

S/W 개발 및 빅데이터 응용과정 국비교육

# 미니 프로젝트

주재영 오병규 이희문 최연웅

# 서울시 파릉이 대여량 예측

# 목 차



## Step 1

- 그룹 및 프로그램 소개
- 목표 정의



## Step 2

- 진행 과정 설명
- 결과 설명



## Step 3

- 스스로 하는 평가
- 출처



**Step 1**

# 그룹 소개



주재영  
분석 및 ppt 작성



오병규  
분석



이희문  
UI 디자인



최연웅  
UI 디자인

# 프로그램

다양한 기후 조건  
서울시 따릉이 대여량을 예측

MainWindow

Data Head

Predict

Graph

Open Path

Test Path

Submission Path

Train Open

Test Open

Sub Open

☐ hour

☐ temperature

☐ precipitation

☐ windspeed

☐ humidity

☐ visibility

☐ ozone

☐ pm10

☐ pm2.5

☐ All

Learning

Train Open

Test Open

Sub Open

☐ hour

☐ temperature

☐ precipitation

☐ windspeed

☐ humidity

☐ visibility

☐ ozone

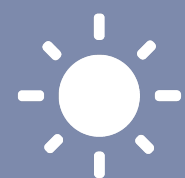
☐ pm10

☐ pm2.5

☐ All

Learning

# 목표



## 첫 번째 목표

특정 변수들을 PyQt5를 이용해  
선택하여 따릉이 대여량 예측해보기



## 두 번째 목표

가장 잘 구현한 예측 모델을  
시각화를 통해 알아보기



**Step 2**



# 분석 기법 소개

## 선형회귀 (Linear)

독립변수  $x$ 와 종속변수  $y$ 의 관계를 선형으로 모델링

## 랜덤포레스트 (RandomForest)

배깅 기법의 한 종류로 자료를 랜덤하게 학습시켜 평균낸 값으로 결과를 내는 모델링

## XGBoost

부스팅 기법의 한 종류로 기존 결정트리 방식에서 과적합 방지에 특화된 모델링

### 1) LinearRegression

```
# 모델 생성
model_LR = LinearRegression()

# 학습
model_LR.fit(x_train, y_train)

# 예측값
y_pred_LR = model_LR.predict(x_test)
```

### 2) RandomForestRegressor

```
# 모델 생성
model_RF100 = RandomForestRegressor(n_estimators=100, random_state=0)

# 모델 학습
model_RF100.fit(x_train, y_train)

# 예측값
y_pred_RF100 = model_RF100.predict(x_test)
```

### 3) XGBRegressor

```
# 모델 생성
model_XGB = XGBRegressor(max_depth = 4, n_estimators = 500)

# 모델 학습
model_XGB.fit(x_train.astype(np.float32), y_train.astype(np.float32))

# 예측값
y_pred_XGB = model_XGB.predict(x_test.astype(np.float32))
```

