

Suporte à tomada de decisões em gestão pública municipal por meio de Mineração de Processos reativa e preditiva

Rafael C. Xavier¹, Luiz F. P. Southier¹, Cinglair A. Capello¹,
Edson E. Scalabrin², Jair J. Ferronato³, Marcelo Teixeira¹

¹Universidade Tecnológica Federal do Paraná (UTFPR) Pato Branco

²Pontifícia Universidade Católica do Paraná (PUCPR)

³Instituto Federal de Santa Catarina (IFSC)

Abstract. *In this article, we integrate Process Mining algorithms to enhance decision-making using data from a municipal government. The study aims to predict the time required to complete administrative processes and identify the causes of delays, thus revealing inconsistencies and bottlenecks in the processes and supporting decisions by the specialist. A case study was conducted on real data, and it shows how to identify critical sectors and propose improvements to optimize resource allocation and investments, increasing the efficiency of municipal public services.*

Resumo. *Este artigo apresenta uma abordagem que integra algoritmos de Mineração de Processos para potencializar o grau decisório na gestão de prefeituras. O estudo visa prever o tempo de conclusão de processos administrativos e identificar suas causas de atrasos, servindo assim para revelar inconsistências e gargalos nos processos e subsidiar decisões pelo especialista. Um estudo de caso foi conduzido sobre dados reais e mostrou como identificar setores críticos e propor melhorias para otimizar a alocação de recursos e evitar desperdícios, aumentando a eficiência dos serviços públicos municipais.*

1. Introdução

O setor público, assim como outras áreas, vive a era do *big data*: dados são gerados em abundância pelos sistemas transacionais em múltiplas áreas da administração [Lopes 2019]. Contudo, no contexto do setor público municipal, essa riqueza de dados raramente se converte em inteligência de gestão. Isso ocorre porque os dados são fragmentados por secretarias, armazenados em sistemas heterogêneos e pouco interoperáveis. Para uma visão completa da realidade municipal, o gestor precisa recorrer individualmente a cada secretaria e interpretar os dados, o que inviabiliza análises holísticas, dificulta a construção de decisões estratégicas e impede a identificação eficiente de gargalos e oportunidades [Lupi 2019].

Sem suporte adequado, decisões são frequentemente pautadas em conhecimento empírico, ficando sujeitas à ineficiências e desgastes institucionais pelos riscos de mau uso de recursos, má alocação de pessoal e atendimento precário à população. Um dos exemplos que expõe a necessidade de suporte decisório automatizado vem dos fluxos de protocolos. Eles são originados pelos contribuintes e tramitam entre os setores administrativos. Compreender esse fluxo é fundamental para o balanceamento da força de trabalho e para evitar atrasos nas respostas, prioridade para a administração municipal.

A literatura apresenta diversas estratégias para a modelagem de fluxos, como por meio de redes de Petri [Jensen and Kristensen 2009], máquinas de estados [Wang 2019], *Business Process Model Notation* (BPMN) [Naufalia et al. 2021] e modelos preditivos [Yang et al. 2022]. Essas abordagens, no entanto, dependem de modelagem, que é uma tarefa manual de engenharia não raramente complexa. A Mineração de Processos (MP) surge como alternativa para automatizar a descoberta desses modelos. Bastante difundida nos setores de saúde e negócios [Riz et al. 2016], a MP utiliza dados transacionais para extrair automaticamente modelos formais de processos, que podem ser analisados para identificação de gargalos, padrões, tempos, inconsistências etc. Apesar do potencial aparente, aplicações de MP no setor público ainda são escassas [Melan et al. 2020].

Usando MP, estudos mostram aplicações promissoras no setor público. A estratégia em [Perceptive Software 2014] foca na redução de despesas departamentais, enquanto [Arturo Martínez Escobar 2020] mostra como eliminar gargalos em processos fiscais, tornando a gestão mais ágil e transparente. No Brasil, há iniciativas semelhantes voltadas à descentralização normativa e à busca por maior eficiência [Henrique Pais da Costa 2019]. Porém, esses trabalhos, em geral, se limitam à revelação do modelo e a análises estatísticas convencionais, suficientes para decisões superficiais. No entanto, a combinação da MP com técnicas mais avançadas de análise amplia o potencial informacional dos modelos [dos Santos Garcia et al. 2019]. Destacam-se, nesse sentido, análises reativas, que usam o histórico para identificar causas de problemas, e preditivas, que projetam o comportamento futuro do processo.

