**프로젝트 기획안**

**2021년 09월 15일**

**과정명 : 데이터 사이언스/엔지니어링 전문가**

|  |  |
| --- | --- |
| 팀 명 | BeautifulAnts |
| 팀 원 | 팀장: 김민창  팀원: 강민정, 방병제, 손동기, 최윤호 |
| 프로젝트 주제 | 주가 예측 및 주식 기사 검증을 위한  주가 및 주식 기사 데이터 구축 |
| 프로젝트 수행 방향 | 1. 수집시스템 구축    * 산업별 주가 /ETF 주가/SPY 주가 수집 2. DB 스키마 설계  * 날짜, 산업별 주가, ETF 주가, SPY 주가 * 날짜, 주식기사 본문, 주식기사별 부정 단어 점수, 긍정 단어 점수, 총 점수  1. 데이터 현황 시각화 (2019.08.01-2021.08.31)  * 주식의 수익률(%) – SPY / ETF * 주식기사점수+SPY / 주식기사점수+ETF * 상관관계 분석 그래프 – 단기 / 장기 투자  1. 데이터 분석 예상  * 자연어 처리 (RNN), 주가 데이터 예측, 상관관계 분석 |
| 역할 분담 및 일정 | 1. 역할분담  * 김민창 : 데이터 전처리, 데이터 분산처리, 시각화 * 강민정 : 데이터 수집, ML, 상관관계 분석, DB구축 * 방병제 : 데이터 수집, 상관관계 분석, DB 구축 * 손동기 : 데이터 수집, AWS서버 환경 구축(하둡,스파크 설치), DB 구축 * 최윤호 : 데이터 전처리, 데이터 분산처리, 시각화  1. 일정  * 09/08 ~ 09/09 : 주제 선정 및 일정 수립 * 09/10 ~ 09/12 : 데이터 수집 * 09/13 ~ 09/17 : 데이터 전처리 및 파이프라인 구축 * 09/18 ~ 09/24 : DB 구축 및 시각화 * 09/25 ~ 09/27 : 상관관계 분석 및 ML * 09/28 ~ 09/29 : 발표자료 제작 * 09/30 : 프로젝트 발표 |
| 프로젝트 수행 도구 | ● 수집 기술 패키지  requests, beautifulsoup4  ● 빅데이터 수집, 저장, 분석 시스템  Apache Hadoop, Apache Spark  ● 데이터베이스  MySQL  ● 시각화 기술 패키지  Matplotlib, Seaborn |