

PROJECT

작업자 친화적인 실시간 불량 예측 모델 개발

VotingClassifier 모델을 이용한 공정 데이터 pass, fail 분류 모델 구축

Family AI Kamp

최영민 안태영 허상호

목차

01 프로젝트 개요

02 데이터 분석 및 전처리

03 모델 학습 및 해석

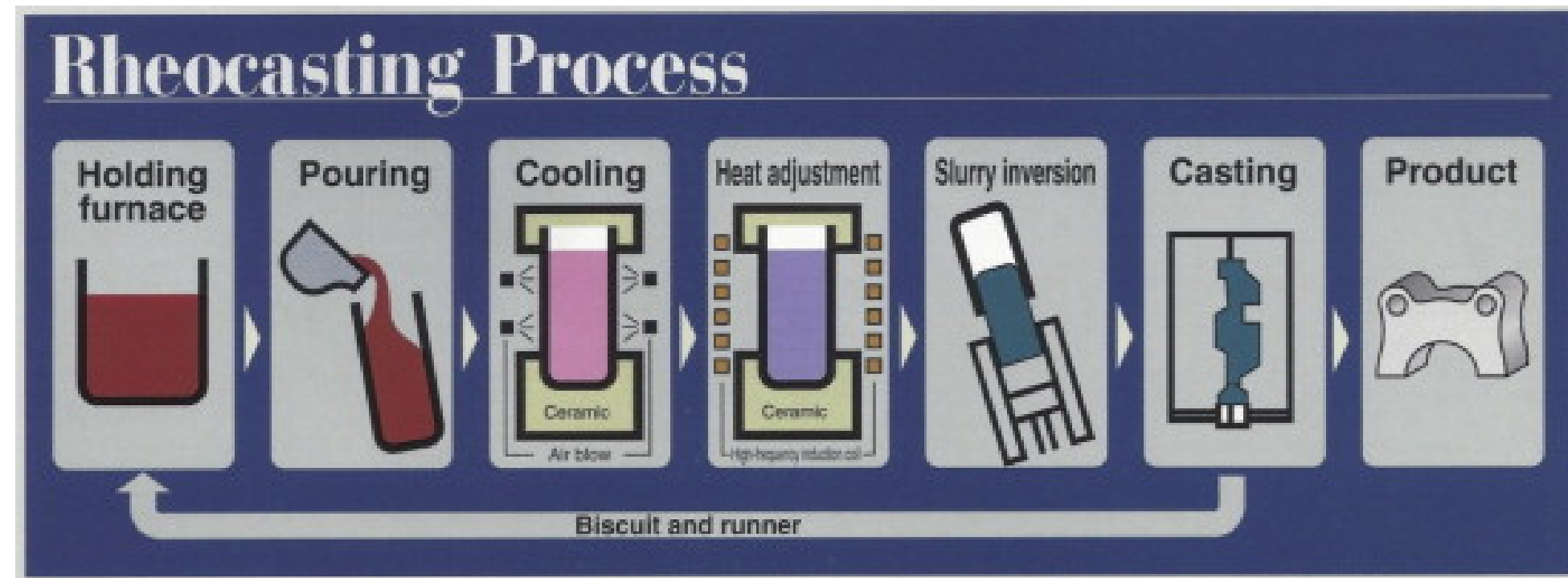
04 결론 및 시사점

01 프로젝트 개요

문제 정의

- **공정 개요** | 반고체 다이캐스팅 중 전자 교반을 활용하는 Rheocasting 방식
- **문제 현황** | 경험에 의존한 설비 운용 및 체계적인 공정 관리 시스템의 부재
불량 원인을 수작업으로 분석, 선진국 대비 낮은 인당 생산량
AI 도입 후 투자 수익률에 대한 불확실성 및 기술 이해 부족
- **목표** | AI를 활용한 작업자 친화적인 실시간 불량 예측 모델 개발

