

ARIMA 모델링 보고서

자기 상관 및 차분 계수에 대한 판단 시 해당 경향성이 공정의 내재적인 부분인지 어떤 이상 원인의 결과인지 알기 위해 훈련 데이터셋의 각 금형코드 별(8412, 8573, 8600, 8722, 8917)로 정상적으로 생산되었던 사이클(~300개)을 각각 두 개씩 추출하여 표본으로 활용.

(한 구간만에서 관측된 자기상관은 그 순간의 이상원인 개입 때문일 수 있기 때문)

1.변수별 자기상관 판단

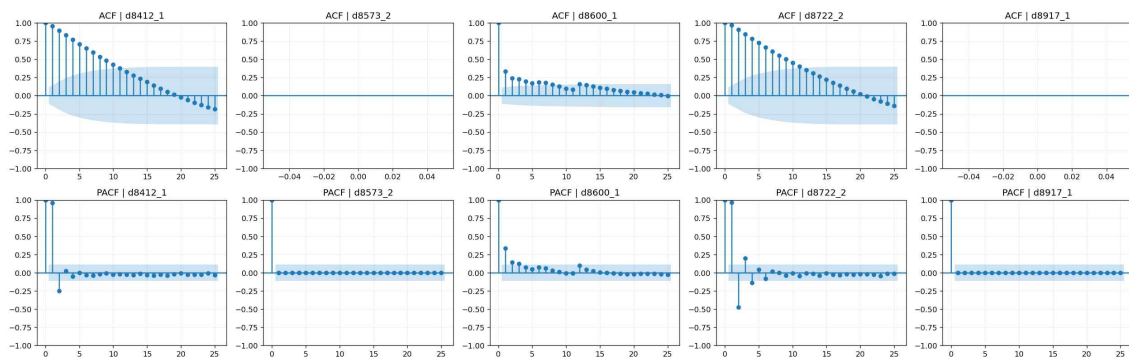
전통적인 슈하르트 관리도의 가장 중요한 가정은 각 관측값들 간의 독립성

특히 데이터가 양의 상관관계가 있을 경우 많은 오경보 발생

이에 따라 금형 코드별 5개 구간 모두에 대해 ACF(자기상관함수), PACF(편자기상관함수)의 그래프를 시차 25구간까지 도출하여 판단에 활용

-용탕량 molten_volume: 자기상관 0

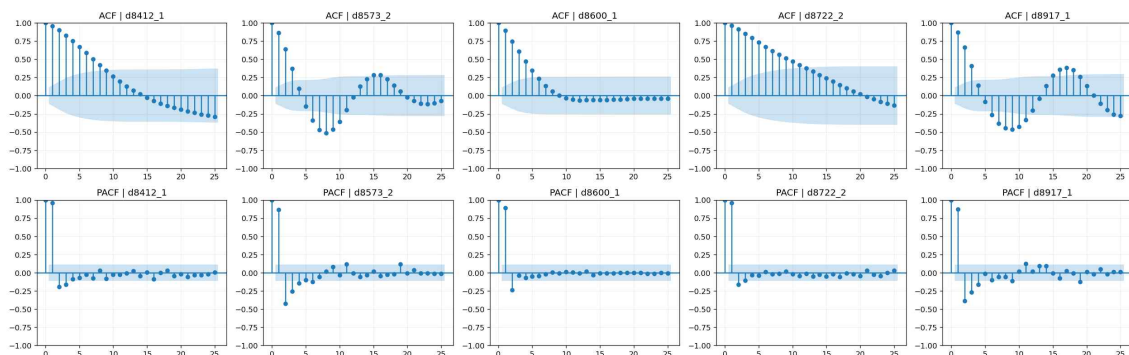
[molten_volume] ACF & PACF (lags=25)



-> 3구간에서 자기 상관 및 비정상성 (ACF 추세)

