
CH-1

1. 분석을 위한 컴퓨팅 환경을 구성할 때, "가상환경"을 만들어 분석에 필요한 프로그램과 모듈을 설치하여 로컬 컴퓨팅 환경과 분리하는 이유가 아닌 것은?
 1. 의존성 격리
 2. 프로젝트 관리의 용이성
 - ♥3) 협업을 위한 형상과 버전관리 용이
 3. 종속성의 명시성
 4. 쉬운 관리 및 유지보수
2. 머신러닝의 학습방법에 따른 일반적인 분류로 학습 데이터의 "라벨:(정답)"의 존재 여부에 따라 분류한 종류에 해당하지 않는 것은?
 1. 지도 학습(Supervised Learning)
 2. 준지도 학습(Semi-Supervised Learning)
 3. 강화 학습(Reinforcement Learning)
 4. 비지도 학습(Unsupervised Learning)
 - ♥5) 자기주도 학습(self-directed learning)
3. Python 버전 관리 도구로 시스템에 여러 버전의 Python을 설치하고 관리하며, 프로젝트별로 특정 버전을 설정할 수 있도록 도와주는 프로그램은 무엇인가?
 1. Docker
 - ♥2) pyenv
 2. git
 3. virtualBox
 4. wsl
4. DIKW 피라미드를 설명한 내용이다. 맞게 설명한 것을 고르시오.
 1. 데이터 (Data) : 지식을 체계화하고 연결하여 얻는 이해
 2. 정보(Information) : 가공되지 않은 객관적인 사실, 숫자, 문자 등
 3. 지식(Knowledge) : 데이터를 체계화하고 연결하여 얻는 이해
 - ♥4) 데이터는 정보의 기초이며, 정보는 지식의 기초, 지식은 지혜의 기초이다.
 4. 피라미드가 올라갈수록 데이터의 가치와 활용 가능성이 감소한다.
5. 다음은 데이터의 유형과 지식유형에 대한 설명이다. 적절히 기술한 것을 선택하십시오.
 1. 정성적 데이터(qualitative data) : 언어•문자 등 비정형 데이터, 상대적으로 적은 비용과 기술적 투자 노력이 수반된다.

2. 정량적 데이터(quantitative data) : 수치•도형•기호 등, 저장•검색•분석 활용에 용이한 데이터이다. ♥
 3. 데이터는 지식경영의 핵심이슈인 암묵지와 형식지의 상호작용에 영향을 주지 않는다.
 4. 암묵지 : 학습과 체험을 통해 개인에게 습득된 무형의 지식으로 공유가 용이한 과학적 발견을 말한다.
 5. 형식지 : 형상화된 지식, 유형의 대상이 있어 지식의 전달과 공유가 매우 어렵다.
6. 다음 분석모델 분류에 대한 설명이다. 설명이 잘못된 것을 선택하시오.
- ♥1) 군집화모델: 데이터 포인트가 특정 카테고리에 속하는지 예측하는 데 사용
 - 2) 예측모델: 과거 데이터를 기반으로 미래 값을 예측하는 데 사용
 - 3) 분류모델: 데이터 포인트가 특정 카테고리에 속하는지 예측하는 데 사용
 - 4) 이상탐지모델: 데이터에서 정상 범위를 벗어나는 데이터 포인트를 식별하는 데 사용.
 - 5) 추천모델: 사용자에게 적합한 제품, 서비스 또는 콘텐츠를 추천하는 데 사용
7. 다음은 PARA 정리법에 관한 설명이다 옳지 않은 것을 고르시오.
1. Project 분류는 목표와 데드라인이 명확한 내용들에 관한 것들이다.
 2. Area 분류는 목표와 데드라인이 없지만 꾸준히 신썬야 할 것들이다.
 3. Resource 분류는 자료나 관심있는 것을 주제별로 모아둠
 4. Archive 분류는 더이상 필요없는 장기 보관되는 내용들을 보관하는 분류이다.
 - ♥5) 한번 지정된 분류는 특별한 일이 없는 이상 다른 분류로 분류하지 않는다.
8. 머신러닝과 일반 연구 방법론의 주요 차이점으로 가장 적절한 것은 무엇인가?
- ♥1) 머신러닝은 데이터 기반 예측에 초점을 맞추고, 일반 연구 방법론은 이론 검증에 초점을 맞춘다.
 - 2) 머신러닝은 가설 설정 및 검증 과정을 반드시 포함하지만, 일반 연구 방법론은 그렇지 않다.
 - 3) 머신러닝은 통계적 유의성을 필수적으로 요구하지만, 일반 연구 방법론은 그렇지 않다.
 - 4) 머신러닝은 결과 해석의 명확성을 최우선으로 고려하지만, 일반 연구 방법론은 그렇지 않다.
 - 5) 머신러닝은 인간의 개입을 최소화하는 것을 목표로 하지만, 일반 연구 방법론은 그렇지 않다.
9. 다음 중 정성적 데이터와 정량적 데이터의 차이에 대한 설명으로 가장 적절하지 않은 것은 무엇인가?
1. 정성적 데이터는 주로 비수치적 형태로 표현되며, 정량적 데이터는 수치적 형태로 표현된다.
 2. 정성적 데이터는 주로 관찰, 인터뷰, 설문조사 등을 통해 수집되며, 정량적 데이터는 측정, 계산 등을 통해 수집된다.
 3. 정성적 데이터는 데이터의 의미와 맥락을 이해하는 데 유용하며, 정량적 데이터는 통계적 분석 및 수치 비교에 유용하다.

4. 정성적 데이터는 범주형 데이터와 순서형 데이터로 나눌 수 있으며, 정량적 데이터는 이산형 데이터와 연속형 데이터로 나눌 수 있다. [3.1. 통계 분석 기법](#)

♥5) 정성적 데이터는 주로 평균, 분산 등의 통계량을 사용하여 분석하며, 정량적 데이터는 빈도, 비율 등의 통계량을 사용하여 분석한다.

10. MLOps에서 ML(Machine Learning) 부분과 OPs(Operations) 부분을 비교할 때, 다음 중 가장 적절하지 않은 설명은 무엇인가?

1. ML 부분은 모델 개발 및 실험에 초점을 맞추고, OPs 부분은 모델 배포 및 운영에 초점을 맞춘다.
2. ML 부분은 데이터 과학자의 전문성이 중요하며, OPs 부분은 엔지니어의 전문성이 중요하다.
3. ML 부분은 모델의 정확도 및 성능 평가에 집중하고, OPs 부분은 모델의 안정성 및 확장성에 집중한다.
4. ML 부분은 코드 버전 관리 및 실험 추적에 중점을 두고, OPs 부분은 인프라 관리 및 배포 자동화에 중점을 둔다.

♥5) ML 부분은 모델 재학습 및 성능 모니터링을 포함하며, OPs 부분은 모델 개발에 필요한 데이터 수집 및 전처리를 담당한다.