반고체 다이캐스팅 공정의 이해



Ube Industries Ltd, Marketing brochure: Japan, 1999

Rheocasting 공정

반고체 상태의 금속을 사용하여 캐스팅을 진행

장점 : 저압 금속 주입, 균일한 금속 분포, 다공성 감소, 높은 표면 품질

분석 프로세스



02 데이터분석 및 전처리

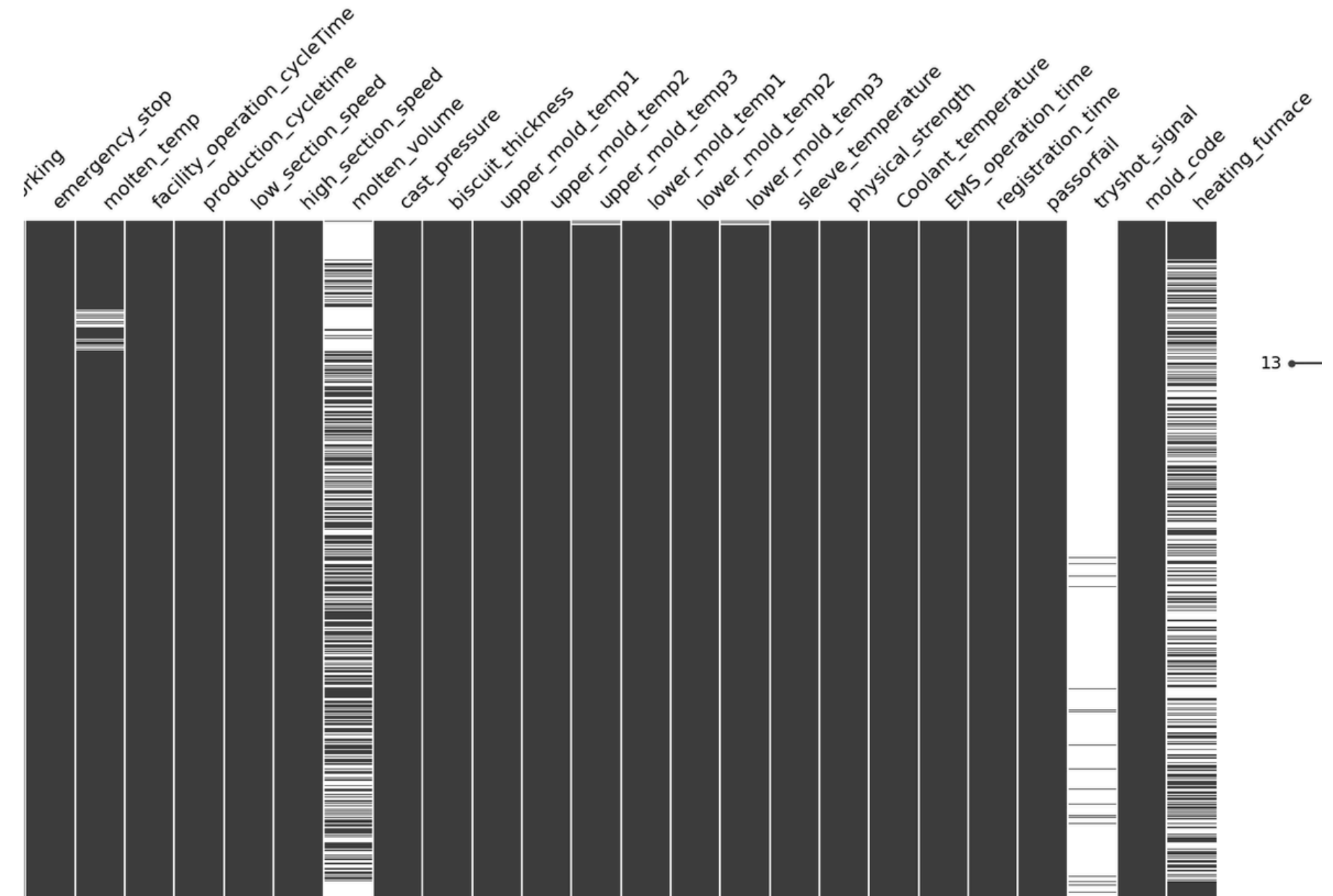
데이터 전처리

- 데이터 수
총 31개의 컬럼 중 21개 사용
92,015개 행으로 구성
- OridinalEncoder를 통한
Object 타입 인코딩

속성(column)	설명	비고
count	일자별 제품 생산 번호	int64
working	가동여부	object
melten_temp	용탕온도	float64
production_CycleTime	제품생산 사이클 시간	int64
low_section_speed	저속구간속도	float64
high_section_speed	고속구간속도	float64
cast_pressure	주조압력	float64
biscuit_thickness	비스킷 두께	float64
upper_mold_temp1	상금형온도1	float64
upper_mold_temp2	상금형온도2	float64
lower_mold_temp1	하금형온도1	float64
lower_mold_temp2	하금형온도2	float64
lower_mold_temp3	하금형온도3	float64
sleeve_temperature	슬리브온도	float64
physical_strength	형체력	float64
Coolant_temperature	냉각수 온도	float64
EMS_operation_time	전자교반 가동시간	int64
tryshot_signal	사탕신호	object
mold_code	금형코드	object
heating_furnace	가열로	object
passorfail	양품불량판정	float64

데이터 전처리

- 결측치 처리
molten_temp, tryshot_signal 등
- 결측 데이터간 관계 분석
molten_volume의 결측치가
heating_furnace와 거의 일치



