

# 입 사 지 원 서

지원분야

데이터 분석

희망연봉

회사내규에따름

## 인적사항



이 름	김 정 호	영 문	Kim Jung Ho
생 년 월 일	1998 년 05 월 22 일 ( 25 세)		
휴대전화	010-6515-2292	E-mail : jungho2292@gmail.com	
주 소	서울시 도봉구 도봉로 104 길 79 103 동 402 호		

## 병역사항 (해당자 기재)

군필 여부	면제	면제사유	건강상의 이유 (일상생활에 지장 없음)
-------	----	------	-----------------------

## 학력사항

재학기간	학교명	전공	학점	구분
2017.09 ~ 2021.07	북경과학기술 대학교	국제경제무역 학과	3.1 / 4.0	졸업
2014.03 ~ 2017.02	선덕 고등학교			

## 경력사항 (경력보유자 기재)

근무기간	회사명 및 부서	직위	담당 업무
2020.06~2020.07	Two hands games (챔피언 스트라이크 게임 부서)	인턴	리서치, 시장조사
2019.06~2020.08	iSoftStone (중국 / 데이터 전처리 부서)	인턴	데이터 수집, 데이터 전처리

## 교육 및 연수사항 (해외연수포함)

기간	교육과정	교육기관	교육내용
2021.11~2022.05	빅데이터 플랫폼 구축(하둡)과 빅데이터 분석(by spark & power Bi) & 자바 웹 개발	더조은 아이티아카데미	- Spring Framework 기반 요구사항 분석 및 설계부터 배포까지 전반적인 과정 이해 - UI 개발 및 데이터 베이스 기술 - R, Python 활용한 빅데이터 분석 및 시각화 - 팀 프로젝트간 협업 방법 - AI 데이터 수집, 가공부터 시각화 및 기대효과까지 전반적인 과정이해
2022.07.04~2022.08.26	LG Aimers : AI 전문가과정	LG Aimers LG AI 연구원	- 해당 과정에서 진행되는 해커톤 문제를 풀기 위한 도메인 지식 (자율주행레이더 공정 과정) - AI 기술 전반에 대한 이해와 함께 최신 트렌드를 반영한 다양한 AI 기술

			- LG 이노텍의 실제 현장 DATA 를 활용한 데이터분석
2022.09.12~2022.10.20	서울 시민 데이터를 활용한 도시문제 해결 경진대회	서울 디지털 재단, DAICON	- 파이썬 (Python)을 활용한 공공데이터 전처리 과정 - 공공데이터 시각화 과정 - 공공데이터 머신러닝

☐ 위의 모든 기재사항은 사실과 다름 없음을 확인합니다.

작성자 : 김정호

□ OA 능력 및 보유기술

한글/MS 워드	파워포인트	엑 셀	인터넷
즉시 업무가능 수준	즉시 업무가능 수준	기본 업무 가능	즉시 업무가능 수준
상세 보유 기술			
분야	기술내역	설명	수준
Web	HTML/CSS	-디자인에 맞게 UI/UX 개발 가능 -UI 퍼블리싱 가능 -오픈 소스 선택 활용가능	중
	JavaScript jQuery	-동적 프로세스 구현 가능 -UX 에 대한 동적 구현 가능	
	Java	-웹 표준 개발 가능	
	JSP/Ajax	-서버 연동 및 서비스 구축 가능 -비동기식, 동기식 연동 가능	
	Spring	-MVC패턴 이해 및 구현 가능 -스프링기반 웹 개발 가능	
데이터베이스	My SQL	-My SQL 기본적인 기능 사용 가능 -ERD 표현 가능	중
서버 (웹 배포)	Linux Oracle Cloud Putty FileZilla	-서버 운영 및 관리 -서버 내 파일 확인, 경로 수정 가능 -Oracle Cloud를 활용한 웹 배포 가능	중
빅데이터 분석	R Python Jupyter notebook	-R studio IDE이용한 데이터 분석 환경 구축 가능 -통계 분석을 이용한 가설 검정 가능 -패키지 이용한 데이터 가공/정제/분석 가능 -데이터 마이닝/시각화 가능	중
	Scikit-learn TensorFlow	-Anaconda 기반 개발 환경 구축 가능 -데이터 읽고 쓰고 문제점 도출 가능 -패키지 및 라이브러리 활용 가능 -시각화 및 디자인 가능 -웹 동적 크롤링(Crawling) 가능	
		머신 러닝, 딥러닝 알고리즘 적용 가능	