Diversos métodos podem ser aplicados nesse tipo de análise; neste trabalho, optou-se pela Análise de Causa Raiz (ACR) [Southier et al. 2023, Teixeira and Cassiani 2010] e pela análise preditiva [Park et al. 2022] devido ao seu caráter complementar. A ACR identifica causas estruturais de falhas e gargalos, enquanto a análise preditiva antecipa riscos e atrasos, ampliando o potencial de diagnóstico e tomada de decisão em comparação a métodos apenas descritivos.

As duas abordagens foram aplicadas e validadas em um processo administrativo real de uma prefeitura, com foco na previsão do tempo de conclusão de protocolos e na identificação das causas de atrasos. Os resultados, confirmados por especialistas, mostraram alta correspondência entre os achados automatizados e a percepção dos analistas.

Estruturalmente, o artigo está organizado da seguinte forma: a Seção 2 aborda conceitos de MP, *log* de eventos e descreve os métodos de predição e ACR; a Seção 3 apresenta a proposta principal; e a Seção 4 apresenta algumas conclusões da pesquisa.

2. Referencial Teórico

A MP é uma abordagem situada entre a Mineração de Dados e a Análise de Processos que visa descobrir, monitorar e melhorar processos reais, extraindo conhecimento de *logs* de eventos construídos a partir de dados brutos armazenados por sistemas de informação [Van Der Aalst 2016]. A partir desses *logs*, a MP descobre modelos formais que descrevem como os processos realmente acontecem, a partir dos quais se pode identificar gargalos, ineficiências, detectar desvios de padrão, prever comportamentos, etc.

Um *log* pode ser compreendido como uma estrutura que centraliza o registro de vários eventos de um processo. Assume-se que cada evento deve conter, no mínimo, três

atributos: (i) a atividade executada no processo; (ii) o *timestamp*, que indica o momento exato da ocorrência do evento; (iii) e um identificador que associa a atividade a um caso específico (instância do processo) [Van Der Aalst 2012].

Com base no *log* é possível aplicar técnicas de descoberta de processos para gerar um modelo formal. Algoritmos como o α -*algorithm* [van der Aalst et al. 2004], *heuristic miner* [Weijters and Ribeiro 2011] e *genetic miner* [de Medeiros et al. 2007] são comumente utilizados para esse fim. O modelo representa, de forma gráfica e estruturada, o fluxo de execução das atividades, incluindo a ordem das atividades e suas relações causais, aspectos muitas vezes não evidentes com os dados na forma textual.

Uma vez calculado o modelo de processo, este pode ser representado em diversas notações formais e semi-formais, tais como Redes de Petri [Jensen and Kristensen 2009], BPMN [Jensen and Kristensen 2009], *Event-driven Process Chain* (EPC) [Scheer 2012], Redes de Fluxo de Trabalho [Van der Aalst 1998] ou Árvores de Processo. O modelo pode então ser analisado computacionalmente, com o objetivo de melhorar, corrigir ou expandir o processo já existente, aproveitando-se dos dados coletados do próprio sistema em operação [Van Der Aalst 2016]. Técnicas de análise usam como entrada os registros de eventos e o modelo de processo para produzir uma versão aprimorada do modelo.

Neste trabalho, a etapa de melhoria será guiada pela aplicação da ACR e da Predição. Para isso, a ACR utiliza técnicas de aprendizado supervisionado, como Árvores de Decisão [Rodriguez 2016]. Além disso, medidas de associação são aplicadas para validar as regras extraídas pelas Árvores de Decisão, assegurando a consistência das inferências geradas [Lehto et al. 2017]. Entre essas medidas, destaca-se o suporte (*sup*), que indica a frequência com que uma regra ocorre no *log*; a confiança (*conf*), que avalia a robustez da relação entre os atributos analisados; e o *lift*, que mede o grau de interdependência entre os itens, permitindo identificar associações estatisticamente relevantes.

