Nikitha Rao

Institute for Software Research
School of Computer Science
Carnegie Mellon University, Pittsburgh, USA

nikithar@andrew.cmu.edu
https://raonikitha.github.io/
Google Scholar

Education

- **Carnegie Mellon University**, Pittsburgh, USA *August 2021*
  Ph.D. in Software Engineering.
  Advisor: Prof. Vincent Hellendoorn
- **PES University**, Bangalore, India *2015 - 2019*
  B.Tech in Computer Science and Engineering with a specialization in Data Science.
  Advisor: Dr. Gowri Srinivasa | Cumulative GPA: 9.48/10.0

Work Experience

- **Microsoft Research, India** - *Research Fellow* *July, 2019 - July 2021*
  Advisors: Chetan Bansal, Dr. Subho Mukherjee, Dr. Nachi Nagappan, and Dr. Tom Zimmermann
  Project Domains: Machine Learning for Software Engineering, Data Science, and Web Search
  Additional Responsibilities: Research Fellow representative for the Diversity and Inclusion committee.
- **Microsoft Research India** - *Research Intern* *January - June, 2019*
  Advisor: Dr. Sreangsu Acharyya
  Project Domain: Data Science
- **Carnegie Mellon University, Pittsburgh** - *Research Intern* *Summer 2018*
  Advisor: Dr. Shawn Blanton
  Project Domain: Machine Learning
- **Indian Institute of Science, India** - *Summer School Program* *July, 2017*
  Was among the youngest students selected for the 5th Summer School Program conducted by the Computer Science and Automation (CSA) Department.

Publications

- **Handling Class Imbalance with POISE: pAUC Optimization in Supervised Experiments** [preprint]
  Nikitha Rao, and Sreangsu Acharyya
  Preprint of full paper available. (9 pages)
  Best Short Paper Award at MLADS-SYNAPSE, 2020.
  Microsoft internal Conference on Machine Learning and Data Science for Asia-Pacific region
  [Acceptance Rate ≈ **8%**]

- **Search4Code: Code Search Intent Classification Using Weak Supervision** [paper]
  Nikitha Rao, Chetan Bansal, and Joe Guan
  *Mining Software Repositories (**MSR**), 2021.* (5 pages)

- **Neural Knowledge Extraction from Cloud Service Incidents** [paper]
  Manish Shetty, Chetan Bansal, Sumit Kumar, **Nikitha Rao**, Nachiappan Nagappan, and Thomas Zimmermann
  *International Conference on Software Engineering (**ICSE - SEIP**), 2021.* (12 pages)
  Featured on VentureBeat: Microsoft's SoftNER AI uses unsupervised learning to help triage cloud service outages.

- **Analyzing Web Search Behavior for Software Engineering Tasks** [paper]
  Nikitha Rao, Chetan Bansal, Thomas Zimmermann, Ahmed Hassan Awadallah, and Nachiappan Nagappan
  *IEEE International Conference on Big Data (**IEEE BigData**), 2020.* (10 pages)
  [Acceptance Rate ≈ **15.5%**]

- **Product Insights: Analyzing Product Intents in Web Search** [paper]
  Nikitha Rao, Chetan Bansal, Subhabrata Mukherjee, and Chandra Maddila
  *International Conference on Information and Knowledge Management (**CIKM**), 2020.* (4 pages)
  [Acceptance Rate ≈ **26%**]

- **Studying Ransomware Attacks Using Web Search Logs** [paper]
  Chetan Bansal, Pantazis Deligiannis, Chandra Maddila, and **Nikitha Rao** *(alphabetical order)*
  *International Conference on Research and Development in Information Retrieval (**SIGIR**), 2020.* (4 pages)
  [Acceptance Rate ≈ **30%**]

Patents

- **Identification of Content Gaps based on Relative User-Selection Rates between Multiple Discrete Content Sources** filed with the USPTO. *October 16, 2020*
  Co-inventors: Chetan Bansal, Junia George, Casey Gossard, Dung Nguyen, Dave Ludwig, and Curtis Anderson.
- **ExtraQuery Context-Aided Search Intent Detection** filed with the USPTO. *October 9, 2020*
  Co-inventors: Chetan Bansal, Joe Guan, Mark Wilson-Thomas, Nachiappan Nagappan, and Thomas Zimmermann.
- **Automatic Recognition of Entities Related to Cloud Incidents** filed with the USPTO. *June 19, 2020*
  Co-inventors: Manish Shetty, Chetan Bansal, Sumit Kumar, Nachiappan Nagappan, and Thomas Zimmermann.

