

증권 금융 상품 거래 고객의 이탈 예측 및 원인 추론

A Securities Company's Customer Churn Prediction Model and Causal Inference with SHAP Value

나광택 · 이진영[†] · 김은찬 · 이효찬

데이터애널리틱스랩*

요약

산업 분야를 막론하고 머신러닝의 관심이 매우 높아지고 있으나, 머신러닝이 지닌 설명 불가능성은 여전히 문제로 남아있어 적극적인 업무 적용에 어려움이 있다.

본고에서는 증권사 금융 고객을 대상으로 이탈예측 모델 개발 사례를 소개하고 SHAP Value 기법을 사용하여 설명 가능한 머신러닝 모델 개발 시도와 해석 가능성 도출에 대한 연구 결과를 소개한다. 총 6가지 고객 이탈 모델을 비교 분석하였으며, SHAP Value와 고객의 자산 변화에 따른 유형 분류 및 데이터 분석을 통해 고객 이탈 원인을 추론한다. 본 연구 결과를 토대로, 향후 마케팅 담당자의 실제 고객 마케팅 수행에 있어 원인 추론이 가능한 이탈 예측 결괏값을 사용하고 고객별 마케팅 여부를 점검하는 등의 종합적 판단 지표로 활용할 수 있을 것으로 판단된다.

- 중심어 : 고객 이탈 예측모델, 머신러닝, 설명가능성, 원인 추론, SHAP Value

Abstract

The interest in machine learning is growing in all industries, but it is difficult to apply it to real-world tasks because of inexplicability.

This paper introduces a case of developing a financial customer churn prediction model for a securities company, and introduces the research results on an attempt to develop a machine learning model that can be explained using the SHAP Value methodology and derivation of interpretability. In this study, a total of six customer churn models are compared and analyzed, and the cause of customer churn is inferred through the classification and data analysis of SHAP Value and the type of customer asset change. Based on the results of this study, it would be possible to use it as a basis for comprehensive judgment, such as using the Value of the deviation prediction result that can infer the cause of the marketing manager's actual customer marketing in the future and establishing a target marketing strategy for each customer.

- Keyword : Churn Prediction Model, Machine Learning, XAI, Causal Inference, SHAP Value

I. 서 론

4차 산업혁명 도래로 인해 학계 전반에 걸쳐 인공지능 및 머신러닝 알고리즘에 대한 관심이 매우 높아지고 있다. 산업계에서도 머신러닝 알고리즘이 업무 전반에 적용되고 있으며, 사업개발, 기존 업무 개선, 리스크 관리, 마케팅 등 다양한 업무 분야에 적용 가능한 예측 모델 개발이 진행되고 있다. 특히 금융권에서는 고객의 과거 금융 거래 이력 등의 행태 정보를 통해 미래 행태를 예측하고, 그 결과값을 다양한 고객 관리 업무와 신규 상품 개발에 활용하는 추세[4, 9]이다.

하지만, 이러한 추세에도 불구하고 머신러닝이 지닌 고유한 특성 중 설명 불가능성으로 인해 적용 가능한 업무 영역의 확대와 실무 활용에 있어 어려움이 존재한다. 이에 따라 소위 “블랙박스”라 불리는 설명 불가능한 영역을 해결하기 위한 학계의 지속적인 연구가 진행되고 있고 산업계 또한 머신러닝의 활용 범위를 확대하기 위해 다방면으로 설명 가능한 인공지능 및 머신러닝의 적용 방안을 모색하고 있다[8, 9, 10, 11, 12, 14, 15].

[21, 22, 23, 24]에서 이탈 원인을 설명하려는 시도가 있었다. 하지만 대부분이 변수 중요도를 추출하는 기법에 지나지 않아 고객이 이탈하는 진짜 원인을 규명하는 것에는 한계를 보인다. 본 연구에서는 이탈에 영향을 미치는 중요 변수를 찾는 것에 머물지 않고 더 나아가 이탈 원인을 찾는 것에 그 목적이 있다. [13]에서 제안한 SHAP 기법을 통해 중요 변수를 찾아내고, 추가적인 데이터 분석을 통해 나온 결과를 이와 결합하여 고객의 숨은 이탈 원인을 추론한다. 우리가 알고 있는바에 의하면 변수 중요도에 그치지 않고 더 나아가 증권 금융 거래 고객의 숨은 이탈 원인을 찾고자 하는 시도는 본 연구가 처음이다.

본고에서는 실제 증권사의 온라인 대 고객 마케팅(CRM)에 활용 중인 고객 이탈 예측 모델 개

발 사례를 소개하고, SHAP Value 기법을 사용하여 설명 가능한 머신러닝 모델 개발 시도와 해석 가능성 도출에 대한 연구 결과를 소개한다.

II. 선행연구

2.1 이탈 예측 모델

금융권 고객 이탈 예측 모델의 선행연구를 살펴보면, 국내에서는 금융 데이터에 대한 학술적 접근이 해외에 비해서 상대적으로 어려운 면 때문에 대체로 고객 규모가 큰 은행과 카드 업권을 중심으로 다양한 목적에 맞는 예측 모델 연구가 진행되었다[6]. 선행연구 대부분은 다양한 머신러닝 알고리즘을 적용하여 예측 성능을 올리는 방법을 제시[1, 3, 6] 하거나 예측된 결과를 활용하여 대 고객 마케팅 지표 세분화에 응용하는 연구로[2, 5, 7] 구분된다. 알고리즘 측면에서 살펴보면, [21]은 통신업 분야에서 일반적인 인공 신경망과 self-organizing maps(SOM)을 결합한 하이브리드 모델을 이용해 고객 이탈을 예측하였다. 인공신경망은 학습 데이터의 차원 축소 역할을 수행하는 데 이때 노이즈를 감소시키는데 활용하였으며, SOM은 차원 축소된 데이터를 활용하여 이탈 예측을 수행하였다. 하지만 차원 축소 과정에서 데이터 손실이 발생하는 큰 단점이 있다. [22]는 은행업 분야에서 SVM을 이용해 이탈 고객 예측 방법을 제안하였다. 데이터 세트의 불균형 문제를 해결하고자 랜덤 샘플링 방법을 적용하였으나 단순 랜덤 샘플링은 데이터 세트의 분포를 유지하지 못한다는 단점이 있다. [23]에서 제안된 모델은 랜덤 포레스트를 활용하여 통신 고객의 이탈 고객을 예측하였다. 또한 분류 후, 제안된 모델은 코사인 유사성을 이용하여 그룹별로 이탈 고객을 분류하여 그룹 기반 Retention 제안을 제공함으로써 이탈 고객의 데이터를 세분화한다. 또한 이탈의 근본 원인을 결정하는데 필수

적인 이탈 요인을 규명하였다. 하지만 이 방법은 변수 중요도 추출과 유사하며 개인별 이탈 원인이 아닌 그룹별 이탈 원인을 찾는 것으로 정확한 이탈 요인을 구분하기에는 한계가 있다. [24]에서는 SVM을 이용해 이탈 고객을 예측하였는데, 모델 설명을 위해 Recursive feature Elimination 기법을 사용해 중요 변수를 추출하였다. 하지만 이는 변수 중요도에 지나지 않으며 이탈 고객의 정확한 원인을 이해하긴 힘들다.

