

발간 등록 번호

11-1543000-004557-01

# 축산 스마트팜 ICT 데이터 표준화 및 데이터 수집활용체계 개선방안 연구

2024. 03.

이문용

산업및시스템공학과, 교수  
지식혁신연구센터, 센터장

**KAIST**



# 목 차

<b>제 1 장. 연구개요</b> .....	<b>1</b>
1.1 연구배경 .....	1
1.2 연구의 필요성 .....	4
1.3 연구의 목적 .....	5
<b>제 2 장. 연구방법</b> .....	<b>7</b>
2.1 분석 방법론 .....	7
2.1.1 비교 사례 연구 (Comparative Case Study) .....	7
2.1.2 시나리오 분석 (Scenario Analysis) .....	8
2.2 분석 대상 .....	9
2.2.1 ICAR (International Committee for Animal Recording) .....	9
2.2.2 스마트팜 데이터마트 .....	13
<b>제 3 장. 스마트팜 데이터마트 데이터 문제점 분석</b> .....	<b>20</b>
3.1 분석방법 .....	20
3.2 한우 데이터 .....	21
3.3 양돈 데이터 .....	60
<b>제 4 장. 스마트팜 데이터마트 데이터 연결성 문제점 분석</b> .....	<b>77</b>
4.1 집계 데이터 연결성 분석 .....	77
4.1.1 한우 집계 데이터 셋 분석 .....	77
4.1.2 양돈 집계 데이터 셋 분석 .....	78
4.2 로우 데이터 연결성 분석 .....	80
4.2.1 한우 로우 데이터 셋 분석 .....	80
4.2.2 양돈 로우 데이터 셋 분석 .....	82
4.3. 시나리오 기반 데이터 상호 연결성 분석 .....	83
4.3.1 관련 연구 사례 .....	83
4.3.2 시나리오 분석 .....	85
<b>제 5 장. 결론 및 제언</b> .....	<b>94</b>
<b>제 6 장. 참고 문헌</b> .....	<b>100</b>
<b>제 7 장. 부록</b> .....	<b>102</b>

# 표 목 차

<b>제 2 장. 연구방법</b> .....	7
표 1. AI Hub 축산데이터의 등재된 데이터 종류 및 유형.....	14
표 2. 스마트팜 데이터마트 축산 데이터 기초 통계량.....	18
표 3. 스마트팜 축산 데이터 셋 리스트 .....	19
<b>제 3 장. 스마트팜 데이터마트 데이터 문제점 분석</b> .....	<b>20</b>
표 4. ICAR 표준과 스마트팜 데이터마트 비교 표 - 개체별 번식이력(한우) .....	23
표 5. ICAR 표준과 스마트팜 데이터마트 비교 표 - 연도별 번식성적(한우).....	26
표 6. ICAR 표준과 스마트팜 데이터마트 비교 표 - 개체별 출하정보(한우).....	28
표 7. ICAR 표준과 스마트팜 데이터마트 비교 표 - 연도별 출하성적(한우).....	31
표 8. ICAR 표준과 스마트팜 데이터마트 비교 표 - 개체별 혈통족보(한우).....	33
표 9. ICAR 표준과 스마트팜 데이터마트 비교 표 - 개체별 혈통성적(한우).....	36
표 10. ICAR 표준과 스마트팜 데이터마트 비교 표 - 개체별 EPD(한우).....	38
표 11. ICAR 표준과 스마트팜 데이터마트 비교 표 - 체중(한우).....	41
표 12. ICAR 표준과 스마트팜 데이터마트 비교 표 - 환경관리기(한우) .....	43
표 13. ICAR 표준과 스마트팜 데이터마트 비교 표 - 사료빈(한우).....	45
표 14. ICAR 표준과 스마트팜 데이터마트 비교 표 - 일반급이기(한우).....	47
표 15. ICAR 표준과 스마트팜 데이터마트 비교 표 - 군사급이기(한우).....	49
표 16. ICAR 표준과 스마트팜 데이터마트 비교 표 - 음수관리기(한우).....	51
표 17. ICAR 표준과 스마트팜 데이터마트 비교 표 - 포유기(한우).....	53
표 18. ICAR 표준과 스마트팜 데이터마트 비교 표 - TMR 급이기(한우).....	55
표 19. ICAR 표준과 스마트팜 데이터마트 비교 표 - 발정탐지기(한우) .....	57
표 20. ICAR 표준과 스마트팜 데이터마트 비교 표 - 환경제어기(한우).....	59
표 21. ICAR 표준과 스마트팜 데이터마트 비교표 - 개체별 번식로우(양돈).....	62
표 22. ICAR 표준과 스마트팜 데이터마트 비교표 - 연도별 번식통계(양돈).....	64
표 23. ICAR 표준과 스마트팜 데이터마트 비교표 - 월별 번식통계(양돈).....	66
표 24. ICAR 표준과 스마트팜 데이터마트 비교표 - 산차별 번식통계(양돈).....	68
표 25. ICAR 표준과 스마트팜 데이터마트 비교표 - 급이기(양돈).....	70
표 26. ICAR 표준과 스마트팜 데이터마트 비교표 - 음수기(양돈).....	72

표 27. ICAR 표준과 스마트팜 데이터마트 비교표 - 환경정보(양돈).....	74
표 28. ICAR 표준과 스마트팜 데이터마트 비교표 - 포유모돈 급이번식 데이터(양돈).....	76

# 그림 목 차

<b>제 1 장. 연구개요</b> .....	<b>1</b>
그림 1. 2016 년~2022 년 연도별 농가 인구 및 가축 수 추이 .....	1
그림 2. 2016 년~2022 년 전국 1 인당 사육 가축 수 추이.....	2
그림 3. 스마트팜 데이터마트 축산 정형데이터 건수 .....	3
<b>제 2 장. 연구방법</b> .....	<b>7</b>
그림 4. AI Hub 홈페이지 .....	13
그림 5. AI Hub 축산데이터 등재된 11 건의 데이터 유형 분포 .....	14
그림 6. KADX 홈페이지 .....	15
그림 7. 스마트팜 R&D 빅데이터 플랫폼 홈페이지 .....	16
그림 8. 스마트팜 데이터마트 홈페이지 .....	17
<b>제 3 장. 스마트팜 데이터마트 데이터 문제점 분석</b> .....	<b>20</b>
그림 9. 한우 데이터 셋 단계별 선택.....	21
그림 10. 개체별 번식이력(한우) 데이터 셋.....	22
그림 11. 개체별 번식이력(한우) 데이터 일부분 .....	24
그림 12. 연도별 번식성적(한우).....	25
그림 13. 개체별 출하정보(한우).....	27
그림 14. 연도별 출하성적(한우).....	30
그림 15. 개체별 혈통축보(한우).....	32
그림 16. 개체별 혈통축보(한우) 데이터 셋의 데이터 일부분 .....	34
그림 17. 개체별 혈통성적(한우).....	35
그림 18. 개체별 EPD(한우).....	37
그림 19. ICT 데이터-체중(한우).....	40
그림 20. ICT 데이터-환경관리기(한우) .....	42
그림 21. ICT 데이터-사료빈(한우).....	44
그림 22. ICT 데이터-일반급이기(한우).....	46
그림 23. ICT 데이터-군사급이기(한우) .....	48
그림 24. ICT 데이터-음수관리기(한우) .....	50
그림 25. ICT 데이터-포유기(한우).....	52

그림 26. ICT 데이터-TMR 급이기(한우) .....	54
그림 27. ICT 데이터-발정탐지기(한우) .....	56
그림 28. ICT 데이터-환경제어기(한우) .....	58
그림 29. 양돈 데이터 셋 .....	60
그림 30. 개체별 번식로우(양돈) .....	61
그림 31. 연도별 번식통계(양돈) .....	63
그림 32. 월별 번식통계(양돈) .....	65
그림 33. 산차별 번식통계(양돈) .....	67
그림 34. ICT 데이터-급이기(양돈) .....	69
그림 35. ICT 데이터-음수기(양돈) .....	71
그림 36. ICT 데이터-환경정보(양돈) .....	73
그림 37. 포유모돈 급이번식 데이터(양돈) .....	75
<b>제 4 장. 스마트팜 데이터마트 데이터 연결성 문제점 분석</b> .....	<b>77</b>
그림 38. 한우 데이터의 집계 데이터의 상호 연결성 .....	78
그림 39. 양돈 데이터의 집계된 데이터의 상호 연결성 .....	79
그림 40. 한우 데이터 셋의 상호 연결성 .....	80
그림 41. 한우 데이터의 ICT 데이터 셋의 상호 연결성 .....	81
그림 42. 양돈 데이터의 로우 데이터 셋의 상호 연결성 .....	82
그림 43. 시나리오 1 - 번식 프로그램 최적화 .....	85
그림 44. 시나리오 2 - 사료 급이 시스템 최적화 .....	88
그림 45: 시나리오 3 - 축산 농장 환경 관리 .....	91
그림 46. ICAR 데이터 표준 준수비율 (한우 데이터) .....	95

# 1. 연구개요

## 1.1 연구배경

인공지능(AI)과 빅데이터는 4차 산업혁명을 견인하는 가장 중요한 기술로 자리잡고 있으며, 전 산업 분야에 걸쳐 혁신적인 변화를 가져오고 있다. 특히, 빅데이터의 분석과 인공지능의 응용은 다양한 산업 분야에서 맞닥뜨리고 있는 복잡한 문제들을 해결하기 위한 효율적인 수단으로 부상하고 있다. 이러한 맥락에서 본 연구는 축산업을 중심으로 현재 산업이 직면한 도전과제를 조명하고 인공지능 기반의 빅데이터 분석을 용이하게 할 수 있는 토대 마련에 기여하는 것을 목표로 한다.

최근 발간된 2023년 농림축산식품 통계연보의 통계자료에 따르면 농가 인구는 2016년 250만명에서 2022년 217만명으로 꾸준히 감소하는 추세에 있는 반면, 같은 기간 사육되는 가축의 수는 증가해왔다. 예를 들어, 소는 296만마리에서 373만마리로, 돼지는 1037만마리에서 1112만마리로 늘어났다 (그림 1 참조). 이는 농가 인구 한 사람이 관리해야 하는 가축 수가 지속적으로 증가하고 있음을 의미하며 축산업의 업무 과중 및 노동력 부족 문제가 심화되고 있음을 추론할 수 있다 (그림 2 참조). 이러한 점을 미루어 보았을 때, 현재 축산업에서 지속적으로 발생하고 있는 가축의 분뇨 처리, 악취 문제, 탄소 배출 증가, 그리고 가축 복지와 방역에 대한 이슈들은 앞서 제시한 축산업의 현상황에서 기인한 문제라고 볼 수 있다.

그림 1. 2016년~2022년 연도별 농가 인구 및 가축 수 추이

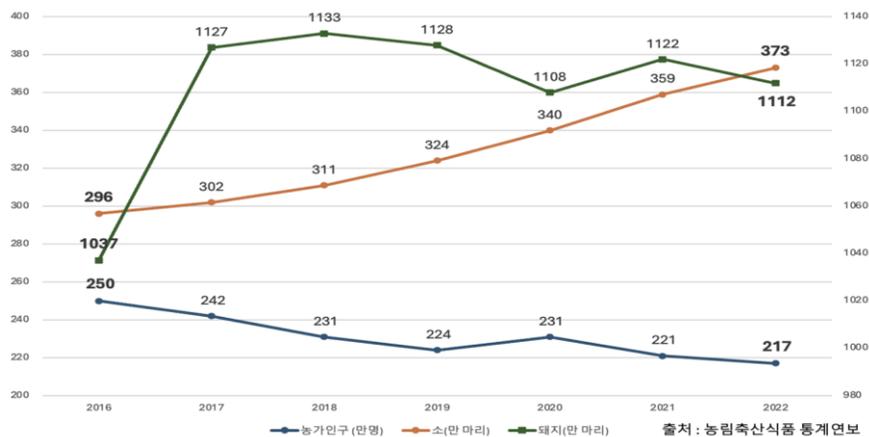
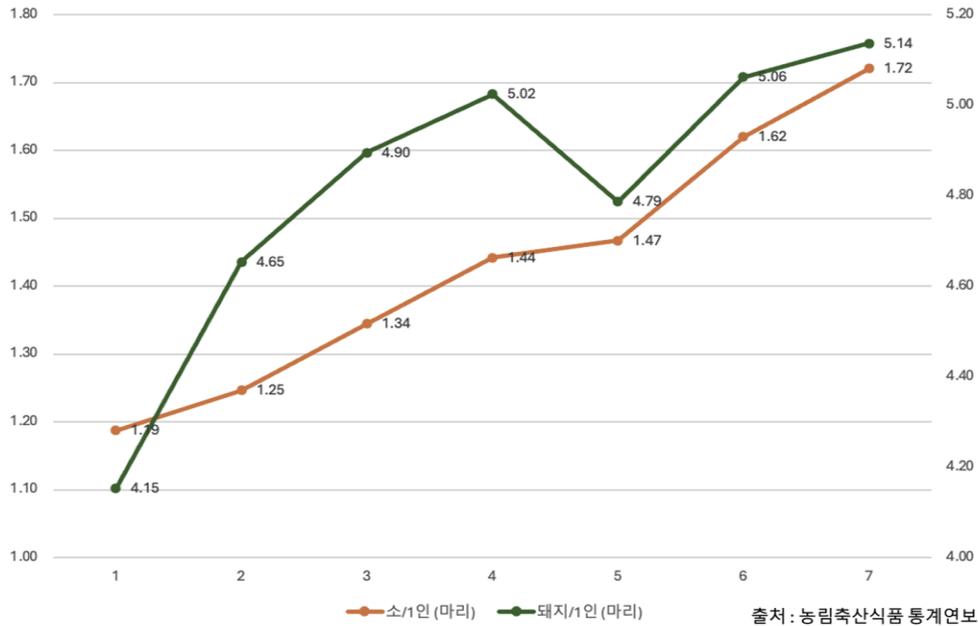


그림 2. 2016 년~2022 년 전국 1 인당 사육 가축 수 추이

(분석 대상: 전국, 대상 가축: 소(초록색), 돼지(갈색))



이에 따라 축산업의 노동력 절감 및 생산성 향상의 필요성이 증대되었고 이를 달성하기 위해 스마트 축산에 대한 관심이 증대되고 있다. 이러한 상황에 대응하기 위해 정부는 2014 년부터 스마트팜을 농업의 핵심 성장동력으로 보고, 스마트팜 보급 사업을 본격적으로 추진해왔다<sup>1</sup>. 사업의 초기에는 자동 급이 시스템, 온습도 측정 장비 등의 ICT 장비를 농가에 보급하여 축산 작업의 자동화를 통해 노동력을 절감하고 생산성을 향상시켰다. 2018년에는 임대형 스마트팜 조성으로 청년들의 스마트팜 창업을 지원하고 스마트팜 확산과 혁신의 거점으로 「스마트팜 혁신밸리」를 조성하고자 하였다<sup>2</sup>. 또한 2019 년부터 스마트 팜 도입을 추진하는 농가에 스마트 장비를 지원하여 스마트팜 경쟁력 향상과 함께 스마트팜 산업 생태계를 조성하기 위한 노력에 박차를 가했다<sup>3</sup>. 2020년에는 한국판 뉴딜 종합 계획의 일환으로 디지털 뉴딜 정책에 따라 인공지능 학습 데이터 구축 사업으로 이를 더 확대하여 추진하기 시작했다<sup>4</sup>. 이어서 2021년에는 1 차 산업의 디지털 전환의 밑거름을 마련하기 위한 데이터 구축 분야로 교통, 물류 분야와 함께 농축수산 분야를 선정하여 지정공모를 통해 추가 농축수산 데이터를

<sup>1</sup> “[기획]농업은 미래성장 산업이다(4)ICT(정보통신기술) 융복합”, 서훈, 농민신문, 2014.10.15

<sup>2</sup> “농업 혁신성장, 스마트팜이 선도한다”, 농림축산식품부, 2018.04.16

<sup>3</sup> “스마트한 축산으로/정부, 스마트팜 지원정책은 2020년까지 스마트팜 5천 750호 구축”, CSN 축산신문, 2019.10.18

<sup>4</sup> “정부, 뉴딜정책 유망분야 ‘스마트팜’에 3867억 투자한다”, 인더뉴스, 2020.12.10

구축하였다<sup>5</sup>. 그리고 2022년에는 농업의 미래 성장 산업화를 명시하고, 스마트 농업을 고도화 하는 동시에 관련 기술의 농가 보급을 촉진하고 스마트팜 혁신밸리를 조성하여 온실, 축산 중심으로 스마트 농업 보급 확산을 추진해왔다. 또한 빅데이터, 인공지능 기반 스마트농업 확산을 위해 인프라 구축과 거점육성 및 보급 확산, 기술개발, 인력양성, 기업 경쟁력 강화, 한국형 스마트팜 수출 활성화 등을 전략으로 포함시켰다<sup>6</sup>. 이와 같은 정부의 적극적인 지원 아래 스마트팜 데이터마트의 축산 분야에서는 한우, 낙농, 양돈, 양계에 대한 방대한 양의 데이터가 축적되었다. 이는 각각 약 890 만 건, 450 만 건, 320 만 건, 160 만 건에 달하는 상당한 양의 빅데이터에 해당한다 (그림 3 참조).

그림 3. 스마트팜 데이터마트 축산 정형데이터 건수 (2024년 3월 24일 기준)

축산 데이터 종류	데이터 양
 한우	8,919,421건
 낙농	4,520,618건
 양돈	3,185,359건
 양계	1,618,940건

이렇게 구축된 축산 빅데이터는 민간 또는 관련 산업체가 인공지능(AI) 분석이나 데이터 기반의 솔루션을 개발할 수 있는 기회를 제공하고 있다. 이는 데이터를 기반으로 한 새로운 가치 창출과 기술 혁신을 가능하게 하는 중요한 전환점으로 작용할 수 있다. 그럼에도 불구하고, 이러한 데이터의 풍부함이 실질적인 활용으로 이어지지 못하는 현상이 나타나고 있다<sup>7</sup>. 수집된 데이터가 방대하다 하더라도, 이를 효과적으로 분류하고 저장하는 과정에서의 표준화가

<sup>5</sup> “[창간특집 인터뷰] ③ 디지털 뉴딜 전도사 문용식 NIA 원장, ICT에 역량 집중”, 조선신문, 2021.04.28

<sup>6</sup> “스마트농업 고도화 전략 마련...육성·확산으로 미래농업 전환”, 데일리안, 2021.12.23

<sup>7</sup> 국내 ‘스마트축산’, “빅데이터와 ICT 융복합 접목 미흡”, 애플경제, 2021.10.04

이루어지지 않으면 데이터의 질과 가치는 크게 떨어질 수밖에 없다. 데이터의 수집, 분류, 저장에 관한 표준화 미비, 데이터설명서의 부재, 품질 관리 기준의 부재 등은 데이터의 활용성을 크게 저해하며, 이는 결국 데이터의 민간 부문에서의 활용도를 현저히 감소시킨다<sup>8</sup>. 이에 대한 명확한 해결책과 정책적 지원이 시급하며, 또한 빅데이터의 가치를 최대화하고 실질적으로 활용할 수 있도록 하는 데이터 표준화 연구가 절실하게 요구되는 상황이다.

## 1.2 연구의 필요성

최근에는 빅데이터와 인공지능이 맞물려 사회 곳곳의 문제를 해결하려는 연구와 시도가 활기를 띠고 있다. 정부의 데이터 개방 정책을 바탕으로 한 다양한 인공지능 빅데이터 분석 대회가 개최되는 것이 이러한 움직임을 상징적으로 보여주고 있다<sup>9</sup>. 이런 대회들은 사회의 복잡한 이슈들에 대한 혁신적인 솔루션을 찾기 위해 데이터의 힘을 빌리고 있다. 하지만, 인공지능을 이용한 데이터 분석의 성공은 고품질 데이터의 존재에 직결되어 있다. 데이터의 호환성과 상호운용성을 보장하는 고품질 데이터는 데이터 분석을 더욱 정확하고 효율적으로 만들어 준다. 이를 위해 데이터 표준화는 필수적이다. 이 과정은 데이터를 일관된 형식으로 유지시켜 분석 과정 중 발생할 수 있는 오류를 최소화하고, 결과의 신뢰성을 높이는 데 중요한 역할을 한다. 데이터 표준화는 관리 측면에서도 많은 이점을 가지고 있다. 일단 데이터가 일관된 형식으로 수집되면 이후의 다양한 태스크나 프로젝트에서 재활용이 가능해져 시간과 자원을 절약할 수 있다. 또한, 표준화된 방식으로 저장된 데이터는 사용자가 원하는 정보를 빠르고 쉽게 찾을 수 있게 함으로써 업무의 효율성을 크게 향상시킨다<sup>10</sup>.

이러한 맥락에서, 스마트팜 데이터마트의 데이터 표준화 연구는 매우 중요하다. 데이터 표준화와 인공지능을 활용한 빅데이터 분석은 축산 분야에서 혁신적인 발전을 가져올 수 있는 잠재력을 가지고 있기 때문이다. 축산 분야에서 고품질의 표준화된 데이터를 활용한다면 다음과 같은 긍정적인 효과를 기대할 수 있다. 우선, **정밀 축산이 가능하다**. 이는 각 개체의

---

<sup>8</sup> 김송강, 남기포, 「스마트팜 구축을 위한 데이터수집의 문제점 분석」, 한국융합보안학회, 제 22 권 제 5 호, 2022, p 69-80

<sup>9</sup> "2024 공모전 트렌드 예측 10, 인공지능, 솫폼, 콘텐츠 뜬다!", Thinkgood(<https://www.thinkcontest.com/>)

<sup>10</sup> "데이터 표준화의 필요성" 데이터온에어 (<https://dataonair.or.kr/>) 2021.02.10)

건강 상태, 성장 패턴, 영양 상태 등을 실시간으로 모니터링하고 분석하여 최적의 사육 조건을 제공할 수 있도록 한다. 인공지능은 다양한 데이터 소스에서 수집된 정보를 종합하여 각 동물에 맞는 맞춤형 사료 급여 계획을 수립하거나 질병을 조기에 예측하고 예방할 수 있도록 도와줄 수 있다. 또한, **사육 환경의 지속 가능성을 향상시킬 수 있다.** 자원의 효율적 사용을 위한 빅데이터 분석을 통해 물과 사료의 낭비를 줄이고, 축산물의 생산성을 극대화하는 동시에 환경 영향을 최소화할 수 있다. 게다가 표준화된 데이터와 인공지능 분석을 통해 **생산 효율성이 높아지고, 이는 축산농가의 비용 절감 및 소득 증대로 이어질 수 있다.** 이처럼 장기간 수집되어 온 축산 빅데이터를 표준화시키는 연구를 통해 현재 문제가 되고 있는 민간의 데이터 활용이 저조한 상황을 해결할 수 있고 더 나아가 축산업의 노동력 부족, 가축의 분노 처리, 악취 문제, 탄소 배출 문제, 가축 복지 및 방역에 대한 문제 및 축산 농가의 소득 문제 등을 해소하는데 큰 도움이 될 것으로 보인다

### 1.3 연구의 목적

데이터 표준화는 다양한 데이터 셋의 연결성을 보장함으로써 인공지능을 활용한 데이터 분석 과정에서 데이터가 가진 가치를 최대한으로 이끌어 낼 수 있도록 하는데 중추적 역할을 수행한다. 조사한 바에 따르면, 스마트팜 데이터 마트는 현재 축산업과 관련하여 방대한 데이터를 축적하고 있음에도 해당 데이터의 표준화가 상대적으로 미비하여 데이터의 활용이 원활히 이뤄지지 않고 있다. 본 연구는 이러한 문제를 해결하기 위해 데이터 표준화의 중요성을 탐구하고, 축산 데이터의 수집 및 저장에 관한 국제 표준을 제시하고 있는 ICAR(International Committee for Animal Recording)<sup>11</sup>의 가이드 라인을 바탕으로 현재 스마트팜 데이터 마트의 데이터 셋을 상세하게 비교하여 분석한다. 이 과정을 통해 현재 스마트팜 데이터마트의 데이터 셋이 국제 표준을 준수하고 있는지 확인하고 만약 준수하지 않는다면 어떤 데이터가 누락되어 있는지 구체적으로 분석하여 향후 추가 및 보완해야 하는 데이터를 제시한다. 여기서 더 나아가 스마트팜 데이터마트의 다양한 데이터 셋을 통합 및 연결하여 축산업 종사자 및 데이터 분석가들이 분석하고자 하는 목적에 적합한 시나리오들을 제시한다. 해당 시나리오에서

---

<sup>11</sup> <https://www.icar.org/>

제시한대로 데이터 분석이 원활히 이뤄질 수 있도록 하기 위해 현재 스마트팜 데이터마트의 데이터 셋의 미비점을 찾아내고 수정 및 보완할 점을 제안한다. 궁극적으로 해당 데이터의 정확도와 신뢰성을 개선하여 데이터 품질의 향상을 도모하고 연구자와 실무자들로 하여금 스마트팜 데이터마트의 데이터 셋을 보다 활발히 활용할 수 있게 하여 축산업에서 발생하는 복잡한 문제들을 효율적으로 대처하고 해결할 수 있도록 하는데 기여한다. 결론적으로 이 연구는 스마트팜 데이터 마트의 데이터 표준화 관련 문제점을 진단하고, 그 나아갈 바를 제시함으로써, 양질의 축산 데이터를 축적할 수 있도록 도와 스마트팜 데이터의 잠재력을 끌어올리고, 축산업의 혁신을 촉진하는 데 목적이 있다.

## 2 연구방법

### 2.1 분석 방법론

#### 2.1.1 비교 사례 연구 (Comparative Case Study)

비교 사례 연구는 두 개 이상의 개별 사례를 상세히 조사하고 분석하여, 그 사이의 유사성과 차이점을 이해하고 해석하기 위한 연구 방법론이다. 이 연구 방법론은 연구 문제에 대한 다면적인 이해를 가능하게 하여 각 연구 대상에 대한 이해도를 높이고 특정 대상의 문제를 다면적으로 파악하는데 적절하다. 이 연구에서는 국내 축산 빅데이터 플랫폼인 스마트팜 데이터마트와 농장 동물 데이터의 측정 및 기록에 대한 표준을 제정하는 ICAR을 연구 대상으로 하여 국내 데이터 플랫폼과 국제 표준 사이의 괴리를 해소하기 위한 연구 내용을 제시한다. 비교 연구 대상으로 ICAR을 선택한 것에는 구체적으로 다음과 같은 이유가 있다.

1. **국제적 표준 제공:** ICAR은 전세계 가축의 특성을 측정하는 국제 표준을 수립하는 데 있어 중추적인 역할을 하고 있다. 이러한 표준은 이 후 다른 축산 데이터 플랫폼과의 비교 분석 시 기준으로 활용될 수 있다.
2. **광범위한 영향력:** ICAR은 55개국 이상의 회원으로 구성되어 있으며, 이는 국제적으로 인정받는 기구의 권위와 네트워크를 반영한다. 다양한 국가와 환경에서의 데이터 측정 및 기록 방법에 대한 비교 분석에 있어서 타당한 비교 대상이 된다.
3. **다양한 동물 종에 대한 데이터:** ICAR은 육우, 젖소, 알파카, 염소, 양 등 다양한 동물 종에 대한 데이터 기록 방법을 표준화하고 있다. 이는 특정 동물 종에 국한되지 않고 넓은 범위의 비교 분석을 가능하게 한다.
4. **데이터의 정확성과 상호 운용성 보장:** ICAR의 표준화 가이드라인은 데이터의 정확성을 보장하며, 데이터 간의 상호 운용성을 증진시킨다.

## 2.1.2 시나리오 분석 (Scenario analysis)

시나리오 분석 방법은 복잡한 문제 상황을 해결하기 위한 효과적인 접근법 중 하나로 특정 대상이나 상황에 대해 가능한 다양한 시나리오를 제시하고 각 시나리오에서 발생할 수 있는 문제를 식별하여 문제를 해결하기 위한 솔루션을 모색하는 과정을 포함한다. 이와 같은 분석 과정을 스마트팜 데이터마트에서 제공하고 있는 축산 데이터 셋에 대해 적용시키는 경우 축산업에 종사하는 연구자들과 관련 업계 전문가들이 직면한 다양한 문제들을 해결하는 데 중요한 역할을 할 수 있다. 시나리오 분석에 앞서 관련 연구 사례 분석을 통해 이들이 관심을 가지고 있는 문제들을 정의하고, 이 문제들을 해결하기 위한 시나리오를 개발함으로써 스마트팜 데이터마트에서 제공하는 데이터 셋이 어떻게 활용될 수 있는지 구체적으로 탐색한다. 예를 들어, 가축 번식 프로그램 최적화, 동물 건강 관리, 사료 효율 최적화 등 다양한 도전 과제를 정의하고 이러한 문제들에 대해 스마트팜 데이터마트의 데이터 셋을 활용하여 해결하기 위한 구체적인 솔루션을 하나의 시나리오로 구성하여 분석한다. 이와 같은 과정을 통해 현재 스마트팜 데이터마트가 가지고 있는 데이터 셋의 한계점을 파악하고 문제 해결을 위해 데이터 셋이 어떻게 개선되어야 하는지에 대한 중요한 통찰을 얻을 수 있을 것이다.

