

PREDIÇÃO DE OCORRÊNCIAS ARQUEOLÓGICAS EM OBRAS DE INFRAESTUTURA

AUTORES

José Antoniel Campos Feitosa

Árlon Facynek de Oliveira Carvalho

REALIZAÇÃO

RESUMO

O artigo apresenta um modelo matemático capaz de prever ocorrências de sítios arqueológicos em empreendimentos de infraestrutura, minorando assim o risco de comprometimento do patrimônio histórico nacional, bem como o de embargos por parte dos órgãos de controle, por inobservância da legislação. É mostrado o panorama atual das obras paralisadas no país, decorrente de pendências ambientais e arqueológicas. São analisadas as duas abordagens clássicas utilizadas na elaboração de modelos com esse fim, a indutiva e a dedutiva, descrevendo as potencialidades e limitações de cada uma, ressaltando-se a dificuldade de obtenção de dados completos na abordagem indutiva e a forte subjetividade presente na dedutiva. Em decorrência, é apresentada uma abordagem alternativa — e inovadora no contexto dos modelos de predição arqueológica —, baseada nas respostas da própria amostra, o que possibilitou elaborar um modelo com uma acurácia de 81%, que se caracteriza como de forte poder preditivo. Estudos de caso são apresentados onde a importância de um modelo preditivo é realçada.

Palavras-chave: modelo preditivo, sítios arqueológicos, regressão logística.

ABSTRACT

The article presents a mathematical model capable of predicting occurrences of archaeological sites in infrastructure enterprises, thus undergoing the risk of commitment of the national historical heritage, as well as that of embargoes by the control bodies, due to non-compliance with the legislation. The current panorama of the works paralyzed in the country is shown, due to environmental and archaeological pending. The two classic approaches used in the elaboration of models for this purpose, the inductive and the deductive, describe the potential and limitations of each one, emphasizing the difficulty of obtaining complete data in the inductive approach and the strong subjectivity present in the deductive. As a result, an alternative - and innovative approach is presented in the context of archaeological prediction models - based on sample responses, which made it possible to develop a model with an accuracy of 81%, which is characterized as strong predictive power. Case studies are presented where the importance of a predictive model is highlighted.

Keywords: predictive model, archaeological sites, logistics regression.

REALIZAÇÃO



1. INTRODUÇÃO

As ocorrências arqueológicas durante a execução de obras de infraestrutura tem sido uma realidade no país. A providência imediata a ser tomada, nos casos de achados fortuitos desses artefatos, é o imediato comunicado ao órgão responsável pelo patrimônio histórico e artístico nacional¹, no caso, ao IPHAN – Instituto do Patrimônio Histórico e Artístico Nacional. Como consequência, a área no entorno da ocorrência deve ser preservada, até que estudos de prospecção possam ser realizados por profissionais especializados, com posterior pronunciamento e deliberação da autoridade.

Em breve pesquisa na internet, pode-se colher informações que evidenciam tanto a regularidade das ocorrências arqueológicas, quanto as consequências daí advindas.

Quadro 1 – Ocorrências e consequências de achados arqueológicos

Data	UF	Manchete
out-19	PA	Órgão responsável pela obra disse que achado de possível sítio arqueológico teria embargado o empreendimento.
jan-21	SP	Justiça manda paralisar obras da linha 2-verde do Metrô de SP
nov-21	AL	Vestígios históricos encontrados no Centro podem ter de 100 a 150 anos
jun-22	PR	Iphan autoriza início da duplicação de trecho da PR-445; obras estão atrasadas por impasse com sítios arqueológicos
jun-22	SP	Metrô toca em sítio arqueológico; parada Saracura/Vai-Vai
ago-22	SC	Obras de pavimentação da estrada de Congonhas são embargadas pelo IPHAN
jan-23	RR	Trecho da BR-174/RR paralisado: necessidade de realizar acompanhamento arqueológico solicitado pelo Iphan
jan-23	PA	Sítio arqueológico: MPF pede paralisação de obras no PA entre Óbidos e Oriximiná

Fonte: elaborado pelo autor, a partir de pesquisa na internet

Fato é que, para o contexto da obra de infraestrutura em andamento, havendo ou não o resgate dos artefatos encontrados, a necessidade de proteção da área direta e indiretamente afetada, provocará impedimentos no andamento normal do empreendimento, com reflexos no equilíbrio contratual pactuado entre as partes, caso

¹ Lei 3.924/1961, Art 18. A descoberta fortuita de quaisquer elementos de interesse arqueológico ou pré-histórico, histórico, artístico ou numismático, deverá ser imediatamente comunicada à Diretoria do Patrimônio Histórico e Artístico Nacional, ou aos órgãos oficiais autorizados, pelo autor do achado ou pelo proprietário do local onde tiver ocorrido.

esse risco não esteja alocado a algum dos contratantes. Em todo caso, mesmo com a alocação do risco, o que supriria a resposta financeira, o impacto no tempo é uma consequência esperada e inevitável. Necessidade de novas programações para as frentes de obras, com deslocamentos adicionais de equipes e possíveis ressequenciamentos de atividades não estão descartadas, tudo em desfavor da produtividade inicialmente planejada.

Na outra ponta do problema, registre-se as inúmeras obras paralisadas decorrentes de pendências as mais variadas possíveis, e muitas delas relacionadas ao patrimônio histórico e artístico nacional, conforme será abordado adiante.

Nesse contexto, a acurácia de um modelo preditivo, ou seja, a sua capacidade de prever quanto a existência ou não de sítios arqueológicos nos locais onde serão executadas obras de infraestrutura, é de fundamental importância para o desempenho contratual, uma vez que um baixo poder preditivo do modelo fará com que a ocorrência arqueológica só se manifeste quando das escavações da referida obra, o que trará a necessidade de isolamento da área diretamente afetada, com reflexo no ritmo e na sequência das atividades previstas no cronograma do empreendimento.

Considerando este cenário, seria conveniente que essa possibilidade de se deparar com uma ocorrência arqueológica, durante a execução de uma obra de infraestrutura fosse, na impossibilidade de ser evitada, pelo menos mitigada.

Nesse sentido, é objetivo do presente artigo apresentar um modelo matemático de previsão de registros arqueológicos, capaz de identificar as áreas mais propensas a essas ocorrências, como medida primeira a ser tomada ainda na fase de definição da localização do empreendimento.

Segundo Tsuchiya (2002, p. 16), “Um modelo preditivo pode prever as prováveis áreas de ocupações pretéritas, baseando-se nas preferências do *habitat* e em variáveis do universo físico, tornando-se útil na etapa de prospecção.”

Para obtenção desse modelo, será revista a teoria que lhe respalda, notadamente as abordagens indutiva e dedutiva.

No tocante à metodologia proposta para elaboração de um modelo preditivo de ocorrências arqueológicas, será mostrado um viés inovador, baseado nas respostas vindas da amostra de estudo.

Dois estudos de caso serão apresentados, referentes à construção de um contorno rodoviário e a uma duplicação de rodovia, contextualizando tais obras de infraestrutura com as ocorrências arqueológicas, onde a importância de um modelo preditivo é evidenciada.

2. REFERENCIAL TEÓRICO

2.1 Da magnitude do problema a ser enfrentado

É praticamente impossível que obras de infraestrutura como construção ou adequação de rodovias ou ferrovias; utilização de recursos hídricos; geração,

transmissão e distribuição de energia elétrica; etc., não se depare com alguma ocorrência arqueológica na sua fase de execução.

O principal levantamento de obras paralisadas no país reporta-se ao conteúdo constante no Acórdão nº 1.079/2019 – TCU – Plenário, que analisou cerca de 38 mil obras, tendo constatado a paralisação, ou o baixíssimo andamento, de aproximadamente 14 mil obras, ou pouco mais de um terço de todas as obras analisadas, conforme resumo na tabela a seguir.

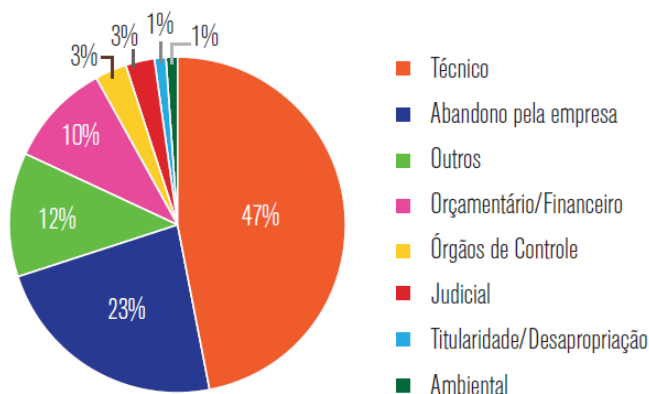
Tabela 1 – Situação das obras analisadas

Situação	Obras analisadas		Investimento inicialmente previsto	
	Qtde	% relativo	R\$	% relativo
Em reformulação	590	1,54%	204.988.942,85	0,03%
Adiantada	950	2,47%	957.961.227,72	0,13%
Atrasada	2.700	7,03%	4.105.680.314,13	0,57%
Normal/Em execução	19.728	51,36%	575.829.146.944,31	79,37%
Obra iniciada s/ medição	41	0,11%	44.541.721,12	0,01%
Paralisada/Inacabada	14.403	37,50%	144.314.132.476,62	19,89%
Total Geral	38.412	100,00%	725.456.451.626,74	100,00%

Fonte: Tribunal de Contas da União – Acórdão 1.079/2019 – TCU – Plenário

São R\$ 144 bilhões de obras paralisadas, o equivalente a aproximadamente 20% do total de recursos investidos. Desse total de obras, 1% tem motivação de cunho ambiental, onde se enquadra a questão das ocorrências arqueológicas, conforme gráfico a seguir.

