

R을 활용한 토마토 농가 생산량 예측 알고리즘 연구

양광호, 여현*

*국립순천대학교

yg8686@naver.com, *yhyun@snu.ac.kr

A Study on the Tomato Farm Production Prediction Algorithm Using R

Yang Kwang Ho, Yoe Hyun*

*Suncheon National Univ.

요약

온실 환경 내에서 토마토 생산을 효과적으로 예측하는 것은 수확량 및 자원 관리를 최적화하는 데 중요하다. 본 연구에서는 온도, 습도, 토마토 생산량 간의 관계를 분석하기 위해 통계 프로그래밍 환경 R을 활용한 예측 알고리즘을 연구하는 것을 목표로 한다. 이러한 환경 요인에 대한 데이터는 농촌진흥청 공공데이터를 활용하였으며 통계 분석을 거쳐 토마토 수확량과의 상관 관계를 확인했다. 알고리즘은 유의미한 결과를 나타냈다. 이 연구는 농업에 통계 모델링 기술을 통합하여 예측 정확성과 운영 효율성을 향상시킬 수 있는 가능성을 강조한다. 연구 결과는 온실 환경 조건을 정밀하게 모니터링하고 제어하면 토마토 수확량 예측을 크게 향상시켜 농업에서 더 많은 정보를 바탕으로 의사 결정을 내릴 수 있음을 시사한다.

I. 서론

스마트 농업 분야에서는 온실과 같은 통제된 환경 내 조건을 최적화하는 것이 수확량과 자원 효율성을 극대화하는 중요하다. 토마토는 전 세계적으로 가장 중요한 원예 작물 중 하나로서 경제적 가치와 영양적 중요성으로 인해 중요한 관심 분야이다. 그러나 최적의 생산을 달성하는 것은 종종 온도와 습도 등 다양한 환경 요인의 영향을 받는 등 여러 가지 어려움이 따른다[1, 2, 3].

최근 데이터 분석 및 환경 모니터링 기술의 발전으로 상세한 실시간 데이터를 통해 수확량 예측을 향상할 수 있게 되었다. 본 연구에서는 온실 환경 내에서 토마토 생산을 예측하기 위해 온도 및 습도 데이터를 사용하는 예측 알고리즘을 연구함으로써 이러한 기술적 이점을 활용하려고 한다 [4]. 이 연구는 통계 도구를 농업과 통합함으로써 이러한 환경 변수와 토마토 수확량 간의 상관 관계를 검증할 뿐만 아니라 미래 생산 추세를 예측하기 위한 신뢰할 수 있는 모델을 연구하는 것을 목표로 한다.

이 논문은 토마토 중에서도 완숙 토마토를 중심으로 복잡한 데이터 세트를 처리, 분석 및 모델링하는 데 능숙한 통계 컴퓨팅 및 그래프를 위한 다목적 도구인 R을 사용하여 연구하였다. 본론에서는 분석 결과를 제시하고 온실 관리 및 토마토 재배에 미치는 영향을 논의하며 연구 중인 예측 모델의 향후 연구 및 기대효과에 대한 사항으로 결론을 내릴 것이다.

II. 토마토 생산량 예측 알고리즘

본 논문에서는 토마토 농가의 생산량을 예측하기 위해 R언어를 사용하였으며 농촌진흥청의 공공데이터 포털에 있는 농가의 환경데이터와 판매 데이터를 활용하였다. 여러 지역의 데이터 중 강원도 평창의 특정 농가와 특정 작기를 필터링하여 분석을 진행하였다. 출하 일자를 보면 출하는 매일 진행되나 하루에도 여러 번 반복되어 명확한 결과를 위해 재차 필터링이 필요했다. 같은 날의 출하량은 모두 합쳐주었으며 그 외에 필요 없는 데이터는 제외시켰다(그림 1).

```
A data.frame: 139 × 2
  Group.1      x
  <chr>    <dbl>
1 2021-06-22    265
2 2021-06-23   100
3 2021-06-24   135
4 2021-06-25   375
5 2021-06-26  3100
6 2021-06-27   525
```

그림 1. 출하 일자 별 출하량 필터링

Fig 1. Filtering shipments by ship date

필터링된 데이터를 토대로 X축은 출하 일자, Y축은 총 출하량으로 그림 2와 같이 그래프를 작성하였다.

