

최악의 미세먼지 원인, 중국일까 한국일까…

HOME > 오피니언 > 대장간 > 철강

포스코가 미세먼지 발생 주범 웬 말?

미세먼지의 원인이 한국의 철강산업이라면,
미세먼지 수치가 철강 기업의 실적을 예측할 수 있지 않을까?

미세먼지의 원인이 한국에 있다는 이견이 제시되면서, 국내 자체에서 그 원인을 찾으려도 시도가 지속적으로 이뤄져 왔다. 그리고 가장 주된 원인으로 철강 산업이 지목되었다. 만약 철강 산업이 한국의 미세먼지의 주 원인이라면, 그것은 철강의 '생산' 과정에서 나올 것이다. 그렇다면

미세먼지가 비정상적으로 높다는 것은 철강 기업들이 수요에 맞춰 생산량을 기준보다 증가시켰다는 신호이고, 미래의 실적 증가를 기대할 수 있다. 즉, 미세 먼지 정보가 철강 기업의 사정을 미리 알 수 있는 내부자 정보로써 역할을 한다는 것이다.

1. 알고리즘

미세먼지 데이터로 철강산업의 2일 뒤 종가를 예측하고, 예측 가격이 현재 종가보다 높으면 매수

【Input Data】

철강산업에 영향을 줄 것으로 기대되는 요소들



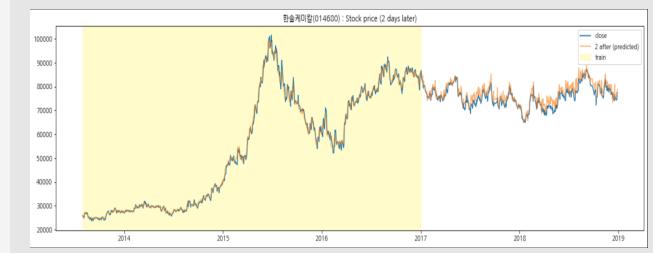
【Model】

Predict Close (2 days later)

dmlc
XGBoost

【Predicted Price】

Predicted Close (2 days later)



2일뒤 예측 가격 > 종가 → 매수

매수로부터 2일 뒤 → 매도



'예측해야 할 N일뒤의 가격'을 Hyperparameter로 놓고 Training Set에서 Bayesian Optimization 수행 결과, 2일의 예측력이 가장 높았음

2. 사용 데이터

1

기상청 서울특별시 대기정보 데이터



| 기상정보 | 대기정보 |
|------|------|
| 평균기온 | |
| 최고기온 | |
| 일강수량 | |
| 평균풍속 | |
| 최다풍향 | 미세먼지 |
| 습도 | |
| 일조시간 | |
| 가조시간 | |
| 평균운량 | |

2

배달음식 주문량 데이터

18년 9월 서울시 치킨 판매업종 이용 통화량
서울지역 치킨집에 대한 2018년 9월 한달 통화량을 기반으로 지역별 이용 현황 데이터를 제공합니다.

다운로드 수: 77

▶서울지역 치킨집에 대한 한달 통화량을 기반으로 지역별 이용 데이터를 제공합니다.
▶SK 텔레콤의 통화량 데이터와 일종 데이터를 기반으로 추출한 데이터로 전체 치킨집 서비스 이용 현황이 활용하시기 바랍니다.
▶본데이터는 월 1회 정기적으로 업데이트 됩니다.



