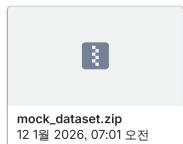


가상 데이터셋 목업 모델

1. 개요

본 문서는 유가보조금 부정수급을 방지하기 위한 탐지 알고리즘의 로직을 상세히 기술합니다. 이 시스템은 DTG(Digital Tachograph) 데이터, 차량 정보, 주유 내역을 종합 분석하여 비정상 패턴을 4단계 등급(HIGH/MEDIUM/LOW/NONE)으로 분류합니다.



CSV 파일들은 데이터베이스 접근 전 임의로 만든 가상의 데이터입니다.

2. 시스템 아키텍처

2.1 전체 처리 흐름

시스템은 다음과 같은 5단계 프로세스로 구성됩니다:

1단계: 데이터 수집 및 전처리

- DTG, 차량정보, 주유내역 데이터 통합

2단계: 개별 탐지

- 5개 탐지 룰 적용 및 플래그 생성

3단계: 신호 강도 분류

- STRONG FLAGS / WEAK FLAGS 구분

4단계: 위험도 등급 산출

- HIGH / MEDIUM / LOW / NONE 등급 결정

5단계: 이상 사유 태깅 및 리포팅

- 탐지된 모든 이상 패턴 기록

2.2 입력 데이터 구성

시스템에서 활용하는 데이터는 다음과 같습니다:

DTG 데이터

- 주요 필드: 일자별 주행거리(km)
- 활용 목적: 실제 운행량 측정

차량정보

- 주요 필드: 평균연비(km/L), 탱크용량(L)

- 활용 목적: 기대 연료량 산출 기준

주유내역

- 주요 필드: 주유량(L), 주유일시, 주유소ID
- 활용 목적: 실제 주유 패턴 분석

3. 핵심 지표 산출 로직

3.1 실제 주행거리 집계

차량별 전체 기간 동안의 주행거리를 합산합니다.

계산식:

```
1 total_distance_km = Σ(차량별 전체 기간 일일 주행거리)
```

처리 방식:

- 차량번호(vehicle_id) 기준으로 DTG 데이터 그룹핑
- 각 날짜별 주행거리를 모두 합산
- 결측치는 0으로 처리

3.2 기대 연료 사용량 계산

실제 주행거리와 차량의 평균 연비를 기반으로 이론적으로 필요한 연료량을 계산합니다.

계산식:

```
1 expected_fuel_liters = total_distance_km ÷ avg_efficiency_km_per_l
```

계산 예시:

- 총 주행거리: 1,000km
- 차량 평균 연비: 5km/L
- 기대 연료량: 200L

이 값은 정상적인 운행 시 필요한 이론적 연료량을 의미합니다.

3.3 허용 오차 범위 설정

시스템은 측정 오차, 운전 습관 차이, 계절적 요인 등을 고려하여 ±10%의 허용 범위를 적용합니다.

계산식:

```
1 expected_low = expected_fuel × 0.9 (하한선)
2 expected_high = expected_fuel × 1.1 (상한선)
```

범위 의미:

- expected_low (정상 범위 하한): 기대량의 90%
- expected_high (정상 범위 상한): 기대량의 110%

4. 개별 탐지 를 상세

시스템은 5개의 독립적인 탐지 룰을 적용하여 다각도로 이상 패턴을 탐지합니다.

4.1 Rule A: 탱크 용량 초과 주유 (OVER_TANK)

탐지 기준:

```
1 flag_over_tank = ANY(주유량 > 차량 탱크용량)
```

주유 내역 중 단 1건이라도 차량의 물리적 탱크 용량을 초과하는 경우 플래그가 발생합니다.

부정수급 가능성:

- 물리적으로 불가능한 주유 = 카드깡 또는 타 차량 주유 확정
- 가장 강력한 부정 신호

탐지 예시:

- 차량 탱크 용량: 60L
- 주유 기록: 75L
- 결과: flag_over_tank = True (즉시 이상 확정)

4.2 Rule B: 하루 4회 이상 주유 (MANY_REFUELS_PER_DAY)

탐지 기준:

```
1 flag_day_over_4 = MAX(일별 주유 횟수) ≥ 4
```

특정 날짜에 4회 이상 주유한 이력이 존재하는 경우 플래그가 발생합니다.

