

2021 금융데이터 활용 경진대회

Human plus AI 단기 종가 예측

Sonewn팀



01 팀 소개

» 각기 다른 분야의 3인이 융합된 프로젝트 팀입니다.

소눈

산업공학
prophet

Cafelatte

컴퓨터공학+경영학
Linear

최선웅

수학과 석사
Xgboost



✓ 배경

1. 알고리즘 기반 예측의 이익 유도 가능성

-인공 지능 측면에서 개발된 예측 모델 **실제 매매에 있어서 성능의 한계**

-직접 모델 개발을 통해 이익 유도 가능성 확인
및 투자전략 연구

2. 미래 주가에 영향을 주는 데이터 탐색

-하나의 모델을 기준으로 어떤 데이터가 미래 주가에
가장 영향력 가지는지 파악

✓ 필요성

한정된 기업 또는 종목내에서의 예측x



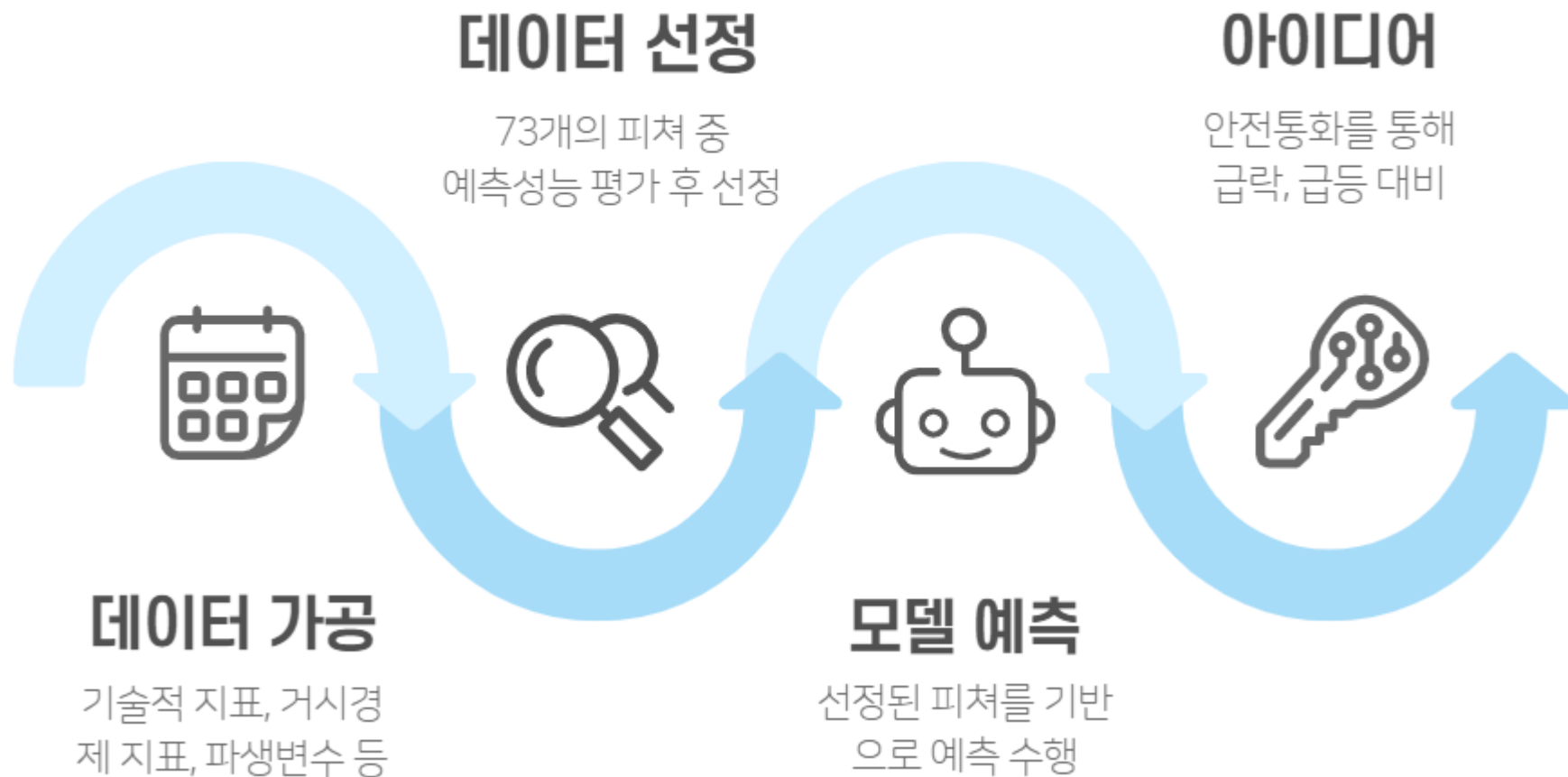
여러 종목에 **범용** 가능한 모델 필요

+

종가에 대한 실질적 **영향 변수**

03 종료 가격 예측 전략

» 종료가격예측 예측 전략을 작성해 주세요





1. Linear Regression

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \varepsilon_t,$$

public score 기준 가장 뛰어난 성능을 가져 선정

Linear public = 2.8xxx
XGboost public = 4.1xxx

+

2. Close_Smoothing

smoothing 공식: $A * (1 + B * \text{time_ngap})$

A: 주가 등락 상하한 (=0.1)
B: target 일자 거리에 대한 multiplier (=0.2)
time_ngap: target과 train의 거리 (1~5)

내부검증 결과 약 -2.5%의 성능 향상

+

3. 안전통화 기반 후처리

안전자산 선호심리 관련 시장트렌드 반영

-> usdjpy 값으로 prediction 조정

-가벼운 변동에 대해서는 계산되지 않도록 반올림 연산 적용

```
