

한이음 ICT멘토링 프로젝트 중간보고서

프로젝트 정보

프로젝트명	자기소개서 데이터 분석
프로젝트 소개	<p>4차 산업혁명으로 빅데이터와 인공지능이 대두되면서 그 활용은 채용 시장에서도 나타나고 있다. 본 프로젝트는 채용시장 중 대학생들이 가장 어려워하는 자기소개서를 주제로 한다. 그 중에서 취업용 자기소개서를 중심으로 프로젝트를 진행한다. 그 이유는 현재 대학생들과 기업에서는 인사채용 및 취업 시 많은 비용을 지출하는데 이를 줄이고자 프로젝트를 진행하게 되었다.</p> <p>빅데이터와 인공지능을 이용해 자기소개서의 내용과 트렌드를 파악하여 사용자들에게 제공해준다. 또한, 입력한 자기소개서를 수치화 하여 사용자들에게 핵심 역량 및 직무 성향을 제공할 것이다.</p>
구성도	<pre> graph LR subgraph "데이터 수집 및 정제" A[합격 자소서 데이터 (잡코리아)] --> B[합격 스펙 관련 데이터 (잡코리아)] B --> C[데이터 클린징] C --> D[텍스트 정제] C --> E[변수 일반화] D --> F[분석단위로 통합] E --> F F --> G[전공, 충분류, 소분류 • 학점, 자격증 등 일반화 • 질문 범주화] end G --> H[2. 자소서 질문 범주화 (동기, 목표 등), 자소서 답변] H --> I[3. 토익, 토스, 오피, 자격증, 해외경험, 인턴, 수상, 동아리, 교내활동, 사회활동, 자원봉사 스펙 데이터] I --> J[4. 평가 (아쉬운점, 좋은점, 총평 데이터)] J --> K[분석 및 모델링] K --> L[2. 자소서 답변 데이터 • 단어-문서 행렬 생성 • 토픽 모델링 • 평가 데이터 • 스코어링] L --> M[분석 활용] M --> N[3. 결과 제공] N --> O[4. 결과 제공] </pre> <p>Legend:</p> <ul style="list-style-type: none"> 범례 Activity 데이터 (출처) Activity (데이터셋 생성) 분석데이터셋 결과물 Activity (폐단/현황 도출) 활용분야

개발배경 및 필요성	<ul style="list-style-type: none"> ▷ 자기소개서와 핵심 역량의 트렌드 파악 <ul style="list-style-type: none"> ○ 현재 취업용 자기소개서 연구가 활발히 진행되고 있지만 대학생들이 어떻게 작성하면 도움되는지에 대한 정보공유가 미비하다. 그래서 저희가 직접 자기소개서에 대한 연구를 진행하고 웹과 앱을 통해 자기소개서에 관한 트렌드를 학생들에게 공유하고자 한다. ▷ 자기소개서 요약 및 핵심 역량에 대한 스코어링 <ul style="list-style-type: none"> ○ 자기소개서를 작성할 때 자신의 자기소개서가 좋은 자소서인지 확인하고 판단하기 어렵다. 또한, 사람이 직접 하면 비용이 많이 사용된다. 만드는 프로그램을 이용하면 자기소개서 핵심 역량과 해당 자기소개서 스코어를 빠르고 저렴하게 확인할 수 있다..
특 · 장점	<ul style="list-style-type: none"> ▷ 컴퓨터가 자동으로 알고리즘에 의해 자기소개서를 분석하고 객관적이고 일관되게 자기소개서를 평가가 가능하다. <ul style="list-style-type: none"> ○ 오프라인으로 자기소개서 평가를 받을 시 평가하는 사람에 따른 능력 및 컨디션에 따라 자기소개서 평가에 대한 퀄리티가 천차만별이다. 하지만 저희 자체개발한 알고리즘에 따라 객관적이고 일관되게 자기소개서를 평가할 수 있다. ▷ 현재 자기소개서 평가가 온라인을 이용해 접근성이 향상되고 누구나 이용할 수 있다. <ul style="list-style-type: none"> ○ 웹과 프로그램을 이용해 쉽게 이용하고 많은 사람이 이용할 수 있다.
주요 기능	<ul style="list-style-type: none"> ▷ 자기소개서와 핵심 역량의 트렌드 파악 <ul style="list-style-type: none"> ○ 자기소개서를 자체 학습법을 연구한 것을 토대로 텍스트 및 동영상을 통해 자기소개서 작성법을 올릴 예정입니다. ▷ 자기소개서 요약 및 핵심 역량에 대한 스코어링 <ul style="list-style-type: none"> ○ 사용자가 자기소개서를 입력하면 자체 알고리즘을 통해 사용자의 자기소개서 요약과 스코어를 제공한다.
기대효과 및 활용분야	<ul style="list-style-type: none"> ▷ 자기소개서가 핵심 역량과 관련된 정도를 점수화하여 직무 성향을 파악하고 숨겨진 인사이트를 발견할 수 있다. <ul style="list-style-type: none"> ○ 자기소개서를 통해 지원자의 상태와 능력을 파악할 수 있어 사용자들의 인사채용결정 프로세스에 많은 도움을 줄 것이라고 기대하고 있다. ▷ 기업의 인사 채용에 대한 의사결정을 지원할 수 있다. <ul style="list-style-type: none"> ○ 현재 기업들은 자체 인사채용과정에서 많은 인적 코스트가 발생하고 있다. 하지만 본 프로젝트에서 개발한 프로그램 및 웹을 사용하면 저렴하고 간편하게 이용할 수 있다. ▷ 자기소개서의 내용을 요약, 전달하여 AI 면접관이 관련된 질문을 생성할 수 있다.

