



# 포르투갈 은행 마케팅 캠페인 분석



정기 예금 가입 예측을 통한 전략 제안



# 목차



01 프로젝트 개요

02 관련 연구 소개 및 요약

---

03 데이터 소개

04 EDA 주요 인사이트

---

05 모델링 및 결과

06 전략 제안 및 결론



# ■ 프로젝트 개요

## 배경 및 접근법

- 캠페인 효율성 향상 필요
- 예금 가입 유도 전략 수립
- 데이터 분석을 통한 전략적 접근
- 2008~2010 은행 마케팅 캠페인 데이터



## 최종 목표

- 고객의 정기 예금 가입 가능성 예측
- 마케팅 담당자로서 정기 예금과 관련 있는 요소 파악
- 고객의 행동 이해
- 모델을 통해 도출한 인사이트를 바탕으로 비즈니스 전략 제시

# ■ 관련 연구 요약

## 연구 목적

- 텔레마케팅 캠페인의 성공(예금 가입)에 대한 정확한 예측
- 의사결정지원시스템(DSS) 개발
- 고객 선별을 통해 마케팅 효율성 향상 및 비용 절감 목표

## 데이터

- 08~13년 사이 포르투갈 은행의 실제 텔레마케팅 데이터 (총 52,944건)
- 성공률(예금 가입률): 12.38%
- 입력 변수: 총 150개

## 모델링 및 주요 결과

- 사용 모델: LR, DT, NN, SVM
- 평가 지표: ROC AUC, ALIFT
- 주요 결과
- 신경망(NN)이 가장 높은 성능 보임  
AUC = 0.80
- ALIFT = 0.67 (최신 데이터 기반 평가)
- 상위 50% 예측 점수 고객 대상으로 성공 고객의 79% 포착 가능

## 결론

- 기존 대비 예측 기반 타겟팅을 통해 효율성과 고객 반응률 모두 향상
- 절반의 고객에게만 전화 걸어도 약 29%p 더 많은 가입 유도 가능
- 마케팅 전략 수립 및 실행에 실질적 도움
- 향후 연구
- 입금 금액 예측, 다중 기간 분석 등

# ■ 데이터 소개

## 고객 정보 변수

- 연령, 직업
- 대출 여부
- 기타 인구통계학적 정보

## 캠페인 정보 변수

- 통화 횟수
- 캠페인 기간
- 연락 방법 (전화, 이메일 등)

