

카드산업에서 휴면 고객 예측

이동규* · 신민수**

목 차

요약	3.1 고객 휴면 예측 및 모형 평가
1. 서론	3.2 인구통계학적 정보의 조절효과
1.1 고객 이탈	4. 결론
1.2 카드산업에서 고객 휴면의 정의	References
2. 선행연구 및 이론적 배경	Abstract
2.1 고객 이탈과 휴면	
2.2 통합기술수용이론(UTAUT)	
3. LSTM을 활용한 고객 휴면 예측	

요약

고객 기반의 산업에서 고객 Retention은 기업의 경쟁력이라 할 수 있으며, 고객 Retention을 높이는 것은 기업의 경쟁력을 높이는 것이라 할 수 있다. 따라서, 미래 휴면 고객을 잘 예측하여 관리하는 것은 기업의 경쟁력을 높이는데 무엇보다 중요하다. 왜냐하면, 신규 고객을 유치하는데 필요한 비용이 기존 고객을 Lock-in 시키는데 드는 비용 보다 많은 것으로 알려져 있기 때문이다. 특히, 수 많은 카드사가 존재하는 국내 카드 산업의 휴면 카드를 관리하고자 정부에서 휴면 카드 자동 해지 제도를 도입하고 있으며, 카드 산업에서 휴면 고객을 관리하는 것이 무엇보다 중요한 과제로 떠오르고 있다.

본 연구에서는 카드 산업에서 휴면 고객을 예측하기 위해 Recurrent Neural Network (RNN)방법론을 사용하였으며, RNN방법론 중에서 긴 시간을 효율적으로 학습할 수 있는 Long-Short Term Memory (LSTM)을 활용하였다. 또한, 통합기술수용이론 (UTAUT)을 입각하여 카드 산업에서 휴면 고객을 예측하는데 필요한 변수를 재정의하였다. 그 결과 안정된 모형의 정확도와 F-1 score를 얻을 수 있었으며, Hit-Ratio를 통하여 모형의 안정된 결과를 입증하였다. 기존 연구에서 지적된 통합기술수용이론 (UTAUT)에서 발생 될 수 있는 인구통계학적 정보의 조절 효과도 발생 되지 않은 것을 보였으며, 이로 인해 통합기술수용이론(UTAUT)를 이용한 변수 선정 모형에서 LSTM을 이용한 휴면 고객 예측 모형은 편향되지 않고 안정된 결과를 가져다 줄 수 있다는 것을 입증하였다.

표제어: 고객 이탈, 고객 휴면, 통합기술수용이론(UTAUT), 머신러닝, 딥러닝, RNN, LSTM

접수일(2023년 02월 09일), 수정일(2023년 3월 14일), 게재확정일(2023년 04월 20일)

* 제1저자, 한양대학교 비즈니스인포매틱스학과 박사과정, harry8256@naver.com

** 교신저자, 한양대학교 경영학과 교수, minsooshin@hanyang.ac.kr

1. 서론

1.1 고객 이탈

산업이 급격하게 변화하고 발전되어옴에 따라 산업이 다양해지고 경쟁이 치열해짐에 따라 유효 고객을 관리하는 것이 중요해지고 있다. 특히 기업들은 고객을 기업의 가장 중요한 자산으로 인식하고 있으며 고객 보유율을 높이기 위해 많은 노력과 비용을 지출한다. (Reichheld and Sasser, 1990)가 다양한 업종을 대상으로 조사한 결과 고객의 이탈을 5% 줄이면 동일 업종에 속한 기업들 모두가 25%~85% 이익이 향상되는 효과를 발견하였다. 또한 (Xia and Jin, 2008)의 연구에서 새로운 고객을 확보하는 것은 기존의 고객을 유지하는 것보다 5~6배 비용이 발생한다고 하였다. 따라서 기업의 측면에서 볼 때 새로운 고객을 확보하는 것도 중요하지만 기존의 고객을 유지하여 발생하는 비용을 절감하는 것이 필요하다.

특히, 본 연구에서 살펴볼 산업인 카드 산업에서는 고객의 이탈을 기업과 회원 간의 계약이 종료되거나 끝나는 ‘회원 탈퇴’와 같은 의미로 널리 인식되고 있으나, 고객의 휴면과 관련하여 명확한 정의가 없는 것으로 파악된다. 또한 금융당국의 규제 정책 변화로 인하여 신용카드 산업에서 고객의 휴면을 재정의하여 현실화할 필요성이 있다.

1.2 카드산업에서 고객 휴면 정의

금융 당국에서는 12개월 연속 미이용 카드에 대하여 자동 해지를 강제하고 있다. 고객의 신용카드 미사용 상태가 지속되면, 카드사와 회원간의 계약 관계가 자동으로 정지된다. 따라서 고객의 보유율이 높은 상태로 유지되기 위해서 휴면 고객에 관한 관리가 더욱 중요하다. 이러한 변화로 신용카드업에서 휴면 고객의 재정의가 필요하며 휴면에 돌입하기 전

고객을 정확하게 예측하는 것이 무엇보다 중요해졌다. 위와 같은 휴면 고객 예측의 필요성에 따라 본 연구에서는 신용카드사의 고객 데이터를 이용하여 휴면 고객 예측 연구를 진행하였다.