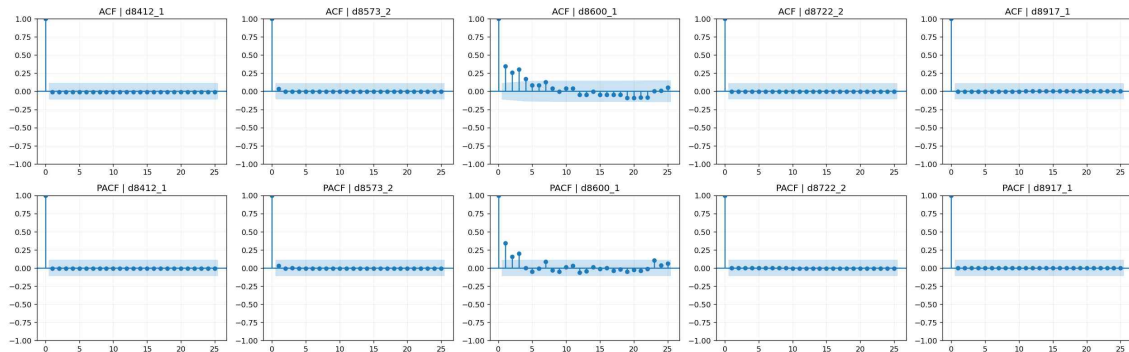
-용탕온도 molten_temp: 자기상관 0(모든 구간 자기상관, 비정상, 주기성)

[molten_temp] ACF & PACF (lags=25)



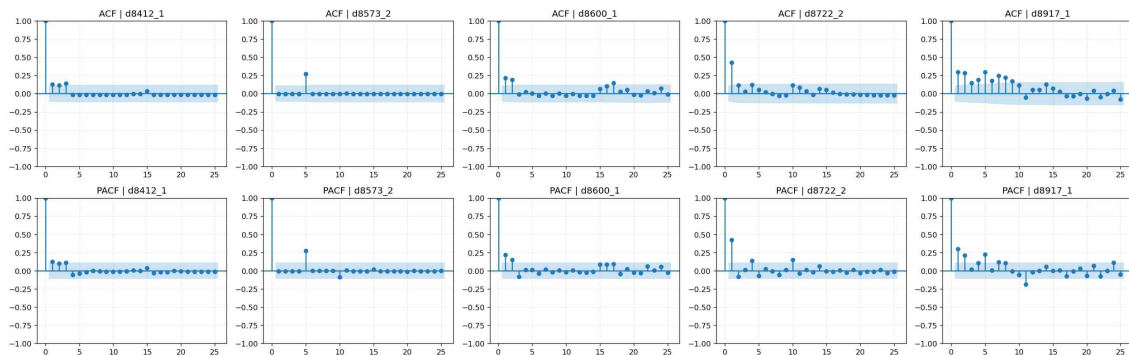
-저속구간속도 low_section_speed: 자기상관 X(4구간에서 자기상관 없음)

[low_section_speed] ACF & PACF (lags=25)



-고속구간속도 high_section_speed: 자기상관 0 (Ljung-Box 검정 실행)

[high_section_speed] ACF & PACF (lags=25)



