

Education

Ph.D., in Computer Science (GPA 3.9) at *Rochester Institute of Technology, NY*. Present
Related Courses in Data Science, Machine Learning, Deep Learning, and Quantitative Foundations

Bachelor of Science (Honors), in Computer Science at *University of Kelaniya, Sri Lanka* March 2017

Recent Research Experience

Fairness in AI - Population level annotator learning for human disagreement prediction August 2018 - Present
Rochester Institute of Technology, NY

Tools used - Python (Scikit-learn, Keras, and TensorFlow), BERT, MongoDB for data storage, and Google Cloud.

- Research to reduce cost and human resources for human annotation for supervised machine learning pipelines.
- Working with clustering methods (K-Means and LDA), and supervised learning methods (CNN).
- Presently working on modeling annotators from social media posts using neural models.
- Automated hyperparameter tuning across multiple models and instances. Tracked experiments remotely with WandB.
- Experiments on large datasets (approximately 50,000) from Reddit, Facebook, and Twitter on hate speech.
- Research work accepted at ECAI 2020* with new methods;
 - Semi-supervised (neighborhood based pooling) learning method for label pooling.
 - Statistical measures for the hyper-parameter selection for clustering and label pooling methods.

LKEMR - Cloud-based Electronic Medical Record System, Sri Lanka March 2018 - Present
Collaborative project with the Faculty of Medicine, and Colombo North Teaching Hospital, Sri Lanka. Tools used - PHP, MySQL, HTML, CSS, and UNIX. Python for data mining. Hosted on DigitalOcean and AWS.

- Web based electronic patient record management system, tailored for Sri Lanka using PHP backend, and MySQL database.
- EMR currently holds records for over 1,000 patients across several cities in Sri Lanka.
- Currently been used to identify potential patients with COVID-19, as it is the only EMR widely used in Sri Lanka.
- The work also contributes to open sourced EMR Project, Open-EMR.

Food Data Aggregation - Project to aggregate nutritional information for crowdsourced recipes from FDA Summer 2019
HEALS Project (part of IBM AI Horizon) at *Rensselaer Polytechnic Institute, NY*.

Tools used - Python for data collection and SpaCy for NLP aggregation.

- Research based on natural language processing and information retrieval with the 1M Recipe (1 million) dataset.
- Conducted research on aggregating food related data sources for food knowledge graph, which is used with IBM Watson.
- The tool was able to generate the nutritional content per FDA guidelines from a crowdsourced recipe.

KeyXtract:Twitter Model - Keyword Extraction Method using syntactic NLP rules. March 2016 - March 2017

Tools used - Stanford CoreNLP (Java) & NLTK (Python) for the keyword extraction & used Python for Data Mining.

- Collaborated with a domain expert from Department of English on gathering & building a model for analyzing Tweets.
- Evaluated measures using standard measures (F1 Score) and human based evaluation (A/B Test).
- Published research on Top-Tier Computer Science conferences in Sri Lanka*,
 - NITC 2016 (National Information Technology Conference). Only 16 papers selected for publication annually.

Recent Leadership Experience

President of the **Doctoral Students Association**, in *Rochester Institute of Technology, NY*. Present

- Raised \$1000 in funding to start a funding opportunity for Conference Attendance for PhD students.
- Represented PhD Student Population at;
 - Graduate Student Communication COVID-19 Task Force.
 - Assistant Dean search committee for the RIT Graduate School.

Infrastructure Lead at *EACL 2021 (European Chapter of the Association for Computational Linguistics)* Feb - April

- Built infrastructure for the virtual conference platform on Amazon AWS and integrated with live stream and Rocket.Chat.
- Integrated an external registration system with AWS Cognito user management for single sign on (OAuth2) through Auth0.
- Leading a team to manage requests from Conference chairs, conference sponsors, and attendees regarding the platform.

Skills

Data Management: Tableau, Jupyter Notebooks (Colab), and Excel

Programming: Python, R, Java, C, SQL, and PHP

Machine Learning: Matlab, Keras, TensorFlow, and sklearn

Cloud: Google Cloud, AWS, and DigitalOcean.