위에서 설명된 바와 같이 금융 당국의 정책으로 인하여 카드 산업에서는 고객의 이탈과 휴면의 경계가 모호해졌다. (Hyun, 2014)의 연구에서 12개월 동안 실적이 없는 고객을 휴면 및 이탈 회원으로 간주하지만 신용카드사에서 공식적으로 발표할 때 휴면 회원을 6개월간 연속 실적이 없는 고객으로 발표하는 경우가 있었다. 왜냐하면 과포화 되어있는 신용카드 시장의 경우 1명의 고객이 1장 이상의 신용카드를 소지하고 있을 가능성이 높으며, 6개월의 기간이라면 대체 카드를 적어도 한 번 이상 사용될 기간으로 충분할 것으로 판단된다. 또한 12개월 실적이 없는 고객을 휴면이라 정의하고 예측 분류를 진행한다면, 잠재 휴면 고객에 의해 발생하는 카드사의 손실이 긴 시간 동안 발생할 것으로 보인다. 따라서, 휴면 고객을 금융 당국 정책보다 엄격히 정의하는 차원에서 본 연구에서는 카드사에서 공식화 하고 있는 6개월 연속 실적이 없는 고객에 대하여 휴면 고객으로 정의하고 딥러닝 방법론인 LSTM 알고리즘을 활용하여 예측 분석을 진행하였다.

2. 선행연구 및 이론적 배경

2.1 고객 이탈과 휴면

Blattberg et al.(2008)의 연구에서 고객의 이탈에 관하여 정의한 바가 있었다. 이 연구에 따르면 기업이 고객 가치를 증대하기 위한 여러 가지 활동을 하는 중에도 고객은 언제든지 서비스나 재화를 이용 중단하기로 결정할 수 있으며, 이를 ‘고객 이탈(Customer Churn)’이라 하였다. 고객 이탈 및 휴면 고객을 예측한 선행 연구의 업종과 방법론은 Tab. 2-1에서와 같다.

Tab. 2-1에서 보이는 바와 같이 이탈 고객을 예측하는 연구는 여러 산업계에서 진행되고 있다. 2000년대 초반에는 Machine Learning(이하 머신러닝) 방법론 중 하나인 SVM을 이용한 예측 분석이 주를 이루었으며, Perceptron의 발전과 함께 인공신경망(Neural Network)방법론 또한 이탈 고객을 예측하는데 많이 사용되었다.

고객 데이터의 경우 대부분 시간의 흐름에 따라 변화하는 경우가 많다. 하지만 기존 연구에서 사용된 분류 알고리즘들은 정적 예측 분류 모형으로써 시계열 형태로 존재하는 고객 데이터를 활용하여 예측하기엔 적절하지 않다. (Holtrop et al., 2017; Yang et al., 2005)의 연구에서 시계열로 존재하는 데이터를 정적 모형에 적용하기 위해 알맞은 형태로 데이터를 가공해야 하며 전처리 과정에서 데이터의 손실이 불가피하다고 지적된 바 있었다. 또한 (Morik and Kopcke, 2004)의 연구에서는 고객 데이터가 시간에 따라 변화하기 때문에 시간의 속성을 고려한 분석이 중요하다고 하였다. 위의 연구에서 지적된 한계점을 극복하기 위해 본 연구에서는 시간

의 흐름을 잘 학습 시킬 수 있는 방법론인 Recurrent Neural Network(RNN)을 활용하였으며, 기존 RNN의 한계를 보완할 수 있는 Long Short-Term Memory(LSTM)을 사용하였다.

Tab. 2-1과 같이 여러 분야에서 머신러닝, 딥러닝을 활용한 고객 이탈 연구는 활발히 진행되고 있으나, 카드 산업에서의 고객 휴면 이탈 예측연구는 지지부진한 것으로 보인다. 따라서, 카드 산업에서 고객 휴면을 예측한 본 연구는 학술적으로 큰 의미를 가질 수 있을 것으로 판단된다.

2.2 통합기술수용이론 (UTAUT)

통합기술수용이론(UTAUT)이 의미하고자 하는 바는 사람의 태도가 사용 의도를 결정하고 사용 의도가 결국 사용 행위를 결정한다는 것이다. 따라서, 새로운 정보기술이 등장할 때 연구자들이 기술의 사용을 알아보고자 통합기술수용이론(UTAUT)을 통하여 요인을 분석하곤 했다. 현금결제만 이루어졌던 시기에서 신용카드의 등장은 결제의 편의를 가져온

Tab. 2-1 Literature Review about Customer Churn

연구자 및 선행연구	분야	방법론
Morik and Kopcke, (2004)	보험업	SVM
Yang et al., (2005)	통신업	SVM
Hur and Lim, (2005)	보험업	Neural Network
Xia and Jin, (2008)	통신업	Neural Network
Xie et al., (2009)	금융업	Random Forests
Tsai and Lu, (2009)	통신업	Neural Network
Wong, (2011)	통신업	Cox Regression
Sharma and Panigrahi, (2011)	통신업	Neural Network
Chen et al., (2012)	통신업	Random Forests
Altas and Gulpinar, (2013)	통신업	Neural Network
Ascarza and Hardie, (2013)	구독	Hidden Markov Model
Ismail et al., (2015)	통신업	Perceptron
Lee et al., (2017)	통신업	Neural Network

새로운 기술의 등장이라 할 수 있다. 또한, 신용카드사는 각기 다른 여러 가지 서비스를 제공하고 있으며, 이를 통해 고객들이 카드를 이용하는 것은 서비스를 같이 수용하는 것으로 판단할 수 있다. 단순히 결제의 편의로 인해 사용되고 인식되었던 과거의 신용카드에서 현대에 들어서는 결제 편의성 기술 수용과 동시에 신용카드사에서 제공하는 서비스를 채택하는 것이다. 따라서 기술 수용과 동시에 서비스 채택에 관하여 신용카드 사용 요인을 변수로써 재정립하는 것은 신용카드 기업의 입장에서 매우 중요한 과제이다.

