

Donia Beshar

donia.a.beshar@gmail.com | [Data Science Portfolio](#) | [Google Scholar](#) | [GitHub](#) | [+971506102788](#) | Abu Dhabi, UAE

Education

M.S./Ph.D. in Statistics (Data Science Track) **08/2026 - Present**
King Abdullah University of Science and Technology (KAUST) Thuwal, Saudi Arabia

Supervisor: [Prof. Raphaël Huser](#)

Research Focus: Modeling complex spatiotemporal processes and predicting extreme events using deep learning architectures.

Selected Coursework: Statistics of Extremes Bayesian Statistics Machine Learning Spatial Data Science with R

Bachelor's Degree in Mathematics **09/2022 - 05/2025**
[Sorbonne Université](#) Paris, France

Grade: Mention Très Bien (Highest Honors, GPA 4.0/4.0 equivalent)

Specialization Modules: Data Science & Machine Learning Multivariate Data Analysis Statistical Inference Probability

Research Experience

Research Assistant **05/2024 - 08/2026**
Sorbonne University, Center for Applied Mathematics, Statistics and Data Science Abu Dhabi, UAE

Supervisor: [Prof. Tanujit Chakraborty](#)

Deep Generative Transformers for Probabilistic Time Series & Spatiotemporal Forecasting **Code:** [Genformer](#)

- Developed the Genformer Python package, a modular PyTorch framework for lightweight deep generative Transformer-based probabilistic forecasting, supporting training, inference, and uncertainty-aware prediction.
- Developed the Enformer and Graph-Enformer (GENformer) models, integrating stochastic noise injection, Transformer architectures, graph convolution, and Energy Score optimization to learn conditional predictive distributions.

Nonlinear Time Series Framework for Financial Forecasting in Emerging Economies **Code:** [NARFIMA](#)

- Designed and implemented the NARFIMA (Neural AutoRegressive Fractionally Integrated Moving Average) model to forecast exchange rates of BRIC (Brazil, Russia, India, and China) economies, incorporating macroeconomic indicators.
- Performed large-scale computational experiments, including:
 - Causal inference analyses to assess the impact of external factors on currency markets.
 - Conducting ablation studies to justify the integration of ARFIMA residuals.
 - Hyperparameter optimization and evaluation of 16 baseline models using robustness and statistical significance tests.
 - Uncertainty quantification through conformal and simulation-based prediction intervals.

Interpretable Probabilistic Forecasting and Driver Analysis of Climate Policy Uncertainty **Code:** [CPU](#)

- Developed the first framework for forecasting Climate Policy Uncertainty (CPU), with explicit emphasis on policy-relevant interpretability and uncertainty quantification.
- Identified key macro-financial and behavioural drivers of CPU through causal analyses.
- Integrated Google Search indicators to capture early shifts in public attention and enhance forecast accuracy.
- Demonstrated the role of CPU forecasting in supporting SDG 13 (Climate Action) and SDG 8 (Decent Work and Economic Growth) by informing climate strategies that balance emissions reduction with economic resilience.

Publications

- Beshar, D.**, Panja, M., Sengupta, S., Chakraborty, T., (2026+) Neural ARFIMA for Forecasting BRIC Exchange Rates with Long Memory. *Submitted to **Studies in Nonlinear Dynamics & Econometrics***. Preprint available at [arXiv:2509.06697](#).
- Beshar, D.**, Sengupt, A., Chakraborty, T. (2026+) Probabilistic Forecasting of Climate Policy Uncertainty: The Role of Macro-financial Variables and Google Search Data. *In Revision to **Journal of Environmental Management***. Preprint available at [arXiv:2507.12276](#).

Research Software & Open Source Packages

- [Genformer](#) - Developer** (1400+ downloads): PyTorch package implementing lightweight Transformer architectures (Enformer and GENformer) for probabilistic multivariate and spatiotemporal forecasting using energy score optimization.
- [narfima](#) - Developer & Maintainer** (2000+ downloads): CRAN R package for nonlinear long-memory time series forecasting with support for exogenous variables, preprocessing, uncertainty quantification, and benchmarking.

Projects

[Will They Get the Job? Ensemble Learning and Skew Kernels for Imbalanced Data](#) **04/2025 - 05/2025**

Project: Developed the “Placement Puzzle: Crack the Hiring Code” [Kaggle-winning machine learning model](#) to classify job placement outcomes from a limited, imbalanced dataset with nonlinear separability, multicollinearity, and missing values.

