

디딤돌 보고서 초안

[개요](#)

[ERD\(Entity Relationship Diagram\)](#)

[양육비 예측 모델](#)

[개요](#)

[서비스 구상도](#)

[서비스 데이터 파이프라인 구축](#)

[양육비 예측 절차 구상도\(Sequence Diagram\)](#)

[양육비 예측 추론 정의](#)

[데이터 정의](#)

[데이터 전처리](#)

[시스템 구조](#)

[요건](#)

[구현\(스펙 정리\)](#)

[학습 시스템 구조](#)

[추론 시스템 구조](#)

[예측 모델 실험](#)

[LightGBM\(<https://lightgbm.readthedocs.io/en/stable/>\)](#)

[LightGBM0417 결과](#)

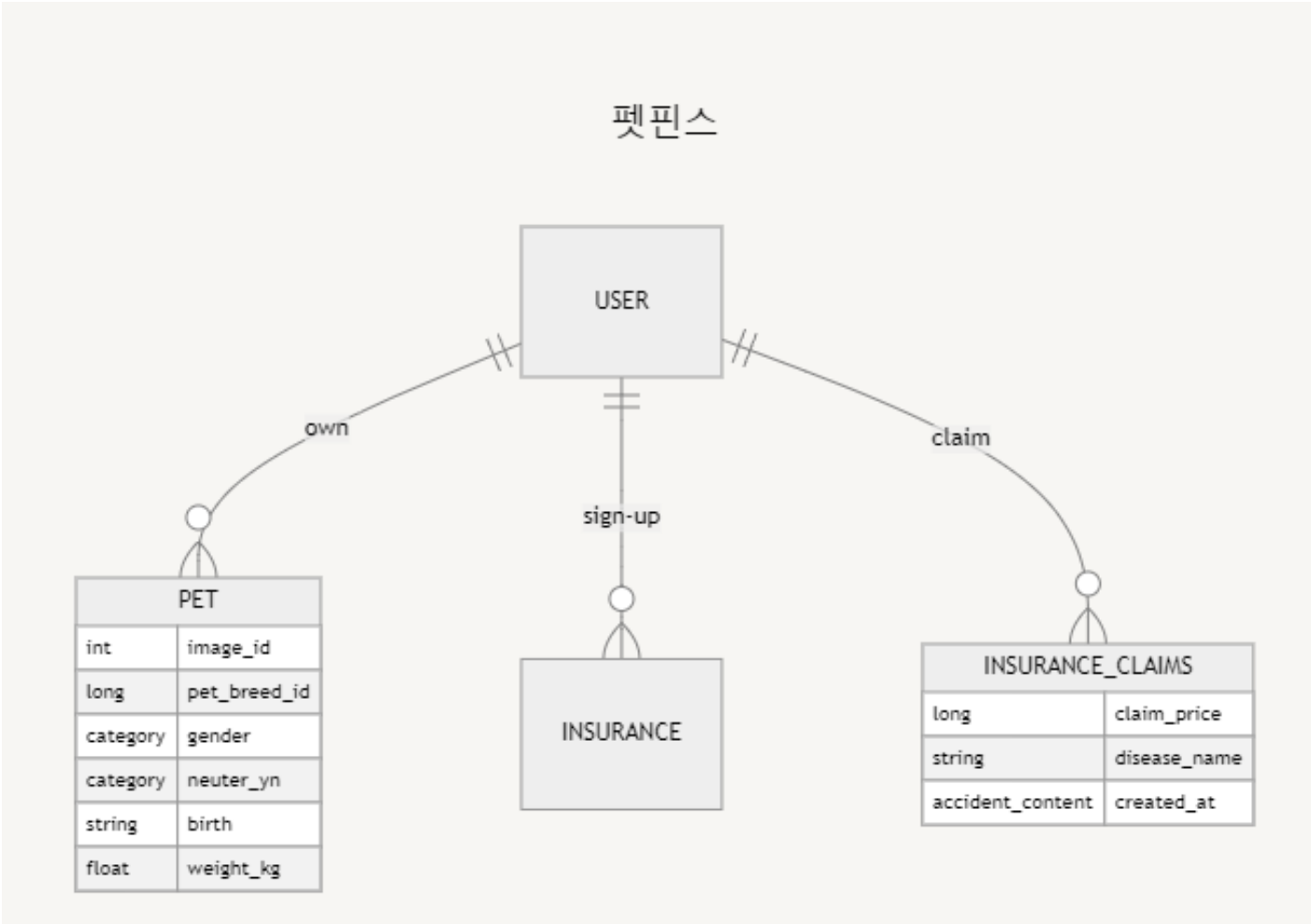
TODO

- ☐ 월 평균 몇 건
- ☐ 양육비 정의 : 현재 데이터 치료비, 배상책임만 포함
- ☐ 추론 시스템 - 모델 로드
- ☐ 시스템 리팩토링
- ☐ 서버 올리기
- ☐ 보험 추천 모델

개요

- 펫핀스는 반려동물 보험 플랫폼 - 월평균 강아지 보험 청구**{# 청구건}**의 데이터가 쌓이는 플랫폼임
