Satvik Kishore

Education

Duke University Durham, NC
Masters in Interdisciplinary Data Science *Expected: May 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 *Jul 2013 – May 2017*
Relevant Courses: Probability, Decision Modeling, Pattern Recognition, Operations Research, Computational Engineering

Technical Skills:

- **Python:** PyTorch, Tensorflow (Keras), Pandas, Scikit-Learn, Numpy, Panas, Matplotlib, Flask
- **R:** data.table, ggplot
- SQL, Git, Docker, Amazon Web Services, Tableau
- **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 Geospatial Machine Learning models to calculate small-area estimates of child undernutrition
- 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 Computer Vision Models for diagnosing breast cancer *Jun 2022 – Present*

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

Professional Experience

Data+ Rhodes Information Initiative *May 2022 – Aug 2022*

- Developed a prototype Machine Learning algorithm that will be installed in Earthquake Early warning sites in Nepal.

- Trained multi-task Gaussian Process Regressions and Linear Regression models from Japanese seismological data.
- Implemented a novel bivariate model testing framework. Achieved 30% increase in R^2 over the SOTA

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.

Lighthouse Datalabs *May 2016 – Aug 2016*

- Developed machine learning models to help a healthcare blog client perform demographic based ad targeting.
- Trained tree-based models to classify age and sex of users, improving the AUC by 10% and ad click-through rate by 5%.

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 and optimized CT imaging on a lesion dataset, creating synthetic data at varying CT projection count.
- Achieved a decrease in cost by 75% by succesfully training Yolo Object Detection models with acceptable accuracy.

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*

- Engineered an AWS cloud solution to generate new artificial tweets everyday based on relevant current topics.
- Implemented an NLP model that trains on newly scraped and cleaned everyday data costing less than $1/month.
- 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.