**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 | 2025.06.05 |

**5) 다음 5개 문서를 Bag of Words 기법을 사용하여 벡터화하시오. 단, 대소문자의 차이와 구두점은 무시하시오.**

|  |
| --- |
| 문서 1) To be or not to be, that is the question. |
| 문서 2) You need to be prepared. |
| 문서 3) Do you have any questions? |
| 문서 4) That was not my question. |
| 문서 5) I want to be a violinist. |

1. 문서 전처리(소문자화 / 구두점 제거)

|  |
| --- |
| to be or not to be that is the question |
| you need to be prepared |
| do you have any questions |
| that was not my question |
| i want to be a violinist |

1. 단어 목록

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 문서1 | 문서2 | 문서3 | 문서4 | 문서5 |
| a | 0 | 0 | 0 | 0 | 1 |
| any | 0 | 0 | 1 | 0 | 0 |
| be | 2 | 1 | 0 | 0 | 1 |
| do | 0 | 0 | 1 | 0 | 0 |
| have | 0 | 0 | 1 | 0 | 0 |
| i | 0 | 0 | 0 | 0 | 1 |
| is | 1 | 0 | 0 | 0 | 0 |
| my | 0 | 0 | 0 | 1 | 0 |
| need | 0 | 1 | 0 | 0 | 0 |
| not | 1 | 0 | 0 | 1 | 0 |
| or | 1 | 0 | 0 | 0 | 0 |
| prepared | 0 | 1 | 0 | 0 | 0 |
| question | 1 | 0 | 0 | 1 | 0 |
| questions | 0 | 0 | 1 | 0 | 0 |
| that | 1 | 0 | 0 | 1 | 0 |
| the | 1 | 0 | 0 | 0 | 0 |
| to | 2 | 1 | 0 | 0 | 1 |
| violinist | 0 | 0 | 0 | 0 | 1 |
| want | 0 | 0 | 0 | 0 | 1 |
| was | 0 | 0 | 0 | 1 | 0 |
| you | 0 | 1 | 1 | 0 | 0 |