Choudur K. Lakshminarayan
9390 Research Boulevard
Teradata Labs
Austin, TX 78759
512-993-3121 (o)
512-983-3879 (c)
choudur.lakshminarayan@gmail.com

Fields of Interest

Research Areas

Large-Scale Machine Learning, Mathematical Statistics, Applied Mathematics, Biostatistics, Statistical Signal Processing, Data Mining, Text Mining, Machine Learning, Parallel Computing and Industrial Mathematics

Application Areas

Highly Scalable Algorithms, Sensor Analytics, Healthcare and Pharmaceuticals, Energy, Semiconductor manufacturing, Database Technology, Large Scale Data centers, Finance, Marketing Optimization, Web analytics, Digital Marketing, Cyber-Security, Clinical trials and Pharma

High Dimensional Statistics, Sparsity and Regularization, Numerical Methods (Gradient Descent methods and Function Approximations, Kernel Methods, Bayesian Learning, Time Series)

Academic Degrees

Doctor of Philosophy, Mathematical Sciences, The University of Texas, 1990
Master of Science, Mathematical Sciences, The University of Texas, 1986

Professional Experience

Industrial:

- ***Chief Data Scientist (Advanced Analytics), Engineering Fellow, Sr. Director***, Teradata Labs, 2016-
- *Principal Research Scientist*, Big Data Advanced R&D Center, Hewlett-Packard Laboratories, 2013-2016
- *Principal Research Scientist*, Hewlett-Packard Laboratories, 2007-2013

- *Senior Research Scientist*, Hewlett-Packard Laboratories, 2002-2007
- *Sr. Member Technical Staff*, E-Business Research, Compaq Computer 2001-2002
- *Member Technical Staff*, Advanced Product Development and Research Labs, Motorola, 2000- 2001
- *Member Technical Staff*, Texas Instruments, 1992- 2000
- *Analyst/Biostatistician*, Johnson and Johnson, 1988-1989
- *Scientific Advisor*, Imanage, a document analytics company, *Chicago, Illinois, 2016-*

Startups:

- Co-Founder, Chief Scientific Officer, Pattern Sciences LLC: An Austin based Startup involved in Data Sciences and Machine Learning in Healthcare applications, and Department of Defense contracts.

- Mentor, Capital Factory, incubator, accelerator for start-ups in Austin, Texas, 701 Brazos St, Austin, TX 78701, 2016-present.

Academic:

- Adjunct Assistant Professor, Department of Statistics and Data Sciences, The University of Texas at Austin, 2018-Present
- Faculty Affiliate (SSI), Department of Statistics and Data Sciences, The University of Texas at Austin, September 2012-Present
- Visiting Professor, Indian Institute of Technology, Hyderabad, Department of Electrical Engineering, Spring, 2012- June 2013
- Professor, Indian Institute of Information Technology, Bangalore, 09/2003-06/2004
- Adjunct Lecturer, Department of Mathematical Sciences, University of Texas at Dallas,1992, 1998
- Adjunct Faculty Member and Lecturer, University of Texas at Arlington, Department of Management Sciences and Information Systems, 1989-1992
- Graduate Research Assistant, Departments of Mathematics, Finance, and Information Systems, The University of Texas, 1986-1989
- Graduate Teaching Assistant, Department of Mathematics, The University of Texas, 1983-1986
- Research Scholar, Production and Quantitative Methods Department, The Indian Institute of Management, Ahmedabad, 1981-1983

Industrial Training (Graduate Level)

1. Design and Analysis of Statistical Experiments, Texas Instruments Incorporated
2. Statistical Process and Quality Control, Texas Instruments Incorporated

Consulting

1. *LiveOak Inc*
   - Machine Learning and Forecasting models for predicting luxury real-estate properties going on sale in Southern California
2. *7X.Energy, Solar Energy*
   - Battery Storage Optimization and Bidding for Wind and Solar at Sub Stations in Texas
3. *Rollick, Austin, Texas*
   - *Campaign Optimization for Forklifts by Toyota*
4. *BrilliantMD, An Austin based Health Analytics Company, 2016-2017*
   - *Analytical pipelines for optimized care-giving from acute care to rehabilitation*
5. *Imanage, a document analytics company, Chicago, Illinois, 2016-*
   - Machine Learning for Document Clustering and Classification. Detection of fraud by attorneys at Law firms.
6. *Move Inc, A Newscorp Company, Santa Clara, California*
   - Machine Learning and Forecasting models for predicting real-estate properties going on sale in the United States of America.
7. *Hologic Systems, Marboro, Massachusetts, 2017*
   - Image recognition for automated Detection of Cancer of the Cervix.
8. *Alberta Open Data Institute*
   - *A Canadian government initiative for analytics of publicly available data related to Healthcare, Energy, and other verticals with consequences to public policy*
9. *McKinsey & Company, Institute for Corporate Excellence, Washington, D.C., 2007*
   - Consulting on Text analytics and mining methods to analyze enduring corporate excellence
10. *BHEL, Research and Development, Government of India, 2007*
    - Adviser on the project related to Gas turbine diagnostics and monitoring via machine learning
11. *B2K Corporation and Brickwork, Bangalore, India, 2003-2005*
    - Adviser related to Quantitative methods, Analytics, and Training in Finance, retail, Business process outsourcing, and Engineering
12. *National Bank of Kuwait, Kuwait City, Kuwait, 2004*
    - Workshop on *"Analytics in Banking,"* and advise on project related to customer recruitment and Retention, Kuwait City, Kuwait
13. *Government of Karnataka, Revenue Department, 2003*
14. Adviser related to Mining government lands records data to detect trends in crop yields, mutation time for deed transfers, and other aspects of government owned lands and tracts

