AXA 지원용

데이터 분석 포트폴리오

DATA ANALYTICS PORTFOLIO

김소윤

데이터 분석가

joysoyoon.kim@gmail.com

010-4936-3150

이력사항

학력

고려대학교 – 통계학과 재학, 2023.03 - 현재

교환학생 – 싱가포르 난양공과대학교 (2025-2)

자격증

- ADSP (데이터분석 준전문가)
- TOEIC 925
- OPIC AL
- 컴퓨터활용능력 2급

핵심 역량

- 예측 모델링
- 통계적 추론
- 고객 분석 및 비즈니스 인텔리전스

목차 및 핵심 역량

핵심 가치 제안

1

비즈니스 중심 분석

측정 가능한 임팩트를 창출하는 리텐션, 리스크, 비용 절감 중심의 데이터 분석

2

도메인 전문성

헬스케어 리스크 예측, 고객 이탈 관리, 포트폴리오 리스크 분석 경험

3

기술 스택

Python/R/SQL, 머신러닝 모델, 통계 분석, 시각화 도구 활용 역량

프로젝트 구성

PROJECT 1

당뇨병 리스크 예측

헬스케어 데이터 기반 고위험군 식별 및 조기 개입 전략

PROJECT 2

고객 이탈 분석

은행 고객 이탈 예측 모델 및 리텐션 전략 수립

PROJECT 3

주식시장 리스크 분석

Copula 모델 기반 포트폴리오 상관관계 및 극단 리스크 평가

PROJECT 1

싱가포르 Healthier SG 정책 지원 - 문제 정의 및 데이터 설계

아시아인 특화 당뇨 예방 모델링을 통한 선제적 진단 시스템 구축

Context: 정책 배경

Healthier SG 정책 지원을 위한 데이터 기반 솔루션 개발

- 고령화로 인한 의료비 급증 문제
- 2030년 65세 이상 인구 23.9% 전망
- 만성질환 관리 비용 증가 대응 필요
- 데이터 기반 예방 의학 체계 구축

S$2.3B → S$9.8B

10년간 의료비 300% 증가 (2007-2017)

Data Overview

250,000개 건강검진 데이터에서 **아시아인 코호트** 추출

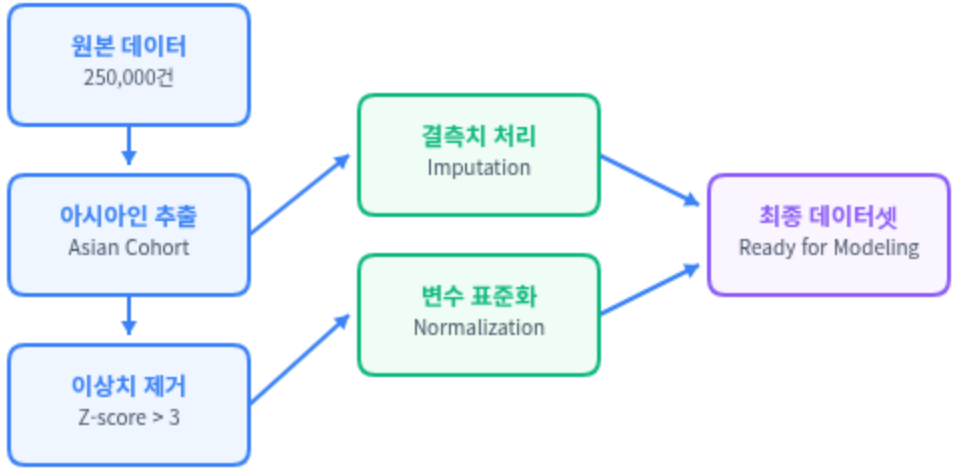
- 아시아인 특화 모델링 타겟 데이터
- 당뇨병 유병률 및 위험 요인 분석
- 불균형 데이터 처리 전략 수립

Data Strategy & Preprocessing

아시아인 코호트 추출 및 **통계적 이상치 제거** 수행

- **Z-score > 3** 기준 극단값 식별 및 제거
- 임상적 유효성을 고려한 데이터 정제
- 보건 데이터 특성 반영한 전처리
- 결측치 처리 및 변수 표준화

데이터 전처리 프로세스 플로우



PROJECT 1

이진 및 다항 분류를 통한 단계별 진단 고도화

통계 모델(Logistic)과 머신러닝(RF, CART)의 결합을 통한 예측 정교화

PHASE 1

Binary Classification

- **목표:** 당뇨 유무 판별 (Diabetic vs Non-Diabetic)
- **통계적 접근:** 로지스틱 회귀로 변수 유의성 검정 수행
- **머신러닝:** Random Forest 모델로 예측 성능 극대화

