

인공지능 알고리즘 기반 한우 경락 가격 예측 모델 분석

정가영¹, 안보경¹, 오태훈², 김민지³, 조석헌*

국립안동대학교¹, 계명대학교², 국립금오공과대학교³, *University of California, San Diego
jky303101@gmail.com, anbokyeong2@gmail.com, taehun2511@gmail.com,
minzzi8040@gmail.com, *justinshcho@gmail.com

Analysis for Korean Beef Auction Price Prediction Model based on Artificial Intelligence Algorithms

Gayeong Jung¹, Bokyeong An¹, Taehun Oh², Minji Kim³, Seokheon Cho*
Andong National University¹, Keimyung University², Kumoh National Institute of Technology³,
*University of California, San Diego (UCSD)

요약

Meat consumption levels increase every year in South Korea. In addition, Hanwoo, which is beef from cows raised natively in Korea, becomes one of the most popular sources of meat to Korean. Thus, it is an important task for the Korean government to maintain a stable supply of Hanwoo at a consistent price. In this study, a Hanwoo auction price prediction model was developed based on Artificial Intelligence algorithms using Hanwoo auction information. We considered four Artificial Intelligence algorithms, which are Random Forest Regression, Multiple Linear Regression, Gradient Boost, and eXtreme Gradient Boost, to predict the auction price of Korean beef. The results of the prediction model analysis indicated that the prediction model based on the eXtreme Gradient Boost algorithm exhibits the best performance.

I. 서론

대한민국의 생활 수준이 향상됨에 따라 대한민국에서의 육류 소비량이 쌀 소비량을 넘어설 만큼 육류의 소비량이 나날이 증가하고 있다. 20대에서 60대까지 3000명을 기준으로 한 한우 고기 소비 동향 모니터링에 따르면 한국인의 68.6%가 육류 중 한우를 가장 선호할 만큼 한우고기는 맛, 안전성 그리고 고영양의 이유로 소비 증가세를 보이고 있다 [1]. 그러나 최근 한우의 수요와 공급의 불안정으로 인해 안정적인 가격으로 한우 고기 조달에 어려움이 있다. 한우의 가격이 여러 요인에 영향을 받아 가격을 예측하기 까다롭다는 점도 불안정을 야기한 이유 중 하나이다. 따라서 본 연구에서는 한우의 안정적인 공급을 위해 한우의 출하 정보를 이용하여 한우 경락 가격을 예측하는 모델을 제안하고자 한다.

HuRak Park *et al.*은 대한민국 경남지방 한우의 경락 가격에 영향을 미치는 요인들을 찾고 이것들이 경락 가격에 영향을 미치는 비중을 대하여 연구하였다 [2]. 연구 결과, 근내지방이 95.41%로 경락가격에 가장 큰 비중을 나타냈으며 도체중 또한 경락 가격을 결정짓는 데 큰 비중을 차지하였다. Jaesung Cho *et al.*은 외부 요인이 끼치는 영향 분석으로 돼지고기 가격을 예측하기 위해 Bidirectional LSTM (BiLSTM), Recurrent Neural Network (RNN) 및 Long Short Term Memory (LSTM) 등 3개의 모델들을 고려하였다 [3]. 여기에서 BiLSTM이 다른 모델에 비해 가장 우수한 예측 성능을 보여주었다. Zaixing Ma *et al.*은 돼지고기 가격 예측을 위하여 Dynamic Bayesian Network (DBN), Support Vector Machine (SVM) 그리고 Autoregressive Integrated Moving Average (ARIMA) 알고리즘들을 사용하였으며 DBN 모델이 최고 예측 성능을 보여주었다 [4]. 본 연구에서는 한우의 기본적인 정보, 도축 정보 및

등급명을 이용하여 4개의 인공지능 알고리즘 기반 한우 경락 가격 예측 모델을 제안하였다.

