**2023. 05.30(화) ~ 10. 27(금) / 시상식 11.30(목) \* 일정은 상황에 따라 변경 될 수 있습니다.**

* 사과의 품질(크기, 단맛, 쓴맛, 무른 정도 등)은 지역별로 차이가 나는지, 난다면 어떤 부분이 어느 정도로 차이가 나는지 기상(날씨)을 이용하여 분석합니다.
* 사과의 품질(크기, 단맛, 쓴맛, 무른 정도 등)은 어느 시기(특정 계절, 월, 일 등)의 어떤 기상요소(온도, 일사, 강수량 등)가 가장 큰 영향을 미치는 지 찾아서 사과의 품질을 예측하는 AI 모델을 개발해 주십시오.

1. 사과 종류 2가지로 나눠서 2개의 데이터프레임 만들기
2. 논문을 토대로 사과의 품질 기준 분류 - 타겟 변수 만들기 (논문 찾아보기)
3. 8개 지역 기상 데이터 가져오기(군위, 청송, 영주, 장수, 거창, 충주, 포천, 화성)

- 월별 최저기온, 최고기온, 평균기온, 평균풍속, 최대풍속, 일조시간, 평균 강수량, 최대 강수량, 습도, 일사량, 일조율, 풍향, 평균지면온도

1. 새로운 변수 생성
2. 수확 ‘월’ 별로 분류 변수 (화아분화/성숙착색/자발휴면) -> 연속형 변수
3. 홍수, 장마기간, 폭염등 이상기후, 자연재해 유무 변수 (홍수, 장마 기간, 태풍 ,폭염 횟수)
4. 지역마다 수확시기에 영향이 있기 때문에 지역&수확 ‘월’을 합쳐 칼럼을 생성한다

<EDA>

0. 품질이 지역별로 차이가 나는지, 난다면 어떤 부분이 어느 정도로 차이가 나는지 확인 (제공해준 사과 품질 데이터로 EDA)

1. 봄철 기온이 낮으면 착과 수와 과실의 무게에 부정적 영향

2. 과실 수확기의 많은 강수량은 과실 비대에 긍정적 영향

3. 과실 비대기의 고온은 과실 비대에 부정적 영향

4. 여름철 일조 시간은 과실 비대에 좋은 영향을 주며, 당도가 높아짐

5. 수확기 최저 기온이 높으면 병충해로 인해 과실의 품질이 저하됨

<전처리>

* 모델링 1,2번 방법에 따라서 변수 상관관계 분석 후 최종적으로 사용할 변수 선택

1. 결측값, 이상치, 중복값 처리
2. 정규화/표준화
3. 인코딩 ( 원핫, 라벨, 랭크인코딩 등)
4. 일자데이터변경(datetime 타입)
5. 아웃라이어 확인 및 제거결정
6. 자료의 불균형 해소를 위한 오버샘플링 or 언더샘플링
7. train / validation / test 분리

<모델>

1. 상관분석에 따른 모델: 사과의 품질과 각 월별 기상 요인의 상관 관계를 분석

* 아래 2번의 데이터프레임에서 corr 메서드를 사용해 사과 품질 칼럼과 다른 열과의 상관계수 추출 (이때 피어슨 상관계수는 두 변수가 모두 연속형 변수일 때만 사용 가능하므로, 스피어만 상관계수나 켄달의 타우를 구해야 함, 이는 순위형 변수의 상관계수를 구하는 것)
* 상관계수가 가장 큰 달 5개와 각 달의 기상요소 한가지를 뽑는다.

1. 요인 분석에 따른 모델

* 요인 분석은 관측 변수 간의 분산을 설명하고 관측 변수 집합을 요인이라고 하는 관측되지 않은 변수로 압축하는 데 사용된다. 요인 또는 잠재 변수는 공통적인 반응 패턴을 갖는 여러 관측 변수와 관련이 있다. 각 요인은 관측된 변수의 특정 분산 양을 설명한다. 요인 분석은 변수의 수를 줄여서 데이터 해석에 도움이 된다.
* KMO 검정은 요인 분석에 대한 데이터의 적합성을 측정한다. 각 관측 변수와 전체 모형에 대한 적합성을 결정한다. KMO는 모든 관측 변수 간의 분산 비율을 추정한다. KMO 값의 범위는 0에서 1 사이이다. 0.6 미만의 KMO 값은 부적합한 것으로 간주된다.
* 관련 코드

import pandas as pd

from sklearn.datasets import load\_iris

from factor\_analyzer import FactorAnalyzer

import matplotlib.pyplot as plt

from factor\_analyzer.factor\_analyzer import calculate\_kmo

kmo\_all,kmo\_model=calculate\_kmo(df)

