Ebrahim Jahanshiri
Data Science and Analytics
ej@cropbase.co.uk

Summary

Ebrahim Jahanshiri is a highly experienced practitioner in:

- Predictive Statistical Modeling, Machine Learning, Deep Learning, Large Lanuguage Models (LLMs) fine-tuning and RAG solutions, inferencing and application development, MLOps, and Git
- Data Management, Data Lakes and Data Warehousing, Cloud Data Management, DVC, and Cloud Computation and storage (GCP, AWS)
- Geographic Information Systems (GIS), Remote Sensing, Statistics, Machine Learning, AI for Geographic Data, Mapping, and GIS Software Development

Career highlights:

- 15 years of experience in geospatial analysis, data science, predictive analytics, and building data and AI systems
- Pioneered the use of new big data analysis tools within the agricultural industry, enabling more efficient data processing and uncovering valuable hidden patterns
- Improved revenue by securing European projects to develop software systems and databases
- Effectively communicated complex data insights to non-technical stakeholders, influencing key business decisions and driving positive outcomes

Geospatial data scientist

Ebrahim's particular expertise lies in the design and development of analytics for geospatial data. He has experience in handling multi-source geospatial data and delivering projects that utilize data from GPS and satellite remote sensing, as well as software platforms leveraging geospatial programming (ArcGIS, ArcObjects SDK for VBA development, QGIS, R). He has also worked extensively with modern geospatial data handling systems on the cloud, such as Google Earth Engine.

Data Science and Analytics Engineer

Ebrahim's experience encompasses managing data science projects, algorithm development, coding in Python and R, AI modeling, and dashboarding and analytics for a wide range of industries. He has developed a variety of machine learning and deep learning models (Clustering, Random Forest, Deep Learning, Generative Neural Networks, and LLMs for knowledge discovery). He also has experience with end-to-end model implementation, CI/CD, MLOps, and Git on Google Cloud and AWS. He has extenstive experience with NLP solutions, LLM fine-tuning and RAG (Retrieval Augmented Generation) solutions.

Data Specialist

As a Data Manager, Ebrahim's experience focuses on developing data models, database implementation, and delivery. He has operationalized data warehouses and data lakes for data acquisition, mining, curation, and front-end development for data portals. He has worked with SQL (Postgres, MySQL), GCP (Cloud APIs, SQL Maestro), is familiar with Databricks lakehouses, and is well-versed in Data Version Control (DVC).

As a data and AI expert, Ebrahim thrived in diverse teams tasked with creating innovative AI-driven systems, all while adhering to agile principles of continuous collaboration and improvement. He has international experience working with, managing, and training teams with diverse backgrounds, coordinating product development efforts and funding. His commitment to continuous learning and the application of new skills fuels his passion for data, especially in areas with the potential to impact profitability and sustainability.

Technical Competencies

Data Science Expertise:

- ***Programming languages***: Python, R, SQL
- ***Statistical analysis and modeling***: Hypothesis testing, regression analysis, time series analysis, dimensionality reduction, etc.
- ***Machine learning and deep learning algorithms***: Supervised learning (classification, regression), unsupervised learning (clustering, dimensionality reduction), and familiarity with advanced techniques like deep learning and ensemble methods. Large-scale model development for deep learning (CNNs, Generative Adversarial Neural Networks, LLM fine-tuning (using Llama suit of models)
- ***Experience with various data science libraries and frameworks***: Pandas, NumPy, scikit-learn, TensorFlow, PyTorch, Spark MLlib, etc.
- ***Big data technologies***: Hadoop, Spark, distributed computing frameworks.
- ***Conducted data science experiments***: Formulating hypotheses, collecting and cleaning data, building and evaluating models, interpreting results.
- ***Data Science Software development:*** Developed Dashboards using, R and Python, implemented DevOps and MLOps using Git and GitLab.

Data Management and Engineering:

- ***Data pipelines and architectures***: Desinged and built data pipelines for data ingestion, transformation, and storage.
- ***Cloud platforms***: AWS, Google Cloud Platform for data storage, processing, and deployment.
- ***Databases and data warehousing***: SQL databases (relational and NoSQL), data warehouses like Redshift, Snowflake.
- ***Data security and privacy***: Basic data anonymization, encryption, and access control mechanisms.

