



주린이 주목! 지금 팔아?! 말아?!

# 주식 초보자의 리스크 관리를 위한 매수/매도 추천 서비스

# INDEX

---

**1** 1주차 요약

**2** 모델링 과정

**3** 최종 모델

**4** 결론

# 1

1주차 요약



### 주식 시장의 상황

주식 시장이 급격히 안 좋아지며  
많은 사람이 투자 손해를 보게 됨



### 주식 초보자들의 상황

투자 정보를 알지 못하는  
주식 초보자들이 큰 손해를 봄



주식 초보자도 쉽게 참고할 수 있는  
투자 인사이트를 제공하기 위한 서비스를 만들자!



### 주식 시장의 상황

주식 시장이 급격히 안 좋아지며  
많은 사람이 투자 손해를 보게 됨



### 주식 초보자들의 상황

투자 정보를 알지 못하는  
주식 초보자들이 큰 손해를 봄

### 시계열자료분석팀 주제분석

주식 초보자의 리스크 관리를 위한 매수/매도 추천 서비스



SK 하이닉스



신한금융지주회사

신한지주



HYUNDAI

현대 자동차

8개의 개별 지표

11개의 공통 지표

### 개별 지표

- ✓ 주식 시세 추이 데이터
- ✓ 투자자별 거래 실적 데이터
- ✓ 외국인 보유량 데이터
- ✓ 공매도 데이터
- ✓ 국내 기사 데이터
- ✓ 영문 기사 데이터
- ✓ 네이버 증권 토론방 데이터
- ✓ 네이버 검색량 데이터

### 공통 지표

- ✓ 코스피 데이터
- ✓ 비트코인 거래 데이터
- ✓ 경제 심리 지수
- ✓ 뉴스 심리 지수
- ✓ 산업 생산 지수
- ✓ 소비자 물가 지수
- ✓ 소비자 신뢰 지수
- ✓ 소비자 심리 지수
- ✓ 실업률
- ✓ 한국은행 기준금리
- ✓ 환율

### 개별 지표

- ✓ 주식 시세 추이 데이터
- ✓ 투자자별 거래 실적 데이터
- ✓ 외국인 보유량 데이터
- ✓ 공매도 데이터
- ✓ 국내 기사 데이터
- ✓ 영문 기사 데이터
- ✓ 네이버 증권 토론방 데이터
- ✓ 네이버 검색량 데이터

### 공통 지표

- ✓ 코스피 데이터
- ✓ 비트코인 거래 데이터
- ✓ 경제 심리 지수
- ✓ 농산물 가격 지수
- ✓ 산업 생산 지수
- ✓ 소비자 물가 지수
- ✓ 소비자 신뢰 지수
- ✓ 소비자 심리 지수
- ✓ 실업률

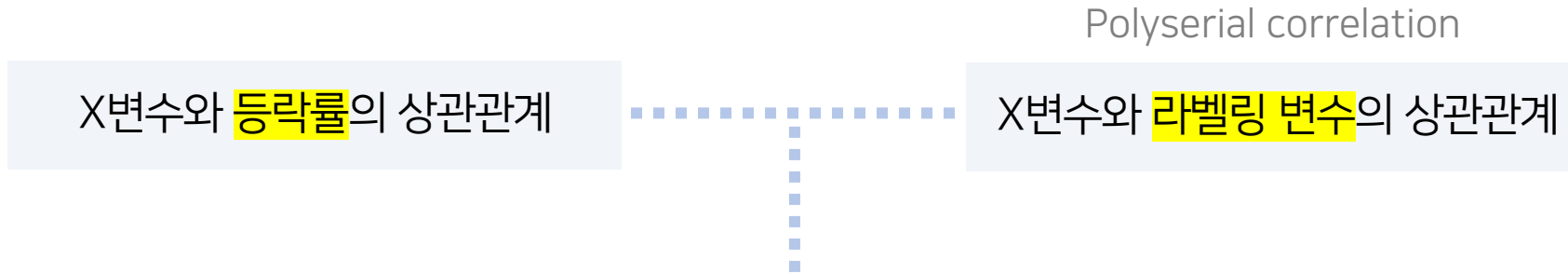


여론 및 투자자 심리에 따른 등락을 파악하기 위해

여론을 반영할 수 있는 데이터 활용 !







두 상관관계가 비슷해지는 지점을 기준으로 라벨링 !



1일 등락률: 3% 기준 라벨링

3일 등락률: 5% 기준 라벨링

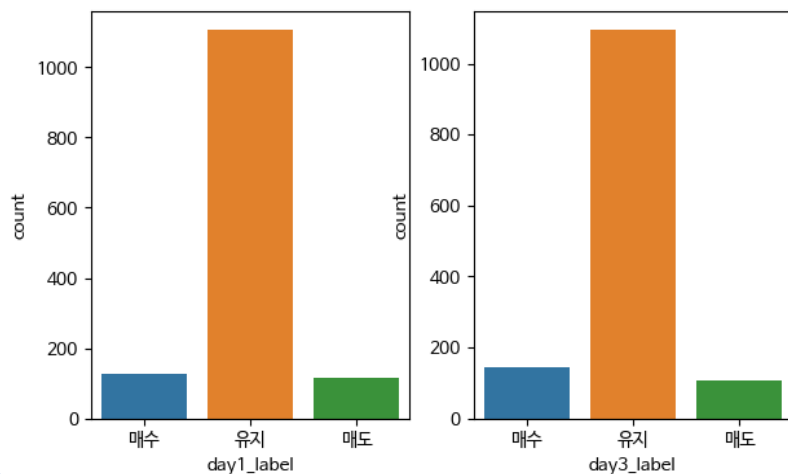
1일치(내일) 등락률  $\leq -3\%$  : 매도 (sell)

1일치(내일) 등락률  $\geq 3\%$  : 매수 (buy)

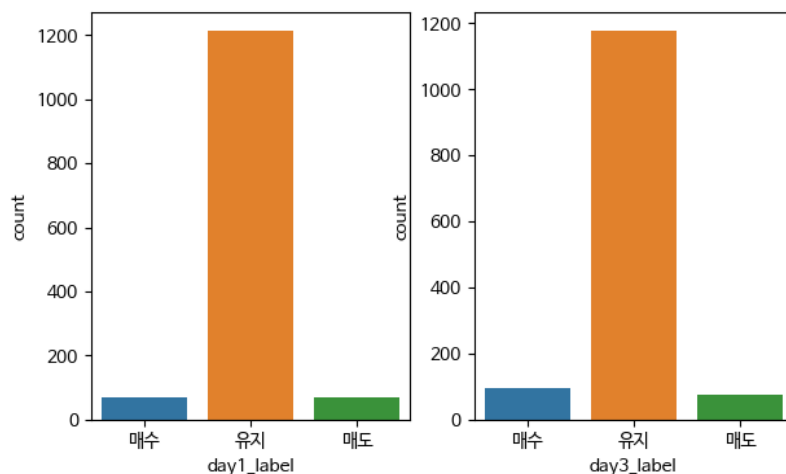
Otherwise : 유지 (maintain)

## 종목별 매도 / 유지 / 매수 라벨링 결과

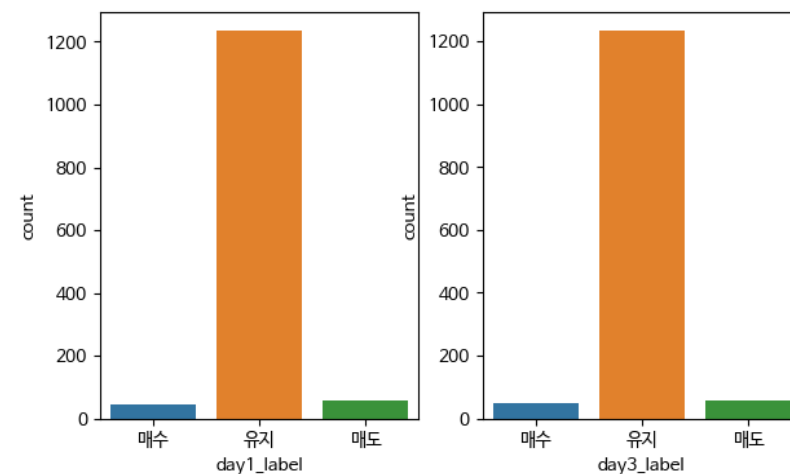
SK 하이닉스 라벨 분포



현대차 라벨 분포



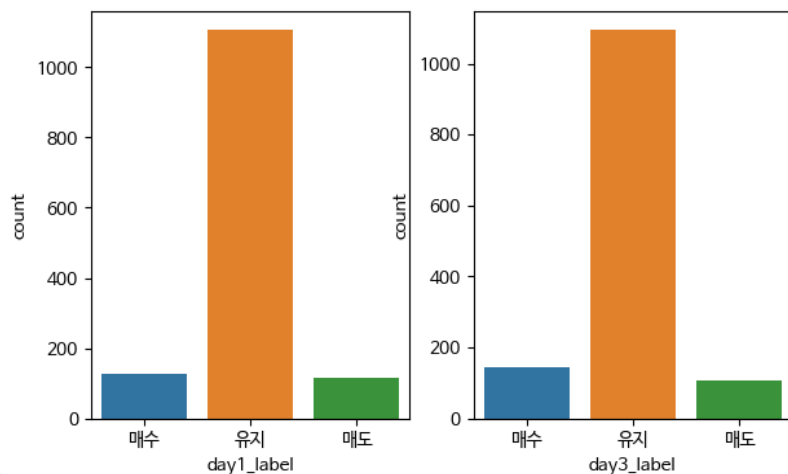
신한지주 라벨 분포



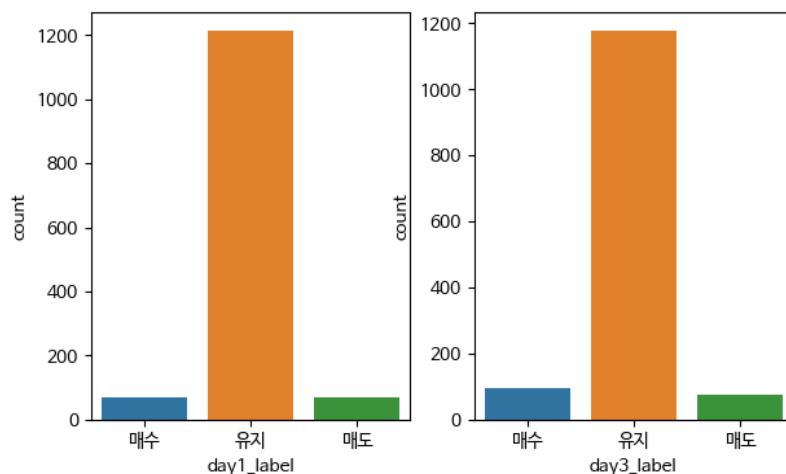
모든 종목에서 '유지' 대비  
'매수' / '매도'의 클래스 불균형 심각

## 종목별 매도 / 유지 / 매수 라벨링 결과

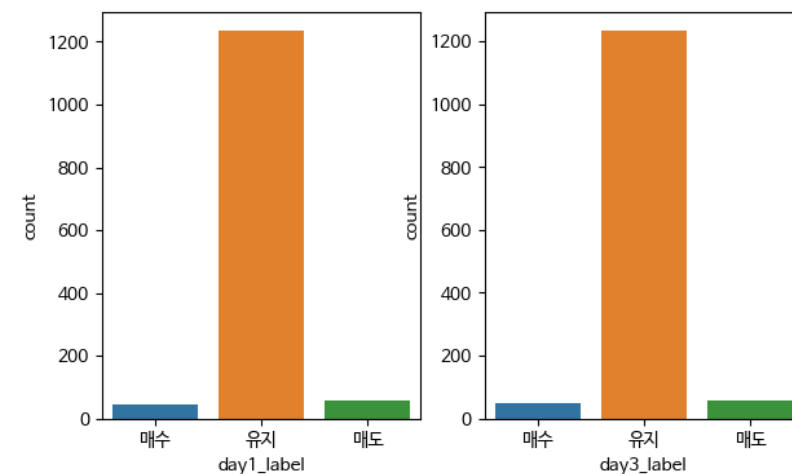
SK 하이닉스 라벨 분포



현대차 라벨 분포



신한지주 라벨 분포

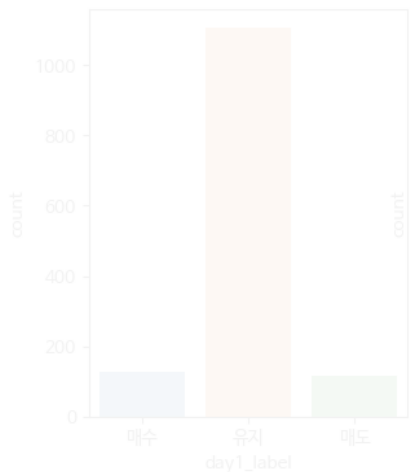


전체 정확도는 높지만, 모델 예측이 전체적으로 '유지'에 편향되어

'매수'와 '매도'를 제대로 예측하지 못하는 문제 발생

## 🚨 시계열 데이터 증강의 어려움 🚨

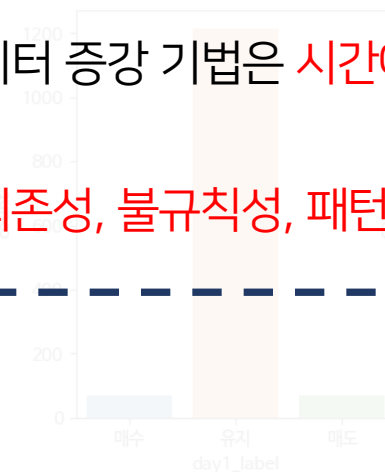
SK 하이닉스 라벨 분포



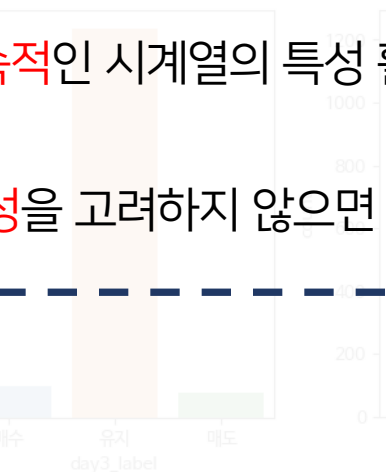
현대차 라벨 분포



현대차 라벨 분포



현대차 라벨 분포

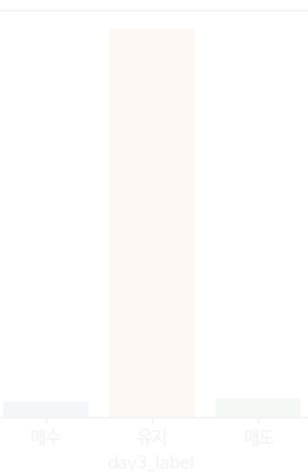


현대차 라벨 분포

신한지주 라벨 분포



현대차 라벨 분포



기존 데이터 증강 기법은 **시간에 종속적**인 시계열의 특성 활용 X



**시간적 의존성, 불규칙성, 패턴 복잡성**을 고려하지 않으면 과적합 발생



전체 정확도는 높지만, 모델 예측이 전체적으로 '유지'에 편향되어

**매도와 매수를 제대로 예측하지 못하는 문제 발생**



# 2

## 모델링 과정

# 2 모델링 과정

5일 라벨링



SK 하이닉스



신한금융지주회사

신한지주



현대자동차

모델링 결과 1일 등락률 예측에 대해서는  
성능이 떨어지는 문제 발생

딥러닝 팀장 ???



