Job Market 데이터 분석

Data & ML Scientist

강종구

분석 절차

Q: 2019년 구직자의 특성과 합격률 그리고 지원한 회사의 이미지는?

데이터 수집 및 생성

구직자 데이터 분석

면접 후기 분석

1. 데이터 수집 및 생성

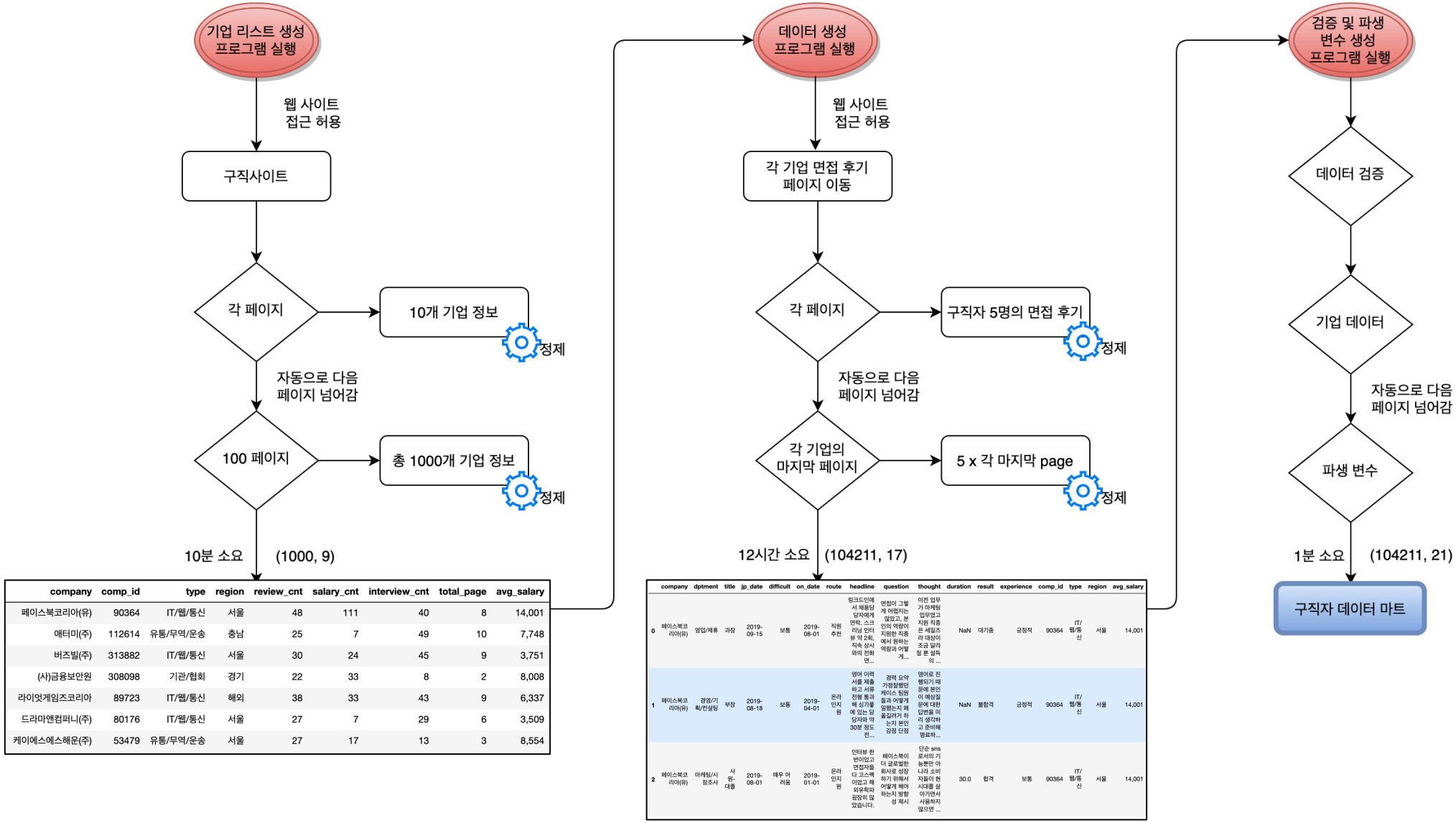
구직 사이트 선택

웹 크롤링으로 데이터 생성/정제

데이터 생성을 위한 아키텍처 설계

데이터 긁어 오기 위한 프로그램 구현

데이터 수집 및 생성



company	comp_id	type	region	review_cnt	salary_cnt	interview_cnt	total_page	avg_salary
페이스북코리아(유)	90364	IT/웹/통신	서울	48	111	40	8	14,001
애터미(주)	112614	유통/무역/운송	충남	25	7	49	10	7,748
버즈빌(주)	313882	IT/웹/통신	서울	30	24	45	9	3,751
(사)금융보안원	308098	기관/협회	경기	22	33	8	2	8,008
라이엇게임즈코리아	89723	IT/웹/통신	해외	38	33	43	9	6,337
드라마앤컴퍼니(주)	80176	IT/웹/통신	서울	27	7	29	6	3,509
케이에스에스해운(주)	53479	유통/무역/운송	서울	27	17	13	3	8,554

	company	dptment	title	jp_date	difficult	on_date	route	headline	question	thought	duration	result	experience	comp_id	type	region	avg_salary
