

# Dhawal Joharapurkar

Contact Information 485, Engineering 2  
Computer Science Department  
University of California, Santa Cruz  
Santa Cruz 95060  
**email:** dhawal@ucsc.edu  
**web:** <https://dhawaljoh.github.io>

Education	<b>University of California</b> , Santa Cruz Ph.D. in Computer Science <b>Advisor:</b> Prof. Lise Getoor	2016 - Present
-----------	--	----------------

**Manipal Institute of Technology**, Manipal 2011 - 2015  
B.Tech in Computer Science & Engineering  
**Bachelor Thesis:** Temporal Scoping of Facts in a Knowledge Base

Work Experience **Xerox Research Centre India**, Bangalore June '15–Jan '16  
 Research Intern, Machine Learning and Statistics Group  
 Mentors: Dr. Vaibhav Rajan and Sumit Negi

- Analysis of clinical notes of ICU patients using Topic Models
- Created a plotting tool to highlight different statistics across classes

**Indian Institute of Science**, Bangalore December '14–May '15  
Project Trainee, Supercomputer Education and Research Centre (SERC)  
Mentor: Prof. Partha Talukdar

- Temporal scoping and ordering of relations in a knowledge base
- Entity linking and disambiguation in large text corpora

**Indian Institute of Technology**, Kharagpur May '14–June '14  
 Summer Research Intern, Dept. of Computer Science & Engineering  
 Mentor: Prof. Sudeshna Sarkar

- Implemented DBSCAN algorithm to find traffic stoppage points and segmented roads based on their speed profiles – automatic profiling of drivers
- Modified “simplekml” Python module to plot the GPS points on Google Maps

**DataWeave Software Pvt. Ltd.**, Bangalore May '13–Jun '13  
 Summer Intern

- Created data crawlers using Python that aggregated and stored content in JSON dumps
- Content available via APIs, a few listed at <http://dataweave.in/apis>

Projects	<b>Question Category Classification</b>	July '15
	<b>Mentor:</b> Prof. Manish Shrivastava	IASNLP, IIIT, Hyderabad
	Question category classification on TREC data using topic composition features generated by LDA.	

<b>Detecting Fibrous Regions in Protein Sequences</b>	November '13 – May '14
<b>Mentor:</b> Dr. Smitha Nair	Manipal Institute of Technology, Manipal
Detection of fibrous regions in protein sequences using Support Vector Machines and Bee Colony Optimization for PCA.	

**Online competition: Photo Tagger – Multi-class classification** March '14  
**Leaderboard:** 84 out of 644 CSA, IISc, Bangalore  
 SVM whose parameters were optimized by GridSearchCV, to classify photos into various classes (people, cars, shoes, buildings, flowers).

**Craigslist Post Classification** October '13  
**Accuracy:** 81% Manipal Institute of Technology, Manipal  
 Vectors created by bag of words model and then tf-idf, SVM to classify posts on Craigslist into sections based on the product description.

Publications	<b>Online Adspace Posts' Category Classification</b> Dhawal Joharapurkar, Vaishak Salin, Vishal Krishna 12 <sup>th</sup> International Conference on Natural Language Processing, December, 2015	
Talks	<b>From Big Text to Big Knowledge</b> SERC Open Day 2015	Feb '15 IISc, Bangalore
Honors	<b>Regents' Fellowship</b> , UC Santa Cruz	2016-17
Online Courses	<b>The Data Scientist's Toolkit</b> coursera.org, 100%	June '14 Johns Hopkins University
	<b>Design and Analysis of Algorithms</b> Massively Empowered Classrooms, 100%	May '14 Microsoft Research
	<b>Machine Learning</b> coursera.org, 100%	April '14 Stanford University
	<b>Algorithms: Design and Analysis, Part 1</b> coursera.org, 98%	July '13 Stanford University
Skills	<ul style="list-style-type: none"> <li>• <b>Languages:</b> C/C++, Python, Octave, SQL, L<sup>A</sup>T<sub>E</sub>X</li> <li>• <b>Operating Systems:</b> UNIX, Microsoft Windows</li> <li>• <b>Tools:</b> Vim, Sublime Text, Enthought Canopy, IPython</li> <li>• <b>Version Control System:</b> git</li> </ul>	