Sungmin Kim

San Jose, CA, 95128
Phone : 312-885-0448 Email : kimx3129@gmail.com Linkedin : https://www.linkedin.com/in/sungminkim510
Github : https://github.com/kimx3129/Simon_Data-Science Portfolio: https://sungmin-kim.dev/

PERSONAL STATEMENT

Passionate about building intelligent workflow systems using Python, AWS, GCP, and modern AI tools to solve real-world problems. Strong interest in AI agents, Harness Engineering, workflow orchestration, and scalable cloud-based automation systems. Driven by continuous learning, practical innovation, and creating impactful AI-powered products that improve productivity, decision-making, and user experience.

EDUCATION

Master of Science in Data Science
Class of 2017, *Depaul University*, Chicago, IL

Bachelor of Science in Computer Science
Class of 2014, College of Science and Engineering, *University of Minnesota*, Minneapolis, MN

SKILLS

- **Computer Languages** : Python, SQL, JAVA, R, C, C++, Jinja Template, NodeJS
- **Softwares** : AWS, GCP, Airflow, Splunk, JIRA Board, Docker, Github, DBT, Jenkins, Opencv2, Concourse CI
- **Foreign Languages** : Fluent in Korean and English, intermediate level of Chinese

INDUSTRY EXPERIENCES

Tempus Inc, Chicago, IL

Senior Software Engineer, 3/22/2023 – Present

- Optimized crosswalk service performance, reducing runtime from 15 hours to 2 hours by leveraging BQ partition key to minimize scan size across the source dataset
- Implemented Data Loss Prevention(DLP) on GCP to safeguard against potential PHI leakage during the de-identification process
- Developed a DPL Automation platform using Airflow, BQ, GSpread API to streamline the ingestion of large-scale tables, improving scalability and operational efficiency
- Mentored a junior engineer, providing technical guidance and support for professional development

Software Engineer, 10/7/2021 – 3/21/2023

- Implemented de-identification process using DBT and Jinja templating, establishing it as a core framework for data ingestion and transformation
- Developed data pipelines for AWS-based architectures, including batch instances on ECS with ECR, and monitored infrastructure using CloudWatch
- Created a Splunk Dashboard to provide real-time visibility into RedShift disk and CPU usage, enhancing observability and system reliability

Egen Solutions, Chicago, IL

Data Engineer, 5/6/2019 – 10/6/2021

- Implemented automated Slack notifications and alerts, customizing messages to include relevant details such as CloudWatch URLs for any failed batch tasks
- Automated manual Jira requests by creating Jira service account and integrating with the Jira API, significantly improving response time and operational efficiency

CarVi Inc, Chicago, IL

Data Scientist, 2/20/2017 – 1/9/2019

- Analyzed driving data and applied Machine Learning techniques such as Decision tree, Naïve Bayes and K-Means Clustering to generate interactive dashboards using Angular JS
- Processed large-scale real-time IoT driving data (80-100K per day) by implementing automated data preprocessing pipelines using AWS Serverless techniques such as Lambda, CloudWatch, and DynamoDB

- Developed and validated a Crash Detection Algorithm in C, leveraging motion sensor data and testing on Raspberry PI for real-world reliability
- Implemented Video On Demand (VOD) feature to enable cloud-based video uploads from CarVi device, using Python and Chalice for API development

Depaul University, Chicago, IL

Research Assistant, 8/29/2016 – 5/26/2017

- Led a data science project focused on Computer-Aided Diagnosis to support clinical decision-making
- Applied Text Mining techniques, including NLP and TF-IDF, to analyze radiologist report and group similar cases using Hierarchical Clustering based on frequent term patterns
- Developed a deep learning-based image classifier in Python to detect whether lung nodule in a medical medical images were benign or malignant

First Analytics, Chicago, IL

Data Scientist Intern, 6/21/2016 – 8/19/2016

- Performed Sentiment Analysis for Animal-Health company blog in order to figure out the user with most negative review and applied K-means Cluster to group similar users and find the most N frequent words people used

Digi International, Minnetonka, MN

Software Engineer Intern, 1/30/2013 – 8/31/2013

- Optimized wireless router software to minimize runtime through Dynamic Programming
- Wrote a unit test for quality assurance in Python

PROJECTS

AI-Native Portfolio Website

AI Engineering Project, May 2026

- Built and deployed an AI-native portfolio website using React and Vercel, implementing responsive UI components and modern web development workflows.
- Leveraged multi-agent development workflows to accelerate prototyping, iterative feature refinement, and AI-assisted debugging throughout the development lifecycle
- Used GitHub as a main version control and deployment pipelines while integrating AI-assisted engineering practices into the project workflow

Jira Sprint Starter Automation

AI Engineering Project, May 2026

- Developed an event-driven Jira sprint automation platform integrating GitHub Actions cron workflows and Jira REST APIs for automated backlog analysis and sprint preparation.
- Built multi-agent orchestration pipelines using LangGraph to perform issue classification, estimation assistance, and sprint draft generation.
- Implemented scheduled and event-based CI automation triggers for recurring sprint workflows, GitHub issue events, and deployment orchestration.

OTHER EXPERIENCES

Inflearn

Online Instructor, May 2020 - Present

- Developed and taught online courses on software engineering and cloud platforms such as AWS, GCP, AI Engineer, and Python in Korean
- Applied leadership and mentoring skills by guiding learners through complex topics and helping them develop problem-solving mindsets

Computer Vision Projects

Participant, March 2019

- Implemented an object detection program to identify human fingers in real-time using a webcam, leveraging OpenCV2 features such as Threshold, Convex Hull, Contour Detection in Python

Data Science Hackathon, Baltimore, MA

Participant, September 2015

- Gained hands-on experiences in data analysis by working on real-world problems using R
- Implemented an interactive Shiny R application for daily nutrition and calories recommendation per day based on user information

PUBLICATIONS

- P Deshpande, A Rasin, J Son, **S Kim**, E Brown, J Frust, DS Raicu (2020), *Ontology-Based Radiology Teaching File Summarization, Coverage, and Integration*, Journal of digital imaging, 33(3), 797-813. https://pmc.ncbi.nlm.nih.gov/articles/PMC7256159/#_ci93_
- **S Kim**. (2023). *Introduction to AWS for immediate use at work*, Hanbit Media. ISBN: 9791169216579

CERTIFICATION

- **GCP Certificate – Professional Data Engineer**, September 2020
- **AWS Certified Developer – Associate**, October 2019
- **Data Science Specialization** (Johns Hopkins University), August 2015
- **IBM SPSS Statistics Level 1 v2**, July 2015
- **SCJP** (Sun Certified Programmer for the Java 2 Platform), June 2009