0	페이스북코리아(유)	영업/제휴	과장	2019-09-15	보통	2019-08-01	직원 추천	링크드인에서 채용담당자에게 연락, 스크리닝 인터뷰 약 2회, 직속 상사와의 전화면...	면접이 그렇게 어렵지는 않았고, 본인의 역량이 지원한 직종에서 원하는 역량과 어떻게...	이전 업무가 마케팅 업무였고 지원 직종은 세일즈라 대상이 조금 달라질 뿐 실적의 ...	NaN	대기중	긍정적	90364	IT/웹/통신	서울	14,001
1	페이스북코리아(유)	경영/기획/컨설팅	부장	2019-08-18	보통	2019-04-01	온라인지원	영어 이력서를 제출하고 서류전형 통과해 상가를에 있는 당장와 약 30분 정도 전...	경력 요약 가정잘했던 케이스 팀원들과 어떻게 일했는지 왜 옮길려거 하는지 본인 강점 단점	영어로 진행되기 때문에 본인이 예상질문에 대한 답변을 미리 생각하고 준비해 명료하...	NaN	불합격	긍정적	90364	IT/웹/통신	서울	14,001
2	페이스북코리아(유)	마케팅/시장조사	사원-대졸	2019-08-01	매우 어려움	2019-01-01	온라인지원	인터뷰 한 번이었고 면접자들 다 고스펙이었고 해외유학파 굉장히 많았습니다.	페이스북이 다 글로벌한 회사로 성장하기 위해서 어떻게 해야 하는지 방향성 제시	단순 sns 로서의 기능뿐만 아니라 소비자들이 현 시대를 살아가면서 사용하지 않으면 ...	30.0	합격	보통	90364	IT/웹/통신	서울	14,001

