## **📄데이터 전처리 결과서**

### **1. 기본 정보**

* **프로젝트명:**
* **작성자: 조성재, 송지훈**
* **작성일: 2025년 6월 5일**
* **데이터 출처:** [Kaggle - Credit Card Customer Churn Prediction](https://www.kaggle.com/datasets/rjmanoj/credit-card-customer-churn-prediction)

### **2. 원본 데이터 개요**

* **데이터 파일명:** Churn\_Modelling.csv
* **행(Row) 수: 10000**
* **열(Column) 수: 14**
* **컬럼 목록 및 설명:**

| **열 이름** | **한국어 설명** |
| --- | --- |
| **RowNumber** | 행 번호 (데이터셋에서 각 행의 고유 번호) |
| **CustomerId** | 고객 ID (고객을 구분하기 위한 고유 식별자) |
| **Surname** | 성 (고객의 성씨) |
| **CreditScore** | 신용 점수 (신용 평가 점수로, 일반적으로 높을수록 신용도가 좋음을 의미) |
| **Geography** | 지역 (France, Spain, Germany) |
| **Gender** | 성별 (남성: Male, 여성: Female) |
| **Age** | 나이 (고객의 나이) |
| **Tenure** | 거래 기간 (해당 은행에서 거래한 연수) |
| **Balance** | 계좌 잔고 (고객 계좌의 현재 잔액) |
| **NumOfProducts** | 이용 중인 금융상품 수 (고객이 은행에서 이용하고 있는 상품의 개수) |
| **HasCrCard** | 신용카드 보유 여부 (1: 있음, 0: 없음) |
| **IsActiveMember** | 활성 고객 여부 (1: 활동 중인 고객, 0: 비활성 고객) |
| **EstimatedSalary** | 추정 연봉 (고객의 연간 소득 추정치 달러) |
| **Exited** | 이탈 여부 (1: 계좌 해지 고객, 0: 계좌 유지 고객) |

### **3. 결측치 처리**

* **결측치 발생 컬럼: 없음**

### **4. 이상치 처리**

* **이상치 탐지 기준: 통계적 분포의 사분위수(Quantile)을 기반으로 한 IQR**
* **처리 방식: 이상치의 크기가 크지 않고 양이 많지 않아서 처리하지 않음**

### **5. 스케일링 / 인코딩 등 전처리 작업**

| **처리 항목** | **적용 대상 컬럼** | **방법** | **설명** |
| --- | --- | --- | --- |
| 스케일링 | Age | 로그변환 | 연속형 변수인 Age 의 분포가 왼쪽으로 치우쳐져 있어서 이상치 영향을 줄이기 위해 로그변환을 통한 스케일링을 수행함 |
| 스케일링 | Balance | 로그변환 | 연속형 변수인 balance 의 분포가 0이 너무 많아 왼쪽으로 치우쳐져 있어서 이상치 영향을 줄이기 위해 로그변환을 통한 스케일링을 수행함 |
| 인코딩 | Geography | Label | LightGBM모델을 수행하기 위해 라벨인코딩을 수행함 |
| 인코딩 | Gender | Label | LightGBM모델을 수행하기 위해 라벨인코딩을 수행함 |

### **7. 컬럼 추가/삭제**

| **작업** | **컬럼명** | **사유** |
| --- | --- | --- |
| 추가 | LogBalance, LogAge | 스케일링을 수행한 칼럼에 기존칼럼을 대체 |
| 삭제 | Balance, Age | 스케일링을 수행한 칼럼에 대체됨 |
| 삭제 | 'RowNumber', 'CustomerId', 'Surname','CreditScore',  'Tenure',  'HasCrCard',  'EstimatedSalary' | Feature\_importance를 통해 importance가 낮은 칼럼들을 모델의 성능에 따라 제거 |

### **8. 전처리 후 데이터 요약**

* **최종 행(Row) 수: 10000개**
* **최종 열(Column) 수: 7개**