

PORTFOLIO

서울시 성북구 공공도서관 별 '인기 아동 도서'와
유사도 높은 '미대출 아동 도서' 추천 시스템 프로젝트

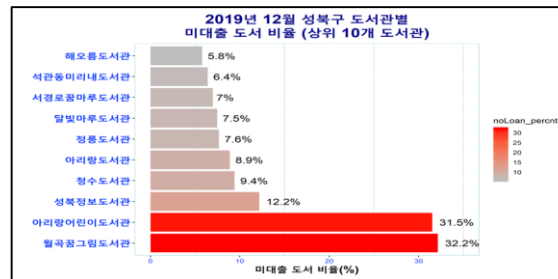
김외솔

PORTFOLIO

프로젝트 취지

“공공도서관 미대출 도서 활성화”

- 도서관 별로 개관 이후 한 번도 대출이 되지 않은 도서가 다수 존재
- 미대출 도서 대책은 ‘도서관 정보나루’ [by 문화체육관광부 & 한국과학기술정보연구원] 빅데이터 사업의 주요 주제



프로젝트 내용

“이용자가 찾는 인기도서가대출 중인 경우, 미대출 도서 중 콘텐츠 유사도가높은 도서 추천!”

- 데이터 수집 (1): 성북구 11개 도서관 각각의 2019년 12월 장서 & 대출 건수 / 최근 3개월 인기 대출 도서 Data
- 데이터 수집 (2): 각 도서의 키워드 & 가중치 Data (여기서 가중치는 ‘제목, 목차, 소개 글에서 단어의 출현 빈도를 의미)
- 데이터 전처리: 2019년 12월 장서 & 대출 건수 데이터에서,
개관 이후 누적 대출 건수가 0 & 아동 도서 & 인기 대출 도서와 주제분류 번호가 일치한 도서 추출
- 사용 알고리즘: 다차원 척도법 (MDS)
 - > 인기 아동 도서에 대한 관련 미대출 아동 도서 각각의 유사도(거리) 계산
 - > 유사도가 높은 순서대로 미대출 도서 추천

PORTFOLIO

프로젝트 상세

- 기간: 2019.11.18 ~ 2020.01.04
- 장소: Multicampus
- 사용언어: R, Python
- 맡은역할: 데이터 크롤링, 전처리, 탐색적 데이터 분석, 도메인 리서치

프로젝트 결과

- Multicampus '서비스 산업 데이터를 활용한 머신러닝 분석' 세미프로젝트 6개 팀 중 1위