# adjustment
if abs(val_adj_usdjpy) != 0:
    tmp_pred = model.predict(val_x[:, np.newaxis]).flatten() * (1 + val_adj_usdjpy)
else:
    tmp_pred = model.predict(val_x[:, np.newaxis]).flatten()
```

05 종료 가격 예측 데이터 분석

2019년 ~ 2021년 11월 26일까지의 아래와 같은 데이터를 수집

해외 index	보조지표	이동평균	파생변수
<ul style="list-style-type: none"> • USDKRW=X • USDAUD=X • USDAUD=X • USDJPY=X) • EURUSD=X • S&P 500 • DJI • IXIC(나스닥지수) • 유로 스톡스50 • 필라델피아 반도체지수 • 상해종합주가지수 • CSI 300 Index • MSCI Emerging Markets 	<ul style="list-style-type: none"> • MFI • MFI_signal • OBV • Stochastic 파생변수 • 기관 연속 순매수 • 외국인 연속 순매수 • • 	<ul style="list-style-type: none"> • Close 5, 10, 20 • Volume 5, 10, 20 • Trading_amount 5,10, 20 • Kospo 5, 10, 20 • 기관 5, 10, 20 	<ul style="list-style-type: none"> • 가격 파생변수 <ul style="list-style-type: none"> (1) 1~5일 전 종가 (2) close_return (3) std_20 (4) close_mean2~3 (5) close_diff 등 • 날짜 파생변수 <ul style="list-style-type: none"> (1) 월별 주기성 sin, cos (2) 요일 변수 (3) 월 변수 • 지표 파생변수 <ul style="list-style-type: none"> (1) MFI 파생변수 (2) OBV 파생변수

