| [ 인공지능 특화 과정] 프로젝트 계획서 | | | | |
| --- | --- | --- | --- | --- |
| 작성일 | 2020년 2월 23일 | | 팀 명 | 감/언/이/설 : **감**정이 담긴 **언**어를 **이**해하고 **설**명한다 |
| 주제 | 영어 문장을 6개 감정으로 분류하기 (NLP Classification task) | | | |
| 배경  및 목표 | 여러 가지 딥러닝 모델을 이용해 입력된 영어 문장에 대한 적절한 감정 표현을 출력하고, 여러 모델 간의 성능 및 특장점을 비교한다.  {고객 니즈/문제}를 해결할 {핵심 가치}를, {서비스/제품 형태}로 제공한다. {고객 대상}이 {원하는 목표}를 달성하도록, {우리 서비스의 차별화된 가치}를 제공한다.  {고객 문제 상황}에서, {경쟁사와 차별화된 해결 방식}을 통해 {구체적 혜택}을 준다. | | | |
| 일정 | 2월 24일 ~ 2월 25일 | 탐색적 자료조사 | | |
| 2월 26일 ~ 2월 27일 | 데이터 수집 및 전처리 | | |
| 2월 28일 ~ 3월 03일 | 모델 설계 및 비교 | | |
| 3월 04일 ~ 3월 05일 | 프로젝트 발표자료 준비 | | |
| 3월 06일 ~ 3월 06일 | 프로젝트 결과 발표 | | |
| 데이터 수집 및 전처리 | 데이터 수집 - 도큐먼트(영어 문장)-감정 표현 데이터세트 활용  데이터 수집 - <https://www.kaggle.com/praveengovi/emotions-dataset-for-nlp>  전처리 :- 데이터 시각화(word cloud, graph)  - 토큰화  - stop-words 설정  - BoW 생성  - padding 사용  - label 데이터 변환 : label데이터(감정 단어) 숫자 데이터로 변환 | | | |
| 모델 구현 | LSTM - keras의 LSTM 클래스 활용  SimpleRNN - keras의 SimpleRNN 클래스 활용  GRU - keras의 GRU 클래스 활용  Naive Bayes Classification - sklearn의 MultinomialNB 클래스 활용  BERT - keras의 TFBertModel 클래스 활용 | | | |
| 개발 환경 | * 운영 체제 : Windows * 언어 : Python * 도구 : Google Colab * 라이브러리 - 전처리 : Pandas, Numpy, Nltk,   라이브러리 - 분석 : Tensorflow, Sklearn  라이브러리 - 시각화 : Matplotlib, Seaborn | | | |
| 참고 | 원핫인코딩된 도큐먼트(영어 문장)-감정 표현 데이터세트 | | | |