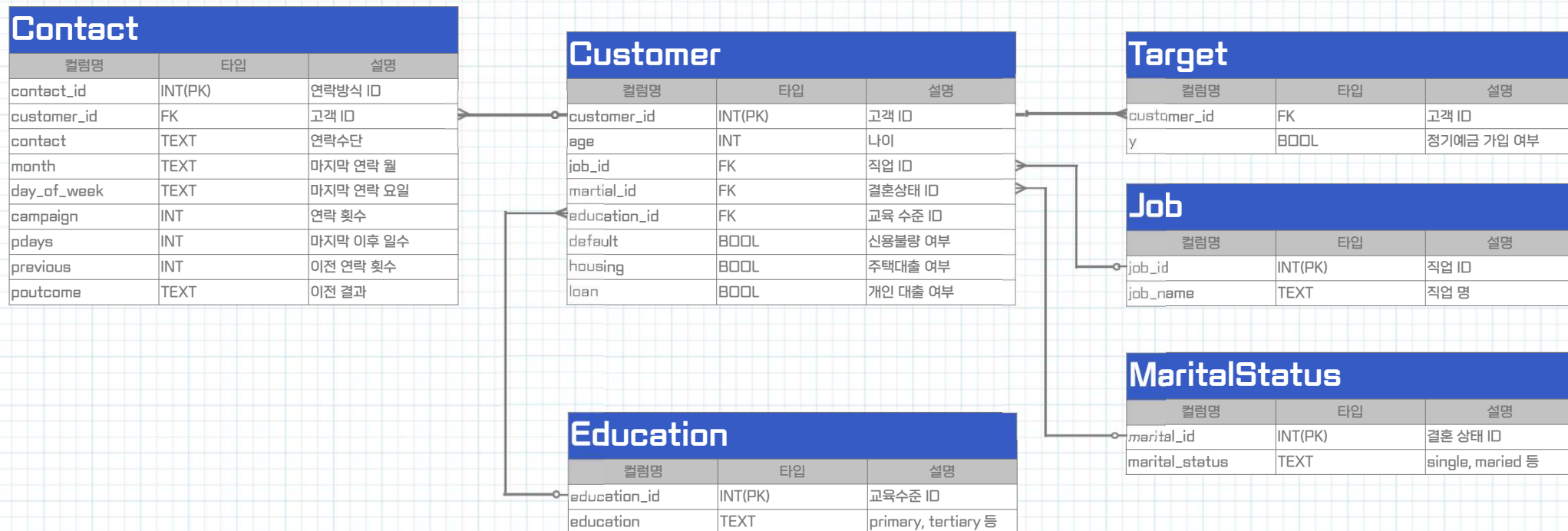
## 특성

- 고객 정보 변수  
age, job, marital, education, default, housing, loan
- 캠페인 정보 변수  
contact, month, day\_of\_week, duration, campaign, pdays, previous, poutcome
- 그 외 사회/경제 지표
- 총 150개 특성 중 선택

## 타겟 변수

- y: 정기 예금 가입 여부
- 이진 분류 문제

# 데이터 소개



# ■ EDA: 주요 인사이트 (1)

## 고객 연령별 분포

- 연령대별 가입 성향 파악
- 주요 타겟 연령층 식별
- 연령대별 마케팅 전략 수립

## 직업군별 가입률

- 직업별 예금 가입률 차이
- 고가입률 직업군 파악
- 각 직업별 접근 방식 개발

## 월별 캠페인 성과 변화

- 월별 캠페인 효과성 분석
- 성과가 높은 시기 파악
- 계절적 요인과 예금 가입 관계 조사

## 이전 통화 이력에 따른 가입률

- 이전 통화 여부에 따른 가입률 비교
- 재접촉 전략의 효과성 평가
- 고객 관계 관리의 중요성 강조

## EDA: 주요 인사이트 (2)

### Euribor 금리 영향

- Euribor 금리와 예금 가입률 관계
- 금리 변동에 따른 고객 반응 분석

### 대출 여부 영향

- 대출 고객과 비대출 고객의 예금 가입 성향 비교
- 금융 상품 간 연관성 분석
- 교차 판매 기회 발굴
- 고객 세그먼트별 맞춤 전략 수립

### 통화 기간

- 최적 통화 기간 도출
- 고객 관심도와 통화 시간 관계
- 효율적인 통화 전략 수립

### 캠페인 접촉 횟수

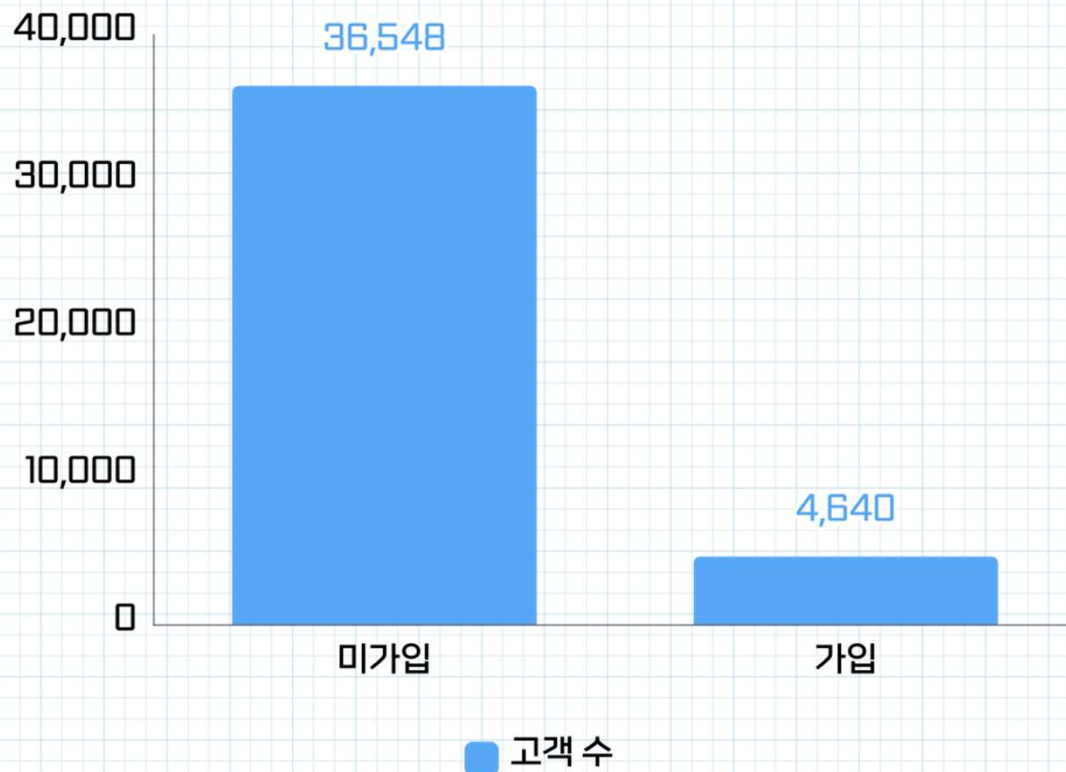
- 접촉 횟수와 가입률 관계 분석
- 최적 접촉 빈도 도출
- 고객 피로도 고려한 전략

