Eithon Cadag

Experience

Jun 2016 - Jul 2021
Principal Data Science Manager, *Microsoft Corp.*, Redmond, WA.
- Manage the team responsible for research, development and testing of the statistical forecasting models that drive the Azure demand signal across the entire cloud; influence $XB of cloud infrastructure and operations investments annually.
- Led team-wide adoption of formal engineering processes for statistical model development, expediting new algorithm development, testing and deployment from weeks to hours.
- Standardized model library and forecast tools; result was widespread adoption of statistical forecasts within the organization and improvements in predictive signal across services.
- Developed generalized extensions to statistical forecasting models, reducing bullwhip effect on supply chain and forecast volatility by >XX%.
- Created and improved core Azure statistical forecasting models; impact was a XX% reduction in forecast error with an estimated $XXXMM of capital expenditure impact over several years.

Mar 2015 - Jun 2016
Director of Data Science, *Capella Biosciences, Inc.*, Palo Alto, CA.
- Standardized statistical model development practices and led adoption of formal engineering processes.
- Led design and development of molecular dynamics and systems biology modeling for drug discovery, and conducted analyses for characterizing drug toxicity.
- Designed and ran computational experiments resulting in discovery and validation of compounds for adverse event prophylaxis and new intellectual property.
- Worked in tandem with CEO in development and cultivation of company culture, including hiring/recruitment process and creation of corporate values.

Aug 2014 - Mar 2015
Data Scientist, *Amino Inc.*, San Francisco, CA.
- First data scientist; developed analytical infrastructure to analyze over one billion medical encounters.
- Built analytical stack on top of Hadoop utilizing Hive, Impala and R. Shipped consumer data product for personalized medical procedure modeling, including event prediction and cost estimation.
- Developed techniques for predicting patient longitudinal history for complex diseases and conditions from noisy and incomplete data, resulting in creation of new intellectual property.
- Continued as a consulting data scientist through March 2016, advising on model development and recruiting.

Jul 2012 - Aug 2014
Manager & Principal Data Scientist, *Ayasdi Inc.*, Menlo Park, CA.
- Led research, development, prototyping and testing of a novel approach for data-driven surgical management using Python. Work resulted in novel intellectual property, a $XMM largest-at-time customer acquisition, and launched the company's marquee efforts in health and medicine.
- Served as product manager for pharmaceutics application from specification through launch. Later assumed formal product management and development responsibilities for company's existing products.
- Research and development in Java and Python for the flagship product. Implemented methods for automating a common workflow using graph theory.
- Led customer projects in life sciences, spanning scoping, analytical execution and delivery for biomarker discovery, toxicology and clinical trials analysis.
- Managed data science team responsible for the pharmaceutics, healthcare and financial services verticals.

Sep 2010 – Jul 2012
Postdoctoral Research Staff, *Lawrence Livermore National Laboratory*, Livermore, CA.
- Led development of statistical methods for pathogen diagnostics on protein immune arrays using R.
- Developed algorithms and analyses for interaction/metabolic network construction in Python.
- Devised metrics and code for large scale protein structure comparisons using Python.

Mar 2010 – Sep 2010
Statistician/Operations Researcher, *US Department of Defense*, Dugway, UT.
- Developed adaptive sampling procedures to meet testing requirements at lower cost.
- Served as subject matter expert for statistical modeling and analysis of bio-related defense data.
- Collaborated with scientists, agencies and other stakeholders to ensure mission needs were met.

Sep 2004 – Dec 2009
Graduate Research Assistant, *University of Washington*, Seattle, WA.
- Developed methodology for query-based statistical classification, validated via protein function prediction.
- Created tools for large scale data integration; devised novel SQL dialect and wrote related lexer, parser for querying uncertain data over graph representations.
- Developed evaluation methods for medical text mining; wrote a natural language tagging system in Python used by academic consortium for annotating medical notes.

Education

Dec 2009 **Ph.D., Biomedical & Health Informatics**, *University of Washington*, Seattle, WA.

Jun 2006 **M.S., Biomedical & Health Informatics**, *University of Washington*, Seattle, WA.

Jun 2004 **B.S., Informatics; B.A., Business Administration**, *University of Washington*, Seattle, WA.
Graduated *cum laude*

Technical Skills

OS & DB Linux, Unix , MS Windows (3+), Apple OS 10+, Postgres (7+), Hadoop/HDFS, AWS

Programming Python/numpy/matplotlib/scipy, R, Java, bash, SAS, SQL/Hive/Impala

Honors, Awards & Affiliations

2019 Microsoft Cloud Capacity Planning Collaboration Award

2018 Microsoft Cloud Infrastructure & Planning Leadership Award

2017 Microsoft Forecasting and Capacity Supply Chain Award

2009 SMART Scholar, US Department of Defense/American Society for Engineering Education

2009 GSFEI Travel Award, University of Washington

2008 Best Poster Award, American Medical Informatics Association Symposium

2007 PSB Travel Award, National Institutes of Health

2007 Graduate student representative for the Biomedical and Health Informatics admissions committee, University of Washington

