



CARD MANAGEMENT SYSTEM

TEAM

김래원, 김정훈, 박선하, 박종호, 우병준, 임병남, 장현, 황태현

프로젝트 목적

고객의 신용카드 및 대출 이력 데이터를 분석하여 연체금액, 연체일수, 연체 빈도 등을 기준으로 고객별 금융 위험군을 분류하는 것이 핵심 목표입니다.

활용방안

분류된 고객 위험군은 추심(채권 회수), 고객 상담, 사전 모니터링 등의 실무적 조치의 우선순위 결정에 활용될 수 있습니다.

기대 효과

업무 효율
성 향상

금융사 내부
시스템 연계
가능성

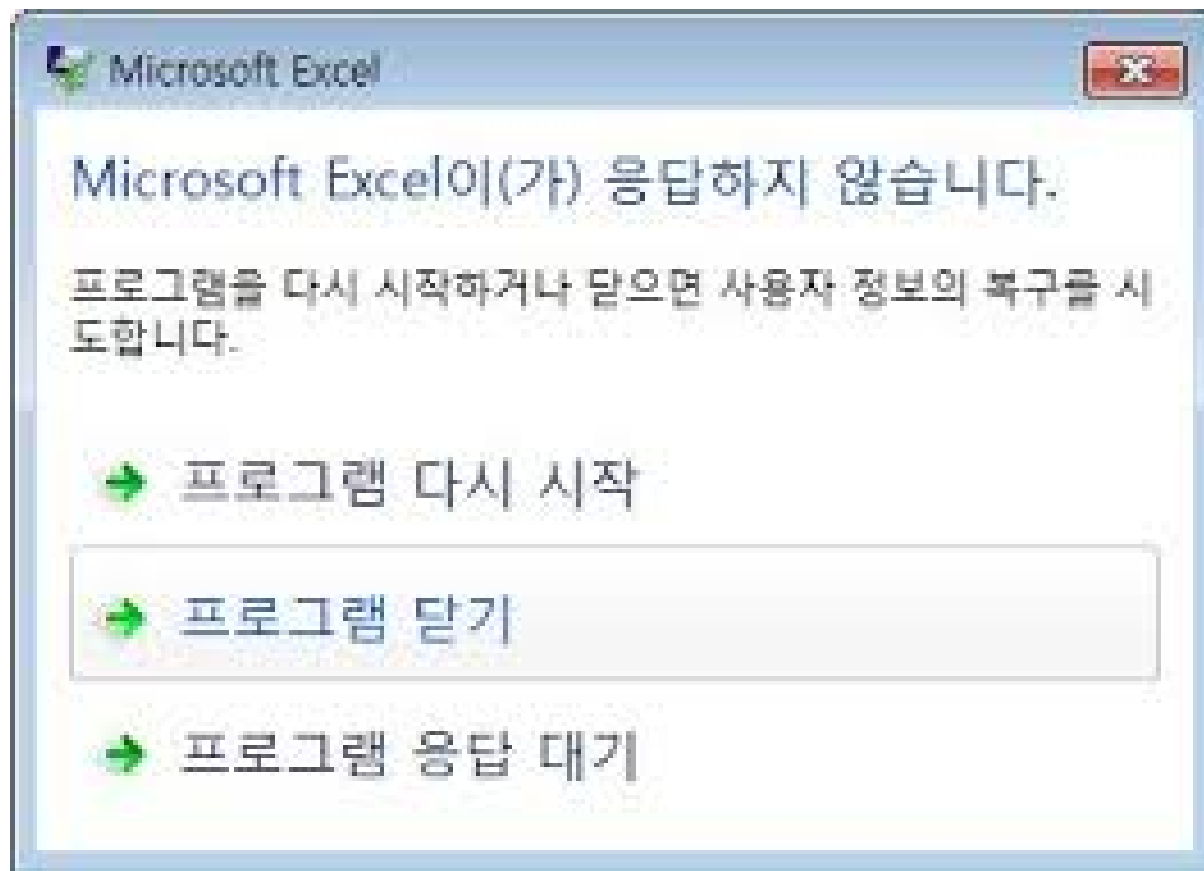
리스크 대
응력 강화

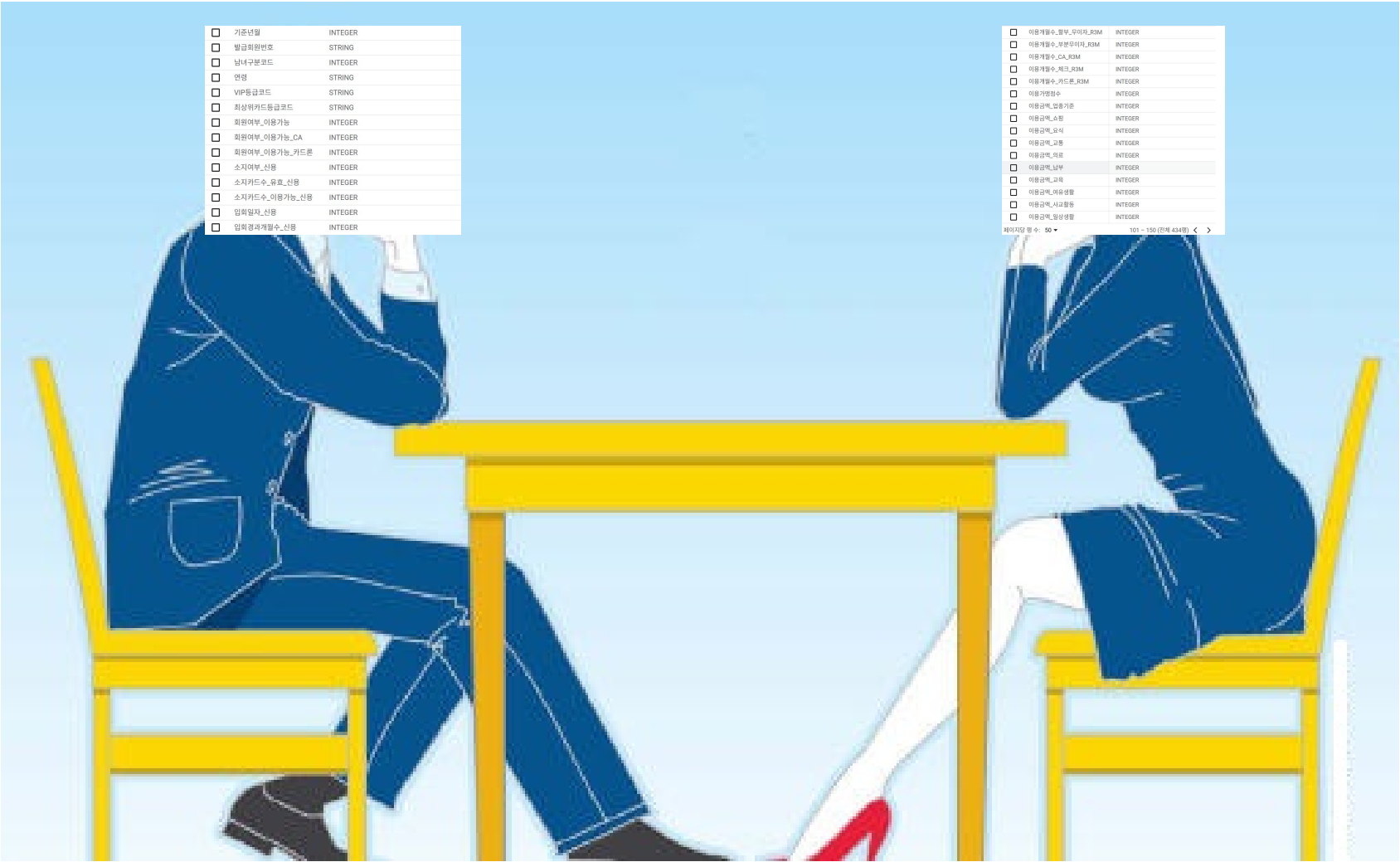
<input type="checkbox"/>	기준년월	INTEGER
<input type="checkbox"/>	발급회원번호	STRING
<input type="checkbox"/>	남녀구분코드	INTEGER
<input type="checkbox"/>	연령	STRING
<input type="checkbox"/>	VIP등급코드	STRING
<input type="checkbox"/>	최상위카드등급코드	STRING
<input type="checkbox"/>	회원여부_이용가능	INTEGER
<input type="checkbox"/>	회원여부_이용가능_CA	INTEGER
<input type="checkbox"/>	회원여부_이용가능_카드론	INTEGER
<input type="checkbox"/>	소지여부_신용	INTEGER
<input type="checkbox"/>	소지카드수_유효_신용	INTEGER
<input type="checkbox"/>	소지카드수_이용가능_신용	INTEGER
<input type="checkbox"/>	입회일자_신용	INTEGER
<input type="checkbox"/>	입회경과개월수_신용	INTEGER

