

# ABCD Project

학교	○○고등학교
학번	○○○○○
이름	○○○
프로젝트 이름	

## Step D. 관심 분야 설정 및 문제 정의

### Step D-1. 관심 분야 설정

특히 흥미를 가지고 있는 주제 또는 진로와 관련하여 평소 관심을 두었던 분야가 있다면 찾아서 적어보자.

[예시] 데이터 사이언스를 접목할 수 있는 다양한 분야들

1. 의료 및 생명과학
2. 환경공학 및 지속 가능성
3. 기후 변화와 경제
4. 금융 및 경제
5. 교육
6. 사회과학 및 정책
7. 법률 및 윤리
8. 도시 계획 및 스마트 시티
9. 스포츠 분석 및 건강관리
10. 예술

## Step D-2. 관심 분야에서 해결하고 싶은 문제 정의하기

10년 후 설정한 관심 분야에서 전문가가 되었다고 생각해보자. 해당 분야의 전문가로서 해결하고 싶은 어떠한 문제가 있는가?

꼭 관심 분야가 아니더라도 평소에 해결하고 싶었던 어떠한 문제가 있는가? 또는 미래 사회에 우리가 부딪히게 될 문제 중 해결하고 싶은 무언가가 있는가? 어떠한 종류의 문제든 본인이 해결하고 싶은 현재 또는 미래 사회의 문제를 찾아보거나 직접 정의해보자.

### [예시] 분야별 데이터 사이언스를 활용하여 해결할 수 있는 문제들

#### 1. 의료 및 생명과학

- AI로 심장질환을 조기에 예측하고 발견하여 환자들에게 위험성을 경고할 수 있는가?
- AI를 활용해 의료 영상을 분석하여 암을 조기에 진단할 수 있는가?
- 병원 내 대기 시간을 예측하고 관리하여 효율성을 높일 수 있는가?

#### 2. 환경공학 및 지속 가능성

- AI로 재활용 쓰레기를 자동으로 분류하고 재활용률을 높일 수 있는가?
- AI 기반 작물 모니터링을 통해 스마트 농업에서 자원 사용을 최적화할 수 있는가?
- 기후 데이터를 분석해 기후 변화의 장기적 영향을 예측하고 환경 보호 정책 수립에 기여할 수 있는가?

#### 3. 기후 변화와 경제

- 다양한 산업에서 발생하는 탄소 배출량을 예측하고 줄일 수 있는 방법을 제시할 수 있는가?
- 기후 변화가 각 산업에 미치는 경제적 영향을 분석하여 정책 제안을 위한 데이터를 제공할 수 있는가?
- 재생 에너지의 효율성을 분석하여 가장 적합한 에너지 솔루션을 제시할 수 있는가?

#### 4. 금융 및 경제

- AI를 통해 주식 시장의 과거 데이터를 분석해 미래 주식 가격 변동을 예측할 수 있는가?
- 개인의 투자 성향과 재정 상태를 분석하여 맞춤형 투자 전략을 제안할 수 있는가?
- 금융 거래 데이터를 분석하여 사기 거래를 탐지할 수 있는가?

#### 5. 교육

- 학생들의 학습 패턴을 분석해 학업 성과를 예측하고 학습 방법을 개선할 수 있는가?
- 학생의 학습 성향을 분석하여 개인 맞춤형 학습 자료와 콘텐츠를 추천할 수 있는가?
- 학생들의 출석 및 참여도를 분석하여 학습 집중도를 높이는 방안을 제시할 수 있는가?

#### 6. 사회과학 및 정책

- AI로 범죄 데이터를 분석하여 특정 지역의 범죄 발생 가능성을 예측하고 예방할 수 있는

가?

- 소셜 미디어 데이터를 통해 여론을 분석하고 사회적 문제에 대한 대중의 반응을 예측할 수 있는가?
- 사회적 정책이 경제 및 사회에 미치는 영향을 예측하여 정책 효과를 평가할 수 있는가?

#### 7. 법률 및 윤리

- AI로 법률 문서를 분석하여 위험 요소나 주요 조항을 자동으로 탐지할 수 있는가?
- 과거 법률 판례를 분석하여 특정 사건의 법적 판결을 예측할 수 있는가?
- AI 기술 발전에 따른 윤리적 문제를 해결하기 위한 AI 윤리 기준을 설정할 수 있는가?

#### 8. 도시 계획 및 스마트 시티

- 실시간 교통 데이터를 분석해 도로 혼잡을 줄이고 최적의 경로를 제시할 수 있는가?
- 스마트 시티 내 건물의 에너지 소비 패턴을 분석하여 에너지 절약 방안을 제시할 수 있는가?
- CCTV 영상을 분석하여 도시 내 공공 안전을 실시간으로 관리하고 위험 상황을 예측할 수 있는가?

#### 9. 스포츠 분석 및 건강관리

- 선수의 신체 데이터를 분석하여 부상 위험을 예측하고 맞춤형 운동 프로그램을 제시할 수 있는가?
- 경기 데이터를 분석해 팀의 경기 전략을 최적화하고 승률을 높일 수 있는가?
- 개인의 건강 데이터를 기반으로 맞춤형 운동 및 건강 관리 프로그램을 제시할 수 있는가?

#### 10. 예술

- AI를 활용해 다양한 장르의 음악을 자동으로 생성하고 이를 창작 활동에 활용할 수 있는가?
- 딥러닝을 통해 예술 작품이나 그림을 자동으로 생성하여 전시나 창작 활동에 활용할 수 있는가?
- 영화 대본 데이터를 분석하여 새로운 스토리나 시나리오를 자동으로 생성할 수 있는가?

### Step D-3. 자료/통계를 근거로 문제의 중요도 설명하기

앞서 정의한 문제가 얼마나 심각한 문제인가? 본인이 찾은 문제가 왜 해결되어야 한다고 생각하는지 정보를 검색하여 자료/통계를 근거로 설명해보자.

### Step D-4. 현재 수준에서 검증해볼 수 있는 문제로 축소하기

앞서 정의한 문제가 이 프로젝트를 통해서 해결할 수 있는 수준인가? 문제의 수준이 너무 높거나 혹은 정확히 어떤 문제를 해결하겠다는 것인지 잘 이해가 안 되도록 모호하게 정의하지는 않았는가? 만약 그렇다면 현재 나의 수준에서 비교적 쉽게 검증해볼 수 있는 문제로 축소해보자. 즉, 범위를 좁히거나, 규모를 축소하거나, 난도를 낮추는 등의 방식으로 문제를 매우 구체적으로 서술해보자.

## Step A. 인공지능 설계

### Step A-1. 문제를 해결하기 위한 인공지능 해결책(아이디어) 제시

본인이 제기한 문제가 왜 아직까지 해결이 안 되었는가? AI를 이용한다면 더 효과적으로 해결할 수 있는가? 앞에서 정의한 문제를 해결하기 위해 어떤 인공지능이 필요하다고 생각하는지 인공지능을 활용한 해결책을 구체적으로 표현해보자.

단, 상상력을 발휘하는 것은 좋지만 실현가능성도 어느 정도 고려해서, 해당 인공지능이 어떤 기능이 있는지, 그 기능이 어떤 방식으로 작동하여 문제를 해결할 수 있는 것인지 등을 구체적으로 설명해보자.

#### [예시] 분야별 문제에 대한 인공지능 해결책

\* 아래 제시된 것들이 실제로 구현가능한지에 대한 여부는 정확히 알 수 없음에 주의

#### 1. 의료 및 생명과학

##### • AI로 심장질환 조기 예측 시스템 구축

- 기능 : 환자의 생체 데이터를 분석하여 심장질환 발병 위험성을 조기에 예측.
- 작동 방식 : 환자의 심박수, 혈압, 콜레스테롤 수치 등 생체 데이터를 입력받아, 기계학습 모델이 정상 범위와 위험 패턴을 비교 분석. 일정 위험 수준에 도달하면 환자와 의료진에게 경고 알람을 보냄.

##### • 의료 영상 분석을 통한 암 조기 진단

- 기능 : X-ray 또는 MRI 영상을 분석하여 암 조직을 조기에 감지.
- 작동 방식 : 영상 내 암 조직의 패턴을 학습한 모델이 새로운 영상 데이터를 분석하여 의심스러운 부분을 하이라이트. 검사 중 이상이 발견되면 의료진에게 즉각 경고.

##### • 병원 내 대기 시간 단축을 위한 AI 솔루션 개발

- 기능 : 환자의 진료 대기 시간을 실시간으로 예측 및 관리.
- 작동 방식 : 병원 시스템에서 환자의 진료 내역과 병원 내 혼잡도를 실시간으로 분석. 예상 대기 시간을 계산하여, 환자에게 알람을 보내고 진료 일정을 조정.

#### 2. 환경공학 및 지속 가능성

##### • AI를 활용한 쓰레기 분리수거 최적화 시스템

- 기능 : 쓰레기를 자동 분류하여 재활용 가능성을 높임.
- 작동 방식 : 쓰레기 이미지 데이터를 분석하여 재활용품과 일반 쓰레기를 구분하는 AI 모델이 분리수거 장비에 장착됨. 분류된 결과에 따라 재활용품과 일반 쓰레기를 각각 다른 통으로 분리.

##### • 스마트 농업을 위한 AI 기반 작물 모니터링 시스템

- 기능 : 작물의 성장 상태를 예측하고 물과 비료 사용을 최적화.
- 작동 방식 : 날씨 데이터와 작물의 생육 상태 데이터를 분석하여 수분과 영양분의 적정량을 계산. AI가 토양 상태에 따라 적정량의 물과 비료를 자동으로 조절하

여 투입.

