# Arpit Kapoor

PhD Candidate | Data Scientist

Phone: +61 412 982 005 | Email: arpit.kapoor@unsw.edu.au

Website: www.arpit-kapoor.com | GitHub: arpit-kapoor | Google Scholar: Profile

---

## Professional Summary

Data Scientist with **4+ years of experience in Machine Learning, Deep Learning, and AI, developing end-to-end ML pipelines for real-world applications**. Expertise in Bayesian methods, time-series forecasting, anomaly detection, and MLOps. Strong research background with **multiple peer-reviewed publications and experience in data-driven decision-making for business applications**. Seeking a Data Scientist role to apply advanced analytics for impactful solutions.

---

## Education

***Doctor of Philosophy (PhD)***                              Aug 2022 – Expected Feb 2026

University of New South Wales, Sydney

- Thesis title: ***Enhancing process-based hydrological modelling with deep learning***
- Jointly funded by UNSW Sydney & ARC Training Centre in Data Analytics for Resources and Environment (DARE Centre)

***Bachelor of Technology (B.Tech)***                              Jul 2015 – May 2019

SRM Institute of Science and Technology, Chennai

- Major in **Computer Science & Engineering** with **CGPA: 9.01**
- Team Leader: SRM Team Humanoid (the student-led humanoid robotics team)

---

## Professional Experience

**Bureau of Meteorology, Australia**                              (Feb 2023 – Present, Part-Time)

Research Support Scientist

- Developed multivariate bias correction methods for climate modeling.
- Built a Python wrapper for the **MRNBC bias correction method** written in FORTRAN, *improving scalability and efficiency* for the Australian Climate Service (ACS).

**Quince, Hyderabad, India**                             (Mar 2022 – Aug 2022, Full-Time)
Data Scientist - II

- Developed customer retention models, increasing repeat customer engagement by 15%.
- Built regression models for logistics cost optimization, reducing costs by 20%.

**3Qi Labs, Hyderabad, India**                           (Nov 2019 – Nov 2021, Full-Time)
Data Scientist

- Designed an **LSTM-based anomaly detection system**, improving accuracy by **40%**.
- Automated ML workflows for **anomaly detection in Extract Transform & Load (ETL)** data pipelines.
- Optimised machine learning based **anomaly detection model into Data Quality Monitoring Software** developed by the organisation
- Integrated **MLOps practices** to optimize the ML lifecycle.

**Bomotix, Hyderabad, India**                            (Jan 2019 – Nov 2019, Full-Time)
Machine Learning Engineer

- Developed Deep Learning models for computer vision based analysis of sports videos
- This includes **player detection and tracking, player pose estimation** (for form analysis) in both 2D and 3D using various **deep learning models (trained in-house)**
- Cleaned, processed, and curated video data for model training and validation.
- Optimised model training on GPU clusters using **mixed-precision for efficiency**.
- **Maintained CI/CD pipelines for deep learning** model development and deployment.

**The University of Sydney, Australia**                  (Jun 2018 – Jan 2019, Part-time)
Research Intern (Machine Learning)

- Contributed to research and development of Bayesian methods for neural networks and geoscientific models.
- Implemented **Parallel Markov Chain Monte Carlo (MCMC)** methods for Bayesian Neural Networks.
- Developed a transfer-learning approach for Bayesian Neural Networks based on a parallel MCMC scheme (published in Elsevier's Neurocomputing Journal)

---

## Technical Skills

**Programming Languages**: Python, C/C++, R, SQL

**Machine Learning & AI:**

- Supervised & Unsupervised Learning: Random Forest, Gradient Boosting, SVM
- Deep Learning: CNN, LSTM, Autoencoders, Seq2Seq models, Bayesian Deep Learning
- Reinforcement Learning: Deep Q-Learning, Policy Gradient Methods
- Time-Series Forecasting & Anomaly Detection

**Frameworks & Tools:**

- Deep Learning: TensorFlow, PyTorch, Flax
- ML & Numerical Computing: Jax, Scikit-learn, Spark MLlib
- MLOps: MLflow, Docker, Kubernetes
- Model Interpretability: SHAP, LIME
- Big Data Processing: Dask, Apache Spark

**Cloud & DevOps:** AWS (EC2, S3, Lambda), Azure, High-Performance Computing (HPC), Bash Scripting

---

## Leadership & Achievements

- Selected to participate in the Information Resilience PhD School (2024) at the University of Queensland, Brisbane
- Facilitated Theyr Challenge at the Data Study Group (May 2024, Alan Turing Institute, London).
- PhD scholarship recipient (Australian Research Council, DARE Centre) and UIPA grant at UNSW
- Finalist in IEEE/RSJ IROS 2017 Humanoid Application Challenge, Vancouver, Canada.
- Robot Programmer and Team Leader (2018-2019) at SRM Team Humanoid (2015-2019): Won Gold, 2 Silver, and 1 Bronze in RoboGames'17, USA
- Recipient of University Excellence Scholarship (Top 1% of students, SRM University)

---

## Key Publications

- Kapoor, A., Pathiraja, S., Marshall, L., & Chandra, R. (2023). DeepGR4J: A deep learning hybridization approach for conceptual rainfall-runoff modeling. *Environmental Modelling & Software, 169, 105831*.
- Kapoor, A., Negi, A., Marshall, L., & Chandra, R. (2023). Cyclone trajectory and intensity prediction using Bayesian Variational RNNs. *Environmental Modelling & Software, 162, 105654.*
- Kapoor, A., Nukala, E., & Chandra, R. (2022). Bayesian neuroevolution using distributed swarm optimization and tempered MCMC. *Applied Soft Computing, 129, 109528.*
- Chandra, R. and Kapoor, A., (2020). Bayesian neural multi-source transfer learning. *Neurocomputing*, *378*, pp.54-64.

---