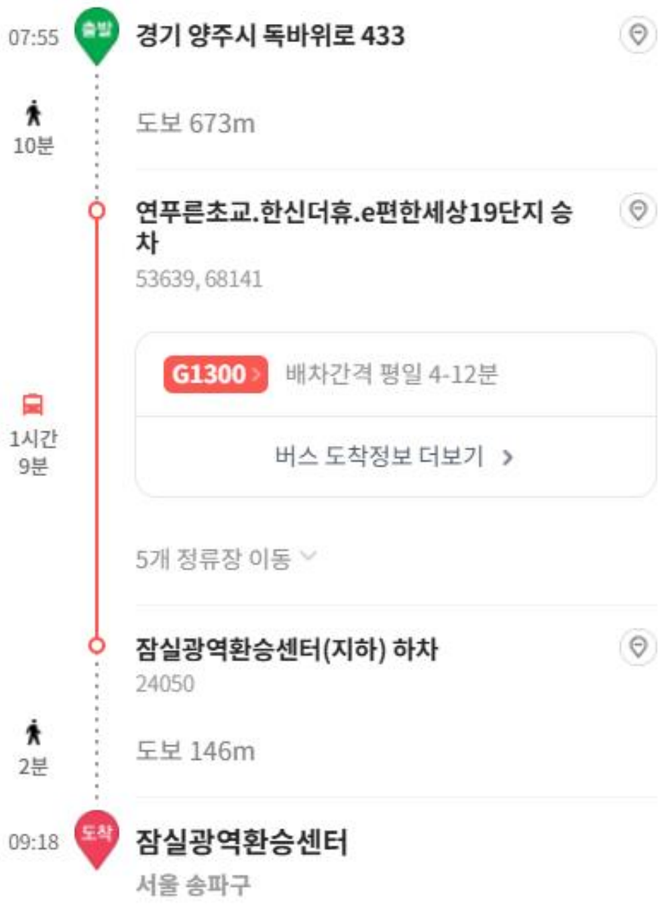
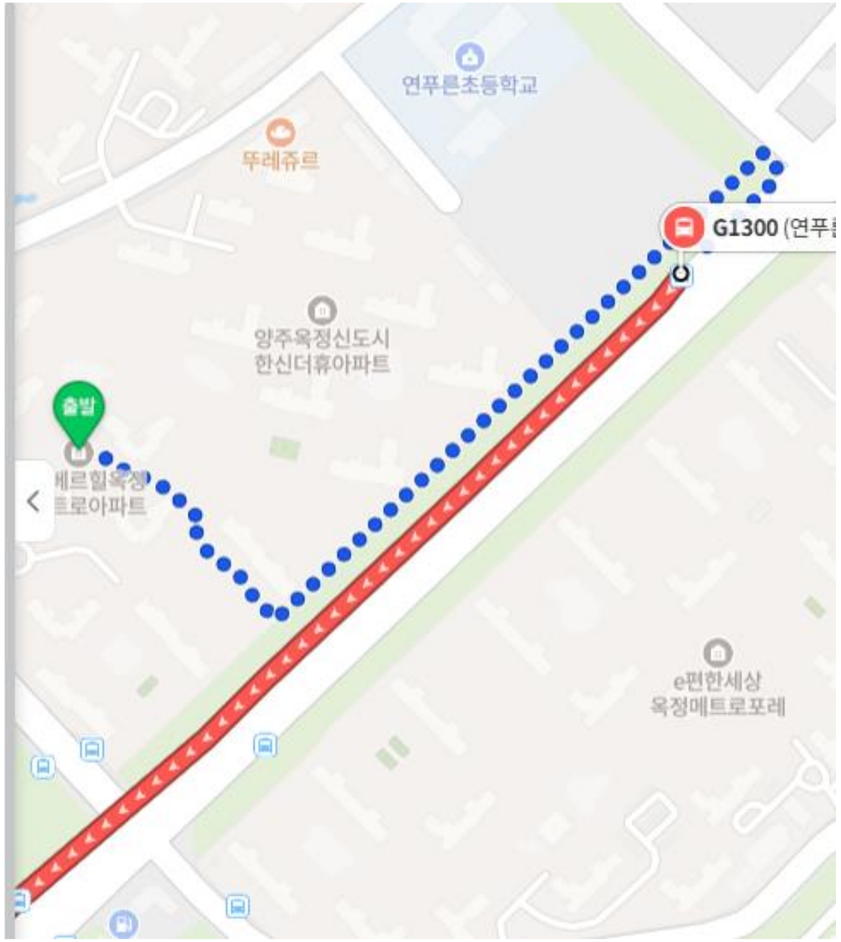
ETA 기반 실시간 대중교통 출발 시점 추천 서비스

