

Netflix 고객 이탈률 예측 모델 산출물

III. 학습된 인공지능 모델

I. 모델 검증 및 평가

1-1. 데이터 분할 및 검증 전략

- 최종 모델 : AdaBoostClassifier
- 학습/검증 : Stratified Split(Train:Test = 8:2) 으로 클래스 비율 유지
- 교차 검증: 5-Fold Cross Validation 기반으로 성능 안정성 확인
- 재현성 : random_state=42 고정으로 결과 재현성 확보

1-2. 평가 지표

- Accuracy, Precision, Recall, F1-score, ROC-AUC 사용
- Overfitting Index = Train Accuracy – Test Accuracy (과적합 추정치)
- Cross-validation 표준편차로 모델 일관성 평가

1-3. 정량 결과

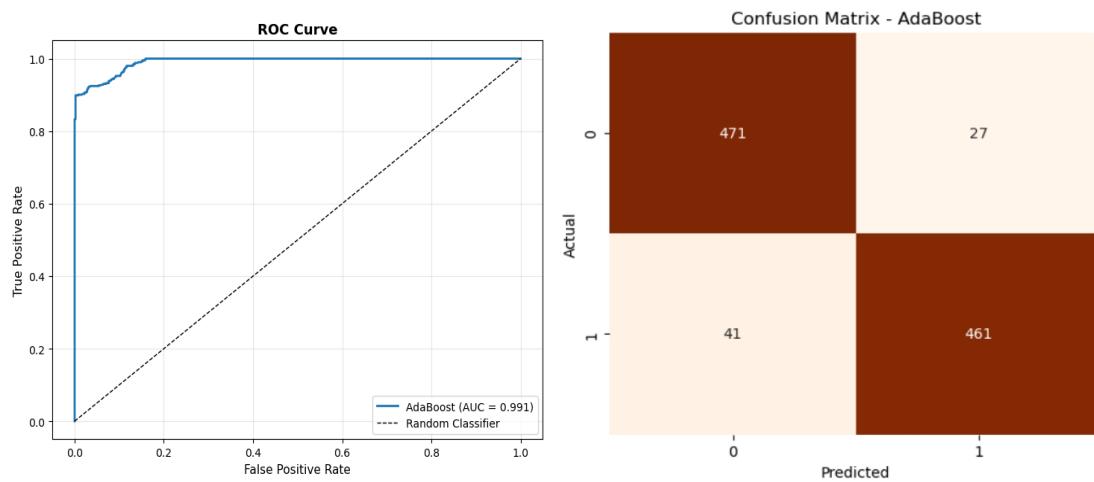
- Train Accuracy : 0.9487
- Test Accuracy: 0.9370
- Overfitting : 0.0412 → 0.0117
- Cross-val std : ±0.008
- Test Classification Report:

	precision	recall	f1-score	support
Not Churned	0.91	0.97	0.94	498
Churned	0.97	0.90	0.93	502
accuracy			0.94	1000
macro avg	0.94	0.94	0.94	1000
weighted avg	0.94	0.94	0.94	1000

1-4. 성능 비교 및 해석

- AdaBoost 모델은 기존 단일 모델(SVC 79.2%) 대비 명확한 성능 향상을 보임.

모델명	Accuracy	F1-Score	Overfitting	특징
KNN	75.5%	0.75	-	정확도 낮음
SVC	77.3%	0.77	-	정확도 낮음
Logistic Regression	79.2%	0.79	-	정확도 낮음
Random Forest	92.7%	0.91	0.041	베이스라인, 과적합 위험 있음
Voting (Soft)	93.2%	0.93	0.020	안정적이나 향상폭 제한
AdaBoost (최종)	93.7%	0.94	0.0117	가장 안정적이며 일반화 우수



II. 모델 개선 및 최적화

1-1. 하이퍼파라미터 튜닝

- GridSearchCV 를 활용하여 주요 하이퍼파라미터 최적화.

