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, 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
- 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 Machine Learning Lab, Duke University under Dr. Cynthia Rudin Durham, NC
Jun 2022 – Present

- Designed computer vision breast cancer detection model based on case based reasoning architecture in PyTorch.
- Achieved SOTA AUC scores of 0.95 while improving the interpretability of the model by making it more sensitive to fine-grained features in images.

Professional Experience

Data+ Rhodes Information Initiative Durham, NC
Machine Learning Engineer Intern *May 2022 – Aug 2022*

- Developed Gaussian Process Regression Machine Learning models for earthquake early warning systems in Nepal.

- Implemented pipeline that takes in input the seismograph wave data, transforms the data to wave features, then uses the trained model to predict intensity of oncoming earthquakes to trigger alarms.
- Improved upon the SOTA R2 scores by 30%. The model is set to be installed in earthquake detection facilities in Nepal.

St. John's National Academy of Health Sciences Bengaluru, India
Data Scientist *Jul 2017 – May 2021*

- Standardized and centralized health data from multiple sources into a single repository to be used by multiple institutions.
- Engaged with government stakeholders to develop data science solutions to tackle child undernutrition in India.
- Secured $200,000 in funding from the Bill and Melinda Gates Foundation to lead a project on analyzing the impact of improved household income on child undernutrition in India.
- Modelled child undernutrition in the country through Geospatial Machine Learning Models using PyTorch.

Lighthouse Datalabs Pune, India
Data Science Intern *May 2016 – Dec 2016*

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

Projects

Brain Tumor image segmentation using U-Nets *Spring 2022*

- Developed an AI tool for brain tumor image segmentation for precise neurosurgery, using TensorFlow.
- Implemented a "U-Net" Deep Neural network architecture to achieve a DICE similiarity score of 81%.

CT Scan Imaging Optimization *Spring 2022*

- Conducted a study on optimization of radiation exposure in CT scans to balance image quality and lesion detection.
- Trained YOLO models to reduce radiation exposure by 50% while maintaining lesion detection capability.

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 Powered Tweet Generator. *Fall 2021*

- Implemented and dockerized NLP model to retrain on fresh daily data and deployed to generate trend-relevant tweets.

- Engineered AWS cloud solution to schedule NLP docker containers to run everyday. Brought down costs to < $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.