Nicholas (Nick) Ball

Chief Data Scientist ▪ Director of Data Science

Solving real-world problems through data science and machine learning.

- **Versatile data scientist driving product quality, user value and research.** Articulate technical generalist with a background in astrophysics and a perspective of data science software platform development.
- **Synthesize and create roadmaps to design data science platforms for end-to-end data science from raw data through to production.** Highly systematic approach.
- **Confident introvert able to clearly communicate with audiences of all types** through presentations, publications, blogs, technical product demos and user documentation. Train, mentor and lead teams.
- **Openminded, grasp both big ideas and detail, highly organized, responsible**, straightforward, honest, logical and even-tempered.

DATA SCIENCE EXPERIENCE

Principal Data Scientist (Product) — **Dotscience**
HQ in UK / Fully Remote Team — Sep 2019 – May 2020

Dotscience was founded and funded in 2017 by DDN, the world's largest privately held storage company, to build an end-to-end data science platform using MLOps (principles of software DevOps for machine learning) for faster deployment and to hold AI accountable for its decision making. (Product launched July 2019; company closed May 2020 due to COVID-19.)

Served as de facto chief data scientist and head of product, working with CEO and VP Engineering to shape the development of the Data Science platform with underlying MLOps architecture to deliver specific value to data scientist users.

- **Provided strategic input to product direction, developed user documentation and tutorials, prepared product demos, and wrote and presented on product capabilities**. Served as the company go-to and manager for data science and machine learning.
- **Owned and drove company roadmap to focus on MLOps,** creating and prioritizing 80+ roadmap items across all aspects of data science, driven by what data scientists need to solve business problems from end to end, including production.
- **Developed demonstrations of Dotscience functionality for data scientists**, e.g. using TensorFlow and H2O for deep learning with hyperparameter tuning, creating real machine learning models for demos or presentations.
- **Conducted extensive product testing from advanced user point of view. Gathered from and shared feedback with product**, front- and back-end engineering, design (UX), sales, marketing, and the CEO.
- **Wrote 5 technical blog entries on topics such as why data scientists need DevOps for machine learning**; how to deploy models to production in 60 seconds; Dotscience competitive advantages; and coding with the Dotscience Python library.

Lead Data Scientist — **Oracle Corporation**
Redwood Shores, CA — Oct 2018 – May 2019

Database giant Oracle built their own cloud infrastructure (OCI) and acquired datascience.com as part of building the new Oracle data science platform to be launched in 2019. (Part of large-scale layoff.)

Joined group led by former datascience.com CEO to become one of two data scientists working on the product team. Worked closely with the ODSS (Oracle Data Science Service) engineering team, bridging the outward-facing (customer) and inward-facing (product and engineering) aspects of the group's data science work.

- **Learned the ODSS product and OCI environment** and worked with other Oracle teams on product directions on algorithms, the data science process, product integrations, and research.
- **Produced demo content on fraud detection and a talk on data science pitfalls**. Supported customer conversations and visits. Staffed product booth at annual Oracle conference.

650.690.2675
San Francisco Bay Area, CA 94030
linkedin.com/in/nickballdatascientist
www.nickballdatascience.com
nick@nickballdatascience.com

TOOLBOX

Data Science End-to-End (full stack)

Machine Learning & Artificial Intelligence

Customer Data Science Projects

Product Roadmaps for Data Science

Academic Research for Publications

Public Speaking, Publications and Communications

EDUCATION

Ph.D. Astronomy
University of Sussex, UK
2004

Studied the large-scale structure of the universe. Thesis: Galaxy Types, Luminosity Functions & Environments in the Sloan Digital Sky Survey

M.Sc. Astronomy
University of Sussex, UK
Distinction | 2001

B.A. Natural Sciences (Geological Sciences)
Cambridge University, UK
2000

In final year, organized field project mapping an area of the Rocky Mountains in Montana, resulting in co-authoring my first publication

DATA SCIENCE EXPERTISE

Automated machine learning (auto-ML)
Data featurization
Data preparation
Decision trees (gradient boosting, random forest)
Dotscience
H2O.ai
Large datasets
Machine learning models in production
Neural networks & deep learning
Skytree
Supervised machine learning
TensorFlow
Unsupervised machine learning

(continued)

Principal Data Scientist — Infosys

Palo Alto, CA — Mar 2017 – Oct 2018

Infosys provides IT services, primarily in India, and wanted to add AI products to its service portfolio for its global customer base. It acquired Skytree's technology and team in 2017.