- **기후 변화 예측 모델 구축**

- 기능 : 기후 데이터를 분석해 기후 변화의 장기적 영향을 예측.
- 작동 방식 : 온도, 습도, 강수량 등의 기후 데이터를 학습한 모델이 미래의 기후 변화를 시뮬레이션하고, 예상되는 기후 변화 패턴을 제공하여 정책 결정에 활용.

### 3. 기후 변화와 경제

- **AI를 활용한 탄소 배출량 예측 및 감소 솔루션 개발**

- 기능 : 다양한 산업의 탄소 배출량을 예측하고 최적의 감축 방안을 제시.
- 작동 방식 : 산업별 생산량과 에너지 소비 데이터를 학습하여 탄소 배출량을 계산. AI가 감축 가능한 영역을 탐색하고, 감축 시나리오에 따른 탄소 감소 효과를 예측.

- **기후 변화가 경제에 미치는 영향 분석**

- 기능 : 기후 변화가 각 산업에 미치는 경제적 영향을 예측.
- 작동 방식 : 기후 변화 지표와 경제 지표를 연관 분석하여 산업별 영향도를 평가. 기후 변화에 따른 경제 변화를 예측해 정책 결정에 참고 자료 제공.

- **재생 에너지 효율 분석 시스템**

- 기능 : 태양광, 풍력 등 재생 에너지의 효율성을 분석하여 최적화.
- 작동 방식 : 에너지 생산량 데이터를 분석하여 에너지 효율을 계산. AI가 기후 조건에 따른 최적의 에너지 생산 전략을 제시하고, 비효율적인 부분을 개선하는 방안을 제안.

### 4. 금융 및 경제

- **AI를 통한 주식 시장 예측**

- 기능 : 과거 데이터를 바탕으로 주가 변동을 예측.
- 작동 방식 : 과거 주가, 거래량, 경제 지표 등의 데이터를 분석하여 패턴을 학습한 AI가 미래 주가 추이를 예측. 예측된 결과에 따라 투자자에게 알림을 제공.

- **개인 맞춤형 투자 전략 제안 시스템**

- 기능 : 개인의 투자 성향과 재정 상태를 기반으로 맞춤형 투자 전략을 제안.
- 작동 방식 : 개인의 투자 성향, 리스크 선호도 등을 분석하여 최적의 투자 조합을 구성하고, 개별 투자 전략을 제공해 사용자에게 투자 방향을 제안.

- **AI 기반 금융 사기 탐지 시스템**

- 기능 : 금융 거래 데이터를 분석하여 사기 거래를 실시간으로 탐지.
- 작동 방식 : 정상 거래와 사기 거래의 패턴을 학습한 모델이 금융 데이터를 실시간으로 분석하여 의심스러운 거래를 탐지. 사기 가능성이 높은 경우 경고를 발송.

### 5. 교육

- **AI 기반 학습 성과 예측 시스템**

- 기능 : 학생의 학습 패턴을 분석하여 학업 성과를 예측하고 학습 전략을 개선.
- 작동 방식 : 학생의 출석률, 과제 제출, 시험 성적 등을 분석하여 학업 성과를 예측하

고, 개선이 필요한 영역을 제안.

• **맞춤형 학습 추천 AI 시스템**

- 기능 : 학생의 학습 성향을 분석해 개인 맞춤형 학습 자료와 콘텐츠를 추천.
- 작동 방식 : 학습 성향 데이터를 분석한 후 적합한 학습 자료를 선택하여 추천. 학생이 부족한 부분에 맞춰 관련 학습 자료를 제공하여 학습 효과를 극대화.

• **학생 출석 및 참여도 분석 시스템**

- 기능 : 학생들의 출석 및 수업 참여도를 분석하여 학습 집중도를 높일 방안을 제시.
- 작동 방식 : 학생의 출석 기록과 참여 데이터를 분석해 학습 집중도 패턴을 파악하고, 낮은 참여도를 보일 때 경고 알림 발송.

**6. 사회과학 및 정책**

• **AI를 이용한 범죄 예측 및 예방 시스템**

- 기능 : 특정 지역에서 발생할 가능성이 높은 범죄를 예측하고 예방책을 제시.
- 작동 방식 : 과거 범죄 데이터를 분석하여 위험도가 높은 지역을 탐지. 예측된 범죄 유형에 따라 예방책과 경고를 해당 지역 사회에 제공.

• **AI 기반 여론 분석**

- 기능 : 소셜 미디어 데이터를 분석해 주요 여론의 방향성을 예측.
- 작동 방식 : 소셜 미디어 데이터를 분석해 주요 키워드와 주제에 대한 여론 방향을 파악. 여론 변화를 실시간으로 모니터링하여 정책 수립 시 참고 자료로 제공.

• **정책 효과 분석을 위한 AI 모델 구축**

- 기능 : 특정 정책이 경제와 사회에 미치는 영향을 분석하고 예측.
- 작동 방식 : 정책 도입 전후 데이터를 비교 분석하여 정책 효과를 예측. 사회적 변화를 시뮬레이션하여 정책의 장기적 영향을 파악.

**7. 법률 및 윤리**

• **AI를 활용한 법률 문서 자동 분석 시스템**

- 기능 : 법률 문서 내 주요 조항과 위험 요소를 자동으로 탐지.
- 작동 방식 : 법률 문서의 내용을 분석해 주요 조항을 추출하고, 위험도가 높은 부분을 표시. 문서 검토 시간을 줄여줌으로써 법률 업무를 효율화.

• **법적 판결 예측 AI 모델**

- 기능 : 특정 사건에 대한 법적 판결을 예측.
- 작동 방식 : 과거 판례를 학습한 모델이 새로운 사건의 데이터를 입력받아 예상 판결을 제시. 법적 위험을 미리 파악하고 전략적 대응이 가능.

**8. 도시 계획 및 스마트 시티**

• **AI 기반 교통 최적화 시스템**

- 기능 : 실시간 교통 데이터를 분석해 도로 혼잡을 줄이고 최적의 경로를 제공.
- 작동 방식 : 차량의 위치, 속도, 교통 혼잡도 등을 실시간으로 수집하여 AI가 최적 경로를 추천. 교통량 예측 모델이 특정 시간대의 혼잡을 미리 예측하고 우회 경로를 제안하여 도시 내 교통 체증을 감소.

- **스마트 시티를 위한 에너지 효율 관리 시스템**

- 기능 : 건물 내 에너지 소비 패턴을 분석하여 에너지를 절약할 수 있는 최적의 방안을 제공.
- 작동 방식 : 건물의 시간대별 전기 사용량, 기온, 조명 사용 패턴 등을 분석하여 에너지 소비가 높은 구간을 찾아냄. AI가 불필요한 전력 소비를 줄이기 위한 최적의 에너지 사용 패턴을 제시하여 전력 효율성을 향상.

- **공공 안전 모니터링 시스템**

- 기능 : 도시 내 CCTV 영상을 분석하여 실시간으로 공공 안전을 모니터링하고, 위험 상황을 감지하여 경고.
- 작동 방식 : AI가 CCTV 영상을 실시간으로 분석해 사람이나 차량의 이상 행동, 혼잡도 증가 등을 감지. AI가 이상 행동을 발견하면 해당 지역의 안전 관리팀에 경고 알림을 보내 즉각적인 조치가 가능하게 함. 예를 들어, 군중이 갑자기 모여들거나 위험한 행동이 포착되면 경고를 발송하여 실시간으로 상황을 관리.

## 9. 스포츠 분석 및 건강관리

- **AI를 활용한 선수 부상 예측 시스템**

- 기능 : 선수의 신체 데이터를 분석해 부상 위험을 예측하고 예방책을 제시.
- 작동 방식 : 운동 중 발생하는 신체 데이터(심박수, 운동 강도 등)를 분석해 부상 위험이 높은 경우 경고. 예방 운동을 추천하여 부상 방지.

- **경기 전략 최적화를 위한 AI 분석**

- 기능 : 경기 데이터를 분석하여 팀의 경기 전략을 최적화하고 승률을 높임.
- 작동 방식 : 경기 중 발생하는 실시간 데이터를 분석해 최적의 전략을 실시간으로 제안. 각 선수의 위치나 동작을 분석하여 전략을 조정.

- **AI 기반 피트니스 개인화 프로그램 개발**

- 기능 : 개인의 건강 데이터를 기반으로 맞춤형 운동 및 건강 관리 프로그램을 제공.
- 작동 방식 : 개인의 운동 기록, 신체 상태 등을 분석해 건강 목표에 맞춘 맞춤형 운동 계획을 제시. 운동 후 성과를 기록하고 다음 운동을 조정하여 최적의 결과를 유도.

## 10. 예술

- **AI 기반 음악 생성 시스템**

- 기능 : 다양한 장르의 음악을 생성하여 작곡가들에게 새로운 아이디어를 제공.
- 작동 방식 : 생성적 적대 신경망(GAN)을 사용해 주어진 음악 스타일에 맞는 멜로디나 리듬을 생성. 사용자는 생성된 음악을 창작에 활용.

- **AI를 이용한 그림 및 이미지 생성**

- 기능 : 딥러닝을 활용해 새로운 예술 작품이나 그림을 자동 생성.
- 작동 방식 : 스타일 GAN을 통해 특정 예술 스타일을 학습하고, 새로운 이미지나 그림을 생성하여 사용자에게 제공.

- **AI를 활용한 영화 시나리오 생성**

- 기능 : 영화 대본 데이터를 분석하여 새로운 시나리오를 자동 생성.

- 작동 방식 : 자연어 생성 모델을 통해 다양한 시나리오 구조와 문장 패턴을 학습한 뒤, 새로운 스토리라인을 생성하여 영화 제작에 활용.

