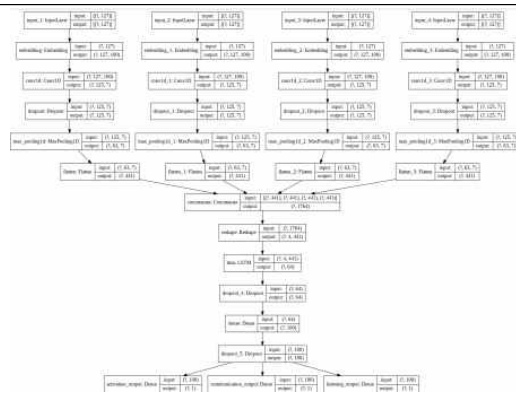


# 프로젝트 기술서

빅데이터 분석 기반 비즈니스 개발자 양성과정 팀프로젝트 수행

▶ 프로젝트 일정	2020.09.14.~10.13
▶ 프로젝트명	딥러닝을 활용한 상담품질평가 알고리즘 구현
▶ 개발환경	Colab in google, Jupyter lab, Khaiii(형태소 분석)
▶ 인원	4명
▶ OS 및 DB	Window10
▶ 사용언어	Python
▶ 프로젝트 소개	<p>딥러닝을 활용한 상담 콜 품질평가 알고리즘 구현을 주제로 콜센터 클라우드 서비스를 개선하는 것에 초점을 둔 프로젝트입니다. 병원 상담 데이터는 기업 브로드씨엔에스에서 제공하였으며, 한글로써 자연어처리 딥러닝을 사용하였습니다. 프로젝트 기획 측면에서는 상담원의 CRM 화면과 관리자의 관리화면 대시보드를 재구성하였고, 개발파트에서는 이를 실현할 수 있는 품질평가 딥러닝 알고리즘을 구축하였습니다.</p>
▶ 본인 역할	<p>1. 관련 기술 조사</p> <ul style="list-style-type: none"><li>- 자연어처리 과정 중 필요한 전처리 관련 기술 및 모델링 관련 기술 조사를 진행하였습니다.</li><li>1) 한국어 자연어처리 전처리 과정에서 언어의 형태소 분석 및 언어의 토큰화 및 패딩 관련 라이브러리 이용법 및 원리를 조사하여 프로젝트에 어떤 방식을 적용할지 선택하였습니다.</li><li>2) 모델링 관련해서는 CNN, LSTM, CNN+LSTM 방식 논문 및 다양한 사례를 조사하며 어떤 모델 구축 방법을 조사하였습니다. 데이터가 Multi-Input &amp; Multi-Output 형태였으며 이에 해당하는 모델링 방법도 조사하였습니다.</li></ul> <p>2. 데이터 전처리</p> <ul style="list-style-type: none"><li>1) 기획팀에서 4문장 단위로 라벨링이 가능하도록 변형하였습니다.</li><li>2) Khaiii 형태소 분석기를 이용해 주요 키워드 단위로 데이터 추출하였습니다.</li><li>3) Tokenizer 라이브러리를 통해 데이터 토큰화 및 일정 길이로 Padding 작업을 하였습니다.</li></ul> <p>3. 모델링</p> <ul style="list-style-type: none"><li>1) CNN, LSTM, CNN+LSTM 등 다양한 모델을 구축해보며 목표 기준을 만족하는 Output이 나오도록 학습하였습니다.</li><li>2) Parameter들을 높은 수준의 성능이 나오도록 여러번 수치 조정을 해보며 학습을 시켰습니다.</li></ul>



적극성 정확도 : 0.978604177279674  
 의사소통 정확도 : 0.9612837493632196  
 경청 정확도 : 0.9816607233825777

#### 4. 예측값의 수치화

- 기존의 점수가 0~1 사이의 예측값을 최소값 : 0, 최대값 : 100 이 되도록 스케일 조정을 하였습니다.

##### 실제 점수

	적극성 점수	의사소통 점수	경청 점수
average	100	84.21	100

##### 모델 예측 점수

	적극성 점수	의사소통 점수	경청 점수
average	98.00	85.30	97.81