2025.07.23

광운대학교
2025 광운회사 최종보고서

Customer Churn Analysis in Telecommunication

고객 이탈 예측 모델 개발

광운대학교 국제통상학부
이상민

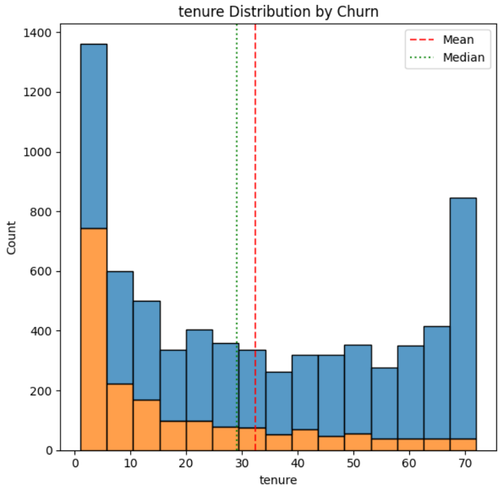
목차

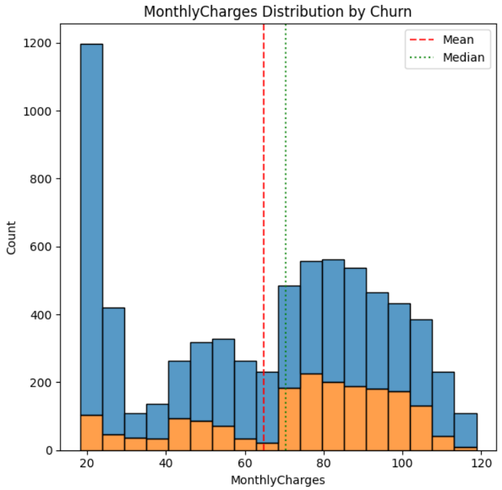
CONTENTS

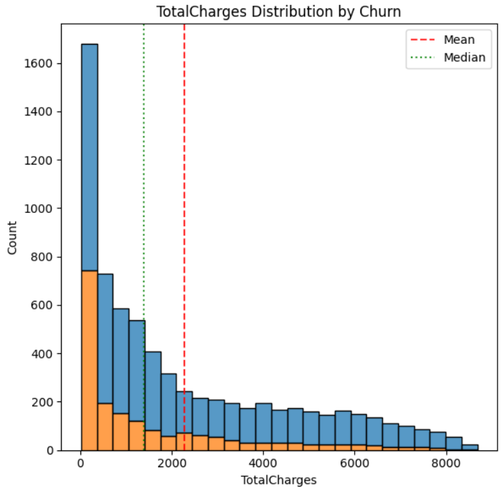
01 분석과제정의서

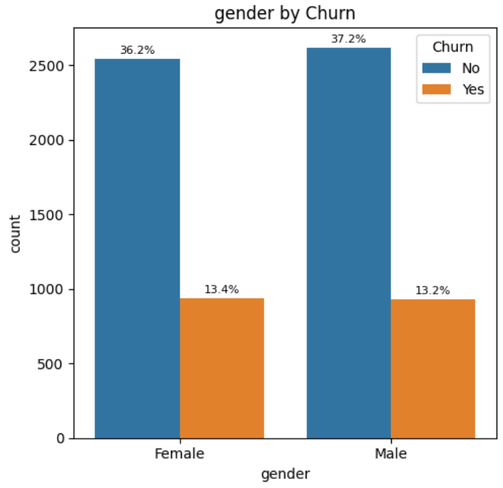
과제명	통신사 고객 이탈 예측 모델 구축을 통한 고객 이탈율 5% 저감			
과제 목표	• 고객의 다양한 속성(인구통계, 이용행태 등)을 반영한 이탈 예측 모델 구축 • 이탈 고위험 고객군을 사전 식별하여 맞춤형 유지 전략 설계 • 예측 모델 기반 관리 체계를 통해 이탈율 5%p 저감			
분석결과활용 프로세스	1. 고객 데이터 수집 및 전처리 : 기본 정보, 요금제, 서비스 이용 패턴 등 2. 이탈 예측 모델 학습 및 평가: 머신러닝 기반 분류모델 활용 (XGBoost, Random Forest등) 3. 이탈 확률이 높은 고객 분류 및 등급화 4. 마케팅/고객센터 연계 프로모션 자동화 적용			