SK빅데이터허브 전화주문량

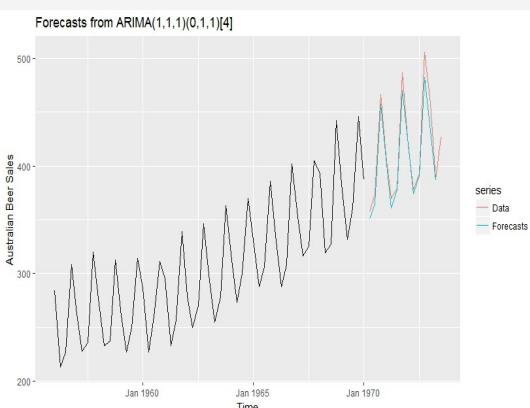
SK빅데이터허브 제공
치킨, 피자, 족발, 중국음식
전화주문량 데이터

농림축산식품부
국내 외식트렌드 조사 보고서

월별 배달 서비스 어플리케이션 점유율

3

시계열 변수 데이터



아래와 같은 요소 반영
Seasonality
• 월
Trend
• Dayofyear, 날짜
Cycle
• 요일, 날짜

4 KOSPI 종가 데이터



KOSPI 종가 데이터

Training Period
상장일 ~ 2016.12.31

Validation Period
2017.01.01 ~ 2018.12.31

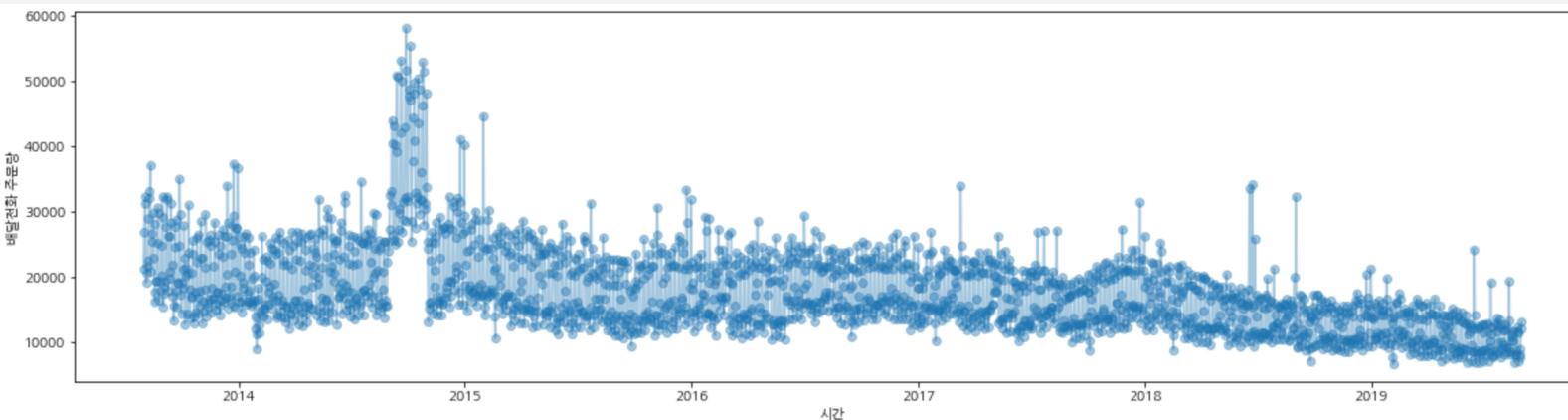
전체 KOSPI를 대상으로 하는 것이
아닌 미세먼지에 영향을 받을 것으로
예상되는 기업 선정하여 분석

3. 데이터 선정 이유

철강산업과 배달 주문데이터의 관계

Investement Sentiment 반영

- 주가를 예측하기 위해서는 해당 기업실적과 같은 객관적인 지표도 물론 중요하지만 전반적인 시장의 investment sentiment를 함께 고려해야 함
- 기업 실적이 좋아도 시장이 얼어있다면 주가 상승 폭이 크지 않을 수 있기 때문
- 전반적인 소비자와 시장의 Sentiment를 반영하는 요소가 배달음식 주문량 사용
- 배달 서비스의 거래액이 매 분기 약 100퍼센트 상승하는 추세를 반영
- 배달 주문 데이터가 전반적인 소비자 심리를 반영하는 합리적인 데이터라고 판단해 변수로 선정



- 2014년 아시안게임 특수를 제외하고는 배달음식 전화주문량은 계속 하락하는 추세
- 배달어플의 점유율이 늘어감에 따라 전체 배달 음식 시장은 커지고 있으나, 전화주문량은 하락함
- Sentiment를 반영하는 지표로써 활용하기 위해 이를 보정할 추가적인 작업 필요

배달앱, 아게임 특수 있다…'배달통' 모바일 주문량 28% 증가 머니S | 2014.09.24. | 네이버뉴스 |

모바일결제 메뉴별 주문량을 살펴보면 대표적인 응원메뉴인 '치킨'이 전체 주문량의 38%로 1위를 차지했고, 2위에는 '한식/분식'이 21%로 뒤를 이었으며, 3위는 17%로 '중식'이 차지했다. 주문 증가와 함께...

연일 들려오는 금메달 소식과 한국 선수단의 선전 속을 지켜보면서 치킨, 피자 등을 배달시키는 사례가 급증하면서 '배달통', '배달의 민족', '요기요' 등 배달앱 업체들은 주문량이 급증하는 아시안게임 '특수'를 톡톡히 누리고 있다.

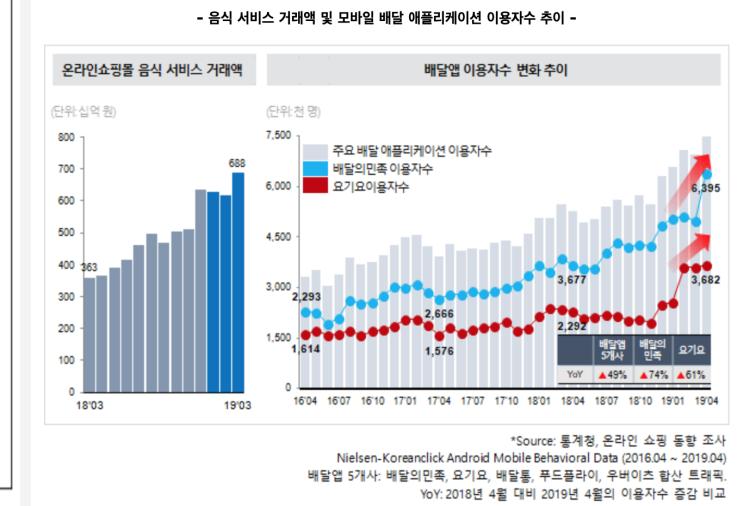
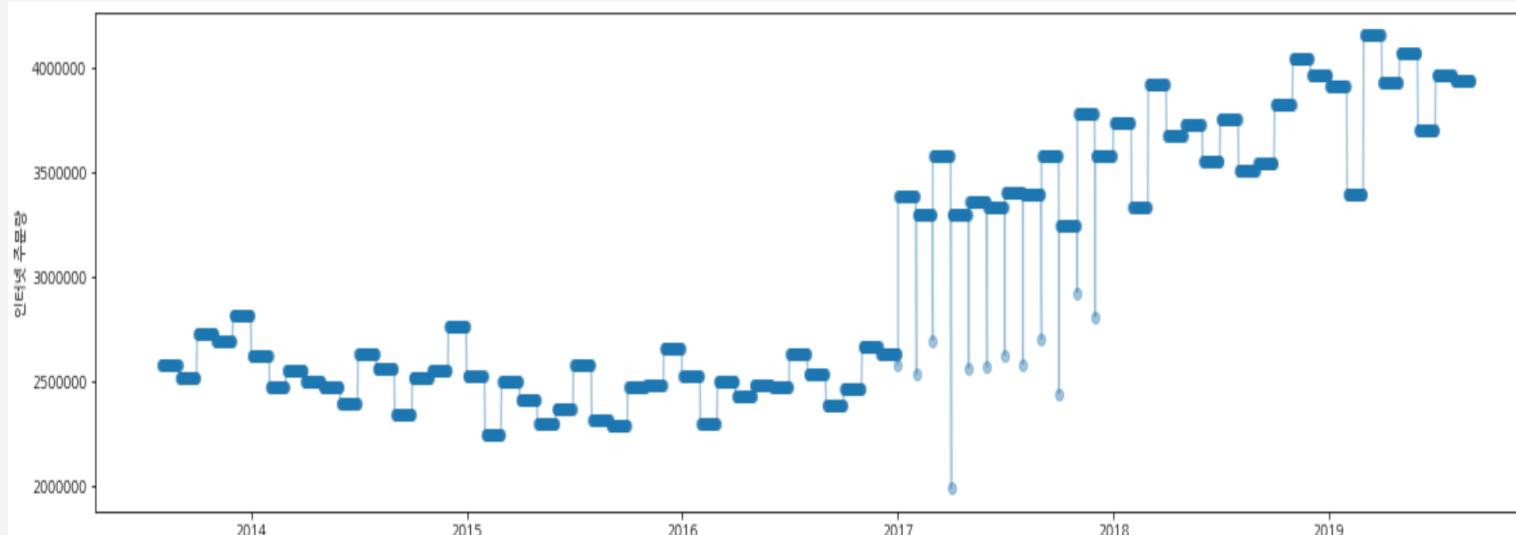
특히 '배달통'의 경우 지난 19일부터 28일까지 10일간 모바일결제 주문이 전주 대비 37% 증가하는 등 가파른 상승세를 보였다. 배달통은 이같은 상승세를 이룰 수 있었던 배경으로 타 업체보다 많은 '등록 배달업체 수'를 꼽았다.

