

Dr. Rob Lyon

Address Redacted, United Kingdom

rob@scienceguyrob.com | <http://www.scienceguyrob.com> | Contact on Request



NATIONALITY

BRITISH & IRISH (EU NATIONAL)

EDUCATION

UNIV. OF MANCHESTER

PH.D. MACHINE LEARNING

Sept. 2011 - Sept. 2015

UNIV. OF LIVERPOOL

M.Sc. ADVANCED COMPUTER SCI.

Aug. 2010 - Sept. 2011 | Distinction

B.Sc. SOFTWARE DEVELOPMENT

Sept. 2004 - July. 2008 | 1st Class

SKILLS

GENERAL EXPERTISE

Data Analysis • Performance Analysis • Independent Research • Technical Writing • Software Engineering & Testing • SAFe Agile Certified • Public Speaking • HPC • Leadership

PROGRAMMING / TOOLS

Advanced Proficiency:

Java • Python • Matlab • \LaTeX

• HTML • Apache Storm • Scikit Learn

Intermediate Proficiency (or rusty!):

C++ • SQL • .NET • C • C# • Cuda

Basic Proficiency:

R • Scala • OpenCL

MACHINE LEARNING

Classifier design • Data exploration • Feature Engineering • Optimisation • Stream mining • Prediction • Pipeline design • High volume streams • Real-time Machine Learning • Data visualisation

LINKS

Github:// [scienceguyrob](#)

LinkedIn:// [roblyon86](#)

Twitter:// [@scienceguyrob](#)

Skype: On request.

SUMMARY

Hi I'm Rob. I've been working in data science/machine learning for almost 10 years now. I'm highly skilled in multiple areas - from predictive analytics and software engineering, to bleeding-edge scientific research. I've taken part in large industrial projects, led teams, managed juniors, and managed my own independent projects. In recent years I've helped architect the analytic pipeline for the worlds most advanced scientific instrument, the Square Kilometre Array (SKA). I'm ambitious, hard-working, and pride myself on attention to detail. This has helped me become the go-to expert in my field.

EXPERIENCE

DEPARTMENT OF COMP. SCI. | LECTURER IN AI & ROBOTICS

Sept. 2019 – Present | Edge Hill University, UK

- Module leader for various courses teaching subjects from introductory programming to machine learning.
- Author and teacher of machine learning, data science and Python courses for the Institute of Code's TechUP Women programme.
- Delivered master classes in machine learning to students at all levels.

SCHOOL OF PHYSICS & ASTRONOMY | ML RESEARCHER

Sept. 2015 – Sept. 2019 | University of Manchester, UK

- Machine learning lead for the Square kilometre Array (SKA) group. Developed new machine learning methods for radio astronomy.
- Helped design the analytic pipelines for the SKA telescope.
- Reported SKA project progress to consortia leads, managed 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 design documents.
- Developed prototype processing pipelines for the design consortia.
- Re-architected off-line processing procedures, enabling their on-line execution across heterogeneous compute facilities.
- Assisted procurement of prototype HPC cluster (value approx. £250,000).
- Co-supervised PhD and masters students working in machine learning.
- Organised and led three data science workshops for medical researchers.
- Produced learning materials aimed at helping students understand machine learning.

SCHOOL OF COMPUTER SCIENCE | RESEARCHER

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

- Awarded a fully funded position (awarded by EPSRC) in the first UK-based centre for doctoral training specialising in computer science.
- Studied the challenges associated with isolating signals of interest in radio data, as part of the Physics department in Manchester.
- Devised machine learning classifiers able to accurately identify signals of interest in high volume data streams.
- Discovered many new Radio Pulsars (a rare type of star) via my work.
- Specialised in stream mining, imbalanced learning, and feature engineering.

OPEN SOFTWARE

Some software arising from my work:

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.

AI4Astro Masterclass

A master-class that teaches programming and machine learning: doi:10.5281/zenodo.3525993.

IOC-Techup Content

Teaching material written for the IOC-Techup Women programme.

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

2008

Best Final Year Software Project, University of Liverpool.

2007

Deloitte Award for Best Group Project, University of Liverpool.

REFEREES

Available on request.

APPSENSE INC. (NOW IVANTI) | SOFTWARE ENGINEER

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.

PUBLICATIONS

- 2020 "Imbalance learning for variable star classification", MNRAS, Volume 493, Issue 4, doi:10.1093/mnras/staa642.
- 2019 "Comparing Multi-class, Binary and Hierarchical Machine Learning Classification schemes for variable stars", MNRAS, Volume 488, Issue 4, doi:10.1093/mnras/stz1999.
- 2019 "A Big Data Pipeline for High Volume Scientific Data Streams", Astronomy & Computing Volume 28, doi:10.1016/j.ascom.2019.100291.
- 2018 "Single-pulse classifier for the LOFAR Tied-Array All-sky Survey", MNRAS, Volume 480, Issue 3, doi:10.1093/mnras/sty2072.
- 2018 "Imbalanced Learning In Astronomy", EWASS, April 4-6.
- 2017 "Ensemble candidate classification for the LOTAAS pulsar survey", MNRAS, Volume 474, Issue 4, doi:10.1093/mnras/stx3047.
- 2017 "Pulsar Searches with the SKA", Proc. IAU Symposium No. 337.
- 2017 "50 Years of Candidate Pulsar Selection - What next?", Proc. 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", MNRAS, 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

First author for various architectural documents for the SKA telescope e.g. "SKA1 CSP Low Pulsar Search Sub-element Requirement Specifications (ED-1a)" or the "SKA1 CSP Low Pulsar Search Sub-element Test Specification (ED-3a)".

ACADEMIC RESPONSIBILITIES

- Reviewer for the Genetic and Evolutionary Computation Conference (GECCO), Astronomy & Astrophysics, and MNRAS.
- Co-supervising one interdisciplinary PhD students at present, working in the machine learning domain. Leader of 4 taught modules.
- Equality and diversity committee member, internal REF panel reviewer.

GRANTS

- Amazon **Astro Compute Grant** (value equal to £20,000). Outcomes include the paper, "A Big Data Pipeline for High Volume Scientific Data Streams".
- Principal investigator for a £16,000 initiative funded by the STFC, which aims to utilise machine learning for radiotherapy treatment (more information available at the project website: <http://radiotherapymlnetwork.co.uk/>).