현황	• 자사 유무선 가입자 기준 연간 고객 이탈율은 평균 18.9% • 이탈 고객의 72%가 계약 종료 1개월 이내 동일 요금제 사용 중지		• 이탈 이후 재가입 유도 마케팅 비용이 고객 1인당 평균 4.1만원 소요 • 이탈 고객 분석은 VOC, 클레임 등 정성적 설문 기반에 머무름	
문제점	• 이탈 사전 예측체계 부재로 고위험 고객을 조기 식별하지 못함 • 전 고객 대상 일괄 프로모션으로 비용 대비 효과 낮음으로 마케팅 자원 낭비		• 이탈 요인을 정량적으로 파악한 구조 부재로 데이터 기반 관리가 미흡 • 이탈 관리 성과에 대한 KPI 미정의 및 추적 어려움	
분석주요내용	**EDA** 고객 속성별 이탈률 비교, 변수 간 상관관계 확인	**Feature Engineering** 파생 변수 생성 (고객 충성도 지수, 요금제 변경 이력 등)	**모델링 및 평가** 머신러닝 다수 모델별 비교 이탈률, 저감률 등 성과 판단	**해석** SHAP 분석을 통한 주요 이탈 요인 도출 및 고객 특성별 설명 가능성 확보
기대효과	• 고객 이탈률 18.9%에서 13.9%로 감소 • 고객 1인당 평균 4.1만원 재유치 비용 절감 기대		• 이탈 고위험군 대상 조기 대응으로 연간 3만명 이상 고객 유지 효과 • 고객 유지 전략에 활용가능한 지속적 예측, 진단 체계 마련	

