

「2022년 충청북도 공공데이터 활용 창업경진대회」 제품 및 서비스 개발 사업계획서

참가자(팀)명	우리집은 남동향
제품(서비스)명	DH_SERVICE (감성분석 기반 동향분석 서비스)

1 제품 및 서비스 개요

1-1. 제품 및 서비스의 기능/특징

구분	기능/특징	적용 기술	UI
1	키워드 검색, 날짜 선택 기능	- Flask - HTML - CSS	
2	긍·부정률 그래프 구현 기능	- BootStrap - JavaScript - Python	
3	키워드(사람,기관)에 대한 기간별 댓글 수집 기능	- Selenium - Beautiful Soup - Python - Multi threaded Programming	
4	수집된 댓글의 감성 분석 및 긍·부정률 판단 기능	- 딥러닝 BERT model - 합성곱 신경망(CNN) - Python	
5	인기 검색어 기능	- HTML - CSS - BootStrap	
6	검색 키워드에 대한 긍·부정률 나타내는 최빈단어 제시 기능	- JavaScript - Python - DataBase	

1-2. 제품 및 서비스의 개발 배경 및 공공데이터 활용 적절성

○ 서비스 아이디어 제안 배경

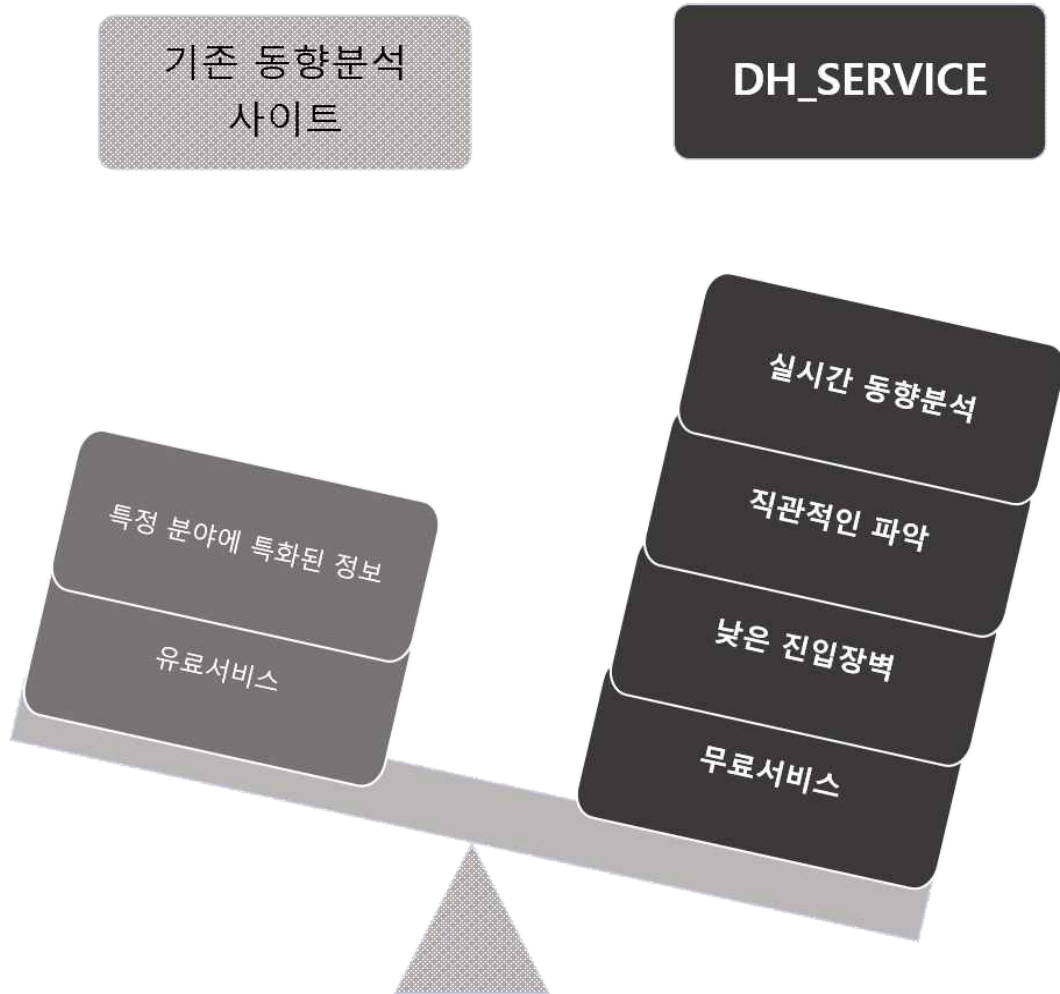
- 빠르게 변화하는 요즘 시대의 동향을 빠르게 파악하고 적응하는 것은 중요함
- 기존에 동향을 파악하는 방식은 조사기관을 통한 금전적인 지출과 많은 시간이 필요함
- 자유롭게 의견을 표출하는 인터넷의 데이터로 보다 다양한 정보 수집 가능
- 접근이 쉬운 웹과 앱을 통해서 어디서든 쉽게 동향파악에 도움이 되는 데이터 습득

