

# 주식 게임

- 대주주 및 임원의 지분공시를 통한 미래의 주식 가격 추이 예측 -

김예슬, 김정우, 김학선, 손용원, 손태산, 최지웅

# 목차

---

1. 개요
2. 가설설정
3. 데이터수집 및 전처리
4. EDA
5. 모델링
6. 결론



개요

# 지분 공시

시장의 투명성 제고 및 투자자 보호  
목적으로 상장회사 주식등의 소유 및 변동  
정보를

임원 및 주요주주의 미공개정보 이용을 사  
전에 예방하기 위한 제도.

---

# 카카오페이, 카카오 CEO 내정자 사퇴

## 카카오 급락, 류영준 내정자 사퇴... '시총 1조 증발'

메트로신문 구남영 기자 | 2022-01-10 15:53:39



<https://www.hani.co.kr> > arti > economy > finance ▼

### '카카오페이 먹튀' 막는다...임원 주식 매각 30일 전 공시해야 : 금융

2022. 9. 13. — 내부자 대량 주식매각으로 주가 급락...투자자 피해상장회사 임원, 주요주주 매매계획 사전공시 의무화.

<https://www.hani.co.kr> > arti > economy ▼

### '주식 먹튀' 논란 카카오 CEO 내정자 사퇴...“임원 스톡옵션 잔치 ...

2022. 1. 10. — 카카오의 차기 공동대표로 내정됐던 류영준 카카오페이 대표이사가 '주식 먹튀' 논란으로 사퇴했다. 그는 카카오 대표이사 내...

<https://www.seoul.co.kr> > news > newsView ▼

### '먹튀 논란' 카카오, 스톡옵션 이익 기업 1위...개인 5위권에 3 ...

2022. 10. 26. — 2020년부터 국내 주요 기업 89곳의 전·현직 임원들이 스톡옵션으로 벌어들인 이익은 1조원에 육박했다. '카카오 먹통 사태' 관련 플랫폼노동자 기자 ...

<https://byline.network> > 2022/01 ▼

### '먹튀 논란' 깜짝 놀란 카카오, 임원 주식 매도 규정 마련

2022. 1. 13. — 계열사인 카카오페이의 최고 경영진이 상장 한달만에 스톡옵션을 대량 매각, 시세차익을 크게 얻어 논란을 일으킨 카카오 그룹이 '임원 주식 매도 ...

<https://biz.chosun.com> > stock > finance > 2022/01/20 ▼

### 카카오페이 '먹튀' 논란 임원 일괄 사의 표명... 3명 물러난다

2022. 1. 20. — 카카오페이 먹튀 논란 임원 일괄 사의 표명 3명 물러난다 류영준·장기주·이진, 카카오페이서 물

# 프로젝트 목표

---

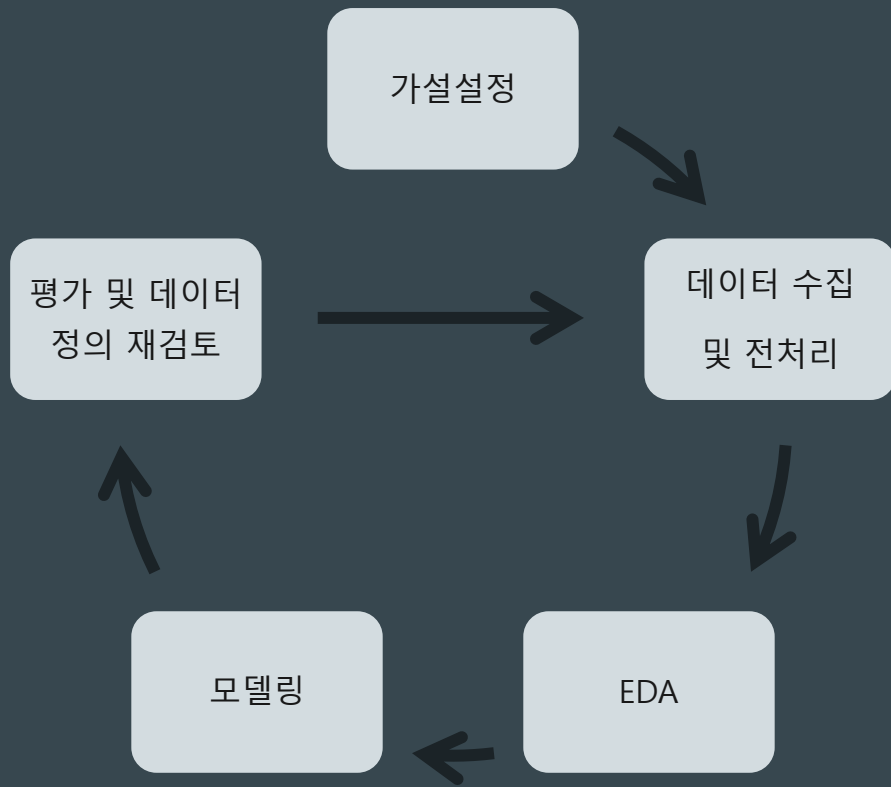
주요주주 및 임원의 지분공시 분석을 통한  
미래의 주식가격 추이 분류

# 분석 개요

# 분석 과정

---

- 프로토타이핑 접근법을 통한 가설검정
- 프로토타입을 이용하여 문제를 이해하고 이를 바탕으로 구체화.





## 가설

지분 공시 이벤트가 주식 가격에 영향을 미칠 것이다.

- 특정 회사에 대한 지분 공시 데이터로부터 지분 공시가 주가에 미치는 영향을 확인할 수 있을 것.
- 그 영향력을 파악함으로써, 지분 공시 데이터로부터 주가가 상승할 지, 하락할 지를 예측 / 판단할 수 있을 것.

