

코스피 예측 모델

2020/10/07 이문형 정기미팅 발표자료

1. 데이터 수집 및 전처리 (코스피, 나스닥, 금값, 원유시세, 거시경제지표 및 글로벌 주가지수 등) ta_practice3.ipynb

-> 대부분의 거시경제지표의 경우, 데이터 시간 단위가 연/분기/월임.

선행 논문에서도 거시경제지표를 활용해 코스피 지수를 예측할 경우 월 단위로 진행함.

본 프로젝트의 도메인은 일자별 예측이라 판단해서 거시경제지표는 반영하지 않음.

-> 일자별 데이터만 사용함.

1) 코스피 차트 데이터 - 출처 : 한국거래소

지표명	표현
년/월/일	date
종가	kospi_close
배당수익률	kospi_dividend_yield_ratio
주가수익비율	kospi_per
주가순자산비율	kospi_pbr
시가	kospi_open
고가	kospi_high
저가	kospi_low
거래량	kospi_volume
거래대금	kospi_value
시가총액	kospi_stock_value

2) 환율 데이터 - 출처 : 한국거래소, 한국은행

지표명	표현
원/달러 종가	exchange_rate_usd_close
원/유로 종가	exchange_rate_eur_close
원/엔 종가	exchange_rate_jpy_close
원/43개국 화폐	exchange_rate_won_나라이름

3) 금값시세 - 출처 : 한국거래소

지표명	표현
금 현물	krx_gold_close
종가/거래량/거래대금	krx_gold_volume krx_gold_value
국제 금 시세	world_gold_won_close
원/oz	world_gold_oz_close

4) 유가 데이터(원자재) - 출처 : 한국거래소, 야후 파이낸스

지표명	표현
국내 휘발유 가중평균(경쟁)/ 거래량/거래대금/ 가중평균(협약)	krx_gasoline_wa_compt krx_gasoline_volume krx_gasoline_value krx_gasoline_wa_agree
국내 등유 가중평균(경쟁)/ 거래량/거래대금/ 가중평균(협약)	krx_diesel_wa_compt krx_diesel_volume krx_diesel_value krx_diesel_wa_agree
국내 경유 가중평균(경쟁)/ 거래량/거래대금/ 가중평균(협약)	krx_kerosene_wa_compt krx_kerosene_volume krx_kerosene_value krx_kerosene_wa_agree
금 종가/거래량	comex_gold_close comex_gold_volume
은 종가/거래량	comex_silver_close comex_silver_volume
동(구리) 종가/거래량	comex_copper_close comex_copper_volume
플래티넘 종가/거래량	platinum_close platinum_volume
팔라듐 종가/거래량	palladium_close palladium_volume
휘발유 종가/거래량	rbob_gasoline_close rbob_gasoline_volume
천연가스 종가/거래량	natural_gas_close natural_gas_volume
난방유 종가/거래량	heating_oil_close heating_oil_volume
원유 종가/거래량	crude_oil_close crude_oil_volume

5) 금리 데이터 - 출처 : 한국은행, 야후 파이낸스

지표명	표현
기준금리	base_rate
콜금리	call_rate
KORIBOR (3개월)	koribor_3m
CD수익률 (91일)	cd_91d
국고채수익률 (1년)	korea_treasury_bond_1y
통안증권수익률 (91일)	tongan_91d
회사채수익률 (3년)	corporate_bond_3y_aa_
미국 국채수익률 (13주)	treasury_13w_close
미국 국채수익률 (5년)	treasury_5y_close
미국 국채수익률 (10년)	treasury_10y_close
미국 국채수익률 (30년)	treasury_30y_close

6) 배출권

지표명	표현
koc 증가	koc_close

7) 글로벌 주가지수 - 출처 : 야후 파이낸스, investing.com

지표명	표현
vix 증가	vix_close
vkospi 증가	vkospi_close
bitcoin_usd 증가/거래량	bitc_close bitc_volume
s&p 500 증가/거래량	snp500_close snp500_volume
daw jones 증가/거래량	daw_jones_close daw_jones_volume
nsadaq 증가/거래량	nasdaq_close nasdaq_volume
nyse 증가/거래량	nyse_close nyse_volume
nyse amex 증가	amex_close
russell 2000 증가/거래량	russell_close russell_volume
dax 증가/거래량	dax_close dax_volume
nikkei 225 증가/거래량	nikkei_close nikkei_volume
hang seng 증가/거래량	hangseng_close hangseng_volume
sse 증가/거래량	sse_close sse_volume
estx 50 증가/거래량	estx_close estx_volume
euronext 100 증가/거래량	euronext_close euronext_volume