<input type="checkbox"/>	이용개월수_할부_무이자_R3M	INTEGER
<input type="checkbox"/>	이용개월수_부분무이자_R3M	INTEGER
<input type="checkbox"/>	이용개월수_CA_R3M	INTEGER
<input type="checkbox"/>	이용개월수_체크_R3M	INTEGER
<input type="checkbox"/>	이용개월수_카드론_R3M	INTEGER
<input type="checkbox"/>	이용가맹점수	INTEGER
<input type="checkbox"/>	이용금액_업종기준	INTEGER
<input type="checkbox"/>	이용금액_쇼핑	INTEGER
<input type="checkbox"/>	이용금액_요식	INTEGER
<input type="checkbox"/>	이용금액_교통	INTEGER
<input type="checkbox"/>	이용금액_의료	INTEGER
<input type="checkbox"/>	이용금액_납부	INTEGER
<input type="checkbox"/>	이용금액_교육	INTEGER
<input type="checkbox"/>	이용금액_여유생활	INTEGER
<input type="checkbox"/>	이용금액_사교활동	INTEGER
<input type="checkbox"/>	이용금액_일상생활	INTEGER

페이지당 행 수: 50 ▼

101 - 150 (전체 434행) < >





<input type="checkbox"/>	기준년월	INTEGER
<input type="checkbox"/>	발급회원번호	STRING
<input type="checkbox"/>	남녀구분코드	INTEGER
<input type="checkbox"/>	연령	STRING
<input type="checkbox"/>	VIP등급코드	STRING
<input type="checkbox"/>	최상위카드등급코드	STRING
<input type="checkbox"/>	회원여부_이용가능	INTEGER
<input type="checkbox"/>	회원여부_이용가능_CA	INTEGER
<input type="checkbox"/>	회원여부_이용가능_카드론	INTEGER
<input type="checkbox"/>	소지여부_신용	INTEGER
<input type="checkbox"/>	소지카드수_유효_신용	INTEGER
<input type="checkbox"/>	소지카드수_이용가능_신용	INTEGER
<input type="checkbox"/>	입회일자_신용	INTEGER
<input type="checkbox"/>	입회경과개월수_신용	INTEGER

<input type="checkbox"/>	기준년월	INTEGER	NULLABLE	-	-	-
<input type="checkbox"/>	발급회원번호	STRING	NULLABLE	-	-	-
<input type="checkbox"/>	최종이용일자_기본	INTEGER	NULLABLE	-	-	-
<input type="checkbox"/>	최종이용일자_신판	INTEGER	NULLABLE	-	-	-
<input type="checkbox"/>	최종이용일자_CA	INTEGER	NULLABLE	-	-	-
<input type="checkbox"/>	최종이용일자_카드론	INTEGER	NULLABLE	-	-	-
<input type="checkbox"/>	최종이용일자_체크	INTEGER	NULLABLE	-	-	-
<input type="checkbox"/>	최종이용일자_일시불	INTEGER	NULLABLE	-	-	-
<input type="checkbox"/>	최종이용일자_할부	INTEGER	NULLABLE	-	-	-
<input type="checkbox"/>	이용건수_신용_BOM	INTEGER	NULLABLE	-	-	-
<input type="checkbox"/>	이용건수_신판_BOM	INTEGER	NULLABLE	-	-	-
<input type="checkbox"/>	이용건수_일시불_BOM	INTEGER	NULLABLE	-	-	-
<input type="checkbox"/>	이용건수_할부_BOM	INTEGER	NULLABLE	-	-	-

전처리 : 파생변수 사용

- `def preprocessing_data(data):`
- `data['이용건수_BOM'] = data[['이용건수_일시불_BOM','이용건수_체크_BOM','이용건수_할부_BOM']].sum(axis=1) -> vip`
- `data['이용금액_BOM'] = data[['이용금액_일시불_BOM','이용금액_체크_BOM','이용금액_할부_BOM']].sum(axis=1)`
- `data['최종이용일자'] = data[['최종이용일자_일시불','최종이용일자_체크','최종이용일자_할부']].max(axis=1)`
- `data['이용후경과월'] = data[['이용후경과월_일시불','이용후경과월_체크','이용후경과월_할부']].sum(axis=1)`
 -

```
from xgboost import XGBClassifier
model = XGBClassifier(
    n_estimators=300,
    max_depth=4,
    learning_rate=0.05,
    objective='multi:softprob',
    use_label_encoder=False,
    eval_metric='mlogloss',
    random_state=42
)
model.fit(X_train,y_train)
```


