

판별분석을 통한
신용카드 대출 연체 예측

제출일자: 2024.06.07

지도교수: 김헌주 교수님

이름: 정희광

학번: 21600673

1전공: 경영학

Abstract

본 논문은 신용카드 대출 연체율의 급증에 관해서 신용카드 대출 데이터를 가지고 4가지의 요인들로 분류하여 분석하고, 분류된 요인들로 판별분석을 진행하여 대출받은 돈을 잘 갚는 사람인지 아닌지에 관한 예측을 진행했습니다. 그리고, 진행한 판별분석의 결과인 판별함수를 가지고 새로운 데이터인 평가 데이터에 이 함수를 대입시켜 봄으로써 연체여부를 잘 판별하고 있는지 확인해 보았습니다. 분석을 진행했을 때, 본 연구에는 총 4가지의 요인이 있으며, 대출, 소득, 부채, 연체 변수로 나뉩니다. 또한, 데이터에서 진행했던 판별함수는 $LD1 = -0.2829*RC1 + 0.4202*RC2 + 0.8656*RC3 + 0.0644*RC4$ 이고 예측한 값과 실제 연체여부의 값의 일치 여부를 판별점수로 했을 때, 92.90점이 나왔습니다. 또한, 평가 데이터에 이 판별함수를 대입했을 때, 판별함수는 매우 유사한 형태로 나왔으며, 판별점수는 92.94점이 나온 것을 확인했습니다. 이런 연구 결과를 바탕으로, 신용카드사는 대출 연체율의 급증을 막기 위해서는 신용카드 대출 신청자에게 조심성 원칙을 적용하거나 지난 결제 이력 및 신용 한도를 확인을 통한 연체 리스크를 막는 것이 필요합니다.

1. 서론

(1) 연구 배경

현대 사회에서, 신용카드는 소비자 금융에 있어 중요한 도구로 자리 잡았습니다. 신용카드는 소비자들에게 편리한 결제 수단을 제공하며, 필요할 때 신속하게 자금을 조달할 수 있는 수단으로도 활용됩니다. 김수래(2022)는 신용카드회사들이 제1금융권인 은행들처럼 고객에게 돈을 빌려주는 일을 하고 있다고 연구 배경을 설명했습니다. 또한, 2013년부터 2022년까지의 대출태도지수와 대출수요지수를 활용했을 때, 대출태도를 강화하면 대출수요는 떨어지고, 대출태도를 약화하면 대출수요가 강화되는 반비례 관계가 나타나고 있는 것을 보였습니다. 그러나 이러한 편리함 뒤에는 과도한 대출과 상환 부담이라는 잠재적인 위험이 발생할 수 있습니다. 특히, 최근 경제 불확실성과 금융 시장의 변동성 증대 등의 요인들로 인해 신용카드 대출 상환 문제 및 신용카드 연체자들이 늘고 있는 상황입니다. 민경록(2007)은 이러한 신용카드 사용자들의 연체 요인에 대해서 4가지 요인을 사용하여 실증연구를 진행했습니다. 해당 요인들은 선별위험요인, 속성위험요인, 기대소득요인, 부도비용요인으로 분류되었습니다. 이러한 요인들로 다기간 로짓 모형으로 분석했을 때, 선별위험요인, 기대소득요인은 연체율과 음의 상관관계를 이루었고 일부 속성위험요인은 양의 상관관계를 이루었고, 또 다른 속성위험요인은 음의 관계를 이루었습니다. 마지막으로, 부도비용요인은 연체율과의 관계가 있다는 증거를 찾지 못했다는 결과를 얻었습니다. 이러한 연구 외에도, 서지용(2017)은 은행의 신용위험 결정요인에 관한 분석기간별 비교연구를 했을 때, 금융위기가전부터 금융위기 진행기간에서는 경기확장이 이루어질 때, 신용 위험이 증가했으며, 금융위기 이후부터 최근까지의 기간에서는 인플레이션 현상이 일어날 때, 신

용 위험이 증가하는 현상이 나타난다고 했습니다. 또한, 위 기간과 관련없이 주택가격 상승과 은행 수익성 개선 등의 변수들로 인해서 신용위험은 낮아지는 결과를 얻었다고 분석했습니다.