15. *World Omni Financial Corporation (Subsidiary of Toyota corporation), Deerfield Beach, Florida, 2000*
16. Business Intelligence and Customer Relationship Management: Analytical solutions for customer segmentation to determine leasing and buy-out strategies for loss mitigation in the automobile industry
17. *Office of Civil Rights, United States Government, 1992*
18. Study discriminatory practices in granting athletic scholarships in colleges and universities within the State of Texas

Short Courses and Workshops

1. 14th Summer Statistics Institute, The University of Texas at Austin, 2020, May 23th-May 26st.
2. Large-Scale Machine Learning and Parallel computation, Teradata Labs, Spring, 2020.
3. 13th Summer Statistics Institute, The University of Texas at Austin, 2020, May 26th-May 29st.
4. American Statistical Association, Short Course on Big Data Analytics, American University, Washington D.C., June, 2019
5. 12th Summer Statistics Institute, The University of Texas at Austin, 2019, May 28th-May 31st.
6. 11th Summer Statistics Institute, The University of Texas at Austin, 2018, May 21th-May 24th.
7. 10th Summer Statistics Institute, The University of Texas at Austin, 2017, May 22th-May 25th
8. 9th Summer Statistics Institute, The University of Texas at Austin, 2016, May 22th-May 25th
9. 8th Summer Statistics Institute, The University of Texas at Austin, 2015, May 26th-May 29th
   - *"Big Data Analytics: Structured, Semi-Structured, and Unstructured"*
10. 2nd International Conference on Big Data Analytics, Mysore, India, 2013
    - *"Tutorial on High Dimensional Big Data,"*
11. 7th Summer Statistics Institute, The University of Texas at Austin, 2014
    - *"Multivariate Methods in Big data: Theory and Methods"*
12. 6th Summer Statistics Institute, The University of Texas at Austin, 2013
    - *"Predictive Analytics and Big Data: Theory and Methods"*
13. SAS Institute, Beijing China
    - Workshop on "*Marketing Optimization, From Theory to Practice*," SAS, Dec 21st-22nd, 2009, Beijing, China
14. SAS Institute, Shanghai China

- Workshop on "*Data Mining in Financial Applications*," SAS, China, Dec 17st-21nd, 2008, Shanghai, China

15. SAS Institute, Hong Kong, China
- Seminar on *"Data Mining, From Theory to Practice in Banking and Finance,"* 2007

16. SAS Institute, India
- Seminar on *"From Analytics to Profit,"* Bangalore, India

17. National Bank of Kuwait, Kuwait
- Tutorial on *"Predictive Analytics and Statistical modeling in Banking,"* 2008

18. McKinsey & Company, Institute for Corporate Excellence, Washington, D.C.2006
- Seminar on *"Text Mining and Unstructured Data analysis. Mining annual reports, Quarterly earnings reports, News reports, Analyst reports to extract insights about corporate performance indices,"* Washington, D.C.

19. Semiconductor Manufacturing Technology Consortium (SEMATECH), Austin, Texas
- Workshop on *Intelligent Data Analysis in Semiconductor Manufacturing and Research*, 1999

Publications

Patents Issued

1. *Incrementally Updating a Database Statistic, 10,922,314*

2. *Singular Value Decompositions, 10,012,1527*

3. *Segments of Contacts, 10,565,603*

4. *Managing Database with Counting Bloom Filter, 10,452,676*

5. *Incrementally Updating Statistics, 10,430, 411*

6. *Estimating Unique Entry Counts Using a Counting Bloom Filter, 9, 465,826.*

5. *Generating a Qury plan for estimating a number of unique attributes, in a database,* 9,576,027

6. *Energy Based Wavelet Thresholding, 9,405,731*

7. *Anomaly Detection in Streaming Data, 9,218,527*

8. *System and Method for Ranking Anomalies, 9,141 914*

9. *Estimate the number of unique values in a list, 9,158, 815*

10. *Detecting Regime Change in Streaming Data, 9,078,629*

11. *Classifying Unclassified Samples, 9,037,518*
12. *Estimation of Unique Database Values, 8,549,004*
13. *Website Analysis using Quantitative and Qualitative Data, 8,396,737*
14. *Method and System for Site Path Evaluation Using Web Session Clustering, 8,572, 223*
15. *Method and apparatus for identifying anomalies of a signal, 8,620,609*
16. *A system for accurately predicting revenues given a portfolio of investments, 8,180,693*
17. *A system and Method for incorporating expert opinion in statistical models, 8,180,694*
18. *System and method for optimizing financial performance generated by marketing investments under budget constraints, 8,027,897*
19. *Multi Regime Detection in Streaming Data, 8,620,987*
20. *Anomaly Detection in Data Centers, 8,668,620*
21. *Classification of a signal in a time domain,8,706,203*
22. *Compression of Non-dyadic Sensor data Organized within a Non-dyadic Hierarchy, 8,898,209*
23. *Process Capability Metrics in On-line Monitoring, (US7,002,106,82): Defensive Publication*
24. *System and Method for estimating unique attributes in a database, 7,428,550*
25. *Autoregressive Model for Time Series Model, 832, 09810 (Allowed for Grant)*
26. *Methods and Systems for Processing Data Arrays Using Bloom Filters, 828,37112,(Allowed for Grant)*

