



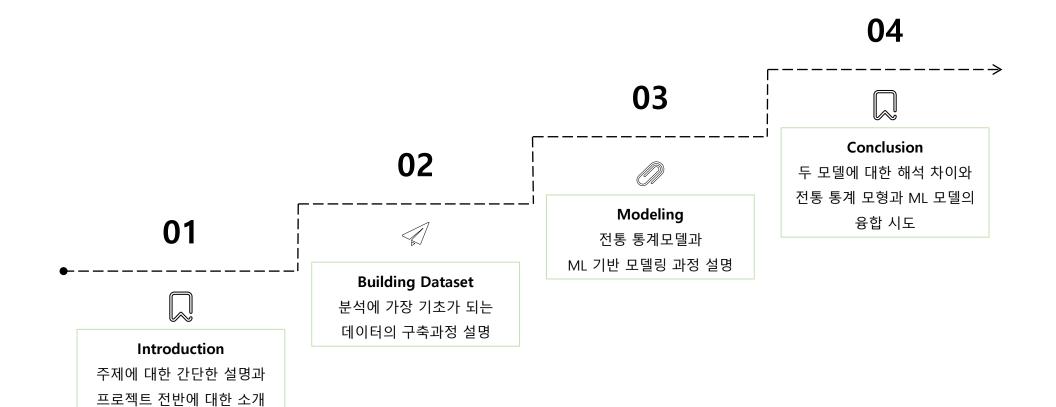
# 시계열 통계모델과 머신러닝 모델의 변수 유의성 판단 비교 및 시사접

Team Deepconomics B782040 정성은 C035006 강서연

# 목차















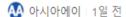


#### 세계 금융계 동향

#### Computing Power의 향상 + 머신러닝 기술



#### 🦈 경제학계에서도 많은 활용

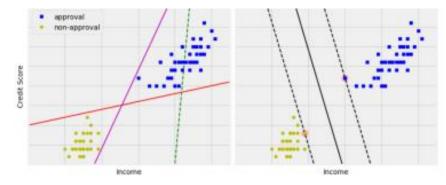


농협은행, 경기예측모형 도입...경제 변수 최대 30개 반영 각 경제 변수에 관해 머신러닝 기법 활용, 발생확률 예측, 경기충격 파급효과를 고 려한다. 이미 미국 중앙은행인 연방준비제도(Fed)와 유럽 중앙은행 등에서 LBVA...



NH농협은행, 경기예측모형 도입 뉴시스 1일 전 1네이버뉴스 농협은행, 은행권 최초 LBVAR 모형 도입 한국경제 | 1일 전 | 네이버뉴스 농협은행, 국내 은행권 최초 LBVAR 모형 도입 서울경제 11일 전 1네이버뉴스 농협은행, 최신 경기예측 모형 도입...스트... 머니투데이 11일 전 1네이버뉴스

관련뉴스 25건 전체보기 >









## 머신러닝

- ♥ '예측'의 문제 해결에 탁월한 성능
- ▼ 모델에 대한 별다른 가정 없이 특성변수를 기반으로 반응변수의 예측치의 정확도를 획기적으로 높일 수 있는 알고리즘을 제공
- ♥ 특성변수와 반응변수와의 관계에 대해 구조적인 해석이 쉽지 않은 'Black Box'

