Data science consultancies at NeSI A whirlwind tour of case studies

Maxime Rio and Alexander Pletzer contact: maxime.rio@nesi.org.nz

NeSI / NIWA

eResearch NZ 2021





support@nesi.org.nz



Source: Drew Conway (CC BY-NC)

NZ RSE 2020 Hackathon: bringing people together

- ML challenge at NZ RSE 2020 (Testimonia network in the second sec
- Dataset from QuakeCoRE: predict job runtime given job parameters
- Learning opportunity for everyone, sharing EDAs and models



Best results from Nick Young (CeR)

Influenza forecasting: HPC for hyperparameters tuning

with Nooriyah Lohani (University of Auckland)

- Forecast influenza-like illness, 3+1 years of data (train/test)
- Fit a Prophet model (Tacebook/prophet) to be used as a baseline
- Use Bayesian optimization to select hyperparameters
- Configure HPC as a backend to accelerate temporal cross-validation
- Upskilling: running notebooks in JupyterHub/Lab (jupyter.nesi.org.nz)



Rainfall ML: workflow management for machine learning with Abha Sood, Olaf Morgenstern, Stephen Stuart and Neelesh Rampali (NIWA)

- Machine learning for downscaling seasonal weather prediction
- NIWA's climate station data, monthly precipitations from 1945 to 2017
- Comparison of multiple models (from linear regression to neural nets)
- Models lifecycle organized using Snakemake for workflow managment



Rainfall ML: workflow management for machine learning with Abha Sood, Olaf Morgenstern, Stephen Stuart and Neelesh Rampali (NIWA)

- Machine learning for downscaling seasonal weather prediction
- NIWA's climate station data, monthly precipitations from 1945 to 2017
- Comparison of multiple models (from linear regression to neural nets)
- Models lifecycle organized using Snakemake for workflow managment



Car crash predictions: exploratory data analysis with Nick Young (Centre for e-Research, University of Auckland)

- Open data from NZTA: Crash Analysis System, \approx 700000 records
- 20 years of data, record weather condition, road features, vehicles, etc.
- EDAs and models available online (The neon-ninja/crash_prediction)





Car crash predictions: exploratory data analysis with Nick Young (Centre for e-Research, University of Auckland)

- Open data from NZTA: Crash Analysis System, \approx 700000 records
- 20 years of data, record weather condition, road features, vehicles, etc.
- EDAs and models available online (The neon-ninja/crash_prediction)



Porting EchoCV: Deep Learning on NeSI

with Nathan Russell (University of Otago)

- Porting o rahuldeo/echocv to run on NeSI's GPUs
- Deep learning toolbox for echocardiography image segmentation
- Code reuse challenges: Python 2.7, "old" Tensorflow code, non-generic code...
- Models sensitivity to input data





Eye tracking: from data acquisition to data analysis with Dr. Matthew McDonald (University of Auckland / Mātai)

- Eye tracking using a wearable device (glasses with cameras)
- Analysis of gaze trajectories to diagnose concussion
- Acquisition code: realtime stimulus presentation
- Analysis code: post-processing step to segment gaze trajectories





Eye tracking: from data acquisition to data analysis with Dr. Matthew McDonald (University of Auckland / Mātai)

- Eye tracking using a wearable device (glasses with cameras)
- Analysis of gaze trajectories to diagnose concussion
- Acquisition code: realtime stimulus presentation
- Analysis code: post-processing step to segment gaze trajectories



Eye tracking: from data acquisition to data analysis with Dr. Matthew McDonald (University of Auckland / Mātai)

- Eye tracking using a wearable device (glasses with cameras)
- Analysis of gaze trajectories to diagnose concussion
- Acquisition code: realtime stimulus presentation
- Analysis code: post-processing step to segment gaze trajectories



Conclusion

For the consultancy service, doing data science means:

- bringing people together,
- HPC for hyperparameters tuning,
- workflow management for machine learning,
- exploratory data analysis,
- deep Learning on NeSI,
- data acquisition and data analysis.

Conclusion

For the consultancy service, doing data science means:

- bringing people together,
- being passionate about their research,
- empowering scientists with new analysis tools,
- helping people to get the best out of their data.





NZ Research Software Engineers Conference

Help us plan the programme! Email <u>events@nesi.org.nz</u> to get involved.

Spring 2021

Who attends:

- Researchers and academics who code
- Software engineers & system admins working in the research domain
- Generalists who bring together the research and technical domains
- Developers, IT managers, coding enthusiasts, and big data analysts from Crown Research Institutes, universities, and other public sector organisations



NeSI @ eResearch NZ - Talks & Workshops:

Wednesday 10 Feb

13:00 - 17:00 - Maxime Rio - Machine Learning on NeSI 101

13:20 - 13:40 - Jun Huh - Taonga: building a data repository for genomics research in New Zealand

13:20 - 13:40 - Dinindu Senanayake -Paving the way for Bioinformatics excellence in New Zealand

14:20 - 15:00 - Brian Flaherty - Moving data: getting up to speed with Globus and Science DMZ

15:50 - 16:50 - Jana Makar - Challenge Accepted: Responding to community feedback for supporting diversity in HPC & eResearch

Thursday 11 Feb

11:00 - 11:20 - Maxime Rio - Data science consultancies at NeSI: A whirlwind tour of case studies

13:30 - 13:50 - Chris Scott - GPUs on NeSI

13:50 - 14:10 - Georgina Rae - Building Partnerships for eResearch

14:10 - 14:30 - Wolfgang Hayek - NeSl Consultancies - Evolving a Scientific Programming Service

14:40 - 15:00 - Albert Savary - Software on NeSI

15:00 - 15:20 - Jeff Zais - Taking Advantage of Technology Innovations in the Next Generation of NeSI HPC Infrastructure

15:20 - 15:40 - Callum Walley - Virtual Desktops for HPC



Thursday 11 Feb (cont.)

15:20 - 15:40 - Robin Bensley - Staying connected in an evolving eResearch ecosystem

16:00 - 17:00 - Megan Guidry - Sowing the seeds of capability: Experience what Carpentries instructor training is all about

Friday 12 Feb

11:20 - 12:30 - Nick Jones - Future of eResearch

12:20 - 12:30 - José Filipe Gonçalves Higino -Coaching great practices of describing a problem

13:30 - 14:30 - Blair Bethwaite - Embracing cloud-native architectures

13:30 - 14:30 - Alexander Pletzer and Nooriyah Lohani - Who needs research software engineers?

13:30 - 14:30 - Georgina Rae - FAIR for Research Software