

# Sauptik Dhar

Robert Bosch LLC  
384 Santa Trinita Ave  
Sunnyvale, CA 94085

Cell : +1 (612) 876-6429  
Email: [sauptik.dhar@gmail.com](mailto:sauptik.dhar@gmail.com)  
Website: <http://sauptik.000webhostapp.com>

## OBJECTIVE

Research and develop machine learning algorithms for analysis of large-scale, high-dimensional data from various business domains (viz., industrial, automotive, healthcare, education etc.)

## EDUCATION

**PhD** (Aug 2008 – Jan 2014): Electrical and Computer Engineering, University of Minnesota.

**Thesis:** Analysis and Extensions of Universum Learning.

**Areas of Interest:** Machine Learning, Optimization Theory.

**B. Tech** (2002-2006): Electronics and Telecommunication Engineering, N.I.T, Silchar, India.

## INDUSTRIAL EXPERIENCE & COLLABORATIONS

### 1. ROBERT BOSCH LLC

**Position:** Senior Research Scientist (Jan 2016 - Present)

Use-Cases:

- *Autonomous Driving:* Lead the effort for developing scalable HD maps for autonomous cars.
- *Thermotechnology:* Research and Development of predictive models for event prediction pertaining to boiler failures. The predictive models guaranteed the KPI measures set by the Business Unit.

**Position:** Research Scientist (Dec 2013 – Dec 2016)

Use-Cases:

- *Automotive:* Research and Development of predictive models for optimal content of vehicle diagnostic tools.
- *Healthcare:* Research and Development of cost-savings model towards Patient Identification for Tele-health Programs. Research results published in international conferences.

**Position:** Research Intern (May 2013 – Sept 2013)

Evaluation of various Big Data Tools in current market, and provide recommendation for the choice of scalable tools for big-data problems within Bosch. The work lead to > \$500,000 license agreement between Bosch and Big Data Tools providers.

### 2. MINNESOTA DEPARTMENT OF TRANSPORT (July 2008-June 2010)

**Project:** Statistical Analysis of the Soil Chemical Survey Data

**Position:** Research Assistant

**Technology:** Matlab, C++

**OS:** Windows.

**Description:**

- Data-Analytic Modeling of the Minnesota Soil Chemical data produced by the 2001 metro soil survey and by the 2003 state-wide survey.
- Visualization of the soil chemical concentration trends within the metro area and in the state of Minnesota using a machine learning tool called "Self Organizing Map" (SOM)

### 3. 3M (May 2009-Aug 2009)

**Project:** Product Uniformity Analysis from Video Images

**Position:** Research Assistant

**Technology:** Matlab.

**OS:** Windows/UNIX.

**Description:** Uniformity Analysis from Video images. The designed software is used for quality management based on Video images.

### 4. WIPRO TECHNOLOGIES (June 2006- May 2008)

**Project:** Optical Manager 5.0

**Position:** Project Engineer (Module lead)

**Technology:** C, C++, Java, SQL.

**OS:** Solaris, HP-UX

**Description:** Module lead for the Network Monitoring System (NMS) for High-Density Cross Connect Network Elements supporting 1.2 Tbps connections at backbone networks.

## RESEARCH EXPERIENCE

*Research Scientist* (Dec 2013- Present): Bosch Center for Artificial Intelligence, Robert Bosch LLC

1. Scalable Machine Learning for Big Data.
2. Advanced Kernel machines for high-dimensional data.
3. Scalable HD maps for autonomous cars

*Research Assistant* (Aug 2008- Dec 2013): Predictive Learning Research Group, University of Minnesota.

1. Learning under Universum Settings.
  - Classification (Standard/Cost-Sensitive/Multi-class), Regression, Single Class (novelty detection) etc.
2. Understanding Black Box Data-Analytic Models: Provide techniques to interpret and visualize the black-box data analytic models like, Support Vector Machines (SVM), Random Forests, and Advanced Kernel Methods.

## RESEARCH PUBLICATIONS

### JOURNAL PAPERS/BOOK CHAPTERS

- [1] V. Cherkassky, **S. Dhar**, "Interpretation of Black-Box Predictive Models". *Measures of Complexity: Festschrift for Alexey Chervonenkis* (Editors: V. Vovk, H. Papadopoulos, A. Gammerman), Oct 2015.
- [2] **S. Dhar**, V. Cherkassky, "Development and Evaluation of Cost-Sensitive Universum SVM", *IEEE Transactions on Systems, MAN, and Cybernetics PART B: Cybernetics*, vol. 45, no. 4, pp. 806-817, Apr 2015.
- [3] A. Jonson, E. Dickson, H. Shiao, V. Cherkassky, **S. Dhar**, L. Downs Jr, "Machine learning as a tool to predict survival outcomes for carcinosarcoma of the female genital tract", *Gynecologic Oncology*, Volume 123 (2), November 2011.
- [4] V. Cherkassky, **S. Dhar**, and W. Dai, "Practical Conditions for Effectiveness of the Universum Learning," *IEEE Transactions on Neural Networks*, vol.22, no. 8, pp. 1241-1255, Aug 2011.