Communication and Collaboration:

- ***Effective communication***: Ability to explain complex technical concepts to both technical and non-technical audiences.
- ***Leadership qualities***: Motivated and mentored team members while fostering collaboration. Built a positive team culture.
- ***Project management***: Defining project goals, setting timelines, managing resources, tracking progress.
- ***Presentation***: Effectively communicated data insights and recommendations to stakeholders.

Additional Skills:

- ***Business acumen***: Basic understanding the business context and objectives of data science projects.
- ***Problem-solving and critical thinking skills***: Identify and solve complex data-related problems.
- ***Staying updated with the latest advancements in data science***: Actively and continuous learning and exploration of new tools and techniques.
- ***Openness to feedback and collaboration***: Willingness to learn from others and adapt to new approaches.

Technologies

Programming Languages:

- Python (libraries like Pandas, NumPy, Matplotlib, Seaborn), R (libraries like dplyr, ggplot2), SQL with basic knowledge of Scala, Julia, Java

Statistical Analysis & Modeling:

- Libraries: scikit-learn, statsmodels, TensorFlow, PyTorch, XGBoost, LightGBM
- Techniques: Regression, classification, clustering, dimensionality reduction, time series analysis, Bayesian statistics

Machine Learning:

- Frameworks: TensorFlow, PyTorch, scikit-learn, Spark MLlib (for big data)
- Deep Learning Libraries: TensorFlow, PyTorch, Keras, MXNet
- Fine-tuning: Transformers, PEFT
- LLM APIs: OpenAI, Gemini, Calude, Llama

- Algorithms: Supervised learning (classification, regression), unsupervised learning (clustering, dimensionality reduction), reinforcement learning

Data Management & Engineering:

- Databases: SQL databases (MySQL, PostgreSQL), NoSQL databases (MongoDB), Data Lakehouse (DataBricks)
- Data Integration & Pipelines: Airflow, Luigi, Prefect, Apache Spark
- Cloud Platforms: AWS, Google Cloud Platform (for data storage, processing, deployment)
- Data Visualization & Communication:
    - Libraries: Matplotlib, Seaborn, Plotly, Bokeh, Tableau, Power BI
    - Interactive Dashboards: Shiny (R), Dash (Python), Flask (Python)

Big Data Analytics:

- Frameworks: Apache Spark, Hadoop, Hive
- Distributed Computing: Spark MLlib, Dask

Additional Tools:

- Version Control: Git
- Collaboration: Jupyter Notebook, Google Colab
- Notebooks: Jupyter Notebook, RStudio Notebooks
- Experiment Tracking & Management: MLflow
- Natural Language Processing (NLP): NLTK
- Computer Vision: OpenCV, TensorFlow Lite

Career History

Niab Ltd, UK	April 2023 – March 2025	Data Science Manager
CFF UK	May 2020 – March 2023	Geospatial Data Scientist
Chalmers University of Technology, Sweden	Dec 2021– July 2022	Data Scientist (Master thesis)
University of Nottingham Malaysia	Nov 2017 – May 2020	Geospatial Data Scientist
University of Nottingham Malaysia	Aug 2014 – Nov 2017	Data Manager
EJDataWorks	Sep 2008 – May 2014	Geospatial Lead
National Univesity of Singapore	Jan 2008 – Jul 2008	Data Entry Specialist
EJDataWorks	Jun 2006 - Present	Data Scientist (part-time)
Universiti Putra Malaysia	Nov 2003 – Aug 2008	Geospatial Analyst
Pooladin Inc.	Sep 2001 – Aug 2002	Director of the Administrative Affairs
Yekanbazr Agro-Industry Co., Pvt.	Jan 2000 - Dec 2000	Project Engineer
Ferdowsi University of Mashhad	Nov 1995 - Dec 1999	IT Support Specialist
Ferdowsi University of Mashhad	Nov 1995 - Dec 1999	Undergraduate Research Assistant

Qualifications

MSc, Applied Data Science, Department of Computer Science and Engineering, University of Gothenburg and the Chalmers University of Technology. Sweden. (Axel-Adler scholarship)

PhD, GIS and Geomatics Engineering, Faculty of Engineering, Universiti Putra Malaysia.