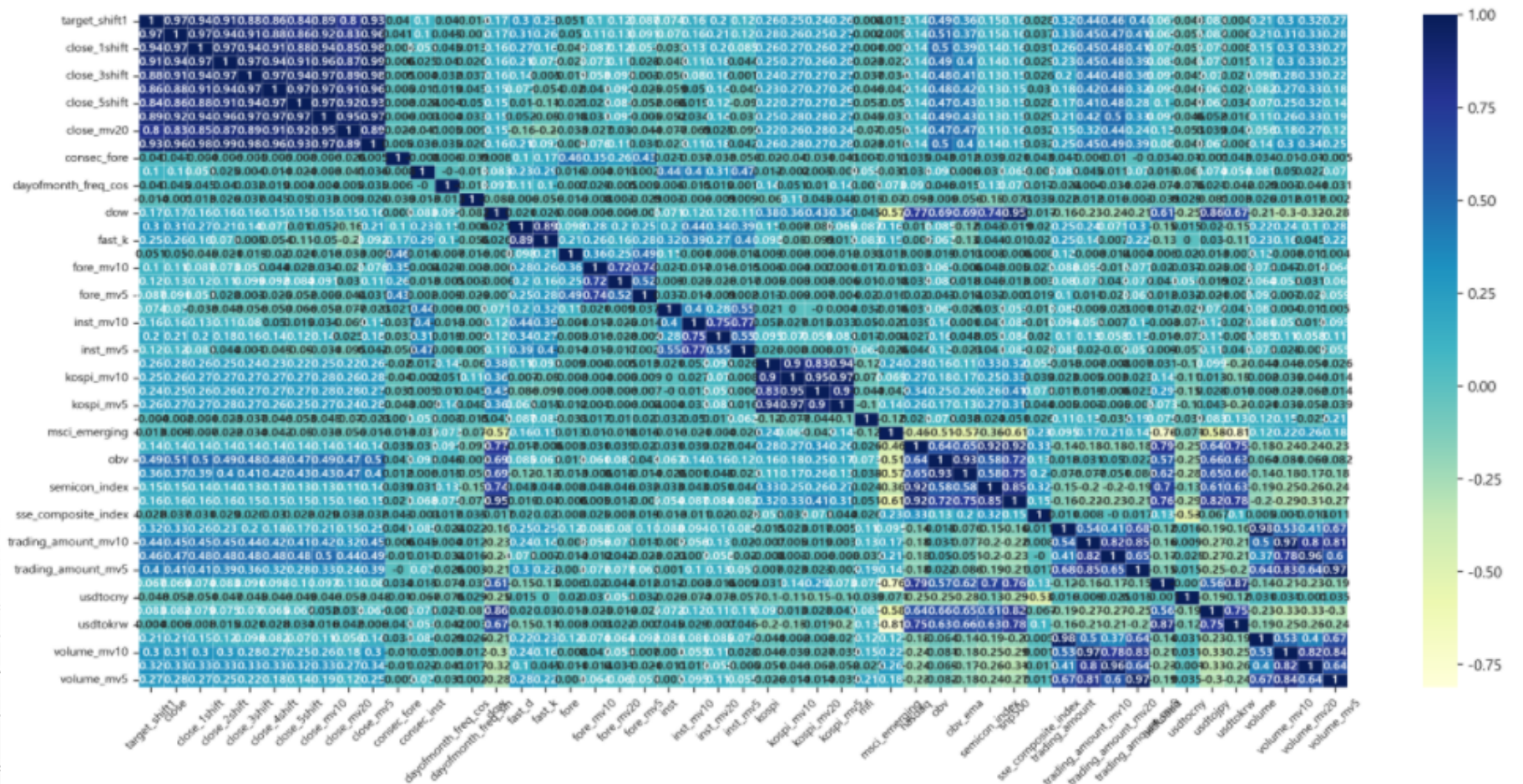
총 73개의 피쳐 생성 후 49가지 조합으로 테스트 진행

06 종료 가격 예측 데이터 처리 결과

종료 가격과의 각 데이터들의 상관관계의 평균, 표준편차, 분산을 파악

Mean Correlation Visualization

(D+1일 기준)



06 종료 가격 예측 데이터 처리 결과

상관관계 요약

1. 오늘의 주가가 가장 큰 영향을 준다.

(value 자체가 가장 근접하기 때문이라고 추정)

2. 기존 인식과는 다르게 외국인과 기관이 종가와 수치적으로는 큰 상관성이 없다.

