**중간보고서**

**인공지능을 이용한 농산물 가격 예측 및 시각화.**

보고서 및 논문 윤리 서약

1. 나는 보고서 및 논문의 내용을 조작하지 않겠습니다.

2. 나는 다른 사람의 보고서 및 논문의 내용을 내 것처럼 무단으로 복사하지 않겠습니다.

3. 나는 다른 사람의 보고서 및 논문의 내용을 참고하거나 인용할 시 참고 및 인용 형식을 갖추고 출처를 반드시 밝히겠습니다.

4. 나는 보고서 및 논문을 대신하여 작성하도록 청탁하지도 청탁받지도 않겠습니다.

나는 보고서 및 논문 작성 시 위법 행위를 하지 않고, 명지인으로서 또한 공학인으로서 나의 양심과 명예를 지킬 것을 약속합니다.

과 목 : 캡스톤디자인1

담당교수 : 이강선 교수님

**GitHub URL:** <https://github.com/Kimsejin97/MJ_Capstone>

<1조>

- 김세진 60161594

- 김진혁 60170339

- 신성범 60165060

**목차**

1. 문서개요

1.1 목적

1.2 범위

1.3 문서구조

2. 프로젝트 개요

3. 분석

3.1 농산물 위험도 분석

3.1.1 분석 문항

3.1.2 데이터

3.1.3 분석 과정

3.1.4 결론

3.2 농산물 가격 예측

3.2.1 분석 문항

3.2.2 데이터

3.2.3 분석 과정

3.2.4 결론

4. 어플리케이션 구조

4.1 어플리케이션 UI화면

4.2 구현기능

**1. 문서개요**

**1.1 목적**

본 문서는 ‘인공지능을 이용한 농산물 가격예측 및 시각화’프로젝트를 위한 문서이다. 이는 프로젝트를 시작하기 위해 필요한 분석을 문서화하였다. 후의 프로젝트 구현 진행 시 문서를 참고하는데 목적을 둔다.

**1.2 범위**

본 문서는 ‘인공지능을 이용한 농산물 가격예측 및 시각화’ 프로젝트의 분석에 대한 문서로, 농산물 위험도 분석, 농산물 가격예측 분석으로 구분된다. 또한 분석을 통해 구현하게 될 어플리케이션의 기능과 구현 방법을 포함하고 있다.

**1.3 문서구조**

본 문서는 프로젝트 개요를 시작으로 농산물 위험도 분석과 농산물 가격예측의 분석문항, 데이터, 분석과정, 결과를 기술하며 어플리케이션 구조를 기술한다.

**2. 프로젝트 개요**

1. 농산물 위험도 분석

농업에 있어 기상은 농산물 생육에 특히 많은 영향을 미친다. 인위적인 방법을 동원하여 환경적 요소를 조정할 수 있지만, 기상은 일단위로 변하거나 계절단위 변화에 따라 생육조건이 달라지고 결과적으로 다른 환경적 요소보다 수확량에 절대적인 영향을 미치게 된다. 즉, 기상은 인간의 힘으로는 통제할 수 없는 부분이기 때문에 적재적소에 필요한 정보가 제공되지 않는다면 농산물 생산량에 큰 타격을 줄 수 있다,

이 프로젝트에서는 농산물 생산량과 기상기후의 상관관계가 농산물 생산량에 어느 정도 영향을 주는지를 분석하고, 농산물마다 중요한 기상정보가 무엇인지를 구분한다. 또한 주산지의 평년대비 현재 기상을 분석해 농산물 위험도를 파악한다.

2. 농산물 가격 예측

농산물 가격 변동이 심해 농민들의 피해가 깊어지고 있다. 농산물 가격은 비농산물 가격과 비교해 높은 변동성을 가지고 있다.(Ahn & Kim, 2008) 이는 소비자 입장에서는 구매 부담을, 생산자 입장에서는 수입의 불안정성을 유발한다. 농산물 가격을 미리 예측하고 적정한 시점에 공급한다면 농산물 가격 안정에 도움이 될 것이다.

이 프로젝트에서는 농산물의 공급량, 생산량, 수입량, 수출량의 상관관계를 파악한 뒤 농산물 가격을 예측하는 모델을 개발한다. 가격 예측 모델을 이용해 가격을 예측하고 농산물 가격 추이를 시각화해 표현한다.

**3. 분석**

**3.1 농산물 위험도 분석**

**3.1.1 분석 문항**

1. 농산물(배추,무,고추, 마늘, 양파)별 생산량과 생육 기간 전체의 기상요인(기온, 강수량, 일사량, 습도, 적설량, 순간최대풍속)의 상관관계 조사하기

1.1.1) 8~10월 강수량은 무 생산량과 얼마만큼의 상관관계가 있는가?

