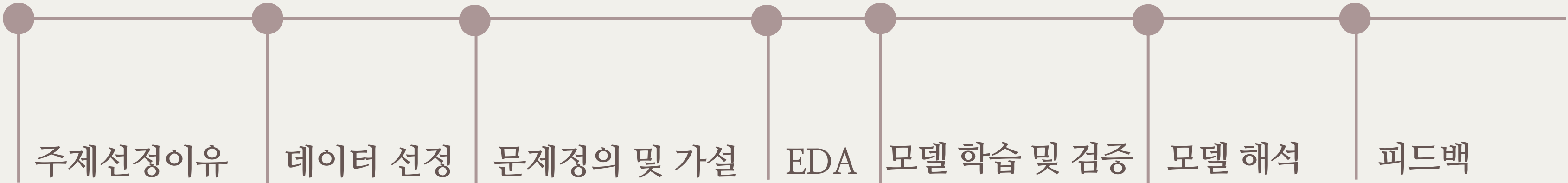
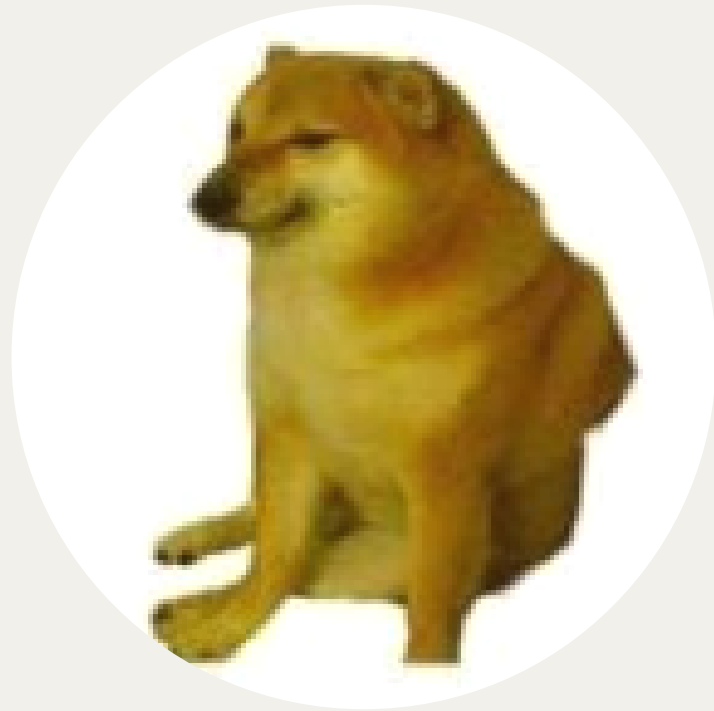


# 기술적 요인과 거시적요인에 따른 코스피 지수의 상승 여부 분류모델 구현

# CONTENTS



# 주제 선정이유



지수의 움직임 예측(시작모델)

상승하면 지수와 상관관계가 높은  
종목, 하락이 예측되면 상관관계가  
낮은 종목



최종 모델

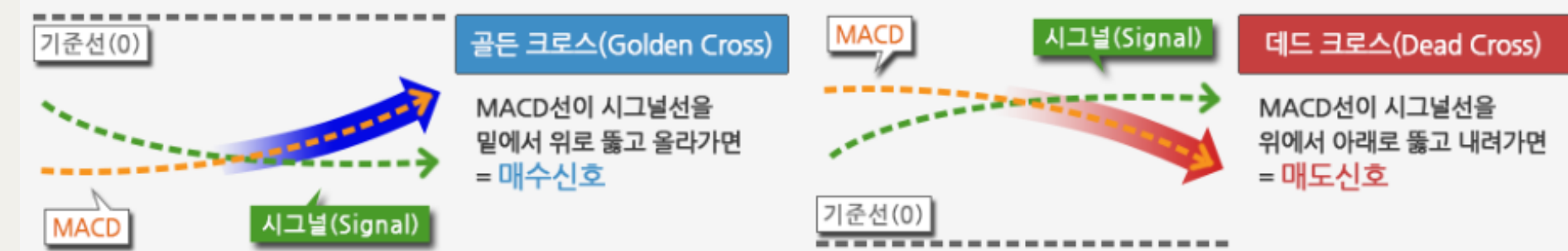
개인 투자자들이 개별주식 종목이나 지수  
를 따르는 ETF상품들의 일정기간 동안  
목표 수익률을 안정적으로 달성할 수 있  
도록 매매타이밍을 추천하는 모델