Patents pending

1. *System and methods for acceleration of machine learning functions, Pending*

2. *A Graph Theoretic Learning Model for Generating Edges and Missing Value Imputation,* (7,002, 21953), Disclosed
3. *Information Theoretic Models for Statistical Classification,* (US8,422, 4822), Pending
4. *Anomaly detection in utility clouds using parametric and non-parametric statistical approaches* (US201005668) Filed
5. *Using Wavelet Transforms for Smoothing in The Estimation Of Regression Parameters For Applications In Marketing*, (US200905071), Filed
6. *Wavelet Compression with Bootstrap Sampling* (US20110184934), Published
7. *Clustering and Analysis of Electronic Medical Records,* (83166859), Filed
8. *Dataset Compression,* (20130191309), Published

Invention Disclosures

1. *A System and Method for Analytical Processing and Image Recognition, DN 17-1054*
2. *A System and Method for Predicting Customer Behavior, DN 17-1053*
3. *Scalable Solution for Dimensionality Reduction by Maximizing Mutual Information, DN 18-1055*

Papers

1. *"Social Determinants of Recidivism: A Machine Learning Solution,"* (With Vikrant Shirvaikar), arXiv:2011.11483
2. *"Topological Data Analysis in Digital Marketing," 2018,* (With Mingzhang, Yin), Applied Stochastic Models in Business and Industry,John Wiley & Sons, vol. 36(6), pages 1014-1028, November.
3. *Enterprise-wide Machine Learning Enterprise Solution using Teradata Vantage: An Integrated Analytics Platform,"* (Choudur Lakshminarayan Khaled Bouaziz, Awny Al Omari, Thyagarajan, Ramakrishnan, Faraz Ahmad, Srinivasan Raghavan, Prama Agarwal*), 2019, IEEE International Conference on Big Data, 2019, 2043-2046*

4. *"Model Management and Handwritten Character Recognition, (*Choudur Lakshminarayan Khaled Bouaziz, Awny Al *Omari, Thyagarajan, Ramakrishnan, Faraz Ahmad, Srinivasan Raghavan, 6110-6112*
5. *Character Recognition by deep Learning: An Enterprise Solution,* (Khaled Bouaziz, Thyagarajan Ramakrishnan, Srinivasan Raghavan, Kyle Grove, Awny Al Omari, Choudur Lakshminarayan), 2018, IEEE International Conference on Big Data, pp:1717-1727, 2018
6. *"Modeling Complex Clickstream Data by Stochastic Models: theory and Practice," WWW (Companion Volume) pp. 879-884, 2016* (With Ram Kosuru, Meichun Hsu)
7. *"Automatic Classification of Heartbeats,"European Signal Processing conference, 2014, pp 1542-1546,* (With Tony Basil)
8. *"High Dimensional Big Data and Pattern Analysis: A Tutorial,"* Lecture Notes in Computer Science, 2013, Volume 8302, pp 68-85
9. "*Pattern Recognition in Large-Scale Data Sets: Application in Integrated Circuit Manufacturing,"*, Lecture Notes in Computer Science, 2013, Volume 8302, 2013, pp 185-196.
10. *"Adaptive Estimators of Process Capability Indices Using Preliminary Test,"* In JSM Proceedings, Section on Statistical Quality and Productivity. Montreal, Canada: American Statistical Association. 2013, (With Chien-Pai Han)
11. *"Robust versions of the Tukey Boxplot with Their Application to Detection of Outliers,"* Proc. of IEEE, ICASSP, 2013, Page(s): 6506-6510 (With Georgy Shevlyakov, Kliton Andrea, Pavel Smirnov, Alexander Ulanov, Natalia Vassileva)
12. "*Detection of Classes of Heart Arrhythmias based on Heartbeat Morphology Patterns,"* Proceedings of 2nd International Workshop on Analytics for Cyber-Physical Systems, 2013 (With Tony Basil, C. Krishnamohan)
13. *"A comparison of Statistical Machine Learning Methods in Heartbeat Detection and Classification*," Big Data Analytics, Lecture Notes in Computer Science, Volume 7678, 2012, pp. 16-25, (With Tony Basil, Bollepalli Chandra)

14. *"Nearest Neighbor Distributions for Imbalanced Classification,"* Proc. Of IEEE-INNS International Joint Conference on Neural Networks, 2012, Page(s):1-5, (With Jose C. Principe, Evan Kriminger)
15. *"Mixture of designer-experts for multi-regime detection in streaming Data,"* in Proceedings of 20th European Signal Processing Conference (EUSIPCO, 2012), Page(s):410-414, (With Jose C. Principe, Evan Kriminger),
16. *"Ultra low power automaton for heartbeat classification based on integrate and fire sampler,"* To appear in Machine Learning and Signal Processing Conference (MLSP), Spain, 2012, (With Gabriel Nallathambi, Jose C. Principe)
17. *"Live Operational Intelligence- Characterizing Operational Processes,"* HP Business Intelligence Conference, 2012 (With Chetan Gupta et al)
18. *"Enabling Partial Data Cube Computations using the Bloom Filter,"* Proc. of the 1st International Workshop on Analytics for Cyber-Physical Systems, 2012, Anaheim, USA, (With Ram Kosuru)
19. *"Ranking Anomalies in Data Centers,"* Proc. Of 13th IEEE Network Operations and Management Symposium, NOMS 2012, Page(s): 79-87 (with Chengwei, W.; Viswanathan, K.; Talwar, V.; Wade, S.; Macdonald, G.)
20. *Time-based Compression and Classification of Heartbeats*, IEEE Transactions on Biomedical Engineering, Volume:59, Issue:6, Pages:1641-1648, (With Alexander Singh Alvarado, Jose C. Principe)
21. *"Early Detection of Anomalous Patterns in Sensor Streams,"* Tech Con '12 Washington, D.C., USA, 2012. (With Alan Benson, Ram Kosuru)
22. "*Modified embedding for multi-regime detection in non-stationary streaming data,*" Proc. of IEEE, ICASSP, 2011, 2256-2259, (With Jose C. Principe, Evan Kriminger)
23. "*Time encoding using the Integrate and fire sampler: Discriminative Representation for Neural Action Potentials*," Proc. of International Conference on Sampling Theory and Applications SampTA'11, 2011, Singapore. (With Alexander Singh Alvarado, Jose C. Principe)
24. *"Local Frequency-Based Estimators for Anomaly Detection in Oil and Gas Applications,"* In JSM Proceedings, Section on Statistical Learning and Data

