

# Modelo de Suporte à Decisão para Obras Públicas com Base em Mineração de Dados

Thaylor Vieira

*tvm2@cin.ufpe.br*

*Centro de Informática*

*Universidade Federal de Pernambuco*

Recife, Brasil

Saulo Henrique

*shna@cin.ufpe.br*

*Centro de Informática*

*Universidade Federal de Pernambuco*

Recife, Brasil

Márcio Oliveira

*mob2@cin.ufpe.br*

*Centro de Informática*

*Universidade Federal de Pernambuco*

Recife, Brasil

Anderson Felinto

*afb4@cin.ufpe.br*

*Centro de Informática*

*Universidade Federal de Pernambuco*

Recife, Brasil

Bezaleel Lira

*bls2@cin.ufpe.br*

*Centro de Informática*

*Universidade Federal de Pernambuco*

Recife, Brasil

**Resumo**—A paralisação de obras públicas representa um grave problema socioeconômico, causando desperdício de recursos e impactos negativos à sociedade. Este trabalho apresenta um modelo de suporte à decisão baseado em mineração de dados para identificar obras públicas com alta probabilidade de paralisação antes do início da execução física. Utilizando dados do Tribunal de Contas da União (TCU), Portal da Transparência e IBGE, foram aplicados algoritmos de regressão logística, árvore de decisão e indução de regras (JRip). O modelo JRip obteve os melhores resultados, com precisão de 56,7% e F1-score de 46,5% para a classe minoritária. As variáveis duração da vigência e execução financeira foram identificadas como os principais fatores preditivos. Os resultados demonstram a viabilidade de um sistema preditivo para apoiar gestores públicos na tomada de decisões proativas, contribuindo para a otimização dos recursos públicos e redução de obras paralisadas.

**Index Terms**—mineração de dados, obras públicas, suporte à decisão, algoritmos de classificação, gestão pública

## I. INTRODUÇÃO

O presente artigo detalha o desenvolvimento e os resultados de um modelo de suporte à decisão focado na identificação de obras públicas passíveis de paralisação antes mesmo do início de sua execução física. A paralisação de obras públicas é um problema crônico e de grande impacto socioeconômico, gerando prejuízos financeiros substanciais, desperdício de recursos públicos e negativos impactos sociais, uma vez que as comunidades não se beneficiam da infraestrutura planejada. Muitas dessas paralisações ocorrem em fases iniciais, antes da obra sair do papel, o que frequentemente aponta para deficiências em gestão e planejamento.

A gestão de obras públicas no Brasil é um desafio complexo, permeado por fatores que vão desde a burocracia excessiva até questões de planejamento e execução. O fenômeno das obras paralisadas representa um entrave significativo ao desenvolvimento nacional e à efetividade dos investimentos públicos. A capacidade de prever a probabilidade de paralisação de uma obra, antes mesmo que ela tenha qualquer execução

física, permite que os órgãos de controle e gestão atuem proativamente, evitando ou mitigando os impactos negativos mencionados. Essa proatividade é a principal força motriz deste projeto, buscando transformar dados em conhecimento açãoável para otimizar o uso dos recursos públicos.

O objetivo central deste projeto é desenvolver um **modelo de suporte à decisão** para auxiliar na identificação de obras públicas com alta probabilidade de serem paralisadas, focando especificamente aquelas onde ainda não houve nenhuma execução física. Este modelo visa fornecer uma ferramenta preditiva que apoie a tomada de decisões estratégicas e operacionais, permitindo intervenções tempestivas e mais eficientes.

O problema de pesquisa abordado é como construir um modelo preditivo robusto, baseado em técnicas de mineração de dados, capaz de identificar padrões e características que aumentam a probabilidade de paralisação de uma obra pública em suas fases iniciais, ou seja, antes da execução física. Isso inclui a seleção de variáveis pertinentes e a aplicação de algoritmos que gerem regras claras e interpretáveis para o suporte à decisão.