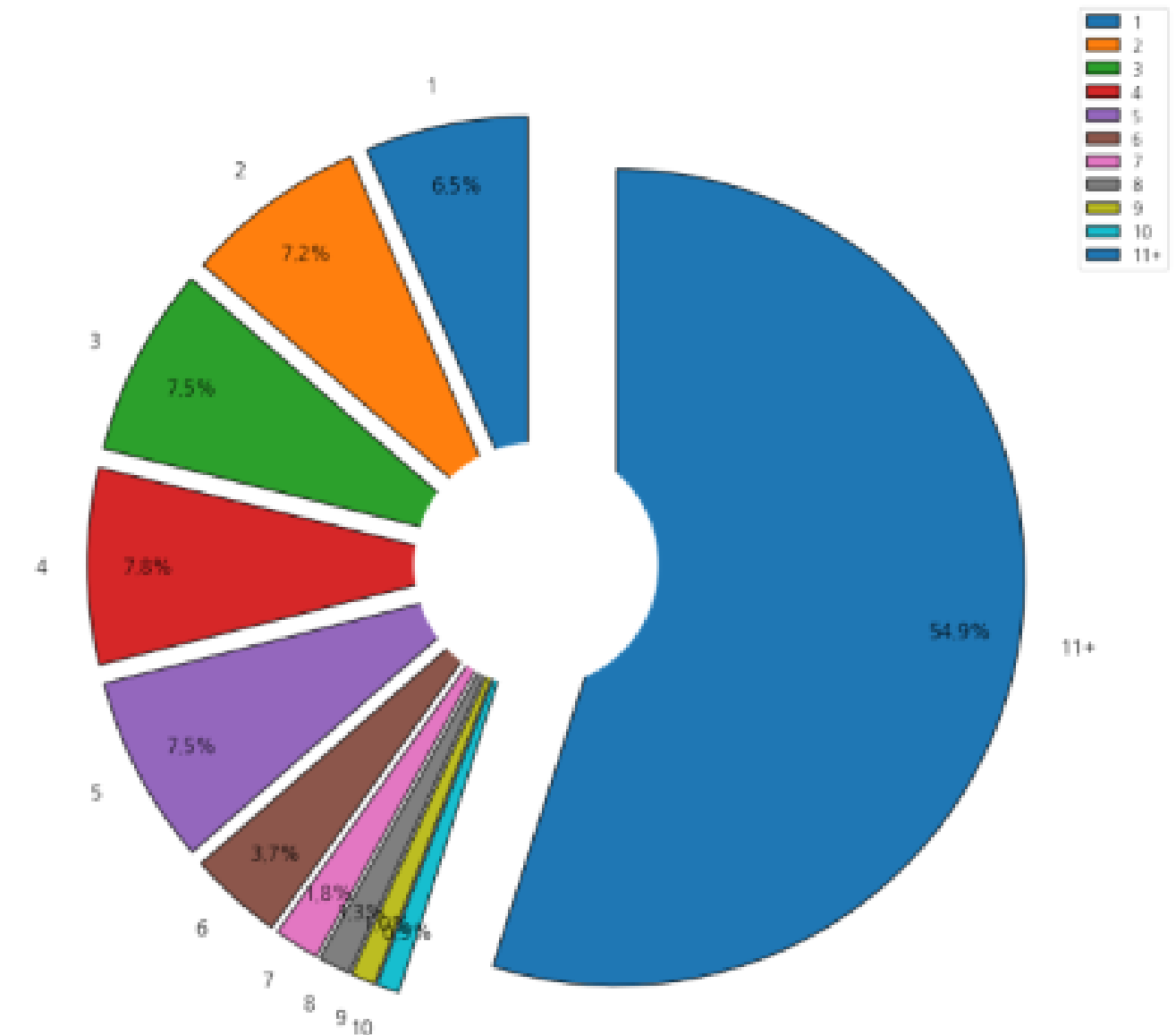
전처리 후 통계

• 데이터 분포 및 왜도

속성(column)	count	mean	std	min	50%	max	왜도
molten_temp	88569	718.7719	50.7567	0	728	735	-13.5018
production_cycletime	88569	122.2745	12.1565	0	121	462	5.5669
low_section_speed	88569	108.7370	7.9386	0	110	140	-7.4522
high_section_speed	88569	111.7759	9.2837	0	112	354	10.7329
cast_pressure	88569	324.9986	25.6672	139	330	345	-6.1023
biscuit_thickness	88569	50.5012	15.4915	0	50	422	22.0015
upper_mold_temp1	88569	186.9405	45.7599	19	195	337	-0.5306
upper_mold_temp2	88569	166.6410	28.3461	15	173	243	-0.5620
lower_mold_temp1	88569	205.3395	53.2084	20	211	367	-0.2692
lower_mold_temp2	88569	200.9323	43.5820	26	199	499	0.0845
lower_mold_temp3	88569	1441.9420	54.6192	638	1449	1449	-8.0316
sleeve_temperature	88569	417.0282	115.5642	33	458	1449	-0.7093
physical_strength	88569	699.7702	48.1839	0	703	736	-13.3081
Coolant_temperature	88569	32.1588	2.6030	17	32	48	-0.0328
EMS_operation_time	88569	18.3018	8.4260	0	23	25	-1.2893
count	88569	148.9747	87.7641	1	147	334	0.0828
working	88569	0.9997	0.0181	0	1	1	-55.2377
tryshot_signal	88569	0.0190	0.1365	0	0	1	7.0483
mold_code	88569	3.6324	2.4405	0	5	6	-0.4680
heating_furnace	88569	1.2823	0.8256	0	2	2	-0.5640
passorfail	88569	0.0397	0.1952	0	0	1	4.7159

• Count 별 fail 비율

공정 가동 초반에 fail 비율 높음



전처리 후 통계

• 데이터 품질 지수

구분	품질지수	가중치	가중치 지수	오류율		구분	품질지수	가중치	가중치 지수	오류율
완전성(누락)	97.51%	25.00%	24.377%	0.623%	➡	완전성(누락)	100.00%	25.00%	25.00%	0.00%
유일성(중복)	99.97%	10.00%	9.997%	0.003%		유일성(중복)	100.00%	10.00%	10.00%	0.00%
유효성(유효)	98.95%	15.00%	14.843%	0.158%		유효성(유효)	100.00%	15.00%	15.00%	0.00%
일관성(표현)	100%	20.00%	20%	0%		일관성(표현)	100.00%	20.00%	20.00%	0.00%
무결성	33.33%	30.00%	9.999%	20.001%		무결성	100.00%	30.00%	30.00%	0.00%
품질지수	85.95%	100.00%	79.216%	14.05%(20.784%)		품질지수	100.00%	100.00%	100.00%	0.00%

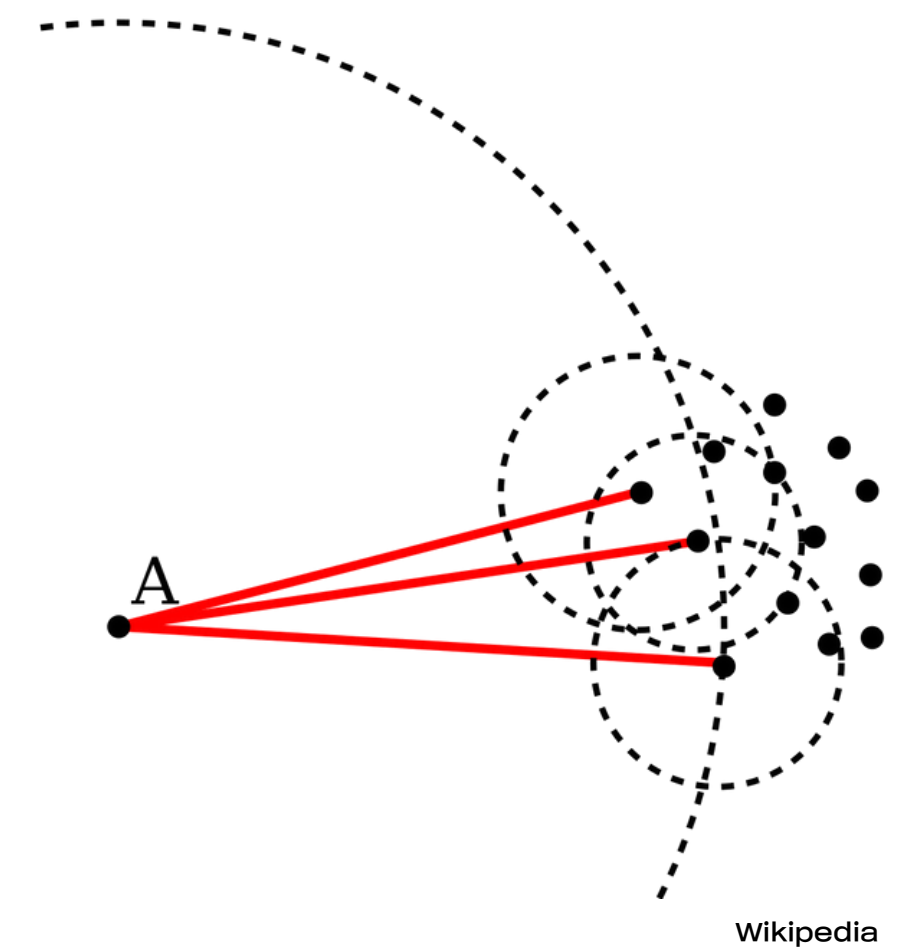
전처리 전 품질 지수

전처리 후 품질 지수

- 사용할 데이터셋의 품질을 평가
- 각 지표별 중요도 부여 : 무결성 > 완전성 > 일관성 > 유효성 > 유일성 순

이상치 제거 (LOF, T-test)

- LocalOutlierFactor(LOF)
- 데이터의 밀집도를 측정해 지역적 이상치를 찾아내는 비지도 학습 기법
- 해당 데이터는 PLC를 통해 공정 데이터를 수집하고 있으며, 순간적인 변화(Hunting)가 기록될 수 있음
- 공정상의 극단적 이상치 제거를 위해 n_neighbors=10 설정
- 양품집단과 불량집단간의 T-test
- pass와 fail 각 그룹에 대한 t-검정 시행
- p-value 0.05로 설정
- '설비작동사이클시간' 및 '상금형온도3' 컬럼 제거