* 요인 수 선택

fa = FactorAnalyzer()

fa.analyze(df, 25, rotation=None)

ev, v = fa.get\_eigenvalues()

ev

* 요인 분석 수행

fa = FactorAnalyzer()

fa.analyze(df, 6, rotation="varimax")

fa.loadings

* LSTM : 위 모델은 시계열 데이터에 대한 예측 성능이 우수하기 때문에, 시계열 기상 데이터에 적합할 것으로 예측하여 선정함
* AdaBoost : 위 모델은 과적합 발생 확률이 낮고, 모델 수행 시간이 빠르며, 우수한 성능을 나타내기 때문에 선정
* XGboost : 위 모델은 현재 kaggle과 같은 데이터 분석 사이트에서 우수한 예측 성능을 보이고 있어 선택

<성능 향상 추가 전처리>

1. 변수 가중치

운영체제: Windows 11

패키지 : NumPy, Pandas, Matplotlib, Scikit-learn, seaborn

프로그래밍 언어: Python

* 총 정리 시각적 자료 만들어서 넣기

<예상하는 결과물 형태와 내용>

* 예시 데이터 프레임을 표로 만들기

<1 - 처음 기상 데이터프레임>

|  | 1월~12월 최저기온 | 1월~12월 최고기온 | … |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 지역+  수확년도 |  |  |  |  |  |  |  |  |
| ex) 2021 충주 |  |  |  |  |  |  |  |  |
| 2020 영주 |  |  |  |  |  |  |  |  |

<2 - 기상 데이터프레임 + 사과 품질 데이터>

|  | 1월~12월 최저기온 | 1월~12월 최고기온 | … |  |  |  |  | 사과 품질 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 지역+  수확년도 |  |  |  |  |  |  |  |  |
| ex) 2021 충주 |  |  |  |  |  |  |  | 상 |
| 2020 영주 |  |  |  |  |  |  |  | 특 |

<3 - 새로운 변수 추가 or 상관분석 후 변수 수정?>

<4 - 최종 프레임>

|  | 과실 비대기 최고기온 | 여름철  일조시간 | … |  |  |  |  | 사과  품질 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 지역+  수확년도 |  |  |  |  |  |  |  |  |
| ex) 2021 충주 |  |  |  |  |  |  |  |  |
| 2020 영주 |  |  |  |  |  |  |  |  |

사과품종별로 나눠서 두 가지 예측값 추출

타겟변수인 ‘사과의 품질’ 은 특, 상 , 일반으로 구분된다. 분류 예측

칼럼중요도로 어떤 기상요소가 가장 큰 영향을 미치는지 분석

결과물의 내용: 재배 지역과 그 지역의 기상 데이터를 토대로 그 해 사과의 품질을 예측

<모델 기대 수준 >

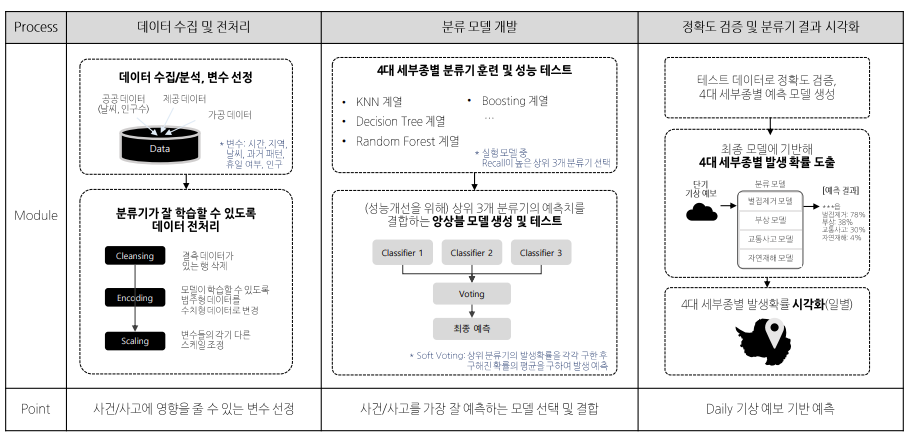
해당 모델의 성능 평가 지표로는 정확도(accuracy)를 사용할 것이다. 모델 개발을 통해 기대하는 수준은 정확도 0.7~0.9사이로 설정하였다. 별도의 test data가 제공되지 않는다면, 주어진 데이터를 이용하여 검증 데이터세트를 설정해 학습 과정 중 성능을 측정할 것이다.

해당 모델을 이용하여 높은 정확도를 통해 소비자와 공급자의 신뢰를 높일 수 있을 것이다.

