

**INSTITUTO FEDERAL DE EDUCAÇÃO CIÊNCIA E TECNOLOGIA DE GOIÁS
CAMPUS GOIÂNIA
DEPARTAMENTO ACADÊMICO DA ÁREA III
COORDENAÇÃO DE ENGENHARIA CARTOGRÁFICA E DE AGRIMENSURA**

ALESSANDRA CRISTINA PEREIRA

**SIG E *MACHINE LEARNING* APLICADOS A PREDIÇÃO DE SÍTIOS
ARQUEOLÓGICOS EM SERRANÓPOLIS, GOIÁS, BRASIL**

Goiânia, GO

2022

ALESSANDRA CRISTINA PEREIRA

**SIG E *MACHINE LEARNING* APLICADOS A PREDIÇÃO DE SÍTIOS
ARQUEOLÓGICOS EM SERRANÓPOLIS, GOIÁS, BRASIL**

Trabalho de Conclusão de Curso apresentado a Coordenação de Geomática, do Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Campus Goiânia, como parte dos requisitos necessários para a obtenção do título de Engenharia Cartográfica e de Agrimensura.

Orientador: Prof. Dr. Édipo Henrique Cremon

Coorientadora: Prof.^a Dra. Rosiclér Theodoro da Silva

Goiânia, GO

2022

P4141s Pereira, Alessandra Cristina.

SIG e *Machine Learning* aplicados a predição de sítios Arqueológicos em Serranópolis, Goiás, Brasil / Alessandra Cristina Pereira. – Goiânia: Instituto Federal de Educação, Ciência e Tecnologia de Goiás, 2022.
65 f.: il.

Orientador: Prof. Dr. Édipo Henrique Cremon.

Coorientadora: Prof.^a Dra. Rosiclér Theodoro da Silva.

TCC (Trabalho de Conclusão de Curso) – Curso de Engenharia Cartográfica e de Agrimensura, Departamento de Áreas Acadêmicas III, Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

1. Geotecnologias. 2. Inteligência Artificial. 3. Sistema de Informações Geográficas.

I. Cremon, Édipo Henrique (orientador). II. Silva, Rosiclér Theodoro da (coorientadora).

III. Instituto Federal de Educação, Ciência e Tecnologia de Goiás. IV. Título.

CDD 526.1285

Ficha catalográfica elaborada pela Bibliotecária Karol Almeida da Silva CRB1/ 2.740
Biblioteca Professor Jorge Félix de Souza,
Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia.

**TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO
NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG**

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

Identificação da Produção Técnico-Científica

- | | |
|--|---|
| ☐ Tese | ☐ Artigo Científico |
| ☐ Dissertação | ☐ Capítulo de Livro |
| ☐ Monografia – Especialização | ☐ Livro |
| ☒ TCC - Graduação | ☐ Trabalho Apresentado em Evento |
| ☐ Produto Técnico e Educacional - Tipo: _____ | |

Nome Completo do Autor: Alessandra Cristina Pereira

Matrícula: 20162011190030

Título do Trabalho: SIG e *Machine Learning* aplicados a predição de sítios arqueológicos em Serranópolis, Goiás, Brasil

Autorização - Marque uma das opções

- ☐ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso aberto);
- ☒ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data 31/12/2022 (Embargo);
- ☐ Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso restrito).

Ao indicar a opção **2 ou 3**, marque a justificativa:

- ☐ O documento está sujeito a registro de patente.
☒ O documento pode vir a ser publicado como livro, capítulo de livro ou artigo.
☐ Outra justificativa: _____

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

- o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

Goiânia, 07/02/2022.
Local Data



Assinatura da Autora

TERMO DE APROVAÇÃO



INSTITUTO FEDERAL
Goiás

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
CÂMPUS GOIÂNIA

ATA Nº ECA 04 - TCC/2021.2

ENGENHARIA CARTOGRÁFICA E DE AGRIMENSURA / Câmpus Goiânia

ATA DA SESSÃO PÚBLICA VIRTUAL DE APRESENTAÇÃO E DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

Aos sete dias do mês de fevereiro do ano de dois mil e vinte e dois, às 13:00 horas, na sala virtual meet.google.com/edr-dbwv-msp, foi realizada a sessão pública de apresentação e defesa do Trabalho de Conclusão de Curso da Graduanda **Alessandra Cristina Pereira (matricula 20162011190030)** do curso de Bacharelado em Engenharia Cartográfica e de Agrimensura, do Departamento III, Câmpus Goiânia, do Instituto Federal de Goiás, no segundo semestre do ano de dois mil e vinte e um. A banca foi composta pelos seguintes membros:

- Édipo Henrique Cremon (Professor-Orientador)
- Rosiclér Theodoro da Silva (Professora Coorientadora)
- Nilton Ricetti Xavier de Nazareno
- Max Well de Oliveira Rabelo

sob a presidência da(o) primeira(o). O Trabalho de Conclusão de Curso tem como título **"SIG e machine learning aplicados a predição de sítios arqueológicos em Serranópolis, Goiás, Brasil"**, sob orientação do Prof. **Édipo Henrique Cremon** e Coorientação da Profa. Dra. **Rosiclér Theodoro da Silva**. Após a apresentação do Trabalho de Conclusão de Curso, tendo sido a autora arguida pela Banca Examinadora, a nota obtida foi 9,9 (nove vírgula nove).

Encerrou-se a sessão às 14 horas e 40 minutos. Eu, Profa. Marina Alberti Macedo, dato e assino a presente ata que segue assinada por todos os membros da Banca Examinadora.

Profa. Marina Alberti Macedo (Assinado eletronicamente)

Prof. Édipo Henrique Cremon (Assinado eletronicamente)

Profa. Rosiclér Theodoro da Silva (Assinado eletronicamente)

Prof. Nilton Ricetti Xavier de Nazareno (Assinado eletronicamente)

Prof. Max Well de Oliveira Rabelo (Assinado eletronicamente)

Documento assinado eletronicamente por:

- Rosiclér Theodoro da Silva, ROSIKLER THEODORO DA SILVA - DOCENTE - PUC / GOIAS (81587669980171), em 07/02/2022 20:10:13.
- Max Well de Oliveira Rabelo, PROFESSOR ENS BASICO TECH TECNOLÓGICO, em 07/02/2022 14:31:13.
- Nilton Ricetti Xavier de Nazareno, PROFESSOR ENS BASICO TECH TECNOLÓGICO, em 07/02/2022 14:30:45.
- Édipo Henrique Cremon, PROFESSOR ENS BASICO TECH TECNOLÓGICO, em 07/02/2022 14:30:21.

Este documento foi emitido pelo SIGAP em 05/02/2022. Para comprovar sua autenticidade, leia a leitura do QRCode ao lado ou acesse <https://sigap.ifg.edu.br/autenticar/> documento e forneça as datas abaixo:

Código Verificador: 243210

Código de Autenticação: c05E073a8f



AGRADECIMENTO

Agradeço a minha família pelo incentivo e apoio, pelas orações, pela paciência e compreensão as minhas ausências, em especial ao meu marido, companheiro no dia a dia.

Ao professor Dr. Édipo Henrique Cremon, meu orientador, inicialmente por ter aceitado o convite e ter embarcado nessa jornada comigo, por dedicar seu tempo a me orientar em todas etapas deste trabalho, por todas as contribuições, correções e revisões.

A professora Dra. Rosiclér Theodoro da Silva e ao professor Dr. Júlio Cezar Rubin de Rubin, meus coorientadores, um oficial e outro extraoficial, por também ter aceitado essa missão e por colaborar com as experiências quanto as questões arqueológicas. Pelo tempo dedicado a mim, pelas correções e sobretudo pela amizade ao longo desses anos.

Aos docentes do Instituto Federal de Educação, Ciência e Tecnologia de Goiás, em especial aos do curso de Engenharia Cartográfica e de Agrimensura pelos conhecimentos repassados durante todo o curso, dentro e fora da sala de aula.

Aos colegas e amigos do IFG pela amizade nesses cinco anos, que fizeram dos momentos mais difíceis os mais “interessantes”.

Ao Instituto Goiano de Pré-História e Antropologia, nas pessoas dos pesquisadores e meus coorientadores, Rosiclér Theodoro e Júlio Rubin pela disponibilização dos dados e informações utilizados neste trabalho.

Ao Instituto do Patrimônio Histórico e Artístico Nacional, nas pessoas de seu superintendente do IPHAN-GO, Allyson Ribeiro e Silva Cabral e da coordenadora técnica substituta Ana Paula Carvalho, pela disponibilização dos dados dos sítios arqueológicos de Serranópolis.

Por fim, um agradecimento a todas as pessoas, que de algum modo, contribuíram para a concretização de mais uma etapa da minha vida.

RESUMO

SIG E MACHINE LEARNING APLICADOS A PREDIÇÃO DE SÍTIOS ARQUEOLÓGICOS EM SERRANÓPOLIS, GOIÁS, BRASIL

AUTORA: Alessandra Cristina Pereira

ORIENTADOR: Prof. Dr. Édipo Henrique Cremon

COORIENTADORA: Prof.^a Dra. Rosiclér Theodoro da Silva

O uso de geotecnologias para fins arqueológicos, vem se popularizando nas últimas décadas. Modelos preditivos para localização de sítios arqueológicos a partir de dados geográficos, têm sido beneficiados por essa tecnologia, em especial o Sistema de Informação Geográfica (SIG) e aprendizado de máquina (*machine learning*). Neste contexto, o objetivo deste trabalho foi prever locais de sítios arqueológicos, por meio de modelos preditivos utilizando SIG e *machine learning*, na região de Serranópolis, Goiás, por possuir um dos mais importantes complexos arqueológicos do Brasil. Foi utilizado, 54 pontos amostrais de presença de sítios arqueológicos e 60 amostras aleatórias de pseudoausências de sítios, além de 18 covariáveis topográficas derivadas diretamente de modelo digital de elevação COP-30. A predição foi realizada por três algoritmos de classificação supervisionada baseados em árvore de decisão: *Random Forest* (RF), C5.0 e *Extreme Gradient Boosting* (XGBoost), considerando três cenários de distribuição aleatória de pontos de pseudoausência. Estes algoritmos foram aplicados seguindo a validação cruzada do tipo *k-fold* para o treinamento e calibração dos modelos com 70% do conjunto amostral e 30% foi separado para teste (validação com amostras independentes). Para aferir a confiabilidade dos resultados da classificação, foram utilizadas como métricas de performance a *Receiver Operator Characteristic Curve* (ROC), a *Area Under the Curve* (AUC), a sensibilidade, a especificidade e *F-score*. Por fim, foram elaborados mapas temáticos dos modelos preditivos. Os principais resultados demonstraram que a curva ROC do RF, em todos os cenários, apresentou a taxa de sensibilidade um pouco melhor que os demais modelos. O RF também apresentou os melhores resultados de valores da área sobre a curva ROC, com AUC de 0,932, o que significa que este algoritmo apresentou melhor habilidade preditiva para esse estudo, dentre os algoritmos analisados. Em relação as métricas da sensibilidade, especificidade e *F-score*, novamente o RF no cenário 1 apresentou o melhor resultado. As covariáveis topográficas mais importantes para a predição de sítios arqueológicos foram associadas a rugosidade, declividade, orientação de vertente, fator de vista para o céu, depressão topográfica, índice de umidade topográfica, amplitude do relevo e índice multirresolução para cristas em áreas elevadas. As métricas de validação constataram que o RF apresentou o melhor resultado, sobretudo no cenário 1, apontando como o modelo de melhor poder preditivo para este trabalho.

Palavras-chave: Inteligência Artificial. *Random Forest*. Arqueologia. Dado geográfico.

ABSTRACT

GIS AND MACHINE LEARNING APPLIED TO THE ARCHAEOLOGICAL SITES PREDICTION IN SERRANÓPOLIS, GOIÁS, BRAZIL

AUTHOR: Alessandra Cristina Pereira

ADVISOR: Dr. Édipo Henrique Cremon

CO- ADVISOR: Dr. Rosiclér Theodoro da Silva

Geotechnologies support for archaeological purposes has become more popular in the last decades. Predictive models for archaeological sites location from geographic data have benefited from this technology, especially the Geographic Information System (GIS) and machine learning methods. In this work the aim was to predict archaeological sites, through predictive models using GIS and machine learning methods, at the Serranópolis City, State of Goiás, as it has one of the most important archaeological complexes in Brazil. 54 presence sample points of archaeological sites and 60 random samples of pseudo-absence was used for prediction, in addition to 18 topographic covariates derived from the COP-30 digital elevation model. The prediction was performed by three supervised classification algorithms based on decision tree: Random Forest (RF), C5.0, and Extreme Gradient Boosting (XGBoost), considering three scenarios of random distribution of pseudo-absence points. These algorithms were applied following k-fold cross-validation for training and calibration of the models with 70% of the dataset and 30% was split for testing (validation with independent samples). Accuracy was evaluated by Receiver Operator Characteristic Curve (ROC), the Area Under the Curve (AUC), the sensitivity, the specificity, and the F-score. Finally, thematic maps of the predictive models were developed. Main results have shown that the RF ROC curve, in all scenarios, presented a slightly better sensitivity rate than the other models. RF also presented the best results of the AUC values of 0.932, which means that this algorithm presented the best predictive ability for this study and among the algorithms analyzed. Regarding the sensitivity, specificity, and F-score metrics, again the RF in scenario 1 presented the best result. The most important topographic covariates for archaeological sites prediction were associated to Vector Ruggedness Measure, Slope, Aspect, Sky View Factor, Topographic Depression, Topographic Wetness Index, Local Relief, and Multiresolution Index of Valley of the Ridge Top Flatness. The validation metrics found that the RF presented the best result, especially in scenario 1, pointing as the model of best predictive power for this work.

Keywords: Artificial intelligence. Random Forest. Archaeology. Geographic data.

LISTA DE FIGURAS

Figura 1 – Exemplo de um dado <i>raster</i>	25
Figura 2 – Exemplo de extração de valores múltiplos para pontos amostrais.	26
Figura 3 – Evolução dos algoritmos baseados em árvore de decisão.	28
Figura 4 – Fases de treinamento e classificação do RF.	29
Figura 5 – Esquematização do algoritmo XGBoost.	31
Figura 6 – Esquematização do método de reamostragem por abordagem de validação cruzada <i>k-fold</i>	32
Figura 7 – Exemplo de curvas ROC.	33
Figura 8 – Mapa de localização da área de estudo – Serranópolis, Goiás.	35
Figura 9 - Fluxograma do procedimento metodológico adotados no estudo.	38
Figura 10 – Mapa hipsométrico com localização geográfica dos sítios arqueológicos.	41
Figura 11 – Mapa de localização dos cenários das pseudoausências e raio de influência dos sítios arqueológicos.	42
Figura 12 – Curva ROC dos três algoritmos, nos três cenários.	46
Figura 13 – Importâncias das covariáveis topográficas para a predição de sítios arqueológicas usando o classificador C5.0, nos três cenários.	48
Figura 14 – Importâncias das covariáveis topográficas para a predição de sítios arqueológicas usando o classificador RF, nos três cenários.	49
Figura 15 – Importâncias das covariáveis topográficas para a predição de sítios arqueológicas usando o classificador XGBoost, nos três cenários.	50
Figura 16 – Gráficos de <i>Violin Plots</i> das variáveis mais preditoras com distribuição dos dados de pseudoausência e presença de sítios arqueológicos.	53
Figura 17 – Mapas preditivos de possíveis locais com sítios arqueológicos no cenário1, para os algoritmos C5.0, RF e XGBoost.	55
Figura 18 – Mapas preditivos de possíveis locais a encontrar sítios arqueológicos no cenário 2, para os algoritmos C5.0, RF e XGBoost.	56
Figura 19 – Mapas preditivos de possíveis locais a encontrar sítios arqueológicos no cenário 3, para os algoritmos C5.0, RF e XGBoost.	57

LISTA DE TABELAS

Tabela 1 – Valores da AUC para os algoritmos C5.0, RF e XGBoost, nos três cenários.	45
Tabela 2 – Valores da taxa de sensibilidade dos algoritmos C5.0, RF e XGBoost, nos três cenários.	46
Tabela 3 – Valores da taxa de especificidades dos algoritmos C5.0, RF e XGBoost, nos três cenários.	47
Tabela 4 – Valores de <i>F-score</i> dos algoritmos C5.0, RF e XGBoost, nos três cenários.	47

LISTA DE QUADROS

Quadro 1 – Parâmetros de especificações do MDE COP-30.	24
Quadro 2 – Atributos topográficos utilizados neste estudo para prever regiões de sítios arqueológicos e suas definições.....	39

LISTA DE ABREVIATURAS E SIGLAS

AP	Antes do Presente
AUC	<i>Receiver Operator Characteristic Curve</i>
CNSA	Cadastro Nacional de Sítios Arqueológicos
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DLR	<i>Deutsches Zentrum für Luft-und Raumfahrt</i>
EMQ	Erro Médio Quadrático
ESA	Agência Espacial Europeia
FAPEG	Fundação de Amparo à Pesquisa do Estado de Goiás
GIS	<i>Geographic Information System</i>
IA	Inteligência Artificial
ICAHM	<i>International Scientific Committee on Archaeological Heritage Management</i>
ICOMOS	<i>International Council on Monuments and Sites</i>
IGPA	Instituto Goiano de Pré-História e Antropologia
InSAR	<i>Synthetic Aperture Radar Interferometry</i>
IPHAN	Instituto do Patrimônio Histórico e Artístico Nacional
LOOCV	<i>Leave-One-Out</i>
MDE	Modelo Digital de Elevação
MDS	Modelo Digital de Superfície
OOB	<i>Out-Of-Bag</i>
PDI	Processamento Digital de Imagens
PUC	Pontifícia Universidade Católica
RF	<i>Random Forest</i>
ROC	<i>Receiver Operator Characteristic Curve</i>
SAGA	<i>System for Automated Geoscientific Analyzes</i>
SIG	Sistema de Informação Geográfica
UTM	<i>Universal Transversa Mercator coordinate</i>
XGBoost	<i>Extreme Gradient Boosting</i>

SUMÁRIO

1.	INTRODUÇÃO.....	14
2.	OBJETIVOS.....	16
2.1.	OBJETIVO GERAL	16
2.2.	OBJETIVOS ESPECÍFICOS	16
3.	REVISÃO DA LITERATURA.....	17
3.1.	GEOARQUEOLOGIA	17
3.1.1.	Patrimônio arqueológico	18
3.1.2.	Estudos arqueológicos na região de Serranópolis.....	19
3.1.3.	Sinergismo entre as covariáveis ambientais e sítios arqueológicos	19
3.2.	GEOPROCESSAMENTO.....	21
3.2.1.	Importância do geoprocessamento nos estudos arqueológicos.....	22
3.2.2.	Modelo Digital de Elevação (MDE)	23
3.2.3.	Atributos topográficos.....	24
3.2.4.	Machine Learning e algoritmos de classificação	26
3.2.5.	Treinamento e teste da classificação	31
4.	LOCALIZAÇÃO E ASPECTOS GERAIS DA ÁREA	35
5.	MATERIAL E MÉTODOS.....	37
5.1.	MATERIAL.....	37
5.2.	PROCEDIMENTOS METODOLÓGICOS	37
5.2.1.	Aquisição e processamento dos atributos topográficos	38
5.2.2.	Aquisição dos dados de sítios arqueológicos.....	40
5.2.3.	Amostras de treinamentos	41
5.2.4.	Classificação, validação e predição	42
6.	RESULTADOS	45
7.	DISCUSSÃO.....	58

8. CONCLUSÃO.....	61
REFERÊNCIAS	62
APÊNDICE A – Quadro de informações, relevantes ao estudo, dos sítios arqueológicos selecionados.....	71
ANEXO A – Ofício nº 717/2021/COTEC IPHAN-GO/IPHAN-GO-IPHAN	74

1. INTRODUÇÃO

O município de Serranópolis, Goiás, possui um dos mais importantes complexos arqueológicos do Brasil. São mais de 30 sítios associados a populações caçadores-coletores e agricultores-ceramistas que ocuparam a região há aproximadamente 11 mil anos Antes do Presente (AP)¹ (IPHAN, 2020a). Rubin *et al.* (2020) destaca que as pesquisas arqueológicas nesta região iniciaram nos anos de 1970 e vêm nas últimas décadas apresentando dados importantes, resultantes de investigações arqueológicas, consistindo assim, uma área com significância histórico-cultural para a compreensão dos processos ocupacionais do Planalto Central Brasileiro.

Dada a importância científica e turística da região, o interesse na localização de sítios arqueológicos, faz que o uso de geotecnologias, em especial o Sistema de Informação Geográfica (SIG), se torne um importante instrumento de apoio, não somente na contextualização geográfica dos sítios arqueológicos, mas na interpretação geral dos elementos constituintes da paisagem sobretudo no contexto geoarqueológico (BITENCOURT, 2008). Logo, os aspectos associados ao meio físico são fundamentais na contextualização do ambiente.

Entender o processo de ocupação das populações humanas do passado e suas relações ambientais traz à discussão o conceito de modelo preditivo arqueológicos, ou seja, prever potenciais locais de sítios em áreas nunca pesquisadas, com base em padrões locais de sítios já conhecidos (SEQUEIRA, 2009). Os modelos preditivos arqueológicos produzidos até o momento, no geral, utilizam-se de uma abordagem metodológica baseada em técnicas estatísticas multivariadas como regressão logística, considerando dados arqueológicos e variáveis ambientais (CASTIELLO; TONINI, 2021). Diante disso, com o avanço da inteligência artificial, em especial de modelos de *Machine Learning* (aprendizado de máquina), modelos preditivos a partir de dados geográficos para fins arqueológicos têm sido beneficiados, a exemplo do trabalho de Castiello e Tonini (2021) que os utilizam para modelagem preditiva de sítios arqueológicos na Suíça. Outro estudo é apresentado por Yaworsky *et al.* (2020), onde é avaliado alguns métodos de *machine learning*, em comparação à modelos de regressão, em estudos arqueológicos no Monumento Nacional *Grand Staircase-Escalante*, nos Estados

¹ Antes do Presente (AP) é uma marcação de tempo utilizada na arqueologia, para cálculo da idade radiocarbônica de uma determinada amostra fóssil pelo método de carbono-14 e tem como base de referência o ano de 1950 no calendário gregoriano.

Unidos. Na Espanha, o estudo apresentado por Stubert *et al.* (2020) faz uso de variáveis topográficas para modelagem preditiva arqueológica. Não obstante, ainda há poucos trabalhos que exploram essa ferramenta para a predição de sítios arqueológicos no Brasil.

Em 2019, pesquisadores do Instituto Goiano de Pré-História e Antropologia (IGPA), vinculado à Pontifícia Universidade Católica de Goiás (PUC Goiás) com apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG) e autorização do Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN), retomaram as pesquisas arqueológicas na região de Serranópolis (IPHAN, 2020a), objetivando identificar possíveis vestígios culturais, sob a perspectiva da geoarqueologia, ampliando as inúmeras discussões sobre a ocupação do Planalto Central Brasileiro. Assim, Serranópolis consiste em um local ideal para a aplicação de técnicas de *machine learning* utilizando dados geográficos para a predição de sítios arqueológicos.

Este trabalho parte do pressuposto que variáveis geomorfométricas associadas ao meio físico possam estabelecer um modelo preditivo que possibilite selecionar potenciais áreas de sítios arqueológicos na região de Serranópolis, empregando inteligência artificial, em especial o *Machine Learning* aliado ao SIG, de forma que possa contribuir para as pesquisas arqueológicas, considerando os poucos trabalhos que exploram essas ferramentas para a predição de sítios arqueológicos.

2. OBJETIVOS

O objetivo geral e os objetivos específicos do presente trabalho serão apresentados a seguir.

2.1. OBJETIVO GERAL

Predizer locais de sítios arqueológicos, utilizando Sistema de Informação Geográfica (SIG) e aprendizado de máquina (*machine learning*) na região de Serranópolis, Goiás.

2.2. OBJETIVOS ESPECÍFICOS

- Analisar a importância das covariáveis geomorfométricas relacionadas aos atributos topográficos da paisagem e do contexto geoarqueológico na predição de sítios arqueológicos;
- Definir o modelo preditivo;
- Verificar a aplicabilidade do modelo preditivo para a região tendo como referência preliminar os sítios já identificados;
- Identificar, por meio das análises estatísticas de validação, a performance do melhor modelo preditivo;
- Avaliar diferentes cenários em função das amostras de treinamento para a predição de locais prováveis de sítios arqueológicos.

3. REVISÃO DA LITERATURA

3.1. GEOARQUEOLOGIA

O termo geoarqueologia pode ser definido como uma inter-relação entre a Arqueologia e as Geociências ou Ciência da Terra. Tal conceito evoluiu e se tornou necessário nas pesquisas arqueológicas nas últimas décadas, uma vez que a interação dos conceitos, métodos e técnicas das geociências aos aspectos paleontológicos, geomorfológicos, geológicos e pedológicos aprimora a análise espacial e temporal do sítio arqueológico no que tange as faces da evolução a compreensão da ocupação humana do passado com o ambiente (BITENCOURT, 2008).