2. 코스피 예측 모델 (feature selection model)

- 피쳐 조합을 변경해가며 autoML을 통해 실험을 진행함.
- 2018.01.01.~2020.09.18 (train 60, val 20, test 20) -> best 3 model -> 앙상블 -> 전체 트레인 셋으로 재훈련 -> 예측

feature combination
(1) 코스피 차트 데이터
(2) 코스피 차트 데이터 + 보조지표
(3) 코스피 차트 데이터 + 글로벌 주가지수
(4) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리
(5) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리 + 글로벌 주가지수
(6) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리 + 글로벌 주가지수 + 보조지표

1) 분류 모델 autoML_practice2.ipynb

분류 모델에서는 모든 피쳐에 보조지표를 반영했을 때의 정확도가 높았음.

글로벌 주가지수 보다는 환율 + 금값시세 + 유가 + 금리를 반영했을 때 예측 정확도가 높았음.

feature combination	score																							
(1) autoML_practice2_char	<table><tr><th>Model</th><th>Accuracy</th><th>AUC</th><th>Recall</th><th>Prec.</th><th>F1</th><th>Kappa</th><th>MCC</th></tr><tr><td>0 Voting Classifier</td><td>0.677</td><td>0</td><td>0.5333</td><td>0.7018</td><td>0.6061</td><td>0.3409</td><td>0.3501</td></tr></table>								Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	0 Voting Classifier	0.677	0	0.5333	0.7018	0.6061	0.3409	0.3501
Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC																	
0 Voting Classifier	0.677	0	0.5333	0.7018	0.6061	0.3409	0.3501																	
(2) autoML_practice2_tech	<table><tr><th>Model</th><th>Accuracy</th><th>AUC</th><th>Recall</th><th>Prec.</th><th>F1</th><th>Kappa</th><th>MCC</th></tr><tr><td>0 Voting Classifier</td><td>0.8872</td><td>0</td><td>0.8226</td><td>0.9273</td><td>0.8718</td><td>0.7718</td><td>0.7762</td></tr></table>								Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	0 Voting Classifier	0.8872	0	0.8226	0.9273	0.8718	0.7718	0.7762
Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC																	
0 Voting Classifier	0.8872	0	0.8226	0.9273	0.8718	0.7718	0.7762																	
(3) autoML_practice2_4	<table><tr><th>Model</th><th>Accuracy</th><th>AUC</th><th>Recall</th><th>Prec.</th><th>F1</th><th>Kappa</th><th>MCC</th></tr><tr><td>0 Voting Classifier</td><td>0.697</td><td>0</td><td>0.5246</td><td>0.7442</td><td>0.6154</td><td>0.3775</td><td>0.3932</td></tr></table>								Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	0 Voting Classifier	0.697	0	0.5246	0.7442	0.6154	0.3775	0.3932
Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC																	
0 Voting Classifier	0.697	0	0.5246	0.7442	0.6154	0.3775	0.3932																	
(4) autoML_practice2_3	<table><tr><th>Model</th><th>Accuracy</th><th>AUC</th><th>Recall</th><th>Prec.</th><th>F1</th><th>Kappa</th><th>MCC</th></tr><tr><td>0 Voting Classifier</td><td>0.7403</td><td>0</td><td>0.75</td><td>0.7105</td><td>0.7297</td><td>0.4801</td><td>0.4807</td></tr></table>								Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	0 Voting Classifier	0.7403	0	0.75	0.7105	0.7297	0.4801	0.4807
Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC																	
0 Voting Classifier	0.7403	0	0.75	0.7105	0.7297	0.4801	0.4807																	
(5) autoML_practice2_2	<table><tr><th>Model</th><th>Accuracy</th><th>AUC</th><th>Recall</th><th>Prec.</th><th>F1</th><th>Kappa</th><th>MCC</th></tr><tr><td>0 Voting Classifier</td><td>0.626</td><td>0</td><td>0.5333</td><td>0.6038</td><td>0.5664</td><td>0.2397</td><td>0.2412</td></tr></table>								Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	0 Voting Classifier	0.626	0	0.5333	0.6038	0.5664	0.2397	0.2412
Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC																	
0 Voting Classifier	0.626	0	0.5333	0.6038	0.5664	0.2397	0.2412																	
(6) autoML_practice2	<table><tr><th>Model</th><th>Accuracy</th><th>AUC</th><th>Recall</th><th>Prec.</th><th>F1</th><th>Kappa</th><th>MCC</th></tr><tr><td>0 Voting Classifier</td><td>0.8879</td><td>0</td><td>0.875</td><td>0.875</td><td>0.875</td><td>0.7733</td><td>0.7733</td></tr></table>								Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	0 Voting Classifier	0.8879	0	0.875	0.875	0.875	0.7733	0.7733
Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC																	
0 Voting Classifier	0.8879	0	0.875	0.875	0.875	0.7733	0.7733																	

feature combination
(1) 코스피 차트 데이터
(2) 코스피 차트 데이터 + 보조지표
(3) 코스피 차트 데이터 + 글로벌 주가지수
(4) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리
(5) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리 + 글로벌 주가지수
(6) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리 + 글로벌 주가지수 + 보조지표

2) 회귀 모델 autoML_practice.ipynb

회귀 모델에서는 보조지표만 반영했을 때 예측 오차가 낮았음.

