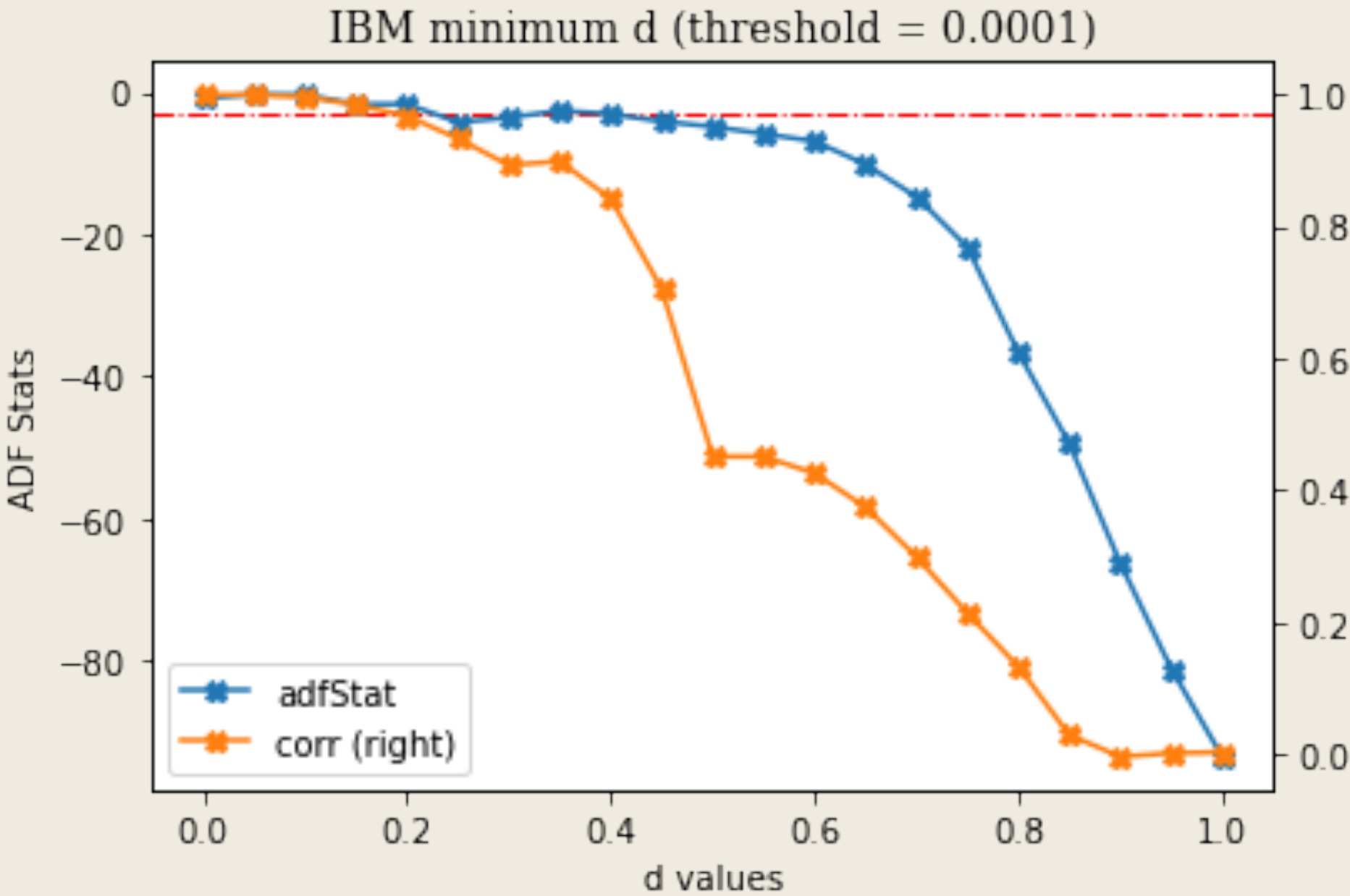


Final Projects

KIM



Set d = 0.45

	adfStat	p-value	lags	nObs	95% conf	corr
0.00	-0.892083	7.905684e-01	1.0	14597.0	-2.861738	1.000000
0.05	-0.325455	9.218210e-01	1.0	4.0	-4.474365	0.999007
0.10	-0.381350	9.131663e-01	1.0	4.0	-4.474365	0.996126
0.15	-1.720336	4.206663e-01	1.0	6.0	-3.646238	0.984297
0.20	-1.707622	4.271898e-01	1.0	8.0	-3.367187	0.967079
0.25	-4.244201	5.541346e-04	1.0	13.0	-3.127149	0.932946
0.30	-3.615943	5.463453e-03	1.0	19.0	-3.031227	0.892474
0.35	-2.630868	8.678379e-02	1.0	30.0	-2.964071	0.898378
0.40	-3.103146	2.632127e-02	1.0	47.0	-2.925338	0.841688
0.45	-4.141121	8.275645e-04	1.0	75.0	-2.900925	0.704957
0.50	-4.875349	3.893941e-05	1.0	121.0	-2.885739	0.450778
0.55	-5.884133	3.029544e-07	1.0	199.0	-2.876176	0.450273
0.60	-6.930087	1.089753e-09	1.0	331.0	-2.870312	0.425243
0.65	-10.169004	7.171881e-18	1.0	561.0	-2.866706	0.373761
0.70	-14.731459	2.655691e-27	1.0	965.0	-2.864540	0.298526
0.75	-21.812931	0.000000e+00	1.0	1682.0	-2.863260	0.213255
0.80	-36.617560	0.000000e+00	1.0	2936.0	-2.862525	0.131831
0.85	-49.424505	0.000000e+00	1.0	5020.0	-2.862116	0.028248
0.90	-66.265741	0.000000e+00	1.0	8073.0	-2.861898	-0.004566
0.95	-81.075013	0.000000e+00	1.0	11613.0	-2.861789	0.000493
1.00	-93.448562	0.000000e+00	1.0	14596.0	-2.861738	0.002246

Final Projects

KIM

```
from FinancialMachineLearning.features.microstructure import *
#Corwin-Schultz Spread Estimator : 하루 중 최고가와 최저가를 통해 당일 거래 스프레드 추정
spread = CorwinSchultz(high = data.High, low = data.Low)
cs_estimator = spread.corwin_schultz_estimator(window = 20)
data['corwin_schultz'] = cs_estimator

#Becker-Parkinson Volatility Estimator: 일일 최고가와 최저가를 사용해 변동성 추정
bp_vol = spread.becker_parkinson_vol(window = 20)
data['becker_parkinson'] = bp_vol
```

Add technical Features and Microstructure features

MFI : money flow index

OBV : on balance volume

RSI : relative strength index

VWAP : volume weighted average price