Park and Ahn (2012)에서는 통합기술수용이론(UTAUT)은 기술의 수용에 관한 요인을 연구함과 동시에 사람들이 서비스를 수용하는 요인에 대하여 연구하였다. 또한 많은 연구에서 새로운 기술에 관하여만 통합기술수용이론(UTAUT)을 활용하는 것

이 아닌 사람들의 서비스 수용에 관하여도 많이 활용되는 것을 알 수 있다. 또한 새로운 상품을 출시하는 것은 새로운 기술의 동반을 의미한다. 예를 들어, Payco 결제 서비스의 경우 모바일 애플리케이션을 통한 핀테크 결제 기술과 동시에 새로운 방식의 결제 서비스로서 무형의 상품을 제공하는 것이다. 이러한 이유에서 새로운 기술을 수용하는 것과 서비스를 수용하는 것은 이원화하여 바라볼 것이 아니라 함께 고려해야 할 필요성이 있다.

신용카드는 무형의 상품이며 기존의 현금 결제와 비교했을 때 새로운 기술을 가진 서비스이다. 고객이 신용카드를 사용하는 것은 기술의 수용과 동시에 무형의 상품인 서비스를 함께 선택하는 것이라 생각할 수 있다. 따라서 신용카드의 경우에도 기술과 서비스의 수용에 관하여 함께 생각하여야 한다.

Tab. 2-2 Define Variables with UTAUT

연구 변수	변수 설명	활용 변수	관련 연구
성과 기대	카드 결제로 높은 가치의 상품을 구매하는데 도움을 줄 수 있는 할부 소비 정보	할부 금액 할부 건수 무이자 할부	(Davis, 1989) (Davis et al., 1992) (Venkatesh et al., 2003)
노력 기대	상품이나 서비스를 구매할 때 신용카드를 이용한 소비 정보	소비 금액 이용 건수 현금 서비스	(Davis, 1989) (Davis et al., 1992) (Venkatesh et al., 2003) (Davis et al., 1989)
사회적 영향	주변인들의 카드사용에 영향을 줄 수 있는 혜택 정보	캐쉬백 포인트 마일리지	(Mathieson, 1991) (Thompson and Higgins, 1991) (Venkatesh et al., 2003)
촉진 조건	카드 결제시 본인의 카드 이용 활동을 지원할 수 있는 조직적 환경이 조성되어 있는 정보	소비 라이프스타일 App 이용	(Ajzen, 1991) (Venkatesh et al., 2003)
사용 의도		카드 휴면 미 휴면	(Davis et al., 1989) (Venkatesh and Goyal, 2010)

Tab. 2-3 LSTM Variables

독립변수 / 종속변수	LSTM 활용 변수
예측 활성 고객	성별
	연령
	수입
	소비 금액
	이용 건수
	할부 금액
	할부 건수
	Cash back (캐시백)
	포인트 사용
	라이프 스타일 소비율(쇼핑, 건강, 레저, 주유소 등)
종속변수	휴면 / 미휴면

위와 같은 이유로 통합기술수용이론(UTAUT)은 카드 산업에서 활용되어 고객들이 카드를 사용하는 요인에 대하여 파악할 필요성이 있다. 기존 통합 기술수용이론(UTAUT)에서 정의된 연구 변수에 의하여 신용카드 산업에서 재정의하고 고객의 카드 사용에 대한 변수를 선정하였다.

신용카드 산업에서 휴면 고객을 포착하고 예측하는 것은 결국 카드의 이용과 관련이 있는 이슈이다. 고객들은 신용카드를 변수에 의해서 사용하고 사용하지 않는가에 대한 것을 알아보는 사용 요인 선정을 통해 변수를 선정하는 것은 고객 휴면을 예측하는 연구에서 가장 중요한 부분이 될 수 있다. 따라서 본 연구에서는 머신 러닝과 딥러닝을 통하여 고객 휴면 예측을 하되 변수의 선정에 있어 통합기술수용이론(UTAUT)을 활용하였다.

Tab. 2-2에서 보이는 바와 같이 각 연구 변수에 의한 카드 고객 변수를 설정할 수 있다. 성과 기대 연구 변수에서 현금 서비스 등 여러 실적을 나타낼 수 있는 변수가 존재하지만 소비 금액이라는 변수에 포함되는 변수이다. 따라서 본 연구에서는 변수들이 서로 중복되지 않게 변수의 포함 관계를 고려하여 설정하였다.

