

PBL 콘텐츠 설계서

PROJECT 1: 의료데이터 분석



2021 March 30

매직에코



**X**

# 목차

[목차 1](#_Toc69301010)

[백그라운드 스토리 3](#_Toc69301011)

[Mission 1: No Show 데이터 분석 3](#_Toc69301012)

[Mission 2: 뇌 CT 분석 3](#_Toc69301013)

[학습해야 하는 키워드에 대한 설명 4](#_Toc69301014)

[Mission 1: 환자 데이터를 분석한 노쇼 예측 4](#_Toc69301015)

[Mission 2: CT 이미지를 분석한 뇌출혈 예측 4](#_Toc69301016)

[프로젝트의 목적 4](#_Toc69301017)

[Mission 1: 텍스트 데이터 분석 및 학습 4](#_Toc69301018)

[Mission 2: 이미지 데이터 분석 및 학습 4](#_Toc69301019)

[수준 구분 5](#_Toc69301020)

[관련 기술 5](#_Toc69301021)

[Mission 1: No Show 데이터 분석 5](#_Toc69301022)

[Knowledge Skill 5](#_Toc69301023)

[Tech Skill 5](#_Toc69301024)

[Soft Skill 5](#_Toc69301025)

[Mission 2: 뇌 CT 분석 6](#_Toc69301026)

[Knowledge Skill 6](#_Toc69301027)

[Tech Skill 6](#_Toc69301028)

[Soft Skill 6](#_Toc69301029)

[미션 별 목표 및 평가기준 7](#_Toc69301030)

[Mission 1-1 7](#_Toc69301031)

[Mission 1-2 8](#_Toc69301032)

[Mission 2-1 9](#_Toc69301033)

[Mission 2-2 10](#_Toc69301034)

[모범 답안 12](#_Toc69301035)

[Mission 1: Medical No Show 12](#_Toc69301036)

[Mission 2: Head CT-Hemorrhage 12](#_Toc69301037)

[데이터 파일 설명 12](#_Toc69301038)

[Mission 1: Medical No Show 12](#_Toc69301039)

[Mission 2: Head CT-Hemorrhage 12](#_Toc69301040)

[학습 지원 정보 13](#_Toc69301041)

[학습 리소스 13](#_Toc69301042)

[포럼 13](#_Toc69301043)

# 백그라운드 스토리

## Mission 1: No Show 데이터 분석

예약 부도(No Show): 진료·수술 일시를 잡고 미리 병·의원에게 알리지 않고 환자가 본인 마음대로 예약을 취소하는 행위, 다른 환자가 진료받을 기회를 빼앗아 간다. 검사나 수술이 시급한 환자들의 빠른 치료 기회가 줄어든다는 점이 가장 큰 문제이다.

오늘도 바쁜 XX병원 데스크. 환자 응대하랴, 걸려오는 예약 전화 받으랴 정신이 없다. 오늘따라 방문을 희망하는 환자들이 많은데 이를 어쩌나, 오전과 오후 시간대 모두 수술 환자가 꽉 잡혀 있어서 다른 환자의 진료가 어렵다. 수술 예약 환자 방문을 10분 앞두고 전화가 또 걸려온다. "너무 아파서 그런데... 지금 바로 가서 진료 좀 볼 수 있을까요?" 하지만 도리없다. "조금 뒤 원장님이 수술을 들어가야 해서, 오늘은 진료가 힘드실 것 같습니다." 응대하는 데스크도 찝찝하다. 너무 아프다고 하는데 진료를 못 봐주니, 방법을 찾아주고 싶지만 불가능하다. 그리고 10분 뒤, 수술 예약한 환자는 오지도 않고 연락도 되지 않는다.

이러한 상황이 반복되니 XX병원 혁신팀은 노쇼를 방지하고자 환자들에게 '예약 선입금'을 시도하려 했지만, 신규 환자 유치를 막는다는 이유로 병원장이 퇴짜를 놓았다. 그리하여 혁신팀은 차선책으로 '예약 대기 환자 시스템'을 제안했다. 예약 대기 환자 시스템이란 부득이 예약이 취소되는 경우를 대비하여 대기 예약 환자를 배정하는 정책이다. 물론 먼저 예약했던 환자가 제시간에 도착한다면 예약 대기 환자는 진료를 받을 수 없지만, 이전의 예약이 부도가 난다면 (노쇼가 발생한다면) 우선 순위는 대기환자에게 돌아가 진료를 받을 수 있게 한다.