=== Ljung-Box 결과 (lag=5) : high_section_speed ===

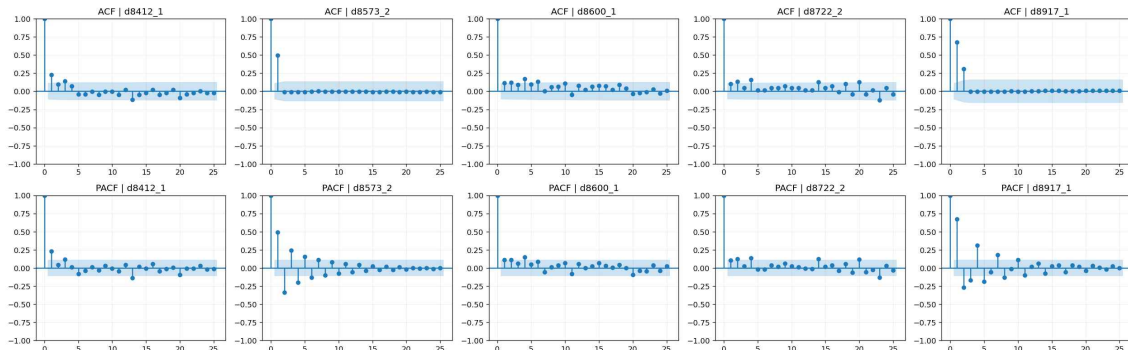
dataset	lag	lb_stat	lb_pvalue	decision
d8412_1	5	14.064191	1.520677e-02	reject_H0_autocorr_present
d8573_2	5	22.942083	3.462812e-04	reject_H0_autocorr_present
d8600_1	5	25.314576	1.211361e-04	reject_H0_autocorr_present
d8722_2	5	65.312822	9.651813e-13	reject_H0_autocorr_present
d8917_1	5	97.039813	2.221630e-19	reject_H0_autocorr_present

=== Ljung-Box 결과 (lag=10) : high_section_speed ===

dataset	lag	lb_stat	lb_pvalue	decision
d8412_1	10	14.549392	1.493741e-01	fail_to_reject_H0_no_evidence
d8573_2	10	22.988553	1.078891e-02	reject_H0_autocorr_present
d8600_1	10	26.005341	3.733009e-03	reject_H0_autocorr_present
d8722_2	10	69.991966	4.449646e-11	reject_H0_autocorr_present
d8917_1	10	154.575486	4.259473e-28	reject_H0_autocorr_present

-주조압력 cast_pressure: 자기상관 0 (Ljung-Box 검정 실행)

[cast_pressure] ACF & PACF (lags=25)



=== Ljung-Box 결과 (lag=5) : cast_pressure ===

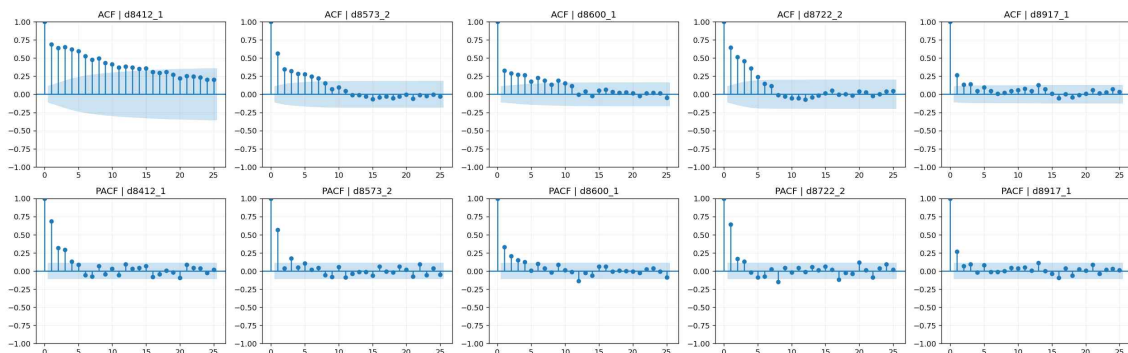
dataset	lag	lb_stat	lb_pvalue	decision
d8412_1	5	26.223235	8.076544e-05	reject_H0_autocorr_present
d8573_2	5	74.550115	1.154771e-14	reject_H0_autocorr_present
d8600_1	5	22.511230	4.184773e-04	reject_H0_autocorr_present
d8722_2	5	17.071710	4.365786e-03	reject_H0_autocorr_present
d8917_1	5	170.846459	4.813868e-35	reject_H0_autocorr_present