Research Interests

Machine learning, Label distribution learning, Deep learning, Data science, Algorithms, Statistics, Probabilistic learning, Human computation, Crowdsourcing, Information retrieval, Natural language processing, Sociolinguistics, NFT, and Blockchain.

Selected Recent Publications

Weerasooriya, T. C., Ororbia, A.G., & Homan, C.H. (2021), *Improving Label Quality by Jointly Modeling Items and Annotators*. arXiv preprint arXiv:2106.10600

Weerasooriya, T. C., Liu, T., & Homan, C.H. (2020). *Neighborhood-based Pooling for Population-level Label Distribution Learning*. 24th European Conference on Artificial Intelligence (ECAI 2020).

Mendis, K., **Weerasooriya, T. C.**, Withana, S., Liyanage, P., Silva, A. W., Wickramasinghe, R., & Weerabaddna, C. (2019). *Cloud-Based Open Source Primary Care Electronic Patient Record System for Sri Lankan Citizens*. 2019 National Information Technology Conference (NITC), pp. 41–46.

Weerasooriya, T., Perera, N., & Liyanage, S. R. (2017). *A Framework for Automated Corpus Compilation for KeyXtract : Twitter Model*. 2017 International Conference on Advances in ICT for Emerging Regions (ICTer), pp. 43–48.

Weerasooriya, T., Perera, N. and Liyanage, S.R. (2016). *A Method to Extract Essential Keywords from a Tweet Using NLP Tools*. 2016 International Conference in Advances in ICT for Emerging Regions (ICTer). pp. 29-34.

Perera, N., & **Weerasooriya, T.** (2016). *The Racecourse of Then and Now: Evolution of the Sri Lankan English Vocabulary Over Two Generations of SLE Speakers*. VISTAS Journal, 10, pp. 1–23..

Recent Work Experience

Adjunct at *Rochester Institute of Technology, NY*. Summer 2020

- Taught CSCI 635 - Introduction to Machine Learning.

Research Assistant for HEALS Project (part of IBM AI Horizon) at *Rensselaer Polytechnic Institute, NY*. Summer 2019

- Conducted research on NLP methods to aggregating food related data sources for food knowledge graph.
- Worked on a way to extract information from ADA guidelines based on the queries using doc2vec.

Intern at *Research and Development Team, Lesova Holdings Private Limited, Sri Lanka* February 2016 - December 2016

- Implemented and maintained customer relations management (CRM) System using Zendesk and Freshdesk.
- Designed and Developed Official Website for T-Shirts.LK Maldives Website.

Teaching Assistant at *Department of Computer Science, University of Kelaniya, Sri Lanka* March 2017 - August 2018

- Assisted in teaching Programming in C, Object Oriented Programming and Data Structures and Algorithms for approximately 150 students.

Research Assistant (Analytical Chemistry) at University of Peradeniya, Sri Lanka December 2014 - March 2015
Science and Technology Research Partnership for Sustainable Development Laboratory at *University of Peradeniya, Sri Lanka*

- Experienced in the analysis of TOC, BOD and COD in water samples collected from different lakes around Kandy.

Leadership Experience

HacKLN 2017, Hackathon organized by the *Department of Computer Science, University of Kelaniya, Sri Lanka*. 2017

- Guided the students as the teaching assistant in charge for the Department for Vidya 50, the 50th Celebratory Exhibition of Faculty of Science.

Gavel Club affiliated with *Toastmasters International, USA*. November 2014 - November 2017

- Organizing Committee Member of Inter University Best Speaker Contest 2014 - 2017.
- Designed, developed and maintained the official Club Website, Contest Website and the online account management system for the executive committee.

Career Fair organized by the *Career Guidance Unit, University of Kelaniya, Sri Lanka*. January 2015 - December 2016

- Designed, developed and maintained the event website and coordinated the registrations for 600+ undergraduates annually.

Awards and Certifications

Google Cloud Research Credit, awarded by *Google* 2020

Competent Gavelier Leader, awarded by *Toastmasters International USA* 2017

Competent Gavelier Communicator, awarded by *Toastmasters International USA* 2017

Social and Behavioral Research Certificate, by *Citi Program* 2018