# 분석 대상

---

대상 기업

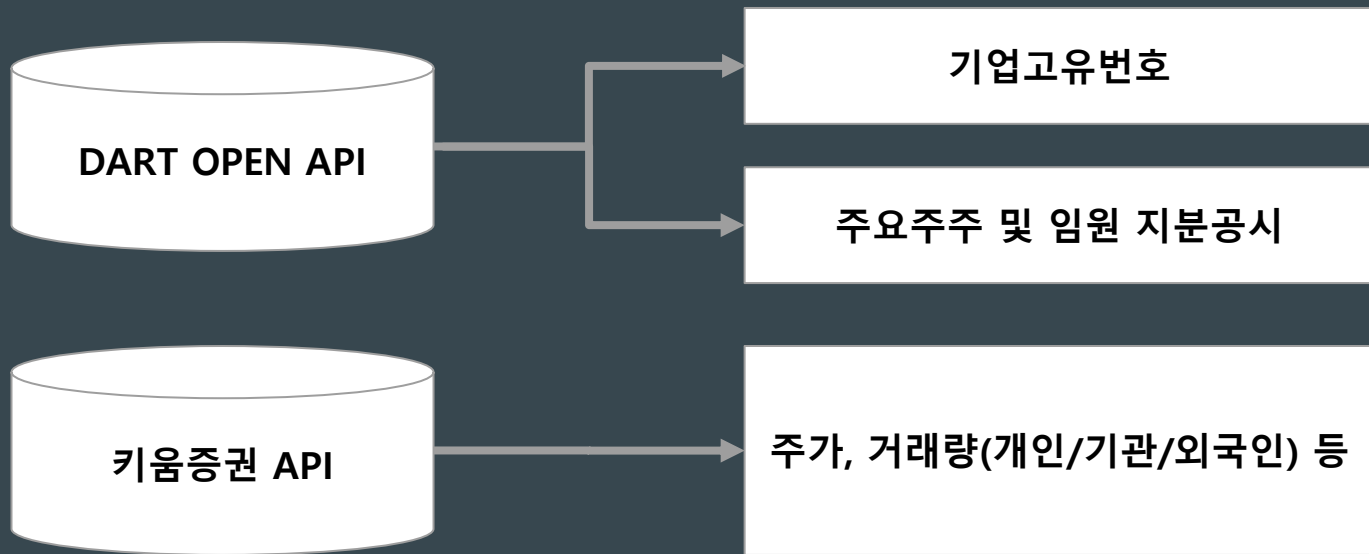
삼성전자

데이터 기간

2015.01.01  
~ 2022.11.11

# 분석 데이터 구성

---



# 수집한 데이터 셋

Index	Variables	Description	Data Type
당일 주가정보	종가	-	float64
	시가	-	float64
	고가	-	float64
	저가	-	float64
	전일비	전일 기준 최종가격 상승 및 하락 비율	float64
	등락률	현재주가 대비 전일 주가 증감율	float64
	거래량	당일 거래된 총 거래량	float64
	금액(백만)	당일 거래된 총 거래액	float64
	신용비	총 발행주식 대비 신용으로 매수된 비율	float64
거래주체 정보	개인	개인이 매매한 총 거래량	float64
	기관	기관이 매매한 총 거래량	float64
	외인수량	외국인이 매매한 총 거래량	float64
	외국계	외국계(헤지펀드 등)이 매매한 총 거래량	float64
	프로그램	프로그램 매매의 총 거래량	float64
	외인비	외국인 보유 비중	float64

# 수집한 데이터 셋

Variables	Description	Unique Value	Data Type
발생주식 총수	해당 회사의 발행 주식의 총 수	-	int64
보고구분	신규 / 변동	2	object
보고자 구분	개인 / 기관 / 외국인	3	object
임원	임원 종류 구분	3	object
직위명	직위 종류 구분	14	object
주요주주	대주주 / 임원 구분	3	object
변동사유	변동사유 상세구분	14	object
특정증권등의 종류	보통주 / 우선주 등	3	object
변동전	보고자의 이전 보유주식 수	-	int64
증감	보고자 주식 보유 변동	-	int64
변동후	보고자의 이후 보유주식 수	-	int64
보고일자	공시일	-	datetime

# 데이터 전처리

# 데이터 전처리

## 결측치 처리

- 액면분할 전, 3일 간의 거래정지로 인한 결측값.

변수	전처리 방법
당일개인거래비율	해당 row 제거
당일기관거래비율	해당 row 제거
당일프로그램거래비율	해당 row 제거

'대장주' 삼성전자의 주식 액면분할에 따른 거래정지 기간이 3거래일로 단축된다.

한국거래소는 삼성전자 액면분할에 따른 시장 영향을 최소화하기 위해 태스크포스(TF)를 구성해 논의한 결과, 올해 정기주주총회부터 액면분할 상장사의 주권 매매거래 정지 기간을 3거래일로 운영하기로 했다고 12일 밝혔다.

그동안은 통상 21일(15거래일)간 거래가 정지됐다.

# 데이터 전처리

## 각 변수 별 전처리

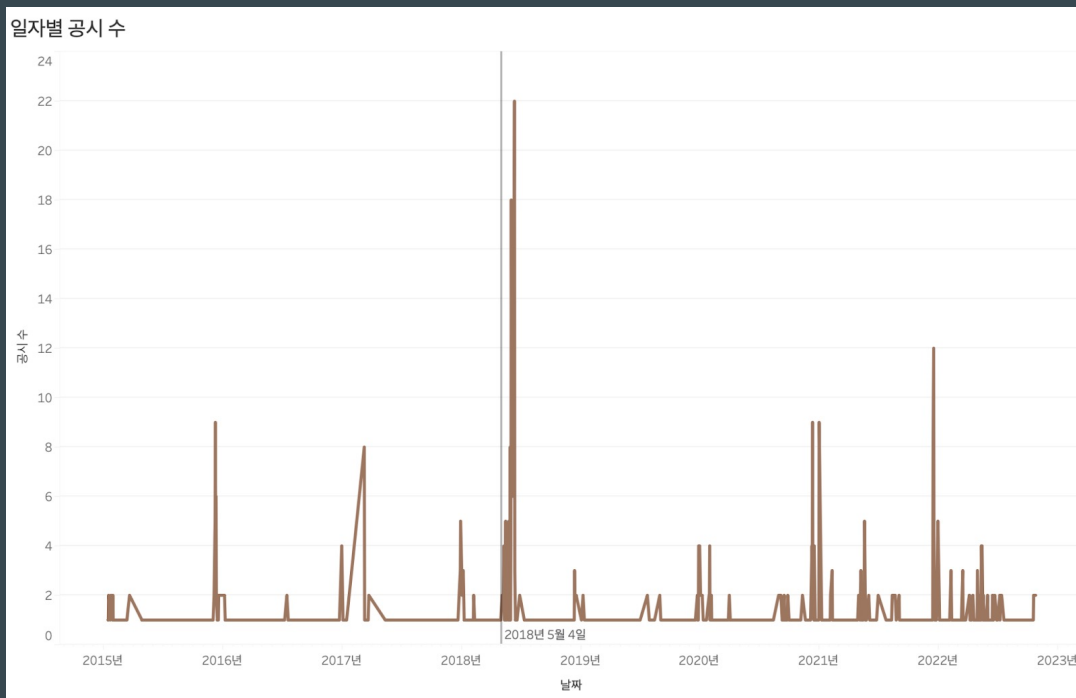
변수	전처리 방법	비고
종가	상승(3%이상) - 2 보합(0~3%) - 1 하락(0%이하) - 0	단기간 투자 시, 유의미한 투자 성과로서 3%를 설정
날짜	년 / 월 / 일 / 요일로 구분	년/월/일 및 요일에 의한 영향을 보기 위함.
직위명	임원진 / 대표 / 고문 / 없음으로 그룹화	동의를 그룹화
거래량 / 발행주식 총수	파생변수 생성	
당일개인거래비율	개인 / 거래량으로 변수 생성	당일 거래 참여자 중 개인비율
당일기관거래비율	기관 / 거래량으로 변수 생성	당일 거래 참여자 중 기관비율
당일프로그램거래비율	프로그램 / 거래량으로 변수 생성	당일 거래 참여자 중 프로그램비율
각 category 변수	One-hot 인코딩 적용	-

