

**김정훈 김민정 김상태 박정현 이윤희**



# 목차



1. 지난 주 리뷰



2. 추가 변수 수집



3. 감성분석



4. 모델링 및 결과



5. 한계 및 의의



# 추가 변수 수집

07

거시경제 변수



엔터 3사



환율



감성 score

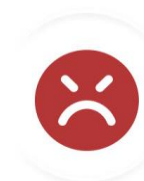


유가



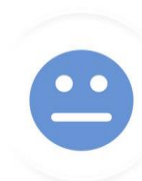
상품 선물

## Sentiment Analysis



**Negative**

I'm dissatisfied with your customer service.  
No one was able to help me with the  
problems I had with using your product.



**Neutral**

The product has multiple features  
that are suitable for users with different  
levels of experience.



**Positive**

I really enjoy how easy this product  
is to use and how it successfully helps  
my team complete their day-to-day tasks.



당연히 고려해야지.  
주가 예측을 위해서라면 말이야...





# 감성 분석



## 감성 분석이란?

**#텍스트 마이닝의 한 종류**

**#어떤 주제에 대한 감정, 인상, 태도 등  
개인의 의견들을 분석하는 기법**

**#주로 긍정, 부정 **이진형식****





# 감성 분석

11



## 감성 분석 결과

### #감성 점수 변수

#주가가 오르면 1, 아니면 0

#주가가 오를 확률(1 일 확률) 산출

#파생 변수로 사용

#이거 하나 만드려고 무슨 고생을.....

	date	sm_title	sm_content	SM_y	sm_score
0	2014-01-01	새해,매치,동방신기,퍼포먼스,강자,이승기,열애,윤아,특파원,칼럼,아베,정령,동방신기...	큐브,DC,매치,다그,비와,레이니,미로,Mirotic,매치,비와,장르,대중,작사,작...	0	0.911425
1	2014-01-02	이승기,열애,소녀시대,윤아,이승기,열애,연애,테이프,윤아,소녀시대,공개,빅스타,대전...	배우,이승기,열애,이승기,후크엔터테인먼트,이승,기씨,분이,연애,죽도,열애,이승기,꽃...	1	0.973506
2	2014-01-03	열애,정경호,수영,아버지,반응,한밤,열애,소식,농구랑이,수영,윤아,완전,정경호,열애...	팀강,아리,배우,정경호,열애,소식,해진,정경호,인터뷰,연애,인생은아름다워,엄마가뵈났...	1	0.997220
3	2014-01-05	토론토,방집,미로,외국인,내비게이션,모두,문화,제공	캐나다,토론토,방집,미로,Mirotic,스페인,농구,강남,스타일,강남,스타일,해달,...	1	0.927656
4	2014-01-06	조권,이특,내세,애도,2AM,2PM,찬성,물결,조부모,이특,유키스,수현,내세,부친상...	연예,조권,조부모,이특,박정수,애도,트위터,이특,내세,이특,이특,윤인,이특,현석,경...	1	0.979145
5	2014-01-07	박근혜,혁신,하이슈,대통령,강조,이특,치매,비극,슈퍼주니어,가족,신비주의,이특,행렬...	이특,조부모,박근혜,기자회견,박근혜,창조경제,청와대,기자회견,부처,창조경제,창조경제...	0	0.133000
6	2014-01-08	국민연금,모드,LG,패션,매수,러브콜,한국,패션,대부,신우,모델,스토리,국민연금공단...	분기,국민연금,국민연금,분기,패션,하나투어,한미약품,LS산전,LG이노텍,국민연금,국...	0	0.044628
7	2014-01-09	아우디,LG,탄성,삼성,연발,정시,실용음악,학과,서울예술전문학교,눈길,학년,경쟁,이...	플뢰르멜르랑,한국,프랑스,루퍼트,독일,아우디,글로벌,유명,LG전자,중소기업,혁신,미...	0	0.972662
8	2014-01-10	의장,대박,불어,주식,프리미어,클래식,영국,패밀리,예술의전당,큐브,SAC,보노,믹스...	진의원,이사회,의장,주식,가치,클럽,주식,부자,의장,의장,순위,메신저,라인,의장,이...	1	0.815603
9	2014-01-11	8월의크리스마스,한석규,심은하,영화,연극	자회사,연극,예술의전당,목요일,재현,연출,영화,8월의크리스마스,제작,무대,연극,제작	1	0.537733