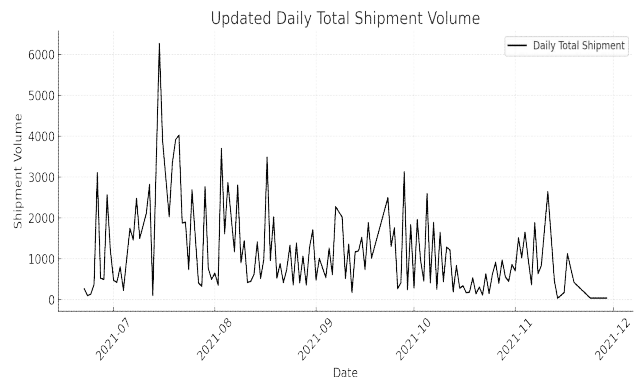


그림 2. 출하 일자 별 총 출하량

Fig 2. Total shipments by ship date

그래프에서 출하량은 균일하지 않고 편차가 심한 결과를 보였다. 온실 환경 요인과 비교를 위하여 다시한번 총 출하량을 주 단위로 합산하여 정리했으며, 온실 내부 온도와 습도는 주 단위로 평균값을 구하였으며 그림 3의 위에서부터 총출하량, 온도, 습도의 그래프이다.

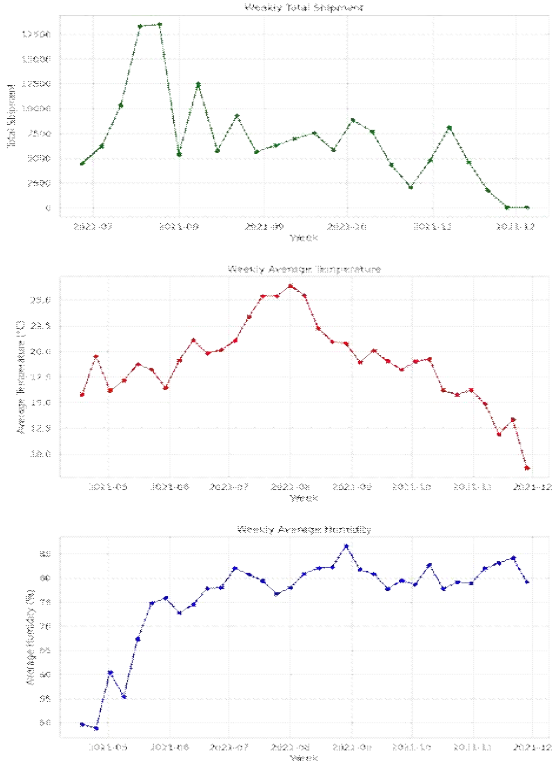


그림 3. 총출하량, 온도, 습도 그래프
Fig 3. Graph of total shipment, temperature and humidity

이 세 데이터를 가지고 상관관계 분석을 진행하였으며 온도와 출하량과의 상관계수가 약 0.717로, 온도가 높아질수록 출하량이 증가하는 양의 상관관계를 보였다(그림 4).

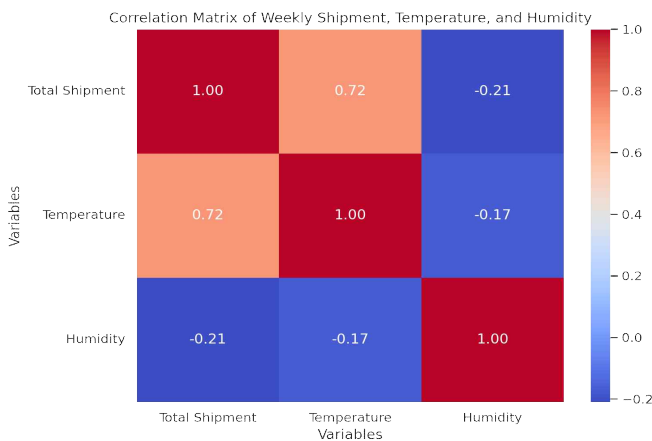


그림 4. 총출하량과 온도, 습도와의 상관관계 매트릭스
Fig 4. Correlation matrix of shipment, teemperature, and humidity

그림 5는 산점도 행렬 결과이며 출하량과 온도는 우상향하는 경향을 보이며 양의 상관관계와 일치하는 것으로 나타났다.