1.1.2) 겨울동안의 적설량은 내년 마늘 및 양파 생산량과 상관관계가 있는가?

1.1.3) 5~7월 일조량은 고추 생산량과 상관관계가 있는가?

**3.1.2 데이터**

**<사용데이터>**

1. 생산량데이터

KOSIS(국가통계포털)- 농작물 생산조사

<http://kosis.kr/index/index.do>

-연도별

농산물: 배추, 무, 건고추, 마늘, 양파

구분: 도별

기간: 2000년~ 2019년

2. 기상데이터 (기온, 강수량, 일조시간, 일사량, 습도, 적설량, 순간최대풍속)

KREI (농업관측본부)- 기상자료

<http://oasis.krei.re.kr/index.do>

구분: 시군구별

기간: 2000/01/01 ~ 2020/05/21

**<데이터 전처리>**

1. 생산량데이터

농산물에 따라 배추와 무는 전라북도, 건고추는 경상북도, 마늘은 전라남도와 경상북도, 양파는 전라남도 지역 데이터만 추출한다.

2.기상데이터

주산지 기상데이터만 추출한다. 배추와 무는 전라북도 고창군. 건고추는 경상북도 안동시. 마늘은 전라남도 고흥군, 경상북도 의성군. 양파는 전라남도 무안군. 주산지의 기상데이터를 추출한 뒤 일별로 되어있는 데이터를 년도와 월별로 groupby한 뒤 값을 평균을 낸다.