A justificativa para este trabalho reside na urgência de se combater o desperdício de recursos públicos e os prejuízos sociais causados pelas obras paralisadas. Ao oferecer um mecanismo preditivo que alerta para riscos de paralisação em estágio precoce, o modelo proposto permite uma gestão mais eficiente, a realocação de recursos, a revisão de projetos e, em última instância, a conclusão de obras que beneficiarão diretamente a população. A identificação de problemas de gestão e planejamento antes da execução física da obra é crucial para aprimorar o processo de investimento público como um todo.

## II. METODOLOGIA E EXPERIMENTO

A construção de um modelo preditivo eficaz para a paralisação de obras públicas exige uma base de dados abrangente

e representativa. O processo de coleta e preparação dos dados foi uma etapa fundamental, envolvendo a integração de informações de diversas fontes oficiais.

#### A. Fontes de Dados

Para o desenvolvimento do modelo, foram utilizadas bases de dados de importantes órgãos e portais de informação do governo brasileiro:

- **Base de Dados do Tribunal de Contas da União (TCU):** Esta foi a fonte primária de dados, utilizada em quatro períodos distintos (agosto/22, abril/23, abril/24 e abril/25). O TCU é o principal órgão de controle externo do governo federal, e sua base de dados oferece informações cruciais sobre convênios e contratos de obras públicas.
- **Base de Dados do IBGE:** Utilizada para capturar informações sobre os biomas das cidades onde as obras estão localizadas. A característica do bioma pode influenciar as dificuldades logísticas, as condições climáticas e a infraestrutura local, fatores que impactam a execução de obras.
- **Portal da Transparência do Governo:** Complementou os dados do TCU, fornecendo informações sobre os convênios que a base do TCU não disponibiliza, como as datas de início e fim de vigência. Essa informação é vital para calcular a duração da vigência e outras variáveis temporais.
- **Outras Bases de Dados Pesquisadas:** Bases como o Índice de Desenvolvimento de Cidade Sustentável (IDCS) e o Índice de Gestão Municipal - Conselho Federal de Administração (IGM-CFA) foram pesquisadas para complementar o modelo. Contudo, os valores dos índices dessas bases não demonstraram relevância estatística significativa e foram descartados da análise final.

#### B. Dados Utilizados no Modelo e Justificativa

As variáveis selecionadas para compor o modelo foram escolhidas por sua relevância potencial em predizer a paralisação de obras, refletindo aspectos financeiros, temporais, geográficos e sociais:

- **duracao\_vigencia (em meses):** Representa o tempo de vigência do convênio ou contrato em meses. Períodos de vigência muito longos podem indicar falta de pressão para a conclusão e, potencialmente, problemas de planejamento ou execução que levam à paralisação. Questões políticas como trocas de mandatos dos prefeitos também podem influenciar muito na paralisação das obras, pois um novo prefeito pode não ter interesse em continuar projetos iniciados pelo antecessor, especialmente se forem de partidos rivais ou se o projeto não trouxer capital político direto.
- **vigencia\_inicio\_dezembro:** Uma variável binária que indica se a obra teve sua vigência iniciada em dezembro. Obras que iniciam em dezembro têm um período muito curto para o empenho, liquidação e pagamento de despesas antes do encerramento do ano fiscal (31 de dezembro).

Essa urgência pode levar a decisões apressadas ou à incapacidade de cumprir os prazos, tornando-as mais vulneráveis à paralisação.

- **tipo\_bioma:** Descreve o bioma da cidade onde a obra está sendo executada (semiárido, amazônico ou outros). Esta variável é crucial, pois reflete dificuldades logísticas próprias da região, a presença de clima extremo e a qualidade da infraestrutura local. Essas condições ambientais e de infraestrutura podem dificultar o avanço das obras e contribuir para sua paralisação.
- **saudade\_esporte\_cat\_descricao:** Indica se a obra pertence à área da saúde ou esporte. Obras nestas categorias tipicamente possuem maior visibilidade social e política, gerando pressão popular por sua entrega. Além disso, tendem a ter prioridade na alocação de recursos, o que, em tese, poderia reduzir o risco de paralisação.
- **proporcao\_contrapartida:** Ajuda a identificar se uma obra possui apoio local suficiente (via contrapartida) ou se depende quase exclusivamente de repasses federais. Uma alta dependência de repasses federais pode tornar a obra mais vulnerável à paralisação, caso haja atrasos ou cortes nesses repasses.
- **investimento\_mensal\_medio:** Representa o valor médio do investimento por mês. Um valor alto nesta variável pode indicar uma urgência na entrega da obra, enquanto um valor menor sugere menos pressão, justificando potenciais paralisações.
- **exec\_financeira:** Exibe o percentual que a obra já pagou, em uma escala de 0 a 1. Quanto maior este valor, menor a chance de a obra ser paralisada, pois indica um avanço financeiro significativo.

