**실무 프로젝트 기획안**

**2020년 8월 21일**

**과정명: 서비스 산업 데이터를 활용한 빅데이터 분석 실무**

|  |  |
| --- | --- |
| **팀 명** | **일조** |
| **팀 원** | ***김도훈(팀장), 김지은, 정혜린 \* 팀장 별도 표시*** |
| **프로젝트 타이틀** | 지역별 초등 돌봄교실 수요 시각화를 통한  초등돌봄교실 우선지역 배치 |
| **프로젝트 주제**  **및 내용**  해결하고자 하는 문제  최종 산출물의 청사진 | **문제인식 및 필요성:**  1) 코로나 19 위기 발발로 초등돌봄 수요가 증가하였으나 공급 이 부족한 상황  2) 돌봄교실을 운영하는데 필요한 전문인력과 지자체의 예산 부족 현상  **분석목적:**  지역별 초등 돌봄교실 수요에 대한 수치를 기반으로 돌봄교실이 부족한 지역을 파악하여 우선 배치 지역 및 학교를 선정한다.  **최종 산출물 :**  서울시의 구별 현황을 트레이닝 모델(training model)로 삼아, 전국 광역별 돌봄교실 수 부족 정도를 파악(testing model)하여 돌봄교실이 필요한 곳을 시각화(visualization)한다. |
| **팀원간 역할**  **분담 및 일정** | **일정:**  **1주차** 8/21(금)기획안 작성/ 데이터 파악,데이터 전처리 1  8/24(월) 데이터 전처리 2  8/25(화) git hub 특강  8/26(수) git hub 특강  8/27(목) 데이터 전처리 3  8/28(금) 데이터 전처리 4  **2주차** 8/31(월) 데이터 전처리 5  9/1(화) 데이터 전처리 6 + 프로그래밍  9/2(수) 데이터 전처리 7 + 프로그래밍  9/3(목) 결과조사 (배경조사)/ 논리구축, 뒷받침 자료찾기  9/4(금) 결과조사 (배경조사)/ 논리구축, 뒷받침 자료찾기  **3주차** 9/7(월) **취업특강**  9/8(화) 마무리 작업  9/9(수) **발표**  **역할분담:**  기획안 작성: 김도훈, 김지은, 정혜린  데이터 파악: 김도훈, 김지은, 정혜린  데이터 전처리: 김도훈, 김지은, 정혜린,  시각화: 김도훈, 정혜린  회귀 알고리즘 진행: 김지은  자료조사: 김도훈, 정혜린  결과보고서 작성: 김도훈, 김지은, 정혜린 |
| **프로젝트 수행 방향**  수행 방법/도구 | 종속변수로 지역별 돌봄교실 수를, 독립변수로 지역별 저소득 수급자 수, 한부모 가정 수, 학생수 등으로 설정하여 R 회귀분석을 한다.  서울시의 구별 자료를 이용하여 트레이닝 모델(training model)을 구축하고, 다른 광역시를 테스팅 모델 (testing model)로 적용하여 돌봄교실의 현황과 부족한 정도를 파악한다.  이를 토대로 돌봄교실 수 증가가 필요한 지역의 우선순위를 정해 전국 지도에 시각화한다. |