목적 반응변수의 추정치 예측의 정확성 향상







전통 ML 모델, 머신러닝을 활용해,

외국인 투자에 미치는 다양한 요인에 대한 분석

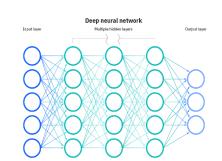






머신러닝

결론에 도달한 과정을 파악하기 적절함 해석의 차이를 얻을 수 있음



딥러닝

판단에 대한 설명을 얻기 까다로움 결론에 도달한 과정을 파악하기 부적절함

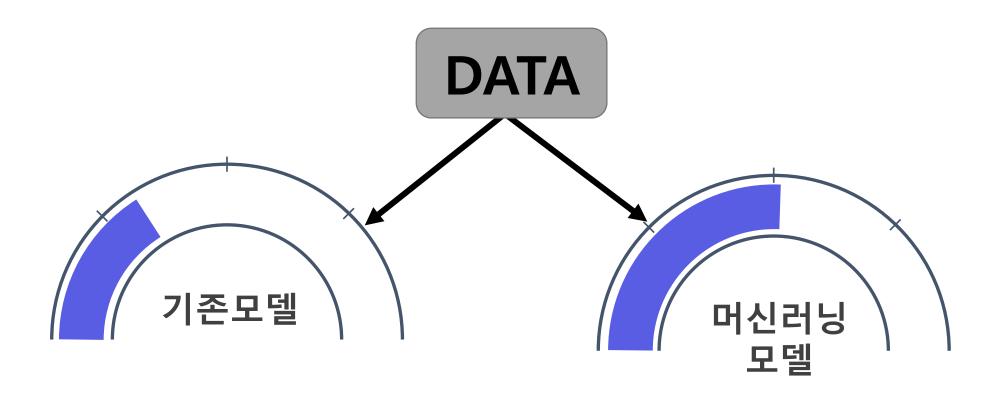


**"해석"** 의 차이를 얻는 것이 주된 목적이기에 머신 러닝 모델 채택









두 모델이 중요하게 바라보는 요인 분석

두 모델의 성능 비교 및 두 분야가 공존할 수 있는 방향에 대해 탐색







#### Dataset 구축

교과서, 여러 논문 참고



외국인 투자액에 영향을 미치는 다양한 변수 선정





#### Feature(선정된 독립변수)

MSCI지수, FTSE지수, 외환보유액, 외평채 스프레드, 국고채, CD, 콜금리, 기준금리, 미연준금리













#### 학습 속도, 활용성 고려

# 월별 데이터 이용 2004년 ~ 2022년 9월까지의 데이터를 종합하여 구축

	외국인_보	외환보유 <sup>민</sup> F	TSE	MSCI	외평채_20 외	평채_20 외평	명채_20 외	평채_20 외원	3재_20 외원	제_20 외원	3재_20 의원	[제_20] 외	8재_20 외원	채_20. 외평	채_20. 외평	8채_20 외	방재_20 국	고재_3년 =	: 과재_5년 =	t ᆚ재_10 의	사제_3년 대	)	콜금리 7	기준금리	연준금리 /	AAII	VIX
0408월	165.6	1705	103.86	237.37	69	0	0	0	0	0	0	0	0	0	0	0	0	3.82	4.01	4.33	4.38	3.7	3.6	3.5	1.429	2.975	15.2
0409월	173.1	1745	107.45	245.87	75	0	97	0	0	0	0	0	0	0	0	0	0	3.59	3.75	4.14	4.1	3.52	3.52	3.5	1.605	15.94	13.3
0410월	171.3	1784	107.08	243.79	56	0	73	0	0	0	0	0	0	0	0	0	0	3.54	3.67	4.05	4.01	3.52	3.52	3.5	1.76	19.5	16.2
0411월	179.1	1926	111.54	252.75	62	0	85	0	0	0	0	0	0	0	0	0	0	3.42	3.53	3.96	3.86	3.42	3.34	3.25	1.932	34.9	13.2
0412월	179.7	1991	113.59	256.4	66	0	82	0	0	0	0	0	0	0	0	0	0	3.28	3.39	3.85	3.73	3.39	3.27	3.25	2.156	34.6	13.2
0501월	190.4	1997	117.43	269.87	61	0	72	0	0	0	0	0	0	0	0	0	0	3.65	3.89	4.42	4.11	3.47	3.26	3.25	2.279	-3.075	12.8
0502월	206.4	2022	127.22	290.05	61	0	70	0	0	0	0	0	0	0	0	0	0	4.19	4.47	4.88	4.63	3.59	3.26	3.25	2.501	11	12.0
0503월	195	2055	120.84	275.98	74	0	83	0	0	0	0	0	0	0	0	0	0	4.04	4.29	4.77	4.48	3.55	3.24	3.25	2.629	-0.82	14.0
0504월	182.6	2064	112.99	257.79	78	0	85	0	0	0	0	0	0	0	0	0	0	3.86	4.11	4.69	4.28	3.52	3.25	3.25	2.784	-10.6	15.3
