

FarmTech

소프트웨어융합학과 2018111395 박서영

소프트웨어융합학과 2018111575 김가을

CONTENTS



Contents 01

주제의 목표 및 주요 내용



Contents 02

관련 연구



Contents 03

전체 진도 계획 및 현재 진도 사항



Contents 04

향후 진행 계획

01

Contents 01

주제의 목표 및 주요 내용

농산물 가격 예측 AI 모델 개발 및
이를 활용한 소비 품목 제안 어플리케이션

주요 내용은 크게 세 가지로 나눌 수 있습니다.

데이터 수집
및 전처리

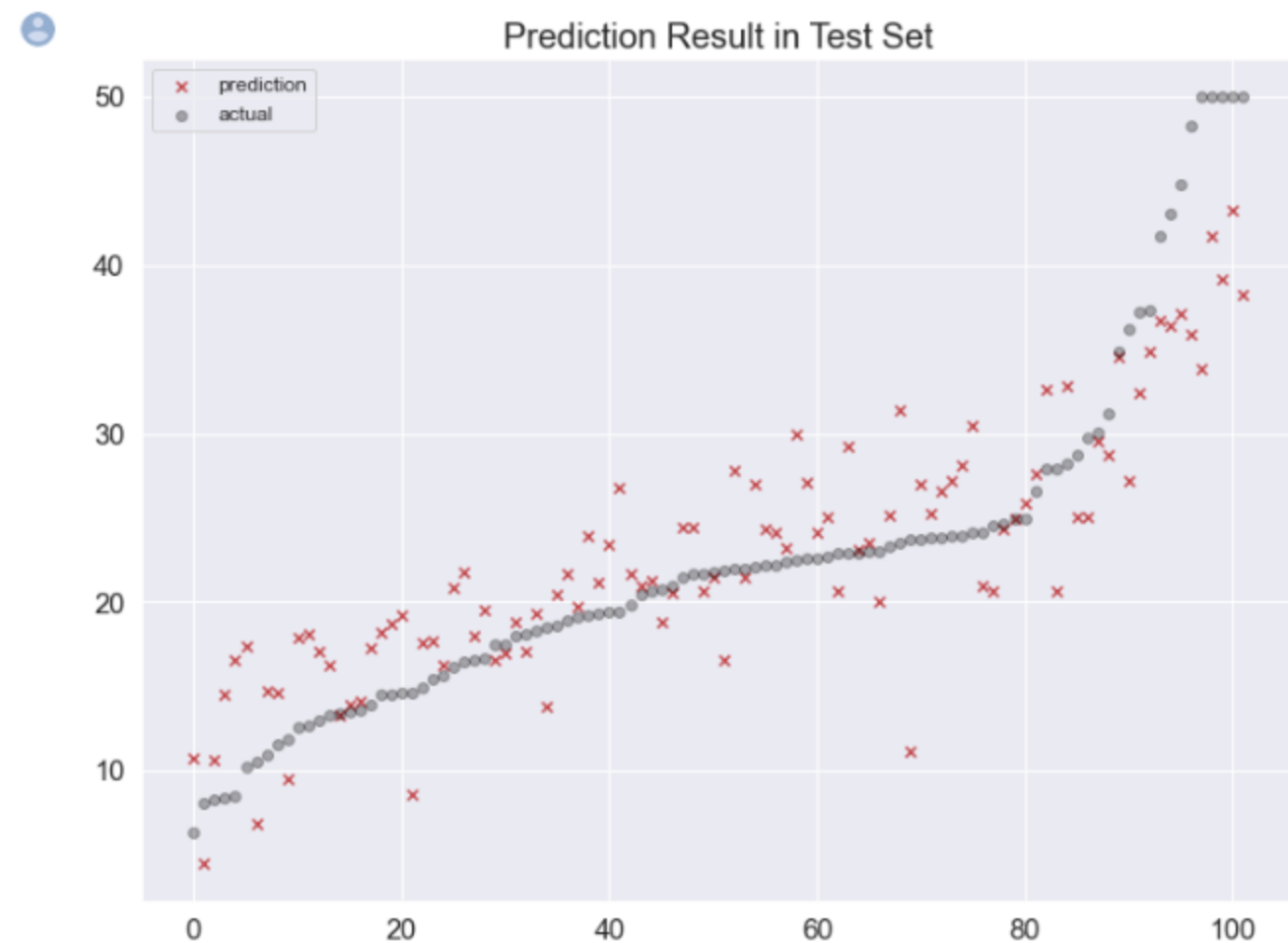
선형회귀와
추천알고리즘을
이용해 AI모델 개발

Tensorflow mobile
or lite(android)
안드로이드
앱 개발

02 Contents 02

관련 연구

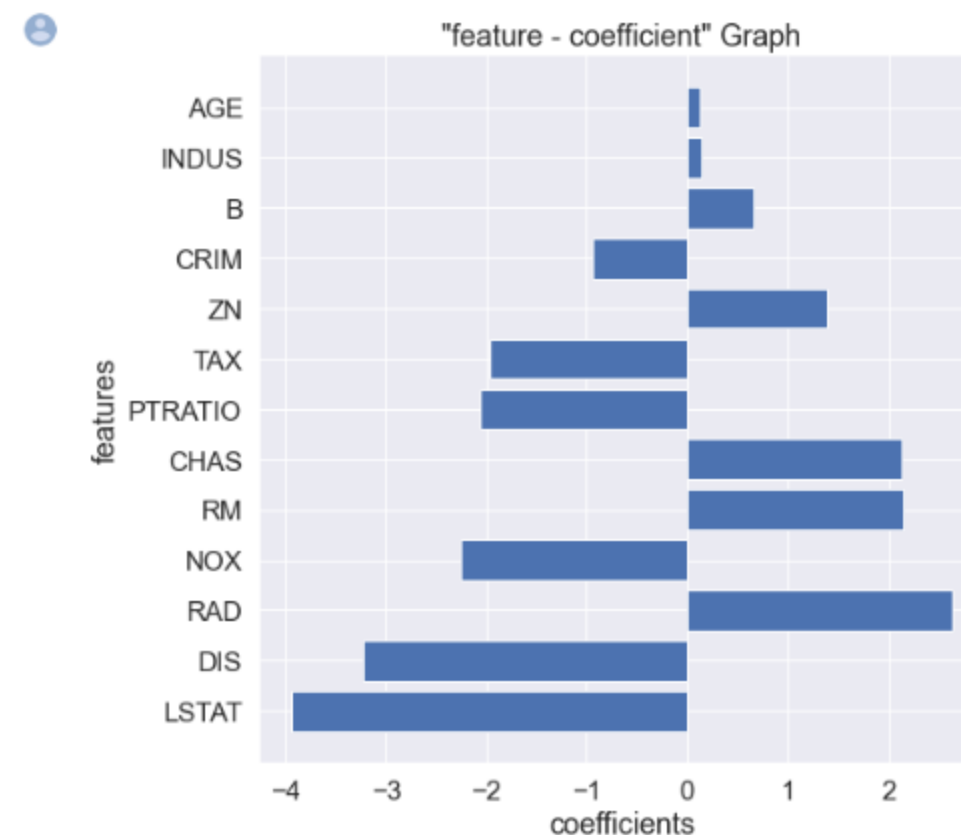
```
plt.figure(figsize=(12, 9))
plt.scatter(df.index, df['prediction'], marker='x', color='r')
plt.scatter(df.index, df['actual'], alpha=0.3, marker='o', color='black')
plt.title("Prediction Result in Test Set", fontsize=20)
plt.legend(['prediction', 'actual'], fontsize=12)
plt.show()
```



- 모델 성능 평가 (R square 와 RMSE)

```
### coefficients 를 시각화 합니다.
### figure size
plt.figure(figsize = (8, 8))

### bar plot : matplotlib.pyplot 모듈의 barh() 함수를 사용해서 수평 막대 그래프를 그릴 수 :
plt.barh(coefs_new['feature'], coefs_new['coefficients'])
plt.title("feature - coefficient" Graph')
plt.xlabel('coefficients')
plt.ylabel('features')
plt.show()
```



- 유의성 검정을 합니다. 통계에서 변수가 유의하다는 의미입니다.

