

선형회귀모델을 사용한 쌀 가격 예측 및 쌀 가격에 영향을 미치는 날씨의 시기 탐색

최다정⁰, 서진경^{*}, 고광호^{**}, 백주련(교신저자)^{*}

⁰평택대학교 데이터정보학과,

^{*}평택대학교 데이터정보학과,

^{**}평택대학교 스마트자동차학과

e-mail: (dajung2020⁰, kwangho^{**}, jrpaik^{*})@ptu.ac.kr, sjk2700@naver.com^{*}

Prediction of Rice Prices and Search for a Period of Weather Affecting the Prices Based on a Linear Regression Model

Da-jeong Choi⁰, Jin-kyeong Seo^{*}, Kwang-Ho Ko^{**}, Juryon Paik(Corresponding Author)^{*}

⁰Dept. of Digital Information & Statistics, Pyeongtaek University,

^{*}Dept. of Digital Information & Statistics, Pyeongtaek University,

^{**}Dept. of Smart Mobility, Pyeongtaek University

● 요약 ●

농산물의 산지 가격이나 도매가격이 등락하면, 즉시 또는 일정한 시차 이후에 소비자가격도 등락한다. 본 논문에서는 선형회귀모델을 통해 쌀 가격을 예측하고 쌀 가격에 영향을 미치는 날씨의 시기를 찾아보고자 한다. 이에 따라 KAMIS, 기상자료개방포털, KOSIS에서 수집한 날씨, 생산량, 그리고 소비자물가 등락률 데이터를 이용하여 쌀 가격 예측을 수행하고, 날씨 데이터와 쌀 가격 데이터의 날씨 간격을 두어 날씨가 쌀 가격에 영향을 미치는 시기를 알아보았다. 모델 평가 결과, 2개월 간격을 두고 예측한 RMSE가 164.135로 가장 큰 영향을 미쳤다. 본 연구를 기반으로 향후 다른 농산물의 가격 예측도 가능할 것이며 농산물에 영향을 미치는 변수의 시기도 예측할 수 있을 것으로 기대한다.

키워드: 쌀 가격(Rice price), 날씨 조건(Weather condition), 선형회귀(Linear regression)

I. Introduction

재배면적이 감소하고 기상 여건이 악화되면서 2020년산 쌀의 생산량은 2019년 대비 6.4% 감소하였으며 이는 쌀 가격의 상승으로 이어졌다. 이처럼 농산물 대다수는 날씨의 영향을 많이 받으며 농산물의 산지 가격이나 도매가격이 등락하면, 즉시 또는 일정한 시차 이후에 소비자가격도 등락한다[1].

본 논문은 선형회귀모델을 통해 쌀 가격을 예측하고 쌀 가격 형성에 영향을 미치는 날씨의 시기를 알아보려고 한다.

으며, 기상자료개방포털[3]에서 2012년부터 2021년까지의 지역별 일일 종관기상관측(ASOS) 데이터 (평균 기온, 평균 풍속, 평균 지면 온도, 일 강수량, 합계 일조 시간, 합계 일사)와 농업기상관측(AAOS) 데이터 (평균 습도, 일 토양수분)를 날씨 변수로 수집하였다.

KOSIS 국가통계포털[4]에서 전년도 생산량 데이터를 위해 2011년부터 2020년까지의 지역별 연간 쌀 생산량과 2012년부터 2021년까지의 월별 소비자물가 등락률 데이터 (총지수, 생활물가지수, 신선식품지수, 농산물 및 석유류 에너지 제외 지수, 식료품 및 에너지 제외 지수)를 수집하였다.

날씨 변수의 결측값을 각 데이터의 평균으로 변환하였으며, 일 강수량 데이터는 결측값을 0으로 변환하였다. 전년도 생산량 데이터는 연도 별 생산량의 합을 사용하였다. 또한 날씨 변수 및 전년도 생산량 데이터를 정규화하였다. 그리고 모든 데이터를 동일한 날짜를 기준으로 병합하였으며, 결측치는 제거하였다. 모델을 학습하고 이를 평가하기

II. The Proposed Scheme

1. 데이터 수집 및 처리

본 논문에서 쌀 가격을 예측하기 위해 KAMIS 농산물 유통정보[2]에서 2011년부터 2021년까지의 일일 평균 쌀 가격 데이터를 수집하였

위해서 훈련용 데이터(80%)와 테스트용 데이터(20%)로 분리하였다.

2. 모델 수립, 예측 및 평가

쌀 가격을 예측하기 위해서 선형회귀모델을 사용하였다. 학습용 데이터를 통해 모델을 학습하고 테스트용 데이터로 예측 수행 후 테스트용 데이터의 실제 가격과 비교하였다. 평가는 RMSE (평균 제곱근 오차)가 작을수록 더 나은 예측 결과를 도출한다고 보았다. 모든 데이터를 동일한 날짜로 병합하였을 때의 RMSE는 166.076이다. 하지만 본 논문에서 수집한 쌀 가격 데이터는 선물 가격이 아닌 현물 가격이므로 같은 날의 쌀 가격에 같은 날 날씨가 영향을 미칠 가능성이 작다. 그러므로 같은 날짜의 날씨 데이터와 쌀 가격 데이터가 아니라, 개월 수를 늘려가며 날씨 데이터와 쌀 가격 데이터에 날짜 간격을 두어 날씨가 쌀 가격에 영향을 미치는 시기를 찾아보고자 한다.

3. 날씨와 쌀 가격 사이의 날짜 간격 조절을 통한 예측

쌀 가격 데이터를 n개월 간격으로 미뤄 예측한 결과는 Table 1과 같다. 0개월부터 2개월까지 낮아지고, 3개월부터 다시 높아짐을 확인할 수 있다. 따라서 날씨와 쌀 가격의 날짜 간격이 2개월일 때 날씨가 쌀 가격에 가장 영향을 많이 미친다고 할 수 있다.

Table 1. 날씨 간격 조절 후 RMSE

날씨와 쌀 가격의 날짜 간격	RMSE
0개월	166.076
1개월	165.435
2개월	164.135
3개월	172.586
4개월	178.842
5개월	185.630
6개월	192.543
7개월	195.750
8개월	198.282
9개월	201.674
10개월	203.520

III. Conclusions

본 논문에서는 한국의 2012년부터 2021년까지 10년을 기준으로 선형회귀모델을 사용해 날씨, 생산량 등을 통한 쌀 가격을 예측하였다. 이를 근거로 향후 다른 농산물의 가격 예측도 가능할 것이며 농산물에 영향을 미치는 특정 변수의 시기 예측도 가능하리라 생각한다. 본 연구에서는 농산물 중 쌀을 이용하여 가격을 예측하였지만 차후 연구에서는 다른 농산물과 농산물에 영향을 미치는 데이터를 더 추가하고 선형회귀 이외의 모델을 사용해 비교하며 모델의 예측력을 더 높일 수 있는 방향으로 나아가고자 한다.

ACKNOWLEDGEMENT

이 논문은 2021년도 정부 (과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 이공분야기초연구사업임 (NRF-2021R1F1A1064073).

REFERENCES

- [1] Kook Seung-yong, Seo Hong-seok, Seo Dong-ju, Kwon Sang-wook, Kim Kyung-jin. (2021). Recent Changes in Agricultural Prices and Implications. Korea Rural Economic Institute agricultural focus, (), 1-25.
- [2] KAMIS, <https://www.kamis.or.kr/>
- [3] Open MET Data Portal, <https://data.kma.go.kr/>
- [4] KOSIS, <https://kosis.kr/>