

# 위(WE)로(ROAD)

---

AI 기반 직업 검색엔진

 김재형 유진영 조한영 정지은

# 목차

## 01

### 프로젝트 개요

- 주제 선정 배경
- 시장 분석

## 02

### 프로젝트 진행

- 데이터 수집
- 데이터 전처리
- 모델 탐색

## 03

### 프로젝트 결과

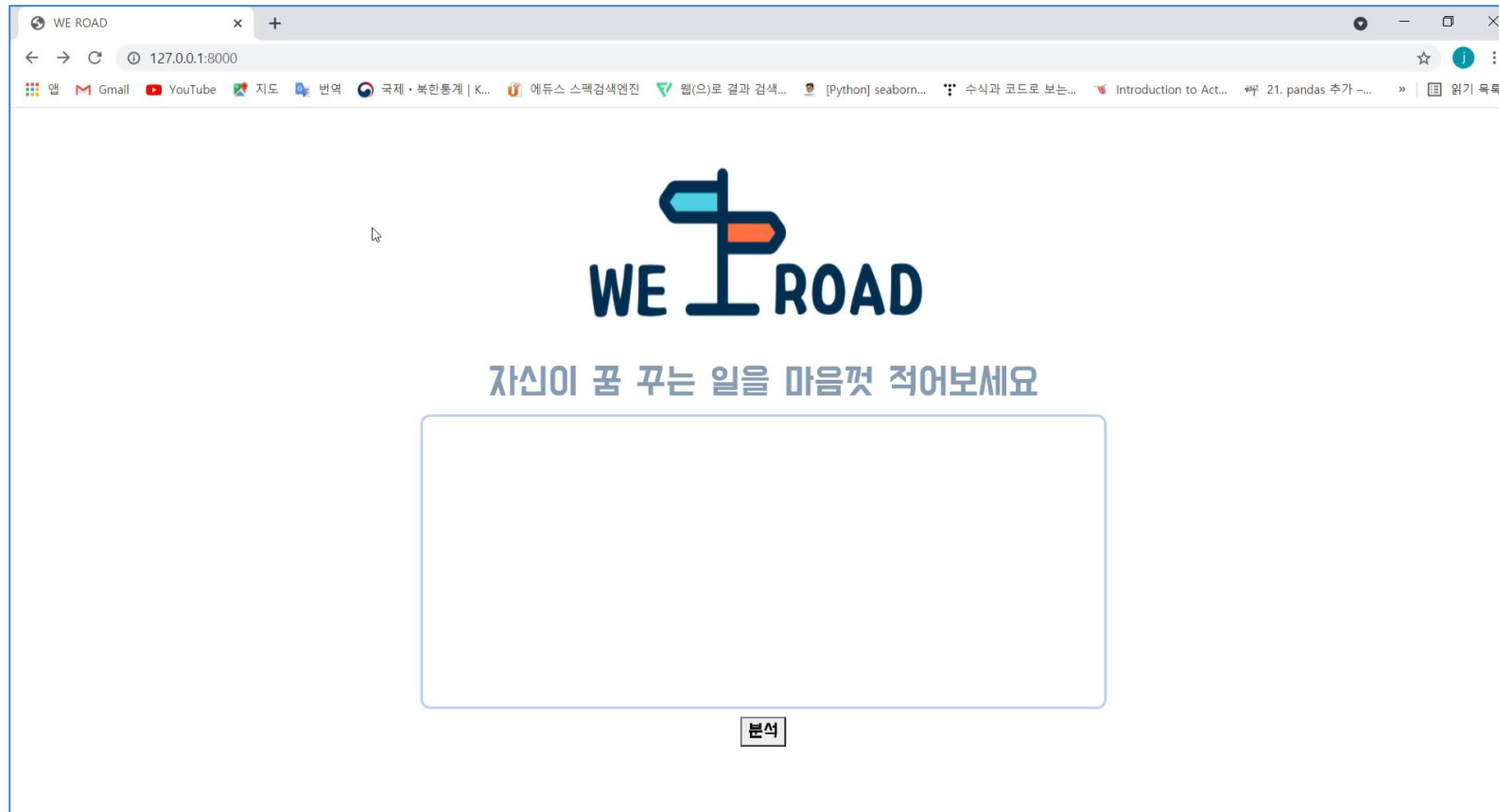
- 웹 구현
- 활용방안
- 한계점

## 04

### 참고사항

- 개발도구
- Q&A

# 구현



## 프로젝트 개요

# 주제 선정 배경



# 주제 선정 배경

## IT 비전공 구직자 36%, IT 개발 직무 취업 하고파!

구직자 1,131명 설문조사 [자료제공: 사람인]

