

데이터 마이닝 프로젝트

B

1 주제 선정 이유

2 데이터 소개 및 전처리

3 감성 분석 결과

4 문제점 및 활용방안



Part 1

# 주제 선정 이유





## 데이터 소개

ID

타입: 정수형

의미: 네이버에서 제공된

아이디

NA: 0

### Document

타입: 문자형

의미: 실제 리뷰(문장)

NA: 5

### Label

타입: 정수형

의미: 리뷰의 감정 분류

0 == 부정적, 1 == 긍정적

NA: 0

### 데이터 전처리

리뷰가 없는 문장 제거

한글과 공백 제외 단어 제거

공백만 있거나 빈 값이 된 리뷰 제거

테스트 데이터도 위와 같이 진행

불용어 정의

형태소 분석기를 사용해 토큰화 및 불용어 제거

테스트 데이터도 마찬가지로 진행

단어 집합 생성 및 단어 빈도수 확인

빈도가 낮은 단어(빈도수:2)를 단어 집합에서 제외 및 남은 단어들로 구성

텍스트 시퀀스를 숫자 시퀀스로 변환 및 패딩 적용

### 데이터 전처리

LSTM 모델 설계 및 조기 종료와 모델 체크포인트 생성

모델 컴파일(손실 함수, 옵티마이저, 평가 지표)

모델 훈련 및 정확도 측정

새로운 문장을 입력받아 감성 예측을 수행하는 함수 정의

예측 수행

Part 3

# 감성 분석 결과

### 감성 분석 결과

```
테스트 정확도

{
    print(paste("테스트 정확도: ", model_accuracy)
    "테스트 정확도: 0.8601"
}
```

```
감성 예측 수행
{
  "좀 별로네요"
  99.17% 확률로 부정 리뷰입니다.

  "재밌어요"
  99.21% 확률로 긍정 리뷰입니다.

  "나쁘진 않아요"
  57.29% 확률로 긍정 리뷰입니다.
}
```

Part 4

# 문제점 및 활용방안

### 문제점 및 활용방안

### 반어적 리뷰 확인

{

("이 영화를 보고 암이 나았습니다 감사합니다 ㅠㅠ") 97.87% 확률로 긍정 리뷰입니다.

("당신이 이 영화를 보지 않았다면 아직 살아있을 이유하나를 간직하고 있는 것이다.") 92.59% 확률로 긍정 리뷰입니다.

("모니터도 울고 외장하드도 울고 숨어있던 바이러스도 울었다")

96.79% 확률로 긍정 리뷰 입니다.

}

### 활용 방안

#### 1. 고객 서비스 개선

- 1. 고객 피드백 분석
- 2. 고객 만족도 추적

#### 2. 마케팅 전략 최적화

- 1. 캠페인 효과 분석
- 2. 타겟 마케팅

#### 3. 브랜드 관리

- 1. 위기 관리
- 2. 브랜드 평판 모니터링

# 감사합니다.