|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호 001 |
| 실습 목적 | 주제 선정 | |
| 날짜 | 23.06.08 | |
| 팀원 | 김윤아 / 김희동 / 백두현 / 엄지우 | |
| 활동 내용 | 1. 팀명 확정 2. 주제 논의 및 데이터 수집 방법  * 데이터 수집 방법론 : 텍스트 마이닝 * 주제 논의중… | |
| 활동 결과 | - | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 김윤아 | 점검자 김희동 | 점검자 백두현 | 점검자 엄지우 |
| 일자 2023년 06월 08일 | 일자 2023년 06월 08일 | 일자 2023년 06월 08일 | 일자 2023년 06월 08일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호002 |
| 실습 목적 | 주제 선정 | |
| 날짜 | 23.06.09 (금) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | 1. 주제 선정 2. 데이터 수집처 3. 타겟 선정 4. 피처 선정 | |
| 활동 결과 | 1. 주제 방향성 수정 2. 개괄적인 분류 기준 결정 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 09일 | 일자 2023년 06월 09일 | 일자 2023년 06월 09일 | 일자 2023년 06월 09일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호003 |
| 실습 목적 | 기획발표 | |
| 날짜 | 23.06.12 (월) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | * + - 1. 주제 재탐색       2. 기획안 작성       3. 기획 발표 | |
| 활동 결과 | 주제에 대한 피드백  방법 수정 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 12일 | 일자 2023년 06월 12일 | 일자 2023년 06월 12일 | 일자 2023년 06월 12일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호004 |
| 실습 목적 | 데이터 수집 방안 탐색 | |
| 날짜 | 23.06.13 (화) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | 1. 주가 예측 수행할 기업 선정 2. 데이터 수집 방안 탐색  * 뉴스 기사 * 금, 유가 데이터  1. 뉴스 기사 활용방안 논의 2. 자연어 처리, 예측 모델에 활용할 알고리즘 학습 분배  * Kobert * Kr – Fin BERT * LSTM * ResNET | |
| 활동 결과 | 1. 기업 선정 – 대한항공 2. 뉴스 기사 데이터 수집(BigKinds 뉴스 기사 다운로드) 3. 금, 유가 데이터 수집([Investing.com](https://kr.investing.com/currencies/usd-krw-historical-data)) | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 13일 | 일자 2023년 06월 13일 | 일자 2023년 06월 13일 | 일자 2023년 06월 13일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호005 |
| 실습 목적 | 데이터 수집 및 전처리 | |
| 날짜 | 23.06.14 (수) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | 1. 자연어 처리, 예측 모델에 활용할 알고리즘 학습 공유 2. 뉴스 데이터 전처리  * 대한항공 키워드가 들어가지 않은 뉴스 기사 제외 * 각 뉴스별 핵심 키워드 단어 추출  1. Bekins 사이트에서 뉴스 본문 크롤링 | |
| 활동 결과 | 1. 모델 학습 공부 공유 2. 뉴스 제목 데이터 전처리 완료 3. 사이트 크롤링 코드 생성 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 14일 | 일자 2023년 06월 14일 | 일자 2023년 06월 14일 | 일자 2023년 06월 14일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호006 |
| 실습 목적 | 데이터 전처리, 뉴스 기사 본문 크롤링 코드 작성, 감성 분석 방법 논의 | |
| 날짜 | 23.06.15 (목) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | 1. 뉴스 데이터 전처리  - ‘대한항공’ 을 포함하지 않는 기사 삭제  - 일자별 언급 횟수 1위 포함하지 않는 기사 삭제  2. 언론사별 크롤링 코드  3. 감성 분석 방법론 논의 | |
| 활동 결과 | 1. 뉴스 기사 선별 완료  2. 언론사별 크롤링 코드 작성 완료  2. 감성 분석 방법론에 대한 논문 공유 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 15일 | 일자 2023년 06월 15일 | 일자 2023년 06월 15일 | 일자 2023년 06월 15일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호007 |
| 실습 목적 | 뉴스 제목 데이터 토큰화 | |
