



머신러닝 과제

학과 : 인공지능

이름 : 김지유(202078402)

2024.06.04

01. 사용한 캐글 데이터셋

- 사용한 캐글 데이터셋과 분석 내용

<https://www.kaggle.com/datasets/merishnasuwal/breast-cancer-prediction-dataset/data>

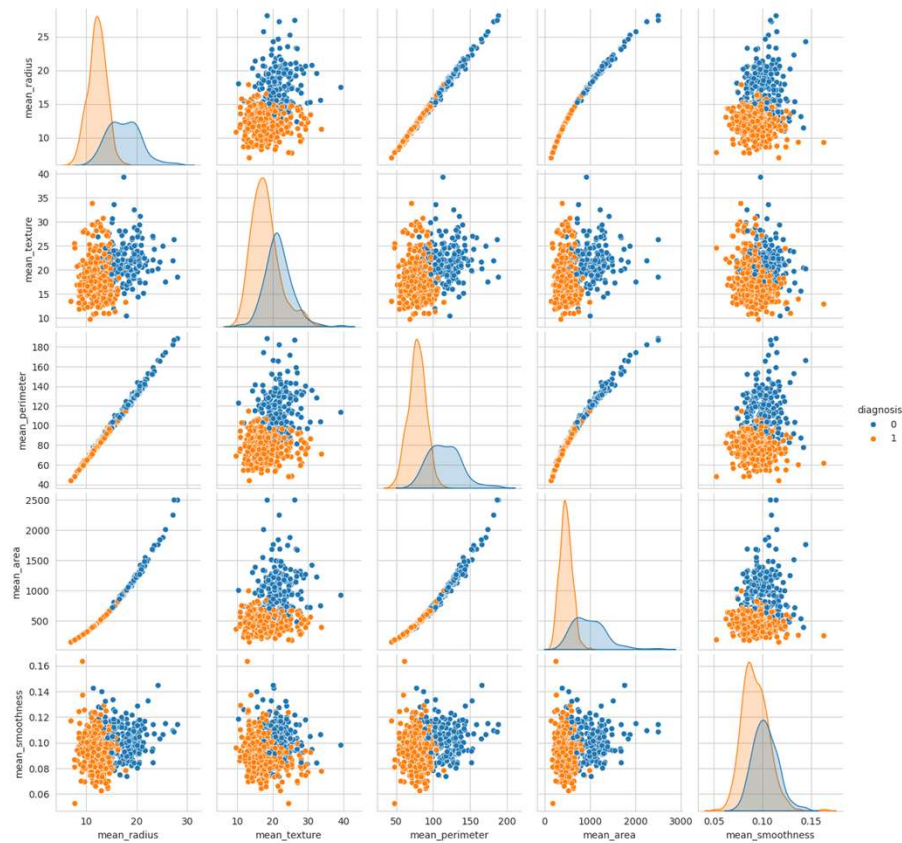
- 유방암 예측

- 'mean_radius', 'mean_texture', 'mean_perimeter', 'mean_area', 'mean_smoothness' 를 통해 종양이 양성인지 악성인지를 분류하는 것이 목표

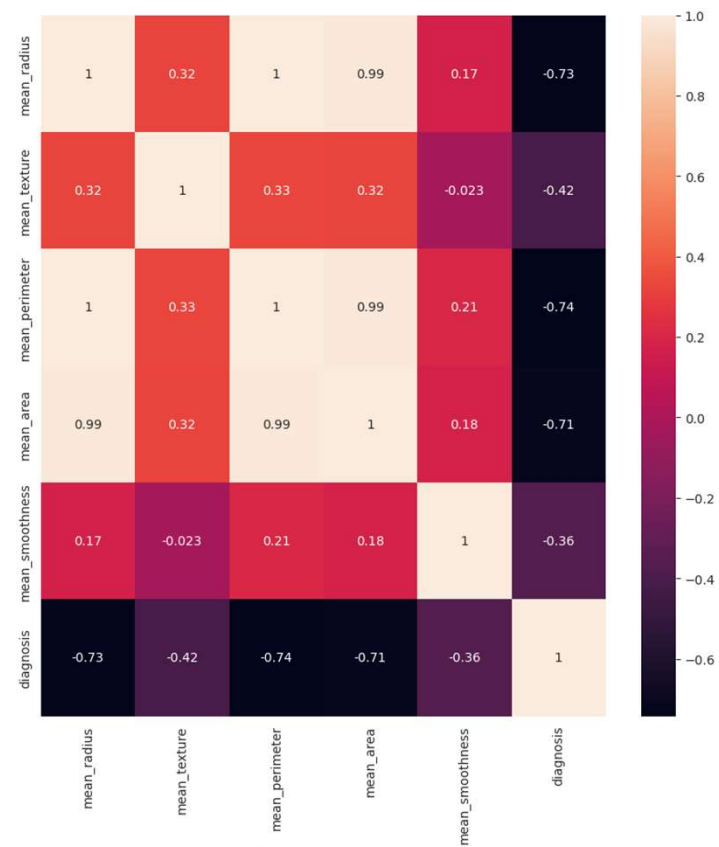
- LinearSVC, SVC, RBF, PCA 등 여러 학습 기법들을 활용해 정확도 비교

