

PROJECT / 머신러닝

1.이탈 예측 모델링을 통한 고객 이탈 최소화 및 내부 고객 관리 강화 전략 수립

프로젝트 개요

기획 의도	엔데믹 영향과 고물가로 인해,이커머스 소비시장이 분산되고, 최저가 경쟁의 이유등으로 기업들은 경쟁적인 상황에 직면하고 있다. ⇒ 이로 인해, 고객 이탈율 상승 및 내부 고객 관리의 필요성 대두
문제 정의	1.고객 특성별 세분화된 마케팅 전략 수립 및 이탈 예측 모델링이 필요하다. 2.도시등급 ,결제 방식 등의 데이터를 활용하여, 비즈니스 전략 수립을 통해 장기적인 가치 창출이 필요하다.
분석 목표	1.고객 이탈 예측 모델 개발 2.이탈에 영향을 주는 특성 확인 3.내부 고객 관리 강화 및 이탈 위험 고객 관리 액션 플랜 도출
Language	Python
Tool	Pandas

프로세스



데이터

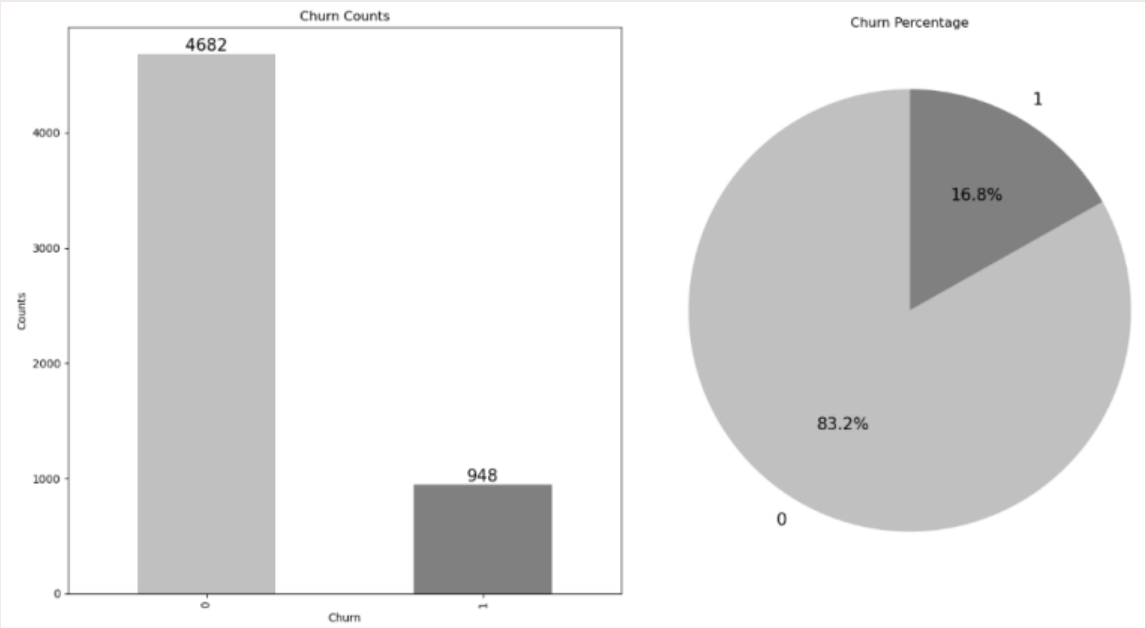
데이터 설명	
데이터 이름	Ecommerce Customer Churn Analysis and Prediction
데이터 출처	Kaggle
데이터 사이즈	5630 rows x 20 columns

