

A word cloud centered around the phrase "BIG DATA" in large, bold, red capital letters. Surrounding this central text are various related terms in black, bold, capital letters of varying sizes. The words are arranged in a circular pattern around the center. The words include: CONSUMER, STORAGE, SAMPLE, RESEARCH, TECHNOLOGY, SIZE, INTERNET, INFORMATION, BEHAVIOR, ANALYTICS, MARKETING, COMPUTERS, and BYTES.

CONSUMER STORAGE
COMPUTERS MARKETING SAMPLE
BYTES **BIG DATA** RESEARCH
BEHAVIOR ANALYTICS TECHNOLOGY
INFORMATION SIZE INTERNET



박성용 SeongYong Park

☎ 010-9522-0172

✉ pty04290@gmail.com

🐙 <https://github.com/pty902/Resume>

Introduce 팀 프로젝트에서 PM을 맡아 Ideation을 담당하였으며 전체적인 일정과 계획을 짜서 팀원들에게 업무 분담을 하였습니다.

또한 OpenAPI, WebCrawling 등의 기술을 사용해서 팀 프로젝트의 자료수집 및 전처리 파트를 담당하였습니다.

SKILL (1~5)

Python 3
Linux 3
SQL 3
R 3
HTML5 3
CSS3 3
JavaScript 3



MariaDB



ORACLE
DATABASE



Interests

TensorFlow
Machine Learning
Deep Learning



- 프로젝트 서론
- 프로젝트 본론
- 프로젝트 결론
- 현재 연구 진행상황
- 추후 연구 계획

- 프로젝트 서론
- 프로젝트 본론
- 프로젝트 결론
- 현재 연구 진행상황
- 추후 연구 계획

프로젝트 명 : Fine Dust

개발기간 : 2018.12.06 ~ 2019.01.23

참여인원 : 3명

담당업무 : PM (자료수집 및 전처리)

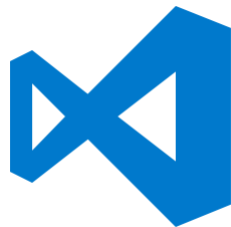
개발환경 : Windows7 64bit

Python 3.6.4 R 3.5.1

MariaDB 10.3 MYSQL 5.7

사용도구 : Jupyter 5.0.0, VSCode , Github Desktop
R Studio 1.1.4, 공공데이터 OpenAPI, Sublime text3
AWS EC2, Xshell 6, Ubuntu 17.10 64 bit

사용기술 : Python R HTML5 CSS3 JavaScript



서울경제

미세먼지에도 마스크 안쓰는 노인

미세먼지에도 마스크 안쓰는 노인

면역력 약해 질병발생 높는데
"귀찮아서" 악영향 인식 안해

기획 보기

오지현 기자 | 2018-01-21 17:19:13 | 기획·연재 31면

역랑

인기사육률



초미세먼지 주의보가 발령된 지난 20일 종로구 탑골공원을 찾은 노인들 중 대다수가 마스크를 쓰지 않은 채 야외에서 담소를 나누고 있다. /오지현기자

