Sang (Tony) Young Chun | Phone: +82-010-6822-**** (KR) / +1-734-904-**** (US) | E-mail | GitHub | LinkedIn | Google Scholar

EDUCATION

University of Michigan, Ph.D. in Bioinformatics (2017) | Ann Arbor, MI | GPA: 3.6/4

University of Michigan, M.Sc. in Bioinformatics (2010) | Ann Arbor, MI | GPA: 3.7/4

Kalamazoo College, B.A. in Biology (2001) | Ann Arbor, MI | GPA: 3.1/4

EXPERIENCE

Wider Planet, Data Scientist, Manager | Feb 2021 ~ Present

- Developed a purchase prediction algorithm for targeted ad campaigns using LightGBM in GCP based on Spark, Big Query, Docker, Kubernetes and custom GPU clusters.
- Designed and implemented a deep feature synthesis methodology that reduced the production feature space by over 99% while increasing prediction accuracy by 5%.

CJ CheilJedang BIO, Senior Scientist | Data Science Team | Oct 2018 ~ Jan 2021

- Built an end-to-end solution for classification of patent text using GCP, BigQuery and Spark; attained over 87% accuracy using BERT in Keras and TensorFlow with multiple NVIDIA Tesla V100 GPUs.
- Developed a fuzzy string matching algorithm using NLTK to extract redundant sequences in a primer order database; extracted user, target and project trends for executive review.
- Created **calenjar**, an automated monthly scheduling system in Python to allocate resources based on variable constraints including personnel availability; built an interactive app in Dash to enable manual curation by non-programmers.
- Designed and implemented a Random Forest classifier to identify specific markers of infection by an under-studied pathogen in commercial agricultural sites; multi-million dollar technology transfer agreement pending.
- Wrote *Magellan* in Python, BASH and Docker to standardize and automate processing and visualization of sequence data; reduced delivery time of completed results by nearly 95%.
- Integrated legacy and newly purchased server resources into a fully scalable HPC using Kubernetes; enforced best practices for software developemnt, version control and deployment using Git/GitLab, Jupyter and Docker.

University of Michigan, Post-doctoral Research Fellow | Nov 2017 ~ Sep 2018

University of Michigan, Graduate Student Instructor | Aug 2016 ~ Dec 2016

- Created interactive laboratory modules on version control and statistical computing in R, directed lecture recitations, and prepared all graded exams and assignments.

University of Michigan, Research Technician Associate | Jan 2002 ~ Jul 2009

SKILLS (NOTE: bold, regular, and *italic* text denotes proficiency from highest to lowest))

Programming

- **Python**, **R**, **BASH**, SQL, AWK, *Java*, *C++*

Environments

- Linux, OS X, Windows, Git/GitHub/GitLab, BitBucket, Slack, Jira, Confluence, Docker/Swarm, Kubernetes, GCP (GCE/GKE, Big Query, AI-Platform), AWS, SGE, Slurm

Machine Learning / AI

- Regression, Random Forest, XGBoost, LightGBM, CatBoost, NLP, PCA, Clustering, Cross-validation, Sampling, Classification, NLTK, Hugging Face, Transformers, TensorFlow, Keras, ktrain, scikit-learn, SciPy

Data Wrangling / Visualization

- Pandas, Dask, Spark, NumPy, Plotly, Dash, Matplotlib, Seaborn, ggplot, Tableau, Bokeh

Languages

- English (native), Korean (conversational), French (basic)

RESEARCH AND TEACHING

- Created interactive laboratory modules on version control and statistical computing in R, directed lecture recitations, and prepared all graded exams and assignments.

University of Michigan, Graduate Student Research Assistant | 2013 ~ 2017

- Worked with Ryan Mills to create **SPECtre**, a signal processing algorithm that integrates multiple types of sequence data to estimate efficiency of protein synthesis in cancer cells. Designed a weighted sampling method to simulate imperfections in sequence data. Developed expertise in the aggregation, extraction, and integration of large data sets and databases including the 1000 Genomes Project, TCGA, COSMIC, GEO, and Ensembl.

University of Michigan, Graduate Student Research Assistant | 2011 ~ 2012

- Designed and implemented methods for the analysis of small RNA regulation using sequencing and mass spectrometry data in the lab of John Kim.

University of Michigan, Graduate Student Research Assistant | 2009 ~ 2011

- Under the supervision of Arul Chinnaiyan, identified high-impact mutations in sequencing data derived from prostate cancer patients.

- CJ CheilJedang BIO, Research Excellence Award | 2020

- Proteome Informatics of Cancer Training Grant, National Cancer Institute | 2013 ~ 2016

- Bioinformatics Training Program, National Institutes of Health | 2011 ~ 2013

- United States, Citizen

- Republic of Korea, F-4 Visa (valid through 2023)