갯세끼

상태 심정

5



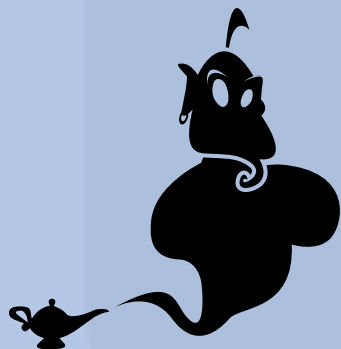


# 사용한 모델

시계열 모델

**LSTM**

**ARIMAX**



- 원하는 거 뭐든 말만 해~

머신러닝 모델

**CatBoost**

**LGBM**

**Random Forest**

**XGBoost**

**NGBoost**



# ARIMAX

시계열 + 선형회귀모델

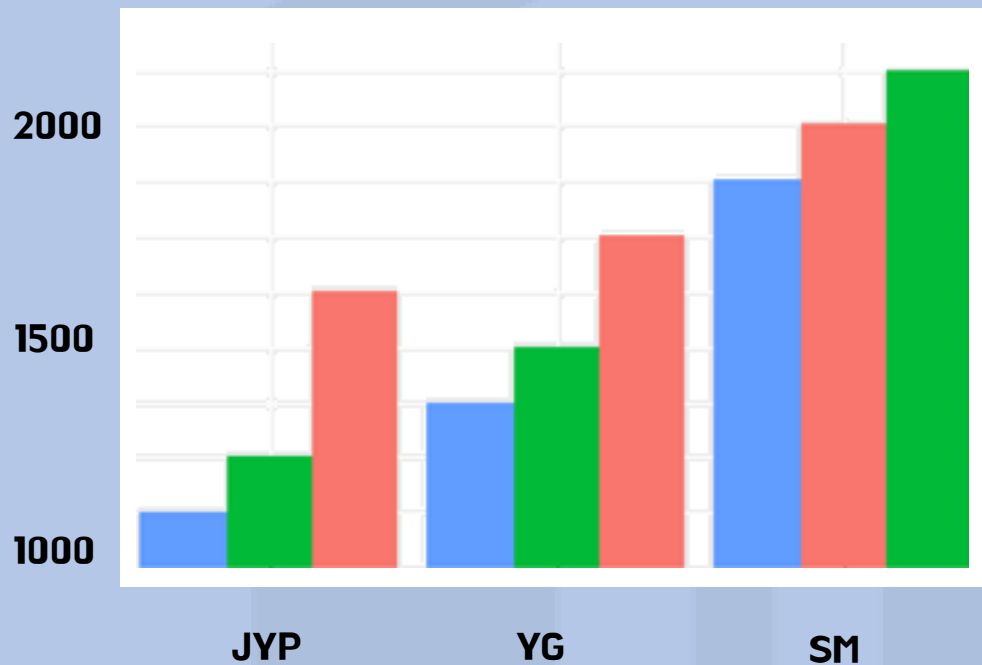
OUTLINE

MODELING

▶ **OUTPUT**

**OUTPUT**

: RMSE



- 거시 지표
- 거시 지표 + 감성 점수
- 거시 지표 + 감성 점수 + 아티스트

거시 지표 + 감성 점수 + 아티스트 관련 변수  