미세먼지에대한 인식미비

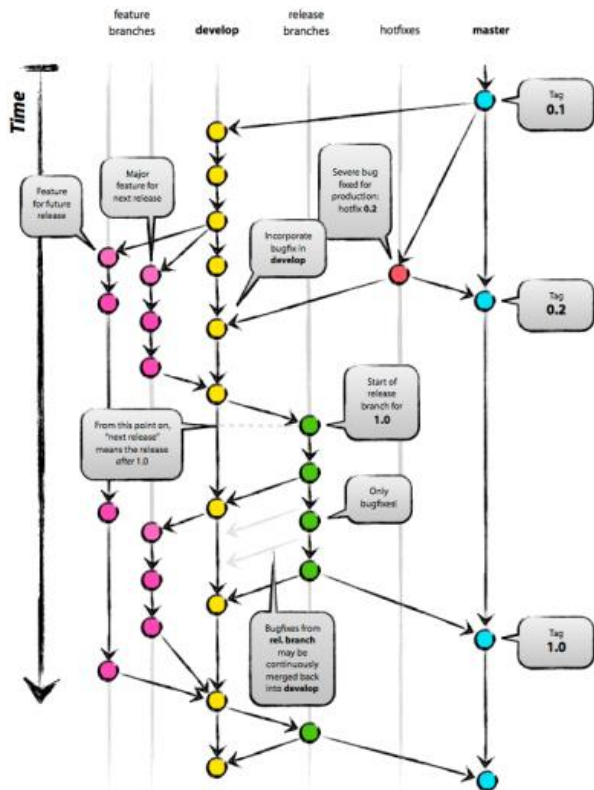
미세먼지 개의치 않는 취약계층인 노인들
삼겹살 먹으면 미세먼지가 배출된다는 잘못된 상식
식약처 인증을 받은 마스크 같은 관련 제품 필요성 인식 부족

프로젝트 프로세스

1. 미세먼지 관련 제품 업체에서
매출 상승을 위한 분석 의뢰를 받았다고 가정
2. 미세먼지와 관련 질병과의 상관관계를 분석하여
소비자들에게 관련 제품의 필요성을 인식시키고
객관적인 데이터에 기반한 관련 제품 판매전략 수립
3. 현업이 정말 필요로 하고 가치 있어 할 만한
분석을 하기 위해 업체와 Co-Work를 하고
분석 결과를 이해서키며 피드백을 받는 프로세스 추가

출처 : 서울경제

<https://www.sedaily.com/NewsView/1RUIJOXL33>



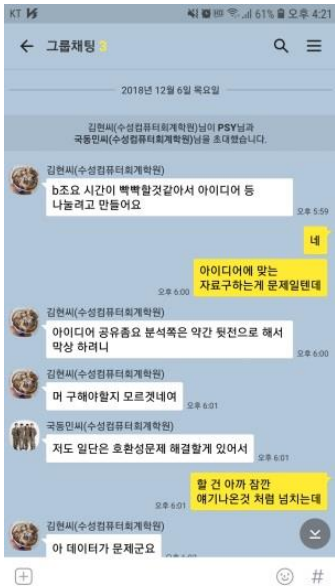
Branch를 5개로 나누어 버전 관리하였습니다.

작업물 Merge시 Conflict를 최대한 방지할 수 있었고

작업과정을 일목요연하게 볼 수 있었습니다.

출처 : nvie.com

<https://nvie.com/posts/a-successful-git-branching-model/>



Initial commit	Verified	a41bc7d	<>
ks5721 committed on 12 Dec 2018			
Add files via upload	Verified	bd34227	<>
pty902 committed on 17 Dec 2018			
유아교육 업체관련 정보	Verified	29b78b8	<>
kdm0904 committed on 19 Dec 2018			
maps	Verified	a54dc79	<>
kdm0904 committed on 7 Jan			
Fine Dust PPT For Presentation ver 1.0	Verified	56f3447	<>
pty902 committed on 20 Jan			
Commits on Jan 23, 2019			
html		d8727dd	<>
ks5721 committed 27 days ago			
Fine Dust PPT For Print	Verified	5739716	<>
pty902 committed 27 days ago			
Fine Dust PPT For Presentation	Verified	7fc1b26	<>
pty902 committed 27 days ago			