### CONFERENCE PAPERS

- [1] Y. Park, **S. Dhar**, S. Boyd, M. Shah, "Variable Metric Proximal Gradient Method with Diagonal Barzilai-Borwein Step size", NIPS 2017 (Optimization for Machine Learning).
- [2] **S. Dhar**, V. Cherkassky, "Universum Learning for SVM Regression", IJCNN 2017.
- [3] **S. Dhar**, N. Ramakrishnan, V. Cherkassky, M. Shah, "On Multiclass Universum Learning", NIPS 2016 (Learning in High-Dimensions with Structure).
- [4] **S. Dhar**, C. Yi, N. Ramakrishnan, and M. Shah, "ADMM based Scalable Machine Learning on Spark", *IEEE Big Data 2015*, Santa Clara, 2015.
- [5] M. Ganser, **S. Dhar**, U. Kurup, C. Cunha, and A. Gacic, "A Data-Driven approach towards Patient Identification for Tele-health Programs", *IEEE Big Data 2015*.
- [6] M. Ganser, **S. Dhar**, U. Kurup, C. Cunha, and A. Gacic, "Patient Identification for Telehealth Programs", ICMLA 2015.
- [7] **S. Dhar**, and V. Cherkassky, "Cost-Sensitive Universum-SVM", *ICMLA*, 2012.
- [8] **S. Dhar** and V. Cherkassky, "Market Timing of International Mutual Funds: A Decade after the Scandal", *IEEE Computational Intelligence for Financial Engineering & Economics*, 2012.
- [9] **S. Dhar**, V. Cherkassky, "Practical Analysis of the Universum SVM Learning", *Snowbird Learning Workshop*, April 2011.
- [10] **S. Dhar**, V. Cherkassky, "Application of SOM to Analysis of Minnesota Soil Survey Data", IJCNN 2011.
- [11] V. Cherkassky, **S. Dhar**, "Simple Method for Interpretation of High-Dimensional Nonlinear SVM Classification Models", *Proceedings of the 2010 International Conference on Data Mining*, July 2010.
- [12] **S. Dhar**, V. Cherkassky, R. Edstrom, J. Seaberg, S. Hennes, "Exploring the Pattern of Clustering within Minnesota Soil Survey Data using Self Organizing Maps", *21st Annual CTS Transportation Research Conference*. River Centre, St. Paul, MN. April 2010.
- [13] **S. Dhar** and M. Paul, "Adaptive Pseudo-OFDM based WLAN Systems using Neural Network", *IEEE International conference on Wireless Communication & Sensor Networks*, December 17-19, 2006.

### TECHNICAL REPORTS/POSTERS/TUTORIALS/TALKS

- [1] S. Dhar, M. Shah, "ADMM based Scalable Machine Learning on Apache Spark", Spark Summit 2017.
- [2] **S. Dhar**, N. Ramakrishnan, V. Cherkassky, M. Shah, "Universum Learning for Multiclass SVM"
- [3] V. Cherkassky and **S. Dhar**, "Advances in Universum Learning", IJCNN 2015. (tutorial)

- [4] S.Dhar et. al, “ADMML: ADMM based ML Toolkit on Spark” <https://github.com/DL-Benchmarks/ADMML> (Report)
- [5] S. Dhar, V. Cherkassky, “Statistical Analysis of the Soil Chemical Survey Data”, Report no. Mn/DOT 2010-22, June 2010. (Technical Report)

## PROFESSIONAL ACTIVITIES

- **Associate Editor (Journal)** : Neural Processing Letters (2016-current)
- **Program Committee (Conferences)**: KDD (2016-2017), ICMLA (2016-2017), SDM 2018, HIS 2017, Baylearn 2017, ICPRAI 2018.
- **Reviewer (Journals)**: Neural Networks, Neural Processing Letters, Pattern Recognition, Neurocomputing, PLoS ONE, IEEE Systems, Man and Cybernetics.
- **Reviewer (Conferences)**: IJCNN (2009 – 2017), NIPS 2016, KDD 2016, ICWSM 2015

## AWARDS & HONORS

- Gold Medalist in the Mathematics Olympiad (CAT-II) 2001 organized by the Assam Academy of Mathematics.
- State Highest in Mathematics, First Year, Higher Secondary Council, Assam.
- Second Prize (Silver Medal) in the Mathletics Competition 2000 organized by the Assam Academy of Mathematics.

## SOFTWARE SKILLS

**OS:** MS-DOS, Windows, UNIX (Solaris/HP-UX), CentOS.

**Programming / Simulation Languages:** MATLAB, R, Python, Java, Scala.

**Data Base / Storage Systems:** Hadoop (HDFS), Oracle, MySQL, MongoDB.

**Platforms / Software Management Tools:** Source Insight / Eclipse / Netbeans (Java), Komodo/Spyder (Python), RStudio (R), IBM Clear Case/Git (Code Management), Navicat (database).

### Research Relevant Tools:

- *Machine Learning*: Tensorflow, Weka, Spider, STPRtool, PRTools, UniverSVM, libSVM, SVMlight, SOM Toolbox etc.
- *Optimization Theory*: CVX, YALMIP, SeDuMi, SDPT3, CPLEX, TOMLAB.
- *Big Data Tools*: Alpine, Revolution R, MLlib (Spark), PhotonML.

## REFERENCES

Available on request.