Mining. Miami, Florida: American Statistical Association. 2011, (With Alexander Singh Alvarado, Evan Kriminger).

25. "*Statistical Techniques for Online Anomaly Detection in Data Centers,*" M 2011: Proc. of 12th IFIP/IEEE International Symposium on Integrated Network Management, 2011:385-392, (With Chengwei, W.; Viswanathan, K.; Talwar, V.; Wade, S.; Karsten, S.)
26. *"Better Drilling Through Sensor Analytics: A Case Study in Live Operations management*," Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data, Page(s):8-15, (With Umeshwar Dayal, Krishnamurthy, Vishwanathan, Chetan Gupta, et al)
27. *"Finding Needles in Haystacks: Analytic Techniques for Anomaly Detection in Data Centers,"* Tech Con'11, Orlando, Florida, 2011, (With Nigel Cook, Jeff Hilland, Partha Ranganathan, Karsten Schwan, Krishnamurthy, Vishwanatha, Vanish Talwar, Chengwei Wang)
28. *"Applications of Sensor Analytics over streaming Data in Oil Production*," in Proc. of 17th International Conference on Management of Data (COMAD 2011), Bangalore, India, (With Umeshwar Dayal, Krishnamurthy, Vishwanathan, Chetan Gupta, et al)
29. "*Online Detection of Utility Cloud Anomalies Using Statistical Metric Distributions*," Tech Con '10 Phoenix, Arizona, USA, 2010. (With Nigel Cook, Jeff Hilland, Partha Ranganathan, Karsten Schwan, Ram, Swaminathan, Vanish Talwar, Chengwei Wang)
30. "*Non-Dyadic Haar Wavelets for Streaming and Sensor Data,*" Proc. of IEEE 26th International Conference on Data Engineering. March 1-6, 2010. Los Angeles, USA, pp- 569-580, 2010(With Chetan Gupta, Song Wang, Abhay Mehta)
31. "*Using page sequence analytics to understand complex web behavior,"* In *JSM Proceedings*, Section on Statistical Learning and Data Mining. Alexandria, VA: American Statistical Association, pp- 3926-3929,2009, (With Alan Benson)
32. "*Marketing Innovation: Budget Optimization for Increased Revenues*," Proceedings of the 15th Annual Workshop of the HP Software University Association (HP-SUA), 2008, page(s): 247-253, Marrakech, Morocco.

33. *"A New UEC Estimator for Neoview: Enhancing Query Throughput,"* Tech Con'08, Boston, 2008, (With Vinay Deolalikar, Hernan Laffitte, et al).
34. *"Session Centric page sequence clustering for improved web experience on hp.com ,"* Tech Con'05, Phoenix, Arizona, USA, 2005. (With Alan Benson)
35. *"Improving Customer Comments via Text Mining,"* Lecture Notes in Computer Science, Page(s): 288-299, Springer Berlin/Heidelberg, 2005. (With Alan Benson, Qingfeng Yu)
36. *"Automated Comment Analysis,"* Tech Con'04, Orlando, Florida, USA, 2004. (With Alan Benson), : Acceptance rate <10%.
37. *"Pattern Recognition in IC diagnostics Using the Linear Discriminant Classifier and Artificial Neural networks,"* Accepted, 2000, IEEE transactions in semiconductor Manufacturing. (With Michael Baron, Zhenwu, Chen)
38. *Markov Random fields in Pattern Recognition for Semiconductor Manufacturing,"* Technometrics, Vol. 43, No. 1, pp.66-72, 2001. (With Michael Baron and Zhenwu Chen)
39. *On estimating the Mean in a Bivariate Normal Distribution,"* Journal of Statistical Simulation and Computation, pp-155-170, Vol. 58, 1997. (With Chien-Pai Han)
40. *"A Preliminary Test estimator of Process Capability Index,"* Texas Instruments Technical journal. 1998. (With Chien-Pai Han).
41. *"Knowledge Discovery and Data Mining for Improved Manufacturing Quality-An Overview,"* Texas Instruments Technical journal. 1998.
42. *"Quantitative Web Analytics and Qualitative Usability research-Strange Bedfellows,"* Tech Con'03, Keystone, Colorado, USA, 2003. (With Alan Benson).
43. *"Signature Analysis: Statistical Methods and Their Applications to Failure Analysis,"* Proceedings of the 22nd International Symposium for Testing and Failure Analysis, Los Angeles, California, Page(s):183-187, 1997. (With Seshu Pabbisetty and Chien-Pai Han)
44. *"Signature Analysis based IC Diagnostics,"* Proceedings of the 6th International Symposium on the Physical and Failure analysis of Integrated Circuits, IEEE, Singapore, Page(s):167-171, 1997. (With Seshu Pabbisetty and Chien-Pai Han)

