

빅데이터 분석 정의서

주제 : 날씨 데이터를 활용한 농작물 가격예측

2023.08.16

I. 개요

1. 아이디어 주제

: 날씨데이터를 활용하여 농작물의 일주일 후 가격을 예측하기

2. 개발 목표

: 텃밭 중계 사이트에서 재미/정보적인 요소로 활용

3. 개발 내용

: 농작물별로 주산지의 날씨데이터를 통해 학습

: 중기예보를 통해 일주일 후의 날씨데이터를 수집하여 일주일 후의 농작물 가격예측

II. 기능별 빅데이터 분석 명세서

기능명	가격 예측
1. 데이터 준비	
데이터 정의	<p>주산지 한지역의 비중이 최소 70% 이상인 농작물과 그지역의 날씨, 가격 정보</p> <ul style="list-style-type: none"> - 감자, 당근, 대파, 양배추, 양파 총 5가지 작물 각각의 주산지 날씨, 가격, 일주일뒤날씨 총 15개
데이터 획득 방법	<p>1. 농산물유통정보 (https://www.kamis.or.kr/customer/main/main.do)</p> <ul style="list-style-type: none"> - 감자 가격: 2011.01~2023.07 - 당근 가격: 2011.01~2023.07 - 대파 가격: 2012.01~2023.07 - 양파 가격: 2011.01~2023.07 - 양배추 가격: 2013.01~2023.07 <p>2. 농업날씨365 (http://weather.rda.go.kr/w/)</p> <ul style="list-style-type: none"> - 감자 (강원도 평창군 여만리) 날씨: 2011.01~2023.07 - 당근 (경상남도 밀양시 상남면) 날씨: 2011.01~2023.07 - 대파 (강원도 홍천군 화동리) 날씨: 2012.01~2023.07 - 양파 (전라남도 무안군 청계면) 날씨: 2011.01~2023.07 - 양배추 (강원도 강릉 안반덕이) 날씨: 2013.01~2023.07 <p>3. 공공데이터포털 (https://www.data.go.kr/) 기상청_중기예보 조회서비스</p>

- 감자 (평창) 예보날씨
- 당근 (밀양) 예보날씨
- 대파 (홍천) 예보날씨
- 양파 (무안) 예보날씨
- 양배추 (강릉) 예보날씨

1. 농산물유통정보

기간	지역	부류	품목	품종	등급
시작일: 2013-01-0	전체	식량작물	콩	전체	전체
2013 1월	서울	채소류	팥	수미(노지)	상품
6 7 8 9 10 11 12	부산	특용작물	녹두	대지마	중품
13 14 15 16 17 18 19	대구	과일류	매일		
20 21 22 23 24 25 26	광주	수산물	고구마		
27 28 29 30 31	대전		감자		
종료일: 2013-12-3					kg 단위원산 (동그라미 체크)

■ 이용안내

- 도매가격(가락시장)은 **농넷** 등에서 조회하실 수 있습니다.
- 중도매가격은 중도매인 상회에서 소상인 및 실수요자에게 판매하는 가격입니다.
- 농산물 특성상 크기와 색상 등이 다양하고 저장기간, 기후변화에 따라 동일 등급에도 다소 차이가 있을 수 있음.
- 1개월전, 1년전은 해당일자 기준 5일 이동평균 가격임. (이동평균은 해당일 기준 전후로 4-5일의 평균값)
- 평년은 5년간(금년 제외) 해당일에 대한 최고값과 최소값을 제외한 3년 평균값. (해당일 평균가격은 5일 이동평균값을 적용)
- 조사단위가 중량이 아닌 품목(포기, 개, 점, 마리 등)은 kg 단위 원산이 제공되지 않음.
- 본 가격자료는 전국 주요 시장에서 조사된 도·소매 평균가격으로 개별 판매처 및 산지, 브랜드, 규격 등에 따라 가격이 다를 수 있으므로 실거래에 있어서 참고 자료로만 활용하여 주시기 바랍니다.

* (중도매인 판매가격 - 기간별) 식량작물 / 감자 / 수미(노지), 등급: 중품, 단위: 1kg

(단위: 원)

데이터저장

구분	01/02	01/03	01/04	01/07	01/08	01/09	01/10	01/11	01/14	01/15	01/16	01/17	01/18	01/21	01/21
평균	1,210	1,210	1,220	1,290	1,290	1,290	1,280	1,280	1,280	1,280	1,280	1,270	1,270	1,200	1,1

수집 데이터

2. 농업날씨365

· 관측지점: 강원도 평창군 여مان리

· 기간구분: ☒ 년 ☐ 월 ☐ 일 ☐ 기간

· 관측기간: 2013 년

Q 조회

수치자료

그래프자료

농업기상기간별자료

인쇄

저장하기

10분별	시간별	일 별	반순	순별	월별		
날짜 (년-월-일)	평균기온 (℃)	최고기온 (℃)	최저기온 (℃)	습도 (%)	풍향 (방향)	풍속 (m/s)	강수량 (mm)
평균	9.8	16.3	4.0	72.7	-	1.2	3.7
최대	27.0	32.8	24.1	99.4	-	3.8	138.0
최저	-17.7	-9.7	-24.5	32.4	-	0.4	0.0
합계	-	-	-	-	-	-	1346.0
2013-01-01	-11.1	-4.2	-17.9	85.5	서	0.5	1.5
2013-01-02	-10.3	-3.3	-18.6	49.1	서	2.3	2.0
2013-01-03	-17.7	-9.7	-23.6	64.3	남남동	1.1	0.0
2013-01-04	-17.4	-6.6	-24.5	69.0	남	0.6	0.0
2013-01-05	-14.1	-2.9	-23.1	71.5	서	0.8	0.0

3. 공공데이터포털

최저/최고기온(℃)

그래프를 클릭하면 최저/최고기온과 변동성 정보를 확인하실 수 있습니다.