AUC 성능 **0.814**

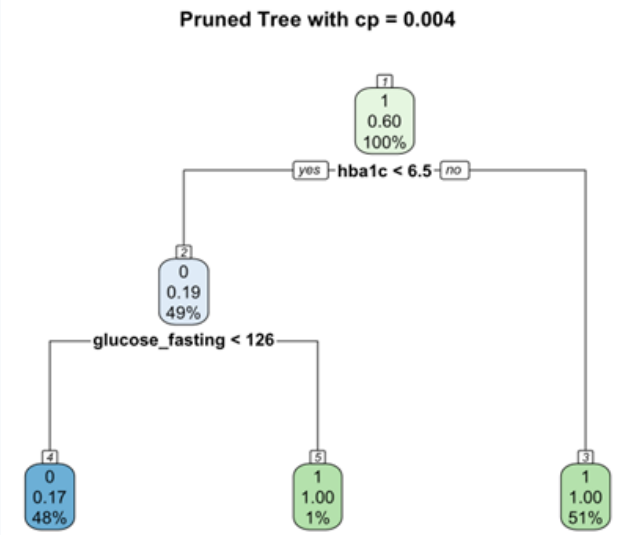
PHASE 2

Multi-class Classification

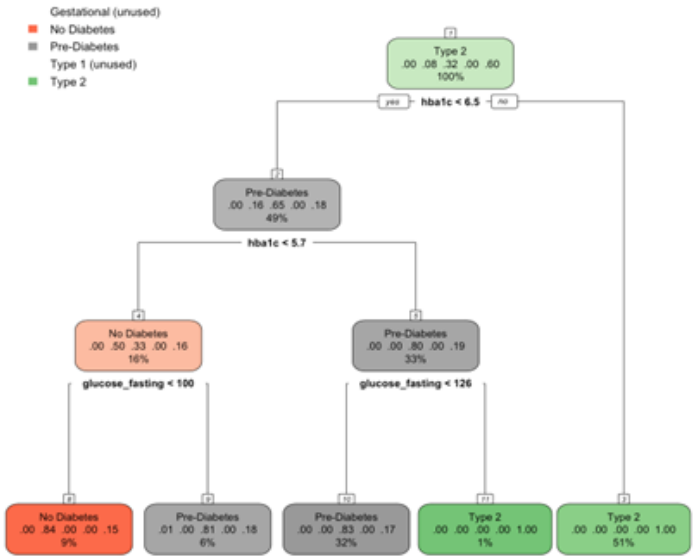
- **목표:** 당뇨 진행 단계 예측 (Normal, Pre-Diabetes, Diabetes)
- **모델링:** CART(Classification and Regression Tree) 적용
- **활용:** 위험도별 차별화된 예방 프로그램 설계 가능

Overall Accuracy **72.7%**

Pruned Decision Tree (Binary Classification)



Gestational Diabetes Classification Tree



PROJECT 2

10,000명의 고객 데이터 분석을 통한 이탈 패턴 규명 및 타겟팅 전략

문제 의식

신규 고객 유치 비용이 기존 고객 유지 비용보다 훨씬 높은 상황에서, **이탈 방지**는 은행의 수익성과 고객 생애가치(LTV) 극대화를 위한 핵심 전략입니다.

핵심 발견 (The Big One)

데이터 분석 결과, 독일 지역 비활성 고객의 이탈률이 프랑스·스페인 대비 **약 2배 높은 것**으로 확인되었습니다.