○ 공공데이터 활용

주관	데이터명	데이터 특성	데이터 활용
한국지능 정보사회 진흥원 (AI Hub)	감성 대화 말뭉치 https://aihub.or.kr/aidata/7978	일반인 1,500명 대상으로 한, 음성 10,000 문장 및 코퍼스 27만 문장	감성별 문장을 활용하여 검색키워드에 대한 감성분석에 활용
	한국어 대화 요약 https://aihub.or.kr/aidata/30714	한국어 대화 원문 35만 건, 한 문장으로 요약된 생성 요약문 35만 건, 개인정보 비식별화, 메타정보(화자정보, 대화 유형, 대화 주제) 부착	정확도 향상을 위한 추가데이터 학습
	한국어 SNS 대화 데이터 https://aihub.or.kr/aidata/30718	한국어 일상대화 200만 건, 말차례 수: 대화당 최소 4회, 평균 8회 내외, 총 1,600만회 내외, 발화 수: 대화당 최소 8회, 총 3,200만회 내외, 개인정보 비식별화, 메타정보(화자정보, 대화 유형, 대화 주제) 부착	정확도 향상을 위한 추가데이터 학습

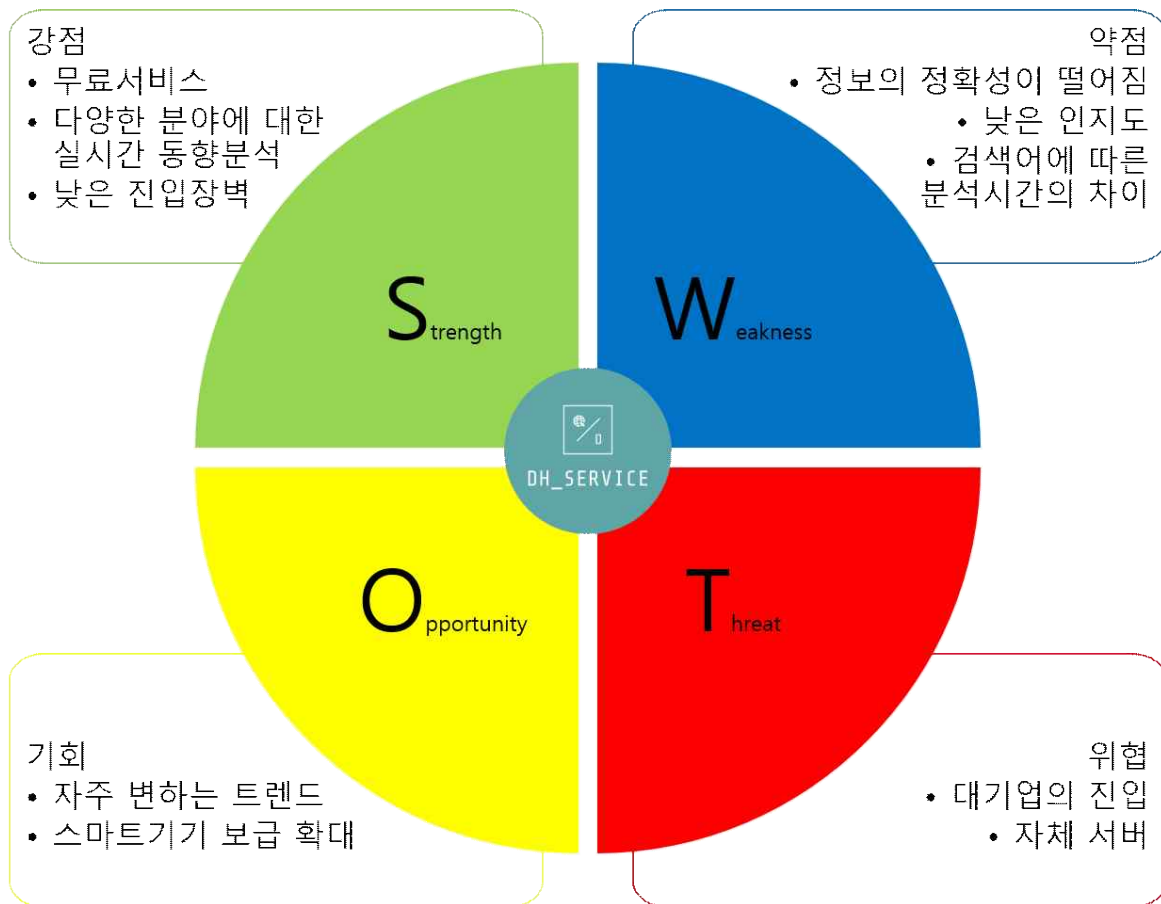
1-3. 제품 및 서비스의 독창성 및 차별성

○ 유사 서비스 비교

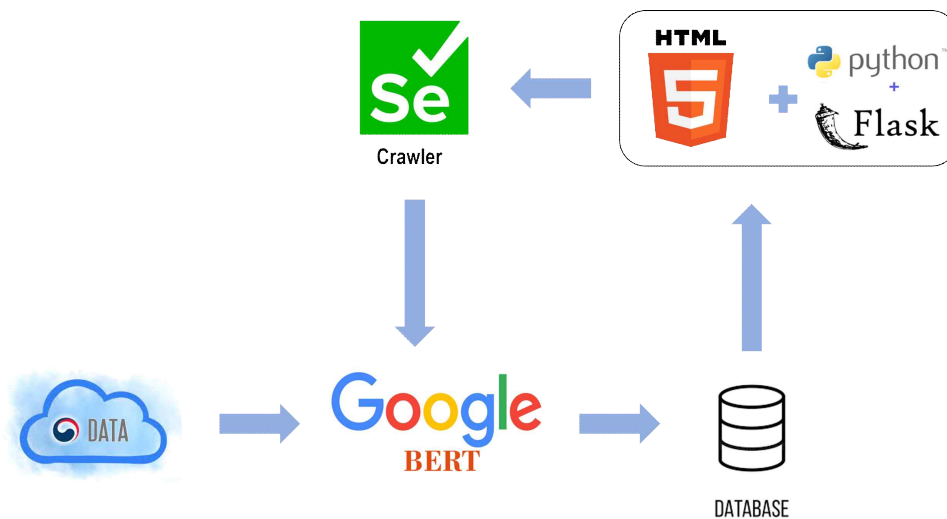


	기존 동향분석 사이트	DH_SERVICE
특징	<ul style="list-style-type: none"> 특정분야에 특화된 정보 <ul style="list-style-type: none"> - 특정 분야 흐름 및 시장성을 분석함(ex. 경제전망, 시장지표 등) 유료서비스 <ul style="list-style-type: none"> - 가입하고 결제후 사용하는 형식 	<ul style="list-style-type: none"> 실시간 동향분석 <ul style="list-style-type: none"> - 다양한 분야의 대중들의 반응을 원하는 기간을 