

빅데이터 분석 웹 서비스 개발 프로젝트 중 데이터 분석 및 예측 프로젝트 세부 사항

작성자: 한상곤, tykim@pusan.ac.kr
문서 버전 : v0.55

1. 데이터 분석 및 예측 프로젝트 개요(Overview)

- 1) <빅데이터 분석 - 웹 서비스 개발 프로젝트(이하 프로젝트)>는 3가지 구성 요소인 프론트 엔드(Front-End, FE), 백엔드(Back-End, 이하 BE) 그리고 데이터 분석(Data Analysis, 이하 DA)으로 진행됩니다. DA의 경우 별도의 프레임워크(Framework)가 존재하지 않아서 다음과 같은 간소화된 데이터 분석 및 예측 프레임워크를 적용하도록 하겠습니다.

과정	목표	도구
1. 데이터 정의	데이터를 확인 후 문제해결에 필요한 데이터 분석 방법 및 과정을 문서화	- Google Docs - Markdown
2. 데이터 전처리	문제 해결을 위해서 불필요한 데이터 및 누락 데이터를 처리하여 모델에 적용할 수 있도록 전처리를 진행	- Pandas
3. 모델 작성 및 평가	전처리된 데이터를 사용해서 예측 결과를 제공할 수 있는 모델을 작성하고 평가를 통해서 모델의 성능 지표를 확인	- Scikit-Learn - TensorFlow
4. 반응형 시각화	제공된 데이터의 특성과 특징을 공유하기 위해서 간단한 형태의 반응형 웹 시각화를 작성성	- Plotly
5. REST API 제공	예측 모델을 활용할 수 있도록 간단한 API를 제공	- Flask

- 2) DA의 경우 다음과 같은 2가지 목표로 진행할 예정입니다. 아래 2가지 목표는 DA에게 요구되는 핵심적인 능력이자 지표로 정의하였습니다. 해당 사항에 대한 내용은 “3. 데이터 분석 및 예측 프로젝트에 새롭게 도입되는 기술”에서 논의하고 있습니다.

<ul style="list-style-type: none">- 시각화- 예측 결과를 제공하는 REST API 제공

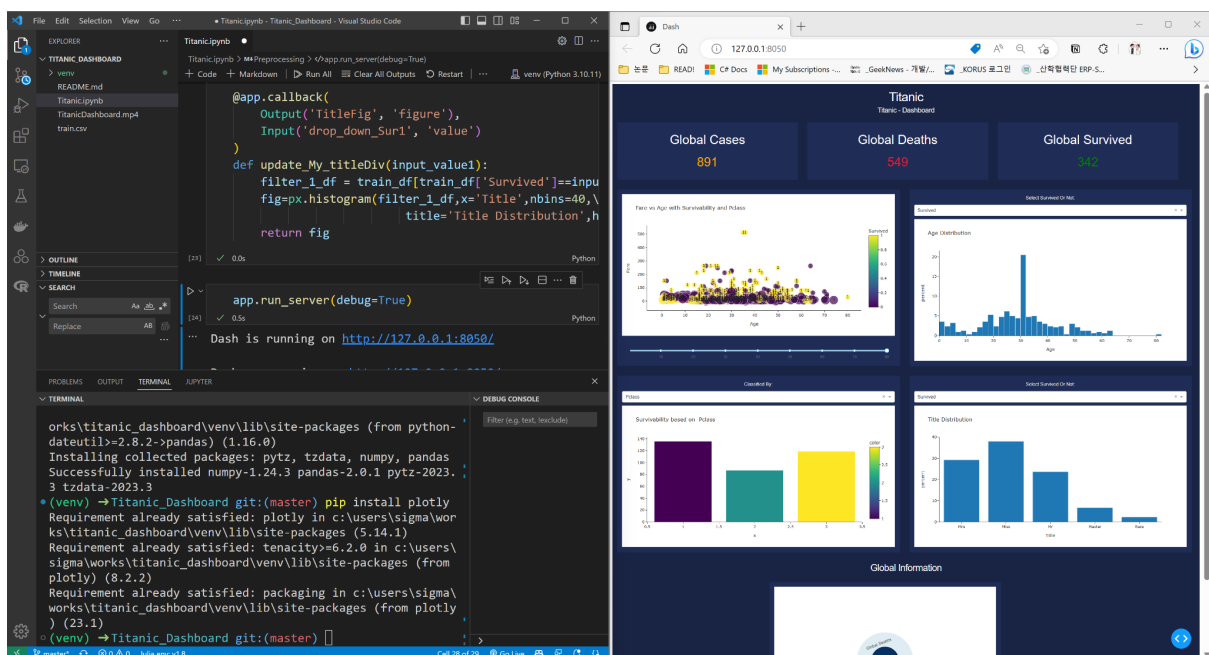
2. 점진적인 개선과 빠른 진행

프레임워크에서 제시하는 절차(1~5)를 1회 반복하는 것을 수행횟수(Iteration, 이터레이션)라 정의하고, 프로젝트는 2~3번의 Iteration을 진행하여 점진적인 개선을 빠르게 진행하는 것을 DA 프로젝트의 핵심적인 진행 방향입니다.