예측 결과를 바탕으로 지역별 농업 생산성 향상 방안을 도출합니다. 어떤 지역에서 어떤 시기에 특정 기후 조건이 가장 적합한 사과를 재배해야 하는지 등을 제안합니다. 농업 생산뿐만 아니라 소비자 입장에서도 예측 결과를 활용하여 사과 구매에 도움을 주는 정보를 제공할 수 있습니다.

accuracy, recall, precision, f1score 등

* 참고 자료



* **해야 할 것**
* ~~선택 이유(현재 사회적 이슈 , 사과에 관한 내용 반영하여 생각/ 공모전 문제 의도파악)~~
* ~~활용 계획 및 기대 효과~~
* ~~논문찾아보기(타겟변수 분류 기준, 월별 특징 담은 변수 기준)~~
* ~~표 만들기~~
* ~~모델링 구체화(상관 분석, 요인분석)~~

—---------------------25 자정에 중간점검

* 2,3번 계획서 총정리~~~~/ 총정리 시각적자료 만들기

3

* 선호 사과 품질 추천시스템 추가
* 농업하시는 분들에게 앞으로의 개선 방향을 제안해주는 ai 모델
* **선택이유**

해당 문제를 선택한 이유는 사과가 세계적으로 많이 재배되고 소비되는 중요한 작물 중 하나이며, 기후 변화와 관련하여 사과의 품질에 큰 영향을 미치기 때문입니다. 사과의 품질과 기후 요소들 간의 상관관계를 파악함으로써 농가들은 더욱 효율적인 생산 방법을 도입하고 농작물 관리에 있어 더 나은 결정을 내릴 수 있습니다. 또한, 소비자들은 예측 가능한 품질의 사과를 구매함으로써 만족도가 향상되며, 이는 사과 시장에 긍정적인 영향을 끼칩니다. 따라서 농가와 소비자 모두에게 다양한 이점을 제공할 수 있어 의미있는 연구라고 생각하여 이 문제를 선택하였습니다.

+) 문제 선택 이유와 활용 방안에 겹치는 내용이 많아져서 다 활용방안으로 뺐거든…? 그래서 조금 짧ㅇ ㅏ…..ㅓ…….

+) 지연이가 말한 것 처럼 기후데이터를 활용하여 분석하는 것에 대한 '흥미'도 넣을지말지 고민

< 최근 기사에 따르면, –한 기후 변화가 사과의 품질에 영향을 미쳤다고 한다. 이뿐만 아니라 최근 이상 기후 현상을 보이며 사과 재배에 어려움을 겪고 있다. 따라서 ~ > 이런거 추가

* **활용방안**

개발한 AI 모델은 다음과 같은 방안으로 활용될 수 있습니다:

* 농업 생산성 향상: AI 모델을 농가들에게 제공하여 사과 품질에 영향을 미치는 최적의 기상 요소와 시기를 알려줌으로써 생산성을 향상시킬 수 있습니다. 이를 통해 농가들은 더 나은 사과를 생산하고 농작물 관리를 최적화함으로써 수익을 증대시킬 수 있습니다.
* 자원 효율성 증대: AI 모델을 활용하여 효과적인 농작물 관리 방안을 제시함으로써 물, 비료 등의 자원을 효율적으로 사용할 수 있습니다. 이는 환경 보호와 더 지속 가능한 농업을 실현하는 데 도움이 됩니다.
* 소비자 만족도 향상: AI 모델을 사용하여 사과 품질을 예측하고 향상시킬 수 있으므로, 소비자들이 더 맛있고 신선한 사과를 즐길 수 있습니다. 이는 소비자 만족도를 높이고 사과 시장에서의 경쟁력을 강화하는 데 도움이 됩니다.
* 농작물 관리와 위기 대응: AI 모델을 활용하여 사과 품질 예측과 기후 변화에 대응하는 방안을 제시함으로써 농가들은 예상치 못한 기후 변화에 대응할 수 있습니다. 이는 농작물 관리와 생산 안정성을 향상시키는 데에 도움이 됩니다.

향후 연구 방향은 다음과 같습니다:

* 기후 예측 데이터 포함: 기후 예측 데이터를 모델에 포함하여 더 정확한 사과 품질 예측을 가능하게 합니다. 과거의 기상 데이터 뿐만 아니라 미래의 기후 예측 정보를 활용함으로써 모델의 성능을 향상시킬 수 있습니다.
* 개별 농가에 맞는 맞춤형 조언 제공: AI 시스템은 개별 농가의 생산 환경과 조건을 고려하여 맞춤형 조언을 제공해야 합니다. 각 농가의 특성에 따라 최적의 기상 요소와 관리 방안을 추천함으로써 개별 농가의 생산성을 최대화할 수 있습니다.

이러한 방향으로 연구를 진행함으로써 사과 산업과 농가, 소비자들에게 더 큰 이익과 가치를 제공하는 AI 모델을 개발할 수 있을 것입니다.

+) 내용이 너무 난잡하고 길지는 않은지… 더 넣고 싶은 내용이 있다면 말해줘… 고쳐둘겟!