0505월	194	2061	120.01	273.48	71	0	77	0	0	0	0	0	0	0	0	0	0	3.71	3.89	4.42	4.12	3.51	3.29	3.25	3.002	2.225	13.2
0506월	200.7	2050	123.44	281.07	64	0	73	0	0	0	0	0	0	0	0	0	0	3.79	3.97	4.46	4.18	3.51	3.27	3.25	3.035	25.92	12.0
0507월	223.8	2057	137.45	313.56	60	0	68	0	0	0	0	0	0	0	0	0	0	4.12	4.39	4.89	4.52	3.51	3.24	3.25	3.262	30.025	11.5
0508월	218.9	2067	134.56	307.52	63	0	74	0	0	0	0	0	0	0	0	0	0	4.33	4.67	5.08	4.75	3.5	3.24	3.25	3.499	6.575	12.
0509월	244	2067	151.19	345.21	57	0	68	0	0	0	0	0	0	0	0	0	0	4.52	4.81	5.15	4.92	3.7	3.24	3.25	3.622	5.26	11.9
0510월	229.2	2073	142.12	324.28	62	0	72	23	0	0	0	0	0	0	0	0	0	4.81	5.06	5.36	5.2	3.93	3.45	3.5	3.779	1.65	15.3
0511월	254.5	2082	158.61	360.28	65	0	74	25	0	0	0	0	0	0	0	0	0	5.1	5.39	5.68	5.5	3.96	3.49	3.5	3.999	31.025	12.0
0512월	270.2	2104	170.85	386.26	67	0	74	23	0	0	0	0	0	0	0	0	0	5.06	5.33	5.6	5.49	4.04	3.7	3.75	4.157	19.42	12.0
0601월	281.7	2169	176.16	398.14	64	0	65	18	0	0	0	0	0	0	0	0	0	5.02	5.29	5.58	5.5	4.15	3.74	3.75	4.285	12.65	12.9
0602월	276.9	2160	171.3	387.84	62	0	63	15	0	0	0	0	0	0	0	0	0	4.87	5.03	5.28	5.33	4.25	3.92	4	4.491	9.075	12.8
0603월	277.2	2173	170.44	386.03	76	0	75	19	0	0	0	0	0	0	0	0	0	4.93	5.13	5.42	5.3	4.27	3.97	4	4.589	11.4	11.3
0604월	287.5	2229	178.12	400.15	65	0	65	22	0	0	0	0	0	0	0	0	0	4.95	5.16	5.45	5.24	4.33	3.97	4	4.79	10.25	11.5
0605월	264.6	2247	165.18	370.41	67	0	70	24	0	0	0	0	0	0	0	0	0	4.78	4.93	5.2	5.05	4.36	3.96	4	4.94	5.75	16.4
0606월	259	2244	162.68	365.32	80	0	81	24	0	0	0	0	0	0	0	0	0	4.89	5.03	5.25	5.17	4.47	4.17	4.25	4.991	-15.02	13.0
0607월	256.5	2257	164.19	367	71	0	73	20	0	0	0	0	0	0	0	0	0	4.89	5	5.16	5.18	4.63	4.21	4.25	5.235	-12.475	14.9
0608월	261.9	2270	169.7	378.38	69	0	70	19	0	0	0	0	0	0	0	0	0	4.77	4.83	4.94	5.1	4.68	4.42	4.5	5.246	-1.66	12.3
0609월	263.4	2282	172.49	383.95	68	0	69	17	0	0	0	0	0	0	0	0	0	4.69	4.76	4.89	5.02	4.64	4.47	4.5	5.254	13.65	11.9