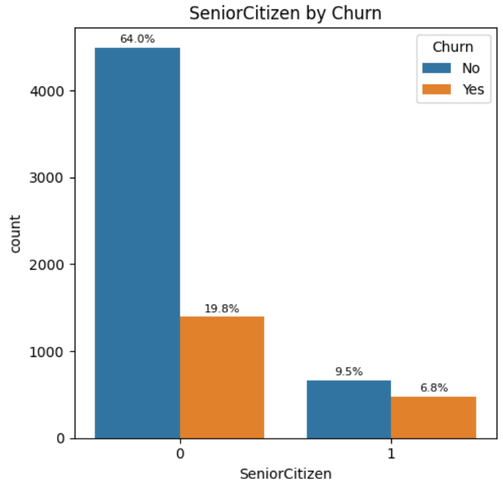
<주요 변수 분포>

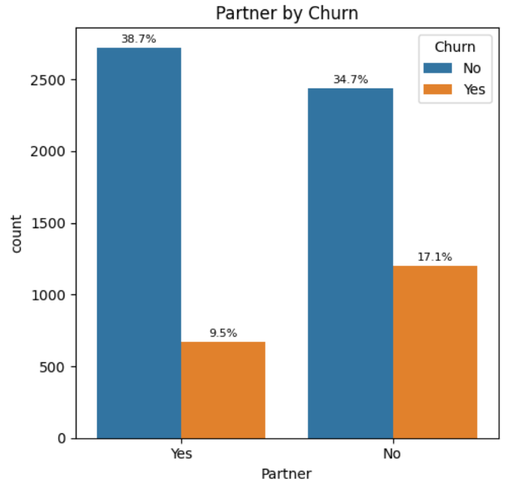

tenure Distribution by Churn
Mean
Median
Count
tenure

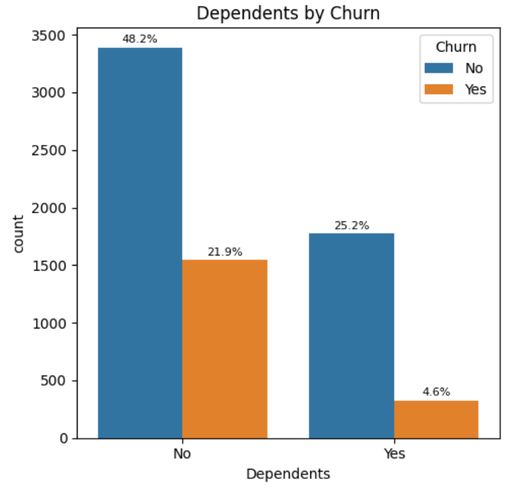

MonthlyCharges Distribution by Churn
Mean
Median
Count
MonthlyCharges


TotalCharges Distribution by Churn
Mean
Median
Count
TotalCharges

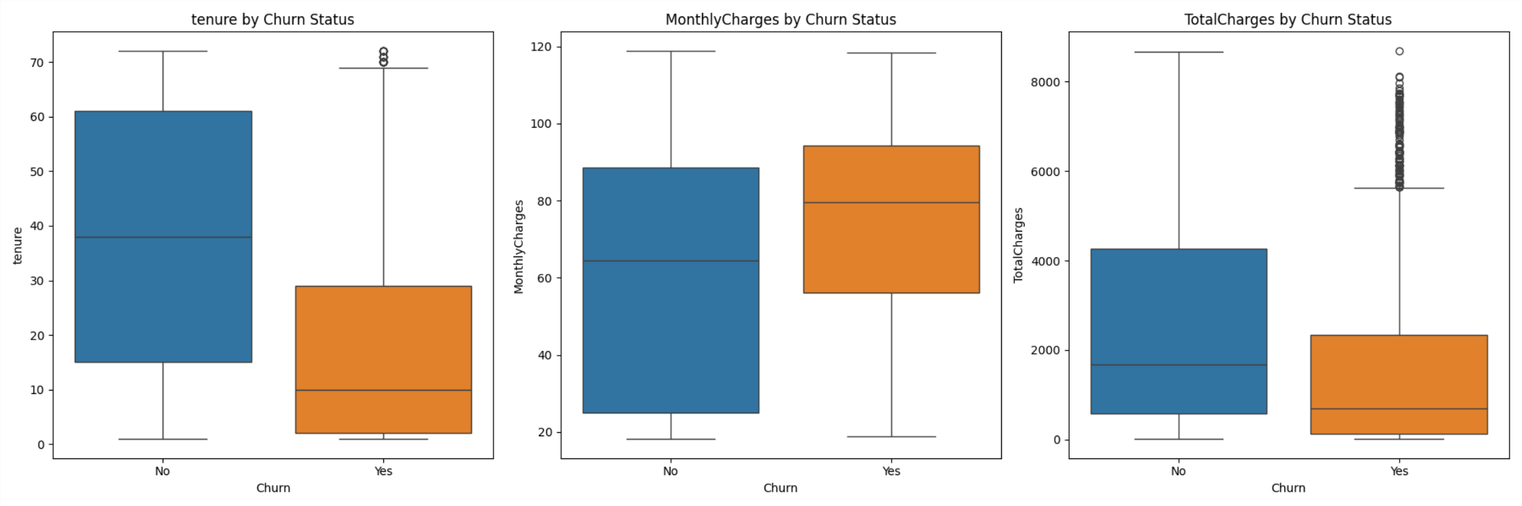

gender by Churn
Churn
No
Yes
36.2%
13.4%
37.2%
13.2%
count
Female
Male
gender


SeniorCitizen by Churn
Churn
No
Yes
64.0%
19.8%
9.5%
6.8%
count
0
1
SeniorCitizen

