

<1P>

안녕하세요 저희는 12조 이김했조이고 발표를 맡은 김소영입니다.
저희 프로젝트명은 나는 착유한다, 고로 존재한다'입니다.
프로젝트명에서 알 수 있듯이,
스마트 축산 착유기 데이터를 기반으로 프로젝트를 진행했습니다.

<2P 목차>

목차는 다음과 같습니다

<3P 컬럼 설명>

저희는 데이터를 세 가지 기준으로 나누었습니다.
종합 정보는 농장아이디, 개체번호가 있고
기계정보는 착유회차, 착유 시작/종료 일시, 공기흐름이 있고
건강 정보는 착유량, 전도도, 혈액흐름, 온도, 유지방, 유단백이 있습니다.
여기서 온도는 착유기를 통해 측정된 값이므로 우유 온도라고 판단했고,
이를 직장 체온과의 관계를 이용해 소의 체온을 추정하여
건강 지표로 활용했습니다.
저희는 표시된 컬럼을 대상으로 착유량 중심 분석을 진행했습니다.

<4P 문제 정의 및 가설설정>

자동화 중심의 국내 스마트 축산 환경에서,
착유량 향상을 위해 농장이 어디에 중점을 두어야 하는지
방향성을 제시하고자 했습니다.
또한, 여러 논문에서 제시하는 변수들의 범위가 과연 정확한지 검증하고자 했습니다.
따라서 목표는 다음과 같습니다.
착유기 센서 데이터를 기반으로 착유량에 영향을 미치는 요인을 분석하고
이를 바탕으로 착유량 예측 머신러닝 모델을 구축하고
농장주가 착유 시스템을 쉽게 파악할 수 있는 대시보드를 구현하는 것입니다

< 5P 문제 정의 및 가설설정>

저희가 세운 가설은 다음과 같습니다.
첫번째, 착유회차가 높을수록 평균 착유량은 감소할 것이다
두번째, 온도가 높을수록 평균 착유량은 감소할 것이다.
세번째, 전도도가 높을수록 평균 착유량은 감소할 것이다

<6P 분석(전처리)>

가설 검증 전에 전처리를 진행했습니다.
사용하지 않는 컬럼을 제거하고,
유지방과 유단백에서 결측치가 존재하는 52개의 행을 제거하였습니다.
또한 농장아이디와 개체번호를 제외한
모든 컬럼이 중복되는
9000개의 행을 제거하여 분석의 정확도를 높였습니다.
수치형 컬럼인 농장아이디, 개체번호를 카테고리형으로 변환하였고
혈액흐름은 T/F의 문자형 데이터라 원-핫 인코딩으로 변환하였습니다.
마지막으로 영양상태를 진단하는 기준인 유단백 대비 유지방 비율을 PFR 컬럼으로 정의하였고
착유종료일시에서 착유시작일시를 뺀 값을 분으로 환산하여 착유소요시간(분)으로 저장하였습니다

<7P 분석(가설검증)>

전처리 후 가설검증을 진행하였습니다.

첫번째 가설, 착유회차가 높을수록 평균 착유량이 감소할 것이다

실제로 착유회차가 높아질 수록 평균 착유량은 감소하는 경향이 나타나, 가설과 일치했습니다.

<8P 분석(가설검증)>

두번째 가설, 온도가 높을수록 평균착유량이 감소할 것이다

그 결과 35도대까지는 착유량 감소 경향이 있었으나 이후 점점 증가하는 추세를 보였으며

42도대에서 최고 착유량이 나타났습니다.

<9P 분석(가설검증)>

세번째 가설, 전도도가 높을수록 착유량이 감소할 것이다

그래프를 보시면 전도도가 높아질수록 약간 증가하다가 8.6 이후로 감소 추세를 보였으나,

9.1과 10처럼 예외적으로 튀는 값도 존재했습니다.

<10P 분석(이상탐지)>

논문에서 제시하는 범위가 우리 데이터에도 맞는지 검증하기 위해 이상탐지를 진행했습니다.

먼저 건강위험 지표들에 대한 이상탐지 결과입니다.

저희는 유방염, 대사문제, 번식장애 등 소의 건강 상태에 영향을 주는

온도, 전도도, PFR을 건강 기준 컬럼으로 설정하고,

Isolation Forest로 분석한 결과 189개의 이상치가 발견되었습니다.

<11P 분석(이상탐지 - 실제지표와 비교)>

이상탐지 결과를 실제 지표와 비교하였습니다.

그 결과 실제 지표에서의 정상데이터는

모델도 거의 대부분 정상으로 예측하였습니다.

하지만 모델이 이상치라고 예측한 것 중

실제로 이상치인 경우는 6%라는 결과를 도출하였습니다.

저희는 오탐지를 무시할 수 없다는 결론을 내렸고

건강 위험 판단은 실제 지표 범위를

최종 기준으로 설정하였습니다.

<12P 분석(이상탐지)>

다음으로 기계오류에 대한 이상탐지를 하였습니다.