데이터 전처리

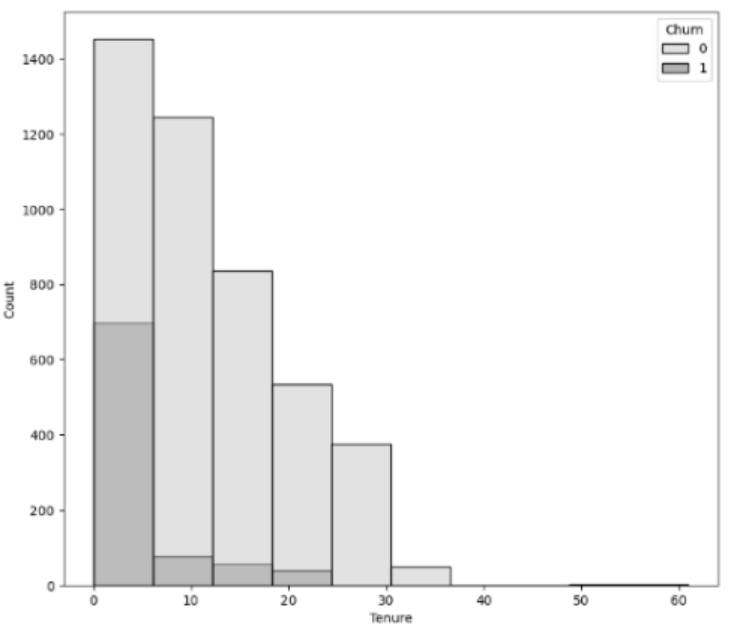
데이터 전처리	
결측치	4% 미만으로,단순 제거
인코딩	순서가 없는 범주형 데이터로, ONE-HOT-Encoding 채택

EDA

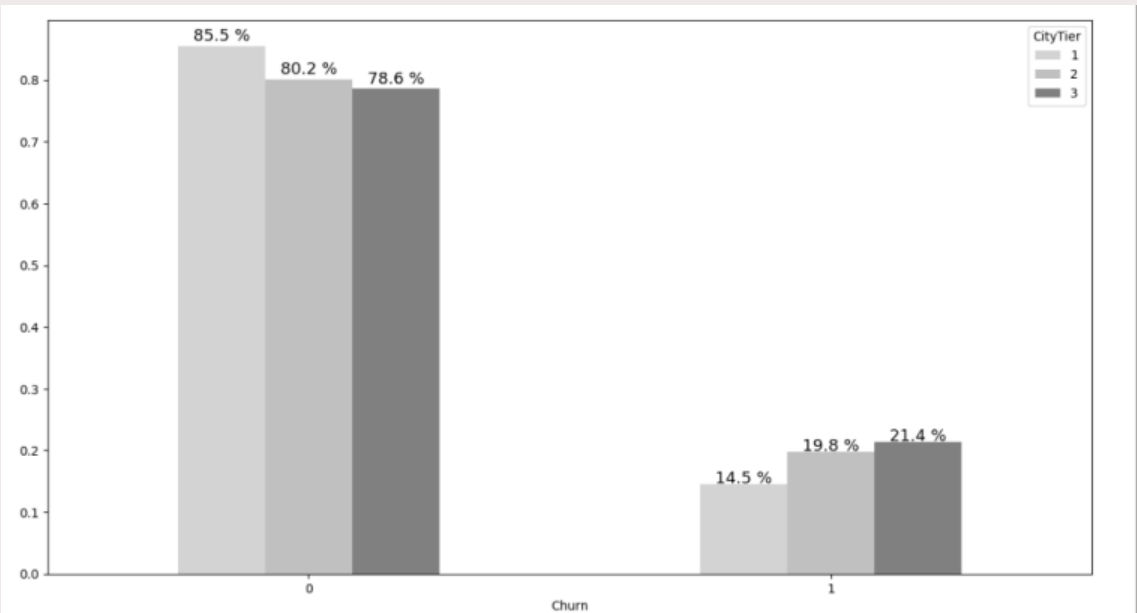
이탈 VS 잔존 고객 데이터의 불균형
이탈 고객 16.8%, 잔존 고객 83.2%로 타겟 데이터의 불균형이 심하다.



