를 관리하는 프로세스”라고 정의하였다. 한편 한국 SW 산업협회에서는 “기업이 보유하고 있는 고객 데이터를 수집, 통합, 가공, 분석하여 고객 개개인의 특성에 맞게 마케팅 활동을 계획 수행, 평가, 수정하는 일련의 과정”이라고 정의하였다.

이러한 정의들로 미루어 볼 때, 고객관계관리에 있어 가장 중요한 것은 고객 정보이다. 고객 정보는 다양한 고객 접점에서 획득되는 고객에 관한 사항들을 가공한 것을 말하며 이름, 연락처와 같은 기본적인 고객 데이터와 고객과의 거래를 통해 얻어지는 거래 데이터, 고객의 불만사항, 질문 등과 같은 피드백 정보 등이 포함된다. 기업은 이러한 정보 분석을 통하여 고객 지식을 얻을 수 있다. 고객 지식은 고객에 대한 깊이 있는 정보를 말하며 고객 요구, 만족도, 충성도, 수익성, 관계 단계, 고객평생가치(LTV) 등을 포함한다. 고객 정보에서 고객 지식을 얻어내기까지는 상당 기간의 고객과의 상호작용이 필요하다(정현주 등, 2002).

인터넷의 발달과 더불어 90년대 후반 이후 인터넷을 활용한 e-CRM 방법론이 개발되었다. e-CRM은 온라인상에서 고객관리를 통해 고객 충성도를 높이고, 고객과의 거래를 통한 수익을 최대화하려는 일련의 활동이다. 특히 e-CRM은 마케팅, 판매, 서비스 지원 및 지식 관리의 차원에서 고객과의 관계 지속에 대한 비즈니스 프로세스의 자동화 및 효과 증진에 초점을 두고

있다. 이처럼 인터넷 기반으로 CRM이 확장됨에 따라 기존의 CRM은 프로세스 상의 고객관리로 발전하였고, 고객이 직접 주는 정보뿐만 아니라, 로그 데이터 분석을 통하여 고객에 대한 부가적인 데이터를 얻을 수 있게 되었다(김종완, 오기욱, 2002).

2000년대 들어서 정보통신기술의 고도화와 모바일 기기의 확산으로 전반적인 인터넷 접근성이 높아졌으며, 개인화된 정보의 생산이 확대되고 있다. 이에 따라 기업의 분석 대상이 세분 집단에서 개인으로 한층 세밀해짐으로써 수익률과 충성도가 높은 고객을 더욱 정확하게 선별할 수 있게 되었고, 더 차별적인 마케팅이 가능하게 되었다.

기업이 활용할 수 있는 마케팅 전략을 크게 두 가지로 나누면 대량 마케팅과 표적 마케팅으로 나눌 수 있다(Kotler, 2002). 대량 마케팅이란 시장 전체에 표준화된 제품 및 서비스를 제공하는 마케팅 전략을 의미한다. 대량 마케팅에서는 개별 소비자의 선호도는 중요하지 않으며 시장 규모와 평균적 특성을 찾고 그에 적합한 제품을 대량 생산하여 신속하게 공급하는 것이 핵심이다.

표적 마케팅은 소비자 전체 집단을 동질적인 여러 개의 하위 집단으로 구분하고 각각의 세분 집단의 욕구에 맞게 차별화된 제품 및 마케팅을 제공하는 전략이다. 대량 마케팅 전략이 흔히 표적 마케팅 전략과 비교하여 열등한 전략으로 묘사되는 경우가 많지만, 대량 마케팅은 경영학 이론의 근간이 되는 대량 생산 및 ‘규모의 경제(economies of scale)’를 통한 비용 절감에 기초를 두고 있다. 이것으로 볼 때 기업은 시장별 고객만족도는 높아지지만, 비용이 증가

하는 표적 마케팅보다 투입할수록 비용 절감이 되는 대량 마케팅을 더 선호할 수 있다. 그러나 기업의 일방적이고 무차별적인 커뮤니케이션의 낭비적 요소를 배제하고 표적 고객 선별을 통하여 지속적인 마케팅 활동의 관계를 유지할 수 있다면, 표적 마케팅 활동의 효율성은 높아지고 매출도 증가하게 될 것이다.

2.2 고객 분류예측을 위한 기계학습 기법

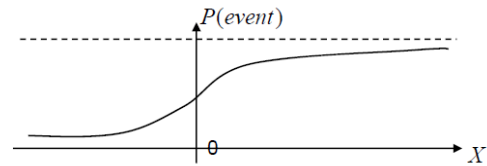
분류예측에는 다양한 종류의 기계학습 방법이 적용될 수 있다. 본 절에서는 고객 분류예측에 적용될 수 있는 다양한 기계학습 기법들에 대해 간략하게 살펴보고자 한다.

2.2.1 로지스틱 회귀분석

(Logistic Regression: LOGIT)

고객분류예측을 포함하여, 대부분의 분류문제는 0 또는 1의 종속변수값을 분류예측하는 이분류(binary classification) 문제이다. 그러나 이와 같이 이항 확률을 가진 종속변수를 통계적으로 설명하고자 할 때, 종속변수가 어떤 값이든 가질 수 있는 기존의 선형회귀분석은 바로 적용하기에 어려움이 있다(안현철, 2002).

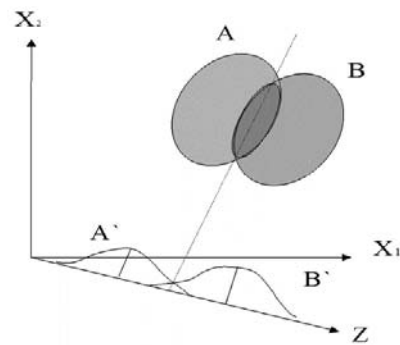
이러한 문제를 해결하기 위해 사용되는 기법이 로지스틱 회귀분석이다. 회귀분석의 결과를 로지스틱 함수를 통해 변환하게 되면, <그림 1>과 같이 정규분포의 누적확률분포와 유사하게 0과 1 사이의 값을 갖게 된다. 때문에, LOGIT은 앞서 설명한 이분법적인 분류 문제를 해결하는 모형으로 적합하다(Shmueli et al., 2016).



<그림 1> 로지스틱 회귀 함수

2.2.2 다중판별분석(Multiple Discriminant Analysis: MDA)

판별분석은 미리 정해진 그룹간의 차이를 잘 설명하여 줄 수 있는 변수들의 선형 결합을 찾아 새로운 항목을 이들 클래스 중의 하나로 분류한다. 판별분석은 로지스틱 회귀분석과 마찬가지로 이분법적인 종속변수의 집단 소속확률에 대한 추정치를 제공한다. 다음 <그림 2>는 판별분석의 기하학적 의미를 표현한 것이다 (Shmueli et al., 2016).



