

변수 자르기 대작전

방법

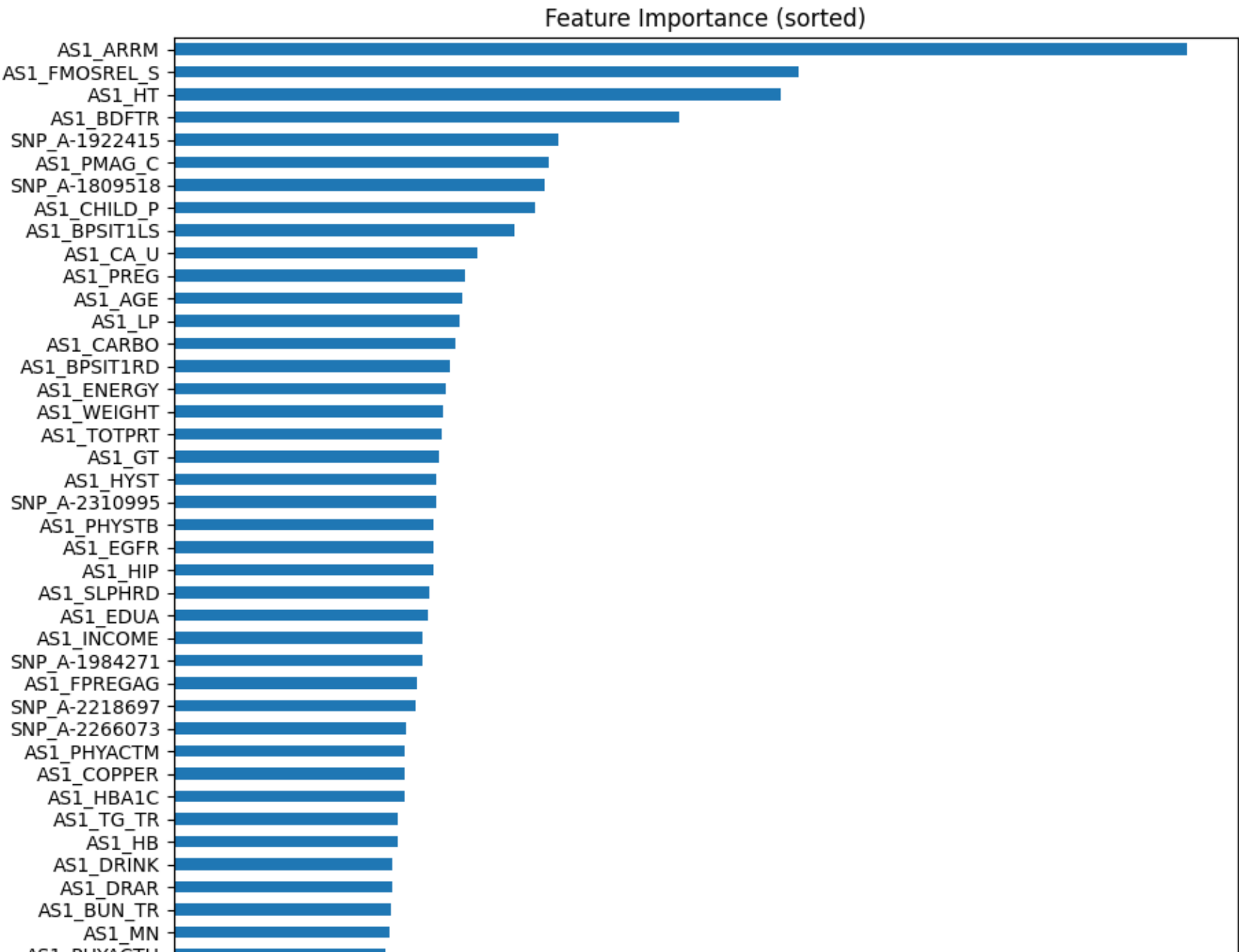
- 1. 전체 x 증강 후 xgboost 학습
- 2. xgboost의 feature importance 추출
- 3. 상위 20개만을 이용하여 다시 학습
- 4. 1번과 3번의 f1-score 비교