하지만 이 시스템을 적용하기에도 아직 어려움이 남아있었다. 모든 시간에 예약 대기환자를 받는다면 예약 대기환자에게 우선순위가 돌아가는 확률은 매우 낮을 것으로 예상되어 예약 대기 환자 시스템을 운영하는 실효성이 떨어진다. 따라서 혁신팀은 환자들의 데이터를 활용하여 특정 환자가 예약을 부도낼 확률, 즉 노쇼 확률(No - Show Probability)를 계산하여 그 확률이 높은 환자들의 예약에만 대기 환자를 받자고 제안했다.

## Mission 2: 뇌 CT 분석

XX병원은 노쇼 문제를 해결하자 더욱 많은 손님들로 붐비게 되었다. 특히 XX병원의 뇌신경센터는 전국적으로 유명하여 항상 환자로 센터가 가득 찼지만, 노쇼 문제가 해결되자 터져나오는 손님으로 센터가 마비가 될 정도였다. 넘쳐나는 환자들은 뇌신경센터에 인력난을 불러왔고, 병원 운영팀은 인력 보충을 약속했지만 곧바로 전문적인 인력이 보충되기는 말처럼 쉬운 일이 아니었다.

이러한 상황에서 의료사고가 발생했다. 하루에 30명 이상의 환자들을 상대하던 방사선 전문의 A는 뇌졸증 환자를 뇌경색으로 판단하였고 결국 환자는 뇌사상태에 빠졌다. 뇌졸증의 치료는 시작부터 끝까지 모든 순간이 중요하지만 환자의 뇌 컴퓨터단층촬영(CT) 이미지로부터 환자에게 뇌출혈이 있는지 판단하는 과정이 가장 중요하다. 이렇게 출혈을 감지하는 작업은 특별히 훈련된 방사선 전문의들에게도 높은 집중력을 필요로 하는 까다로운 작업이다.

이러한 문제가 발생하자 병원 혁신팀은 뇌출혈을 정확하여 의료사고를 예방하고 의료진들의 피로를 덜어주고자 뇌출혈 판단 모델을 만들기로 결정하였다. 이제까지의 뇌출혈 환자들의 CT사진을 식별하여 뇌출혈 판단 모델을 만들고 해당 모델을 활용하여 새로운 환자의 CT사진만 갖고도 그 환자가 뇌출혈을 갖고 있는지 빠르게 판단하고, 의사들로 하여금 결정을 도울 수 있도록 하였다.

# 학습해야 하는 키워드에 대한 설명

## Mission 1: 환자 데이터를 분석한 노쇼 예측

환자에게 적절하고 정확한 의료 서비스를 제공하고 병원 자원을 효율적으로 사용하기 위 하여 대부분의 병원은 예약 진료 시스템을 사용하고 있다. 그러나 사회가 고령화 되어감에 따라 의료 서비스 요구가 증가하고, 이는 곧 예약의 증가, 예약 대기의 증가, 그리고 예약 부도 (no-show)의 증가로 이어지고 있다. 이러한 상황에서 환자들은 불필요하게 길어진 예약 대기 때문에 건강 상태의 불확실성이 증가하게 되며, 불필요한 응급 진료를 받는 경우가 발생하기도 한다. 또한, 병원에서는 의료진들의 진료 효율성이 저하되고 병원 자원을 비효율적으로 운영하게 되어 병원의 수익성이 악화될 수 있다.

no-show를 설명하고 예측하는 연구는 실제 데이터를 이용하여 no-show에 영향을 줄 수 있는 요인들을 나열하고 통계적 분석 등의 방 법을 통해 핵심 요소들을 찾는다. 그리고 이를 이용하여 병원 예약 관리에 시사점을 제시하거나, 대안을 제시한 뒤 적용 사례연구를 통해 대 안의 효과에 대해 분석하고, 환자 별 no-show 확률에 대해서 예측하기도 한다.