O termo geoarqueologia é definido por diversos autores, entre eles Butzer (1982, p. 35): *“Geo-archaeology implies archaeological research using the methods and concepts of the earth sciences. The term is not synonymous with archaeological geology, and it is not necessarily linked to geology²”*, Gladfelter (1981, p. 343): *“In broadest terms, geoarchaeology deals with earth history within the time frame of human history³”* e Cardoso (1996, p. 70):

Interessam à Geoarqueologia todos os métodos, susceptíveis de aplicação a artefactos ou sítios arqueológicos, de forma a recuperar toda a informação de ordem geológica, petrográfica ou mineralógica que potencialmente possam fornecer.

O fato da arqueologia e das geociências se relacionarem está na razão que ambas atuam na coleta de dados visando a sua interpretação a partir de elementos presentes na superfície ou subsuperfície terrestre (ANGELUCCI, 2003). Bitencourt (2008) ainda destaca que os diversos métodos das geociências são utilizados na pesquisa arqueológica nas mais abrangentes escalas de observação desde análise espacial envolvendo sensoriamento remoto até análise de solos arqueológicos, o que torna um importante instrumento de apoio na caracterização dos elementos da paisagem.

² Tradução da autora: A geoarqueologia é a pesquisa arqueológica que utiliza os métodos e conceitos das ciências da Terra. O termo não é sinônimo de geologia arqueológica, e não está necessariamente ligado à geologia.

³ Tradução da autora: Em termos mais amplos, a geoarqueologia lida com a história da Terra dentro do período de tempo da história humana.

3.1.1. Patrimônio arqueológico

O termo sítio arqueológico é usado para definir lugares onde se encontram vestígios de atividades humana que ali viveram. Os vestígios podem ser o esqueleto do indivíduo evidenciando sua existência, ou objetos, tais como ferramentas e utensílios feitos de pedra, madeira, osso, cerâmica; restos de fogueira e de alimentos; remanescentes de residência; pinturas rupestres, entre outros. Tais vestígios são úteis, não somente para reconstruir o modo de vida, mas para conhecimento da sociedade e o valor cultural e histórico deixado (LIMA, 2016).

No que compete ao patrimônio arqueológico brasileiro, este está sujeito a um regime jurídico especial, consolidado por princípios e estabelecidos também por convenções internacionais, como a Carta de Lausanne, recomendado pelo ICOMOS/ICAHM⁴ em 1990:

O patrimônio arqueológico engloba todos os vestígios da existência humana e interessa a todos os lugares onde há indícios de atividades humanas, não importando quais sejam elas; estruturas e vestígios abandonados de todo o tipo, na superfície, no subsolo ou sob as águas, assim como o material associado (Carta de Lausanne, art. 1º, In: IPHAN, 2020b).

A Carta de Lausanne, destaca que a proteção do patrimônio arqueológico é de obrigação moral de todos, pública coletiva, de caráter frágil, não renovável e pertence a toda humanidade. Sua proteção, contudo, deve ser por meio de uma legislação adequada e de forma eficaz de preservação (LIMA, 2016). No Brasil, a legislação de acesso e proteção ao patrimônio cultural e, por conseguinte, o arqueológico está também sedimentada na Constituição Federal de 1988, a qual determina o artigo 216 que:

Constituem patrimônio cultural brasileiro os bens de natureza material e imaterial, tomados individualmente ou em conjunto, portadores de referência à identidade, à ação, à memória dos diferentes grupos formadores da sociedade brasileira, nos quais se incluem: I—as formas de expressão; II—os modos de criar, fazer e viver; III—as criações científicas, artísticas e tecnológicas; IV—as obras, objetos, documentos, edificações e demais espaços destinados às manifestações artístico-culturais; V—os conjuntos urbanos e sítios de valor histórico, paisagístico, artístico, arqueológico, paleontológico, ecológico e científico (BRASIL, 1988, art. 216).

⁴ Conselho Internacional de Monumentos e Sítios (ICOMOS) é um organismo não governamental, consultor do Comitê do Patrimônio Mundial da UNESCO. Sua missão é promover a conservação, a proteção, o uso e a valorização de monumentos, centros urbanos e sítios. O Comitê Científico Internacional sobre Gestão do Patrimônio Arqueológico (ICAHM) assessora o ICOMOS e o Comitê do Patrimônio Mundial em questões que dizem respeito a todos os aspectos da gestão de sítios arqueológicos e paisagens. Isso inclui a formulação e propagação de padrões e melhores práticas para pesquisa arqueológica e gestão de recursos culturais.

3.1.2. Estudos arqueológicos na região de Serranópolis

O interesse arqueológico pela região de Serranópolis, no estado de Goiás, decorre da década de 1970, quando descobertos os primeiros indícios de ocupação humana em abrigos rochosos e a céu aberto. Os primeiros resultados das escavações apontaram para uma ocupação humana que data de 11.000 anos AP até meados do período colonial. Para os pesquisadores, essa região se destaca pelo seu importante papel nos estudos da chegada do homem e das populações caçadoras do período Holoceno e por compor um relevante patrimônio arqueológico do continente americano (NOGUEIRA, 2015). De acordo com Schmitz *et al.* (1989, p. 9):

A região de Serranópolis é um lugar privilegiado pela convergência de uma grande extensão de espaços abrigados, pela proximidade de ambientes diferenciados e pela abundância de matéria-prima útil para produção de artefatos e utensílios, condições que possibilitam uma ocupação constante, transformando o local em um centro de referência para populações caçadoras, coletoras e cultivadoras incipientes, estas ainda fortemente dependentes de caça e coleta [...] as abundantes pinturas e gravuras nos abrigos marcaram o espaço da concentração das populações e o sinalizam com os símbolos e a história do grupo.

Ainda segundo Schmitz (1984), a ocupação espacial de Serranópolis possui referência na história do povoamento central do Brasil, pois a mesma região fora ocupada sistematicamente, abrigando inicialmente gerações de caçadores e em seguida agricultores-ceramistas, vindos das outras regiões, constituindo ali suas aldeias.

3.1.3. Sinergismo entre as covariáveis ambientais e sítios arqueológicos

Sequeira (2009) afirma que para alguns arqueólogos, as variáveis culturais, sociais religiosos de cada população, fazem parte da criação de padrões de ocupação territorial e que modelos preditivos de sítios arqueológicos, baseiam-se somente em três pressupostos:

1. a escolha do local foi influenciada por elementos ambientais, naturais e físicos favoráveis a fixação humana, criando um padrão espacial específico;
2. as variáveis ambientais que condicionaram a fixação humana encontram-se, directa e indirectamente, representada na paisagem actual;
3. sítios que se localizam numa mesma área de características ambientais (físicas e naturais) análogas têm tendência de igual forma do meio envolvente (SEQUEIRA, 2009, p. 26)

Embora as variáveis ambientais não sejam os únicos fatores determinísticos, elas, na sua maior parte, influenciam a escolha do espaço para fixação de áreas residenciais ou logísticas. Entre as variáveis ambientais mais frequentes destacam-se a geologia, pedologia, geomorfologia, vegetação, hidrografia, dentre outros. Para a arqueologia, tais elementos são fundamentais, pois os grupos humanos selecionavam locais para assentamento, considerando a proximidade aos recursos e local seguro (THOMAZ *et al.*, 2002).

Dentre as Ciências da Terra que tem relação com a arqueologia, destaca o conceito de geologia arqueológica, como geologia realizada com vistas a uma aplicação arqueológica. Esse termo está implícito no conceito amplo da Geoarqueologia, pois, ambos os campos atuam na interface entre a arqueologia e geologia. Os vestígios de atividade humana pretérita estão contidos num contexto essencialmente geológico e sua recuperação arqueológica e a análise desse material envolvem procedimentos geoarqueológicos (ARAÚJO, 2018).

A gênese dos solos além de possuir características específicas do material de origem, apresenta variáveis que expressam os efeitos de fatores ambientais, tais como clima, organismo, relevo e tempo, ou seja, o solo se desenvolve pela ação integrada dos agentes climáticos (intemperismo) e organismos, condicionados pelo relevo. As ações destes fatores ambientais podem resultar em modificações e corresponder a uma sequência de eventos ou a um complexo de reações e/ou disposições de materiais, provocando alterações que se refletem nas características morfológicas dos solos (PALMIERI; LARACH, 2010; CREMON *et al.*, 2021). Ainda segundo os autores, as unidades de relevo têm uma inter-relação direta com as características das classes do solo e com os aspectos fitofisionômicos da vegetação natural.

Diante disso, na arqueologia o estudo dos solos está associado, em parte, com a ocupação humana na região, uma vez que tal ocupação pode alterar as propriedades físicas e química do solo em razão de depósitos de rejeitos, como fragmentos de artefatos cerâmicos e líticos e, restos de fauna e flora (KÄMPF e KERN, 2005).

Em vista dessas afirmações, os processos de formação e transformação da geologia e dos solos fornecem elementos que possibilitam definir determinadas características da ocupação e a interpretação do lugar onde os sítios estão inseridos.

O relevo, no processo de ocupação humana, exerceu um papel importante, uma vez que barreiras naturais, ambientes inóspitos, terrenos acidentados e planícies, influenciaram na escolha do local de fixação das populações (GARCIA, 2015). O autor ainda considera que a heterogeneidade da superfície terrestre, apresenta variedades em escala espaciais e temporais, logo é possível distinguir as diversas formas que constituem o relevo por meio da

geomorfologia, cujo objetivo é entender os processos e fenômenos de transformação da paisagem.

Conforme pode ser observado, há uma certa relação entre os sítios arqueológicos com as variáveis geológica, pedológica e geomorfológica. Assim como há forte relação entre variáveis geomorfométricas derivadas de modelos digitais de elevação com a geologia (ANDRADES FILHO *et al.*, 2014; VALERIANO; ROSSETTI, 2017), pedologia (CREMON *et al.*, 2021) e geomorfologia (CAMARGO *et al.*, 2012). Por isso, neste estudo é explorado o uso de variáveis geomorfométricas como covariáveis para a predição de sítios arqueológicos.

3.2. GEOPROCESSAMENTO

Conceitualmente, geoprocessamento é um conjunto de ferramentas computacionais voltadas para um determinado objetivo específico por meio de coleta e tratamento das informações espaciais. Dentre essas ferramentas, está o Sistema de Informação Geográfica (SIG) ou *Geographic Information System* (GIS), que permite não somente criar um banco de dados georreferenciado a partir da integração de fontes diversas, como também possibilita automatizar a produção cartográfica (CÂMARA; DAVIS, 2004).

Câmara e Davis (2004), afirmam “Se **onde** é importante para seu negócio, então Geoprocessamento é sua ferramenta de trabalho”. Neste sentido, no planejamento ambiental e na gestão territorial, o geoprocessamento é utilizado para tratamento de grandes volumes de dados, sendo disponibilizados variáveis geotopológicas para fins de análises e utilização imediata. Logo, em especial o SIG, consiste em uma poderosa ferramenta de trabalho na área da geoarqueologia, uma vez que se trabalha com os sítios arqueológicos georreferenciados.

Silva (2009, p. 42), considera ainda o geoprocessamento como:

Um conjunto de conceitos, métodos e técnicas que, atuando sobre bases de dados georreferenciados, por computação eletrônica, propicia a geração de análises e sínteses que consideram, conjugadamente, as propriedades intrínsecas e geotopológicas dos eventos e entidades identificados, criando informação relevante para apoio à decisão quanto aos recursos ambientais.

3.2.1. Importância do geoprocessamento nos estudos arqueológicos

Câmara e Monteiro (2004) afirmam que “geoprocessamento é uma tecnologia interdisciplinar, que permite a convergência de diferentes disciplinas científicas para o estudo de fenômenos ambientais e urbanos”. Nessa perspectiva, diversos campos científicos como a Geografia, Geologia, entre outros, têm-se intensificado no uso de ferramentas ligadas ao SIG para atividades de planejamento e gestão territorial.

Segundo Nazareno (2005), os estudos arqueológicos são realizados por meio de procedimentos técnicos e metodológicos já consagrados utilizando a interdisciplinaridade de outras ciências e cada vez mais vem beneficiando-se do uso de recursos tecnológicos.

Na arqueologia, o SIG vem sendo utilizado desde a década de 1960, e considerando o aumento dos trabalhos científicos relativos à temática de geoprocessamento aplicado à arqueologia apresentados em revistas especializadas e em congressos, mostram que essas ferramentas amplia o âmbito de conhecimento, ao mesmo tempo que agiliza as pesquisas de campo arqueológicas (NAZARENO, 2005; THOMAZ *et al.*, 2002), pois permite investigar diferentes perspectivas espaciais, como criação de modelos preditivos e análises geomorfológicas e hidrológicas de sítios, reconstruções paleoambientais, detecção de áreas com potencial arqueológico, interpretação de sítios e suas relações ambientais, alcançando resultados amplos na interpretação de contextos arqueológicos (INFANTINI, 2015).

O conceito de modelos preditivos, refere-se ao método de prever a probabilidade de ocorrência de uma determinada variável dependente em uma região não testada, utilizando variáveis independentes. No âmbito arqueológico, sua aplicabilidade é a predição de potenciais sítios em locais nunca pesquisados, com base nas características físicas e ambientais de sítios já conhecidos (padrões locais). Este modelo preditivo de ocupação territorial denota como parte integrante no processo de entender a realidade das populações pré-históricas (SEQUEIRA, 2009; SILVA, 2007) e para citar alguns exemplos Castiello e Tonini (2021) destaca diversos trabalhos que aborda estudos e implementação de modelos preditivos arqueológicos.

A análise das características de um dado conjunto de sítios arqueológicos em uma mesma região, segundo Sequeira (2009), procura estabelecer padrões de forma a compreender o peso que as características ambientais tiveram no momento da escolha do sítio, sendo possível definir uma relação, de forma genérica, por meio de uma função entre a variável dependente e as variáveis independentes (Equação 1):

$$y = f(x_1, x_2, x_3, \dots, x_n) \quad (1)$$

onde:

y é a localização dos sítios arqueológicos (variável dependente); e

x são as características que condicionam a localização (variáveis independentes).

O autor afirma ainda, que o modelo preditivo “complementa os estudos arqueológicos, na tentativa de codificar aspectos do comportamento humano, porém, este nunca será completamente reconstruído, e a razão das suas ações nunca completamente conhecida”.

3.2.2. Modelo Digital de Elevação (MDE)

Garcia (2015) afirma que avanço tecnológico permitiu uma maior participação dos meios computacionais nas análises e mapeamento da superfície terrestre:

Dados topográficos obtidos por meio de radar ou por imagem de satélite ou aerotransportadas constituem uma nova ferramenta para a geomorfologia, pois possibilitam a obtenção de dados do meio físico através de Modelos Digitais de Elevação (GARCIA, 2015, p. 2).

Um novo MDE de versão gratuita foi lançado em dezembro de 2020 pela Agência Espacial Europeia (ESA), por meio do Programa Copernicus DEM, denominado de MDE COP-30 (também nomeado como GLO-30). Consiste em um modelo digital de superfície (MDS) que representa a superfície da Terra incluindo edificações, infraestrutura e vegetação (ESA, 2021).

Este MDS é derivado de interferometria de radar de abertura sintética (*Synthetic Aperture Radar Interferometry* - InSAR) em banda X da missão TanDEM-X, onde o processamento InSAR, calibração e mosaico foram feitos pelo Centro Aeroespacial Alemão (*Deutsches Zentrum für Luft-und Raumfahrt* - DLR) formando o MDE TanDEM-X com tamanho de pixel de 0,4 segundo de arco (aproximadamente 12 m na linha do Equador) com altitude relativa ao sistema de referência WGS84. Com base nesse MDE, a *Airbus Defence and Space*, em parceria público-privada com a DLR, fez o reproprocessamento convertendo-o para altitude ortométrica com base no modelo geoidal EGM2008, o qual passou a se chamar de *WorldDEM_{Core}*. Em seguida esse MDE foi refinado com preenchimento de vazios, edição das linhas costeiras e litorâneas, aplainando das áreas de corpos d'água, tornando o MDE hidrologicamente consistente, mas mantendo o mesmo tamanho de pixel de 0,4 segundo de

arco, e passa a ser chamado *WorldDEMTM*. Os dados *WorldDEMTM* foram reamostrados e distribuídos para a ESA para compor o programa Copernicus DEM (AIRBUS, 2019; 2020; ESA, 2021). O Quadro 1 apresenta alguns parâmetros e especificações do MDE COP-30.

Quadro 1 – Parâmetros de especificações do MDE COP-30.

Formato	GeoTIFF e DTED
Tipo de dado	32 bits, <i>floating point</i> (formato DGED), isto é, a altitude é dada em números decimais 16 bits, inteiro (formato DTED), isto é, a altitude é fornecida em números inteiros (metro a metro)
Quadrículas	1° x 1° latitude/longitude
Sistema de Coordenadas	Geográficas
Datum horizontal	WGS84 (EPSG 4326)
Datum vertical	EGM2008 (EPSG 3855)
Tamanho do pixel	30 metros

Fonte: AIRBUS (2020).

O programa Copernicus DEM, além do COP-30 (GLO-30) oferece outra versão igualmente gratuita e de cobertura global denominada de COP-90 (GLO-90). As diferenças nas versões estão no tamanho de pixel, a COP-90 é a reamostragem do pixel para 3 segundos de arco o que corresponde à aproximadamente 90 m em sistema de coordenadas métricas (ESA, 2021).

3.2.3. Atributos topográficos

O estudo da morfologia do terreno visa identificar e caracterizar as formas do relevo por meio de dados qualitativos (morfografia) e quantitativos (morfometria), descrevendo o relevo e suas principais feições, geralmente, pela criação de classes e subclasses agrupadas de acordo com suas características semelhantes (GARCIA, 2015). A disponibilidade de bases topográficas digitais associadas a métodos computacionais torna possível a extração de atributos topográficos, também denominados de variáveis geomorfométricas. O uso de MDE para extração desses atributos são comumente usados para cálculo de parâmetros de superfície com objetivo de identificar classes e características do relevo (WILSON, 2012).

Os atributos topográficos são classificados como locais ou regionais. Os atributos locais descrevem características pontuais do terreno considerando sua vizinhança, enquanto os atributos regionais descrevem características de uma região maior. Os atributos topográficos também podem ser classificados como primários, quando obtidos diretamente por meio das

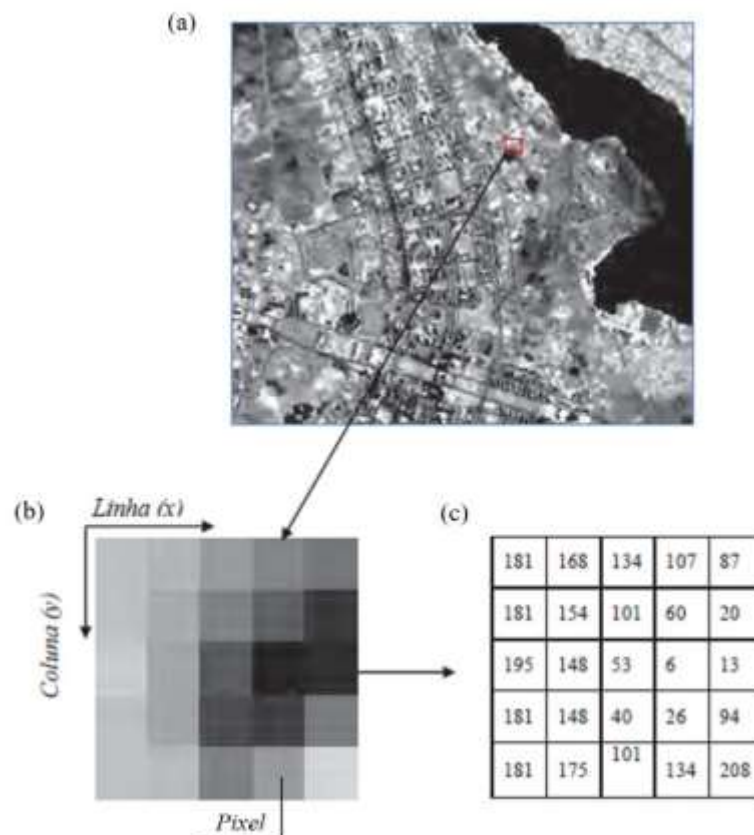
derivações do MDE, enquanto os secundários são por combinação de duas ou mais variáveis primárias ou associados a algum índice (MOORE *et al.*, 1991).

Os atributos topográficos mais usuais utilizados em classificação e predição, estão descritos nas publicações de Moore *et al.* (1991), Riley *et al.* (1999), Gallant e Wilson (2000), Sappington *et al.* (2007), Dietrich e Böhner (2008), Muñoz e Valeriano (2014), Dornik *et al.* (2018) e Cremon *et al.* (2021).

3.2.3.1. Relação de dados matriciais à dados geográficos

Os atributos topográficos extraídos do MDE, são dados matriciais (*raster*). De maneira simplificada, um *raster* consiste em uma matriz de células (*pixels*) organizadas em linhas e colunas, no qual cada *pixel* possui um valor digital (MENESES, 2012), ou seja, a cada *pixel* é atribuído uma identidade de feição única e estes valores representam o fenômeno retratado pelo conjunto de dados *raster* (Figura 1).

Figura 1 – Exemplo de um dado *raster*.

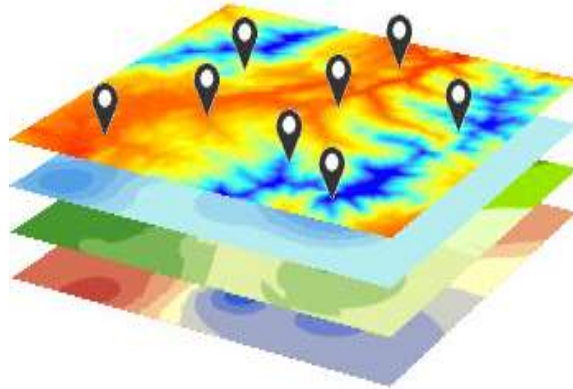


Fonte: Adaptado de Meneses (2012).

Nota: Imagem digital (a) com destaque de um grupo de *pixels* (b) e correspondentes valores digitais (c).

Para identificar determinada característica geográfica, em locais de pontos específicos, como um ponto amostral, é necessário a extração dos múltiplos valores dos *pixels*. Os valores pontuais extraídos das várias camadas *raster* é então anexando à camada do ponto amostral de entrada (ESRI, 2017), como demonstrado na Figura 2.

Figura 2 – Exemplo de extração de valores múltiplos para pontos amostrais.



Fonte: Esri (2017).

3.2.4. Machine Learning e algoritmos de classificação

Machine Learning, é uma das técnicas de Inteligência Artificial (IA) para realizar a tarefa de aprendizado. Consiste na utilização de algoritmos matemáticos e estatísticos para extrair informações de dados brutos e representá-los por meio de algum modelo que será usado para fazer inferências, ou previsões, a partir de novos conjuntos de dados (ESCOVEDO; KOSHIYAMA, 2020, p. 2):

Machine Learning se concentra na descoberta de padrões ou de fórmulas matemáticas que expliquem o relacionamento entre os dados, sem necessariamente se preocupar com seu grau de utilidade ou aplicação ao negócio, e estuda formas de automatização de tarefas inteligentes que seriam difíceis de serem realizadas por humanos.

A depender do tipo de dados de entrada e quais resultados espera-se obter, o treinamento dos algoritmos pode ser basicamente em duas abordagens: supervisionado e não supervisionado. Esses conceitos também são aplicáveis em Processamento digital de imagens (PDI) no sensoriamento remoto.

Escovedo e Koshiyama (2020) esclarece que os algoritmos de aprendizagem supervisionados são utilizados para induzir modelos preditivos por meio da observação de um conjunto de dados rotulados, isto é, “um conjunto de dados de entrada que se sabe de antemão

a saída esperada”. Esta abordagem ainda pode ser subdividida em aprendizagem por ‘classificação’ e ‘regressão’. Por classificação, o objetivo é classificar itens ou amostras de acordo com as características observadas fazendo que a ‘máquina’ possa procurar por padrões e com isso gerar o modelo preditivo, e o resultado desse processo é do tipo categórico. O algoritmo supervisionado de regressão, é semelhante ao processo do supervisionado por classificação, porém o objetivo é analisar as variáveis e prever um valor, e o resultado é do tipo numérico.

No algoritmo de aprendizagem não supervisionado, o treinamento é quase que totalmente automatizado pela ‘máquina’, o aprendizado não há rótulos, isto é, não há entrada de informação de dados reais, o algoritmo se encarrega de agrupar os pixels de características similares (*clusters*) para melhor entendimento da estrutura, ou fazer a simplificação dos dados (MENESES; SANO, 2012; ESCOVEDO; KOSHIYAMA, 2020).