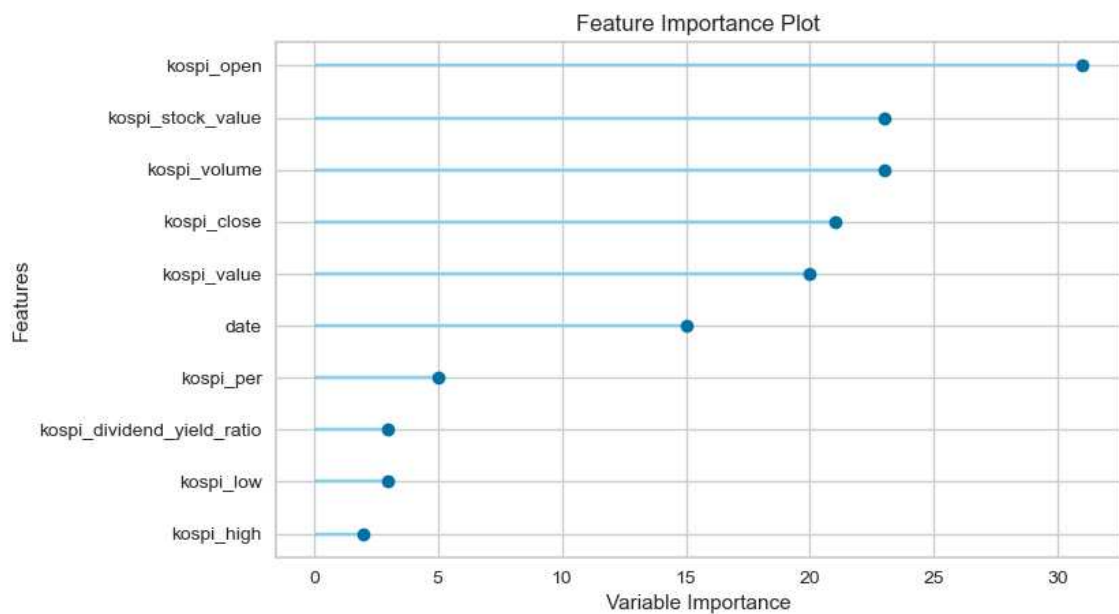
feature combination	score						
(1) autoML_practice_chart	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
	0 Voting Regressor	0.0301	0.0026	0.0513	0.9976	0.0238	0.0581
(2) autoML_practice_tech	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
	0 Voting Regressor	0.0108	0.0003	0.0161	0.9997	0.0088	0.0081
(3) autoML_practice_4	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
	0 Voting Regressor	0.0338	0.0242	0.1556	0.9701	0.0658	-0.0365
(4) autoML_practice_3	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
	0 Voting Regressor	0.0242	0.0027	0.0523	0.9964	0.0377	-0.0725
(5) autoML_practice_2	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
	0 Voting Regressor	0.022	0.0008	0.0286	0.9989	0.0169	0.0113
(6) autoML_practice	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
	0 Voting Regressor	0.0244	0.0013	0.0364	0.9985	0.0199	-0.0347

3) feature importance (light gbm)