본 논문의 구성은 다음과 같다. 제 II 장에서는 본 논문에서 고려하는 원본 데이터 세트와 데이터 전처리 과정을 설명한다. 제 III 장에서는 연구에 활용한 알고리즘들과 예측 결과를 판단할 성능 지표를 제시한다. 제 IV 장에서는 앞서 언급한 인공지능 알고리즘들을 기반으로 한 예측 성능 결과를 성능 지표별로 비교한다. 최종적으로, 제 V 장에서는 본 연구의 결과를 설명하며 마무리 짓고자 한다.

II. 원본 데이터 세트 설명 및 전처리 과정

2.1 한우 개체별 출하 정보 데이터 세트 설명

본 연구에서는 스마트 팜 데이터 마트에서 제공하는 한우 개체별 출하 정보 데이터 세트를 이용한다 [5]. 한우 개체별 출하 정보 데이터 세트는 2012년부터 2024년까지의 출하 정보로 구성된 데이터를 포함하고 있다. 한우 경락 가격 예측 모델 성능을 높이기 위하여 최근 5년간의 데이터인 2018년 1월 1일부터 2024년 1월 31일까지의 데이터를 활용한다.

표 1은 한우 개체별 출하 정보 데이터 세트에 포함된 특성들과 특성별 데이터 타입을 나타내고 있다. 고려하는 특성들에 대해 설명하자면 다음과 같다. 한우 농장이 위치한 지역 (Region), 각 농장당 소유하고 있는 한우 수 (Cow scale), 농장별 이름 (Cow farm), 개체 번호 (Cow object num), 출생일자 (Cow birthday), 품종 (Cow breed), 성별 (Cow gender), 월령 수 (Cow age), 도축장 이름 (Slaughterhouse), 도축 일자 (Date of slaughter), 등급판정일자 (Date of grade), 도체중 (Cow carcass weight), 등지방 두께 (Cow thickness of backfat), 등심 면적 (Cow sirloin

area), 육량지수 (Cow weight index), 근내지방 (Cow marbling score), 조직감 (Cow tissue), 성숙도 (Cow maturity), 육색 (Cow meat color), 지방색 (Cow color of fat), 육량등급 (Cow weight grade), 육질등급 (Cow quality grade), 등급명 (Cow Total grade), 및 한우 경락 가격 (Cow auction price) 등을 포함한 총 24 개의 변수들을 포함하고 있다.

표 1. 한우 개체별 출하정보의 원본 데이터 세트 및 특성별 데이터 타입

Region [String], Cow scale [String], Cow farm [String], Cow object num [String], Cow birthday [Integer], Cow breed [String], Cow gender [String], Cow age [Integer], Slaughterhouse [String], Date of slaughter [Local Date], Date of grade [Local Date], Cow carcass weight [Double], Cow thickness of backfat [Double], Cow sirloin area [Double], Cow weight index [Double], Cow marbling score [Integer], Cow tissue [Integer], Cow maturity [Integer], Cow meat color [Integer], Cow color of fat [Integer], Cow weight grade [String], Cow quality grade [String], Cow Total grade [String], Cow auction price [Double]

2.2 원본 데이터 세트 전처리 과정

표 1 의 원본 데이터 세트에 있어서 다음과 같은 5 개의 변수들을 한우 경락 가격 예측 시 제외시켰다. 한우의 도체 값을 측정하지 못해 Cow auction price 에 결측값이 있는 데이터 샘플들은 예측 모델의 정확성을 위해 삭제하였다. Cow breed 은 모두 한우이기 때문에 품종 변수를 제외하였다. Cow object num 에는 Cow birthday 에 관한 정보가 포함되어 있기 때문에 같은 의미를 가지는 변수를 중복으로 사용하지 않기 위해 이 개체제 번호 변수를 제외하였다. 또한, Region, Cow scale 그리고 Cow farm 과 같은 특성들은 한우 경락 가격에 영향을 주는 정도가 미미하다고 판단하여 제외하였다.