설정하여 실시간으로 파악이 가능함 직관적인 파악 <ul style="list-style-type: none"> - 그래프와 키워드를 분석하여 해결책을 직관적으로 파악 가능 낮은 진입장벽 <ul style="list-style-type: none"> - 로그인 없이 개인 정보 유출 위험이 없고, 간단한 UI를 통해 접근이 쉬움 무료 서비스 <ul style="list-style-type: none"> - 무료서비스로 누구나 쉽게 이용 가능함

○ DH_SERVICE SWOT 분석



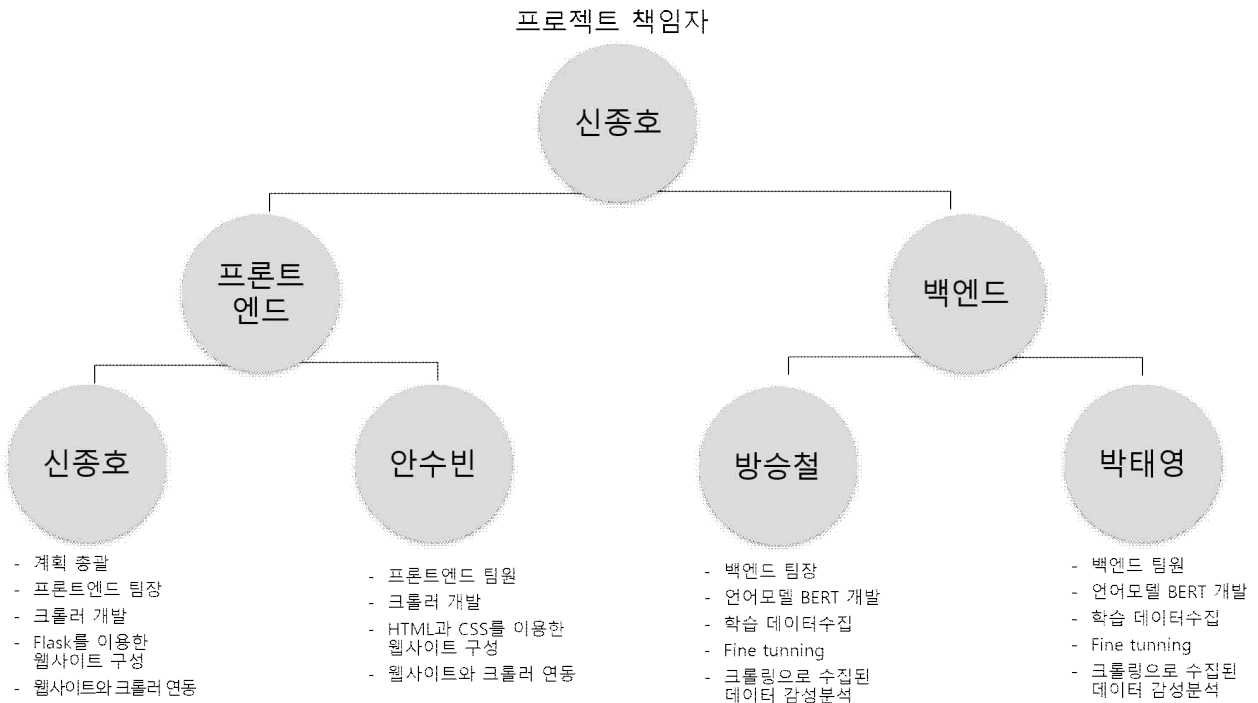
○ 독창성



- 70만개의 문장으로 된 공공데이터를 이용한 딥러닝 BERT model 학습
- Multi threaded 크롤러를 이용한 빠른 정보 취득
- 학습된 모델로 여러 문장의 긍·부정 판단
- 데이터베이스를 이용해 검색순위, 단어 빈도수, 최빈단어 등 다양한 형식 서비스 제공 가능

