

## VRAJ SHAH

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

3230 EBU3B CSE  
9500 Gilman Drive  
La Jolla, CA 92093

EDUCATION	<b>University of California</b> , San Diego, CA <i>PhD (In transfer from MS)</i> , Computer Science. GPA: 3.74/4. Thesis Advisor: Arun Kumar	Sept 2016 - Present
	<b>Indian Institute of Technology</b> , Indore, India <i>Bachelor of Technology</i> , Computer Science & Engineering. Major GPA: 9.47/10.	Aug 2012 - June 2016
EXPERIENCE	<b>Graduate Student Researcher</b> <i>University of California, San Diego</i>	Sept 2016 - Present
	<ul style="list-style-type: none"><li>• <b>SpeakQL</b>: Developing 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. Project Website: <a href="https://adalabucsd.github.io/speakql">https://adalabucsd.github.io/speakql</a></li><li>• <b>Hamlet</b>: Analyzed the accuracy effects of joins on machine learning algorithms, when learning over normalized data to improve performance and usability for analysts. Project Website: <a href="https://adalabucsd.github.io/hamlet">https://adalabucsd.github.io/hamlet</a></li></ul>	
	<b>Software Engineer, Intern</b> <i>Infor Corporation</i>	June 2017 - Sept 2017
	<ul style="list-style-type: none"><li>• Worked with the analytics team to integrate machine learning algorithm - neural networks inside the existing LogicBlox forecasting engine.</li></ul>	
	<b>Research Assistant</b> <i>University of Alberta, Canada</i>	May 2015 - July 2015
	<ul style="list-style-type: none"><li>• 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.</li><li>• Developed a tool in python that identifies influential communities in social network platforms like Twitter, GitHub and Stack Overflow to design effective advertisement campaigns.</li></ul>	
PUBLICATIONS	<i>Are Key-Foreign Key Joins Safe to Avoid when Learning High Capacity Classifiers?</i> Vraj Shah, Arun Kumar, Xiaojin Zhu. In Proceedings of Very Large Data Bases Conferences, VLDB 2018.	
	<i>SpeakQL: Towards Speech-driven Multi-modal Querying.</i> Dharmil Chandarana, Vraj Shah, Arun Kumar, Lawrence Saul. In Proceedings of Human In Loop Data Analytics, SIGMOD 2017.	
	<i>GitHub's Big Data Adaptor: An Eclipse Plugin.</i> Ali Sajedi, Vraj Shah, Eleni Stroulia. In Proceedings of 25th Annual International Conference on Computer Science and Software Engineering. IBM CASCON 2015.	
PROJECTS	<b>Amazon Helpfulness Prediction</b> : Developed a recommender system for Amazon reviews using random forest regressor that predict whether the review of an item will be considered helpful.	
	<b>Real Time Twitter Opinion Mining and Visualization</b> : Used naive Bayes Classifier and Dictionary based approach to perform sentiment analysis of government policies	

posted on Twitter. Handled challenging scenarios like sarcasm with more than 70 % accuracy.

**Open Access Academic Journal System:** Developed a web-based journal management and publishing system that allows submission of articles through a peer review process.

## SKILLS

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

**Web Development:** HTML, CSS, JavaScript, PHP.

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

## AWARDS

Microsoft Travel Award to attend SIGMOD 2017.

University of Alberta Research Experience program honor for poster presentation at *University of Alberta International Research Symposium* 2015.

MITACS Globalink Research Award 2015.

DAAD WISE fellowship 2015.

## POSTERS

*Are Key-Foreign Key Joins Safe to Avoid when Learning High Capacity Classifiers?*  
UCSD Research Expo 2017

*SpeakQL: Towards Speech-driven Multi-modal Querying.*  
ACM SIGMOD 2017

*GitHub's Big Data Adaptor: An Eclipse Plugin.*  
University of Alberta Research Symposium 2015

## TALKS

*SpeakQL: Towards Speech-driven Multi-modal Querying.* June 2017  
ACM SIGMOD, UCSD Database Seminar.

## REFERENCES

Arun Kumar  
Assistant Professor, University of California, San Diego  
arunkk@eng.ucsd.edu