<그림 2> 판별분석의 기하학적 의미

2.2.3 의사결정나무(Decision Tree: DT)

의사결정나무는 의사결정 규칙을 나무 구조로 도표화하여 분류 예측을 수행하는 방법으로, 마디(node)라고 불리는 구성 요소들로 이루어져 있다. 의사결정 나무는 다른 방법에 비해 결

과가 명료하며 결과에 대한 해석이 쉽다는 장점을 가지고 있지만, 과적합화에 민감하다(안현철 2002; Shmueli, et al., 2016).

2.2.4 사례기반추론 (Case Based Reasoning: CBR)

사례기반추론은 과거 경험이나 사례에 기반하여 현재 주어진 문제에 대한 해결책을 제시하는 기법으로 가장 가까운 이웃 k개를 찾아 거리를 측정하는 k-NN(k-nearest neighbor)이 주로 사용된다(한현웅, 안현철, 2017). 장점은 모델이 단순하며 적은 데이터로도 결과제시가 가능하지만, 예측력이 상대적으로 낮다는 한계를 가진다(Jiawei et al., 2011).

2.2.5 인공신경망 (Artificial Neural Networks: ANN)

인공신경망은 생물학적인 뇌의 작동 원리를 그대로 모방한 방법으로, 입력값을 받는 노드들로 구성된 입력층과 앞 층으로부터 입력을 받는 노드들로 구성된 연이은 층들로 이루어진 회로망으로 설계된다. 각 층의 노드 출력은 다음 층 노드에 입력되며, 마지막 층은 출력층이라 불린다. 인공신경망은 계산된 출력값과 목표 출력값을 비교하고 오차를 최소화시킬 수 있도록 반복하여 조절한다. 장점은 복잡하고, 비선형적인 자료를 분석할 수 있으며, 강력한 성능을 보인다. 단점으로는 최종적으로 학습된 결과가 복잡하여 관계를 직관적으로 설명하지 못하는 점, 과적합화의 위험이 높은 점, 직관적으로 결정해야 할 설계요소가 많은 점 등을 지적할 수 있다(안현철, 2002; Jiawei et al., 2011).

2.2.6 SVM (Support Vector Machine)

SVM은 러시아의 통계학자 Vapnik가 개발한 기계학습 기법으로, 이분류에 특화된 기법이다. SVM에서는 먼저 훈련 데이터를 고차원 공간으로 사상하고, 늘어난 차원 수를 바탕으로 초평면(hyperplane)을 나누어주는 최적의 선을 찾음으로 초평면상의 두 데이터 사이에 항상 적절한 분할 경계를 찾는다. SVM의 장점은 정확성이 뛰어나며, 복잡한 데이터에도 적용할 수 있다는 점이다. 또한 구조적 위험을 최소화함으로써 과대적합문제를 벗어날 수 있다는 장점이 있다(안현철 등, 2005; 황운호, 강신재, 2007; Jiawei et al., 2011).

2.3 무료 샘플 판촉

판매촉진(sales promotion; 판촉)이란 기업이 구매를 유도하기 위해 고객에게 인센티브(예: 할인쿠폰, 무료샘플 배포, 사은품 증정, 세일)를 제공하는 것을 의미한다(박찬수, 2014). 판촉은 특정 제품 혹은 브랜드의 현저성(salience)을 증가시켜 대상 제품 혹은 브랜드의 구매를 유도하는 역할을 하므로 기업의 즉각적이고 가시적인 성과로 연결되기 쉽다. 오늘날 기업은 판촉 비용을 지속적으로 증가시키고 있다. 그 이유는 기술의 발전으로 방대한 고객 데이터 입수가 가능해졌고, 다양한 판촉 매체가 발달함으로써 기업이 효과적인 판촉 활동을 실행하는 것이 용이 해졌기 때문이다(최지은, 박종철, 2013).

판촉 활동 중 기업에서 가장 쉽게 활용할 수 있는 것이 무료 샘플 판촉이다. 무료 샘플은 지난 몇 세기 동안 마케팅 실무자들에 의해 제품과 서비스를 홍보하는 데 사용되어왔다. 소비자

는 무료 샘플을 통해 구매에 대한 부담없이 제품을 시도해볼 수 있게 되는데, 이 때의 경험이 추후 구매 결정을 내릴 때 중요한 기준이 된다. 기업에게 있어 무료 샘플은 대중적이지만 가장 유용한 판촉 도구 중 하나이다(Shultz et al, 1997; Heiman, et al., 2001). 무료 샘플을 통해 기업은 특별한 마케팅 전략을 세우지 않아도 다양한 고객에게 쉽게 접근하여 신제품이나 브랜드를 홍보할 수 있으며, 구매 시 경쟁 제품에서 자사 제품으로 전환할 수 있도록 환기시켜 시장 점유율 제고를 도모할 수 있다(Joanne and Warren, 2015).

판촉을 통하여 무료 샘플을 사용해 볼 기회를 가지게 된 소비자들은 무료 샘플 경험 이후 제품을 구매할 의향을 가지게 되거나 충성도가 높은 고객이 될 잠재성을 가지게 된다(Juliano and Michael, 2013). 또한 무료 샘플은 단기적으로 소비자의 구매 확률을 높이는 등 즉각적인 구매를 가져오며, 장기적으로 제품과 브랜드에 대한 충성심을 통하여 부가가치를 창출할 수 있다(Heiman et al, 2001; Villas-Boas, 2004; Mikaela and Maria, 2012).

그러나 잘못된 무료 샘플 판촉은 마케팅 비용을 증가시킬 뿐만 아니라 제품 비용 또한 증가시켜 투자 수익률(return on investment: ROI)을 낮춘다. 그 원인은 샘플 판촉에 (1)적합하지 않은 제품을 선택하였거나 (2)판촉 대상을 무분별하게 선정하거나 해당 제품 분야에 대한 지출이 적은 소비자, 즉 잘못된 표적을 선정했기 때문이다(Mikaela and Maria, 2012; Joanne and Warren, 2015).

