



Ministério da Educação
UNIVERSIDADE FEDERAL DO PARANÁ
SEPT – Setor de Educação Profissional e Tecnológica
Especialização em Inteligência Artificial Aplicada



ALTAIR MALINOWSKI
CLEANTO EVARISTO AMORIM
GABRIEL PAULO MAI
HELDER AGOSTINHO DIAS MORAIS
PEDRO VITOR SPIRANDELLI GOMES
RENATO FERREIRA SOARES

TRABALHO DE IAA003 – LINGUAGEM R

Trabalho apresentado ao Curso de Pós-Graduação em Inteligência Artificial Aplicada, na disciplina Linguagem R, como requisito parcial para obtenção de nota.

2025

Questão 1 Pesquisa com Dados de Satélite (Satellite)

A escolha do modelo **SVM (Support Vector Machine)** para a classificação dos tipos de solo e vegetação foi baseada na análise comparativa das matrizes de confusão dos modelos Random Forest (RF), SVM e Redes Neurais Artificiais (RNA). Os critérios considerados incluíram **acurácia geral, sensibilidade, especificidade por classe**, e o **coeficiente Kappa**, refletindo a concordância entre as previsões e os valores reais.

O SVM obteve a **maior acurácia (87,07%)**, superando o RF (84,19%) e a RNA (80,84%), indicando melhor desempenho global. Destacou-se ainda pela alta sensibilidade na detecção das classes **"grey soil" (95,94%)** e **"very damp grey soil" (84,39%)**, superando os demais modelos. Em contraste, a RNA apresentou desempenho insatisfatório na previsão de **"damp grey soil"**, com **sensibilidade de 0%**.

Em relação à especificidade, o SVM também se destacou, especialmente para a classe **"very damp grey soil" (96,24%)**, o que indica menor incidência de falsos positivos. Ademais, o modelo apresentou o **maior coeficiente Kappa (0,8399)**, evidenciando maior consistência entre as previsões e os valores observados.

Dessa forma, o SVM foi escolhido como o modelo mais adequado por apresentar o **melhor equilíbrio entre sensibilidade e especificidade, alta acurácia global e maior confiabilidade na classificação** dos diferentes tipos de solo e vegetação.

Questão 2 Estimativa de Volumes de Árvores

Com base nos resultados obtidos, o modelo de **Random Forest (RF)** se destaca como o mais adequado para a estimativa do volume de árvores. Ele apresentou o maior coeficiente de determinação (**$R^2 = 0,8535647$**), o que revela um alto poder explicativo. Além disso, o erro padrão da estimativa (**$S_{yx} = 0,1445527$**) foi o menor entre os modelos analisados, indicando maior precisão nas previsões.

O modelo de **Máquinas de Vetores de Suporte (SVM)** também demonstrou bom desempenho (**$R^2 = 0,8484652$**), com uma diferença relativamente pequena em relação ao RF. No entanto, a **robustez e flexibilidade** do Random Forest frente a variações nos dados e ruídos o tornam mais confiável para aplicações práticas.

Em contraste, a **Rede Neural Artificial (RNA)** obteve um **R^2 negativo**, o que evidencia um



desempenho insatisfatório, inclusive pior do que simplesmente utilizar a média como predição. Esse resultado sugere que o modelo de RNA não conseguiu captar adequadamente os padrões nos dados utilizados.

Já o **modelo alométrico tradicional** mostrou um desempenho razoável, mas ainda inferior ao do RF, o que reforça a superioridade dos métodos de aprendizado de máquina mais avançados nesse tipo de tarefa.

Dessa forma, considerando tanto a **acurácia (R^2)** quanto a **precisão (Syx)**, além da **robustez do modelo**, o **Random Forest** se configura como a melhor escolha para a estimativa de volume de árvores neste estudo.