



결과 정리

주제 및 배경

- **배경** : 최근 음악 스트리밍 사이트 등에서 음악들을 자체적으로 분류하여 고객들에게 추천을 하거나, 수많은 음악들을 효과적으로 분류할 수 있는 방법들에 대한 수요와 시도가 지속적으로 증가하고 있음. 따라서 이러한 문제를 해결하는 첫걸음으로, 음악 샘플들의 특징으로부터 장르를 자동적으로 분류하는 AI 모델을 생성해보고자 함
- **주제** : 음악 장르를 분류하는 AI 알고리즘 개발
- **카테고리** : 알고리즘 | 정형 | 분류 | 음악 | Macro F1 Score
- **사용 데이터** : 데이콘 Basic 음악 장르 분류 AI 경진대회
- **평가** : Macro F1 Score

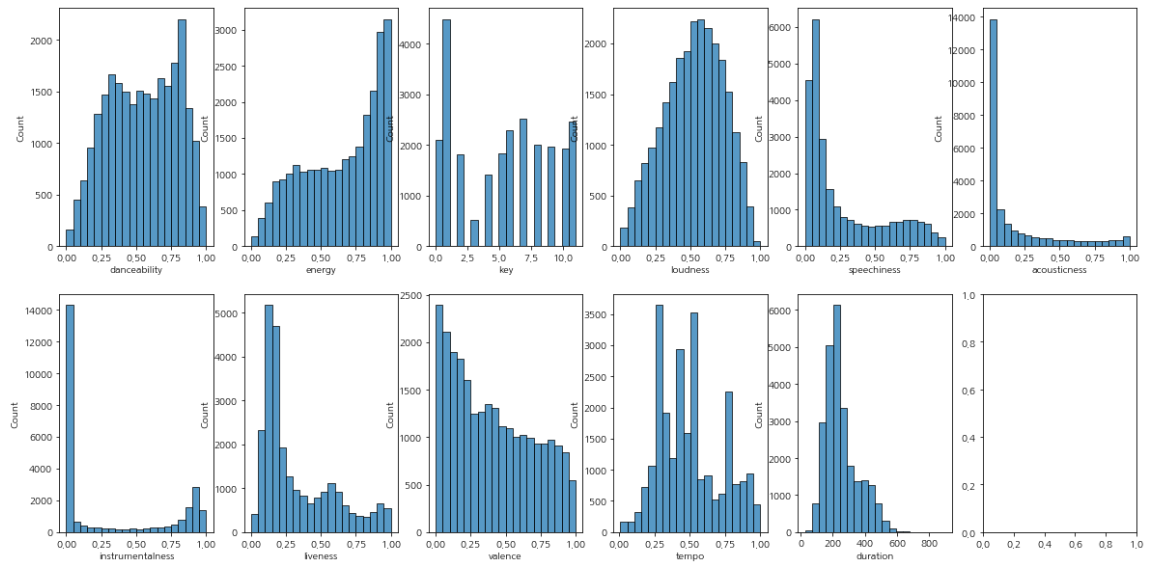
EDA

▼ Columns

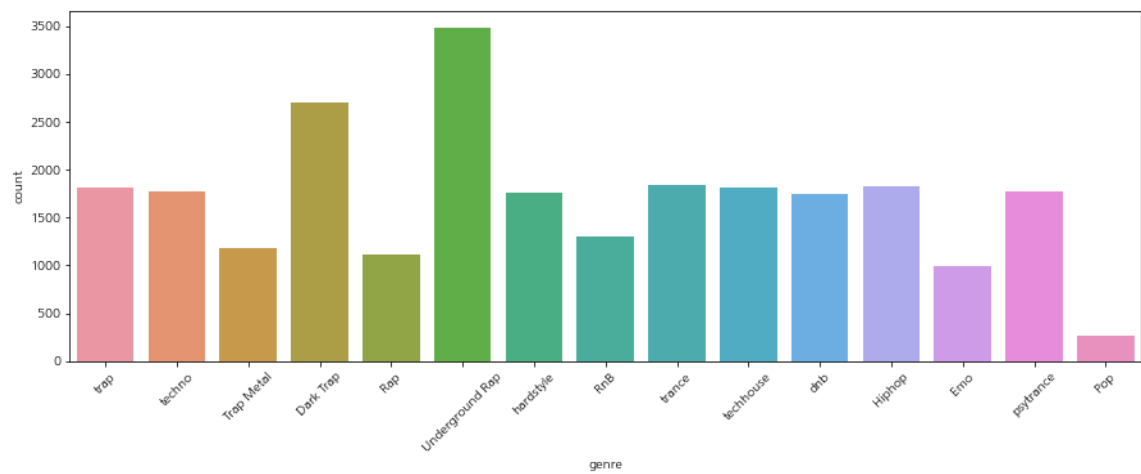
- **ID** : 음악 샘플 고유 ID
- 음악 샘플의 특징 정보danceability, energy, key, loudness, speechiness, acousticness, instrumentalness, liveness, valence, tempo
- **duration** : 음악 재생 시간(길이) (초)
- **genre** : 음악 장르 (총 15개 종류)Underground Rap, Dark Trap, trance, Hiphop, trap, techhouse, techno, psytrance, hardstyle, dnb, RnB, Trap Metal, Rap, Emo, Pop

