기후 변화에 따른 쌀 생산량 분석 및 예측 보고서

3-A 202244035 이승예



INDEX

01 Lv.1 데이터 수집

05 Lv.5 쌀 생산량 예측

02 Lv.2 쌀 생산량 데이터

06 Lv.6 결론

03 Lv.3 기후 데이터

04 Lv.4 시각화



Lv.1 데이터 수집

쌀 생산량 데이터

전국 시군별 논벼 생산량 (Excel)

재배면적(ha)과 생산량(톤) 포함

Lv.1

Lv.2

Lv.3

01

Lv.4

Lv.5

Lv.6

Lv.7

Lv.8

수집 데이터 종류

02

기후 데이터

연간 기후 데이터 (CSV) 월별 기후 데이터 (CSV) 03

대상 지역

전라남도 충청남도 전라북도 경상북도

01



Lv.2

Lv.3

Lv.4

Lv.5

Lv.6

Lv.7

Lv.8

Lv.2 쌀 생산량 데이터 - 1) 원본 데이터 구조 이해

쌀 생산량 데이터 원본 형태

행정구역별(1)	2008		2009		2010		
867 42(1)	논벼:재배면적 (ha)	논벼:생산량 (톤)	논벼:재배면적 (ha)	논벼:생산량 (톤)	논벼:재배면적 (ha)	논벼:생산량 (톤)	
전국	927,995	4,825,078	917,990	4,898,725	886,516	4,281,729	
서울특별시	502	2,560	286	1,371	267	1,161	
부산광역시	4,112	21,245	4,112	20,084	3,644	16,944	
대구광역시	3,976	20,863	3,477	19,146	3,217	15,612	
인천광역시	13,778	70,351	13,256	62,221	12,588	54,193	
광주광역시	6,442	31,020	6,528	31,758	6,182	28,410	

문제점

- 연도별 데이터가 가로로 나열됨
- 재배면적과 생산량이 반복되는 구조
- 분석하기 어려운 길게 늘어진 형식



Lv.2

Lv.3

Lv.4

Lv.5

Lv.6

Lv.7

Lv.8

Lv.2 쌀 생산량 데이터 - 2) 데이터 재구조화

데이터 재구조화: 각 행정구역별, 연도별로 데이터 재구성

```
restructured_data = []
for idx, row in raw_data.iloc[2:].iterrows():
   region_name = row.iloc[0]
   for i, year in enumerate(unique_years):
       area idx = i * 2 + 1 # 재배면적 컬럼 인덱스
       prod_idx = i * 2 + 2 # 생산량 컬럼 인덱스
       restructured_data.append({
           '행정구역': region name,
           '일시': year,
           '재배면적(ha)': row.iloc[area_idx],
           '생산량(톤)': row.iloc[prod_idx]
       })
```

변환 전:

```
행정구역별(1) 2008 2009
논벼:재배면적 (ha) 논벼:생산량(톤) 논벼:재배면적 (ha) 논벼:생산량(톤)
전국 927,995 4,825,078 917,990 4,898,725
```

변환후:

행정구역	일시	재배면적(ha)	생산량(톤)
전라남도	2008	183630	888531
전라남도	2009	183359	902089
충청남도	2008	162244	895112
충청남도	2009	160952	914388



Lv.2

Lv.3

Lv.4

Lv.5

Lv.6

Lv.7

Lv.8

Lv.2 쌀 생산량 데이터 - 3) 대상 지역 필터링

대상 지역 필터링

1) 필터링을 위한 초기 설정 : 전국 지역 중 쌀 생산량이 많은 전라남도, 충청남도, 전라북도, 경상북도 지역을 필터링

```
target_regions = ['전라남도', '충청남도', '전라북도', '경상북도']
processor = DataProcessor(target_regions)
```

2) 필터링 및 정렬 함수

```
def _filter_and_sort_data(self, df: pd.DataFrame) -> pd.DataFrame:
    """
    데이터프레임을 필터링하고 정렬하는 헬퍼 메서드
    """
    if self.target_regions:
        df = df[df['행정구역'].isin(self.target_regions)]
    return self._sort_dataframe(df, '행정구역')

def _sort_dataframe(self, df: pd.DataFrame, location_col: str) -> pd.DataFrame:
    """데이터프레임 정렬을 위한 통일된 메서드"""
    return df.sort_values([location_col, '일시']).reset_index(drop=True)
```