2-1. 참가자(팀)의 개발 및 사업 역량

○ 인력 편성



이름	경력
신중호	1. 딥러닝 기초 구현능력 함양을 위한 파이토치 초급 과정 참여 2. 현대자동차 전기차 컨버터, 인버터 조립 및 품질 관리 3. 의료적 관리가 필요한 장애인을 위한 재활운동기기 개발 및 임상 중개연구 - 참여 연구원
방승철	1. 2021 동계어드벤처디자인경진대회 입상 2. 2021년 가속기장치 선도기술개발 사업 - 참여연구원 3. 장애인 재활운동 및 체육에 관한 데이터 통합 및 부처연계 방안 연구 - 참여 연구원
박태영	1. '2019 청주대학교 창업경진대회' 사업화부문 장려상 2. '2019 소셜벤처경연대회 대전 권역' 입상 3. '2019 공공데이터 활용 창업교육' 수료 4. 청주상공회의소 'IP창업Zone 6기' 수료 5. 인공지능(음성인식/출력) 기반 전동 높낮이 조절 안마베드 개발 - 참여 연구원
안수빈	1. 2018 HTML교육 수강 2. 대학생 코딩캠프 7기 3. 2021년 가속기장치 선도기술개발 사업 - 참여연구원
전 인원	1. ACIN Lab 학부연구생 2. 인공지능소프트웨어학과 재학

2-2. 제품 및 서비스 개발 실적

○ 주요 개발 일정 및 성과

번호	구 현 계 획	추진 일정	인 원	완료여부
1	프로젝트 계획 보고서 작성	05.10 ~ 05.13	4	완료
2	프로젝트 일정 작성	05.14 ~ 05.16	4	완료
3	웹 크롤러 개발	05.17 ~ 06.07	2	완료
4	BERT 모델 개발	05.17 ~ 06.07	2	완료
5	Flask를 이용한 웹사이트 구성	06.01 ~ 06.10	2	완료
6	학습데이터 수집	06.07 ~ 06.18	2	완료
7	웹사이트와 크롤러 연동	06.11 ~ 06.16	2	완료
8	BERT모델 Fine tuning	06.19 ~ 06.23	2	완료
9	댓글 감성분석 테스트	06.24 ~ 06.26	2	완료
10	긍·부정 데이터 웹사이트에 구현	06.27 ~ 06.28	2	완료
11	알파 테스트	06.29 ~ 06.30	4	완료
12	베타 테스트	07.01 ~	4	진행
13	어플리케이션 개발	계획	4	계획
14	서비스 속도, 정확도 향상	계획	4	계획
15	수익모델 재구축	계획	4	계획
16	부가기능 추가	계획	4	계획

2-3. 개발된 제품 및 서비스의 개선/보완 및 확장 계획

○ 어플리케이션 개발

- 어플 형식으로 전환하여 서비스 사용의 편의성을 증진함
- 지속적인 유지보수를 통해서 모델 정확도 개선 및 데이터 출력 데이터 다양화

○ 서비스 속도 향상

- DataBase에 키워드별로 날짜, 댓글, 긍·부정률을 저장하여 추후에 같은 검색어와 겹치는 기간이 발생하면 DataBase에 저장된 데이터를 사용함
- Crawler Multi threaded programing 성능향상
- AWS 가상 서버 호스팅 사용

○ 정확도 향상

- 검색 키워드와 댓글의 내용이 연관되어있는지 파악하는 딥러닝 모델을 만들어 결과에 대한 신뢰도, 정확도를 높임
- 다양한 플랫폼별로 많은 양의 데이터를 이용해 정보의 객관성, 정확성을 높임