(2) 연구 목표

현재 연구하고자 하는 데이터에 연체위험 요인들로 분류함으로써 어떤 요인에서 연체위험이 증가하고 감소하는지에 대한 분석을 진행해 보고자 합니다. 이러한 요인들을 신용카드 대출 데이터 안의 독립변수들을 통해서 4가지 요인으로 분류할 것이며, 해당 요인들을 통한 판별분석을 진행할 것입니다. 이 판별분석은 신용카드로 대출을 진행한 사람들의 연체여부를 종속변수로 하여 대출을 잘 갚고 있는 사람인지, 대출을 잘 갚고 있지 않는 사람인지를 판별점수를 통해서 나타내는 것입니다. 그리고 더 나아가서, 대출을 잘 갚고 있지 않는 연체자들에게 있어서 어떠한 해결책이 필요한지에 대한 분석을 진행해 보고자 합니다.

(3) 분석하려고 하는 데이터 설명

본 연구에서 사용한 데이터는 데이콘이 주최했던 신용카드 고객 대출 등급 분류 콘테스트에 쓰였던 데이터입니다. 해당 데이터는 훈련 데이터와 평가 데이터가 따로 있으며, 대출 등급 분류에 관한 내용이기 때문에 훈련 데이터에는 대출 등급 변수가 존재하며, 평가 데이터에는 대출 등급 변수가 존재하지 않습니다. 대출 등급 변수 외에도, 해당 데이터에는 ID, 대출금액, 대출기간, 근로기간, 주택소유상태, 연간소득, 부채 대비 소득 비율, 총 계좌 수, 대출 목적, 총상환연금, 총상환이자, 총연체금액, 연체계좌 수, 이렇게 14가지의 변수가 있습니다. 이 변수들을 활용하여, 훈련 데이터의 요인분석 및 판별분석을 진행할 것이며, 해당 판별분석에서 나온 판별함수를 활용하여 평가 데이터에 적용하여 판별분석을 진행해서 나온 정확도(Accuracy)를 구해보려고 합니다.

2. 연구 내용

(1) 판별분석이란?

판별분석이란 데이터 분석 기법 중 하나로, 변수 간의 관계를 분석하여 주어진 데이터의 그룹을 구분하는데 사용됩니다. 또한, 예측이나 분류 관련 문제를 해결하는 데에 우수한 성능을 가지는 분석 방법입니다. 마지막으로, 변수들 간의 관계를 통해 분류하려고 하는 그룹 간의 차이를 명확히 이해하고자 할 때 사용됩니다.

판별분석의 주요 목적은 데이터 안의 많은 변수들이 있고, 그 변수들을 그룹화했을 때, 그룹 간의 차이를 최대화하는 변수를 찾아내는 것입니다. 이를 찾아내기 위해서 판별분석은 그룹 간의 분산과 그룹 내 분산을 비교함으로써 그룹을 잘 구분할 수 있는 변수를 찾아냅니다. 이러한 과정을 통해 사용하려고 하는 데이터가 각 그룹에 속할 확률을 추정하고,

이 추정으로 평가 데이터나 새로운 데이터의 그룹을 예측하는 데 사용합니다.

판별분석을 위해서는 각 데이터들이 소속된 그룹에 얼마나 잘 판별되는가에 대한 판별력을 측정해야 합니다. 그 판별력을 ‘판별 점수’라는 지표를 이용하여 높을수록 그룹을 더 잘 판별한다고 할 수 있습니다. 이 판별력을 위해서는 판별 함수가 필요합니다. 판별 함수란, $Y = a_1 * F_1 + a_2 * F_2 + \dots + a_k * F_k$ ($k = \text{number of Factors}$)를 수식으로 하며, 해당 함수를 이용하여 판별분석을 진행합니다.