I. 프로젝트 개요

1. 프로젝트 소개

- 다양한 분야의 자기소개서를 수집하고 그 내용을 요약, 분석하여 족업 준비생에게 적절한 피드백을 제공하는 서비스를 창출한다.
- 자기소개서에 대한 핵심 역량 및 직무 성향을 점수와 그래프로 표현함으로서 기업의 인사 채용에 대한 의사결정을 지원한다.

2. 추진배경 및 필요성

- 다량의 자기소개서를 작성하거나 읽을 때 노력이 따르게 된다. 이를 효율화하기 위해 자기소개서 내용을 신속하고 정확하게 파악할 수 있는 첨단 기술이 요구된다.
- 핵심 역량 중심의 채용 트렌드와 개인의 직무 성향을 파악하는 것이 더욱 중요해지고 있다. 다양한 측면으로 자기소개서를 분석하고 숨겨진 인사이트를 발견해야 한다.

3. 국내 · 외 기술 현황

- Text Mining은 이미 국내외를 막론하고 광범위하게 활용되고 있으며, 현재 대표적으로 널리 응용 되고 있는 분야는 소셜미디어와 이메일이다.

Facebook과 Twitter는 테러, 폭력, 마약과 같은 관련 용어를 쓰는 문제의 계정을 관리자에게 보고하고 해당 계정을 폐쇄토록 해 범죄를 사전 차단하는데 주력하고 있으며,

이와 마찬가지로 이메일 서비스를 제공 하는 Microsoft, Naver 등의 회사도 자사 Text Mining 기술을 이용해, 스팸메일이 가지고 있는 문자적 특성을 파악하고 Filtering 해 User가 스팸메일을 읽지 않도록 도와주고 있다.

4. 개발목표 및 내용

- 최종 개발목표
 - 인사처에 채용 의사 결정에 대한 도움을 줄 수 있다.
 - 기존 시간 및 인력 비용을 아껴줄 수 있다.
 - 자기소개서 내용을 요약하고 핵심 역량과 관련된 정도를 점수화하여 인사 채용에 발생하는 비용을 최소화 한다.
 - 다수의 자기소개서와 관련 된 텍스트 데이터를 수집, 분석하여 채용 시장의 트렌드를 파악한다.

○ 주요 개발내용(기능중심)

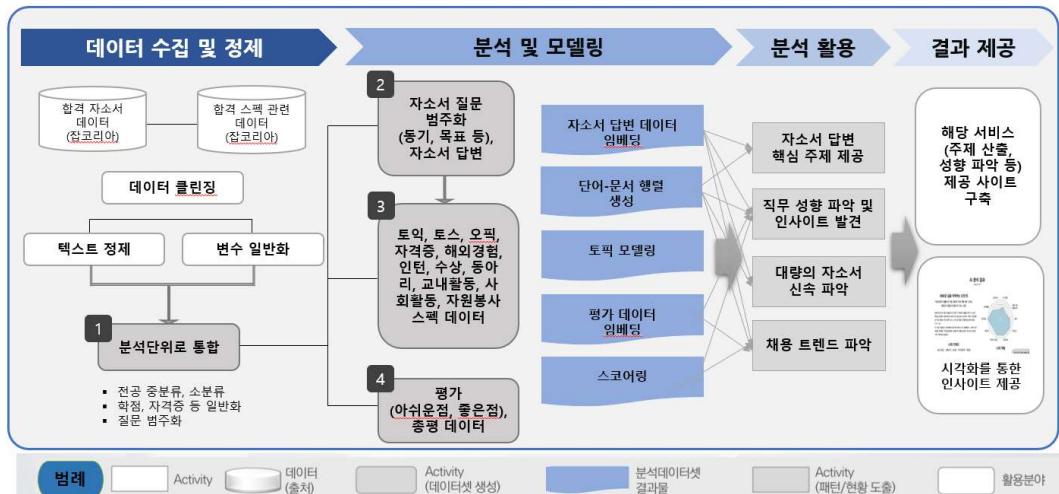
- 데이터 수집을 위한 크롤링 기술(Python의 BeautifulSoup, Selenium) 모듈 활용과 형태소 분석기 적용, 제작 (Kkma, Komoran, Mecab, Stanford NLP 등), 자기소개서 트렌드 파악을 위한 토플 모델링(LDA, LSA 등) 핵심 역량 키워드 매칭을 위한 단어 및 임베딩 기술(FastText, GloVe, ElMo, Bert 등) 대시보드를 이용한 분석 결과 시각화(Wordcloud, Rader Chart, Tableau 등), 시스템 자동화 및 구현을 위한 개발(Python)의 기능을 이용한다.