4. 배달앱 주문량 보정

배달앱 점유율 데이터를 변수로 활용하여, 총 배달음식 주문량을 보정

외식배달업의 성장률을 반영하여 해당 지표 보정

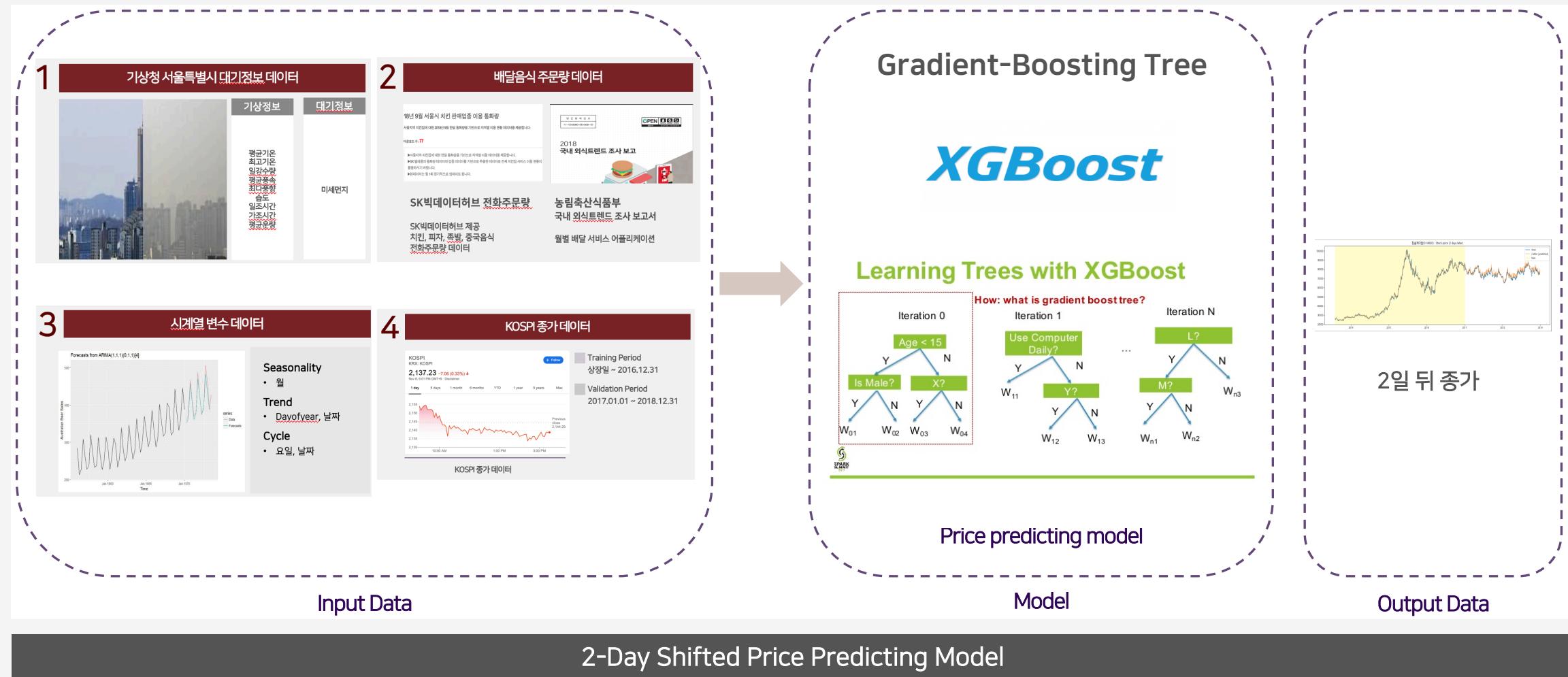
- 아래 데이터를 추가로 변수로 활용하여, 모델이 일별 주문량으로 스케일링하도록 학습
 - '2018 국내 외식 트렌드 조사 보고서', 농림축산식품부
 - '2013 - 2019 온라인 쇼핑몰별 운영형태/상품군별거래액', 통계청
 - 'Android Mobile App '배달의민족', '요기요' 4월 이용 행태 분석', Nielsen Koreanclick



