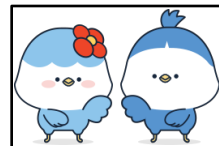




iM DiGital Banker Academy



## 데이터분석 전문가 양성과정

### 프로젝트 기획서

#### 1. 팀 소개

팀명	최다은조
팀원 명	손효은, 염기진, 이우태, 최다은
주제	신용카드 고객 세그먼트 분류기 개발 및 활용성 검토
요약	<p>본 프로젝트는 신용카드 고객의 다양한 금융 활동 데이터를 기반으로, 고객을 행동 및 특성에 따라 그룹화(세그먼트) 하는 AI 기반 분류 알고리즘을 개발하는 것을 목표로 한다. 회원 정보, 신용 정보, 승인 매출, 잔액, 채널 이용, 마케팅 및 성과 데이터 등 다양한 속성 데이터를 활용하여, 고객 간의 유사성과 차이점을 반영한 정교한 세그먼트 분류 모델을 설계한다.</p> <p>먼저, 비지도 학습 기반의 클러스터링 기법을 적용하여 신용카드 고객을 다차원적으로 분류하고, 이를 통해 향후 마케팅, 상품 전략, 리스크 관리 등에 활용 가능한 기초 고객 그룹 정보를 확보한다. 또한, 도출된 잠재 세그먼트를 바탕으로 지도 학습 기반의 세그먼트 분류 예측 모델을 구축하고, SHAP 기반 인사이트 분석을 통해 고객 세분화 결과를 실제 마케팅, 상품 추천, 리스크 관리 등 다양한 실무 분야에 활용할 수 있도록 지원한다.</p>

2. 역할 분담

이름	역할
손효은	데이터 전처리 및 EDA
염기진	Feature Engineering, 모델 학습 및 평가
이우태	리포트 작성 및 시각화, 발표자료 제작
최다은	전체 프로젝트 총괄

3. 추진 일정

단계	기간	주요 활동
사전 기획	4 월 15 일 ~ 4 월 17 일	프로젝트 기획서 작성
데이터 구조 이해 및 EDA 수행	4 월 18 일 ~ 4 월 19 일	데이터 구조 파악 및 타겟 변수 검토
전처리 및 파생 변수 정의	4 월 19 일 ~ 4 월 20 일	주요 변수 전처리 및 파생 변수 생성
모델링 및 초기 성능 평가	4 월 21 일 ~ 4 월 22 일	분류 모델 적용 및 베이스라인 성능 확보

모델 고도화 및 인사이트 도출	4 월 23 일 ~ 4 월 24 일	SHAP 기반 중요 변수 도출 및 성능 개선
최종 발표	4 월 25 일	최종 결과 시각화 및 발표 자료 작성

4. 제안 내용

추진 배경

1. 디지털 전환 속 맞춤형 금융 서비스의 중요성 대두

최근 금융산업에서는 고객의 세부 성향을 파악하고, 이를 기반으로 한 맞춤형 금융 서비스 제공이 점점 더 중요해지고 있다. 특히 디지털 전환이 가속화되며, 마케팅 전략 역시 정교한 고객 세분화(Segmentation)를 중심으로 전환되고 있는 추세다.

이에 대한 실무적 인사이트는 [DITODAY: 고객의 마음을 훔치는 효과적인 방법, 개인화 마케팅](#) 기사에서도 제시된 바 있으며, 고객의 세그먼트에 따라 오픈율, 클릭률 등 마케팅 반응률이 유의미하게 달라지는 점을 통해 맞춤형 접근의 중요성을 확인할 수 있다.

이에 따라 카드사에서는 고객의 다양한 금융 활동 데이터를 기반으로 세분화된 고객 관리 전략을 수립하고자 하는 니즈가 커지고 있으며, 이는 곧 **데이터 기반의 세그먼트 예측 모델 개발**로 이어진다. 예를 들어 연령대별 고객 분포, Segment 별 카드 이용 패턴, 승인 매출 수준 등을 확인하면 각 세그먼트의 고유 특성과 차이를 명확히 확인할 수 있으며, 이는 실무적인 인사이트 도출로도 직결된다.