=== Ljung-Box 결과 (lag=10) : cast_pressure ===

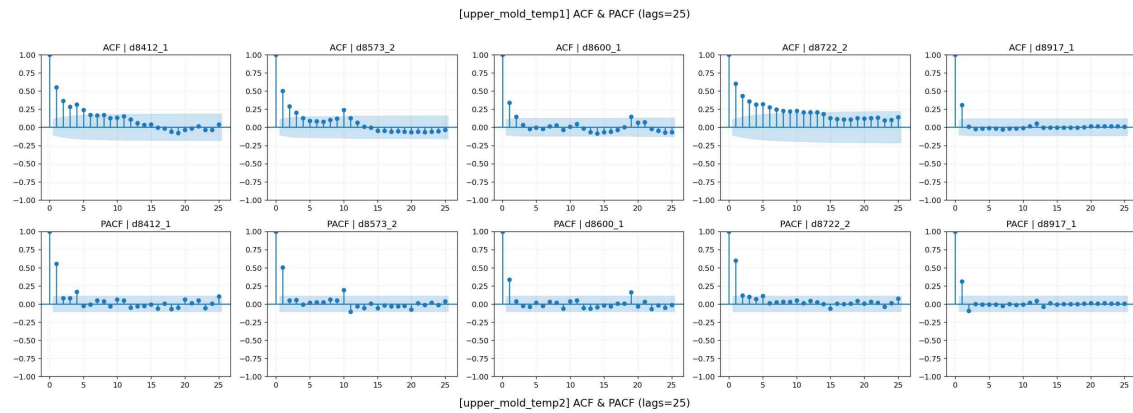
dataset	lag	lb_stat	lb_pvalue	decision
d8412_1	10	27.446449	2.212462e-03	reject_H0_autocorr_present
d8573_2	10	74.579317	5.744588e-12	reject_H0_autocorr_present
d8600_1	10	34.162502	1.733486e-04	reject_H0_autocorr_present
d8722_2	10	20.772113	2.273989e-02	reject_H0_autocorr_present
d8917_1	10	170.852789	1.847212e-31	reject_H0_autocorr_present

-비스킷 두께 biscuit_thickness: 자기상관 0 (모든 구간에서 관측)

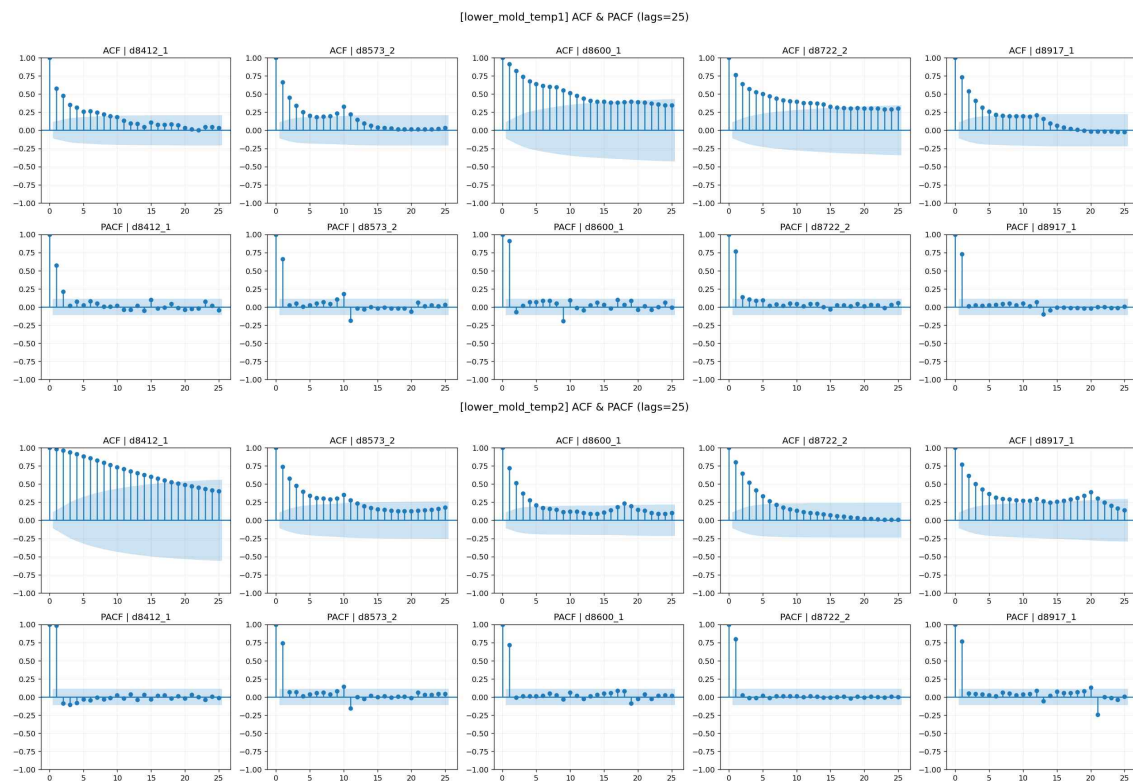
[biscuit_thickness] ACF & PACF (lags=25)