'Android Mobile App '배달의민족', '요기요' 4월 이용 행태 분석'

Nielsen Koreanclick

5. 사용 모델



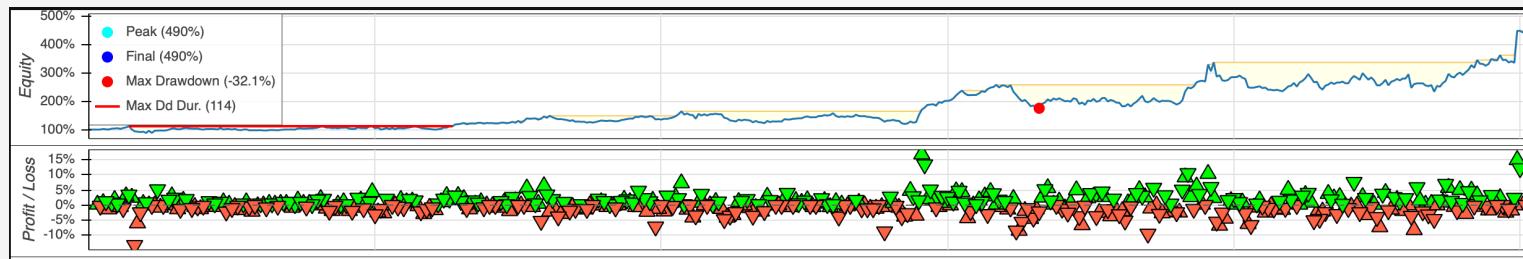
"미세먼지와 그 외 데이터, 당일 종가를 활용하여 철강업계 기업들의 2일뒤 종가 예측"

6-1. Backtesting 결과 (이화산업 000760)



Price Prediction

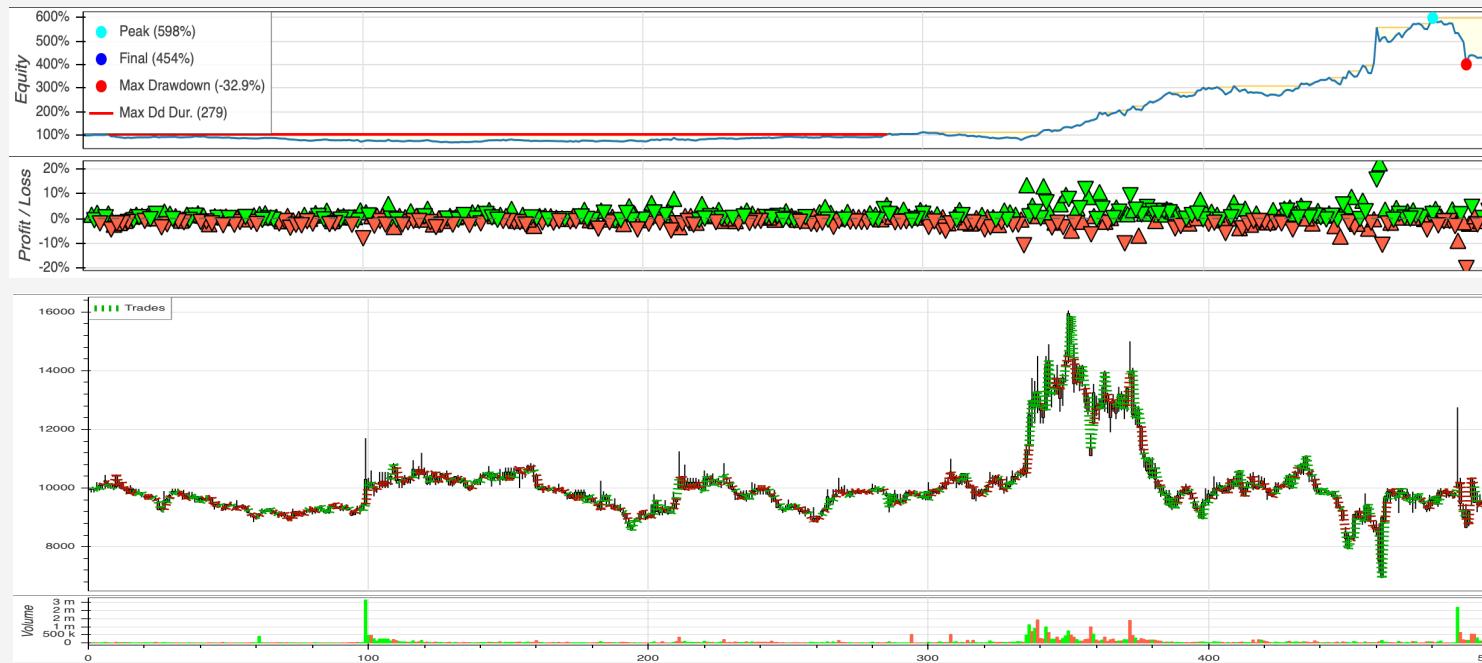
- Training Period : 2013.08.01 ~ 2016.12.31
- Validation Period : 2017.01.01 ~ 2018.12.31
- Training R2 Score : 0.9987
- Validation R2 Score : 0.7461



Backtesting Result

- End Equity : 490%
- Maximal Drawdown : -32.1%
- Sortino Ratio : 0.15
- Sharpe Ratio : 0.10
- Exposure : 99.67%
- Win Rate : 49.01%