이번 프로젝트에서는 금융 데이터를 활용해 고객 별 일정 기간의 데이터를 바탕으로 카드사

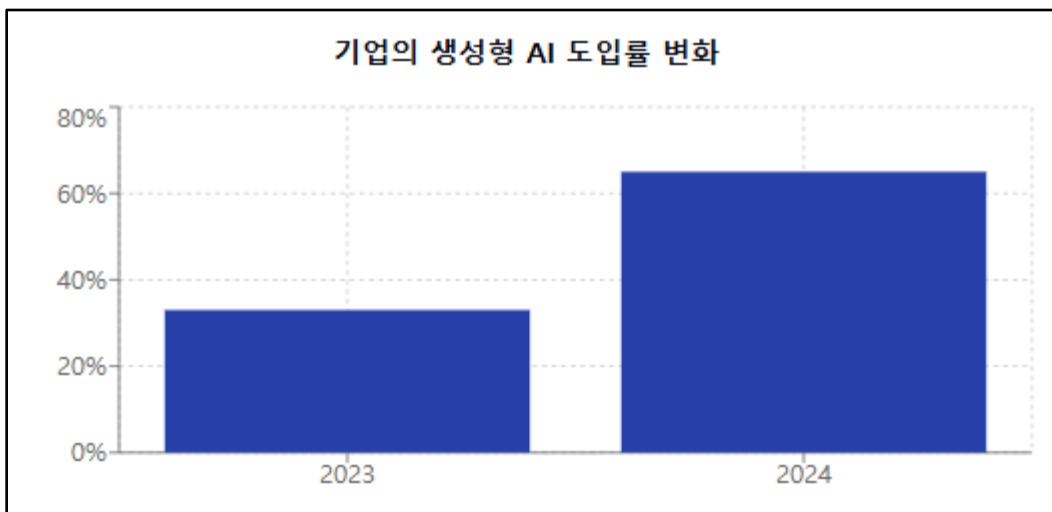
신용카드 고객 세그먼트(Segment)를 분류하는 AI 알고리즘을 개발하는 것을 목표로 한다. 고객의 신용정보, 승인매출, 청구, 잔액, 채널, 마케팅, 성과 등 다양한 데이터를 바탕으로 각 고객들의 세그먼트 분류를 예측하는 모델을 설계하며, 예측된 세그먼트 결과는 시각화를 통해 고객군 특성 파악 및 마케팅 전략 수립 방향성과 연계하여 제시될 예정이다.

신용카드 고객 세그먼트를 정확하게 분류할 수 있는 AI 알고리즘을 개발하고, 이를 기반으로 카드사에서 고객 맞춤형 서비스와 마케팅 전략을 효과적으로 실행할 수 있도록 기여하는 것이 본 프로젝트의 최종 목표이다.