#### C. Preparação dos Dados

1) *Definição da Variável-Alvo (Target):* O target foi definido ao filtrar obras que, em um dos quatro períodos distintos da Base de Dados do TCU, estavam rotuladas como "em execução" e com 0% de execução física. Essa etapa é muito importante, pois não deve-se filtrar obras rotuladas como "paralisada", pois isso é caracterizado como informação a posteriori. O objetivo do modelo é dar suporte à decisão antes que a paralisação ocorra. Após esse primeiro filtro, verificou-se se essas mesmas obras estavam rotuladas como "paralisada" em um período de tempo posterior ao inicial. Em caso afirmativo, essas obras são rotuladas como `target_vai_paralisar = sim`. Após isso, foi feito um segundo filtro para definir o `target_vai_paralisar = nao`. Para definir essa parte do target, foi necessário verificar se as obras foram concluídas sem nenhuma paralisação.

2) *Limpeza e tratamento dos dados:* Foi realizada a limpeza dos *missing values* ao excluir exemplos da amostra que não tinham valores referentes ao `cod_instrumento`. Esta variável é o ID da obra e foi por meio dela que o nó *Reference Row Filter* no KNIME foi usado para complementar a nossa base de dados do TCU com as informações da Base de Dados do Portal da Transparência do Governo.

Após a definição das classes do target\_vai\_paralisar, foi realizada uma verificação de integridade para garantir que não houvesse obras (identificadas pelo cod\_instrumento) classificadas simultaneamente como "sim" e "não". Nenhuma inconsistência foi encontrada, assegurando a não contaminação do target.

Para realizar o tratamento de dados da Base do IBGE, foi aplicada uma padronização da variável municipio\_uf com o uso do nó *String Manipulation*, removendo acentos e convertendo os nomes para letras minúsculas. Essa padronização foi necessária para permitir a correta junção dos dados sobre biomas com os dados principais do projeto.

Durante a tentativa de integrar a base do IGM-CFA, identificou-se um problema decorrente da mudança de nomes de alguns municípios brasileiros ao longo do tempo. Isso gerou uma quantidade limitada de registros com *missing data* após a unificação. Como o número de casos era pequeno (cerca de 10 a 20 municípios), realizou-se uma correção manual com base em pesquisa sobre as alterações de nome, garantindo a integridade dos dados nesse ponto.

3) *Feature Engineering*: Nesta etapa, foram criadas variáveis derivadas com o objetivo de capturar informações mais relevantes para a previsão da paralisação de obras públicas. A transformação de variáveis permite extrair relações mais significativas dos dados e reduzir ruídos que poderiam prejudicar a performance dos modelos.

As principais variáveis construídas foram:

- **proporcao\_contrapartida** (**valor\_investimento/valor\_contrapartida**): definida como a razão entre o valor da contrapartida e o valor total do investimento da obra. Essa variável foi criada para unificar, em um único indicador, o poder de análise do valor\_contrapartida e do valor\_investimento, permitindo avaliar o grau de comprometimento financeiro do ente local em relação ao total investido na obra.
- **investimento\_mensal\_medio** (**valor\_investimento/duracao\_vigencia**): calculado como o valor total do investimento dividido pela duração de vigência da obra, em meses. Para lidar com a presença de outliers extremos, aplicou-se a winsorização nos percentis de 5% e 95%, limitando os valores fora desse intervalo aos próprios limites. Essa técnica foi adotada para suavizar distorções sem a necessidade de excluir exemplos da amostra, pois acima do percentil 95%, muitos valores estavam escalonando de forma exponencial, chegando em valores como R\$6.937.692,00/mês. Sendo assim, a técnica acaba preservando a integridade da base e graças a ela, a amostra tem valores entre R\$4.221,43/mês e R\$179.616,72/mês.
- **vigencia\_inicio\_dezembro**: variável binária criada a partir da data de início da vigência da obra, assumindo valor 1 quando o mês de início é dezembro, e 0 caso contrário.