2. 구직자 데이터 분석

현재

구직자의

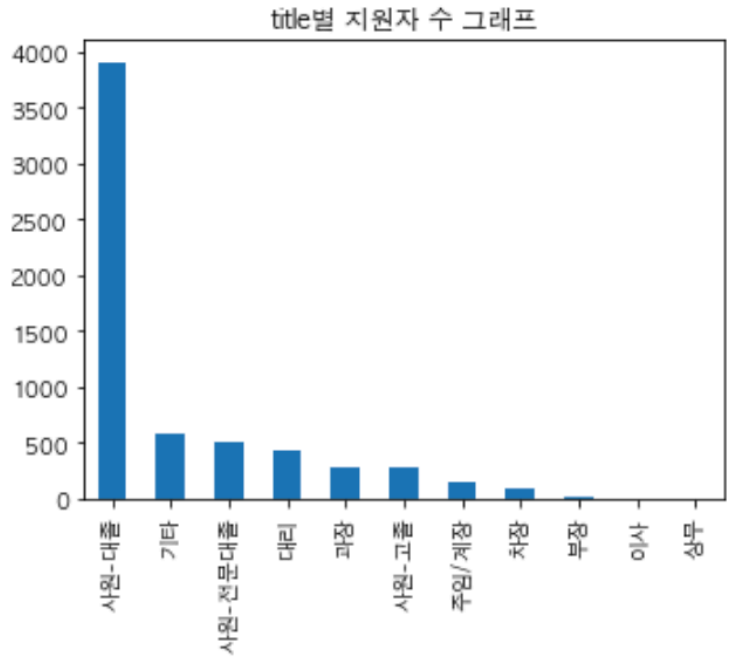
특성 시각화

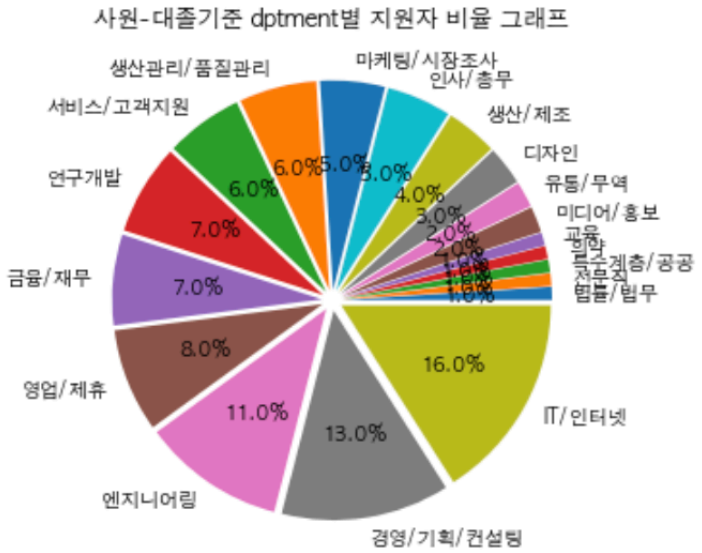
2019년

데이터만 추출

(6000개)

