Journal of Next-generation Convergence Information Services Technology Vol.14, No.6, December (2025), pp.743-752
ISSN 2384-101X(p) 2672-1163(e) *http://dx.doi.org/10.29056/jncist.2025.12.01*

설명가능한 인공지능을 활용한 개인 신용카드 채무불이행 예측 연구

A Study on Credit Default Prediction using eXplainable Artificial Intelligence

이우식[1]

Woosik Lee[1]

요 약

본 연구는 개인 신용카드 채무불이행 예측을 위해 로지스틱 회귀 모형과 XGBoost 모형의 성능을 실증 비교하고, 설명가능 인공지능(eXplainable Artificial Intelligence, XAI) 기법을 통해 모형의 해석 가능성을 강화하는 것을 목적으로 하였다. 이를 위해 캐글의 아메리칸 익스프레스 데이터를 활용하여 계수 기반, gain 기반, SHAP(SHapley Additive exPlanations) 기반 방법으로 변수 중요도를 비교 분석한 결과는 다음과 같다. 첫째, 두 모형 모두 우수한 예측 성능을 보였으나, XGBoost는 혼동행렬 분석에서 False Negative(제2종 오류)가 발생하지 않아 실제 채무불이행 고객을 완벽히 식별하였다. 이는 금융기관의 손실을 방지하는 리스크 관리 측면에서 중요한 실무적 의의를 지닌다. 둘째, SHAP 분석을 통해 신용점수, 과거 연체 이력, 신용한도 사용률이 채무불이행 예측에 핵심 변수임을 확인하였다. SHAP는 샤플리값을 기반으로 변수 간 상호작용을 고려하면서 개별 예측에 대한 변수별 기여도를 명확히 분해하여, 전통적 방법론인 단순 계수 또는 gain 기반 분석 대비 해석 가능성을 크게 향상시켰다. 이는 금융 실무에서 요구되는 모형의 투명성과 신뢰성을 확보하는 데 중요한 근거를 제공한다.

핵심어 : 비즈니스 애널리틱스, 설명가능한 인공지능, 계량금융, 금융 AI, 신용 위험 관리

Abstract

This study aimed to empirically compare the performance of the logistic regression model and the XGBoost model for predicting individual credit card default, and to enhance the interpretability of the models through eXplainable Artificial Intelligence (XAI). Using American Express data from Kaggle, I analyzed feature importance through coefficient-based, gain-based, and SHAP (SHapley Additive exPlanations) - based methods with the following results. First, while both models demonstrated excellent predictive performance, XGBoost achieved zero False Negatives (Type II error) in the confusion matrix analysis, thereby perfectly identifying all actual default customers. This has important practical implications in risk management as it helps prevent financial losses for financial institutions. Second, the SHAP analysis confirmed that credit score, previous defaults, and credit limit usage are key variables in predicting default. SHAP, based on Shapley values, clearly decomposes the contribution of each variable to individual predictions while considering interactions among variables, leading to significantly improved interpretability

1 Gyeongsang National University, Jinju, Korea [Professor]
e-mail: woosiklee@gnu.ac.kr

Received(October 2, 2025), Review Result(1st: October 23, 2025), Accepted(December 12, 2025), Published(December 31, 2025)

compared with traditional methodologies such as coefficient-based and gain-based analysis. This provides important grounds for securing the transparency and credibility of models required in financial practice.

Keyword : Business Analytics, eXplainable Artificial Intelligence, Quantitative Finance, Financial AI, Financial Risk Management

1. 서론

신용 위험 관리는 금융기관의 핵심 업무로, 대출 위험을 최소화하고 건전한 자산 운용을 보장하는 데 필수적이다 [1]. 이는 금융기관의 수익성과 직결되는 대출자산의 건전성을 확보하는 과정으로, 연체와 같은 채무불이행을 사전에 예방하기 위해 다양한 평가와 관리 체계가 활용된다. 이러한 맥락에서 채무불이행 예측은 신용위험 관리의 중요한 영역으로 자리 잡고 있으며, 특히 신용카드 산업에서는 소비자 대출 부문 위험 관리의 중심축을 이루고 있다.