[25, 26]에서는 SHAP 기법을 활용한 통신사 고객의 이탈을 설명하는 연구를 제시하였는데, 다른 유사 연구와 마찬가지로 단순 변수 중요도 파악에 그치고 있다. 또한, 모형이 예측한 결과에 대한 변수 중요도를 평균으로 산출해 냈을 뿐, 개인 고객별 이탈을 설명하는 방법은 제시하지 못하였다. 본 연구에서는 변수 중요도 파악에 그치지 않고 고객 개인별 이탈 원인을 찾아내는 방법론을 제시함으로써, 마케팅 업무 현장에서 활용 가능한 수준의 이탈 예측 모형 활용 가능성을 도모한다.

2.2 설명 가능한 인공지능/머신러닝

설명 가능한 인공지능 및 머신러닝 연구에 있어 가장 핵심은 결괏값의 해석에 대한 문제인데, 두 가지 대표적인 기법이 학계에서 널리 사용되고 있다. 첫째는 LIME(Local Interpretable Model-agnostic Explanations)인데, 이 알고리즘은 입력값에 변화를 주었을 때 모델의 예측 값이 변화하는 수준에 따라 변수 중요도를 확인한다. 즉, 다른 변수를 통제하여 사용된 각각의 독립변수 값을 조정했을 때 모델의 예측 값이 크게 바뀌는 순으로 변수의 중요도가 높다고 분석[15] 한다. 결괏값의 근사한 값(Local Result)을 통해 역으로 모델의 작동법을 설명하려는 방식이다.

둘째는 SHAP (SHapley Additive exPlanations)은 게임이론에서 샐플리 값(Shapley Value)으로

합리적 분배 방식을 결정하는 방식에서 차용된 기법이다. 이 기법은 LIME 과 마찬가지로 머신러닝 알고리즘을 적용하여 개발된 모델의 결괏값을 설명하기 위해 고안되었으며 사용된 변수 유무에 따른 평균적 결괏값 변화를 통해 영향도를 추정하는 방식[13]을 따른다.

우리는 상기 두 기법 중, SHAP(이하 SHAP Value)을 이용해 금융상품 거래 고객 이탈에 영향을 미치는 변수를 찾아내고 이에 대한 데이터 분석을 통해 이탈 원인을 추론하는 연구 결과를 소개한다. 예를 들어, 투자 실패 후 이탈하는 고객은 단순 투자 실패와 원금 회복형 투자 실패 등 다양한 거래 패턴이 나타나는데 이에 대한 해석 가능한 원인을 추론하여 고객 이탈 전 마케팅 실무자들이 사전 대응할 수 있는 방안을 제시한다. SHAP Value를 본 연구에 채택한 이유는 선형 모델뿐만 아니라, 인공지능망의 비선형 모델까지 적용할 수 있기 때문이다. LIME 기법은 머신러닝 모델이 지역적으로 선형적인 특성을 나타낸다고 가정하지만, 왜 그렇게 가정하였는지에 대한 이론과 설명이 전무하다. 반면에, SHAP Value는 대조 설명(Contrastive Explanations)이 가능하여, LIME의 지역적 모델(Local Model)에 없는 장점을 가지고 있다.

III. 활용 방법론

3.1 변수 선택 및 추출

변수 선택과 추출은 모델링에서 가장 중요한 요소 중 하나로서 모델의 예측 성능에 많은 영향을 미친다. 우리는 기본적인 인구통계학 정보뿐만 아니라 금융상품 거래 고객의 특징을 잘 나타낼 수 있는 자산 정보, 거래 정보 등을 주요 변수로 포함하였다.

인구통계학 정보 : 가장 기본적인 고객 정보로서 고객의 나이, 성별, 고객 주기, 고객 등급, 계좌

개설일, 무료 계좌 여부 등이 포함되어 있다.

거래 정보 : 고객 접촉 채널별 매수, 매도 횟수 및 금액 등이 포함되어 있으며, 거래 고객 이탈 예측에 있어 중요한 지표로 활용된다. 여기서 채널이란 MTS¹⁾, HTS²⁾, WTS³⁾, ARS⁴⁾ 등이 포함된다.

자산 정보 : 고객 전체 자산을 비롯해, 세부 내역이 포함되어 있다. 예수금, 주식 자산, 선물, 채권, CD/CP⁵⁾, RP⁶⁾, 파생상품, 펀드, 신탁, WRA⁷⁾, 퇴직연금 등으로 구성되어 있다.

온라인 행태정보 : 채널별 접촉 이력으로 채널별 접속 횟수, 체류 시간 등으로 구성되어 있다. 각각의 고객은 성향에 따라 다른 채널 접촉 행태를 보인다. 일반적인 이탈 고객의 경우 채널 접속 횟수와 체류시간이 점차 감소한다. 반면 체리피커(Cherry Picker⁸⁾)의 경우 예수금 입고 후 아무런 행태를 보이지 않다가 특정 기간이 지나면 이탈해 버린다. 충성도가 높은 고객은 자신이 선호하는 채널을 이용해 꾸준한 활동을 보인다.

파생 변수 : 주로 자산 정보를 이용해 파생 변수를 생성하였으며, $\frac{\text{금융상품 자산}}{\text{총자산}}$ 과 같은 비율 형식으로 구성하였다. 단기 투자성 고객은 주로 RP에 많은 자산을 보유하여, 다른 금융상품 대비 RP 자산 보유 비율이 높다.

3.2 정규화 및 전처리

변수 선택 및 추출에서 살펴본 여러 변수들은 그 형식과 스케일이 서로 다르다. 인공신경망과 같은 모델은 범주형 변수를 바로 사용할 수 없으며, 변수 간의 스케일에도 많은 영향을 받는다. 이를 해결하기 위해 범주형 변수는 이진수로 변

환하였으며, 변수 간 스케일 문제를 해결하기 위해 정규화 하였다.

인구통계학 정보 : 고객 등급과 같은 범주형 변수는 원-핫 인코딩(One-hot encoding)을 적용하였으며, 누락되거나 이상치가 존재하는 변수는 이상치를 평균 또는 최솟값으로 보간(Interpolation) 하였다.

거래 정보, 자산 정보, 온라인 행태정보 : 고객 간 거래 횟수 및 자산 정도의 차이가 매우 커 Z-Score 정규화를 사용하였으며 수식은 다음과 같다.

$$\frac{x - \mu}{\sigma} = z$$

x : 거래정보, μ : 평균, σ : 표준편차

Z-Score 정규화를 통하면 $z \sim N(0,1)$ 를 따르게 되어 변수 간 스케일 차이는 극복된다. 하지만 범주형 변수와의 스케일 문제는 여전히 남아있어 Sigmoid 함수를 이용해 값의 범위를 [0,1] 사이로 조정한다.