**FULL MODEL 이 BEST 이다!**



# CATBOOST

OUTLINE

▶ **MODELING**

OUTPUT



## CATBOOST

**엔터 3사**



SM  
ENTERTAINMENT



ENTERTAINMENT



ENTERTAINMENT

LEADER IN ENTERTAINMENT



**변수 추가하여 모델 적합**



거시적 지표



감성 점수



아티스트



거시적 지표



감성 점수



거시적 지표



**튜닝 & 학습**



파라미터 튜닝



모델 학습





# LGBM

## ▶ OUTLINE

MODELING

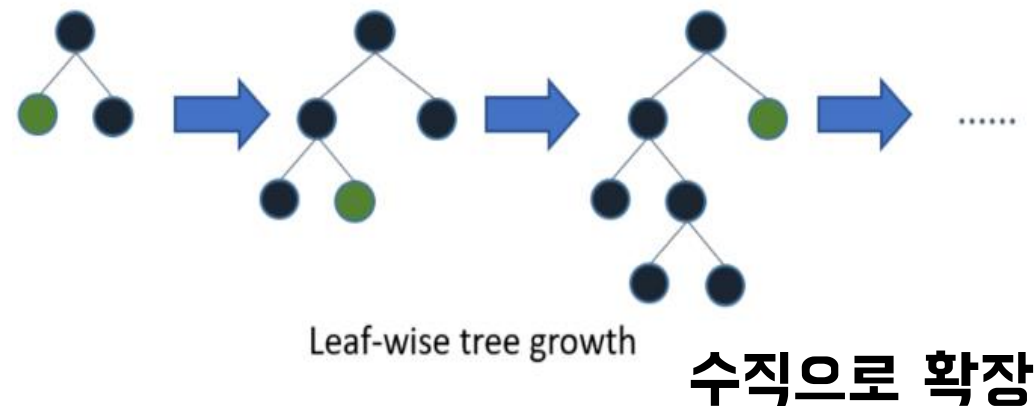
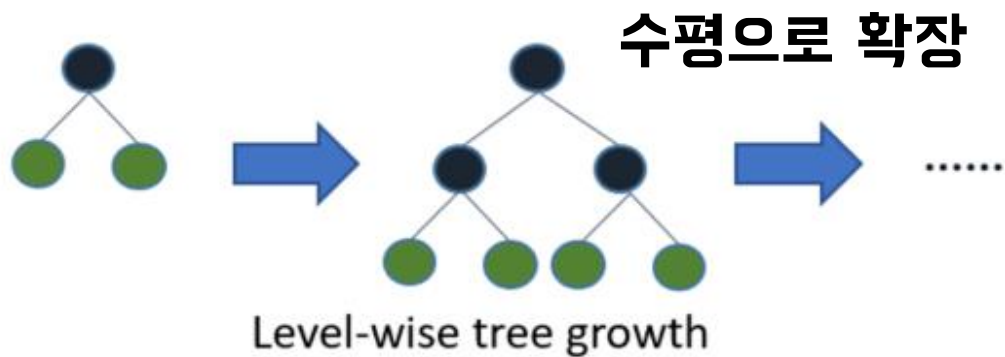
OUTPUT

