

Previsão das propriedades de solos agrícolas através da utilização do sensor NIR

Aprendizagem Automática Aplicada

Docente: Manuel Campagnolo

Ana Carolina Santos Menoita nº 25953

Maria Rafaela da Silva Navalho nº 26219

Categoria do Projeto: Dados Tabulados (Regressão)

Este projeto propõe investigar a viabilidade de prever propriedades físico-químicas do solo utilizando exclusivamente dados espectrais NIR (Near Infrared Reflectance). Contrariamente às análises laboratoriais tradicionais, demoradas e exigentes em mão-de-obra. Esta abordagem pretende simplificar o processo tradicional de caracterização do solo num dado local. Uma caracterização mais rápida do solo permite acelerar a avaliação das necessidades do solo, otimizar o uso de fertilizantes, reduzir desperdícios e minimizar a dependência de análises laboratoriais dispendiosas, tornando a agricultura mais acessível, sustentável e eficiente.

Identificar relações entre os espectros NIR e as propriedades do solo, pode ser um desafio, por estarmos a lidar com dados multidimensionais e ruidosos. É necessário ter em conta também a heterogeneidade natural dos solos que pode dificultar a identificação de padrões. O maior desafio será garantir que o modelo mantém precisão quando aplicado a diferentes tipos de solo o que requiere validação rigorosa com amostras diversas.

Os dados a utilizar foram recolhidos manualmente através de amostragens de solo em diferentes locais de uma herdade. Posteriormente, estas amostras foram

submetidas a análises laboratoriais: Análises químicas como o pH, condutividade elétrica, razão C:N, azoto total e outros nutrientes; Análise LOI (Loss On Ignition) para estimar a matéria orgânica; Medição espectral utilizando um sensor NIR. Todos os dados estão organizados numa tabela com colunas para comprimentos de onda NIR (*Predictor Variable*) e propriedades do solo medidas em laboratório (*Target Variables*).

Para prever as propriedades do solo a partir das leituras NIR propomos utilizar técnicas de regressão multivariada como: a regressão PLSR por ser amplamente utilizada em análise espectral e testar algoritmos como “Random Forest” e “XGBoost” para capturar relações não lineares nos dados. Os modelos serão ajustados e comparados com base no desempenho e interpretabilidade.

A avaliação do modelo será realizada através de métricas de regressão como: RMSE (*Root Mean Squared Error*) e MAE (*Mean Absolute Error*) para quantificar a magnitude dos erros de previsão, bem como o Coeficiente de Determinação (R^2) para avaliar a proporção da variabilidade explicada pelo modelo.

A robustez dos resultados será garantida através de validação cruzada *k-fold*, assegurando a generalização do modelo a diferentes subconjuntos de dados. Adicionalmente, será analisada a importância das variáveis para identificar as bandas NIR mais influentes na previsão de cada parâmetro do solo, permitindo uma interpretação fundamentada dos resultados.