Lv.2

Lv.3

Lv.4

Lv.5

Lv.6

Lv.7

Lv.8

Lv.2 쌀 생산량 데이터 - 4)필터링 결과

필터링 전 (일부 데이터):

행정구역	일시	재배면적(ha)	생산량(톤)
전국	2008	927995	4825078
서울	2008	502	2560
부산	2008	4112	21245
전라남도	2008	183630	888531
충청남도	2008	162244	895112
제주도	2008	32	149

필터링 후 (일부 데이터):

행정구역	일시	재배면적(ha)	생산량(톤)
경상북도	2008	123678	658429
경상북도	2009	122441	679872
전라남도	2008	183630	888531
전라남도	2009	183359	902089
전라북도	2008	141161	761262
전라북도	2009	138445	797338
충청남도	2008	162244	895112
충청남도	2009	160952	914388



Lv.2

Lv.3

Lv.4

Lv.5

Lv.6

Lv.7

Lv.8

Lv.3 기후 데이터 - 1) 원본 데이터 구조 이해

연간 기후 데이터(연간기후전체지역.csv)

	지점	성 지점	명	일시	평균기	온(°C) 합계	강수량(mm)	평균 상대습도(%)	합계 일조시간(hr)	
0	90	속초	2008		13.0	1415.0	63.0	2224.6		
1	90	속초	2009		12.8	1420.1	61.1	2185.2		
2	90	속초	2010		12.3	1283.6	62.3	2065.7		
3	90	속초	2011		12.1	1656.1	62.0	2043.2		
4	90	속초	2012		12.0	1217.7	67.4	2101.0		

월별 기후 데이터 (월별기후전체지역.csv)

	지점	성 지점	명	일시	평균.	기온(°C)	평균상대습도(%)	월힏	;강수량(00~24h만)(mm)	합계 일조시간(hr)
0	90	속초	2008-01		0.3	50.0	7	1.1	161.5	
1	90	속초	2008-02		1.3	36.0		0.6	216.0	
2	90	속초	2008-03		7.3	63.0	10	3.9	156.0	
3	90	속초	2008-04		13.1	62.0	1	3.5	210.8	
4	90	속초	2008-05		16.8	67.0	7	0.0	205.1	

문제점

- 1. 컬럼명이 통일되지 않음
- 2. 결측치 존재



Lv.2

Lv.3

Lv.4

Lv.5

Lv.6

Lv.7

Lv.8

Lv.3 기후 데이터 - 2) 기후 지표 통일화 및 지역 매핑

연간/월간 데이터의 컬럼명 통일화

```
CLIMATE_COLUMNS = {
    'annual': {
        '평균기온(°c)합계': '평균기온(°c)',
        '강수량(mm)': '합계 강수량(mm)',
        '평균 상대습도(%)합계': '평균 상대습도(%)',
        '일조시간(hr)': '합계 일조시간(hr)'
    },
    'monthly': {
        '평균기온(°c)': '평균기온(°c)',
        '평균상대습도(%)': '평균 상대습도(%)',
        '월합강수량(00~24h만)(mm)': '합계 강수량(mm)',
        '합계 일조시간(hr)': '합계 일조시간(hr)'
    }
}
```

컬럼명 변경 적용

```
def _prepare_climate_data(self, df: pd.DataFrame, data_type: str) -> pd.DataFrame:
# 컬럼명 통일화

df = df.rename(columns=self.CLIMATE_COLUMNS[data_type])

# 지역 정보 메핑

df['행정구역'] = df['지점명'].map(self.REGION_MAPPING)

# 대상 지역 필터링

return df[df['행정구역'].sin(self.target_regions)]
```

RECION_MAPPING:

42개 관측지점을 4개 광역시도로 매핑 예: '목포', '여수', '순천' → '전라남도'