+) 생산량의 문제가 아니라 품질의 문제이기 때문에 농업 생산성은 빼도 될거 같기도…

* **지연) 월별 특징 담은 변수 기준 논문**

의견1) 어차피 주어질 데이터가 수확 데이터니까 날짜가 12월~ 5월 까지는 없을 것으로 예상됨. ㅇㅇ,, 그래서 수확 가능성이 있는 달을 크게 잡았을 때 6월~ 11월 일 것 같아서 그 사이만 있어도 될듯.

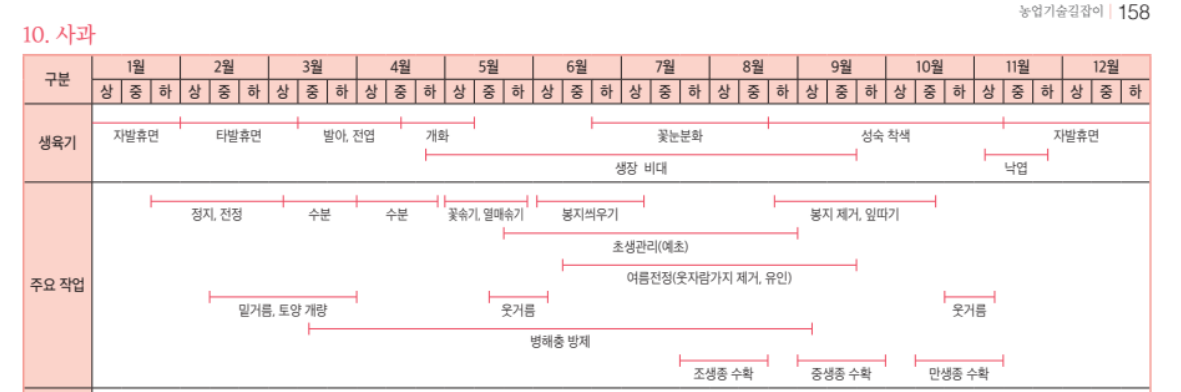
의견2) 논문 찾다가 보니까 지역별로 다를 수 있대. 그래서 지역&달(월) 이렇게 봐야지 의미가 있지 달(월) 하나만 보는 것도 의미가 있나?싶긴함.

그래서 만약 우리 전처리 할 때 지역과 월을 합친 칼럼 하나 만들어도 괜찮을듯

서울 9 / 서울 10/ 서울 11/ 청주 9/ 청주 10/ 청주 11 이런식으로

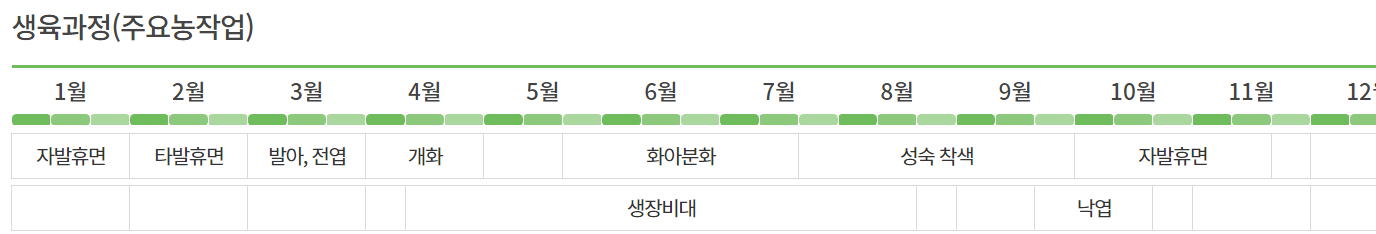
* 농촌 진흥청 자료 (2019)

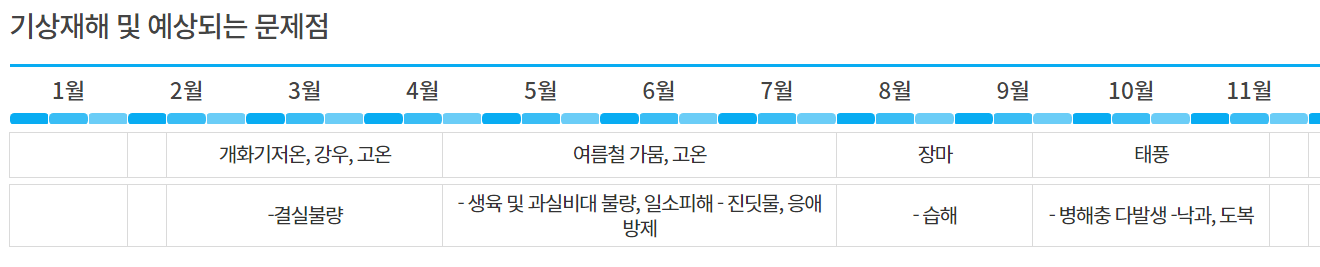
[농업기술길잡이 문서뷰어 | 농사로 (nongsaro.go.kr)](https://www.nongsaro.go.kr/portal/ps/psb/psbx/cropEbookFileViewPop.ps?indexPage=158&indexBasePage=-1&cropsEbookFileNo=00001&ebookCode=119)



* 농사로 자료(이건 홈페이지라.. 이게 더 갱신된 걸수도?)