Joined with the acquisition of Skytree, staying with the team to continue working on the world-class machine learning product, with the potential for it to become global. Served as technical data science representative with salespeople and engineers, occasionally leading calls.

- **Led PoC (Proof of Concept) project for a bank for fraud detection** using supervised machine learning to predict patterns for fraudulent transactions.
- **Project lead for the group's data science work, developing product roadmap and generating dozens of Jira tickets** and epics for engineers. Led testing and feedback from data scientist user point of view.
- **Updated Skytree demo material showcasing machine learning usage** including credit risk, customer churn, fraud detection, lead scoring and spam filtering, with videos for prospects.

Staff Data Scientist — Skytree

San Jose, CA — Nov 2012 – Mar 2017

Skytree was founded in 2010 as a platform to do machine learning, based on the work of the FASTLab group at Georgia Tech. The company grew from 20 to 50 people in my time there. (Acquired by Infosys.)

Recruited into a world-class group of data scientists by the CEO after being one of Skytree's first customers while at the Herzberg Institute. Combined research responsibilities and work on PoC projects and product roadmaps, feedback and demos.

- **Led Skytree benchmarking project with data science and engineering colleagues** to prove that Skytree was more than 100x faster than competitors.
- **Served as lead data scientist on 20+ projects including 5 of Skytree's 10 paying customers,** 10 other PoCs and pre-PoC projects. Generated product feedback totaling hundreds of Jira tickets (roadmap epics, technical, user experience, presentation), significantly improving the product.
- **Worked on two astronomy projects;** one for SETI to use machine learning to classify signals received and the other with Lawrence Berkeley Lab on large-scale structures in the universe.
- **Generated videos and blogs for the website about Skytree,** including technical differentiators, how to download and install, speed and automation, and demos of machine learning across topics. Outlined, scripted, did live demos. Represented Skytree solo at trade show.

ACADEMIC RESEARCH EXPERIENCE

Assistant Research Officer — Herzberg Institute of Astrophysics

Victoria, BC, Canada — Jul 2009 – Jul 2012

Herzberg Institute of Astrophysics is the astronomy institute of Canada's National Research Council.

Hired for the CANFAR project (a precursor to cloud computing, but for astronomy) including the Next Generation Virgo Cluster Survey (NGVS). In a newly created position, was the only scientist with an overlap of astronomy and data science, focused on data science for publication.

- **Completed "Data Mining and Machine Learning in Astronomy,"** one of the first major reviews in the world on this topic (invited review, International Journal of Modern Physics D, 61 pages).
- **Created one of the first large-scale astronomy data mining systems in the world,** via collaboration with Skytree: CANFAR+Skytree. Negotiated with Skytree to use their $300K software for $3K.
- **Led the Luminosity Function Science Working Group** of the NGVS and also the Photometric Redshifts subgroup. Showed machine learning could resolve which galaxies are in the Virgo Cluster of Galaxies.

Postdoctoral / Research Scientist — University of Illinois at Urbana-Champaign

Urbana, IL — Nov 2004 – Jun 2009

UIUC's astronomy department and the National Center for Supercomputing Applications (NCSA) were part of the nascent astronomy data mining community.

As a postdoc, directed research and publications and served as Principal and Co-Investigator on funding grants including from NASA.

- **Produced 3 publications of original research in the *Astrophysical Journal***, including the first sample of quasars with accurate distances from just images. Created some of the first applications of machine learning to astronomical datasets of over 100 million objects, using NCSA supercomputers.

