Satvik Kishore

Education

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

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, et al. Modeling the potential impacts of improved monthly income on child stunting in India: a subnational geospatial perspective. BMJ Open 2022;12:e055098. https://doi.org/10.1136/bmjopen-2021-055098
    - Trained Gaussian Process Regression Models to calculate small-area estimates of child undernutrition in India.
    - Estimated impact of improved income on undernutrition via injection of supplementary monthly income.
- Shivakumar N, Kashyap S, Kishore S, et al. Protein-quality evaluation of complementary foods in Indian children. American Journal of Clinical Nutrition. 109:5. May 2019. Pages 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*

- Developed a prototype Machine Learning algorithm to predict Earthquakes and associated intensities in Nepal.

- Trained Gaussian Process Regression Models using Seismologic wave data in a multi-task prediction framework.
- Implemented new testing standards. The trained model achieved a 30% increase in R^2 over the SOTA.

Professional Experience

St. John's Research Institute | Public Health Research Bengaluru, India
Data Scientist *2017–2021*

- Engaged with government stakeholders to develop data science solutions to tackle child undernutrition in India.
- Ideated and developed research projects to discover efficiency of different methods to tackle undernutrition.
- Raised $200,000 in funding from the Bill and Melinda Gates Foundation and led a project that analyzed impact of improved household monthly income on child undernutrition prevalence.
- Developed and trained geospatial models using Gaussian Process Regressions on large datasets using PyTorch.
- Compiled public health data from different sources into a data repository that was adopted by multiple institutions.

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.