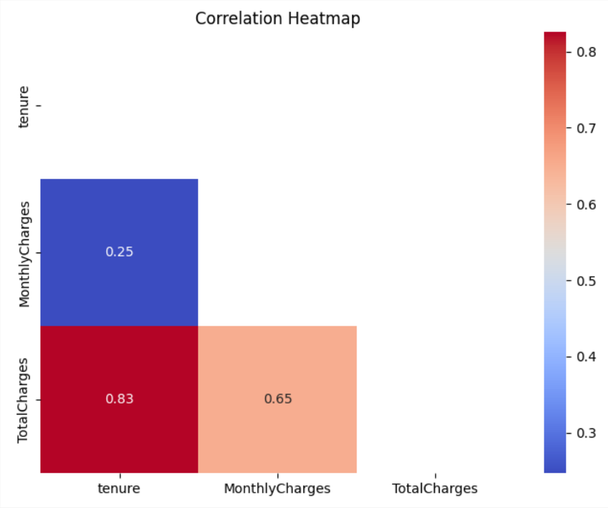

Partner by Churn
Churn
No
Yes
38.7%
9.5%
34.7%
17.1%
count
Yes
No
Partner


Dependents by Churn
Churn
No
Yes
48.2%
21.9%
25.2%
4.6%
count
No
Yes
Dependents

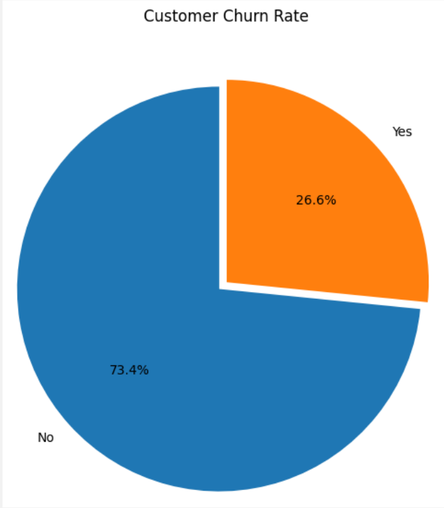
<주요 변수 분포>



<수치형 변수 상관 계수>



<고객 이탈 비율>

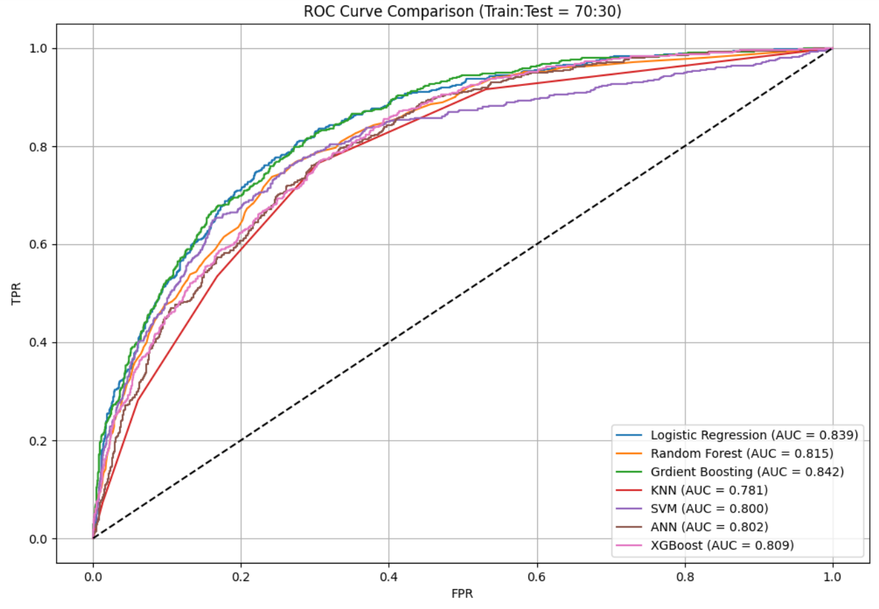


분석 모델 개발 전 EDA를 통한 인사이트에서
단기 계약(Month-to-Month) 고객, 월 요금이 높은 고객,
가입 초기 고객, Fiber Optic 서비스를 이용하는 고객이
이탈 할 가능성이 높은 것으로 나타남