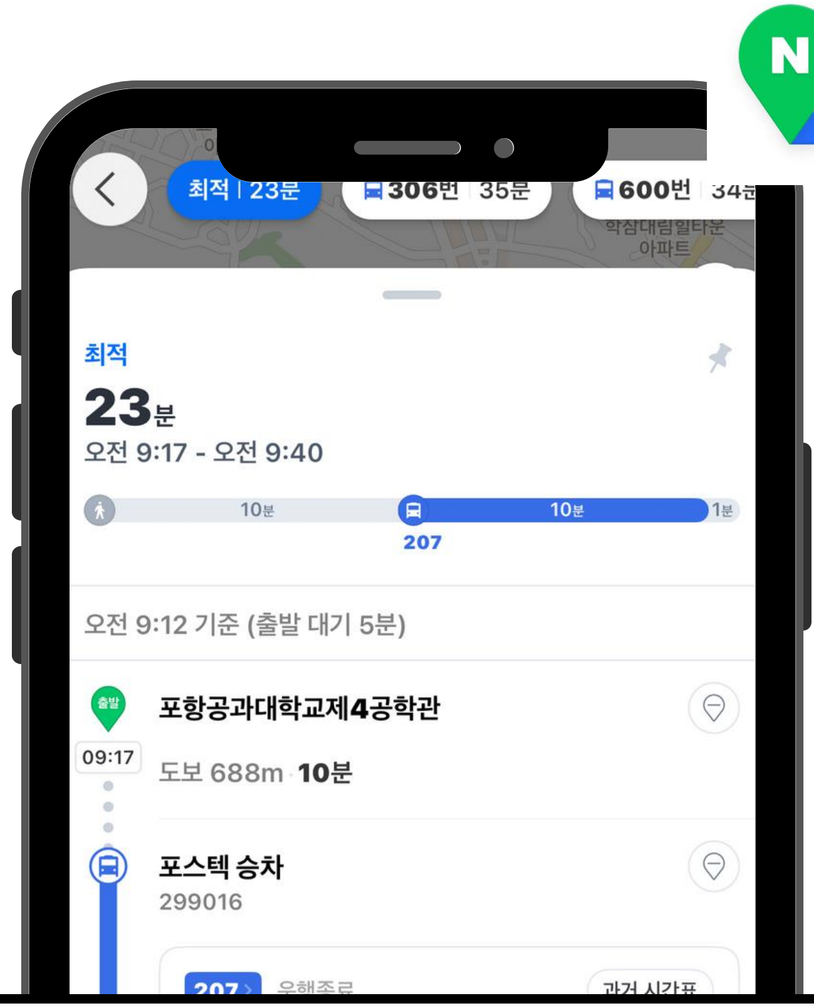
IDSC501-01 Data Engineering
변민우 박주연 김유진

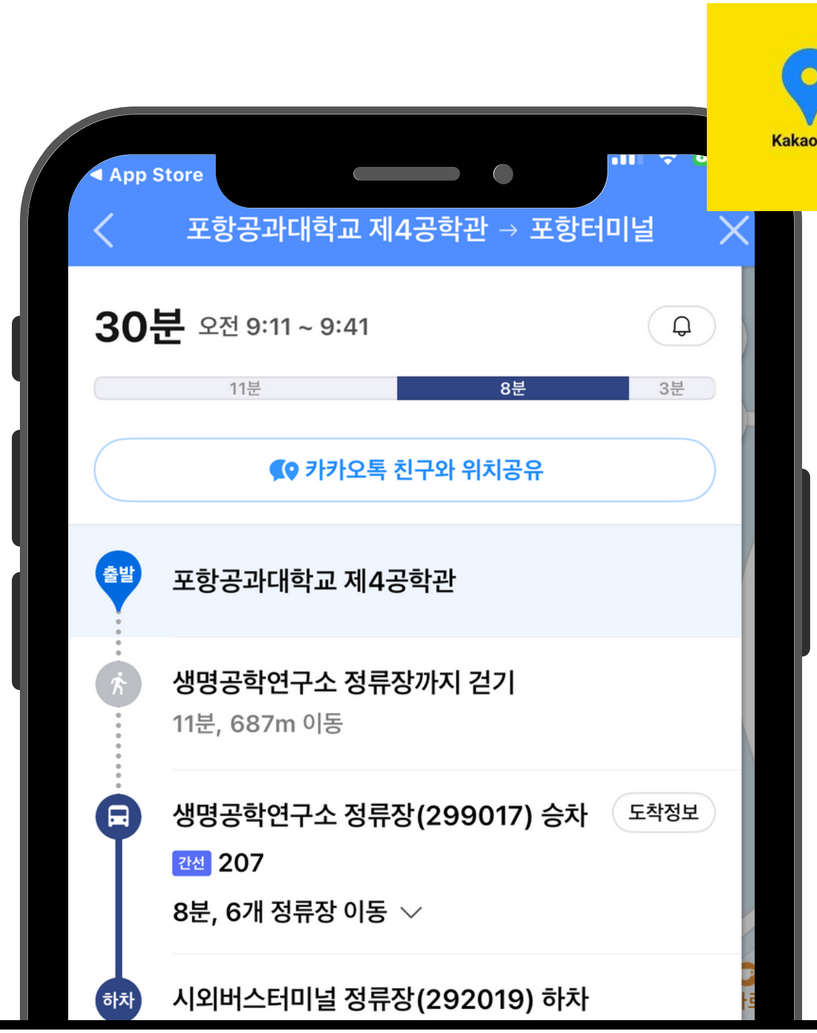




*ETA: Estimated Time of Arrival

1.문제 상황


최적 | 23분
306번 35분
600번 34분
최적
23분
오전 9:17 - 오전 9:40
10분
10분
1분
207
오전 9:12 기준 (출발 대기 5분)
출발
09:17
포항공과대학교제4공학관
도보 688m 10분
포스텍 승차
299016


App Store
포항공과대학교 제4공학관 → 포항터미널
30분 오전 9:11 ~ 9:41
11분
8분
3분
카카오톡 친구와 위치공유
출발
포항공과대학교 제4공학관
생명공학연구소 정류장까지 걷기
11분, 687m 이동
생명공학연구소 정류장(299017) 승차
도착정보
간선 207
8분, 6개 정류장 이동
하차
시외버스터미널 정류장(292019) 하차