0610월	260.8	2295	171.06	380.6	67	0	67	17	0	0	0	0	0	0	0	0	0	4.62	4.67	4.79	4.93	4.57	4.48	4.5	5.245	12.1	11.
0611월	271	2343	178.39	396.38	64	0	68	17	0	0	0	0	0	0	0	0	0	4.72	4.79	4.91	5.05	4.6	4.47	4.5	5.245	7.94	10.9
0612월	272.8	2390	178.26	395.18	60	0	64	23	70	0	0	0	0	0	0	0	0	4.82	4.88	4.95	5.18	4.76	4.5	4.5	5.238	1,775	11.50
0701월	277.6	2402	170.96	379.3	66	0	66	19	72	0	0	0	0	0	0	0	0	4.97	5	5.04	5.32	4.92	4.63	4.5	5.248	16,675	10.4
0702월	291.8	2428	178.59	395.57	64	0	64	26	73	0	0	0	0	0	0	0	0	4.91	4.94	5.01	5.28	4.95	4.6	4.5	5.258	20	15.4
0703월	297.9	2439	182.26	404.75	63	0	62	24	72	0	0	0	0	0	0	0	0	4.79	4.82	4.94	5.2	4.94	4.6	4.5	5.256	0.68	14.6
0704월	314.3	2473	191.65	424.43	59	0	60	20	68	0	0	0	0	0	0	0	0	4.94	4.96	5.07	5.34	4.95	4.69	4.5	5.249	3.5	14.2
0705월	336.4	2507	208.11	458.21	58	0	58	16	66	0	0	0	0	0	0	0	0	5.1	5.13	5.22	5.49	5.04	4.61	4.5	5.253	-7.32	13.0
0706월	343.1	2507	213.08	469.6	65	0	63	17	69	0	0	0	0	0	0	0	0	5.24	5.35	5.48	5.64	5.03	4.54	4.5	5.254	1.1	16.2
0707월	369.9	2548	235.74	518.36	94	0	95	17	100	0	0	0	0	0	0	0	0	5.35	5.43	5.52	5.75	5.05	4.67	4.75	5.258	9.275	23.5
0708월	348.4	2553	229.65	505.53	97	0	87	30	98	0	0	0	0	0	0	0	0	5.28	5.33	5.42	5.7	5.21	4.92	5	5.022	0.32	23.37
709월	359.8	2573	241.45	529.84	93	0	74	29	86	0	0	0	0	0	0	0	0	5.39	5.44	5.52	5.85	5.34	4.98	5	4.938	5.875	1/
0710월	372.1	2601	254.43	558.47	93	0	73	29	86	0	0	0	0	0	0	0	0	5.42	5.49	5.56	5.98	5.34	5	5	4.755	11,175	18.5
7711월	344.7	2619	235.33	516.72	131	0	100	43	117	0	0	0	0	0	0	0	0	5.54	5.59	5.61	6.16	5.43	5	5	4.486	-15.64	22.8
0712월	344.2	2622	235.96	516.89	130	0	98	45	110	0	0	0	0	0	0	0	0	5.89	5.9	5.82	6.73	5.73	499	5	4.244	-4.65	22.0
00012	270.2	2022	200.00	447.04	140	0	172	45 F0	120	0	0	0	0	0	0	0	0	5.05	5.5	5.02	0.75	5.75	4.00		2.04	20.22	201