45. *"An Auto-Logistic Model for the Joint Distribution of Failed Chips on a regular Lattice,"* Proceedings of the American Statistical Association section on Physical and Engineering Sciences, pp-75-80, Chicago, USA, 1996. (With Michael Baron)
46. *"An Adaptive Estimator of Location based on the t-family,"* Communications in Statistics, Theory and Methods, Vol. 23, pp-747-761, 1996. (With D.L. Hawkins)
47. *"Detecting Outliers in Semiconductor Wafer yields and Manufacturing,"* Texas Instruments Technical journal. Page(s):52-57, 1993.
48. *"Estimation of Regression Coefficients Under Multicollinearity Condition,"* Non-Parametric Statistics and Related Topics, Saleh (Ed.), Elsevier Publishers, Page(s): 297-309, 1992 (With Chien-Pai Han)
49. "*Adaptive wavelet filtering in database applications*, International Conference on Time-Frequency Analysis," 2009, Strobl, Austria.
50. "*Marketing Optimization for Increased Revenues*," M2009, 12th Annual Data Mining Conference, Las Vegas, Nevada.
51. "*Vector space based Text Mining Models*," M2005, 7th Annual Data Mining Conference, Las Vegas, Nevada.
52. *"A biased estimator of the process capability index,"* International Indian Statistical Association, 1998, McMaster University, Hamilton, Ontario, Canada
53. *"Markov Random Fields in Pattern Recognition for Semiconductor Manufacturing,"* Special invited paper, Institute of Mathematical Statistics/American Statistical Association, Spring Research Conference (2001)
54. "**Spatial Statistical Interpolation Methods for Image enhancement*," INRIA, Sophia Antopolis, Nice, France, 1999 (With Mukul Shirvaikar)
55. *"Statistical Techniques for Online Anomaly Detection in Data Centers,"* HPL-2011-8, HP Labs Technical Report, 2011, (With Wang, Chengwei; Viswanathan, Krishnamurthy; Talwar, Vanish; Satterfield, Wade; Schwan, Karsten)
56. "*Parallelizing Statistical Operators in a Database Management System Framework,"* HP Labs Technical report, HPL-2011-84, 2011, (With Subbiah, Suresh, Wehrmeister, Robert)

57. *"On Wavelet Compression and Cardinality Estimation in Enterprise Data,"* HPL-2010-132. HP Labs Technical Report, 2010, (With Dayal, U.; Chetan, G.; Swaminathan, R.)
58. "*A new composite estimator of distinct values that performs well in a wide range of skewness,"* HP Labs Technical report HPL-2008-218, (With Deolalikar, Vinay, Laffitte, Hernan)
59. *"Introducing Dependency in Integrated Circuit Signature Analysis Models,"* Technical report, No. 238, The University of Texas at Dallas, 1997. (With Michael Baron)

*This is a paper-level presentation that was not ultimately published due to my departure from Texas Instruments.

Lecture notes and Manuals and Books

1. *Data mining and analytics for decision making*
2. *Data mining and analysis: Theory to practice*
3. *From Analytics to Profit using Statistical Modeling*
4. *"Statistics in data mining: exploration, modeling, and analysis," Unpublished manuscript*
5. *"Pattern Analysis, Predictive Analytics, and Big Data: Theory and Methods," Manuscript under preparation, CRC Press, Taylor & Francis, due Spring 2021, ISBN: 978-1-4987-5492-7*

Invited Talks

Data Science colloquiim, College of Natural Sciences, The University of Texas at Arlington, March 25th, 2020, Capital Factory Innovation Forum, Austin, Texas, 2019, The Fall Seminar Series, Department of Statistics and Data Sciences, The University of Texas at Austin, Loyola College, Loyola Institute of Business Administration, Chennai India, 2018, College of Business, Distinguished Lecture, Department of Industrial Engineering, Northwestern University, 2017, The University of Texas at Arlington, 2017, The University of Madras, Department of Statistics, S. Rajagopalan Endowed Lecture, 2017, Department of Mathematics, American University, Washington D.C. 2016, Department of Operations Management and Management Sciences, Distinguished Lecture, The University of Texas at Arlington, 2016, The University of Texas at San Antonio 2016, The University of Texas at Tyler (National Science Foundation Sponsored lecture), 2016, Department of Computer Science, The Hyderabad Central University, 2013, The

Numerical and Harmonic Analysis Group, University of Vienna, 2012, The Department of Electrical Engineering, Department of Mathematics, The Indian Institute of Technology, 2012, The Indian Institute of Science, 2012, The Department of Computer Science, The University of North Texas 2010, Department of Computer Science, The Imperial College, London 2002, Department of Statistics, University of Madras, 2002, Department of Operations Research, McCombs School of Business 2001, INRIA, Sophia Antopolis, Nice, France, 2000, Department of Statistics, Osmania University, 2000

Conferences

Over 50 Conference presentations around the world

Principal Investigator and Co-PI of Research Grants

1. *"*Borrowing Digital Data across studies in Clinical Trials Designs*" (with Dr. Peter Mueller, The University of Texas, Dr. Robertino Mera, Gilead Sciences)*