## 2. 고객 세분화 전략의 도입과 성과

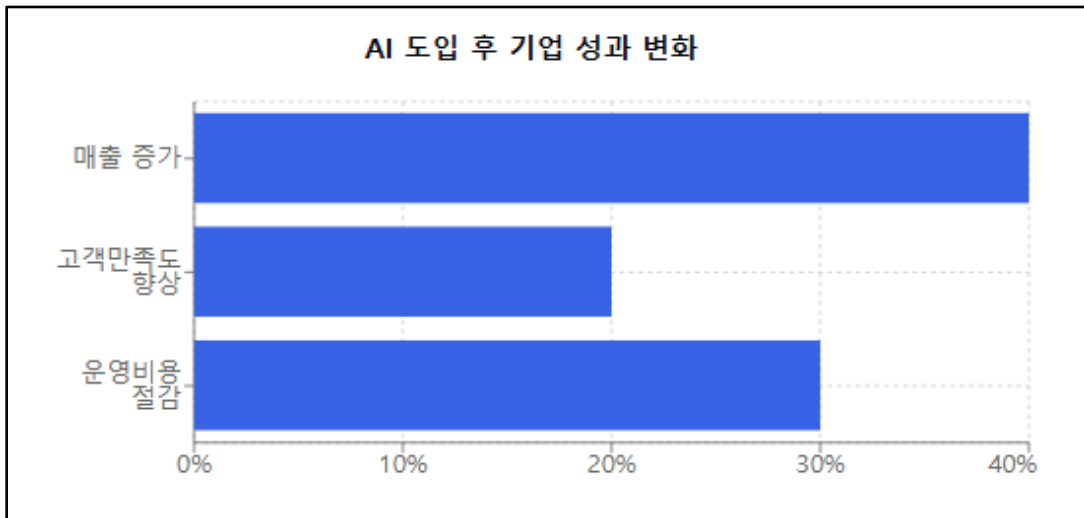
### 1) 초개인화 AI 마케팅이 만들어낸 실질 성과

기업들이 앞다투어 AI 기반 초개인화 마케팅 전략을 도입하고 있다. 맥킨지 글로벌 서베이에 따르면, 기업들의 생성형 AI 도입률은 전년도 33%에서 올해 65%로 급격히 증가하였으며, 특히 마케팅·세일즈와 제품·서비스 개발 분야에서 AI 활용이 매출 증대 및 고객 만족도 향상으로 이어지고 있다.



이러한 변화는 국내 기업의 성과에서도 잘 나타난다. G마켓은 AI 기반 초개인화 서비스를 모바일 앱에 도입한 결과, 홈 화면에서의 구매 전환율이 40% 이상 증가했다고 밝혔다.

CJ 그룹 또한 AI CENTER 를 설립하여 전사적인 디지털 혁신에 속도를 내고 있다.



이는 단순한 상품 추천을 넘어서, 고객의 실시간 행동 패턴과 선호도를 기반으로 한 세분화된 타겟팅 전략이 기업의 핵심 성장 동력이 되고 있음을 보여주는 대표적인 사례다.

## 2) 식음료 브랜드의 정교한 세그먼트 타겟팅 성공 사례

국내 식음료 브랜드 W 사는 자사 고객 데이터를 마케팅에 보다 정밀하게 활용하기 위해, 기존 보유 고객 DB 를 세분화하여 오디언스 타겟팅 캠페인을 시도했다. 기존에는 고객 전체를 대상으로 일괄적인 메시지를 전달하는 방식이었지만, 해당 캠페인에서는 구매 가능성이 높은 고객군만을 추출하여 개별화된 메시지를 발송했다. 이를 위해 W 사는 고객의 과거 구매 이력, 선호 카테고리, 구매 주기 등 다양한 행동 데이터를 기반으로 세그먼트를 나누고, 각 세그먼트별로 맞춤형 쿠폰, 이벤트 메시지, 신제품 알림 등을 발송했다. 특히, '최근 3 개월간 2 회 이상 구매한 고객', '비회원 중 장바구니 이탈 이력이 있는 고객' 등 실질적 행동 지표에 기반한 세분화 전략을 도입하였다.

이 전략은 단순히 메시지 발송 효율을 높인 것을 넘어, 도달률과 반응률에서 큰 향상을 이뤄냈다. 실제로 'Begin 카카오 친구톡 캠페인' 기간 동안 관련 제품군의 매출이 약 3 배 가까이 증가하는 성과로 이어졌다. 또한, 고객 이탈률이 낮아지고, 브랜드 재구매율이 유의미하게 상승하는 부수적 효과도 나타났다. 이 사례는 고객 세분화가 단순한 분류 작업을 넘어 실질적인 비즈니스 성과로 직결될 수 있음을 보여주며, 데이터 기반 마케팅 전략 수립에 있어 '정교한 고객군 정의'의 중요성을 강조하는 대표적인 국내 사례로 평가받고 있다.

## Begin 카카오 친구톡 캠페인 성과

측정 결과

• 순위

초스도

**빅인 캠페인으로 인한 매출 상승**

- 당일 전체 매출 중 47% 비중 차지
- 전일 대비 일매출 194.18% 상승

begin\_talk

2023.07.12

사용자 목록 탐색

날짜	begin_talk	기타
07.09	~100	~400
07.10	~100	~400
07.11	~100	~400
07.12	~2100	~400
07.13	~100	~400
07.14	~100	~400

© 2023 Biginsight Inc. All rights reserved.

위와 같은 추세를 보았을 때, 카드사의 다양한 고객 데이터를 분석해 연체 위험, 소비 수준 등을 반영한 고객 세그먼트 분류 모델을 만드는 것은 보다 효과적이고 실질적인 전략이다. 본 프로젝트는 이를 통해 향후 카드사의 타겟 마케팅, 리스크 관리, 상품 개발의 방향성을 데이터에 기반하여 설계할 수 있는 토대를 마련하고자 한다.