(2) 판별분석을 하기 위한 종속변수 설정

판별분석을 진행할 때 가장 먼저 고려했던 것은 각 데이터들이 소속된 그룹에 얼마나 잘 판별되는지에 대해서 고민한 것이 아니라, 그 소속된 그룹에 어떤 종속변수가 들어가야 하는지에 대한 것이었습니다. 그렇기 때문에 대출받은 것을 잘 갚는지에 대해서 확인하기 위해서 가장 중요한 것은 최근 몇 년동안 그 사람이 연체되지 않고 꾸준히 갚고 있는 지에 대한 내용입니다. 즉 현재 갖고 있는 데이터 내에서 이와 관련된 변수인 “최근 2년간 연체 횟수”를 통해서 현재 그 사람이 잘 갚고 있는지를 확인해 볼 수 있습니다.

하지만 단순히 연체 횟수를 종속변수로 사용한다면, 0번 연체된 사람부터 20번 연체된 사람을 판별하는 방법을 사용하는 것인데, 해당 변수로 분석한다면 판별 점수가 매우 낮게 나올 것입니다. 하지만 이 변수를 “연체 여부”라는 변수로 변경하여 연체 횟수가 없는 사람과 연체 횟수가 있는 사람으로 구분하고, 이 종속변수를 판별분석에 적용해 본다면, 판별함수 및 판별분석이 가능할 것이라고 생각했습니다. 또한, 연체여부를 판별할 때, 최근 2년동안 연체가 1회밖에 되지 않았다면, 연속으로 연체가 된 상태가 아니기 때문에 연체상태라고 할 수 없습니다. 따라서, 연체 1회는 연체가 되지 않은 상태로 지정하고 1회 이하와 2회 이상으로 구분하여 0과 1로 종속변수를 나타내려고 했습니다.

(3) 신용카드 대출 데이터 전처리 과정

- 이상치 데이터 삭제 - 부채 대비 소득 비율이 9999인 경우, 주택소유상태에 해당하는 값이 1개 밖에 없는 경우
- 근로기간 변수 삭제 - 근로기간에는 na값이 5671개가 있는 것을 확인했으며, 이 na값이 판별분석을 하는 데에 있어서 코딩 에러가 많이 일어나서 삭제하게 되었습니다.

(4) 신용카드 대출 상환 판별분석 전 독립변수들의 요인분석

신용카드 대출 상환에 대한 판별분석을 진행하기 전, 총 9개의 독립변수들을 가지고 있는 데이터에서 판별분석을 진행하기에는 무리가 있다고 판단했습니다. 해당 변수들을 가지고 판별분석을 진행한다면, 해당 변수들간의 관계에 대해서 제대로 알 수 없으며, 코딩하는데 걸리는 시간이 길어질 수 있기 때문입니다. 따라서, 9개의 독립변수를 가지고 요인분석을 진행한다면, 3~4가지 요인으로 나눌 수 있을 것이라고 생각했습니다. 요인분석이란, 관측된 변수들 간의 상관 관계를 바탕으로 데이터의 구조를 파악하는 분석입니다. 변수들

간의 상관 관계가 높을 때, 그 변수들을 묶어서 요인이라고 부르고, 요인들을 통해서 데이

Principal Components Analysis

Call: principal(r = loan2, nfactors = 4, rotate = "varimax", scores = TRUE)

