

축산물 가격 예측 모델을 기반으로 한 레시피 추천 서비스 개발

2022년 축산데이터 분석·ICT 활용 아이디어 경진대회

팀명: 음~메로나

팀장: 함양훈

팀원1 : 양정미

팀원2 : 이윤혁

제출일: 22.11.14

목차

서론.....	3
1. 데이터 분석	4
1-1. 전처리.....	4
1-1. 분석 프로세스	5
1-2. 분석 방법.....	6
가. 가격 예측 모델.....	6
나. 레시피 추천 모델	7
2. 분석 활용 전략.....	9
2-1. 추천 모델 프로세스	10
2-1. 추천 모델 활용 방향	10
3. 기대효과 및 시사점.....	11

서론

국내 축산물 업계는 코로나 팬데믹에 의해 매우 큰 타격을 입어 어려움을 겪고 있다. 2019 년 코로나에 의한 대체 수요 증가로 37.2%의 소고기 자급률 상승을 보였지만, 이 현상은 일시적인 효과라고 전문가들은 말한다. 다음으로, 최근 2 년 사이 배합사료 가격이 31.5~33.4%, 조사료 가격이 30.6% 폭등했고 낙농가의 실질생산비가 1000 원 내외를 육박했으며 2001 년 수입자유화 이후 소고기 수입량은 2015 년부터 2020 년까지 총 41.1% 증가했다. 그리고, 소비자들의 수입에 대한 인식과 선호도 또한 중요한 요소로 꼽히고 있다. 그 이유는, 한우 가격이 수입산에 비해 너무나 비싼 가격이 큰 이유라 말한다.



그림 1 은 2017 년부터 2022 년까지 원산지 별 고기 평균 가격 변화 그래프이다.

(그림 1)

이처럼 국내 축산업은 여러 외부적인 요소에 의해 가격은 점차 증가하지만 소비는 줄어 들고 있다.

따라서, 본 프로젝트는 저조한 국내 축산물 소비에 문제점을 파악하고 소비자의 입장에서 소비를 촉진시킬 수 있는 방안을 찾고자 하였고 그에 따른 추천 모델을 제시하고자 한다.

1. 데이터 분석

본 프로젝트는 축산물 가격 예측 모델과 레시피 추천 모델로 이루어져 있다. 예측 모델에서 사용한 축산품 가격 데이터는 KADX 농식품 빅데이터 거래소의 '농식품 일자별 도소매 가격' 데이터를 사용하였다. 추천 모델에는 KADX 농식품 빅데이터 거래소의 '무료 레시피' 데이터와 축산물품질평가원에서 제공한 시계열 데이터인 '소비자

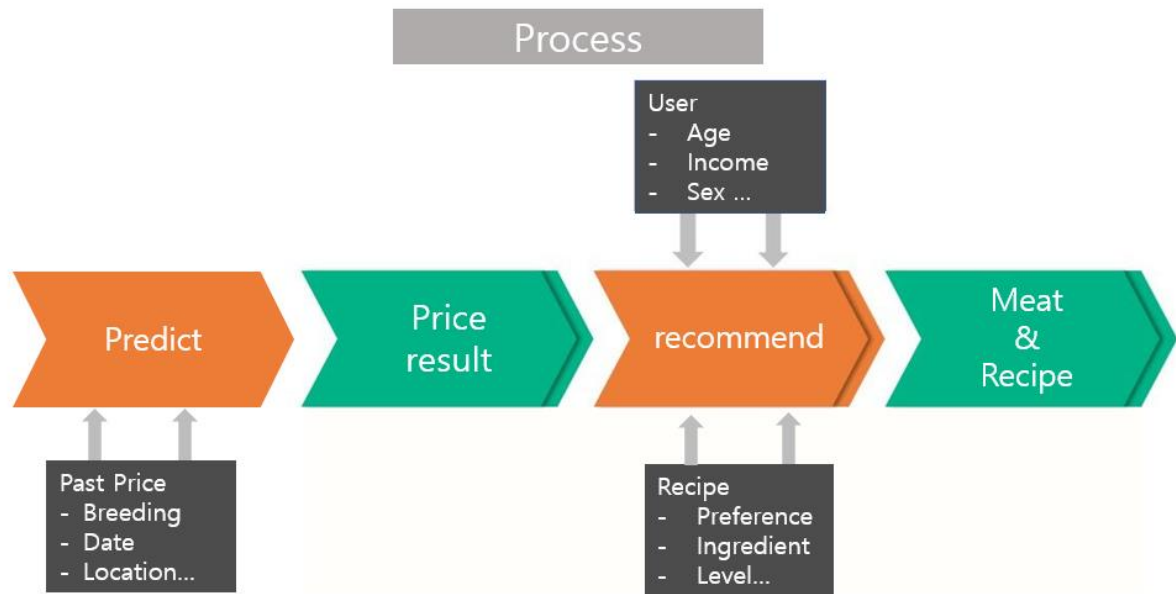
패널'데이터를 전처리 하여 활용하였다. 소비자 패널 데이터는 개개인의 취향에 알맞은 고기를 추천할 수 있도록 한 사람마다 여러 가지 선호도가 포함된 속성들이 있기에 이 데이터를 사용한 것이다.

