**파이널 프로젝트 기획안**

**2020년 3월 28일**

**서비스 산업 데이터를 활용한 머신러닝 분석 (B반)**

|  |  |
| --- | --- |
| 팀 명 | 노구리 |
| 팀 원 | **구민주, 노소영, 이리애, \*이혁재(PM)  *\* 팀장 별도 표시*** |
| 프로젝트 명  *\* 가제* | 머신러닝을 이용한 21대 국회의원 총선거 예측 |
| 담당 멘토 | 박인근, 홍지후, 윤석용 멘토 |
| 프로젝트 주제/내용   * 기획 의도 * 프로젝트 목표 * 주요 기능 | 기존 대선과 총선, 각종 선거와 관련하여 이를 다양한 센서스와 새로운 방법을 접목시켜 효과적인 여론조사결과를 내려는 시도는 이미 다양하게 이루어졌다고 생각합니다. 머신 러닝 분야 역시 예외는 아닙니다.  이와 더불어 기존 모델링이 선거 결과와 대조하였을 때 필연적으로 갖게 되는 오차와 모델이 갖는 치명적인 허점을 보완하고 향후 유사한 시도에서 결정적인 지표를 제공할 모델을 설정하는 것이 저희의 이번 프로젝트 목표입니다.  프로젝트로 최종 구현될 기능은 우선  1) 서울시 선거구별 우위후보의 예측,  2) 선거 결과예측에 대한 시각화  3) 특정 후보자에 대한 요약정보 표시  로 정리되겠습니다. |
| 기대효과   * 산출 결과의 유용성 | 정치는 우리의 삶과 밀접한 연관이 있습니다.  국회의원 총선거는 대선과 더불어 가장 큰 선거행사의 일환입니다.  그 결과에 대해서 사전에 정확히 파악할 수 있다면 직간접적으로 연관된 사람들의 다양한 대응에 시사하는 바가 큽니다.  머신러닝 모델링을 통해 더 정확한 선거예측이 가능하다면   1. 선거패턴을 체계적으로 구조화하고 객관적이고 과학적인 선거 연구 유도 2. 대규모 여론 조사 및 센서스에 대한 사회적 비용 절감 3. 효율적인 선거운동과 더불어 유권자에게 좀 더 객관적인 후보자를 선택할 수 있는 정보를 제공. |
| 프로젝트 수행 방향   * 팀원간 역할 분담 * 프로젝트 수행 일정 * 수행 방법/도구 및 * 개발 환경 | * 팀원간 역할 분담   데이터수집/ 전처리 : 구민주, 이리애  모델링 학습 및 분석 : 구민주 노소영 이리애 이혁재  데이터 시각화 : 노소영, 이혁재  스캐줄 관리 및 조율: 이혁재   * 프로젝트 수행 일정   3/28~4/14 데이터 수집 및 전처리  4/2~4/14 모델링 학습 및 분석, 모델 수정  4/14~4./30 결과물 시각화 및 보완/수정   * 수행 방법/도구 및 개발환경   가장 접근성이 좋고 최근 빈번하게 정치적 소통의 창구로 활용되는 Youtube 댓글란을 crawling하여 이를 형태소 분석을 통한 긍/부정어 사전 분석을 데이터 수집 원안으로 설정하였습니다. 또한 지역구별 후보자의 검색어 유입 트래픽 수, 관련 기사 등의 자료역시 대상으로 설정할 예정입니다.  이를 python 및 환경에서 NLP분석이 가능한 Keras라이브러리 참고 및 AWS, google Colab등의 IDE 환경을 사용하여 모델링을 설정하고, 해당 코드를 Tomcat을 이용한 가상 웹 환경으로 시각화를 구현 및 서비스를 할 예정입니다. |
|  |