-상금형 온도1,2 upper_mold_temp1, 2: 자기상관 0 (모든 구간에서 관측)

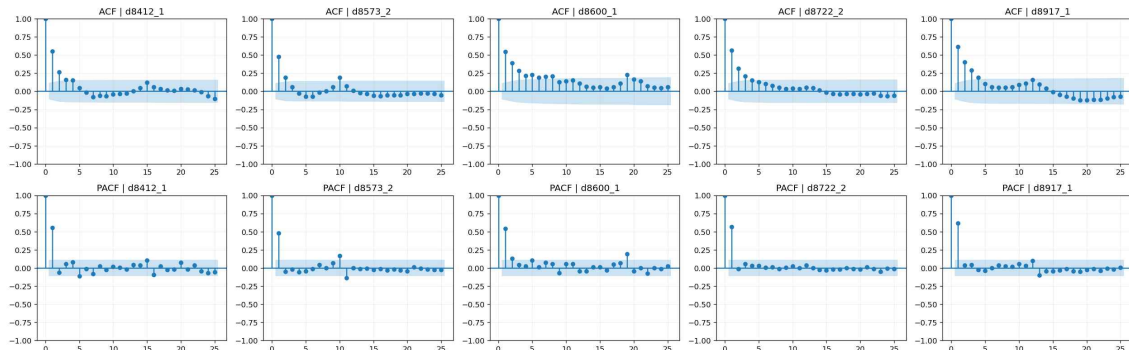


-하금형 온도1, 2 lower_mold_temp1, 2: 자기상관 0 (모든 구간에서 관측)



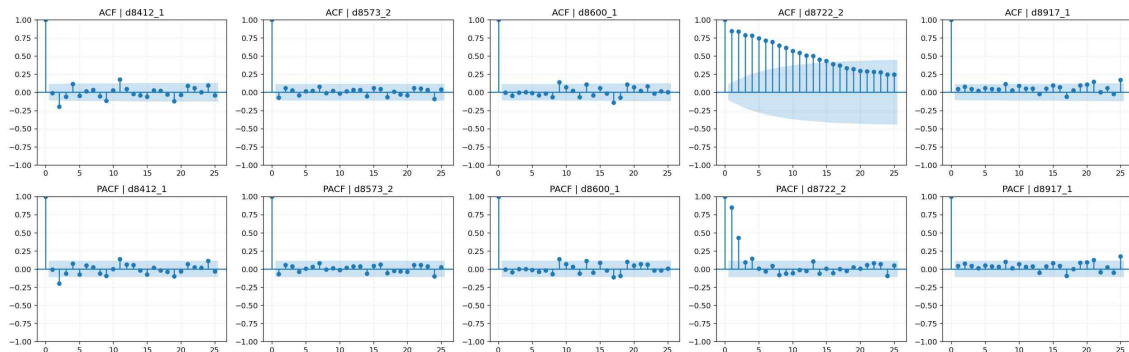
-슬리브온도 sleeve_temperature: 자기상관 0 (모든 구간에서 관측)