2018.12.06 프로젝트 팀 결성

2018.12.12 그룹 Github 생성

2018.12.17 Feature Branch 자료저장

인터넷검색 OpenAPI WebCrawling 등으로 수집한 자료들과 초기 분석 결과를 홈페이지 저장

2018.12.19 Develop Branch 자료저장

전처리된 자료들과 CSS3 JavaScript 처리된 홈페이지, 완성된 분석 결과물 (어린이집 경로당 관련 질병 등) 저장

2019.01.07 Release Branch 자료저장

초기 PPT, Develop 자료 중 회의를 거쳐 선별된 어린이집 경로당 관련 질병 관련 제품 분석 결과물을 코드와 함께 저장

2019.01.20 Hotfix Branch 자료저장

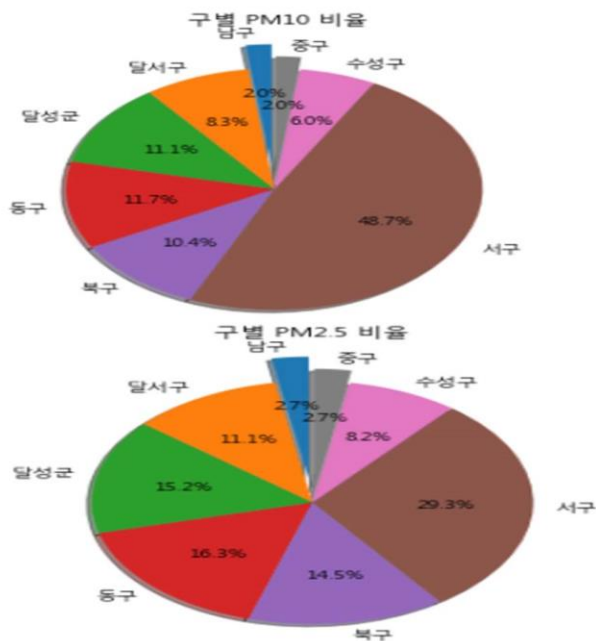
PPT, 분석 결과물 수정 및 아이디어 추가, 코드 정리

2019.01.23 Master Branch 자료저장

PPT 홈페이지 마무리 후 프로젝트 종료

- 프로젝트 서론
- 프로젝트 본론
- 프로젝트 결론
- 현재 연구 진행상황
- 추후 연구 계획

구별 비율



미세먼지를 가장 많이 배출하는 곳은 서구이고 가장 적게 배출하는 곳은 남구 중구임을 알 수 있다.

2015년 기준

출처 : 통계청

matplotlib.pyplot 파이썬 모듈을 사용하여 파이 그래프로 분석 결과 시각화

https://github.com/Bgroupbb/b-store/blob/Hotfix/findustemssion/finedust_emssion_rate_daegu.py In [17] 참고

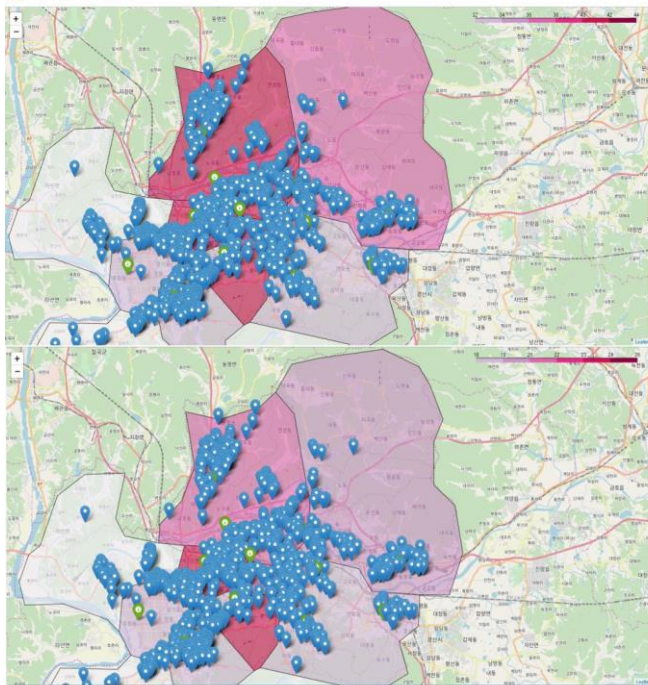
핵심코드

```
sizes2 = [26.238, 108.183, 147.926, 158.099, 140.682, 284.892, 79.259, 26.249]
explode = (0.2, 0, 0, 0, 0, 0, 0, 0.1)

plt.figure(figsize=(20, 15))
fig2, ax2 = plt.subplots()
ax2.pie(sizes2, explode=explode, labels=labels, autopct='%1.1f%%',
        shadow=True, startangle=90)
ax2.axis('equal')

plt.title('구별 PM2.5 배출 비율')
plt.tight_layout()
plt.legend(labels, loc='center left',
           bbox_to_anchor=(1, 0.5))
plt.show()
```