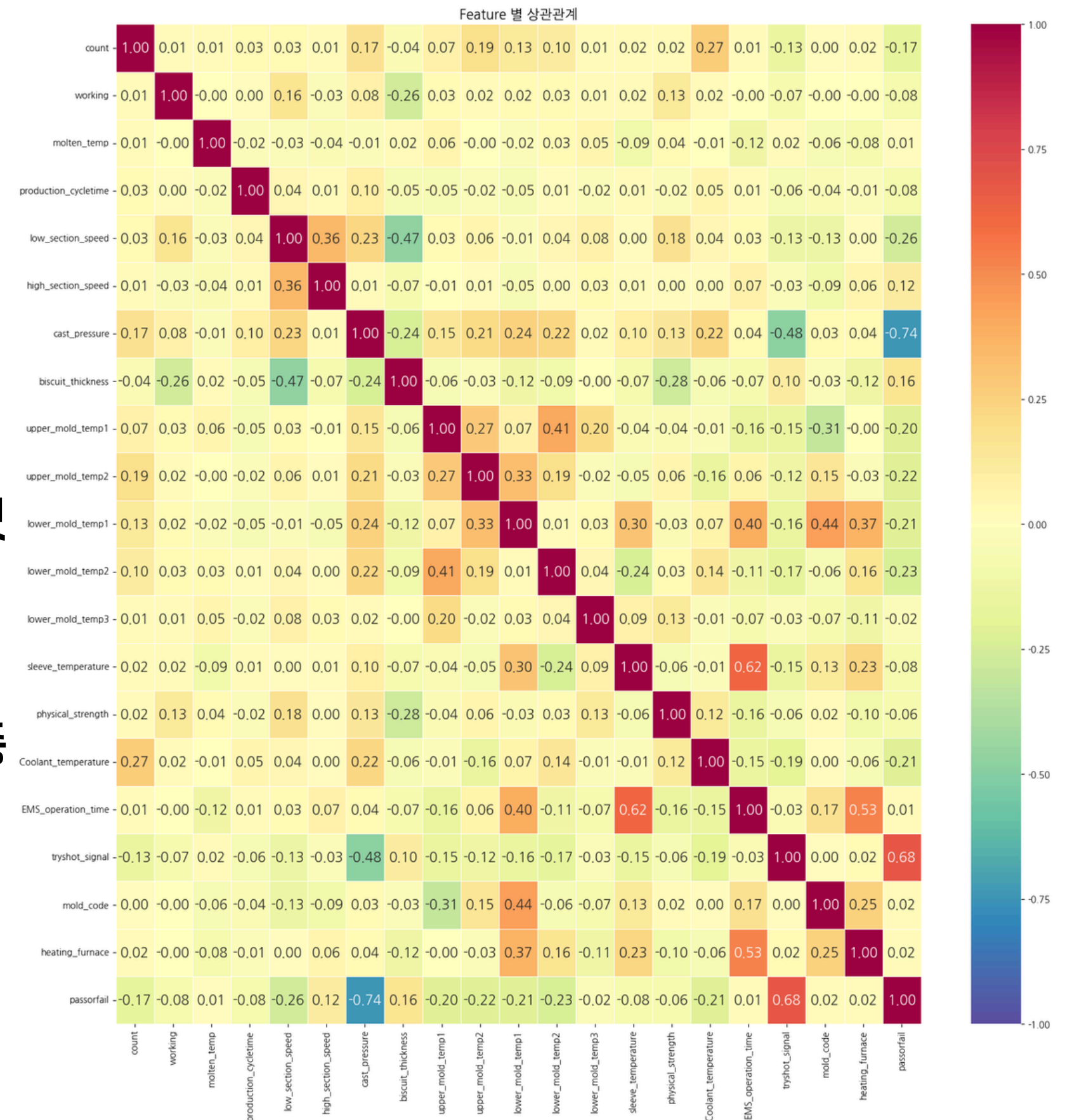


LOF의 원리

상관관계 분석

• 변수간 상관관계

- 종속변수 [passorfail]은 독립변수 [cast_pressure]와 -0.74의 강한 음의 상관관계가 나타나며,
- 독립변수 [tryshot_signal]과 0.68의 중간 정도의 양의 상관관계가 나타나고 있다.
- 종속변수 [EMS_operation_time]는 종속변수 [sleeve_temperature]와 0.62, 종속변수 [heat_furance]와 0.53의 중간 정도의 양의 상관관계가 나타나고 있다.