#### D. Modelagem

Durante a etapa de modelagem, foram inicialmente utilizados os algoritmos de regressão logística, árvore de decisão, e JRip para indução de regras. Vale ressaltar que o conjunto de dados trabalhado é desbalanceado, com aproximadamente 75% das obras classificadas como "não" e apenas 25% como "sim" para o risco de paralisação.

A divisão da base de dados, composta por 4.675 exemplos no total, foi feita em 75% para treinamento (3.506 exemplos) e 25% para teste (1.169 exemplos), com estratificação pelo target em todos os algoritmos.

### III. RESULTADOS

#### A. Regressão Logística

A regressão logística foi aplicada por sua robustez na análise de variáveis contínuas e sua capacidade de fornecer probabilidades associadas à ocorrência da paralisação, além de permitir a interpretação dos coeficientes como medidas de impacto.

A regressão logística gerou coeficientes estatisticamente significativos para todas as variáveis incluídas no modelo, conforme evidenciado pelos valores de *p* inferiores a 0,05. As variáveis duracao\_vigencia (coef. 5.145) e exec\_financeira (coef. -2.771) se destacaram como os principais fatores associados à paralisação de obras. A primeira apresentou o maior coeficiente positivo, indicando que obras com vigência mais longa tendem a ter maior risco de paralisação, possivelmente por envolverem prazos dilatados que favorecem atrasos ou mudanças de gestão. Já a exec\_financeira apresentou o maior coeficiente negativo, evidenciando que quanto maior o percentual já executado financeiramente, menor a chance de paralisação. O modelo mostra-se consistente, com todos os *z-scores* elevados e coeficientes com interpretações coerentes com a realidade do problema. A Figura 1 ilustra os valores dos coeficientes e estatísticas das variáveis utilizadas:

Variable	Coeff. ↓	Std. Err.	z-score	P> z
String	Number (double)	Number (double)	Number (double)	Number (double)
duracao_vigencia	5.145	0.36	14.31	0
sim_vigencia_inicio_dezembro	0.425	0.09	4.714	0
Amazônico_tipo_bioma	0.371	0.129	2.869	0.004
Semiárido_tipo_bioma	0.258	0.103	2.498	0.013
Saude_Esporte_cat_descricao	-0.882	0.191	-4.627	0
proporcao_contrapartida	-1.226	0.335	-3.661	0
investimento_mensal_medio	-1.383	0.256	-5.413	0
Constant	-1.624	0.156	-10.414	0
exec_financeira	-2.771	0.191	-14.492	0

Figura 1. Coeficientes e estatísticas da Regressão Logística

A Curva ROC para a Regressão Logística apresentou um AUC (Area Under the Curve) de 0.773, indicando uma boa capacidade discriminatória do modelo. A Figura 2 ilustra a curva ROC:

Além da análise da AUC, foi realizado um ajuste no threshold de classificação da regressão logística, alterando o valor padrão de 0.5 para 0.4. Essa modificação teve como objetivo aumentar a sensibilidade do modelo à classe minoritária, ou seja, às obras com risco de paralisação. A decisão se

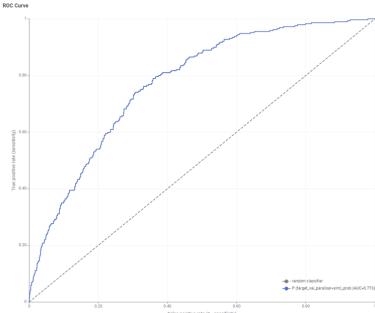


Figura 2. Curva ROC - Regressão Logística

baseou na lógica de que o prejuízo de não alertar sobre um risco real é maior do que o de emitir um falso positivo, isto é, prever que uma obra irá paralisar quando ela não paralisaria. Com o novo ponto de corte em 0,4, o modelo se torna mais conservador e proativo, favorecendo seu uso como ferramenta de alerta precoce para gestores públicos.

