

Ethan Kim

(617) 758-9553 | ethan_kim@college.harvard.edu | 308 Amory St, Boston, MA 02130

Github: <https://github.com/ethankim00/> | Website: <https://ethankim00.github.io/>

EXPERIENCE

BGI LLC

Data Science Intern.

Sep '20 - Present

- Use Natural Language Processing to analyze cockpit voice recordings and improve aviation performance
- Develop transfer learning techniques for speech to text transcription using Mozilla DeepSpeech
- Fine tune deep speech model on Air Traffic Control audio data to achieve over 90% character accuracy rate

BRAINSPEC

Data Science Intern

May '20 – August '20

- Developed visualizations to present Magnetic Resonance Spectroscopy data from brain scans for clinical use
- Used d3 and Angular to visualize spectra for brain metabolites
- Created Python data processing pipeline to analyze changes in brain spectra due to PTSD and concussive injury

HARVARD COLLEGE DATA ANALYTICS GROUP

Director of Consulting

Aug '20 - Present

- Oversee 7 Data Science consulting teams working with nonprofits, startups and Fortune 500 companies
- Developed technical standards for teams
- Sourced clients to help student club grow to 50+ members

Case Team Leader

Jan '20 – May '20

- Led team of six on Data Science research project for a sports analytics company
- Explored low latency win probability prediction models for soccer matches
- Achieved high ranked probability score with faster prediction time than existing models

Coronavirus Response Team

May '20 – June '20

- Collaborated with Brigham and Women's Hospital to create visualizations for a high impact COVID paper (JAMA IM)
- Performed initial data processing and consulted on effective visualization techniques

ACTIVITIES

HARVARD ACADEMIC RESOURCE CENTER

Peer Tutor

Dec 2018 - Present

- Tutored Students in Math 21A Multivariable Calculus and Stat 111, Introduction to Statistical Inference

Interests

Sports analytics, reading, football, Harvard Men's Club Soccer, Harvard Square Homeless Shelter

EDUCATION

HARVARD UNIVERSITY

Bachelor's in Statistics, Secondary in Computer Science

GPA: 3.76/4.0

SAT: 2370/2400 (V800, M800, W770)

COURSES

Advanced Topics in Data Science, Data Structures and Algorithms Data Science, Time Series, Probability, Inference, Linear Models

SKILLS

Python, Scikitlearn, TensorFlow, Dask, SQL, R, d3.js, Angular.js, git, Docker, PySpark

PROJECTS

Impact of Weather on Social Distancing

- Scraped county level weather data to evaluate relationship with social distancing indices
- Explored effect of weather on Coronavirus reproductive rates
- Use LSTM models to predict rates of coronavirus spread

Predicting Types of Crime in Boston

- Using spatiotemporal data to predict category of crimes in Boston
- Achieved significant increase in accuracy using census and location data and Random Forest models

Detection of Cell Cycle Genes and Regulators in *S. Cerevisiae*

- Time series analysis on mRNA expression data in yeast to detect cyclic genes
- Autocorrelation technique correctly identified key cell cycle proteins and regulators

Training an AI to play Incan Gold

- Implemented playable version of the board game, Incan Gold, using Python, Flask and JavaScript
- AI trained on winning moves from thousands of games achieved 150% increase in win percentage

Covid Drug Discovery

- Analyze chemical properties of small molecules to predict antiviral properties