IT 개발 직무  
취업 원한다 **36.3%**

원하지 않는다 **63.7%**

\* IT 개발 직무로 취업하고 싶은 이유 (복수응답)

앞으로 계속 유망한 직무여서 **71.8%**

취업이 잘 될 것 같아서 **39.7%**

타 직무보다 연봉이 높아서 **24.8%**

IT 회사의 주요 직무를 하고 싶어서 **16.1%**

특기, 적성에 맞는 것 같아서 **11.4%**

saramin



코드스테이츠, IT 개발자로 커리어 전환하는 수강생 86% 컴퓨터공학 비전공자

정부, 비전공 개발자 '18만명' 육성...전문가도 긍정 전망

IT 업계 호황에 비전공자도 "개발자 이직 관심 폭증"

# 시장 분석

## ✚ 페르소나 선정



- 이름 : 김근호 (29살)
- 학력 : 4년제 문과 졸업생

- 졸업 후 취업이 되지 않아 새로운 분야로 전향하고자 함
- 최근 IT 개발자 채용이 많아지고, 연봉이 높다는 기사를 보고 개발자 직군으로 취업하고자 함
- 개발자에 대해 알아보니 전문용어로 된 수 많은 정보로 인해 세부분야를 결정하는데 어려움을 느낌
- 간단한 검색을 통해 자신이 하고 싶은 일이 어떤 분야인지 알고 싶음

# 목표

텍스트 입력



분류



결과

“디자인 하는게 너무 재밌는데 이번에 뉴스를 보니까 그림도 컴퓨터로 표현할 수 있다고 하네?”