어린이집 분포



출처 : 통계청

Folium 파이썬 모듈을 사용하여 Google Maps에
Json 파일을 이용한 대구 미세먼지 구별 분포를 표시하고 어린이집 위치를 Marker로 표시

PM10

대구 어린이집 위치와
미세먼지 구별 분포도

2018 연평균 기준

PM2.5

핵심코드

```
mapping = folium.Map(location=[df['lat'].mean(),  
df['lng'].mean()], zoom_start=11)+  
+  
mapping.choropleth(geo_data=geo_data, +  
data = gu_data['연_평균값'], +  
columns = [gu_data.index,  
'연_평균값'], +  
fill_color = 'PuRd', +  
key_on = 'feature.id')+  
+  
for n in df.index: +  
if df['도로명주소'][n] != 'Multiple': +  
folium.Marker([df['lat'][n], df['lng'][n]],  
popup=df['상호명'][n]).add_to(mapping)+  
+  
for x in air.index: +  
if air['측정소'][x] != 'Multiple': +  
folium.Marker([air['lat'][x],  
air['lng'][x]], popup=air['측정소'][x],  
icon=folium.Icon(color='green', icon='info-  
sign')).add_to(mapping)+  
+  
mapping+
```

■ 미세먼지농도와 관련질병환자수의 상관관계

	PM10	PM2.5	C34	I20	I50	I64	J18
PM10	1.0000000	0.8315156	-0.387220458	0.20192382	0.3975649	0.29913714	0.82596620
PM2.5	0.8315156	1.0000000	-0.302565221	0.27783479	0.2957580	0.31671769	0.71479554
C34	-0.3872205	-0.3025652	1.000000000	0.53847864	0.5239977	0.34968087	-0.08172730
I20	0.2019238	0.2778348	0.538478641	1.000000000	0.9060248	0.84382789	0.08399057
I50	0.3975649	0.2957580	0.523997696	0.90602484	1.000000000	0.83164942	0.32191412
I64	0.2991371	0.3167177	0.349680875	0.84382789	0.8316494	1.000000000	0.05281766
J18	0.8259662	0.7147955	-0.081727302	0.08399057	0.3219141	0.05281766	1.000000000
J44	0.7916848	0.7347171	-0.007118199	0.61951078	0.7522035	0.65015171	0.65359746
J45	0.8240417	0.7834161	-0.128731840	0.12967552	0.2927814	0.07010298	0.96087105
	J44	J45					
PM10	0.791684843	0.82404169					
PM2.5	0.734717149	0.78341608					
C34	-0.007118199	-0.12873184					
I20	0.619510783	0.12967552					
I50	0.752203480	0.29278144					
I64	0.650151714	0.07010298					
J18	0.653597458	0.96087105					
J44	1.000000000	0.66214793					
J45	0.662147935	1.000000000					

출처 : 통계청 중앙응급의료센터

cor 함수를 사용하여 미세먼지 농도와 관련질병환자수의 상관관계를 시각화

https://github.com/Bgroupbb/b-store/blob/Hotfix/ill%26dust_cor/illdust.R Line 1 ~ 10 참고