## 2.2 분석 대상

### 2.2.1 ICAR (International Committee for Animal Recording)



ICAR 은 가축의 식별, 데이터 측정 및 기록 그리고 유전 평가의 발전을 목적으로 하는 세계적인 비정부 기구이다. 1951 년 유럽의 12 명의 전문가들에 의해 창립된 이후, 전세계적으로 가축의 특성을 측정하는 국제 표준을 수립하는 데 중추적인 역할을 해왔다. 이 기구는 55 개 이상의 국가에서 온 130 명의 회원으로 이루어져 있으며, 다양한 동물에 대한 데이터 측정 및 기록을 위한 방법론을 표준화하고 있다. ICAR 은 육우, 젖소, 알파카, 염소, 양과 같은 다양한 동물 종에 대한 데이터 기록 방법을 표준화하여 전 세계 동물 사육자들로 하여금 신뢰할 수 있는 정보에 기반하여 결정을 내릴 수 있도록 지원한다. 2022 년 6 월자로 최신화된 ICAR 의 데이터 측정 및 수집 표준화 가이드라인은 가축 데이터의 정확성을 보장하고, 데이터 간의 상호 운용성을 증진시키며 국제적으로 일관된 정보 교류를 가능하게 만든다. 이는 축산업 기술의 발전, 농장 경영의 효율성 개선, 그리고 동물 유전 자원의 관리와 보전에 기여하고 있다. ICAR 의 웹사이트에 공개된 자료는 전 세계적으로 가축에 대한 데이터를 표준화하는 데 있어 가장 최신의 리소스를 제공하고 있으며 회원국들에게 지속적인 교육과 지원을 제공하여 농장 동물 데이터 기록 및 평가에 대한 전문성을 강화하고 있다. 이러한 국제 표준을 통해 가축의 관리가 더욱 체계적이고 정밀하게 이루어질 수 있도록 지원한다. 현재 ICAR 웹사이트에 나열된 구체적인 지침은 다음과 같다.

- 일반 규칙: 모든 동물에 대한 고유한 식별을 의무화한다. 이는 태그나 전자식 장치 등 다양한 식별 수단이 포함될 수 있다. 수입된 동물의 경우, 해당 동물이 새로운 국가의 규정을 준수하는 한 기존의 식별정보를 그대로 사용할 수 있다. 또한 인공수정 절차에 대한 기록 방법에 대해 세부 지침을 명시하고 있으며 인공수정에 사용된 소의 식별과 번식주기 내 인공수정 시도 순서를 기록할 것을 명시한다. 이러한 규칙은 동물 데이터의

수집 및 교환 과정에서 일관성과 정확성을 보장하며 효과적인 유전적 평가를 위해 필수적이다.

- 낙농 기록: 우유 생산 특성에 대한 정보는 낙농 무리의 관리와 번식에 매우 중요하다. 우유 기록 과정은 해당 젖소의 식별, 분만 날짜, 우유의 양과 제공된 시간 등의 데이터 수집으로 시작하며 이 때, 우유 샘플을 채취할 수 있다. 얻어진 우유 샘플은 우유 성분 에 대해 분석되고 분석 결과와 우유 생산량 및 착유 시간에 대한 데이터는 데이터베이스에 저장된다. 그 후, 다양한 매개변수, 누적 생산량 및 지표가 계산되어 추가적으로 데이터베이스에 저장되며 마지막으로 농부에게 보고된다.
- 육우 기록:
  - 육우 생산자와 사육자가 국가 간에 효율적으로 소통할 수 있도록 하는 육우 기록 체계에 대한 공통된 이해를 제공하고 육우 기록에 대한 글로벌 표준을 제시한다.
  - 새로운 국가의 육우 기록 체계의 설립을 위한 조언과 도움을 제공한다.
  - 육우 특성의 유전적 평가를 위한 견고한 데이터 인터페이스를 제공한다.
  - 적절한 데이터 구조를 구현함으로써 유전 증거의 신뢰성 향상을 도모한다.
  - 중요한 비유전적 효과의 식별 및 기록을 통한 유전 증거의 정확성 향상을 도모한다.
  - 효율적인 국가 및 국제 데이터 교환이 가능하도록 육우를 위한 국제 데이터 사전 을 설립한다.
  - 유전 평가 프로그램에 관련된 기록 및 번식 기구에 대해 지원한다.
  - 신뢰할 수 있는 실천 규약을 제시한다.
- DNA 기술: 상업적 가축 개체군에서 유전 정보를 성공적이고 유익하게 찾기 위한 기술을 제시한다.
- 동물 표준 특성: 낙농, 육우 등에 대한 특성 정의 및 데이터 수집 및 모니터링 분류자의 투명성과 개선에 대한 권장사항을 제시한다.

- **인공수정 및 배아 이전:** 소 정액 스트로우(Straw)에 인쇄해야 할 최소 정보와 추가 정보를 인쇄할 경우 사용자를 돕기 위한 일반 권장사항을 제공한다.
- **동물 건강 특성:** 이 섹션은 동물 복지, 식품 안전, 질병 관리에 대한 우려가 커지고 있는 것에 대응하여, 전통적인 생산 특성과 함께 건강 정보를 기록하고 모니터링하는 것의 중요성을 강조한다. 특정 기록 방법은 개별 기록 기구에 맡겨져 있지만, 지침에서는 발굽 건강, 절름발이 등 기록되어야 할 필수적인 측면을 개관한다. 이 섹션은 또한 이러한 건강 데이터에 기반한 국가 모니터링 프로그램의 가치를 강조하여, 당국, 소비자, 생산자 모두에게 이점을 제공한다.
- **품질 인증:** ICAR 품질 인증의 목적은 ICAR 회원 기구가 ICAR의 지침이 있는 활동 분야에서 높은 서비스 품질을 유지하고, ICAR의 작업 범위 내의 다른 분야에서도 서비스 품질의 지속적인 향상을 보장하는 것을 지원하기 위한 것이다.
- **유전 평가:** 국가 및 국제 수준에서 낙농 소의 유전 및 유전체 평가와 관련하여 현재 시행되고 있는 관행에 대한 일반적인 개요를 제공한다. 개요는 현재 국가 유전 평가 센터(NGEC)에서 제공하는 유전/유전체 평가 시스템(GES)을 바탕으로 한다.
- **식별 장치 인증:** ICAR에 의해 개발된 동물 식별 장치의 테스트 및 인증을 위한 원칙과 절차에 대한 일반적인 소개이다.
- **측정, 기록 및 샘플링 장치의 테스트, 승인 및 검사:** ICAR에서 개발한 측정, 기록 및 샘플링 장치와 센서 시스템의 테스트 및 인증을 위한 원칙과 절차를 소개한다.
- **우유 분석:** 개별 젖소의 우유에서 지방, 단백질, 유당, 요소 및 체세포 수 결정을 위한 방법들을 제시한다.

- 데이터 교환:
  - 데이터 처리 시스템 개발자들에게 ICAR 회원들이 사용하는 전 범위의 동물 데이터 정보 시스템과 쉽고 신뢰성 있게 연결될 수 있도록 보장하는 표준 및 프로세스를 제공한다.
  - 서로 다른 정보 시스템 간 정보 교환을 가능하게 하기 위해 교환되는 데이터의 정의를 제공한다.
  - 가축에 대해 전 세계적으로 합의된 데이터 사전을 개발한다.
  - 정보 시스템과 농장 장비 간 데이터 교환을 지원하는 표준화된 시스템의 개발과 설치를 지원한다.
  
- 품종 협회: 품종 협회 개발을 위한 표준 프레임워크를 제공하고, 필요한 절차와 프로토콜을 정의한다.
  
- 급여: 유전체 및 유전 평가를 위한 사료 섭취 정보의 기록 및 처리에 관한 지침을 제공한다.
  
- 메탄 배출: 현재 지침의 첫 번째 목표는 개별 동물의 대규모 메탄 배출 측정 방법의 적합성을 검토하고 비교하는 것으로 방법들 간의 정확성, 정밀성 및 상관관계 평가가 포함된다. 또한 다른 나라와 데이터 셋을 결합하는 것은 측정하기 어려운 특성을 찾아내는데 성공적인 전략이 될 수 있다 (de Haas et al. 2017). 다른 출처에서 온 데이터는 결합될 때 적절히 가중치를 두거나 조정되어야 하므로, 다른 방법들에 의해 얻어진 결과들 간의 상관관계를 검토하고, 궁극적으로 배출량이 높거나 낮은 개별 동물을 선택하기 위한 신뢰 한계 추정으로 이어진다 (Garnsworthy et al. 2019).

## 2.2.2 스마트팜 데이터마트

조사한 바에 따르면, 스마트팜 데이터마트와 비견될 수 있는 국내 농업 및 축산 분야에 관한 데이터 플랫폼은 한국지능정보사회진흥원의 "AI Hub", 한국농수산식품유통공사의 "KADX 농식품빅데이터거래소", 스마트팜연구개발사업단의 "스마트팜 R&D 빅데이터 플랫폼"이 있다. 먼저 스마트팜 데이터마트에 대한 소개에 앞서 각 데이터 플랫폼에 대해 간략히 설명하고자 한다.

한국지능정보사회진흥원의 AI Hub(aihub.or.kr)는 농축수산 뿐만 아니라 헬스케어, 교통물류, 스포츠 등 다양한 분야의 데이터를 제공하고 있다. 현재 농축수산과 관련된 데이터는 총 71 건이고 이 중 11 건의 축산 데이터가 일반인에게 제공되고 있다 (그림 4 참조). 현재 제공되고 있는 축산 데이터의 종류는 아래 표 1 과 같고, 그림 5 에 보여지는 축산 데이터의 유형은 모두 비정형 데이터만 제공되고 있다. 그림 5 는 데이터 셋 세부 구성 요소를 대상으로 단순 계수에 기반한 상대적인 비율을 나타낸다.

그림 4. AI Hub 홈페이지 (aihub.or.kr)

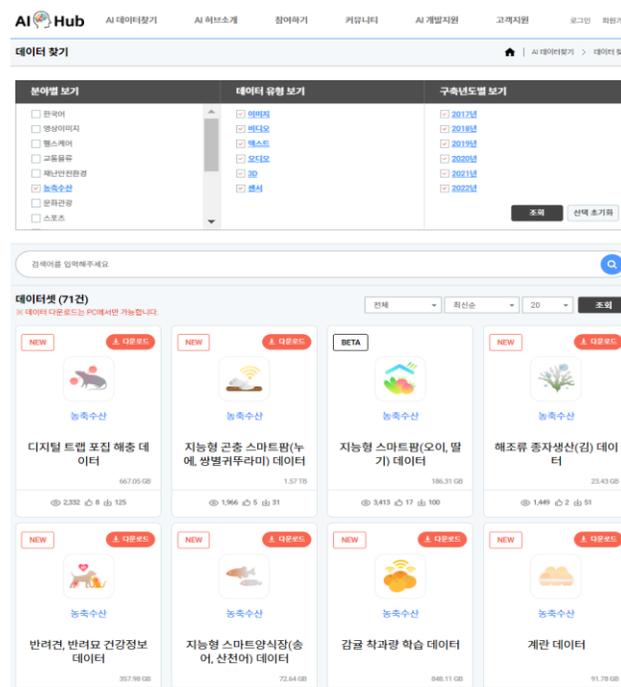
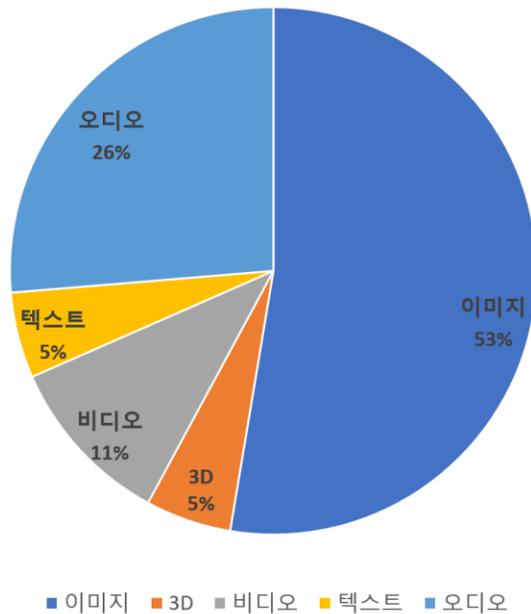


표 1. AI Hub 축산데이터의 등재된 데이터 종류 및 유형

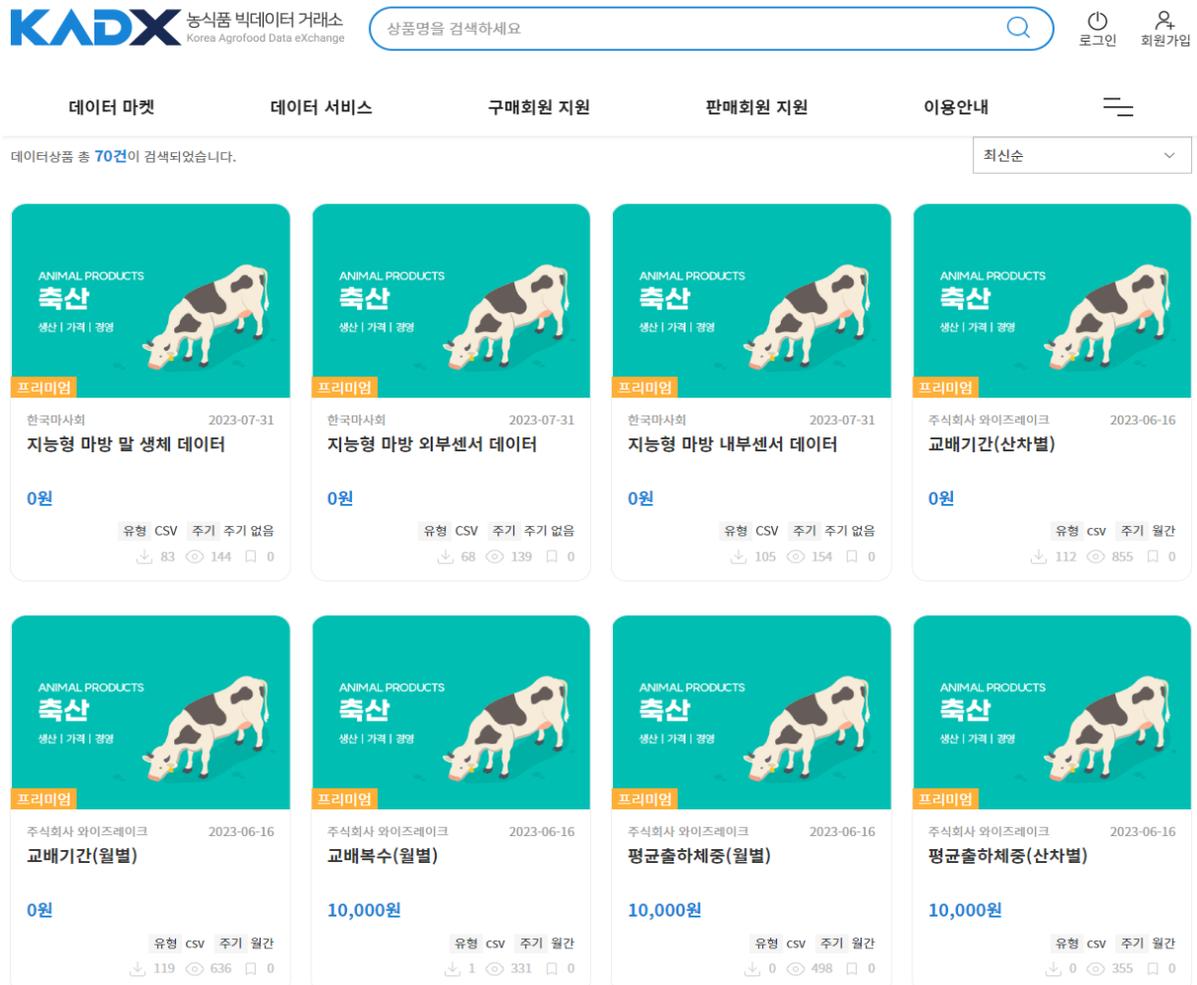
축산 데이터 종류	데이터 유형
계란데이터	이미지
소 및 돼지 발정 행동 데이터	3D, 이미지, 비디오, 텍스트, 오디오
양돈 생체 에너지 데이터	영상, 이미지
축산 기자재(소, 돼지) 3D 데이터	이미지
지능형 스마트축사 통합데이터(양돈)	이미지, 오디오
지능형 스마트축사 통합데이터(한우)	이미지, 오디오
지능형 스마트 축사(젖소) 데이터	이미지, 오디오
지능형 스마트 축사 데이터(육계, 산란계)	이미지, 오디오
한우 신체충실지수 등급 데이터	이미지
가축 행동 영상	이미지
축산물 품질(QC) 이미지	이미지

그림 5. AI Hub 축산데이터 등재된 11 건의 데이터 유형 분포



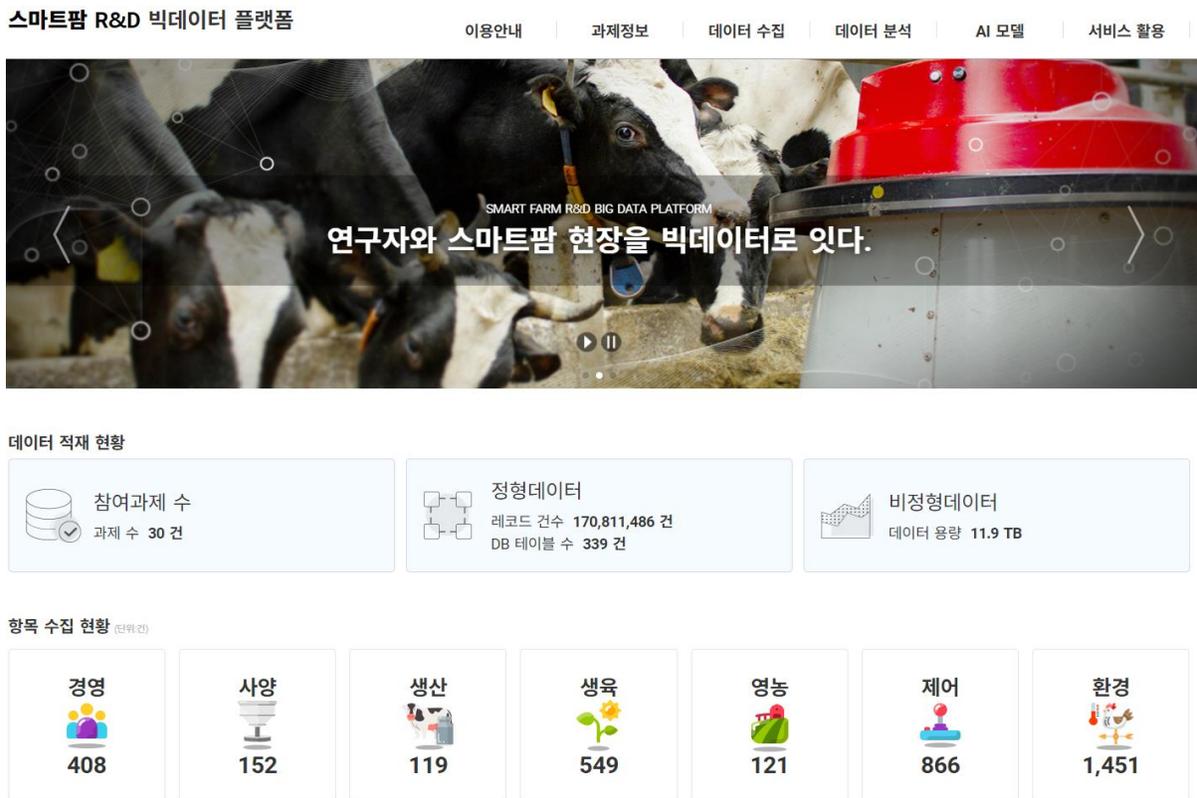
한국농수산물유통공사 KADX 농식품빅데이터거래소(kadx.co.kr)는 농식품 데이터를 온라인으로 판매하거나 구매할 수 있는 플랫폼으로 기존의 타 플랫폼과 달리 거래소 역할이 포함된 유료 플랫폼이다. 농식품관련 데이터에 대해 판매회원과 구매회원으로 구분되어 있으며 판매자와 구매자는 서로의 필요에 맞게 해당 홈페이지를 이용할 수 있다. 현재 판매자의 데이터를 기반으로 제공되고 있는 데이터 서비스는 출하반장, 레시피 추천서비스, 농산물 종합 분석, 양돈농가 자가진단 서비스가 있다. 현재 축산과 관련해서 구매할 수 있는 데이터는 총 70 건으로 모두 정형데이터만 제공되고 있다 (그림 6 참조).

그림 6. KADX 홈페이지 (kadx.co.kr)



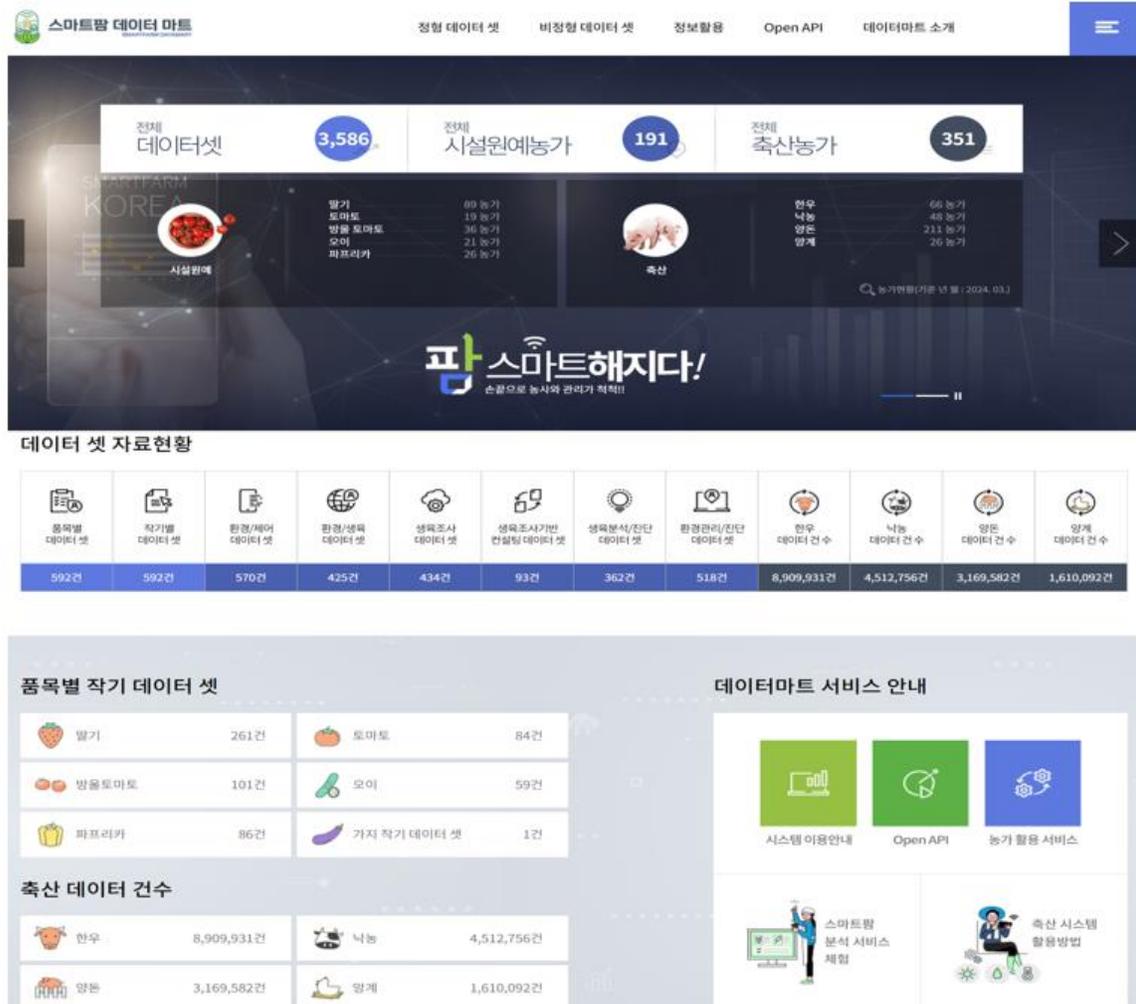
스마트팜연구개발사업단의 스마트팜 R&D 빅데이터 플랫폼(rnd.bigdata-smartfarm.kr)은 스마트팜 다부처 패키지 혁신기술개발사업에서 생산된 결과물(AI 모델, 플랫폼, 소프트웨어, 하드웨어, 기타)의 정보를 제공하는 플랫폼이다. 참여하는 과제의 수가 아직까지는 제한적이며, 데이터 적재현황은 아래 그림 7에 보여지고 있다. 데이터가 현재도 업로드 되어 구축되는 중에 있다.

그림 7. 스마트팜 R&D 빅데이터 플랫폼 홈페이지 (rnd.bigdata-smartfarm.kr)



스마트팜 데이터마트(data.smartfarmkorea.net)는 대한민국 농업 및 축산 분야의 혁신을 촉진하는 중추적인 플랫폼으로 전국의 농장들로부터 수집한 방대한 양의 축산 및 농업 데이터를 관리하고 공유하기 위한 목적으로 설립되었다 (그림 8 참조).

그림 8. 스마트팜 데이터마트 홈페이지 (data.smartfarmkorea.net)



현재 축산과 관련해서는 정형데이터 와 비정형 데이터를 제공하고 있으나 정형데이터에 상대적으로 더 많은 초점이 맞춰져 있는 관계로 본 연구의 대상은 정형데이터를 중점적으로 분석한다. 이 플랫폼은 농림수산물교육 문화정보원이 스마트팜 데이터 및 웹페이지를 관리하고 운영하고 있으며 이 중 스마트 축산 데이터에 대해서는 별도로 축산물 품질평가원이 관리 및 운영하고 있다. 스마트 축산 데이터는 한우, 젓소, 양돈, 닭에 관한 포괄적인 데이터 베이스를 구축하고 있으며 사용자가 쉽게 접근하여 활용할 수 있도록 테이블 형태로 제공되는 정형데이터와 이미지, 영상, 음성 등의 형태로 제공되는 비정형 데이터로 구성되어 있다.

공개된 데이터는 개인정보 보호를 위해 일부 비식별화 처리가 되어 있으며, 누구나 인증없이 접근할 수 있다. 다만 OPEN API 를 활용하여 데이터를 수신 받으려는 경우 관리자의 승인 아래 제한적으로 공유 받을 수 있다. 스마트팜 데이터마트는 이러한 농장의 빅데이터를 민간, 산업체 등에 제공하여 축산 산업에 있어서 기술의 발전과 데이터 기반 의사결정을 지원하여 축산업의 지속 가능한 발전에 기여하고자 하는 목적으로 운영되고 있다. 다음 표 2 는 2024 년 3 월 24 일까지 스마트팜 데이터마트에 의해 수집된 데이터에 대한 기초 통계량을 보여준다.

표 2. 스마트팜 데이터마트 축산 데이터 기초 통계량 (2024 년 3 월 24 일 기준)

품종	데이터 총 건수	데이터셋 개수	총 동물 수	총 농장 수	총 센서 수
한우	8,919,421	17	32,150	136	10
낙농	4,520,618	20	6,282	94	15
양돈	3,185,359	8	45,095	87	4
닭	1,618,940	9	-	57	4

표 2 에 따르면 한우 데이터는 약 890 만 건의 데이터로 스마트팜 데이터마트가 보유하고 있는 데이터 셋 중 가장 많은 데이터를 가지고 있다. 이러한 풍부한 데이터는 보다 신뢰할 수 있는 연구 결과와 정책적 제언을 이끌어낼 수 있는 정보를 많이 포함할 수 있으므로 ICAR 에서 제시하는 데이터 수집, 측정 표준과의 비교연구 대상에 포함시킨다. 더 나아가 소와는 다른 특성을 가지고 있는 양돈 데이터를 연구 대상에 포함하여 더 일반적인 결과를 이끌어 낼 수 있도록 한다. 이 두 농장 동물의 데이터 셋에 초점을 맞추어 다음의 표 3 과 같이 데이터 셋에 대한 연구 범위를 제한한다.

표 3. 스마트팜 축산 데이터 셋 리스트

품종	번호	데이터 구분	데이터 타입 (Raw Vs. Aggregated)
한우	1	번식이력	Raw
	2	번식성적	Aggregated
	3	출하정보	Raw
	4	출하성적	Aggregated
	5	혈통족보	Raw
	6	개채별 혈통성적(한우)	Aggregated
	7	EPD	Raw
	8	ICT 데이터 - 체중	Raw
	9	ICT 데이터 - 환경관리기	Raw
	10	ICT 데이터 - 사료빈	Raw
	11	ICT 데이터 - 일반급이기	Raw
	12	ICT 데이터 - 군사급이기	Raw
	13	ICT 데이터 - 음수관리기	Raw
	14	ICT 데이터 - 포유기	Raw
	15	ICT 데이터 - 체중 TMR 급이기	Raw
	16	ICT 데이터 - 발정탐지기	Raw
	17	ICT 데이터 - 환경제어기	Raw
양돈	1	번식로우데이터	Raw
	2	번식통계데이터 - 년도별	Aggregated
	3	번식통계데이터 - 월별	Aggregated
	4	번식통계데이터 - 산차별	Aggregated
	5	ICT 데이터 - 급이기	Raw
	6	ICT 데이터 - 음수기	Raw
	7	ICT 데이터 - 환경정보	Raw
	8	포유모돈 급이 번식 데이터 - 월별	Aggregated

### 3. 스마트팜 데이터마트 데이터 문제점 분석

#### 3.1 분석방법

현재 스마트팜 데이터마트에서 수집 및 저장하고 있는 데이터의 종류와 ICAR의 축산과 관련된 데이터의 저장 및 측정 표준이 어떻게 다른지를 비교한다. 구체적으로 스마트팜 데이터마트의 데이터 셋과 관련한 ICAR 표준을 찾고 해당 표준에서 제시하는 데이터 수집 항목을 스마트팜 데이터마트의 데이터 셋이 갖추고 있는지를 파악한다. ICAR 표준에서 제시하는 데이터 수집 항목은 다른 데이터 셋 간의 연결성 및 분석 가능성을 고려한 최소한의 수집항목들을 의미하므로 이것이 누락되어 있는 경우 데이터의 활용도가 저하될 수 있다. 따라서 어떤 데이터 항목이 스마트팜 데이터마트에서 누락되어 있는지 일목요연하게 파악할 수 있도록 표를 통해 나열하고 누락 데이터가 1개 이상 존재하는 경우 해당 데이터 셋이 ICAR 표준을 준수하고 있지 않음으로 정의하여 표기한다. ICAR에서 다루고 있지 않은 데이터 셋의 수집 항목에 대해서는 구체적인 비교대상이 없기에 따라 ICAR 준수 여부를 판단하지 않도록 한다. 이와같은 데이터 비교 분석 내용은 스마트팜 데이터마트의 각 데이터 셋마다 정리된 표를 통해 나타내며 한우 데이터 셋과 양돈 데이터 셋에 각각 적용시킨 결과를 제시한다.

