Shivani Rao

shivaniraoweb@gmail.com
www.shivanirao.info
http://myresearchdiaries.blogspot.com

Objective	To solve challenging machine learning problems and data problems in various domains of Data Science at a data-driven company that values teamwork, growth and learning. Outside out of the R&D work, I love writing, teaching, mentoring and giving talks on topics that I am most passionate about.
Areas of Interests	Recommender Systems, Information Retrieval and Text Mining, Large Scale Machine Learning, Real-time analytics, Time Series Analysis, Content Understanding, Natural Language Processing, Speech and Vision related data problems
Highlights	• Strong background in Machine Learning Algorithms, Probability and Statistics, Bayesian Inference. • Extensive experience in Data acquisition, data cleaning and data analysis, feature extraction • Ability to apply existing ML algorithms to new domains, and survey competing approaches
Technical Skills	Languages: C, C++, R, Python, Matlab, (Beginner level in) Scala, Pig, Hive, SQL Technologies: Hadoop, Spark, (Familiarity with) Kafka, Storm and OpenTSDB, Airpal
Work Experience	**Senior Applied Researcher, Lynda Relevance Team @ LinkedIn, San Francisco, CA** **Nov'15 – Present** In this role, I am foraying into the field of Machine Learning for Ed-Tech (Education Technology), in particular, I will be working on Recommender Systems. **Software Engineer, Analytics and Machine Learning Team @Box Inc, Los Altos, CA,** **Nov '13 –Oct ' 15** Involved in design and implementation of Machine Learning algorithms to improve user experience for Box customers. Our algorithms are being developed and testing using Apache Spark compute framework **Research Assistant, Robot Vision Lab @ Purdue University, West Lafayette, IN** **May '08 – Oct '13** Applied Machine Learning and Information Retrieval algorithms to build software engineering tools. Proposed, Experimented with and Published novel approaches to improving the efficiency and effectiveness of code-search tools Presented quarterly report on findings and progress to the funding agency **Software Engineering Intern, Street-View Team @ Google Inc, Mountain View, CA** **May '10 –Aug '10** Built and Tested a prototype of the solution to large-scale Loop-Closing problem to correct the error in GPS data using Computer Vision Algorithm. Tested the solution on the street-view data captured in cities using MapReduce. **Software Engineer, Video-codec Team @ Nvidia Inc, Pune, India** **Aug '06 – June '07** Develop and debug drivers to support streaming and playback of H.264 and Blu-ray formats on Windows XP with Nvidia GPUs
Education	**PhD, Electrical and Computer Engineering, Purdue University, IN, USA** **Jan '08 -Dec '13,** CGPA – 3.92/4.0 Areas: Information Retrieval, Machine Learning, Data Mining, Software Engineering **MS, Computer Science and Engineering, IIT Madras, Chennai, India** **Jan '04 – May '06,** CGPA – 9.8/10 Areas: Shape from Texture, Texture Analysis, Image Processing, Computer Vision **BS Information Technology, University of Mumbai, Mumbai, India** **Aug '99 – Aug '03,** Grade-Distinction Areas: Computer Vision, Robotics, Simulation and Modeling, Algorithms, Data Structure

Relevant Projects	**Lynda Course Recommendations @ LinkedIn (Current)** At LinkedIn, I will be foraying into the world of Machine Learning for Education Technology. In particular, I will be involved in researching and building (and possibly publishing about) recommender systems to help improve "learner" engagement. Technologies: Pig, Scalding **Search Relevance Analysis @ Box.com (A primarily Data Mining Project)** Search logs are the primary source of information that reveals the quality of a search engine and the level/kind of engagement of a user and the search tool. I mined the search logs, used mining and machine learning techniques to compute optimal session length. The sessionized logs were then used to compute basic engagement/search quality metrics. Technologies used: Hive, Python, SQLite, Postgres SQL. **DLP using Machine Learning@ Box (Machine Learning project)** Data Loss Prevention (DLP) means protecting sensitive information (documents containing Personal Identification Information like credit card numbers, SSN, sales numbers of an organization). The current DLP systems use rudimentary algorithms based on regular expression based matching or keyword-based matching to identify these documents. As a part of this project, I developed and built statistical modeling techniques to detect confidential documents. Technologies used: Apache Spark, Python **Information Retrieval based Bug Localization, (PhD) (Data Mining, Machine Learning, Information Retrieval)** During my PhD, I worked on the problem of bug localization – locating source files related to bugs search-based tools. I mined the commit history and bug tracking system of popular open source software tools to create our gold datasets. I researched various ML models, text-mining approaches, proposed ways to evaluate and improve retrieval accuracy and efficiency of these search tools. Several publications and public datasets resulted from this work. Technologies: R, git, perl, Python, Matlab
Personal Projects	• Anomaly Detection (analyzed several techniques to detect anomalies in aggregates like average number of files downloaded from Box users and so on) • Churn Analysis (to predict whether an enterprise is going to churn their Box account) • Gibbs Sampler Implementation (to learn the topic model represented by LDA) • Large scale SVD using Gensim (for incremental update of model with evolving data) • Experimented with algorithms for Automatic Keyword Extraction (RAKE, TextRank) and also played with word2vec tool of Gensim as an additional feature.
Publications	Please visit www.shivanirao.info/publications for the full list of publications
Teaching Experience	Computer Vision, Pattern Classification, Digital Signal Processing, Computer System Design, Computer Graphics (at Purdue and IIT Madras)
Community Involvement	**Program Committee Member:** ICSME (International Conference on Software Maintenance and Engineering) Tools Demo, ICPC (International Conference on Program Comprehension) ERA track, SEDE (Software Engineering and Data Engineering), GHC (Grace Hopper Celebration of Women in Computing), GHC India (Grace Hopper India) **Book Review:** Machine Learning for R, PACKT publication **Industry Talks:** • Panel on Apache Spark at GHC, Oct 2015, • Technovation World Pitch Conference, June 2015 • Student Opportunity Lab at GHC, Oct 2014 **Women in Tech related**: Bay Area Girl Geek Dinner of 2015, Box Representative at the HackBright Academy's Career Day summer 2014, Mentor at the Introduce a Girl to Engineering Day (IGED) at Purdue University's Women in Engineering Program (WIEP) Feb 2009, Feb 2011