핵심코드

```

dat <-
read.csv('C:/Users/USER/Desktop/프로젝트/csv/R 그래프용/ill.csv')
cor(dat)
library(ggplot2)
ggplot(dat, aes(J18, PM10))+ geom_point() + geom_smooth(method =
"lm")
ggplot(dat, aes(J18, PM2.5))+ geom_point() + geom_smooth(method
= "lm")
ggplot(dat, aes(J44, PM10))+ geom_point() + geom_smooth(method =
"lm")
ggplot(dat, aes(J44, PM2.5))+ geom_point() + geom_smooth(method
= "lm")
ggplot(dat, aes(J45, PM10))+ geom_point() + geom_smooth(method =
"lm")
ggplot(dat, aes(J45, PM2.5))+ geom_point() + geom_smooth(method
= "lm")
corrplot(dat)

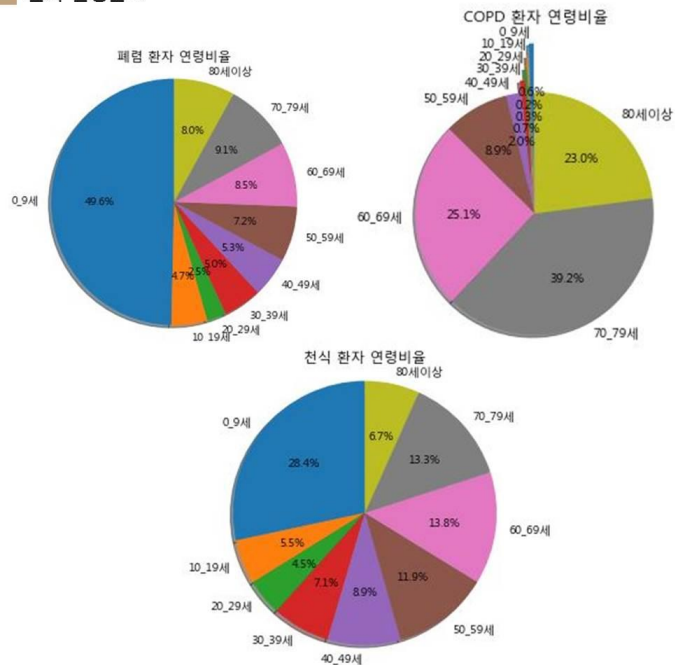
```

J18 폐렴
J44 만성 폐색성 폐질환
J45 천식

0.1이상 무시 선형관계
0.3이상 약한 선형관계
0.7이상 뚜렷한 선형관계
1.0이상 강한 선형관계

2017년 기준

환자 연령분포



미세먼지 관련 질병별
취약계층 분석 결과
폐렴은 어린이
COPD는 노인
천식은 노약자 모두
발병률이 높음을 확인

2017년 기준

핵심코드

```
labels = ['0_9세', '10_19세',
          '20_29세', '30_39세', '40_49세',
          '50_59세', '60_69세', '70_79세',
          '80세이상']
sizes = [3863, 1169, 1855, 4048, 12084,
         53292, 149978, 234384, 137240]
explode = (0.4, 0.3, 0.2, 0.1, 0, 0, 0,
           0, 0)
fig, ax = plt.subplots()
ax.pie(sizes, explode=explode,
       labels=labels, autopct='%1.1f%%',
       shadow=True, startangle=90)
ax.axis('equal')

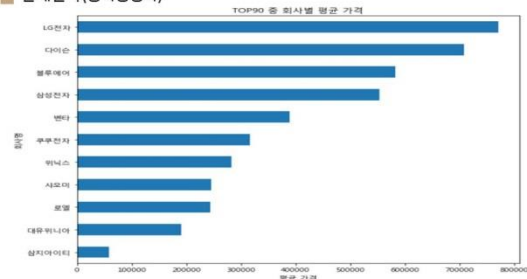
plt.title('COPD 환자 연령비율')
plt.tight_layout()
plt.show()
```

출처 : 보건 의료 빅데이터 개방 시스템

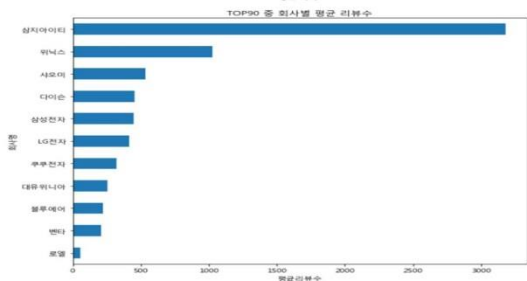
matplotlib.pyplot 파이썬 모듈을 사용하여 파이 그래프로 분석 결과 시각화

https://github.com/Bgroupbb/b-store/blob/Hotfix/all_ills.ipynb In [8] 참고

판매전락(공기청정기)



삼지아이티가 가격도 싸고
리뷰수(구매수)도
압도적으로 많다.



