

Batu Yildirim, Ph.D.

email: by256@pm.me | web: by256.github.io | phone: +447903689916

Education

University of Cambridge – Ph.D. in Physics and Machine Learning Cambridge, UK 10/2018 – 07/2023
Thesis: *Machine Learning for Structural Characterisation and Generation: Applications to Small-Angle Scattering and Electron Microscopy*. Thesis submitted in 01/2023, graduated in 07/2023.

Queen Mary, University of London – M.Sc. in Data Science (73%) London, UK 09/2017 – 09/2018

University of Manchester – M.Eng. in Materials Science and Engineering (First class) Manchester, UK 09/2013 – 06/2017

Work Experience

Machine Learning Research Engineer II – InstaDeep London, UK 03/2023 – Present

- Conducted research, developed software and employed machine learning to predict SARS-CoV-2 fitness and immune escape. My model predictions were confirmed by experimental validation, enabling me to contribute to the proposal of over fifty vaccine designs that are being trialed preclinically as future vaccine candidates.
- Formulated experiments and employed hypothesis testing methods for analysis of data, experimental outcomes and comparative analysis of machine learning models.
- Constructed data pipelines, including the curation and refinement of data for experiments, analyses and model training/evaluation.
- Actively engaged with collaborators and stakeholders to consistently communicate and present my research methods, results and findings.

Research Scientist (Intern) – STFC Harwell, UK 09/2019 – 09/2020

- Conducted research into the use of machine learning methods for materials characterisation using electron microscopy and small-angle scattering data, resulting in two publications.

Data Scientist (Intern) – StatusToday London, UK 06/2018 – 10/2018

- Created structured datasets from unstructured data and implemented ML models to classify user activity from automated system activity, leading to more accurate insights computed by the company's platform.

Competitions

Numer.ai Remote 05/2021 – Present

- Machine learning quantitative finance competition, where I earned the ranks of master (2nd) in 2022 and expert (17th) in 2023. ([profile](#))
- Developed time-series models on financial data consisting of millions of rows. Performed statistical testing at multiple stages of my pipeline, notably during feature and model selection. Addressed non-stationarity and co-dependence between features and reduced my model's exposure to volatile features.

Citadel Datathon Dublin, IE 01/2019

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

Open-Source Projects

Core Developer

- [rdfpy](#): a Python module for fast computation of 2-D and 3-D radial distribution functions. 06/2020 – 01/2023
- [ImageDataExtractor](#): a Python framework for electron microscopy image quantification. 12/2019 – 01/2023

Contributor

- [Ensemble-PyTorch](#): a unified ensemble framework for PyTorch. I extended the API by implementing the capability to use arbitrary objective functions when training ensemble models. (See commit [bb7b988](#).) 08/2021

Technical Skills

Python (advanced): highly experienced in building applications, writing scripts and data analysis; C++ (basic): reasonable grasp of syntax and capable of writing straightforward scripts and programs; SQL (basic): capable of writing queries with the aid of online resources and setting up databases for web applications.

PyTorch (advanced): ability to build complex model architectures and train models; **NumPy/SciPy** (advanced): numerical computing, encompassing array and matrix operations, optimisation and advanced linear algebra techniques for scientific computing; **Pandas** (advanced): data manipulation, transformation, grouping, aggregation and merging for dataset construction and analysis; **scikit-learn** (advanced): predictive and statistical modelling/analysis; **Django/Flask** (proficient): web development, deployment of machine learning models and data dashboards.

Git (proficient): regular use for version control, collaboration and maintaining projects; **Cloud Platforms (AWS/GCP)** (proficient): capable of creating and working on instances, deploying applications and managing resources; **Distributed computing** (proficient): ability to implement and manage distributed computing processes, including parallel computations, data partitioning, and process synchronisation across multiple nodes and clusters; **Docker** (basic): ability to containerise applications and perform basic container management tasks such as building images and running containers.

Publications

1. B. Yildirim, J. Douth, and J. M. Cole, "Multi-Task Scattering-Model Classification and Parameter Regression of Nanostructures from Small-Angle Scattering Data" *RSC Digit. Discov.*, **2024**, [10.1039/D3DD00225J](#).
2. B. Yildirim, A. Washington, J. Douth, and J. M. Cole, "Calculating Small-Angle Scattering Intensity Functions from Electron Microscopy 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. C. J. Court*, B. Yildirim*, 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](#). *Equal contribution.
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](#).

Honours and Awards

| | |
|--|---------|
| Fitzwilliam College, Senior Scholarship: awarded in recognition of significant research progress made during COVID-19. | 11/2020 |
| Fitzwilliam College, Senior Scholarship: awarded on the basis of excellent work. | 11/2019 |
| 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. | 08/2017 |

Additional Information

Languages: English (native), Turkish (bilingual).