## □ 수행 프로젝트

포트폴리오 주소

<http://146.56.44.116:8080>

프로젝트 명	머신러닝과 딥러닝(LSTM)을 활용한 전력수요 예측
수행기간	2021.04.14 ~ 2021.05.23
수행인원	5 명
프로젝트 목표/소개	코로나 엔데믹이 다가오면서 전력수요가 코로나 이전 시절로 돌아올 것이라 예상하였다. 코로나 이전의 데이터들을 활용하여 앞으로의 전력 수요량을 분석하고 예측하는 웹 페이지를 제공하고자 하였다.
개발환경 (사용도구/언어)	<p>데이터 분석 : python (사용 도구 : jupyter notebook, scikit-learn, pandas, numpy, seaborn, matplotlib, TensorFlow)</p> <p>웹 개발 : java (사용 도구 : Eclipse, JS, Maven, My Batis, jQuery, JSP, Spring boot, My SQL)</p> <p>웹 배포 : Linux (사용 도구 : Oracle Cloud, FileZilla, Apache Tomcat, PuTTY)</p>
담당역할	<p>데이터 분석 및 웹 배포 (데이터 분석 : 데이터 정의서 작성, 데이터 수집, 데이터 전처리, 데이터 분석 및 시각화)</p> <p>(웹 배포 : Oracle cloud 를 활용한 web server 구축)</p>
수행 결과물	<div> <div> <h2>02 데이터 분석</h2> <ol style="list-style-type: none"> <li>데이터 정의서</li> <li>데이터 수집</li> <li>데이터 전처리</li> <li>데이터 분석 및 시각화</li> </ol> </div> <div> <h2>데이터 전처리 – 2<sup>nd</sup>, 3<sup>rd</sup>, Final LSTM</h2> <h3>2) 중복값 제거</h3> <h4>03) 중복값 제거</h4> <div> <pre> 1 df_dup = df[df.duplicated(subset='date')] 2 df_dup = df.drop_duplicates(subset='date') </pre> </div> <div> <pre> 1 df_dup = df.drop_duplicates(subset='date') 2 df_dup = df.drop_duplicates(subset='date') 3 df_dup = df </pre> </div> </div> </div> <div>38</div> <p>화면/기능설명: 데이터 수집 이후 전처리 과정 중 일부만</p>

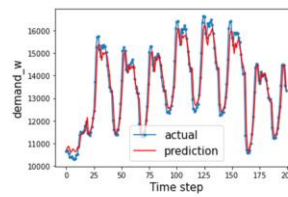
## 02 데이터 분석

- 01. 데이터 정의서
- 02. 데이터 수집
- 03. 데이터 전처리
- 04. 데이터 분석  
및 시각화

73

## 데이터 분석 및 시각화 – Final LSTM

### 2) 데이터 분석 - n\_train\_time 변경한 모델의 결과



```
In [26]: 1 model.summary()
```

Model: "sequential"		
Layer (type)	Output Shape	Param #
-----		
lstm (LSTM)	(None, 100)	64400
dropout (Dropout)	(None, 100)	0
dense (Dense)	(None, 1)	101
-----		
Total params: 64,501		
Trainable params: 64,501		
Non-trainable params: 0		