[sleeve_temperature] ACF & PACF (lags=25)



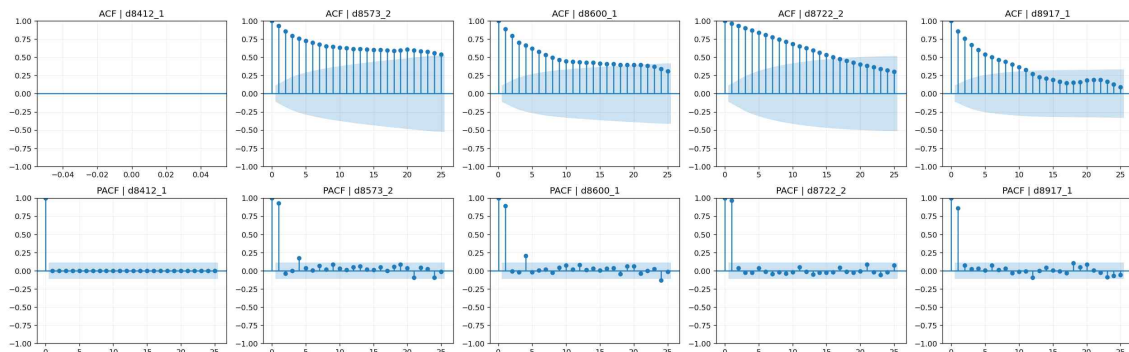
-형체력 physical_strength: 자기상관 X (한 구간 제외 관측 X)

[physical_strength] ACF & PACF (lags=25)



-냉각수 온도 Coolant_temperature: 자기상관 0 (한 구간 제외 관측)

[Coolant_temperature] ACF & PACF (lags=25)



결과: 저속구간속도, 형체력 변수 제외 11개 변수 공정 내재적 자기상관 관측

-> 11개 변수에 대해서는 ARIMA+잔차 관리도,

저속구간속도, 형체력 변수에 대해서는 기본 슈하르트 관리도 중 하나인 X-R 관리도

2.변수별 정상성 판단

1차적으로 위에서 도출한 ACF 및 PACF 그래프를 통해 비정상 시계열의 패턴(ACF가 느리게 추세를 띄며 감소, PACF가 특정 시차 이상부터 빠르게 0으로 수렴X)을 띄지 않는 변수들을 정상 시계열로 판단

-정상 시계열 판단 변수: 주조압력, 고속구간속도, 비스켓 두께, 슬리브 온도

-> 해당 변수에 대해서는 차분 진행하지 않고, ARMA(p,q) 모델로 결정

-비정상 시계열 판단 변수:

