

머신러닝과 산업응용 과제2

반도체 공정에서 웨이퍼의 품질 상태를 판정하기 위한 데이터가 있다. 이 데이터는 총 474개의 독립변수와 1개의 종속변수로 구성되어 있으며, 독립변수는 공정 중 모니터링된 센서 값이다. 종속변수는 웨이퍼의 품질 상태를 나타내며, 정상일 경우 -1, 불량일 경우 +1로 표시된다. 전체 데이터는 254개이며, 이러한 센서 값을 이용하여 공정이 완료된 후의 웨이퍼 품질을 예측하고자 한다.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	...	X465	X466	X467	X468	X469
0	0.373044	0.500048	0.318866	0.113673	0.181737	0.500728	0.543103	0.595661	0.218789	0.451876	...	0.410448	0.145439	0.434028	0.017065	0.019531
1	0.659672	0.367706	0.179527	0.029298	0.215474	0.537828	0.568966	0.366978	0.417800	0.592170	...	0.283280	0.246376	0.673611	0.017491	0.025391
2	0.352476	0.401779	0.463937	0.126285	0.083402	0.491896	0.284483	0.473938	0.483313	0.543230	...	0.283280	0.246376	0.697917	0.024104	0.033203
3	0.249908	0.331418	0.361890	0.089591	0.091509	0.530866	0.560345	0.818620	0.320148	0.699837	...	0.283280	0.246376	0.704861	0.033276	0.031250
4	0.197750	0.390574	0.222614	0.132807	0.154885	0.430756	0.500000	0.463091	0.483313	0.632953	...	0.283280	0.246376	0.902778	0.039676	0.034180
5	0.763736	0.351900	0.411149	0.427761	0.176048	0.317722	0.663793	0.360048	0.411619	0.316476	...	0.283280	0.246376	0.791667	0.007466	0.005859
6	0.521581	0.393437	0.235276	0.223348	0.175878	0.405294	0.422414	0.733655	0.504326	0.339315	...	0.283582	0.498101	0.809028	0.007892	0.008789
7	0.453824	0.445588	0.716094	0.581922	0.224718	0.408723	0.534483	0.409762	0.358467	0.270799	...	0.283280	0.246376	0.562500	0.010879	0.013672
8	0.522225	0.416382	0.470929	0.454682	0.241245	0.373771	0.534483	0.271467	0.524104	0.484502	...	0.283280	0.246376	0.756944	0.081485	0.074219
9	0.164343	0.571191	0.324472	0.401357	0.249864	0.229873	0.448276	0.592648	0.498146	0.340946	...	0.201493	0.214622	0.621528	0.020904	0.018555
10	0.622768	0.436559	0.324346	0.269970	0.325159	0.260426	0.551724	0.470021	0.447466	0.306688	...	0.283280	0.246376	0.701389	0.017918	0.019531

- 1) 과제 코드 파일에 작성법을 참고하여 전처리 및 변수 선택을 진행하시오.
- 2) 각자의 아이디어로 교차검증 및 하이퍼파라미터 최적화를 진행하시오.
- 3) Random Forest, Gradient Boosting, LightGBM, XGBoost, CatBoost 총 5개의 분류 모델을 구축하고 테스트 데이터의 정상 / 불량 유무를 판별하시오.
- 4) 3번에서 구축한 다섯 가지 모델의 테스트 데이터의 정상 유무를 예측해낸 결과를 아래 **양식에 맞춰** CSV 파일로 제출하시오.

RF	GBM	LightGBM	XGBoost	CatBoost
-1		
...		
1		

1. 제출 자료

- 분석 보고서 (Word 형식)
 - 분석 보고서에는 전처리, 변수선택 선정 이유와 사용한 교차검증 및 하이퍼파라미터 최적화 방법을 포함시키고, 핵심 내용만을 작성.
- 소스 코드
 - 소스 코드 미제출 시 감점 **-30점**
 - 코딩이 없는 상용 SW 시 0점 처리
- 예측값 CSV 파일
 - 파일명: 학번_이름_prediction.csv

- 예측값 미제출시 감점 **-30점**

2. 평가

- 테스트 F1 Score를 통해 최종 평가 순위를 정함 (30%)
- 분석 보고서 (70%)

3. 제출 기한

- 12월 3일 23시59분 (LearnUs에 제출 자료를 업로드 하시오).
- 이 시간 이후 제출시 감점 부여 (5분당 늦게 제출할 때마다 감점 **-5점**)
 - 예) 00시 03분 제출 시 5점 감점
 - 예) 00시 07분 제출 시 10점 감점