

Initial Project Draft:

Analyzing the Impact of Solar Radiation on Agricultural Soil Conditions Using Data Science and AI Models

Problem Definition:

As we all know, agriculture plays a great part in our daily lives. It provides us with valuable food. However, the climate and many other factors affect the agricultural conditions. For instance, solar radiation plays a crucial role in determining soil temperature and moisture levels, which in turn significantly affects agricultural productivity. This project aims to investigate the impacts of varying levels of solar radiation on soil temperature and moisture, using data science and machine learning techniques. In addition, I also wish to evaluate the carbon costs throughout the process, such as kilograms of CO2 equivalent per kilogram of crop produced, to evaluate the environmental sustainability of agricultural practices. By understanding these impacts, this project seeks to offer insights into optimizing agricultural practices for better crop yields and more sustainable farming methods in the context of changing climate conditions.

Data Collection and Refinement:

Data will be sourced from VisualCrossing with the weather query builder. Different datasets will be obtained for 10 distinct locations. I will focus on different places in America. To be more specific, these locations will include, but are not limited to, areas in North America (e.g., Arlington, Troy), Western areas such as Napa Valley and the Palouse, Eastern areas like Loess Plateau, and Southern areas like Pampas region and Nile Delta, ensuring an effective perspective that captures varying soil types, crop varieties, and climate conditions in America. Then, I aim to look for different combinations of datasets consisting of solar radiation levels, weather, wind energy, and agricultural research databases for soil temperature and moisture content across various geographic locations and crop types. Other sources might also be included and well documented throughout the analysis.

After getting the related datasets, I will utilize different preprocessing steps, including data cleaning, such as removing outliers, handling missing values, normalization, and potentially transformation, to align the scales of measurement and temporal resolution across different data sources. In addition, exploratory data analysis will be conducted to see the correlations for feature selections before using models. I will also make training and validation sets for later use if the model requires them. However, these are my planned data collection and refinement steps; the actual process might be different but will be documented if changes are made.

Commented [DL1]: Overall, this proposal has all the right elements in the 4 main sections (problem, data, implementation, evaluation). 2 out of 2 points. Please take into account the comments below for your refined proposal.

Commented [DL2]: How many different locations? Across which geographic areas?

Implementation:

The project will employ regression or classification models and time series analysis to examine the relationship between solar radiation levels and soil conditions. Machine learning techniques, such as Random Forests and Gradient Boosting Machines, will be used to predict soil temperature and moisture based on historical solar radiation data. The reason is that I think I might find the relationship between solar energy and soil conditions to not be linear. Therefore, Random Forests can help me handle non-linear relationships and Gradient Boosting Machines can give me optimization for accuracy. For the prediction of crop yields, the ARIMA model can be utilized. In addition, I also want to utilize different deep learning methods such as LSTM, RNN, and GRU to make forecast comparisons on the solar radiation levels so that we can know further about solar energy and the soil conditions. All the processes will be done through Python and recorded through the GitHub repository.

Evaluation:

I will use evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to evaluate the accuracy of the predictive models. Additionally, model performance will be assessed in terms of its ability to generalize across different agricultural zones and crop types. Also, confusion matrix can be included to see how well the classification method performs. The models' predictions will also be validated against actual measured changes in soil temperature and moisture, with a particular emphasis on their usefulness in predicting outcomes under various weather conditions such as wind, solar, air moisture, and so on. Therefore, basically, the evaluation will consist of tables using numeric values such as accuracy, MAE, RMSE and figures such as confusion matrix and validation plots.

In addition, since in the project description we also need to address carbon costs, this project will introduce carbon efficiency metrics by utilizing CodeCarbon. I believe that this can estimate the amount of CO₂ produced by the computational resources used during model training. With insights into the carbon impact of our model training processes as well as the other metrics such as MAE, RMSE, we can make informed decisions to optimize our code and choose more energy-efficient methods, potentially reducing the overall carbon footprint.

Conclusion:

This project proposal aims to give a basic idea of my topic, datasets, models, and methods that I will conduct in the near future. My goal is to discover and analyze the agricultural and environmental sciences by providing a data-driven understanding of how solar radiation affects soil conditions critical to crop growth. Through the application of DS/ML/AI techniques, this project will provide valuable insights into soil temperature and moisture dynamics, supporting the development of more sustainable and resilient agricultural practices. The findings could also inform adaptive farming strategies that optimize irrigation and planting schedules

Commented [DL3]: Can you say more about this? How will you do it?

When I suggested to address carbon costs, I meant the carbon cost of the ML models you're training. So please specify how you'll do that (e.g. use CodeCarbon).

If you were going to report on the carbon efficiency of crops, just to address my instruction to measure carbon costs, you don't need to. I meant to measure the carbon cost of the ML, not of the climate topic you're addressing.

based on predicted soil conditions and solar radiation levels, enhancing crop resilience to environmental stressors and contributing to food security.

Reference:

Visual Crossing. (n.d.). Weather Data Services. Retrieved February 21, 2024, from <https://www.visualcrossing.com/weather/weather-data-services>