학습률(learning_rate)과 트리 깊이(max_depth)를 조정해 오버피팅을 최소화하고, 일반화 성능을 개선.

파라미터	튜닝 전	튜닝 후
learning_rate	0.1	0.05
max_depth	3	4
n_estimators	100	250

- 튜닝 결과 정확도 0.9320 → 0.9370 으로 향상
 - 오버피팅 지수는 0.0412 → 0.0117 로 감소
 - 성능 최적화와 함께 모델의 안정성 강화

```

=====
# 1. AdaBoost Hyper Parameter Tuning
=====

print("+" * 80)
print("1. AdaBoost Hyper Parameter Tuning")
print("+" * 80)

# AdaBoost 파라미터 그리드
adaBoost_param_grid = {
    'n_estimators': [10, 150, 200, 250, 300],
    'learning_rate': [0.05, 0.1, 0.15, 0.2],
    'estimator_max_depth': [2, 3, 4]
}

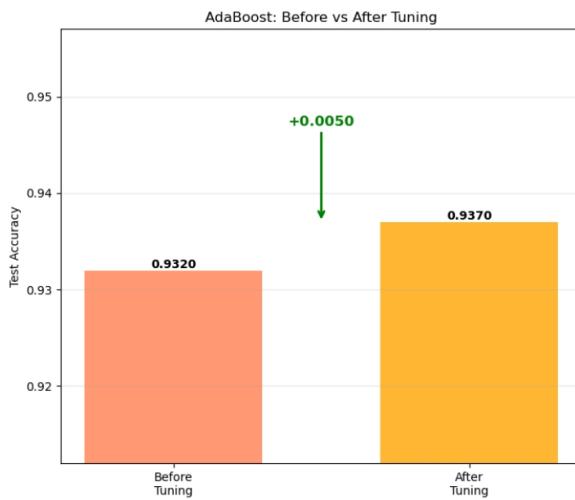
# GridSearchCV with tdm
start_time = time.time()

cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

# GridSearchCV 설정
scoring = {
    'accuracy': make_scorer(accuracy_score),
    'f1': make_scorer(f1_score),
    'roc_auc': 'roc_auc'
}

adaBoost_grid = GridSearchCV(
    AdaBoostClassifier(),
    estimator=DecisionTreeClassifier(random_state=42, class_weight='balanced', criterion='gini', splitter='best'),
    random_state=42,
    param_grid=adaBoost_param_grid,
    cv=cv,
    refit='accuracy', # accuracy를 기준으로 최적 모델 선택
    verbose=1,
    n_jobs=-1,
    scoring=scoring
)

```



III. 모델 설명 및 해석

1-1. AdaBoost 의 feature_importances Top5

순위	변수명	영향도(%)	해석
1	watch_hours	43	시청 시간이 많을수록 이탈률 감소
2	last_login_days	33	로그인 간격이 길수록 이탈률 증가
3	number_of_profiles	8	프로필 갯수가 적을수록 이탈률 증가
4	subscription_type	6	구독타입에 따른 이탈률 변화
5	age	5	나이차이에 따른 이탈률 변화

2-2. 모델 결과 해석 및 시사점

- 모델은 시청 행동(watch_hours) 과 활동성(last_login_days) 을 핵심 지표로 학습하였으며, 이는 실제 고객 유지 전략과 직결되는 요인임.
- 요금제(subscription_type) 변수의 중요도는 구독 구조 및 만족도가 이탈률에 직접적 영향을 미침을 의미함.
- 지역(region), 선호 장르(favorite_genre) 등은 세그먼트 기반 마케팅(지역·취향별 맞춤 캠페인) 기획에 활용 가능함.