# 2 모델링 과정

5일 라벨링

## 1일 등락률 예측 결과

LGBM

	매수	유지	매도
매수	0	22	0
유지	0	201	0
매도	0	15	0

XGB

	매수	유지	매도
매수	0	22	0
유지	9	186	6
매도	1	14	0

Logistic

	매수	유지	매도
매수	1	2	19
유지	4	38	159
매도	0	2	13

1일 등락률 예측에 대해서 모델링 성능이 좋지 않았음  
(딥러닝 모델들도 별반 다르지 않았음)



# 2 모델링 과정

5일 라벨링

3일 등락률 예측 결과

LGBM

	매수	유지	매도
매수	3	24	0
유지	3	186	3
매도	0	15	4

XGB

	매수	유지	매도
매수	11	16	0
유지	5	186	1
매도	0	15	4

Logistic

	매수	유지	매도
매수	8	24	0
유지	3	222	1
매도	0	26	1

3일 등락률은 어느 정도 맞추는 것을 확인

# 2 모델링 과정

5일 라벨링



3일 등락률에 대한 예측 성능보다

1일 등락률에 대한 예측 성능이 부진한 이유는...?

LGBM

XGB

Logistic

	매수	매도	추정 원인
매수	3	0	3일 등락률에서는 데이터간 종속성이 더 커지지만 1일 등락률은 그에 비해 X변수에 덜 의존하고 랜덤에 가까운 특성을 갖기 때문이 아닐까...
유지	3	0	
매도	0	0	추정 원인

	매수	유지	매도
매수	11	10	0
유지	5	186	1
매도	0	15	4

	매수	유지	매도
매수	8	24	0
유지	3	222	1
매도	0	26	1

1일 등락률이 3일 등락률보다 클래스 불균형이 심하기 때문에  
클래스 불균형으로 인해 성능이 덜 나온 게 아닐까...

3일 등락률은 어느 정도 맞추는 것을 확인



# 2 모델링 과정

5일 라벨링



3일 등락률 예측 결과

3일 등락률에서의 성능이 1일 등락률보다 낮다면

더 긴 기간에서의 등락률을 예측해보는 것도 유의미하지 않을까?

LGBM

	매수	유지	매도
매수	3	24	0
유지	3	186	3
매도	0	15	4

XGB

	매수	유지	매도
매수	11	16	0
유지	5	186	1
매도	0	15	4

Logistic

	매수	유지	매도
매수	8	24	0
유지	3	222	1
매도	0	26	1

5일 등락률도 시도해보기로 결정



넌 꼭 살아  
남아야 해!!  
제발.....

3일 등락률은 어느 정도 맞추는 것을 확인

# 2 모델링 과정

## 5일 라벨링

```
SK['day5_label'] = SK['5일 등락률'].apply(lambda x: 'maintain' if abs(x) < 5 else 'buy' if x >= 5 else 'sell')
```

2주차와 마찬가지로 polyserial 상관계수 구해  
등락률과의 상관계수와 비교하여  
라벨링 threshold 결정



3일 등락률과 마찬가지로 5%로 결정

# 2 모델링 과정

## 변수선택

1

[ 인과관계 검정]

2

[ VIF]

3

[ Feature Importance]

4

[ KS 검정 ]

5

[ Full Model]

## 시도1. 인과관계 검정 &amp; 상관분석

## 인과관계 검정

동일한 시간축의 범위를 가진 두 데이터가 있을 때  
한 데이터를 다른 데이터로 선형회귀를 할 수 있고 유의하다면 그래인저 인과관계 존재

## 시도1. 인과관계 검정 &amp; 상관분석

## 인과관계 검정

동일한 시간축의 범위를 가진 두 데이터가 있을 때  
한 데이터를 다른 데이터로 선형회귀를 할 수 있고 유의하다면 그래인저 인과관계 존재



1주차 상관분석과 인과관계 검정을 토대로 변수 선택 진행

## 시도1. 인과관계 검정 &amp; 상관분석

## 인과관계 검정

동일한 시간축의 범위를 가진 두 데이터가 있을 때  
한 데이터를 다른 데이터로 선형회귀를 할 수 있고 유의하다면 그래인저 인과관계 존재



변수 선택 결과
등락률, 종가, 거래량, 거래대금, 시가총액, 순매수기관, 순매수 개인, 순매수 외국인, 기사감성점수



## 시도2. VIF

## Variance Inflation Factor

분산 팽창 인수, 다중회귀분석에서 독립변수가

다중 공선성의 문제를 갖고 있는지 판단하는 기준으로 보통 10보다 크면 다중공선성이 있다고 판단

$$\begin{aligned} VIF_i > 10 &\Leftrightarrow \frac{1}{1 - r_i} > 10 \\ 1 &> 10 - 10r_i \\ r_i &> 0.9 \end{aligned}$$



i 번째 독립변수가 빠져도  
나머지 변수들이 반응 변수를  
90%이상 설명한다는 의미 !

## 시도2. VIF

## Variance Inflation Factor

분산 팽창 인수, 다중회귀분석에서 독립변수가  
다중 공선성의 문제를 갖고 있는지 판단하는 기준으로 보통 10보다 크면 다중공선성이 있다고 판단



수집한 전체 변수에서 VIF기준 10 이상인 변수를 순서대로 제거

## 시도2. VIF

## Variance Inflation Factor

분산 팽창 인수, 다중회귀분석에서 독립변수가  
다중 공선성의 문제를 갖고 있는지 판단하는 기준으로 보통 10보다 크면 다중공선성이 있다고 판단



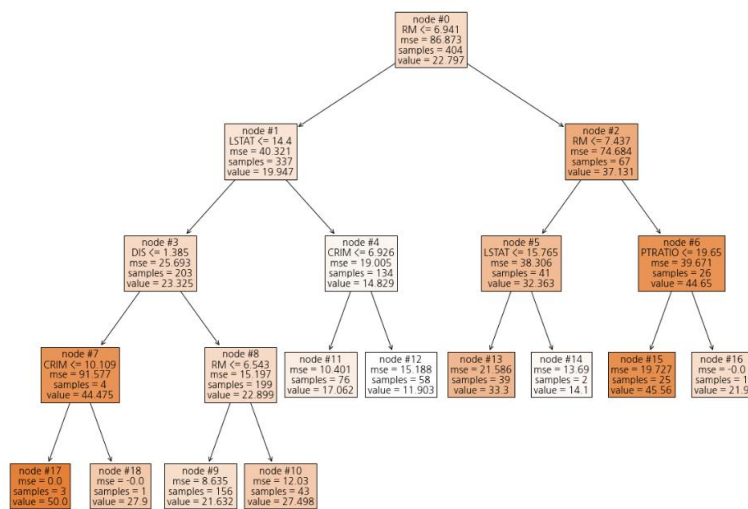
## 변수 선택 결과

경제심리지수(순환변동치),시가총액,비트코인종가,원/미국달러,소비자심리지수,코스피거래대금,  
순매수 개인,순매수 외국인, 산업생산지수, 코스피거래량, 뉴스심리지수, 원/유로, 토론방, 실업률,  
원/일본엔(100엔), 거래량, 기사감성점수, 외국인 보유수량, 코스피등락률, 경제활동참가율,  
검색어, 보도량, 순매수 기관, 비트코인거래량, 비트코인변동

## 시도3. Feature Importance

feature importance

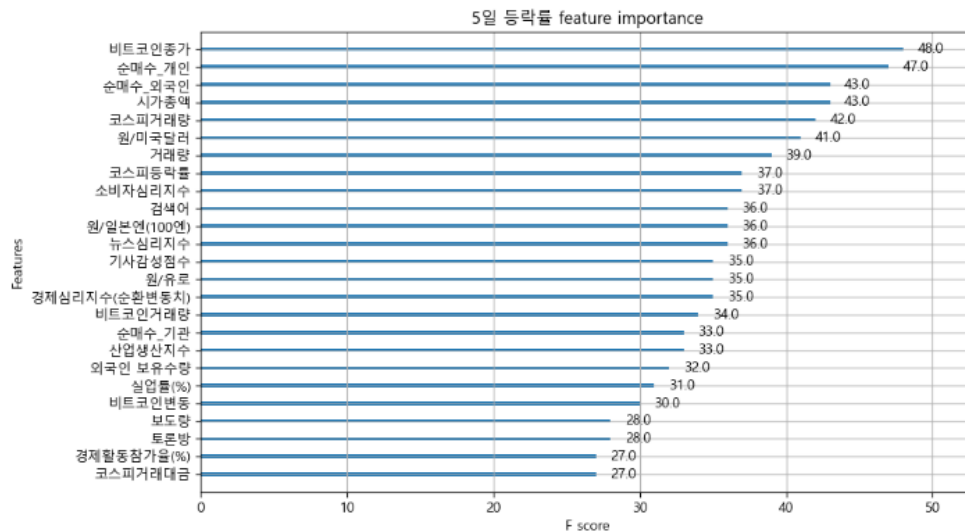
트리 기반 모델에서 분류를 진행할 때, 변수가 분기에 얼마나 자주 중요하게 사용되는지에 대한 값으로 분기에 쓰인 변수의 랭킹이라 볼 수 있음



## 시도3. Feature Importance

## feature importance

트리 기반 모델에서 분류를 진행할 때, 변수가 분기에 얼마나 자주 중요하게 사용되는지에 대한 값으로 분기에 쓰인 변수의 랭킹이라 볼 수 있음



feature importance에서 유의미한 차이가 없어

변수 선택 진행 X

## 시도3. Feature Importance

## feature importance

트리 기반 모델에서 분류를 진행할 때, 변수가 분기에 얼마나 자주 중요하게 사용되는지에 대한 값으로 분기에 쓰인 변수의 랭킹이라 볼 수 있음



다중공선성 사전에 제거 후 진행!

## 변수 선택 결과

순매수\_개인, 순매수\_외국인, 순매수\_기타법인, 거래량, 비트코인변동, 토론방, 뉴스심리지수, 순매수\_기관, 보도량, 기사감성점수, 원/위안, 검색어, 코스피거래량, 등락률, 외국인 보유수량, 거래대금, 대비, 비트코인종가, 원/유로, 종가, 비트코인거래량, 코스피등락률, 원/미국달러, 원/일본엔, 코스피대비, 코스피거래대금

## 시도4. KS test

## Kolmogorov Smironov Test

비모수적 방법으로 비교하고자 하는 두 분포의  
경험적분포(Empirical distribution)의 차이를 특정 기준과 비교하여 기각 여부를 결정

신용평가모형에서 우/불량 고객 선별을 위해 자주 사용하는 기법 !

신용평가모형의 목표가 본 주제분석의 모델링 목표와 유사하고,  
비모수적 접근법이니 선형 관계만을 활용한 변수 선택법의 한계를 극복할 수 있지 않을까?!





## 시도4. KS test

## 경험적 분포 함수란?

Kolmogorov Smirnov Test

$$F_n(x) = \frac{\sum_{i=1}^n 1(X_i \leq x)}{n} \quad (\text{단, } n \text{은 observation의 총 개수})$$

경험적분포(Empirical distribution)의 차이를 특정 기준과 비교하여 기각의 여부를 결정

모집단  $X$ 의 CDF의 정의는  $F_n(x) = P(X \leq x)$ 인데, 상당히 유사

$$P(\lim_{n \rightarrow \infty} |F(x) - F_n(x)| = 0) = 1$$

글리벤코-칸텔리 정리에 따르면 EDF는 CDF에 확률 수렴

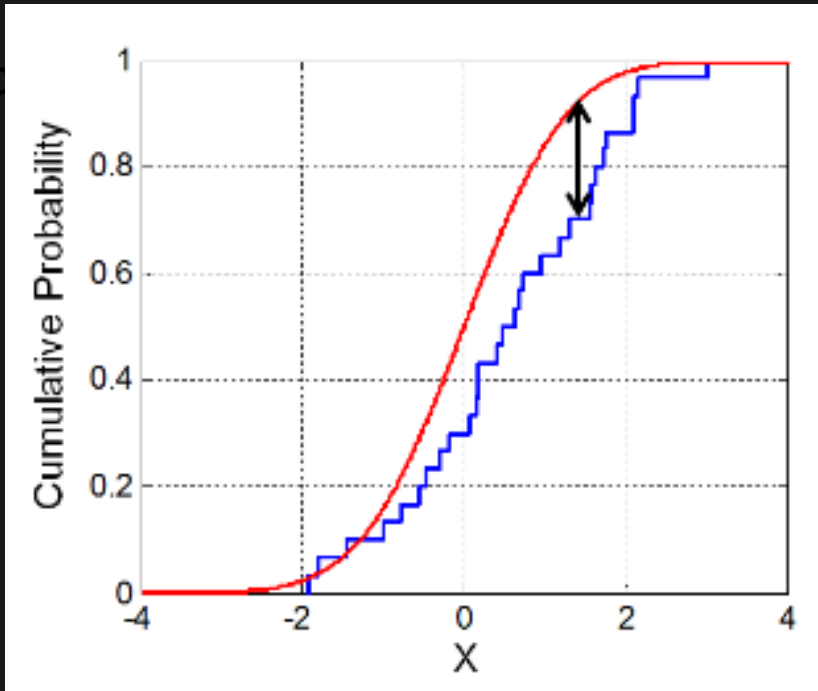
즉, EDF는 모집단의 CDF에 대응되는 통계량을 정의하기 위해 사용



# 2 모델링 과정

변수선택

시도4. KS test



그래프에는 하나의 CDF와 EDF가 보이고

그 사이에 있는 검정색 화살표가 KS 통계량을 나타냄

법으로 비교하고자 하는 두 분포의

on)의 차이를 특정 기  $D_{n,m} = \sup(|F_{1,n} - F_{2,m}(x)|$

$$D_{n,m} > c(\alpha) \sqrt{\frac{n+m}{nm}}$$

분포가 다르게 나타나는 변수만을 선택한다

력 향상에 도움이 되지 않을까?