Paralelamente, técnicas de predição podem ser aplicadas para estimar a duração restante dos processos e avaliar o impacto das ações corretivas. Neste trabalho, a predição implementada utiliza Sistemas de Transição (ST) anotados com informações temporais [Van der Aalst et al. 2011], que permitem prever o tempo de execução com base no estado atual do processo. A predição é realizada a partir de traços parciais mapeados a estados específicos do ST, utilizando o histórico de execuções anteriores como base comparativa para estimar o tempo restante até a conclusão [Van Der Aalst 2016].

3. Abordagem Proposta

Esta pesquisa se inspira em uma metodologia de MP chamada *Process Mining Project Methodology* (PM²) [Van Eck et al. 2015]. A PM² visa sistematizar um conjunto de passos que, quando bem conduzidos, tendem a culminar em melhorias efetivas no processo. A versão da PM² adaptada para este artigo é mostrada na Figura 1.

A metodologia se estrutura em seis fases distintas, cada uma com entradas, procedimentos e saídas bem definidas. Nas subseções a seguir, detalham-se as atividades realizadas em cada etapa, bem como os resultados obtidos no contexto deste projeto.

3.1. Planejamento

A fase inicial consiste na formação da equipe responsável pelo projeto de MP, composta por, no mínimo, um proprietário do processo, um especialista da área e um analista de

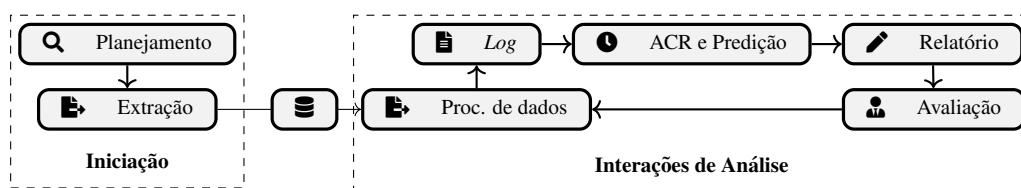


Figura 1. Versão da PM² adaptada para a proposta desse artigo.

processos. Juntos, esses atores definem a questão de pesquisa, que irá orientar as etapas subsequentes, de extração de dados e as análises a serem conduzidas.

Esse estudo analisou a estrutura organizacional real de uma prefeitura conveniada. Por meio de um acordo formal de cooperação¹, obteve-se acesso às bases de dados e à permissão de uso anonimizado. A equipe foi constituída pela prefeitura, no papel de proprietária detentora do processo, um especialista da prefeitura com domínio pleno da tramitação de processos e a equipe do Grupo de Pesquisas em Processamento de Informação: Engenharia de Dados e de Processos, da UTFPR, no papel de analistas de dados. Em conjunto, a equipe chegou à seguinte questão de pesquisa: “Os processos internos da prefeitura estão dentro do prazo de 20 dias?”.

O prazo de 20 dias foi definido com base na Lei nº 12.527/2011, que define um período legal para o atendimento de solicitações de informações [Portal Gov.br 2023]. Embora esta lei seja aplicável às esferas federais e o município não possua uma legislação específica que estipule prazos legais para as tarefas, é importante utilizá-la como referência.

3.2. Extração de dados

Com a equipe formada, a etapa seguinte consiste na definição do escopo dos dados e sua posterior extração das bases. Essa fase abrange a definição da granularidade dos dados, do período de análise, dos atributos relevantes e das possíveis correlações entre os dados. Ressalta-se que, no caso de prefeituras, os dados são distribuídos em diferentes fontes, em geral uma para cada secretaria, e a fragmentação de backup é, em geral, anual, o que se converte em um esforço não trivial de extração e preparação. Além disso, as bases de dados são heterogêneas, multi-plataformas, incluem páginas web, e-mails e documentos semi ou desestruturados. Assim, se mostrou fundamental que o especialista transmitisse aos analistas informações sobre o contexto e a estrutura dos dados.

