데이터 사이언스 프로젝트 계획서

201600779 김영민

201700255 강민정

**팀 이름** : ME (mathematics & Economics)

**팀원 이름** : 김영민(경제학과 201600779), 강민정(수학과 201700255)

**제목**: 병원 개/폐업 예측

**<수정 사항>**

|  |  |
| --- | --- |
| **모델&입력&출력** | |
| 수정 전 | CNN(합성곱 신경망)모델을 사용할 것입니다. |
| 수정 후 | 인공신경망 모델을 사용할 것이다. |
| **구현기술** | |
| 수정 전 | CNN 모델링을 사용할 때에는 |
| 수정 후 | 인공신경망 모델로 모델링을 할 때에는 |

**목적:** 병원의 개/폐업 예측을 통해 다른 산업 분야에도 적용하여 사업의 존망을 예측 할 수 있다. 또한, 폐업의 중요 파라미터를 고려하여 기존에 사업을 하는 사람들에게 유의점을 줄 수 있다.

**프로젝트 요약:** 프로젝트는 KDD 분석 절차(Knowledge Discovery in Database)를 따라서 분석을 할 것이다. KDD분석 방법론은 기술과 데이터베이스를 중심으로 한 Insight를 발견하기 위한 분석 방법론이다. 이 분석 방법은 데이터 선택, 데이터 전처리, 데이터 변환, 데이터 마이닝, 데이터 평가 순으로 5가지 절차로 이뤄져 있다. 이에 추가로 간단한 EDA와 시각화를 통해 데이터의 특징을 파악할 것이다.

그리고 데이터 평가 방법은 accuracy score를 이용할 것이다.

**모델&입력&출력:** 인공신경망 모델을 사용할 것이다. Input은 의료기관의 최근 2개년의 재무 정보와 병원 기본 정보이고 output은 병원의 폐업 여부다.

**장점:** 팀원 각자의 주 전공 지식을 기반으로 하여 회계와 통계학에 대한 지식을 갖추고 있기 때문에 회계 데이터에 대한 이해도가 높다. 이는 주어진 데이터를 통해 보다 병원 개/폐업에 대한 보다 높은 정확도를 예측할 수 있을 것이다

**고유성:** 머신러닝(RandomForestClassifier, XGBoostClassifier)등의 모델과 파이토치의 모델의 성능을 비교하여 정확도를 예측할 것이다.

**구현기술:** EDA 및 Preprocessing은 python으로 구현하고 비교를 위해 머신 러닝을 사용할 때에도 python으로 구현할 것이다.

그리고 인공신경망 모델로 모델링을 할 때에는 Pytorch으로 구현을 할 것이다.

**데이터세트 확보 방법 및 특징:** Dataset은 한국의 캐글 이라고 불리우는 데이콘(<https://dacon.io/>) 이라는 사이트에서 획득하였습니다.

이 데이터의 특징은 float type 64개, object type 4개 , int64 type 4개로 구성되어 있습니다. 또한, shape는 row=301개 column = 58개로 구성되어 있습니다.

이 데이터는 회계 데이터로 주로 구성되어 있습니다.