- 최종 데이터 형태: **399** rows X **74** columns



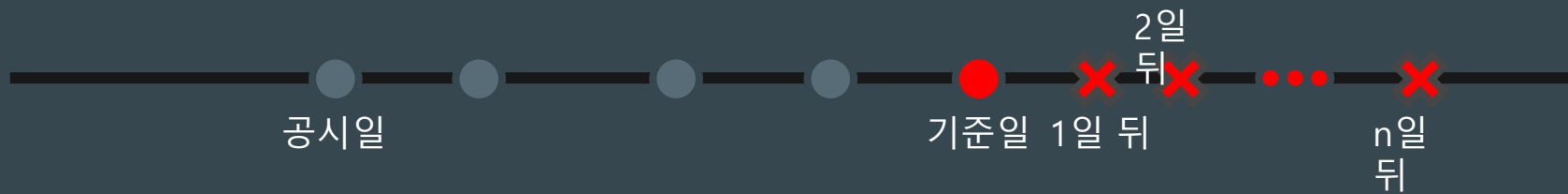
# 이상치 처리

- 특정 날짜에 공시가 몰리는 현상 발견.
- 2018년 6월 ... 삼성전자의 액면분할 이슈로 인한 공시로 인해 펌핑.

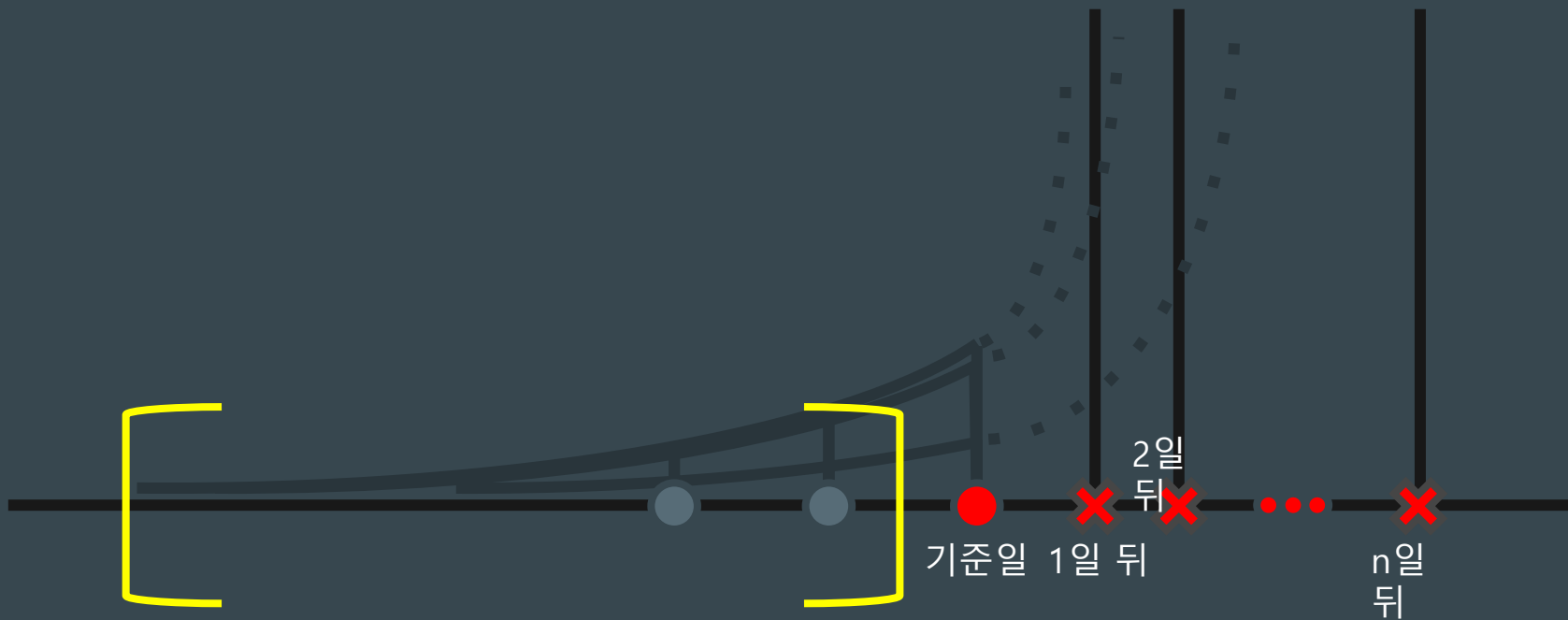


# 기간에 따른 가중치 적용

---

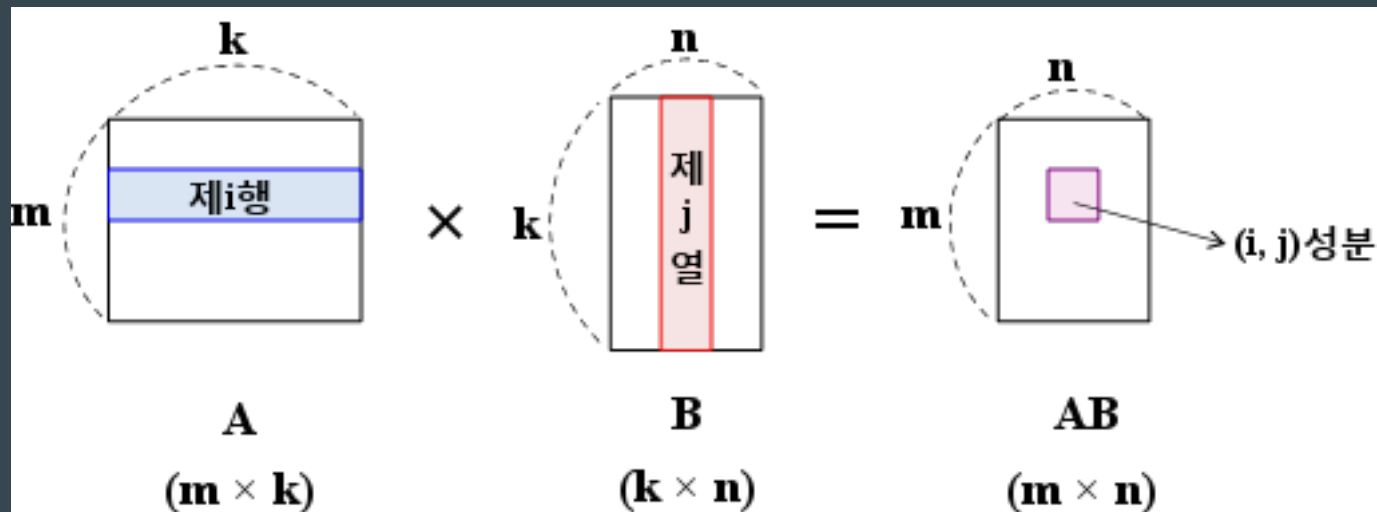


# 기간에 따른 가중치 적용



- 기준일 이전의 데이터의 영향력이 시간이 오래 지날수록 미치는 영향이 적도록 함,
- 가중치 =  $1 / (\text{예측 날짜} - \text{해당 날짜})$

## 기간에 따른 가중치 적용



# 기간에 따른 가중치 적용

- (예측일 - 공시일) 에 대한 역수를 가중치로 적용함.