○ 수익모델 재구축

- 서비스 부분 유료화 전환
- 결과보고서 주기적 발송 서비스 제공

○ 부가기능 추가

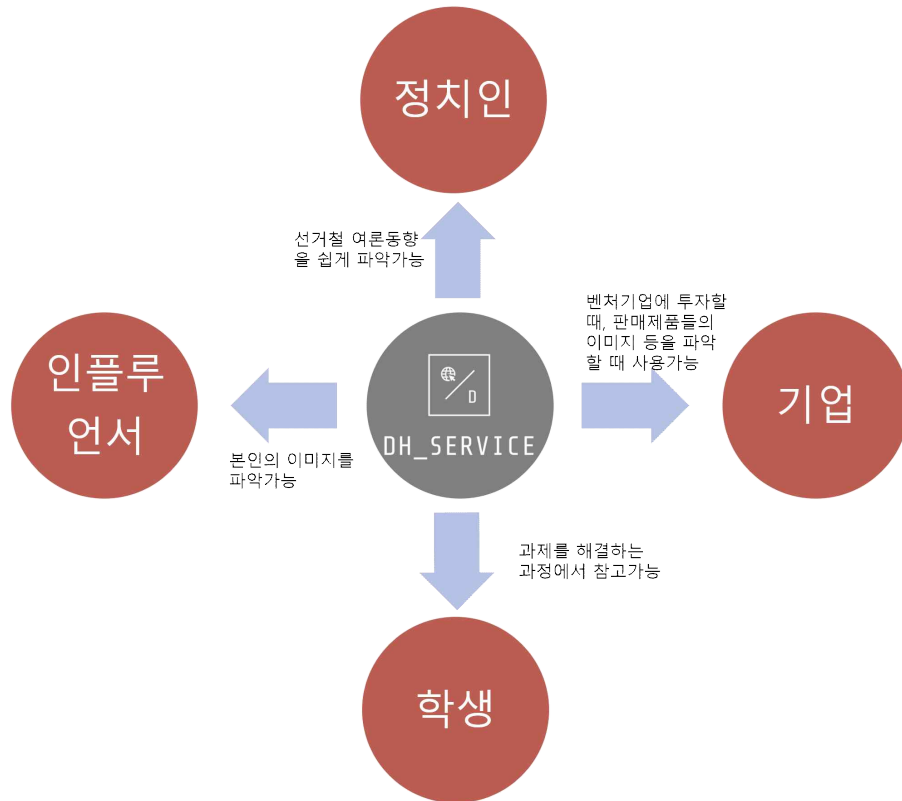
- 검색어 관련 동향분석 보고서 크롤링 및 다운로드 기능
- 다양한 감성 분석 결과
ex) 가장 많이 나온 단어 파악 그래프, 댓글 수에 따른 관심도 그래프
- 유튜브, facebook, instagram 등, SNS 댓글의 동향파악 기능
- 많이 언급된 긍·부정 단어 나열