#### B. Árvore de Decisão

Outro algoritmo de aprendizado de máquina explorado, que cria um modelo preditivo baseado em uma série de decisões sobre as variáveis. Para evitar ramos muito profundos e regras excessivamente específicas, foi limitado o número mínimo de registros por nó ("Min number records per node") para 230.

Observou-se que as variáveis duracao\_vigencia e exec\_financeira foram as mais utilizadas na árvore (Raiz e nível 1 da árvore), reforçando as suas respectivas importâncias com base no que já foi constatado na regressão logística através de seus coeficientes. Alterar o número mínimo de 230 para um valor inferior gerava árvores com pelo menos 1 nível a mais de profundidade, resultando em regras mais específicas, difíceis de interpretar e de baixa cobertura. A Figura 3 ilustra a árvore de decisão gerada a partir do modelo:

Algumas variáveis constatadas com relevância estatística significativa na regressão logística não apareceram na árvore de decisão, gerando assim uma curva ROC com a AUC=0.702, um valor um pouco menor do que o da regressão logística. A Figura 4 ilustra a Curva ROC gerada pela árvore de decisão:

#### C. Indução de Regras (JRip)

O JRip foi particularmente eficiente em extraír padrões claros para a paralisação, fornecendo regras interpretáveis que podem ser utilizadas diretamente em contextos de apoio à decisão por gestores públicos.

A decisão por JRip foi estratégica, pois ele é particularmente eficaz na criação de regras para a classe alvo minoritária. Este é um aspecto crucial do nosso dataset, que é desbalanceado. Por outro lado, como o JRip tende a ignorar a classe majoritária durante o processo de indução, optamos também por empregar o algoritmo PART (que tende a ignorar a classe minoritária em casos de datasets desbalanceados) de forma complementar, com o objetivo de gerar regras mais abrangentes para a classe "não".

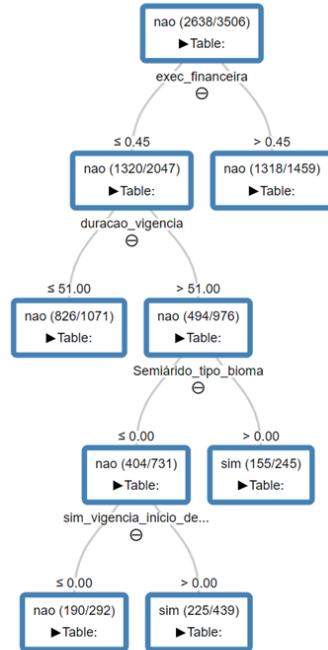


Figura 3. Árvore de Decisão

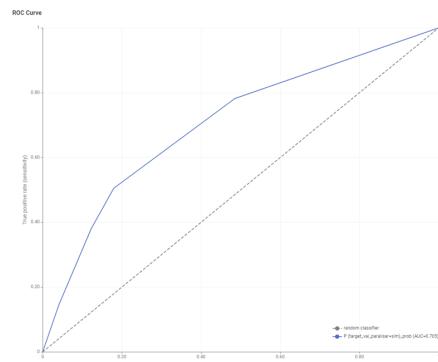


Figura 4. Curva ROC - Árvore de Decisão

A avaliação da indução de regras na amostra de teste demonstrou que o modelo classificou corretamente 907 exemplos (77.587%) e incorretamente 262 exemplos (22.412%). Outras métricas incluíram um Kappa statistic de 0.3293, Mean Absolute Error de 0.323, Root Mean Squared Error de 0.4065, e uma cobertura de casos de 100%. A Figura 5 apresenta as estatísticas geradas pelo JRip:

1) *Regras geradas:* **Regra 1:** duracao\_vigencia  $\geq 67$  AND exec\_financeira  $\geq 0.096$  AND exec\_financeira  $\leq 0.376 \Rightarrow$  "sim"

**Explicação:** SE a obra possui vigência igual ou superior a 67 meses E a execução financeira está entre 9,6% e 37,6%, ENTÃO há 59,52% de chance (confiança) de que ela venha a paralisar, pois esse perfil combina longo tempo de vigência com avanço financeiro lento, o que indica risco de estagnação.

