

Seungjun Kim

Website: seungjun-data-science.github.io | LinkedIn: /kimseungjun | Github: /seungjun-data-science

Education

Vassar College

Poughkeepsie, NY | May 2020

Bachelor of Arts (B.A.) in *Mathematics (with Departmental Honors)*; Minors in *Computer Science, Chinese*

- **Coursework:** Applied Statistical Modeling, Data Science and Statistical Learning, Computational Linguistics, Bayesian Inference, Real Analysis, Econometrics, Linear Algebra, Analysis of Algorithms

Oxford University, St. Catherine's College

Oxford, United Kingdom | Jan 2019 – Jun 2019

Junior Year Abroad Visiting Students Program

- **Coursework:** Statistical Machine Learning, Quantitative Economics, Finance, Corporate Social Responsibility

Experience

Institute for Social Research, University of Michigan, *Research Area Specialist*

Ann Arbor, MI | July 2020 – Present

- Writes STATA and Python scripts for cleaning and harmonizing of criminal justice data to a standardized schema specified by the Criminal Justice Administrative Records System (CJARS).
- Builds automated Python program for data quality assurance that detects errors and distributional anomalies in the data.
- Creates Python program that produces statistical reports from the criminal justice database for agencies such as Department of Corrections, Sheriff and Police organizations.
- Performs literature review to be included in grant proposals for institutions including the National Institute of Justice (NIJ) and the National Science Foundation (NSF).
- Identifies victims of domestic violence and estimates their caseloads using the lab's criminal justice database and residence and relations crosswalks built from U.S. Census Bureau's survey and administrative data in a long term research project with the Center for Disease Control (CDC) of measuring the cost of violence in the U.S.

Spotify, *Data Science Intern*

New York, NY | Jun 2019 – Aug 2019

- Pulled user and web log data for the Mobile Web Player (MWP) with BigQuery and performed user segmentation analysis with Python to better understand user behavior and needs.
- Used anomaly detection techniques and identified deep link error which affected, on average, 14.8% of daily active users.
- Predicted growth projections of new features or marketing methods in collaboration with the UX Researcher and Product Manager to make proposals for roll out.

Vassar College Economics Department, *Research Assistant*

Poughkeepsie, NY | May 2018 – Dec 2018

- Researched with Dr. Esteban Argudo on two topics – Expected impact of 2017 Repatriation Tax Holiday on U.S. economy and measuring the impact of macroeconomic effects on loans – using 10-K data collected from the Securities and Exchange Commission for fiscal years 09'-17' on 10,148 unique companies.

Family Services Inc., *Data Analytics Intern*

Poughkeepsie, NY | Jan 2018 – May 2018

- Wrangled employee data with Python and investigated factors associated with employee turnover using Pearson correlation, partial dependence plots, SHAP values, permutation importance and LIME.
- Created an ensemble model with python that averaged the predicted probability of turnover from three solo models, Logistic Regression, Random Forest and Adaptive Boost, and classified employees into different turnover risk bins.
- Human Resources Manager used this model to preemptively address dissatisfaction of employees who are predicted to have high likelihood of quitting the job.

Skills

Data Analysis, ML, DB: Python, Tensorflow, Pytorch, R, STATA, Tableau, SQL, BigQuery

Other: Git, LaTeX, Asana, Microsoft Suite, Google Suite, Prezi, Canva

Language: Korean (native), Chinese (intermediate; HSK Level 5 Certificate)

Leadership

Vassar Haiti Project (VHP), *Student Leader and Treasurer*

Poughkeepsie, NY | Jun 2019 – May 2020

- Led general board members as one of the student leaders at the VHP, a 501(c) non-profit supporting the welfare of artists and sustainable development in Cher maître, Haiti through auction and sale of Haitian art and handicrafts.
- Allocated budget to events, digitized receipts on QuickBooks, managed money flow, worked as cashier during fundraiser art sales using the Square app, and educated new treasurers.

Korean Defense Intelligence Command, *Army Sergeant and Squad Leader*

Anyang, South Korea | Aug 2015 – May 2017

- Led 35 squad members during military training sessions as squad leader.
- Implemented a peer review system for the English to Korean translation operations of editorials and think-tank reports regarding military, national security, politics, and international relations of North Korea and the U.S.

Community Service

Liberty in North Korea (LINK) Vassar Chapter, *Volunteer*

Poughkeepsie, NY | Sep 2014 – May 2015

- Raised \$2,000+ funds for rescuing North Korean refugees through the Asian Night Market sales and the Benefit Concert.
- Increased awareness on the humanitarian crisis in North Korea by inviting North Korean refugees for talks and screening documentaries or movies on human rights abuses in North Korea.

Gwan-Ak Community Center, *Volunteer*

Seoul, South Korea | Jul 2007 – Aug 2013

- Delivered lunch boxes for the elderly living alone or the disabled who are unable to visit the community center for meals.
- Visited homes of the elderly on a regular basis for counseling.
- Initiated the *Hand-Written Letter Program* to enhance emotional health of the elderly which matched the senior citizens living alone with volunteers who sent hand written letters to them.

Competitions, Honors and Awards

Vassar College DataFest, *Best Overall Analysis Award*

2020

- Participated in a 24 hour long competition with the freedom to choose any COVID-19 related topic, sponsored by the American Statistical Association (ASA).
- Compared depression levels of people between during the 2008 Subprime Mortgage Crisis and the 2020 COVID-19 pandemic by building a depression classifier trained on Reddit data that used majority voting of three solo models – Naïve Bayes, Logistic Regression and Light Gradient Boosting Machine.
- Examined whether responses from the US government coincided with a decrease in depression levels.
- Used Tweets from 2009 and the early COVID period, Reddit data under r/depression and r/AskReddit sections and the Oxford COVID-19 Government Response Dataset to address these research questions.

Kaggle, *Instant Gratification*

2019

- Binary classification competition with synthetic data provided by Kaggle. Ranked 328 out of 1832 (Top 18%).

Kaggle, *Santander Customer Transaction Prediction*

2019

- Binary classification competition using Santander's customer transaction data to predict which customers will make a specific transaction in the future. Ranked 1354 out of 8802 (Top 16%).

Vassar College, *Tananbaum Fellowship*

2019

- Awarded to 20 junior year students selected based on merit.
- Provided \$5,000 worth of financial support for internships and career guidance.

National Model United Nations (NMUN) NY, *Outstanding Delegation & Position Paper Awards*

2015

- Received the Outstanding Delegation Award as a member of the delegation (Hudson River Group - Dutchess Community College and Vassar College Joint Team) representing Australia.
- Received the Outstanding Position Paper Award at the Nuclear Non-Proliferation Treaty Committee.