Higher School of Agriculture of Mograne

جـامعة قــرطاج
Université de Carthage

Homework Project

Group: 3^{rd} Engineer in agronomy

Prof. Bilel AMMOURI
Period: 2022 – 2023

Homework Project: Predicting African Soil Fertility using Dry Chemistry Methods and Machine Learning

Background:

The goal of this project is to explore the correlation between soil elemental analysis obtained through cost-effective dry chemical methods (specifically XRF and FTIR spectroscopy) and soil fertility. Traditional wet chemical methods for soil nutrient measurements can be expensive, posing a challenge in scaling soil surveillance. The implementation of lower-cost dry chemical methods, coupled with machine learning, provides an opportunity to infer nutrient concentrations and assess soil quality efficiently and affordably (Ref. (Git repository).

Data

The dataset hosted at arn:aws:s3:::afsis contains field and laboratory measurements of soil samples collected through the Africa Soil Information Service (afsis) project, spanning from 2009 through 2018. Georeferenced soil samples were collected from various countries throughout Sub-Saharan Africa. The nutrient content of these samples was analyzed using both wet chemistry methods (e.g., Mehlich-3) and dry chemistry methods (e.g., infrared spectroscopy, x-ray fluorescence).

The dataset encompasses the following variables: M3 Ca, M3 K, M3 Al, M3 P, M3 S, pH, Psa asand, Psa asilt, Psa aclay, Volfr, Awc1, Lshrinkpct, $Flash2000_{Np}pm$, Acidified carbon, pH, $Leco_{Np}pm$, C % Org, ICP OES K mg/kg, ICP OES P mg/kg, P, K, S, Ca, Mg, Cu, Cl, Zn, Fe, Mn, Mo, Latitude, Longitude, Cluster, Depth, Country, Cultivated. The target variable is $Flash2000_{Np}pm$.

Objectives:

1. **Data Acquisition**:
   - Obtain a dataset that includes soil elemental analysis data obtained through XRF and FTIR spectroscopy, along with corresponding information about soil fertility and crop presence.
2. **Data Exploration and Preprocessing**:
   - Perform exploratory data analysis (EDA) to understand the characteristics of the dataset.
   - Handle missing values, outliers, and any other data preprocessing steps deemed necessary.
3. **Feature Engineering**:
   - Explore the possibility of creating new features from the existing data that might enhance the predictive power of the model.
4. **Model Development**:
   - Split the dataset into training and testing sets.
   - Develop a machine learning model to predict soil fertility based on the dry chemistry measurements.

- Experiment with different algorithms (e.g., regression, classification) and evaluate their performance.

5. **Model Evaluation and Interpretation**:
   - Assess the model's accuracy, precision, recall, and other relevant metrics.
   - Interpret the results and understand the significance of the features in predicting soil fertility.
6. **Optimization and Fine-Tuning**:
   - Explore hyperparameter tuning and other optimization techniques to enhance the model's performance.
7. **Communication**:
   - Prepare a comprehensive report or presentation summarizing the findings, methodology, and implications of the model.
   - Clearly communicate the potential applications of the developed algorithm for quick and cost-effective screening of soil quality.

Deliverables:

1. Jupyter Notebook or R Markdown document containing the code, analysis, and visualizations.
2. A written report or presentation summarizing the project, methodology, results, and conclusions.

Deadline:

Please submit your completed project by **December 25, 2023**.