COMPUTING TECHNOLOGY

C
Cloud (Amazon AWS, Google GCP, Microsoft Azure, Oracle OCI)
LaTeX typesetting
MATLAB
Microsoft Office (Excel, PowerPoint, Word)
PySpark
Python
Shell scripting (Bash)
SQL
UNIX/Linux

ACADEMIC RESEARCH

Astronomy & astrophysics
Astroinformatics
Geology
Scientific writing & publications

TRANSFERABLE SKILLS

Collaborative and cross-functional problem solving
Communication to technical and nontechnical audiences (presentations, reports, speaking, writing)
Mentoring and teamwork
Technical writing
Training

ON THE SIDE

Toastmasters International Competent Communicator

World Scrabble Championship team member for UK, USA, and Canada, Peak rankings include top 50 world ranking and top 10 in the UK

Certified tournament director for the North American Scrabble Players Association

Wrote the wiki pages for NASPA on how American players can play with the International/English dictionary

Lifetime interest in astronomy, science and tech

Led tours and demonstrations of historical telescopes for the Cambridge University Astronomical Society

Global citizen: I'm English (US Green Card), my wife is Taiwanese, our son is American

Nick Ball Addendum

PUBLICATIONS AND PATENTS: SUMMARY

- Data Mining and Machine Learning in Astronomy, International Journal of Modern Physics D, 2010
- Lead author, 6 publications in major peer-reviewed astronomical journals (Monthly Notices of the Royal Astronomical Society; Astrophysical Journal)
- Lead or co-author, 15 other publications, beginning aged 20
- Total citations 1263; 850 as first author; h-index = 11 (Google Scholar, Jun 2020)
- Coauthor of 2 patents for Skytree

SELECTED PUBLICATIONS: DETAILS

Raichoor A., et al., 2014, The Next Generation Virgo Cluster Survey. XV. The Photometric Redshift Estimation for Background Sources, ApJ 797 102

Ball N.M., 2013, CANFAR+Skytree: A Cloud Computing and Data Mining System for Astronomy. Astronomical Data Analysis Software & Systems (ADASS) XXII, ASP Conference Proceedings, eds. Friedel D., Freemon M., Plante R. (San Francisco: ASP)

Ball N.M., 2012, Utilizing Astroinformatics to Maximize the Science Return of the Next Generation Virgo Cluster Survey. Astrostatistics and Data Mining in Large Astronomical Databases, Springer

Ferrarese L.F., et al., 2012, The Next Generation Virgo Cluster Survey (NGVS). I. Introduction to the Survey, ApJS 200 4

Gaudet S., Hill N., the CADC team, 2010, CANFAR: the Canadian Advanced Network for Astronomical Research. Proc. SPIE, Software and Cyberinfrastructure in Astronomy, eds. Radziwill N.M. and Bridger A., 7740-1L

Ball N.M. & Brunner R.J., 2010, Data Mining and Machine Learning in Astronomy, International Journal of Modern Physics D, 19(7), 1049

Myers A.D., White M. and **Ball N.M.**, 2009, Incorporating photometric redshift probability density information into real-space clustering measurements, MNRAS 399 2279

Ball N.M., et al., 2008, Robust Machine Learning Applied to Astronomical Datasets III: Probabilistic Photometric Redshifts for Galaxies and Quasars in the SDSS and GALEX, ApJ 683 12

Ball N.M., Loveday J. & Brunner R.J., 2008, Galaxy Colour, Morphology, and Environment in the Sloan Digital Sky Survey, MNRAS 383 907

Ball N.M., et al., 2007, Robust Machine Learning Applied to Astronomical Datasets II: Quantifying Photometric Redshifts for Quasars Using Instance-Based Learning, ApJ 663 774

Ball N.M., et al., 2006, Bivariate Galaxy Luminosity Functions in the Sloan Digital Sky Survey, MNRAS 373 845

Ball N.M., et al., 2006, Robust Machine Learning Applied to Astronomical Datasets I: Star-Galaxy Classification of the SDSS DR3 Using Decision Trees, ApJ 650 497

Ball N.M., et al., 2004, Galaxy Types in the Sloan Digital Sky Survey Using Artificial Neural Networks, MNRAS 348 1038

Moore J.R., Watts M.D., **Ball N.M.**, Ratcliff M.D., 2000, The Geology of the Ford Creek Area, Northern Montana Thrust Belt. Montana/Alberta thrust belt and adjacent foreland. 50th Anniversary Symposium of the Montana Geological Society, Volume 1, pp 151–168, Montana Geological Society, Billings, Montana, USA