금융기관은 대출 심사 과정에서 상환 가능성을 예측하여 대출 실행 여부를 결정하며 [2], 이를 통해 부실 발생 가능성을 최소화하고자 한다. 개인 신용대출의 부도 가능성을 정확히 예측하는 능력은 전통적으로 금융기관의 경쟁력과 안정성을 좌우하는 핵심 역량이었으며, 최근 연체율 상승과 같은 금융환경 변화 속에서 그 중요성이 더욱 부각되고 있다. 그러나 개인 신용카드 채무불이행을 정밀하게 예측하는 일은 쉽지 않은 과제이다. 대출 조건, 거시경제 상황, 개인의 행동 양식 등 시간에 따라 변동하는 요소들이 복합적으로 작용하기 때문이다.

기존의 신용 위험 평가는 주로 통계적 방법에 의존해왔으나, 최근에는 기계학습과 인공지능의 발전으로 데이터 기반의 신용 위험 관리가 주목받고 있다. Yang과 Zhang [3]은 UCI 데이터를 활용하여 신용 카드 채무불이행 예측에서 로지스틱 회귀, 신경망, 서포트 벡터 머신, XGBoost, LightGBM등을 비교·분석하였다. Chen과 Zhang [4]은 k-mean SMOTE와 BP 신경망을 결합하여, 신용카드 채무불이행 예측에서 서포트 벡터 머신, 로지스틱 회귀, 랜덤포레스트 모델보다 높은 예측 정확도를 보였다. Gao 외 2명의 연구 [5]는 XGBoost와 LSTM 모델을 결합해 신용카드 거래 흐름 데이터가 신용카드 연체를 효과적으로 예측할 수 있음을 보여주었다. 김창효와 이군희의 연구 [6]는 트랜스포머 모형을 신용대출 부도 예측 문제에 적용하였으며, XGBoost와 유사한 수준의 예측 성능을 보였다.

그러나 대부분의 기존 연구는 기계학습 모형의 성능 비교에 치중하고 있으며, 높은 예측력을 보이더라도 모형의 해석 가능성이 부족하다는 구조적 한계를 지닌다. 이로 인해 예측 결과의 신뢰성과 실무 적용성이 제한되는 문제가 존재한다.

이에 본 논문은 로지스틱스 회귀 모형과 XGBoost 모형 기반 예측 모형을 구축하고, XAI를 활용하여 각 입력 변수가 예측에 기여하는 영향을 정량적으로 분석한다. 이를 통해 개인 신용카드 채무불이행 리스크를 설명 가능하도록 규명하고, 선행연구와의 차별성을 확보하고자 한다.

본 논문은 다음과 같이 구성된다. 제1장에서는 연구의 배경과 필요성을 설명하고, 제2장에서는 XGBoost 모형과 SHAP 기법에 대한 이론적 배경을 고찰한다. 제3장에서는 연구 설계와 실증 분석 과정을 제시하고 결과를 보고한다. 마지막으로 제4장에서는 주요 결과를 바탕으로 연구의 의의와 시사점을 논의하며 결론을 제시한다.

2. 이론적 배경

2.1 XGBoost

XGBoost(eXtreme Gradient Boosting) [7]는 앙상블 기법을 활용한 기계학습 알고리즘으로, 회귀 및 분류 작업에 모두 적용 가능하며 다양한 산업에서 뛰어난 성과를 입증하고 있다.

XGBoost는 손실 함수의 기울기를 따라 약한 트리를 점진적으로 생성하고 통합하는 방식으로 작동한다. 모형의 예측함수는 아래와 같이 표현된다 [8].

$$\hat{Y}_i^T = \sum_{k=1}^{T} f_k(x_i) = \hat{y}_i^{T-1} + f_T(x_i)$$

where

$\hat{Y}_i^T$: 첫 번째부터 T번째 트리까지의 예측값을 합산한 최종 결과

$f_k(\bullet)$: k번째 회귀 트리

x_i : 입력 특징 벡터

$\hat{y}_i^{T-1}$: 이전 단계까지의 누적 예측값

$f_T(x_i)$: 새롭게 학습된 트리의 예측값

XGBoost는 예측 정확도와 모형 복잡도를 동시에 최적화하는 목표함수를 최소화한다 [9].

$$L(\phi) = \sum_i L(y_i, \hat{y_i}) + \sum_k \Omega(f_k)$$

where

$L(\phi)$: 전체 목적 함수

$\sum_i L(y_i, \hat{y_i})$: 실제값 y_i와 예측값 $\hat{y_i}$의 손실 함수

$\Omega(f_k)$: k번째 트리 f_k에 대한 복잡도 페널티

ϕ : 모형의 모든 파라미터

복잡도 페널티 $\Omega(f_T)$는 트리의 복잡성을 조절하여 과적합을 방지하고, 모형의 일반화 성능을 개선한다 [9].