MSc, Precision Farming, Faculty of Engineering, Universiti Putra Malaysia, Malaysia.

B.S. Agricultural Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad.

CAREER EXAMPLES:

Ottomatter LLM Software Developer Jul 2024-Present

EJDataWorks

Project Name: Shortlisting

Category: Generative AI

Project Description: This project involves developing an application that utilises infrencing generative models to develop automated criteria for scoring applicants who apply for jobs in Princess Alexandra Hospital in the UK.

My role:

- Fine-tuned Llama 3.1 with Salesforce customer conversation data
- Developed the front-end application using Python Flask engine
- Built complete CI/CD pipline for the software development
- Worked with prompting major Large Language Models (LLMs): OpenAI, Gemini and Claude
- **Highlight:** The client is now able to more efficiently score the applicants based for myriads of jobs at the PAH (more than 500 jobs yearly).

European Comission Data Science Manager Jul 2023-Nov 2023

Niab, UK

Project Name: LandSupport Agriseum

Category: Data Mining & Knowledge Management

Project Description: This project involved creating a centralized database to store and make accessible data from agronomic studies conducted across European agro-climatic zones between 1972 and 2022. The project aimed to facilitate research and knowledge sharing in the agricultural domain.

My role:

- Managed a team of 5 data curators and data mining experts to develop a database for more than 500 agronomic studies that were published across European agro-climatic zones: https://zenodo.org/records/8406516
- Developed a front-end data data portal with PHP and MYSQL for the database
- The project is implemented on Google Cloud Platform compute engine.
- **Highlight:** The database has been and used in 55 projects globally and has become a defacto crop modelling dataset for ARMOSA modelling.

European Commission Geospatial Data Scientist Jun 2017-Mar 2023

Nottingham University

Project Name: LANDSUPPORT ProResilience

Category: Data Science & Climate Modelling

Project Description: This project, titled LANDSUPPORT ProResilience, aimed to evaluate the potential of proso millet as a climate-resilient crop that could promote sustainable agricultural diversification in Sri Lanka. The project made extensive use of data management, simulations and modelling across a range climate data for the currect and future scenarios.

My role:

- Developed a modelling framework to assess the potential of proso millet as a climate-resilient crop for sustainable diversification. https://zenodo.org/records/7456224
- Calibrated Agricultural Production Systems Simulator (APSIM) model for 95 locations in Sri Lanka.

- Managed a large simulation modelling and mapping cluster for climate computation on GCP
- **Highlight:** The project equips extension workers with the knowledge that is required for guiding millet farmers in Sri Lanka.

European Commission Data Science Manager Sep 2021-present

Niab, UK

Project Name: RADIANT CropDB

Category: Knowledge Management & Data Infrastructures

Project Description: This project combines aspects of knowledge management and data infrastructures. It establishes a centralized repository (knowledge base) for information on plant species relevant to agricultural diversification, making this data accessible to a wider audience. The project also highlights the technical expertise involved in managing and deploying such a large-scale dataset on the Google Cloud Platform.

My role:

- Developed a Global knowledge base for more than 2700 species for agricultural diversification with MYSQL, PHP https://github.com/geoej/RADIANT-DB
- Information on over 9 million data points related to the plant species was used
- Fully managed the service on the Google Cloud Platform (Compute Engine)
- **Highlight:** Deliverable 1.5 of the pan European RADAINT project for knowledge dissemination. The RADIANT CropDB serves as a valuable resource for researchers, policymakers, and farmers seeking to explore alternative crops and promote agricultural diversification.

European Commission Data Science Manager Jan 2024-present

Niab, UK

Category: Agritech & NLP Integration

Project Description: This project combines elements of agricultural technology (agritech) and Natural Language Processing (NLP). It bridges the gap between structured agricultural data and a large language model, aiming to generate human-readable insights through NLP techniques. The project also falls under the umbrella of agritech by focusing on data and models that can be beneficial for the agricultural domain.