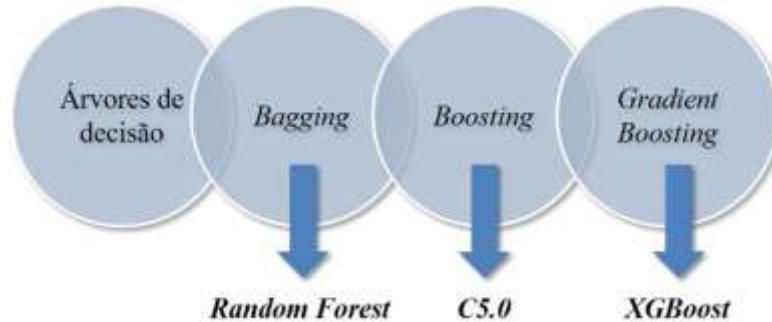
No processo de classificação, as informações das imagens são extraídas de forma a reconhecer padrões e objetos homogêneos, com objetivo de mapear áreas da superfície terrestre que correspondem aos temas de interesse. Segundo Meneses e Sano (2012), Santiago e Silva (2018) os classificadores podem ser divididos em classificadores por pixel ou ‘pixel a pixel’ e classificadores por ‘regiões’. Na classificação por pixel a pixel, o pixel é rotulado dentro de alguma das classes definidas e nesta classificação é considerada apenas o atributo de cada pixel. Na classificação por regiões, é necessário que a imagem a ser classificada passe por um processo de segmentação para que os pixels que compartilham características em comuns são agrupados dentro do mesmo segmento (sub-região) da imagem sendo usado como unidade de classificação.

Novos métodos vêm sendo desenvolvidos para a alcance de informações ambientais, utilizando a modelagem com ferramenta estratégica. Métodos de *machine learning* supervisionado, tais como árvores de decisão, redes neurais artificiais, máquinas de vetores de suporte, entre outros, podem fornecer soluções que auxiliem a extração automática de informações a partir de um conjunto de dados existentes (CHAGAS *et al.*, 2010).

O aprendizado por conjunto também denominado de métodos de *Ensemble Learning*, combinam várias árvores de decisão para produzir melhor desempenho preditivo do que a utilização de uma única árvore de decisão. O princípio básico por trás do *Ensemble Learning* é que um grupo de aprendizes simples se reúne para formar um aprendiz mais robusto (ROKACH, 2019).

Diversos algoritmos usam esta abordagem, com destaque para o *Random Forest* (RF), o C5.0 e o *Extreme Gradient Boosting* (XGBoost), e os que os diferenciam são suas características operacionais de treinamento (Figura 3).

Figura 3 – Evolução dos algoritmos baseados em árvore de decisão.



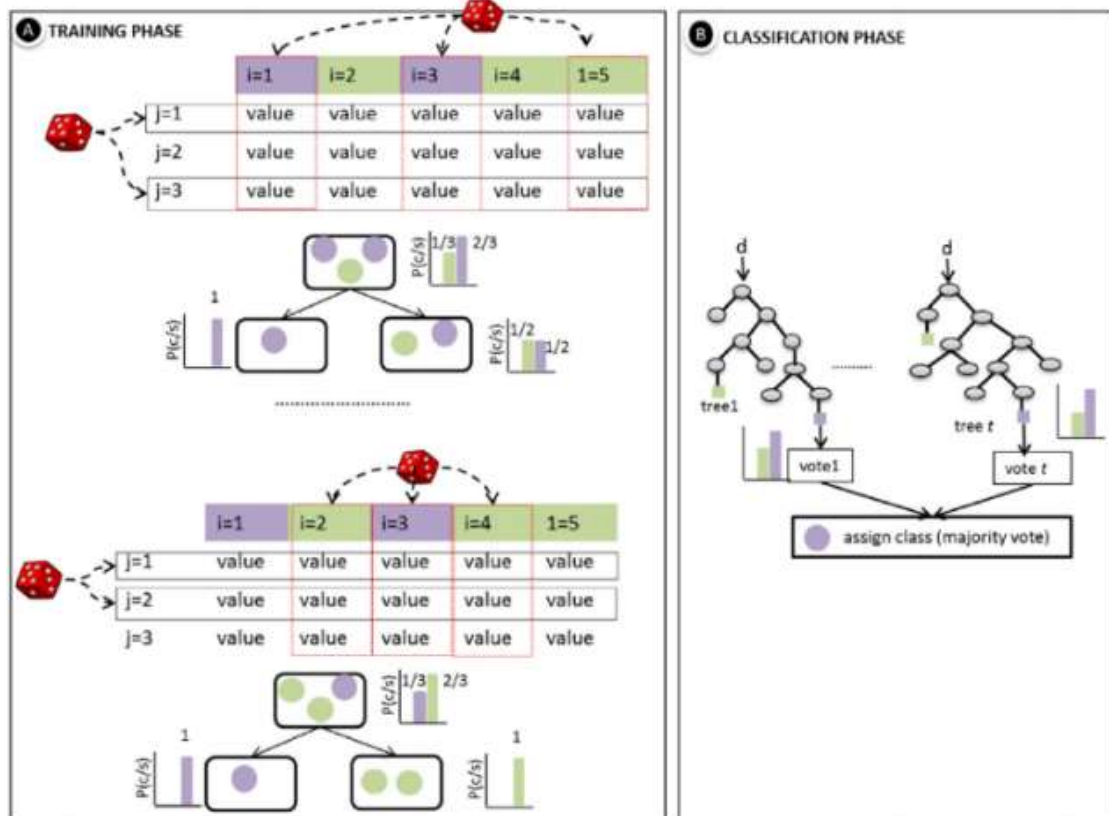
Fonte: Adaptado de Espinosa-Zuniga (2020).

3.2.4.1. *Random Forest*

O classificador RF vem se destacando na comunidade científica para fins de classificação em diversas áreas (GISLASON *et al.*, 2006; RODRIGUEZ-GALIANO *et al.*, 2012). Trata-se de um classificador supervisionado, não-paramétrico, isto é, ele não assume premissas sobre a distribuição estatísticas dos dados, baseado no conceito de árvores de decisão do tipo *Bagging*.

A classificação consiste em uma combinação de várias árvores de decisão preditoras, no qual essas árvores são criadas de forma independentes por meio de um subconjunto de amostras de treinamento (BREIMAN, 2001), como demonstrado na Figura 4.

Figura 4 – Fases de treinamento e classificação do RF.



Fonte: Diniz (2019).

Nota: Sendo i = amostras; j = variáveis; p = probabilidade; c = classe; s = dado; t = número de árvores; d = novo dado para ser classificado e valor = valores que a variável j pode assumir.

São definidos dois parâmetros de entrada para produzir as árvores de decisão: o '*ntree*' que representa o número de árvores a serem criadas e o '*mtry*' que é número de variáveis que serão testadas (DINIZ, 2019). Cada subconjunto é criado por meio de uma amostragem com reposição de 2/3 do conjunto de dados, que dará origem a uma árvore de decisão. O restante das amostras (1/3) que não estão presentes no subconjunto de calibração, são incluídas como parte de outro subconjunto denominado *Out-Of-Bag* (OOB) que será usado para validação de erros do modelo e medição da importância das variáveis de entrada (PAVANELLI, 2016).

A definição da quantidade de árvores na floresta (*ntree*) e dos parâmetros (*mtry*) interfere na acurácia do modelo (BREIMAN, 2001). Os valores padrões de *ntree* e *mtry* do pacote RF do pacote estatístico R são 500 árvores e a raiz quadrada do número total de variáveis ($\sqrt{N}$) de entrada, respectivamente.

Adicionalmente, o RF pode estimar a importância de cada variável para o modelo classificatório, utilizando duas medidas: a primeira denominada importância baseada no erro, mede o aumento quando se permutam os valores do atributo de interesse. A segunda,

denominado importância de *Gini*, baseado na soma dos decréscimos do índice de *Gini* em todos os nós rotulados pelo atributo (BREIMAN, 2001; BASTOS *et al.*, 2014).

De acordo com Rodriguez-Galiano *et al.* (2012), o classificador apresenta muitas vantagens na sua aplicação, dentre elas é um classificador robusto capaz de lidar com inúmeras variáveis de entrada ao mesmo tempo, executado de forma eficiente em grandes bancos de dados; fornece estimativas de quais variáveis são mais importantes na classificação; gera uma estimativa imparcial interna do erro de generalização (erro OOB); calcula proximidades entre pares de casos que podem ser usados na localização de *outliers*; e é computacionalmente mais leve do que outros métodos de conjunto de árvores.

3.2.4.1. C5.0

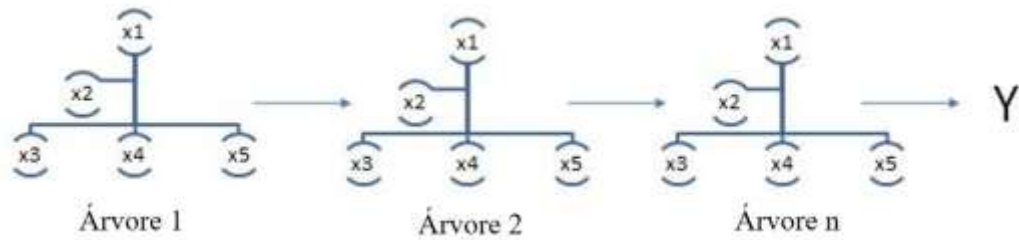
O algoritmo baseado em árvore de decisão C5.0, usa a aprendizagem por reforço (*Boosting*), cuja finalidade é gerar uma série sequencial de árvores de decisão. Para cada modelo (árvore de decisão) criado, há a dependência do resultado do modelo anterior, o que possibilita a seleção de atributos preditivos mais informativos.

Inicialmente os dados de treinamento são divididos, aleatoriamente, em duas partes iguais. Uma parte é gerado o modelo, que tem como propósito avaliar a utilidade dos atributos preditivos. A outra parte das amostras de treinamento, que não foi utilizada para gerar o modelo, é empregada para estimar a taxa de erro do modelo. A taxa de erro é calculada com e sem o atributo preditivo. Caso essa taxa aumente com a inserção do atributo preditivo, sua remoção do conjunto de dados é realizada. Atributos, que não estão em nenhuma divisão da árvore, são considerados sem importância e podem ser removidos do conjunto de dados (RUIZ *et al.*, 2017).

3.2.4.2. Extreme Gradient Boosting

O XGBoost representa uma categoria de algoritmo baseado em árvores de decisão com aumento de gradiente (*Gradient Boosting*), apresentada por Chen e Guestrin (2016). O algoritmo consiste em uma montagem sequencial de árvores de decisão (Figura 5). Esse sequenciamento, tem como finalidade o aprendizado considerando os resultados das árvores anteriores, corrigindo o erro por elas produzido, até que o erro não possa mais ser corrigido, conhecido como gradiente descendente (ESPINOSA-ZUNIGA, 2020).

Figura 5 – Esquemática do algoritmo XGBoost.



Fonte: Adaptado de Espinosa-Zuniga (2020).

Matematicamente o algoritmo funciona assim: Uma árvore inicial F_0 é obtida para prever a variável alvo 'Y', o resultado é associado a um erro residual ($Y - F_0$). Uma nova árvore h_1 é obtida e se ajusta ao erro residual da etapa anterior. Os resultados de F_0 e h_1 são combinados para obter a árvore F_1 , onde a raiz do erro quadrático médio de F_1 será menor do que F_0 . Este processo é continuado iterativamente até que o erro residual seja minimizado ao máximo (ESPINOSA-ZUNIGA, 2020).

Ainda segundo Espinosa-Zuniga (2020), como características para evitar *overfitting* ou enviesamento do modelo, ele emprega processamento paralelo, poda de árvore, tratamento de valores ausentes e regularização (otimização que penaliza a complexidade dos modelos).

As principais vantagens do algoritmo XGBoost são: pode ser utilizado para grandes bancos de dados com múltiplas variáveis; pode lidar com valores ausentes; possui resultados muito precisos; e velocidade de execução excelente. Em oposição, suas principais desvantagens são no maior recurso computacionais em grandes bancos de dados; os parâmetros do algoritmo devem ser ajustados corretamente para evitar *overfitting* do modelo e minimizar o erro de precisão; e funciona apenas com vetores numéricos, portanto, é necessário converter previamente os tipos de dados não numéricos (ESPINOSA-ZUNIGA, 2020).

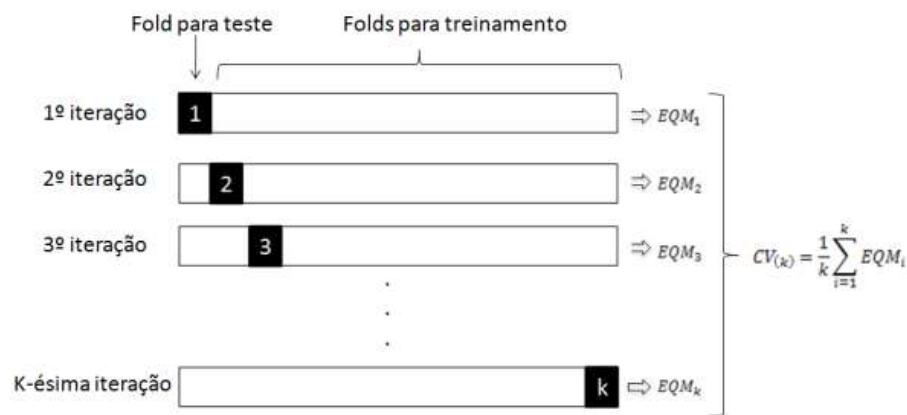
3.2.5. Treinamento e teste da classificação

Na calibração do modelo, usualmente 70% das amostras são utilizadas para treino e 30% das amostras para teste. Ainda na parte de treinamento, é possível trabalhar com métricas associadas a validação cruzada ou *cross-validation*. Está é uma abordagem amplamente utilizada para estimar o erro de teste associado ao ajuste de um determinado método de aprendizagem estatística em um conjunto de observações. É útil na calibração e pré-avaliação do modelo e ajuda a garantir que este não esteja superajustado (*overfitting*).

Há diversos tipos de validação cruzada: *Holdout* que divide o conjunto de observações em duas partes (um conjunto de treinamento e um conjunto de validação); *Leave-One-Out (LOOCV)*, parecido com o *Holdout*, porém utiliza somente uma observação na fase de validação; e *k-fold* que faz uma divisão em *k*-partes (AZEVEDO, 2018). Contudo, a abordagem de divisão das observações em duas partes pode trazer resultados divergentes, dependendo da informação do conjunto de dados de entrada, em especial se os dados são escassos. A abordagem de validação cruzada por *k-fold* minimiza esses problemas (FERREIRA, 2018).

A validação cruzada do tipo *k-fold* consiste em dividir aleatoriamente um conjunto de amostras em *k* partes (*fold*), usando uma parte para validação e o restante (*k*–1) para treinamento, fazendo isso *k* vezes (Figura 6).

Figura 6 – Esquematização do método de reamostragem por abordagem de validação cruzada *k-fold*.



Fonte: Azêvedo (2018).

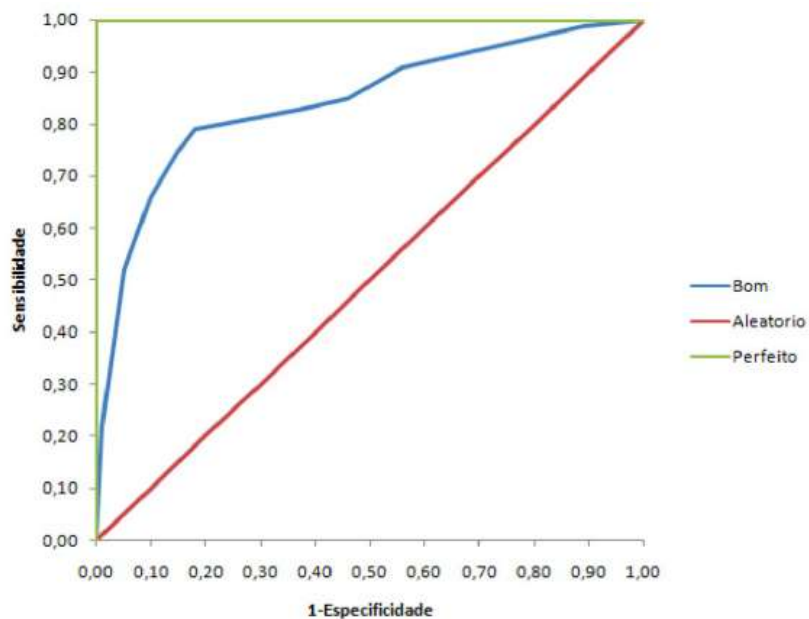
O modelo é ajustado utilizando o conjunto de treinamento. Então, é medido a exatidão preditiva dos modelos por meio do cálculo do Erro Médio Quadrático (EMQ), em caso de regressão e a acurácia em caso de classificação, no conjunto de testes.

O processo de validação cruzada é repetido *k* vezes, de modo que cada um dos *k* grupos sejam utilizados exatamente uma vez como dado de teste para validação do modelo. Esse processo resultará em *k* estimativas do erro de teste, EMQ_1 ; EMQ_2 ; EMQ_k . A estimativa de validação cruzada *k-fold* é calculada pela média destes valores e o resultado final desse processo é o desempenho médio do classificador nos *k* testes. O objetivo de repetir os testes múltiplas vezes é aumentar a confiabilidade da estimativa da precisão do classificador (AZEVEDO, 2018). Na literatura, os valores mais usuais são *k* = 2, 5 ou 10, porém a escolha desse valor depende da quantidade de amostras (MATOS *et al.*, 2009).

Uma das últimas etapas no processo de classificação é a avaliação do resultado. Nessa etapa se avalia o desempenho preditivo da classificação. Métricas como a acurácia, quantifica percentual de acerto da classificação e depende de uma série de fatores, desde as resoluções espaciais, complexidade do terreno, a escolha do próprio algoritmo de classificação a ser utilizado, escolha das classes e o conjunto de dados amostrais a ser representado como verdade terrestre (MENESES; SANO, 2012). O que buscamos é a mensuração de performance para os dados categóricos que quantificam a eficiência da classificação, a exemplo da análise da curva ROC (*Receiver Operator Characteristic Curve*), a análise da AUC (*Area Under the Curve ROC*), a sensibilidade, a especificidade e *F-score*.

Escovedo e Koshiyana (2020), esclarece que a curva ROC, curva característica de operação do receptor, é um gráfico que permite avaliar o desempenho do classificador. Essa visualização leva em consideração a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos (especificidade). Quanto mais a curva estiver próxima ao canto superior esquerdo, melhor é o classificador, como demonstrado na Figura 7.

Figura 7 – Exemplo de curvas ROC.



Fonte: Escovedo e Koshiyana (2020).

A curva ROC pode ser avaliada pela métrica AUC, área sob a curva ROC, e se trata do cálculo da área formada abaixo da curva. O valor da AUC varia entre 0 e 1. Quanto maior esse valor, melhor é o modelo preditivo (SILVA, 2019).

A sensibilidade ou *recall* é a proporção de verdadeiros positivos, que é capacidade de prever corretamente o resultado classificado como positivo (ESCOVEDO; KOSHIYAMA, 2020) sendo calculada pela Equação 2.

$$\text{Sensibilidade} = \text{VP} / (\text{VP} + \text{FN}) \quad (2)$$

onde:

VP é Verdadeiro Positivo (são observações cujo valor real é positivo e o valor previsto é positivo, isto é, o modelo acertou);

FN é Falso Negativo (são casos em que o resultado correto é positivo, entretanto o resultado obtido é negativo, isto é, o modelo errou).

Todavia, a especificidade é a proporção de verdadeiros negativos, que é capacidade do sistema em prever corretamente a ausência da condição, ou seja, o resultado classificado como negativo (ESCOVEDO; KOSHIYAMA, 2020) dada pela Equação 3.

$$\text{Especificidade} = \text{VN} / (\text{VN} + \text{FP}) \quad (3)$$

onde:

VN é Verdadeiro Negativo (são observações cujo valor real é negativo e o valor previsto é negativo, isto é, o modelo acertou);

FP é Falso Positivo (são casos em que o resultado correto é negativo, entretanto o resultado obtido é positivo, isto é, o modelo errou).

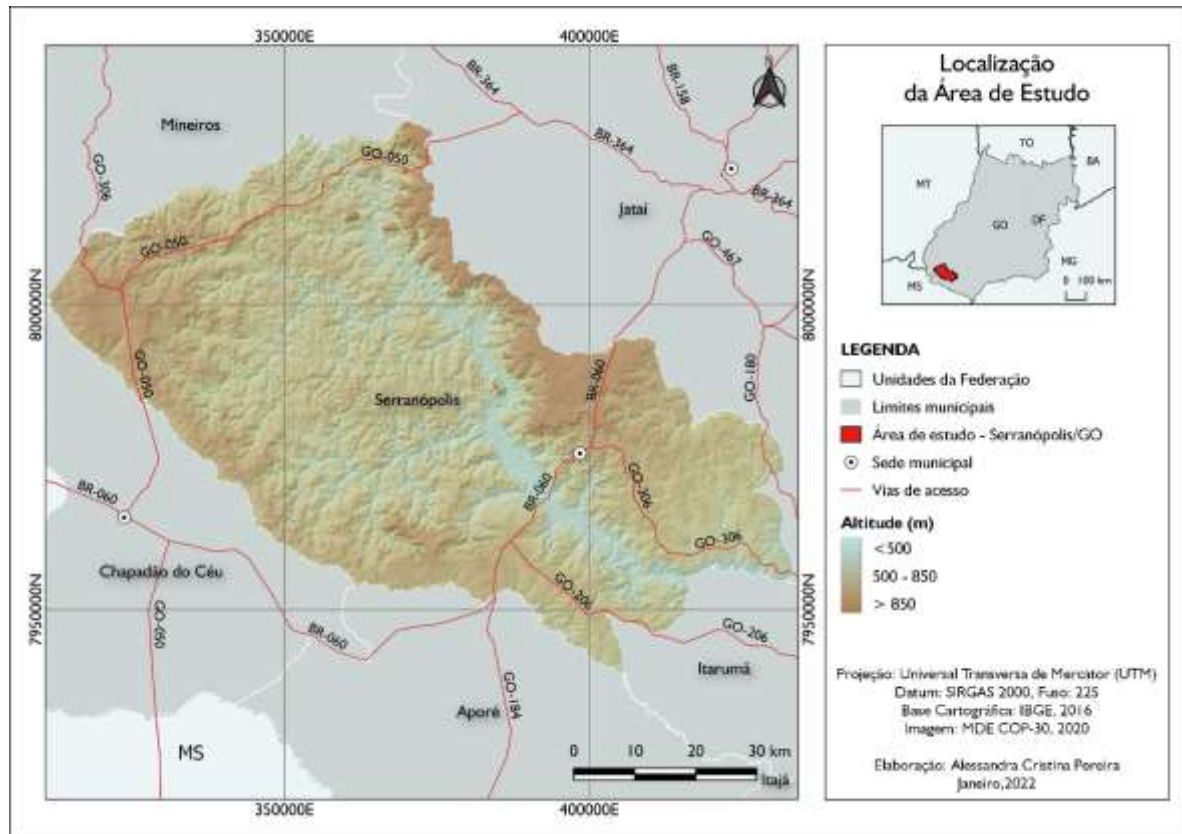
A métrica *F-score* é uma média harmônica calculada com base na precisão e na sensibilidade, que torna a métrica mais sensível a desproporções das taxas (SILVA, 2019). Ela pode ser obtida Equação 4:

$$\text{F-score} = 2 \times (\text{Precisão} \times \text{Sensibilidade}) / (\text{Precisão} + \text{Sensibilidade}) \quad (4)$$

4. LOCALIZAÇÃO E ASPECTOS GERAIS DA ÁREA

Situada na região sudoeste do estado de Goiás, a área de estudo abrange o município de Serranópolis, com 5.526,4 km² e distante cerca de 450 km da capital, Goiânia (Figura 8).

Figura 8 – Mapa de localização da área de estudo – Serranópolis, Goiás.



Fonte: A autora (2022).

No contexto ambiental, o município apresenta um clima do tipo tropical subquente e úmido, com invernos secos e verões chuvosos, enquadrado como Aw, na classificação de Köppen. Quanto à temperatura, é predominantemente elevada na primavera-verão variando de 24° a 32°C e amena no inverno (MARIANO, 2005). A variação da temperatura média anual está correlacionada ao relevo, onde localidades com altitudes inferiores a 450 m, como nos vales dos rios Aporé, Corrente, Verde, Claro, Preto e Bois, apresentam temperaturas mais elevadas. Em localidades com altitudes superiores a 800 m, apresentam temperaturas médias inferiores a 21°C (VARGAS, 1989).