표 1 에서 정의하고 있는 변수들 중에서 몇 개의 변수들은 다음처럼 서로 밀접한 관계를 가지고 있다. 무엇보다도 먼저, Cow carcass weight, Cow thickness of backfat 그리고 Cow sirloin area 에 대하여 각각 등급을 판정 후 이를 통해 Cow weight index 라는 수치를 결정하고 이 수치를 등급으로 단순히 환산하여 Cow weight grade 을 결정한다. 또한, Cow marbling score, Cow tissue, Cow maturity, Cow meat color 및 Cow color of fat 에 대하여 각각 등급을 판정하고 이 등급들을 조합하여 Cow quality grade 을 결정한다. 최종적으로 Cow weight grade 과 Cow quality grade 을 결합하여 만들어낸 변수가 Cow Total grade 이다. 본 연구에서는 Cow Total grade 을 결정하는 데 사용되는 하위 변수 11 개 (Cow carcass weight, Cow thickness of backfat, Cow sirloin area, Cow weight index, Cow weight grade, Cow marbling score, Cow tissue, Cow maturity, Cow meat color, Cow color of fat 및 Cow quality grade)를 한우 경락 가격 예측 시 제외시키고 대신에 등급명을 사용한다.

위의 전처리 과정을 마친 후 한우 경락 가격 예측 모델을 위해 고려할 특성들에 대한 데이터 세트를 Constructed Feature Dataset (CFD)라 정의하고 이를 표 2에 나타내었다. 독립 변수들은 Cow birthday, Cow gender, Cow age, Slaughterhouse, Date of slaughter, Date of grade 및 Cow total grade 등 총 7개의 특성들은

독립 변수로 사용한다. 또한, Cow auction price을 종속 변수로 사용한다. 한우 출하 정보의 데이터 샘플의 총개수는 6,580개이다.

표 2. 한우 경락 예측 모델을 위해 고려하는 데이터 세트 (Constructed Feature Dataset: CFD)

Independent Variables	Cow birthday, Cow gender, Cow age, Slaughterhouse, Date of slaughter, Date of grade, Cow total grade
Dependent Variables	Cow auction price [KRW/kg]

이때, 모델의 성능을 높이고 복잡성을 줄이기 위하여 변수들 간의 순서가 필요 없는 변수들은 명목형 변수 (nominal variable)로 변환하였다. Cow gender에서 거세 소와 암소를 각각 'Castration cow = 1' 과 'Female cow = 2' 로 설정하였고, 16개의 도축장 이름 정보를 가지고 있는 Slaughterhouse에 있어서 각각의 도축장을 1에서 16까지의 숫자로 분류하였다. Cow total grade는 이를 구성하는 Cow quality grade와 Cow weight grade들을 명목형 변수로 설정하는 과정에서 자동으로 명목형 변수로 변환된다. 예를 들면, 1++값의 Cow quality grade을 명목형 변수값 1으로 변경하고 B값의 Cow weight grade을 명목형 변수값 2으로 변경한다면, 1++B값의 Cow Total grade는 명목형 변수값 12으로 자동 변환되는 것이다. 또한, 종속 변수인 Cow auction price는 경락 가격을 측정할 때 1kg당 한국 화폐인 Korean won (KRW)를 단위로 한다.

III. 알고리즘 및 성능 지표

본 연구에서는 한우 경락 가격 예측을 위해 다음과 같은 4 개의 인공지능 알고리즘을 사용한다. 여러 개의 의사결정 트리를 결합하여 회귀분석을 수행하는 Random Forest Regression (RFR) 알고리즘, 여러 개의 독립 변수와 하나의 종속 변수의 선형 관계를 모델링하는 Multiple Linear Regression (MLR) 알고리즘, 앙상블 학습 기법으로 이전 트리의 오차를 보완해 새로운 트리를 구성하여 학습하는 Gradient Boost (GB) 과 여러 개의 결정 트리를 조합하여 예측 모델을 형성하는 eXtreme Gradient Boosted (XGBoost) 알고리즘들이다.

또한, 각 알고리즘들에 대하여 다수의 테스트를 통해 과적합을 방지하고 최고의 성능을 보이는 알고리즘별 하이퍼 파라미터는 다음과 같다. RFR 알고리즘 최적의 트리 깊이 (tree depth)와 트리 모델 개수는 각각 20 과 200 이다. GB 알고리즘 최적의 트리 깊이, 트리 모델 개수와 Learning rate 는 각각 10, 200 그리고 0.8 이다. 마지막으로, XGBoost 알고리즘 최적의 Boosting rounds 와 Base score 는 각각 100 과 0.7 이다.

