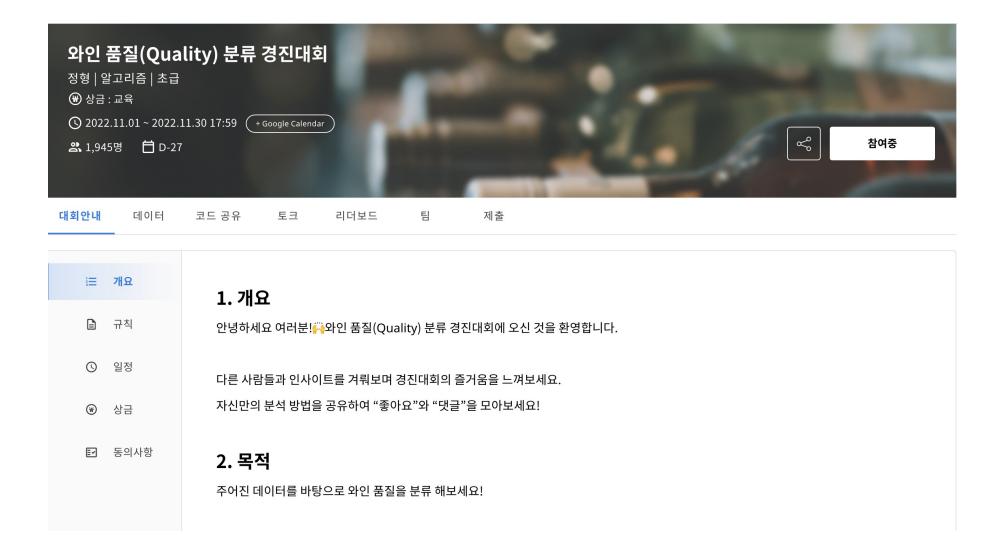


대회 소개



https://www.dacon.io/competitions/open/235610/overview/description

대회 소개

1. 규칙

- 제출 횟수 및 최대 팀원
 - a. 1일 최대 제출 횟수: 10회
 - b. 팀 최대 인원 : 5명
- 리더보드
 - a. 평가 산식 : ACCURACY

2. 외부 데이터 및 사전학습 모델

- 공공 데이터와 같이 누구나 얻을 수 있고 법적 제약이 없는 외부 데이터 허용
- 사전 학습 모델의 경우 사전 학습에 사용된 데이터를 명시해야 함

3. 개인 및 팀 병합 규정

- 개인 참가 방법 : 팀 신청 없이 자유롭게 제출탭에서 제출 가능
- 팀 참여 방법 : 팀 탭에서 가능, 상세 내용은 팀 탭에서 팀 병합 정책 확인
- 하나의 대회는 하나의 팀으로만 등록 가능
- 팀 병합 후 해체 및 개인 참가 불가

데이터 살펴보기

	quality	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pН	sulphates	alcohol	type
index													
0	5	5.6	0.695	0.06	6.8	0.042	9.0	84.0	0.99432	3.44	0.44	10.2	white
1	5	8.8	0.610	0.14	2.4	0.067	10.0	42.0	0.99690	3.19	0.59	9.5	red
2	5	7.9	0.210	0.39	2.0	0.057	21.0	138.0	0.99176	3.05	0.52	10.9	white
3	6	7.0	0.210	0.31	6.0	0.046	29.0	108.0	0.99390	3.26	0.50	10.8	white
4	6	7.8	0.400	0.26	9.5	0.059	32.0	178.0	0.99550	3.04	0.43	10.9	white

- index 구분자
- fixed acidity 산도
- volatile acidity 휘발성산
- citric acid 시트르산
- residual sugar 잔당 : 발효 후 와인 속에 남아있는 당분
- chlorides 염화물
- free sulfur dioxide 독립 이산화황
- total sulfur dioxide 총 이산화황
- density 밀도
- pH 수소이온농도
- sulphates 황산염
- alcohol 도수
- type 종류

Input: 12

Output: 1

모델 추천

Extra Trees Classifier

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html

Random Forest Classifier

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Gradient Boosting Classifier

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html

KNN, Logistic Regression

와인 품질 분류 경진대회

Hint

- 이상치 확인: sns.boxplot(x = "column name", data = df)
- Data scaling
- data.corr()을 통해 correlation 확인
- validation set 만들기
- 모델 파라미터 튜닝
- Kfold
- Ensemble

참고 코드

https://drive.google.com/file/d/1IGV2k6eaLcL6lyBAleukwsYotDtsJY3j/view?usp=sharing https://dacon.io/competitions/open/235610/codeshare/4154?page=1&dtype=recent

