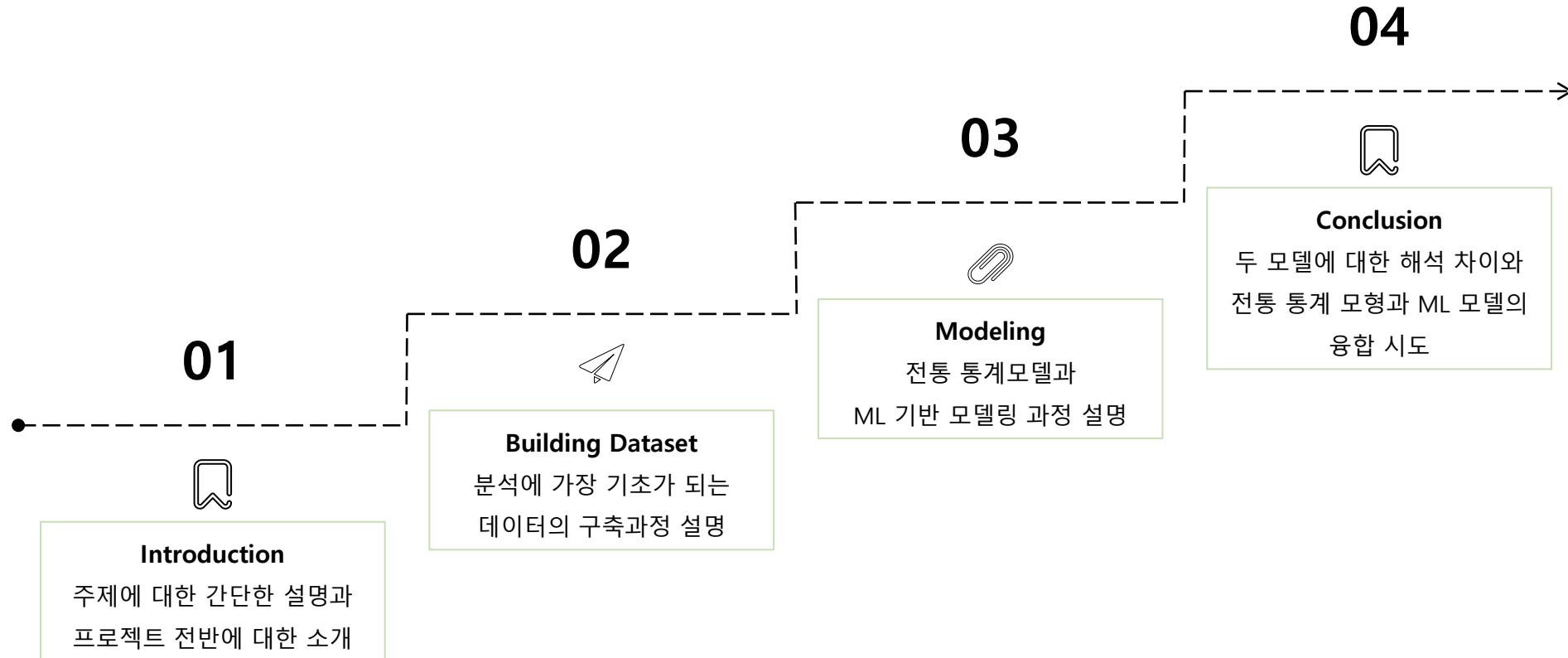




시계열 통계모델과 머신러닝 모델의 변수 유의성 판단 비교 및 시사점

Team Deepconomics
B782040 정성윤 C035006 강서연



세계 금융계 동향

Computing Power의 향상 + 머신러닝 기술

➡ 경제학계에서도 많은 활용

AA 아시아에이 | 1일 전

농협은행, 경기예측모형 도입...경제 변수 최대 30개 반영

각 경제 변수에 관해 머신러닝 기법 활용, 발생확률 예측, 경기충격 파급효과를 고려한다. 이미 미국 중앙은행인 연방준비제도(Fed)와 유럽 중앙은행 등에서 LBVA...