TSI : true strength index

EOM : ease of movement

VPT : volume price trend

Corwin schultz spread estimator

Becker Parkinson volatility estimator

```
from ta.volatility import BollingerBands
from ta.volume import money_flow_index, on_balance_volume
from ta.momentum import rsi
from ta.volume import volume_weighted_average_price
from ta.momentum import tsi
from ta.volume import ease_of_movement
from ta.volume import volume_price_trend

#ibm 수익률
data['return'] = data_ret

#20일 이동평균선 기준布林저 밴드
indicator_bb = BollingerBands(close = data["Close"], window = 20, window_dev = 1)

data['bb_bbm'] = indicator_bb.bollinger_mavg() #중간
data['bb_bbh'] = indicator_bb.bollinger_hband() #상단
data['bb_bbl'] = indicator_bb.bollinger_lband() #하단

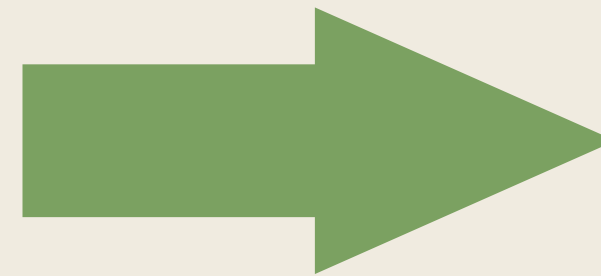
#MFI(Money Flow Index, 자금 흐름 지수)
#주가와 거래량을 결합해 계산. 매수와 매도 압력을 측정
data['mfi'] = money_flow_index(
    high = data.High, low = data.Low,
    close = data.Close, volume = data.Volume, window = 20
)
#OBV(On-Balance Volume, 잔고량 지표)
#거래량의 변화를 통해 주가의 상승 또는 하락 추세를 예측하는 지표
data['obv'] = on_balance_volume(close = data.Close, volume = data.Volume)
#RSI(Relative Strength Index, 상대 강도 지수)
#주식이 과매수 혹은 과매도 상태인지 판단하는 지표
data['rsi'] = rsi(close = data.Close, window = 21)
#VWAP(Volume Weighted Average Price, 거래량 가중 평균 가격)
#주어진 기간동안의 거래량을 고려한 평균 가격
data['vwap'] = volume_weighted_average_price(
    high = data['High'], low = data['Low'],
    close = data['Close'], volume = data['Volume'], window = 20)
#True Strength Index(TSI, 진정한 강도 지수)
#가격 모멘텀을 기반으로 주식의 추세를 판단한 지표
data['tsi'] = tsi(close = data['Close'], window_fast = 13, window_slow = 15)
#Ease of Movement(EOM, 용이성 지수)
#주가의 움직임이 얼마나 쉬웠는지를 나타내는 지표. 거래량과 가격 변화 모두 고려
data['eom'] = ease_of_movement(
    high = data['High'], low = data.Low, volume = data.Volume, window = 20
)
#Volume Price Trend(VPT, 거래량 가격 추세)
#거래량의 변화와 가격 움직임을 결합해 추세의 강도를 측정
data['vpt'] = volume_price_trend(
    close = data.Close,
    volume = data.Volume
)
```