용탕량, 용탕온도, 상금형온도1~2, 하금형온도1~2, 냉각수온도

-> 해당 변수에 대해서는 1차 차분 진행 후 단위근 검정(ADF 검정) 진행

2.1) 비정상 시계열 1차 차분 이후 단위근 검정 진행 결과

dataset	variable	n(D1)	ADF_stat	ADF_p	판정
d8412_1	molten_volum e	292	-13.2486	8.87E-25	정상
d8573_2	molten_volum e	301			계산불가
d8600_1	molten_volum e	299	-8.11073	1.23E-12	정상
d8722_2	molten_volum e	304	-6.93146	1.08E-09	정상
d8917_1	molten_volum e	306			계산불가
d8412_1	molten_temp	292	-9.5986	1.95E-16	정상
d8573_2	molten_temp	301	-9.64896	1.45E-16	정상
d8600_1	molten_temp	299	-14.4935	6.09E-27	정상
d8722_2	molten_temp	304	-14.7976	2.12E-27	정상
d8917_1	molten_temp	306	-9.94009	2.68E-17	정상
d8412_1	biscuit_thickn ess	292	-13.1314	1.49E-24	정상
d8573_2	biscuit_thickn ess	301	-10.4956	1.12E-18	정상
d8600_1	biscuit_thickn ess	299	-8.9167	1.07E-14	정상
d8722_2	biscuit_thickn ess	304	-7.51948	3.82E-11	정상
d8917_1	biscuit_thickn ess	306	-8.84583	1.62E-14	정상
d8412_1	upper_mold_t emp1	292	-8.18479	7.95E-13	정상
d8573_2	upper_mold_t emp1	301	-10.7435	2.78E-19	정상
d8600_1	upper_mold_t emp1	299	-9.30694	1.07E-15	정상
d8722_2	upper_mold_t emp1	304	-8.91521	1.08E-14	정상
d8917_1	upper_mold_t emp1	306	-7.98546	2.56E-12	정상

d8412_1	upper_mold_t emp2	292	-7.94579	3.22E-12	정상
d8573_2	upper_mold_t emp2	301	-7.40779	7.27E-11	정상
d8600_1	upper_mold_t emp2	299	-8.87906	1.33E-14	정상
d8722_2	upper_mold_t emp2	304	-10.8066	1.96E-19	정상
d8917_1	upper_mold_t emp2	306	-8.15539	9.45E-13	정상
d8412_1	lower_mold_t emp1	292	-10.4145	1.77E-18	정상
d8573_2	lower_mold_t emp1	301	-7.68186	1.5E-11	정상
d8600_1	lower_mold_t emp1	299	-10.1392	8.51E-18	정상
d8722_2	lower_mold_t emp1	304	-8.50738	1.19E-13	정상
d8917_1	lower_mold_t emp1	306	-8.25876	5.15E-13	정상
d8412_1	lower_mold_t emp2	292	-5.15687	1.08E-05	애매
d8573_2	lower_mold_t emp2	301	-7.64907	1.81E-11	정상
d8600_1	lower_mold_t emp2	299	-8.71241	3.56E-14	정상
d8722_2	lower_mold_t emp2	304	-12.0052	3.28E-22	정상
d8917_1	lower_mold_t emp2	306	-8.1377	1.05E-12	정상
d8412_1	sleeve_tempe rature	292	-8.28691	4.37E-13	정상
d8573_2	sleeve_tempe rature	301	-11.3225	1.17E-20	정상
d8600_1	sleeve_tempe rature	299	-8.5154	1.14E-13	정상
d8722_2	sleeve_tempe rature	304	-10.4067	1.85E-18	정상
d8917_1	sleeve_tempe rature	306	-8.22747	6.19E-13	정상
d8412_1	Coolant_temp erature	292			계산불가
d8573_2	Coolant_temp erature	301	-9.30596	1.08E-15	정상
d8600_1	Coolant_temp erature	299	-13.4897	3.12E-25	정상
d8722_2	Coolant_temp erature	304	-8.32625	3.46E-13	애매
d8917_1	Coolant_temp erature	306	-10.3225	2.98E-18	정상

결과: 대부분 생산 사이클 구간에서 모든 변수 1차 차분 이후 정상 시계열로 판정

ARIMA(p,1,q) : 용탕량, 용탕온도, 상금형온도1~2, 하금형온도1~2, 냉각수온도 변수

3. AUTO-ARIMA 모듈 활용하여 p,q 차수 결정

앞서 구한 변수별 차분 차수를 고정시키고 AR(p), MA(q) 모델의 차수를 결정짓기 위해 정상성 및 자기상관 판단에 활용되었던 10개의 데이터셋에 auto-arima 모듈을 적용시켜 AIC, BIC를 기준으로 각각 최적의 차수를 결정한 후 voting(평균)을 통해 11개의 변수에 대해 각각 하나의 최적 ARIMA(p,d,q) 차수를 결정하였다.