실험

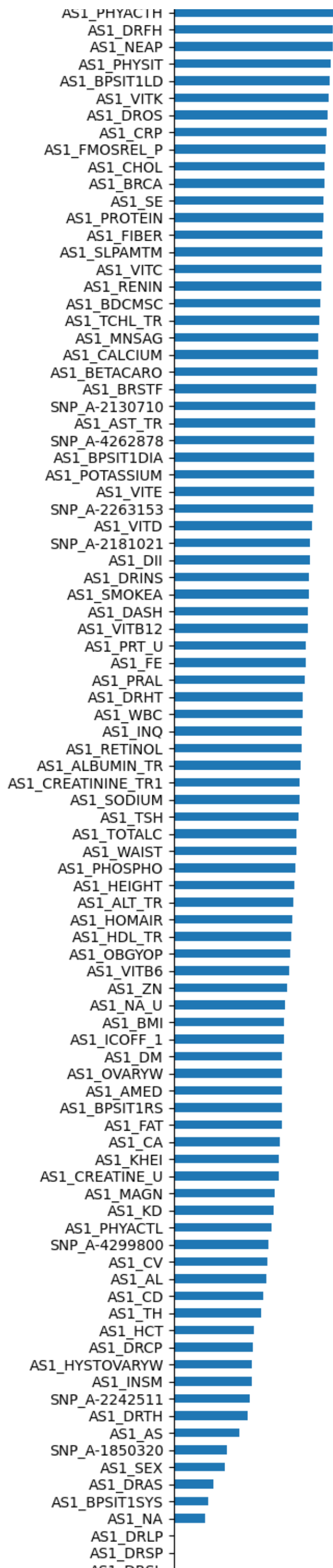
f1 score 등

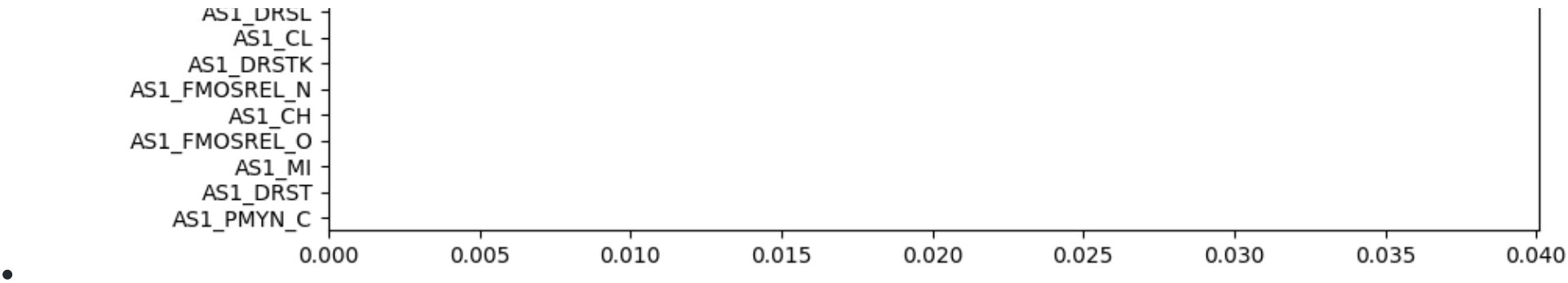
- x_train만 증강 후 xgboost : 자료 삭제
최소: 0.3 최대: 0.34782608695652173 평균: 0.3262824444034863
- x_train, test, val 각각 증강
Confusion Matrix: [[285 85] [102 268]] [[TP FN] [FP TN]] Accuracy: 0.747 Precision: 0.759 Recall : 0.724 F1 Score: 0.741
- 전체 증강 후 xgboost
Confusion Matrix: [[292 77] [89 281]] [[TP FN] [FP TN]] Accuracy: 0.775 Precision: 0.785 Recall : 0.759 F1 Score: 0.772
- 전체 증강 후 xgb
Confusion Matrix: [[290 80] [93 277]] [[TP FN] [FP TN]] Accuracy: 0.766 Precision: 0.776 Recall : 0.749 F1 Score: 0.762

feature importance



Feature





상위 20개 목록

S1_HYST 0.009873 AS1_GT 0.009998 AS1_TOTPRD 0.010084 AS1_WEIGHT 0.010143 AS1_ENERGY 0.010235 AS1_BPSIT1RD 0.010397
AS1_CARBO 0.010594 AS1_LP 0.010786 AS1_AGE 0.010885 AS1_PREG 0.010990 AS1_CA_U 0.011435 AS1_BPSIT1LS 0.012847
AS1_CHILD_P 0.013625 SNP_A-1809518 0.013976 AS1_PMAG_C 0.014113 SNP_A-1922415 0.014502 AS1_BDFTR 0.019050 AS1_HT
0.022881 AS1_FMOSREL_S 0.023556 AS1_ARRM 0.038178

['SNP_A-2242511', 'AS1_CA_U', 'AS1_INCOME', 'AS1_BPSIT1RD', 'AS1_WAIST', 'AS1_SLPAMTM', 'AS1_EDUA', 'AS1_PHYSTB', 'AS1_SLPHRD',
'SNP_A-1922415', 'AS1_PHYACTM', 'SNP_A-2181021', 'AS1_BRCA', 'SNP_A-2218697', 'AS1_CHILD_P', 'AS1_DRHT', 'AS1_AGE', 'SNP_A-
1984271', 'AS1_HT', 'AS1_ARRM']

top20 추출 후

Confusion Matrix: [[287 82] [79 291]] [[TP FN] [FP TN]] Accuracy: 0.782 Precison: 0.780 Recall : 0.786 F1 Score: 0.783