**REPORT**

(CHAPTER 8)

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 | 2025.06.11 |

**6) 자연어를 처리할 때, Bag of Words 등의 기법을 사용하여 텍스트를 벡터 형태로 변환하는 이유는 무엇인가?**

- 인공지능 모델을 구성하기 위해서는 텍스트 데이터를 벡터 공간에 수치로 표현하는 임베딩 작업이 필요하다.

* Bag of Words는 비슷한 주제의 문서에서 등장하는 단어가 유사할 것으로 가정하는 방식
* 컴퓨터는 문자열 자체를 이해하지 못함
* 텍스트 데이터를 숫자 벡터로 바꿔야 머신러닝 알고리즘에 입력 가능

**7) 8.4장에서 감성 분석을 수행하였다. 아마존에서 알렉사를 구입하고 남긴 리뷰 데이터를 활용하여 다음과 같은 추가 분석을 진행하시오.**

- 기존의 결측치를 제거하는 Select Rows에 조건 추가

- rating 점수가 5점인 데이터만 선택

1) rating 점수 5점에 해당하는 리뷰 데이터를 워드 클라우드로 구성

|  |  |  |
| --- | --- | --- |
|  |  |  |

- 전체 인스턴스 수가 2241개로 감소

- love > great > echo등의 긍정적인 단어들이 상위에 등장함

2) rating 점수 1점에 해당하는 리뷰 데이터를 워드 클라우드로 구성

|  |  |  |
| --- | --- | --- |
|  |  |  |

* 전체 인스턴스 수가 146개로 감소
* work > echo 등 기능과 관련된 단어들이 상위에 등장함
* 기능적인 평가에 대한 리뷰가 있었을 것으로 추정
* rating 점수에 따른 인스턴스 수로 보았을 때, 긍정적인 리뷰가 훨씬 많음