# 데이터셋 선정

\*이동평균선: 이동평균선은 n일간의 (수정)주가를 평균낸 값

\*MACD: 단기 이동평균 - 장기 이동평균



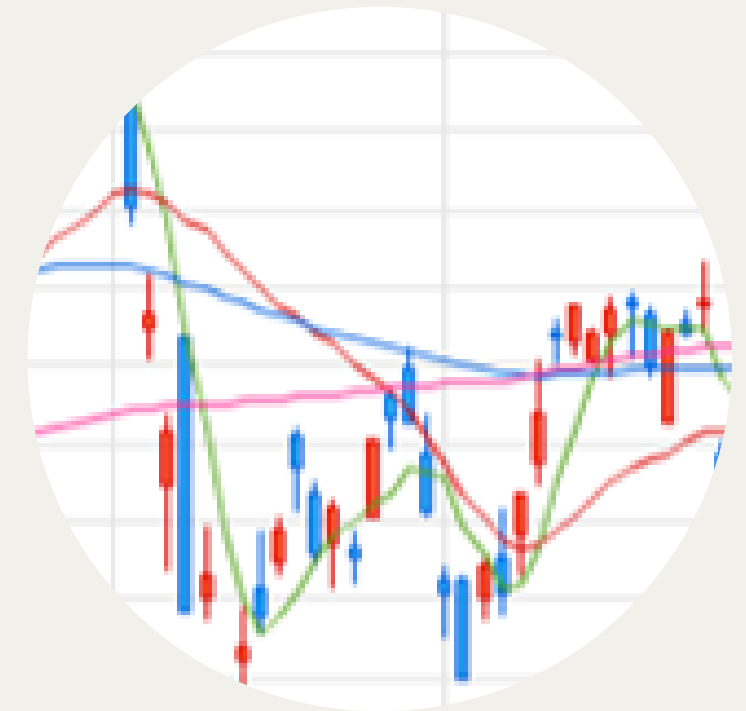
코스피 지수 증가



거시경제 지표

나스닥 지수의 증가, 미국과 한국의 국채 10년 금리, 미국과 한국의 전월 대비, 전년 대비 CPI 증감율

- 출처: yfinance



기술적 지표

코스피 지수의 이평선들과 단기 이평선과 장기 \*이평선의 차이인 \*MACD

- 출처: Federal Reserve Economic Data, 연방준비은행 경제 데이터베이스

# 문제정의 및 가설

---

- 1.나스닥 지수와 코스피 지수는 비례 관계를 가질 것이다.
  - 2.금리와 코스피 지수는 반비례 관계
  - 3.CPI 증감율과 코스피 지수는 반비례 관계
  - 4.MACD가 작아지면 지수가 상승하거나 하락.
    - 단기 이평선이 장기 이평선을 뚫고 지나가는 골든 크로스 지점이거나 역으로 하락하면서 갭이 좁아지는 경우.
- 코스피의 상승여부는 현재종가에서 이전종가를 빼서 양수면 1, 음수, 즉, 하락하면 0.

# EDA, 데이터 처리

- 데이터 누락이나 시장 영업일 차이로 생긴 결측치는 앞 데이터로 채우기.
- 또한, 라벨 데이터 이외에는 모두 수치형이기 때문에 특성 간의 단위를 균등하게 보정해주는 표준화 진행.
- 독립변수들 간의 상관관계가 너무 높은 이평선들의 일부는 **\*과적합**을 방지하기 위해 macd를 계산하는데에 쓰고 나중에삭제