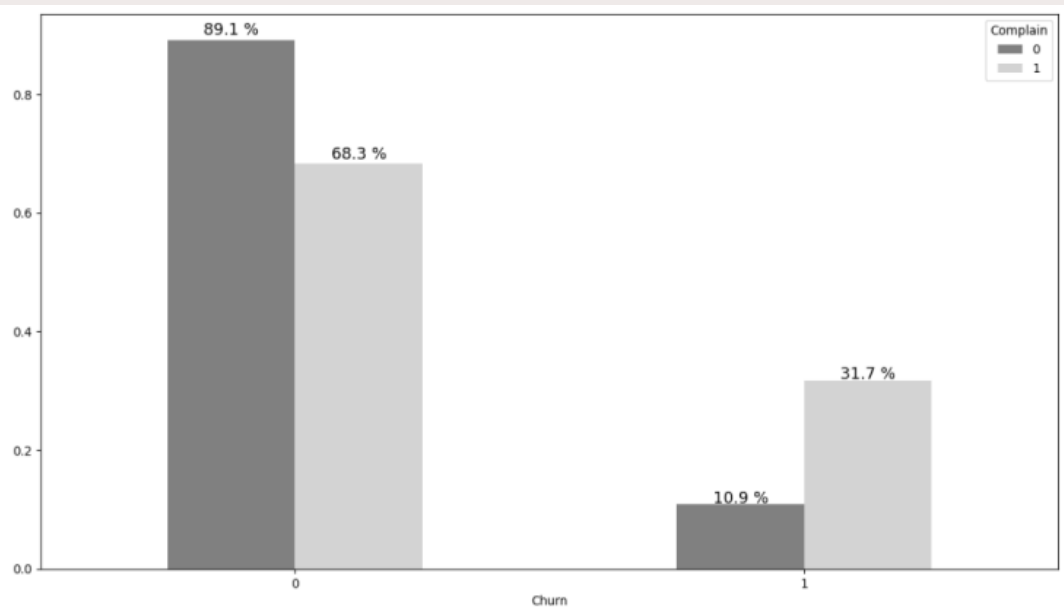
사용기간이 짧을수록 이탈율이 높다.
사용 기간이 10일 이내인 고객의 이탈율이 24.4%로 가장 높다.



도시등급이 높아질수록 이탈율이 높다.
도시등급 3은 이탈율이 가장 높은 지역으로, 이탈율은 약 21.4%



컴플레인 경험이 있을수록 이탈율이 높다.
컴플레인 경험이 있는 고객의 이탈율은 약 21% 높다.



모델링

	결측치	인코딩	이상치	스케일링	오버샘플링	분류기명	precision	Recall	F1 Score
Data 1	제거	ONE-HOT-Encoding	X	X	SMOTE	XGBClassifier	0.9423	0.9159	0.9289
Data 2	제거	ONE-HOT-Encoding	X	MinMaxScaler	SMOTE	XGBClassifier	0.9423	0.9159	0.9289
Data 3	제거	ONE-HOT-Encoding	IQR	X	SMOTE	XGBClassifier	0.9512	0.8387	0.8914
Data 4	제거	ONE-HOT-Encoding	IQR	MinMaxScaler	SMOTE	XGBClassifier	0.9412	0.8602	0.8989
Data 5	제거	ONE-HOT-Encoding	isolation_forest	X	SMOTE	XGBClassifier	0.9649	0.9649	0.9649
Data 6	제거	ONE-HOT-Encoding	isolation_forest	MinMaxScaler	SMOTE	XGBClassifier	0.9524	0.8621	0.9050

모델 비교

전처리,모델링에 따른 '36'가지 경우의 수 비교 결과, Data 5의 F1 Score이 0.9649로 가장 우수하다.