2006 – 2009 Predoctoral Fellow, National Institutes of Health/National Library of Medicine

2004 Member, Phi Beta Kappa

2002 Evert McCabe Endowed Scholar, University of Washington, Michael G. Foster School of Business

Selected Publications & Patents

Theses

[T.1] "Automated learning of protein involvement in pathogenesis using integrated queries" Doctoral Dissertation; University of Washington. Seattle, Washington, Dec. 2009. Advisor: Peter J. Myler, Ph.D.; Committee: Peter Tarczy-Hornoch, M.D., William S. Noble, Ph.D., Ira J. Kalet, Ph.D, Evan E. Eichler, Ph.D.

[T.2] "High-throughput inference-supported protein characterization utilizing the BioMediator data integration platform" Masters Thesis; University of Washington. Seattle, Washington, Jun. 2006. Advisors: Peter J. Myler, Ph.D., Peter Tarczy-Hornoch, M.D.

Peer-reviewed Manuscripts

[M.1] Leung, E., Huang, A., **Cadag, E.**, Montana, A., Soliman, J.L., Zhou, C.L.E., "Protein Sequence Annotation Tool (PSAT): A centralized web-based meta-server for high-throughput sequence annotations", BMC Bioinf, Vol 17, Issue 43, January 2016

[M.2] Sarikonda, G., Pettus, J., Phatak, S., Sachithanantham, S., Miller, J., Wesley, J., **Cadag, E.**, Chae, J., Ganesan, L., Mallios, R., Edelman, S., Peters, B., von Herrath, M., "CD8 T-cell reactivity to islet antigens is unique to type 1 while CD4 T-cell reactivity exists in both type 1 and type 2 diabetes", J Autoimmun, Vol 50, May 2013

[M.3] **Cadag, E.**, Tarczy-Hornoch, P., Myler, P.J., "Learning pathogenic proteins from integrated query networks", BMC Bioinf, Vol. 13, Issue 321, December 2012

[M.4] **Cadag, E.**, Vitalis, E., Lennox, K., Zhou, C., Zemla, A., "Computational analysis of pathogen-borne metallo β-lactamases reveal discriminating structural features between B1 types", BMC Res Notes, Vol. 5, Issue 96, February 2012

[M.5] Halgrim, S., Xiag, F., Solti, I., **Cadag, E.**, Uzuner, Ö., "A cascade of classifiers for extracting medication information from discharge summaries", J Biomed Sem, Vol. 2 (Suppl. 3), Issue 1, July 2011

[M.6] **Cadag, E.**, Tarczy-Hornoch, P., "Supporting retrieval of diverse biomedical data using evidence-aware queries", J Biomed Inform, Vol. 43, Issue 6, December 2010

[M.7] Uzuner, Ö., Solti, I., **Cadag, E.**, " Extracting Medication Information from Clinical Text", J Am Med Inform Assoc, Vol. 17, Issue 6, 2010

[M.8] Shen, T., Tarczy-Hornoch, P., Detwiler, T., **Cadag, E.**, Carlson, C.S., "Evaluation of Probabilistic and Logical Inference for a SNP Annotation System", J Biomed Inform, Vol. 43, Issue 3, June 2010

[M.9] Halgrim, S., Xia, F., Solti, I., **Cadag, E.**, Uzuner, Ö., "Extracting Medication Information from Discharge Summaries" 2nd Louhi Workshop on Text and Data Mining of Health Documents (workshop to Conference of the North American Chapter of the Association for Computational Linguistics), June 2010

[M.10] **Cadag, E.**, Tarczy-Hornoch, P., Myler, P.J., "On the reachability of trustworthy information from integrated exploratory biological queries", Proc. Data Integration in the Life Sciences (DILS), LNCS 5647:55-70, July 2009

[M.11] **Cadag, E.**, Louie, B., Myler, P.J., Tarczy-Hornoch, P., "BioMediator data integration and inference for the functional annotation of anonymous sequences", Proc. Pacific Symposium on Biocomputing (PSB), 12:343-354, January 2007

[M.12] El-Sayed, N., *et al.* "Comparative genomics of Trypanosomatid parasitic protozoa" Science, Vol 309, Issue 5733, July 2005

[M.13] Ivens A.C., *et al.* "The genome of the kinetoplastid parasite, Leishmania major" Science, Vol 309, Issue 5733, July 2005

[M.14] El-Sayed N.M., *et al.* "The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas' disease" Science, Vol 309, Issue 5733, July 2005

[M.15] Schilit, B., LaMarca, A., McDonald, D., Tabert, J., **Cadag, E.**, Boriello, G., Griswold, W. "Bootstrapping the Location-enhanced World Wide Web" Proceedings of the Location-Aware Computing Workshop, UbiComp 2003