딥러닝 모델

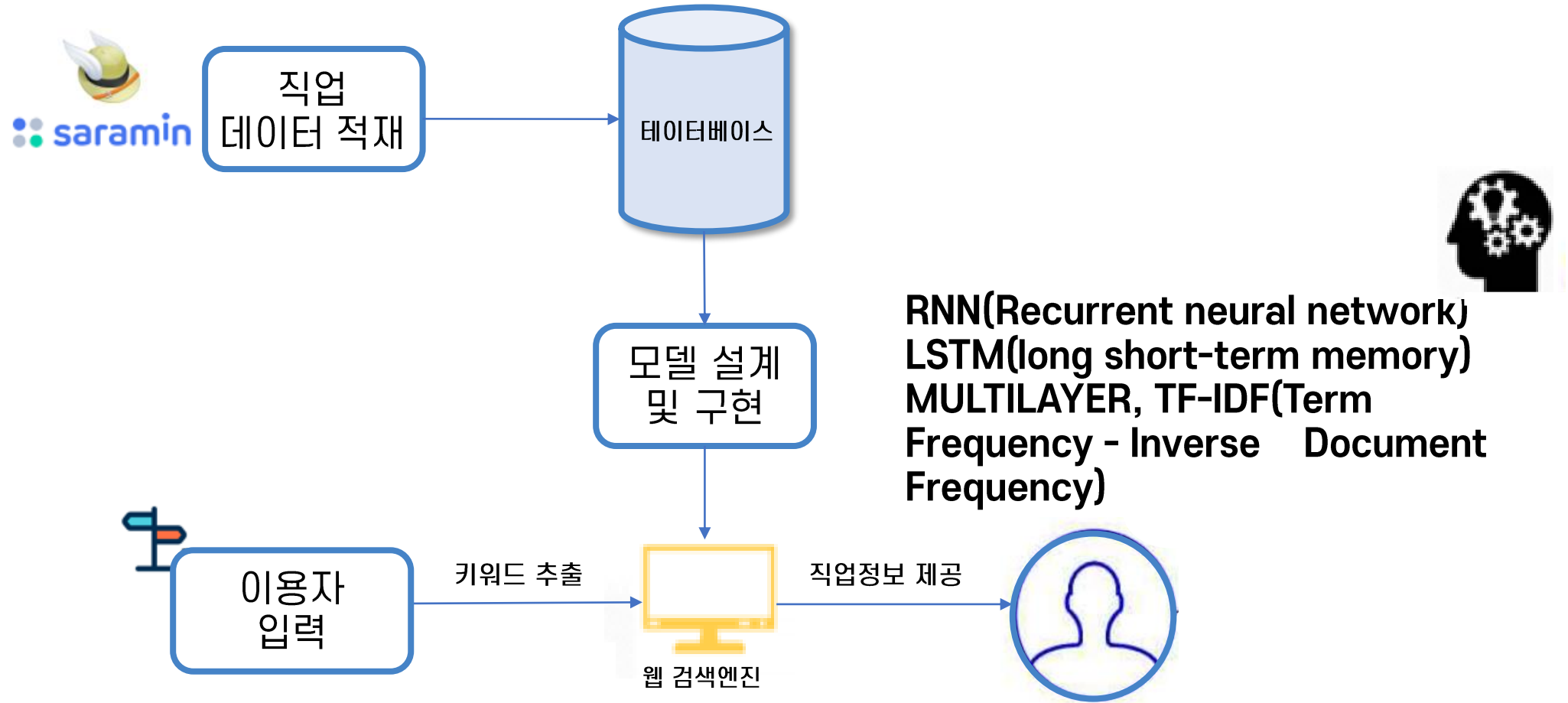
웹 프로그래머 직업에 대한 정보를 제공한다

하고싶은 일을 입력하면(단, IT 전문용어가 아니어도 됨)  
나에게 잘 맞는 직업정보(하는 일, 훈련정보 등)를 찾아주는 웹 검색엔진.



## 프로젝트 진행

# 아키텍처



# 요구사항 정의서

No ▼	REQ ID ▼	구분 ▼	명칭 ▼	요구사항 상세 ▼
1	REQ-FUN-001	기능	키워드 분석	키워드를 입력 받고 그에 맞는 직업 추천
2	REQ-FUN-002	기능	입력	사용자가 문장 또는 키워드 입력
3	REQ-FUN-003	기능	모델 사용데이터	자연어 처리
4	REQ-FUN-004	기능	알고리즘 구현	RNN, LSTM
5	REQ-FUN-005	기능	데이터 수집	BeautifulSoup 크롤링
6	REQ-FUN-006	기능	웹구현	Django 웹 프레임워크
7	REQ-NON-001	비기능	품질	테스트 정확도 90% 이상
8	REQ-NON-002	비기능	유지보수	직업에 따른 정보등을 6개월 단위 업데이트
9	REQ-NON-003	비기능	인터페이스 구현	홈 화면에서 사용자의 문장을 입력, 다음 화면에서 연관정보 출력
10	REQ-NON-004	비기능	보안	별도의 인증 없이 이용(최소의 보안 : 로봇인지 사람 확인)
11	REQ-NON-005	비기능	장비구성	컴퓨터 4대, 아이리포 IP 이용
12	REQ-NON-006	비기능	성능	입력 후 직업 결과 report 제공까지 최대 30초 이내
13	REQ-NON-007	비기능	데이터 요구사항	사람인 직업 사전, 네이버 백과사전, 블로그 검색결과 이용