## Step A-2. 인공지능 구현을 위한 원리 조사

아이디어로 제시한 인공지능을 기계학습으로 구현하기 위해 어떤 기계학습을 사용해야 하는지 조사해보자.

- ① 지도학습      ② 비지도학습      ③ 강화학습      ④ 2가지 방법 이상      ⑤ 기타

### [예시] 사례별 원리

\* 아래 제시된 것이 고정된 정답은 아님에 주의

#### 1. 의료 및 생명과학

##### • AI로 심장질환 조기 예측 시스템 구축

→ ① 지도학습 : 심장질환 예측 문제는 과거 환자의 데이터(심박수, 혈압 등)와 해당 환자가 질환을 겪었는지 여부(정답 데이터)를 함께 사용해 예측 모델을 학습. 정상과 비정상 상태에 대한 레이블이 있어 지도학습에 적합.

##### • 의료 영상 분석을 통한 암 조기 진단

→ ① 지도학습 : 암 진단 문제에서는 X-ray 또는 MRI 영상 데이터(문제 데이터)와 암 여부에 대한 레이블(정답 데이터)을 함께 사용해 모델을 학습. 레이블된 이미지 데이터가 있어 지도학습에 적합.

##### • 병원 내 대기 시간 단축을 위한 AI 솔루션 개발

→ ① 지도학습 : 대기 시간 예측 문제는 환자의 진료 기록과 과거 대기 시간 데이터(문제 데이터)와 실제 대기 시간(정답 데이터)을 이용해 예측 모델을 학습. 예측하고자 하는 결과에 대한 정답이 있어 지도학습에 적합.

#### 2. 환경공학 및 지속 가능성

##### • AI를 활용한 쓰레기 분리수거 최적화 시스템

→ ① 지도학습 : 쓰레기 분류 문제는 다양한 종류의 쓰레기 이미지 데이터(문제 데이터)와 각 이미지에 대한 분류 레이블(정답 데이터)을 사용해 학습. 라벨이 있는 이미지 분류 문제이므로 지도학습에 적합.

##### • 스마트 농업을 위한 AI 기반 작물 모니터링 시스템

→ ② 비지도학습 : 작물 상태 모니터링은 주로 정상 상태의 데이터만을 수집해 학습한 뒤, 새로운 데이터에서 비정상 패턴을 감지. 정답 없이 정상 상태 패턴만을 학습하고 이상치를 감지하므로 비지도학습에 적합.

##### • 기후 변화 예측 모델 구축

→ ① 지도학습 : 기후 예측 문제는 과거의 기온, 강수량, 습도 등 기후 데이터(문제 데이터)와 미래 기후 상태(정답 데이터)를 기반으로 학습. 예측하고자 하는 결과에 대한 정답이 있으므로 지도학습에 적합.

#### 3. 기후 변화와 경제

##### • AI를 활용한 탄소 배출량 예측 및 감소 솔루션 개발

→ ① 지도학습 : 탄소 배출 예측 문제는 다양한 산업에서 발생한 과거의 배출량 데이터(문제 데이터)와 실제 배출량(정답 데이터)을 학습하여 예측 모델을 구축. 과거 데

이터에 배출량 레이블이 있으므로 지도학습에 적합.

- **기후 변화가 경제에 미치는 영향 분석**

- ① 지도학습 : 경제 데이터와 기후 데이터를 함께 사용해 특정 경제 지표에 대한 기후 변화의 영향을 예측. 경제 지표의 변화(정답 데이터)에 대한 기후 요인(문제 데이터)을 학습하는 지도학습 문제.

- **재생 에너지 효율 분석 시스템**

- ② 비지도학습 : 재생 에너지 시스템의 효율성을 모니터링하고 비정상적인 에너지 소비 패턴을 감지하기 위해, 비정상 상태의 레이블 없이 정상적인 에너지 소비 패턴을 학습. 정답 없이 이상치를 감지하는 방식이므로 비지도학습에 적합.

#### 4. 금융 및 경제

- **AI를 통한 주식 시장 예측**

- ① 지도학습 : 과거의 주식 시장 데이터(문제 데이터)와 실제 주가 변동(정답 데이터)을 학습하여 미래의 주가 변동을 예측하는 문제. 미래 주가를 예측하기 위한 정답 데이터가 있어 지도학습에 적합.

- **개인 맞춤형 투자 전략 제안 시스템**

- ① 지도학습 : 개인의 투자 이력과 성향 데이터(문제 데이터)와 맞춤형 투자 전략에 대한 레이블(정답 데이터)을 학습하여 개별 투자 전략을 추천. 예측하고자 하는 정답이 있어 지도학습에 적합.

- **AI 기반 금융 사기 탐지 시스템**

- ④ 2가지 방법 이상 : 지도학습으로 사기 거래와 정상 거래를 학습하여 사기 패턴을 감지하고, 비지도학습으로 새로운 이상 거래를 탐지.

#### 5. 교육

- **AI 기반 학습 성과 예측 시스템**

- ① 지도학습 : 학생의 학습 데이터를 바탕으로 성과를 예측하는 문제로, 학습 기록(문제 데이터)과 학업 성과(정답 데이터)를 함께 학습. 정답 데이터가 있어 지도학습에 적합.

- **맞춤형 학습 추천 AI 시스템**

- ① 지도학습 : 학생의 학습 성향 데이터를 기반으로 적합한 학습 자료를 추천하기 위해, 과거 추천 결과(정답 데이터)를 학습. 정답이 있는 문제이므로 지도학습에 적합.

- **학생 출석 및 참여도 분석 시스템**

- ② 비지도학습 : 출석 및 참여도에서 비정상적인 패턴을 감지하기 위해, 정상 출석 데이터를 학습하고 이상 패턴을 탐지. 정답 데이터가 없는 문제이므로 비지도학습에 적합.

#### 6. 사회과학 및 정책

- **AI를 이용한 범죄 예측 및 예방 시스템**

- ① 지도학습 : 과거의 범죄 데이터(문제 데이터)와 발생 여부(정답 데이터)를 학습하여 특정 지역의 범죄 위험을 예측. 레이블된 데이터가 있으므로 지도학습에 적합.

- AI 기반 여론 분석
  - ② 비지도학습 : 소셜 미디어 데이터를 기반으로 여론을 분석하여 주요 패턴을 추출. 레이블 없이 주요 주제를 발견하고 분석하므로 비지도학습에 적합.
- 정책 효과 분석을 위한 AI 모델 구축
  - ① 지도학습 : 과거 정책의 도입과 그에 따른 사회적 변화를 레이블 데이터로 학습하여 새로운 정책의 영향을 예측. 정답 데이터가 있으므로 지도학습에 적합.

## 7. 법률 및 윤리

- AI를 활용한 법률 문서 자동 분석 시스템
  - ② 비지도학습 : 법률 문서의 주요 항목을 분석하기 위해, 레이블 없이 문서의 주요 패턴을 탐색하고 자동으로 특징을 추출. 정답 데이터가 없으므로 비지도학습에 적합.
- 법적 판결 예측 AI 모델
  - ① 지도학습 : 과거 법률 판례 데이터(문제 데이터)와 판결 결과(정답 데이터)를 학습하여 판결을 예측. 예측할 판결 레이블이 있어 지도학습에 적합.
- AI 윤리 기준 개발 프로젝트
  - ⑤ 기타 : 기계학습보다는 AI 기술의 윤리적 기준을 수립하는 연구로, 학습 알고리즘이 필요하지 않음.

## 8. 도시 계획 및 스마트 시티

- AI 기반 교통 최적화 시스템
  - ① 지도학습 : 과거 교통 흐름 데이터(문제 데이터)와 실제 교통 혼잡도(정답 데이터)를 학습하여 혼잡을 줄이는 최적 경로를 예측. 정답 데이터가 있으므로 지도학습에 적합.
- 스마트 시티를 위한 에너지 효율 관리 시스템
  - ② 비지도학습 : 건물 내 에너지 소비 패턴에서 비정상적인 에너지 소비를 감지하기 위해 정상 패턴만을 학습. 정답 데이터가 없으므로 비지도학습에 적합.
- 공공 안전 모니터링 시스템
  - ④ 2가지 방법 이상 : 지도학습으로 과거의 사고 데이터를 학습하고, 비지도학습으로 새로운 이상 패턴을 탐지하여 실시간으로 위험 상황을 감지.

## 9. 스포츠 분석 및 건강관리

- AI를 활용한 선수 부상 예측 시스템
  - ① 지도학습 : 선수의 신체 데이터와 부상 발생 여부를 학습하여 부상 위험을 예측. 예측할 정답 데이터가 있으므로 지도학습에 적합.
- 경기 전략 최적화를 위한 AI 분석
  - ② 비지도학습 : 경기 데이터를 바탕으로 유사한 전략 패턴을 탐색하여 최적의 전략을 도출. 정답 데이터 없이 패턴을 분석하므로 비지도학습에 적합.
- AI 기반 피트니스 개인화 프로그램 개발
  - ① 지도학습 : 개인의 건강 데이터를 학습하여 최적의 운동 프로그램을 추천. 추천할 정답 데이터가 있으므로 지도학습에 적합.

## 10. 예술

- AI 기반 음악 생성 시스템

- ⑤ 기타 : 생성 모델(GAN)을 사용해 새로운 음악을 생성. 기존 학습 유형과 달리 음악 생성에 특화된 모델이 필요.

- AI를 이용한 그림 및 이미지 생성

- ⑤ 기타 : 딥러닝의 생성 모델(GAN)을 통해 새로운 이미지를 생성. 비정형 데이터를 생성하는 특수 모델이 필요.

- AI를 활용한 영화 시나리오 생성

- ⑤ 기타 : 자연어 생성 모델을 사용해 영화 시나리오를 생성. 딥러닝 기반의 자연어 생성 모델이 필요.