화면/기능설명: 전처리 및 스케일링 후 모델 Fitting 결과 중 일부분

□ 수행 프로젝트

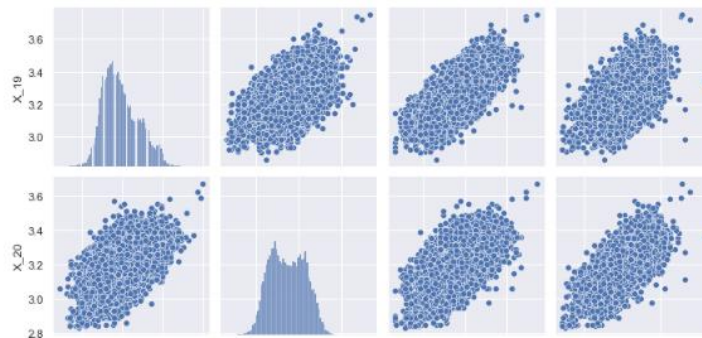
프로젝트 명	자율주행 센서의 안테나 성능 예측 AI 경진대회
수행기간	2022.08.14 ~ 2022.08.26
수행인원	4 명
프로젝트 목표/소개	공정 데이터를 활용하여 Radar 센서의 안테나 성능 예측을 위한 AI 모델 개발
개발환경 (사용도구/언어)	데이터 분석 : python (사용 도구 : jupyter notebook, scikit-learn, pandas, numpy, seaborn, matplotlib) (사용 알고리즘 : Regressor(xgb, lgbm, gb)
담당역할	데이터 분석 (데이터 분석 : EDA 확인, 데이터 전처리, 데이터 분석)

## 수행 결과물

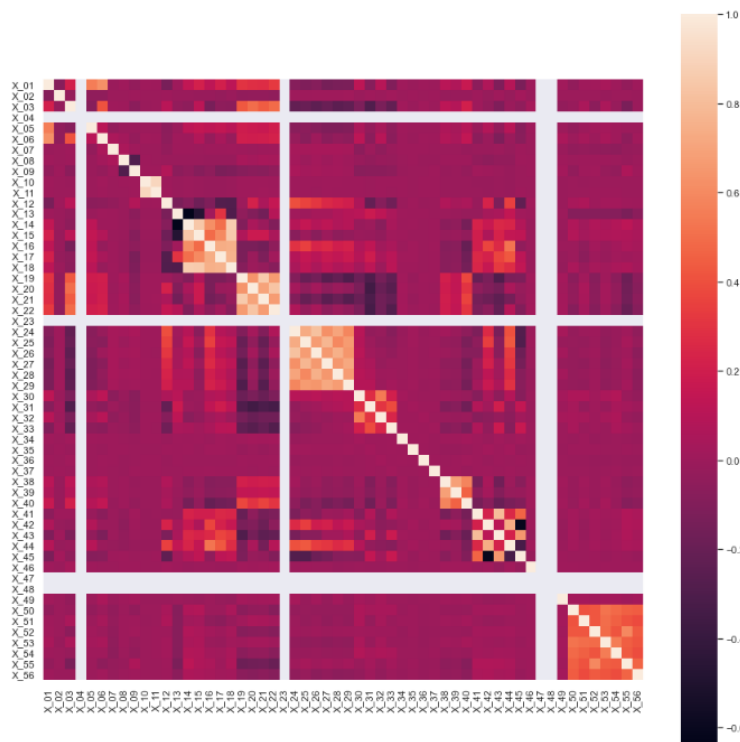
```
In [45]: 1 # Group_8 : N번 스크류 삽입 깊이
2 g8_df = train_x[["X_19", "X_20", "X_21", "X_22"]]
3
4 cols = range(0,4)
5 i = 1
6 groups = cols
7 values = g8_df.values
8
9 plt.figure(figsize = (30,30))
10 for group in groups :
11     plt.subplot(len(cols),1,i)
12     plt.hist(values[:,group], bins=200)
13     plt.title(g8_df.columns[group], y=0.75, loc="right")
14     i+=1
15 plt.show()
```

```
In [46]: 1 #scatterplot
2 sns.set()
3 sns.pairplot(g8_df, size = 2.5)
4 plt.show();
```

