**REPORT**

(CHAPTER 2)

로고, 상징, 등록 상표, 폰트이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

|  |  |
| --- | --- |
| 이름 | 김민서 |
| 학과 | 컴퓨터인공지능공학부 |
| 학번 | 202213083 |
| 과목명 | 인공지능입문 |
| 교수님 | 진예지 교수님 |
| 분반 | 105분반 |
| 제출일 | 2025.03.21 |

**1. 오렌지는 데이터 타입을 4가지로 구분하여 사용한다. 44가지 타입에 대해 설명하시오.**

|  |  |
| --- | --- |
| 데이터 타입 | 설명 |
| 수치형(numeric) | 숫자로 표시되는 대부분의 데이터 |
| 날짜(datetime) | 날짜 |
| 텍스트(text) | 문자열, 문장, 문단 |
| 범주형(categorical) | 정해진 개수의 범주(카테고리)로 나타나는 값 |

**2. 다음 데이터가 어떤 타입에 속하는지를 근거를 들어 답하시오.**

1) 한국에 거주하는 20대 남녀의 체중

- Numeric : 체중은 연속적인 숫자 데이터

2) 서울대학교에 재학 중인 학생들의 생일

- Datetime : 생일은 날짜 데이터

3) 유엔에 정회원으로 가입한 193개국 국가명

- Text : 193개의 국가명은 모두 텍스트 데이터

4) 13자리의 주민등록번호

- Text : 주민등록번호는 숫자로 표시되지만, 그 기능은 문자에 가깝기에 텍스트 데이터

5) 핸드폰 번호

- Text : 주민등록번호와 같이 숫자로 표시되지만, 그 기능은 문자에 가깝기에 텍스트 데이터

**3. 아래에 제시된 위젯은 Data 카테고리에 속해 있다. 각 위젯의 이름과 기능을 비교하여 설명하시오.**

|  |  |  |
| --- | --- | --- |
| 위젯 | 이름 | 기능 |
|  | Datasets | 오렌지의 온라인 저장소에 있는 데이터셋을 불러오는 위젯 |
|  | File | 외부에 있는 데이터셋을 불러오는 위젯  (엑셀이나 구글 스프레드시트 등) |

**4. 아래에 제시된 위젯은 Transform 카테고리에 속해있다. 이 위젯의 이름과 기능을 설명하시오.**

|  |  |  |
| --- | --- | --- |
| 위젯 | 이름 | 기능 |
|  | Pivot Table | 열을 기준으로 데이터 테이블 모양을 변경하는 위젯 |

**5. 오렌지에서 Feature Statistics 위젯을 사용하면 데이터셋에 대해 7가지 종류의 기초 통계량을 확인할 수 있다. 7가지 기초 통계량의 종류에 대해 설명하시오.**

|  |  |
| --- | --- |
| 기초 통계량 | 설명 |
| Distribution | 값의 범위를 일정한 크기의 구간별로 나누어  빈도수를 나타낸 그래프 |
| Mean | 데이터의 모든 값의 총합을 데이터의 개수로 나눈 값 |
| Median | 데이터의 값을 정렬한 후, 전체 데이터 중 가운데에 있는 수 |
| Dispersion | 값들이 평균값에 밀집되어 있는지 퍼져 있는지를 나타내는 척도 |
| Min | 데이터 중 제일 작은 값 |
| Max | 데이터 중 제일 큰 값 |
| Missing | 데이터에 값이 없음, 누락된 데이터 |

**6. 다음에 제시된 그림은 2.2장에서 사용한 학생 성적 데이터셋이다. 이 그림에서 1) 데이터 샘플, 2) 변수, 3) 메타 변수, 4) 특징, 5) 타겟을 찾고 각 용어에 대해 설명하시오.**

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.

|  |  |  |
| --- | --- | --- |
| 용어 | 데이터 | 설명 |
| 데이터 샘플 | 5개 각각의 가로 행 | 표에서 하나의 행을 가리킴 |
| 변수 | 학번, 이름, 학과,  나이, 성별, 생년월일,  출석횟수, 총점, 성적 | 표에서 각 열의 이름을 가리킴 |
| 메타 변수 | 학번, 이름 | 인스턴스의 이름, 데이터 분석의 대상이 아님 |
| 특징 | 출석횟수, 총점 | 원인이 되는 변수 |
| 타겟 | 성적 | 특징 변수의 값에 따라 변화할 것으로 추정되는 변수 |

* 학과, 나이, 성별, 생년월일은 데이터 분석의 대상이 될 수는 있으나,

타겟에 직접적인 원인이 되는 변수는 아님.

**7. 2.1장에서 사용한 Iris 데이터셋은 인공지능 분야에서 오랫동안 활용되어 온 유명 데이터셋이다. 해당 데이터셋에 대해 설명하고 1) 데이터 샘플, 2) 변수, 3) 특징, 4) 타겟에 대해 설명하시오.**

**텍스트, 스크린샷, 소프트웨어, 번호이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.**

* Iris 데이터셋은 통계학자이자 생물학자인 로널드 피셔가 1936년에 소개한 것으로, 꽃잎과 꽃받침의 길이/너비 정보에 따라 붓꽃의 종류를 예측하는 인공지능 모델을 만드는 것이 목표인 데이터셋
* 붓꽃의 종류별로 50개씩, 총 150개 꽃에 대한 데이터가 정리되어 있음

1) 데이터 샘플 : 150개 각각의 가로 행

2) 변수 : iris, sepal length, sepal width, petal length, petal width

3) 특징 : sepal length, sepal width, petal length, petal width

4) 타겟 : iris

**8. 2.1장의 Step 4에서 Iris-setosa 데이터를 선택하여 기초 통계량을 확인하였다. Iris-versicolor와 Iris-verginica 데이터 각각의 기초 통계량을 확인하시오.**

|  |  |
| --- | --- |
| Iris-versicolor | Iris-verginica |
| 텍스트, 스크린샷, 소프트웨어, 컴퓨터 아이콘이(가) 표시된 사진  AI가 생성한 콘텐츠는 부정확할 수 있습니다. | 텍스트, 스크린샷, 소프트웨어, 컴퓨터 아이콘이(가) 표시된 사진  AI가 생성한 콘텐츠는 부정확할 수 있습니다. |

**9. 오렌지의 Dataset 위젯을 이용하여 Course Grades 데이터셋을 불러올 수 있다. Data Table 위젯을 이용하여 이 데이터셋의 정보 1)~3)을 설명하시오.**

1) 데이터 샘플 : 16개 각각의 가로 행

2) 변수 : Student, English, French, History, Algebra, Biology, Physics, Physical

3) 메타 변수 : Student

**10. 9번 문제에서 Course Grades 데이터셋을 불러온 후, Data Table 위젯을 이용하여 해당 데이터셋의 정보를 파악해보았다. 이번에는 Feature Statistics 위젯을 이용하여 각 과목의 기초 통계량을 분석하시오.**

텍스트, 스크린샷, 소프트웨어, 직사각형이(가) 표시된 사진

AI가 생성한 콘텐츠는 부정확할 수 있습니다.