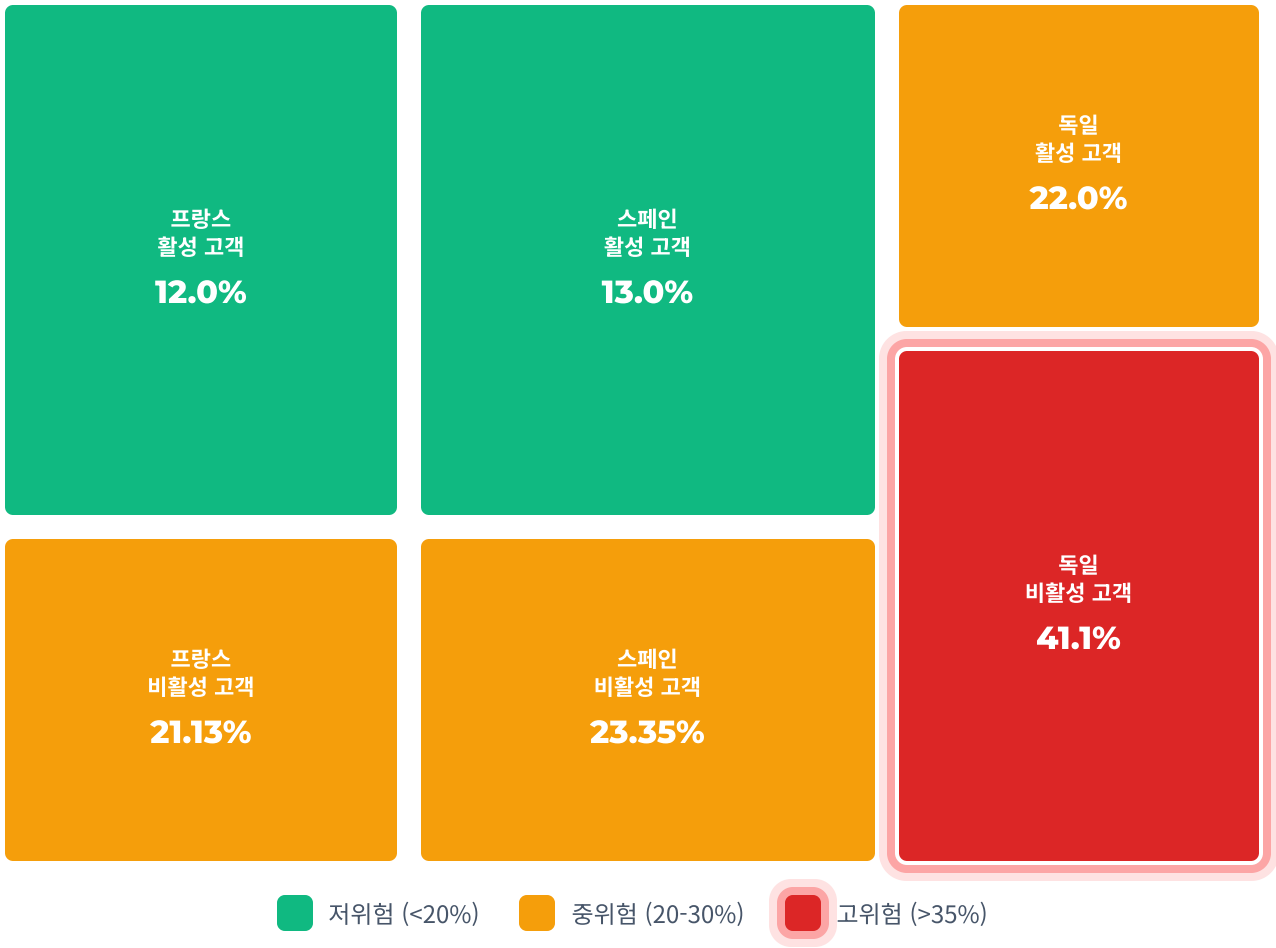
21~23%
프랑스·스페인
비활성 이탈률

→

41.1%
독일 비활성
이탈률

독일 지역의 비활성 고객 관리가 전체 이탈률 방어의 핵심 레버리지임을 확인

모자이크 플롯: 지역 × 활동 여부에 따른 이탈률 분포



PROJECT 2

이탈에 영향을 미치는 핵심 변수 도출 및 예측 모델 구축

통계적 검정 → 모델 구축 → 모델 최적화 → 성능 검증

카이제곱 검정

- 범주형 변수의 이탈 관계 검정 $p < 0.05$
- **지역:** 독일 지역이 이탈과 강한 상관관계
- **활동 여부:** 비활성 고객의 이탈 위험 유의

T-검정

- 연속형 변수의 집단 간 평균 차이 검정
- **나이:** 이탈 고객의 평균 연령이 유의하게 높음
- **잔액:** 이탈 고객의 평균 잔액이 더 많음

로지스틱 회귀 모델링

- 통계적 유의 변수 활용 예측 모델 구축

```
Exited ~ Age + Balance + IsActiveMember
```

- 다중공선성 검정 VIF ≤ 1.34
- 분석 도구: R (MASS, glm)

모델 최적화 및 성능

- 임계값 조정 0.5 → 0.3으로 Recall 개선
- "떠날 사람을 놓치지 않는 것"을 중시

AUC 성능 **0.746**

PROJECT 2

고위험군 집중 관리를 통한 이탈 방지 솔루션 (Action Plan)

PRIMARY TARGET

독일 지역의 비활성 고객

이탈률 **41.1%**

1

전환 캠페인

독일 내 비활성 고객을 대상으로 활동성을 유도하는 리워드 프로그램 시행

- 첫 거래 시 현금 리워드 또는 수수료 면제 혜택 제공
- 로그인 및 앱 활동 시 포인트 적립 프로그램
- 금융 활동성 강화를 위한 맞춤형 인센티브
- 단계적 혜택 증대로 지속적 참여 유도