이 식으로 계산하여 검정 통계량 값이 기준치 보다

높으면, 귀무가설인 “같은 모분포 출신”을 기각!

## 시도4. KS test

## Kolmogorov Smironov Test

비모수적 방법으로 비교하고자 하는 두 분포의

경험적분포(Empirical distribution)의 차이를 특정 기준과 비교하여 기각 여부를 결정



3가지 종목 모두에서 매도/유지, 유지/매수, 매도/매수의 분포가 다 유의하게 달랐던 변수의 개수 < 9가지의 분포 조합 중, 6번 이상 유의하게 다른 분포를 갖는 변수 선택

## 시도4. KS test

## Kolmogorov Smironov Test

비모수적 방법으로 비교하고자 하는 두 분포의

경험적분포(Empirical distribution)의 차이를 특정 기준과 비교하여 기각 여부를 결정



## 변수 선택 결과

종가, 등락률, 거래량, 거래대금, 시가총액, 토론방, 순매수기관, 순매수기타법인,  
순매수개인, 순매수외국인, 검색어, 보도량, 기사감성점수, 코스피등락률

# 2 모델링 과정

## 변수선택

1

[ 인과관계 검정 ]

2

[ VIF ]

3

[ Feature Importance ]

4

[ KS 검정 ]

5

[ Full Model ]

But...

대부분의 경우에선

Full model의  
성능이 가장 좋았다...



# 2 모델링 과정

변수선택

1

[ Full model ]

2



[ 교과서 ]

3

[ Feature Importance ]

4

변수선택 조합과 함께 모델링 가보자고!

[ VIF ]

[ KS 검정 ]

## Input

현재 시점의 X변수 (minmax scaled / full model)

## Output

앞으로 5일 이내의 주가 등락 예측에 따른 현재 시점에서의 매수/유지/매도 추천

- t+1 시점부터 t+6 시점까지, 5일 동안 등락률이 5%이상 상승할 것으로 예상되면 매수 추천
- t+1 시점부터 t+6 시점까지, 5일 동안 등락률이 5%이상 하락할 것으로 예상되면 매도 추천
- t+1 시점부터 t+6 시점까지, 5일 동안 등락률의 절댓값이 5% 이내로 예상되면 유지 추천

### Input

현재 시점의 X변수 (minmax scaled / full model)

### Output

앞으로 5일 이내의 주가 등락 예측에 따른 현재 시점에서의 매수/유지/매도 추천

- t+1 시점부터 t+6 시점까지, 5일 동안 등락률이 5%이상 상승할 것으로 예상되면 매수 추천
- t+1 시점부터 t+6 시점까지, 5일 동안 등락률이 5%이상 하락할 것으로 예상되면 매도 추천
- t+1 시점부터 t+6 시점까지, 5일 동안 등락률의 절댓값이 5% 이내로 예상되면 유지 추천



클래스 불균형이 가장 덜한 SK하이닉스 데이터로 모델 적합 후, 동일한 모델을 3종목 모두에 적용



모델 선정 기준: 매수/매도를 얼마나 잘 맞추는가? (매수/매도 정확도 높은 것 중 f1-score 높은 모델)

Confusion matrix

Confusion matrix

PREDICTIVE VALUES

POSITIVE (1)

NEGATIVE (0)

ACTUAL VALUES

POSITIVE (1)

TP

FN

NEGATIVE (0)

FP

TN

### 정확도 (Accuracy)

전체 예측된 결과 중  
올바르게 예측한 결과의 비율  
 $(TP+TN)/(TP+TN+FP+FN)$

### 정밀도 (Precision)

모델이 양성(Positive)로 예측한 결과 중  
실제로 양성인 결과의 비율  
 $TP/(TP+FP)$



Confusion matrix

		PREDICTIVE VALUES	
		POSITIVE (1)	NEGATIVE (0)
ACTUAL VALUES	POSITIVE (1)	TP	FN
	NEGATIVE (0)	FP	TN

## 재현율 (Recall)

실제로 양성인 샘플 중  
모델이 양성으로 예측한 샘플의 비율  
 $TP/(TP+FN)$

## F1 score

정밀도와 재현율의 조화 평균  
 $2(Precision*Recall)/(Precision+Recall)$

### Optuna

하이퍼파라미터 공간을 탐색하여 목적 함수값을 최소화 or 최대화 하는 **파라미터 조합을 찾는 과정**

### Optuna

하이퍼파라미터 공간을 탐색하여 목적 함수값을 최소화 or 최대화 하는 **파라미터 조합을 찾는 과정**

### Optuna score 커스텀

최적화할 **목적 함수**의 평가지표를 사용자가 직접 정의하고 제공하는 것



Optuna가 각 하이퍼파라미터 조합에 대해 결과를 반환하는 함수

## Optuna

하이퍼파라미터 공간을 탐색하여 목적 함수값을 최소화 or 최대화 하는 **파라미터 조합을 찾는 과정**

## Optuna score 커스텀

최적화할 목적 함수의 평가지표를 사용자가 직접 정의하고 제공하는 것



F1 score + 3가지의 커스텀을 통해 optuna 진행

### 01 매수, 매도, 유지 정확도의 평균

```
cm=confusion_matrix(y_test, y_pred)
bacc=cm[0,0]/sum(cm[0]) # 매수 정확도
sacc=cm[2,2]/sum(cm[2]) # 매도 정확도
macc=cm[1,1]/sum(cm[1]) # 유지 정확도
rst=np.mean([bacc,sacc,macc])
```

### 02 매수, 매도 정확도의 평균

```
cm=confusion_matrix(y_test, y_pred)
bacc=cm[0,0]/sum(cm[0]) # 매수 정확도
sacc=cm[2,2]/sum(cm[2]) # 매도 정확도
rst=np.mean([bacc,sacc])
```

### 03 매수, 매도 정확도, F1 Score의 평균

```
cm=confusion_matrix(y_test, y_pred)
bacc=cm[0,0]/sum(cm[0]) # 매수 정확도
sacc=cm[2,2]/sum(cm[2]) # 매도 정확도
f1=sum(scores)/len(scores)
rst=np.mean([bacc,sacc,f1])
```

### 04 매수, 매도의 정확도, 정밀도의 평균

```
cm=confusion_matrix(y_test, y_pred)
bacc=cm[0,0]/sum(cm[0]) # 매수 정확도
sacc=cm[2,2]/sum(cm[2]) # 매도 정확도
bpre=cm[0,0]/np.sum(cm, axis=0)[0] #매수 정밀도
spre= cm[2,2]/np.sum(cm, axis=0)[2] #매도 정밀도
rst=np.mean([bacc,sacc,bpre,spre])
```

### 모델 종류

- LSTM
- CNN
- SVM
- Logistic regression
- 나이브 베이즈
- XGB
- LGBM
- LGBM 회귀
- LGBM-CNN 회귀

## 모델 종류

- LSTM
- CNN
- SVM
- Logistic regression
- 나이브 베이즈
- XGB
- LGBM
- LSTM 회귀
- LSTM-CNN

할일을 눈앞에 산더미처럼 쌓아 놓으니  
더 하기 싫어진다



SK 하이닉스 기준 4가지 score로  
optuna 진행해 최적의 모델 선택

# 3

최종 모델



# 3 최종 모델

## 최종 모델 소개

### XGB classifier

데이터 : SK 하이닉스

변수 : Full Model

평가지표 : custom한 optuna score

분류 : 다중 분류(매수, 매도, 유지)

### LSTM regressor

데이터 : SK 하이닉스

변수 : VIF로 선별된 변수

특징 : 라벨 인코딩 한 결과를

회귀로 예측하고

Threshold 결정 함수로 분류

# 3 최종 모델

## 최종 모델 소개

### XGB classifier

데이터 : SK 하이닉스

변수 : Full Model

평가지표 : custom한 optuna score

분류 : 다중 분류(매수, 매도, 유지)

### LSTM regressor

데이터 : SK 하이닉스

변수 : VIF로 선별된 변수

특징 : 라벨 인코딩 한 결과를

회귀로 예측하고

Threshold 결정 함수로 분류

## 1. 변수 선택

데이터 : SK 하이닉스

변수 : Full Model

### 변수

X : '종가', '대비', '등락률', '거래량', '거래대금', '시가총액', '외국인 보유수량', '외국인 지분율', '토론방', '순매수\_기관', '순매수\_기타법인', '순매수\_개인', '순매수\_외국인', '검색어', '보도량', '기사감성점수', '뉴스심리지수', '비트코인종가', '비트코인거래량', '비트코인변동', '코스피종가', '코스피등락률', '코스피거래량', '코스피거래대금', '코스피시가총액', '한은금리', '원/미국달러', '원/위안', '원/일본엔(100엔)', '원/유로', '경제심리지수(원계열)', '경제심리지수(순환변동치)', '산업생산지수', '물가상승률', '소비자신뢰지수', '소비자심리지수', '경제활동참가율(%)', '실업률(%)', '고용률(%)', '코스피대비'

Y : 'day5\_label'

## XGB Classifier

데이터 : SK 하이닉스



클래스 불균형이 가장 심하지 않은 SK하이닉스 데이터를 기준으로  
하이퍼파라미터 튜닝 후 나머지 종목에 적용할 예정 !

[illegible]

## 2. 라벨인코딩

target label (day5\_label)에 대한 라벨인코딩 진행

buy(매수)	0
maintain(유지)	1
sell(매도)	2

## 3. MinMax Scaling

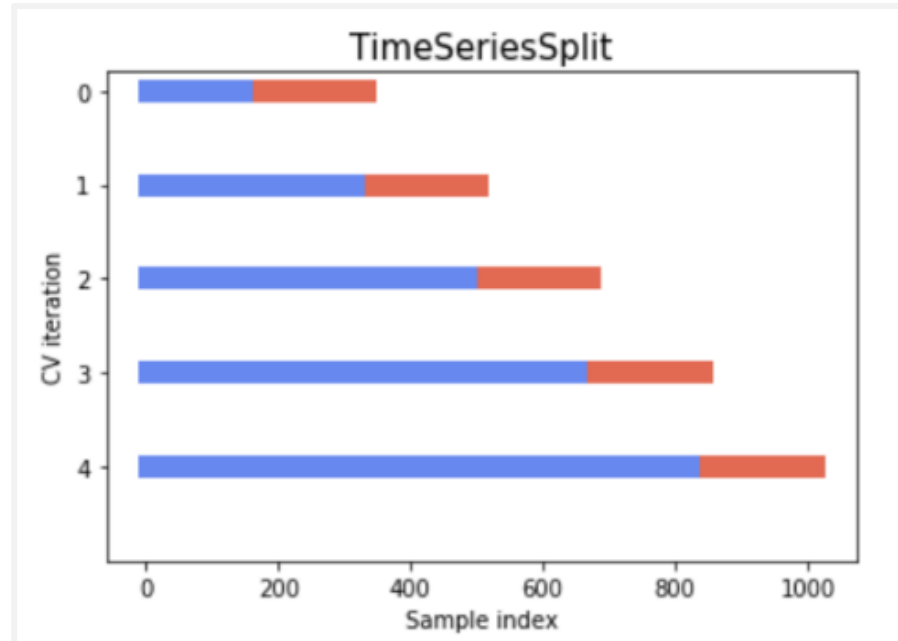
$$\frac{x - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)}$$

- 모든 연속형 X 변수에 대해서 MinMax scaling 진행
- 변수의 범위를 바꿔주는 정규화 스케일링 ( 범위 : [0, 1] )
- 종목 간 스케일 차이를 줄여 동일한 하이퍼파라미터에 적합하기 위함

# 3 최종 모델

## XGB Classifier

### 3. Expanding Window CV

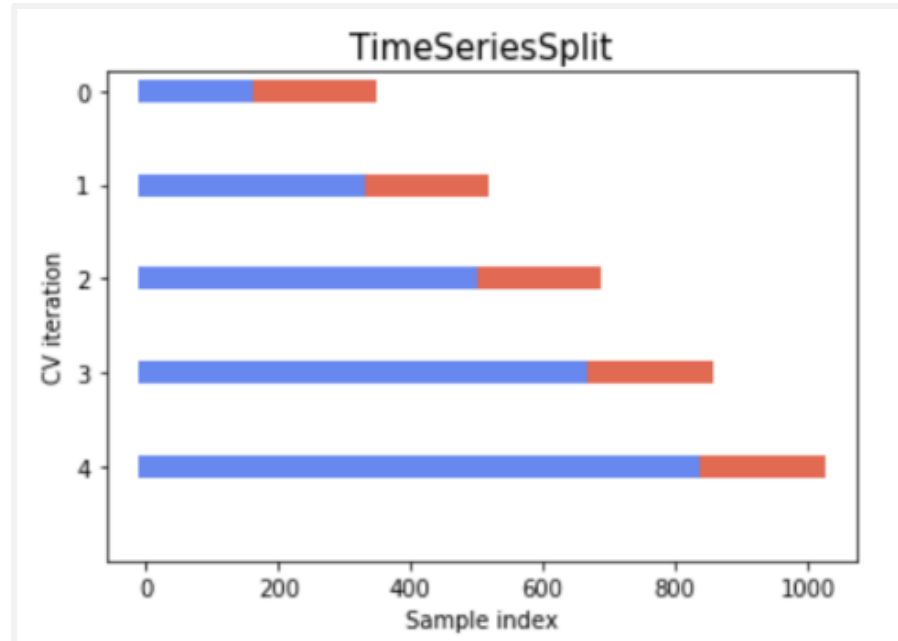


동일한 크기의 window를 누적하여 이동하는 시계열 교차검증 기법으로  
이전 단계의 train set과 validation set을 현재 단계의 train set으로 사용

# 3 최종 모델

## XGB Classifier

### 3. Expanding Window CV



**n\_splits = 4** 인 Expanding Window CV 활용 !

Split 횟수가 많아지면 validation set의 크기가 작아지기 때문에,  
하나의 validation set에서의 클래스 불균형이 너무 심각해지는 문제를 예방하기 위함

# 3 최종 모델

XGB Classifier

## 4. 클래스 가중치



클래스 불균형 문제를 어떻게 해결할 것인가???