지역	20일 (일)	21일 (월)	22일 (화)	23일 (수)	24일 (목)	25일 (금)	26일 (토)	27일 (일)	기온범위
서울 인천 경기도	서울	25 / 31	25 / 31	25 / 30	25 / 31	24 / 31	24 / 31	24 / 31	그래프
	인천	24 / 30	25 / 30	25 / 29	24 / 30	24 / 30	24 / 30	24 / 30	그래프
	수원	24 / 32	24 / 31	24 / 30	24 / 31	24 / 31	24 / 31	24 / 31	그래프
	파주	23 / 31	23 / 31	23 / 29	23 / 29	23 / 31	22 / 31	22 / 31	그래프
	이천	23 / 32	23 / 31	23 / 30	23 / 30	23 / 31	23 / 31	23 / 31	그래프
	평택	23 / 32	23 / 32	24 / 30	24 / 30	24 / 31	24 / 31	24 / 31	그래프

2. 전처리

전처리 과정

예전날씨와 가격은 엑셀파일로 받음

가격의 날짜가 2023년으로 되어있어 해당년으로 바꿈

날씨가 필요없는 내용으로 묶여있어 제거함

4년에 한번 있는 2월29일을 제거함

중간에 정보가 비어있는 경우 제거함

중기예보를 통해 얻을 수 있는 최고/최저 기온을 제외한 데이터를 제거함

전처리 데이터

5]: 1 Prepro('감자', 2013, 2023)

C:\Users\ggjaischool\AppData\Local\Temp\ipykernel_6736
A value is trying to be set on a copy of a slice from
Try using .loc[row_indexer,col_indexer] = value instea

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/tutorials/10min/05-indexing.html>

df_price['평균'] = df_price['평균'].apply(remove_co

C:\Users\ggjaischool\AppData\Local\Temp\ipykernel_6736
A value is trying to be set on a copy of a slice from

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/tutorials/10min/05-indexing.html>

df_weather.replace("-", 0, inplace=True)

5]:

날짜	최고기온	최저기온	월	일	연도	가격
2013-01-02	-3.3	-18.6	1	2	2013	1210.0
2013-01-03	-9.7	-23.6	1	3	2013	1210.0
2013-01-04	-6.6	-24.5	1	4	2013	1220.0
2013-01-07	-1.3	-17.7	1	7	2013	1290.0
2013-01-08	-0.2	-17.1	1	8	2013	1290.0
...
2023-08-09	28.8	21.4	8	9	2023	1712.0
2023-08-10	22.9	20.5	8	10	2023	1702.0
2023-08-11	29.8	20.3	8	11	2023	1712.0
2023-08-14	32.8	21.1	8	14	2023	1714.0
2023-08-16	31.4	19.4	8	16	2023	1691.0

2614 rows × 6 columns

3. 데이터 분석	
데이터 분석 목표	일주일 후의 날씨데이터를 가지고 해당 일의 해당 작물 가격예측
데이터 분석 시나리오	이전까지의 날씨와 가격데이터를 학습시키고, 중기예보를 통해 일주일 후의 최대/최소가격을 가져온 후 이를 입력하여 원하는 작물의 가격을 예측한 값을 얻기
데이터 분석 결과	미래정보로 얻을 수 있는 최고/최저, 강수확률 중에서 불정확한 강수확률은 제거하고 최고/최저 기온만 사용
4. 모델 생성 및 학습	
모델링 목표	원하는 날씨의 작물가격 예측
학습 모델	DecisionTreeRegressor, RandomForestRegressor, ExtraTreesRegressor
학습 방법	3가지 모델에 학습시켜보고 결과가 좋은 모델을 사용
모델 예측 결과	<pre>In [54]: 1 machine('감자')</pre> <p>Decision Tree R2 Score: 0.9218191768129006 Random Forest R2 Score: 0.9671206439331559 Extra Trees R2 Score: 0.981710955724767 Decision Tree MSE: 31065.544342507645 Random Forest MSE: 13064.777936697246 Extra Trees MSE: 7267.243970489298 Decision Tree RMSE: 176.25420375840017 Random Forest RMSE: 114.30125955866474 Extra Trees RMSE: 85.24813177125525</p>
5. 검증	
분석 검증 방안	<pre>26 print(m1.predict([[HT, LT, month, day, year]])) 27 print(m2.predict([[HT, LT, month, day, year]])) 28 print(m3.predict([[HT, LT, month, day, year]]))</pre> <pre>In [59]: 1 machine_p('감자',29,20,8,24,2023)</pre> <p>0.9253771462822343 0.9661982566384242 0.9799946735397579 [1691.] [1779.26] [1784.14]</p>
분석 평가 결과	테스트케이스의 오차가 RMSE내의 범위로 원하는 오차범위내의 결과가 나왔다.