3

창업 및 사업화 계획

3-1. 개발된 제품 및 서비스를 활용한 창업(사업) 계획

○ 예상 고객

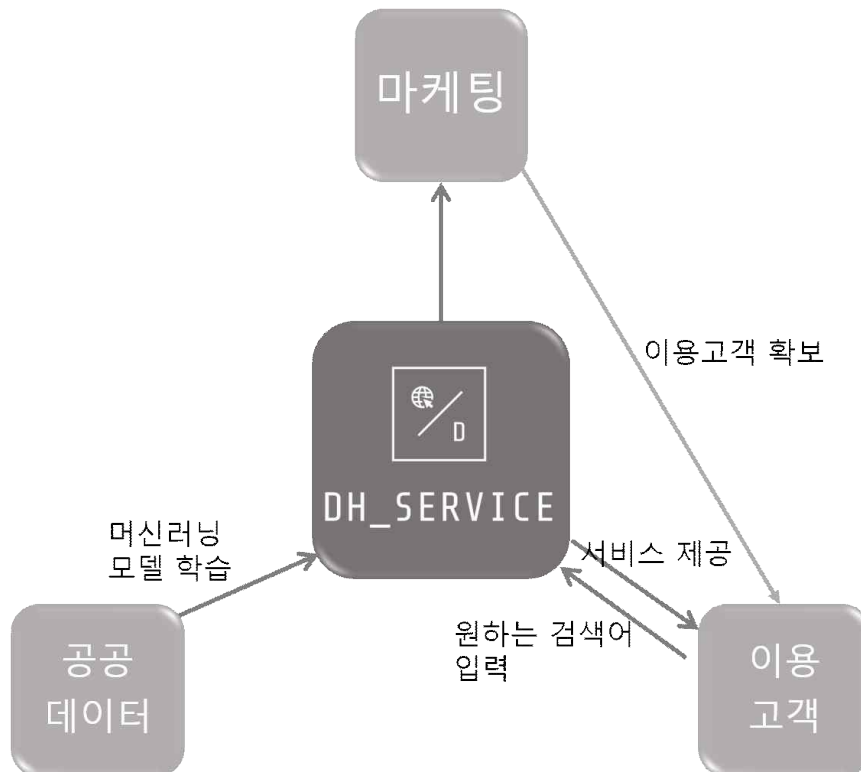


○ 시장성

- 동향, 통계 분석 사이트는 대부분 유료 및 회원서비스임
☞ "DH_SERVICE"는 무료서비스로 누구나 쉽게 이용 가능함
- 기존 동향, 통계 분석 사이트는 특정 분야 흐름 및 시장성을 분석함
☞ "DH_SERVICE"는 다양한 분야의 대중들의 반응을 원하는 기간을 설정하여 실시간으로 파악이 가능함
- 최근 급변하는 트렌드
☞ "DH_SERVICE"는 트렌드에 맞춰 변하는 대중들의 반응을 표나 그래프를 통해 한눈에 파악이 가능해 트렌드를 맞춰갈 수 있음
- 스마트기기의 보급 확대
☞ "DH_SERVICE"는 여러 스마트기기에서 사용할 수 있을 뿐만 아니라 보급이 확대됨에 따라 포털사이트, SNS 등 발생하는 댓글 수가 증가하여 정보의 객관성이 높아짐

3-2. 개발된 제품 및 서비스의 사업화 계획

○ 비즈니스 모델



○ 공공데이터

- 70만 문장으로 이루어진 공공데이터를 전처리하여 딥러닝 모델에 학습시켜 문장 및 단어들의 감성분석을 함

○ 이용고객

- 원하는 검색어와 기간을 입력하면 해당 검색어 기사들의 댓글들을 뽑아내어 딥러닝 모델이 긍·부정을 판단하여 그래프 형식으로 동향 분석 서비스를 제공함

○ 마케팅

- 기업과 인플루언서들에게 우선적으로 시범 운영한 후, 3개월 후 유료로 전환하는 방식
- 유료 서비스의 금액이 부담되는 이용자에게는 친구 추천하는 방식으로 3명당 1개의 보고서를 열람할 수 있는 방식

○ 수익 창출

- 웹페이지 내에 광고를 띄워 수익을 창출할 수 있음.
- 동향분석서비스 특성을 살려 웹 페이지 내에 의뢰자의 긍정적인 뉴스를 홍보할 수 있으며 해당 홍보로 수익을 창출할 수 있음.
- 기업에게는 사용료를 부과하여 수익을 창출하며, 대가로 유료 서비스 사용 가능.
- 일반 사용자의 경우 유료 서비스로 직접 전환하여 수익을 창출할 수 있음.
= 유료 서비스 : 일주일마다 동향 분석 보고서를 사용자에게 보내줌