**[ 9월 1주차 프로젝트 수행 일지 ]**

| **프로젝트 타이틀** | **뉴스 매체와 유튜버의 주식가격 등락 예측 정확도 분석** | | |
| --- | --- | --- | --- |
| **프로젝트 팀명** | 업빛투 | 프로젝트 팀원 | (팀장) (정) 정길종, (부) 김형림  (팀원) (사이언스) 윤보람, 채길호  (팀원) (엔지니어링) 인태우 |

※ 프로젝트 수행 일지는 매 주 금요일 23:59 전까지 최종본을 저장해주세요

**1. 프로젝트 수행 계획 및 현황**

| **1. 프로젝트 수행 계획 및 현황** | | |
| --- | --- | --- |
| **이름** | **금주 프로젝트 수행 내용** | **차주 프로젝트 수행 계획** |
| **데이터**  **사이언스** | **□ LSTM**  **- FinanceDataReader로 주식 데이터 로드 후, 테스트 수행**  **□ KoNLPy**  **- 수집한 2만 여 개 뉴스 데이터 분류 작업**  **□ 감성 어휘 사전**  **- 수집한 종목 뉴스 데이터로 테스트 수행**  **- VADER 감성사전이 종목 뉴스 기사를 대부분 긍정으로 분류하는 문제점 발견**  **- 주식가격 변동으로 자체 감성 어휘 사전을 구축 후 감성점수 확인** | **□ EDA**  **- 데이터 이상치 제거**  **- 데이터 속성 관계 분석**  **□ Py-Hanspell**  **- 네이버 한글 맞춤법 검사기 전처리 패키지(YouTube 스크립트)**  **□ 감성 어휘 사전**  **- 주식가격 분석에 적합한 감성 사전 서칭** |
| **데이터**  **엔지니어링** | **□ 뉴스 크롤링**  **- 4개 종목 2020~2021년 뉴스 데이터 수집(매일경제)**  **- 타 언론사 크롤링에 문제(머니투데이)**  **- YouTube 스크립트 크롤링 실습 및 결과 확인**  **□ AWS 내 작업환경 구축**  **- 필요 라이브러리 설치**  **□ MySQL DB 구축**  **- 연결 환경** | **□ 뉴스 크롤링**  **- 필요 종목에 대한 2018~2021년 뉴스 데이터 수집**  **- 타 언론사(머니투데이) 새 크롤링 코드 작성 및 데이터 수집**  **□ YouTube 스크립트 크롤링**  **- 주식 채널 크롤링 데이터 수집**  **□ 데이터 전처리**  **- 데이터 전처리**  **- 데이터 정규화** |

**2. 강사님 피드백**

| **2. 강사님 피드백** | |
| --- | --- |
| **데이터사이언스**  **이경미 강사님** | - 텍스트 데이터 활용한 주가 예측 참고 문헌 추가 조사  - 주식과 관련된 감성사전 조사  - 서비스 타겟과 내용에 대한 고민 필요 |
| **데이터엔지니어링**  **김성환 강사님** | - 뉴스 데이터 수집 기간을 객관적 기준 정하기  - 크롤링 데이터 전처리 후 DB 적재 |

**[ 9월 2주차 프로젝트 수행 일지 ]**

| **프로젝트 타이틀** | **뉴스 매체와 유튜버의 주식가격 등락 예측 정확도 분석** | | |
| --- | --- | --- | --- |
| **프로젝트 팀명** | 업빛투 | 프로젝트 팀원 | (팀장) (정) 정길종, (부) 김형림  (팀원) (사이언스) 윤보람, 채길호  (팀원) (엔지니어링) 인태우 |

※ 프로젝트 수행 일지는 매 주 금요일 23:59 전까지 최종본을 저장해주세요

**1. 프로젝트 수행 계획 및 현황**

| **1. 프로젝트 수행 계획 및 현황** | | |
| --- | --- | --- |
| **이름** | **금주 프로젝트 수행 내용** | **차주 프로젝트 수행 계획** |
| **데이터**  **사이언스** | **□ 감성 어휘 사전**  **- 재무분석에 사용하는 KOSELF 감성사전 확보**  **- KOSELF 감성사전은 애널리스트 리포트로부터 만들어졌기 때문에, 뉴스 기사의 감성 분석에 완벽하게 적합하지는 않다는 문제점 발견 → KOSELF 감성사전 보완 방안 논의**  **□ TF-IDF**  **- 주가 변동률을 Label로 하여 텍스트를 TF-IDF 벡터화 후 Logistic Regression을 적용한 결과 예측 정확도는 50% 정도**  **□ LSTM**  **- 주식 데이터 추출 라이브러리를 FinanceDataReader로 통일**  **- 재무정보로 재무지표를 계산하여 주식데이터에 추가**  **- LSTM 수행 결과 MSE 확인**  **□ Tableau EDA**  **- 주가 데이터 시각화** | **□ 감성 어휘 사전**  **- 뉴스 기사를 통해 자체 제작한 감성사전을 KOSELF에 추가한 후 감성 분석 진행**  **□ TF-IDF**  **- 모델의 예측 정확도 향상을 위한 하이퍼파라미터 튜닝**  **□ 유튜브 데이터 분석**  **- 벡터화된 단어 간 유사도 확인**  **□ 주식 데이터를 통한 종가(Close) 예측**  **- 머신러닝 기법과 강화학습을 통한 종가 예측**  **- 등락 정확도 파악**  **□ Tableau 시각화**  **- 분석 결과 시각화** |
| **데이터**  **엔지니어링** | **□ 크롤링**  **- 4개 종목 2018~2021년 뉴스 데이터 수집 완료(매일경제)**  **- 4개 종목 2018~2021년 뉴스 데이터 수집 완료(아시아경제)**  **- YouTube 3개 채널 스크립트 크롤링 데이터 수집 완료**  **□ AWS 내 Django**  **- 서버 구축 완료**  **- 태블로 시각화 자료 Django 연결 테스트 완료**  **□ 데이터 수집 스케줄링**  **- 데이터 DB에 적재되는 것까지 확인** | **□ Django**  **- Django-MySQL 연동**  **- 태블로 연동**  **- DB 데이터 장고 대시보드 연결**  **□ 유튜브**  **- 유튜브 전체에서 4개 종목 일별 데이터 수집**  **□ 데이터 수집 스케줄링**  **- 주식, 뉴스, 유튜브 데이터 수집 스케줄링** |