### 3.2 한우 데이터

스마트팜 데이터마트에서 제공하고 있는 한우 데이터 셋은 번식이력, 연도별번식성적, 출하정보, 출하성적, 혈통족보, 혈통족보성적, EPD, ICT-체중, ICT-환경관리기, ICT-사료빈, ICT-일반급이기, ICT-군사급이기, ICT-음수관리기, ICT-포유기, ICT-TMR 급이기, ICT-발정탐지기, ICT-환경제어기 총 17 개의 데이터 셋을 제공하고 있다. 이는 그림 9 의 한우 데이터 셋 단계별 선택과정 중 데이터 구분을 선택하는 단계에서 확인할 수 있다. 각각의 데이터 셋에 대해 ICAR 표준과 비교하여 분석한 내용은 다음과 같다.

그림 9. 한우 데이터 셋 단계별 선택

#### 한우데이터셋



데이터 셋 검색

STEP 단계별로 선택하세요.

Q

STEP 1 지역 선택	STEP 2 규모 선택	STEP 3 데이터 구분	STEP 4 상세구분
<input checked="" type="checkbox"/> 전국 <input type="checkbox"/> 강원특별자치도 <input type="checkbox"/> 경기도 <input type="checkbox"/> 경상남도 <input type="checkbox"/> 경상북도 <input type="checkbox"/> 광주광역시 <input type="checkbox"/> 대구광역시 <input type="checkbox"/> 대전광역시	<input checked="" type="checkbox"/> 전체 <input type="checkbox"/> 100두 미만 <input type="checkbox"/> 100~200 <input type="checkbox"/> 200~300 <input type="checkbox"/> 300~400 <input type="checkbox"/> 400~500 <input type="checkbox"/> 500~1000 <input type="checkbox"/> 1000두 이상	<input checked="" type="checkbox"/> 번식이력 <input type="checkbox"/> 번식성적 <input type="checkbox"/> 출하정보 <input type="checkbox"/> 출하성적 <input type="checkbox"/> 혈통족보 <input type="checkbox"/> 혈통족보 성적 <input type="checkbox"/> EPD <input type="checkbox"/> ICT데이터	<input checked="" type="checkbox"/> 개체별

## 1) 개체별 번식이력(한우)

개체별 번식이력(한우) 데이터 셋은 한우의 번식 이력, 성과, 출생과 관련된 세부 정보를 추적하기 위해 설계된 포괄적인 데이터 기록이다. 한우 데이터 셋의 단계별 선택 과정을 통해 제공받을 수 있는 개체별 번식이력(한우) 데이터 셋의 구조는 그림 10 과 같다.

그림 10. 개체별 번식이력(한우) 데이터 셋

### 개체별 번식이력(한우)

한우 개체별 번식이력, 번식성적, 출생 송아지 정보를 확인할 수 있는 데이터셋

총 86,272 건이 검색되었습니다.

다운로드

번식별 데이터				번식 이력						번식		
지역	규모	농장	개체번호	산차	수정차수	수정일	수정유형	KPN	분만일	임신기간(일)	총부회수	공
3000000000	200~300	0021163	20120517010287	1	1	2015-04-09	인공수정	KPN980	20160119	285	1	
3000000000	200~300	0021163	20120625010286	1	1	2015-03-27	인공수정	KPN980	20160106	285	1	
3000000000	200~300	0021163	20130411010285	2	1	2015-08-22	인공수정	KPN855	20160602	285	1	
3000000000	200~300	0021163	20130411010285	1	1	2014-05-24	인공수정	KPN774	20150305	285	1	
3000000000	200~300	0021163	20130517010284	1	1	2015-10-14	인공수정	KPN960	20160725	285	1	
3000000000	200~300	0021163	20170905010288	3	1	2021-06-11	인공수정	KPN1274			1	
3000000000	200~300	0021163	20170905010288	2	1	2020-01-18	인공수정	KPN1278	20201110	297	1	
3000000000	200~300	0021163	20170905010288	1	1	2018-11-12	인공수정	KPN1187	20190824	285	1	
3000000000	200~300	0021163	20180125010127	2	2	2021-05-18	인공수정	KPN1283			2	
3000000000	200~300	0021163	20180125010127	2	1	2020-03-26	인공수정	KPN1016			1	

1 2 3 4 5 6 7 8 9 10 > >>

개체별 번식이력(한우)과 관련된 ICAR 표준은 번식 상태 데이터 기록 표준 [ICAR-03-2.3.4]이다. 해당 가이드 라인에 따르면, 번식 이벤트에는 교배, 인공수정, 배아이전, 암컷의 출산 및 분만, 수컷의 거세와 같은 이벤트를 포함한다. 또한 교배 기간 동안 암소가 하나 또는 여러 마리의 수소와 같이 있을 경우 교배 가능한 모든 수소를 해당 교배 시간에 기록해야 한다고 정한다. [ICAR-03-2.3.4]에서 제안하는 수집 데이터는 아래와 같으며 이를 표 4 를 통해 스마트팜 데이터마트와 비교한다.

- a. 동물 ID
- b. 날짜
- c. 기록자 ID
- d. 실제 위치: 농장 ID (해당되는 경우 농장 내 관리 그룹)

e. 번식 이벤트를 설명하는 코드

f. 관련된 다른 동물(들)의 ID (예: 교배 파트너, 송아지, 양육 송아지 등; 해당되는 경우)

표 4. ICAR 표준과 스마트팜 데이터마트 비교 표 - 개체별 번식이력(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-03-2.3.4] <sup>12</sup> 동물 ID 날짜 기록자 ID 농장 ID 생식 이벤트 관련동물 ID (ex. 교배 파트너, 송아지, 양육송아지 등)	지역 규모 농장 개체제번호 산차 수정차수 수정일 수정유형 KPN 분만일 임신기간(일) 종부횟수 공태일수 분만간격(일) 초산일령 송아지성별 출생체중(kg)	기록자 ID 관련동물 ID	No

표 4 에서 볼 수 있듯이 스마트팜 데이터마트에서 얻은 한우의 개체별 번식이력 데이터 셋은 [ICAR-03-2.3.4] 표준을 충족하지 않는다. 특히, 데이터를 기록한 기록자의 ID 와 교배 또는 번식

<sup>12</sup> ICAR Guidelines-Section 03-2.3.4 "Reproductive status of the animal"

이벤트에 관련된 다른 동물의 ID(예: 교배 파트너)와 같은 필수 데이터가 누락되어 있다. 또한, 각 필드의 의미를 설명하는 데이터설명서도 없어 데이터를 올바르게 사용하기 위한 필요한 정보가 부족하다. 예를들어, 스마트팜 데이터마트의 KPN 과 같은 용어에 대한 데이터설명서가 없기 때문에 데이터 사용자가 해당 용어를 파악하기 어렵고 수정차수와 중복횟수 모두 동일한 값을 표시하고 있는 것에 대한 설명이 없어 모호함을 유발한다. 우선 수정차수와 중복횟수에 대한 데이터를 혼동을 방지하고 중복되는 값을 제거하는 것이 필요하다.

또한, 그림 11 에 표시된 것처럼 식별 번호가 20031006010392 인 개체는 세 개의 다른 산차 값과 네 개의 수정차수 값을 포함한 다섯 개의 기록으로 나열되고 있다. 전체 데이터 셋을 면밀히 살펴보면, 이 개체의 열한 번째 산차는 세 번의 수정차수 후에 달성된 것으로 나타난다. 이는 열한 번째 산차에 대해 각각 고유한 수정차수(1, 2, 3)와 해당 수정일이 있는 세 개의 항목으로 분명해진다. 반면, 열두 번째와 열세 번째 산차는 각각 한 번씩 기록되어, 첫 번째 수정차수 시도가 성공적이었음을 암시한다. 이러한 데이터는 특히 항목이 체계적으로 정리되지 않았을 때 혼란스러울 수 있으며, 데이터설명서와 명확한 데이터 설명의 중요성, 그리고 불필요한 데이터의 중복을 피하기 위해 번식이력과 수정이력 데이터를 다른 데이터 셋으로 분리하는 것(normalization)의 중요성을 보여준다. 이렇게 분리 되었을 때 그림 11 에 보여지는 데이터는 수정이력에 해당된다.

그림 11. 개체별 번식이력(한우) 데이터 일부분

개체번호	산차	수정차수	수정일
20031006010392	13	1	2016-10-02
20031006010392	12	1	2015-11-14
20031006010392	11	3	2014-11-30
20031006010392	11	2	2014-11-06
20031006010392	11	1	2014-09-05

## 2) 연도별 번식성적(한우)

연도별 번식성적(한우) 데이터 셋은 국내 다양한 한우 농장에서의 소의 연간 평균 번식 성과를 평가하기 위한 데이터를 축적하고 있다. 각 데이터는 해당하는 분만 년도 내 개체들에 대한 모든 관련 측정치를 취한 후 이를 평균 내어 계산된 평균값을 나타내고 있다. 한우 데이터 셋의 단계별 선택 과정을 통해 제공받을 수 있는 연도별 번식성적(한우) 데이터 셋의 구조는 그림 12 과 같다.

그림 12. 연도별 번식성적(한우)

### 연도별 번식성적(한우)

한우 농가의 연도별 전체 개체 평균 번식성적을 확인할 수 있는 데이터셋

총 1,478 건이 검색되었습니다.

다운로드

비식별 데이터			년도	번식 성적			
지역	규모	농장	년도	초산일령	공태일수	분만간격(일)	임신기간(일)
3000000000	200~300	0021163	2015	694.0			285.0
3000000000	200~300	0021163	2016	1166.0	170.0	455.0	285.0
3000000000	200~300	0021163	2019	719.0			285.0
3000000000	200~300	0021163	2020	734.0	147.0	444.0	291.0
3000000000	200~300	0021163	2021	683.0			291.5
3611000000	100~200	0021161	2014	790.0			285.0
3611000000	100~200	0021161	2015		51.0	335.0	284.0
3611000000	100~200	0021161	2016		155.0	440.0	285.0
3611000000	100~200	0021161	2017		60.0	343.0	283.0
3611000000	100~200	0021161	2018	705.5	52.0	344.0	286.8

1 2 3 4 5 6 7 8 9 10 > >>

연도별 번식성적(한우) 데이터 셋과 관련된 ICAR 의 표준은 [ICAR-03-3.3.3]이다. 이 가이드라인에 따르면, 암컷의 번식 능력은 단순히 수정하고 배아를 하기까지의 능력 뿐만 아니라 살아 있는 송아지를 출산하고 출생 후 송아지의 성장을 위한 양육 환경을 제공하는 능력을 포함한다. 이러한 능력은 동물의 생애 이력과 번식 및 임신과 같은 번식 이정표를 검토함으로써 결정된다. 번식 성적의 철저한 분석을 위해서는 모든 관련된 번식 정보가 필요하다. 이에는 분만일, 초산일령(일 또는 월로 표현됨), 그리고 차후 분만에서의 나이가 포함된다. 또한, 다른 번식 이벤트들 사이의 시간 간격이 중요하다. 이에는 분만에서 첫 번식 시도까지의 간격, 분만에서 수정까지의 간격, 전체 분만 간격, 평균 생애 분만 간격, 분만까지의

평균 일수, 그리고 임신 기간이 포함될 수 있다. [ICAR-03-3.3.3] 표준과 스마트팜 데이터마트의 연도별 번식성적(한우) 데이터 구조의 비교는 아래 표 5 와 같다.

표 5. ICAR 표준과 스마트팜 데이터마트 비교 표 - 연도별 번식성적(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-03-3.3.3] <sup>13</sup> 분만일 초산일령 차후 분만 시 연령	지역 규모 농장 년도 초산일령 공태일수 분만간격(일) 임신기간(일)	차후 분만 시 연령	No

표 5 에 따르면 스마트팜 데이터마트에서 제공하는 번식 성능 데이터 셋은 분만년도, 초산일령, 분만간격, 공태일수, 임신기간 등의 세부 정보를 포함하여 [ICAR-03-3.3.3] 표준에서 제안하는 데이터 종류를 부분적으로 충족하고 있음을 알 수 있다. 그러나 차후 분만 시의 나이에 대한 데이터가 누락되어 있으며 번식 이력 데이터 셋과 마찬가지로 각 데이터에 대한 데이터설명서가 누락되어 있다.

<sup>13</sup> ICAR Guidelines-Section 03-3.3.3 "Female reproduction"

### 3) 개체별 출하정보(한우)

개체별 출하정보(한우) 데이터 셋은 한우 개체의 출하성적을 저장하고 있다. 이는 농장에서부터 시장에서 판매되기까지의 과정을 포괄적으로 추적할 수 있게 하여 최종적으로 제품의 품질과 가치에 영향을 미치는 요소에 대한 분석을 가능하게 한다. 이는 즉, 한우 산업에서 공급망 분석, 시장 가격 전략, 품질 관리에 유용하게 사용될 수 있는 데이터 셋이다. 한우 데이터 셋의 단계별 선택 과정을 통해 제공받을 수 있는 개체별 출하정보(한우) 데이터 셋의 구조는 그림 13 과 같다.

그림 13. 개체별 출하정보(한우)

#### 개체별 출하정보(한우)

비식별화된 한우 개체의 출하성적을 확인할 수 있는 데이터셋

총 17,783 건이 검색되었습니다.

다운로드

비식별 데이터				개체정보				
지역	규모	농장	개체번호	출생일자	품종	성별	월령	
3611000000	100~200	0021161	20211020010129	20211020	한우	거세	29	대전
3611000000	100~200	0021161	20210717010132	20210717	한우	거세	31	대전
3611000000	100~200	0021161	20210401010091	20210401	한우	거세	34	대전
3611000000	100~200	0021161	20160309010052	20160309	한우	암	89	농협
3611000000	100~200	0021161	20180329010041	20180329	한우	암	65	농협
3611000000	100~200	0021161	20200930010039	20200930	한우	거세	34	농협
3611000000	100~200	0021161	20201111010024	20201111	한우	거세	32	농협
3611000000	100~200	0021161	20210220010008	20210220	한우	거세	29	농협
3611000000	100~200	0021161	20180620010026	20180620	한우	암	57	농협
3611000000	100~200	0021161	20180612010034	20180612	한우	암	57	농협

1 2 3 4 5 6 7 8 9 10 > >>

개체별 출하정보(한우) 데이터 셋과 관련된 ICAR 의 표준은 [ICAR-03-3.11]과 [ICAR-03-4.4]이다. 해당 가이드라인에 따르면, 도체 평가와 관련하여 도축장에서 기록을 얻기 위한 필수 전제 조건은 동물의 ID 가 도체와 함께 유지되거나, 해당 동물의 ID 와 도체 데이터를 보고할 수 있는 시스템이 사용되어야 한다. 필수적으로 기록되어야 하는 데이터의 종류에는 도체중, 예상 육량(퍼센트 또는 점수로 기록됨), 그리고 등급이 있다. 또한 상업적 도축 데이터와 관련하여, 각 동물에 대해 최소한 사육 농장의 ID, 동물 ID 번호, 도체 무게, 도축 날짜, 국가

등급 시스템에 따른 도체등급이 기록되어야 하며, 추가적으로 수축 중량, 고기 수율 결정에 영향을 미치는 도체 세부 사항, 고기 수율, 지방 비율, 형태 점수 및 지방 점수의 결정을 가능하게 하는 영상 데이터 결과 등을 포함할 수 있다. 이에 대해 스마트팜 데이터마트와 비교한 표는 다음의 표 6 와 같다.

표 6. ICAR 표준과 스마트팜 데이터마트 비교 표 - 개체별 출하정보(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-03-3.11] <sup>14</sup> 동물 ID 도체중 예상 육량 (백분율 또는 점수로 기록) 등급  [ICAR-03-4.4] <sup>15</sup> 사육 농장 ID 동물 ID 도체 중량 도축 날짜 도체 등급	지역 규모 농장 개체제번호 출생일자 품종 성별 월령 도축장 도축일자 육량등급 육질등급 도체중(kg) 등급판정일자 육량지수 등급명 등지방두께(mm) 등심면적(cm <sup>2</sup> )	없음	Yes

<sup>14</sup> ICAR Guidelines-Section 03-3.11 "Carcass assessments"

<sup>15</sup> ICAR Guidelines-Section 03-4.4 "Commercial Slaughter Data"

	근내지방 육색 지방색 조직감 성숙도 경락가격		
--	---	--	--

표 6에서 볼 수 있듯이 스마트팜 데이터마트의 출하정보 데이터 셋은 ICAR의 도체 평가 [ICAR-03-3.11] 및 상업적 도축 데이터 [ICAR-03-4.4]에 의해 설정된 표준을 만족하고 있다. 그러나 일부 열에 누락된 값이나 잘못된 문자가 포함되어 있는 등 데이터 품질에 문제가 있다 (그림 14 참조). 또한, 데이터설명서가 부재하여 데이터베이스 항목에 대한 측정 단위나 허용 가능한 값 범위가 정확히 무엇인지 알 수 없다.

#### 4) 연도별 출하성적(한우)

출하성적(한우) 데이터 셋은 연간 기준으로 국내 한우 농장에서의 고기 품질 등급의 분포 및 비율을 분석하기 위해 설계되었다. 이 데이터 셋은 국내 농장에서 생산된 한우의 품질 분포를 이해하고 해마다 어떻게 변화하는지 관찰하는 데 가치가 있다. 이는 업계 이해관계자들이 표준을 지속적으로 모니터링하고 소비자들이 이를 인식하여 한우 품질데이터에 대해 신뢰하며 연구자들이 한우 품질의 추세를 연구하는 데 사용될 수 있다. 한우 데이터 셋의 단계별 선택 과정을 통해 제공받을 수 있는 연도별 출하성적(한우) 데이터 셋의 구조는 그림 14 와 같다.

그림 14. 연도별 출하성적(한우)

#### 연도별 출하성적(한우)

한우 농가의 연도별 육질등급별 마릿수와 비율을 확인할 수 있는 데이터셋

총 676 건이 검색되었습니다.

다운로드

비식별 데이터			년도	육질 등급별 마릿수와 비율										
지역	규모	농장	년도	1++(마리)	1+(마리)	1(마리)	2(마리)	3(마리)	D(마리)	총합(마리)	1++ 비율(%)	1+ 비율(%)	1(마리)	
3611000000	100~200	0021161	2020	0	1	2	1	0	0	4	0.0%	25.0%	5	
3611000000	100~200	0021161	2021	0	5	2	0	1	0	8	0.0%	62.5%	2	
3611000000	100~200	0021161	2022	8	9	1	3	1	0	22	36.4%	40.9%	4	
3611000000	100~200	0021161	2023	2	1	2	2	1	0	8	25.0%	12.5%	2	
3611000000	100~200	0021161	2024	0	0	3	0	0	0	3	0.0%	0.0%	10	
4100000000	100두 미만	0022912	2016	0	1	1	0	0	0	2	0.0%	50.0%	5	
4100000000	100두 미만	0022912	2019	0	1	0	0	0	0	1	0.0%	100.0%	0	
4100000000	100두 미만	0022912	2022	6	13	13	2	1	0	35	17.1%	37.1%	3	
4100000000	100두 미만	0022912	2023	3	16	8	5	0	0	32	9.4%	50.0%	2	
4100000000	1000두 이상	0020192	2019	1	1	0	0	0	0	2	50.0%	50.0%	0	

연도별 출하성적(한우) 데이터 셋과 관련하여 ICAR의 표준은 찾을 수 없다. ICAR은 육우의 도체 품질에 대한 집계 기록이나 성과에 대해 특정한 지침을 제공하지 않는다. 따라서 해당 데이터 셋이 표준화된 기준에 부합하는지 여부는 판단할 수 없다 (표 7 참조). 게다가, 이 데이터 셋 역시 데이터들에 대한 명확한 정의나 설명이 없어 연도별 출하성적이 도축일자의 연도를 의미하는지 등급판정일자의 연도를 의미하는지 알 수 없다. 물론 모든 데이터 셋 값을 면밀히 살펴보면 도축일자에 대해 구성된 것으로 추론할 수 있지만 이와같은 과정은 사실 해당 데이터에 대한 명확한 설명이 있으면 쉽게 해결될 수 있는 문제로 현재 데이터 이용자로 하여금 불필요한 노력을 요구하게 한다.

표 7. ICAR 표준과 스마트팜 데이터마트 비교 표 - 연도별 출하성적(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
구체적 항목 없음	지역 규모 농장 연도 1++(마리) 1+(마리) 1(마리) 2(마리) 3(마리) D(마리) 총합(마리) 1++ 비율(%) 1+ 비율(%) 1 비율(%) 2 비율(%) 3 비율(%) D 비율(%)	없음	판단불가

## 5) 개체별 혈통족보(한우)

개체별 혈통족보(한우)는 한우의 혈통족보와 출하성적을 평가하고 예측하기 위해 맞춤 제작되었으며, 어미소가 훌륭한 품질의 소인지 식별하고 그들의 자손에 대한 결과를 예측하는데 중점을 둔다. 이 데이터 셋은 자세한 혈통 정보 내역과 출하 성과 지표를 통합하여 데이터를 제공하고 있으며, 한우 데이터 셋의 단계별 선택 과정을 통해 제공받을 수 있는 개체별 혈통족보 (한우) 데이터 셋의 구조는 그림 15 과 같다.

그림 15. 개체별 혈통족보(한우)

### 개체별 혈통족보(한우)

개체의 혈통족보와 출하성적을 결합하여 어미소가 고능력 우인지 여부와 자손들의 출하성적을 예상할 수 있는 데이터셋

총 100,061 건이 검색되었습니다.

다운로드

비식별 데이터			비식별 혈통족보								출하 성적			
지역	규모	농장	모개체	1세대	2세대	3세대	4세대	5세대	6세대	7세대	생년월일	씨수소명	도체중	육질등급
4500000000	200~300	0021173	암											
4500000000	200~300	0021173		암							2009-12-23	KPN531	435.000	2
4500000000	200~300	0021173			수						2012-08-10	KPN710	487.000	2
4500000000	200~300	0021173			수						2013-09-14	KPN828	499.000	2
4500000000	200~300	0021173			수						2014-12-18	KPN969	474.000	2
4500000000	200~300	0021173			수						2016-03-13	KPN866	518.000	2
4500000000	200~300	0021173			암						2017-04-24	KPN828		
4500000000	200~300	0021173				수					2019-05-16	KPN1115	533.000	2
4500000000	200~300	0021173				수					2020-05-12	KPN1100		
4500000000	200~300	0021173				암					2021-11-25	KPN1302		

1 2 3 4 5 6 7 8 9 10 > >>

개체별 혈통족보(한우)와 관련이 있는 ICAR 표준은 [ICAR-03-3.1.2]이 해당된다. 이 ICAR 표준의 가이드 라인에 따르면 일반적으로 혈통 정보는 동물의 식별 번호와 그 부모의 식별 번호, 품종, 성별, 출생일을 포함한 정보를 나타낸다. 또한 번식 집단의 유전 구조에 기여하는 모든 동물들을 포함시켜서 생각해야 한다. 기록에 있어서는 가능한 경우 수소와 어미소의 신원 번호, 이름, 품종, 교배 이벤트, 인공 수정이 사용되거나 자연 교배가 목격된 경우의 교배 날짜를 포함시켜야 한다. 만약 교배가 목격되지 않은 경우에는 어미와 수소가 함께 있었던 기간을 기록해야 한다. 이와 관련하여 스마트팜 데이터마트와 비교한 표는 다음 표 8 과 같다.

표 8. ICAR 표준과 스마트팜 데이터마트 비교 표 - 개체별 혈통축보(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-03-3.1.2] <sup>16</sup> 모개체 ID 씨수소명 손자 개체제번호 수정일	지역 규모 농장 모개체 1 계대 2 계대 3 계대 4 계대 5 계대 6 계대 7 계대 생년월일 씨수소명 도체중 육질등급 도축일 월령	모개체 ID 손자 개체제번호 수정일	No

표 8 에 따르면 스마트팜 데이터마트에서 제공하는 개체별 혈통축보 데이터 셋은 ICAR 표준의 [ICAR-03-3.1.2] 가이드라인을 충족하지 못한다. 씨수소명은 포함되어 있지만 모개체 ID 를 생략하고 모개체 열에서는 오직 모개체의 성별만을 제공하고 있는데 이는 특별히 유용한 정보가 아니다. 또한 수정일과 같은 중요한 데이터가 부재하다. 게다가 명확한 데이터 설명이 없어 데이터를 해석하는데 어려움이 있다. 예를 들어, 아래의 그림 16 을 보면 농장 0020224 의

<sup>16</sup> ICAR Guidelines-Section 03-3.1.2 "Parentage recording"

경우, 2 계대에 걸쳐 자손을 가진 두 모개체가 나열되어 있지만 모개체 ID, 손자 개체제번호가 없어서 어떤 모개체에 속하는지 추적하여 분석하는 것이 불가능하다. 또한, 해당 데이터 셋의 날짜 형식(YYYY-MM-DD)은 다른 데이터 셋에서 사용되는 날짜형식(YYYYMMDD)과 다른 포맷을 사용하고 있다. 이는 데이터 사용자로 하여금 불필요한 데이터 전처리 과정을 늘리게 되어 비효율을 유발하기 때문에 되도록 모든 데이터 셋에 대해 공통된 날짜 형식을 사용해야 한다. 또한, 혈통족보 데이터 셋의 구조 역시 사용자 친화적이지 않다. 계대를 하나하나를 별도의 열로 제시하기보다 계대 번호를 기록하는 단일 열을 가지는 것이 분석에 더 효율적일 것이다.

그림 16. 개체별 혈통족보(한우) 데이터 셋의 데이터 일부분

농장	모개체	1계대	2계대
0020224	암		
0020224		암	
0020224		수	
0020224		수	
0020224		수	
0020224		암	
0020224			거세
0020224			거세
0020224			암
0020224			암
0020224			거세
0020224			암
0020224			거세
0020224			거세
0020224			암
0020224			암
0020224		암	
0020224		암	
0020224		수	
0020224	암		
0020224		수	
0020224		암	
0020224		거세	
0020224		암	
0020224			수
0020224			거세
0020224			암
0020224			거세
0020224			거세
0020224			암

## 6) 개체별 혈통성적(한우)

개체별 혈통성적(한우) 데이터 셋은 한우 자손의 출하 성적을 종합적으로 추적하는데 활용될 수 있는 데이터를 모으고 있으며, 특히 최종적으로 생산된 고기의 품질에 대한 데이터를 중점적으로 제공한다. 이는 다양한 한우 품질 등급에 따라 분류된 도체의 수와 백분율을 상세히 나열하고 개별 한우의 혈통 관계와 연결한다. 한우 데이터 셋의 단계별 선택 과정을 통해 제공받을 수 있는 개체별 혈통성적(한우) 데이터 셋의 구조는 그림 17 과 같다.

그림 17. 개체별 혈통성적(한우)

### 개체별 혈통성적(한우)

모개체별 자손의 출하성적에서 육질등급별 마릿수와 비율을 확인할 수 있는 데이터셋

총 12,542 건이 검색되었습니다.

다운로드

비식별 데이터			모개체별 자손의 육질등급별 마릿수									육질		
지역	규모	농장	모개체	1++(마리)	1+(마리)	1(마리)	2(마리)	3(마리)	D(마리)	총합(마리)	1++ 비율(%)	1+ 비율(%)	1 비율(%)	
3000000000	200~300	0021163	1	0	1	1	1	0	0	3	0.00	33.30	33.30	
3000000000	200~300	0021163	2	1	0	2	0	0	0	3	33.30	0.00	66.70	
3000000000	200~300	0021163	3	0	0	1	0	0	0	1	0.00	0.00	100.00	
3000000000	200~300	0021163	4	0	0	0	0	0	0	0	0.00	0.00	0.00	
3000000000	200~300	0021163	5	0	1	0	0	0	0	1	0.00	100.00	0.00	
3000000000	200~300	0021163	6	0	0	0	0	0	0	0	0.00	0.00	0.00	
3000000000	200~300	0021163	7	0	0	1	0	0	0	1	0.00	0.00	100.00	
3000000000	200~300	0021163	8	0	0	0	0	0	0	0	0.00	0.00	0.00	
3000000000	200~300	0021163	9	0	0	0	0	0	0	0	0.00	0.00	0.00	
3000000000	200~300	0021163	10	0	0	1	1	0	0	2	0.00	0.00	50.00	

1 2 3 4 5 6 7 8 9 10 > >>

개체별 개체별 혈통성적(한우)(한우)와 관련하여 ICAR 표준 가이드라인은 육우의 도체 품질이나 혈통 성적에 대한 데이터를 집계하는 것에 대해 구체적인 지침을 제공하지 않는다. 따라서 해당 데이터 셋이 ICAR 표준을 준수하는지 여부는 판단할 수 없다 (표 9 참조). 해당 데이터 셋 역시 제공하고 있는 데이터에 대한 명확한 정의를 제공하지 않고 있으며 특히 모개체 열이 단일 모개체 ID 를 참조하는지 아니면 모개체의 총 수를 참조하는지에 대해서 모호함을 유발하고 있다.