02. 데이터 시각화

pairplot



heatmap



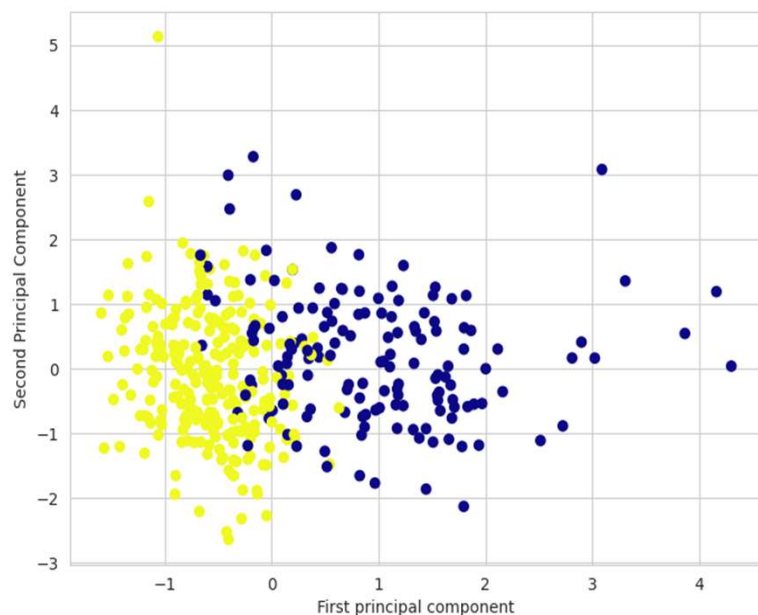
03. 기계학습 기법 비교

| ML 학습기법 | train | test |
|-----------------------|--------|--------|
| LinearSVC | 80.90% | 80.70% |
| SVC | 91.71% | 94.15% |
| rbf | 92.71% | 94.74% |
| Linear Kernel SVM | 92.71% | 94.74% |
| Polynomial Kernel SVM | 73.62% | 75.44% |
| Radial Kernel SVM | 93.97% | 94.74% |
| GridSearchCV | 92.71% | 94.74% |

SVC모델 중 가장 정확도가 높은 것은 “Radial Kernel SVM”이며, 가장 정확도가 낮은 학습모델은 “Polynomial Kernel SVM” 로 확인

03. 기계학습 기법 비교

- PCA를 활용하여 유방암 진단에 중요하지 않은 정보를 제거하고, 주요 정보만 추출해서 재 진행하여 정확도가 TRAIN -> 92.46%, TEST-> 94.15로 확인됨



```
Fitting 5 folds for each of 126 candidates, totalling 630 fits
Best params: {'C': 10, 'gamma': 1, 'kernel': 'linear'}
Train Result:
=====
Accuracy Score: 92.46%

CLASSIFICATION REPORT:
=====

```

| | 0 | 1 | accuracy | macro avg | weighted avg |
|-----------|------------|------------|----------|------------|--------------|
| precision | 0.928058 | 0.922780 | 0.924623 | 0.925419 | 0.924756 |
| recall | 0.865772 | 0.959839 | 0.924623 | 0.912806 | 0.924623 |
| f1-score | 0.895833 | 0.940945 | 0.924623 | 0.918389 | 0.924056 |
| support | 149,000000 | 249,000000 | 0.924623 | 398,000000 | 398,000000 |

```

Confusion Matrix:
[[129 20]
 [ 10 239]]

Test Result:
=====
Accuracy Score: 94.15%

CLASSIFICATION REPORT:
=====

```

| | 0 | 1 | accuracy | macro avg | weighted avg |
|-----------|-----------|------------|----------|------------|--------------|
| precision | 0.934426 | 0.945455 | 0.94152 | 0.939940 | 0.941391 |
| recall | 0.904762 | 0.962963 | 0.94152 | 0.933862 | 0.941520 |
| f1-score | 0.919355 | 0.954128 | 0.94152 | 0.936742 | 0.941317 |
| support | 63,000000 | 108,000000 | 0.94152 | 171,000000 | 171,000000 |

```

Confusion Matrix:
[[ 57  6]
 [  4 104]]
```