1-1. 전처리

2017년부터 2022년까지 가격 데이터는 농축산 데이터가 같이 포함되어 있어 고기 데이터만 추출해 활용했다. 정확한 분석을 위해 국내산과 수입산을 분리했고 수입 고기 데이터는 여러 무게 단위가 섞여 있어 100g으로 통일시켜 전처리를 진행하였다. 소비자 패널 데이터는 구매한 고기, 구매 목적, 재구매 의사, 성별, 나이 등 변수 중요도가 높은 속성들이 대부분 문자형으로 구성 되어있어 수치형으로 변환하여 분석을 진행했다.

고기 가격 데이터는 원산지별 세부 부위로 나뉘져 데이터가 가진 크기와 편차가 다르다. 그 때문에 한 속성의 특징을 너무 많이 반영하거나 패턴을 찾아내는데 문제가 발생하기 때문에 정규화를 진행한다. 해당 모델에 사용한 스케일러는 데이터들을 0~1사이 값으로 변환시켜주는 'MinMaxScaler'로 스케일을 진행하였다.


1-2. 분석 프로세스



예측 모델은 수집된 육종 가격 데이터를 분석하여 소비자에게 최적의 가격으로 나올 고기들을 예측하여 그 결과값을 추천 모델에 적용한다. 다음으로, 추천 모델은 소비자 패널 데이터로 군집분석을 진행하여 레시피 데이터와 융합한다. 합리적인 가격과 소비자의 선호도를 고려하여 원하는 고기가 사용된 레시피를 추천해준다. 따라서 소비자들의 축산물 구매를 유도하는 모델을 생성하는 것이다.