**3.1.3 분석 과정**

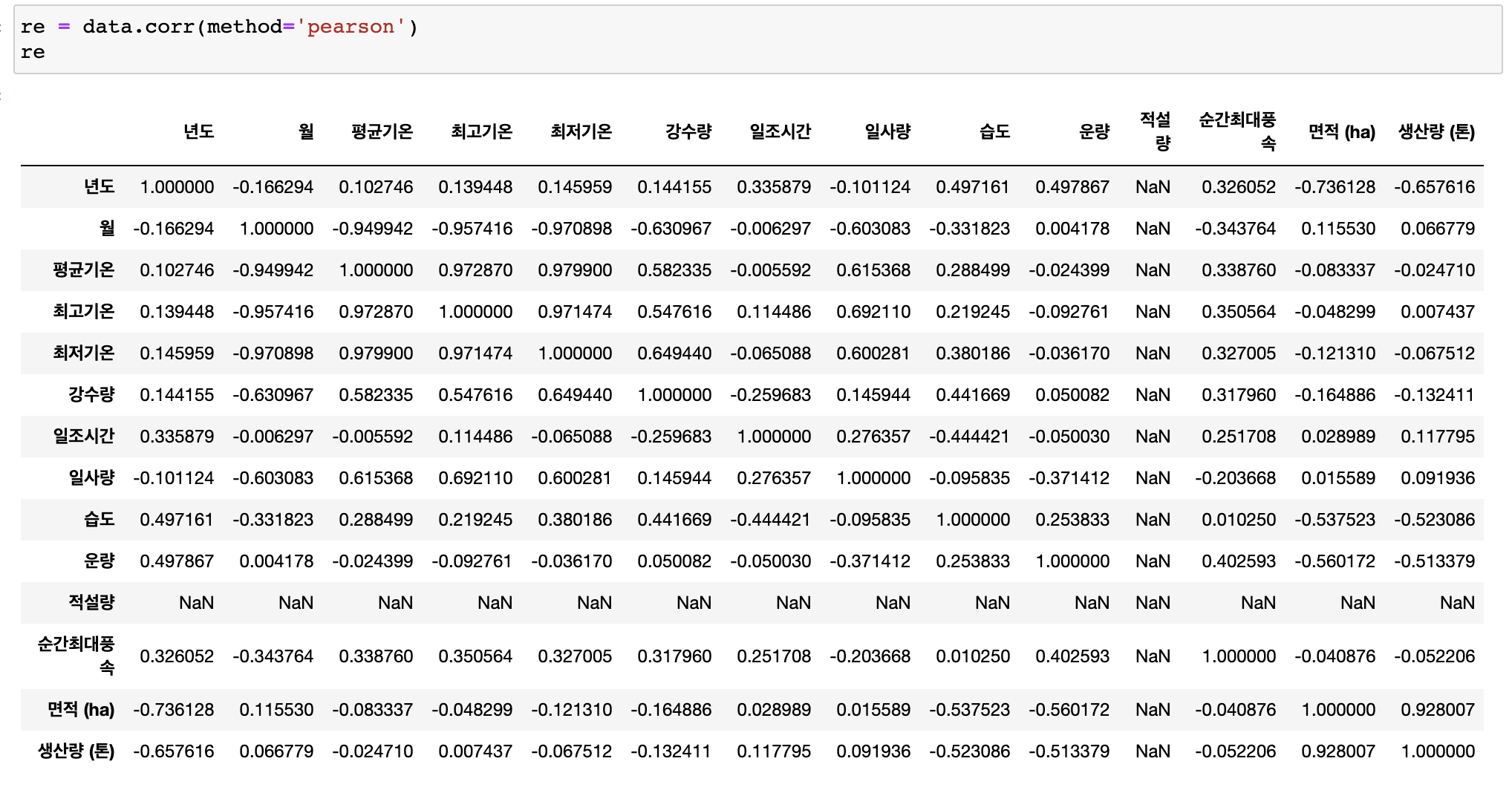
**<1.1번 분석문항>**

1. 데이터 병합



기상데이터와 생산량데이터를 합친다.

2. 데이터 상관관계 분석을 한다.



**3.1.4 결론**

상관분석 결과 강수량은 무 생산량과 상관관계가 없었으며, 습도만이 생산량과 음의 상관관계가 나타난다. 생육기간 동안 습도가 높았다면 무 생산량이 감소한다는 결론을 얻을 수 있다. 또한 고추 생산량과 일조시간이 음의 상관관계를 나타냈다. 고추는 일조시간이 많을수록 생산량이 적어지는 것으로 보인다. 조사해 본 결과 고추는 민감한 농산물이라 여러 변수의 영향을 많이 받아 상관관계가 명확하지 않다고 나온다. 일조시간을 제외한 다른 변수들과는 뚜렷한 상관관계가 없다. 전체적인 결론은 작물별로 구체적인 기상요인은 다르지만 전체적으로 농산물 생산량과 생육 기간의 기상요인은 상관관계가 있다는 결론을 얻을 수 있다.

**3.2 농산물 가격 예측**

**3.2.1 분석 문항**

1. 농산물(고추, 마늘, 양파) 공급량이 가격에 영향을 끼치는가?

2.1.1) 농산물의1년간 공급량은 가격에 영향을 끼치는가?

2.1.2) 농산물의 한달간 공급량은 가격에 영향을 끼치는가?

2. 농산물(고추, 마늘, 양파) 생산량이 가격에 영향을 끼치는가?

2.2.1) 고추의 전년산 생산량은 당해 연도 8월 가격에 영향을 끼치는가?

2.2.2) 마늘의 전년산 생산량은 당해 연도 6월 가격에 영향을 끼치는가?

2.2.3) 양파의 전년산 생산량은 당해 연도 4월 가격에 영향을 끼치는가?

3. 농산물(고추, 마늘, 양파) 수입량, 수출량이 가격에 영향을 끼치는가?

2.3.1) 농산물의 평년 수입량은 가격에 영향을 끼치는가?

2.3.2) 농산물의 평년 수출량은 가격에 영향을 끼치는가?

2.3.3) 농산물의 월별 수입량은 가격에 영향을 끼치는가?

2.3.4) 농산물의 월별 수출량은 가격에 영향을 끼치는가?

**3.2.2 데이터**

**<사용데이터>**

1. 연도별 농산물(고추, 마늘, 양파)의 가격 데이터

KAMIS(농산물유통정보)- 월별, 연도별 가격 데이터

<https://www.kamis.or.kr/customer/price/wholesale/period.do?action=monthly>

-월별

농산물: 건고추(양건, 화건), 마늘(피마늘-난지), 양파

기간: 2000/01월 ~ 2019/12월

-연도별

농산물: 건고추(양건, 화건), 마늘(피마늘-난지), 양파

기간: 2000년 ~2019년

1. 연도별 농산물(고추, 마늘, 양파)의 공급량 데이터

농업 관측 통계 시스템 - 반입 물량 데이터

<http://oasis.krei.re.kr/basicInfo/wholesale/qty.do>

2000년 ~ 2020년, 시군구별

1. 연도별 농산물(고추, 마늘, 양파)의 생산량 데이터

KOSIS(국가통계포털)- 농작물 생산조사

<http://kosis.kr/index/index.do>

농산물: 건고추, 마늘, 양파

기간: 2000년~ 2019년

1. 연도별 농산물(고추, 마늘, 양파)의 수입 및 수출량 데이터

KATI(농식품수출정보)- 월별, 연도별 수입,수출량

<https://www.kati.net/statistics/periodPerformance.do>

-월별

농산물: 고추, 마늘, 양파

기간: 2010/01월 ~ 2019/12월

-연도별

농산물: 고추, 마늘, 양파

기간: 2010년 ~2019년

**<데이터 전처리>**

2. 공급량 데이터

연도별 반입 물량을 하나의 DataFrame으로 통합한다. 거래년월일에서 연도와 월을 분리해서 별도의 변수로 만든다. 거래량에서 콤마(,)를 제거하고, integer형으로 변환한다. 가격 데이터와 공급량 데이터를 merge한다.