## Mission 2: CT 이미지를 분석한 뇌출혈 예측

CT 검사 시 뇌출혈의 부위는 심각한 정도에 따라 인지하기 어려운 경우도 있으며, 응급상황에서 숙련이 되지 않은 의료진에게는 부담을 준다. 응급상황에서 검사와 동시에 뇌출혈부위를 자동으로 빨리 파악하고 정량적인 정보를 제공하는 보조적인 역할은 필요하며, 컴퓨터를 이용한 자동 검출 및 인식 시스템은 출혈부위 진단에 매우 큰 도움을 줄 수 있다. 본 연구에서는 출혈부위를 문턱치 값 처리, 모폴로지 연산, 원형률 계산을 접목하여 뇌 출혈부위의 자동검출에 성공하였고 검출 후보군에서 잘 못 선정된 영역을 판정하기위한 주성분분석을 이용한 분류기 개발에 성공하였다. 개발된 시스템을 응급상황의 뇌출혈 환자에게 적용한다면 의료진에게 수술계획을 위한 유용한 정보가 될 것으로 사료된다.

프로젝트의 목적

Mission 1: 텍스트 데이터 분석 및 학습

|  |  |
| --- | --- |
| 구분 | 설명 |
| **1** | 데이터 저장 및 준비 등 데이터 전처리 과정을 경험할 수 있다. |
| **2** | 데이터를 자세하게 분석하고 해당 내용에 대해 설명할 수 있다. |
| **3** | 모델링을 경험해볼 수 있다. |

Mission 2: 이미지 데이터 분석 및 학습

|  |  |
| --- | --- |
| 구분 | 설명 |
| **1** | 데이터 저장 및 준비 등 데이터 전처리 과정을 경험할 수 있다. |
| **2** | 데이터 전처리 과정 중 이미지 변형에 대해 학습할 수 있다. |
| **3** | 모델링 과정을 통해 CNN을 경험할 수 있다. |

# 수준 구분

* AI/Data분석/Edu4
* AI/모델개발/Edu4
* AI/AIOps/Edu4

# 관련 기술

## Mission 1: No Show 데이터 분석

### Knowledge Skill

* Python
* Jupyter Notebook
* TensorFlow
* NumPy
* Pandas
* scikit-learn
* CNN

### Tech Skill

|  |  |  |
| --- | --- | --- |
| 분류 | 경험 | 난이도 |
| Data분석(데이터 추출 및 가공 역량) | 수집된 데이터를 특정 포맷으로 가공하고 정렬할 수 있다, 수집된 데이터에서 잘못된 데이터를 찾아 수정 삭제할 수 있다, 수집된 데이터에서 불필요한 차원을 제거할 수 있다. | 2 |
| Data분석(데이터 추론 역량) | 모집단으로부터 추출된 표본을 바탕으로 통계적 추론이 가능하다, 대용량의 자료에서 데이터의 관계, 패턴, 규칙을 탐색하고 활용할 수 있다, 모델링을 위한 성능 평가 지표를 정의하고 구현할 수 있다. | 3 |
| Data분석(데이터 재가공 및 전산화 역량) | 데이터의 특성을 파악해, 수집된 데이터를 재 가공할 수 있다. | 2 |
| AI모델개발(ml/dl/기초 지식 이해력) | 머신러닝과 관련된 공개된 학습모델(사이킷런 라이브러리의 문서 등)을 따라서 모델을 돌려볼 수 있다, 과적합에 대한 정의를 이해하고 설명할 수 있다. | 4 |

### Soft Skill

|  |  |  |
| --- | --- | --- |
| 역량명 | 키워드 | 행동 |
| 비판적 사고 Critical Thinking | 핵심 정보 파악 | 비판적 사고/핵심 정보 파악/다양한 정보나 아이디어를 무작정 수용하지 않고, 업무에 도움이 되는 것과 그렇지 않은 것을 구별할 수 있다. |
| 비판적 사고 Critical Thinking | 근본 원인 규명 | 비판적 사고/근본 원인 규명/문제 발생 시, 문제 해결을 위해 근본 원인을 해결할 수 있다. |

