**세미프로젝트 기획안**

**과목명: AI 서비스 개발 프로젝트**

**기획안 작성일자 : 2022-08-03**

|  |  |
| --- | --- |
| 팀 명 | 1조 : Deeply |
| 프로젝트 조직  (구성원 및 역할) | **공동 작업 : EDA 분석 및 데이터 전처리, 모델링 작업**  **팀장 : 천승진 (회의 진행 및 기록, 팀 발표)**  **부팀장 : 주혜린 (소스 및 협업 툴 관리)**  **팀원 : 권혁재 (일정수립 및 검토)**  **팀원 : 박준하 (ppt 제작, 자료 수집)**  **팀원 : 장현문 (포트폴리오 작성 및 검수)** |
| 프로젝트 주제 및 개요  해결하고자 하는 문제  최종 산출물의 청사진 | 주제 : NLP with Disaster Tweets and more  스마트 폰의 대중화로 인해 트위터와 같은 sns는 중요한 커뮤니케이션 채널이 됨. 이러한 트위터의 재난 관련 메시지 중 실제 재난 상황을 알리는 메시지를 파악하는 서비스를 개발. 사람들이 작성한 메시지의 내용이 실제로 재난을 알리는 지 여부를 이진 분류 구조로 판단 하고 자 함. 서비스 개발 후 별도로 수집한 데이터를 서비스에 적용하여 응용 가능성을 파악하고 실무에서 활용되고 있는 BERT, HuggingFace 모델을 적용. 이러한 과정을 통해 배운 내용을 바탕으로 자연어 처리에 대한 기본적인 실습 경험을 쌓고자 함. |
| 프로젝트 수행 방향  수행 방법/도구 | 수행 방법   1. 데이터 수집 2. EDA 및 시각화 3. 데이터 전처리 : 4. 정규 표현 식을 활용한 텍스트 정제 5. 중복 요소 제거 6. 공백(결측치) 채우기 7. 자연어만이 발생 가능한 문제 해결 : 맞춤법 검사, 의미 해석을 제대로 하지 못한 요소 수정(예 : 동음이의어), 영어가 아닌 타국어 번역 후 재활용   도구  • 전처리 패키지(nltk, pyspellchecker등)을 활용한 데이터 전처리  • 시각화 패키지(matplotlib, seaborn등)을 활용한 데이터 시각화  • 모델(RNN, LSTM 등)을 활용하여 모델링 작업 수행  • 모델(BERT, HuggingFace)을 활용하여 Transformer 모델링 작업 수행 |
| 프로젝트 추진 일정 | 8/2~8/4 : 프로젝트 기획 안 작성 및 일정 수립  8/5~8/8 : EDA 및 데이터 전처리  8/9~8/12 : 모델링 및 성능 개선 작업  8/13~8/16 : 직접 수집한 데이터를 바탕으로 추가 작업  8/17~8/18 : 포트폴리오 검토 및 ppt작성  8/19 : 팀 프로젝트 발표  (전반적인 일정은 유동적이며 팀원들의 의견을 반영하여 변경 가능함) |