[illegible]

## 1. 금융권의 고객 세분화 전략의 실효성

### 1) KB 국민은행 AI 기반 마케팅 자동화 플랫폼 'AIMs'



최근 금융권에서는 고객 데이터를 기반으로 한 세분화 전략과 맞춤형 마케팅이 활발히 도입되고 있으며, 이는 실제로 마케팅 효율성과 고객 반응률 향상으로 이어지고 있다.

대표적인 사례로 KB 국민카드는 'AI 기반 마케팅 자동화 플랫폼 Aims'를 도입해 고객을 133 개의 세그먼트로 세분화하고, 각 고객의 소비 수준과 혜택 선호도에 맞춰 맞춤형 메시지를 발송하였다. 예를 들어, 어떤 고객에게는 월 50 만 원 이상 카드 사용 시 커피 쿠폰을, 다른 고객에게는 20 만 원 이상 사용 시 편의점 쿠폰을 제공하는 등 이용 수준별 차등 보상 구조를 설계하였다. 이 전략을 통해 고객은 자신에게 맞는 실현 가능한 혜택을 받아 캠페인 참여율이 증가했고, 카드사는 과도한 혜택 제공 없이도 마케팅 효율을 높여 비용 절감 효과를 거둘 수 있었다.

이처럼 정교한 고객 세그먼트 분류와 예측 모델은 실제 금융 서비스 현장에서 성과와 직결되는 핵심 기술로 자리잡고 있다.

## 2)뱅크샐러드 마이데이터 기반 카드 추천 서비스

또 다른 사례로는 뱅크샐러드의 마이데이터 기반 카드 추천 서비스가 있다. 뱅크샐러드는 금융 마이데이터 도입 이후 고객의 소비 습관과 선호도를 분석하여 개인 맞춤형 카드 상품을 추천하는 서비스를 제공해왔으며, 해당 서비스의 누적 이용자 수는 2년 만에 100만 명을 돌파했다. 이 서비스는 130여 개 카드 상품을 실시간 비교해 고객이 가장 많은 혜택을 받을 수 있는 카드를 제안하며, 실제로 카드 재발급률이 높은 수준을 기록하는 등 고객 반응이 매우 긍정적이다.

이는 마이데이터 기반의 정밀한 고객 분석이 금융상품 추천 정확도와 고객 만족도를 동시에 끌어올릴 수 있음을 보여주는 사례이며, 고객 세분화 기반의 예측 모델이 실제 서비스 성과로 이어질 수 있다는 것을 보여준다.

## 3) 업계 1위 신한카드의 개인 마케팅 전략

### 신한카드 맞춤대출

한번의 조회로 적합한 상품을 추천해 드리는 고객 맞춤형 서비스

한도 10만원~5,000만원

이자율 연 5.2%~19.9%

한도조회는 신용평점에 영향을 주지 않습니다.

한도조회 및 신청

### 카드사별 7월 누적 신용판매액

단위: 원

신 용 판 매 액		
신	한	90조1639억
연	대	80조8461억
삼	성	79조445억
K B	국 민	70조6673억
롯데	데	46조5783억
N H	농 협	34조1952억
우 리		32조7416억
하 나		29조22억
비 씨		12조3507억
합	계	475조5899억

자료: 여신금융협회 대한금융신문

국내 카드사들이 판촉 비용은 줄이고 비용 절감 효과를 얻을 수 있는 개인 맞춤 마케팅에 집중하고 있다. 수천만명의 고객 금융 데이터를 통해 고객이 꼭 필요한 혜택을 제공하고 카드



이용을 유도하려는 전략이다.

신한카드는 2018 년부터 맞춤형 마케팅 강화를 위한 내부 교육을 실시해왔다. 고객의 구매 이력과 유사 연령대의 선호도를 분석해 가맹점을 추천하고, 개인별로 최적화된 쿠폰을 발급하는 전략을 개발해온 것이다.