클래스 불균형이 있는 데이터라면,  
scale\_pos\_weight 파라미터를 사용하면 되는 거 아니까요??



# 3 최종 모델

XGB Classifier

## 4. 클래스 가중치



클래스 불균형 문제를 어떻게 해결할 것인가???



클래스 불균형이 있는 데이터라면,  
scale\_pos\_weight 파라미터를 사용하면 되는 거 아니까요??



scale\_pos\_weight 파라미터는 이진 분류에만 사용 가능 ...

# 3 최종 모델

## XGB Classifier

### 4. 클래스 가중치



클래스 불균형 문제를 어떻게 해결할 것인가???



클래스 불균형이 있는 데이터라면,  
scale\_pos\_weight 파라미터를 사용하면 되는 거 아니까요??



다중 분류에서는 fit 함수의 파라미터로 **sample\_weight**를 사용 가능 !!!

# 3 최종 모델

## XGB Classifier

### 4. 클래스 가중치

각 클래스별 비율의 역수를 클래스별 샘플 가중치로 활용 !

```
class_weights = class_weight.compute_sample_weight( class_weight='balanced', y=y_train )
```

▶ 불균형한 학습 데이터에 대해 클래스별 샘플 가중치를 계산해주는 함수

```
xgb_model=xgb.XGBClassifier(**params, random_state = 42)  
xgb_model.fit(x_train, y_train, sample_weight=class_weights)
```

이런 식으로 fit 함수에서 활용 가능 !



# 3 최종 모델

## XGB Classifier

### 5. Optuna 하이퍼파라미터 튜닝

#### XGBoost Classifier 하이퍼파라미터

- max\_depth : 트리의 최대 깊이, 깊어질수록 복잡한 모델
- learning\_rate : 학습율
- n\_estimators : 트리의 개수
- min\_child\_weight : 분기에 필요한 최소 hessian 가중치, 작을수록 복잡한 모델
- gamma : 분기에 필요한 최소한의 loss 감소 비율, 작을수록 복잡한 모델
- subsample : 각 트리마다의 데이터 샘플링 비율
- colsample\_bytree : 각 트리마다의 feature 샘플링 비율
- reg\_alpha : L1 가중치, 작을수록 복잡한 모델
- reg\_lambda : L2 가중치, 작을수록 복잡한 모델

# 3 최종 모델

## XGB Classifier

### 5. Optuna 하이퍼파라미터 튜닝

정확도 (Accuracy)

$$\frac{TP + TN}{TP + TN + FP + FN}$$

정밀도 (Precision)

$$\frac{TP}{TP + FP}$$

Optuna 평가 지표

매수 정확도, 매수 정밀도, 매도 정확도, 매도 정밀도의 평균

## 5. Optuna 하이퍼파라미터 튜닝

정확도 (Accuracy)

거래 손익에 직접적 영향을 줄 수 있는 '매수', '매도'를 잘 맞추는 모델을 만들기 위해  
클래스 개수가 많은 유지를 제외한 매수, 매도의 **정확도**를 평가 지표에 포함

정밀도 (Precision)

Optuna 평가 지표

**매수 정확도**, 매수 정밀도, **매도 정확도**, 매도 정밀도의 평균

## 5. Optuna 하이퍼파라미터 튜닝

정확도 (Accuracy)

$$\frac{TP + TN}{TP + TN + FP + FN}$$

모델이 과도하게 매수, 매도로만 예측하는 것을 예방하고

'유지'에 대한 예측력도 유지하기 위하여

매수, 매도의 정밀도를 평가 지표에 포함

정밀도 (Precision)

$$\frac{TP}{TP + FP}$$



## Optuna 평가 지표

매수 정확도, 매수 정밀도, 매도 정확도, 매도 정밀도의 평균

# 3 최종 모델

## XGB Classifier

### 5. Optuna 하이퍼파라미터 튜닝

#### Best Parameters

- max\_depth : 14
- learning\_rate : 0.01843
- n\_estimators : 604
- min\_child\_weight : 6
- gamma : 0.13475
- subsample : 0.14467
- colsample\_bytree : 0.96275
- reg\_alpha : 8.16062e-05
- reg\_lambda : 2.07248e-08



# 3 최종 모델

## XGB Classifier

### 5. Optuna 하이퍼파라미터 튜닝

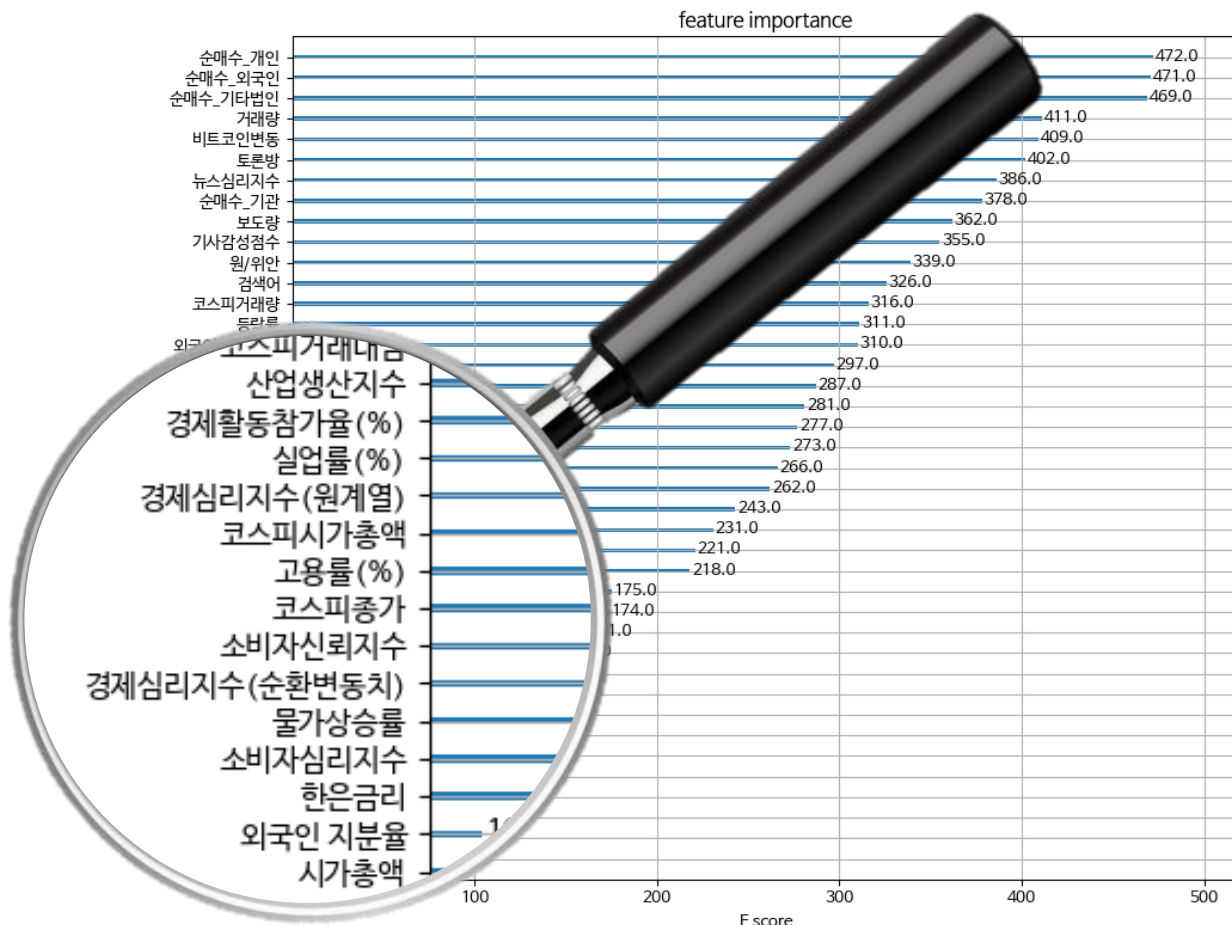


투자자별 순매수량, 비트코인변동, 토론방 게시글 수, 뉴스심리지수, 기사 보도량, 기사감성점수, 검색어량



투자자별 순매수량 데이터와  
여론 및 투자자 심리 관련 데이터가  
중요한 변수로 작용

## 5. Optuna 하이퍼파라미터 튜닝



산업생산지수, 경제활동참가율, 실업률, 코스피, 경제심리지수, 고용률, 물가상승률, 한은금리, ...



반면, 거시경제 관련 데이터는 상대적으로 덜 중요한 변수로 나타남

## 6. 예측

- with test set

## [ SK하이닉스 ]

```
===== SK 하이닉스 =====
[[27  7  3]
 [45 90 63]
 [ 6  7 33]]
```

전체 정확도 : 0.5338078291814946  
전체 f1-score : 0.5563170430999059

매수 정확도 : 0.7297297297297297  
매도 정확도 : 0.717391304347826  
유지 정확도 : 0.45454545454545453

F1 score : 0.56

매수 정확도 : 0.73

매도 정확도 : 0.72

## [ 현대차 ]

```
===== 현대차 =====
[[ 2  1  0]
 [24 117 11]
 [ 0  3  3]]
```

전체 정확도 : 0.7577639751552795  
전체 f1-score : 0.8229783067649851

매수 정확도 : 0.6666666666666666  
매도 정확도 : 0.5  
유지 정확도 : 0.7697368421052632

F1 score : 0.82

매수 정확도 : 0.66

매도 정확도 : 0.5

## [ 신한지주 ]

```
===== 신한지주 =====
[[ 10  8  1]
 [ 40 163 37]
 [ 0  13 13]]
```

전체 정확도 : 0.6526315789473685  
전체 f1-score : 0.6975956808520171

매수 정확도 : 0.5263157894736842  
매도 정확도 : 0.5  
유지 정확도 : 0.6791666666666667

F1 score : 0.69

매수 정확도 : 0.52

매도 정확도 : 0.5

# 3 최종 모델

## 최종 모델 소개

### XGB classifier

데이터 : SK 하이닉스

변수 : Full Model

평가지표 : custom한 optuna score

분류 : 다중 분류(매수, 매도, 유지)

### LSTM regressor

데이터 : SK 하이닉스

변수 : VIF로 선별된 변수

특징 : 라벨 인코딩 한 결과를

회귀로 예측하고

Threshold 결정 함수로 분류

## 1. 변수 선택

데이터 : SK 하이닉스

변수 : VIF를 통해서 선별한 변수

## 변수

X : '경제심리지수(순환변동치)', '시가총액', '비트코인종가', '원/미국달러', '소비자심리지수', '코스피거래대금', '순매수\_개인', '순매수\_외국인', '산업생산지수', '코스피거래량', '뉴스심리지수', '원/유로', '토론방', '실업률(%)', '원/일본엔(100엔)', '거래량', '기사감성점수', '외국인 보유수량', '코스피등락률', '경제활동참가율(%)', '검색어', '보도량', '순매수\_기관', '비트코인거래량', '비트코인변동', '5일 등락률'

Y : 'day5\_label'

## 1. 변수 선택

데이터 : SK 하이닉스

변수 : VIF를 통해서 선별한 변수

변수
<p>X : '경제심리지수(순환변동치)', '시가총액', '비트코인종가', '원/미국달러', '소비자심리지수', '코스피거래대금', '순매수_개인', '순매수_외국인', '산업생산지수', '코스피거래량', '뉴스심리지수', '원/유로', '토론폰방', '실업률(%)', '원/일본엔(100엔)', '거래량', '기사감성점수', '외국인 보유수량', '코스피등락률', '경제활동참가율(%)', '검색어', '보도량', '순매수_기관', '비트코인거래량', '비트코인변동', '5일 등락률'</p> <p>해당 날짜 5일 후의 등락률 예측을 바탕으로 결정한 범주형 변수 buy, sell, maintain의 3가지 범주로 구성</p> <p>Y : 'day5_label'</p>

# 3 최종 모델

## LSTM Regressor

### 2. 라벨인코딩

target label (day5\_label)에 대한 라벨인코딩 진행

buy(매수)	0
maintain(유지)	1
sell(매도)	2

### 3. MinMax Scaling

$$\frac{x - \text{Min}(X)}{\text{Max}(X) - \text{Min}(X)}$$

- 모든 연속형 X 변수에 대해서 MinMax scaling 진행
- 변수의 범위를 바꿔주는 정규화 스케일링 ( 범위 : [0, 1] )
- 분류 모델보다 **회귀 모델**에 적합