Final Projects

KIM

PRIMARY MODEL

- Set side using Bollinger band
- 1 : up
- 0 : sideways
- -1 : down



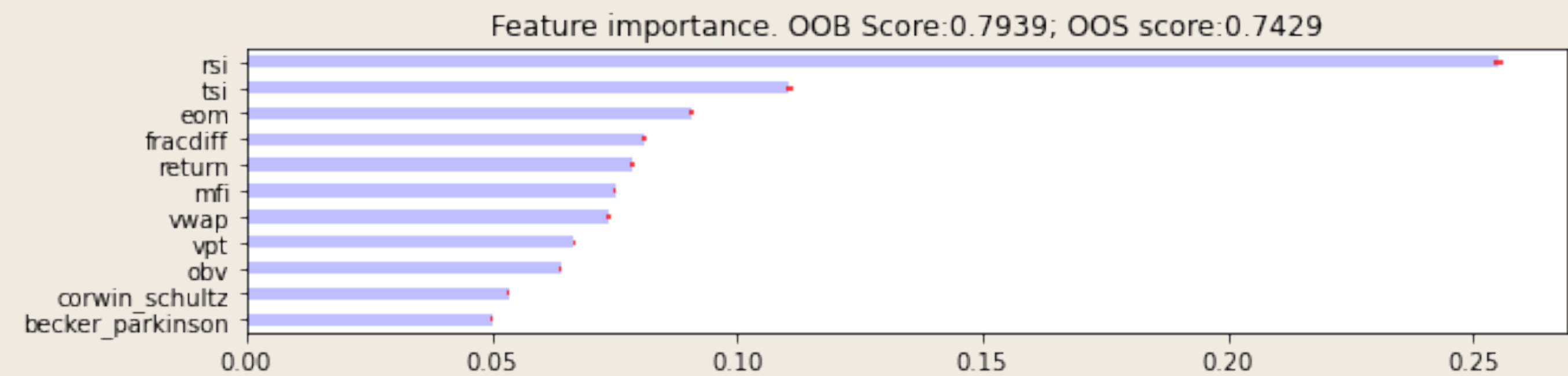
META MODEL

- Based on primary model, meta model only *predict trading* signals

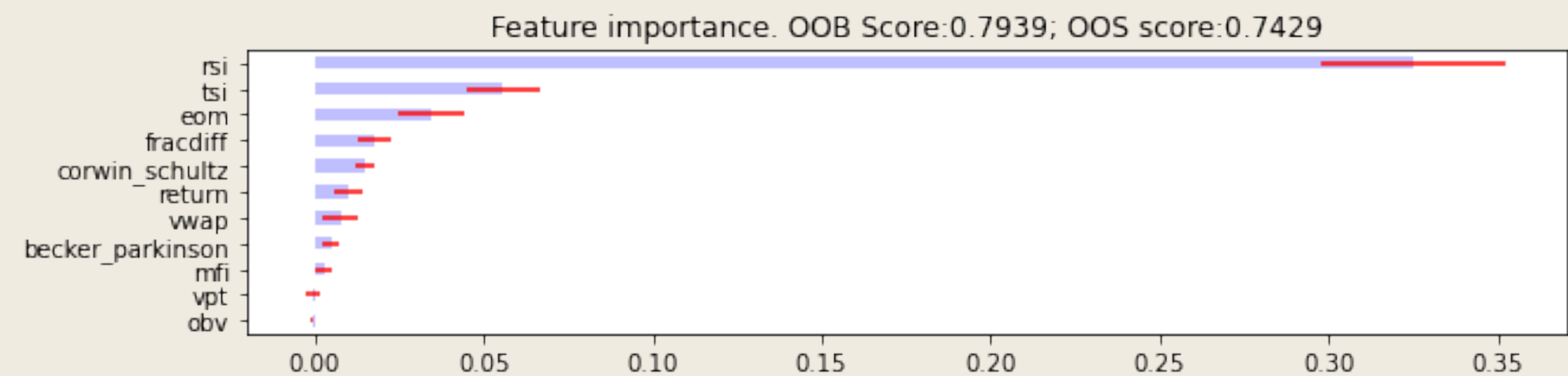
```
criterion = 'entropy',  
class_weight = 'balanced_subsample',  
min_weight_fraction_leaf = 0.0,  
random_state = 42,  
n_estimators = 300,  
max_features = 6,  
oob_score = True,  
n_jobs = 1
```


Final Projects

KIM



MDI feature importance



MDA feature importance

Final Projects

KIM

	precision	recall	f1-score	support
0	0.71	0.66	0.68	450
1	0.74	0.78	0.76	567
accuracy			0.73	1017
macro avg	0.72	0.72	0.72	1017
weighted avg	0.73	0.73	0.73	1017

Recall score of label 1 : **0.78**
F1 score of label 1 : **0.76**
AUC : **0.82**