A vegetação característica da região é rasteira e rala, típica do bioma Cerrado. No que concerne a hidrografia, a rede de drenagem está inserida na Bacia Hidrográfica do rio Paranaíba, porção sudoeste do Estado, sub-bacia dos rios Aporé, Corrente, Verde e Claro (GOIÁS, 2021).

Geologicamente, a região apresenta arenitos da Formação Botucatu e basaltos da Formação Serra Geral (SCHMITZ *et al.*, 2004). A Formação Botucatu favorece o surgimento de abrigos, alguns dos quais foram ocupados por grupos pré-coloniais de caçadores-coletores e posteriormente por agricultores-ceramistas. Em razão do recuo das escarpas por erosão diferencial, formando as “*cuestas*”, o arenito foi coberto por lavas basálticas, metamorfoseando superficialmente, formando os tetos e, às vezes, as paredes dos abrigos.

O solo da região é resultante da decomposição dos arenitos Botucatu nas chapadas e interflúvios, enquanto nos vales, entalhados pela ação fluvial, ocorrem os latossolos vermelhos associados ao latossolo roxo. Sobre os chapadões predominam o latossolo vermelho-amarelo associado ao vermelho. Ocorrem, ainda, com menor expressão espacial em manchas esparsas, argissolos, solos hidromórficos e neossolos quartzarênicos (GUERRA, 1989).

A geomorfologia local, insere-se no contexto das paisagens dominadas por forma e relevo convexas e tabulares, com diferentes ordens de grandezas e de aprofundamento de drenagens pertencentes ao domínio do Planalto Setentrional da Bacia Sedimentar do Paraná. A hipsometria encontra-se na faixa de 600 a 800 metros de altura, podendo, em algum caso, atingir altitudes em torno de 1.000 m (SCHMITZ *et al.*, 2004).

Os sítios arqueológicos encontram-se distribuídos em altitude diversas entre 750 e 560 m, declividade variando entre 4° e 35°, e relevo variando de suave-ondulado a montanho, constituídos por arenito silicificados da Formação Botucatu, no vale do rio Verde, afluentes do rio Paranaíba (BITENCOURT, 2008).

5. MATERIAL E MÉTODOS

Nesta sessão são apresentados os materiais utilizados para a realização do trabalho e os procedimentos metodológicos.

5.1. MATERIAL

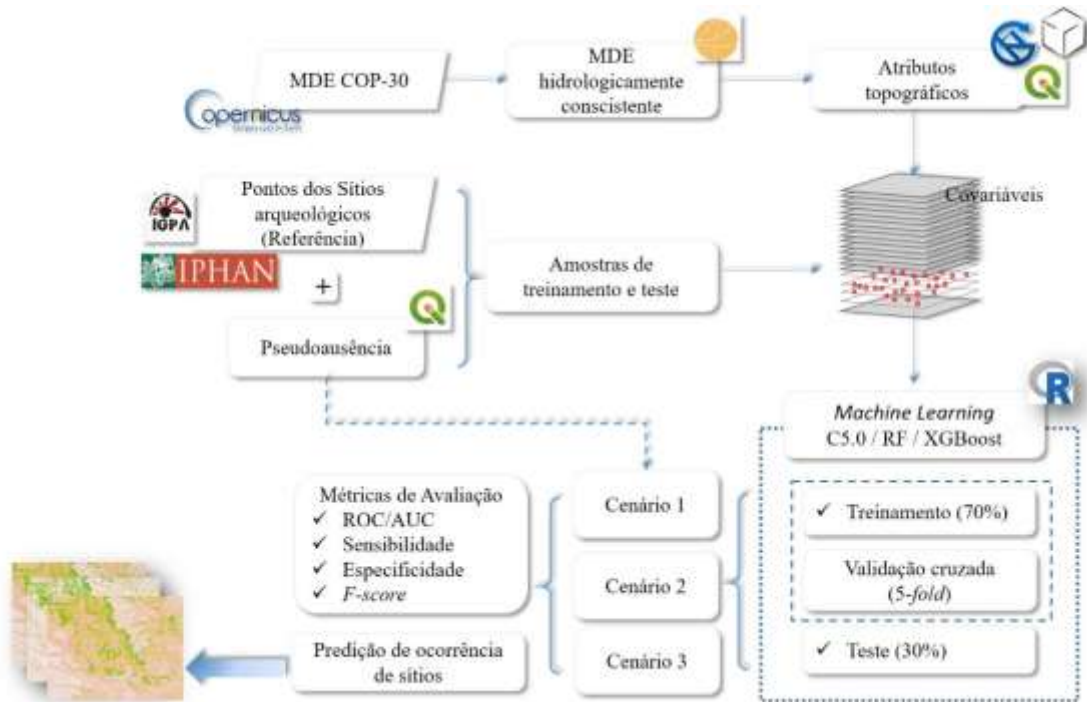
Para a realização desse estudo foram necessários:

- Arquivo em formato vetorial: os arquivos vetoriais possuem fontes, natureza e escalas diversas. O conjunto de arquivos vetoriais, foi composto por arquivos de pontos dos sítios arqueológicos e arquivos de limites políticos, divisão administrativa e sede municipal, malha viária e hidrografia, na escala 1:100.000 disponível em www.ibge.com.br, para compor os produtos cartográficos;
- Arquivo em formato matricial: MDE COP-30 para extração das covariáveis topográficas para a análise geomorfométrica;
- *Softwares*: *System for Automated Geoscientific Analyzes* (SAGA-GIS) v.2.3.2; *Whitebox Geospatial Inc* v.2.1.0; *TerraViewHidro* v.4.2.1; *QGIS Development Team* v.3.16.8; e *RStudio* v.2021.09.0, todos de acesso livre.

5.2. PROCEDIMENTOS METODOLÓGICOS

Os procedimentos metodológicos para o trabalho, iniciaram com a aquisição e processamento dos atributos topográficos, usados como covariáveis independentes. Aquisição dos pontos dos sítios arqueológicos, usados como variável dependente e criação de pontos de pseudoausência de sítios arqueológicos. A predição foi realizada pelos algoritmos de classificação supervisionada baseados em árvore de decisão RF, C5.0 e XGBoost. Estes algoritmos foram aplicados seguindo a validação cruzada do tipo *k-fold* para o treinamento e calibração do modelo. Para aferir a confiabilidade dos resultados da classificação, foram utilizadas como métricas de performance: análise da curva ROC, AUC, a sensibilidade, a especificidade e *F-score*. Tais procedimentos são apresentados na Figura 9 e detalhados a seguir.

Figura 9 - Fluxograma do procedimento metodológico adotados no estudo.



Fonte: A autora (2022).

5.2.1. Aquisição e processamento dos atributos topográficos

O estudo utilizou um conjunto de 18 atributos topográficos (Quadro 2) usados como covariáveis geomorfométricas independentes para inferências de sítios arqueológicos, extraídos diretamente do MDE COP-30. Esse MDE está disponível em coordenadas geográficas com datum horizontal WGS-84 e tamanho de pixel de 1 segundo de arco, que foram reprojetaadas para coordenadas métricas no sistema de coordenadas *Universal Transversa de Mercator* (UTM) fuso 22, projeção Sul com tamanho de pixel de 30 m, datum horizontal SIRGAS2000, pelo método de convolução cúbica com tamanho de pixel de 30 m, etapa foi realizada no *software* QGIS (QGIS DEVELOPMENT TEAM, 2021). Em seguida, foi realizado o processo de correção do MDE para eliminação dos vazios (fossos), tornando o MDE hidrológicamente consistente, utilizando o *software* TerraViewHidro. Por fim, a extração dos atributos foi realizada por meio de algoritmos disponíveis nos *softwares* SAGA-GIS, QGIS e Whitebox. Neste trabalho, optou-se por adotar o nome das variáveis geomorfométricas em inglês, assim como está disponível na maioria dos *softwares* e literatura.

Quadro 2 – Atributos topográficos utilizados neste estudo para prever regiões de sítios arqueológicos e suas definições.

Atributos	Definição
<i>Slope</i>	Declividade. Inclinação entre o plano horizontal e um plano tangencial para a superfície do terreno em porcentagem (MOORE <i>et al.</i> , 1991).
<i>Profile curvature</i>	Curvatura vertical. Corresponde a segunda derivada da altitude representada por um MDE, caracterizando as propriedades convexo/côncavo da superfície quando analisado em perfil (MOORE <i>et al.</i> , 1991).
<i>Plan curvature</i>	Curvatura horizontal. Representa as características de convergência ou divergência de fluxo na superfície (MOORE <i>et al.</i> , 1991).
<i>Aspect</i>	Aspecto. Representa a orientação da inclinação da vertente em graus no sentido horário a partir do Norte para cada pixel em um MDE. É representado em graus, variando de 0° a 360° (MOORE <i>et al.</i> , 1991).
<i>Circular variance of aspect</i>	Variância circular do aspecto. Esta variável calcula a variância da orientação da vertente (aspecto) dentro de uma janela móvel circular. É, portanto, uma medida da complexidade do formato da superfície ou textura. Terá um valor de 0,0 para locais lisos e próximo de 1,0 em áreas de alta rugosidade de superfície ou topografia complexa (GROHMANN <i>et al.</i> , 2010).
<i>Depression</i>	Depressão. Probabilidade de cada célula (pixel) pertencer a uma depressão.
<i>Dissection</i>	Dissecação. Corresponde a elevação abaixo do ponto mais alto dentro de uma janela móvel (MUÑOZ, 2009).
<i>Local relief</i>	Relevo local (Amplitude). Corresponde à diferença de elevação entre o ponto mais baixo e o mais alto dentro de uma janela móvel (MUÑOZ, 2009; HESSE, 2010).
<i>Relative Topographic Position (RTP)</i>	Posição topográfica relativa. Índice de posição topográfica local, ou seja, quão elevado ou baixo é um local em relação a sua vizinhança (NEWMAN <i>et al.</i> , 2018).
<i>Vertical Distance to Channel Network (VDCM)</i>	Distância Vertical para a Rede do Canal. É a diferença entre as elevações originais e a interpolação da elevação do nível base da rede de canais, em metros (DORNIK <i>et al.</i> , 2018).
<i>SAGA Wetness Index (TWI)</i>	Índice de umidade topográfica. Calculado a partir do declive e da área de contribuição modificado (DORNIK <i>et al.</i> , 2018).
<i>Multi-Resolution Index of the Ridge Top Flatness (MRRTF)</i>	Índice de Resolução Múltipla do Topo de Cume Plano. Índice projetado para identificar áreas planas altas em uma variedade de escalas (crista/divisor de drenagem). Valores 0 indicam áreas que são íngremes ou baixas, com valores 1 e maiores, indicando áreas progressivamente elevadas e planas (DORNIK <i>et al.</i> , 2018).
<i>Multi-Resolution Index of Valley Bottom Flatness (MRVBF)</i>	Índice de Resolução Múltipla de Nivelamento do Fundo do Vale. Este índice classifica os graus de achatamento do fundo do vale, que podem estar relacionados com a profundidade do depósito. O índice também pode ser usado para identificar as restrições das águas subterrâneas e para delinear as unidades hidrológicas e geomórficas (GALLANT, 2003).
<i>Topographic height</i>	Altura topográfica. Corresponde a elevação acima do ponto mais baixo dentro de uma janela móvel (MUÑOZ, 2009).
<i>Vector Ruggedness Measure (VRM)</i>	Medição da Rugosidade Vetorial. Quantifica a rugosidade do terreno medindo a dispersão de vetores ortogonais à superfície do terreno. Esse método sintetiza a variabilidade da declividade e da orientação de vertente em uma única medida. Os valores de VRM são baixos tanto em áreas planas, mas os valores são elevados em áreas que com terreno íngreme e acidentado. Um valor de 0 representa nenhuma variação do terreno (ou menor rugosidade), enquanto um valor de 1 representa uma variação completa do terreno (rugosidade máxima) (SAPPINGTON <i>et al.</i> , 2007).
<i>Positive openness</i>	Abertura positiva. Expressa a dominância (positiva) de um local de paisagem. A abertura tem sido relacionada com a amplitude que uma paisagem pode ser vista de qualquer posição (YOKOYAMA <i>et al.</i> , 2002).
<i>Negative openness</i>	Abertura negativa. Expressa o enclausuramento (negativa) de um local de paisagem (YOKOYAMA <i>et al.</i> , 2002).
<i>Sky View Factor (SVF)</i>	Fator de visibilidade do céu. Define a proporção hemisférica do céu visível do solo (não obstruído por edifícios, terrenos ou árvores). Varia de 0 a 1 (HÄMMERLE <i>et al.</i> , 2011; ZAKŠEK <i>et al.</i> , 2011).

Fonte: Adaptado de Cremon *et al.* (2021).

5.2.2. Aquisição dos dados de sítios arqueológicos

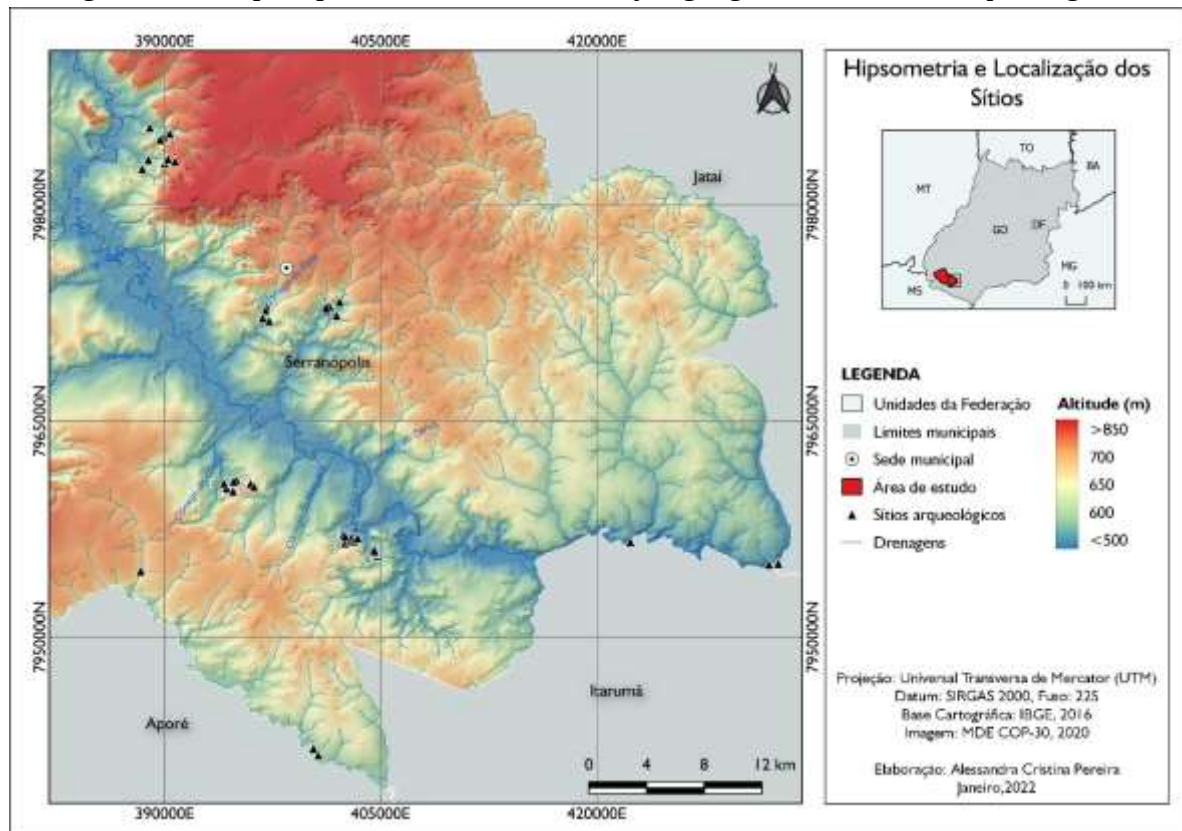
Os sítios arqueológicos usados neste estudo representam a variável dependente no modelo. Os registros foram levantados da base de dados do Cadastro Nacional de Sítios Arqueológicos (CNSA) disponibilizados pela superintendência do IPHAN em Goiás (BRASIL, 2021)⁵, além de outros pontos de localização e informações cedidos pelos pesquisadores do IGPA/PUC Goiás. Estas fontes permitiram reunir 69 registros de sítios arqueológicos.

Após um refinamento para eliminar dados repetidos, sítios fora da área de interesse e sem informação de geolocalização, 54 sítios foram selecionados e usados para amostras de treinamento e teste do modelo preditivo (APENDICE A). Os perfis completos dos sítios selecionados estão descritos nas fichas catalográficas disponibilizadas pelo IPHAN.

A respeito das principais características dos sítios selecionados, dos 54 sítios, 37 são classificados como abrigo sob rocha e 17 a céu aberto; 50 registros foram classificados como sítios pré-colonial e quatro como sítios históricos; e quanto a relevância científica dos sítios, 30 são classificados com alta relevância científica, 3 como média e 21 foram classificados como baixa relevância. A Figura 10 ilustra a distribuição geográfica dos sítios selecionados para este trabalho.

⁵ ANEXO A: Ofício nº 717/2021/COTEC IPHAN-GO/IPHAN-GO-IPHAN

Figura 10 – Mapa hipsométrico com localização geográfica dos sítios arqueológicos.



Fonte: A autora (2022).

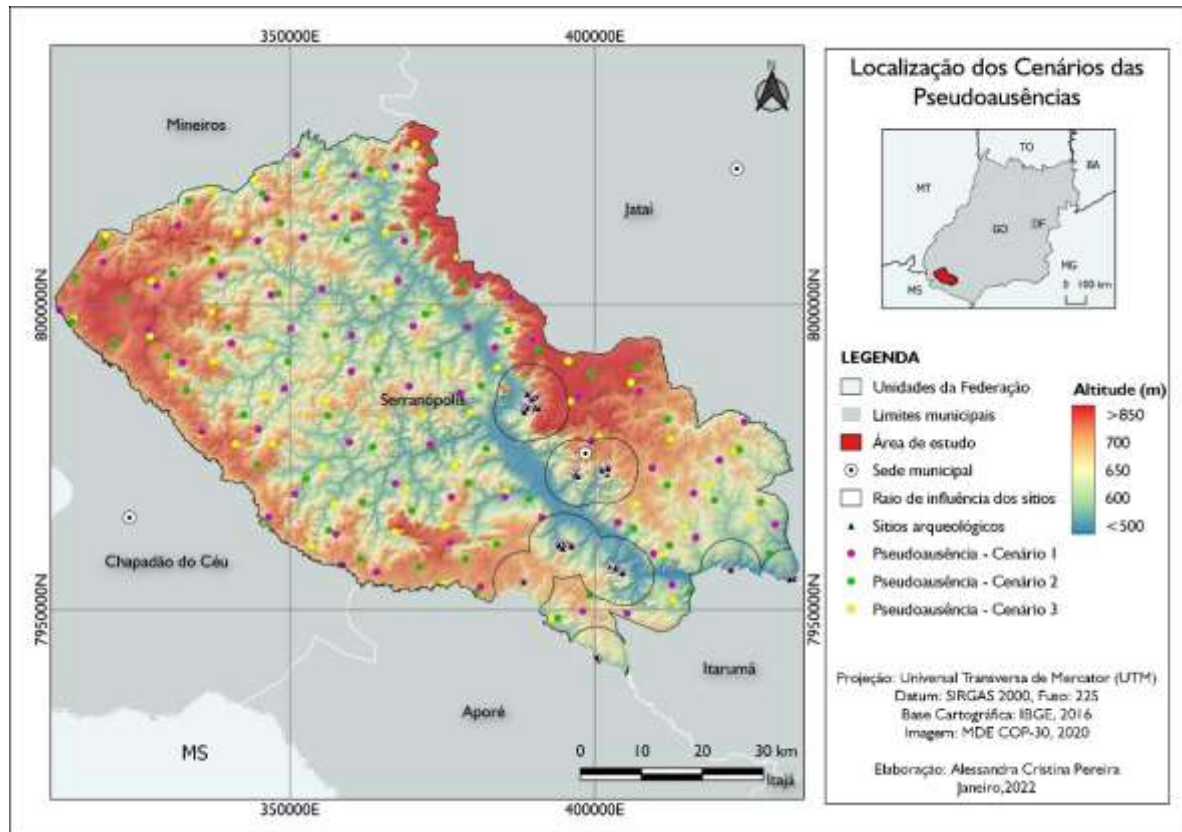
5.2.3. Amostras de treinamentos

Em função dos pontos de presença de sítios arqueológicos, foram definidos os pontos de pseudoausência de sítio e assumidos como não sítio, ou seja, correspondem às alternativas para os reais pontos de ausência. A definição de pontos de pseudoausência é comum em modelagens onde se conhece a presença espacial de um fenômeno ou alvo a ser predito, mas a ausência é desconhecida, sendo bastante utilizada em modelagens de distribuição de espécies na ecologia (ZURELL *et al.*, 2020).

A partir de um raio de 5 km dos pontos de ocorrência de sítios, foram gerados, aleatoriamente, um conjunto de 60 pontos de pseudoausência com distância mínima de cinco quilômetro entre eles. Uma vez que a distribuição desses pontos de pseudoausência é aleatória, logo seu arranjo espacial influencia diretamente no resultado da modelagem efetuada, sendo assim, para avaliar a influência dessa distribuição na modelagem, foram considerando três cenários com diferentes distribuições de pontos de pseudoausência (Figura 11). As amostras foram classificadas com valor '0' para as pseudoausência e valor '1' para a presença de sítios arqueológicos. Esse processo resultou em um conjunto de 114 pontos de dados geográficos

binários balanceados de presenças e pseudoausências para sítios, necessário para fins da modelagem. Essa etapa foi realizada no *software* QGIS (QGIS DEVELOPMENT TEAM, 2021).

Figura 11 – Mapa de localização dos cenários das pseudoausências e raio de influência dos sítios arqueológicos.



Fonte: A autora (2022).

5.2.4. Classificação, validação e predição

A predição empregada foi com a abordagem de classificação supervisionada pixel a pixel, utilizando os algoritmos classificadores baseados em árvore de decisão C5.0, RF e XGBoost, executado no pacote de *machine learning caret* (KUHN *et al.*, 2021), no *software* R pela interface do *RStudio*.

Na calibração do modelo e na validação-cruzada dos resultados, neste trabalho, foi considerado 70% dos pontos de presença e pseudoausência de sítios arqueológicos e com base nesse pontos foi utilizada a abordagem *k-fold* na classificação, com $k = 5$. Nessa abordagem, o conjunto de amostras de treinamento é dividido em cinco subconjuntos (*5-fold*). Destes, um subconjunto é utilizado para derivar métricas de validação-cruzada e os outros quatro subconjuntos restantes foram usados para treinar os modelos de classificação. Esse processo é

repetido cinco vezes até que todos os subconjuntos tenham sido utilizados para treinamento e validação-cruzada, ou seja, todas os subconjuntos fazem parte do grupo de validação-cruzada alguma vez nesse processo de repetição, conforme esquematizado na Figura 6.

Para aferir a confiabilidade dos resultados da classificação, 30% dos pontos de presença e pseudoausência de sítios arqueológicos, não utilizados no treinamento e calibração, foram utilizadas como amostras independentes para teste da modelagem, onde foram avaliadas as métricas de performance: curva ROC, AUC, sensibilidade, especificidade e *F-score*.

Para a avaliação de quais atributos foram mais importantes no processo de predição de possíveis locais de ocorrência de sítios arqueológicos, foi usado via pacote *caret* (KUNH *et al.*, 2021) métricas para cada algoritmo, normalizados em valores entre 0 e 100.

Para o algoritmo RF é utilizado o critério de impureza de *Gini*, para quantificar a redução de impurezas em cada nó na árvore de decisão (BREIMAN, 2001). Essa métrica quantifica a redução de impurezas em cada nó para a variável preditora utilizada para dividir os dados nas várias árvores de decisão que o método gera, fornecendo um *ranking* da importância das variáveis preditoras (BREIMAN, 2001). No algoritmo C5.0, a importância das variáveis é determinada pela porcentagem de amostras do conjunto de treinamento que recaem em todos os nós terminais após a partição da árvore de decisão. Com isso, a variável preditora na primeira partição tem automaticamente uma medida de importância de 100%, já que todas as amostras são afetadas por esta partição, enquanto nos nós terminais da árvore de decisão, por cobrirem apenas uma pequena parcela de amostras do conjunto de treinamento, a pontuação de importância pode ser próxima de zero (KUNH *et al.*, 2021). Por sua vez, o classificador XGBoost é baseado em uma métrica denominada de Ganho, no qual representa a contribuição fracionária de cada recurso (variável preditora) para o modelo com base no ganho total de divisões desses recursos (KUNH *et al.*, 2021). Embora cada algoritmo use uma métrica diferente para computar a importância das variáveis, os dados normalizados entre 0 e 100, permitem uma comparação relativa entre os algoritmos. Adicionalmente, todas as métricas têm como princípio atribuir maior peso para as variáveis que estão nas partes mais altas da árvore de decisão.