My role:

- The project leverages over 2.5 GB of structured data.
- The project utilizes the capabilities of the Gemini large language model.
- Ongoing project that is aimed at linking the domain specific knowledge (with more than 2.5 GB of structured data) with Gemini large language model.
- The project will will be implemented on AWS with an online interface. https://github.com/geoej/AiSQLquery
- **Highlight:** The augmented Large Language Model can accurately generate human-readable adaptation measurement for any location

EJDataWorks Data Scientist Dec 2023-Apr 2024

EJDataWorks

Project: MLFLOW for health

Category: Machine Learning Engineering

Project Description: This project combines medical image analysis with machine learning engineering. It utilizes a deep learning model for chest cancer prediction, highlighting the application of AI in medical diagnosis. Additionally, the project incorporates best practices in machine learning workflows with Mlflow and DVC, showcasing expertise in machine learning engineering.

My role:

- Trained and implemented an end to end Deep Learning application for prediction of chest cancer using Mlflow experiment tracking, DVC integration for pipeline tracking and user app creation https://github.com/geoej/EndtoEndDVC
- The model can successfully identify the cancer cases using trained model.
- **Highlight:** Deep learning module simplifies the process of chest cancer diagnosis.

European Commission Geospatial Data Scientist Dec 2020-Nov 2021

CFF UK

Project Name: LandSupport ViableCrop

Category: Land Management & Geospatial Data Analysis

Project Description: This project falls under the umbrella of sustainable land management by developing a framework to assess the feasibility of introducing new land uses while considering economic viability and potential climate impacts. The project leverages geospatial data analysis techniques using R and Google Earth Engine, showcasing their utility in this domain.

My role:

- Developed a multifaceted land suitability analysis approach for feasibility assessment based on a data from different sources (MySQL databases, Google Earth Engine).
- Performed detail mapping and land use planning using R Geospatial packages.
- Conducted feasibility assessment for profitability and ROI.
- **Highlight:** First ever end to end feasibility for a new land use option that includes economics and projection agains climate change

CFF UK Geospatial Data Scientist Sep 2022-May 2023

CFF UK

Project Name: UK Crop Diversification

Category: Data Science & Sustainable Food Systems

Project description: This project combines agricultural data science with the goal of promoting sustainable food systems. It utilizes diverse data sources and geospatial analysis techniques to identify climate-resilient crops that could contribute to a more diversified and secure food supply for the UK.

My role:

- Developed an analysis framework for diversification in the United Kingdom.
- Climate, soil, economics, and market demand data was used.
- API database of 2700 crops, combined GIS and data mining to determine the most promising climate resilitent crop for the UK: https://www.mdpi.com/2077-0472/13/4/787

Highlight: he project's significance was recognized by its selection for presentation to HRH Princess Ann on October 10th, 2023, focusing on solutions for UK food and nutrition security. (https://www.mdpi.com/2077-0472/13/4/787)

European Commission. Geospatial Data Scientist Nov 2019-July 2022

Nottingham University

Project: LandSupport CropDivMap

Category: Data Science & Spatial Planning for Sustainability

Project Description: This project pioneered a data-driven approach to identify underutilized crops (UCs) with potential for sustainable agricultural diversification in Italy. This project falls under the umbrella of agricultural

data science and spatial planning for sustainability. It leverages data analysis techniques to identify underutilized crops with potential for diversification, promoting a more resilient and sustainable agricultural system in Italy.

My role:

- Developed a data-driven approach for shortlisting underutilized crops (UCs) for sustainable agricultural diversification across Italy: https://www.mdpi.com/2073-4395/12/7/1636
- Lead the team for data management and mapping.
- Published the first every map for diversificatio of crops in a European country: https://www.mdpi.com/agronomy/agronomy-12-01636/article_deploy/html/images/agronomy-12-01636-g003.png
- **Highlight:** The first ever map for crop diversification was produced for ITALY. This shortlist provides valuable insights for policymakers, researchers, and farmers who are interested in promoting crop diversification.

EJDataWorks. Geospatial Data Scientist Dec 2022-Feb 2023

Stockholm Environment Institute

Project Name: TRASE: Data-Driven Transparency Initiative for Trade Mapping

Project category: Environmental Monitoring & Geospatial Data Analysis

Project Description: This project, TRASE, focused on promoting transparency in international trade and financing, with a specific application to deforestation in the Amazon rainforest.