## 4. Window dataset 생성

## 일반적인 데이터

데이터셋을 train, test dataset으로  
랜덤하게 나누어서 평가 진행



## 시계열 데이터

랜덤하게 데이터셋을 나누면 시계열적  
특성 반영하지 못함



## 4. Window dataset 생성

## 일반적인 데이터

데이터셋을 train, test dataset으로  
랜덤하게 나누어서 평가 진행



## 시계열 데이터

랜덤하게 데이터셋을 나누면 시계열적  
특성 반영하지 못함

sliding window 방식을 통해서 window dataset 생성



## sliding window

설정된 window size를 바탕으로 전 시간 단계를  
학습에 사용하고, 이를 바탕으로 다음 시간 단계를 예측하는 방식

## 4. Window dataset 생성

일반적인 데이터

데이터셋을 train, test dataset으로

랜덤하게 나

window size

하나의 값을 예측하기 위해 학습에 사용할 시점의 개수

시계열 데이터

랜덤하게 데이터셋을 나누면 시계열적

지 못함

sliding window 방식을 통해서 window dataset 생성



sliding window

설정된 window size를 바탕으로 전 시간 단계를  
학습에 사용하고, 이를 바탕으로 다음 시간 단계를 예측하는 방식

## 4. Window dataset 생성

EXAMPLE ) window size = 3

sliding window

날짜	비트코인종가	실업률	거래량	검색어	보도량	...	day5_label
2017-07-11	2324.3	3.4	3187332	8.10396	58	...	1
2017-07-12	2403.1	3.4	3462150	8.16834	65	...	1
2017-07-13	2362.4	3.4	5432312	11.22361	90	...	1
2017-07-14	2234.2	3.4	2931832	9.64898	72	...	0
2017-07-17	2233.4	3.4	2804598	9.12856	50	...	0
2017-07-18	2320.2	3.4	2066194	7.92513	76	...	1
2017-07-19	2282.6	3.4	2009799	7.69511	42	...	1
2017-07-20	2866.0	3.4	1647153	7.71154	31	...	1

# 3 최종 모델

## LSTM Regressor

### 4. Window dataset 생성

EXAMPLE ) window size = 3

sliding window

날짜	비트코인종가	실업률	거래량	검색어	보도량	...	day5_label
2017-07-11	2324.3	3.4	3187332	8.10396	58	...	1
2017-07-12	2403.1	3.4	3462150	8.16834	65	...	1
2017-07-13	2362.4	3.4	5432312	11.22361	90	...	1
2017-07-14	2234.2	3.4	2931832	9.64898	72	...	0
2017-07-17	2233.4	3.4	2804598	9.12856	50	...	0
2017-07-18	2320.2	3.4	2066194	7.92513	76	...	1
2017-07-19	2282.6	3.4	2009799	7.69511	42	...	1
2017-07-20	2866.0	3.4	1647153	7.71154	31	...	1

→ X\_train[0]

→ y\_train[0]

# 3 최종 모델

## LSTM Regressor

### 4. Window dataset 생성

EXAMPLE ) window size = 3

sliding window

날짜	비트코인종가	실업률	거래량	검색어	보도량	...	day5_label
2017-07-11	2324.3	3.4	3187332	8.10396	58	...	1
2017-07-12	2403.1	3.4	3462150	8.16834	65	...	1
2017-07-13	2362.4	3.4	5432312	11.22361	90	...	1
2017-07-14	2234.2	3.4	2931832	9.64898	72	...	0
2017-07-17	2233.4	3.4	2804598	9.12856	50	...	0
2017-07-18	2320.2	3.4	2066194	7.92513	76	...	1
2017-07-19	2282.6	3.4	2009799	7.69511	42	...	1
2017-07-20	2866.0	3.4	1647153	7.71154	31	...	1

→ X\_train[1]

→ y\_train[1]

# 3 최종 모델

## LSTM Regressor

### 4. Window dataset 생성

EXAMPLE ) window size = 3

sliding window

날짜	비트코인종가	실업률	거래량	검색어	보도량	...	day5_label
2017-07-11	2324.3	3.4	3187332	8.10396	58	...	1
2017-07-12	2403.1	3.4	3462150	8.16834	65	...	1
2017-07-13	2362.4	3.4	5432312	11.22361	90	...	1
2017-07-14	2234.2	3.4	2931832	9.64898	72	...	0
2017-07-17	2233.4	3.4	2804598	9.12856	50	...	0
2017-07-18	2320.2	3.4	2066194	7.92513	76	...	1
2017-07-19	2282.6	3.4	2009799	7.69511	42	...	1
2017-07-20	2866.0	3.4	1647153	7.71154	31	...	1

→ X\_train[2]

→ y\_train[2]

# 3 최종 모델

## LSTM Regressor

### 4. Window dataset 생성

EXAMPLE ) window size = 3

sliding window

날짜	비트코인종가	실업률	거래량	검색어	보도량	...	day5_label
2017-07-11	2324.3	3.4	3187332	8.10396	58	...	1
2017-07-12	2403.1	3.4	3462150	8.16834	65	...	1
2017-07-13	2362.4	3.4	5432312	11.22361	90	...	1
2017-07-14	2234.2	3.4	2931832	9.64898	72	...	0
2017-07-17	2233.4	3.4	2804598	9.12856	50	...	0
2017-07-18	2320.2	3.4	2066194	7.92513	76	...	1
2017-07-19	2282.6	3.4	2009799	7.69511	42	...	1
2017-07-20	2866.0	3.4	1647153	7.71154	31	...	1

→ X\_train[3]

→ y\_train[3]

# 3 최종 모델

## LSTM Regressor

### 4. Window dataset 생성

EXAMPLE ) window size = 3

sliding window

날짜	비트코인종가	실업률	거래량	검색어	보도량	...	day5_label
2017-07-11	2324.3	3.4	3187332	8.10396	58	...	1
2017-07-12	2403.1	3.4	3462150	8.16834	65	...	1
2017-07-13	2362.4	3.4	5432312	11.22361	90	...	1
2017-07-14	2234.2	3.4	2931832	9.64898	72	...	0
2017-07-17	2233.4	3.4	2804598	9.12856	50	...	0
2017-07-18	2320.2	3.4	2066194	7.92513	76	...	1
2017-07-19	2282.6	3.4	2009799	7.69511	42	...	1
2017-07-20	2866.0	3.4	1647153	7.71154	31	...	1

→ X\_train[4]

→ y\_train[4]



# 3 최종 모델

## LSTM Regressor

### 4. Window dataset 생성

Window size = 10으로 window dataset 생성

기존 data size : (1409, 27)



window sliding 적용

data size : (1399, 10, 27)



10개의 time step으로 하나의 time step을 예측하기 때문에  
앞의 10개의 time step은 label에서 제외됨

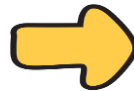
# 3 최종 모델

## LSTM Regressor

### 4. Window dataset 생성

Window size = 10으로 window dataset 생성

기존 data size : (1409, 27)



window sliding 적용

data size : (1399, 10, 27)

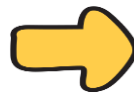


window size

## 4. Window dataset 생성

Window size = 10으로 window dataset 생성

기존 data size : (1409, 27)



window sliding 적용

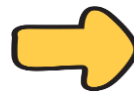
data size : (1399, 10, 27)



window size

## 5. train, validation, test split

data size : (1399, 10, 27)



Train set : (1120, 10, 27)

Validation set : (140, 10, 27)

Test set : (139, 10, 27)

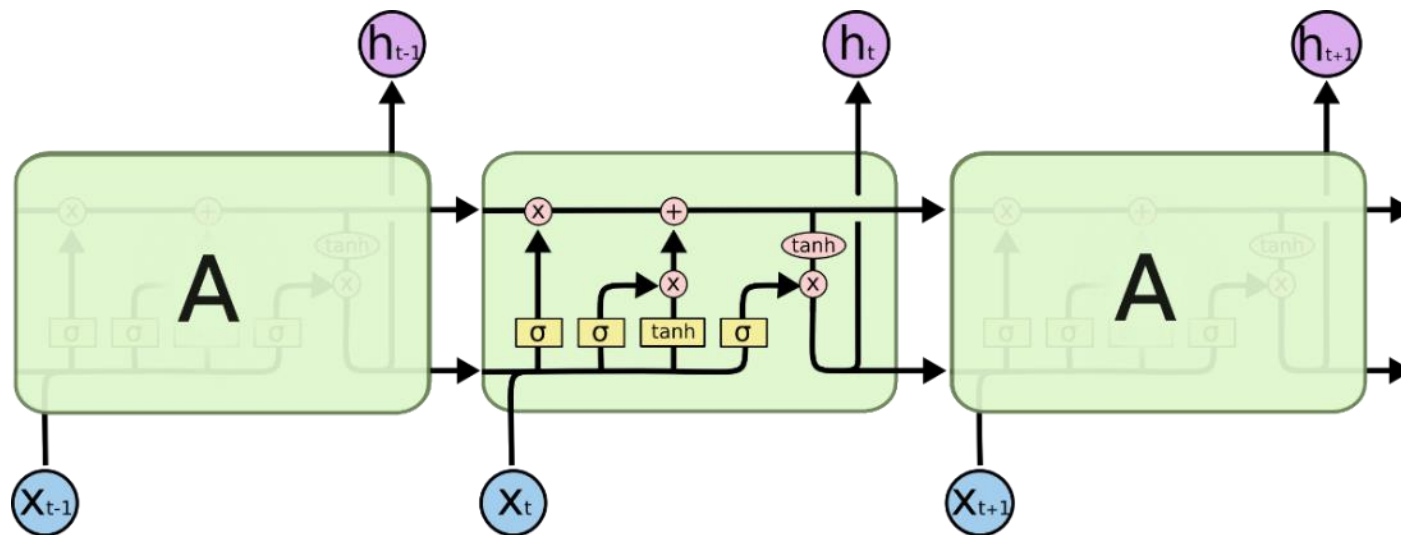
# 3 최종 모델

## LSTM Regressor

### 6. LSTM Regressor

- with train set

LSTM 관련 자세한 설명은 딥러닝팀 클린업 3주차 참고!



RNN의 장기 의존성 문제 해결을 위해 등장한 구조로  
입력 게이트, 망각 게이트, 출력 게이트로 구성되어 장단기 기억력이 뛰어난 모델

## 6. LSTM Regressor

- with train set

hidden_size	2
num_layers	1
learning_rate	0.0001
loss function	MSE loss
optimizer	Adam
epoch	8000



데이터의 y label은 0, 1, 2로 구성된 범주형 변수지만 회귀를 통해서 예측

→ 최적의 모델 체크포인트로 저장

## 6. LSTM Regressor

- with validation set

	매도 정확도	매수 정확도	유지 정확도	매도 정밀도	매수 정밀도	f1 score	평균
조합 1	0	0	X	0	0	X	
조합 2	0	0	0	X	X	X	★BEST★
조합 3	0	0	X	X	X	0	
조합 4	0	0	0	X	X	0	

각 예측값을 범주형 데이터로 분리하기 위한 threshold(경계값) 결정



데이터의 y label은 0, 1, 2로 구성된 범주형 변수지만 회귀를 통해서 예측

→ 예측값이 범주형이 아닌 수치형 데이터

## 6. LSTM Regressor

- with validation set

	매도 정확도	매수 정확도	유지 정확도	매도 정밀도	매수 정밀도	f1 score	평균
조합 1	0	0	X	0	0	X	
조합 2	0	0	0	X	X	X	★BEST★
조합 3	0	0	X	X	X	0	
조합 4	0	0	0	X	X	0	



## Threshold 함수의 역할

- Validation set 기준, 매수 정확도, 매도 정확도, 유지 정확도의 평균이 최대가 되는 조합으로 threshold 결정
- 결정된 threshold 바탕으로 매수/매도/유지 예측 자동화

## 7. 예측

- with test set

## [ 신한지주 ]

```
===== 신한지주 =====
[[ 23  15   0]
 [ 24 529  62]
 [   1  15  39]]
```

```
전체 정확도 : 0.8347457627118644
전체 f1-score : 0.8503657789754228
```

```
매수 정확도 : 0.6052631578947368
매도 정확도 : 0.7090909090909091
유지 정확도 : 0.8601626016260162
```

F1 score : 0.85

매수 정확도 : 0.61

매도 정확도 : 0.71

## [ SK하이닉스 ]

```
===== SK하이닉스 =====
[[104  18   0]
 [113 313  45]
 [   3  27  76]]
```

```
전체 정확도 : 0.7052932761087267
전체 f1-score : 0.7165113879684445
```

```
매수 정확도 : 0.8524590163934426
매도 정확도 : 0.7169811320754716
유지 정확도 : 0.6645435244161358
```

F1 score : 0.72

매수 정확도 : 0.85

매도 정확도 : 0.67

## [ 현대차 ]

```
===== 현대차 =====
[[ 9  9  3]
 [ 47 248 57]
 [ 4  7 14]]
```

```
전체 정확도 : 0.6809045226130653
전체 f1-score : 0.7416229778038823
```

```
매수 정확도 : 0.42857142857142855
매도 정확도 : 0.56
유지 정확도 : 0.7045454545454546
```

