Rahul Yedida, Data Scientist
Website :: GitHub :: LinkedIn :: Google Scholar
rahul@ryedida.me
+1 (206) 660-7542

Education

North Carolina State University — Raleigh, NC
Ph.D. Computer Science - GPA: 3.9/4.0 — *Aug. 2019 – Present*

PES University — Bangalore, India
B.E. Computer Science - GPA: 7.9/10.0 — *Aug. 2015 – May 2019*

Employment

Amazon — Bellevue, WA
Software Dev Engineer Intern — *May 2022 – Jul 2022*

- Worked on publishing announcements in scorecards used by delivery service partners (DSPs).
- **Technology:** React, Redux Saga, AWS (ECS, DynamoDB), Java

North Carolina State University — Raleigh, NC
Graduate Research Assistant — *Jan. 2020 – May 2022*

- Better, faster deep learning for software engineering; V&V for AI systems; reuse in SE

Graduate Teaching Assistant — *Aug. 2019 – Jan. 2020*

- Held office hours for 54 undergraduate students and delivered lectures on C++.

Indian Institute of Astrophysics — Bangalore, India
Research Intern — *Jul. 2018 - Mar. 2019*

- **Image denoising:** Worked on image restoration of globular clusters using convolutional neural networks.
- **Research:** Proposed novel adaptive learning rate scheme for deep neural networks.

Publications

Yedida, R., Kang, H. J., Tu, K., Lo, D., & Menzies, T. (2022). How to Find Actionable Static Analysis Warnings. *arXiv preprint arXiv:2205.10504*

Yedida, R., Menzies, T. (2022). How to Improve Deep Learning for Software Analytics (a case study with code smell detection). In *2022 IEEE/ACM 19th International Conference on Mining Software Repositories (MSR). IEEE, 2022.*.

Yedida, R., Krishna, R., Kalia, A., Menzies, T., Xiao, J., & Vukovic, M. (2022). Partitioning Cloud-based Microservices (via Deep Learning). *arXiv preprint arXiv:2109.14569*.

Baldassarre, M. T., Ernst, N., Hermann, B., Menzies, T., & **Yedida, R.** (2021). (Re)use of Research Results (is Rampant). *arXiv preprint arXiv:2108.06821*

Yedida, R., & Saha, S. (2021). Beginning with Machine Learning: A Comprehensive Primer. *The European Physical Journal Special Topics: 1-82.*

Agrawal, A., Yang, X., Agrawal, R., **Yedida, R.**, Shen, X., & Menzies, T. (2021). Simpler Hyperparameter Optimization for Software Analytics: Why, How, When?. *IEEE Transactions on Software Engineering, doi: 10.1109/TSE.2021.3073242*

Yang, X., Chen, J., **Yedida, R.**, Yu, Z., & Menzies, T. (2021). Learning to recognize actionable static code warnings (is intrinsically easy). *Empirical Software Engineering, 26(3), 1-24.*

Yedida, R., & Menzies, T. (2021). On the Value of Oversampling for Deep Learning in Software Defect Prediction. *IEEE Transactions on Software Engineering, doi: 10.1109/TSE.2021.3079841*

Yedida, R., Krishna, R., Kalia, A., Menzies, T., Xiao, J., & Vukovic, M. (2021). Lessons learned from hyper-parameter tuning for microservice candidate identification. *Proceedings of the thirty-sixth IEEE/ACM International Conference on Automated Software Engineering (ASE).*

Yedida, R., Yang, X., & Menzies, T. (2021). Old but Gold: Reconsidering the value of feedforward learners for software analytics. *arXiv preprint arXiv:2101.06319*.

Saha, S., Nagaraj, N., Mathur, A., **Yedida, R.**, & Sneha, H. R. (2020). Evolution of novel activation functions in neural network training for astronomy data: habitability classification of exoplanets. *The European Physical Journal Special Topics, 229(16), 2629-2738*.

Yedida, R., Michael-Beasly, J., Korn, D., Abrar, S. M., Melo-Filho, C., Muratov, E., Graedon, J., Graedon, T., Chirkova, R., & Tropsha, A. (2020). Text Mining of the People's Pharmacy Radio Show Transcripts Can Identify Novel Drug Repurposing Hypotheses. *arXiv preprint arXiv:2011.07959*.

Yedida, R., Saha, S., & Prashanth, T. (2020). LipschitzLR: Using theoretically computed adaptive learning rates for fast convergence. *Applied Intelligence, 1-19*.

Sridhar, S., Saha, S., Shaikh, A., **Yedida, R.**, & Saha, S. (2020, July). Parsimonious Computing: A Minority Training Regime for Effective Prediction in Large Microarray Expression Data Sets. In *2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE*.

Khaidem, L., **Yedida, R.**, & Theophilus, A. J. (2019, November). Optimizing Inter-nationality of Journals: A Classical Gradient Approach Revisited via Swarm Intelligence. In *International Conference on Modeling, Machine Learning and Astronomy (pp. 3-14). Springer, Singapore*.

Funding

$5,000, Google Cloud Academic Research Grant, 2022.

Honors

2022, Google Cloud Research Innovator

Relevant Projects

Google/Meta Data Mining Feb. 2021 – May 2021
Python, Keras *GitHub*

Data science project to use Google Takeout and Meta user data to suggest products to advertise to a user from Amazon best sellers using DistilGPT-2 and achieved 0.6 F-1 score.

RAISE Aug. 2020 – Present
Python, Keras *GitHub :: PyPI*

Sole developer for a PEP8/PEP257-compliant, ML Python package used by our research lab and others for replicable results. Downloaded 12k times.

Personalized Chatbot May 2019
Python, Keras *GitHub*

Fine-tuned a GPT-2 345M model on 730k messages from Telegram logs to create a personalized chatbot.

Intelligent Tutoring System Sep. 2018 – May 2019
Python *GitHub*

Implemented an Intelligent Tutoring System backend using Bayesian Knowledge Tracing and a novel question selection algorithm.

Human Activity Data Project Oct. 2018 – Nov. 2018
Python, Keras *GitHub*

Collected personal activity data for 9 months, grouped tasks into 21 categories. Analyzed most productive hours of the day and built a 2-layer predictive LSTM model, achieving 42% top-5 accuracy.

Skills

Languages: Python, JavaScript, Java, C++

Frameworks: Flask, Keras, PyTorch, Node.js, React

Databases: SQL, MongoDB, DynamoDB