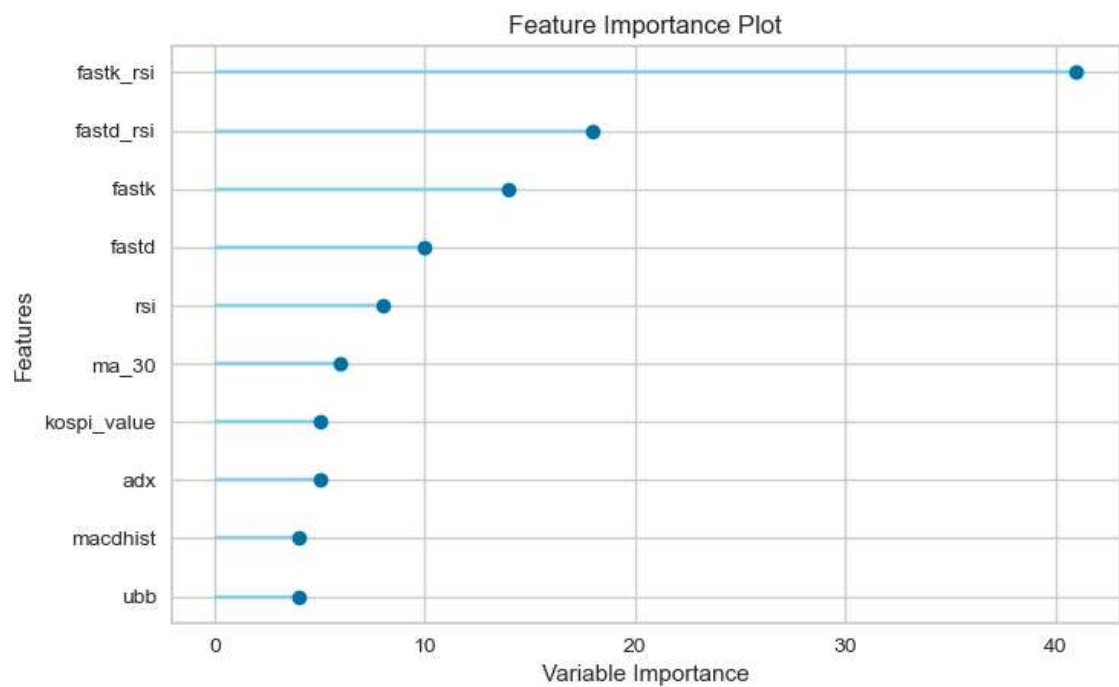
-> (참고) 모든 경우에서 트리 모델이 best가 아님. 분류에서는 트리 모델 성능이 좋음.
모델 학습 시, feature_selection, remove_multicollinearity, ignore_low_variance 파라미터 조합에 따라 결과가 다른 것을 확인(현재는 feature_selection = True, ignore_low_variance = True) -> 추후 이에 대한 실험 + 강화학습에 적용해보기