2. *"Talent Management and Pilot Training Analytics,"* US Airforce, $100,000, Pattern Sciences LLC.
3. *"Sensor data, Sparse sampling, and signal reconstruction*," Department of Electrical Engineering, The Indian Institute of Science, Bengaluru. $10,000, Collaborator; Dr Seelamankula Chandrasekhar
4. *"Comparative Effectiveness for Evidence-based HealthCare Advances,"* Granting Body: Innovation Research program, HP Laboratories, 2012-2013, $75,000 (With Professor Eva K. Lee, Georgia Institute of Technology, Georgia, USA)
5. *"Real-Life Data Analysis: Distribution Model Fitting and Anomaly Detection,"* Granting Body: Innovation Research program, HP Laboratories, 2012-2013, (With Georgy Shevlyakov, St. Petersburg State Polytechnic University, Russia)
6. *"Anomaly Detection in Multivariate Data Streams using Kernel Methods and Information Theoretic Cost Functions,"* Granting Body: Innovation Research program, contract CW221761, HP Laboratories, 2010-2011, $50,000 (With Professor Jose C. Principe, The University of Florida, Gainseville, USA)
7. *"Sparse sampling, Stochastic Point Processes, and Classification of Signals,"* Granting Body: Innovation Research program, contract CW221761, HP Laboratories, 2011-2012, $50,000 (With Professor Jose C. Principe, The University of Florida, Gainseville, USA)
8. "*Statistical Models for Signature Analysis,"* Granting Body: Texas Instruments Corporate Technical Council, 1996-1997, $25,000, (With Professor Chien-Pai Han, the University of Texas at Arlington, USA)
9. *"Statistical Pattern Recognition for Yield enhancement and Cycle Time Reduction,"* Granting Body: Texas Instruments Corporate Technical Council, 1997-1998, $25,000, (With Professor Michael Baron, the University of Texas at Dallas)

10. *"Applying Markov Random Fields to model failed Chip Patterns in Semiconductor Manufacturing,"* Advanced Technology program of Texas University System. Ranked 20/480, 1998. (With Professor Michael Baron, the University of Texas at Dallas, USA)

Significant projects and programs in the Industry

1. **Digital Marketing**

Patented technologies for predicting online user behavior using page navigation sequences on enterprise websites using stochastic models and computational topology.

2. **Scientific Computations (Patents Pending)**

Highly Scalable and Parallel Algorithms for Large-Scale Machine Learning, Efficient implementation of algorithms in Big Data Analytics. Examples: Machine Leraning Models with Regularization.

3. **Randomized Clinical Trials**

Statistical designs which combine cohorts who volunteered for an experimental therapy (treatment group) with a synthetically generated control group of patients found in real world data. Random assignment of patients to treatment and control groups yield comparable results. In contrast, a non-random selection process results in misleading comparisons when patients exposed to the treatment group differ systematically from observed data. So, adjustment for lack of randomization is paramount. Borrowing Digital Data across studies in Clinical Trials Designs (Under review at National Institutes of Health).

4. **Business Intelligence and Live Analytics (Patents pending)**

Lead statistician and member of the team to create a unified data and analytics platform, called "Live Business Intelligence (BI)" that shifts BI from the traditional back-office, report-generation orientation, to an enabler for delivering data and computationally - intensive mathematical methods that transform operational business processes and customer interactions for near real time decision making. The Live BI Platform leverages a new paradigm by massive parallel processing, and supports analytics over both streaming and historical data. A prototype is completed which includes pattern detection by non-linear dynamical systems, detectors based on time-frequency procedures, auto-regressive processes, and anomalies. This program generated several patent pending technologies, mentions in the popular press, and show cased in trade shows and conferences

2 **Comparative Effectiveness Research (Patents Pending)**

The aim of the study is an exhaustive assessment, analysis, and future directions of comparative effective research (CER) for improving healthcare quality and efficient delivery using the latest scientific and technological advances. It is anticipated that the evidence based research will produce a blueprint for personalized medical care of patients with chronic illnesses with the best possible clinical outcome while controlling escalating costs and increasing patient satisfaction.

5. **Sensors, Sensing, and Analytics (Patents Pending)**

The aim is to investigate embedding sampling algorithms for compression at the sensor level as well as perform analytics in the compressed domain. We envision implementation of event-based samplers in low power analog VLSI circuits for sampling and subsequently low-bandwidth transmission. The data generated in the compressed domain are a set of events (spikes). These are treated as realizations of an underlying stochastic point process. Allowing us to employ tools in the point processes literature for real-time signal classification, detection, and prediction.

6. **Patented technology for query optimization in relational database systems**

When a query optimizer in a large database prepares query plans, it requires an estimate of the number of distinct values in specified columns of the database. The number of distinct values, also known as unique entity count is a fundamental component of the query optimization process. A patented technology based on a family of generalized Jackknife estimators and others is implemented and productized. The resulting query throughput due to this algorithm was a market differentiator. Product is now used by New York stock exchange, *Yahoo!*, WalMart and other major companies.

7. **Multi-patent technology for Marketing Optimization**

The goal of the program is to more effectively manage the $500 Million marketing budget. The objective is to determine investments in marketing instruments to maximize revenue and profit. We used a combination of regression analysis, signal processing, mathematical optimization, and qualitative analysis to build patents pending solutions and a framework to optimally allocate investments to generate higher revenue and profits. The program was implemented in Germany, Spain, Netherlands, Canada, Korea, Australia, and Mexico. The incremental revenue generated across countries is ~$30 million. I served as the lead scientist to develop, implement and operationalize this awards winning program.