1.문제 상황

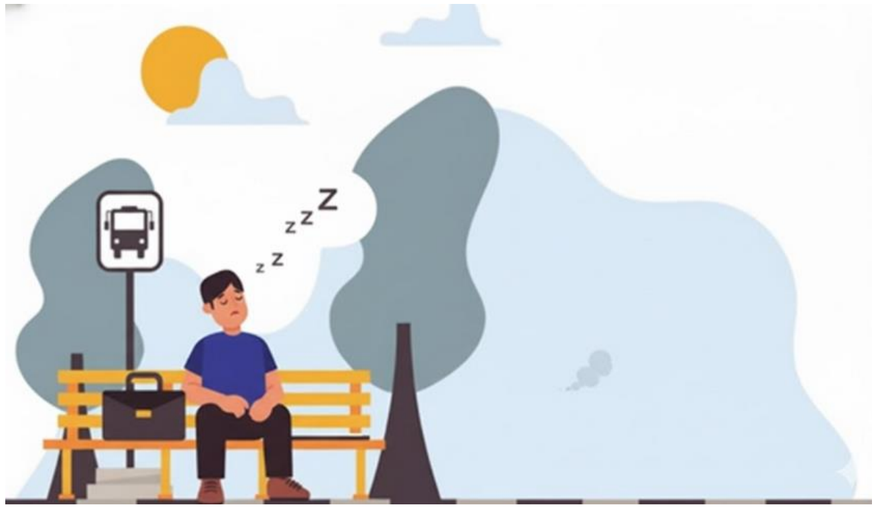




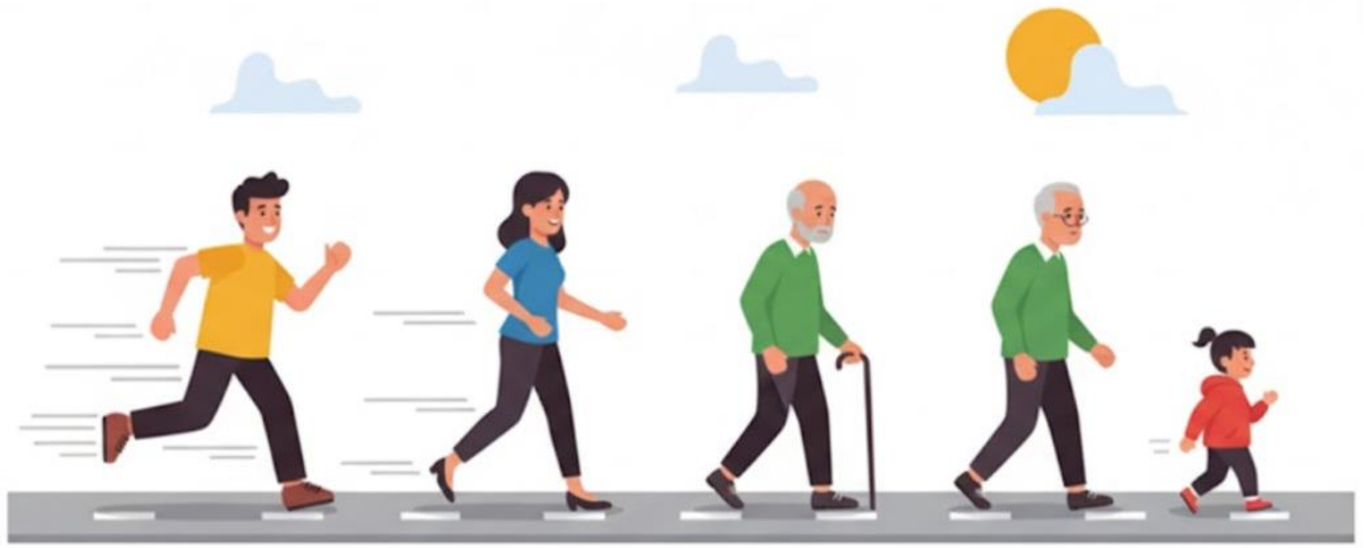
too late

too early





1.문제 상황



2.주제

언제 출발해야 놓치거나 오래 기다리지 않고 탈 수 있을까?

2.주제

“나는”

언제 출발해야 놓치거나 오래 기다리지 않고 탈 수 있을까?

2.주제

“나는”

언제 출발해야 놓치거나 오래 기다리지 않고 탈 수 있을까?