F1 score : 0.74

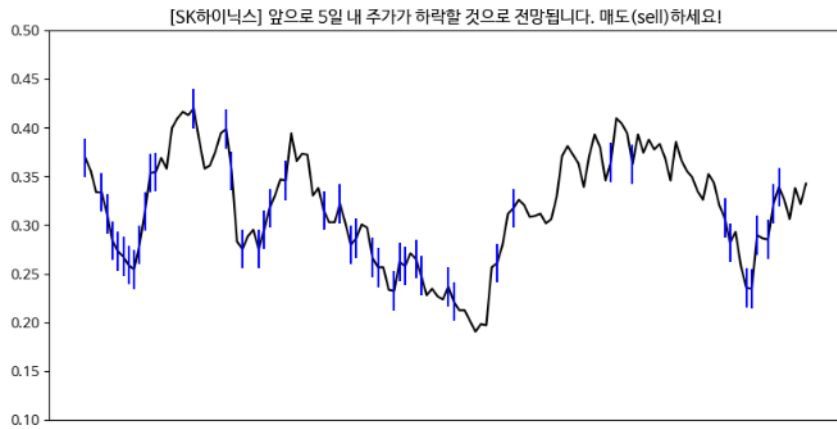
매수 정확도 : 0.43

매도 정확도 : 0.56

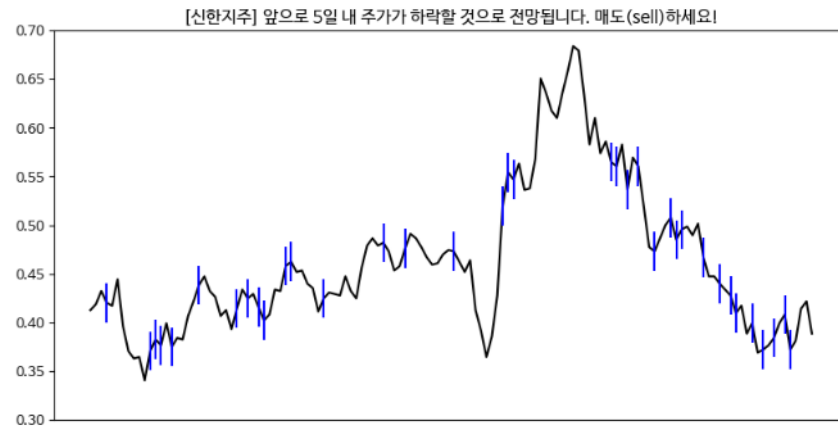


# 3 최종 모델

## 예측 결과 시각화



[ SK하이닉스 ]



[ 신한지주 ]



[ 현대차 ]

# 3 최종 모델

## 예측 결과 시각화



[ SK하이닉스 ]



[ 신한지주 ]



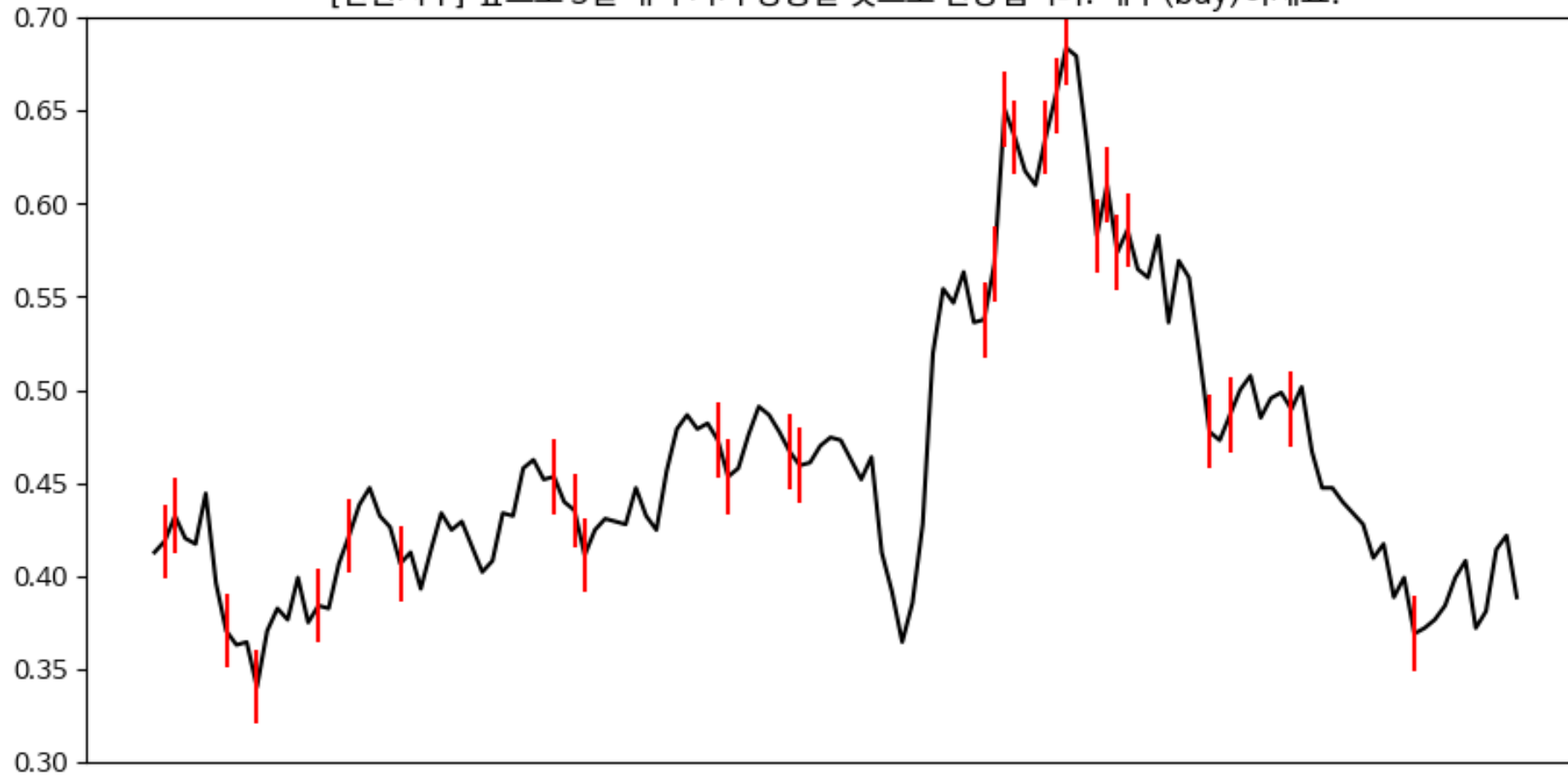
[ 현대차 ]

예측 결과 차트 톱아보러 가보자고 !!!

# 3 최종 모델

## 예측 결과 시각화

[신한지주] 앞으로 5일 내 주가가 상승할 것으로 전망됩니다. 매수(buy)하세요!

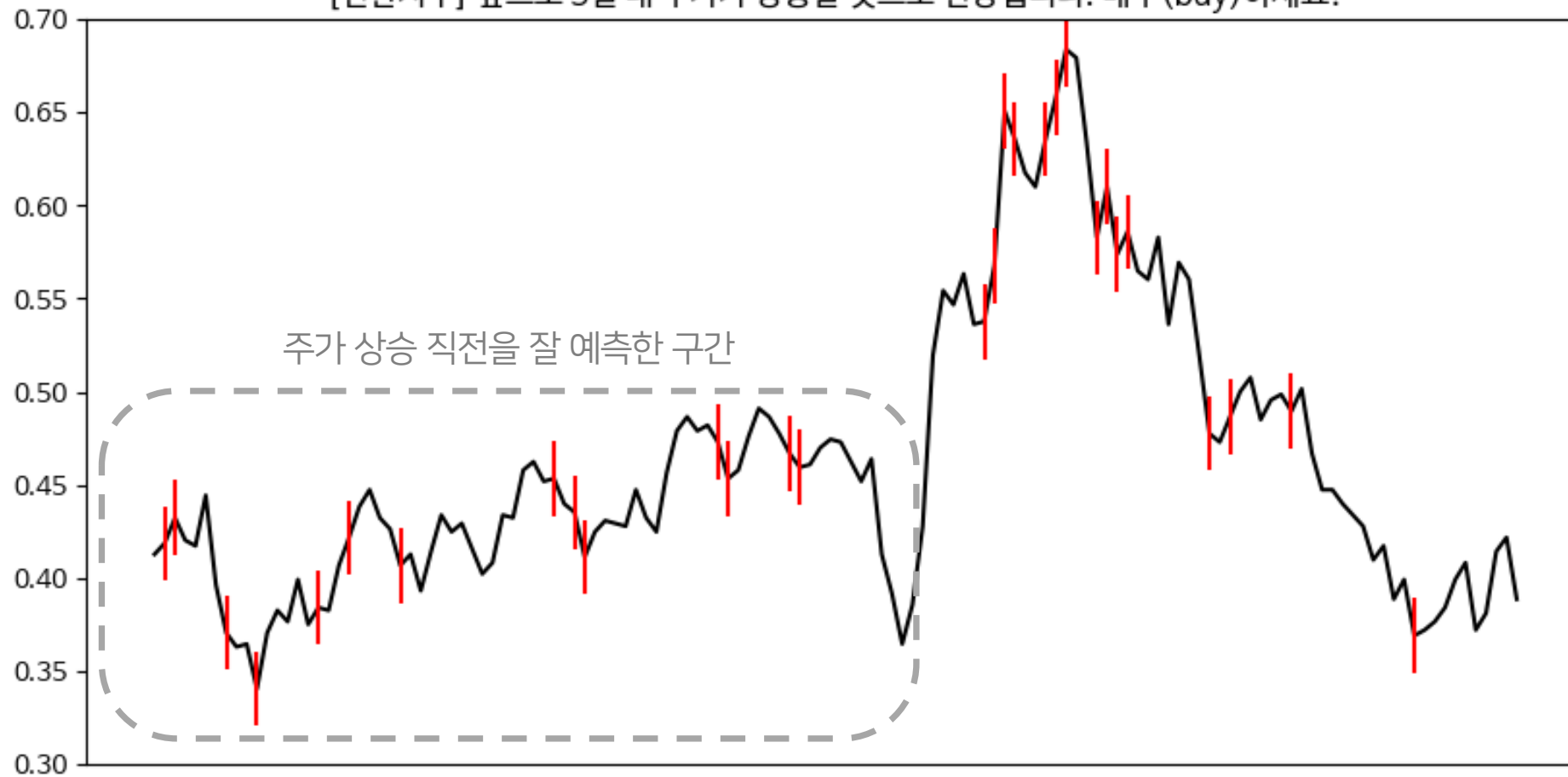


빨간색 지점을 기준으로  
5일 이내 주가가 5% 이상  
상승될 전망  
▶ 매수 추천

# 3 최종 모델

## 예측 결과 시각화

[신한지주] 앞으로 5일 내 주가가 상승할 것으로 전망됩니다. 매수(buy)하세요!



빨간색 지점을 기준으로  
5일 이내 주가가 5% 이상  
상승될 전망  
▶ 매수 추천

# 3 최종 모델

## 예측 결과 시각화

[신한지주] 앞으로 5일 내 주가가 상승할 것으로 전망됩니다. 매수(buy)하세요!

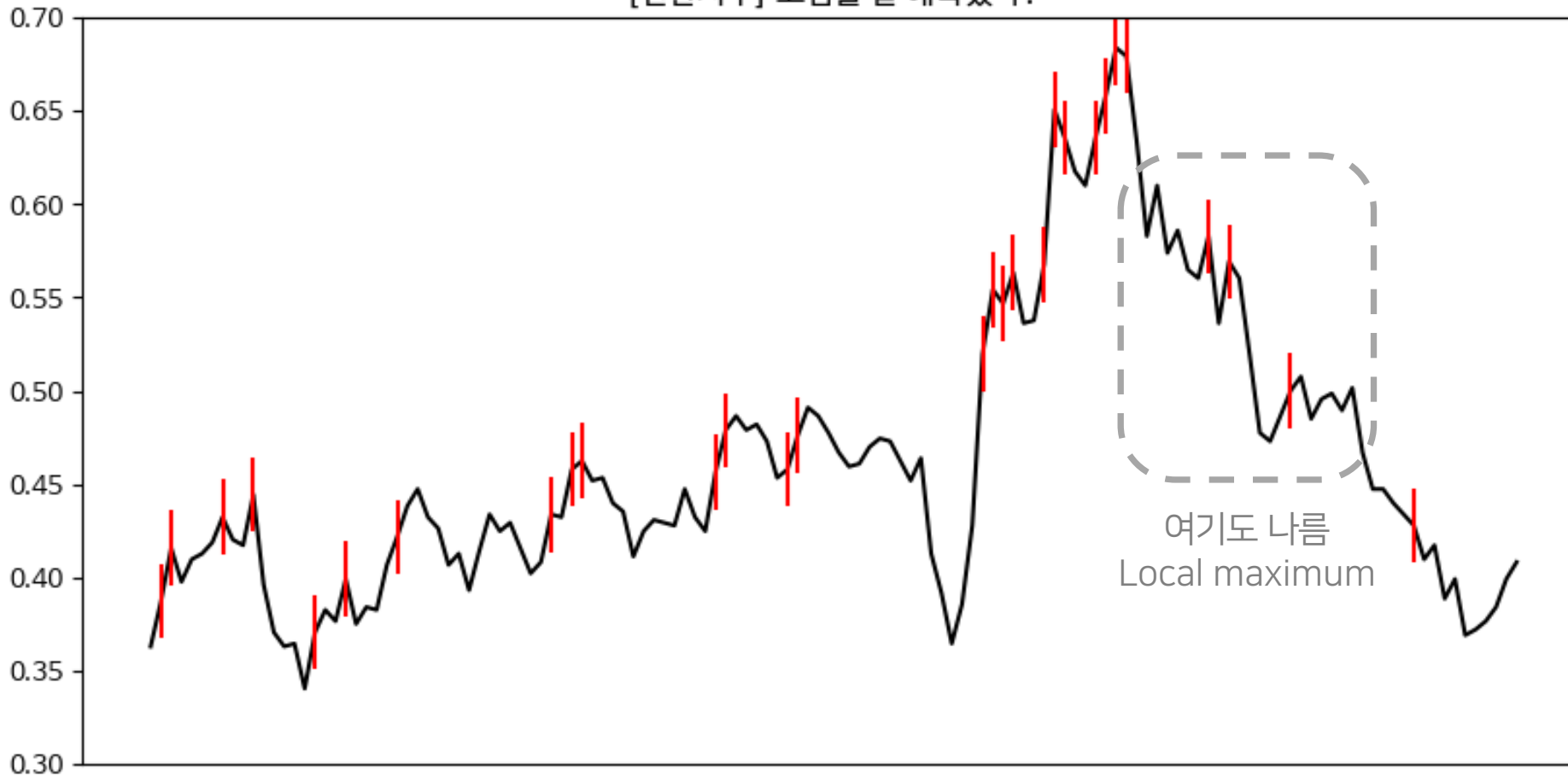


빨간색 지점을 기준으로  
5일 이내 주가가 5% 이상  
상승될 전망  
▶ 매수 추천

# 3 최종 모델

## 예측 결과 시각화

[신한지주] 고점을 잘 예측했나?