## Mission 2: 뇌 CT 분석

### Knowledge Skill

* Python
* Jupyter Notebook
* TensorFlow
* NumPy
* Pandas
* scikit-learn
* CNN

### Tech Skill

|  |  |  |
| --- | --- | --- |
| 분류 | 경험 | 난이도 |
| Data분석(데이터 추출 및 가공 역량) | 수집된 데이터를 특정 포맷으로 가공하고 정렬할 수 있다, 수집된 데이터에서 잘못된 데이터를 찾아 수정 삭제할 수 있다, 수집된 데이터에서 불필요한 차원을 제거할 수 있다. | 2 |
| Data분석(데이터 추론 역량) | 모집단으로부터 추출된 표본을 바탕으로 통계적 추론이 가능하다, 대용량의 자료에서 데이터의 관계, 패턴, 규칙을 탐색하고 활용할 수 있다, 모델링을 위한 성능 평가 지표를 정의하고 구현할 수 있다. | 3 |
| Data분석(데이터 재가공 및 전산화 역량) | 데이터의 특성을 파악해, 수집된 데이터를 재 가공할 수 있다. | 2 |
| AI모델개발(ml/dl/기초 지식 이해력) | 머신러닝과 관련된 공개된 학습모델(사이킷런 라이브러리의 문서 등)을 따라서 모델을 돌려볼 수 있다, 과적합에 대한 정의를 이해하고 설명할 수 있다. | 4 |

### Soft Skill

|  |  |  |
| --- | --- | --- |
| 역량명 | 키워드 | 행동 |
| 비판적 사고 Critical Thinking | 핵심 정보 파악 | 비판적 사고/핵심 정보 파악/다양한 정보나 아이디어를 무작정 수용하지 않고, 업무에 도움이 되는 것과 그렇지 않은 것을 구별할 수 있다. |
| 비판적 사고 Critical Thinking | 근본 원인 규명 | 비판적 사고/근본 원인 규명/문제 발생 시, 문제 해결을 위해 근본 원인을 해결할 수 있다. |

# 미션 별 목표 및 평가기준

## Mission 1-1

환자 데이터를 읽어오고 데이터를 전처리하고 분석하라, 분석한 내용을 토대로 보고서를 작성하고 데이터 분석에 관한 발표 동영상을 만들어라

**키워드**

Numpy, Pandas

matplotlib, seaborn, data analysis

**목표**

1. Data Store: 구글드라이브에 환자 진료 데이터를 저장하고 구글 드라이브를 구글 Collaboratory에서 사용할 있도록 마운트하라
2. Data Explore & Preprocess: 주어진 환자 진료 데이터를 분석하기 위해 전처리를 진행하라
3. Data Analysis: 전처리를 완료한 환자 데이터를 분석하라

|  |  |
| --- | --- |
| 세부 목표 | 구분 |
| 1. 구글드라이브에 환자 진료 데이터를 저장하고 구글 드라이브를 구글 Collaboratory에서 사용할 있도록 마운트하라 | Data Store |
| 2. 데이터에는 오타가 존재한다. 오타를 고쳐라. (optional) | Data Explore & Preprocess |
| 3. 학습에 필요 없는 데이터가 있으면 삭제하라. | Data Explore & Preprocess |
| 4. 학습을 위한 데이터 타입으로 변환하라. | Data Explore & Preprocess |
| 5. 데이터 에러가 있는 지 검사하고 있다면 수정하라. | Data Explore & Preprocess |
| 6. 학습에 필요하다고 판단되는 새로운 데이터를 생성하라. (optional) | Data Explore & Preprocess |
| 7. 데이터들을 그래프로 그리고 분석하여 본인의 의견을 적는다. | Data Analysis |
| 8. 분석한 보고서를 팀원들 앞에서 발표한다. | Data Analysis |

**결과물**

* 데이터 분석 보고서 (형태: 자유)
* 데이터 분석 발표 동영상

**평가 기준**

|  |  |
| --- | --- |
| 멘토 평가 항목 | 구분 |
| 1. 구글드라이브에 저장했는가? | Data Store |
| 2. try except의 사용으로 데이터 엑세스 에러에 대처했는가? | Data Store |
| 1. 오타를 고쳤는가? | Data Explore & Preprocess |
| 2. 학습에 필요없는 데이터를 삭제했는가? | Data Explore & Preprocess |
| 3. 데이터 타입을 변환했는가? | Data Explore & Preprocess |
| 4. 데이터 에러를 확인하려는 노력을 했는가? (검출하지 못했어도 다양하게 노력한 것으로만 점수 부여) | Data Explore & Preprocess |
| 5. 새로운 데이터를 생성했는가? (높은 점수 부여) | Data Explore & Preprocess |
| 6. 많은 데이터를 분석할수록 높은 점수 부여 | Data Analysis |
| 7. 보고서를 잘 만들수록 높은 점수 부여 | Data Analysis |
| 8. 발표를 잘할수록 높은 점수 부여 | Data Analysis |