따라서 무료 샘플 판촉의 효과적 활용을 위하여 두 가지가 필요하다. 첫째, 화장품처럼 샘플

사용량이 많은 제품을 샘플 판촉 제품으로 선정해야 한다. 무료 샘플 판촉의 목적은 제품의 현저성을 증가시켜 구매로 유도하는 것이기 때문에, 기본적으로 샘플 사용량이 많은 제품을 선택해야 소비자의 반응을 쉽게 유발할 수 있다. 대표적인 예가 화장품 분야이다. 김미경과 김주덕(2015)의 ‘성인 여성들의 화장품 구매행태에 관한 연구’에 따르면 성인 여성이 화장품을 구매할 때, 추천보다는 직접 경험을 선호하는 것으로 밝혀졌고, Amor and Guibert(2009)의 연구에서도 화장품 샘플을 받은 후 72%의 고객이 구매로 연결되었다. 이처럼, 화장품은 경험재적 성격을 띠기 때문에 샘플에 대한 사용량이 많고, 무료 샘플 판촉의 효과 또한 크다고 할 수 있다.

둘째, 샘플 판촉 이후 반응이 일어날 가능성이 높은 적절한 고객층을 목표로 해야 한다. 앞서 설명한 것과 같이 대량 마케팅이 여전히 효과적이었던 시기에는 무료 샘플 판촉 또한 그 대상이 시장 전체 혹은 세분 집단 전체였다. 그러나 소비자가 다양해지고 경쟁이 치열해진 표적 마케팅 시대에 샘플 판촉의 남발은 마케팅 비용을 증가시킬 위험이 있다. 때문에 많은 판촉 담당 실무자들이 무료 샘플 판촉의 대상이 선별적으로 가려지기를 원하고 있다. 특히 무료 샘플 판촉 결과의 측정이 어려울 뿐만 아니라, 자체로는 효과성 입증에 위한 명확한 근거가 제공되지 않기 때문에 이들은 무료 샘플 판촉의 결과에 대한 분석이 제시되길 원하고 있는 실정이다(Mikaela and Maria, 2012; Joanne and Warren, 2015; Kapil and Robert, 2004).

정보통신기술의 발전은 지난 30년간 마케팅을 혁신 시켜 왔으며, 무료 샘플 판촉도 인터넷

과 모바일과 같은 비전통적인 마케팅 채널로 확장되었다. 그러나 현재 선행 연구를 살펴보면 무료 샘플 판촉에 관한 연구는 주로 무료 샘플 판촉의 전통적인 형태와 효과에만 집중하고 있다. 온라인을 통한 무료 샘플 판촉에 초점을 맞춘 연구는 제한적이며, 판촉 효과를 분석하기 위하여 개발된 모델도 대단히 드물다(Joanne and Warren, 2015).

Ⅲ. 제안 모형

본 연구는 이상의 선행 연구를 참고하여 무료 샘플 판촉의 효과적 활용을 위한 기계학습 기반의 고객분류예측 모형을 제안하고자 한다.

Kapil and Robert(2004)에 따르면 무료 샘플 판촉에는 세 가지 잠재적 효과가 존재한다. 첫째는 가속 효과로, 무료 샘플을 경험한 신규 고객이 무료 샘플을 받지 못한 고객보다 브랜드를 빨리 구매할 수 있다는 것이다. 두 번째는 자기시장잠식(cannibalization) 효과로, 판촉을 통하여 무료 샘플을 받은 고객이 브랜드를 이미 구매한 고객일 경우, 샘플을 사용하면서 해당 고객의 재구매 횟수가 줄어들 수 있다는 것이다. 세 번째는 확장 효과로, 무료 샘플 판촉이 없었다면 브랜드 구매를 시도하지 않았을 구매자의 수를 늘리는 것이다.

많은 기업이 무료 샘플 판촉을 하는 이유는 가속 효과와 확장 효과를 통하여 매출 증대를 얻기 위함이다. 그러나 이론적 배경에서 밝힌 바와 같이 무료 샘플 판촉에 대한 결과를 분석하고 그것을 기업에 유의미하게 제시할 수 있는 모형은 그동안 거의 논의되지 못했다. 이에

본 연구에서는 판촉을 통하여 무료 샘플을 경험한 후, 확장 효과를 나타낼 고객을 선별해 낼 수 있는 기계학습 기반의 고객분류예측모형을 제안하고자 한다.

본 연구에서는 무료 샘플 판촉을 분석하기 위하여 전술한 것과 같이 제품 샘플의 사용량이 많은 화장품 업체 D사의 온라인 쇼핑몰 화장품 구매 데이터를 활용하였다. D사는 구매와 동시에 카테고리별 무료 샘플을 제시하고 고객에게 그 중 하나를 선택하게 하여 샘플 판촉을 진행하고 있다. 본 연구는 구매자들과 최대한 많은 접촉을 보인 샘플의 확장효과를 알아보기 위하여 D사의 판매량이 가장 많은 상품들로 구성된 패키지 형태의 특정 무료 샘플(이하 B 샘플)을 선정하였다. 이후 B 샘플에 영향을 받은 집단을 분류하고 그 집단을 이용하여 고객 예측모형을 구축하였다. 다음의 <그림 3>은 본 연구에서 제안하는 모형의 4단계 작동 과정을 시각화하여 제시하고 있다.



<그림 3> 제안 모형의 작동 과정

1단계. 데이터 수집 및 전처리 단계이다. D사의 2016년 12월부터 2017년 7월까지의 데이터를 본 연구의 목적에 부합하도록 분류하여

D-dataset을 구축한다. 그 후 분석이 가능하도록 전처리를 거치고 분석에 사용할 변수를 추출한다.

2단계. 분류예측모형을 구축한다. 본 연구에서는 D사의 B샘플을 분석 대상으로 선택하고 있는데, B샘플의 수량이 해당 샘플의 본 제품으로 구매가 이어진 고객, 즉 확장 효과를 나타낸 고객과 그렇지 않은 집단의 고객을 종속변수로 설정하여 고객 분류예측모형을 구축한다 (Kapil and Robert, 2004).

3단계. 이렇게 하여 데이터 전처리 및 모형 구축 작업이 끝나고 나면, 데이터를 학습용과 검증용으로 나누어 기계학습 기법에 적용한다. 다양한 기계학습 기법을 활용함으로써 여러 개의 모형을 구축하게 된다.

4단계. 기계 학습을 통하여 모형들이 구축되면, 예측정확도의 모형 간 결과 비교를 통하여 가장 우수한 모형을 선정하는 작업이 이루어진다. 이후, 추가적으로 통계검정을 수행하여 선정 모형의 성능 차이가 통계적으로 유의한 수준인지 확인한다. 마지막으로 선정된 최적 모형을 표적 마케팅 전략에 이용하는 방안을 모색한다.

을 위한 설정을 설명하고, 4.2 절에서는 정제된 데이터를 이용한 변수 추출과 실험을 설명한다. 4.3 절에서는 실험 결과를 제시하며, 마지막 4.4 절에서는 결과를 활용한 표적 마케팅 전략 모색 방안을 제시한다.