[사과 | 농사로 (nongsaro.go.kr)](https://www.nongsaro.go.kr/portal/ps/psb/psbl/workScheduleDtl.ps?menuId=PS00087&cntntsNo=30663&sKidofcomdtySeCode=210002&totalSearchYn=Y)





* 11월, 12월, 1월, 2월 휴면상태

- 3월은 발아기

- 4월은 개화기

- 5월은 착과기 (개화 및 생장 비대기)

- 6월, 7월, 8월은 비대기

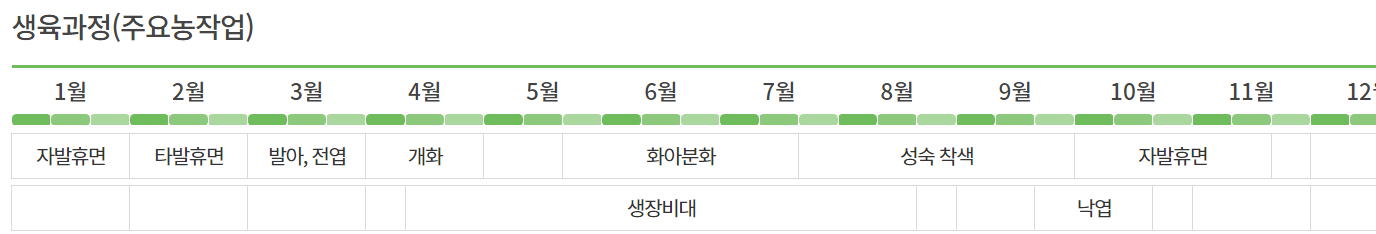
- 9월, 10월은 성숙 및 착색기

크게는 이렇게 나눌 수 있는데, 우리가 수확 날자만 있다면(8월~ 11월처럼 적으면)

더 세분화 하는 게 좋을 듯.