My role:

- Conducted land cover analysis to assess the deforestation extent within the Amazon region. This likely involved utilizing satellite imagery and geospatial data analysis techniques.
- Developed automated routines using Google Earth Engine (cloud-based geospatial analysis platform) and R spatial packages (statistical software for spatial data analysis). These automated routines likely streamlined the analysis process and facilitated efficient monitoring of deforestation patterns.
- **Highlight:** The project received positive feedback for its delivery and write-up, indicating successful completion and clear communication of the findings.

European Commission Geospatial Data Scientist Jan 2021-Jan 2022

CFF UK

Project Name: China Crop Diversification

Category: Spatial Agricultural Planning & Sustainable Food Systems

Project Description: This project explored the potential for diversifying China's agricultural landscape by investigating the suitability of bambara groundnut as a new crop.

My role:

- Carried out spatial suitability analysis of bambara groundnut in China using stochastic environmental research and risk assessment, sub-daily swat models, cluster analysis, economic data, climate data, soil data, crop production data, and FAOSTAT data: https://onlinelibrary.wiley.com/doi/full/10.1002/fes3.35 8
- Lead the team for data collection and mapping and analysis.
- **Highlight:** The project's findings have demonstrably made an impact, influencing land-use planning for food and nutrition security in both Africa and China.

EJDataWorks Geospatial Data Scientist July 2019-Dec 2020

EJDataWorks

Project Name: 3D Digital Maps

Category: Spatial Data Science & Open Data Infrastructure

Project description: This project combines elements of spatial soil science and open data infrastructure. It leverages geostatistical analysis and spatial modeling techniques to created a 3D digital soil database, providing a freely accessible resource for further research on soil properties and related agricultural and environmental applications. The project's pioneering nature and lasting impact within Sri Lanka make it a notable contribution to the field.

My role:

- Lead the development of The first version of nation-wide open 3D soil database for Sri Lanka: https://www.sciencedirect.com/science/article/pii/S235234092031235X
- Performed getostatistical analysis with R and open source GIS software.
- Prediction was made for all locations across the country and across the depth of 100cm.
- **Highlight:** First attempt to develop digital soil data for an Asian country. The database has seince been used for soil carbon stock and agronomic biofortfication in Sri Lanka

Chalmers Universithy of Technology Sweden Data Scientist Dec 2021-Jul 2022

EJDataWorks

Project Name: Spatial Generative Adversarial Neural Networks

Category: Environmental Machine Learning

Project description: This project explored the application of Spatial Generative Adversarial Neural Networks (SGANs) for predicting soil properties in unsampled areas. Due to the high cost of traditional soil sampling methods, this project aimed to develop a more cost-effective and accurate approach. By utilizing SGANs in conjunction with machine learning models, the project successfully demonstrated the ability to improve the accuracy of soil property data in unsampled locations. Additionally, the project implemented SGANs for data mining from published resources, showcasing the potential of this technique for augmenting traditional data collection methods.

My role:

- Utilised the Spatial Generative Adversarial Neural Networks (SpaceGAN) and machine learning models to predict soil properties at unsampled areas (due to high cost). The results show that this method can improve the accuracy of soil information.
- Spatial Generative Adversarial Networks (SpaceGANs) for data mining from published resources and machine learning models using Python Pytorch https://gupea.ub.gu.se/handle/2077/72200?show=ful
- **Highlight:** This project used Generative Machine learning to develop and augment traditional data

European Commission Geospatial Data Scientist Nov 2017 - May 2020

Nottingham University

Project Name: CropBASE: A Suite of Apps for Agricultural Diversification

Category: AI-Powered AgTech for Diversification

Project description: This project involved the development of CropBASE, a suite of mobile and web applications designed to promote agricultural diversification. As a Geospatial Data Scientist at the European Commission's Joint Research Centre (JRC), I led a team of 24 data specialists and software developers in creating CropBASE. The project utilized PHP and MySQL for the database and data portal, and further implemented the platform on Google Cloud Platform to ensure scalability and accessibility.

My role:

- Lead a team of 24 data specialists, and software developers to develop CropBASE suite of Apps for agricultural diversification: https://www.cropbase.co.uk
- The database data portal was implemented based on PHP and MYSQL and further implemented on the Google Cloud Platform
- **Highlight:** CropBASE's success is evidenced by the EUR 290,000 investment it received from the European Commission's RADIANT project. This funding signifies the project's potential to contribute significantly to the advancement of agricultural diversification initiatives.