Figura 1 – Motivos de paralisação das obras



Fonte: Tribunal de Contas da União – Acórdão 1.079/2019 – TCU – Plenário

Importante ressaltar que 1% dos recursos investidos e que se encontram impactados por paralisações, representa aproximadamente R\$ 1,5 bilhão.

Detalhando os contratos que envolvem obras paralisadas, Paula (2022, p.167-585) traz um resumo dos empreendimentos paralisados, cujo objeto é o gerenciamento de atividades relacionadas com arqueologia, conforme quadro a seguir.

Quadro 2 – Contratos paralisados envolvendo estudos ambientais e arqueológicos

OBJETO	UF	VALOR INVESTIMENTO R\$	% REALIZADO	DATA DE INÍCIO	DATA DE TÉRMINO DA OBRA	SITUAÇÃO	JUSTIFICATIVA
ELABORAÇÃO DE PROJETO DE ESTUDO DE IMPACTO AMBIENTAL, PLANO BÁSICO AMBIENTAL, COMPONENTE INDÍGENA, SERVIÇO DE ARQUEOLOGIA E ESTUDO PARA OBTENÇÃO DE AUTORIZAÇÃO DE SUPRESSÃO DE VEGETAÇÃO	MG	R\$ 2.875.168,65	64%	02/10/09	02/02/15	PARALISADO	INTERESSE ADMINISTRATIVO
ELABORAÇÃO DE ESTUDO AMBIENTAL - EA, PLANO BÁSICO AMBIENTAL - PBA, SERVIÇOS DE ARQUEOLOGIA E ESTUDOS PARA OBTENÇÃO DA AUTORIZAÇÃO DE SUPRESSÃO DE VEGETAÇÃO - ASV, VISANDO O LICENCIAMENTO AMBIENTAL NA BR-158/MT, TERRA INDÍGENA MARAIWAT	MT	R\$ 2.209.591,54	45%	04/02/11	31/01/15	PARALISADO	INTERESSE ADMINISTRATIVO
PRESTAÇÃO DE SERVIÇOS TÉCNICOS ESPECIALIZADOS PARA A EXECUÇÃO DE PROSPECÇÃO ARQUEOLÓGICA, DO MONITORAMENTO ARQUEOLÓGICO E A ELABORAÇÃO DO PROGRAMA DE EDUCAÇÃO PATRIMONIAL PARA OBTENÇÃO DA ANUÊNCIA DO IPHAN, NA RODOVIA BR-493/RJ	RJ	R\$ 367.975,58	0%	20/05/11		PARALISADO	INTERESSE ADMINISTRATIVO

OBJETO	UF	VALOR INVESTIMENTO R\$	% REALIZADO	DATA DE INÍCIO	DATA DE TÉRMINO DA OBRA	SITUAÇÃO	JUSTIFICATIVA
ELABORAÇÃO DE ESTUDOS DE IMPACTO AMBIENTAL (EIA), RELATÓRIO DE IMPACTO AMBIENTAL (RIMA), PLANO BÁSICO AMBIENTAL, SERVIÇOS ARQUEOLOGIA E ESTUDOS PARA OBTENÇÃO DA AUTORIZAÇÃO DE SUPRESSÃO DE VEGETAÇÃO PARA LICENCIAMENTO AMBIENTAL DAS OBRAS DE DUPLICAÇÃO BR262	ES	R\$ 1.390.001,70	50%	30/08/13	12/09/15	PARALISADO	NECESSIDADE DE ELABORAR UM NOVO CRONOGRAMA
ELABORAÇÃO DE ESTUDO DE IMPACTO AMBIENTAL, RELATÓRIO DE IMPACTO AMBIENTAL, PLANO BÁSICO AMBIENTAL, COMPONENTE INDÍGENA E QUILOMBOLA, SERVIÇOS DE ARQUEOLOGIA E ESTUDOS PARA OBTENÇÃO DA AUTORIZAÇÃO DE SUPRESSÃO DE VEGETAÇÃO PARA LICENCIAMENTO DA BR 101/BA.	BA	R\$ 2.476.823,55	48%	16/09/13	04/03/16	PARALISADO	INTERESSE ADMINISTRATIVO
ELABORAÇÃO DE EIA, RIMA, PBA, SERVIÇOS DE ARQUEOLOGIA E ESTUDOS PARA OBTENÇÃO DA AUTORIZAÇÃO DE SUPRESSÃO DE VEGETAÇÃO PARA O LICENCIAMENTO AMBIENTAL DAS OBRAS DE IMPLANTAÇÃO E PAVIMENTAÇÃO E OBRAS DE ARTES ESPECIAIS DA RODOVIA BR-393/ES.	ES	R\$ 1.336.266,29	37%	16/01/14	07/09/15	PARALISADO	PARALISAÇÃO DO PROJETO DE ENGENHARIA PARA IMPLANTAÇÃO DA RODOVIA BR-393/ES,
PREPARAÇÃO DE SÍTIOS ARQUEOLÓGICOS E GEOLÓGICOS, NO MUNICÍPIO DE JACOBINA	BA	R\$ 500.000,00	42%	01/06/15		PARALISADO	
EXECUÇÃO DOS SERVIÇOS DE RESGATE ARQUEOLÓGICO, MONITORAMENTO ARQUEOLÓGICO, EDUCAÇÃO PATRIMONIAL E PROTEÇÃO AO PATRIMÔNIO HISTÓRICO, CULTURAL E ARQUEOLÓGICO JUNTO ÀS OBRAS DE IMPLANTAÇÃO E PAVIMENTAÇÃO NA RODOVIA BR-080/GO.	GO	R\$ 3.544.810,12	4%	16/11/15	04/12/17	PARALISADO	INDEFINIÇÃO DO PROJETO EXECUTIVO DA BR-080/GO, IMPOSSIBILITANDO CONHECER O TRAÇADO DEFINITIVO E SUA ÁREA DE INFLUÊNCIA DIRETA E INDIRETA, ONDE SERÃO CONFIRMADOS OU NÃO O RESGATE DOS SÍTIOS ARQUEOLÓGICOS DEFINIDOS.

Fonte: adaptado de Paula (2022)

Os contratos acima são denominados de contratos acessórios, tendo em vista que sua existência decorre de um contrato dito principal, cujo objeto é a obra de infraestrutura, a qual a paralisação é decorrente das causas elencadas na Figura 1, acima. Uma vez paralisado o contrato principal, a não paralisação do contrato acessório impõe justificativas sólidas por parte do agente público, sob pena de sofrer sanções dos órgãos de controle por gestão temerária, o que o faz optar por, também, paralisar o contrato acessório. Ocorre que essa paralisação do contrato acessório impede a continuação das sondagens de ocorrências de artefatos de interesse arqueológico, atividade que deve anteceder as ações do contrato principal, ou seja, a própria execução da obra de infraestrutura, o que trará impacto mesmo quando da retomada do contrato principal. É um ciclo vicioso.

É no desempenho do objeto do contrato acessório onde são realizadas as predições de sítios arqueológicos, por meio de modelos matemáticos elaborados na forma descrita neste artigo, possibilitando, então, uma melhor eficiência nas sondagens propriamente ditas, as quais, uma vez confirmando a ocorrência arqueológica, possibilitará a tempestiva tomada de decisão quanto à sua preservação ou mesmo resgate, na forma da legislação vigente.

2.2 Das abordagens teóricas de construção de modelos preditivos

Algumas premissas são adotadas na formulação de modelos matemáticos voltados à predição de sítios arqueológicos. Segundo Tsuchiya (2002, p. 17), essas premissas podem ser elencadas da seguinte forma:

- a) Primeira: a suposição de que a escolha do local dos acampamentos dos povos pré-históricos foi influenciada por elementos ambientais, naturais e físicos;
- b) Segunda: essas variáveis ambientais sobreviveram e podem ser representadas por dados atuais. Estes dados podem estar em mapas, monografias ou podem ser coletados no campo;
- c) Terceira: considera-se que as correlações entre locais arqueológicos e as características do ambiente físico/natural, observado por arqueólogos, são fatos que indicam as decisões dos povoamentos pretéritos.

Essas três condições formam a base das formulações voltadas às predições de locais com interesse arqueológico, mediante a elaboração de um modelo que congregue as variáveis ambientais, naturais e físicas que determinaram a escolha do local e a permanência, ainda que residual, dessas variáveis no tempo presente, validando as correlações consideradas pelos especialistas.

Merencio (2020, p. 30) comenta que a modelagem preditiva arqueológica (MPA) é anterior à popularização do SIG, sendo associada aos estudos de padrão e

sistemas de assentamentos de 1950. Aduz que “modelos são hipóteses formuladas que buscam explicar um universo empírico observado, ao mesmo tempo que possui caráter preditivo”, opinião compartilhada por Kipnis (1997, p. 4), quando afirma que modelos são hipóteses, ou um conjunto de hipóteses que simplifica observações complexas ao mesmo tempo em que oferece um quadro preditivo exato que estrutura estas observações. Este autor ressalta a dupla vantagem na utilização da MPA: na fase de planejamento, fornece a *localização* com maiores probabilidades de ocorrências arqueológicas, o que proporciona levantar os *custos* das diversas alternativas de locação do empreendimento.