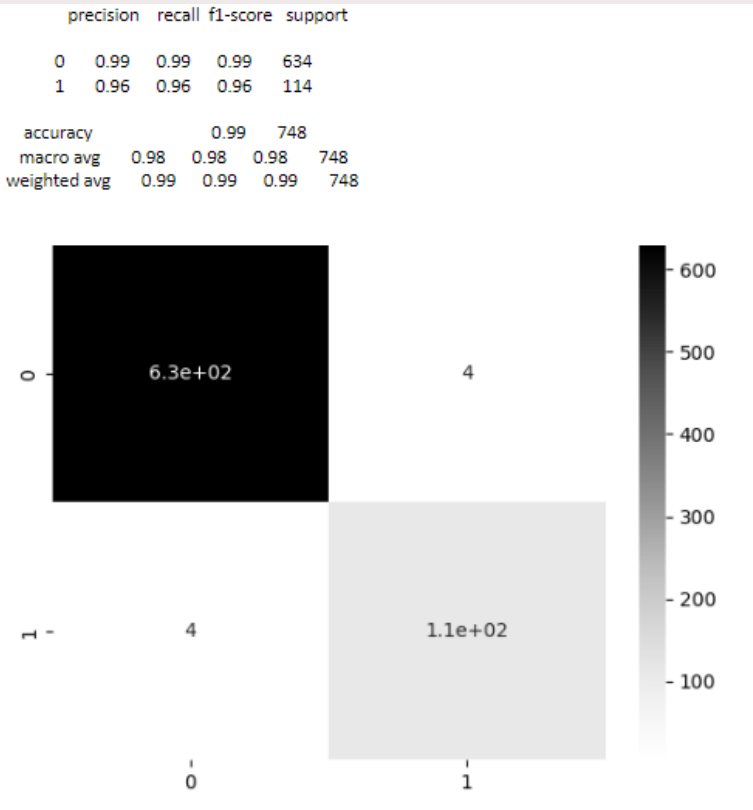
고객 이탈 예측을 위해 precision 과 Recall 의 조화 평균인 F1 Score 중점으로 비교

XGBClassifier 모델의 성능이 가장 우수하다

Dataframe	Sampling	Classifier	Accuracy	Precision	Recall	F1 Score	ROC AUC
df5	SMOTE	XGBClassifier	0.9893	0.9649	0.9649	0.9649	0.9793
df5	SMOTE	Random Forest	0.9853	0.9813	0.9211	0.9502	0.9589
df5	SMOTE	Decision Tree	0.9559	0.8092	0.9298	0.8653	0.9452
df5	SMOTE	GradientBoostingClassifier	0.9225	0.7500	0.7368	0.7434	0.8463
df5	SMOTE	Logistic Regression	0.8810	0.6147	0.5877	0.6009	0.7607
df5	SMOTE	Support Vector Machine	0.7219	0.3357	0.8421	0.4800	0.7712

모델 적용

F1 Score와 train 데이터와 test 데이터의 gap 이 0.01로 가장 적은 Data5 전처리 데이터의 XGBClassifer 선정
파라미터 튜닝 시도하였으나, 튜닝 후 성능이 오히려 저하되어 기본 베이스라인 모델 선정



결과

	CustomerID	Predict	Predcit_proba	rank
0	53402	1	99.993881	1
1	52821	1	99.984932	1
2	50787	1	99.982224	1
3	52257	1	99.980667	1
4	55072	1	99.977165	1

High risk 이탈 위험 고객 은 이탈 예상률 60%~90%로, 이탈 등급 (rank 1) 부여

액션 플랜

High risk 이탈 위험 고객 식별 데이터 + 고객 데이터 분석을 통한 인사이트 를 통한 CRM 팀 액션플랜 제시

- 서비스 이용 기간에 따른 멤버십 혜택 부여
 - 내부 고객 : 장기 이용 고객 대상 정기적인 감사 이벤트를 통해 충성도 강화
 - 이탈 위험 고객 : 이용 기간이 짧은 신규 고객에게 멤버십 혜택 소개 및 특별 리워드 증정
- 컴플레인 해결 및 고객 응대 강화
 - 내부 고객:컴플레인 고객 응대 및 문제 해결을 위한 컴플레인 해결 프로세스를 개선하고, 컴플레인 피드백을 반영하여 제품 및 서비스의 개선 방향 도출
 - 이탈 위험 고객: 이탈 위험 고객 전용 서비스 팀을 구성하여,불만사항에 대해 우선적으로 대응,주요 이슈에 대해 지속적인 모니터링
- 도시 등급 별 선호하는 결제 방식에 따른 프로모션 진행
 - 내부 고객: 도시 등급1은 "직불 카드", 도시 등급 2는 "은행 간 계좌이체", 도시 등급 3은 '전자 지갑 시스템'에 맞춰,맞춤형 혜택 제공
 - 이탈 위험 고객 : 도시 등급 3의 높은 이탈율에 주목하여, 해당 지역에 특화된 혜택 및 프로모션 진행을 통해 이탈 고객 유지 및 고객 경험 개선