통합기술수용이론(UTAUT)의 성과 기대, 노력 기대,

사회적 영향, 촉진 조건의 정의를 고려하여 신용 카드 사용 및 신용 카드의 서비스 이용에 대하여 재정의 하였다. 성과 기대는 ‘새로운 기술을 사용하는 것이 업무의 성과를 달성하는데 도움을 줄 것이라 믿는 정도’라 정의했었다. 이를 카드 산업에 맞추어 보았을 때 업무의 성과 달성은 신용카드 사용으로 가치가 높은 상품이나 서비스를 구매하는 것으로 볼 수 있다. 따라서 Tab. 2-2와 같이 성과 기대를 정의하였다. 노력 기대는 ‘새로운 기술을 사용하는 것이 쉬울 것이라고 인식하는 정도’로 정의되었다. 이를 카드 산업에 빚대어 본다면 신용카드를 쉽게 이용할 수 있어 상품을 편리하게 구입할 수 있다고 표현할 수 있고 신용 카드의 사용 실적과 같이 생각할 수 있다. 위와 같은 현상에 입각하여 노력 기대에 대한 요인으로 카드 고객 변수로써 실적 정보를 선정하였다. 사회적 영향은 ‘개인에게 많은 영향을 미치는 주변인들의 입장에서 판단했을 때 본인이 새로운 기술을 수용하고 활용해야 한다고 믿는 정도’라 정의되었다. 신용 카드 산업에서 주변인이 사용하는 카드의 혜택이 개인에게 소비 생활에 있어 도움을 준다고 판단하여 신용카드를 사용하는 경우가 있다. 따라서 카드 고객의 변수를 신용카드의 혜택 정보를 선정하였다. 촉진 조건은 ‘새로운 기술을 활

용하는데 필요한 조직이나 기술적인 기반을 갖추고 있다고 믿는 정도'라 정의된다. 신용카드사는 여러 가맹점을 통해 고객들에게 서비스를 제공한다. 예를 들어 외식 산업에서 어떤 가게를 이용하면 포인트를 쌓을 수 있거나 할인을 받고 캐시백 혜택을 받을 수 있다. 이러한 소비 환경에서 본 연구자는 월별 고객이 어떤 범주의 가맹점에서 이용하였는지에 대한 정보 즉, 쇼핑 이용 비율, 건강 관련 이용 비율, 레저 관련 이용 비율, 주유소 이용 비율, 자기 관리 이용 비율, 소액 결제 이용 비율, 외식 관련 이용 비율을 선정하여 모형의 독립변수로 활용하였으며, 모형에 활용된 변수는 Tab. 2-3와 같다.

3. LSTM을 활용한 고객 휴면 예측

3.1 고객 휴면 예측 및 모형 평가

본 연구에서는 국내 대기업 신용카드 고객 데이터에 LSTM을 적용하여 휴면 고객 예측 분석을 진행하였다. 신용카드 고객 데이터는 월별 연속적이며 고객이 한 달간 소비한 정보가 나타나있다.

데이터는 약 700만 명의 고객 중 유효 고객 20,000명 휴면 고객 20,000명 총 40,000명의 고객을 랜덤 샘플링하였다. 2017년 3월부터 2018년 2월 총 12개월의 데이터를 학습하는데 사용하였으며 휴면 포착 기간은 2018년 3월부터 2018년 8월 까지 총 6개월로 설정하여 예측하였다.

유효 고객과 휴면 고객의 월평균 신용카드 이용 추이는 다소 차이가 있는 것으로 나타났다. 전체적으로 하락하는 추세를 따르는 것은 두 집단 모두 비슷하였지만 유효 고객의 신용카드 이용에서 상승과 하락이 반복되며 전체적으로 하락한 반면 휴면 고객의 경우 단조 감소하는 형태의 그래프를 볼 수 있었다. 이러한 현상은 휴면 고객과 유효 고객을 분류하는데 독립변수로서 큰 비중을 차지할 것으로 보인다. 반면 월별 소비 금액과 이용 건수 이외의 변수

들에 관하여 효과가 미비할 수 있다는 것이다. 따라서 본 연구에서는 휴면의 포착 시점인 2018년 3월과 시차를 둔 후 학습 개월 수를 줄여 가며 몇 개월 데이터를 학습하였을 때 모형의 분류 성능이 어떻게 변화하는지 살펴보았다. 모형의 평가 지표로 정확도(Accuracy)와 F-1 score를 활용하였으며, 모형의 안정성을 살펴보기 위하여 Hit-ratio를 활용하여 모형의 평가지표로 활용하였다.

고객의 12개월 데이터를 시작으로 1개월씩 줄여 가며 예측하였다. LSTM 알고리즘의 경우 학습된 개월 수가 많을수록 높은 정확도를 기록하였다. 즉, 휴면에 임박한 정보를 담고 있을수록 휴면 고객을 분류하는데 더욱 효과를 발휘하는 것으로 보였다. 하지만 학습되는 데이터가 줄어들수록 LSTM의 결과는 가파르게 하강하는 것을 볼 수 있었으며, 12개월 학습 결과와 2개월 학습 결과의 차이가 크게 나타났다. 이러한 결과로 보아 LSTM 알고리즘은 학습되는 개월 수가 학습 결과에 큰 영향을 끼친다는 것을 알 수 있었다.