# 데이터 수집

웹 직업군



검색



게임 직업군



검색



모바일 직업군



검색



✚ 네이버 검색결과

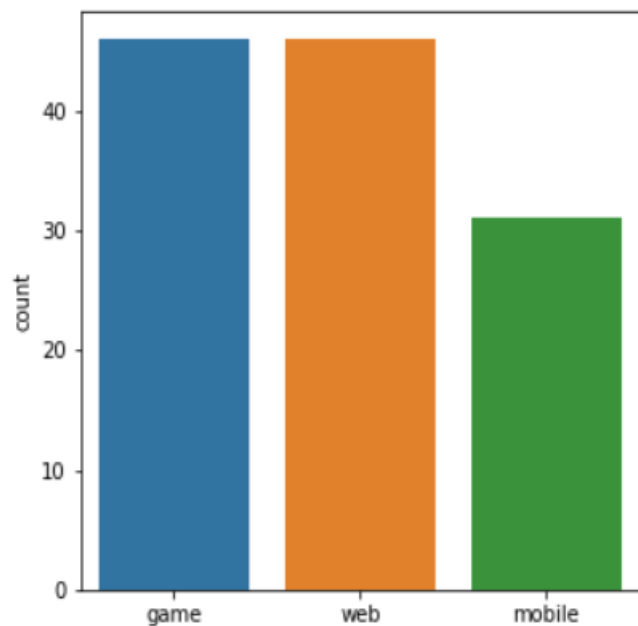
- 지식백과, 블로그

✚ 직업 사전

✚ 채용 공고

# 데이터 분석

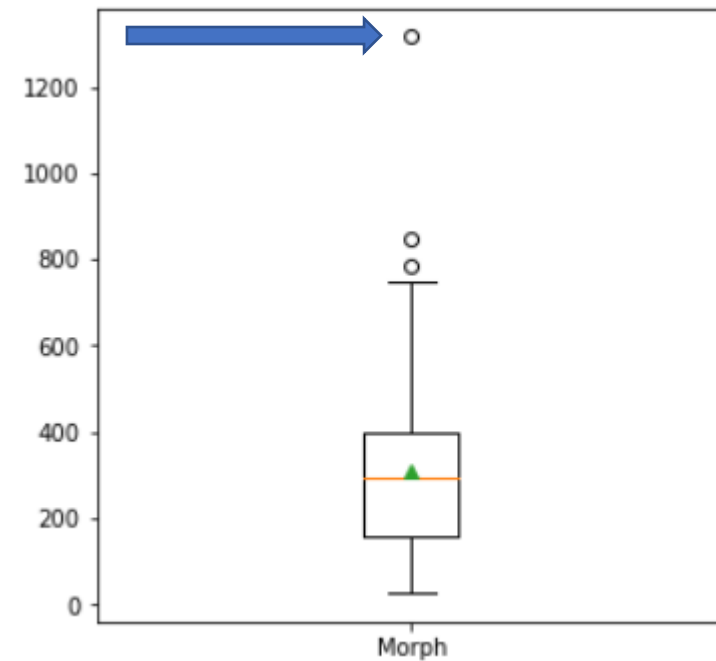
라벨 별 데이터 분포



형태소 길이

형태소 최대길이: 1316  
형태소 최소길이: 26  
형태소 평균길이: 308.64  
형태소 길이 표준편차: 206.85  
형태소 중간길이: 292.0  
형태소 1/4 퍼센타일 길이: 157.0  
형태소 3/4 퍼센타일 길이: 398.0