Evaluation:			
Correctly Classified Instances	907	77.5877 %	
Incorrectly Classified Instances	262	22.4123 %	
Kappa statistic	0.3293		
Mean absolute error	0.323		
Root mean squared error	0.4065		
Coverage of cases (0.95 level)	100 %		
Total Number of Instances	1169		

Figura 5. Métricas de Avaliação do modelo JRip na amostra de teste

A regra apresenta cobertura de 7,2%. Obras com essas características têm 2,4 vezes mais (lift) chance de paralisar do que a média.

**Regra 2:**  $\text{duracao_vigencia} \geq 50 \text{ AND } \text{Semiárido_tipo_bioma} \geq 1 \text{ AND } \text{exec_financeira} \leq 0.4512 \Rightarrow \text{"sim"}$

**Explicação:** SE a obra possui vigência igual ou superior a 50 meses, está localizada no bioma Semiárido E a execução financeira é inferior ou igual a 45,12%, ENTÃO há 60,24% de chance (confiança) de que ela venha a paralisar, pois esse perfil representa obras com prazo razoavelmente longo, inseridas em uma região com histórico de vulnerabilidades operacionais, e com execução financeira ainda limitada — o que pode refletir dificuldades estruturais ou de gestão. A regra apresenta cobertura de 7,1%. Obras com essas características têm 2,44 vezes mais (lift) chance de paralisar do que a média.

**Regra 3:**  $\text{exec_financeira} \leq 0.2744 \text{ AND } \text{duracao_vigencia} \geq 50 \text{ AND } \text{vigencia_inicio_dezembro} \geq 1 \text{ AND } \text{investimento_mensal_medio} \leq 95934.84 \Rightarrow \text{"sim"}$

**Explicação:** SE a obra possui execução financeira igual ou inferior a 27,44%, vigência igual ou superior a 50 meses, início previsto para o mês de dezembro E o investimento mensal médio é igual ou inferior a R\$ 95.934,84, ENTÃO há 59% de chance (confiança) de que ela venha a paralisar, pois esse perfil reúne obras com baixo avanço financeiro, vigência prolongada, início no final do ano fiscal — o que dificulta o uso imediato dos recursos — e valores de investimento mensal que não indicam uma pressão iminente para finalização, como ocorre em projetos com recursos muito altos e prazos curtos. A ausência dessa pressão pode levar à despriorização ou inércia na execução. A regra apresenta cobertura de 10%. Obras com essas características têm 2,39 vezes mais (lift) chance de paralisar do que a média.

**Regra 4:**  $\text{exec_financeira} > 0.4619 \Rightarrow \text{"nao"}$

**Explicação:** SE a obra possui execução financeira superior a 46,19%, ENTÃO há 90% de chance (confiança) de que ela não venha a paralisar, pois esse nível de execução indica que a obra já avançou significativamente no aspecto financeiro, o que geralmente reflete estabilidade contratual, prioridade institucional e menor risco de interrupção. A regra apresenta cobertura de 40%. Obras com essas características têm 1,2 vezes mais (lift) chance de não paralisar do que a média.

**Regra 5:**  $\text{duracao_vigencia} < 51 \text{ AND } \text{exec_financeira} \leq 0.4619 \Rightarrow \text{"nao"}$

**Explicação:** SE a obra possui vigência inferior a 51 meses E a execução financeira é igual ou inferior a 46,19%, ENTÃO há 78,5% de chance (confiança) de que ela não venha a paralisar, pois esse perfil representa obras com prazos curtos e execução financeira ainda em estágio inicial ou intermediário, mas que, por sua curta duração, têm maior chance de serem finalizadas rapidamente mesmo com pouco avanço atual. Essa regra complementa a anterior (Regra 4), cobrindo a faixa restante de execução financeira. A regra apresenta cobertura de 30%. Obras com essas características têm 1,04 vezes mais (lift) chance de não paralisar do que a média.