3.3 SHAP Value를 이용한 영향 변수 분석

실무자 입장에서 해석이 어려운 머신러닝 모델의 경우 마케팅 실행 영역에 적용하기 쉽지 않다. 마케팅은 타깃군에 대한 이해를 바탕으로, 상세한 기획안(e.g. 실행 채널, 마케팅 메시지, 오피, 캠페인 기간 등) 이 필요하다. 하지만, 머신러닝 모델의 예측 결과값은 모델링에 활용된 각 고객 변수에 대한 설명이 없기 때문에, 마케팅 기획 및

1) Mobile Trading System, 스마트폰 어플 기반 트레이딩 시스템
 2) Home Trading System, 컴퓨터 프로그램 기반 트레이딩 시스템
 3) Web Trading System, 웹 기반 트레이딩 시스템
 4) Automatic Response System, 전화 기반 트레이딩 시스템
 5) CD: 양도성정기예금증서, CP: 기업 어음
 6) 환매조건부 매매
 7) 일임자산운용전무가가 투자목적과 성향에 따라 포트폴리오 제공
 8) 제품이나 서비스 중 특정 요소만 골라 경제적 이득을 취하려는 소비 현상

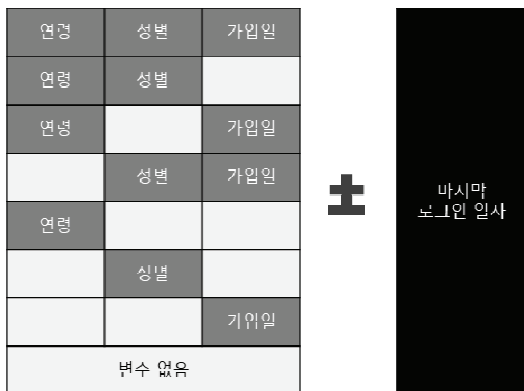
실행에 활용하기 어렵다. 하지만, SHAP Value 활용을 통해 모델링에 사용된 변수의 예측 영향도를 파악할 수 있어, 마케팅 분야에서는 예측 결과에 대한 이해를 바탕으로 한 전략 수립을 시도해 볼 수 있다. [12].

SHAP Value는 모델링 시 모든 변수 조합에 대해서 하나의 변수 기여도를 종합적으로 판단한다. 예를 들어, 특정 서비스를 이용 중인 고객이 3개월 뒤 이탈할 가능성을 예측하는 모델을 개발한다고 가정하자. 모델에 사용하는 변수는 총 4개(연령, 성별, 가입일, 마지막 로그인 일자)이고, 마지막 로그인 일자의 예측 기여도를 알고자 한다. 가능한 변수의 조합은 그림 1과 같이 총 8개이고, 각 조합에 대해서 마지막 로그인 일자 변수를 포함했을 때와 포함하지 않았을 때의 예측 값의 차이를 각각 계산하고 이들의 가중평균을 구한다.

이를 수식으로 아래와 같이 표현할 수 있다.[13]

$$\phi_j(v) = \sum_{S \subseteq \{x_1, \dots, x_p\} \setminus \{x_j\}} \frac{|S|!(p-|S|-1)!}{p!} (v(S \cup \{x_j\}) - v(S))$$

* $v(\cdot)$ 는 Value 로 모델의 예측 값



〈그림 1〉 SHAP Value 계산을 위한 가능한 변수 조합

여기서 S 는 모델에 사용된 변수들의 부분 집합이고, x_j 는 기여도를 관측하고자 하는 j 번째 변수,

p 는 활용 가능한 변수의 수이다. 오른쪽 항 $(v(S \cup \{x_j\}) - v(S))$ 은 x_j 변수가 포함되지 않았을 때와 포함되었을 때 예측치 차이에 의한 기여도이고, 그 왼쪽 항 $\frac{|S|!(p-|S|-1)!}{p!}$ 은 모든 가능한 변수의 조합에서 x_j 변수의 가중 평균이다.

이와 같이 SHAP Value는 고객별 각 변수에 대한 영향도를 전체와 비교하여 측정할 수 있다. 이를 SHAP Value의 주요 특징인 대조 설명(Contrastive Explanations)이라고 한다. 만약, 특정 고객 군에 마지막 로그인 일자가 이탈에 가장 많은 영향을 미쳤고, 로그인하지 않은 기간이 길수록 기여도가 더 높은 것으로 파악되었다면, 고객의 서비스 방문율을 높여 이탈 방어 캠페인 전략 수립을 구상해볼 수 있다. 그러나, 이러한 변수 영향도 하나로 개별 고객의 이탈 원인을 설명하기에는 다소 무리가 있는데, 그 원인 및 해결 방안에 대해서는 4.7 이탈 추론 결과에서 다루도록 하겠다.

3.4 이탈 예측 방법론

이탈 예측과 관련된 예측 모델은 다양하나 본 연구에서는 모델을 크게 3가지 범주로 나누어 적용하였다. 먼저 시계열 특성을 잘 살릴 수 있는 순환 신경망(이하 RNN) 계열의 장단기 메모리(이하 LSTM)와 변수에 가중(이하 Attention)을 두어 번역에 주로 활용되는 시퀀스-투-시퀀스(이하 Seq2Seq with Attention) 알고리즘을 적용하였다. 또한, 인접한 데이터의 지역 정보(Local Information)을 활용할 수 있는 합성곱 신경망(이하 CNN)과 그 응용인 잔차 신경망(이하 ResNet)을 적용하였다. 본 연구에서는 인접한 시계를 인접한 지역 정보로 가정한다. 마지막으로 데이터를 통계 처리하여 적용할 수 있는 랜덤 포레스트(이하 Random Forest)와 서포트 벡터머신(이하 SVM)을 이용해 이탈 예측 모델링을 진행하였다. 채택한 6가지 모델의 개요는 다음과 같다.

3.4.1 LSTM

LSTM은 RNN 알고리즘의 응용 모델로 데이터에 장기간 의존성이 존재할 때 특히 유용하다 [16]. RNN 은 장기간 의존성을 해결하지 못하는 문제점이 있으나 이를 해결하는 것이 LSTM의 Cell State라는 요소이다. 이는 필요 없는 정보를 없애거나 새로 추가된 정보를 저장하는데 활용된다. 이러한 Cell State는 Gradient Highway 와 같은 역할도 수행해 기울기 소실도 방지한다. LSTM은 총 4단계로 구분되어 작동된다. 먼저 Forget gate로 과거 정보의 저장 유무를 결정한다.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Forget gate는 이전 은닉 상태(hidden state, h_{t-1})와 현재 데이터(x_t)를 입력으로 받는다. 다음은 Input gate로 현재 데이터의 저장 유무를 결정한다.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

현재의 cell state값에 새로운 정보를 얼마나 저장할지 결정하게 된다. 세번째로 Cell state를 업데이트하는 과정을 거친다.