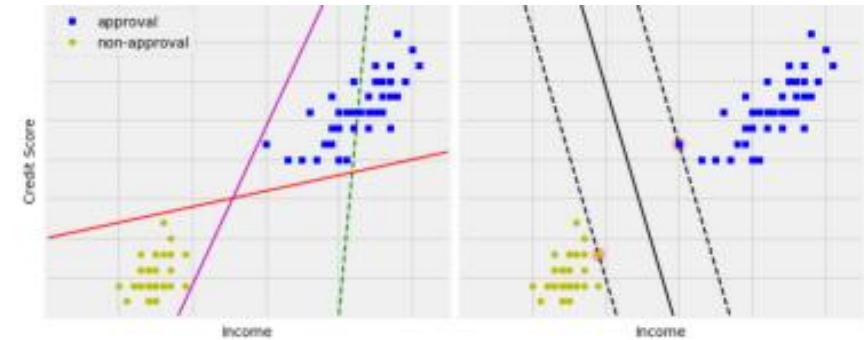
NH농협은행, 경기예측모형 도입 뉴시스 | 1일 전 네이버뉴스

농협은행, 은행권 최초 LBVAR 모형 도입 한국경제 | 1일 전 네이버뉴스

농협은행, 국내 은행권 최초 LBVAR 모형 도입 서울경제 | 1일 전 네이버뉴스

농협은행, 최신 경기예측 모형 도입...스트... 머니투데이 | 1일 전 네이버뉴스

관련뉴스 25건 전체보기 >



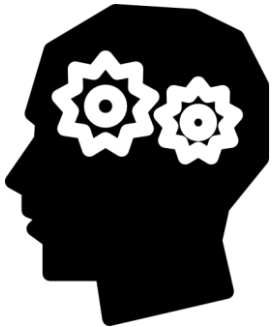
머신러닝

- ✓ '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

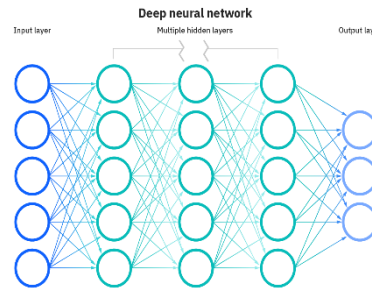
반응변수의 추정치 예측의 정확성 향상

전통 ML 모델, 머신러닝을 활용해,
외국인 투자에 미치는 다양한 요인에 대한 분석



머신러닝

결론에 도달한 과정을 파악하기 적절함
해석의 차이를 얻을 수 있음

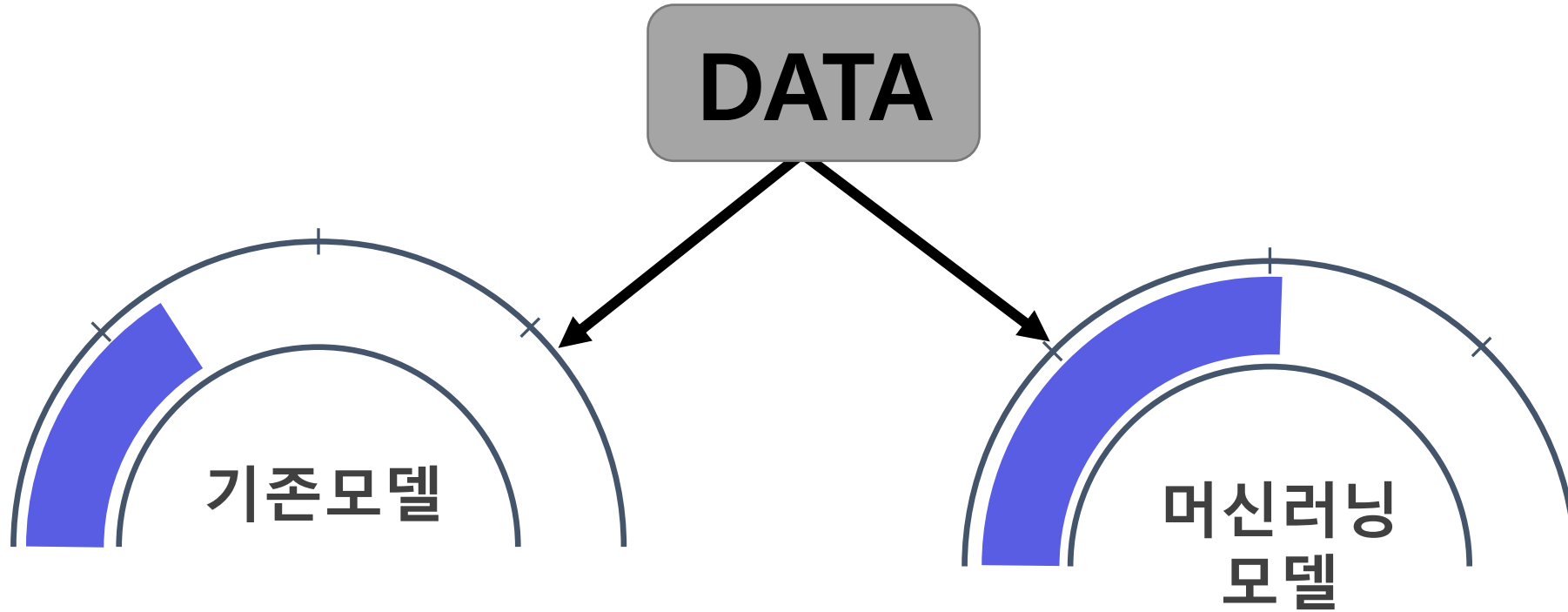


딥러닝

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 **공존**할 수 있는 **방향**에 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





Building Dataset



학습 속도, 활용성 고려

유니콘 데이터 이용

2004년 ~ 2022년 9월까지의 데이터를 종합하여 구축

[illegible]

