Arnob Mukherjee

Uppsala, Sweden; Eligible to work in Sweden
arnob.mukherjee@physics.uu.se | +46 72 225 7142
linkedin.com/in/arnobmukherjee1988
github.com/arnobmukherjee1988

Professional Summary

Data Scientist with 5+ years of experience building and deploying production-grade forecasting systems, risk assessment models, and analytical solutions for data-driven decision support. Proven expertise in time-series forecasting, credit risk modelling, anomaly detection, and A/B testing frameworks. Skilled in Python, SQL, machine learning (Logistic Regression, Random Forest, XGBoost, CNN), and currently learning the MLOps practices. Experienced in ETL pipeline development, REST API integration, and learning cloud infrastructure (AWS, GCP). Strong collaborator with international teams, translating complex analytical requirements into scalable deliverables. Holds a PhD in Computational Physics with deep expertise in statistical modelling, Monte Carlo methods, and numerical optimization.

Technical Skills

Programming:	Python (Pandas, NumPy, SciPy, Scikit-learn, Statsmodels, TensorFlow, Keras), SQL (basic), Shell scripting
Time-Series:	EDA, ARIMA, SARIMA, Prophet, seasonal decomposition, stationarity testing (ADF), backtesting
Machine Learning:	Logistic Regression, Random Forest, XGBoost, Neural Networks (CNN), model evaluation (AUC-ROC, precision-recall), hyperparameter tuning, feature engineering
Statistical Methods:	Hypothesis testing, A/B testing, experimental design, Bayesian inference, Monte Carlo methods, MCMC, regression analysis, power analysis
MLOps & Infrastructure:	GitHub, model deployment, performance monitoring, AWS (EC2, Lambda)
Data Engineering:	ETL pipeline design, REST API integration, data quality validation, workflow automation, batch processing, HPC systems (SLURM), Cron scheduling
Visualization & Tools:	Matplotlib, Seaborn, Plotly, Tableau (learning), data presentation, interactive dashboards

Professional Experience

Data Scientist & Machine Learning Researcher ***November 2023 – November 2025***

Uppsala University, Sweden

- Developed production ETL pipeline integrating REST APIs for CO_2 emission forecasting with automated validation, transformation, and scheduled ingestion workflows, enabling time-series forecasting (Prophet, ARIMA, SARIMA) with **28% error reduction** through iterative model improvements **[GitHub]**
- Built comprehensive credit risk classification system processing **887K loan records** with end-to-end pipeline: data validation, exploratory analysis (distribution analysis, correlation studies, outlier detection), feature engineering, and model deployment (Logistic Regression, Random Forest, XGBoost), achieving **94% AUC-ROC** and demonstrating expertise in financial services domain **[GitHub]**
- Designed Monte Carlo stress-testing framework (**10K simulations**) utilizing Bayesian uncertainty quantification (MCMC methods) to project default rate scenarios, directly informing capital adequacy planning and risk-based pricing automation for stakeholder discussions
- Led cross-functional collaboration on end-to-end data pipeline processing **100K+ images**: Monte Carlo generation $\rightarrow$ transformation $\rightarrow$ parallel processing $\rightarrow$ CNN model deployment (TensorFlow, Keras), achieving **97% accuracy** for anomaly detection in large-scale datasets **[GitHub]**
- Designed and executed A/B testing framework for e-commerce free shipping strategy, conducting statistical analysis (hypothesis testing, segmentation), uncovering Simpson's Paradox, and delivering data-driven recommendations generating **321% ROI (+109K BRL projected profit)** **[GitHub]**
- Collaborated with cross-functional stakeholders to translate business questions into analytical approaches, define metrics, and present findings through clear visualizations and structured narratives, enabling informed decision-making

Data Engineer & Research Software Engineer ***April 2022 – July 2023***

University of Tennessee, Knoxville, USA

- Developed modular Python components for large-scale Monte Carlo simulation pipelines processing terabytes of data, establishing reusable computational workflows with automated convergence checks and statistical validation [**GitHub**]
- Architected scalable data processing pipelines using Python and Shell scripting, improving computational efficiency by **40%** through algorithmic optimization (vectorization, caching) and data synchronization enhancements
- Implemented batch processing automation with Cron-based orchestration and SLURM on HPC systems, enabling reliable large-scale job scheduling with reduced manual oversight
- Developed and deployed open-source Python CLI tool (GPTPlot) to PyPI for automated exploratory data analysis (EDA) and scientific visualization supporting multiple file formats (CSV, DAT, TXT, etc.), demonstrating end-to-end product development capability. Currently, integrating off-the-shelf LLM for AI-assistance **[PyPI**, **GitHub]**
- Worked with international research teams (USA, Canada, China), acting as coordination point by translating analytical requirements into scalable technical deliverables and facilitating cross-team communication through regular meetings

Doctoral Researcher – Computational Physics ***August 2014 – November 2021***

IISER Mohali, India

- Developed custom gradient descent optimization algorithms in high-dimensional parameter spaces, demonstrating understanding of the principles underlying neural network training and convergence behavior
- Applied Fourier Transform techniques for signal processing and spectral analysis, directly relevant to feature engineering for time-series anomaly detection
- Conducted extensive Monte Carlo simulations and statistical sampling methods for complex system modeling, building foundation for Bayesian inference and uncertainty quantification
- Published peer-reviewed research in computational physics and delivered graduate-level courses in statistical mechanics and numerical methods

Education

- **PhD in Computational Physics**, IISER Mohali, India — August 2014 – November 2021
- **MSc in Applied Physics**, IIT Dhanbad, India — August 2010 – July 2012

Certifications

- **Physics and Finance**, Uppsala University — 2024
- **Machine Learning Specialization**, DeepLearning.AI / Stanford University — 2023
- **Machine Learning A-Z: Hands-On Python & R in Data Science**, Udemy — 2022

Additional Information

Languages: English (Fluent), Swedish (Beginner, A1-A2 level), Hindi & Bengali (Native)
Work Authorization: Fully eligible to work in Sweden; available for roles in Uppsala, Stockholm, and surrounding areas
Availability: Immediate start possible; open to hybrid and on-site arrangements
Interests: Chess, Badminton, Cooking, Table Tennis