$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t$$

Forget gate를 통해서 얼마나 정보를 제거할지, Input gate를 통해서 정보를 얼마나 추가할지 결정하게 된다. 마지막은 Output Gate로 어떤 출력값을 출력할지 결정하게 된다.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \circ \tanh(C_t)$$

3.4.2 Seq2Seq with attention

주로 언어 번역에 많이 사용되는 seq2seq [17] 알고리즘은 RNN으로 구성된 인코더-디코더로 이루어진다. 영어-한글 번역을 예로 인코더는 입력 영어 문장 시퀀스를 하나의 벡터(Context Vector)로 인코딩(압축) 하고 디코더는 인코딩된 벡터를 다시 디코딩(재현) 하여 한글로 번역하는 역할을 수행한다. 하지만, 하나의 벡터로는 입력 문장의 모든 시퀀스 정보를 담기 어렵다는 단점이 있다. 이러한 문제를 해결하기 위해 [18]에서는 seq2seq 알고리즘에 Attention 기법을 도입하였다. Attention 은 입력 문장 시퀀스의 개별 단어에 가중치를 두어 목표(Target) 단어와 중요도 관계를 찾는 역할을 수행한다.

3.4.3 CNN

CNN 은 인간의 시신경을 모사한 모델로 AlexNet [19] 이후로 딥러닝의 새로운 도약을 알렸다. Convolution과 Pooling 계층을 통해 인접한 데이터의 지역 정보 관계를 찾는데 좋은 성능을 보이며 본 연구에서는 시계열 데이터의 지역 정보를 활용하기 위해 CNN을 활용하였다.

3.4.4 ResNet

일반적인 CNN을 포함한 심층 신경망은 계층이 깊어지면 기울기 소실, 폭발 및 Degradation 이 발생한다. 기울기 소실 및 폭발 문제는 출력 층의 오차 값에 대한 미분 값이 입력 층까지 도달하지 못하고 중간에 소실되거나 폭발해 버리는 현상을 말한다. Degradation 문제는 비교적 최근에 발견된 문제로 층이 깊어지면 오히려 성능이 하락하는 현상을 말한다. 일반적인 심층 신경망은 항등 매핑(Identity Mapping)에 어려움이 있어 이러한 현상이 일어나는데, 이 문제를 해결하기 위해 [20]에서 예측 값과 관측 값의 차이인 잔차를 학습하는 방법을 제안하였다. 이러한 잔차 학습은 항등 매핑을 용이하게 해 신경망의 계층을 깊게 쌓을 수 있다.

3.4.5 Random Forest

하나의 결정 트리는 과적합에 매우 취약하지만, 앙상블 학습을 통해 과적합 문제를 해소할 수 있다. Random Forest는 Bagging이라는 과정을 통해 데이터와 특징(변수)의 부분집합을 뽑아 여러 개의 트리를 학습한다. 각 트리에서 분류한 결과에 대해 투표를 실시하여 최종 결과를 선택한다. Random Forest는 과적합 방지뿐만 아니라 변수의 스케일 등에 영향을 받지 않는다는 장점이 있다.

3.4.6 Support Vector Machine

SVM은 훈련 집합의 부분 집합인 서포트 벡터 간의 거리를 최대로 하는 초평면을 찾는 것을 목적으로 한다. 서포트 벡터는 각각의 클래스에 속한 데이터 중에서 초평면과 가장 가까운 데이터이다. 만약 데이터가 선형 분리가 불가능하다면, 커널 트릭을 이용해 고차원 공간으로 매핑하여 초평면을 찾아낸다. SVM 최적화 문제는 다음과 같이 나타낼 수 있다.

$$\min \frac{1}{2} \|W\|^2$$

$$s.t. y_i(w \cdot x_i + b) \geq 1$$

여기서 y_i 는 레이블이며 x_i 는 입력 데이터이다. 선형 분리가 가능한 데이터에선 SVM을 다음과 같이 나타낼 수 있다.

$$f(x) = w^T x + b = \sum_{i=1}^N y_i \alpha_i x_i + b$$

하지만 위 수식은 에러를 허용하지 않는다. 에러를 허용하기 위해 slack variable ξ 를 도입하면 다음과 같이 표현된다.

$$\min \frac{1}{2} \|W\|^2 + C \sum_{i=1}^N \xi_i$$

$$s.t. y_i(w \cdot x_i + b) \geq 1$$

여기서 C 는 벌점 계수로 잘못 분류된 샘플에 벌점을 가한다. 위 수식을 Lagrangian을 활용해 듀얼 문제로 전환할 수 있다.

$$\min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) - \sum_{i=1}^N \alpha_i$$

$$s.t. \sum_{i=1}^N y_i \alpha_i = 0, a \leq \alpha_i \leq C$$

비선형 문제를 풀기 위해 커널을 도입해 $(x_i \cdot x_j)$ 을 $K(x_i, x_j)$ 로 대체하면 다음과 같다.

$$\min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j K(x_i \cdot x_j) - \sum_{i=1}^N \alpha_i$$

$$s.t. \sum_{i=1}^N y_i \alpha_i = 0, a \leq \alpha_i \leq C$$

커널 $K(x_i, x_j)$ 은 설계자가 정하며 선형 커널과 RBF 커널이 자주 사용된다.

IV. 실험

4.1 데이터 세트

본 연구에서는 증권사의 실제 데이터를 활용해 모델을 학습하였다. 특히 코로나 시대를 맞이하여 비대면 고객을 대상으로 이탈 예측 분석 및 모델 학습을 진행하였다. 샘플링을 통해 고객 기

준 약 15,000 건의 데이터 세트를 확보하였고 인구 통계학 정보, 자산 및 거래 정보 등 총 70 여개의 변수를 활용하였다.

고객 관리 입장에서 자산의 규모에 따라 이탈 위험도가 달라진다. 고액 자산가일수록 이탈 시 손실이 크기 때문이다. 비대면 고객을 관리하는 담당 부서의 고객 관리체계를 반영하여 자산규모 별로 4단계로 구분하였다. 증권사 대외비 정보이기 때문에, 그룹별 고객 분포 및 고객수는 생략한다.

<표 1> 자산 규모별 그룹

그룹	자산 규모
그룹 1	자산 100만원 이하
그룹 2	자산 100만원 ~1,000만원
그룹 3	자산 1,000만원 ~3,000만원
그룹 4	자산 3,000만원 이상

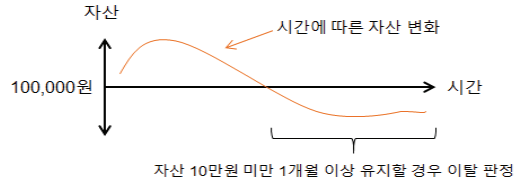
전체 데이터 세트에서 이탈 고객과 정상 고객을 기준으로 계층적 샘플링을 통해 약 2,000명의 고객을 테스트 세트로 샘플링 하였다.