시계열조치

지표명	외환보유액		
통계표명	외환보유액 ▼	초기화	
주기	월 ▼	기간	199701 ▼ ~ 202210 ▼ 조회

○ 통계표

[단위 : 억달러, %] 엑셀저장

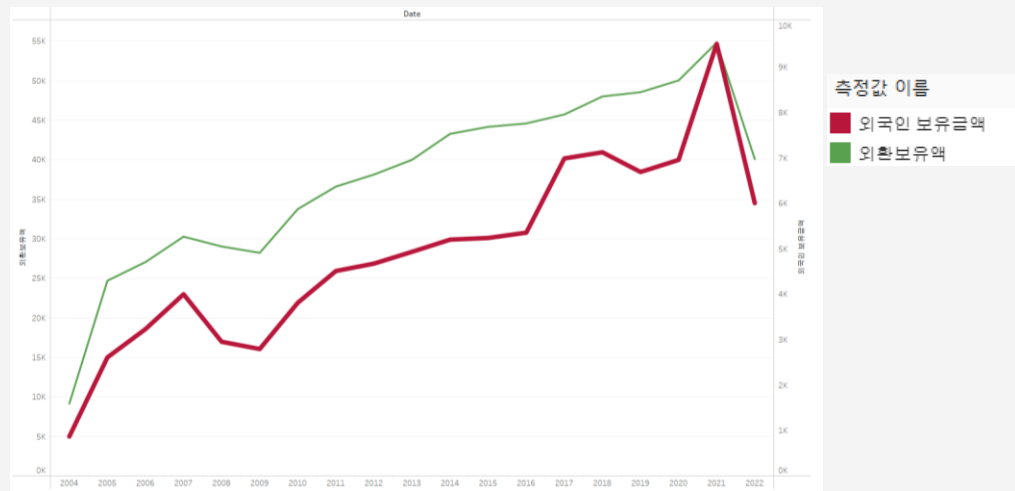
	1997 01월	1997 02월	1997 03월	1997 04월	1997 05월	1997 06월	1997 07월	1997 08월	1997 09월	1997 10월	1997 11월	1997 12월	1998 01월	1998 02월	1998 03월	1998 04월	1998 05월
	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000	000
외환보유액	272	228	211	218	239	253	257	231	224	223	73	89	124	185	242	308	

 출처 : 한국은행 (ECOS 경제통계시스템)

 주석 : * 매 연도/분기/월말 기준

- * 한국은행 경제통계정보 시스템(ECOS) 통계 및 통계청 통계와 차이가 나는 것은 한국은행의 경우 1999년 8월까지의 과거 통계기준인 총회원보유액 기준으로 작성하고 그 이후는 가용회원보유액 기준(총회원보유액에서 국내은행의 해외점포 예치금 등을 제외한 금액)으로 작성하였으나 e-나라 지표의 경우 통계 기준의 일관성 유지를 위해 1999년 8월이전이라도 현재의 기준인 가용회원보유액 기준으로 작성하여 게시함에 따른 차이일 뿐으로서 연구 목적이나 필요한 통계 사용시에 따라 취사 선택이 필요
- * 외환은행의 시계열을 제외한 보조지표는 연도별 기준으로 편제함