Selected Presentations & Abstracts

[A.1] Pike, N., Kim, K., Park, E., Armstrong, C., Pham, L., **Cadag, E.**, "A Unique Combination of SMarTR Bioinformatics and Cardiomyocyte Screening Technology Allows for Development of Novel Cardio-protective Drugs", Basic Cardiovascular Sciences, Phoenix, Arizona, July 2016

[A.2] Radecki, R.P., **Cadag, E.**, Petrossian, T., "Topological Visualization Uncovers Novel Clinically Relevant Clusters", Proceedings of the AMIA, Washington DC, November 2013

[A.3] **Cadag, E.**, Huang, A., Navid, A., Zhou, C., Dhaeseleer, P., "Integration of whole-genome functional annotations from multiple sources, applied to human pathogens", Proceedings of the AMIA Summit on Translational Bioinformatics, San Francisco, California, March 2012

[A.4] **Cadag, E.**, Vitalis, E., Lennox, K., Zhou, C., Zemla, A., "Methods for Structure-Based Computational Analysis of Protein Variation Applied to the Emerging Antibiotic Resistance Threat NDM-1 ", Defense Threat Reduction Agency Chemical and Biological Defense Science and Technology Conference, Las Vegas, NV, November 2011

[A.5] Gunnell, M., Slater, S., Brunjes, D., **Cadag, E.**, Bennett, B., Madonna, A., "An Approach to BioWatch Gen-3 Testing: Methodology, Instrumentation, and Statistical Analysis", Department of Homeland Security National BioWatch Conference; Milwaukee, WI; August 2010

[A.6] Halgrim, S.R., Xia, F., Solti, I., **Cadag, E.**, Uzuner, Ö., "Statistical extraction of medication information from clinical records" Proceedings of the AMIA Summit on Translational Bioinformatics, San Francisco, California, March 2010

[A.7] **Cadag, E.**, Tarczy-Hornoch, P., "Supporting retrieval of diverse biomedical data using evidence-aware queries" American Medical Informatics Association (AMIA), November 2009. San Francisco, California

[A.8] Uzuner, Ö., Solti, I., **Cadag, E.**, "i2b2 NLP Challenges" i2b2 2009 NLP Challenge Workshop, November 2009 (workshop of AMIA). San Francisco, California

[A.9] **Cadag, E.**, Tarczy-Hornoch, P., Myler, P. J., "Learning pathogenic proteins across fractured and heterogeneous data", Proc. of the American Medical Informatics Association; Wash. D.C.; November 2008, (*Best Poster Award*)

[A.10] **Cadag, E.**, Aggarwal, G,. Anupama, A., Myler, PJ "Polycistronic-intelligent consensus and syntenous gene-finding in Trypanosomes" TriTryp Genomes Meeting, September 2004. Seattle, Washington

Patents

[P.1] Park, E., Lam, J., Armstrong, C., Zhong, Z., Lum, P., **Cadag, E.**, Pham, L. "Pharmaceutical Compositions and Methods for Countering Induced Cardiotoxicity ", USPTO App. No. US 15/424,743, 3 Feb 2017

[P.2] Vivero, D., Billerey-Mosier, R., Shah, M., Shah, S., **Cadag, E.**, Caballero, J., Chanthasiriphan, K., Yarlagadda, A., Frandsen, E. "Entity Cohort Identification and Entity Profiling.", USPTO App. No. US14/990,766, 7 Jan 2016

[P.3] Vivero, D., Caballero, J, Othman, A.M., Jamshidian, F., **Cadag, E.**, Yarlagadda, A., Fingal, J., Dunkman, N., Shah, K.D., Strauser, W., Lin, M., Hampden, M., Shah, S., Ackermann, R. "Data Analysis Mechanism for Generating Statistics, Reports and Measurements for Healthcare Decisions.", USPTO App. No. US 14/990,744, 7 Jan 2016

[P.4] Lum, P.Y, **Cadag, E.**, Grahnen, J., Lewis, J. "Consensus Sequence Identification.", USPTO App. No. US 14/597,156, 14 Jan 2015

Media and Service

[S.1] Invited panelist for the 9th Annual Berkeley-Haas Healthcare Conference, session "Data Science in Health: Lessons from the Field", Nov 2015; San Francisco, California

[S.2] Featured interviewee for *The Data Science Handbook* (ISBN 978-0-692-43487-1), "Ch. 11: Eithon Cadag – Bridging the Chasm: From Bioinformatics to Data Science" (p.128-142) by C. Shan, H. Wang, W. Chen and M. Song., 2015

[S.3] Invited panelist for the 4th Annual Clinical Genome Conference, session "Hadoop and High-Dimensional Data Mining in the Era of the Clinical Genome", Jun 2014; San Francisco, California

[S.4] Co-organizer for the Informatics for Integrating Biology & the Bedside (i2b2) Natural Language Processing Challenge Workshop, 2009; Washington DC