예측 모델들에 대한 성능 평가 지표로는 0 과 1 사이의 범위에서 계수의 값이 1 에 가까울수록 성능이 좋다고 평가하는 R-squared (R^2) 과 실제 값과 예측한 값의 차이를 표준편차로 나타내는 Root Mean Squared Error (RMSE)를 사용하였다.

IV. 한우 경락 가격 예측 분석 및 결과

한우 경락 가격 예측 모델을 위해 표 2에서 정의하고 있는 CFD에 대하여 학습 데이터 세트와 테스트 데이터 세트로 각각 80%와 20%로 분할하여 사용하였다. 또한, 제시하는 예측 모델에 대한 성능

신뢰성을 확보하기 위해서 다섯 번의 교차 검증을 실시하는 K겹 교차 검증 (K-fold cross validation) 기법을 사용하였다.

그림 1은 본 연구에서 고려하고 있는 4개의 알고리즘 기반 한우 경락 가격 예측 모델들로부터 얻은 R²값이다. 4개의 알고리즘 중 XGBoost 알고리즘은 R²값이 0.857로 가장 우수한 예측 모델 성능을 보여주는 반면에, MLR 알고리즘은 R²값이 0.699로 예측 성능이 가장 저조함을 확인할 수 있었다. XGBoost, RFR, GB 그리고 MLR 순서로 모델 성능이 떨어짐을 알 수 있다.

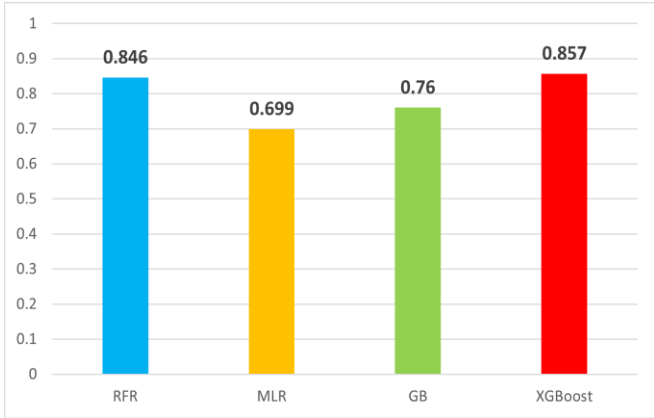


그림 1. 한우 경락 가격 예측 모델별 분석 결과 (R²)

그림 2는 한우 경락 가격 예측 모델들에 대한 또 다른 성능 평가 지표인 RMSE 값을 나타낸 것이다. XGBoost 기반 모델이 가장 작은 RMSE 값을 가지기 때문에 가장 우수한 성능을 보임을 확인할 수 있다. 특히, XGBoost 기반 모델로부터 얻게 되는 RMSE 값이 1,652.17 [KRW/kg로 이는 XGBoost 기반 예측 모델을 이용할 경우 실제 한우 경락 가격들과 비교하였을 때 1 kg당 약 1,652.17원의 오류가 발생함을 알 수 있다. 표 2에서 정의하고 있는 데이터 세트에서 평균 한우 경락 가격이 약 18,664 [KRW/kg]인 것을 고려했을 때, 실제 평균 가격 대비 XGBoost 기반 모델의 RMSE값 비율은 8.85%이다. 또한, RMSE값을 살펴보았을 때 XGBoost, RFR, GB 그리고 MLR 순서로 모델 성능이 저하됨을 알 수 있다. 이는 그림 1의 R² 성능 지표를 통해 확인하였던 모델들의 우수한 성능에 대한 순서와 동일함을 알 수 있다.