**3.2.3 분석 과정**

**<2.1번 분석문항>**

1. 연도별 농산물(고추, 마늘, 양파)의 가격 데이터에 공급량 데이터를 붙이고, 상관분석을 수행함, 이후 그래프와 히스토그램으로 시각화한
2. 연도별 농산물(고추, 마늘, 양파)의 가격 데이터에 생산량 데이터를 붙이고, 상관분석을 수행함
3. 연도별 농산물(고추, 마늘, 양파)의 가격 데이터에 수출입 물량 데이터를 붙이고, 연도별 월별 데이터로 나눠서 상관분석을 함, 분석 결과를 히트맵으로 나타냄

**<2.2번 분석문항>**

예시)

2.2.1) 고추의 전년산 생산량은 당해 연도 8월 가격에 영향을 끼치는가?

1. 데이터 추출

output = pepper.iloc[:19,3] #2019년을 제외하고 연도별 공급량 추출

price\_yang = pepper\_yang\_price[pepper\_yang\_price['월']==8].reset\_index() #8월 추출

price\_yang = price\_yang.iloc[1:,3].reset\_index(drop=True) #2000년을 제외한 8월가격

2. 전년산생산량과 당해연도 8월가격 합치기

data = pd.concat([output, price\_yang], axis=1)

3. 상관관계 분석

corr = data.corr(method='pearson')

corr

**<2.3번 분석문항>**

1. 데이터 병합



월별 수입수출량데이터와 월별 농산물 가격 데이터 병합



연도별 수입수출량데이터와 연도별 농산물 가격데이터 병합

2. 상관관계 분석

onion\_year\_data[['수입(중량)', '평균가격']].corr(method='pearson’)

onion\_year\_data[['수출(중량)', '평균가격']].corr(method='pearson’)

onion\_data[['수입(중량)', '가격']].corr(method='pearson’)

onion\_data[['수출(중량)', '가격']].corr(method='pearson')

**3.2.4 결론**

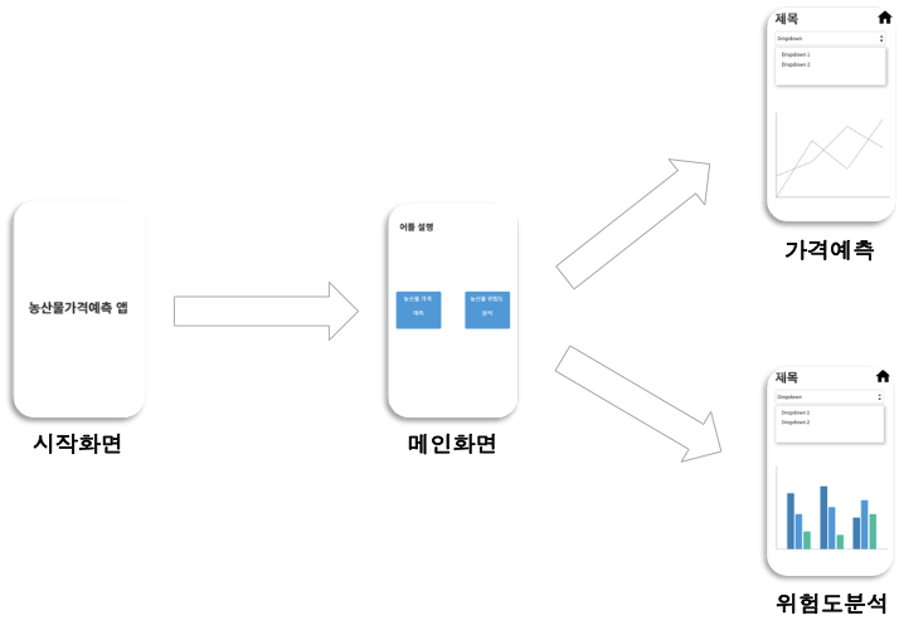
고추의 가격과 공급량은 상관관계가 매해 달라지기 때문에 일관성이 있지 않다. 다만 강한 양의 상관관계 또는 강한 음의 상관관계가 나타나는 해가 있다. 마늘의 가격과 공급량 상관관계 그래프는 대체적으로 음의 상관관계쪽으로 치우쳐져 있다. 그렇기 때문에 마늘의 공급량이 증가하면 가격이 낮아지고, 공급량이 줄어들면 가격이 높아진다. 양파의 가격과 공급량의 상관관계는 마늘보다 더 강한 음의 상관관계를 나타내고 있다. 그렇기 때문에 양파는 마늘보다 공급량에 의한 가격의 변동이 더 심할 것으로 보인다.