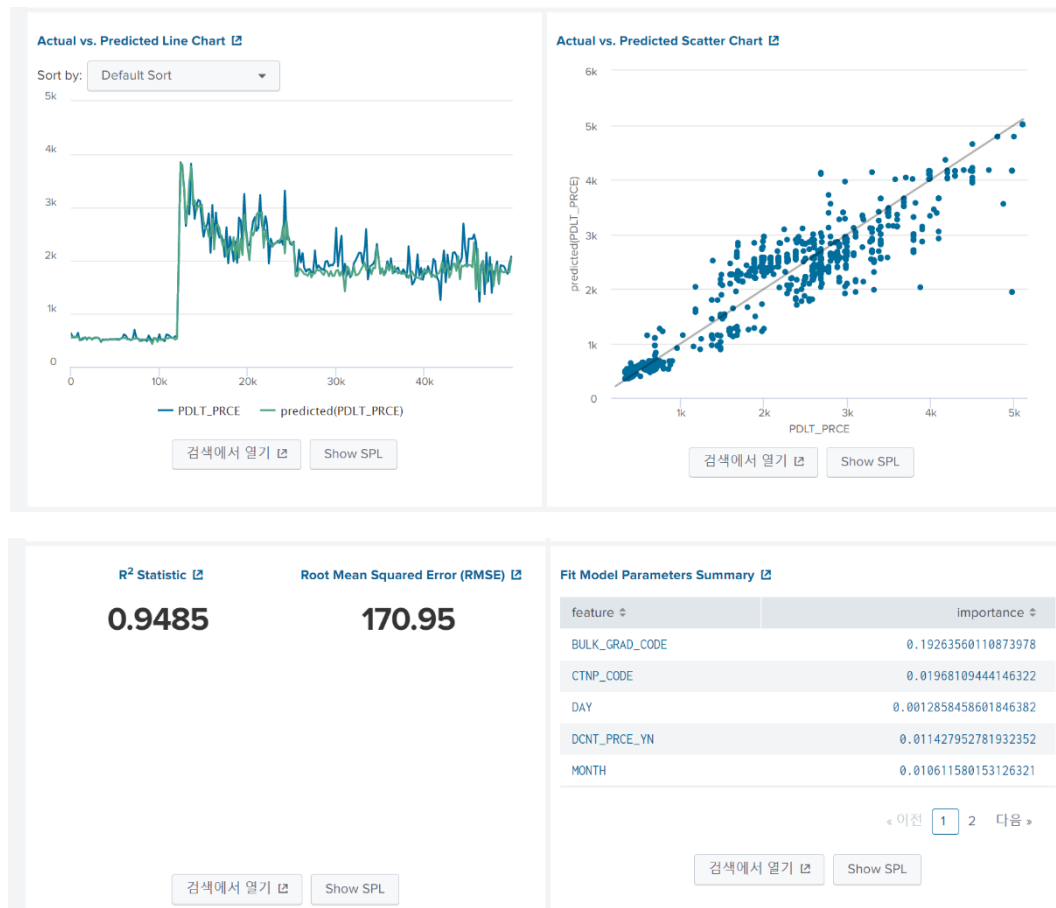
1-3. 분석 방법

가. 가격 예측 모델 가격 예측 모델 생성 시 4개(Catboost, SVM, LGBM, random forest)의 부스트 기법을 사용하였다. 각 모델 생성 후 모델 성능 측정 지표인 결정계수(r^2 score)를 이용하여 최종 사용 모델을 선정하였다.

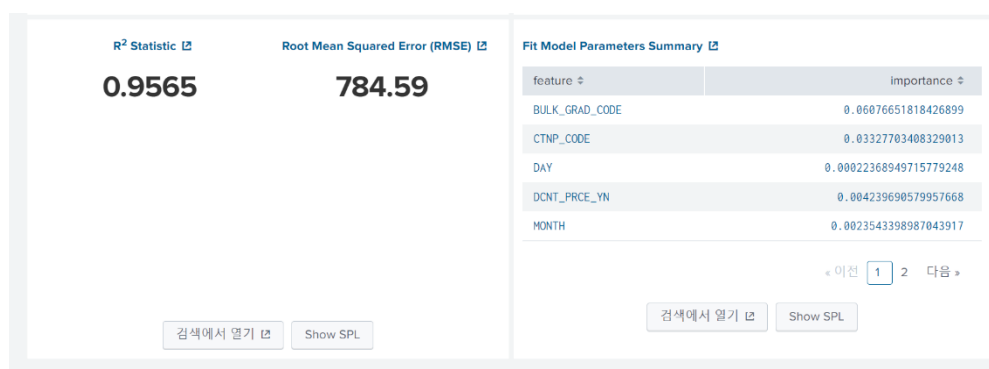
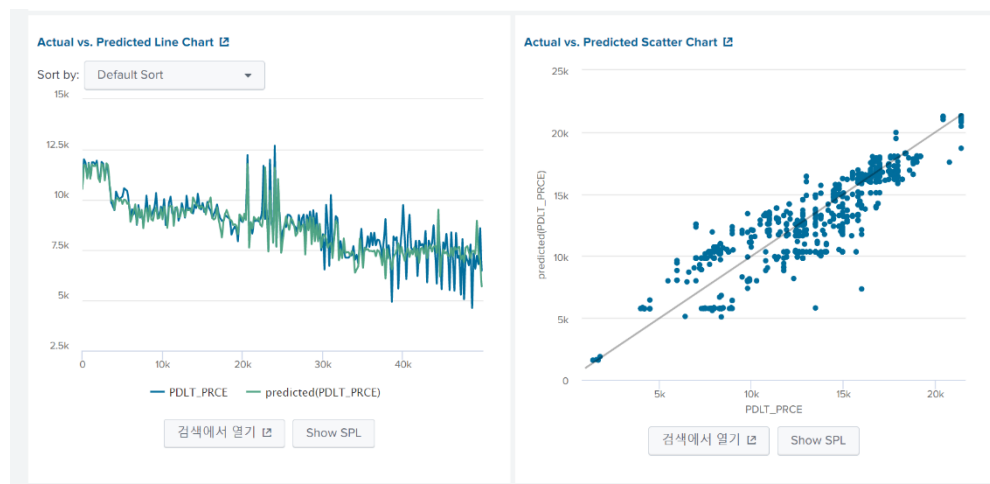
Period	Random Forest	LGBM	CatBoost	SVM
 score	99%	97%	98%	42%

