|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **2차 프로젝트 최종 보고서** | | | | | | | | |
| 프로젝트 명 | 국문 | | | 증권뉴스의 감성분석을 통한 주가 예측 프로젝트 | | | | |
| 영문 | | |  | | | | |
| 수행기간 | 2022 . 10 . 05 ~ 2022. 10. 18 | | | | | | | |
| 참여자 | 소속 | | | | 참여자성명 | 휴대전화 | 이메일 |
| 4조 | | | | 유청환 | 010-4158-1389 | katimere15@gmail.com |
| **프로젝트 요약** | |  | | | | | | |
|  | |  | | | | | | |
| 1. 개발 목표  * 뉴스의 감성 분석을 통한 주가 예측 * 제작 모델의 정확도 65%이상을 목표  1. 개발 내용 개요  * 주가에 영향을 미치는 데이터가 담긴 뉴스 정보를 분석해 다음날 주가를 예측한다.  1. 개발 과정    1. yahoo finance의 3년치 코스피 지수 데이터 수집    2. 팍스넷, 네이버의 증권 뉴스 데이터 수집    3. 전처리 진행    4. 뉴스 데이터의 토큰화 진행    5. 토큰 데이터 전처리 진행    6. TF-idf와 로지스틱 회귀를 이용해 머신러닝 모델 제작, 저장    7. 저장 모델을 불러와 결과물 출력 2. 최종 개발 결과물      1. 프로젝트 진행 중 애러사항  |  |  |  | | --- | --- | --- | | 순번 | 프로젝트 단계 | 발생 문제 | | 1 | 데이터 수집 | 문제 : 네이버의 뉴스데이터를 수집하는데 있어서 뉴스의 미리보기 이미지 유무에 따라 select가 달라지는 문제  해결 : 에러코드를 분석해 try except문을 사용해 문제 해결 | | 2 | 데이터 분석 | 문제 : 분석 결과가 모두 부정(“0”)이 나오는 문제  해결 : 코드 분석 결과 모델의 return값이 int타입인 것을 확인하였고 결과를 출력하는 if문을 수정해 해결함 | | | | | | | | | |
| **프로젝트 회고** | |  | | | | | | |
|  | | 날이 지나면 자동으로 분석해 보여주는 것이 아닌 유저가 직접 뉴스데이터를 가져와 코드를 실행해야 한다는 것이 매우 아쉬웠다. 다음 프로젝트에 이 분석 모델을 사용해 새로운 뉴스가 업로드 되면 자동으로 수집, 분석을 통해 결과를 저장하는 플랫폼을 제작하고 싶다.  뉴스를 업로드한 곳에 따라 그 데이터에 느끼는 감정이 다르다는 것을 깨달았으며 각 언론사별로 뉴스 데이터를 분류해 분석을 진행했으면 좋았을 것 같다 | | | | | | |
|  | | | | | | | | |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | 주요내용 | 추진 일정 | | | | | 5~7 | 11~12 | 13~15 | 17~18 | | 계획서 제작 및 설계 | ⊙ |  |  |  | | 데이터 수집 및 전처리(뉴스데이터) |  | ⊙ |  |  | | 데이터 분석(감성분석, 키워드 분석) |  |  | ⊙ |  | | 문서화 |  |  |  | ⊙ | | | | | | | | | |
| **결과활용계획** | | |  | | | | | |
|  | | | 1. 다음의 프로젝트에서 빅데이터 플랫폼을 구축해 자동으로 데이터를 수집, 적재, 분석을 진행하는 프로젝트를 진행할 것이다. 2. 팍스넷, 네이버 뿐만 아니라 다른 언론사의 데이터를 수집해 머신러닝 모델의 정확도를 올리도록 할 것이다. | | | | | |
|  | | | | | | | | |