2.2 SHAP

SHapley Additive exPlanations (SHAP) [10]은 게임이론의 샤플리 값을 기반으로 한 설명 가능 인공지능 기법이다. SHAP은 모든 특성 조합을 분석하여 예측 결과 $\hat{f}(x)$에 대한 각 특성이 미치는 영향을 수치적으로 계산한다.

$$E[f_{S\cup i}(x_{S\cup i})] - E[f_S(x_S)]$$

$$= E[f(x_1,x_2,\cdots,x_i,\cdots,x_{n-1},x_n)] - E[f(x_1,x_2,\cdots,X_i,\cdots,x_{n-1},x_n)]$$

where
S : 특성의 부분집합
i : 분석대상 특성의 인덱스
$f_S(x_S)$: 부분집합 S의 특성만으로 예측한 모델 출력값
$f_{S\cup i}(x_{S\cup i})$: 부분집합 S와 특성 i를 포함하여 예측한 모델 출력값
$E[\bullet]$: 기대값

위 공식에서 S는 전체 특성 집합, x는 각 특성의 실제 관측값 그리고 X는 해당 특성이 가질 수 있는 모든 값을 나타낸다. 개별 특성의 영향력 총합은 모형의 최종 예측 결과 $\hat{f}(x)$로 아래 식과 같이 표현된다 [9].

$$\hat{f}(x) = \phi_0 + \sum_{i=1}^{n}\phi_i$$

where
$\hat{f}(x)$: 모델의 예측 결과
ϕ_0 : 기준값
ϕ_i : 특성 i의 SHAP값
n : 전체 특성의 개수

위 식에서 예측 결과 $\hat{f}(x)$는 각 특성의 기여도 ϕ_i와 모든 기여도 계산의 기준이 되는 기준값

ϕ_0의 합으로 이루어진다. 이러한 방식을 통해 SHAP은 복잡한 모델의 예측 과정을 개별 특성의 기여도로 분해하여 입력 변수가 예측 결과에 미치는 영향을 정량적으로 밝힌다 [9].

3. 실증분석

3.1 자료의 구성

본 연구에서 사용된 데이터셋은 캐글(Kaggle)에서 제공하는 아메리칸 익스프레스(American Express)의 자료로, 개인의 신용평가와 관련된 다양한 특성을 포함한다. 구체적으로 데이터는 식별 변수, 인구통계학적 변수, 자산 및 소유 변수, 고용 관련 변수, 신용 관련 변수, 신용 이력 변수 그리고 타겟 변수를 포함하고 있다.

데이터 품질의 확보와 분석 적합성을 위해 다음과 같은 전처리를 수행하였다. 먼저, 7개의 변수(owns_car, no_of_children, no_of_days_employed, total_family_members, migrant_worker, yearly_debt_payments, credit_score)에서 결측값이 발생한 관측치를 제거하였으며, 이는 전체 데이터의 약 4.4%에 해당한다. 또한 성별 변수의 이상값(XNA) 1개를 제거하고, 예측에 불필요한 식별 변수(customer_id, name)도 제외하였다. 이 과정을 통해 최종적으로 43,508명의 고객 데이터와 17개 변수를 확보하였다.

범주형 변수에는 라벨 인코딩(Label Encoding)을 적용하였다. 성별, 자동차와 주택 소유 여부, 이주 노동자 여부와 같은 이진 변수는 0과 1로 인코딩하였으며, 직업 유형은 19개 범주를 0부터 18까지의 연속된 정수로 매핑하였다.

타겟 변수 분포 확인 결과, 정상 고객 39,977명(91.88%)과 채무불이행 고객 3,531명(8.12%)으로 클래스 불균형이 존재하였다. 이를 보완하기 위해 언더샘플링(Random Under Sampling)을 적용하여 정상 고객 수를 소수 클래스 수준으로 조정하였고, 결과적으로 총 7,062명(각 클래스 3,531명)의 균형 데이터를 분석에 활용하였다.