3. 데이터 분석 및 예측 프로젝트에 새롭게 도입되는 기술

3.1 시각화

- 1) 우리는 프로젝트를 진행함에 있어서 DA의 가장 핵심적인 능력은 다른 영역의 개발자들과 원활한 소통이 중요하다고 정의하였습니다.
- 2) 데이터 전처리 및 분석 결과를 시각적으로 표현하여 다른 구성원과 의견을 교류하고 분석 과정을 효율적으로 할 수 있는 기술을 적용하기로 하였습니다. 이를 위해서 시각적인 표현 능력 뿐만 아니라 데이터를 다양한 형태로 조작할 수 있는 도구가 필요하다고 판단하여 [plotly](#) 이용한 웹 기반 데이터 시각화 기술을 활용하기로 하였습니다.
- 3) [plotly](#)는 웹 기반의 간단한 반응형 데이터 시각화 기술이며, 무료로 활용할 수 있습니다. 현재는 Meta(전 Facebook)에서 제공하는 시계열 제품인 [Prophet](#) 등에서 무료로 사용하고 있을만큼 많은 데이터 분석가들이 사용하고 있습니다.



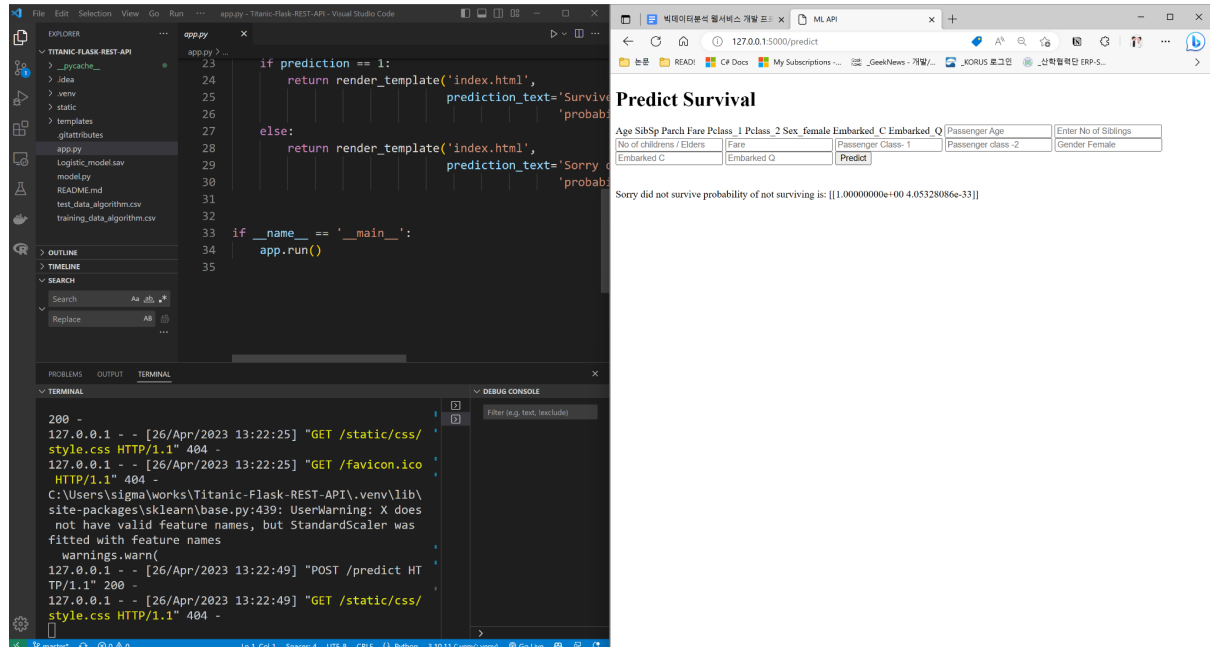
<그림 1. Plotly를 사용해서 Titanic 생존자 예측 데이터를 시각화 한 예제>

3.2 REST API 제공

- 1) 우리는 원활한 소통을 기반으로 프로젝트를 효율적으로 진행하기 위해선 다른 구성원에게 예측 결과를 손쉽게 확인할 수 있는 간단한 도구를 제공하는 능력이 필요하다고 정의하였습니다.
- 2) 보안, 세션 연결 등과 같은 BE에서 처리하는 능력을 요구하기 보다는 빠르게 예측 결과를 다른 구성원들이 확인 및 사용할 수 있도록 경량 마이크로서비스를

지향하며 Python에서 매끄럽게(seamless)하게 작동하는 [flask](#)를 활용하기로 하였습니다.

- 3) flask의 경우 학습 곡선이 가파르지 않고, Python으로 구동되기 때문에 [Scikit-Learn](#) 및 [TensorFlow](#)를 활용하는데 적합합니다.



<그림 2. Flask를 활용한 간단한 REST API 예제>

4. 문서 변경 사항

버전	변경사항
v0.55	- 프로젝트 관련하여 이미지(스크린캡처) 추가 - 기타 오타자 및 문구 수정 - 제목을 수정
v0.1	초안작성