

이탈 분석

1. 배경

B2C 서비스를 판매하는 기업에게 이탈자 관리는 매출 증대와 비용 감소로 이어지기에 주요 KPI 중 하나다. 만일 고객생애가치(Customer Lifetime Value)가 높은 유저가 이탈한다면 해당 유저로부터 얻을 수 있던 기대가치, 즉 매출을 잃게 된다. 또한 한번 이탈한 고객을 다시 재유입시키는 데 드는 비용은 기존에 있던 고객이 이탈하지 않도록 잡아두는 비용보다 크기 때문에 이탈자 관리는 비용 감소를 이끌어낼 수 있다. 이탈 분석을 통한 매출 증대와 비용 감소에 대한 기업 니즈를 충족시키기 위해 텐디는 이탈 분석 서비스를 자체 DMP에 탑재 또는 커스텀 분석 서비스로 제공한다.

2. 차별성

시장에 나온 여러 지표분석 툴과 타사 DMP도 이탈 분석 서비스를 제공한다. 하지만 유저가 이탈하게 된 원인이 아니라, 이탈이 발생한 후에 이탈 유저 그룹의 특성을 보여주는 것에 그친다.

예를 들어 "이탈그룹은 A event를 80% 사용한다"라는 결과를 보여준다. 하지만 실제 A event가 이탈로 이끌었는지는 알 수 없다. A event가 아닌 B event가 오히려 이탈에 더 큰 영향을 미친 변수(원인)일 수 있으므로 위의 결과를 인과관계로 속단해서는 안된다. 다른 예시로 "과거 1달동안 0월 00일에 이탈이 가장 많이 발생했다"라는 결과를 보여준다. 이 역시 마케터가 마케팅 인사이트로 활용하기에는 너무 불친절한 결과다. '왜' 이탈이 발생하는가에 대한 진단이 있어야, 그 결과를 바탕으로 이탈자를 관리하여 궁극적인 기업의 매출 및 비용 관리로 이어질 수 있다.

텐디의 이탈 분석이 바로 그러하다. 텐디는 단순히 이탈자에 대한 지표분석과 특성 나열에 그치는 것이 아니라, 유저의 모든 행태데이터 로그를 분석하여 이탈에 영향을 미치는 변수(원인)를 제공하고, 더 나아가 해당 변수(원인)의 영향력까지도 보여줄 수 있다. 예를 들어 "최근 3일간 이용시간이 절반으로 줄어들수록 이탈확률은 10% 증가한다"라는 결과를 보여준다면 앞선 예시보다 마케터가 활용하기에 훨씬 더 직관적이고 이해하기 쉽다.

3. 분석방법

3-1. 데이터

고객사의 서비스 앱/웹에서 수집되는 로그 데이터(행태 데이터)를 분석에 활용한다. 행태 데이터에는 유저가 앱/웹에 접속하여 서비스를 시작한 시점부터 종료한 시점까지 모든 행태(event, screen, 시간, referral, 노출 등)가 기록된다. 고객 행태 데이터를 분석하는 것은 가정이나 임의의 데이터를 생성해낼 필요가 없다. 즉, 현실 속에서 실제로 발생하는 고객의 행동을 들여다보는 것이므로 고객을 진정으로 이해하고 고객 자신조차도 미처 몰랐던 행동 패턴과 심리를 유추해낼 수 있는 가장 좋은 방법이다.

텐디가 제공하는 이탈 분석 서비스의 개요는 다음과 같다. 과거의 행태 데이터를 학습하여 그 속에 숨겨진 유의미한 이탈 패턴을 찾아낸다. 그리고 현 시점에 수집된 행태 데이터에 학습한 패턴을 적용하여 현재 해당 고객이 앞으로 이탈할지를 확률로 예측한다.

3-2. 변수

텐디의 이탈 분석 서비스에 사용되는 변수는 이용시간, event, screen, 광고로 크게 4가지로 분류된다. 하지만 단순히 “이용시간이 몇시간인가”, “A event를 몇번 클릭했는가”에 그치는 것이 아니라 보다 자세하고 깊이있는 변수를 생성한다. 왜냐하면 단순한 이용시간의 총합, 클릭의 횟수의 총합은 실제 고객의 행태를 다 표현할 수 없고 너무도 많은 정보손실이 발생하기 때문이다. 예를 들어 꾸준히 10분씩 5번 접속한 A 유저와, 한번에 50분을 접속한 B 유저는 분명 전혀 다른 행태를 보인다고 해석해야 한다.

따라서 한 유저의 평균 이용시간을 비롯하여 최대/최소값, 최근 n일간의 평균 이용시간, 이용시간의 분산 등 약 50여개의 이용시간 관련 변수를 추가로 생성하여 분석에 활용한다. 마찬가지로 event, screen, 광고와 관련된 변수들도 생성하여 총 120여개의 변수를 분석에 활용한다.

3-3. 알고리즘

- Time Series Analysis (시계열 분석) :

특정 기간 동안의 유저 행태를 분석해야하므로 시간의 흐름을 고려한 분석 방법을 사용한다.

- Multiple Logistic Regression (다중 로지스틱 회귀) :

이탈과 비이탈을 분류하기 위해 Logistic Regression method를 사용하고 주요 변수와 영향력을 추출한다.

- Tree based Classifiers (Random Forest, LGBM, CatBoost)(의사결정나무 기반의 머신러닝 알고리즘):
예측력을 높이기 위해 다양한 머신러닝 분류 모델을 사용한다.

4. 결과물

4-1. 이탈그룹

현 시점으로부터 향후 7일 이내에 이탈할 것으로 예상되는 유저의 비중과 수를 보여준다. 그리고 해당 유저들의 adid를 다운로드 받아 리타겟팅 모수로 활용할 수 있다.

4-2. 이탈에 영향을 미치는 변수와 영향력

이탈에 영향을 미치는 변수를 영향력 순으로 제시하고, 카테고리별로 제시한다. 마케터는 이를 리타겟팅 인사이트로 활용해 적절한 마케팅 전략을 수립할 수 있다.

4-3. 이탈그룹과 비이탈그룹의 차이 시각화/수치화

이탈하는 유저와 이탈하지 않는 유저는 어떠한 특성 차이를 보이는지 제시한다. 예를 들어, 이용시간의 차이를 시각화하거나, 각 그룹에서 A event를 실행한 유저의 비중을 수치로 표현할 수 있다. 이 역시 마케터가 인사이트로 활용할 수 있는 결과물이다.

4-4. 일자별 이탈률 추이 그래프

텐디의 이탈 분석 서비스는 일별로 제공한다. 만일 1달 동안 이탈 분석 서비스를 이용했다면 과거 1달간 고객 이탈률이 어떻게 달라졌는지 추이를 그래프로 시각화한다. 추이 그래프에서 이탈률이 급격하게 상승한 구간이 있다면 해당 일자의 시장상황, 서비스 품질 등을 사후적으로 점검할 수 있다.

4-5. 일자별 이탈 예측의 정확도 추이

과거 1달간 이탈자를 예측한 모델의 정확도 추이를 그래프로 시각화한다. 시간의 흐름에 따라 정확도가 상승한다면 텐디의 이탈 분석 서비스의 신뢰도가 상승하는 것이고, 하락한다면 이탈 예측 모델의 피드백으로 활용하여 성능 개선을 꾀한다.