IV. 인공지능 실제 적용 가능성 및 확장성

1-1. 실제 적용 가능성

- 본 모델은 고객의 이탈(Churn) 여부를 사전에 예측하여, 기업의 마케팅 및 고객 관리 전략 수립에 즉시 활용 가능한 수준의 실무적 완성도를 확보하였음.
- 예측 결과는 CRM 시스템 및 마케팅 자동화 플랫폼과 직접 연동되어 고객 유지율 제고 및 비용 효율화에 기여할 수 있음

① 고객 이탈 조기 경보 시스템

- 시청시간 및 로그인 패턴 변화가 감지될 경우, 이탈 위험 고객을 실시간으로 탐지 가능
- 예측 결과를 기반으로 고객 상담센터 및 유지 캠페인과 자동 연계하여 대응 효율을 향상시킴

② 맞춤형 요금제 추천 엔진

- KNN 기반 유사 고객 분석을 통해, 이탈 확률이 높은 고객에게 최적의 구독제(Basic → Standard, Premium) 추천 가능
- 고객 세그먼트별 개인화된 프로모션·콘텐츠 제공으로 만족도 향상

③ 고객 유지 캠페인 자동화

- 예측된 이탈 위험군을 대상으로 맞춤형 프로모션, 할인 쿠폰, 개인화 콘텐츠 등을 자동 발송 가능
- CRM(Customer Relationship Management) 시스템과 연동하여 실시간 고객 상태 모니터링 가능

④ 마케팅 효율화 및 비용 절감 효과

- 이탈 고위험군(상위 10%)만 집중 관리하여 마케팅 비용을 효율적으로 배분
- 재유입 비용 대비 이탈 예방 비용이 낮다는 점을 고려할 때, 본 모델은 ROI(투자 대비 효율) 극대화에 기여

⑤ 고객 생애가치(LTV) 기반 자원 최적화

- 유지비용과 이탈비용을 분석하여 VIP 고객 유지에 우선순위를 둘 수 있음
- 장기 고객 확보 및 구독 매출 안정화에 기여

🎯 맞춤 플랜 찾기

AI 분석 시스템이 고객 이탈 위험도를 예측하고, 최적의 구독 타입을 추천해드립니다.

고객 정보 입력

나이	30	월 평균 시청 시간	1.0
최종 로그인 경과일	28	사용하는 프로필 수	4
성별	기타	현재 구독 타입	스탠다드
지역	주요 접속 디바이스	선택 항목	
아프리카	TV	액션	

🔥 이탈 위험도 예측 및 최적 플랜 찾기

이탈 위험도 분석 결과

고객님의 이탈 예측 결과는 **✖️ 높음 (이탈 위험)**입니다.

이탈 예측 결과	✖️ 높음 (이탈 위험)	이탈 확률	64.97%
----------	---------------	-------	--------

구독 타입 추천

⭐️ 추천 구독 타입

Premium

월 17,000원

IV. 결론 및 요약

- AdaBoost 기반의 고객 이탈 예측 모델
 - 정확도(Accuracy) 93.7%, F1-score 0.94로 기존 단일 모델 대비 성능이 크게 향상됨
 - Overfitting 지수 0.0117로 과적합 위험이 매우 낮아 안정적이고 일반화된 예측 성능을 확보하였음.
- 하이퍼파라미터 최적화(GridSearchCV) 및 교차 검증(5-Fold)을 통해 모델의 신뢰성과 재현성을 강화함.
- 주요 영향 변수로는 시청시간(watch_hours), 최근 로그인 일수(last_login_days) 등이 도출되어 고객의 이용 행동 패턴이 이탈 여부에 직접적인 영향을 미침을 확인하였음.
- 본 모델은 성능·안정성·해석력을 모두 갖춘 고객 이탈 예측용 실무형 인공지능 모델로 평가됨.
- CRM 연계 및 마케팅 자동화 시스템에 즉시 적용 가능한 수준의 완성도를 확보하였음.
- Netflix 등 구독형 서비스의 고객 이탈 예측, 개인화된 추천, 마케팅 자동화 시스템에 즉시 활용 가능.
- 향후 시청 로그, 콘텐츠 선호도, 고객 지원 이력 등 정성 데이터를 포함한 딥러닝 기반 하이브리드 모델(CNN+LSTM)로 확장 가능성을 보유.