#### 시계열조회

지표명	외환보유액		
통계표명	외환보유액 🗸 초기화		
주기	월 🗸	기간	[199701 ✔] ~ [202210 ✔] 조희

#### ○ 통계표

[단위 : **억달러, %**] 📳 엑셀저장

	1997 01월	1997 02월	1997 03월	1997 04월	1997 05월	1997 06월	1997 07월	1997 08월	1997 09월	1997 10월	1997 11월	1997 12월	1998 01월	1998 02월	1998 03월	1998 04월	1998 05월
외환보유액	272	228	211	218	239	253	257	231	224	223	73	89	124	185	242	308	34

🎒 출처 : 한국은행 (ECOS 경제통계시스템)

☑ 주석: \* 매 연도/분기/월말 기준

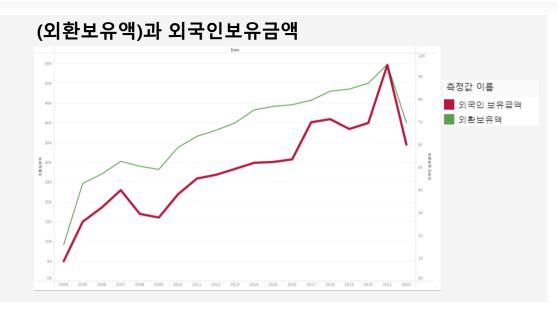
• 한국은행 경제통계정보 시스템(ECOS)통계 및 통계청 통계와 차이가 나는 것은 한국은행의 경우 1999년 8월까지는 과거 통계기준인 총외환보유액 기 준으로 작성하고 그 이후는 가용외환보유액 기준(총외환보유액에서 국내은행의 해외점포 예치금 등을 제외한 금액)으로 작성하였으나 e-나라 지표의 경우 통계 기준의 일관성 유지를 위해 1999년 8월이전이라도 현재의 기준인 가용외환보유액 기준으로 작성하여 게시함에 따른 차이일 뿐임으로 연구 목적이나 필요한 통계 요구사항에 따라 취사 선택이 필요

\* 외환보유액 시계열을 제외한 보조지표는 연도말 기준으로 편재함

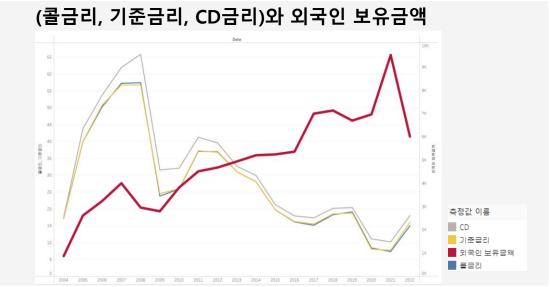


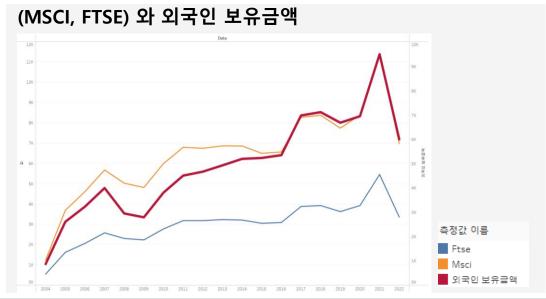










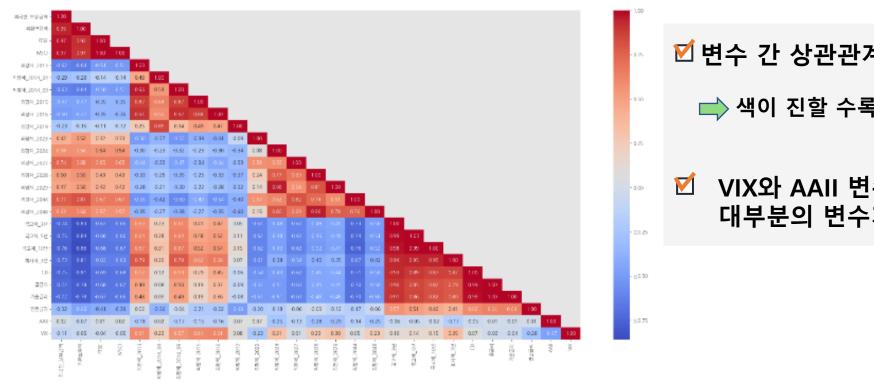






# 다중 공선성 파악 1. Correlation matrix Heatmap으로 시각화

#### **Correlation matrix Heatmap**



ゼ 변수 간 상관관계를 파란색~빨간색으로 표현

➡ 색이 진할 수록, 두 변수 간의 상관관계가 높다

VIX와 AAII 변수를 제외하고는 대부분의 변수가 다른 변수와 상관관계 존재



# 다중 공선성 파악 2. VIF(Variance Inflation Factor, 분산팽창요인)

#### **VIF**

- ☑ 1~ 무한대의 값을 가짐
- ☑ 5를 넘기면 주의
- ☑ 10을 넘기면 다중 공선성이 존재한다고 판단

변수명	VIF	변수명	VIF
FTSE	50811.62	외평채 _2029	54.29
MSCI	43556.42	외평채 _2014_09	46.47
국고채_5년	15621.93	외평채 _2013	40.94
국고채_3년	6252.25	외평채 _2016	31.26
국고채_ 10년	4787.25	VIX	23.33
콜금리	2831.92	연준금리	21.52
외국인 보유금액	2473.36	외평채 _2015	18.08
기준금리	1895.49	외평채 _2044	17.33
CD	1260.44	외평채 _2019	15.33
외환보유액	547.05	외평채 _2027	15.12
회사채_3년	447.78	외평채 _2014_04	15.12
외평채_ 2048	92.83	외평채 _2023	10.81
외평채_ 2028	91.60	AAII	2.09
외평채_ 2024	62.70		