그 중에서 가장 정확도가 높게 나온 랜덤 포레스트를 사용하였다. 랜덤 포레스트

트는 지도 학습 머신 러닝 알고리즘 중 하나이다. 정확성, 단순성 및 유연성으로 인해 가장 많이 사용되는 알고리즘 중 하나이고, 예측, 분류 및 회귀 작업에 사용된다. 아래 그림은 Splunk의 MLTK 앱 기능을 활용하여 동일한 데이터와 예측 모델 사용시 결과를 비교한 그래프이다. 그림 4는 각각 수입산, 국내산 고기 가격 예측 모델의 정확도이다.



(그림 3)



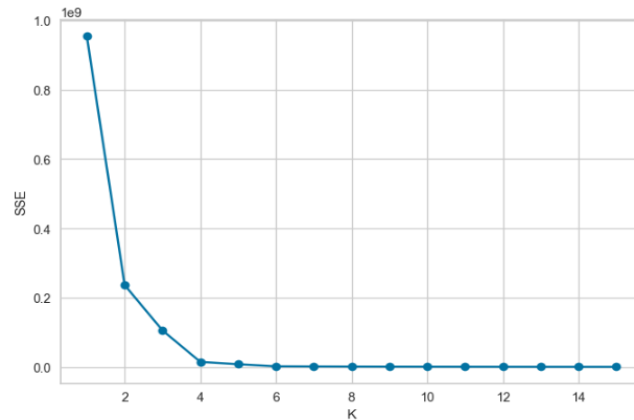
(그림4)

각각의 가격 예측 모델의 정확도는 국내산 95%, 수입산 94%이다. Splunk와 파이썬으로 만든 모델 모두 정확도에서 큰 차이가 없어 모델 최적화가 잘되어 있음을 알 수 있다.

나. 레시피 추천 모델

소비자 패널 데이터를 추천 모델에 사용할 수 있도록 군집화 모형을 생성해야 한다. 여기서 군집화 모형은 데이터의 유사도를 바탕으로 개별 데이터를 그룹에 할당하는 (라벨을 만들어내는) 비지도 학습인 K-Means 군집화 알고리즘을 사용

했다. K-Means 알고리즘으로 군집화 하기에 앞서, 군집 개수인 k값을 탐색한다. 오차제곱합(SSE)의 값이 최소가 될 때 군집의 중심을 결정해 나가는 방법인 elbow 기법으로 분석해본 결과(그림5), k값은 3에서 6사이의 값이라는 것으로 파악됐다.



(그림5)

이제 k값으로 군집의 품질을 정량적으로 계산해주는 silhouette 기법을 활용하여 정확한 값을 도출한다. k값이 3,4,5,6일 때 silhouette_score 값은 0.769, 0.891, 0.908, 0.912, 0.833 이다. silhouette_score 값은 1에 가까울수록 군집의 품질이 좋으며 정확하게 형성된 것이다.

따라서 silhouette_score 값 중 가장 높은 값은 0.912로 k값은 6으로 선정했다.