구직자 데이터 분석

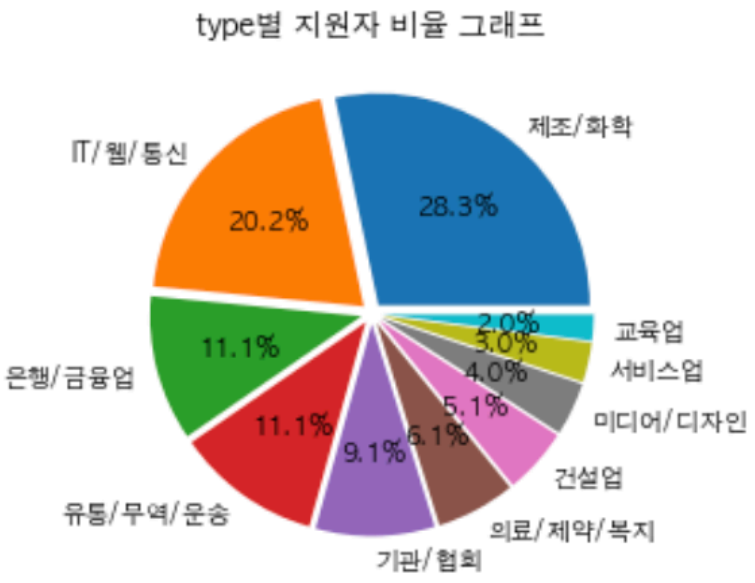


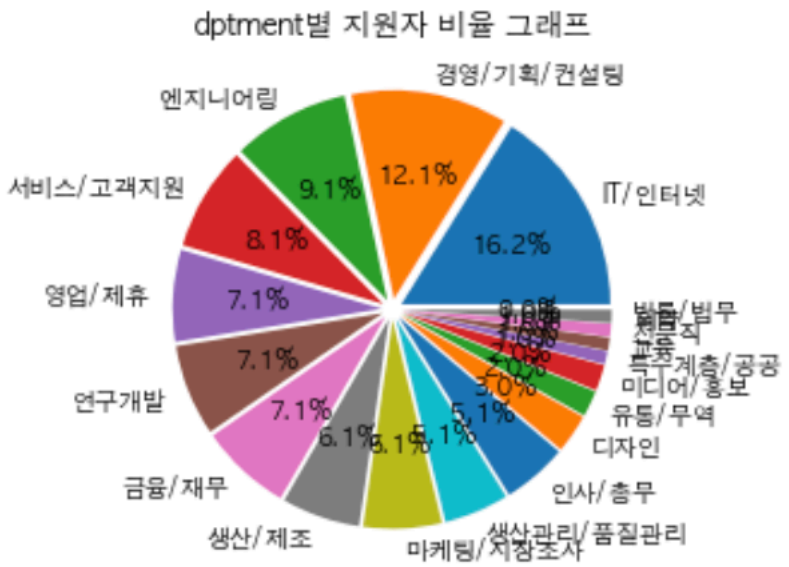


- 2019년 총 6244명의 구직자 중 대졸 출신의 사원 직급 구직자가 3908명이고 이는 2019년 총 구직자의 64%에 해당

- 2019년 총 3098명의 대졸 출신 사원 구직자는 [IT/인터넷], [엔지니어링], [경영/기획/컨설팅] 부서에 높은 지원율을 나타남

구직자 데이터 분석

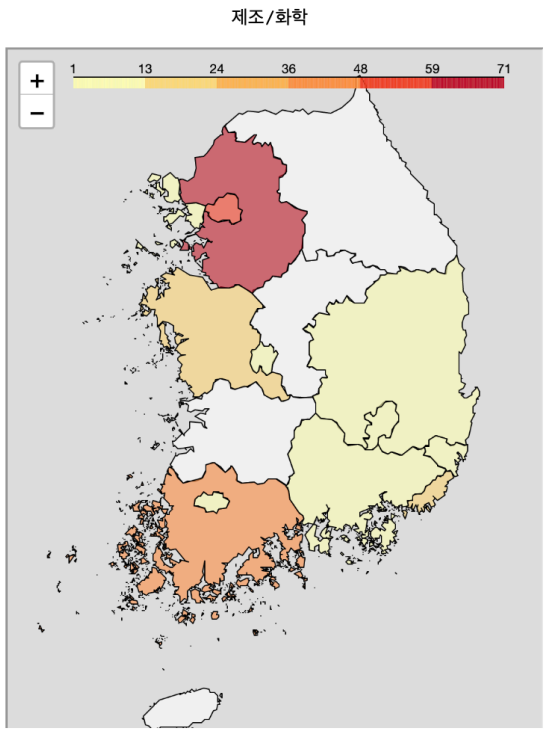




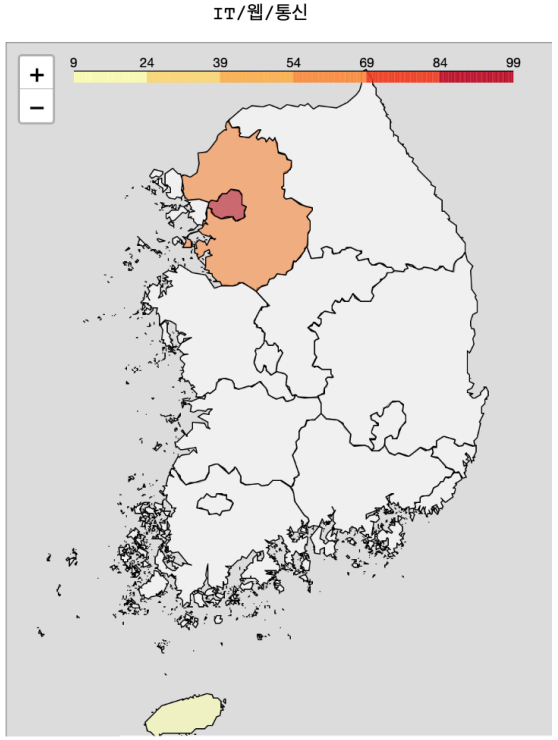
- 2019년 총 6244명의 구직자 중 1750명의 구직자는 [제조/화학] 산업과 [IT/웹/통신] 산업에 높은 지원율을 보엿으며, [교육] 산업에 가장 낮은 지원율을 보임