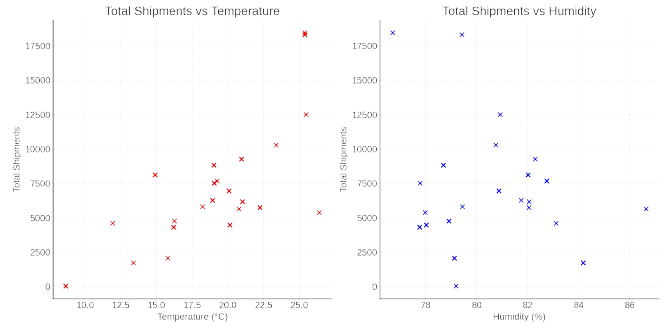


그림 5. 총출하량과 온도(좌), 습도(우)와의 산점도 분석
Fig 5. Scatter plot analysis of total shipment volume, temperature (left), and humidity (right)

III. 결론

본 논문에서는 온실 환경 조건, 특히 온도, 습도와 토마토 생산량 사이의 복잡한 관계를 연구했다. R 프로그래밍 언어를 사용한 종합적인 데이터 분석을 통해 이러한 환경 요인과 완속 토마토 수확량 사이에 유의미한 상관관계가 있음을 확인했다.

연구 결과에 따르면 온도는 온실 환경에서 토마토 생산에 영향을 미치는 중요한 요소로 보여졌다. 이러한 매개변수를 기반으로 한 예측 모델은 유의미한 결과를 예측할 수 있었으며, 이를 통해 농업 운영의 계획 및 의사결정을 위한 귀중한 도구를 제공할 수 있을 것이다.

또한, 이 연구는 다른 작물 및 조건에 적용할 수 있는 방법론적 프레임워크를 제공함으로써 정밀 농업 분야에 기여 할 수 있다. 향후 연구에서는 작물 수확량에 큰 영향을 미칠 수 있는 CO₂, 일사량, 토양 수분등과 같은 추가 변수를 통합하여 보다 정확한 예측 결과를 얻을 수 있을 것으로 예상된다. 또한, 토마토 뿐만 아니라 온실의 여러 작물에 적용 가능한 신뢰도 높은 알고리즘을 개발할 계획이다.

ACKNOWLEDGMENT

“본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT 연구센터사업의 연구결과로 수행되었음” (RS-2024-00259703)

참 고 문 헌

- [1] 정경희, 진효정, 안재욱, 윤혜숙, 오상석, 임채신, 엄영철, 김희대, 홍광표, 박성민. “수량예측을 위한 ‘Cupra’, ‘Fiesta’ 파프리카의 생육특성 및 수확량 패턴 분석”, 시설원예·식물공장, 27(4), 349-355, 2018.
- [2] 이태석, 김진구, 박석호, 이재한, 한길수, 문종필, “외부 광환경 및 생육도일온도가 딸기 생산량에 미치는 영향”, 생물환경조절학회지, 31(4), 432-437, 2022.
- [3] 이태석, 김진구, 한길수, “관부 난방 시스템 적용으로 인한 고설 딸기의 재배 환경 변화와 그에 따른 출퇴, 개화 및 수확량 비교 분석”, 생물환경조절학회지, 32(4), 449-455, 2023.
- [4] 강소라, 조원환, 나명환, 나인섭, “시계열 자료 분석법을 이용한 재배 환경 관리에 따른 참외의 수확량 예측 모델 개발”, 한국정보과학회 학술발표논문집, 2021.
- [5] 최호길, 안희화, 정이나, 이병관, “IoT 및 딥 러닝 기반 스마트 팜 환경 최적화 및 수확량 예측 플랫폼”, 한국정보전자통신기술학회 논문지, 12(6), 672-680, 2019.