European Commission Geospatial Data Scientist Jan 2018-May 2020

Nottingham University

Project Name: SelectCrop

Category: AI-Powered Precision Agriculture

Project Descriptions: This project, undertaken as a Geospatial Data Scientist at the European Commission's Joint Research Centre (JRC), involved the development of a prototype application called SelectCrop. This web-based tool allows users to perform on-the-fly analysis of crop suitability for any terrestrial location. I led the team in creating the prototype, utilizing a combination of PHP, R, and SQL for the backend development. The project leveraged Google Cloud Platform (GCP) to ensure scalability and global accessibility.

My role:

- Lean the team to develop Crop suitability mapper prototype available a https://www.cropbase.co.uk/src/cropgrids/cropgrids.php
- The backend is developed based on PHP, R, SQL and is implemented on GCP
- **Highlight:** SelectCrop's key strength lies in its ability to provide real-time crop suitability analysis for any location on Earth. This empowers farmers and agricultural stakeholders to make informed decisions regarding crop selection based on specific environmental conditions.

European Commission Data Scientist July 2019-present

Nottingham University

Project Name: AssessCrop

Category: EU Nutrigenomics Data Platform

Project Description: This project aimed to establish a robust evidence base for the potential of underutilized crops by analyzing their nutritional value. AssessCrop offers a data portal showcasing a comprehensive database of food products and their corresponding nutritional properties across the European Union. The backend development utilized MySQL on Google Cloud Platform (GCP), ensuring scalability and accessibility.

My role:

- Lead the team to develop an evidence-base for new products and their nutrition from underutilised crops: https://cropbase.co.uk/src/assesscrop/assesscrop/product_compare.php
- The backend is developed based on MYSQL on GCP and a data portal based on Sql Maestro technology
- **Highlight:** AssessCrop serves as a valuable resource for the European Union, promoting the exploration and utilization of underutilized crops as potential sources of nutrition. This data platform empowers researchers, policymakers, and food producers to make informed decisions regarding crop diversification and dietary health.

European Commission Data Manager Nov 2014-May 2020

Nottingham University

Project Name: Global Knowledge Base

Category: Open-Access AgDiversity Knowledge Base

Project Description: This groundbreaking project represents the first-ever attempt to establish a digital infrastructure dedicated to promoting agricultural diversification. The knowledge base, accessible at https://cropbase.co.uk/, provides open and accessible information on underutilized crops, empowering farmers and stakeholders worldwide.The project utilized MySQL on Google Cloud Platform (GCP) for the backend development and Sql Maestro technology for the data portal, ensuring scalability and ease of use.

My role:

- Lead the development of global knowledge base for underutilised crops : https://cropbase.co.uk/cropbasev5/
- The backend is developed based on MYSQL on GCP and a data portal based on Sql Maestro technology.
- This is the first ever attempt to create a digital infrastructure that is open and accessible for agricutlural diversification
- **Highlight:** The project's impact extends beyond Europe. The suite of software tools and information within the knowledge base has been adopted by over 10,000 farmers in Southeast Asia, demonstrating its effectiveness in promoting diversification practices on a global scale.

Nottingham University Data Scientist Sep 2016- Jan 2020

Nottingham University

Project Name: SVM for Economics

Category: AI-Powered Economic Assessment for Underutilized Crops

Project Description: The project proposes a framework to assess the economic feasibility of cultivating underutilized crops in rural areas. This framework leverages Support Vector Machines (SVM) and Genetic Algorithms (GA) to analyze data and make rapid assessments suitable for rural development programs.

My role:

- Participate in the developmen of Support Vector Machines algorithm for the economic feasibility assessment of underutilised crops: https://www.sciencedirect.com/science/article/abs/pii/S0168169918306665
- Acquire data of farm income for the analysis for more than 60 rurual economic area in Sri Lnaka
- **Highlight:** The project resulted in a general economic model for rural villages that demonstrates good classification accuracy in assessing the economic feasibility of underutilized crops.