날짜	2022-10-24	2022-10-18	2022-10-17	2022-09-29	2022-09-28	2022-09-21	2022-09-16	2022-09-08	2022-09-06	2022-08-17	...
1일 뒤											
2022-10-25	1.0	0.142857	0.125	0.038462	0.037037	0.000000	0.000000	0.000000	0.000000	0.0	...
2022-10-19	0.0	1.000000	0.500	0.050000	0.047619	0.035714	0.000000	0.000000	0.000000	0.0	...
2022-10-18	0.0	0.000000	1.000	0.052632	0.050000	0.037037	0.000000	0.000000	0.000000	0.0	...
2022-09-30	0.0	0.000000	0.000	1.000000	0.500000	0.111111	0.071429	0.045455	0.041667	0.0	...
2022-09-29	0.0	0.000000	0.000	0.000000	1.000000	0.125000	0.076923	0.047619	0.043478	0.0	...
...	...	...	...	...	...	...	...	...	...	...	...
2015-01-23	0.0	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	...
2015-01-21	0.0	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	...
2015-01-16	0.0	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	...
2015-01-15	0.0	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	...
2015-01-14	0.0	0.000000	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	...

399 rows × 399 columns

	날짜	시가	고가	저가	종가	전일비	등락률	거래량	금액(백만)	신용비	...
0	2022-10-24	57000.0	57500.0	56700.0	57500.0	1600.0	2.86	16944503.0	968892.0	0.15	...
1	2022-10-18	56700.0	57100.0	55800.0	56500.0	-100.0	-0.18	14820702.0	835745.0	0.14	...
2	2022-10-17	55800.0	57000.0	55700.0	56600.0	300.0	0.53	13641878.0	769638.0	0.14	...
3	2022-09-29	53300.0	53700.0	52600.0	52600.0	-300.0	-0.57	13882080.0	737731.0	0.17	...
4	2022-09-28	53900.0	54400.0	52500.0	52900.0	-1300.0	-2.40	19991129.0	1064666.0	0.17	...
...	...	...	...	...	...	...	...	...	...	...	...
394	2015-01-22	27960.0	27980.0	27560.0	27560.0	-17000.0	-1.22	17957700.0	499074.0	0.00	...
395	2015-01-20	27000.0	27440.0	26900.0	27440.0	29000.0	2.16	13278050.0	360703.0	0.00	...
396	2015-01-15	26900.0	26980.0	26580.0	26680.0	-11000.0	-0.82	14103900.0	378298.0	0.00	...
397	2015-01-14	26780.0	27100.0	26700.0	26900.0	6000.0	0.45	14332250.0	385455.0	0.00	...
398	2015-01-13	26280.0	26800.0	26000.0	26780.0	23000.0	1.75	12293400.0	324625.0	0.00	...

399 rows × 65 columns

## 기간에 따른 가중치 적용

- 기준일에서 특정 구간만 가중치를 적용하여 그 구간의 데이터만 고려함.
- 이번 프로토 타입에선 해당 구간을 30일로 결정.
- 30일에 대한 가중치를 온전히 받지 않는 행은 제거.

[illegible]

# 프로토 타입 모델링

# 프로토타입 모델링

## 평가지표 선정 – Accuracy

: 투자자들의 입장에서, 해당 공시로 인해 주가가 상승할 지, 하락할 지를 파악하고자 함.

- 상승할 것으로 예측한다면 매수를,
- 하락할 것으로 예측한다면, 원래 보유 중이던 주주들은 매도를 고려할 수 있음.
- TP와 FN 모두 중요하므로 Accuracy를 평가 지표로 선정.

표 1. 주가 경향 예측 연구 비교 분석

Ref.	Year	Metrics	Data composition information
[1]	2019	hit ratio, recall, specificity	not presented
[3]	2014	accuracy, precision	presented (entire data)
[6]	2019	F-score, sharpe ratio	not presented
[7]	2016	accuracy, return rate	not presented
[8]	2018	accuracy, return per trade	not presented
[9]	2018	accuracy, profit rate	not presented
[10]	2010	accuracy	presented (entire data)
[11]	2013	accuracy	presented (test)
[12]	2015	accuracy, F-score	presented (training, test)
[13]	2018	accuracy	presented (training, validation, test)
[14]	2019	accuracy, profit rate	presented (training, test)
[15]	2010	area under ROC	not presented
[16]	2010	accuracy	not presented
[17]	2011	accuracy, recall, specificity	not presented
[18]	2011	accuracy	not presented
[19]	2015	accuracy	not presented
[20]	2015	accuracy	not presented
[21]	2013	accuracy	not presented
[22]	2014	accuracy	not presented
[23]	2013	accuracy	not presented
[24]	2003	accuracy	not presented
[25]	2013	accuracy	not presented
[26]	2014	accuracy, recall, specificity	not presented
[27]	2012	accuracy, profit rate	not presented
[28]	2014	accuracy	not presented
[29]	2000	accuracy	not presented
[30]	2016	accuracy, recall, specificity, sharpe ratio	not presented
[31]	2009	accuracy	not presented
[32]	2011	accuracy	not presented
[33]	2016	accuracy, recall, precision, F-score, joint prediction error	not presented

참고자료 ) 주가 경향 예측 모델의 공정한 성능 평가 방법, 임정수, 한국교통대학교 전자공학과, 한국콘텐츠학회논문지 '20 Vol. 20 No. 10,, 2020.08.25



