Dr. Rob Lyon

rob@scienceguyrob.com | http://www.scienceguyrob.com

EDUCATION

UNIV. OF MANCHESTER
PH.D. MACHINE LEARNING
Sept. 2015

UNIV. OF LIVERPOOL
M.SC. ADVANCED COMPUTER SCI.
Sept. 2011 | Distinction

B.SC. SOFTWARE DEVELOPMENT
Sept. 2008 | First-class honours (1st)

SKILLS

GENERAL EXPERTISE
Data Analysis • Performance Analysis • Independent Research • Technical Writing• Software Engineering & Testing • Agile Software Development • Public Speaking • HPC • Leadership

PROGRAMMING
Advanced Proficiency:
Java • C++ • C • C# • Python • Matlab • LaTeX •HTML • Apache Storm • Scikit Learn
Intermediate Proficiency:
JavaScript • SQL • .NET • iOS • Hadoop • Cuda
Basic Proficiency:
R • Scala • Julia • OpenCL

MACHINE LEARNING
Classifier design • Data exploration • Feature Engineering • Optimisation • Stream mining • On-line learning • Pipeline design • High volume streams • Real-time Machine Learning

LINKS

Github:// **scienceguyrob**
LinkedIn:// **roblyon86**
Twitter:// **@scienceguyrob**

ABOUT ME

Hi I'm Rob. I'm a post-doctoral researcher. I work in the School of Physics & Astronomy, at the University of Manchester (SKA Group). I develop resource efficient machine learning methods for use with the worlds largest radio telescope, the Square Kilometre Array (SKA). I'm a member of the Central Signal Processor (CSP) and Science Data Processor (SDP) design consortia, both tasked with designing the telescope. I'm an active member of the pulsar and machine learning research communities. I'm also a software engineer with industry experience. Using that experience I've authored multiple software design documents for the pulsar search sub-element of the SKA. I'm an enthusiastic public speaker. I've been lucky enough to deliver talks around the globe. I passionately believe in sharing knowledge and helping others. My research interests include imbalanced learning problems and real-time classification tasks. I'm especially interested in helping solve problems impacting the lives of ordinary people. I firmly believe that everyone should benefit from advances in machine intelligence, particularly if it improves their health and well-being.

EXPERIENCE

SCHOOL OF PHYSICS & ASTRONOMY | POST-DOC RESEARCHER
Sept. 2015 – Present | University of Manchester, UK

- Non-imaging processing machine learning lead.
- Designed the analytics pipeline for the SKA, the world's most sophisticated scientific instrument.
- Tasked with reporting project progress to stakeholders (in UK and RSA).
- Liaised with academic colleagues and industry stakeholders (Arup, HP, Lenevo, IBM) when tackling SKA design challenges.
- Developed prototype processing pipelines for the CSP and SDP consortia.
- Re-architectured off-line processing procedures, enabling their on-line execution across heterogeneous compute facilities.
- Worked on the procurement of a prototype HPC cluster for use in the South African desert (value approx. £250,000).
- Co-supervised PhD and masters students working in machine learning.
- Supervised a summer project in musical genre classification.

CENTRE FOR DOCTORAL TRAINING | RESEARCH STUDENT
Sept. 2011 – Sept. 2015 | University of Manchester, UK

- Studied in the first UK-based Centre for Doctoral Training (CDT) specialising in computer science. The research was fully funded by The Engineering and Physical Sciences Research Council (EPSRC), for which I'm very grateful.
- My research focused on devising machine learning classifiers for the SKA. The aim was to produce a classifier able to identify pulsar signals in high throughput astronomical data streams.
- Worked extensively on data stream mining, imbalanced learning, and feature creation & selection.
- The research has helped discover >1% of all pulsars known to science.

MERSEYSTEM | STEM AMBASSADOR
Nov. 2011 – August 2012 | Liverpool, UK

- Supported a weekly STEM club in secondary school, running computer science workshops which taught programming principles.

OPEN SOFTWARE

Some software arising from my work:

SKA Data Models

A Jupyter notebook that describes SDP data rates & volumes: doi:10.5281/zenodo.836715.

Pulsar Feature Lab

A python library useful for extracting machine learning features: doi:10.6084/m9.figshare.1536472.v1.

Stuffed

Enables classifier testing and evaluation on unlabelled data streams: doi:10.6084/m9.figshare.1536471.v1.

OPEN DATA

Pulsar Survey Database

This database lists every major pulsar survey conducted since 1967: doi:10.6084/m9.figshare.3114130.v1.

HTRU2

A sample of pulsar candidates obtained using the Parkes telescope: doi:10.6084/m9.figshare.3080389.v1.

REFEREES

Available on request.

APPSENSE | Performance Analyst
July 2008 – August 2010 | Daresbury, UK

- Responsible for analysing software performance and scalability.
- Designed and developed test harnesses that stressed terminal server/IIS web server/SQL server applications in conjunction and isolation.
- Responsible for analysing performance data, making recommendations, and producing summary reports/white-papers for non-technical staff.
- Led the performance team prior to leaving for further university study.

PUBLICATIONS

2018 **"A Big Data Pipeline for High Volume Scientific Data Streams"**, submitted to Data Mining & Knowledge Discovery.

2018 "**Imbalanced Learning In Astronomy"**, EWASS, April 4-6 (accepted).

2017 **"Ensemble candidate classification for the LOTAAS pulsar survey"**, Monthly Notices of the Royal Astronomical Society, Volume 474, Issue 4, doi:10.1093/mnras/stx3047.

2017 **"Pulsar Searches with the SKA"**, Pulsar Astrophysics - The Next 50 Years Proceedings IAU Symposium No. 337.

2017 **"50 Years of Candidate Pulsar Selection – What next?"**, Pulsar Astrophysics - The Next 50 Years Proceedings IAU Symposium No. 337.

2016 **"Why are pulsars so hard to find?"**, University of Manchester.

2016 **"Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach"**, Monthly Notices of the Royal Astronomical Society, 459 (1): 1104-1123, doi:10.1093/mnras/stw656.

2014 **"Hellinger Distance Trees for Imbalanced Data Streams"**, ICPR, pp.1969-1974, doi:10.1109/ICPR.2014.344.

2013 **"Classification in Imbalanced and Partially-Labelled Data Streams"**, in Simple and Effective Machine Learning for Big Data, Special Session, IEEE International Conference on Systems, Man, and Cybernetics, doi:10.1109/SMC.2013.260.

ACADEMIC RESPONSIBILITIES

- Reviewer for the Genetic and Evolutionary Computation Conference (GECCO), 2015-2018, Evolutionary Machine Learning track.
- Reviewer for the Monthly Notices of the Royal Astronomical Society.
- Co-supervising two interdisciplinary PhD students at present.

GRANTS

- Awarded an Amazon **Astro Compute Grant** along with colleagues (value equal to £20,000). Research outcomes include the paper, **"A Big Data Pipeline for High Volume Scientific Data Streams"**, which is listed above.

AWARDS

2011 Best Presentation, University of Liverpool.
2008 Best Final Year Software Project, University of Liverpool.
2007 Deloitte Award for Best Group Project, University of Liverpool.