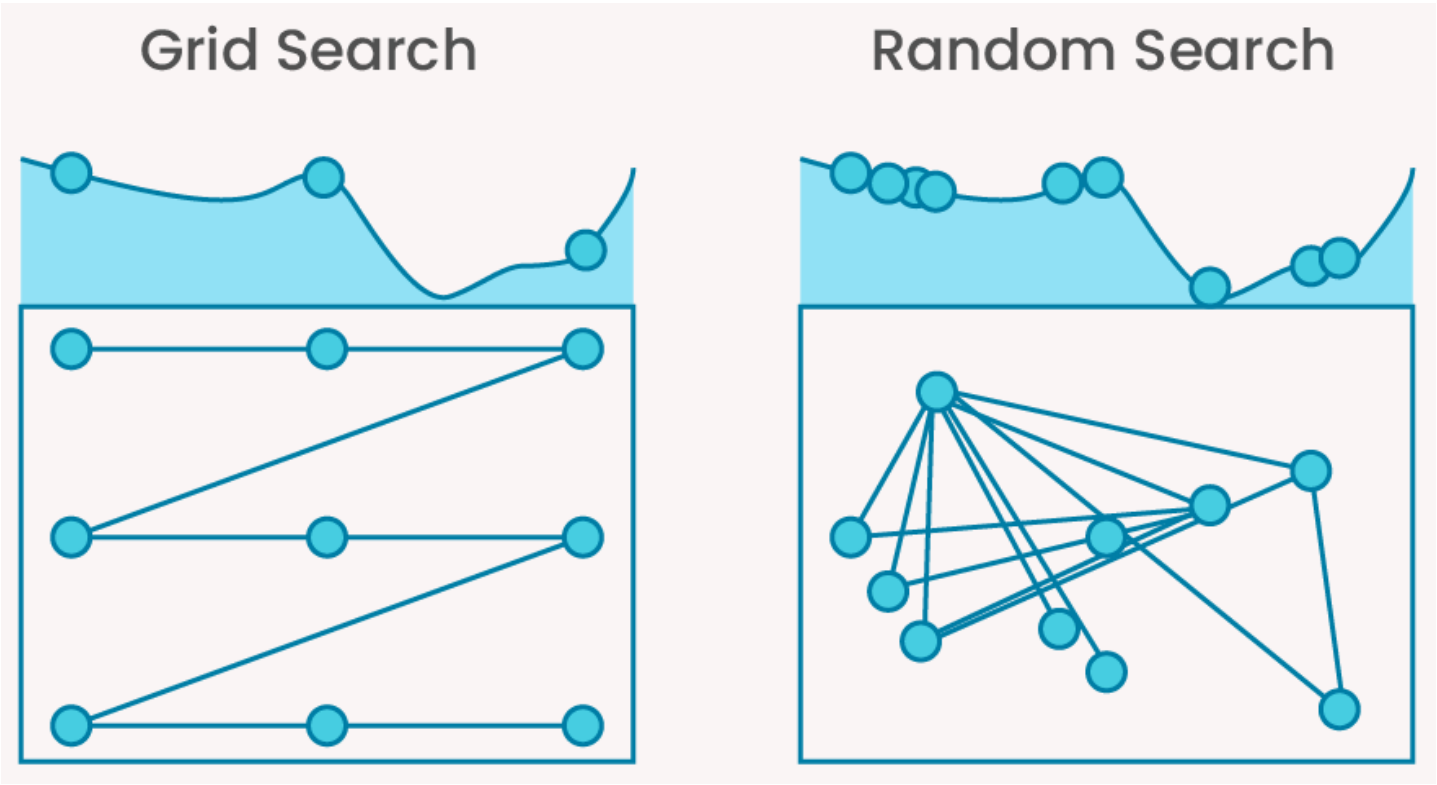


03 모델 학습 및 해석

하이퍼파라미터 튜닝

하이퍼파라미터 랜덤 서치 진행

- 랜덤 서치를 통해 각 모델 하이퍼파라미터 튜닝 진행
- 각 모델 별 최적의 하이퍼파라미터



	n_estimators	max_depth	Learning_rate	Boosting_type	l2_leaf_reg	iterations	depth
lightGBM	400	-1	0.1	gbdt	-	-	-
XGBoost	300	12	0.1	-	-	-	-
catBoost	-	-	0.1	-	5	1000	8

모델 학습 결과

XGBoost

```

classification report :
              precision    recall  f1-score   support

      0.0           1.00       1.00       1.00      17011
      1.0           0.97       0.94       0.95        703

   accuracy                   1.00      17714
  macro avg           0.98       0.97       0.98      17714
 weighted avg           1.00       1.00       1.00      17714

```

CatBoost

```

classification report :
              precision    recall  f1-score   support

      0.0           1.00       1.00       1.00      17011
      1.0           0.97       0.93       0.95        703

   accuracy                   1.00      17714
  macro avg           0.99       0.97       0.98      17714
 weighted avg           1.00       1.00       1.00      17714

```

LightGBM

```

classification report :
              precision    recall  f1-score   support

      0.0           1.00       1.00       1.00      17011
      1.0           0.98       0.95       0.96        703

   accuracy                   1.00      17714
  macro avg           0.99       0.97       0.98      17714
 weighted avg           1.00       1.00       1.00      17714

```

VoatingClassifier

```

classification report :
              precision    recall  f1-score   support

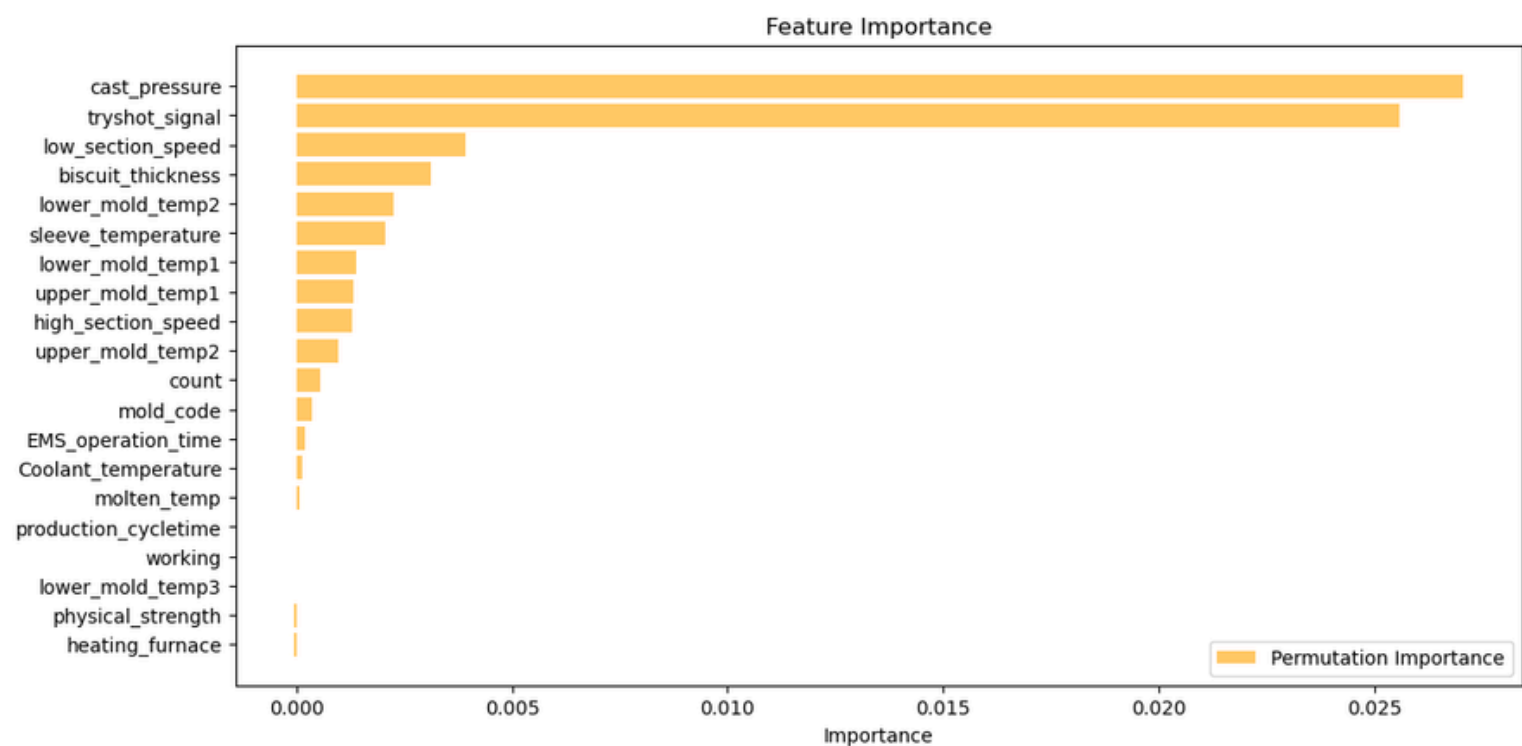
      0.0           1.00       1.00       1.00      17011
      1.0           0.98       0.95       0.97        703

   accuracy                   1.00      17714
  macro avg           0.99       0.97       0.98      17714
 weighted avg           1.00       1.00       1.00      17714

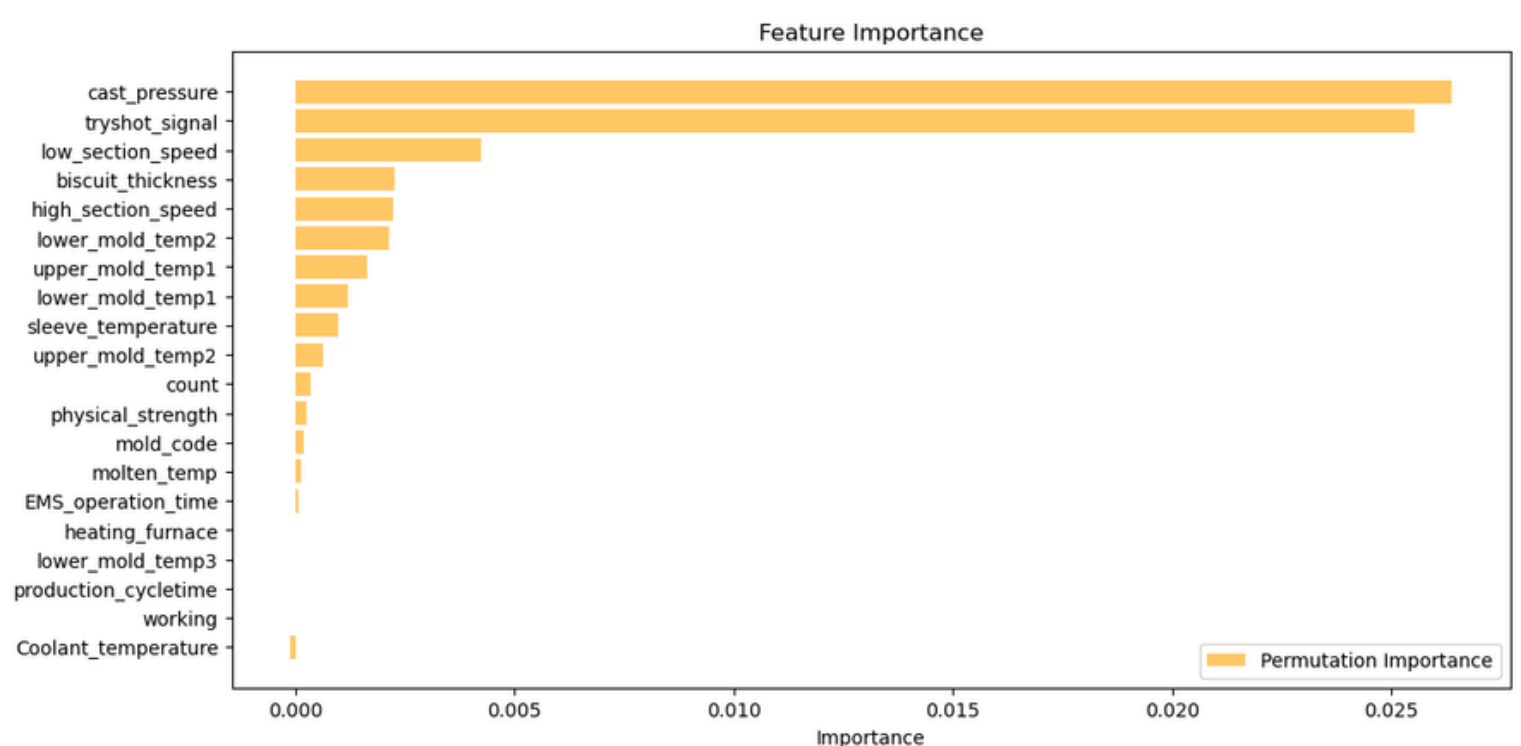
```