VIF측정 결과







# 정상성 검정 - ADF Test(Augmented Dickey-Fuller Test) & KPSS(Kwiatkowski-Phillips-Schmidt-Shin) 교차 검정

변수명	KPSS	ADF	정상성 여부	변수명	KPSS	ADF	정상성 여부
외국인_ 보유금액	0.01	0.566	FALSE	외평채 _2028	0.01	0.243	FALSE
외환보유액	0.01	0.569	FALSE	외평채 _2029	0.01	0.25	FALSE
FTSE	0.01	0.297	FALSE	외평채 _2044	0.01	0.797	FALSE
MSCI	0.01	0.27	FALSE	외평채 _2048	0.01	0.841	FALSE
외평채 _2013	0.01	0.171	FALSE	국고채_3년	0.01	0.477	FALSE
외평채 _2014_04	0.045	0.051	FALSE	국고채_5년	0.01	0.517	FALSE
외평채 _2014_09	0.01	0.137	FALSE	국고채_ 10년	0.01	0.565	FALSE
외평채 _2015	0.02	0.103	FALSE	회사채_3년	0.01	0.483	FALSE
외평채 _2016	0.01	0.525	FALSE	CD	0.01	0.326	FALSE
외평채 _2019	0.061	0.006	TRUE	콜금리	0.01	0.346	FALSE
외평채 _2023	0.01	0.279	FALSE	기준금리	0.01	0.349	FALSE
외평채 _2024	0.01	0.514	FALSE	연준금리	0.011	0.042	FALSE
외평채 _2027	0.01	0.518	FALSE	AAII	0.1	0	TRUE
				VIX	0.1	0	TRUE

**ADF Test** 

☑ P-value < 0.05 인 경우, 정상성 만족한다고 판단

**KPSS** 

☑ P-value > 0.05 인 경우, 정상성 만족한다고 판단

3개를 제외하고는 정상성 충족 못함

PCA 기법 & 차분(Differencing)기법 사용

정상성 검정 결과

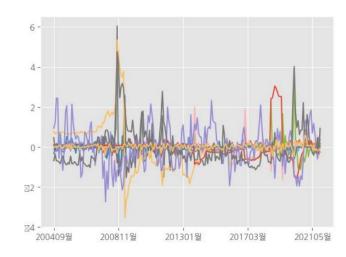


# **Final Dataset**



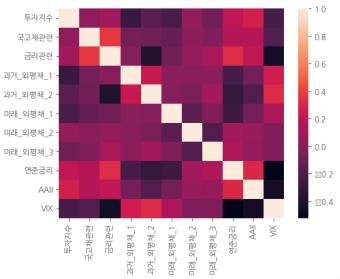


변수명	KPSS	ADF	정상성 여부	변수명	KPSS	ADF	정상성 여부
투자지수	0.1	0.000	TRUE	미래 외평채 1	0.1	0.000	TRUE
국고채관련	0.1	0.004	TRUE	미래 외평채 2	0.1	0.000	TRUE
금리관련	0.1	0.000	TRUE	미래 외평채 3	0.1	0.000	TRUE
과거 외평채 1	0.1	0.000	TRUE	AAII	0.1	0.000	TRUE
과거 외평채 2	0.063	0.034	TRUE	VIX	0.1	0.000	TRUE
연준금리	0.1	0.005	TRUE				



정상성 문제 해결 완료! ✔

변수명	VIF	변수명	VIF
VIX	1.84	과거 외평채 1	1.257
금리관련	1.67	국고채 관련	1.21
연준금리	1.63	미래 외평채 1	1.20
과거 외평채 2	1.38	미래 외평채 2	1.15
AAII	1.33	미래 외평채 3	1.09
투자지수	1.258		