대표적인 빅데이터 처리 문제인 '집 값 예측 모델'을 통해 현재 프로젝트와 연결한다.

03

Contents 03

전체 진도 계획 및 현재 진도 사항

전체 진도 계획

01 EDA

02 Grouping, Clustering

03 선형 회귀

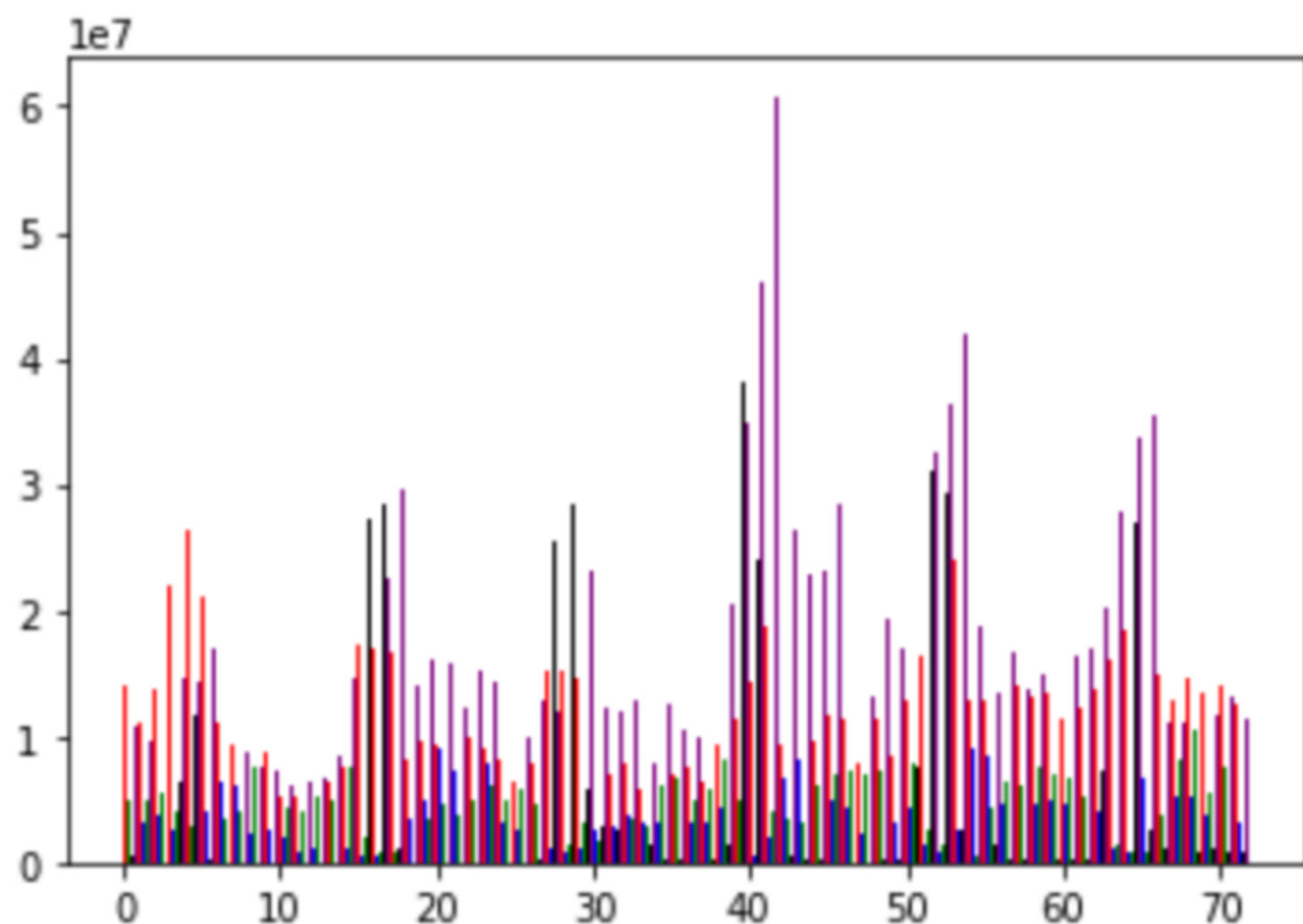
04 AUC 처리

05 랜덤 포레스트(앙상블)