	variable	avg_p	avg_d	avg_q	opt_p	opt_d	opt_q
0	Coolant_temper	1	1	2	1	1	2
1	biscuit_thickness	1.4	0	1.4	1	0	1
2	cast_pressure	0.4	0	0.8	0	0	1
3	high_section	0.6	0	1.2	1	0	1
4	lower_mold_temp2	1	1	4.6	1	1	5
5	lower_mold_temp1	1	1	5	1	1	5
6	mold_temperature	1	1	3	1	1	3
7	molten_viscosity	0.4	1	2	0	1	2
8	sleeve_temperature	1	0	1	1	0	1
9	upper_mold_temp1	1.4	1	5	1	1	5
10	upper_mold_temp2	1	1	5	1	1	5

-> 해당 차수를 기준으로 가장 최근의 생산 사이클에서 정상적인 사이클 구간을 추출한 후 ARIMA 모델 적합. (5개 금형코드, 11개 변수, 총 55개 모델 적합)

*금형코드 8413, 8576의 데이터는 test셋에서 가장 초반부 60개를 추출하여 적합하고, 실제 스트리밍 할 데이터에서는 해당 샘플들 제거

4. 잔차 독립성 검정(Ljung-Box Test)

모델 적합 이후, 각 모델 별로 각각 용-박스 검정을 진행하여 ARIMA 모델 활용을 통해 자기상관성을 제거한 후에도 여전히 잔차 간 자기상관을 갖는지 검정을 진행하였다.

ld_code	del_file	variable	prde	rde	rde	n_o	aic	bic	model_p	lb_lag	all_lags	(alpha=0.05)
d8412_4	a8412_low	lower_mold_temp2	1	1	5	297	2179.093848	2186.475	./saved_n	24	FALSE	
d8413_4	a8413_cas	cast_pressure	0	0	1	65	569.5858264	573.9036	./saved_n	13	FALSE	
d8413_4	a8413_low	lower_mold_temp2	1	1	5	65	409.4239978	413.7418	./saved_n	13	FALSE	
d8413_4	a8413_up	upper_mold_temp1	1	1	5	65	553.4188457	562.1164	./saved_n	13	FALSE	
d8576_4	a8576_up	upper_mold_temp1	1	1	5	56	534.4540045	542.5554	./saved_n	11	FALSE	
d8722_4	a8722_cas	cast_pressure	0	0	1	282	2437.636432	2444.913	./saved_n	24	FALSE	
d8722_4	a8722_up	upper_mold_temp1	1	1	5	282	1703.482087	1718.05	./saved_n	24	FALSE	

결과: 주조압력, 상금형온도1, 하금형온도 변수의 경우 몇몇 금형코드 데이터 셋에서 잔차 독립성 검정을 통과하지 못해, auto-arima 모듈의 최적 차수 도출 기준에 Ljung-Box Test 검정 통과 조건을 추가하여 다시 도출함으로써 (p,d,q) 차수를 조정하였다.

5. 최종 ARIMA(p,d,q) 계수

1. 비정상(1차 차분)

-molten_volume: ARIMA(0,1,2)

-molten_temp: ARIMA(1,1,3)

-upper_mold_temp1: ARIMA(1,1,1)

-upper_mold_temp2: ARIMA(1,1,5)

-lower_mold_temp1: ARIMA(1,1,5)

-lower_mold_temp2: ARIMA(1,1,2)

-Coolant_temperature: ARIMA(1,1,2)

2. 정상(차분 X)

-cast_pressure: ARIMA(1,1,1)

-high_section_speed: ARIMA(1,0,1)

-biscuit_thickness: ARIMA(1,0,1)

-sleeve_temperature: ARIMA(1,0,1)

3. 자기 상관 x (X-R 차트)

-physical_strength

-low_section_speed

-> ARIMA (p,d,q) 계수를 기준으로 5개 금형코드, 11개 변수에 대해 총 55개 모델 적합