- 2019년 총 6244명의 구직자들의 지원 부서는 [It/인터넷]에 높은 지원율을 보였으며, [법률/법무]에 가장 낮은 지원율을 보임

구직자 데이터 분석

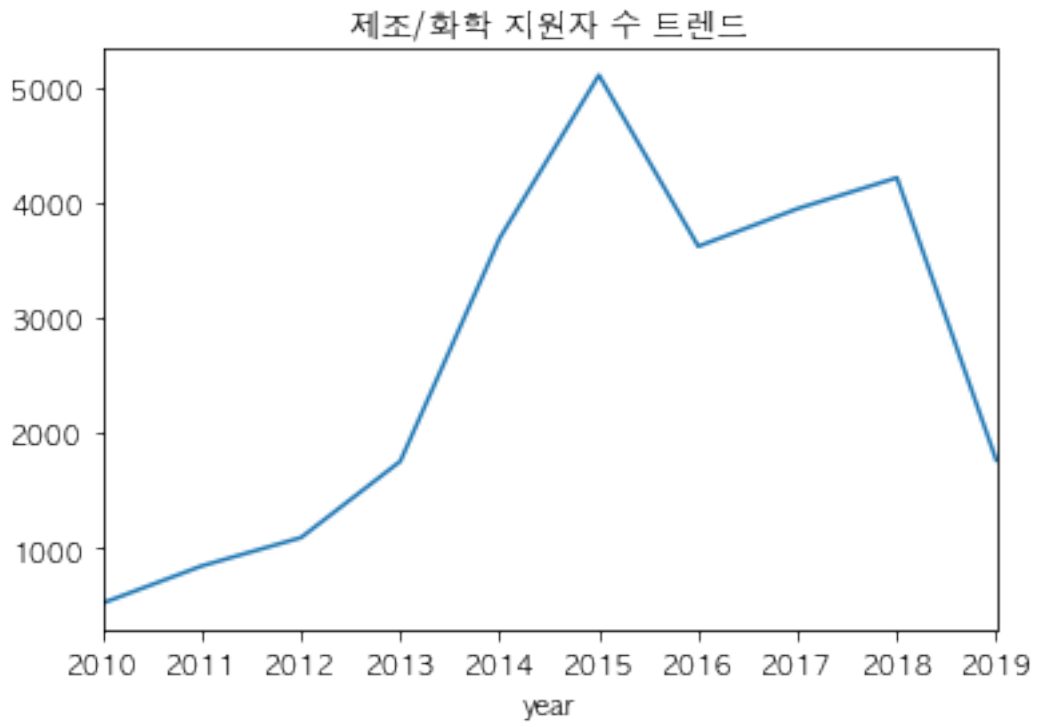


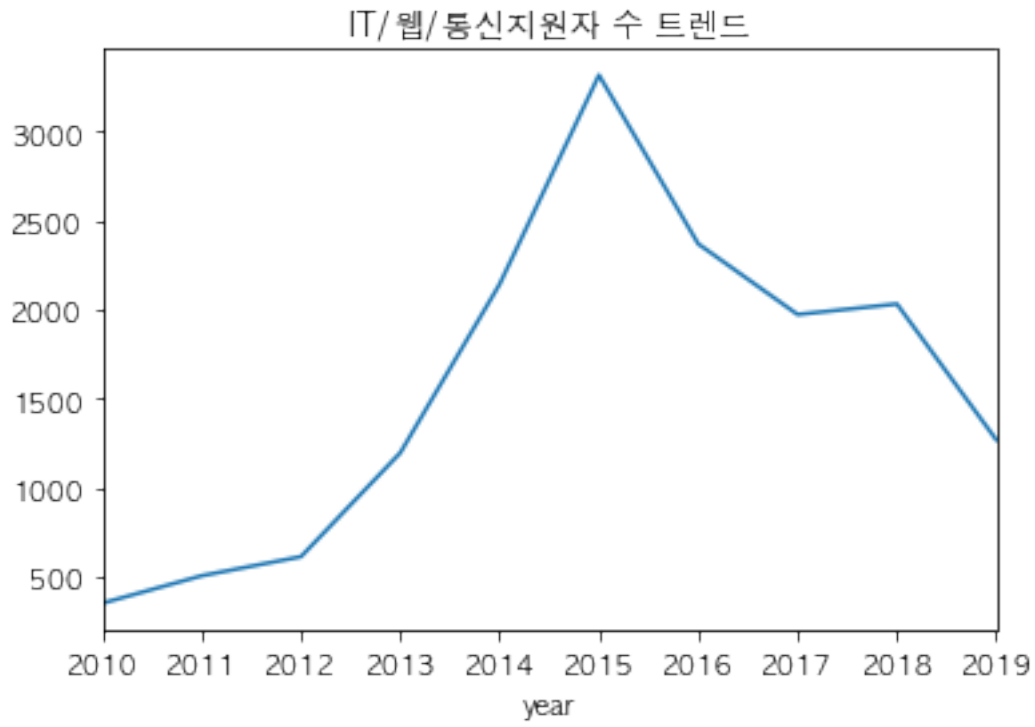
- 2019년 구직자들이 지원한 [제조/화학] 산업의 기업들은 수도권과 대기업 대표 공장이 위치한 지방에서 지원율이 보였고, 특히 경기도에서 높은 지원율이 보임