3. 분석한 기술적 지표 중 가장 큰 상관성을 가지는 지표는 OBV이다.

close	0.621622
close_1shift	0.135135
close_2shift	0.127027
close_3shift	0.110811
close_4shift	0.083784
close_5shift	0.067568
close_mv10	0.143243
close_mv20	0.197297
close_mv5	0.310811
consec_fore	0.091892
consec_inst	0.116216
dayofmonth_freq_cos	0.143243
dayofmonth_freq_sin	0.154054
dow	0.221622
fast_d	0.159459
fast_k	0.083784
fore	0.113514
fore_mv10	0.140541
fore_mv20	0.191892
fore_mv5	0.121622
inst	0.110811
inst_mv10	0.186486
inst_mv20	0.167568
inst_mv5	0.162162
kospi	0.102703
kospi_mv10	0.191892
kospi_mv20	0.22973
kospi_mv5	0.127027
mfi	0.127027
msci_emerging	0.156757
nasdaq	0.227027
obv	0.175676
obv_ema	0.232432

추가자료 : 선형회귀 결과에 대해 전체 기업
중 유의성을 보이는 변수의 비율
(alpha = 0.05 일 때, p-value 판별)

06 종료 가격 예측 데이터 처리 결과



```
**** Linear ****
```

	time_gap_1	time_gap_2	time_gap_3	time_gap_4	time_gap_5	avg	std	running_time
feature_seed_1	2.796144	4.706079	6.901019	8.488088	10.609497	6.700165	3.071549	0.011222
feature_seed_2	2.026291	3.392397	4.715429	6.370653	8.476691	4.996292	2.523604	0.005134
feature_seed_3	1.695163	2.64663	3.1698	4.089943	5.448304	3.409968	1.431616	0.005006
feature_seed_4	1.56596	2.238617	2.662146	3.546419	4.696821	2.941993	1.215655	0.004437
feature_seed_5	1.593868	2.326421	2.739494	3.532726	4.714973	2.981496	1.196035	0.004315
feature_seed_6	1.60205	2.369558	2.752485	3.582873	4.748709	3.011135	1.205206	0.004421
feature_seed_7	1.605063	2.168434	2.736571	3.617371	4.684247	2.962337	1.216294	0.004913
feature_seed_8	1.636137	2.440626	2.936945	3.667301	4.608374	3.057877	1.139343	0.004913
feature_seed_9	1.579857	2.179939	2.736244	3.581358	4.64124	2.943728	1.201529	0.004755
feature_seed_10	1.600961	2.250382	2.75946	3.606332	4.683282	2.980067	1.201657	0.004883
feature_seed_11	1.663527	2.312746	2.85867	3.785244	5.01969	3.127975	1.312662	0.00488
feature_seed_12	1.612603	2.156792	2.703604	3.548018	4.605413	2.925286	1.180412	0.004898
feature_seed_13	1.625723	2.302694	2.74289	3.54635	4.600041	2.96354	1.150322	0.004748
feature_seed_14	1.614418	2.179179	2.705167	3.569271	4.613439	2.936295	1.181892	0.004932
feature_seed_15	1.615	2.224333	2.656951	3.484694	4.476024	2.8914	1.117158	0.004963
feature_seed_16	1.633615	2.490536	2.975008	3.962054	4.991939	3.21063	1.303718	0.004822
feature_seed_17	1.621284	2.296611	2.880341	3.852536	4.8271	3.095574	1.26781	0.004988
feature_seed_18	1.617846	2.253198	2.658485	3.462642	4.409959	2.880426	1.085729	0.004972
feature_seed_19	1.637454	2.183739	2.671446	3.470519	4.435847	2.879801	1.100178	0.004821
feature_seed_20	1.672647	2.20649	2.861931	3.852481	5.112562	3.141223	1.369353	0.005132
feature_seed_21	1.654205	2.250048	2.755276	3.534167	4.587067	2.956153	1.14324	0.004843
feature_seed_22	1.677899	2.244194	2.956871	4.037641	5.272054	3.237732	1.438524	0.004978
feature_seed_23	1.635899	2.131128	2.689558	3.587904	4.638283	2.936554	1.195989	0.004886
feature_seed_24	1.62447	2.153669	2.670228	3.625528	4.64899	2.944577	1.205426	0.004883
feature_seed_25	1.598376	2.254541	2.807868	3.721165	4.77828	3.032046	1.248439	0.004918
feature_seed_26	1.62718	2.241912	2.683266	3.42793	4.400008	2.876059	1.075406	0.004817
feature_seed_27	1.661745	2.208219	2.760886	3.577986	4.692419	2.980251	1.190737	0.004838
feature_seed_28	1.646334	2.201205	2.703606	3.469609	4.452044	2.89456	1.099058	0.005069
feature_seed_29	1.64203	2.331433	2.774425	3.448744	4.462183	2.931763	1.079122	0.005034
feature_seed_30	1.693713	2.247314	3.185879	4.262045	5.452107	3.368212	1.51947	0.005509