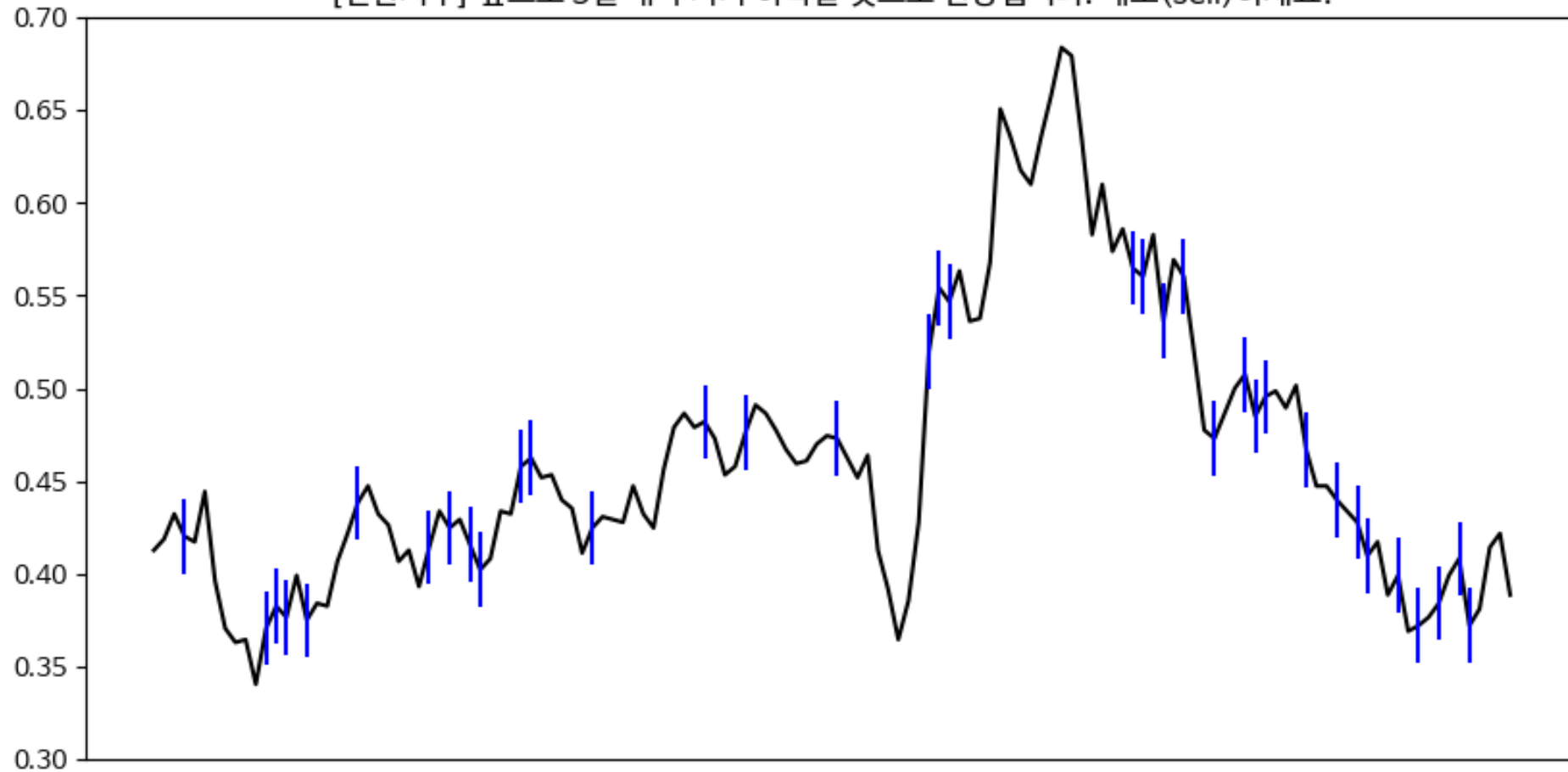


빨간색 지점을 5일 이후  
시점으로 이동한 결과,  
대체로 주가가 국소적 극대를  
기록하는 지점에 잘 매칭됨!

# 3 최종 모델

## 예측 결과 시각화

[신한지주] 앞으로 5일 내 주가가 하락할 것으로 전망됩니다. 매도(sell)하세요!

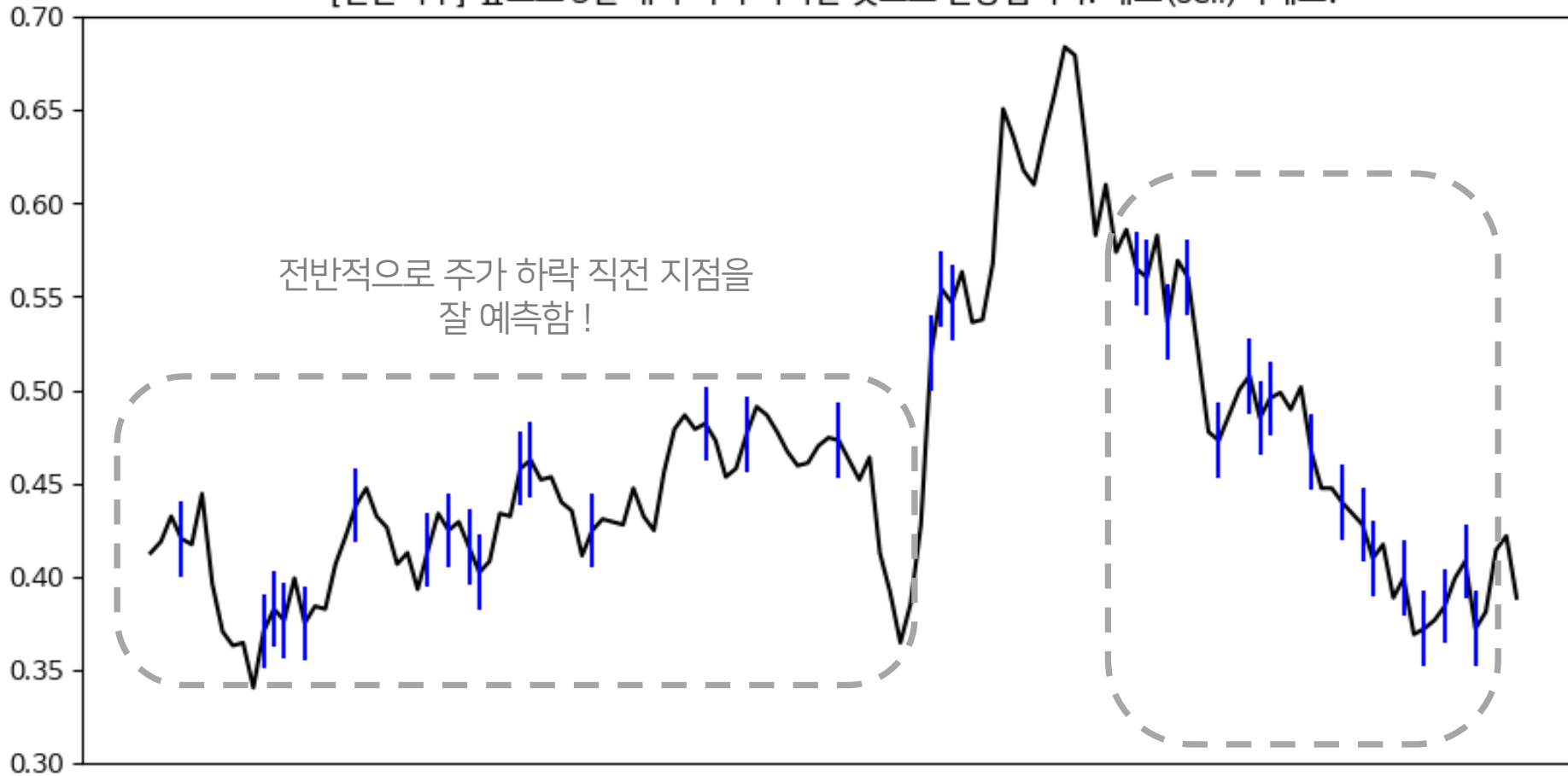


파란색 지점을 기준으로  
5일 이내 주가가 5% 이상  
하락할 전망  
▶ 매도 추천

# 3 최종 모델

## 예측 결과 시각화

[신한지주] 앞으로 5일 내 주가가 하락할 것으로 전망됩니다. 매도(sell)하세요!



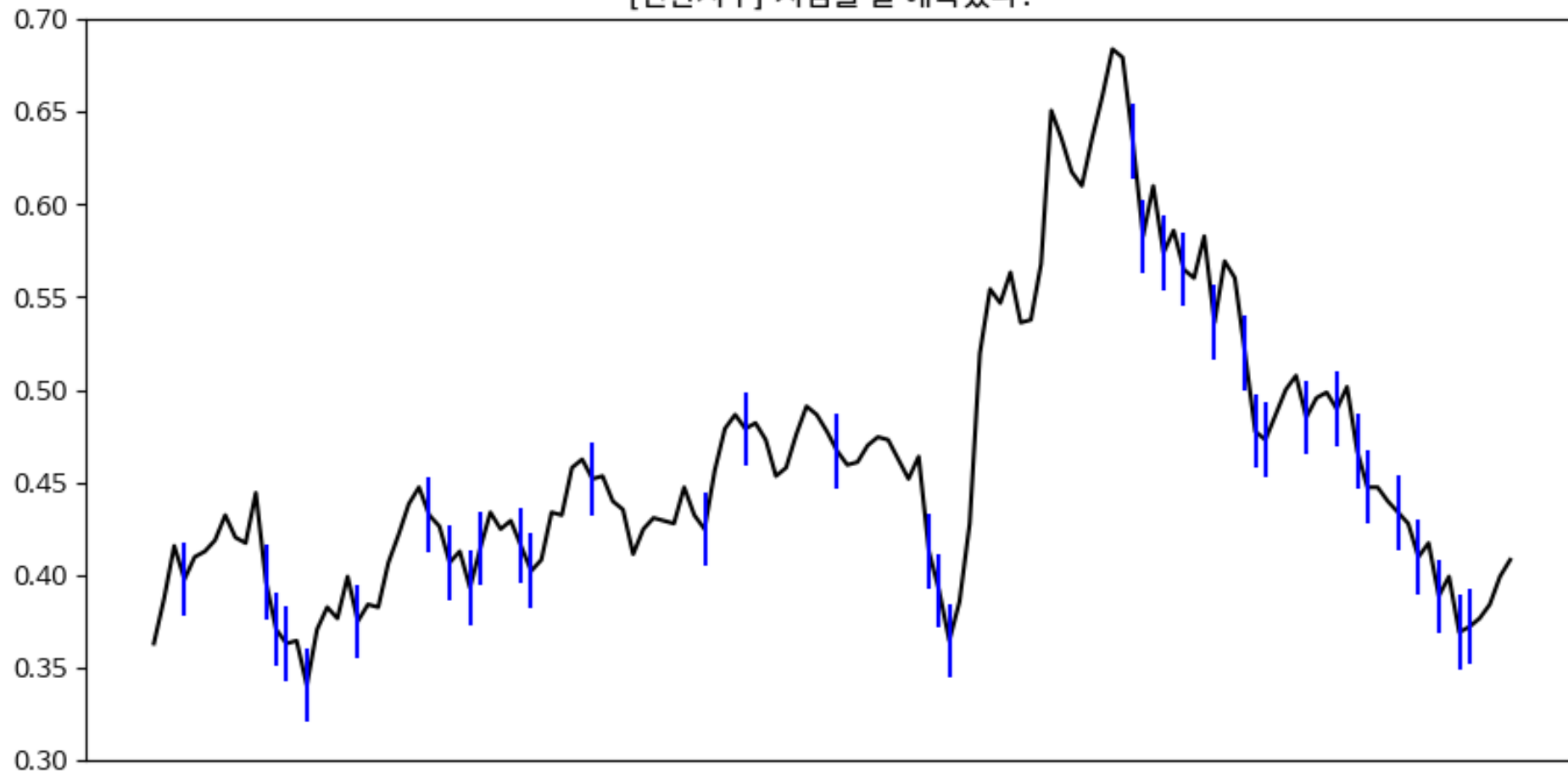
파란색 지점을 기준으로  
5일 이내 주가가 5% 이상  
하락할 전망  
▶ 매도 추천



# 3 최종 모델

## 예측 결과 시각화

[신한지주] 저점을 잘 예측했나?



파란색 지점을 5일 이후  
시점으로 이동한 결과,  
대체로 주가가 국소적 극소를  
기록하는 지점에 잘 매칭됨!

# 4

결론

### 주식 초보자의 리스크 관리를 위한 매수/매도 추천 서비스

주식 등락률에 영향을 미치는 데이터를 모델에 입력 → 모델이 매도/유지/매수를 학습  
투자자들에게 현재 상황에 따른 매도 또는 매수를 추천!