- 2019년 구직자들이 지원한 [IT/웹/통신] 산업의 기업들은 수도권과 제주 지역에 지원율이 보였고, 서울에 특히 높은 지원율이 보임

구직자 데이터 분석





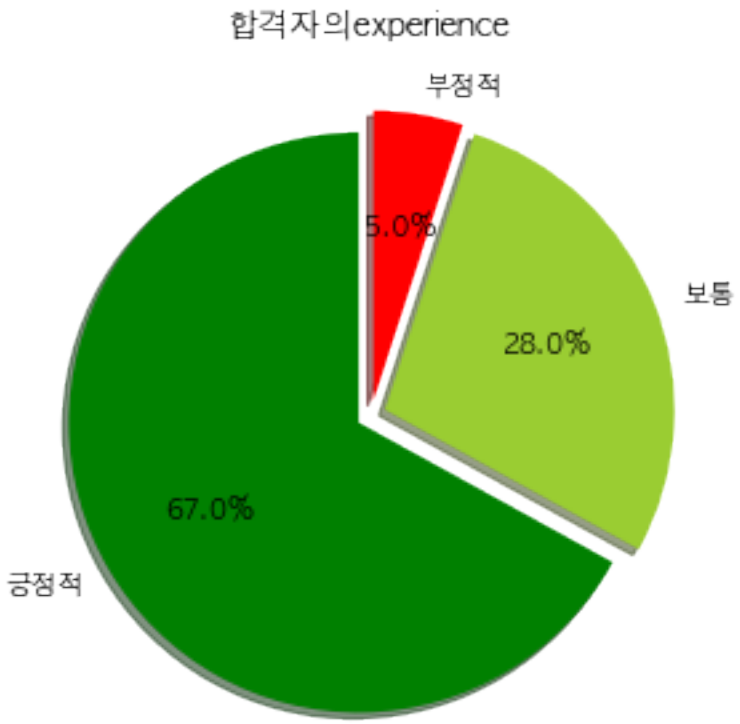
- [제조/화학] 산업의 지원자 수는 지속적으로 증가하여 2015년에 제일 많은 인원이 지원하였고, 2018년에 그 다음으로 높은 지원율을 보임
- [IT/웹/통신] 산업의 지원자 수는 지속적으로 증가하여 2015년에 제일 많은 인원이 지원하였고, 2016년에 그 다음으로 높은 지원율을 보임

