

Sabuj C. Bhowmick

Data Scientist | Machine Learning Researcher | Computational Biology

 sabujcb  sabuj-bhowmick-34920874

Profile

MSc Data Analytics student at the University of Turku with an interdisciplinary background in statistics, mathematical sciences, public health, IoT engineering, machine learning, and computational research. Focused on reproducible and interpretable data science, rigorous model validation, bioinformatics, multi-omics integration, and research-oriented software development.

Education

MSc	University of Turku , Information and Communication Technology - Data Analytics - Graduate studies in machine learning, data analytics, statistical modeling, and data-intensive systems. - Focus on reproducible analysis, model validation, and research-oriented computing.	2025 – present
BEng	Savonia University of Applied Sciences , Internet of Things Engineering studies in connected systems, data acquisition, software development, and applied analytics.	2024
MSc	Norwegian University of Science and Technology , Mathematical Sciences Graduate training in mathematical reasoning, quantitative modeling, and scientific problem solving.	2019
MPhil	University of Oslo , International Community Health Research training in public health, epidemiological reasoning, and quantitative health research.	2015
MPS	University of Dhaka , Population Sciences	2013
MSc	University of Dhaka , Statistics Advanced training in probability, statistical inference, regression, and applied data analysis.	2011
BSc	University of Dhaka , Statistics	2010

Experience

University of Turku - Department of Computing , Project Researcher Conducting data analytics for the NoveLyme project, supporting computational research on Lyme disease-related data within the Department of Computing. - Prepare, analyze, and interpret project data in support of NoveLyme research objectives. - Develop transparent analytical outputs and communicate findings within the project research team.	Sept 2026 – present
University of Turku - Department of Computing , Research Assistant Contributed to the development and validation of Joint-RPCA workflows for microbiome multi-omics data in the R/Bioconductor mia ecosystem. - Prepared aligned metagenomic and metatranscriptomic inputs, shared-sample sets, deterministic train/test splits, and reproducible filtering workflows. - Compared R/Bioconductor mia and Python/Gemelli outputs across sample scores, feature loadings, explained variation, cross-validation results, and distance structure. - Improved research software through metadata-access refactoring, runtime messaging, tests, documentation, and reproducible Quarto reports with software provenance. - Collaborated through GitHub with international professor-led research teams.	May 2026 – Aug 2026
Savonia University of Applied Sciences , Project Worker - 5G to Boost Digitization Developed applied data-analysis and forecasting workflows within a 5G- and IoT-oriented digitalization project. - Built solar-energy production analysis and forecasting workflows. - Compared statistical and machine learning models, including ARIMA, random forest, SVR, and XGBoost. - Integrated weather data and prepared reproducible technical reporting.	June 2023 – Aug 2023

Voimatel Oy / Savonia University of Applied Sciences, Intern - Starlink Feasibility Testing
Evaluated satellite connectivity for remote operational scenarios and compared it with mobile-network alternatives.

Sept 2023 – Jan 2024

- Measured latency, bandwidth, signal reliability, and field performance.
- Documented equipment setup, test conditions, results, and practical limitations.

University of Florida-associated research collaboration, Senior Reviewer - Interdisciplinary
Cardiac Arrest Review

2018 – 2022

- Contributed to structured evidence assessment and interdisciplinary scientific review.
- Applied methodological scrutiny to research interpretation and reporting.

Skills

Programming and research software: Python, R, SQL, Bash, Git, GitHub, Quarto, R Markdown, LaTeX

Machine learning: scikit-learn, PyTorch, XGBoost, random forest, SVM, regression, classification, cross-validation

Statistics: applied statistics, statistical inference, time-series analysis, biostatistics, model evaluation

Bioinformatics: Bioconductor, microbiome analysis, compositional data, multi-omics integration, Joint-RPCA

Engineering practices: reproducible pipelines, automated testing, data validation, version control, technical documentation

Projects

Joint-RPCA for Multi-omics Integration

Cross-language validation of R/Bioconductor and Python/Gemelli workflows for microbiome multi-omics data.

- Developed and tested consistent input filtering for multi-assay biological count data.
- Compared sample scores, feature loadings, explained variation, and cross-validation outputs.
- Produced reproducible Quarto workflows with fixed configurations and interoperable outputs.

Solar Energy Forecasting

Comparative forecasting with statistical and machine learning models.

- Integrated photovoltaic production and weather data.
- Compared ARIMA/ETS, random forest, SVR, and XGBoost using chronological validation.

Protein-Drug Affinity Prediction

Group-aware validation study for affinity prediction.

- Compared leave-one-observation, leave-one-protein, and leave-one-drug validation.
- Identified a substantial loss of generalization for previously unseen proteins.

Languages

Bangla: Native

English: Professional working proficiency

Finnish: Basic

Interests

Research interests: computational biology, bioinformatics, applied AI, biostatistics, scientific computing, reproducibility