Reshmi Ghosh

Websites: https://www.linkedin.com/in/reshmi-ghosh/

Email : gsh.reshmi@gmail.com
Mobile : (412)499-5751

Education

Carnegie Mellon University — Pittsburgh, PA

Doctor of Philosophy(Ph.D.) - **Focus: Deep & Machine Learning for Climate Change**; GPA: 3.74/4.00 — *Oct 2021*

M.S. Engineering & Public Policy; **Focus: Data Science for Policy** — *May 2020*

M.S. Civil & Environmental Engineering - **Focus: Machine Learning for Infrastructure Systems**; GPA: 3.70/4.00 — *Dec 2017*

University of Mumbai — Mumbai, India

B.Tech Civil Engineering; GPA: 3.80/4.00 — *May 2016*

Relevant Coursework

Data Structures & Algorithms, Ph.D. Deep Learning, Practical Data Science, Machine Learning, Data Analytics, Business Intelligence, Decision Analytics, Risk Analysis, Python for Developers, Data Warehousing, A/B Testing (Udacity), Applied Data Analysis (Statistics), Reinforcement Learning

Relevant Professional Experience

Microsoft — Cambridge, MA

Data and Applied Scientist II, Microsoft AI — *Oct 2021 - Present*

A team spanning across multiple MS organizations to empower products with AI and ML capabilities

● FY22H1 rotation: Azure Quality Infrastructure - Signal Processing Team

1. Deployed an end-to-end pipeline in Python and PySpark to monitor anomalies in Azure services by consuming terabytes of Service Level Indicator data in Synapse VM. Improved existing system by parallelizing spark jobs to reduce iteration time by 15%, leading to significant compute cost reduction.

2. Analyzed customer impacting Azure service outages by developing an open-source python package to determine statistical correlation metrics between customer responses and SLI service anomalies.

3. Created data-ingestion pipelines using Python and Spark to generate daily anomalies and monitor health of multiple Azure services.

● FY22H2 rotation: Office Experiences Organization (OXO) - Docs Machine Learning Team

1. Supported OXO's effort to simplify commanding in O365 applications (PowerPoint & Word) by developing database quering scripts in C# to analyze Gigabytes of usage data. Investigated frequently used command probabilistic distributions in Python by consuming the extracted usage data.

2. Researched and engineered a novel 'vectorized reward & penalty' based loss function in Python for Machine Learning models in production to improve command predictions and model performance for 4million customers worldwide.

3. Improved ranking of predictions of existing Machine Learning models, by prototyping a new 'Counterfactual Learning to Rank' approach in PyTorch to help surface rarely used commands to the user. New model increased performance metrics by 7%.

Microsoft Hackathons

● **NLP to predict root cause**: Supported Azure CXP to determine the root cause for incident tickets based on the ticket Title/Description from the P360 dataset. Developed a multi-hierarchical NLP model using BERT, LSTM, and Support Vector Machines. The Support Vector Machines with radial basis kernel outperformed complex BERT and LSTM models with minimal fine-tuning (89%) accuracy.

● **Fix Hack Learn - Loop auto-complete**: Leveraged Bing API search queries to implement auto-complete commanding feature by developing a BERT based NLP model.

Relevant Projects

- **Attention based Transcript generation from speech**: Deployed an attention based encoder-decoder language model using Pyramidal BiLSTMs and LSTMs with regularization techniques like locked dropouts and teacher forcing rate, in PyTorch, to generate text from input speech data. Used Beam search for decoding and obtained a Levenshtein distance of 17.95.
- **Utterance to Phoneme mapping in Speech data**: Implemented a LSTM model to predict phonemes associated with a given set of utterances. The model was trained on unaligned phonemes labelled data. The predictions resulted in a Levenshtein distance of 9.84.
- **Deep Learning research project**: Trained models in a team of four that can be optimized to build an android application to help the blind in grocery shopping. Deployed the VizWiz dataset (collection of 20k images captured by blind people through their mobile phones) to implement MobileNetv3 & an LSTM based language model to develop a Visual Question Answering based framework. Achieved an accuracy of 74%.
- **Ph.D. thesis**: Researched the impact of long-term inter-annual variance of historical temperature records on anomalous demand signals by reconstructing 40 years of hourly load data using BiLSTM based time-series model with fixed effects in PyTorch, and compared the performance against GBM and CatBoost. Explored the implications of the anomalous predictions by developing a stochastic models leveraging Monte-Carlo Simulation.
- **Unsupervised Machine Learning project**: Conducted end-to-end data analysis to examine user behavior on Instagram about climate change by using unsupervised clustering techniques on web-scraped images & associated hashtags. Determined that only 10% of all users posted relevant climate change related content between November-December'19.

Programming Skills

Languages: Python(PyTorch, TensorFlow NumPy, Pandas, scikit-learn, NLTK, Matplotlib, PySpark), KQL, SCOPE, SQL, Matlab, C#*

Technologies: Google Colab, GCP, AWS, Azure Databricks, Azure ML, Synapse ML, Visual Studio, AWS, Git, MySQL, PostgreSQL   *Familiar

Invited Speaker

- ODSC East 2022 Speaker for Women in Ignite Session Panel   *April 2022*
- Guest Lecturer (Deep Reinforcement Learning), School of Computer Science, Carnegie Mellon University   *April 2022*

Achievements

- Semi-Finalist CTO Open AI CODEX Challenge, Microsoft   *Jan 2022*
- Dean's Fellowship recipient, Carnegie Mellon University   *Aug 2018*
- Carnegie Mellon Merit Scholarship   *March 2017*
- CMU Civil and Environmental Engineering department scholarship   *Jan 2016*

Services

- Organizer, Women in Machine Learning and Data Science, Boston chapter
- Oxford University Artificial Intelligence Society Mentor
- Machine Learning Engineering Mentor and PoC, Microsoft Silicon Testing Project
- Conference Reviewer, Microsoft Machine Learning & Data Science Conference
- Scholarship Committee Reviewer, Women @ Microsoft

Publications

- Reconstruction of long-term historical demand data, <u>Reshmi Ghosh</u>, Michael Craig, H. Scott Matthews, Constantine Samaras; Tackling Climate Change with AI workshop, ICML'21
- Topic Segmentation in the Wild - Topic Modeling for Unstructured Data & Conversations; Sharanya Kamath, Dhuri Shrivastava, Harjeet Singh Kajal, <u>Reshmi Ghosh</u>, Samyadeep Basu, Soundaranjan Srinivasan (in review SCAI'22)