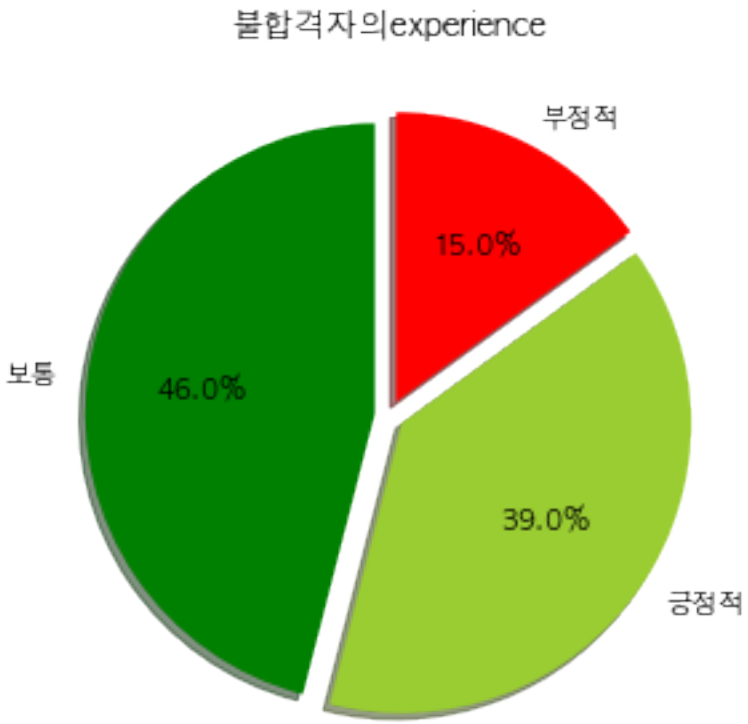
3. 면접 후기 분석

데이터 기반

머신 러닝 / 딥 러닝

구직자 데이터 분석 - 데이터 기반





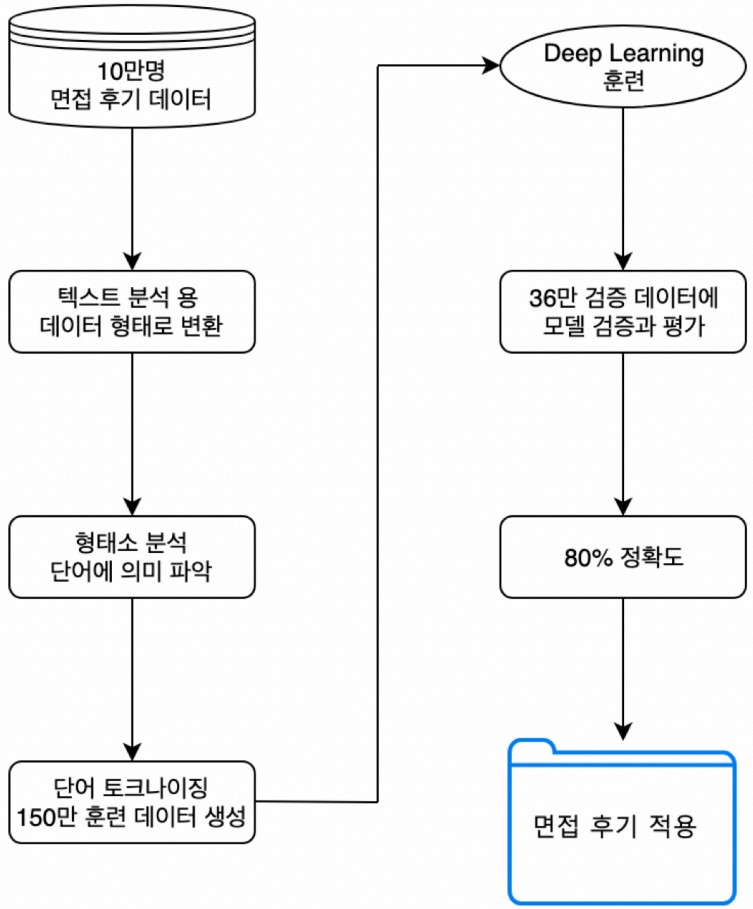
- 2019년 취직에 성공한 구직자들의 67%는 면접 과정이 긍정적이었고, 5%만이 부정적 임. 면접 과정에서 부정적인 느낌이 적을 때 취직 성공이 크다.

- 2019년 취직일 하지 못한 구직자들의 39%는 면접 과정이 긍정적이었고, 15%는 면접 과정이 부정적이었다. 긍정적인 느낌이 성공자의 반으로 떨어진 것을 확인

구직자 데이터 분석 - 데이터 기반

- 80%의 정확도 Deep Learning 모델 생성
- 모델이 면접 후기 텍스트 읽어 드림
- 텍스트가 긍정인지 부정인지 모델이 산출
- 면접자가 직접 밝힌 긍정/부정/보통 만족도와 비교 검증