Para a construção de modelos preditivos, há duas abordagens clássicas: a indutiva e a dedutiva, cujas características principais estão descritas no quadro a seguir.

Quadro 3 – Abordagens usuais em MPA

INDUTIVA	DEDUTIVA
Orientada apenas por dados empíricos tratados de forma estatística para obtenção de padrões	A modelagem é orientada pelas experiências prévias do arqueólogo com áreas preferenciais de localização de sítios arqueológicos.
Parte de dados empíricos para verificação das correlações entre sítios arqueológicos e variáveis ambientais e culturais, através da aplicação de testes estatísticos.	Formulação de hipóteses sobre a localização dos sítios, como por exemplo, proximidade de água ou elevação.
Uso de técnicas inferenciais, como a regressão logística	Para cada variável considerada relevante, atribui-se diferentes pesos através de uma equação matemática.
Possibilidade de extrapolar os resultados para uma área maior	Não há a criação de novos conhecimentos e sim, a utilização de um conhecimento prévio.

Fonte: adaptado de Merencio (2020) e Deblasis e Kozlowski (2019)

Fazendo-se uma analogia entre essas abordagens e o tratamento de dados conforme descrito nas Normas ABNT da série 14.653², referente a avaliação de bens, a abordagem indutiva seria o equivalente ao “tratamento científico”, ao passo que a abordagem dedutiva estaria representada pelo “tratamento por fatores”.

No tocante às variáveis comumente utilizadas, a literatura elenca as seguintes:

² Item 8.2.1.4 da Norma ABNT NBR 14.653-2:2011

Quadro 4 – Variáveis utilizáveis em MPA

VARIÁVEIS	RELEVÂNCIA
Distância da água	Representa a maior ou menor distância à fonte de água e sua influência na escolha de locais de habitação
Declividade	Representa a energia dispendida até às fontes (água, alimentos)
Elevação	representa a altimetria do terreno em relação ao nível do mar
Vegetação	Representa o tipo de vegetação predominante, pretérita e atual
Clima	informa quanto aos períodos de seca e médias térmicas
Radiação solar	representa o nível de exposição à claudicância
Geologia	estudo da crosta terrestre, da matéria que a compõe, o seu mecanismo de formação, as alterações que ocorrem desde a sua origem e a estrutura que a sua superfície possui atualmente
Geomorfologia	estuda a origem e a estrutura das formas de relevo. A formação de elementos da superfície terrestre é identificada pela natureza das rochas, pelo clima e por fatores endógenos e exógenos.
Pedologia	Refere-se ao estudo dos solos no seu ambiente natural.
Litologia	estudo especializado em rochas e suas camadas
Hidrogeologia	identifica a presença de aquíferos
Rugosidade do terreno	Identifica solos planos ou irregulares

Fonte: adaptado de Merencio (2020), Peres, Afonso e Mota (2018) e Deblasis e Kozlowski (2019)

As variáveis independentes são utilizadas nos modelos preditivos, cuja variável dependente é uma resposta entre 0 e 1, indicativa da probabilidade de se encontrar sítios arqueológicos no local em estudo. Esse modelo pode ser obtido por meio da abordagem dedutiva, que é aquela onde são empregados “pesos” às variáveis, ou pela abordagem indutiva, que utiliza técnicas estatísticas, como a regressão logística.

Para Gonzales (2018, p. 15), A regressão logística é uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, frequentemente binária, em função de uma ou mais variáveis independentes contínuas e/ou binárias. Então, a partir desse modelo gerado é possível calcular ou prever a probabilidade de um evento ocorrer, dado uma observação aleatória. Para cada dado da amostra, a variável resposta é binária: sim (1) ou não (0).

Autores como Deblasis e Kozlowski, consideram que a variável resposta pode advir de uma incerteza:

O conjunto de dados geográficos necessários para a elaboração dos modelos é composto por dois elementos: os sítios conhecidos e um mesmo número de locais aleatórios onde não há presença de sítios arqueológicos, pois a regressão logística opera com valores binários de certeza ou incerteza da presença de sítios.

[DEBLASIS e KOZLOWSKI, 2019, p. 81]

Mas para Tsuchiya (2002, p. 60), a presença ou ausência de ocorrências arqueológicas não é uma incerteza, mas sim uma condição necessária para que as variáveis independentes possam explicar a ausência ou a presença dessas ocorrências: “na regressão logística utiliza-se variáveis independentes, que explicam a variável dependente, a presença/ausência de ocorrências arqueológicas.”

Alex de la Cruz Huayanay Também comunga desse entendimento, de que a variável resposta só comporta valores certos, quer seja de “sucesso” ou de “fracasso”:

Os modelos de regressão binária são aqueles em que a variável de interesse, habitualmente denominada variável resposta, pode assumir somente dois valores denotados geralmente por 1 para a ocorrência do evento de interesse (“sucesso”) e 0 para a ocorrência do evento complementar (“fracasso”).

[HUAYANAY, 2019, p. 33]

Em abordagem puramente teórica quanto à técnica estatística em comento — a regressão logística —, Gujarati (2019, p.195) apresenta uma alternativa ao uso dessa técnica em situações em que a variável resposta é dicotômica, ou seja, assume apenas dois resultados possíveis: “sim” ou “não”, como é o caso do presente estudo. A estratégia consiste em agrupar os dados conforme a frequência das variáveis utilizadas. Dessa forma, tem-se que a variável resposta tem por fim estimar, conforme a combinação dessas variáveis, o total de ocorrências arqueológicas. Sabendo-se, a priori, o total de ocorrências, a frequência relativa de cada grupo responderia pela probabilidade de ocorrências para aquela determinada combinação de variáveis. Estabelecida assim essa nova amostra, pode-se usar então os Mínimos Quadrados Ordinários – MQO, para obtenção do modelo de regressão.

3. METODOLOGIA

Das duas abordagens analisadas, a indutiva e a dedutiva, é incontestável que a segunda traz maior subjetividade na aplicação, decorrente da seleção tanto das variáveis que comporão o modelo, quanto dos “pesos” que essas terão, ao passo que, na perspectiva indutiva, tanto as variáveis são validadas por meio de testes estatísticos, quanto os seus coeficientes são também obtidos sem a interferência do pesquisador.

Ocorre que, para a abordagem indutiva, e por extensão, o uso da regressão logística, há que se ter uma amostra representativa composta por respostas “sim” e “não”, o que significa, na prática, que não basta apenas a caracterização do local em que se encontra o sítio arqueológico, mas igualmente se faz necessário que tenha havido a prospecção com respostas “não”, ou seja, sondagens fracassadas de pesquisa por artefatos para, daí então, caracterizar o ambiente com as variáveis apropriadas. Critica-se a estratégia de selecionar locais aleatórios e considerá-los como de respostas “não”, pelo fato de não se ter a certeza de que, de fato, não há registros arqueológicos ali. A regressão logística trabalha com certezas, tanto para respostas “sim”, quanto para respostas “não”.

Tal é a dificuldade de se utilizar o que seria a técnica de excelência para o fim de detectar sítios arqueológicos, que é a regressão logística.

No outro pórtico, tem-se, como já citado, a alternativa de utilizar a abordagem dedutiva, com a limitação imposta pela subjetividade do arqueólogo, por mais experimentado que seja.

Assim, entre esses extremos, optou-se por adotar uma estratégia que se configura como intermediária entre as duas abordagens clássicas, que é a trazida por Gujarati (2019, p. 195), ao aproveitar o manancial de informações contidas na amostra composta por respostas “sim”, ou seja, o próprio universo das ocorrências de sítios arqueológicos da região de interesse, tirando da amostra as probabilidades de ocorrências arqueológicas, aqui denominado de risco arqueológico (R_a), permitindo estabelecer uma métrica para o potencial arqueológico, do tipo “Alto”, “Médio” ou “Baixo”, conforme o risco arqueológico calculado para o local de interesse.

3.1 Região de interesse

Foram estudadas as ocorrências de sítios arqueológicos no Estado do Rio Grande do Norte, conforme mapeamento constante na figura a seguir.

Figura 2 – Mapeamento de sítios arqueológicos do RN



Fonte: IPHAN, 2023.

Foram selecionados 85 sítios (Δ) que continham a localização geográfica, sendo que um deles, o de nº 36, era uma ocorrência marítima, tendo sido descartado. A amostra final é composta, portanto, por 84 dados.