2. 분석 활용 전략

2-1. 추천 모델 프로세스

소비자 패널 데이터 군집 결과를 바탕으로 레시피 데이터를 혼합하여 해당 소비자에게 적합한 레시피를 추천하는 추천 모델을 개발했다. 추천 모델의 메커니즘

을 살펴보면 다음과 같다. 새로운 소비자의 정보가 들어오면 해당 소비자는 6개의 군집 중, 어떤 군집에 속하는지 파악해야 한다. 새로운 소비자의 정보는 Random Forest 모델에 적용하여 어떤 군집에 속하는지 예측한다. 그 후 해당하는 군집이 선호하는 고기의 레시피 데이터를 제공해주는 것이다. 레시피 데이터를 제공할 때는 인기순, 난이도순으로 제공해줄 수 있다. 또한 가격 예측 모델에서 나온 결과 리스트 중 저렴한 고기를 바탕으로 인기있는 레시피 데이터를 제공할 수 있다. 요약하자면, 추천 모델은 해당 소비자의 선호도, 요리 난이도를 고려하여 적합한 레시피를 추천할 수 있으며, 저렴한 고기를 예측하여 해당 고기의 인기있는 레시피를 추천할 수 있다. 소비자 측면에서는 자신에게 선호도가 높을 수 있는 고기 레시피와 구매할 저렴한 고기의 레시피 정보를 받을 수 있는 것이다.

2-2. 추천 모델 활용 방향

본 프로젝트에서 만든 모델은 현재 저조한 국산 축산물 소비에 도움을 주고자 하는 공익적인 목표에서 시작하였다.

오늘의 추천 고기: 앞다리살 예상 가격(100g 기준): 1270.0358333333331

추천 레시피 목록

음식명: 돼지고기호박까나페 난이도: 초급

추천 레시피: [재료] 돼지고기 뒷다리살 500g| 애호박 [양념] 고추장 2스푼| 조청(물엿이나 설탕으로 대체가능) ;

음식명: 콩불 난이도: 초급

추천 레시피: [재료] 돼지고기 적당량| 콩나물 1~2주먹| 고추장 2큰술| 설탕 or 올리고당 3큰술| 양파 1/2~1개| 0

음식명: 마파두부 난이도: 초급

추천 레시피: [식재료] 다진돼지고기 250g| 두부 500g| 감자 1주먹크기| 대파 1단소| 홍고추 2개| 전분 2T [양념]

(그림6)

그림6은 완성된 모델의 결과를 보여주고 있다. 완성된 모델의 사용법에 대해 설명을 하자면, 소비자가 원하는 고기가 가장 싼 시기를 미리 예측하여 그 가격을

보여준다. 그리고 해당 고기가 사용되는 음식의 레시피를 추천함으로써 소비 촉진에 적극적으로 맞추어져 있음을 볼 수 있다.

이처럼 완성된 모델은 추후 식품 쇼핑몰에 활용될 수 있다. 예를 들어, ‘ssg몰’, ‘이마트몰’, ‘온라인 하나로마트’ 등 온라인 쇼핑몰의 시스템에 활용되어 이용자가 시기 적절하게 구매할 수 있는 식품을 확인하고 기념일, 명절 혹은식사 자리에 어울리는 음식의 레시피를 추천함으로써 이용자에게 편리함과 서비스의 풍부함을 제공해 줄 수 있을 것이다.

3. 기대효과 및 시사점

본 프로젝트의 예측 모델에서 나온 결과인 축산물 예상 가격데이터를 추천 모델에 적용하여 원하는 정보를 소비자에게 제공함으로써 소비와 생활 부분에 양질의 서비스를 제공할 수 있을 것이다. 개발한 모델로 소비자와 축산업계 양측이 이득을 볼 수 있는 메커니즘을 다시 살펴보면 다음과 같다. 예측 모델은 소비자들이 원하는 고기를 합리적인 가격에 구매를 할 수 있도록 도와주고, 추천 모델은 원하는 고기가 사용된 레시피 정보를 제공함으로써 축산품 소비에 다양성을 향상시키고, 촉진 효과까지 기대할 수 있을 것이다. 그리고, 축산업계의 소매업 측면에서 폐 고기 발생 비율이 감소하고 고기 판매량이 증가하는 등 축산업계의 수익 향상에 도움이 될 것이다.

본 프로젝트는 개발한 모델을 활용하여 축산업계의 경제적 활성화를 이바지하고자 하는 것이다. 넓은 시선으로 바라본다면, 축산업계의 한정적인 경제적 활성

화뿐만 아니라 다양한 시장의 경제적 활성화 측면에서도 도움이 될 수 있을 것
이라 기대한다.