6-2. Backtesting 결과 (한국주철관공업 000970)



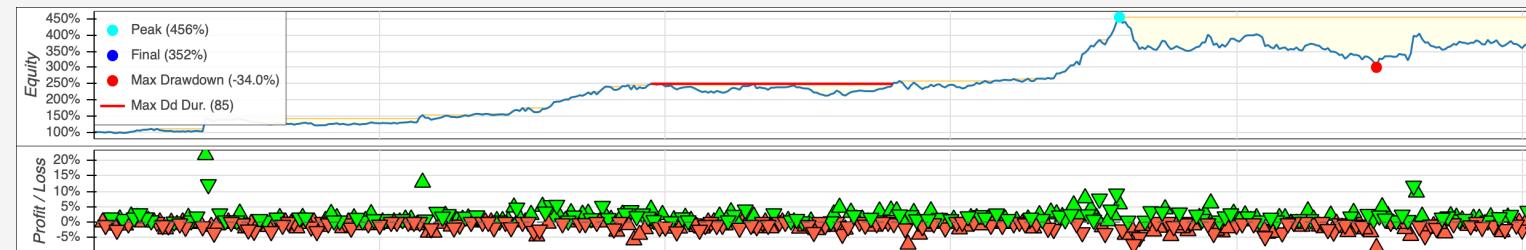
Price Prediction

- Training Period : 2013.08.01 ~ 2016.12.31
- Validation Period : 2017.01.01 ~ 2018.12.31
- Training R2 Score : 0.9994
- Validation R2 Score : 0.7682

Backtesting Result

- End Equity : 454%
- Maximal Drawdown : -32.9%
- Sortino Ratio : 0.09
- Sharpe Ratio : 0.06
- Exposure : 99.60%
- Win Rate : 46.41%

6-3. Backtesting 결과 (대장 012800)



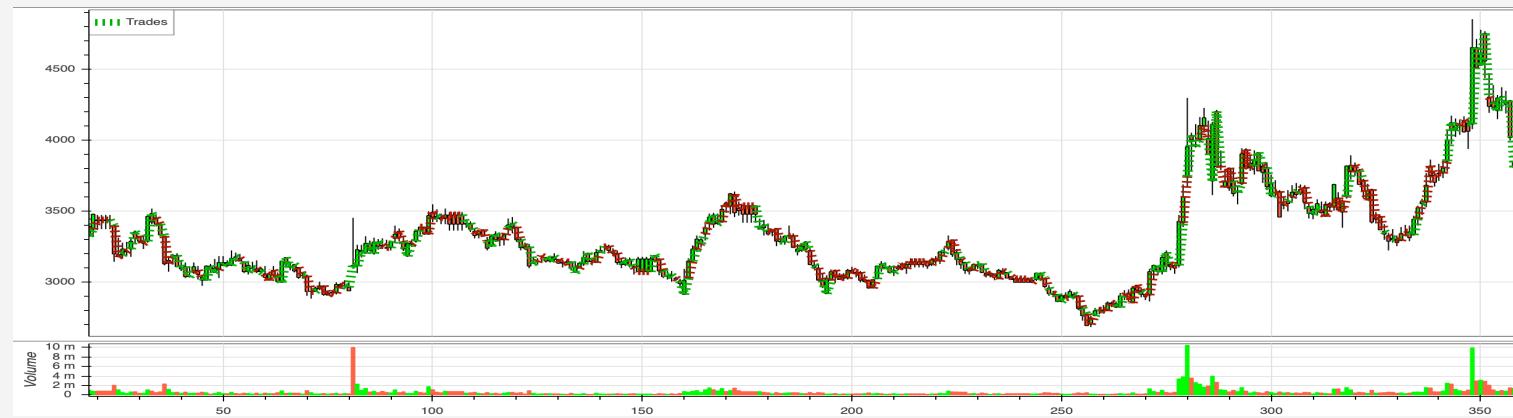
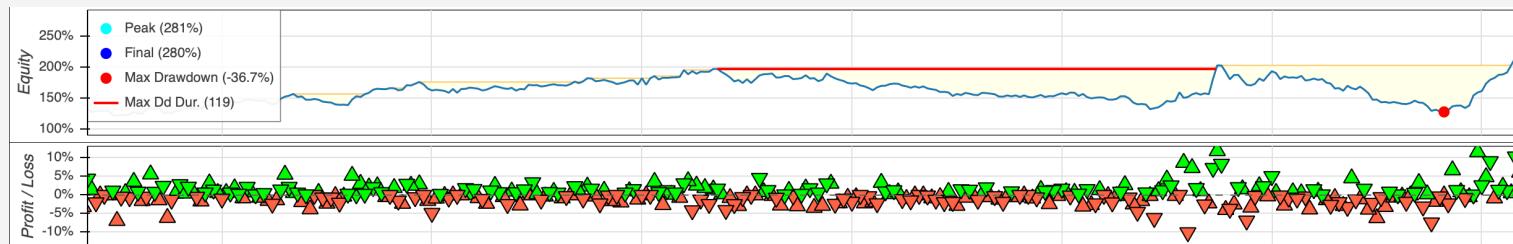
Price Prediction

- Training Period : 2013.08.01 ~ 2016.12.31
- Validation Period : 2017.01.01 ~ 2018.12.31
- Training R2 Score : 0.9966
- Validation R2 Score : 0.7792

Backtesting Result

- End Equity : 456%
- Maximal Drawdown : -34.0%
- Sortino Ratio : 0.18
- Sharpe Ratio : 0.09
- Exposure : 99.60%
- Win Rate : 44.62%

6-4. Backtesting 결과 (GS글로벌 001250)