형태소 길이



# 데이터 분석

## WORD CLOUD



## 웹 직업군



## 게임 직업군

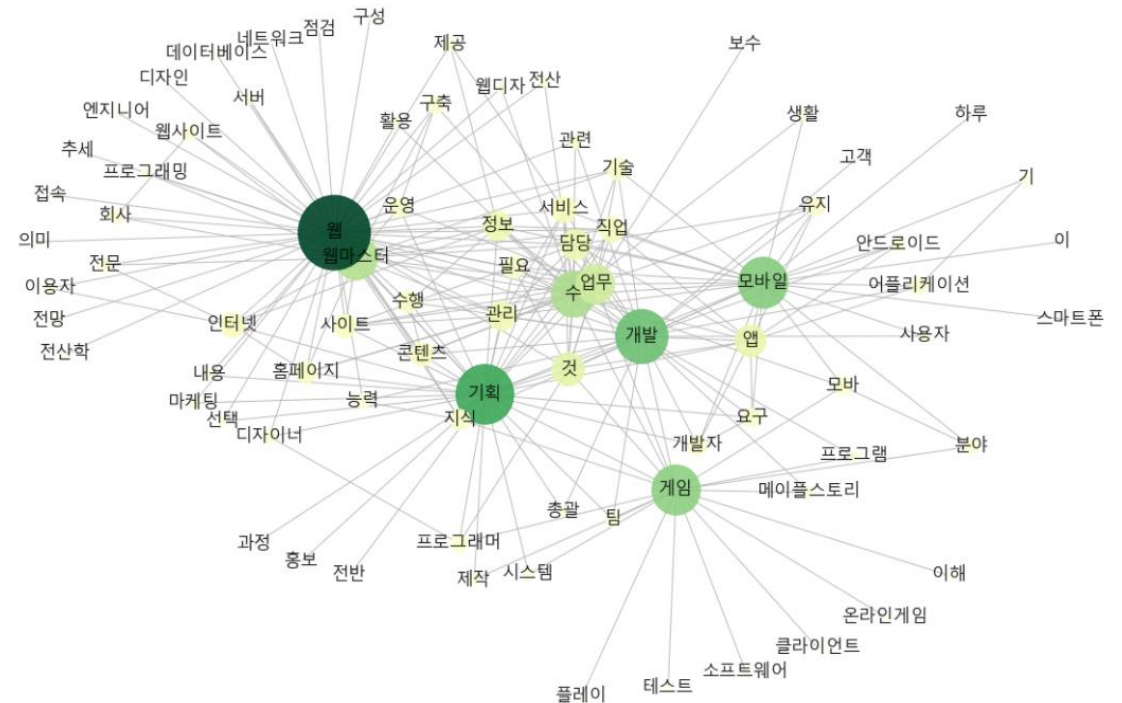


## 모바일 직업군

# 데이터 분석

## 연관 그래프

- Apriori algorithm 이용
- 문서 내 단어들의 연관 패턴 추출
- 연관 키워드 추출 가능



# 데이터 전처리

## ✚ 토큰화 Tokenizer

- 형태소 분석기 : Okt
- 구두점 (punctuation) 제외
- 마침표(.), 쉼표(,), 물음표(?), 세미콜론(;), 느낌표(!) 제거

## ✚ 불용어 처리 Stop word

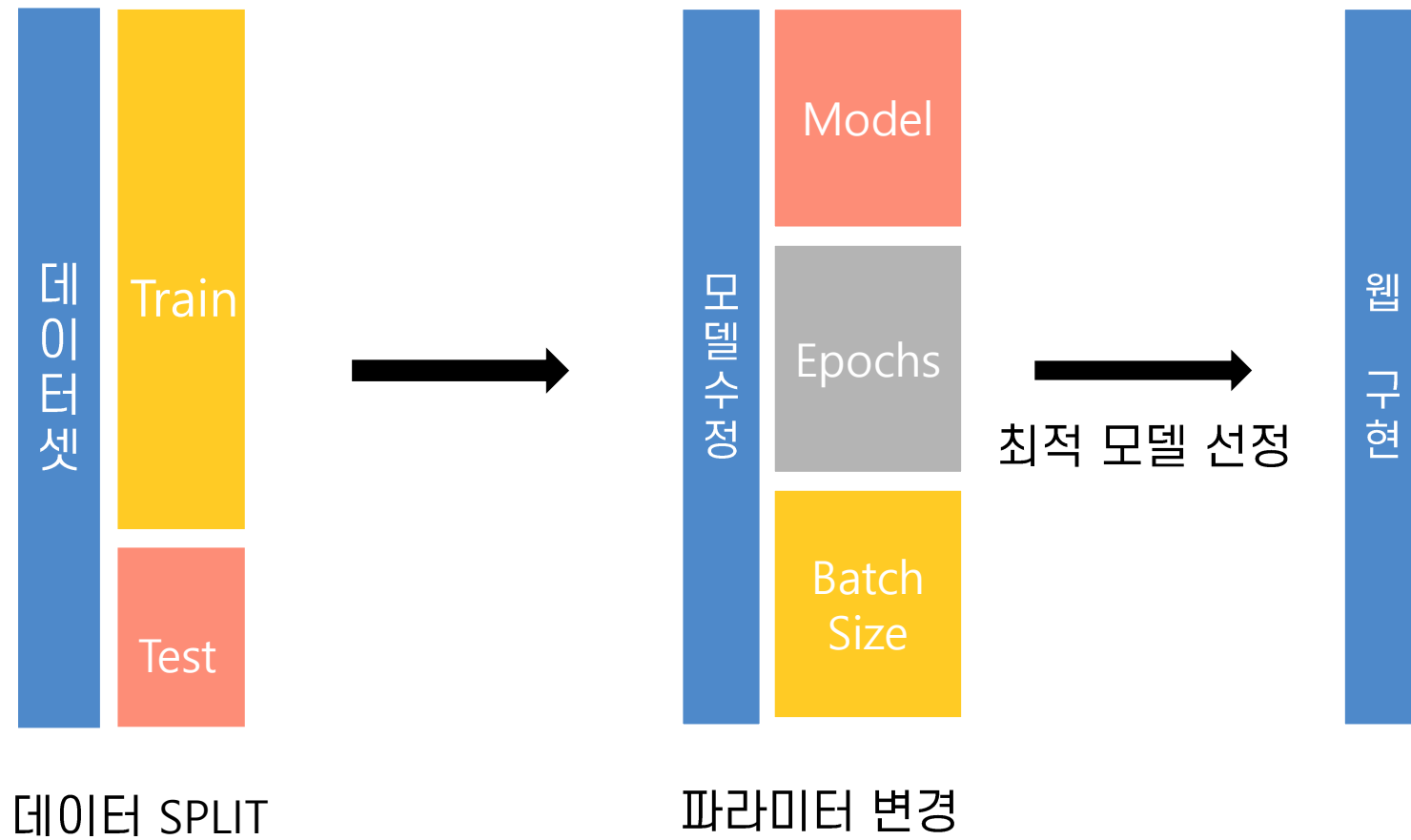
- Word cloud 및 한국어 불용어 사전 참고
- 조사(을,를,이,가 …) 제거