4.2 이탈 기준

대면 고객과 비대면 고객은 자산 운용 규모와 거래 빈도가 다르다. 대면의 경우 자산 100만 원 미만으로 특정 기간 유지되면 이탈로 판정하는 반면, 비대면의 경우 자산 10만 원 미만으로 특정 기간 유지 시 이탈로 판정된다. 우리는 비대면 고객을 대상으로 연구를 진행하였으므로 10만 원 미만 1개월 이상 유지 시 이탈로 판단하였다.

<표 2> 정상 및 이탈 고객 분류 기준

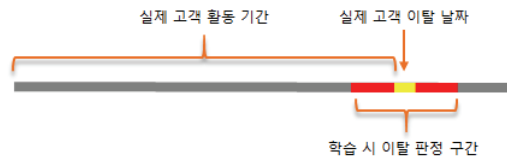
구분	정의
정상고객	자산 10만원 이상 유지 고객
이탈고객	자산 10만원 미만 1달 이상 유지 고객



<그림 2> 자산 변화에 따른 이탈 판정

4.3 정상 고객 / 이탈 고객 Labeling

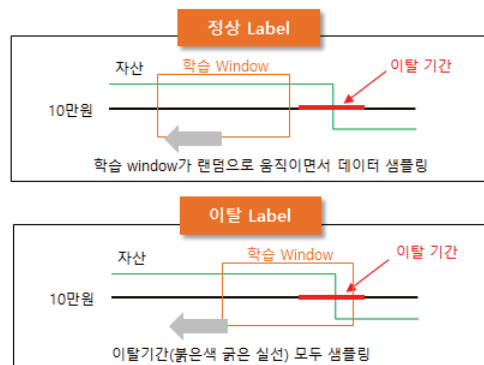
고객이 이탈할 날짜를 예측한다는 것은 매우 어려운 문제이다. 따라서 진행된 연구에서는 실제 이탈 일을 기준으로 ± 3 일 만큼 이탈 레이블을 표기하였다. 예를 들어, 3월 5일이 실제 이탈 일이라면 이 고객의 이탈 레이블은 3월 2일~8일 까지가 된다.



<그림 3> 실제 이탈 및 이탈 판정구간

4.4 Data Augmentation / 불균형

이탈 예측은 정상 고객에 비해 이탈 고객이 매우 적은 문제가 있다. 따라서, Data Augmentation



<그림 4> 정상 및 이탈 레이블 별 샘플링

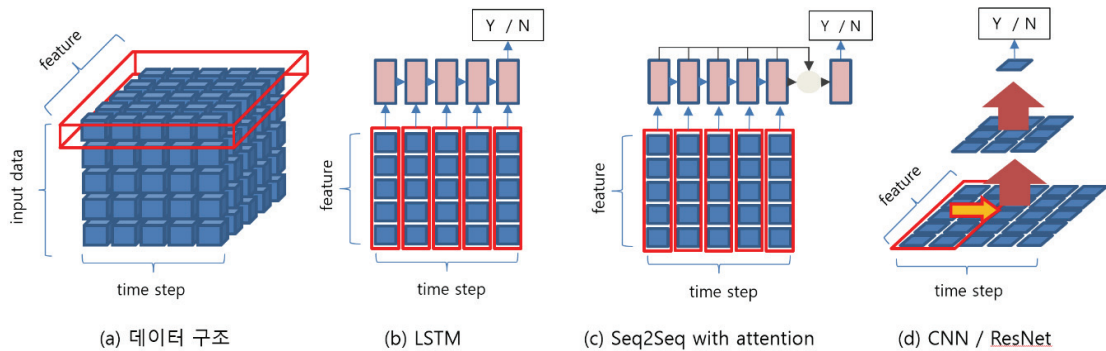
이 필요한데, 먼저 데이터 학습 Window를 따라 샘플링을 진행하였다. 학습 Window는 입력 데이터의 기간으로 본 연구에서는 35 일을 기준으로 삼았다. 그림 4처럼 이탈 구간의 데이터는 모두 샘플링 하였으며, 정상 구간은 랜덤 샘플링을 진행하였다. 이때, 정상 구간의 데이터를 2배 이상 더 랜덤 샘플링 하였다.

또한 SMOTEEN 기법을 이용해 데이터 클래스간 불균형 문제를 해결하였다. SMOTEEN은 SMOTE 기법(소수 클래스에서 임의의 데이터를 샘플링 한 후 이웃과의 차이를 계산하고 이 값에 난수를 곱하여 해당 위치에 새로운 데이터를 생성)과 ENN 기법(다수 클래스의 각 데이터 주변 k 개의 데이터가 모두 다수 클래스에 속하지 않

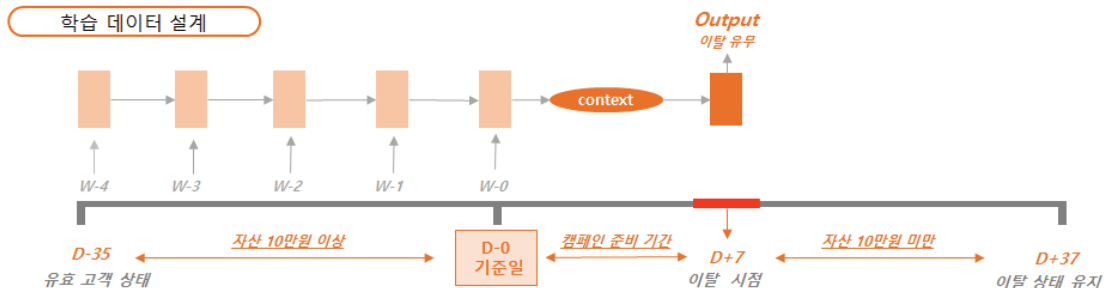
으면 삭제)을 조합한 기법이다.

4.5 입력 데이터 형식

사용된 신경망 계열 알고리즘과 일반 머신러닝 알고리즘의 입력 데이터 형식이 다르다. 신경망 계열의 알고리즘은 일별 데이터를 활용하였으면 일반 머신러닝 알고리즘은 주간 별 통계량을 활용하였다. 또한 이탈 예측을 한 후 이탈 방지를 위한 캠페인 준비 기간을 위해 입력 데이터 기준 7일 이후를 예측하였다. 그림 5는 입력 데이터와 모델 별 작동원리를 도식화하였으며, 그림 6은 LSTM을 예로 입력 데이터 형식을 나타냈다.



〈그림 5〉 데이터 구조 및 모델별 작동 원리 시각화 (a) 입력데이터 구조로 붉은 직육면체가 하나의 고객을 나타낸다. (b), (c) LSTM과 Seq2Seq with attention으로 (a)의 붉은 직육면체가 배치 형태로 입력된다. (d) CNN과 ResNet으로 convolution단계에서 필터가 time step을 이동하면서 데이터가 입력된다.



