250103\_ML0

# HoML Machine Learning

본 문서는 Aurélien Géron, Hands-On Machine Learning, O'REILLY 를 바탕으로 작성되었습니다.

## [1. 개발 환경 설정](https://startedourmission.github.io/environment-setting-for-machine-learning.html)

* Google Colab
* Git / Github
* Kaggle

* 목표: 머신러닝 수행
* 핵심 학습 내용:
  + 데이터 탐색 및 시각화 기초
  + 결측치 처리와 기본 전처리
  + 로지스틱 회귀 이해
  + 결정 트리 기초
  + 모델 평가 지표 (정확도, 정밀도, 재현율)
  + 기본적인 교차 검증

## 3. 회귀 분석 마스터하기 : House Prices

* 목표: 다양한 회귀 모델과 고급 전처리 기법 학습
* 핵심 학습 내용:
  + 다양한 유형의 피처 처리
  + 선형 회귀와 규제(Ridge, Lasso, ElasticNet)
  + 다항 회귀
  + 고급 피처 엔지니어링
  + 회귀 모델 평가 지표 (MSE, RMSE, MAE)
  + 데이터 스케일링과 정규화

## 4. 불균형 데이터 다루기 : Credit Card Fraud Detection:

* 목표: 불균형 데이터 처리와 이상치 탐지
* 핵심 학습 내용:
  + 불균형 데이터 처리 기법
  + 언더샘플링과 오버샘플링
  + SMOTE 등 고급 샘플링 기법
  + 비용 기반 학습
  + ROC 곡선과 AUC
  + 이상치 탐지 알고리즘

## 5. 시계열 기초와 회귀 : Store Sales Forecasting

* 목표: 시계열 데이터의 기본적인 처리와 예측
* 핵심 학습 내용:
  + 시계열 데이터 기초
  + 이동평균과 지수평활
  + 시계열 피처 엔지니어링
  + 시계열 교차 검증
  + 회귀 기반 예측
  + 시계열 평가 지표

## 6. 비지도 학습 실전 : Customer Segmentation

* 목표: 군집화와 차원 축소 학습
* 핵심 학습 내용:
  + K-means 클러스터링
  + 계층적 군집화
  + PCA를 통한 차원 축소
  + 실루엣 분석
  + 군집화 평가 방법
  + 고객 세그먼테이션 전략

## 7. 앙상블 학습 : Porto Seguro's Safe Driver Prediction

* 목표: 다양한 앙상블 기법 마스터
* 핵심 학습 내용:
  + 랜덤 포레스트 심화
  + Gradient Boosting (XGBoost, LightGBM)
  + 배깅과 부스팅의 원리
  + 하이퍼파라미터 튜닝
  + 스태킹과 블렌딩
  + 모델 해석기법 (SHAP, Feature Importance)