Standardized loadings (pattern matrix) based upon correlation matrix

	RC1	RC2	RC3	RC4	h2	u2	com
대출금액	0.76	0.41	0.09	0.01	0.76	0.24	1.6
대출기간	0.79	-0.27	0.02	-0.19	0.74	0.26	1.4
연간소득	0.04	0.69	0.10	-0.23	0.54	0.46	1.3
부채_대비_소득_비율	0.17	-0.41	0.67	0.06	0.66	0.34	1.8
총계좌수	0.09	0.23	0.81	-0.11	0.73	0.27	1.2
총상환원금	0.25	0.65	0.01	0.33	0.59	0.41	1.8
총상환이자	0.81	0.25	0.01	0.18	0.76	0.24	1.3
총연체금액	0.03	-0.08	0.02	0.86	0.75	0.25	1.0
연체계좌수	-0.10	0.08	0.25	0.18	0.11	0.89	2.5

	RC1	RC2	RC3	RC4
SS loadings	1.98	1.44	1.19	1.02
Proportion Var	0.22	0.16	0.13	0.11
Cumulative Var	0.22	0.38	0.51	0.63
Proportion Explained	0.35	0.26	0.21	0.18
Cumulative Proportion	0.35	0.61	0.82	1.00

Mean item complexity = 1.5

Test of the hypothesis that 4 components are sufficient.

The root mean square of the residuals (RMSR) is 0.12

with the empirical chi square 99940.74 with prob < 0

Fit based upon off diagonal values = 0.61

터의 차원을 축소하거나 변수 간의 패턴을 파악할 수 있습니다.

주어진 요인분석 결과를 기반으로 변수들을 4가지 요인으로 분류하려면 각 변수들의 주성분 값을 고려해야 합니다. 주성분 적재값은 변수들이 각 요인과 어떤 관계를 가지는지를 나타내며, 적재값이 높은 요인에 해당하는 변수들은 그 요인과 관련이 크다고 볼 수 있습니다. 따라서 요인분석을 하면서 4가지 요인으로 분류했습니다.

요인1: 대출금액, 대출기간, 총상환이자 -> 대출과 관련된 변수들

요인2: 연간소득, 총상환원금 -> 소득과 관련된 변수들

요인3: 부채 대비 소득 비율, 총계좌수 -> 부채와 관련된 변수들

요인4: 총연체금액, 연체계좌수 -> 연체와 관련된 변수들

총 이렇게 4가지 요인으로 분류해볼 수 있습니다.

(5) 신용카드 대출 상황 요인 간의 판별분석 결과

판별분석 전 요인분석을 통해 요인 4가지를 추출했습니다. 따라서, 이 4가지 요인을 가지고 판별함수를 만들어서 분석을 진행합니다. 판별분석을 위해서는 4가지 요인 외에도 각 요인이 주어진 관측치에 대해 얼마나 강하게 나타나는지를 나타내는 값인 요인 점수가 필요합니다. 또한, R의 lda 함수를 활용한다면, 종속변수인 연체여부와 요인점수를 통해서 판별분석을 진행해 볼 수 있습니다. lda 함수를 활용한 결과입니다.

```

> discriminant_model <- lda(dependent_variable ~ ., data =
factor_scores)
> discriminant_model
Call:
lda(dependent_variable ~ ., data = factor_scores)

Prior probabilities of groups:
      0      1
0.92915299 0.07084701

Group means:
      RC1      RC2      RC3      RC4
0  0.007740035 -0.01149786 -0.0236870 -0.001763415
1 -0.101509954  0.15079355  0.3106532  0.023127050

Coefficients of linear discriminants:
      LD1
RC1 -0.28285270
RC2  0.42017910
RC3  0.86562044
RC4  0.06444243

```

이 결과로 인해서 4가지 요인의 선형 판별함수를 구할 수 있습니다. 제 연구 결과로 구한 판별함수는 $LD1 = -0.2829 \cdot RC1 + 0.4202 \cdot RC2 + 0.8656 \cdot RC3 + 0.0644 \cdot RC4$ 입니다. 판별함수를 구한 뒤, 판별분석에서 이뤄져야 하는 작업은 연체여부를 예측하는 것입니다. 판별분석 시 연체여부를 예측하기 위해서 사용하는 함수는 predict() 함수를 사용했습니다.

```

factor_scores <- data.frame(factor_analysis$scores[, 1:4])
dependent_variable <- loan$연체여부2
discriminant_model <- lda(dependent_variable ~ ., data = factor_scores)
discriminant_model
predictions <- predict(discriminant_model)
predicted_classes <- predictions$class #0과 1로 구분된 판별 분석 분류.
# 실제 값 추출
actual_classes <- factor(loan$연체여부2)
accuracy <- mean(predicted_classes == actual_classes)
accuracy
#93%

```

해당 분석 및 코드를 실행해 본 결과입니다.