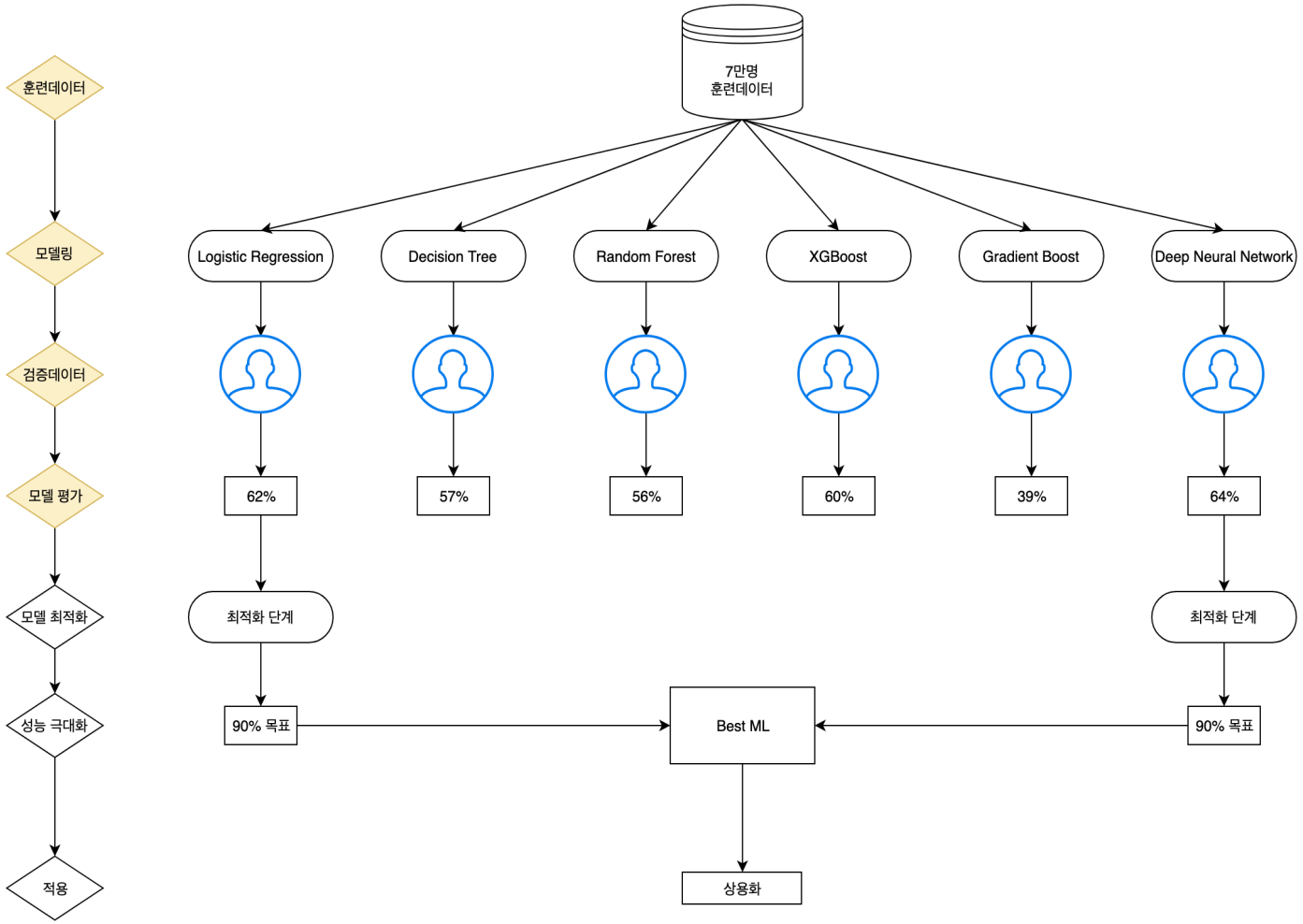
```
Real: 부정적 Real Value
ML says: 부정 23.46 %
  뭐 아는대로 대답했긴 했지만 썩 표정이 안좋았다.

Real: 보통 Real Value
ML says: 긍정 96.06 %
  기술했던 경력 중 최초 아이디어부터 진행까지 모든 부분에 인볼브했던 경험를 자세하게 답변

Real: 부정적 Real Value
ML says: 긍정 93.66 %
  생각을 솔직하게 이야기했습니다면접 질문에 정답은 없다고는 하지만,항상 정답은 있는 기분입니다.그 정답을 찾기 위해 고민했습니다

Real: 보통 Real Value
ML says: 긍정 98.86 %
  중간중간 짜투리 시간을 잘 활용하여 효율적으로 일하겠다고 밝힘

Real: 보통 Real Value
ML says: 긍정 99.8 %
  그냥 질문에 대한 적절한 답변들을 했던 것 같음 분위기는 나쁘지 않았음. 면접관님이 하신 말씀 중에.. 회사가 너무 좋아서 하루종일 회사에서 머물고 상주하며, 같이 일하고 야근하는게 일상같고 행복해서 집에 가고싶지가 않다면 이회사는 평점이 5점만점이어야 하는데 왜 .0밖에 안될까요? .. 너무 가식적이어도 보기 좋지 않습니다.

Real: 부정적 Real Value
ML says: 부정 99.99 %
  실무진 태도가 매우 불량. 기분나쁘게 사람 훑고 비웃기를 반복함. 노트북 화면 보면서 타자 치는데, 표정이나 눈치가 채팅하는 것 처럼 보였음. 각종 후기에서 봤던 안좋은 인상들을 나역시 첫만남에 확인할 수 있었음
```



ML / DL Application

- ML / DL 활용
  - 합격률 높은 지원자 구분

ML/DL application



- 웹 크롤링 데이터를 기반으로 합격/불합격 분류 모델 생성
- 현재 정확도 60% 전후의 정확도
- 모델 최적화 단계 실시 시 성능을 90% 이상 올릴 수 있음 (시간 비용이 큼)
- 완성 시 합격률이 높은 지원자를 구분하여 먼저 면접 진행 가능