UTAUT 정의에 입각하여 선정된 독립변수들을 LSTM 알고리즘에 적용하여 예측한 결과 약 83%의 정확도를 기록했다 (Tab. 3-2). 또한 Recall, Precision, F-1 score의 결과와 같이 정확도와 큰 차이를 보이지 않았으며, 이는 LSTM의 예측력이 안정적이라 할 수 있다. 즉, 휴면 고객에 초점을 맞추어 본 결과에서 더욱 세밀하게 휴면 고객을 잘 예측하였다. 기존 정적 모형으로 예측되었던 고객 이탈 및 휴면 예측에 있어서 각 고객들의 평균치 데이터를 사용하였다. 이러한 결과로 실적이 없는 개월 수 정보를 예측 모형에 담지 못하는 경우가 발생한다. 하지만 LSTM의 경우 12개월 동안 발생한 모든 실적 정보를 이용하기 때문에 실적이 없는 고객의 패턴 정보를 담을 수 있다. 따라서 월별로 발생하는 고객 정보를 학습하기 위해선 LSTM 알고리즘이 더욱 효과적이며 알맞은 예측 방법이 될 수 있다.

Hit-ratio는 1등급에서 10등급으로 갈수록 실제

휴면 고객의 수가 줄어들어야 모형이 안정적이라 할 수 있다. LSTM의 Hit-ratio를 살펴보면 실제 휴면 고객의 수가 상승과 하락이 반복되는 것이 아닌 계속 감소하는 형태를 보였다 (Tab. 3-4). 여러 평가 지표를 보아 통합기술수용이론에 입각하여 선정된

독립변수들을 이용하여 휴면 고객을 예측할 때 LSTM을 사용하여 예측하는 것이 더욱 안정적인 결과를 가져다 준다는 것을 알 수 있었다.

Tab. 3-1 LSTM Confusion Matrix

Confusion Matrix	실제 휴면 고객	실제 활성 고객
예측 휴면 고객	3,814	806
예측 활성 고객	582	3,194

Tab. 3-2 LSTM Performance

측정 지표	Performance
Accuracy	0.827
Recall	0.799
Precision	0.846
Specificity	0.856
F-1 Score	0.823

Tab. 3-3 Performance by Number of Months Learned

학습 개월 수	Accuracy
12개월	0.827
11개월	0.818
10개월	0.800
9개월	0.783
8개월	0.768
7개월	0.753
6개월	0.736
5개월	0.725
4개월	0.709
3개월	0.692

Tab. 3-4 LSTM Hit-ratio

Grade	실제 휴면 고객	Ratio
1	769	0.961
2	736	0.920
3	721	0.901
4	628	0.785
5	435	0.544
6	270	0.338
7	150	0.188
8	133	0.166
9	97	0.121
10	61	0.076

3.2 인구통계학적 정보의 조절효과

통합기술수용이론(UTAUT)를 통해 고객의 월별 변수를 선정하여 예측 분석에 활용했다. 그 결과 LSTM 알고리즘의 타 알고리즘 대비 예측력이 우수했으며, 타 알고리즘의 비교 결과는 Tab. 3-5와 같다. 하지만 (Park and Ahn, 2012; Jeon et al, 2011)의 연구에서 주장된 바와 같이 인구통계학적 정보는 조절 변수로써 역할을 할 수 있다. 조절 변수에 따른 조절 효과는 어떤 한 변수의 클래스에 따라 다른 변수들의 영향력이 달라져 모형의 결과가 달라지는 효과를 의미한다. 예를 들어 성별에 따라 분류 알고리즘의 예측력이 달라진다면 성별은 휴면 고객을 예측하는 모형에 있어 조절 효과를 가져오는 조절 변수인 것이다. 따라서 본 연구에서 사용된 고객의 인구통계학적 정보의 조절 효과를 알아보고자 성별, 연령, 추정 수입을 클래스 별로 나누어 기존의 모든 변수를 활용한 모형의 예측력과 비교하고자 하였다. 가장 안정된 예측력을 보였던 LSTM모형을 사용하였으며, 12개월 모든 데이터를 이용하여 예측 성능을 비교하였다.

먼저 성별에 따라 결과가 어떻게 달라지는지 남성, 여성의 예측력 차이와 선정된 모든 변수를 활용

한 LSTM 알고리즘의 예측력 차이를 비교해 보았다. 총 40,000명의 데이터에서 남성, 여성의 성비는 20398 : 19602로 약 5.1 : 4.9였다. 기존 실험 방법과 똑같은 조건으로 Training set: Validation set: Test set을 6:2:2로 나누어 진행하였으며, 모형의 예측 결과는 다음 Tab. 3-6과 같았다.

아래 결과 남성은 0.833, 여성은 0.841의 정확도를 기록하였다. 앞서 진행한 모든 변수를 포함한 모형의 결과는 0.842로 0.001 ~ 0.009 정도의 예측력 차이가 있었다. 모든 변수를 이용하여 예측한 모형이 근소하게 높은 정확도를 기록하였지만 눈에 띄는 차이를 볼 수 없었다. F-1 Score의 경우 남성은 0.822, 여성은 0.840을 기록하였으며, 기존 0.837보다 남성은 0.015 낮았으며, 여성은 0.003 높았다. 하지만 전반적으로 예측 정확도와 F-1 Score 모두 큰 차이를 발견할 수 없었다.