모델 학습 결과

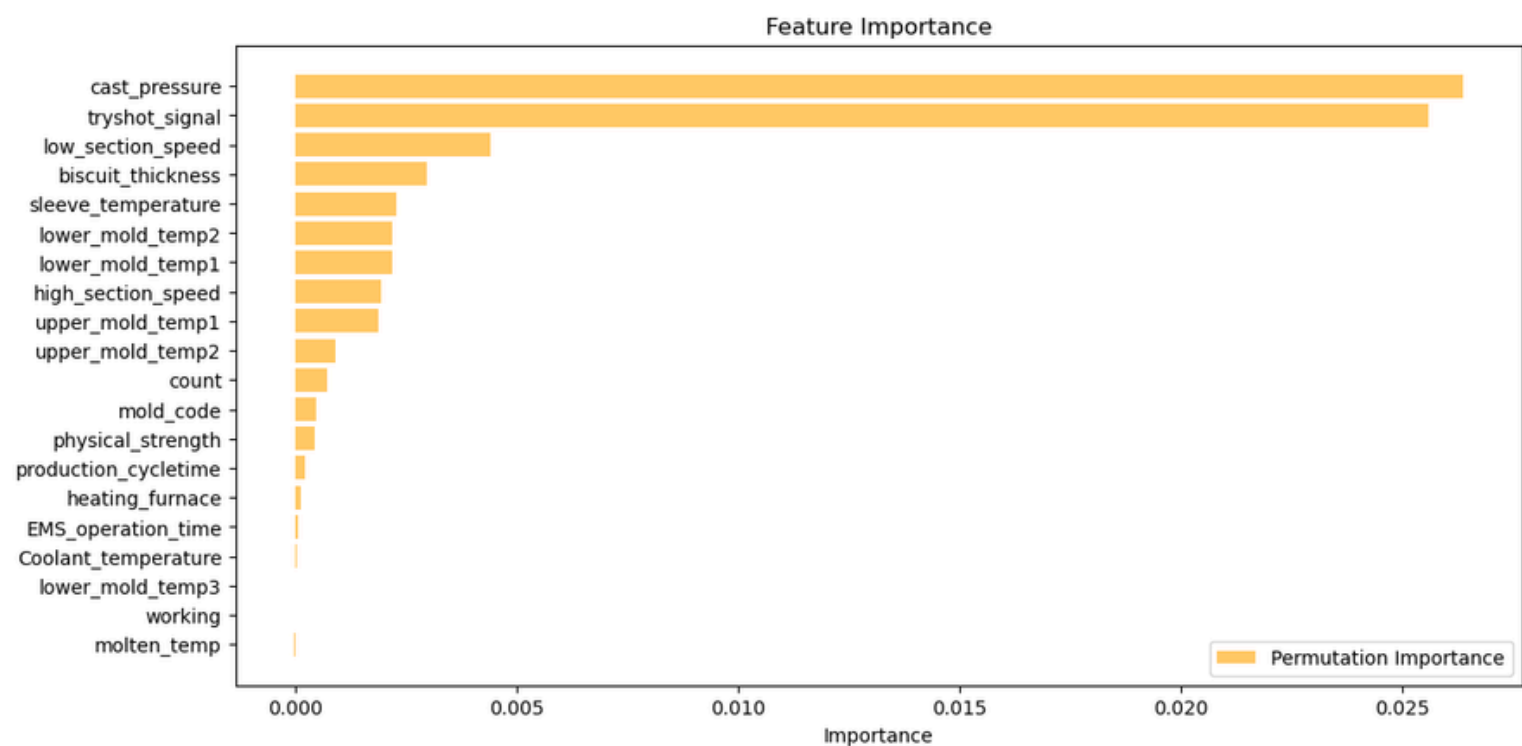
XGBoost



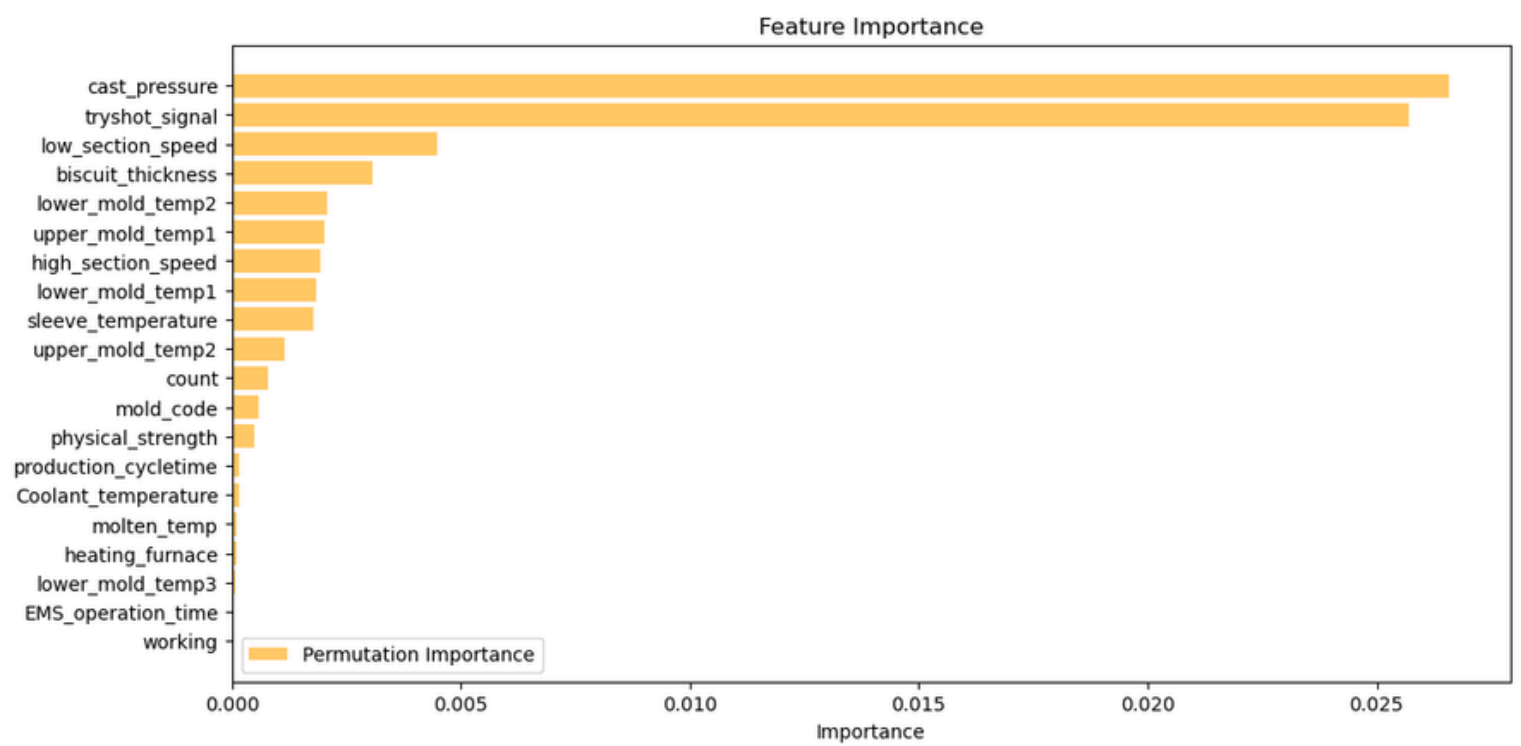