	Close	Adj Close	Volume	m5	m12	m20	m26	m60	m120	t10_rate	k10_rate	nsd_close
Close	1.000000	1.000000	-0.054316	0.997920	0.994130	0.989770	0.986540	0.970349	0.926464	-0.587137	-0.751693	0.767264
Adj Close	1.000000	1.000000	-0.054316	0.997920	0.994130	0.989770	0.986540	0.970349	0.926464	-0.587137	-0.751693	0.767264
Volume	-0.054316	-0.054316	1.000000	-0.056503	-0.062736	-0.069325	-0.074309	-0.094434	-0.092037	-0.164945	-0.082408	0.014057
m5	0.997920	0.997920	-0.056503	1.000000	0.997695	0.993780	0.990671	0.974433	0.930668	-0.588673	-0.752229	0.768179
m12	0.994130	0.994130	-0.062736	0.997695	1.000000	0.998244	0.995886	0.980566	0.937399	-0.590438	-0.752751	0.769463
m20	0.989770	0.989770	-0.069325	0.993780	0.998244	1.000000	0.999245	0.986532	0.944509	-0.592506	-0.753630	0.771003
m26	0.986540	0.986540	-0.074309	0.990671	0.995886	0.999245	1.000000	0.990287	0.949412	-0.594276	-0.754635	0.772287
m60	0.970349	0.970349	-0.094434	0.974433	0.980566	0.986532	0.990287	1.000000	0.969343	-0.609610	-0.763697	0.780138
m120	0.926464	0.926464	-0.092037	0.930668	0.937399	0.944509	0.949412	0.969343	1.000000	-0.624573	-0.761869	0.780172
t10_rate	-0.587137	-0.587137	-0.164945	-0.588673	-0.590438	-0.592506	-0.594276	-0.609610	-0.624573	1.000000	0.838747	-0.482218
k10_rate	-0.751693	-0.751693	-0.082408	-0.752229	-0.752751	-0.753630	-0.754635	-0.763697	-0.761869	0.838747	1.000000	-0.831258
nsd_close	0.767264	0.767264	0.014057	0.768179	0.769463	0.771003	0.772287	0.780138	0.780172	-0.482218	-0.831258	1.000000
전월의 미국 CPI%change	0.212202	0.212202	-0.104539	0.213346	0.215736	0.216933	0.217498	0.211183	0.148570	0.223293	0.058179	0.104368
전월의 전년대비 미국 CPI%change	-0.210327	-0.210327	-0.178481	-0.206278	-0.198402	-0.189222	-0.181806	-0.142155	-0.101725	0.416327	0.379606	-0.173199
전월의 한국 CPI%change	-0.021713	-0.021713	-0.066832	-0.020470	-0.019055	-0.018733	-0.018968	-0.022812	-0.030923	0.207762	0.224498	-0.183388
전월의 전년대비 한국 CPI%change	-0.456821	-0.456821	0.009538	-0.455010	-0.451144	-0.446268	-0.442165	-0.419841	-0.377253	0.331166	0.637724	-0.640787
exchange_rate	0.019724	0.019724	0.346551	0.021836	0.025662	0.029823	0.033065	0.055560	0.112514	-0.543595	-0.339756	0.113480
MACD(12-26)	0.053888	0.053888	0.129577	0.047532	0.015360	-0.041022	-0.075309	-0.136718	-0.160707	0.060084	0.043395	-0.054276
MACD(mean(5,20)-mean(60,120))	0.022257	0.022257	0.130461	0.017397	-0.004061	-0.034205	-0.058830	-0.173507	-0.340823	0.201886	0.160996	-0.125644

**\*과적합**: 훈련데이터의 학습에 몰두한 나머지 검증, 테스트 데이터의 학습 일반화 성능에 지장이 가는 것.

# 모델 학습 및 검증 - 데이터 분리



훈련 데이터

---

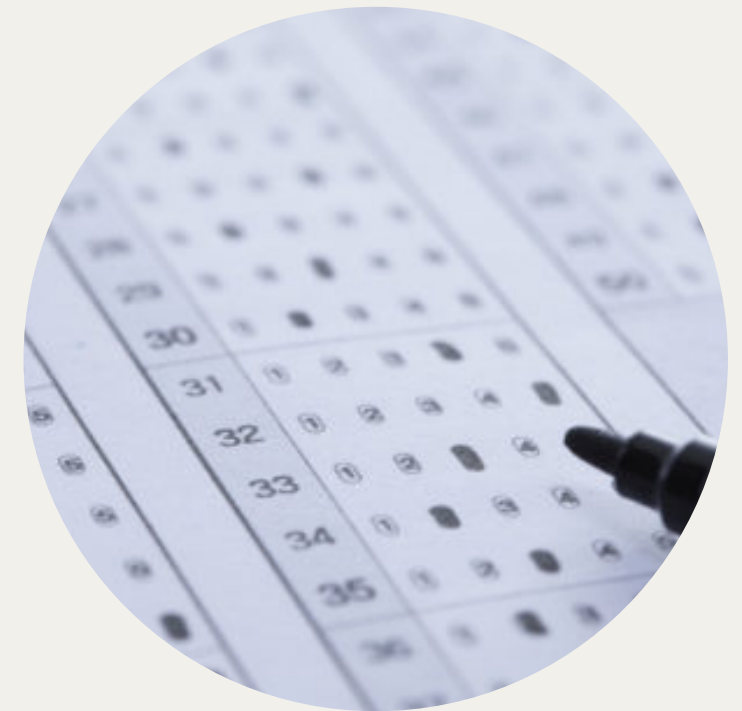
2017년 11월 22일 이전 데이터들을  
7:3으로 나누고 7의 비중에 해당.



