VRAJ SHAH

Phone: (+1) 858-291-2697   *Web:* pvn25.github.io
Email: vps002@eng.ucsd.edu   *LinkedIn*: linkedin.com/vraj

EDUCATION	**University of California**, San Diego, CA	Sept 2016 - March 2022 (Expected)

PhD, Computer Science.
Thesis Advisor: Prof. Arun Kumar

Indian Institute of Technology, Indore, India   Aug 2012 - June 2016
Bachelor of Technology, Computer Science & Engineering.

RESEARCH EXPERIENCE

Graduate Student Researcher   Sept 2016 - Present
University of California, San Diego

- **SortingHat.** Automate data preparation tasks for AutoML by formalizing and benchmarking them to objectively validate and improve AutoML platforms.
- **SpeakQL.** Developed a system for making spoken SQL querying effective and efficient. The speech driven interface allows the users to query in any domain with infinite vocabulary using interactive query correction.
- **Hamlet.** Analyzed the accuracy effects of joins on machine learning algorithms, when learning over normalized data to improve performance and usability for analysts.

Research Assistant   May 2015 - July 2015
University of Alberta, Canada

- Developed a Big Data Adaptor as Eclipse plugin which efficiently handles GitHub's data dump. The tool makes the data more appropriate for a variety of downstream purposes like analytics.

INDUSTRY EXPERIENCE

Research Intern   June 2018 - Sept 2018
Microsoft

- Implemented computational graph-level optimizations inside Microsoft's deep learning system for the inference task.
- Implemented analytical cost model for Machine Learning operators in the inference serving system and used learning to predict the operator execution time.

Research Intern   June 2017 - Sept 2017
Infor Corporation

- Worked with the analytics team to integrate machine learning algorithms inside the existing LogicBlox forecasting engine.

PUBLICATIONS

Towards Benchmarking Feature Type Inference for AutoML Platforms.
Vraj Shah, Jonathan Lacanlale, Premanand Kumar, Kevin Yang, Arun Kumar.
SIGMOD 2021 | Paper.

SpeakQL: Towards Speech-driven Multimodal Querying of Structured Data.
Vraj Shah, Side Li, Arun Kumar, Lawrence Saul.
SIGMOD 2020 | Paper.

Demonstration of SpeakQL: Speech-driven Multimodal Querying of Structured Data.
Vraj Shah, Side Li, Kevin Yang, Arun Kumar, Lawrence Saul.
SIGMOD (*Demo track*) 2019 | Paper.

The ML Data Prep Zoo: Towards Semi-Automatic Data Preparation for ML.
Vraj Shah, Arun Kumar.
DEEM Workshop, SIGMOD 2019 | Paper.

SpeakQL: Towards Speech-driven Multi-modal Querying.
Vraj Shah.
Student Research Competition, SIGMOD 2019 | Awarded *Second Runner-up* | Paper.

Are Key-Foreign Key Joins Safe to Avoid when Learning High Capacity Classifiers?
Vraj Shah, Arun Kumar, Xiaojin Zhu.
VLDB 2018 | Paper.

SpeakQL: Towards Speech-driven Multi-modal Querying.
Dharmil Chandarana, **Vraj Shah**, Arun Kumar, Lawrence Saul.
HILDA Workshop, SIGMOD 2017. | Paper

GitHub's Big Data Adaptor: An Eclipse Plugin.
Ali Sajedi, **Vraj Shah**, Eleni Stroulia.
IBM CASCON 2015. | Paper

RESEARCH IMPACT	Improving Feature Type Inference Accuracy of TFDV with SortingHat **Vraj Shah**, Kevin Yang, and Arun Kumar. Technical Report.
	Models from project SortingHat explored for production use by TensorFlow Data Validation in collaboration with Google
PATENTS	Speech Based Structured Querying Arun Kumar, **Vraj Shah**, Dharmil Chandarana

AWARDS

Second Runner-up, ACM SIGMOD Student Research Competition	2019
NSF Travel Award to attend ACM SIGMOD	2019
NSF Travel Award to attend VLDB	2018
Microsoft Travel Award to attend ACM SIGMOD	2017
Research Experience program honor for poster at *University of Alberta*	
International Research Symposium	2015
MITACS Globalink Research Award	2015
DAAD WISE Fellowship	2015

SERVICE

Program Committe: VLDB 2022

External Reviewer: VLDB 2019, VLDB Demo 2018

SKILLS

Languages: C, C++, Java, Python, R , SQL.

Web Development: HTML, CSS, JavaScript, PHP.

Tools and Libraries: Scikit-learn, Keras, Tensorflow, Matlab, Eclipse, Git.

TEACHING EXPERIENCE	Teaching Assistant - Advanced Data Analytics Systems	Winter 2020
	Teaching Assistant - Database System Implementation	Spring 2020
MENTORSHIP EXPERIENCE	Francisco Cornejo-Garcia, BS, Cypress College	*Summer 2020*
	Kevin Yang, BS, UCSD.	*Fall 2019 - Spring 2020*
	Jonathan Lacanlale, BS, California State University, Northridge	*Summer 2019*
	Thomas Parashos, BS, California State University, Northridge	*Summer 2019*