### 시각화 방법

- 히스토그램: 연령, 소득 분포
- 박스플롯: 직업별 가입률
- 바차트: 월별 캠페인 성과



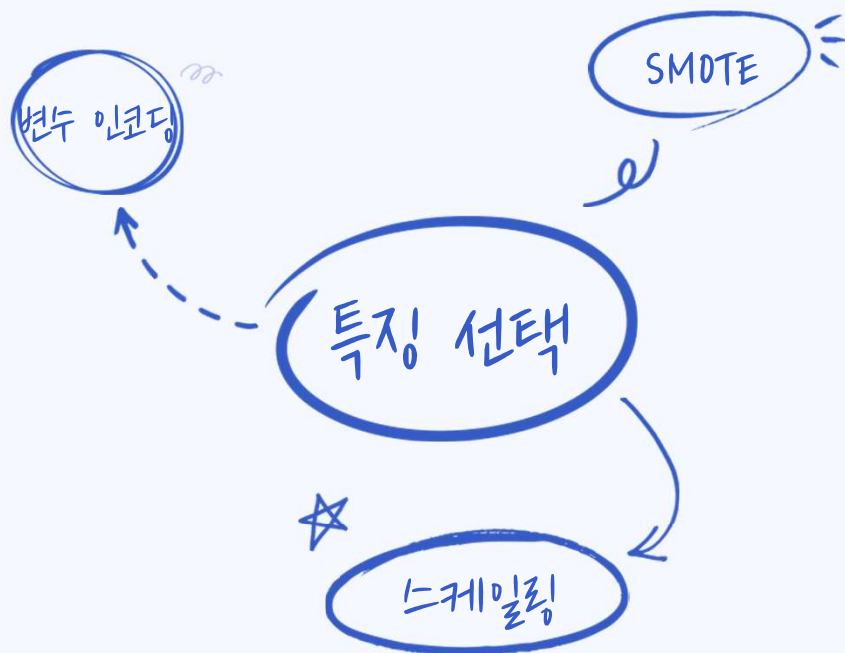
## EDA: 주요 인사이트 (3)



### 전체 가입/미가입 고객 비교

- 전체 고객 중 가입 고객은 약 11.26%로 매우 낮은 비율
- 가입 비율 늘리는 것이 급선무
- 텔레 마케팅은 요구하는 비용과 인력 리소스 ↑
- Precision과 Recall의 균형(F-1 score)  
혹은 Recall 중심 전략이 적합
- Threshold 기본값(0.5)에서 약간 하향 조정  
0.2, 0.3, 0.4 비교

# ■ 전처리 및 특징 선택



## ✓ 데이터 전처리 과정

- 불균형 데이터 처리: SMOTE 기법 적용
- 문자형 변수의 수치화: 원-핫 인코딩
- 불필요한 변수 제거로 모델 복잡성 감소
- 중요 특징 선택: 상관관계 분석, 특징 중요도 평가
- 데이터 스케일링: 표준화 또는 정규화 적용
- 결측치 처리: 평균값 또는 중앙값으로 대체
- 이상치 처리: 이상치에 민감하지 않은 모델 사용  
(Decision Tree, Random Forest)



## 사용된 머신러닝 모델과 성능 지표

본 프로젝트에서는 다양한 머신러닝 모델을 적용하여 정기 예금 가입 여부를 예측하였습니다.