Lv.2

Lv.3

Lv.4

Lv.5

Lv.6

Lv.7

Lv.8

Lv.3 기후 데이터 - 2) 기후 지표 통일화 및 지역 매핑 결과

연간 데이터 변환 결과:

변환 전:

지점 지점명 평균기온(°C)합계 강수량(mm) 평균 상대습도(%)합계 일조시간(hr)

변환 후:

지점 지점명 평균기온(°C) 합계 강수량(mm) 평균 상대습도(%) 합계 일조시간(hr)

월별 데이터 변환 결과:

변환 전:

지점 지점명 평균기온(°C) 평균상대습도(%) 월합강수량(00~24h만)(mm) 합계 일조시간(hr)

변환 후:

지점 지점명 평균기온(°C) 평균 상대습도(%) 합계 강수량(mm) 합계 일조시간(hr)



Lv.2

Lv.3

Lv.4

Lv.5

Lv.6

Lv.7

Lv.8

Lv.3 기후 데이터 - 3) 결측치 처리

처리된 결측치 유형

- 1. 연속적 결측: 특정 연도 전체 데이터 누락
- 2. 산발적 결측: 일부 월별 데이터 누락
- 3. 부분 결측: 특정 측정값만 누락

2	59 강진군	2017	816.2		2367.3
		\vee			
259 강진군	2017	14	816.2	72.5	2367.3 전라남도



결측치 처리 방법 => fill_missing_years()

- 1. 각 지역별로 누락된 연도/월 확인
- 2. 해당 지역의 평균값으로 결측치 대체
- 3. 월별 데이터의 경우 동일 월의 평균값으로 대체 (계절성 고려)
- 4. 데이터의 일관성을 위해 소수점 첫째자리까지 표준화

Lv.4 시각화(1) - 생산량 변화율 분석



Lv.1

Lv.2

Lv.3

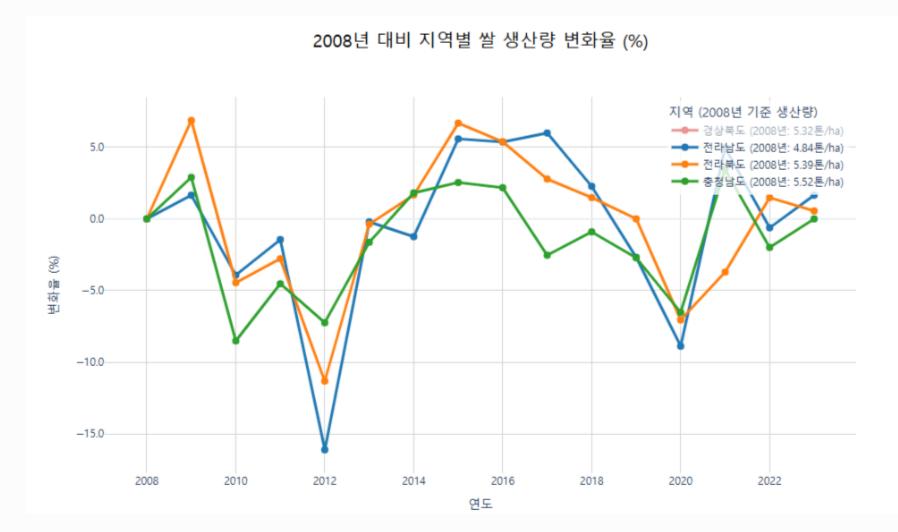
Lv.4

Lv.5

Lv.6

Lv.7

Lv.8



분석 목적

- 2008년 대비 지역별 쌀 생산량(단위 면적당)의 변화율 추적
- 지역간 생산량 변동 패턴 비교
- 장기적 생산량 트렌드 파악

분석 결과

- 2012년에 전라남도의 급격한 감소(-15%)와 2016년 전국적인 생산량 증가
- 2020년 모든 지역의 동반 하락이 주요 변동 시점이었으며, 2022년에는 대체로 회복세를 보임.