### Step A-3. 인공지능망 선택 및 설계

기계학습을 구현하기 위해 인공지능망 알고리즘을 사용하려고 한다. 인공지능망은 어떻게 설계하느냐에 따라서 다층퍼셉트론(MLP), 합성곱신경망(CNN), 순환신경망(RNN) 등, 다양한 인공지능망들이 있으며, 어떤 인공지능을 구현하고자 하는지에 따라 적절한 인공지능망을 선택하여 사용한다.

앞에서 조사한 기계학습을 구현하기 위해 어떤 인공지능망이 필요한지 조사해보고, 적절한 인공지능망을 선택하여 자신의 아이디어를 딥러닝으로 구현하기 위한 인공지능망을 설계해보자.

#### [예시] 사례별 인공지능망 설계

\* 아래 제시된 것이 고정된 정답은 아님에 주의

##### 1. 의료 및 생명과학

###### • AI로 심장질환 조기 예측 시스템 구축

- 다층 퍼셉트론(MLP) : 환자의 심박수, 혈압 등 다양한 수치형 데이터를 입력으로 받는 문제이므로 다층 퍼셉트론(Multi-Layer Perceptron)을 사용.
- 구조 예시 : 입력층(심박수, 혈압 등) → 은닉층 2개(각 64, 32개 뉴런, ReLU 활성화 함수) → 출력층(2개의 뉴런, Softmax 활성화 함수)

###### • 의료 영상 분석을 통한 암 조기 진단

- 합성곱 신경망(CNN) : X-ray 또는 MRI와 같은 이미지 데이터를 처리하기 위해 CNN을 사용.
- 구조 예시 : 입력층(이미지) → Conv2D(32, 3x3 필터, ReLU) → MaxPooling2D → Conv2D(64, 3x3 필터, ReLU) → MaxPooling2D → 플래튼 → 완전 연결층(128 뉴런, ReLU) → 출력층(Softmax)

###### • 병원 내 대기 시간 단축을 위한 AI 솔루션 개발

- 다층 퍼셉트론(MLP) : 수치형 데이터 기반 예측 문제이므로 MLP를 사용.
- 구조 예시 : 입력층(환자의 진료 기록 등) → 은닉층 2개(각 64, 32개 뉴런, ReLU) → 출력층(1개 뉴런, 대기 시간 예측)

##### 2. 환경공학 및 지속 가능성

###### • AI를 활용한 쓰레기 분리수거 최적화 시스템

- 합성곱 신경망(CNN) : 이미지 기반 분류 문제이므로 CNN을 사용하여 쓰레기 이미지를 분류.
- 구조 예시 : 입력층(이미지) → Conv2D(32, 3x3 필터, ReLU) → MaxPooling2D → Conv2D(64, 3x3 필터, ReLU) → MaxPooling2D → 플래튼 → 완전 연결층(128 뉴런, ReLU) → 출력층(Softmax)

###### • 스마트 농업을 위한 AI 기반 작물 모니터링 시스템

- 오토인코더 : 비지도학습을 통해 정상 작물 상태를 학습하고, 비정상 상태를 감지하기 위해 오토인코더 사용.
- 구조 예시 : 입력층(작물 상태 데이터) → 인코더(Dense 64, 32, ReLU) → 디코더

(Dense 32, 64, ReLU) → 출력층(입력과 동일 차원)

• 기후 변화 예측 모델 구축

- 다층 퍼셉트론(MLP) : 수치형 데이터 기반 예측 문제이므로 MLP를 사용.
- 구조 예시 : 입력층(기온, 강수량, 습도 등) → 은닉층 3개(128, 64, 32 뉴런, ReLU) → 출력층(미래 기후 예측)

### 3. 기후 변화와 경제

• AI를 활용한 탄소 배출량 예측 및 감소 솔루션 개발

- 다층 퍼셉트론(MLP) : 산업별 수치형 데이터를 기반으로 탄소 배출을 예측하므로 MLP를 사용.
- 구조 예시 : 입력층(산업 데이터) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(배출량 예측)

• 기후 변화가 경제에 미치는 영향 분석

- 다층 퍼셉트론(MLP) : 수치형 데이터 기반 예측 문제로 MLP 사용.
- 구조 예시 : 입력층(기후, 경제 지표) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(경제 지표 예측)

• 재생 에너지 효율 분석 시스템

- 오토인코더 : 비지도학습을 통해 정상 에너지 소비 패턴을 학습하고, 비정상 소비 패턴을 감지하기 위해 오토인코더 사용.
- 구조 예시 : 입력층(에너지 소비 데이터) → 인코더(Dense 64, 32, ReLU) → 디코더(Dense 32, 64, ReLU) → 출력층(입력과 동일 차원)

### 4. 금융 및 경제

• AI를 통한 주식 시장 예측

- 순환 신경망(RNN) : 주식의 과거 시계열 데이터를 활용하므로 RNN을 사용해 시계열 데이터 학습.
- 구조 예시 : 입력층(과거 주가 데이터) → LSTM(64 유닛) → LSTM(32 유닛) → 출력층(주가 예측)

• 개인 맞춤형 투자 전략 제안 시스템

- 다층 퍼셉트론(MLP) : 개인의 투자 성향을 분석하는 수치형 데이터 기반 예측 문제로 MLP 사용.
- 구조 예시 : 입력층(투자 성향 데이터) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(맞춤형 투자 전략)

• AI 기반 금융 사기 탐지 시스템

- 다층 퍼셉트론(MLP) + 오토인코더 : 지도학습으로 사기 패턴을 학습하고, 비지도학습 오토인코더로 새로운 이상 거래 감지.
- 구조 예시 : 지도학습 - MLP 사용 (사기/정상 예측), 비지도학습 - 오토인코더 사용

### 5. 교육

• AI 기반 학습 성과 예측 시스템

- 다층 퍼셉트론(MLP) : 학생의 학습 데이터를 기반으로 성과를 예측하므로 MLP 사용.

- 구조 예시 : 입력층(학습 패턴 데이터) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(학습 성과 예측)

• **맞춤형 학습 추천 AI 시스템**

- 다층 퍼셉트론(MLP) : 학생의 학습 성향을 분석하고 추천하므로 MLP 사용.
- 구조 예시 : 입력층(학습 성향 데이터) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(추천 학습 자료)

• **학생 출석 및 참여도 분석 시스템**

- 오토인코더 : 정상 출석 및 참여도를 학습하고, 비정상 패턴을 감지하기 위해 오토인코더 사용.
- 구조 예시 : 입력층(출석 데이터) → 인코더(Dense 64, 32, ReLU) → 디코더(Dense 32, 64, ReLU) → 출력층(입력과 동일 차원)

## 6. 사회과학 및 정책

• **AI를 이용한 범죄 예측 및 예방 시스템**

- 다층 퍼셉트론(MLP) : 수치형 데이터를 기반으로 특정 지역의 범죄를 예측하므로 MLP 사용.
- 구조 예시 : 입력층(범죄 데이터) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(범죄 발생 예측)

• **AI 기반 여론 분석**

- 순환 신경망(RNN) : 소셜 미디어 텍스트의 시계열 패턴을 분석하므로 RNN 사용.
- 구조 예시 : 입력층(텍스트 데이터) → LSTM(64 유닛) → LSTM(32 유닛) → 출력층(여론 주제 분석)

• **정책 효과 분석을 위한 AI 모델 구축**

- 다층 퍼셉트론(MLP) : 정책과 사회적 데이터 기반 예측 문제로 MLP 사용.
- 구조 예시 : 입력층(정책 및 사회 데이터) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(정책 효과 예측)

## 7. 법률 및 윤리

• **AI를 활용한 법률 문서 자동 분석 시스템**

- 다층 퍼셉트론(MLP) : 문서의 주요 조항 분석이 주로 수치 데이터로 변환 가능하므로 MLP 사용.
- 구조 예시 : 입력층(문서 특징) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(조항 분석)

• **법적 판결 예측 AI 모델**

- 다층 퍼셉트론(MLP) : 판례의 수치형 특성을 입력으로 받는 문제로 MLP 사용.
- 구조 예시 : 입력층(판례 데이터) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(판결 예측)

## 8. 도시 계획 및 스마트 시티

• **AI 기반 교통 최적화 시스템**

- 순환 신경망(RNN) : 교통 데이터의 시계열 패턴을 분석하여 RNN 사용.

