Sambadi Majumder, Ph.D.

Data Scientist

South Royalton, Vermont 05068, USA| +1 321-202-1464 |sambadimajumder@gmail.com |
LinkedIn: linkedin.com/in/sam-majumder-25720870 **| Portfolio Website:**
https://sambadimajumder.quarto.pub/spatialdatachronicles/**| GitHub**: github.com/SamMajumder

Core Competencies

- **Programming languages:** R and Python
- **Version Control Software:** Git
- **Machine Learning algorithms:** Random Forest, Gradient Boosting Machines, XG Boost, Convolutional Neural Network, KNN, DBSCAN
- **GIS Software and libraries:** ArcGIS, ArcMap, QGIS, GDAL
- **Writing and Documentation:** Rmarkdown, Quarto
- **Dashboard frameworks:** R Shiny, Streamlit, flexdashboard
- **Cloud platforms:** Databricks, AWS, Domino Data Lab, Google Colab

Professional Experience

Data Scientist | Global Water Security Center | Remote   *2023-Present*

Designed data engineering and geospatial machine learning workflows addressing issues pertaining to environmental security

- Engineered robust and reproducible automated Python-based data pipelines for efficient data collection and processing leading to improvements in data quality and reliability

- Optimized multidimensional data storage and retrieval processes to enhance system performance and scalability leading to efficient data analysis

- Collaborated with interdisciplinary teams to integrate data-driven solutions into water security research

Data Scientist | University of Central Florida | Orlando, FL   *2019-2023*

Designed geospatial machine learning workflows addressing U.S. food security under climate change for the next 60 years as it pertains to oil seed sunflower.

- Streamlined spatial and temporal data wrangling by designing reproducible workflows using R and ArcGIS to extract, compile, aggregate, and clean large multidimensional datasets from publicly available databases including USDA and NOAA

- Discovered 10 most influential weather factors impacting production in 177 counties using machine learning algorithms by incorporating target oriented cross-validation within feature selection

- Forecasted a 20% overall national production decline over the next 60 years and identified regions most vulnerable to the impacts of climate change

- Built an open-source R Shiny web application called "SunScope" with interactive visualizations delivering actionable insights to hundreds of farmers, policymakers, and scientists nationwide

Data Scientist | Bayer Crop Science | Remote   *2023*

Spearheaded a comprehensive analysis of national corn cultivation patterns in the U.S. to provide evidence-based recommendations for improving corn yield.

- Executed SQL queries to extract and summarize data from 10+ Google BigQuery datasets, enabling efficient analysis of large agricultural and multidimensional weather datasets from 4000 sites

- Deployed R Shiny dashboards with interactive visualizations to illustrate the economic importance of key corn traits across 20 states to plant breeders and agronomists for targeted breeding strategies
- Leveraged interpretable machine learning frameworks to identify influential climate patterns and agricultural practices relative to corn yield across 100 counties

NASA DEVELOP Geospatial Data Scientist | Space Systems and Innovations Lab Ltd | Remote   *2022*

Applied geospatial machine learning to create actionable data-informed conservation practices in the Grand Valley region of Colorado.

- Formulated a combined ArcGIS and R programming data wrangling workflow to process and clean remote sensing data (Landsat, MODIS, SRTM) and spatiotemporal weather datasets
- Successfully identified influential factors driving vegetation dynamics and forecasted the anticipated shifts in greenery over the next two decades

Data Scientist | Becton Dickinson and Company | Raleigh, NC   *2022*

Implemented an interpretable machine learning framework to uncover hidden patterns in bacterial genomes related to infection.