Neste estudo, a análise foi direcionada aos dados extraídos da plataforma 1DOC [Softplan 2023], de acesso pelo especialista da prefeitura. Foram obtidos aproximadamente 167 mil registros distribuídos em 12 tabelas, com informações fragmentadas sobre os processos administrativos. Após a extração dos dados, a equipe concentrou-se em compreender a estrutura e aplicabilidade desses dados. Em conjunto com o especialista, foram definidos os processos de interesse, como os protocolos, que apresentam início e fim claramente definidos, enquanto outros, como circulares, foram descartados por sua natureza contínua e indefinida.

Esta fase também incluiu a definição da granularidade dos eventos, limitada ao nível de transição entre setores da prefeitura. Não foi possível obter informações detalha-

¹Processo 23064.051783/2022-64.

das sobre as atividades internas realizadas dentro de cada setor, como quem realizou. Ou seja, com os dados existentes, da forma como estão sendo lançados, a visão de análise é, no máximo, por setor. Porém, em caso de modificação mais granular da forma como a prefeitura registra sua tramitação, a análise automaticamente pode ser transferida para a internalização dos setores e, conseqüentemente, para análises mais pontuais.

3.3. Processamento de Dados

Na etapa de processamento, os dados extraídos são organizados e convertidos em um *log* de eventos, que serve como base para a geração do modelo de processo. Esse modelo permite uma análise preliminar, voltada a confirmar se as informações extraídas condizem com a realidade operacional da prefeitura.

A construção do *log* foi fundamentada em planilhas fornecidas pela prefeitura, extraídas por consultas às bases transacionais. Em conjunto com o especialista, foi realizada uma análise exploratória para limpeza e seleção de atributos relevantes. Os principais atributos extraídos foram: **Case_id** (identificador do processo), **Setor**, **Sigla**, **Timestamp**, **Fluxo** e **Assunto**. A Tabela 1 apresenta um recorte desse *log*.

Tabela 1. Recorte do log de eventos.

Case_id	Setor	Sigla	Timestamp	Tipo de Fluxo	Assunto
63	Setor de Fiscalização	SEFISC	2021-06-23 14:52:25	Memorando	Cartazes dentro da Faixa de Domínio da RFFSA
63	Setor de Fiscalização	SEFISC	2021-07-08 13:37:10	Memorando	Cartazes dentro da Faixa de Domínio da RFFSA
63	Setor de Fiscalização	SEFISC	2021-07-08 15:29:54	Memorando	Cartazes dentro da Faixa de Domínio da RFFSA
63	Setor de Eng. e Arquitetura	SEA	2021-07-12 11:34:55	Memorando	Cartazes dentro da Faixa de Domínio da RFFSA
63	Setor de Fiscalização	SEFISC	2021-07-12 15:40:41	Memorando	Cartazes dentro da Faixa de Domínio da RFFSA
63	Obras e Posturas	SFOP	2021-10-26 12:04:10	Memorando	Cartazes dentro da Faixa de Domínio da RFFSA
63	Procuradoria Municipal	PMGER	2021-10-26 12:12:09	Memorando	Cartazes dentro da Faixa de Domínio da RFFSA
63	Setor de Fiscalização	SEFISC	2021-10-26 14:21:27	Memorando	Cartazes dentro da Faixa de Domínio da RFFSA
63	Obras e Posturas	SFOP	2021-10-26 16:21:21	Memorando	Cartazes dentro da Faixa de Domínio da RFFSA
63	Secretaria de Planejamento	SMP	2021-10-28 10:44:18	Memorando	Cartazes dentro da Faixa de Domínio da RFFSA
63	Obras e Posturas	SFOP	2021-11-23 17:04:20	Memorando	Cartazes dentro da Faixa de Domínio da RFFSA
72	Consultoria Tributária	CT	2021-06-23 15:24:56	Proc. Administrativo	Adesivagem - Vidraça interna - Engenharia
72	Orçamentos	DC	2021-07-01 13:55:23	Proc. Administrativo	Adesivagem - Vidraça interna - Engenharia
72	Consultoria Tributária	CT	2021-07-01 15:19:34	Proc. Administrativo	Adesivagem - Vidraça interna - Engenharia
...

Em seguida, utilizou-se a plataforma Disco [Fluxicon 2023] para gerar o modelo de processo a partir do *log* construído. Durante a consulta com o especialista, foi constatado que a prefeitura desconhecia a estrutura do fluxo processual e não tinha visibilidade sobre quais setores apresentavam os maiores índices de atraso.