출처 : 다나와

matplotlib.pyplot 파이썬 모듈을 사용하여 **바 그래프**로 분석 결과 시각화

https://github.com/Bgroupbb/b-store/blob/Hotfix/shoppingmall_goods.ipynb In [19] 참고

핵심코드

```
pv2 = pd.pivot_table(df, columns=["company"],
values=['reviews'], aggfunc=[np.mean], fill_value=0)+
pv2+
pv2.to_excel(r'C:\Users\USER\Desktop\프로젝트\csv\크롤
링파일\공기청정기\companyreviews.xlsx', encoding='UTF-
8')+
cr
=
pd.read_excel(r'C:\Users\USER\Desktop\프로젝트\csv\크롤
링파일\공기청정기\companyreviews2.xlsx', encoding='UTF-
8')+
cr.set_index('company', inplace=True)+
cr.head()+
cr['reviews'].sort_values().plot(kind='barh',
figsize=(9, 8))+
plt.title('TOP90 중 회사별 평균 리뷰수')+
plt.ylabel('회사명')+
plt.xlabel('평균리뷰수')+
plt.show()+
```

결론



홈페이지 <https://bgroub.net>

실시간정보를 홈페이지를 통해서 얻을 수 있다.
다양한 정보 및 인터넷 쇼핑 구매가이드 미세먼지관련 어플들을 소개를 하고있다.

Ubuntu 17.10, Apache 2.9.1, AWS EC2를 사용하여 서버를 만들고
Sublime Text3로 HTML5 CSS3 JavaScript를 코딩하여 만든 홈페이지

<https://github.com/Bgroupbb/b-store/blob/master/html/index.html> Line 2 ~ 25 참고

핵심코드

```
<head>+
  <meta charset="UTF-8">+
  <meta name="description" content="">+
  <meta http-equiv="X-UA-Compatible" content="IE=edge">+
  <meta name="viewport" content="width=device-width, initial-scale=1,
  shrink-to-fit=no">+
  <!-- Title -->+
  <title>B 조 단체구매</title>+
+
  <link rel="icon" type="image/png" sizes="32x32" href="/img/core-
  img/favicon.ico" />+
  <!-- Style CSS -->+
  <link rel="stylesheet" href="style.css">+
  <link
href="https://fonts.googleapis.com/css?family=Janum+Pen+Script&subset=
+Korean" rel="stylesheet">+
  <link href="https://fonts.googleapis.com/css?family=Bugl"
rel="stylesheet">+
  <link href="https://fonts.googleapis.com/css?family=Black+Han+Sans"
rel="stylesheet">+
  <link href="https://fonts.googleapis.com/css?family=Okdo"
rel="stylesheet">+
  <link href="https://fonts.googleapis.com/css?family=Outa+Font"
rel="stylesheet">+
  <link href="https://fonts.googleapis.com/css?family=Janum+Pen+Script"
rel="stylesheet">+
  <link href="https://fonts.googleapis.com/css?family=Noto+Sans+KR"
rel="stylesheet">+
  <link href="https://fonts.googleapis.com/css?family=Leon+Bang"
rel="stylesheet">+
  <link href="https://fonts.googleapis.com/css?family=Jua"
rel="stylesheet">+
+
</head>+
```