▼ train

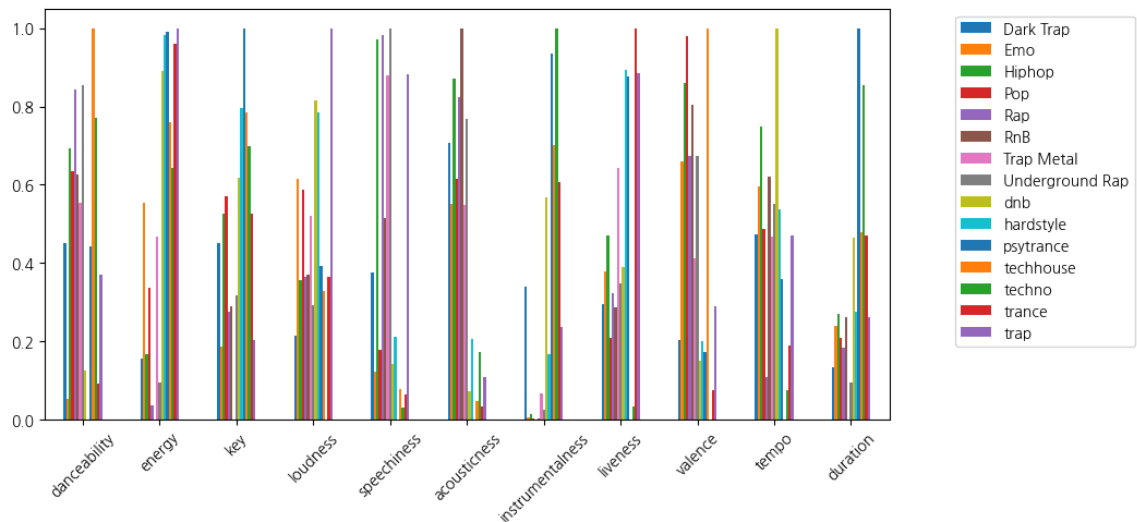
- **shape** : (25383, 13)
- 변수별 분포 확인



- target 변수 분포 확인



- genre 별 변수들의 분포 확인



Model 후보 및 결과

※ 데이콘 대회 기간 종료로 Submission이 불가능해 고정된 Validation set을 Test set 대신 사용해 점수 도출

- 팀원 별 2개씩 총 8개의 모델을 선정하여 모델링 및 하이퍼파라미터 튜닝 진행
- **Model 자료 조사 (Model 이름_f1 score)**

 Random Forest_0.631

 Catboost_0.648

XGBoost_0.658

SVM_0.613

 LightGBM_0.649

 KNN_0.5265

 Logistic Regression_0.533

 Gradient Boosting Classifier_0.648

최종 모델 선정

▼ XGBoost - 0.658

- tree_method = 'exact'

- `gpu_id = 0`
- `booster = 'dart'`
- `n_estimators = 1000`
- `max_depth = 5, max_leaves = 31`
- `learning_rate = 0.01`
- `subsample = 0.7`
- `colsample_bytree = 0.7`
- `n_jobs = 8`

사용 코드

음악장르분류대회 - Google Drive

<https://drive.google.com/drive/folders/12ViguCR8gDoug0Nqe8FMNtAjzC4SBCho>