4.1 실험 데이터

실험 데이터로는 화장품 업체 D사의 2016년 12월부터 2017년 7월까지 8개월 동안, 총 18,609명 회원의 온라인 쇼핑몰 구매 데이터를 활용하였다. 이 중, 확장 효과를 보인 고객 집단을 분류하기 위하여 ACE(Acceleration Cannibalization Expansion; 가속/ 시장잠식/ 확장) 모델을 참고하였다. ACE 모델은 앞서 설명한 무료 샘플 판촉의 세 가지 효과의 상대적 크기에 따라 판촉의 결과가 긍정적이거나 부정적일 수 있음을 시사하며 다음과 같은 세 가지 고객 세분 집단이 있다고 가정한다. (1)이미 샘플에 해당하는 브랜드를 시도한 집단, (2)무료 샘플 없이 브랜드를 시험 구매한 집단, (3)무료 샘플이 없었다면 브랜드를 구매하지 않을 집단 (Kapil and Robert, 2004).

이러한 ACE 모델을 참고하여 <그림 4>와 같이 전체 18,609명의 회원을 분류하였다.

IV. 실증 분석

본 장은 제안 모형의 적용 가능성을 확인하기 위해 수행된 실증 분석에 대하여 설명한다. 먼저 4.1 절에서는 실험 데이터의 구성과 실험

P1 : 9,281명	P2 : 3,503명
	P3 : 5,196명
	P4 : 629명

<그림 4> 데이터 상 고객 분류

P1: B샘플을 경험해 보지 않은 고객 (9,281명)

P2: B샘플을 경험하였으나, 그 이후 구매로 연결되지는 않은 고객 (3,503명)

P3: B샘플을 경험하고 나서, 본 제품을 구매하였으나, B샘플을 경험해 보기 이전에 이미 본 제품을 구매했던 경험이 있는 고객 (5,196명)

P4: B샘플을 경험한 이후, 처음으로 해당 샘플의 본 제품을 온라인으로 구매한 고객 (629명)

P1 집단은 B샘플에 대한 관심이 없거나 아직 경험하지 못한 집단으로 추후 확장효과를 보일 가능성이 있는 집단이다.

P2 집단은 B샘플을 경험하였지만 구매로 연결되지 않은 고객이며, P3 집단은 B샘플을 받기 이전 시점부터 본 제품을 구매한 경험이 있는 고객으로 두 집단은 ACE 모델 상 (1) 이미 샘플에 해당하는 브랜드를 시도한 집단이며, 그 중에는 (2) 무료 샘플 없이 브랜드를 시험 구매한 집단(P3)도 포함된다.

P4 집단 고객 629명은 B샘플을 경험한 이후 샘플의 본 제품 구매로 연결된 고객으로서, 본 연구에서 초점을 맞추고자 하는 확장 효과를 보인 집단이다. 본 연구 상에서 확장효과는 온라인 쇼핑몰에서 샘플의 본 제품 주문내역이 없었으나, B샘플을 선 수령 후 사용한 뒤 최초 구매가 일어난 구매 건으로 정하였다.

전술했듯이, 무료 샘플 판촉의 효과적 활용을 위해서는 (1)샘플의 사용량이 많은 제품을 선정하고, (2) 잠재적 구매 가능성이 높은 고객층을 표적화하여야 한다. 본 실증분석에서는 화장품 사례를 사용하고 있기 때문에, 첫 번째 조

건(샘플의 사용량이 많은 제품 선정)은 명백하게 충족된다. 따라서, 잠재적 구매 가능성이 높은 고객층의 표적화만 하면 되는데, 현재 확보된 데이터셋에서는 P4가 바로 이 ‘잠재적 구매 가능성이 높은 고객층’이라 할 수 있다. 이들은 기존에는 브랜드 사용 경험이 없다가 샘플 경험 이후, 비로소 본 제품 구매로 이어진 고객들이기 때문이다.

이러한 맥락에서 본 연구에서는 ‘P4 집단에 속하였는가’의 여부를 종속변수로 설정하였다. 또한 학습의 왜곡을 막기 위해 집단 별 빈도 비율을 1:1(629:629)로 맞추기 위하여 B샘플을 경험하지 않은 P1 집단에서 629건의 표본을 무작위 추출하였다. 그리하여 총 1,258건의 데이터로 고객분류 예측모형을 구축하였다.

4.2 실험 설계

본 실험에서 종속변수는 ‘B샘플을 사용한 후 해당 본 제품을 구매한 고객 집단(P4)에 포함되는가’의 여부로 정하였다. 과적합화 여부를 확인하기 위하여 비율을 8:2로 적용하여 80%의 해당하는 데이터(총 1,008건)를 학습용 데이터 그리고 나머지 20%(총 250건)를 검증용 데이터로 사용하였다.

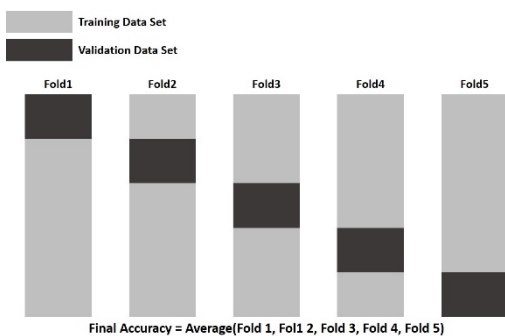
입력 변수의 후보군은 크게 인구통계학적 변수, 구매패턴 및 웹 로그 데이터 변수 그리고 샘플 주문 이전 구매 패턴의 변수 등 크게 세 가지 형태로 이루어져 있다. 이 중 인구통계학적 변수는 성별, 나이, 거주지 등이며, 파생변수를 포함하여 총 8개로 구성되어 있다. 구매패턴 및 웹 로그 데이터 변수는 회원가입일, 주문 횟수, 상품군 구매 여부, 접근 경로와 지불수단 등

