Disrupting Carbon Footprint

Location: Remote / In-office (Bangalore)

Role Type: Summer Internship

About Varaha:

Varaha is a climate tech venture focused on developing economies. Established in 2022, the company collaborates with smallholder farmers and local communities to implement carbon removal projects aimed at combating climate change by capturing greenhouse gas emissions.

Role: AI\ML Intern

What you'll be doing

- Understand and define object identification use cases by analyzing problem requirements, data characteristics, and end-user expectations.
- Develop a standardized training and annotation pipeline, ensuring consistency across datasets and enabling efficient labeling workflows.
- Design and implement case-specific, state-of-the-art computer vision models, and perform comparative analysis to evaluate their performance across key metrics.
- Establish robust model tracking and analysis mechanisms, including accuracy assessment, performance benchmarking, and time/resource profiling.
- Conduct comprehensive model testing and use-case validation, documenting results, limitations, and potential improvements.

- Refactor code into modular, maintainable components following Python PEP standards and containerize the complete workflow using Docker for portability and reproducibility.

Key Responsibilities

- Implement and explore existing machine learning models for various use cases (computer vision, classification, regression, clustering, etc.)
- Work with open datasets — clean, preprocess, and prepare them for model training and evaluation
- Use State of the art ML libraries for model deployment
- Analyze model performance and suggest basic improvements (hyperparameter tuning, feature engineering, etc.)
- Document work done — maintain clear code, comments, and reports explaining model logic and results
- Collaborate with senior team members on exploratory data analysis (EDA) and visualization
- Optionally, support data collection and annotation if needed for training new models

Required Skills:

- Strong proficiency in Python — must be comfortable writing scripts, working with files, and using libraries
- Familiarity with ML and DL algorithms
- Hands-on experience with at least one popular ML library (e.g., scikit-learn, TensorFlow, PyTorch)
- Good understanding of model evaluation techniques (accuracy, precision, recall, RMSE, etc.)
- Ability to read and implement research papers, online repositories & documentation (Hugging Face, PaperwithCode and similar) and understand existing implementations

Who Should Apply

- Students pursuing or recently completed B.E/B.Tech/M.Tech in Computer Science, Data Science, or related fields.
- Candidates who have worked on ML projects as part of coursework, hackathons, or internships.
- Self-starters who can independently research, learn and apply new concepts during the internship.

Role: RS\AI Intern

What you'll be doing

- Build automated workflows for a few feasibility models.
- Develop a structured reporting framework that automatically compiles analysis outputs into standardized reports. This includes generating maps, time-series plots, and summary statistics directly from processed data.
- Enable flexibility by allowing users to define parameters (AOI, time period, crop type, etc.)
- Automatically track analysis progress and link generated reports back to corresponding Notion tickets, ensuring seamless traceability and collaboration.
- Implement basic quality checks (missing data, projection mismatches) and trigger alerts if any issues are detected during the automation run.

Key Responsibilities:

- Implement and explore existing machine learning models / rule based models using geospatial and remote sensing data for sustainability applications.
- Work with raster, vector, and time-series geospatial datasets — perform preprocessing, cleaning, and transformation.
- Use libraries such as geemap, ee, scikit-learn, TensorFlow, PyTorch, and geospatial libraries like geopandas, rasterio, GDAL, xarray, etc.

- Support development of image processing, classification, regression, and clustering models using remote sensing data (e.g., deforestation, canopy metric estimation).
- Perform spatial analysis using Google Earth Engine (GEE) and visualization using tools like QGIS, Python (matplotlib, plotly), etc.
- Assist in developing workflows to integrate remote sensing data into machine learning pipelines.
- Document processes, workflows, and model performance for future reference.
- Collaborate with the team on exploratory spatial data analysis (EDA) and generation of spatial insights.
- Optionally, support data collection and annotation if needed for training new models

Required Skills

- Strong proficiency in Python — must be comfortable with data handling, scripting, and geospatial libraries.
- Understanding of basic machine learning algorithms and ability to implement existing models
- Experience working with geospatial data formats (shapefiles, GeoTIFFs, NetCDF, etc.).
- Familiarity with libraries like geemap, ee, geopandas, rasterio, GDAL, xarray, scikit-learn, tensorflow, or pytorch.
- Basic experience handling remote sensing data (optical, multispectral, radar) and time-series analysis is expected.

Who Should Apply

- Students pursuing or recently completed B.E/B.Tech/M.Sc/M.Tech in fields such as Computer Science, Data Science, Remote Sensing, GIS, Geoinformatics, Environmental Science, or related disciplines.
- Candidates who have worked on ML/Geospatial projects in academic coursework, internships, hackathons, or personal projects.

- Passionate learners who are excited to work at the intersection of AI, Earth Observation, and Sustainability.

How to Apply

Send your CV along with details of your notable projects and accomplishments to careers@varahaag.com, using the subject line "**Top20 College Internships Program - Application.**"