<세분화 하면>

여기서 화아분화/성숙착색/자발휴면 으로 나누기



혜현조사 부분



=> 우리가 얻는 사과의 정보는 과중, 종경, 횡경, L/D 비율, 경도1, 경도2, 경도 평균, 당도, 산도, 착색 정보 3가지

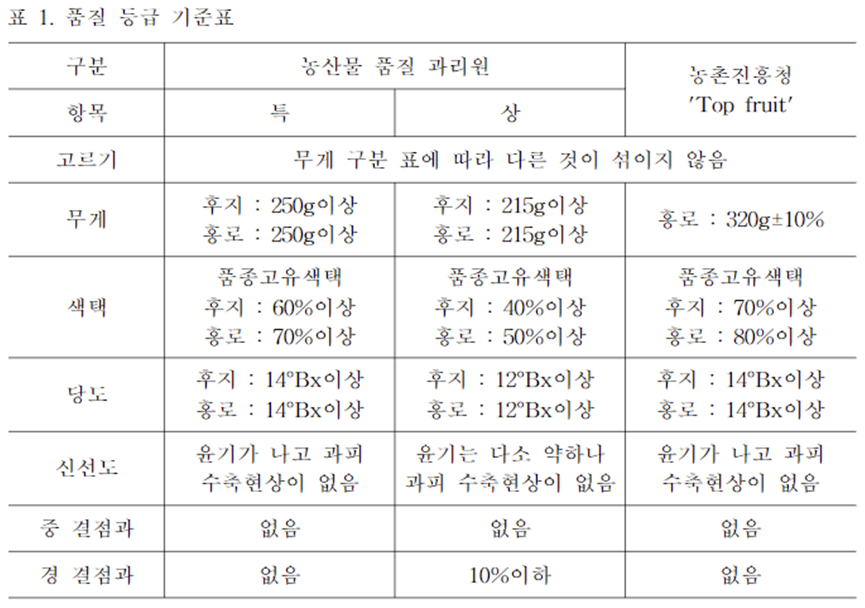
· 과중(g): 과실의 무게 · 종경(mm): 세로 길이, 과일의 크기

· 횡경(mm): 가로 길이, 과일의 크기 · L/D: 종경과 횡경의 비율, 과일의 형태

· 경도(N/ø11mm): 단단하거나 무른 정도 · 당도(˚Brix): 당의 함량, 단맛

· 산도(%): 산의 함량, 쓴맛 · 착색(Lab): 과일 색을-60 ~ 60으로 표현

**<방법 1> - 도현이가 조사해준 논문**



=> 후지와 홍로를 따로 구분할 수 있음

=> 과중, 종경, 횡경, L/D 비율, 경도1, 경도2, 경도 평균, 당도, 산도, 착색 정보 3가지 이용 가능

**<방법 2> - ‘소비자의 선호를 반영한 등급 표준화’라는 논문 참고**



=> 후지에 대한 자료만 있음

=> 과중, (종경, 횡경, L/D 비율) => 모양, (경도1, 경도2, 경도 평균) => 경도, 당도, 산도 => 당산비, 착색 정보 3가지 고려 가능

+) 국립농산물품질관리원의 농산물 표준 규격

<https://www.naqs.go.kr/contents/contentsTab.do>

-=> 농산물 표준 규격을 살펴보니 상자를 단위로 평가해서 이는 힘들듯

<https://fruit.nihhs.go.kr/apple/qlityInfo_frutQlity.do>

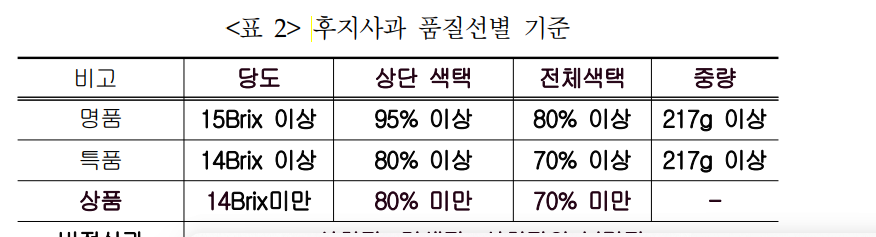
**0726 자정 회의**

<할 것>

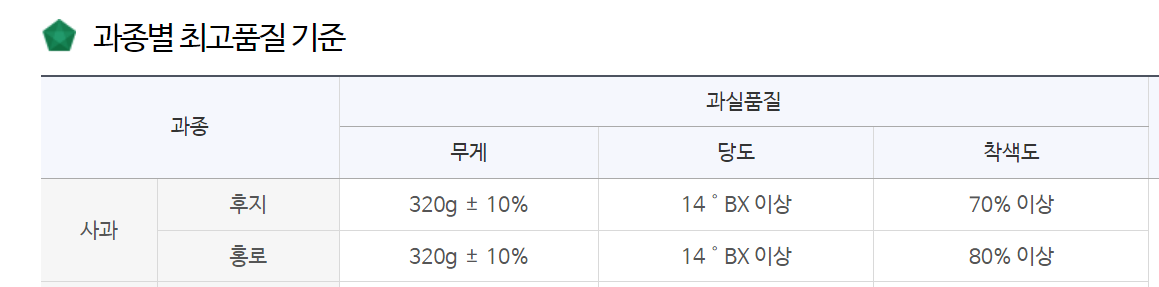
* ~~논문 더 찾아보기(타겟칼럼) 사과 품질 어떻게 구분할지 -지연수미~~
* ~~결과물 어떻게 구성할건지 -혜현도현~~

**논문 더 찾아보기(타겟칼럼) 사과 품질 어떻게 구분할지 -지연수미**

이것도 후지사과만 있네..



[RISS 검색 - 학위논문 상세보기](http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=494b2fa1cf0cbcebffe0bdc3ef48d419&keyword=%EC%82%AC%EA%B3%BC%20%ED%92%88%EC%A7%88%20%EB%93%B1%EA%B8%89)

ㅠ

[탑프루트란? - 농업기술센터 (chungju.go.kr)](https://www.chungju.go.kr/cja/contents.do?key=1050)

찾아도 뭐가 막 더 잘 나오진 않았음 ㅜㅜ여러개 검색해도 후지 , 홍로 같이 있는 것은 무게, 당도, 착색도만 나와있는듯. 농촌진흥청 자료가 2006년부터 뭘 수집해서 결정했대 글서 거기 맞춰서 다 이 ㅅㅔ개만 쓴듯

방법 1) 칼럼 별로 히스토그램을 그려 (예) 상위 30%안에 드는 것들을 특으로 분류한다.

방법 2) 혜현이가 어제 말한 1번 방법으로 (알아본 것들 중 최신 자료로) 세가지 조건만을 and로 연결시키고 나머지 조건에서 outlier만 뽑아낸다.

-> 이상치가 발견되면 …

무른정도가 정말 너무 무르거나 너머어어어무 단단한 이상치.. 이상한애들…

하나의 사과는 무른정도랑 가로길이가 이상치야. 이상치 개수=2 ,,...

이상치가 품질 등급 (특/상/보통) 에 무슨 영향을 줘야될까? 한단계씩 낮추기…?