- 프로젝트 서론
- 프로젝트 본론
- 프로젝트 결론
- 현재 연구 진행상황
- 추후 연구 계획

미세먼지관련상품판매 기획안 결론

1. 판매지역 : 대구광역시 서구, 북구, 남구, 중구

2. 타겟 연령층 : 30~50대

3. 홍보 및 판매전략 집중이 필요한 주요 회사상품 :

마스크 : 3M, 유한킴벌리, 파인텍, 웰킵스, 장정산업

가성비 - 장정산업 애니가드 방진마스크 VC201G

최고 인기상품 - 파인텍 네펜어 황사방역용 마스크 KF94 대형

공기청정기 : 삼성, LG, 위닉스, 삼지아이티

가성비 - 삼지아이티 PIsnet 피스넷 에어, 퓨어 360

최고 인기상품 - 삼성전자 블루스카이 AX60N5580WDD

홈페이지를 통해 '다나와' 사이트를 연결해서

자세한 인기상품을 볼 수 있음

구매 가이드라인 세일정보등 도움되는 정보들을 제공



<http://bgroup.ml/>



<https://github.com/Bgroupbb/b-store/blob/master/Fine%20Dust.pdf>

이 프로젝트로 미세먼지와 관련질병의 상관관계를 분석하려했으나 R을 잘 다루지 못 해 통계적인 내용을 구하기 힘들었습니다.

그리고 머신러닝을 하지 못 하여 미세먼지 향후 예측을 못 했습니다.

그래서 R 과 머신러닝을 개인적으로 공부하여 수학과 통계적인 부분에서 설득력을 가질 수 있도록 하겠습니다.

1. 판매지역 우선순위 설정 할 수 있다.

- 미세먼지 분포도가 심한 지역
- 취약계층 분포가 많은 지역

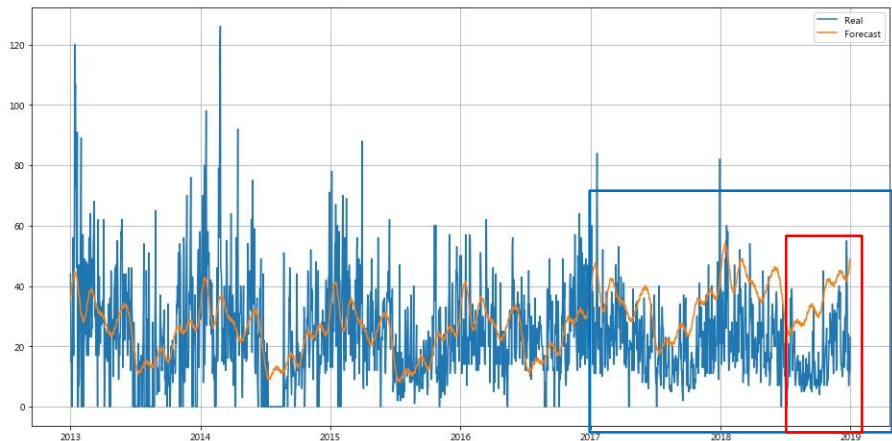
2. 판매대상 우선순위 설정 할 수 있다.

- 취약계층 분포가 많은 지역

3. 판매상품 우선순위 설정 할 수 있다.

- 인기상품
- 가성비상품

- 프로젝트 서론
- 프로젝트 본론
- 프로젝트 결론
- 현재 연구 진행상황
- 추후 연구 계획



Prophet 파이썬 모듈을 사용하여 730일(2년) 치의 자료를 예측

[https://github.com/pty902/Resume/blob/master/Fine%20Dust%20Project/Fine%20Dust\(Prophet\).ipynb](https://github.com/pty902/Resume/blob/master/Fine%20Dust%20Project/Fine%20Dust(Prophet).ipynb) In [30~34] 참고