1) 분류 모델

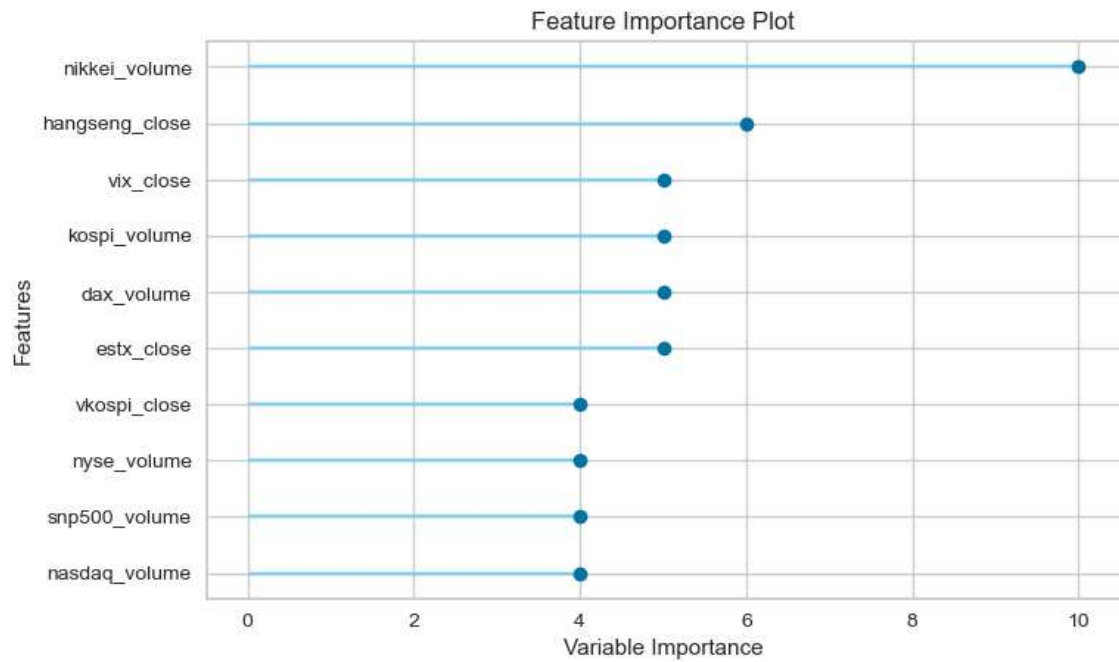
(1) 코스피 차트 데이터



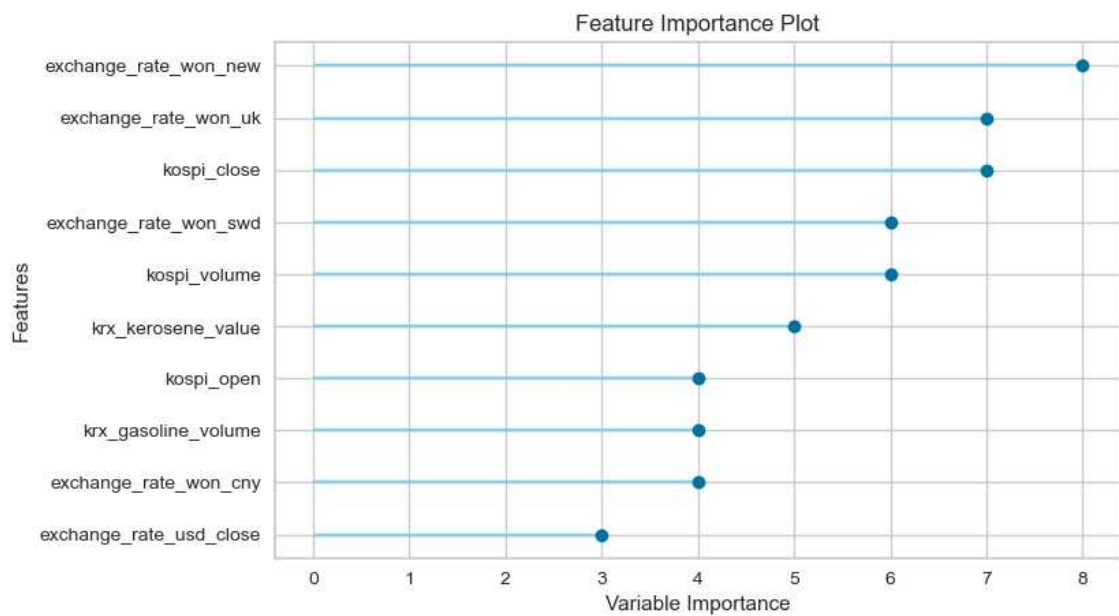
(2) 코스피 차트 데이터 + 보조지표