CatBoost



LightGBM

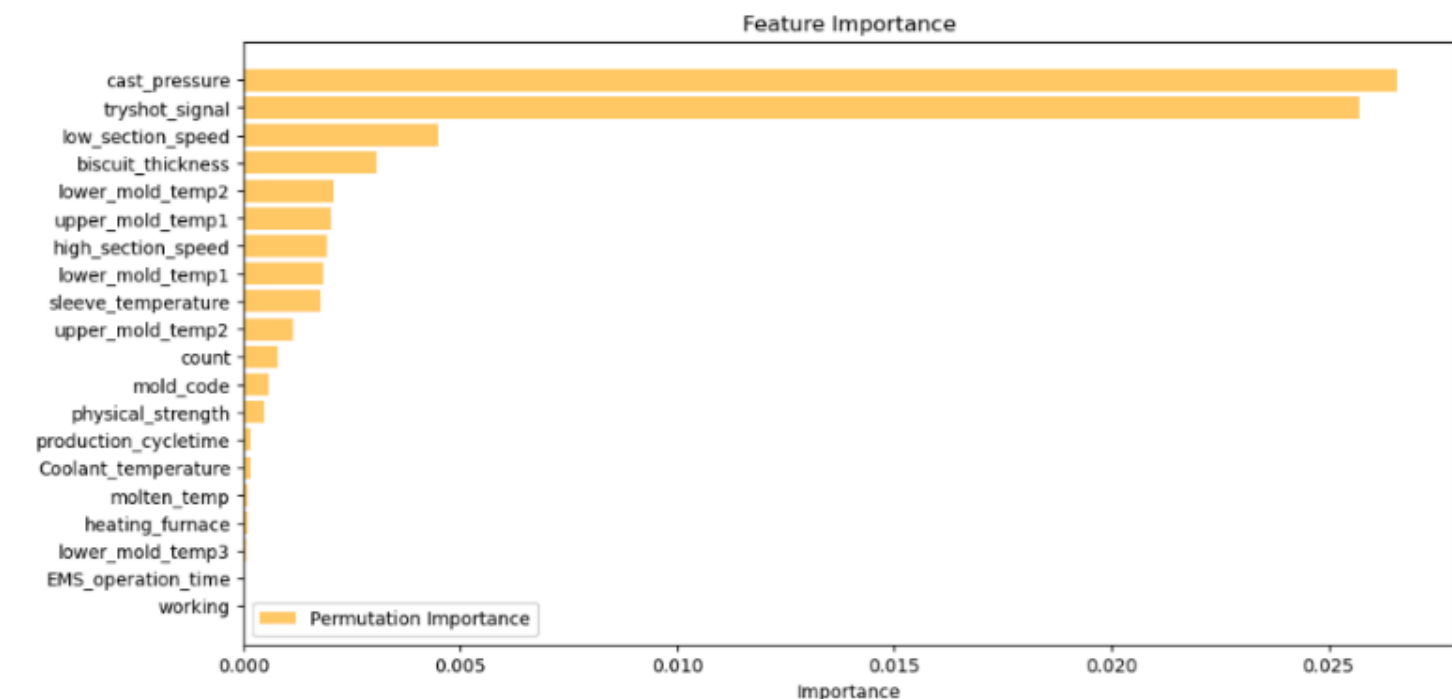
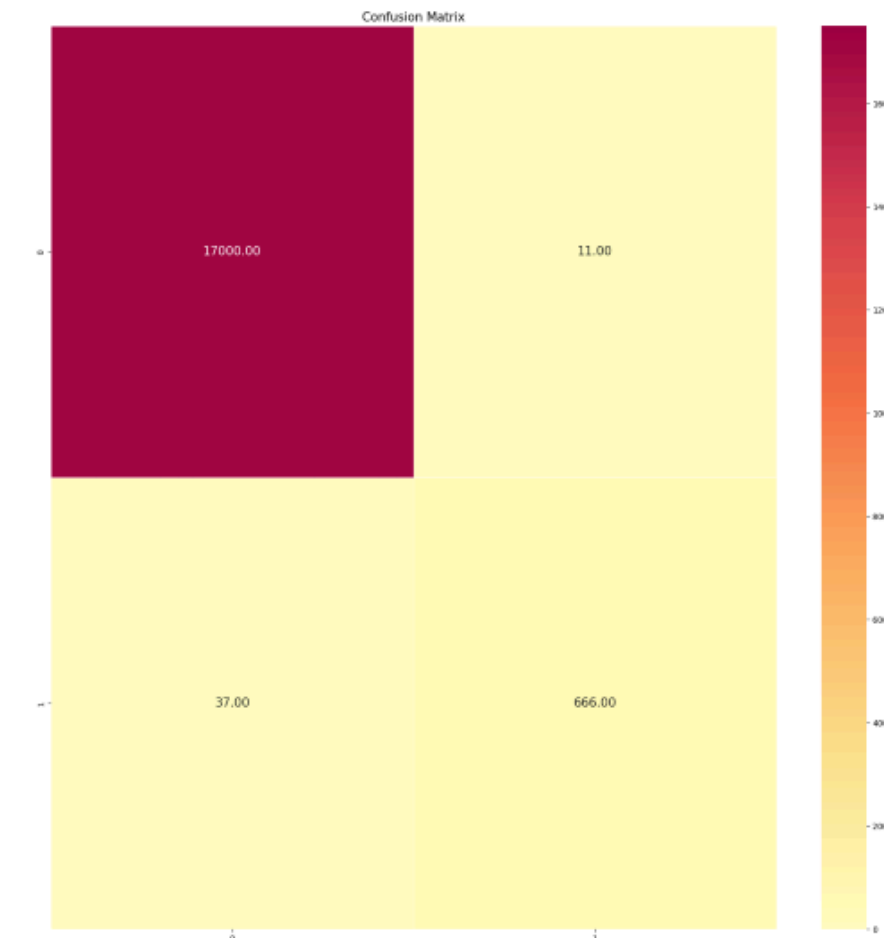