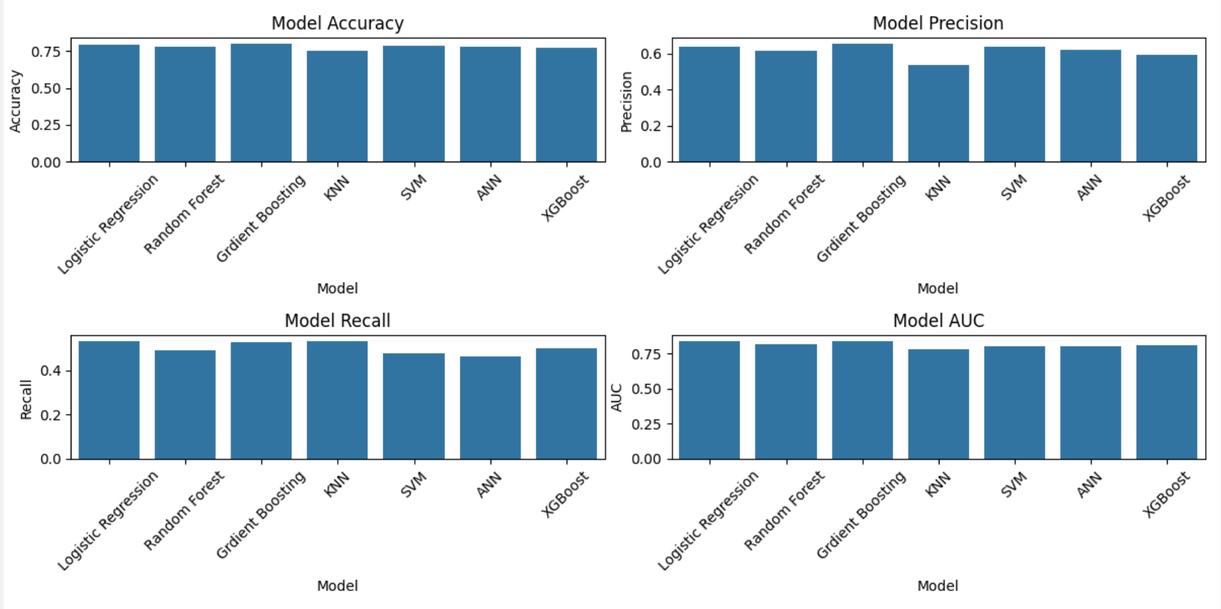
또한, 시니어 고객과 전자결제 방식 고객 등
일부 특성군에서도 이탈률이 다소 높게 나타났으며,
이는 고객군별 맞춤 전략의 필요성을 보여줌.

이러한 결과는 단순히 이탈을 예측하는데 그치지 않고,
실제 마케팅 전략 및 고객 관리 정책 수립에 활용될 수 있는
실질적인 인사이트를 제공함.

03 모델 비교 및 성능 평가



<ROC Curve>



<지표별 모델 비교>

- Y 변수: Churn (이탈 여부: Yes / No)
- 데이터 분할: 훈련 30% / 테스트 70%
- 비교 모델: Logistic Regression, Random Forest, Gradient Boosting, KNN, SVM, ANN, XGBoost
- 비교 지표: Accuracy, Precision, Recall, F1 Score

04 성능 향상 방향

<비교를 위한 하이퍼파라미터>

- 트리 개수 (n_estimators): 100, 200
- 학습률 (learning_rate): 0.05, 0.1, 0.2
- 트리 깊이 (max_depth): 3, 4, 5