| 날짜 | 23.06.16(금) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | 1. 뉴스 데이터 전처리  * 제목 데이터 명사 뽑아 토큰화 * 뉴스 날짜에 맞게 대한항공 종가로 상승 하락 target변수 생성 * 뉴스별 대한항공 뉴스 내용 크롤링 | |
| 활동 결과 | 1. 뉴스 제목 ‘명사’로 나누어 토큰화 2. 뉴스 날짜에 맞춰 주가 데이터 합치기 3. 뉴스 내용 크롤링 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 15일 | 일자 2023년 06월 15일 | 일자 2023년 06월 15일 | 일자 2023년 06월 15일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호008 |
| 실습 목적 | 뉴스 제목 데이터 토큰화 | |
| 날짜 | 23.06.19(월) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | 1. 결측치대체 2. 뉴스 본문 크롤링 | |
| 활동 결과 | 1. 주말 결측치(종가, 금종가, 유가종가 결측치) 다음 월요일 값으로 대체 2. 뉴스 본문 데이터 크롤링 자동화 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 19일 | 일자 2023년 06월 19일 | 일자 2023년 06월 19일 | 일자 2023년 06월 19일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호009 |
| 실습 목적 | 뉴스 제목 데이터 토큰화 | |
| 날짜 | 23.06.20(화) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | 1. 중간발표 ppt만들기 2. 본문 데이터 전처리하기 | |
| 활동 결과 |  | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 19일 | 일자 2023년 06월 19일 | 일자 2023년 06월 19일 | 일자 2023년 06월 19일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호010 |
| 실습 목적 | 중간발표 후 피드백 | |
| 날짜 | 23.06.21(수) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | 뉴스는 초단기 주가에 영향 미침  제목 고쳐 ; 기업->대한항공  여러 거시경제 지표 중 금, 유가가 가장 영향력을 미쳤다는 것을 수치적으로 제시해야 올바른 논리 ; 거시경제 지표 여러개 후보로 놓고 그중에서 추리는 것도 좋을듯  뉴스 뽑아온 카테고리(우리는 경제 섹터) 명시해야  망하지 않은 기업의 경우 뉴스기사는 대부분 긍정적 기사  기사출력시간을 통해 기사가 주가에 미치는 영향을 좀 더 세분화  ex) 3시30분 이전 발행된 기사는 당일날 종가에 영향  3시30분 이후 발행된 기사는 부득이하게 다음날 종가에 영향  기업 선정 이유도 확실히 | |
| 활동 결과 |  | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 19일 | 일자 2023년 06월 19일 | 일자 2023년 06월 19일 | 일자 2023년 06월 19일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호011 |
| 실습 목적 | 데이터 처리 다시시작 | |
| 날짜 | 23.06.22(목) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | 1. 주제 명확히 다시 정하기 2. 기업 2 하나 더 선정하기 3. 거시 지표 x => 거시지표 뉴스 데이터 | |
| 활동 결과 | 1. 기업: 대한항공, LG화학 2. 유가 -> 유가 뉴스 데이터로 대체 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 19일 | 일자 2023년 06월 19일 | 일자 2023년 06월 19일 | 일자 2023년 06월 19일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호012 |
| 실습 목적 | 데이터 처리 | |
| 날짜 | 23.06.22(목) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | 1. 대한항공 데이터 뉴스 시간 가져오기 2. LG화학 데이터 크롤링 및 전처리 3. 유가 뉴스 데이터 크롤링 및 전처리 | |
| 활동 결과 | 1. 총 데이터 병합 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 19일 | 일자 2023년 06월 19일 | 일자 2023년 06월 19일 | 일자 2023년 06월 19일 |

|  |  |  |
| --- | --- | --- |
| 실습 제목 | 텍마하트 | 일련번호013 |
| 실습 목적 | 데이터 전처리 | |
| 날짜 | 23.06.26(월) | |
| 팀원 | 백두현 / 김희동 / 엄지우 / 김윤아 | |
| 활동 내용 | 1. 대한항공 데이터 뉴스 시간 결측치 수동으로 가져오기 2. LG화학 데이터 뉴스 시간 결측치 수동으로 가져오기 3. DL 지수(대한항공 + LG화학) 산출 | |
| 활동 결과 | 1. 지수 산출 완료 | |

|  |  |  |  |
| --- | --- | --- | --- |
| 기록자 백두현 | 점검자 김희동 | 점검자 김윤아 | 점검자 엄지우 |
| 일자 2023년 06월 26일 | 일자 2023년 06월 26일 | 일자 2023년 06월 26일 | 일자 2023년 06월 26일 |