## LGBM

Light GBM : 트리기반의 학습 알고리즘인 gradient boosting 방식의 프레임 워크

다른 트리기반 알고리즘

LGBM 알고리즘





# 랜덤포레스트

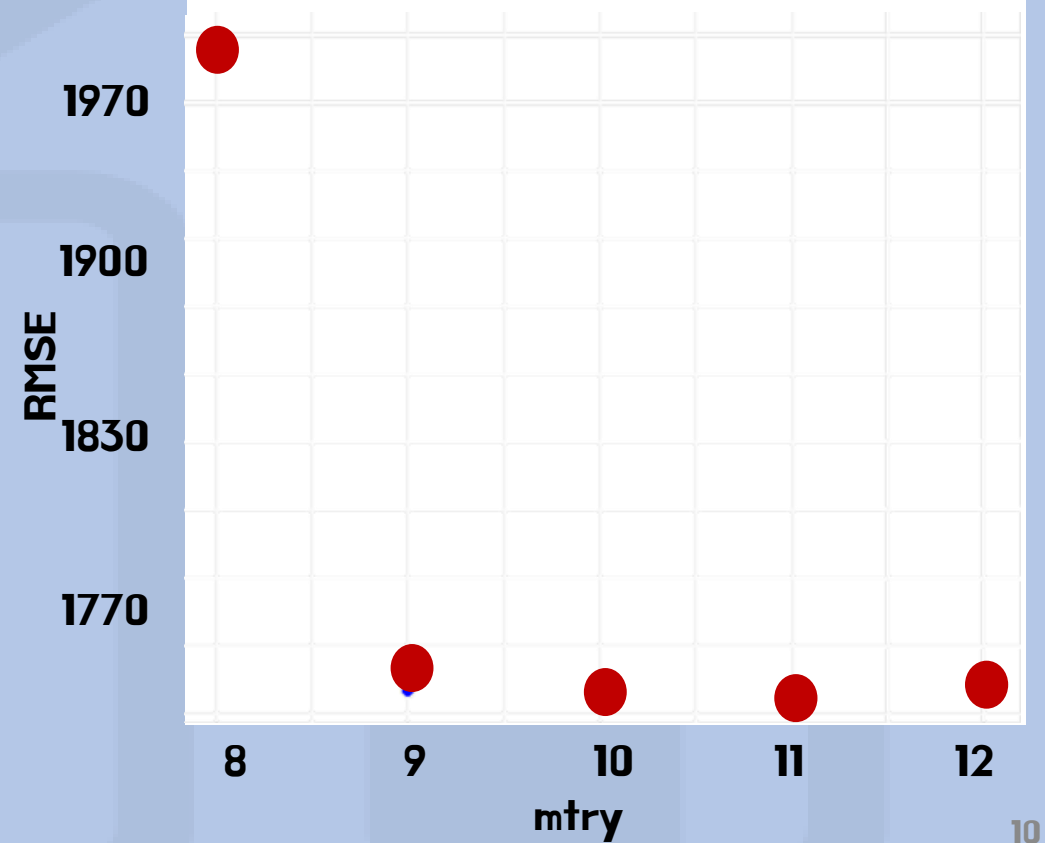
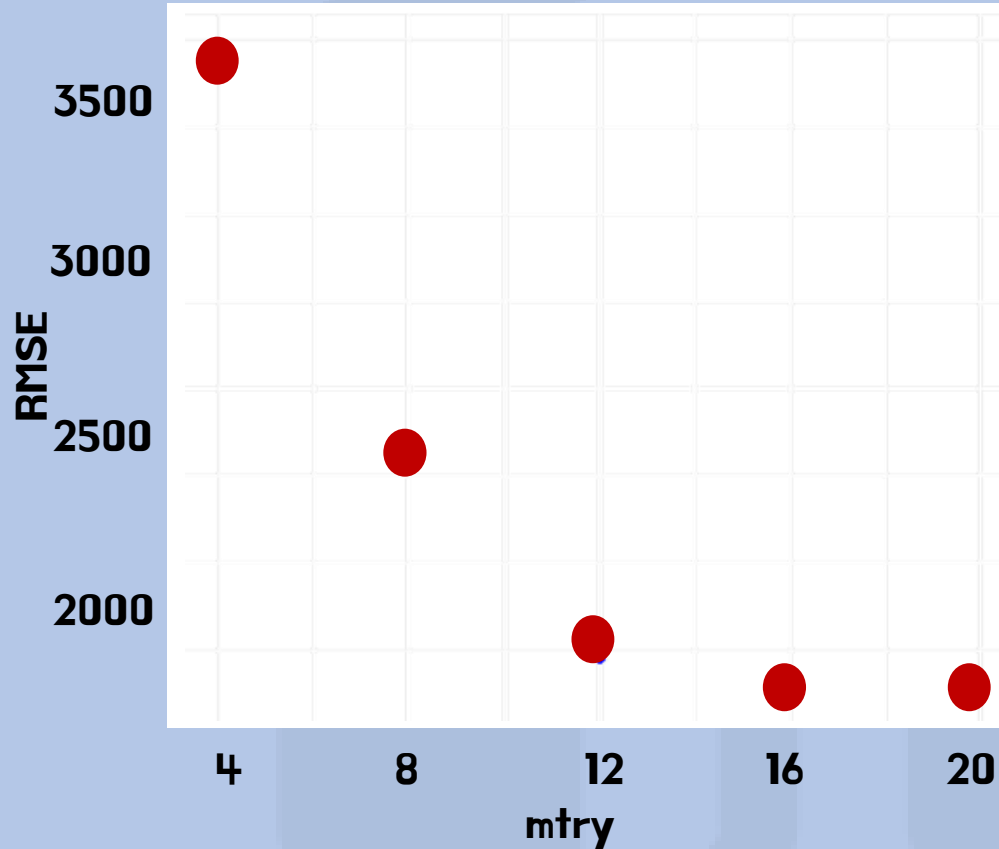
OUTLINE