<GridSearchCV를 활용한 최적 하이퍼파라미터>

- 트리 개수: 100
- 학습률: 0.1
- 트리 깊이: 3

<기본 모델 성능>

	Precision	Recall	F1-Score	Support
유지	0.83	0.90	0.86	1549
탈퇴	0.64	0.50	0.56	561

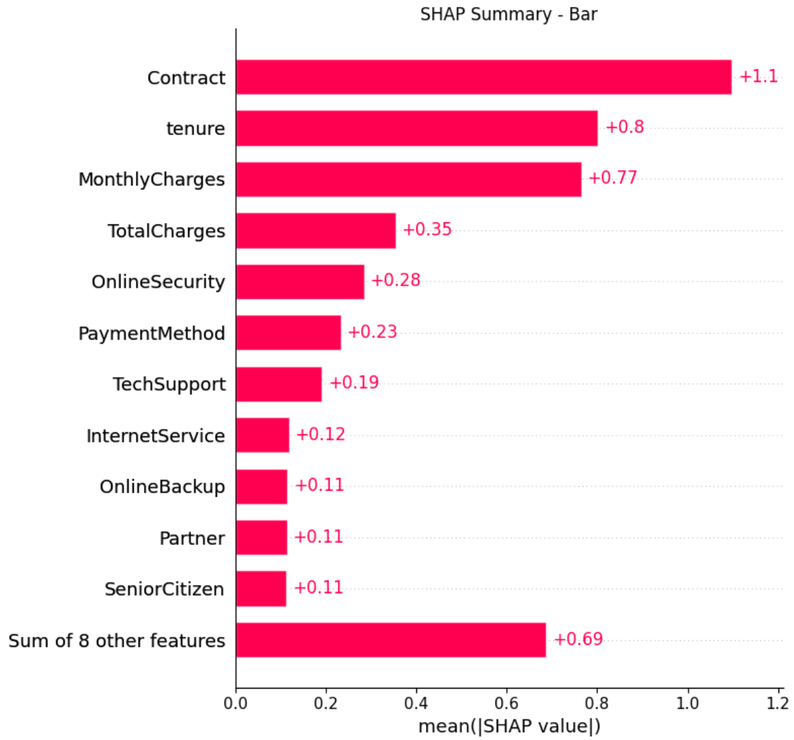
<최적 모델 성능 (Train Set)>

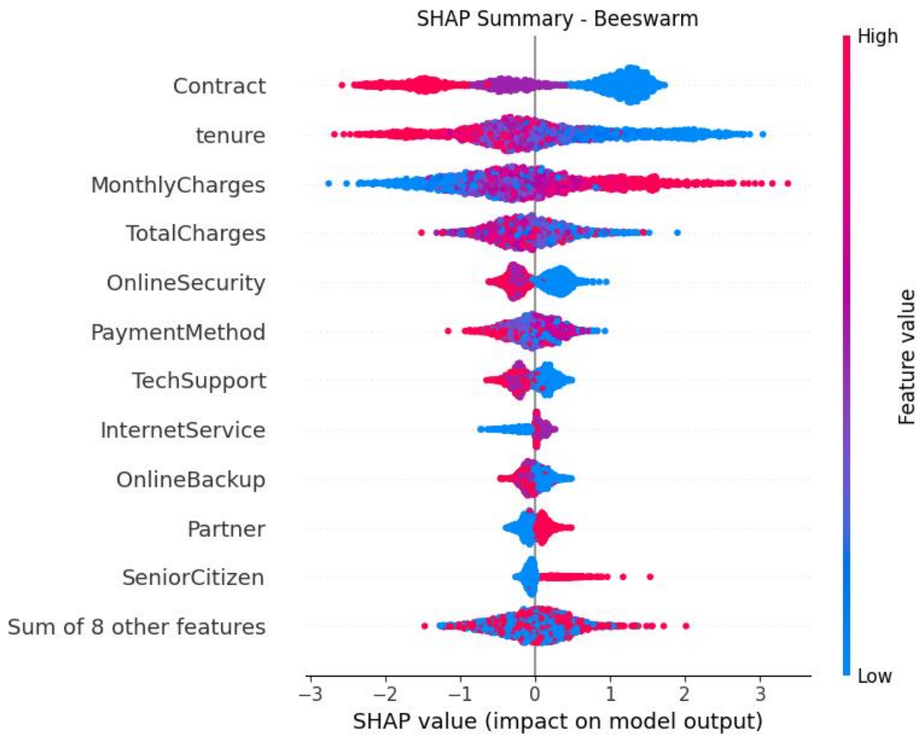
	Precision	Recall	F1-Score	Support
유지	0.88	0.93	0.90	1549
탈퇴	0.77	0.63	0.69	561