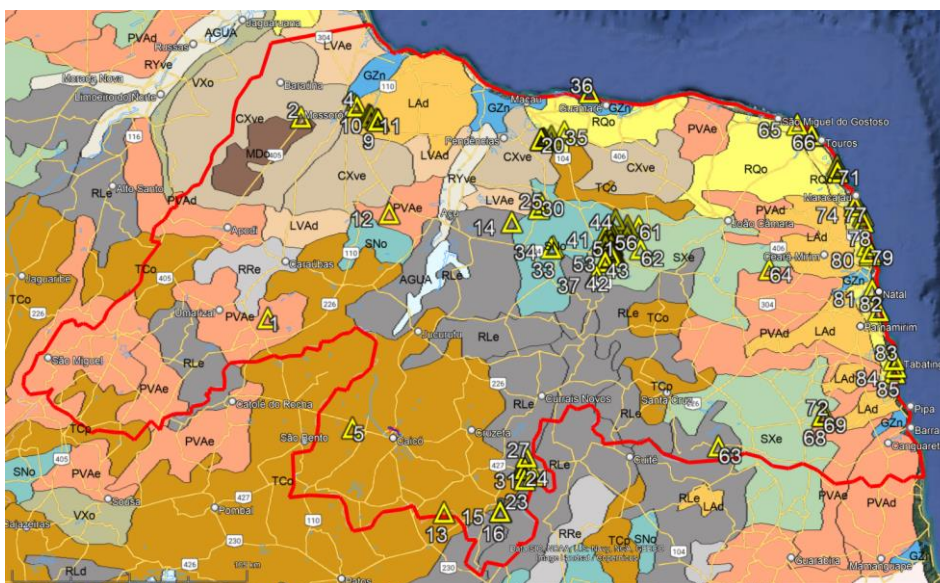
3.2 Variáveis utilizadas

Foram provisoriamente selecionadas as seguintes variáveis independentes:

- **Distância da água:** distância euclidiana entre o sítio arqueológico e a formação aquática mais próxima;
- **Classes de solo:**
 - Chernossolo
 - Neossolos
 - Planossolos
 - Latossolos
 - Cambissolos
 - Argissolos
 - Luvissolos
- **Declividade:** inclinação do terreno no sentido sítio/fonte aquática.

Os tipos de solos encontrados no RN estão mapeados conforme figura a seguir:

Figura 3 – Mapeamento dos tipos de solos do RN



Fonte: SANTOS (2011)

Na classificação brasileira de solos, os tipos de solos encontrados no RN pertencem às seguintes classes:

Quadro 5 – Tipos e classes de solo do RN

Símbolo	Tipo de solo	Classe
MDo	Chernossolos Rendzicos Órticos	Chernossolo
CXve	Cambissolos Háplicos Ta eutróficos	Cambissolos
LAd	Latossolos amarelos distróficos	Latossolos
PVAe	Argissolos vermelho-amarelos eutróficos	Argissolos
RLe	Neossolos litólicos eutróficos	Neossolos
RQo	Neossolos quartzarênicos órticos	Neossolos
SNo	Planossolos nátricos órticos	Planossolos
SXe	Planossolos háplicos eutróficos	Planossolos
TCo	Luvisolos crômicos órticos	Luvisolos

Fonte: SANTOS (2011)

A amostra está representada na tabela a seguir.

Tabela 2 – Sítios arqueológicos do RN (amostra)

Sítio	LONG	LAT	Dist.Água (m)	Elev.Sítio (m)	Elev.Água (m)	Cota.Água (m)	Declividade	Solo
1	-37,605	-6,012	317	193	179	14	0,0442	PVAe
2	-37,513	-5,213	32	72	72	0	0,0000	MDo
3	-37,303	-5,152	430	4	5	-1	-0,0023	LAd
4	-37,292	-5,160	454	5	5	0	0,0000	LAd
5	-37,246	-6,428	430	169	159	10	0,0233	TCo
6	-37,244	-5,191	216	13	5	8	0,0370	LAd
7	-37,238	-5,195	267	4	7	-3	-0,0112	LAd
8	-37,233	-5,198	585	13	8	5	0,0085	LAd
9	-37,212	-5,207	2766	28	9	19	0,0069	LAd
10	-37,205	-5,209	3501	40	9	31	0,0089	LAd
11	-37,205	-5,212	3560	43	9	34	0,0096	LAd
12	-37,142	-5,575	157	91	87	4	0,0255	PVAe
13	-36,866	-6,742	42	256	255	1	0,0238	TCo
14	-36,654	-5,586	627	83	90	-7	-0,0112	RLe
15	-36,643	-6,726	2126	526	428	98	0,0461	RLe
16	-36,640	-6,726	1836	510	428	82	0,0447	RLe
17	-36,563	-6,552	178	359	323	36	0,2022	RLe
18	-36,562	-6,552	201	370	323	47	0,2338	RLe
19	-36,556	-6,553	337	404	334	70	0,2077	RLe
20	-36,556	-5,256	1175	20	13	7	0,0060	CXve
21	-36,555	-5,256	1068	20	13	7	0,0066	CXve
22	-36,554	-6,554	300	363	330	33	0,1100	RLe
23	-36,553	-6,596	1457	383	354	29	0,0199	RLe
24	-36,552	-6,553	427	359	330	29	0,0679	RLe
25	-36,550	-5,527	250	121	120	1	0,0040	SNo
26	-36,549	-5,255	475	16	14	2	0,0042	CXve
27	-36,549	-6,506	1107	491	465	26	0,0235	RLe
28	-36,546	-5,252	112	15	15	0	0,0000	CXve
29	-36,545	-5,253	10	14	14	0	0,0000	CXve
30	-36,544	-5,506	2074	89	58	31	0,0149	SNo
31	-36,517	-6,577	0	616	616	0	0,0000	RLe
32	-36,516	-5,241	3541	47	13	34	0,0096	CXve
33	-36,493	-5,680	200	134	130	4	0,0200	SNo
34	-36,486	-5,675	850	145	133	12	0,0141	SNo
35	-36,465	-5,220	1911	37	28	9	0,0047	RQo
37	-36,309	-5,781	618	228	219	9	0,0146	SXe
38	-36,280	-5,645	598	201	192	9	0,0151	SXe
39	-36,279	-5,660	610	211	192	19	0,0311	SXe
40	-36,279	-5,657	563	214	189	25	0,0444	SXe

Sítio	LONG	LAT	Dist.Água (m)	Elev.Sítio (m)	Elev.Água (m)	Cota.Água (m)	Declividade	Solo
41	-36,279	-5,616	1435	234	212	22	0,0153	SNo
42	-36,277	-5,722	0	232	232	0	0,0000	SXe
43	-36,276	-5,720	0	230	230	0	0,0000	SXe
44	-36,268	-5,594	0	207	207	0	0,0000	SNo
45	-36,266	-5,630	290	219	205	14	0,0483	SXe
46	-36,263	-5,633	990	217	212	5	0,0051	SXe
47	-36,259	-5,612	1757	237	219	18	0,0102	SXe
48	-36,258	-5,658	125	208	201	7	0,0560	SXe
49	-36,256	-5,626	162	220	219	1	0,0062	SXe
50	-36,254	-5,653	424	210	202	8	0,0189	SXe
51	-36,254	-5,698	209	208	204	4	0,0191	SXe
52	-36,253	-5,657	414	202	198	4	0,0097	SXe
53	-36,253	-5,693	500	211	204	7	0,0140	SXe
54	-36,251	-5,688	1196	214	197	17	0,0142	SXe
55	-36,248	-5,686	1214	208	195	13	0,0107	SXe
56	-36,248	-5,617	1052	239	228	11	0,0105	SXe
57	-36,244	-5,555	1282	203	201	2	0,0016	SXe
58	-36,244	-5,685	642	204	197	7	0,0109	SXe
59	-36,197	-5,578	730	234	207	27	0,0370	SXe
60	-36,195	-5,620	683	199	185	14	0,0205	SXe
61	-36,151	-5,578	845	203	196	7	0,0083	SXe
62	-36,140	-5,665	689	171	158	13	0,0189	SXe
63	-35,791	-6,425	563	325	251	74	0,1314	RLe
64	-35,626	-5,719	1950	115	96	19	0,0097	SXe
65	-35,544	-5,142	456	0	-6	6	0,0132	RQo
66	-35,485	-5,180	2598	15	13	2	0,0008	RQo
67	-35,482	-5,182	2865	15	13	2	0,0007	RQo
68	-35,383	-6,284	419	76	54	22	0,0525	PVAe
69	-35,381	-6,284	168	60	54	6	0,0357	PVAe
70	-35,377	-5,295	1414	5	11	-6	-0,0042	RQo
71	-35,375	-5,321	2878	14	4	10	0,0035	RQo
72	-35,374	-6,280	528	70	57	13	0,0246	PVAe
73	-35,373	-5,317	3357	18	4	14	0,0042	RQo
74	-35,299	-5,450	136	14	7	7	0,0515	RQo
75	-35,296	-5,437	295	6	0	6	0,0203	RQo
76	-35,264	-5,486	1542	27	2	25	0,0162	RQo
77	-35,262	-5,512	1022	20	-3	23	0,0225	RQo
78	-35,240	-5,605	366	30	12	18	0,0492	RQo
79	-35,229	-5,630	1658	30	2	28	0,0169	RQo
80	-35,228	-5,640	605	10	2	8	0,0132	RQo
81	-35,212	-5,777	107	0	0	0	0,0000	RQo
82	-35,182	-5,860	2271	34	29	5	0,0022	RQo
83	-35,110	-6,050	3181	21	3	18	0,0057	RQo

Sítio	LONG	LAT	Dist.Água (m)	Elev.Sítio (m)	Elev.Água (m)	Cota.Água (m)	Declividade	Solo
84	-35,104	-6,101	1789	10	4	6	0,0034	RQo
85	-35,101	-6,078	239	1	2	-1	-0,0042	RQo

Fonte: IPHAN (2023); Google Earth (2023) e adaptado de SANTOS (2011)

A identificação dos sítios arqueológicos e as coordenadas geográficas foram fornecidas pelo IPHAN (2023); a distância da água ao sítio (m), a elevação do sítio (m) e a elevação da massa aquática (m) mais próxima ao sítio foram obtidos diretamente no Google Earth. A cota da água é a diferença entre a elevação do sítio e a elevação da água. Dividindo-se a cota pela distância da água ao sítio, tem-se a declividade predominante. O tipo de solo vem de Santos (2011).

