**주제**

Red Wine Quality에 영향을 미치는 요인 분석

**개요**

와인의 물리화학적 요소를 기반으로 데이터를 분석하여 Red wine의 품질을 측정합니다. 다양한와인의 특성을 분석하여 와인 품질에 영향을 미치는 주요 요소를 식별하고자 합니다. 이를 통해 와인 제조업체나 소비자에게 와인 품질을 향상시키는 데 도움이 되는 인사이트를 제공할 수 있습니다.

**데이터 정보**

**[Kaggle data link]**

<https://www.kaggle.com/datasets/lovishbansal123/red-wine-quality>

**[Column 정보]**

1) fixed acidity: 고정 산도

2) volatile acidity: 휘발성 산도

3) citric acid: 시트르산

4) residual sugar: 잔류 설탕

5) chlorides: 염화물

6) free sulfur dioxide: 자유 이산화황

7) total sulfur dioxide: - 총 이산화황

8) density: 밀도

9) pH: pH

10) sulphates: 황산염

11) alcohol: 알코올

12) quality (score between 0 and 10) : 품질(사람의 주관적인 판단에 의한)

**분석 계획**

1. 데이터 수집

Kaggle의 데이터셋을 사용하여 레드 와인의 물리화학적 특성과 데이터를 수집합니다.

1. EDA

각 column의 의미를 파악하고, 데이터의 분포를 시각화하여 이상치나 결측치를 확인합니다.

1. 데이터 전처리

이상치나 결측치를 처리하고, 필요하다면 데이터의 스케일을 조정하거나 범주형 변수를 처리합니다.

1. 요인 분석

품질에 영향을 미치는 주요 요인을 분석하기 위해 상관 분석 등을 수행합니다.

1. 품질 예측 모델 구축

회귀 분석, 결정 트리, 랜덤 포레스트 등의 모델을 사용하여 와인 품질을 예측할 수 있는 모델을 설계합니다.

1. 모델 학습

선택한 모델을 와인 데이터에 적용하여 모델을 학습시킵니다.

1. 모델 평가

테스트 데이터에 모델을 적용하여 예측을 수행하고, 평가 지표를 사용하여 모델의 성능을 평가합니다. AUC 값을 통해 모델의 예측 성능을 확인합니다.

1. 결과 해석

결과를 해석하여 각 특성이 와인 품질에 미치는 영향을 파악하고, 와인 제조 과정이나 품질 개선을 위한 제안을 도출합니다.