▶ **MODELING**

OUTPUT

## TUNING

: 최적의 MTRY 값 찾기





# XGBOOST

## ▶ OUTLINE

MODELING

OUTPUT

Model의 residual에 새로운 트리를 fit

이 과정에서 **Gradient Descent**를 이용해  
Loss function(rmse, classification error) 최적화

$$y = f(x) + \varepsilon_1 \text{ Iteration 1}$$

$g(x)$  :  $\varepsilon_1$ 에 적합시킨 모델

$h(x)$  :  $\varepsilon_2$ 에 적합시킨 모델

$$= f(x) + g(x) + \varepsilon_2 \text{ Iteration 2}$$

$$= f(x) + g(x) + h(x) + \varepsilon_3 \text{ Iteration 3}$$



# NGBOOST

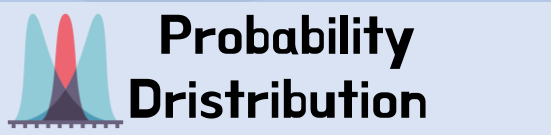
OUTLINE

▶ **MODELING**

OUTPUT

NGBOOST 튜닝

**다른 트리 모델과 다른 부분!**



**Regression**



Normal, LogNormal,  
Exponential

**Classification**



Multinomial, Bernouli



**Score Rule**

**MLE**



(Maximum Likelihood Estimation)

**CPRS**



(Continuous Ranked Probability Score)



