



Análise Comparativa da Acurácia dos Algoritmos Random Forest e Support Vector Machine na Classificação Supervisionada de Cobertura do Solo Utilizando o Google Earth Engine.

Modalidade: Pesquisa | Grande área: Ciências Humanas e Sociais | Área temática: Geografia

Matheus Guimarães Cancela Silva - Graduando em Geografia - UFV | matheus.cancela@ufv.br
André Luiz Lopes de Faria - Professor do departamento de Geografia (orientador) - UFV | andre@ufv.br

Palavras-chave: Classificação de cobertura da terra, Google Earth Engine, Análise comparativa

Introdução

A classificação supervisionada de cobertura da terra desempenha um importante papel na análise e monitoramento da superfície terrestre. Nesse contexto, algoritmos de aprendizado de máquina (*machine learning*) têm sido amplamente utilizados para obter resultados mais precisos e confiáveis, juntamente com plataformas de sensoriamento remoto e geoprocessamento.

Dentre as possíveis ferramentas disponíveis, destaca-se o *Google Earth Engine* (GEE). Plataforma focada em análises espaciais e que disponibiliza múltiplas ferramentas integradas, dentre elas, os algoritmos *Random Forest* (RF) e *Support Vector Machine* (SVM).

Objetivos

O objetivo deste estudo foi o de realizar uma análise comparativa da acurácia e qualidade visual entre os algoritmos *Random Forest* e *Support Vector Machine*, aplicados à classificação supervisionada de cobertura do solo de imagens do Sentinel-2 com a utilização da plataforma *Google Earth Engine*.

Como principais objetivos do estudo, destacam-se:

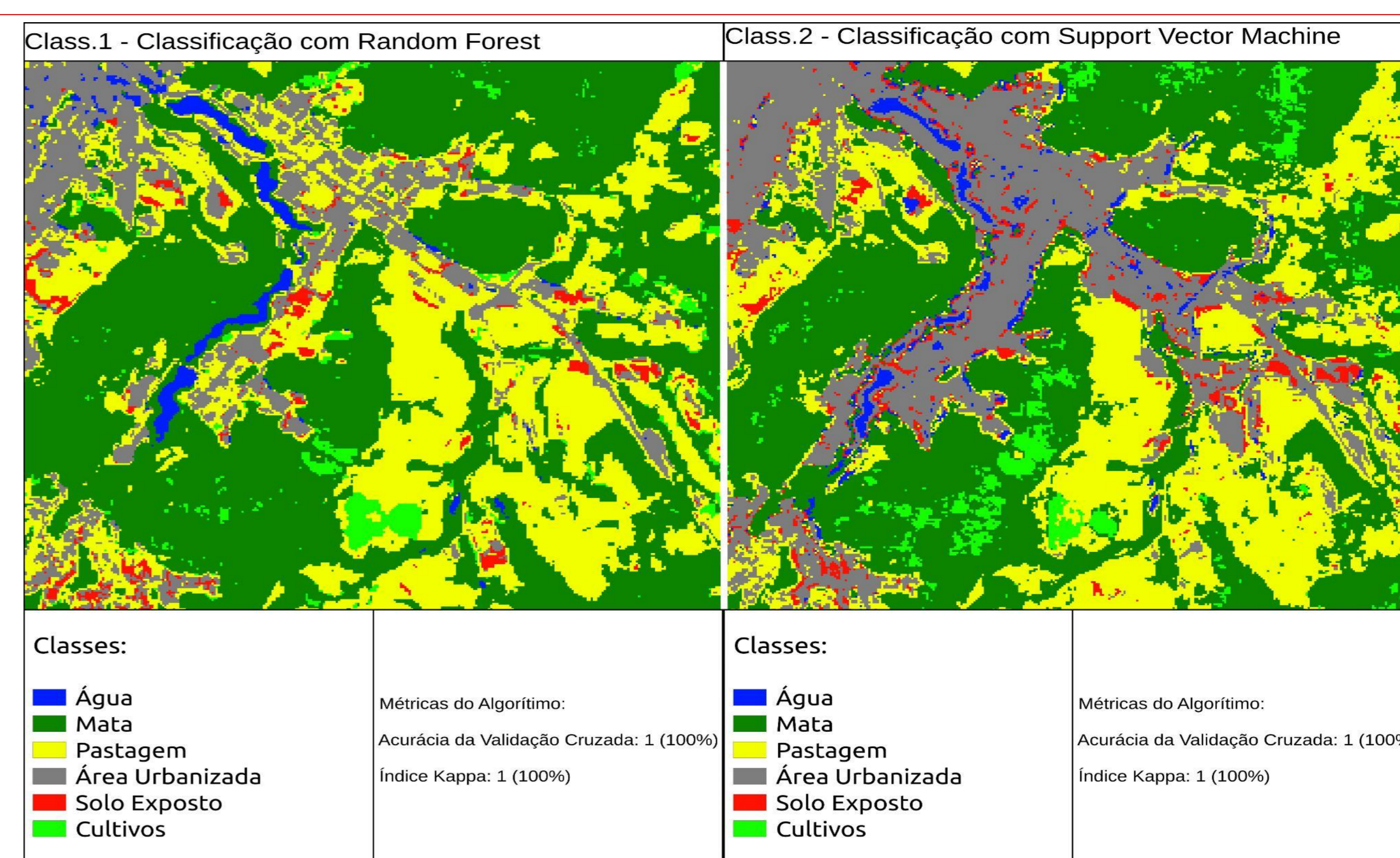
- Definir os melhores parâmetros para cada classificador.
- Aplicar métodos de validação e acurácias básicas.
- Analisar as métricas de cada modelo, juntamente com a qualidade visual apresentada.