3.2 모형의 추정 및 분석

본 연구에서는 개인 신용카드 채무불이행 예측성능을 비교하기 위해 로지스틱스 회귀 모형과 XGBoost 모형을 구현하였다. 전체 데이터의 80%를 훈련용으로, 20%를 테스트용으로 나누었으며, 계층적 추출을 통해 클래스 비율을 일관되게 유지하였다. 두 모형 모두에 대해 그리드 탐색과 3겹 교차검증을 활용하여 하이퍼파라미터 최적화를 수행하였다. 그리고 성능 평가는 정확도, 정밀도, 재현율, F1 점수, ROC AUC 등 다양한 지표를 활용하였다. 성능 평가 결과는 [표 1]와 [그림 1]에 제시하였다. 로지스틱 회귀 모형의 경우, 최적 파라미터는 규제 강도는 1, 반복 횟수는 1000, 규제

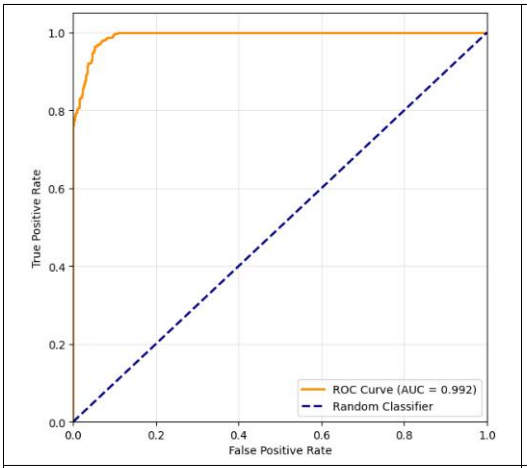
방식은 L1, 최적화 알고리즘은 liblinear로 선정되었다. 테스트 데이터 기준 성능은 정확도, 정밀도, 재현율 및 F1 점수에서 모두 0.95를 기록하였으며, ROC AUC는 0.992로 나타나 우수한 분류 성능을 보였다. XGBoost 모형의 경우, 최적 파라미터는 추정기 개수 100개, 최대 트리 깊이는 3, 학습률은 0.1, 부분표본추출 비율은 1.0, 특성 부분표본추출 비율은 0.8로 도출되었다. 테스트 데이터 기준 성능은 정확도, 정밀도, 재현율 및 F1 점수에서 모두 0.98을 기록하였으며, ROC AUC는 0.995로 나타나 로지스틱 회귀 대비 다소 더 우수한 분류 성능을 기록하였다. 혼돈행렬 분석 결과도 XGBoost 모형은 거짓 음성(False Negative)이 전혀 발생하지 않은 반면, 로지스틱 회귀 모형에서는 32건의 거짓 음성이 발생하여 오분류 건수 측면에서도 XGBoost 모형이 더 우수한 성능을 보였다.

[표 1] 테스트 데이터에 대한 로지스틱 회귀 모형과 XGBoost 모형의 성능 평가

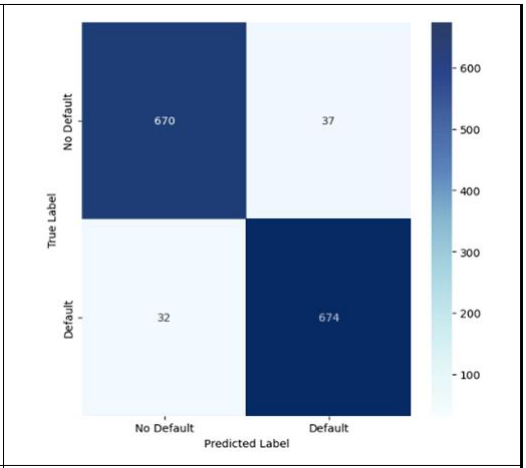
[Table 1] Logistic Regression and XGBoost Model Performance Evaluation on Test Dataset

Logistic Regression	Precision	Recall	F1-Score	Support
No Default	0.95	0.95	0.95	707
Default	0.95	0.95	0.95	706
Accuracy			0.95	1413
Macro Avg.	0.95	0.95	0.95	1413
Weighed Avg.	0.95	0.95	0.95	1413