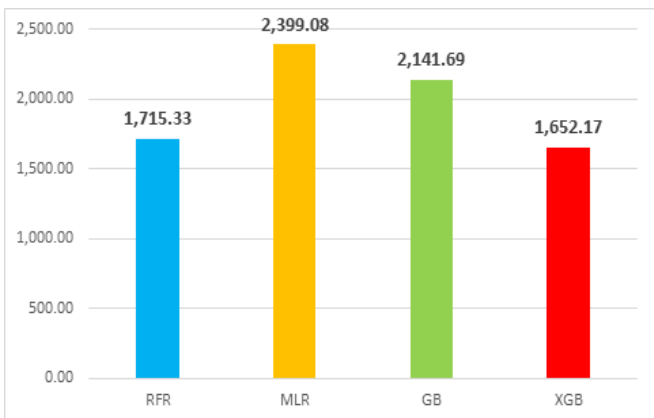


그림 2. 한우 경락 가격 예측 모델별 분석 결과 (RMSE)

V. 결론

본 연구에서는 한우의 출하 정보를 이용하여 인공지능 알고리즘 기반 한우 경락 가격을 예측하는 모델들을 제시하였다. 이를 위해 스마트 팜 데이터 마트에서 제공하는 한우 개체별 출하 정보 데이터 세트를 이용하였다. 한우 경락 가격 예측 모델을 위해 사용될 변수들을 선택하기 위해서 한우 개체별 출하 정보 데이터 세트에 포함된 변수들에 대하여 전처리를 수행하였다. 또한, 한우 개체별 출하 정보 데이터 세트에는 서로간의 밀접한 관계를 가지고 있는 변수들이 있는데, 이러한 관계성에 따라 도출되는 한우별로 주어지는 등급명이라는 변수를 독립변수로 고려하였다. Random Forest Regression, Multiple Linear Regression, Gradient Boost와 eXtreme Gradient Boosted 등 총 4 개의 인공지능 알고리즘들을 한우 경락 가격 예측을 위해 사용하였다. 성능 지표로 고려한 R²와 RMSE값 모두를 살펴보았을 때 총 4개의 알고리즘 기반 모델들 중에서 XGBoost 알고리즘 기반 모델이 가장 우수한 성능을 보였다. XGBoost 알고리즘 기반 모델 다음으로 RFR, GB 그리고 MLR 기반 모델들의 순서로 성능이 떨어짐을 확인하였다.

한우 경락 가격은 한우의 도축 날짜와 등급판정일자 사이의 시간 차이 뿐만 아니라 대중들의 한우 소비 경향과 같이 여러 다양한 외부요인의 영향을 받는다. 그러므로 이렇게 끊임없이 변화하는 한우 경락 가격을 예측하는 모델의 성능을 향상시키기 위해서 향후 연구로서 다양한 외부요인들을 함께 고려하는 새로운 접근 방식이 필요하다.

ACKNOWLEDGMENT

Following are results of a study on the "Leaders in Industry-university Cooperation 3.0" Project, supported by the Ministry of Education and National Research Foundation of Korea.

참 고 문 헌

- [1] Hanwoo Board Organization, "2022 Hanwoo Consumption Trend Survey," May. 2023.
- [2] H.R. Park, S.H. Eum, J.H. Park, J.K. Seo, S.K. Cho, T.S. SHIN, B.W. CHO, H.C. Park, E.J. Lee, D.W. Sun, H.T. Lim, J.G. Lee, and B.W. Kim, "Contribution Analysis of Carcass Traits on Auction Price in Gyeongsangnam-do Hanwoo," Journal of Agriculture and Life Science, vol. 49, no. 6, pp. 187-195, Oct. 2015.
- [3] J.S. Choi and M.K. Kim, "Comparative Study of Deep Learning Methods for Predicting Pork Prices," Proceedings of Symposium of the Korean Institute of communications and Information Sciences, pp. 1,176-1,177, Feb. 2020.
- [4] Z. Ma, Z.M. Chen, T.T. Chen, and M.W. Du, "Application of Machine Learning Methods in Pork Price Forecast," In Proceedings of the 2019 11th International Conference on Machine Learning and Computing, pp. 133- 136, Feb. 2019.
- [5] Smart Farm Data Mart, "Korean Beef Dataset," Feb. 2024.