위 실행 결과처럼, 위에 언급했던 판별함수는 일치함을 알 수 있고, 정확도는 92.91% 정도 예측할 수 있는 것을 알 수 있습니다.

(6) 훈련 데이터의 요인분석 및 판별분석 결과로 평가 데이터 연체여부 예측 결과

훈련 데이터로 진행되었던 판별분석 결과를 토대로, 평가 데이터의 연체여부를 예측해 보는 코드 및 결과입니다.

```

> discriminant_model
Call:
lda(dependent_variable ~ ., data = factor_scores)

Prior probabilities of groups:
      0      1
0.92915299 0.07084701

Group means:
      RC1      RC2      RC3      RC4
0  0.007740035 -0.01149786 -0.0236870 -0.001763415
1 -0.101509954  0.15079355  0.3106532  0.023127050

Coefficients of linear discriminants:
      LD1
RC1 -0.28285270
RC2  0.42017910
RC3  0.86562044
RC4  0.06444243
> predictions <- predict(discriminant_model)
> predicted_classes <- predictions$class #0과 1로 구분된 판별분석 분류.
> # 실제값 추출
> actual_classes <- factor(loan$연체여부2)
> accuracy <- mean(predicted_classes == actual_classes)
> accuracy
[1] 0.9290907
> #93%
#훈련 데이터의 요인분석을 활용한 평가 데이터의 요인분석.
factor_scores_test <- predict(factor_analysis,
                             loan_test2[,c('대출금액', '대출기간', '연간소득',
                                              '부채_대비_소득_비율', '총계좌수',
                                              '총상환원금', '총상환이자', '총연체금액',
                                              '연체계좌수')]))[, 1:4]

factor_scores_test <- data.frame(factor_scores_test)
head(factor_scores_test)

discriminant_model_test <- lda(dependent_variable_test ~ ., data = factor_scores_test)
discriminant_model_test
# 선형 판별분석 모델을 사용하여 테스트 데이터 예측
predictions <- predict(discriminant_model_test, newdata = factor_scores_test)
# 예측된 클래스
predicted_classes <- predictions$class
# 실제값 추출
actual_classes <- factor(loan_test$연체여부2)

# 정확도 계산
accuracy <- mean(predicted_classes == actual_classes)
accuracy
#93점.

```

이 결과를 놓고 봤을 때, 판별분석을 통해서 연체자를 분류할 때, 92.94%의 판별점수를 얻을 수 있었습니다.


```

> discriminant_model_test
Call:
lda(dependent_variable_test ~ ., data = factor_scores_test)

Prior probabilities of groups:
      0      1
0.92946601 0.07053399

Group means:
      RC1      RC2      RC3      RC4
0  0.007128161 -0.01039612 -0.02300765 -0.0007255956
1 -0.093931779  0.13699553  0.30318478  0.0095615810

Coefficients of linear discriminants:
      LD1
RC1 -0.27154384
RC2  0.40557708
RC3  0.87604231
RC4  0.02403069
> # 선형 판별분석 모델을 사용하여 테스트 데이터 예측
> predictions <- predict(discriminant_model_test, newdata = factor_scores_test)
> View(predictions)
> # 예측된 클래스
> predicted_classes <- predictions$class
> # 실제값 추출
> actual_classes <- factor(loan_test$연체여부2)
>
> # 정확도 계산
> accuracy <- mean(predicted_classes == actual_classes)
> accuracy
[1] 0.9294037
> #93점.