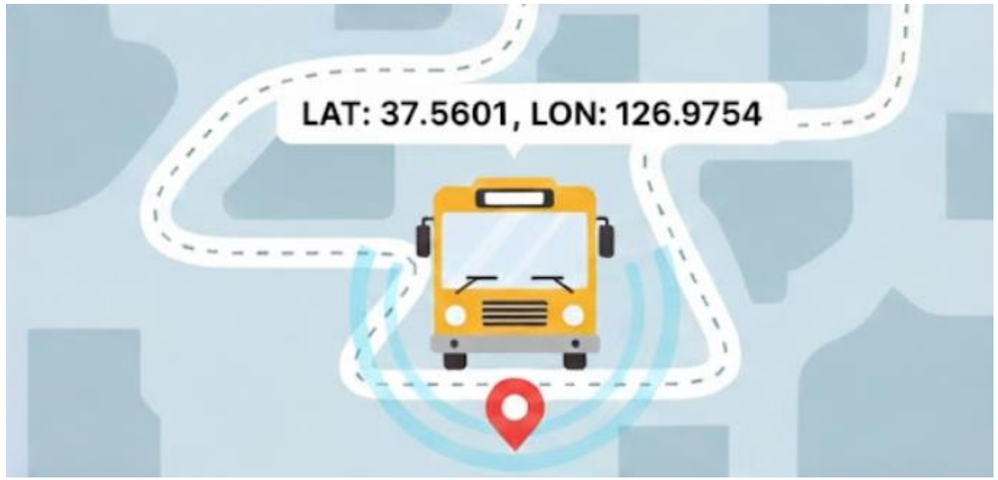
사용자 별로 정확히 언제 출발해야

X% 확률로 제시간에 대중교통을 탈 수 있는지

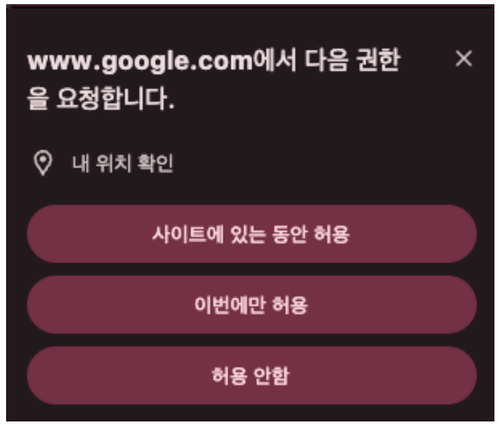
실시간으로 예측하는 서비스

3.데이터 수집

공공데이터 포털
Open API

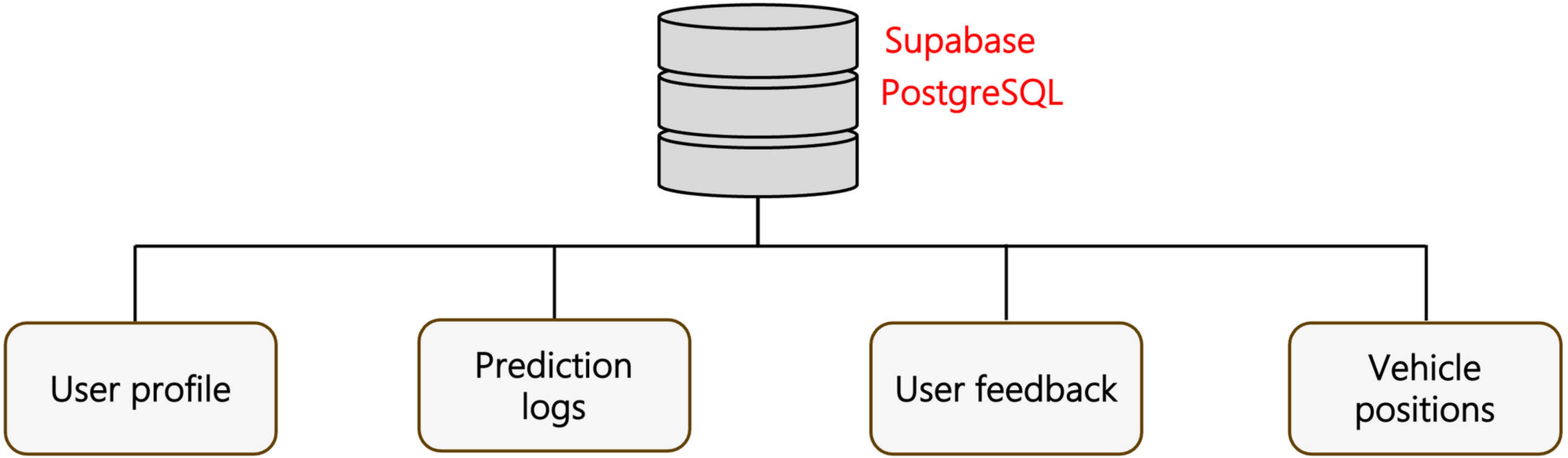


Geolocation API





4.데이터 저장/관리



사용자별 설정/개인화

- 보행 속도
- 100m 기록
- 페이스

예측 요청/결과 로그

- 출발지 / 정류장 / 노선
- walk / vehicle ETA
- miss prob@50/95 등)