- 월평균 {# }건** 쌓이는 보험 청구 데이터를 활용하여 “반려동물의 정보를 기반으로 반려동물을 양육하는데 필요한 금액을 예측하는 머신러닝 모델 서비스”를 제공하고, 나아가 “**알맞는 보험을 추천하는 보험 추천 모델 서비스**”를 제공하고자 함.
 - 개발 모델**
 - 양육비 예측 모델
 - 보험 추천 모델
 - 해당 머신러닝 시스템 개발을 통해 반려동물 보험 가입시 어느 정도의 금액을 절약할 수 있을지에 대한 정보를 제공하여 사용자에게 알맞는 보험가입을 돕고, 나아가 지속적으로 반려동물 관련 금융 데이터를 축적하여 좀 더 통찰력있는 부가서비스들을 창출하는 것이 목표임.
 - 지속적으로 유입되는 데이터를 반영하여 머신러닝 모델의 성능을 개선하기 위해 지속적 학습(Continuous Training)이 가능한 머신러닝 시스템을 구축함
- 펫핀스 시스템 엔티티는 아래의 ERD를 참고
- 펫핀스 양육비 예측 모델은 아래의 Sequence Diagram 참고

ERD(Entity Relationship Diagram)



- 사용자는 0마리 이상의 반려동물을 키운다. 1: optional many (예비 반려인 포함)
- 사용자는 0개 이상의 보험을 가입한다.
- 사용자는 0개 이상의 보험청구를 청구한다.