고객의 취미를 건강, 골프, 쇼핑 등 16 개 분야로 분류하고, 각 영역에 대한 소비 패턴을 점수화해 개인에게 가장 적합한 혜택을 제공하는 방식을 직원들이 직접 체험하며, 개인화 마케팅의 중요성을 체감했다.

또한, 2024 년 10 월부터는 전 고객에게 동일한 광고를 노출하는 방식에서 벗어나, 각 고객에게 맞춤형 광고를 제공할 수 있는 시스템을 구축하였다.

이렇게 축적된 고객 데이터를 바탕으로, 높은 마케팅 효율을 유지하면서도 비용을 최소화하는 전략을 지속적으로 발전시키고 있다.

## 목표

### **1 차 목표: 데이터 정합성 확보 및 세그먼트 모델 초기 구축**

- iM뱅크 데이터에 포함된 주요 피처와 유사한 속성을 데이콘 데이터에서 추출
- 양측 데이터를 연계 분석 가능한 형태로 정비하고, 초기 세그먼트 예측 모델을 구축하여 데이터 기반 분류 가능성 탐색

### **2 차 목표: 고객 세그먼트 분류 모델 정교화 및 전략 도출**

- 데이콘 데이터 기반으로 소비 패턴과 고객 특성에 따른 세그먼트 분류 모델 고도화
- 각 세그먼트의 주요 특성과 소비 트렌드를 분석하여, 고객군별 맞춤형 전략 방향 제시

### **최종 목표: 정밀 고객 분류 기반 실전형 고객분류 AI 모델 완성**

- iM뱅크가 해당 모델을 기반으로 고객 맞춤형 마케팅, 상품 추천, 리스크 관리 전략을 구체화할 수 있도록 실질적 방향성 제공

**예상 데이터 출처 및 수집 방안**

본 프로젝트는 두 종류의 데이터셋을 기반으로 진행되며, 실제 금융업 적용 가능성과 일반화 가능성을 동시에 확보할 수 있는 고객 세그먼트 분류 모델을 개발하는 데 중점을 둔다.

### ① Dacon - 신용카드 고객 세그먼트 분류 AI 경진대회 데이터

#### ● 출처:

→ [Dacon 경진대회 공식 페이지](#)

#### ● 데이터 구성:

- 기간: 2018 년 7 월 ~ 2018 년 12 월
- 대상: 약 40 만 명 고객
- 분할: **train / test** (월별, 고객별 분리 구조)
- 파일 수: 총 96 개 (8 개 유형 × 6 개월 × 2 splits)
- 데이터 유형:
  - **회원정보, 신용정보, 승인매출정보, 청구입금정보, 잔액정보, 채널정보, 마케팅정보, 성과정보**
- 데이터 형식: **.parquet** → **pandas.read\_parquet()**로 로딩
- 데이터 구조 및 설명: [신용카드 고객 세그먼트 분류 AI 경진대회 데이터 명세.xlsx] 참고

#### ● 타겟 변수:

- **Segment** (성과정보 테이블 내 포함)
- 고객을 A~E 그룹으로 분류 (다중 분류 문제)

#### ● 데이터 처리 및 분석 방식 (Baseline 기준):

- 월별 **.parquet** 파일을 전처리 후 하나의 통합 데이터프레임으로 병합
- **회원정보, 신용정보, 승인매출정보** 등 주요 테이블을 기준으로 ID 와 **기준년월**을 기준으로 Merge
- 주요 파생 변수 생성:
  - 예: 승인매출 합계, 카드한도 대비 승인금액 비율, 연체 이력 유무, Life

### Stage 그룹화 등

- 범주형 변수에 대해 Label Encoding 적용
- 모델링: XGBoostClassifier 활용 (baseline 기준)
- 평가 지표: Macro F1-score (클래스 불균형 대응 포함)

#### ● 활용 목적:

- 다양한 고객 속성과 금융 행동 데이터를 기반으로 신용카드 고객을 정교하게 세분화
- 타겟 값(Segment)과의 연관성을 해석함으로써 맞춤형 마케팅 전략 기반 구축

## ② iM بانک 고객 및 카드 거래 실사용 데이터

#### ● 제공자: iM بانک

#### ● 데이터 수집 기간: 2021 년 1 월 ~ 2023 년 12 월

#### ● 데이터 파일 구성: 총 6 개 파일

- 고객 데이터: 연도별 고객 정보 파일 (3 개)
- 카드 데이터: 연도별 카드 승인 거래 정보 파일 (3 개)