〈그림 6〉 LSTM을 예로 든 데이터 입력 및 결과 출력 과정 그림상에는 주 별 데이터가 입력되는 것처럼 보이나 실제로는 일별 데이터가 사용되었으며, D-0를 기준으로 D-35까지의 데이터가 입력으로 사용되었다. 또한 D-0를 기준으로 D+7을 예측일로 설정해 실무자의 마케팅 대응시간을 확보하였다.

4.6 모델별 아키텍처 및 성능

표 3은 모델 성능 지표로 본 연구에서는 정확도(Accuracy)인 $\left(\frac{TP+TN}{TP+FP+TN+FN}\right)$ 를 기준으로 하였으며, 각 그룹별 인원수로 가중 평균하였다.

〈표 3〉 모델 성능 지표

	정상	이탈
정상	TP (True Positive)	FP (False Positive)
이탈	FN (False Negative)	TN (True Negative)

TP: 실제 정상을 정상으로 예측, FP: 실제 이탈을 정상으로 예측, FN: 실제 정상을 이탈이라고 예측, TN: 실제 이탈을 이탈로 예측

LSTM의 경우 5 layers (512x256x128x32x2)로 설계하였고 Seq2Seq 모델은 4 layers (256x128x32x2)로 설계하였다. CNN은 (conv → conv → pooling)을 3 계층 쌓았고 이후에 완전 연결 계층 5 layers (2024x512x128x64x2)를 쌓았다. ResNet은 (ResNet Layer → pooling)을 4 계층 쌓았고 이후에 완전 연결 계층 3 layers (512x128x2) 개 쌓았다. 모델의 구조가 간단한 이유는 실무자가 향후 모델 운영에 있어 업무 적용 및 활용에 용이하며 파라미터 튜닝에 어려움을 최소화하기 위함이다. 또한, 추후 모델 재개발에 필요한 인력 및 자원을 최소화하고 업무 효율 극대화를 위해 Shallow한 모델 구조를 선택하였다. 그림 7은 모형 개발

프로세스로 복수의 모델을 비교 학습하여 우수한 모델을 선택하였다.

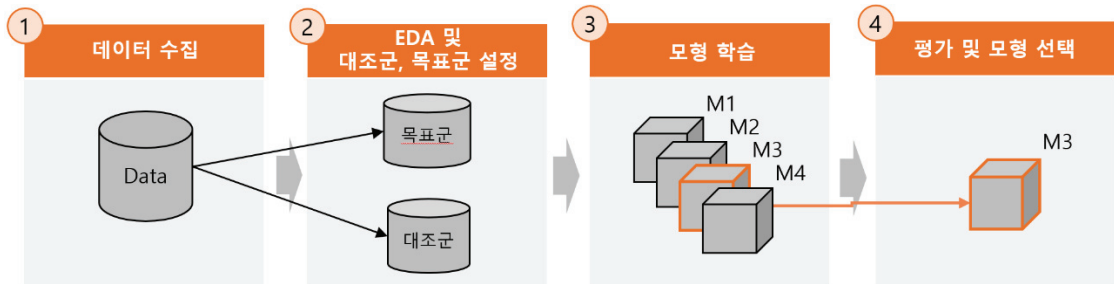
아래 표 4에서 나타낸 것처럼 성능 지표 상으로는 CNN이 가장 좋은 성능을 보였으며 다음으로는 Seq2Seq with Attention이 높았다. 다른 모델은 성능이 유사하게 나타났다. 이러한 이유는 모델을 가볍게 설계하였기 때문인데, ResNet처럼 깊게 쌓아야 좋은 성능이 나오는 모델들이 제 성능을 발휘하지 못했을 가능성이 크다. 모델 성능은 실제 증권사의 고객 정보가 포함되어 있고, 본 연구는 모델 성능보다 모델의 아웃풋에 SHAP Value를 결합한 이탈 원인 추론이 주 목적이므로 정확도(Accuracy)만 제시한다.

〈표 4〉 모델 별 성능

	모델명	정확도(%)
1	LSTM	81.59
2	Seq2Seq with attention	82.23
3	CNN	86.54
4	ResNet	81.83
5	Random Forest	81.23
6	SVM	80.43

4.7 이탈 추론 결과

모델이 예측한 금융상품 거래 고객의 이탈 확률에 원인을 추론하기 위해 SHAP Value를 통한



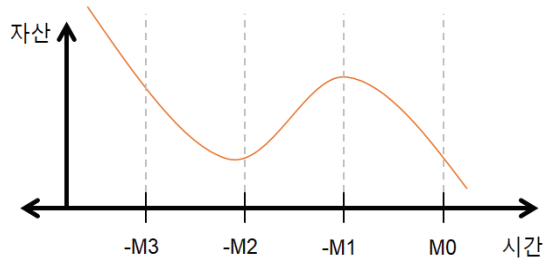
〈그림 7〉 모형 개발 프로세스 1) 데이터 수집 단계, 목적에 적합한 데이터를 수집 및 추출, 2) 데이터에 대한 EDA를 통해 목표군(이탈 고객)과 대조군(비 이탈 고객)을 구분하고 유의미한 변수를 추출, 3) 모형학습 단계로 여러 모델을 학습하여 성능 비교, 4) 성능이 뛰어난 모델을 선택 후 테스트 및 운영

변수 중요도와 더불어 표 5에 정의된 개별 고객의 이탈 직전 Fact 데이터 분석을 하였다. SHAP Value는 특정 변수가 이탈 확률에 어느 정도 영향을 주었는지 확인 가능하지만, 이 사실 하나로 이탈 원인까지 추론하는 것은 한계가 있다. 가령, 한 고객의 ‘주식거래 횟수’가 가장 중요한 변수로 나타났다면, 그 횟수가 빈번하여 수익실현을 하고 이탈로 예상되는지, 거래 횟수가 적어 향후 거래 중단을 할 것인지 알 수 없다. 즉, SHAP Value는 특정 변수의 성질에 따른 중요도 파악이 어렵다. 따라서, 사전에 가공한 개별 고객의 Fact 데이터를 모델의 결괏값(이탈 확률)과 SHAP Value에 각각 결합하여 고객의 이탈 원인을 추론한다.