핵심코드

```
m = Prophet(daily_seasonality=True)
m.fit(df2)

future = m.make_future_dataframe(periods=730)
future.tail()

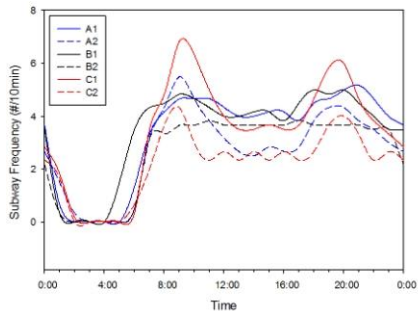
forecast = m.predict(future) forecast[['ds', 'yhat', 'yhat_lower',
'yhat_upper']].tail()

plt.figure(figsize=(16,8))
plt.plot(df3['ds'], df3['y'], label='Real')
plt.plot(forecast['ds'], forecast['yhat'], label='Forecast')
plt.grid()
plt.legend()
plt.show()
```

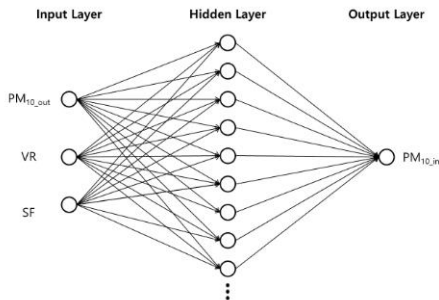
2017~2018년 실제값과
2017~2019년 예측값(파란박스) 을 같이 두고 그래프화

2018년 7~10월, 12월값(빨간박스)을
제외하곤 대체적으로 비슷한 모습
2018년 7월 고기압 2018년 12월 러시아발 고기압때문에
중국으로부터의 미세먼지 유입이 없었기 때문이었음

- 프로젝트 서론
- 프로젝트 본론
- 프로젝트 결론
- 현재 연구 진행상황
- 추후 연구 계획

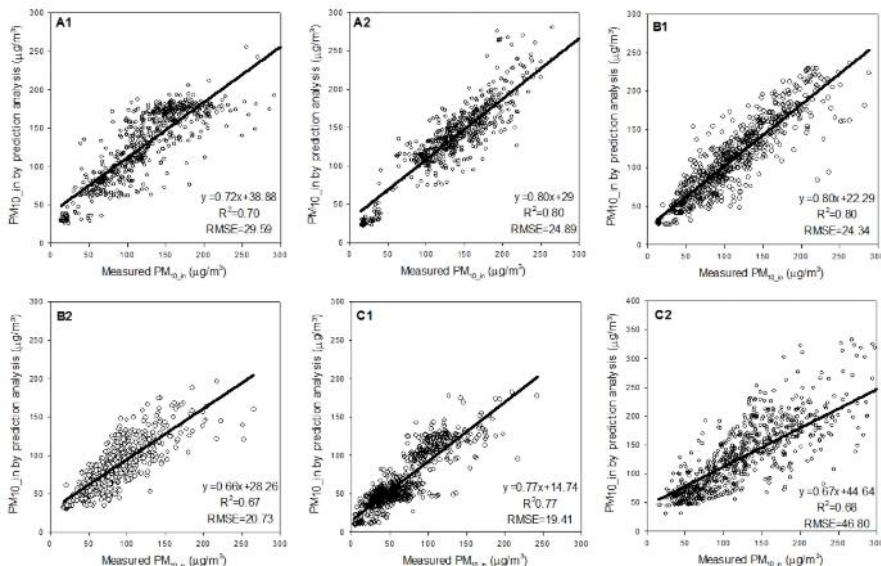


[그림 3] 지하철 미세먼지 예측 ANN모델 구조와 입력변수 데이터 패턴*7



머신러닝을 공부를 하여 모델을 훈련시켜보겠습니다.
그래서 향후 미세먼지 농도 예측을 다양하게 해보고
최적의 모델을 찾아서 실용성을 높여보겠습니다.

관련자료 출처 : 카카오시리포트 권순박
<https://brunch.co.kr/@kakao-it/219>



[그림 4] 서울시 6개 지하철역사 내 ANN모델을 활용한 미세먼지 예측 정확도*8

Thank You!

