

**MAPAS PEDOLÓGICOS DETALHADOS PARA APRENDIZADO DE  
MÁQUINA NO MAPEAMENTO DIGITAL DE SOLOS DA BACIA DO RIO  
JAÚ EM NÍVEL DE RECONHECIMENTO**

Fabiano de Jesus Ribeiro <sup>1</sup>; Ricardo Marques Coelho <sup>2</sup>

1 Discente do Programa de Pós-Graduação em Agricultura Tropical e Subtropical (Gestão de Recursos Agroambientais) nível de mestrado/doutorado. E-mail: fabiano.ribeiro.floretal@gmail.com

2 Orientador, Centro de Conservação de Solos, Campinas, São Paulo

A gestão dos recursos hídricos superficiais e subterrâneos demanda o conhecimento das propriedades e da distribuição dos diferentes tipos solos presentes nas bacias hidrográficas. Os mapas pedológicos em escala 1:100.000 existentes para o estado São Paulo apresentam estas informações dos solos, são compatíveis com o planejamento em bacias hidrográficas, mas cobrem apenas 15 % do estado. Os mapas pedológicos digitais (MPD) obtidos por meio de mineração de dados apresentam grande potencial para ampliar a área do estado de São Paulo coberta com mapas pedológicos em escala 1:100.000. Durante a mineração, algoritmos de aprendizado identificam relações entre variáveis preditivas e variável meta (que se pretende predizer). Para geração do modelo de predição em MPD, variáveis ambientais, geralmente relacionadas ao relevo e ao material de origem são utilizadas como preditoras, enquanto classes de solo (ou atributos) são utilizadas como variável meta. Uma vez construído, o modelo pode ser extrapolado para predizer classes de solos em áreas fisiograficamente semelhantes. A falta de informações de solo passíveis de serem utilizadas como variável meta, se apresenta como uma barreira para o emprego dessa abordagem. O objetivo do presente trabalho consiste na elaboração de mapa pedológico digital para a Bacia do Rio Jaú em nível de reconhecimento utilizando, para o aprendizado de máquina, mapas de solo detalhados elaborados de forma descontínua em área de lavoura de cana-de-açúcar. Para obtenção do modelo de predição foram adotadas como preditores as variáveis geomorfométricas obtidas do modelo digital de elevação hidrologicamente consistente, as formações geológicas do mapa geológico e as informações geomorfológicas (morfoescultura e modelado de relevo). Os mapas de solos detalhados foram utilizados como variável meta.

Por se tratarem de documentos para atender ao setor sucroenergético, a legendas precisou ser adequada ao Sistema Brasileiro de Classificação de Solo. O conjunto de dados (variáveis preditivas e variável meta) foi submetido a três algoritmos de aprendizado: J48, *Random Forest* e *Naive Bayes*. A validação cruzada mostrou o modelo gerado pelo *Random Forest* como o mais eficiente em prever classes de solo, apresentando taxa de acerto igual 89.8%, sendo então, o escolhido para predição de classes de solos nas áreas não mapeadas localizadas no interior da bacia, gerando o mapa pedológico digital acrescido de informação de textura. Comprovada a capacidade de predição do modelo, a próxima etapa será a validação do mapa digital por verdade de campo.

**Palavras-chave:** mapa pedológico digital, mineração de dados, pedometria.

**Agradecimentos:** À CAPES, pela concessão da bolsa, e ao Instituto Agrônômico de Campinas, pela oportunidade.