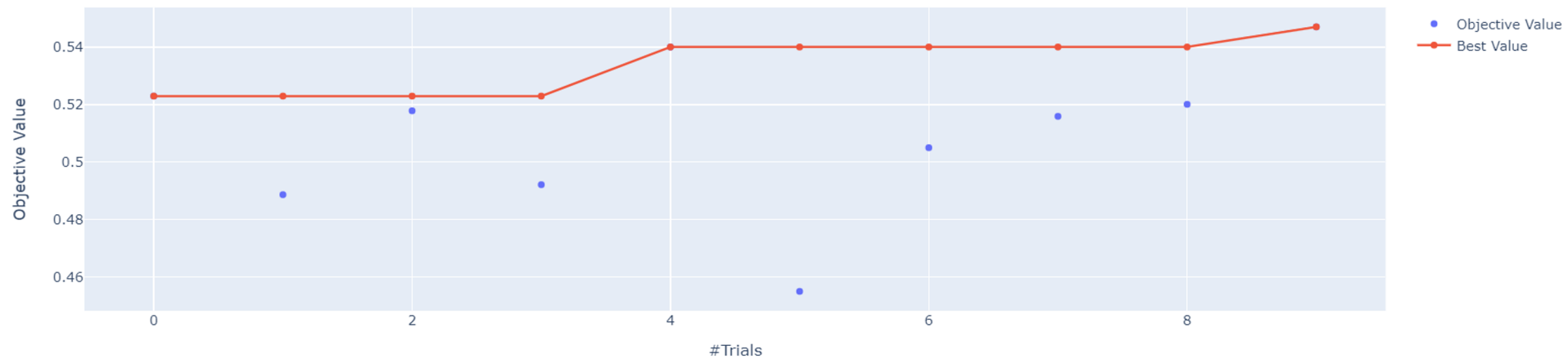
## 프로토타입 모델링

- LightGBM을 이용하여 빠르게 모델링.
- 9일 뒤를 기준으로 진행.
- MinMax / Robust Scaling 한 데이터에 대해서 진행했을 때, 더 낮게 나옴.
- Train\_test\_split을 이용하여 train / test 분리 (8:2)
- Optuna를 통해 HPT 진행.
- LightGBM 결과,
  - Train set에 대해 0.54의 Accuracy
  - Test set에 대해 0.51의 Accuracy

LGBM 평균 ACC(1일 뒤 기준):	0.49232462877624167
LGBM 평균 ACC(2일 뒤 기준):	0.46373783922171014
LGBM 평균 ACC(3일 뒤 기준):	0.48140296979006647
LGBM 평균 ACC(4일 뒤 기준):	0.5084895033282131
LGBM 평균 ACC(5일 뒤 기준):	0.5127752176139273
LGBM 평균 ACC(6일 뒤 기준):	0.514726062467998
LGBM 평균 ACC(7일 뒤 기준):	0.4961802355350742
LGBM 평균 ACC(8일 뒤 기준):	0.5054275473630313
LGBM 평균 ACC(9일 뒤 기준):	0.556415770609319
LGBM 평균 ACC(10일 뒤 기준):	0.5298771121351766