- 구조 예시 : 입력층(교통 데이터) → LSTM(64 유닛) → LSTM(32 유닛) → 출력층(최적 경로 예측)
- **스마트 시티를 위한 에너지 효율 관리 시스템**
  - 오토인코더 : 정상 에너지 소비 패턴을 학습하고 이상 패턴 감지에 오토인코더 사용.
  - 구조 예시 : 입력층(에너지 데이터) → 인코더(Dense 64, 32, ReLU) → 디코더(Dense 32, 64, ReLU) → 출력층(입력과 동일 차원)
- **공공 안전 모니터링 시스템**
  - 다층 퍼셉트론(MLP) + 오토인코더 : 지도학습을 통해 과거 사고 데이터를 학습하여 특정 유형의 위험을 분류하고, 비지도학습을 통해 새로운 이상 패턴을 감지할 수 있도록 오토인코더를 사용. 지도학습과 비지도학습을 결합하여 실시간 위험 상황을 감지하는 시스템을 구축.
  - 구조 예시 :
    - 지도학습 모델 (MLP) : 과거 사고 데이터에 대한 레이블이 있는 경우 사용하여 특정 위험 유형(예 : 사고, 화재, 범죄 등)을 분류.
    - 입력층(안전 관련 수치 데이터 : 위치, 시간대, 사건 유형 등) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(위험 유형별 Softmax 분류)
    - 비지도학습 모델 (오토인코더) : 기존 정상적인 패턴의 안전 데이터를 학습하여 이상치 탐지를 수행. 새로운 데이터가 들어왔을 때 재구성 오차를 통해 이상 상황을 감지.
    - 입력층(안전 관련 수치 데이터) → 인코더(Dense 64, 32, ReLU) → 디코더(Dense 32, 64, ReLU) → 출력층(입력과 동일 차원)

## 9. 스포츠 분석 및 건강관리

- **AI를 활용한 선수 부상 예측 시스템**
  - 다층 퍼셉트론(MLP) : 선수의 다양한 신체 데이터(예 : 운동 강도, 심박수, 운동 시간 등)와 부상 발생 여부를 기반으로 부상 위험을 예측하므로 MLP 사용.
  - 구조 예시 : 입력층(신체 데이터) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(부상 발생 예측)
- **경기 전략 최적화를 위한 AI 분석**
  - 순환 신경망(RNN) : 경기 중 발생하는 일련의 시계열 데이터를 분석하여 팀 전략을 최적화하므로 RNN 사용.
  - 구조 예시 : 입력층(시계열 경기 데이터) → LSTM(64 유닛) → LSTM(32 유닛) → 출력층(최적 전략 제안)
- **AI 기반 피트니스 개인화 프로그램 개발**
  - 다층 퍼셉트론(MLP) : 개인의 운동 기록과 신체 데이터를 바탕으로 최적의 운동 프로그램을 추천하기 위해 MLP 사용.
  - 구조 예시 : 입력층(개인 건강 및 운동 데이터) → 은닉층 2개(64, 32 뉴런, ReLU) → 출력층(개인화된 운동 추천)

## 10. 예술

- AI 기반 음악 생성 시스템

- 생성적 적대 신경망(GAN) : GAN을 사용하여 다양한 장르의 음악을 생성하는 방식. GAN은 두 개의 네트워크(생성자와 판별자)가 서로 경쟁하며 학습하는 방식으로, 창의적인 음악을 생성하는 데 적합.

- 구조 예시 :

- 1) 생성자 : 입력층(노이즈 벡터) → Dense(256) → ReLU → Dense(512) → ReLU → 출력층(음악 벡터)

- 2) 판별자 : 입력층(음악 벡터) → Dense(512) → LeakyReLU → Dense(256) → LeakyReLU → 출력층(실제/가짜 분류)

- AI를 이용한 그림 및 이미지 생성

- 생성적 적대 신경망(GAN) : 다양한 스타일의 그림과 이미지를 생성하기 위해 GAN 사용. 특히, 스타일을 특징하게 생성할 경우 스타일 GAN을 사용할 수 있음.

- 구조 예시 :

- 1) 생성자 : 입력층(노이즈 벡터) → Dense(256) → ReLU → Conv2DTranspose(128, 4x4) → ReLU → 출력층(이미지)

- 2) 판별자 : 입력층(이미지) → Conv2D(64, 4x4) → LeakyReLU → Dense(1) → Sigmoid (실제/가짜 분류)

- AI를 활용한 영화 시나리오 생성

- 순환 신경망(RNN) : 시나리오는 긴 문장 구조와 문맥을 고려하여 생성되어야 하므로, 텍스트 데이터를 처리하는 LSTM 기반 RNN 사용.

- 구조 예시 : 입력층(텍스트 시퀀스) → LSTM(128 유닛) → LSTM(64 유닛) → Dense(어휘 크기, Softmax) (다음 단어 예측)

## Step B. 데이터 수집 및 처리

### Step B-1. 데이터 특징 및 속성 정의

아이디어로 제시한 인공지능을 딥러닝으로 구현하기 위해 필요한 데이터를 수집해야 한다. 데이터 수집에 앞서, 다음을 참고하여 데이터가 가져야 할 특징과 세부 속성을 정의해보자. 그리고 왜 그러한 속성이 필요하다고 생각하는지 설명해보자.

기계학습의 종류에 따라 수집해야 하는 데이터의 특징 및 속성

#### 1. 지도학습

- 특징 : 지도학습은 모델이 주어진 입력 데이터와 정답(레이블)을 통해 학습하는 방식이다. 따라서 레이블이 있는 데이터가 필요하다. 데이터는 문제를 설명할 수 있는 속성(입력 특성)과, 각 데이터에 대한 정답이 포함된 출력 값(레이블)을 함께 가지고 있어야 합니다.
- 필요한 데이터 속성
  - 입력 특성 : 문제를 설명하는 다양한 입력 속성 (예: 이미지 분류를 위한 픽셀 값, 예측 대상에 관련된 환경 변수 등).
  - 레이블(정답 데이터): 각 입력에 대한 정확한 결과 값 (예: 이미지에 레이블된 객체 종류, 시간대별 주가 변동 등).

#### 2. 비지도학습

- 특징 : 비지도학습은 레이블이 없는 데이터를 사용하여 데이터의 구조나 패턴을 찾는 학습 방법이다. 주로 데이터의 군집화, 이상치 탐지 등을 목적으로 사용되므로, 특정 패턴이나 유사성을 파악할 수 있는 다양한 속성을 포함한 데이터가 필요하다.
- 필요한 데이터 속성
  - 다양한 입력 특성: 데이터를 그룹화하거나 유사성을 파악할 수 있도록 여러 가지 속성을 포함 (예: 생체 신호, 거래 내역 등).
  - 레이블이 없음: 정답이나 출력값이 필요하지 않으며, 입력 데이터 자체에 의미 있는 패턴이나 특징이 존재.

#### 3. 강화학습

- 특징 : 강화학습은 에이전트가 특정 환경에서 행동을 학습하여 보상을 극대화하는 방식이다. 상태, 행동, 보상 데이터를 포함한 데이터가 필요하다. 주로 시뮬레이션 환경에서 작동하며, 에이전트가 매 순간의 결과를 학습하고 다음 행동을 결정하는 데 사용된다.
- 필요한 데이터 속성
  - 상태(State): 현재 에이전트가 위치한 환경의 상태 정보를 나타내는 속성 (예: 게임의 현재 상황, 로봇의 위치).
  - 행동(Action): 에이전트가 취할 수 있는 다양한 행동 선택지 (예: 게임에서 움직임, 로봇의 다음 동작).
  - 보상(Reward): 특정 행동을 했을 때 주어지는 결과값. 긍정적 보상은 목표 달성을 돕고, 부정적 보상은 피해야 할 행동을 알림.
- 예시 :
  - 게임 AI 개발 : 게임 속 캐릭터가 취할 수 있는 다양한 움직임(행동)과 각 상황의 상태

정보, 행동에 따른 보상(예: 점수 증가나 감소)을 기록한 데이터를 수집.

- 이유 : 캐릭터가 목표를 달성하기 위한 최적의 행동을 학습하기 위해 상태와 보상 데이터가 필요.
- 로봇 경로 최적화: 로봇의 위치(상태), 다음 이동 방향(행동), 이동 성공 여부(보상)를 수집하여 로봇이 목적지에 도달하는 최적 경로를 학습.
- 이유: 로봇이 효율적으로 이동하도록 보상을 최대화하는 경로를 찾기 위해 각 상태와 행동의 보상이 필요.

## [예시] 사례별 데이터 속성

\* 아래 제시된 것이 고정된 정답은 아님에 주의

### 1. 의료 및 생명과학

- **AI로 심장질환 조기 예측 시스템 구축**
  - 문제 데이터 속성 : 심박수, 혈압, 콜레스테롤 수치, 연령, 성별 등의 생체 데이터.
  - 특징 : 수치형 데이터로, 실시간 모니터링이 가능하며 주로 병원에서 수집된 건강 기록 데이터.
  - 정답 데이터 속성 : 심장질환 여부 레이블 (정상/위험).
  - 특징 : 질병 진단 결과로 구성된 범주형 레이블.
- **의료 영상 분석을 통한 암 조기 진단**
  - 문제 데이터 속성 : X-ray 또는 MRI 영상 데이터.
  - 특징 : 고해상도 이미지 데이터로, 주로 흑백 또는 컬러 이미지 형태로 제공되며 픽셀 수준의 세밀한 정보가 포함됨.
  - 정답 데이터 속성 : 암 유무 레이블.
  - 특징 : 이미지에 대한 이진 레이블로, 특정 질환 유무를 나타냄.
- **병원 내 대기 시간 단축을 위한 AI 솔루션 개발**
  - 문제 데이터 속성 : 진료 종류, 진료 시간, 환자 수, 과거 대기 시간 등의 정보.
  - 특징 : 수치형 및 범주형 데이터로 구성되며, 과거 기록을 바탕으로 시간 예측에 사용.
  - 정답 데이터 속성 : 실제 대기 시간.
  - 특징 : 수치형 데이터로, 대기 시간을 정확히 예측하기 위한 레이블.

### 2. 환경공학 및 지속 가능성

- **AI를 활용한 쓰레기 분리수거 최적화 시스템**
  - 문제 데이터 속성 : 다양한 쓰레기 이미지.
  - 특징 : 이미지 데이터로, 색상 및 형태 정보가 포함되며 재활용 가능 여부에 따라 다양한 쓰레기 이미지가 수집됨.
  - 정답 데이터 속성 : 쓰레기 종류 레이블 (예 : 플라스틱, 유리 등).
  - 특징 : 범주형 레이블로, 이미지 분류에 사용됨.
- **스마트 농업을 위한 AI 기반 작물 모니터링 시스템**
  - 문제 데이터 속성 : 작물 상태, 온도, 습도, 토양 pH 등 다양한 환경 데이터.
  - 특징 : 수치형 데이터로, 계절 및 날씨에 따른 변화를 포함하며 정상 상태 패턴을 학습

하기 위해 사용.

