Satvik Kishore

Education

Duke University Durham, NC
Masters in Interdisciplinary Data Science *2021–2023*
Relevant Courses: Statistics, Machine Learning, Cloud Engineering, Natural Language Processing, Computer Vision and Medical Imaging, Causal Inference.

Indian Institute of Technology Madras Chennai, India
B.Tech., Materials Engineering; Minor in Industrial Engineering *2013–2017*
Relevant Courses: Probability, Decision Modeling, Pattern Recognition, Operations Research, Computational Engineering.

Technical Skills:

- **Python:** PyTorch, Tensorflow (Keras), Pandas, Scikit-Learn, Flask
- **R:** data.table, ggplot
- SQL, Git, Docker, Amazon Web Services
- **Certifications:** AWS Machine Learning Specialty

Publications

- Kishore, S., Thomas, T., Sachdev, H., Kurpad, A. V., & Webb, P. (2022) Modeling the potential impacts of improved monthly income on child stunting in India: a subnational geospatial perspective. *BMJ Open*, 12:e055098. https://doi.org/10.1136/bmjopen-2021-055098
- Shivakumar, N., Kashyap, S., Kishore, S., Thomas, T., Varkey, A., Devi, S., Preston, T., Jahoor, F., Sheshshayee, M. S., & Kurpad, A. V. (2019). Protein-quality evaluation of complementary foods in Indian children. The *American journal of clinical nutrition*, 109(5), 1319–1327. https://doi.org/10.1093/ajcn/nqy265

Research Experience

Interpretable AI for diagnosing breast cancer *Summer 2022–Present*

- Building convolutional neural networks with additional prototype layers.
- These models provide diagnoses on malignancy of breast tissue along with informing the physician why the model has made the prediction.

Earthquake Early Warning Detection, Duke University *Summer 2022*

- Built a prototype Machine Learning Model and framework to detect incoming earthquakes and classify their severity from seismological data.
- Explored the efficacy and performance of Gaussian Process Regression to model seismological data.
- Improved upon state the current state of the art model evaluation metrics by models by 30%.

Professional Experience

St. John's Research Institute Bengaluru, India
Data Scientist, Data Science & Public Health Research *2017–2021*
My role at St. John's included ideating research proposals, performing analyses, and publishing and presenting results from different projects.

- Modeling potential impacts of improved monthly income on child stunting in India — a geospatial perspective
    - Raised $100,00 awarded by the Gates Foundation and the Government of India as part of The Grand Challenges - India.
    - The aim was to help policymakers identify optimal strategies to reduce child undernutrition (stunting) in India.
    - Trained and applied Gaussian Process Regressions Models to calculate small-area estimates of stunting prevalence and potential benefits.
    - Tools used: Python, PyTorch, GPyTorch
- Protein-quality evaluation of complementary foods in Indian children
    - Tested statistical differences in digestibility across multiple amino acids in a clinical research experiment.
    - Estimated digestibility corrected Amino Acid intake scores for states of India. Tested for associations with child growth failure.
- Nutrition Value chain Data Repository
    - Collated, preprocessed, and set up a centralized database of about 100+ harmonized datasets across the nutrition value chain (crop production to health outcomes through distribution, purchase, consumption, and absorption.
    - Built a dashboard for easy exploration and extraction of the data for the Institute.

Projects

Brain Tumor Segmentation *Spring 2022*

- Built Image Segmentation Neural Network Models (U-net) that scan 3-dimensional brain MRI images and identiy regions of brain associated with a tumor.
- The model can be used to aid neurosurgeons to precisely locate damaged regions in the brain.

Optimizing CT scan slice count through Lesion detection using YOLO *Spring 2022*

- Simulated CT imaging on images of lesions to create images that would have been obtained from lower resolution CT Scanning.
- Optimzed number of CT slices required during imaging that allow acceptable object detection capabilities on Object-Detection Neural Network Models while minimizing costs.

Does Airbnb listing’s annual revenue vary by with host status? *Spring 2022*

- Analyzed AirBnb data from American cities to determine if superhosts are able to generate more revenue than regular hosts.
- Used Causal Inference principles to balance data and determined that superhosts are indeed more profitable.

AWS Cloud Tweet Generator *Fall 2021*

- Built and deployed a end-to-end cloud-based data pipeline that generates tweets everyday relevant to current trending topics.
- The pipeline collects and cleans tweets everyday for training data and then re-trains and deploys a generative LDA model.
- Used AWS services: Lambda, S3, EC2, ECR, and Batch. The pipeline was deployed using Infrastructure as Code (AWS CDK).

Star Trek: Analysis of Episodes *Fall 2021*

- A statistical analysis of IMDb data from four star trek TV shows to evaluate which characters are perceived more favourably.
- Engineered features from script data to create a proxy for character-screentime in each episode.