양육비 예측 모델

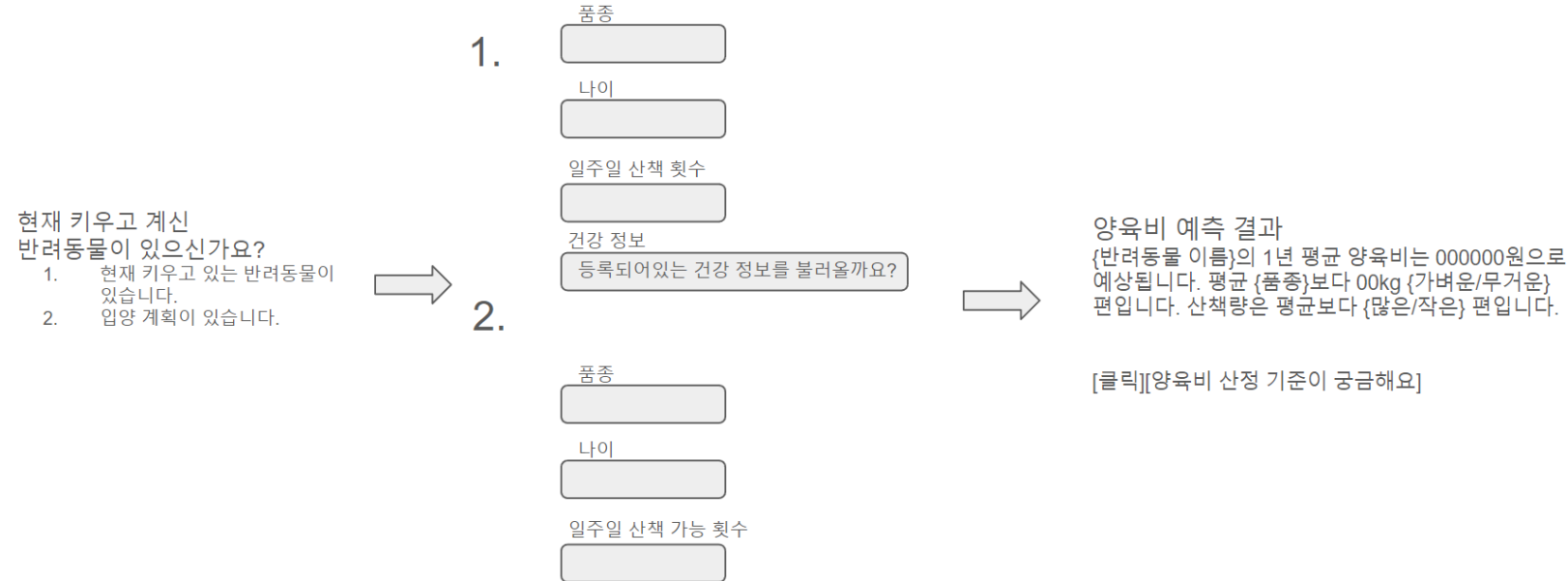
개요

- 반려동물 관련 정보를 기반으로 연평균 양육비를 머신러닝 모델을 활용하여 예측함
- 양육비 정보 : [치료비, 사료 구입, 미용비, 기타]