<최적 모델 성능 (Test Set)>

	Precision	Recall	F1-Score	Support
유지	0.86	0.92	0.89	3614
탈퇴	0.73	0.58	0.65	1308

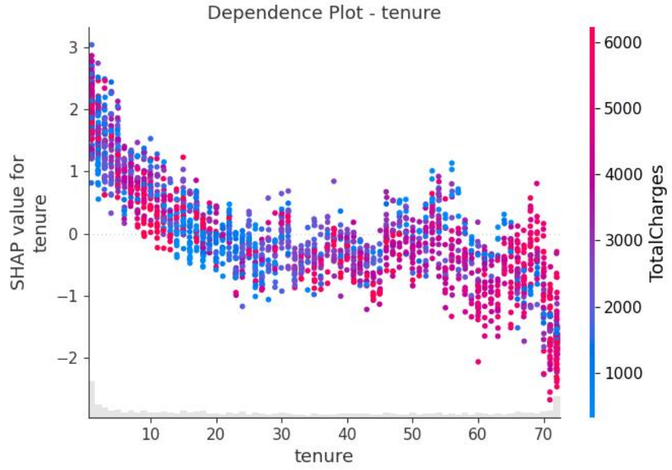
05 결과 해석

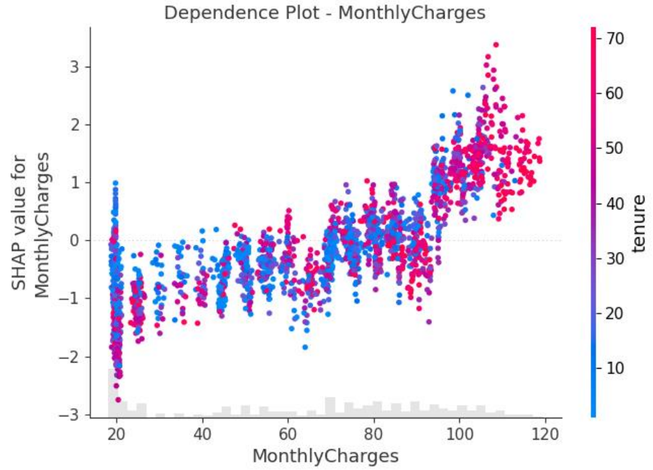


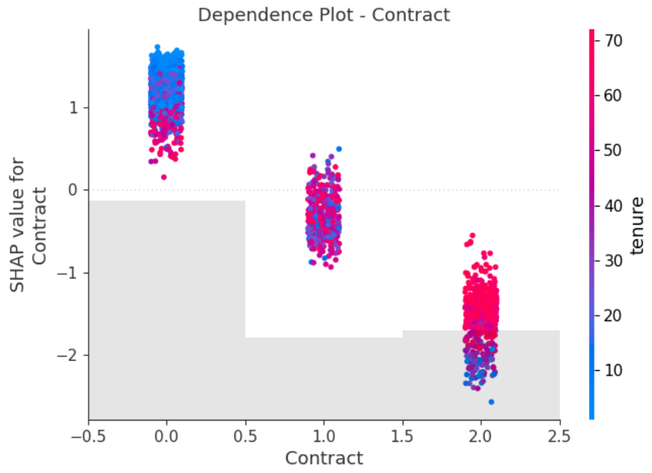


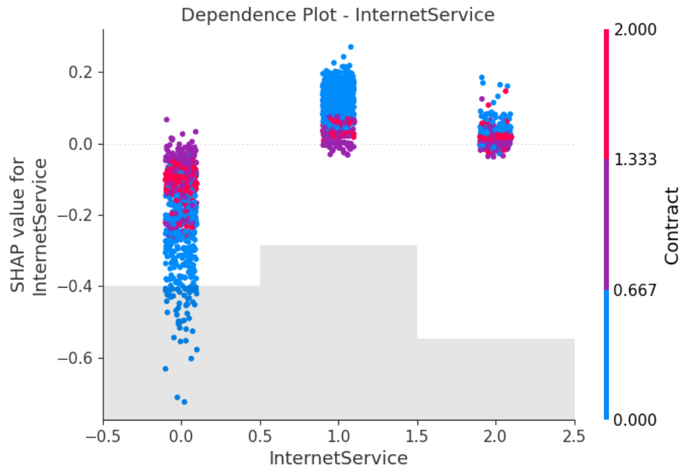
- SHAP (SHapley Additive exPlanations) 기법을 활용하여, 각 특성(feature)이 예측에 미치는 영향을 정량적으로 분석함.
- SHAP Bar Plot: 계약 유형, 계약 기간(tenure), 월 요금(Monthly Charges) 등 주요 특성들이 이탈 예측에 큰 영향을 미치는 것으로 분석됨
- Beeswarm Plot: 계약 유형이 단기 일수록, 가입 기간이 짧을 수록 이탈 가능성 증가에 기여