Price Prediction

- Training Period : 2013.08.01 ~ 2016.12.31
- Validation Period : 2017.01.01 ~ 2018.12.31
- Training R2 Score : 0.9998
- Validation R2 Score : 0.7914

Backtesting Result

- End Equity : 280%
- Maximal Drawdown : -36.7%
- Sortino Ratio : 0.1
- Sharpe Ratio : 0.06
- Exposure : 99.60%
- Win Rate : 47.21%

7. Backtesting 결과 해석

검증 지표

- Profit의 Variance를 고려한 Sharpe Ratio, Sortino Ratio를 사용
- 높은 Profit을 내는 몇 번의 거래로 인해 Variance가 커져 해당 지표는 좋지 않음
- 추가로 End Equity, Max.Drawdown, Win Rate를 활용

하락장에 강한 알고리즘

- 선정된 업종들이 Validation 기간 내에 특별한 상승 추세를 보이지 않음에도 수익을 창출하는 알고리즘
- 대부분의 수익이 하락시의 반등에서 나타나고 있음

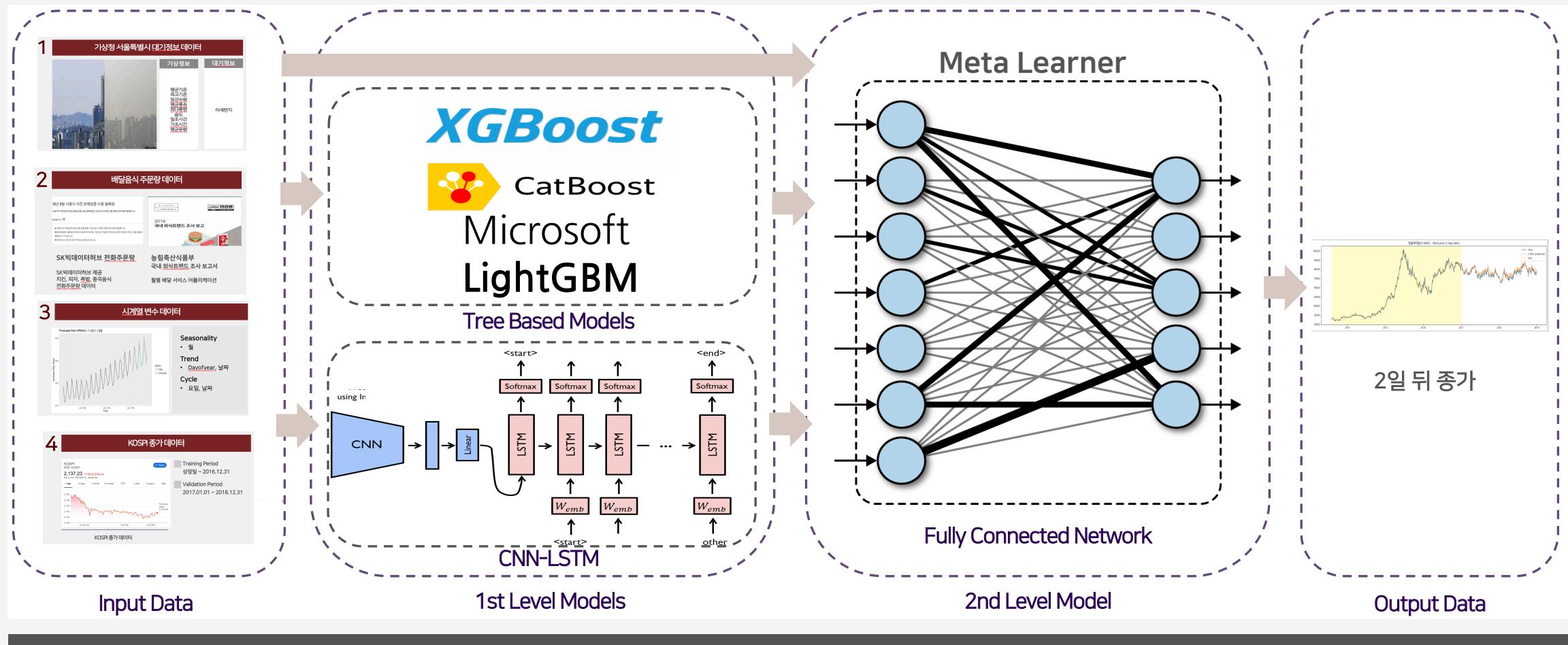
낮은 Max. Drawdown

- 낮은 Max. Drawdown을 보이며 높은 수익을 기대할 수 있는 알고리즘
- 전체 기간과 Maximum Profit을 고려했을 때 Max. Drawdown의 비율이 Max. Profit의 10% 수준