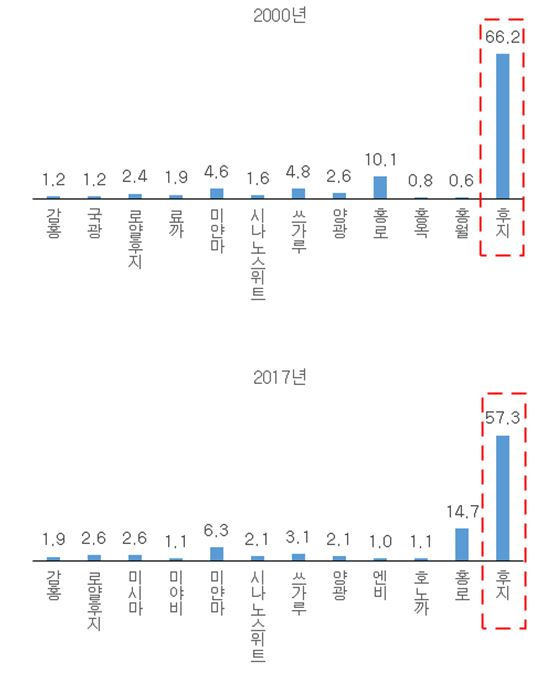
방법 3) 후지는 구분 기준을 많이/ 홍로는 3개만.. (서로 다른 기준으로 분류.. )

후지가 한국 사과의 약 57%, 홍로는 약 15% -> 소비자의 수요가 많을거아냐 -> 소비자의 선호도를 고려한 사과 품질 등급 기준을 사용 -> 이거에 따라서 후지는 나누고

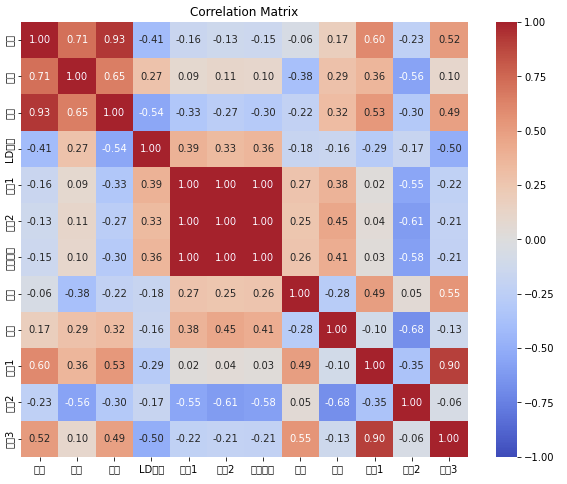
홍로는 비교적 적은 비중 -> 소비자 수요 기준에 따른 조사가 불충분하여 농촌진흥청 자료 활용

* 사과 품종 비율 뉴스 기사

<http://www.fsnews.co.kr/news/articleView.html?idxno=28554>



+) 일단 주어진 데이터만을 활용했을 때는 경도1,경도2,경도평균의 correlation이 거의 1이기 때문에 하나만을 사용해도 될듯….



+) 칼럼별로 히스토그램도 그려보긴했는데 표본이 너무 적어서 의미 없음.

+) 그리고 사과가 동그란게 좋지 않을까 해서 L/D는 1에 가까운게 좋지 않을까... 근거는 못 찾음....

그리고 이런게 있긴한데 별 도움 안될듯… 혹시 아이디어에 도움이 될까 해서 첨부했어…

* 전북 장수 사과 당도 품질 데이터
* <https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=490>
* 기상요인이 사과 '홍로'의 과실품질에 미치는 영향

<https://www-dbpia-co-kr-ssl.access.ewha.ac.kr/journal/articleDetail?nodeId=NODE06135918>

* 착색에 씁향을 주는 안토시아닌 함량은 수확 1달 전인 8월의 최고기온이 가장 씁향을 미쳤는데 8월중 최고기온이 낮을수록 안토시아닌 함량이 높아지는 부의 상관을 나타내었다.
* 과실의 경도는 5월부터 8월까지의 평균기온이 낮을수록 경도가 높아지는 부의 상관을 나타내었다.
* 당도는 5월부터 8월까지의 최저기온이 낮을수록 당도가 올라가는 부의 상관을 나타내었다.

결과가 위에 나온 것 처럼 8월 최저기온이 낮을수록 당도가 높아지기 때문에 품질이 좋아진다. 처럼 나와야 하는게 아닐까 싶기도 해…. 8월 최저기온이 품질(우리가 새로 만든 변수)에 영향을 끼친다. 까지는 될 거 같은데 이게 품질에서 당도에 영향을 끼친다 를 보일 수 있나???