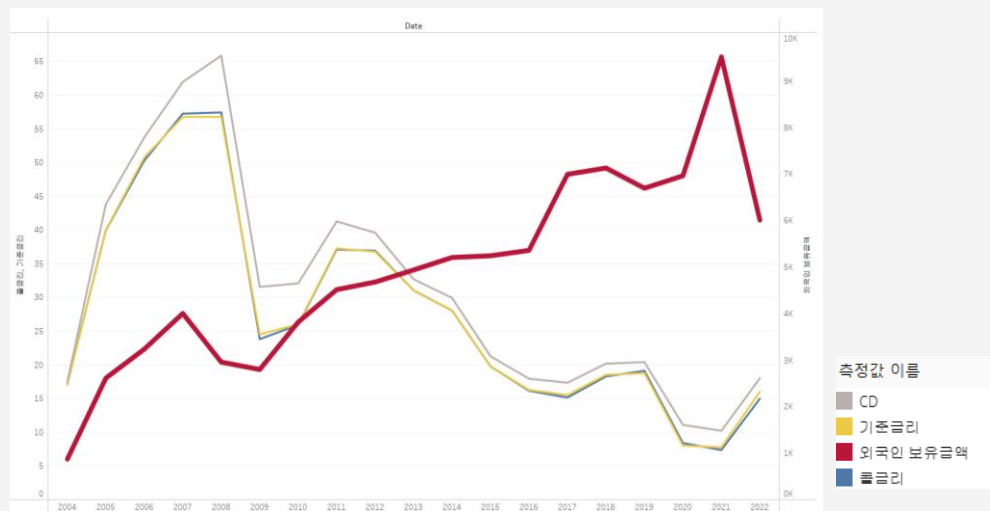
(외환보유액)과 외국인보유금액



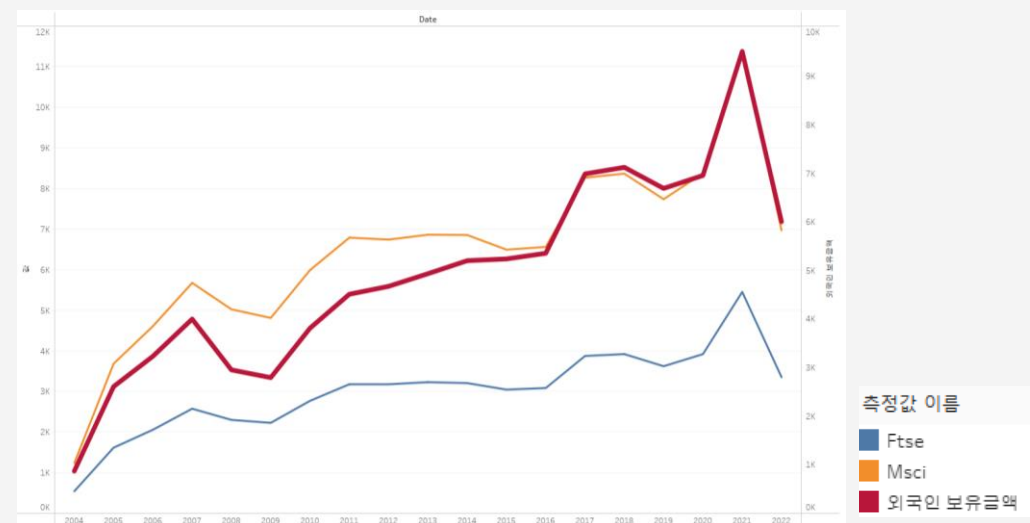
(국고채 10년,3년,5년 /회사채 3년)과 외국인보유금액



(콜금리, 기준금리, CD금리)와 외국인 보유금액

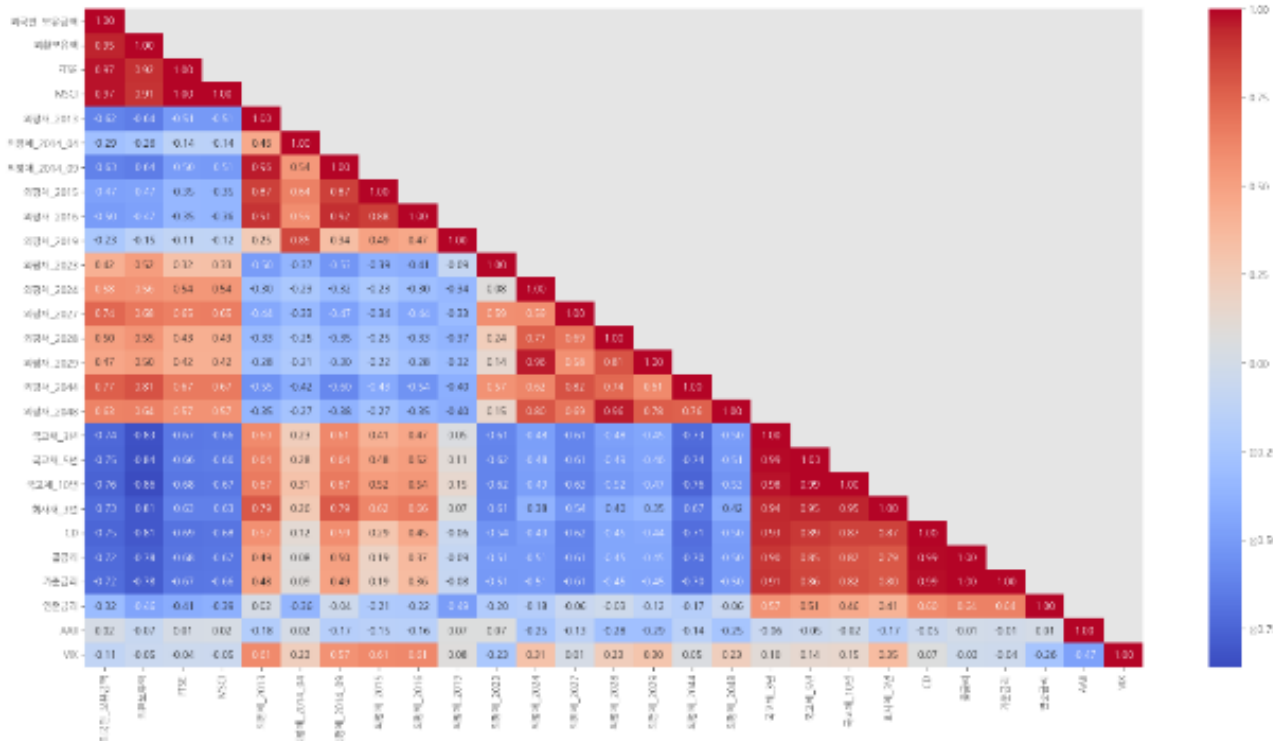


(MSCI, FTSE) 와 외국인 보유금액



다중 공선성 파악 1. Correlation matrix Heatmap으로 시각화

Correlation matrix Heatmap



- ☑ 변수 간 상관관계를 **파란색**~**빨간색**으로 표현
 - ➡ 색이 진할 수록, 두 변수 간의 상관관계가 높다
- ☑ VIX와 AII 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
물금리	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