3.3 Agrupamento de variáveis

A variável “Distância da água” foi segregada em dois grupos: “próximo” e “distante”, sendo a divisão estabelecida por clusterização por k-means, a fim de evitar subjetividades, sendo o ponto de mudança igual a 1.500 m. Dessa forma, a variável terá tratamento dicotômico, sendo considerada “próximo”, caso o sítio esteja a menos de 1.500 m, ou “distante”, em caso contrário.

Da mesma forma, a variável “Declividade” foi segregada em dois grupos: “baixa” e “alta”, conforme a inclinação do terreno seja menor que 0,090 (“baixa”) ou igual ou superior (“alta”) a esse ponto de mudança, segundo a clusterização realizada.

Para a variável “Classe de solos”, observou-se inicialmente a frequência de cada classe, a partir da qual estabeleceu-se também um tratamento dicotômico caso o solo pertença à classe “Neossolos” ou “Planossolos”.

O quadro resumo das variáveis é o seguinte:

Quadro 6 – Segregação de variáveis

Dist.Água (A1)		Declividade (D1)		Classe de solo			
< 1.500 m		< 0,090		Neossolos (S1)		Planossolos (S2)	
Sim	Não	Sim	Não	Sim	Não	Sim	Não
1	0	1	0	1	0	1	0

Fonte: elaborado pelo autor

Combinando-se as possibilidades do Quadro 4 com a amostra constante na Tabela 2, obtém-se a seguinte amostra agrupada:

Tabela 3 – Amostra agrupada

#	A1	D1	S1	S2	Quant.	Frequência (Ra)
1	1	1	1	0	14	0,1667
2	1	1	0	1	28	0,3333
3	1	1	0	0	18	0,2143
4	1	0	1	0	5	0,0595
5	1	0	0	1	0	0,0000
6	1	0	0	0	0	0,0000
7	0	1	1	0	12	0,1429
8	0	1	0	1	3	0,0357
9	0	1	0	0	4	0,0476
10	0	0	1	0	0	0,0000
11	0	0	0	1	0	0,0000
12	0	0	0	0	0	0,0000
Total:					84	

Fonte: elaborado pelo autor

A frequência representa o risco arqueológico (Ra) de se encontrar um sítio, de acordo com a combinação de variáveis presentes no local; ou seja, é a variável resposta buscada.

Com essa configuração amostral, pode-se aplicar o MQO, para obtenção do valor de “Ra”, o que será visto mais adiante.

A correlação entre as variáveis independentes e a variável dependente “Ra” está consignada no quadro a seguir.

Quadro 7 – Matriz de correlações

	A1	D1	S1	S2	Ra
A1	1				
D1	0	1			
S1	0	0	1		
S2	0	0	-0,5	1	
Ra	0,4402	0,7081	0,0609	0,0609	1

Fonte: elaborado pelo autor

Vê-se que a variável independente mais fortemente correlacionada com a variável dependente “Ra” é a “Declividade”, seguida da “Distância da água”. As variáveis identificadoras do tipo de solo predominante possuem baixa correlação.

Todas as correlações são diretas, o que significa que, para a variável “Distância da água”, a correlação aumenta se a distância for inferior a 1.500 (assume o valor 1) m e diminui, caso a distância seja igual ou superior a 1.500 m. O mesmo ocorre com a variável “Declividade”: para uma inclinação do terreno inferior a 0,090, a correlação aumenta; e diminui, caso a declividade seja igual ou superior a 0,090.

3.4 Modelo de predição de sítios arqueológicos

Os dados componentes da amostra agrupada foram tratados por Estatística Inferencial, que é a parte da ciência estatística que permite extrair conclusões sobre uma população a partir dessa amostra.

Dos 12 dados agrupados, todos foram utilizados efetivamente na modelagem.

Foi utilizada a técnica de Regressão Linear, uma vez que se buscava estudar o comportamento de uma variável dependente — no caso, o risco de ocorrência arqueológica, Ra — em relação a outras variáveis, chamadas independentes, que são responsáveis pela variabilidade observada na probabilidade da ocorrência.

Para a elaboração dos cálculos, utilizou-se o software SisDea Home, v. 1.60.

O modelo de regressão de melhor ajuste, ou seja, o que melhor representa a probabilidade de ocorrência de sítio arqueológico, é dado pela equação:

$$Ra = -0,01871785714 + 0,05640952381 \times A1 + 0,1119761905 \times D1 + 0,0268 \times S1 - 0,02551071429 \times S2$$

(Eq. 1)

Onde:

Ra: Probabilidade de encontrar um sítio arqueológico no local de interesse;

A1: Distância da massa aquática ao sítio arqueológico, em metros, assumindo o valor “1”, caso a distância seja < 1.500 m, ou valor “0”, caso a distância seja ≥ 1.500 m;

D1: Declividade predominante no local de interesse, adimensional, assumindo o valor “1”, para uma declividade < 0,090, ou valor “0”, caso a declividade seja ≥ 0,090;

S1: Classe de solo. Assume o valor “1”, caso o solo seja da classe Neossolos, ou valor “0”, caso não seja;

S2: Classe de solo. Assume o valor “1”, caso o solo seja da classe Planossolos, ou valor “0”, caso não seja.

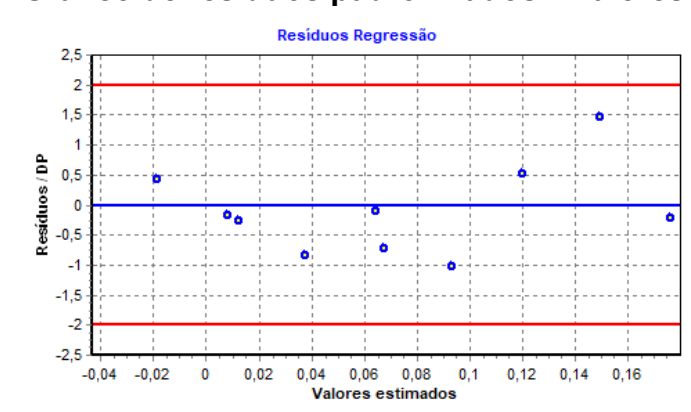
3.5 Análise dos pressupostos básicos do modelo de regressão linear

O modelo não possui “outliers” (pontos atípicos na amostra).

A correlação parcial máxima observada entre as variáveis independentes foi de 0,60, o que permite garantir a ausência de multicolinearidade (correlação parcial > 0,80), que é o fenômeno onde a influência de uma variável é camuflada devido a uma correlação entre essa variável e uma outra.

A análise gráfica dos resíduos mostra que não há heterocedasticidade, que é o fenômeno em que a variância dos erros não é constante:

Figura 4 – Gráfico de resíduos padronizados x valores estimados



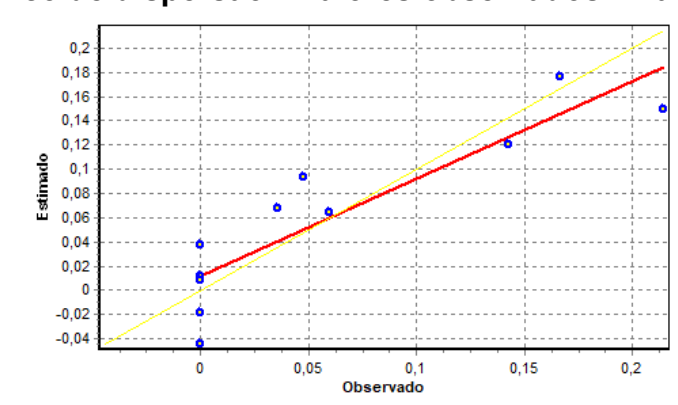
Fonte: modelagem via SisDea Home

3.6 Poder de explicação do modelo

O coeficiente de correlação “r” do modelo foi de 0,897, o que demonstra que as variáveis independentes (A1, D1, S1 e S2) estão fortemente correlacionadas com a variável dependente (Ra).

O coeficiente de determinação “r²” do modelo foi de 0,804, o que demonstra um forte poder de explicação da variação da variável dependente por parte das variáveis independentes, conforme mostra o gráfico a seguir.

Figura 5 – Gráfico de dispersão – valores observados x valores estimados



0,068; portanto, para um risco arqueológico (R_a) menor que 0,068 (6,8%), considera-se baixo o potencial de ocorrência. Para valores intermediários a esses pontos de mudança, considera-se que o risco é médio para a ocorrência de sítio arqueológico. Esses horizontes constam na tabela a seguir.

Tabela 4 – Potencial arqueológico

Potencial arqueológico	
Risco (R_a)	Potencial
$< 0,068$	Baixo
$0,068 \leq R_a \leq 0,124$	Médio
$> 0,124$	Alto

Fonte: elaborado pelo autor

Estabelecido esses horizontes, aplicou-se o modelo e esses critérios de potenciais arqueológicos à amostra total (Tabela 2), cujo resultado consta na tabela a seguir.

