

KoGES 골다공증 예측

증강지능 연구실 황승현

2023-10-17

목차

- 이전 내용 정리
- 골다공증 분석 모델 소개 및 문제점
- 지적사항 수정 후
- 결론

골다공증 분석 모델

소개 그리고 문제점

여러 알고리즘으로 나온 결과

	식이패턴	식이패턴+유전자	유전자
DT	0.729	0.734	0.679
KNN	0.749	0.749	0.575
SVM	0.667	0.608	0.262
XGB	0.785	0.776	0.713
LGBM	0.789	0.775	0.682

- 그럴듯한 결과?
- 문제점 있음

문제점

- Data Augmentation 오류
 - Train 데이터 뿐만 아니라, Test 데이터도 증강
 - 모델 학습의 순수성 보장 x
- Data Scaling 오류
 - 전체 데이터로 scaling 후 train test 분리...?

지적사항 수정

지적사항 해결

기존

- Data Augmentation
 - X, y 모두 증강
- Data Scaling
 - X, y 전체를 보고 Scaling

개선

- Data Augmentation
 - $X_{\text{train}}, y_{\text{train}}$ 만 증강
 - $X_{\text{test}}, \text{val}$ 등은 증강 X
- Data Scaling
 - Train, test 각각 Scaling

폐경 전, 식이패턴 있음, 유전체 있음

Confusion Matrix:
[[282 12]
[11 1]]

Accuracy: 0.925
Precision: 0.077
Recall : 0.083
F1 Score: 0.080

Confusion Matrix:
[[258 36]
[11 1]]

Accuracy: 0.846
Precision: 0.027
Recall : 0.083
F1 Score: 0.041

Confusion Matrix:
[[294 0]
[12 0]]

Accuracy: 0.961
Precision: 1.000
Recall : 0.000
F1 Score: 0.000

Confusion Matrix:
[[224 10]
[9 1]]

Accuracy: 0.922
Precision: 0.091
Recall : 0.100
F1 Score: 0.095

Confusion Matrix:
[[290 4]
[10 2]]

Accuracy: 0.954
Precision: 0.333
Recall : 0.167
F1 Score: 0.222

Confusion Matrix:
[[290 4]
[10 2]]

Accuracy: 0.954
Precision: 0.333
Recall : 0.167
F1 Score: 0.222

폐경 후, 식이패턴 있음, 유전체 있음

Confusion Matrix:
[[273 97]
[84 60]]

Accuracy: 0.648
Precision: 0.382
Recall : 0.417
F1 Score: 0.399

Confusion Matrix:
[[280 90]
[105 39]]

Accuracy: 0.621
Precision: 0.302
Recall : 0.271
F1 Score: 0.286

Confusion Matrix:
[[119 251]
[8 136]]

Accuracy: 0.496
Precision: 0.351
Recall : 0.944
F1 Score: 0.512

Confusion Matrix:
[[197 98]
[68 48]]

Accuracy: 0.596
Precision: 0.329
Recall : 0.414
F1 Score: 0.366

Confusion Matrix:
[[370 0]
[144 0]]

Accuracy: 0.720
Precision: 1.000
Recall : 0.000
F1 Score: 0.000

Confusion Matrix:
[[370 0]
[144 0]]

Accuracy: 0.720
Precision: 1.000
Recall : 0.000
F1 Score: 0.000

주요 수치 정리

이전	폐경 전	폐경 후
XGBoost	0.959	0.682
LGBM	0.974	0.713

현재	폐경 전	폐경 후
XGBoost	0.222	0.000
LGBM	0.222	0.000



전체 데이터로 학습했을 때

이전

- Train, test 100% 나왔음
- 학습을 할 수 있는 문제

현재

```
[62] 1 # Defining the hyper parameters
      2 hps = {
      3     'max_depth': 5,
      4     'min_samples_split': 4
      5 }
      6
      7 # Loading the tree object
      8 tree = DecisionTreeClassifier(**hps)
      9 tree.fit(x, y)
```

```
DecisionTreeClassifier
DecisionTreeClassifier(max_depth=5, min_samples_split=4)
```

```
[63] 1 y_pred = tree.predict(x)
      2
      3 SAMCGS(y, y_pred)
```

```
Confusion Matrix:
[[1764  82]
 [ 535 186]]
```

```
[[TP  FN]
 [ FP  TN]]
```

```
Accuracy: 0.760
Precision: 0.694
Recall   : 0.258
F1 Score: 0.376
```

결론 및 향후 계획

- 왜 이렇게 됐을까..?
- 스케일링 방법 변경
 - 기존: QuantileTransformer
 - 변경 예정..?
 - RobustScaler
 - sklearn.preprocessing Normalizer