C:\Users\user\anaconda3\lib\site-packages\seaborn\axisgrid.py:2078: UserWarning: The 'size' parameter has been renamed to 'height'; please update your code.  
warnings.warn(msg, UserWarning)



```
In [31]: 1 #correlation matrix
2 cormat = X.corr()
3 f, ax = plt.subplots(figsize=(15, 15))
4 sns.heatmap(cormat, vmax=1, square=True);
```



화면/기능설명: EDA 확인 다중공선성 확인

```

In [:] 1 # Group_2 방열 재료 무게 (X_10, X_11) drop
2 train_X = train_X.drop(columns=['X_10', 'X_11'])
3
4 # Group_3 통과 여부 (X_04, X_23, X_47, X_48) drop
5 train_X = train_X.drop(columns=['X_04', 'X_23', 'X_47', 'X_48'])
6
7 # Group_4 방열 재료 면적 (X_09)
8 # 9번 이상치 평균으로 바꾸기
9 train_x09 = train_X['X_09']
10 idx = train_X[(abs((train_x09-np.mean(train_x09))/np.std(train_x09)))>1.96].index
11 train_x09[idx] = 0
12 med_score = train_x09[train_x09 != 0].mean()
13 train_x09[idx] = med_score
14 train_X['X_09'] = train_x09
15
16 # Group_7 안테나 패드 위치
17 train_X = train_X.drop(columns=['X_14', 'X_15', 'X_16', 'X_17', 'X_18'])
18
19 # Group_10 스크류 삽입 깊이
20 train_X = train_X.drop(columns=['X_30', 'X_31', 'X_32', 'X_33'])
21
22 # Group_11 스크류 체결 시 분당 회전수
23 train_X = train_X.drop(columns=['X_34', 'X_35', 'X_36', 'X_37'])

```

```

In [51]: 1 with tf.device("/device:GPU:0"):
2         score = rmsle_cv(model_gb)
3         gb_score = score.mean()
4         print("GradientBoostingRegressor score: {:.4f} ({:.4f})".format(score.mean(), score.std()))

```

GradientBoostingRegressor score: 1.2089 (0.1305)

```

In [94]: 1 with tf.device("/device:GPU:0"):
2         score = rmsle_cv(model_xgb)
3         xgb_score = score.mean()
4         print("XGBRegressor score: {:.4f} ({:.4f})".format(score.mean(), score.std()))

```

...

```

In [90]: 1 with tf.device("/device:GPU:0"):
2         score = rmsle_cv(model_lgbm)
3         lgbm_score = score.mean()
4         print("LGBMRegressor score: {:.4f} ({:.4f})".format(score.mean(), score.std()))

```

...

```

In [102]: 1 total_weight = (1. / xgb_score) + (1. / gb_score) + (1. / lgbm_score)
2 preds = (pred_xgb * (1. / xgb_score) + pred_gb * (1. / gb_score) + pred_lgbm * (1. / lgbm_score)) / total_weight

```

```

In [103]: 1 submit = pd.read_csv('./sample_submission.csv')

```

```

In [104]: 1 for idx, col in enumerate(submit.columns):
2         if col=='ID':
3             continue
4         submit[col] = preds[:,idx-1]
5         print('Done.')