정상성 검정 결과

ADF Test

✓ P-value < 0.05 인 경우, 정상성 만족한다고 판단

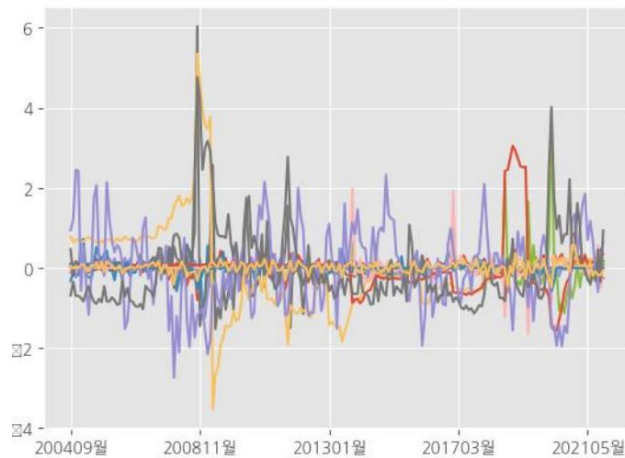
KPSS

✓ P-value > 0.05 인 경우, 정상성 만족한다고 판단

3개를 제외하고는 정상성 충족 못함

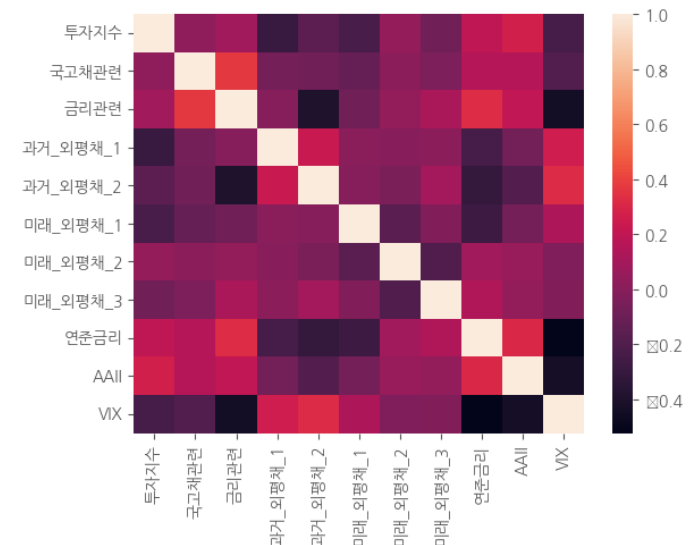
➡ PCA 기법 & 차분(Differencing)기법 사용

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! ✓

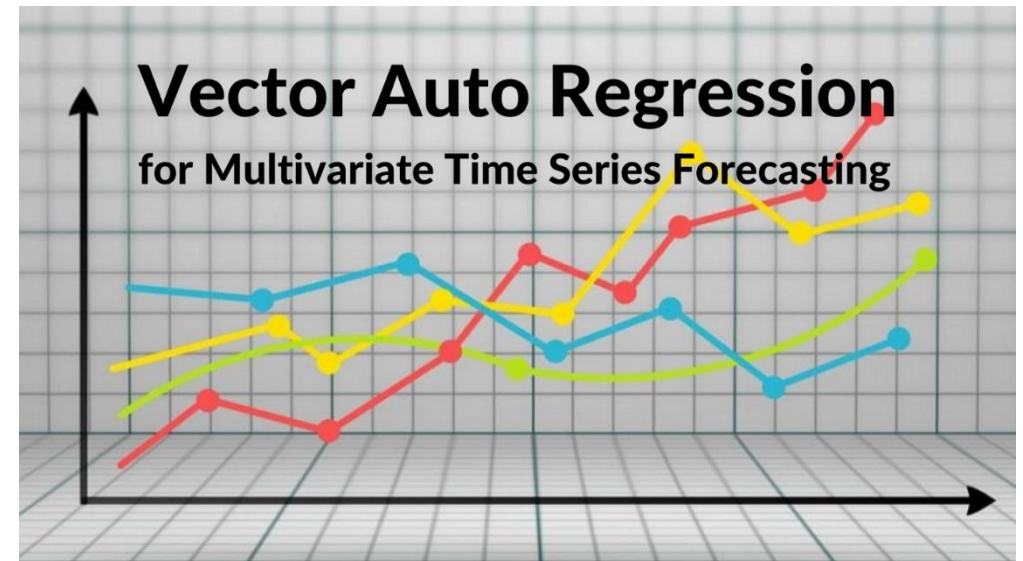
변수명	VIF	변수명	VIF
VIX	1.84	과거 외평채 1	1.257
금리관련	1.67	국고채 관련	1.21
연준금리	1.63	미래 외평채 1	1.20
과거 외평채 2	1.38	미래 외평채 2	1.15
AAII	1.33	미래 외평채 3	1.09
투자지수	1.258		



