

# Sungbin Cho

DATA SCIENTIST · RESEARCHER

18-7 Room301, Wausanro23-Gil, Seogyo-Dong, Mapo-Gu, Seoul, Rep. of KOREA

☎ (+82) 10-7530-1211 | ✉ kadensungbincho@gmail.com | 📷 kadensungbincho

"All life is problem solving"

## Summary

Problem solver, breakthrough promoter constructing efficient analysis system. Expertise in Python-based analysis modules(Pandas, Scikit-Learn) with DB(SQL, mongo DB) and Web(Flask, Basic html, php) Knowledge. 2 Years experience as a food-related researcher. I love to make something useful with coding.

## Work Experience

### Jardin Corp.

DEVELOPER & RESEARCHER

Seoul, S.Korea

Aug. 2015 - Jul. 2017

- Develop powder and liquid drink product for B2B & B2C market
- Control and devise process & factory environment for new products
- Documents management of product formula, process description, regulation
- Regular market research and presentation

## Projects

### Seoul Commercial Real-Estate Market Analysis

Seoul, S.Korea

FOCUSED ON API TO WEB SERVICE DEVELOPMENT PROCESS

Dec. 2017 - Present

- Develop Seoul Market Cluserter & Comparison Service
- Building scripts A-Z backend which could be automated just with scheduler
- Using Gov's Franchise list, Naver API, Mongo DB, HDBScan Clustering Algorithm, GeoPandas & Shapely & Folium for mapping, and AWS & Flask with nginx, wsgi
- <http://ec2-13-125-106-155.ap-northeast-2.compute.amazonaws.com/>

### Leaf Classification

Seoul, S.Korea

KAGGLE COMPETITION - CLASSIFICATION OF LEAF SPECIES WITH IMAGE, NUMERIC DATASET

Nov. 2017

- Score logloss of '0.02296' which ranks about 258/1598 teams (Late Submission)
- Using linear regression for core prediction algorithm
- Using python's pandas, scikit-learn packages for data handling and OpenCV for image processing(image information extraction)
- [https://github.com/kadensungbincho/Leaf\\_Classification\\_20171101\\_20171122](https://github.com/kadensungbincho/Leaf_Classification_20171101_20171122)

### Sberbank Russian Housing Market

Seoul, S.Korea

KAGGLE COMPETITION - BUILDING PREDICTIVE MODEL FOR RUSSIAN HOUSING MARKET

Oct. 2017

- Score rmse of '0.5062' which ranks about 3078/3274 teams (Late Submission)
- Using self-made algorithm which is based on Pearson's Correlation & P-value
- Using python's pandas, scikit-learn, statsmodels for data handling
- [https://github.com/kadensungbincho/Sberbank\\_Russian\\_Housing\\_Market\\_20170928\\_20171025](https://github.com/kadensungbincho/Sberbank_Russian_Housing_Market_20170928_20171025)

## TechSkill

### Backend & Analysis & Frontend

DATABASE, VERSION CONTROL, AND ANALYSIS TOOLS TO WEB

- SQL, MongoDB, AWS, Git, Github
- Python & Python packages Pandas, Scikit-Learn, Numpy, OpenCV, Shapely, Folium
- Flask, Basic understanding of HTML, CSS, PHP, Django

## Education

### Cyber University of Korea

Seoul, S.Korea

B.S. IN SOFTWARE ENGINEERING

Mar. 2017 - Present

- Specialized in software developer curriculum

## Fastcampus

DATA SCIENCE SCHOOL

- Python-based data science school covering basic python, statistics, modules for analysis, machine & deep-learning

Seoul, S.Korea

Aug. 2017 - Exp. Dec. 2017

## KyungHee University of Seoul

B.S. & B.B.A. IN FOOD ENGINEERING & FOOD SERVICE MANAGEMENT

- Got double degrees and one semester exchange course on Finland Seinäjoki University of Applied Sciences

Seoul, S.Korea

Mar. 2009 - Exp. Aug. 2015

## Others

---

### Certificates

LANGUAGES & COMPUTER RELATED

- English : Opic (AL), TOEIC (880)
- Craftsman Information Processing (written), Survey Analyst Junior (written)

### Writing : Flow - 1pagebookreview

FOUNDER & WRITER

- Phrase collection of books about economics, science, management, etc
- <https://www.facebook.com/1pagebookreview>

Facebook Page

Mar. 2013 - PRESENT

# 자기소개서

---

## 1) 데이터 사이언스 / 엔지니어링 직무를 지원하게 된 동기는 무엇인가요?

식품연구원으로 근무하는 동안 수많은 서류를 관리하며 정보관리에 점점 빠져들게 되었습니다. 원료, 공정, 법규 등 수많은 문서에서 빠르게 제가 원하는 정보를 찾아 보고서를 작성하기 위해 저는 많은 시간 '정보정리'에 대해 고민하였습니다. 한번은 잦은 카페인 함량 요청에 업무가 중단되는 경우가 많았습니다. 저는 그로 인해 업무 효율이 떨어지는 것을 막기 위해 업무 후 남아 원료에서 카페인 함량을 자동으로 구하는 엑셀 시트를 만들어 놓았습니다. 간단한 수식이었지만 그러한 개선이 저는 물론 팀원까지 편의까지 가져주었습니다.