```

Done.

```

In [105]: 1 submit.to_csv('./ensemble_13(gb+xgb+lgbm)_test2.csv', index=False)

```

화면/기능설명: 데이터 전처리 과정 및 모델 학습 및 앙상블 기법 사용 후 제출용 파일로 변환



## □ 수행 프로젝트

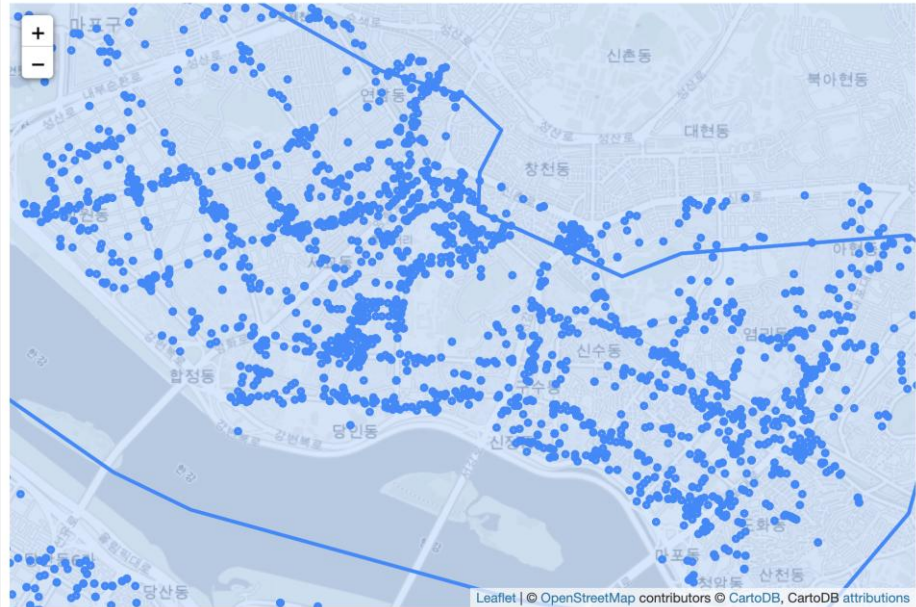
프로젝트 명	시민 안전과 전동 킥보드 수거 효율화를 위한 불법 주정차 킥보드 위치 데이터 분석
수행기간	2022.09.12~2022.10.20
수행인원	3 명
프로젝트 목표/소개	<ul style="list-style-type: none"> <li>- 공공 데이터를 활용하여 서울 도시 문제 해결 방안 제시</li> <li>- 시민 안전과 수거 효율화를 위한 불법 주정차 전동 킥보드 위치 데이터 분석</li> </ul>
개발환경 (사용도구/언어)	<p>데이터 분석 : python</p> <p>(사용 도구 : jupyter notebook, scikit-learn, pandas, numpy, seaborn, matplotlib, folium)</p> <p>(사용 알고리즘 : Regressor (Random Forest, Decision Tree, Linear)</p>
담당역할	데이터 수집, 전처리 및 시각화
수행 결과물	<div> <pre> In [1]: 1 from selenium import webdriver  In [1]: 1 # 크롬드라이버 실행 (경로 예: "/Users/Ray/Downloads/chromedriver") 2 driver = webdriver.Safari() 3 4 #크롬 드라이버에 url 주소 넣고 실행 5 driver.get('https://address.dawul.co.kr/') 6 7 # 페이지가 완전히 로딩되도록 3초동안 기다림 8 time.sleep(2) 9 10 driver.switch_to.window(driver.window_handles[0]) # 메인 윈도우 가기  In [1]: 1 from tqdm import tqdm, tqdm 2 3 lat = [] # 위도 4 log = [] # 경도 5 6 for num in range(len(songpa_1)): 7     search_box = driver.find_element(By.XPATH, r'//*[@id="input_juso"]') 8     search_box.send_keys(songpa_1["주소"][num]) 9     search_box.send_keys(Keys.RETURN) 10    for j in range(len(songpa_1["주소"][num])): 11        search_box.send_keys(Keys.BACK_SPACE) 12        time.sleep(1) 13        result = driver.find_elements(By.XPATH, r'//*[@id="insert_data_5"]') 14 15    for i in result: 16        title = i.text 17        res = re.split('[: .]', title) 18        lat.append(res[1]) 19        log.append(res[2]) </pre> <p>화면/기능설명 : 데이터 Crawling</p> <pre> In [16]: 1 # 구청별대량 그룹화 2 adr = board.groupby("구명").count().sort_values("주소", ascending = False)[["주소"]]  In [17]: 1 plt.figure(figsize = (15, 8)) 2 sns.barplot(y = adr.index, x = adr["주소"]) 3 plt.show() </pre> <p>화면/기능설명 : 데이터 전처리 과정 및 데이터 시각화</p> </div>

```

In [107]: 1 # 서울 행정구역 json raw파일(githubcontent)
          2 r = requests.get('https://raw.githubusercontent.com/southkorea/seoul-maps/master/kostat/2013/json/seoul_municipalities_geo_simple.json')
          3 c = r.content
          4 seoul_geo = json.loads(c)

In [108]: 1 fm = folium.Map(location=[lat, long], zoom_start=12, tiles='cartodbpositron',
          2                   width = 750, height = 500)
          3
          4 folium.GeoJson(
          5     seoul_geo,
          6     name='시각화'
          7 ).add_to(fm)
          8
          9 for _, rows in df.iterrows():
          10     folium.CircleMarker(location = [rows['위도'], rows['경도']],
          11                        tooltip = rows['명칭'],
          12                        fill = True, radius = 2).add_to(fm)
          13
          14 fm