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

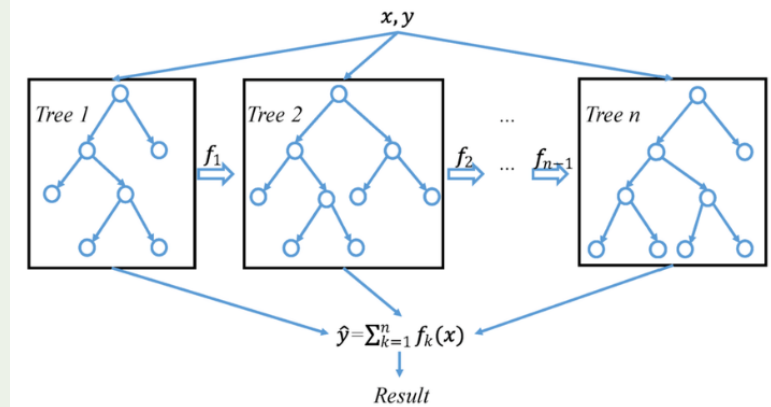
Tabular Data를 다루는 업무에서 주로 사용

✓ Tree 기반의 모델

장점) 데이터의 주어진 범위 내에서의 예측 능력 ↑

단점) 데이터가 가지고 있는 범위 외의 예측을 해야 하는 경우, 예측 성능 ↓

✓ 정상성을 띄도록 데이터를 가공해주어야 한다는 주의점



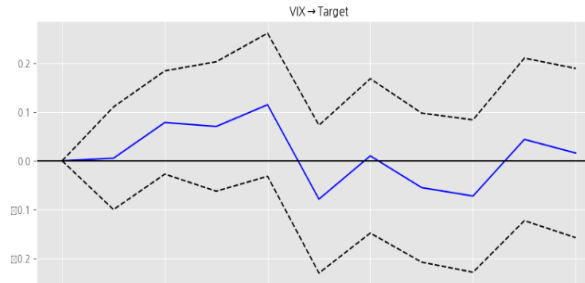
$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t$$