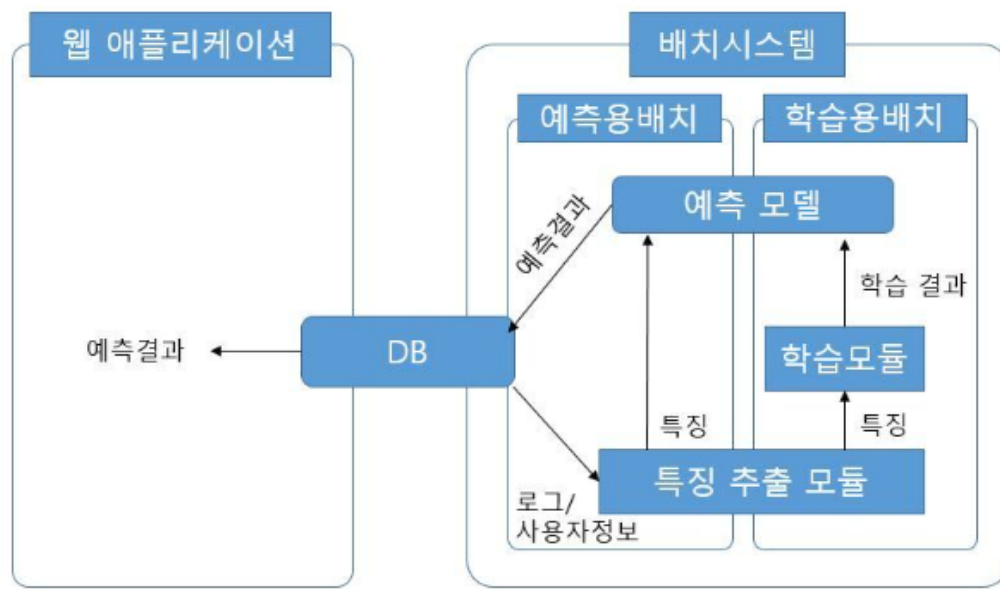
서비스 구상도

양육비 예측

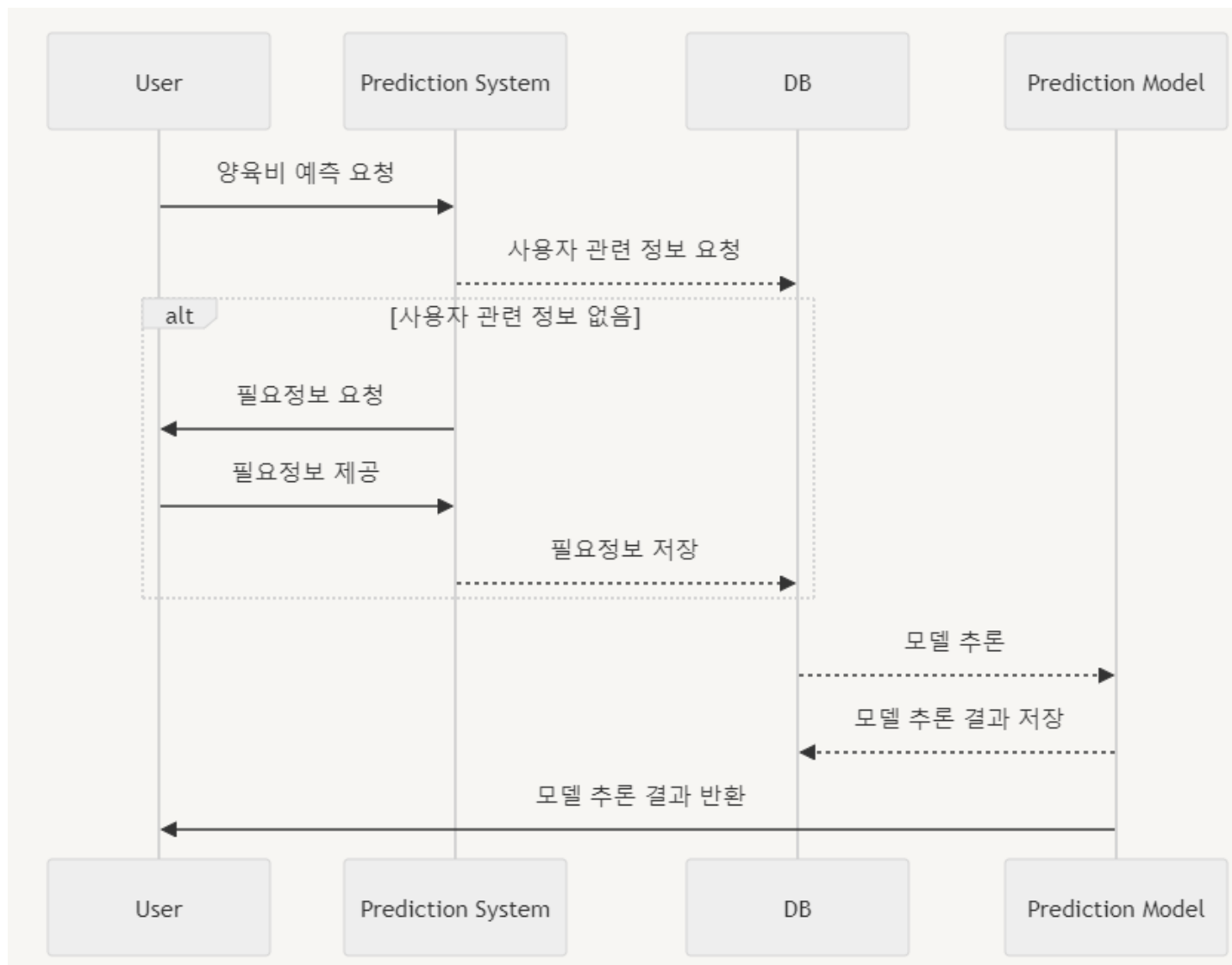
： {반려동물}:강아지, 고양이. 우선 강아지로 종류 제한
예측 서비스 사용 대상: 현재 키우고 있거나, 입양 계획이 있을 경우