```



화면/기능설명 : Folium 을 활용한 위도, 경도 데이터 시각화

# 자 기 소 개 서

직무 역량	지원분야/회사에 지원하는 지원동기는?
	<p>대학 시절 수업에서 처음 접한 '빅데이터' 분야에 흥미를 갖게 되어, 전공과 관련된 일보다는 빅데이터와 관련된 분야에서 일을 하고 싶었습니다. 졸업을 위해 인턴생활을 해야 할 때, 데이터를 다루는 회사에서 데이터 전처리를 하는 인턴 경험을 가졌습니다. 졸업 이후에는 국비 지원 수업을 찾아 참여하며 이 분야에 대해 지식을 더 쌓고자 노력하였습니다. 빅데이터에 대해 공부를 하다 보니 데이터 내에서 중요한 정보들을 찾아, 그것을 가공하여 더 효율적으로 활용하는 방식에 관심이 많아졌습니다. 그때 이런 생각을 하였습니다.</p> <p>"나중에 내가 일하고 싶은 회사에서 내가 하고 싶어 하는 데이터 분석의 일을 한번 해보자" 하는 다짐을 하게 되었습니다. 이후, 저의 다짐과 일치하는 일을 하고 싶었으며, 알맞은 기업을 찾기 위해 구직 중입니다.</p>
	지원분야에서 활용할 수 있는 직무 수행 경험은? (보유기술이 어필될 수 있도록 상세 경험 작성)
	<p>국비 지원 수업을 들을 때, LSTM 시계열 모형 알고리즘을 활용한 '전력 수요 예측'이란 주제를 가진 최종 프로젝트를 진행하였습니다. 과거 전력 데이터와 날씨 데이터들을 토대로 1년 치 데이터를 예측하는 TensorFlow의 LSTM 모델을 만들어, 웹을 제작하여 배포하였습니다.</p> <p>LG Aimers 가 주관하는 자율주행 센서의 안테나 성능 예측 AI 경진대회에 참여하였습니다. 공정 데이터와 제품 성능간 상관 분석을 통해 제품의 불량률 예측/분석하는 AI 모델을 설계했습니다. 해당 대회에서 상위 15% 안에 진입하였습니다. 해당 대회에서 장려상(4위)을 수상하였습니다.</p> <p>서울 디지털 재단에서 주관하는 서울 시민 데이터를 활용한 도시문제 해결 경진대회에 참여하였습니다. 다양한 공공 데이터를 활용하여 서울 시민 안전과 전동 킥보드 수거 효율화를 위한 불법 주정차 킥보드 위치 데이터 분석을 진행하였습니다.</p> <p>이런 경험을 바탕으로 귀사에 보탬이 될 수 있는 인재가 되겠습니다.</p>
	입사 후 포부를 제시하고, 그에 따른 노력을 구체적으로 작성하십시오.
	<p>조직원은 본인의 직무에 최선을 다하며, 끊임없는 노력과 자기 개발을 통해 회사와 함께 성장해야 한다고 생각합니다. 업무에 관련된 전반적인 지식을 배우고, 이해하는 데 있어서 적극적이고 주도적인 자세로 임하겠습니다. 모든 업무에 최선을 다해 자신의 가치를 높여, 나아가서는 귀사의 가치를 높이는데 귀중한 한몫을 해내고 싶습니다.</p> <p>빠르게 변하는 IT 분야에 걸맞게 시대의 트렌드를 놓치지 않고 다른 기업들과 경쟁이 더욱 치열해질 미래를 대비하여 많은 경험과 공부를 통해 꾸준히 발전하도록 노력할 것입니다.</p>
	어렵거나 힘들었던 문제의 상황을 설명하고 어떠한 과정을 거쳐 해결하였는지 설명하십시오.