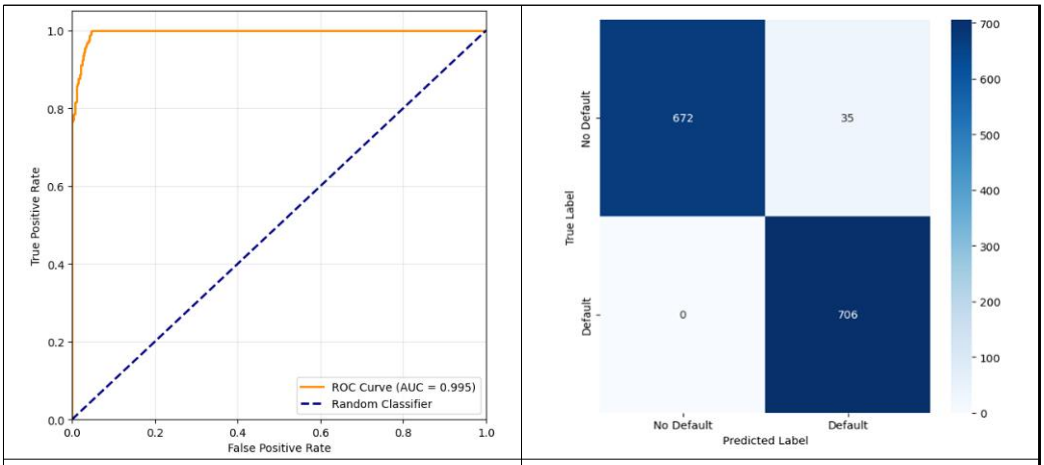
XGBoost	Precision	Recall	F1-Score	Support
No Default	1.00	0.95	0.97	707
Default	0.95	1.00	0.98	706
Accuracy			0.98	1413
Macro Avg.	0.98	0.98	0.98	1413
Weighed Avg.	0.98	0.98	0.98	1413



(a) ROC Curve(Logistic Regression)



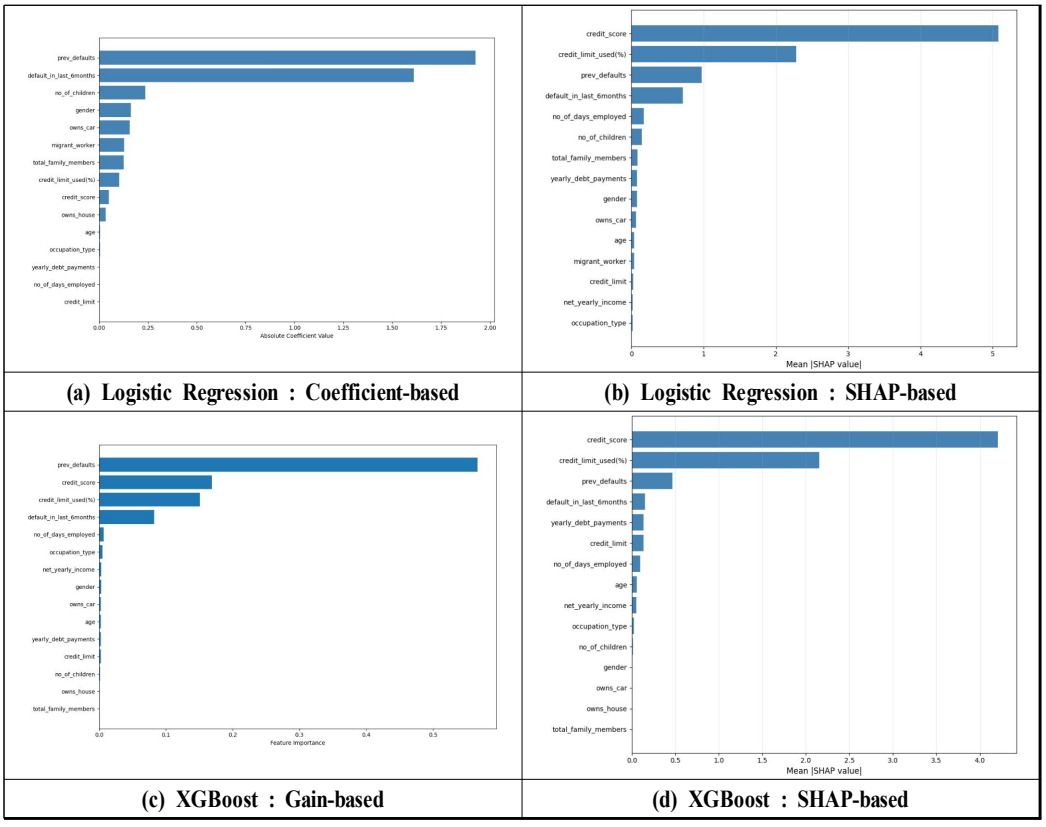
(b) Confusion Matrix(Logistic Regression)