다음으로 연령에 따라 예측력이 어떻게 변화하는지 알아보았다. 연령의 경우 20, 30대, 40대, 50대, 60대 이상으로 분할하였으며, 그 비율은 11,997 : 10,612 : 9,212 : 8,179로 약 3 : 2.7 : 2.3 : 2였다. 연령대의 경우도 위의 성별과 마찬가지로 비교적 균형된 데이터였다. LSTM의 조건은 위와 동일했으며 예측력은 Tab. 3-7과 같았다. 연령대에 따

른 모형의 예측 정확도는 20, 30대의 경우에서 가장 높은 것으로 나타났다. 하지만 전반적인 예측력과 큰 차이를 보이는 F-1 Score를 기록하였다. 반면 60대 이상의 경우에서 정확도는 20, 30대 보다 낮았지만 F-1 Score는 높은 수치를 도출하였다. 모든 변수를 사용한 모형의 결과와 비교해 보면 20, 30대의 결과 보다 0.001 낮았으며, F-1 Score는 0.014 높은 수치였다. 50대의 결과 보다 정확도는 0.021 높았으며, F-1 Score도 0.015 높았다. 하지만 성별의 결과와 마찬가지로 예측력과 F-1 Score에 있어서 큰 차이를 찾아볼 수 없었다.

마지막 인구 통계학적 변수인 추정 수입에 관하여 실험을 진행해 보았다. 먼저 3,000만 원 미만, 3,000만 원대, 4,000만 원 이상으로 분할하여 예측 분석을 하였다. 분할된 데이터의 비율은 11,141 : 18,403 : 10,456 이었으며, 약 2.6 : 4.6 : 2.8 이었다. 앞서 성별과 연령대 정보를 분할한 데이터 보다 비교적 불균형된 데이터였으며, 예측력은 Tab. 3-8과 같으며, 이어서 3,000만원 대 고객들을 3,500만 원을 기준으로 나누어 예측한 결과는 Tab. 3-9과 같다.

Tab. 3-8의 결과에서 추정 수입이 3,000만 원 미만인 고객들의 예측력과 F-1 Score는 모든 변수를 활용한 결과 보다 0.015, 0.037가량 높았으며 조절 효과를 측정하고자 한 실험에서 가장 우수한 분류 성능을 볼 수 있었다. 각 추정 수입을 바탕으로 분할된 모형들의 결과에서도 차이가 있었다. 3,000만 원 미만의 고객 데이터를 활용한 모형의 예측력이 가장 좋았으며, 정확도는 0.04 ~ 0.023,

F-1 Score는 0.083 ~ 0.048 정도 차이가 있었다. 3,500만 원을 기준으로 3,000만원 대 고객을 나누는 데이터에서 비율은 3,000만원 이상 3,500만원 미만 : 3,500만원 이상 : 4,000만원 미만 = 9,366 : 9,037로 약 5:5로 분할될 수 있었다. Tab. 3-9의 결과와 같이 두 그룹의 예측력은 차이가 없었으며, F-1 Score의 경우 약 0.009 가량 3,500만 원 이상 4,000만 원 미만의 그룹이 더 높았다. 모든 변수를 이용한 모형의 분류 예측 성능과 정확도는 0.023 차이가 있었으며, F-1 Score는 0.033 ~ 0.024만큼 차이를 나타내었다.

통합기술수용이론(UTAUT)에서 발생할 수 있는 인구통계학적 정보의 조절효과에서 대해서 실험해 보았다. 그 결과 성별과 연령대를 분할한 모형의 결과는 전체 변수를 모두 활용한 모형의 결과와 크게 다르지 않았다. 조절 효과를 본 실험에서 가장 좋은 분류 예측력을 보였던 그룹은 추정 수입에 따라 분할된 데이터의 3,000만 원 미만의 그룹 모형 결과였다. 하지만 모든 결과를 종합적으로 비교해 보았을 때 모든 변수를 이용한 모형의 결과와 각 인구통계학적 변수를 바탕으로 분할된 데이터의 결과는 크게 다르지 않았다. 이러한 수치 비교를 보았을 때 인구통계학적 정보는 뚜렷하게 큰 차이를 주는 조절 변수로서 효과는 없었던 것으로 판단된다. 결론적으로 통합기술수용이론(UTAUT)를 바탕으로 선정된 변수들을 이용할 때 인구통계학적 변수를 조절 변수로서 조정하는 것이 아닌 모든 변수를 독립변수로 활용하여 LSTM에 적용하여도 큰 차이가 없는 결과를 얻을 수 있을 것으로 보인다.

Tab. 3-5 Logistic Regression, SVM, Random Forest, LSTM Accuracy & F-1 Score

	Logistic Regression	SVM	Random Forest	LSTM (RNN)
Accuracy	0.773	0.761	0.769	0.827
F-1 Score	0.770	0.752	0.762	0.823

Tab. 3-6 LSTM Accuracy & F-1 Score by Gender

	남성	여성	Total
Accuracy	0.833	0.841	0.842
F-1 Score	0.822	0.840	0.837

Tab. 3-7 LSTM Accuracy & F-1 Score by Age

Confusion Matrix	20,30대	40대	50대	60대 이상	Total
Accuracy	0.843	0.822	0.821	0.829	0.842
F-1 Score	0.823	0.816	0.822	0.833	0.837

Tab. 3-8 LSTM Accuracy & F-1 Score by Income(1)

	3,000만 원 미만	3,000만 원대	4,000만 원 이상	Total
Accuracy	0.857	0.834	0.817	0.842
F-1 Score	0.874	0.826	0.791	0.837

Tab. 3-9 LSTM Accuracy & F-1 Score by Income(2)

	3,000만 원 ~ 3,500만 원	3,500만 원 ~ 4,000만 원	Total
Accuracy	0.819	0.819	0.842
F-1 Score	0.804	0.813	0.837