표 9. ICAR 표준과 스마트팜 데이터마트 비교 표 - 개체별 혈통성적(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
구체적 항목 없음	지역 규모 농장 모개체 1++(마리) 1+(마리) 1(마리) 2(마리) 3(마리) D(마리) 총합(마리) 1++ 비율(%) 1+ 비율(%) 1 비율(%) 2 비율(%) 3 비율(%) D 비율(%)	없음	판단불가

## 7) 개체별 EPD(한우)

개체별 EPD(한우) 데이터 셋은 개별 한우의 예상 유전 전달 능력에 대한 데이터를 제공한다. EPD(Expected Progeny Differences)는 동물 번식에 사용되며 개체의 부모로서의 유전적 잠재력의 추정치에 해당한다. 이 데이터 셋은 미래의 자손의 품질 특성에 대해 EPD 데이터를 기반으로 번식할 동물을 선택하려는 사육자와 유전학자들에게 필수적인 데이터 셋이다. 이는 정보에 기반한 번식 결정을 내리는 데 중요한 개별 동물의 다양한 신체적 및 품질 특성의 스냅샷을 제공하고 있으며 해당 데이터 셋은 한우 데이터 셋의 단계별 선택 과정을 통해 그림 18 과 같이 제공받을 수 있다.

그림 18. 개체별 EPD(한우)

### 개체별 EPD(한우)

개체의 EPD정보를 확인할 수 있는 데이터셋

총 13,835 건이 검색되었습니다.

다운로드

비식별 데이터				EPD정보							
지역	규모	농장	개체식별번호	생년월일	성별구분	계대	냉도체중등급	냉도체중지수	등심단면적등급	등심단면적지수	
4100000000	100두 미만	0022913	20220317010358	2022-03-17	암	3	A	19.52790	B	2.98480	
4100000000	100두 미만	0022913	20220317010357	2022-03-17	암	6	B	10.97060	D	1.37530	
4100000000	100두 미만	0022913	20220319010356	2022-03-19	수	4	A	17.38330	B	3.28850	
4100000000	100두 미만	0022913	20220322010355	2022-03-22	수	3	A	24.40700	A	4.10440	
4100000000	100두 미만	0022913	20220420010354	2022-04-20	수	8	B	11.41120	C	1.85760	
4100000000	100두 미만	0020348	20180801010001	2018-08-01	암	1	D	3.74730	B	2.44800	
4100000000	100두 미만	0020348	20181112010031	2018-11-12	암	1	D	4.00650	D	0.00950	
4100000000	100두 미만	0020348	20181207010016	2018-12-07	암	3	D	4.39050	D	1.02610	
4100000000	100두 미만	0020348	20190328010009	2019-03-28	암	6	C	9.82320	D	0.69460	
4100000000	100두 미만	0020348	20190424010004	2019-04-24	암	1	D	3.25430	D	-0.63820	

1 2 3 4 5 6 7 8 9 10 > >>

ICAR 은 EPD 기록이나 보고를 위한 구체적인 지침을 제공하지 않고 있다. EPD 는 일반적으로 ICAR 보다는 품종 협회에 의해 관리된다. 구체적인 정보는 품종 협회, EPD 의 사용 목적(예: 특정 특성에 대한 선택), 개별 생산자의 선호에 따라 다를 수 있음을 주목하는 것이 중요하다. ICAR 이 EPD(Expected Progeny Differences)의 기록이나 보고를 위한 구체적인 지침을 제시하고 있지 않지만, 정확하고 일관된 데이터 기록을 위한 원칙은 여전히 관련이 있다.

따라서 스마트팜 데이터마트에서 제공하는 EPD 데이터는 ICAR 표준의 [ICAR-03-2.3.5.4] 가이드라인을 따를 수 있으며 이 가이드 라인에는 다음 항목들이 포함되며 이를 스마트팜 데이터마트와 비교한 내용은 표 10 에서 제시하고 있다.

- a. 동물 ID (해당되는 경우, 동물 그룹)
- b. 평가 날짜
- c. 기록자 ID
- d. 실제 위치, 예를 들어 농장 ID (해당되는 경우, 농장 내 관리 그룹)
- e. 특성 이름 또는 코드
- f. 특성의 값

표 10. ICAR 표준과 스마트팜 데이터마트 비교 표 – 개체별 EPD(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-03-2.3.5.4] <sup>17</sup> 동물 ID 평가일자 기록자 ID 농장 ID 형질 이름 형질 값	지역 규모 농장 개체식별번호 생년월일 성별구분 계대 냉도체중등급	기록자 ID	No

<sup>17</sup> ICAR Guidelines-Section 03-2.3.5.4 "Recorded data"

	냉도체중지수 등심단면적등급 등심단면적지수 등지방두께등급 등지방두께지수 근내지방등급 근내지방지수 갱신일자		
--	--	--	--

표 10 에 따르면 현재 스마트팜 데이터마트의 EPD 데이터 셋은 [ICAR-03-2.3.5.4] 가이드라인을 대부분 충족하고 있지만 기록자 ID 에 대한 정보가 누락되어 있다. 또한, 각 데이터에 대한 명확한 정의를 제공하지 않아 계대와 같은 특정 데이터에 대해 다른 의미로 해석할 수 있는 여지가 있다. 게다가 생년월일과 갱신일자 데이터 열 내의 날짜 형식(YYYY-MM-DD) 역시 스마트팜 데이터마트 데이터베이스의 다른 날짜 관련 열(YYYYMMDD)과 일치하지 않고 있다. 명확성과 일관성을 향상시키기 위해서는 모든 데이터 셋에서 통일된 날짜 형식을 사용해야 한다.

## 8) ICT 데이터-체중(한우)

ICT 데이터-체중(한우) 데이터 셋은 ICT 기기를 활용하여 기록한 한우의 체중을 시계열로 기록하여 저장하기 위한 데이터 셋이다. 해당 데이터 셋은 한우 데이터 셋의 단계별 선택 과정을 통해 제공받을 수 있으나, 그림 19 과 같이 현재 기록되어 있는 데이터는 없다.

그림 19. ICT 데이터-체중(한우)

**ICT데이터 체중(한우)**  
 체중측정기의 측정일시, 체중 ICT RAW 데이터를 확인할 수 있는 데이터셋  
 총 0 건이 검색되었습니다. ↓ 다운로드

비식별 데이터				장비 정보		장비 측정 데이터	
지역	규모	농장	개체번호	제조사아이디	장비번호	측정일시	체중(kg)
데이터가 없습니다.							

현재 스마트팜 데이터마트에 저장된 데이터는 없지만 이 데이터 셋과 관련한 ICAR 표준이 존재하기 때문에 제공하고자 하는 데이터 항목에 대해 비교 분석이 가능하다. 관련 있는 ICAR 표준은 [ICAR-03-3.8.2.11.4], [ICAR-03-3.5], [ICAR-03-2.3.1.3]에 해당하며 스마트팜 데이터 마트와 비교한 내용은 표 11 에 제시되어 있다.

표 11 에 따르면 스마트팜 데이터마트의 데이터는 [ICAR-03-3.8.2.11.4]에 명시된 최소 표준(날짜, 대상 ID, 체중 값으로 이루어진 간단한 데이터 기록)을 충족하였다. 하지만 [ICAR-03-3.5]에 따르면 다양한 종류의 중량(출생 체중, 이유 체중, 이유 후 체중, 마무리 체중, 테스트 체중, 추정 체중 또는 조정 체중)이 기록되어야 한다고 지정하고 있다. 따라서 스마트팜 데이터마트 데이터가 어떤 종류의 체중이 기록되고 있는지를 나타내는 데이터가 추가로 제공되어야 한다. 또한, 소의 체중 측정하는 방법에는 다양한 방식이 존재하기 때문에 어떤 방식으로 체중을 측정하였는지에 대한 설명 역시 필요하다. 마지막으로, [ICAR-03-2.3.1.3]에 따르면 누가 소의 체중을 측정하였는지를 기록하는 것도 요구된다.

표 11. ICAR 표준과 스마트팜 데이터마트 비교 표 - 체중(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-03-3.8.2.11.4] <sup>18</sup> 동물 ID 체중 측정일 체중  [ICAR-03-3.5] <sup>19</sup> 증량 측정방법 생시총체 이유체중 이유전체중, 이유후체중, 테스트 체중 추정 체중 수정 체중  [ICAR-03-2.3.1.3] <sup>20</sup> 기록자 ID	지역  규모  농장  개체번호 제조사아이디  장비번호  측정일시  체중(kg)	측정 체중 종류 체중 측정 방식  기록자 ID	No

<sup>18</sup> ICAR Guidelines-Section 03-3.8.2.11.4 "Weight details of animals on test"

<sup>19</sup> ICAR Guidelines-Section 03-3.5 "Live animal weights"

<sup>20</sup> ICAR Guidelines-Section 03-2.3.1.3 "Invariant recording personnel data"

## 9) ICT 데이터-환경관리기(한우)

ICT 데이터-환경관리기(한우) 데이터 셋은 농장에서 사용되는 다양한 환경관리기의 온도, 습도, CO2, 암모니아에 대해 측정된 시계열 데이터를 기록한다. 이 데이터 셋은 한우 데이터 셋의 단계별 선택 과정을 통해 그림 20 과 같이 제공받을 수 있다.

그림 20. ICT 데이터-환경관리기(한우)

### ICT데이터 환경관리기(한우)

환경관리기의 온도, 습도, CO2, 암모니아 ICT RAW 데이터를 확인할 수 있는 데이터셋

총 51,523 건이 검색되었습니다.

다운로드

비식별 데이터			장비 정보			장비 측정 데이터	
지역	규모	농장	장비종류	제조사아이디	장비번호	측정일자	측정값(°C · % · ppm)
3611000000	100~200	0021161	온도	FSSYSTEM	1	2022-04-05	11
3611000000	100~200	0021161	온도	FSSYSTEM	1	2022-04-04	11
3611000000	100~200	0021161	온도	FSSYSTEM	1	2022-04-03	9
3611000000	100~200	0021161	온도	FSSYSTEM	1	2022-04-02	9
3611000000	100~200	0021161	온도	FSSYSTEM	1	2022-04-01	10
3611000000	100~200	0021161	온도	FSSYSTEM	1	2022-03-31	13
3611000000	100~200	0021161	온도	FSSYSTEM	1	2022-03-30	12
3611000000	100~200	0021161	온도	FSSYSTEM	1	2022-03-29	10
3611000000	100~200	0021161	온도	FSSYSTEM	1	2022-03-27	10
3611000000	100~200	0021161	온도	FSSYSTEM	1	2022-03-26	13

1 2 3 4 5 6 7 8 9 10 > >>

ICAR 에서는 온도, 습도, CO2, 암모니아의 기록에 대해 명시적인 가이드라인을 제안하지 않는다. 유일하게 사용 가능한 ICAR 표준은 온도 기록에 대한 [ICAR-3.7.4.2]과 사료 섭취를 바탕으로 한 메탄 측정 [ICAR-6.5]이 있다. 따라서 이 두 가지 가이드라인을 바탕으로 스마트팜 데이터마트의 데이터와 비교한 내역은 표 12 과 같다. 다만 명확히 제공되는 항목이 없어 ICAR 준수 여부를 판단할 수 없으며 스마트팜 데이터 마트는 측정 시간에 대한 정보를 추가로 제공하는 것이 필요하다. 현재 마지막 데이터 열의 값이 계산되는 위치(일일 평균인지 또는 하루 중 특정 시간의 실제 측정 값인지)를 나타내지 않기 때문이다.

표 12. ICAR 표준과 스마트팜 데이터마트 비교 표 - 환경관리기(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-3.7.4.2] <sup>21</sup> 구체적 항목 없음  [ICAR-20.6] <sup>22</sup> 구체적 항목 없음	지역 규모 농장 장비종류 제조사아이디 장비번호 측정일시 측정값(°C · % · ppm)	없음	판단불가

<sup>21</sup> ICAR Guidelines-Section 03-3.7.4.2 "Facilities"

<sup>22</sup> ICAR Guidelines-Section 20.6 "Methane measurement methods"

## 10) ICT 데이터-사료빈(한우)

ICT 데이터-사료빈(한우) 데이터 셋은 개별 사료빈에서 측정된 사료빈의 잔여량, 배출량에 대한 로우 데이터를 저장하고 있으며 이를 통해 사료 효율을 계산할 수 있다. 이 데이터 셋은 한우 데이터 셋의 단계별 선택 과정을 통해 그림 21 와 같이 제공받을 수 있다.

그림 21. ICT 데이터-사료빈(한우)

### ICT데이터 사료빈(한우)

사료빈의 잔여량, 배출량 ICT RAW 데이터를 확인할 수 있는 데이터셋

총 277,276 건이 검색되었습니다.

다운로드

비식별 데이터			장비정보		장비 측정 데이터		
지역	규모	농장	제조사아이디	장비번호	측정일시	잔여량(kg)	배출량(kg)
4100000000	100두 미만	0022344	SAMWOO	1	2021-12-02 23:26:05	0	1,156
4100000000	100두 미만	0022344	SAMWOO	1	2021-12-02 22:26:04	0	1,156
4100000000	100두 미만	0022344	SAMWOO	1	2021-12-02 21:26:04	0	1,156
4100000000	100두 미만	0022344	SAMWOO	1	2021-12-02 20:26:04	0	1,156
4100000000	100두 미만	0022344	SAMWOO	1	2021-12-02 19:26:04	0	1,156
4100000000	100두 미만	0022344	SAMWOO	1	2021-12-02 18:26:04	0	1,156
4100000000	100두 미만	0022344	SAMWOO	1	2021-12-02 17:21:05	0	1,156
4100000000	100두 미만	0022344	SAMWOO	1	2021-12-02 16:21:04	0	1,156
4100000000	100두 미만	0022344	SAMWOO	1	2021-12-02 15:21:04	0	1,156
4100000000	100두 미만	0022344	SAMWOO	1	2021-12-02 14:21:03	0	842

1 2 3 4 5 6 7 8 9 10 > >>

사료 효율은 소비된 사료 양과 배출량의 양을 비교하여 분석할 수 있다. ICT 데이터-사료빈 데이터 셋의 특징들은 [ICAR-19]에서 명시한 최소 표준(표 13 참조)과 일치한다. ICAR 에서는 잔여량과 배출량 기록이 서로 상관관계가 높음에도 불구하고 분리되어 있다. [ICAR-19-10]에 따르면 개별 급이기를 사용하지 않는 경우에는 사료빈 당 동물 수에 특히 관심을 가지고 있다. 이러한 점을 보아 국내 스마트팜 데이터에 대해서도 사료빈 당 동물의 수에 대한 정보를 추가로 제공하는 것이 필요해 보인다.

표 13. ICAR 표준과 스마트팜 데이터마트 비교 표 - 사료빈(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-19] <sup>23</sup> 건초 배합비율(%) 사료 제공량  [ICAR-19-10] <sup>24</sup> 사료빈당 동물수	지역 규모 제조사 ID 장비번호 측정일시 잔여량(kg) 배출량(kg)	없음	Yes

<sup>23</sup> ICAR Guidelines-Section 19 "Guidelines for recording feed intake for genetic evaluation"

<sup>24</sup> ICAR Guidelines-Section 19-10 "Number of animals per bin"

## 11) ICT 데이터-일반급이기(한우)

ICT 데이터-일반급이기(한우) 데이터 셋은 일반급이기를 사용할 때 급이량에 대한 측정 데이터를 저장하고 있다. 각 데이터 포인트는 지정된 시간에 사료빈에서 급이된 급이량을 나타낸다. 이 데이터 셋은 한우 데이터 셋의 단계별 선택 과정을 통해 그림 22 과 같이 제공받을 수 있다.

그림 22. ICT 데이터-일반급이기(한우)

### ICT데이터 일반급이기(한우)

일반급이기의 급이량 ICT RAW 데이터를 확인할 수 있는 데이터셋

총 60,580 건이 검색되었습니다.

다운로드

비식별 데이터			장비 정보		장비 측정 데이터	
지역	규모	농장	제조사아이디	장비번호	측정일시	급이량(kg)
4100000000	100두 미만	0022044	미래씨엔씨	4	2023-12-31	1820.00
4100000000	100두 미만	0022044	미래씨엔씨	4	2023-12-29	1820.00
4100000000	100두 미만	0022044	미래씨엔씨	4	2023-12-28	1820.00
4100000000	100두 미만	0022044	미래씨엔씨	4	2023-12-27	1820.00
4100000000	100두 미만	0022044	미래씨엔씨	4	2023-12-23	1820.00
4100000000	100두 미만	0022044	미래씨엔씨	4	2023-12-22	1820.00
4100000000	100두 미만	0022044	미래씨엔씨	4	2023-12-21	1820.00
4100000000	100두 미만	0022044	미래씨엔씨	4	2023-12-19	1820.00
4100000000	100두 미만	0022044	미래씨엔씨	4	2023-12-15	1820.00
4100000000	100두 미만	0022044	미래씨엔씨	4	2023-12-12	1820.00

1 2 3 4 5 6 7 8 9 10 > >>

일반급이기 데이터 셋은 이전의 사료빈 데이터 셋과 유사하다. 이 데이터 셋의 모든 특징은 사료빈 데이터 셋과 연결되어 있다. 따라서 이 데이터의 특징도 [ICAR-19 & ICAR-3.8.2.11.3]에서 명시한 최소 표준(표 14 참조)을 만족하고 있다. 특히 이 데이터 셋의 급이량 데이터의 특징은 표준 [ICAR-3.8.1]과도 일치하고 있다. 사료빈 데이터 셋과 마찬가지로 일반적으로 여러 마리의 한우가 함께 사용하는 일반급이기에 대해서 해당 사료빈 당 한우의 수에 대한 데이터가 포함될 필요가 있다고 판단된다.

표 14. ICAR 표준과 스마트팜 데이터마트 비교 표 - 일반급이기(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-19] <sup>25</sup> 사료량 제공량 잔여량 관련 건자재(%) 사료빈당 동물 수 [ICAR-3.8.2.11.3] <sup>26</sup> 동물 ID 기록 날짜 기간 동안의 사료 섭취량 [ICAR-3.8.1] <sup>27</sup> 사료 섭취량 EBV	지역 규모 농장 제조사아이디 장비번호 측정일시 급이량(kg)	없음	Yes

<sup>25</sup> ICAR Guidelines-Section 19 "Guidelines for recording feed intake for genetic evaluation"

<sup>26</sup> ICAR Guidelines-Section 03-3.8.2.11.3 "Intake details of animals on test"

<sup>27</sup> ICAR Guidelines-Section 03-3.8.1 "Feed intake"

## 12) ICT 데이터-군사급이기(한우)

ICT 데이터-군사급이기(한우) 데이터 셋은 군사급이기를 통해 측정된 섭취량, 설정량 등의 급이 데이터를 저장하고 있다. 여기서 군사급이기는 일반적으로 한 급이기당 한 마리의 소가 배정되기 때문에, 해당 소를 식별하기 위한 대상 번호 기능이 데이터 셋에 포함되어 있다. 이 데이터 셋은 한우 데이터 셋의 단계별 선택 과정을 통해 그림 23 와 같이 제공받을 수 있다.

그림 23. ICT 데이터-군사급이기(한우)

### ICT데이터 군사급이기(한우)

군사(개별)급이기의 섭취량, 설정량 ICT RAW 데이터를 확인할 수 있는 데이터셋

총 243,070 건이 검색되었습니다.

다운로드

비식별 데이터				장비 정보		장비 측정 데이터				
지역	규모	농장	개체번호	제조사아이디	장비번호	측정일시	섭취량(kg)	설정량(kg)	급이시작일시	급이종료일시
500~1000	0020418	20060730011031	주식회사 다운	3	2023-08-16	0	9	2023-08-16 08:38:00	2023-08-16 08:38:00	
500~1000	0020418	20060730011031	주식회사 다운	3	2023-08-16	0	9	2023-08-16 10:23:00	2023-08-16 10:23:00	
500~1000	0020418	20060730011031	주식회사 다운	3	2023-08-16	0	9	2023-08-16 08:37:00	2023-08-16 08:37:00	
500~1000	0020418	20060730011031	주식회사 다운	3	2023-08-07	0	9	2023-08-07 12:24:00	2023-08-07 12:24:00	
500~1000	0020418	20060730011031	주식회사 다운	3	2023-08-07	0	9	2023-08-07 09:36:00	2023-08-07 09:36:00	
500~1000	0020418	20060730011031	주식회사 다운	3	2023-08-07	2	9	2023-08-07 12:19:00	2023-08-07 12:19:00	
500~1000	0020418	20060730011031	주식회사 다운	3	2023-08-07	0	9	2023-08-07 15:29:00	2023-08-07 15:29:00	
500~1000	0020418	20060730011031	주식회사 다운	4	2023-07-29	0	8	2023-07-29 01:08:00	2023-07-29 01:08:00	
500~1000	0020418	20060730011031	주식회사 다운	4	2023-07-29	2	8	2023-07-29 06:47:00	2023-07-29 06:47:00	
500~1000	0020418	20060730011031	주식회사 다운	4	2023-07-29	1	8	2023-07-29 00:37:00	2023-07-29 00:37:00	

1 2 3 4 5 6 7 8 9 10 > >>

표 15 에서 볼 수 있듯이, 전반적으로 이 데이터 셋은 [ICAR-3.8.2.11.3] 가이드 라인의 최소 요구 사항을 충족하고 있다. 다만 급이시작일시, 급이종료일시는 ICAR 가이드 라인에서 명시하고 있지는 않지만 군사급이기 시스템에서 직접 나온 데이터로 [ICAR-3.8.2.11.3]에 명시된대로 군사급이기 시스템에서 직접 나온 데이터로 판단된다.

표 15. ICAR 표준과 스마트팜 데이터마트 비교 표 - 군사급이기(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-3.8.2.11.3] <sup>28</sup> 동물 ID 기간간 섭취	지역 규모 농장 개체번호 제조사아이디 장비번호 측정일시 섭취량(kg) 설정량(kg) 급이시작일시 급이종료일시	없음	Yes

<sup>28</sup> ICAR Guidelines-Section 03-3.8.2.11.3 "Intake details of animals on test"

### 13) ICT 데이터-음수관리기(한우)

ICT 데이터-음수관리기(한우) 데이터 셋은 음수관리기의 센서에서 기록된 데이터를 저장하고 있으며 그림 24 과 같이 음수관리기 장비 정보와 관련된 데이터와 장비 측정 데이터를 포함한다.

그림 24. ICT 데이터-음수관리기(한우)

#### ICT데이터 음수관리기(한우)

음수관리기의 측정일시, 측정값 ICT RAW 데이터를 확인할 수 있는 데이터셋

총 39,926 건이 검색되었습니다.

다운로드

비식별 데이터			장비 정보		장비 측정 데이터	
지역	규모	농장	제조사아이디	장비번호	측정일시	측정값(L)
	100두 미만	0022743	(주)애그리로보텍	371	2023-12-21	1.00
	100두 미만	0022743	(주)애그리로보텍	371	2023-12-21	4.00
	100두 미만	0022743	(주)애그리로보텍	371	2023-12-21	3.00
	100두 미만	0022743	(주)애그리로보텍	371	2023-12-21	4.00
	100두 미만	0022743	(주)애그리로보텍	371	2023-12-21	7.00
	100두 미만	0022743	(주)애그리로보텍	371	2023-12-21	5.00
	100두 미만	0022743	(주)애그리로보텍	371	2023-12-21	1.00
	100두 미만	0022743	(주)애그리로보텍	371	2023-12-21	4.00
	100두 미만	0022743	(주)애그리로보텍	371	2023-12-21	10.00
	100두 미만	0022743	(주)애그리로보텍	371	2023-12-21	12.00

1 2 3 4 5 6 7 8 9 10 > >>

ICAR 은 음수와 관련하여 명시적인 가이드라인을 제공하고 있지 않다. 하지만 이 음수 데이터는 소의 체중 측정하는데 고려되어야 하는 데이터이다. [ICAR-3.6] 가이드라인에 따르면 체중 측정은 마지막 사료 또는 물 섭취 후 수 시간 이내에 수행되어야 하는 제약이 있기 때문이다. 하지만 이 데이터 셋은 개별적인 소를 식별할 수 있는 정보가 없어서 다른 데이터 셋과 연결하여 분석하는데 무리가 있다.

표 16. ICAR 표준과 스마트팜 데이터마트 비교 표 - 음수관리기(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-3.6] <sup>29</sup> 구체적 항목 없음	지역 규모 농장 제조사아이디 장비번호 측정일시 측정값(L)	없음	판단불가

<sup>29</sup> ICAR Guidelines-Section 03-3.6 "Live animal assessment"

#### 14) ICT 데이터-포유기(한우)

ICT 데이터-포유기(한우) 데이터 셋은 포유기의 측정일시, 포유량, 포유온도 등의 로우(Raw) 데이터를 제공하여 어미 소의 포유 능력을 나타내기 위해 사용된다. 이 데이터 셋은 그림 25 와 같이 측정 장비 정보와 측정 데이터를 포함하고 있다.

그림 25. ICT 데이터-포유기(한우)

**ICT데이터 포유기(한우)**  
 포유기의 측정일시, 포유량, 포유온도 등 ICT RAW 데이터를 확인할 수 있는 데이터셋  
 총 1,716 건이 검색되었습니다. [다운로드]

비식별 데이터			장비 정보		장비 측정 데이터				
지역	규모	농장	제조사아이디	장비번호	측정일시	포유량	포유시작일시	포유종료일시	포유온도(°C)
4500000000	100두 미만	0022249	주식회사 다운	0	2022-05-30	1	2022-05-30 00:50:00	2022-05-30 00:51:00	
4500000000	100두 미만	0022249	주식회사 다운	0	2022-05-30	1	2022-05-30 09:13:00	2022-05-30 09:22:02	
	300~400	0021165	(주)동조	1	2022-01-04	6			0
	300~400	0021165	(주)동조	1	2022-01-04	5			0
	300~400	0021165	(주)동조	2	2022-01-04	3			0
	300~400	0021165	(주)동조	2	2022-01-04	4			0
	300~400	0021165	(주)동조	2	2022-01-04	4			0
	300~400	0021165	(주)동조	2	2022-01-04	3			0
	300~400	0021165	(주)동조	2	2022-01-04	3			0
	300~400	0021165	(주)동조	2	2022-01-04	5			0

1 2 3 4 5 6 7 8 9 10 > >>

ICT 데이터-포유기(한우) 데이터 셋과 관련 있는 ICAR 표준 가이드라인은 [ICAR-3.3.3.8] 이다. 이는 어미 소의 포유 측면에 대한 정보를 언급하고 있다. 그러나 ICAR 에서는 이에 대해 구체적인 항목에 대해서는 제공하고 있지 않다. 따라서 해당 데이터 셋이 ICAR 표준을 지키고 있는지 여부를 판단할 수 없다 (표 17 참조).

표 17. ICAR 표준과 스마트팜 데이터마트 비교 표 - 포유기(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-3.3.3.8] <sup>30</sup> 구체적 항목 없음	지역 규모 농장 제조사아이디 장비번호 측정일시 포유량 포유시작일시 포유종료일시 포유온도(°C)	없음	판단불가

<sup>30</sup> ICAR Guidelines-Section 03-3.3.3.8 "Mothering aptitude"

## 15) ICT 데이터 - TMR 급이기(한우)

ICT 데이터-TMR 급이기(한우) 데이터 셋은 소에게 주어진 사료의 비율을 설명하기 위해 측정일시, 급이량, 사료코드, 배합비율 등의 데이터를 저장하고 있다 (그림 26 참조). 이 데이터 셋과 관련된 ICAR 표준의 가이드라인은 [ICAR-3.8.2]에 해당한다. 다만, 표 18 에 따르면 [ICAR-3.8.2] 가이드라인에서 명시한 최소 요구 사항을 충족하고 있으나 현재 이 데이터 셋은 사료코드와 배합비율에 대한 완전한 정보를 포함하고 있지는 않으므로 보완이 필요하다.

그림 26. ICT 데이터-TMR 급이기(한우)

### ICT데이터 TMR급이기(한우)

TMR급이기의 측정일시, 급이량 등 ICT RAW 데이터를 확인할 수 있는 데이터셋

총 172,121 건이 검색되었습니다.