Awards and Honors

- **Graduate Dean's Scholar Award**, Computer Science, UCLA *(declined in favor of CMU).* *2021*
- **Computer Science Excellence Fellowship**, Computer Science, UIUC *(declined in favor of CMU).* *2021*
- **Dean's Distinguished Graduate Fellowship**, Computer Science, UC Davis *(declined in favor of CMU).* *2021*
- **Best Short Paper Award** at MLADS-SYNAPSE. *2020*
- **Best Outgoing Student Award** for class of 2019 (360 students), Computer Science, PES University. *2019*
- Five time recipient of the **CNR Rao Scholarship**, Computer Science, PES University. *2016 - 2019*

Projects

- **Automatic Detection of Bugs During Code Review** *August, 2020 - July 2021*
  Advisors: Chetan Bansal, and Dr. Subho Mukherjee, *Microsoft Research*
  - Working on automatic detection of bug related pull request comments using weak supervision and few-shot learning.
  - We are building a taxonomy for various types of bugs that are found during code review.
  - Goal is to automatically detect bugs and generate bug fixes using machine learning models.

- **Partial-AUC Optimization to Handle Class Imbalance** *January, 2019 - July 2021*
  Advisor: Dr. Sreangsu Acharyya, *Microsoft Research*
  - Developed a novel partial AUC maximization algorithm to handle class imbalance using a game theory formulation.
  - We also developed a novel task-specific vector embedding technique that captures the geometry induced by decision trees to extend the algorithm to non-linearly separable data. This improved the overall performance by $\approx 40\%$.
  - We empirically showed that our approach significantly outperformed the existing baselines by 5% to 9% on learning to rank task and 16% to 67% in binary classification tasks in domains where the class imbalance is a prominent issue.

- **Code Search Intent Classification Using Weak Supervision** *June, 2020 - January, 2021*
  Advisors: Chetan Bansal, *Microsoft Research*
  - Worked on automatic detection of code search intent for C# and Java queries using weak supervision techniques.
  - We released 'Search4Code', the first web query dataset for code search.
  - This work has been accepted at *MSR, 2021.*

- **Search Insights: Analysing Web Search Behavior to Mine Insights** *August, 2019 - July, 2020*
  Advisors: Chetan Bansal, Dr. Subho Mukherjee, Dr. Nachi Nagappan, and Dr. Tom Zimmermann, *Microsoft Research*
  - Millions of search logs from Bing are analyzed to characterize user intent and behavior.
  - We studied the distribution of intents across various web search metrics and other trend characteristics.
  - The work done on ransomware attacks has been accepted at *SIGIR, 2020.*
  - The work done on product search has been accepted at *CIKM, 2020.*
  - The work done on software engineering tasks has been accepted at *IEEE BigData, 2020.*

- **Neural Knowledge Extraction from Cloud Service Incidents** *January, 2020 - July, 2020*
  Advisors: Chetan Bansal, Dr. Nachi Nagappan and Dr. Tom Zimmermann, *Microsoft Research*
  - We developed SoftNER, a framework for unsupervised knowledge extraction from cloud service incidents that leverages structural patterns for bootstrapping the training data.
  - We have built a novel multi-task learning based BiLSTM-CRF model that uses both the semantic context and data-types for named-entity extraction.
  - We show that using the knowledge extracted by SoftNER significantly helps improve model performance for important downstream tasks like incident triaging.
  - This work has been accepted at *ICSE - SEIP, 2021.*
  - This work has been featured on **VentureBeat** - *Microsoft's SoftNER AI uses unsupervised learning to help triage cloud service outages.*

- **Retinopathy of Prematurity – Feature Engineering and Predictive Analysis**. *August, 2018 - July, 2019*
  Undergraduate Thesis, Computer Science Department, PES University.
  Advisor: Dr. Gowri Srinivasa
  - Features extracted from the retinal images of prematurely born infants are used to build a rule-based model to automatically classify the severity of the disease.
  - This is in collaboration with Narayana Nethralaya (specialized hospital) and Rx Digi Health Platform (start-up).

- **Defining the Level of Hardware Obfuscation using Machine Learning Techniques** *June - July, 2018*
  Advisor: Dr. Shawn Blanton, *Carnegie Mellon University, Pittsburgh*
  - We analyze the patterns in input-output sequences of various obfuscated circuits to define a metric to quantify the level of obfuscation in a circuit using machine learning techniques.