No modelo gerado, cada setor da prefeitura é representado por um nó (caixa), enquanto as conexões entre eles (setas) indicam as transições de processos. A intensidade das setas reflete o tempo médio de tramitação entre os setores: quanto mais espessa a seta, maior o tempo médio registrado.

Mesmo na visualização inicial [Xavier 2025], já é possível identificar gargalos no fluxo processual. Um dos principais ocorre no Departamento de Atendimento ao Cidadão, como mostra a Figura 2. Foram observados tempos médios de 17 e 15 dias em determinadas interações. Considerando-se que o cidadão é o destinatário final do serviço e que o prazo de referência é de 20 dias, esses valores indicam a possibilidade real de extrapolação do limite legal.

Outros setores, como Tesouraria e Desenvolvimento Humano e Ocupacional, também apresentam tempos médios elevados. Em particular, o tempo de tramitação no setor de Tesouraria chega a ser o dobro do tempo de referência, conforme ilustra a Figura

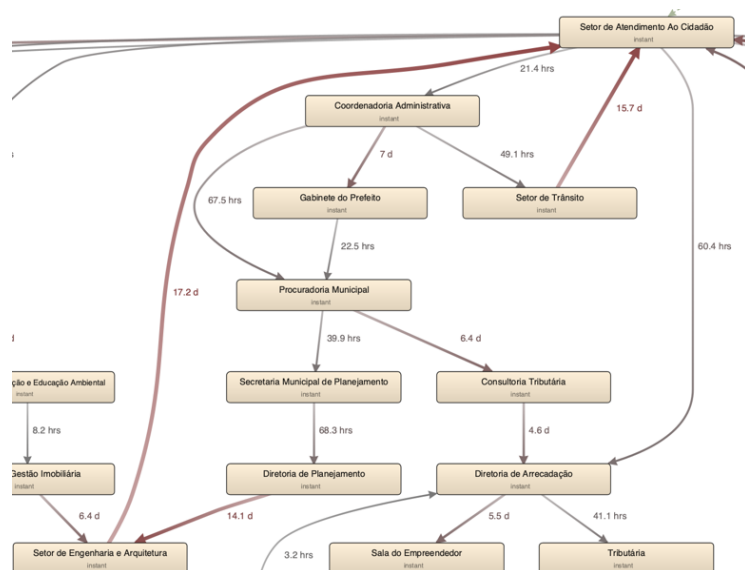


Figura 2. Recorte do Modelo: Departamento de Atendimento ao Cidadão.

3. Esses achados reforçam a relevância da análise para identificar e validar os principais pontos de ineficiência (ou de sobrecarga).

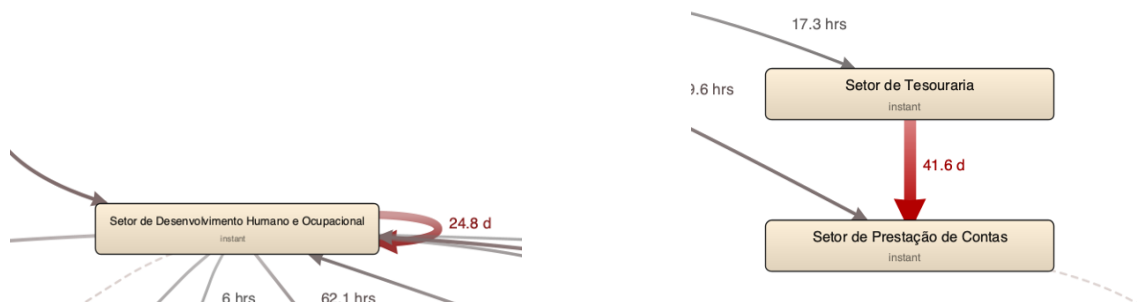


Figura 3. Recorte do Modelo: Tesouraria e Desenvolvimento Humano.

3.4. ACR e Predição

Nesta fase, realiza-se a análise dos modelos extraídos a partir do *log* de eventos, com o objetivo de identificar padrões, gargalos e desvios nos fluxos processuais. Com base nesse mesmo *log*, são aplicadas as técnicas propostas para identificar oportunidades de melhoria nos processos administrativos.