[다운로드](#)

비식별 데이터			장비 정보		장비 측정 데이터			
지역	규모	농장	제조사아이디	장비번호	측정일시	급이량(kg)	사료코드	배합비율(%)
3000000000	100두 미만	0021057	(주)길소프트	1	2023-10-29	362.00		
3000000000	100두 미만	0021057	(주)길소프트	1	2023-10-28	360.00		
3000000000	100두 미만	0021057	(주)길소프트	1	2023-10-27	350.00		
3000000000	100두 미만	0021057	(주)길소프트	1	2023-10-26	349.00		
3000000000	100두 미만	0021057	(주)길소프트	1	2023-10-25	354.00		
3000000000	100두 미만	0021057	(주)길소프트	1	2023-10-24	360.00		
3000000000	100두 미만	0021057	(주)길소프트	1	2023-10-23	360.00		
3000000000	100두 미만	0021057	(주)길소프트	1	2023-10-22	368.00		
3000000000	100두 미만	0021057	(주)길소프트	1	2023-10-21	369.00		
3000000000	100두 미만	0021057	(주)길소프트	1	2023-10-20	358.00		

1 2 3 4 5 6 7 8 9 10 > >>

표 18. ICAR 표준과 스마트팜 데이터마트 비교 표 - TMR 급이기(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR-3.8.2] <sup>31</sup> 동물 ID 기록일자 급이량	지역 규모 농장 제조사아이디 장비번호 측정일시 급이량(kg) 사료코드 배합비율(%)	없음	Yes

<sup>31</sup> ICAR Guidelines-Section 03-3.8.2 "Feed efficiency"

## 16) ICT 데이터-발정탐지기(한우)

ICT 데이터-발정탐지기(한우) 데이터 셋은 소의 발정 징후를 탐지하기 위해 측정일시, 운동량, 반추량, 체온, pH 등의 로우데이터를 기록하고 있다 (그림 27 참조). 표 19 에서 볼 수 있듯이 ICAR 은 발정상태와 관련된 기록에 대해 구체적인 가이드라인을 제공하지 않는다. 따라서 현재 스마트팜 데이터마트의 데이터와 비교할 수 있는 글로벌 표준이 없으므로 ICAR 표준 준수여부는 판단할 수 없다.

그림 27. ICT 데이터-발정탐지기(한우)

### ICT데이터 발정탐지기(한우)

발정탐지기의 측정일시, 운동량, 반추량, 체온, pH 등 ICT RAW 데이터를 확인할 수 있는 데이터셋

총 272,184 건이 검색되었습니다.

다운로드

비식별 데이터				장비 정보		장비 측정 데이터					
지역	규모	농장	개체번호	제조사아이디	장비번호	측정일시	운동량(회)	반추량(회)	체온(°C)	PH	발정상태값
4300000000	100~200	0022294	20171008010149	상육축산	121186664	2023-01-08					48
4300000000	100~200	0022294	20171008010149	상육축산	121186664	2023-01-01					68
4300000000	100~200	0022294	20191222010070	상육축산	146569204	2023-01-21					54
4300000000	100~200	0022294	20191222010070	상육축산	146569204	2023-01-12					47
4300000000	100~200	0022294	20201219010036	상육축산	159792738	2023-01-18					51
4300000000	100~200	0022294	20201219010036	상육축산	159792738	2023-01-15					54
4300000000	100~200	0020995	20130326010154	상육축산	304802827	2023-01-29	353.0	394.000			0
4300000000	100~200	0020995	20130326010154	상육축산	304802827	2023-01-28	325.0	579.000			0
4300000000	100~200	0020995	20130326010154	상육축산	304802827	2023-01-27	333.0	585.000			0
4300000000	100~200	0020995	20130326010154	상육축산	304802827	2023-01-26	353.0	502.000			0

1 2 3 4 5 6 7 8 9 10 > >>

표 19. ICAR 표준과 스마트팜 데이터마트 비교 표 - 발정탐지기(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
구체적 항목 없음	지역 규모 농장 개체번호 제조사아이디 장비번호 측정일시 운동량(회) 반추량(회) 체온(°C) pH 발정상태값 건강상태값 알람일시	없음	판단불가

## 17) ICT 데이터-환경제어기(한우)

ICT 데이터-환경제어기(한우) 데이터 셋은 농장 환경에 영향을 줄 수 있는 장비에 대해 제어와 관련된 데이터를 기록하여 저장하고 있다. 각 데이터 포인트는 특정 측정 날짜에 해당하는 환경제어기의 on/off 여부 또는 측정값을 저장하고 있으며 일 단위로 기록되고 있다. 이와 같은 데이터 셋은 그림 28 과 같다.

그림 28. ICT 데이터 환경제어기(한우)

### ICT데이터 환경제어기(한우))

환경제어기의 측정일자, 측정값 ICT RAW 데이터를 확인할 수 있는 데이터셋

총 17,663 건이 검색되었습니다.

다운로드

비식별 데이터			장비 정보			장비 측정 데이터					
지역	규모	농장	장비종류	제조사아이디	장비번호	측정일자	측정값01(% · on/ff)	측정값02	측정값03	측정값04	측정값05
3611000000	100~200	0021161	송풍팬	에프에스	1	2023-04-16	26	0			
3611000000	100~200	0021161	송풍팬	에프에스	1	2023-04-15	28	0			
3611000000	100~200	0021161	송풍팬	에프에스	1	2023-04-14	28	0			
3611000000	100~200	0021161	송풍팬	에프에스	1	2023-04-13	23	0			
3611000000	100~200	0021161	송풍팬	에프에스	1	2023-04-12	20	0			
3611000000	100~200	0021161	송풍팬	에프에스	1	2023-04-11	14	0			
3611000000	100~200	0021161	송풍팬	에프에스	1	2023-04-10	27	0			
3611000000	100~200	0021161	송풍팬	에프에스	1	2023-04-09	20	0			
3611000000	100~200	0021161	송풍팬	에프에스	1	2023-04-08	17	0			
3611000000	100~200	0021161	송풍팬	에프에스	1	2023-04-07	30	0			

1 2 3 4 5 6 7 8 9 10 > >>

ICT 데이터-환경제어기(한우) 데이터 셋은 조사해본 결과 현재 송풍팬의 측정만을 기록하고 있다. 환경제어기에는 송풍팬 이외에도 다양한 장비가 해당될 수 있기 때문에 이외의 다른 환경제어기에 대한 데이터도 측정되어야 한다. 이 데이터 셋은 환경관리기 데이터 셋과 유사하게 ICAR 에서 구체적인 표준을 제시하고 있지 않다. 따라서 ICAR 준수 여부에 대해서는 판단할 수 없다. (표 20 참조).

표 20. ICAR 표준과 스마트팜 데이터마트 비교 표 - 환경제어기(한우)

ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
구체적 항목 없음	지역 규모 농장 장비종류 제조사아이디 장비번호 측정일시 측정값 01(% on/ff), 측정값 02, 측정값 03, 측정값 04, 측정값 05	없음	판단불가

### 3.3 양돈 데이터

스마트팜 데이터마트에서 제공하고 있는 양돈 데이터 셋은 개체별 번식로우데이터, 연도별 번식통계, 월별 번식통계, 산차별 번식통계, ICT 데이터 급이기, ICT 데이터 음수기, ICT 데이터 환경정보, 포유모돈 급이번식 데이터까지 총 8 개의 데이터 셋을 제공하고 있다. 이는 그림 29의 양돈 데이터 셋 단계별 선택과정 중 데이터 구분을 선택하는 단계에서 확인할 수 있다. ICAR 지침은, 소고기 및 낙농 소의 경우와는 다르게, 양돈에만 적용되는 구체적인 지침을 명시하지 않고 있다. 하지만 [ICAR 1.1-1.2]<sup>32</sup> 가이드라인에 따르면 소 데이터에 적용되는 식별, 친자 및 번식의 일반 규칙에 대해서는 양돈 데이터에도 충분히 적용될 수 있다. 아래는 비교할 수 있는 ICAR 가이드라인이 있다면 이와 비교하여 표에 제시하였고, 그렇지 않은 경우에는 스마트팜 데이터마트에서 제공하는 데이터 셋에 대해서만 나열했다.

그림 29. 양돈 데이터 셋

양돈데이터셋 🖨️ 🔄

---

데이터 셋 검색

STEP 단계별로 선택하세요.

🔍

STEP 1 지역 선택	STEP 2 규모 선택	STEP 3 데이터 구분	STEP 4 상세구분	STEP 5 날짜선택
<input checked="" type="checkbox"/> 전국 <input type="checkbox"/> 강원특별자치도 <input type="checkbox"/> 경기도 <input type="checkbox"/> 경상남도 <input type="checkbox"/> 경상북도 <input type="checkbox"/> 광주광역시 <input type="checkbox"/> 대구광역시 <input type="checkbox"/> 대전광역시	<input checked="" type="checkbox"/> 전체 <input type="checkbox"/> 100두 미만 <input type="checkbox"/> 100~200 <input type="checkbox"/> 200~300 <input type="checkbox"/> 300~400 <input type="checkbox"/> 400~500 <input type="checkbox"/> 500~1000 <input type="checkbox"/> 1000두 이상	<input checked="" type="checkbox"/> 번식통계데이터 <input type="checkbox"/> 번식로우데이터 <input type="checkbox"/> ICT데이터 <input type="checkbox"/> 포유모돈 급이 번식 데이터	<input checked="" type="checkbox"/> 연도별 <input type="checkbox"/> 월별 <input type="checkbox"/> 산차별	시작 날짜 <input type="text" value="2000"/> 종료 날짜 <input type="text" value="2023"/>

<sup>32</sup> ICAR Guidelines-Section 03-1 "Introduction"

## 1) 개체별 번식 로우 데이터(양돈)

개체별 번식 로우 데이터(양돈) 데이터 셋은 개체별 번식이력(한우)과 매우 유사하며 각 어미 돼지의 번식 이력을 제공한다. 다만, 한우 데이터의 경우 임신 종료와 관련된 데이터가 포함되어 있지 않았으나 양돈 데이터에는 번식과 관련된 이벤트에 분만이나 임신의 종료를 포함하고 있다는 점이 차이점이다. 이 데이터 셋은 양돈 데이터 셋의 단계별 선택 과정을 통해 그림 30 과 같은 형태로 제공받을 수 있다.

그림 30. 개체별 번식로우(양돈)

### 개체별 번식로우(양돈)

농가정보를 비식별화 한 개체별 지표 데이터셋

총 162,812 건이 검색되었습니다.

다운로드

비식별 데이터				모든 번식 ROW데이터								
지역	규모	농장	개체번호	산차	교배일	임신사고일	분만일	총산	실산	생시총체	이유일	이유두수
경기도	100~200	0002615	2	0	20141030	61.0						
경기도	100~200	0002615	2	0	20150123		20150519	14	13		20150611	13
경기도	100~200	0002615	2	1	20150616		20151011	14	14		20151029	9
경기도	100~200	0002615	2	2	20151103		20160227	6	6		20160317	6
경기도	100~200	0002615	2	3	20160321		20160715	11	10		20160719	0
경기도	100~200	0002615	3	0	20150113		20150507	14	14		20150611	13
경기도	100~200	0002615	3	1	20150616	24.0						
경기도	100~200	0002615	3	1	20150801		20151123	15	14		20151224	11
경기도	100~200	0002615	3	2	20151229		20160421	11	11		20160512	10
경기도	100~200	0002615	3	3	20160517		20160909	14	14		20161006	10

1 2 3 4 5 6 7 8 9 10 > >>

관련된 ICAR 표준인 ICAR 1.1(식별 방법)과 ICAR 1.2(친자 기록 방법)에 따르면, 동물 개체를 식별할 수 있는 동물의 ID 는 반드시 기록되어야 하며 가축 무리 또는 농장 그룹 내에서 유일해야 한다. 또한, 교배된 동물의 ID 와 출생 당일 새끼의 성별 및 ID 가 기록되어야 한다. 하지만 스마트팜 데이터마트에서 제공하고 있는 데이터 셋에는 출산된 새끼돼지들의 신원이나 성별이 포함되지 않고 있다. 이는 곧 돼지의 혈통을 추적할 수 없게 하여, 결국 효과적인 데이터 분석을 할 수 없도록 한다 (표 21 참조). 또한 개체별 번식이력(한우)과 유사하게 현재 양돈 번식 데이터 셋에 교배, 분만, 임신일, 그리고 이유가 포함되어 있는데 이 데이터들을 다른 데이터

셋으로 분리할 필요가 있다. 이러한 과정을 통해 불필요하게 중복되거나 비어 있는 데이터를 효과적으로 없앨 수 있을 것이다.

표 21. ICAR 표준과 스마트팜 데이터마트 비교표 – 개체별 번식로우(양돈)

관련 ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
[ICAR 1-1 <sup>33</sup> , 1-2 <sup>34</sup> ] 동물 ID 동물 성별 개체 식별 방법 날짜 인공수정 수컷 ID	지역 규모 농장 개체제번호 산차 교배일 임신사고일 임신사고일 분만일 총산 실산 생시총체 이유일 이유두수	동물 성별 개체 식별 방법 인공수정 수컷 ID	No

<sup>33</sup> ICAR Guidelines-Section 1-1 "Methods of identification"

<sup>34</sup> ICAR Guidelines-Section 1-2 "Parentage recording methods"

## 2) 연도별 번식통계(양돈)

연도별 번식통계(양돈) 데이터 셋은 다양한 양돈 농장에 대한 연간 번식 통계를 집계하여 제공한다. 이는 국내 양돈 농업 업계에서 사용되는 미리 정의된 다양한 측정 항목들(ex. MSY, PSY, 모돈회전율, 평균비생산일수 등)을 포함하고 있다. 이는 그림 31 에서 확인할 수 있다.

그림 31. 연도별 번식통계(양돈)

### 연도별 번식통계(양돈)

농가정보를 비식별화 한 연도별 지표 데이터셋

총 1,014 건이 검색되었습니다.

다운로드

비식별 데이터				연도별 통계 데이						
년도	지역	규모	농장	상시모돈수(두)	MSY(두)	PSY(두)	모돈회전율(회전)	평균비생산일수(일)	평균총산(두)	평균실산(두)
2019		100~200	PIGGO_156	143.70	17.2	19.7	2.13	74.60	13	11
2018		100~200	PIGGO_156	146.10	19.2	21.7	2.18	69.40	13	12
2021		500~1000	0021749	636.90		29.2	2.52	25.30	14	12
2021		400~500	0021437	453.60		30.9	2.29	43.90	16	14
2021	경상남도	200~300	0021398	251.50		28.6	2.48	30.60	14	13
2021	경상남도	500~1000	0020254	762.40		26.2	2.30	38.50	15	12
2021	경기도	300~400	0021289	352.50		22.5	2.24	51.00	14	12
2021		200~300	0021947	222.40		26.3	2.33	37.30	13	12
2021	제주특별자치도	100~200	0021283	155.60		23.1	2.27	46.70	13	12
2021	제주특별자치도	100~200	0021284	165.10		19.3	2.30	47.80	10	9

1 2 3 4 5 6 7 8 9 10 > >>

조사한 바에 따르면, 이 데이터 셋에는 많은 수의 데이터가 비어 있는 것을 확인할 수 있었으나 해당 데이터가 비어 있는 이유는 불분명하다. 이에 대해 이전의 한우 데이터 셋과 유사하게 데이터 기록자를 포함하고 데이터에 대한 스키마를 지정하여 비어 있는 값을 가지면 안되는 데이터에 대해 강제로 지정하여 데이터의 품질을 높일 수 있다. 또한 데이터의 의미에 대한 모호함을 없애기 위해 데이터설명서의 첨부 역시 필요하다. 그리고, 표 22 에 보여지듯 연도별 번식통계(양돈) 데이터 셋과 대응되는 ICAR 표준이 없기 때문에 ICAR 준수여부는 판단할 수 없다.

표 22. ICAR 표준과 스마트팜 데이터마트 비교표 - 연도별 번식통계(양돈)

관련 ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
구체적 항목 없음	년도 지역 규모 농장 상시모돈수 MSY(두) PSY(두) 모돈회전율(회전) 평균비생산일수(일) 평균총산(두) 평균실산(두) 평균이유(두) 분만율(%) 초교배일령(일) 임신기간(일) 포유기간(일) 재귀발정일(일)	없음	판단불가

### 3) 월별 번식통계(양돈)

월별 번식통계(양돈) 데이터 셋은 다양한 양돈 농장의 월별 통계를 포함하고 있다. 현재 측정되어 저장되고 있는 통계 데이터에는 분만율, 평균이유, 임신사고(불규칙), 7 일내재귀율, 임신사고(도태), 모돈회전율, 임신사고(공태), 임신사고(2 차), 임신사고(조기), 임신사고(불임), 임신사고(폐사), 평균총산, 총산합계, 사산합계, 임신사고(유산), 임신사고(지연), 임신사고(1 차), 미라합계, 그리고 평균실산이 있으며 이는 그림 32 에서 확인 가능하다.

그림 32. 월별 번식통계(양돈)

#### 월별 번식통계(양돈)

농가정보를 비식별화 한 월별 지표 데이터셋

총 14,344 건이 검색되었습니다.

다운로드

비식별 데이터				월별 통계 데이터												
년도	지역	규모	농장	데이터구분	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
2012		100두 미만	0020069	임신사고(조기)	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2012	충청북도	100~200	0020069	임신사고(조기)	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2012	충청북도	100~200	0020069	임신사고(조기)	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2012		100두 미만	0020069	임신사고(폐사)	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
2012	충청북도	100~200	0020069	임신사고(폐사)	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
2012	충청북도	100~200	0020069	임신사고(폐사)	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
2012		100두 미만	0020069	임신사고(1차)	2.00	0.00	3.00	0.00	3.00	0.00	0.00	0.00	0.00	0.00	3.00	2.00
2012	충청북도	100~200	0020069	임신사고(1차)	2.00	0.00	3.00	0.00	3.00	0.00	0.00	0.00	0.00	0.00	3.00	2.00
2012	충청북도	100~200	0020069	임신사고(1차)	2.00	0.00	3.00	0.00	3.00	0.00	0.00	0.00	0.00	0.00	3.00	2.00
2012		100두 미만	0020069	임신사고(불규칙)	1.00	4.00	2.00	3.00	0.00	3.00	3.00	12.00	6.00	15.00	5.00	1.00

1 2 3 4 5 6 7 8 9 10 > >>

현재 월별 번식통계(양돈) 데이터 셋의 경우 1 월부터 12 월까지 시간을 나타내는 데이터를 데이터 셋의 열로 구성하고 데이터 유형을 행으로 하여 일반적인 데이터 구조와 반대로 구성하였다. 날짜와 관련된 데이터를 하나의 열로 만들고, 각 데이터 유형을 별도의 열 또는 데이터 셋으로 분리하는 것이 권장된다. 이렇게 하면 불필요하게 비어 있는 데이터를 없앨 수 있다. 또 다른 문제는 번식 데이터 셋에 로우 데이터를 기록하지 않고 있다는 점이다. 이 데이터 셋에 저장된 통계는 개체별 번식 로우(양돈) 테이블에서 수행된 계산과 정확히 일치하고 있지 않다. 예를 들어, 2020 년 1 월 농장 0020069 에 대한 평균실산 값은 11.1 이지만, 로우 테이블에서 같은 값을 계산할 때 값은 11.125 이다. 마찬가지로, 2020 년 2 월에 통계 데이터

셋은 12.6, 로우 데이터 셋은 12.5 로 일관된 정보를 제공하고 있지 않는다. 추가적으로, 같은 농장에 대해 규모가 다르게 적혀 있는 것은 추가적인 조사가 필요하다. 표 23 과 같이 월별 번식통계(양돈) 데이터 셋과 대응되는 ICAR 표준이 없기 때문에 ICAR 준수여부는 판단할 수 없다.

표 23. ICAR 표준과 스마트팜 데이터마트 비교표 - 월별 번식통계(양돈)

관련 ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
구체적 항목 없음	년도 지역 규모 농장 데이터구분 1~12 월	없음	판단불가

#### 4) 산차별 번식통계(양돈)

산차별 번식통계(양돈) 데이터 셋에는 각 농장별로 산차(첫 번째, 두 번째..., 임신)에 따라 집계된 16 개의 측정값(분만율, 평균실산, 미라율, 평균실산(재귀 3 일), 평균총산(재귀 3 일), 평균실산(재귀 4 일), 평균총산(재귀 4 일), 평균실산(재귀 5 일), 평균총산(재귀 5 일), 평균실산(재귀 7 일), 평균총산(재귀 7 일), RT8LB, RT8TB, 사산율, 평균총산, 평균이유, 평균실산(재귀 6 일), 평균총산(재귀 6 일))이 포함되어 있다. 이는 그림 33 에서 확인 가능하다.

그림 33. 산차별 번식통계(양돈)

#### 산차별 번식통계(양돈)

농가정보를 비식별화 한 산차별 지표 데이터셋

총 2,197 건이 검색되었습니다.

[다운로드](#)

비식별 데이터			산차별 통계 데이터								
지역	규모	농장	데이터구분	1산	2산	3산	4산	5산	6산	7산	8산이상
	100두 미만	0020069	분만율	60.30	53.50	86.70	89.70	83.00	77.50	64.70	0.00
	100두 미만	0020069	평균실산	9.30	9.30	10.70	11.70	10.40	10.40	11.60	
	100두 미만	0020069	미라율	3.30	2.40	0.60	2.20	1.00	2.10	2.20	
	100두 미만	0020069	평균실산(재귀3일)			2.00	16.00				
	100두 미만	0020069	평균총산(재귀3일)			2.00	16.00				
	100두 미만	0020069	평균실산(재귀4일)		5.50	12.70	12.50	9.50	10.30	12.00	
	100두 미만	0020069	평균총산(재귀4일)		6.00	14.10	14.10	12.50	13.00	13.50	
	100두 미만	0020069	평균실산(재귀5일)		9.00	9.70	12.10	11.20	10.80	11.30	
	100두 미만	0020069	평균총산(재귀5일)		11.00	11.60	13.40	12.80	12.30	12.00	
	100두 미만	0020069	평균실산(재귀7일)			11.00	8.50			12.00	

1 2 3 4 5 6 7 8 9 10 > >>

이 데이터 셋의 경우 8 개의 산차를 열로 제시하고 8 번째 이후 산차의 측정값을 하나의 열에 공통적으로 기록하고 있는데, 이보다는 산차를 하나의 열로 구성하는 것이 나아보인다. 데이터구분 열은 각 측정값의 단위가 다르기 때문에 별도의 열 또는 데이터 셋으로 만들어져야 한다. 또한 표 24 와 같이 산차별 번식통계(양돈) 데이터 셋과 대응되는 ICAR 표준이 없기 때문에 ICAR 준수여부는 판단할 수 없다.

표 24. ICAR 표준과 스마트팜 데이터마트 비교표 - 산차별 번식통계(양돈)

관련 ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
구체적 항목 없음	지역 규모 농장 데이터구분 1~7 산 8 산이상	없음	판단불가

## 5) ICT 데이터-급이기(양돈)

ICT 데이터 급이기(양돈) 데이터에는 자동 급이기를 사용하여 양돈에게 주어진 사료의 양과 관련된 정보가 포함되어 있다 (그림 34 참조). 이 데이터 셋에 대해서는 별다른 문제가 없다. 그러나 1) 지역과 개체번호에 대한 빈 값이 존재해서 동물을 식별하기 어렵게 만드는 문제와 2) 현 데이터 셋의 목적과 적합하지 않은 정보인 교배일과 분만일을 포함하고 있는 문제는 해결되어야 한다. 또한 표 25 와 같이 ICT 데이터 급이기(양돈) 데이터 셋과 대응되는 ICAR 표준이 없기 때문에 ICAR 준수여부는 판단할 수 없다.

그림 34. ICT 데이터-급이기(양돈)

### ICT데이터 급이기(양돈)

모든의 급이 섭취량을 확인할 수 있는 급이기 데이터셋

총 905,236 건이 검색되었습니다.

다운로드

비식별 데이터				장비 정보		장비 측정 데이터				
지역	규모	농장	개체번호	장비구분	장비번호	섭취일	설정량(kg)	섭취량(kg)	교배일	분만일
	100두 미만	0020091		포유모돈 급이기	61	2020-01-17	7.80	0.00	2019-09-09	2020-01-05
	100두 미만	0020091		포유모돈 급이기	67	2020-01-17	8.40	0.00	2019-09-09	2020-01-04
	100두 미만	0020091		포유모돈 급이기	37	2020-02-19	2.40	2.40	2019-10-20	2020-02-16
	100두 미만	0020091		포유모돈 급이기	37	2020-02-20	3.00	3.00	2019-10-20	2020-02-16
	100두 미만	0020091		포유모돈 급이기	24	2020-02-19	1.80	1.80	2019-10-22	2020-02-17
	100두 미만	0020091		포유모돈 급이기	24	2020-02-20	2.40	2.40	2019-10-22	2020-02-17
	100두 미만	0020091		포유모돈 급이기	30	2020-02-19	3.60	3.60	2019-10-22	2020-02-14
	100두 미만	0020091		포유모돈 급이기	30	2020-02-20	4.20	4.20	2019-10-22	2020-02-14
	100두 미만	0020091		포유모돈 급이기	43	2020-02-19	6.00	6.00	2019-10-14	2020-02-10
	100두 미만	0020091		포유모돈 급이기	43	2020-02-20	6.60	6.60	2019-10-14	2020-02-10

1 2 3 4 5 6 7 8 9 10 > >>

표 25. ICAR 표준과 스마트팜 데이터마트 비교표 - 급이기(양돈)

관련 ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
구체적 항목 없음	지역 규모 농장 개체제번호 장비구분 장비번호 섭취일 설정량(kg) 섭취량(kg) 교배일 분만일	없음	판단불가

## 6) ICT 데이터-음수기(양돈)

ICT 데이터-음수기(양돈) 데이터 셋과 유사하게 ICT 데이터-음수기(양돈) 데이터 셋에는 음수기를 사용하여 주어진 물의 양에 대한 정보를 포함하고 있다 (그림 35 참조). 그러나 개별 어미 돼지의 식별 정보와 그에 해당하는 물 섭취량에 대한 정보를 포함하고 있지 않다. 이 데이터 셋에 대해 특별한 문제를 찾을 수는 없지만, 빈 데이터에 대한 처리를 위해 데이터 스키마를 지정하여 유효하지 않은 값이 데이터 셋에 저장되는 상황을 막고 데이터 셋의 유용성을 높일 것을 권장한다. 또한 표 26 과 같이 ICT 데이터-음수기(양돈) 데이터 셋과 대응되는 ICAR 표준이 없기 때문에 ICAR 준수여부는 판단할 수 없다.

그림 35. ICT 데이터-음수기(양돈)

### ICT데이터 음수기(양돈)

모든의 음수량을 확인할 수 있는 음수기 데이터셋

총 215,621 건이 검색되었습니다.

다운로드

비식별 데이터			장비 정보		장비 측정 데이터	
지역	규모	농장	장비구분	장비번호	섭취일	섭취량(kg)
충청남도		0020090	음수기	1	2023-11-14	23
전라북도		0022258	음수기	01-02	2023-08-31	1990
전라북도		0022258	음수기	01-01	2023-08-31	1760
전라북도		0022258	음수기	01-01	2021-12-13	1450
전라북도		0022258	음수기	01-02	2021-12-13	1386
전라북도		0022258	음수기	01-01	2021-12-13	1450
전라북도		0022258	음수기	01-01	2021-12-13	1450
전라북도		0022258	음수기	01-01	2022-01-31	0
전라북도		0022258	음수기	01-02	2022-01-31	0
전라북도		0022258	음수기	01-02	2021-12-13	1386

1 2 3 4 5 6 7 8 9 10 > »

표 26. ICAR 표준과 스마트팜 데이터마트 비교표 - 음수기(양돈)

관련 ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
구체적 항목 없음	지역 규모 농장 장비구분 장비번호 섭취일 섭취량(kg)	없음	판단불가

## 7) ICT 데이터-환경정보(양돈)

ICT 데이터-환경정보(양돈) 데이터 셋에는 6 개의 다른 측정 장치 중 하나에서 측정한 값이 포함되어 있다. 측정값에는 온도, 습도, 복합악취, 이산화황, 암모니아, CO2 가 있다. 해당 데이터 셋에는 두 가지 주요 문제가 있다. 첫째, 장비구분에 빈 데이터가 있어 관련된 데이터에서 어떠한 결론도 도출할 수 없도록 한다. 둘째, 측정값 옆에 저장된 데이터가 다른 단위를 가지고 있기 때문에 이들을 서로 다른 열 또는 다른 데이터 셋으로 분리할 것을 권장한다. 또한 표 27 과 같이 ICT 데이터-환경정보(양돈) 데이터 셋과 대응되는 ICAR 표준이 없기 때문에 ICAR 준수여부는 판단할 수 없다.

그림 36. ICT 데이터-환경정보(양돈)

### ICT데이터 환경정보(양돈)

농장에 설치된 환경정보(온도, 습도, CO2, 암모니아 등)를 확인할 수 있는 데이터셋

총 283,275 건이 검색되었습니다.

다운로드

비식별 데이터			장비 정보		장비 측정 데이터	
지역	규모	농장	장비구분	장비번호	측정일	측정값(온습도 :도 · CO2 : PPM)
4400000000		0020090	습도	1	2023-07-11	54.00
4400000000		0020090	온도	1	2023-07-12	26.00
4400000000		0020090	습도	1	2023-07-12	73.00
4400000000		0020090	온도	1	2023-07-13	25.00
4400000000		0020090	습도	1	2023-07-13	76.00
4400000000		0020090	온도	1	2023-07-14	25.00
4400000000		0020090	습도	1	2023-07-14	79.00
4400000000		0020090	습도	1	2023-07-17	74.00
4400000000		0020090	온도	1	2023-07-17	25.00
4400000000		0020090	온도	1	2023-07-18	25.00

1 2 3 4 5 6 7 8 9 10 > >>

표 27. ICAR 표준과 스마트팜 데이터마트 비교표 - 환경정보(양돈)

관련 ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
구체적 항목 없음	지역 규모 농장 장비구분 장비번호 측정일 측정값 (온습도 : 도, CO2 : PPM)	없음	판단불가

## 8) 포유모돈 급이번식 데이터(양돈)

포유모돈 급이번식 데이터(양돈) 데이터 셋에는 특정 모돈에게 주어진 사료뿐만 아니라 어미 돼지의 번식 정보도 포함되어 있다 (그림 37 참조). 번식 및 급이 데이터를 두 개의 데이터 셋으로 분리할 것을 권장한다. 또한 데이터 셋의 모든 열에 대한 명확한 설명을 강력히 권장한다. 왜냐하면 각 데이터 열이 무엇을 의미하는지 분명하지 않아 데이터 사용자로 하여금 혼돈을 일으킬 수 있기 때문이다. 또한 표 28 과 같이 포유모돈 급이번식 데이터(양돈) 데이터 셋과 대응되는 ICAR 표준이 없기 때문에 ICAR 준수여부는 판단할 수 없다.