(c) ROC Curve(XGBoost) **(d) Confusion Matrix(XGBoost)**

[그림 1] ROC 곡선과 혼동행렬

[Fig. 1] ROC Curve and Confusion Matrix



(a) Logistic Regression : Coefficient-based **(b) Logistic Regression : SHAP-based**

(c) XGBoost : Gain-based **(d) XGBoost : SHAP-based**

[그림 2] 특성 중요도

[Fig. 2] Feature Importance

더불어 [그림 2]를 통해 로지스틱 회귀 모형과 XGBoost 모형의 변수 중요도를 각각 계수 기반 및 gain 기반 방법으로 도출하고, SHAP 값과도 함께 종합적으로 비교 분석하였다. 로지스틱 회귀 모형에서 계수 기반 분석 [그림 2(a)]은 과거 연체 이력(prev_defaults)을 가장 중요한 변수로 제시한 반면, SHAP 기반 분석 [그림 2(b)]에서는 신용점수(credit_score)가 가장 높은 중요도를 보이며, 신용한도 사용률(credit_limit_used%), 과거 연체 이력(prev_defaults)이 뒤따랐다. XGBoost 모형에서도 유사한 패턴이 관찰되었다. Gain 기반 분석 [그림 2(c)]에서는 과거 연체 이력, 신용점수, 신용한도 사용률 순으로 나타난 반면, SHAP 기반 분석 [그림 2(d)]에서는 신용점수, 신용한도 사용률, 과거 연체 이력 순으로 중요도가 재배열되었다. 분석 방법에 따라 변수 중요도의 순위는 다소 차이가 있었으나, 모든 분석에서 신용 관련 지표가 채무불이행 예측의 핵심 변수로 일관되게 확인되었다. 한편, 인구통계학적 변수, 자산 및 소유 변수, 고용 관련 변수, 신용 이력 변수들도 상대적으로 낮은 중요도를 보였지만, 모형의 예측력 향상에 부분적으로 기여하는 것으로 나타났다.

4. 결론

본 연구는 통계모형과 기계학습 기반 개인 신용카드 채무불이행 예측 모형의 성능을 실증적으로 검증하고, 설명가능한 인공지능(XAI)기법을 통해 모형의 해석 가능성을 확보하고자 하였다. 캐글의 아메리칸 익스프레스 데이터를 활용하여 로지스틱 회귀 모형와 XGBoost 모형의 변수 중요도를 각각 계수 기반 및 gain 기반 방법으로 도출하고, SHAP값과 함께 종합적으로 비교 분석한 결과는 다음과 같다.

첫째, 로지스틱스 회귀 모형과 XGBoost 모형 모두 우수한 분류 예측 성능을 보였다. 특히 XGBoost 모형은 대부분의 성능 지표에서 로지스틱 회귀를 상회하며 전반적으로 더 우수한 분류 성능을 보였다. 이러한 결과는 계수 기반의 선형 분석을 수행하는 로지스틱 회귀와 트리 기반의 비선형 분석을 수행하는 XGBoost 모형이 개인 신용카드 채무불이행 예측 문제에서 모두 유효한 판별력을 지니고 있음을 시사한다. 하지만 혼동행렬 분석에서 두 모형 간 중요한 차이가 발견되었다. XGBoost 모형은 False Negative(제2종 오류)가 발생하지 않아 실제 채무불이행 고객을 모두 정확히 식별한 반면, 로지스틱 회귀 모형에서는 일부 채무불이행 사례가 정상으로 오분류되었다. 금융 리스크 관리 관점에서 False Negative는 금융기관의 직접적인 손실로 이어질 수 있으므로, XGBoost 모형의 이러한 특성은 실무적으로 매우 중요한 강점이라 할 수 있다.

둘째, SHAP 분석을 통해 개인 신용카드 채무불이행 예측의 핵심 결정 요인을 식별하였다. 분석 결과, 신용점수, 과거 연체 이력 그리고 신용한도 사용률이 모형 예측에 가장 큰 영향을 미치는 것으로 확인되었다. SHAP은 게임 이론의 샤플리값을 기반으로 각 변수의 한계 기여도(Marginal Contribution)를 개별 관측치 수준에서 계산한다. 이는 전통적인 계수 기반(로지스틱 회귀) 또는 gain

기반(XGBoost) 중요도와 달리, 변수 간 상호작용 효과를 고려하면서도 각 예측에 대한 변수별 기여도를 명확히 분해할 수 있다는 장점이 있다. 따라서 SHAP은 예측 결과에 대한 해석 가능성을 크게 향상시킨다.