# 데이터 전처리

- ✚ TF-IDF Vectorization
  - Sklearn TfidfVectorizer()
- ✚ 라벨 인코더 Label Encoder
  - 3개 라벨 분류
  - One-Hot encoding

# 모델 개요



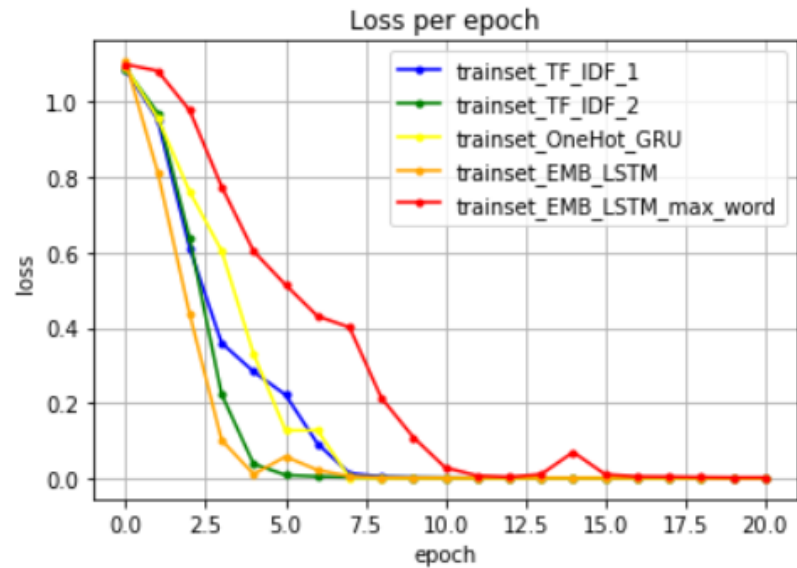
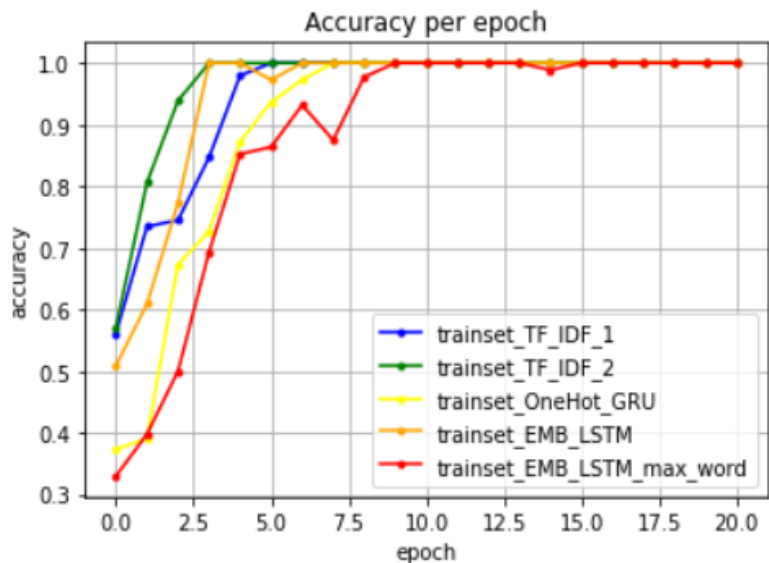
# 모델 탐색

