

엔코아 프로젝트 수행계획서

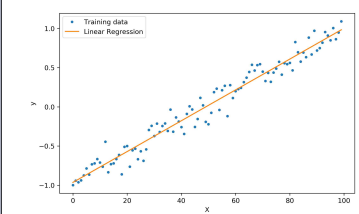

I. 팀 정보

팀 명	동수야 일어나조 (임시)	
팀 원	이 름	소 속
멘 토	정좌연	LG CNS
멘티	팀장	조상혁
	팀원1	김용진
	팀원2	문주혁
	팀원3	조동수

팀 사진



II. 프로젝트 정보

프로젝트명	Daily 민심 체크	
기술분야	■ 디지털 콘텐츠 ■ DB ■ 머신러닝	
주제영역	■ 생활	
성과목표	■ 엔코아 프로젝트 ■ 공모전	
수행예상기간	2020. 9. 1. ~ 2020. 12. 04.	
프로젝트소개 및 제안배경	최근 포털 사이트별 정치 성향이 나뉘는 경향을 띄고 있어, 정부 지지도, 특정 정책 및 정치사건에 대해서 전반적인 여론을 확인하기 어려워짐. 이에 특정 정치성향에 편향되지 않은 대중들의 의견과 찬/반 정도를 확인할 수 있는 사이트의 필요성을 느낌.	
주요기능	키워드 추출	· 여러 포털의 정치 기사 댓글을 수집/정제/저장/클린징 하여 키워드 수집
	감성/문서분석	· 감성사전을 직접 작성 후, 수집한 키워드와 비교/대조하여 감정 점수 산출
	관계 분석	· 과거 댓글의 감정 점수와 정부 지지도의 관계를 밝힘으로써, 현재의 사건/정책에 대한 여론을 예측함.
	시각화	· 시각화 도구를 활용한 차트 조회 및 분석
	서버구축	· 서버를 구축하고 모델링 결과 및 시각화 차트를 웹 서비스화
적용기술	데이터 수집	· 공개 데이터 수집, 데이터 크롤링 · 활용 : Python
	데이터 정제/적제	· 사전 전처리, 정제를 통해서 필요한 데이터 적제 · 활용 : (Elastic Search or Mysql + AWS or Hadoop)
	데이터 분석	· 시계열 분석, 회귀 분석, 머신 러닝 분석을 통한 유의미 정보 추출 · 활용 : R, Python
	데이터 시각화	· 오픈소스를 활용한 데이터 시각화 개발 · 활용 :
	서버 구축	· 활용 : (Kibana or AWS)
예상결과물		
	-댓글분석에 적합한 머신러닝 모델 생성-	-웹 서버에 분석 결과 시각화 -

기대효과 및 활용분야	빅데이터 활용 능력 제고	- 데이터 파이프라인 구축 직접 수행해 봄으로써 빅데이터 역량을 키움
	정치적 객관성 유지	- 자주 방문하는 포털의 정치성향에 매몰되지 않고, 전반적인 여론 확인함으로써 정치에 대한 자기 판단력을 기름.
	공모전 참여	- 분석결과에 대한 논문, 공모전 참여
필요기자재	(AWS 서버비용)	

III. 프로젝트 수행방법

가. 멘티(참여학생) 업무분장

번호	이름	역할	담당업무
1	조상혁	팀장	
2	김용진	팀원	
3	문주혁	팀원	
4	조동수	팀원	

나. 프로젝트 추진일정

구분	추진내용	추진일정			
		9월	10월	11월	12월
계획	아이디어 및 컨셉 결정, 학습 대상 데이터 수집				
분석	데이터 전처리 작업 수행				
	확보한 데이터 이상치, 결손 여부 점검				
	확보한 데이터 특징 분석 후 프로젝트 목표에 부합하는지 점검				
설계	특성 공학 수행을 통해 머신러닝 알고리즘에 사용할 특징 추출				
	머신러닝 알고리즘 선정 및 상세 파라미터 및 옵션 결정				
개발 / 테스트	효율적인 데이터 적재 프로세스 고안 및 개발(배치 or 온라인)				
	머신러닝 학습 수행 및 주기적으로 결과 로깅 및 분석				
	학습 결과 시각화 프로세스 고안 및 개발				
	학습 완료된 모델의 정확도 검증				
서비스 개발	학습 모델을 이용하여 심화성과 차별성을 갖춘 웹서비스 개발				
공모전 참가	개발된 웹서비스와 개발 과정을 기록한 문서를 통해 공모전 참가				
종료	프로젝트 전과정을 다시 살펴보고 반성 및 개선점 모색				
오프라인 미팅계획	매주 금요일 아이디어 회의 진행				

다. 의사소통방법

- 주 1회 이상 주기적으로 오프라인 미팅을 통해 아이디어 공유 및 개발 현황 점검
- 오프라인 미팅이 불가할 시 줌을 통한 온라인 미팅으로 대체
- 카카오톡을 통한 온라인 피드백은 상시 진행

라. 프로젝트 Ground Rule (기본원칙)

- 미팅 시간은 철저히 준수한다.

IV. 기대효과 및 활용분야

2. 기대효과

- 엔코아 빅데이터 교육에서 배운 내용들을 활용하여 빅데이터 직무에 대한 전반적인 이해도를 높인다.
- 데이터 수집(크롤링), 데이터 전처리, 시각화로 이어지는 데이터 파이프라인을 직접 구축을 통해 추후 빅데이터 전문가로써 실무에서 활용될 스킬들을 익힘
- 여러 머신러닝 모델들을 테스트 해봄으로써 데이터 분석 과정을 익힘
- 프로젝트 기획부터 개발까지 전단계를 직접 수행함으로써 실무에서 활용될 경험을 쌓음

3. 활용분야

- 댓글 감정 분석은 정치 여론분석뿐만 아니라, 상품에 대한 소비자의 평가 분석을 통한 마케팅, 미디어 콘텐츠에 대한 반응 분석을 통한 추천 알고리즘 등 활용할 분야가 매우 광범위하다고 판단됨.