Accuracy : 0.8989083333333333

Precision: 0.7928580773718682

Recall : 0.7160552468528755

F1 Score : 0.7482546671213121

Classification Report:

precision recall f1-score support

0	0.94	0.97	0.95	480367
1	0.75	0.53	0.62	32276
2	0.69	0.64	0.67	87357

accuracy			0.90	600000
----------	--	--	------	--------

macro avg 0.79 0.72 0.75 600000

weighted avg 0.89 0.90 0.89 600000

Xgboost

사용모델

Accuracy : 0.88491

Precision: 0.7581466425054462

Recall : 0.6768148704744555

F1 Score : 0.7088617137684281

Classification Report:

precision recall f1-score support

0	0.93	0.96	0.95	480366
1	0.70	0.46	0.55	32277
2	0.64	0.61	0.62	87357

accuracy			0.88	600000
----------	--	--	------	--------

macro avg 0.76 0.68 0.71 600000

weighted avg 0.88 0.88 0.88 600000

랜덤포레스트

추심 전처리 (최근, b1m , b2m)

숫자 필터(F) >

검색

- ☒ (모두 선택)
- ☒ -999999999
- ☒ 62
- ☒ 66
- ☒ 67
- ☒ 68
- ☒ 69
- ☒ 70
- ☒ 71
- ☒ 72
- ☒ 73
- ☒ 74
- ☒ 75

확인 취소

숫자 필터(F) >

검색

- ☒ (모두 선택)
- ☒ -999999999
- ☒ 0
- ☒ 10
- ☒ 15
- ☒ 16
- ☒ 18
- ☒ 19
- ☒ 20
- ☒ 21
- ☒ 22
- ☒ 23
- ☒ 24

확인 취소

숫자 필터(F) >

검색

- ☒ (모두 선택)
- ☒ -999999999
- ☒ 0
- ☒ 48
- ☒ 50
- ☒ 51
- ☒ 52
- ☒ 53
- ☒ 54
- ☒ 55
- ☒ 56
- ☒ 57
- ☒ 58

확인 취소

0일? -999999999? / 0일은 당일 환납(구독, 정기결제 등) , -999999999는 사용 이력없음

추심 전처리 : 결측 논리

연체일수_최근 	연체원금_B1M	연체원금_B2M	잔액_현금서비스_B1M	잔액_현금서비스_B2M
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

연체일수_최근	연체원금_B1M	연체원금_B2M	잔액_현금서비스_B1M	잔액_현금서비스_B2M
-99999999	0	0	2793073	2458063
-99999999	0	0	473255	554310
-99999999	0	339590	0	0
-99999999	0	0	864988	882394

-99999999는 사용 이력없음인데 연체 및 잔액 발생 = 입력 누락으로 최종 규정

추심 전처리 : 파생변수 생성

- 주 연체원금(`overdue_principal`)
 - 최근 1개월(`overdue_principal_b1m`) 값을 기본으로 사용하도록 한다.
 - 최근 1개월 값이 0이면 최근 2개월(`overdue_principal_b2m`) 값으로 대체한다.
- 주 연체일수(`overdue_days`)
 - 최근 1개월(`overdue_days_b1m`) 값을 기본으로 사용하도록 한다.
 - 최근 1개월 값이 0이면 최근 2개월(`overdue_days_b2m`) 값으로 대체한다.

추심 : 위험군 라벨 생성

```
def classify_risk(row):
    principal = row['overdue_principal']
    days = row['overdue_days']
    repeat = row['last_overdue_months_r15m']

    if principal >= 300000 or days >= 60:
        group = '위험군'
    elif principal >= 100000 or days >= 30:
        group = '준위험군'
    elif principal > 0 or repeat >= 3:
        group = '관찰군'
    else:
        group = '정상군'

    if group in ['정상군', '관찰군']:
        if row['avg_loan_month'] >= 500000 or row['avg_ca_month'] >= 300000:
            group = '준위험군' if group == '관찰군' else '관찰군'
        if row['revolving_carryover_count_6m'] >= 3:
            group = '준위험군' if group == '관찰군' else '관찰군'
```

추심모델 결과

[가데이터 정확도]: 0.9454285714285714

[분류 성능 보고서]

	precision	recall	f1-score	support
관찰군	0.86	0.92	0.89	1609
위험군	1.00	0.99	0.99	952
정상군	0.87	0.64	0.74	301
준위험군	0.97	0.97	0.97	4138
accuracy			0.95	7000
macro avg	0.93	0.88	0.90	7000
weighted avg	0.95	0.95	0.94	7000

김래원

김정훈

박선하

박종호

우병준

임병남

장현

황태현