서비스 데이터 파이프라인 구축



양육비 예측 절차 구상도(Sequence Diagram)



양육비 예측 추론 정의

- 실시간성(batch X), 단일 추론

데이터 정의

- 자사 플랫폼 펫핀스에서 취득가능한 데이터 정보는 다음과 같음.
 - 반려동물 관련 정보: [품종, 생년월일, 체중, 중성화 여부, 성별]
 - 보험 청구 정보: [청구 유형, 청구 금액, 청구 상세 정보] e.g) [치료비, 217,000원, '슬개골 탈구']

데이터 전처리

- 머신러닝 모델에 각 특성 정보를 반영하기 위해서는 전처리 단계가 필요함.

컬럼명	인코딩 방법
품종(pet_breed_id)	pet_breed_id 갯수 457, 이 중 등장한 것들만 사용하여 인코딩
나이(birth, created_at)	created_at - birth로 나이 측정. days or months? ⇒ days , min_max_scaling
체중(weight_kg)	min_max_scaling
중성화 여부(neuter_yn)	one-hot encoding
성별(gender)	one-hot encoding
청구 금액(claim_price)	각 pet_id에 대해 1년 단위 청구금액 합산

시스템 구조

요건

- 사용자 요청 시, 반려동물 정보를 불러와 단일 건에 대한 실시간 처리가 가능해야함.
- 보험 청구 데이터가 갖는 계절성과, 데이터가 추가적으로 확보될 수 있는 머신러닝 모델의 특성상 지속적인 모델 학습 (Continuous Training)이 가능하도록 설계해야함
- 모델 학습은 데이터가 일정량 축적되면 주기적으로 업데이트하고, 학습된 모델은 실험 결과를 확인하기 쉽게 구성함

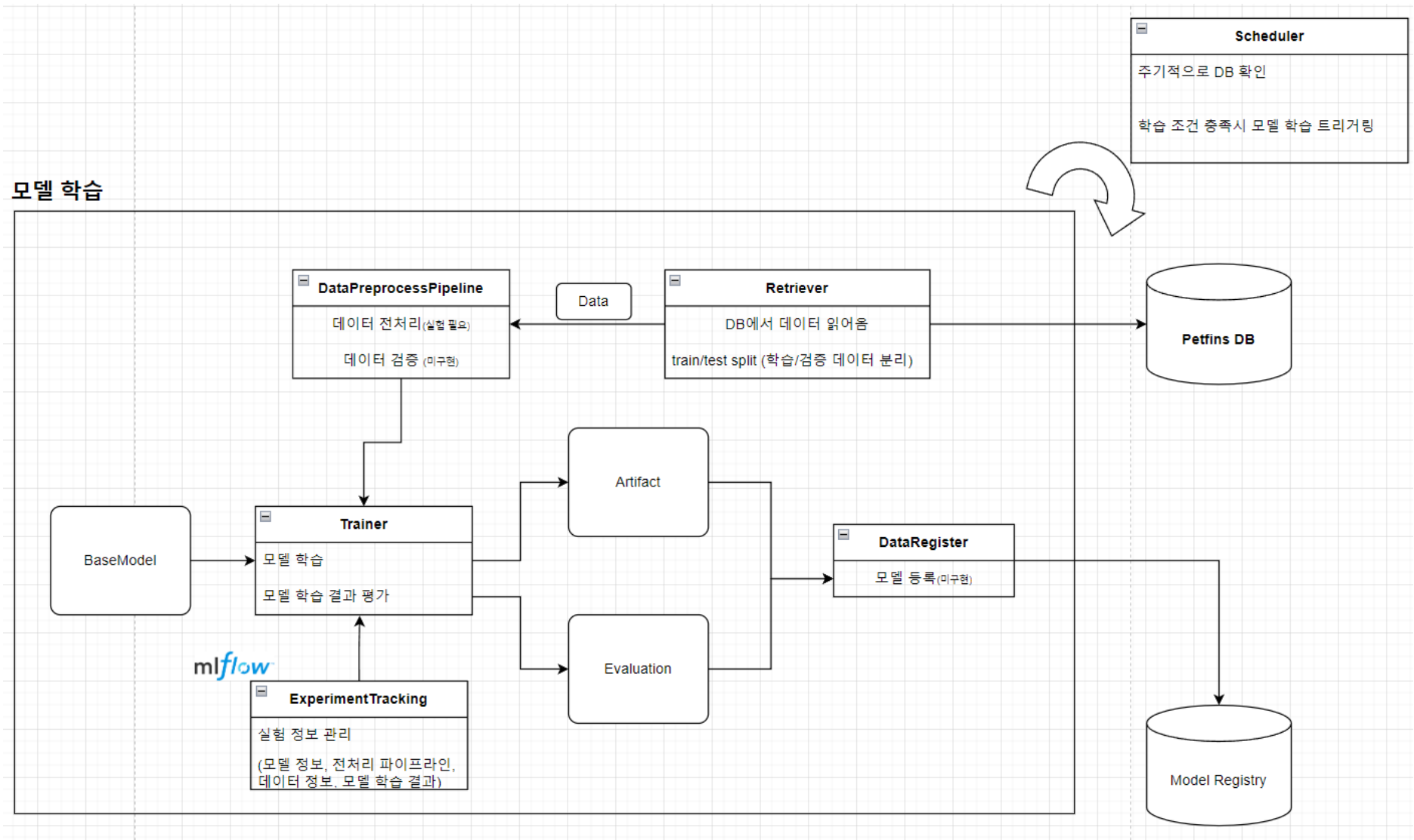
구현(스펙 정리)

