Batuhan Yildirim

Email: by256@cam.ac.uk | **Web:** by256.github.io

I am a Research Engineer at InstaDeep in the BioAI team. Previously, I did a PhD in the Molecular Engineering group at the University of Cambridge, where I wrote software and researched machine learning methods for characterization and generation of atomic and nano structures. My thesis was titled *Machine Learning for Structural Characterisation and Generation: Applications to Small-Angle Scattering and Electron Microscopy*.

Education

University of Cambridge	**PhD** Physics and Machine Learning	*Cambridge, UK (Oct 2018 - Oct 2022)*
Queen Mary, University of London	**MSc** Data Science	*London, UK (Sep 2017 - Sep 2018)*
University of Manchester	**MEng** Materials Science	*Manchester, UK (Sep 2013 - Jun 2017)*

Employment Experience

InstaDeep

London, UK (Mar 2023 - Present)

Research Engineer

- Designing and implementing neural networks and machine learning methodologies for molecular modelling and protein design in the BioAI team.

StatusToday

London, UK (Jun 2019 - Oct 2019)

Data Scientist (Intern)

- Created structured datasets from unstructured data and implemented ML models to classify user activity from automated system activity, leading to more accurate insights and features computed by StatusToday.

Competitions

Numer.ai

Remote (May 2021 - Present)

- ML competition where I previously ranked in the top 0.1% participants and currently rank in the top 3% .
- Developing ML time-series models; performing statistical testing at multiple stages of my pipeline, most notably during feature selection and model selection; dealing with non-stationarity and co-dependence between features; and reducing a model's exposure to volatile features, resulting in stable performance over time.

Citadel Datathon

Dublin, IE (Jan 2019)

- 2nd place at the 2019 Citadel Dublin Data Open. $5000.

Honours and Awards

Fitzwilliam College, Senior Scholarship: awarded in recognition of significant research progress made during COVID-19. *(Nov 2020)*

Fitzwilliam College, Senior Scholarship: awarded on the basis of excellent work. *(Nov 2019)*

Rolls-Royce/Tin-Plate Workers Award: awarded for achieving first-class honours and finishing top of my class at the University of Manchester during my third year. *(Aug 2017)*

Open-Source Projects

Core Developer

- rdfpy: a Python module for fast computation of 2D and 3D radial distribution functions. *(Jun 2020 - Present)*
- ImageDataExtractor: a Python framework for electron microscopy image quantification. *(Dec 2019 - Present)*

Contributor

- Ensemble-PyTorch: a unified ensemble framework for PyTorch. I extended the API by implementing the capability to use arbitrary loss functions when training ensemble models (see commit bb7b988). *(Aug 2021)*

Publications

1. **B. Yildirim**, J. Doutch, and J. M. Cole, "Multi-Task Scattering-Model Classification and Parameter Regression of Nanostructures from Small-Angle Scattering Data" Under review, **2023**.
2. **B. Yildirim**, A. Washington, J. Doutch, and J. M. Cole, "Calculating Small-Angle Scattering Intensity Functions from Electron Micrsocopy Images" *RSC Adv.*, vol. 12, pp. 16656–16662, **2022**, 10.1039/D2RA00685E.
3. **B. Yildirim**, J. M. Cole, "Bayesian Particle Instance Segmentation for Electron Microscopy Image Quantification" *J. Chem. Inf. Model.*, vol. 61, no. 3, pp. 1136–1149, **2021**, 10.1021/acs.jcim.0c01455.
4. **B. Yildirim**, C. J. Court, J. M. Cole, "3-D Inorganic Crystal Structure Generation and Property Prediction via Representation Learning" *J. Chem. Inf. Model.*, vol. 60, no. 10, pp. 4518–4535, **2020**, 10.1021/acs.jcim.0c00464.
5. K. T. Mukaddem, E. J. Beard, **B. Yildirim**, J. M. Cole, "ImageDataExtractor: A Tool To Extract and Quantify Data from Microscopy Images" *J. Chem. Inf. Model.*, vol. 60, no. 5, pp. 2492–2509, **2020**, 10.1021/acs.jcim.9b00734.

Fun

Generative Art

I enjoy creating computational art that exploits randomness, motion and colour to produce diverse sets of examples. I have sold many of my pieces and have created and deployed Twitter bots that randomly generate pieces and post several times a day.