<표 1> 전체 입력변수 후보군

변수	유형	변수명	설명
독립변수	인구통계학적	gender	성별
		Age	연령
		Target_CS_1	주 고객 여부(35세~39세 여성)
		Target_CS_2	주 고객 여부(40세~44세 여성)
		ADDR1	거주지 서울 여부
		ADDR2	거주지 경기 여부
		ADDR3	거주지 광역시 여부
	ADDR4	거주지 지방 여부	
	구매 패턴 및 웹 로그 데이터	B	BEST 상품군 구매 여부
		S	SPECIAL 상품군 구매 여부
		ETC	기타 상품군 구매 여부
		Hf	모공 상품군 구매 여부
		Ela	탄력 상품군 구매 여부
		White	화이트닝 상품군 구매 여부
		Ac	AC케어 상품군 구매 여부
		Bs	BEST 샘플 선택 여부
		Ss	SPECIAL 샘플 선택 여부
		ETCs	기타 샘플 선택 여부
		Hfs	모공 샘플 선택 여부
		Elas	탄력 샘플 선택 여부
		Whites	화이트닝 샘플 선택 여부
		Acs	AC케어 샘플 선택 여부
		Now_date_reg	회원가입일로부터 현재일까지의 계산일수
		Order_cnt	주문 횟수
		vst_cnt	온라인 쇼핑몰 방문 횟수
		ISMAILING	메일 수신 여부
		ISSMS	SMS 수신 여부
		Method1	지불수단 카드 여부
		Method2	지불수단 가상계좌 여부
		Method3	지불수단 카카오페이 여부
		Method4	지불수단 페이나우 여부
		Method5	지불수단 포인트결제 여부
		Method6	지불수단 휴대폰 결제 여부
		Access1	접근경로 PC 여부
		Access2	접근경로 휴대폰 여부
		Access3	접근경로 전화 여부
		M1_A1	접근경로가 PC이면서 지불수단이 카드인 고객 여부
		M2_A1	접근경로가 PC이면서 지불수단이 가상계좌인 고객 여부
		M6_A1	접근경로가 PC이면서 지불수단이 휴대폰 결제인 고객 여부
		M1_A2	접근경로가 휴대폰이면서 지불수단이 카드인 고객 여부
		M2_A2	접근경로가 휴대폰이면서 지불수단이 가상계좌인 고객 여부
	M6_A2	접근경로가 휴대폰이면서 지불수단이 휴대폰 결제인 고객 여부	
	MS_A1	접근경로가 PC이면서 문자수신에 동의한 고객 여부	
MA_A1	접근경로가 PC이면서 이메일 수신에 동의한 고객 여부		
MA_A2	접근경로가 모바일이면서 메일 수신에 동의한 고객 여부		
샘플 주문 이전 구매 패턴	Bfsp_odcnt	B샘플 선택 주문 이전 주문횟수	
	Bfsp_sim	B샘플 선택 주문 이전 샘플 상품군과 유사한 품목 구매 여부	
	Bfsp_date	B 샘플 선택 주문 이전 거래일로부터 직전 주문까지의 일 수	
	Bfsp_top50	B샘플 선택 주문 이전 쇼핑몰 판매 상위 50개 품목 구매 여부	
중속변수		Y	B 샘플을 사용한 후 해당 제품을 구매한 고객의 여부

이며 파생변수를 포함하여 총 37개로 구성되어 있다. 샘플 주문 이전 구매 패턴 변수는 B샘플이 포함된 주문 이전의 구매 패턴을 파생하여 만들어낸 4개의 변수로 구성되어 있다. 총 49개의 변수 중 연속형 변수를 제외한 나머지를 모두 더미(dummy) 변수로 변환하였다. 이후 독립표본 T 검정과 카이 제곱(χ^2) 검정을 이용하여 신뢰수준 95% 하에서 유의한 변수를 추출하였다. 그 다음 각 분석기법에 적용할 최종 독립변수를 선정하기 위하여 로지스틱 회귀분석을 적용하였다.

본 연구의 실험 데이터에 가장 잘 맞는 기계학습 기법을 선정하기 위해 로지스틱 회귀분석과 다중 판별 분석, 의사결정나무, 사례기반추론, SVM, 인공신경망을 활용하여 분류 예측 모델을 학습하고, 이를 검증용 데이터셋에 적용하여 각 기법 간 예측 정확도를 비교하였다. 또한 실증분석 결과의 신뢰도 제고를 위하여 k-겹 교차검증(cross validation)도 적용하였다.



<그림 5> 5-겹 교차검증

본 연구에서는 <그림 5>와 같은 방법으로 5겹으로 나누어 교차검증을 진행하였는데, 이 때 학습용 데이터로 약 80%에 해당하는 데이터

(총 1,008건)를 그리고 검증용 데이터로 약 20%에 해당하는 데이터(총 250건)를 사용하였다.

본 연구의 분석에 사용한 LOGIT, MDA, DT, CBR은 IBM SPSS Statistics 20을, ANN은 NeuroShell2 4.2를, SVM은 LibSVM v3.0을 활용하여 분석하였다.

4.3 실험 결과

<표 1>은 분석에 사용한 전체 입력변수 후보군을 나타내고 있다. 이러한 전체 49개의 입력변수 후보군에서 독립표본 T 검정과 카이 제곱(χ^2) 검정을 이용하여 신뢰수준 95% 하에서 유의한 변수를 추출하였다. 그 결과 <표 2>에 제시된 총 17개의 변수가 먼저 추출되었다.

이후 분석 기법에 적용할 최종 독립변수를 선정하기 위하여 추출된 17개의 변수를 LOGIT에 적용하였다. 전방 탐색(forward selection) 방법을 LOGIT에 적용한 결과, 17개 변수 중 총 6개의 변수가 최종 독립변수로 선정되었다. 최종 선정된 6개 독립변수들은 다음의 <표 3>에 제시되어 있다.

최종 선정된 독립변수들을 제시된 기계학습 기법에 적용하여 실험하였다. 전술한 대로 5-겹 교차검증을 적용하였기 때문에, 데이터의 80%(1,008건)를 학습용, 20%(250건)를 검증용으로 사용하였다. <표 4>에는 기계학습별 5-겹 실험 결과가 제시되어 있다. <표 4>의 결과를 살펴보면, 검증용 데이터셋에 대하여 SVM을 적용한 경우의 평균 예측 정확도가 75.36%로 가장 우수한 예측 정확도를 산출하고 있음을 알 수 있다. 평균 예측 정확도가 높은 순서는

<표 2> 유의한 독립 변수

변수	유형	변수명	설명
독립변수	인구통계학적	gender	성별
		Age	연령
		Target_CS_1	주 고객 여부(35세~39세 여성)
	구매 패턴 및 웹 로그 데이터	S	SPECIAL 상품군 구매 여부
		ETC	기타 상품 구매여부
		Now_date_reg	회원가입일로부터 현재일까지의 계산일수
		Order_cnt	주문 횟수
		vst_cnt	온라인 쇼핑몰 방문 횟수
		ISMAILING	메일 수신 여부
		Access1	접근경로 PC 여부
		Access2	접근경로 Mobile 여부
		M2_A1	접근경로가 PC이면서 지불 수단이 가상계좌인 고객 여부
		M2_A2	접근경로가 모바일이면서 지불수단이 가상계좌인 고객 여부
		MA_A2	접근경로가 모바일이면서 메일 수신에 동의한 고객여부
	샘플 주문 이전 구매 패턴	Bfsp_odcnt	B샘플 선택 주문 이전 주문횟수
		Bfsp_sim	B샘플 선택 주문 이전 샘플 상품군과 유사한 품목 구매 여부
		Bfsp_top50	B샘플 선택 주문 이전 쇼핑몰 판매 상위 50개 품목 구매 여부
종속변수		Y	B 샘플을 사용한 후 해당 제품을 구매한 고객의 여부