**<결과물>**

XAI 알고리즘 (https://blog.naver.com/pbk11260/222143017876)

: 인공지능 모델이 특정 결론을 내리기까지 어떤 근거로 의사결정을 내렸는지를 알 수 있게 설명 가능성을 추가하는 기능

XAI 알고리즘의 SHAP 기법

: XGBoost의 feature importance는 변수간의 의존성을 간과한다. 하지만 SHAP은 의존성까지 고려해 모델에 미치는 영향력을 계산해준다. 또한 음의 영향력도 계산할 수 있어 넓은 범위를 볼 수 있음. 따라서, 더욱 적합한 설명력을 가짐

우리의 독립변수들은 기상데이터이기 때문에 변수간의 의존성이 약간 존재할 수 밖에 없음. 따라서 SHAP을 사용하여 변수의 영향력을 살펴보고, 사과 품질에 영향을 미친 feature를 제공

1. 사과 품질 예측 모델
2. 한 해의 기상 데이터를 입력한 경우

=> 사과의 품질(상, 특, 탑, 보통 등), 사과 품질의 비율 그래프(원 그래프 출력), feature importance를 통해 사과의 품질에 영향을 미친 요인 알려주기.

=> 이를 통해 얻을 수 있는 것은 농가들에게 제공하여 사과 품질에 영향을 미쳤던 최적의 기상 요소와 시기를 알려줄 수 있고 내년 농업에 반영할 수 있음.

1. 여러 해의 기상 데이터를 입력한 경우

=> 사과의 품질 비율을 비교해주며 사과의 품질이 향상되고 있는지 떨어지고 있는지, 각각의 주된 이유를 함께 알려주기(feature importance로?)

=> 이전 기후 데이터 하나를 넣었을 때는 해당 연도의 사과 품질에 대해서만 알 수 있었다면, 이를 통해 사과 품질의 흐름을 파악할 수 있음

2. 수확하기 전, 사과를 기른 날까지의 기상 데이터를 통한 사과 품질 예측 ai모델

1월을 거쳐 발아기부터 수확기까지 기상 데이터가 누적되어갈 때 품질에 크게 영향을 미친 부분이 있었다면,

해당 기상요소를 거쳤으나 품질이 좋았던 과거 데이터를 토대로 앞으로의 대처 방안을 계획할 수 있도록 함

-> 예를 들어 어떤 ’월’이 품질에 영향을 미치는지 사전 분석 후, 농부가 사과를 기른 1월~5월까지 기상 데이터를 넣었을 때 3월의 평균 기온이 기준치보다 낮았다면, 기존에 학습한 모델에서 3월의 평균 기온이 낮았던 데이터들을 모아 그 사과들의 품질을 분류하고, 높은 품질이 있다면 어떤 기상 요소가 다시 품질을 향상시켰는지를 분석해 대처 방안으로 사용하도록 함.

1. 상관분석을 통해 사과 품질에 영향을 미치는 각 월의 기상요인 분석

-> ex) 1월 습도, 2월 최저온도, 3월 최고온도, 4월 강수량

1. 주어진 데이터에서 사과의 품질별로 그룹을 나눈 뒤, 1)의 기상요인이 어떻게 분포되어있는지 확인

-> ex) 특/상/보통 집단의 기상요인(습도)에 대한 시각화 진행 후 두드러진 기준 설정

1. 품질을 나누는 기상요소(ex 습도)의 기준치를 정한다.
2. 입력한 데이터에서 중요한 기상요인이 기준치 이하인 것이 있는지 확인한다.
3. 있을 경우, 지난 데이터에서 해당 기상요인이 비슷한 양상을 보이는 데이터 중 가장 품질이 좋은 데이터들(ex. 특/상 이렇게 있다면 특들이 선택되게)만 추출한다.
4. 추출한 데이터를 예측 모델을 돌리고, SHAP 기법을 사용하여 영향을 미친 feature를 도출한다.
5. 해당 분석을 농가에 제공함으로써 사과의 상태와 사과 품질 향상을 위한 개선 방향성을 제공한다.

이해: 1월 습도가 영향을 미치긴 했는데, 정확하게 분류를 하기 위해서 1월 습도가 어느정도 높았을 때 품질에 영향을 주었는지 그 기준치 결정 ( 사과 품질과 해당 칼럼의 데이터를 관찰 후 결정 )

**0726 23시 회의**

< 할 것 >

* ~~참가 신청서 작성 27일까지 -지연혜현~~
* 개요 만들기 28일까지 + 제출 - 수미도현
* +시간남으면 데이터프레임

<참가 신청서> 0,1,4 혜현/ 2,3 지연

0. 프로젝트 간략 소개

1. 문제 선택 이유 + 수정

2. 문제해결 모델 개발 계획

3. 예상하는 결과물 형태와 내용 (구체적으로 작성)

4. 활용 방안