여기서는 착유기계종류, 착유방법, 착유시설, 기계 압력 등에 영향을 받는

공기흐름, 착유소요시간을 기계 오류 기준으로 설정하고

Isolation Forest로 분석한 결과 944개의 이상치가 발견되었습니다.

두 컬럼은 정확한 범위를 찾지 못해서 이상탐지를 통해 찾은 범위를 데이터에 적용 하였습니다.

<13P 분석(상관관계 - 히트맵)>

모델 분석에 들어가기전 히트맵으로 확인한 결과,

변수 간 강한 선형 관계가 없었으므로 비선형 모델로 분석했습니다.

<14P 분석(모델링)>

그 다음으로

어떤 모델이 저희 데이터에 적합한지 판단하기 위해

LightGBM, CatBoost, RandomForest, XGBoost

이렇게 네 가지 모델을 대상으로 학습을 진행해 보았습니다.

그 결과 **CatBoost**가 **R²** 값이 가장 높고 과적합이 적어

저희 데이터에 가장 적합하다고 판단하였습니다.

<15P 분석(모델링)>

이후 **optuna**를 이용해 하이퍼파라미터를 최적화하여

과적합을 줄이고 성능을 높인 최종 모델을 선정했습니다.

그 결과 **R²** 값은 **1.5%** 상승했고,

MAE 값은 **0.042** 감소하여 초기 모델보다 약간의 성능 향상을 확인할 수 있었습니다.

<16P 해석 및 적용>

저희는 이렇게 완성된 모델을 스트림릿으로 구현하여,

농장주가 변수 값을 입력하면 착유량을 예측할 수 있는 서비스를 만들었습니다.

<17P 해석 및 적용>

또한 앞서 진행한 이상탐지를 바탕으로 범위를 최종적으로 설정했습니다.

기계는

비정상 범위를 공기 흐름이 **0.9**에서 **5.9**를 벗어나거나,

착유 시간이 **3분**에서 **13분**을 벗어날 경우로 정의했고,

두 조건을 모두 벗어나면 기계 오류로 의심했습니다.

건강은 실제지표를 이용하였고

온도는 **41도** 이상일때,

전도도는 **7.3**이상일 경우,

PFR은 **0.7~1.0**을 벗어나는 것이 비정상 범위였습니다.

모두 정상범위이면 정상,

세 개 중 하나라도 벗어나면 관심필요,

모두 벗어나면 조치필요로 분류하였습니다.

이 범위를 통해 분류를 진행하고

태블로를 이용하여 농장주가 현황을 쉽게 파악할 수 있도록 대시보드를 제작하였습니다.

저희가 제작한 대시보드를 소개하겠습니다.

<소 대시보드>

먼저 소 대시보드입니다.
로봇 착유기를 사용하는 농장주의 입장에서 생각했을 때,
가장 궁금한 것은
‘이 착유기의 언제, 어디서 문제가 발생했는지’와
‘내가 키우는 소들의 현재 상태와 질병 의심 여부’일 것이라고 판단했습니다.
그래서 시트를 농장별, 날짜별, 개체번호별로 확인할 수 있게 구성했고,
특히 개체별 건강 분류 시트에서는
착유회차별 개체의 착유현황을 확인하다가
다음과 같은 문제를 발견했을때
이것을 선택하면
아래에 어떤 문제로 조치필요라는 결과가 나왔는지 확인할 수 있습니다.
또한 각각의 시트 제목 앞에 아이콘을 배치해
농장주가 원하는 정보를 빠르게 찾을 수 있도록 했습니다.
예를 들어, 착유 아이콘이 있는 시트에서는 착유기 관련 정보를,
병원 아이콘이 있는 시트에서는 건강 상태를,
소 아이콘이 있는 시트에서는 착유량이 높은 소들의 특징을 확인할 수 있습니다.

<우유 대시보드>

다음은 우유 대시보드입니다.
‘우리 목장에서 착유한 우유가 다른 농장과 비교해서 어떤 수준일까’,
‘내가 생산한 우유가 좋은 품질의 우유일까’를 농장주가 궁금해할 수 있습니다.
그래서 소 대시보드와 동일한 필터들에 더해
우유 등급별로 특징을 확인할 수 있는 필터를 만들고,
착유 회차별, 온도별, 착유 시간별로 특징을 분석할 수 있도록 구성했습니다.
여기있는 우유 등급 필터를 활용하면 이 대시보드를 가장 효과적으로 사용할 수 있습니다.

<착유량 예측 대시보드>

마지막으로 착유량 예측 대시보드입니다.
PPT에서 보여드린 스트림릿 서비스를 대시보드에도 통합하여,
농장주가 각 변수 값을 입력하고 ‘예측하기’를 누르면
해당 조건에서의 착유량과 우유 등급을 바로 확인할 수 있도록 했습니다.
이를 통해 소 관리나 농장 운영의 효율성을 높이는 데
도움을 줄 수 있을 것이라 생각합니다.

발표를 마치겠습니다 감사합니다