

# AIS 선박 데이터 메타데이터

## 1. 개요

이 문서는 2020년 3월부터 2021년 2월까지 대한민국에서 수집된 AIS(자동 식별 시스템) 데이터의 메타데이터로, 데이터는 한국 표준시(KST) 기준 1초 간격으로 기록돼 있음. 추가로, 데이터의 범위는 VTS(관제센터) 기준으로 한반도 전체의 선박을 포함함.

## 2. 데이터 스키마

AIS 데이터셋은 다음과 같은 속성들로 구성되어 있음.

칼럼명	데이터 타입	설명
mmsi	TEXT	선박의 고유 식별자(해상 이동 무선 식별번호)
timestamp	timestamp	AIS 신호 전송의 날짜와 시간(KST, 초 단위)
latitude	REAL	선박위 북위 좌표(WGS-84 기준)
longitude	REAL	선박의 동경 좌표(WGS-84 기준)
sog	REAL	대지 속도(노트 단위)
cog	REAL	대지 방위각(0-360도)
heading	REAL	선박의 선수 방향(진북 기준, 0-360도)

## 3. 추가 메타데이터 고려 사항

- 가. 공간 범위 : 대한민국 영해
- 나. 시간 범위 : 2020년 3월 1일 - 2021년 2월 28일
- 다. 시간대 : 한국 표준시(UTC+9)
- 라. 데이터 빈도 : 최소 1초 간격, 동일 시간대의 다른 MMSI 데이터 존재 가능
- 마. 좌표 참조 시스템 : WGS-84
- 바. 데이터 출처 : 대한민국 해역에서 운항하는 선박의 AIS 트랜스폰더

## 4. 데이터 품질 및 전처리

- 가. 데이터 정제 : 중복되거나 오류가 있는 기록 제거 (예: 유효하지 않은 위도, 경도)
- 나. 이상치 탐지 : 통계적 방법(예: 박스 플롯) 및 IQR을 활용한 비정상적인 SOG 및 COG 값 탐지
- 다. 데이터 저장 형식 : 효율적인 쿼리 및 분석을 위해 PostgreSQL RDBMS 사용

## 5. 적용 분야

- 가. 자동 경로 예측 : 실시간 경로 예측을 위한 딥러닝 모델 개발
- 나. AIS 데이터 마이닝 : 주기, 분기별로 선박 AIS 데이터의 패턴 발굴