유저 피드백

- 실제 도착시간
- 탑승 성공여부
- 추가 대기(초)
  등

실시간/준실시간 차량 위치

- 노선 / 차량ID / 정류장 ID
- 좌표, ETA

4.데이터 저장/관리

- **user_profile**

Columns	Data type	Role
user_id	uuid	각 행을 고유하게 식별
display_name	text	사용자의 표시 이름
walking_speed_ms	numeric	사용자의 평균 보행 속도 (m/s)
pace_style	text	사용자의 보행 스타일
hundred_meter_sec	numeric	사용자의 100m달리기 기록
created_at	timestamptz	프로필이 생성된 시간

4.데이터 저장/관리

- **prediction_logs**

Columns	Data type	Role
id	bigserial	로그 레코드를 고유하게 식별
user_id	uuid	예측 요청을 한 사용자 id
created_at	timestamptz	로그 레코드가 생성된 시간
origin_lat	double precision	예측 요청 시점의 출발 위도
origin_lon	double precision	예측 요청 시점의 출발 경도
stop_id	text	목표 정류장역 고유 ID
route_id	text	노선 ID
walk_eta_sec	integer	출발지에서 정류장까지 걸리는 예상 시간
vehicle_eta_sec	integer	목표 이동수단이 정류장에 도착할 때 까지 걸리는 예상 시간
start_in_sec	integer	추천하는 출발 시간
miss_prob_50	numeric	놓칠 확률이 50% 넘지 않도록 하는 출발 시간 컷오프 값
miss_prob_95	numeric	놓칠 확률이 95% 넘지 않도록 하는 출발 시간 컷오프 값
model_version	text	예측모델 버전

4.데이터 저장/관리

- **user_feedback**

Columns	Data type	Role
id	bigserial	피드백 기록 고유 id
user_id	uuid	사용자 id
prediction_id	bigint	피드백과 관련된 특정 예측기록 id
arrived_at	timestamptz	사용자가 정류장에 도착한 실제 시간
boarded	boolean	탑승 성공 여부
extra_wait_sec	integer	탑승 위한 추가 대기 시간
notes	text	추가 상세 피드백
created_at	timestamptz	피드백 기록 시각

4.데이터 저장/관리

- **vehicle_positions**

Columns	Data type	Role
id	bigserial	위치 기록 고유 식별 번호
collected_at	timestamptz	위치정보 수집 시각
route_id	text	노선 고유ID
vehicle_id	text	탑승수단의 고유 ID
stop_id	text	목적 정류장 ID
lat	double precision	현재 위도
lon	double precision	현재 경도
eta_sec	integer	stop_id에 도착하는 예상 시간

5.데이터 가공

실시간 처리

- 실시간 위치 데이터
- 도보 ETA 계산 (OSMR)
- 추천 출발 시점 계산

데이터 기반 예측 계산

추후 데이터 처리

Daily마다 실행

- 라벨 생성(놓침/초과대기)
- 데이터 피처 엔지니어링
- 모델 학습
- 개인화 파라미터 갱신

개인화를 위한 데이터 & 모델 업데이트

5.데이터 가공

Data Extraction

대중교통 이용 피드백

사용자 입력 정보

Labeling

miss_label = (boarded != True)

excess_wait_min = clip((actual_board_ts - pred_board_ts), 0, 20)