각 고객별 Fact 데이터의 평균 수익률은 3개월간 금융상품 거래 합계 평균 수익률이며, 수익 유형은 지난 3개월간 유형을 분석하여 표 6과 같이 15개 체계로 분류하였다. ‘수익실현-일반’ 유형은 3개월 동안 매월 수익을 달성하여 수익 실현을 하고 이탈로 예측된 고객이며, ‘수익실현-CUP’ 유형은 처음 1개월(-M3에서 -M2) 동안 손

〈표 5〉 Fact 데이터 항목 및 세부 정의

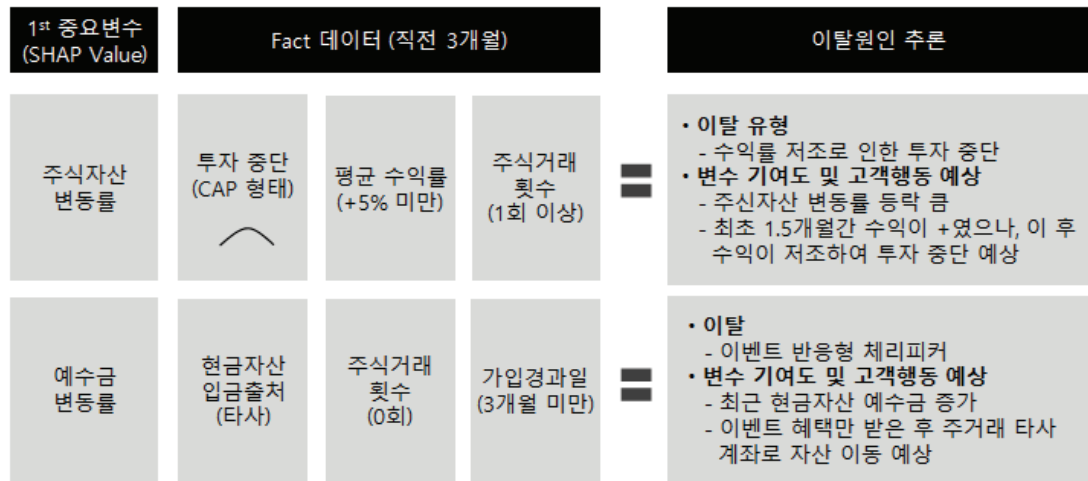
Fact 데이터 항목	설명
수익 유형	직전 3개월간 1개월 단위 수익률을 측정한 수익유형
평균 수익률	이탈예측 시점 기준 직전 3개월 동안평균 수익률
자산 입금 출처	최근 계좌 입금 출처(은행, 증권사 등)
자산 입출금 횟수	직전 3개월 동안 입출금 횟수
금융상품 거래 횟수	직전 3개월 동안 거래 횟수



〈그림 8〉 투자실패 CAP-N

〈표 6〉 Fact 데이터 항목 중 투자 수익 유형 정의 M0는 예측 시점이며 -M3는 M0기준 3개월 전

수익 구분	이탈 추론	수익유형	수익유형 정의
직전 3개월간 순수 자산 증가, 합계 수익률 5% 이상			
수익 실현		수익 실현-일반	M0 기준 과거 3개월간 매월 수익률 + 달성
		수익 실현-CAP	-M3에서 -M2 까지 수익률 증가 후 감소
		수익 실현-CUP	-M3에서 -M2 또는 -M3에서 -M1 까지 수익률 감소 후 증가
		수익 실현-CAP-N	-M3에서 -M2 까지 수익률 감소, -M2에서 -M1 까지 증가, -M1에서 M0 까지 감소
		수익 실현-CUP-N	-M3에서 -M2 까지 수익률 증가, -M2에서 -M1 까지 감소, -M1에서 M0 까지 증가
직전 3개월간 순수 자산 감소, 합계 수익률 -5% 이하			
투자 실패		투자 실패-일반	M0 기준 과거 3개월간 매월 수익률 -달성
		투자 실패-CAP	-M3에서 -M2 까지 수익률 증가 후 감소
		투자 실패-CUP	-M3에서 -M2 또는 -M3에서 -M1 까지 수익률 감소 후 증가
		투자 실패-CAP-N	-M3에서 -M2 까지 수익률 감소, -M2에서 -M1 까지 증가, -M1에서 M0 까지 감소
		투자 실패-CUP-N	-M3에서 -M2 까지 수익률 증가, -M2에서 -M1 까지 감소, -M1에서 M0 까지 증가
직전 3개월간 합계 수익률 ±5%미만			
투자 중단		투자 중단-일반	M0 기준 3개월간 매월 수익률 변동 미미
		투자 중단-감소	M0 기준 3개월간 매월 수익률 소폭 감소
		투자 중단-증가	M0 기준 3개월간 매월 수익률 소폭 증가
		투자 중단-CAP	-M3에서 -M2 또는 -M3에서 -M1 까지 소폭 증가 후 감소
		투자 중단-CUP	-M3에서 -M2 또는 -M3에서 -M1 까지 소폭 감소 후 증가



〈그림 9〉 SHAP Value 기반 이탈 원인 추론 과정 예시

실을 보다가 이후 2개월(-M2에서 M0) 간 수익을 달성하여 합계 평균 수익률이 양수인 이탈 예정 고객이다. ‘투자실패-CAP’ 유형은 처음 1개월 수익을 달성했으나, 이후 2개월 동안 손실이 발생하여 합계 평균 수익률이 음수로 이탈 예정인 고객 유형이다. 현금자산 입금 출처는 단기 이벤트로 유입되어 체리피커 형 이탈 여부를 구분하기 위해 사용하였다. 가입 경과 일이 짧고 타 증권사로부터 유입된 고객은 이벤트 반응형 즉 체리피커 형 고객일 확률이 높다. 입출금 횟수와 금융상품 거래 횟수는 투자 활동성을 나타내는 지표로, 두 항목의 거래 횟수가 특정 기간 동안 없을 시 자산을 방치하여 자연 이탈 예정인 고객 유형이다. 그림 8은 ‘투자 실패 CAP-N’의 예시로 -M2 시점에-M3 대비 자산이 감소하다가 다시 -M1 시점에 회복세로 돌아선다. 하지만 M0에 결국 -M3 대비 자산이 감소한 형태를 띈다.

그림 9 은 개별 고객의 SHAP Value와 Fact 데이터를 매핑한 후 Rule에 근거한 대표적인 이탈 원인 추론 과정이다. 첫 번째 Case는 최근 주식자산 변동률이 이탈 원인에 가장 크게 기여한 것으로 나타났으나, 그 성질을 알 수 없으므로 Fact 데이터를 매핑 한 결과, 3개월 동안 최초 1개월간 수익을 보았으나 최근 투자 결과가 안 좋아 이탈

할 것으로 추론이 가능해진다. 두 번째 Case는 예수금 변동률이 기여도가 가장 큰 것으로 나타났고, Fact 데이터 매핑을 통해 최근 타 증권사로부터 현금 자산이 유입되었으며, 주식거래는 0 회로 단순 이벤트 혜택을 위해 유입된 고객으로 혜택 수령 후 이탈할 것으로 예상할 수 있다. 당 연구에서는 이러한 이탈 원인을 총 22개 분류 체계로 개발하였으며, 마케팅 담당자는 이탈 원인 유형에 따라 맞춤형 고객 Retention 캠페인을 기획 및 실행할 수 있도록 최종 설계하였다.

V. 결론

코로나 국면의 장기화로 인해 전 산업에 걸쳐 비대면 고객 마케팅이 점차 중요해지고 있다. 특히 금융 산업은 대면 중심에서 비대면 중심으로 서비스 제공 형태가 변화하고 있기에, 다양한 비대면 채널 관리와 고객 특성별 마케팅 수행의 중요성이 매우 높아지고 있다.