Model	GRU	LSTM_1	LSTM_2	MULTILAYER_1	MULTILAYER_2
Preprocess	One-Hot ENC	Embedding	Embedding	TF-IDF	TF-IDF
Epochs	10	20	30	40	50
Batchsize	2	5	10	20	30

- GRU , LSTM (RNN) : 문서 내 단어들의 흐름이 문서를 분류하는데 있어 의미가 있다는 접근
- MULTILAYER : 문서 내 단어를 벡터화 하고 단어 모두를 속성으로 한 분류가 가능하다는 접근
- 그 외 구현해본 모델 : 나이브 베이즈 모델

# 모델 탐색

## Loss 및 Accuracy



Epochs :  
학습 횟수

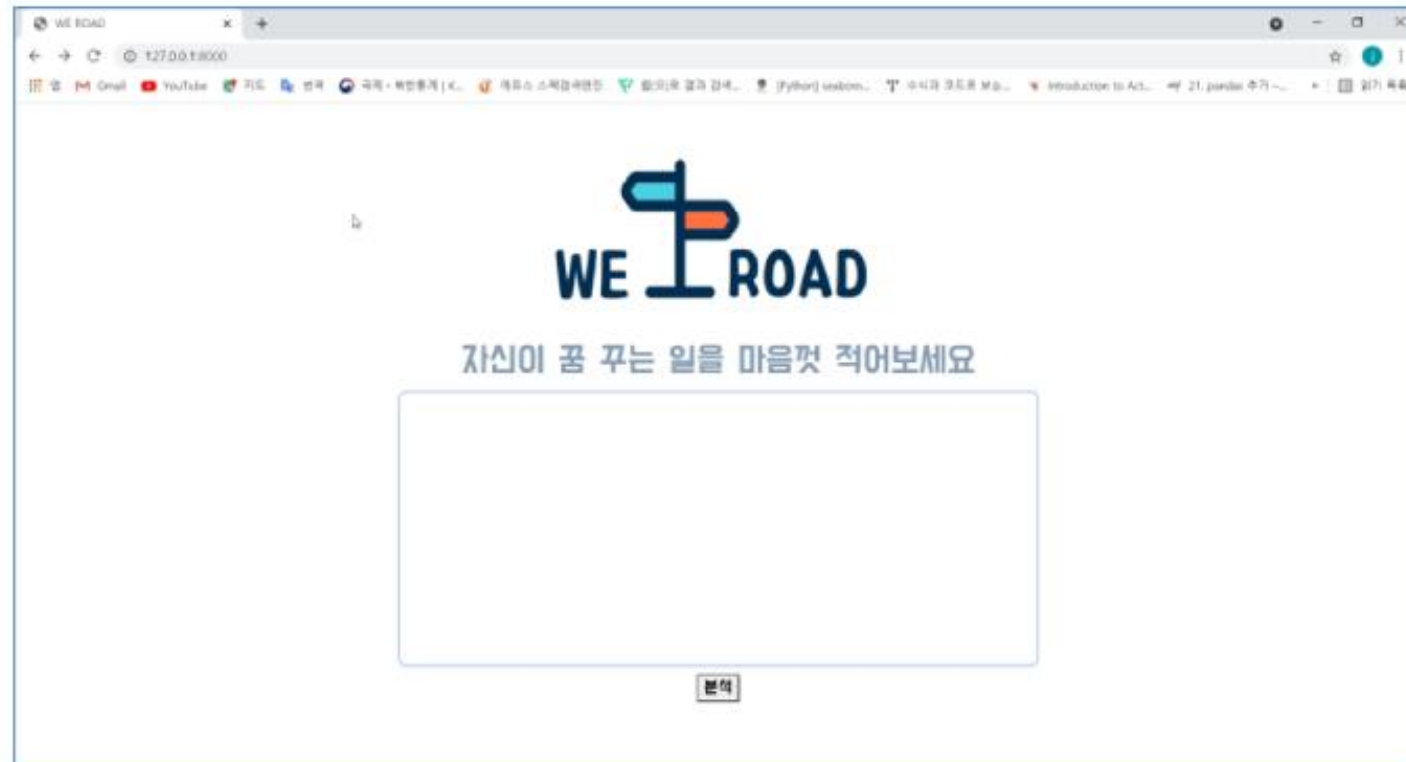
Accuracy :  
모델 정확도

Loss :  
모델 예측값 과  
정답의 차이

- 최종 모델 : TF\_IDF\_2
- Accuracy 96% Loss 0.2

## 프로젝트 결과

# 웹 검색엔진



# 활용방안 및 기대효과

## 활용방안

- 코딩 교육 사이트 교육방향 추천
- 학생 진로 방향성 추천 / 상담

## 기대효과

- 직업정보에 대한 접근성 상승

# 한계점

## ✚ 자연어 처리

- 형태소 분석 한계
- 불용어 기준 설정

## ✚ 데이터

- 직업군 제한의 아쉬움

## ✚ 서비스

- 학습된 데이터 부족으로 다양한 키워드 처리한계



# 시장 분석

## 경쟁 제품



### 분류별 찾기

아래의 9가지 분류 기준으로 직업을 선택하여 원하는 직업정보를 검색할 수 있습니다.  
 관리·경영·금융·보험 | 교육·연구·법률·보건 | 사회복지·문화·예술·방송 | 운송·영업·판매·경  
 미용·숙박·여행·스포츠·음식 | 건설·기계·재료·화학·섬유 | 전기·전자·정보통신 | 식품



### 지식별 찾기

여러분이 가지고 있는 지식과 가장 적합한 직업정보를 검색할 수 있습니다.  
 자신이 가지고 있는 지식을 가능한 많이 선택해야 정확한 직업정보를 검색할 수 있습니다.  