Para comparar o conjunto de dados e suas dispersões nas covariáveis mais preditoras, foi usado o gráfico de *box-plot* combinado ao de *violin plots*, considerando o cenário de melhor desempenho estatístico. Os gráficos de *box-plot* permitem visualizar os valores mínimos e máximos, mediana, primeiro e terceiro quartil, e *outliers* do conjunto de dados. Enquanto o *violin plot* permite visualizar a densidade da probabilidade dos dados. Regiões mais largas do gráfico indicam maior densidade de pontos e que estes valores ocorrem com mais frequência,

enquanto que regiões mais estreitas indicam menor densidade de pontos, ou seja, valores que ocorrem com menos frequência. A largura da caixa é proporcional ao tamanho amostral. Todos os gráficos foram elaborados no *software* R pela biblioteca *ggplot2*.

Por fim, com a predição das ocorrências de sítios arqueológicos, utilizando o *software* QGIS (QGIS DEVELOPMENT TEAM, 2021), foram elaborados mapas temáticos, que em função dos dados utilizados, é compatível com a escala 1:50.000.

6. RESULTADOS

No teste dos resultados da predição foram utilizadas como métricas de performance a análise da curva ROC e AUC, a sensibilidade, a especificidade e *F-score*.

Na comparação entre os algoritmos utilizados neste trabalho, no cenário 1 o algoritmo C5.0 apresentou uma AUC de 0,799; o RF foi 0,932; e o XGBoost foi de 0,879. No cenário 2, a AUC no algoritmo C5.0 foi de 0,865; o RF foi 0,931; e o XGBoost foi de 0,896. E por fim no cenário 3, o algoritmo C5.0 apresentou uma AUC de 0,844; RF apresentou uma AUC de 0,875; e o XGBoost foi de 0,747 (Tabela 1).

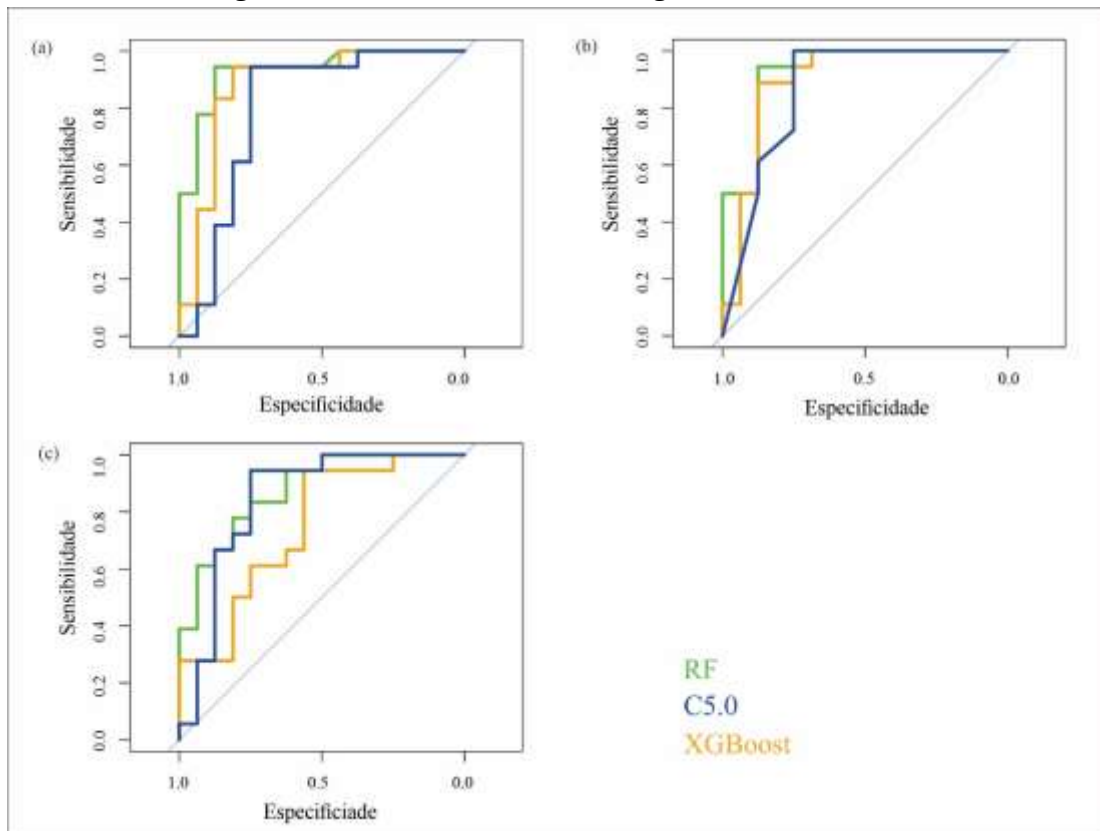
Tabela 1 – Valores da AUC para os algoritmos C5.0, RF e XGBoost, nos três cenários.

Cenário / Algoritmo	C5.0	RF	XGBoost
Cenário 1	0,799	0,932	0,879
Cenário 2	0,865	0,931	0,896
Cenário 3	0,844	0,875	0,747

Nas três simulações de cenários (Figura 12) são demonstradas as curvas ROC, que expressa a relação entre a especificidade (falso positivo, eixo x) e a sensibilidade (verdadeiro positivo, eixo y).

Em geral, observa-se que curva do RF, em cor verde, se encontra mais próxima ao eixo y nos três cenários, indicando um melhor resultado de verdadeiro positivo. Nos cenários 1 e 2 a curva do XGBoost (cor laranja) apresentou melhores resultados perante a curva do C5.0 (cor azul), com exceção do cenário 3, onde o C5.0 apresentou melhor resultado que a curva do XGBoost.

Figura 12 – Curva ROC dos três algoritmos, nos três cenários.



Fonte: A autora (2022).

Nota: Curva ROC do C5.0, RF e XGBoost, no cenário 1 (a), cenário 2 (b) e cenário 3 (c).

Com a métrica de sensibilidade (Tabela 2), observa-se que o algoritmo RF apresentou o maior valor no cenário 1 (0,850) e cenário 2 (0,810). No cenário 3, o RF apresentou o mesmo valor que o algoritmo C5.0 (0,778). O algoritmo XGBoost, apresentou um desempenho pouco abaixo, comparado aos demais. Esses valores representam a proporção de verdadeiros positivos, ou seja, quanto que o modelo acertou a presença de sítios arqueológicos.

Tabela 2 – Valores da taxa de sensibilidade dos algoritmos C5.0, RF e XGBoost, nos três cenários.

Cenário / Algoritmo	C5.0	RF	XGBoost
Cenário 1	0,810	0,850	0,810
Cenário 2	0,765	0,810	0,750
Cenário 3	0,778	0,778	0,682

Quanto a análise da métrica de especificidade (Tabela 3), cujo valores representam a proporção de verdadeiros negativos, ou seja, quanto que o modelo foi capaz de prever corretamente a ausência de uma determinada condição, no caso os sítios arqueológicos, observa-se que o algoritmo RF, apresentou o maior valor no cenário 1 (0,929). No cenário 2, o

algoritmo XGBoost apresentou melhor valor (1,000). No cenário 3, os três algoritmos apresentaram os mesmos valores (0,750).

Tabela 3 – Valores da taxa de especificidades dos algoritmos C5.0, RF e XGBoost, nos três cenários.

Cenário / Algoritmo	C5.0	RF	XGBoost
Cenário 1	0,923	0,929	0,923
Cenário 2	0,706	0,923	1,000
Cenário 3	0,750	0,750	0,750

A análise da métrica *F-score* (Tabela 4), observa-se que o algoritmo RF, apresentou o maior valor no cenário 1 (0,866) e cenário 2 (0,831). No cenário 3, o RF apresentou o mesmo valor que o algoritmo C5.0 (0,771). O algoritmo XGBoost, apresentou valores pouco abaixo, comparado aos demais.

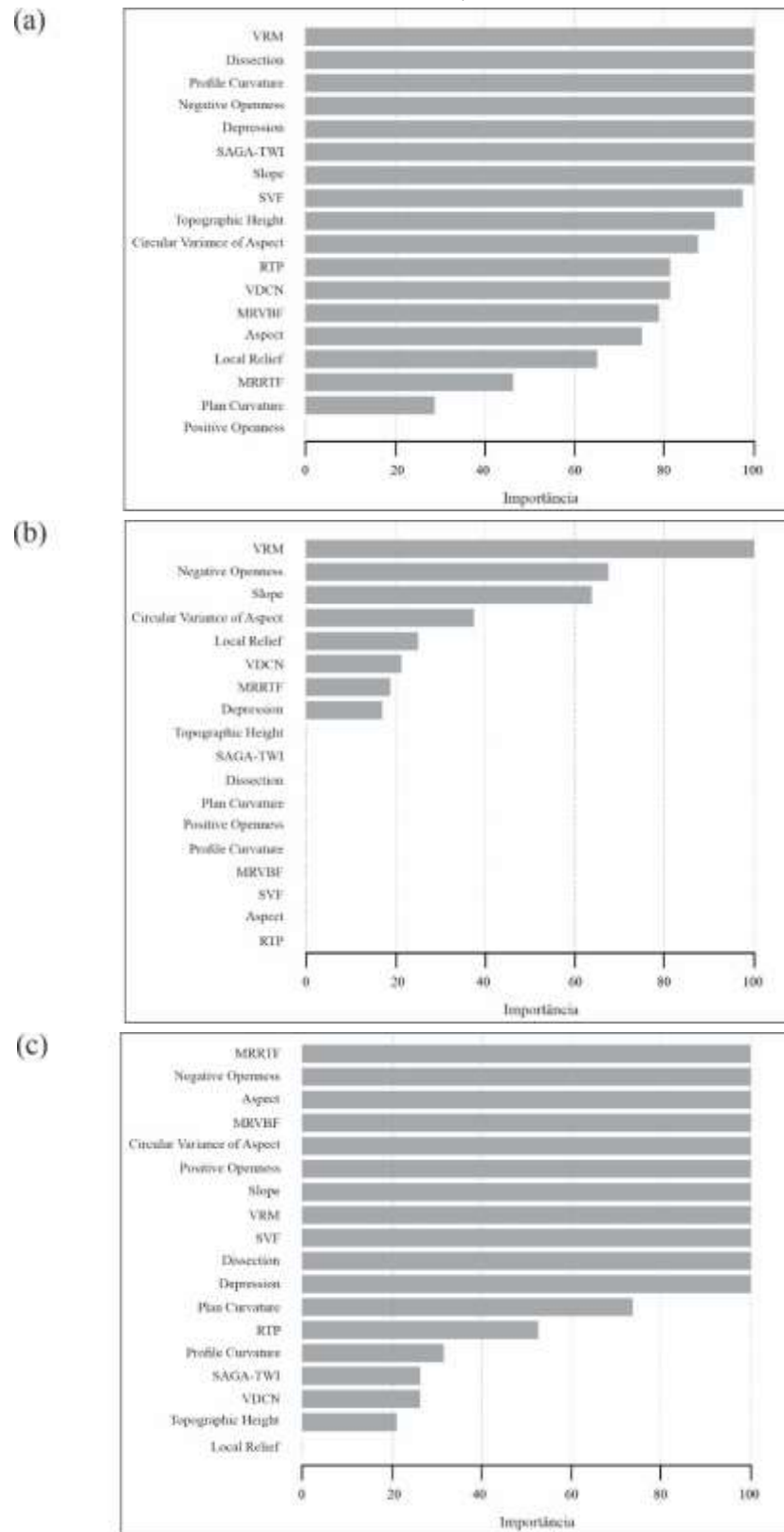
Tabela 4 – Valores de *F-score* dos algoritmos C5.0, RF e XGBoost, nos três cenários.

Cenário / Algoritmo	C5.0	RF	XGBoost
Cenário 1	0,831	0,866	0,831
Cenário 2	0,750	0,831	0,785
Cenário 3	0,771	0,771	0,694

A seguir são demonstrados os *rankings* de importância das covariáveis topográficas. Uma matriz de medida de importância, as linhas representam cada variável preditoras e as colunas as diferentes medidas de frequências. A porcentagem de frequência, representa o número relativo de vezes que um recurso foi usado nas árvores.

No algoritmo C5.0 (Figura 13) as cinco covariáveis mais preditoras no cenário 1 foram: *VRM*, *Dissection*, *Profile curvature*, *Negative openness* e *Depression*, com o mesmo grau de importância, seguidos ainda por *SAGA-TWI* e *Slope*. No cenário 2, assim como no cenário 1, a covariável *VRM* se destaca como a mais importante, seguida por *Negative openness*, *Slope*, *Circular variance of aspect* e *Local relief*. No cenário 3, a covariável *MRRTF* aparece em primeira colocada no *ranking*, entretanto as covariáveis *Negative openness*, *Aspect*, *MRVBF*, *Circular variance of aspect*, *Positive openness*, *Slop*, *VRM*, *VSF*, *Dissection* e *Depression* também aparecem na mesma categoria de importância.

Figura 13 – Importâncias das covariáveis topográficas para a predição de sítios arqueológicos usando o classificador C5.0, nos três cenários.

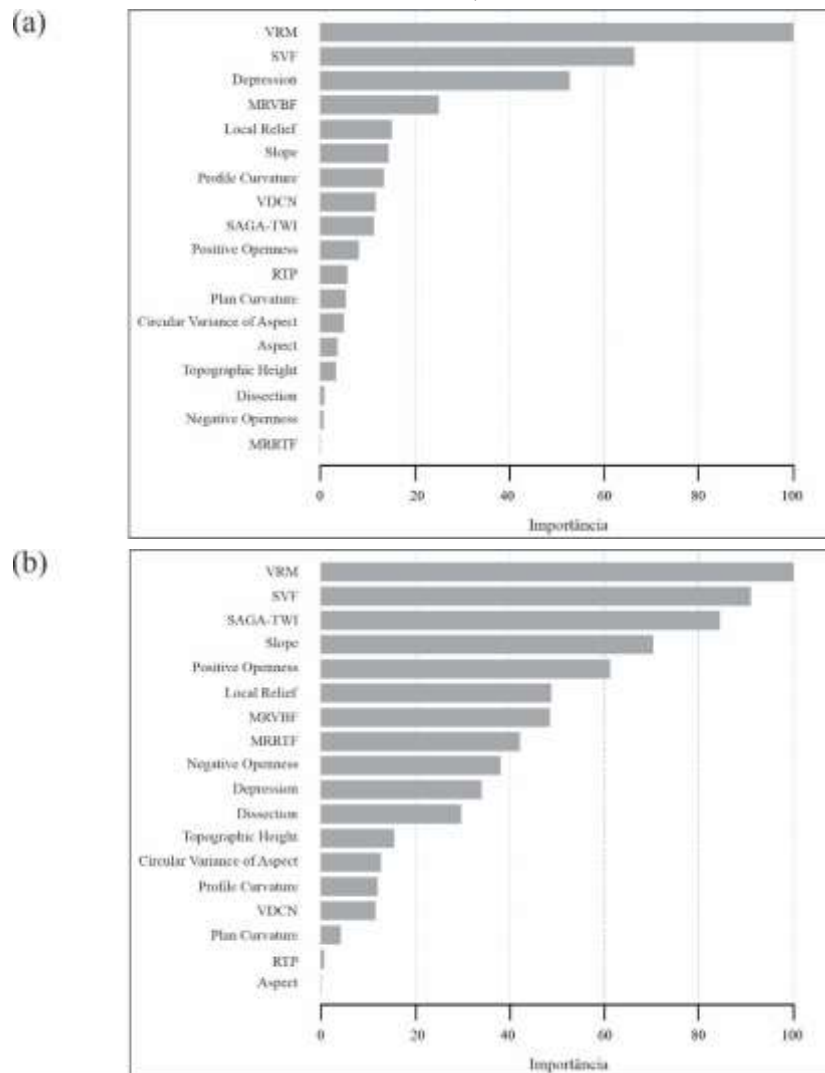


Fonte: A autora (2022).

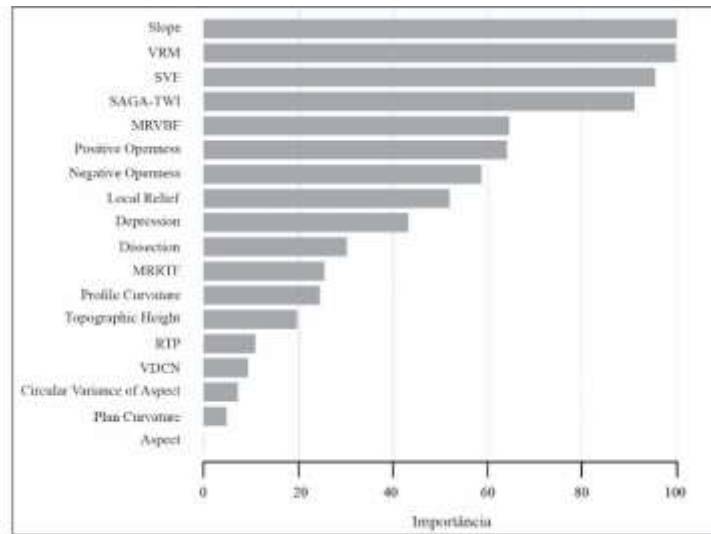
Nota: Atributos topográficos mais preditores no cenário 1 (a), cenário 2 (b) e cenário 3 (c) para o algoritmo C5.0.

As covariáveis mais importantes, no classificador RF são demonstradas na Figura 14. No cenário 1 destaca-se as covariáveis *VRM*, *SVF*, *Depression*, *MRVBF* e *Local relief*, como as cinco mais preditoras. No cenário 2, as covariáveis mais preditoras foram: *VRM*, *SVF*, *SAGA-TWI*, *Slope* e *Positive openness*. No cenário 3, nota-se que as covariáveis *Slope*, *VRM*, *SVF*, *SAGA-TWI* se repetem dentre as cinco mais preditoras, acrescido da *MRVBF*.

Figura 14 – Importâncias das covariáveis topográficas para a predição de sítios arqueológicas usando o classificador RF, nos três cenários.



(c)



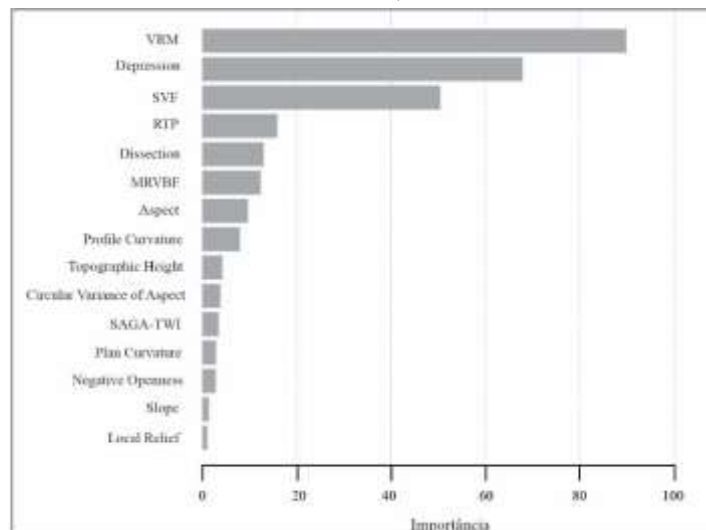
Fonte: A autora (2022).

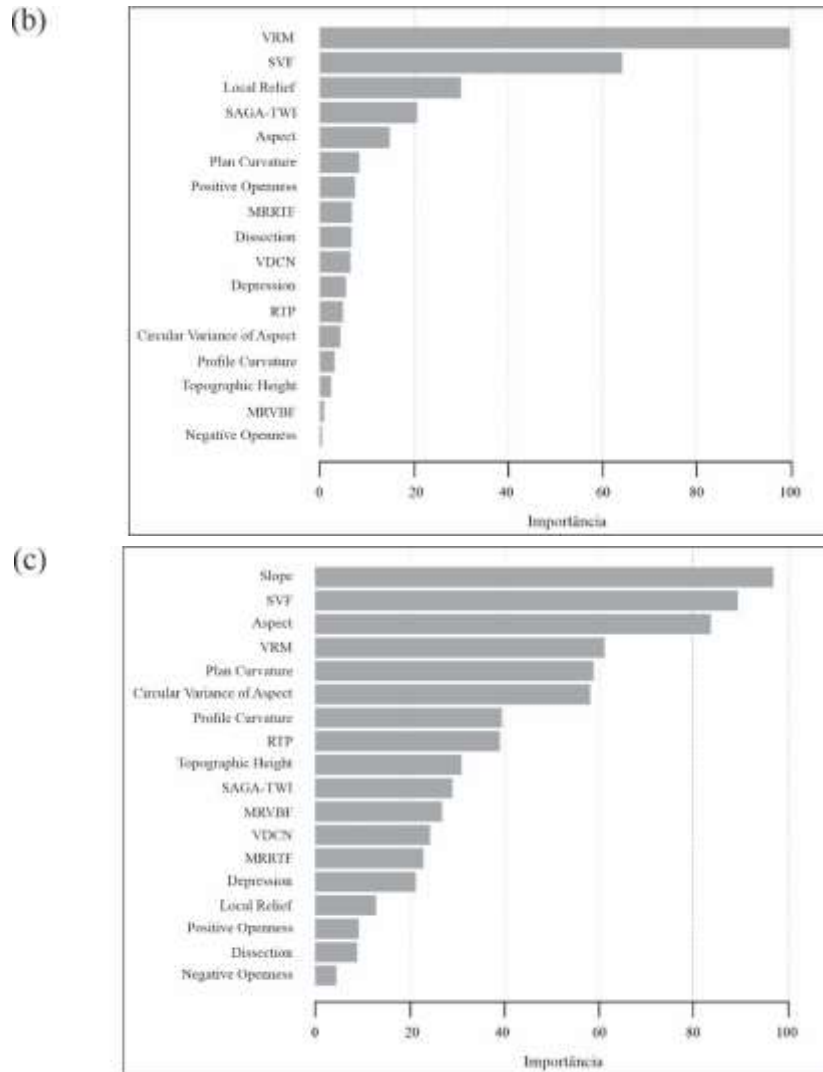
Nota: Atributos topográficos mais preditores no cenário 1 (a), cenário 2 (b) e cenário 3 (c) para o algoritmo RF.

Da mesma forma, na Figura 15, o classificador XGBoost cria uma matriz de importância dos recursos. No cenário 1, as cinco principais covariáveis mais preditoras foram *VRM*, *Depression*, *SVF*, *RTP* e *Dissection*. No cenário 2 assim como no cenário 1, as covariáveis *VRM* e *SVF* se destacam entre as cinco mais preditoras, acrescido ainda das covariáveis *Local relief*, *SAGA-TWI* e *Aspect*. No cenário 3, nota-se que a variável *Slope* se destaca como sendo a mais preditora, seguida por *SVF*, *Aspet*, *VRM* e *Plan curvature*.

Figura 15 – Importâncias das covariáveis topográficas para a predição de sítios arqueológicas usando o classificador XGBoost, nos três cenários.

(a)





Fonte: A autora (2022).

Nota: Atributos topográficos mais preditores no cenário 1 (a), cenário 2 (b) e cenário 3 (c) para o algoritmo XGBoost.

Em relação a distribuição estatística das covariáveis mais importantes na predição dos sítios arqueológicos, dado pelos gráficos de *box-plot* e *violin plot* considerando o cenário 1 (Figura 16), para a covariável *Slope* (Figura 16-a) observa-se que a distribuição da densidade de probabilidade do conjunto de Pseudoausência de sítios arqueológicos (esquerda) apresenta uma menor variabilidade de dispersão, com locação dos dados variando entre 0 e 10% de declividade, e maior densidade acima da mediana. O conjunto de dados de Presença (direita) apresenta uma distribuição bimodal e alta dispersão dos valores, pois o gráfico mostra-se mais alongado, com locação dos dados variando entre 10 e 40% de declividade. Observa-se que o gráfico se apresenta mais largo abaixo da mediana.

A distribuição da densidade de probabilidade da covariável *VRM* (Figura 16-b) no conjunto de Pseudoausência (esquerda) apresenta uma menor variabilidade de dispersão, com locação dos dados variando entre 0 e 0,005. A região que apresenta mais densa está acima da

mediana. Quanto ao conjunto de dados de Presença (direita), apresenta uma alta dispersão dos valores, pois o gráfico mostra-se mais alongado, com dados variando entre 0,015 e 0,025. Nota-se que a forma do gráfico apresenta mais arredondada no entorno da mediana, portanto com maior densidade de probabilidade nessa região. Isso sugere que a ocorrência dos sítios arqueológicos está em áreas mais rugosas topograficamente considerando a covariável *VRM*.