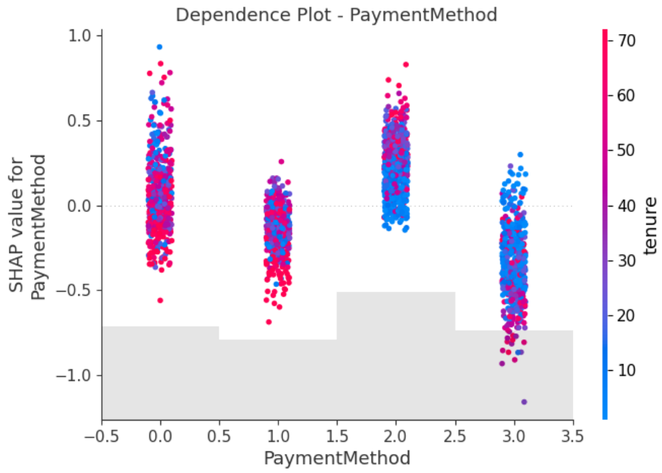
05 결과 해석











- 특정 변수 수준에 따른 예측 기여도 분석: 요금 및 계약 유형에 따라 SHAP 값 분포 차이 확인됨

06 결론 및 향후 개선 방향

<결과>

이번 분석은 Telco 고객 데이터를 기반으로 Gradient Boosting 모델을 활용하여 고객 이탈 여부를 예측하고,
SHAP 해석을 통해 예측 결과에 영향을 미친 주요 요인을 분석함.
모델 학습 결과, Contract(계약 유형), tenure(가입 기간), Monthly Charges(월 요금) 변수가
이탈 가능성에 주요한 기여를 하는 것으로 나타났으며, 단기 계약자, 가입 초기 고객, 월 요금이 높은 고객에서
이탈 확률이 상대적으로 높게 확인됨.
SHAP 기법을 통해 예측의 설명력을 확보함으로써, 모델의 투명성을 높이고 실무 적용 가능성을 확장할 수 있는 기반을 마련함

<향후 개선 방향>

1. 비즈니스 연계 변수 추가 필요성
   - 고객의 서비스 사용 행태, 최근 문의 및 민원 기록 등 시계열 기반의 행동 데이터를 추가적으로 수집 및 반영함으로써 예측 정확도 높임

2. 하이퍼파라미터 최적화 범위 확대
   - Learning_rate, max_depth, subsample 등 핵심 파라미터에 대해 GridSearchCV, RandomizedSearchCV, Bayesian Optimization등의 기법을 활용한 보다 정교한 최적화가 요구됨

3. 이탈 유형 구분 분석으로의 확장
   - 단순한 이진 분류를 넘어, 이탈 사유(요금 불만, 품질 불만, 경쟁사 이동 등)를 다중 클래스 분류로 예측함으로써 정밀한 맞춤 대응 전략 수립 가능

4. 시간 기반 예측 분석 적용 필요
   - 고객이 언제 이탈할지를 예측하는 생존 분석(Survival Analysis)이나 시계열 기반 접근을 통해 마케팅 및 고객 유지 전략을 선제적으로 계획

5. 다양한 모델 간 비교 및 앙상블 적용
   - LightGBM, CatBoost, Random Forest 등의 모델을 병렬 적용 및 성능 비교함으로써, 정확도와 해석력, 학습 속도 간의 균형 도출 가능