**2. 강사님 피드백**

| **2. 강사님 피드백** | |
| --- | --- |
| **데이터사이언스**  **이경미 강사님** | [9/8 수]  - 서비스 구체화  - 범용적인 감성사전 제작 (종목별이 아닌)  - KOSELF 감성사전과 자체 제작 감성사전 단어들 간의 유사도 확인  - 유튜브 스크립트 Raw Data 그대로 분석 (요약 전처리 X)  - 시계열 다양한 분석 비교해볼 것 (ARIMA, Prophet, LSTM, RL)  [9/10 금]  - 수집 데이터 기간 설정의 객관적인 이유   (현재 삼성전자 액면 분할 이후 시점인 2018년 1월 1일부터 데이터 수집)  - 감성 사전을 위한 종목 추가 필요  - PPT 데이터 구축 및 모델 요약 장표 ***Good-!!*** |
| **데이터엔지니어링**  **김성환 강사님** | [9/10 금]  - 등락 정확도 파악 |

**[ 12월 2주차 프로젝트 수행 일지 ]**

| **프로젝트 타이틀** | 웹 반응 데이터 감성분석 | | |
| --- | --- | --- | --- |
| **프로젝트 팀명** | 16조\_살아남아조 | **프로젝트 팀원** | (팀장) 민채원  (팀원) 이예슬 |

※ 프로젝트 수행 일지는 매 주 토요일 23:59 전까지 최종본을 저장해주세요

**1. 프로젝트 수행 계획 및 현황**

| **과정명 및 이름** | **금주 프로젝트 수행 내용** | **차주 프로젝트 수행 계획** |
| --- | --- | --- |
| **(빅데이터)**  **홍길동**  **홍길순** | - 월별 기초 통계량 그룹핑 및 시각화  - Bert, NB 등 모델 정확도 비교 분석 | - 월별 분석 통한 데이터 수치화 및   긍/부정 비교 분석  - 모델 정확도 개선에 필요한   word2vector 만들기 |
| **(AI)**  **홍길동** | -  - | -  - |
| **(IoT)**  **홍길동** | -  - | -  - |
| **(클라우드)**  **홍길동** | -  - | -  - |

**2. 강사님 피드백**

| **빅데이터**  **김정현 강사님** | 완성된 긍정어 부정어 사전을 이용하여 긍정문장과 부정문장을 예측하여 정확도를 확인해보고  문제점과 보완점을 점검해주세요 |
| --- | --- |
| **AI**  **문성훈 강사님** |  |
| **IoT**  **김구수 강사님** |  |
| **클라우드**  **박창렴 강사님** |  |

**[ 12월 3주차 프로젝트 수행 일지 ]**

| **프로젝트 타이틀** | 웹 반응 데이터 감성분석 | | |
| --- | --- | --- | --- |
| **프로젝트 팀명** | 16조\_살아남아조 | **프로젝트 팀원** | (팀장) 민채원  (팀원) 이예슬 |

※ 프로젝트 수행 일지는 매 주 토요일 23:59 전까지 최종본을 저장해주세요

**1. 프로젝트 수행 계획 및 현황**

| **과정명 및 이름** | **금주 프로젝트 수행 내용** | **차주 프로젝트 수행 계획** |
| --- | --- | --- |
| **(빅데이터)**  **홍길동**  **홍길순** | - 월별 기초 통계량 그룹핑 및 시각화  - Bert, NB 등 모델 정확도 비교 분석 | - 월별 분석 통한 데이터 수치화 및   긍/부정 비교 분석  - 모델 정확도 개선에 필요한   word2vector 만들기 |
| **(AI)**  **홍길동** | -  - | -  - |
| **(IoT)**  **홍길동** | -  - | -  - |
| **(클라우드)**  **홍길동** | -  - | -  - |

**2. 강사님 피드백**

| **빅데이터**  **김정현 강사님** | 완성된 긍정어 부정어 사전을 이용하여 긍정문장과 부정문장을 예측하여 정확도를 확인해보고  문제점과 보완점을 점검해주세요 |
| --- | --- |
| **AI**  **문성훈 강사님** |  |
| **IoT**  **김구수 강사님** |  |
| **클라우드**  **박창렴 강사님** |  |