다중공선성 문제 해결 완료! 🗸







# **VAR** - Vector Autoregression

- ✓ 다변량 시계열 모델 중 기본
- ✓ 예측할 변수의 과거값 + 다른 변수들까지 고려 변수들 사이의 의존성, 상관관계 파악 가능
- ✓ 충격반응함수와 Variance Decomposition 기법과 함께 내부 변수들 간의 영향력 파악 가능

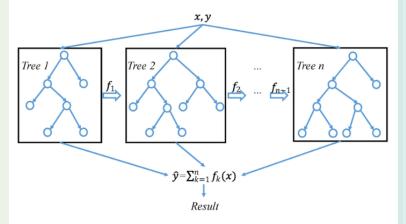






# **XGBoost** – Extream Gradient Boosting

- ✓ CART 기반의 Regression 문제, Classification 문제에서 뛰어난 성적
- Tabular Data를 다루는 업무에서 주로 사용
- ✓ Tree 기반의 모델
  - 장점) 데이터의 <mark>주어진 범위 내</mark>에서의 예측 능력<del>1</del> 단점) 데이터가 가지고 있는 범위 외의 예측을 해야 하는 경우, 예측 성능
- ✓ 정상성을 띄도록 데이터를 가공해주어야 한다는 주의점









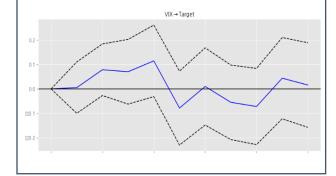
$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + e_t$$

#### 지연계수(Lag Order) 결정

- · p=1~p=20의 AIC 접수 파악
- · 14까지 낮아지고, 15부터 다시 반등하는 현상 발견
- · 이에 p=14로 결정하여 학습 진행
- · 실질적인 최적값인지 확인을 위해 forecasting까지 진행한 결과, 제일 MSE가 낮은 결과를 보임

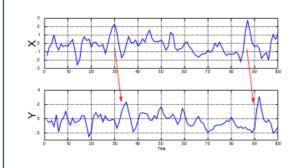
#### **Impulse Response Analysis**

- · 특정 변수의 충격(Impulse)이 다른 변수 에 미치는 영향을 분석하기 위함
- · Target 변수에 큰 영향을 준 변수를 파악 하는 데에 사용



#### **Granger Causality**

- · 변수간의 선후관계를 파악하기 위함.
- · 선후관계까지 고려하여 Target Data에 대한 변수의 영향력을 파악하고자 사용



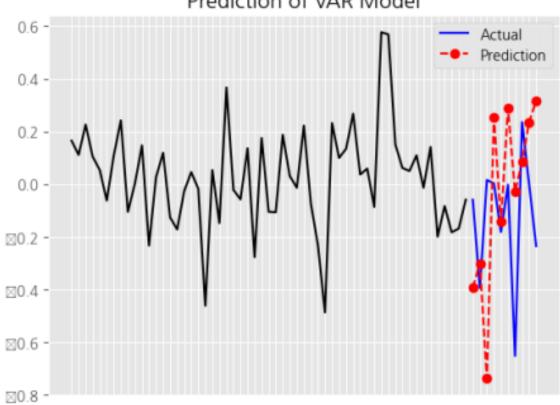






#### VAR모델의 예측 결과









예측을 함에 있어 변화의 방향성과 그 정도를 일부 맞춘 것으로 보여짐

> **But, '예측을 잘 한다**'라고 판단하기엔 **무리가 있음.**

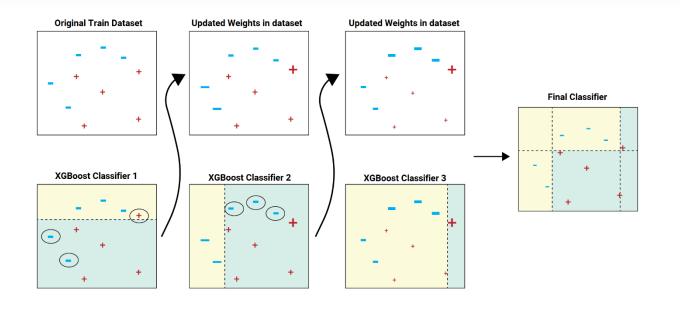
MSE: 0.161



#### **XGBoost**







- CART 기반 모델로 Regression에서 활용 가능
- Tree의 Regression : Entropy나 Gini 계수가 아닌 Error를 기준으로 Node Split Leaf에 할당 되어 있는 데이터들의 '평균값'을 최종 예측값으로 활용!
- Feature Importance와 Permutation Importance 기법을 적용하여 예측에 있어 중요한 변수 IIP 가능

