



# 지진 기록 탐색적 데이터 분석 및 예측 모델링 (Earthquake EDA and Prediction)

2025년 7월 31일

---

작성자 김동훈

## 개요

역사상 기록된 지진 활동들 중 수치적으로 활용이 가능한 자료를 최대한 활용하여 탐색적 데이터 분석(EDA)을 통하여 지난 지진 활동에 대한 데이터를 정량화, AI 모델 학습용 정제, 시각화 작업을 시행한 이후, 미래의 지진 활동 예측 모델을 설계한다. AI 모델을 통해 도출된 지진 예측 정보는 웹 페이지를 통하여 실시간으로 시각화된 자료로 제공할 수 있도록 하여 재해에 선제적으로 대비할 수 있도록 돕는 역할을 한다. 여기에 새로운 이론을 적용하여 지진 예측을 보다 정확하게 수행할 수 있도록 서브 프로젝트를 진행하는데, 이는 과거의 지진 자료들을 탐색적 데이터 분석을 통해 단순히 지진의 발생 지역과 시기, 주기, 규모 등의 일반적인 지진 측정 데이터 뿐만 아니라 진원지 주변의 자기장 변화와 지각의 온도 변화 데이터를 추가하여 상관관계를 이해하고 최종적으로 이 이론을 논문으로 작성한다.

## 목표

1. 실제 발생했던 지진 기록들을 탐색적 데이터 분석을 통해 시각화 및 예측에 필요한 유의미한 자료로 재구성한다.
2. AI 기술을 활용하여 지진을 예측할 수 있는 모델을 개발하고, 새로운 이론을 적용하여 지진 발생과의 상관관계에 대해서 이해한다.
3. 예측 가능한 지진의 발생 시기와, 규모, 위치 등의 정보를 제공하는 사이트를 배포한다.
4. 진원지의 자기장 변화와 지각의 온도 변화 데이터와 지진 발생의 상관관계를 명확히 하여 논문으로 작성한다.

## 기대효과

역사적으로 기록된 지진 활동의 시기와 주기, 발생 지역, 지진의 규모, 피해 규모 등의 자료들을 탐색적 데이터 분석 방법을 통하여 수치적, 시각적으로 활용이 용이하게 만들고, AI 지진 예측 모델을 통해 지진을 미리 예측하여 피해를 최소화 할 수 있으며, 새로운 이론을 적용하여 지진 예측의 정확도를 높일 수 있다.

## 탐색적 데이터 분석 (EDA)

데이터 분석을 통해 보다 정량화된 데이터와 시각화 자료들을 제공하여 지진 예측을 위한 AI 모델 학습용 데이터셋을 제공할 수 있을 뿐만 아니라, 정제된 데이터 자체만을 활용하여 지진 연구에 활용하거나 교육용으로 사용할 수 있다.

## 지진 예측 AI 모델링

AI 모델을 통하여 주어진 데이터를 기반으로 하여 미래에 발생 가능한 지진의 위치와 시기, 규모 등을 선제적으로 파악하고 대비할 수 있도록 하여, 해당 지역 및 주변 지역의 피해를 최소화할 수 있다.

## 새로운 이론의 도입

이전의 지진 연구와 예측에서 일반적으로 사용되었던 지진의 발생 주기만으로는 정확한 지진 발생 시기와 규모 등을 파악하기 힘들었지만, 여기에 지자기장 변화와 지각 온도 변화라는 새로운 데이터를 고려 요소로 적용하여 지진과의 상관관계 분석을 통하여 보다 정확한 지진 예측에 기여한다. 새로운 이론의 적용이 지진 예측과 상관관계가 전혀 없다고 하더라도 해당 분야의 연구에서 오답을 소거하는데 기여할 수 있다.

## 마일스톤

### I. 데이터 수집 및 정제

공개 데이터셋 (예: USGS Earthquake Catalog, NOAA NCEI 등)을 활용하여 데이터셋을 확보하고, 필요에 맞게 정제한다.

지자기장/지열 데이터를 수집하여 전처리 및 정제 과정을 실시한다.

전처리된 데이터는 CSV파일로 산출되도록 하고, 데이터 정의서를 작성한다.

### II. EDA 및 시각화 결과 도출

1차적으로 정제된 데이터를 탐색적 데이터 분석을 통해 '지진 예측'이라는 목적에 맞게 2차 정제를 실시한다. 이후 데이터를 시각화하여 데이터의 분포 및 분석이 용이하게 한다.

해당 과정을 수행한 후, 분석 리포트와 시각화 그래프를 산출물로 제출한다.

### III. AI 모델 구축 및 검증

지진 예측에 AI 모델링 작업을 시작한다. LSTM, Random Forest, Transformer 등의 다양한 모델을 활용하여 정제된 데이터를 학습시키고, 결과를 비교 분석한 후 모델 성능 평가를 통해 최적의 모델을 선정한다.

모델 코드와 결과 리포트를 작성한다.

### IV. 지자기장/지열 변수 적용 및 이론 분석

지자기장과 지열 변수를 포함하여 모델을 재학습 시키고, 상관관계를 분석한다.

모델 개선 리포트를 작성한다.

### V. 웹 서비스 연동 및 시각화 시스템 개발

웹 기반 시각화 페이지를 구축하고, 실시간 데이터 연동을 시도한다.

웹 데모 버전과 사용자 메뉴얼 등을 작성한다.

### VI. 지자기장/지열 이론 분석 및 논문화 시도

데이터셋 정제와 EDA과정, AI 모델링과 모델 재학습 과정을 모두 포함하여 연구 보고서 및 프로젝트 문서를 정리한다. 또한, 지자기장과 지열 변수를 포함한 지진 예측의 전반적인 프로세스와 이론 분석 및 상관관계를 포함한 논문을 작성한다.

산출물로는 논문 초안, 포스터, 발표자료 등이 있다.

## AI 모델 설계 예시

### 목표

입력 데이터: 시계열 기반 지진 관측값, 위치 정보, 지자기장 및 지열 지표

예측 대상: 미래 1개월 내 발생 가능한 지진의 규모 및 위치

### 모델 구조

입력: 시계열 특성 (시간, 진도, 위치 등), 공간적 특성 (위도/경도), 지자기/지열 변화 지표

→ 다층 LSTM 혹은 GRU

→ Dense Layer

출력: 발생 가능성, 예상 위치, 규모

(대안 모델)

- Random Fores 혹은 XGBoost: 해석력 강화
- Autoencoder + 이상치 탐지: 비정상적인 패턴 예측 (지진의 전조 탐지)
- Hybrid CNN-LSTM: 지표 변화의 시각적 패턴 반영 가능