8. **Patented system for Total customer experience and satisfaction models**

The total customer experience (TCE) group is focused on customer satisfaction on www.hp.com. Towards that end, they capture customer related issues via on-line surveys.

While the data was collected, there was no formal way to read, collate, and respond to unstructured (textual) customer comments. We introduced text mining as a technology in HP to process and understand customer concerns. We used latent semantic indexing, the EM algorithm, numerical linear algebra procedures and Gaussian mixture models for clustering customer complaints and identified issues that were previously undetected. The use of non-parametric density estimation in statistical classification helped track customer satisfaction by problem areas. The program increased customer satisfaction significantly up by 5%. I served as the lead scientist to develop, implement and operationalize this award winning program.

9. **On-line customer behavioral models and analysis**

Lead statistician of a team that set up a database ~10 Tb of on-line customer information, originally via Keylime and later Omniture web analytics system. This is the first time that client-side on-line customer click-stream data was captured in a central repository at Hewlett- Packard. The customer database provided the basis to tap into the on-line customer market more effectively. Provided and developed key technical ideas, web metrics, and mathematical methods to find predictors of purchase which helped to double revenues.

10. **Integrated circuit Fault Diagnostics**

In the early 90's statistical pattern recognition and data mining were very incipient. We developed an automated system for detecting faults in integrated circuit (IC) manufacturing in a modern wafer fabrication facility. Using artificial neural network systems and generalized Markov random fields, we built a framework to analyze manufacturing defects and patterns of defects to find the root cause. The program resulted in significant cost savings in the order of millions and reduction in failure analysis cycle times from days to hours. I served as the lead scientist to develop, implement and productize the ground breaking program.

11. **Motorola Inc, Advanced Products and Research development Labs (APRDL)**

Lead the effort to set up a Statistics laboratory for analysis of semiconductor processing and measurement data. The role of the lab was to create a centralized database of chip manufacturing and failure analysis data. Built an on-line system to identify process problems in real-time to minimize faulty product, minimize cycle-time, and improve yield

12. **Texas Instruments**

- Lead Statistician to develop software tools for intelligent data analysis and pattern recognition.
- Develop algorithms for IC fault diagnostics, Process control, and Quality control
- Develop algorithms for Image enhancement in Digital Signal Processing chips

Ongoing Projects

1. "*Sampling, encoding and decoding of continuous time series using stochastic point processes and generalized linear models."*
2. *"Customer Segmentation Models in Digital Marketing."*
3. *"Touch Point Optimization Using Time Series Models in Digital Marketing."*
4. *"Detection and Classification of Heart Arrhythmias Using ECG Signals."*

Principal Investigator and Co-PI of Research Grants

1. *"*Borrowing Digital Data across studies in Clinical Trials Designs*" (with Dr. Peter Mueller, The University of Texas, Dr. Robertino Mera, Gilead Sciences)*

2. *"Talent Management and Pilot Training Analytics,"* US Airforce, $100,000, Pattern Sciences LLC.
3. *"Sensor data, Sparse sampling, and signal reconstruction*," Department of Electrical Engineering, The Indian Institute of Science, Bengaluru. $10,000, Collaborator; Dr Seelamankula Chandrasekhar
4. *"Comparative Effectiveness for Evidence-based HealthCare Advances,"* Granting Body: Innovation Research program, HP Laboratories, 2012-2013, $75,000 (With Professor Eva K. Lee, Georgia Institute of Technology, Georgia, USA)
5. *"Real-Life Data Analysis: Distribution Model Fitting and Anomaly Detection,"* Granting Body: Innovation Research program, HP Laboratories, 2012-2013, (With Georgy Shevlyakov, St. Petersburg State Polytechnic University, Russia)
6. *"Anomaly Detection in Multivariate Data Streams using Kernel Methods and Information Theoretic Cost Functions,"* Granting Body: Innovation Research program, contract CW221761, HP Laboratories, 2010-2011, $50,000 (With Professor Jose C. Principe, The University of Florida, Gainseville, USA)
7. *"Sparse sampling, Stochastic Point Processes, and Classification of Signals,"* Granting Body: Innovation Research program, contract CW221761, HP Laboratories, 2011-2012, $50,000 (With Professor Jose C. Principe, The University of Florida, Gainseville, USA)
8. "*Statistical Models for Signature Analysis,"* Granting Body: Texas Instruments Corporate Technical Council, 1996-1997, $25,000, (With Professor Chien-Pai Han, the University of Texas at Arlington, USA)
9. *"Statistical Pattern Recognition for Yield enhancement and Cycle Time Reduction,"* Granting Body: Texas Instruments Corporate Technical Council, 1997-1998, $25,000, (With Professor Michael Baron, the University of Texas at Dallas)
10. *"Applying Markov Random Fields to model failed Chip Patterns in Semiconductor Manufacturing,"* Advanced Technology program of Texas University System.