그렇게 저는 상황에 맞는 프로그램이 가져다주는 편리함과 그 가능성에 감탄하였고 사내 교육을 통해 인프라구조, C 언어를 공부하며 어떻게 팀의, 회사의 효율 향상을 가로막는 문제를 풀 수 있을지 배우고 경험하기 위해 노력해왔습니다.

사이버대학 교육을 통한 SQL, 컴퓨터 및 네트워크 기초에 대해 공부를 하며 저는 효율적이고 분석적인 데이터 관리의 힘을 믿게 되었습니다. 사내 인트라넷의 구조가 데이터베이스의 스키마로 친근하게 다가오면서 회사를 그만둘 결심을 하고 패스트캠퍼스의 데이터사이언스 스쿨 과정을 신청하였습니다.

파이썬, SQL, 크롤링과 생활코딩의 프론트-백엔드 구조를 배우고 경험하며 코딩이 점점 익숙해지는 만큼 IT 업계에서의 제 모습을 꿈꿔왔습니다. 스쿨 과정 중 매주의 통계분석, 크롤링, SQL 과 AWS 과제를 진행하며 데이터 전문가로 가는 기반을 닦았습니다. 또한, Kaggle 의 부동산 집값 예측, 나뭇잎 종류 분류문제 등의 프로젝트를 진행하며 실무에 적응할 수 있는 기초를 준비하였습니다.

데이터 수집(크롤링), 데이터 전처리(imputation), 데이터 구조화(SQL, mongoDB), 데이터 분석 및 시각화(Python, Pandas, Scikit-Learn)로 이어지는 기본 사이클에 꾸준한 공부와 경험으로 탄실하게 성장해 나가겠습니다.

## 2) 성격의 장점과 단점 각각에 대해 자세히 설명해주세요.

저는 꼼꼼하고 신중한 성격의 사람이라고 생각합니다. 이런 제 성격은 정확하고 신중하게 진행하는 업무에 있어서 많은 도움이 되었습니다. 실제로 연구원으로 근무하는 동안에도 법규 개정 후 2년 동안 발견하지 못했던 식품표기법을 발견하여 즉각적인 시정조치로 회사의 피해를 막기도 하였습니다. 또한, 보고서 작성 시 하나의 오타도 놓치지 않으려 했던 꼼꼼함이 정확한 타이핑을 요하는 함수 코딩이나 쿼리문, HTML 문서 작성 시에도 많은 도움이 되고 있습니다.

신중하고 꼼꼼하여서 자칫 까다로운 인상을 줄 수 있다는 것이 단점이 될 수 있을 것 같습니다. 연구 또는 분석 업무로 온종일 앉아 업무를 진행하며 그런 성향이 강해지는 것 같아 저는 마라톤을 취미로 활동적인 습관을 길러왔습니다. 연구원으로 근무하면서도 연구도 중요하지만 일을 진행하는 관계자들과의 커뮤니케이션 또한 대단히 중요하다고 생각하였습니다. 타 부서와 함께 마라톤 대회에 참가하고 매주 꾸준한 운동은 제가 2년 동안 역동적으로 업무에 임할 수 있던 원동력이었습니다.

저의 꼼꼼함과 신중함을 바탕으로 매일의 분석과 프로그래밍에서 효율적으로 수행하고 꾸준한 운동으로 다져진 활기로 프로젝트를 원활히 추진해나가는 팀원이 되고 싶습니다.

## 3) 직무와 관련된 경험이나 보유한 역량에 관해 설명해주세요.

도서 및 사이버대학 강의를 통해 컴퓨터 공학의 기초를 다지고 패스트 캠퍼스 데이터 사이언스 과정의 실무 중심 프로젝트를 진행하며 분석가로의 역량을 쌓아왔습니다.

먼저, 러시아 집값 예측 모델 분석을 진행하며 데이터 분석의 기초를 실제로 적용하였습니다. 21580 x 293 의 train 데이터를 pandas 로 전처리하며 데이터프레임 핸들링은 물론 전처리가 어떻게 진행되고 최종 결과에 어떤 영향을 미치는지 경험할 수 있었습니다. 또한, 짧은 기간 동안 최적의 결과를 얻기 위해 단방향인 아닌 결과를 피드백으로 보고 이전 단계를 재조정하는 쌍방향 방식으로 진행하였습니다. 그렇기에 전처리부터 최종 모델까지의 단계를 직접 짰 코드로 자동화하여 신속하고 빈번하게 다양한 전처리, 인코딩, 모델을 시도할 수 있었습니다. 또한, XGBoost, Gradient 등의 모델이 아니라 greedy 방식의 back-elimination 모델과 제가 직접 짰 알고리즘을 통해 변수 선택션 모델을 만들며 전체 구조를 명확히 파악할 수 있었습니다.

