

[빅데이터] 데이터 사이언스 담당자 코딩테스트

다음 2가지 유형의 문제 중에서 한 가지를 선택하여 답안을 제출해 주세요.

- 답안은 인재영입 홈페이지 혹은 이메일을 통해 제출하여 주시길 바랍니다. (문제 아래에 제출방법이 기재되어 있습니다.)
- 분석에 사용하는 언어는 Python, JAVA, Scala, R로 한정되어 있습니다.
- 분석 결과가 재현 가능하도록 개발 환경 및 설정 방법 등을 포함하여 제출해 주세요.
- 시간이 부족하여 완결하지 못한 경우, 문제 해결 방안을 간략하게 정리해 주세요.

유형 1. 네트워크 분석

○ 데이터셋

<https://snap.stanford.edu/data/sx-stackoverflow.html>

Stack Overflow 사이트에서 사용자간에 질문, 답변, 코멘트가 이루어진 관계를 기록한 데이터이며, 자세한 설명 및 다운로드를 위 주소에서 확인이 가능합니다.

○ 문제

주어진 stackoverflow 관계 데이터를 사용하여 다음과 같이 분석하고 그 과정을 설명해 주세요.

1. 2가지 이상 관점(평판, 영향력, 전문성 등)에서 각 사용자를 특징지을 수 있는 점수를 구하시오.
2. 사용자들을 군집화하고 그 특징을 분석하시오.
3. 자주 나오는 질문-답변-코멘트의 패턴(motif)을 찾고 그 특징을 분석하시오.
4. 앞 3번에서 찾은 패턴들 중에서 어뷰징과 같은 이상행위패턴을 찾아서 그 특징을 분석하시오.

유형 2. 추천 시스템

○ 데이터셋

<https://www.kaggle.com/ashukr/movie-rating-data>

영화에 대한 평점 데이터이며 자세한 설명 및 다운로드를 위 주소에서 확인이 가능합니다.

○ 문제

고객에게 영화를 추천하는데 아래 항목 중 3가지 이상을 고려한 모형을 만드세요.

- Novelty (새로운 경험: 과거에 본 영화와는 다른)
- Diversity (추천의 다양성)
- Recentness (최근의 경험)
- Freshness (영화의 최신성)

○ 참고 논문

다음의 참고 논문은 이해를 돕기 위함이고, 모형은 자유롭게 설계 가능합니다.

- <http://ir.ii.uam.es/~alejandro/2017/recsys.pdf>

- http://iridia.ulb.ac.be/~rdevooght/papers/UMAP__Long_and_short_term_with_RNN.pdf

- <https://static.googleusercontent.com/media/research.google.com/ko/pubs/archive/45530.pdf>

○ **제출기한** : 2019년 4월 9일 화요일 오후 11시 59분

○ **제출방법** : 인재영입 홈페이지 > 마이페이지 > '과제제출' 란에 첨부파일 업로드

(첨부파일의 크기가 50MB 이상인 경우, recruit@kakaobank.com 으로 제출해주시길 바랍니다.)

문의사항이 있으시다면, 인재영입담당자에게 연락해 주세요.

카카오뱅크 인재영입담당자 Chloe.hyo

Mail. chloe.hyo@kakaobank.com

Phone. 02-6288-6244

Mobile. 010-3735-8821

(주말에 급한 문의가 생기셨을 경우, 핸드폰으로 연락주셔도 됩니다.)