<표 3> 최종 선정된 독립 변수

변수	변수명	설명
독립변수	gender	성별
	S	SPECIAL 상품군 구매 여부
	Now_date_reg	회원가입일로부터 현재일까지의 계산일수
	Order_cnt	주문 횟수
	MA_A2	접근경로가 모바일이면서 메일 수신에 동의한 고객여부
종속변수	Bfsp_top50	B샘플 선택 주문 이전 쇼핑몰 판매 상위 50개 품목 구매 여부
	Y	B 샘플을 사용한 후 해당 본 제품을 구매한 고객의 여부

<표 4> 실험결과 종합

Dataset		LOGIT	MDA	DT	CBR	ANN	SVM
Fold 1	Train	71.53%	69.84%	75.20%	70.44%	74.21%	75.49%
	Valid	70.80%	70.40%	73.20%	74.40%	74.80%	75.20%
Fold 2	Train	70.34%	68.35%	75.00%	73.21%	75.53%	69.54%
	Valid	76.40%	76.00%	74.80%	72.00%	75.60%	77.60%
Fold 3	Train	71.63%	70.14%	74.40%	71.73%	75.13%	75.60%
	Valid	70.00%	66.80%	67.20%	72.80%	72.80%	76.00%
Fold 4	Train	71.53%	69.54%	74.60%	72.82%	74.21%	76.19%
	Valid	70.00%	70.00%	74.40%	74.00%	69.20%	74.00%
Fold 5	Train	71.92%	70.24%	74.31%	71.43%	76.72%	76.49%
	Valid	68.00%	66.00%	68.80%	67.20%	72.80%	74.00%
Average	Train	71.39%	69.62%	74.70%	71.93%	75.16%	74.66%
	Valid	71.04%	69.84%	71.68%	72.68%	73.04%	75.36%

SVM > ANN > CBR > DT > LOGIT > MDA 순으로 나타났다.

마지막으로 모형 간 예측 정확도의 차이가

통계적으로 유의한 지를 검증하기 위하여, 이표본 비울검정을 수행하였다. 본 연구에서 적용된

이표본 비울검정의 귀무가설 H_0 는 $P_A = P_B$, 대

<표 5> 이표본 비울검정 결과(Z-Values)

	MDA	DT	CBR	ANN	SVM
LOGIT	0.6574	-0.3539	-0.5763	-1.1141*	-2.4384***
MDA		-1.0113*	-1.2336*	-1.7711**	-3.0941***
DT			-0.2224	-0.7603	-2.0851***
CBR				-0.5379	-1.8630**
ANN					-1.3256**

* Statistical Significant at 10%, ** Statistical Significant at 5%, *** Statistical Significant at 1%

립가설 H_a 는 $P_A > P_B$ (P_X : 모형 X의 검증용 데이터셋에 대한 평균 예측 정확도 비율)이다.

<표 5>는 이러한 이표본 비울검정의 결과를 나타내고 있다. 이 표에 제시된 바와 같이 SVM은 95% 신뢰수준 하에서 CBR, ANN과 통계적으로 유의한 성과 차이를 보이는 것으로 나타났으며, LOGIT, MDA, DT와는 99% 신뢰수준 하에서 통계적으로 유의한 성과 차이를 보이는 것으로 나타났다.

4.4 표적 마케팅 전략 모색

지금까지 제안 모형을 분석하여 최적 모형을 선정하고 결과를 검증하였다. 본 연구에서 온라인 쇼핑물 거래 데이터를 통하여 고객분류예측 모형을 만들고 분석한 모든 과정은 ‘고객관계 관리를 위하여 어떻게 효과적인 무료 샘플 관측을 전개해 나갈 것인가’에 대한 질문으로 귀결된다. 따라서 모형을 온라인 쇼핑물에 적용하여 본다면, 신규 고객이나 특정 상품군의 구매 경험이 없는 기존 고객을 고객분류예측모형을 통하여 예측하고 예측된 표적 고객에게 적절한 표적 마케팅 전략을 펼칠 수 있다.

예를 들어, 각 샘플 카테고리별로 고객 분류 예측모형을 만들고 신규 고객 가입 시 가장 유사하게 예측된 샘플 카테고리를 고객에게 먼저 제시하여 추천하거나, 샘플을 사용하도록 유도

할 수 있을 것이다. 또한 각 모형으로 예측된 표적 집단의 특성을 분석하여 그것을 마케팅 전략의 세분 집단으로 활용 할 수 있으며, 모형을 무료 샘플 관측 이후 재분석함으로써 관측의 결과 확인과 효과성을 입증하는 과정에도 활용될 수 있다.

결과적으로 관측 담당 실무자가 필요로 하는 무료 샘플 관측 대상의 기준점과 관측 결과 분석의 두 가지 요구사항을 이 모형을 통하여 제시할 수 있다고 할 수 있다(Mikaela and Maria, 2012; Joanne and Warren, 2015; Kapil and Robert, 2004).

V. 결론 및 시사점

본 연구는 앞서 밝힌 바와 같이 온라인쇼핑물 무료 샘플 관측에 있어, 특정 샘플을 선택하여 사용한 후 추후 구매 시 해당 본 제품을 구매한 집단, 즉 무료 샘플 관측의 목적 중 하나인 확장 효과를 보인 집단을 분류하고, 그 집단을 바탕으로 예측 모형을 제안하였다. 또한 제안된 모형을 구현하는데 있어 로지스틱 회귀분석, 다중 판별분석, 의사결정나무, 사례기반추론, SVM, 인공지능망 등의 다양한 기계학습 기법의 활용을 제안하였으며, 이렇게 구축된 고객분

류예측 모형이 최종적으로는 표적 마케팅에 어떻게 활용될 수 있는지 제시하였다. 구체적으로 본 연구가 갖는 의미를 살펴보면 다음과 같다.