Tabela 5 – Teste do modelo preditivo

Sítio	A1	D1	S1	S2	QUANT	R_a calc	Potencial
1	1	1	0	0	1	0,1497	Alto
2	1	1	0	0	1	0,1497	Alto
3	1	1	0	0	1	0,1497	Alto
4	1	1	0	0	1	0,1497	Alto
5	1	1	0	0	1	0,1497	Alto
6	1	1	0	0	1	0,1497	Alto
7	1	1	0	0	1	0,1497	Alto
8	1	1	0	0	1	0,1497	Alto
9	0	1	0	0	1	0,0933	Médio
10	0	1	0	0	1	0,0933	Médio
11	0	1	0	0	1	0,0933	Médio
12	1	1	0	0	1	0,1497	Alto
13	1	1	0	0	1	0,1497	Alto
14	1	1	1	0	1	0,1765	Alto
15	0	1	1	0	1	0,1201	Médio
16	0	1	1	0	1	0,1201	Médio
17	1	0	1	0	1	0,0645	Baixo
18	1	0	1	0	1	0,0645	Baixo
19	1	0	1	0	1	0,0645	Baixo
20	1	1	0	0	1	0,1497	Alto
21	1	1	0	0	1	0,1497	Alto
22	1	0	1	0	1	0,0645	Baixo
23	1	1	1	0	1	0,1765	Alto

Sítio	A1	D1	S1	S2	QUANT	Ra_calc	Potencial
24	1	1	1	0	1	0,1765	Alto
25	1	1	0	1	1	0,1242	Alto
26	1	1	0	0	1	0,1497	Alto
27	1	1	1	0	1	0,1765	Alto
28	1	1	0	0	1	0,1497	Alto
29	1	1	0	0	1	0,1497	Alto
30	0	1	0	1	1	0,0677	Baixo
31	1	1	1	0	1	0,1765	Alto
32	0	1	0	0	1	0,0933	Médio
33	1	1	0	1	1	0,1242	Alto
34	1	1	0	1	1	0,1242	Alto
35	0	1	1	0	1	0,1201	Médio
37	1	1	0	1	1	0,1242	Alto
38	1	1	0	1	1	0,1242	Alto
39	1	1	0	1	1	0,1242	Alto
40	1	1	0	1	1	0,1242	Alto
41	1	1	0	1	1	0,1242	Alto
42	1	1	0	1	1	0,1242	Alto
43	1	1	0	1	1	0,1242	Alto
44	1	1	0	1	1	0,1242	Alto
45	1	1	0	1	1	0,1242	Alto
46	1	1	0	1	1	0,1242	Alto
47	0	1	0	1	1	0,0677	Baixo
48	1	1	0	1	1	0,1242	Alto
49	1	1	0	1	1	0,1242	Alto
50	1	1	0	1	1	0,1242	Alto
51	1	1	0	1	1	0,1242	Alto
52	1	1	0	1	1	0,1242	Alto
53	1	1	0	1	1	0,1242	Alto
54	1	1	0	1	1	0,1242	Alto
55	1	1	0	1	1	0,1242	Alto
56	1	1	0	1	1	0,1242	Alto
57	1	1	0	1	1	0,1242	Alto
58	1	1	0	1	1	0,1242	Alto
59	1	1	0	1	1	0,1242	Alto
60	1	1	0	1	1	0,1242	Alto
61	1	1	0	1	1	0,1242	Alto
62	1	1	0	1	1	0,1242	Alto
63	1	0	1	0	1	0,0645	Baixo
64	0	1	0	1	1	0,0677	Baixo
65	1	1	1	0	1	0,1765	Alto
66	0	1	1	0	1	0,1201	Médio
67	0	1	1	0	1	0,1201	Médio
68	1	1	0	0	1	0,1497	Alto
69	1	1	0	0	1	0,1497	Alto
70	1	1	1	0	1	0,1765	Alto
71	0	1	1	0	1	0,1201	Médio

Sítio	A1	D1	S1	S2	QUANT	Ra_calc	Potencial
72	1	1	0	0	1	0,1497	Alto
73	0	1	1	0	1	0,1201	Médio
74	1	1	1	0	1	0,1765	Alto
75	1	1	1	0	1	0,1765	Alto
76	0	1	1	0	1	0,1201	Médio
77	1	1	1	0	1	0,1765	Alto
78	1	1	1	0	1	0,1765	Alto
79	0	1	1	0	1	0,1201	Médio
80	1	1	1	0	1	0,1765	Alto
81	1	1	1	0	1	0,1765	Alto
82	0	1	1	0	1	0,1201	Médio
83	0	1	1	0	1	0,1201	Médio
84	0	1	1	0	1	0,1201	Médio
85	1	1	1	0	1	0,1765	Alto

Fonte: elaborado pelo autor

A acurácia do modelo é medida pela quantidade de acertos em classificar como “alto” o potencial de ocorrência de sítios arqueológicos, uma vez que todos os dados da amostra se reportam a ocorrências positivas. Assim, computando-se as quantidades “baixo”, “médio” e “alto”, tem-se o seguinte cenário:

Baixo	8
Médio	16
Alto	60
Total	84

A acurácia do modelo é igual à soma das quantidades “Alto”, mais 50% das quantidades “Médio”, uma vez que para esse potencial há igual probabilidade de ocorrência ou não de sítios arqueológicos; assim, a acurácia é calculada da seguinte forma: $(60 + 0,50 \times 16) / 84 = 0,81$, ou 81%, o que se pode considerar um forte poder preditivo, pois o modelo “acerta” 81% dos seus prognósticos.

4. ESTUDO DE CASO

Serão apresentados dois estudos de caso, o primeiro referente a uma obra de infraestrutura já executada e o segundo relativo a uma obra em execução; em ambas, a predição de ocorrências arqueológicas mostrou-se e mostra-se como uma ferramenta indispensável à preservação do patrimônio histórico, ao passo que otimiza as ações de prospecção dos potenciais sítios arqueológicos.

4.1 Primeiro estudo de caso: Construção de contorno rodoviário

Obra: Construção de contorno rodoviário, com 16 km.

O órgão ambiental estadual notificou o executante da obra, impedindo parcialmente a sua execução, devido ao projeto previsto para a construção do contorno rodoviário invadir uma área de preservação permanente (APP), numa extensão de 6 km, o que obrigou a alteração do traçado. Na figura a seguir, consta o detalhamento da ação.

Figura 6 – Contorno rodoviário (2012)



Fonte: imagem de satélite (Google Earth)

Na figura acima, o contorno rodoviário está representado pela linha amarela. A massa aquática, definidora da APP, está representada pela linha azul e o traçado original está representado pela linha vermelha, que teve de ser alterada para o traçado atual.

Efetivada a alteração do projeto para o atual traçado, o mesmo foi aprovado pelo órgão ambiental. Calculou-se, então, o risco arqueológico para os dois cenários, conforme figura a seguir.

Figura 7 – Alternativas de traçado do Contorno rodoviário (2012)



Fonte: imagem de satélite (Google Earth)

O que se observou foi que, tanto na situação anterior (traçado invadindo APP, ícones na cor amarelo), quanto na atual (ícones na cor vermelha), embora a questão ambiental tivesse sido contornada, o risco de ocorrências arqueológicas continuou inalterada, no sentido de apresentar, em ambos os cenários, um alto potencial de ocorrência de sítios arqueológicos, conforme o potencial detectado pelo risco arqueológico (R_a), calculado com o uso do modelo da Eq. 1.

Tabela 6 – Potencial arqueológico das alternativas de traçado do contorno rodoviário

TRAÇADO ORIGINAL						
Ponto	A1	D1	S1	S2	R_a	Potencial
1	1	1	0	0	0,1497	Alto
2	1	1	0	0	0,1497	Alto
3	1	1	0	0	0,1497	Alto
4	1	1	0	0	0,1497	Alto
5	1	1	0	0	0,1497	Alto
6	1	1	0	0	0,1497	Alto
7	1	1	0	0	0,1497	Alto
8	1	1	0	0	0,1497	Alto
9	1	1	0	0	0,1497	Alto
10	1	1	0	0	0,1497	Alto
11	1	1	0	0	0,1497	Alto
12	1	1	0	0	0,1497	Alto

TRAÇADO MODIFICADO						
Ponto	A1	D1	S1	S2	R_a	Potencial
1	1	1	0	0	0,1497	Alto
2	1	1	0	0	0,1497	Alto
3	1	1	0	0	0,1497	Alto
4	1	1	0	0	0,1497	Alto
5	1	1	0	0	0,1497	Alto
6	1	1	0	0	0,1497	Alto
7	1	1	0	0	0,1497	Alto
8	1	1	0	0	0,1497	Alto
9	1	1	0	0	0,1497	Alto
10	1	1	0	0	0,1497	Alto
11	1	1	0	0	0,1497	Alto
12	1	1	0	0	0,1497	Alto

TRAÇADO ORIGINAL						
Ponto	A1	D1	S1	S2	Ra	Potencial
13	1	1	0	0	0,1497	Alto
14	1	1	0	0	0,1497	Alto
15	1	1	0	0	0,1497	Alto
16	1	1	0	0	0,1497	Alto
17	1	1	0	0	0,1497	Alto
18	1	1	0	0	0,1497	Alto
19	1	1	0	0	0,1497	Alto
20	1	1	0	0	0,1497	Alto
21	1	1	0	0	0,1497	Alto
22	1	1	0	0	0,1497	Alto
23	1	1	0	0	0,1497	Alto
24	1	1	0	0	0,1497	Alto
25	1	1	0	0	0,1497	Alto
26	1	1	0	0	0,1497	Alto
27	1	1	0	0	0,1497	Alto
28	1	1	0	0	0,1497	Alto
29	1	1	0	0	0,1497	Alto
30	1	1	0	0	0,1497	Alto
31	1	1	0	0	0,1497	Alto
32	1	1	0	0	0,1497	Alto
33	1	1	0	0	0,1497	Alto
34	1	1	0	0	0,1497	Alto
35	1	1	0	0	0,1497	Alto
36	1	1	0	0	0,1497	Alto
37	1	1	0	0	0,1497	Alto
38	1	1	0	0	0,1497	Alto
39	1	1	0	0	0,1497	Alto