검증 데이터

---

2017년 11월 22일 이전 데이터들을  
7:3으로 나누고 3의 비중에 해당.



테스트 데이터

---

훈련:테스트 = 7:3의 비중으로  
2017년 11월 22일 이후 데이터부터  
2023년 4월 10까지의 데이터까지.

# 검증 vs 테스트 데이터 모델 비교

훈련 정확도 0.8146093388811836  
검증 정확도 0.5900755124056095  
차이 0.22453382647557407

	precision	recall	f1-score	support
0	0.56	0.52	0.54	430
1	0.61	0.64	0.62	497
accuracy			0.59	927
macro avg	0.58	0.58	0.58	927
weighted avg	0.58	0.59	0.58	927

## 검증데이터

토론하고자 하는 내용을  
간략하게 설명해 주세요.

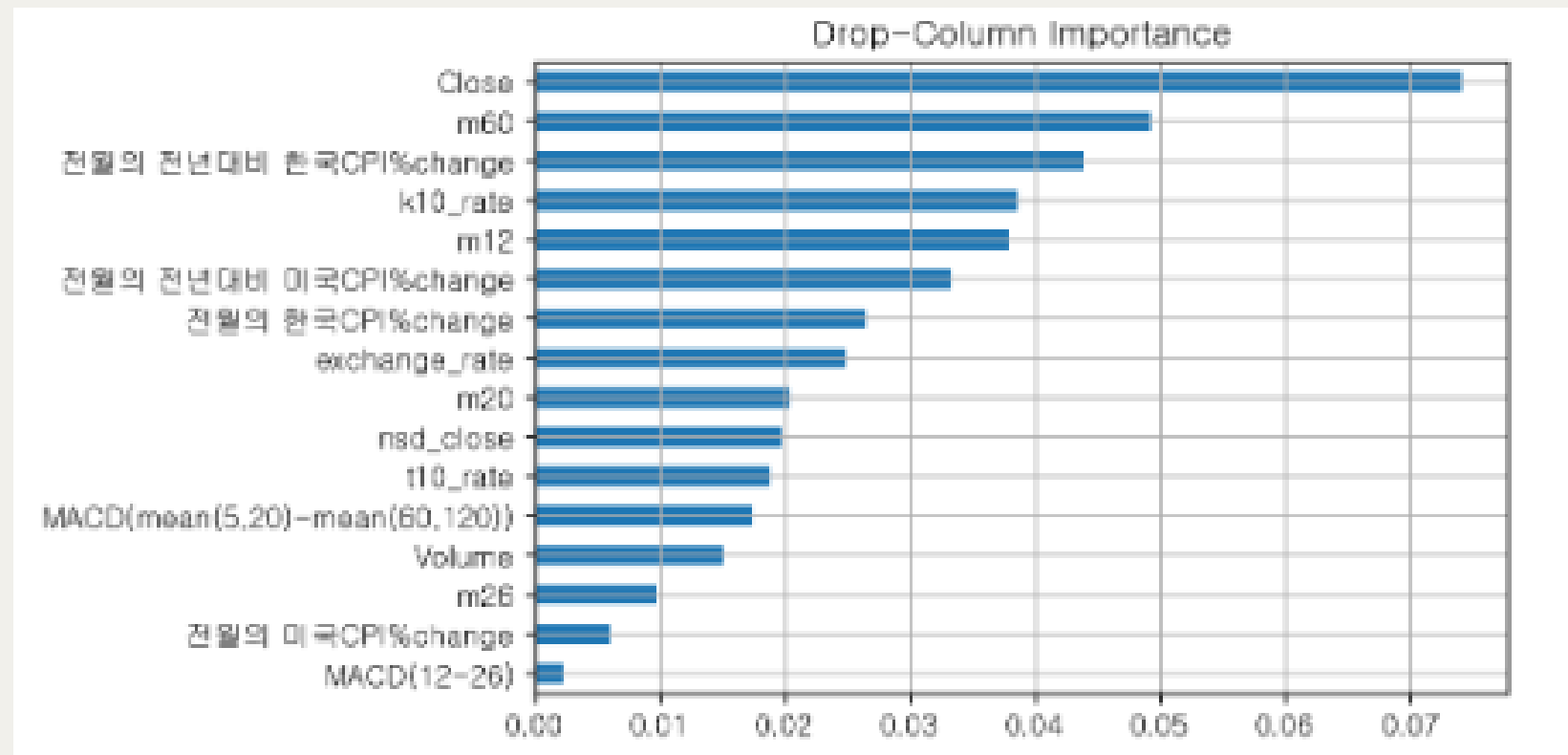
	precision	recall	f1-score	support
0	0.47	0.31	0.38	608
1	0.55	0.70	0.61	714
accuracy			0.52	1322
macro avg	0.51	0.51	0.50	1322
weighted avg	0.51	0.52	0.51	1322

