Nikitha Rao

Research Fellow
Microsoft Research
Bangalore, India

t-nirao@microsoft.com
https://raonikitha.github.io/
Google Scholar

Education

- **PES University (previously known as PES Institute of Technology or PESIT)** *2015 - 2019*
  B.Tech in Computer Science and Engineering with a specialization in Data Science.
  Advisor: Dr. Gowri Srinivasa GPA: 9.48/10
  Won the **Best Student Award** for demonstrating academic excellence (out of 360 students).

Work Experience

- **Microsoft Research, India** - *Research Fellow* *July, 2019 - Present*
  Advisors: Chetan Bansal, Dr. Subho Mukherjee, Dr. Nachi Nagappan and Dr. Tom Zimmermann
  Project Domains: Machine Learning for Software Engineering, Web Search and Information Retrieval
  Additional Responsibilities: Research Fellow representative for the Diversity and Inclusion (DNI) 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

* - equal contribution

- **Code Search Intent Classification Using Weak Supervision**
  Nikitha Rao, Chetan Bansal, Joe Guan
  Under review at ICSE - NIER, 2021.
- **Neural Knowledge Extraction from Cloud Service Incidents** [arXiv]
  Manish Shetty, Chetan Bansal, Sumit Kumar, **Nikitha Rao**, Nachiappan Nagappan and Thomas Zimmermann
  Under review at ICSE - SEIP, 2021.
  Featured on VentureBeat - Microsoft's SoftNER AI uses unsupervised learning to help triage cloud service outages.
- **Handling Class Imbalance with POISE: pAUC Optimization in Supervised Experiments**
  Nikitha Rao and Sreangsu Acharyya
  Best Short Paper Award at MLADS-SYNAPSE, 2020 **(Presenter)**
  Microsoft internal Conference on Machine Learning and Data Science for Asia-Pacific region
  [Acceptance Rate ≈ 8%]
- **Analyzing Web Search Behavior for Software Engineering Tasks** [arXiv]
  Nikitha Rao, Chetan Bansal, Thomas Zimmermann, Ahmed Hassan Awadallah and Nachiappan Nagappan
  *IEEE International Conference on Big Data (**IEEE BigData**), 2020* **(Presenter)**.
  [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* **(Presenter)**.
  [Acceptance Rate ≈ 26%]
- **Studying Ransomware Attacks Using Web Search Logs** [paper]
  Chetan Bansal*, Pantazis Deligiannis*, Chandra Maddila* and **Nikitha Rao***
  *International Conference on Research and Development in Information Retrieval (**SIGIR**)*, 2020 **(Presenter)**.
  [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, 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, 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, Thomas Zimmermann.

Projects

- **Automatic Detection of Bugs During Code Review** *August, 2020 - Present*
  Advisors: Chetan Bansal and Dr. Subho Mukherjee *Microsoft Research*
  - Build a taxonomy for bug related pull request comments.
  - Build a dataset of buggy and bug fixed code for the major categories of bugs.
  - Automatic bug detection and bug fix generation using machine learning models.
- **Code Search Intent Classification Using Weak Supervision** *June, 2020 - Present*
  Advisors: Chetan Bansal and Dr. Nachi Nagappan *Microsoft Research*
  - Automatic detection of code search intent for natural language queries using weak supervision techniques.
  - Build a data set of natural language queries having code intent for C# and Java.
  - Paper under review at *ICSE - NIER, 2021.*
- **Partial-AUC Optimization to Handle Class Imbalance** *January, 2019 - Present*
  Advisor: Dr. Sreangsu Acharyya, *Microsoft Research*
  - Worked on a partial AUC maximization algorithm to handle class imbalance that uses a game theory formulation for the cost function.
  - The novel use of learned decision tree paths as vector embeddings captures any inherent non-linearity in the data.
  - We demonstrate empirically that our approach is both fast and scalable. We also show that our method not only performs well for learning to rank tasks but can also generalize to other problem spaces where partial AUC or accuracy at the top is crucial.
- **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 analysed to characterize user intent and user behavior.
  - A study on distribution of intents across various web search metrics and other trend characteristics is conducted.
  - 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 have built SoftNER, a framework for unsupervised knowledge extraction from cloud service incidents that leverages structural patterns for bootstrapping the training data.
  - We build 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.
  - Paper under review 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 and Engineering, 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 identify the stage of the disease.
  - This is in collaboration with Rx Digi Health Platform, a start-up based in Bangalore, India.
- **Defining the Level of Hardware Obfuscation using Machine Learning Techniques** *June - July, 2018*
  Advisor: Dr. Shawn Blanton, *Carnegie Mellon University, Pittsburgh*
  - Analysis of various patterns in the input-output sequences of various obfuscated circuits to define a metric to quantify the level of obfuscation in a circuit using machine learning techniques.

Selected Awards and Honors

- **Best Short Paper Award** at **MLADS - SYNAPSE '20**, a Microsoft internal Conference on Machine Learning and Data Science for Asia-Pacific region *2020*
- Won the **Best Student Award** for demonstrating academic excellence, for the graduating class of 2019, in the Computer Science Department, PES University. *2019*
- Five time recipient of the **CNR Rao Scholarship** for academic performance in the Computer Science Department, PES University. *2016-2019*