그림 37. 포유모돈 급이번식 데이터(양돈)

### 포유모돈 급이번식 데이터(양돈)

농장에 설치된 급이기 장비 정보와 수기 입력한 번식정보를 융합한 데이터셋

총 256,162 건이 검색되었습니다.

다운로드

비식별 데이터			급이 장비 정보										
지역	규모	농장	품종	산차	급이횟수	급이설정량	섭취량	급이일자	분만일로부터의 기간	분만일	총산	실산	분만시
	1000두 이상	PF_0020082	F1	7	3	7.50	7.50	2022-01-18	13	2022-01-05	11	10	19.0
	1000두 이상	PF_0020082	F1	7	3	8.00	8.00	2022-01-19	14	2022-01-05	11	10	19.0
	1000두 이상	PF_0020082	F1	7	3	8.00	8.00	2022-01-20	15	2022-01-05	11	10	19.0
	1000두 이상	PF_0020082	F1	7	3	8.00	8.00	2022-01-21	16	2022-01-05	11	10	19.0
	1000두 이상	PF_0020082	F1	7	3	8.00	8.00	2022-01-22	17	2022-01-05	11	10	19.0
	1000두 이상	PF_0020082	F1	7	3	8.50	8.50	2022-01-24	19	2022-01-05	11	10	19.0
	1000두 이상	PF_0020082	F1	7	3	8.50	8.50	2022-01-25	20	2022-01-05	11	10	19.0
	1000두 이상	PF_0020082	F1	7	3	8.50	6.80	2022-01-26	21	2022-01-05	11	10	19.0
	1000두 이상	PF_0020082	F1	7	3	8.50	0.00	2022-01-28	23	2022-01-05	11	10	19.0
	1000두 이상	PF_0020082	F1	8	2	3.80	0.00	2022-05-17	-10	2022-05-27	14	13	17.0

1 2 3 4 5 6 7 8 9 10 > >>

표 28. ICAR 표준과 스마트팜 데이터마트 비교표 - 포유모돈 급이번식 데이터(양돈)

관련 ICAR 표준	SmartFarm 제공 데이터	누락 데이터	ICAR 준수 여부
구체적 항목 없음	지역 규모 농장 품종 산차 급이횟수 급이설정량 섭취량 급이일자 분만일로부터의 기간 분만일 총산 실산 분만시 BCS 이유일 이유두수 이유총체 이유평체 재귀발정일	없음	판단불가

## 4. 스마트팜 데이터마트 데이터 연결성 문제점 분석

이 섹션에서는 스마트팜 데이터마트의 데이터 셋이 어떻게 연결되는가에 대해 살펴보고자 한다. 데이터 셋은 크게 두 가지 유형인 로우 데이터 셋과 집계된 데이터 셋으로 나눌 수 있다. 로우 데이터 셋은 기본 데이터이며, 집계된 데이터 셋은 이러한 로우 데이터 셋에서 편집된 데이터를 뜻한다. 집계된 데이터 셋의 경우, 오리지널 로우 데이터로 다시 되돌아갈 수 있는지 그리고 정확성이 로우 데이터와 일치하는지 확인해야 한다. 또한, 몇개의 주된 시나리오를 기반으로, 스마트팜 데이터마트 데이터 셋을 서로 연결하여 보다 상세하거나 포괄적인 분석을 제공할 수 있는지 살펴보고자 한다.

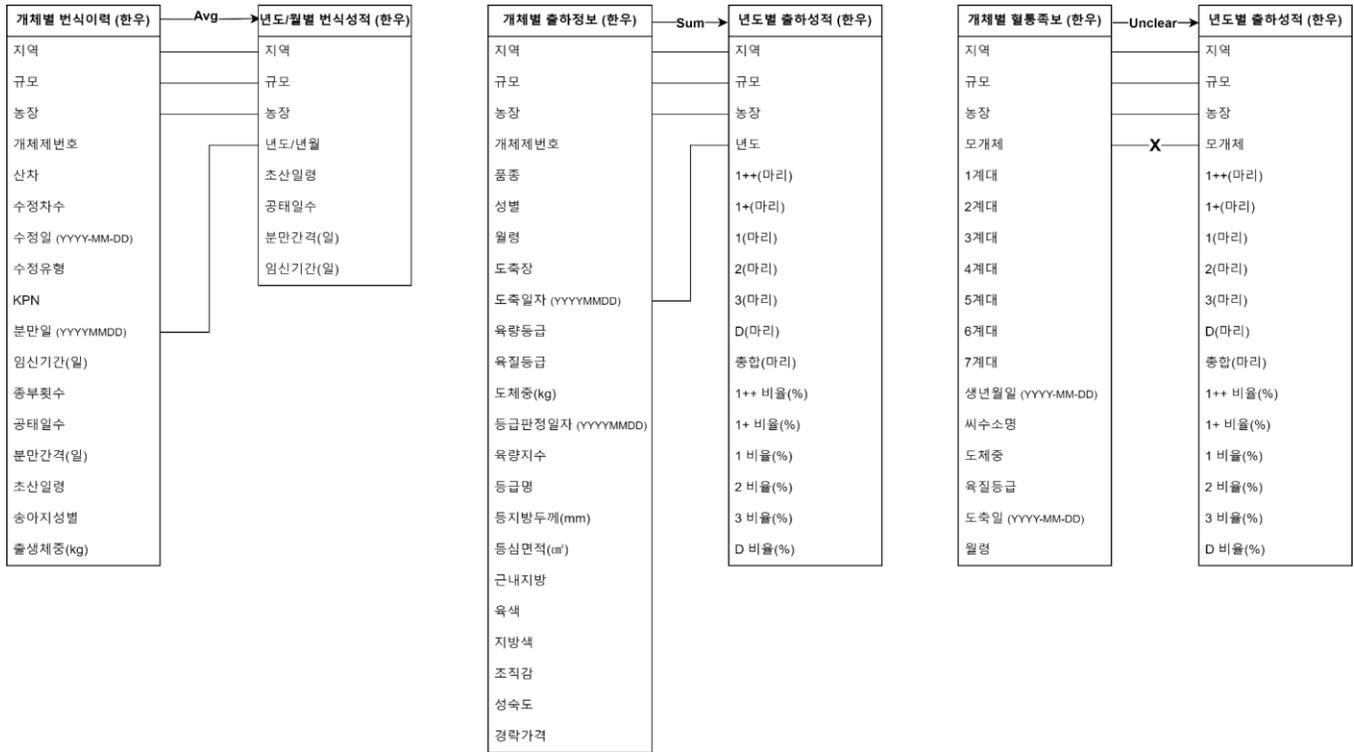
### 4.1 집계 데이터 연결성 분석

스마트팜 데이터마트에서 제공하고 있는 집계 데이터 셋은 일정 주기 또는 목적에 따라 로우 데이터를 가공하고 집계하여 제공하는 데이터 셋이다. 다음 섹션에서는 이와 같은 집계 데이터 셋에 대해 한우 및 양돈 집계 데이터를 중심으로 이들 데이터 셋 간의 연결성을 분석한 내용을 제시한다.

#### 4.1.1 한우 집계 데이터 셋 분석

한우 데이터 셋에 대해 스마트팜 데이터마트는 네 가지 종류의 집계 데이터를 제공하고 있다. 연도별 번식성적(한우), 월별 번식성적(한우), 연도별 출하성적(한우), 그리고 개체별 혈통축보(한우)가 해당한다. 그림 38 에 나와 있는 것처럼, 연도별 번식성적(한우)과 월별 번식성적(한우)은 개체별 번식이력 테이블에서 지역, 규모, 농장 및 분만일의 연도/연월을 상호참조 키(Key)로 결합하여 집계된다. 개체별 번식이력 데이터 셋의 각 값은 그룹화되어 평균을 내어 앞서 언급한 집계된 데이터를 생성해야 한다. 마찬가지로, 연도별 출하성적(한우) 데이터는 출하정보와 지역, 규모, 농장, 년도를 결합한 키(Key)를 사용하여 연결된다. 이 데이터도 병합하기 전에 키(Key)에 따라 그룹화되고 합산되어야 한다.

그림 38. 한우 데이터의 집계 데이터의 상호 연결성 (X: 연결 불가)



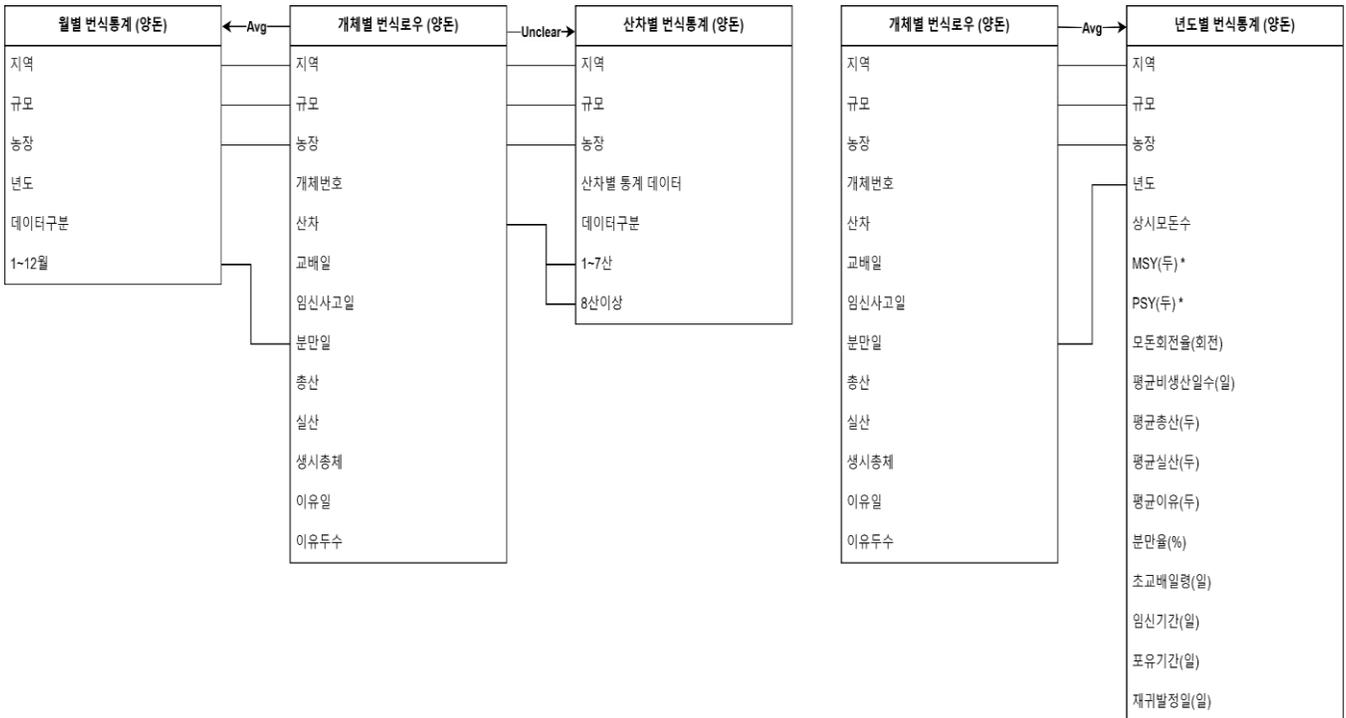
이에 대해서 분석해 본 결과 집계된 데이터 값은 해당하는 로우 데이터를 집계한 결과와 근접하게 일치하고 있다. 하지만 개체별 혈통족보(한우) 데이터는 이를 만족하고 있지 않다. 개체별 혈통족보(한우) 데이터 셋에서는 모개체에 대한 고유 식별자가 없어 연도별 출하성적 데이터를 추적할 수 없기 때문이다. 현재 개체별 혈통족보(한우) 데이터 셋에는 모개체의 성별만 명시되어 있다.

#### 4.1.2 양돈 집계 데이터 셋 분석

양돈 데이터에 대해 스마트팜 데이터마트는 월별 번식통계(양돈), 연도별 번식통계(양돈), 산차별 번식통계(양돈)의 3 종류의 집계 데이터를 제공한다. 그림 41 에서 볼 수 있듯이, 월별 번식통계(양돈)와 연도별 번식통계(양돈)는 개체별 번식로우(양돈) 데이터 셋에서 지역, 규모, 농장, 그리고 분만일의 년/월을 키로 사용하여 집계된다. 개체별 번식로우(양돈) 데이터 셋의 각 값은 그룹화되어 평균을 내어 앞서 언급한 집계 데이터를 생성해야 한다. 비슷하게, 산차별

번식통계(양돈) 데이터를 생성하기 위해선, 개체별 번식로우(양돈) 데이터 셋이 지역, 규모, 농장, 그리고 산차에 따라 그룹화되어야 한다.

그림 39. 양돈 데이터의 집계된 데이터의 상호 연결성 (\*: 추적할 수 없는 추가 정보)



분석 결과 연도별 번식통계(양돈) 데이터 셋의 집계 데이터는 가장 가까운 정수로 반올림된다. 예를 들어, 2019년 농장 0020069의 이유두수는 9.412로 계산되어 연도별 번식통계(양돈) 데이터 셋의 평균이유(두) 필드에서 9로 반올림된다. 그러나 월별 번식통계(양돈)와 산차별 번식통계(양돈)의 일부 값이 일치하지 않는 것을 발견했다. 예를 들어, 월별 번식통계(양돈)에서 2020년 1월 농장 0020069의 평균이유 값은 9.4인 반면, 동일한 값이 로우 테이블에서 계산된 경우 값은 9.875이다. 마찬가지로, 2020년 2월의 값은 통계 데이터 셋과 로우 데이터 셋에서 각각 10.0과 11.285714이다. 또한, 산차별 번식통계(양돈)의 평균실산은 농장, 규모 및 산차로 그룹화된 개체별 번식로우(양돈)의 평균실산과 일치하지 않는다. 강조해야 할 또 다른 문제는 연도별 번식통계(양돈) 데이터 셋이 MSY와 PSY 같이 개체별 번식로우(양돈)로 추적할 수 없는 추가 정보를 포함한다.

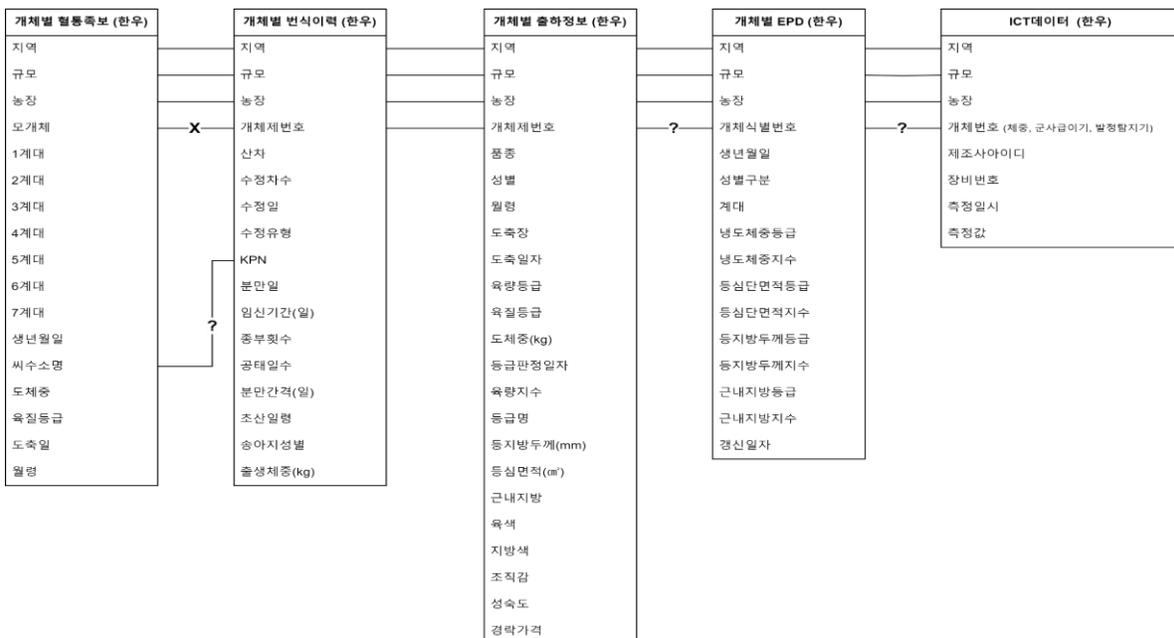
## 4.2 로우 데이터 연결성 분석

스마트팜 데이터마트에서 제공하고 있는 로우 데이터 셋은 특정 목적에 맞게 측정된 데이터를 가공하지 않은 상태로 그 정보를 온전히 저장하고 있는 데이터 셋이다. 다음 섹션은 이와 같은 로우 데이터 셋에 대해 한우 및 양돈 로우 데이터를 중심으로 이들 데이터 셋 간의 연결성을 분석한 내용을 제시한다.

### 4.2.1 한우 로우 데이터 셋 분석

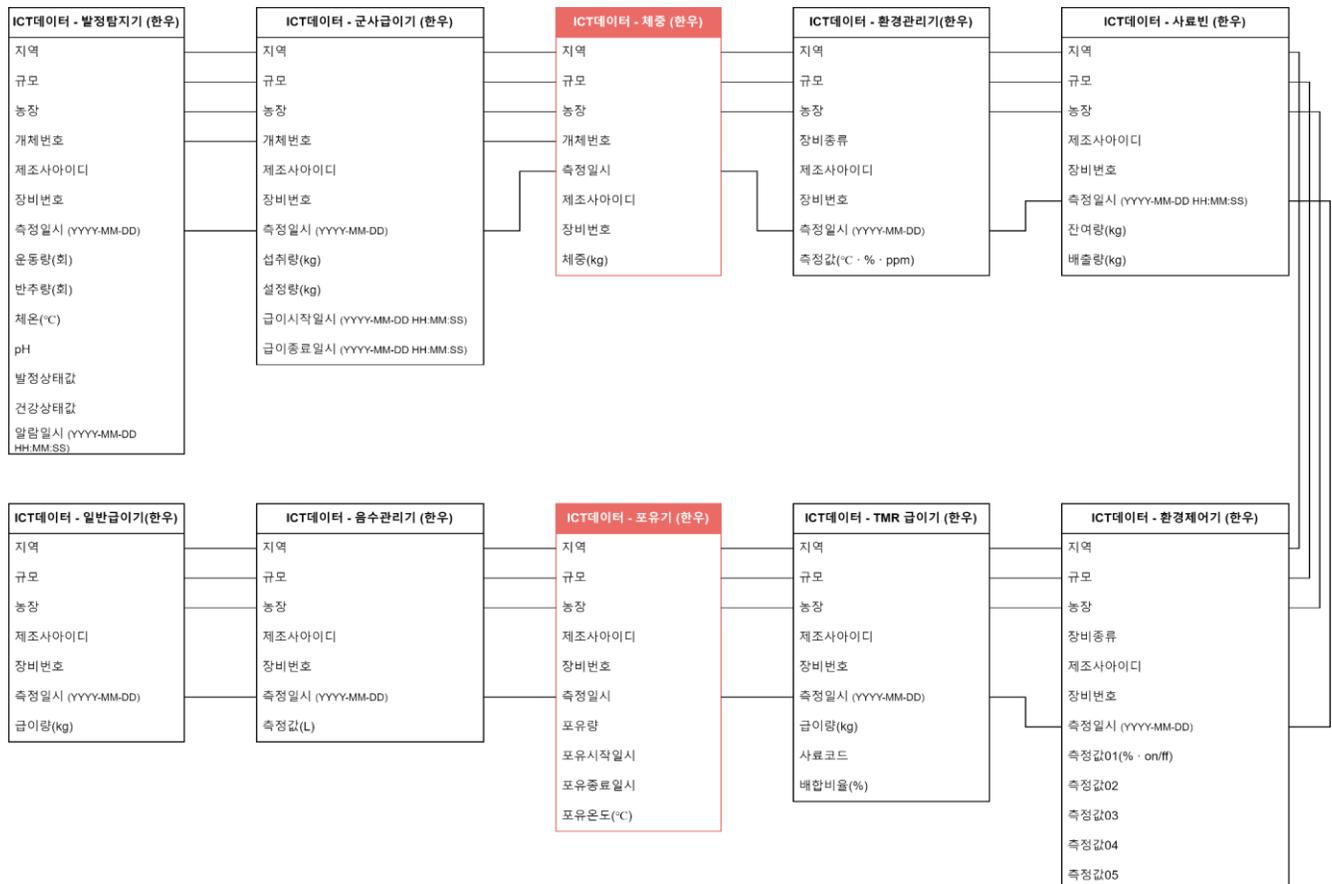
한우 로우 데이터 셋에는 개채별 번식이력(한우), 개채별 출하정보(한우), 개채별 혈통족보(한우), 그리고 개채별 EPD(한우)와 같은 ICT 데이터와 비-ICT 데이터가 포함된다. 대부분의 비-ICT 데이터 셋은 지역, 규모, 농장, 그리고 개체제번호의 조합을 사용하여 상호 연결될 수 있다 (그림 39 참조). 그러나 개채별 EPD(한우) 데이터 셋은 개체제번호 대신 개체식별번호를 사용하고 있다. 비-ICT 데이터와 ICT 데이터 셋을 연결하기 위해 개체번호 대신 현재 개체제번호가 사용된다. 이와 같은 데이터 셋 연결은 각 동물의 번식, 출하, EPD 및 ICT 데이터를 포괄적으로 추적할 수 있도록 한다.

그림 40. 한우 데이터 셋의 상호 연결성 (X: 연결 불가, ?: 열 이름이 다르더라도 연결 가능)



예외적으로, 개채별 혈통축보(한우) 데이터 셋은 모개체 ID 정보의 부재로 인해 직접적인 연결을 할 수 없다. 그러나 개채별 혈통축보(한우)의 씨수소명 열과 개채별 번식이력(한우)의 KPN 열을 일치시켜 연결할 수 있다. 일부 씨수소명 ID 가 번식이력의 KPN 과 일치하지만, 서로 다른 열 이름과 명확한 정의 부재로 인해 명확한 결론을 내리기 전에 추가 검증이 필요하다. 이러한 연결을 분석할 때 간접적인 연결로 인한 주의가 필요하다. 또한 그림 41 에서 볼 수 있듯이, 모든 ICT 데이터는 지역, 규모, 농장, 그리고 측정일시를 사용하여 상호 연결된다. 체중, 군사급이기, 그리고 발정탐지기와 같은 개별 지표 데이터 셋의 경우, 개체번호를 매칭하는 것이 필요하다. 그러나 이러한 데이터 셋을 매칭하는 것은 측정일시 세부 정보의 차이로 어려울 수 있다. 예를 들어, 사료빈은 날짜만 기록하는 다른 데이터와 달리 시간까지 기록한다. 이러한 불일치로 인해 특정 사료빈 작업을 환경제어기와 같은 데이터 셋과 정확하게 일치시키는 것이 복잡해진다. 또한, 체중과 포유기 ICT 데이터가 기록되지 않은 것으로 나타나며 (아래 그림 41 에서 빨간색으로 표시됨), 이로 인해 이러한 데이터 셋에 대한 분석이 제한되게 된다.

그림 41. 한우 데이터의 ICT 데이터 셋의 상호 연결성 (빨간색: 내용이 빈 데이터 셋)

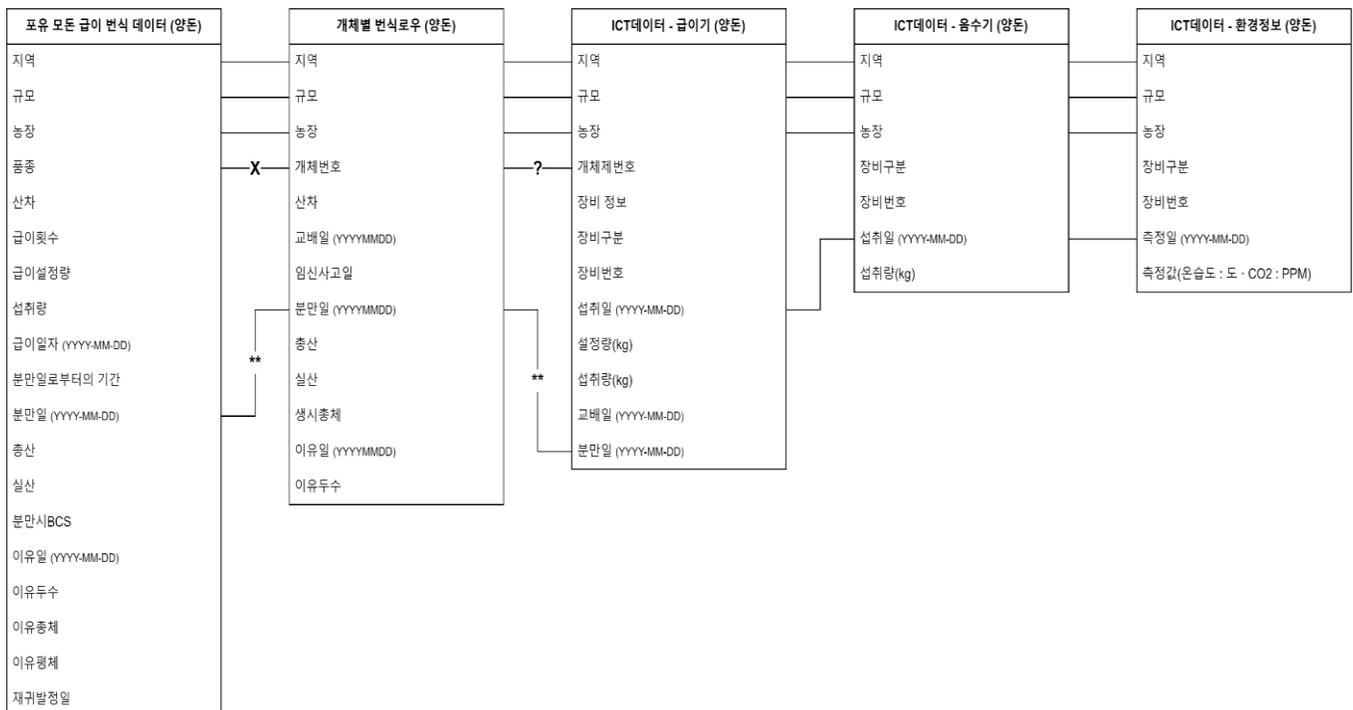


## 4.2.2 양돈 로우 데이터 셋 분석

한우 데이터와 마찬가지로, 양돈 데이터에는 ICT(급이기, 음수기, 환경정보) 및 비-ICT(개체별 번식로우 및 포유 모돈 급이 번식) 데이터가 포함된다. 하지만 포유 모돈 급이 번식(양돈) 데이터 셋에는 개체번호가 없어 개체별 번식로우(양돈) 데이터 셋과 연결할 수 없다. 다만, 개체별 번식로우(양돈)는 지역, 규모, 농장, 개체번호 및 분만일의 조합을 사용하여 ICT 데이터 - 급이기(양돈)와 연결될 수 있다. 그러나 두 데이터 셋의 분만일 값이 서로 다른 형식을 가지고 있기 때문에 두 데이터 셋 사이의 분만일 값을 연결하기 위해서는 추가 처리가 필요하다 (개체별 번식로우는 YYYYMMDD 형식이고 ICT 데이터 - 급이기는 YYYY-MM-DD 형식이다). 또한, ICT 데이터 - 급이기(양돈)에서 개체번호 대신 개체제번호가 사용된다 (그림 42 참조). ICT 데이터에 대해서는 지역, 규모, 농장, 그리고 섭취/측정일을 사용하여 상호 연결할 수 있다. 그러나 ICT 데이터 - 음수기(양돈)에서 동일한 섭취일, 장비구분 및 장비번호에 대해 여러 레코드가 있을 수 있으며, 이는 다른 데이터 셋과의 연결을 복잡하게 만든다. 이 문제는 추가 데이터 처리 또는 보다 상세한 섭취일 기록이 필요할 수 있다.

그림 42. 양돈 데이터의 로우 데이터 셋의 상호 연결성

(X: 연결 불가, ?: 열 이름이 다르더라도 연결 가능, \*\*: 추가 처리 후 연결 가능)



### 4.3. 시나리오 기반 데이터 상호 연결성 분석

현재 스마트팜 데이터마트에서 사용 가능한 축산 데이터 셋 중 가장 많은 비중을 차지하고 있는 한우 데이터 셋에 초점을 맞추어 시나리오를 기반으로 데이터 간의 상호 연결성을 분석한다. 이에 앞서 축산업과 관련하여 기존의 연구자들이 특히 관심을 가지고 있는 문제들에 대해 정리하고 관련 연구 사례들을 간략히 제시한다. 이어서 이와 같은 문제들에 대해 하나의 시나리오 형식으로 구성하여 스마트팜 데이터마트의 데이터 셋을 통합하고 연결하여 어떤 솔루션을 제공할 수 있는지 구체적으로 설명한다. 이와 같은 과정에서 현재 스마트팜 데이터 셋이 가지고 있는 한계로 인해 발생하는 문제점을 찾아내고 주어진 시나리오가 성공적으로 수행되기 위해서는 어떠한 문제점이 해결되어야 하는 지를 분석한다.

