

Executive Summary

Data Analysis of Soil Parameters at a Local Land Preserve used a data analysis approach to analyze data previously collected in a study titled *Environmental Restoration Target Estimation Around Engquist Nature Preserve* by Jack Colwell and other students at Valparaiso University. This project's objective is to use our various soil parameters as predictors for SOC in the previously collected soil samples. The SOC content in soil varies, influencing the ability to manage healthy agriculture and the global carbon cycle. Understanding these SOC levels helps farmers make more informed decisions on how to maintain healthy soil levels and mitigate climate change.

The data was cleaned and errors were removed using Python in a Google Colab environment. Then a correlation matrix was created to narrow down the highest correlated variables to use in our prediction for SOC content. From this matrix, it was found that two variables from this dataset are the best predictors to use: Carbonate-Free Mineral Sample (g) and Mineral Sample (g). By using these two variables, several regression models were run to output the RMSE (Root Mean Square Error) and R^2 value of each regression model, as these results influence the decision on which model is best to apply.

Key findings indicate that the random forest regressor, which tends to resist overfitting, is the best model for this situation and data. That predictive model outputs are as follows:

≈ 0.760 RMSE

≈ 0.914 R^2

This means that approximately 91.4% of the variance in the data can be explained by our model. That is to say, using these two variables, the SOC content can be reliably predicted, without a direct measure of SOC content itself. We validated that the random forest regressor is not overfitting as we checked the R^2 value in the training section and it resulted in 0.8731. When dealing with an overfitted model the R^2 value is extremely close to 1.

These findings demonstrate that accurate SOC predictions can be achieved without measuring all variables, significantly reducing costs and time. By optimizing soil testing practices, farmers can improve crop management and enhance resilience to climate change. Moreover, this research aligns with broader goals in climate-smart agriculture by supporting efforts to increase soil carbon stocks and reduce greenhouse gas emissions.