# UI 소개

## 각 항목 기능

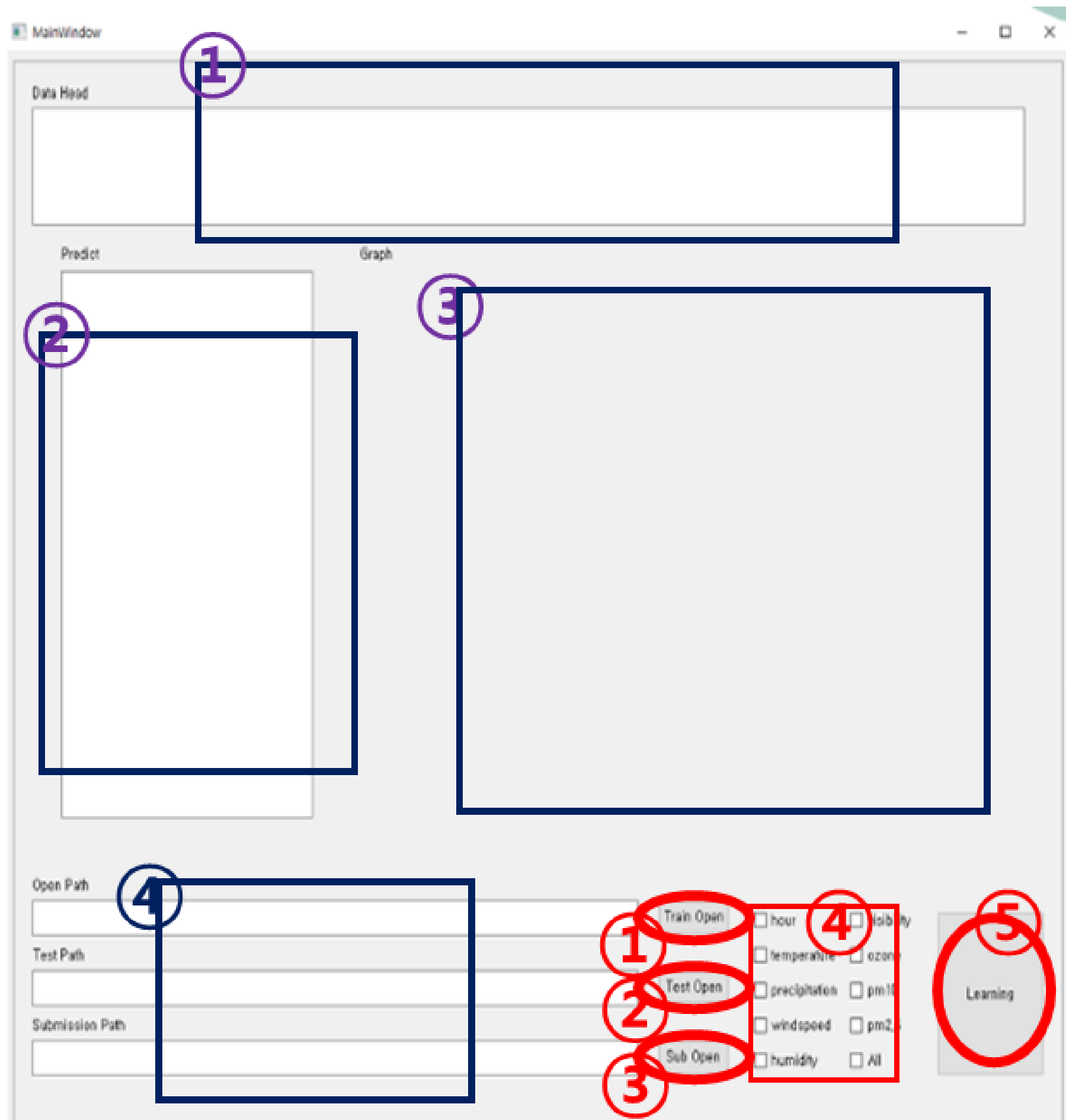
- Data Head : 불러온 csv 데이터의 head 값을 미리볼 수 있다.
- Predict : 분석한 데이터의 예측값 출력 공간
- Graph : 분석한 데이터의 heatmap 시각화 출력 공간
- Train Path : Trainig 데이터의 full path
- Test Path : Test 데이터의 full path
- Submission Path : Submission 데이터의 full path

## 버튼 기능

- Train Open, Test Open, Sub Open 버튼으로 계산에 필요한 각 데이터를 dialog 탐색으로 불러올 수 있다
- 체크박스 : 분석시 체크된 데이터로만 예측값을 계산할 수 있다
- Learning : 계산 실행 버튼

The screenshot shows a software window titled 'MainWindow'. The interface is divided into several sections:

- Data Head**: A large empty rectangular box at the top.
- Predict**: A large empty rectangular box on the left side.
- Graph**: A large empty rectangular box on the right side.
- Open Path**: Three text input fields for 'Train Path', 'Test Path', and 'Submission Path'.
- Buttons**: Three buttons labeled 'Train Open', 'Test Open', and 'Sub Open' are positioned to the right of their respective input fields.
- Checkboxes**: A group of checkboxes on the right side, including 'hour', 'visibility', 'temperature', 'ozone', 'precipitation', 'pm10', 'windspeed', 'pm2.5', 'humidity', and 'All'.
- Learning Button**: A large button labeled 'Learning' is located at the bottom right of the window.



- ① 훈련 데이터 셋을 불러 옵니다.
- ② 테스트 데이터 셋을 불러 옵니다.
- ③ 서브미션 데이터를 불러 옵니다.
- ④ 특정 환경 변수를 지정합니다.
- ⑤ 학습을 통하여 결과를 보여줍니다.

- ① 각 각의 데이터의 헤드 값을 보여줍니다.
- ② 결과 값을 나타냅니다.
- ③ 결과 값으로 그린 히트맵을 보여줍니다
- ④ 파일 경로를 나타냅니다.



**Step 3**

### 주재영

다양한 분석 기법을 더 잘 구현하지 못한 아쉬움이 남지만 첫 프로젝트를 잘 마무리 했고 많은 것들을 배워 뿌듯하다.

### 오병규

프로젝트를 진행하면서 기획 디테일의 중요성을 알게 되었고, 혼자 코드를 짜는 것과는 다른 협업의 즐거움과 고충을 알게되어 좋은 경험이 되었다.

## 한 줄 평가

### 이희문

혼자서 해왔던 프로그래밍에서 다같이 서로 보완할 수 있는 기회가 되어 많이 배웠습니다.

### 최연웅

QT designer를 활용한 UI 구성에 있어 완벽한 구현을 시키진 못했지만 이외에도 많은 기능을 배운 것 같아 좋은 경험이 되었습니다.

# 데이터 출처

데이콘 ( 서울시 따릉이 대여량 예측 경진대회 )

# 참고 자료

- 01 딥러닝 텐서플로 교과서 (길벗)  
빅데이터 분석 및 인공지능 (인피니티박스)
- 02 PyQt5 유튜브 자료
- 03 PPT 템플릿  
<https://www.canva.com/>



# 감사합니다!

저희 그룹에 궁금하신 게 있나요?

