Joshua Chen

Portfolio: chenjoshua7.github.io | GitHub: /chenjoshua7 | Brooklyn, New York | chen.joshua98@gmail.com

Education

M.Sc. in Data Science Methodology - Barcelona School of Economics   2023 - 2024
Selected Courses: Statistical Modeling and Inference, Machine Learning, Deep Learning, Advanced-Natural Language Processing, Causal Inference, Optimization, Reinforcement Learning

B.A. in Mathematics - Boston College   2016 - 2020
Morrissey College of Arts and Sciences Honors Program, Dean's List First Honors
Selected Courses: Econometrics (Doctorate Level), Mathematical Modelling, Statistics, Probability

Selected Projects

- **Comparing CNNs and Vision Transformers (Masters Thesis)** - *Pytorch, Transfer Learning, Computer Vision*
  I compared the mathematics and intuition of CNN and Transformer architectures for vision classification. I fine-tuned ViT and ResNet models on various datasets, evaluating their performance and rotational invariance and then utilized Centered Kernel Alignment (CKA) to analyze and compare layer similarities between the models.
- **BERT for Topic Classification (Course Capstone)** - *Transformers, Data Augmentation, Zero-shot Learning*
  I fine-tuned DistilBERT from Hugging Face to classify journal articles into 10 categories. Despite having only 32 total training samples, I boosted accuracy by 20% using feature engineering, data augmentation, and generating additional data with the OpenAI API.
- **Hospital Mortality Prediction (Kaggle)** - *Random Forest, XGBoost, SVM, Feature Engineering*
  I achieved a 0.92 AUC in predicting mortality using the Beth Israel dataset. This success was driven by comprehensive feature engineering and the strategic use of metadata. I applied PCA and clustering techniques to effectively reduce the dimensionality of patient vital signs data.

Work Experience

Data Science Intern
World Bank - Development Impact (DIME)   April 2024 - Present
- Support the development of an R package for querying Facebook Marketing API
- Utilized clustering techniques and conducted statistical tests to create R vignettes with the package, aimed at analyzing the rural and urban divide.

11th Grade Team Lead/AP Calculus Teacher   August 2020 - July 2023
Uncommon Preparatory Charter High School
- Led a cross-department team of 12 staff and over a hundred students - resulting in the highest attendance and homework completion rates across all grades in the school.
- Achieved the district's second-highest Calculus test scores by implementing engaging and challenging coursework and implemented data-driven strategies that led to a 15% improvement in Algebra 2 scores.

Skills

Programming:
- **Tasks:** Natural Language Processing, Computer Vision, Text/Image Classification, Deep Learning Optimization
- **Tech Stack:** Python, R, PyTorch, SQL, HTML, CSS, Keras, Tensorflow, Tableau, Sci-kit Learn

Other

M.A. in Secondary Mathematics Education - Relay Graduate School of Education   2020 - 2022
Mentor - Big Brothers Big Sisters of Massachusetts Bay   2017 - Present
Languages - English (Native), Chinese (Fluent)