TRAÇADO MODIFICADO						
Ponto	A1	D1	S1	S2	Ra	Potencial
13	1	1	0	0	0,1497	Alto
14	1	1	0	0	0,1497	Alto
15	1	1	0	0	0,1497	Alto
16	1	1	0	0	0,1497	Alto
17	1	1	0	0	0,1497	Alto
18	1	1	0	0	0,1497	Alto
19	1	1	0	0	0,1497	Alto
20	1	1	0	0	0,1497	Alto
21	1	1	0	0	0,1497	Alto
22	1	1	0	0	0,1497	Alto
23	1	1	0	0	0,1497	Alto
24	1	1	0	0	0,1497	Alto
25	1	1	0	0	0,1497	Alto
26	1	1	0	0	0,1497	Alto
27	1	1	0	0	0,1497	Alto
28	1	1	0	0	0,1497	Alto
29	1	1	0	0	0,1497	Alto
30	1	1	0	0	0,1497	Alto
31	1	1	0	0	0,1497	Alto
32	1	1	0	0	0,1497	Alto
33	1	1	0	0	0,1497	Alto
34	1	1	0	0	0,1497	Alto
35	1	1	0	0	0,1497	Alto
36	1	1	0	0	0,1497	Alto
37	1	1	0	0	0,1497	Alto

Fonte: elaborado pelo autor

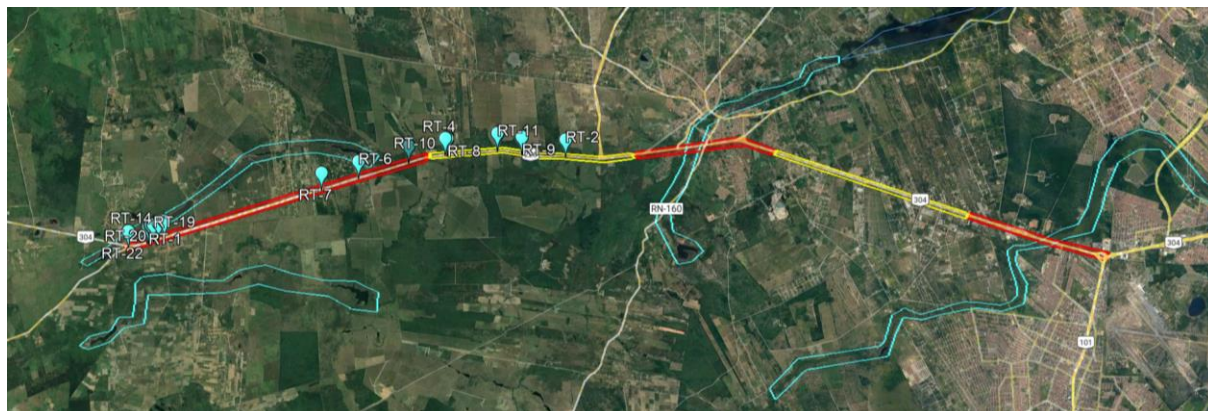
Com essa constatação, observa-se que, mesmo estando uma obra de infraestrutura obedecendo a legislação ambiental, o potencial de surgimento de sítios arqueológicos deve ser verificado e, para tanto, o uso de um modelo preditivo mostra-se de fundamental importância.

4.2 Segundo estudo de caso: Duplicação de rodovia

Obra: Duplicação de uma rodovia, com 27 km.

Foi aplicado o modelo (Eq. 1) em toda a extensão da obra rodoviária de duplicação, tendo sido obtida a seguinte configuração de potencial arqueológico.

Figura 8 – Potencial arqueológico de uma duplicação de rodovia (2022)



Fonte: elaborado pelo autor, com imagens de satélite do Google Earth

Os segmentos considerados como de “Alto” potencial arqueológico estão destacados em vermelho, ao passo que os segmentos de “Médio” potencial estão indicados em amarelo. Não foi detectado segmento de “Baixo” potencial arqueológico.

A primeira constatação que se faz é de que as áreas de “Alto” potencial arqueológico são aquelas próximas às massas aquáticas (polígonos na cor azul), o que ressalta a importância da variável “Distância da água”, presente no modelo.

Atualmente, encontra-se em execução da duplicação rodoviária, os primeiros segmentos de “Alto” e “Médio” potencial, a partir da esquerda da figura. Nesses dois segmentos, o monitoramento arqueológico em curso já detectou 22 ocorrências, sendo 15 localizadas nos segmento de “Alto” potencial e 7 ocorrências detectadas no segmento de “Médio” potencial. Os dados amostrais constam na tabela a seguir.

Tabela 7 – Potencial arqueológico da duplicação de rodovia

#	Em execução?	A1	D1	S1	S2	Ra	Potencial	Ocorrência
1	Sim	1	1	0	0	0,1497	Alto	RT-22
2	Sim	1	1	0	0	0,1497	Alto	
3	Sim	1	1	0	0	0,1497	Alto	RT-1; RT-12; RT-13; RT-15; RT-16; RT-17; RT-18; RT-19
4	Sim	1	1	0	0	0,1497	Alto	RT-14; RT-20; RT-21
5	Sim	1	1	0	0	0,1497	Alto	
6	Sim	1	1	0	0	0,1497	Alto	
7	Sim	1	1	0	0	0,1497	Alto	

#	Em execução?	A1	D1	S1	S2	Ra	Potencial	Ocorrência
8	Sim	1	1	0	0	0,1497	Alto	
9	Sim	1	1	0	0	0,1497	Alto	
10	Sim	1	1	0	0	0,1497	Alto	
11	Sim	1	1	0	0	0,1497	Alto	
12	Sim	1	1	0	0	0,1497	Alto	
13	Sim	1	1	0	0	0,1497	Alto	
14	Sim	1	1	0	0	0,1497	Alto	
15	Sim	1	1	0	0	0,1497	Alto	
16	Sim	1	1	0	0	0,1497	Alto	
17	Sim	1	1	0	0	0,1497	Alto	
18	Sim	1	1	0	0	0,1497	Alto	RT-7
19	Sim	1	1	0	0	0,1497	Alto	
20	Sim	1	1	0	0	0,1497	Alto	
21	Sim	1	1	0	0	0,1497	Alto	RT-6
22	Sim	1	1	0	0	0,1497	Alto	
23	Sim	1	1	0	0	0,1497	Alto	
24	Sim	1	1	0	0	0,1497	Alto	
25	Sim	1	1	0	0	0,1497	Alto	
26	Sim	1	1	0	0	0,1497	Alto	RT-5
27	Sim	1	1	0	0	0,1497	Alto	
28	Sim	0	1	0	0	0,0933	Médio	
29	Sim	0	1	0	0	0,0933	Médio	RT-4; RT-8; RT-10
30	Sim	0	1	0	0	0,0933	Médio	
31	Sim	0	1	0	0	0,0933	Médio	
32	Sim	0	1	0	0	0,0933	Médio	
33	Sim	0	1	0	0	0,0933	Médio	
34	Sim	0	1	0	0	0,0933	Médio	RT-3; RT-11
35	Sim	0	1	0	0	0,0933	Médio	
36	Sim	0	1	0	0	0,0933	Médio	RT-9
37	Sim	0	1	0	0	0,0933	Médio	
38	Sim	0	1	0	0	0,0933	Médio	
39	Sim	0	1	0	0	0,0933	Médio	
40	Sim	0	1	0	0	0,0933	Médio	RT-2
41	Sim	0	1	0	0	0,0933	Médio	
42	Sim	0	1	0	0	0,0933	Médio	
43	Sim	0	1	0	0	0,0933	Médio	
44	Sim	0	1	0	0	0,0933	Médio	
45	Sim	0	1	0	0	0,0933	Médio	
46	Sim	1	1	0	0	0,1497	Alto	
47	Sim	1	1	0	0	0,1497	Alto	
48	Não	1	1	0	0	0,1497	Alto	
49	Não	1	1	0	0	0,1497	Alto	
50	Não	1	1	0	0	0,1497	Alto	
51	Não	1	1	0	0	0,1497	Alto	
52	Não	1	1	0	0	0,1497	Alto	
53	Não	1	1	0	0	0,1497	Alto	