첫째, 본 연구는 기계학습 방법론에 기반한 분류모형을 화장품 기업의 온라인 판촉에 적용하였다는 점에서 의의가 있다. 그간 화장품의 판촉과 관련해서는 구매 행동, 선호도와 같은 행동분석연구와 거래 데이터를 활용한 수요 예측이나 추천시스템과 같은 연구들이 주로 발표되어 왔으며, 무료 샘플을 대상으로 수행된 연구들은 전통적 판촉에서의 효과성 등을 주로 다루어 왔다. 반면 국내는 물론 해외에서도 무료 샘플을 중심으로 하는 온라인 판촉활동에 대한 연구는 그 사례를 거의 찾아볼 수 없는 실정이다. 때문에, 화장품 기업의 온라인 판촉을 대상으로 하고 있는 본 연구는 학술적으로 그리고 실무적으로 상당히 의의가 있다고 할 수 있다.

둘째, 화장품 업체들이 관례적으로 시행해 온 무료 샘플 판촉의 효과를 측정하고, 그 분석 결과를 제시하였다는 점에서 의의가 있다. 선행 연구에서 무료 샘플 판촉이 어떤 유형의 효과를 가져올 수 있는지는 이론적으로 제시된 바 있지만, 그것이 실제적으로 어느 정도 효과를 가져오는지 실제 데이터에 기반해 확인한 연구는 찾아보기 어렵다. 이러한 상황에서 비록 확장 효과에 국한되기는 하였지만, 무료 샘플 판촉이 실제 기업에서 어느 정도 효과를 보이는지 수치로 제시하고 있는 본 연구는 현업에서 관련 업무를 수행하고 있는 실무자들에게 유용한 지침이 될 수 있을 것으로 예상된다.

셋째, 대량 마케팅의 개념 하에서 시장 혹은 세분 집단 전체를 대상으로 진행하던 무료 샘

플 판촉을 표적 마케팅으로 적용할 수 있는 방안을 제시하였다는 점에서 의의가 있다. 고객관계관리에서 고객관계의 획득보다는 유지 및 확장을 더 중요하게 강조하고 있음을 감안할 때, 무료 샘플 판촉을 고객관계의 확장 수단으로 활용할 수 있는 1:1 표적 마케팅을 위한 고객분류 모형을 제안하고 있는 본 연구는 실무적 활용가치가 상당히 높은 연구라 사료된다.

반면, 본 연구는 다음과 같은 한계점들을 갖고 있다. 첫째, 데이터의 양적 측면에서 확장 효과를 보인 데이터 양이 충분하지 않았으며, 질적 측면에서 고객에 대한 개인 고유 데이터가 충분치 않았다. 때문에 향후 데이터의 기간 범위를 확장하고, 고객에 대한 추가적인 데이터를 확보하여 실험할 필요가 있다. 또한, 모형에 이용된 기계학습 기법 외에 새로운 기법을 적용하는 연구도 필요할 것이다.

둘째, 본 연구는 무료 샘플 판촉의 효과 중 하나인 가속효과가 고려하지 못하였다. 확장효과와 가속효과를 명확히 구분해내는 것은 어렵지만, 가속효과 역시 무료 샘플 판촉의 긍정적인 효과 중 하나로서 향후 연구에서 보다 심도 있게 논의될 필요가 있다.

셋째, 본 연구는 온라인 무료 샘플 판촉에 한정하여 다루고 있어, 오프라인에서의 샘플 및 제품 경험의 가능성을 통제 하지 못하고 있는 한계가 있다. 또한 2016년 12월부터 2017년 7월까지의 데이터를 사용하는데 있어 확장 효과의 기간을 사전에 설정하지 못하였다. 후속 연구에서는 확장효과에 대한 엄격하고 명확한 설정을 통해 앞서 언급한 외생변수들을 보다 체계적으로 통제, 관리할 필요가 있다.

넷째, 본 연구에는 다수의 샘플 중에서 하나

의 샘플을 선택한 고객 데이터를 바탕으로 모형을 구축하고 있는데, 표적 마케팅은 엄선된 하나의 샘플을 일방적으로 제안하는 형태로 구현된다. 따라서, 본 연구의 제안 모형이 표적 마케팅에서 효과를 보이기 위해서는 샘플을 선택할 수 있는 상황에서 보여지는 소비자의 반응과 샘플이 사전에 정해진 상태에서 제안되었을 때 보여지는 소비자의 반응이 동일 또는 유사할 것이라는 전제가 있어야만 한다. 하지만 실제로 이런 전제가 유효한 지에 대해서는 본 연구에서 확인하지 못하였다. 소비자 행동 이론의 관점에서 본 연구의 전제가 유효할 지에 대한 검증과 확인이 후속 연구에서는 반드시 이루어져야 할 것으로 사료된다.

참고문헌

- 김미경, 김주덕, “성인 여성들의 화장품 구매행태에 관한 연구,” 한국화장품미용학회지, 제5권, 제2호, 2015, pp. 125-135.
- 김상수, 문준연, “마케팅과 정보기술의 통합적 활용 효과에 관한 실증연구,” 정보시스템연구, 제7권, 제1호, 1998, pp. 99-128.
- 김신곤, “데이터 마이닝 기법(CHAD)을 이용한 효과적인 데이터베이스 마케팅에 관한 연구,” *Journal of Information Technology Application & Management*, 제6권, 제1호, 1999, pp. 89-101.
- 김종완, 오기욱, “E-CRM을 위한 개인화 마케팅 기법에 관한 연구,” 한국컴퓨터정보학회 논문지, 제7권, 제2호, 2002, pp. 179-186.
- 김주호, “소비자 반응과 가치를 중시한 개별 마케팅(Individual Marketing)전략,” 마케팅, 1998, pp. 56-61.
- 박주석, “정보기술과 마케팅의 변화,” IE매거진, 제7권, 제2호, 2000, pp. 28-33.
- 박찬수, 마케팅 원리, 제5판, 서울: 법문사, 2014.
- 서현석, 주형준, “e-CRM 구성요인이 관계품질, 지각된 전환비용 및 지속된 이용의도에 미치는 영향,” 인터넷전자상거래연구, 제11권, 제3호, 2011, pp. 171-193.
- 안현철, “데이터 마이닝을 활용한 인터넷 쇼핑물의 상품 추천시스템 개발,” 한국과학기술원, 석사학위논문, 2002.
- 안현철, “사례기반추론의 유사 임계치 및 커버리지 최적화,” 정보처리학회논문지: 소프트웨어 및 데이터 공학, 제2권, 제8호, 2013, pp. 535-542.
- 안현철, 김경제, 한인구, “Support Vector Machine을 이용한 고객구매예측모형,” 지능정보연구, 제11권, 제3호, 2005, pp. 69-81.
- 원지성, “가격탄력성과 생산비용에 기초한 대량 마케팅과 표적시장 마케팅의 비교 분석,” 유통과학연구, 제11권, 제4호, 2013, pp. 61-72.
- 정현주, 고준, 김영걸, “고객관계관리(CRM)에서 고객정보/고객지식 품질에 영향을 미치는 요인: 서비스 산업을 중심으로,” 경영과학, 제19권, 제2호, 2002, pp. 1-24.
- 조대현, 김병수, 석경하, 이종언, 김종성, 김선