모든 농산물은 전년산 생산량과 유의미한 상관관계를 가지고 있었다. 고추와 마늘은 음의 상관관계를 가지고 있었고, 양파는 양의 상관관계를 가지고 있었다. 전년도 생산량이 많은 경우 고추와 마늘의 가격은 내려가고, 양파의 가격은 오히려 오를것으로 보인다.

수출입량과 가격과의 상관관계를 조사한 결과 고추는 전반적으로 수출보다는 수입량과 가격은 양의 상관관계가 있고, 마늘은 수출입량과 가격이 유의미한 상관관계가 있지는 않고 양파는 수출은 음의 상관관계, 수입은 양의 상관관계가 있다.

**4. 어플리케이션 구조**

**4.1 어플리케이션 UI 화면**



**4.2 구현 기능**

1. 시작 화면: 어플리케이션 데이터 로딩 전 나타나는 화면
2. 메인 화면: 가격 예측 / 위험도 분석 기능을 선택할 수 있는 버튼
3. 가격 예측

농산물 가격예측 분석문항을 통해 각 농산물별 농산물 공급량, 생산량, 수입량, 수출량이 농산물 가격에 끼치는 영향을 확인했다. 각 농산물별 가격에 영향을 끼치는 요인을 중심으로 변수를 선정해 예측모델을 개발한다. 가격예측 모델은 LSTM 또는 RNN을 사용하며 가격변수를 기준으로 가격 변동성을 설명할 수 있는 예측 변수를 적용한다.

1. 위험도 분석

주산지 날씨 데이터를 가져와서 위험도를 표시한다. ( 안정-초록 / 주의-주황 / 위험-빨강)

농산물 위험도 분석 결과로 인해 기상기후와 농산물 생산량간의 상관관계가 성립되었다. 기상요인 중 ‘기온, 강수량, 일조시간’을 이용해 농산물 위험도분석 기능을 구현한다. 농산물 위험도 분석기능은 주산지 날씨데이터 주산지 평년(5년간 데이터 평균)을 이용해 비교한다. 주산지의 경우는 배추 – 전라북도 고창, 무 – 전라북도 고창, 건고추 – 경상북도 안동, 마늘 – 전라남도 고흥, 경상북도 의성, 양파 – 전라남도 무안으로 정의한다. 주산지의 날씨 7일간의 동향을 주산지 평년(5년간 데이터 평균)날씨와 비교한다.

ex) 2020년 5/24일 배추의 농산물 위험도를 분석할경우

주산지 날씨(a): 5/17~5/23일 7일간의 기상요인(기온, 강수량, 일조시간) 평균

주산지 평년 날씨(b): 2015~ 2019년도 5/17~5/23일 7일간의 기상요인(기온, 강수량, 일조시간) 평균

a-b가 주산지 날씨와 주산지 평년 날씨의 비교이다.

농산물 위험도 구분은 안정-초록색, 주의-주황색, 위험-빨간색으로한다. 각 위험도의 기온과 일조시간 기준은 |a-b|의 차이가 1.5이하일경우 ‘안정’, 1.5 < |a-b| <= 3.0일경우 ‘주의’, 차이가 3.0초과일 경우 ‘위험’으로 지정한다. 강수량의 경우 기준은 |a-b|의 차이가 3.0이하일경우 ‘안정’, 3.0 < |a-b| <= 6.0일경우 ‘주의’, 차이가 6.0초과일 경우 ‘위험’으로 지정한다.

주산지 날씨(a) - 주산지 평년 날씨(b)

|a-b|

기온 |a-b| <= 1.5 ‘안정’

1.5 < |a-b| <= 3.0. ‘주의’

|a-b| > 3.0. ‘위험’

강수량 |a-b| <= 3.0 ‘안정’

3.0 < |a-b| <= 6.0. ‘주의’

|a-b| > 6.0. ‘위험’

일조시간 |a-b| <= 1.5 ‘안정’

1.5 < |a-b| <= 3.0. ‘주의’

|a-b| > 3.0. ‘위험’