전처리, 인코딩, 변수 선택 모델링, 다항 회귀분석과 크로스밸리데이션과 정규화로 이어지는 분석 프로젝트는 실제 실무를 조금이나마 경험할 수 있었던 값진 교육이었습니다.

두 번째로 진행한 나뭇잎 분류 프로젝트에서는 OpenCV 패키지를 사용하며 기본적인 이미지처리에 대하여 학습할 수 있었습니다. 첫 번째 프로젝트보다 데이터가 작고 Null 값이 없어 프로젝트 구성을 체계적이고 비교적 순차적으로 진행하였습니다. 이미지 정보를 제외한 숫자 features 를 정리한 후 이미지에서 정보를 추출하는 부분이 중점이 되었습니다. 이미지 크기, 나뭇잎 이미지의 길이, 넓이, 사각형 넓이, convex 넓이, PCA 등을 통해 이미지에서 정보를 추출하였습니다. 또한, 모델 선택 시 10 여 종의 모델을 시도하고 결과가 좋은 모델을 위주로 최적화하여 프로젝트를 논리적인 절차에 따라 진행하여 보았습니다. 최종적으로 1598 팀 중 258 등의 결과(late submission)를 얻을 수 있었습니다(logloss).

마지막으로 개인프로젝트를 진행하여 네이버 API 요청부터 웹 서비스(사이트)로 이어지는 파이썬 제품을 만들었습니다. 초기 입력 값으로 정부에서 발표한 가맹점 리스트와 요청 대상 지역 CSV 파일을 넣으면 API 요청-Preprocessing-Mongo DB 저장 후 불러오기-GeoProcessing-html Mapping까지의 과정을 코드로 구현하였습니다. 또한, 부트스트랩을 이용하여 flask 로 웹 서비스를 만들고 nginx 와 wsgi 로 AWS EC2 상의 웹 서버를 만들었습니다.

이 프로젝트를 진행하며 folium, shpely, flask, geopandas 등 새로운 패키지를 시도해 보았습니다. 또한, 많은 새로운 패키지를 시도하며 ‘구현을 위해 필요한 것을 어떻게 빠르게 배우는가’에 대해 많은 고민을 통해 체감할 수 있었습니다.

분석 제품을 만든다는 관점으로 프로젝트를 진행하며 꾸준히 코딩하고 성장하고 싶습니다. 그러한 반복을 통해 학습한 데이터 수집-최종의 분석 및 시각화의 단단한 기초를 통해 신뢰받는 팀원의 역할을 수행하고 그 기초 위에 전문성을 쌓아 올리고 싶습니다.

#### **4) 앞으로 해당 직무를 수행하며 어떤 방향으로 성장할 것인지에 대해 말해주세요.**

데이터 사이언스에는 프로그래밍, 커뮤니케이션 능력, 분석력, 데이터 관리 능력 등 어느 하나도 놓치고 싶지 않은 많은 분야가 연관되어 있다고 생각합니다. 또한, 분석업무를 진행하는 상황에 따라서도 사용하는 툴도 진행하는 방식도 차이가 있기에 현재로는 어떤 상황에도 빠르게 적응할 수 있는 기초를 쌓는 데 주력하고 있습니다. 또한, 어떤 상황에도 시너지를 일으킬 수 있는 하드웨어에 대한 이해를 바탕으로 한 프로그래밍 능력과 문제에 대한 수학 및 통계적 해결 역량을 기르고 싶습니다. 두 가지를 바탕으로 기하급수적으로 생산성을 높여나가는 데이터 사이언티스트가 되고 싶습니다.

다른 부분으로는 다양한 부서의 실무자들과 소통하는 데이터 전문가가 되고 싶습니다. 연구원으로 근무하는 동안 공장 출장을 가게 되면 저는 운동을 통해 다져진 체력으로 생산자분들과 함께 옆에서 뛰며 업무를 진행하였습니다. 공장 안 가보지 않은 곳이 없을 정도로 구석구석 돌아다니며, 함께 땀 흘리며 듣는 작업자분들의 한 마디는 사무실에 앉아서 얻을 수 없는 현장 경험이었습니다. 연계된 업무를 처리하며 현장경험에서 나오는 타 부서에 대한 세밀한 배려로 공장분들과의 신뢰 관계를 쌓았고 그렇기에 갑작스러운 진행에도 공장의 협조를 얻을 수 있었습니다. 또한, 현장에 대한 이해는 문제 발생 시 전체 프로세스에서 어떤 부분이 원인인지 빠르게 찾아낼 수 있게 하였습니다.

데이터 사이언티스트 역시도 광범위한 시각에서 바라보는 전체 프로세스에 대한 이해와 살아 숨쉬는 세밀한 현장의 이해가 필요한 직무라고 생각합니다. 수많은 관계부서, 다양한 고객과 바쁘게 변하는 시장에서 변화를 위한 킥핀을 찾아내고 해결하는 데이터 전문가로 성장하겠습니다.