<best feature set>
feature_seed_26

✓ 검증에서 가장 성능이 좋은 데이터 조합

- 날짜
- 전일 증가
- 5일전 증가
- 필라델피아 반도체 지수
- 상해종합 주가 지수
- USD/KRW, EUR/USD
- 요일 데이터



✓ Public Score가 가장 좋은 데이터 조합

- 날짜
- 전일 증가
- 요일 데이터

06 종료 가격 예측 데이터 처리 결과



✓

Public Score가 가장 좋은 데이터

+

Model

✓

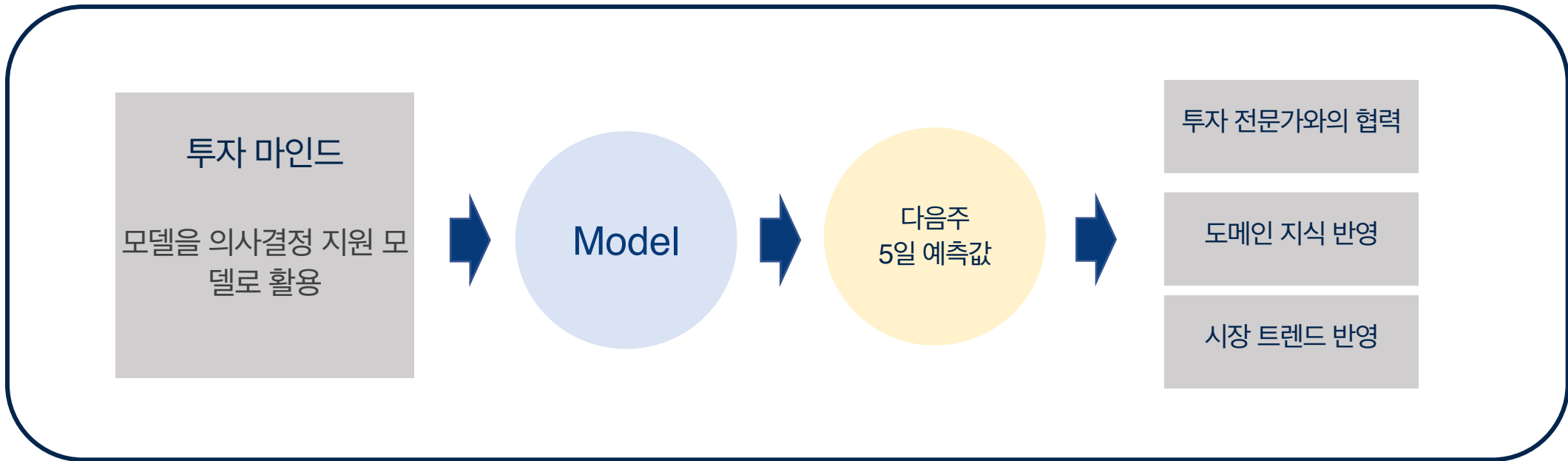
val_adj_usdjpy 반영



종가 예측 정확도

Dacon Private score
3.71564

Human plus AI 단기투자 전략



모델을 통해 의사결정 과정을 데이터화 해 심리편향 없는 투자를 지원



국토교통부



전라북도



전북창조경제혁신센터

마무리 멘트

Thank you