- Devised a reproducible data wrangling/cleaning pipeline to extract, aggregate, and preprocess complex Next-Generation Sequencing data from 2000 bacterial samples
- Developed a novel machine learning-oriented method to successfully identify the most relevant genomic characteristics and improved the overall accuracy of prediction by 27%
- Constructed R Shiny dashboards with interactive visualizations to efficiently showcase meaningful insights to healthcare researchers

Education

University of Central Florida – Orlando, FL, USA

PhD in Integrative and Conservation Biology | 2019-2023

- Dissertation title: "Applications of interpretable machine learning methods in plant ecology and crop science"

University of Giessen – Giessen, Hesse, Germany

MSc in Agro-Biotechnology | 2015-2018

- Thesis title: "Investigating the relationships between vegetative traits and seed yield amongst old and modern oilseed rape genotypes"

West Bengal University of Technology – Kolkata, West Bengal, India

BTech in Biotechnology | 2010-2014

Peer-reviewed Publications

- Bahmani K, A Robinson, **S Majumder**, A LaVardera, JA Dowell, EW Goolsby, CM Mason. Broad diversity in monoterpene-sesquiterpene balance across wild sunflowers: implications of leaf and floral volatiles for biotic interactions. Am J Bot. 2022 Nov 1. doi: 10.1002/ajb2.16093. Epub ahead of print. PMID: 36317693.

Publications in review

- **Majumder S**, CM Mason. A Machine Learning approach to study plant functional trait divergence. bioRxiv. 2023 Mar. doi: https://doi.org/10.1101/2023.03.16.533012
- **Majumder S,** CM Mason. Applying an interpretable machine learning approach to assess intraspecific trait variation under landscape-scale population differentiation. bioRxiv. 2023 Apr. doi: https://doi.org/10.1101/2023.04.07.536012

Publications in preparation

- **Majumder S,** CM Mason. Interpretable Machine Learning for modeling weather impacts on oilseed sunflower yields: Unravelling historical patterns and forecasting under climate change.

Selected Conference Presentations

Conference of Florida Graduate Schools

- 2023, **Oral Presentation:** AI Harvest: Forecasting Sunflower Crop Yields in a Changing Climate

Botanical Society of America Annual Conference

- 2022, **Oral Presentation:** A Machine Learning approach to gain insight into crop sunflower cultivation under climate change and forecast future trends in the United States.
- 2022, **Oral presentation:** Using a Machine Learning approach to study functional trait diversification in Helianthus.
- 2021, **Oral presentation:** Machine Learning approaches to investigating trait divergence under species diversification.

NASA Science Applications Week

- 2022, **Oral Presentation**: Feasibility Study Snapshot: Grand Valley Ecological Forecasting II

UCF Research Week

- 2020, **Poster Presentation:** Machine Learning frameworks for analyzing codon differences in plant metabolite genes

Awards and Recognitions

- Awarded the Graduate Presentation Fellowship 2023 in recognition of representing UCF at the Statewide 3MT Thesis competition held at the University of Miami
- Awarded 1st Place at the 5th Annual UCF College of Graduate Studies 3MT (Three Minute Thesis) competition for presenting a research talk on the applications of Artificial Intelligence in the improvement of Sunflower cultivation. (Link:https://www.youtube.com/watch?v=YEMwz9wW7qc&list=PLG5Mcg1rBGB1Ub XAiMCWw1_OZw4-xKt25&index=7)
- Awarded the Spring 2022 Graduate Research/Travel award for research pertaining to the improvement of sunflower cultivation through geospatial data science.
- Featured in the UCF College of Graduate Studies Annual Report 2021 for research pertaining to machine learning in plant science. (Link: https://graduate.ucf.edu/profile/sam-majumder)
- Featured in UCF TODAY "Research in 60 Seconds" blog and video series. Blog (Link: https://www.ucf.edu/news/research-in-60-seconds-using-algorithms-to-study-sunflowers/) Video series (https://www.youtube.com/watch?v=259o1ADpEeI)

Volunteer Activities

- Represented graduate and international students as a member of the 2021-22 President's Student Advisory Council at the University of Central Florida (Link: https://www.ucf.edu/leadership/president/psac/advocates/)
- Mentored undergraduate students teaching them methods of scientific research, specifically experimental design and analysis of data.
- Served as a committee member for an undergraduate thesis
- Volunteered at Citizen Science GIS as a GIS Analyst to process drone imagery for flood modeling in Alaska
- Volunteered through the Skype A Scientist nonprofit and spoke with middle and high school-aged students answering questions pertaining to science, careers in science, and college applications