주요 모델로는 Decision Tree, 랜덤 포레스트 등을 사용하였으며, 모델의 특성을 고려하여 Threshold 조정, 하이퍼파라미터 튜닝 등을 통해 최적의 성능을 도출하고자 하였습니다.

모델 성능 평가를 위해 정확도(Accuracy), F1-점수, AUC(Area Under the Curve) 등의 지표를 활용하였습니다. 이를 통해 각 모델의 예측 능력을 다각도로 분석하고, 실제 비즈니스 상황에서의 적용 가능성을 평가하였습니다.

## 피쳐 선택



### duration 컬럼 유무에 따른 모델 성능 비교

항목	duration 포함	duration 제거
Accuracy	0.92	0.90
Precision	0.71	0.57
Recall	0.48	0.29
F-1 Score	0.57	0.39
ROC AUC	0.95	0.78
TP (가입 예측 성공)	446	273
FN (가입 놓침)	482	655

# 피쳐 선택



## duration 컬럼 유무에 따른 모델 성능 비교

---

- Accuracy: 정확도 2% 하락
- Precision / Recall: 14% / 21% 하락
- ROC AUC: 17% 하락
  
- 전체적인 성능 약 17% 하락
- Precision과 Recall Trade-off 관계이지만  
duration 제거 시 동시에 하락 → Target Leakage 가능성 시사

# 피쳐 선택



## duration 컬럼 제거

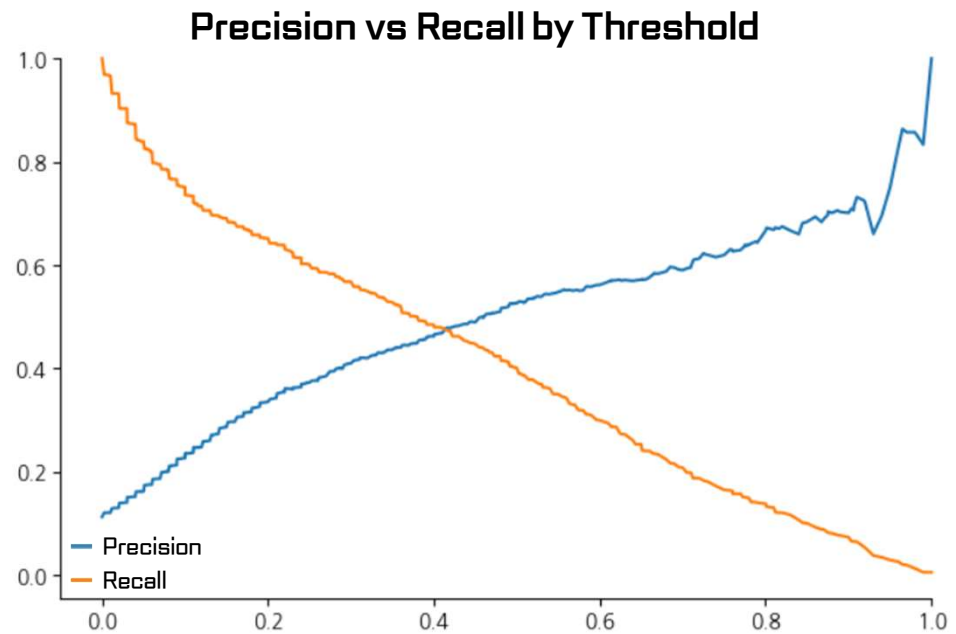
---

- duration: 마지막 통화의 길이(sec)
- 제거 이유 (Target Leakage)
  - Target Leakage: 머신러닝 모델이 예측하려는 정답(target)과 직접적으로 연결되거나 이후 발생하는 정보가 학습에 포함되어 성능이 부풀려지는 현상
- duration은 "사후적 변수"
- 고객과 통화한 결과로서의 시간 길이이므로 예측 시점에는 알 수 없음
- duration 포함 모델: 높은 성능, 실제 마케팅 전략에는 사용 불가
- duration 제거 모델: 낮은 성능, 현실적으로 타당한 예측 모델