A covariável *Aspect* (Figura 16-c) apresenta a distribuição da densidade de probabilidade do conjunto de Pseudoausência (esquerda) com locação dos dados variando entre 150 e 300° e nota-se que a forma do gráfico apresenta mais arredondada no entorno da mediana. Enquanto que o conjunto de dados de Presença (direita) apresenta uma distribuição bimodal com locação dos dados entre 50 e 350°, no entanto apresenta forma mais larga, ou seja, com maior densidade de pontos, acima da mediana.

Local relief (Figura 16-d) apresenta a variabilidade de dispersão dos dados no conjunto de Pseudoausência (esquerda) variando entre 0 e 50 m, com região mais larga abaixo da mediana, próximo a 25 m. Enquanto que no conjunto de dados de Presença (direita) a locação dos dados, varia entre 50 e 150 m, com maior densidade abaixo da mediana. Isso mostra que os sítios arqueológicos de Serranópolis se concentram em locais com maior amplitude de relevo, não ocorrendo em áreas mais planas.

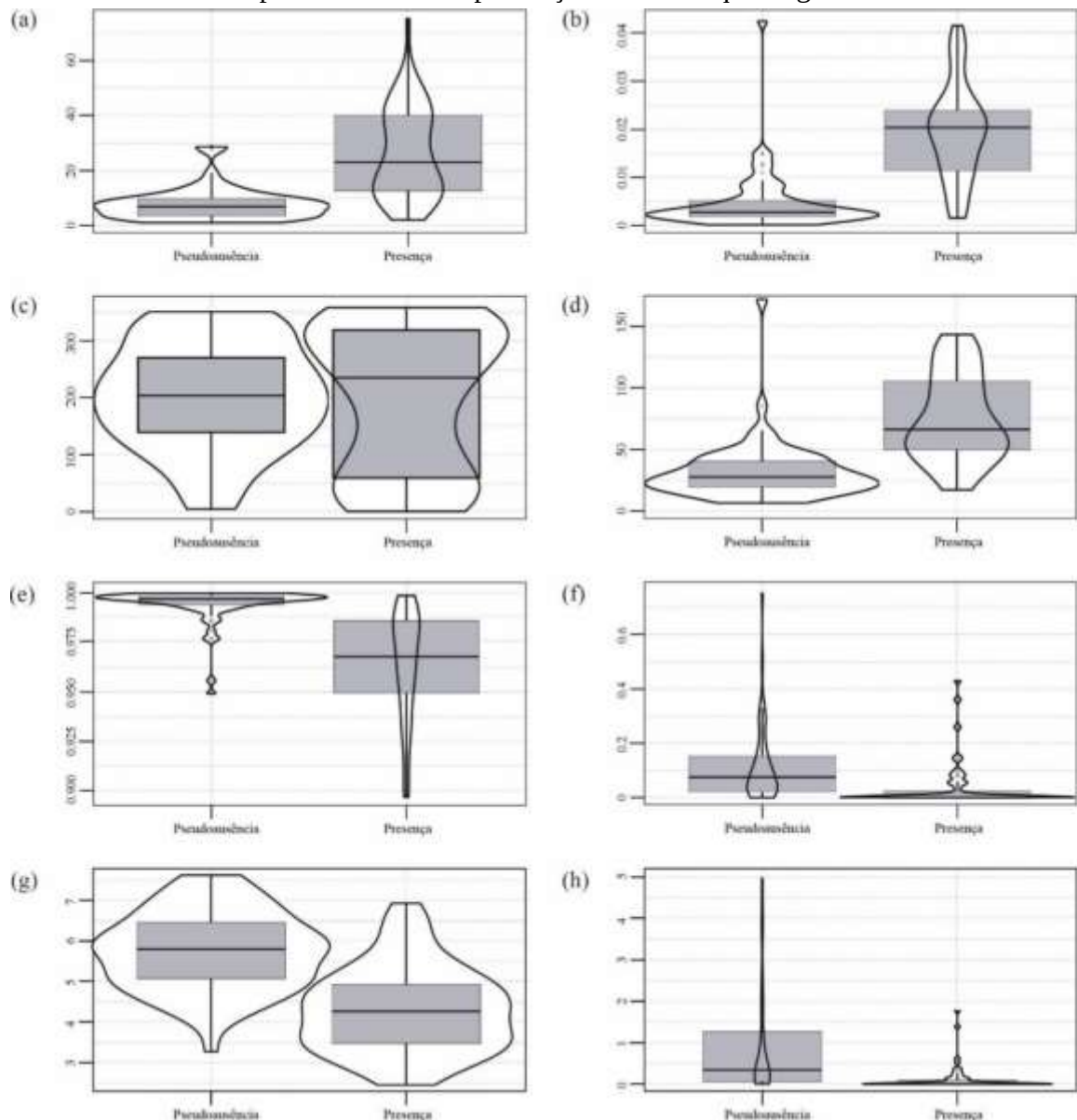
A distribuição da densidade de probabilidade da covariável *SVF* (Figura 16-e) no conjunto de Pseudoausência (esquerda) apresenta uma menor variabilidade de dispersão, mas uma alta frequência, variando entre 0,988 e 1. O conjunto de dados de Presença (direita) apresenta alta dispersão dos valores, pois o gráfico mostra-se mais alongado, apresentando a região que apresenta maior densidade de probabilidade acima da mediana, com locação dos dados variando entre 0,950 e 0,988, indicando que os sítios não ocorrem em regiões com total luminosidade.

Depression apresenta uma maior variabilidade de dispersão no conjunto de Pseudoausência (esquerda) com locação dos dados variando entre 0 e 0,1 (Figura 16-f) e região mais larga abaixo da mediana. Enquanto que no conjunto de dados de Presença (direita) apresenta uma menor variabilidade e com densidade, próxima a 0 (Figura 16-f), indicando assim que os sítios não ocorrem em áreas com depressão topográfica.

A covariável *SAGA-TWI* (Figura 16-g) apresenta a distribuição assimétrica da densidade de probabilidade do conjunto de Pseudoausência (esquerda) com locação dos variando entre 5 e 6,5, e maior probabilidade acima da mediana. Quanto ao conjunto de dados de Presença (direita) a locação dos dados varia entre 3 e 5 e com maior probabilidade abaixo da mediana, indicando que a presença dos sítios está em áreas menos úmidas.

A covariável *MRRTF* (Figura 16-h) apresenta locação dos dados de Pseudoausência (esquerda) variando entre 1 e 0,5, com região de maior probabilidade abaixo da mediana. Enquanto o conjunto de dados de Presença (direita) apresenta uma menor densidade, com mediana próxima a 0, indicando assim que a presença dos sítios não ocorre em áreas de crista e divisores de drenagem.

Figura 16 – Gráficos de *Violin Plots* das variáveis mais preditoras com distribuição dos dados de pseudoausência e presença de sítios arqueológicos.



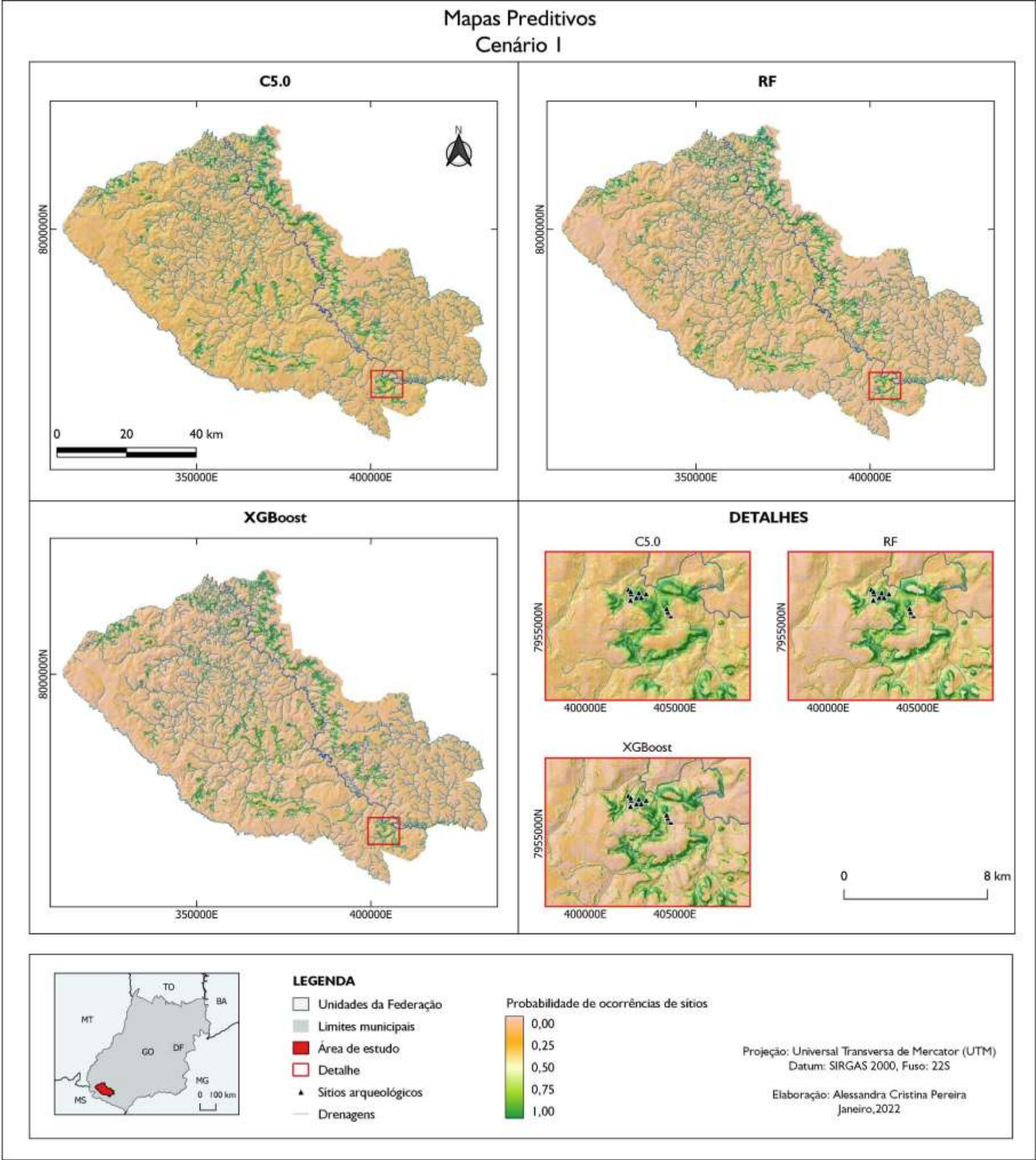
Fonte: A autora (2022).

Nota: Gráficos de *violin plot* da distribuição dos dados de pseudoausência (esquerda) e presença (direita) nas covariáveis topográficas mais preditoras: *Slope* (a), *VRM* (b), *Aspect* (c), *Local relief* (d), *SVF* (e), *Depression* (f), *SAGA-TWI* (g) e *MRRTF* (h).

As análises a seguir comparam os mapas finais preditos (Figura 17, Figura 18 e Figura 19) considerando os três cenários e os algoritmos classificadores proposto neste trabalho. Em detalhes, a comparação com algumas ocorrências de sítios arqueológicos, como referência, para cada algoritmo classificador.

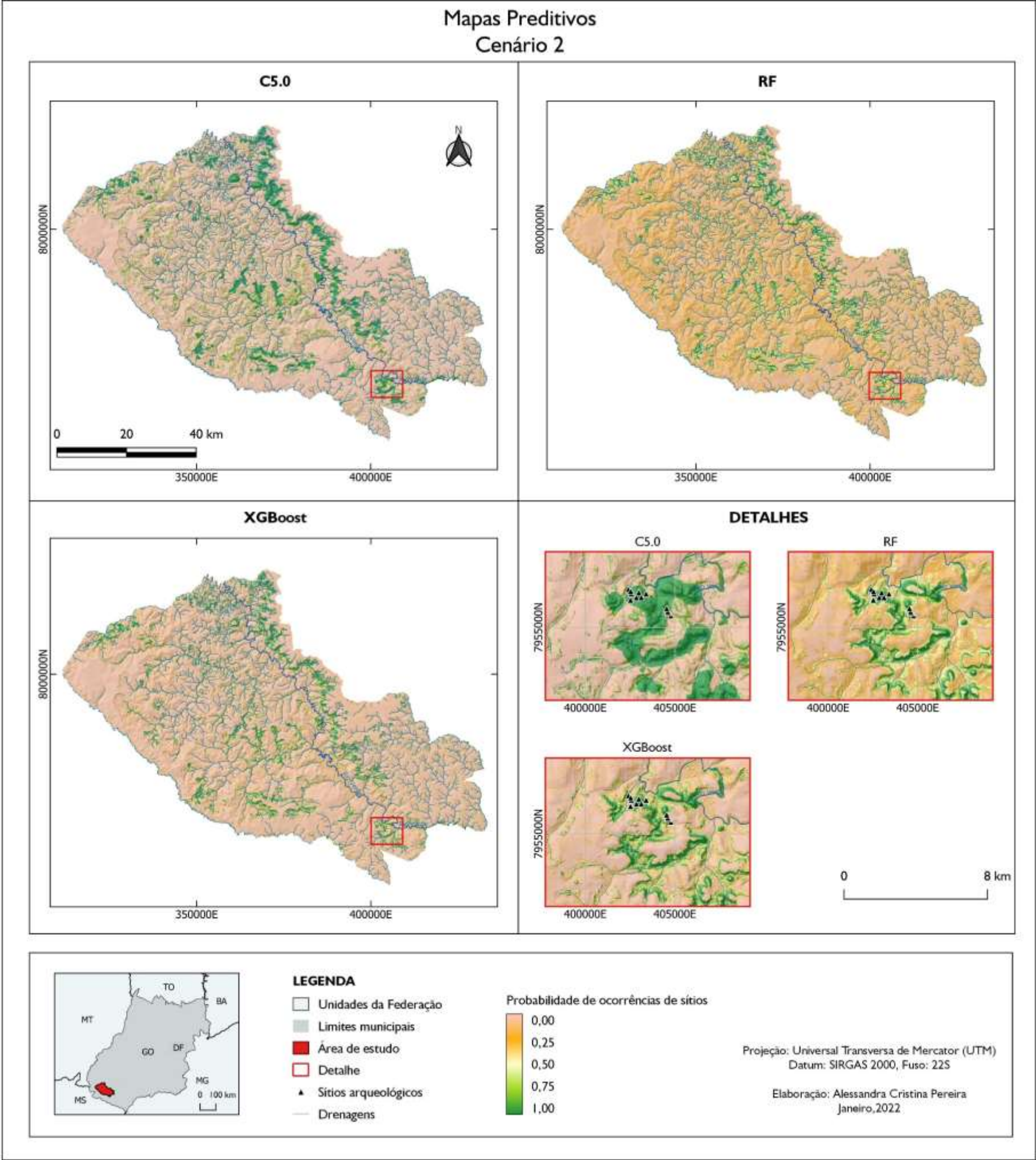
A probabilidade de ocorrência em uma determinada região varia de 0 a 1, onde 0 é a menor probabilidade e 1 a maior probabilidade. Para facilitar a visualização, foi definida uma rampa de cor. Os resultados dos classificadores são consistentes e corroboram com os resultados das métricas de performance.

Figura 17 – Mapas preditivos de possíveis locais com sítios arqueológicos no cenário1, para os algoritmos C5.0, RF e XGBoost.



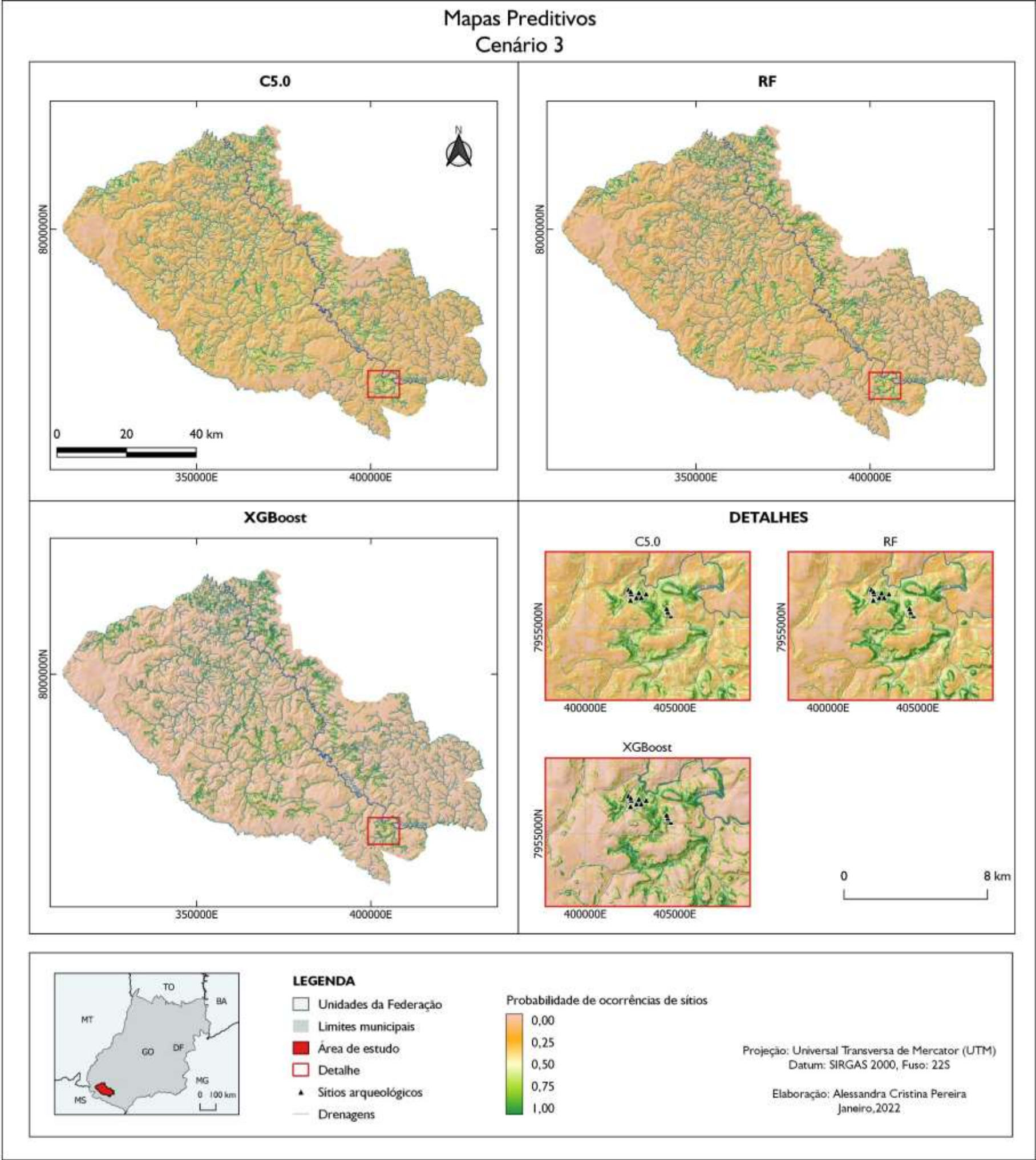
Fonte: A autora (2022).

Figura 18 – Mapas preditivos de possíveis locais a encontrar sítios arqueológicos no cenário 2, para os algoritmos C5.0, RF e XGBoost.



Fonte: A autora (2022).

Figura 19 – Mapas preditivos de possíveis locais a encontrar sítios arqueológicos no cenário 3, para os algoritmos C5.0, RF e XGBoost.



Fonte: A autora (2022).

7. DISCUSSÃO

Quanto as características definidas mais relevantes a este estudo, predominam os sítios do tipo abrigo sob rocha, utilizados por populações no período pré-colonial, com um grau de relevância científica de média a alta. Segundo Silva (2007), Serranópolis caracteriza-se pela presença de abrigos naturais e o interesse pela ocupação, estaria relacionada a oferta de matéria-prima para confecção de ferramentas e artefatos (instrumentos líticos) muito utilizada pelos grupos de caçadores-coletores.

Quanto a análise da curva ROC, observa-se que o RF em todos os cenários, se encontra mais próxima ao eixo y, demonstrando que esse algoritmo apresentou a taxa de sensibilidade um pouco melhor que os demais modelos, o que corrobora com as análises dos valores da AUC no qual, demonstra também que apresentou os melhores resultados, o que significa que este algoritmo apresenta uma habilidade preditiva mais alta, para esse estudo.

Em relação as métricas da sensibilidade, especificidade e *F-score*, nota-se que o RF apresentou o melhor resultado no cenário 1. Constatou-se, que os diferentes arranjos espaciais das amostras, influenciaram diretamente nos resultados das modelagens, uma vez que os algoritmos classificadores, utilizados neste estudo, apresentaram resultados distintos, embora aproximados. Acerca deste ponto, percebe-se que os únicos dados de entrada diferentes foram os pontos de pseudoausência, e a falta de dados de entrada de ausência verdadeira de sítios arqueológicos, sendo necessário inferir pontos de pseudoausência, podem influenciar na classificação.

De modo geral, todos os classificadores tiveram bons desempenhos, mostrando que a estratégia de *ensemble* dos algoritmos baseados em árvore de decisão representa uma importante ferramenta para predição (GHIMIRE *et al.*, 2012; WEN; HUGHES 2020; MOSAVI *et al.*, 2021). O melhor desempenho do classificador RF, baseado em *bagging*, em relação aos classificadores C5.0 e XGBoost, baseados em *boosting*, não é uma novidade na literatura. O mesmo ocorreu em estudo efetuado por Mosavi *et al.* (2021), entretanto isso não é uma regra. Há estudos no qual o desempenho de algoritmos baseados em *boosting* possuem desempenho superior do que os de *bagging* para classificação binária (SINGHAL *et al.*, 2018). Logo, sempre é importante avaliar mais casos, mas no momento o RF tem se mostrado uma poderosa ferramenta preditiva para sítios arqueológicos como demonstrado neste trabalho e em outros (YAWORSKY *et al.*, 2020; CASTIELLO; TONINI, 2021).

No que se refere as covariáveis geomorfológicas mais importantes para prever locais de sítios arqueológicos, neste estudo, destaca-se *VRM*, apontada pelos modelos de *machine learning*, como a mais preditiva. Outras também se destacaram: *Slope*, *SVF*, *Depression*, *SAGA-TWI*, *Local relief*, *aspect*, e *MRRTF*.

VRM quantifica a rugosidade do terreno, onde uma menor rugosidade representa nenhuma variação do terreno, ao passo que rugosidade máxima representa uma maior variação do terreno. *Slope* (declividade) representa a inclinação da superfície no terreno, enquanto *Aspect* (orientação de vertente) mostra a direção da inclinação. Segundo Zheng *et al.* (2020), variáveis como *Slope* e *Aspect* são utilizadas em muitos modelos preditivos arqueológicos, uma vez que estudos comprovam que estas feições topográficas foram usadas como fatores determinantes para escolha de local para assentamento.

O *SVF* calcula o fator de visibilidade do céu, baseada na iluminação difusa, influenciada pelo relevo. Zakšek *et al.* (2011) destacam que é uma ferramenta muito útil em arqueologia, pois melhora o reconhecimento das características do relevo, bem como é um fator importante para estimativa da radiação solar. Terrenos mais abertos, como em feições expostas (planos e picos) são mais visíveis e, portanto, recebem maior quantidade de luminosidade provenientes do céu claro, enquanto terrenos com relevos mais profundos e partes mais baixas de vales, por exemplo, quase nenhum céu é visível, logo, recebem pouca radiação solar. Uma variável que pode ser relacionada ao *SVF* é o *MRRTF*, pois esse índice identifica áreas planas em grandes altitudes.

As covariáveis geomorfológicas *Depression* e *Dissection* destaca a unidade de relevo de baixa altitudes. *SAGA-TWI* é um índice topográfico muito utilizado para caracterizar a hidrologia, representada pelo *déficit* de umidade. *Local relief* gera um modelo do relevo local, acentuando a visibilidade das depressões, independentemente do ângulo de iluminação. Segundo Hesse (2010), o uso de um modelo utilizando o *Local relief*, vem sendo empregado para prospecção arqueológica onde há variações de relevo positivo e/ou negativo a exemplo também das covariáveis *Positive openness* e *Negative openness* utilizadas nesse trabalho.

É importante salientar que a região de Serranópolis se encontra inserida em chapadões, resultantes dos processos erosivos de arenitos da Formação Botucatu e áreas de aplainamento caracterizado por morros tabulares de bordas escarpadas, vales encaixados e drenagem de média densidade. No contexto arqueológico, sobretudo os sítios do tipo abrigo sob rocha, se beneficiam do recuo das escarpas pela erosão diferencial. Logo podemos aferir que os sítios

arqueológicos estão localizados em regiões onde o relevo é bem delimitado e com menor índice de umidade, o que justifica o *rol* destas covariáveis como as mais importantes.