**확인 질문**

|  |  |  |
| --- | --- | --- |
| 분류 | 문제 | 정답 |
| Data Explore & Preprocess | 판다스 데이터 프레임을 간단하게 요약해서 보여주며, 주로 데이터 타입을 확인할 때 사용하는 메소드는? | info() (sample(), describe()을 적었을 경우 점수 반만 부여) |
| Data Explore & Preprocess | 열에 어떤 데이터들이 있는지 확인해 보고 싶을 때 사용하는 메소드는? | set(), unique() (둘중 하나만 적어도 정답) |
| Data Analysis | 판다스 데이터의 개수를 세기 위해서 사용하는 메소드. 예를 들어 no show의 yes와 no의 개수를 세기 위해서 어떤 메소드를 사용해야 하는가? | size() |
| Data Analysis | 병원은 쉬는 요일이 있다. 언제인가? | 토요일, 일요일 |

## Mission 1-2

분석한 환자 데이터를 기반으로 CNN모델을 개발하는 단계

**키워드**

Decision Tree Classifier, Random Forest Classfier, Grid Search CV, Bayes Classifier, XGBoost, Logistic Regression

**목표**

Modeling: 현재까지 가공 및 분석한 환자 데이터를 토대로 CNN모델을 개발하라

|  |  |
| --- | --- |
| 세부 목표 | 구분 |
| 1. 데이터를 2개 이상의 알고리즘으로 모델링하고 score를 확인하라. (optional) | Data Modeling |
| 2. 팀원들 앞에서 사용한 알고리즘을 간단히 설명하고 모델링 결과를 설명하라. | Data Modeling |

**결과물**

* 코드
* 모델링 결과 발표 동영상

**평가 기준**

|  |  |
| --- | --- |
| 멘토 평가 항목 | 구분 |
| 1. 많은 알고리즘으로 모델링할 수록 높은 점수 부여 | Data Modeling |
| 2. score가 높은 순서대로 높은 점수 부여 | Data Modeling |
| 3. 발표를 잘할수록 높은 점수 부여 | Data Modeling |
| 4. 알고리즘에 대해서 잘 설명한 순서대로 높은 점수 부여 | Data Modeling |

**확인 질문**

|  |  |  |
| --- | --- | --- |
| 분류 | 문제 | 정답 |
| Data Modeling | 모델의 학습을 위해 data type이 object인 것을 숫자 값으로 변환해 주어야 한다. 어떤 메소드를 어떻게 사용해야 하는지 서술하라. | sklearn의 labelEncoder()와 fit\_transform()을 사용하여 범주형 데이터를 정수형 유효값으로 치환한다. |
| Data Modeling | 모델의 정확도를 측정하기 위해 사용하는 메소드는? | score() |

## Mission 2-1

CT이미지를 읽어오고 데이터를 분석한다.

(01\_Data Storage 🡪02\_Data Analysis 🡪 03\_Data Split 🡪 04\_Image Data Generator순으로 미션 진행)

**키워드**

numpy hist, image to numpy array, opencv

ImageDataGenerator

train\_test\_split

**목표**

1. Data Store: 구글드라이브에 환자 진료 데이터를 저장하고 구글 드라이브를 구글 Collaboratory에서 사용할 있도록 마운트하라
2. Data Analysis: 데이터들을 다양한 그래프로 출력하여 분석한다. 학습을 위해 이미지를 128, 128 크기로 저장한다.
3. Data Split: 학습을 위해 데이터를 train 80%, validation 10%, test 10%로 나누어라.
4. Image Data Generator: 데이터 변환으로 추가데이터를 확보한다는 것은 train과 validation 데이터를 가공하여 추가 데이터를 확보한다는 의미이다. test데이터는 가공할 이유가 없으므로 추가데이터 확보는 반드시 데이터를 train과 test로 구분하고 난 뒤에 train에 대해서만 가공한다.

|  |  |
| --- | --- |
| 세부 목표 | 구분 |
| 1. 구글드라이브에 환자 진료 데이터를 저장하고 구글 드라이브를 구글 Collaboratory에서 사용할 있도록 마운트하라 | Data Store |
| 2. 이미지는 128, 128 크기로 저장했는가? | Data Analysis |
| 3. 학습을 위해 데이터를 train 80%, validation 10%, test 10%로 나누어라. | Data Split |
| 4. ImageDataGenerator를 사용하여 이미지를 변형하라. (optional) | Image Data Generator |
| 5. 이미지를 출력하여 확인하라. | Image Data Generator |
| 6. ImageDateGenerator의 인수 rescale, shear\_range, rotation\_range, width\_shift\_range, height\_shift\_range, horizontal\_flip에 대해서 자세히 공부하고, 발표하라. | Image Data Generator |