- 정답 데이터 속성 : 없음 (정상 상태 패턴 학습 후 이상치 감지).
- **기후 변화 예측 모델 구축**
  - 문제 데이터 속성 : 과거의 기온, 강수량, 습도 등의 기후 데이터.
  - 특징 : 시계열 수치형 데이터로, 장기적 추세와 변동성을 포함함.
  - 정답 데이터 속성 : 미래 기후 상태.
  - 특징 : 예측 목표로 설정된 수치형 데이터로, 미래 특정 기후 변수의 값.

### 3. 기후 변화와 경제

- **AI를 활용한 탄소 배출량 예측 및 감소 솔루션 개발**
  - 문제 데이터 속성 : 산업별 생산량, 에너지 소비량.
  - 특징 : 수치형 데이터로, 산업별 활동에 따라 변화하며 탄소 배출과의 상관 관계를 포함함.
  - 정답 데이터 속성 : 과거의 탄소 배출량.
  - 특징 : 수치형 레이블로, 예측 목표인 탄소 배출량의 실측 값.
- **기후 변화가 경제에 미치는 영향 분석**
  - 문제 데이터 속성 : 기후 지표(온도, 강수량 등)와 경제 지표(성장률, 수익 등).
  - 특징 : 수치형 데이터로, 기후 변화가 경제 활동에 미치는 영향을 학습하기 위해 사용됨.
  - 정답 데이터 속성 : 경제적 영향(예 : 매출 변화).
  - 특징 : 수치형 레이블로, 기후 변동에 따른 경제 지표 변화.
- **재생 에너지 효율 분석 시스템**
  - 문제 데이터 속성 : 에너지 소비량, 기후 조건, 생산량.
  - 특징 : 수치형 시계열 데이터로, 정상적인 에너지 소비 패턴을 학습하고 비정상 소비 감지에 활용.
  - 정답 데이터 속성 : 없음 (정상 상태의 에너지 소비 패턴을 학습).

### 4. 금융 및 경제

- **AI를 통한 주식 시장 예측**
  - 문제 데이터 속성 : 과거 주가, 거래량, 경제 지표.
  - 특징 : 시계열 수치형 데이터로, 주가 변동 추세와 시장 변동성을 반영.
  - 정답 데이터 속성 : 미래 주가 변동.
  - 특징 : 예측 목표인 수치형 데이터로, 향후 주가의 변동 값을 포함.
- **개인 맞춤형 투자 전략 제안 시스템**
  - 문제 데이터 속성 : 개인의 투자 이력, 리스크 선호도, 재정 상태.
  - 특징 : 수치형 및 범주형 데이터로, 투자 성향과 위험 선호도를 반영함.
  - 정답 데이터 속성 : 맞춤형 투자 전략 레이블.
  - 특징 : 범주형 레이블로, 개별 투자 성향에 맞춘 전략 추천.
- **AI 기반 금융 사기 탐지 시스템**
  - 문제 데이터 속성 (지도학습) : 거래 시간, 거래 금액, 계좌 정보.
  - 특징 : 수치형 및 범주형 데이터로, 거래 패턴을 학습하여 정상/사기 구분.

- 정답 데이터 속성 (지도학습) : 정상 거래/사기 거래 레이블.
- 특징 : 이진 분류 레이블로, 거래 데이터의 정상/사기 여부를 표시.
- 문제 데이터 속성 (비지도학습) : 거래 정보.
- 특징 : 레이블 없이 비정상 거래를 감지하기 위한 패턴 탐색에 사용.
- 정답 데이터 속성 (비지도학습) : 없음 (이상치 패턴 탐지).

## 5. 교육

### • AI 기반 학습 성과 예측 시스템

- 문제 데이터 속성 : 출석률, 과제 점수, 시험 성적 등 학습 관련 데이터.
- 특징 : 수치형 및 범주형 데이터로, 학생의 학업 성과를 예측하는 주요 특성을 포함.
- 정답 데이터 속성 : 학업 성과.
- 특징 : 수치형 또는 범주형 레이블로, 예측 목표인 학업 성과.

### • 맞춤형 학습 추천 AI 시스템

- 문제 데이터 속성 : 학생의 학습 성향, 선호 과목, 학습 시간.
- 특징 : 범주형 데이터로, 각 학생의 학습 성향을 반영.
- 정답 데이터 속성 : 추천 학습 자료.
- 특징 : 범주형 레이블로, 학습 성향에 맞춘 추천 자료.

### • 학생 출석 및 참여도 분석 시스템

- 문제 데이터 속성 : 출석 시간, 참여도, 과제 제출 여부 등의 데이터.
- 특징 : 수치형 및 범주형 데이터로, 정상 패턴을 학습해 비정상 출석 패턴을 탐지.
- 정답 데이터 속성 : 없음 (정상 패턴을 학습하여 이상 감지).

## 6. 사회과학 및 정책

### • AI를 이용한 범죄 예측 및 예방 시스템

- 문제 데이터 속성 : 지역, 시간대, 과거 범죄 유형 및 발생 빈도.
- 특징 : 범주형 및 수치형 데이터로, 특정 지역의 범죄 패턴을 학습.
- 정답 데이터 속성 : 범죄 발생 위험.
- 특징 : 수치형 레이블로, 특정 시간대와 지역의 범죄 발생 가능성을 나타냄.

### • AI 기반 여론 분석

- 문제 데이터 속성 : 소셜 미디어 텍스트, 주요 키워드.
- 특징 : 텍스트 데이터로, 주요 주제와 감정 패턴을 추출하는 데 사용.
- 정답 데이터 속성 : 없음 (주요 패턴과 감정 분석을 통해 여론의 흐름 파악).

### • 정책 효과 분석을 위한 AI 모델 구축

- 문제 데이터 속성 : 정책 시행 전후의 경제 지표, 사회 지표.
- 특징 : 수치형 데이터로, 정책 변화에 따른 사회적 반응을 포함.
- 정답 데이터 속성 : 정책의 사회적 영향.
- 특징 : 예측 목표로 설정된 수치형 데이터로, 정책 효과를 분석하는 레이블.

## 7. 법률 및 윤리

### • AI를 활용한 법률 문서 자동 분석 시스템

- 문제 데이터 속성 : 계약서, 법률 문서.

- 특징 : 텍스트 데이터로, 문서의 주요 조항을 자동으로 탐색하고 분류하는 데 사용.
- 정답 데이터 속성 : 없음 (문서 내 주요 패턴과 특징을 탐지).

• **법적 판결 예측 AI 모델**

- 문제 데이터 속성 : 과거 판례 문서, 사건의 특성.
- 특징 : 텍스트 데이터로, 법적 사건과 판례의 특징을 포함.
- 정답 데이터 속성 : 판결 결과.
- 특징 : 범주형 레이블로, 판례에 따른 판결 결과를 학습.

**8. 도시 계획 및 스마트 시티**

• **AI 기반 교통 최적화 시스템**

- 문제 데이터 속성 : 차량 위치, 교통량, 시간대 등 교통 흐름 데이터.
- 특징 : 시계열 수치형 데이터로, 시간대별 교통 혼잡 패턴을 포함함.
- 정답 데이터 속성 : 최적 경로 및 교통 혼잡도.
- 특징 : 수치형 레이블로, 최적 경로와 혼잡도를 예측.

• **스마트 시티를 위한 에너지 효율 관리 시스템**

- 문제 데이터 속성 : 건물의 시간대별 전력 소비량, 기후 조건.
- 특징 : 시계열 수치형 데이터로, 정상적인 에너지 소비 패턴을 학습하여 이상치를 탐지.
- 정답 데이터 속성 : 없음 (정상 패턴을 학습해 비정상 패턴을 감지).

• **공공 안전 모니터링 시스템**

- 문제 데이터 속성 (지도학습) : CCTV 영상, 위치, 시간대, 사건 유형.
- 특징 : 영상 데이터로, 특정 상황 발생 시 비상 알림 기능을 추가.
- 정답 데이터 속성 (지도학습) : 사고/비사고 레이블.
- 특징 : 범주형 레이블로, 위험 상황을 분류.
- 문제 데이터 속성 (비지도학습) : CCTV 영상 및 공공 안전 관련 데이터.
- 특징 : 비정형 데이터로, 레이블 없이 이상 패턴을 탐지.
- 정답 데이터 속성 (비지도학습) : 없음 (새로운 이상 행동 패턴 탐지).

**9. 스포츠 분석 및 건강관리**

• **AI를 활용한 선수 부상 예측 시스템**

- 문제 데이터 속성 : 심박수, 운동 강도, 운동 시간 등 신체 데이터.
- 특징 : 수치형 데이터로, 운동 중 부상 위험을 분석하는 데 사용.
- 정답 데이터 속성 : 부상 여부.
- 특징 : 이진 분류 레이블로, 부상 여부를 예측.

• **경기 전략 최적화를 위한 AI 분석**

- 문제 데이터 속성 : 경기 데이터 (예 : 위치, 선수 간 거리, 시간).
- 특징 : 시계열 데이터로, 경기 중 발생하는 특정 패턴을 분석하여 최적화.
- 정답 데이터 속성 : 없음 (유사한 패턴을 통해 최적 전략을 도출).

• **AI 기반 피트니스 개인화 프로그램 개발**

- 문제 데이터 속성 : 개인의 운동 기록, 체력 수준, 건강 목표.
- 특징 : 수치형 데이터로, 맞춤형 운동 계획을 수립하는 데 사용.

- 정답 데이터 속성 : 최적의 운동 프로그램.
- 특징 : 범주형 레이블로, 맞춤형 운동 추천을 학습.