# Threshold 조정



## Precision Recall Curve



- Precision Recall Curve: 약 0.4에서 교차
- Recall 중심 전략: 0.2 ~ 0.4 사이 조정



# Threshold 조정



## Threshold 조정

지표	Threshold=0.5 default	Threshold=0.4	Threshold=0.3 <sup>✓</sup>	Threshold=0.2
Accuracy	0.89	0.88	0.86	0.83
Precision	0.53	0.46	0.41	0.34
Recall	0.39	0.48	0.57	0.65
F-1 Score	0.45	0.47	0.48	0.44
ROC AUC	0.7849	0.7849	0.7849	0.7849

- 적당한 정확도 (0.86), 높은 Recall값 (0.57), 가장 높은 F-1 Score (0.48)
- Threshold=0.3 으로 설정 후 모델링 진행



# 최종 모델 성능 결과



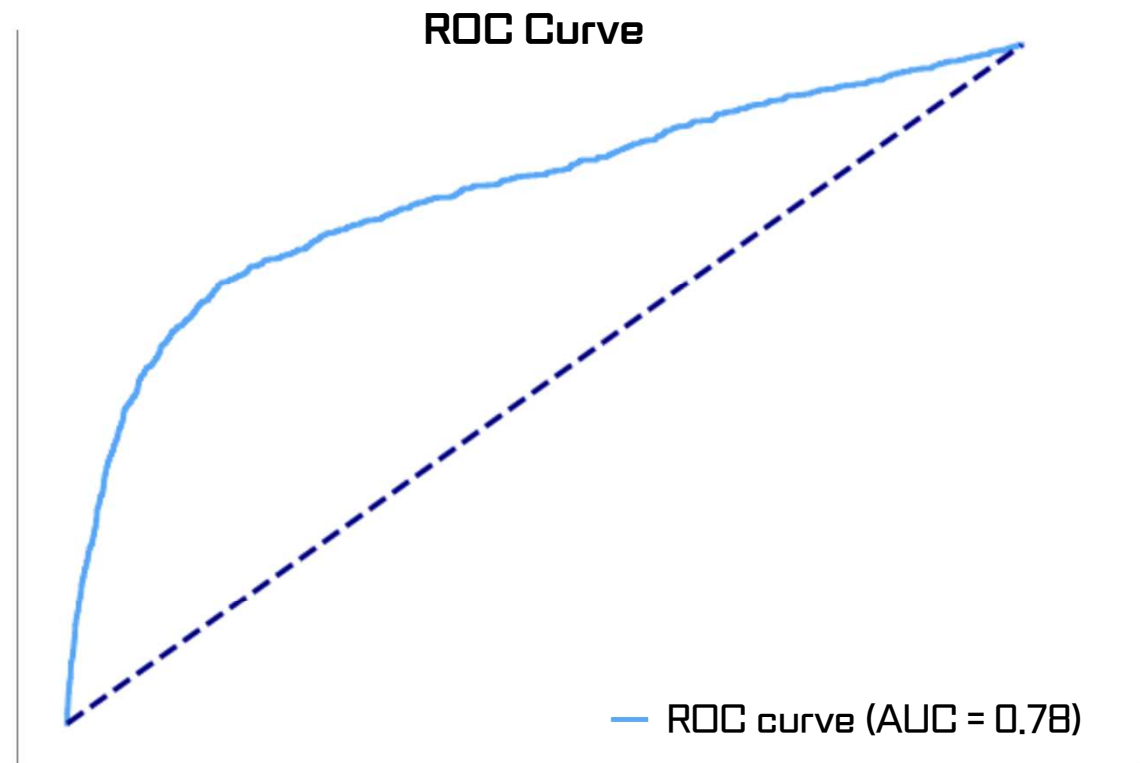
## Confusion Matrix (Threshold=0.3)

Predicted Values	Actual Values	
	Positive	Negative
Positive	6538	772
Negative	403	525

