

금융서비스·마케팅 분야 분석 종합 프로젝트 기획안

2022년 1 월 3 일

과정명: 서비스 산업 데이터를 활용한 빅데이터 분석

프로젝트 주제	악성댓글 감정분석
팀 명	Writer's Warden
팀 원	김나래(조장), 김성훈, 김은지, 이동근, 장유림
프로젝트 내용 해결하고자 하는 문제 최종 산출물의 청사진	<p>악성 댓글이 많지만 효율적으로 차단하는 방법이 마땅치 않고 오인제재도 많다. 감정에 의해 욕해서 악성댓글을 남기거나 반복적으로 악성댓글을 제지하는 이도 별도로 관리할 필요가 있다고 생각했다.</p> <p>악플을 종류별로 분석해 점수를 매겨서 댓글에 해당되는 점수를 판별할 수 있도록 한다.</p> <p>특정 점수 이상의 악플을 쓸 경우 해당되는 라벨과 연관된 법안을 노출시켜 댓글을 쓰기 전 확인을 누를 수 있도록 해 악성댓글을 우발적으로 작성하지 못하게 만든다.</p> <p>특정 점수 이상의 악플을 누적해서 작성할 경우 악성 댓글 작성자에게 점수가 누적되어 댓글이 하단에 노출되거나 닉네임 옆에 특별 마크가 붙는 등의 제재도 가할 수 있다.</p>
팀원간 역할 분담 및 일정	<p>[역할]</p> <p>김나래 - ☆팀장 : 자료 수집 / 모델 학습 / 자료 정리</p> <p>김성훈 - 모델 학습 및 평가</p> <p>김은지 - 데이터 수집 및 전처리</p> <p>이동근 - 자연어 처리 모델 생성 및 자연어 전처리</p> <p>장유림 - 데이터 수집 및 전처리</p> <p>[일정]</p> <p>12-27 ~ 12-31 : 1차 주제 선정</p> <p>12-31 : 멘토링</p> <p>01-01 ~ 01-03 : 2차 주제 선정</p> <p>01-14 : 멘토링</p>
프로젝트 수행 방향 수행 방법/도구	<p>사용 툴: 주피터 노트북, MySQL Workbench, Google Colab</p> <p>사용 언어: Python, MySQL, HTML, JavaScript</p> <p>사용 패키지: Tensorflow, BeautifulSoup, Komoran, Selenium</p>

	<p>사용 데이터 출처 : AI허브</p> <p>사용 모델 : CNN</p>
<p>프로젝트 범위 (상세 구현 기능)</p>	<p>데이터의 수집 범위는 소셜, 기사, 커뮤니티 등 다양한 주제의 댓글로 선정한다.</p> <p>댓글의 악플 여부를 판단하고 악플의 공격성 점수를 계산한다.</p> <p>입력한 글이 악플일 경우 글의 공격성에 따라 점수를 판별하고 어떤 종류의 악플인지 분류하는 시스템을 구현한다.</p>
<p>프로젝트 관리 방안</p>	<ol style="list-style-type: none"> 1. 모든 데이터의 전처리, 분석 과정은 ipynb 파일로 정리하도록 한다. 2. 모델 파일의 공유는 Google Drive를 통해 이루어진다. 3. 모델 형성 과정은 아마존 웹 서버의 주피터 노트북을 활용한다.
<p>상세 일정</p>	<ol style="list-style-type: none"> 1. AWS에 개발환경 구축 (Jupyter Notebook, tensorflow) 2. AIHUB에서 다운받은 데이터를 학습시키기 3. 새롭게 웹 크롤링 한 데이터를 적용하여 라벨링하기 4. 라벨링한 결과를 바탕으로 악플의 정도 타당성 보완하기 5. 악플로 감지된 댓글에 대한 솔루션 만들기 6. 실제 가상 웹 환경에서 솔루션 적용해보기
<p>기대 효과</p>	<ol style="list-style-type: none"> 1. 악플을 종류별로 분석하여 어떤 종류의 악플인지 알아낼 수 있다. 2. 어떤 종류의 악플인지 파악하여 그에 대응하는 해결책을 제시할 수 있다. 3. 점수제의 도입을 통해 악플을 많이 쓴 사람의 댓글을 제한하고 하위노출을 시킴으로 반복적인 악플작성을 예방할 수 있다. 4. 악성 댓글을 입력했을 때, 댓글의 점수와 연관 법령을 노출하여 작성자가 한번더 생각할 수 있도록 함으로써, 악성 댓글을 절감할 수 있을 것으로 예상된다.