- 단일건에 대해 전처리와 모델을 로드해 실시간으로 처리함(최적화 처리 필요)
- 플랫폼 DB를 주기적으로 모니터링하는 스케줄러를 두고, 학습 조건 충족시 모델 학습을 트리거링함
- mlflow와 hydra 사용하여 모델 실험 관리(결과 기록)

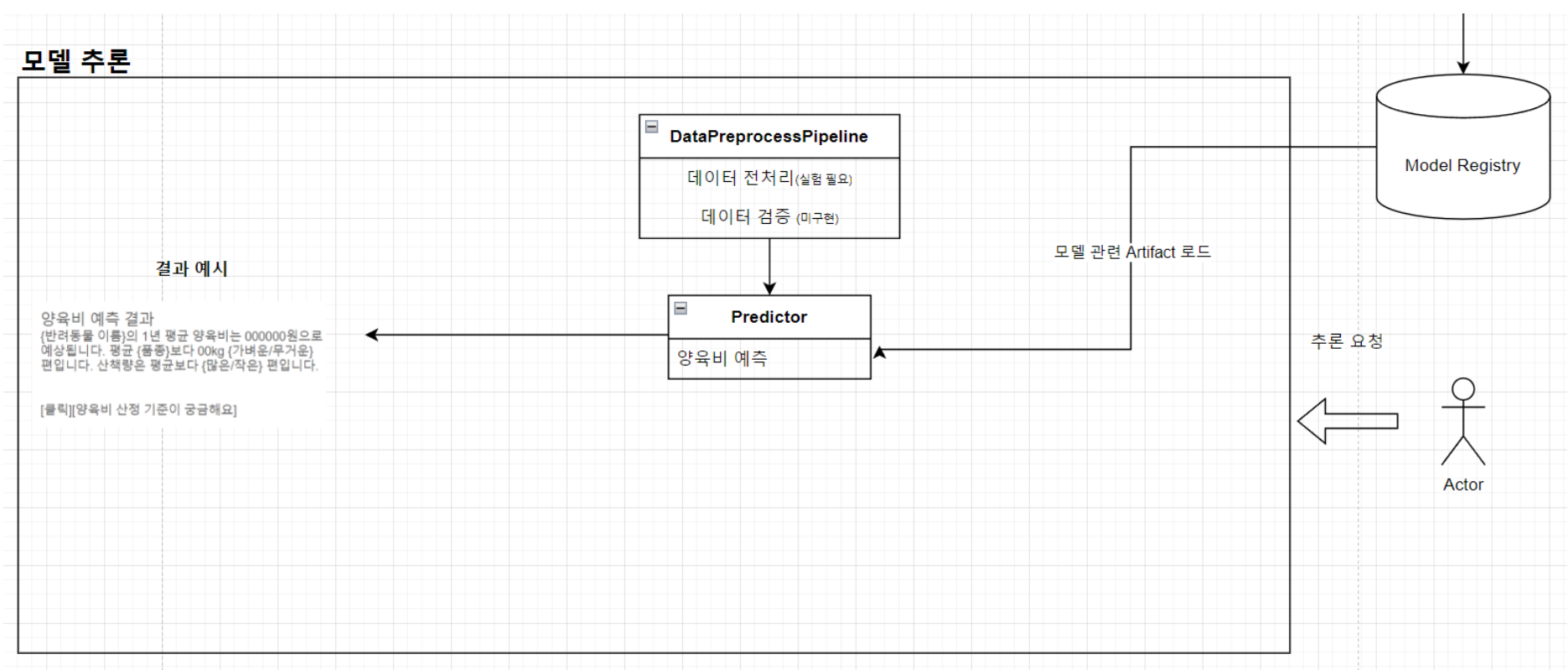
스펙정리

- 언어: python3.10
- 의존성 관리툴: poetry
- 환경: Docker

학습 시스템 구조



추론 시스템 구조



예측 모델 실험

LightGBM(<https://lightgbm.readthedocs.io/en/stable/>)

- 트리 기반의 Gradient Boosting Framework

LightGBM0417 결과

- 학습 데이터 : 44건

- 테스트 데이터: 11건
- 읽는 방법: (첫번째 행)사용자 A는 2341일 중성화하지 않은 무게 10kg의 남자 프렌치불독을 키운다. 이 사용자의 한달 예측 양육비는 38만9천원이었고, 실제 양육비는 11만 1천원이었다.

실제값(원)	예측값(원)	실제값 - 예측값	에러비율 (실제값 - 예측값) / 실제값	성별	중성화 여부	품종	무게(kg)	age(일수)
111000.0	389217.966437	-278217.966437	-2.506468	남자	n	프렌치불독 (1224)	10.0	2341.0
49500.0	371151.275764	-321651.275764	-6.498006	남자	n	말티즈(1109)	3.0	313.0
111000.0	389217.966437	-278217.966437	-2.506468	남자	n	프렌치불독 (1224)	10.0	2341.0
138030.0	389217.966437	-251187.966437	-1.819807	여자	y	프렌치불독 (1224)	10.0	2310.0
56100.0	371151.275764	-315051.275764	-5.615887	남자	y	말티푸(1452)	4.0	344.0
151000.0	389217.966437	-238217.966437	-1.577602	남자	n	골든리트리버 (1301)	43.0	4015.0
35200.0	371151.275764	-335951.275764	-9.544070	여자	y	프렌치불독 (1224)	8.3	776.0
56000.0	371151.275764	-315151.275764	-5.627701	여자	n	비송프리제(1121)	0.9	282.0
70800.0	389217.966437	-318417.966437	-4.497429	남자	n	프렌치불독 (1224)	10.0	2474.0
63800.0	389217.966437	-325417.966437	-5.100595	남자	n	골든리트리버 (1301)	43.0	3739.0
151000.0	389217.966437	-238217.966437	-1.577602	남자	n	골든리트리버 (1301)	43.0	4068.0