

Ciência da Computação / Sistemas de Informação

Geração de Dados Sintéticos com SMOTE para Predição do Potencial Hídrico em Cafeeiros

Ana Clara Carvalho Nascimento - 6 módulo de Ciência da Computação, UFLA, iniciação científica voluntária.

Meline de Oliveira Santo - EPAMIG

Vânia Aparecida Silva - EPAMIG

Margarete Marin Lordelo Volpato - EPAMIG

Marluce Rodrigues Pereira - Orientador DCA, UFLA. - Orientador(a)

Resumo

O manejo hídrico adequado é essencial para a produtividade e a qualidade do café. O potencial hídrico (Ψ_w) é um dos principais indicadores do estado fisiológico das plantas. Sua medição tradicional, por meio da câmara de pressão de Scholander, é precisa, mas trabalhosa, cara e pouco escalável, resultando em bases de dados pequenas e desbalanceadas. Este estudo teve como objetivo avaliar o impacto da técnica SMOTE (Synthetic Minority Over-sampling Technique) na geração de dados sintéticos para posteriormente serem processados por algoritmos de aprendizado de máquina e prever o Ψ_w a partir dos índices foliares. O banco de dados inicial foi obtido por meio de espectrômetro portátil, que forneceu diversos índices espectrais, e por medições de Ψ_w com a câmara de pressão. Inicialmente, a variável contínua Ψ_w foi categorizada em três classes: alta (maior igual $0,5$ MPa), média ($1,5 < \Psi_w < 0,5$ MPa) e baixa (menor igual $1,5$ MPa), definidas com base em critérios agrônômicos da EPAMIG. Em seguida, aplicou-se a técnica SMOTE para balancear o conjunto de treinamento, gerando exemplos sintéticos das classes minoritárias. Foram treinados os algoritmos Naive Bayes, Decision Tree, Random Forest e suas versões otimizadas com GridSearch, sendo avaliados por acurácia, precisão, recall e F1-score usando dados de teste separados. Os resultados mostraram que o efeito do SMOTE variou conforme o algoritmo. O maior ganho foi observado na Decision Tree com GridSearch, cuja acurácia aumentou para cerca de 72% após o balanceamento. Em contrapartida, a Random Forest apresentou melhor desempenho sem o SMOTE, evidenciando que métodos de conjunto baseados em árvores são menos sensíveis ao desbalanceamento. Conclui-se que o SMOTE pode melhorar o desempenho de algoritmos mais sensíveis ao desbalanceamento, como árvores únicas, mas não é vantajoso para todos os modelos. Como trabalhos futuros, pretende-se ampliar o estudo, aplicando o SMOTE em novas bases de dados de áreas cafeeiras obtidas por drones e satélites e aplicando outros algoritmos. Desse modo, os resultados de previsão de Ψ_w serão mais precisos para dar suporte na tomada de decisão de produtores de café.

Palavras-Chave: potencial hídrico, smote, aprendizado de máquina.

Instituição de Fomento: FAPEMIG, Consórcio Pesquisa Café, CNPq

Link do pitch: <https://youtu.be/3qvU3Is3YIA>