Ranked 20/480, 1998. (With Professor Michael Baron, the University of Texas at Dallas, USA)

Other Collaborations

1. Anomaly detection and Prediction in Data Centers, Department of Electrical engineering, Georgia Institute of Technology, Collaborators: Professor Karsten Schwan, Dr. Vanish Talwar (Intelligent Infrastructure Lab, HP)
2. Time-frequency analysis of continuous time series data, Department of Mathematics, IIT Chennai, collaborators; Dr. Radha Ramakrishnan, Dr. S.H. Kulkarni

Technical positions (By election)

1. Nominated to the **Fellowship of the American Statistical Association**, 2019
2. **Engineering Fellow,** 2017-Present, Teradata Labs
3. **Master technologist**, 2003 – 2016, Technical career path, Hewlett-Packard Labs, Hewlett Packard Company
4. **Senior member of technical staff**, 2001-2003, Advanced Technical career ladder, The Compaq Computer Corporation

Program Committee Member

1. Quality and Productivity Research Conference (QPRC), American Society for Quality, 2019
2. SIAM International conference on data mining,
3. Joint Statistical meetings
4. HP Tech Con'11, Review Committee Member
5. International Conference on Big Data Analytics, 2012, New Delhi, India

Service to Profession

1. IEEE Computer Science, and EMBS, Austin Chapter, Communications Director, 2013-16
2. American Statistical Association, Austin Chapter, Informal Support for recruiting speakers, Vice President nominee, 2018-

Referee Activities

1. Technometrics (Journal of American Society of Quality and Technology)
2. Communications in Statistics, Theory and Methods
3. IEEE Transactions
4. VLDB (International conference on Very Large Databases)
5. ICDE (International conference on Data engineering)
6. ICDM(International conference on Data Mining)

7. DEXA (Conference on Databases and Expert Systems)
8. DNIS (Databases and Networked Information Systems)
9. DKE (Databases and Knowledge Engineering)
10. SIAM (Society for Industrial and Applied Mathematics)
11. EDBT (Enterprise Database Technology Conference)
12. HP Tech Con(Annual Technical conference of HP technologists)
13. American Society of Quality
14. SIGMOD, Association of Computing Machinery
15. ACM SIGKDD

Advising

1. *"Human Extremity Detection and Its Applications in Action Detection and Recognition," Ph.D, Thesis*, Qingfeng Yu, Department of Electrical and Communication Engineering, The University of Texas at Austin, 2009

2. *"Features for Detection of Heart Arrhythmias," M. Tech Thesis, "*Tony Basil, Department of Computer Science and Engineering, The Indian Institute of Technology, Hyderabad, August, 2013

3. Erin Burgoon, Ph.D (2014), Department of Psychology, The University of Texas at Austin

Mentorship

1. Harshita, Vemula, PhD Student, Department of Industrial Engineering and Operations Research, The University of Texas at Austin, 2018-
2. Ramakrishnan, Thiagarajan, Software Engineer, Teradata Labs, Austin, Texas, 2018-
3. Nikhil Chakravarthula, Software Engineer, Teradata Labs, Hyderabad, India, 2019-
4. Dr. Faraz Ahmad, Sr. Architect, Teradata Labs, Santa Clara, 2017-
5. Vikrant Shirvaikar, Student at The University of Texas at Austin and Year round Intern 2017-
6. Mingzhang, Yin, Ph.D student at The University of Texas at Austin and Summer Intern at HP R&D, 2016 Summer
7. Erin Burgoon, Ph.D student at The University of Texas at Austin and Summer Intern at HP R&D
8. Matthew Hagen, Ph.D (2013, expected), School of Industrial and Systems Engineering, Georgia Institute of Technology
9. AlexanderSingh Alvarado, Ph.D (2011), Department of Electrical Engineering, The University of Florida, Gainesville
10. Evan Kriminger, Ph.D (2011), Department of Electrical Engineering, The University of Florida, Gainesville
11. Qingfeng Yu, 2001-2005, Ph.D (2009). Department of Electrical Engineering, The University of Texas at Austin, Texas

12. Bharat, A., Ph.D (2009). Birla Institute of Technology and Sciences, Pilani, India
13. Greg Nolder, 2003, M.S. Department of Management sciences, Southern Methodist University, Dallas, Texas, and Hewlett-Packard
14. Siddique, Kashif, 2005, M.S. in Computer Science, Non-stop Engineering Division, HP, The University of Texas at Austin
15. Gutzler, Jay, 1997-1999, Ph.D., Texas Instruments, Department of Mathematics, University of Texas at Arlington
16. Chen, Zhenwu, 1999-2000, Ph.D., Department of Mathematics, University of Texas at Dallas, and Texas Instruments

Awards and Honors

1. "Topological analysis in Digital Marketing," Special invited paper, International Society of Business and Industrial Statistics (ISBIS) Conference, 2021.
2. HP Innovation Award, 2013
3. Return on marketing Investment *HP Star Award* (2006), HP Awards (2006, 2007)
4. Return on marketing Investment, *Marketing Circle Award* finalist in the category of Marketing Innovation, (2006)
5. Selected among the 19 HP employees from a total of 433 entries from around the world as finalists in the annual competition for contributions to Best Innovations, (2006)
6. Marketing Optimization, A standardized platform for optimal marketing, HP Employee Award 2007
7. A robust estimator of number of unique attributes in a database, HP Employee Award 2007
8. Update Statistics for query Optimization in SQL/MX Databases, HP Employee Award, 2008
9. Pattern Recognition for IC failure Diagnostics, Texas Instruments, TI IDEA program Award, (1998)
10. Image Enhancement using Gaussian Kernel methods, Texas Instruments, TI IDEA program Award, (2000)
11. Markov Random Fields in Pattern Recognition for Semiconductor Manufacturing, Special invited paper, Institute of Mathematical Statistics/American Statistical Association, Spring Research Conference, Alexandra, Viriginia, (2001)

Personal

Married, with no children, United States Citizen