# Joshua Chen

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

## Summary


- Extensive experience in developing both traditional and deep learning models, with particular expertise in predictive modeling, Natural Language Processing, and Computer Vision.
- Strong statistical and mathematical background with a deep theoretical understanding of data-related challenges. Hands-on experience in Linear Programming, Statistical Inference, and Causal Inference.
- Highly proficient working with Python, R, and many machine learning packages. Skilled technical communicator with data visualization skills with Tableau, Plotly, Leaflet, and more.


## Selected Projects

**Song Lyrics Classification Web App (Personal Project)** *MLOps, Azure, Docker, FastAPI, JavaScript*

Developed an end-to-end machine learning project where I fine-tuned DistilBERT on a dataset of 50,000 songs. The model's progress was tracked using MLFlow, and it was deployed as an API with FastAPI, containerized with Docker, and hosted on an Azure server. The application is featured in my portfolio.

**Comparing CNNs and Vision Transformers (Master's Thesis)** *Pytorch, Tensorflow, Computer Vision*

Detailed the mathematical foundations and intuitive differences between CNN and Transformer architectures for vision classification. Fine-tuned ViT and ResNet models on multiple datasets, assessing their performance and rotational invariance. Used CKA to analyze layer similarity and transfer learning affects.

**Hospital Mortality Prediction (Kaggle)** *Random Forest, XGBoost, SVM, PCA, 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.

**Additional Projects:** MovieLens Recommendation System, Primavera Sound effect on Barcelona Hotel Prices

## Skills

**Tasks:** NLP, Computer Vision, Text/Image Classification, ML Prediction, Statistical Inference, Optimization
**Tech Stack:** Python, R, SQL, MLFlow, FastAPI, Docker, Azure, PyTorch, TensorFlow, XGBoost, Pandas/Numpy
**Additional:** Tableau, Web Development (HTML, CSS, JavaScript), Git, Gurobi, Leaflet, Plotly, Geospatial Data

## Work Experience

**Data Science Intern**

*World Bank - Development Impact (DIME)* April 2024 - August 2024

- Developed an R package for querying the Facebook Marketing API, with demonstrated applications in studying the urban/rural divide in Nigerian regions through a published R markdown vignette.
- Applied unsupervised learning, statistical techniques, and created interactive graphs and geospatial visualizations using leaflet and ggplot2 to effectively communicate derived insights.

**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.

## Education

**M.Sc. in Data Science Methodology** - Barcelona School of Economics August 2023 - July 2024

**Selected Courses:** Machine Learning for Finance, Statistical Modeling and Inference, Machine Learning, Deep Learning, Advanced Natural Language Processing, Causal Inference

**B.A. in Mathematics** - Boston College August 2016 - May 2020

*Morrissey College of Arts and Sciences Honors Program, Dean's List First Honors*

**Selected Courses:** Econometrics (Doctorate Level), Mathematical Modeling, Statistics, Probability