- Tackled severe class imbalance in a limited dataset by applying GHOST (Generalized tHreshOld ShifTing) and SMOTE.
- Engineered features to improve separability, selected features to reduce multicollinearity, and preprocessed for missing values.
- Benchmarked XGBoost, CatBoost, kernelized SVM, MLP, Random Forest, Decision Trees, Logistic Regression, and KNN, ensembled models using bagging or boosting where appropriate, and optimized hyperparameters via Bayesian optimization.
- Achieved superior predictive performance using ensemble learning and a custom kernel for asymmetric data distribution.

Credit Card Debt Prediction: Ensemble Regression Modeling

03/2024 - 04/2024

Project: Developed the “Credit Card Debt Prediction” [Kaggle-winning regression model](#) for predicting credit card debt.

- Conducted a thorough analysis of causal relationships and identified key socio-economic predictors for the model.
- Addressed multicollinearity, nonlinearity, outliers, and nonnormal residuals by implementing L_1 and L_2 regularization, LOESS (Locally Estimated Scatterplot Smoothing), Huber loss, and Box-Cox transformation, respectively.
- Developed an ensemble model combining insights from ten distinct frequentist and Bayesian models.

Analysis of COVID-19 Mortality Data: A Probabilistic Perspective

03/2023 - 05/2023

Project: Conducted a probabilistic analysis of COVID-19 mortality data to identify the underlying distribution and provide insights for pandemic management strategies.

- Investigated COVID-19 mortality distribution by fitting parametric models, including Marshall-Olkin extended distributions and Generalized Exponential distributions, and evaluated the goodness of fit of the distributions.
- Applied Maximum Likelihood Estimation (MLE) to estimate distribution parameters, utilizing optimization techniques when MLE lacked a closed-form solution.

Outreach and Engagement

Women in STEM organized by CCI France UAE

11/2024

- Delivered a talk on the role of time series forecasting and AI in advancing Sustainable Development Goals (SDGs), focusing on monetary policy for SDG 8 (Decent Work and Economic Growth) and climate challenges for SDG 13 (Climate Action).

Peer Tutor, Sorbonne University

09/2024 - 06/2025

- Empowered L2 students to master data science and statistics by fostering a collaborative learning environment and providing tailored one-on-one sessions, leading to significant academic improvement and increased confidence.

Future of Energy Hackathon hosted by TotalEnergies, Institut Français, and La Fresque du Climat

06/2022

- Investigated the causes and implications of climate change, designed sustainable development strategies for École 42, contributed to innovative solutions such as a self-generating energy keyboard, and presented proposals to a panel of experts.

Science Communicator: “The Simplicity of Complexity”, Luxembourg Pavilion, Expo 2020

12/2021

- Facilitated interactive board games based on open mathematical problems designed by the University of Luxembourg and addressed inquiries from the public regarding mathematical concepts.

Key Skills

Programming Skills:

- Python (pandas, NumPy, scikit-learn, TensorFlow, PyTorch, DARTS, imbalanced-learn, statsmodels, matplotlib, seaborn)
- R (devtools, dplyr, tidyverse, lubridate, ggplot2, forecast, caret, glmnet)
- LaTeX
- Microsoft Office

Data Science Skills:

- Probabilistic Modeling
- Statistical Modeling
- Data Analysis
- Data Visualization
- Causal Inference
- Quantifying Uncertainty
- Ensemble Methods
- Forecasting
- Classification
- Regression

Soft Skills:

- Time Management
- Communication
- Adaptability
- Problem-solving
- Team Player
- Attention to Detail
- Critical Thinking
- Work Ethic
- Independence
- Creativity

Languages: • English: Fluent

• French: Elementary

• Arabic: Native

Honors, Awards, and Recognition

- 75% merit-based scholarship by Sorbonne University (2025)
- Certificate of Excellence for ranking 1st in Kaggle competitions (2024, 2025)
- 50% merit-based scholarship by Sorbonne University (2023, 2024)
- Certificate of Achievement for participating in the 2022 Youth Hackathon on Climate Change (2022)
- Scholarship for outstanding academic potential by Sorbonne University corporate partner, The Ardian Foundation (2022)
- Certificate of Appreciation for participating in the organization of the event “The Simplicity of Complexity” (2021)