No que tange as análises exploratórias dos dados estatísticas de distribuição das densidades da probabilidade do conjunto de dados de pseudoausência e presença de sítios arqueológicos nas covariáveis mais preditoras, observa-se quanto a variabilidade de dispersão e o intervalo interquartil dos dados, que o conjunto ‘Presença’, apresentou-se maior, no geral, em relação ao conjunto ‘Pseudoausência’, com exceção nas covariáveis *Depression* e *MRRTF*. Quanto a locação dos dados (mediana) e assimetria, no geral, não houve similaridade entre os conjuntos. A análise gráfica permite inferir que os sítios arqueológicos estão localizados em áreas que não são cristas ou divisores de drenagem, são áreas secas, bem drenadas, e estão associadas às áreas topograficamente rugosas, declivosas e com maior amplitude de relevo, indicando um fato de proteção para a população que vivia nesses locais no tempo passado.

Para cada modelagem foi gerado um mapa preditivo. O desempenho do classificador avaliado por meio das informações das curvas ROC e a avaliação da qualidade do ajuste dos mapas preditivos analisado por meio da AUC, demonstram que os valores destas métricas foram muito próximos, isso reflete na interpretação visual das predições tendo que se atentar principalmente às análises estatísticas e aos detalhes de uma determinada região, para definir qual modelo apresentou o melhor resultado.

Analisando os detalhes dos mapas classificados e comparando-os às reais ocorrências de sítios arqueológicos, nota-se que estas coincidiram em regiões com maiores probabilidades de ocorrência de sítios arqueológicos (áreas verdes), o que demonstra que as performances dos classificadores, foram satisfatórias.

8. CONCLUSÃO

Os modelos para prever potenciais regiões de ocorrências de sítios arqueológicos, foram implementados nos algoritmos C5.0, RF e XGBoost, em três cenários, considerando todas as 18 covariáveis geomorfométricas. Desta forma, com o uso de SIG e *machine learning* foi possível neste estudo, estabelecer um modelo preditivo que melhor se adequasse as variáveis de entrada, na região de Serranópolis.

Analisando estas variáveis topográficas, constatou-se que isoladamente não são capazes de visualizar todas as características de relevo conhecidas e que, portanto, a combinação dessas covariáveis permite obter uma quantidade melhor de informações quanto a potenciais regiões com sítios arqueológicos.

Outro ponto importante a se destacar, foi a utilização de métricas de validação que permitissem avaliar graficamente e estatisticamente o desempenho e a eficiência desses classificadores. Estas análises, constatou que o RF apresentou o melhor resultado, o que significa que esse algoritmo apontou ser mais adequado para modelagem preditiva arqueológica neste estudo, sobretudo no cenário 1.

Acerca dos diferentes cenários, podemos afirmar que os arranjos espaciais das amostras, influenciaram diretamente nos resultados da modelagem dos diferentes algoritmos classificadores, uma vez que houve resultados distintos, ainda que pequenas variações. Contudo, a abordagem de modelagem com o maior poder preditivo (RF), não é uma resposta definitiva, ele apenas define um roteiro. É uma ferramenta extremamente útil que pode ser utilizada para auxiliar os pesquisadores nas prospecções e estudos arqueológicos.

Importante frisar quanto as informações dos sítios, que o total utilizado para este estudo (54) não esgota a possibilidade da existência de mais ocorrência na região, visto que o a extensa área territorial do município de Serranópolis ainda é objeto de estudos, o que faz que este sistema seja retroalimentado a cada campanha de pesquisa, em virtude do seu potencial arqueológico.

Tendo em vista a significância cultural para a compreensão dos processos ocupacionais do planalto Central Brasileiro, almeja-se que este modelo preditivo (RF no cenário 1), aqui apresentado, poderá contribuir para as pesquisas nos âmbitos acadêmicos, de arqueologia preventiva e na gestão ambiental e patrimonial de áreas arqueológicas, reforçando assim a importância de pesquisas interdisciplinares.

REFERÊNCIAS

- AIRBUS. **WorldDEM™: Technical Product Specification, version 2.5**. Airbus Defence and Space. 2019. Disponível em: https://airbusus.com/wp-content/uploads/2020/06/r51492_9_2019-04_worldDEM_technicalspecs_version2.5_i1.0.pdf. Acesso em: 27 jun. 2021.
- AIRBUS. **Copernicus Digital Elevation Model – Product Handbook**. Airbus Defence and Space. 2020. Disponível em: https://spacedata.copernicus.eu/documents/20126/0/GEO1988-CopernicusDEM-SPE-002_ProductHandbook_I1.00.pdf. Acesso em: 27 jun. 2021.
- ANDRADES FILHO, C.O. ROSSETTI, D.F.; BEZERRA, F.H.R.; MEDEIROS, W.E.; VALERIANO, M.M.; CREMON, É.H.; OLIVEIRA, R.G. Mapping Neogene and Quaternary sedimentary deposits in northeastern Brazil by integrating geophysics, remote sensing and geological field data. **Journal of South American Earth Sciences**, v. 56, p. 316-327, 2014.
- ANGELUCCI, D.E. A partir da Terra: a contribuição da Geoarqueologia. In: José E. Mateus e Marta Moreno-García (org.). **Paleoecologia Humana e Arqueociências. Um Programa Multidisciplinar para a Arqueologia sob a Tutela da Cultura**. Lisboa: Ed. Trabalhos de Arqueologia 29, 2003. p. 36-84. Disponível em: https://www.researchgate.net/publication/306253823_Paleoecologia_Humana_e_Arqueociencias_Um_programa_multidisciplinar_para_a_arqueologia_sob_a_tutela_da_Cultura. Acesso em: 26 jun. 2021.
- ARAÚJO, R.P. Por uma abordagem geoarqueológica: A interface entre a arqueologia e as ciências da Terra e a ênfase no contexto espacial. **Caderno do Lepaarq**, v. 15, n. 29, p. 20-39, 2018. Disponível em: <https://periodicos.ufpel.edu.br/ojs2/index.php/lepaarq/article/download/12298/8464>. Acesso em: 05 jul. 2021.
- AZEVEDO, L.L.A. de. **Métodos Estatísticos em Aprendizado de Máquinas para problemas de Classificação**. 2018. Dissertação (Mestrado em Estatística) – Universidade de Brasília, Brasília, Distrito Federal, 2018. Disponível em: https://repositorio.unb.br/bitstream/10482/33025/1/2018_LuanaL%C3%BAciaAlvesdeAzev%C3%AAdo.pdf. Acesso em: 28 jun. 2021.
- BASTOS, D.G.D., NASCIMENTO, P.S., LAURETTO, M.S. Análises empíricas de desempenho de quatro métodos de Seleção de características para Random Forest. **Revista Brasileira de Sistemas de Informação**, Rio de Janeiro, v. 7, n. 2, p. 25-57, 2014. Disponível em: <http://www.seer.unirio.br/isy/article/download/3309/3842/22210>. Acesso em: 15 jul. 2021.
- BITENCOURT, A.L.V. Princípios, métodos e algumas aplicações da geoarqueologia. In: Julio Cezar Rubin de Rubin e Rosiclér Theodoro da Silva (org). **Geoarqueologia: Teoria e Prática**. Goiânia: Ed. da UGC, 2008. p. 41-69.
- BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 2016. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 05 mai. 2021.

BRASIL. Ministério do Turismo. Instituto do Patrimônio Histórico e Artístico Nacional – IPHAN. **Ofício n.717/2021/COTEC IPHAN-GO/IPHAN-GO-IPHAN**. Goiânia, GO: Ministério do Turismo, 23 ago. 2021. Disponível em: https://sei.iphan.gov.br/sei/modulos/pesquisa/md_pesq_processo_pesquisar.php?acao_externa=protocolo_pesquisar&acao_origem_externa=protocolo_pesquisar&id_orgao_acesso_externo=0. Acesso em: 24 ago. 2021.

BREIMAN, L. Random Forest. **Machine Learning**, v. 45 (1), p. 5-32, 2001. Disponível em: <https://link.springer.com/article/10.1023/A:1010933404324>. Acesso em: 27 mai. 2021.

BUTZER, K.W. **Archaeology as human ecology: Method and theory for a contextual approach**. Cambridge: Cambridge University Press, p. 35-42. 1982. Disponível em: cambridge.org/core/books/archaeology-as-human-ecology/C7ED4A520324242015F56A2CA8EF6BB7. Acesso em: 26 jun. 2021.

CÂMARA, G.; DAVIS, C. Introdução à ciência da geoinformação. In: Gilberto Câmara, Clodoveu Davis e Antônio Miguel Vieira Monteiro (org.). **Introdução à ciência da geoinformação**. São José dos Campos: INPE, 2004. p. 1-5. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/introd/index.html>. Acesso em: 27 mai. 2021.

CÂMARA, G.; MONTEIRO, A.M.V. Conceitos básicos em ciência da geoinformação. In: Gilberto Câmara, Clodoveu Davis e Antônio Miguel Vieira Monteiro (org.). **Introdução à ciência da geoinformação**. São José dos Campos: INPE, 2004. p. 2-35. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/introd/index.html>. Acesso em: 27 mai. 2021.

CAMARGO, F.F.; ALMEIDA, C.M.; COSTA, G.A.O.P.; FEITOSA, R.Q.; OLIVEIRA, D.A.B.; HEIPKE, C.; FERREIRA, R.S. An open source object-based framework to extract landform classes, **Expert Systems with Applications**, v. 39 (1), p. 541-554, 2012.

CARDOSO, J.L. **A Geoarqueologia: fundamentos e métodos. Sua aplicação em Portugal**. Almada: Al-madan, n. 5 (II), p. 70-77. 1996. Disponível em: https://www.researchgate.net/publication/267514004_A_geoarqueologia_fundamentos_e_metodos_Sua_aplicacao_em_Portugal. Acesso em: 26 jun. 2021.

CASTIELLO, M.E.; TONINI, M. An Explorative Application of Random Forest Algorithm for Archaeological Predictive Modeling. A Swiss Case Study. **Journal of Computer Applications in Archaeology**, 4(1), p.110–125. 2021. Disponível em: <http://doi.org/10.5334/jcaa.71>. Acesso em: 07 dez. 2021.

CHAGAS, C.da S.; FERNANDES FILHO, E.I.; VIEIRA, C.A.O.; SCHAEFER, C.E.G.R.; CARVALHO JÚNIOR, W.de. Atributos topográficos e dados do LandSat7 no mapeamento digital de solos com uso de redes neurais. **Embrapa Solos**, Brasília, v.45, p.497-507. 2010. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/858546/atributos-topograficos-e-dados-do-landsat-7-no-mapeamento-digital-de-solos-com-uso-de-redes-neurais>. Acesso em: 11 jul. 2021.

CHEN, T.; GUESTRIN, C. *XGBoost: A Scalable Tree Boosting System*. In: *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016. p.785–794. San Francisco. **Anais [...]**. Disponível em: <https://doi.org/10.1145/2939672.2939785>. Acesso em: 07 dez. 2021.

CREMON, E.H.; PEREIRA, A.C; PAULA, L.D.L de; NUNES, E.D. Geological and terrain attributes for predicting soil classes using pixel- and geographic object-based image analysis in the Brazilian Cerrado. **Geoderma**, v. 401, 2021, 115315. Disponível em: <https://doi.org/10.1016/j.geoderma.2021.115315>. Acesso em: 06 jul. 2021.

DIETRICH, H.; BÖHNER, J. Cold Air Production and Flow in a Low Mountain Range Landscape in Hessia (Germany). **SAGA - Seconds Out, Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie**, p. 42, 2008. Disponível em: http://sourceforge.net/projects/saga-gis/files/SAGA%20-%20Documentation/HBPL19/hbpl19_05.pdf/download?use_mirror=kent Acesso em: 04 jun. 2021.

DINIZ, J.M.F.S. **Avaliação do potencial dos dados polarimétricos sentinel-1A para mapeamento do uso e cobertura da Terra na região de Ariquemes - RO**. 2019. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, São Paulo, 2019. Disponível em: <http://mtc-m21c.sid.inpe.br/col/sid.inpe.br/mtc-m21c/2019/01.28.12.01/doc/publicacao.pdf>. Acesso em: 27 mai. 2021.

DORNIK, A.; DRĂGUȚ, L.; URDEA, P. Classification of Soil Types Using Geographic Object-Based Image Analysis and Random Forests. **Pedosphere**, v. 28, n. 6, p. 913-925, 2018. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1002016017603771>. Acesso em: 04 jun. 2021.

ESCOVEDO, T.; KOSHIYAMA, A. **Introdução a Data Science: Algoritmos de Machine Learning e métodos de análises**. São Paulo: Casa do Código, 2020. E-book.

ESPINOSA-ZUNIGA, J. J. Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. **Ingeniería, Investigación y Tecnología**, Ciudad de México, v.21, n.3, dez. 2020. Disponível em: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-77432020000300002&lng=es&nrm=iso. Acesso em: 15 dez. 2021.

ESRI. **ArcGIS Blog**: How to extract raster values at point locations? 2017. Página inicial. Disponível em: <https://www.esri.com/arcgis-blog/products/arcgis-pro/analytics/how-to-extract-raster-values-at-point-locations/>. Acesso em: 20 dez. 2021.

EUROPEAN SPACE AGENCY - ESA. **Copernicus Global Digital Elevation Model**. OpenTopography. Sinergise, 2021. Disponível em: <https://doi.org/10.5069/G9028PQB>. Acesso em: 27 jun. 2021.

FERREIRA, E.V. **Machine Learning para cientista de dados: Métodos de reamostragem**: Notas de Aula. Vídeo aula. Laboratório de Estatística e Geoinformação/Departamento de Estatística da Universidade Federal do Paraná (LEG/UFPR). 2018. Disponível em: <http://cursos.leg.ufpr.br/ML4all/>. Acesso em: 06 jul. 2021.

GALLANT, J.C.; WILSON, J.P. Digital terrain analysis. **Terrain Analysis: Principles and Applications**, p. 1–27, 2000. Disponível em: <https://www.researchgate.net/publication/303543739>. Acesso em: 04 jun. 2021.

GALLANT, J.C. Multi-Resolution Index of Valley Bottom Flatness for Mapping Depositional Areas. **Water Resources Research**, v. 39, n. 12, p.1347, 2003. Disponível em: https://www.researchgate.net/publication/248808482_A_Multi-Resolution_Index_of_Valley_Bottom_Flatness_for_Mapping_Depositional_Areas. Acesso em: 27 dez. 2021.

GARCIA, G.P.B. **Elaboração de mapas geomorfológicos a partir de Modelos Digitais de Elevação**. 2015. Dissertação (Mestrado em Recurso Minerais e Hidrologia) – Instituto de Geociências, Universidade de São Paulo, São Paulo, 2015. Disponível em: https://www.teses.usp.br/teses/disponiveis/44/44137/tde-17112015-150344/publico/Mestrado_Guilherme.pdf. Acesso em: 04 jul. 2021.

GHIMIRE, B.; ROGAN, J.; GALIANO, V.R.; PANDAY, P.; NEETI, N. An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA. **GIScience Remote Sens**, v. 49, p. 623–643, 2012.

GISLASON, P.O.; BENEDIKTSSON, J. A.; SVEINSSON, J. R. Random Forest for land cover classification. **Pattern Recognition Letters, Elsevier**, v. 27, n. 4, p. 294–300, 2006. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0167865505002242>. Acesso em: 28 mai. 2021.

GLADFELTER, B.G. Developments and directions in geoarchaeology. In: Michael B. Schiffer (org.). **Advances in Archaeological Method and Theory**, v. 4, p. 343-364. 1981. Disponível em: <https://www.sciencedirect.com/book/9780120031047/advances-in-archaeological-method-and-theory#book-info>. Acesso em: 26 jun. 2021.

GOIÁS. Secretaria de Estado de Meio Ambiente e Desenvolvimento Sustentável – SEMAD. **Comitês de Bacia Hidrográfica**. 2021. Disponível em: <https://www.meioambiente.go.gov.br/conselhos-e-c%C3%A2maras/comit%C3%AAs-das-bacias-hidrogr%C3%A1ficas.html>. Acesso em: 05 fev. 2021.

GROHMAN, C.H.; SMITH, M.J.; RICCOMINI, C. Multiscale analysis of topographic surface roughness in the Midland Valley, Scotland. **IEEE Transactions on Geoscience and Remote Sensing**, v. 49, n. 4, p. 1200-1213, 2010. Disponível em: https://www.researchgate.net/publication/224163264_Multiscale_Analysis_of_Topographic_Surface_Roughness_in_the_Midland_Valley_Scotland. Acesso em: 27 dez. 2021.

GUERRA, A.J.T. **Um estudo do meio físico com fins de aplicação ao planejamento do uso agrícola da Terra no Sudoeste de Goiás**. Rio de Janeiro: IBGE, Departamento de Recursos Naturais e Estudos Ambientais, 1989. 210p. (Projeto Cerrado II; Convênio IBGE/Embrapa). Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=282387>. Acesso em: 28 fev. 2021.

HÄMMERLE, M.; GÁL, T.; UNGER, J.; MATZARAKIS, A. Comparison of models calculating the sky view factor used for urban climate investigations. **Theoretical and Applied Climatology**, v. 105, n. 3, p. 521-527, 2011. Disponível em: https://www.researchgate.net/publication/226606861_Comparison_of_models_calculating_the_sky_view_factor_used_for_urban_climate_investigations. Acesso em: 27 dez. 2021.

HESSE, R. LiDAR-derived Local Relief Models – a new tool for archaeological prospection. **Archaeological Prospection**, v. 17, n. 2, p. 67-72, 2010. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/arp.374>. Acesso em: 06 jan. 2022.

INFANTINI, L. Sistemas de Informação Geográficos (SIG) em Arqueologia. **Revista Arqueologia Pública**, Campinas, v. 9, n. 3 (13), p. 114–121, 2015. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rap/article/view/8641308>. Acesso em: 1 jul. 2021.

INSTITUTO DO PATRIMÔNIO HISTÓRICO E ARTÍSTICO NACIONAL – IPHAN. Em Serranópolis (GO), estudos arqueológicos vão ampliar o conhecimento sobre o Brasil central. **IPHAN-Goiás**, Goiânia, 28 out. 2020. 2020a. Disponível em: <http://portal.iphan.gov.br/go/noticias/detalhes/5816/em-serranopolis-go-estudos-arqueologicos-vao-ampliar-o-conhecimento-sobre-o-brasil-central>. Acesso em: 15 dez. 2020.

INSTITUTO DO PATRIMÔNIO HISTÓRICO E ARTÍSTICO NACIONAL – IPHAN. **Carta de Lausanne, 1990**. IPHAN-Goiás, 2020b. Disponível em: <http://portal.iphan.gov.br/uploads/ckfinder/arquivos/Carta%20de%20Lausanne%201990.pdf>. Acesso em: 15 jan. 2021.

KÄMPF, N.; KERN, D.C. O solo como registro da ocupação humana pré-histórica na Amazônia. **Tópicos em Ciência do solo**, Viçosa: Sociedade Brasileira de Ciência do Solo, ed. 1 v. 4, p. 277-320, 2005. Disponível em: <https://rosalux.org.br/wp-content/uploads/2015/10/Dados-TPA-Top-V4-N7.pdf>. Acesso em: 05 jul. 2021.

KUHN, M.; WING, J.; WESTON, S.; WILLIAMS, A. KEEFER, C.; ENGELHARDT, A.; COOPER, T.; MAYER, Z.; KENKEL, B.; R Core Team; BENESTY, M.; LESCARBEAU, R.; ZIEM, A.; SCRUCICA, L.; TANG, Y.; CANDAN, C.; HUNT, T. **Caret: classification and regression training. R package (version 6.0-90)**. 2021. Disponível em: <https://cran.r-project.org/web/packages/caret/index.html>. Acesso em: 18 jan. 2022.

LIMA, A.P.G. de. **Patrimônio arqueológico de Serranópolis/GO: Reflexões para a busca de uma tutela compartilhada e efetiva**. 2016. Tese (Doutorado em Arqueologia) – Programa de Pós-Graduação do Museu de Arqueologia e Etnologia da Universidade de São Paulo, São Paulo, 2016. Disponível em: <https://teses.usp.br/teses/disponiveis/71/71131/tde-29082016-151736/publico/TeseAnaPaulaLimaFINALREVISADA.pdf>. Acesso em: 15 jan. 2021.

MARIANO, Z.F. **A importância da variável climática na produtividade de soja no Sudoeste de Goiás**. 2005. Tese (Doutorado em Geografia) – Instituto de Geociências e Ciências Exatas, Rio Claro, São Paulo, 2005. Disponível em: https://bdtd.ibict.br/vufind/Record/UNSP_5e1ad35890ba91f90edab2de27217327. Acesso em: 26 jun. 2021.

MATOS, P.F.; LOMBARDINI, L. de O.; CIFERRI, R.R.; PARDO, A.S.; CIFERRI, C.D. de A.; VIEIRA, M.T.P. **Relatório técnico “Conceito sobre aprendizado de máquinas”**. p. 6-14, 2009. Disponível em: <https://sites.icmc.usp.br/taspardo/techreportufscar2009b-matosetal.pdf>. Acesso em 07 jul. 2021.

MENESES, P.R. Formato das imagens de sensoriamento remoto. *In: Paulo Roberto Meneses e Tati de Almeida (org.). Introdução ao processamento de imagens de sensoriamento remoto*. Brasília, 2012, p. 77 - 81.

MENESES, P.R.; SANO, E.E. Classificação pixel a pixel de imagens. *In: Paulo Roberto Meneses e Tati de Almeida (org.). Introdução ao processamento de imagens de sensoriamento remoto*. Brasília, 2012, p. 191-208.

MOSAVI, A.; SAJEDI HOSSEINI, F. ; CHOUBIN, B. et al. Ensemble Boosting and Bagging Based Machine Learning Models for Groundwater Potential Prediction. **Water Resour. Manage**, v 35, p. 23–37, 2021.

MOORE, I.D.; GRAYSON, R.B.; LADSON, A.R. Digital Terrain Modelling: a review of hydrological, geomorphological, and biological applications. **Hydrological Processes**, v. 5, p. 3-30, set. 1991. Disponível em: <https://pdfs.semanticscholar.org/bfbd/cb38c3ad4d86c0ad5c8f3d05db8a4897dfcb.pdf> Acesso em: 04 jun. 2021.

MUÑOZ, V.A. **Análise Geomorfométrica de dados SRTM aplicada ao estudo das relações solo-relevo**, 2009. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2009. Disponível em: <http://urlib.net/sid.inpe.br/mtc-m18@80/2009/03.25.15.09>. Acesso em: 04 nov. 2021.

MUÑOZ, V.A.; VALERIANO, M.M. Mapeamento de planície de inundação por processamento de dados de elevação de sensoriamento remoto. *In: Eulógio Pardo-Igúzquiza, Carolina Guardiola-Albert, Javier Heredia, Luis Moreno-Merino, Juan Durán, José Antonio Vargas-Guzmán. Mathematics of Planet Earth (org.). Lecture Notes in Ciências da Terra sistema, Springer*. Berlim: Heidelberg, 2014, p. 481-484. Disponível em: <https://link.springer.com/book/10.1007/978-3-642-32408-6>. Acesso em: 06 jul. 2021.

NAZARENO, N.R.X. de. **SIG arqueologia: aplicação em pesquisa arqueológica**. 2005. Tese (Doutorado em Arqueologia) – Museu de Arqueologia e Etnologia, Universidade de São Paulo, São Paulo, 2005. Disponível em: <https://www.teses.usp.br/teses/disponiveis/71/71131/tde-22082006-144612/en.php>. Acesso em: 01 jul. 2021.

NEWMAN, D.R., LINDSAY, J.B., COCKBURN, J.M.H. Evaluating metrics of local topographic position for multiscale geomorphometric analysis. **Geomorphology**, v. 312, n. 1, p. 40-50, 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0169555X18301508>. Acesso em: 27 dez. 2021.

NOGUEIRA, R.A.S. Arqueologia da paisagem, Serranópolis na interpretação dos espaços sociais. **Revista Habitus**, Goiânia, v. 13, n. 1, p. 89-112, 2015. Disponível em: <http://seer.pucgoias.edu.br/index.php/habitus/article/download/4244/2435>. Acesso em: 15 jan. 2021.

PALMIERI, F.; LARACH, J.O.I. Pedologia e geomorfologia. *In: Antonio José Teixeira Guerra e Sandra Baptista da Cunha (org.). Geomorfologia e meio ambiente*. Rio de Janeiro: Bertrand Brasil, 2010, ed. 9, p. 74-90.

PAVANELLI, J.A.P. **Integração de dados ópticos e SAR para caracterização da paisagem de tensão ecológica em Roraima**. 2016. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, São Paulo, 2016. Disponível em: <http://mtc-m21b.sid.inpe.br/col/sid.inpe.br/mtc-m21b/2016/02.26.16.17/doc/publicacao.pdf>. Acesso em: 28 mai. 2021.