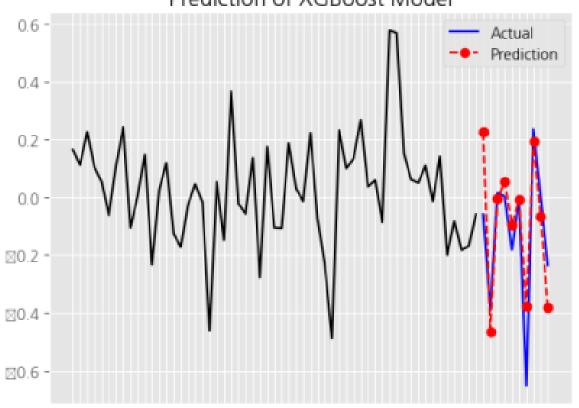






#### XGBoost 모델의 예측결과

#### Prediction of XGBoost Model







예측의 방향과 그 값을 VAR에 비해 잘 잡아내는 모습

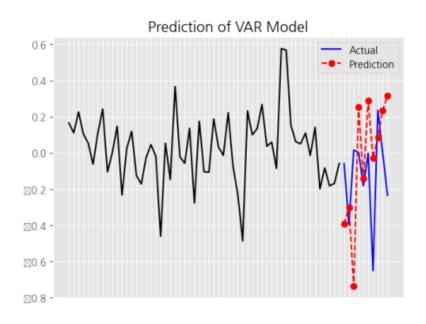
적절한 전처리를 거친 XGBoost 모델은 정상성을 가진 시계열 데이터에 대해 훌륭한 예측력을 보임

MSE: 0.020





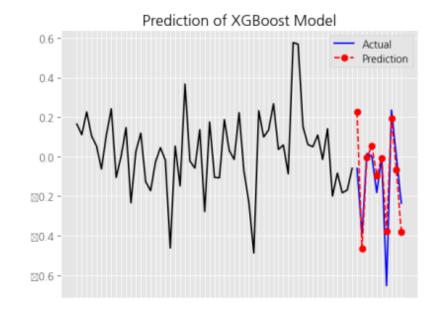




미래 외평채, 금리관련, AAII

변수들의 과거가 가지는 영향력을 포함하여 변수 사이의 상호 관계를 여러 통계적 기법을 통해 검증

이론적, 논리적 기반이 마련 되어 있는 결론



투자지수, 과거 외평채

데이터들이 가진 패턴을 파악하는 과정에서 나온 '부산물'로서 변수 중요도 파악

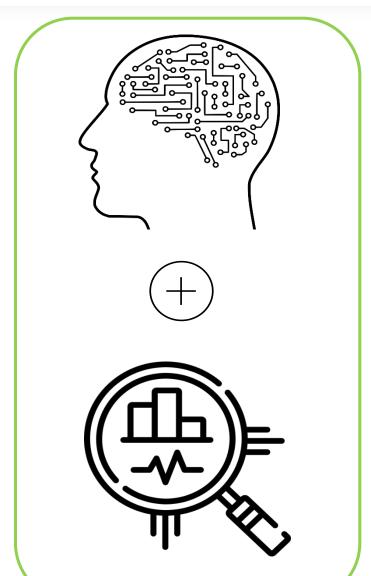
예측은 훌륭하지만 결과론적인 결론



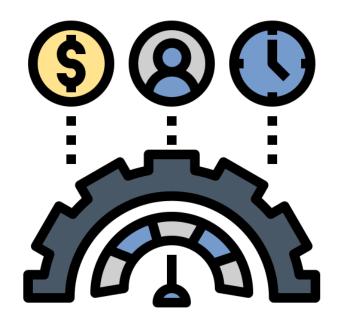
# **Conclusion**











Better Efficiency!

# 질문 있으신가요?













# Thank you!