

Dr. Rob Lyon

Merseyside, United Kingdom

rob@scienceguyrob.com | <http://www.scienceguyrob.com> | Phone number provided on request



NATIONALITY

BRITISH

EDUCATION

UNIV. OF MANCHESTER

PH.D. MACHINE LEARNING
Sept. 2016

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 • Teaching

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](#)

Skype: [rob_lyon](#)

EXPERIENCE

SCHOOL OF PHYSICS & ASTRONOMY | POST-DOC RESEARCHER

Sept. 2015 – Present | University of Manchester, UK

- Machine learning lead for the Square kilometre Array (SKA) group. Developed new machine learning methods for pulsar and transient search.
- Helped design the analytic and signal search pipelines for the SKA.
- Reported SKA project progress to consortia leads, managed the SKA pulsar search sub-element requirements and test specifications.
- Collaborated with academic colleagues and industry stakeholders to tackle SKA design challenges, delivered talks describing design work when required.
- Authored and co-authored multiple SKA design documents.
- Developed prototype processing pipelines for the SKA design consortia.
- Re-architected 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).
- Principal investigator for STFC funded (initial value £16,000) radiotherapy & machine learning project, aimed at improving cancer treatment in the NHS.
- Currently co-supervising a DARA Big Data PhD student (1st year), and another radio astronomy PhD student (3rd year).

SCHOOL OF COMPUTER SCIENCE | RESEARCH STUDENT

Sept. 2011 – Sept. 2015 | University of Manchester, UK

- Studied the challenges associated with isolating and correctly classifying signals of interest in astronomical data processing pipelines.
- Devised machine learning classifiers able to accurately identify pulsar signals in high volume data streams.
- Increased the known pulsar population by >1% via application of my research.
- Specialised in stream mining, imbalanced learning, and feature engineering.

MERSEY SYSTEM | STEM AMBASSADOR

Nov. 2011 – August 2012 | Liverpool, UK

- Held workshops which taught school students computer programming.
- Delivered talks aimed at encouraging young people pursue a science career.

APPSENSE (NOW IVANTI) | PERFORMANCE ANALYST

July 2008 – August 2010 | Daresbury, UK

- Responsible for analysing software performance and scalability.
- Developed test harnesses for multi-tier client-server architectures.
- Studied performance data, produced summary reports/white-papers for non-technical staff, made recommendations to management.
- Led the performance team prior to leaving for further university study.

REFEREES

On request.

Dr. Rob Lyon

Merseyside, United Kingdom

rob@scienceguyrob.com | <http://www.scienceguyrob.com> | Phone number provided on request

OPEN SOFTWARE

Some software arising from my work:

Imbalanced Learning in Astronomy

A Jupyter notebook supporting my EWASS talk & paper:
doi:10.5281/zenodo.1212631.

Big Data Pipeline

A Jupyter notebook describing algorithms supporting the paper, "A Big Data Pipeline for High Volume Scientific Data Streams":
doi:10.5281/zenodo.1116302.

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.

Test Vector Pipeline

A Docker-based pipeline that produces reusable SKA test-vectors:
doi:10.5281/zenodo.1165435.

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.

AWARDS

2011

Best M.Sc. Presentation, University of Liverpool.

2008

Best Final Year Software Project, University of Liverpool.

2007

Deloitte Award for Best Group Project, University of Liverpool.

PUBLICATIONS

- 2018 "Single-pulse classifier for the LOFAR Tied-Array All-sky Survey", submitted to Monthly Notices of the Royal Astronomical Society.
- 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.
- 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", Proceedings IAU Symposium No. 337.
- 2017 "50 Years of Candidate Pulsar Selection – What next?", 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.

TECHNICAL REPORTS

Much of my research output produced since 2015, has consisted of SKA technical reports and design documents. I've listed the most important such work below.

- 2017 "SKA1 CSP Low Pulsar Search Sub-element Requirement Specifications (ED-1a)", submitted to the SKA Office.
- 2017 "SKA1 CSP Mid Pulsar Search Sub-element Requirement Specifications (ED-1b)", submitted to the SKA Office.
- 2017 "SKA1 CSP Low Pulsar Search Sub-element Test Specification (ED-3a)", submitted to the SKA Office.
- 2017 "SKA1 CSP Mid Pulsar Search Sub-element Test Specification (ED-3b)", submitted to the SKA Office.

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, both working in the pulsar/transient search domain.

GRANTS

- Awarded STFC Radio Therapy "Sandpit" funds (initial £16,000) - on-going.
- Awarded an Amazon **Astro Compute Grant** (value equal to £20,000). Yielded the paper, "A Big Data Pipeline for High Volume Scientific Data Streams".