4. 결론

본 연구에서는 과포화 상태에 직면해 있는 대한민국의 카드산업에서 휴면 고객을 예측하는 연구를 진행하였다. 서비스의 이용 및 기술 수용과 관련된 이론인 통합기술수용이론(UTAUT)에 입각하여 신용카드 고객의 카드 활용 요인을 독립변수로 재정의하여 예측 분석을 실시하였다. 휴면 고객을 예측하기 위해 딥러닝 방법론 중 하나인 RNN을 사용하였으며, 그중 LSTM알고리즘을 활용하였다. 또한 정확도와 F-1 score를 평가 지표로써 사용하였으며, 알

고리즘을 통해 나온 휴면 예측 확률이 높은 고객부터 낮은 고객까지 10분위로 나누어 실제 휴면 고객 적중률을 볼 수 있는 Hit-ratio를 활용하였다. 마지막으로 통합기술수용이론(UTAUT)에서 발생할 수 있는 조절효과에 대해서 알아봄으로써 통합기술수용이론(UTAUT)과 머신러닝을 접목하여 활용할 것인가에 대해서도 알아보았다.

기존 연구들과 달리 신용카드 산업에서 고객들의 활동 정보를 통합기술수용이론(UTAUT)에 따라 재정의 함으로써 앞으로 있을 카드 산업에서 사용될 독립변수의 선정에 큰 기여를 할 수 있을 것이라 예상된다.

통합기술수용이론(UTAUT)을 활용하여 통계적 분석을 진행했을 때, 인구통계학적 정보에서 조절효과가 발생하는 것으로 지적되었다. 본 연구에서 통계적 방법론을 활용하여 분석된 것이 아니라, 통합기술수용이론을 비롯하여 여러 이론에 입각하여 선정된 변수와 머신 러닝, 딥러닝 방법론과 결합한다면, 강건한(Robust) 분석 결과를 얻을 수 있을 것으로 판단되며 앞으로의 연구 방향에 기여가 있을 것으로 보인다.

또한, 본 연구에서는 휴면 포착 시점 3개월 전 데이터까지 활용하여 예측하였다. 이는 휴면에 돌입하기 3개월 전에 휴면 고객을 예측할 수 있다는 것이다. 이러한 연구 및 분석 방법을 통해 실제 기업에서 휴면으로 돌입할 고객을 3개월 전에 먼저 인지하고, 3개월간 휴면으로 예측된 고객에게 CRM 너지, 등 다양한 마케팅 활동을 통해 휴면에 돌입할 고객을 다시 활성 고객으로 전환 시킬 수 있을 것으로 보이며, 고객 관리 측면에서 한층 고도화 가능할 것으로 판단된다.

[References]

- [1] Ajzen, I.(1991), The Theory of Planned Behavior, *Organizational Behavior and Human Decision Processes*, 50(2), 179-211
- [2] Altas, D. and Gulpinar, V.(2013), Customer Churn Analysis through Artificial Neural Networks in Turkish Telecommunications Market, *International Journal of Economic Perspectives*, 7(4), 63-80
- [3] Ascarza, E., and Hardie, B.G.S.(2013), A Joint Model of Usage and Churn in Contractual Settings, *Marketing Science*, 32(4), 570-590
- [4] Blattberg, R. C., Kim, B.D and Scott A.N. (2008), Database Marketing: Analyzing and Managing Customers. International Series in Quantitative Marketing, Springer
- [5] Chen, Z.Y., Fan, Z.P. and Sun, M.(2012), A Hierarchical Multiple Kernel Support Vector Machine for Customer Churn Prediction Using Longitudinal Behavioral Data, *European Journal of Operational Research*, 223(2), 461-472
- [6] Davis, F.D.(1989), Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly*, 13(3), 319-340
- [7] Davis, F.D., Rechar, P.B. and Paul R.W. (1992), Extrinsic and Intrinsic Motivation to Use Computers in the Workplace, *Journal of Applied Social Psychology*, 22(14), 1111-1132
- [8] Holtrop, N., Wieringa, J.E. and Gijsenberg, M.J.(2017), No Future without the Past? Predicting Churn in the Face of Customer Privacy, *International Journal of Research in Marketing*, 34(1), 154-172
- [9] Hur, Y. and Lim, S.H.(2005), Customer Churning Prediction Using Support Vector Machines in Online Auto Insurance Service, *Advances in Neural*