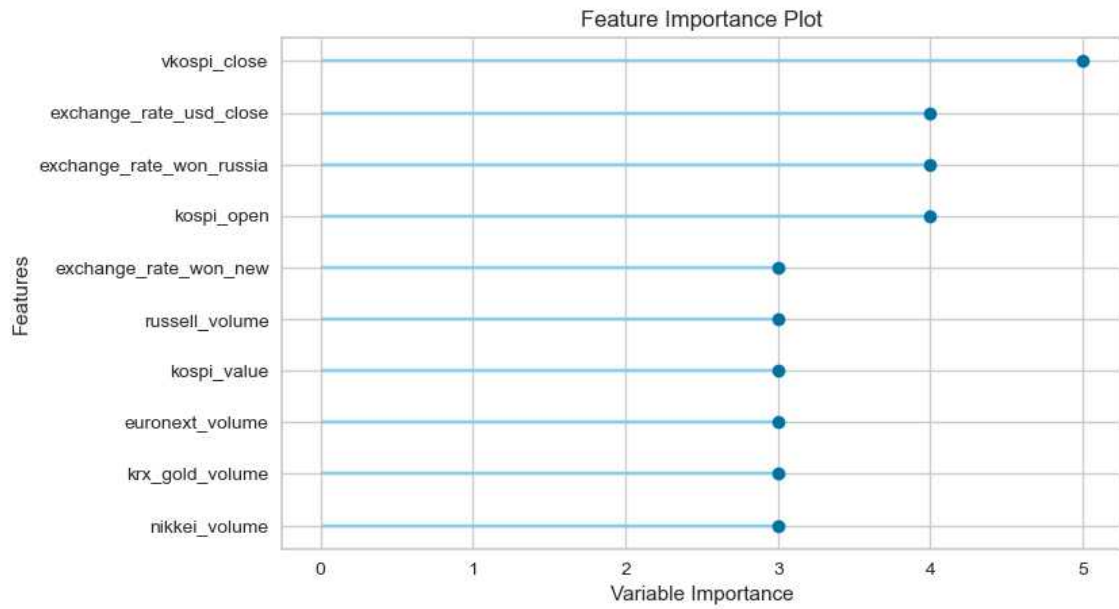
(3) 코스피 차트 데이터 + 글로벌 주가지수



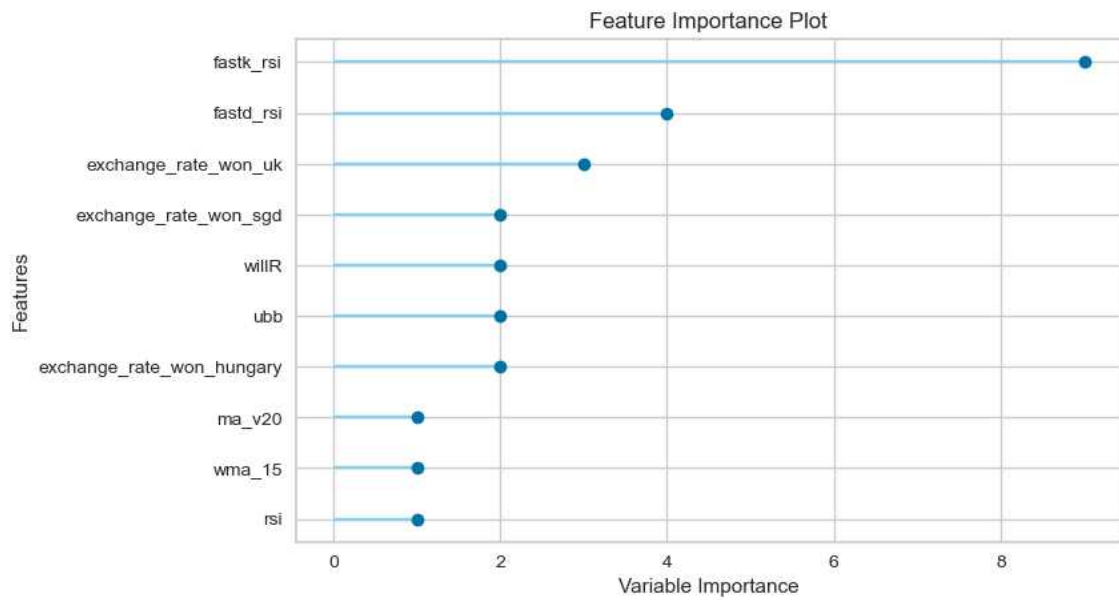
(4) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리