큰 수익거래를 잡아내는 알고리즘

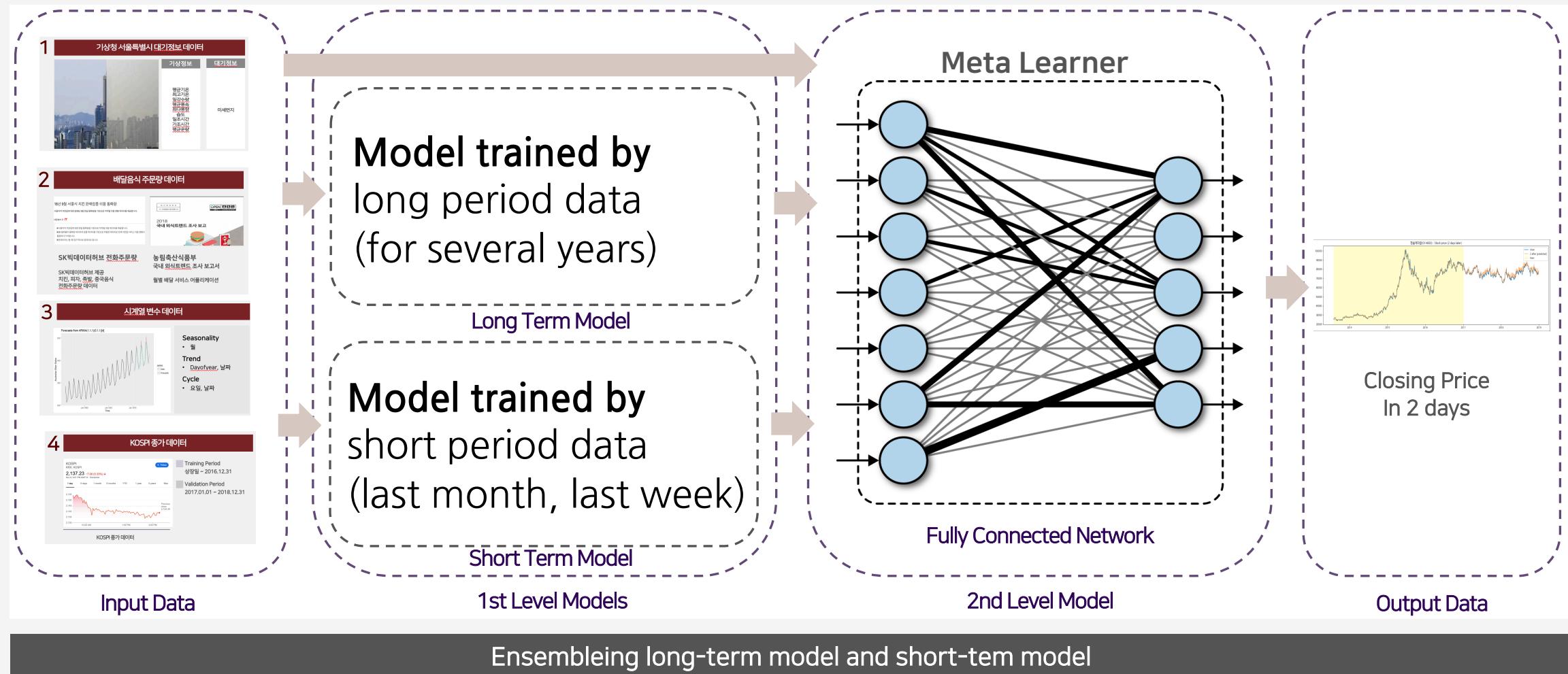
- 15~20%대의 수익을 내는 거래를 잡아내는 알고리즘
- 반면 손해를 보는 경우에는 최대 10% 수준이며, 그 빈도가 15~20%의 수익보다 훨씬 적음

8-1. 2차 기간중 보완할 내용



Stacked Regressor를 활용한 Price-Predicting Model 성능 개선

8-2. 2차 기간중 보완할 내용



Improving price-predicting model performance with considering long term and short term trend

8-3. 2차 기간중 보완할 내용

기상청 미세먼지 데이터보다 세부적인 항목을 제공하는 에어코리아 대기오염정보 데이터 사용으로 정확도 향상

기상청 데이터

기상정보와 함께 미세먼지라는 단일 항목만 제공

| 날짜 | 시군구 | 성별 | 연령대 | 총건수 | 월 | 요일 | 평균기온 | 최고기온 | 최저기온 | 일강수량 | 평균풍속 | 최다풍향 | 평균상대습도 | 기조시간 | 일조시간 | 강설량 | 평균운량 | 미세먼지 | |
|----|------------|-----|-----|-----|-----|----|------|------|------|------|------|------|--------|------|------|-----|------|------|----|
| 0 | 2013-08-01 | 강남구 | 남 | 10대 | 71 | 8 | 목요일 | 28.0 | 24.5 | 32.1 | 0.0 | 1.7 | 50.0 | 67.3 | 14.1 | 6.0 | 0.0 | 7.1 | 34 |
| 1 | 2013-08-01 | 강남구 | 남 | 20대 | 241 | 8 | 목요일 | 28.0 | 24.5 | 32.1 | 0.0 | 1.7 | 50.0 | 67.3 | 14.1 | 6.0 | 0.0 | 7.1 | 34 |
| 2 | 2013-08-01 | 강남구 | 남 | 30대 | 276 | 8 | 목요일 | 28.0 | 24.5 | 32.1 | 0.0 | 1.7 | 50.0 | 67.3 | 14.1 | 6.0 | 0.0 | 7.1 | 34 |
| 3 | 2013-08-01 | 강남구 | 남 | 40대 | 142 | 8 | 목요일 | 28.0 | 24.5 | 32.1 | 0.0 | 1.7 | 50.0 | 67.3 | 14.1 | 6.0 | 0.0 | 7.1 | 34 |
| 4 | 2013-08-01 | 강남구 | 남 | 50대 | 44 | 8 | 목요일 | 28.0 | 24.5 | 32.1 | 0.0 | 1.7 | 50.0 | 67.3 | 14.1 | 6.0 | 0.0 | 7.1 | 34 |
| 5 | 2013-08-01 | 강남구 | 남 | 60대 | 28 | 8 | 목요일 | 28.0 | 24.5 | 32.1 | 0.0 | 1.7 | 50.0 | 67.3 | 14.1 | 6.0 | 0.0 | 7.1 | 34 |
| 6 | 2013-08-01 | 강동구 | 남 | 10대 | 45 | 8 | 목요일 | 28.0 | 24.5 | 32.1 | 0.0 | 1.7 | 50.0 | 67.3 | 14.1 | 6.0 | 0.0 | 7.1 | 34 |
| 7 | 2013-08-01 | 강동구 | 남 | 20대 | 91 | 8 | 목요일 | 28.0 | 24.5 | 32.1 | 0.0 | 1.7 | 50.0 | 67.3 | 14.1 | 6.0 | 0.0 | 7.1 | 34 |
| 8 | 2013-08-01 | 강동구 | 남 | 30대 | 136 | 8 | 목요일 | 28.0 | 24.5 | 32.1 | 0.0 | 1.7 | 50.0 | 67.3 | 14.1 | 6.0 | 0.0 | 7.1 | 34 |
| 9 | 2013-08-01 | 강동구 | 남 | 40대 | 71 | 8 | 목요일 | 28.0 | 24.5 | 32.1 | 0.0 | 1.7 | 50.0 | 67.3 | 14.1 | 6.0 | 0.0 | 7.1 | 34 |