#### 4.3.1 관련 연구 사례

한우는 본래 농산물의 생산과 운송 목적으로 활용되었다. 하지만 시간이 지나면서 본래의 목적 외에 육우로도 활용되기 시작했다. 비교적 높은 가격에도 불구하고 한우는 마블링된 지방의 특성, 독특한 풍미 등으로 한국을 넘어서 외국 소비자에게도 극찬을 받고 있으며 이러한 광범위한 수요로 인해 한우 사육은 수세기에 걸쳐 한국 농업의 중요한 부분이 되어 왔다. 한우 사육자들은 지속적으로 커져가는 국내외 시장의 수요를 충족시키기 위해 고품질 한우를 생산하기 위한 연구를 지속하고 있다. 이를 위해 사육자들은 특별히 설계된 번식 프로그램을 시행하고 구체적인 관리 시스템 및 사료 급이 전략 등을 채택할 필요가 증대되고 있다 (Lee et al., 2014).

한국에서는 좋은 품질의 송아지를 낳을 수 있는 씨를 가지고 있는 수소를 선택하기 위해 해당 수소에 대해 다양한 유전적 테스트를 하여 검증한다. 이러한 테스트에는 체중, 근육 크기, 지방 두께, 마블링과 같은 특성을 살펴보는 것을 포함한다. 어린 소의 경우에는 주로 체중과 마블링에 초점을 맞추며 성체에 대해서는 근육 크기, 지방 두께, 마블링을 주로 고려한다. 이러한 테스트는 사육자가 좋은 품질의 자손을 낳을 수 있는 수소를 선택하는 데 도움을 준다 (Song, 1994).

좋은 품질의 소에 대한 번식과 관련하여 초기에는 젊은 수소의 유전적 평가를 통해 우월한 소를 선발하고 이 소의 자손인 송아지에 대해 추적 관찰하여 품질에 대한 테스트를 진행하였다. 검증된 수소들로부터 번식된 송아지는 체중, 평균 일일 체중 증가량 등의 관찰된 특성을 바탕으로 농장에서 키워져 도축될 때까지 지속적으로 추적 관찰했다 (Lee et al., 2014; Park et al., 2013).

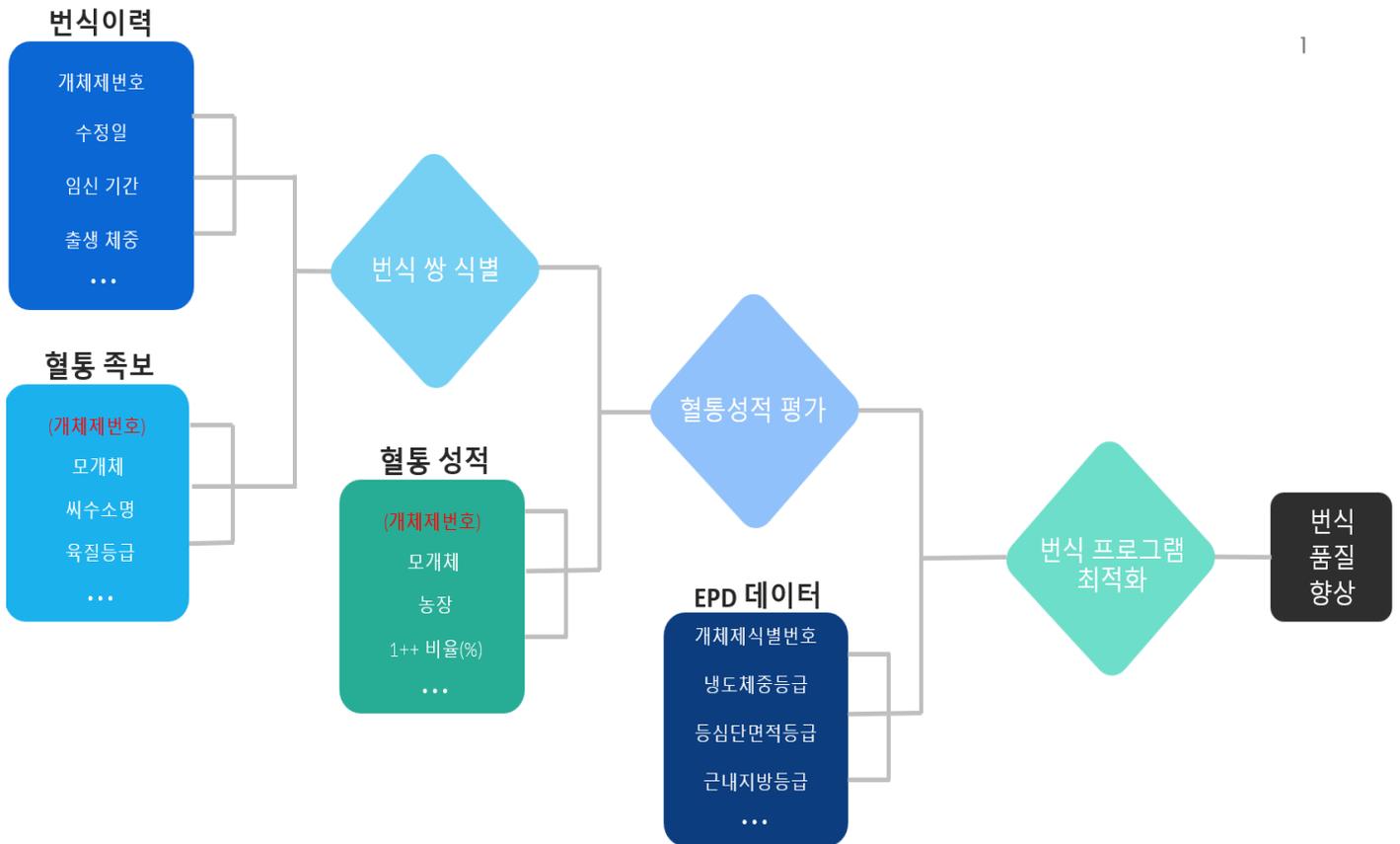
번식 과정 외에도, 사료 급이 시스템은 한우 소고기의 도체 특성과 전반적인 품질을 결정하는데 중요한 역할을 한다. 한우 사료 급이 시스템은 효율적인 성장을 보장하면서 소고기 품질을 최적화하도록 설계되었으며 이 시스템은 일반적으로 특정 영양 함량을 가진 제한된 농축사료와 목초지에 대한 자유로운 접근을 보완하는 방식으로 진행된다 (Park et al., 2013). 또한 한우 농장의 환경 관리 방법은 소가 얼마나 잘 성장하고 건강을 유지하는지에 큰 영향을 미칠 수 있다. 지나치게 덥거나 추운 극단적인 온도는 한우에게 스트레스를 유발하여 사료 섭취 감소, 체중 증가 부진, 그리고 면역 기능 저하로 이어질 수 있다. 따라서 극단적인 기상 조건의 부정적인 영향을 완화하고 소의 편안함을 보장하기 위해 적절한 농장의 환경 관리 시스템은 필수적이다 (Choi et al., 2018; Kim et al., 2003; Lee et al., 2012).

### 4.3.2 시나리오 분석

#### 시나리오 1: 번식 프로그램 최적화

한우의 번식 프로그램을 최적화하기 위해 스마트팜 데이터마트에서 제공하는 번식이력, 개체별 혈통족보, 개체별 개체별 혈통성적(한우), EPD 데이터 셋을 통합 활용하여 아래와 같은 프로세스로 번식 프로그램 최적화를 달성할 수 있다 (그림 43 참조)

그림 43. 시나리오 1 - 번식 프로그램 최적화



#### 1) 번식이력 데이터 셋 활용

- 스마트팜 데이터 마트의 개체별 번식이력(표 4 참조)을 활용하여 번식 이벤트와 관련된 상세 데이터(ex. 개체제번호, 수정일, 임신 기간, 출생 체중 등)에 접근한다.

## 2) 번식 쌍 식별

- 개체별 혈통족보 표(표 8 참조)를 활용하여 개체별 번식이력과 관련된 번식 쌍을 식별한다.
- 하지만, 현재 **개체별 혈통족보 표에서는 개체별 번식이력에서 찾을 수 있는 개체제번호와 연결될 수 있는 식별자가 존재하지 않아 연결이 불가능하다.** 모개체, 씨수소명 모두 현재 개체제번호와 연결되지 않는다. (그림 43 에 “(개체제번호)”로 나타내어 해당 데이터 셋에 필요한 데이터가 없음을 의미)
- 이와 같은 문제가 해결된다면 번식 이력과 관련된 번식 쌍을 식별하고 출하 성적 데이터와 연결 분석하여 우수한 번식 혈통을 식별할 수 있을 것이다.

## 3) 혈통성적 평가

- 개체별 혈통성적 표(표 9 참조)를 연결하여 2)에서 식별한 혈통별로 자손의 품질을 평가 및 분석할 수 있다. 이는 어떤 혈통 또는 번식 쌍이 고품질의 자손을 번식시킬 수 있는지 식별하는데 도움이 된다.
- 현재 **혈통성적 평가 데이터 셋 역시 개체별 식별이 불가능하기 때문에 이같은 분석은 불가능하다.** (혈통성적 평가의 모개체 데이터는 번식이력 데이터 셋의 개체제번호와 대응되지 않는다)

## 4) EPD 데이터 셋 활용

- EPD 데이터 셋(표 10 참조)을 활용하면 도체 관련 데이터(ex. 냉도체중등급, 등심단면적등급, 근내지방등급 등)를 얻을 수 있는데 이는 소 개체별로 가진 육질 특성을 파악할 수 있다.
- 번식 이력의 개체제번호와 EPD 데이터 셋의 개체식별번호를 대응하면 소 마다 가지고 있는 다양한 특성을 분석할 수 있다. 이는 앞서 제시한 혈통과 함께 분석하게 된다면 우수한 혈통이 가지고 있는 소의 육질 특성을 분석할 수 있을 것이다.

## 5) 번식 프로그램 최적화

- 이전의 분석과정으로 우수한 품질의 자손을 번식할 수 있다고 평가되는 혈통과 번식 쌍을 식별해 냈다면 해당 번식 쌍 데이터를 적극적으로 활용하여 인공 수정 등의 방식으로 한우 자손을 번식시킨다.
- 이후, 해당 한우 자손의 품질 데이터와 관련하여 식별 가능하도록 추적 기록하여 지속적으로 데이터를 평가하고 한우의 번식 품질 향상을 위해 위의 과정을 반복 실행하여 최적화한다.

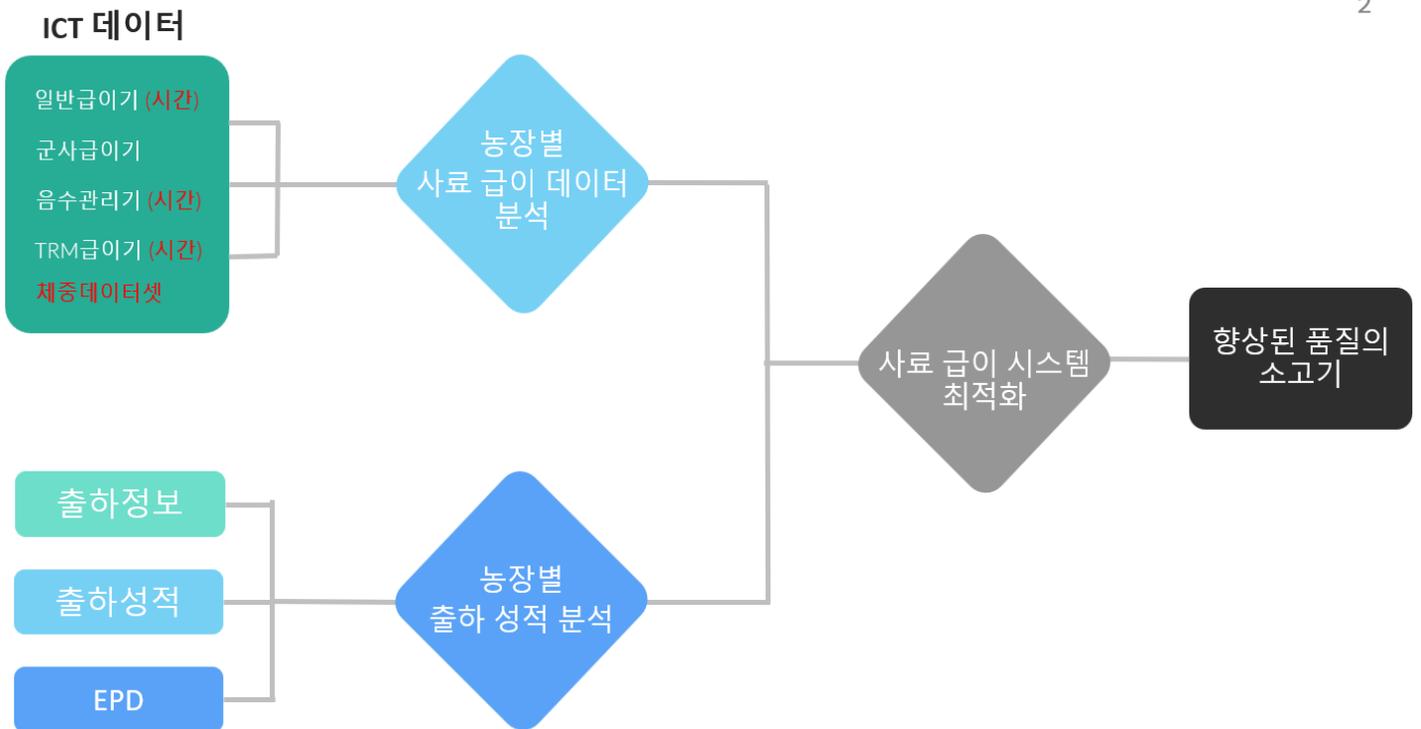
이러한 데이터 분석 프로세스를 통해 한우 사육자들은 좋은 유전적 특성을 가지고 있는 수소와 암소 쌍을 식별할 수 있게 되어 기존보다 효율적인 방식으로 소의 번식 프로그램을 최적화할 수 있으며 결과적으로 향상된 품질의 소고기를 생산하여 농가의 생산성 향상에 기여할 것이다. 하지만 현재 **번식이력과 EPD 데이터에만 개체제 식별번호가 있어서 혈통 족보와 혈통 성적이 연결되지 않아 번식 프로그램을 최적화를 하는데 한계점이 있다.**

## 시나리오 2: 사료 급이 시스템 최적화

고품질의 한우 소고기를 얻는 것이 최종 목적이라고 가정한다면 해당 한우 소고기를 얻기 위한 과정에서 소가 섭취하는 사료의 양, 사료의 섭취 시간, 사료의 종류 등의 데이터와 소의 체중, 출하 성적, 소고기 품질 등의 데이터 셋과 연결해 분석해야 한다. 이러한 데이터 분석 과정을 통해 가장 효율적인 방식의 사료 급이 시스템을 찾아내고 최적화시킨다. 이 시나리오는 스마트팜 데이터 마트에서 제공되는 ICT 데이터 - 일반급이기, 군사급이기, 음수관리기, TMR 급이기, 체중 데이터 셋과 출하정보, 출하성적, EPD 데이터셋을 연결하여 분석함으로써 달성할 수 있다 (그림 44 참조)

그림 44. 시나리오 2 - 사료 급이 시스템 최적화

2



### 1) 사료 급이 데이터 분석

- ICT 데이터 - 일반급이기, 군사급이기, 음수관리기, TMR 급이기(표 14, 15, 16, 18 표 참조) 데이터 셋을 활용하여 측정일시, 급이량, 사료코드, 배합비율, 개체번호, 급이시작일시, 급이종료일시 등의 급이 관련 데이터에 접근한다.

- 다만, 급이와 관련하여 소가 하루에 몇 회, 정확히 언제 먹는지 추적 관찰할 필요가 있기 때문에 측정일시와 관련된 데이터는 최소한 분 단위로 기록되어야 한다. 이러한 관점에서 현재 군사급이기 데이터는 충분한 정보를 제공하고 있지만 일반급이기, 음수관리기, TMR 급이기는 측정일시에 대해 일 단위로 제공되고 있다는 점에서 해당 데이터로부터 얻을 수 있는 정보가 한정되어 있다. (그림 45 에서 (시간)으로 표기하여 해당 데이터에 문제가 있음을 나타냄)
- 또한 군사급이기를 제외한 나머지 급이 관련 데이터는 개체별 데이터가 아닌 농장별 데이터로 제공되므로 소의 육질과 관련한 데이터는 개체별이 아닌 농장별 분석을 하여야 한다.

## 2) 농장별 출하 성적 분석

- 출하정보, 출하성적, EPD, ICT 데이터-체중(표 6, 7, 10, 11 참조) 데이터 셋을 통해 농장 별로 제공되는 육량등급, 육질등급, 육량지수, 체중 등 소의 품질과 관련한 데이터를 얻을 수 있다.
- 이러한 데이터를 활용하여 농장별로 동일하게 적용되는 급이 시스템을 분석한 내용과 그 결과로 제시될 수 있는 농장별 출하 성적을 분석함으로써 급이시스템과 소고기의 품질의 인과성을 파악할 수 있다.
- 단, ICT 데이터-체중 데이터는 현재 측정되고 있는 데이터가 없다. 소의 체중 데이터는 소고기의 품질과 직접적 연관이 있으며 급이 시스템의 시계열적 추적 관찰을 통해 소의 체중의 변화를 추적하는데 중요한 데이터가 될 수 있으므로 향후 체중 데이터 측정이 필요할 것으로 보인다.

## 3) 사료 급이 시스템 최적화

- 사료 급이 데이터의 측정 주기가 분단위로 수정이 되고 어떤 사료를 사용했는지, 사료의 종류는 무엇인지, 어떤 배합으로 사용되었는지, 얼마나 많은 양의 사료가 제공되었는지 등의 데이터와 농장별 소의 품질과 관련된 데이터가 적절히 연결될 수 있다면 위의 1), 2) 과정을 반복하여 농장의 사료 급이 시스템을 조절하여

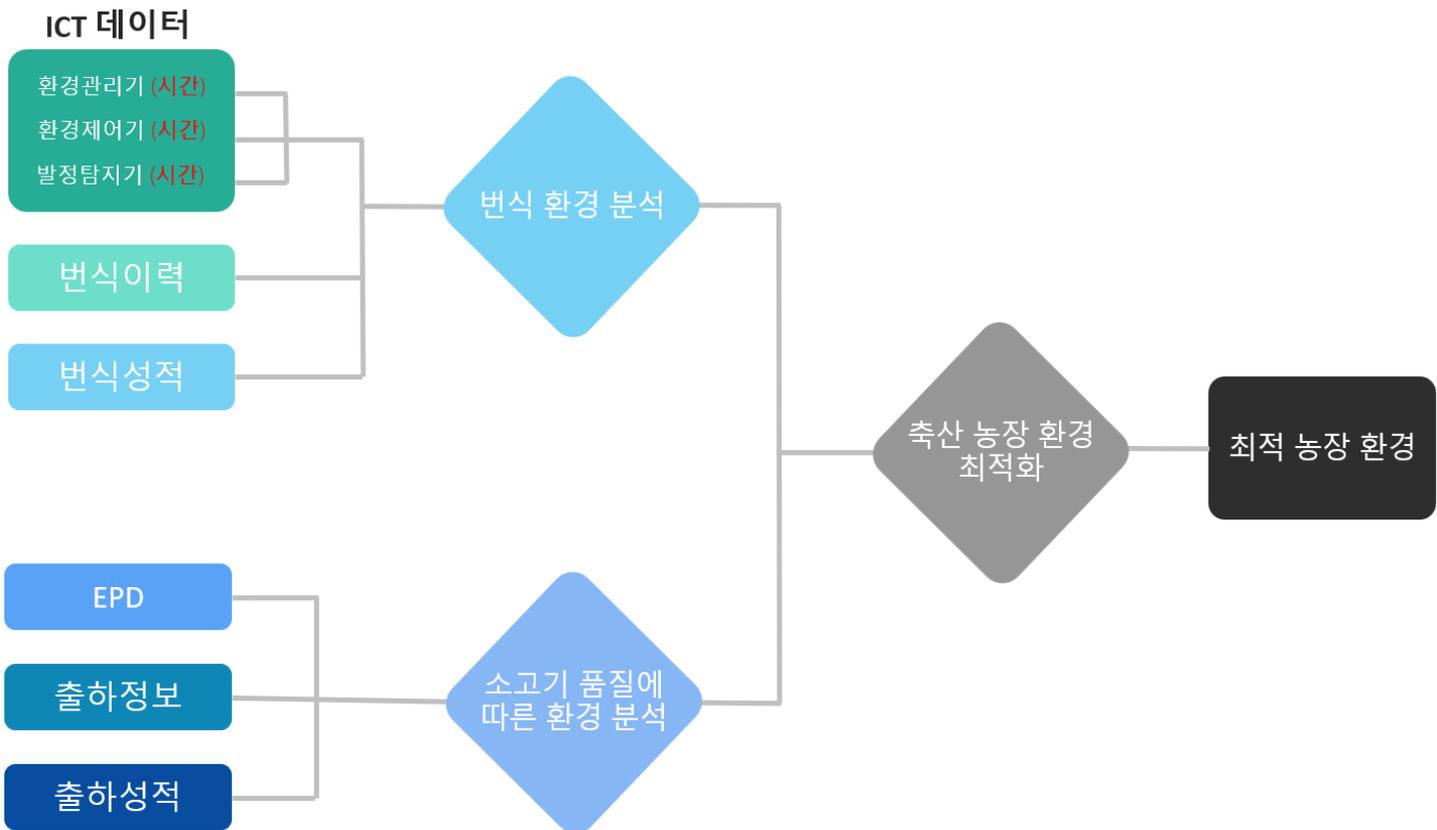
사료의 효율을 향상시키고 비용을 절감하여 최적화된 사료 급이 시스템을 찾을 수 있을 것이다.

이와 같은 시나리오가 달성된다면, 농장은 기존보다 적은 비용으로 최적의 효율을 내는 사료 급이 시스템을 달성할 수 있고 결과적으로 향상된 품질의 소고기를 생산할 수 있어 농가의 수익성을 향상시키는데 도움이 될 것이다. 하지만 현재 ICT 데이터에서 각 급이기 또는 관리기에서 전달되는 데이터 시간이 통일화 되어 있지않아서 통일성이 필요하며, 소의 육질과 관련된 데이터는 개체별 데이터가 아닌 농장별 데이터로 측정해서 분석할 수 있도록 데이터 수집 기준의 변화가 필요하다.

### 시나리오 3: 축산 농장 환경 관리

축산 농장의 환경적인 요인은 소의 스트레스 및 건강 상태에 밀접한 영향을 주며 이는 궁극적으로 소의 번식, 소고기의 품질에 지대한 영향을 미칠 수 있다. 이러한 점을 미루어 보아 축산 농장 환경이 어떻게 관리되고 있는지 추적 관찰하는 것이 필요하며 소의 성장, 건강, 복지에 관련하여 최적의 환경을 제공할 수 있도록 스마트팜 데이터마트에서 제공되고 있는 데이터들을 활용하여 분석해본다. 현재 제공되고 있는 ICT 데이터의 환경관리기, 환경제어기, 발정탐지기와 번식이력, 번식성적, 출하정보, 출하성적, EPD 데이터 셋을 통합 및 연결하여 아래와 같은 분석 과정을 통해 최적의 축산 농장 환경 관리 솔루션을 구현할 수 있다 (그림 45 참조).

그림 45: 시나리오 3 - 축산 농장 환경 관리



### 1) 축산 농장 환경 데이터 수집

- ICT 데이터-환경관리기, 환경제어기(표 12, 20 참조) 데이터 셋에서 해당 농장의 온도, 습도, CO2, 암모니아 데이터에 대해 접근
- 현재 분석하고자 하는 목적에 적합하려면 환경 관련 데이터는 반드시 최소한 분 단위의 데이터가 제공되어야 한다. 하지만 현재 제공되고 있는 ICT 데이터-환경관리기의 측정 일자는 최소 일 단위로 제공되고 있으며 이마저도 어떤 시점에 해당 데이터가 측정되었는지 알 수 없어 환경 데이터로서의 활용도가 떨어진다. (그림 46 에서 (시간)으로 표기하여 해당 데이터가 문제가 있음을 나타냄)
- ICT 데이터- 환경관리기 (표 11 참조), ICT 데이터 환경제어기 (표 19 참조)

### 2) 번식 관련 데이터 연결

- 번식이력, 번식성적, ICT 데이터- 발정탐지기(표 4, 5, 19 참조) 데이터 셋을 통해 번식과 관련한 다양한 데이터에 접근할 수 있다. 이와 같은 번식 데이터를 축산 농장 환경 데이터와 연관지어 분석하면, 환경적 요인이 번식 성과에 어떤 영향을 줄 수 있는지 분석이 가능하다.
- 다만, 특정 시점의 온도, 습도, CO2, 암모니아 상태 등 농장 환경에 대해 대응하는 소의 발정 이벤트를 추적해야 하므로 발정 탐지기 역시 최소한 분 단위의 데이터를 제공해야 한다. 하지만 현재 발정탐지기 데이터는 최소 일 단위로 제공되고 있어 실시간 발정탐지에 대한 추적 관찰이 불가능한 상황이다.
- 또한, 번식 이력, 번식 성적에 대해서도 일 단위로 제공되고 있는 환경 데이터가 특정 시점에 일관되게 측정된 데이터임을 확신할 수 없기 때문에 이 두 데이터 셋 상호간 의미있는 연결이 불가능하다.

### 3) 소의 출하 성적 데이터 연결

- 출하정보, 출하성적, EPD (표 6, 7, 10 참조) 데이터 셋을 통해 농장 별로 제공되는 육량등급, 육질등급, 육량지수 등 소의 품질과 관련한 데이터에 접근한다.
- 이와 같은 데이터를 축산 농장 환경 데이터와 연결하여 분석함으로써 농장의 환경 요인이 소고기의 품질과 어떠한 관련이 있는지 분석할 수 있다.

- 하지만, 현재 출하성적은 최소 년도별로 측정되는 데이터로 일 단위로 측정되는 농장 환경 데이터와 대응되기엔 그 기간의 차이가 너무 길어서 최소 월 단위의 데이터는 제공되어야 의미 있는 결과를 얻을 수 있을 것이다.
- EPD 정보의 경우 개체별로 갱신일자를 기준으로 시계열 데이터를 얻을 수 있으므로 특정 농장에 속해 있는 개체들에 대해 환경 데이터와 소의 품질 변화를 연결 지어 분석할 수 있다. 다만, 이 경우에도 환경 데이터가 특정 시점에 일관되게 측정되었는지 여부가 중요하고 이와 같은 문제가 해결되어야 분석이 가능할 것이다.

#### 4) 최적의 축산 농장 환경 식별

- 이처럼 환경관리기 및 환경 제어기 데이터 셋을 활용하여 온도, 습도, CO2 농도, 암모니아 농도 등의 환경 데이터를 얻고 해당 환경데이터의 변화가 소의 번식 및 출하 성적에 어떤 영향을 주는지 분석하여 최적의 축산 농장 환경을 식별해 낼 수 있다.
- 추가적으로 소의 스트레스 측정 등과 관련된 직접적인 데이터를 스마트팜 데이터마트 데이터 셋으로 추가한다면 현재보다 더 직접적인 환경요인의 연관성을 파악할 수 있을 것으로 생각된다.

이와 같은 축산 환경 관리 최적화가 달성된다면, 농장은 최적의 농장 환경 관리를 통해 소의 스트레스 감소와 건강 상태 증진이 가능하게 되어 향상된 품질의 소고기를 생산할 수 있을 것이다. 하지만 현재 ICT 데이터에서 측정되고 있는 시간에 대한 정보가 없어서 데이터 활용도가 떨어지고 있어 ICT 데이터 수집 시간에 대한 기준이 필요하다.

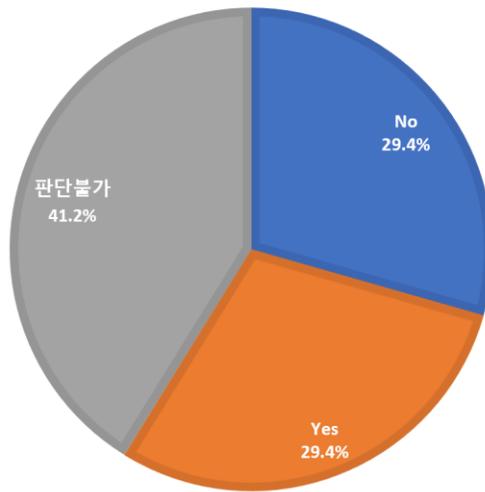
## 5. 결론 및 제안

최근 농가 인구는 시간이 지나며 점점 감소하고 있는 반면 노동인력 1인당 관리해야 하는 가축 수가 지속적으로 증가함에 따라 축산업에 있어서 업무 과중 및 노동력 부족 문제가 심화되고 있다. 이에 더하여 악취 문제, 탄소 배출 증가, 가축 복지 및 방역 등 다양한 문제가 발생하고 있다. 정부는 이러한 문제를 해결하는 것에서 더 나아가 축산업의 근본적인 경쟁력 강화를 위해 스마트팜과 관련하여 지속적인 정책적 지원을 하였다. 2014년 축산농가에 ICT 장비를 지원하는 것을 시작으로 임대형 스마트팜 조성, 스마트팜 혁신밸리조성, 축산 인공지능 학습 데이터 구축 사업 등을 통해 빅데이터와 인공지능을 기반으로 하는 스마트 농업의 확산을 위해 인프라를 구축해왔다. 그 결과 스마트팜 데이터마트는 축산 분야에 상당한 양의 빅데이터를 축적했고, 이렇게 축적된 축산 빅데이터를 민간 또는 관련 산업체에 무상으로 제공하여 다양한 데이터 기반 솔루션을 개발할 수 있는 기회를 제공했다.