○ 기존 기술 활용여부 및 차별성

- 자기소개서가 핵심 역량과 관련된 정도를 점수화하여 직무 성향을 파악하고 숨겨진 인사이트를 발견할 수 있다.
- 자기소개서의 내용을 요약, 전달하여 AI면접관이 관련된 질문을 생성할 수 있다.
- 기업의 인사 채용에 대한 의사결정을 지원할 수 있다.

II. 프로젝트 내용

1. 구성도



2. 주요기능

○ 전체 기능 목록

구분	기능	설명	현재진척도(%)
S/W	자기소개서와 핵심역량의 트렌드 파악	다양의 자기소개서와 채용 공고의 텍스트 데이터를 수집, 분석하여 채용 시장의 트렌드를 파악한다.	40%
	자기소개서 요약 및 핵심 역량에 대한 스코어링	자기소개서 내용을 요약하고 핵심 역량과 관련된 정도를 점수화하여 인사 채용에 발생하는 비용을 최소화 한다.	40%

○ S/W 주요 기능

기능	설명
자기소개서와 핵심역량의 트렌드 파악	다양의 자기소개서와 인터넷 상의 데이터를 수집과 분석하여 채용 시장의 유행을 파악한다.
자기소개서 요약 및 핵심 역량에 대한 스코어링	자기소개서 내용을 요약하고 핵심 역량과 관련된 정도를 수치화하여 인사 채용에 발생하는 비용을 감소시킨다.

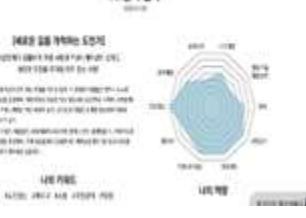
○ H/W 주요 기능

기능/부품	설명
N/A	N/A
N/A	N/A

3. 적용기술

- 데이터 수집을 위한 크롤링 기술: Python의 BeautifulSoup, Selenium 모듈 활용
- 형태소 분석기 적용: Kkma, Komoran, Mecab, Twitter, Hannanum 등
- 형태소 분석기 제작: Stanford NLP, Soynlp, SentencePiece 등
- 자기소개서 트렌드 파악을 위한 토픽 모델링: LDA, LSA 등
- 핵심 역량 키워드 매칭을 위한 단어 및 임베딩 기술: FastText, GloVe, ElMo, Bert 등
- 대시보드를 이용한 분석 결과 시각화: Wordcloud, Rader Chart 그래프, Tableau를 이용한 대시보드
- 시스템 자동화 및 구현을 위한 개발: Python

4. 예상 결과물

예상 결과물 이미지	설명
	자기소개서 데이터를 수집을 한 후 Python의 라이브러리인 BeautifulSoup, Pandas 등을 사용을 하여 전처리를 했서 데이터를 생성했다.
	개인의 직무 성향을 간결히 파악 할 수 있도록 LDA, FastText등등을 사용해 자기소개서를 분석해 그래프로 표현하여 대시보드를 제공한다.

III. 프로젝트 수행내용

1. 프로젝트 수행일정

프로젝트 기간 (한이음 사이트 기준)		2020.00.00. ~ 2020.00.00.											
구분	추진내용	프로젝트 기간											
		1 월	2 월	3 월	4 월	5 월	6 월	7 월	8 월	9 월	10 월	11 월	12 월
계획	세부 진행사항 기획												
설계	활용 데이터 식별												
	분석 방법 및 시각화 기법 연구												
	프로젝트 위험 요인 파악												
	데이터 수집 및 분석 설계												
	시각화 및 개발 설계												
	최종 보고서 설계												
분석	데이터 수집 및 전처리												
	데이터 분석 및 시각화 진행												
	개발												
테스트	최종 테스트 진행												
종료	보고서 작성												
오프라인 미팅	월 1회 오프라인 미팅 진행												

2. 프로젝트 수행 과정에서의 문제점 및 애로사항

- 코로나 바이러스와 그에 따른 정부 정책 준수로 인한 오프라인 미팅 최소화와 그에 따른 어려움

VI. 기대효과 및 개선사항

1. 기대효과

- 다양한 분야의 자기소개서를 수집하고 그 내용을 요약, 분석하여 취업 준비생에게 적절한 피드백을 제공하는 서비스를 창출한다.
- 자기소개서에 대한 핵심 역량 및 직무 성향을 점수와 그래프로 표현함으로써 기업의 인사 채용에 대한 의사결정을 지원한다.
- 팀원 간 협력을 통해 문제해결능력을 함양한다.
- 실무 중심의 프로젝트 경험을 통해 실제 업무 프로세스를 체험한다.
- 다양한 분석 및 시각화 기법을 학습하고 구현해보는 경험을 쌓는다.
- 여러 협업 툴의 사용 방법을 익힌다.

2. 개선사항

- 코드의 Optimization
- 기타 각종 버그/오류 수정
- 각종 Module을 이용한 feature 추가