VoatingClassifier



모델 학습 결과

- VoatingClassifier 구성
- 가중치
- catboost : 1.0, lgbm : 2.5, xgb : 1.0
- Feature Importance
- cast_pressure, try_shot_signal이 0.025 이상으로 가장 큰 영향
- 불합격에 대한 F1-Score
- 0.965



04 결론 및 시사점

AI 데이터셋 분석 시사점

- 최종 모델의 분석 결과, 불량 예측에 많은 영향을 미치는 주요 인자들은 Permutation Importance 기법을 통해 확인할 수 있다. 이 분석을 통해 공정에서 관리 가능한 변수로는 cast_pressure, low_section_speed, biscuit_thickness, upper_mold_temp1 등이 중요한 영향을 미친다는 것을 확인할 수 있었다.
- 또한, 공정 또는 장비 문제를 의미하는 tryshot_signal 변수가 불량 예측에 상당한 영향을 주는 중요한 요소로 나타났다. 이러한 분석 결과를 바탕으로, 공정에서는 cast_pressure 변수에 대한 세밀한 통제가 필요하며, 이를 위한 철저한 관리 방법을 도입이 필요하다.
- 제조 공정의 다른 장비에 AI 분석을 적용하면, 변수별 수준에 따른 장비의 차이를 확인할 수 있다. 이러한 방법은 눈에 보이지 않는 변화를 빠르게 감지할 수 있어, 문제를 해결을 위한 인력 투입을 효율적으로 절감할 수 있을 것으로 예상된다.

AI 데이터셋 분석 시사점

- Permutation Importance 기법을 통해 확인한 결과, count 변수는 불량 예측에 중요한 영향을 미치지 않는 것으로 나타났다. 그러나 데이터를 분석한 결과, 장비 재가동 후 안정화까지 5~6회의 작업이 필요한 것을 확인하였다. 이를 바탕으로, 장비의 정상화를 위한 Dummy 작업을 진행한다면 비용 절감과 함께 total loss를 줄일 수 있을 것으로 예상된다.
- 장비상의 PLC에서 기록되는 데이터를 이용하여 분석을 진행하였다. 향후 각종 불량 유형을 이후 공정에서 라벨링하게 된다면, 이를 통해 불량 원인을 보다 정확하게 판단할 수 있을 것이라 예상된다.

중소제조기업에 미치는 파급효과

- 손쉬운 사용을 통해 다양한 공정에 적용
 - 제조 현장에서는 공정에 대한 지식은 있지만 AI에 대한 이해가 부족해 AI 도입이 어려운 상황입니다. 이를 해결하기 위해, 본 팀은 코드를 함수화하여 사용자가 코딩 지식 없이도 변수만 변경하면 다양한 공정에 쉽게 적용할 수 있도록 프로그램을 개발했습니다.
- 업무 환경 개선
 - 품질 관리를 사람의 손이 아닌 공정 데이터를 기반으로 예측함으로써, 품질 관리와 유지보수 효율이 개선되고 생산성 향상이 기대됩니다. 또한, 중소기업의 제조 경쟁력을 강화하고 지속적인 성장을 이끌어낼 수 있습니다.

중소제조기업에 미치는 파급효과

- **생산성 증가 및 품질 관리 비용 절감**
 - 공정 최적화를 통해 생산 비용을 절감하고 출시 기간을 단축할 수 있습니다. 기계 센서 데이터를 활용한 AI 기반 품질 검사는 불량률을 줄이고 검사 시간을 단축할 수 있습니다. 또한, 설비 고장 패턴 분석을 통해 고장을 사전에 감지하고 예방할 수 있어 유지보수 비용 절감이 가능합니다. AI 분석을 통해 불량 공정의 원인을 경험이 아닌 데이터로 기반으로 찾아내어, 시행착오를 줄이고 공정 효율성 및 제품 개발 비용 절감에 기여할 것입니다.