본 연구의 학술적 의의는 다음과 같다. 기존 연구가 주로 모형 간 성능 비교에만 집중한 반면, 본 연구는 XAI기법을 활용하여 통계모형과 기계학습 모형의 의사결정 과정에 대한 해석 가능성을 확보했다는 점에서 학술적 기여도를 갖는다. 더불어 이는 금융 실무에서 요구되는 모형의 투명성과 신뢰성을 확보하는 데 중요한 근거를 제공한다. 즉, 이러한 해석 가능성은 금융기관이 실제 의사결정 과정에서 리스크 요인을 명확히 파악하고, 신용정책 수립이나 조기 경보 체계 구축에 활용할 수 있는 기반을 마련한다는 점에서 학문적·실무적 의의를 지닌다.

본 연구의 한계점과 향후 연구방향은 다음과 같다. 첫째, 본 연구에서 사용된 데이터가 캐글에서 제공하는 특정 시점의 아메리칸 익스프레스 자료에 한정되어 있어, 표본 대표성이 제한적이며 시간에 따른 고객 신용 행동 변화나 장기적 채무불이행 위험을 충분히 반영하지 못한다. 둘째, 본 연구는 신용점수, 연체 이력, 신용한도 사용률 등 정형 데이터를 중심으로 수행되었다. 그러나 최근 고객 상담 데이터 등 비정형 데이터까지 고려하면, 본 연구의 데이터 범위 제약은 한계로 지적될 수 있다.

References

[1] M. J. Ariza-Garzón, J. Arroyo, A. Caparrini, M. J. Segovia-Vargas, “Explainability of a machine learning granting scoring model in peer-to-peer lending”, IEEE Access, vol. 8, March 2020, pp. 64873-64890, doi: 10.1109/ACCESS.2020.2984412.

[2] S. Lee, E. Y. Ryu, “Analysis and practical implications of default rates of peer-to-peer loans and loan products”, Korean Management Consulting Review, vol. 20, no. 1, February 2020, pp. 275-283.

[3] S. I. Yang, H. Zhang, “Comparison of several data mining methods in credit card default prediction”, Intelligent Information Management, vol. 10, September 2018, pp. 115-122, doi: 10.4236/iim.2018.105010.

[4] Y. Wang, Z. Xu, K. Ma, “Credit default prediction with machine learning: A comparative study and interpretability insights”, Proceedings of the 2024 4th International Conference on Communication Technology and Information Technology (ICCTIT), December 27-29, 2024, Guangzhou, China, pp. 1-6, doi: 10.1109/ICCTIT64404.2024.10928657.

[5] J. Gao, W. Sun, X. Sui, “Research on default prediction for credit card users based on XGBoost-LSTM model”, Discrete Dynamics in Nature and Society, vol. 2021, December 2021, pp. 1-13, doi: 10.1155/2021/5080472.

[6] C. H. Kim, G. Lee, “Personal credit loan default prediction model using the transformer methodology based on tabular data”, Korean Management Consulting Review, vol. 24, no. 4, August 2024, pp. 159-170.

[7] K. Ostrowski, K. Birman, “Extensible web services architecture for notification in large-scale systems”,

Proceedings of the 2006 IEEE International Conference on Web Services (ICWS'06), September 18-22, 2006, Chicago, IL, USA, pp. 383-392, doi: 10.1109/ICWS.2006.63.

[8] K. Tissaoui, T. Zaghdoudi, A. Hakimi, O. Ben-Salha, L. B. Amor, “Does uncertainty forecast crude oil volatility before and during the COVID-19 outbreak? Fresh evidence using machine learning models”, Energies, vol. 15, no. 15, August 2022, pp. 5744, doi: 10.3390/en15155744.

[9] W. S. Lee, “A study on KOSPI volatility prediction using explainable artificial intelligence”, Journal of the Korean Society of Industry Convergence, vol. 28, no. 3, June 2025, pp. 747-754, doi: 10.21289/KSIC.2025.28.3.747.

[10] S. M. Lundberg, S. I. Lee, “A unified approach to interpreting model predictions”, Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), December 4-9, 2017, Long Beach, CA, USA, pp. 4768-4777, doi: 10.5555/3295222.3295230.