인성 역량	<p>국비 지원 수업 수료를 위해, 데이터 분석과 관련된 최종 프로젝트를 진행할 때였습니다. 처음 보는 알고리즘들을 접하고, 머신러닝뿐 아니라 딥러닝 모델을 접하면서 어느 모델이 저희 프로젝트에 알맞을지 결정하는 것은 결코 쉬운 일이 아니었습니다. 프로젝트 초기부터 어려움을 맞닥트린 저희 조를 본 강사님은 딥러닝 모델은 어려우니 다른 쉬운 주제를 찾으라 권유했습니다. 오히려 이 권유가 자극이 되어 수업 시간 이후에도 조원들을 모아 알고리즘에 대해 공부하고, 끈기 있게 노력한 결과, 딥 러닝 모델을 활용한 최종 프로젝트를 성공적으로 완성할 수 있었습니다.</p> <p>목표한 바를 이루기 위해 포기하지 않는 강한 인내심과 끈기는 사회 진출 후 저에게 주어지는 어떠한 문제라도 해결할 수 있는 큰 힘이 되어 주리라 믿어 의심치 않습니다.</p>
	<p><b>성격의 장/단점을 보여줄 만한 구체적인 사례를 작성하십시오. (단점 작성 시, 보완노력 포함)</b></p>
	<p>저의 장점은 강한 책임감과 유연한 사고입니다. 선후배나 주변 동료들의 조언을 잘 새겨들어 고칠 줄 알고, 새로운 것을 잘 수용하는 오픈 마인드를 가지고 있습니다. 이러한 성격 때문에 학창 시절부터 종종 임원직과 조별 과제 조장을 자주 맡았습니다. 다른 사람과의 소통이 원활하고, 여러 의견을 잘 수용하여 융합할 수 있어서 어느 모임에서도 중심적인 역할을 맡았습니다.</p> <p>하지만, 이러한 성격 때문에 선택을 내릴 때 너무 신중한 것이 단점입니다. 선택을 내리기로 결정했어도, 더 나은 선택지가 나오게 된다면 선택을 미루게 됩니다. 선택에 따른 책임의 중압감을 이겨 내기 위해 수많은 시행착오를 겪었습니다. 간단한 조별 과제뿐 아니라 학생회 임원, 프로젝트 조장을 맡으면서 많은 성공과 실패를 겪었습니다. 리더로서 선택에 대한 결과들을 보며, 성공을 통해 겸손과 경험을 얻었고, 실패를 반면교사로 삼아 더 나은 사람이 되기 위해 꾸준히 노력했습니다.</p>