

Seungjun “Josh” Kim

76000 Verano Road South, Irvine, CA 92617 | kims17@uci.edu

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

EDUCATION

University of California, Irvine | Doctor of Philosophy (PhD) in *Informatics* Irvine, CA | June 2028 | GPA: N/A

- **Advisor:** [Dr. Sean D. Young](#)
- **Research Interests:** Developing ML/DL and NLP methods applied on unstructured big data (e.g. social media feeds) to address medical misinformation, bias in healthcare and inform public health policymakers.

Vassar College | Bachelor of Arts (B.A.) in *Mathematics*; Minor in *Computer Science* Poughkeepsie, NY | May 2020 | GPA: 3.80

- **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 | Visiting Students Program Oxford, United Kingdom | Jan – Jun 2019

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

Skills

- Comfortable: Python, STATA, Supervised ML, Unsupervised ML, BigQuery, MySQL, Sentiment Analysis, Topic Modelling, NER
- Working Proficiency: R, Tableau, Git, Pytorch, RNN, LSTM, XAI, Language Models (BERT, RoBERTa etc.), Diff-in-Diff
- Basic: HTML, CSS, Java, D3.js, Django, Tensorflow, Streamlit
- Languages: Korean (native), English (native), Chinese (intermediate; HSK Level 5 Certificate)

RESEARCH EXPERIENCE

Institute for Social Research, University of Michigan, [CJARS](#), *Research Area Specialist* Ann Arbor, MI | July 2020 – June 2022

- Objective of the project was to measure the cost of domestic violence but mostly intimate partner violence (IPV) in the U.S.
- Wrote scripts in STATA and Python to:
 - Identify perpetrators of domestic violence based on offense description from the Criminal Justice Administrative Records System (CJARS) database.
 - Link these offenders with family relations and residence data from the U.S. Census Bureau and implemented logics to identify victims of IPV.

Vassar College Economics Department, *Research Assistant* Poughkeepsie, NY | May – Dec 2018

- Worked with Dr. Esteban Argudo to describe the impact of the 2017 Repatriation Tax Holiday on loan supply by analyzing 10-K data for fiscal years 09'-17' on 10,148 unique companies and lending behavior of multinational and domestic banks.

WORK EXPERIENCE

Institute for Social Research, University of Michigan, [CJARS](#), *Research Area Specialist* Ann Arbor, MI | July 2020 – June 2022

- Harmonized criminal justice data based on pre-defined schema and built pipelines for data processing in Python and Stata.
- Built an automated data quality control program in Python that applies anomaly detection algorithms and custom logic.
- Wrote SQL queries that produces recidivism, health and labor outcome statistics of criminal offenders as part of the Django program that generates statistical reports from the lab database.
- Performed literature review to be included in grant proposals for the National Institute of Justice (NIJ) and the National Science Foundation (NSF).

Spotify, Mobile Web Player Team, *Data Science Intern*

New York, NY | Jun – Aug 2019

- Performed user cohort analysis in Python on terabytes of user data to better understand user behavior and needs.
- Used anomaly detection techniques to identify deep link error and presented solutions to resolve it, resulting in a 5% reduction of affected daily active users (14.8% → 9.2%).

- Identified factors associated with employee turnover using partial dependence plots, SHAP values, and LIME.
- Built an ensemble model 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.

PROJECTS

(2020) **Comparing Depression Levels of People during 2008 Subprime Mortgage Crisis v.s. 2020 COVID-19 Pandemic**

(2020) **Real, Fake and Sarcastic News Detection in a Class Imbalance Setting**

(2020) **Potential Gentrification Trends of New York City Airbnbs and Listing Price Prediction Using Machine Learning**

(2019) **Do players in certain positions (e.g. Center, Guard, Forward) have more Instagram followers, on average, than those in other positions? Analysis using Bayesian Inference**

TEACHING

University of California, Irvine (UCI)

2022 Fall—*Teaching Assistant*; ICS 139W: CRITICAL WRITING ON INFORMATION TECHNOLOGY

Vassar College

2018 Spring—*Python Tutor*; Google Ignite-CS Program (Vassar College Chapter)

Korean Minjok Leadership Academy

2014 Summer, 2013 Winter, 2011 Winter—*English Debate Teaching Assistant*; Global Leadership Program for Students (GLPS)

COMPETITIONS, HONORS, AWARDS

Vassar College DataFest, *Best Overall Analysis Award* 2020

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.

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 Chermaitre, Haiti through auction and sale of Haitian art and handcrafts.
- 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.
- Translated open source intelligence (e.g. editorials and think-tank reports) regarding military, national security, politics, and international relations of North Korea and the U.S. from English to Korean and vice versa.
- Implemented a peer review system for the translations.

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 via 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 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.