부정수급 가능성:

- 정상적인 운행 패턴에서는 극히 드문 현상
- 소액 쪼개기 결제를 통한 카드 한도 우회 의심
- 여러 차량에 분산 주유 의심

탐지 예시:

2025년 1월 10일 주유 내역:

- 09:00 → 15L
- 11:30 → 20L
- 14:00 → 18L
- 16:45 → 12L
- 총 4회 주유 발생
- 결과: flag_day_over_4 = True

4.3 Rule C: 특정 주유소 집중도 80% 이상 (STATION_CONCENTRATION)

탐지 기준:

```
1 flag_station_conc_80p = (최다 이용 주유소 횟수 ÷ 전체 주유 횟수) ≥ 0.8
```

전체 주유의 80% 이상이 단일 주유소에서 발생한 경우 플래그가 발생합니다.

부정수급 가능성:

- 주유소 사업자와 차주 간 공모 가능성
- 허위 주유 또는 금액 부풀리기 의심
- 단독으로는 약한 신호 (정상적인 단골 이용 패턴일 수도 있음)

탐지 예시:

전체 주유 20회 중:

- A 주유소: 17회 (85%)
- B 주유소: 2회
- C 주유소: 1회
- 결과: flag_station_conc_80p = True

4.4 Rule D: 기대 연료량 초과 (FUEL_OVER_EXPECTED)

탐지 기준:

```
1 flag_fuel_over_expected = actual_fuel > expected_high
```

실제 주유량이 허용 상한(expected_high)을 초과하는 경우 플래그가 발생합니다.

부정수급 가능성:

- 주행거리 대비 연료 사용량 과다
- 주유한 연료의 타 용도 전용 의심 (재판매, 타 차량 사용 등)

탐지 예시:

- 총 주행거리: 1,000km
- 차량 연비: 10km/L
- 기대 연료량: 100L
- 허용 상한: 110L
- 실제 주유량: 150L
- 결과: flag_fuel_over_expected = True

4.5 Rule E: 기대 연료량 미달 (FUEL_UNDER_EXPECTED)

탐지 기준:

```
1 flag_fuel_under_expected = (actual_fuel > 0) AND (actual_fuel < expected_low)
```

실제 주유량이 허용 하한(expected_low) 미만이지만 0이 아닌 경우 플래그가 발생합니다.

부정수급 가능성:

- DTG 주행거리 조작(부풀리기) 의심
- 또는 개인 자비 주유 병행 (부정수급과 무관할 수도 있음)
- 약한 신호 (참고 지표로만 활용)

탐지 예시:

- 총 주행거리: 1,000km
- 차량 연비: 10km/L
- 기대 연료량: 100L
- 허용 하한: 90L
- 실제 주유량: 60L
- 결과: flag_fuel_under_expected = True

5. 신호 강도 분류 체계

탐지된 플래그는 신호의 강도에 따라 STRONG과 WEAK 두 가지 범주로 분류됩니다.

5.1 STRONG FLAGS (강한 이상 신호)

다음 3개 플래그는 강한 부정수급 신호로 분류됩니다:

- flag_over_tank (탱크 용량 초과)
- flag_day_over_4 (하루 4회 이상 주유)
- flag_fuel_over_expected (기대 연료량 초과)

특징:

- 이 중 하나라도 True면 실질적 부정수급 가능성이 높음
- 단독으로도 조사 대상 선정 근거가 됨

5.2 WEAK FLAGS (약한 의심 신호)

다음 2개 플래그는 약한 의심 신호로 분류됩니다:

- flag_station_conc_80p (특정 주유소 집중)
- flag_fuel_under_expected (기대 연료량 미달)

특징:

- 단독으로는 부정수급 확정 불가
- STRONG FLAGS와 결합 시 위험도 상승
- 정황 증거 또는 보조 지표로 활용

5.3 anomaly_confirmed (이상 확정 여부)

```
1 anomaly_confirmed = ANY(STRONG_FLAGS) == True
```

결과 해석:

- True: 실질적 이상 패턴 확정, 조사 필요
- False: 경미한 의심 또는 정상 범위

6. 위험도 등급 분류 알고리즘

6.1 등급 결정 로직

시스템은 발생한 플래그의 조합에 따라 4단계 위험도 등급을 부여합니다.