## 10. 예술

### • AI 기반 음악 생성 시스템

- 문제 데이터 속성 : 다양한 장르의 음악 데이터.
- 특징 : 비정형 데이터로, 생성 모델을 통해 새로운 음악을 학습.
- 정답 데이터 속성 : 없음 (음악 생성에 특화된 모델 활용).

### • AI를 이용한 그림 및 이미지 생성

- 문제 데이터 속성 : 다양한 스타일의 그림 및 이미지 데이터.
- 특징 : 이미지 데이터로, 생성 모델을 통해 새로운 스타일 이미지를 생성.
- 정답 데이터 속성 : 없음 (이미지 생성에 특화된 모델 활용).

### • AI를 활용한 영화 시나리오 생성

- 문제 데이터 속성 : 영화 대본, 시나리오 텍스트 데이터.
- 특징 : 텍스트 데이터로, 자연어 생성 모델을 통해 새로운 시나리오 작성.
- 정답 데이터 속성 : 없음 (새로운 시나리오 생성에 특화된 모델 활용).

## Step B-2. 데이터 수집

앞서 정의한 데이터 속성을 바탕으로 본격적으로 딥러닝을 위한 데이터를 수집해보자. 그리고 보고서에 캡처하거나 표의 형태로 데이터의 일부를 이곳에 작성해보자.

데이터를 직접 수집하는 방법도 있지만 정부나 민간 기관에서 제공하는 데이터를 사용할 수 있다. 다음과 같은 데이터 제공 사이트에 방문하여 어떤 데이터가 있는지 살펴보자. 나의 진로와 관련된 데이터도 찾고, 해결하고 싶은 문제에 대한 힌트도 얻을 수 있다. 단, 데이터 수집은 저작권, 개인 정보 보호법 등을 주의하며 수집하도록 한다.

<참고할 수 있는 데이터 수집 사이트>

데이터 종류	사이트
정형 데이터 중심	공공 데이터 포털 ( <a href="https://www.data.go.kr">https://www.data.go.kr</a> )
비정형 데이터 중심	AI Hub ( <a href="https://www.aihub.or.kr">https://www.aihub.or.kr</a> )
국내 민간 정형·비정형 데이터	데이콘 ( <a href="https://www.dacon.io">https://www.dacon.io</a> )
외국 민간 정형·비정형 데이터	캐글 ( <a href="https://www.kaggle.com">https://www.kaggle.com</a> )

### [참고]

의료 영상 데이터를 확보하기 어려운 경우, 다음과 같은 방법으로 데이터 접근성을 높일 수 있다.

1. **공개 데이터셋 활용** : 여러 연구 기관과 대학이 의료 영상 데이터셋을 공개하고 있다. 대표적인 의료 영상 데이터셋은 다음과 같다.
  - NIH Chest X-ray Dataset : 미국 국립보건원(NIH)에서 제공하는 흉부 X-ray 이미지 데이터셋으로, 여러 질병에 대한 레이블이 포함되어 있다.
  - ISIC Archive : 피부암 검출 연구를 위한 피부 이미지 데이터셋을 제공하는 플랫폼으로, 다양한 피부 질환에 대한 이미지와 레이블이 포함되어 있다.
  - LUNA16 : 폐암 스크리닝을 위한 CT 이미지 데이터셋으로, 폐 결절 데이터가 포함되어 있다.
2. **모의 데이터셋 사용** : 의료 영상 데이터를 모사한 가상 데이터나 합성 데이터를 사용해 학습하는 방법도 있다. 이러한 데이터는 원본 이미지와 유사하게 만들어져 학습 목적으로 유용하게 활용될 수 있다.
3. **데이터 증강(Data Augmentation)** : 제한된 수의 의료 영상 데이터가 있는 경우, 이미지 회전, 크기 변경, 밝기 조정 등을 통해 기존 이미지를 변형하여 데이터의 양을 늘릴 수 있다. 이는 데이터 다양성을 높이고, 모델의 일반화 성능을 향상하는 데 도움을 준다.
4. **전이 학습(Transfer Learning)** : 전이 학습은 사전 학습된 모델을 활용하여 적은 데이터셋에서도 좋은 성능을 낼 수 있도록 하는 방법이다. 대규모 이미지 데이터셋(예:

ImageNet)으로 미리 학습된 모델을 사용해 의료 영상 데이터에 맞춰 재학습(미세 조정)하는 방식으로, 데이터가 적어도 모델 성능을 높일 수 있다.

5. 시뮬레이션 또는 오픈 소스 데이터 생성기 : 의료 영상 데이터를 생성하는 오픈 소스 도구나 시뮬레이터를 사용하는 방법이다.

<참고 사이트>

1. <https://www.syntho.ai/ko/>
2. [https://m.blog.naver.com/sdt\\_wiki/222941105005](https://m.blog.naver.com/sdt_wiki/222941105005)

### Step B-3. 데이터 분석

앞서 수집한 데이터는 ‘날것의 데이터(raw data)’이므로 학습을 위해 적절히 가공할 필요가 있는데, 이를 ‘전처리’라고 한다. 예컨대, 여러분이 학습을 위해 구매하는 문제집(데이터)은 학문적 지식을 학교 선생님들이 잘 가공한 결과물이다.

데이터 전처리를 하려면 우선 데이터를 분석하여 데이터의 속성, 특징 등을 파악하고 데이터가 가진 의미를 해석하는 작업이 필요하다. Step B-1에서 내가 필요하다고 생각했던 속성들을 가진 데이터를 잘 수집하였는지 확인해보고, 내가 속성을 잘 정의한 것이 맞는지 판단해보자. 또한 수집한 데이터를 시각화하여 데이터의 속성을 파악해보고, 이를 통해 학습에 사용할 속성과 제외할 속성을 결정해보자.

#### ※ 참고

- 정형 데이터는 수집이 상대적으로 수월하지만, 학습에 사용할 속성과 그렇지 않은 속성을 결정하는 것이 중요하다. 반면 비정형 데이터는 학습에 사용할 속성과 그렇지 않은 속성을 결정하는 과정이 생략될 수 있다.
- 데이터의 의미를 해석하기 어려울 때는 데이터의 맥락을 잘 이해하고 있는 전문가에게 자문을 구할 수 있다.

## Step B-4. 데이터 전처리

데이터 전처리의 목적은 데이터의 품질과 일관성을 확보하는 데 있다. 데이터 분석 결과를 토대로 데이터를 적절하게 가공해보자.

### [예시] 데이터 전처리를 위한 기준

데이터 분석 과정에서 확인한 누락된 값이나 이상한 값을 처리하고, 학습에 적합한 형태로 데이터를 가공한다. 자연어 데이터를 형태소나 품사 단위로 처리할 수도 있고, 이미지 데이터의 경우 부족한 데이터를 늘릴 수도 있다. 이러한 과정을 '전처리'라고 하며, 일반적으로 전체 문제 해결 과정 중 많은 시간이 소요된다.

다음은 데이터 전처리를 위한 기준으로 자주 사용되는 예시이다.

#### [기준 1] 누락된 값은 없는가?

- 누락된 값(NaN, null 등)은 모델 학습에 방해.
- 데이터에 NaN(Not a Number), null, 빈 값 등이 존재하는지 확인. 누락된 값이 있다면, 제거하거나 다른 값으로 대체할 방법을 결정.
- 처리 방법 :
  - 삭제 : 데이터의 일관성에 큰 영향을 미치지 않는 경우 해당 샘플이나 열을 제거.
  - 대체 : 중요한 특성이라면 평균, 중앙값, 최빈값, KNN 또는 모델에 따라 학습 가능한 대체 방법으로 채움.

#### ※ KNN을 이용한 대체 방법

KNN(K-Nearest Neighbors)을 이용해 누락된 값을 대체하는 방법은, 주변 데이터의 유사한 값들을 참고하여 누락된 값을 채우는 방식이다. KNN 알고리즘은 가장 가까운 K개의 이웃 데이터를 기준으로 해당 값들을 평균 또는 최빈값으로 채워준다.

<예시>

다음과 같은 학생 성적 데이터가 있다.

학생	수학 성적	영어 성적	과학 성적
A	90	85	88
B	85	NaN	82
C	92	80	84
D	70	88	77

여기서 학생 B의 영어 성적이 누락되어 있다. 이때, KNN을 이용하여 이 값을 다음과 같이 채울 수 있다.

#### 1. KNN 적용

수학 성적과 과학 성적을 기반으로 B 학생과 가장 유사한 학생들 K명(K=2 또는 3)을 찾는다. 예를 들어, 학생 A와 학생 C가 B와 성적 패턴이 비슷한 경우가 많다면, 이들이 B 학생의 "이웃"으로 간주된다.

## 2. 누락된 값 대체

B 학생의 이웃으로 선택된 A와 C 학생의 영어 성적(85와 80)의 평균을 계산하거나 최빈값을 선택한다. 즉, 평균값(82.5)을 채우거나, 이 값들을 참고하여 B 학생의 영어 성적 누락값을 82로 대체할 수 있다.

이처럼 KNN을 활용해 가장 유사한 이웃들의 값으로 누락된 데이터를 채워, 데이터의 일관성을 유지하면서 누락값을 보완할 수 있다.

### [기준 2] 이상값(outliers)은 없는가?

- 이상값은 모델의 학습 성능을 저하.
- 비정상적으로 너무 크거나 작은 값이 있는지 확인. 이상값은 예측 모델 성능에 큰 영향을 미칠 수 있음.
- 처리 방법 : IQR(Interquartile Range)이나 Z-score 개념을 활용하여 이상값을 식별하고 제거하거나 적절한 값으로 대체.