A Figura 6 ilustra os valores compactados das métricas de cada regra gerada:

Regra String	Confiança Number (double)	Cobertura Number (double)	Lift Number (double)
Regra_1_sim	0.595	0.072	2.408
Regra_2_sim	0.602	0.071	2.437
Regra_3_sim	0.59	0.1	2.386
Regra_4_nao	0.9	0.402	1.196
Regra_5_nao	0.785	0.299	1.043

Figura 6. Métricas das Regras Induzidas

#### D. Comparação entre Modelos

A Tabela I apresenta a comparação dos modelos focando na classe minoritária "sim":

Tabela I  
COMPARAÇÃO DE DESEMPENHO DOS MODELOS

Métrica	Reg. Logística	Árvore de Decisão	JRip
Recall (Sensibilidade)	<b>0,415</b>	0,381	0,394
Precisão	0,504	0,505	<b>0,567</b>
F1-score	0,455	0,434	<b>0,465</b>
Cohen's Kappa	0,299	0,281	<b>0,329</b>

O JRip superou os outros dois modelos em praticamente todas as métricas, com destaque para:

- Precisão (0,567): menor quantidade de falsos positivos ao prever paralisação.
- F1-score (0,465): equilíbrio entre recall e precisão.
- Kappa (0,329): melhor consistência do modelo ao lidar com desbalanceamento.

A Regressão Logística teve o melhor desempenho em recall entre os modelos clássicos (0,415), importante em contextos onde o custo de não prever o risco é alto. A Árvore de Decisão ficou um pouco abaixo nos três critérios principais.

#### IV. CONCLUSÃO

Este projeto demonstrou ser possível desenvolver um modelo preditivo eficaz para antecipar a paralisação de obras públicas antes do início da execução física, utilizando mineração de dados e técnicas de indução de regras. As variáveis duração da vigência e execução financeira foram consistentemente identificadas como os principais fatores associados ao risco de

paralisação, tanto pela regressão logística quanto pela árvore de decisão e pelas regras geradas via JRip.

A regressão logística se destacou pela robustez e interpretabilidade, principalmente no tratamento de variáveis contínuas. A árvore de decisão contribuiu para a visualização das decisões em níveis, embora com desempenho levemente inferior. O algoritmo JRip, por sua vez, mostrou-se o mais eficaz na geração de regras voltadas para a classe minoritária (obras com risco de paralisação), fornecendo insights comprehensíveis e acionáveis.

Além disso, foi adotado um ajuste no limiar de classificação da regressão logística (de 0,5 para 0,4), justificando-se pela importância de priorizar a sensibilidade do modelo, dado que o custo de não alertar sobre um risco real de paralisação é maior do que o de gerar um alerta falso.

Por fim, ressalta-se que o modelo ainda possui margem para evolução. Com uma base de dados maior e mais atualizada, seria possível treinar modelos ainda mais robustos e com maior poder de generalização. No entanto, devido à limitação na disponibilidade pública de informações históricas completas sobre a execução das obras em diferentes períodos, o trabalho aqui apresentado representa, até o momento, o melhor aproveitamento dos dados acessíveis pesquisados.

#### REFERÊNCIAS

- [1] Base de Dados do TCU. Disponível em: <https://paineis.tcu.gov.br/pub/?workspaceId=8bfbd0cc-f2cd-4e1c-8cde-6abfdffea6a8&reportId=013930b6-b989-41c3-bf00-085dc65109de>
- [2] Base de Dados sobre os convênios do Portal da Transparência. Disponível em: <https://portaldatransparencia.gov.br/download-de-dados/convenios>
- [3] Base de Dados do IBGE. Disponível em: [https://ftp.ibge.gov.br/Pib\\_Municípios/2021/base/](https://ftp.ibge.gov.br/Pib_Municípios/2021/base/)
- [4] Base de Dados do IDCS. Disponível em: <https://idsc.cidadessustentaveis.org.br/rankings/>
- [5] Base de Dados do IGM-CFA. Disponível em: <https://igm.cfa.org.br/bi/>