#### ● 데이터 규모:

- 고객 정보: 연간 약 2 천만 건 (총 약 6 천만 건)
- 카드 승인: 연간 약 6 천만 건 (총 약 1 억 8 천만 건)

#### ● 주요 변수:

- 고객 정보: 연령대, 성별, 고객등급, 자택\_시도, 수신\_요구불예금, 대출금액 등 금융자산 속성

- 카드 정보: 승인건수, 승인금액, 가맹점 업종명, 가맹점 지역, 거래년월 등 소비 패턴 속성
- 변수 라운딩 및 범주화 처리:
  - 예) 승인건수는 5 구간 범주화, 승인금액은 최대 1 천만원 미만까지 허용

● 활용 방식:

- 고객 ID 를 기준으로 고객 속성과 카드 사용 이력을 병합
- 고객의 연령대, 금융 자산 대비 소비 규모, 업종/지역별 소비 집중도 등의 파생변수를 생성
- 고객의 소비 여력 및 연체 가능성을 고려한 세그먼트 분류 특성 정의에 활용
- ※ [참고 문서: 변수정의서.pdf]에 따라 데이터 전처리 및 해석 기준 수립

본 프로젝트는 비지도 및 지도학습 알고리즘을 조합하여 고객 세그먼트를 분류하고, 주요 인사이트를 도출하는 단계적 머신러닝 분석 구조로 설계된다.

단계	설명
1 단계. 데이터 전처리 및 통합	<ul style="list-style-type: none"><li>- 고객별 신용정보, 승인매출, 청구, 잔액, 마케팅 반응, 성과 등 테이블 통합</li><li>- 고객 ID 기준 병합, 결측치 처리, 수치형/범주형 구분 및 인코딩</li><li>- 이상치 처리 및 변수 스케일링</li></ul>
2 단계. 클러스터링 기반 잠재 세그먼트 분석	<ul style="list-style-type: none"><li>- 필요 시 K Means, GMM, 계층적 클러스터링 적용</li><li>- 군집별 특성 파악</li></ul>
3 단계. 지도학습 기반 세그먼트 분류 예측 모델 구축	<ul style="list-style-type: none"><li>- LightGBM, XGBoost, RandomForest 등 분류 알고리즘 사용</li><li>- 평가 지표: <b>F1 Score</b> 중심</li><li>- 변수 중요도 분석 및 성능 비교</li></ul>
4 단계. SHAP 기반 인사이트 도출	<ul style="list-style-type: none"><li>- SHAP 을 활용해 주요 변수 영향력 분석</li><li>- 세그먼트별 특징 파악 및 고객 이해도 향상</li></ul>
5 단계. 예측 결과 생성	제공된 테스트셋에 예측 모델 적용

## 기대 효과

### ● [DITODa 고객데이터 관리 중요한 세그먼트로 모델 확보](#) 마케팅

→ 단순한 소비 패턴 분석을 넘어, 신용도, 회원정보 등 다양한 고객 성향을 반영한 소비자평가

실질적 그룹화 가능

### [매드타임스\(MADTimes\)](#)

→ 마케팅, 리스크 관리 등 다양한 실무 영역에서 활용도 높은 고객 분류 지원

[\[단독\] KB 국민카드, AI 추천 마케팅 '적중'...고객 반응↑ 업무 시간↓ | 아주경제](#)

### ● [맞춤형 상품 설계 및 리스크 예측 지원](#)

[뱅크샐러드 마이데이터 기반 카드 추천 이용자 100 만명 돌파](#)

→ 그룹별 신용 수준 차이, 소비 여력 등을 반영한 금융 서비스 제공 가능

[빅데이터 · 코딩... 카드사 '디지털' 열풍](#)

→ 차별화된 금융 서비스 제공 및 사전 리스크 대응 전략 수립 가능

[Dacon 경진대회 공식 페이지](#)

### ● [모델 성능 및 해석력 확보](#)

→ SHAP 분석 기반으로 변수 중요도 및 영향도 시각화

→ 모델의 투명성 확보와 함께, 실행 가능한 실무 인사이트 도출 가능