Lv.4 시각화(2) - 지역별 생산 특성 비교



Lv.1

Lv.2

Lv.3

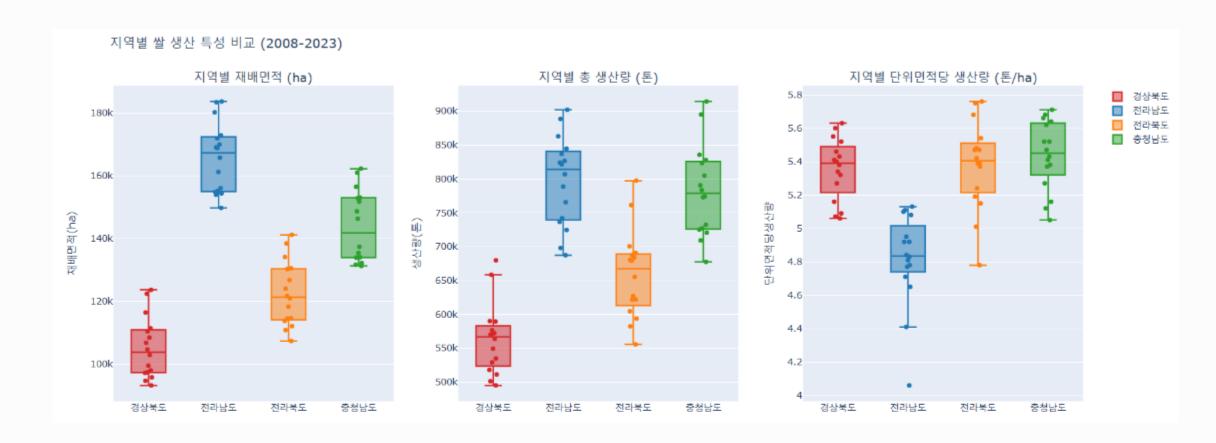
Lv.4

Lv.5

Lv.6

Lv.7

Lv.8



분석 목적

- 재배면적, 생산량, 단위면적당 생산량의 지역별 분포 파악
- 생산 효율성 비교
- 지역별 생산 규모 차이 분석

분석 결과

- 재배면적이 큰 지역이 반드시 단위면적당 생산성이 높은 것은 아님
- 충청남도는 재배면적은 중간 수준이나 단위생산성이 가장 높아 효율적인 생산 체계를 가진 것으로 보임
- 전라남도는 가장 넓은 재배면적을 보유했으나 단위생산성이 낮음