**Decision Tree**

depth, colsamples 등  
기존 트리 모델 parameter



# 모델 결과비교(JYP)

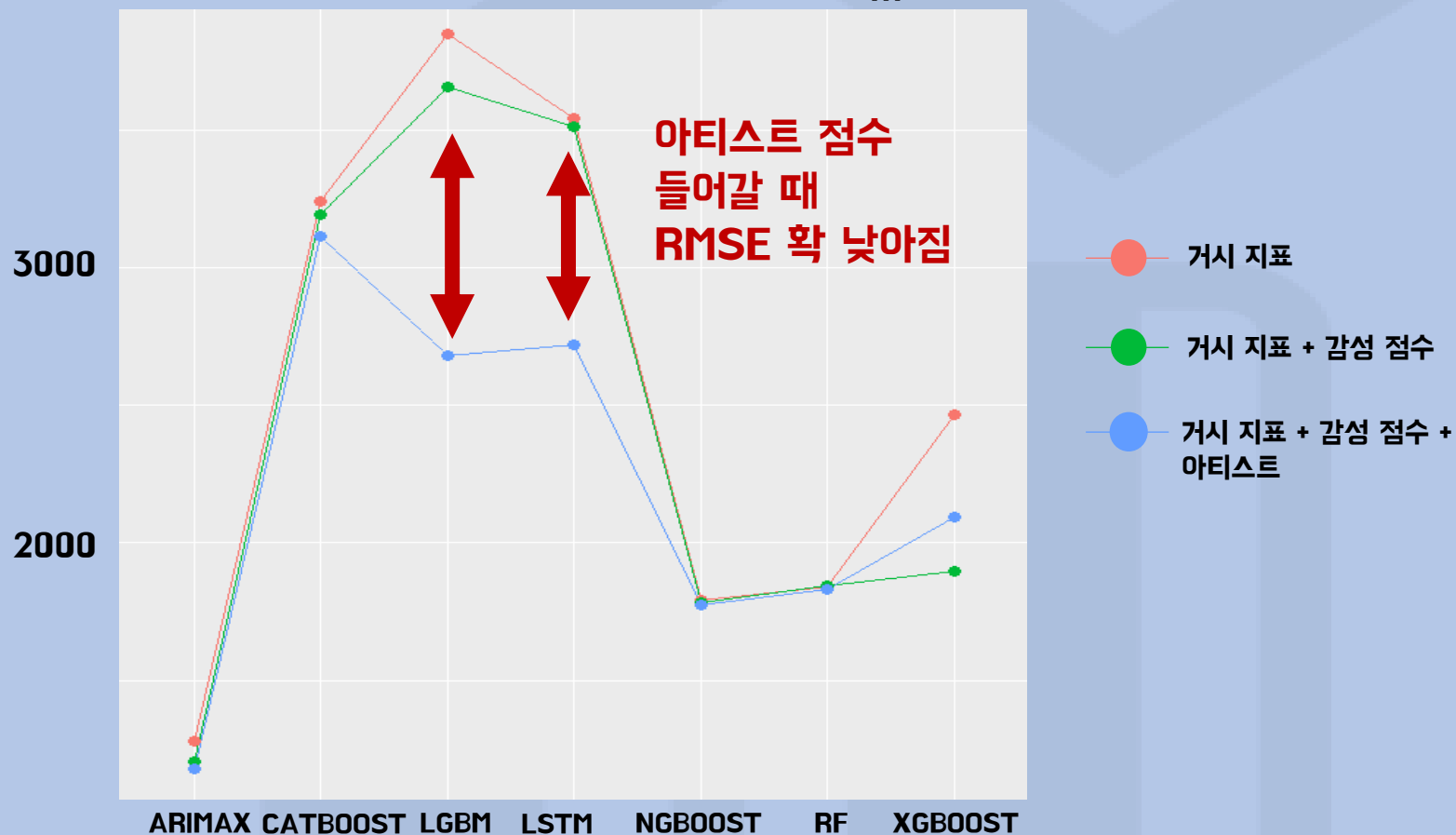
항목	NG Boost	Cat Boost	LSTM	ARIMAX	Random Forest	LGBM	XG Boost
거시	787.06	833.87	3896.70	1265.01	752.20	783.76	759.97
거시 + 갠성점수	735.39	750.74	2997.88	1173.12	749.67	803.72	751.42
거시 + 갠성점수 + 아티스트점수	730.49	664.81	2294.18	1168.38	745.19	668.59	672.01

(RMSE 기준)



# 모델 결과비교(SM)

#주식 가격의 큰 변동 #큰 RMSE #흠...



변동이 컸지만  
그래도 SM 먹여살리는 건  
우리 가수들 뿐...





# 한계 및 의의



P-SAT 최초로 주식 데이터를 다룸



각자 모델 최소 1개씩 맡아서 Flex함



열심히 설계한 파생변수들이 다 의미있음



근데 주식 어렵다... **가슴**으로 이해하자^^