03

Contents 03

전체 진도 계획 및 현재 진도 사항



```
df.fillna(df.mean())
```

	거래일자	양파(전체)	반입량	만생양파	반입량	저장양파	반입량	조생양파	반입량	기타	반입량
0	2015-01-01		13916582		3786154		4935804		401340		10696722
1	2015-02-01		11005449		3175228		5064789		0		9499536
2	2015-03-01		13611617		3660763		5389042		2324		12055334
3	2015-04-01		22077825		2638003		4038779		6362671		14524703
4	2015-05-01		26358838		2842755		2746511		11667414		14394715
...
67	2020-08-01		13006304		5187926		8030901		895903		11205384
68	2020-09-01		14683073		5187342		10586685		910505		11712069
69	2020-10-01		13567706		3826548		5546410		989646		11722141
70	2020-11-01		14104641		4433809		7665432		939633		13126886
71	2020-12-01		12713812		3282039		9478399		807799		11410239

72 rows × 6 columns

```
[220] df.head()
```

	거래일자	양파(전체)	반입량	만생양파	반입량	저장양파	반입량	조생양파	반입량	기타	반입량
0	2015-01-01		13916582		3786154		4935804		401340		10696722
1	2015-02-01		11005449		3175228		5064789		0		9499536
2	2015-03-01		13611617		3660763		5389042		2324		12055334
3	2015-04-01		22077825		2638003		4038779		6362671		14524703
4	2015-05-01		26358838		2842755		2746511		11667414		14394715