### [기준 3] 각 특성(feature)의 데이터 타입이 모델에 적합하게 설정되었는가?

- 모든 특성(feature)이 수치형 데이터로 변환되어야 함. 수치형 데이터는 모델이 연산에 사용할 수 있는 형식이므로, 텍스트나 범주형 데이터는 수치형으로 변환.
- 모든 열의 데이터 타입을 확인하고, 목적과 상황에 맞게 단순 수치형, One-Hot Encoding, Embedding 등의 방법을 사용하여 변환.
- 처리 방법 :
  - 수치형 데이터 : 문자열로 저장된 수치형 데이터는 정수형(Integer) 또는 실수형(Float)으로 변환하여 계산 가능하도록 함.
  - 범주형 데이터 : 범주형 데이터는 One-Hot Encoding 또는 Embedding을 사용하여 수치형으로 변환하여, 모델이 각 범주의 의미를 왜곡 없이 학습할 수 있도록 함.
- 예시 :
  - 1) 숫자가 포함된 문자열 데이터 → 수치형 데이터 변환
    - 고객 연령 정보 : age = ["20", "35", "42", "50"]
    - 문제점 : 숫자처럼 보이지만 '문자열(string)'로 되어 있어 계산이나 분석이 어려움.
    - 변환 방법 : 숫자처럼 보이는 문자열을 '정수형(int)' 또는 '실수형(float)'으로 변환.
    - 변환 후 결과 : 수치형 데이터 age = [20, 35, 42, 50]
  - 2) 범주형 데이터 → 원핫인코딩 변환
    - 고객 성별 정보 : gender = ["남성", "여성", "여성", "남성", "남성"]
    - 문제점 : 문자열로 표현된 카테고리형 데이터로, 계산과 분석에 비효율적.
    - 변환 방법 : 남자는 [1, 0], 여자는 [0, 1]로 변환.
    - 변환 후 결과 : 수치형 데이터 gender = [[1, 0], [0, 1], [0, 1], [1, 0], [1, 0]]
  - 3) 날짜/시간 데이터를 적절한 수치형 형식으로 변환
    - 날짜/시간 데이터는 분석에 유용하지만 문자열로 저장되어 있으면 학습에 불리하므로 날짜를 연, 월, 일, 시간 등으로 분리하여 추가 특성으로 사용하거나, 시간 간격 등을 계산해 활용.

- 예시 : "2023-01-01"을 연(2023), 월(1), 일(1)로 분리.

- ▶ 참고로, "남자는 0, 여자는 1"로 간단하게 수치형 데이터로 변환( Label Encoding)할 수도 있지만, 딥러닝 모델에서 성별이 0과 1로 표시된 수치형 데이터로 입력될 경우, 모델은 이 데이터를 연속적 수치로 이해할 가능성이 있다. 이로 인해 의미가 왜곡될 수 있다. 연속적인 수치형 데이터는 딥러닝 모델에서 크기나 순서에 따라 해석된다. 예를 들어, 모델은 1이 0보다 크므로, 남성이 여성보다 더 큰 값을 가진다고 해석할 수 있다. 이는 성별이 가지고 있는 본래 의미와는 맞지 않다.

#### ※ 딥러닝 관점에서 데이터 타입 설정의 의미

- 딥러닝 모델은 수치형 입력을 필요로 하므로, 범주형 데이터는 반드시 수치형으로 변환되어야 한다. 그러나 단순한 수치형 변환이 아닌 적절한 인코딩 방법을 통해 데이터의 본래 의미를 보존하면서 의미 있는 수치형 형태로 변환하는 것이 중요하다. 따라서 "데이터 타입을 올바르게 설정"한다는 것은 단순히 데이터를 수치형으로 만드는 것이 아니라, 데이터의 의미를 반영할 수 있는 방식으로 수치화하여 모델이 더 잘 이해하고 학습할 수 있도록 준비하는 것을 의미한다.

#### [기준 4] 중복된 데이터가 있는가?

- 중복된 데이터는 모델 학습에 불필요한 정보를 추가하고, 과적합을 유발.
- 같은 데이터가 반복되는 경우 중복된 행이나 열을 확인하여 제거.
- 처리 방법 : `drop_duplicates()` 등 함수를 통해 중복을 제거.

#### [기준 5] 수치형 특성에 대한 스케일링(정규화 또는 표준화)이 필요한가?

- 딥러닝에서는 데이터 스케일이 중요. 특히, 신경망에서 큰 값을 처리할 때 학습 효율에 영향을 줄 수 있음. 수치형 특성의 값 범위가 크게 차이 나면 모델 학습에 방해.
- 스케일링 방법 :
  - 정규화(Normalization) : 모든 값을 0~1 사이로 맞추는 방식.
  - 표준화(Standardization) : 평균이 0, 표준편차가 1이 되도록 맞추는 방식.
- 예시 : 소득 데이터가 [0, 10000] 범위일 때 [0, 1] 범위로 정규화.

#### [기준 6] 데이터의 분포를 확인하여 균형을 맞추었는가?

- 데이터의 분포를 확인하여, 편향이 없는지 확인.
- 특정 클래스나 값에 편향이 있을 경우 모델의 성능에 영향. 즉, 클래스가 불균형하면 모델이 특정 클래스에 편향될 수 있음.
- 처리 방법 : 데이터의 분포를 시각화하고, 불균형이 심할 경우 언더샘플링 또는 오버샘플링으로 데이터 균형 조정.
- 예시 : 분류 문제에서 양성(positive)과 음성(negative) 클래스의 비율을 맞추기.

#### [기준 7] 불필요한 특성(FEATURE)은 없는가?

- 불필요한 특성(중복되거나 관련이 없는 데이터)은 모델의 복잡성만 높이며 학습 속도 및 성능 저하 유발.
- 처리 방법 :

- 상관관계 분석 또는 도메인 지식을 활용해 비효율적 특성을 제거.
- feature selection 기법을 통해 유의미한 특성만 남김.
- 예시 : 중복된 고객 ID 열이나 날짜 형식의 문자열.

#### [기준 8] 데이터 샘플의 순서가 무작위로 섞였는가?

- 데이터가 정렬된 상태로 학습될 경우 모델이 학습 데이터에 편향될 가능성.
- 처리 방법 : shuffle을 통해 순서를 무작위로 섞은 후 학습.
- 예시 : 시간 순으로 정렬된 데이터라면 랜덤 셔플링하여 학습.

#### [기준 9] 클래스 불균형을 고려했는가?

- 분류 문제에서 클래스 간 비율이 너무 차이 나면, 즉 특정 클래스의 데이터가 과소 대표 될 경우 예측 성능에 영향을 미침.
- 처리 방법 :
  - 오버샘플링 : 적은 클래스의 데이터를 증강.
  - 언더샘플링 : 많은 클래스의 데이터를 감소.
- 예시 : 사기 탐지 데이터에서 사기(positive)와 정상(negative) 비율을 조정.

## Step C. 인공지능 구현 [Optional]

### Step C-1. No-Code Tool을 이용한 딥러닝

Orange3 프로그램을 활용하여 인공신경망을 설계해보고, 준비된 데이터를 학습시켜보자. 원하는 성능을 얻었는가? 얻지 못했다면 그 이유를 찾아보고 원하는 성능이 나올 때까지 하이퍼 파라미터를 조정해보자. 그 과정에서 발견한 최적의 하이퍼 파라미터를 설명해보자.

### Step C-2. Python을 이용한 딥러닝 [Challenge]

지금까지 수행한 프로젝트 전 과정을 파이썬을 이용하여 수행해보자. 특히 Orange3 프로그램에서 제공하지 않았던 다양한 하이퍼 파라미터들을 추가해보고, 어떤 결과 차이가 있는지 탐구해보자.

## Step Final. 프로젝트 결과 분석 및 외고

### Step 1. 프로젝트 수행 과정 복기 및 외고

다음 순서에 따라 프로젝트를 수행한 과정을 정리해보자.

1. 이 프로젝트를 위해 정의한 문제를 제기하게 된 동기는 무엇이었는가?  
평소에 그와 관련해서 특별한 문제의식이 있었는가?
2. 문제를 해결하기 위한 아이디어는 어떻게 떠올릴 수 있었는가?  
그 아이디어를 떠올릴 수 있었던 특별한 배경이 있는가?
3. 데이터를 수집하고 처리하는 과정은 원활하게 잘 진행되었는가?  
아니라면 이유가 무엇이라고 생각하는가?
4. 결론적으로 이 프로젝트를 통해서 최초로 설정했던 목표를 달성했다고 생각하는가? 원하는 결과를 얻지 못했거나 중간 단계에서 문제가 생겨서 프로젝트를 끝까지 완수하지 못했을 수도 있다. 프로젝트 과정에서 아쉬운 부분이 있다면 무엇이고, 그 이유는 무엇이라고 생각하는가? 또, 어떤 부분이 가장 어려웠는가?

## Step 2. 의미/시사점 도출

최종적으로 이 프로젝트를 통해서 무엇을 배웠고, 무엇을 얻을 수 있었는가? 지식을 얻었을 수도 있고, 중요한 통찰을 얻었을 수도 있으며, 진로와 관련하여 내가 앞으로 나아가야 할 방향을 설정하는 계기가 되었을 수도 있다. 무엇이 되었든 이 프로젝트를 통해 배우고 얻은 것에 대해서 정리해보자. 그리고 이 프로젝트가 자신에게 어떤 의미였는지 본 프로젝트의 의미 또는 시사점을 적어보자.

## Step 3. 토론 및 발표

프로젝트 결과를 친구들과 공유하고 토론하면서 피드백도 받아보자. 마지막으로 생각을 가다듬고 정리해서 자신 있게 발표해보자.

## 참 고 문 헌

프로젝트를 수행하면서 참고했던 주요 자료들의 출처를 적어주세요.