EJDataWorks Data Scientist Jan 2020-May 2020

Nottingham University

Project Name : Nutrition Security for sustainable developemnt

Category: AI for Sustainable Food Security with Underutilized Crops

Project Description: This innovative project represents the first attempt to visually translate research data on underutilized food crops into actionable insights for decision-makers. The tool, accessible at https://geoprocessing.shinyapps.io/underutilised/, leverages the power of R and the Shiny framework to create an interactive and user-friendly platform.

My role:

- Develop A Nutrition Security Tool demonstrating the importance of Underutilised Food Crops for Sustainable Nutrition Security https://geoprocessing.shinyapps.io/underutilised/
- The backend was developed based on R and Shiny and implmented on Shinyapp cloud
- **Highlight:** This is the first attempt to visualise research data for decision making using the power of R and Shiny

EJDataWorks Data Scientist Jan 2020-May 2020

EJDataWorks

Project Name: CropModel Input Generator: Streamlining Crop Modeling Workflows

Category: AI-Powered Crop Modeling Platform

Project Description: This innovative web application, accessible at https://geoprocessing.shinyapps.io/DSSAT-INPUT-Shiny/, represents the first attempt to utilize R and the Shiny framework to create a user-friendly platform specifically designed to generate inputs compatible with DSSAT crop modeling software.

My role:

- DSSAT-INPUT: an online DSSAT-compatible crop modelling input generatorhttps://geoprocessing.shinyapps.io/DSSAT-INPUT-Shiny/
- The backend was developed based on R and Shiny and implmented on Shinyapp cloud
- **Highlight:** The CropModel Input Generator simplifies the process of creating crop model inputs, a traditionally time-consuming and complex task. This web-based tool empowers researchers and agricultural professionals to streamline their workflows and dedicate more time to analysis and interpretation of modeling results.

Government of Malaysia Geospatial Lead Sep 2010 - Nov 2013

EJDataWorks

Project Name: Spatial Temporal Regression for Real Estate Valuation

Category: Advanced Real Estate Valuation with Spatiotemporal Analysis

Project Description: This project delves into the comparison of linear regression and spatial-temporal autoregressive models for mass appraisal of single-storey residential properties. An R code had to be developed to analyze the specifications of these models and assess their effectiveness in valuation.

My role:

- Developed R code to compare specifications of linear regression and spatial-temporal autoregressive models in mass appraisal valuation for single storey residential property: http://psasir.upm.edu.my/id/eprint/60054
- **Highlight:** Improved Accuracy in Mass Appraisal: By comparing and potentially identifying a superior model (spatial-temporal autoregressive model), the project could have implications for improving the accuracy and efficiency of mass appraisal processes.

Government of Malaysia Geospatial Lead Jan 2014 - Sep 2014

EJDataWorks

Project Name: Geostatistical web mapping for precision farming

Category: Precision Agriculture Platform (PAP)

Project Description: As Geospatial Lead at EJDataWorks for the Government of Malaysia (Jan 2014 - Sep 2014), I spearheaded the development of PFMap, an innovative web-based application designed to empower precision farming practices. PFMap leverages R statistical libraries to create a user-friendly platform for geostatistical soil mapping.

My role:

- Developed Shiny PFMap software, soil mapper using R statistical libraries
  https://geoprocessing.shinyapps.io/geoprocessing/
  data for testing the App: http://b.link/a69

- Precision Soil Analysis: PFMap allows farmers to upload or utilize existing data to generate high-resolution soil maps, identifying spatial variability in soil properties crucial for informed decision-making.
- Data Visualization: The application utilizes interactive map visualizations to effectively communicate complex soil data, enabling farmers to easily identify areas requiring specific fertilizers, amendments, or other crop management practices.
- **Highlight:** PFMap represents a significant advancement in precision agriculture for Malaysia. By providing a user-friendly and accessible tool for geostatistical soil mapping, PFMap empowers farmers to optimize resource allocation, improve crop yields, and promote sustainable agricultural practices. A demo of the PFMap software here: https://geoprocessing.shinyapps.io/geoprocessing/. Sample data for testing the application is available here: http://b.link/a69.