## 테스트 데이터

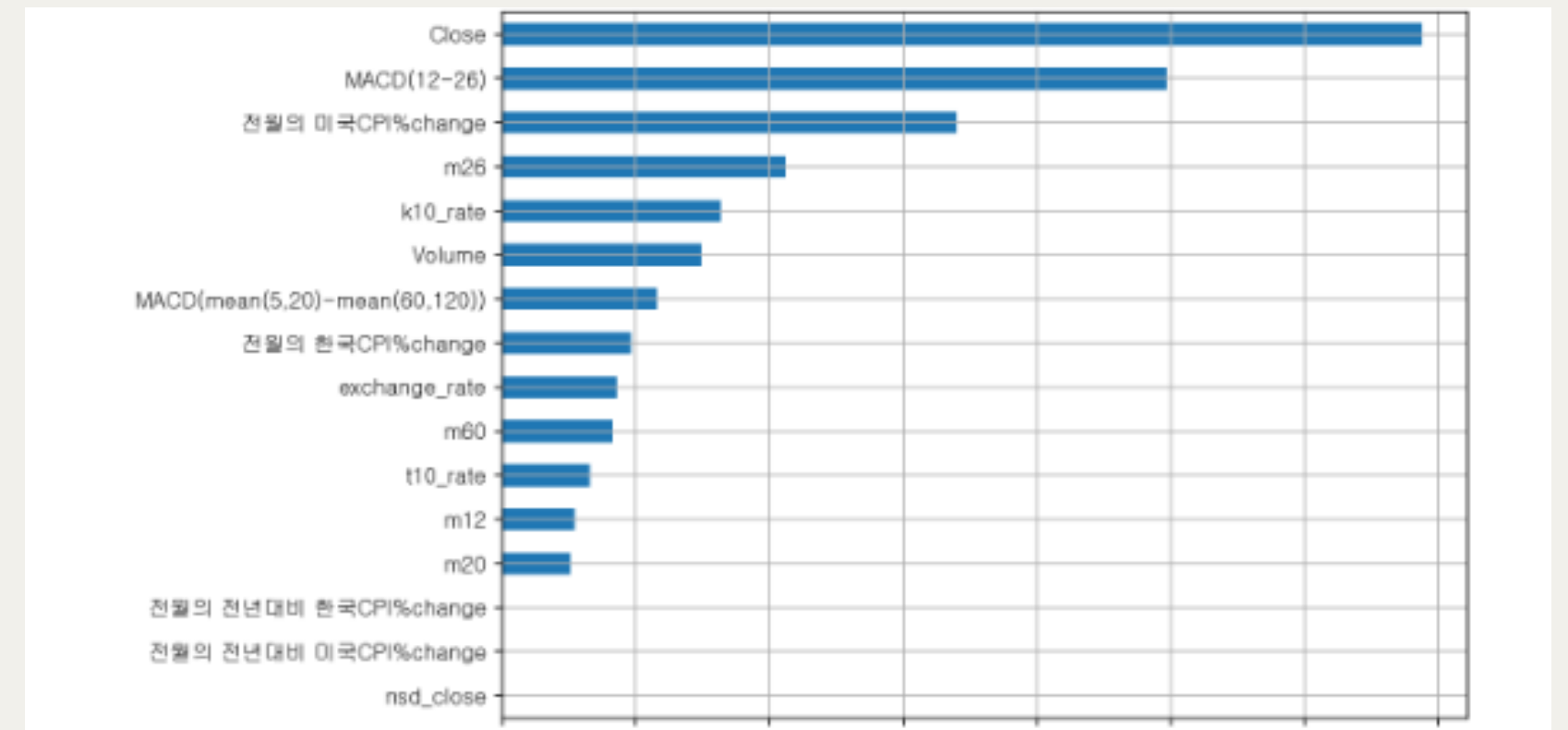
토론하고자 하는 내용을  
간략하게 설명해 주세요.



# 모델 해석-1



Drop-Column Importance



Permutation Importance



# 모델 해석-2