Em consulta com o especialista e com base na Lei nº 12.527/2011, estabeleceu-se a expectativa de que todos os processos deveriam ser concluídos em até 20 dias. No entanto, a aplicação das técnicas de MP revelou que 11% dos 11.300 protocolos analisados excederam esse limite, evidenciando uma discrepância entre a expectativa normativa e a execução prática dos processos. A partir dessa constatação, aplicou-se a ACR para investigar as causas dos atrasos identificados.

A ACR foi conduzida com base no *log* de eventos, por meio da geração de uma Árvore de Decisão com profundidade limitada a dois níveis, de forma a identificar as

variáveis mais relevantes associadas aos atrasos. A variável resposta utilizada foi a “Duração”, categorizada entre curta e longa, com o intuito de modelar a relação entre os setores envolvidos e o tempo de execução. Os resultados obtidos estão apresentados na Tabela 2, destacando os setores e atividades mais associados aos atrasos.

Tabela 2. Métricas para as regras descobertas

A			B	sup	conf	lift
S. Processo Seletivo			Longa	1.43	71.64	0.10
Sem S. Processo Seletivo	S. Controle Interno		Longa	0.78	71.03	0.10
	Sem S. Controle Interno	Prestação de Contas	Longa	0.67	47.10	0.07
		Sem Prestação de Contas	Curta	92.02	93.34	0.01

A Tabela 2 ilustra que o setor mais responsável pela demora na resposta do atendimento é o “Setor de Processo Seletivo”, seguido pelo “Setor de Controle Interno” e, posteriormente, pela atividade “Prestação de Conta” desse mesmo setor. Então, procedeu-se com a aplicação das medidas de associação para a validação da Árvore de Decisão encontrada, resultando nas métricas mostradas na Tabela 2. Para a regra $A \rightarrow B$, associada aos diversos setores e atividade, foram identificados suportes abaixo de 1.5%, para longa duração. Isso se deve ao fato de que considerou-se como longa duração processos que tenham demorado um desvio-padrão acima da média de duração de todos os processos.

Além disso, as regras apresentam uma confiança de 71% para os setores e 47% para a atividade, ou seja, 71% das vezes que ocorre o Setor A , houve duração “Longa” e a atividade é uma porcentagem próxima a 50%. Adicionalmente, foi calculado um *lift* de aproximadamente 0.1% na maioria dos casos, o que evidencia que as variáveis A e B não são independentes.

Após a identificação dos principais pontos de atraso nos processos, foi treinado um modelo preditivo com o objetivo de estimar o tempo restante para a conclusão dos processos em aberto. Além disso, o modelo permite avaliar a probabilidade de os setores responsáveis pelos atrasos manterem esse padrão de comportamento ao longo do tempo.

Para avaliar a precisão da predição, é possível percorrer parcialmente cada traço, avançando em intervalos de 10% do caminho total e aplicar esse mesmo avanço no ST. Com base na média e desvio-padrão da anotação temporal do estado do ST, é possível estimar o tempo necessário para a conclusão do processo e, comparando com o tempo restante real, calcular o erro associado. A Figura 4 mostra a média e desvio padrão do erro quadrático da predição comparado com o valor real, em relação à porcentagem de amostragem do processo em avaliação.

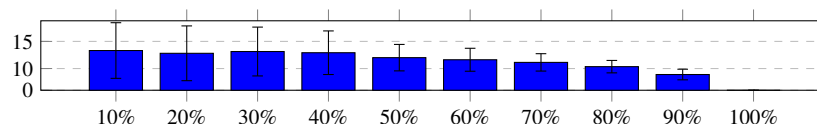


Figura 4. Média e desvio padrão do erro (dias) por porcentagem de amostragem

Ao aplicar esta técnica, verificou-se que, a taxa de acerto do modelo preditivo mostrou ser mais precisa para processos com menor variabilidade nos tempos de execução, como documentação e certidões, enquanto processos mais variados devido a sua natureza, como recursos humanos e folha de pagamento, apresentaram maior variabilidade e taxas

de erro mais elevadas, e na maioria dos casos, as taxas de erro e os desvios padrão de predição ao conforme avançamos no traço diminuíram, indicando processos mais consistentes e previsíveis. Essa estabilidade pode ser atribuída a rotinas mais padronizadas, menor influência de variáveis externas nos prazos e procedimentos bem definidos, o que facilita a gestão e a execução eficiente.