# 최종 모델 성능 결과



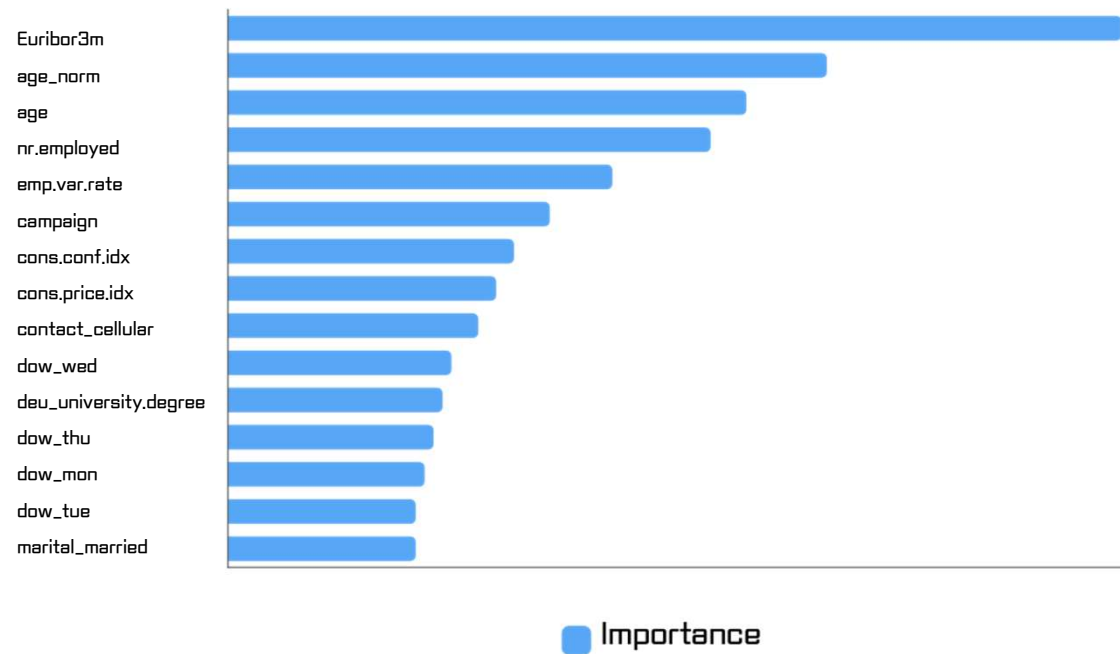
## ROC Curve (Threshold=0.3)

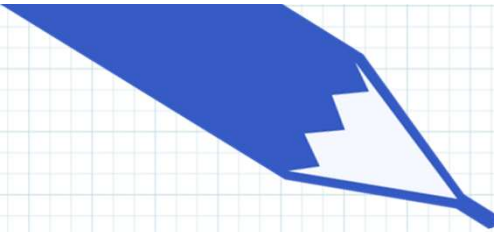


# 최종 모델 성능 결과



## Feature Importance





# 전략 제안 및 결론



## 마케팅 전략

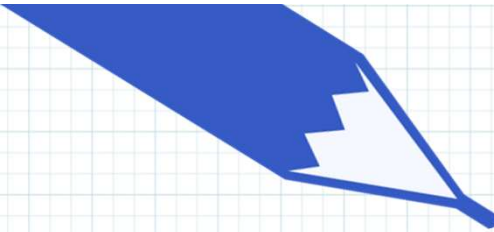
### 1. 비효율 마케팅 방지

불필요한 콜 비용 절감, 고객 불만 해소, 상담사 리소스 효율화

### 2. 마케팅 전략 차별화

Threshold 조정을 통한 공격형/방어형 마케팅 전략 수립





# 전략 제안 및 결론



마케팅 효과 분석

ROI 측정 기반 제공  
예측 모델 기반으로 성과 추적 용이





# 전략 제안 및 결론



## 고객 세분화

### 1. 예측 결과 기반 고객 등급 부여

모델 예측 결과를 바탕으로 등급을 부여하여 각 등급별 맞춤 마케팅 전략 수립

### 2. 이탈 가능성 높은 고객 사전 차단

가입 확률 낮은 고객에 대해서 적극적 마케팅 지양  
관심도 낮은 고객의 피로도 감소