# 프로토타입 모델링

Optimization History Plot



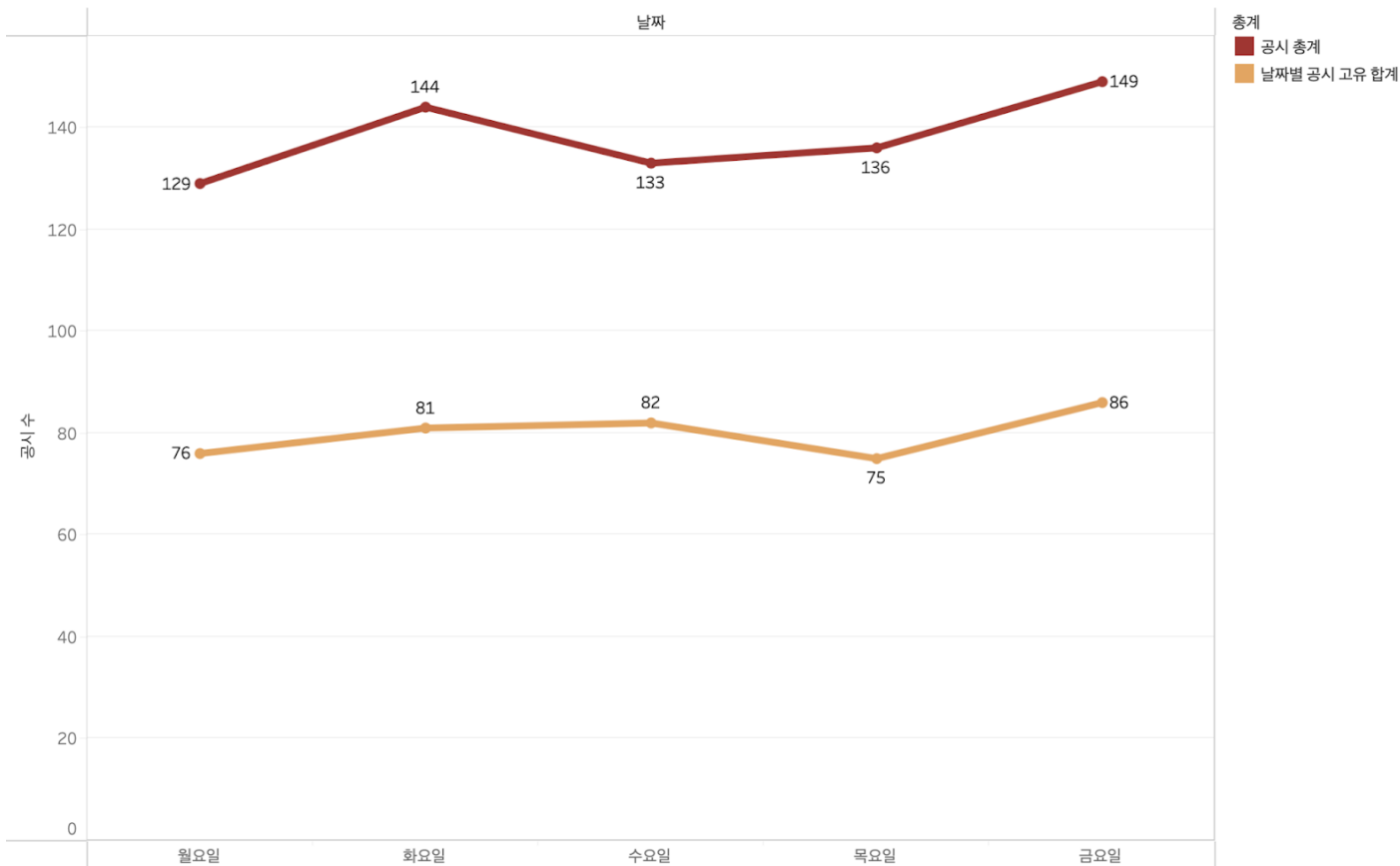
변수 선정

# 트리기법 변수선택

	SelectFromModel	Permutation Importance	최종	구분
1	변동후	신용비	변동후	
2	전일비	월	증감	
3	등락률	요일_월	전일비	
4	변동전	직위명_임원진	등락률	
5	변동사유_장내매도(-)	직위명_대표	신용비	
6	보고구분_변동	특정증권등의 종류_우선주	요일_월	
7	개인	당일개인거래비율	요일_금	
8	보고구분_신규	변동사유_장내매수(+)	월	
9	요일_화	변동사유_신규보고(+)	변동사유_장내매도(-)	
10	당일기관거래비율	거래량/발행주식 총수	변동사유_장내매수(+)	
11	보고자 구분_개인(국내)	외국계	변동사유_신규보고(+)	
12	발행주식 총수	증감	보고자 구분_개인(국내)	
13	직위명_대표	임원_등기임원	거래량/발행주식 총수	
14	기관	고가	당일개인거래비율	
15	거래량	요일_금	당일기관거래비율	
16	임원_등기임원	개인	당일프로그램거래비율	
17	외인비	보고자 구분_개인(국내)	보고구분_변동	
18	특정증권등의 종류_보통주	주요주주_-	보고구분_신규	
19	당일프로그램거래비율	당일기관거래비율	특정증권등의 종류_우선주	
20	신용비	변동사유_기타(+)	특정증권등의 종류_보통주	