```


3. 결론

(1) 결과 요약

신용카드 대출 데이터 안의 독립변수들을 통해서 4가지 요인을 요인분석을 통해서 구하였는데, 그 요인들은 대출, 소득, 부채, 연체 관련 요인으로 구분했습니다. 그리고, 이 4가지 요인들로 판별분석을 진행하기 전, 판별함수를 구하였을 때의 판별함수는 $LD1 = -0.2829 \cdot RC1 + 0.4202 \cdot RC2 + 0.8656 \cdot RC3 + 0.0644 \cdot RC4$ 이란 값이 나왔습니다. 이 중 대출 관련 변수인 RC1의 계수 값은 음수로 나왔습니다. 따라서, 대출금액이 커지거나 대출기간이 길어질수록 연체될 확률은 줄어드는 것으로 해석해 볼 수 있습니다. 연체 관련 요인들에게는 0.10보다 낮은 계수 값이기 때문에 다른 변수들보다 연체여부에 적은 영향을 끼치고 있는 것을 확인했습니다. 다른 두 값들은 데이터의 값이 더 클수록, 연체여부가 상승하고 있다는 결과를 보여줍니다. 마지막으로, 평가 데이터의 판별함수 또한 구했을 때, $LD1 = -0.2715 \cdot RC1 + 0.4056 \cdot RC2 + 0.8760 \cdot RC3 + 0.0240 \cdot RC4$ 의 값이 나왔고, 이는 훈련 데이터의 판별함수와 매우 비슷한 결과치임을 알 수 있습니다.

판별분석을 통해서 얻은 결과입니다. 훈련 데이터를 통한 훈련 데이터의 판별분석을 진행했을 때의 정확도는 92.90%의 결과를 얻었습니다. 그리고 훈련 데이터의 요인분석을 평가 데이터에 대입했을 때 구한 평가 데이터의 4가지 요인들에 관한 데이터프레임을 가지고 판별분석을 진행했을 때, 정확도는 92.94%가 나왔습니다.

(2) 결과를 활용한 결론 도출

신용카드 대출 상환 예측을 위한 판별분석을 진행한 결과, 확인할 수 있었던 것은, 96292개의 데이터 중 7%에 가까운 2회 이상 연체되고 있는 사람들을 예측했을 때의 판별 점수 '93점' 입니다. 하지만, 예측된 연체자들에게 있어 중요한 변수가 무엇인지 확인 절차를 진행해 보지 않았고 연체자들을 어떻게 하면 줄일 수 있는지에 대한 해결책을 제시하는 것이 필요합니다. 몇 가지 선행연구를 통해 해결책을 제시해보려고 합니다.

Pratama, D. R.(2018)은 사회학적 · 법학적 접근을 통한 은행 대출의 주요 요인 및 요인의 해결책을 제시했습니다. 은행 대출의 주요 요인에는 채무자의 사업 실패, 고객의 경제 활동 감소 및 고금리, 채무자의 신용 능력 부족 등이 있다고 했습니다. 이러한 상황에서, 고객에게 해결책을 제시하는 것이 아니라, 은행을 중심으로 해결책을 제시했습니다. 이 논문의 해결책에는 채무자와 대출 계약을 실행할 때 조심성의 원칙을 적용하거나 법적으로 채무자에게 감시 및 징벌을 시행해야 한다고 했습니다. 또한, Alam, T. M. (2020)은 신용카드사는 고객에게 신용카드를 발급할 때, 해당 사람의 지난 결제 이력을 확인해 봐야 한다고 주장했습니다. 그 이유는, 이미 연체되고 있는 자들은 지불을 지연시킬 가능성이 높고, 상환할 수 없기 때문입니다. 또한, 신용카드 신청자의 신용 한도를 살펴보는 것이 중요하다고 했는데, 그 이유는 연체되지 않고 돈을 갚는 사람들은 신용점수가 높은 경향이 있으므로, 은행은 이런 사람들에게 신용 한도를 늘리면서 리스크를 감소시키는 것이 필요하

다고 했습니다.