양파에 경우 x축은 날짜, y 데이터는 양파 품목별 데이터를 각 막대 그래프에 넣어 visualization을 실행

03

Contents 03

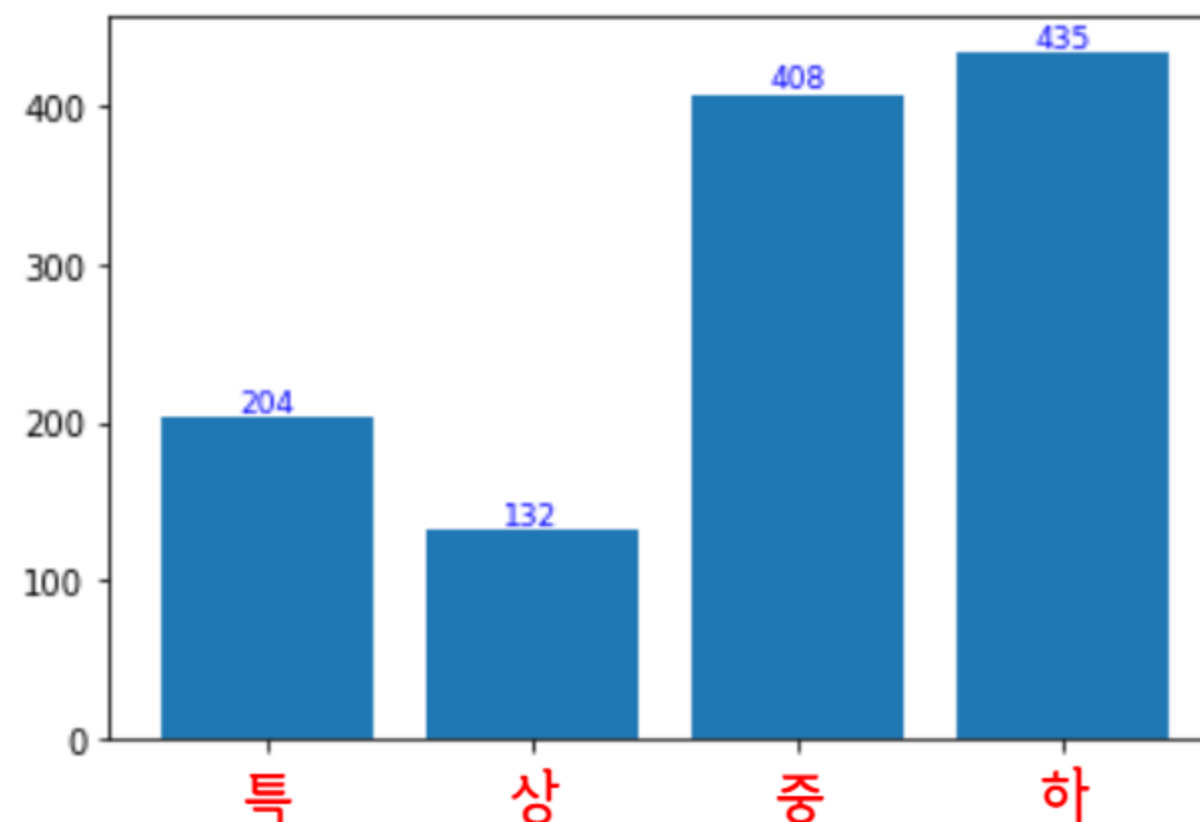
전체 진도 계획 및 현재 진도 사항



데이터 전처리 파트

```
# nan 값이 얼마나 있는지 column별로 확인하기
df.isnull().sum()
#####
```

거래일자	0
봄배추 10.0 kg 상자 /특	44
봄배추 10.0 kg 상자 /상	36
봄배추 10.0 kg 상자 /중	98
봄배추 10.0 kg 상자 /하	105
봄배추 반입량	1
여름배추 10.0 kg 상자 /특	69
여름배추 10.0 kg 상자 /상	8
여름배추 10.0 kg 상자 /중	116
여름배추 10.0 kg 상자 /하	120
여름배추 반입량	1
김장(가을)배추 10.0 kg 상자 /특	80
김장(가을)배추 10.0 kg 상자 /상	77
김장(가을)배추 10.0 kg 상자 /중	112
김장(가을)배추 10.0 kg 상자 /하	120
김장(가을)배추 반입량	6
월동배추 10.0 kg 상자 /특	11
월동배추 10.0 kg 상자 /상	20
월동배추 10.0 kg 상자 /중	82
월동배추 10.0 kg 상자 /하	90
월동배추 반입량	1



[특, 상, 중, 하] 배추 별로 NaN값의 수를 확인했더니 [중, 하] 이 두 품목이 압도적으로 NaN값의 수가 많아서 크게 처리할 필요가 없는 이상치라 판단하여 제거함

03

Contents 03

전체 진도 계획 및 현재 진도 사항



AI 모델 개발 파트

```
[15] # split dataset into training & test
X = df['날짜']
y = df['양파(전체) 반입량']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
```

```
[16] X_train.shape, y_train.shape

((57,), (57,))
```

```
[17] X_test.shape, y_test.shape

((15,), (15,))
```

```
from sklearn import linear_model

# fit regression model in training set
lr = linear_model.LinearRegression()
model = lr.fit(X_train, y_train)

# predict in test set
pred_test = lr.predict(X_test)
```

```
ValueError                                Traceback (most recent call last)
<ipython-input-19-9bad770332ba> in <module>()
      3 # fit regression model in training set
      4 lr = linear_model.LinearRegression()
----> 5 model = lr.fit(X_train, y_train)
      6
      7 # predict in test set
```

선형 회귀 모델 사용을 위해 데이터 대입 시, shape으로 인한 오류 생김
해당 사항 개선 진행 중

04

Contents 04

향후 진행 계획

적은 데이터로 효율적인 학습을 위해 K-fold를 사용하여 활용 가능한 데이터셋을 늘려 학습시키기로 함.
11월 초까지 AUC 처리까지 마무리한 후 최적 알고리즘을 찾아낼 예정

감사합니다