**프로젝트 기획안**

**2020년 10월 13일**

**딥러닝 기반 AI 엔지니어링**

|  |  |
| --- | --- |
| 팀 명 | 뽑아조 |
| 팀 원 | **김진원** , 이동재 , 이지윤 , 박지원 |
| 프로젝트 명 | 미국 대선 관련 뉴스 기사 및 트위터 텍스트 마이닝을 통한 감성 분석 및 예측 |
| 프로젝트 주제  및 내용  해결하고자 하는 문제  최종 산출물의 청사진 | 2020년 11월 3일은, 역대 59번째 대통령 선거날이다. 현재 코로나 사태로 인하여 세계가 많이 혼란스러운 상태에 있으며, 미국 경찰의 인종 차별적인 채포 과정과 트럼프의 북한 방문 등 평화를 위해 노력하는 모습 등, 많은 사건들이 있었다.  그에 따라 우리는 이번 대통령 선거에서 트럼프 대통령과 바이든 부통령의 sns 또는 언론사 글에 대해 자연어 처리를 통한 감성분석 및 데이터 분석을 실시하여, 당선될 가능성이 누가 더 높은 지 예측해보려 한다.  텍스트이(가) 표시된 사진  자동 생성된 설명 |
| 주제 선정 배경  주제 선정의 정당성 산출될 결과의 유용성 | 최근 치러진 미국 대선에서는 머신러닝, 딥러닝 기술이 선거 예측에 활발히 사용되고 있다. 지난 2016년 미국 대선 결과를 맞힌 것은 여론조사가 아니라 구글 검색량이라는 빅데이터였다. 당시 도널드 트럼프 공화당 후보는 여론조사에서 힐러리 클린턴 민주당 후보에게 열세를 면치 못하는 가운데, 대다수는 클린턴 후보의 승리를 점치던 상황이었다. 하지만 빅데이터 결과는 트럼프 후보의 승리를 이미 예견하고 있었던 것이 선거 후 밝혀졌다.  이와 같이 대선을 예측함에 있어 빅데이터의 중요성이 다시 한번 주목받는 시점에서 이번 대선 또한 빅데이터를 활용한 분석 및 예측 과정을 진행하려 한다. |
| 프로젝트 수행 방향  팀원간 역할 분담  프로젝트 수행 일정  수행 방법/도구 | **팀원간 역할 분담**   |  |  | | --- | --- | | 김진원 | 트위터 크롤링, EDA, Modeling, Visualization | | 박지원 | 민주당 지지 뉴스 크롤링, EDA, Modeling, Visualization | | 이동재 | 공화당 지지 뉴스 크롤링, EDA, Modeling, Visualization | | 이지윤 | 중도 뉴스 크롤링, EDA, Modeling, Visualization |   **프로젝트 수행 일정**   |  |  | | --- | --- | | 10/29 ~ 10/30 | 주제 선정 | | 10/31 ~ 11/1 | 주제 구체화 | | 11/2 ~ 11/3/ | 데이터 수집 | | 11/4 ~ 11/6 | EDA , 데이터 분석(+감성사전) | | 11/7 ~ 11/9 | 시각화 및 모델링 | | 11/10~11/11 | 발표 자료 작성 |   **수행 도구**   * Jupyter notebook * Colab * Python * NLTK * vaderSentiment (Pos / Neg dictionary) |