Networks-ISBNN 2005, 928-933

- [10] Hyun, Y.J.(2014), Policies to Increase Profitability of the Credit Card Industry in Korea, *Industrial Innovation Research*, 30(1), 37-68 (현영진(2014) 우리나라 신용카드 산업의 수익성 활성화 방안. 산업혁신연구, 30(1), 37-68)
- [11] Ismail, M.R., Awnag, M.K., Rahman, M.N. and Mokhairi. M.(2015), A Multi-Layer Perceptron Approach for Customer Churn Prediction, *International Journal of Multimedia and Ubiquitous Engineering*, 10(7), 213-222
- [12] Jeon, S.H., Park, N.R. and Lee, C.C. (2011), Study on the Factors Affecting the Intention to Adopt Public Cloud Computing Service, *Entrue Journal of Information Technology*, 10(2), 97-112 (전세하, 박나래, 이중성, (2011) 공공부문 클라우드 컴퓨팅 서비스 사용의도에 영향을 미치는 요인에 관한 연구. *Entrue Journal of Information Technology*, 10(2), 97-112)
- [13] Lee, E.B, Kim, J.H. and Lee, S.G.(2017), Predicting Customer Churn in Mobile Industry Using Data Mining Technology, *Industrial Management & Data System*, 117(1), 90-109
- [14] Mathieson, M.(1991), Predicting User Intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior, *Information Systems Research*, 2(3), 173-191
- [15] Morik, K. and Kopcke, H.(2004), Analysing Customer Churn in Insurance Data - A Case Study, *Knowledge Discovery in Databases: PKDD 2004*, 325-336
- [16] Park, I.S. and Ahn, H.C.(2012), A Study on the User Acceptance Model of Mobile Credit Card Service based on UTAUT, *The E-Business Studies*, 13(3), 551-574 (박일순, 안현철, (2012) UTAUT 기반 모바일 신용카드 서비스의 사용자 수용 모형에 관한 연구. *e-비즈니스연구*, 13(3), 551-574)
- [17] Reichheld, F.R. and Sasser, W.E.(1990), Zero Defection: Quality Comes to Service, *Harvard Business Review*, Sep-Oct
- [18] Sharma, A. and Panigrahi, P.K.(2011), A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services, *International Journal of Computer Applications*, 27(11), 26-31
- [19] Thompson, R.L. and Higgins, C.A. (1991), Personal Computing: Toward a Conceptual Model of Utilization, *MIS Quarterly*, 15(1), 125-143
- [20] Tsai, C.F. and Lu, Y.H.(2009), Customer churn prediction by hybrid neural networks, *Expert Systems with Applications*, 36(10), 12547-12553
- [21] Venkatesh, V., Morris, M.G., Davis, M.G.

- and Davis, F.D.(2003), User Acceptance of Information Technology: Toward a Unified View, *MIS Quarterly*, 27(3), 425-478
- [22] Venkatesh, V. and Goyal, S.(2010), Expectation Disconfirmation and Technology Adoption: Polynomial Modeling and Response Surface Analysis, *MIS Quarterly*, 34(2), 281-303
- [23] Wong, K.K.(2011), Using Cox Regression to Model Customer Time to Churn in the Wireless Telecommunications Industry, *Journal of Targeting, Measurement and Analysis for Marketing*, 19(1), 37-43
- [24] Xia, G.E and Jin, W.D.(2008), Model of Customer Churn Prediction on Support Vector Machine, *Science Direct*, 28(1), 71-77
- [25] Xie, Y.Y., Li, X., Ngai, E.W.T. and Ying, W.Y.(2009), Customer Churn Prediction Using Improved Balanced Random Forests, *Expert Systems with Applications*, 36(3), 5445-5449
- [26] Yang, Y., Yang, Q., Lu, W., Pan, J., Pan, R., Lu, C., Li, L. and Qin, Z.(2005), Preprocessing Time Series Data for Classification with Application to CRM, *AI 2005: Advances in Artificial Intelligences*, 133-142



Lee, Dong-Kyu (harry8256@naver.com)

After completing his master's degree in Business Informatics at Hanyang University, Lee Dong-kyu worked as a data analyst at Cheil Pengtai, a digital marketing company. He is currently working as a business analyst for the CLASS101 business strategy team, which is leading the educational platform, and is in the process of conducting a Ph.D. program in Business Informatics at Hanyang University. His research interest focused on solving business problems using quantitative data and applying business problems of computing algorithms.



Shin, Min-Soo (minsooshin@hanyang.ac.kr)

Minsoo Shin is a professor at the school of Business Hanyang University. He received his Ph.D. from Decision Support Group at Engineering Department of Cambridge University, UK. His current research interests include AI-based strategic decision making and contents industry analysis.

Prediction of Dormant Customer in the Card Industry

DongKyu Lee* · Minsoo Shin**

ABSTRACT

In a customer-based industry, customer retention is the competitiveness of a company, and improving customer retention improves the competitiveness of the company. Therefore, accurate prediction and management of potential dormant customers is paramount to increasing the competitiveness of the enterprise. In particular, there are numerous competitors in the domestic card industry, and the government is introducing an automatic closing system for dormant card management. As a result of these social changes, the card industry must focus on better predicting and managing potential dormant cards, and better predicting dormant customers is emerging as an important challenge.

In this study, the Recurrent Neural Network (RNN) methodology was used to predict potential dormant customers in the card industry, and in particular, Long-Short Term Memory (LSTM) was used to efficiently learn data for a long time. In addition, to redefine the variables needed to predict dormant customers in the card industry, Unified Theory of Technology (UTAUT), an integrated technology acceptance theory, was applied to redefine and group the variables used in the model. As a result, stable model accuracy and F-1 score were obtained, and Hit-Ratio proved that models using LSTM can produce stable results compared to other algorithms. It was also found that there was no moderating effect of demographic information that could occur in UTAUT, which was pointed out in previous studies. Therefore, among variable selection models using UTAUT, dormant customer prediction models using LSTM are proven to have non-biased stable results.

This study revealed that there may be academic contributions to the prediction of dormant customers using LSTM algorithms that can learn well from previously untried time series data. In addition, it is a good example to show that it is possible to respond to customers who are preemptively dormant in terms of customer management because it is predicted at a time difference with the actual dormant capture, and it is expected to contribute greatly to the industry.

Keywords: Customer Churn, Prediction, UTAUT, Machine Learning, Deep Learning, LSTM

* First Author, Ph.D. Student, Business School, Hanyang University

** Corresponding Author, Professor, Business School, Hanyang University