- 화, “화장품구매 자료를 통한 고객 구매 행태 분석,” 한국데이터정보과학회지, 제 20권, 제4호, 2009, pp. 615-627.
- 최지은, 박종철, “관측의 종류와 제시형태가 소비자 반응에 미치는 영향-소비자의 인지속구와 설득지식의 조절역할을 중심으로,” 한국심리학회지: 소비자. 광고, 제14권, 제3호, 2013, pp. 385-408.
- 한현웅, 안현철, “절대 유사 임계값 기반 사례기반추론과 유전자 알고리즘을 활용한 시스템 트레이딩,” 정보시스템연구, 제26권, 제3호, 2017, pp. 63-90.
- 홍태호, 박지영, “사례기반추론을 이용한 다이렉트 마케팅의 고객반응예측모형의 통합,” 정보시스템연구, 제18권, 제3호, 2009, pp. 375-399.
- 황운호, 강신재, “SVM 기계학습을 이용한 웹문서의 자동 의미 태깅,” 한국산업정보학회논문지, 제12권, 제2호, 2007, pp. 49-59.
- Amor, B. I. and Guilbert, F., “Influences on Free Samples Usage within the Luxury Cosmetic Market,” *Direct Marketing: An International Journal*, Vol. 3, No. 1, 2009, pp. 67-82.
- Heiman, A., McWilliams, B., Shen, Z., and Zilberman, D., “Learning and Forgetting: Modeling Optimal Product Sampling Over Time,” *Management Science*, Vol. 47, No. 4, 2001, pp. 532-546.
- Jiawei, H., Micheline, K., and Jian, P., *Data mining: Concepts and Techniques*, 3rd Edition, Elsevier Inc., 2011.
- Joanne, K., and Warren, W., “Something for Nothing? An Investigation into Online Product Sampling,” *International Journal of Information, Business and Management*, Vol. 7, No. 2, 2015, pp. 97-106.
- Juliano, L., and Michael, T., “An Investigation of the Effectiveness of Uncertainty in Marketing Promotions Involving Free Gifts,” *Journal of Marketing*, Vol. 77, No. 2, 2013, pp. 112-123.
- Kapil, B., and Robert, S., “The Effects of Free Sample Promotions on Incremental Brand Sales,” *Marketing Science*, Vol. 23, No. 3, 2004, pp. 345-363.
- Mikaela, E., and Maria, J., “How to Create Value through Strategic Product Sample Promotions - a Case Study of L’Oreal Sweden, Uppsala University, Sweden, Master thesis, 2012.
- Kotler, P., *Marketing Management*, 11th edition, Englewood Cliffs, NJ: Prentice-Hall, Inc., 2002.
- Rajiv, K., Frank, P., Tim, E., Tom, V., Marylou, S., and Gary, B., “Managing Customer Relationships through E-business Decision Support Application: A Case of Hospital Physician Collaboration,” *Decision Support System*, Vol. 32, 2001, pp. 171-197.
- Schultz, D. E., William, A. R., and Lisa, A. P., *Sales Promotion Essentials*, NTC

Business Books, Lincolnwood, IL., 1997.

Shmueli, G., Peter C. B., Nitin R. P., "Data Mining for Business Analytics," 3rd Edition, WILEY Inc., 2016.

Villas-Boas, J. M., "Consumer Learning, Brand Loyalty, and Competition," *Marketing Science*, Vol. 23, No. 1, 2004, pp. 134-145.

안 현 철 (Ahn, Hyunchul)



현재 국민대학교 경영대학 경영정보학부 부교수로 재직 중이다. KAIST에서 산업경영 학사를 취득하고, KAIST 테크노경영대학원에서 경영정보시스템을 전공하여 공학석사와 박사를 취득하였다. 주요 관심분야는 금융 및 고객관계 관리 분야의 인공지능 응용, 정보시스템 수용과 관련한 행동모형 등이다. *Annals of OR*, *Applied Intelligence*, *Applied Soft Computing*, *Expert Systems with Applications*, *Information and Management*, *International Journal of Electronic Commerce* 등의 국제학술지와 *Asia Pacific Journal of Information Systems*, *Journal of Information Technology Applications and Management*, *지능정보연구* 등의 국내학술지에 연구논문을 발표한 바 있다.

원 하 람 (Won, Ha-Ram)



한라대학교에서 경영학사 학위를 취득하였으며, 현재 국민대학교 비즈니스 IT 전문대학원에서 Business Analytics 트랙으로 석사과정에 재학 중이다. 주요 관심 분야는 Business Analytics과 CRM과 데이터 기반 마케팅 분석 등이다.

김 무 전 (Kim, Moo-Jeon)



부산대학교 경영학사와 연세대학교 경영학 석사학위를 취득하였다. 현재 국민대학교 비즈니스 IT전문대학원에서 Business Analytics 트랙으로 박사과정 재학 중이며, DHC 코리아 대표이사로 재직 중이다. 주요 관심분야는 Business Analytics, Data-driven Marketing, IoT Beauty Care 등이다.

<Abstract>

A Machine Learning-based Customer Classification Model for Effective Online Free Sample Promotions

Won, Ha-Ram · Kim, Moo-Jeon · Ahn, Hyunchul

Purpose

The purpose of this study is to build a machine learning-based customer classification model to promote customer expansion effect of the free sample promotion. Specifically, the proposed model classifies potential target customers who are expected to purchase the products included in the free sample promotion after receiving the free samples.

Design/methodology/approach

This study proposes to build a customer classification model for determining customers suitable for providing free samples by using various machine learning techniques such as logistic regression, multiple discriminant analysis, case-based reasoning, decision tree, artificial neural network, and support vector machine. To validate the usefulness of the proposed model, we apply it to a real-world free sample-based target marketing case of a Korean major cosmetic retail company.

Findings

Experimental results show that a machine learning-based customer classification model presents satisfactory accuracy ranging from 70% to 75%. In particular, support vector machine is found to be the most effective machine learning technique for free sample-based target marketing model. Our study sheds a light on customer relationship management strategies using free sample promotions.

Keyword: Online Free Sample Promotion, Machine Learning, Customer Relationship Management, Customer Expansion

* 이 논문은 2018년 7월 31일 접수, 2018년 9월 11일 1차 심사, 2018년 9월 25일 게재 확정되었습니다.