# Appendix) 변수선택 판단 근거 - 요일

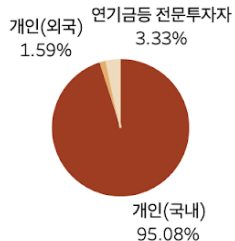
공시 요일



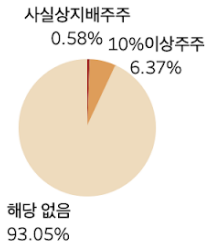
EDA

# 보고자 유형별 공시 수

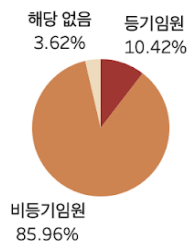
보고자 구분



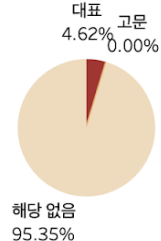
주요주주 구분



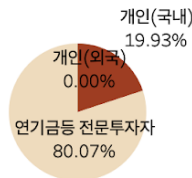
임원 구분



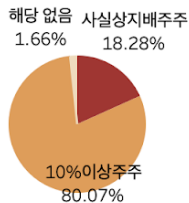
직위명 구분 증감



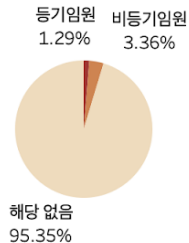
보고자 구분 증감



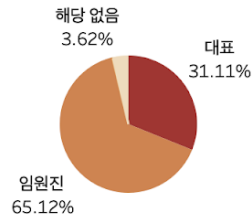
주요주주 구분 증감



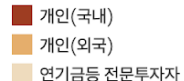
임원 구분 증감



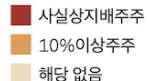
직위명 구분



보고자 구분



주요주주



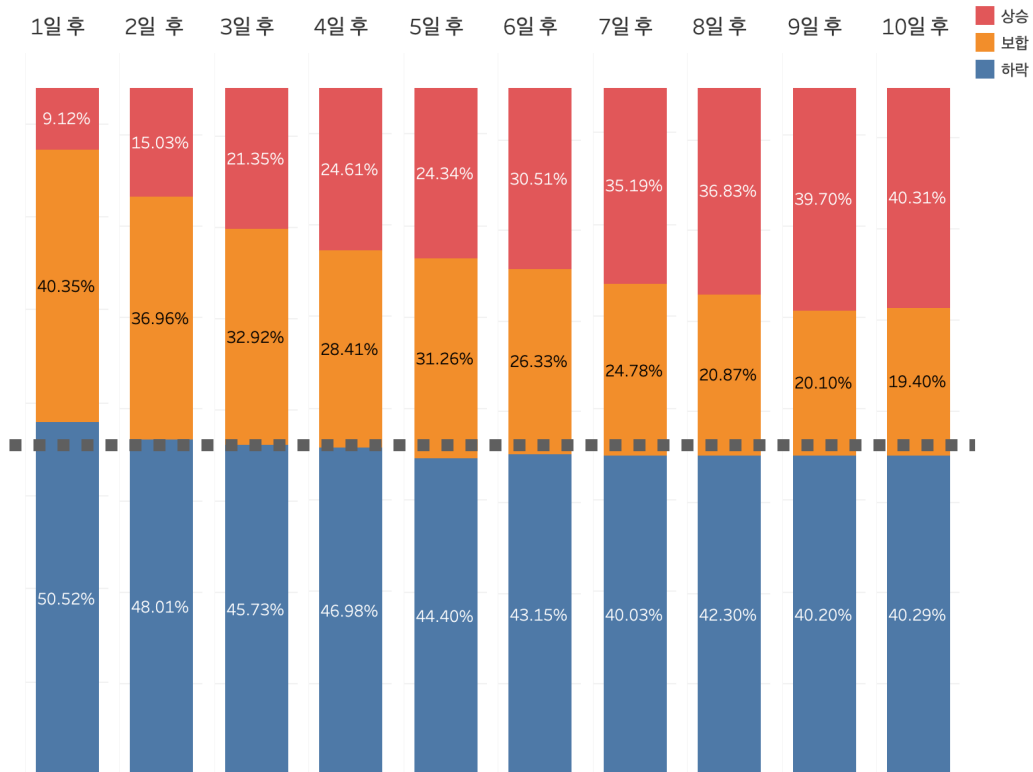
임원



직위명

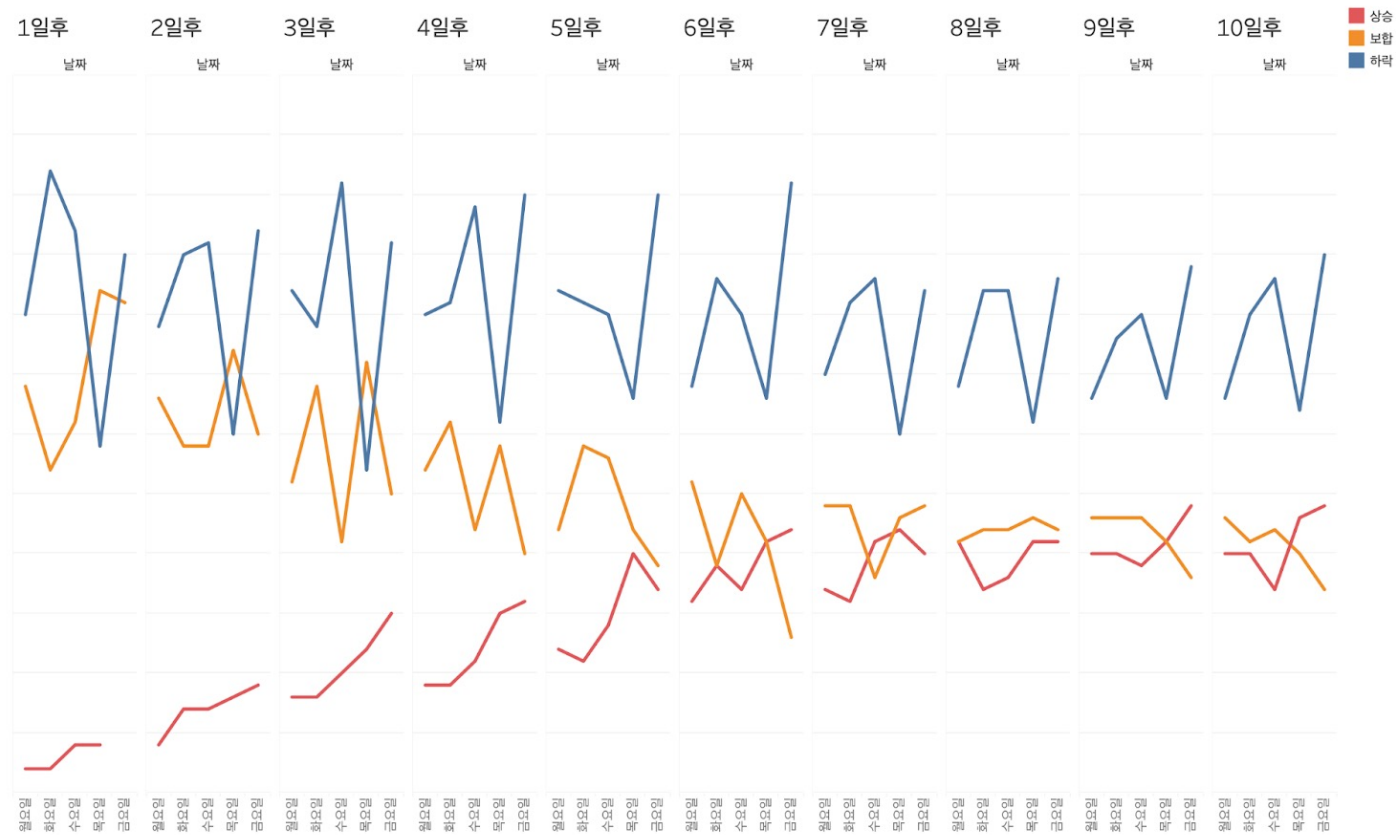


# 시기 별 증가 증감 현황



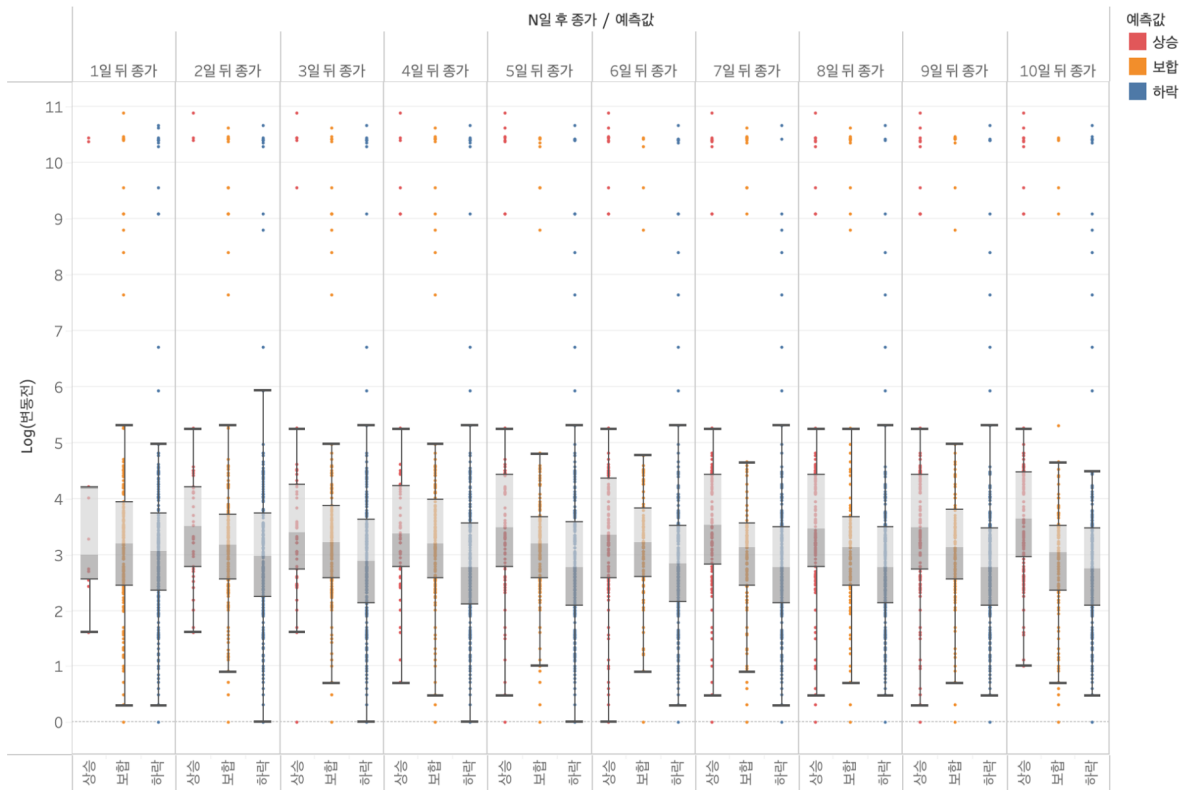