Lv.4 시각화(3) - 기후 취약성 분석



Lv.1

Lv.2

Lv.3

Lv.4

Lv.5

Lv.6

Lv.7

Lv.8



분석 목적

- 기후 변수별 지역 취약성 평가
- 생산량 변동성과 기후 요인의 관계 분석
- 지역별 기후 리스크 파악

분석 결과

모든 지역에서 강수량의 변동성이 가장 높아 쌀 생산량의 주요 리스크 요 인임을 알 수 있음

Lv.1

Lv.2

Lv.3

Lv.4

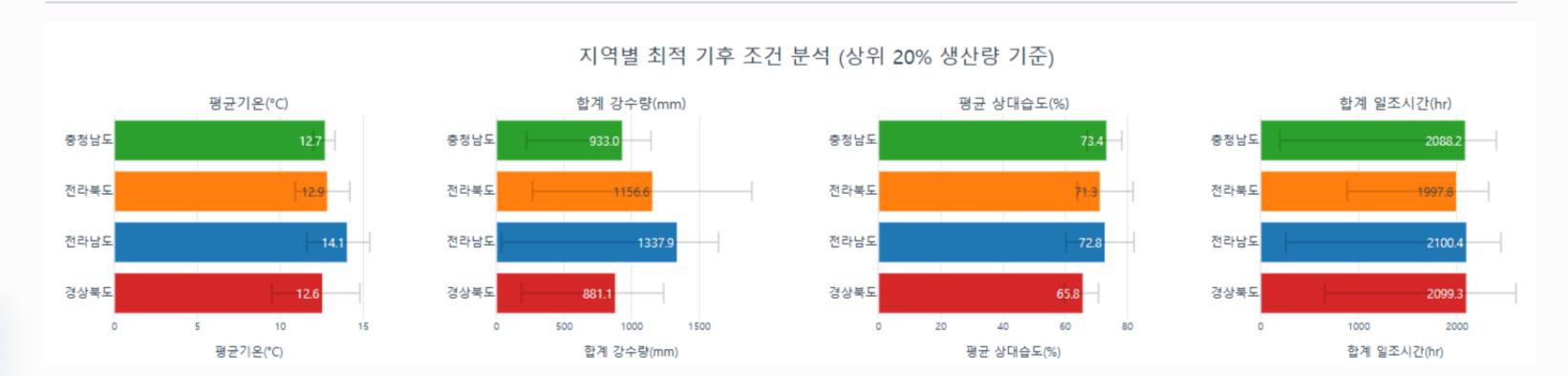
Lv.5

Lv.6

Lv.7

Lv.8

Lv.4 시각화(4) - 최적 재배 조건 분석



분석 목적

- 상위 20% 생산량 기준 최적 기후 조건 도출
- 지역별 최적 생산 조건 비교
- 기후 요인별 허용 범위 파악
- 향후 기후변화에 따른 쌀 생산량 변화를 예측 하고 대비하기 위함

분석 결과

전라남도가 더 높은 온도와 강수량에서 최적 생산성을 보이는 반면, 충청남도와 경상북도는 상대적으로 더 낮은 기후 조건에서도 높은 생산성을 달성할 수 있음