3.5. Avaliação

Nesta fase, realiza-se um diagnóstico dos resultados obtidos, interpretando-os e distinguindo comportamentos anormais ou inesperados dos padrões normais ou esperados. Além disso, avaliam-se possíveis novas questões de pesquisa que possam surgir a partir desses mesmos dados.

Para assegurar a aplicabilidade prática dos resultados de ambas as técnicas, o especialista do processo realizou uma validação final. O especialista contrastou as análises com o cenário real da prefeitura e validou os padrões de fluxo observados. Ou seja, as análises refletem os padrões e tendências esperados nos processos municipais.

4. Conclusões e perspectivas

A abordagem de MP apresentada neste artigo integra as análises reativa e preditiva e se aplica ao gerenciamento de setores administrativos que estejam sujeitos a atrasos na condução dos seus processos. Tal gerenciamento se aplica tanto no contexto passado, para identificar locais e causas de atrasos, quanto no contexto futuro, para estimar o tempo de conclusão de trâmites. Em conjunto, essas abordagens permitem balancear a alocação de recursos (humanos e financeiros) e melhor atender ao contribuinte, apesar de desafios devido à variabilidade e heterogeneidade de contextos nos quais estão imersos os processos.

Ao aplicar a abordagem nos dados reais de uma prefeitura, identificou-se que aproximadamente 11% dos protocolos excederam o prazo de 20 dias, indicando ineficiências significativas. A ACR permitiu identificar setores específicos como principais responsáveis pelos atrasos. Este método revelou causas subjacentes dos problemas, possibilitando ações corretivas focadas.

Por outro lado, a abordagem preditiva, utilizando modelos baseados em ST, proporcionou previsões sobre o tempo de conclusão dos processos com uma precisão maior para aqueles com menos variabilidade, como documentação e certidões, que atingiram um erro médio de menos de 2 semanas, mesmo para os piores casos.

A integração das abordagens se mostrou adequada para uma visão mais completa dos processos. Além disso, a validação das análises por especialistas do processo reforçou a confiabilidade dos achados. Dessa forma, a abordagem desponta como promissora para aprimorar a tomada de decisões e otimizar a alocação de recursos no setor público. Para trabalhos futuros, prevê-se a aplicação das mesmas técnicas em outras prefeituras, com o objetivo de avaliar a recorrência dos padrões identificados. Caso necessário, poderão ser incorporados métodos complementares, a fim de aprimorar a profundidade e a qualidade das análises conduzidas.

Agradecimento

Este trabalho foi apoiado por CNPq, CAPES, FINEP, Fundação Araucária e UTFPR.