지연계수(Lag Order) 결정

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

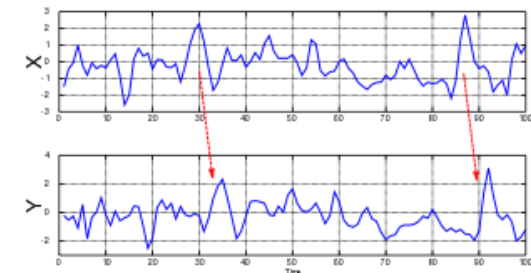
Impulse Response Analysis

- 특정 변수의 충격(Impulse)이 다른 변수에 미치는 영향을 분석하기 위함
- Target 변수에 큰 영향을 준 변수를 파악하는 데에 사용



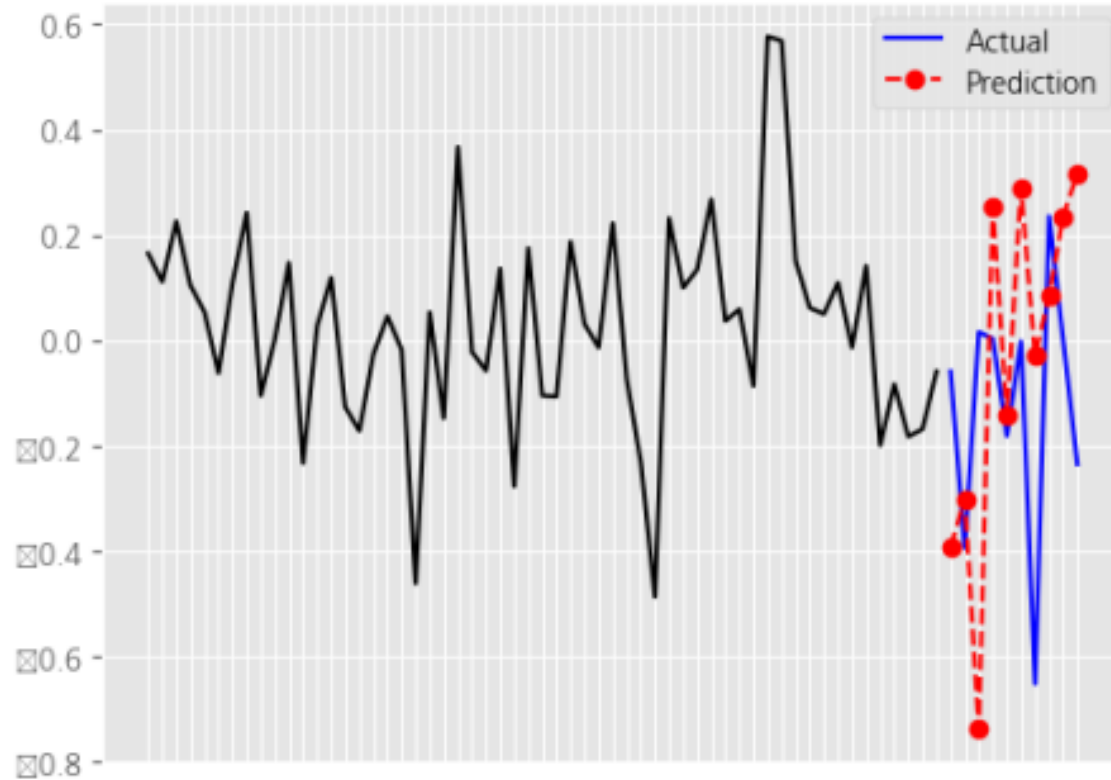
Granger Causality

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



VAR모델의 예측 결과

Prediction of VAR Model

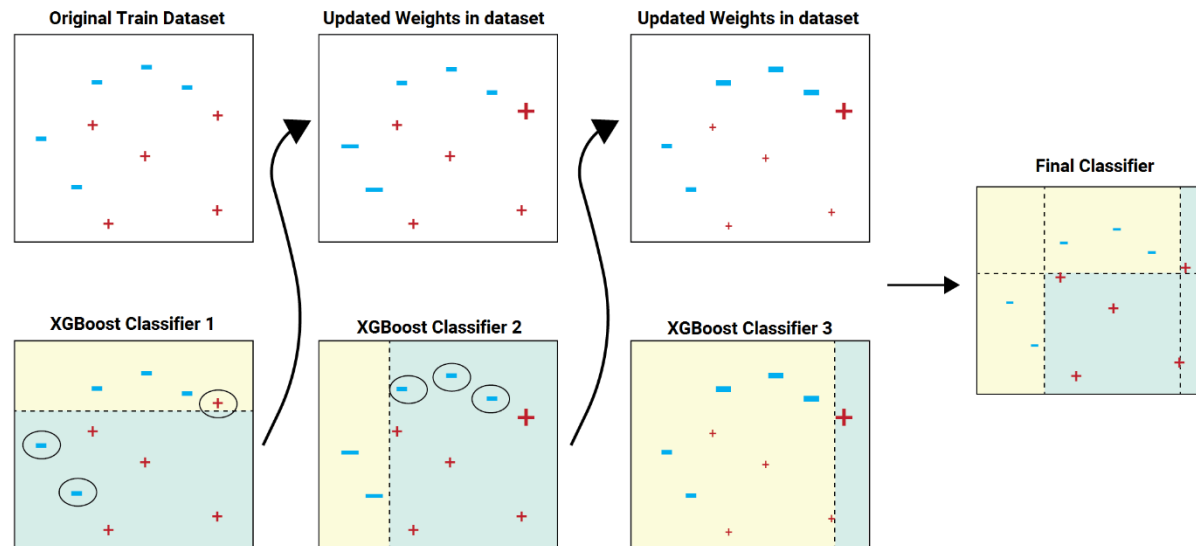


MSE : 0.161



예측을 함에 있어 변화의 방향성과
그 정도를 일부 맞춘 것으로 보여짐

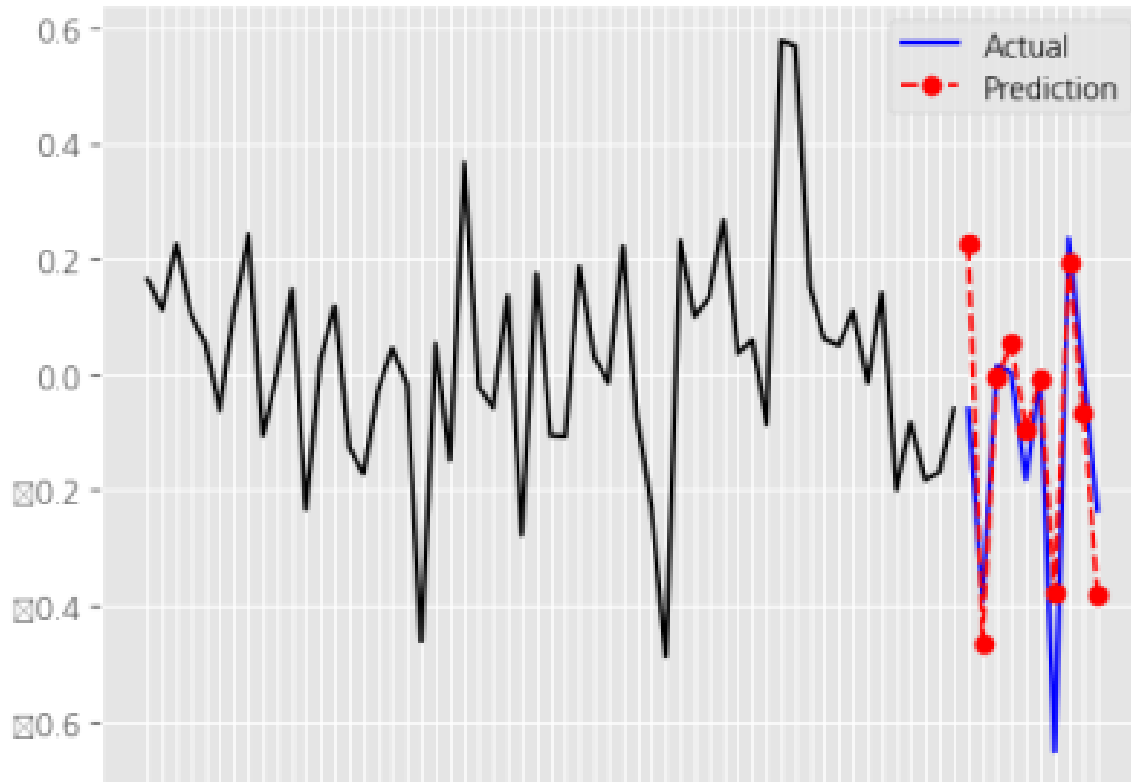
**But, '예측을 잘 한다'라고
판단하기엔 무리가 있음.**



- 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 파악 가능

