

버즈니 AI Lab 데이터 분석 엔지니어 과제

문제 1.2 모두 필수로 풀어야 합니다.

문제 1.

다음은 홈쇼핑모아의 키워드 마케팅 운영 데이터입니다.(하단 [데이터 설명] 부문 참조)

운영 데이터는 홈쇼핑모아에서 발생한 사용자의 로그를 기반으로 구축된 속성 데이터와 해당 키워드를 통하여 유입된 앱 설치수로 구성되었습니다. 이 데이터를 분석하여, 다음 요구사항을 만족하도록 문제를 풀어주세요.

< 요구사항 >

- 주어진 데이터를 시계열 데이터(time series)로 접근하여 각 Token(키워드)의 Y(설치 수)를 예측하는 모델을 개발해 주세요. 본 문제에서 Y 값은 각 row의 날짜 기준 미래 2일간의 누적 Y 값으로 재설정 해주시기 바랍니다.
예를 들면 Token X의 20190701부터 20190705까지 과거 5일간의 시계열 데이터로 미래 2일간 Y값의 합, 즉 $SUM(Y_{20190706}, Y_{20190707})$ 을 예측하는 것입니다. 과거 며칠간의 데이터를 사용할지는 자유롭게 테스트해볼 수 있습니다. 예측값은 미래 2일간 합으로 정해졌습니다.
 - 주어진 데이터를 학습 셋과 테스트 셋으로 나누고 모델을 개발해 주세요.
 - 개발한 예측 모델의 알고리즘에 대해 설명해 주세요.
 - 본인이 개발한 예측 모델을 적절한 평가 기준으로 평가해 주시고, 평가 결과가 어떤 의미를 갖는지 설명해주세요.
 - 핵심 키워드: Time series analysis model

안내:

- 파이썬 기반의 라이브러리를 자유롭게 사용하여 문제를 해결하시면 됩니다.
- 문제풀이에 대한 설명 문서와 실행 가능한 소스 코드(python 혹은 ipython notebook 코드)를 제출해주세요.
- [과제 데이터 다운로드](#)

데이터 설명:

- 데이터 이름 : 데이터분석_엔지니어과제_데이터.csv
- 데이터 포맷 : CSV
- 데이터 요약 : 총 35,841 row
 - 필드 설명
 - idx: row 인덱스
 - Token: 키워드의 해시 값
 - Date: 각 키워드의 이벤트 발생 날짜
 - a ~ h: 각 키워드의 속성 값
 - Y: 해당 키워드로 유입된 앱 설치수

문제 2.

홈쇼핑모아에는 수십만에서 수천만에 달하는 다양한 키워드들이 존재합니다. 수많은 키워드 중 홈쇼핑모아 앱을 다운로드 하지 않은 사용자가 다운로드 할 만한 키워드를 추출하고자 합니다. 문제1에서 **학습한 모델을 활용하여** 다음 요구사항을 만족하도록 문제를 풀어주세요.

< 요구사항 >

- 과거 N일간의 시계열 데이터를 입력으로 하고, 예측된 미래 2일간의 앱 설치 수를 출력으로 하는 **API 서버**를 개발해 주세요.
 - N은 문제 1에서 정한 값을 그대로 사용해 주세요.
 - 문제 2에서 생성한 모델을 이용하여 API의 출력을 생성해주세요.
 - API의 URI와 URL은 자유롭게 정해주세요.
- 각 키워드별로 2019년 8월 1일 ~ 2019년 8월 2일 이틀동안 얼마나 많은 다운로드가 발생 할 지 **예측**을 수행해 주세요. 예측된 결과는 **데이터베이스에 저장**해 주시기 바랍니다.
 - 어떤 데이터베이스에 어떤 형태로 저장할 지는 자유입니다.
 - 예측은 반드시 앞서 개발한 모델 예측 API 서버를 활용해 예측을 수행해 주세요.
- 키워드별 다운로드 예측 결과를 활용하기 위해, 다음과 같은 기능을 수행하는 **별도의 API 서버**를 개발해주세요.
 - **생성**: 문제 1에서 등장하지 않은 키워드의 다운로드 수를 예측하여 데이터베이스에 추가 저장
 - **조회**: 데이터베이스에 저장 된 키워드 중 가장 다운로드 수가 많을 것으로 예상되는 키워드 N개 조회
 - **수정**: 키워드별 예측된 다운로드 수의 변경
 - **삭제**: 특정 키워드의 정보 삭제

안내:

- 파이썬 기반의 라이브러리를 자유롭게 활용하여 개발해주세요.
- 문제 1에서 등장하지 않은 키워드에 대한 속성 정보는 무작위로 생성하여도 무방합니다.
- 첫 번째 요구사항과 세 번째 요구사항에서 개발되는 API 서버는 별도로 존재해야 하며, 세 번째 요구사항에서 개발한 API는 첫 번째 요구사항에서 개발한 API를 활용하도록 개발해주세요.
- 문제풀이에 대한 설명 문서와 실행 가능한 소스 코드를 함께 제출해주세요.

추가 요구사항:

- Docker와 Docker-compose를 활용하여 환경 관리 시, 추가 가산점이 있습니다.