2

VIP 관리 체계

잔액이 높을수록 이탈률이 증가하는 패턴에 대응하는 고잔액 고객 전담 관리

- 고잔액 고객 전담 매니저 배치 및 1:1 상담
- 거래 수수료 할인 및 우대 금리 제공
- 맞춤형 자산 관리 서비스 및 금융 컨설팅
- VIP 전용 금융 상품 및 특별 혜택 제공

3

예측 시스템 활용

로지스틱 회귀 모델을 통한 이탈 위험군 실시간 모니터링 및 선제적 대응

- 모델 기반 이탈 위험군 리스트 주간 단위 생성
- 영업팀에 고위험 고객 정보 자동 전달
- 장기 미사용 계좌 자동 리마인더 발송
- 데이터 기반 선제적 고객 이탈 방지 프로세스

PROJECT 3

데이터 및 모델링: 5개년 시계열 상관구조 분석

분석 개요: 최근 5년간의 주식 수익률 데이터를 활용하여 단순 상관계수로는 파악하기 어려운 비선형적 의존 구조를 분석합니다. 개별 주식의 분포(Marginal Distribution)를 추정한 후, Copula 함수를 통해 결합 분포를 모델링하여 섹터 간 리스크 전이 가능성을 진단합니다.

Data Scope
분석 대상 데이터

- 기간: 2019.10 ~ 2024.10 (5년)
- 주기: Weekly Adj. Close 수익률
- 목적: 장기 시계열 기반 구조적 특징 파악

Comparative Strategy
비교 분석 전략

- Within-Sector: 동조화(Synchronization) 분석
- Cross-Sector: 다변화(Diversification) 검증
- 금융 위기 시 동반 하락 위험 측정

Modeling Process
모델링 절차

- 1. Marginal 분포 추정 (각 종목별)
- 2. Probability Integral Transform (u, v 변환)
- 3. 최적 Copula 모델 파라미터 추정

Comparison Portfolios Composition

- **Within-Sector (Finance):**
  BAC(Bank of America) & JPM(JPMorgan Chase)
  → 산업군 내 강력한 상관관계 및 시스템 리스크 분석
- **Cross-Sector (Diversification):**
  BAC(Finance) & SHEL(Energy, Shell plc)
  → 섹터 간 낮은 의존성을 통한 리스크 분산 효과 검증

Applied Copula Families

- **Elliptical Copulas:**
  Gaussian (선형 상관), Student's-t (꼬리 의존성 반영)
- **Archimedean Copulas:**
  Clayton (하방 위험), Gumbel (상방 위험), Frank (대칭적 강도)
- * 각 쌍(Pair)에 대해 AIC/BIC 기준 최적 모델을 탐색하여 선정

PROJECT 3

분석 결과: 섹터 다변화의 필요성 검증

비교 포트폴리오

Within-Sector

BAC (금융) & JPM (금융)
동일 산업군 내 동조화 분석

High Correlation

Cross-Sector

BAC (금융) & SHEL (에너지)
섹터 간 리스크 분산 효과 검증

Diversification

핵심 지표

Tail Dependence (λ): 극단적 시장 상황에서의 동반 하락 확률
Correlation (ρ): 일반적 상관관계 측정

의존성 구조 분석 결과 (λ 수치 증명)

BAC & JPM (금융 섹터 내) — Best: Student's-t Copula

상관계수(ρ)가 매우 높을 뿐 아니라, **꼬리 의존성(λ) 수치가 유의미하게 높음**. 금융 위기 발생 시 두 종목이 **동시에 급락할 확률이 매우 높음**을 의미하며, Systemic Risk에 취약함.

BAC & SHEL (섹터 간) — Best: Frank Copula

Student's-t 대비 **현저히 낮은 Tail Dependence**를 보임 (Frank Copula 채택). 금융주가 급락할 때 에너지주는 **독립적으로 움직일 가능성**이 커서, 위기 상황에서 실질적인 리스크 분산이 가능함.

기술 스택 및 분석 도구

데이터 분석 및 통계 모델링을 위한 핵심 기술

프로그래밍 언어

R | dplyr | ggplot2 | tidyr | SQL | SAS | Python (기초)

주요 도구: R, SQL, SAS

통계 분석

회귀분석 | 로지스틱 회귀 | Copula 분석 | 가설검정 | 다중공선성 검정 | 유의성 검정 | VIF 분석

통계학 전공 기반 분석

데이터 시각화

ggplot2 (R) | R 시각화 | 데이터 전처리 | 탐색적 분석 | PPT 프레젠테이션

R 중심 시각화

감사합니다

김소윤

EMAIL
joysoyoon.kim@gmail.com

PHONE
010-4936-3150

포트폴리오 및 프로젝트 코드는 요청 시 제공 가능합니다.
AXA와 함께 데이터 기반 인사이트로 비즈니스 가치를 창출하고 싶습니다.