Lv.2

Lv.3

Lv.4

Lv.5

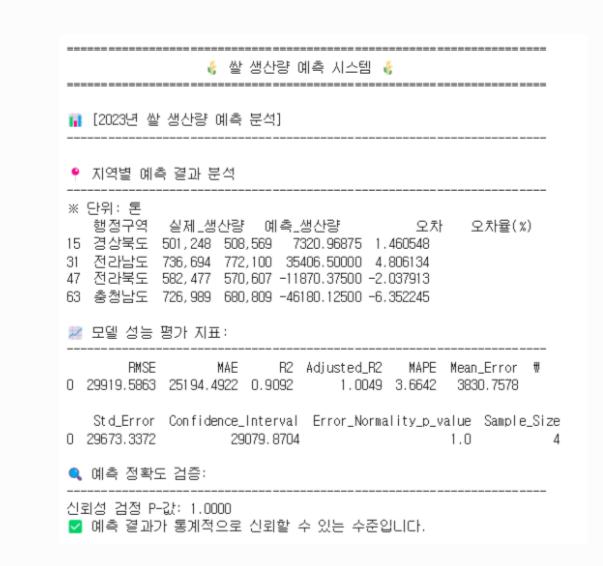
Lv.6

Lv.7

Lv.8

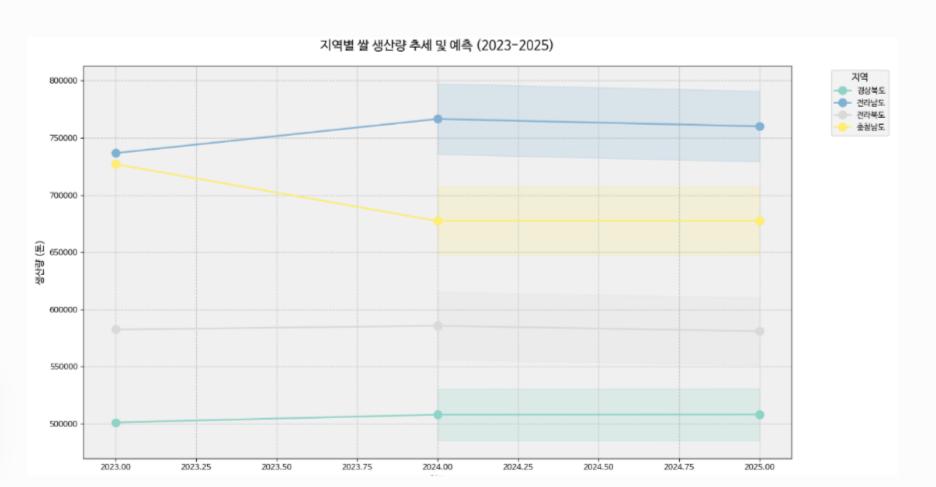
Lv.5 예측시스템 - 2023년 예측 결과 및 모델 정확도 분석

```
def train_model(self, target_year=2023):
   """모델 학습"""
       data = self.prepare_data()
       # 학습 데이터와 테스트 데이터 분리
       train_data = data[data['일시'] < target_year]
       test_data = data[data['일시'] == target_year]
       if len(test_data) == 0:
          raise ValueError(f"{target_year}년 데이터가 없습니다.")
       # 특성과 타겟 분리
       self.feature_columns = [col for col in data.columns if col not in ['생산량(톤)', '일시']]
       X_train = train_data[self.feature_columns]
       y_train = train_data['생산량(톤)']
       X_test = test_data[self.feature_columns]
       y_test = test_data['생산량(톤)']
       # 특성 스케일링
       X_train_scaled = self.scaler.fit_transform(X_train)
       X_test_scaled = self.scaler.transform(X_test)
       #모델 학습
       self.model = XGBRegressor(
          n_estimators=100.
          learning_rate=0.1,
          max_depth=5,
          random_state=42
       self.model.fit(X_train_scaled, y_train)
       # 예측 및 평가
       y_pred = self.model.predict(X_test_scaled)
```



연간 생산량을 예측하기 전 2023년 데이터를 이용하여 오차율을 줄이고 정확도를 높였다.

Lv.5 예측시스템 - 2024,2025년 예측 분석



🔃 2024년, 2025년 예측 결과

▶ 경상북도

2023년 실제 생산량: 501,248 톤 2024년 예측 생산량: 508,046 톤 2025년 예측 생산량: 508,200 톤 2024-2025 변화율: 0.03%

▶ 전라남도

2023년 실제 생산량: 736,694 톤 2024년 예측 생산량: 766,530 톤 2025년 예측 생산량: 760,122 톤 2024-2025 변화율: -0.84%

▶ 전라북도

2023년 실제 생산량: 582,477 톤 2024년 예측 생산량: 585,759 톤 2025년 예측 생산량: 581,020 톤 2024-2025 변화율: -0.81%

▶ 추청난도

2023년 실제 생산량: 726,989 톤 2024년 예측 생산량: 677,366 톤 2025년 예측 생산량: 677,284 톤 2024-2025 변화율: -0.01%

Lv.6

Lv.7

Lv.1

Lv.2

Lv.3

Lv.4

Lv.5

Lv.8

목적: 지역별 생산량의 변화 추세

결과: 4개 지역 모두 변화율 ±1% 이내로 매우 안정적인 생산량 유지 예상.

전반적으로 네 지역 모두 2025년에는 생산량이 소폭 감소하거나 변동이 크지 않을 것으로 예측



Lv.2

Lv.3

Lv.4

Lv.5

Lv.6

Lv.7

Lv.8

Lv.6 결론

결론적으로, 우리나라의 쌀 생산 시스템은 기후변화에 대해 상당한 회복력과 일부 취약성을 동시에 보여주고 있다 강점:

- 최악의 기후 시나리오에서도 생산량 변동폭이 ±2~3%로 안정적
- 2012년 -15% 급감 후 2014년 빠른 회복세를 보이는 등 극단적 상황에서도 높은 회복력 입증
- 2025년 예측에서도 모든 지역이 ±1% 이내의 안정적 생산량 유지

취약점:

- 지역별 기후 취약성 분석 결과, 강수량의 변동계수가 0.23~0.28로 가장 높아 주요 리스크 요인으로 확인
- 전라남도는 최적 생산을 위해 상대적으로 높은 평균기온(14.1°C)과 강수량(1337.9mm) 필요

이러한 분석을 종합할 때, 우리나라 쌀 생산은 강수량 변동성에 대한 대비가 필요하지만, 전반적으로 기후변화에 대한 적응력이 뛰어나며 향후 에도 안정적인 생산이 가능할 것으로 전망된다.

감사합니다