▶ 누구나 쉽고 간편하게 참고할 수 있는 투자 인사이트를 제공하는 서비스



주식 초보자, 이른바 주린이들이  
투자 시 참고할 수 있는 쉽고 간단한 투자 지표 제공

# 4 결론

## 기대효과 및 확장성

### 기대효과

- 어플로 제작하여 본 서비스를 제공한다면, **주식을 막 시작하는 20대 고객 유치** 가능
- 대부분의 사회 초년생 및 주식 초보자 고객은 처음 가입한 플랫폼을 지속적으로 사용하기 때문에, **충성도가 높은 고객 확보** 가능

### 확장성

- 등락률 **라벨링 기준**을 원하는 값으로 **입력 받아**, 개인의 **취향을 반영**한 매수/매도 추천 가능
- 투자**성향**에 따라 **고객 맞춤형** 매수/매도 추천 서비스 제공 가능

# 4 결론

## 의의 및 한계

### 의의

- 등락률 예측을 위한 데이터로 정형 및 비정형 데이터 및 다양한 데이터를 사용하여 유의미한 결과 도출
- 주식 데이터의 특성(imbalanced data, time series data)을 반영하여 분석 진행
- 도메인에 크게 영향을 받지 않고 일정한 정확도를 보이는 강건한 모델

### 한계

- 실생활에 적용하기 위해서는 데이터 수집 자동화 필요
- 데이터셋으로 사용했던 3가지 종목 외에 더 다양한 종목에 적용 가능한지 시험해보지 않은 것

# 4 결론

## 주제분석 후기



[ 김 민 ]

피셋 마지막 학기가 정말 쏠살같이 지나가버렸네요... 이번 시계열팀에 임하면서 개인적으로도 능력적 한계를 마주하고 여러모로 부족하다는 생각을 했지만, 그럼에도 제가 할 수 있는 최선을 다했다고 생각합니다...! 클린업도 주제분석도 잘 마무리할 수 있을지 스스로도 확신을 못했는데, 능력 있는 팀원들 덕분에 성공적으로 마무리할 수 있었던 것 같습니다. 이번 학기 우리 팀, 수린 언니, 동환 오빠, 유진이, 다연이와 함께하지 않았다면 이만큼의 결과를 내지 못했을 거라고 생각해요. 비록 제가 팀장이긴 하지만 팀원들에게 더 배워가는 게 많았던 학기였던 것 같습니다. 함께 피셋 졸업하는 수린 언니는 피셋 1년 잘 버텼으니 앞으로 어디서든, 뭐든 잘 해내리라 믿어 의심치 않고, 한 학기 남은 동환 오빠, 유진이, 다연이는 뭐 너무 잘해서 남은 한 학기도 멋지고 의미 있게 보낼 수 있으리라고 말하기가 입만 아픕니다. 모두에게 이번 한 학기 활동이 좋은 기억, 의미 있는 결과물로 남았으면 좋겠어요... 다들 수고 정말 많았고, 시계열팀으로 같이 활동해줘서 고맙습니다!! 피셋 끝나도 저한테 연락해주세요ㅠㅠ 마지막으로 무엇보다도 건강이 제일 중요하니 다들 항상 건강합시다!!

# 4 결론

## 주제분석 후기



드디어 psat이 끝나게 되었네요.... 패키지와 주분으로부터 벗어난다니 기쁜 한편 이렇게나 맛있고 좋은 사람들과 열심히 활동할 수 있는 기회가 또 있을까 아쉬운 생각도 듭니다. 첫 학기 때는 r markdown도 쓸 줄 모르는 코린이였는데... 허허허... 이런 저에게 정말 큰 힘이 되어준 지영, 지윤, 지훈오빠, 희지까지 정말 최고였어요!! 선대팀을 만난 덕분에 psat의 첫 단추를 잘 끼울 수 있었던 것 같습니다. 그리고 아직 너무나 부족한 실력인데도 불구하고 기존 기수가 되어 부담스러웠는데... 우리 짱시계열팀을 만나 한 학기 부담 없이 즐겁게 보낸 것 같습니다. 어려운 시계열 내용을 척척 설명해준 민이, 첫 학기 활동인데도 뭐든지 척척 해내는 동환, 유진, 다연이까지 정말 사랑합니다!!! 99즈와 01즈의 조합은 역시 ♥ 이지~~~ 비록 공모전 결과는 아쉽지만 5명이 한 마음 한 뜻이 되어 최선을 다한 건 두고두고 잊지 못할 겁니다. 물론 같이 국밥집으로 회식 간 것도 잊지 못할것 같네요 ㅎㅎ 그 동안 우리 짱시계열팀 진짜 수고했엉~~~ 앞으로도 꼭 친하게 지내요 여러분!!! 마지막으로 아직 한 한기 더 남은 동환, 유진, 다연을 포함한 31기 분들 모두 화이팅입니다!!!! psat FOREVER ♥ ♥ ♥



[ 이수린 ]

# 4 결론

## 주제분석 후기



피셋에 들어온지 얼마 안 된 것 같은데 벌써 한 학기가 지나가다니  $\pi\pi\pi$  후기를 쓰고 있는데도 실감이 나지 않아요...  $\pi\pi$  휴학 중이었는데 피셋을 하게 되어서 휴학 생활을 조금 더 의미있게 보낼 수 있었던 것 같아요! 들어와서 보니 정말 많은 분들이 열심히, 또 치열하게 시간들을 보내고 있다고 느꼈어요. 그래서 정말 많이 배우고 느꼈던 한 학기였습니다!! 일단 시계열팀! 너무 좋은 팀장과 팀원들 덕분에 즐겁고 알차게 한 학기를 보낼 수 있었던 것 같아요! 이번 학기 휴학 중이어서 학교에서 자주 마주치지 못했었는데, 휴학을 하지 않고 학교에서 수업도 같이 듣고 했으면 어땠을까 아쉽기도 해요  $\pi\pi$  먼저 나랑 동갑인데도 이렇게 많은 걸 알고 할 수 있나? 싶었고 배울 점도 정말 많았고 그래서 정말 많이 의지가 되었던 민이, 팀장을 제외하면 유일한 기존 기수라서 고생했을텐데 내색 않고 많이 알려주고 맡아준 수린언니, 처음 맡아보는 일도 너무 척척 잘 해와서 항상 놀라게 했던 현 팀원 곧 팀장 다연이, 주분 아이디어부터 시작해서 프로젝트 전반에 아이디어를 정말 많이 내준, 모른다고 하지만 알고 보면 다 할 줄 아는 동환오빠까지 다들 너무 감사드려요...♥♥♥ 내향인들끼리 잘 맞아서 너무 좋았고 덕분에 많이 배웠습니다!! 우리 짱시계열을 포함한 모든 피셋분들께 감사하고, 다음 학기에는 기존 기수가 되는데 더 열심히 공부해서 지금의 기존 기수처럼 많이 배우고 많이 알려줄 수 있는 사람이 되도록 노력하겠습니다!!



[ 서유진 ]



# 4 결론

## 주제분석 후기



[ 김동환 ]



시계열팀 모두 고생 많았습니다. 처음에 솔직하게 혼자 남자여서 이래저래 걱정이 많았습니다. 특히 처음 만나서 밥먹기로 한 날, 가서 조용히 쓸데없는 소리 안하고 밥만 먹고 와야겠다고 다짐한 기억이 납니다. 다행히도 다들 착하고(?) 성격이 좋아서(흠흠) 그리고 알게 모르게 잘 어울릴 수 있게 배려해 준 거도 있는거 같고, 해서 한학기 잘 해 낼 수 있었습니다. 감사의 인사 전합니다.(—)

우리 미모와 지성을 겸비한(^\_^) 팀장 민아, 처음 전화와서 어색하게 어 근데요... 제가 그럴라고 그런건 아닌데, 뽕다보니 혼자 남자 시더라고요 괜찮으세요? 라고 하던게 벌써 한참 전이구나 ㅋㅋ. 되게 다양한 인간 군상을 만나온거 같은데, 정말 앞으로 내가 회사생활을 하게 된다면 내 상사는 너같은 사람이면 좋겠다는 생각을 자주 했어. 적당히 귀찮아하지만? 매우 열심히하는, 또 뭔가 맨날 열심히 하고 있는 뭔가 모순같은 사람이라고 느꼈거든. 또 어느 조직이든 팀이든 머리에는 다른 사람의견 잘 수용해주고 전체가 하나로 잘 녹아들수 있도록하는 사람이 있어야 잘 굴러가는 법인데, 그런 점에서 너가 굉장히 돋보였던거 같아. 앞으로 꼭 너같은 사람 상사로 만나길 바랄게. 취준하느라 바쁜 거같은데, 가서도 잘 살어 응원할게.

집갈때 제일 행복한 수린이 처음에 다들 나보다 어린 줄 알았는데, 나랑 동갑이 있더라고?? 혼자 내적 친밀감을 느끼고있었던단... 항상 피피티 종합해줘서 고마워, 고생 많았습니다. 근데 나는 진짜 열심히 만들었거든? 후.. 근데 그게 마음대로 안되더라고. 완성본 피피티 보면 항상 고마움과 미안함을 같이 느꼈어. 근데 지금 생각해보니까 난 나름 괜찮게 만들었는데, 민이가 날 가스라이팅 한 거 같아. 이게 맞다. 한 학기동안 너랑 나랑 수업이 많이 겹쳐가지고 얼굴은 젤 많이 본거같아 ㅋㅋ 오며가며 인사도 많이하고 하다보니까 느껴지는게 항상 집가고 싶어하는 눈이야. 근데 수업은 또 항상 일찍와서 열심히 듣더라고? 내가 분석했을 때 너는 현대 직장인의 삶에 가장 적합한 인간인거같아. 일을 하기 싫은 마음은 굴뚝같지만, 몸이 거부하지 못하고 출근하고 있는 어때 무슨 느낌인지 알겠지? 한 학기 동안 시계열팀에서 활동하면서 뭔가 난 하고있는 줄도 몰랐는데, 민이가 카톡에 "아 수린 언니가 ~했는데 결과가 ~" 라고 꽤나 자주 말을 했던 기억이 있어. 그런 거 볼 때마다 참 든든했어. 기존 기수들이 잘하니까 팀도 나름 잘 굴러간거같아. 나도 다음학기 신입 기수들에게 든든한 사람이 되도록 노력해볼게. 덕분에 한학기 잘 지내고 갑니다. 간간히 연락하면서 지내자. 꼭 원하는데로 알파팀이고 학회장팀이고 하지말고 행복하게 끝낼 수 있길 바래.

# 4 결론

## 주제분석 후기



[ 김동환 ]



ㅋㅋ결국 팀장을 하기로 했구나 다연아. 난 진짜 그냥 하는 소린줄 알았는데, 진짜였다니... 항상 수업시간에 맨 앞에 앉아서 열심히 듣는 너를 보면서 나는 대충 2주차부터 너를 팀장 후보로 정해놨었어. 아니 근데 점점 학기가 지나갈 수록 앞에 앉던 사람이 앞에 안 보이더라고?? ㅋㅋ첫 인상하고 가장 차이가 많은 사람이 너인거같아. 한 학기 서로 모르는거 문답을 많이 했는데, 진짜 도움 많이 받고 또 준거 같아. 근데 다음학기 배신하고 혼자 휴학하는게 맞나? 난 모르는거 누구한테 물어보냐? 다음학기 아마 다른팀 되겠지만 피셋 같이 잘 해보자 ^\_^

뭔가 많이 들고 다니는 유진아, 항상 멀리서 오느라 고생했다. 근데 뭔가 많이 들고 다녀서 더 고생하는거 같긴해 ㅋㅋ. 세미나 들을 때 항상 크런키를 까먹던 너... 오물오물 잘 먹더라. 그러다 눈마주치면 오빠도 줘?ㅋㅋㅋ 가끔 길가다 생각나면 하나씩 사다줄게. 주분 때 다들 시험기간이 겹쳐서 일정이 좀 빠듯했는데, 너가 자기 휴학생은 무적이라며 혼자 자처해서 내가 다 하겠다 했을 때 정말 감격스러웠어. 아직도 잊혀지지않는다. 진짜 서유진은 신이다. 다음학기도 휴학하고 나랑 같은팀 해줘.

우리 팀은 다들 뭔가 조용조용하고 힘이 없는 거 같다고 느끼고 있었는데, 점점 마지막 주차로 향해 다가 갈수록 "아, 다들 연기하고 있었던 거구나" 라는 생각을 했어. 정신없이 패키지하고 피피티 만들고 학교 공부하고 시험보고 하다 보니 벌써 한학기가 끝나네. 이렇게 정신없이 살았던게 정말 오랜만인거 같은데, 좋은 사람 좋은 팀 만나서 이번학기 잘 버티고 또 잘 보낼 수 있었어. 마음 같아선 이 팀 그대로 한번 더하고 싶은데, 상황이 상황인게 참 아쉬울 따름이야. 그리고 뭔가 우리팀은 온라인으로 많이해서 더 아쉽고 그런거같아. 다음 학기도 갈 사람은 다른데 가서도 잘하고, 피셋에 남는 사람은 또 남는데로 화이팅하자~ 짱시계열 다들 고생많았 습니다!!

# 4 결론

## 주제분석 후기



[ 장다연 ]

피셋 면접을 보고 민이로부터 합격 전화를 받은 게 어제 같은데 벌써 피셋에서의 한 학기가 끝나다니 정말 믿기지가 않네요!! 피셋에서 활동하면서 처음 생각했던 것보다 더 많은 것들을 얻어가는 것 같습니다. 많은 멋진 사람들을 만나고 그 사이에서 활동하면서, 어디에서도 할 수 없는 것들을 경험하고 배우며 성장했습니다. 그리고 이 모든 건 역시 우리 짱시계열 덕분이겠종~~ 장다연 선정 피셋 역사상 가장 친절한 팀장 민이, 항상 묵묵하고 또 깔끔하게 피피티 취합 등 여러 가지를 맡아서 해준 수린언니 덕분에 너무 든든했습니당!!! 휴학생이라는 이유만으로 엄청난 많은 일들을 해내준 유진이, 아이디어뱅크이자 나의 챗지피티 동환오빠까지! 신입기수라고는 믿을 수 없는 실력자들 덕분에 우리팀이 더 잘 굴러갈 수 있던 것 같고, 이런 사람들과 함께 든든하면서도 편안한 한 학기 보낼 수 있어 행복했습니다!!! ㅎㅎ 한 학기 동안 시계열 금쪽이를 잘 챙겨주고 이끌어줘서 정말 고마웠어 이 기억들과 배움을 통해 다음 학기에는 더 멋진 기존기수로 활동하겠습니다!!!! 민이랑 수린언니는 탈출을 축하하고, 유진이랑 동환오빠는 나랑 한 학기 더 고생하자!!!! 팀은 해체되지만... 영원히 짱시~~~♥ 피셋 끝나고도 자주 보자구요 다들!!! 쪽~🎯 항상 학회원들 생각해주고 좋은 패키지 문제 내주신 학회장팀도 감사했습니다! 클린업과 주제분석을 통해 다양한 이론 및 실제 분석 과정을 알려준 다른 팀 분들에게도 감사하다는 말 전하고 싶습니당 ㅎㅎㅎ 다들 한 학기 수고 많으셨습니다!!!

감사합니다!!!



지금까지  
짱시계열이었습니다~~