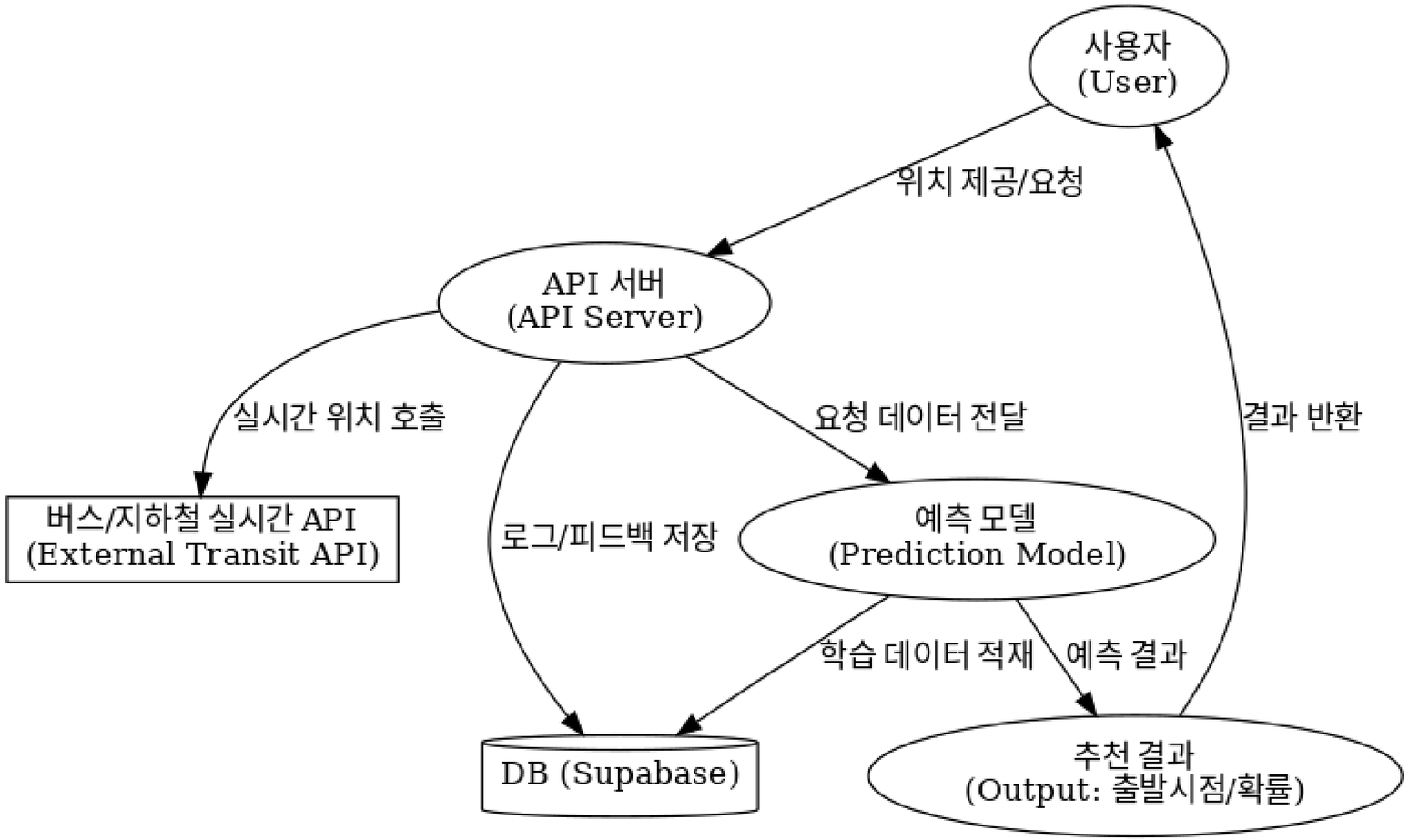
Parameter Updates

walking_speed_mps_new = α·(최근 관측속도) + (1-α)·old

Generate Table

정류장×노선×시간대별 excess_wait_min의 Q50/Q95 계산

전체적인 서비스 흐름


사용자
(User)
위치 제공/요청
API 서버
(API Server)
실시간 위치 호출
요청 데이터 전달
결과 반환
버스/지하철 실시간 API
(External Transit API)
로그/피드백 저장
예측 모델
(Prediction Model)
학습 데이터 적재
예측 결과
DB (Supabase)
추천 결과
(Output: 출발시점/확률)

6.Final 단계까지 처리할 작업

예측 로직 고도화
(ML)

개인화 서비스
수준 향상
(사용자 입력 + 로그)



날씨, 교통량 등의
외부변수 추가

서비스 데모

감사합니다