**최소 5개 이상의 지식을 선택하시기 바랍니다.**



### 업무수행능력별 찾기

여러분이 가지고 있는 업무수행 능력과 가장 적합한 직업정보를 검색할 수 있습니다.  
 자신이 가지고 있는 업무수행능력을 가능한 많이 선택해야 정확한 직업정보를 검색할 수 있습니다.  
**최소 7개 이상의 업무수행능력을 선택하시기 바랍니다.**



### 업무수행능력 선택

업무수행능력 선택	업무수행능력에 대한 설명
<input type="checkbox"/> 읽고 이해하기	업무와 관련된 문서를 읽고 이해한다
<input type="checkbox"/> 듣고 이해하기	다른 사람들이 말하는 것을 집중해서 듣고 상대방이 말하려는 요점을 이해하거나 적절한 질문을 한다
<input type="checkbox"/> 글쓰기	글을 통해서 다른 사람과 효과적으로 의사소통한다
<input type="checkbox"/> 말하기	자기가 알고 있는 것을 다른 사람에게 조리있게 말한다
<input type="checkbox"/> 수리력	어떤 문제를 해결하기 위해 수학을 사용한다
<input type="checkbox"/> 논리적 분석	문제를 해결하기 위해(혹은 의사결정을 하기 위해) 체계적으로 이치에 맞는 생각을 해낸다
<input type="checkbox"/> 창의력	주어진 주제나 상황에 대하여 독특하고 기발한 아이디어를 산출한다
<input type="checkbox"/> 범주화	기준이나 법칙을 정하고 그에 따라 사물이나 행위를 분류한다

## 워크넷 직업 정보 찾기

## 참고 사항

# 참고

🔗 분석 언어 및 도구



# 프로젝트 일정

[illegible]

# 팀원 소개

데이터 수집  
데이터 분석  
데이터 시각화  
연관규칙 모델 설계



조한영



정지은

데이터 수집, 전처리  
LSTM 모델 설계  
발표

데이터 수집, 전처리  
LSTM, NaiveBayes 모델 설계  
모델 보완



김재형



유진영

데이터 수집, 전처리  
TF-IDF, GRU 모델 설계  
Django 웹 구현

# Q&A

감사합니다 ➡