Referências

- Arturo Martínez Escobar, Nemury Silega Martínez, M. N. (2020). Analyzing the complaints process at granada city council. <https://tinyurl.com/4d635k75>.
- de Medeiros, A. K. A., Weijters, A. J., and van der Aalst, W. M. (2007). Genetic process mining: an experimental evaluation. *Data mining and knowledge discovery*, 14:245–304.
- dos Santos Garcia, C., Meincheim, A., Junior, E. R. F., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., Santos, E. A. P., and Scalabrin, E. E. (2019). Process mining techniques and applications—a systematic mapping study. *Expert Systems with Applications*, 133:260–295.
- Fluxicon (2023). Disco. <https://fluxicon.com/disco>.
- Henrique Pais da Costa, Julian Marcondes Viana de Assis, C. C. d. V. (2019). Government process mining in the brazilian executive branch. <https://tinyurl.com/53ddkm9j>.
- Jensen, K. and Kristensen, L. M. (2009). *Coloured Petri nets: modelling and validation of concurrent systems*. Springer.
- Lehto, T., Hinkka, M., and Hollmén, J. (2017). Focusing business process lead time improvements using influence analysis. In *Int. Symposium on Data-Driven Process Discovery and Analysis*, pages 54–67. CEUR.
- Lopes, M. A. (2019). *Aplicação de aprendizado de máquina na detecção de fraudes públicas*. PhD thesis, Universidade de São Paulo.
- Lupi, L. (2019). City data plan: The conceptualisation of a policy instrument for data governance in smart cities. *Urban Science*, 3(3):91.
- Melan, R. L., da Silva, H. F., Amâncio-Vieira, S. F., and Baccarin, E. (2020). Mineração de processos em um órgão público: estudo da diretoria de loteamentos da prefeitura municipal de londrina (PR). *Revista Inovação, Projetos e Tecnologias*, 8(2):221–234.
- Naufalia, R., Usman, S. A., and Bambang, C. L. (2021). Analysis and development of company business processes using business process model notation. *Journal of Soft Computing Exploration*, 2(2):135–142.
- Park, G., Küsters, A., Tews, M., Pitsch, C., Schneider, J., and van der Aalst, W. M. P. (2022). Explainable predictive decision mining for operational support. <https://arxiv.org/pdf/2210.16786.pdf>.
- Perceptive Software (2014). Process mining case study within bolton council: Discovering process inefficiencies in city government. <https://abrir.link/ZtnJP>.
- Portal Gov.br (2023). Aspectos gerais da lei de acesso à informação. <https://tinyurl.com/3busf93d>. Publicado em 13/05/2014, Atualizado em 23/05/2023.
- Riz, G., Santos, E. A. P., and Loures, E. d. F. R. (2016). Análise de conformidade na Área de saúde com o suporte da mineração de processos. In *Anais do XII Simpósio Brasileiro de Sistemas de Informação*, pages 052–059. SBC.
- Rodriguez, J. (2016). Step 6: part i—root cause analysis. *CAPA in the Pharmaceutical and Biotech Industries*.

- Scheer, A.-W. (2012). *Business process engineering: reference models for industrial enterprises*. Springer Science & Business Media.
- Softplan (2023). 1doc by softplan. <https://1doc.com.br/>. Acessado em: jul. 2025.
- Southier, L. F. P., Krugger, E. M. R., Garcia, C. D. S., and Scalabrin, E. E. (2023). Process mining and root cause analysis for detecting inefficiencies in business processes: an applied case in a brazilian telecommunications provider. In *Int. Conf. on Computer Supported Cooperative Work in Design (CSCWD)*, pages 53–58. IEEE.
- Teixeira, T. C. A. and Cassiani, S. H. D. B. (2010). Análise de causa raiz: avaliação de erros de medicação em um hospital universitário. *Revista da Escola de Enfermagem da USP*, 44:139–146.
- Van Der Aalst, W. (2012). Process mining: Overview and opportunities. *ACM Trans. on Management Information Systems*, 3(2):1–17.
- Van Der Aalst, W. (2016). *Data science in action*. Springer.
- van der Aalst, W., Weijters, T., and Maruster, L. (2004). Workflow mining: discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142.
- Van der Aalst, W. M. (1998). The application of petri nets to workflow management. *Journal of circuits, systems, and computers*, 8(01):21–66.
- Van der Aalst, W. M., Schonenberg, M. H., and Song, M. (2011). Time prediction based on process mining. *Information systems*, 36(2):450–475.
- Van Eck, M. L., Lu, X., Leemans, S. J., and Van Der Aalst, W. M. (2015). Pm: a process mining project methodology. In *Int. Conf. on advanced information systems engineering*, pages 297–313. Springer.
- Wang, J. (2019). *Formal methods in computer science*. Chap & Hall/CRC.
- Weijters, A. and Ribeiro, J. (2011). Flexible heuristics miner. In *IEEE Symp. on Comp. Intelligence and Data Mining (CIDM)*, pages 310–317.
- Xavier, R. C. (2025). Modelo completo descoberto a partir do log de eventos gerado pelos processos administrativos da prefeitura. <https://zenodo.org/records/15850934>. Publicado no Zenodo.
- Yang, Y., Xia, X., Lo, D., Bi, T., Grundy, J., and Yang, X. (2022). Predictive models in software engineering: Challenges and opportunities. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 31(3):1–72.