(5) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리 + 글로벌 주가지수

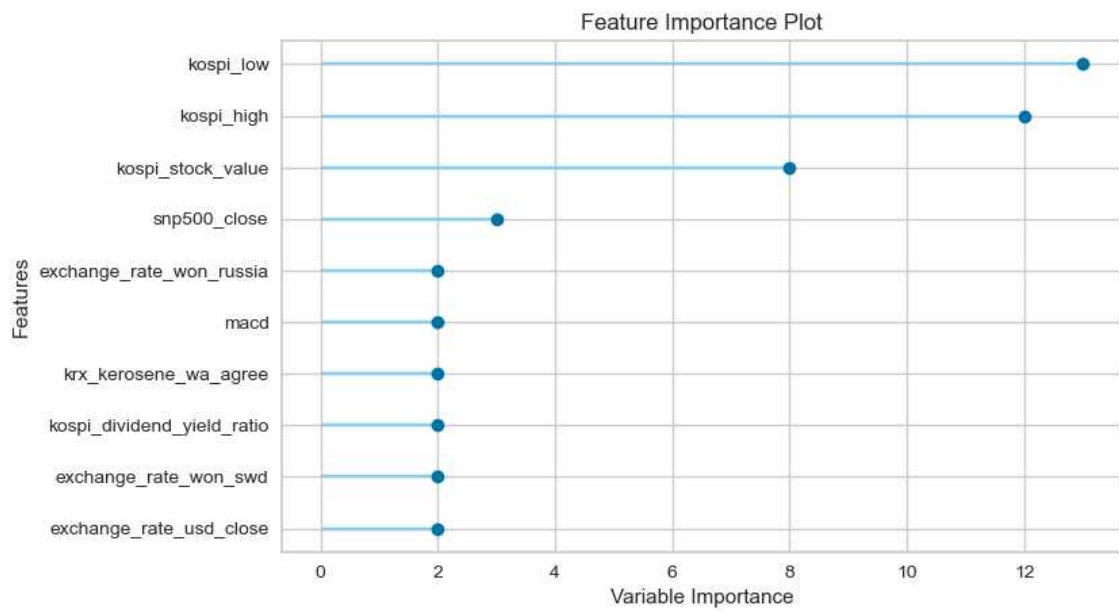


(6) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리 + 글로벌 주가지수 + 보조지표

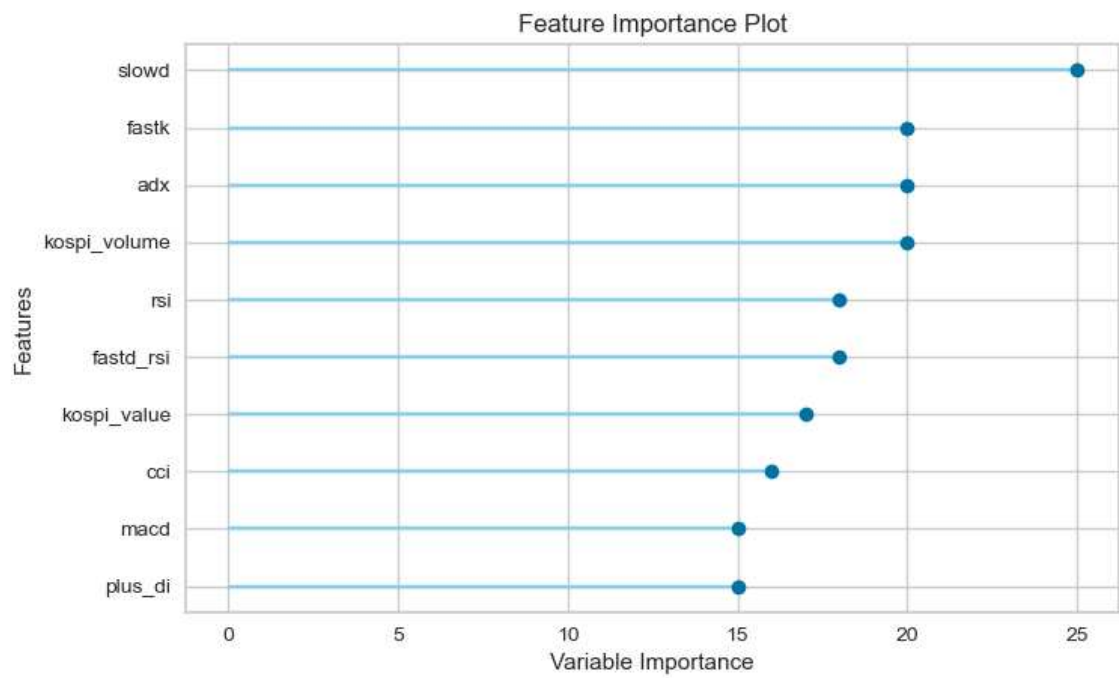


2) 회귀 모델

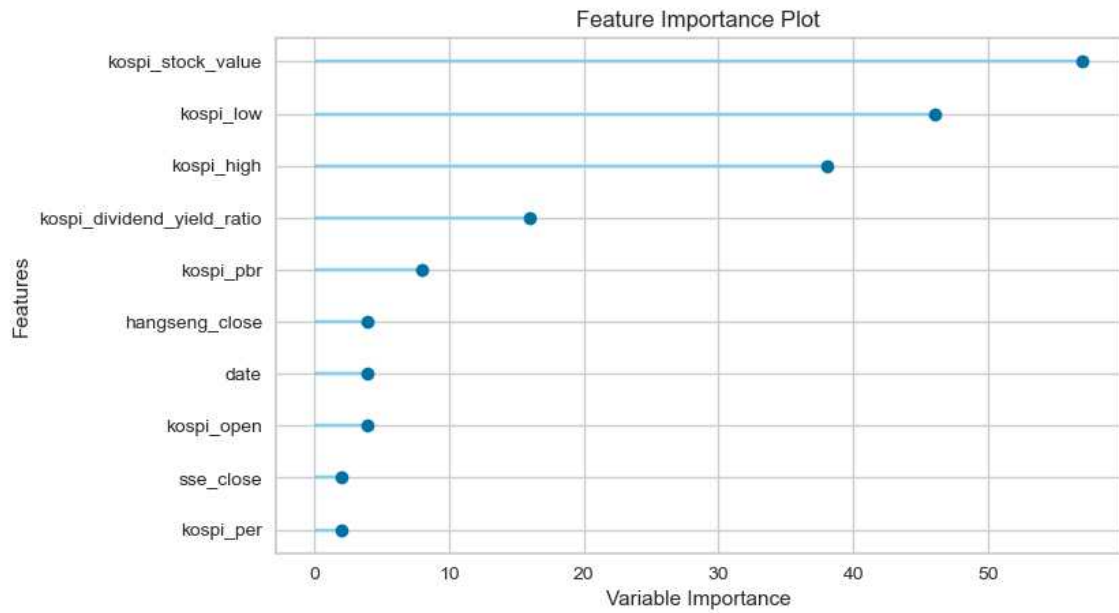
(1)