이러한 선행연구를 바탕으로, 신용카드사는 예측된 연체자들의 지난 결제 이력 및 신용 한도를 살펴보고, 연체되거나 갚지 않은 기록이 있다면, 조심성의 원칙 적용 및 감시를 시행합니다. 감시 이후, 갚지 않을 것이라고 판단된다면, 징벌 및 제공된 계약을 이행시키기 위한 대비책이 있어야 할 것으로 결론을 제시합니다.

(3) 추가 연구가 필요한 부분이나 보완해야 할 부분

실제 데이터 값과 예측한 데이터 값이 얼마나 차이가 있는지 확인했을 때 92.90점이 나왔다는 것은 꽤 높은 수치로 예측한 값이라고 할 수 있습니다. 하지만, 데이터 전처리 부분에서 사용하지 않고 삭제했던 변수인 근로기간 변수 또한 사용한다면, 예측값이 더 정확하거나 높을 수 있다고 판단했습니다. 근로기간 변수는 결측값이 5671개 값이 있었으며, 해당 결측값인 데이터를 삭제하고 진행하거나 근로기간 변수 자체를 삭제하는 두 가지 방향 중 한 가지를 선택했어야 했고, 후자를 선택하게 되었습니다. 근로기간 변수는 연구 초기에 요인분석에 포함되어서 분석했을 때, 부채 관련 요인의 핵심 변수로 분류되었었기 때문에, 이 변수를 삭제하게 되다면 부채 관련 요인 간의 관계가 약해질 것이라고 판단해 볼 수도 있습니다. 따라서, 추가 연구 시 근로기간 변수 또한 요인분석 및 판별분석을 할 수 있도록 데이터를 정제하는 과정이 필요하다고 분석했습니다.

참고 문헌

- 김수래, 김주일. (2022). 신용카드회사의 대출행태에 있어서 신용위험과 대출수요와의 연관성에 관한 연구. 신용카드리뷰, 16(1), 54-69.
- 민경록(2007). 국내 신용카드 사용자의 연체요인에 관한 실증연구. 대한경영학회지, 20(4), 1953-1976.
- 서지용. (2017). 은행의 신용위험 결정요인에 관한 분석기간별 비교연구: 대출자산 구성변화를 중심으로. Journal of The Korean Data Analysis Society, 19(2), 797-808.
- 이민준, 이정환(2019), “대출 포트폴리오와 대출금 변수를 중심으로 본 국내은행 신용위험의 결정요인”, 「산업경제연구」, 32(1), 49-76.
- Alam, T. M. (2020). An investigation of credit card default prediction in the imbalanced datasets. Ieee Access, 8, 201173-201198.
- Gross, David B. and Nicholas S. Souleles,(2002), “An Empirical Analysis of Personal Bankruptcy and Delinquency”, The Review of Financial Studies, (Spring),Vol 15, No 1, pp. 319-347.
- H. Kim, H. Cho, and D. Ryu, (2018). “An empirical study on credit card loan delinquency,” Econ. Syst., vol. 42, no. 3, pp. 437-449
- Pratama, D. R., & Purnawan, A. (2018). Default In And Credit Agreement And Implementation Of Solution Efforts (A Case Study Of Decision 336/Pdt/G/2016/Pn. Smg). Jurnal Daulat Hukum, 1(2), 345-350.
- P. Xu, Z. Ding, and M. Pan. (2018) “A hybrid interpretable credit card users default prediction model based on RIPPER,” Concurrency Comput. Pract. Exper., vol. 30, no. 23, p. e4445
- Zribi, N. and Boujelbene, Y.(2011), The factors influencing bank credit risk: the case of Tunisia, Journal of Accounting and Taxation, 3(4), 70-78.