QGIS Development Team. **QGIS Geographic Information System (versão 3.16)**. 2021. Disponível em: <http://qgis.osgeo.org>. Acesso em: 01 ago. 2021.

RILEY, S.J.; DeGLORIA, S.D.; ELLIOT, R.A. Terrain Ruggedness Index that Quantifies Topographic Heterogeneity. **Internation Journal of Science**, v. 5, n. 1-4, p. 23-27, 1999. Disponível em: https://www.researchgate.net/publication/259011943_A_Terrain_Ruggedness_Index_that_Quantifies_Topographic_Heterogeneity. Acesso em: 29 mai. 2021.

RODRIGUEZ-GALIANO, V.F.; GHIMIRE, B.; ROGAN, J.; CHICA-OLMO, M.; RIGOL-SANCHEZ, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 67, p. 93–104, 2012. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0924271611001304>. Acesso em: 27 mai. 2021.

ROKACH, L. **Ensemble Learning: Pattern Classification Using Ensemble Methods**. Negev: Word Scientific, 2 ed., 2019.

RUBIN, J.C.R.; VIANA, S.A.; SILVA, R.T.; BARBERI, M.; RESENDE, F.E.C.P.e.; RIBEIRO-FREITAS, J.E.; SOUZA, M.G.; VARGAS, E.R. Cazadores-recolectores y el paisaje en Serranópolis, Goiás, Brasil. **Boletín de Arqueología PUC Peru**. n. 29, p. 129-158, 2020.

RUIZ, L.F.C.; GUASSELLI, L.A.; CATEN, A.T. Árvore de decisão e análise baseada em objetos na classificação de imagens com resolução especial submétrica adquiridas por VANT. **Boletim de Ciências Geodésicas**, 2017. Disponível em: <https://www.scielo.br/j/bcg/a/jV8P7fqTn5NszbGJnvFDTCh/?lang=pt>. Acesso em: 15 dez. 2021.

SANTIAGO, L.F.P.; SILVA, W.B. Avaliação de métodos de classificação por regiões em imagens de alta resolução do sensor Rapideye. **Revista Militar de Ciências e Tecnologia**, Rio de Janeiro, v. 35, n. 2, p. 22-26, 2018. Disponível em: http://rmct.ime.eb.br/arquivos/RMCT_2_tri_2018/RMCT_37817.pdf. Acesso em: 10 jul. 2021.

SAPPINGTON, J.M.; LONGSHORE, K.M.; THOMPSON, D.B. Quantifying Landscape Ruggedness for Animal Habitat Analysis: A Case Study Using Bighorn Sheep in the Mojave Desert. **Journal of Wildlife Management**, v. 71 (5), p. 1419–1426, 2007. Disponível em: <https://wildlife.onlinelibrary.wiley.com/doi/abs/10.2193/2005-723>. Acesso em: 29 mai. 2021.

SCHMITZ, P.I. **Caçadores e coletores antigos no Sudeste, Centro-Oeste e Nordeste do Brasil**. São Leopoldo: Instituto Anchietano de Pesquisas/IAP-UNISINOS, 1984.

SCHMITZ, P.I.; BARBOSA, A.S.; JACOBUS, A.L.; RIBEIRO, M.B. **Arqueologia nos cerrados do Brasil Central. Serranópolis I**. São Leopoldo: Instituto Anchieta de Pesquisas/IAP-UNISINOS, 1989.

SCHMITZ, P.I.; ROSA, A.O.; BITENCOURT, A.L.V. **Arqueologia nos cerrados do Brasil Central. Serranópolis III**. Pesquisas, Antropologia 60. São Leopoldo: Instituto Anchieta de Pesquisas, 2004.

SEQUEIRA, C.M.A.G. **Modelos preditivos em SIG na localização de sítios arqueológicos de cronologia mesolítica no Vale do Tejo**. 2009. Dissertação (Mestrado em Arqueologia) – Departamento de História, Arqueologia e Patrimônio da Faculdade de Ciências Humanas e Sociais, Universidade do Algarve, Faro, 2009. Disponível em: https://scholar.google.pt/citations?view_op=view_citation&hl=pt-PT&user=Zt1CsxUAAAAJ&citation_for_view=Zt1CsxUAAAAJ:9ZIFYXVOiuMC. Acesso em: 15 jan. 2021.

SINGHAL, Y.; JAIN, A.; BATRA, S.; VARSHNEY, Y.; RATHI, M. Review of bagging and boosting classification performance on unbalanced binary classification. *In: 2018 IEEE 8th International Advance Computing Conference (IACC)*, p. 338-343, 2018.

SILVA, R.T. da. Zoneamento geoarqueológico aplicado à gestão de recursos culturais. Tese (Doutorado em Geociências) – Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista, Rio Claro, 2007.

SILVA, J.X. da. O que é Geoprocessamento? **Revista do Crea-RJ**, Rio de Janeiro, v. 79, p. 42-44, 2009. Disponível em: <http://www.ufrj.br/lga/tiagomarino/artigos/oqueegeoprocessamento.pdf>. Acesso em: 27 mai. 2021.

SILVA, M. Matriz de Confusão e AUC ROC. **Data Hackers** – Medium. 2019. Disponível em: <https://medium.com/data-hackers/matriz-de-confus%C3%A3o-e-auc-roc-f7e446dca107>. Acesso em: 15 dez. 2021.

STUBERT, L.; OLIVEIRAS, A.M i.; MÄRKER, M.; SCHERNTHANNER, H.; VOGEL, S. Viticulture in the Laetanian Region (Spain) during the Roman Period: Predictive Modelling and Geomatic Analysis. **Geosciences**, v. 10, n. 6, p. 206. 2020. Disponível em: <https://doi.org/10.3390/geosciences10060206>. Acesso em: 07 dez. 2021.

THOMAZ, R.C.C.; IMAI, N.N.; SILVA, E.A.; TSUCHIYA, I. Visualização cartográfica como ferramenta de suporte à implementação de um modelo preditivo em arqueologia: um estudo de caso. *In: Simpósio Brasileiro de Geomática*, 2002, p.574-584, Presidente prudente. **Anais [...]**. Disponível em: https://www.researchgate.net/profile/Erivaldo-Silva/publication/268367231_VISUALIZACAO_CARTOGRAFICA_COMO_FERRAMENTA_DE_SUPORTE_A_IMPLEMENTACAO_DE_UM_MODELO_PREDITIVO_EM_ARQUEOLOGIA_UM_ESTUDO_DE_CASO/links/5631c31d08ae0530378d33f6/VISUALIZACAO-CARTOGRAFICA-COMO-FERRAMENTA-DE-SUPORTE-A-IMPLEMENTACAO-DE-UM-MODELO-PREDITIVO-EM-ARQUEOLOGIA-UM-ESTUDO-DE-CASO.pdf. Acesso em: 02 jul. 2021.

VALERIANO, M.M.; ROSSETTI, D.F. Regionalization of local geomorphometric derivations for geological mapping in the sedimentary domain of central Amazônia. **Computers & Geosciences**, v. 100, p. 46-56, 2017.

VARGAS, M. Clima. In: GUERRA, A.J.T. *et al.* **Um estudo do meio físico com fins de aplicação ao planejamento do uso agrícola da Terra no Sudoeste de Goiás**. Rio de Janeiro: IBGE, Departamento de Recursos Naturais e Estudos Ambientais, 1989.

WEN, L.; HUGHES, M. Coastal Wetland Mapping Using Ensemble Learning Algorithms: A Comparative Study of Bagging, Boosting and Stacking Techniques. **Remote Sensing**, v. 12, n.10, p. 1683, 2020.

WILSON, J.P. Digital terrain modeling. **Geomorphology**, v. 137, p. 107-121. 2012. Disponível em: <https://johnwilson.usc.edu/wp-content/uploads/2018/03/Digital-terrain-modeling.pdf>. Acesso em: 12 jul. 2021.

YAWORSKY, P.M.; VERNON, K.B.; SPANGLER, J.D.; BREWER, S.C.; CODDING, B.F. Advancing predictive modeling in archaeology: An evaluation of regression and machine learning methods on the Grand Staircase-Escalante National Monument. **PLoS ONE**, v. 15, n. 10. 2020. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0239424>. Acesso em: 07 dez. 2021.

YOKOYAMA, R., SHIRASAWA, M., PIKE, R.J. Visualizing topography by openness: a new application of image processing to digital elevation models. **Photogrammetric engineering and remote sensing**, v. 68, n. 3, p. 257-266, 2002. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.451.9798&rep=rep1&type=pdf>. Acesso em: 27 dez. 2021.

ZAKŠEK, K.; OŠTIR, K.; KOKALJ, Ž. Sky-View Factor as a Relief Visualization Technique. **Remote Sensing**, v. 3, n. 2, p.398-415, 2011. Disponível em: <https://www.mdpi.com/2072-4292/3/2/398/htm>. Acesso em: 06 jan. 2022.

ZHENG, M.; TANG, W.; OGUNDIRAN, A.; YANG, J. Spatial Simulation Modeling of Settlement Distribution Driven by Random Forest: Consideration of Landscape Visibility. **Sustainability**, v. 12, n. 11, p. 4748. 2020. Disponível em: <https://doi.org/10.3390/su12114748>. Acesso em: 12 jan. 2022.

ZURELL, D.; FRANKLIN, J.; KÖNIG, C.; BOUCHET, P.J.; DORMANN, C.F.; ELITH, J.; FANDOS, G.; FENG, X.; GUILLERA-ARROITA, G.; GUIAN, A.; LAHOZ-MONFORT, J.J.; LEITÃO, P.J.; PARK, D.S.; PETERSON, A.T.; RAPACCIUOLO, G.; SCHMATZ, D.R.; SCHRÖDER, B.; SERRA-DIAZ, J.M.; THUILLER, W.; YATES, K.L.; ZIMMERMANN, N.E.; MEROW, C. A standard protocol for reporting species distribution models. **Ecography**, v. 43, p. 1261-1277, 2020.

APÊNDICE A – Quadro de informações, relevantes ao estudo, dos sítios arqueológicos selecionados.

ID	CNSA	CNSA/GO	Núcleo	Nome do sítio	Sigla	Fuso	X	Y	Datum	X fornecido	Y fornecido	Datum fornecido	Categoria	Exposição	Relevância do sítio	Ano de registro
1	57705	Não informado	E	Abrigo da Pedra - Aglomerado Jerônimo Rufino	Aglomerado Jerônimo Rufino	22	402571,97	7956893,76	SIGAS2000	402621	7956938	SAD-69	Pré-colonial	Abrigo sob rocha	Média	2010
2	57707	Não informado	E	Abrigo do Morro - Aglomerado Jerônimo Rufino	Aglomerado Jerônimo Rufino	22	402526,97	7956540,76	SIGAS2000	402576	7956585	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	2010
3	45226	GO00147=GO00159	A	Diogo Lemes da Silva (1)	GO-JA-001	22	390023,96	7984900,88	SIGAS2000	390073	7984945	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
4	57609	Não informado	A	Diogo Lemes da Silva - GO-JA-01a	GO-JA-001a	22	389956,96	7984821,88	SIGAS2000	390006	7984866	SAD-69	Pré-colonial	Céu aberto	Alta	1975
5	57610	Não informado	A	Diogo Lemes da Silva - GO-JA-01b	GO-JA-001b	22	390098,96	7984923,88	SIGAS2000	390148	7984968	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
6	57612	Não informado	A	Diogo Lemes da Silva - GO-JA-01c	GO-JA-001c	22	390104,96	7984927,88	SIGAS2000	390154	7984972	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
7	57613	Não informado	A	Diogo Lemes da Silva - GO-JA-01d	GO-JA-001d	22	390111,96	7984930,88	SIGAS2000	390161	7984975	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
8	57614	Não informado	A	Diogo Lemes da Silva - GO-JA-01e	GO-JA-001e	22	390119,96	7984934,88	SIGAS2000	390169	7984979	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
9	57615	Não informado	A	Diogo Lemes da Silva - GO-JA-01f	GO-JA-001f	22	390295,96	7984975,88	SIGAS2000	390345	7985020	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
10	57616	Não informado	A	Diogo Lemes da Silva - GO-JA-01g	GO-JA-001g	22	390358,96	7984971,88	SIGAS2000	390408	7985016	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
11	57617	Não informado	A	Diogo Lemes da Silva - GO-JA-01h	GO-JA-001h	22	389941,96	7984743,88	SIGAS2000	389991	7984788	SAD-69	Pré-colonial	Céu aberto	Alta	1975
12	45227	GO00148	A	Diogo Lemes da Silva (2) - Gruta do Diogo II	GO-JA-002a	22	389682,96	7984539,88	SIGAS2000	389732	7984584	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
13	56380	GO00150=GO00805	D	Manoel Braga (1) Abrigo 1001	GO-JA-003	22	394353,00	7960300,00	WGS-84	394353	7960300	WGS-84	Pré-colonial	Abrigo sob rocha	Alta	1975
14	57733	GO00802	A	GO-JA-03A	GO-JA-003a	22	394296,05	7960385,12	SIGAS2000	394349	7960388	Córrego Alegre	Pré-colonial	Abrigo sob rocha	Média	1999
15	45230	GO00151	D	Manoel Braga (2)	GO-JA-004	22	394917,00	7960866,00	WGS-84	394917	7960866	WGS-84	Pré-colonial	Abrigo sob rocha	Alta	1975
16	57460	GO00152	D	Manoel Braga (3)	GO-JA-004a	22	395003,00	7960891,00	WGS-84	395003	7960891	WGS-84	Pré-colonial	Abrigo sob rocha	Média	1975
17	57461	GO00153	C	Gruta do Japonês	GO-JA-005, ex-Fazenda Lindomar Ribeiro Franco	22	401927,99	7972340,79	SIGAS2000	401977	7972385	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
18	56892	GO00156	A	Reduzino (1)	GO-JA-008	22	390002,97	7982917,86	SIGAS2000	390052	7982962	SAD-69	Pré-colonial	Abrigo sob rocha	Média	1975

ID	CNSA	CNSA/GO	Núcleo	Nome do sítio	Sigla	Fuso	X	Y	Datum	X fornecido	Y fornecido	Datum fornecido	Categoria	Exposição	Relevância do sítio	Ano de registro
9	57462	GO00157	A	Reduzino (2)	GO-JA-008a	22	390233,97	7983207,86	SIGAS2000	390283	7983252	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
2	57618	Não informado	A	Fazenda Canguçu	GO-JA-008b	22	390731,97	7983056,86	SIGAS2000	390781	7983101	SAD-69	Histórico	Céu aberto	Média	2010
21	45240	GO00160	F	Aglomerado Simplício Canjerana (1)	GO-JA-010	22	404775,97	7955671,77	SIGAS2000	404825	7955716	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
22	45241	GO00161	C	Lindomar Ribeiro Franco (1)	GO-JA-011-1a	22	401445,99	7972969,79	SIGAS2000	401495	7973014	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
23	57464	GO00162	C	Lindomar Ribeiro Franco (2)	GO-JA-011-2b	22	401213,99	7972738,79	SIGAS2000	401263	7972783	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
24	45244	GO00164	F	Aglomerado Simplício Canjerana (2)	GO-JA-012	22	404647,97	7955829,77	SIGAS2000	404697	7955874	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
25	45245	GO00165	E	Aglomerado Jerônimo Rufino (1)	GO-JA-013	22	402409,97	7957136,76	SIGAS2000	402459	7957181	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
26	57466	GO00166	E	Aglomerado Jerônimo Rufino (2)	GO-JA-013c	22	402537,97	7957021,76	SIGAS2000	402587	7957066	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
27	45247	GO00167	F	Aglomerado Simplicio Canjerana (3)	GO-JA-014	22	404590,97	7955916,77	SIGAS2000	404640	7955961	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
28	45248	GO00168	F	Aglomerado Simplício Canjerana (4)	GO-JA-015	22	404607,97	7955899,77	SIGAS2000	404657	7955944	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1975
29	45249	GO00169	F	Aglomerado Simplício Canjerana (5)	GO-JA-015a, Lajedo	22	404545,97	7956063,77	SIGAS2000	404595	7956108	SAD-69	Pré-colonial	Céu aberto	Alta	1975
30	45253	GO00170	B	Aglomerado Jair Gonçalves (1)	GO-JA-020	22	397054,99	7972978,77	SIGAS2000	397104	7973023	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1976
31	45254	GO00171	B	Aglomerado Jair Gonçalves (2)	GO-JA-021, Casarão de Pedra	22	397009,99	7972645,77	SIGAS2000	397059	7972690	SAD-69	Pré-colonial	Abrigo sob rocha	Média	1976
32	57468	GO00172	B	Aglomerado Jair Gonçalves (3)	GO-JA-021a	22	396996,99	7972700,77	SIGAS2000	397046	7972745	SAD-69	Pré-colonial	Céu aberto	Alta	1982
33	45256	GO00173	B	Aglomerado Jair Gonçalves (4)	GO-JA-022	22	396800,99	7972163,77	SIGAS2000	396850	7972208	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1982
34	57471	GO00175	B	Jair Gonçalves de Assis	GO-JA-024	22	397256,99	7971959,77	SIGAS2000	397306	7972004	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1982
35	57472	GO00176	D	(Núcleo D)	GO-JA-025	22	394801,98	7960839,75	SIGAS2000	394851	7960884	SAD-69	Pré-colonial	Abrigo sob rocha	Média	1982
36	57473	GO00177	D	Manoel Braga (4)	GO-JA-026	22	394124,98	7960692,75	SIGAS2000	394174	7960737	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	1982
37	57474	GO00178	D	Manoel Braga (5)	GO-JA-027	22	394750,98	7960129,75	SIGAS2000	394800	7960174	SAD-69	Pré-colonial	Abrigo sob rocha	Média	1982
38	57475 = 45262	GO00179	D	Manoel Braga (6)	GO-JA-028	22	395966,98	7960669,75	SIGAS2000	396016	7960714	SAD-69	Pré-colonial	Abrigo sob rocha	Média	1982
39	Não informado	GO01268	Não informado	Sítio Maracajá	GO-JA-113	22	432583,88	7955120,99	SIGAS2000	432633	7955165	SAD-69	Pré-colonial	Céu aberto	Média	2010

ID	CNSA	CNSA/GO	Núcleo	Nome do sítio	Sigla	Fuso	X	Y	Datum	X fornecido	Y fornecido	Datum fornecido	Categoria	Exposição	Relevância do sítio	Ano de registro
40	Não informado	GO01272	Não informado	Sítio Ponte entre Rio	GO-JA-116	22	422318,92	7956611,89	SIGAS2000	422368	7956656	SAD-69	Histórico	Céu aberto	Média	2011
41	Não informado	GO01271	Não informado	Sítio Cemitério	GO-JA-118	22	431942,89	7955060,97	SIGAS2000	431992	7955105	SAD-69	Histórico	Céu aberto	Média	2011
42	57706	Não informado	C	Abrigo do Cachorro	Lindomar Ribeiro Franco (4)	22	401210,99	7972912,79	SIGAS2000	401260	7972957	SAD-69	Pré-colonial	Abrigo sob rocha	Alta	2010
43	56746	Não informado	Não informado	Rio Bonito	RB	22	388338,00	7954601,00	WGS-84	388338	7954601	WGS-84	Histórico	Céu aberto	Média	2005
44	Não informado	Não informado	Não informado	Gercilene 1	Gercilene 1	22	400669,94	7941830,74	SIGAS2000	400719	7941875	SAD-69	Pré-colonial	Céu aberto	Média	Não informado
45	Não informado	Não informado	Não informado	Gercilene 2	Gercilene 2	22	400325,94	7942280,74	SIGAS2000	400375	7942325	SAD-69	Pré-colonial	Céu aberto	Média	Não informado
46	Não informado	Não informado	A	A5	Não informado	22	388972,96	7985384,89	SIGAS2000	389022	7985429	SAD-69	Pré-colonial	Céu aberto	Baixa	Não informado
47	Não informado	Não informado	A	A6	Não informado	22	388871,97	7983154,87	SIGAS2000	388921	7983199	SAD-69	Pré-colonial	Céu aberto	Baixa	Não informado
48	Não informado	Não informado	A	A7	Não informado	22	388432,97	7982512,87	SIGAS2000	388482	7982557	SAD-69	Pré-colonial	Céu aberto	Baixa	Não informado
49	Não informado	Não informado	C	C5	Não informado	22	402149,98	7973288,8	SIGAS2000	402199	7973333	SAD-69	Pré-colonial	Abrigo sob rocha	Média	Não informado
50	Não informado	Não informado	D	D8	Não informado	22	396203,98	7960466,75	SIGAS2000	396253	7960511	SAD-69	Pré-colonial	Céu aberto	Média	Não informado
51	Não informado	Não informado	E	E3	Não informado	22	402858,97	7956682,76	SIGAS2000	402908	7956727	SAD-69	Pré-colonial	Abrigo sob rocha	Média	Não informado
52	Não informado	Não informado	E	E4	Não informado	22	402994,97	7956953,76	SIGAS2000	403044	7956998	SAD-69	Pré-colonial	Céu aberto	Média	Não informado
53	Não informado	Não informado	E	E5	Não informado	22	403129,97	7956682,76	SIGAS2000	403179	7956727	SAD-69	Pré-colonial	Abrigo sob rocha	Média	Não informado
54	Não informado	Não informado	E	E6	Não informado	22	403399,97	7956885,77	SIGAS2000	403449	7956930	SAD-69	Pré-colonial	Céu aberto	Média	Não informado

Fonte: A autora 2022.

Nota: Adaptado a partir de BRASIL (2021) e dos dados cedidos pelos pesquisadores do IGPA/PUC Goiás.

ANEXO A – Ofício nº 717/2021/COTEC IPHAN-GO/IPHAN-GO-IPHAN

30/09/2021 09:29

SEI/IPHAN - 2906160 - Ofício



MINISTÉRIO DO TURISMO
SECRETARIA ESPECIAL DE CULTURA
INSTITUTO DO PATRIMÔNIO HISTÓRICO E ARTÍSTICO NACIONAL
Coordenação Técnica do IPHAN-GO

Ofício Nº 717/2021/COTEC IPHAN-GO/IPHAN-GO-IPHAN

Ao Senhor
ÉDIPO HENRIQUE CREMON
Professor Doutor
Instituto Federal de Educação, Ciência e Tecnologia de Goiás - IFG-GO
Rua 75, nº 46 - Setor Central
74.055-110 - Goiânia-GO
E-mail: edipo.cremon@ifg.edu.br

Assunto: Dados dos sítios arqueológicos localizados em Serranópolis-GO

Referência: Caso responda este, indicar expressamente o Processo nº 01516.000352/2021-51.

Senhor Professor,

Em atendimento à solicitação apresentada no Ofício nº 001/2021 *Instituto Federal Goiás* (doc. IPHAN/SEI! 2806170), inserimos no presente processo, para download, as fichas de cadastro de sítio arqueológico situados no município de Serranópolis-GO (doc. IPHAN/SEI! 2903836), levantadas nos arquivos desta Superintendência do Iphan em Goiás. Para consultar os documentos, o interessado deverá acessar o [Sistema Eletrônico de Informação \(SEI\) do Iphan](#), clicar em [Consulte seu Processo](#) e digitar o número 01516.000352/2021-51 no campo "Nº do Processo ou Documento:", e o código de verificação no canto superior direito da página. O processo encontra-se aberto para consulta pública e o interessado poderá baixar o arquivo contendo os documentos solicitados - doc. IPHAN/SEI! 2903836, denominado "Ficha CNSA Sítios Arqueológicos Serranópolis".

Oportunamente, solicitamos encarecidamente o envio dos resultados dos estudos desenvolvidos pela aluna Alessandra Cristina Pereira, quando finalizados, especialmente as ferramentas e meios para a aplicação do modelo preditivo de sítios arqueológicos, que podem contribuir sobremaneira com as análises dos processos de licenciamento ambiental pelos arqueólogos desta Superintendência do Iphan-GO, bem como com a gestão do rico patrimônio arqueológico da região de Serranópolis-GO.

Atenciosamente,

ANA PAULA CARVALHO
Coordenadora Técnica substituta da Superintendência do Iphan em Goiás

30/09/2021 09:29

SEI/IPHAN - 2906160 - Ofício

ALLYSON RIBEIRO E SILVA CABRAL
Superintendente do Iphan em Goiás



Documento assinado eletronicamente por **Ana Paula Carvalho, Coordenadora Técnica Substituta do IPHAN-GO**, em 23/08/2021, às 10:05, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Allyson Ribeiro e Silva Cabral, Superintendente do IPHAN-GO**, em 23/08/2021, às 12:06, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site <http://sei.iphan.gov.br/autenticidade>, informando o código verificador **2906160** e o código CRC **0F8618E8**.

Praça Doutor Pedro Ludovico Teixeira Nº 210 Setor Central , Goiânia. CEP 74.003-010
Telefone: (62) 3224-6402 | Website: www.iphan.gov.br