등급 결정 알고리즘:

1. STRONG FLAGS가 있고 집중도 플래그도 있는 경우 → **HIGH** 등급
2. STRONG FLAGS만 있는 경우 → **MEDIUM** 등급
3. 집중도 플래그만 있는 경우 → **LOW** 등급
4. 어떤 플래그도 없는 경우 → **NONE** 등급

6.2 등급별 의미 및 권장 조치

HIGH 등급

- 조건: STRONG FLAGS + 집중도 플래그

- 의미: 복합적 부정 패턴, 공모 가능성 높음
- 권장 조치: 즉시 현장 조사, 보조금 지급 중단 검토

MEDIUM 등급

- 조건: STRONG FLAGS만 존재
- 의미: 명백한 이상 패턴 존재
- 권장 조치: 정밀 데이터 분석, 소명 요청

LOW 등급

- 조건: 집중도 플래그만 존재
- 의미: 경미한 의심 수준
- 권장 조치: 모니터링 강화, 패턴 추적

NONE 등급

- 조건: 플래그 없음
- 의미: 정상 범위
- 권장 조치: 통상 처리

6.3 등급별 실제 사례

HIGH 등급 사례:

차량 A의 탐지 내역:

- 탱크 용량 60L인데 80L 주유 기록 발생 (OVER_TANK)
- 전체 주유의 90%가 특정 주유소에서 발생 (STATION_CONCENTRATION)
- 판정: 주유소 공모 카드깡 의심, 즉시 조사 필요

MEDIUM 등급 사례:

차량 B의 탐지 내역:

- 특정 날짜에 하루 5회 주유 (MANY_REFUELS_PER_DAY)
- 주유소는 여러 곳에 분산되어 있음
- 판정: 소액 쪼개기 의심, 공모는 불명확하나 조사 필요

LOW 등급 사례:

차량 C의 탐지 내역:

- 모든 주유가 집 근처 단골 주유소에서 발생 (STATION_CONCENTRATION)
- 다른 이상 신호 없음
- 판정: 정상적 이용 패턴일 가능성 높음, 모니터링만 실시

7. 이상 사유 태깅 시스템

7.1 사유 코드 생성 로직

시스템은 발생한 모든 플래그를 문자열로 기록하여 추적 가능성을 확보합니다.

생성 프로세스:

1. 빈 리스트 생성

2. 각 플래그가 True인 경우 해당 코드 추가:

- flag_over_tank → "OVER_TANK"
- flag_day_over_4 → "MANY_REFUELS_PER_DAY"
- flag_fuel_over_expected → "FUEL_OVER_EXPECTED"
- flag_station_conc_80p → "STATION_CONCENTRATION"
- flag_fuel_under_expected → "FUEL_UNDER_EXPECTED"

3. 파이프(|) 문자로 연결하여 최종 문자열 생성

7.2 복합 사유 표현 예시

여러 플래그가 동시에 발생한 경우:

```
1 "OVER_TANK|FUEL_OVER_EXPECTED|STATION_CONCENTRATION"
```

이 방식의 장점:

- 한눈에 모든 이상 패턴 파악 가능
- 조사 우선순위 판단 용이
- 후속 분석 시 특정 패턴만 필터링 가능
- 통계 분석 및 추세 파악 용이

8. 데이터 흐름 다이어그램

시스템 내 데이터의 흐름은 다음과 같습니다:

입력 단계:

- DTG 데이터 (주행거리) 수집
- 차량정보 (연비, 탱크용량) 수집
- 주유내역 (주유량, 패턴) 수집

처리 단계:

- DTG + 차량정보 → 기대 연료량 계산
- 기대 연료량 → 허용 범위 산출 ($\pm 10\%$)
- 주유내역 + 허용 범위 → 실제값 vs 기대값 비교

분석 단계:

- 5개 탐지 를 적용 → 개별 플래그 생성
- 플래그 → STRONG/WEAK 분류
- 분류 결과 → 위험도 등급 산출

출력 단계:

- 최종 등급 + 이상 사유 태깅
- 조사 대상 목록 생성