하지만 예상과는 달리 스마트팜 데이터마트의 데이터 셋에 대해 실질적인 활용까지 이어지지 못하고 있는 상황으로 우리는 이러한 문제 상황이 데이터의 표준화가 제대로 되어 있지 않다는 점에서 기인한다고 판단했다. 이와 같은 상황에서 해당 문제의 원인을 보다 면밀히 분석하고 그에 대한 알맞은 해법을 제시하고자 우선 현재 국제적으로 통용되는 축산 데이터의 측정 및 저장에 관한 표준을 제공하고 있는 ICAR의 지침에 대해서 조사했다. 이렇게 조사한 ICAR의 표준 가이드라인 지침을 스마트팜 데이터마트의 각 데이터 셋과 비교 분석하여 데이터 셋의 어떤 부분이 표준 가이드라인을 따르고 있지 않는지 상세하게 분석하여 그에 대한 내용을 섹션 3에 제시하였다. 이에 대한 결과를 한눈에 제시하고 있는 그림 46에 따르면 한우 데이터 셋이 ICAR 데이터 표준을 준수하고 있는 비율은 현재 29.4%(전체 17개 데이터 셋 중 5개)에 해당한다. 이외에 29.4%(전체 17개 데이터 셋 중 5개)에 해당하는 데이터 셋은 ICAR의 관련된 데이터 표준과 비교하였을 때 이를 따르고 있지 않는 경우에 해당하고 나머지 판단불가로 표기된 41.2%(전체 17개 데이터 셋 중 7개)의 데이터 셋은 해당 데이터 셋에 대응되는 ICAR의 표준이 없어 해당 가이드라인의 준수 여부를 판단할 수 없는 경우를 나타낸다. **판단불가의 경우를 제외하면 준수와 비준수의 비율은 각각 50%에 해당한다.**

그림 46. ICAR 데이터 표준 준수비율 (한우 데이터)

ICAR 데이터 표준 준수 비율 (한우)



이처럼 스마트팜 데이터마트의 데이터 셋이 ICAR의 데이터 표준을 따르고 있지 않는 것을 확인한 후 데이터가 표준화 되어있지 않음으로써 발생할 수 있는 다양한 문제 상황에 대해서 분석해보았다. 이는 관련 연구 사례를 조사하여 연구자와 관련 산업 종사자가 관심을 가지고 있는 문제 상황을 파악하고 이와 같은 문제를 해결할 수 있는 시나리오를 설정하여 해당 시나리오를 풀어 나가는 과정에서 스마트팜 데이터마트의 데이터를 기반으로 데이터 표준화의 미비로 인해 발생하는 이슈들을 분석하였다. 이에 대해 섹션 4에서 번식프로그램 최적화, 사료급이시스템 최적화 및 축산농장 환경관리에 대한 3가지 시나리오를 정의하였고, 이러한 시나리오를 분석하는 과정을 통해 데이터 표준화 미비로 데이터 셋의 통합 및 연결이 불가능하여 실제로 활용할 수 있는 정도가 한정적임을 파악할 수 있었다. 데이터 연결성을 높이기 위해 스마트팜 데이터마트의 데이터 셋에 대해 종합적인 관점에서 필요한 표준화 작업이 무엇인지 식별하였고 이러한 분석 내용에 따라 다음과 같이 6가지 항목에 대해서 데이터의 개선이 이루어질 필요가 있음을 제안한다.

### 1. 데이터 스키마 정의

데이터 스키마는 데이터베이스 내의 데이터 구조를 정의하는 일종의 설계도와 같다. 구체적으로 각 데이터 항목에 대해서 입력될 수 있는 데이터 타입(ex. 숫자, 문자열 등)을 미리 지정할 수 있고 저장할 수 있는 데이터의 크기와 빈 공백 데이터를 넣는 것이

가능한지 여부를 강제할 수 있다. 이처럼 데이터 스키마를 지정하면 그림 13 과 같이 축산 농장의 ICT 기기의 잘못된 측정으로 인해 발생한 오류 데이터가 데이터 셋에 포함되는 것을 방지할 수 있고 반드시 값을 가져야 하는 키 데이터에 대한 일관성을 보장하여 다른 데이터 셋과의 관계를 쉽게 정의하고 연결할 수 있다. 이는 궁극적으로 데이터의 무결성을 유지할 수 있도록 도움을 주며 데이터 셋에 저장된 정보의 일관성과 정확성을 보장할 수 있도록 한다.

## 2. 데이터설명서 추가

데이터설명서는 데이터 셋의 각 항목에 대한 정의와 상세한 설명을 포함한 정보이다. 현재 스마트팜 데이터마트에서 제공하고 있는 모든 데이터 셋에 대해 해당 데이터 셋이 포함하고 있는 데이터 항목에 대한 설명이 부재하다. 이는 일반인들이 쉽게 사용하지 않는 단어(ex. 생식체중, MSY, PSY 등) 또는 측정항목의 단위(ex. 체중, 온도 등)에 대해 잘못 해석할 수 있는 여지를 남길 수 있다. 또한 온도, 습도, 체중 등 다양한 ICT 장비를 활용한 측정 데이터를 저장하고 있는데 이에 대해 ICT 장비가 공통적으로 어느 위치에 설치되었으며 측정된 데이터가 하루 중 언제 측정되었는지와 같은 정보를 데이터설명서로 제공하여야 해당 데이터가 동일한 조건에서 측정되고 있음을 명확히 알 수 있을 것이다. 따라서 스마트팜 데이터마트의 데이터 셋을 활용하는 모든 사람이 공통된 이해를 하고 있음을 보장하기 위해서 데이터설명서는 반드시 필요한 정보이며 현재 포함하고 있는 모든 데이터 항목에 대해 데이터설명서를 포함시켜야 한다.

## 3. 시간 데이터 포맷 표준화 및 세분화

시간 데이터 포맷의 표준화와 세분화는 스마트팜 데이터마트의 데이터 분석의 정확성과 효율성을 대폭 향상시킬 수 있다. 현재 스마트팜 데이터마트에서 제공되는 다양한 데이터 셋이 서로 다른 날짜 포맷을 사용하고 있어, 데이터를 통합하고 분석하는 과정에서 추가적인 데이터 전처리 작업이 필요한 상황이다. 예를 들어, 개체별 번식로우 데이터는 YYYYMMDD 형태로 제공되는 반면, ICT 데이터-급이기 등의 데이터 셋은 YYYY-MM-DD 형태로 제공되고 있다. 이는 가지고 있는 정보는 같을 수 있으나 날짜를 기준으로 데이터 셋을 연결하는 과정에서 불필요한 데이터 전처리 과정을 포함시키도록

강제하여 데이터 분석에서 비효율을 유발한다. 또한, 현재 제공되는 데이터 중 대부분이 일 단위로만 측정되고 있어 분석의 정밀도와 활용성을 제한하고 있다. 예를 들어, 농장의 온도, 습도, 암모니아 수준, 발정 탐지기, 자동 급이기 등의 데이터를 더 짧은 시간 단위(시, 분, 초)로 제공받을 수 있다면 농장 관리자와 연구자들은 가축의 건강 상태나 행동 패턴을 더 세밀하게 모니터링하고 분석할 수 있게 될 것이다. 이를 통해 가축의 스트레스 요인을 신속히 식별하고 사료 급여 시간을 최적화하며 질병 발생을 조기에 감지하는 등의 조치를 취할 수 있을 것이다. 따라서 스마트팜 데이터마트의 시간 데이터 포맷 표준화와 시간 데이터 세분화가 필요하다.

#### 4. 지역 정보 제공

비슷한 기후와 환경적 조건을 공유하는 지역 내의 농장에서 길러진 가축은 유사한 성장 패턴, 건강 상태, 생산성 등의 공통된 특징을 보일 가능성이 높다. 이러한 지역 정보를 데이터 분석에 포함시킴으로써 지역에 따라 가축의 사료 배합, 건강 관리 프로그램, 번식 전략을 다르게 적용하여 지역별 최적화를 달성할 수 있다. 또한 지역정보는 질병 발생의 지역적 패턴을 식별하고 예방 조치를 마련하는 데 중요한 역할을 할 수 있다. 특정 지역에서 발생하는 질병에 대한 데이터를 분석함으로써 해당 지역 또는 유사한 환경을 가진 다른 지역의 가축 질병 예방 및 관리에 필요한 정보를 얻을 수 있다. 이처럼 지역 정보는 축산업에 있어서 다양한 의사 결정을 지원하는 데 필수적인 요소이다. 따라서 스마트팜 데이터마트에서 제공하고 있는 지역 정보를 가능하면 현재보다 작은 범위(시, 군, 구, 읍, 면, 리)로 제공한다면 해당 데이터로부터 더 많은 정보를 얻을 수 있을 것이다.

## 5. 개체 식별번호

개체 식별번호는 농장에서 키워지는 동물들을 개별적으로 관리하고 추적하는 데 필수적인 요소이다. 이는 각 동물이 태어나는 순간부터 부여되며 생애 주기 전반에 걸쳐 그 동물의 건강 상태, 생산성, 이동 경로 등 다양한 정보를 연결시킬 수 있는 매개체가 된다. 예를 들어, 개체 식별번호를 사용하여 개별 동물의 성장 과정을 모니터링하게 되면 특정 사료가 동물의 성장률에 미치는 영향을 정확히 평가할 수 있고 이를 바탕으로 사료 공급 계획을 최적화할 수 있다. 또한 동물의 건강 이력을 추적함으로써 조기에 질병을 감지하고 적절한 예방 조치를 취할 수 있을 것이다. 게다가 개체 식별번호를 통해 정확한 혈통 정보를 관리하면서 우수한 유전자를 가진 동물을 선별하여 번식시킬 수 있으며 소비자는 특정 육류 제품이 어떤 동물로부터 왔는지, 그 동물이 어떤 환경에서 자랐는지, 어떤 사료를 먹었는지 등의 정보를 알 수 있다. 이와 같은 분석이 가능하려면 해당 데이터 셋들을 개체 식별번호를 기준으로 연결할 수 있어야 한다. 따라서 스마트팜 데이터마트가 모으는 관련 있는 모든 데이터 셋에 대해 개체 하나하나를 일관되게 식별할 수 있는 개체 식별번호를 포함시켜야 한다.

## 6. 집계데이터와 로우데이터 제공

집계 데이터를 제공할 때는 반드시 이를 뒷받침하는 로우 데이터가 제공되어야 한다. 집계 데이터는 특정 기간 동안의 데이터를 집계하여 만들어진 통계적 정보를 포함하고 있기 때문에 로우 데이터에 비해 상대적으로 얻을 수 있는 정보가 한정적이다. 따라서 다양한 사용자가 데이터를 보다 효과적으로 활용할 수 있도록 하기 위해서 집계 데이터뿐만 아니라 그 기반의 로우 데이터도 함께 제공할 필요가 있다. 이 과정에서 집계 데이터와 로우 데이터의 상호 연결성을 강화하는 것이 필요하다. 집계 데이터를 생성할 때 적용된 반올림 방법, 통계적 수식, 그리고 선택된 데이터 집계 기간 등의 정보를 명확하게 제공하여 사용자로 하여금 집계데이터와 로우데이터간의 연결 고리를 분명하게 이해할 수 있도록 해야 한다. 이를 통해 집계 데이터에서 나타나는 특정 결과 값 뒤에 있는 로우 데이터의 세부 사항을 탐색할 수 있도록 하여 차이가 있다면 사용자가 그 차이를 이해하고 납득할 수 있도록 할 수 있다.

## 7. 데이터의 상호연결성을 고려한 데이터 셋 설계

단독으로 존재하는 개별 데이터 셋이 담고 있는 정보는 그 범위가 제한적이다. 이를 다른 데이터 셋과 통합하여 분석할 때, 비로소 빅데이터의 진정한 가치를 발휘할 수 있다. 이와 같은 관점에서 데이터 간의 상호연결성을 고려한 데이터 셋 설계는 매우 중요하다. 효과적인 데이터 셋 설계를 위해 먼저 다양한 관련 연구 사례들을 면밀히 조사하고 현재 연구자와 실무자들이 주목하고 있는 문제 상황들을 발굴해야 한다. 이렇게 발굴된 문제 상황을 바탕으로 다양한 시나리오를 개발하고 이 시나리오들을 통해 시뮬레이션을 진행함으로써 필요한 데이터 셋을 식별할 수 있다. 이 과정은 실제 산업 현장에서 요구되는 가치 있는 데이터 셋을 선별하고 확보하는데 효과적으로 작용할 것이다. 또한 이런 접근 방식으로 구축된 데이터 셋은 시나리오 분석을 통해 어떻게 서로 연결될 수 있는지를 명확하게 정의할 수 있어서 데이터의 상호 연결성을 향상시키는 데 중요한 역할을 할 수 있다. 이러한 접근법은 데이터가 단순히 정보의 집합이 아니라 서로 연결되어 더 큰 가치를 생성할 수 있는 자원임을 인식할 수 있다. 데이터 셋의 설계와 활용에서 데이터 간의 상호연결성을 고려함으로써 보다 포괄적이고 심도 있는 분석을 가능하게 하며 이는 궁극적으로 실세계의 복잡한 문제를 해결하는 데 커다란 도움이 될 것이다.

## 6. 참고 문헌

1. Animal Production Statistics (apro\_anip). (2023). European Commission. [https://ec.europa.eu/eurostat/cache/metadata/en/apro\\_anip\\_esms.htm](https://ec.europa.eu/eurostat/cache/metadata/en/apro_anip_esms.htm)
2. European Food Safety Authority (EFSA), Aminalragia-Giamini, R., Dhollander, S., Gibin, D., Mur, L., Papanikolaou, A., & Zancanaro, G. (2022). Guidance for reporting animal population data using the SIGMA data model. EFSA Supporting Publications, 19, 7568E. <https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2022.EN-7568>
3. European Food Safety Authority (EFSA), Zancanaro, G., Antoniou, S. E., Bedriova, M., Boelaert, F., Gonzales Rojas, J., Monguidi, M., Roberts, H., Saxmose Nielsen, S., & Thulke, H.-H. (2019). SIGMA Animal Disease Data Model. EFSA Journal, 17(1), e05556. <https://doi.org/10.2903/j.efsa.2019.5556>
4. Google, Microsoft, Yahoo, & Yandex. (2015). Schema.org. <https://schema.org>
5. 농림수산식품교육문화정보원. (n.d.). 스마트팜 데이터마트. Retrieved March, 2024, from <https://data.smartfarmkorea.net>
6. ICAR. (2018, 03). Section 3 - ICAR Guidelines for Beef Cattle Production Recording. <https://www.icar.org/Guidelines/03-Beef-Cattle-Recording.pdf>
7. ICAR. (2018, 10). Section 1 - General Rules for ICAR Guidelines (18.04). <https://www.icar.org/Guidelines/01-General-Rules.pdf>
8. ICAR. (2020, 6). Section 19 - Guidelines for Recording Feed Intake for Genetic Evaluation (1.02). <https://www.icar.org/Guidelines/19-Recording-Feed-Intake-for-Genetic-Evaluation.pdf>
9. ICAR. (2020, 6). Section 20 - Recording Dairy Cattle Methane Emission for Genetic Evaluation (1.02). <https://www.icar.org/Guidelines/20-Recording-Dairy-Cattle-Methane-Emission-for-Genetic-Evaluation.pdf>
10. The International Committee for Animal Recording. (n.d.). Retrieved March, 2024, from <https://www.icar.org>

11. Choi, T.-J., Lee, S.-J., Park, J.-E., Lim, D., Cho, Y.-M., & Park, B. (2018). Analysis of environment effects on the carcass traits Hanwoo cows using ultrasonic measurement. Korean Journal of Agricultural Science (농업과학연구), 45(1), 66-73.  
<https://doi.org/10.7744/kj014oas.20180014>
12. Kim, Y. S., Yun, S. K., Song, Y. H., & Lee, S. K. (2003, 4). Effect of season on color of Hanwoo (Korean native cattle) beef. Meat Science, 63(4), 509-513. [https://doi.org/10.1016/S0309-1740\(02\)00112-2](https://doi.org/10.1016/S0309-1740(02)00112-2)
13. Lee, S.-H., Park, B.-H., Sharma, A., Dang, C.-G., Lee, S.-S., Choi, T.-J., Choy, Y.-H., Kim, H.-C., Kim, K.-J., Kim, S.-D., Yeon, S.-H., Park, S.-B., & Kang, H.-S. (2014). Hanwoo cattle: origin, domestication, breeding strategies and genomic selection. Journal of Animal Science and Technology, 56(2). <https://doi.org/10.1186/2055-0391-56-2>
14. Lee, S. M., Kim, J. Y., & Kim, E. J. (2012). Effects of Stocking Density or Group Size on Intake, Growth, and Meat Quality of Hanwoo Steers (*Bos taurus coreanae*). Asian-Australas J Anim Sci., 25(11), 1553–1558. 10.5713/ajas.2012.12254
15. Mwangi, F. W., Charmley, E., Gardeniner, C. P., Malau-Aduli, B. S., Kinobe, R. T., & Malau-Aduli, A. E. O. (2019). Diet and Genetics Influence Beef Cattle Performance and Meat Quality Characteristics. Foods, 8(12), 648. <https://www.mdpi.com/589578>
16. Park, B., Choi, T., Kim, S., & Oh, S.-H. (2013, 2). National Genetic Evaluation (System) of Hanwoo (Korean Native Cattle). Asian-Australasian journal of animal sciences, 26(2), 151-156. 10.5713/ajas.2012.12439
17. Song, C.-W. (1994). THE KOREAN HANWOO BEEF CATTLE. Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animales, 14, 61 - 71.  
<https://doi.org/10.1017/S1014233900000341>

## 7. 부록

### 스마트팜 데이터마트 한우 데이터 데이터 셋

#### 1. 번식이력

이름	샘플 데이터	문제
지역	3611000000	결측치가 존재하면 안되는 데이터에 결측치의 존재
규모	100~200	
농장	21161	
개체제번호	20120415010115	
산차	3	
수정차수	2	
수정일	2015-10-16	
수정유형	인공수정	
KPN	KPN944	명확하지 않은 정의
분만일	20160727	
임신기간(일)	285	
종부횟수	2	수정차수(Parity)과의 차이가 불분명, 데이터 설명 필요
공태일수	155	
분만간격(일)	440	
초산일령		
송아지성별	수컷	
출생체중(kg)	0	

2. 년도 별 번식성적(한우)

이름	샘플 데이터	문제
지역	5100000000	결측치가 존재하면 안되는 데이터에 결측치의 존재
규모	500~1000	
농장	20418	
년도	2016	
초산일령	774.6	결측치 존재
공태일수	81.3	결측치 존재
분만간격(일)	368.3	결측치 존재
임신기간(일)	287.2	

### 3. 출하정보

이름	샘플 데이터	문제
지역	4100000000	결측치가 존재하면 안되는 데이터에 결측치의 존재
규모	100 두 미만	
농장	22912	
개체제번호	20210418011660	
출생일자	20210418	
품종	한우	결측치가 존재하면 안되는 데이터에 결측치의 존재
성별	거세	결측치가 존재하면 안되는 데이터에 결측치의 존재
월령	33	
도축장	도드람엘피씨공사	결측치 존재
도축일자	20231205	
육량등급	A	결측치 존재
육질등급	1+	
도체중(kg)	475	
등급판정일자	20231206	성적결정일자
육량지수	62.77	결측치 존재
등급명	1+A	
등지방두께(mm)	6	0 또는 NaN 값을 통일해서 표기
등심면적(㎠)	88	0 또는 NaN 값을 통일해서 표기
근내지방	62	단위 불분명

육색	5	결측치 존재
지방색	3	결측치 존재
조직감	2	결측치 존재
성숙도	2	
경락가격	18999	

#### 4. 출하성적

이름	샘플 데이터	문제
지역	4100000000	결측치가 존재하면 안되는 데이터에 결측치의 존재
규모	100 두 미만	
농장	22912	
년도	2016	도축 연도인지 등급 연도인지 불분명
1++(마리)	0	
1+(마리)	1	
1(마리)	1	
2(마리)	0	
3(마리)	0	
D(마리)	0	
총합(마리)	2	
1++ 비율(%)	0.00%	
1+ 비율(%)	50.00%	
1 비율(%)	50.00%	
2 비율(%)	0.00%	
3 비율(%)	0.00%	
D 비율(%)	0.00%	

5. 혈통족보

이름	샘플 데이터	문제
지역	4100000000	결측치가 존재하면 안되는 데이터에 결측치의 존재
규모	300~400	
농장	20212	
모개체	암	
1 계대	암	
2 계대	암	
3 계대	암	
4 계대	암	
5 계대	암	
6 계대	암	
7 계대	암	
생년월일	2009-08-02	결측치가 존재하면 안되는 데이터에 결측치의 존재
씨수소명	KPN622	결측치 존재
도체중	252	단위 불분명
육질등급	3	결측치 존재
도축일	2013-02-18	
월령	43	

6. 개채별 혈통성적(한우)

이름	샘플 데이터	문제
지역	4100000000	결측치가 존재하면 안되는 데이터에 결측치의 존재
규모	100 두 미만	
농장	22913	
모개체	256	
1++(마리)	1	
1+(마리)	2	
1(마리)	1	
2(마리)	0	
3(마리)	0	
D(마리)	0	
총합(마리)	4	
1++ 비율(%)	25	
1+ 비율(%)	50	
1 비율(%)	25	
2 비율(%)	0	
3 비율(%)	0	
D 비율(%)	0	

7. EPD

이름	샘플 데이터	문제
지역	4100000000	결측치가 존재하면 안되는 데이터에 결측치의 존재
규모	100 두 미만	
농장	22913	
개체식별번호	20220317010358	
생년월일	2022-03-17	
성별구분	암	
계대	3	0 또는 NaN 값을 통일해서 표기
냉도체중등급	A	결측치 존재
냉도체중지수	19.5279	결측치 존재
등심단면적등급	B	결측치 존재
등심단면적지수	2.9848	결측치 존재
등지방두께등급	C	결측치 존재
등지방두께지수	-0.0396	결측치 존재
근내지방등급	D	결측치 존재
근내지방지수	0.0754	결측치 존재
갱신일자	2023-02-01	결측치 존재

8. ICT – 체중

이름	샘플 데이터	문제
지역	-	결측치 존재
규모	-	결측치 존재
농장	-	결측치 존재
개체번호	-	결측치 존재
제조사아이디	-	결측치 존재
장비번호	-	결측치 존재
측정일시	-	결측치 존재
체중(kg)	-	결측치 존재

9. ICT – 환경관리기

이름	샘플 데이터	문제
지역	4100000000	
규모	100 두 미만	
농장	22044	
장비종류	온도	
제조사아이디	미래씨엔씨	
장비번호	1	
측정일시	2023-12-31	
측정값(℃ · % · ppm)	2	데이터 값의 의미 불분명

10. ICT – 사료빈

이름	샘플 데이터	문제
지역	4100000000	
규모	100 두 미만	
농장	22344	
제조사아이디	SAMWOO	
장비번호	1	
측정일시	2021-12-02 23:26:05	
잔여량(kg)	0	
배출량(kg)	1156	

11. ICT – 일반급이기

이름	샘플 데이터	문제
지역	4100000000	
규모	100 두 미만	
농장	22044	
제조사아이디	미래씨엔씨	
장비번호	4	
측정일시	2023-12-31	
급이량(kg)	1820	

12. ICT – 군사급이기

이름	샘플 데이터	문제
지역		
규모	500~1000	
농장	0020418	
개체번호	20090512010957	
제조사아이디	주식회사 다운	
장비번호	1	
측정일시	2023-01-31	
섭취량(kg)	2	
설정량(kg)	9	
급이시작일시	2023-01-31 14:12:00	
급이종료일시	2023-01-31 14:17:24	

13. ICT – 음수관리기

이름	샘플 데이터	문제
지역		
규모	100 두 미만	
농장	0022743	
제조사아이디	(주)에그리로보텍	
장비번호	371	
측정일시	2023-01-31	
측정값(L)		결측치

14. ICT – 포유기

이름	샘플 데이터	문제
지역	-	결측치 존재
규모	-	결측치 존재
농장	-	결측치 존재
제조사아이디	-	결측치 존재
장비번호	-	결측치 존재
측정일시	-	결측치 존재
포유량	-	결측치 존재
포유시작일시	-	결측치 존재
포유종료일시	-	결측치 존재
포유온도(°C)	-	결측치 존재

15. ICT – TMR 급이기

이름	샘플 데이터	문제
지역	3000000000	
규모	100 두 미만	
농장	0021057	
제조사아이디	(주)길소프트	
장비번호	1	
측정일시	2023-01-12	
급이량(kg)	406.00	
사료코드		결측치 존재
배합비율(%)		결측치 존재

16. ICT – 발정탐지기

이름	샘플 데이터	문제
지역	4300000000	
규모	100~200	
농장	0022294	
개체번호	20171008010149	
제조사아이디	상육축산	
장비번호	121186664	
측정일시	2023-01-08	
운동량(회)	-	결측치 존재
반추량(회)	-	결측치 존재
체온(°C)	-	결측치 존재
pH	-	결측치 존재
발정상태값	48	
건강상태값	100	
알람일시	2023-01-08 03:25:11	결측치 존재

17. ICT – 환경제어기

이름	샘플 데이터	문제
지역	3611000000	
규모	100~200	
농장	0021161	
장비종류	송풍팬	
제조사아이디	에프에스	
장비번호	1	
측정일시	2023-01-31	
측정값 01(% · on/ff)	0	
측정값 02	-	결측치 존재
측정값 03	-	결측치 존재
측정값 04	-	결측치 존재
측정값 05	-	결측치 존재

## 스마트팜 데이터마트 양돈 데이터 데이터 셋

### 1. 개체별 번식로우(양돈)

이름	샘플 데이터	문제
지역	경기도	
규모	100~200	
농장	0002615	
개체제번호	2	
산차	1	의미가 불분명한 음수 값을 포함
교배일	20141030	
임신사고일	61.0	
분만일	20141030	
총산	14	
실산	13	
생시총체	7.0	의미가 불분명
이유일	20150611	결측치 존재
이유두수	13	결측치 존재

## 2. 년도별 번식통계(양돈)

이름	샘플 데이터	문제
년도	2019	
지역	경상남도	
규모	100~200	
농장	PIGGO_156	
상시모돈수	146.10	
MSY(두)	19.2	
PSY(두)	21.7	
모돈회전율(회전)	2.18	의미가 불분명
평균비생산일수(일)	69.40	의미가 불분명
평균총산(두)	13	의미가 불분명
평균실산(두)	12	의미가 불분명
평균이유(두)	10	의미가 불분명
분만율(%)	76.30	의미가 불분명
초교배일령(일)	263.1	의미가 불분명
임신기간(일)	116.0	의미가 불분명
포유기간(일)	20.2	의미가 불분명
재귀발정일(일)	12.9	의미가 불분명

### 3. 월별 번식통계(양돈)

이름	샘플 데이터	문제
년도	2012	
지역	충청북도	
규모	100~200	
농장	0020069	
데이터구분	임신사고(조기)	
1~12 월	3.00	

### 4. 산차별 번식통계(양돈)

이름	샘플 데이터	문제
지역	경상남도	
규모	100~200	
농장	0021290	
데이터구분	평균이유	
1~7 산	10.20	
8 산이상	11.8	

5. ICT 데이터 급이기(양돈)

이름	샘플 데이터	문제
지역	2020-02-15	
규모	100 두 미만	
농장	0020091	
개체제번호	8	결측치 존재
장비구분	포유모돈 급이기	
장비번호	22	
섭취일	2020-02-19	
설정량(kg)	3.00	
섭취량(kg)	3.00	
교배일	2019-10-21	의미가 불분명
분만일	2020-02-15	

6. ICT 데이터 음수기(양돈)

이름	샘플 데이터	문제
지역	충청남도	
규모		결측치 존재
농장	0020090	
장비구분	음수기	음수기만 해당, 불필요한 열
장비번호	1	
섭취일	2023-11-14	
섭취량(kg)	23	결측치 존재

7. ICT 데이터 환경정보(양돈)

이름	샘플 데이터	문제
지역	4600000000	
규모	200~300	
농장	0022871	
장비구분	온도	
장비번호	04	
측정일	2023-07-12	
측정값(온습도 : 도 · CO2 : PPM)	25.00	

8. 포유모돈 급이변식 데이터(양돈)

이름	샘플 데이터	문제
지역		모두 빈 값
규모	1000 두 이상	
농장	PF_0020082	하나의 농장만 사용
품종	F1	한 품종만 사용
산차	7	
급이횟수	3	
급이설정량	7.50	
섭취량	7.50	
급이일자	2022-01-18	
분만일로부터의 기간	13	잘못된 데이터 포함(ex.음수)
분만일	2022-01-05	
총산	11	
실산	10	
분만시 BCS	19.000	의미가 불분명
이유일	2022-01-28	
이유두수	8	의미가 불분명
이유총체	52.800	의미가 불분명
이유평체	6.600	의미가 불분명
재귀발정일	4.0	의미가 불분명