## 요일별 N일 후 증가 차이



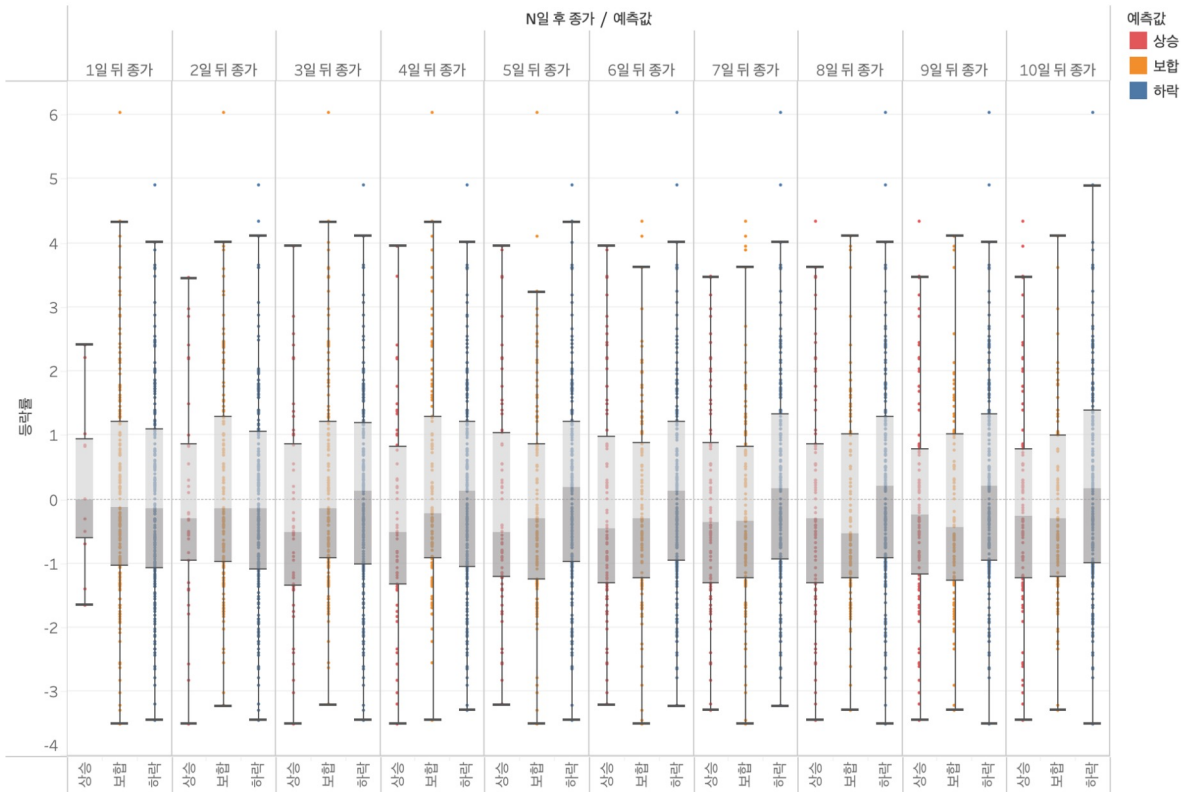
# 변동전에 따른 결과 추이

변동전



# 등락률에 따른 주가 추이

등락률



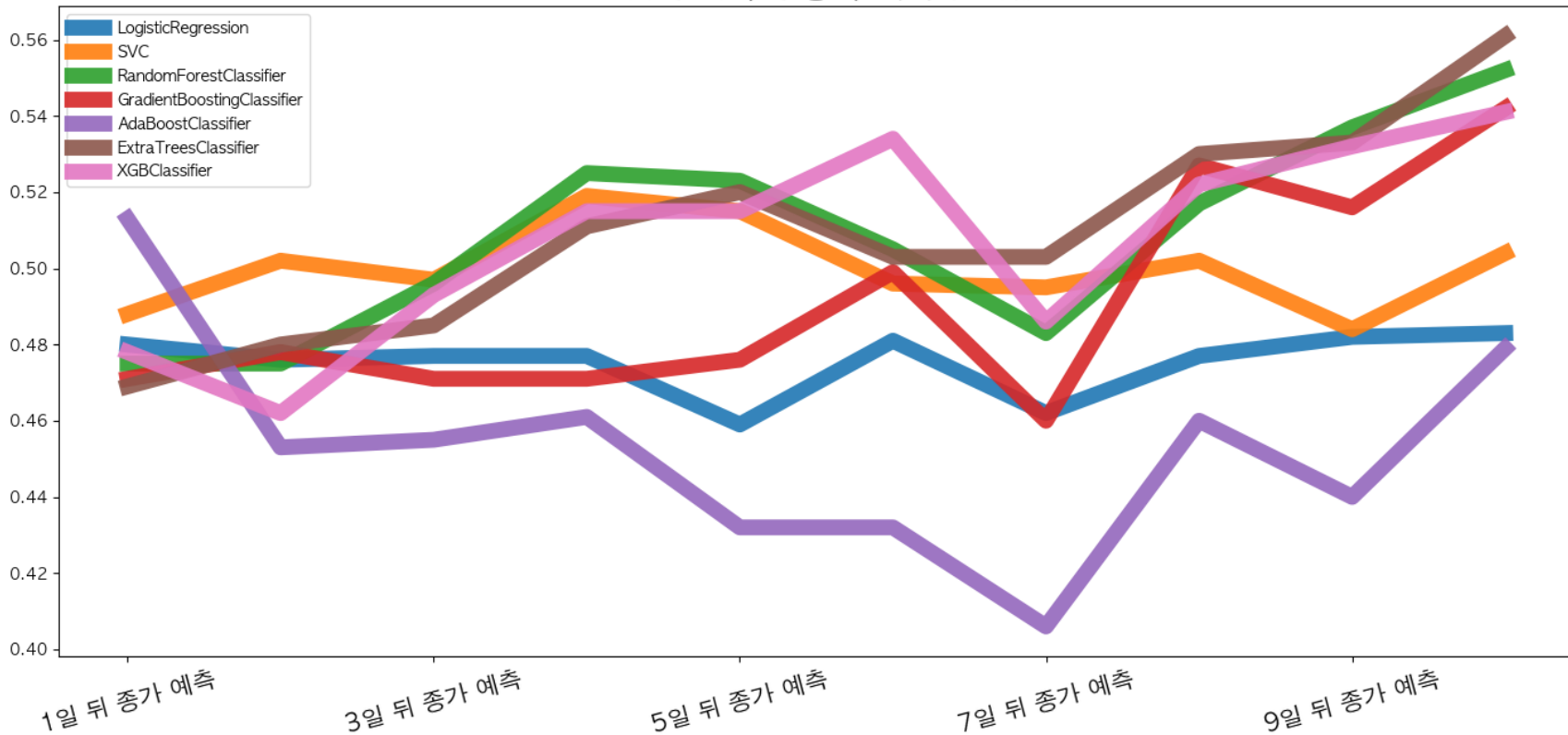
# 모델링 결과

## 모델링 결과

학습한 모델	정확도			
LightGBM	0.59			
XGBoost_1	0.53			
XGBoost_2	0.63			
XGBoost_3	0.54			
Logistic Regression	0.54			
Random Forest_1	0.35			
Random Forest_2	0.59			
KNN	0.54			
SVM_1	0.54			
SVM_2	0.60			

# 모델별 종가 예측 결과값

모델별 일자별 종가 예측 결과값



# 결론

# 결론

---

## 결과

- Accuracy가 0.6 대 이므로, 주요 주주 및 임원들의 지분공시가 주가 변동과 연관이 있음.

## 한계점

- 특정 한 기업에만 맞춰진 모델이므로, 다른 기업들에도 적용되는지 파악이 필요.
- 액면분할 이슈에 대해 상관성을 파악하지 않음.
- 추후 발생할 실제 공시에 대해 테스트 필요.
- 다른 MACRO적인 요인에 대해 고려하여 좀 더 심도 높은 EDA 필요.
- 목적을 바꾸어서, 다른 평가 지표에 대해서도 고려해볼직 함. (Recall / Precision / F1-Score)



감사합니다.