XGBoost 모델의 예측결과

Prediction of XGBoost Model

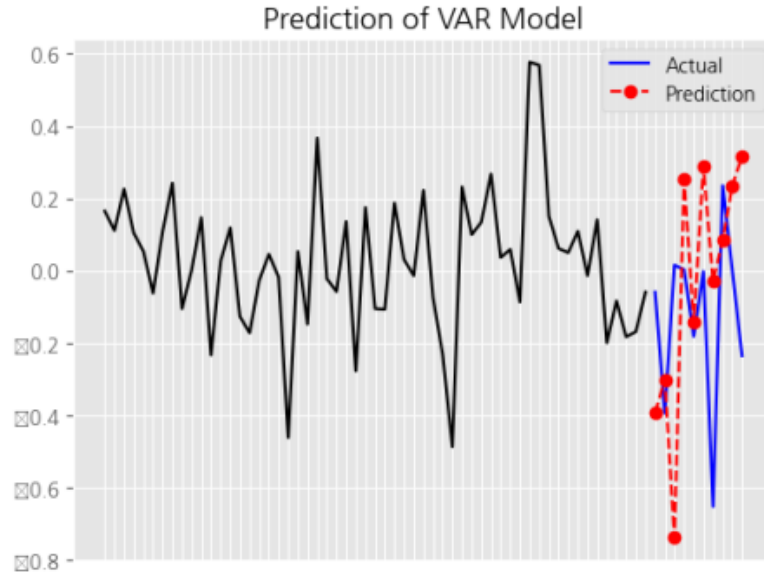


MSE : 0.020



예측의 방향과 그 값을
VAR에 비해 잘 잡아내는 모습

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

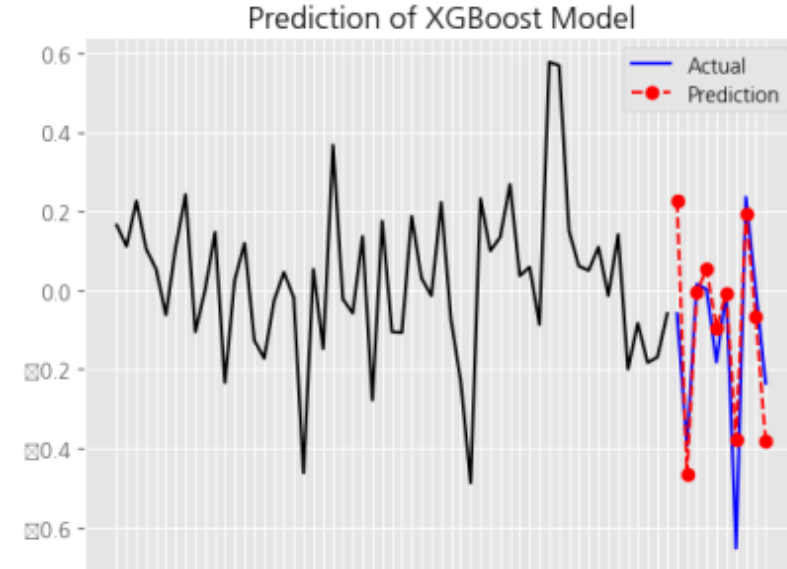


미래 외평채, 금리관련, AAI

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

이론적, 논리적 기반이 마련 되어 있는 결론

≠

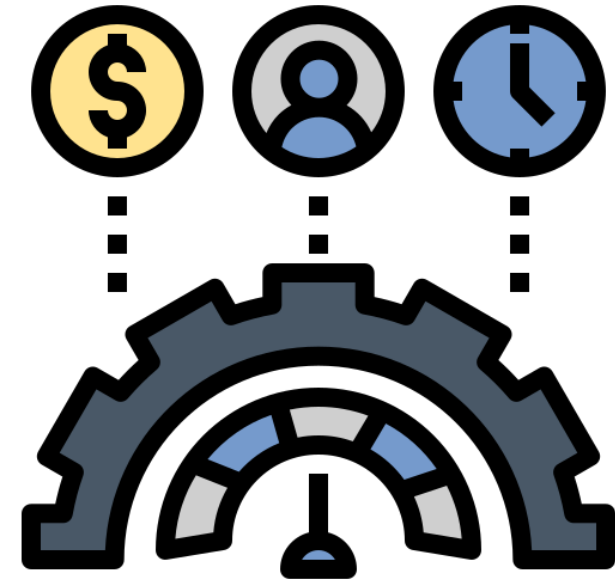
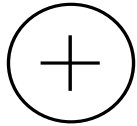
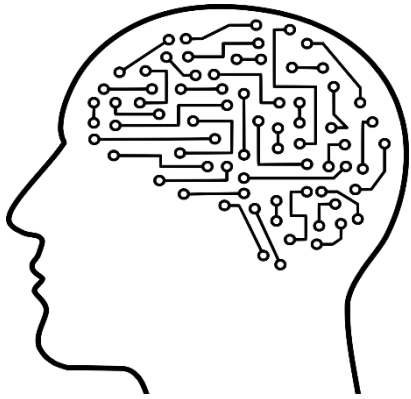


투자지수, 과거 외평채

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

예측은 훌륭하지만 결과론적인 결론

Conclusion



Better Efficiency!





**Thank
you!**