Material e Método

No estudo, foi empregada uma metodologia que envolveu a coleta de amostras supervisionadas para treinamento de algoritmos de classificação. Os parâmetros dos algoritmos foram ajustados para otimizar o desempenho, incluindo análises visuais e estatísticas, juntamente com validação cruzada, onde 70% das amostras foram usadas para treinamento e 30% para validação. Os dados de imagens de satélite do Sentinel-2 no GEE foram utilizados para a análise comparativa, que se concentrou em uma área de interesse em Viçosa-MG.

Inicialmente, foi realizado um levantamento da área de interesse para coletar amostras estratificadas em seis classes de cobertura da terra. Os dados foram divididos em conjuntos de treinamento e validação para garantir a representatividade em todas as classes e avaliar a capacidade de generalização dos modelos. Além disso, os parâmetros específicos de cada algoritmo foram configurados, como o número de árvores para o RF e a seleção do *kernel* e parâmetros de regularização para o SVM. Com os modelos treinados, a classificação da cobertura da terra na área de interesse foi realizada.

Resultados e Discussão



Como ilustrado na imagem acima, ambos os modelos apresentaram ótimas acurácias a partir da validação cruzada, chegando a uma pontuação igual a 1 nas duas métricas utilizadas. O que na prática indica que, os 30% amostrais utilizados para validar, foram corretamente generalizados pelos 70% utilizados no treinamento.

Como já indicado, além das análises estatísticas, também se utilizou de critérios visuais, e nestes, foi possível observar uma diferença relevante entre os modelos, sendo o *Random Forest* mais eficiente na identificação das classes e suas respectivas proporções. Já o modelo de *Support Vector Machine* apresentou boas generalizações, mas superestimou a área percentual de muitas delas, principalmente as classes de áreas urbanizadas e solo exposto.

Conclusões

Esta análise demonstra que, apenas métricas básicas e estatísticas de uma modelagem de dados, podem não ser o suficiente para apurar sua real assertividade, sendo necessário modelos de validação mais confiáveis além da validação cruzada simples.

É imperativo observar que nem todas as modelagens com *Machine Learning* envolvem dados espaciais que podem ser contrapostos com a realidade de forma instantânea, portanto, nesses casos, métricas confiáveis podem ser ainda mais importantes.

Bibliografia

- HEUNG, Brandon et al. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, v. 265, p. 62-77, 2016.
- AHMED, Nesreen K. et al. An empirical comparison of machine learning models for time series forecasting. *Econometric reviews*, v. 29, n. 5-6, p. 594-621, 2010.
- OSISANWO, F. Y. et al. Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, v. 48, n. 3, p. 128-138, 2017.