#	Em execução?	A1	D1	S1	S2	Ra	Potencial	Ocorrência
54	Não	1	1	0	0	0,1497	Alto	
55	Não	1	1	0	0	0,1497	Alto	
56	Não	1	1	0	0	0,1497	Alto	
57	Não	1	1	0	0	0,1497	Alto	
58	Não	0	1	0	0	0,0933	Médio	
59	Não	0	1	0	0	0,0933	Médio	
60	Não	0	1	0	0	0,0933	Médio	
61	Não	0	1	0	0	0,0933	Médio	
62	Não	0	1	0	0	0,0933	Médio	
63	Não	0	1	0	0	0,0933	Médio	
64	Não	0	1	0	0	0,0933	Médio	
65	Não	0	1	0	0	0,0933	Médio	
66	Não	0	1	0	0	0,0933	Médio	
67	Não	0	1	0	0	0,0933	Médio	
68	Não	0	1	0	0	0,0933	Médio	
69	Não	0	1	0	0	0,0933	Médio	
70	Não	0	1	0	0	0,0933	Médio	
71	Não	0	1	0	0	0,0933	Médio	
72	Não	0	1	0	0	0,0933	Médio	
73	Não	0	1	0	0	0,0933	Médio	
74	Não	0	1	0	0	0,0933	Médio	
75	Não	0	1	0	0	0,0933	Médio	
76	Não	1	1	0	0	0,1497	Alto	
77	Não	1	1	0	0	0,1497	Alto	
78	Não	1	1	0	0	0,1497	Alto	
79	Não	1	1	0	0	0,1497	Alto	
80	Não	1	1	0	0	0,1497	Alto	
81	Não	1	1	0	0	0,1497	Alto	
82	Não	1	1	0	0	0,1497	Alto	
83	Não	1	1	0	0	0,1497	Alto	
84	Não	1	1	0	0	0,1497	Alto	
85	Não	1	1	0	0	0,1497	Alto	
86	Não	1	1	0	0	0,1497	Alto	
87	Não	1	1	0	0	0,1497	Alto	
88	Não	1	1	0	0	0,1497	Alto	

Fonte: elaborado pelo autor

Computando-se as quantidades “RT Alto” e “RT Médio”, tem-se o seguinte cenário:

RT "Alto"	15
RT "Médio"	7
RT total	22

A acurácia do modelo é igual à soma das quantidades “Alto”, mais 50% das quantidades “Médio”, uma vez que para esse potencial há igual probabilidade de ocorrência ou não de sítios arqueológicos; assim, a acurácia é calculada da seguinte forma: $(15 + 0,50 \times 7) / 22 = 0,84$, ou 84%, o que significa dizer que o modelo “acertou” 84% dos seus prognósticos. Esse percentual confirma a acurácia medida para o modelo (81%), quando aplicado na totalidade dos sítios arqueológicos do RN.

Uma vez que a obra de duplicação da rodovia encontra-se ainda em execução, especial atenção deve ser dada aos segmentos de “Alto” potencial arqueológico ainda não iniciados, conforme indicados na Tabela 7.

Essa antecipação da ação de monitoramento arqueológico só se faz possível com o amparo de um modelo preditivo como o ora utilizado.

5. CONCLUSÃO

Grande parte das obras paralisadas no país tem como causa as falhas no seu planejamento, com destaque à inexistência de prazos no cronograma físico para a regularização ambiental do empreendimento, categoria em que se enquadram as demandas do necessário monitoramento arqueológico. O impacto na produtividade pactuada, e na consequente extensão de prazo, é um prejuízo certo para as partes, mas em especial à sociedade, que se privará, por tempo incerto, da respectiva obra estruturante.

Como forma de minorar esses impactos, o estágio da arte da estatística inferencial permite elaborar modelos matemáticos capazes de prever a existência de sítios arqueológicos nos locais previstos para os empreendimentos de infraestrutura, informação essa que se reveste de fundamental importância desde a fase de concepção do projeto até à de execução da obra.

Das abordagens clássicas para elaboração de um modelo matemático com essa finalidade, tem-se a dedutiva, onde prevalece a experiência do arqueólogo em definir tanto as variáveis que serão utilizadas no modelo, quanto os “pesos” que cada variável deve ter, o que termina por agregar certo grau de subjetividade em sua formulação. No outro extremo, tem-se a abordagem indutiva, que utiliza a regressão logística, a partir de amostras com resultados “sim” (existência de sítios arqueológicos) e resultados “não” (inexistência de sítios arqueológicos), onde as variáveis são testadas e seus coeficientes advêm da própria técnica regressiva, o que afasta a subjetividade do operador em sua formulação. Como limitador da utilização dessa abordagem, tem-se o fato de que são escassos os registros de resultados “não”, ou seja, sondagens que redundaram em fracasso no descobrimento de sítios arqueológicos. Em que pese os resultados “não” serem inclusive em maior monta que os resultados “sim”, é a inexistência desses registros, com todas as informações necessárias, que dificulta o uso da regressão logística. No mesmo diapasão, rejeita-se, nesse artigo, a estratégia de se eleger locais aleatórios como se fossem de

resultados “não”, por distorcer a essência da técnica da regressão logística, que trabalha com certezas tanto para respostas “sim”, quanto para respostas “não”.

Como alternativa intermediária, adotou-se no presente artigo a estratégia sugerida por Gujarati (2019), de se utilizar os dados da amostra agrupados segundo a frequência das variáveis independentes, o que redundou em um modelo matemático capaz de prever as ocorrências arqueológicas com uma acurácia de 81%, o que se pode considerar um forte poder preditivo.

Aplicado o modelo em dois estudos de caso, no primeiro deles mostrou-se relevante em mostrar que o cumprimento da legislação ambiental, no tocante à preservação de áreas de proteção permanente (APPs), não desobriga o monitoramento arqueológico, pois se tratam de disciplinas que, embora complementares em um empreendimento de infraestrutura, são de naturezas distintas. No segundo caso, por se tratar de uma obra ainda em execução, foi possível mensurar em 84% a acurácia do modelo, no tocante aos achados arqueológicos já ocorridos no segmento em execução, ao passo que, para o segmento ainda a ser executado, delineou as áreas tidas como de “Alto” e “Médio” potencial arqueológico, orientando assim as ações tanto de monitoramento, quanto de movimento de terras próprios do empreendimento.

Para estudos futuros visando a otimização do modelo ora obtido, sugere-se a ampliação da amostra, com introdução de novas variáveis independentes, utilizando-se testes de especificação, como o teste RESET, de Ramsey, e do Multiplicador de Lagrange.

REFERÊNCIAS BIBLIOGRÁFICAS

BRASIL. TRIBUNAL DE CONTAS DA UNIÃO. Acórdão 1.079/2019 – TCU – Plenário. Brasília: TCU, 2019.

DEBLASIS, Paulo e KOZLOWSKI, Henrique de Sena. **Perspectivas de modelagem preditiva arqueológica: estudo de caso na Bacia do Rio Tubarão – Santa Catarina**. In: Especiaria - Caderno de Ciências Humanas. p. 72-94, mar. 2019.

GONZALEZ, Leandro de Azevedo. **Regressão Logística e suas Aplicações**. 2018. 106 f. Monografia (Graduação) - Curso de Ciência da Computação, Centro de Ciências Exatas e Tecnológicas, Universidade Federal do Maranhão, São Luís, 2018.

GUJARATI, Damodar N. **ECONOMETRIA: princípios, teoria e aplicações práticas**. 1 ed. São Paulo: Saraiva Educação, 2019. 522 p.

HUAYANAY, Alex de La Cruz. **Modelos De Regressão Para Resposta Binária Na Presença De Dados Desbalanceados**. 2019. 91 f. Dissertação (Mestrado) - Curso de Estatística, Centro de Ciências Exatas e Tecnologia, UFSCar-Usp, São Carlos, 2019.

IPHAN. Instituto do Patrimônio Histórico e Artístico Nacional. < <https://www.gov.br/iphan/pt-br/patrimonio-cultural/patrimonio-arqueologico/cadastro-de-sitios-arqueologicos> > (acesso em 12/04/2023).

KIPNIS, Renato. **O uso de modelos preditivos para diagnosticar recursos arqueológicos em áreas a serem afetadas por empreendimentos de impacto ambiental**. IN: CALDARELLI, Solange Bezerra. Atas do Simpósio sobre Política Nacional do Meio Ambiente e Patrimônio Cultural. Goiânia: Instituto Goiano de Pré-História e Antropologia, 1997, p. 34–51.

MERENCIO, Fabiana Terhaag. **A modelagem preditiva na Arqueologia: um olhar crítico a partir do estudo de caso no território Xetá**. Cadernos do Lepaarq, v. XVII, n.33, p. 25-45, Jan-Jun. 2020.

PAULA, Eduardo Henrique Andrade de. **INSTITUTIONAL VOIDS**: uma análise taxonômica e fenomenológica de sua influência na paralisação de obras públicas no Brasil. 2022. 727 f. Tese (Doutorado) - Curso de Administração, Ciências Empresariais, Universidade Fumec, Belo Horizonte, 2022.

PEREZ, Glauco Constantino; AFONSO, Marisa Coutinho e MOTA, Lúcio Tadeu. **Métodos de análise espacial para sítios arqueológicos**: um modelo preditivo para o estado de São Paulo. In: Cadernos do Lepaarq, v. XV, n.30., p. 98-120, Jul-Dez. 2018.

SANTOS, Humberto Gonçalves dos. **O novo mapa de solos do Brasil**: legenda atualizada. Rio de Janeiro: Embrapa Solos, 2011. 67 p. Modo de acesso: < <http://www.cnps.embrapa.br/solosbr/publicacao.html> >. (acesso em 12/04/2023).

TSUCHIYA, Ítalo. **Regressão Logística Aplicada Na Análise Espacial De Dados Arqueológicos**. 2002. 103 f. Dissertação (Mestrado) - Curso de Ciências Cartográficas, Cartografia, Universidade Estadual Paulista, Presidente Prudente, 2002.