1Spatial Data Scientist Jan 2014 – Sep 2014

EJDataWorks

Project Name: Automatic valuation system for real properties (stand alone software)

Category: AI-Powered Real Estate Valuation (Standalone Software)

Project Description: This innovative tool utilizes geospatial regression techniques to automate the valuation process for real properties. The software is built on R programming language for statistical analysis and leverages TCL/TK for a user-friendly graphical interface.

My role:

- Developed An Automated Valuation System for real estate appraisal based on geospatial regression techniques.
- **Highlight:** The standalone software automates the valuation process, potentially leading to significant improvements in efficiency and scalability compared to traditional manual methods. By creating a user-friendly software application, this project could potentially make real estate valuation more accessible to a wider range of users, such as appraisers, investors, and homeowners.

Government of Malaysia Geospatial analyst Nov 2003 – Aug 2008

EJDataWorks

Project Name: Optimizing Geospatial Data Collection and Analysis for Resource Management

Category: Efficient Geospatial Sampling and Analysis for Resource Management

Project Description: Development of a novel geostatistical analysis and mapping algorithm. This innovative approach resulted in a substantial reduction (70%) in the number of samples required for spatial mapping, significantly improving efficiency and reducing costs associated with data collection. The

My role:

- Developed a complete algorithm for geostatistical analysis and mapping that would reduce the number of samples required for mapping by 70%: https://link.springer.com/article/10.1007/s12517-015-1912-6
- Developed object-oriented approach for image segmentation using LandSat data (UGSS)
- Granted RM 135,000 (Ringgit Malaysia). "Developing a cellular automata-based planning decision support system" (01-01-04-SF0979), Ministry of Science, Technology and Innovation
- **Highlight:**
    - The innovative sampling algorithm demonstrates a significant achievement with real-world impact. By reducing the number of samples needed for accurate mapping, this project has the potential to optimize resource allocation and streamline data collection processes in various government sectors responsible for resource management, such as environmental monitoring, forestry, and land-use planning.

- The project titled "Developing a cellular automata-based planning decision support system" received a grant of RM 135,000 (Ringgit Malaysia) from the Ministry of Science, Technology and Innovation (Ref: 01-01-04-SF0979). This further highlights innovative contributions in the field of geospatial analysis and planning.

National University of Singapore Data Entry Specialist Jan 2008 - Jul 2008

EJDataWorks

Project Name: Land Cover Change in Southeast Asia (Data Entry Specialist)

Category: Remote Sensing Data Management

Project Description: To support the research on land-use land-cover change (LULC) in Southeast Asia the project needed a metadata system to enable tracking the datasets (sattellite and ground data). A literature review was also needed to be conducted to develop self-oranising maps for LULC concepts.

My role:

- Designed and implemented a metadata system to organize and manage remote sensing data used in the LULC project. This system likely involved creating detailed descriptions of the data source, format, acquisition date, and other relevant information, ensuring efficient data access and analysis for researchers.
- Conducted a literature search for the project titled "Land-use land-cover change research explored using self-organizing map." This likely involved identifying relevant academic publications, summarizing key findings, and potentially creating a bibliography to support the research efforts.
- **Highlight:** The work provided a critical foundation for these environmental research projects. By establishing a robust metadata system, transcribing important data records, and conducting literature reviews, the accuracy, organization, and accessibility of information, ultimately contributing to a deeper understanding of LULC patterns and their environmental consequences in Southeast Asia.

National University of Singapore Data Entry Specialist Jan 2008 - Jul 2008

EJDataWorks

Project Name: Forest Change in Western New York (Data Entry Specialist)

Category: Ecological Data Management

Project Description: This project focused specifically on data transcription for the research investigating "Effects of clearance and fragmentation on forest compositional change and recovery after 200 years in western New York."

My role:

- Transcribed records for a separate project investigating the "Effects of clearance and fragmentation on forest compositional change and recovery after 200 years in western New York." This task involved meticulously converting handwritten recordings into digital text, ensuring accurate data preservation and accessibility for further analysis.

Highlight: The meticulous data transcription efforts ensured the preservation and accessibility of valuable ecological data. This data likely documented the composition of forest ecosystems over time, including changes in tree species abundance and diversity due to past human activities like clearance and fragmentation. By accurately converting handwritten or audio records into digital format, I facilitated further analysis and contributed to a better understanding of forest recovery processes in western New York.