**결과물**

* ImageDateGenerator의 발표 동영상

**평가 기준**

|  |  |
| --- | --- |
| 멘토 평가 항목 | 구분 |
| 1. try except의 사용으로 데이터 엑세스 에러에 대처했는가? | Data Store |
| 2. 구글드라이브에 저장했는가? | Data Store |
| 3. 그래프를 많이 만들어서 데이터 분석을 많이 할수록 높은 점수 부여 | Data Analysis |
| 4. 이미지는 128, 128 크기로 저장했는가? | Data Analysis |
| 5. 데이터를 train, validation, test로 나누었는가? (train\_test\_split을 사용하지 않았어도 감점 요소가 아님) | Data Split |
| 6. ImageDateGenerator를 사용하였는가? | Image Data Generator |
| 7. 발표를 잘할 수록 높은 점수 부여 | Image Data Generator |

**확인 질문**

|  |  |  |
| --- | --- | --- |
| 분류 | 문제 | 정답 |
| Data Analysis | 이미지를 pyplot에서 출력하기 위해서 사용하는 메소드는? | imshow() |
| Data Split | 퀴즈1: train\_test\_split을 사용할 때, 실행 시 마다 같은 학습 데이터를 얻고 싶다. 어떤 인자 값을 어떻게 해야 하는가? | random\_state에 임의의 값을 넣어준다. 시드를 고정하기 때문에 실생 시마다 같은 학습 데이터를 얻는다. |

## Mission 2-2

분석한 뇌 CT 이미지를 기반으로 CNN 모델을 개발한 후 평가한다.

**키워드**

CNN 기초

**목표**

1. CNN Modeling: CNN 기초를 공부한다. 구현 중심으로 코드를 공부할 것
2. Estimation: train, validation, test의 accuracy를 구하고 history를 이용하여 모델의 accuracy와 loss에 대한 그래프를 그리고 평가하라.

|  |  |
| --- | --- |
| 세부 목표 | 구분 |
| 1. CNN으로 모델을 만들어라. 여러 개의 Conv Layer를 사용해 보고 최적의 모델링을 하라. (optional) | CNN Modeling |
| 2. 모델을 훈련시켜라. | CNN Modeling |
| 3. relu, maxpooling, sigmoid, binary\_crossentropy에 대해서 조사하고 발표하라. | CNN Modeling |
| 4. 여러 알고리즘을 사용하여 각각 모델링 해보자 (optional) | CNN Modeling |
| 5. train, validation, test의 accuracy를 구하고 history를 이용하여 모델의 accuracy와 loss에 대한 그래프를 그리고 평가하라. | Estimation |
| 6, accuracy와 loss 외의 그래프를 추가하여 결과를 설명하라. (optional) | Estimation |
| 7. 그래프를 해석하고 결과를 설명하는 발표를 하라. | Estimation |

**결과물**

* CNN 발표 동영상 & 결과 발표 동영상
* accuracy와 loss에 대한 그래프와 evaluate 결과에 대한 스크린 샷
* 코드

**평가 기준**

|  |  |
| --- | --- |
| 멘토 평가 항목 | 구분 |
| 1. 여러 개의 Conv Layer를 시도해 봤는가? (학생 코드 관찰 필요) | Data Analysis |
| 2. 발표를 잘할 수록 높은 점수 부여 | Data Analysis |
| 3. 스크린샷에 accuracy 그래프, model loss 그래프, eavluste 결과가 포함되었는가? (각각 점수 부여) | Estimation |
| 4. test 데이터로 모델을 평가하여 높은 정확도 순으로 높은 점수 부여 | Estimation |
| 5. 결과 발표를 잘한 순으로 높은 점수 부여 | Estimation |
| 6. 2개 이상의 모델을 만들어서 비교하여 좋은 모델을 선택한 경우 높은 추가 점수 부여. 많은 모델을 비교할 수록 높은 점수 부여 (코드 관찰이 필요 - 멘토 )  추가 평가 | Estimation |