금번 연구는 금융업권 중에서도 주식거래 고객의 특성상 고객의 이탈이 매우 빈번하게 발생하는 실제 증권사의 비대면 채널 고객 행태정보를 기반으로, 다양한 머신러닝 알고리즘을 적용한 실험 연구를 통해 이탈 예측 모델을 개발하였

다는데 학문적 의의가 있다. 또한, 기존 머신러닝 모델이 지닌 설명 불가능 문제를 해결하기 위해 게임 이론을 기반으로 고안된 SHAP Value 기법을 적용하고, 마케팅 전문가의 도메인 지식을 반영한 22 개의 고객 이탈 원인을 실증 및 도출하였다는 점에서 학문적 기여가 크다고 할 수 있다.

그동안 머신러닝 알고리즘의 운영을 통해 고객 관리를 수행해 온 마케팅 실무자는 알고리즘의 결과 해석에 있어 많은 어려움을 겪어 왔다. 특히 머신러닝 기반 고객 이탈 예측 모델을 통해 이탈 확률이 높은 고객을 대상으로 이탈 방지 마케팅을 수행할 때, 고객의 이탈 원인을 확인할 길이 없으면 마케팅 수행 방안 마련에 있어 어려움이 컸다. 즉, 이탈 원인을 알아야 고객 맞춤형 마케팅을 수행할 수 있기에 당 연구에서는 이탈에 영향을 미치는 주요 변수를 찾는 것에 머물지 않고, 나아가 이탈 원인을 밝혀 현업 실무자들에게 직관적인 정보를 제공하였다는데 학문적 시사점이 있다.

이러한 노력을 통해 얻어진 연구 결과를 토대로, 향후 마케팅 담당자의 실제 고객 마케팅 수행에 있어 원인 추론이 가능한 이탈 예측 결과값을 사용하고, 각 고객별 타겟 마케팅 전략을 수립하는 데 있어 이를 종합적 판단 근거로 활용할 수 있을 것이다.

참 고 문 헌

- [1] 김경태, & 이지형. (2018). 딥 러닝과 Boosted Decision Tree 를 활용한 고객 이탈 예측 모델. 한국지능시스템학회 논문지, 28(1), 7-12.
- [2] 김은미, & 박지영. (2019). 온라인 P2P 환경에서 딥러닝을 적용한 다분류 기반 개인신용등급 예측 모델. 인터넷전자상거래연구, 19(4), 43-57.
- [3] 김승수, & 김종우. (2018). 비정형 정보와 CNN 기법을 활용한 이진 분류 모델의 고객 행태 예측: 전자상거래 사례를 중심으로. 지능정보연구, 24(2), 221-241.
- [4] 서민교. (2013). 국내 금융기관의 빅 데이터 (Big Data) 활용 사례에 관한 연구. 전자무역연구, 11(4), 115-134.
- [5] 이동규, & 신민수. (2018). 카드 산업에서의 고객 휴면 예측. 한국 IT 서비스학회 학술대회 논문집, 2018, 404-407.
- [6] 이세희, & 이지형. (2016). RNN 을 이용한 고객 이탈 예측 및 분석. 한국컴퓨터정보학회 학술발표논문집, 24(2), 45-48.
- [7] 장민석, & 김형중. (2018). 빅데이터를 활용한 은행권 고객 세분화 기법 연구. 한국디지털콘텐츠학회 논문지, 19(1), 85-91.
- [8] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138-52160.
- [9] Ahn, Y., Kim, D., & Lee, D. J. (2019). Customer attrition analysis in the securities industry: a large-scale field study in Korea. International Journal of Bank Marketing.
- [10] Binder, A., Montavon, G., Lapuschkin, S., Müller, K. R., & Samek, W. (2016, September). Layer-wise relevance propagation for neural networks with local renormalization layers. In International Conference on Artificial Neural Networks (pp. 63-71). Springer, Cham
- [11] Gunning, D. (2017). Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web, 2(2).
- [12] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature machine intelligence, 2(1), 2522-5839.
- [13] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In

- Advances in neural information processing systems (pp. 4765-4774).
- [14] Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73, 1-15.
- [15] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).
- [16] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- [17] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems* (pp. 3104-3112).
- [18] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- [19] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.
- [20] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [21] Tsai, C. F., & Lu, Y. H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 36(10), 12547-12553.
- [22] He, B., Shi, Y., Wan, Q., & Zhao, X. (2014). Prediction of customer attrition of commercial banks based on SVM model. *Procedia Computer Science*, 31, 423-430.
- [23] Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & 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.
- [24] Li, Y., & Xia, G. E. (2010, November). The explanation of support vector machine in customer churn prediction. In *2010 International Conference on E-Product E-Service and E-Entertainment* (pp. 1-4). IEEE.
- [25] Pamina, J., Raja, B., SathyaBama, S., Sruthi, M. S., & VJ, A. (2019). An effective classifier for predicting churn in telecommunication. *Jour of Adv Research in Dynamical & Control Systems*, 11.
- [26] DUMITRACHE, A., NASTU, A. A., & STANCU, S. Churn Prediction in Telecommunication Industry: Model Interpretability.

저자 소개



나 광택(Kwangtek Na)

- 2013년 : 인하대학교 토목공학 (학사)
- 2017년: 인하대학교 컴퓨터공학(석사)
- 2017년~ 현재: 인하대학교 전기컴퓨터공학(박사과정)
- 2018년~ 현재: 데이터애널리틱스랩 (한화 그룹) 선임연구원
- 관심분야: 머신러닝, 딥러닝, XAI, 베이지안 추론, 추천시스템, 시계열



이 진 영(Jinyoung Lee)

- 2008년: University of Washington, Seattle 정보공학 (학사)
- 2020년 ~ 현재: 연세대학교 인공지능학 (석사과정)
- 2019년 ~ 현재: 데이터애널리틱스랩 (한화 그룹) 책임연구원

·관심분야: 머신러닝, 딥러닝, XAI, CRM



김 은 찬(Eunchan Kim)

- 2012년: University of Minnesota, Twin Cities 경제학 (학사)
- 2017년: 서울대학교 경영학 (경영정보) (석사)
- 2019년 ~ 현재: 서울대학교 지능정보융합학 (박사과정)

·2018년 ~ 현재: 데이터애널리틱스랩 (한화 그룹) 책임연구원

·관심분야: 머신러닝, 예측 모델링, XAI, 인지과학, HCI



이 효 찬(Hyochan Lee)

- 1997년: 중앙대학교 무역학과 (학사)
- 1999년: 중앙대학교 일반대학원 국제경제학 (경영학 석사)
- 2011년: 중앙대학교 일반대학원 국제경제학 (경영학 박사)

·2018년 ~ 현재: 데이터애널리틱스랩 (한화 그룹) 수석연구위원

· 관심분야: 디지털라이제이션, 머신러닝, 시계열 분석, UI/UX