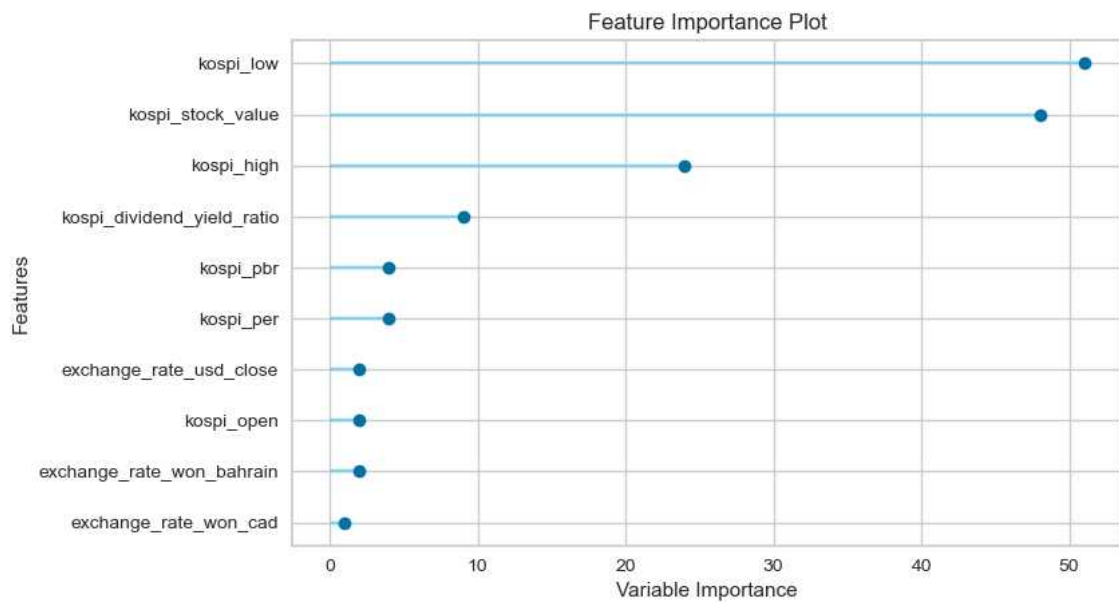
(2) 코스피 차트 데이터 + 보조지표



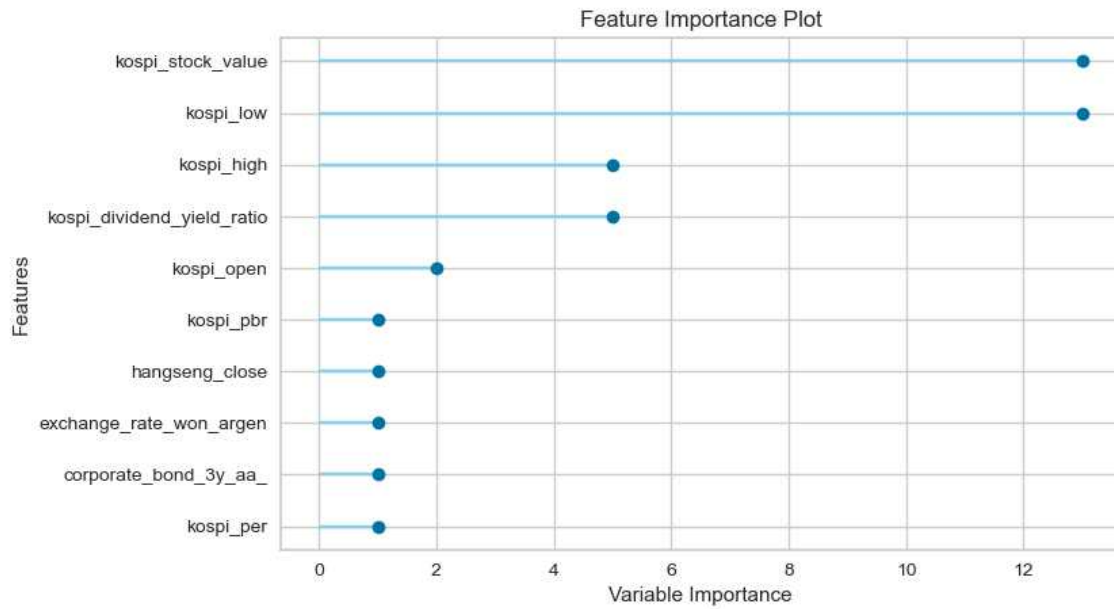
(3) 코스피 차트 데이터 + 글로벌 주가지수



(4) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리



(5) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리 + 글로벌 주가지수



(6) 코스피 차트 데이터 + 환율 + 금값시세 + 유가 + 금리 + 글로벌 주가지수 + 보조지표