에어코리아 측정자료

PM10, PM2.5, 오존, 이산화질소, 일산화탄소, 아호아산가스 등
더 세세한 데이터 제공

| 날짜(년-월-일) | PM ₁₀ ($\mu\text{g}/\text{m}^3$) | PM _{2.5} ($\mu\text{g}/\text{m}^3$) | 오존(ppm) | 이산화질소(ppm) | 일산화탄소(ppm) | 아황산가스(ppm) |
|------------|---|--|---------|------------|------------|------------|
| 2019-01-01 | 36 | 25 | 0.02 | 0.03 | 0.6 | 0.003 |
| 2019-01-02 | 33 | 22 | 0.015 | 0.036 | 0.6 | 0.004 |
| 2019-01-03 | 35 | 22 | 0.007 | 0.048 | 0.7 | 0.004 |
| 2019-01-04 | 55 | 38 | 0.003 | 0.068 | 1 | 0.006 |
| 2019-01-05 | 60 | 38 | 0.021 | 0.03 | 0.6 | 0.005 |
| 2019-01-06 | 42 | 20 | 0.011 | 0.041 | 0.6 | 0.004 |
| 2019-01-07 | 51 | 32 | 0.01 | 0.053 | 0.9 | 0.005 |
| 2019-01-08 | 38 | 18 | 0.025 | 0.023 | 0.5 | 0.005 |
| 2019-01-09 | 47 | 22 | 0.009 | 0.041 | 0.7 | 0.004 |
| 2019-01-10 | 51 | 30 | 0.005 | 0.052 | 0.7 | 0.005 |
| 2019-01-11 | 71 | 48 | 0.004 | 0.061 | 1 | 0.005 |
| 2019-01-12 | 89 | 63 | 0.01 | 0.058 | 1 | 0.005 |
| 2019-01-13 | 107 | 73 | 0.018 | 0.05 | 1.1 | 0.005 |
| 2019-01-14 | 155 | 113 | 0.005 | 0.073 | 1.4 | 0.006 |
| 2019-01-15 | 122 | 72 | 0.024 | 0.029 | 0.8 | 0.005 |
| 2019-01-16 | 43 | 17 | 0.017 | 0.026 | 0.5 | 0.003 |
| 2019-01-17 | 39 | 20 | 0.014 | 0.038 | 0.6 | 0.004 |
| 2019-01-18 | 53 | 32 | 0.01 | 0.054 | 0.8 | 0.005 |
| 2019-01-19 | 87 | 58 | 0.01 | 0.059 | 1 | 0.007 |
| 2019-01-20 | 59 | 30 | 0.029 | 0.019 | 0.5 | 0.004 |

더 자세한 미세먼지 데이터

각 대기오염 요소가 예측 가격에 미치는 영향을 파악하여 더 자세한 모델 학습 가능

■ 9. 활용 전망

NEWSIS

철강업계 "고로 대기오염 물질 배출 현재로선 해결 어려워" 호소

김지은 2019.06.04. 13:06

2019년도에 들어 각종 철강산업들이 미세먼지 감축에 신경을 쓰고 있다. 하지만 위의 뉴스와 같이 실질적으로 대기오염 물질 배출 문제 해결은 쉽지 않다. 즉 사람들의 관심을 많이 받는 미세먼지는 감축이 가능 할 수 있을 지라도 다양한 철강산업과 관련된 다양한 대기오염지수는 여전히 존재 할 것이다. 이를 이용하여 미세먼지 외에도 오존지수, 이산화질소, 일산화질소, 아황산가스 지수를 철강산업 예측에 이용할 수 있을 것으로 예상된다.

10. 참고자료

신문기사

- '최악의 미세먼지 원인, 중국일까 한국일까.. 확인 가능하다', 송채경화, 한겨레
- '미세먼지 오염 주 원인은 지나친 과장', 윤철주, 철강금속신문
- '철강업계 "고로 대기오염 물질 배출 현재로선 해결 어려워" 호소', 김지은, NEWSIS

참고문헌

- '알고리즘 트레이딩', 장순용
- '알고리즘 트레이딩. 2.' 금융 투자 리스트, 장순용
- '파이썬으로 배우는 알고리즘 트레이딩', 조대표
- '머신러닝을 이용한 알고리즘 트레이딩 시스템 개발', 안명호 외 1인

Python Package Documentation

- Backtesting.py Documentation
- XGBoost Documentation

논문

- 'A Study on the Number of Domestic Food Delivery Services (2015)', Jaeyoung Kwion et al.

보고서

- '2018 국내 외식 트렌드 조사 보고서', 농림축산식품부
- '2013 – 2019 온라인 쇼핑몰별 운영형태/상품군별거래액', 통계청
- 'Android Mobile App '배달의민족', '요기요' 4월 이용 행태 분석', Nielsen Koreanclick