**확인 질문**

|  |  |  |
| --- | --- | --- |
| 분류 | 문제 | 정답 |
| Estimation | accuracy 그래프에서 train과 test의 그래프가 차이가 커지는 것은 무엇을 의미하는가? | 오버피팅이 증가함을 의미한다. |

# 모범 답안

## Mission 1: Medical No Show

<https://github.com/edu10make/Project-X_/blob/main/Project1-Medical%20Analysis/001_Medical%20No%20Show/001_Medical%20No%20Show.ipynb>

## Mission 2: Head CT-Hemorrhage

<https://github.com/edu10make/Project-X_/blob/main/Project1-Medical%20Analysis/002_Head%20CT-Hemorrhage/002_Head%20CT-Hemorrhage.ipynb>

# 데이터 파일 설명

## Mission 1: Medical No Show

**Medical no show.csv:** 브라질 환자들의 병원 예약 정보와 예약 취소 여부에 대한 데이터

* PatientId: 환자 ID
* AppointmentID: 예약 ID
* Gender: 성별
* DataMarcacaoConsulta: 예약 날짜
* DataAgendamento: 진료 날짜
* Age: 나이
* Neighbourhood: 지역
* Scholarship: 장학금
* Hipertension: 고혈압
* Diabetes: 당뇨
* Alcoholism: 알콜중독
* Handcap: 장애
* SMS\_received: 환자에게 보낸 문자
* No-show: 예약취소

## Mission 2: Head CT-Hemorrhage

**이미지 파일**

* 정상 뇌 사진 100장
* 뇌출혈 뇌 사진 100장

**labels.csv:** 뇌출혈 여부 데이터

# 학습 지원 정보

## 학습 리소스

* <https://en.wikipedia.org/wiki/Convolution_neural_network>  
  : Convolution Neural Network(합성곱 신경망)에 대한 일반적인 이해 (위키피디아)
* <https://en.wikipedia.org/wiki/Scikit-learn>  
  : scikit-learn에 대한 일반적인 이해 (위키피디아)
* <https://en.wikipedia.org/wiki/CT_scan>  
  : CT scan에 대한 일반적인 이해 (위키피디아)
* [Tensorflow KR 페이스북 그룹](<https://www.facebook.com/groups/TensorFlowKR>) : tensorflow를 학습하거나, tensorflow, 다른 Machine Learning 도구에 관한 연구, 응용 제작을 하는 분들의 모임. 질의 응답이 매우 활발함
* •[Machine Learning Yearning](<https://www.deeplearning.ai/machine-learning-yearning/>) : 앤드유 응교수의 ML 프로젝트를 진행할 때 고려해야 할 우선순위, 성능에 대한 고려 사항 (딥러닝 기술 자체를 다룬 책은 아니라 필독). PDF 버전을 인터넷에서 쉽게 구할 수 있음
* •[모두를 위한 딥러닝]([https://www.inflearn.com/course/기본적인-머신러닝-딥러닝-강좌#](https://www.inflearn.com/course/기본적인-머신러닝-딥러닝-강좌%23)) : 머신 러닝과 딥러닝에 대해 더 이해하고 본인들의 문제를 Tensorflow를 이용하여 풀 수 있게 도와주는 동영상 강의. 이 강좌는 수학이나 컴퓨터 공학적인 지식이 없어도 어렵지 않게 볼 수 있음.
* <https://www.youtube.com/playlist?list=PLS8gIc2q83OjStGjdTF2LZtc0vefCAbnX&ab_channel=NeoWizard>  
  : 머신 러닝, 딥러닝
* [https://www.youtube.com/watch?v=AHdGc4nR5f4&list=PLS8gIc2q83OhM0RTktKDitgZGX5dHo7Vs&ab\_channel=NeoWizard](https://www.youtube.com/watch?v=AHdGc4nR5f4&list=PLS8gIc2q83OhM0RTktKDitgZGX5dHo7Vs&ab_channel=NeoWizard%20)   
  : TensorFlow

# 포럼

* 학습 리소스: <https://github.com/edu10make/Project-X_>
* 토론과 Q&A: https://prj.innoaca.kr/forum/project/2/forum
* 코드리뷰: https://prj.innoaca.kr/forum/project/2/codereview
* 동료평가: https://prj.innoaca.kr/forum/project/2/evaluation
* 멘토: https://prj.innoaca.kr/forum/project/2/mentor
* 동료목록: https://prj.innoaca.kr/forum/project/2/peer