

인공지능 과제

- Logistic regression -

컴퓨터 정보통신 공학과 214683

ㅅ ㅅ ㅅ

목차



1. 모델링 배경 및 목적
2. 데이터 설명
3. 모델링 과정
4. 결과 및 성능

1. 모델링 배경 및 목적

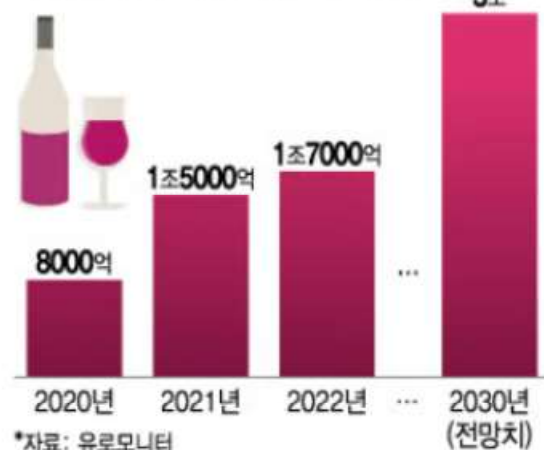


모델링 배경 및 목적

와인 시장 규모

글로벌 와인 시장 규모는 2024년에 4,635억 달러 로 평가되었으며, 2025년에는 4,902억 3,000만 달러 에서 2033년에는 7,491억 달러 로 성장할 것으로 예상되며, 예측 기간(2025-2033) 동안 연평균 성장률 5.5% 로 성장할 것입니다.

국내 와인 시장 규모 (단위: 원)



*자료: 유로모니터
그래픽: 윤선정 디자인가자

현재 와인 시장은 점점 커지고 있으며, 앞으로도 증가 할 가능성이 굉장히 높습니다.

와인에 대해 입문하고 싶었던 마음도 한편에 존재 하였기 때문에 관심이 갔습니다.

그중 레드 와인은 무슨 요소가 와인의 질에 영향을 주는지 분석해보고자 주제로 선정하게 되었습니다.

2. 데이터 설명



데이터 설명

```
[4]: wine = pd.read_csv("winequality-red.csv")
      wine.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1599 entries, 0 to 1598
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   fixed acidity         1599 non-null   float64
 1   volatile acidity      1599 non-null   float64
 2   citric acid           1599 non-null   float64
 3   residual sugar        1599 non-null   float64
 4   chlorides             1599 non-null   float64
 5   free sulfur dioxide   1599 non-null   float64
 6   total sulfur dioxide  1599 non-null   float64
 7   density              1599 non-null   float64
 8   pH                   1599 non-null   float64
 9   sulphates            1599 non-null   float64
10   alcohol              1599 non-null   float64
11   quality              1599 non-null   int64  
dtypes: float64(11), int64(1)
memory usage: 150.0 KB
```

1600개의 데이터 셋
11개의 독립변수, 1개의 종속변수 존재
종속 변수 : quality

- 1 - fixed acidity : 휘발성
- 2 - volatile acidity : 비휘발성 산도
- 3 - citric acid : 유기산
- 4 - residual sugar : 당분
- 5 - chlorides : 염화물
- 6 - free sulfur dioxide : 유리 이산화황
- 7 - total sulfur dioxide : 총 이산화황
- 8 - density : 밀도
- 9 - pH : 산성
- 10 - sulphates : 황산염
- 11 - alcohol : 알콜
- 12 - quality

3. 모델링 과정



모델링 과정

1. 데이터 분리

```
[5]: y=wine['quality']  
     x=wine.drop('quality', axis=1)
```

y축 : 예측값(질)

x축 : quality를 제외한 모든 요소들

2. 종속 변수 이진화로 변경 및 데이터셋 분리

```
[8]: y = np.where(y>5, 1, 0)  
     x_train, x_test, y_train, y_test = train_test_split(x,y,test_size = 0.2, random_state = 42)
```

3. 스케일링

```
[35]: scaler=StandardScaler()  
     x_train = scaler.fit_transform(x_train)  
     x_test = scaler.transform(x_train)
```


모델링 과정

4. 모델학습

```
[40]: num_epochs = 100
      losses = []

      model = SGDClassifier(
          loss='log_loss', max_iter = 1, tol = 0,
          learning_rate = 'constant', eta0=0.01,
          warm_start=True, random_state=42
      )
      for epoch in range(num_epochs):
          model.fit(x_train, y_train)
          y_proba = model.predict_proba(x_train)[:,-1]

          epsilon = 1e-15
          y_proba = np.clip(y_proba, epsilon, 1-epsilon)
          log_loss = -np.mean(y_train * np.log(y_proba) + (1-y_train) * np.log(1-y_proba))
          losses.append(log_loss)
```

4. 결과 및 성능



결과 및 성능

```
[53]: pred = np.concatenate([model.intercept_, model.coef_.reshape(-1)]).round(2)
      odds = np.exp(pred).round(2)
      interpret = np.where(pred>0, 'good', 'bad')
```

	beta	exp(beta)	interpret
const	0.17	1.19	good
fixed acidity	0.32	1.38	good
volatile acidity	-0.62	0.54	bad
citric acid	-0.29	0.75	bad
residual sugar	0.05	1.05	good
chlorides	-0.11	0.9	bad
free sulfur dioxide	0.28	1.32	good
total sulfur dioxide	-0.57	0.57	bad
density	-0.1	0.9	bad
pH	-0.02	0.98	bad
sulphates	0.49	1.63	good
alcohol	0.83	2.29	good

와인에 좋은 영향을 끼치는 요인 :

- 비휘발성산, ɔ %, 우리 이산화황, 황산염, 알코올

와인에 안좋은 영향을 끼치는 요인

- 휘발성산, 구연산, 염화물, 총 이산화황, 밀도, ð ~

참고자료

- <https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009>
- <https://straitsresearch.com/ko/report/wine-market>
- <https://news.mt.co.kr/mtview.php?no=2023052415183742823>